Севастопольський інститут банківської справи
Української академії банківської справи Національного банку України

Кафедра економічної кібернетики

ЗАТВЕРДЖУЮ
Завідувач кафедри
канд. екон. наук, доцент
_____________ С.О. Хайлук
03.01.2012

КОНСПЕКТ ЛЕКЦІЙ
з дисципліни
«Дослідження операцій»
освітньо-професійної програми підготовки бакалаврів за напрямом
6.030502 – «Економічна кібернетика»

Укладач: канд. екон. наук, доц. _________________ І.С. Кондіус
03.01.2012

Розглянуто та схвалено на засіданні кафедри, протокол від 03.01.2012 № 1

Севастополь – 2012

ПЕРЕДМОВА

Конспект лекцій розроблений відповідно до вимог робочої навчальної програми з дисципліни «Дослідження операцій» нормативної компоненти освітньо-професійної програми підготовки за напрямом 6.030502 – «Економічна кібернетика» галузі знань 0305 «Економіка і підприємництво», які розроблені Науково-методичною комісією з галузі "Економіка та підприємництво" Міністерства освіти і науки України та Київським національним економічним університетом.

Конспект лекцій містить зміст лекційного курсу, завдання до самостійного вивчення теоретичного матеріалу курсу, що вивчається в позааудиторний час, список рекомендованої літератури і ресурсів Інтернет, ілюстративний матеріал до лекцій.

ЗМІСТ

ВСТУП

Дисципліна «Дослідження операцій» належить до нормативної компоненти освітньо-професійної програми підготовки бакалаврів за напрямом 6.030502 – «Економічна кібернетика» галузі знань 0305 «Економіка і підприємництво».

Мета: формування системи теоретичних знань та практичних навичок формалізації задач управління з використанням спеціалізованих оптимізаційних методів.

Завдання: надання фундаментальних знань щодо суті та етапів дослідження операцій; основних принципів та прийомів математичного моделювання операцій; принципів підбору математичного забезпечення практичної реалізації задач;. та формування умінь: самостійного опрацювання літератури; самостійно розширювати свої знання, розвивати логічне і алгоритмічне мислення.

Предметом є методологічні й методичні засади та інструментарій кількісного аналізу та управління організаційно-економічними системами.

Зміст дисципліни розкривається в наступних темах:

1. Дослідження операцій як науковий підхід до аналізу економічних об’єктів.

2. Оптимізаційні задачі управління запасами.

3. Застосування теорії мереж в економіці.

4. Прийняття рішень в умовах невизначеності.

5. Теорія ігор та ігрове моделювання.

6. Нечіткі множини в системах керування.

Поточний контроль знань та одержаних навиків здійснюється при тестуванні, оцінці виконання лабораторної роботи та захисту лабораторної роботи, оцінці виконання домашніх завдань обов’язкового та вибіркового характеру.

Підсумковий контроль знань здійснюється у вигляді підсумкового екзамену і полягає в оцінюванні засвоєння студентами програмного матеріалу, набуття ними вміння та практичних навичок в даній предметній галузі, здатності опрацювання та письмового викладу конкретних питань дисципліни та проведення розрахунків за тематикою курсу.

На вивчення дисципліни навчальним планом передбачено 108 годин, з них 19% відведено на лекційні заняття, 30% на лабораторні заняття. На самостійне вивчення відводиться 51% часу.

Послідовність вивчення розділів дисципліни задається тематикою лекцій та завдань для самостійної роботи, що доводяться на кожній лекції. Перелік рекомендованої літератури, на яку дається посилання під час вивчення дисципліни, подано в наступній таблиці.

Таблиця 1 – Перелік рекомендованої літератури

№ пор.	Бібліографічний опис	Кількість примір.	УДК бібліотеки
Основна література			
1	Зайченко, Ю. П. Дослідження операцій [Текст]: Підручник для студ. вищих навч. закл., що навч. за напрямками "Прикладна математика" та "Комп'ютерні науки". / Ю.П. Зайченко – 4.вид., перероб. і доп. – К. : ЗАТ "ВІПОЛ", 2000. – 687с. ISBN 966-8407-64-4	1	519.8 (075.8)
2	Кутковецький, В.Я. Дослідження операцій [Текст]: підручник / В.Я. Кутковецький Миколаївський держ. гуманітарний ун-т ім. Петра Могили комплексу «Києво-Могилянська академія». – Миколаїв: Видавництво МДГУ, 2007. – Т.1 – 311 с., Т.2 – 270 с. ISBN 966-8556-04-6	1	517.93
3	Венцель, Е.С. Исследование операцій: задачи, принципы, методология [Текст]. /Е.С. Венцель. – 2-е изд., стер. – М.: Наука. 1988. – 208 с.	1	519,8

№ пор.	Бібліографічний опис	Кількість примір.	УДК бібліотеки
	ISBN 5-02-013900-9		
4	Таха Хэмди А. Введение в исследование операций [Текст] / Таха Хэмди А 7.изд.:Пер. с англ. – М.: Издательский дом «Вильямс», 2007. – 912с. ISBN 5-8459-0740-3	1	681.3.07
Додаткова література			
5	Ларіонов, Ю.І. Дослідження операцій Частина II [Текст]: Навчальний посібник /Ю.І. Ларіонов, Л.С. Манченко, М.А. Хажмурадов. – Х.: ВД «ІНЖЕК», 2005. – 288 с. ISBN 966-8515-91-9	1	681.5:658.012.012.56
6	Карагодова, О. О. Дослідження операцій [Текст]: Навч. посібник / О. О. Карагодова, В. Р. Кігель, В. Д. Рожок – К.: Центр учбової літератури, 2008 – 256с.		
7	Колемаев, В.А. Практикум по исследованию операций в экономике [Текст]: Учебное пособие для вузов. / В.А. Колемаев. – М., 2007. —192с.		
8	Кремер, Н. Ш. Исследование операций в экономике [Текст]: Учебн. пособие для вузов / Н. Ш. Кремер, Б. А. Прутко, И. М. Тришин, М. Н. Фридман: [Под ред. проф. Н. Ш. Кремера – М.: Банки и биржи. ЮНИТИ, 2003 —407 с. ISBN 5-85173-092-7	1	330.45 (075.8)

ЛЕКЦІЯ 1 «Дослідження операцій як науковий підхід до аналізу економічних об’єктів»

Анотація

Принципи застосування математики в економіці. Переваги застосування математичних методів в економіці. Поняття моделі та моделювання. Характеристики матеріальних та ідеальних моделей. Математична модель як метод дослідження економічних процесів. Класифікація математичних моделей, використовуваних в економіці. Загальні відомості про дослідження операцій. Поняття операції та дослідження операцій. Типізація задач дослідження операцій. Основні етапи дослідження операцій. Історія розвитку методів дослідження операцій. Методи ДО.

1.1 Принципи застосування математики в економіці

Використання математичних методів у економіці почалося досить давно. Перша у світі економічна модель була створена у XVIII столітті французьким економістом Ф. Кене. У XX столітті його “Економічна таблиця” послужила основою для побудови й розвитку численних моделей суспільного відтворення. Так, міжгалузева модель “Витрати–випуск” В. Леонтьєва є подальшим логічним кроком у продовження економічної таблиці Ф. Кене.

Розквіт застосування математичних методів у економіці ознаменувало XX століття. З їх використанням пов‘язанні роботи практично всіх вчених, відзначених Нобелівською премією з економіки, наприклад, Д.Хікс, Р.Солоу, В.Леонтьєв, П.Самуельсон.

Перші роботи із застосуванням математики в економіці не виходили за рамки найпростіших обробок результатів спостережень. Подальший розвиток мікро– і макроекономіки, прикладних економічних дисциплін пов‘язаний з дедалі вищим рівнем їх формалізації. Основу для цього заклав прогрес у самій математиці особливо в галузі прикладної математики.

Яку ж конкретно роль відіграють математичні методи **в** економіці? Використання їх, тією мірою, якою самі моделі адекватні об'єкту дослідження, дає можливість:

– точно і компактно викладати положення економічної теорії;

– виділяти і формально описувати найістотніші зв'язки економічних змінних і характеристик;

– одержувати висновки про функціонування об'єкта;

– отримувати нові знання про об'єкт;

– передбачати майбутню поведінку об'єкта у разі зміни якихось його параметрів.

Природно, що використання математичних методів і побудова на їх основі математичних моделей супроводжувалися розвитком відповідних понять. Різні дослідники давали своє тлумачення вже сформованим термінам. Розглянемо їх коротко.

Модель – це такий матеріальний або уявлюваний об'єкт (об'єкт-замінник), який у процесі дослідження заміщає об'єкт-оригінал так, що його безпосереднє вивчення дає нові знання про об'єкт-оригінал.

Модель потрібна для того, щоб

– зрозуміти з чого складається конкретний об‘єкт;

– навчитись керувати об‘єктом (процесом) і визначати найкращі способи управління при заданих умовах;

– прогнозувати прямі і непрямі наслідки реалізації заданих форм впливу на об‘єкт.

Процес побудови, вивчення і застосування моделей називають — *моделюванням.*

Існує декілька прийомів моделювання, які умовно можна поєднати у дві групи: матеріальне (предметне) і ідеальне:

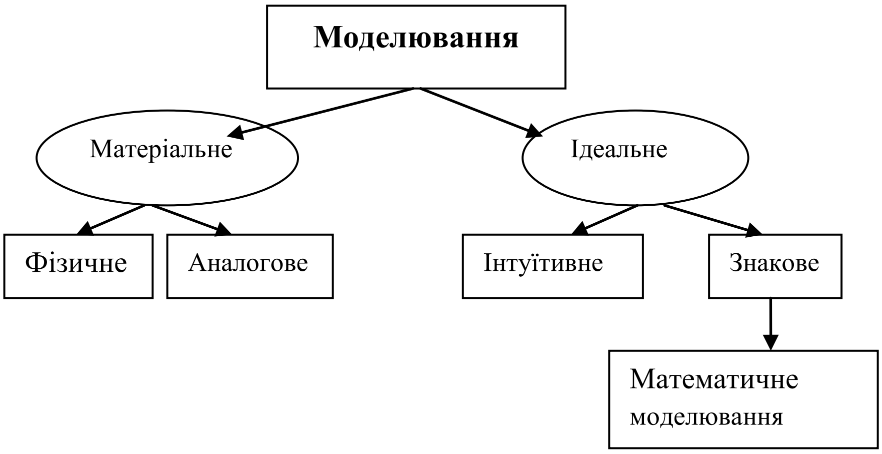


Рисунок 1.1 – Поділ прийомів моделювання

Відповідно усі моделі поділяються на два великих класи: моделі *матеріальні* і моделі *ідеальні*. Характерний представник моделей першого класу – фізичні моделі. До другого класу належать, у принципі, всі створені людиною мислені уявлення про навколишній світ, зокрема й математичні моделі.

Об'єктом дослідження математичного моделювання в економіці є економічна система.

Математична модель економічного об‘єкта (системи) – це його спрощений образ, поданий у вигляді сукупності математичних співвідношень (рівнянь, нерівностей, логічних співвідношень, графіків тощо).

У широкому сенсі *математичне моделювання* – це метод дослідження, базований на аналогії процесів різної природи, але описуваних однаковими математичними залежностями.

Необхідність використання моделювання визначається тим, що багато об'єктів і пов'язані з ними проблеми дослідити безпосередньо або зовсім неможливо, або ж їхнє дослідження вимагає так багато сил і часу, що вже з цієї причини стає неможливим.

За своїм визначенням будь-яка економічна модель абстрактна, отже, неповна, оскільки, виділяючи визначальні закономірності, вона абстрагується від інших факторів, які, незважаючи на їх відносну малість, у сукупності чи за певних умов можуть визначати не тільки відхилення в поведінці об'єкта дослідження, а й саму поведінку. Однак при цьому методі пізнання не залишається нічого іншого, як припускати, що невраховані фактори справляють на об'єкт незначний вплив, або ж уводити їх у модель і робити врахованими, якщо це можливо.

Наприклад, у найпростішій моделі попиту вважається, що попит на товар визначається його ціною і доходом споживача. Насправді ж на попит впливають і інші фактори: смаки й очікування споживачів, ціни на інші товари, реклама, мода і так далі. І іноді роль останніх буває визначальною.

Для математичних моделей, використовуваних в економіці, застосовуються різні види класифікації. Основні з них подано на рис. 1.2.

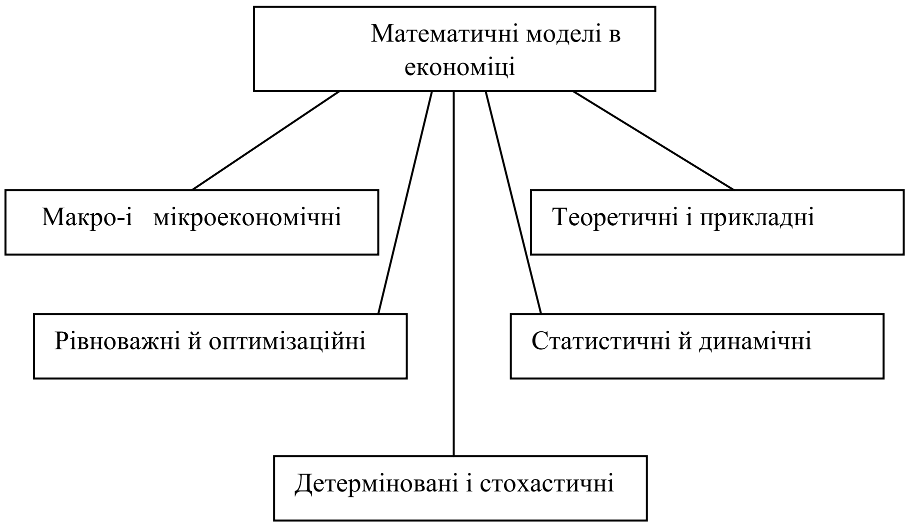


Рисунок1.2 – Класифікація математичних моделей, використовуваних в економіці

Будуючи моделі, виділяють істотні фактори й відкидають деталі, щоб виокремити спільне й суттєве для всіх принципово однакових, але таких різних у деталях явищ. Приклади економічних моделей наведено на схемі, рис. 1.3.

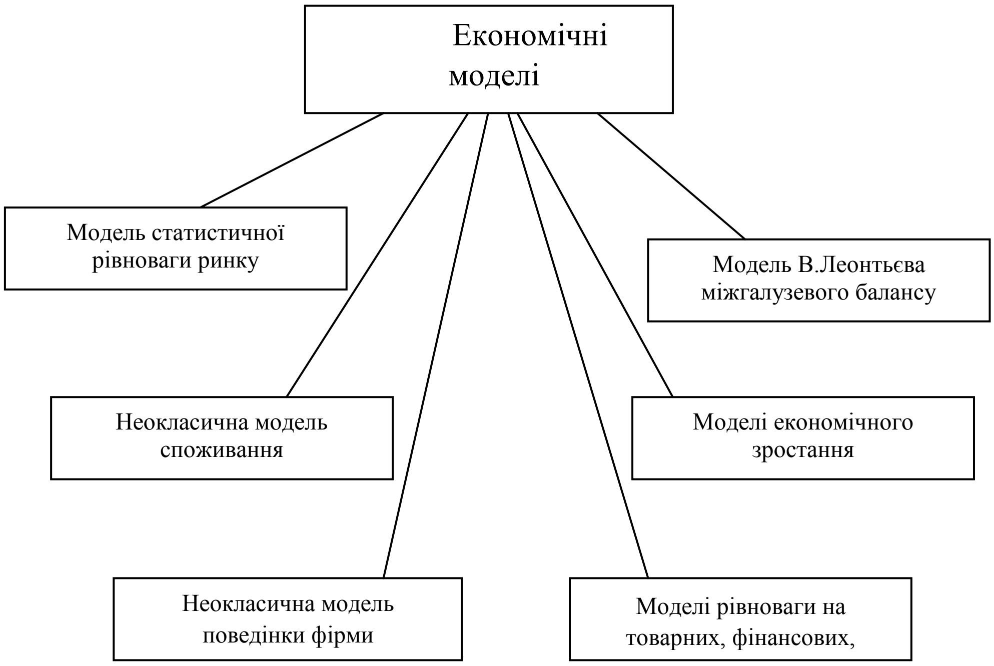


Рисунок 1.3 – Приклади економічних моделей

Проникнення математики в економіку пов'язане з подоланням значних труднощів. Головні причини лежать у специфіці економічної науки, а також у природі економічних процесів, для яких характерні масовість, динамічність і стохастичність. Крім того, більшість об'єктів, досліджуваних економічною наукою, характеризується поняттям «складна система». Складність економічних процесів часто розглядається як обґрунтування неможливості їх формалізації і моделювання засобами математики. Хибність такої точки зору доводиться самою «живучістю» моделей. Моделювати можна об'єкт будь-якої природи і складності, тому що теза про принципову неможливість моделювання рівносильна твердженню про принципову непізнаваність об'єкта.

І саме складні об'єкти становлять найбільший інтерес для моделювання. Саме тут моделювання може дати і дає результати, які не можна одержати жодними іншими способами дослідження.

У наш час моделювання набуло нової якісної форми, що тісно пов'язано з використанням можливостей, наданих комп'ютерами.

Інформатизація суспільства–закономірний процес. Наприклад, у США злам припав на 1991 рік, коли вперше витрати на придбання інформаційної техніки

(112 млрд дол.) перевищили витрати на придбання промислового обладнання (107 млрд дол.). Цей рік можна вважати першим роком інформаційної ери. Відтоді різниця між зазначеними витратами постійно збільшується. Зростання ролі знань, сучасних технологій, добування нової важливої для керування інформації притаманне інформатизованим суспільствам. Це видно на прикладі компаній "IBM" та "Microsoft". Так, наприкінці 1996 року ринкова вартість компанії "Microsoft" становила 85,5, а "IBM" – 70,7 млрд дол., хоча остання продавала набагато більше продукції. Окрім того, вартість основних виробничих засобів та устаткування "IBM" сягає 16,6 млрд дол, а "Microsoft" – не перевищує 930 млн дол. Отже, з позиції індустріального суспільства основною вартістю "Microsoft" є "повітря" – ідеї, думки, набутий працівниками досвід. Престижне ім‘я, можливості, особливо можливості перспективні а також розумні й творчі голови службовців.

Зауважимо, що в розвинутих країнах чисельність працівників, зайнятих у сфері виробництва з року в рік зменшується. Так, нині у США частка таких працівників становить приблизно 10%, а в інтелектуальній сферах зайнято вже 60%. При цьому рівень рентабельності у виробничій сфері не перевищує 5 – 15 %, а в інтелектуальній – 1000 – 2000%.

1.2 Загальні відомості про дослідження операцій. Історія розвитку методів дослідження операцій

Під *операцією* розуміють будь-яку діяльність людини, що спрямована на досягнення якоїсь мети (у виробництві, у військовій операції, у перевезенні вантажів, у плануванні робіт, у прийнятті політичного рішення та ін.).

Припустимо, що людина приймає рішення (часто дуже важливе, бо від нього залежить її доля, доля її підприємства, доля військової операції, напрям розвитку держави). Виникає питання: наскільки це рішення є правильним? Виникає потреба об'єктивної *кількісної* оцінки прийнятого рішення.

Дослідження психологів показали, що людина почуває себе невпевнено, якщо при прийнятті рішення потрібно врахувати понад 10 змінних або суперечливих факторів. Але в реальних умовах виробництва на процеси впливають сотні (а іноді й тисячі) факторів. Тому науковий підхід до кількісної оцінки прийнятого рішення за допомогою методів дослідження операцій є дуже важливим.

Дослідження операцій - це теорія використання наукових кількісних методів для прийняття найкращого рішення у різних галузях діяльності людини. Ця наука дає об'єктивні, кількісні рекомендації з управління цілеспрямованими діями людини.

Як самостійний науковий напрям, дослідження операцій оформилося на початку 40-х років минулого століття. Перші публікації з досліджень операцій з'явилися у 1939-1940 рр. А на період Другої світової війни США використовували науковців, які давали поради військовим щодо прийняття рішень при аналізі та дослідженні військових операцій. Звідси і виникла назва дисципліни.

Пізніше принципи і методи дослідження операцій (ДО) стали використовуватися у цивільній сфері: у промисловості, для управління фінансами, у сільському господарстві та ін.

Метою ДО є *наукове кількісне обґрунтування рішень*, які приймаються щодо управління в господарських, військових та державних справах. У деяких випадках (наприклад, у багатьох комбінаторних задачах) отримати оптимальне рішення неможливо, і тому приймається субоптимальне (не найгірше) рішення.

Виникає питання філософського характеру: наскільки впливають методи ДО на наше життя? Відповідь на це дає скорочений перелік питань, які вирішуються за допомогою методів ДО: плани у політиці (у Канаді та США створені так би мовити "електронні уряди"), плани розвитку народного господарства (тобто ми живемо за планами, визначеними ЕОМ), розвиток військових справ та військових операцій, фінансові справи. На перший погляд, ЕОМ у цих випадках лише "дає поради", а "рішення приймає людина". Але певною мірою це самообман, бо перевірити розв'язок машини людина може, знову ж таки, лише за допомогою іншої машини. І виходить, що доля людства залежить від розв'язку машини, затвердженого людиною.

Предметом дослідження операцій є: військові операції; рішення у політиці та виробництві, сільському господарстві, фінансових справах і т.п. Ми будемо розглядати виробничі процеси у господарській діяльності людини.

Типовими класами задач дослідження операцій є:

Розподіл ресурсів. Ресурси - це гроші, матеріали, людська праця і т.п. Ресурси завжди обмежені і в різних виробах забезпечують різний прибуток. Наприклад, ми маємо матерію, з якої можна виготовити або чоловічий, або жіночий, або дитячий одяг за різними цінами та прибутками. Виникає проблема розподілу людей, матерії та інших ресурсів між виробами з метою отримання найбільшого прибутку.

Управління запасами. Із збільшенням запасів створюються умови для більш ритмічної роботи виробництва. Запас - це гарантія можливості виконання будь-якого замовлення. Якщо запасів не вистачає, то можливі значні збитки за рахунок невиконання зобов'язань. Але разом із збільшенням запасів збільшується змертвілий капітал і витрати на зберігання. Недаремно існують підприємства, які зовсім не мають складів: їх замінюють майданчики для розвантаження отриманої та відвантаження виготовленої продукції. Виникає проблема управління запасами при найменших витратах.

Задачі мережного планування і управління розглядають співвідношення між термінами закінчення великого комплексу операцій і моментами початку всіх операцій комплексу. Потрібно знайти мінімальні тривалості комплексу операцій, оптимальні співвідношення вартості і термінів виконання.

Мережні задачі полягають у оптимізації процесу обслуговування на мережах чи самої структури мережі.

Задачі планування і розміщення пов'язані з визначенням оптимального числа і місця розміщення нових об'єктів з урахуванням їх взаємодії з наявними об'єктами і між собою.

Задачі дослідження конфліктних ситуацій полягають у виборі оптимальних стратегій поведінки учасників конфлікту.

Задачі масового обслуговування: розглядають питання створення та функціонування черг (на заводському конвеєрі; у залізничній касі; для літаків над аеропортом, що йдуть на посадку; клієнтів в ательє побутового обслуговування; абонентів міської телефонної станції тощо). Потрібно розв'язати проблеми якісного обслуговування при мінімальних витратах на обладнання.

Задачі складання розкладів (календарного планування) полягають у визначенні оптимальної черговості виконання операцій на різних видах устаткування чи при певному способі надання послуг.

Ремонт та заміна устаткування. Застаріле обладнання вимагає витрат на ремонт і має знижену продуктивність. Потрібні розрахунки для прийняття рішення щодо термінів ремонту та заміни обладнання, які забезпечують найбільший прибуток.

Задача рюкзака: рюкзак (вантажна машина, вагон, судно, літак) має обмежену вантажопідйомність. Потрібно так заповнити рюкзак, щоб отримати максимальний прибуток.

Задачі комівояжера, створення сумішей, наймання / звільнення робітників, мережевого планування робіт, порядку обробки кількох різних деталей, комбіновані задачі та ін. - усім цим займається наука "*Математичні методи дослідження операцій*".

1.3 Основні поняття дослідження операцій

Як і кожна сформована наука, дослідження операцій має свою власну систему понять. Розглянемо основні.

При цьому під *операцією* розуміється будь-який керований захід, спрямований на досягнення мети. Результат операції залежить від способу її проведення чи організації, інакше – від вибору деяких параметрів.

Будь-який вибір набору параметрів *називається рішенням. Оптимальними* вважаються ті рішення, що в обговореному заздалегідь сенсі мають переваги над іншими. Виходячи з мети цієї теорії, можна сказати, що основним завданням дослідження операцій є знаходження оптимальних рішень у рамках обраної моделі.

Модель операції – це якомога точніший опис операції за допомогою математичного апарата.

Ефективність операції – це ступінь її пристосованості до виконання поставленої мети, що кількісно виражається у вигляді цільової функції. Вибір критерію ефективності визначає практичну цінність дослідження.

У процесі формування як стратегічних, так і тактичних рішень керівник змушений брати до уваги численні, нерідко взаємосуперечливі вимоги і спиратися на складні критерії досягнення кінцевих цілей. У цих умовах для досягнення високого рівня управління йому далеко не завжди вистачає професійних знань, власного досвіду, інтуїції й організаторських здібностей у їх традиційному розумінні. Потрібні науково обґрунтовані й точні методи прийняття рішень.

Однак зауважимо, що сам реальний процес ухвалення рішення виходить за рамки науки дослідження операцій і належить до компетенції особи (частіше групи осіб), що приймає рішення (ОПР). Неодмінна присутність людини не скасовується навіть у разі повної автоматизації системи управління.

Основною особливістю дослідження операцій є побудова математичних моделей і використання для їх аналізу математичного апарата. Це насамперед означає, що хоча б деякі дані, які фігурують у формулюванні задачі, мають мати кількісне вираження. Міркування якісного характеру є своєрідним тлом для використовуваної моделі і враховуються додатково.

Основні етапи дослідження операцій:

1. Отримання змісту задачі у вигляді текстового (технічного) завдання. Збір даних, їх аналіз. Формулювання задачі з точки зору Замовника. Консультації із Замовником. Виявлення факторів, які впливають на процес. Уточнення мети (варіантів мети).

2. Формалізація задачі у вигляді математичної моделі, яка складається з функції мети (показника якості або ефективності процесу) $F = F(X, У) = \max (\min)$

при обмеженнях $g_i(X, Y) \le b_i$

де X – вектор керованих змінних (ними розпоряджається керуюча сторона); У

- вектор некерованих аргументів (некеровані, невизначені, випадкові фактори); $g_i(X,$ У) -функція споживання *i-го* ресурсу; b_i - величина *i-го* ресурсу (вага ресурсу, сума грошей, фонд машинного часу верстата та ін.).

За допомогою обмежень знаходять область припустимих розв'язків, а функція мети дозволяє визначити оптимальну точку в цій області. Отримати оптимальний розв'язок означає знайти такі величини *X,* при яких функція мети F досягає оптимуму при одночасному дотриманні нерівностей.

3. Розв'язання задачі виконується такими найбільш розповсюдженими методами:

• лінійного програмування, якщо *F = F(X, У)* та $g_i(X, У)$ - лінійні функції відносно ***X*,** У;

• нелінійного програмування, якщо *F = F(X, У)* та $g_i(X, У)$ - нелінійні функції відносно *X*, У;

• динамічного програмування, якщо *F = F(X, У)* є адитивною або мультиплікативною функцією від змінних *X*, У;

• дискретного програмування, якщо на змінні -Х, У накласти умови дискретності (наприклад, цілочисленості);

• стохастичного програмування, якщо У - випадкова величина, а замість функції мети F = F*(X,* У) розглядають її математичне очікування.

4. Перевірка та корегування моделі. Перевірка виконується порівнянням поведінки моделі з фактичним поводженням.

5. Реалізація на практиці.

Отримане на основі дослідження операцій рішення має свої особливості:

1. *Наукове кількісне обґрунтування* рекомендованої варіанту рішення із визначенням: найкращого способу дії повноти досягнення мети і ціни досягнутої мети, ступеня ризику.

2. *Системний підхід:* будь-яка задача розглядається з точки зору її впливу на критерії функціонування всієї системи.

3. *Дорогий фізичний експеримент* замінюється відносно дешевим математичним моделюванням, яке дає відповідь на багато питань і дозволяє

прийняти оптимальне рішення. При цьому використовується ЕОМ.

4. *Рекомендуючий характер* висновків із дослідження операцій: рішення приймає людина, яка повинна нести повну відповідальність за наслідки цих рішень.

Методи ДО вміщують цілий арсенал математичним засобів:

• теорію лінійного, нелінійного, дискретного (цілочисленого, бінарного, неподільного), динамічного, стохастичного програмування;

• теорію ігор;

• теорію систем масового обслуговування;

• прийняття рішень в умовах нечіткої інформації;

• теорію експертних систем;

• теорію ефективності та ін.

У принципі, будь-який розрахунок можна розглядати як дослідження операцій, бо він дозволяє прийняти обґрунтоване оптимальне рішення у багатофакторній області. Але традиційно дослідження операцій стосується більш вузького кола питань: організації взаємодій та оптимального функціонування складних систем з множиною рішень і при умовах дотримання вказаної форми математичної моделі.

ЛЕКЦІЯ 2 «Оптимізаційні задачі управління запасами»

Анотація

Загальна постановка задачі та типи моделей управління запасами. Основні характеристики моделей управління запасами. Класифікація задач управління запасами за видом попиту.

Для діяльності будь-якої організації необхідні якісь запаси. Якщо їх не буде, то при найменшому порушенні збуту вся діяльність зупиниться. Зберігати ж занадто багато запасів економічно невигідно. Знаходженню балансу між цими двома крайностями присвячені завдання управління запасами.

2.1 Загальна постановка завдання

Завдання полягає в наступному: Необхідно скласти план випуску деякого виду виробів на період, що складається з N відрізків. Передбачається, що для кожного з цих відрізків є точний прогноз попиту на продукцію що випускається. Для різних відрізків попит неоднаковий. Причому, продукція, що виготовляється протягом відрізку часу t, може бути використана для повного або часткового покриття попиту протягом цього відрізка. Крім того, розміри виготовлених партій продукції впливають на економічні показники виробництва. У зв'язку з цим буває доцільно виготовляти протягом деякого періоду обсяг продукції, що перевищує його попит в межах цього періоду і зберігати ці надлишки до задоволення подальшого попиту. Проте, зберігання запасів пов'язане з витратами (плата за складські приміщення, страхові внески і витрати на утримання запасів і тощо).

Мета підприємства - розробити таку програму, при якій загальна сума витрат на виробництво і утримання запасів мінімізується при умови повного і своєчасного задоволення попиту на продукцію. Для забезпечення безперервного і ефективного функціонування практично будь-якої організації необхідне створення запасів, наприклад, в виробничому процесі, торгівлі, медичному обслуговуванні і т.д. В залежно від ситуації під запасами можуть матися на увазі: готова продукція, сировина, напівфабрикати, верстати, інструмент, транспортні засоби, готівка та ін.. Невірний розрахунок необхідних запасів може привести як до незначного збитку (втрата частини доходу від дефіциту товару), так і до катастрофічних наслідків (при помилковій оцінці запасів палива на літаку).

До економічного збитку приводить як надмірне наявність запасів, так і їх недостатність. Так, якщо деяка компанія має товарні запаси, то капітал, матеріалізований у цих товарах, заморожується. Цей капітал, який не можна використовувати, представляє для компанії втрачену вартість у формі невиплачених відсотків або не використовуваних можливостей інвестування. Крім того, запаси, особливо продукти що швидко псуються, вимагають створення спеціальних умов для зберігання. Для цього необхідно виділити певні площі, найняти персонал, застрахувати запаси. Все це спричиняє певні витрати.

З іншого боку, чим менше рівень запасу, тим більша ймовірність виникнення дефіциту, що може принести збитки внаслідок втрати клієнтів, зупинки виробничого процесу і т.д.

Крім того, при малому рівні запасів доводиться часто постачати нові партії товару, що призводить до великих витрат на доставку замовлень.

Звідси випливає важливість розробки та використання математичних моделей, що дозволяють знайти оптимальний рівень запасів, що мінімізують суму всіх описаних видів витрат.

Розглянемо основні характеристики моделей управління запасами.

Попит. Попит на продукт що запасається може бути *детермінованим* (в найпростішому випадку – постійним в часі) або *випадковим*. Випадковість описується або випадковим моментом попиту, або випадковим об’ємом попиту в детерміновані або випадкові моменти часу.

Поповнення складу. Поповнення складу може відбуватися або періодично через конкретні інтервали часу, або по мірі вичерпаності запасів, тобто зниження їх до деякого рівня.

Об’єм замовлення. При періодичному поповненні та випадковому вичерпанні запасів об’єм заказу може залежати від того стану, який спостерігається в момент подачі замовлення. Замовлення як правило подається на одну і ту ж величину при досягнення запасів заданого рівня – так званої *точки замовлення*.

Час доставки. В ідеалізованих моделях управління запасами передбачається, що замовлене поповнення поставляється на склад миттєво. В інших моделях розглядається затримка поставок на фіксований або випадковий інтервал часу.

Вартість поставки. Як правило, передбачається, що вартість кожної поставки складається з двох компонент – разових витрат, незалежних від об'єму партії що замовляється, та затрат, що залежать (частіше всього – лінійно) від об'єму партії.

Витрати на зберігання. В більшості моделей управління запасами вважають об'єм складу практично необмеженим, а в якості контролюючої величини служить об'єм запасів що зберігаються.

Штраф за дефіцит. Любий склад створюється для того, щоб попередити дефіцит конкретного виду виробів в системі що обслуговується. Відсутність запасів в потрібний момент приводить до збитків, зв'язаних з простоєм обладнання, неритмічністю виробництва і т.п. ці збитки в подальшому будемо називати *штрафом за дефіцит.*

Номенклатура запасу. В самих простих випадках передбачається, що на складі зберігається запас однотипних виробів або однотипної продукції. В більш складних випадках розглядається *багатономенклатурний запас*.

Я якості критерію ефективності прийнятої стратегії управління запасами виступає *функція затрат (витрат)*, що представляє сумарні затрати на зберігання та поставку запасає мого продукту (в тому числі втрати від псування продукту при його зберіганні і його морального старіння, втрати прибутку від омертвіння капіталу и т. п.) та затрати на штрафи.

Управління запасами заклечається в пошуку такої стратегії поповнення та витрат запасами, при якому функція затрат приймає мінімальне значення.

2.2 Типи моделей управління запасами.

Узагальнена модель управління запасами, розглянута вище, є досить простою. Але *різноманітність моделей* цього класу й *методів розв'язування відповідних задач*, які базуються на різному математичному апараті - від простих схем диференціального і інтегрального числення до складних алгоритмів динамічного і інших видів математичного програмування, - визначається *характером попиту*, який може бути *детермінованим* або *стохастичним*. На рис. 2 наведена схема класифікації попиту, який, зазвичай, використовується в моделях управління запасами.

Детермінований попит може бути *статичним,* в тому сенсі, що інтенсивність споживання залишається незмінною з часом, або *динамічним,* коли попит відомий достовірно, але змінюється в залежності від часу.

Стохастичний попит є *стаціонарним,* якщо функція щільності імовірності попиту незмінна в часі, і *нестаціонарним,* коли функція щільності попиту змінюється в часі.

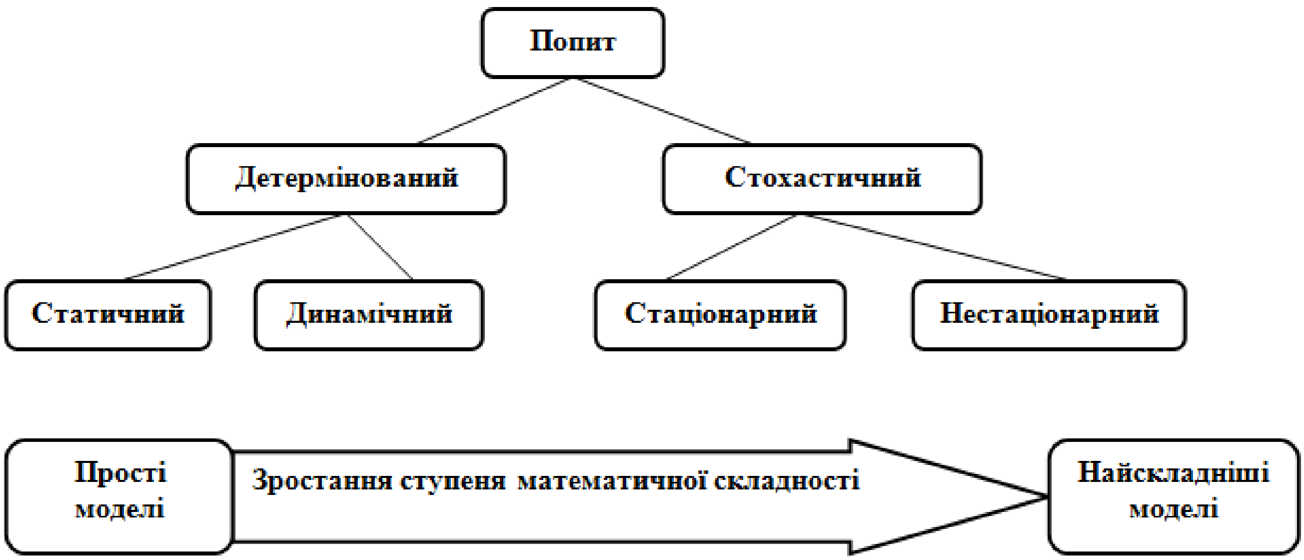


Рисунок 2.1 — Класифікація задач управління запасами за видом попиту

В реальних умовах випадок детермінованого статичного попиту зустрічається доволі рідко. Такий випадок є найпростішим. Так, *наприклад*, хоча попит на деякі продукти масового споживання, як хліб, може змінюватися від одного дня до іншого, ці зміни можуть бути настільки незначними, що припущення про статичність попиту несуттєво спотворює дійсність.

Найточніше характер попиту може бути описаний за допомогою *нестаціонарних розподілів вірогідностей*. Однак із математичної точки зору модель значно ускладнюється, особливо при збільшенні періоду часу, що розглядається. На рис. 2 ілюструється зростання математичної складності моделі управління запасами при переході від *детермінованого статичного попиту* до *вірогіднісного нестаціонарного попиту*. По суті, цю класифікацію можна вважати представленням *рівнів абстрагування* при описанні попиту.

На першому рівні абстрагування припускається, що розподіл вірогідності попиту стаціонарний у часі. Це означає, що для описання попиту протягом усіх періодів часу, що досліджуються, використовується одна і та ж функція розподілу вірогідностей. При такому припущенні вплив квартальних коливань попиту в моделі не враховується.

На другому рівні абстракції враховуються зміни попиту від одного періоду до іншого. Однак при цьому функції розподілу не застосовуються, а потреби в кожному періоді описуються середньою величиною попиту. Це спрощення означає, що елемент ризику в управлінні запасами не враховується. Однак це дозволяє досліджувати квартальні коливання попиту, які внаслідок аналітичних і обчислювальних труднощів не можна врахувати у вірогіднісній моделі.

Іншими словами, тут виникає певний компроміс: можна використовувати, з одного боку; стаціонарні розподіли ймовірностей, а з іншого - змінну, але відому функцію попиту при припущенні "визначеності".

На третьому рівні спрощення виключаються як елементи ризику, так і зміни попиту. Тим самим попит протягом будь-якого періоду вважається рівним середньому значенню відомого (за припущенням) попиту для всіх періодів, що розглядаються. В результаті цього спрощення попит можна оцінити його *сталою інтенсивністю*.

Хоча характер попиту є одним з основних факторів при побудові моделі управління запасами, є й інші фактори, що впливають на вибір типу моделі, а саме:

Запізнення надходжень виконання замовлень. Після розміщення замовлення воно може бути поставлене відразу ж або буде потрібний деякий час на його виконання. Інтервал часу між моментом розміщення замовлення і його надходженням називається запізненням замовлення або терміном виконання замовлень. Ця величина може бути детермінованою або стохастичною.

Поповнення замовлення. Хоча система управління запасами може функціонувати при запізненні надходжень, процес збільшення запасу може здійснюватися *миттєво* або *рівномірно в часі*. *Миттєве збільшення запасу* може бути реалізоване за умови, коли замовлення надходять від зовнішнього джерела. *Рівномірне збільшення* може бути тоді, коли продукція, що запасається, виробляється самою організацією. В загальному випадку система може функціонувати при позитивному запізненні надходження і рівномірному збільшенні запасу.

Період часу визначає інтервал, протягом якого здійснюється зміна рівня запасу. В залежності від проміжку часу, на якому можна надійно прогнозувати *період*, що розглядається, вважається *скінченим* або *нескінченим*.

Кількість пунктів накопичення запасу. До складу системи управління запасами може входити декілька пунктів зберігання запасу. В деяких випадках ці пункти організовані таким чином, що один є постачальником для іншого. Ця схема деколи реалізується на різноманітних рівнях, так що пункт-споживач одного рівня може стати пунктом-постачальником на іншому. В такому випадку це є система управління запасами з розгалуженою структурою.

Кількість видів продукції. В системі управління запасами може фігурувати *більше ніж один вид продукції*. Цей фактор враховується за умови наявності деякої залежності між різними видами продукції. Так, для різних виробів може використовуватися одне і те ж складське приміщення або ж виробництво може здійснюватися при обмеженнях на загальні виробничі фонди.

2.3 Прості моделі управління запасами.

Нехай функції *A(t), B(t),* та *R(t)* виражають відповідно поповнення запасів, їх витрати та попит на продукт що запасається на проміжку часу [0, t]. В моделях управління запасами використовуються похідні цих функцій за часом *a(t), b(t), r(t)*, що називаються відповідно *інтенсивностями поповнення, витрат та попиту.*

Якщо функції *a(t), b(t), r(t),* - не випадкові величини, то модель управління запасами вважається *детермінованою*, якщо хоча б одна із них носить випадковий характер – *стохастичною.* Якщо всі параметри моделі не змінюються в часі, вона називається *статичною*, інакше – *динамічною.* Статичні моделі використовуються, коли приймається разове рішення про рівень запасів на певний період, а динамічні – у випадку прийняття послідовних рішень про рівні запасу чи корегування раніше прийнятих рішень з врахуванням змін що відбуваються.

Рівень запасу в момент t визначається основним рівнянням запасі

$$J(t) = J_0 + A(t) - B(t), \tag{2.1}$$

де J_0 – початковий запас в момент $t=0$.

Рівняння (2.1) частіше використовується в інтегральній формі:

$$J(t) = J_0 + \int_0^t a(t)\,dt - \int_0^t b(t)\,dt. \tag{2.2}$$

Приклад 1. Інтенсивність надходження деталей на склад готової продукції цеха складає на початку зміни 5 дет./хв., на протязі першої години лінійно росте, досягаючи до кінця її 10 дет./хв., а потім залишається постійною. Припускаючи, що надходження деталей на склад відбувається безперервно на протязі всіх семи годин зміни, а вивіз деталей зі складу виконується лише в кінці роботи, записати вираз для рівня запасу в довільний момент часу та, використовуючи його, знайти кількість деталей на складі: а) через 30 хв. після початку роботи; б) в кінці зміни.

Рішення. За умовою протягом зміни не відбувається видача деталей зі складу, тобто $b(t)=0$. Інтенсивність поповнення запасу на протязі першої години лінійно зростає, тобто $a(t)=kt+b$. Враховуючи, що $a(0)=5$, отримуємо $b=5$. Так як в кінці першої години, тобто при $t=60$ $a(60)=10$, тоді $10=k*60+5$,звідки $k=1/12$. Таким чином, для першої години зміни $a(t)=(1/12)t+5$, а тоді $a(t)=10$. Враховуючи тривалість зміни (7 год. = 420 хв.) та співвідношення (2.2), отримаємо:

$$J(t) = \int_0^t (t/12+5)\,dt = t^2/24 + 5t, \qquad \text{якщо } 0 \le t \le 60\,,$$

і

$$J(t) = \int_0^{60} (t/12+5)\,dt + \int_{60}^{t} 10\,dt = (t^2/24 + 5t)\Big|_0^{60} + 10t\Big|_{60}^{t} = 450 + 10t - 600 = 10t - 150, \qquad \text{якщо } 60 \le t \le 420\,.$$

Кількість деталей на складі через 30 хв. після початку роботи: $J(30)=900/24+5*30=187,5$, а в кінці зміни: $J(420)=10*420-150=4050$.

2.4 Статична детермінована модель без дефіциту

Припущення про те, що дефіцит не допускається, означає повне задоволення попиту на продукт, що запасається, тобто співпадання функцій $r(t)$ та $b(t)$. Нехай загальне споживання продукту що запасається за розглянутий інтервал часу θ дорівнює N. Розглянемо просту модель, в якій передбачається, що витрати запасу відбуваються безперервно з постійною інтенсивністю, тобто $b(t)=b$. Цю інтенсивність можна знайти, розділивши загальне споживання продукту на час, на протязі якого він витрачається:

$$b = \frac{N}{\theta}. \tag{2.3}$$

Поповнення заказу відбувається партіями однакового об’єму, тобто функція $a(t)$ не являється безперервною: $a(t)=0$ при всіх t, крім моментів поставки продукту, коли $a(t)=n$, де n – об’єм партії. Так як інтенсивність витрат дорівнює b, то вся партія буде використана за час

$$T = \frac{n}{b}. \quad (2.4)$$

Якщо відлік часу почати з моменту надходження першої партії, то рівень запасу в початковий момент дорівнює об’єму цієї партії n,тобто $J(0)=n$. Графічно рівень запасу в залежності від часу представлено на рис.2.2.

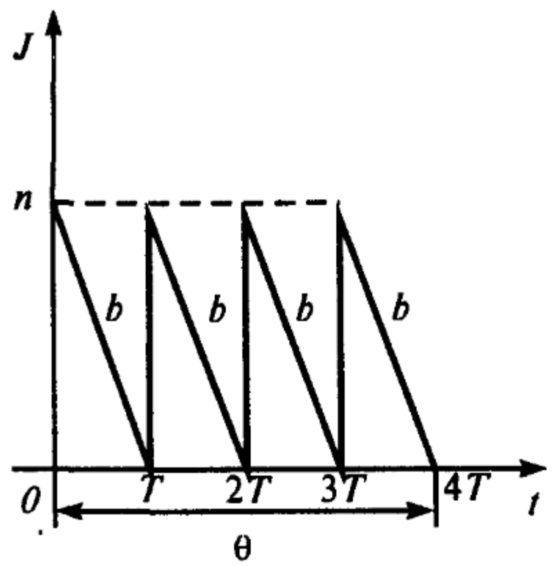


Рисунок 2.2 – Залежність рівня запасу від часу

На протязі часового інтервалу *[0,T]* рівень запасу зменшується по прямій $J(t)=n\text{-}bt$ від значення n до нуля. Так як дефіцит не допускається, то в момент T рівень запасу миттєво поповнюється до попереднього значення n за рахунок надходження партії заказу. Таким чином процес зміни $J(t)$ повторюється на кожному часовому інтервалі тривалістю T (див. рис. 2.2)

Задача управління запасами виражається у визначенні такого об’єму партії n, при якому сумарні затрати на створення та зберігання запасу були б мінімальними.

Позначимо сумарні затрати через C, затрати на створення запасу – через C_1, а затрати на зберігання запасу – через C_2 та знайдемо ці величини за весь проміжок часу T.

Нехай затрати на доставку однієї партії продукту, незалежно від об’єм партії, дорівнюють c_1, а затрати на зберігання однієї одиниці продукції за одиницю часу – c_2. Так як за час θ необхідно запастись N одиницями продукту, який поставляється партіями об’єму n, то число таких партій k дорівнює:

$$k = \frac{N}{n} = \frac{\theta}{T}. \tag{2.5}$$

Звідси отримуємо

$$C_1 = c_1 k = c_1 \frac{N}{n}. \tag{2.6}$$

Моментальні затрати зберігання запасу в момент часу t дорівнюють $c_2 J(t)$. Значить за проміжок часу $[0,T]$ вони складуть

$$c_2 \int_0^T J(t)\, dt = c_2 \int_0^T (n - bt)\, dt$$

або, враховуючи (2.4):

$$c_2 \int_0^T J(t)\, dt = c_2 \int_0^T \left(n - \frac{n}{T} t\right) dt = c_2 \left(nt - \frac{nt^2}{2T}\right)\Big|_0^T = \frac{c_2 n T}{2}.$$

Середній запас за проміжок $[0,T]$ дорівнює $nT/2$, тобто затрати на зберігання всього запасу при лінійних (за часом) його витратах дорівнюють затратам на зберігання середнього запасу.

Враховуючи періодичність функції *J(t)* (всього за проміжок часу θ буде k=*N/n* «зубців», аналогічних тим, що розглядалися на відрізку *[0,T]*), та формулу (2.5), отримуємо, що затрати зберігання запасу за проміжок часу θ дорівнюють:

$$C_2 = \frac{c_2 n T}{2} k = \frac{c_2 n T}{2} * \frac{N}{n} = \frac{c_2 T N}{2} = \frac{c_2 \theta n}{2}. \qquad (2.7)$$

Неважко помітити, що затрати C_1 обернено пропорційні, а затрати C_2 прямо пропорціональні об'єму партії *n*. Графіки функцій $C_1(n)$ та $C_2(n)$, а також функції сумарних затрат

$$C = \frac{c_1 N}{n} + \frac{c_2 \theta}{2} n \qquad (2.8)$$

представлені на рис. 2.3.

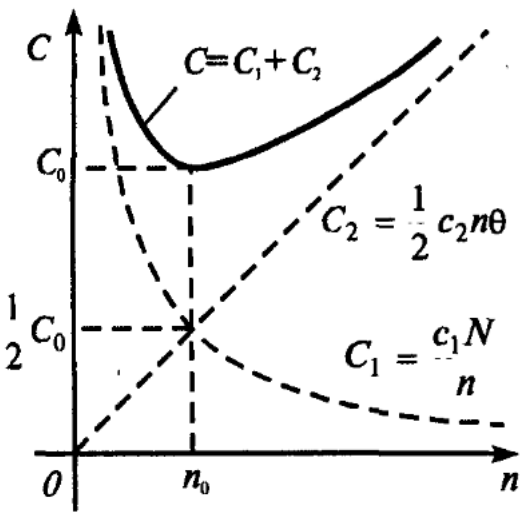


Рисунок 2.3 – Функція сумарних затрат та функцій $C_1(n)$ та $C_2(n)$

В точці мінімуму функції *C(n)* її похідна $C'(n) = -\frac{c_1 N}{n^2} + \frac{c_2 \theta}{2} = 0$, звідки

$$n = n_0 = \sqrt{\frac{2 c_1 N}{c_2 \theta}} \qquad (2.9)$$

або враховуючи (2.3):

$$n_0 = \sqrt{\frac{2c_1 b}{c_2 p}}. \tag{2.10}$$

Формула (2.10) називається *формулою Уілсона* або *формулою найбільш економічного об'єму партії*, широко використовується в економіці. Ця формула може бути отримана і іншим способом, якщо врахувати, що добуток $C_1C_2=0{,}5c_1c_2N\theta$ є величина постійна, незалежна від n. В цьому випадку, як відомо, сума двох величин приймає найменше значення, коли вони рівні, тобто $C_1=C_2$ або

$$\frac{c_1 N}{n} = \frac{c_2 n \theta}{2}, \tag{2.11}$$

звідки отримуємо (2.9).

Із (2.11) випливає, що мінімум загальних затрат задачі управління запасами досягається тоді, коли затрати на створення запасу дорівнюють затратам на збереження запасів. При цьому мінімальні сумарні затрати

$$C_0 = C\ n_0 = \frac{2c_1 N}{n}, \tag{2.12}$$

звідки, враховуючи (2.9) та (2.3), отримаємо $C_0 = \sqrt{2c_1c_2\theta N}$

або $$C_0 = \theta\sqrt{2c_1c_2 b}. \tag{2.13}$$

Число оптимальних партій за час θ з урахуванням (2.5), (2.9) і (2.3) дорівнює:

$$k_0 = \frac{N}{n_0} = \sqrt{\frac{c_2 N\theta}{2c_1}} = \theta\sqrt{\frac{c_2 b}{2c_1}}.$$

Час витрати оптимальної партії на основі (2.4) з урахуванням (2.9) та (2.3) дорівнює

$$T_0 = \frac{n_0}{b} = n_0 \frac{\theta}{N} \quad (2.14)$$

або

$$T_0 = \sqrt{\frac{2c_1\theta}{c_2 N}} = \sqrt{\frac{2c_1}{c_2 b}}. \quad (2.15)$$

Приклад 2. Потреба складального підприємства в деталях певного типу складає 120000 деталей на рік, причому ці деталі витрачаються в процесі виробництва рівномірно та безперервно. Деталі замовляються раз на рік та поставляються партіями однакового об'єму, вказаного в замовлені. Зберігання деталі на складі коштує 0,35 грошових одиниць на добу, а поставка партії – 10000 грош. одиниць. Затримка виробництва із-за відсутності деталей недопустима. Встановити найбільш економічний об'єм партії та інтервал між поставками, які необхідно вказати в замовлені (постачальник не допускає затримки поставок).

Рішення. За умовою затрати на одну партію складають $c_1=10000$ грош.од., загальний проміжок часу $\theta=1$ рік =365 днів, а загальний об'єм запасу за цей період $N=120000$ деталей. За формулою (2.9) $n_0 = \sqrt{\frac{2*10000*120000}{0,35*365}} \approx 4335\ \textit{дет.}$, а за (2.14) $T_0 = n_0 \frac{\theta}{N} = 13,2 \approx 13\ \textit{днів}$.

Таким чином, найбільш економічний об'єм партії дорівнює 4335 деталей, а інтервал між поставками ≈13 днів.

На практиці ж об'єм партії може відрізнятися від отриманого n_0, розрахованого за (2.9). Так, в попередній задачі може виявитися зручним замовити партії по 4500 або навіть по 5000 деталей і виникає питання, як при цьому змінюються сумарні затрати.

Для відповіді на це запитання розложимо функцію $C(n)$ в ряд Тейлора навколо точки n_0, обмежившись першими трьома членами ряду при достатньо малих змінах об'єму партії Δn:

$$C(n) = C(n_0) + C'(n_0)\Delta n + \frac{c''(n_0)}{2!}\Delta n^2 + \cdots .$$

Враховуючи, що при $n=n_0$ $C'(n_0) = 0$, $C''(n_0) = \frac{2c_1N}{n_0^3}$, а $C_0=C(n_0)$ визначається за формулою (2.12), знайдемо:

$$\frac{\Delta C}{C_0} = \frac{C(n) - C(n_0)}{C(n_0)} \approx \frac{C''(n_0)\Delta n^2}{2C(n_0)} = \frac{2c_1N\Delta n^2}{n_0^3(\frac{2c_1N}{n_0})}$$

або

$$\frac{\Delta C}{C_0} \approx \frac{1}{2}\left(\frac{\Delta n}{n_0}\right)^2. \qquad (2.16)$$

Формула (2.16) свідчить про певну стійкість сумарних затрат у відношенні до найбільш економічного об'єму партії, так як при малих Δn відносні зміни затрат приблизно на порядок менші відносної зміни об'єму партії в порівнянні з оптимальним.

Приклад 3. За умовою задачі 2 визначити, на скільки процентів збільшаться затрати на створення та зберігання запасу в порівнянні з мінімальними затратами при об'ємі замовляємих партій 5000 деталей.

Рішення. Відносна зміна об'єму партії в порівнянні з оптимальним $n_0=4335$ складає $\Delta n/n_0=(5000-4335)/4335=0,153$. У відповідності до (2.16) відносна зміна сумарних затрат складе $\Delta C/C_0=0,153^2/2\approx0,012$, або всього 1.2%.

Приклад 4. В умові задачі 3 припустимо, що замовляються не всі партії відразу, а кожна окремо, причому строк виконання замовлення дорівнює 16 днів. Визначити точки замовлення, тобто при якому рівні запасу слід замовляти наступну партію.

Рішення. Так як за результатами рішення задачі 2 довжина інтервалу між поставками дорівнює 13,2 днів, то замовлення в умовах неналежного виробництва потрібно відновити, коли рівень запасу достатній для задоволення потреби на 16-13,2=2,8 дні. Так як щоденна необхідність (інтенсивність витрат запасу) дорівнює за формулою (2.3) b=120000/365=329 деталей, то замовлення повинні робитися регулярно при досягненні рівня запасу 329*2,8≈922 деталі.

ЛЕКЦІЯ 3 «Статистична детермінована модель з дефіцитом»

Анотація

Статична модель економічного замовлення з наявністю дефіциту запасів. Стохастичні моделі управління запасами. Стохастичні моделі управління запасами з фіксованим часом затримки поставок.

3.1 Статистична детермінована модель з дефіцитом

В розглянутій моделі будемо вважати наявність *дефіциту*. Це означає, що при відсутності запасоємного продукту, тобто при $J(t)=0$ попит зберігається з тією ж інтенсивність $r(t)=b$, але споживання запасу відсутнє – $b(t)=0$, внаслідок чого накопичується дефіцит із швидкістю b. Графік зміни рівня запасу в цьому випадку представлений на рис. 3.1.

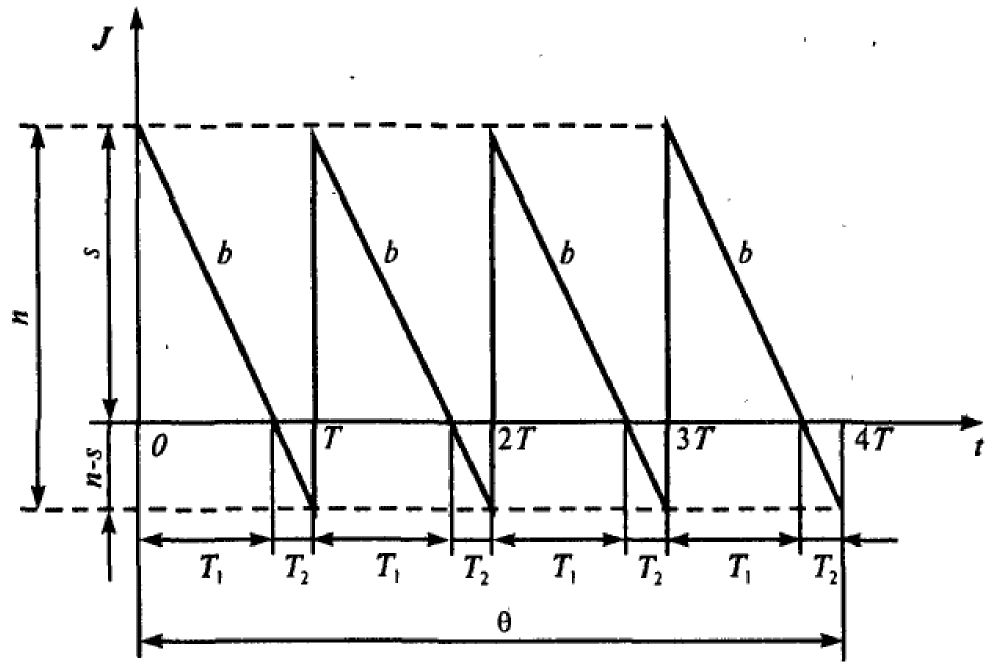


Рисунок3.1 – Графік зміни рівня запасу

Спадання графіку нижче осі абсцис в область від'ємних значень на відміну від графіка на рис 3.1 характеризує накопичення дефіциту.

З рис.3.1 видно, що кожний період «пилки» $T = \frac{n}{b}$ розбиваються на два часових інтервали, тобто $T = T_1 + T_2$, де T_1 – час, на протязі якого відбувається споживання запасу, T_2 – час, коли запас відсутній та накопичується дефіцит, який буде перекритий в момент надходження наступної партії.

Необхідність покриття дефіциту приводить до того, що максимальний рівень запасу s в момент надходження кожної партії тепер не дорівнює його об'єму n, а менше його на величину дефіциту n-s, накопиченого за час T_2 (див. рис.3.1).

З геометричних міркувань легко встановити, що

$$T_1 = \frac{s}{n}T, \quad T_2 = \frac{n-s}{n}T. \tag{3.1}$$

В даній моделі в функцію сумарних затрат C поряд із затратами C_1 (на поповнення запасу) та C_2 (на зберігання запасу необхідно ввести затрати C_3 – на штраф із-за дефіциту, тобто C=C_1+C_2+C_3.

Затрати C_1, як і раніше, знаходимо за формулою (2.11). В попередньому розділі було показано, що затрати C_2 при лінійних витратах запасу дорівнюють затратам на зберігання середнього запасу, який на час споживання T_1 дорівнює sT 2; тому з урахуванням (2.7) та (2.5) ці затрати складають

$$C_2 = \frac{c_2 s T_1}{2}k = \frac{c_2 s * sT}{2} * \frac{\theta}{T} = \frac{c_2 s^2 \theta}{2n}. \tag{3.2}$$

При розрахунку затрат C_3 будемо вважати, що штраф за дефіцит становить за одиницю часу c_3 на кожну одиницю продукту. Так як середній рівень дефіциту за період T_2 дорівнює $(n$-$s)T_2/2$, то штраф за цей період T_2 складає $\frac{1}{2}c_3(n-s)T_2$, а за весь період θ з урахуванням (2.7) і (2.3) –

$$C_3 = \frac{1}{2}c_3 \; n-s \; T_2 k = \frac{1}{2}c_3 \; n-s \; \frac{n-s}{n}T\frac{\theta}{T} = \frac{c_3\theta \; n-s \; ^2}{2n}. \tag{3.3}$$

Тепер, враховуючи (2.12), (3.2) та (3.3), сумарні затрати дорівнюють

$$C = c_1\frac{N}{n} + \frac{c_2\theta s^2}{2n} + \frac{c_3\theta\ \ b-s\ ^2}{2n}. \qquad (3.4)$$

Неважко помітити, що при n=s формула (3.3) співпадає з отриманою раніше (3.2) в моделі без дефіциту.

Розглянута *задача управління запасами зводиться до відшукання такого об'єму партії* n *та максимального рівня запасу* s, *при яких функція* C (2.16) *приймає мінімальне значення.* Іншими словами, необхідно дослідити функцію двох змінних *C(n, s)* на екстремум. Прирівнюючи часткові похідні $\partial C\ \ \partial n,\ \ \partial C\ \ \partial s$ до нуля, отримаємо після перетворення систему рівнянь:

$$\begin{gathered} n^2c_3 - \ \ c_2 + c_3\ \ s^2 = \frac{2c_1N}{\theta}, \\ s = n\frac{c_3}{c_2+c_3}. \end{gathered} \qquad (3.5)$$

Вирішуючи систему, отримуємо формули найбільш економічного об'єму партії n_0 та максимального рівня запасу s_0 для моделі з дефіцитом:

$$n_0 = \sqrt{\frac{2c_1N}{c_2\theta}}\ \sqrt{\frac{c_2+c_3}{c_3}} = \sqrt{\frac{2c_1b}{c_2}}\ \sqrt{\frac{c_2+c_3}{c_3}}, \qquad (3.6)$$

$$s_0 = \sqrt{\frac{2c_1N}{c_2\theta}}\ \sqrt{\frac{c_3}{c_2+c_3}} = n_0\frac{c_3}{c_2+c_3}. \qquad (3.7)$$

Величина

$$p = \frac{c_3}{c_2+c_3} \qquad (3.8)$$

називається *щільністю збитків із-за незадоволеного попиту* і відіграє важливу роль в управлінні запасами. Відмітимо, що $0 \leq p \leq 1$. Якщо значення c_3 дорівнює з c_2, то величина p наближається до нуля: коли c_3 значно перевищує c_2, то p

близька до 1. Недопустимість дефіциту рівноцінна припущенню, що $c_3 = \infty$ або *p=1*.

Використовуючи (3.8), основні формули (3.6) і (3.7) можна записати компактніше:

$$n_0 = \overline{\frac{2c_1b}{c_2p}}, \tag{3.9}$$

$$s_0 = n_0p. \tag{3.10}$$

Необхідно врахувати, що в силу (3.1) та (3.10) $T_1 \ \ T = s_0 \ \ n_0 = p$ і $T_2 \ \ T = n_0 - s_0 \ \ n_0 = 1 - p$. Тому твердження про те, що щільність збитків із-за незадоволеного попиту дорівнює p, означає , що на протязі *(1-p)100%* часу від повного періоду *T* запас продукту буде відсутній.

Із порівняння формул (3.9) та (2.10) випливає, що оптимальні об’єми партій для задач з дефіцитом та без дефіциту при однакових параметрах пов’язані співвідношенням

$$n_0 = \frac{n_0}{\overline{p}}, \tag{3.11}$$

звідки випливає, що *оптимальний об’єм партії в задачі з дефіцитом завжди більший (в* $\frac{1}{p}$ *раз), ніж в задачі без дефіциту.*

Приклад 5. Знайти найбільш економічний об’єм партії та інтервал між поставками, зберігаючи умови задачі 2, окрім недопустимості дефіциту, якщо відомо, що відсутність на збірці кожної деталі приносить на добу збитки в розмірі 3,5 грош. од.

Рішення. За умовою c_3*=3,5*. Раніше за формулою (2.9) було отримано n_0*=4335* та за (2.15) T_0*=13,2*. Знайдемо щільність збитків із-за незадоволеного попиту за формулою (2.24): *p=3,5/(0,35+3,5)=0,909*, тобто *100(1-0,909)=9,1%* часу між поставками деталі на збірці будуть відсутні.

Тепер оптимальний розмір партії за формулою (3.11) $n_0 = 4335 \quad \overline{0{,}909} = 4547$. В силу (2.15) пропорційно збільшено n_0 повинен збільшитись інтервал між поставками, тобто

$T_0 = T_0 \quad \overline{p} = 13{,}2 \quad \overline{0{,}909} = 13{,}8 \approx 14$ днів.

3.2 Стохастичні моделі управління запасами

Розглянемо стохастичні моделі управління запасами, в яких попит являється випадковим. *Ц*ей факт істотно відбивається на характері відповідних моделей та значно ускладнює їх аналіз, тому будемо розглядати найпростіші моделі.

Припустимо, що попит *r* за інтервал часу *T* являється випадковим і задано його закон (ряд) розподілу *p(r)* або щільність ймовірностей *φ(r)* (як правило функції *p(r)* та *φ(r)* оцінюються на основі дослідних або статистичних даних). Якщо попит *r* нижчий рівня запасу*s*, то придбання (зберігання, продаж) надлишку продукту потребує додаткових затрат c_2 на одиницю продукту; навпаки, якщо попит *r* вище рівня запасу *s*, то це приводить до штрафу за дефіцит c_3 на одиницю продукції.

В якості функції сумарних витрат, що в стохастичних моделях являється випадковою величиною, розглядають її середнє значення або математичне очікування.

В моделі, що розглядається, при дискретному випадковому попиті *r*, що має закон розподілення *p(r)*, математичне очікування сумарних витрат має вигляд:

$$C(s) = c_2 \sum_{r=0}^{s} (s-r)\,p(r) + c_3 \sum_{r=s+1}^{\infty} (r-s)\,p(r) \qquad (3.12)$$

У виразі (3.12) перший доданок враховує витрати на придбання (зберігання) надлишку *s*-*r* одиниць продукту (при *r* ≤ *s*), а другий доданок – штраф за дефіцит на *r*-*s* одиниць продукту (при *r* > *s*).

У випадку неперервного випадкового процесу, що задається щільністю ймовірностей *φ(r)*, вираз *C(s)* прийме вигляд:

$$C(s) = c_2 \int_0^s (s-r)\,\varphi(r)\,dr + c_3 \int_0^s (r-s)\,\varphi(r)\,dr \qquad (3.13)$$

Задача управління запасами заключається в знаходженні такого запасу s, при якому математичне очікування сумарних витрат (2.28) та (2.29) приймають мінімальне значення.

В роботах А. Кафмана та У. Черчмена доведено, що при дискретному випадковому процесі r вираз (2.28) мінімальний при запасі s_0, що задовольняє нерівностям

$$F(s_0) < p < F(s_0+1), \qquad (3.14)$$

а при неперервному випадковому процесі r вираз (2.29) мінімальний при значення s_0, визначеному із рівняння

$$F(s_0) = p, \qquad (3.15)$$

де

$$F(s) = p(r < s) \qquad (3.16)$$

є функцією розподілення попиту r, $F(s_0)$ та $F(s_0 + 1)$ – її значення; p – щільність збитків через незадовільний попит, що встановлюється за (3.8).

Отриманий запас s_0 при безперервному попиті за даним значенням p може бути знайдений і графічно (рис. 3.2)

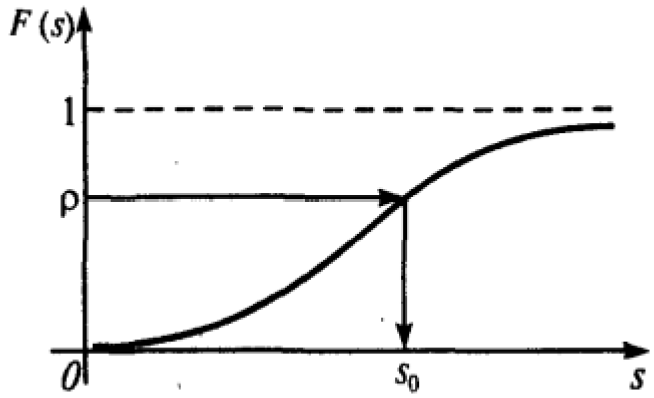


Рисунок 3.2 – Графічне представлення запасу s_0 при безперервному попиті

Задача 6. Підприємство закуповує агрегат із запасними блоками до нього. Ціна одного блоку дорівнює 5 грош. один. У випадку виходу агрегату із ладу через поломку блока, що відсутній в запасі, простій агрегату та терміну замовлення нового блоку до нього обійдеться в 100 грош. од. Експериментальне розподілення агрегатів за числом блоків, що потребували заміни, представлені в табл.3.1.

Таблиця 3.1— Експериментальне розподілення агрегатів за числом блоків, що потребували заміни

Число замененных блоков r	0	1	2	3	4	5	6
Статистическая вероятность (доля) агрегатов $p(r)$, которым потребовалась замена r блоков	0,90	0,05	0,02	0,01	0,01	0,01	0,00

Необхідно встановити оптимальне число блоків, які потрібно придбати разом з агрегатом.

Рішення. За умовою $c_2=5$, $c_3=100$. Розрахуємо щільність збитків через недостатність запасних блоків за формулою (2.24) $p = 100 / (5 + 100)=0,952$.

Враховуючи (2.32), найдемо значення функції розподілення попиту (табл.3.2)

Таблиця 3.2 – Функції розподілення попиту

s	0	1	2	3	4	5	6	>6
r	0	1	2	3	4	5	6	>6
$F(s)$	0,00	0,90	0,95	0,97	0,98	0,99	0,99	1,00

З таблиці очевидно, що оптимальний запас становить s_0=2, та як він задовольняє нерівності (2.30): $F(2) < 0,952 < F(3)$.

Задача 7. Розв’язати задачу 6 при умові безперервного випадкового процесу r, розподіленого за показниковим законом з функцією розподілення $F(r)=1-e^{-\lambda r}$ при $\lambda=0,98$.

Рішення. Оптимальне число запасних блоків s_0 знайдемо з рівняння (2.31): $1 - e^{-\lambda s_0} = p$, звідки $e^{-\lambda s_0} = p$ та $s_0 = -\frac{1}{\lambda}\ln(1-p)$. При λ*=0,98* s_0 = -*(1/0,98)ln0,02≈4* (блока).

В умовах моделі що розглядається, припустимо, що витрачання запасу відбувається безперервно з однаковою інтенсивністю. Таку ситуацію можна представити графічно (рис.3.3)

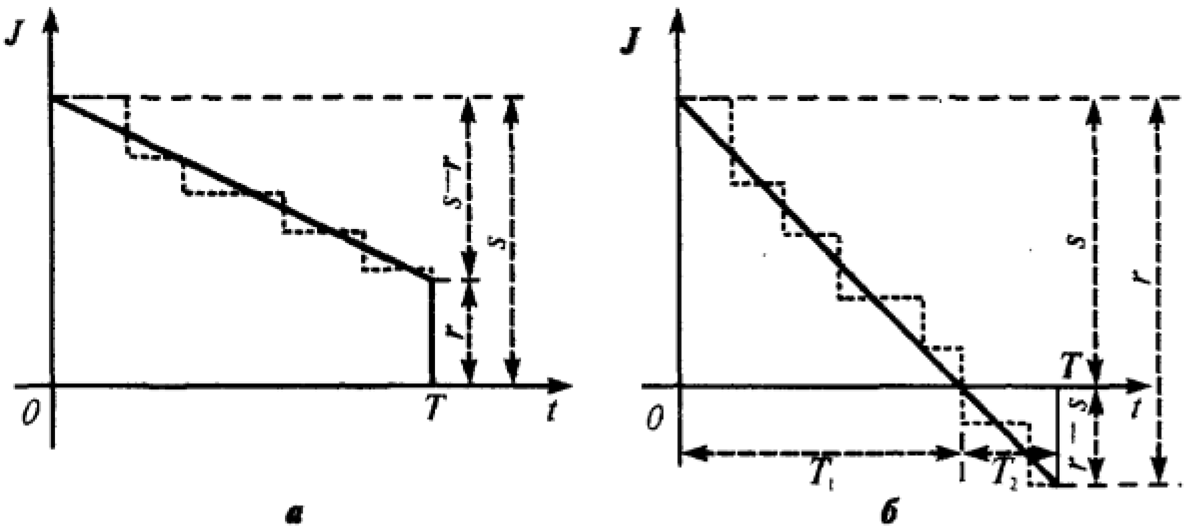


Рисунок 3.3 – Графік безперервного витрачання запасу з однаковою інтенсивністю

Рис.3.3, *а* відповідає випадку $r \leq s$, коли попит не перевищує запас, а Рисунок2.5, *б* – випадок, попит перевищує запас, тобто $r > s$. Слід відмітити, що насправді графік *J(t)* представляє ступінчату ломану, показану на рис.2.5 пунктиром, але для дослідження моделі нам простіше розглядати *J(t)* у вигляді прямої, що згладжує цю ламану.

Середній запас, відповідний рис. 3.3, а дорівнює

$$s_1 = \frac{1}{2}\ s + \ s + r \quad = s - \frac{1}{2}r. \qquad (3.17)$$

Середній запас, відповідний рис.3.3, б з урахуванням формули (3.1), в якій вважаємо $n=r$, складає

$$s_2 = \frac{1}{2} s \frac{T_1}{T} = \frac{1}{2} \frac{s^2}{r} \tag{3.18}$$

Середній дефіцит продукту за період T_2 для випадку, що відповідає рис.3.3, б з урахуванням (3.1), де $n=r$, дорівнює

$$s_3 = \frac{1}{2} (r - s) \frac{T_2}{T_1} = \frac{1}{2} \frac{(r-s)^2}{r} \tag{3.19}$$

Математичне очікування сумарних витрат складе:

$$C(s) = c_2 \sum_{r=0}^{s} \left(s - \frac{r}{2}\right) p(r) + c_2 \sum_{r=s+1}^{\infty} \frac{1}{2} \frac{s^2}{r} p(r) + c_3 \sum_{r=0}^{\infty} \frac{1}{2} \frac{(r-s)^2}{r} p(r) \tag{3.20}$$

А. Кафманом та У. Черчменом доведено, що в цьому випадку математичне очікування (3.20) мінімальне при запасі s_0, що задовольняє нерівності

$$L(s_0) < p < L(s_0 + 1) \tag{3.21}$$

де p знову ж таки встановлюється за формулою (3.8):

$$L(s) = F(s) + \left(s - \frac{1}{2}\right) \sum_{r=s}^{\infty} \frac{p(r)}{r} \tag{3.22}$$

$L(s_0)$ та $L(s_0+1)$ – значення функції (2.38), а $F(s)$ знаходиться у відповідності до визначення (3.16).

Приклад 8. Вироби, що знаходяться на складі, рівномірно використовуються на протязі місяця. Затрати на зберігання одного виробу складають 5 грош. од., а штраф за дефіцит одного виробу обходиться в 100 грош. од. Вивчення попиту дало розподілення числа споживаних за місяць виробів, представлене в табл. 3.3.

Таблиця 3.3 – Розподілення числа споживаних за місяць виробів

Спрос r	0	1	2	3	4	5	≥6
Статистическая вероятность $p(r)$	0.1	0.2	0.2	0.3	0.1	0.1	0.0

Необхідно, визначити оптимальний місячний запас складу.

Рішення. Так же як і в задачі 6, $c_2=5$, $c_3=100$, $p=0{,}952$.

Значення функції $L(r)$ встановимо за допомогою таблиці 3.4.

Таблиця 3.4 – Алгоритм встановлення функції $L(r)$

s	r	$p(r)$	$\frac{p(r)}{r}$	$\sum_{r=s}^{\infty}\frac{p(r)}{r}$	$\left(s-\frac{1}{2}\right)\sum_{r=s}^{\infty}\frac{p(r)}{r}$	$F(r)$	$L(r)$
0	0	0,1	—	—	—	0,0	—
1	1	0,2	0,200	0,445	0,2225	0,1	0,3225
2	2	0,2	0,100	0,245	0,3675	0,3	0,6675
3	3	0,3	0,100	0,145	0,3625	0,5	0,8625
4	4	0,1	0,025	0,045	0,1575	0,8	0,9575
5	5	0,1	0,020	0,020	0,0900	0,9	0,9900
≥6	≥6	0,0	0,000	0,000	0,0000	1,0	1,0000

Очевидно, що оптимальний запас виробів $s_0 = 3$, так як він задовольняє умові (2.37): $L(3) < 0{,}952 < L(4)$.

3.3 Стохастичні моделі управління запасами з фіксованим часом затримки поставок

В розглянутих вище ідеалізованих моделях управляння запасами передбачалось, що поповнення запасів відбувається практично миттєво. Проте в багатьох задачах *час затримки* поставок може виявитись на стільки значним, що його необхідно враховувати в моделі.

Нехай за час затримок поставок Θ вже замовлено n партій по одній в кожний із n періодів тривалістю $T=\Theta/n$.

Позначимо:

$s_{нз}$ – початковий рівень запасу (до початку першого періоду);

s_i – запас за i-й період;

r_i – попит за i-й період;

q_i – поповнення запасу за i-й період.

Тоді до кінця n-го періоду на склад поступить ${}_{i=1}^{n} q_i$ одиниць продукту, а витрачено буде ${}_{i-1}^{n} r_i$ одиниць, тобто

$$s_n = s_{нз} + {}_{i=1}^{n} q_i - {}_{i=1}^{n} r_i, \tag{3.23}$$

або

$$s_n = s - r \tag{3.24}$$

де

$$s = s_{нз} + {}_{i=1}^{n} q_i, \tag{3.25}$$

$$r = {}_{i=1}^{n} r_i. \tag{3.26}$$

Потрібно знайти оптимальний об’єм партії замовлення, який необхідно зробити за останній n-й період, попередній надходженню зробленого раніше замовлення.

Математичне очікування сумарних витрат в цьому випадку встановлюється за формулою (3.12), а оптимальний запас s знаходиться за формулою (3.14), тобто

$$F(s_0) < p < F(s_0 + 1) \tag{3.27}$$

Знайшовши оптимальний запас s_0 та знаючи $q_1, q_2, \ldots, q_{n-1}$, можна розрахувати q_n за формулою (3.25), тобто

$$q_n = s_0 - \ s_{нз} + \ {}_{i=1}^{n-1} q_i \ . \qquad (3.28)$$

Завдання 9. Швидкопсувний товар що замовляється щоденно поступає в магазин через 7 днів після замовлення. В момент чергового замовлення запас товару склав 10 грош. од. Товар в день виготовлення приносить прибуток 0,95 грош. од., а не проданий в цей день товар може бути пізніше реалізований із збитками 0,10 грош.од.

На основі експериментальних даних отримано наступне розподілення попиту на даний товар (табл. 3.5)

Таблиця 3.5 – Розпреділення попиту на товар

r	0	10	20	30	40	50	60	70	80	90	100
p(*r*)	0,00	0,00	0,01	0,02	0,05	0,08	0,11	0,12	0,14	0,13	0,10
r	110	120	130	140	150	160	170	180	190	200	>200
p(*r*)	0,08	0,05	0,03	0,02	0,02	0,01	0,01	0,01	0,01	0,00	0,00

Необхідно встановити оптимальний об'єм замовляємого товару q_7 на сьомий день після замовлення.

Рішення. Щільність збитків через дефіцит товару за формулою (3.8) дорівнює $p = 0{,}95 \ / \ (0{,}10 + 0{,}95) = 0{,}905$. Враховуючи умови (3.16), знайдемо значення функції розподілення попиту (табл. 3.6)

Таблиця 3.6 – Значення функції розподілення попиту

s	r	$F(s)$	s	r	$F(s)$	s	r	$F(s)$	s	r	$F(s)$
0	0	0,00	50	50	0,16	100	100	0,86	150	150	0,96
10	10	0,00	60	60	0,27	110	110	0,84	160	160	0,97
20	20	0,01	70	70	0,39	120	120	0,89	170	170	0,98
30	30	0,03	80	80	0,53	130	130	0,92	180	180	0,99
40	40	0,08	90	90	0,76	140	140	0,94	≥ 190	≥ 190	1,00

Умову (3.27) задовольняють $s_0 = 120$ так як $F(120) < 0,905 < F(130)$.

Таким чином, оптимальний запас товару за 7 днів повинен бути на суму 120 грош. од., звідки оптимальний об'єм замовленого товару на сьомий день за (3.28) складе: $q_7 = 120 – (10 + (10 + 20 +10 + 10 + 20 +10)) = 30$ грош. од.

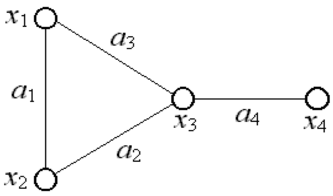
ЛЕКЦІЯ 4 «Оптимізаційні задачі управління економічними системами на основі застосування теорії мереж»

Анотація

Основні поняття теорії графів. Алгоритм побудови мінімального покриваючого дерева. Алгоритми визначення найкоротшого шляху (Дейкстри та Флойда) між вузлами мережі. Задача про максимальний потік.

4.1 Основні поняття теорії графів

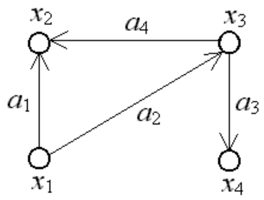
Графом $G = (X, A)$ називається пара об’єктів $X = \{x_1, x_2 ., x_n\}$ і $A = \{a_1, a_2 ., a_m\}$, де X – множина вершин, а A – множина ребер графа. Якщо ребра з множини *A орієнтовані*, то вони називаються *дугами*, а граф називають *орієнтованим*. Якщо ребра *не мають орієнтації*, то граф називають *неорієнтованим*. Інакше граф є *змішаним*. На рис.4.1 – 4.6 приведені неорієнтований і орієнтований графи відповідно.



$X = \{x_1, x_2, x_3, x_4\}$,

$A = \{a_1, a_2, a_3, a_4\}$.

Рисунок 4.1



$X = \{x_1, x_2, x_3, x_4\}$,

$A = \{a_1 = (x_1, x_2), a_2 = (x_1, x_3), a_3 = (x_3, x_4), a_4 = (x_3, x_2)\}$.

Рисунок 4.2

Якщо зіставити кожному ребру число з множини C, тоді граф називають *зваженим*.

Граф можна задати матрицями *суміжності* і *інцидентності*. Елементи матриці суміжності S графа задаються так:

$$s_{ij} = \begin{cases} 1, & \text{якщо існує ребро (дуга), що сполучає вершини } x_i \text{ та } x_j; \\ 0 & \text{інакше,} \end{cases}$$

$(i, j=1, 2,\ldots, n)$.

Елементи матриці інцидентності $U = \left|u_{ij}\right|_{m \times n}$ для графа G, що складається з n вершин і m дуг, визначаються як:

$$u_{ij} = \begin{cases} 1, & \text{якщо вершина } x_i \text{ – початок дуги } a_j; \\ -1, & \text{якщо вершина } x_i \text{ – кінець дуги } a_j; \\ 0, & \text{якщо вершина } x_i \text{ не інцидентна дузі } a_j. \end{cases}$$

$(i=1, 2,\ldots,n; \quad j=1, 2,\ldots, m)$.

Для графа, приведеного на рис.4.3, матриця суміжності приведена на рис.4.4, а матриця інцидентності – на рис.4.5.

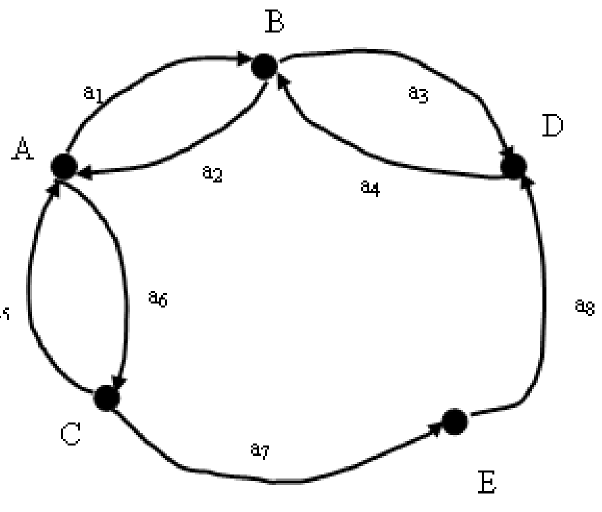


Рисунок 4.3

$$
\begin{array}{c|c}
 & A \quad B \quad C \quad D \quad E \\
\hline
\begin{array}{r} A \\ B \\ S = C \\ D \\ E \end{array} &
\begin{pmatrix}
0 & 1 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0
\end{pmatrix}
\end{array}
$$

Рисунок 1.4 – Матриця суміжності

$$
\begin{array}{c|c}
 & a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7 \quad a_8 \\
\hline
\begin{array}{r} A \\ B \\ U = C \\ D \\ E \end{array} &
\begin{pmatrix}
1 & -1 & 0 & 0 & -1 & 1 & 0 & 0 \\
-1 & 1 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 \\
0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 1
\end{pmatrix}
\end{array}
$$

Рисунок 4.5 – Матриця інцидентності

Якщо граф містить петлі, тобто дуги вигляду (x_i, x_i) тоді елементи матриці інцидентності, відповідні дугам, створюючим петлі, одночасно рівні 1 і –1, що приводить до неоднозначності матриці інцидентності.

Нехай G – неорієнтований граф. *Маршрутом* в графі G називається така послідовність (кінцева або нескінченна) ребер a_1, a_2, ... a_n..., що кожні два сусідні ребра a_i та a_{i+1} мають загальну інцидентну вершину. Одне і те ж ребро може зустрічатися в маршруті кілька разів. У кінцевому маршруті (a_1, a_2, ... a_n) є перше ребро a_1 і останнє ребро a_n. Вершина x_1, інцидентна ребру a_1, але не інцидентна ребру a_2, називається початком маршруту, а вершина x_n, інцидентна ребру a_n, але не інцидентна ребру a_{n-1}, називається кінцем маршруту.

Довжиною маршруту називається число ребер, що входять в маршрут, причому кожне ребро рахується стільки раз, скільки воно входить в даний маршрут.

Замкнутий маршрут називається *циклом.*

Маршрут (цикл), в якому всі ребра різні, називається *простим ланцюгом (циклом)*. Маршрут (цикл), в якому всі вершини (окрім першої і останньої) різні, називається *елементарним ланцюгом (циклом)*.

На рис.4.6 зображено два маршрути з вершини x_1 до вершини x_4: $M_1 = (a_1, a_2, a_4)$ та $M_2 = (a_1, a_2, a_5, a_6)$. Довжина маршруту M_1 дорівнює 3, а довжина маршруту M_2 дорівнює 4.

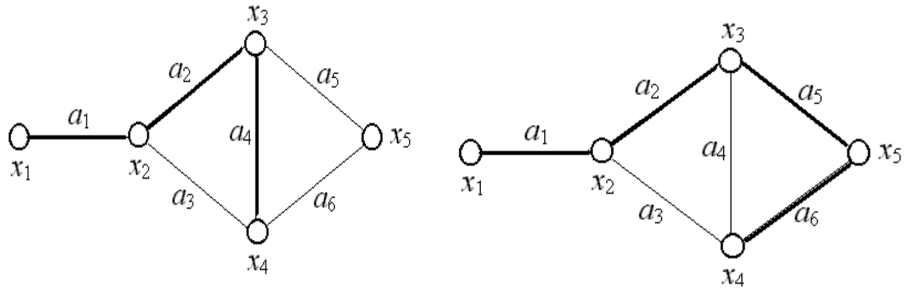
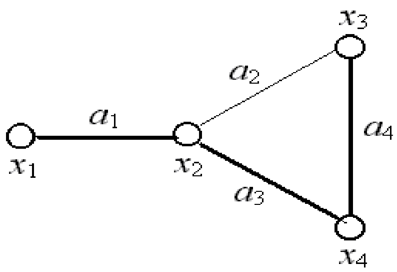
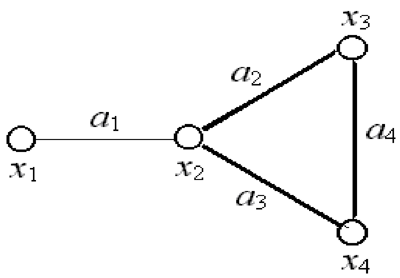


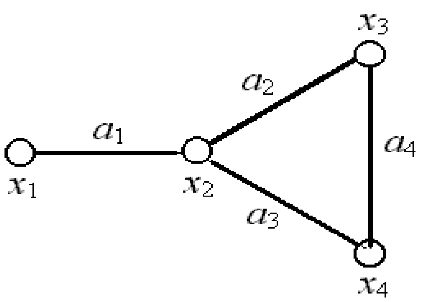
Рисунок 4.6 – Зображення маршрутів з вершини x_1 до вершини x_4



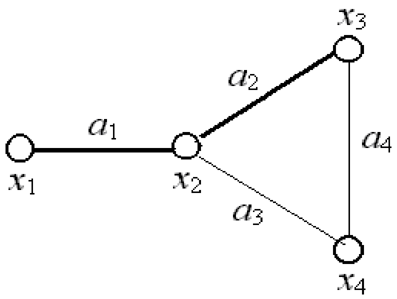
(a_1,a_3,a_4) – простий елементарний ланцюг довжини 3, оскільки всі ребра і вершини попарно різні.



(a_2,a_4,a_3) – простий елементарний цикл, оскільки це замкнутий маршрут, у якого всі ребра і вершини, окрім першої і останньої, різні.



(a_1,a_2,a_4,a_3) – ланцюг, який є простим, але не елементарним, оскільки всі ребра різні, але вершина x_2 зустрічається двічі.



(a_1, a_2, a_2) – маршрут довжини 3, що не являється ні простим, ні елементарним ланцюгом, оскільки ребро a_2 і вершина x_2 зустрічаються двічі.

Поняття шляху і контура в орієнтованому графі аналогічні поняттям маршруту і циклу в неорієнтованому графі.

Шляхом в орієнтованому графі називається послідовність дуг, в якій кінцева вершина всякої дуги, відмінної від останньої, є початковою вершиною наступної дуги.

Число дуг шляху називається *довжиною* шляху.

Шлях називається *контуром*, якщо його початкова вершина співпадає з кінцевою вершиною.

Шлях (контур), в якому всі дуги різні, називається *простим*.

Шлях (контур), в якому всі вершини, окрім першої і останньої, різні, називається *елементарним.*

Поняттям ребра, маршруту, ланцюга, циклу в неорієнтованому графі відповідають поняття дуги, шляху, орієнтованого ланцюга, контура в орієнтованому графі.

Неорієнтований граф	Орієнтований граф
ребро	дуга
маршрут	Шлях
цикл	контур

Граф називається *зв'язним*, якщо кожна пара різних вершин може бути сполучена, принаймні, одним ланцюгом.

Орієнтований граф називається *навантаженим,* якщо дугам цього графа поставлені у відповідність ваги, так що дузі (x_i, x_j) зіставлено деяке число $c(x_i, x_j)$ =

c_{ij}, зване *довжиною* (або *вагою,* або *вартістю* дуги). *Довжиною* (або *вагою* або *вартістю*) шляху s, що складається з деякої послідовності дуг (x_i, x_j), називається число $l(s)$, що дорівнює сумі довжин дуг, що входять в цей шлях, тобто

$$l(s) = \sum c_{ij},$$

причому підсумовування ведеться по всіх дугах $(x_i, x_j) \in s$.

Матриця $C = (c_{ij})$ називається *матрицею довжин дуг* або *матрицею вагів*.

Підграфом неорієнтованого графа G називається граф, всі вершини і ребра якого містяться серед вершин і ребер графа G. Підграф називається *власним*, якщо він відмінний від самого графа. Аналогічно визначається підграф орієнтованого графа.

Компонентою зв'язності неорієнтованого графа називається його зв'язний підграф, що не є власним підграфом ніякого іншого зв'язного підграфа даного графа.

Неорієнтованим деревом (або просто *деревом*) називається зв'язний граф без циклів.

Остовним деревом (деревом-остовом, покриваючим деревом, скелетним деревом) зв'язного графа G називається будь-який його підграф, що містить всі вершини графа G і що є деревом.

4.2 Алгоритм побудови мінімального покриваючого дерева

Нехай G – зв'язний навантажений граф. Завдання побудови *мінімального остовного дерева* полягає в тому, щоб з множини остовних дерев знайти таке, у якого сума довжин ребер мінімальна.

Приведемо типові випадки, коли виникає необхідність побудови мінімального остовного дерева графа:

а) Необхідно з'єднати n міст залізничними лініями (або автомобільними дорогами, лініями електропередач, мережею трубопроводів і так далі) так, щоб сумарна довжина ліній або вартість була б мінімальною.

б) Потрібно побудувати схему електричної (комп'ютерної) мережі, в якої клеми (вузли мережі) повинні бути сполучені за допомогою проводів найменшої загальної довжини.

Задачу побудови мінімального дерева-остову можна вирішити за допомогою *алгоритму Краскала*. Приведемо опис алгоритму по кроках.

Крок 1. Відсортуємо ребра графа по неубуванню вагів.

Крок 2. Вважаємо, що кожна вершина відноситься до своєї компоненти зв'язності.

Крок 3. Проходимо ребра в «відсортованому» порядку. Для кожного ребра виконуємо наступну перевірку:

а) якщо вершини, що сполучаються даним ребром, лежать в різних компонентах зв'язності, то об'єднуємо ці компоненти в одну, а дане ребро додаємо до мінімального дерева-остову;

б) якщо вершини, що сполучаються даним ребром, лежать в одній компоненті зв'язності, то виключаємо ребро з розгляду, оскільки при включенні даного ребра утворюється цикл.

Крок 4. Якщо є ще нерозглянуті ребра і не всі компоненти зв'язності об'єднані в одну, то переходимо до кроку 3, інакше алгоритм завершує роботу:

а) якщо при цьому проглянуті всі ребра, але не всі компоненти зв'язності об'єднані в одну, то для початкового графа неможливо побудувати покриваюче дерево;

б) якщо проглянуті всі ребра, і всі компоненти зв'язності об'єднані в одну, то для початкового графа побудовано мінімальне покриваюче дерево.

Приклад 4.1. Граф G містить 5 вершин. Відстані між вершинами задані таблицею 4.1. Знайти його мінімальне дерево-остов (мінімальне покриваюче дерево).

Таблиця 4.1 – Відстані між вершинами

	1	2	3	4	5
1	0	5	8	2	7
2	5	0	9	2	5
3	8	9	0	10	10
4	2	2	10	0	7
5	7	5	10	7	0

Рішення. Оскільки матриця симетрична, можна розглядати, наприклад, тільки елементи, розташовані вище або нижче за головну діагональ:

Початкові дані

ребро	длина
(1,2)	5
(1,3)	8
(1,4)	2
(1,5)	7
(2,3)	9
(2,4)	2
(2,5)	5
(3,4)	10
(3,5)	10
(4,5)	7

Впорядковані по довжині ребра

ребро	длина	Включить в дерево-остов
(1,4)	2	Да
(2,4)	2	Да
(1,2)	5	Нет
(2,5)	5	Да
(1,5)	7	Нет
(4,5)	7	Нет
(1,3)	8	Да
(2,3)	9	Нет
(3,4)	10	Нет
(3,5)	10	Нет

1. Включаємо в дерево-остов ребро (1, 4). Множина вершин, включених в дерево-остов V={1, 4} (рис.4.7).

2. Наступним кандидатом на включення в дерево-остов є ребро (2, 4). Додавання вершини 2 до множини V і ребра (2, 4) до дерева не створює циклу, оскільки вершина 2 не входить в множину V. Після додавання ребра (2, 4) до дерева множина вершин, включених в дерево-остов V={1, 2, 4} (рис.4.8).

3. Додавання ребра (1, 2) приведе до утворення циклу (рис.4.9). Тому не включаємо це ребро в дерево-остов.

4. Наступним кандидатом на включення в дерево-остов є ребро (2, 5). Його включення не створює циклу, тому V={1, 2, 4, 5} (рис.4.10).

5. Додавання ребра (1, 5) приведе до утворення циклу (рис.4.11). Тому не включаємо це ребро в дерево-остов.

6. Наступним кандидатом на включення в дерево-остов є ребро (4, 5). Проте його додавання приведе до утворення циклу. Тому не включаємо це ребро в дерево-остов.

7. Включення ребра (1, 3) не створює циклу, тому *V*={1, 2, 3, 4, 5} (рис.4.12).

8. Оскільки всі вершини графа увійшли до дерева, то отримано покриваюче дерево з мінімальною вагою, рівною 17.

Рисунок 4.7

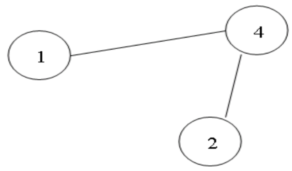


Рисунок 4.8

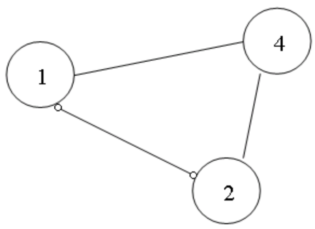


Рисунок 4.9

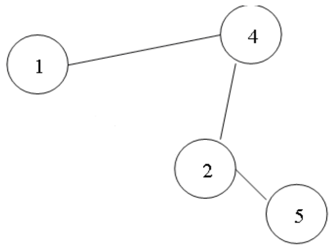


Рисунок 4.10

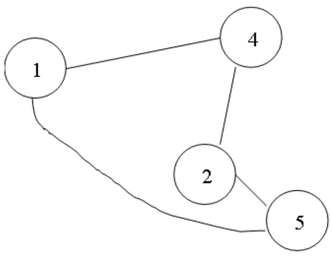


Рисунок 4.11

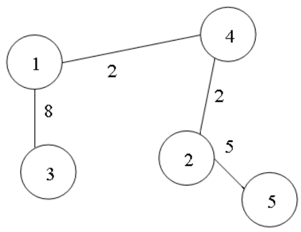


Рисунок 4.12

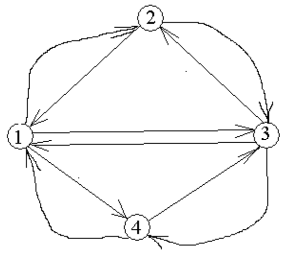
4.3 Алгоритми визначення найкоротшого шляху (Дейкстри та Флойда) між вузлами мережі

Пошук шляхів із заданою кількістю дуг

Нехай $G = (X, A)$ – зв'язний граф. Для визначення кількості шляхів, що складаються з k дуг, необхідно звести в k-у ступінь матрицю суміжності. Тоді її елемент s_{ij}^k дасть кількість шляхів довжини k із вершини x_i до вершини x_j.

Приклад 4.2. Для графа, приведеного на рис.4.13 зліва, знайти всі шляхи довжини 3 (тобто, знайти всі шляхи, що містять рівно три дуги).

Рішення. Матриця суміжності для даного графа має вигляд, представлений на рис.4.13 справа.



$$S^1 = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

Рисунок 4.13

Тоді S^2 та S^3 виглядають так:

$$S^2 = \begin{pmatrix} 3 & 1 & 2 & 1 \\ 1 & 2 & 1 & 2 \\ 2 & 1 & 3 & 1 \\ 1 & 2 & 1 & 2 \end{pmatrix}, \; S^3 = \begin{pmatrix} 4 & 5 & 5 & 5 \\ 5 & 2 & 5 & 2 \\ 5 & 5 & 4 & 5 \\ 5 & 2 & 5 & 2 \end{pmatrix}$$

Значення $s_{11}^3 = 4$ означає, що з вершини 1 у вершину 1 існує 4 шляхи довжини 3, $s_{12}^3 = 5$— з вершини 1 у вершину 2 – 5 шляхів довжини 3 і так далі

Щоб виявити ці шляхи, слід позначити дуги, наприклад, так, як на рис.4.14. Замість матриці суміжності введемо в розгляд матрицю, елементами якої є дуги вигляду u_r, $r = 1, 2, \ldots, 10$ (рис.4.15).

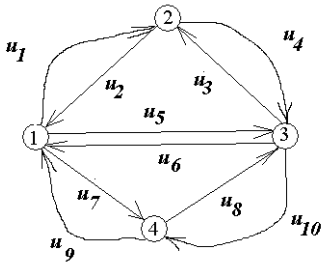


Рисунок 4.14

$$S^1 = \begin{pmatrix} 0 & u_1 & u_5 & u_7 \\ u_2 & 0 & u_4 & 0 \\ u_6 & u_3 & 0 & u_{10} \\ u_9 & 0 & u_8 & 0 \end{pmatrix}$$

Рисунок 4.15

Виконуємо символьне множення матриць:

$$S^2 = \begin{pmatrix} 0 & u_1 & u_5 & u_7 \\ u_2 & 0 & u_4 & 0 \\ u_6 & u_3 & 0 & u_{10} \\ u_9 & 0 & u_8 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & u_1 & u_5 & u_7 \\ u_2 & 0 & u_4 & 0 \\ u_6 & u_3 & 0 & u_{10} \\ u_9 & 0 & u_8 & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} u_1u_2 + u_5u_6 + u_7u_9 & u_5u_3 & u_1u_4 + u_7u_8 & u_5u_{10} \\ u_4u_6 & u_2u_1 + u_4u_3 & u_2u_5 & u_2u_7 + u_4u_{10} \\ u_3u_2 + u_{10}u_9 & u_6u_1 & u_6u_5 + u_3u_4 + u_{10}u_8 & u_6u_7 \\ u_8u_6 & u_9u_1 + u_8u_3 & u_9u_5 & u_9u_7 + u_8u_{10} \end{pmatrix}$$

$$S^3 = \begin{pmatrix} u_1u_2 + u_5u_6 + u_7u_9 & u_5u_3 & u_1u_4 + u_7u_8 & u_5u_{10} \\ u_4u_6 & u_2u_1 + u_4u_3 & u_2u_5 & u_2u_7 + u_4u_{10} \\ u_3u_2 + u_{10}u_9 & u_6u_1 & u_6u_5 + u_3u_4 + u_{10}u_8 & u_6u_7 \\ u_8u_6 & u_9u_1 + u_8u_3 & u_9u_5 & u_9u_7 + u_8u_{10} \end{pmatrix} \times$$

$$\times \begin{pmatrix} 0 & u_1 & u_5 & u_7 \\ u_2 & 0 & u_4 & 0 \\ u_6 & u_3 & 0 & u_{10} \\ u_9 & 0 & u_8 & 0 \end{pmatrix}.$$

У таблиці 4.2 приведена матриця S^3 по стовпцях.

Таблиця 4.2 – Представлення матриці S^3 по стовпцях

$u_5u_3u_2+u_1u_4u_6+u_7u_8u_6+u_5u_{10}u_9$ $u_2u_1u_2+u_4u_3u_2+u_2u_5u_6+u_2u_7u_9+u_4u_{10}$ $u_6u_1u_2+u_6u_5u_6+u_3u_4u_6+u_{10}u_8u_6+u_6u_7$ $u_9u_1u_2+u_8u_3u_2+u_9u_5u_6+u_9u_7u_9+u_8u_{10}$	1 стовпець
$u_1u_2u_1+u_5u_6u_1+u_7u_9u_1+u_1u_4u_3+u_7u_8u_3$ $u_4u_6u_1+u_2u_5u_3$ $u_3u_2u_1+u_{10}u_9u_1+u_6u_5u_3+u_3u_4u_3+u_{10}u_8u$ $u_8u_6u_1+u_9u_5u_3$	2 стовпець
$u_1u_2u_5+u_5u_6u_5+u_7u_9u_5+u_5u_3u_4+u_5u_{10}u_8$ $u_4u_6u_5+u_2u_1u_4+u_4u_3u_4+u_2u_7u_8+u_4u_{10}u_8$ $u_3u_2u_5+u_{10}u_9u_5+u_6u_1u_4+u_6u_7u_8$ $u_8u_6u_5+u_9u_1u_4+u_8u_3u_4+u_9u_7u_8+u_8u_{10}u_8$	3 стовпець
$u_1u_2u_7+u_5u_6u_7+u_7u_9u_7+u_1u_4u_{10}+u_7u_8u$ $u_4u_6u_7+u_2u_5u_{10}$ $u_3u_2u_7+u_{10}u_9u_7+u_6u_5u_{10}+u_3u_4u_{10}+u_{10}u_8$ $u_8u_6u_7+u_9u_5u_{10}$	4 стовпець

Відмітимо, що число доданків в кожному елементі отриманої матриці рівне числу елементів матриці S^3:

$$S^3=\begin{pmatrix}4&5&5&5\\5&2&5&2\\5&5&4&5\\5&2&5&2\end{pmatrix}$$

Розглянемо, наприклад, суму

$$u_5u_3u_2+u_1u_4u_6+u_7u_8u_6+u_5u_{10}u_9\,.$$

Вона відповідає чотирьом шляхам з вершини 1 у вершину 1:

1-u_5-3-u_3-2-u_2-1, 1-u_1-2-u_4-3-u_6-1, 1-u_7-4-u_8-3-u_6-1, 1-u_5-3-u_{10}-4-u_9-1.

Алгоритм Дейкстри для пошуку найкоротшого шляху між заданою парою вершин

Нехай *G = (X, A)* – зв'язний граф, кожній дузі якого приписано деяке число *a(x, y)≥ 0*. Завдання побудови найкоротшого шляху між заданою парою вершин

$s \in X$ і $t \in X$ полягає в тому, щоб з множини шляхів, що сполучають вказані вершини, знайти такий, сумарна довжина дуг якого мінімальна.

Для вирішення завдання можна скористатися *алгоритмом Дейкстри*. Приведемо його опис по кроках.

Крок 1. Перед початком виконання алгоритму всі вершини не помічені. Кожній вершині $x \in X$ в ході виконання алгоритму привласнюється число $d(x)$, рівне довжині найкоротшого шляху з s в x, що включає тільки помічені вершини.

Покласти $d(s)=0$, $d(x)=\infty$ для всіх x, відмінних від s. Помітити вершину s і покласти $y = s$ (y – остання із помічених вершин).

Крок 2. Для кожної непоміченої вершини x таким чином перерахувати величину $d(x)$:

$$d(x) = \min \{d(x), d(x)+a(y, x)\}.$$

Якщо $d(x)=\infty$ для всіх непомічених вершин x, закінчити процедуру алгоритму: у заданому графі відсутні шляхи між вказаною парою вершин. Інакше помітити ту з вершин x, для якої величина $d(x)$ є найменшою. Покласти $y = x$.

Крок 3. Якщо $y = t$, закінчити процедуру: найкоротший шлях з вершини s в вершину t знайдено. Інакше перейти до кроку 2.

Приклад 4.3. В заданому графі G (рис.4.16) знайти найкоротший шлях між вершинами 1 і 10.

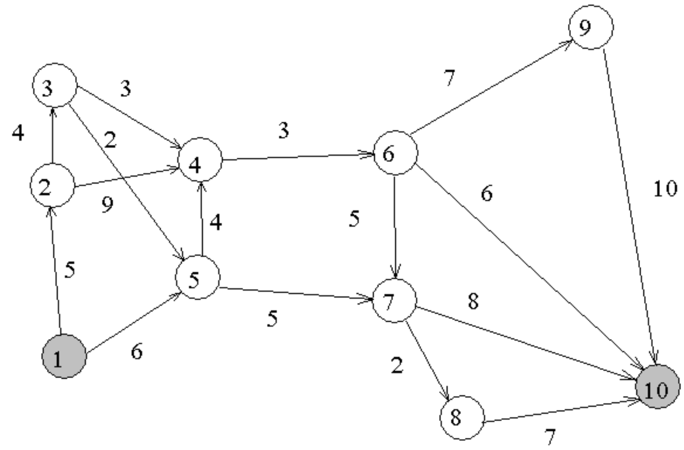


Рисунок 4.16 – Заданий граф

Рішення. Перед початком виконання алгоритму вважаємо $d(1)=0$, $d(x)=\infty$ для всіх $x\neq 1$; вершина 1 – остання із помічених вершин.

$d(2)= min\ \{d(2), d(1)+c(1, 2)\}= min\ \{\infty, 0+5\}=5;$

$d(5)= min\{d(5), d(1)+c(1, 5)\}= min\ \{\infty, 0+6\}=6;$

$d(x)= \infty$ для всіх $x\neq 1, 2, 5$.

Оскільки мінімум випав на вершину 2, то $y=2$ – остання із помічених вершин (рис.4.17).

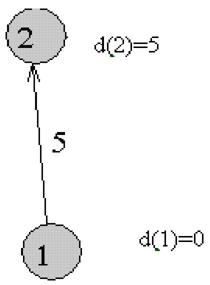


Рисунок 4.17

$d(3)=min\{d(3), d(2)+c(2, 3)\}=min\ \{\infty, 5+4\}=9;$

$d(5)= min\{d(5), d(2)+c(2, 5)\}=min\ \{6, 5+\infty\}=6;$

$d(4)= min\{d(4), d(2)+c(2, 4)\}=min\ \{\infty, 5+9\}=14;$

$d(x)= \infty$ для всіх $x\neq 1, 2, 3, 4, 5$.

Оскільки мінімум випав на вершину 5, то $y=5$ – остання із помічених вершин (рис.4.18).

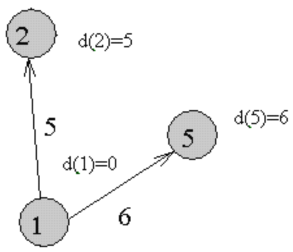


Рисунок 4.18

$d(3)=min\{d(3), d(5)+c(5, 3)\}=min\ \{9, 5+\infty\}=9;$

$d(4)= min\{d(4), d(5)+c(5, 4)\}=min\ \{14, 6+4\}=10;$

d(7)= min{d(7), d(5)+c(5, 7)}=min {∞, 6+5}=11;

d(x)= ∞ для всіх *x≠1, 2, 3, 4, 5, 7.*

Оскільки мінімум випав на вершину 3, то *y=3* – остання із помічених вершин (рис.4.19).

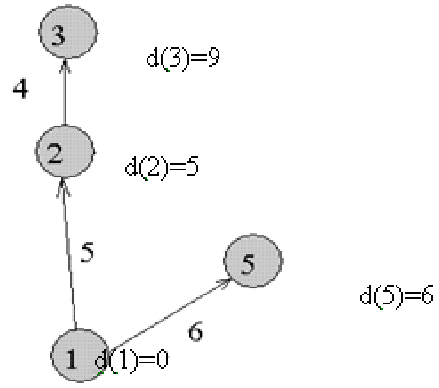


Рисунок 4.19

d(4)=min{d(4), d(3)+c(3, 4)}=min {10, 9+3}=10;

d(5)=min{d(5), d(3)+c(3, 5)}=min {6, 9+2}=6

(оцінка не зменшилася, тобто кращий шлях не знайдений);

d(7)= min{d(7), d(3)+c(3, 7)}=min {11, 9+∞}=11;

d(x)= ∞ для всіх *x≠1, 2, 3, 4, 5, 7.*

Оскільки мінімум випав на вершину 4 (з непомічених), то *y=4* – остання із помічених вершин (рис.4.20).

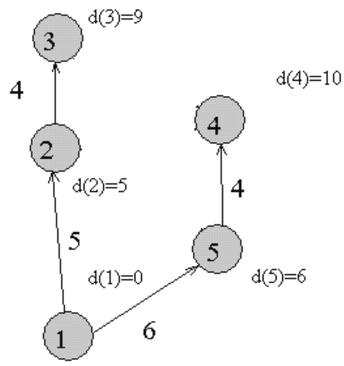


Рисунок 4.20

d(6)=min{d(6), d(4)+c(4, 6)}=min {∞, 10+3}=13;

d(7)= min{d(7), d(4)+c(4, 7)}=min {11, 10+∞}=11;

d(x)= ∞ для всіх *x≠1, 2, 3, 4, 5, 6, 7.*

Оскільки мінімум випав на вершину 7, то *y=7* – остання із помічених вершин (рис.4.21).

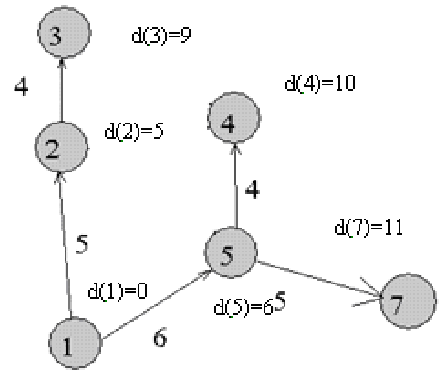


Рисунок 4.21

d(6)=min{d(6), d(7)+c(7, 6)}=min {13,11+∞}=13;

d(8)=min{d(8), d(7)+c(7, 8)}=min(∞,11+2}=13;

d(10)=min{d(10), d(7)+c(7, 10)}=min {∞,11+8}=19;

d(x)= ∞ для *x=9.*

Оскільки мінімум випав на вершину 6, то *y=6* – остання із помічених вершин (рис.4.22).

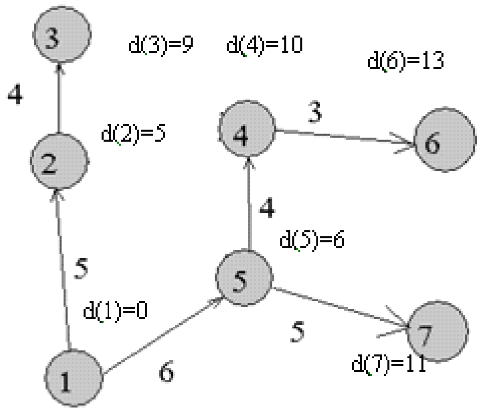


Рисунок 4.22

d(7)= min{d(7), d(6)+c(6, 7)}=min {11,13+5}=11

(оцінка не зменшилася, тобто кращий шлях не знайдений);

d(8)=min{d(8), d(6)+c(6, 8)}=min {13,13+∞}=13;

d(9)=min{d(9), d(6)+c(6, 9)}=min {∞,13+7}=20;

d(10)=min{d(10), d(6)+c(6,10)}=min {19,13+6}=19.

Оскільки мінімум випав на вершину 8 (з непомічених), то *y=8* – остання із помічених вершин (рис.4.23).

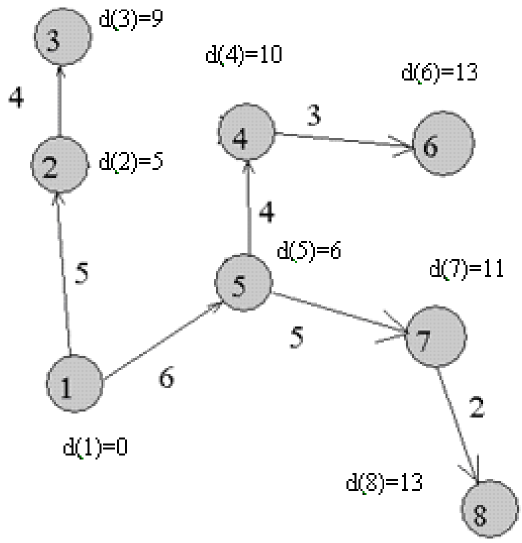


Рисунок 4.23

d(9)=min{d(9), d(8)+c(8, 9)}=min {20,13+∞}=20;

d(10)=min{d(10), d(8)+c(8,10)}=min {19,13+7}=19.

Оскільки мінімум випав на вершину 10, то *y=10*, і на цьому алгоритм Дейкстри закінчує роботу (рис.4.24).

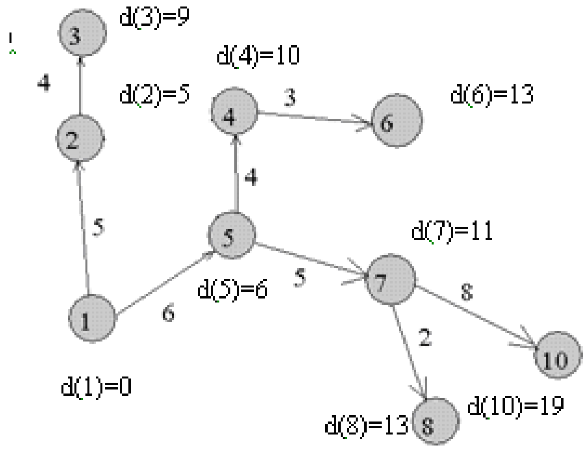


Рисунок 4.24

Таким чином, найкоротший шлях з вершини 1 у вершину 10 проходить через проміжні вершини 5 і 7 (рис.4.25). Довжина цього шляху дорівнює 19.

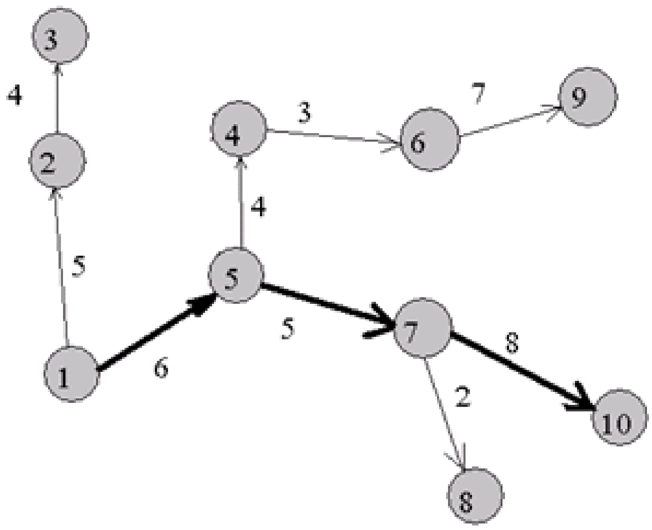


Рисунок 4.25

Пошук всіх найкоротших шляхів (алгоритм Флойда)

Перенумеруємо всі вершини початкового графа цілими числами від 1 до *n*. Позначимо через d_{ij}^{m} довжину найкоротшого шляху з вершини *i* до вершини *j*, який як проміжні може містити тільки перші *m* вершин графа. Якщо ж між вершинами *i* та *j* не існує жодного шляху вказаного типу, то умовно вважатимемо, що $d_{ij}^m = \infty$. З даного визначення величини d_{ij}^m витікає, що величина d_{ij}^0 є довжиною найкоротшого шляху з вершини *i* до вершини *j*, що не має проміжних вершин, тобто, довжиною найкоротшої дуги, що сполучає вершини *i* та *j* (якщо такі дуги присутні в графі). Вважатимемо, що $d_{ij}^0 \geq 0$ для всіх *i* та *j* ($1 \leq i \neq j \leq n$). Для будь-якої вершини *i* покладемо $d_{ii}^0 = 0$.

Позначимо через D^m матрицю розмірності $n \times n$, елемент якої, розташований на перетині *i*-го рядка та *j*-го стовпця, співпадає з d_{ij}^m. Якщо в початковому графові відома довжина кожної дуги, то можна сформувати матрицю D^0. Наша мета полягає у визначенні матриці D^n, що представляє найкоротші шляхи між всіма вершинами початкового графа. У алгоритмі Флойда як початкова виступає матриця D^0. Потім по ній обчислюється матриця D^1, а по ній – матриця D^2 і так далі. Процес повторюється до тих пір, поки не буде обчислена матриця D^n.

Припустимо, що нам відомі:

1)найкоротший шлях з вершини *i* у вершину *m*, в якому як проміжні допускається використання тільки перших (*m-1*) вершин;

2)найкоротший шлях з вершини *m* у вершину *j*, в якому як проміжні допускається використання тільки перших (*m-1*) вершин;

3)найкоротший шлях з вершини *i* у вершину *j*, в якому як проміжні допускається використання тільки перших (*m-1*) вершин.

По припущенню початковий граф не може містити контурів від’ємної довжини. Отже, один з двох шляхів – шлях, що співпадає з шляхом, описаним в пункті 3, або шлях, що є об’єднанням шляхів з пунктів 1–2, – повинен бути найкоротшим шляхом з вершини *i* у вершину *j*, в якому як проміжні допускається використання тільки перших *m* вершин. Таким чином

$$d_{ij}^{m} = \min\left\{d_{im}^{m-1} + d_{mj}^{m-1}, d_{ij}^{m-1}\right\}.$$

Приведемо покроковий опис *алгоритму Флойда*:

Крок 1. Пронумерувати вершини початкового графа цілими числами від 1 до *n*. Визначити матрицю D^0, задавши величину кожного її елементу d_{ij}^0 рівній довжині найкоротшої дуги, що сполучає вершини *i* і *j*. Якщо в початковому графі вказані вершини не з'єднуються дугами, покласти $d_{ij}^0 = \infty$. Крім того, покласти $d_{ii}^0 = 0$ для всіх i ($1 \le i \le n$).

Крок 2. Для цілого *m*, послідовно приймаючого значення 1, 2, ..., *n*, визначити по величинах елементів матриці D^{m-1} величини елементів матриці D^m, використовуючи співвідношення

$$d_{ij}^{m} = \min\left\{d_{im}^{m-1} + d_{mj}^{m-1}, d_{ij}^{m-1}\right\}.$$

При визначенні величини кожного елементу матриці D^m фіксувати відповідний найкоротший шлях. Після закінчення даної процедури величина елементу d_{ij}^n матриці D^n визначає величину найкоротшого шляху, що веде з вершини *i* у вершину *j*.

Рядки і стовпці матриці D^m для яких $i = m$ і $j = m$, називатимемо базовими. Неважко відмітити, що в таких рядках і стовпцях значення матриці можна не перераховувати, оскільки вони повністю співпадають з відповідними значеннями матриці D^{m-1}.

Приклад 4.4. Для графа, приведеного на рис.4.26, знайти найкоротші шляхи між будь-якою парою вершин.

$$D^0 = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 2 & 0 & 7 & \infty \\ 6 & 5 & 0 & 2 \\ 1 & \infty & 4 & 0 \end{pmatrix}$$

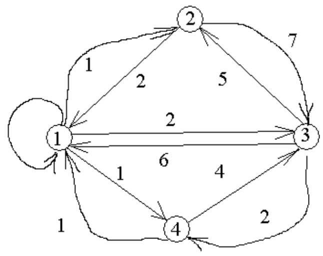


Рисунок 4.26

Рішення. Матриця D^0 для даного графа приведена на рис.4.26 зліва. Весь процес обчислень приведений в таблиці 4.3.

Таблиця 4.3 – Процес обчислень матриці D^1

$d^1_{ij} = \min\{d^0_{i1} + d^0_{1j}, d^0_{ij}\}$	Відповідні шляхи
$d^1_{11} = d^0_{11} = 0$	
$d^1_{12} = d^0_{12} = 1$	(1, 2)
$d^1_{13} = d^0_{13} = 2$	(1, 3)
$d^1_{14} = d^0_{14} = 1$	(1, 4)
$d^1_{21} = d^0_{21} = 2$	(2, 1)
$d^1_{22} = 0$	
$d^1_{23} = \min\{d^0_{21} + d^0_{13}, d^0_{23}\} = \min\{2+2, 7\} = 4$	(2, 1), (1, 3)
$d^1_{24} = \min\{d^0_{21} + d^0_{14}, d^0_{24}\} = \min\{2+1, \infty\} = 3$	(2, 1), (1, 4)
$d^1_{31} = d^0_{31} = 6$	(3, 1)
$d^1_{32} = \min\{d^0_{31} + d^0_{12}, d^0_{32}\} = \min\{6+1, 5\} = 5$	(3, 2)
$d^1_{33} = 0$	
$d^1_{34} = \min\{d^0_{31} + d^0_{14}, d^0_{34}\} = \min\{6+1, 2\} = 2$	(3, 4)
$d^1_{41} = d^0_{41} = 1$	(4, 1)

$d_{42}^{1}=\min\{d_{41}^{0}+d_{12}^{0},d_{42}^{0}\}=\min\{1+1,\infty\}=2$	(4, 1), (1, 2)
$d_{43}^{1}=\min\{d_{41}^{0}+d_{13}^{0},d_{43}^{0}\}=\min\{1+2,4\}=3$	(4, 1), (1, 3)
$d_{44}^{1}=0$	

Аналогічно визначаються матриці D^2, D^3, D^4. Отримані результати приводяться в таблиці 4.4.

Таблиця 4.3 – Визначення матриць D^2, D^3, D^4.

Матриці	Найкоротші шляхи
$D^2=\begin{pmatrix}0&1&2&1\\2&0&4&3\\6&5&0&2\\1&2&3&0\end{pmatrix}$	$\begin{pmatrix} & (1,2) & (1,3) & (1,4)\\(2,1) & & (2,1),(1,3) & (2,1),(1,4)\\(3,1) & (3,2) & & (3,4)\\(4,1) & (4,1),(1,2) & (4,1),(1,3) & \end{pmatrix}$
$D^3=\begin{pmatrix}0&1&2&1\\2&0&4&3\\6&5&0&2\\1&2&3&0\end{pmatrix}$	$\begin{pmatrix} & (1,2) & (1,3) & (1,4)\\(2,1) & & (2,1),(1,3) & (2,1),(1,4)\\(3,1) & (3,2) & & (3,4)\\(4,1) & (4,1),(1,2) & (4,1),(1,3) & \end{pmatrix}$
$D^4=\begin{pmatrix}0&1&2&1\\2&0&4&3\\3&4&0&2\\1&2&3&0\end{pmatrix}$	$\begin{pmatrix} & (1,2) & (1,3) & (1,4)\\(2,1) & & (2,1),(1,3) & (2,1),(1,4)\\(3,4),(4,1) & (3,4),(4,1),(1,2) & & (3,4)\\(4,1) & (4,1),(1,2) & (4,1),(1,3) & \end{pmatrix}$

Для вирішення реальних завдань приведений вище спосіб формування найкоротших шляхів малопридатний.

Зручно ввести в розгляд матрицю маршрутів *R*. На *m-ій* ітерації вона визначається як $R^m=\{r_{ij}^m\}$ де r_{ij}^m — перший проміжний вузол найкоротшого шляху з *i* до *j*, вибраний серед множини {*1, 2., m*} ($i \neq j \neq m$). Алгоритм починає роботу при $R^0 = \left(r_{ij}^0\right)$, де $r_{ij}^0 = j$. На *m-ій* ітерації елемент r_{ij}^m може бути отриманий з наступного співвідношення:

$$r_{ij}^m = \begin{cases} m, \text{ я як } & d_{ij}^{m-1} > d_{im}^{m-1} + d_{mj}^{m-1} \\ r_{ij}^{m-1} & \text{інакше.} \end{cases}$$

Для розглянутого вище прикладу маємо:

$$D^0 = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 2 & 0 & 7 & \infty \\ 6 & 5 & 0 & 2 \\ 1 & \infty & 4 & 0 \end{pmatrix} \qquad R^0 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 \end{pmatrix}$$

$$D^1 = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 2 & 0 & 4 & 3 \\ 6 & 5 & 0 & 2 \\ 1 & 2 & 3 & 0 \end{pmatrix} \qquad R^1 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 4 \end{pmatrix}$$

$$D^2 = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 2 & 0 & 4 & 3 \\ 6 & 5 & 0 & 2 \\ 1 & 2 & 3 & 0 \end{pmatrix} \qquad R^2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 4 \end{pmatrix}$$

$$D^3 = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 2 & 0 & 4 & 3 \\ 6 & 5 & 0 & 2 \\ 1 & 2 & 3 & 0 \end{pmatrix} \qquad R^3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 4 \end{pmatrix}$$

$$D^4 = \begin{pmatrix} 0 & 1 & 2 & 1 \\ 2 & 0 & 4 & 3 \\ 3 & 4 & 0 & 2 \\ 1 & 2 & 3 & 0 \end{pmatrix} \qquad R^4 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 1 & 1 \\ 4 & 4 & 3 & 4 \\ 1 & 1 & 1 & 4 \end{pmatrix}$$

Щоб скористатися матрицею R^4 для визначення шляху, наприклад, з вершини 3 у вершину 2, проглядаємо елемент r_{32}^4. Оскільки він рівний 4, це означає, що вершина з номером 4 є першою проміжною в цьому шляху. Далі проглядаємо елемент r_{42}^{4}. Оскільки він рівний 1, це означає, що вершина з номером 1 є другою проміжною в цьому шляху. Далі проглядаємо елемент r_{12}^{4}. Оскільки елемент рівний номеру кінцевої вершини 2, отримуємо, що шуканий шлях проходить по дугах (3, 4), (4, 1) і (1, 2) (що відповідає результату з таблиці 4.3).

4.4 Задача про максимальний потік та мінімальний розріз в мережі

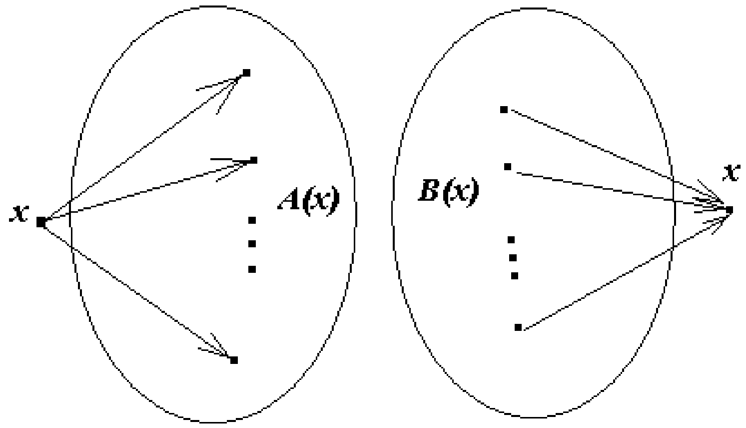
Нехай *G = (N, A)* – зв'язний граф, в якому виділено дві вершини: *s* – джерело, *t* – стік. Кожній дузі (*x*, *y*) графа поставлено у відповідність невід'ємне число *c*(*x*, *y*), яке інтерпретується як максимальна кількість одиниць деякого товару, яке може бути доставлено з вершини *x* у вершину *y* за одиницю часу. Це число прийнято називати пропускною спроможністю дуги. Такий граф називають мережею, а його вершини – вузлами.

Стаціонарний потік величини *v* з *s* в *t* в мережі *G = (N, A)* є функція *f*, що відображає множину *A* в множину невід'ємних чисел, що задовольняють лінійним рівнянням і нерівностям наступного вигляду:

$$\sum_{x\in A(x)} f(x, y) - \sum_{y\in B(x)} f(y, \text{x}) = \begin{cases} v, x = s; \\ 0, \ x \neq s, t; \\ -v, x = t; \end{cases}$$

$$f(x, y) \le c(x, y) \text{ для всіх } (x, y) \in A,$$

де *A(x)* – підмножина множини *N*, що включає вершини *y*, що є кінцями дуг з початком у вершині *x*; *B(x)* – підмножина множини *N*, що включає вершини *y*, що є початком дуг, що входять у вершину *x*.



Розрізом $(X, \overline{X})$ у мережі $G = (N, A)$, що відокремлює вузли s і t, називається множина дуг, де $s \in X, t \in \overline{X}$. При цьому $X \cap \overline{X} = \varnothing,\ X \cup \overline{X} = N$.

Приведемо формулювання *теореми про максимальний потік і мінімальний розріз*.

Теорема. Для будь-якої мережі максимальна величина потоку з s до t рівна мінімальній пропускній спроможності розрізу, відокремлюючого s і t.

Завдання побудови максимального потоку між заданою парою вершин $s \in N$ і $t \in N$ полягає в тому, щоб з множини шляхів, що сполучають вказані вершини, знайти такі, по яких можна пропустити максимальну кількість одиниць потоку в одиницю часу. При цьому повинні дотримуватися наступні обмеження:

- потік по кожній дузі не повинен перевищувати її пропускну спроможність;
- потік з джерела s рівний потоку, що приходить в стік t;
- для проміжних вершин кількість одиниць потоку, що потрапив в цей вузол, повинна в точності дорівнювати кількості одиниць потоку, що вийшов з цього вузла.

Для вирішення завдання можна скористатися *алгоритмом розстановки позначок Форда-фалкерсона*.

Алгоритм може починати роботу з нульового потоку. Потім обчислення розвиваються у вигляді послідовності «розстановки позначок» (операція A), кожна з яких або приводить до потоку з більшою величиною (операція B), або закінчується висновком про те, що даний потік максимальний. Всі вузли мережі знаходяться в одному з наступних станів: не помічений, помічений і не проглянутий, помічений і проглянутий. Спочатку всі вузли не помічені.

Операція A (процес розстановки позначок). Джерело *s* отримує позначку *(-, ε(s)=∞)* (джерело тепер помічене і не проглянуте, решта вузлів не помічена). Вибираємо будь-який помічений і не проглянутий вузол *x*. Хай він має позначку $(x^{\pm}, \varepsilon(x))$. Всім вузлам *y*, які не помічені і для яких *f(x, y) < c(x, y)* приписуємо позначку $(x^{+}, \varepsilon(y))$, де

$$\varepsilon(y) = \min\{\varepsilon(x), c(x, y)\text{-}f(x, y)\}$$

(такі вузли *y* тепер помічені і не проглянуті). Всім вузлам *y*, які після цього не помічені і для яких *f(y, x) > 0* приписуємо позначку $(x^{-}, \varepsilon(y))$, де

$$\varepsilon(y) = \min\{\varepsilon(x),\ f(y, x)\}$$

(такі вузли у тепер помічені і не проглянуті, а вузол *x* після цього помічений і проглянутий).

Цей загальний крок повторюваний до тих пір, поки

1) не виявиться поміченим і не проглянутим стік *t;*

2) або ж до тих пір, поки не можна буде помітити жоден вузол, а стік *t* залишиться непоміченим.

У першому випадку переходимо до операції *B*, а в другому – алгоритм закінчив роботу, оскільки максимальний потік в мережі отриманий.

Операція B (зміна потоку). Хай стік *t* має позначку $(y^{\pm}, \varepsilon(t))$. Якщо він має позначку $(y^{+}, \varepsilon(t))$, то *f(y, t)* замінюємо на *f(y, t)+ε(t);* якщо він має позначку $(y^{-}, \varepsilon(t))$, то *f(y, t)* замінюємо на *f(y, t)-ε(t).* У будь-якому з цих випадків переходимо до вузла *y*. Взагалі, якщо вузол *y* має позначку $(x^{+}, \varepsilon(y))$, то *f(x, y)* замінюємо на *f(x, y)+ε(t),* а якщо він має позначку $(x^{-}, \varepsilon(y))$, то *f(x, y)* замінюємо на *f(x, y)-ε(t)* і переходимо до вузла *x*. Коли ми досягнемо джерела *s*, зміна потоку припиняється. Потрібно стерти всі старі позначки і знов перейти до операції *A*.

Примітка. Пропускні спроможності дуг повинні бути цілими невід'ємними числами.

Приклад 4.5. Знайти максимальний потік і мінімальний розріз, що відокремлює джерело 1 і стік 6 для мережі, приведеної на рис.4.27. Пара чисел біля кожної дуги означає пропускну спроможність і потік по дузі відповідно.

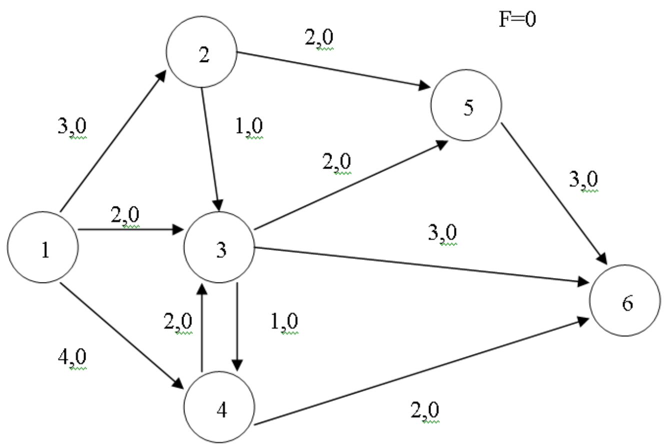


Рисунок 4.27 – Вихідна мережа

Рішення. На рис.4.28 показані позначки, отримані в ході виконання операції *A* на першій стадії роботи алгоритму. Джерело (вузол 1) отримало позначку (-, ∞). Потім з вузла 1 позначаються вузли 2, 3 і 4. З вузла 2 позначається вузол 5, а з вузла 3 – стік (вузол 6). Оскільки стік виявився поміченим, переходимо до операції *B*. Стік отримав позначку (3+, 2). Отже, *f(3, 6)* стане рівним 0+2=2. Переходимо до вузла 3. Оскільки він має позначку (1+, 2), то потік по дузі (1, 3) стане рівним 0+2=2.

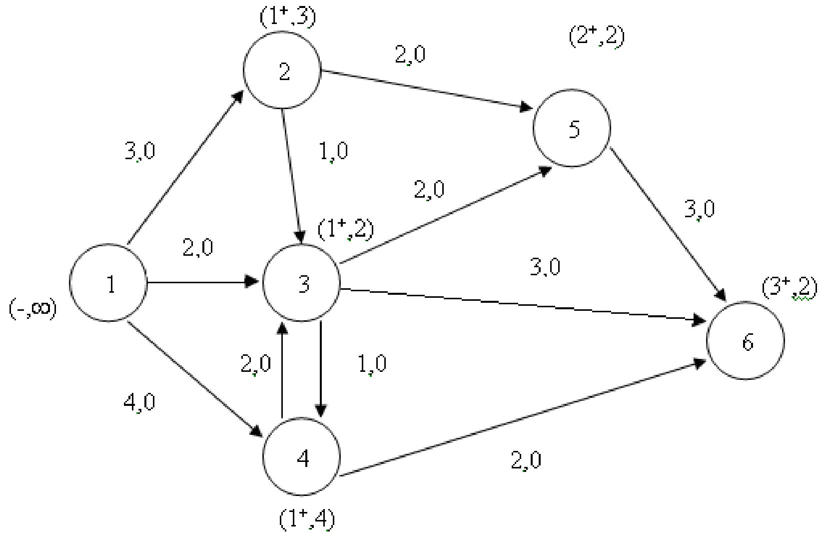


Рисунок 4.28 – Позначки, отримані в ході виконання операції *A* на першій стадії роботи алгоритму

Оскільки в результаті виконання операції *B* досягнуте джерело (вузол 1), то слід стерти старі позначки і знову перейти до операції *A*. Стан мережі приведений на рис.4.29 Жирними стрілками вказані дуги, по яких йдуть дві одиниці потоку.

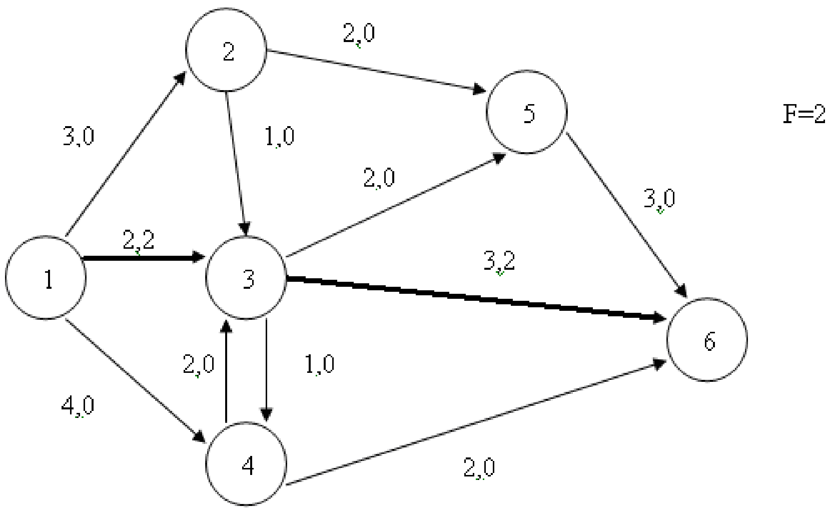


Рисунок 4.29 – Стан мережі після виконання операції *B*

Знову виконуємо операцію *A*. Джерело (вузол 1) отримало позначку (-, ∞). Потім з вузла 1 позначаються вузли 2 і 4 (рис.4.30). З вузла 2 позначаються вузли 3 і 5, а з вузла 3 – стік (вузол 6). Оскільки стік виявився поміченим, переходимо до операції B. Стік отримав позначку (3^+, 1). Отже, *f(3, 6)* стане рівним 2+1=3. Переходимо до вузла 3. Оскільки він має позначку (2^+, 1), то потік по дузі (2, 3) стане рівним 0+1=1. Переходимо до вузла 1 і збільшуємо потік по дузі (1^+, 3), то потік по дузі (1, 2): він стане рівним 0+1=1. Отже, потік в мережі став рівний 3. На рис.4.30 жирними стрілками вказано нове збільшення потоку в мережі.

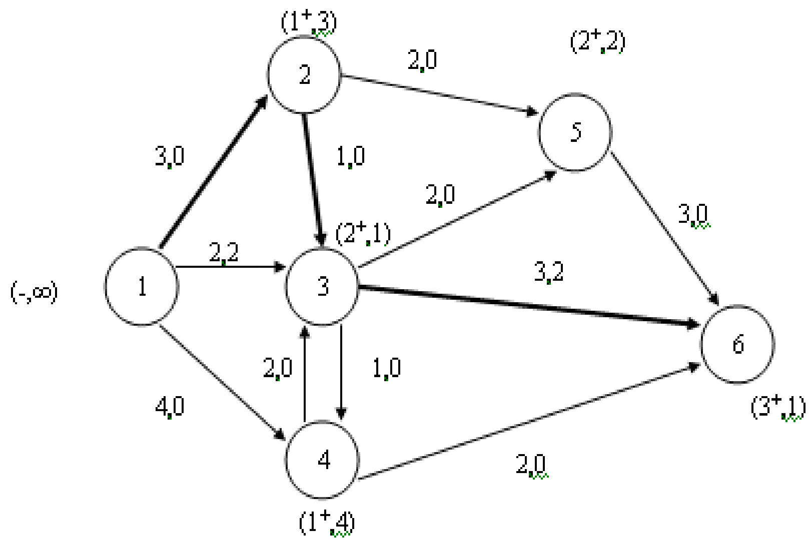


Рисунок 4.30 – Позначення вузлів 2 і 4

На рис.4.31 показаний новий стан мережі з потоком в 3 одиниці.

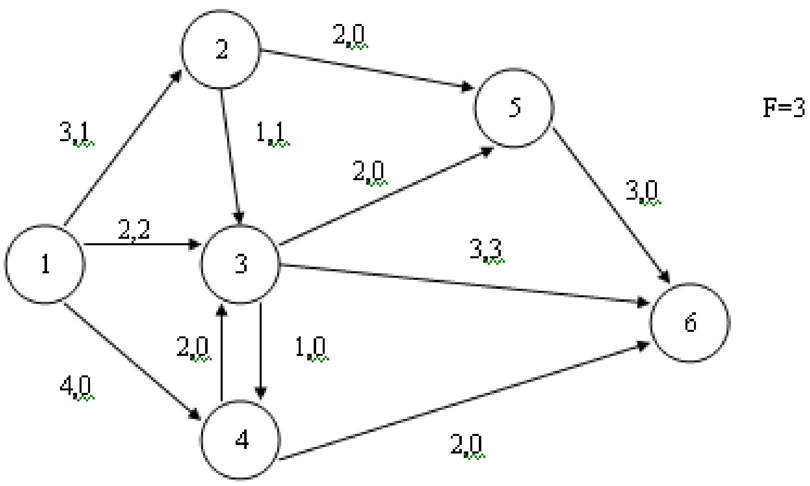


Рисунок 4.31 – Новий стан мережі з потоком в 3 одиниці

На рис.4.32–4.37 показано виконання операцій *A* і *B*, які привели до збільшення потоку в мережі до 8 одиниць.

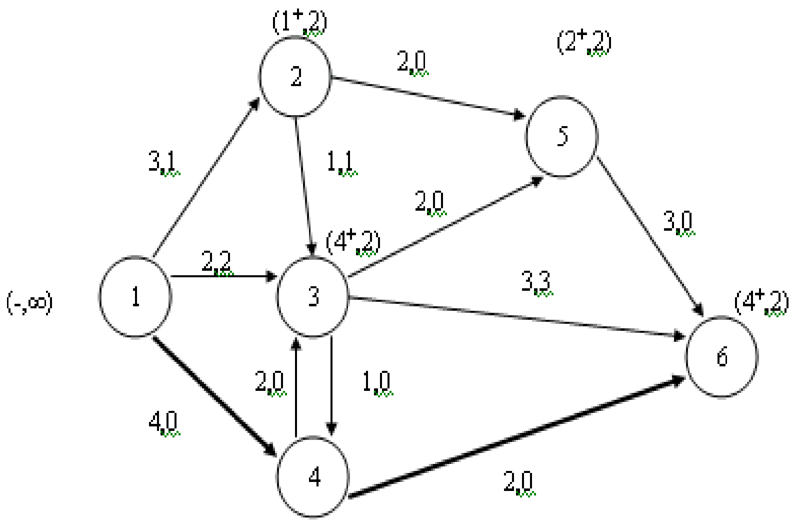


Рисунок 4.32

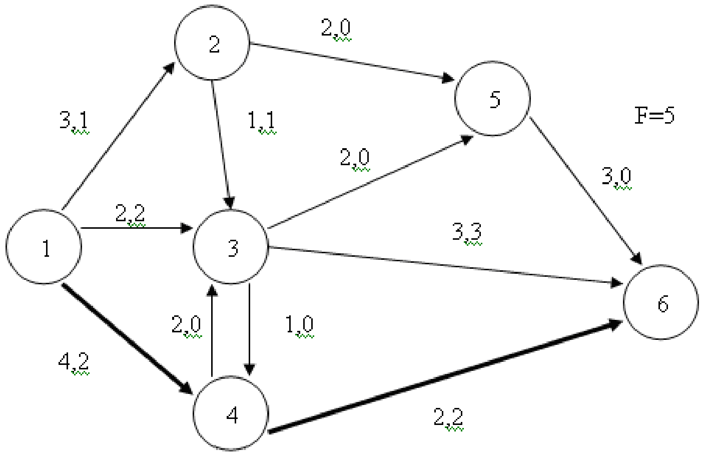


Рисунок 4.33

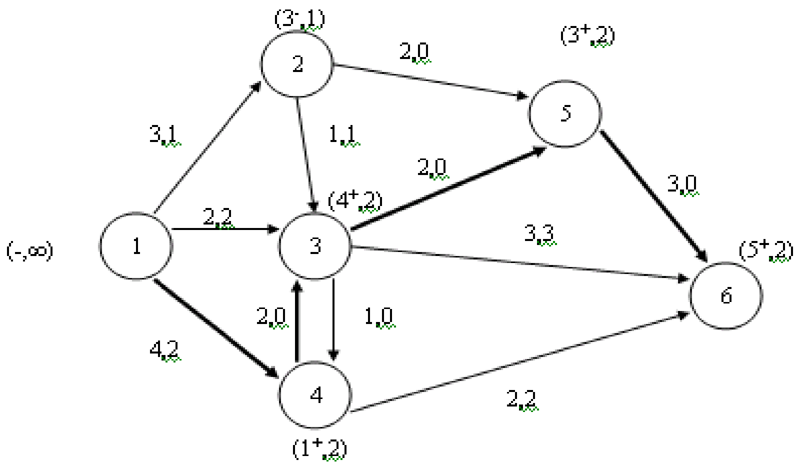


Рисунок 4.34

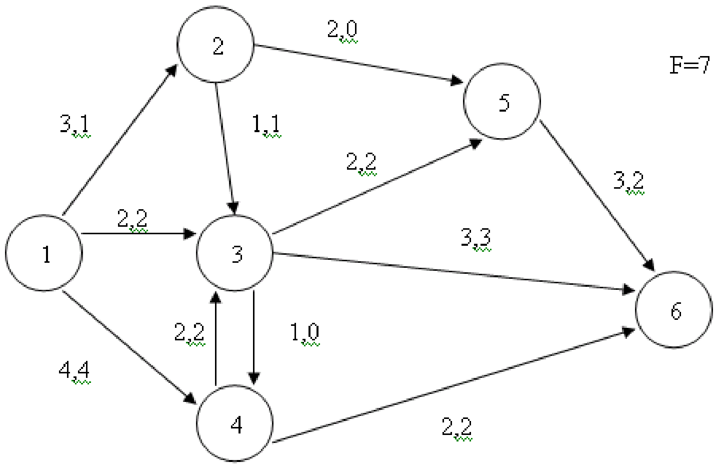


Рисунок 4.35

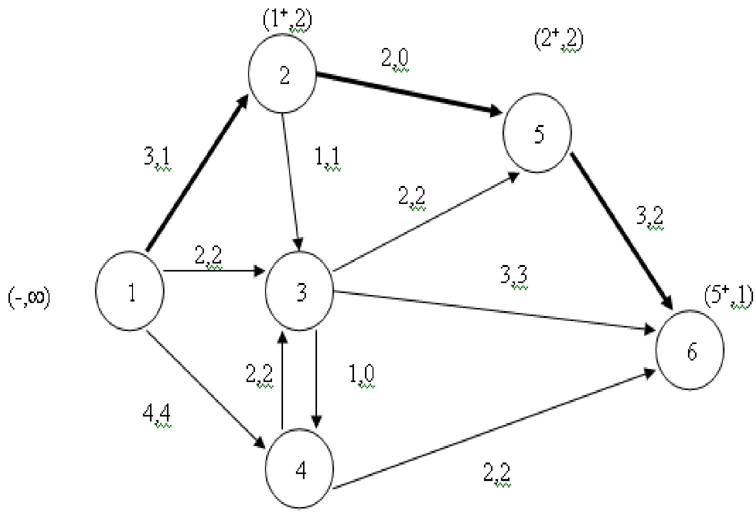


Рисунок 4.36

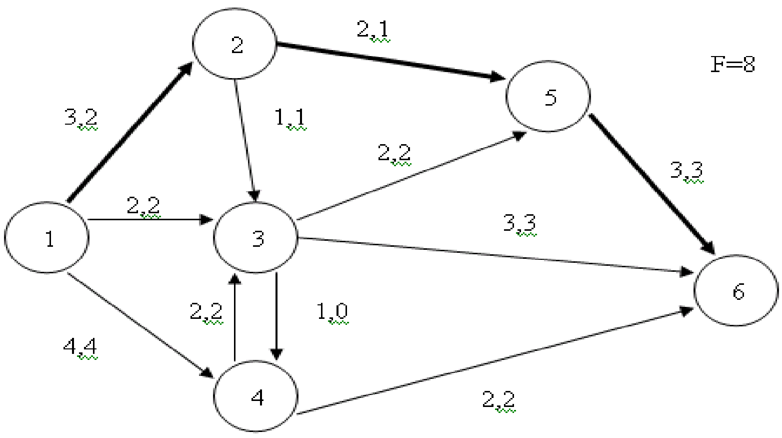
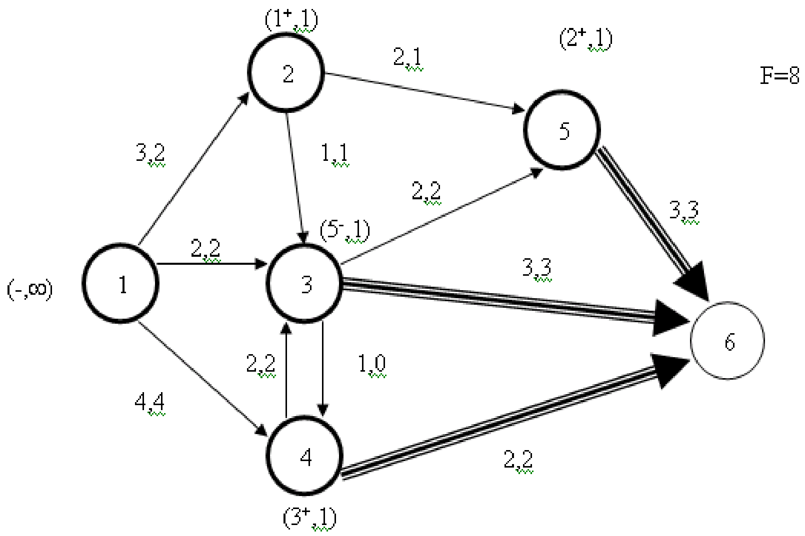


Рисунок 4.37

На рис.4.38 показана спроба знаходження нового прориву в мережі.



$X = \{1,2,3,4,5\};$

$\overline{X} = \{6\};$

$(X,\overline{X}) = \{(5,6),(3,6),(4,6)\}.$

Рисунок 4.38 – Спроба знаходження нового прориву в мережі

При цьому вершини 2 і 5 отримали позначки $(1^+, 1)$ і $(2^+, 1)$ відповідно. Відмітимо, що в даному випадку вдається помітити вузол 3 позначкою (5-, 1), оскільки по дузі (3, 5) є ненульовий потік. Після цього вдається помітити вершину 4 позначкою $(3^+, 1)$. Проте подальше виконання алгоритму неможливе, оскільки дуги (5, 6), (3, 6) і (4, 6) мають нульову залишкову пропускну спроможність. Оскільки не вдається помітити вершину 6 (стік), то на попередньому кроці виконання алгоритму отриманий максимальний потік величини 8.

Оскільки поміченими виявилися вершини 1–5, а непоміченою – вершина 6, то мінімальний розріз, що відокремлює джерело і стік, є сукупністю дуг

$(X,\overline{X}) = \{(5,6),(3,6),(4,6)\}$, де $X = \{1, 2, 3, 4, 5\}$, $\overline{X} = \{6\}$.

Стан мережі приведений на рис.4.38. Жирними стрілками зображений мінімальний розріз $(X,\overline{X})$.

ЛЕКЦІЯ 5 «Моделі мережевого планування та управління»

Анотація

Призначення та області застосування мережевого планування та управління.Побудова правильної нумерації вершин графа. Алгоритм пошуку найкоротшого шляху мережі (графу). Побудова графу планування та управління мережею (ПУМ), порядок та правила побудови графів ПУМ. Упорядкування графу ПУМ, обчислення основних параметрів. Параметри планування й управління мережі за критерієм часу

5.1 Призначення та сфера використання

Темпи виробництва, його масштаби та спеціалізація окремих галузей, багатопрофільні зв'язки обумовлюють необхідність розробки ефективних методів планування та управління, які б давати можливість оцінити змінний стан системи та передбачати її майбутнє, щоб оптимізувати відповідний процес і керувати його перебігом. Системи об'єктів дослідження разом зі зв'язками між ними називаються мережею. Діапазон реального існування мереж дуже широкий: мережі електропостачання, радіо- та телекомунікації, транспортні (залізничні, автомобільні), об'єкти господарювання як в одному господарстві, так і в їх комплексі, плани виконання робіт з реалізації певних проектів і т. ін. Але прикладами таких систем можуть бути також організація поточного виробництва, реконструкція існуючого виробництва, організація капітального виробництва, реконструкція та ремонт існуючих споруд, організація науково-дослідних робіт і т. ін.,де також необхідно узгоджувати та оцінювати зв'язки між окремими елементами. Методи планування та управління мережею забезпечують:

- складання календарного плану виконання певного комплексу робіт;
- оцінку необхідних трудових, матеріальних та фінансових ресурсів, затрат часу;
- контроль комплексу робіт з прогнозуванням і запобіганням можливих зривів при виконанні робіт;

- ефективне управління при чіткому розподілі відповідальності між керівниками різних рівнів і виконавцями робіт;
- оцінку дієздатності та якості системи стосовно певних критеріїв.

Для одних систем зв'язки між об'єктами реалізовані фізично (система комунікацій між населеними пунктами), для інших мають інформаційний характер або є поєднанням як фізичної реалізації , так і інформаційної.

У цій галузі розглядаються дві задачі стосовно управління мережами: планування робіт з реалізації певного проекту та оцінка пропускної спроможності даної мережі.

Математичний апарат, який використовується при дослiджені мереж, розроблений у теорії графів.

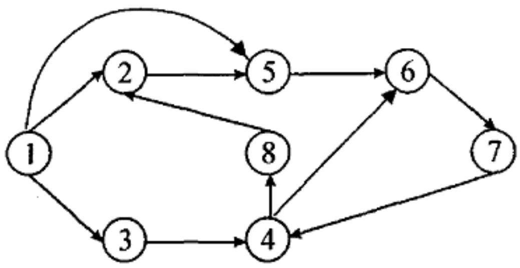
5.2. Основні поняття теорії графів

Побудова математичних моделей розв'язання вказаних задач планування та управління мережами ґрунтується на досліджені попарних (бінарних) зв'язків між об'єктами, які утворюють систему дослідження. Графічне зображення множини досліджуваних об'єктів і зв'язків між ними називається *графом*. На діаграмі об'єкти ображаються пронумерованими точками або кружками, які називаються вершинами. Однобічний орієнтований відрізок називається *дугою*, а графічне зображення неорієнтованих попарних зв'язків між об'єктами – *ребрами*. Граф, вершини якого мають лише однобічні зв'язки, називається *орієнтованим*, або *орграфом*.

Приклад графів.

На рис.5.1 зображено схему використання обладнання при запуску бензинового двигуна внутрішнього згоряння (а) та її граф (б).

а)



б)

Рисунок5.1 – Схема використання обладнання при запуску бензинового двигуна внутрішнього згоряння (а) та її граф (б)

Граф вважається *завантаженим*, якщо він визначений разом з певною функцією на множині його ребер або дуг. Така функція може визначати віддаль між вершинами (карта доріг), час або вартість перевезень між населеними пунктами, пропускну спроможність лінії електропередач або каналу системи зрошення.

Маршрутом називається така послідовність ребер, коли кожна пара сусідніх ребер має одну загальну вершину.

Простим ланцюгом називається маршрут, в якому вершини не повторюються. Ланцюг визначається послідовністю вершин, через які він проходить.

Цикл – це ланцюг, початкова вершина якого співпадає з кінцевою.

Шляхом називається орієнтований ланцюг.

5.3 Побудова правильної нумерації вершин графа

Правильна нумерація вершин виконується за так званим «алгоритмом викреслення дуг».

Алгоритм опишемо, посилаючись на граф (рис. 5.2), для якого у квадратах наводиться довільна нумерація вершин.

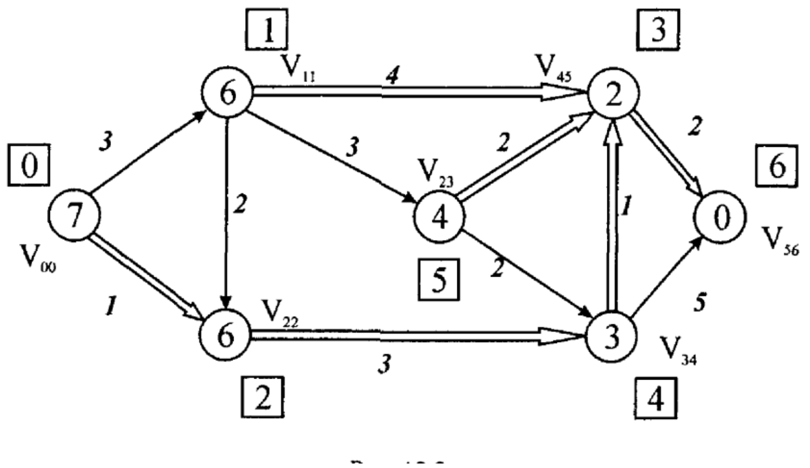


Рисунок5.2 – Алгоритм правильної нумерації вершин графу

Умовно виділимо всі дуги, які виходять з початкової вершини (V_{00} – вершина нульового рангу). Далі розглядаються вершини, в які не входять інші дуги, окрім викреслених. Такі вершини називають вершинами 1-го рангу та пронумеруємо їх у довільному порядку, дотримуючись неперервності в нумерації. Т.ч. кожній вершині надається два індекси: перший – ранг вершини, другий – її порядковий номер серед множини вершин однакового рангу. Повторюючи даний алгоритм, маємо одну вершину першого рангу (V_{11}), дві вершини другого рангу (V_{22}), та (V_{23}), по одній – третього, четвертого і п'ятого рангу відповідно: (V_{34}), (V_{45}), (V_{56}). Виконавши правильно нумерацію вершин графу, в подальшому індекс рангу вершини можна не вказувати.

5.4 Алгоритм пошуку найкоротшого шляху мережі (графу)

Реалізацію алгоритму приведено до побудови послідовності шляхів з однієї дуги, двох дуг, трьох дуг і т. д, які з'єднують вершини певного рангу з кінцевими вершинами. При цьому перегляд вершин виконується в послідовності зменшення їх номерів. На кожному етапі алгоритму відбувається перехід від вершини більш високого рангу до вершини меншого рангу за умови, що із множини всіх шляхів , які починаються в одній і тій же вершині, необхідно залишити для наступного кроку реалізації алгоритму лише найкоротший шлях від цієї вершини до завершальної.

Розглянемо реалізацію цього алгоритму на прикладі графу, зображеного на рис.5.2. Згідно з алгоритмом рухаємося від кінцевої вершини V_{56} до початковоїV$_{00}$. У кружках, що зображають вершини графу записуємо найкоротшу відстань від вершини (n-1)-го рангу до кінцевої вершини. Одночасно найкоротший шлях від кожної вершини до кінцевої будемо відмічати подвійною стрілкою. У кружку останньої кінцевої вершини V_{56} записуємо «0», бо звідси виконуємо відлік відстані. Переходимо до вершини 4-го рангу (V_{45}). Від цієї вершини в кінцеву існує лише один шлях: (V_{45}, V_{56}), його довжина дорівнює двом одиницям виміру, її й запишемо у вершині $V_{45,}$ відповідну дугу позначимо на графові подвійною стрілкою. Переходимо до вершини 3-го рангу V_{34}. З цієї вершини в кінцеву маємо два шляхи: (V_{34}, V_{56}) та (V_{34}, V_{45}, V_{56}). Найкоротшим є останній. Його одержуємо, додаючи дугу (V_{34}, V_{45}) до вже побудованого шляху (V_{45}, V_{56}) та відмічаючи шлях (V_{34}, V_{45}) подвійною стрілкою.

Довжина шляху (V_{34}, V_{45}, V_{56}) дорівнює 1+2=3, записуємо це число у вершині V_{34}. Шлях (V_{34}, V_{56}) при наступних пошуках не використовуємо, тому що з вершини 3-го рангу вже знайдено найкоротший шлях до кінцевої вершини. Алгоритм повторюється аж до вершини $V_{00.}$

Якщо завантаження графу рис. 5.2 тлумачити не як відстань, а як тарифи перевезень вантажу, то ми знайшли шлях найменшої вартості перевезень з вершини V_{00} до V_{56}.

5.5 Побудова графу планування та управління мережею (ПУМ)

Щоб скласти план виконання робіт за проектами , необхідно зобразити його деякою математичною моделлю, яка називається моделлю мережі і є відображенням певних послідовностей виконання робіт і взаємозв’язків між ними з урахуванням необхідних матеріальних ресурсів. Отже, під *моделлю мережі*розуміють план виконання комплексу взаємопов’язаних робіт, представлений у спеціальній формі графу, який називається *графіком мережі*.

Основними елементами графіка мережі є поняття події та роботи.

Термін *робота* використовується у широкому розумінні:

1) це певна реальна робота, яка потребує затрат матеріальних ресурсів та відповідного терміну виконання;

2) чекання – процес у часі, який не потребує ніяких матеріальних затрат (затвердіння бетону, висихання фарби);

3) *фіктивна робота* – це природний логічний взаємозв’язок між двома або кількома роботами чи їх завершенням, який не потребує затрат праці, матеріальних ресурсів або часу. Термін виконання фіктивної роботи

Подія – це фіксація моменту завершення певного етапу виконання проекту.

1) *Вихідна подія* не має попередніх робіт та подій стосовно досліджуваного в моделі комплексу робіт.

2) *Завершальна подія* не може мати наступних подій та робіт.

3) Подій, які не визначають термін виконання певної роботи, називають *початковою* та *кінцевою*.

Події на графі ПУМ зображають кружками (вершини графу), а роботи – орієнтованими дугами, які показують, які роботи потрібно виконати, щоб відбулася певна подія, та які роботи можна виконувати, якщо подія відбулася. Отже, будь яка робота на графі ПУМ позначається двома подіями, між якими вона знаходиться. Фіктивна робота на графі ПУМ позначається штриховою лінією без зазначення часу. На рис. 5.3 наведено граф, який узагальнено описує комплекс будівельних робіт при спорудженні виробничого комплексу.

І на початку, і по завершені реальної роботи повинна бути лише одна подія, бо роботи ідентифікуються за номерами подій, між якими виконуються, тому на графі введено фіктивні роботи (3; 4) та (5; 4).

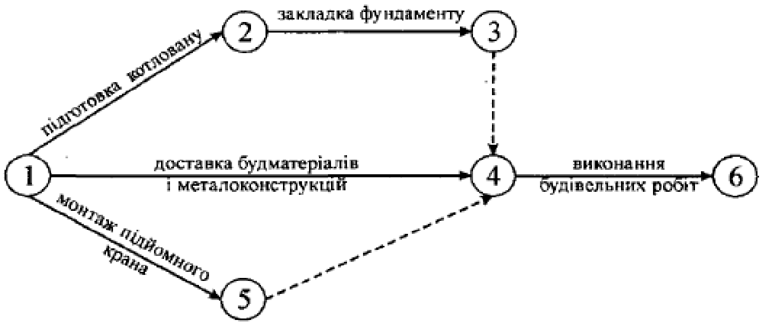


Рисунок 5.3 – Граф, що узагальнено описує комплекс будівельних робіт при спорудженні виробничого комплексу

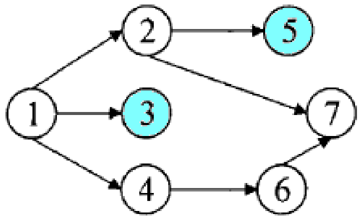
Використовують два принципи побудови графів ПУМ. При реалізації одного з них роботи зображуються дугами, а вершини – подіями, які означають завершення робіт. Це так званий граф «події – роботи». Але використовують й інші підходи до побудови графів ПУМ – без подій. За таким принципом роботи є вершинами графу, а дуги означають залежність між певними роботами в послідовності їх виконання. Побудовані за таким принципом графи ПУМ більш місткі та менш зручні для аналізу та реалізації в ЕОМ. Тому на практиці використовують перший спосіб.

Порядок та правила побудови графів ПУМ.

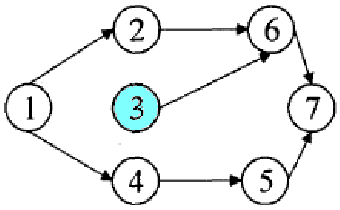
План реалізації проекту виконання комплексу робіт починається з побудови відповідного графу. У плановому процесі виділяються роботи, які відповідають певним етапам, складається реєстр робіт та подій, аналізується послідовність виконання робіт і взаємозв’язки між ними, за роботами закріплюються виконавці. Потім виконують аналіз побудованого графу та його оптимізацію за обраними критеріями.

Вимого до побудови графів ПУМ:

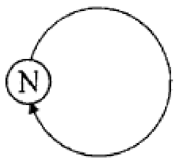
1. Граф ПУМ не повинен мати «глухих кутів», тобто подій, з яких не виходить жодної роботи, окрім завершальної події.

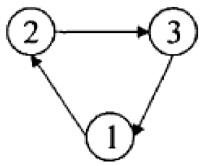


2. На графові не може бути «хвостових» подій (окрім вихідної водії), тобто подій, яким не передує жодна робота.

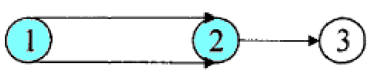


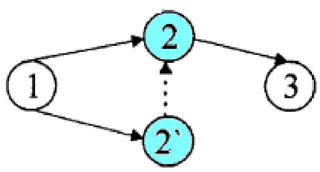
3. Граф не може мати замкнутих контурів і петель, тобто шляхів, які з’єднують певні події з ними ж.



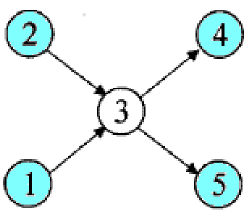


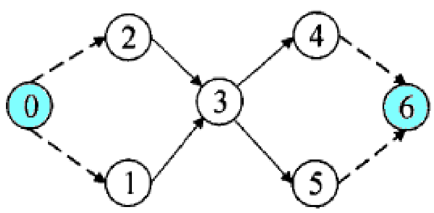
4. Дві довільні події повинні бути безпосередньо пов’язані не більше ніж однією дугою – роботою. Так як роботи позначаються двома індексами (*i, j*), які відповідають подіям «*i*» та «*j*». Якщо в дійсності треба виконати кілька робі, які починаються і завершуються одночасно, при одних і тих же подіях початкових та кінцевих, тоді необхідно ввести фіктивну подію та фіктивні роботи, паралельні роботи при цьому замикаються на фіктивні події.



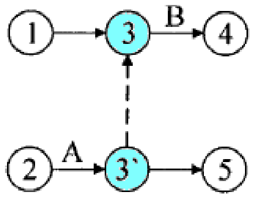


5. На графі ПУМ повинна бути лише одна вихідна і лише одна завершальна події. Якщо це не так (початок реалізації комплексу робіт можна розпочинати паралельно з декількох робіт), то необхідно ввести фіктивні події.

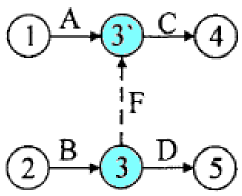




Фіктивні події та роботи можуть вводитися і через інші умови. Однією з таких умов є відображення залежності подій, не пов’язаних з реальними роботами. Наприклад, роботи A та B можуть виконуватись технологічно незалежно одна від одної, але в умовах конкретного виробництва B не може розпочатися раніше, ніж завершиться робота A. Тоді вводиться фіктивна робота C.



Другий випадок – неповна залежність робіт. Наприклад, робота C може бути розпочата лише по завершенні робіт A та B, але робота B – лише по завершенні роботи B і не залежить від роботи A. Тоді необхідно ввести фіктивну роботу F та фіктивну подію 3'.



5.6 Упорядкування графу ПУМ, обчислення основних параметрів

Побудувавши за всіма вимогами граф ПУМ, необхідно зробити правильну нумерацію його вершин. При такому упорядкування графу всі роботи – дуги будуть зорієнтовані зліва на право. Після виконання правильної нумерації індекс рангу вершини можна опустити.

У залежності від характеру та умов виробництва оцінки визначаються за допомогою методів:

- з урахуванням продуктивності праці за умови, що такі роботи виконувались раніше за таких же або близьких обставин;
- за діючими нормами, з урахуванням обсягу роботи, продуктивності праці одного виконавця, кількості працюючих у зміну, кількості змін , тощо;

- методом експертних оцінок, з використанням оцінок, визначених кожним експертом. Наприклад, терміни виконання роботи обчислюється як середня арифметична термінів, названих кожним експертом;
- терміни виконання робіт розглядаються як випадкові величини з деякими законами розподілу та відповідними числовими характеристиками: математичним сподіванням та дисперсією.

Найпростішим розподілом є так званий β-розподіл в математичній статистиці.

Для визначення математичного сподівання терміну виконання певної роботи від виконавців та експертів одержують таку інформацію:

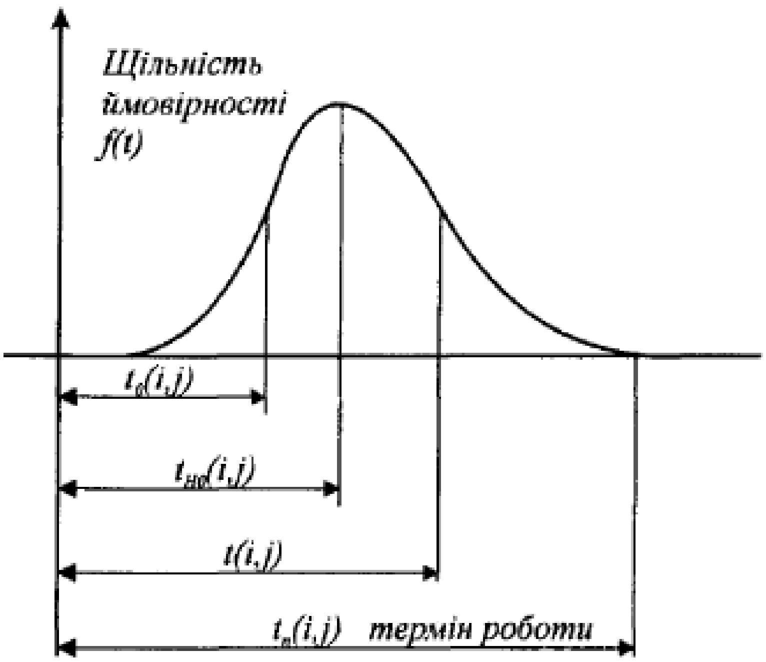


Рис. 5.4 – Інформація, необхідна для визначення математичного сподівання терміну виконання певної роботи

- оптимістичну оцінку t_o, тобто термін виконання роботи за найсприятливіших умов;

- песимістичну оцінку $t_п$, тобто термін виконання роботи за найнесприятливіших умов;

- найбільш ймовірну оцінку $t_{но}$ терміну виконання роботи

Виходячи з гіпотези про β-розподіл величини терміну виконання певної роботи та використовуючи вказану інформацію, отриману від фахівців, математичне сподівання терміну виконання кожної роботи обчислюється за формулою:

$$t(i,j) = \frac{t_o(i,j) + 4t_{но}(i,j) + t_n(i,j)}{6}.$$

Зважаючи на труднощі одержання оцінки $t_{но}$, на практиці використовують більш просту, хоч і менш точну, формулу:

$$t(i,j) = \frac{2t_o(i,j) + 3t_n(i,j)}{5}.$$

Отже для дослідження графу ПУМ необхідно визначити термін виконання кожної роботи. Завантажений граф ПУМ буде мати або детерміновані, або стохастичні величини термінів виконання робіт. Для подальшого аналізу приймемо, що завантажений граф ПУМ має детерміновані оцінки термінів виконання робіт.

5.7 Параметри планування й управління мережі за критерієм часу

Шлях – це будь-яка послідовність робіт, якщо кінцева подія кожної роботи є початковою подією наступної роботи.

Завершений шлях – це будь-який шлях , початком якого є вихідна подія, а закінченням – завершальна.

Термін шляху дорівнює сумі всіх термінів виконання робіт, які створюють шлях. Завершений шлях з найбільшим терміном серед усіх завершених шляхів називається *критичним шляхом*. Роботи, які його створюють, називаються критичними. Граф може мати не один критичний шлях.

Максимальним шляхом між двома подіями «*i*»та «*j*» називається шлях від *i*-ої події до *j*-ої, який має максимальний термін, тобто сума термінів робіт, які складають такий шлях, є не меншою ніж відповідна сума для довільного шляху від *i*-ої події до *j*-ої.

Наприклад, для графу рис. 5.5 завершеними шляхами будуть L_1=[2; 6; 7], L_2=[4; 5; 7],L_3=[3; 6; 7]і т.д.

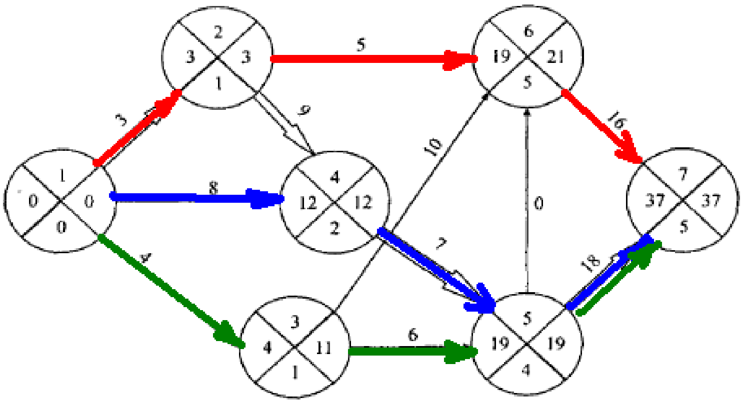
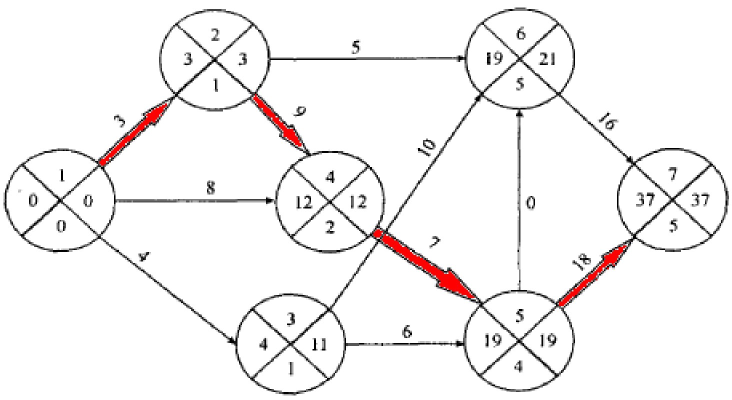


Рисунок 5.5

Їх терміни: $t(L_1)=3+5+16=24$; $t(L_2)=8+7+18=33$;$t(L_3)=4+10+16=30$. Критичний шлях слід визначати за алгоритмом, описаним у п. 5.2, використовуючи завантаження дуги не пропускною спроможністю, а терміном виконання роботи, що відповідає дузі. Критичний шлях для даного графу,$L_{кр}$=[1; 2; 4; 5; 7], а його термін $t(L_{кр})=3+9+7+18=37$.



Критичними роботами будуть: (1; 2), (2; 4), (4; 5), (5; 7).

Критичний шлях має особливе значення для ПУМ, бо роботи, які складають цей шлях, визначають загальний термін завершення всіх робіт комплексу, які плануються в даній системі ПУМ.

Основні параметри ПУМ за критерієм часу наведені в табл. 5.1.

Таблиця 5.1 – Основні параметри планування й управління мережі

Елемент мережі, характеризований параметром	Назва параметра	Умовні позначення параметра
Подія i	Ранній термін звершення події	$t_р(i)$
	Пізній термін звершення події	$t_п(i)$
	Резерв часу події	$R_р(i)$
Робота (i, j)	Термін виконання роботи	$t\ (i, j)$
	Ранній термін початку роботи	$t_{рп}(i, j)$
	Ранній термін завершення роботи	$t_{рз}(i, j)$
	Пізній термін початку роботи	$t_{пп}(i, j)$
	Пізній термін завершення роботи	$t_{пз}(i, j)$
	Повний резерв часу роботи	$R_п(i, j)$
	Частковий резерв першого виду часу роботи	$R_1(i, j)$
	Частковий резерв другого виду часу роботи	$R_2(i, j)$
	Незалежний резерв часу роботи	$R_н(i, j)$
Шлях L	Термін шляху	$t(L)$
	Термін критичного шляху	$t_{кр}$
	Резерв часу шляху	$R_t(L)$

Розглянемо формули обчислення параметрів, вказаних в табл. 5.1.

Результати обчислення параметрів записуються у таблиці на зразок табл.5.2. Якщо граф має велику кількість подій (до двох десятків), то результати можна записувати на графові за схемою, зображеною на рис. 5.6.

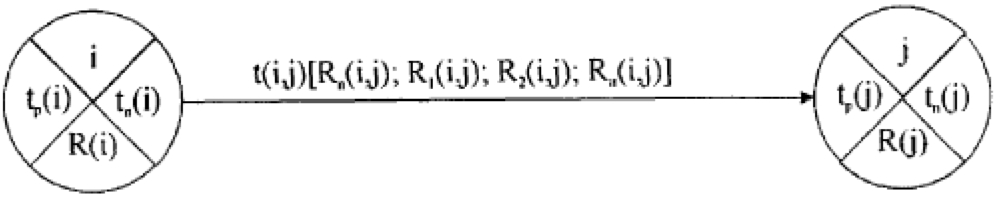


Рисунок 5.6 – Схема запису параметрів планування й управління мережі

Розглянемо параметри подій. Певна подія jне може відбутися раніше, ніж завершаться всі роботи, які їй передують. Отже ранній термін $t_p(j)$ можливого звершення j-ої події визначається терміном максимального шляху:

$$t_p\ j\ = \max\ \ t(L_{bj}) \tag{5.1}$$

L_{bj} – будь-який шлях, який передує j-й подій, тобто шлях від вихідної до j-ої події).

Якщо подія має кілька шляхів, які їй передують, то ранній термін $t_p(j)$ звершення події j зручно обчислювати за формулою:

$$t_p\ j\ = \max\ \ t_p\ i\ + t(i,j)\ . \tag{5.2}$$

Формула (5.2) показує, що обчислення параметру t_p доцільно починати з вихідної події, для якої t_pдорівнює нулю, розглядаючи наступні події в порядку збільшення їх номерів.

Пізній (або граничний) термін (строк) $t_n(i)$ звершення i-ої події обчислюється за формулою:

$$t_n\ i\ = t_{кр} - \max_{L_{iз}} t(L_{iз}) \tag{5.3}$$

де $L_{iз}$ - будь-який шляхL відi-ої події до завершальної.

Якщо подія «i» має кілька наступних шляхів, пов'язана з кількома наступними подіями «j», то пізній строк завершення події «i» обчислюється за формулою:

$$t_n\ i\ = \min_{(i,j)}\ t_n\ j\ - t(i,j)\ . \tag{5.4}$$

Резерв часу $R(i)$ i-ої події обчислюється як різниця пізнього та раннього термінів звершення i-ої події:

$$R\ i\ = t_n\ i\ - t_p(i). \tag{5.5}$$

Резерв часу подій показує, на який допустимий термін можна затримати завершення події, не гальмуючи (не збільшуючи) при цьому термін виконання всього комплексу робіт мережі.

Критичні події не мають резервів часу, бо будь-яка затримка зі звершенням подій, які розташовані на критичному шляху, викличе таку ж затримку у виконанні завершальної події.

Отже, щоб визначити термін критичного шляху, необхідно і достатньо обчислити ранній строк завершальної події, його величина і визначає цей термін.

Події з нульовим резервом часу визначають роботи, які складають критичний шлях.

Приклад. Обчислити параметри часу подій та критичний шлях графу рис.5.5. Результати наведені в табл. 5.2.

Таблиця 5.2 – Результати розрахунків параметрів подій графу (рис.5.5)

Номер події	t_p – ранній термін звершення події	t_n – пізній термін звершення події	R_p – резерв часу події
1	0	0	0
2	0+3=3	$t_n = t_p$(критична подія)	0
3	0+4=4	min{19–6; 21–10}=11	11–4=7
4	max {0+8; 3+9}=12	$t_n = t_p$(критична подія)	0
5	max {4+6; 12+7}=19	$t_n = t_p$(критична подія)	0
6	max {3+5; 4+10; 19+0 }=19	37–16–21	21–19=2
7	max {19+16; 19+18}=37	37	0

Для обчислення ранніх термінів $t_p(i)$ звершення подій вершини графу розглядаємо в порядку зростання їх номерів, які записані у верхній чверті кружка вершини.

Для вихідної вершини $i=1$ очевидно $t_p(i)=0$. Для $i=2$ маємо $t_p(2)=3$ бо існує лише одна робота, завершення якої відповідає події 2. З тієї ж причини $t_p(3)=4$. Для обчислення $t_p(4)$ скористаємося формулою (5.2) та заданими величинами $t(2; 4)=9$ та$t(1; 4)=8$:

$$t_p\ 4\ = max\ t_p\ 1\ + t\ 1;\ 4\ ;\ t_p\ 2\ + t(2;\ 4)\ = max\ 0+8; 3+9\ = 12.$$

$$t_p\ 5\ = max\ t_p\ 3\ + t\ 3;\ 5\ ;\ t_p\ 4\ + t(4;5)\ = max\ 4+6; 12+7\ = 19$$

$$t_p\ 6\ = max\ t_p\ 2\ + t\ 2;\ 6\ ;\ t_p\ 5\ + t\ 5;6\ ;\ t_p\ 3\ + t\ 3;\ 6\ =$$

$$= max\ 3+5; 19+0; 4+10\ = 19$$

$$t_p\ 7\ = max\ t_p\ 5\ + t\ 5;7\ ;\ t_p\ 6\ + t(6;7)\ = max\ 19+18; 19+16\ = 37$$

Отже, термін критичного шляху дорівнює 37 одиницям виміру часу.

При обчислені пізніх термінів здійснення подій $t_n(i)$розглядаємо вершини графу в послідовності зменшення їх нумерації. Для завершальної події пізній термін її завершення дорівнює ранньому, тому $t_n(7)=37$. Пізні терміни інших подій обчислюємо, користуючись формулою (5.4).

$t_n\ 6\ = min\ t_n\ 7\ - t\ 6;\ 7\ \ = min\ 37-16\ = 21$, бо для події 6 існує лише один наступний шлях (6; 7).

$t_n\ 5\ = min\ t_n\ 7\ - t\ 5;\ 7\ \ = min\ 37-18\ = 19.$

$t_n\ 4\ = min\ t_n\ 5\ - t\ 4;\ 5\ \ = 19-7 = 12$, бо для події 4 існує лише один наступний шлях, який починається з дуги (4; 5).

$t_n\ 3\ = min\ t_n\ 5\ - t\ 3;\ 5\ ;\ t_n\ 6\ - t\ 3;\ 6\ \ = min\ 19-6; 21-10\ = 11.$

$t_n\ 2\ = min\ t_n\ 4\ - t\ 2;\ 4\ ;\ t_n\ 6\ - t\ 2;\ 6\ \ = min\ 12-9; 21-5\ = 3.$

$t_n\ 1\ = 0.$

За формулою (5.5) обчислюємо резерв часу для кожної події: $R(1)=0-0=0$; $R(2)=3-3=0$; $R(3)=11-4=7$; $R(4)=12-12=0$; $R(5)=19-19=0$; $R(6)=21-19=2$; $R(7)=37-37=0$.

Відмінний від нуля резерв часу певної події означає, що термін звершення такої події може бути збільшений на величину її резерву часу без затримки терміну виконання всього комплексу робіт.

Аналізуючи результати розрахунків, знаходимо, що події, які лежать на критичному шляху, не мають резерву часу.

Наступним кроком буде *розрахунок параметрів робіт*.

Очевидно, що ранній термін $t_{рп}(i, j)$ початку роботи (i, j) співпадає з раннім терміном здійснення попередньої події i, тому:

$$t_{рп}(i,j) = t_p(i). \tag{5.6}$$

Таким чином, ранній термін $t_{рп}(i, j)$ завершення роботи (i, j) обчислюється так:

$$t_{рз}(i,j) = t_p(i) + t(i,j). \tag{5.7}$$

Для того, щоб фактичний термін виконання певної роботи не збільшив час виконання комплексу робіт, повинна виконуватись вимога: кожну роботу слід завершувати не пізніше допустимого терміну її завершальної події j. За так5их умов пізній термін $t_{пз}(i, j)$ завершення роботи (i, j) визначається формулою:

$$t_{пз}(i,j) = t_n(j). \tag{5.8}$$

а пізній термін $t_{пп}(i, j)$ початку цієї роботи – формулою:

$$t_{nn}(i,j) = t_n(j) - t(i,j). \tag{5.9}$$

Розглянемо *резерви часу шляхів*. Такі резерви мають всі некритичні шляхи. *Резерв часу* $R_t(L)$ шляхуL визначається як різниця термінів критичного та даного шляхів:

$$R_t(L) = t_{кр} - t(L). \tag{5.10}$$

Величина $R_t(L)$ показує на скільки можуть бути збільшені в сумі терміни всіх робіт, які складають шлях L.

Повний резерв $R_n(i, j)$ часу роботи (i, j) обчислюється за формулою:

$$R_n \;\; i,j \;\; = t_n \;\; j \;\; - t_p \;\; i \;\; - t(i,j). \tag{5.11}$$

На рис. 5.7 показано схему розрахунків параметрів часу стосовно робіт.

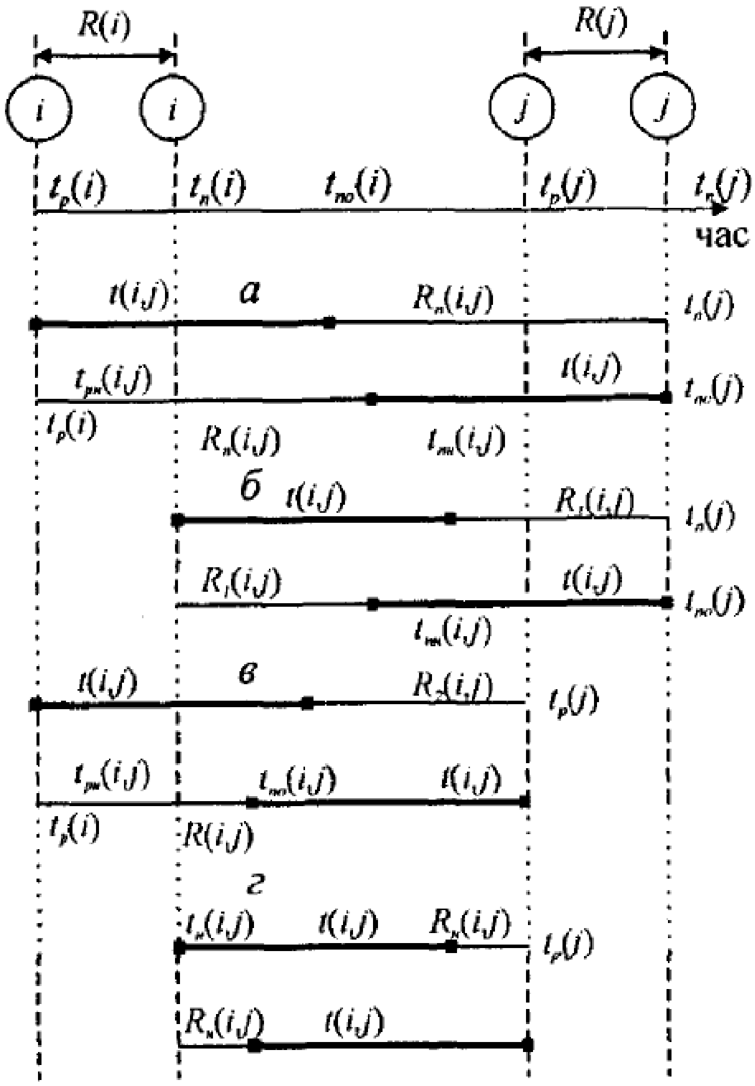


Рисунок5.7 – Схема розрахунків параметрів часу стосовно робіт

Повний резерв часу роботи дорівнює резервові часу максимального із шляхів, які містять цю роботу. Цим резервом можна скористатись при виконанні роботи при умові, що її початкова подія відбудеться найранніший термін, а її завершальна подія – в пізній термін (рис. 5.7, а).

Повний резерв стосується не лише цієї роботи, а й усіх повних шляхів, які містять її. Тоді резерви часу робіт, які належать іншим, але не максимальним за терміном шляхом, що містить цю роботу, відповідно зменшиться на величину використаного резерву.

Частинним резервом першого виду R_1 часу роботи (i, j)є та частина повного резерву часу цієї роботи, на яку можна збільшити термін роботи, не змінивши при цьому пізній термін її початкової події. Цим резервом можна скористатися при виконання певної роботи за умови, що її початкова та завершальні події відбудуться в свої пізні терміни (рис 5.7, б). Величина $R(i, j)$ обчислюється за формулою:

$$R_1(i,j) = t_n(j) - t_n(i) - t(i,j). \tag{5.12}$$

або

$$R_1(i,j) = R_n(i,j) - R(i). \tag{5.13}$$

Частинним резервом другого виду R_2 часу роботи (i, j)є та частина повного резерву часу цієї роботи, на яку можна збільшити термін її виконання, не змінивши при цьому ранній термін її початкової події.

Цим резервом можна користуватись при виконанні певної роботи за умови, що початкова та завершальна події відбудуться в свої ранні терміни (рис. 5.7, в). Величина $R_2(i, j)$ обчислюється за формулою:

$$R_2(i,j) = t_p(j) - t_p(i) - t(i,j). \tag{5.14}$$

або

$$R_2(i,j) = R_n(i,j) - R(j). \tag{5.15}$$

Величина *R_2(i, j)* показує, на скільки можна перенести початок або збільшити термін виконання роботи.

Незалежним резервом часу *R_n*роботи *(i, j)*є та частина повного резерву часу цієї роботи, яка визначається за умови, що попередні роботи завершуються в пізні терміни, а всі наступні розпочинаються в ранні терміни (рис. 5.7, г).

Величина *R_n*обчислюється за формулою:

$$R_n \;\; i,j \;\; = t_p \;\; j \;\; - t_n \;\; i \;\; - t(i,j). \tag{5.16}$$

або

$$R_n \;\; i,j \;\; = R_2 \;\; i,j \;\; - R \;\; i \;\; = R_1 \;\; i,j \;\; - R(j). \tag{5.17}$$

Незалежний резерв часу має сенс лише для тих робіт, які не належать максимальним шляхам, що проходять через її початкові та завершальні події.

Висновки. Частковий резерв першого виду часу роботи може бути використаний на збільшення терміну виконання даної та наступної робіт. Частковий резерв другого виду часу роботи може бути використаний еа збільшення терміну виконання даної та попередніх робіт без зміни резерву часу наступних робіт. Незалежний резерв часу може бути використаний для збільшення терміну виконання лише даної роботи.

Роботи, які складають критичний шлях, та критичні події будь-яких резервів часу не мають. Якщо початкові події *i* роботи *(i, j)*належать критичному шляху, то повний резерв часу

$$R_n \;\; i,j \;\; = R_1(i,j). \tag{5.18}$$

Якщо завершальна подія *j* роботи *(i, j)* належить критичному шляху, але сама робота не належить йому, то повний резерв часу визначається так:

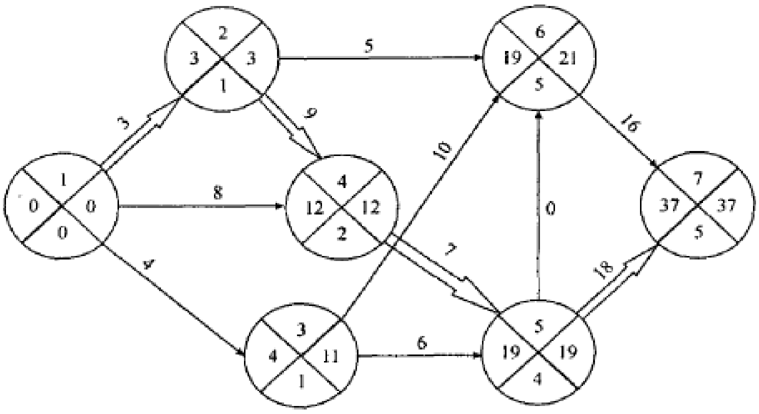
$$R_n \;\; i,j \;\; = R_2(i,j). \tag{5.19}$$

Якщо початкові події i та завершальна подія j належать критичному шляху, але сама робота не належить йому, то повний резерв часу визначається так:

$$R_n\ i,j\ = R_1\ i,j\ = R_2\ i,j\ = R_n\ i,j\ . \tag{5.20}$$

Формули (5.18) - (5.20) доцільно використовувати для контролю вірності розрахунків резервів темпів виконання окремих робіт.

Обчислимо часові параметри робіт графу ПУМ



Результати розрахунків представлені в табл.5.3

Таблиця 5.3 – Таблиця розрахунків термінів графу

Робо-ти	Терміни робіт	R_n (роботи) формула (11)	R_1 (роботи) формула (12)	R_2 (роботи) формула (14)	$R_н$ (роботи) формула (16)
1; 2	3	3–0–3=0	3–0–3=0	3–0–3=0	3–0–3=0
1; 3	4	11–0–4=7	11–0–4=7	4–0–4=0	4–0–4=0
1; 4	8	12–0–8=4	12–0–8=4	12–0–8=4	12–0–8=4
2; 4	9	12–3–9=0	12–3–9=0	12–3–9=0	12–3–9=0
2; 6	5	21–3–5=13	21–3–5=13	19–3–5=11	19–3–5=11
3; 5	6	19–4–6=9	19–11–6=2	19–4–6=9	19–11–6=2
3; 6	10	21–4–10=7	21–11–10=0	19–4–10=5	19–11–10= –2
4; 5	7	19–12–7=0	19–12–7=0	19–12–7=0	19–12–7=0
5; 6	0	21–19–0=2	21–19–0=2	19–19–0=0	19–19–0=0
5; 7	18	37–19–18=0	37–19–18=0	37–19–18=0	37–19–18=0
6; 7	16	37–19–16=2	37–21–16=0	37–19–16=2	37–21–16=0

R_n (3; 6) = -2 свідчить про те, що наступну роботу (6; 7) не можна виконати в ранній термін. Аналіз розрахунків показує, що виконання робіт (1; 4), (2; 6), (3; 5), (3; 6), (6; 7) та наступних за будь-якими шляхами може бути затримано на відповідну кількість часу без витрат термінів виконання попередніх робіт, бо величина R_2для вказаних робіт більша за нуль.

.

ЛЕКЦІЯ 6 «Прийняття рішень в умовах невизначеності»

Анотація

Задачі прийняття рішень в умовах невизначеності. Критерії прийняття рішень в умовах невизначеності:Лапласа, Вальда, Севіджа, Гурвіца (оптимізму-песимізму), Бейєса (максимум середнього виграшу), мінімуму середнього ризику, Ходжеса-Лемана

6.1 Задачі прийняття рішень в умовах невизначеності

Задача прийняття рішень в умовах невизначеності виникає при необхідності діяти в ситуації, яка відома не повністю. Її формулюють переважно як задачу пошуку окремого найкращого (в певному розумінні) рішення на наперед заданій множині допустимих рішень. Основна проблема в тому, що наслідки, пов’язані з прийняттям того чи іншого рішення, залежать від невідомої ситуації. Ступінь неприйнятності цих наслідків прийнято вимірювати умовними одиницями – втратами, яких за припущенням може зазнати активна особа (той, хто приймає рішення). Основною вхідною інформацією, необхідною для розв’язання задач такого типу, є функція втрат, яка являє собою залежність втрат від двох аргументів: рішення та ситуації. Основний крок при розв’язуванні задачі полягає в перетворенні функції втрат на функцію ризику, яка відображає залежність ступеня ризику, на який іде активна особа. Спосіб такого перетворення неоднозначний і залежить від критерію ризику, який вибрала активна особа.

Основними причинами невизначеності є:

- невизначений характер науково-технічного процесу;
- динамічні зміни внутрішніх і зовнішніх умов розвитку економіки;
- неминучі похибки при аналізі складних систем;
- імовірносний характер основних економічних параметрів;
- розвиток і розширення творчої активності працездатного населення;
- необхідність проектування потужних інформаційних потоків на комп’ютерній базі.

В якості ризику ми розглядаємо такі ситуації, при яких настання невідомих подій дуже ймовірне і може бути знайденим. У той же час ситуація, при якій імовірність настання невідомих подій завчасно не може бути нами встановленою, чи не може бути встановленою традиційними засобами, називається невизначеністю.

Поняття господарського ризику та умови його виникнення тісно пов’язані з поняттям невизначеності й ефективності. Ось чому процесу знаходження найбільш ефективного варіанту розвитку деякої виробничої системи властивий господарський ризик. Отже, раціональні методи прийняття рішень в умовах ризику пов’язані з множиною допустимих (збалансованих) планів і їх ефективностями, які є складовими оптимального планування. Тобто, раціональні рішення в умовах ризику є оптимальними.

За наявності ризику, а отже й невизначеності, під збалансованим планом уже недостатньо розуміти план, узгоджений із внутрішніми та зовнішніми параметрами лише за усередненими очікуваними об’ємними показниками, оскільки їх дійсні значення можуть істотно відрізнятися від очікуваних. Тут необхідно враховувати варіацію невизначених параметрів і частоти, з якими вони потрапляють у той або інший інтервал.

Одним із основних способів підвищення ступеня збалансованості плану в умовах невизначеності є формування необхідних резервів.

Для прийняття рішень в умовах невизначеності вхідна інформація задається у вигляді матриці, стрічки якої відповідають можливим альтернативам, а стовпці – станам систем.

Кожній альтернативі та кожному стану системи відповідає результат (наслідок), який визначає виграш (або втрати) при виборі даної альтернативи й реалізації даного стану. Отже, якщо a_i представляє альтернативу i $(i = 1,...,n)$, S_j представляє можливий стан $j (j = 1,…, m)$, то $V(a_i,S_j)$ описує відповідний результат. У загальному випадку $V(a_i, S_j)$ може бути неперервною функцією аргументів a_i та S_j.

У дискретному випадку вказані дані представляються матрицею:

	S_1	…	S_m
a_1	$V(a_1, S_1)$	…	$V(a_1, S_m)$
⋮	⋮	⋮	⋮
a_m	$V(a_n, S_1)$	…	$V(a_n, S_m)$

Така форма представлення в подальшому буде використовуватися при розгляді критеріїв прийняття рішень в умовах невизначеності.

6. 2 Критерій Лапласа

Критерій Лапласа використовується при умові, коли ймовірності можливих станів системневідомі, тобто в умовах повної невизначеності. Даний критерій базується на використанні принципу недостатнього обґрунтування, який стверджує, що стани системи $S_1, S_2, \ldots, S_m$ мають рівні ймовірності. Враховуючи вище сказане, початкову задачу можна розглядати як задачу прийняття рішень в умовах ризику, коли вибирається альтернатива a_i, яка дає найбільш очікуваний виграш R_1 (коли $V(a_i,S_j)$ моделює прибуток) або найменший очікуваний програш R_1(коли $V(a_i, S_j)$ моделює витрати).

Отже, для знаходження величини R_1має місце:

$$R_1 = \begin{cases} \max_i \frac{1}{m} \sum_{j=1}^{m} V(ai,Sj) \;, \textit{якщо } V(ai,Sj) - \textit{прибуток} \\ \min_i \frac{1}{m} \sum_{j=1}^{m} V(ai,Sj) \;, \textit{якщо } V(ai,Sj) - \textit{витрати.} \end{cases} \tag{6.1}$$

де 1/m- імовірність реалізації стану $S_j, j = \overline{1,m}$.

Даний критерій доцільно використовувати в тих випадках, коли різниця між окремими станами системи велика, тобто велика дисперсія значень.

Приклад 6.1. Підприємство повинно визначити рівень виробництва певного виду продукції так, щоби задовольнити потребу споживачів протягом певного періоду часу. Конкретна кількість споживачів невідома, але очікується, що вона може становити одне з п'яти значень: 250, 300, 350, 400, або 450. Для кожного з цих можливих значень існує найкращий рівень пропозиції чи найкраща альтернатива (з точки зору можливих затрат). Відхилення від цих рівнів приводить до додаткових витрат або через перевищення пропозиції над попитом, або через неповне задоволення попиту.

Розмір втрат (тис. грн.) приведений у табл. 6.1. Використовуючи критерій Лапласа, знайти оптимальну альтернативу.

Таблиця 6.1 – Розмір втрат

Альтернатива	Споживачі					$\sum_{j=1}^{m} V(a_i, S_j)$	$\frac{1}{m}\sum_{j=1}^{m} V(a_i, S_j)$	$\min_i$
	S_1	S_2	S_3	S_4	S_5			
a_1	4	22	15	16	29	86	17.2	
a_2	10	15	26	12	10	73	14.6	
a_3	8	19	6	24	5	62	12.4	R_1=12.4
a_4	30	25	5	14	16	90	18.0	
a_5	15	5	30	22	9	81	16.2	

Розв'язання.

Принцип Лапласа припускає, що події S_1, S_2, S_3, S_4, S_5 рівноймовірні. Тобто $p\ S = S_j\ = \frac{1}{5}, j = 1, 5$. Математичні сподівання витрат при різних альтернативах будуть: $M(a_1)=17.\ 2$; $M(a_2)=14.6$; $M(a_3)=12.4$; $M(a_4)=18.0$; $M(a_5)=16{,}2$.

Тоді $R_1 = min\{M(ai)\}= min\{17.2; 14.6; 12.4; 18.0; 16.2\}=12.4$.

Отже, враховуючи критерій Лапласа, найкращою альтернативою буде альтернатива a_3.

6. 3 Критерій Вальда

Даний критерій є найбільш обережним, оскільки він грунтується на виборі альтернативи з усіх найгірш можливих. У зв'язку з цим критерій Вальда часто називають максиміном (мінімаксним).

Якщо результат $V(a_i, S_j)$ відображає втрати особи, що приймає рішення, то для альтернативи a_i найбільші втрати, незалежно від можливого стану S_j, будуть рівними $\max_j V(a_i, S_j)$. Відповідно до мінімаксного критерію найкращою вибирається альтернатива a_i, яка дає $R_2 = \min_i \max_j V(a_i, S_j)$. Аналогічно в тому випадку, коли $V(a_i, S_j)$ відображає виграш, відповідно до максиміного критерію, вибирається альтернатива a_i, яка дає $R_2 = \max_i \min_j V(a_i, S_j)$

Приклад 6.2. Користуючись критерієм Вальда, знайти розв'язок прикладу 6.1.

Розв'язання.

Оскільки $V(a_i, S_j)$ відображає втрати, використаємо мінімаксний критерій. Для знаходження найкращої альтернативи побудуємо таблицю.

Таблиця 6.2 – Використання мінімаксного критерію для знаходження найкращої альтернативи

	S_1	S_2	S_3	S_4	S_5	$\max_j\{V(a_i, S_j)\}$	
a_1	4	22	15	16	29	29	
a_2	10	15	26	12	10	26	
a_3	8	19	6	24	5	24	$\leftarrow \min_i$
a_4	30	25	5	14	16	30	
a_5	15	5	30	22	9	30	

$$R_2 = \min\{29; 26; 24; 30; 30\} = 24.$$

Отже, мінімаксною альтернативою буде a_3. Отриманий результат співпадає з результатом прикладу 6.1.

6. 4 Критерій Севіджа

Використання критерію Вальда інколи приводить до суперечливих висновків. Розглянемо таку матрицю втрат (грн.).

$V(a_i, S_j)=$		S_1	S_2	$\max\limits_j$	
	a_1	50	210	210	
	a_2	150	200	200	$\leftarrow \min\limits_i$

Користуючись критерієм Вальда, приходимо до вибору альтернативи a_2. Інтуїтивно проситься вибрати a_1, оскільки не виключено, що $S = S_1$. Тоді втрати складуть тільки 50 грн. При виборі альтернативи a_2 втрати завжди будуть не меншими 150 грн.

Розглянемо критерій Севіджа, який грунтується на принципі мінімакса наслідків прийнятого помилкового рішення і старається мінімізувати втрачену вигоду. Його зміст полягає у формуванні нової матриці втрат $W(a_i, S_j)$ з допомогою наступної формули:

$$W(a_i, S_j) = \begin{cases} \max_k V(a_k, S_j) - V(a_i, S_j), \text{якщо } V(a_i, S_j) - \text{прибуток} \\ V(a_k, S_j) - \min_k V(a_i, S_j), \text{якщо } V(a_i, S_j) - \text{витрати}. \end{cases} \quad (6.2)$$

Отримані значення показують величину ризику, тому критерій Севіджа називають критерієм мінімального ризику. У першому випадку $W(a_i, S_j)$є різницею найкращого значення в стовпці S_j і значенням $V(a_i, S_j)$. За змістом, $W(a_i, S_j)$виражає «співчуття» особі, що приймала рішення, в зв'язку з тим, що вона не вибрала найкращої дії відносно стану S_j.

У другому випадку $W(a_i, S_j)$відображає різницю $V(a_i, S_j)$ танайгіршого значення в стовпці S_j.

Незалежно від того, є $V(a_i, S_j)$ прибутком або втратами, функція $W(a_i, S_j)$, в обох випадках визначає втрати. Тому до $W(a_i, S_j)$слід використовувати тільки мінімаксний критерій.

Отже, формула для вибору оптимальної альтернативи з допомогою критерія мінімального ризику набуває вигляду:

$$R_3 = \min_i \max_j W(a_i, S_j)$$

Приклад 6.3. Користуючись критерієм Севіджа, знайти розв'язок прикладу 6.1.

Розв'язання.

Відповідно до умови прикладу 6.1 матриця $V(a_i, S_j)$, відображає втрати. Отже, для даного випадку має місце формула

$$W(a_i, S_j) = V(a_i, S_j) - \min_k V(a_i, S_j), \ k = 1,5.$$

Знайдемо числові значення

$$\min_k V(a_k, S_1) = 4; \qquad \min_k V(a_k, S_2) = 5;$$

$$\min_k V(a_k, S_3) = 6; \qquad \min_k V(a_k, S_4) = 12;$$

$$\min_k V(a_k, S_5) = 5, \qquad k = 1,5$$

Тоді шукана величина ризику $W(a_i, S_j)$ приймає вигляд (табл. 6.3).

Таблиця 6.3 – Використання критерію Севіджа для знаходження найкращої альтернативи

		S_1	S_2	S_3	S_4	S_5	$\max_j\{W(a_i, S_j)\}$	
$W(a_i, S_j)=$	a_1	0	17	10	4	24	24	
	a_2	6	10	21	0	5	21	
	a_3	4	14	1	12	0	14	$\leftarrow \min_i$
	a_4	26	20	0	2	11	26	
	a_5	11	0	25	10	4	25	

Отримуємо, $R_3 = \min\{24; 21; 14; 26; 25\} = 14$.

Отже, найкращою альтернативою знову виявилася a_3. Розглянутий критерій досить часто використовується в практичній діяльності при прийнятті управлінських рішень на тривалий період. Наприклад: при розподілі капітальних вкладень на перспективу він дає добрі результати.

6. 5 Критерій Гурвіца (критерій оптимізму-песимізму)

Критерій Гурвіца в своєму алгоритмі охоплює декілька підходів до прийняття рішень: від найбільш оптимістичного до найбільш песимістичного.

При найбільш оптимістичному підході можна вибрати альтернативу, яка дає $\max_i \max_j V(a_i, S_j)$, де $V(a_i, S_j)$являє собою виграш (прибуток).

Аналогічно для найбільш песимістичних припущень вибрана альтернатива відповідає

$$\max_i \max_j V(a_i, S_j). \quad (6.3)$$

Критерій Гурвіца встановлює баланс між випадками крайнього оптимізму й крайнього песимізму, порівнюючи обидві альтернативи з допомогою відповідних коефіцієнтів α, та $(\alpha\text{-}1)$, де $0 \le \alpha \le 1$. Якщо $V(a_i, S_j)$, представляє прибуток, то вибираємо альтернативу, яка дає

$$R_4 = \max_i \{\alpha \max_j V(a_i, S_j) - (1-\alpha) \min_j V(a_i, S_j)\} \quad (6.4)$$

У випадку, коли $V(a_i, S_j)$ представляє втрати, критерій вибирає альтернативу, яка дає

$$R_4 = \min_i \{\alpha \min_j V(a_i, S_j) + (1-\alpha) \max_j V(a_i, S_j)\}. \quad (6.5)$$

Параметр α являє собою показник оптимізму (ступінь впевненості): при $\alpha=1$, критерій дуже оптимістичний; при$\alpha = 0$ – дуже песимістичний. Значення α ($0 \le \alpha \le 1$) може визначитися в залежності від характеру особи, яка приймає рішення, тобто, що їй більш характерно: песимізм або оптимізм. Чим складніша господарська ситуація, чим більше в ній хоче підстрахуватись ОПР, тим ближче до нуля вибирається α. Якщо α наближається до нуля, то збільшується невпевненість при досягненні успіху. Використання даного критерію ускладнюється при відсутності достатньої інформації про величину параметра α, який в силу суб'єктивних причин при різних рішеннях і в різних ситуаціях приймає різні значення. При відсутності інформації про явно виражений характер особи α приймається рівним 0.5.

Припустимо, що $\alpha = 0$, тобто ОПР має мало надії на сприятливий наслідок, тоді отримаємо:

$$R_4 = \max_i \left(0 \cdot \max_j V\left(a_i, S_j\right) + (1-0) \cdot \min_j V\left(a_i, S_j\right) \right) = \max_i \min_j V\left(a_i, S_j\right) = R_2$$

При абсолютній впевненості в досягненні успіху (значення α приймаємо за 1) маємо крайній оптимізм:

$$R_4 = \max_i \left(1 \cdot \max_j V\left(a_i, S_j\right) + (1-1) \cdot \min_j V\left(a_i, S_j\right) \right) = \max_i \min_j V\left(a_i, S_j\right)$$

За умови, що ОПР не має змоги визначити коефіцієнт α, а компроміс між оптимістичним і песимістичним рішенням бажаний використовуємо вираз:

$$R_4 = \begin{cases} \max_i \dfrac{\max_j V\left(a_i, S_j\right) + \min_j V\left(a_i, S_j\right)}{2}, \textit{якщо} V\left(a_i, S_j\right) - \textit{прибуток} \\ \min_i \dfrac{\max_j V\left(a_i, S_j\right) + \min_j V\left(a_i, S_j\right)}{2}, \textit{якщо} V\left(a_i, S_j\right) - \textit{витрати}. \end{cases} \qquad (6.6)$$

Приклад 6.4. Користуючись критерієм Гурвіца, знайти розв'язок прикладу 6.1.

Розв'язання.

Використовуємо критерій Гурвіца до умови прикладу 6.1.

Покладемо α*=0,5*. Для знаходження оптимального рішення побудуємо таблицю:

Таблиця 6.4 – Використання критерію Гурвіца для знаходження найкращої альтернативи

	$\min\limits_j\{V(a_i,S_j)\}$	$\max\limits_j\{V(a_i,S_j)\}$	$\alpha\min\limits_j\{V(a_i,S_j)\}+(1-\alpha)\max\limits_j\{V(a_i,S_j)\}$	
a_1	4	29	16.5	
a_2	10	26	18	
a_3	5	24	14.5	$\leftarrow \min\limits_i$
a_4	5	30	17.5	
a_5	5	30	17.5	

$$R_4 = min\ \{16.5;\ 18;\ 14.5;17.5;17.5\}= 14.5.$$

Отже, оптимальне рішення полягає у виборі альтернативи a_3.

6.6 Критерій Бейєса (максимум середнього виграшу)

Даний критерій використовується за умови, коли відомий розподіл ймовірностей відбуття станів системи. Припустимо, що нам відомі значення ймовірностей$\{p_j, j = 1,m\}$ настання станів системи $S_j, j = 1,m$, які задаються таким розподілом:

S_j	S_1	S_2	…	S_m	$\sum\limits_{j=1}^{m} p_j = 1, 0 \le p_j \le 1$
p_j	p_1	p_2	…	p_m	

Існування закону розподілу ймовірностей станів системи дає можливість визначити математичне сподівання корисності при виборі кожної альтернативи. Оптимальною вважається та альтернатива, яка забезпечує екстремальне (min або max) значення даного математичного сподівання:

$$R_5 = \begin{cases} \max_i \sum_{j=1}^{m} p_j \cdot V(a_i,S_j), \ \text{якщо}\ V(a_i,S_j) - \text{прибуток} \\ \min_i \sum_{j=1}^{m} p_j \cdot V(a_i,S_j), \ \text{якщо}\ V(a_i,S_j) - \text{витрати.} \end{cases} \qquad (6.7)$$

Приклад 6.5. Користуючись критерієм Бейєса, знайти розв'язок прикладу 6.1, якщо відомі ймовірності станів {0.2; 0.15; 0.3; 0.25; 0.1}.

Розв'язання.

Розв'язок задачі представимо таблицею 6.5.

Таблиця 6.5 – Використання критерію Бейєса для знаходження найкращої альтернативи

Альтернатива	S_1		S_2		S_3		S_4		S_5		$\sum_{j=1}^{5} p_j \cdot V_{ij}$	$\min_i$
	V_{i1}	V_{i1}*p1	V_{i2}	V_{i2}*p2	V_{i3}	V_{i3}*p3	V_{i4}	V_{i4}*p4	V_{i5}	V_{i5}*p5		
a_1	4	0.8	22	3.3	15	4.5	16	4.0	29	2.9	15.5	
a_2	10	2.0	15	2.25	26	7.8	12	3.0	10	1.0	16.05	
a_3	8	1.6	19	2.85	6	1.8	24	6.0	5	0.5	12.75	$\leftarrow \min_i$
a_4	30	6.0	25	3.75	5	1.5	14	3.5	16	1.6	16.35	
a_5	15	3.0	5	0.75	30	9.0	22	5.5	9	0.9	19.15	
p_j	0.2		0.15		0.3		0.25		0.1			

Отже, оптимальним рішенням є вибір альтернативи a_3.

6. 7 Критерій мінімуму середнього ризику

Припустимо, що ОПР володіє інформацією про закон розподілу ймовірностей $\{p_j, j = \overline{1,m}\}$ настання станів системи $S_j, j = \overline{1,m}$, і ставить перед собою завдання мінімізувати середній ризик. У даному випадку критерій прийме вид:

$$R_6 = \min_i \sum_{j=1}^{m} p_j W(a_i, S_j) =$$

$$= \min_i \begin{cases} \sum_{j=1}^{m} P_j \left(\max_k V(a_k, S_j) - V(a_i, S_j) \right), & \text{якщо } V(a_i, S_j) - \text{прибуток} \\ \sum_{j=1}^{m} P_j \left(V(a_i, S_j) - \min_k V(a_k, S_j) \right), & \text{якщо } V(a_i, S_j) - \text{витрати.} \end{cases} \qquad (6.8)$$

Приклад 6.6. Користуючись критерієм мінімуму середнього ризику, знайти розв'язок прикладу 6.1, якщо відомі ймовірності станів {0.2; 0.15; 0.3; 0.25; 0.1}.

Розв'язання.

Для розв'язку даної задачі необхідно мати матрицю величин ризику *[W(a_i, S_j)]*, значення якої знайдені в прикладі 6.3. Дальше використаємо формулу (6.8) і представимо розв'язок задачі таблицею 6.6.

Таблиця 6.6 – Використання критерію мінімуму середнього ризику для знаходження найкращої альтернативи

Альтернатива	S_1			S_2			S_3			S_4			S_5			$\sum_{j=1}^{5} p_j \cdot W_{ij}$	$\min_i$
	V_{i1}	W_{i1}	p_1W_{i1}	V_{i2}	W_{i2}	P_1W_{i2}	V_{i3}	W_{i2}	P_3W_{i3}	V_{i4}	W_{i4}	p_4W_{i4}	V_{i5}	W_{i5}	P_5W_{i5}		
a_1	4	0	0	22	17	2.55	15	9	2.7	16	4	1.0	29	24	2.4	8.65	
a_2	10	6	1.2	15	13	1.95	26	20	6.0	12	0	0	10	5	0.5	9.65	
a_3	8	4	0.2	19	14	2.1	6	0	0	24	12	3.0	5	0	0	5.3	$\leftarrow \min_i$
a_4	30	26	5.2	25	23	3.45	5	2	0.6	14	2	0.5	16	11	1.1	10.85	
a_5	15	11	2.2	5	0	0	30	24	7.2	22	10	2.5	9	4	0.4	12.3	
Імовірність	**0.2**			**0.15**			**0.3**			**0.25**			**0.1**				

Отже, оптимальною альтернативою знову буде a_3.

6. 8 Критерій Ходжеса-Лемана

Даний критерій використовує два суб'єктивних показники: закон розподілу ймовірностей $\{p_j, j = 1, m\}$ настання станів системи $S_j, j = 1, m$ і параметр оптимізму α для критерію Гурвіца.

Для загального випадку критерій Ходжеса-Лемана визначається виразом:

$$R_7 = \begin{cases} \max_i \left(\alpha \sum_{j=1}^{m} p_j V(a_i, S_j) + (1-\alpha) \min_j V(a_i, S_j) \right), \text{якщо } V(a_i, S_j) - \text{прибуток} \\ \min_i \left(\alpha \sum_{j=1}^{m} p_j V(a_i, S_j) + (1-\alpha) \max_j V(a_i, S_j) \right), \text{якщо } V(a_i, S_j) - \text{витрати}. \end{cases} \quad (6.9)$$

де $0 \le \alpha \le 1$, $0 \le p_j \le 1$, $\sum_{j=1}^{m} p_j = 1$.

Приклад 6.7. Користуючись критерієм Ходжеса-Лемана, знайти розв'язок прикладу 6.1, якщо відомі ймовірності станів {0.2; 0.15; 0.3; 0.25; 0.1} та значення "параметрів оптимізму" {0.3; 0.6; 0.8}.

Поклавши в основу розрахунків формулу (6.9) і використовуючи результат прикладу 6.5, розв'язок даної задачі представимо таблицею 6.7.

Таблиця 6.7 – Використання критерію Ходжеса-Лемана для знаходження найкращої альтернативи

Альтерн.	$\sum_{j=1}^{5} p_j V(a_i, S_j)$	$\max_j V(a_i, S_j)$	$\alpha \sum_{j=1}^{5} p_j V(a_i, S_j) + (1-\alpha) \max_j V(a_i, S_j)$					
			α=0.3	$\min_i$	α=0.6	$\min_i$	α=0.8	$\min_i$
a_1	15.5	29	24.95		20.9		18.2	
a_2	16.05	26	23.015		20.03		18.4	
a_3	12.75	24	20.625	*min*	17.25	*min*	15.0	*min*
a_4	16.35	30	25.905		21.81		19.08	
a_5	19.15	30	26.745		23.49		21.32	

Результати отриманих розрахунків у табл. 6.7 за критерієм Ходжеса-Лемана для всіх значень α показують, що оптимальною альтернативою буде a_3.

Розглянемо можливі часткові випадки критерію Ходжеса-Лемана:

- $\alpha = 1$, отримуємо формулу критерію Бейєса;

- $\alpha = 0$, отримуємо формулу критерію Вальда.

Оцінюючи необхідну початкову інформацію критерію Ходжеса-Лемана, можна зробити висновок про ступінь його складності.

Основним недоліком розглянутого критерію є те, що в його алгоритмі використовується багато суб'єктивних факторів.

ЛЕКЦІЯ 7 «Теорія ігр та ігрове моделювання»

Анотація

*Основні поняття теорії ігор. Оптимальний розв'язок в іграх двох осіб з нульовою сумою. Змішані стратегії. Графічний метод розв'язку ігор виду 2 × m і n × 2. Ітеративний метод Брауна-Робінсон***.** *Зведення задач теорії ігор до задач лінійного програмування.*

7. 1 Основні поняття теорії ігор

При розв'язуванні ряду економічних задач дуже часто виникають конфліктні ситуації, які породжуються суперечливими інтересами виробничих або зацікавлених структур. Математичним апаратом розв'язку даного типу задач є теорія ігор, яка представляє собою теорію побудови математичних моделей прийняття оптимальних рішень в умовах конфлікту.

Теорія гри – розділ сучасної математики, що вивчає математичні моделі так названих конфліктних ситуацій (тобто ситуацій, при яких інтереси учасників або протилежні і тоді ці моделі називаються «антагоністичними іграми», або не збігаються, хоча і не протилежні, і тоді мова йде про «ігри з протилежними інтересами»). Основоположники теорії Дж. фон Нейман і О. Моргенштерн намагались математично описати характерні для капіталістичної економіки явища конкуренції як деяку «гру». В найбільш простому випадку мова йде про протиборство двох противників, наприклад, двох конкурентів, які ведуть боротьбу за ринок збуту.

Оскільки сторони, що беруть участь у вирішенні конфліктів, зацікавлені у прихованні своїх намірів від супротивника, прийняття рішень в умовах конфлікту виявляється переважно прийняттям рішень в умовах невизначеності. Фактор невизначеності в даній ситуації можна інтерпретувати як супротивника суб'єкта, що приймає рішення.

Логічною основою теорії ігор є формалізація трьох понять, які входять у її визначення й є базовими для всієї теорії: конфлікту, прийняття рішення в ньому та оптимальність цього рішення. *Ситуація називається конфліктною*, якщо в ній приймають участь сторони, інтереси котрих повністю чи частково протилежні. Гра – це дійсний або формальний конфлікт, в якому беруть участь хоч би два учасники (гравці), кожний із яких прагне досягти власної мети. Допустимі дії кожного з гравців, спрямовані на досягнення деякої мети, називаються правилами гри. Кожний гравець має деяку множину (скінченну чи нескінченну) можливих виборів, яка називається *стратегіями*.

Стратегією гравця називається сукупність правил, що визначає вибір його дій при кожному власному ході в залежності від ситуації, яка відбулася. Звичайно в процесі гри при кожному приватному ході гравець робить вибір в залежності від конкретної ситуації. Однак, в принципі можливо, що всі рішення, прийняті гравцем раніше (у відповідь на будь-яку створену ситуацію). Це означає, що гравець обрав визначену стратегію, яка може бути задана у вигляді списку правил або програм (так можливо здійснити гру за допомогою ЕОМ).

Метою теорії гри є визначення оптимальної стратегії для кожного гравця.

Стратегія гравця називається оптимальною, якщо при багатократному повторенні гри вона забезпечує гравцеві максимально можливий середній виграш (або мінімально можливий програш).

В основу класифікації ігор покладено такі ознаки: кількість гравців, кількість стратегій, характер взаємовідносин, характер виграшів, вигляд виграшів, кількість ходів, стан інформації (табл. 7.1).

Залежно від кількості гравців існують ігри: одного гравця, двох гравців, k-гравців. Гра називається парною, якщо в ній приймають участь тільки дві сторони (дві особи). Щодо кількості стратегій ігри діляться на скінченні та нескінченні. Якщо в грі кожен із гравців має скінченне число стратегій, то вона називається скінченною. Якщо ж хоч би один із гравців має нескінченну кількість можливих стратегій, то така гра буде називатися нескінченною. За характером взаємовідносин гравців ігри поділяються на безкоаліційні, кооперативні та коаліційні. Безкоаліційними називають ігри, в яких гравці не мають права домовлятися між собою, тобто утворювати коаліції. У кооперативній грі коаліції наперед відомі.

Таблиця 7.1 – Класифікація ігор

№ п/п	Класифікаційні ознаки	Групи ігор	
1	Число гравців	1. Одного гравця 2. Двох гравців 3. *k*- гравців	
2	Кількість стратегій	1. Скінченні 2. Нескінченні	
3	Характер взаємовідносин	1. Безкоаліційні 2. Кооперативні 3. Коаліційні	
4	Характер виграшів	1. Нульовою сумою 2. Ненульовою сумою	
5	Вигляд функції виграшів	1. Матричні 2. Біматричні 3. Неперервні	4. Випуклі 5. Сепарабельні 6. Типу дуелі
6	Кількість ходів	1. Однокрокові 2. Багатокрокові (позиційні, стохастичні, диференціальні, типу дуелі).	
7	Стан інформації	1. З повною інформацією 2. З неповною інформацією	

За характером виграшів ігри поділяються на ігри з нульовою сумою та ігри з ненульовою сумою. Гра називається грою з нульовою сумою, якщо сума виграшів усіх гравців у кожній її партії дорівнює нулю, тобто загальний капітал усіх гравців не змінюється, а тільки перерозподіляється між гравцями залежно від отриманих наслідків.

Залежно від виду функції виграшів ігри діляться на матричні, біматричні, неперервні, випуклі, сепарабельні, типу дуелі тощо.

Відносно кількості ходів ігри поділяють на однокрокові й багатокрокові. Однокрокові ігри закінчуються після закінчення одного ходу кожного гравця. Багатокрокові ігри поділяються на позиційні, стохастичні, диференціальні, типу дуелі тощо.

Залежно від стану інформації розрізняють ігри з повною та неповною інформацією. Якщо на кожному кроці гри кожному гравцеві відомо, які дії були зроблені гравцями раніше, то така гра називається грою з повною інформацією, якщо ж не все відомо про попередні дії, то – грою з неповною інформацією.

7. 2 Оптимальний розв’язок в іграх двох осіб з нульовою сумою

Розглянемо гру, в якій беруть участь два гравці, один з яких може дотримуватися стратегії i з n своїх можливих стратегій (i=1, 2,..., n), а другий, не знаючи вибору першого, вибирає стратегію j із mсвоїх можливих стратегій (j =1, 2, ..., m).У результаті перший гравець (А) виграє a_{ij}, а другий (В) програє цю величину.

Величини a_{ij} утворюють платіжну матрицю (матрицю гри):

$$\left[a_{ij}\right] = \begin{array}{c} \\ A_1 \\ A_2 \\ \vdots \\ A_n \end{array} \begin{array}{c} \begin{array}{ccc} B_1 & \cdots & B_m \end{array} \\ \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2m} \\ \vdots & \cdots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \end{array}. \qquad (7.1)$$

Рядки матриці [a_{ij}] відповідають стратегіям (A_1, ..., A_n) гравця А. А стовпці – стратегіям (B_1, ..., B_m) гравця В. Дані стратегії називаються *чистими*. Будемо вважати, що при $a_{ij} > 0$ гравець А виграє, а гравець В програє величину a_{ij}. Якщо $a_{ij} < 0$, то навпаки, виграє гравець B_i програє гравець А.

Спочатку знайдемо найкращу із стратегій гравця А, тобто найкращу серед A_1, ..., A_n з урахуванням можливих варіантів відповідей на неї гравця В. При цьому необхідно враховувати те, що на довільну стратегію A_i гравець В відповідає стратегією B_j, для якої виграш гравця А буде мінімальним. Для знаходження стратегії B_j необхідно в i-му рядку платіжної матриці знайти $a_{ij} = \min_j\ a_{ij}$. При зміні стратегії гравця А одночасно будуть змінюватися відповідні їм числа α_i. Зрозуміло, що гравцеві А вигідно завжди зупинятися на такій стратегії A_i, для якої значення $a = max_i\ a_i$, або, враховуючи представлення α_i, отримаємо $a = max_i\, min_j\ a_{ij}$.

Число α називається *нижньою ціною гри* чимаксиміном,а відповідна йому стратегія (рядок) – максимінною.

Якщо гравець А буде дотримуватися максимінної стратегії, то йому, при довільній поведінці гравця В, у будь-якому випадку гарантований виграш, не менший α. Аналогічно можна визначити найкращу стратегію для гравця В, мета якого – привести виграш гравця А до мінімуму. Для цього гравець В прагне для кожної своєї стратегії B_j отримати максимальне значення виграшу при довільній стратегії гравця A_i, тобто він шукає значення $\beta_j = max_i \ a_{ij}$. Проте гравець В не може розраховувати на те, що гравець А дозволить йому отримати будь-який із виграшів β_j. Єдине, на що може розраховувати гравець В, то це на те, щоб отримати виграш, який буде не меншим величини $\beta = min_j \, max_i \ a_{ij}$. Величина β називається *верхньою ціною гри*, чи мінімаксом, а відповідна йому стратегія гравця (стовпець) мінімаксною. Мінімакснастратегія – найобережніша стратегія для гравця В. Вона забезпечує йому в будь-якому випадку програш, не більший β, і відповідно виграш гравцеві А, теж не більший від β. Якщо $\alpha = \beta = \nu$, то число v називається *ціною гри*.

Гра, для якої $\alpha = \beta$, тобто мінімаксне значення рівне максимінному, називається *грою із сідловою точкою*. Для гри зі сідловою точкою розв'язок полягає у виборі максиміної й мінімаксної стратегії, що є оптимальними. “Оптимальність” тут означає, що жоден гравець не прагне змінити свою стратегію, оскільки його суперник може відповісти на це вибором іншої стратегії, яка може дати гірший результат для першого. Взагалі значення гри повинно задовільняти нерівність:

[Максиміннезначення]≤ [Значеннягри]≤ [Мінімаксне значення]

Приклад 7.1. Підприємства А та В виробляють два конкуруючих види продукції. У даний час кожний вид продукції “контролює” 50% ринку. Для покращення якості продукції підприємства планують розгорнути рекламні заходи. Якщо обидва підприємства не будуть цього робити, то стан ринку не зміниться. Обстеження ринку показує, що 50% потенційних покупців отримують інформацію через телебачення, 30% - через пресу й останні 20% - через радіомовлення.

Мета кожного підприємства – вибрати ефективні засоби реклами. Дану задачу необхідно сформулювати як гру двох осіб з нульовою сумою та знайти оптимальні стратегії.

Розв'язання. Учасниками гри є два підприємства А і В. Кожен із гравців має три стратегії використання реклами – телебачення (1), преса (2) й радіомовлення (3). Якщо обидва гравці виберуть для реклами своєї продукції однакові засоби інформації, то їхній вплив на ринок не зміниться. Припустимо, що якщо підприємство А вибрало як засіб реклами телебачення (A_1), то підприємство В може вибрати для рекламителебачення (B_1), пресу (B_2), чи радіомовлення (B_3). У результаті такого вибору вплив на ринок для А в першому випадку не зміниться, в другому й третьому відповідно збільшиться на 20% і 30%. Якщо підприємство А вибере за стратегію рекламу через пресу (A_2), то В може вибрати телебачення (B_1), пресу (B_2), чи радіомовлення. Тоді в першому випадку підприємство А втратить 20% споживачів, у третьому – попитзростена 10%. Аналогічно аналізуємо третю стратегію. Остаточно отримуємо платіжну матрицю a_{ij} :

B min в рядках					
		B_1	B_2	B_3	
A	A_1	0	20	30	0 ← максмін
max →	A_2	-20	0	10	-20
у стовпцях	A_3	-30	-10	0	-30
		0	20	30	
		↑ мінімакс			

Знайдемо нижню та верхню ціни даної гри:

$$\alpha = \max_i \min_j \{a_{ij}\} = \max\{0;-20;-30\} = 0;$$
$$\beta = \min_j \max_i \{a_{ij}\} = \min\{0;20;30\} = 0.$$

Оскільки $\alpha = \beta = 0$, то гра має сідлову точку. Оптимальними будуть стратегії A_1 для підприємства А і B_1 для В, тобто обом підприємствам слід використати для реклами телебачення, як засіб інформації.

7. 3 Змішані стратегії

Наявність у грі сідлової точки дає можливість визначити необхідні оптимальні стратегії. Але деякі ігри не завжди мають сідлові точки, тобто максимінно-мінімаксні стратегії не оптимальні. Це призводить до того, що кожний із гравців може поліпшити своє становище, вибравши іншу стратегію. Уданому випадку виникає потреба у використанні змішаних стратегій.

Змішані стратегії – це математична модель можливої й гнучкої тактики гравця, при якій супротивний йому гравець не може знати наперед ситуацію, з якою йому прийдеться зіткнутись у грі, тому перед кожною партією проводиться випадковий вибір однієї з чистих стратегій з допомогою деякого механізму, який здійснює цей вибір із визначеними й наперед заданими ймовірностями. Розглянемо гру двох осіб, матриця платежів якої має розмірність $n \times m$. Нехай гравець А має n стратегій, а гравець В –m стратегій. Позначимо через $p = (p_1, p_2, \ldots, p_n)$ і $q = (q_1, q_2, \ldots, q_m)$ вектори ймовірностей, з якими гравці А та В відповідно вибирають свої чисті стратегії. Оскільки ці стратегії за умовою гри повністю вичерпують можливі ходи гравців А і В, то вони утворюють повну групу подій.

Тому має місце:

$$ {}_{i=1}^{n} p_i = \quad {}_{j=1}^{m} q_j \qquad (7.2)$$

Якщо a_{ij} – (i, j)-й елемент матриці гри, то платіжна матриця має вигляд

Відповідно до основної теореми теорії ігор кожна скінченна гра має хоч би один розв’язок, який визначає певна змішана стратегія.

			B_1	B_2	…	B_m
			q_1	q_2	…	q_m
A	A_1	p_1	a_{11}	a_{12}	…	a_{1m}
	…	…	⋮		…	⋮
	A_n	p_n	a_{n1}	a_{n2}	…	a_{nm}

Методика визначення розв’язку гри при змішаних стратегіях в основному також грунтується на використанні критерію мінімаксу. Різниця полягає в тому, що гравець А вибирає P_i так, щоб максимізувати найменший сподіваний виграш (математичне сподівання) по стовпцях, тоді як гравець В вибирає q_j з метою мінімізації найбільшого сподіваного виграшу по рядах. Математично критерій мінімаксу для змішаних стратегій описується наступним чином. Гравець А вибирає стратегію A_i, яка дає

$$\max_{P_i}\left\{\min\left(\sum_{i=1}^{n} a_{i1}p_i, \sum_{i=1}^{n} a_{i2}p_i, \ldots, \sum_{i=1}^{n} a_{im}p_i\right)\right\}, \tag{7.3}$$

а гравець В вибирає стратегію B_j, яка дає

$$\min_{q_j}\left\{\max\left(\sum_{j=1}^{m} a_{1j}q_j, \sum_{j=1}^{m} a_{2j}q_j, \ldots, \sum_{j=1}^{m} a_{nj}q_j\right)\right\}. \tag{7.4}$$

Ці величини визначаються відповідно як сподівані максимінні та мінімаксні платежі. Прицьому має місце співвідношення:

$$\begin{bmatrix}\text{Мінімаксний сподіваний}\\ \text{програш}\end{bmatrix} \geq \begin{bmatrix}\text{Максиміний сподіваний}\\ \text{виграш}\end{bmatrix}.$$

Якщо p_i і q_j відповідають оптимальним розв’язкам, тобто виконується строга рівність, то результуюче значення дорівнює сподіваному (оптимальному) значенню гри. Якщо p_i^* і q_j^* оптимальні розв’язки для обох гравців, то кожному елементу платіжної матриці відповідає ймовірність $p_i^* q_j^*$. Отже, оптимальне сподіване значення (ціна) гри, або математичне сподівання плати гравця А гравцеві В (середній виграш гравця В) знаходиться за формулою:

$$v^* = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} p_i^* q_j^* \qquad (7.5)$$

Для знаходження оптимальних стратегій в іграх двох осіб з нульовою сумою можна використати графічний метод (для ігор виду $2 \times m$ або $2 \times n$), а також привести задачу до лінійного програмування.

7.4 Графічний метод розв’язку ігор виду $2 \times m$і $n \times 2$

Розглянемо гру виду $2 \times m$, яка не має сідлової точки.

			B			
			B_1	B_2	…	B_m
			q_1	q_2	…	q_m
A	A_1	p_1	a_{11}	a_{12}	…	a_{1m}
	A_2	p_2=1-p_1	a_{21}	a_{22}	…	a_{2m}

Оскільки гравець А має дві стратегії (A_1 і A_2), то $p_2 = 1 - p_1$. Знайдемо сподівані виграші (математичне сподівання гравця) А залежно від чистих стратегій гравця В:

$$M_j = \ p_1 \ = a_{1j} \cdot p_1 + a_{2j} \cdot p_2 = \ a_{1j} - a_{2j} \ p_1 + a_{2j}, j = 1, m. \qquad (7.6)$$

Отримані результати представимо у вигляді табл. 7.2.

Таблиця. 7.2 – Стратегії гравців *A* та *B*

Чисті стратегії гравця *B*	Сподівані виграші гравця *A*
1	$(a_{11}-a_{21})p_1 + a_{21} = M_1\ (p_1)$
2	$(a_{12}-a_{22})p_1 + a_{22} = M_2\ (p_1)$
⋮	⋮
m	$(a_{1m}- a_{2m})p_1 + a_{2m} = M_m\ (p_1)$

Із таблиці 7.2. випливає, що сподіваний виграш гравця А $M_j(p_1)$ лінійно залежить від p_1 і являє собою пряму лінію.

Мета гравця В полягає в мінімізації виграшу гравця А за рахунок вибору своїх стратегій, тобто $M(p_1) = \min_j M_j(p_1)$. На рис. 7.1 точка M^* означає значення $M(p_1^*)$, яке дістали при $p_1 = p_1^*$. Ціна гри $v=M(p_1^*)$. Таким чином, графічно визначається оптимальна змішана стратегія $p_1 = p_1^*$ першого гравця і пара частих стратегій другого гравця, які в перетині утворюють точку M^* (на рис. 7.1 це відповідає ІІ і ІІІ стратегіям гравця В).

Розв’язавши систему рівнянь $M_j(p_1) = v (j = 2, 3)$, знайдемо точне значення p_1^*.і v. Після цього, надавши $q_j = 0$ для тих j, для яких $M_j(p_1)$ не утворюють точку M^*, можемо знайти значення q_j для тих j, для яких $M_j(p_1)$ утворюють точкуM^*, розв’язавши систему

$$\begin{aligned} a_{12}q_2 + a_{22}q_3 = v \\ a_{13}q_2 + a_{23}q_3 = v \end{aligned} \qquad (7.7)$$

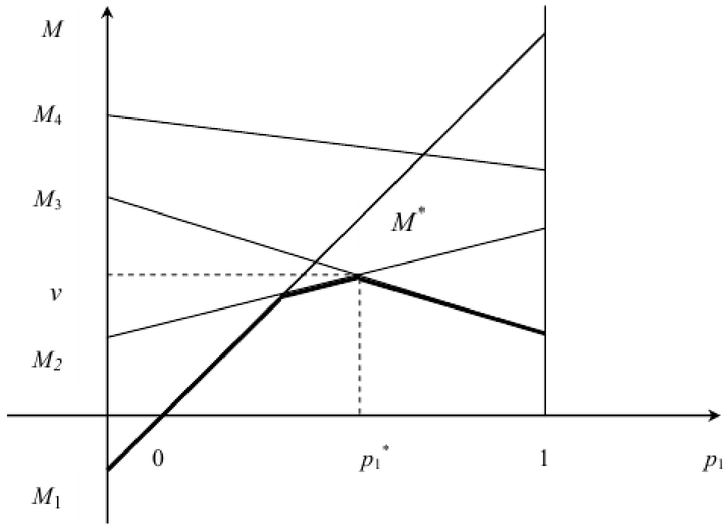


Рисунок 7.1 – Графічне визначення оптимальної змішаної стратегії $p_1 = p_1^*$ першого гравця і пара частих стратегій другого гравця

Приклад 7.2. На основі наявного добового обсягу сировини підприємство має можливість випускати два види продукції, яка швидко псується. Прибуток підприємства залежить від обсягу реалізованої продукції кожного виду, яка в свою чергу залежить від погоди.

Реалізація першого виду продукту вища за теплої погоди, другого – за прохолодної. Стан погоди можна розглядати як такі стратегії природи: день пекучий сухий, пекучий вологий, теплий сухий, теплий вологий, прохолодний сухий, прохолодний вологий. Відома матриця прибутку (тис. грн.) підприємства за кожним видом продукції залежно від стану погоди:

$$\begin{matrix} 12 & 9 & 7 & 5 & 4 & 2 \\ 3 & 5 & 6 & 8 & 9 & 10 \end{matrix}.$$

За допомогою графічного методу знайти оптимальні стратегії по організації випуску продукції підприємством.

Розв'язання. Побудована гра не має сідлової точки в чистих стратегіях, тому для визначення оптимальних стратегій можна скористатися графічним методом. Стратегіями підприємства є виробництво продукції першого або другого виду. Будемо вважати підприємство першим гравцем, а природу – другим. Позначимо через p_1 імовірність використання своєї першої стратегії першим гравцем, через $q(q_1, q_2, \ldots, q_6)$ – змішану стратегію другого гравця. Побудуємо графіки середніх виграшів першого гравця (рис. 7.2.), для цього знайдемо $M_j(p_1)$:

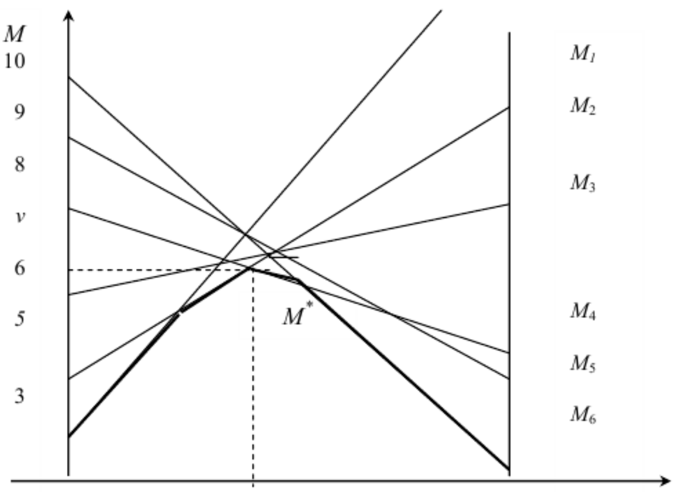


Рисунок 7.2 – Графіки середніх виграшів першого гравця

$$M_1(p_1) = (12-3)p_1 + 3 = 9p_1 + 3; \quad M_2(p_1) = (9-5)p_1 + 5 = 4p_1 + 5;$$
$$M_3(p_1) = (7-6)p_1 + 6 = p_1 + 6; \quad M_4(p_1) = (5-8)p_1 + 8 = -3p_1 + 8;$$
$$M_5(p_1) = (4-9)p_1 + 9 = -5p_1 + 9; \quad M_6(p_1) = (2-10)p_1 + 10 = -8p_1 + 10.$$

Нижня границя множини обмежень зображена на рис. 7.2 жирною лінією. Як бачимо, *maxM(p1)* досягається в точці M^*, що утворюється лініями $M_1(p_1)$ і $M_4(p_1)$. Покладемо $q_2 = q_3 = q_5 = q_6 = 0$. Для знаходження p_1, p_2, q_1, q_4, v необхідно розв’язати такі системи рівнянь:

$$\begin{cases} 12p_1 + 3p_2 = v, \\ 5p_1 + 8p_2 = v, \\ p_1 = 1 - p_2, \end{cases} \qquad \begin{cases} 12q_1 + 3q_4 = v, \\ 5q_1 + 8q_4 = v, \\ q_2 = 1 - q_1. \end{cases}$$

Розв’язками цих систем буде:

$$p_1 = \frac{5}{12};\ p_2 = \frac{7}{12};\ q_1 = \frac{5}{12};\ q_4 = \frac{7}{12};\ v = \frac{27}{4}\text{ млн.грн.}$$

Ми отримали такий оптимальний розв’язок.

Підприємству необхідно 5/12 обсягів сировини використати на виготовлення першого виду продукції, а 7/12 – для другого виду продукції. При цьому отримаємо максимальний прибуток у розмірі 6.75 млн. грн.

7.5 Ітеративний метод Брауна-Робінсон

У попередньому розділі було показано, як знайти точні значення оптимальних змішаних стратегій гравців у матричній грі, що не має сідлової точки. На практиці часто буває достатнім знайти наближення до оптимальних змішаних стратегій, що забезпечують середній виграш, близький до ціни гри. У такому разі можна застосувати один з ітеративних методів розв'язування матричної гри.

Розглянемо ітеративний метод, запропонований Брауном і обґрунтований Робінсон (метод Брауна-Робінсон).

В основі цього методу лежить такий принцип: кожен з гравців прагне збільшити свій виграш, вважаючи, що майбутнє подібне минулому. При цьому вважається також, що жоден з гравців не знає своєї оптимальної змішаної стратегії. Такий принцип приводить до деякої послідовності партій гри, для кожної з яких можна підрахувати наближені значення оптимальних змішаних стратегій кожного з гравців, а також нижню та верхню границі для ціни гри.

У першій партії гравці P1, P2 вибирають довільно свої чисті стратегії, відповідно, i_1 та j_1. Формується m-вимірний вектор $\mathbf{x}^1 = (0 , .. , 0, 1, 0, ..., 0)$, в якому 1 міститься на i_1-у місці і який характеризує частоти вибору гравцем P1 своїх чистих стратегій, та n-вимірний вектор $\mathbf{y}^1 = (0, ..., 0, 1, 0, ..., 0)'$, де 1 розміщена на j_1-у місці і який характеризує частоти вибору чистих стратегій гравцем P2. Нехай після s партій маємо $x^s = (x_1 , ..., x_j^s, ..., x_m^s)$ та $y^s = (y_1^s, ..., y_j^s, ..., y_n^s)'$ – вектори частот вибору чистих стратегій гравцями P1 та P2 відповідно.

Визначаючи свій вибір i_{s+1}у $(s+1)$-й партії, гравець P1 підраховує свій середній програш

7.6 Зведення задач теорії ігор до задач лінійного програмування

Розглянемо матричну гру $n\times m$, задану матрицею $[a_{ij}]$.

Позначимо через $p^* = (p_1^*, p_2^*,..., p_n^*)$, $q^* =(q_1^*, q_2^*,..., q^*_m)$ імовірності оптимальних змішаних стратегій відповідно першого та другого гравців, а через v- ціну гри. Без доведення сформулюємо теорему.

Теорема. Для того, щоб число v було ціною гри, а p^* і q^* - векторам иймовірностей оптимальних стратегій, необхідне й достатнє виконання системи нерівностей:

$$\begin{cases} \sum_{i=1}^{n} a_{ij} p_i^* \geq v, \ j = \overline{1,m}, \\ \sum_{j=1}^{m} a_{ij} q_j^* \leq v, \ i = \overline{1,n}. \end{cases} \quad (7.8\text{-}7.9)$$

Для простоти припустимо, що $v > 0$. Цього завжди можна досягнути завдяки додаванню до всіх елементів матриці $[a_{ij}]$ одного й того ж постійного числа k. Така процедура не вплине на оптимальні стратегії, а тільки збільшить ціну гри на k.

Мета першого гравця полягає в отриманні максимального виграшу, а другого – мінімального програшу. Користуючись сформульованою теоремою, задачу максимізації гарантованого виграшу першого гравця і задачу мінімізації гарантованого програшу другого гравця можна представити як пару взаємнодвоїстих задач лінійного програмування:

Завдання I гравця

$$\begin{aligned}
&Z^* = v \to \max \\
&\sum_{i=1}^{n} a_{ij} p_i^* \geq v,\ j = \overline{1,m} \\
&\sum_{i=1}^{n} p_i^* = 1 \\
&p_i^* \geq 0,\ i = \overline{1,n}.
\end{aligned} \tag{7.10}$$

Завдання II гравця

$$\begin{aligned}
&F^* = v \to \min \\
&\sum_{j=1}^{m} a_{ij} q_j^* \leq v,\ i = \overline{1,n} \\
&\sum_{j=1}^{m} q_j^* = 1 \\
&q_j^* \geq 0,\ j = \overline{1,m}.
\end{aligned} \tag{7.11}$$

Запишемо систему обмежень (7.10) у розширеній формі:

$$\begin{aligned}
&a_{11} p_1^* + a_{21} p_2^* + \ldots + a_{n1} p_n^* \geq v \\
&\vdots \\
&a_{1m} p_1^* + a_{2m} p_2^* + \ldots + a_{nm} p_n^* \geq v \\
&p_1^* + p_2^* + \ldots + p_n^* = 1.
\end{aligned} \tag{7.12}$$

Поділимо праву й ліву частину нерівності на v. Отримуємо:

$$\begin{aligned}
&a_{11} p_1^* / v + a_{21} p_2^* / v + \ldots + a_{n1} p_n^* / v \geq v / v \\
&\vdots \\
&a_{1m} p_1^* / v + a_{2m} p_2^* / v + \ldots + a_{nm} p_n^* \geq v / v \\
&p_1^* / v + p_2^* / v + \ldots + p_n^* / v = 1 / v
\end{aligned} \tag{7.13}$$

Зробимо заміну:$\frac{p_i^*}{v} = x_i, i = 1, n$. Маємо:

$$\begin{aligned} &a_{11}x_1 + a_{21}x_2 + \ldots + a_{n1}x_n \geq 1 \\ &\vdots \\ &a_{1m}x_1 + a_{2m}x_2 + \ldots + a_{nm}x_n \geq 1. \end{aligned} \tag{7.14}$$

Відомо, що $\sum_{i=1}^{n} p_i^* = 1$. Отже, $\sum_{i=1}^{n} \frac{p_i^*}{v} = \sum_{i=1}^{n} x_i$. Звідси: $\sum_{i=1}^{n} x_i = \frac{1}{v}$.

Оскільки метою першого гравця є максимальний виграш, то він має забезпечити $min\frac{1}{v}$. Отже, визначення оптимальної стратегії для першого гравця зводиться до вирішення такої задачі лінійного програмування:

Знайти $Z = x_1 + x_2 + \ldots + x_n \rightarrow min$ при виконанні умов

$$\begin{aligned} &a_{11}x_1 + \ldots + a_{n1}x_n \geq 1 \\ &\vdots \\ &a_{1m}x_1 + \ldots + a_{nm}x_n \geq 1 \\ &x_i \geq 0, i = \overline{1, n}. \end{aligned} \tag{7.15}$$

Аналогічні міркування можна провести відносно задачі другого гравця. Виконавши заміну $\frac{q^*}{v} = y_j$, отримаємо наступну задачу:

Знайти $F = y_1 + y_2 + \ldots + y_m \rightarrow max$ при виконанні умов

$$\begin{aligned} &a_{11}y_1 + \ldots + a_{1m}y_m \leq 1 \\ &\vdots \\ &a_{n1}y_1 + \ldots + a_{nm}y_n \leq 1 \\ &y_m \geq 0, i = \overline{1, m}. \end{aligned} \tag{7.16}$$

Задача другого гравця є двоїстою до задачі першого гравця.

Процес знаходження розв'язку гри з використанням лінійного програмування складається з таких етапів:

I – побудова пари двоїстих задач лінійного програмування, еквівалентних даній матричній грі;

II – знаходження оптимальних планів пари двоїстих задач;

III – знаходження розв'язку гри, використовуючи співвідношення між планами двоїстих задач і оптимальними стратегіями та ціною гри.

Приклад 7.3. Дві фірми можуть вкласти свій наявний капітал у будівництво п'яти об'єктів. Стратегія фірм: *i*-та стратегія полягає у фінансуванні *i*-го об'єкту (i = 1, 2, …,5). Враховуючи особливості вкладів й інші умови, прибуток фірми виражається за допомогою матриці А:

$$A = \begin{bmatrix} 2 & 0 & 1 & -3 & 4 \\ 0 & -2 & -1 & 4 & 3 \\ 2 & 0 & 3 & 2 & -1 \\ 0 & -3 & 1 & 0 & 1 \\ 3 & 0 & 0 & -2 & 2 \end{bmatrix}.$$

Прибуток першої фірми є величиною збитку другої, тобто дану гру можна розглядати як гру двох осіб з нульовою сумою.

Розв'язання. Для зведення даної задачі до задачі лінійного програмування необхідно, насамперед, до кожного елементу матриці додати число k = 4 (необхідно, щоб усі елементи матриці А були додатніми). Отримуємо:

$$A = \begin{bmatrix} 6 & 4 & 5 & 1 & 8 \\ 4 & 2 & 3 & 8 & 7 \\ 6 & 4 & 7 & 6 & 3 \\ 4 & 1 & 5 & 4 & 5 \\ 7 & 4 & 4 & 2 & 6 \end{bmatrix}.$$

Вводимо невідомі величини $x_1, x_2, \ldots, x_5$. Тоді отримуємо таку задачу лінійного програмування:

Знайти $Z = x_1 + x_2 + x_3 + x_4 + x_5 \rightarrow min$

при виконанні умов

$$\begin{aligned}
&6x_1 + 4x_2 + 6x_3 + 4x_4 + 7x_5 \geq 1\\
&4x_1 + 2x_2 + 4x_3 + x_4 + 4x_5 \geq 1\\
&5x_1 + 3x_2 + 7x_3 + 5x_4 + 4x_5 \geq 1\\
&x_1 + 8x_2 + 6x_3 + 4x_4 + 2x_5 \geq 1\\
&8x_1 + 7x_2 + 3x_3 + 5x_4 + 6x_5 \geq 1\\
&x_i \geq 0, \quad i = \overline{1,5}.
\end{aligned}$$

Розв'язком цієї задачі буде

$x_1 = 0; x_2 =0; x_3 = 0.125; x_4 = 0; x_5 = 0.125.$

Оскільки $x_1 + x_2 + x_3 + x_4 + x_5 =1/v$, то

$$v = \frac{1}{x_1 + x_2 + x_3 + x_4 + x_5} = \frac{1}{0.125 + 0.125} = \frac{1}{0.25} = 4.$$

Знаючи те, що $x_i = p_i/v$, отримуємо: $p_i = x_i \cdot v$; $p_1 = 0; p_2 = 0; p_3 = 0.5; p_4 = 0; p_5 = 0.5.$

Побудуємо до даної задачі двоїсту. За невідомі візьмемо y_1, y_2, y_3, :y_4, y_5. Маємо

$$F = y_1 + y_2 + y_3 + y_4 + y_5 \rightarrow max$$

при виконанні умов

$$\begin{aligned}
&6y_1 + 4y_2 + 5yx_3 + y_4 + 8y_5 \leq 1\\
&4y_1 + 2y_2 + 3y_3 + 8y_4 + 7y_5 \leq 1\\
&6y_1 + 4y_2 + 7y_3 + 6y_4 + 3y_5 \leq 1\\
&4y_1 + y_2 + 5y_3 + 4y_4 + 5y_5 \leq 1\\
&7y_1 + 4y_2 + 4y_3 + 2y_4 + 6y_5 \leq 1\\
&y_i \geq 0, \quad i = \overline{1,5}.
\end{aligned}$$

Розв'язком цієї задачі буде

$y_1 = 0; y_2 = 0,25; y_3 = 0;$:$y_4 = 0; y_5 = 0$. Користуючись формулою $y_i = q_i / v$, знайдемо:

$$q_i = y_i \cdot v; q_1 = 0; q_2 = 1; q_3 = 0; q_4 = 0; q_5 = 0.$$

Отже, вектори ймовірностей оптимальних змішаних стратегій відповідно для першої та другої фірми будуть:

$p = (0; 0; 0.5; 0; 0.5); q = (0; 1; 0; 0; 0)$, а ціна початкової гри $v^* = v - 4 = 0$.

ЛЕКЦІЯ 8 «Позиційні ігри та ігри декількох осіб»

Анотація

Поняття про позиційні ігри. Кооперативні ігри та методи їх дослідження.Некооперативна гра двох осіб. Кооперативна гра двох осіб. Переговорна множина. Схема алгоритму переговорної множини.

8.1 Поняття про позиійні ігри

Позиційна гра – це природне розширення матричної гри двох гравців з нульовою сумою, в якій може брати участь скінченна кількість гравців, кожен з яких може робити послідовно скінченну кількість ходів, причому деякі з них можуть бути випадковими, а інформація про них може змінюватися від одного до іншого ходу. Такі ігри можуть бути формалізовані, певним чином перетворені до гри, що еквівалентна деякій матричній грі двох осіб з нульовою сумою. Процес приведення позиційної гри до матричної називається нормалізацією, а отримана матрична гра – грою в нормальній формі.

Розглянемо наступний приклад. Дві корпорації мають бажання встановити між собою ділові зв'язки і вирішити питання про побудову на території другої корпорації виробництва. Гра складається з трьох ходів. Перша корпорація обирає число з множини $\{1, 2\}$. Після цього друга корпорація обирає з множини двох можливих $\{1, 2\}$, знаючи, який вибір здійснила перша корпорація на першому ході. Третій хід робить перша корпорація: знаючи, який хід зробила друга корпорація, та пам'ятаючи про свій вибір на першому кроці, обирає число з множини $\{1, 2\}$. На цьому гра завершується і розподіляються виграші: перший гравець виплачує другому певну суму, визначену функцією $M(x,y, z)$ яка визначена наступним чином в залежності від вибору гравцями 1-го – 3-го ходів x,утаz відповідно:

$M(1, 1, 1) = -2;$ $M(1, 1, 2) = -1;$ $M(1, 2, 1) = 3;$ $M(1, 2, 2) = -4;$

$M(2, 1, 1) = 5;$ $M(2, 1, 2) = 2;$ $M(2, 2, 1) = 2;$ $M(2, 2, 2) = 6.$

Змістова інтерпретація цієї гри є наступною:

Хід 1. 1-а корпорація здійснює вибір з двох альтернатив: $x = 1$ – запропонувати 2-й побудувати на її теорії складальне виробництво комп’ютерів, $x = 2$ – побудувати виробництво мікропроцесорів.

Хід 2. 2-а корпорація, знаючи, яку альтернативу обрала 1-а на першому ході, здійснює вибір з двох альтернатив: $y = 1$– будувати складальне виробництво та запропонувати це 1-й корпорації; $y = 2$ – будувати виробництво мікропроцесорів та запропонувати це 1-й.

*Хід 3***.** 1-а корпорація, знаючи вибір 2-ї на другому ході та пам’ятаючи свій вибір на першому ході, здійснює вибір з двох альтернатив: $z = 1$– погодитися з пропозицією 2-ї,$z = 2$ – не погодитися з пропозицією 2-ї.

Після того, як зроблені ходи, партія зіграна, і 1-а корпорація отримує суму $M(x,y, z)$.

Для нормалізації цієї гри необхідно відтворити стратегіях 1-го та 2-го гравця.

Стратегії 2-го гравця:

B_1– обрати $y = 1$ не зважаючи на вибір 1-го гравця на першому ході;

B_2– обрати $y = 2$не зважаючи на вибір 1-го гравця на першому ході;

B_3– погодитися з вибором 1-го гравця на першому ході, тобто обрати $y = 1$, якщо $x = 1$, і $y = 2$, якщо$x = 2;$

B_4– не погодитися з вибором 1-го гравця на першому ході, тобто обрати $y = 2$, якщо $x = 1$, і $y = 1$, якщо$x = 2$.

Таким чином 2-й гравець має 4 стратегії.

Стратегії першого гравця будуються аналогічно з врахуванням раніше зроблених виборів; вибір на першому кроці дає дві можливості, після вибору другого гравця з’являється чотири варіанти, і реалізація на третьому ході – 8 стратегій дії для 1-го гравця Таким чином, стратегію першого гравця зображатимемо за допомогою трійки (i, i_1, i_2) – де i – вибір 1-го гравця на 1-му ході; i_1–вибір 1-го гравця на 3-му ході за умови вибору на 2-му ході 2-м гравцем$y = 1;$ i_2–вибір 1-го гравця на 3-му ході за умови вибору на 2-му ході 2-м гравцем$y = 2$.

Враховуючи відтворені стратегії, будуємо матрицю цієї гри:

	B_1	B_2	B_3	B_4
A_1(1, 1, 1)	M(1, 1, 1) =-2	M(1, 2, 1) =3	M(1, 1, 1) =-2	M(1, 2, 1) =3
A_2(1, 1, 2)	M(1, 1, 1) =-2	M(1, 2, 2) =-4	M(1, 1, 1) =-2	M(1, 2, 2) =-4
A_3(1, 2, 1)	M(1, 1, 2) =-1	M(1, 2, 1) =3	M(1, 1, 2) =-1	M(1, 2, 1) =3
A_4(1, 2, 2)	M(1, 1, 2) =-1	M(1, 2, 2) =-4	M(1, 1, 2) =-1	M(1, 2, 2) =-4
A_5(2, 1, 1)	M(2, 1, 1) =5	M(2, 2, 1) =2	M(2, 2, 1) =2	M(2, 1, 1) =5
A_6(2, 1, 2)	**M(2, 1, 1) =5**	M(2, 2, 2) =6	M(2, 2, 2) =6	**M(2, 1, 1) =5**
A_7(2, 2, 1)	M(2, 1, 2) =2	M(2, 2, 1) =2	M(2, 2, 1) =2	M(2, 1, 2) =2
A_8(2, 2, 2)	M(2, 1, 2) =2	M(2, 2, 2) =6	M(2, 2, 1) =6	M(2, 1, 2) =2

Розв’язком гри є дві сідлові точки, ціна гри – 5 .

Побудуємо дерево позиційної гри. В цьому дереві вузол позначатиме номер гравця, що робить хід, а другі – його хід. Листя дерева відображатимуть виграші, а кожний шлях від кореня до листка – партію.

Для відображення необхідних даних про зроблені вибори при певних ходах гравців за умов їх різної інформованості на різних ходах на дереві позиційної гри пунктиром позначатимемо інформаційні множини вузлів.

Оскільки 1-й хід робить 1-й гравець, то корінь відповідатиме ходу 1-го гравця та позначається 1. 2-й гравець робить 2-й хід, а тому вузли наступного рівня позначені 2, і так як йому відомий вибір 1-го гравця на першому ході, то він, здійснюючи свій хід, в момент здійснення ходу знає точно, де він (на якій гілці дерева) знаходиться – а тому кожен вузол нижнього рівня утворює окрему інформаційну множину (внаслідок повного знання ходів кожен з вузлів дерева цієї гри є окремою інформаційною множиною).

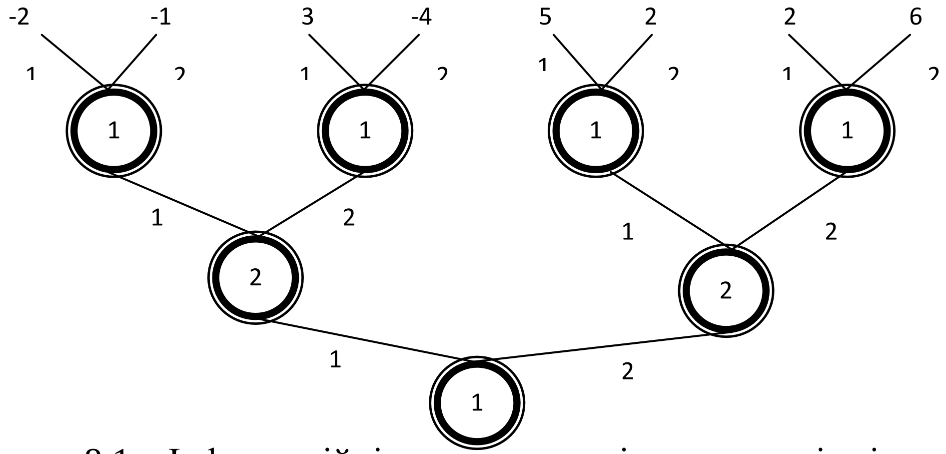


Рисунок8.1 – Інформаційні множини вузлів при знанні всіх ходів

Приклад 2. Порівняно з попереднім прикладом третій хід робить 1-й гравець, але вже не пам'ятаючи про те, який хід він зробив першим та не знаючи, який другий хід зробив другий гравець (1-го гравця можна уявити, як 2 особи, що знаходяться в окремих кімнатах та які не мають змоги обмінюватися інформацією). Відповідне дерево гри з інформаційними множинами наведене на рис.8.2.

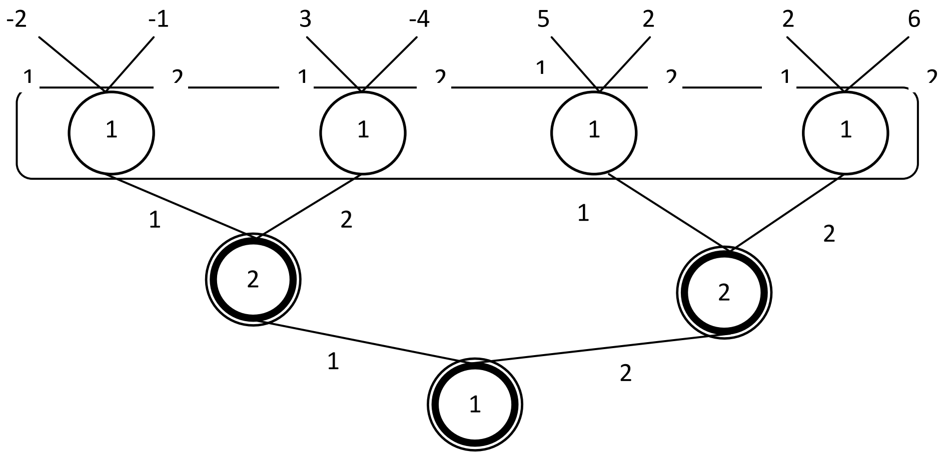


Рисунок8.2 – Інформаційні множини при частковому знанні

Проведемо гру до нормальної форми. У 2-го гравця є 4 таких же стратегії, як і в попередньому випадку. У 1-го гравця можливості зменшуються за рахунок браку інформації: оскільки він на 3-му ході не знає попередніх виборів, то його стратегія складається з пари чисел *(x, z)*, тобто обрати 1 або 2 на 1-му ході та 1 або 2 на 3-му ході. Відповідна матриця гри буде мати наступний вигляд:

	B_1	B_2	B_3	B_4
$A_1(1, 1)$	M(1, 1, 1) =-2	M(1, 2, 1) =3	M(1, 1, 1) =-2	M(1, 2, 1) =3
$A_2(1, 2)$	M(1, 1, 2) =-1	M(1, 2, 2) =-4	M(1, 1, 2) =-1	M(1, 2, 2) =-4
$A_3(2, 1)$	M(2, 1, 1) =5	M(2, 2, 1) =2	M(2, 2, 1) =2	M(2, 1, 1) =5
$A_4(2, 2)$	M(2, 1, 2) =2	M(2, 2, 2) =6	M(2, 2, 2) =6	M(2, 1, 2) =2

Отримана гра не має сідловок точки. Розв’язуючи гру в мішаних стратегіях, отримаємо мішані стратегії 1-го гравця – (0, 0, 4/7, 3/7), 2-го гравця – *(4/7, 3/7, 0, 0)*, та ціна гри становитиме $v = 26/7$. Таким чином, в загальному випадку втрата інформації зменшує ціну гри.

8.2 Кооперативні ігри та методи їх дослідження

Конфлікті ситуації не завжди мають антагоністичнийхарактер. Дуже часто учасники конфлікту, переслідуючи свої цілі, готові вступити в переговори один з одним, укладаючи деякі угоди чи навіть об’єднати свої зусилля в надії отримати з цього переваги.

Одним з найважливіших висновків теорії ігор – це те, що певні форми кооперування гравців за умови зовнішньо різних їх прагнень дійсно мають сенс. Це частково пояснюється великою цінністю інформації., яка може бути передана від одного до іншого учасника гри, зростаючим значенням рішень, що приймаються спільно, синергічним ефектом від хоча б часткового об’єднання ресурсів.

Ігри двох осіб посідають центральне місце у всій теорії ігор. Серед таких ігор виділяються біматричні ігри. Причини цього наступні. Біматричні ігри адекватніше відображають реальні конфлікти двох осіб порівняно з матричними, що використовуються для описання конфліктів з повною суперечністю і несумісністю інтересів, з відсутністю можливості компромісів; в таких конфліктах зростання прибутку одного гравця завжди означає збільшення втрат іншого. В житті такі конфлікти зустрічаються порівняно не частою Можна було б піддатися спокусі розглядати війну як крайній приклад зіткнення інтересів, але, взагалі кажучи, війна не є строгим суперництвом, оскільки обидві сторони вважають нічию кращим результатом, аніж взаємне знищення. Хоча локальне зіткнення або повітряний бій, очевидно, доцільно розглядати як гру зі строгим суперництвом. Але взагалі для описання військових конфліктів оперативно-тактичного плану, для яких приманна наявність не менш ніж двох ієрархічно пов'язаних ланок «напад – оборона», звичайно як адекватні використовуються біматричні ігри.

Біматричні ігри – це скінченні ігри двох осіб з довільною сумою, тобто такі, для яких не обов'язково виконується умова a_{ij} = -b_{ij} (тобто виграш одного з гравців – це програш іншого. Ці ігри описуються або за допомогою двох матриць A та B – виграшів кожного гравця, або ж за допомогою складної матриці, елементами якої є пари (a_{ij}, b_{ij}).

$$\begin{matrix} a_{11}, b_{11} & a_{12}, b_{12} & \cdots & a_{1n}, b_{1n} \\ a_{21}, b_{21} & a_{22}, b_{22} & \cdots & a_{2n}, b_{2n} \\ \cdots & \cdots & & \cdots \quad \cdots \\ a_{m1}, b_{m1} & a_{m2}, b_{m2} & \cdots & a_{mn}, b_{mn} \end{matrix} \qquad (8.1)$$

Біматрична гра – це найпростіша ігрова модель, що припускає можливість співробітництва.

Гра з нестрогим суперництвом, є грою, в якій не має однозначного суперника, тому що існує принаймні одна пара ситуацій x та x^*, така, що один гравець віддає перевагу ситуації x^*перед ситуацією x. Для таких ігор неможливо обрати функції корисності гравців так, щоб сума їх дорівнювала нулю; тому терміни «ігри з нестрогим суперництвом» та «ігри з ненульовою сумою» є еквівалентними. Більшість економічних, політичних і військових конфліктів інтересів можна представити у формі ігор лише в тому випадку, якщо визнати властиве їх нестроге суперництво.

В іграх зі строгим суперництвом гравці не можуть досягати обопільної вигоди посередництвом якого-небудь співробітництва; в іграх з нестрогим суперництвом, навпаки, такий обопільний виграш завжди можливий.

Існують дві різновидності бінарних ігор – безкоаліційні, що забороняють будь-яку співробітництво, та кооперативні, що дозволяють співробітництво.

Результат біматричної безкоаліційної гри рідко можна передбачити, оскільки зв'язки між виграшами сторін відсутні, і з'являється можливість діяти самостійніше.

Значення середніх виграшів гравців A та B в біматричній грі рівні відповідно:

$$M_A\ A, x, y = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j \text{ та } M_B\ B, x, y = \sum_{i=1}^{m} \sum_{j=1}^{n} b_{ij} x_i y_j. \quad (8.2)$$

Ситуація рівноваги для бінарної гри – це така пара *(x, y)*для якої виконується співвідношення

$$\begin{aligned} \sum_{j=1}^{n} a_{ij} y_j &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_i y_j, i = 1, m \\ \sum_{i=1}^{m} b_{ij} x_i &= \sum_{i=1}^{m} \sum_{j=1}^{n} ab_{ij} x_i y_j, j = 1, n \end{aligned} \quad (8.3)$$

та природні обмеження:

$$x_i \geq 0, \sum_{i=1}^{m} x_i = 1, i = 1, m, y_j \geq 0, \sum_{j=1}^{n} y_j = 1, j = 1, n \quad (8.4)$$

Такі ситуації рівноваги в бінарній грі існують завжди – відповідна теорема про існування ситуації рівноваги в мішаних стратегіях для скінчених неантагоністичних ігор двох осіб доведена Нешем (NashJ.F.).

Однак ця теорема не дає інформації про те, яким чином знайти ці ситуації рівноваги. Різноманітні алгоритми для знаходження всіх ситуацій рівноваги запропоновані багатьма авторами. Таким чином, виграші гравців у біматричних іграх задаються звичайними матричними добутками:

$$H_1(X,Y) = XAY^T, \; H_2(X,Y) = XBY^T \tag{8.5}$$

Розглянемо тепер основні ідеї ігор двох осіб з ненульовою сумою.

8.3 Некооперативна гра двох осіб

Нехай задана гра двох осіб з матрицею (8.1).

В теорії розглядаються в основному дві стратегії поведінки гравців – це *максимінна стратегія* і так звана *стратегія загрози*.

Максимінна стратегія– це стратегія украй обережної людини, яка, розраховуючи на якнайгіршу ситуацію, хотіла би мати в цьому випадку максимум можливого.

Якщо один з гравців застосовує свою оптимальну стратегію $x^* = (x_1^*, ..., x_m^*)$, то інший не може покращити своє становище, тобто для оптимальної стратегії справедливі співвідношення

$$\sum_{i=1}^{m} a_{ij}x_i \geq v, j = \overline{1,n}, x_i \geq 0, \; \sum_{i=1}^{m} x_i = 1, i = \overline{1,m} \text{ за умови } v \Rightarrow Max.$$

Перетворимо цю задачу здійснивши підстановку $p_i = \frac{x_i}{v}$, і отримаємо:

$$\sum_{i=1}^{m} a_{ij}p_i \geq v, j = \overline{1,n}, \quad \sum_{i=1}^{m} \frac{x_i}{v} = \frac{1}{v} = \sum_{i=1}^{m} p_i \Rightarrow Min, \text{ тому що } v \Rightarrow Max.$$

Здійснивши підстановку $p_i = \frac{x_i}{v}$ та враховуючи, що гравець *A* прагне максимізувати свій середній виграш, отримаємо задачу лінійного програмування, розв’язання якої дозволить визначити оптимальну стратегію гравця *A*:

$$\frac{1}{v} = \sum_{i=1}^{m} p_i \Rightarrow Min, \; \sum_{i=1}^{m} a_{ij}p_i \geq 1, p_i \geq 0, j = \overline{1,n}, \tag{8.6}$$

Друга можлива стратегія – це *стратегія загрози*, за якої гравець ставить за мету не виграти самому, а дати можливість виграти другому гравцеві якнайменше (при цьому в бінарній грі він і сам може виграти найменше!).

Станемо знову на позицію першого гравця. Хай він знову застосовує мішану мінімаксну стратегію (прагне мінімізувати виграш 2-го гравця) $x^{*} = (x_1^{*}, \ldots, x_m^{*})$. Таким чином, застосовуючи її, він рахує *не свій виграш, а виграш другого гравця*. Якщо другий гравець робить хід jто його середній виграш становитиме $\sum_{i=1}^{m} b_{ij}x_i$. Перший гравець діє за принципом $\max_j \sum_{i=1}^{m} b_{ij}x_i \Rightarrow min$, тобто він *мінімізує максимальний виграш другого гравця*. Якщо позначити виграш другого гравця через w, то ми маємо:

$$\sum_{i=1}^{m} b_{ij}x_i \geq w, j = 1, n, x_i \geq 0, \ \sum_{i=1}^{m} x_i = 1, i = 1, m \ \textit{за умови}\ w \Rightarrow Min.$$

Перетворимо цю задачу, здійснивши підстановку $p_i = \frac{x_i}{w}$, і отримаємо:

$$\sum_{i=1}^{m} b_{ij}p_i \geq v, j = 1, n, \quad \sum_{i=1}^{m} \frac{x_i}{w} = \frac{1}{w} = \sum_{i=1}^{m} p_i \Rightarrow Max, \ \textit{тому що}\ w \Rightarrow Min.$$

Здійснивши підстановку $p_i = \frac{x_i}{w}$ та враховуючи, що гравець прагне мінімізувати середній виграш гравця *B*, отримаємо задачу лінійного програмування, розв’язання якої дозволить визначити оптимальну стратегію гравця *A*:

$$\frac{1}{w} = \sum_{i=1}^{m} p_i \Rightarrow Max, \ \sum_{i=1}^{m} b_{ij}p_i \geq 1, p_i \geq 0, j = 1, n, \tag{8.7}$$

Розглянемо докладніше випадок $n = m = 2$. Тоді матриця виплат гри має вигляд:

$$\begin{array}{cccc} & B_1 & B_2 & \\ A_1 & a_{11}, b_{11} & a_{12}, b_{12} & x_1 \\ A_2 & a_{21}, b_{21} & a_{22}, b_{22} & x_2 = 1 - x_1 \\ & y_1 & y_2 & \end{array} \tag{8.8}$$

Відобразимо геометричну максимінну стратегію і стратегію загрози першого гравця. Почнемо з максимінної стратегії. Нехай перший гравець обирає стратегію A_1 з ймовірністю x_1, $x_2 = 1-x_1$. За умови вибору гравцем B стратегії B_1 середній виграш першого гравця становитиме $a_{11}x_1 + a_{21}(1 - x_1)$.

За умови вибору гравцем B стратегії B_2 середній виграш першого гравця становитиме $a_{12}x_1 + a_{21}(1 - x_1) = a_{12}x_1 + a_{22}(1 - x_1) = v$, отримаємо оптимальне значення середнього виграшу гравцяA у випадку застосування ним своєї максимінної стратегії. Аналогічні рівняння можна виписати також і для випадку стратегії загрози.

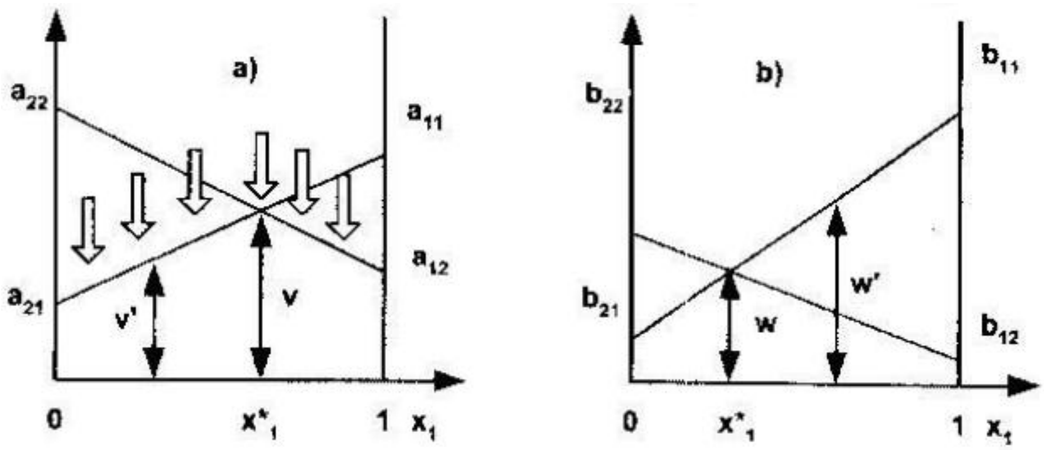


Рисунок 8.3 – Геометрична ілюстрація максимінної стратегії та стратегії загрози

У випадку максимінної стратегії гравець A гарантує собі виграш, не менший за v (рис. 8.3а), але при цьому згоджується з тим, що припускає і більший виграш гравця B порівняно з застосуванням стратегії загрози (значення w'на рис. 8.3б).

І навпаки, діючи з позиції стратегії загрози, гравець A гарантує для гравця B отримання мінімального виграшу, але разом з тим і зменшує свій середній виграш порівняно з максимінною стратегією (відомий підхід з точки зору «нехай мені буде гірше, але щоб у сусіда було найгірше»).

8.4 Кооперативна гра двох осіб. Переговорна множина

Розглянемо гру з наступною матрицею виплат:

$$\begin{array}{ccc} & y & 1-y \\ x & 2,1 & -1,-1 \\ 1-x & -1,-1 & 1,2 \end{array}$$

Нехай гравець *A* використовує мішану стратегію *(x; 1–x)*, *B* – стратегію *(y; 1-y)*. Тоді середні виграші гравців становитимуть:

$$v = 2xy - 1x\ 1-y\ \ -1y\ 1-x\ \ +1(1-x)(1-y),$$

$$w = 1xy - 1x\ 1-y\ \ -1y\ 1-x\ \ +2(1-x)(1-y)$$

Таким чином, ми побудували відображення *x, y*в *v, w*. $0 \le x \le 1$, $0 \le y \le 1$. Відображаючи всі можливі варіанти *x, y*в простір*v, w*, отримаємо наступну фігуру (рис. 8.4), обмежену прямими, що проходять через пари точок (-1; -1), (1, 2) і (-1, -1), (2, 1), а також шматком параболи $5(v-w)^2 = 2(v+w)-1$. В ній є «провал», обмежений саме цією параболою.

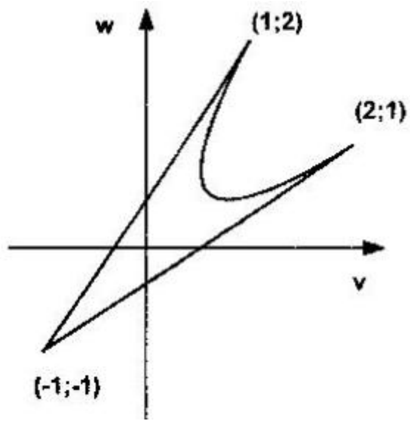


Рисунок 8.4 – Відображення прикладу кооперативної гри за умови повного суперництва

Повернемося до загального випадку гри двох осіб з матрицею виплат (8.1) і припустимо, що гравці мають нагоду домовлятися про сумісні дії. В чому полягають ці сумісні дії?

Раніше стратегія A_i першого гравця обиралася з вірогідністю x_i, стратегія B_jз вірогідністюy_j, і стратегії обох гравців були незалежні, так що комбінація *(A_i, B_j)*з’являлася з вірогідністю. $x_i y_j$. Зараз ходи обираються спільно і тому комбінація стратегій *(A_i, B_j)*з’являється з деякою сумісною вірогідністю p_{ij} Сумісна гра зводиться таким чином, до вибору сумісної мішаної стратегії p_{ij} При цьому очевидно:

$$\forall\ i,j\ : pij \geq 0 \wedge \sum_{i=1}^{m} \sum_{j=1}^{n} pij = 1. \tag{8.9}$$

При такій сумісній мішаній стратегії середні виграші першого і другого гравців рівні відповідно:

$$v = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} p_{ij},\ w = \sum_{i=1}^{m} \sum_{j=1}^{n} b_{ij} p_{ij}. \tag{8.10}$$

Представимо собі множину *(v, w)*.Яку область заповнять значення, що отримані з наведених формул? Виявляється,що ця область є опуклою оболонкою образів точок з координатами *(a_{ij}, b_{ij})*.

Так для наведеного вище прикладу гри область *R*є опукла оболонка точок (-1, -1), (2, 1) і (1, 2) (див. рис. 8.5)

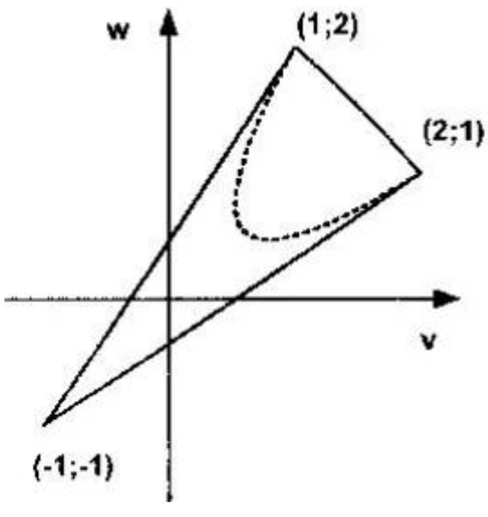


Рисунок 8.5 – відображення прикладу кооперативної гри за умови співробітництва

Порівнявши цю область з тією, яка зображена на рис. 8.4. бачимо, що застосування сумісних стратегій дозволило заповнити ту «западину», яка була при некооперативній грі.

Про що ж тепер можуть домовитися гравці? Нехай v^*іw^*-максимінні виграші першого і другого гравців відповідно (рис. 8.6). Нанесемо на нашу множину *R* точку з координатами точкою (v^*,w^*). Ця точка називається точкою **statusquo.** Очевидно, що жоден з гравців не погодиться одержувати в результаті сумісної гри менше ніж дає йому максимінна стратегія – навіщо йому така домовленість, якщо він може гарантувати собі без жодних домовленостей v^*абоw^*.

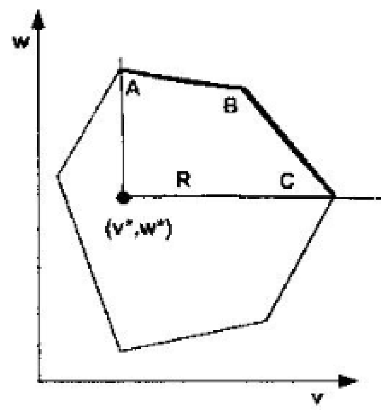


Рисунок 8.6 – Множина Парето-опримальних розв'язків гри

Тому з нашої множини розв'язків відразу зникає область , де $v<v^*$та $w<w^*$.

Розглянемо область, що тепер залишилася і визначається сумісною дією обмежень $(v \geq v^*) \wedge (w \geq w^*)$. Задачу можна розглядати як двокритерійну з критеріями (v^*,w^*).Як нескладно визначити, множину Парето-оптимальних (недомінованих) розв'язків утворюють всі точки, що знаходяться на ламаній *A* – *B* – *C*.

Очевидно, що якщо точка (v,w) домінується точкою (v',w'), то в процесі торгівлі гравці безболісно відмовляться від точки (v,w) на користь точки (v',w'),оскільки при такому переході хоча б одному стає краще, а іншому – не гірше. Очевидно, що точки, які домінують точку (v,w), знаходяться правіше і вище на множині R. Тому множина Парето-оптимальних розв'язків називається також переговорною множиною для гравців A та B. Переговорна множина на рис. 8.5 – це відрізок прямої, що сполучає точки (1, 2) і (2, 1).

До чого гравці домовляться – наперед сказати не можна, оскільки на цій множині інтереси гравців прямо протилежні. Результат залежить від уміння вести переговори і лежить за межами математичного дослідження.

Отже, в певному значенні, вирішити біматричну кооперативну гру двох осіб означає побудувати переговорну множину.

8.5 Схема алгоритму побудови переговорної множини

Крок 1. Побудова образу всіх можливих мішаних стратегій x, y,

$$x = (x_1, \dots, x_m),\ x_i \geq 0,\ \sum_{i=1}^{m} x_i = 1, i = \overline{1,m},$$

$$y = (y_1, \dots, y_n),\ y_j \geq 0,\ \sum_{j=1}^{n} y_j = 1, j = \overline{1,n}$$

в просторі (v,w).

Крок 2. Побудова опуклої оболонки образу припустимих стратегій і просторі (v,w).

Крок 3. Знаходження максимінних виграшів кожного з гравців – точки statusquo.

Крок 4. Побудова множини Парето-оптимальних розв'язків як «північно-східної границі» оптимального образу в просторі (v,w), для якої виконуються умови $(v \geq v^*) \wedge (w \geq w^*)$.

Слід відзначити, що конкретні реалізації алгоритмів для біматричної гри з числом стратегій, що більше 2, є достатньо складними.

ЛЕКЦІЯ 9 «Нечіткі множини»

Анотація

Еволюція теорії нечітких множин. Поняття нечітких множин. Основні характеристики нечітких множин. Функція приналежності. Операції над нечіткими числами. Операції над нечіткими множинами. Наочне представлення операцій над нечіткими множинами. Переваги та застосування нечітких систем

9.1 Еволюція теорії нечітких множин

Теорія множин – розділ математики, в якому вивчаються загальні властивості множин. Теорія множин лежить в основі більшості математичних дисциплін; вона зробила глибокий вплив на розуміння предмету самої математики.

До другої половини XIX століття поняття «множини» не розглядалося як математичне («множина книг на полиці», «множина людських чеснот» і т. д. – все це чисто побутові обороти мови). Положення змінилося, коли німецький математик Георг Кантор розробив свою програму стандартизації математики, в рамках якої будь-який математичний об'єкт повинен був виявлятися тою або іншою «множиною». Наприклад, натуральне число, по Канторові, слід було розглядати як множину, що складається з єдиного елементу іншої множини, званої «натуральним рядом», – який, у свою чергу, сам є множиною, що задовольняє так званим аксіомам Пеано.

Напевно, самою вражаючою людського інтелекту є здатність ухвалювати правильні рішення в умовах неповної і нечіткої інформації. Побудова моделей наближених роздумів людини і використання їх в комп'ютерних системах представляє сьогодні одну з найважливіших проблем науки.

Основи нечіткої логіки були закладені в кінці 60-х років в роботах відомого американського математика Лотфі Заде. Дослідження такого роду було викликано зростаючим незадоволенням експертними системами. Хвалений "штучний інтелект", який легко справлявся із завданнями управління складними технічними комплексами, був безпорадним при простих висловах повсякденному життю, типу "Якщо в машині перед тобою сидить недосвідчений водій – тримайся від неї подалі". Для створення дійсно інтелектуальних систем, здатних адекватно взаємодіяти з людиною, був необхідний новий математичний апарат, який переводить неоднозначні життєві твердження в мову чітких і формальних математичних формул. Першим серйозним кроком в цьому напрямі стала теорія нечітких множин, розроблена Заді. Його робота "Fuzzy Sets", опублікована в 1965 році в журналі "Information and Control", заклала основи моделювання інтелектуальної діяльності людини і стала початковим поштовхом до розвитку нової математичної теорії. Він же дав і назву для нової області науки - "fuzzy logic" (fuzzy – нечіткий, розмитий, м'який). На відміну від булевої алгебри, у котрій існує лише дві величини (0 та 1, правда чи неправда) у нечіткій логіці існують також перехідні величини (стани).

Існує легенда про те, яким чином була створена теорія "нечітких множин". Одного разу Заде мав довгу дискусію зі своїм другом щодо того, чия з дружин привабливіша. Термін "приваблива" є невизначеним і в результаті дискусії вони не змогли прийти до задовільного підсумку. Це змусило Заде сформулювати концепцію, яка виражає нечіткі поняття типу "приваблива" в числовій формі.

Подальші роботи професора Лотфі Заде і його послідовників заклали фундамент нової теорії і створили передумови для впровадження методів нечіткого управління в інженерну практику. Апарат теорії нечітких множин, продемонструвавши ряд багатообіцяючих можливостей застосування - від систем управління літальними апаратами до прогнозування підсумків виборів, виявився разом з тим складним для втілення. Враховуючи наявний рівень технології, нечітка логіка зайняла своє місце серед інших спеціальних наукових дисциплін - десь посередині між експертними системами і нейронними мережами.

Тріумфальний хід нечіткої логіки по світу почався після доказу в кінці 80-х Бартоломеєм Косько знаменитої теореми FAT (Fuzzy Approximation Theorem). У бізнесі і фінансах нечітка логіка отримала визнання після того, як в 1988 році експертна система на основі нечітких правил для прогнозування фінансових індикаторів єдина передбачила біржовий крах. І кількість успішних фаззі-застосувань в даний час обчислюється тисячами.

Третій період почався з кінця 80-х років і до цих пір. Цей період характеризується бумом практичного застосування теорії нечіткої логіки в різних сферах науки і техніки. До 90-го року з'явилося близько 40 патентів, що відносяться до нечіткої логіки (30 - японських). Сорок вісім японських компаній створюють лабораторію LIFE (Laboratory for International Fuzzy Engineering), японський уряд фінансує 5-річну програму по нечіткій логіці, яка включає 19 різних проектів, - від систем оцінки глобального забруднення атмосфери і передбачення землетрусів до АСОВІ заводських цехів. Результатом виконання цієї програми була поява цілого ряду нових масових мікрочіпів, що базуються на нечіткій логіці. Сьогодні їх можна знайти в пральних машинах і відеокамерах, цехах заводів і моторних відсіках автомобілів, в системах управління складськими роботами і бойовими вертольотами.

У США розвиток нечіткої логіки йде по шляху створення систем для великого бізнесу і військових. Нечітка логіка застосовується при аналізі нових ринків, біржовій грі, оцінки політичних рейтингів, виборі оптимальної цінової стратегії і тому подібне. З'явилися і комерційні системи масового застосування.

Зсув центру досліджень нечітких систем у бік практичних застосувань привів до постановки цілого ряду проблем, зокрема: нова архітектура комп'ютерів для нечітких обчислень; елементна база нечітких комп'ютерів і контроллерів; інструментальні засоби розробки; інженерні методи розрахунку і розробки нечітких систем управління, і тому подібне.

Основні зусилля кібернетики зараз направлені на створення штучного інтелекту, що не поступається людському мозку. Оскільки комп'ютери розуміють тільки мову математики, то виникла необхідність представлення нечітких понять у вигляді математичних змінних, названих лінгвістичними.

Сукупність лінгвістичних змінних (мало, багато ...) складає нечітка множина. Наприклад, вік (молодий + не молодий + старий + немолодий + .).

Для опису нечітких множин вводяться поняття нечіткої і лінгвістичної змінних.

Нечітка змінна описується набором *(N,X,A)*, де *N* – це назва змінної, *X* – універсальна множина (область міркувань), *A* – нечітка множина на *X*.

Значеннями лінгвістичної змінної можуть бути нечіткі змінні, тобто лінгвістична змінна знаходиться на більш високому рівні, чим нечітка змінна. Кожна лінгвістична змінна складається з:

— назви;

— множини своїх значень, яка також називається базовою терм–множиною *T*. Елементами базової терм-множини є назви нечітких змінних;

— універсальної множини *X*;

— синтаксичного правила *G*, за яким генеруються нові терми із застосуванням слів природної або формальної мови;

— семантичного правила *P*, яке кожному значенню лінгвістичної змінної ставить у відповідність нечітку підмножину множини *X*.

Розглянемо таке нечітке поняття як «Ціна акції». Це і є назва лінгвістичною змінною. Сформуємо для неї базову терм-множину, яка складатиметься з трьох нечітких змінних: «Низька», «Помірна», «Висока» і задамо область міркувань у вигляді *X=[100; 200]* (одиниць). Останнє, що залишилося зробити – побудувати функції приналежності для кожного лінгвістичного терма з базового терм-множини *T*.

Ступінь приналежності може трактуватися по різному в залежності від завдання, в якій використовується нечітка множина. Можливі трактування ступеня приналежності:

— ступінь відповідності поняттю *A*,

— ймовірність,

— можливість,

— корисність,

— істинність,

— правдоподібність,

— значення функції та ін.

Для кожної трактування ступеня приналежності розроблені свої методи побудови функцій приналежності. У ряді моделей м'яких обчислень функції приналежності задаються досить довільно в параметричному вигляді. Наприклад, функції приналежності нечітких множин можуть спочатку задаватися в моделі так, щоб вони "рівномірно покривали" область визначення *X*, а потім налаштовуватися в результаті зміни їх параметрів в процесі налагодження моделі.

9.2 Поняття нечітких множин

Хай *E* - універсальна множина, *x* - елемент *E*, а *R* – певна властивість. Звичайна (чітка) підмножина *A* універсальної множини *E*, елементи якої задовольняють властивість *R*, визначається як множина впорядкованої пари $A = \{\mu A\ (x)\ /x\}$, де $\mu A(x)$ – характеристична функція, що приймає значення 1, коли *x* задовольняє властивості *R*, і 0 – в іншому випадку.

Нечітка підмножина відрізняється від звичайної тим, що для елементів *x* з *E* немає однозначної відповіді "ні" щодо властивості *R*. У зв'язку з цим, нечітка підмножина *A* універсальної множини *E* визначається як множина впорядкованою парі $A = \{\mu A(x)\ /x\}$, де $\mu A(x)$ - характеристична функція приналежності (або просто функція приналежності), що приймає значення в деякій впорядкованій множині *M* (наприклад, $M = [0, 1]$).

Функція приналежності указує ступінь (або рівень) приналежності елементу *x* до підмножини А. Множину М називають множиною ознак. Якщо $M = \{0,1\}$, тоді нечітка підмножина *A* може розглядатися як звичайна або чітка множина.

Розглянемо множину *X* всіх чисел від 0 до 10. Визначимо підмножину *A* множини *X* всіх дійсних чисел від 5 до 8.

$A = [5, 8]$

Покажемо функцію приналежності множині *A*, ця функція ставить у відповідність число 1 або 0 кожному елементу в *X*, залежно від того, належить даний елемент підмножині *A* чи ні. Результат представлений на рисунку 9.1.

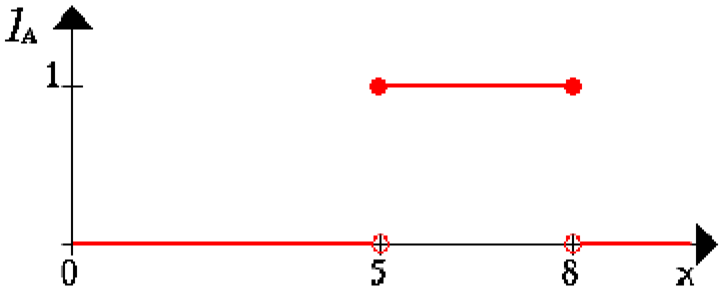


Рисунок9.1 – Зображення приналежності звичайної (чіткої) множини

Можна інтерпретувати елементи, відповідні 1, як елементи, що знаходяться в множині *A*, а елементи, відповідні 0, як елементи, що не знаходяться у множині А.

Ця концепція використовується в багатьох областях. Але існують ситуації, в яких даній концепції не вистачатиме гнучкості.

У даному прикладі опишемо множину молодих людей. Формально можна записати так

$$B = \{\text{ множина молодих людей}\}$$

Оскільки, взагалі, вік починається з 0, то нижня межа цієї множини повинна бути нулем. Верхню межу визначити складніше. Спочатку встановимо верхню межу, скажімо, рівну 20 рокам. Таким чином, маємо *B* як чітко обмежений інтервал, буквально: $B = [0, 20]$. Виникає питання: чому хтось в свій двадцятирічний ювілей – молодий, а відразу наступного дня вже не молодий? Очевидно, це структурна проблема, і якщо пересунути верхню межу в іншу точку, то можна поставити таке ж питання.

Природніший шлях створення множини *B* полягає в ослабленні строгого ділення на молодих і не молодих. Зробимо це, виносячи не тільки чіткі думки "Так, він належить множині молодих людей" чи ні, вона не належить множині молодих людей", але і гнучкі формулювання "Так, він належить до досить молодих людей" чи ні, він не дуже молодий".

Розглянемо як за допомогою нечіткої множини визначити вираз "він ще молодий".

У першому прикладі ми кодували всі елементи множини за допомогою 0 чи 1. Простим способом узагальнити дану концепцію є введення значень між 0 і 1. Реально можна навіть допустити нескінченне число значень між 0 і 1, в одиничному інтервалі *I = [0, 1]*.

Інтерпретація чисел при співвідношенні всіх елементів множини тепер складніша. Звичайно, число 1 відповідає елементу, що належить множині *B*, а 0 означає, що елемент точно не належить множині *B*. Всі інші значення визначають ступінь приналежності до множини *B*.

Для наочності приведемо характеристичну функцію множини молодих людей, як і в першому прикладі.

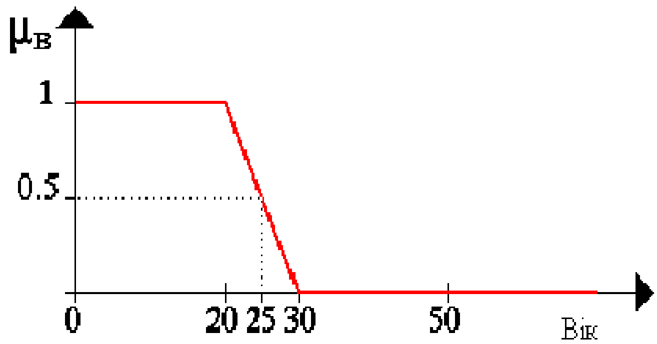


Рисунок 9.2 – Характеристична функція множини молодих людей

Хай *E = {x1, x2, x3, x4, x5}, M = [0, 1]*; *A* – нечітка множина, для якої

μA(x1)=0,3; μA(x2)=0; μA(x3)=1; μA(x4)=0,5; μA(x5)=0,9

Тоді *A* можна представити у вигляді:

A = {0,3/x1; 0/x2; 1/x3; 0,5/x4; 0,9/x5 } або

A = 0,3/x1 + 0/x2 + 1/x3 + 0,5/x4 + 0,9/x5

(знак "+" є операцією не складання, а об'єднання) або

	x_1	x_2	x_3	x_4	x_5
$A =$	0,3	0	1	0,5	0,9

Формалізуємо неточне визначення «гарячий чай». Як *x* (область існування) виступатиме шкала температури в градусах Цельсія. Очевидно, що вона буде змінюється від 0 до 100 градусів. Нечітка множина для поняття «Гарячий чай» може виглядати таким чином:

C = {0/0; 0/10; 0/20; 0,15/30; 0,30/40; 0,60/50; 0,80/60; 0,90/70; 1/80; 1/90; 1/100}.

Так, чай з температурою 60°C належить до множини «Гарячий» зі ступенем приналежності 0,80. Для однієї людини чай при температурі 60°C може опинитися гарячим, для іншого – не дуже гарячим. Саме у цьому і виявляється нечіткість завдання відповідної множини.

9.3 Основні характеристики нечітких множин

Хай *M = [0, 1]* і *A* – нечітка множина з елементами з універсальної множина E і з множиною визначення *M*

— Величина $\mu A(x)$ $\sup\limits_{x \in E}$ називається висотою нечіткої множини *A*. Нечітка множина *A* є нормальною, якщо її висота дорівнює 1, тобто верхня межа її функції приналежності дорівнює 1 ($\sup\limits_{x \in E} \mu A(x) = 1$). При $\mu A(x) < 1$ нечітка множина називається субнормальною.

— Нечітка множина є порожньою, якщо $\forall x \in E\ \mu A(x) = 0$. Не порожню субнормальну множину можна нормалізувати за формулою $\mu A(x) := \frac{\mu A(x)}{\sup\limits_{x \in E} \mu A(x)}$

— Нечітка множина є унімодальною, якщо $\mu A(x) = 1$ лише для одного *x* з *E*.

— Носієм нечіткої множини *A* є звичайна підмножина з властивістю μA(x) >0, тобто носій $A = \{x / \mu A(x) > 0\}\ \forall x \in E$

— Елементи x ∈ E, для яких μA(x)=0,5 називаються точками переходу множини A.

Наведемо приклади нечітких множин з їх характеристиками:

1. Хай *E = {0,1,2..,10}, M =[0, 1].* Нечітку множину "декілька" можна визначити таким чином:

"декілька" = 0,5/3+0,8/4+1/5+1/6+0,8/7+0,5/8;

її характеристики: – висота = 1,

– носій = {3,4,5,6,7,8},

– точки переходу – {3,8}.

2. Хай *E = {0,1,2,3...,n,...}*. Нечітку множину "малий" можна визначити:

$$\mu\text{"малий"}\ x\ =\ \mu_{малий}\ n\ = \frac{1}{1+\ \frac{n}{10}\ ^{2}}/n \tag{9.1}$$

3. Хай *E = {1,2,3...,100}* і відповідає поняттю "вік", тоді нечітку множину "молодий", можна визначити з допомогою

$$\mu\text{"молодий"}\ x\ =\ \begin{matrix} 1,\ \ x\ \in\ 1,25 \\ \frac{1}{1+\ \frac{x-25}{5}\ ^{2}},\ x \geq 25 \end{matrix} \tag{9.2}$$

$$\mu\ x\ = \frac{1}{1+\frac{x-a\ ^{2}}{b}}\ x$$

$$\mu\ x\ = a\ \overline{\frac{x}{x_{max}}\ \ 1-\frac{x}{x_{max}}}$$

Нечітка множина "молодий" на універсальній множині *E' ={ Іванов, Петров, Сидорів...}* задається за допомогою функції приналежності *μ"молодий"(x)* на *E = {1,2,3..100}* (вік), що називається відносно *E'* функцією сумісності, при цьому:

μ"молодий"(Сидорів)= μ"молодий"(x), де x – вік Сидорова.

4. Хай *E = { Запорожець, Жигулі, Мерседес....}* – множина марок автомобілів, а *E′ = [0,μ]* – універсальна множина "вартість", тоді на *E′* ми можемо визначити нечітку множине типу: "для небагатих", "для середнього класу", "престижні", з функціями приналежності типу:

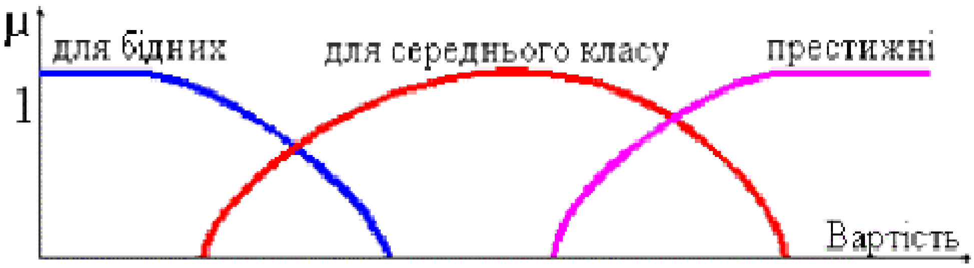


Рисунок 9.3 – Характеристична функція множини вартість автомобіля для людей різного достатку

Маючи ці функції і знаючи ціни автомобілів з *E* в даний момент часу, визначимо на *E′* нечіткі множини з цими ж назвами.

Так, наприклад, нечітка множина "для небагатих", задана на універсальній множині *E = { Запорожець, Жигулі, Мерседес....}* виглядає таким чином:

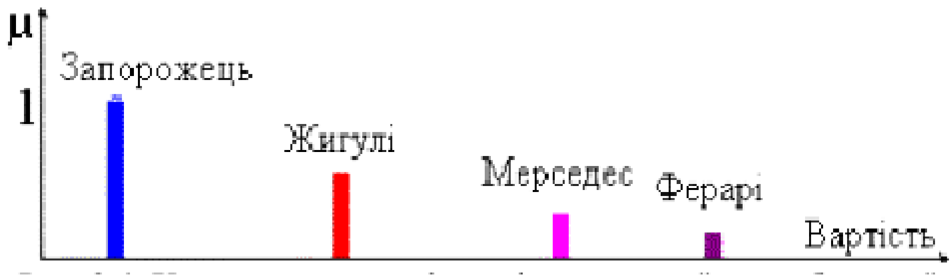


Рисунок 9.4 – Характеристична функція множини "для небагатих"

Аналогічно можна визначити нечітку множину "швидкісні", "середні", "тихохідні" і так далі.

9.4. Функція приналежності

У приведених вище прикладах використані прямі методи, коли експерт або просто задає для будь-якого x ∈ E значення μA(x), або визначає функцію приналежності. Як правило, прямі методи завдання функції приналежності використовуються для вимірних понять, таких як швидкість, година, відстань, тиск, температура і так далі, тобто коли виділяються полярні значення.

У багатьох завданнях при характеристиці об'єкту можна виділити набір ознак і для будь-якого з них визначити полярні значення, що відповідають значенням функції приналежності, 0 або 1.

Наприклад, в завданні розпізнавання особи можна виділити наступні пункти:

Позначення	Опис характеристики	0	1
x_1	висота лоба	низький	широкий
x_2	профіль носа	кирпоносий	горбатий
x_3	довжина носа	короткий	довгий
x_4	розріз очей	вузький	широкий
x_5	колір очей	світлий	темний
x_6	форма підборіддя	гострий	квадратний
x_7	товщина губ	тонкі	товсті
x_8	колір особи	темний	світлий
x_9	овал особи	овальне	квадратне

Для конкретної особи А експерт, виходячи з приведеної шкали, задає $\mu A(x) \in [0, 1]$, формуючи векторну функцію приналежності $\{ \mu A(x1)), \mu A(x2)... \mu A(x9)\}$.

Непрямі методи визначення значень функції приналежності використовуються у випадках, коли немає елементарних вимірних властивостей для визначення нечіткої множини. Як правило, це методи попарних порівнянь. Якби значення функцій приналежності були відомі, наприклад, $\mu A(x_i) = w_i$, $i=1,2,...,n$, тоді попарні порівняння можна представити матрицею відносин $A = \{a_{ij}\}$, де $a_{ij}=w_i / w_j$ (операція ділення).

Існує понад десяток типових форм кривих для завдання функцій приналежності. Найбільшого поширення набули: трикутна, трапецеїдальна функції та функція приналежності Гауса.

Трикутна функція приналежності визначається трійкою чисел *(a,b,c)*, і її значення в точці *x* обчислюється згідно виразу:

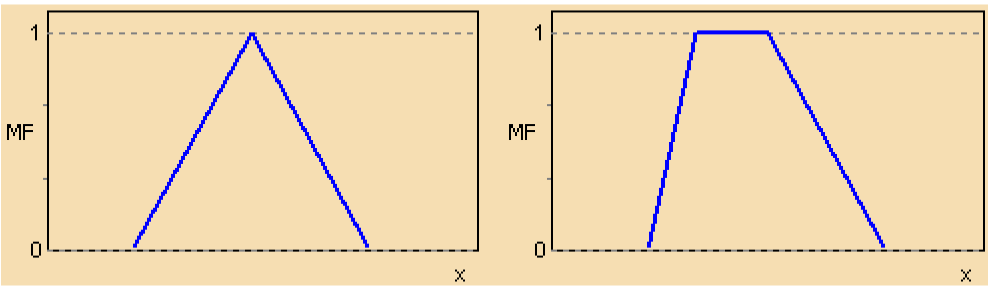
$$MF\ x\ = \begin{cases} 1-\frac{b-x}{b-a}, & a \le x \le b \\ 1-\frac{x-b}{c-b}, & b \le x \le c \\ 0, & \textit{в інших випадках} \end{cases} \tag{9.3}$$

При *(b - a) = (c - b)* маємо випадок симетричної трикутної функції приналежності, яка може бути однозначно задана двома параметрами з трійки *(a,b,c)*.

Аналогічно для завдання трапецеїдальній функції приналежності необхідна четвірка чисел *(a,b,c,d)*:

$$MF\ x\ = \begin{cases} 1-\frac{b-x}{b-a}, & a \le x \le b \\ 1, & b \le x \le c \\ 1-\frac{x-c}{d-c}, & c \le x \le d \\ 0, & \textit{в інших випадках} \end{cases}, \tag{9.4}$$

При *(b - a) = (d - c)* трапецеїдальна функція приналежності приймає симетричний вигляд.



a *b*

Рисунок 9.5 – Типові кусочно-лінійні функції приналежності: трикутна (a) та трапецеїдальна (b).

Функція приналежності гауссова типу описується формулою

$$MF(x) = exp\left(-\left(\frac{x-c}{\sigma}\right)^2\right), \tag{9.5}$$

і оперує двома параметрами. Параметр c позначає центр нечіткої множини, а параметр σ відповідає за крутизну функції.

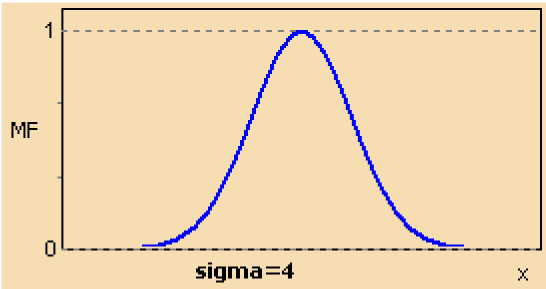


Рисунок 9.6 – Гауссова функція приналежності

Сукупність функцій приналежності для кожного терма з базової терм-множини T зазвичай зображаються разом на одному графіку. На рис. 9.7 приведений приклад описаної вище лінгвістичної змінної «Ціна акції», на рис. 9.8 – формалізація неточного поняття «Вік людини». Так, для людини 48 років ступінь приналежності до множини «Молодий» рівна 0, «Середній» – 0,47, «Вище середнього» – 0,20.

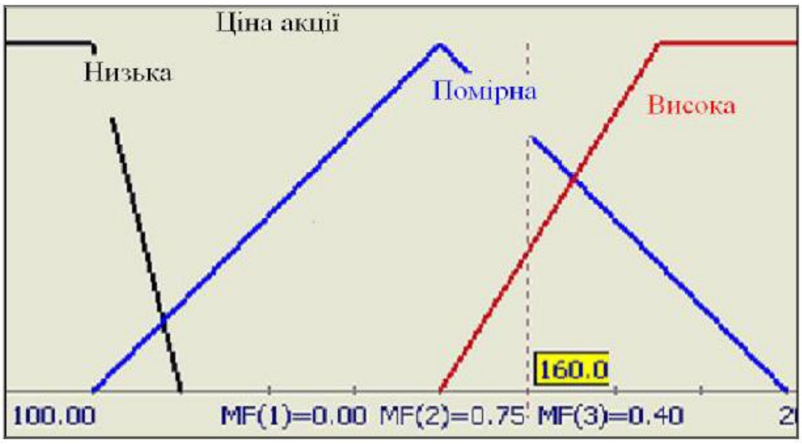


Рисунок 9.7 – Опис лінгвістичної змінної «Ціна акції»

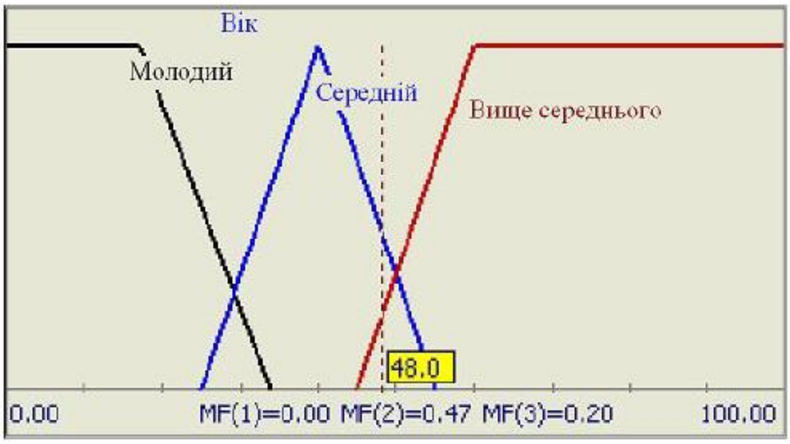


Рисунок 9.8 – Опис лінгвістичної змінної «Вік»

Кількість термів в лінгвістичній змінній рідко перевищує 7.

9.5 Операції над нечіткими числами

Цілий розділ теорії нечітких множин – м'які обчислення (нечітка арифметика) – вводить набір операцій над нечіткими числами. Ці операції вводяться через операції над функціями приналежності на основі так званого сегментного принципу.

Визначимо рівень приналежності α як ординату функції приналежності нечіткого числа. Тоді перетин функції приналежності з нечітким числом дає пару значень, які прийнято називати межами інтервалу достовірності.

Задамося фіксованим рівнем приналежності α і визначимо відповідні йому інтервали достовірності по двох нечітких числах A і B: $[a_1, a_2]$ і $[b_1, b_2]$, відповідно. Тоді основні операції з нечіткими числами зводяться до операцій з їх інтервалами достовірності. А операції з інтервалами, у свою чергу, виражаються через операції з дійсними числами - межами інтервалів:

— операція "складання":

$$[a_1, a_2]\ (+)\ [b_1, b_2] = [a_1 + b_1, a_2 + b_2] \tag{9.6}$$

— операція "віднімання":

$$[a_1, a_2]\ (-)\ [b_1, b_2] = [a_1 - b_2, a_2 - b_1] \tag{9.7}$$

— операція "множення":

$$[a_1, a_2]\ (\times)\ [b_1, b_2] = [a_1 \times b_1, a_2 \times b_2], \tag{9.8}$$

— операція "ділення":

$$[a_1, a_2]\ (/)\ [b_1, b_2] = [a_1 / b_2, a_2 / b_1], \tag{9.9}$$

— операція "піднесення до ступеня":

$$[a_1, a_2]\ (\wedge)\ i = [a_1^i\ , a_2^i]. \tag{9.10}$$

З суті операцій з трапезоїдними числами можна зробити ряд важливих тверджень :

— дійсне число є окремим випадком трикутного нечіткого числа;

— сума трикутних чисел є трикутне число;

— трикутне (трапезоїдне) число, помножене на дійсне число, є трикутне (трапезоїдне) число;

— сума трапезоїдних чисел є трапезоїдне число;

Аналізуючи властивості нелінійних операцій з нечіткими числами (наприклад, ділення), дослідники приходять до висновку, що форма функцій приналежності результуючих нечітких чисел часто близька до трикутної. Це дозволяє апроксимувати результат, приводячи його до трикутного вигляду. І, якщо приводимість в наявності, тоді операції з трикутними числами зводяться до операцій з абсцисами вершин їх функцій приналежності.

Тобто, якщо ми вводимо опис трикутного числа набором абсцис вершин *(a, b, c)*, то можна записати:

$$(a_1, b_1, c_1) + (a_2, b_2, c_2) \equiv (a_1 + a_2, b_1 + b_2, c_1 + c_2) \qquad (9.11)$$

Це – найпоширеніше правило м'яких обчислень.

9.6. Операції над нечіткими множинами

1. Домінування (Вміщення)

Хай A і B - нечіткі множини на універсальній множині E.

Говорять, що A міститься в B, якщо $\forall x \in E\ \mu A(x) < \mu B(x)$. (9.12)

Позначення: $A \subset B$.

Коли використовують термін "домінування", тобто у випадку якщо $A \subset B$, говорять, що B домінує A.

2. Рівність

A і B рівні, якщо $\forall x \in E\ \mu A(x) = \mu B(x)$. (9.13)

Позначення: $A = B$.

3. Доповнення

Хай $\mu = [0, 1]$, A і B - нечіткі множини, задані на E. A і B доповнюють один одного, якщо $\forall x \in E\ \mu A(x) = 1 - \mu B(x)$. (9.14)

Позначення: $B = B$ або $A = A$

Очевидно, що A=A. (Доповнення визначене для $\mu = [0, 1]$, але очевидно, що його можна визначити для будь-якого впорядкованого M).

Доповнення нечіткої множини A позначається символом ¬ A і визначається

$$\neg A = {}_{U} \quad \frac{1 - \mu_A\ u}{} \quad u. \qquad (9.15)$$

Операція доповнення відповідає логічному запереченню.

4. Перетин

Перетин A і B позначається $A \cap B$ і визначається

$$A \cap B = \int_U \mu_A(u) \wedge \mu_B(u) \; / u. \qquad (9.16)$$

Перетин відповідає логічній зв'язці «і». $A \cap B$ – найменша нечітка підмножина, яка міститься одночасно в A і B:

$$\mu A \cap B(x) = \mu in(\mu A(x), \mu B(x)). \qquad (9.17)$$

5. Об'єднання

Об'єднання нечітких множин A і B $(A + B)$

$$A + B = \int_U \mu_A(u) \vee \mu_B(u) \; / u. \qquad (9.18)$$

Об'єднання відповідає логічній зв'язці «або».

$A \cup B$ – найбільша нечітка підмножина, яка включає як A, так і B, з функцією приналежності:

$$\mu A \cup B(x) = \mu ax(\mu A(x), \mu B(x)). \qquad (9.19)$$

6. Різниця

$A - B = A \cap B$ з функцією приналежності:

$$\mu A - B(x) = \mu A \cap B(x) = min(\mu A(x), 1 - \mu B(x)). \qquad (9.20)$$

7. Диз'юнктивна сума

$A \oplus B = (A - B) \cup (B - A) = (A \cap B) \cup A \cap B)$ з функцією приналежності:

$$\mu A - B(x) = max\{[min\{\mu A(x), 1 - \mu B(x)\}]; \\ [min\{1 - \mu A(x), \mu B(x)\}]\} \qquad (9.21)$$

8. Добуток *A* і *B* позначається *AB* і визначається

$$AB = \int_U \mu_A(u)\,\mu_B(u)\,/u \tag{9.22}$$

9. Піднесення до ступеня

$$a > 0,\ A^a = \int_U \left(\mu_A(u)\right)^a / u \tag{9.23}$$

10. Концентрація, частковий випадок піднесення до ступеня:

$$CON\,(A) = A^2. \tag{9.24}$$

11. Розтягування (розмивання):

$$DIL(A) = A^{0.5}. \tag{9.25}$$

Властивості операцій $\cup$ і $\cap$

Хай *A, B, C* - нечіткі множини, тоді виконуються наступні властивості:

— $A \cap B = B \cap A$, $A \cup B = B \cup A$ – комутативність;

— $(A \cap B) \cap C = A \cap (B \cap C)$, $(A \cup B) \cup C = A \cup (B \cup C)$ – асоціативність;

— $A \cap A = A$, $A \cup A = A$ – ідемпотентність;

— $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$, $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ - дистрибутивність;

— $A \cup \emptyset = A$, де $\emptyset$ - порожня множина, тобто $\mu_\emptyset(x) = 0\ \forall x \in E$;

— $A \cap \emptyset = \emptyset$;

— $A \cap E = A$, де *E* - універсальна множина;

— $A \cup E = E$;

— $A \cap B = A \cup B$, $A \cup B = A \cap B$ – теореми де Моргана.

На відміну від чітких множин, для нечітких множин в загальному випадку:

— $A \cap A \neq \emptyset$,

— $A \cup A \neq E$.

Приклади:

Хай*:* *A = 0,4/x1 + 0,2/x2+0/x3+1/x4;*

B = 0,7/x1+0,9/x2+0,1/x3+1/x4;

C = 0,1/x1+1/x2+0,2/x3+0,9/x4.

Тоді, для визначених вище перетворень маємо:

1. $A \subset B$, тобто *A* міститься в *B* або *B* домінує *A*, *C* не зрівняно ні з *A*, ні з *B*, тобто пари *{A, C}* і *{B, C}* – пари не домінуємих нечітких множин.

2. A ≠ B ≠ C.

3. *A*= 0,6/x1 + 0,8/x2 + 1/x3 + 0/x4;

B = 0,3/x1 + 0,1/x2 + 0,9/x3 + 0/x4.

4. A∩B = 0,4/x1 + 0,2/x2 + 0/x3 + 1/x4.

5. A∪C = 0,7/x1 + 0,9/x2 + 0,1/x3 + 1/x4.

6. A - C = A∩ *B* = 0,3/x1 + 0,1/x2 + 0/x3 + 0/x4;

B - A = *A* ∩ C = 0,6/x1 + 0,8/x2 + 0,1/x3 + 0/x4.

7. A ⊕ B = 0,6/x1 + 0,8/x2 + 0,1/x3 + 0/x4.

9.7. Наочне представлення операцій над нечіткими множинами

Для нечітких множин можна застосувати візуальне уявлення. Розглянемо прямокутну систему координат, на осі ординат якої відкладаються значення μA(x), на осі абсцис в довільному порядку розташовані елементи *E*. Якщо *E* за своєю природою впорядковано, то цей порядок бажано зберегти в розташуванні елементів на осі абсцис. Таке уявлення робить наочними прості операції над нечіткими множинами.

Хай *A* нечіткий інтервал між 5 до 8 і *B* нечітке число близько 4, як показано на рис. 9.9.

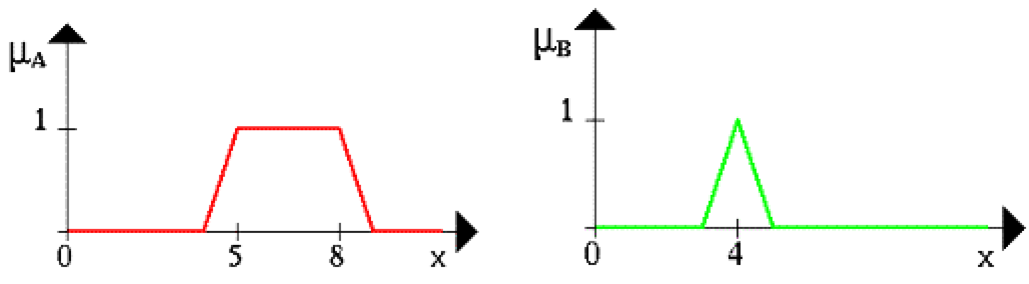
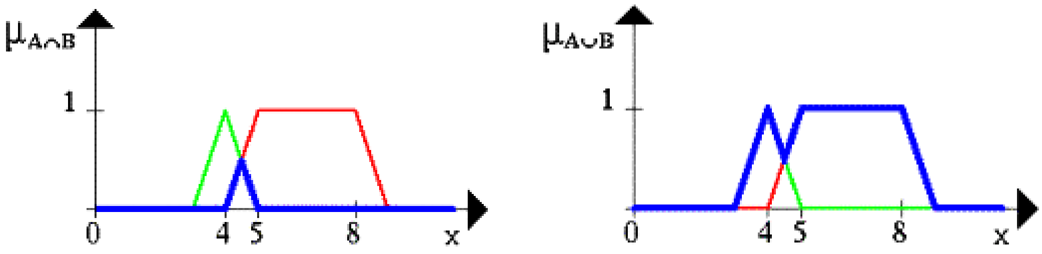


Рисунок 9.9 – Наочне представлення нечіткого інтервалу *A* і нечіткого числа *B*

Проілюструємо нечітку множину між 5 і 8 «І» (AND) близько 4 (синя лінія на рис. 9.10.а). Нечітка множина між 5 і 8 «АБО» (OR) близько 4 показано на наступному рисунку (знову синя лінія на рис. 9.10.б).



а *б*

Рисунок 9.10 – Наочне представлення операцій «І» (а) та «АБО» (б)

Рис.9.11 ілюструє операцію доповнення. Синя лінія - це ДОПОВНЕННЯ нечіткої множини *A*.

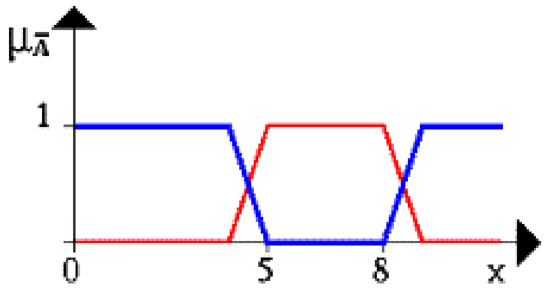


Рисунок 9.11 – Ілюстрація операції доповнення нечіткої множини *A*

На рис. 9.12 заштрихована частина відповідає нечіткій множині *A* і зображає область значень *A* і всіх нечітких множин, що містяться в *A*. Решта рисунків зображає відповідно, A , $A \cap A$, $A \cup A$.

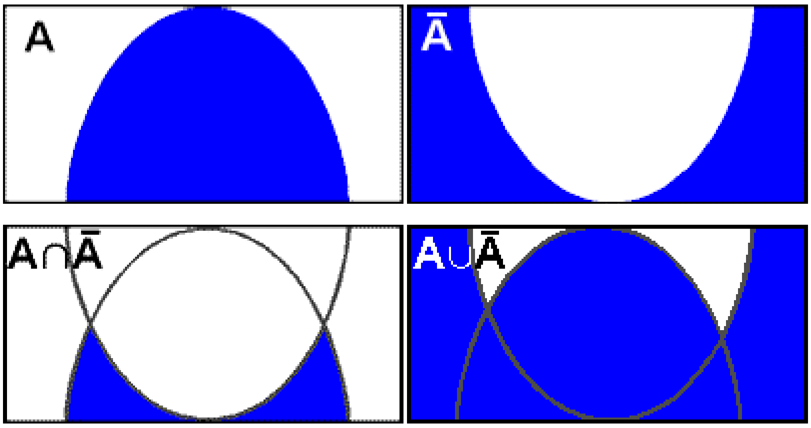


Рисунок 9.12 – Наочне представлення операцій $\bar{A}$, $A\cap \bar{A}$, $A\cup \bar{A}$. з нечіткою множиною A

На рис. 9.13 наведено наочний приклад використання операції концентрації та розмивання для нечіткої множини A.

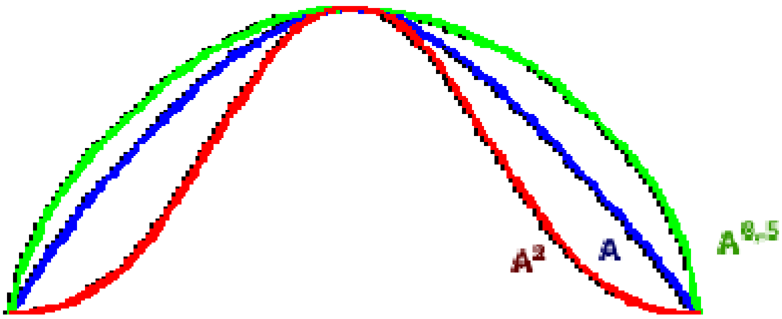


Рисунок 9.13 – Операції концентрації *CON(A)* та розмивання *DIL(A)*

9.8. Переваги та застосування нечітких систем

Коротко перерахуємо переваги fuzzy-систем в порівнянні з іншими:

— можливість оперувати нечіткими вхідними даними: наприклад, значення (динамічні завдання), що безперервно змінюються в часі, значення, які неможливо задати однозначно (результати статистичних опитів, рекламні компанії і так далі);

— можливість нечіткої формалізації критеріїв оцінки і порівняння: операція критеріями "більшість", "можливо", переважно" і т.д.;

— можливість проведення якісних оцінок як вхідних даних, так і вихідних результатів: ви оперуєте не тільки значеннями даних, але і їх мірою достовірності (не плутати з вірогідністю!) і її розподілом;

— можливість проведення швидкого моделювання складних динамічних систем і їх порівняльний аналіз із заданим ступенем точності: оперуючи принципами поведінки системи, описаними fuzzy-методами, ви по-перше, не витрачаєте багато часу на з'ясування точних значень змінних і складання рівнянь, що описують, по-друге, можете оцінити різні варіанти вихідних значень.

Що стосується вітчизняного ринку комерційних систем на основі нечіткої логіки, то його формування почалося в середині 1995 року. Популярними є наступні пакети:

• CubiCalc 2.0 RTC – одна з могутніх комерційних експертних систем на основі нечіткої логіки, що дозволяє створювати власні прикладні експертні системи;

• CubiQuick – дешева "університетська" версія пакету CubiCalc;

• RuleMaker – програма автоматичного витягання нечітких правил зі вхідних даних;

• FuziCalc – електронна таблиця з нечіткими полями, що дозволяє робити швидкі оцінки при неточних даних без накопичення погрішності;

• OWL – пакет, що містить початкові тексти всіх відомих видів нейронних мереж, нечіткій асоціативній пам'яті і так далі.

Основними споживачами нечіткої логіки на ринку СНД є банкіри і фінансисти, а також фахівці в області політичного і економічного аналізу. Вони використовують CubiCalc для створення моделей різних економічних, політичних, біржових ситуацій. Що ж до пакету FuziCalc, то він зайняв своє місце на комп'ютерах великих банкірів і фахівців з надзвичайних ситуацій – тобто тих, для кого важлива швидкість проведення розрахунків в умовах неповноти і неточності вхідної інформації. Проте можна з упевненістю сказати, що епоха розквіту прикладного використання нечіткої логіки на вітчизняному ринку ще попереду.

Сьогодні елементи нечіткої логіки можна знайти в десятках промислових виробів – від систем управління електропоїздами і бойовими вертольотами до пилососів і пральних машин. Без застосування нечіткої логіки немислимі сучасні ситуаційні центри керівників західних країн, де ухвалюються ключові політичні рішення і моделюються різні кризові ситуації. Одним з вражаючих прикладів масштабного застосування нечіткої логіки стало комплексне моделювання системи охорони здоров'я і соціального забезпечення Великобританії (National Health Service – NHS), яке вперше дозволило точно оцінити і оптимізувати витрати на соціальні потреби.

Не обійшли засоби нечіткої логіки і програмні системи, обслуговуючих великий бізнес. Першими, зрозуміло, були фінансисти, завдання яких вимагають щоденного ухвалення правильних рішень в складних умовах непередбаченого ринку. Перший рік використання системи Fuji Bank приніс банку в середньому $770000 на місяць (і це тільки офіційно оголошений прибуток!).

Услід за фінансистами, стурбовані успіхами японців і втратою стратегічної ініціативи, когнітивними нечіткими схемами зацікавилися промислові гіганти США. Motorola, General Electric, Otis Elevator, Pacific Gas & Electric, Ford та інші на початку 90-х почали інвестувати в розробку виробів, що використовикористовують нечітку логіку. Маючи солідну фінансову "підтримку", фірми, що спеціалізуються на нечіткій логіці, дістали можливість адаптувати свої розробки для широкого круга застосувань. "Зброя еліти" вийшла на масовий ринок.

Серед лідерів нового ринку виділяється американська компанія Hyper Logic, заснована в 1987 році Фредом Уоткинсом (Fred Watkins). Спочатку компанія спеціалізувалася на нейронних мережах, проте незабаром цілком концентрувалася на нечіткій логіці. Недавно вийшла на ринок друга версія пакету CubiCalc фірми HyperLogic, яка є одній з щонайпотужніших експертних систем на основі нечіткої логіки. Пакет містить інтерактивну оболонку для розробки нечітких експертних систем і систем управління, а також run-time модуль, що дозволяє оформляти створені користувачем системи у вигляді окремих програм.

Окрім Hyper Logic серед "патріархів" нечіткої логіки можна назвати фірми IntelligenceWare, InfraLogic, Aptronix. Всього ж на світовому ринку представлено більше 100 пакетів, які так чи інакше використовують нечітку логіку.

У трьох десятках СУБД реалізована функція нечіткого пошуку. Власні програми на основі нечіткої логіки анонсували такі гіганти як IBM, Oracle та інші.

На принципах нечіткої логіки створений і один з російських програмних продуктів – відомий пакет "Бізнес-прогноз". Призначення цього пакету – оцінка ризиків і потенційної прибутковості різних бізнес-планів, інвестиційних проектів і просто ідей щодо розвитку бізнесу. "Ведучи" користувача за сценарієм його задуму, програма задає ряд питань, які допускають як точні кількісні відповіді, так і наближені якісні оцінки, – типу "маловірогідно", "ступінь риски високий" і так далі Узагальнивши всю отриману інформацію у вигляді однієї схеми бізнеспроекта, програма не тільки виносить остаточний вердикт про ризиковану проекту і очікуваних прибутків, але і указує критичні крапки і слабкі місця в авторському задумі. Від аналогічних іноземних пакетів "Бізнес-прогноз" відрізняється простотою, дешевизною і, зрозуміло, російськомовним інтерфейсом.

Втім, програма "Бізнес-прогноз" - лише перша ластівка, за якою неминуче з'являться нові розробки учених СНД.

ЛЕКЦІЯ 10 «Нечіткі множини в системах керування»

Анотація

Нечіткі моделі та системи. Поняття системи нечіткого логічного висновку. Основні етапи нечіткого виводу: формування бази правил, фазифікація вхідних змінних, агрегування підумов, активізація під висновків, акумуляція висновків, дефазифікація. Основні алгоритми нечіткого виводу. Приклади використання системи нечіткого виведення в задачах управління

10.1 Нечіткі моделі та системи

Моделі статичних і динамічних систем, побудова, використання та аналіз яких базується на положеннях теорії нечітких множин і нечіткої логіки називають *нечіткими моделями* або *нечіткими системами*.

Метою нечіткого моделювання складних явищ є наближений опис залежності (апроксимація деякої функції)

$$Y = f(X),$$

де Y - вихідна лінгвістична змінна;

X - вектор вхідних лінгвістичних змінних розмірністю n;

f - залежність між X і Y, описувана сукупністю нечітких продукційних правил.

Нечіткі моделі представляють узагальнення інтервально-оцінюваних моделей, які, в свою чергу, є узагальненням чітких моделей.

В основі нечітких продукційних моделей лежать сукупність нечітких правил «ЯКЩО-ТО», що описують залежності між нечіткими змінними предметної області, композиційне правило виведення і спосіб обчислення значень нечітких змінних (спосіб нечіткого виводу).

Модель опису поведінки систем на природній (або близькій до природної) мові у вигляді наближених міркувань в теорії нечітких множин і нечіткої логіки, заснована на композиційному правилі виводу, називається *системою нечіткого логічного висновку*.

У систему нечіткого логічного висновку входять наступні об'єкти (рис. 10.1):

1) сукупність нечітких продукційних правил (база правил);

2) набір функцій належностей бази нечітких змінних (база змінних);

3) блок фазифікації;

4) блок дефазифікації;

5) блок виводу.

База правил зберігає безліч логічних правил виводу, а також їх порядок (ієрархічну структуру) застосування. База нечітких змінних містить назви лінгвістичних термів та параметри їх функцій приналежності. База правил разом з базою нечітких змінних утворюють базу знань (БЗ) системи нечіткого виводу.

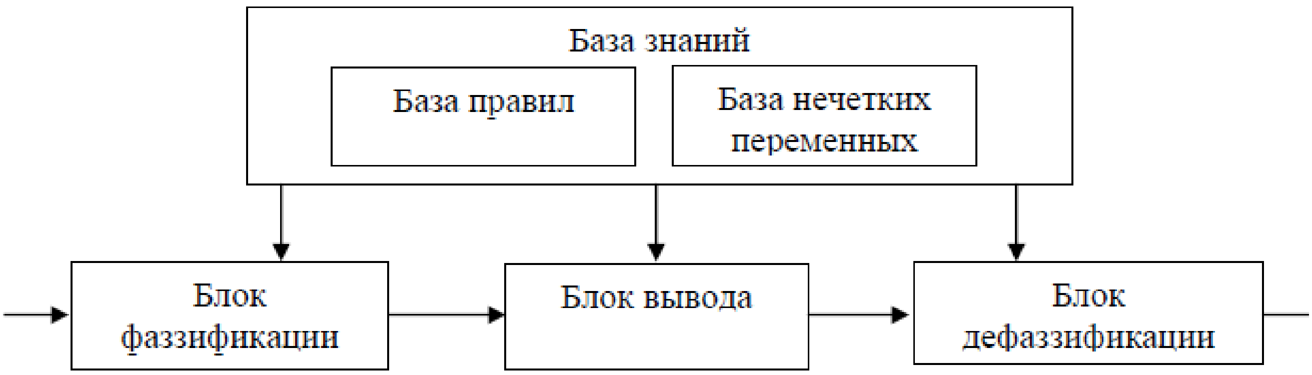


Рисунок 10.1 – Система нечіткого виводу

Найпростіші системи нечіткого логічного висновку засновані на правилах виду:

Ri: Якщо $X \in A_i$ *і* $Y \in B_i$, *то* $Z \in C_i$,

Ri: Якщо $X \in A_i$ *і* $Y \in B_i$, *то* $z = f_i(x, y)$,

де X, Y - вхідні нечіткі змінні;

Z - вихідна нечітка змінна;

A_i, B_i, - вхідні значення (функції належності);

C_i - вихідні нечіткі значення (функції належності);

f_i - деякі речові функції.

При цьому повинні дотримуватися наступні умови:

1) Існує хоча б одне правило для кожного лінгвістичного терма вихідної змінної.

2) Для будь-якого терма вхідної змінної є хоча б одне правило, в якому цей терм використовується як передумови (ліва частина правила).

В іншому випадку має місце неповна база нечітких правил.

10.2 Основні етапи нечіткого виводу

Говорячи про нечітку логіку, найчастіше мають на увазі системи нечіткого виводу, які широко використовуються для управління технічними пристроями та процесами. Розробка і застосування систем нечіткого виведення включають в себе ряд етапів, реалізація яких виконується за допомогою розглянутих раніше основних положень нечіткої логіки.

Інформацією, яка надходить на вхід системи нечіткого виводу, є виміряні деяким чином вхідні змінні. Ці змінні відповідають реальним змінним процесу управління. Інформація, яка формується на виході системи нечіткого висновку, відповідає вихідним змінним, якими є керуючі змінні процесу управління.

Системи нечіткого виводу призначені для перетворення значень вхідних змінних процесу управління у вихідні змінні на основі використання нечітких правил продукцій. Для цього системи нечіткого виводу повинні містити базу правил нечітких продукцій і реалізовувати нечіткий висновок на основі посилок або умов, представлених у формі нечітких лінгвістичних висловлювань.

Таким чином, основними етапами нечіткого виводу є (рис. 10.2).

1) Формування бази правил систем нечіткого виводу.

2) Фазифікації вхідних змінних.

3) Агрегація підумови в нечітких правилах продукцій.

4) Активізація або композиція підзаключень в нечітких правилах продукцій.

5) Акумулювання висновків нечітких правил продукцій.

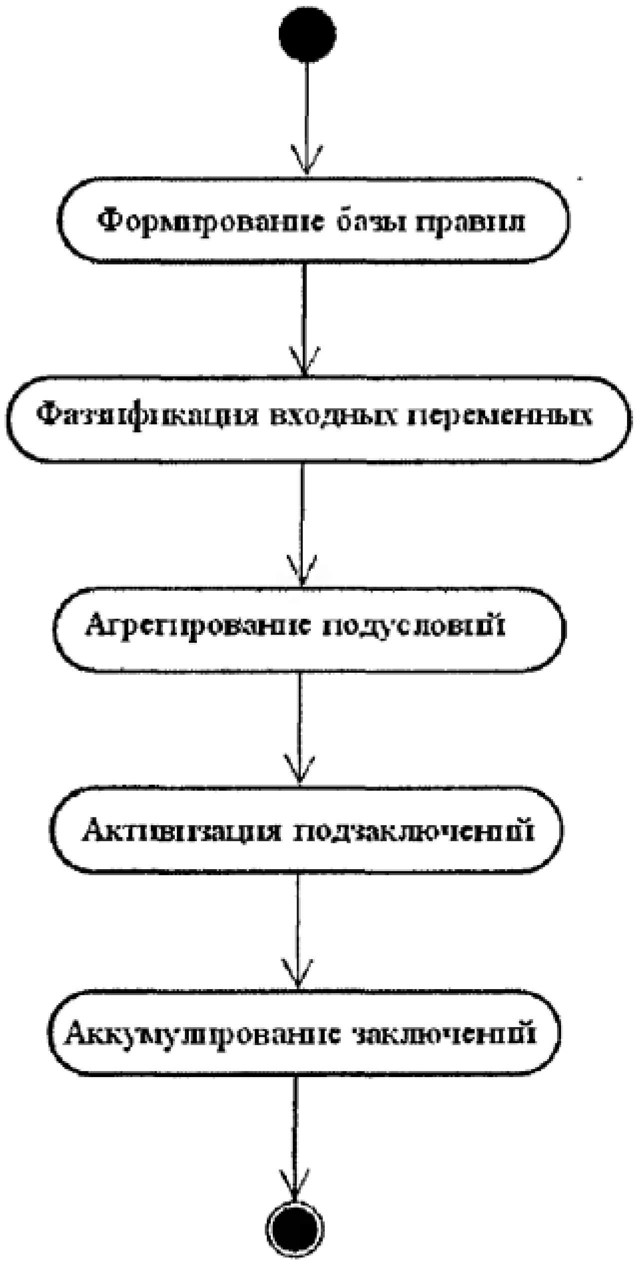


Рисунок 10.2 – Діаграма діяльності процесу нечіткого виведення

Нижче розглядаються основні особливості кожного з цих етапів і наводяться прості приклади їх виконання.

10.2.1 Формування бази правил систем нечіткого виведення

База правил систем нечіткого виведення призначена для формального подання емпіричних знань або знань експертів в тій чи іншій проблемній області. У системах нечіткого виведення використовуються правила нечітких продукцій, в яких умови і висновки сформульовані в термінах нечітких лінгвістичних висловлювань розглянутих вище видів. Сукупність таких правил будемо далі називати базами правил нечітких продукції.

Визначення 10.1. База правил нечітких продукції являє собою кінцеву множину правил нечітких продукцій, узгоджених щодо використовуваних в них лиігвістичних змінних. Найчастіше база правил представляється у формі структурованого тексту:

ПРАВІЛО_1: ЯКЩО "*Умова_1*" ТО "*Висновок_1*" (F_1)

ПРАВІЛО_2: ЯКЩО "*Умова_2*" ТО "*Висновок_2*" (F_2)

... (10.1)

ПРАВІЛО_n: ЯКЩО "*Умова_n*" ТО "*Висновок_n*" (F_n)

Тут через F_i $(i = \{1, 2 ,..., n\})$ позначені коефіцієнти визначеності або вагові коефіцієнти відповідних правил. Ці коефіцієнти можуть приймати значення з інтервалу [0, 1]. У разі якщо ці вагові коефіцієнти відсутні, зручно прийняти, що їх значення рівні 1.

Узгодженість правил щодо використовуваних лінгвістичних змінних означає, що у якості умов та висновків правил можуть використовуватися тільки нечіткі лінгвістичні висловлювання, при цьому в кожному з нечітких висловлювань повинні бути визначені функції приналежності значень терм-множини для кожної з лінгвістичних змінних.

Визначення 10.2. У системах нечіткого виведення лінгвістичні змінні, які використовуються в нечітких висловлюваннях підумови правил нечітких продукцій, часто викликані вхідними лінгвістичними змінними, а змінні, які використовуються в нечітких висловлюваннях підвисновків правил нечітких продукцій, часто називають вхідних лінгвістичними змінними.

Таким чином, при завданні або формуванні бази правил нечітких продукцій необхідно визначити: множину правил нечітких продукцій: $P = \{R_1, R_2, ..., R_n\}$ у формі (10.1), множину вхідних лінгвістичних змінних: $V = \{\beta_1, \beta_2, ..., \beta_m\}$ і множину вихідних лінгвістичних змінних: $W = \{w_1, w_2, ..., w_s\}$. Тим самим база правил нечітких продукцій вважається заданою, якщо задано множини P, V, W.

Нагадаємо, що вхідні $\beta_i \in V$ або вихідна $w_j \in W$ лінгвістична змінна вважається заданою або опреділеною, якщо для неї визначено базову терм-множину з відповідними функціями належності кожного терма, а також дві процедури G і M. Найбільш поширеним випадком є використання в якості функцій належності трикутних або трапецієподібних функцій приналежності.

10.2.2 Фазифікації (Fuzzification)

У контексті нечіткої логіки під фазифікації розуміється не тільки окремий етап виконання нечіткого виведення, але і власне процес або процедура знаходження значень функцій належності нечітких множин (термів) на основі звичайних (не нечітких) вхідних даних. Фазифікації ще називають введенням нечіткості. Метою етапу фазифікації є встановлення відповідності між конкретним (зазвичай - чисельним) значенням окремої вхідної змінної системи нечіткого вивода і значенням функції приналежності відповідного їй терма вхідної лінгвістичної змінної. Після завершення цього етапу для всіх вхідних змінних повинні бути визначені конкретні значення функцій приналежності по кожному з лінгвістичних термів, які використовуються в підумові бази правил системи нечіткого виводу.

Формально процедура фазифікації виконується наступним чином. До початку цього етапу передбачаються відомими конкретні значення всіх вхідних змінних системи нечіткого виведення, тобто множина значень $V' = \{a_1, a_2, \ldots, a_m\}$. У загальному випадку кожне $a_i \in X_i$, де X_i - універсум лінгвістичної змінної β_i.

Далі розглядається кожна з підумови виду «β_i ϵ α'» правил системи нечіткого висновку, де α' - деякий терм з відомою функцією приналежності. При цьому значення a_i використовується як аргумент $\mu(x)$, тим самим знаходиться кількісне значення $b' = \mu(a_i)$. Це значення і є результатом фазифікації підумови «β_i ϵ α'». Етап фазифікації вважається закінченим, коли будуть знайдені всі значення для кожної з підумови всіх правил, що входять у розглянуту базу правил системи нечіткого виводу. Цю множину значень позначимо через $B = \{b_i'\}$. При цьому якщо деякий терм α'' лінгвістичної змінної β_i не присутній ні в одному з нечітких висловлювань, то відповідне йому значення функції приналежності не знаходиться в процесі фазифікації.

Приклад 10.1. Для ілюстрації виконання цього етапу розглянемо приклад процесу фазифікації трьох нечітких висловлювань: "*швидкість автомобіля мала*", "*швидкість автомобілів середня*", "*швидкість автомобіля висока*" для вхідної лінгвістичної змінної β_1 - швидкість руху автомобіля. Їм відповідають нечіткі висловлювання першого виду: "β_1 є α_1", "β_1 є α_2", "β_1 є α_3". Припустимо, що поточна швидкість автомобіля дорівнює 55 км/год., тобто a_1 = 55 км/год.

Тоді фазифікація першого нечіткого висловлювання дає в результаті число 0, яке означає його ступінь істинності і виходить підстановкою значення a_1 = 55 км/год. в якості аргументу функції приналежності терма α_1 (рис. 10.3, а). Фазифікація другого нечіткого висловлювання дає в результаті число 0.67 (наближене значення), яке означає його ступінь істинності і виходить підстановкою значення a_1 = 55 км/год. в якості аргументу функції приналежності терма α_2 (рис. 10.3, б). Фазифікації третього нечіткого висловлювання дає в результаті число 0, що означає його ступінь істинності і виходить підстановкою значення a_1 = 55 км/год. в якості аргументу функції приналежності терма α_3 (рис. 10.3, в).

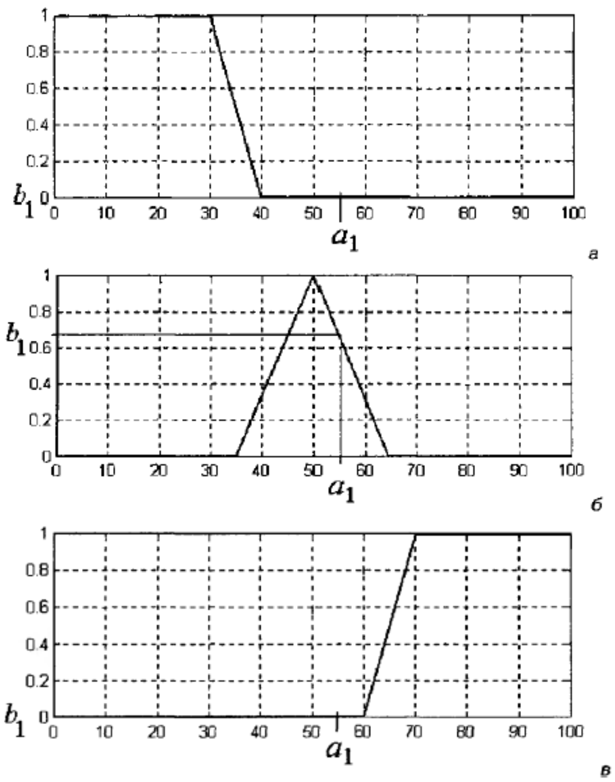


Рисунок 10.3 – Приклад фазифікації вхідної лінгвістичної змінної "*швидкість автомобіля*" для трьох нечітких висловлювань

10.2.3 Агрегація (Aggregation)

Агрегація являє собою процедуру визначення ступеня істинності умов по кожному з правил системи нечіткого виводу.

Формально процедура агрегування виконується наступним чином. До початку цього етапу передбачаються відомими значення істинності всіх підумов системи нечіткого висновку, тобто множина значень $B = \{b_i'\}$. Далі розглядається кожне з умов правил системи нечіткого виводу. Якщо умова правила є нечітке висловлювання, то ступінь його істинності дорівнює відповідному значенню bi'.

Якщо ж умова складається з декількох підумов, причому лінгвістичні змінні в підумові попарно не дорівнюють один одному, то визначається ступінь істинності складного висловлювання на основі відомих значень істинності підумови. При цьому для визначення результату нечіткої кон'юнкції або зв'язки "І" може бути використано одну з формул

$$T(\tilde{A} \wedge \tilde{B}) = \min\{T(\tilde{A}), T(\tilde{B})\}$$

$$T(\tilde{A} \wedge \tilde{B}) = T(\tilde{A}) \cdot T(\tilde{B}).$$

$$T(\tilde{A} \wedge \tilde{B}) = \max\{T(\tilde{A})+T(\tilde{B}) - 1, 0\},$$

а для визначення результату нечіткої диз'юнкції або зв'язки "АБО" може бути використана одна з формул

$$T(\tilde{A} \vee \tilde{B}) = \max\{T(\tilde{A}), T(\tilde{B})\}$$

$$T(\tilde{A} \vee \tilde{B}) = T(\tilde{A}) + T(\tilde{B}) - T(\tilde{A}) \cdot T(\tilde{B})$$

$$T(\tilde{A} \vee \tilde{B}) = \min\{T(\tilde{A})+T(\tilde{B}), 1\}.$$

При цьому значення b_i' використовуються в якості аргументів відповідних логічних операцій. Тим самим знаходяться кількісні значення істинності всіх умов правил системи нечіткого виводу. Етап агрегування вважається закінченим, коли будуть знайдені всі значення b_i'' для кожного з правил R_k, що входять у розглянуту базу правил P системи нечіткого виводу. Цю множину значень позначимо через $B''= \{b_1'', b_2'', ..., b_n''\}$

Приклад 10.2. Для ілюстрації виконання цього етапу розглянемо приклад процесу агрегування двох нечітких висловлювань: "*швидкість автомобіля середня*" І "*кава гаряча*" і "*швидкість автомобіля середня*" АБО "*кава гаряча*" для вхідної лінгвістичної змінної β_1 - швидкість руху автомобіля і β_2 - температура кави. Припустимо, що поточна швидкість автомобіля дорівнює 55 км/год., тобто a_1 = 55 км/год., а температура кави дорівнює a_2 = 70°С.

Тоді агрегування першого нечіткого висловлювання з використанням операції нечіткої кон'юнкції $T(\tilde{A} \wedge \tilde{B}) = \min\{T(\tilde{A}), T(\tilde{B})\}$ дає в результаті число b_1''= 0.67 (наближене значення), яке означає його ступінь істинності і виходить як мінімальне зі значень 0.67 і 0.8 (рис. 10.4, а). Агрегація другий нечіткого висловлювання з використанням операції нечіткої диз'юнкції (10.1) дає в результаті число b_2''= 0.8, що означає його ступінь істинності і виходить як максимальне з значень 0.67 і 0.8 (рис. 10.4, б).

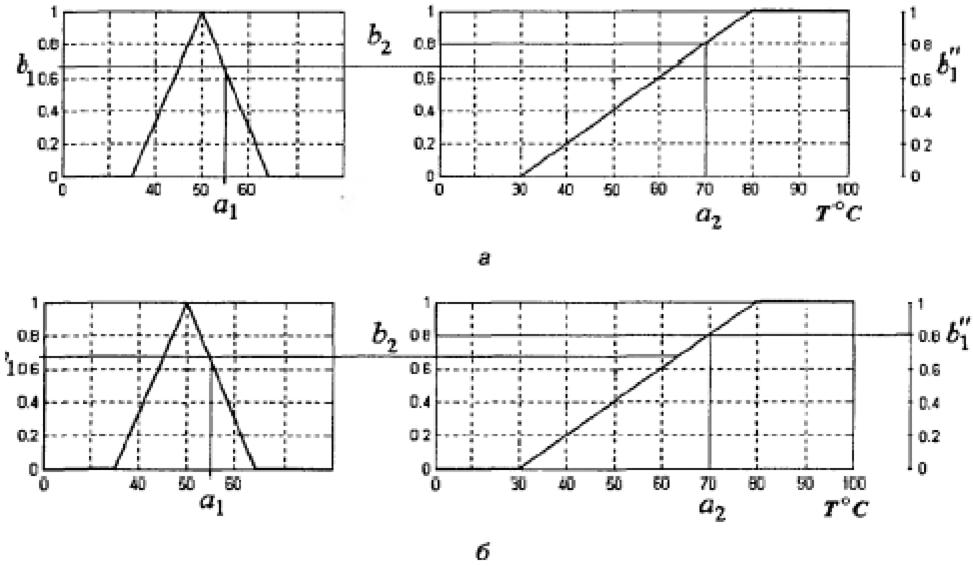


Рисунок 10.4 – Приклади агрегування підумови для двох нечітких висловлювань "*швидкість автомобіля середня*" І "*температура кави висока*" (а) і "*швидкість автомобіля середня*" АБО "*температура кави висока*" (б)

10.2.4 Активізація (Activation)

Активізація в системах нечіткого виведення являє собою процедуру або процес знаходження ступеня істинності кожного з підвисновків правил нечітких продукцій. Активізація в загальному випадку багато в чому аналогічна композиції нечітких відносин, але не тотожна їй. Оскільки в системах нечіткого виведення використовуються лінгвістичні змінні, то формули для нечіткої композиції втрачають своє значення. В дійсності при формуванні бази правил системи нечіткого виводу задаються вагові коефіцієнти F_i для кожного правила (за умовчанням передбачається, якщо ваговий коефіцієнт не заданий явно, то його значення дорівнює 1).

Формально процедура активізації виконується наступним чином. До початку цього етапу передбачаються відомими значення істинності всіх умов системи нечіткого виведення, тобто множина значень $B''= \{b_i''\}$ і значення вагових коефіцієнтів F_i для кожного правила. Далі розглядається кожне з висновків правил системи нечіткого вивода. Якщо висновок правила є нечітке висловлювання, то ступінь його істинності дорівнює алгебраїчному добутку відповідного значення b_1'' на ваговий коефіцієнт F_i. Якщо ж висновок складається з декількох підвисновків виду

ПРАВИЛО <#>: ЕСЛИ "β_1 *есть* α''" ТО "β_2 *есть* α'''" И "β_3 *есть* ν" ,

ПРАВИЛО <#>: ЕСЛИ "β_1 *есть* α''" ТО "β_2 *есть* α'''" ИЛИ "β_3 *есть* ν"

причому лінгвістичні змінні в підвисновках попарно не дорівнюють один одному, то ступінь істинності кожного з підвисновків дорівнює алгебраїчному добутку відповідного значення b_i'' на ваговий коефіцієнт F_i. Таким чином, знаходяться всі значення $c_к$ ступенів істинності підвисновків для кожного з правил R_k, що входять у розглянуту базу правил P системи нечіткого виводу. Цю множину значень позначимо через $C = \{c_1, c_2, ..., c_q\}$, де q-загальна кількість підвисновків в базі правил.

Після знаходження множини $C = \{c_1, c_2, ..., c_q\}$ визначаються функції належності кожного з підвисновків для розглянутих вихідних лінгвістичних змінних. Для цієї мети можна використовувати один з методів, які є модифікацією того чи іншого методу нечіткої композиції:

min-активізація: $\mu'(y) = \min\{c_i, \mu(y)\}$ (10.2)

prod-активізація: $\mu'(y) = c_i * \mu(y)$ (10.3)

average-активізація: $\mu'(y) = 0.5 * (c_i + \mu(y))$ (10.4)

де $\mu(x)$ - функція приналежності терма, який є значенням деякої вихідний змінної w_j, заданої на універсумі Y.

Етап активізації вважається закінченим, коли для кожної з вихідних лінгвістичних змінних, що входять в окремі підвисновки правил нечітких продукцій, будуть визначені функції приналежності нечітких множин їх значень, тобто сукупність нечітких множин: $C_1, C_2, .. ., C_q$, де q-загальна кількість підвисновків в базі правил системи нечіткого виводу.

Приклад 10.3. Для ілюстрації виконання цього етапу розглянемо приклад процесу активізації укладення в наступному правилі нечіткої продукції (це правило навряд чи має цільове застосування і використовується формальним чином):

ЯКЩО "*швидкість автомобіля середня*" ТО "*кава гаряча*"

Вхідною лінгвістичною змінною в цьому правилі є β_1 - швидкість руху автомобіля, а вихідний змінної є β_2 - температура кави. Припустимо, що поточна швидкість автомобіля дорівнює 55 км/год., тобто a_1 = 55 км/год.

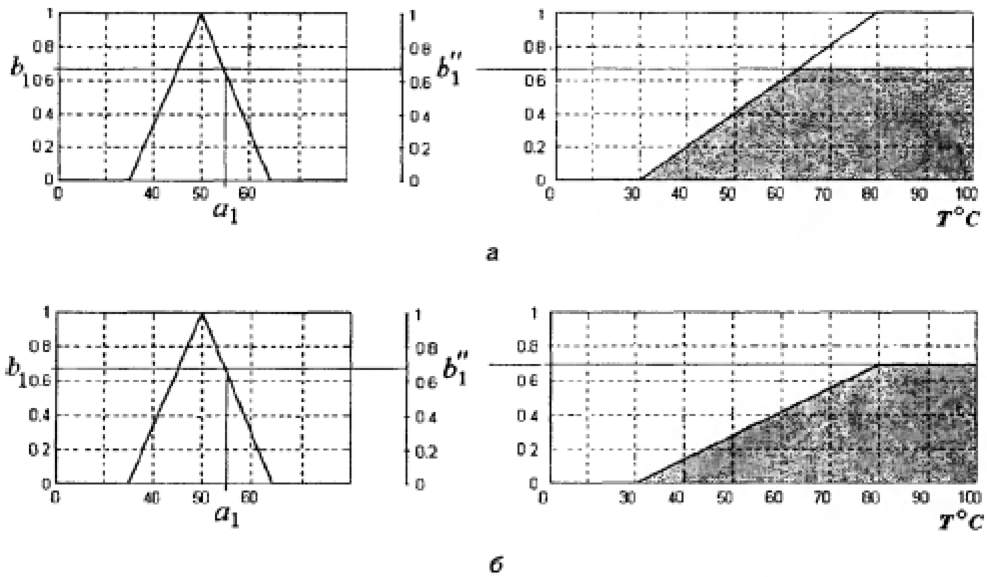


Рисунок 10.5 – Приклад активізації ув'язнення для правила нечіткої продукції

Оскільки агрегування умови цього правила дає в результаті b_1''= 0.67, а ваговий коефіцієнт дорівнює 1 (за замовчуванням), то значення 0.67 буде використовуватися в якості c_1 для отримання результату активізації. Результат, отриманий методом min-активізації (10.2), зображений на рис. 10.5, а більш темним кольором, а результат, отриманий методом prod-активізації (10.3), зображений на рис. 10.5, б більш темним кольором. Слід пам'ятати, що в цьому прикладі на відміну від попереднього "*температура кави*" - вихідна лінгвістична змінна.

10.2.5 Акумуляція (Accumulation)

Акумуляція або акумулювання в системах нечіткого виведення являє собою процедуру або процес знаходження функції приналежності для кожної з вихідних лінгвістичних змінних безлічі $W = \{w_1, w_2, ..., w_s\}$.

Мета акумуляції полягає в тому, щоб об'єднати або акумулювати всі степені істинності висновків (підвисновків) для отримання функції приналежності кожної з вихідних змінних. Причина необхідності виконання цього етапу полягає в тому, що підвисновки пов'язані з однією і тією ж вихідною лінгвістичною змінною, належать різним правилам системи нечіткого виводу.

Формально процедура акумуляції виконується наступним чином. До початку цього етапу передбачаються відомими значення істинності всіх підвисновків для кожного з правил R_k, що входять у розглянуту базу правил P системи нечіткого виводу, у формі сукупності нечітких множин: $C_1, C_2, ..., C_q$, де q - загальна кількість підвисновків в базі правил. Далі послідовно розглядається кожна з вихідних лінгвістичних змінних $w_j \in W$ і пов'язані з нею нечіткі множини: C_{j1}, C_{j2} ,..., C_{jq}. Результат акумуляції для вихідної лінгвістичної змінної w_j визначається як об'єднання нечітких множин C_{j1}, C_{j2} ,..., C_{jq} за однією з формул.

Етап акумуляції вважається закінченим, коли для кожної з вихідних лінгвістичних змінних будуть визначені функції приналежності нечітких множин їх значень, тобто сукупність нечітких множин: C_1', C_2', ..., C_s' де s - загальна кількість вихідних лінгвістичних змінних в базі правил системи нечіткого вивода.

Приклад 10.4. Для ілюстрації виконання цього етапу розглянемо приклад процесу акумуляції висновків для трьох нечітких множин C_{11}, C_{12}, C_{13}, отриманих в результаті виконання процедури активізації для вихідної лінгвістичної змінної "*швидкість руху автомобіля*" в деякій системі нечіткого виводу. Припустимо, що функції приналежності цих нечітких множин зображені на рис. 10.6, а, б, в відповідно.

Акумуляція цих функцій приналежності методом max-об'єднання нечітких множин C_{11}, C_{12}, C_{13} за формулою $\mu_{\bar{D}}(x) = \max(\mu_{\bar{A}}(x), \mu_{\bar{B}}(x)) \quad (\forall \boldsymbol{x} \in \boldsymbol{X})$. дозволяє отримати в результаті функцію приналежності вихідної лінгвістичної змінної "*швидкість руху автомобіля*", яка представлена на рис. 10.6, г. Ця функція належності відповідає нечіткій множині C_1', прийнявши, що розглянута вихідна лінгвістична змінна є w_1.

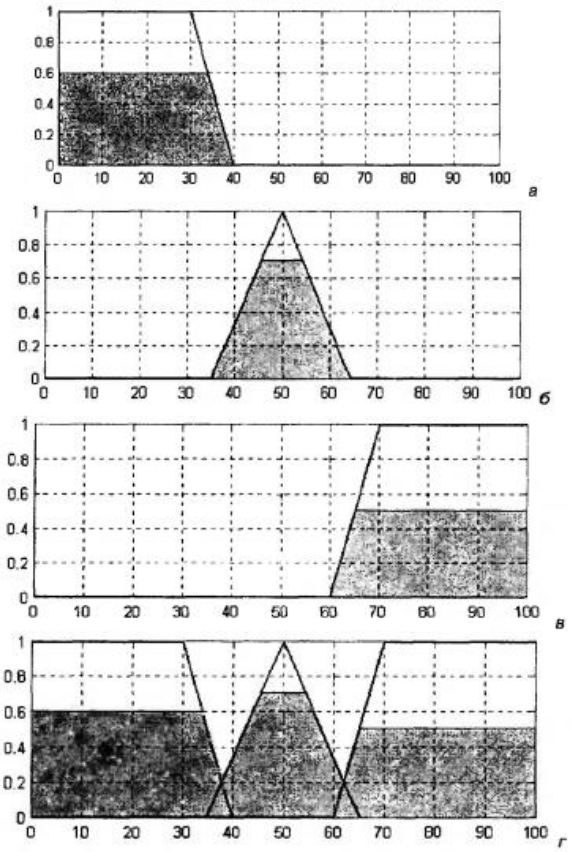


Рисунок 10.6 – Приклад акумуляції висновку для вихідної лінгвістичної змінної "*швидкість руху автомобіля*"

10.2.6 Дефазифікація (Defuzzification)

Дефазифікація в системах нечіткого виведення являє собою процедуру або процес знаходження звичайного (не нечіткого) значення для кожної з вихідних лінгвістичних змінних множини $W = \{w_1, w_2, ..., w_s\}$.

Мета дефазифікації полягає в тому, щоб, використовуючи результати акумуляції всіх вихідних лінгвістичних змінних, отримати звичайне кількісне значення (Crisp value) кожної з вихідних змінних, яке може бути використане спеціальними пристроями, зовнішніми по відношенню до системи нечіткого виводу.

Тому дефазифікацію називають також приведенням до чіткості.

Формально процедура дефазифікації виконується наступним чином. До початку цього етапу передбачаються відомими функції приналежності всіх вихідних лінгвістичних змінних у формі нечітких множин: C_1', C_2',..., C_s', де s - загальна кількість вихідних лінгвістичних змінних в базі правил системи нечіткого виводу. Далі послідовно розглядається кожна з вихідних лінгвістичних змінних $w_j \in W$ до якої відноситься нечітка множина C_j'. Результат дефазифікації для вихідний лінгвістічноъ змінної w_j, визначається у вигляді кількісного значення $y_i \in R$, отриманого за однією з розглянутих нижче формул.

Етап дефазифікації вважається закінченим, коли для кожної з вихідних лінгвістичних змінних будуть визначені кількісні значення у формі деякого дійсного числа, тобто у вигляді y_1, y_2 ,..., y_s, де s - загальна кількість вихідних лінгвістичних змінних в базі правил системи нечіткого виводу.

Для виконання чисельних розрахунків на етапі дефазифікації можуть бути використані такі формули, що отримали назву *методів дефазифікації*.

10.2.6.1 Метод центру ваги (тяжіння)

Центр ваги (CoG, COG, Centre of Gravity) або центроїд площі розраховується за формулою:

$$y = \frac{\int_{Min}^{Max} x * \mu(x)dx}{\int_{Min}^{Max} \mu(x)dx}, \tag{10.5}$$

де y – результат дефазифікації;

x - змінна, відповідна вихідний лінгвістичній змінній w;

$\mu(x)$ – функція приналежності нечіткої множини, відповідної вихідної змінної w після етапу акумуляції;

Min і Max - ліва і права точки інтервалу носія нечіткої множини аналізованої вихідної змінної w.

При дефазифікації методом центру тяжіння звичайне (не нечітке) значення вихідної змінної дорівнює абсцисі центру ваги площі, обмеженою графіком кривої функції належності відповідної вихідної змінної.

Приклад дефазифікації методом центру тяжіння функції належності вихідної лінгвістичної змінної "*швидкість руху автомобіля*" зображений на рис. 10.7. Тут y_1 = 40 км/год. (наближене значення).

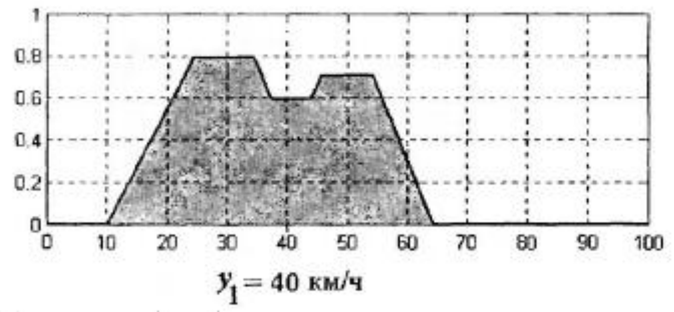


Рисунок 10.7 – Приклад дефазифікації вихідної лінгвістичної змінної "*швидкість руху автомобіля*" методом центру тяжіння

10.2.6.2 Метод центру ваги для одноточкових множин

Центр тяжкості (COGS, Centre of Gravity for Singletons) для одноточкових множин розраховується за формулою:

$$y = \frac{\sum_{i=1}^{n} x_i * \mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)} \qquad (10.6)$$

де n – число одноточкових (одноелементних) нечітких множин, кожне з яких характеризує єдине значення розглянутої вихідний лінгвістичної змінної.

Приклад дефазифікації методом центру тяжіння для одноточкових множин функції належності вихідної лінгвістичної змінної "*швидкість руху автомобіля*" зображений на рис.10.8. У цьому випадку y1 = 41 км /год. (наближене значення).

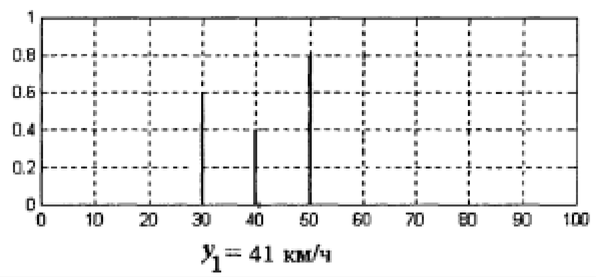


Рисунок 10.8 – Приклад дефазифікації вихідної лінгвістичної змінної "*швидкість руху автомобіля*" методом центру тяжіння для одно точкових множин

10.2.6.3 Метод центру площі

Центр площі (CoA, COA, Centre of Area, Bisector of Area) дорівнює $y = u$, де значення u визначається з рівняння:

$$\int_{Min}^{u} \mu(x)dx = \int_{u}^{Max} \mu(x)dx, \tag{10.7}$$

Іншими словами, центр площі дорівнює абсцисі, яка ділить площу, обмежену графіком кривої функції належності відповідної вихідної змінної, на дві рівні частини. Іноді центр площі називають бісектрисою площі. Цей метод не може бути використаний у разі одноточкових множин.

Приклад дефазифікації методом центру площі функції належності вихідної лінгвістичної змінної "*швидкість руху автомобіля*" зображений на рис. 10.9. Тут y_1 = 35 км/год. (наближене значення).

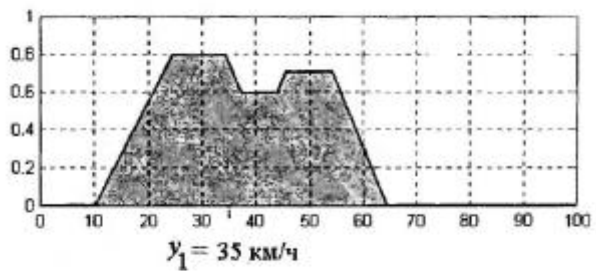


Рисунок 10.9 – Приклад дефазифікації вихідної лінгвістичної змінної "*швидкість руху автомобіля*" методом центру площі

10.2.6.4 Метод лівого модального значення

Ліве модальне значення (LM, Left Most Maximum) розраховується за формулою:

$$y = min\{x_m\}, \tag{10.8}$$

де x_m – модальне значення (мода) нечіткої множини, відповідної вихідної змінної *w* після акумуляції.

Іншими словами, значення вихідної змінної визначається як мода нечіткої множини для відповідної вихідної змінної або найменша з мод (сама ліва), якщо нечітка множина має кілька модальних значень.

Приклад дефазифікації методом лівого модального значення функції приналежності вихідної лінгвістичної змінної "*швидкість руху автомобіля*" зображений на рис. 10.10. У цьому випадку y1 = 24 км/год. (наближене значення).

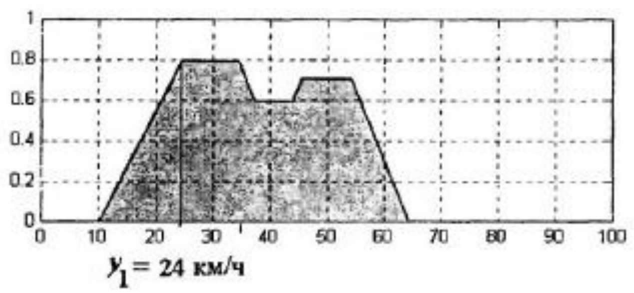


Рисунок 10.10 – Приклад дефазифікації вихідної лінгвістичної змінної "*швидкість руху автомобіля*" методом лівого модального значення

10.2.6.5 Метод правого модального значення

Праве модальне значення (RM, Right Most Maximum) розраховується за формулою:

$$y = max\{x_m\}, \tag{10.9}$$

де x_m – модальне значення (мода) нечіткої множини для вихідної змінної після акумуляції.

У цьому випадку значення вихідної змінної також визначається як мода нечіткої множини для відповідної вихідної змінної або найбільша з мод (сама права), якщо нечітка множина має кілька модальних значень.

Приклад дефазифікації функції належності вихідної лінгвістичної змінної "*швидкість руху автомобіля*" зображений на рис. 10.11. У цьому випадку y1 = 54 км /год. (наближене значення) методом правого модального значення.

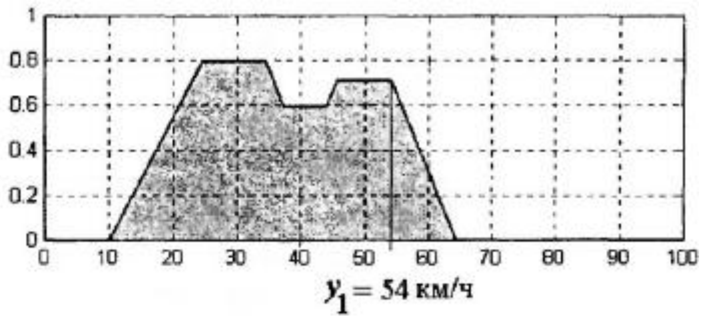


Рисунок 10.11 – Приклад дефазифікації вихідний лінгвістичної змінної "*швидкість руху автомобіля*" методом правого модального значення

10.3 Основні алгоритми нечіткого виводу

Розглянуті вище етапи нечіткого виводу можуть бути реалізовані неоднозначним чином, оскільки включають в себе окремі параметри, які повинні бути фіксовані або специфіковані. Тим самим вибір конкретних варіантів параметрів кожного з етапів визначає певний алгоритм, який в повному обсязі реалізує нечіткий висновок в системах правил нечітких продукцій. До теперішнього часу запропоновано кілька алгоритмів нечіткого висновку. Ті з них, які отримали найбільше застосування в системах нечіткого висновку, розглядаються нижче.

10.3.1 Алгоритм Мамдані (Mamdani)

Алгоритм Мамдані є одним з перших, який знайшов застосування в системах нечіткого виводу. Він був запропонований в 1975 р. англійським математиком Е. Мамдані (Ebra-him Mamdani) як методу для управління паровим двигуном. За своєю суттю цей алгоритм породжує розглянуті вище етапи, оскільки найбільшою мірою відповідає їх параметрами.

Формально алгоритм Мамдані може бути визначений таким чином.

1) Формування бази правил систем нечіткого виводу.

2) Фазифікації вхідних змінних.

3) Агрегація підумови в нечітких правилах продукцій. Для знаходження степені істинності умов кожного з правил нечітких продукцій використовуються парні нечіткі логічні операції. Ті правила, ступінь істинності умов яких відмінна від нуля, вважаються активними і використовуються для подальших розрахунків.

4) Активізація підзаключень в нечітких правилах продукцій. Здійснюється за формулою (10.2), при цьому для скорочення часу виведення враховуються тільки активні правила нечітких продукцій.

5) Акумуляція висновків нечітких правил продукцій. Здійснюється за формулою $\mu_{\tilde{D}}(x) = \max(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)) \quad (\forall \boldsymbol{x} \in \boldsymbol{X})$ для об'єднання нечітких множин, відповідних термам підзаключень, що відносяться до одних і тих же вихідних лінгвістичних змінних.

6) Дефазифікації вихідних змінних. Традиційно використовується метод центру ваги в формі (10.5) - (10.6) або метод центру площі (10.7).

10.3.2 Алгоритм Цукамото (Tsukamoto)

Формально алгоритм Цукамото може бути визначений таким чином.

1) Формування бази правил систем нечіткого виводу.

2) Фазифікації вхідних змінних.

3) Агрегація підумови в нечітких правилах продукцій. Для знаходження степені істинності умов всіх правил нечітких продукцій використовуються парні нечіткі логічні операції. Ті правила, ступінь істинності умов яких відмінна від нуля, вважаються активними і використовуються для подальших розрахунків.

4) Активізація підзаключень в нечітких правилах продукцій. Здійснюється аналогічно алгоритму Мамдані за формулою (10.2), після чого знаходяться звичайні (не нечіткі) значення всіх вихідних лінгвістичних змінних в кожному з підзаключень активних правил нечітких продукцій. У цьому випадку значення вихідної лінгвістичної змінної w_j в кожному з підзаключень знаходиться як розв'язок рівняння:

$$c_i = \mu(w_j) \ \ (\forall i \in \{1,2,\ldots,q\}), \tag{10.9}$$

де q - загальна кількість підвисновків в базі правил.

5) Акумуляція висновків нечітких правил продукцій. Фактично відсутня, оскільки розрахунки здійснюються із звичайними дійсними числами w_j.

6) Дефазифікації вихідних змінних. Використовується модифікований варіант у формі методу центру тяжіння для одноточкових множин:

$$y = \frac{\sum_{i=1}^{n} c_i * w_i}{\sum_{i=1}^{n} c_i}, \tag{10.10}$$

де n - загальна кількість активних правил нечітких продукцій, в підзаключеннях в яких присутня вихідна лінгвістична змінна w_j.

10.3.3 Алгоритм Ларсена (Larsen)

Формально алгоритм Ларсена може бути визначений таким чином.

1) Формування бази правил систем нечіткого виводу.

2) Фазифікації вхідних змінних.

3) Агрегація підумови в нечітких правилах продукцій. Використовуються парні нечіткі логічні операції для знаходження ступеня істинності умов всіх правил нечітких продукцій (як правило, *max*-диз'юнкція і *min*-кон'юнкція). Ті правила, ступінь істинності умов яких відмінна від нуля, вважаються активними і використовуються для подальших розрахунків.

4) Активізація підзаключень в нечітких правилах продукцій. Здійснюється використуванням формули (10.3), за допомогою чого знаходиться сукупність нечітких множин: C), C_1 ,..., C_q, де q - загальна кількість підзаключень в базі правил.

5) Акумуляція висновків нечітких правил продукцій. Здійснюється за формулою $\mu_{\tilde{D}}(x) = \max(\mu_{\tilde{A}}(x), \mu_{\tilde{B}}(x)) \quad (\forall x \in X)$ для об'єднання нечітких множин, відповідних термам підзаключень, що відносяться до одних і тих же вихідниїх лінгвістичних змінним.

6) Дефазифікації вихідних змінних. Може використовуватися будь-який з разглянутих вище методів дефазифікації.

10.3.4 Алгоритм Сугено (Sugeno)

Формально алгоритм Сугено, запропонований Сугено і Такагі, може бути визначений таким чином.

1) Формування бази правил систем нечіткого виводу. У базі правил використовуються тільки правила нечітких продукцій у формі:

ПРАВИЛО <#>: ЯКЩО «β_1 є α» І «β_2 є α» ТО «$w = \varepsilon_1 \cdot a_1 + \varepsilon_2 \cdot a_2$». (10.11)

де ε_1, ε_2 - деякі вагові коефіцієнти. При цьому значення вихідної змінної w у висновку визначається як деяке дійсне число.

2) Фазифікації вхідних змінних. Особливості фазифікації збігаються з розглянутих вище при описі даного етапу.

3) Агрегація підумови в нечітких правилах продукцій. Для знаходження степені істинності умов всіх правил нечітких продукцій, як правило, використовується логічна операція *min*-кон'юнкції. Ті правила, ступінь істинності умов яких відмінна від нуля, вважаються активними і використовуються для подальших розрахунків.

4) Активізація підзаключень в нечітких правилах продукцій. По-перше, із використанням методу (10.2) знаходяться значення ступенів істинності всіх висновків правил нечітких продукцій. По-друге, здійснюється розрахунок звичайних (не нечітких) значень вихідних змінних кожного правила. Це виконується за допомогою формули для висновку (10.11), в яку замість a_1 і a_2 підставляють значення вхідних змінних до етапу фазифікації. Тим самим визначаються множина значень $C = \{c_1, c_2, ..., c_n\}$ і множина значень вихідних змінних $W = \{w_1, w_2, ..., w_n\}$, де n - загальна кількість правил в базі правил.

5) Акумуляція висновків нечітких правил продукцій. Фактично відсутня, оскільки розрахунки здійснюються із звичайними дійсними числами w_j.

6) Дефазифікації вихідних змінних. Використовується модифікований варіант у формі методу центру тяжіння для одноточкових множин (10.10).

10.4. Приклади використання систем нечіткого виведення в задачах управління

Одним з основних напрямків практичного використання систем нечіткого вивода є вирішення завдань управління різними об'єктами або процесами. У цьому випадку побудова нечіткої моделі ґрунтується на формальному поданні характеристик досліджуваної системи в термінах лінгвістичних змінних. Оскільки крім алгоритму управління, основними поняттями систем управління є вхідні і вихідні змінні, то саме вони розглядаються як лінгвістичні змінні при формуванні бази правил в системах нечіткого виводу.

У загальному випадку мета управління полягає в тому, щоб на основі аналізу поточного стану об'єкта управління визначити значення керуючих змінних, реалізація яких дозволяє забезпечити бажану поведінку або стан об'єкта управління. Нині для вирішення відповідних завдань використовується загальна теорія управління, в рамках якої розроблені різні алгоритми знаходження оптимальних законів управління об'єктами різної фізичної природи.

Нечітка модель управління змішувачем води при прийнятті душа

Як перший приклад використання систем нечіткого виведення в задачах управління розглядається задача управління змішувачем води при прийнятті душа. Це завдання є однією з найбільш простих, яке може бути вирішене методами нечіткого моделювання. Для визначеності припустимо, що в якості алгоритму нечіткого виводу буде використовуватися алгоритм Мамдані.

Змістовна постановка задачі

При прийнятті душа на вхід змішувача подається холодна і гаряча вода по відповідним магістральним трубопроводам. Найбільш комфортні умови для душа створюються при наявності на виході змішувача теплої води постійної температури. Оскільки під час прийняття душу може спостерігатися нерівномірна витрата води, температура води на виході змішувача буде коливатися, приводячи до необхідності ручної зміни подачі холодної або гарячої води. Завдання полягає в тому, щоб зробити регулювання температури води автоматичною, забезпечуючи постійну температуру води на виході змішувача (рис. 10.12).

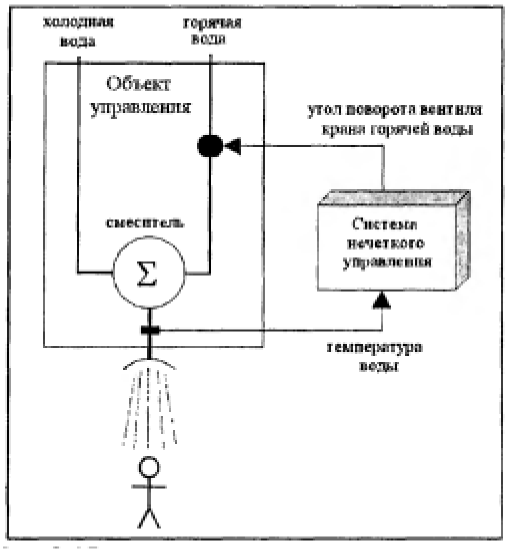


Рисунок 10.12 – Ілюстрація моделі нечіткого управління змішувачем води при прийнятті душа

Досвід прийняття душу дозволяє сформулювати кілька евристичні правила, які ми застосовуємо у разі регулювання температури води на виході змішувача:

1. Якщо вода гаряча, то слід повернути вентиль крана гарячої води на великій кут вправо.
2. Якщо вода не дуже гаряча, то слід повернути вентиль крана гарячої води на невеликий кут вправо.
3. Якщо вода тепла, то залишити вентиль крана гарячої води без впливу.
4. Якщо вода прохолодна, то слід повернути вентиль крана гарячої води на невеликий кут вліво.
5. Якщо вода холодна, то слід повернути вентиль крана гарячої води на великій кут вліво.

Ця інформація буде використовуватися при побудові бази правил системи нечіткого виводу, яка дозволяє реалізувати дану модель нечіткого керування.

Побудова бази нечітких лінгвістичних правил

Для формування бази правил систем нечіткого виведення необхідно попередньо виділити вхідні і вихідні лінгвістичні змінні. Очевидно, в якості вхідної лінгвістичної змінної слід використовувати температуру води на виході змішувача або формально: β_1 - "*температура води*". В якості вихідної лінгвістичної змінної будемо використовувати кут повороту вентиля крана гарячої води або формально: β_2 - "*кут повороту*".

У цьому випадку система нечіткого виведення буде містити 5 правил нечітких продукцій такого вигляду:

ПРАВІЛО_1: ЯКЩО "*вода гаряча*" ТО "*повернути вентиль крана гарячої води на великий кут вправо* "

ПРАВІЛО_2: ЯКЩО "*вода не дуже гаряча*" ТО "*повернути вентиль крана гарячої води на невеликий кут вправо* "

ПРАВІЛО_3: ЯКЩО "*вода тепла*" ТО "*залишити кут повороту крана гарячої води без зміни* "

ПРАВІЛО_4: ЯКЩО "*вода прохолодна*" ТО "*повернути вентиль крана гарячої води на невеликий кут вліво*"

ПРАВІЛО_5: ЯКЩО "*вода холодна*" ТО "*повернути вентиль крана гарячої води на великий кут вліво* "

Фазифікації вхідних змінних

Як терм-множини першої лінгвістичної змінної будемо використовувати множину T_1 = *{"гаряча", "не дуже гаряча", "тепла", "прохолодна", "холодна")* з функціями приналежності, зображеними на рис. 10.13, а.

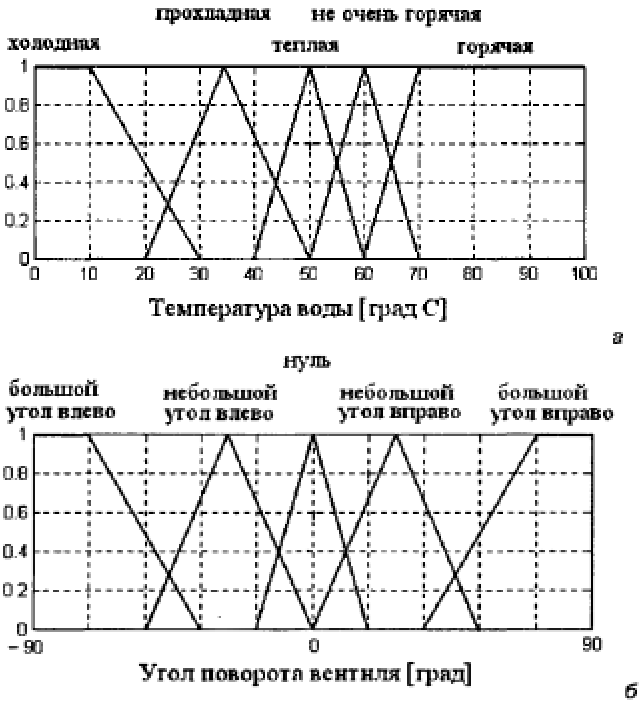


Рисунок 10.13 – Графіки функцій приналежності для термів лінгвістичної змінної "Температура води" (а), та лінгвістичної змінної "Кут повороту вентиля крана" (б)

Як терм-множину другої лінгвістичної змінної будемо використовувати множину T_2 = *{"великий кут вправо", "невеликий кут вправо", "нуль", "невеликий кут вліво","великий кут вліво"}* з кусково-лінійними функціями приналежності, зображеними на рис. 10.13, б.

При цьому температура води вимірюється в градусах Цельсія, а кут повороту - в кутових градусах. В останньому випадку поворот вправо означає позитивний напрям відрахунки, а поворот вліво - негативне.

Використовуючи як алгоритму виведення алгоритм Мамдані, розглянемо приклад його виконання для випадку, коли поточна температура води на виході змішувача дорівнює 55°C. В цьому випадку фазифікація вхідної лінгвістичної змінної приводить до значень степеней істинності 0.5 для правил нечітких продукцій з номерами 2 і 3. Ці правила вважаються активними і використовуються в поточному процесі нечіткого виводу.

Оскільки всі умови в правилах 1-5 задані у формі нечітких лінгвістичних висловлень першого виду, етап їх агрегування тривіальний і залишає ступені істинності 0.5 без зміни.

Наступним етапом нечіткого виводу є активізація висновків в нечітких правилах продукцій. Оскільки всі висновки правил 1-5 задані у формі нечітких лінгвістичних висловлювань першого виду, а вагові коефіцієнти правил за замовчуванням рівні 1, то активізація правил 2 і 3 призводить до нечітких множинам, функції приналежності яких зображені на рис. 10.14, а.

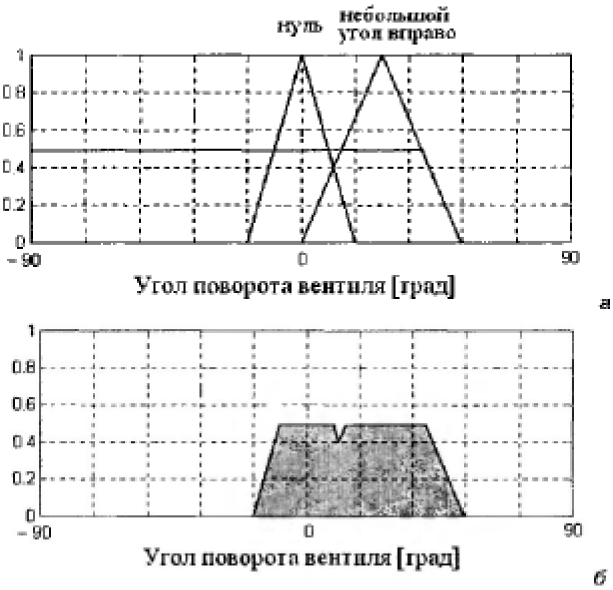


Рисунок 10.14 – Графіки функції приналежності для двох термів вихідної лінгвістичної змінної "Кут повороту вентиля крана" (а), і функції приналежності після акумуляції (б)

Акумулювання висновків нечітких правил продукцій з використанням операції max-диз'юнкції для правил 2 і 3 призводить в результаті до нечіткої множини, функція приналежності якої зображена на рис. 10.14, б.

Дефазифікації вихідний лінгвістичної змінної "*Кут повороту вентиля крана*" методом центру тяжіння для значень функції приналежності, зображеної на рис. 8.17, призводить до значення керуючої змінної, рівному повороту вентиля крана вправо на 16° (наближене значення). Це значення і є результатом виконання завдання нечіткого виводу для поточного значення вхідний лінгвістичної змінної "*Температуру води*".

Для реалізації цього алгоритму нечіткого управління необхідно організувати періодичне вимірювання температури води на виході змішувача в деякі дискретні моменти часу. При цьому, чим менше інтервал вимірювання цієї температури, тим вище виявляється точність регулювання температури води.

Що стосується реалізації власне процедури нечіткого управління, то для цієї цілі необхідно використовувати відповідні програмні або апаратні засоби, спеціально призначені для виконання всіх етапів нечіткого виводу.