

УДК 303.732.4

С.В. Кунцев, ГВУЗ "Українська академія банківського дела НБУ"

КЛАССИФИКАЦИОННЫЕ МОДЕЛИ – ОСНОВОПОЛАГАЮЩИЕ ФОРМЫ ЗНАНИЯ

Рассмотрены некоторые особенности создания и использования классификационных моделей, на которые целесообразно обратить внимание исследователю, эксперту или системному аналитику, занимающимся регистрацией и обработкой опытных данных.

Классификация является одним из фундаментальных процессов в науке, интеллектуальной деятельностью высокого уровня, направленной на понимание законов природы.

Сущность классификации заключается в разделении объектов на классы по признаку "сходство-различие" [1].

Процесс разбиения объектов на классы называют *классифицированием*, результат разбиения – *классификацией*, классификационную сетку – *классификатором*.

В частности, при классификации выявленных проблем целесообразно пользоваться классификатором проблем (табл. 1).

Таблица 1

Классификатор проблем

Признак классификации	Градации признаков				
	Тип решений, которые затрагивает проблема	Стратегические		Тактические	
Функция управления	Анализ	Планирование	Контроль	Организация	Мотивация
Глубина проблемы	Это действительно проблема			Симптом других проблем	
Факторы, обуславливающие проблему	Контролируемые		Частично контролируемые		Не контролируемые
Системный блок и подсистема	Входы:	Выходы:	Процессор:		Внешняя среда
Актуальность проблемы	Очень актуальна			Менее актуальна	

В отличие от числовых моделей, которые описывают один объект, классификационные модели описывают множество различных объектов.

Они являются качественными, хотя в них допускаются количественные переменные.

Исходной информацией для классифицирования являются протоколы наблюдения, представляющие собой матрицы данных.

Протокол наблюдений на классификационном уровне эксперимента содержит результаты измерения ряда признаков X для подмножества A объектов, выбранных из множества Γ : каждый объект $a_i \in A \subseteq \Gamma$ обладает значениями признаков $x_i = (x_{i0}, x_{i1}, \dots, x_{in}) \in \{X_0, X_1, \dots, X_n\} = X$, $i = 1, N$, n – число признаков, N – число объектов в A . Признак характеризует конкретное свойство объекта, поэтому иногда такой протокол называют таблицей “объект – свойство” [2].

Под “измерением” понимается алгоритмическая операция, которая приписывает данному наблюдаемому состоянию объекта, процесса, явления определенное обозначение: число, номер или символ. Измерение в классификационных моделях существенно шире обычного количественного измерения и обладает рядом особенностей:

- наблюдаемые явления, которые в принципе не допускают числовой меры, можно фиксировать в “слабых”, “качественных” шкалах;
- “расплывчатость” некоторых наблюдений признается их неотъемлемым естественным свойством, которое учитывается в виде математических формул;
- неизбежным свойством процесса измерения являются погрешности измерений;
- в процессе измерения временных зависимостей входов и выходов нужно учитывать их статистический характер.

Успех классифицирования во многом зависит от выбора измерительной шкалы [3]. Различают следующие группы шкал, предназначенных для измерения признаков трех видов: *номинальных* (шкала наименований); *порядковых* (шкалы порядка, частичного порядка, рангов, баллов); *количественных* (шкалы интервалов, отношений, абсолютная шкала).

При классификации дискретных по своей природе явлений чаще всего используется так называемая *шкала наименований*, обеспечивающая взаимно однозначное соответствие между классами эквивалентности и действительными числами. Для обозначения классов можно использовать:

- слова естественного языка (географические названия, собственные имена людей);
- произвольные символы (гербы и флаги, эмблемы);

- номера (регистрационные номера автомобилей, официальных документов);
- комбинации (почтовые адреса, экслибрисы личных библиотек, печати).

Названные обозначения эквивалентны простой нумерации. Обозначения классов упрощаются, если вводить их в виде иерархической структуры. Например: почтовые адреса, номера автомобилей.

Шкалу наименований можно использовать (условно) и тогда, когда классифицируемые состояния образуют непрерывное множество. Однако условность введенных классов может рано или поздно проявиться на практике. Например, возникают трудности точного перевода с одного языка на другой при описании цветовых оттенков: в английском языке голубой, лазоревый и синий цвета не различаются. В профессиональных языках имеет место аналогичная ситуация. В медицине названия болезней образуют шкалу наименований. Психиатр, ставя больному диагноз “шизофрения”, “паранойя”, “маниакальная депрессия” или “психоневроз”, использует номинальную шкалу. Название болезни лишь обозначает класс, внутри которого на самом деле имеются различия, так как эквивалентность внутри класса носит условный характер.

Между элементами номинальной шкалы определены следующие отношения:

1. Либо $A = B$, либо $A \neq B$.
2. Если $A = B$, то $B = A$.
3. Если $A = B$ и $B = C$, то $A = C$.

Здесь символ $=$ означает отношение эквивалентности.

Следующей по силе за номинальной шкалой является *порядковая (ранговая) шкала*. Этот класс шкал появляется, если кроме аксиом тождества 1-3 классы удовлетворяют следующим аксиомам упорядоченности:

4. Если $A > B$, то $B < A$.
5. Если $A > B$ и $B > C$, то $A > C$.

Примерами применения такой шкалы являются призовые места, воинские звания.

Порядковые шкалы модифицируются. Для уменьшения относительности порядковых шкал, для учета в них непрерывных величин, предпринимаются попытки их усиления путем введения третьего между двумя шкальными значениями. Примерами таких модификаций являются шкала силы ветра по Бофорту, шкала магнитуд землетрясений по Рихтеру.

Особый интерес представляют балльные шкалы оценки знаний учащихся.

Опытный преподаватель, слушая ответы учащихся или сравнивая их письменные работы, может обнаружить разницу между ними и установить, чьи ответы лучше – это типичное отношение порядка. Сравнение старшеклассника с младшеклассником по степени овладения знаниями проблематично.

Потребность общества в официальном определении степени квалифицированности проходящих обучение независимо от того, где, когда и как они получают образование, способствовала введению общепринятых шкал для оценивания степени овладения знаниями: 2-балльной, 4-балльной, 12-балльной, 100-балльной. Однако нужно понимать, что любая балльная шкала является порядковой. Попытки придать оттенок объективности и количественности балльной шкале введением стандартов и требований к знаниям учащихся не имеют успеха: оценки все равно получаются относительными. Принятая практика учета среднеарифметического балла не имеет смысла в порядковых шкалах. Разумно проводить не конкурсы документов, а конкурсы самих претендентов.

В социологических исследованиях часто оказывается полезным и возможным предложить опрашиваемому не только упорядочить заданный перечень альтернатив, но и указать, хотя бы грубо, силу предпочтения. Для этого используется порядковая шкала Черчмена и Акоффа. Приведем пример.

Пусть имеется четыре предмета. Сначала опрашиваемый упорядочивает их в порядке предпочтения: $A \geq B \geq C \geq D$. Затем его просят поставить в соответствие предметам любые числа между нулем и единицей, выразив грубо “силу” предпочтения. Пусть результат таков:

<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
1,00	0,85	0,75	0,20

С помощью дальнейших вопросов уточняется действительная “сила” предпочтений опрашиваемого.

При обработке опытных данных, зафиксированных в номинальной шкале, непосредственно с самими данными можно выполнять только операцию проверки их совпадения или несовпадения. Ведь обозначения классов – это только символы, даже если для этого использованы номера. Например, с номерами участников соревнований нельзя

обращаться, как с числами. Можно определить их равенство или неравенство.

Для записи этих операций используется символ Кронекера:

$$\delta_{ij} = \{1 : x_i = x_j; 0 : x_i \neq x_j\},$$

где x_i и x_j – записи разных измерений.

С результатами этой операции можно выполнить более сложные преобразования: вычислить количество совпадений, относительные частоты классов, сравнить эти частоты между собой по моде, выполнить различные статистические процедуры.

Для обработки данных в порядковых шкалах допустимыми операциями являются: вычисление δ_{ij} и ранга R_i . Сравнительные сведения о некоторых основных измерительных шкалах приведены таблице 2.

Таблица 2

Сравнение измерительных шкал

Название шкалы	Определяющие отношения	Эквивалентное преобразование шкал	Допустимые операции над данными	Вторичная обработка данных
Номинальная	Эквивалентность	Перестановки наименований	Вычисление символов Кронекера δ_{ij}	Вычисление относительных частот и операции над ними
Порядковая	Эквивалентность; предпочтение	Не изменяющее порядка (монотонное)	Вычисление δ_{ij} и рангов R_i	Вычисление относительных частот
Интервальная	Эквивалентность; предпочтение; сохранение отношений интервалов	Линейное преобразование $y = ax + b$, $a > 0$, $b \in R$	Вычисление δ_{ij} и рангов R_i и интервалов (разностей между наблюдениями)	Арифметические действия над интервалами

Известно, что чем “сильнее” измерительная шкала, тем больше сведений об изучаемом объекте, явлении, процессе дают измерения. Вполне естественно стремление каждого исследователя провести измерения в возможно более сильной шкале. Однако выбор шкалы измерения должен ориентироваться на объективные отношения, которым подчинена наблюдаемая величина, и лучше всего производить

измерения в той шкале, которая максимально согласована с этими отношениями. После измерений исследователь также может преобразовать протокол наблюдений, переведя их из одной шкалы в другую. Однако перевод данных в более слабую шкалу ухудшает качество выводов.

На практике, в реальных протоколах наблюдений встречаются особенности, которые следует учитывать при обработке данных:

1. Большая размерность – когда число объектов N и число признаков n очень велики. Это осложняет задачу, несмотря на применение компьютера.
2. Разнотипность данных. Разные признаки могут измеряться в различных шкалах. В этом случае изменять протокол нежелательно, правильнее применить алгоритмы для обработки разнотипных данных.
3. Пропущенные значения. Универсального способа “восстановления” пропусков нет. Можно исключить из таблицы строку и столбец, на пересечении которых находится пустая ячейка или “восстановить” пропущенные значения, однако это не всегда допустимо. Важно, чтобы критерий “восстановления” и цель обработки были согласованы.
4. “Зашумленность”, заключающаяся в том, что измеряемое значение отличается от измерения, занесенного в протокол, на некоторую случайную величину.

Способ обработки протокола зависит от цели обработки. Часто оказывается, что задача может быть сформулирована как определение по наблюдавшимся значениям признаков $x = (x_1, \dots, x_n)$ значений ненаблюдаемого (“селевого”) признака x_0 . Целевыми признаками, как правило, являются те параметры модели, которые требуется уточнить по экспериментальным данным.

С помощью классификационных моделей можно решать задачи различного типа:

- *упорядочивание объектов*. В этой задаче требуется установить отношения порядка между $x_{10}, x_{20}, \dots, x_{N0}$ (или некоторой их частью) по определенному критерию предпочтения;
- *кластеризация*. Эта задача заключается в поиске “естественной” группировки объектов. Количество классов и границы классов в пространстве признаков не задаются. Требуется определить их, исходя из “близости”, “похожести” или “различия” описаний объектов $x_i = (x_{i1}, \dots, x_{in})$. Компоненты вектора X_0 – признаки кластера, значения которых подлежат определению;

- *классификация* (распознавание образов). В этой задаче число классов задано. В том случае, если также заданы границы между классами, то имеем *априорную классификацию*. Если требуется найти границы, оценить по классифицированным примерам, то такая задача называется *распознаванием образов* по обучающей выборке. Целевой признак X_0 замерен в номинальной шкале;
- *уменьшение размерности модели*. Задача заключается в уменьшении размера классификационной модели с помощью отбора наиболее информативных признаков.

Литература

1. Гносеологический статус классификации как формы познания: Автореф. дис. канд. философ. наук: 09.00.01 – онтология и теория познания / М.В. Стёпкина; Самарский муниципальный университет Наяновой. – Самара, 2006. – 18 с.
2. Лбов Г.С. Методы обработки разнотипных экспериментальных данных. – Новосибирск: Наука, 1981. – 196 с.
3. Пфанцагль И. Теория измерений. – М.: Мир, 1976. – 244 с.

Summary

S. Kuncsev. Classification Models as the fundamental forms of knowledge.

Considered some features of creation and use of classification models. A researcher, an expert or a system analyst engaged in registration and experimental data processing is to pay attention to these features.

Рукопис надіслано до редакції 17.12.2007.

УДК 111(058)

М.Е. Щербина, ГВУЗ "Украинская академия банковского дела НБУ"

ПОЛИМОДАЛЬНОСТЬ ЛОГИКИ В ИСТОРИЧЕСКОМ ПРОСТРАНСТВЕ ПРОЧТЕНИЙ НЕКОТОРЫХ ФРАГМЕНТОВ ВЕТХОГО ЗАВЕТА

Полиmodalность в сфере логики человеческого мышления неоднократно обнаруживалась на разных ступенях развития рефлексивных форм гуманитарного знания; освещались также ее конструктивные социокультурные функции. В данной статье рассмотрены модусы логических конструкций, возникающие при попытках истолкования некоторых "темных" фрагментов текстов Ветхого Завета, смысл которых на сегодняшний день не вполне прояснен с точки зрения их обычного культурологического прочтения.

Сфера логики на протяжении тысячелетий осмысливается как один из атрибутивных уровней развития человеческой сущности. Она является одной из тех очеловеченных форм переработки информации, на которых базируется как духовная, так и материальная культура.

© М.Е. Щербина, 2008