

Особенности автоматизированной обработки тестовых заданий открытого типа в компьютерных обучающих системах

Золотухина О.А.

Донецкий национальный технический университет,
кафедра программного обеспечения интеллектуальных систем,
goldy.oksana@gmail.com

This paper analyzes the methods of automated processing of an open test tasks in the learning computer systems. The main characteristics of text arrays used in open tests are discussed in. Developed an algorithm of the open test tasks responses automated processing. For pre-processing text is proposed to use a stop words list based on the learning system subject domain and keywords synonyms.

ВВЕДЕНИЕ

Современные образовательные процессы широко используют компьютерные информационные технологии. Это проявляется не только в расширении доступа участников образовательного процесса к информации посредством сети Интернет, но и в повсеместном внедрении в учебный процесс таких форм образования, которые используют компьютер и компьютерные программы как средство обучения. Спектр программного обеспечения, которое используется в учебном процессе, достаточно широк – начиная от обычных справочно-информационных систем и заканчивая адаптивными системами обучения, которые используют сложные модели обучения и модели обучаемого и могут в какой-то степени заменить живое общение с учителем. Подсистема контроля знаний является одной из ключевых в составе любой компьютерной обучающей системы, поскольку она позволяет не только оценивать уровень знаний обучаемого, но и управлять обучением на основе данных о результатах деятельности обучаемого. Тестовые задания являются наиболее

распространенной формой организации контроля результатов деятельности обучаемого. Включение в учебный процесс открытых тестовых заданий позволяет максимально приблизить процесс обучения с помощью компьютера к естественному процессу обучения с помощью учителя. Основная сложность автоматизированной обработки ответов в заданиях такого типа возникает в тех случаях, когда требуется не просто сравнить эталонный ответ и ответ обучаемого на точное совпадение, а оценить степень схожести ответа.

ОРГАНИЗАЦИЯ ТЕСТИРОВАНИЯ В КОМПЬЮТЕРНЫХ ОБУЧАЮЩИХ СИСТЕМАХ

Общепринятой является следующая классификация тестовых заданий:

- задания с выбором/определением одного или нескольких правильных ответов;
- задания на определение соответствия;
- задания на определение последовательности.

Каждое из заданий может быть реализовано как задание закрытого или открытого типа. Наиболее часто используются задания закрытого типа, что позволяет существенно упростить процесс их обработки. Однако, это влечет за собой ряд недостатков:

- проверяются только те знания, которые возможно представить в виде закрытых тестов, а для многих предметных

областей набор таких знаний может быть весьма ограниченным и не охватывает весь учебный материал;

– набор возможных вариантов ответов в тестах ограничен некоторым множеством (при формировании произвольных вариантов ответов имеется сильная привязка к предметной области, что требует разработки специализированных алгоритмов автоматизации формирования ответов для каждой конкретной системы);

– снижается уровень творческого подхода к процессу обучения.

При использовании тестовых заданий открытого типа обучаемому предоставляется возможность самостоятельного формирования ответа. Чаще всего ответы формулируются в следующем виде:

– числовое данное, символ или аббревиатура – обработка таких ответов не представляет особой сложности, ввиду простоты сравниваемых элементов и однозначности вариантов правильных ответов;

– слово на естественном или формальном языке – при обработке слова необходимо учитывать возможность опечатки или использования обучаемым синонимичных понятий, но в целом, в большинстве случаев задача сводится к определению идентичности введенного и эталонного значения;

– набор слов на естественном языке (словосочетание, предложение, текст из нескольких предложений) – является наиболее сложным в обработке, поскольку необходимо учитывать не столько варианты полного совпадения, сколько степень релевантности текста, который сформулировал ученик, к ответу (или ответам), которые содержатся в базе данных системы.

ОБРАБОТКА ОТВЕТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ В ТЕСТАХ ОТКРЫТОГО ТИПА

Рассмотрим вариант реализации открытого теста, при котором ответ формулируется как одно или несколько слов естественного языка.

В этом случае задача оценки правильности ответа обучаемого сводится к определению схожести двух текстовых массивов: эталонного (заданного учителем) и введенного учеником. Существует множество методов сравнения текстовой информации, однако, для каждой конкретной задачи они могут давать разные результаты релевантности. Выбор методов обработки существенно зависит от размера текстовых массивов и от самой задачи обработки.

Тестовые задания и ответы к ним в компьютерных обучающих системах характеризуются небольшим размером текстового массива и привязкой к некоторой предметной области. Ответы, даваемые непосредственно учеником, помимо тех же свойств, дополнительно характеризуются наличием ошибок, связанных с некачественным набором текста и/или незнанием правильного написания тех или иных слов. Кроме того, при ответе на открытый вопрос, обучаемый может использовать синонимичные понятия, а также формулировать семантически идентичное, но синтаксически отличающееся от эталонного, предложение.

Начальным этапом при любом анализе текстовой информации является ее предварительная обработка. Традиционно, предобработка текста включает удаление стоп-слов и выделение основ слов. Базы стоп-слов в нашем случае должны быть реализованы с учетом предметной области обучающей системы.

Для выделения основ слов целесообразно применять процедурные методы морфологического анализа (например, стемматизацию) – в отличие от декларативных методов они не используют базы основ слов, что позволяет существенно сократить время формирования основы слова, а также уменьшить размер программы

(что немаловажно, например, при переходе на мобильные платформы). Однако, следует заметить, что в случае достаточных вычислительных ресурсов все-таки предпочтительнее применять декларативные методы, поскольку они используют словари, содержащие все возможные варианты формирования слов.

Независимо от результатов морфологического анализа для повышения качества сравнения текстов словам в исходном корпусе необходимо поставить в соответствие синонимы. Степень синонимичности слов должна регулироваться в зависимости от привязки к конкретной предметной области.

Для учета опечаток и ошибочно написанных слов, предлагается сравнивать не сами слова текста, а их фонетические коды. Фонетический код слова представляет собой цифровую или буквенно-цифровую комбинацию, заменяющую исходное слово.

Поскольку в славянских языках порядок слов не настолько важен, как, например, в английском языке, то при сравнении текстов важным является не столько месторасположение слов в предложении, сколько количество слов (словосочетаний) эталонного текста, встретившихся в ответе обучаемого и взаимное расположение ключевых слов относительно друг друга (семантически связанные цепочки слов). Если представить сравниваемые тексты в виде семантических сетей, содержащих фонетические коды слов и синонимичных им понятий, то задачу можно свести к определению инвариантов графов, описывающих сравниваемые корпуса текстов, однако, в отличие от классической задачи, здесь необходимо помимо связей между вершинами учитывать еще и семантическую близость самих вершин. Семантическая близость вершин может определяться на основе близости фонетических кодов.

Алгоритм автоматизированной обработки ответов в тестовых заданиях открытого типа

в общем виде можно описать следующим образом:

1. Удаление стоп-слов из текста ответа.
2. Выделение основ оставшихся слов.
3. Определение синонимов слов.
4. Построение фонетических кодов слов.
5. Построение семантической сети ответа.
6. Сравнение построенной семантической сети с эталонной.

Предварительные этапы работы подсистемы контроля знаний в целом должны включать формирование и настройку базы стоп-слов и синонимов предметной области.

ВЫВОДЫ

Оценка правильности ответов обучаемого в тестах открытого типа является трудоемким процессом. Рассмотренный подход позволяет расширить возможности автоматизированной обработки ответов в тестах открытого типа и снизить влияние ошибок, допускаемых учеником в ответах непреднамеренно. Повышение качества сравнения текстов также достигается за счет привязки к предметной области при работе со стоп-словами и с синонимами слов.

ЛИТЕРАТУРА

- [1] А. Родишевский. Алгоритмы приближительного сравнения текста.[Электронный ресурс]. – Режим доступа: <http://www.morfoedro.it/doc.php?n=222&lang=ru/>
- [2] P. Willett The Porter stemming algorithm: then and now // Program: Electronic Library and Information Systems. – 2006. – В. 3. – Т. 40. – С. 219-223.
- [3] Неелова Н.В. Исследование лексического метода вычисления схожести строк веб-документа с учетом предварительной обработки / Н.В. Неелова // XXXV Гагаринские чтения: Научные труды Международной молодежной научной конференции в 8 томах – М. : МАТИ, 2009. – [4 т.] – С. 31-32.
- [4] Драль А.А. Классификация коротких текстовых документов /А.А. Драль, Э. Мбайкоджи// Тезисы докладов. Всероссийской конференции. с международным участием Москва, РУДН, 23–27 апреля 2012 года. – М.:РУДН, 2012. – 369 с.:ил. – С.121-123.