

Методы тестирования знаний на основе применения аппарата нейронной сети

Представлены методы тестирования знаний учащихся и алгоритм анализа ответов на примере гуманитарных дисциплин. Данные методы основаны на комбинации традиционных методов тестирования и применении аппарата нейронных сетей. Уникальность алгоритма позволяет использовать его в различных сферах анализа естественно-языковых текстов.

Ключевые слова: контроль знаний, модели и алгоритмы, методы тестирования, компьютерные системы контроля знаний, нейронная сеть.

TESTING METHODS OF KNOWLEDGE ON THE BASIS OF NEURAL NETWORK

The methods of testing students' knowledge and ability to assess the responses to the example of the humanities. These methods are based on a combination of traditional testing methods and apparatus using neural networks. Uniqueness of algorithm allows to use it in various spheres of the analysis of natural language texts.

Keywords: knowledge control, models and algorithms, computer systems of knowledge control, neural networks.

Введение

В современном мире система образования постоянно модернизируется и вносит свои коррективы. В связи с этим растет популярность компьютерных и аналоговых тестов. Под компьютерными тестами будем понимать автоматизированные компьютерные системы тестирования, под аналоговыми – классическое тестирование, т.е. бумажные тесты. При разработке компьютерных и аналоговых тестов возможно два типа представления тестовых заданий – в открытой и закрытой форме.

Классификация тестов основывается на наличии или отсутствии дополнительной информации, вводимой испытуемым. В случае необходимости дополнительной информации тест относится к открытой форме, при ее отсутствии – к закрытой. Тестовые задания закрытой формы, как правило, сводятся к предъявлению тестируемому фиксированного множества тестовых заданий и различных вариантов ответов на каждое из них [1, 2]. Задача тестируемого состоит в выборе одного или нескольких истинных, по его мнению, ответов на каждое тестовое задание [3]. Основу этих

способов составляет оценивание истинности предлагаемых вариантов ответов, что требует от организатора тестирования признать абсолютную истинность одних вариантов ответа и абсолютную ложность других вариантов. Недостаток такого подхода состоит в невозможности учитывать при тестировании неполные или не совсем точные ответы обучаемого [4].

Как правило, тестовое задание закрытой формы состоит из двух частей: вопросной части, содержащей вопрос или утверждение, и ответной, содержащей 3 или 4 варианта ответа, среди которых только один верный. Задания же открытой формы требуют ввода дополнительной информации – дополнения. Как правило, ответы на задания открытой формы анализируют эксперты – опытные специалисты в данной области.

В заданиях с выбором одного ответа большая вероятность угадывания верного ответа. Она равна обратной величине количества вариантов ответов. Задания, называемые заданиями с множественным выбором [5], фактически являются расширенным вариантом заданий с единственным выбором. Веро-

ятность угадывания правильного ответа в этом типе заданий очень мала, поскольку количество вариантов неправильных ответов в заданиях может быть различным.

Задача авторов состоит в анализе и усовершенствовании методов тестирования знаний, а также существующих алгоритмов обработки текстовой информации.

1. Метод уточнения результата

Данный метод является дальнейшим развитием результатов работ [1, 3]. За основу используется вопрос закрытого типа, и в случае, если ответ будет дан неверно, тестируемому задается еще один вопрос, ответ на который требуется дать в открытой форме, т.е. тестируемый сможет объяснить, почему он дал именно этот ответ. Это позволяет исключить элемент угадывания, а также дается возможность исправить результат, если ответ был дан неверно. В результате тестирования выводятся данные, в которых отражается итоговый результат, где тестируемый может увидеть, в каком месте он допустил ошибку.

Рассмотрим простой пример из тестов по философии. Тестируемо-



Артур Александрович Мицель,
д.т.н, профессор каф. АСУ,
Томский государственный
университет систем управления и
радиоэлектроники
www.tusur.ru
Профессор каф. информационных
систем (ИС),
Юргинский технологический
институт НИ ТПУ
uti.tpu.ru
Тел.: 8 (3822) 70-15-36
Эл. почта: maa@asu.tusur.ru

Artur A. Mitsel,
Doctor of Engineering Science, Professor
of Department ASC
Tel.: 8 (3822) 70-15-36
E-mail: maa@asu.tusur.ru
Tomsk State University of Control
systems and Radioelectronics
www.tusur.ru
Professor, Department of Information
Systems (IS), Institute of Technology
Yurginskiy NO TPU
uti.tpu.ru



Алексей Андреевич Погуда,
ст. преподаватель ФИТ
Тел.: 8 (3822) 52-94-98
Эл. почта: alexsmail@sibmail.com
Томский государственный
университет
www.tic.tsu.ru

Aleksey A. Poguda,
senior Lecturer FIT
Tel.: 8 (3822) 52-94-98
E-mail: alexsmail@sibmail.com
Tomsk State University
www.tic.tsu.ru

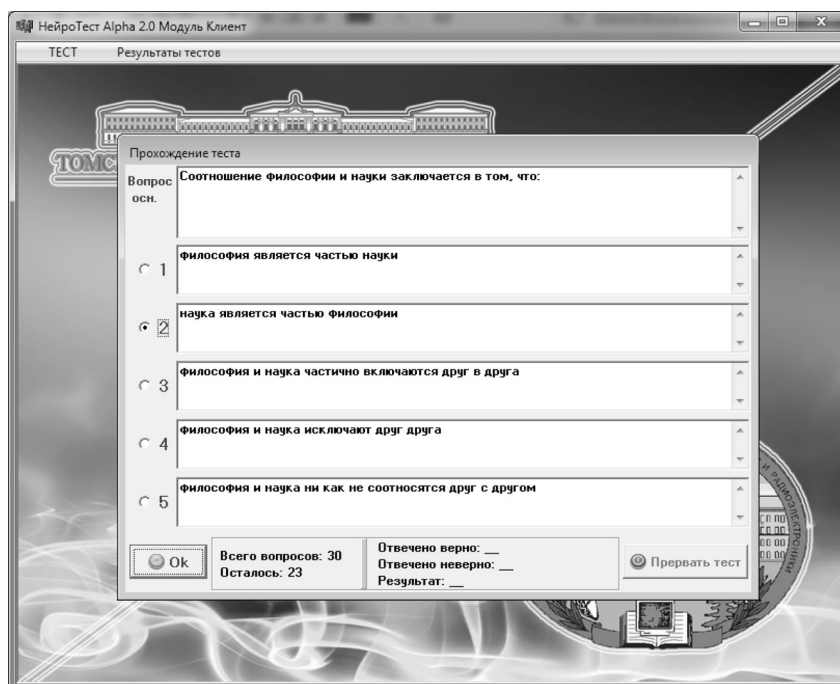


Рис. 1. Пример прохождения теста с применением метода уточнения результата в разработанной системе тестирования «НейроТест»

му задается вопрос «Соотношение философии и науки заключается в том, что...» и предлагаются 5 вариантов ответа (рис. 1). Предположим, что был дан ответ «наука является частью философии», система тестирования распознает ответ как частично неправильный и задает дополнительный вопрос, ответ на который требуется дать в открытой форме (рис. 2).

Рассмотрим предлагаемый метод более подробно. Метод содержит 3 основных блока [3].

В первом блоке содержатся вопросы закрытой формы. Результат ответа первого блока учитывается при окончательном анализе, т.е. после обработки ответа открытой формы. В некоторых тестах закрытой формы встречаются вопросы с «изюминкой», на которые следует

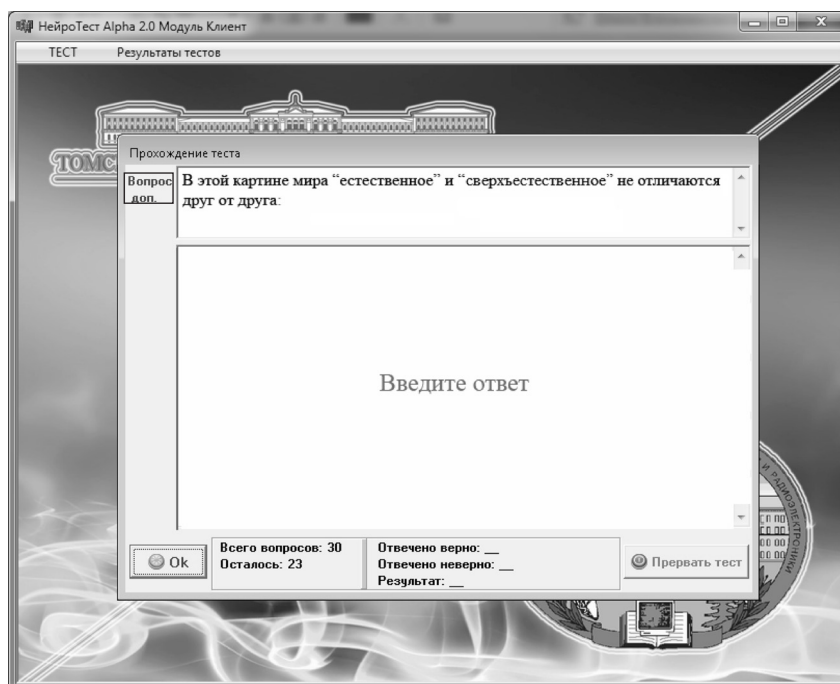


Рис. 2. Ответ на дополнительный вопрос открытой формы



Константин Андреевич Семенов,
студент
Тел.: 8 (3822) 70-15-36
Эл. почта: icetomcat@gmail.com
Томский государственный
университет систем управления и
радиоэлектроники
www.tusur.ru

Konstantin A. Semenov,
student
Tel.: 8 (3822) 70-15-36
E-mail: icetomcat@gmail.com
Tomsk State University of Control
systems and Radioelectronics
www.tusur.ru



Анастасия Евгеньевна Утешева,
студент
Тел.: 8 (3822) 70-15-36
Эл. почта: uae2010@mail.ru
Томский государственный
университет систем управления и
радиоэлектроники
www.tusur.ru

Anastasia E. Utesheva,
student
Tel.: 8 (3822) 70-15-36
E-mail: uae2010@mail.ru
Tomsk State University of Control
systems and Radioelectronics
www.tusur.ru

НейроТест Alpha 2.0 Модуль Клиент

ТЕСТ Результаты тестов

Результаты тестов

Фамилия	Имя	Отчество	Уч.завед	Курс	Группа	Дата/время	К.вопр	Прав	Непр	%	Прерв.
Иванов	Иван	Иванович	ТГУ ФИТ	4	18603	01.07.2011 22:27:40	30	30	0	100	0
Петров	Иван	Семенович	ТГУ ФИТ	4	18603	04.07.2011 11:51:37	31	14	17	46	0
Сидоров	Петр	Петрович	ТГУ ФИТ	5	18501	05.07.2011 10:13:15	0	0	0	0	1
Семенов	Олег	Андреевич	ТГУ	2	17845	05.07.2011 10:17:53	21	9	12	51	0
Ленов	Василий	Львович	ТГУ	5	19808	05.07.2011 10:18:43	33	14	19	48	0
Анохов	Лев	Семенович	ТГУ	5	19808	07.07.2012 12:31:10	20	8	12	45	0
Елизарова	Ирина	Львовна	ТГУ	4	18905	06.07.2012 12:39:48	18	7	11	47	0
Иванов	Андрей	Андреевич	ТУСУР	4	499	09.08.2012 23:47:20	5	2	3	45	0
Рыбанов	Константин	Алексеевич	ТУСУР	4	499	27.08.2012 15:48:38	12	7	5	66	0
Семашко	Иван	Иванович	ТГУ	5	18602	02.10.2012 14:24:29	12	5	7	54	0
Романенко	Ростислав	Иванович	ТУСУР	4	405	02.10.2012 14:24:58	0	0	0	0	1
Колотаев	Иван	Иванович	ТУСУР	5	18003	28.11.2012 16:21:34	0	0	0	0	1

Завершенность теста
☒ Все
☐ Завершённые
☐ Незавершённые
 Применить
 Закрыть

Рис. 3. Результаты тестирования в системе тестирования «НейроТест»

давать строго определенные варианты ответа, но при этом среди выбранных вариантов приводятся и другие, частично правильные ответы. Поэтому для частных случаев возможные значения дополняются определенным заранее параметром, который означает частично верный ответ.

Второй блок отвечает за обработку ответа, который введен тестируемым на вопросы открытой формы, с помощью разработанного алгоритма и применения нейронной сети. Здесь тестируемый либо подробно объясняет, почему он выбрал именно этот вариант ответа в первом блоке, либо система задает ему дополнительный вопрос, касающийся этой темы, например, если ответ был дан неверно (рис. 2). Данный параметр задается орга-

низаторами тестирования. В итоге ответ тестируемого сравнивается с правильным ответом, после чего на выходе мы имеем процентное соотношение истинности. Особенностью этого блока является то, что если тестируемый неверно ответил на вопрос в блоке 1, то здесь ему дается шанс исправиться, так как в этом случае в блоке 2 ответ оценивается независимо от ответа в блоке 1.

В третьем блоке вычисляется итоговый результат тестирования. Оценка в тесте выводится за каждый ответ, и в конце теста выводится результат в виде средней оценки. Результат оценки по каждому вопросу выводится из заранее подготовленных шаблонов, которые могут изменяться для каждого типа вопросов. Так, например, при

Таблица 1

Детальный просмотр заданных вопросов

Вопрос	Формулировка вопроса	Балл за ответ блок А / блок Б	Оценка
6	Соотношение философии и науки заключается в том, что:	50% / 96%	5
1 доп	В этой картине мира «естественное» и «сверхъестественное» не отличаются друг от друга.	—	—
Итоговая оценка за тест			5 (5)

Таблица 2

Детальный просмотр данных ответов

Вопрос	Данный ответ	Правильный ответ	Балл за ответ
6	наука является частью философии;	философия и наука частично включаются друг в друга;	50%
1 доп	в мифологической картине мира	в мифологической	96%

Таблица 3

Детальный просмотр заданных вопросов

ИД	1		
Вопрос	Формулировка вопроса	Балл за ответ блок А / блок Б	Оценка
10	В этой картине мира «естественное» и «сверхъестественное» не отличается друг от друга.	100%	5
Итоговая оценка за тест			5 (5)

Таблица 4

Детальный просмотр данных ответов

ИД	1		
Вопрос	Данный ответ	Правильный ответ	Балл за ответ
10	в мифологической	в мифологической	100%

результате ответа на вопрос более 75% система оценит ответ на «отлично». Более детально информация отображена в табл. 1 и 2.

2. Метод коррекции ошибок

Суть метода заключается в объединении методов оценки ответа на вопросы закрытого и открытого типа. Основное отличие предлагаемого метода от метода оценки ответа закрытой формы заключается в том, что среди ответов закрытой формы присутствует поле «Свой ответ», где тестируемый может предложить свой вариант ответа. Новый метод отличается и от ответов на задания открытой формы, так как отсутствует необходимость ввода своего ответа и можно выбрать уже существующие варианты. К преимуществам данного метода можно отнести и то, что преподаватель при проверке ответов открытой формы может составить представление о том, насколько усвоен пройденный материал, а также проанализировать ответы и переформулировать существующие ответы закрытой формы. При проверке ответов на задания открытой формы преподаватель работает только с полем «вопрос-ответ» и уникальными номерами, которые присваиваются при входе тестируемого в систему тестирования. Такой подход позволяет исключить пред-

взятое отношение преподавателя к тестируемому. Подробная информация отображена в табл. 3 и 4.

Данный метод не предусматривает дополнительные вопросы в случае абсолютно или частично неверного ответа, но организаторы тестирования могут активировать эту опцию.

Рассмотрим приведенный выше метод на конкретном примере. При прохождении тестирования был задан вопрос: «В этой картине мира естественное и сверхъестественное не отличаются друг от друга:», и предлагается 4 варианта ответа на выбор или ввод своего ответа (рис. 4). В данном случае, тестируемый вы-

брал 3 вариант ответа «в мифологической», что является правильным ответом.

Помимо основного преимущества предусматривается возможность динамической базы ответов. Под термином «динамическая база ответов» подразумевается, что к каждому вопросу в базе данных формируется база правильных и неправильных ответов. На начальной стадии преподаватель или организаторы тестирования формируют базу вопросов по той или иной дисциплине, где для каждого вопроса достаточно 2 правильных и 6 неправильных ответов. При прохождении тестирования из базы правильных ответов в случайном порядке выбирается только 1 верный ответ, а из базы неправильных ответов – 3. Если при прохождении теста тестируемый предложит свой вариант ответа и он окажется верным, то система добавляет его в базу данных ответов и помечает маркером потенциально правильный ответ. Аналогично система поступает и в том случае, если ответ будет предложен неверно, за исключением того, что ответ будет помечен как неправильный. После завершения тестирования преподаватель либо соглашается с предложенным вариантом ответа, и тогда тот добавляется в базу правильных ответов, либо засчитывает его неправильным и, если потребуется, заносит в базу неправильных ответов.

Система тестирования «НейроТест» предусматривает возможность прерывания прохождения теста. Во время прохождения тес-

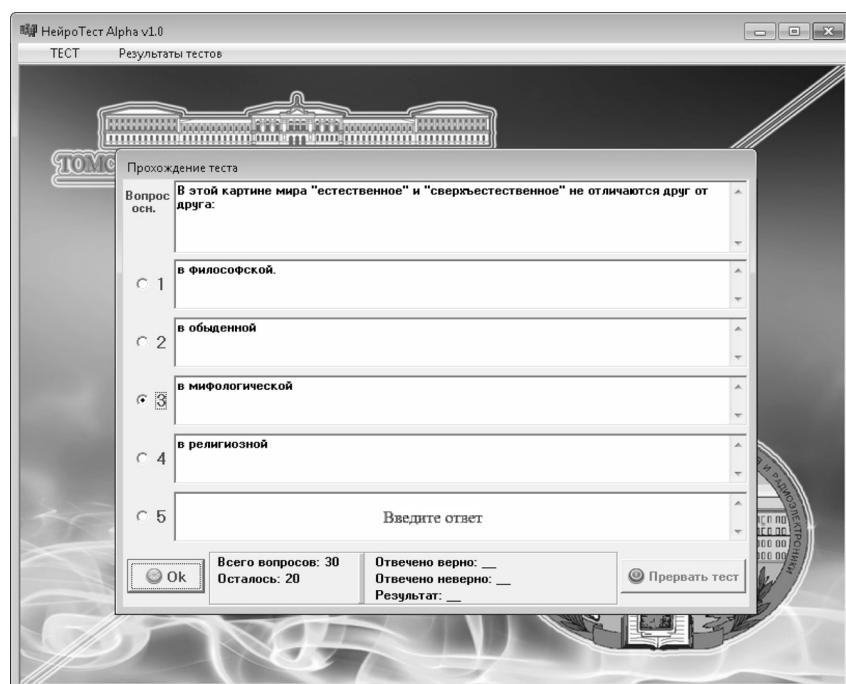


Рис. 4. Пример прохождения теста с применением метода коррекции ошибок

тирования в случае закрытия или прерывания теста в базе данных регистрируется событие о прерывании теста. Данная опция введена для исключения случаев, когда тестируемый выбирает вариант с максимальным итоговым баллом путем многократного прохождения тестирования.

3. Нейронные сети в задачах тестирования

Использование нейронных сетей представляет собой принципиально новый подход к решению задач тестирования и контроля знаний, а также идеально подходит для анализа текстов на естественном языке. Это может максимально приблизить оценивание знаний компьютером к выводам, которые делает преподаватель при проверке устного или письменного задания.

Искусственная нейронная сеть – это набор нейронов, соединенных между собой. Как правило, передаточные функции всех нейронов в нейронной сети фиксированы, а веса являются параметрами нейронной сети и могут изменяться. Некоторые входы нейронов помечены как внешние входы нейронной сети, а некоторые выходы – как внешние выходы нейронной сети. Подавая любые числа на входы нейронной сети, мы получаем какой-то набор чисел на выходах нейронной сети. Нейронные сети могут менять свое поведение в зависимости от состояния окружающей их среды. После анализа входных сигналов (возможно, вместе с требуемыми выходными сигналами) они самонастраиваются и обучаются, чтобы обеспечить правильную реакцию [5]. Обучение нейронной сети заключается в изменении «силы» синаптических связей между нейронами. После того, как проведен анализ входных сигналов, синаптические связи обучаются и настраиваются, чтобы обеспечить правильную реакцию. Обученная сеть, как правило, устойчива к небольшим отклонениям входных данных, это позволяет ей правильно обрабатывать образ, который содержит различные помехи и искажения [6]. Существует множество различных нейросетевых

архитектур, и к наиболее распространенным и изученным архитектурам относятся многослойный персептрон, нейронная сеть с общей регрессией, а также нейронная сеть Кохонена.

Выделяют три основных вида обучения: с учителем, самообучение и смешанный. В первом случае, т.е. с учителем, известны все верные ответы к каждому входному примеру, а для того, чтобы минимизировать ошибку, автоматически подстраиваются веса. Во втором случае при обучении распределяются по категориям формирующиеся образцы за счет раскрытия внутренней структуры и природы данных. В последнем случае используются оба подхода. Существует множество алгоритмов и методов обучения, которые ориентированы на решение широкого круга задач. Среди наиболее эффективных современных алгоритмов выделяют алгоритм обратного распространения ошибки. Идея такого алгоритма заключается в том, что изменение весов синапсов происходит с учетом локального градиента функции ошибки. Разница между реальными и верными ответами нейронной сети распространяется в обратном направлении навстречу потоку сигналов. В этом случае каждый нейрон может определить вклад каждого своего веса в суммарную ошибку сети. Главное при обучении – изменить синоптические веса пропорционально их вкладу в общую ошибку.

Для более корректной работы алгоритма выбранную нейронную сеть следует правильно обучить, т.е. подобрать значения ее весов так, чтобы она работала нужным образом. В нейронных сетях, используемых на практике, количество весов достигает несколько десятков тысяч, поэтому обучение –

это сложный и долгий процесс. Для многих существующих архитектур разработаны специальные алгоритмы обучения, которые позволяют настроить веса нейронной сети определенным образом. Среди популярных алгоритмов можно выделить метод обратного распространения ошибки (Error Back Propagation), как правило, его используют для обучения персептрона. Более подробно с данным алгоритмом можно ознакомиться в работе [3].

Одной из современной и гибкой к обучению является нейронная сеть Кохонена (рис. 5). По своей архитектуре она более подходит для обработки естественно-языковых текстов, так как в такой сети все объекты классифицируются, и представляется в виде некоторого вектора, подающего на вход нейронной сети. Количество нейронов во входном слое определяется количеством компонентов этого входного вектора, а количество выходов определяется количеством классов, но возможна ситуация, когда несколько нейронов относятся к одному классу. Весовые коэффициенты являются объектами того же типа, что и входные данные. Далее вводится функция расстояния между объектами данного типа, в нашем случае, это расстояние Левенштейна. Нейронная сеть Кохонена используется в классическом виде (рис. 6), но вычисление расстояния Левенштейна модифицировано под решение конкретной задачи.

Для проведения семантического анализа ответа открытой формы используется самоорганизующаяся карта (SOM), она состоит из компонентов, называемых узлами или нейронами. Их количество задается аналитиком или меняется в процессе обучения. Каждый из узлов описыва-

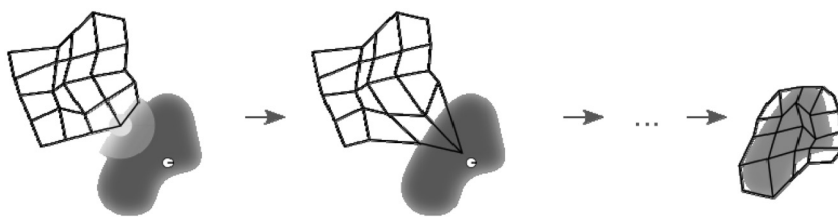


Рис. 5. Процесс самоорганизации из произвольного случайного положения вершин сети

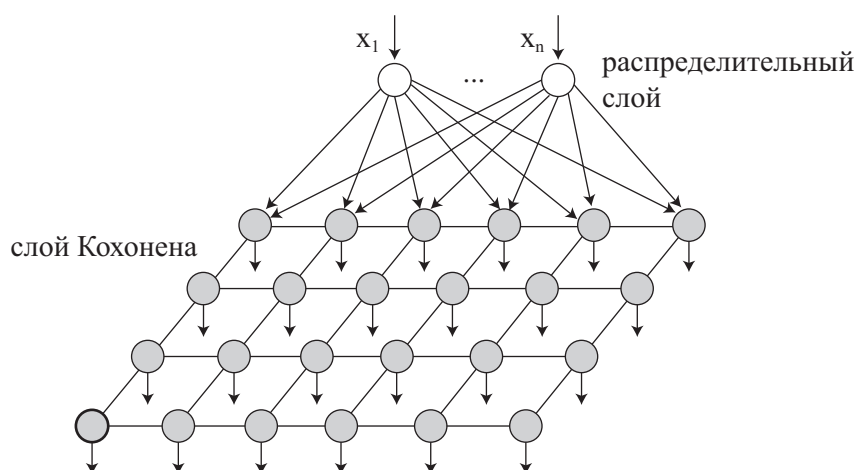


Рис. 6. Топология самоорганизующейся карты Кохонена

ется двумя векторами. Первый – вектор веса, имеющий такую же размерность, что и входные данные. Второй вектор представляет собой координаты узла на карте [3]. Обычно узлы располагают в вершинах регулярной решётки с квадратными или шестиугольными ячейками. В более широком смысле весовые коэффициенты могут являться любым объектом, важно, чтобы была определена функция расстояния между этими объектами.

Для вычисления расстояния между предложениями был выбран алгоритм взвешенного расстояния Левенштейна (LD, Levenshtein distance) [7]. Самым близким к расстоянию Левенштейна по качеству вычисления расстояния является «расстояние максимальной апостериорной вероятности», в работе [7] приведены эксперименты по вычислению обобщённой медианы слов, и расстояние Левенштейна показало лучшие результаты. LD для строк A и B определяется как

$$LD(A, B) = \min \{a(i) + b(i) + c(i)\}.$$

Здесь строка B получается из строки A путём $a(i)$ замен, $b(i)$ вставок и $c(i)$ удалений символов.

Но этого не достаточно, чтобы построить хорошую систему для вычисления семантического расстояния. Семантическое расстояние — это взвешенное расстояние Левенштейна, где веса операций замены, вставки, удаления и транспозиции символов подобраны таким образом, чтобы учитывать их смысловую значимость в предложении.

Для улучшения свойств алгоритма введём понятие веса операции.

Очевидно, что человеку свойственно ошибаться, особенно, если это касается написания текста на русском языке, поэтому необходимо оценивать значимость операции «превращения» строки A в строку B . Запишем взвешенное расстояние Левенштейна (WLD):

$$WLD(A, B) = \min \{pa(i) + qb(i) + rc(i)\},$$

где скалярные коэффициенты p , q и r вычисляются в ходе обучения системы.

Дамерау [7] утверждал, что 80% ошибок при наборе текста человеком являются транспозициями. Если к списку разрешённых операций добавить транспозицию, то такое расстояние называется расстоянием Дамерау – Левенштейна. Данное расстояние обладает следующим недостатком: если вычислить расстояние между совершенно разными короткими словами, то оно оказывается небольшим, в то время как расстояния между очень похожими длинными словами оказываются значительными. Для решения этой проблемы вводим норму расстояния WLD . Для этого необходимо разделить результат WLD на максимальную длину строк, при этом заранее наложив ограничение на весовые коэффициенты, но они не должны превосходить единицы. Таким образом, для невзвешенного нормированного расстояния справедливым становится утверждение, что если $LD(A, B) = 0,5$, то строка A наполовину похожа на строку B ,

и наоборот. Если после получения нормированного расстояния значение выходит за границы интервала $[0, 1]$ или меньше нуля, тогда полагаем его равным нулю, а в случае больше единицы полагаем его равным единице.

Существует ещё один существенный недостаток: при перестановке местами слов или частей слов получаются сравнительно большие расстояния. Для морфем слова эта проблема не существенна, но если переставить местами большие части текста, например обороты в предложении, абзацы или главы книги, то вывод расстояния будет ошибочным. При этом смысл написанного текста может даже не меняться, а расстояние изменяется весьма значительно. Если меняются местами части слова, то это критично для расчета расстояния. Расстояние должно резко увеличиваться, но для набора слов это правило должно работать наоборот.

Для решения этой задачи вводим понятие абстрактного уровня предложения. Абстрактным уровнем называется некий набор блоков символов, полученный после разбиения исходного текста по определённому закону. Счёт уровней начинается от уровня элементарных объектов и выше (рис. 7).

Для дальнейшей работы требуется определить закон разбиения. На первой стадии вводим в алфавит специальные символы-разделители на каждый абстрактный уровень или всего один символ для всех уровней. В случае если не используются символы разбиения, строки разбиваются по зависимостям между символами, которые, в свою очередь, могут быть совершенно любыми объектами, если между ними определено отношение равенства. Также разбивать входные данные можно по несколько фиксированных блоков на уровне до тех пор, пока не будет достигнут определённый предел или деление будет невозможным. В случае если на каждый абстрактный уровень вводится свой символ-разделитель, то не требуется дополнительных операций. Если используется всего один символ-разделитель, то текст можно разбивать на несколько

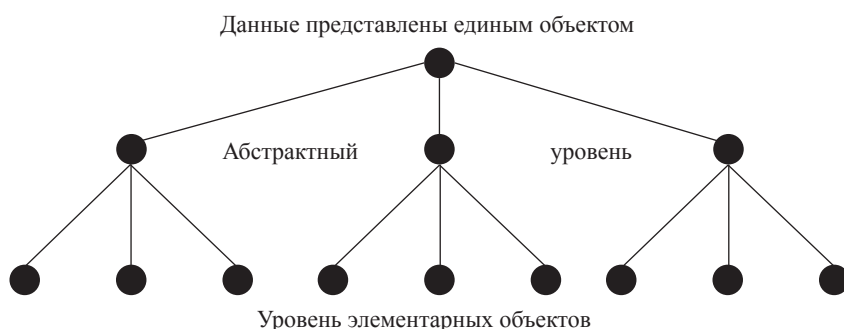


Рис. 7. Абстрактные уровни

фиксированных блоков в месте положения символов-разделителей. Если на уровне не остаётся символов-разделителей, то разбиение продолжается.

В таком представлении (рис. 7) абстрактные уровни не содержат данных, они только указывают уровни, расположенные ниже.

При вычислении расстояния получаем расстояние второго абстрактного уровня, начиная от корня, поэтому корень можно опустить при построении дерева и вычислении расстояния. Если расстояние окажется меньше заданного числа, то можно считать, что объекты равны. Под корнем подразумевается представлением данных как единого объекта.

Вычислять расстояние можно только между одинаковыми абстрактными уровнями. Если исходные предложения не получается разделить на одинаковое количество абстрактных уровней, то необходимо разделить входные данные на максимально возможное количество уровней.

Например, строка *A* делится на 2 абстрактных уровня, а строка *B* – на 5. Очевидно, что строку *B* требуется делить только до второго уровня, так как операция сравнения разных абстрактных уровней не определена. Таким образом, количество элементов в строке *B* на втором уровне или ниже будет больше, чем у строки *A*.

После деления на абстрактные уровни можно определить для каждого уровня цену перестановки блока. На нижних уровнях эта цена высока, но чем выше абстрактный уровень, тем эта цена становится меньше.

Далее рассмотрим проблемы, связанные с вычислением семантического расстояния.

На вход система получает набор текстов. Каждый текст помечен, к какому классу он относится, например оценка за ответ. Далее происходит попарно сравнение текстов из этого набора. Если возникает некоторое несоответствие между классами и расстоянием, то необходимо вносить корректировки в функцию оценки. Этих функций 4, для каждой операции по одной.

В [7] автор рекомендует использовать элементы теории информации для вычисления веса слова. Если слово встречается в тексте очень часто, то оно является «важным», и наоборот. Этот подход обладает основным недостатком – предлоги и союзы в тексте встречаются гораздо чаще, чем «важные» слова, например в произведении «Война и мир» самым часто встречающимся словом является предлог «и», который встречается ~22000 раз, следом за ним предлог «в» – ~11000 раз. Хотя смысла в тексте, возможно, они не несут. Этот недостаток решается путём анализа разницы текстов. Исходя из оценки и расстояния между ними, необходимо сформировать некие правила.

Оценка прямого преобразования хранится в виде:

$x(A)y \Rightarrow x(B)y$, class = any, cost = *S*, так как хранение в обычном виде может повлечь за собой коллизии в оценке расстояния. Эта запись говорит, что преобразование сегмента *A* в *B* в контексте *x* и *y* относится к классу any и имеет цену *S*.

Разделение на классы требуется для того, чтобы разделять специализированное обучение системы, а именно, обучать систему оценивать ответ на конкретный вопрос от общих знаний системы.

Рассмотрим влияние одних блоков абстрактных уровней на другие. Например, отрицание какого-либо слова или факта с помощью частицы «не». В таком случае будет применено правило коррекции цены существования частицы или ее отсутствия. Но бывают ситуации, когда можно это легко вычислить при наличии незначительного шума. Допустим, имеется 2 почти одинаковых предложения, но в одном отрицается важный факт, а в другом – нет. Посчитав разницу между предложениями, применим различные правила. Правило «отрицание последующего слова» гласит, что частица (блок) не имеет веса и отрицает последующее слово, т.е. вес следующего слова становится отрицательным.

Если после повторного пересчёта расстояния это правило пододвинет ближе всего к истинной оценке, то будем считать его правильным и занесём в базу. Если этой корректировки недостаточно, то поднимаемся на один абстрактный уровень выше и пытаемся найти, что ещё можно подкорректировать. Если в тексте встречается несколько одинаковых элементов, например, несколько предлогов «и», то при поиске подходящего правила для его обработки правила применяются ко всем предлогам.

Заключение

Проблема анализа текстовой информации изначально была связана с выявлением в его тематическо-содержательной природе собственно структуры, содержания, правил построения и языковых особенностей. Затем возникла потребность аннотирования и реферирования текстов, создания на их основе новых знаний, справок, отчётов и др. Важными свойствами тестовой информации являются дискретность и «предрасположенность» к формированию внутри себя микротем и микросмыслов, представляющих комбинации лексических, грамматических, синтаксических единиц (текстовых фрагментов). Анализ разных видов информации широко используется в самых различных областях, например, образовательной, социальной, политической и

военной сферах деятельности различных государств и общественных образований.

В данной работе были рассмотрены новые методы и универсальный алгоритм для дисциплин, требующих развернутого ответа.

Уникальность алгоритма и методов заключается в том, что их можно использовать не только в сфере образования, но и в других областях, требующих анализа текста, например системах поиска или проверки на плагиат. На основе предложен-

ных методов и алгоритма разработана система тестирования «Нейро-Тест», которая проходит апробацию на факультете инновационных технологий в Национальном исследовательском Томском государственном университете.

Литература

1. Мицель А.А., Погуда А.А. Модели и алгоритмы для компьютерного контроля знаний // Открытое образование. – 2010. – Вып. № 6. – С. 44–49.
2. Мицель А.А., Погуда А.А. Универсальный алгоритм проверки естественно-языковых тестов // Доклады Томского государственного университета систем управления и радиоэлектроники. – 2010. – № 2 (22), Ч. 2. – С. 290–294.
3. Мицель А.А., Погуда А.А. Нейросетевой подход к задаче тестирования // Прикладная информатика. – 2011. – Вып. № 5 (35). – С. 60–67.
4. Ким В.С. Тестирование учебных достижений. – Уссурийск: УГПИ, 2007. – 214 с.
5. Борисов А., Крумберг И., Федоров И. Принятие решений на основе нечетких моделей. – Рига: Зинатне, 1990. – 352 с.
6. Вербицкий А.А., Гридин В.Н., Солодовников В.И., Солодовников И.В. Использование нейронных сетей в задаче тестирования // Информационные технологии. – 2007. – № 9. – С. 21–26.
7. Кохонен Т. Самоорганизующиеся карты / пер. 3-го англ. изд. – М.: БИНОМ. Лаборатория знаний, 2011. – 655 с.
8. Мицель А.А., Погуда А.А. Технология обработки информации в задачах тестирования на основе нейронной сети // Современное образовательное пространство: пути модернизации: труды Междунар. заоч. научно-практ. конф. – Чебоксары, 2011. – С. 122–127.