УДК 004.02

МЕТОД ПРЕДВАРИТЕЛЬНОЙ ОЦЕНКИ ОТВЕТОВ ОБУЧАЮЩИХСЯ НА ОСНОВЕ ВЕКТОРНОЙ МОДЕЛИ ДОКУМЕНТОВ

Ч. Б. Миннегалиева[1[0000-0003-4648-1623]], Г. А. Сабитова[2[0009-0008-1840-4819]],

А. М. Гаялиев[3[0009-0006-9735-675X]]

[1,2,3] *Институт вычислительной математики и информационных технологий, Казанский федеральный университет, Казань*

[1]mchulpan@gmail.com, [2]gulshat.9922@gmail.com, [3]gayaliev.almaz@mail.ru

Аннотация

Рассмотрено применение векторных моделей для предварительного анализа ответов студентов, сформулированных в свободной форме. Векторные представления слов и документов получены при помощи моделей word2vec, doc2vec, BERT. Сходство ответа, данного обучающимся, и корректного ответа определялось с использованием косинусной меры. Выявлено, что векторные модели позволяют определить явно неверные ответы с достаточной точностью. Для ответов, которые близки по формулировке, предлагается провести дополнительный этап проверки. При помощи word2vec выполнена двоичная классификация ответов на определенные вопросы, приведены оценки точности, полноты, F1-меры.

Ключевые слова*: векторная модель, word2vec, doc2vec, BERT, косинусное сходство, векторное представление*

ВВЕДЕНИЕ

Контроль знаний является важным составляющим процесса обучения, и проблема его автоматизации остается актуальной. Блоки проверки усвоения учебного материала в составе электронных образовательных ресурсов и онлайн-курсов содержат вопросы разных типов. В системах управления обучением, например, в LMS Moodle или Blackboard, преподаватель также имеет возможность разработать качественный тест [1]. Для объективного контроля знаний необходим банк с достаточным количеством заданий. Ранее нами была изучена возможность генерации параметров и структуры тестовых заданий и автоматиче-

© Ч. Б. Миннегалиева, Г. А. Сабитова, А. М. Гаялиев, 2023.
Данная статья распространяется на условиях международной лицензии Creative Commons License Attribution 4.0 International (CC BY 4.0).

ской проверки ответов [2]. С возрастанием популярности чат-ботов они чаще используются в процессе обучения, одно из их основных преимуществ – оптимизация выполнения стандартных повторяющихся задач [3].

Возможности компьютерного тестирования расширяются, специалистами изучаются способы анализа ответов обучаемых, данных в свободной форме. Например, предлагается тестирование, основанное на использовании семантической сети ключевых понятий базы знаний обучающей системы. Система позволяет проводить проверку ответов в виде развернутых предложений, но анализируются не смысл самого предложения, а ключевые слова [4]. В работе [5] авторы предлагают фреймворк, который может помочь автоматическим системам оценки эссе предоставить подробные результаты оценки.

В работе [6] описана процедура, которая получает на вход ответ тестируемого на вопрос и ответы на тот же вопрос, хранящиеся в базе ответов, включая эталонные ответы. На выходе формируется оценка семантической близости введенного ответа пользователя к ответам, хранящимся в базе данных. Данный подход базируется на векторных представлениях word2vec слов на естественном языке, косинусной мере их сходства и семантических моделях русского языка. Также изучается использование других моделей, например, авторы анализировали письменные размышления учителей до поступления на работу при помощи предварительно обученной языковой модели BERT (Bidirectional Encoder Representations from Transformers) [7].

В настоящее время для работы с векторными представлениями слов и документов, а также автоматической обработки текстовой информации используются различные инструменты [8]. Ниже описаны результаты применения моделей word2vec, doc2vec, BERT для предварительного анализа ответов студентов, сформулированных в свободной форме.

ПРЕДВАРИТЕЛЬНЫЙ ЭТАП РАБОТЫ

Подготовка данных. В начале работы были выбраны разделы учебников, соответствующие программе курса «Компьютерная графика и дизайн», и на их основе подготовлен набор текстов по компьютерной графике. Данный курс по выбору предлагается студентам, обучающимся по программе бакалавриата. В ре-

зультате удаления повторяющихся фрагментов и объединения всех текстов получен текстовый файл, содержащий около 19000 предложений. Далее была проведена предобработка данных, удалены стоп-слова, остальные приведены в нормальную форму. Полученный файл в дальнейшем применялся при обучении разных моделей и дообучении моделей, предварительно обученных на больших объемах текстов.

Опрос студентов. Студентам, обучающимся на 3–4 курсах, было предложено дать короткие ответы, состоящие из 1–3 предложений, на вопросы по курсу «Компьютерная графика и дизайн». Студенты отвечали непосредственно после изучения определенной темы, сразу после завершения курса, повторно через месяц. Проверка знаний проводилась в аудиториях, без использования студентами справочных материалов.

Например, студентам было предложено назвать недостатки центральной проекции. В ходе изучения курса был сделан вывод, что центральная проекция объектов непригодна для представления точной формы и размеров объектов. Примеры ответов студентов: «Невозможность узнать точную форму и размер», «В центральной проекции объекты не отражают истинные размеры», «Не подходит для представления точных размеров объекта, а также точной формы». На каждый из 23 вопросов было получено до 40 ответов в зависимости от количества студентов в группе. Ниже ответы студентов приведены без исправления орфографических, грамматических, пунктуационных, речевых ошибок.

ОЦЕНКА ОТВЕТОВ ОБУЧАЮЩИХСЯ

Подготовка моделей. Word2vec принимает текстовый корпус в качестве входных данных и каждому слову сопоставляет вектор на основе статистики по совместному появлению слов в фразах [9]. Полученные векторные представления позволяют определить семантическую близость между словами при помощи косинусной меры. Doc2vec – это инструмент для получения векторного представления документов, который является обобщением word2vec. В настоящее время он используется специалистами для анализа текстовых данных, например, в оценке сходства наборов текстового контента сообществ социальной сети [10]. При работе с алгоритмом BERT на вход также подается текст, результатом является вектор. В данном случае обрабатываются слова и части слов. Модель

BERT была разработана позже, широко применяется в различных целях, в том числе для решения задач сентимент-анализа [11].

Для анализа ответа обучаемых, сформулированных в свободной форме, модели word2vec и doc2vec были обучены на подготовленных текстах по компьютерной графике. Также проверены результаты, которые были получены при использовании готовой модели word2vec, предварительно обученной на большом объеме общих текстовых данных [12]. Были использованы предобученная модель BERT и модель, которая была дообучена на текстах по компьютерной графике. Для обработки данных и обучения моделей использованы язык программирования Python, библиотеки NLTK, Gensim, scikit-learn, pymorphy3, pandas.

Анализ ответов с использованием модели word2vec. При помощи модели word2vec определены вектора слов, составляющих предложение. В таблице 1 приведены шаги алгоритма и результаты выполнения шагов на примере определения косинусного сходства между корректным ответом и ответом студента на вопрос: «Графический редактор – это программа для …» (дописать определение).

Таблица 1. Нахождение косинусного сходства между корректным ответом и ответом студента (вектор предложения вычисляется как среднее векторов слов)

	Корректный ответ	Ответ, данный студентом
Формулировка ответа	обработки, создания, просмотра, редактирования цифровых изображений на компьютере	работы с фотографиями
Предобработка	['обработка', 'создание', 'просмотр', 'редактирование', 'цифровой', 'изображение', 'компьютер']	['работа', 'фотография']
Получение векторов каждого слова	'обработка' → v_1, 'создание ' → v_2, 'просмотр' → v_3, 'редактирование' → v_4, 'цифровой' → v_5, 'изображение' → v_6, 'компьютер' → v_7	'работа' → w_1, 'фотография'→ w_2,
Нахождение вектора предложения	$V = \frac{v_1 + v_2 + v_3 + v_4 + v_5 + v_6 + v_7}{7}$	$W = \frac{w_1 + w_2}{2}$

Определение косинусного сходства	$Cos_sim = \frac{V \cdot W}{\lvert V \rvert \cdot \lvert W \rvert}$

В примере вектор ответа получен как среднее векторов слов. Также при нахождении вектора предложения коэффициенты определялись по схеме взвешивания TF-IDF.

В таблице 2 приведены значения косинусного сходства между корректным ответом и некоторыми ответами студентов на вопрос «Что означает вектор градиента?». Вариант корректного ответа: «Вектор, своим направлением указывающий направление наискорейшего возрастания некоторой величины».

Таблица 2. Косинусное сходство ответов обучающихся и верного ответа

Ответ	Значение косинуса
Вектор, показывающий направление наибольшего возрастания функции.	0.6017
Может быть, направление.	0.5638
Не знаю.	0.0429

Модель word2vec позволяет на первом этапе определить явно неверные (абсурдные) ответы, это демонстрирует и пример в таблице. Первый ответ верный, второй не был преподавателем засчитан как верный. Но значения косинуса для них оказались близки. В этом случае предлагается в ответе выделить слова, являющиеся ключевыми, и проверить их наличие в ответе студента. Схема оценивания приведена на Рис. 1.

Например, для ответа на рассмотренный вопрос ключевыми являются слова «направление», «возрастание», «функция». Если проверить слова, являющиеся ключевыми, второй ответ будет оценен как неверный.

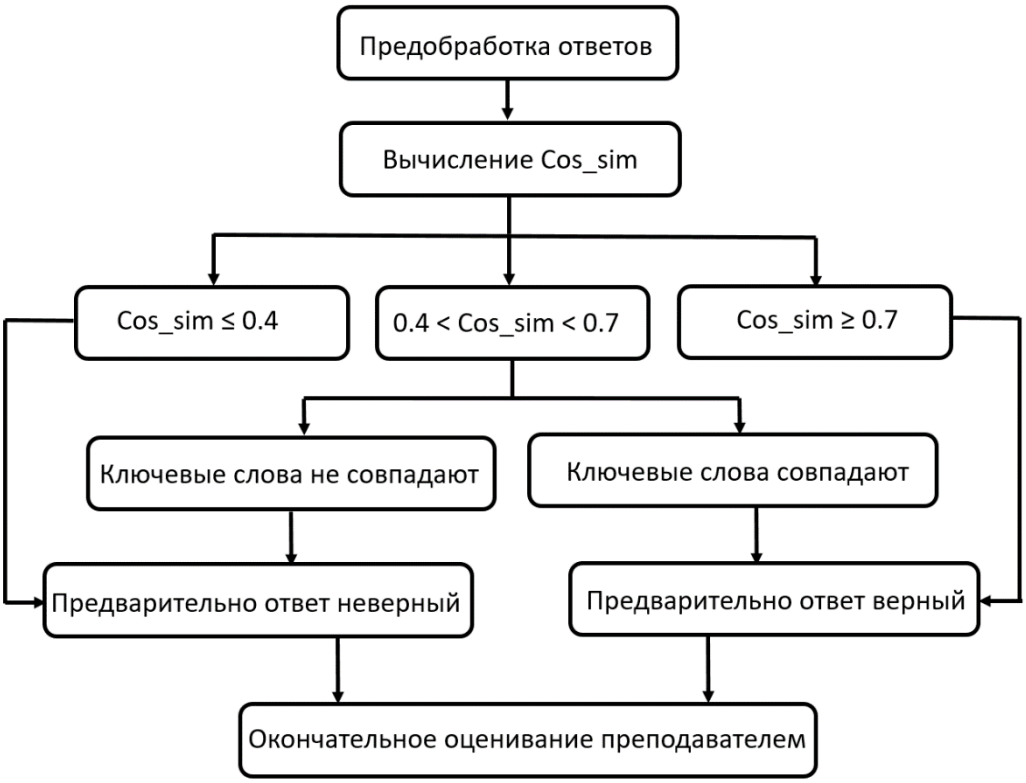

Рис. 1. Схема оценивания ответов

Возможность различения неточных и неполных ответов также зависит от формулировки вопроса и особенностей тестирования (при текущем контроле неточные ответы близки к верным, при проверке остаточных знаний косинусное сходство между ними меньше). Так, при анализе ответов студентов, данных через месяц после изучения курса, оценки, полученные при помощи модели, были точнее. Например, среди вопросов, на которые отвечали студенты, был следующий: «Что называют аналитической моделью поверхности?». Корректный ответ: «Описание поверхности математическими формулами». Примеры ответов студентов и косинусное сходство между ними и корректным ответом приведены в таблице 3.

Таблица 3. Косинусное сходство ответов обучающихся и верного ответа

Ответ	Значение косинуса
Формулы, описание поверхности.	0.9255
Представление фигуры в математической форме, в виде набора координат ребер и вершин.	0.4553

Модель, составленная предварительно, с помощью которой можно что-то узнать про исходный объект.	0.1411
Возможно, вид сверху.	0.3910

Из ответов, приведенных в таблице, только первый является верным, значение косинуса для данного ответа тоже близко к 1. Остальные ответы не являются верными, и значение косинуса меньше 0.5.

Ответы обучающихся были проанализированы с использованием модели word2vec, обученной на большом объеме текстов и подготовленных текстах по компьютерной графике. Обучение модели на текстах по компьютерной графике проводилось при разных значениях параметра. Оценка, определенная моделью, была наиболее близка к оценке, выставленной преподавателем, при следующих параметрах: минимальная частотность слова в корпусе – 1, размер окна наблюдения – 3, размерность вектора – 100.

Применение моделей doc2vec и BERT. Предварительно коэффициенты сходства в диапазоне от 0 до 1 между корректными ответами и ответами студентов были определены преподавателем (0 – полностью неверный ответ, 1 – полностью верный ответ). Далее коэффициент сходства определен при помощи моделей, с использованием косинусной меры. Ниже приведены результаты оценивания для одного из вопросов. Студентам было предложено продолжить предложение «Файлы формата CDR используются для хранения ...».

На графике (Рис. 2) представлены значения абсолютной разности между коэффициентами сходства, определенными преподавателем и моделью (модель doc2vec обучена на текстах по компьютерной графике).

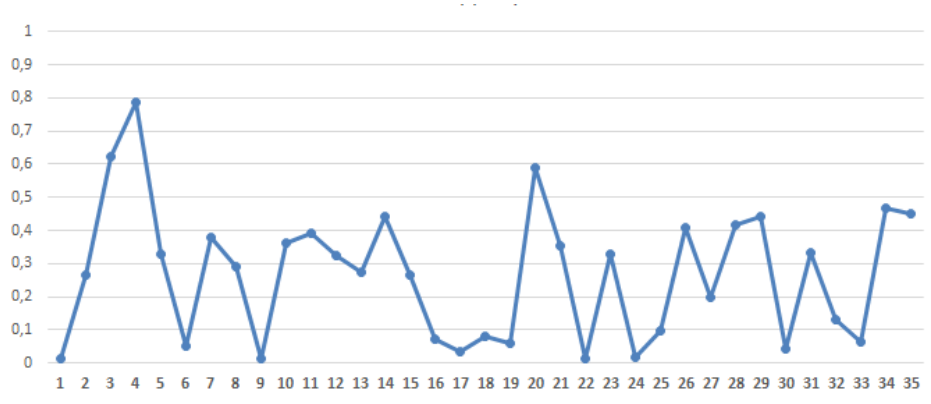

Рис. 2. Значения абсолютной разности между коэффициентами сходства, определенными преподавателем и моделью doc2vec

На графике по горизонтальной оси отложены номера ответов студентов, по вертикальной – значения абсолютной разности коэффициентов. В 26 случаях из 35 коэффициенты отличаются не более, чем на 0.4.

На следующем графике (Рис. 3) представлены значения абсолютной разности между коэффициентами сходства, определенными преподавателем и моделью (предобученная модель BERT).

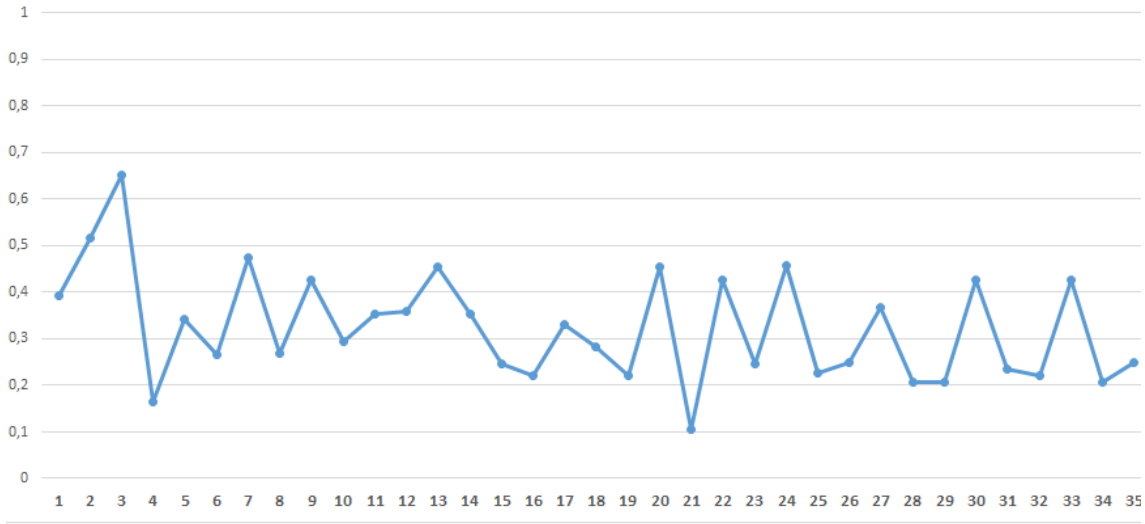

Рис. 3. Значения абсолютной разности между коэффициентами сходства, определенными преподавателем и моделью BERT

Из 35 значений 31 (88,6 %) принадлежат промежутку (0.2, 0.5), 25 значений (71,4 %) – промежутку (0, 0.4).

Далее модель BERT было дообучена на текстах по компьютерной графике. Проанализированы ответы студентов на вопрос «Какой вектор называется вектором нормали плоскости (нормальным вектором)?». Сравнивались ответы студентов с корректными ответами, также сравнивались между собой ответы, данные разными студентами. По графику следует, что в 15 случаях из 21 (71,4 %) коэффициенты, определенные преподавателем и моделью, отличаются не более, чем на 0.15 (Рис. 4). Среди ответов, оценка сходства которых преподавателем и моделью отличается более, чем на 0.5, присутствует следующий: «нулевой вектор, принадлежащий прямой, перпендикулярной рассматриваемой плоскости». Ответ неверный, по оценке модели коэффициент сходства равен 0.8093. Студент ошибся, написав слово «нулевой» вместо слова «ненулевой», что изменило смысл фразы.

Ответ «прямолинейный вектор» является неверным, по оценке модели коэффициент сходства с верным ответом равен 0.6753. После опроса было выяснено, что студент верно представлял вектор нормали, но не смог корректно сформулировать выражение. В этом случае необходим описанный выше этап проверки наличия слов, являющихся ключевыми для ответа на данный вопрос.

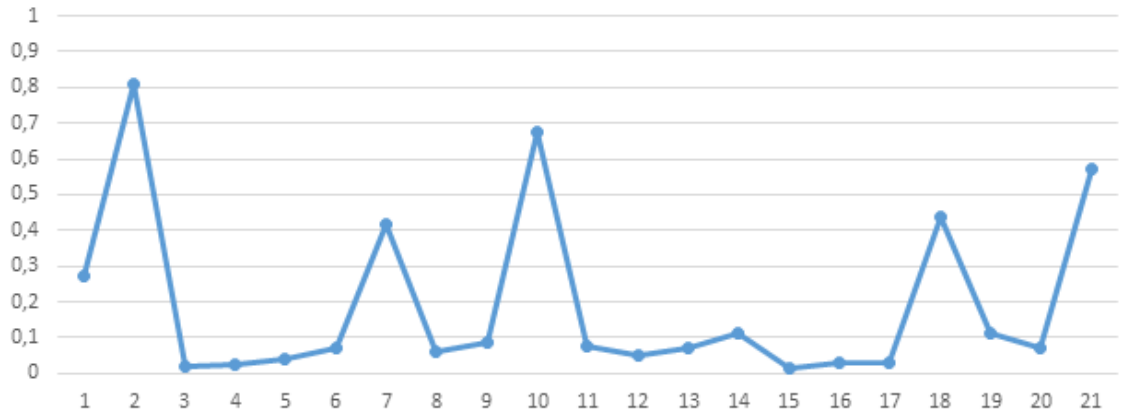

Рис. 4. Значения абсолютной разности между коэффициентами сходства, определенными преподавателем и дообученной на текстах по компьютерной графике моделью BERT

Ниже приведен график, демонстрирующий коэффициенты сходства ответов на вопрос «Какой вектор называется вектором нормали плоскости (нормальным вектором)?», определенные преподавателем и моделью (Рис. 5).

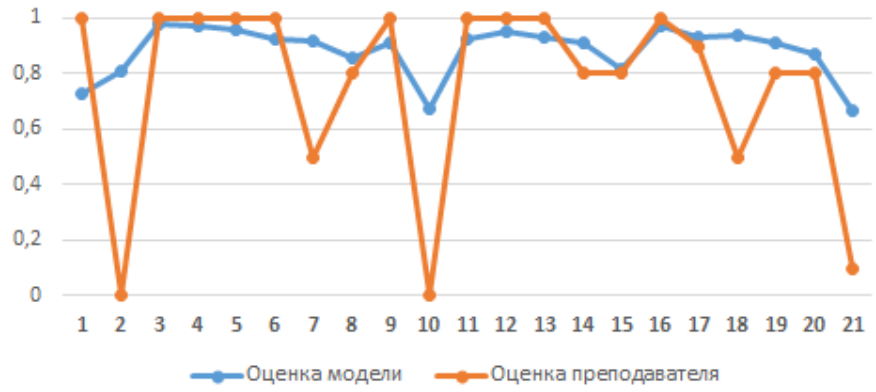

Рис. 5. Значения коэффициентов сходства, определенные преподавателем и дообученной на текстах по компьютерной графике моделью BERT

Пример бинарной классификации для оценки качества модели. В бинарной классификации решается вопрос о принадлежности объекта к одному из двух классов, в рассматриваемом случае это верные и неверные ответы. Например,

обучающиеся должны были дать определение бинаризации изображения. Примеры ответов студентов (приведены без обработки, в том виде, в котором они получены): «Разделение его на 2 чего-то» (неверный); «Перевод изображения в чёрно-белое изображение. Так как бинарность – это представление чего-либо в двоичной системе. Для изображения это чёрные и белые цвета» (верный); «Преобразование исходного изображения в градациях серого в бинарное изображение, элементы которого могут принимать только два значения» (верный).

Первоначально все ответы были оценены преподавателем, рассматривались оценки «верный ответ» и «неверный ответ» (объекты классов X и Y). Классы x и y предсказаны классификатором. В каждом случае выявлялось, совпали или нет предсказанные и реальные классы.

Таблица 3. Структура матрицы ошибок (Confusion Matrix)

	X	Y
x	TP (True Positive)	FP (False Positive)
y	FN (False Negative)	TN (True Negative)

Для оценки качества моделей используются следующие метрики. Precision – это доля объектов, определенных классификатором к данному классу и действительно принадлежащие ему:

$$Precision = \frac{TP}{TP+FP};$$

Recall (Полнота) показывает долю объектов класса определенных классификатор:

$$Recall = \frac{TP}{TP+FN};$$

*F*1 (*F*1-Score) – среднее гармоническое величин Precision и Recall:

$$F1 = \frac{2}{1/Precision+1/Recall} = \frac{2 \cdot Precision \cdot Recall}{Precision+Recall}.$$

По значению Accuracy оценивается точность прогноза, доля верно классифицированных объектов:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}.$$

Для бинарной классификации с помощью word2vec была использована модель случайного леса из библиотеки scikit-learn (язык программирования Python). В этом случае в качестве вектора предложения рассмотрен вектор, являющийся средним векторов, входящих в них слов. Было получено значение Accuracy, равное 0.67, остальные метрики приведены в таблице 5.

Таблица 5. Метрики качества проведенной бинарной классификации.

	Precision	Recall	F1
X1 (верные ответы)	0.60	0.50	0.55
X0 (неверные ответы)	0.70	0.78	0.74

Метрики качества отличаются в зависимости от формулировки вопроса и полученных ответов. Для части вопросов они лучше, чем значения, приведенные в таблице 5. Например, значение Accuracy равнялось 0.74, 0.82, 0.87. Необходимо отметить, что в примере выше были проанализированы ответы 32 студентов, для более точных выводов планируется провести опросы в других группах обучающихся и повторить эксперимент с аугментацией данных.

ЗАКЛЮЧЕНИЕ

Представлен метод предварительной оценки ответов, сформулированных в свободной форме. Для векторизации слов и предложений использовались модели word2vec, doc2vec, BERT. Сходство ответа обучаемого и корректного ответа определялось при помощи косинусной меры. Описан алгоритм оценивания, который состоит из этапов вычисления значения косинуса между векторами ответов и проверки наличия слов, являющихся ключевыми для ответа на данный вопрос. Оценено качество проведенной бинарной классификации.

В дальнейшем планируются расширение множества вопросов, проверка метода с другими группами студентов, а также разработка приложения, позволяющего преподавателю проводить предварительное оценивание ответов.

Благодарности

Работа выполнена за счет средств Программы стратегического академического лидерства Казанского (Приволжского) федерального университета («ПРИОРИТЕТ-2030»)

СПИСОК ЛИТЕРАТУРЫ

1. *Сидоренко А.С.* Оптимизация компьютерного тестирования студентов: минимизация влияния на ответы помощи интернета // Информатика и образование. 2020. № 5(314). С. 50–55.

2. *Миннегалиева Ч.Б., Мухамедшин Д.Р., Русецкий К.В., Паркалов А.В.* Некоторые вопросы автоматизации контроля знаний // Компьютерные инструменты в образовании. 2014. № 6. С. 52–59.

3. *Харламенко И.В., Воног В.В.* Обратная связь как форма контроля в техногенной образовательной среде // Информатика и образование. 2020. № 5(314). С. 44–49.

4. *Веремчук С.Э., Гурин Н.И.* Система тестирования знаний на естественном языке на основе семантической сети обучающей системы // Труды БГТУ. Серия 3: Физико-математические науки и информатика. 2019. № 1(218). С. 51–56.

5. *Xu S., Xu G., Jia P., Ding W., Wu Z., Liu, Z.* Automatic Task Requirements Writing Evaluation via Machine Reading Comprehension // In: Roll I., McNamara D., Sosnovsky S., Luckin R., Dimitrova V. (Eds.) Artificial Intelligence in Education. AIED 2021. Utrecht, The Netherlands, June 14–18, 2021. Lecture Notes in Computer Science, Vol. 12748. Springer, Cham, 2021. P.446–458.

6. *Милов В.Р., Дубов М.С., Калинина Н.А., Салтыкова А.А.* Интеллектуализация тестирования с открытыми вопросами на основе определения семантической близости фраз // Интеллектуальные системы в науке и технике. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сборник статей по материалам Международной конференции «Интеллектуальные системы в науке и технике» и Шестой всероссийской научно-практической конференции «Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века» (г. Пермь, 12–18 октября 2020 г.). 2020. С. 112–115.

7. *Wulff P., Mientus L., Nowak A. et al.* Utilizing a Pretrained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections // International Journal of Artificial Intelligence in Education. 2022.

8. *Белов С.Д., Зрелова Д.П., Зрелов П.В., Кореньков В.В.* Обзор методов автоматической обработки текстов на естественном языке // Системный анализ в науке и образовании. 2020. № 3. С. 8–22.

9. *Mikolov T., Sutskever I., Chen K., Corrado G.S., Dean J.* Distributed representations of words and phrases and their compositionality // Advances in Neural Information Processing Systems. 2013. P. 3111–3119.

10. *Кузнецов С.А., Вильнин А.Д.* Сравнение методов оценки семантического сходства на основе doc2vec и tf-idf // Электронные средства и системы управления. Материалы докладов Международной научно-практической конференции. 2021. № 1–2. С. 166–168.

11. *Ярушкина Н.Г., Мошкин В.С., Константинов А.А.* Применение языковых моделей word2vec и bert в задаче сентимент-анализа текстовых сообщений социальных сетей // Автоматизация процессов управления. 2020. № 3(61). С. 60–69.

12. *Kutuzov A., Kuzmenko E.* WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models // In: Ignatov D. et al. (Eds.) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science, Springer, Cham. 2017. V. 661. P. 155–161.

METHOD OF PRE-ASSESSMENT OF STUDENTS' ANSWERS BASED ON THE VECTOR MODEL OF DOCUMENTS

Chulpan Minnegalieva[1[0000-0003-4648-1623]], Gulshat Sabitova[2[0009-0008-1840-4819]], Almaz Gayaliev [3[0009-0006-9735-675X]]

[1,2,3] *Institute of Computational Mathematics and Information Technologies, Kazan Federal University, Kazan*

[1] mchulpan@gmail.com, [2] gulshat.9922@gmail.com, [3] gayaliev.almaz@mail.ru

Abstract

This article discusses the application of vector models for the preliminary analysis of students' free-form answers. Vector representations of words and documents were obtained using word2vec, doc2vec, and BERT models. The similarity between the answer given by the student and the correct answer was determined using the cosine measure. It was found that vector models allow identifying obviously incorrect answers with sufficient accuracy. For answers that are close in wording, an additional verifica-

tion step is proposed. Using word2vec, binary classification of answers to certain questions was performed, and accuracy, precision, recall and F1-measure estimates were given.

Keywords: *vector model, word2vec, doc2vec, BERT, cosine similarity, vector representation*

REFERENCES

1. *Sidorenko A.S.* Optimizaciya komp'yuternogo testirovaniya studentov: minimizaciya vliyaniya na otvety pomoshchi interneta // Informatika i obrazovanie. 2020. No. 5(314). S. 50–55.

2. *Minnegalieva C.B., Muhamedshin D.R., Ruseckij K.V., Parkalov A.V.* Nekotorye voprosy avtomatizacii kontrolya znanij // Komp'yuternye instrumenty v obrazovanii. 2014. No. 6. S. 52–59.

3. *Harlamenko I.V., Vonog V.V.* Obratnaya svyaz' kak forma kontrolya v tekhnogennoj obrazovatel'noj srede // Informatika i obrazovanie. 2020. No. 5 (314). S. 44–49.

4. *Veremchuk S.E., Gurin N.I.* Sistema testirovaniya znanij na estestvennom yazyke na osnove semanticheskoj seti obuchayushchej sistemy // Trudy BGTU. Seriya 3: Fiziko-matematicheskie nauki i informatika. 2019. № 1 (218). S. 51–56.

5. *Xu S., Xu G., Jia P., Ding W., Wu Z., Liu, Z.* Automatic Task Requirements Writing Evaluation via Machine Reading Comprehension // In: Roll I., McNamara D., Sosnovsky S., Luckin R., Dimitrova V. (Eds.) Artificial Intelligence in Education. AIED 2021. Utrecht, The Netherlands, June 14–18, 2021. Lecture Notes in Computer Science. V. 12748. Springer, Cham, 2021. P.446–458.

6. *Milov V.R., Dubov M.S., Kalinina N.A., Saltykova A.A.* Intellektualizaciya testirovaniya s otkrytymi voprosami na osnove opredeleniya semanticheskoj blizosti fraz // Intellektual'nye sistemy v nauke i tekhnike. Iskusstvennyj intellekt v reshenii aktual'nyh social'nyh i ekonomicheskih problem HKHI veka: sbornik statej po materialam Mezhdunarodnoj konferencii «Intellektual'nye sistemy v nauke i tekhnike» (g. Perm', 12–18 oktyabrya 2020 g.). 2020. S. 112–115.

7. Wulff P., Mientus L., Nowak A. et al. Utilizing a Pretrained Language Model (BERT) to Classify Preservice Physics Teachers' Written Reflections // International Journal of Artificial Intelligence in Education. 2022.

8. *Belov S.D., Zrelova D.P., Zrelov P.V., Koren'kov V.V.* Obzor metodov avtomaticheskoj obrabotki tekstov na estestvennom yazyke // Sistemnyj analiz v nauke i obrazovanii. 2020. No. 3. S. 8–22.

9. *Mikolov T., Sutskever I., Chen K., Corrado G. S. and Dean J.* Distributed representations of words and phrases and their compositionality // Advances in neural information processing systems. 2013. P. 3111–3119.

10. *Kuznecov S.A., Vil'nin A.D.* Sravnenie metodov ocenki semanticheskogo skhodstva na osnove doc2vec i tf-idf // Elektronnye sredstva i sistemy upravleniya. Materialy dokladov Mezhdunarodnoj nauchno-prakticheskoj konferencii. 2021. No. 1–2. S. 166–168.

11. *Yarushkina N.G., Moshkin V.S., Konstantinov A.A.* Primenenie yazykovyh modelej word2vec i bert v zadache sentiment-analiza tekstovyh soobshchenij social'nyh setej // Avtomatizaciya processov upravleniya. 2020. No. 3 (61). S. 60–69.

12. *Kutuzov A., Kuzmenko E.* WebVectors: A Toolkit for Building Web Interfaces for Vector Semantic Models // In: Ignatov D. et al. (Eds.) Analysis of Images, Social Networks and Texts. AIST 2016. Communications in Computer and Information Science, Springer, Cham. 2017. V. 661. P. 155–161.

СВЕДЕНИЯ ОБ АВТОРАХ

МИННЕГАЛИЕВА Чулпан Бакиевна – к. п. н., доцент кафедры информационных систем Института вычислительной математики и информационных технологий, Казанский федеральный университет, г. Казань.

Chulpan Bakievna MINNEGALIEVA – associate professor, Institute of Computational Mathematics and Information Technologies, Kazan Federal University, Kazan.

email: mchulpan@gmail.com
ORCID: 0000-0003-4648-1623

САБИТОВА Гульшат Альфисовна – студентка 2 курса магистратуры, Институт вычислительной математики и информационных технологий, Казанский федеральный университет, г. Казань.

Gulshat Alfisovna SABITOVA – 2nd year master's student, Institute of Computational Mathematics and Information Technologies, Kazan Federal University, Kazan.

email: gulshat.9922@gmail.com
ORCID: 0009-0008-1840-4819

ГАЯЛИЕВ Алмаз Маратович – студент 4 курса, Институт вычислительной математики и информационных технологий, Казанский федеральный университет, г. Казань.

Almaz Maratovich GAYALIEV – 4th year student, Institute of Computational Mathematics and Information Technologies, Kazan Federal University, Kazan.

email: gayaliev.almaz@mail.ru
ORCID: 0009-0006-9735-675X

Материал поступил в редакцию 23 марта 2023 года