

Научная статья

УДК 004.6:004.8

DOI 10.35266/1999-7604-2023-3-17-22

ПРИМЕНЕНИЕ АЛГОРИТМОВ ФОРМИРОВАНИЯ ВОПРОСОВ ПО ТЕКСТУ ДЛЯ АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ТЕСТОВ

Алексей Александрович Васильев^{1✉}, Андрей Сергеевич Нестеров²

^{1, 2} Сибирский государственный университет телекоммуникаций и информатики,

Новосибирск, Россия

¹ systems1@yandex.ru ✉

² nesterov@biik.ru

Аннотация. Представлены результаты исследования ручных, полуавтоматизированных и автоматизированных способов генерации контрольных вопросов, основанных на методах аннотирования, выделения ключевых слов и использования обучающего набора данных (датасета) для формирования тестов к учебному материалу, с описанием алгоритма работы каждого метода, примеров сгенерированных вопросов и оценкой их качества. На тестовых примерах представлены преимущества алгоритма генерации метода с использованием датасета и комбинированных методов и их возможного применения на практике.

Ключевые слова: генерация вопросов по тексту, нейронные сети, сравнение методов генерации

Финансирование: работа выполнена в рамках Государственного задания № 071-03-2023-001.

Для цитирования: Васильев А. А., Нестеров А. С. Применение алгоритмов формирования вопросов по тексту для автоматической генерации тестов // Вестник кибернетики. 2023. Т. 22, № 3. С. 17–22. DOI 10.35266/1999-7604-2023-3-17-22.

Original article

APPLYING TEXT QUESTIONS GENERATION ALGORITHMS FOR AUTOMATIC TEST GENERATION

Aleksey A. Vasilyev^{1✉}, Andrey S. Nesterov²

^{1, 2} Siberian State University of Telecommunications and Information Science, Novosibirsk, Russia

¹ systems1@yandex.ru ✉

² nesterov@biik.ru

Abstract. The article presents findings of manual, semi-automatic, and automatic approaches to generate test questions based on such methods as annotation, keyword extraction, and learning datasets for compiling tests for studying material, along with a description of each method algorithm, examples of generated questions, and their quality assessment. These examples demonstrate the advantages of an algorithm for generating a method using a dataset and a combination of methods, as well as their possible practical application.

Keywords: text questions generation, neural networks, comparison of methods of generation

Funding: the study is conducted in the framework of a State task No. 071-03-2023-001.

For citation: Vasilyev A. A., Nesterov A. S. Applying text questions generation algorithms for automatic test generation. *Proceedings in Cybernetics*. 2023;22(3):17–22. DOI 10.35266/1999-7604-2023-3-17-22.

ВВЕДЕНИЕ

Цифровая экономика требует квалифицированных специалистов и создания цифрового информационного пространства, в том числе в образовательной сфере. В статье рас-

сматриваются способы генерации вопросов для тестов на естественном русском языке, а также некоторые примеры реализаций решений этой задачи с помощью нейросетей, которые можно использовать для автоматизи-

рованной генерации контрольных материалов в виде тестов.

Широкое применение технологий дистанционного обучения поднимает проблему объективной, желательной автоматизированной оценки результатов освоения обучающимися образовательных программ в цифровой образовательной среде. Одним из возможных вариантов решения является использование тестов с большим количеством вопросов для

генерации индивидуальных вариантов. Высокая трудоемкость создания таких тестов существенно ограничивает их применение.

МАТЕРИАЛЫ И МЕТОДЫ

По степени участия человека способы генерации можно разделить на ручные, полуавтоматизированные и автоматизированные (табл. 1) [1, с. 86].

Таблица 1

Способы и методы генерации вопросов

Способы генерации	Методы генерации
Ручной	Параметризованные тесты Единый тест для всех учащихся
Полуавтоматизированный	Аннотирование текста Выделение ключевых слов
Автоматизированный	Применение обученных моделей Использование обученной нейронной сети (датасета)

Примечание: составлено по [1].

В работе исследуются алгоритмы [2, с. 594], позволяющие по тексту – учебному материалу, конспекту лекций и пр., сгенерировать вопросы для тестов [3, с. 7]. Из перечисленных выше методов в практических задачах были рассмотрены три метода генерации вопросов с наименьшей необходимостью ручной доработки результатов: аннотирование текста, выделение ключевых слов и использование датасета [4, с. 1033].

Метод «Выделение ключевых слов» основан на использовании библиотеки *Mystem* для выделения из текста ключевых слов (существительных и глаголов). Для формирования вопроса к найденным ключевым словам добавляется вопросительное слово, которое в свою очередь определяется с помощью искусственной нейронной сети (ИНС), обученной на выборке, содержащей пары «набор ключевых слов – вопросительное слово».

Метод «Выделение ключевых слов с предварительной аннотацией текста» позволяет уменьшить чувствительность предыдущего метода к формированию не соответствующих основному содержанию вопросов из-за наличия в исходном тексте слов, не несущих важной смысловой нагрузки. В данном методе исходный текст разбивается на логические

блоки, в качестве которых принимаются абзацы. Далее для каждого логического блока с помощью библиотеки *Transformers* и модели «*IlyaGusev/mbart_ru_sum_gazeta*» генерируется аннотация, для которой выполняется описанный ранее алгоритм на основе выделения ключевых слов [5, с. 56].

В методе «Использование датасета» формирование вопроса осуществляется без предварительной обработки текста непосредственно с помощью ИНС, реализованной с помощью библиотеки *Scikit-learn*. Составленная вручную обучающая выборка содержит пары «предложение – вопрос». Библиотека *Scikit-learn* обеспечивает использование ИНС типа MLP (Multilayer Perceptron – многослойный перцептрон). Для обработки текстовых данных в библиотеке используется метод оценки важности слов в контексте документа *TfidfVectorizer*, основанный на замене слов векторами в многомерном пространстве. Обучающие датасеты в работе были получены путем ручного формирования вопросов для каждого абзаца текстов, найденных в Интернете по заданной тематике [6, с. 317]. Объем датасета составлял 531 пару «абзац – вопрос».

Для сравнения рассматриваемых методов с их помощью было обработано около 70 тыс. символов текста по теме «устройство компьютера», разбитого на 196 абзацев. Для каждого абзаца текста автоматически были сформированы вопросы и проведена оценка возможности их использования в тестах без дополнительной ручной обработки путем присвоения им значений в интервале 0÷1 со следующей интерпретацией: 0 баллов – «неудовлетворительно, вопрос не может быть использован без значительной ручной доработки»; 0,5 балла – «удовлетворительно, вопрос может быть использован после незначительной ручной доработки»; 1 балл – «хо-

рошо, вопрос может быть использован без ручной доработки».

Для каждого из рассмотренных методов рассчитывались средние арифметические значения оценок сгенерированных вопросов по всей тестовой выборке в 196 абзацев. Полученные значения интерпретировались следующим образом: диапазон 0÷0,33 – «неудовлетворительно», 0,33÷0,66 – «удовлетворительно» и 0,66÷1 – «хорошо». В табл. 2 приведены примеры исходных текстовых данных и сгенерированные по ним вопросы.

В следующей табл. 3 приведены значения рассчитанных средних оценок качества генерации вопросов.

Таблица 2

Пример работы методов генерации вопросов по тексту

Выделение ключевых слов/оценка	Аннотирование текста/оценка	Использование датасета/оценка
<i>Исходный текст:</i> Персональный компьютер, ПК (англ. personal computer, PC), ПЭВМ (персональная электронно-вычислительная машина) – однопользовательская (предназначенная для использования одним пользователем) ЭВМ, имеющая эксплуатационные характеристики бытового прибора и универсальные функциональные возможности. ПК может быть стационарным (чаще всего настольным) или портативным (ноутбук, планшет)		
Как компьютер ПК англ.? – 0 Как ПК может быть всего ноутбук планшет? – 0,5	Как компьютер машина предназначенная использования пользователем имеющая характеристики прибора возможности? – 0	Что такое персональный компьютер? – 1
<i>Исходный текст:</i> Жесткий диск – один из самых популярных устройств массового хранения данных. Он позволяет хранить данные даже в том случае, если питание выключено		
Является ли диск устройств хранения данных? – 0,5 Позволяет ли он хранить данные случае питание выключено? – 0	Является ли диск устройств хранения данных? – 0,5	Является ли диск устройством хранения данных? – 1 Позволяет ли диск хранить данные в случае, если питание выключено? – 1
<i>Исходный текст:</i> Компьютер – устройство, которое выполняет логические операции и обработку данных, может использовать устройства ввода и вывода информации на дисплей и обычно включает в себя центральный процессор (CPU) для выполнения операций		
Компьютер устройство выполняет операции обработку данных может использовать устройства ввода вывода информации дисплей включает себя процессор выполнения операций? – 0	Термин компьютер устройство выполняет операции обработку? – 0	Сколько транзисторов содержит процессор? – 0,5
<i>Исходный текст:</i> Центральный процессор компьютера – это чип, покрытый миллиардами крошечных транзисторов, который встраивается в материнскую плату устройства		
Процессор компьютера чип покрытый миллиардами транзисторов встраивается плату устройства? – 0,5	Компьютере процессор чип покрытый миллиардами транзисторов? – 0	Сколько транзисторов содержит процессор? – 1
<i>Исходный текст:</i> Процессор работает в команде с другими компонентами компьютера: памятью, видеокартой и периферийными устройствами		
Процессор работает команде компонентами компьютера памятью видеокартой устройствами? – 0	Обзоре новинок? – 0	Как подключаются другие основные компоненты системы? – 1

Окончание табл. 2

Выделение ключевых слов/оценка	Аннотирование текста/оценка	Использование датасета/оценка
<i>Исходный текст:</i> Изначально термин центральное процессорное устройство описывал специализированную систему элементов, предназначенных для понимания и выполнения машинного кода компьютерных программ, а не только фиксированных логических операций. Начало применения термина и его аббревиатуры по отношению к компьютерным системам было положено в 1960-е гг. Устройство, архитектура и реализация процессоров с тех пор неоднократно менялись. В современных вычислительных системах все функции центрального процессора обычно выполняет одна микросхема высокой степени интеграции – микропроцессор		
Термин устройство описывал систему элементов предназначенных понимания выполнения кода программ фиксированных операций Начало применения термина аббревиатуры отношению системам было положено годы Устройство архитектура реализация процессоров пор менялись системах функции процессора выполняет микросхема степень интеграции микропроцессор? – 0	Чем управляет операционная система (ОС)? – 0	Чем управляет операционная система (ОС)? – 0
<i>Исходный текст:</i> Процессор Intel 80 386 появился в 1985 г. и привнес улучшенный защищенный режим, 32-битную адресацию, позволившую использовать до 4 Гб оперативной памяти и поддержку механизма виртуальной памяти. Эта линейка процессоров построена на регистровой вычислительной модели		
Процессор появился году привнес улучшенный защищённый режим адресацию позволившую использовать Гб памяти поддержку механизма памяти линейка процессоров построена модели? – 0	Расскажите о лазерных принтерах? – 0	Для чего используется основная память ПК? – 0,5
<i>Исходный текст:</i> Способность выполнения нескольких машинных инструкций за один такт процессора путем увеличения числа исполнительных устройств. Появление этой технологии привело к существенному увеличению производительности, в то же время существует определенный предел роста числа исполнительных устройств, при превышении которого производительность практически перестает расти, а исполнительные устройства простаивают. Частичным решением этой проблемы является, например, технология Hyper-threading		
Способность выполнения инструкций такт процессора увеличения числа устройств появление технологии привело увеличению производительности время существует предел роста числа устройств превышении производительность перестает расти устройства простаивают решением проблемы является технология? – 0	Назовите емкость компакт-дисков? – 0	Сохраняет ли оперативная память свое содержание в отсутствии мощности? – 0,5
<i>Исходный текст:</i> В октябре 2004 года Sun Microsystems выпустила двухъядерный процессор UltraSPARC IV, который состоял из двух модифицированных ядер UltraSPARC III. В начале 2005 г. был создан двухъядерный UltraSPARC IV+		
Октябре года выпустила процессор состоял ядер начале был создан? – 0	В каком году они перешли на x86-совместимые процессоры Intel? – 0,5	Из чего обычно изготавливается? – 0
<i>Исходный текст:</i> Изменение компоновки процессоров. Кристалл процессора переместился с внутренней на внешнюю сторону для лучшего отвода тепла к радиатору системы охлаждения		
Изменение компоновки процессоров Кристалл процессора переместился сторону отвода тепла радиатору системы охлаждения? – 0,5	Что может иметь чип процессора? – 0	Что может иметь чип процессора? – 0

Примечание: составлено авторами на основании данных, полученных в исследовании.

Таблица 3

Оценка качества генерации вопросов

Метод	Среднее значение оценки	Результат интерпретации
Выделение ключевых слов	0,1	неудовлетворительно
Аннотирование текста	0,15	неудовлетворительно
Использование датасета	0,341	удовлетворительно

Примечание: составлено авторами на основании данных, полученных в исследовании.

РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Из табл. 2–3 следует, что метод «Использование датасета» приводит к результату, требующему меньшей ручной доработки. В эксперименте этот метод использовался для генерации вопросов к существенному объему учебных данных (около 70 тыс. символов текста), для которых было сгенерировано 196 вопросов.

В табл. 2 можно встретить сгенерированные вопросы по темам, не содержащимся в исходных текстах. Сам вопрос выглядит корректно, и его появление связано не с ошибкой

алгоритма, а с недостаточным размером обучающей выборки.

Среднее арифметическое оценки качества вопросов составило 0,341, что по введенной шкале соответствует оценке «удовлетворительно». Полученный результат указывает, что метод «Использование датасета» хотя и может применяться на практике, все же требует заметной ручной фильтрации и доработки сгенерированных вопросов. Была проверена возможность повышения качества формируемых данным методом вопросов путем его комбинации с остальными методами (табл. 4).

Таблица 4

Результаты комбинации методов

Вариант	Среднее значение оценки качества
Напрямую по абзацу текста	0,341
Извлечение ключевых слов из абзаца текста	0,351
Аннотация по абзацу текста	0,220
Извлечение ключевых слов из аннотации	0,251

Примечание: составлено авторами на основании данных, полученных в исследовании.

Из данных табл. 4 следует, что на используемом учебном материале извлечение ключевых слов приводит к незначительному улучшению результатов. Невысокое среднее значение оценки можно объяснить недостаточным размером/качеством обучающей выборки [7, с. 61] либо наличием каких-либо других проблем ее формирования [8, с. 15].

ЗАКЛЮЧЕНИЕ

Из рассмотренных методов наилучшие, но все же требующие дополнительной доработки результаты показал метод «Ис-

пользование датасета». Комбинация методов не привела к существенному росту качества сгенерированных вопросов, но даже в текущем виде данные алгоритмы при условии ряда оговорок могут быть использованы для генерации вопросов. Поскольку проблема автоматизированного формирования контрольных материалов в образовательном процессе обладает очевидной привлекательностью и широкой перспективой использования, исследование подобных алгоритмов необходимо продолжать.

Список источников

1. Кручинин В. В., Кузовкин В. В. Обзор существующих методов автоматической генерации задач с условиями на естественном языке // Компьютерные инструменты в образовании. 2022. № 1. С. 85–96.
2. Посов И. А. Обзор генераторов и методов генерации учебных заданий // Образовательные технологии и общество. 2014. Т. 17, № 4. С. 593–609.
3. Куртасов А. М., Швецов А. Н. Метод автоматизированной генерации контрольно-тестовых заданий из текста учебных материалов // Вестник Череповецкого государственного университета. 2014. № 7. С. 7–11.
4. Тарасенко С. В., Рязанова Н. Ю. Анализ методов автоматической генерации вопросов на естественном языке // Инженерный вестник. 2015. № 12. С. 1032–1037.

References

1. Kruchinin V. V., Kuzovkin V. V. Overview of existing methods for automatic generation of tasks with conditions in natural language. *Kompiuternye instrumenty v obrazovanii*. 2022;(1):85–96. (In Russian).
2. Posov I. A. Obzor generatorov i metodov generatsii uchebnykh zadaniy. *Obrazovatelnye tekhnologii i obshchestvo*. 2014;17(4):593–609. (In Russian).
3. Kurtasov A. M., Shvetsov A. N. Metod avtomatizirovannoi generatsii kontrolno-testovykh zadaniy iz teksta uchebnykh materialov. *Cherepovets State University*. 2014;(7):7–11. (In Russian).
4. Tarasenko S. V., Ryazanova N. Yu. Analiz metodov avtomaticheskoi generatsii voprosov na estestvennom iazyke. *Engineering Bulletin*. 2015;(12):1032–1037. (In Russian).

5. Морозов Д. А., Глазкова А. В., Тютюльников М. А. и др. Генерация ключевых слов для аннотаций русскоязычных научных статей // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. 2023. Т. 21, № 1. С. 54–66.
6. Романенко В. В., Аксененко И. О. Обзор технологий автоматизированного создания и публикации тестовых заданий в системах дистанционного обучения // Информационные технологии в науке, управлении, социальной сфере и медицине : сб. науч. трудов IV Междунар. науч. конф., 5–8 декабря 2017 г., г. Томск. В 2 ч. Ч. 1. Томск : Изд-во ТПУ, 2017. С. 317–321.
7. Парасич А. В., Парасич В. А., Парасич И. В. Формирование обучающей выборки в задачах машинного обучения // Информационно-управляющие системы. 2021. № 4. С. 61–70. DOI 10.31799/1684-8853-2021-4-61-70.
8. Кафтаников И. Л., Парасич А. В. Проблемы формирования обучающей выборки в задачах машинного обучения // Вестник Южно-Уральского государственного университета. Серия: Компьютерные технологии, управление, радиоэлектроника. 2016. Т. 16, № 3. С. 15–24. DOI 10.14529/ctcr160302.
5. Morozov D. A., Glazkova A. V., Tyutyulnikov M. A. et al. Keyphrase generation for abstracts of the Russian-language scientific articles. *NSU Vestnik. Series: Linguistics and Intercultural Communication*. 2023;21(1):54–66. (In Russian).
6. Romanenko V. V., Aksenenko I. O. Review of technologies of automated creation and publication of test tasks in the distance learning systems. In: *Proceedings of the IV International Scientific Conference "Informatsionnye tekhnologii v nauke, upravlenii, sotsialnoi sfere i meditsine"*, December 5–8, 2017, Tomsk. In 2 parts. Pt. 1. Tomsk: Publishing House of the Tomsk Polytechnic University; 2017. p. 317–321. (In Russian).
7. Parasich A. V., Parasich V. A., Parasich I. V. Training set formation in machine learning tasks. *Survey. Information and Control Systems*. 2021;(4):61–70. DOI 10.31799/1684-8853-2021-4-61-70. (In Russian).
8. Kaftannikov I. L., Parasich A. V. Problems of training set's formation in machine learning tasks. *Bulletin of the South Ural State University. Ser. Computer Technologies, Automatic Control, Radio Electronics*. 2016;16(3):15–24. DOI 10.14529/ctcr160302. (In Russian).

Информация об авторах

А. А. Васильев – преподаватель.

А. С. Нестеров – кандидат технических наук, преподаватель.

Information about the authors

A. A. Vasilyev – Lecturer.

A. S. Nesterov – Candidate of Sciences (Engineering), Lecturer.