
КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА И ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА

УДК 004.91

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕШЕНИЯ ЗАДАЧИ КЛАССИФИКАЦИИ ДОКУМЕНТОВ НАУЧНО-ОБРАЗОВАТЕЛЬНОГО УЧРЕЖДЕНИЯ

М. Н. Краснянский, А. Д. Обухов, Е. М. Соломатина, А. А. Воякина

Тамбовский государственный технический университет

Поступила в редакцию 03.08.2018 г.

Аннотация. В данной статье рассматривается актуальная задача классификации документов с использованием методов машинного обучения в рамках предметной области научно-образовательного учреждения. Анализ разработок в данной области показал, что достаточной теоретической базы по интеграции существующих методов классификации для анализа документов научно-образовательного учреждения не разработано. Поэтому для решения поставленной задачи сформирован алгоритм классификации документов, учитывающий специфику документов рассматриваемой предметной области научно-образовательного учреждения. В статье рассматривается система признаков, используемая для решения задачи комбинированной классификации. Рассмотрен подход предварительной обработки текста, позволяющий при использовании известных методов машинного обучения повысить точность и быстродействие классификации документов.

Ключевые слова: машинное обучение, классификация документов, системы электронного документооборота, алгоритм предварительной обработки данных.

Annotation. This article discusses the actual problem of classification of documents using machine learning methods in the subject area of research and educational institutions. Analysis of developments in this area showed that there is no sufficient theoretical basis for the integration of existing classification methods for the analysis of documents of research and educational institutions. Therefore, to solve this problem, an algorithm of classification of documents, taking into account the specifics of the documents of the subject area of scientific and educational institutions. The article deals with the system of features used to solve the problem of combined classification. The paper considers the approach of preprocessing of the text, which allows using the known methods of machine learning to improve the accuracy and speed of document classification.

Keywords: machine learning, classification of documents, electronic document management system, data preprocessing algorithm.

ВВЕДЕНИЕ

Правильная организация работы с документами в наши дни имеет большое значение, так как от эффективности реализации документооборота напрямую зависит эффективность работы любой организации. Несмотря на то, что количество электронных документов постоянно растет, при управлении доку-

ментооборотом необходимо учитывать и бумажные, и преобразованные в электронную форму документы.

Системы электронного документооборота (СЭД) формируют новое поколение систем автоматизации предприятий. Основными объектами автоматизации в таких системах являются документы и процессы их движения и обработки. Данный подход к автоматизации предприятий является одновременно и конструктивным, и универсальным, обеспечивая автоматизацию документооборота и

© Краснянский М. Н., Обухов А. Д.,
Соломатина Е. М., Воякина А. А., 2018

всех бизнес-процессов предприятия в рамках единой концепции и единого программного инструментария.

В настоящее время актуальной задачей в области разработки СЭД является их интеграция с технологиями искусственного интеллекта и машинного обучения, что позволяет значительно повысить уровень автоматизации информационных систем, снизить влияние человеческого фактора на процессы управления информационными потоками. Одной из ключевых задач в этой области является классификация документов в соответствии с набором заданных признаков [1, 2]. Однако, в каждой предметной области существуют определенные особенности и ограничения, без учета которых достигнуть высоких показателей эффективности и точности классификации невозможно.

В рамках данной статьи будут рассмотрены основные признаки классификации документов исследуемой предметной области научно-образовательного учреждения и алгоритм, позволяющий решить задачу классификации с использованием технологий машинного обучения.

ЛИТЕРАТУРНЫЙ ОБЗОР

Существует множество методов классификации, которые используют различный математический аппарат и различные подходы при реализации [3–6]. Однако эффективность этих методов зависит от конкретной решаемой задачи. Несмотря на то, что последнее десятилетие коммерческие компании занимаются проблемой повышения качества машинного обучения, на сегодняшний день не существует методов, которые могли бы однозначно эффективно решить задачу классификации.

Можно выделить следующие типы методов классификации: вероятностные, метрические, логические, линейные, логическая регрессия. Обобщенно опишем некоторые из них, указывая преимущества и недостатки каждого из них.

Метод Байеса (Naive Bayes, NB) относится к вероятностным методам классификации

[7–11]. Преимущества метода состоит в следующем: высокая скорость работы, поддержка инкрементного обучения, простая реализация алгоритма в виде программы, легкая интерпретируемость результатов работы алгоритма. Несмотря на приведенные достоинства, метод Байеса имеет так же и минусы в своей реализации. Относительно низкое качество классификации и неспособность учитывать зависимость результата классификации от сочетания признаков являются главными недостатками этого метода.

Метод k ближайших соседей (k Nearest Neighbors, KNN) относится к метрическим методам и считается простейшим классификатором [12–14]. Объект присваивается тому классу, который является наиболее распространенным среди соседей данного элемента. Достоинства данного метода: простая реализация, проработанная теоретическая база, адаптация под нужную задачу выбором метрики или ядра, интерпретируемость. К недостаткам относятся: недостаточная производительность в реальных задачах, так как число соседей, используемых для классификации, будет достаточно большим; трудность в наборе подходящих весов и определением, какие признаки необходимы для классификации; зависимость от выбранной метрики расстояния между примерами.

Метод опорных векторов (Support Vector Machine, SVM) является линейным методом классификации, в настоящее время признан одним из лучших [15, 16]. Потенциальные недостатки метода опорных векторов заключается в следующем: невозможность калибровки вероятности попадания в определенный класс, подходит только для решения задач с 2 классами, параметры модели сложно интерпретировать.

Метод деревьев решений (Decision Trees, DT) относится к логическим методам классификации [17–19]. Деревом решений называют ациклический граф, по которому производится классификация документов, описанных набором признаков. Каждый узел дерева содержит условие ветвления по одному из признаков. У каждого узла столько ветвлений, сколько значений имеет выбран-

ный признак. Главным преимуществом метода является высокая производительность обучения и прогнозирования, такие деревья решений можно легко визуализировать и интерпретировать.

Нейронные сети активно используются в связи с появлением больших объемов данных и больших вычислительных возможностей [20, 21]. Сети с прямой связью являются универсальным средством аппроксимации функций, что позволяет их использовать при решении задач классификации. Их эффективность достаточно высока, потому что они генерируют фактически большое число регрессионных моделей (которые используются в решении задач классификации статистическими методами). Однако, любой метод, основанный на нейронных сетях, никогда не даст классификатор нужного качества, если имеющийся набор примеров не будет достаточно полным для той задачи, с которой придется работать в системе.

Проанализировав перечисленные выше методы классификации [22, 23] и исходя из проведенного сравнения, отметим, что существенно правильного метода не существует, поэтому предлагается адаптация наиболее распространенных и эффективных методов машинного обучения для решения задачи классификации документов научно-образовательного учреждения при помощи оригинального алгоритма, учитывающей специфику предметной области.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЙ

Классификация документов – одна из задач информационного поиска, заключающаяся в определении документа к одной из нескольких категорий на основании его содержания. Классификация может осуществляться вручную либо автоматически с помощью набора правил или применением методов машинного обучения.

Следует отличать классификацию текстов от кластеризации. В кластеризации тексты также группируются по некоторым критери-

ям, но данные критерии и конечные категории заранее не известны.

Существуют три подхода к решению задачи классификации текстов: обучение с учителем, обучение без учителя и обучение с подкреплением. Одним из популярных подходов, представляющий наибольший интерес, является классификация на основе машинного обучения. В данном подходе производится обучение классификатора (системы наименований объектов, каждому из которых соответствует уникальный идентификатор) на наборе исходных обучающих данных в виде документов с заданными для них значениями категорий [24].

Под обучающими данными будем понимать некоторое количество образцов документов из каждого класса, позволяющее с достаточной точностью выделить конкретную категорию. Отметим, что в машинном обучении сохраняется необходимость ручной разметки (указания классов документов), что, тем не менее, является более простой задачей, чем написание системы правил. Приведем пример такой разметки, которая может производиться даже в процессе функционирования системы: в электронной почте может существовать возможность пометить письма как спам, тем самым формируя обучающее множество для классификатора – фильтра нежелательных сообщений. Таким образом, классификация текстов, основанная на машинном обучении, является примером обучения с учителем, где в роли учителя выступает человек, задающий набор классов и размечающий обучающее множество [25, 26].

Итак, несмотря на огромное количество методов, подходов и технологий машинного обучения, используемые для решения задачи классификации документов, особенности предметной области, в которой реализуются данные технологии, накладывают свою ограничения и требования. В рамках данной статьи мы рассматриваем применение машинного обучения для классификации документов научно-образовательного учреждения. Для решения поставленной задачи максимально эффективно необходимо разработать мето-

дику, в соответствие с которой будет осуществляться процесс обработки документов, машинного обучения и классификации (рис. 1). Представленный алгоритм формализует процесс классификации документов научно-образовательного учреждения и позволяет за счет декомпозиции исходной задачи на ряд более простых подзадач распараллелить и ускорить процесс классификации документов [27].

Рассмотрим каждый из этапов представленного алгоритма более подробно:

Анализ предметной области. Документы в научно-образовательном учреждении могут сильно отличаться по структуре и размеру в зависимости от направленности деятельности организации, в рамках которых они существуют. Документы могут быть как достаточно небольшими по размеру (служеб-

ные записки, выписки, счета), так и включать несколько сотен страниц (коллективные записки, дипломные проекты), а их структура может быть как жестко заданной, так и полностью произвольной. Данные особенности крайне негативно влияют на точность и скорость классификатора.

Анализ множества документов. На данном этапе осуществляется формализация структуры информационных потоков в виде наглядной математической модели документооборота. Представим данную математическую модель M в коротком виде [28]:

$$M(U, P, O) = S, \quad (1)$$

где $U = \{u_i | i = 1..nU\}$ – множество объектов документооборота – документов, nU – общее количество объектов;

$P = \{p_q | q = 1..nP\}$ – множество пользователей, nP – общее количество пользователей;



Рис. 1. Алгоритм классификации документов научно-образовательного учреждения на основе машинного обучения

$O = \{o_l \mid l = \overline{1..nO}\}$ – множество операций, выполняемых над документом, nO – общее их количество;

$S = (U, P, O)$ – структура документооборота организации.

Используя данную математическую модель, мы можем формализовать информационные потоки организации, а также описать внутреннюю структуру документов, которую представим в виде упорядоченных пар атрибутов документа и их значений:

$$u_i = \{(a_{ik}, d_{ik})\}, k = \overline{1..nA_i}, \quad (2)$$

где a_{ik} – атрибут, d_{ik} – его значение, nA_i – общее количество атрибутов.

Среди всех атрибутов для решения задачи классификации самую большую роль играет его содержание – текст.

Анализ методов классификации. Важной проблемой организации современного документооборота является то, что множество документов, созданных в процессе функционирования организации, нельзя однозначно и непротиворечиво классифицировать. Не существует общей, научно обоснованной, универсальной классификации. Однако, можно выделить множество признаков и на их основе сформулировать систему классификации. Прежде чем перейти непосредственно к перечислению этих признаков, введем такие понятия, как «иерархический метод» и «фасетный метод», являющиеся ключевыми в системе классификации [1, 5].

Для последовательного разделения множества объектов на группировки используется иерархический метод классификации, позволяющий получить иерархическую древовидную структуру в виде ветвящегося графа, в котором объекты разделяются по группам в соответствии со своими признаками. Для реализации методов машинного обучения более удобным и применимым является фасетный метод, т.к. каждому типу документов мы сможем сопоставить перечень признаков. Теория фасетного метода заключается в следующем: каждому объекту документооборота $u_i \in U$ соответствует некоторое подмножество признаков классификации f_j , на основе которых осуществляется их группировка по

категориям cls_{ij} , где i соответствует идентификатору документов, а j – номеру признака (рис. 2):

$$CLS(u_i, f_j) = cls_{ij}, \quad (3)$$

где CLS – функция-классификатор, ставящая в соответствии документу и признаку классификации конкретное значение категории. Каждому документу u_i должно соответствовать множество значений признаков $u_i \rightarrow \{cls_{ij}\}$.

Таким образом, для решения задачи классификации необходимо получить классификатор CLS для каждого признака f_j .

Выбор признака классификации и определения множества категорий. Сформулируем систему признаков f_j документов научно-образовательного учреждения, по которым их можно классифицировать [29]:

1. по наименованию;
2. по способу фиксации информации;
3. по степени сложности;
4. по степени гласности;
5. по юридической силе;
6. по срокам исполнения;
7. по структурной принадлежности;
8. по срокам хранения;
9. по степени обязательности;
10. по степени унификации;
11. по месту составления;
12. по происхождению;
13. по направлению деятельности организации;
14. по состоянию документа;
15. по архитектуре документа;
16. по категории исполнителей документа.

Используя представленную систему и комбинируя различные сочетания признаков, можно сформировать классификатор требуемой сложности.

Предварительная обработка данных. Машинное обучение является мощным и эффективным инструментом при реализации алгоритмов классификации, маршрутизации, обработки и поиска документов, однако, определяющее значение в этих процессах имеет качество исходных данных [30]. Именно поэтому проведение подготовки исходных документов, их предварительная обработка

позволяет значительно повысить точность результатов, получаемых в ходе применения машинного обучения. Разобьем этот этап на 6 шагов.

1. На вход поступает множество документов определенных форматов (txt, doc, pdf и т. д.) Выбирается библиотека программного кода в зависимости от формата исходного документа и осуществляется извлечение данных из документа в виде неформатированного текста.

2. Исходный текст документа разбивается на несколько частей в соответствии с набором граничных лексем, после чего каждая часть ранжируется в зависимости от положения в документе. Осуществляется фильтрация блоков по заданному пороговому значению.

3. Фильтрация текста по стоп-листу (короткие слова и знаки препинания, не несущие смысловой нагрузки для последующего анализа), что приводит к сокращению объема текста и повышению его смысловой ценности.

4. Лемматизация – процесс приведения слов к леммам, т. е. нормальным словесным формам. Для реализации лемматизации можно использовать библиотеку программного кода Python Stemming Snowball, позволяющую привести все русские и английские слова к нормальной форме. Полученный после выполнения лемматизации набор слов уже может использоваться для проведения машинного обучения и решения конкретных задач классификации, маршрутизации и т.д.

5. Индексация документов – это построение некоторой числовой модели текста, которая переводит текст в удобное для дальнейшей обработки представление. Для индексации могут использоваться модели мешка слов, N-грамм или Word 2VEC [16, 21].

6. Уменьшение размерности пространства признаков. Для эффективной работы классификатора часто прибегают к сокращению числа используемых признаков (терминов). За счет уменьшения размерности пространства терминов можно снизить эффект переобучения – явление, при котором классификатор ориентируется на случайные или ошибочные характеристики обучающих данных,

а не на важные и значимые. Для этого можно использовать метод TF-IDF [29]. Таким образом, этап предварительной обработки позволяет значительно сократить размерность решения задачи и точность классификации.

Выбор метода машинного обучения. После предварительной обработки переходим непосредственно к решению задачи классификации. Первый этап её решения – это выбор функции классификации CLS и признаков f_j , по которым документы будут распределены по категориям. Для этого необходимо провести эксперимент: на основе обработанных на прошлом этапе данных провести обучение и проверку точности классификации по выбранному набору методов машинного обучения. По полученным результатам эксперимента определяется наиболее эффективный в текущих условиях метод.

Реализация процесса обучения. Используя выбранные метод машинного обучения и набор признаков классификации, осуществляется обучение классификатора.

Формирование множества упорядоченных пар «документ-категория». После успешного обучения полученные пары «документ-категория» вносятся в общую базу данных системы электронного документооборота научно-образовательного учреждения. Полученная информация о категориях проанализированных документов и найденные параметры классификатора позволяют проводить классификацию новых документов уже без повторного обучения.

РЕЗУЛЬТАТЫ

Для апробации представленного в статье алгоритма была решена задача классификации учебно-методических документов научно-образовательного учреждения, в качестве признака классификации использовалось «Наименование документа» (т. е. его категория: служебная записка, заявление, рабочая программа, лекционный материал и т. д.), а для решения задачи использовались следующие методы машинного обучения:

- Полиномиальный Наивный Байес.
- Бернулли Наивный Байес.

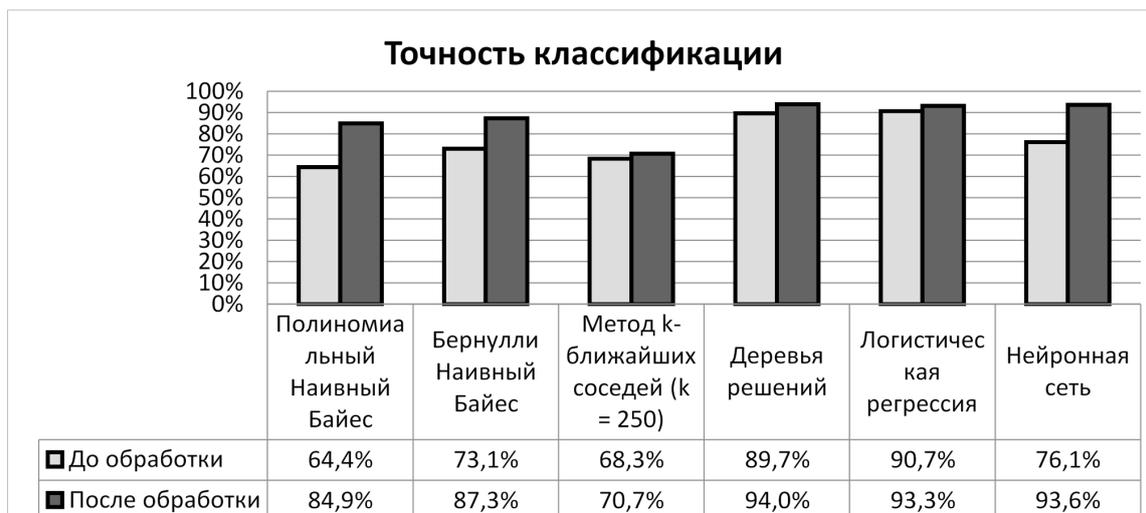


Рис. 2. Сравнение точности классификации документов при различных методах машинного обучения



Рис. 3. Сравнение времени обучение классификатора документов при различных методах машинного обучения

- Метод k -ближайших соседей ($k = 250$).
- Деревья решений.
- Логистическая регрессия.
- Нейронная сеть.

В итоге получены следующие экспериментальные данные, представленные на рис. 2 и 3. При обучении использовался набор из 3000 документов.

Полученные результаты подтверждают оправданность применения предложенного подхода к организации процесса машинного обучения для решения задачи классификации. Достигнуты положительные результаты как по времени обучения классификатора (вплоть до трехкратного улучшения показателя), так и по точности его работы (прирост от 5 до 20 %).

теля), так и по точности его работы (прирост от 5 до 20 %).

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В рамках данной статьи рассмотрено большое количество методов машинного обучения, используемых при решении задачи классификации документов [7–21]. Однако достаточной теоретической базы по интеграции используемых методов классификации для анализа документов научно-образовательного учреждения не существует, используются общие, универсальные подходы [22].

Для решения этой проблемы используется представленный алгоритм классификации документов. Его применение позволило улучшить показатели точности и времени обучения классификатора при анализе документов научно-образовательного учреждения за счет учета особенностей их структуры и предварительной обработки текста. Данный положительный эффект достигнут на большинстве проанализированных методов [7, 12, 15, 17, 20].

Также в данной статье приводится система признаков, по которым можно классифицировать документы. Комбинируя эти признаки, можно организовать сложную систему классификации документов [29]. Однако, представленный набор признаков возможно расширить и дополнить новыми в зависимости от особенностей выбранной предметной области и структуры анализируемых документов.

ЗАКЛЮЧЕНИЕ

В данной статье рассмотрена задача классификации документов в системе электронного документооборота научно-образовательного учреждения. Проведен сравнительный анализ существующих подходов к машинному обучению, на основе которого сделан вывод, что единственно верного и оптимального метода классификации документов не существует, необходимы испытания на конкретных наборах исходных данных.

Поэтому в рамках данной статьи для повышения качества классификации и снижения времени на обучение классификатора разработан алгоритм классификации документов, основанный на применении машинного обучения и учитывающий специфику документов научно-образовательного учреждения.

Для решения задачи классификации также необходимо выбрать определённые признаки, в соответствии с которыми будет распределяться исходный набор документов, для чего предлагается представленная в статье система признаков классификации документов научно-образовательного учреждения.

Рассмотрен подход предварительной обработки текста, позволяющий на известных методах машинного обучения получить значительное улучшение показателей точности и быстродействия.

Таким образом, представленное в статье алгоритмическое обеспечение может использоваться в качестве теоретической базы для интеграции методов машинного обучения при анализе и классификации документов научно-образовательного учреждения.

Работа выполнена при финансовой поддержке Министерства образования и науки РФ в рамках гранта Президента РФ МК-1666.2018.9.

СПИСОК ЛИТЕРАТУРЫ

1. Клячкин В. Н. и др. Применение методов машинного обучения при решении задач технической диагностики // Научный вестник УВАУ ГА (И). – 2016. – Т. 8. – С. 158–161.
2. Баракнин В. Б. и др. Автоматизированная классификация русских поэтических текстов по жанрам и стилям // Вестник Новосибирского государственного университета. Серия: Лингвистика и межкультурная коммуникация. – 2017. – Т. 15. – № 3. – С. 13–23.
3. Бабуцкий В. А., Сидоров И. Д. Методы и средства извлечения ключевых слов в задаче автоматической идентификации потенциально опасных текстов в условиях неопределённости их тематической принадлежности // Успехи современной науки. – 2017. – Т. 1. – № 12. – С. 54–59.
4. Yang W., Boyd-Graber J., Resnik P. A discriminative topic model using document network structure // Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). – 2016. – Т. 1. – С. 686–696.
5. Комарова А. В. и др. Метод автоматизированного извлечения адресов из неструктурированных текстов // International Journal of Open Information Technologies. – 2017. – Т. 5. – № 11. – С. 21–26.
6. Piernik M., Brzezinski D., Morzy T. Clustering XML documents by patterns // Knowl-

- edge and Information Systems. – 2016. – Т. 46. – № 1. – С. 185–212.
7. Tang B. et al. A Bayesian classification approach using class-specific features for text categorization // IEEE Transactions on Knowledge and Data Engineering. – 2016. – Т. 28. – № 6. – С. 1602–1606.
 8. Yoo J. Y., Yang D. Classification scheme of unstructured text document using TF-IDF and naive bayes classifier // Advanced Science and Technology Letters. – 2015. – Т. 3. – С. 263–266.
 9. Feng G. et al. Feature subset selection using naive Bayes for text classification // Pattern Recognition Letters. – 2015. – Т. 65. – С. 109–115.
 10. Wang S., Jiang L., Li C. Adapting naive Bayes tree for text classification // Knowledge and Information Systems. – 2015. – Т. 44. – № 1. – С. 77–89.
 11. Peng F., Schuurmans D. Combining naive Bayes and n-gram language models for text classification // European Conference on Information Retrieval. – Springer, Berlin, Heidelberg, 2003. – С. 335–350.
 12. Bijalwan V. et al. KNN based machine learning approach for text and document mining // International Journal of Database Theory and Application. – 2014. – Т. 7. – № 1. – С. 61–70.
 13. Adeniyi D. A., Wei Z., Yongquan Y. Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method // Applied Computing and Informatics. – 2016. – Т. 12. – № 1. – С. 90–108.
 14. Ingram S., Munzner T. Dimensionality reduction for documents with nearest neighbor queries // Neurocomputing. – 2015. – Т. 150. – С. 557–569.
 15. Jun S., Park S. S., Jang D. S. Document clustering method using dimension reduction and support vector clustering to overcome sparseness // Expert Systems with Applications. – 2014. – Т. 41. – № 7. – С. 3204–3212.
 16. Lilleberg J., Zhu Y., Zhang Y. Support vector machines and word2vec for text classification with semantic features // Cognitive Informatics & Cognitive Computing (ICCI+ CC), 2015 IEEE 14th International Conference on. – IEEE, 2015. – С. 136–140.
 17. Pliakos K., Geurts P., Vens C. Global multi-output decision trees for interaction prediction // Machine Learning. – 2018. – С. 1–25.
 18. Capannini G. et al. Quality versus efficiency in document scoring with learning-to-rank models // Information Processing & Management. – 2016. – Т. 52. – № 6. – С. 1161–1177.
 19. Lucchese C. et al. Quickscore: A fast algorithm to rank documents with additive ensembles of regression trees // Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. – ACM, 2015. – С. 73–82.
 20. Abadi M. et al. TensorFlow: A System for Large-Scale Machine Learning // OSDI. – 2016. – Т. 16. – С. 265–283.
 21. Dekhtyar A., Fong V. RE Data Challenge: Requirements Identification with Word2Vec and TensorFlow // Requirements Engineering Conference (RE), 2017 IEEE 25th International. – IEEE, 2017. – С. 484–489.
 22. Wu X. et al. Top 10 algorithms in data mining // Knowledge and information systems. – 2008. – Т. 14. – № 1. – С. 1–37.
 23. Jain A., Mandowara J. Text classification by combining text classifiers to improve the efficiency of classification // International Journal of Computer Application (2250-1797). – 2016. – Т. 6. – № 2.
 24. Canhasi E., Kononenko I. Multi-document summarization via archetypal analysis of the content-graph joint model // Knowledge and information systems. – 2014. – Т. 41. – № 3. – С. 821–842.
 25. Lan A. S. et al. Mathematical language processing: Automatic grading and feedback for open response mathematical questions // Proceedings of the Second (2015) ACM Conference on Learning@ Scale. – ACM, 2015. – С. 167–176.
 26. Gupta V. et al. A survey of text mining techniques and applications // Journal of emerging technologies in web intelligence. – 2009. – Т. 1. – № 1. – С. 60–76.
 27. Krasnyanskiy M. N., Ostroukh A. V., Karpushkin S. V., Obukhov A. D. Algorithm for Structural and Parametric Synthesis of Electronic Document Management System of Research and Education Institution. Journal of Applied Sciences. – 2016. – Vol. 16, No 7. – P. 332–337. – DOI: 10.3923/jas.2016
 28. Обухов А. Д. Алгоритм структурно-параметрического синтеза системы электронно-

го документооборота научно-образовательного учреждения // Вопросы современной науки и практики. Ун-т им. В. И. Вернадского. – 2016. – № 1(59). – С. 199–209.

29. *Батура Т. В.* Методы автоматической классификации текстов // Программные продукты и системы. – 2017. – Т. 30. – № 1.

Краснянский Михаил Николаевич – профессор, д-р техн. наук, ректор Тамбовского государственного технического университета.

E-mail: kras@tmabov.ru

Обухов Артем Дмитриевич – канд. техн. наук, старший преподаватель кафедры «Компьютерно-интегрированные системы в машиностроении» Тамбовского государственного технического университета.

E-mail: obuhov.art@gmail.com

Воякина Александра Алексеевна – студент кафедры «Системы автоматизированной поддержки принятия решений» Тамбовского государственного технического университета.

Соломатина Екатерина Михайловна – студент кафедры «Системы автоматизированной поддержки принятия решений» Тамбовского государственного технического университета.

30. *Обухов А. Д.* Постановка задачи структурно-параметрического синтеза системы электронного документооборота научно-образовательного учреждения // Вестник ТГТУ. – 2016. – № 2. – С. 217–232. – DOI: 10.17277/vestnik.2016.02.pp.217-232

Krasnyansky Mikhail Nikolaevich – Professor, doctor of technical Sciences, rector of Tambov State Technical University.

E-mail: kras@tmabov.ru

Obukhov Artem Dmitrievich – Candidate of technical Sciences, senior lecturer of the Department «Computer-integrated systems in mechanical engineering» of Tambov State Technical University.

E-mail: obuhov.art@gmail.com

Voyakina Alexandra Alekseevna – student of the Department of «Automated decision support systems» of Tambov State Technical University.

Solomatina Ekaterina Mikhailovna – student of the Department of «Automated decision support systems» Tambov State Technical University.