УДК 004.622 DOI: 10.26102/2310-6018/2021.32.1.010

Обработка текстов и подготовка моделей векторизации для программного комплекса классификации научных текстов

П.Ю. Гусев
Воронежский государственный технический университет
Воронеж, Российская Федерация

Резюме: Задача классификации научной специальности представляет собой сложный процесс, в котором, как правило, задействуется команда специалистов по определенному научному направлению. Одна из наиболее частых ситуаций, при которой возникает подобная задача, это определение научной специальности при защите диссертации. При решении подобной задачи можно использовать уже существующие научные тексты по специальностям. Наиболее показательным набором текстов по определенной специальности является набор авторефератов. Перед созданием интеллектуальной системы классификации научной специальности требуется обработка текстов авторефератов и их векторизация, которая обеспечит возможность обучения моделей. Разные способы обработки текстов оказывают разное влияние на конечный результат. В данной работе проведено сравнение разных способов подготовки текстов. При этом особенное внимание уделено возможности применения способов на разных по размеру наборах данных. Исследование способов подготовки текстов на малом наборе данных, а затем масштабирование этих же способов на большой набор данных обеспечит значительное сокращение затрачиваемого машинного времени на работу с текстами. В результате исследования установлена самая эффективная комбинация способов подготовки текстовых данных. Дальнейшая векторизация текстов возможна разными способами. В работе рассмотрена возможность векторизации методом TF-IDF. Для обеспечения наилучшего результата работы моделей машинного обучения проведены эксперименты по выбору оптимальных гиперпараметров векторизатора. В результате проведения экспериментов оценено влияние различных изменений гиперпараметров на конечный результат работы модели машинного обучения.

Ключевые слова: обработка текста, векторизация, программный комплекс, интеллектуальная система, моделирование.

Для цитирования: Гусев П.Ю. Обработка текстов и подготовка моделей векторизации для программного комплекса классификации научных текстов. *Моделирование, оптимизация и информационные технологии.* 2021;9(1). Доступно по: https://moitvivt.ru/ru/journal/pdf?id=912
DOI: 10.26102/2310-6018/2021.32.1.010

Word processing and preparation of vectorization models for a software package for the classification of scientific texts

P.Y. Gusev
Voronezh State Technical University
Voronezh, Russian Federation

Abstract: The task of classifying a scientific specialty is a complex process in which, as a rule, a team of specialists in a certain scientific direction is involved. One of the most common cases in which such a task arises is the definition of a scientific specialty when defending a dissertation. When solving such a problem, you can use existing scientific texts in specialties. The most indicative set of texts on a particular specialty is a set of abstracts. Before creating an intelligent classification system for a scientific specialty, it is necessary to process the texts of abstracts and their vectorization, which will provide the possibility of training models. Different types of word processing have different effects on the final result. This paper compares different methods of preparing texts. At the same time, special attention is paid to the possibility of using the methods on data sets of different sizes. Investigation of ways of

preparing texts on a small data set, and then scaling the same methods for a large data set will provide a significant reduction in the computer time spent on working with texts. As a result of the research, the most effective combination of methods for preparing text data has been established. Further vectorization of texts is possible in different ways. The paper considers the possibility of vectorization using the TF-IDF method. To ensure the best result of the machine learning models, experiments were carried out to select the optimal hyperparameters of the vectorizer. As a result of the experiments, the influence of various changes in hyperparameters on the final result of the machine learning model was evaluated.

Keyword: word processing, vectorization, software package, intelligent system, modeling.

For citation: Gusev P.Y. Word processing and preparation of vectorization models for a software package for the classification of scientific texts. *Modeling, Optimization and Information Technology*. 2021;9(1). Available from: https://moitvivt.ru/ru/journal/pdf?id=912 DOI: 10.26102/2310-6018/2021.32.1.010 (In Russ).

Введение

Задача определения научной специальности представляет собой трудоемкий процесс, в котором задействуются эксперты из определенных областей знаний. Подобная задача может возникать в следующих случаях:

- при определении принадлежности диссертации к одной из специальностей;
- для принятия решения о публикации научной статьи в специализированном журнале;
- при рецензировании сборников конференций, учебных пособий, монографий и т.д.

Наибольшую трудность при решении подобных задач испытывают молодые ученые, а также специалисты, которые переквалифицируются в смежные специальности. Определить разницу между смежными специальностями для неопытного человека без квалифицированной помощи зачастую практически невозможно. Получить консультацию эксперта тоже не всегда возможно, т.к. опытные эксперты не обладают достаточным количеством времени для консультирования всех желающих.

Один из способов решения проблемы определения научной специальности – разработка интеллектуальной системы, которая объединит уже наработанный квалифицированными экспертами опыт. Система в своей основе может использовать модель машинного обучения, натренированную на уже имеющихся данных.

Но, перед разработкой модели машинного обучения, требуется предварительная обработка данных и разработка векторизатора текста [1-2]. От выполнения этих этапов зависит успешность обучения модели машинного обучения [3]. В рамках данной статьи рассматриваются основные действия по подготовке текстовых данных и их влияние на результат обучения модели машинного обучения.

Вопросы подготовки текстовых данных неоднократно поднимались в ряде трудов [4-5]. Подготовка текстовых данных для разных областей знаний требует специфического подхода, что находит отражение в разнообразии трудов по данной тематике [6-7]. Построение различных моделей векторизации также рассматривалось в научных публикациях [8-9].

Однако поставленная задача имеет свои специфические особенности. Одной из ключевых особенностей задачи является набор данных, представляющий собой авторефераты диссертаций по различным научным специальностям. Таким образом, поставлена цель работы – повышение качества подготовки текстовых данных, используемых для построения моделей машинного обучения. При этом решались следующие задачи:

- выбор способов обработки текстов на естественных языках;
- выбор параметров, отвечающих за векторизацию текстов.

Исходные данные

В качестве используемого набора данных выступила база авторефератов по специальностям группы 05.13.00 - Информатика, вычислительная техника и управление. Всего, при разработке модели машинного обучения, использовано 11000 авторефератов по всем специальностям из исследуемой группы. Но, для проведения настоящего исследования, размер набора данных сокращен. Выполнено это для сокращения времени проведения экспериментов, а также для оценки влияния изменений в подготовке данных на разные наборы данных.

В рамках исследования рассмотрены 6 специальностей из группы 05.13.00. Количество авторефератов по каждой специальности представлено в Таблице 1.

Таблица 1 – Описание исходных данных
Table 1 - Description of the initial data

Специальность	Количество авторефератов	Год публикации первого автореферата
05.13.01	2200	2008
05.13.06	1499	2006
05.13.10	1684	1982
05.13.11	1498	1983
05.13.12	1083	1983
05.13.18	2600	2008

Для получения адекватных данных о подготовке текстов и параметрах моделей введено ограничение года публикации наиболее ранних авторефератов. Данное ограничение обеспечивает анализ наиболее актуальных научных работ.

Эксперименты с подготовкой текстовых данных и векторизацией текстов проводились на языке программирования Python с применением специализированных библиотек.

Обработка текстовых данных

Перед векторизацией требуется предварительная обработка текстовых данных. В некоторых случаях предварительная обработка данных приводит к уменьшению точности создаваемой в конечном итоге модели. Поэтому для каждого набора данных следует определять оптимальные пути работы с текстом.

Эксперименты по предварительной обработке текстов проводились на двух разных наборах данных. Первый набор данных представлял собой 4000 текстов авторефератов научных специальностей 05.13.01 и 05.13.18. На каждую специальность приходилось по 2000 текстов авторефератов. Второй набор данных включал в себя авторефераты специальностей 05.13.01, 05.13.06, 05.13.10, 05.13.11, 05.13.12, 05.13.18. Применение такого разбиения текстов авторефератов на 2 набора данных объясняется необходимостью проверки способов обработки текстов для дальнейшего масштабирования.

Масштабирование выбранных способов обработки текстов потребуется, когда разрабатываемая система классификации научных текстов будет расширена на другие группы специальностей. Применение одинаковых способов обработки текстов для разных по объему наборов данных также покажет возможность сокращения времени

выбора оптимальных способов обработки текстов. Это объясняется высокими затратами машинного времени на экспериментирование.

Для оценки влияния разных способов обработки текстов на результат решено использовать конечную модель машинного обучения, в которой будут использоваться векторизованные тексты. В качестве модели машинного обучения использована логистическая многоклассовая регрессия [10-11]. В качестве метрики использована точность (Accuracy) [12]. При этом модели векторизации строились с использованием стандартных настроек инструмента TfidfVectorizer из пакета sklearn [13].

Первый этап экспериментирования проводился на наборе данных из 4000 авторефератов. В первую очередь проверена точность модели машинного обучения на не подготовленных данных. Из текстов авторефератов, при этом, убран код специальности и название специальности. В противном случае модель машинного обучения настроится на классификацию текстов авторефератов по коду и названию специальностей. В результате тестирования модели неверно классифицировано 45 текстов, и точность составила 0.9427.

Дальнейшее исследование подготовки текстов предусматривало удаление слов, содержащих 3 и менее буквы. Как правило, такие слова не оказывают значительного влияния на научный смысл текста. В результате удаления слов с 3 и менее буквами моделью машинного обучения неверно классифицировано 43 тестовых текста, и точность составила 0.9452. Не смотря на незначительное изменение показателя, можно сделать вывод об увеличении точности.

Учитывая специфику научных текстов по группе специальностей 05.13.00 и наличие множества формул, решено оставить в текстах только буквы и цифры. Специальные символы и знаки в формулах могут мешать работе модели машинного обучения, т.к. модель, в первую очередь, настраивается на специфические слова и их объединения. В результате удаления всех символов и знаков кроме букв и цифр, модель машинного обучения неверно классифицировала 41 тестовый текст. Значение точности, при этом, составило 0.9478, что позволяет сделать вывод о действенном способе обработки текста на рассматриваемом наборе.

Дальнейшее исследование было направлено на проведение аналогичных экспериментов на расширенном наборе данных, содержащем тексты авторефератов по 6 научным специальностям. В первую очередь исследовано применение текстов без подготовки, убраны только код и названия специальностей. В результате проверки модели на тестовых данных неверно классифицировано 62 текста, а точность составила 0.8937. Ввиду того, что количество текстов в наборах данных различно и абсолютные показатели точности несопоставимы, проведено сравнение тенденции изменения точности.

Как и в случае с сокращенным набором данных проверено влияние исключения из текстов слов с количеством букв 3 и менее. При запуске модели машинного обучения на тестовом наборе данных неверно классифицировано 64 текста, точность составила 0.8902. Полученный результат говорит о том, что на расширенном наборе данных удаление коротких слов не принесло ожидаемого результата.

Исследование следующего способа подготовки текстов для векторизации решено проводить без удаления слов длиной менее 4 букв. В результате удаления из текстов расширенного набора данных всех символов кроме букв и цифр, проверка работы модели машинного обучения на тестовых данных показала точность 0.8971, неверно классифицировано 60 текстов.

Таким образом, результат работы машинного обучения улучшился. Но, для обеспечения полноты проведения эксперимента, решено проверить результат работы модели машинного обучения на тестовых данных при объединении способов подготовки

текстов. После удаления слов с длиной менее 4 букв и удаления всех символов кроме букв и цифр, при запуске модели машинного обучения на тестовых данных точность составила 0.8988, неверно классифицировано 59 текстов. Объединение двух способов подготовки текстов показало увеличение точности.

В Таблице 2 представлены результаты экспериментирования с разными способами подготовки текстов для векторизации.

Таблица 2 – Результаты экспериментов по подготовке текстовых данных
Table 2 - Experimental results for preparing text data

Способы подготовки текстов	Результат работы модели машинного обучения			
	4000 текстов по специальностям 05.13.01 и 05.13.18		Расширенный набор данных	
	Количество неверно классифицированных текстов	Точность (Accuracy)	Количество неверно классифицированных текстов	Точность (Accuracy)
Удалены код и название специальности	45	0.9427	62	0.8937
Удалены слова с длиной менее 4 букв	43	0.9452	60	0.8971
Удалены все символы, кроме букв и цифр	41	0.9478	59	0.8988

В результате проведения экспериментов по выбору способов подготовки текстовых данных для векторизации определена возможность масштабирования выбранных способов.

Выбор параметров векторизации текстов

Векторизация текстов при разработке моделей машинного обучения является одним из ключевых этапов, который в значительной степени влияет на результат. В случае неверно выбранного способа векторизации может как значительно уменьшится точность работы модели машинного обучения, так и могут значительно возрасти затраты машинного времени на процесс векторизации уже во время работы с новыми данными.

В данной работе в качестве метода векторизации использована мера TF-IDF. TF-IDF — статистическая мера, используемая для оценки важности слова в контексте документа, являющегося частью коллекции документов или корпуса [14].

TF – обозначает количество вхождений слова в конкретном рассматриваемом документе. IDF позволяет оценить значимость слова применительно ко всему набору рассматриваемых текстов. В результате произведения двух значений вычисляется статистическая мера TFIDF, которая широко применяется в качестве метода векторизации текстов. Применение именно такого способа векторизации текстов обусловлено возможностью его настройки с помощью гиперпараметров.

Для исходного набора данных произведен поиск оптимальных параметров применения сочетания слов при векторизации. При этом исследовано применение как отдельно биграмм, триграмм и т.д., так и сочетания разного количества слов. Применяемый инструмент TfidfVectorizer из пакета sklearn позволяет настраивать

количество анализируемых сочетаний слов с помощью параметров min_gram и max_gram. Параметр min_gram определяет минимальное количество подряд идущих слов, которые будут использованы при векторизации. Max_gram соответственно определяет максимальное количество подряд идущих слов.

Для экспериментов с параметрами min_gram и max_gram были выбраны минимальное и максимальное устанавливаемые значения равные 1 и 5 соответственно. В результате проведения эксперимента перебраны все комбинации этих значений. Для каждой комбинации, также как и при экспериментировании со способами подготовки текстов, производилось обучение модели машинного обучения и дальнейшее ее тестирование. В Таблице 3 представлены результаты экспериментов.

Таблица 3 – Результаты экспериментов по определению количества слов
Table 3 - Word Count Experiment Results

№ эксперимента	Исследуемые параметры		Результаты тестирования модели машинного обучения	
	min_gram	max_gram	Количество неверно классифицированных текстов	Точность (Accuracy)
1	1	1	24	0.8737
2	1	2	27	0.8579
3	1	3	29	0.8474
4	1	4	28	0.8526
5	1	5	28	0.8526
6	2	2	20	0.8947
7	2	3	22	0.8842
8	2	4	22	0.8842
9	2	5	22	0.8842
10	3	3	18	0.9053
11	3	4	18	0.9053
12	3	5	18	0.9053
13	4	4	21	0.8895
14	4	5	21	0.8895
15	5	5	28	0.8526

Как видно из Таблицы 3 – наилучший результат показал эксперимент №10. При этом эксперименты №11-12 имеют аналогичный результат, но т.к. в этих экспериментах применяется большее, в сравнении с 10 экспериментом, количество вариаций слов, то затраты машинного времени на векторизацию буду выше. Таким образом, оптимальным выбором сочетаний слов для исследуемого корпуса текстов являются триграммы.

В рамках настоящего исследования проведен эксперимент по определению оптимального количества токенов, которые будут использоваться при векторизации. Токен – это слово или сочетание слов, используемое при векторизации методом TF-IDF. Оптимальный размер токена для исследуемого набора данных, как показал предыдущий эксперимент – триграмм.

За определение количества токенов для инструмента TfidfVectorizer из пакета sklearn используется параметр max_features. Токены, используемые векторизатором, сортируются по частоте употребления в рассматриваемом корпусе текстов. Параметр max_features выбирает для обработки первые n сортированных токенов. Эксперимент по определению оптимального количества токенов проводился с использованием

триграммов. В Таблице 4 представлены результаты проведения экспериментов по определению оптимального количества токенов.

Таблица 4 – Результаты эксперимента по поиску оптимального количества токенов
Table 4 - Results of an experiment of searching the optimal number of tokens

№ эксперимента	Максимальное количество токенов	Результаты тестирования модели машинного обучения	
		Количество неверно классифицированных текстов	Точность (Accuracy)
1	500	14	0.9263
2	550	14	0.9263
3	600	14	0.9263
4	650	14	0.9263
5	700	13	0.9316
6	750	13	0.9316
7	800	12	0.9368
8	850	13	0.9316
9	900	12	0.9368
10	950	13	0.9316
11	1000	13	0.9316
12	1100	13	0.9316
13	1200	13	0.9316
14	1500	14	0.9263
15	2000	13	0.9316
16	3000	14	0.9263
17	5000	14	0.9263
18	10000	14	0.9263
19	50000	15	0.9211

По результатам эксперимента видно, что значительное влияние на результат оказывает сам факт ограничения количества токенов. При этом точное значение ограничения не оказывает решающего влияния. По представленным в Таблице 4 данным можно сделать вывод о том, что оптимальным количество токенов для используемого набора данных является 800.

Заключение

Настоящая работа посвящена выбору способов подготовки научных текстов для дальнейшей векторизации, а также вопросам подбора параметров векторизатора для дальнейшего обучения моделей машинного обучения. В результате проведения экспериментов по выбору способов подготовки текстов и проверки их масштабируемости на расширенные наборы данных установлено, что выбранные способы обработки одинаково работают при разных размерах наборов данных, но необходимо учитывать особенности комбинации способов обработки. Эксперименты по поиску оптимальных параметров векторизации позволили установить ключевые значения, обеспечивающие наивысшую точность работы модели машинного обучения.

Основными задачами перспективных исследований являются углубленное изучение возможностей применения комбинаций способов подготовки текстов для векторизации и расширение исследуемых параметров векторизатора и их комбинации.

ЛИТЕРАТУРА

1. Иванов Н.Н. Синтаксический разбор предложения для векторизации текста. *Вопросы науки и образования*. 2017;11(12):45-46.
2. Спивак А.И., Лапшин С.В., Лебедев И.С. Классификация коротких сообщений с использованием векторизации на основе elmo. *Известия Тульского государственного университета. Технические науки.* 2019;10:410-418.
3. Флах, П.. *Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных*. Litres, 2019.
4. Бородин А.И., Вейнберг Р.Р., Литвишко О.В. Методы обработки текста при создании чат-ботов. *Хуманитарни Балкански изследвания*. 2019;3(3(5)):108-111. DOI: 10.34671/sch.hbr.2019.0303.0026
5. Кайбасова Д.Ж. Извлечение статистических данных для определения уникальности документов на основе анализ контента учебных программ дисциплин. *The Scientific Heritage*. 2020;44-1(44):57-62.
6. Кротова О.С., Москалев И.В., Хворова Л.А., Назаркина О.М. Реализация эффективных моделей классификации медицинских данных методами интеллектуального анализа текстовой информации. *Известия Алтайского государственного университета*. 2020;1(111):99-104.
7. Исаченко В.В., Апанович З.В. Система анализа и визуализации для кросс-языковой идентификации авторов научных публикаций. *Вестник Новосибирского государственного университета. Серия: Информационные технологии*. 2018;16(2):49-61. DOI: 10.25205/1818-7900-2018-16-2-49-61
8. Жеребцова Ю.А., Чижик А.В. Создание чат-бота: обзор архитектур и векторных представлений текста. *International Journal of Open Information Technologies*. 2020;8(7):50-56.
9. Попова Е.П., Леоненко В.Н.. Прогнозирование реакции пользователей в социальных сетях методами машинного обучения. *Научно-технический вестник информационных технологий, механики и оптики*. 2020;20(1):118-124.
10. Udhayakumar S., Nancy J.S., UmaNandhini D., Ashwin P., Ganesh R. Context Aware Text Classification and Recommendation Model for Toxic Comments Using Logistic Regression. *Intelligence in Big Data Technologies—Beyond the Hype. Springer, Singapore.* 2021;209-217. DOI: 10.1007/978-981-15-5285-4_20.
11. De Cock M., Dowsley R., Nascimento A.C., Railsback D., Shen J., Todoki A. (2021). High performance logistic regression for privacy-preserving genome analysis. *BMC Medical Genomics*. 2021;14(1):1-18. DOI: 10.21203/rs.3.rs-26375/v1.
12. Kumar V., Subba B. (2020, February). A TfidfVectorizer and SVM based sentiment analysis framework for text data corpus. *2020 National Conference on Communications (NCC). IEEE.* 2020;1-6. DOI: 10.1109/ncc48643.2020.9056085.
13. Subba B., Gupta P. A tfidfvectorizer and singular value decomposition based host intrusion detection system framework for detecting anomalous system processes. *Computers & Security.* 2021;100. DOI: 10.1016/j.cose.2020.102084.
14. Абрамов П.С. Извлечение ключевой информации из текста. *Новые информационные технологии в автоматизированных системах*. 2018;21:217-219.

REFERENCES

1. Ivanov N.N. Sintaksicheskii razbor predlozheniya dlya vektorizatsii teksta. *Voprosy nauki i obrazovaniya.* 2017;11(12):45-46. (In Russ)
2. Spivak A.I., Lapshin S.V., Lebedev I.S. Klassifikatsiya korotkikh soobshchenii s ispol'zovaniem vektorizatsii na osnove elmo. *Izvestiya Tul'skogo gosudarstvennogo*

universiteta. Tekhnicheskie nauki. 2019;10:410-418. (In Russ)
3. Flach, P.. *Mashinnoe obuchenie. Nauka i iskusstvo postroeniya algoritmov, kotorye izvlekayut znaniya iz dannykh.* Litres, 2019. (In Russ)
4. Borodin A.I., Veinberg R.R., Litvishko O.V. Methods of text processing when creating chatbots. *Khumanitarni Balkanski izsledvaniya*. 2019;3(3(5)):108-111. DOI: 10.34671/sch.hbr.2019.0303.0026 (In Russ)
5. Kaibasova D.Zh. Izvlechenie statisticheskikh dannykh dlya opredeleniya unikal'nosti dokumentov na osnove analiz kontenta uchebnykh programm distsiplin. *The Scientific Heritage*. 2020;44-1(44):57-62 (In Russ)
6. Krotova O.S., Moskalev I.V., Khvorova L.A., Nazarkina O.M. Realizatsiya effektivnykh modelei klassifikatsii meditsinskikh dannykh metodami intellektual'nogo analiza tekstovoi informatsii. *Izvestiya Altaiskogo gosudarstvennogo universiteta.* 2020;1(111):99-104. (In Russ)
7. Isachenko V.V., Apanovich Z.V. Sistema analiza i vizualizatsii dlya kross-yazykovoi identifikatsii avtorov nauchnykh publikatsii. *Vestnik Novosibirskogo gosudarstvennogo universiteta. Seriya: Informatsionnye tekhnologii.* 2018;16(2):49-61. DOI: 10.25205/1818-7900-2018-16-2-49-61 (In Russ)
8. Zherebtsova Yu.A., Chizhik A.V. Sozdanie chat-bota: obzor arkhitektur i vektornykh predstavlenii teksta. *International Journal of Open Information Technologies.* 2020;8(7):50-56. (In Russ)
9. Popova E.P., Leonenko V.N.. Prognozirovanie reaktsii pol'zovatelei v sotsial'nykh setyakh metodami mashinnogo obucheniya. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics.* 2020;20(1):118-124. (In Russ)
10. Udhayakumar S., Nancy J.S., UmaNandhini D., Ashwin P., Ganesh R. Context Aware Text Classification and Recommendation Model for Toxic Comments Using Logistic Regression. *Intelligence in Big Data Technologies—Beyond the Hype. Springer, Singapore.* 2021;209-217. DOI: 10.1007/978-981-15-5285-4_20.
11. De Cock M., Dowsley R., Nascimento A.C., Railsback D., Shen J., Todoki A. (2021). High performance logistic regression for privacy-preserving genome analysis. *BMC Medical Genomics*. 2021;14(1):1-18. DOI: 10.21203/rs.3.rs-26375/v1.
12. Kumar V., Subba B. (2020, February). A TfidfVectorizer and SVM based sentiment analysis framework for text data corpus. *2020 National Conference on Communications (NCC). IEEE.* 2020;1-6. DOI: 10.1109/ncc48643.2020.9056085.
13. Subba B., Gupta P. A tfidfvectorizer and singular value decomposition based host intrusion detection system framework for detecting anomalous system processes. *Computers & Security.* 2021;100. DOI: 10.1016/j.cose.2020.102084.
14. Abramov P.S. Izvlechenie klyuchevoi informatsii iz teksta. *Novye informatsionnye tekhnologii v avtomatizirovannykh sistemakh.* 2018;21:217-219. (In Russ)

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Гусев Павел Юрьевич, кандидат технических наук, доцент Воронежского государственного технического университета, Воронеж, Российская Федерация.
e-mail: gusevpvl@gmail.com
ORCID: 0000-0002-3752-0152

Pavel Y. Gusev, Ph. D., Associate Professor, Voronezh State Technical University, Voronezh, Russia.