

А. В. Романов, канд. техн. наук, ст. преп., e-mail: a.romanov@yandex.ru,  
 Национальный исследовательский университет "Высшая школа экономики"  
 К. Е. Довгань, студент, e-mail: ke.dovgan@yandex.ru  
 Е. С. Козлова, студентка, e-mail: es.kozlova@yandex.ru

**Применение методов машинного обучения  
 для решения задачи автоматической рубрикации статей по УДК**

*Выявлены особенности применения современных методов машинного обучения в задаче автоматической рубрикации статей УДК научными статьями. В качестве классификаторов рассмотрены линейный метод, метод максимальных правдоподобия, метод опорных векторов, метод опорных машин, метод опорных векторов и др. Рассмотрены проблемы автоматической рубрикации статей по УДК.*

**Ключевые слова:** классификация текстов, машинное обучение, искусственный интеллект, нейронные сети, метод опорных векторов, УДК

**Введение**

На современном этапе задача обработки текстов на естественных языках является одной из ведущих. Она напрямую связана с развитием компьютерной обработки текстовой информации в сети Интернет. Научные статьи, книги, рукописи, журналы, справочники — все это многообразие текстов в бумажном и электронном виде. Вместе с тем возникает проблема презентации в таком большом объеме информации, поиска нужной информации по конкретному вопросу становится все более трудным. В этой связи использование автоматических классификаторов является одним из перспективных способов решения данной проблемы [1].

Задача классификации текстов возникает в различных областях: например, при обработке электронной корреспонденции (фильтры спама), поиске информации, контекстной рекламе, в системах документооборота, при создании тематических каталогов, при извлечении материалов тематических Интернет-порталов и т. д. [2, 3]. Для решения этих задач применяются различные варианты кластеризации, такие как: метрические (SVM, KNN [4]); нейронные (различные перцептроны, сверточные нейронные сети [5], ART, SOM [6]); вероятностные (байесовский классификатор [7]); основанные на решении задачи максимизации правдоподобия (случайный лес [8]) и др. Однако вопрос автоматической классификации научных статей по УДК с помощью машинного обучения в работах по данной проблеме практически не рассматривался. Следовательно, целью данного исследования — разработать систему, которая сможет с высокой точностью генерировать коды УДК для научных статей, основываясь на принципах машинного обучения.

Особенностью большинства методов классификации текстов является то, что алгоритм настройки параметров модели (обучение) требует обучающей выборки (некоторого набора данных), а также метода замера и в случае обучения с учителем. В данном случае в качестве обучающего набора данных выступают тексты научных статей, а в качестве метода замера — код УДК, присвоенный статье. Таким образом, методом исследования является сбор данных по первым этапам исследования [9] (выявление и анализ предметной области исследования) и результате которого была составлена обучающая выборка статей.

**Обработка кодов УДК**

УДК, или универсальная десятичная классификация [10] — это система классификации информации, которая широко используется по всему миру для систематизации произведений науки, литературы и искусства, периодической печати, различных видов документов и организации каталогов. Центральной частью УДК являются основные таблицы, охватывающие всю совокупность знаний и построенные по иерархическому принципу: от общего к частному с использованием шифрового десятичного кода.

Основным разделом УДК по ГОСТ 7.90–2007 [10] 0. Общий отдел. Наука и знание. Информационная Документация. Библиотечное дело. Организация Публикации в целом.

1. Философия. Психология.
2. Религия. Богословие.
3. Общественные науки.

выборки (некоторого набора данных), а также метода замера и в случае обучения с учителем. В данном случае в качестве обучающего набора данных выступают тексты научных статей, а в качестве метода замера — код УДК, присвоенный статье. Таким образом, методом исследования является сбор данных по первым этапам исследования [9] (выявление и анализ предметной области исследования) и результате которого была составлена обучающая выборка статей.

После преобработки данных необходимо выбрать гиперпараметры модели. В рассматриваемой задаче — это количество нейронов в скрытом слое нейронной сети или глубина рекуррентной сети. После проведения данного этапа исследования к обучению и тестированию, на основе результатов которого проводится анализ и делается вывод о применимости той или иной модели в решаемой задаче.

УДК, или универсальная десятичная классификация [10] — это система классификации информации, которая широко используется по всему миру для систематизации произведений науки, литературы и искусства, периодической печати, различных видов документов и организации каталогов. Центральной частью УДК являются основные таблицы, охватывающие всю совокупность знаний и построенные по иерархическому принципу: от общего к частному с использованием шифрового десятичного кода.

Основным разделом УДК по ГОСТ 7.90–2007 [10] 0. Общий отдел. Наука и знание. Информационная Документация. Библиотечное дело. Организация Публикации в целом.

1. Философия. Психология.
2. Религия. Богословие.
3. Общественные науки.

**Обучение моделей классификаторов**

В данной статье рассматриваются три метода машинного обучения: линейный метод, метод опорных векторов и метод опорных машин. Каждый из них имеет свои особенности, но все они направлены на решение задачи максимизации правдоподобия при заданном наборе данных. Рассмотрим подробнее метод опорных векторов.

**Искусственная нейронная сеть.** Несмотря на то, что формально нейрон Максвелла (Питча [11]) представляет собой искусственную перцептронную сеть, на самом деле это устройство, которое реализует по сути, алгоритмический процесс, который можно назвать обучением. Искусственная нейронная сеть состоит из нескольких слоев: входного, скрытого и выходного. В данной работе мы рассмотрим искусственную нейронную сеть с несколькими слоями нейронов, в том числе и с обратными связями [12]. Такая архитектура называется многослойной нейронной сетью.

Практика показывает, что для решения большинства задач достаточно одного скрытого слоя. Однако исследованием множественных слоев выявлено, что с их помощью можно находить сложные зависимости, выявить более сложные кластеры объектов и т. д. [13, 14]. Поэтому в работе рассмотрены также и двухслойные перцептронные нейронные сети, которые используются для решения задач кросс-классификации [15, 16].

В таблице приведены значения метрик качества для нераспространенных и обратных связей. Отметим, что различия между ними практически нет. Не-

**Результаты работы классификаторов**

Классификатор	Тип терминов	Ассиметрический индекс	Прецизионность	Полнота	F1-индекс
Искусственная нейронная сеть (одна скрытая связь)	Слова	0,6982	0,6985	0,7033	0,6990
	Биграммы	0,7033	0,7099	0,7182	0,7133
Искусственная нейронная сеть (два скрытых слоя)	Слова	0,6984	0,6991	0,7037	0,7003
	Биграммы	0,6985	0,7031	0,7247	0,7100
Линейный метод с использованием алгоритма AdaBoost	Слова	0,4771	0,4935	0,4771	0,4821
	Биграммы	0,4718	0,4908	0,4718	0,4787
Линейная регрессия	Слова	0,7650	0,7633	0,7650	0,7629
	Биграммы	0,7090	0,7091	0,7090	0,7022

4. Удаляется правые операции операции соединения. В ходе предыдущих исследований [14] было замечено, что в записи этой операции совершается

3. Принимается допущение, что части кода, связанные символами отношения или двоичного отношения, равнозначны и могут быть записаны как через символы, более широкие по значению коды через символы "?". Это возможно в силу того, что при формировании составных индексов на первом месте в отношении ставится раздел, который является основным для документа. Значит основная тема может быть сформирована достаточно грамотно, чтобы части отношения были связаны слабо.

2. Удаляются квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

1. Удаляются все специальные текстовые обозначения (места, времени и т. д.).

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Удалены квадратные скобки, которые применяются для обозначения равнохарактерных связей и индексы с двумя и более знаками отношения, когда необходимо показать, что один из знаков отношения относится к совокупности других индексов, соединенных этим знаком и являющихся с ним одним понятием, например, 621 311 25(621 362 53 84) — деструктуризация с магнитно-электронными регистраторами.

Алгоритмы выбора могут быть выполнены независимо от дерева, который является результатом. Однако мы часто слышим описания алгоритмов выбора в контексте комбинационных методов. Это объясняется тем, что выбор часто выполняется как часть операции по поиску оптимального решения. Например, поиск оптимального решения задачи выбора в контексте комбинационных методов является частым случаем задачи выбора.

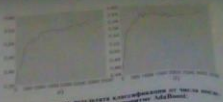


Рис. 2. Зависимость результатов тестирования алгоритма AdaBoost от числа деревьев

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

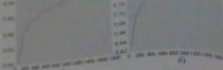


Рис. 3. Зависимость результатов работы логистической регрессии от числа деревьев

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

Генерация кода УДК. Процесс генерации кода УДК начинается с выбора метода кодирования. В зависимости от метода кодирования будет определяться структура дерева. Если используется метод кодирования по частоте встречаемости, то структура дерева будет зависеть от частоты встречаемости символов в тексте. Если используется метод кодирования по информации, то структура дерева будет зависеть от информации, которую несет каждый символ.

В зависимости от метода кодирования будет определяться структура дерева. Если используется метод кодирования по частоте встречаемости, то структура дерева будет зависеть от частоты встречаемости символов в тексте.

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

Вид изображения сети в виде эксперимента обучения по стандартному методу "Машин" в адаптивной структуре обучения, разработанного командой ученых из Австралии (Насиритри и Геррант) [Калман 19].  
Алгоритм Хауэна. Одним из самых популярных алгоритмов выбора является алгоритм Хауэна. Алгоритм Хауэна [20] заключается в том, что очень сложны классификаторы — решение при просмотре дерева с глубиной 1 или глубиной 2 и максимальной глубиной — последовательно рещивать на варианты ошибки, после чего отбрасываются переиспользуемые варианты. В результате остаются только те варианты, которые не используются повторно.

В зависимости от метода кодирования будет определяться структура дерева. Если используется метод кодирования по частоте встречаемости, то структура дерева будет зависеть от частоты встречаемости символов в тексте.

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

Генерация кода УДК

В ходе исследования было разработано свой подход к генерации кода УДК. Первый подход заключается в том, что переход на следующий уровень

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

УДК 62-50  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению  
Исследования по автоматическому обучению

... (text continues) ...

**Список литературы**

1. Furlan L. et al. ESIFC: A Benchmark for Large Scale Text Classification. *arXiv preprint*, 2015. arXiv:1503.03511.
2. Johnson R. *Classification of text documents*. The Proceedings of Learning and Intelligent Systems. Oxford: Elsevier, 2006. P. 2, pp. 431-462.
3. Balaban M. V., Gribina E. G., Ivanov A. D., Staryk V. A. Hybrid system of informational and oblique resources. *Information Science Bulletin*, 2009, no. 3, pp. 23-27.
4. Galdrinich E., Markovich S. Text categorization with many redundant features using adaptive feature selection to make SVMs competitive with C4.5. *Proceedings of the twenty-first international conference on Machine learning*, 2004, pp. 41-48.
5. Johnson R., Zhang T. Effective Use of Word Order for Text Classification with Convolutional Neural Networks. *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the ACL, Denver, Association for Computational Linguistics*, 2013, pp. 103-112.
6. Bai M. E., Yelisevich P. Hierarchical text categorization using neural networks. *Information Retrieval*, 2002, vol. 5, no. 1, pp. 87-118.
7. Furek A. S. Автоматическая классификация текстовых документов. *Методы, классификация текстовых документов. Математические аспекты и приложения*, 2010, no. 1 (21), pp. 65-82.
8. Пронин Е. В., Ягупова Е. В. Адаптивный анализ текстов в задачах рекомендательных систем с-таузис (Aspect analysis of semantic systems for a product recommendation system). *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 23-25.
9. Коновалов А. В., Лопатин К. Е., Козлов Е. С., Колосовский А. Л. Research of Neural Networks Application Efficiency in Automatic Scientific Articles Classification According to UDC. *IEEE 2016 16th International Shorten Conference on Control and Communication (I2C2CC) Proceedings*, Moscow, NRU HSE, 2016, article number 54194.
10. ГОСТ 7.74-96-1997. Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная классификация (система стандартов по информации, библиотечному и издательскому делу). *Известия и публикации. Информационный язык. Термины и стандарты*. М.: ИПК Издательство стандартов, 1997, 38 p.
11. ГОСТ 7.90-2007. Система стандартов по информации, библиотечному и издательскому делу. *Известия и публикации. Информационный язык. Термины и стандарты. Методы, классификация текстовых документов*. Москва, Библиоинформ, 2010, 28 p.
12. Шабалин А. В. Автоматическая классификация научных статей. *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 41-43.
13. Лопатин К. Е., Козлов Е. С., Колосовский А. Л., Романов А. Ю. Применение современных методов классификации текстовых документов к анализу научных статей по УДК. *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 92-95.
14. Терехин А. Т. Нейронные моделирование и обучение в задачах классификации текстов. *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 113-115.
15. Терехин А. Т. Нейронное моделирование и обучение в задачах классификации текстов. *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 113-115.
16. Мешкова Е. В. Методика построения классификатора текста на основе гибридной нейронной сети // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 116-118.
17. Gardner M. W., Darling S. R. Artificial neural networks for multivariate regression - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 1998, vol. 32, no. 14, pp. 2627-2636.
18. LeCun Y. Gradient based learning applied to document recognition. *Proceedings of the IEEE*, 1998, vol. 86, no. 11, pp. 2278-2294.
19. Kluge D., Ba J. Adam: A method for stochastic optimization. *ICLR 2015, arXiv preprint*, San Diego, 2015. arXiv:1512.07545.
20. Hirsch G., Onishi T. Soft margins for AdaBoost. *Machine Learning*, 2001, vol. 42, no. 1, pp. 287-320.
21. Давыдов В. А. Алгоритм обучения нейронной сети на основе гибридной нейронной сети // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 119-121.
22. Zhang E. V. Improving binary logistic regression with neural networks and adaptive bias // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 122-124.
23. Фейсман Т. А., Котельников Е. В. Параллельная реализация алгоритма обучения нейронной сети // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 125-127.

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. Том 23, № 6, 2017**

**А. В. Романов**, PhD, Senior Lecturer, e-mail: a.romanov@phs.ru, National Research University Higher School of Economics, K. E. Lomakin, Student, e-mail: k.e.lomakin@gmail.com, E. S. Kozlov, Student, e-mail: kos.kozlov@gmail.com

**Machine Learning Methods Application to Automatic Scientific Articles Classification According to UDC**

*The paper deals with the applicability of modern machine learning methods to the problem of automatic generation of UDC for scientific articles. As the classifiers, such models as artificial neural networks, logistic regression and boosting are considered. Graph algorithms and a genetic software module for automatic UDC are discussed.*

**Keywords:** text classification, machine learning, artificial neural network, logistic regression, boosting, UDC

**References**

1. Furlan L. et al. ESIFC: A Benchmark for Large Scale Text Classification. *arXiv preprint*, 2015. arXiv:1503.03511.
2. Johnson R. *Classification of text documents*. The Proceedings of Learning and Intelligent Systems. Oxford: Elsevier, 2006. P. 2, pp. 431-462.
3. Balaban M. V., Gribina E. G., Ivanov A. D., Staryk V. A. Hybrid system of informational and oblique resources. *Information Science Bulletin*, 2009, no. 3, pp. 23-27.
4. Galdrinich E., Markovich S. Text categorization with many redundant features using adaptive feature selection to make SVMs competitive with C4.5. *Proceedings of the twenty-first international conference on Machine learning*, 2004, pp. 41-48.
5. Johnson R., Zhang T. Effective Use of Word Order for Text Classification with Convolutional Neural Networks. *Human Language Technologies: The 2013 Annual Conference of the North American Chapter of the ACL, Denver, Association for Computational Linguistics*, 2013, pp. 103-112.
6. Bai M. E., Yelisevich P. Hierarchical text categorization using neural networks. *Information Retrieval*, 2002, vol. 5, no. 1, pp. 87-118.
7. Furek A. S. Автоматическая классификация текстовых документов. *Методы, классификация текстовых документов. Математические аспекты и приложения*, 2010, no. 1 (21), pp. 65-82.
8. Пронин Е. В., Ягупова Е. В. Адаптивный анализ текстов в задачах рекомендательных систем с-таузис (Aspect analysis of semantic systems for a product recommendation system). *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 23-25.
9. Коновалов А. В., Лопатин К. Е., Козлов Е. С., Колосовский А. Л. Research of Neural Networks Application Efficiency in Automatic Scientific Articles Classification According to UDC. *IEEE 2016 16th International Shorten Conference on Control and Communication (I2C2CC) Proceedings*, Moscow, NRU HSE, 2016, article number 54194.
10. ГОСТ 7.74-96-1997. Система стандартов по информации, библиотечному и издательскому делу. Информационно-библиотечная классификация (система стандартов по информации, библиотечному и издательскому делу). *Известия и публикации. Информационный язык. Термины и стандарты*. М.: ИПК Издательство стандартов, 1997, 38 p.
11. ГОСТ 7.90-2007. Система стандартов по информации, библиотечному и издательскому делу. *Известия и публикации. Информационный язык. Термины и стандарты. Методы, классификация текстовых документов*. Москва, Библиоинформ, 2010, 28 p.
12. Шабалин А. В. Автоматическая классификация научных статей. *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 41-43.
13. Лопатин К. Е., Козлов Е. С., Колосовский А. Л., Романов А. Ю. Применение современных методов классификации текстовых документов к анализу научных статей по УДК. *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 92-95.
14. Терехин А. Т. Нейронные моделирование и обучение в задачах классификации текстов. *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 113-115.
15. Терехин А. Т. Нейронное моделирование и обучение в задачах классификации текстов. *Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 113-115.
16. Мешкова Е. В. Методика построения классификатора текста на основе гибридной нейронной сети // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 116-118.
17. Gardner M. W., Darling S. R. Artificial neural networks for multivariate regression - a review of applications in the atmospheric sciences. *Atmospheric Environment*, 1998, vol. 32, no. 14, pp. 2627-2636.
18. LeCun Y. Gradient based learning applied to document recognition. *Proceedings of the IEEE*, 1998, vol. 86, no. 11, pp. 2278-2294.
19. Kluge D., Ba J. Adam: A method for stochastic optimization. *ICLR 2015, arXiv preprint*, San Diego, 2015. arXiv:1512.07545.
20. Hirsch G., Onishi T. Soft margins for AdaBoost. *Machine Learning*, 2001, vol. 42, no. 1, pp. 287-320.
21. Давыдов В. А. Алгоритм обучения нейронной сети на основе гибридной нейронной сети // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 119-121.
22. Zhang E. V. Improving binary logistic regression with neural networks and adaptive bias // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 122-124.
23. Фейсман Т. А., Котельников Е. В. Параллельная реализация алгоритма обучения нейронной сети // *Известия и публикации. Информационный язык. Термины и стандарты. Матрица научных идей XVIII Общероссийской конференции Интернет и информационные технологии (IMS-2015)*, Санкт-Петербург, 2015, pp. 125-127.

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. Том 23, № 6, 2017**















К. Н. Ломотнв. Кс

Н. С. Конова. солснуга. е•mail сопл

обучения

для реп.ения задачи автоматической рубрикации статей по У ДК

при исмамоы•ти

мет.н\м

А

“и•оритиы

МА в.лччы• статей В

'ч•и•'

Д.уч•.сик

сеть .

«5.-чены•.

выборки (некоторого набора ланны)\'. а вста  
эксперта в случае обучения с учителем В валлча обработки текстов ном  
случае в качестве наэ•юа Инны,

На современном выступают тосты научных статей. а в качест— на естеслвснныч ч 'ыкач чьнлштш•н  
олнон из насуш- неша жсперта кол У ДК. присвоенныи ны\'. Она напрямую свя кша с расшупшм колнчсст-  
к•рнст гороч или чолсрлторами ресурса Таким воч [ТКС тонов информации н септ Ин этапом  
исс.лсло•иния ятается

Научныс статьи. киши. руководства. журналы. в спра- ( первым выполнен в коле прслыд»иего  
исслеЛЮБния ночники — все мноахязразне была состипена

спк• своем гцх•лсглышно в нифровом Вместе св вьюорка результате статси.которо10

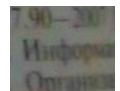
гем вникает ориентации н гаком боль-в ос- После предобработки данных необходимо

п.тгокс информации. поскольку ланные добрать иперпараметры молелеи В нонном нс унорчлочсны.  
и поиск полной информа- мои залаче — это количество нейронов в скрыт, нин по конкретному нонрху  
становится все более слое нейронноИ сети или ' лубина решаюа:тj гул.к•мким, В этой сня нспшьзонанис  
автомаги- После прохождения данного этапа модели гь

рева. ческиз якчястсн одним пл перспек- тоны к  
мЗучению и лесгированию. на ось— И гивных репкния ланной просисмы зультатов которого проводится анииз  
и

Змачи юассификашш п•кстон во шикают в раз- выводы о применимости той или иной модеш личных  
областях: например. при обработке элек- решаемой задаче. тронной корреспонденнип (фильгры спама): поиске  
информаии; контекстной рекламе: в системах Обработка кодов У ДК локумсншо«нЛьорота: при создании  
тематических капри наптнenni магерихтами гематиче- УАК. или универсальная десятичная скич  
Интернел=норгшчов и г. л. р. З'. Для реп.ения каиия ЮI — это система классификации ишач приме•нянтг  
варианты класси- мании. которая широко используется во рашшчные фикаторов. гакие как: метрические  
периснлроны, (SV М. свергочныеК ХМ шля систематизации искусства. периодическоИ произведений  
печати.наухи. нефюе•теныс (ралличныс и нейронные сегн 1.51. АКТ. SOM вероятност- туры видов документов  
и организации ные (байесовский юиссификатор 17 р: основанные ных УДК являются на ренькнних правилах  
(рентлюшие деревья, слу- Центтхгньной частью чайный лес и др. Однако вопрос автоматиче- липы.  
охватывающие всю совокупность скоп классификашш научных стаге” по У ДК с по- построенные по  
иерархическому приниипэ монпю машиннонз обучения н работад по данной ния от общего к частному с  
использование просис•магике практически не рассматривался. Сле- роного десятичного кола.

ловагслно. пель ланноуо исследования — разрабо- Основные рамелы У ДК по ГОСТ гагь  
систему. которая сможет с высокой точностью 0. Общий отдел. Наука и знание 'енсриронать  
колы УПК для науппых статей. осно- Документация. Библиотечное дело нынаясь на  
нриннипах машинною обучения. Пу(уликации в целом.





(особенностью большинства методов классифи- 1. Философия. Психология. кашти  
текстов является ш, что игоритм настройки 2. Религия. Богос-товие. параметрон чолели  
(обучение) тебует обучающс•й

3, Общественные нажи.

418



ИНФОРМАЦИОННЫЕ Пхнолог—, т—

§

науки

листе И

6

Мслипнна 1 ечноч.нмя

РЛ

(пор [

Кол ЛК птх•лелштне1 сосняч посл.слов•лге ннфр. тзн че• ка—лыс нолс•мс опрелсли соо'нсг-ль свои ра н некоторыв

шннфра. нифра “ (Г. цифра е апострофом. нифра \_н•фисом'. ванисъ лиана уем пора •лелитсся “ кода УАК опрелслены онерапии отношения олного кола Э к тру•ому н сослннсния(сим• лноИного отнокн•ння

В общем случае исполь•онагься опр• к.титс.тн языка. этносов и наролнос места (с возможностью .ток.тлизипии локумснга. н нрсле•ллх времени.алмнни,-п-огивной слннины•. зон и рлионов, формы Также имененся так кшматических “обшис опрсле.штели”. которые начиназывается с символов “—0“ и могут означат. свойства. наюгся или личные характеристики 21\_ материалы

Генерировать кол такой сложной структуры. имея лишь гексг научнон стагъи.может справиться - эло ктлача, с лажекоторой мчастую с человек трудом Шя автомач сис• полготоубтснныВ тем она все сше остается и разработки невыполнимой. н области НССМОЛРЯобработ-

на эксперименты на естественных я лыках Поэтому кн текстов работы колы УПК УПК для в гекстон контексте из обучающей данного исслелова-выборки

бытн приведены к более простому виду по следуюишу правилам: удаляются все сненихтьные текстовые опре- летитеш (места. времени и т. л.)У у лалются кн.ттрагные скобки. которые применяются шля обозначения разнохарактерных связей в иоексе с двумя и более знаками отношения, когда

ненных этим знаком и янляюшихся слиным понягием. например, 621 В1 125:162 [362:531 ,841 мектростаниии с магнитогилролинамическим нера тором .

3. Принимается лопушение. что части кола. сваяинные символом отношения или двойного отношения. равнозначны и могут быть записаны как отдельные более широкие по значению коды через символ Это возможно в силу ОЛНО10 из правил образования составных индексов: на первом мсте в отношении ставится раздел, когорый яюшется основным шшш документа. Зачастую основная тема может быть сформулирована достаточно размыто, чтобы части отношения были связаны слабо.

4.

В припечены ктаксифиешоо«н• и ИА Чс [РИК“ для релли ивла•ш

в велу• к

неиЗИИМО показать. что олин из знаков отношения относится к совокупности других индексов. соеди-

Обучение моделей вдаееифи:в.торо.

лелей име•с••• свои (0 не но умолчании.

и полра •умсвае•еи

Сле•я при описании милели, поскольку не лис. научно“ нови питавные молсли классици•кллоров подробно

Искусственная нейронная сеть. Несмо•ри на то, что (формальный нейрон Маккллоел Пи гп:д реали вус•. по суш, ие:тичсскуи• рел реесив•. ком- по линия таких совер•нснно новую молсль. в нско•орых аспекза» неао раГне мо нат Нейронные примон• рае проелрране•нии н ное лелнсе время (Алрели Гн.лыиуиз нонулярность и их уснсп•но примсннтл для клас сицшкании рав•ичныя объев. [он. н гом числе и гекстон 15. В мы рассмацливли нейронную сс н. с нес колькими слоями нейронов, сель ВАОЛНОИ слой. выколной. а

•акжс олив или нско.лько скрытых 191. 1 акая архшектурл называется мнонлелюным

Практика показывает. шли ремне•нии большшнслма залк•ч лост;ночно олно'о скрытою слои, (лнако исслеловании МН'ЯОСЦПВ'ЫА снууктур свилетсльстнуун о '10М. •пос ич номошью можно находить неявные ;анисимостм, разслятг. более сложныс кластеры объсктон и г л 171 По ному в раб'оте рассмојрсны также и лиухсловные перпе•илроны. Больше число слоев, как праи•ью. испольгуется для рннсния узкого круга спенифичсских валач ' Я,

В т•аблице приведены значения метрик качссгва для нейросети с олним и днумя слоями, (очсвидно, что раишпы между ними практически нет.

Классификатор Тшш термов

о дою

Искусстенля нейронная сеть  
Киграммы (Олин скрытый слои)  
Иску«твснндя нейроннля сеть  
Бицх.ччы (два скрытых слоя)  
Деревья с исполкилванием

Биграммы шпоритма О.ИЧ,

.Лшж•тическа

Удаляются правые операнды операции соединения. В ходе предыдущих исследований [14] было отмечено, что в записи этой операции совершается



неф—:•

й: хт— шта пои -  
Ь

ф» нытси —

па ока ылась примспно н. О 5 чсч  
прихлснные в

мовет являтья  
мя •вгомтичесыно гмпри  
ТнКтвкнии системы класси-

Чалт:ь по метолу ••Adam- с адат“ ИВнов сковнк”ть«»  
(Мучения, рата(нпанному коман-

А. •ыт— буствнг. Одним из сачыд популярньп  
ансзмотсвы• истолон ТАЯСТСЯ апоритч  
аланЛус•гингз .AdaB0(N | 20' Заик»частся он в том, что  
очень слабые классификаторы - - решаюше пня  
'решаюше•е лерегю [лубяной или лечатенькой  
глубиной последовательно

уч.гся на взаимны. пшибках. после чего их  
огнеты (былинякпся оптклсленныч образом. В  
результате получается тсьма хорошо  
настраиваемая модель, приченичая пгтько на  
деревьях с очень мштенькой поскшькь она склонна  
к переобучению В каков- момент ре вульаты  
ЛОЛАНЫ сибили пнжаться. что и присолит у  
Ю1ассификаторгов. испотыуюющих слова (рис 2).  
Шя юлассификагоров. инщыуюющих биграммы.  
стабилизация не тшикаег на этом нромежуткс. что  
является ири внаком сложных нелинейных  
ивисимостен ре •у.тьгата ог признаков.

Логнстическая регрессия. Стелуноншм исполь-  
јусмыч южсификатором стата логистическая ре'  
д•ссия. предсгабяющая собой линейную  
регрессию. ;кзульаты которой являются  
аргументами логистической функции. что по  
шоляет определять вероятность принадлежности  
текста к определенному классу 1221 Яким образом.  
получен достаточно фА•ктинный к.лассификатор,  
результаты тести-

рования которого принслены на рис. 3. Отдельв,  
стоит отметить. что этот классификатор покани  
мые высокие результаты по всем используемым  
метрикам. что следует из приведенной габлины

Анализ результатов тестирования классификато.  
ров. Слелуст отметить. что лучшие РС вульаты  
даг, логистическая регрессия шля слот ней Неболг  
пшм отставанием идут лот истическая регреии;  
л\_ля биграмм и нейронные сети. мгоритм Аса  
Воом, построенный на решаюших деревьях, Поч.  
хулший результат.

В большинстве случаен при равном числе  
мов биграммы показывают худшие  
результаты. чеч слова, практически по всем  
метрикам качества Тем не менее рост качества  
в зависимости от чисе термов у них  
продолжается дольше, это позволям слелать  
вывод о том, что потенциал их применений на  
большем объеме данных у них выше.

Можно заметить. что качество лучших мода;  
приближается к отметке 0,7. Вероятно, это но  
ошибками в экспертных метках юласса и нес.  
вершенстном метола отбора ключевых термов  
сюда следует вывод о том, это наравне с  
гиперк& раметрами модели следует улучшать  
качестыј ных и вспомогательных алгоритмов.

### Генерация кода УДК

В холе исследования было ра 'работано  
чатся в том, что переход на следующий

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. том В

ТО  
ШИИ  
гек

би

ни

кл

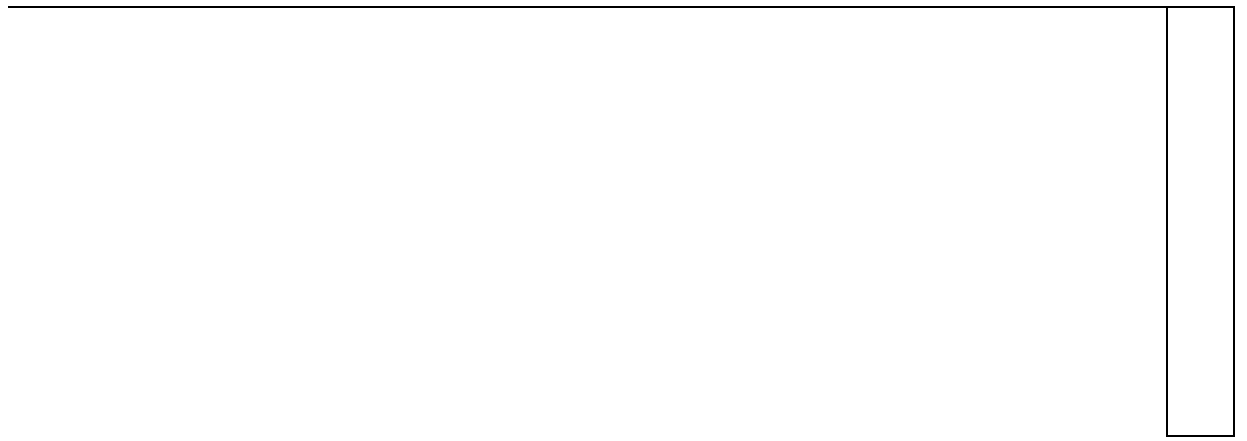
е;

е

с







·  
·  
|

сле л.ф.вторы в аиле к Шону или иному в нск.Н'Ч\*Но локалл\со• принадно\ол Чо•но СЛЧЫА ла.ЛЬНИ\ лрсне по форчылировка кослслуыннч образом текс нс — нан•и нрпшну к когороИ от нелст ребро е самым

Снульшич ите-

к

иннилснч• ю п•ьун.ен. и принять ее ы текущую гавот полчочла чалнстсн высокая 910

' ле

«рена конечна невелика. а 'акжс тк•ч что олно•о 'скстл часть вес» клас.•ификшороп с нс то качсстгно очень важно нос при тожегак как конечный поль и•вагсль Г»от.

-танны\ н Котор»А) ин система к или человек )

не

.ннеп• в гсчснис јглиттльн«но нрсчени В сле час. копи классификаторы ннпчрированы в более систему валержк;• может нарушить ес раже снсгсма классификации работаеш на сервере. го ссрнер. вероятное всего. разорвет сос,шнненне по [аимау. лавный нелостлгкок алгоритма состоит в том. что он не гарантирует нако•аснис лучшего ре јуль• тата: нс происходит обмол нс.мо дерева с присвовеса реорам. поэтому учитывакугся лишь веса иннилент•ных гекушей вершине, [сли у вепш Дерена показатели распо внавания „теине. го они будут проигнорированы. н то время как из низким покатате.теи нижних уроннеи тлет выбран наибольшии. хогь и не предельно оптичальный. в контексте всего дерева,

Второй полхолл состоит в построении ориентированного нзвешснного дерева всей струкг.ры УЛК. заложенной в систему при обучении. Системой выполняется опенка приналтежносги текста кем массам и полютассам. пока ители ставится в соотнегстние ребрам. вхоляшим н ршину. а ваттм выбирается лист. шля пути к которому лос-гшаецсн максимум некоторой меры принаиежности (наприиер. среднего значения ве•сов пупо. Хнор.нм. реми зуюший втрой подход. вы\дялит слелуюющим образом: пля всех листон получить ответ масснфikatorа шля каждой на каждом увинс. ммировать ре вульгаты опенки ши кех путей но уровням и разделить на номер уровнв. на котором

начолится лист данной ветви. и выЗбрать лист, на пути к которому среднее значение максимально.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГЪМ. том В. В. РИ•мы п.

тьно

ы

сл ап.Асны

икно

Рзиские Пре о нвхч•ипании и л о

святык в лскнд Китм• В

полхолл тысячами

Постткнис тене'ьи•и

(ктановить спуск. АКИМ ЧО•НО О СМ:.

стигм. колы хля которы• быть —ты« скнерированы пнько -ьыстрымтолько “ ме.иенным" шлпттом Пса решение это“ лту.т,темы состоит в горитмов, Например, сначала каецся на несколько

н«.чгеы

• р.н\н

шля

н • всН\сВ

отче-•ння н  
'емЫ ь кон е но лк к ллсеи

шодлив'  
(ли

но о Рь ьнс.п, клчсс мсм

рикзм'

Н проект:с иселе•лонанвн бы ли рлрлрл(Ылны

графоные колл УЛК комьянани

от:шчлечся н потенп.тлом

улу•ппсннн ( )лннм и

вспре•јультавоп работы стал пропннн

но рнрннну чешкой речн:сснн и способный

'енсррон;нь пернут нифру колл УЛК

Стслует гнмсгнть. рсзутпгг:тты ллнннВ япчяяся

для нровсдения лаль•

неин•ид исследовани, в колс выполнения кшнорых

рассмотренис огнпирной проблс•

чатнви. которая включает :

- проблемы балнса скорости и при иснотьлов.лнни а.•норп 1мон 'сне•рании кола УПК.
- перспективные архитектуры нейронных сетей шт рассмтривасмой проблем,лнлки и возможногти примениии метолов искуссртзснннц интеллска к управлению построением системы классификашров: • возможности рекурсиннсно огоматического построения сиск:мы классификаторов шля дерена УД К:
- возможности применения мелоло параллельных вычислении ускорения алчјритмов классифик:шии н [шнерапии УПК:
- применение семантических моделей текста. а также проведение эксперимен гон по комбиниронанию их со статистическими метолами;
- возможности п.ччерании спе•нишьных определиге.тсй УЛК (места, даты, этноса и т. д.):
- адаптация результатов исслс•лошјний к другим мдачам об:хи5отхи гексгов на естественных языках. Итогом эгид

исследований станет создание программнош комплекса, который сможет самостоятельно загружать текстоные ;шокументы требуемого

22

ПЛ [ мне

54

„т, МЮ. „„„„А НЯН„пе.

НИВ. и иллагелы-кпиу ле т, ин.у.р?,:; во но лоиееопыс 1997 АИТермины и И

[ЧК Т 7.90—2007 СЛ«тема с.гавластгпв и и и вдаге;н.ьыгму лелу

ная клас. ификапии С [рук тура. правила и напни. М тллатпнформ. 2010 2И

2 ( )бшая методика ирмевевия уни—вальс••

классификаии.дата травнения 25 ! 2 20 (

' 3 Заболеева- Зотова А. В.. К). А. Автоматв.ш., млншическ«ио аналила локуиентании технического компьютерных и информиионных тегно,мини

14. Ломотип К. Е.. Козлова Е. С.. Колееии•по А. Х. нов

А. К). Применение статеи соитк•мнннх по У.ПК истоцов

// Инновационнну„чассифеца» рубрикапии наушыя

(Бормапи• •ннне и коммуникаиионные технологии гураон МП Межлъяролнои научно- практической нии М. Жуковскоп». Ассониипии 2016 выпускников С 92—95.и сотрудницы ВЫ

им проф

5 Терехин А. Т. Нейросетевое моделигювание функций мо на: об «кноинных илие слслонания 2009л 2 httpwpsystuay ния: 25 12.2016) 16 Мешкова Е. В. Методика

Южною федеральною университета 17

Gardner M. W., Oog\* S. И.

A. Yu. Romanov. PhD. Senior lecturer. e-mail: a.romanov@hse.ru.  
 National Research University Higher School Of Economics. K.  
 E. Lomotin, Student. e-mail: ke.lomotin@gmail.com.  
 E. S. Kozlova. Student. e-mail: hse.kozlovaes@gmail.com

### Machine Learning Methods Application to Automatic Scientific Articles Classification According to UDC

The paper deals with the applicability of modern machine learning methods to the problem of automatic generation of UDC & scientific articles, the classification of such model as artificial neural networks, logistic regression and boosting. Graph algorithms and prototype software module to generate are developed.

Keywords: text classification, machine learning, artificial neural network, regression, boosting, UDC

#### References

Partalas I., et al. ISITCJ A Benchmark for Large-Scale Text Classification. *arXiv preprint*, 2015. arXiv no 1

2 Sebastiani and F. Ingwers. Classification of text. Edition, 2006. Oxford: The Elsevier, Encyclopaedia 2006. [Language vol 2. pp 457—462

J. Bulgakov M. V., Gridina E. G., Ivannikov D., Staryh V. A. 'System of standards of information resources in the federal system of informational and educational resources'. *Informatsionnyye Resursy*, 2009, no 2. pp 25—27

4. Gabrilovich E., Markovitch S. Text categorization with many redundant lemmings: Using aggressive feature selection to make SVMs with C4.5. *Proceedings of the twenty-first international conference on Machine learning*. 2004. pp 41—48

Johnson R., Zhang T. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. *Human Language Technologies at the 2015 Annual Conference of the North American Chapter of the ACL*. Denver Association for Computational Linguistics. 2015. pp 103—112.

6 Ruiz M. E. P. Hierarchical text categorization using neural networks. *Information Retrieval*, 2002, vol 5, no 1. pp. 87—118. Eprev A. S. Avromaticheskaya klassifikatsiya tekstovykh dokumentov (Automatic classification of text documents), *Struktural'nyy modelirovaniye*. 2010. no. 1 (21). pp 65—82

8, Pronoza E. V., Jagunova E. V. Aspektnyy analiz otzyvov o restoranah dlya rekomendatsionnykh sistem v turizme 'Aspect analysis of restaurant reviews for e-tourism recommendation systems'. *Sbornik nauchnykh A" ob edinennoy konferentsii / Internet sovremennoe (MS20/5)*. Sankt-Peterburg. 2015. pp. 23—25.

9 Romanov A. Yu., Lomotin K. E., Kozlova E. S., Kolesnichenko A. L. Research of Neural Networks Application Efficiency in Automatic Scientific Articles Classification According to IEEE 20/6 Inam-ammo/ Sinenun Conference on Control and Communications (SIBCOS) Proceedings. Moscow: NRU HSE. 2016. arXiv:1604.04444v1 [cs.LG]

1997. Sifema standartov po informatsii. *biyetyet: anvmu i icdate'skomu delu. Informatsionno-ptskovve;azyki. iemmy opredeleniya (System of standards on Information. librarianship and publishing Information retrieval languages. Terms and conditions. M. WK Izdatelstvo standartov. 1997, 38 p.*

! i GO.ST 7.90—2007. 'Sistema Standartov po Informatsii. biblioteka'skomu delu Unisvvlhaja devatichnay'tl klasifikatsiya. Struktura pravda nedemja i indeksirovaniya System of standards on information. Publishing. Universal Decimal Classification; Structure. codes of administration and indexing'. Moscow, Inform, 2010, 28 p.

ТЕХНОЛОГИИ, Том 23, 6, 2017

Объединяет методика применения автоматизации на основе ИИ классификации документов с помощью системного анализа. <http://www.rscd.ru>

Guide to the use of the 25 12

1 A. V., Ju. A. Avtomaticheskaya tekhnologiya. [aiaru@yandex.ru](mailto:aiaru@yandex.ru)

ernant anal'iz is oi I:vesnp 3. 52—Si

K. E., Kozlova E. S., Romanov A. Yu. Primeneniye matritsnykh klastrov nauchnykh statei po aplikatsii modernizatsionnoy tekhnologii. *Informatsionnyye Resursy*, 2010, no. 1 (21). pp 65—82

tonnye Ybomik trudov 'fe+tdunarudnot nauchno-pruknchesko; Muq-ow. vypusnikov somninkov I'V/4am nruko•vsko.»

IS A. T. 'veirosete. oc k•æ.nbr.nyh funkcij moz&i 0b''' osngy•vnyh 'de; Neurai of hr-am functtons an of the hasu dovanva. 2fY)9, no 2 avaiabie ar [http ru](http://ru) (date of Ecess: 25 1 2 2016)

16 Meshkova E. V. Metoda avtomaticheskoy klassifikatsii kamora teksta na osnove ybr'dnoj neimsetevoy metodiki: chnzquc of text clasifier construction based on the neural network model. taesnja Juzhnoy federal'noy unversiteta. *Tekhnicheskie nauki*. 2(Y)8.

17 Gardner M. W., Darting S. R. Artificial neural networks (the multilayer perceptron) as a paradigm for the environment. 32 *IEEE Trans. on Systems, Man, and Cybernetics*. 1998. vol. 28, no. 1, pp. 56–65.

19 D., Ba J. Adam. A method for  
ICLR 2015. arXiv preprint. Sun Otego. 2015. no. 1.

20 Ritsch G., T. Soft margins for learning. vol. 42. no. 4. pp. 287–320

21 V. I. Aieor-rtmy ohuchen»a. na resha'ushhh Machine learning algorithms on construction of decision trees), Zhurnal matematicheskoy fiziki 1982. vol. 22. no. 4. pp.

22. V. zovarye binarnyye funktsii

Ocherk teorii

Slon to the quality equation of the test. & zrsnenn.unnersueta. Iomsk.

. 23 Peski\*eva T. K«el'naov E. V. gortma ohucheniia stgemy tek.stovo' mentatton of learning algorithm leu (fimskoep

U,ra• UGATU. IS. m.