

УДК 004.032.26

DOI: 10.26102/2310-6018/2019.25.2.027

А.Г. Кравец, А.С. Бурмистров, П.А. Задорожный  
**ЭКСПЕРИМЕНТАЛЬНОЕ ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНЫХ  
ПАРАМЕТРОВ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ЗАДАЧ  
КЛАССИФИКАЦИИ ПАТЕНТОВ**

*Волгоградский государственный технический университет,  
Волгоград, Россия*

*Показатели патентной активности в настоящее время нередко используются при технологическом прогнозировании и в конкурентной разведке. Важную роль приобретает прогнозирование развития патентных трендов в отдельно взятых странах и во всем мире, позволяющее выявить основные приоритетные направления развития технологий. Одной из основных операций анализа уровня техники для патентной заявки является поиск патентов аналогов. Международная патентная классификация устарела, большинство исследований носит междисциплинарный характер. Появляется необходимость в выделении и создании новых классов. Цель данного исследования – анализ параметров, влияющих на результаты работы рекуррентной нейронной сети, предназначенной для тематической классификации патентного массива. Проведение анализа выявленных параметров, влияющих на качество работы нейросети и подбор оптимальных значений. Были определены оптимальные параметры нейронной сети: количество слоев, размер слоёв, значение исключающего параметра, размер партии для обучения сети, осуществлен выбор оптимизатора библиотеки Keras. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-07-01200.*

**Ключевые слова:** тренд, классификация, патент, рекуррентная нейронная сеть, исключаящий слой, анализ оптимизаторов, размер партии.

### **Введение**

Показатели патентной активности в настоящее время нередко используются при технологическом прогнозировании и в конкурентной разведке. Важную роль приобретает прогнозирование развития патентных трендов в отдельно взятых странах и во всем мире, позволяющее выявить основные приоритетные направления развития технологий [1].

Одной из основных операций анализа уровня техники для патентной заявки является поиск патентов аналогов, которые могли бы опровергнуть новизну заявки и сделать невозможной выдачу по ней патента. На проведение поиска во многих случаях изобретатель тратит десятки часов и при этом просматривает тысячи существующих патентов и других документов [2].

В целом, на первом месте по количеству поданных заявок уже 39-й год подряд остаются базирующиеся в США заявители, на долю которых приходится приблизительно одна четверть (24,3%) из 233 тыс. заявок, поданных по процедуре Договора о патентной кооперации (РСТ) ВОИС

(Всемирная организация интеллектуальной собственности), в то время как само по себе число выросло на 7,3% по сравнению с предыдущим годом (Рисунок 1). Следующие два места заняли заявители в Японии (19,4%) и Китае (18,5%), причем последние являются главной движущей силой общего увеличения спроса. Наиболее высокую активность демонстрировал сектор цифровой связи и компьютерной техники.



Рисунок 1 – Рост количества заявок

Наиболее высокая доля опубликованных заявок по процедуре РСТ (8,5%) приходится на сектор цифровой связи, за которым следуют области компьютерной техники (8,2%), электронного оборудования (6,9%) и медицинской техники (6,8%).

Среди десяти ведущих отраслей техники наиболее высокие темпы роста в 2016 г. были зафиксированы в областях медицинской техники (+12,8%), оптики (+12,7%) и цифровой связи (+10,7%).

Цель данного исследования (Рисунок 2) – анализ параметров влияющих на результаты работы рекуррентной нейронной сети, предназначенной для тематической классификации патентного массива.

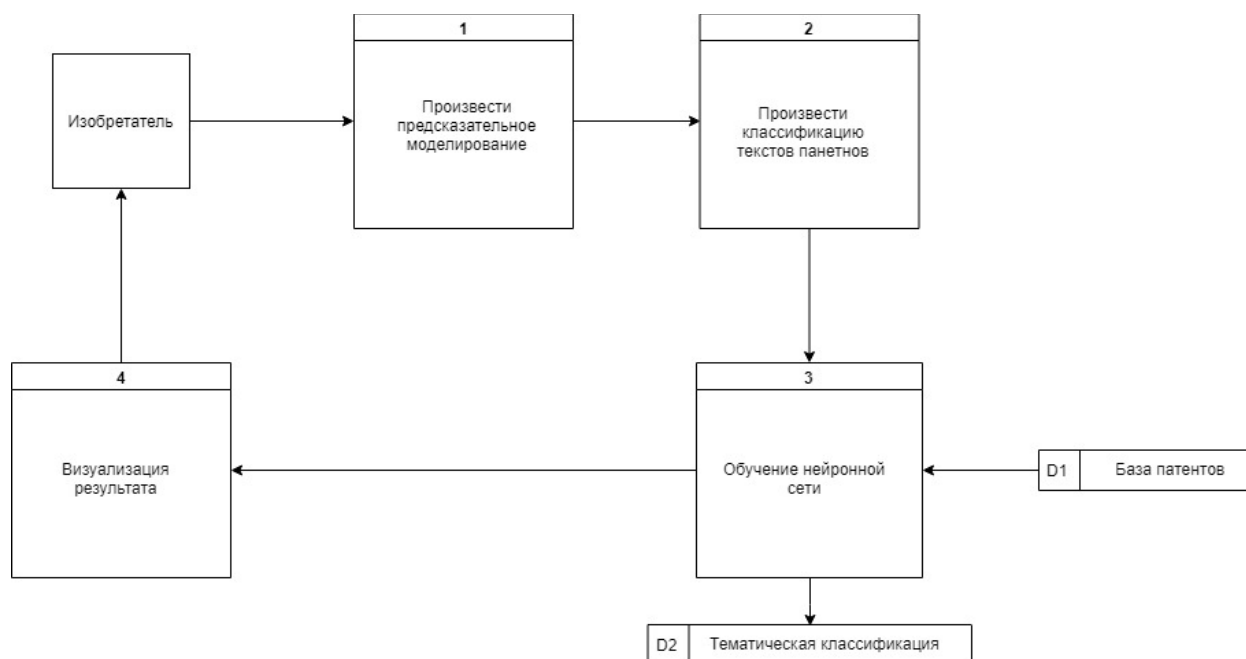


Рисунок 2 – Общая информационная модель исследования

## Материалы и методы

Процесс анализа патентов изобретателем не начинается с чистого листа. При получении информации, человек понимает каждое слово, картинку, график, основываясь на понимании предыдущих знаний об этих объектах. Мозг не выбрасывает все накопленные знания и не начинает “думать с нуля”. Генерируемые мысли имеют в числе своих качеств постоянство [3].

Традиционные нейронные сети не обладают этим свойством, и в этом их главный недостаток. Традиционная нейронная сеть не может использовать рассуждения о предыдущих событиях, чтобы получить информацию о последующих.

Решить эту проблемы помогают рекуррентные нейронные сети (Recurrent Neural Networks, RNN). Это сети, содержащие обратные связи и позволяющие сохранять информацию.

Поскольку, для задач классификации текстов патентов и предсказаний трендов технологического развития используется нейронная сеть с долгой-краткосрочной памятью, то появляется необходимость в подборе оптимальных параметров этой сети, для получения более качественных результатов.

Нейронная сеть была создана с помощью библиотеки Keras. Нейронная сеть состоит из одного или нескольких уровней LSTM, каждый из которых сопровождается исключаяющим слоем, который помогает упорядочить обучение и предотвратить переобучение сети [4]. Для оценки обучения нейронной сети, было решено использовать функцию потерь бинарной кросс-энтропии (Log-Loss) вместо классической среднеквадратичной ошибки. Это связано с природой рассматриваемой проблемы, близкой к проблеме классификации по нескольким меткам, и поскольку бинарная кросс-энтропия даёт многозначную классификацию, а также даёт оценку потерь, более точно отражающую точность модель [5]. Чтобы оценить производительность сети, исходный набор данных был разделён на 80% обучающих данных и 20% проверочных данных. Используется F-мера, для вычисления точности сети [6].

Эксперименты помогут определить наиболее подходящие параметры нейронной сети, а именно:

- количество слоёв нейронной сети;
- размер слоёв нейронной сети;
- значение исключаяющего параметра;
- размер партии для обучения сети;
- тестирование и выбор оптимизатора библиотеки Keras.

### Методика оценки результатов экспериментов

Метриками для оценки алгоритмов извлечения информации являются Precision (точность) и Recall (полнота). Данные метрики используются в качестве основы для производных метрик, F-мера или R-Precision.

От уровня точности и полноты, зависит качество обучения системы, однако максимальные значения точности и полноты не могут одновременно достигаться и возникает задача нахождения оптимального баланса [7].

Именно, F-мера является такой метрикой.

F-мера стремится к нулю, если точность или полнота стремится к нулю.

$$F = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (1)$$

Данная формула придает одинаковый вес точности и полноте, поэтому F-мера будет падать одинаково при уменьшении и точности и полноты.

F-мера является хорошим кандидатом на формальную метрику оценки качества классификатора. Она сводит к одному числу две других основополагающих метрики: точность и полноту. Имея в своем распоряжении подобный механизм оценки вам будет гораздо проще принять решение о том являются ли изменения в алгоритме в лучшую сторону или нет.

### Определение количества слоев

Это очень важный этап в применении многослойных нейронных сетей. От него зависит качество решения задачи. В настоящее время нет каких-либо жестких правил для выбора количества скрытых слоев или выбора количества нейронов, входящих в каждый слой. Однако существуют ограничения, помогающие принимать решения:

1. Если функция определена на конечном множестве точек, то 3-ех слойный персептрон способен ее аппроксимировать.
2. Если функция непрерывна и определена на компактной области, то 3 - слойный персептрон способен ее аппроксимировать.
3. Остальные функции, которым могут быть обучены нейронные сети, могут быть аппроксимированы 4-ех слойным персептрон.

Таким образом, теоретически максимальное количество слоев, которое необходимо – четыре, или два скрытых слоя. Хотя при решении реальных задач может использоваться большее количество слоев.

### Определение размера слоев

Большое влияние на работу нейросети оказывает и размер её слоёв. Существует прямая зависимость переобучения сети от размера слоя. Чем больше количество нейронов в слое, тем больше сеть способна сохранить данных обучающего набора в весах. На Рисунке 3 отображены кривые для 5 размеров слоев: 32 (бледно-голубая линия), 64 (розовая линия), 128 (зеленая линия), 256 (серая линия), 512 (оранжевая линия).



Рисунок 3 – Результаты эксперимента по тестированию размера слоев где а – тренировочная F-мера, b – тренировочное значение потерь, с – тестовая F-мера, d – тестовое значение потерь.

Тренировочная F-мера эпох – F-мера, рассчитанная по тренировочным данным (ось X), нанесенным на график по отношению к количеству эпох (ось Y).

Тренировочное значение потерь – значение потерь, рассчитанное по тренировочным данным (ось X), построенных по отношению к эпохе (ось Y).

Тестовая F-мера эпох – F-мера, рассчитанная по тестовым данным (ось X), нанесенным на график по отношению к количеству эпох (ось Y).

### **Подбор исключаящего параметра**

Для предотвращения разрыва соединения синапсов между двумя слоями нейронной сети и упорядочивания обучения используется специальная техника, именуемая как, исключаящий параметр. При помощи исключаящего параметра, можно распределять информацию по различным направлениям, что в свою очередь побуждает сеть использовать разные пути. Суть данной техники заключается в исключении переобучения нейронной сети. Однако совсем исключить переобучение не удалось, но из графиков и экспериментов заметно, что происходит процесс задержки переобучения.

Рассмотрев Рисунок 4 можно выделить, что добавление исключаящего слоя повлекло медленную сходимость сети к тренировочным данным, тем самым увеличивая обобщение за счёт уменьшения запоминания модели [6].

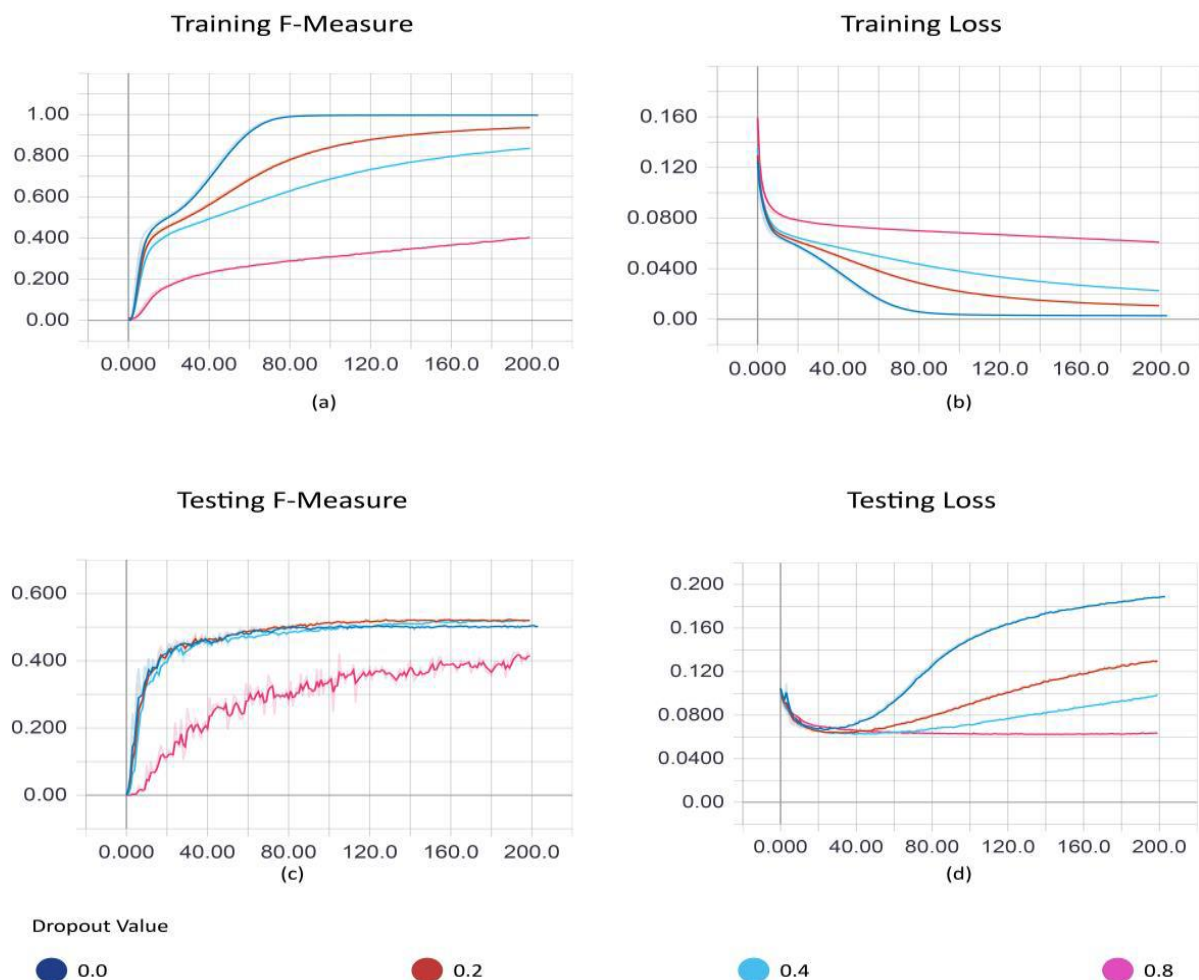


Рисунок 4 – Результаты эксперимента по тестированию значения исключающего параметра, где a – тренировочная F-мера, b – тренировочное значение потерь, c – тестовая F-мера, d – тестовое значение потерь.

Значения исключения: 0.8 (розовая линия), 0.4 (бледно-голубая линия), 0.2 (красная линия), 0.0 (синяя линия).

### Выбор оптимизатора Keras

Оптимизатор берет на себя необходимые вычисления, требующиеся для адаптации весов и переменных отклонений нейросети в ходе обучения. Эти вычисления ведут к подсчетам так называемых градиентов, которые обозначают направление необходимого изменения отклонений и весов для минимизации функции.

Разработка стабильного и быстрого оптимизатора — одна из основных задач создателей нейронных сетей. В данном случае использовались одни из наиболее распространенных оптимизаторов в сфере машинного обучения [7].

Тесты проводились на оптимизаторах Keras, каждый из которых обладает определёнными характеристиками. Для тестирования нейронной сети, были исследованы различные оптимизаторы. На Рисунке 5 представлены результаты их сравнения.

Перечень используемых оптимизаторов:

- ADAGrad (зеленая линия);
- стохастический градиентный спуск (розовая линия);
- Adam (красная линия);
- RMSProp (светло-синяя линия).

В результате наиболее подходящим для поставленных задач, был выбран оптимизатор RMSProp. Минусом данного оптимизатора является тенденцию к переобучению, так как происходит быстрое прибытие на этап обучения, где происходит переобучение.



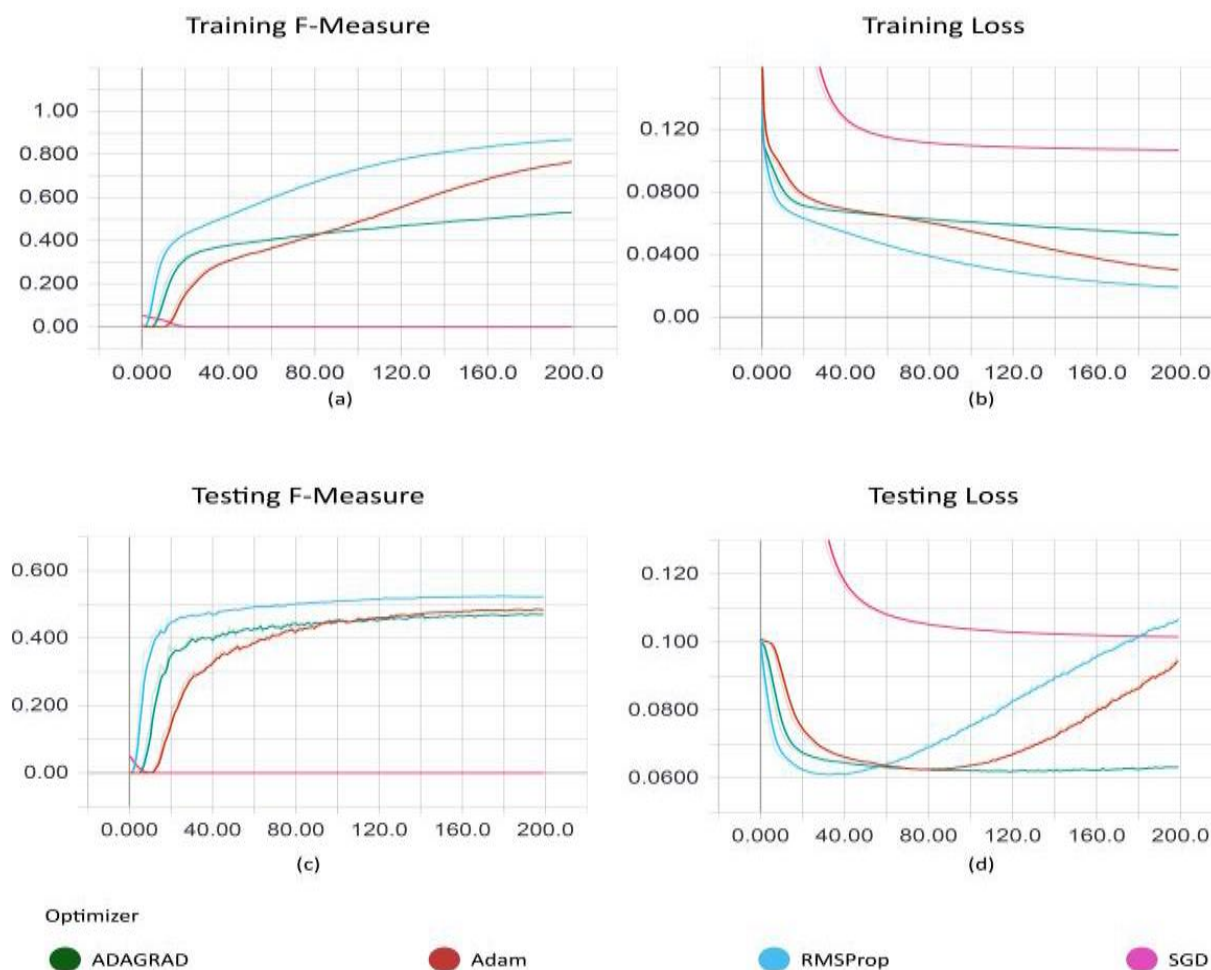


Рисунок 5 – Тестирование различных оптимизаторов, где а – тренировочная F-мера, b – тренировочное значение потерь, с – тестовая F-мера, d – тестовое значение потерь.

### Размер партии (batch size) для обучения сети

Чтобы эффективно работать с большими объемами данных, требуется использовать параметры (epoch, batch size, итерации), так как зачастую нет возможности загрузить сразу все данные в обработку. Патентная база данных представляет собой сверхбольшой информационный массив. Для преодоления проблемы его обработки данные делят на части меньшего размера, загружают их по очереди и обновляют веса нейросети в конце каждого шага, подстраивая их под данные [10].

На Рисунке 6 изображены результаты различных значений размера партии. Чем меньше размеры партии, тем модель более эффективна, но небольшие размеры партии делают обучение очень длительным.

Размер партии 512 показал лучший результат при сохранении относительно быстрого обучения.

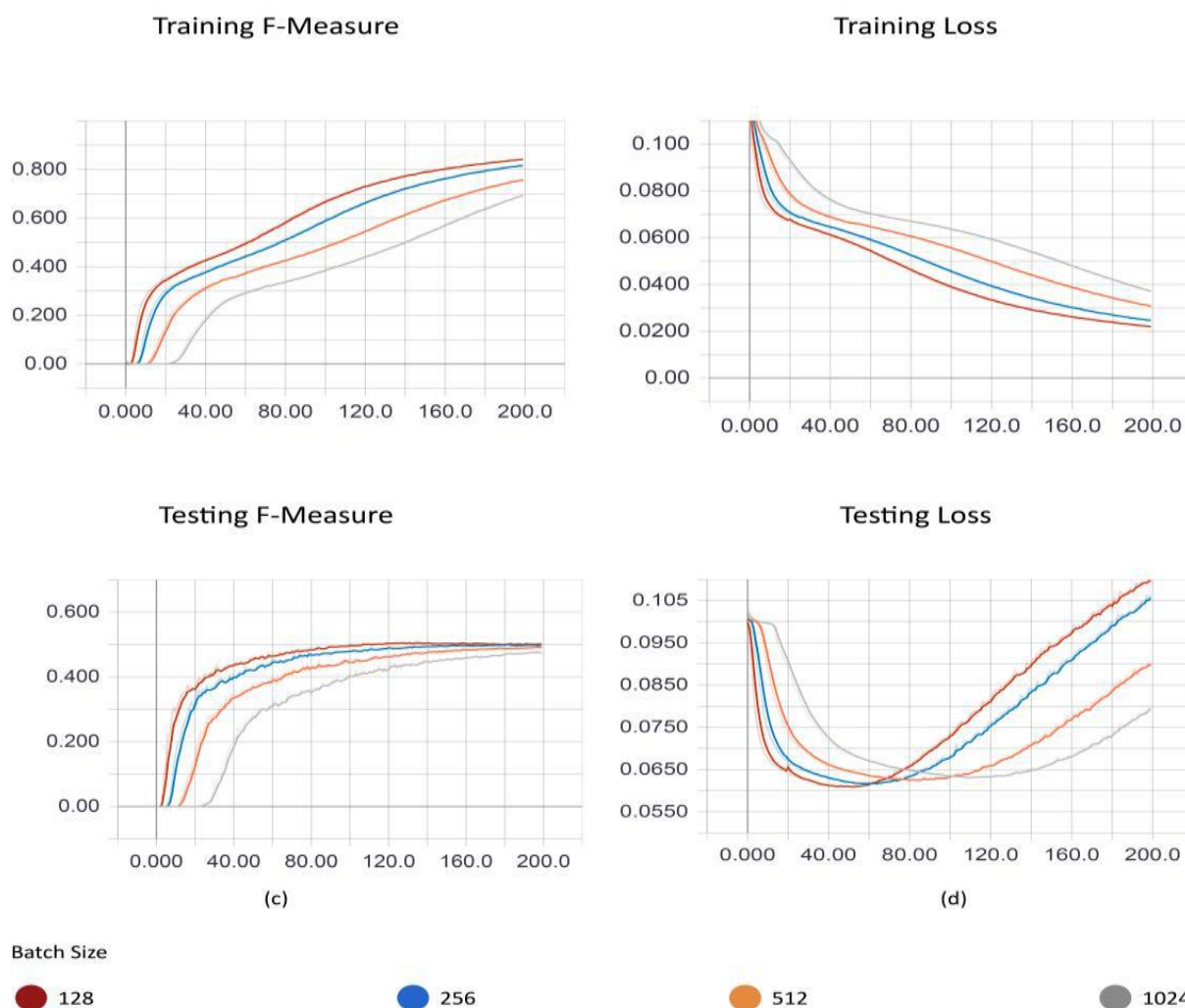


Рисунок 6 – Тестирование размера партии для обучения сети, где а – тренировочная F-мера, b – тренировочное значение потерь, с – тестовая F-мера, d – тестовое значение потерь.

### Заключение

В результате проведенного исследования были получены следующие результаты (Таблица 1):

1. Подобрана модель нейросети;
2. Выделены параметры, влияющие на работу нейронной сети;
3. Выявлен каждый параметр улучшающей результат работы нейросети;
4. Составлена оптимальная конфигурация рекуррентной сети для поставленной задачи.

Задачей анализируемой рекуррентной нейронной сети является устранение проблем, носящих междисциплинарный характер, которые проявляются в результате работы с патентными документами.

Таблица 1 - Оптимальная конфигурация рекуррентной нейронной сети

Количество слоев	Размер слоёв	Значение исключающего параметра	Размер партии для обучения сети	Выбор оптимизатора Keras
4	512	0.6	512	RMSProp

Таким образом, построенную конфигурацию рекуррентной сети можно использовать в рамках задач классификации текстов патентов и предсказания трендов технологического развития.

## ЛИТЕРАТУРА

1. Methods of Statistical and Semantic Patent Analysis / Д.М. Коробкин, С.А. Фоменков, А.Г. Кравец, С.Г. Колесников // Creativity in Intelligent Technologies and Data Science. Second Conference, CIT&DS 2017 (Volgograd, Russia, September 12-14, 2017) : Proceedings / ed. by A. Kravets, M. Shcherbakov, M. Kultsova, Peter Groumos ; Volgograd State Technical University [et al.]. – [Germany] : Springer International Publishing AG, 2017. – P. 48-61. – (Ser. Communications in Computer and Information Science ; Vol. 754).
2. Определение патента и организация прав патентообладателя [Электронный ресурс]. – Режим доступа: [http://economic-definition.com/Economic\\_and\\_legal\\_terminology/Patent\\_Patent\\_\\_eto.html](http://economic-definition.com/Economic_and_legal_terminology/Patent_Patent__eto.html) (дата обращения 01.06.2019).
3. Кондратьева Т.Н., Прогнозирование тенденции финансовых временных рядов с помощью нейронной сети LSTM // Интернет-журнал Науковедение. 2017. Т. 9. № 4. С. 61.
4. Кравец, А.Г. Patents Images Retrieval and Convolutional Neural Network Training Dataset Quality Improvement [Электронный ресурс] / А.Г. Кравец, Н.С. Лебедев, М.С. Легенченко // Proceedings of the IV International research conference «Information technologies in Science, Management, Social sphere and Medicine» (ITSMSSM 2017) / ed. by O.G. Berestneva [et al.]. – [Published by Atlantis Press], 2017. – P. 287-293. – (Ser. Advances in Computer Science Research (ACSR) ; Vol. 72). – URL :

- <https://www.atlantis-press.com/proceedings/itsmssm-17>.
5. Kline, D. M., Revisiting squared-error and cross-entropy functions for training neural network classifiers. [Электронный ресурс]. – 2005. – Режим доступа: <https://link.springer.com/article/10.1007/s00521-005-0467-y> (Дата обращения: 04.05.2019).
  6. Lance G. N., Willams W. T. A general theory of classification sorting strategies. 1. hierarchical systems // *Comp. J.* — 1967. — no. 9. — Pp. 373–380.
  7. Jain A., Murty M., Flynn P. Data clustering: A review // *ACM Computing Surveys.* — 1999. — Vol. 31, no. 3. — Pp. 264–323.
  8. Eck, D., Schmidhuber, J. A First Look at Music Composition using LSTM Recurrent Neural Networks. – Manno, Switzerland: Istituto Dalle Molle di studi sull' intelligenza artificiale. [Электронный ресурс]. – 2002. – Режим доступа: <http://people.idsia.ch/~juergen/blues/IDSIA-07-02.pdf> (Дата обращения: 04.05.2019).
  9. The Python Deep Learning library [Электронный ресурс]. – Режим доступа: <https://keras.io/> (дата обращения 10.05.2019)
  10. «Smart Queue» Approach for New Technical Solutions Discovery in Patent Applications / А.Г. Кравец, Н. Шумейко, В. Lempert, N. Salnikova, Н.Л. Щербакова // *Creativity in Intelligent Technologies and Data Science. Second Conference, CIT&DS 2017 (Volgograd, Russia, September 12-14, 2017) : Proceedings / ed. by A. Kravets, M. Shcherbakov, M. Kultsova, Peter Groumos ; Volgograd State Technical University [et al.]. – [Germany] : Springer International Publishing AG, 2017. – P. 37-47. – (Ser. Communications in Computer and Information Science ; Vol. 754).*

A.G. Kravets, A.S. Burmistrov, P.A. Zadorozhny  
**EXPERIMENTAL DETERMINATION OF THE OPTIMAL  
PARAMETERS OF THE RECURRENT NEURAL NETWORK FOR  
THE TASKS OF PATENT CLASSIFICATION**

*Volgograd State Technical University,  
Volgograd, Russia*

*Indicators of patent activity are now often used in technological forecasting and competitive intelligence. An important role is to predict the development of patent trends in individual countries and around the world, which allows to identify the main priority directions of technology development. Analysis of patents in the field of analog technology was fulfilled. The international patent classification is outdated, most studies are interdisciplinary. There is a need to select and create new classes. The purpose of this study is to analyze the parameters that affect the results of the recurrent neural network, designed for the thematic classification of the patent array. The analysis of the identified parameters affects the quality of the neural network and the selection of optimal values. The optimal parameters of the neural network were*

*determined: the number of layers, the size of the layer, the value of the exclusion parameter, the batch size for training in the network, the choice of the Keras library optimizer was made. The reported study was funded by RFBR according to the research project № 19-07-01200.*

**Keywords:** trend, classification, patent, recurrent neural network, exclusion layer, optimizer analysis, batch size.

## REFERENCES

1. Methods of Statistical and Semantic Patent Analysis / D.M. Korobkin, S.A. Fomenkov, A.G. Kravets, S.G. Kolesnikov // Creativity in Intelligent Technologies and Data Science. Second Conference, CIT & DS 2017 (Volgograd, Russia, September 12-14, 2017): Proceedings / ed. by A. Kravets, M. Shcherbakov, M. Kultsova, Peter Groumpos; Volgograd State Technical University [et al.]. - [Germany]: Springer International Publishing AG, 2017. - p. 48-61. - (Ser. Communications in Computer and Information Science; Vol. 754).
2. Determination of the patent and the organization of the rights of the patent holder [Electronic resource]. - Access mode: [http://economic-definition.com/Economic\\_and\\_legal\\_terminology/Patent\\_Patent\\_\\_eto.html](http://economic-definition.com/Economic_and_legal_terminology/Patent_Patent__eto.html) (circulation date 01.06.2018).
3. Kravets, A.G. Patents Images Retrieval and Convolutional Neural Network Training Dataset Quality Improvement [Electronic resource] / A.G. Kravets, N.S. Lebedev, M.S. Legenchenko // Proceedings of the IV International Research Conference on Information Technologies in Science, Management, Social Sphere and Medicine (ITSMSSM 2017) / ed. by O.G. Berestneva [et al.]. - [Published by Atlantis Press], 2017. - p. 287-293. - (Ser. Advances in Computer Science Research (ACSR); Vol. 72). - URL: <https://www.atlantispress.com/proceedings/itsmssm-17>.
4. Kondratieva TN, Forecasting the tendency of financial time series using the LSTM neural network // Internet journal Naukovedenie. 2017. V. 9. № 4. P. 61.
5. Kline, D.M., Revisiting squared-error and cross-entropy functions for training neural network classifiers. [Electronic resource]. - 2005. - Access mode: <https://link.springer.com/article/10.1007/s00521-005-0467-y> (Contact date: 05/04/2019).
6. Lance G. N., Willams W. T. A general theory of classification sorting strategies. 1. hierarchical systems // Comp. J. - 1967. - no. 9. - Pp. 373–380.
7. Jain A., Murty M., Flynn P. Data clustering: A review // ACM Computing Surveys. - 1999. - Vol. 31, no. 3. - Pp. 264–323.
8. Eck, D., Schmidhuber, J. A First Look at Music Composition using LSTM Recurrent Neural Networks. - Manno, Switzerland: Instituto Dalle Molle di

- studi sull 'intelligenza artificiale. [Electronic resource]. - 2002. - Access mode: <http://people.idsia.ch/~juergen/blues/IDSIA-07-02.pdf> (Revised: 05/04/2019).
9. The Python Deep Learning library [Electronic resource]. - Access mode: <https://keras.io/> (request date 10.05.2019)
10. "Smart Queue" Approach for New Technical Solutions Discovery in Patent Applications / A.G. Kravets, N. Shumeiko, B. Lempert, N. Salnikova, N.L. Shcherbakova // Creativity in Intelligent Technologies and Data Science. Second Conference, CIT & DS 2017 (Volgograd, Russia, September 12-14, 2017): Proceedings / ed. by A. Kravets, M. Shcherbakov, M. Kultsova, Peter Groumpos; Volgograd State Technical University [et al.]. - [Germany]: Springer International Publishing AG, 2017. - p. 37-47. - (Ser. Communications in Computer and Information Science; Vol. 754).