



Нейросетевой анализ публикационной активности в отечественной базе данных научного цитирования eLIBRARY

В. О. Орлов, Г. А. Карнуп, О. А. Тельминов, С. Л. Фоттеллер
АО «Научно-исследовательский институт молекулярной электроники», г. Зеленоград, Москва, Россия

Предложен алгоритм и программное обеспечение на основе работы нейронной сети для анализа работ автора, выработки рекомендаций по формированию коллектива исполнителей НИОКТР и прогнозированию достижений автора. Для апробации идеи использована отечественная база данных eLIBRARY. Связь разработанного прототипа программного обеспечения на языке Python с базой данных осуществляется через приложение API (application programming interface).

Ключевые слова: наукометрические показатели, прогноз, eLIBRARY, API, машинное обучение

Статья поступила: 14.11.2022

Статья принята: 28.11.2022

ВВЕДЕНИЕ

В настоящее время в области фотоники и микроэлектроники наблюдается активный рост в части постановки и выполнения научно-исследовательских, опытно-конструкторских и технологических работ (НИОКТР) [1]. Основной критерий научной значимости работ научного коллектива – публикационная активность. Критерий измеряется набором наукометрических показателей, общепринятых научным сообществом. Отдел научно-технической информации (ОНТИ или его производные) предприятий традици-

Neural Network Analysis of Publication Activities in the Russian eLIBRARY Science Citation Database

V. O. Orlov, G. A. Karnup, O. A. Telminov, S. L. Fotteller
Research Institute of Molecular Electronics Stock Company,
Zelenograd, Moscow, Russia

An algorithm and software based on a neural network operation are proposed for analyzing the author's work, providing recommendations for creation of the R&D team and predicting the author's achievements. The national eLIBRARY database was used for the concept validation. Connection between the developed software prototype using the Python language and the database is established through API (application programming interface).

Keywords: research chops, forecast, eLIBRARY, API, machine learning

Received on: 14.11.2022

Accepted on: 28.11.2022

INTRODUCTION

At present, there is an active growth in terms of establishment and performance of various research, development, design and process works (R&D) in the field of photonics and microelectronics [1]. The main criterion for scholarly importance of the research team's work is publication activity. The criterion is measured by a set of research chops generally accepted by the academic community. The scientific and technical information division (STID or its derivatives) of the enterprises usually supports the ongoing exploratory research and provides a comfortable environment for the work of specialists. According to its functionality, the STID, at the request of the heads of departments, selects scientists and engineers with the required experience, confirmed by their publications. Moreover, the STID can participate in consideration of requests for setting

онно поддерживает проводимые поисковые исследования и обеспечивает комфортную среду для работы специалистов. Согласно своему функционалу ОНТИ по запросу руководителей подразделений подбирает в состав команды исполнителей НИОКТР ученых и инженеров, имеющих необходимый опыт, подтверждаемый их публикациями. Также ОНТИ может участвовать в рассмотрении запросов на определение приоритетов при подаче результатов интеллектуальной деятельности к регистрации, идентификации публикационной зрелости сотрудников к защите диссертационной работы, подготовке проектов заявок на участие команды в научно-технических конкурсах.

Задачи, связанные с определением публикационной активности, слабо формализуемы. Если формализация начальных и граничных условий обязательна для разработки классических алгоритмов, то при решении слабо формализуемых задач широко применяется аппарат нейронных сетей. Основными направлениями реализации нейронных сетей являются классификация (отнесение входного образа к одному из классов) и предсказание (определение очередных значений на будущий период, вычисленных на основе уже известных значений).

Фактически нейронная сеть представляет собой нелинейное математическое выражение с эмпирически определяемыми коэффициентами, число которых составляет от сотен в простых задачах до сотен миллиардов в наиболее сложных. Основным элементом нейронной сети является цифровой аналог биологического нейрона, который взвешивает входные сигналы в соответствии с синаптическими весами (эмпирическими коэффициентами), добавляет некоторое смещение и пропускает результат через функцию активации. Такая функция может быть представлена одной из нелинейных зависимостей. Нейроны организуются в слои: входной, несколько скрытых и выходной. При наличии нескольких слоев в такой нейронной сети она называется глубокой нейронной сетью.

Отдельным вопросом является процесс обучения нейронной сети. На вход нейросети подается очередной образ, и нейросеть вычисляет результат. Далее результат сравнивается с заранее известным значением, и в случае расхождения в нейросети корректируются значения эмпирических параметров (синаптических весов). Процесс машинного обучения заключается в подстройке весов таким образом, чтобы через несколько сотен, тысяч или более итераций процент ошибок классификации

priorities when submitting the intellectual activity results for registration, identifying the publication maturity of the employees for defending a thesis work, preparing draft applications for the team participation in the research and technical competitions.

The tasks related to determination of publication activity are poorly formalized. If the formalization of initial and boundary conditions is obligatory for the development of typical algorithms, then the neural networks are widely used when solving poorly formalizable problems. The main fields of neural network implementation include classification (assigning the input image to one of the classes) and prediction (determining the next values for the future period, calculated on the basis of well-known values).

In fact, a neural network is a non-linear mathematical expression with the empirically determined coefficients, the number of which ranges from hundreds in the simple tasks to hundreds of billions in the most comprehensive ones. The main element of the neural network is a digital analogue of a biological neuron that weighs the input signals in accordance with the synaptic weights (empirical constants), adds some bias and passes the result through an activation function. Such a function can be represented by one of the non-linear dependencies. The neurons are arranged into the layers, such as input, several hidden and output ones. If there are several layers in such a neural network, it is called a deep neural network.

A separate issue is the neural network training process. The image is provided to the neural network input, and the neural network calculates the result. The result is then compared with a previously known value. In the case of discrepancy in the neural network, the values of empirical parameters (synaptic weights) are adjusted. The machine learning process consists in the weight adjustment in such a way that after several hundreds, thousands or more iterations, the classification or prediction error percentage (depending on the neural network purpose) is at the lowest possible level.

The neural network development technology for determining the publication activity criteria involves the selection of neural network architecture, namely the sequence and specifications of its layers. It is also necessary to use a labeled database with the images provided to the neural network input and the correct answer corresponding to each image.

The routine work of a STID specialist can be significantly facilitated due to automation by simplifying access to the initial scientific citation data, their subsequent processing to obtain the statistics and predictive values. In this study, the national eLIBRARY data-



или прогнозирования (в зависимости от назначения нейросети) был минимальным.

Технология разработки нейросетей для решения задачи определения критериев публикационной активности заключается в выборе архитектуры нейросети – последовательности и характеристик ее слоев. Также необходима размеченная база данных с образцами, подаваемыми на вход нейросети и соответствующим каждому образцу корректным ответом.

Рутинная работа специалиста ОНТИ может быть существенно облегчена за счет автоматизации путем упрощения доступа к предоставлению исходных данных научного цитирования, их последующей обработки для расчета статистики и прогнозных значений. В настоящем исследовании для апробации идеи использована отечественная база данных eLIBRARY, крупнейшей российской электронной библиотеки научных публикаций на русском языке в мире. Библиотека интегрирована с Российским индексом научного цитирования (РИНЦ) – аналитическим инструментом измерения публикационной активности ученых и организаций [3]. Связь разработанного прототипа программного обеспечения на языке Python с указанной базой осуществляется через программный интерфейс приложения (API – application programming interface). Из базы извлекаются организации, авторы, публикации с соответствующей аналитикой, наукометрические показатели. В ходе исследования методов машинного обучения получены результаты по указанным выше видам прогнозов.

В итоге разработан инструмент – программное средство для сотрудника ОНТИ, реализующий описанный выше функционал для данных, полученных через API eLIBRARY в режиме тестового доступа. Удобным функционалом может являться и голосовое управление [4]. В дальнейшем планируется развить эти исследования, используя инновационный инструментарий в виде грантовой поддержки.

ОПРЕДЕЛЕНИЕ *i*-ИНДЕКСА АВТОРА НА ОСНОВЕ АНАЛИЗА ДАННЫХ В eLIBRARY.RU

В основе системы eLIBRARY лежит библиографическая реферативная база данных, в которой индексируются статьи, публикуемые в российских научных журналах, доклады конференций, монографии, учебные пособия, патенты, диссертации. База содержит сведения о выходных данных статей, авторах публикаций, местах их работы, ключевых

base, the world's largest Russian electronic library of academic publications in Russian, was used for concept verification. The library is integrated with the Russian Science Citation Index (RSCI), an analytical tool for measuring the publication activity of researchers and organizations [3]. Connection between the developed software prototype using the Python language and this database is established through API (application programming interface). The entities, authors, publications with the relevant analytics, research chops are drawn from the database. During the study of machine learning methods, the results were obtained in terms of the above predictions.

As a result, we have developed a tool, namely a software tool for a STID employee that implements the functions described above for data received through the eLIBRARY API in the test access mode. The voice-activate control can also be a convenient feature [4]. In the future, it is planned to develop these studies using the innovative tools in the form of a grant support.

DETERMINATION OF THE AUTHOR'S *i*-INDEX BASED ON DATA ANALYSIS IN eLIBRARY.RU

The eLIBRARY system is based on a bibliographic abstract database that indexes the articles published in the Russian academic periodicals, conference papers, monographs, textbooks, patents, or academic dissertations. The database contains information about the article imprints, authors of publications, places of their work, keywords and domains, as well as the abstracts and references. In the RSCI interface, it is possible to simultaneously obtain the number of publication citations in the RSCI, Web of Science and Scopus.

At present, eLIBRARY contains the texts of more than 38 million academic publications and patents [3]. The specified capacity is more than adequate for training of a neural network that analyzes the publication activities. However, there is an issue on selection of the aggregate author's activity criterion that makes it possible to take a decision on his/her reasonable inclusion in the professional R&D team.

A generally accepted measure is the eLIBRARY main research chop value, namely the Hirsch index, originally introduced to assess the scientific efficiency of physicists. The index is based on the distribution of citations of the given researcher's papers.

We propose to use a complex factor, namely an *i*-index (smart index) that includes the following author and publication specifications in eLibrary (Table 1).

The resulting *i*-index is obtained by weighing each of the above specifications with the subsequent sum-

словах и предметных областях, а также аннотации и списки литературы. В интерфейсе РИНЦ можно увидеть одновременно число цитирований публикации в РИНЦ, Web of Science и Scopus.

В настоящее время в eLIBRARY содержатся тексты более 38 миллионов научных публикаций и патентов [3]. Указанный объем более чем достаточен для обучения нейросети, реализующей анализ публикационной активности. Однако встает вопрос о выборе совокупного критерия активности автора, позволяющего принять решение о его обоснованном включении в команду специалистов, реализующих конкретный проект или НИОКТР.

Общепринятой мерой является величина основного наукометрического показателя eLIBRARY – индекса Хирша, первоначально введенного для оценки научной продуктивности физиков. Индекс вычисляется на основе распределения цитированных работ данного исследователя.

Мы предлагаем к использованию комплексный показатель – *i*-индекс (умный индекс), включающий в себя следующие характеристики автора и публикации в eLibrary (табл. 1).

Результирующий *i*-индекс получается взвешиванием каждой из перечисленных выше характеристик с последующим суммированием. Впоследствии планируется переход от специализации автора к более обобщенному анализу путем проведения качественной оценки аффилиаций и соавторов статьи, а также выполнения синтаксического анализа полнотекстовых публикаций для составления семантического ядра по каждому автору.

Дальнейшее развитие оценки направлено на учет теоретических разработок – на основе анализа публикаций, прикладных решений – на основе анализа патентов, а также применимости – на основе анализа использования патентов в государственных и коммерческих заказах. Возможен учет и результаты оценки публикаций с помощью системы плагиата.

Выбор весовых коэффициентов для теоретических разработок, прикладных решений и применимости позволит выполнить построение научно-практико-внедренческого портрета автора и соответствующее ранжирование для формирования команды НИОКТР выбранного профиля.

mation. Subsequently, it is planned to move from the author's specialization to a more generalized analysis by a qualitative assessment of affiliations and article's co-authors, as well as a syntactic analysis of full-text publications to compile a semantic kernel for each author.

Further development of the assessment is aimed at consideration of theoretical insights based on the analysis of publications, applied solutions based on the analysis of patents, as well as applicability based on the analysis of the patent application in the public and commercial orders. It is possible to record and evaluate publications using the plagiarism system.

The selection of weight coefficients for theoretical insights, applied solutions and applicability will make it possible to obtain an academic, practical and innovative portrait of the author and the relevant ranking for the R&D team formation with the required profile.

DEVELOPMENT OF A TOOL FOR PREDICTIVE ANALYSIS OF THE AUTHOR'S PUBLICATION ACTIVITY

The prediction of the author's publication activity is possible based on the feature analysis of publications available in eLIBRARY. The specifications/features under consideration include the number of publications, authors and their affiliations, the source rating, the RSCI parameters and the RSCI kernel.

The paper describes a developed tool for predictive analysis of the author's publication activity. Moreover, a neural network analysis of the above specification values over the past 5 years is performed, on the basis

Таблица 1. Характеристики автора и публикации в eLibrary
Table 1. Specifications of the author and publication in eLibrary

Характеристики автора Specifications of the author	Характеристики публикации Specifications of the publication
основная специализация main specialization; наличие ученой степени availability of academic degree; период деятельности period of activity; количество качество аффилиаций quantity / quality of affiliations; количество соавторов number of co-authors	количество публикаций number of publications; ссылочная масса reference weight; объем публикаций scope of publications; тип публикаций (тезисы, доклады, статьи и т. д.) type of publications (abstracts, reports, articles, etc.); рейтинг источника source rating; количество аффилиаций в публикации number of affiliations in the publication; количество соавторов в публикации number of co-authors in the publication

РАЗРАБОТКА ИНСТРУМЕНТА ПРЕДИКТИВНОЙ АНАЛИТИКИ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ АВТОРА

Прогноз публикационной активности автора возможен на основе анализа характеристик публикаций, имеющих в eLIBRARY. В качестве рассматриваемых характеристик используются число публикаций авторов и их аффилиаций, рейтинг источника, параметры РИНЦ и ядра РИНЦ.

В работе разработан инструмент предиктивной аналитики публикационной активности автора. При этом производится нейросетевой анализ значений указанных выше характеристик за последние 5 лет, на основе которого предсказывается количество публикаций на год, следующий за анализируемыми. Структура данных для каждого автора по одной теме представлена в табл. 2, при этом каждая характеристика темы усредняется за год и заносится в таблицу.

Подобные таблицы создаются для каждого автора. Важно, что темы направлений публикаций у каждого автора должны совпадать с учетом очередности. Если некоторый автор не имеет публикаций в какой-либо тематике, то все значения соответствуют нулевым.

Для обучения нейросети необходимо подготовить обучающую и валидационную выборку. Обе выборки формируются из базы eLIBRARY.RU. Причем первая выборка (обычно 80% от обрабатываемых данных) используется для настройки параметров нейросети, а вторая – для оценки корректности работы обученной на первой выборке нейросети.

Структура обучающей выборки состоит из слоев. Каждый слой является матрицей, каждый столбец которой содержит значения параметров для конкретной тематики в конкретный год. Таким образом, если в таблице представлены 5 направлений, то, с учетом указанных параметров, матрица имеет размер 8×5 . Подобные слои формируются для каждого года и объединяются в один тензор входных данных. Более подробно структура входных данных показана на рисунке.

Обучение нейросетевого алгоритма построено таким образом, что на вход подается тен-

of which the number of publications for the year following the analyzed one is predicted. The data structure for each author in terms of one topic is provided in Table 2, while each topic specification is averaged over the year and entered in the table.

The similar tables are generated for each author. It is important that the themes of publications for each author should coincide with due regard to the priority. If some author has no publications in any area, then all values are equal to zero.

To train the neural network, it is necessary to prepare a training and validation set. Both sets are developed using the eLIBRARY.RU database. Moreover, the first set (usually 80% of the processed data) is used to configure the neural network parameters, and the second one is used to assess the work soundness and correctness of the neural network trained using the first set.

The training set structure includes the layers. Each layer is a matrix, each column of which contains the parameter values for a specific topic during a specific year. Thus, if the table contains 5 areas, then the matrix has a size of 8×5 with due regard to the indicated parameters. Similar layers are generated for each year and combined into one input data tensor. The input data structure is shown in more detail in the figure.

Training of the neural network algorithm is performed in such a way that the input data tensor for 5 years is provided and the vector of publications for the year following the specified period is predicted. For example, if there is information in relation to the author for 7 years (Table 1), then data for 1-5 years can be submitted to the neural network and the features for the 6th year can be predicted. Then, based on the

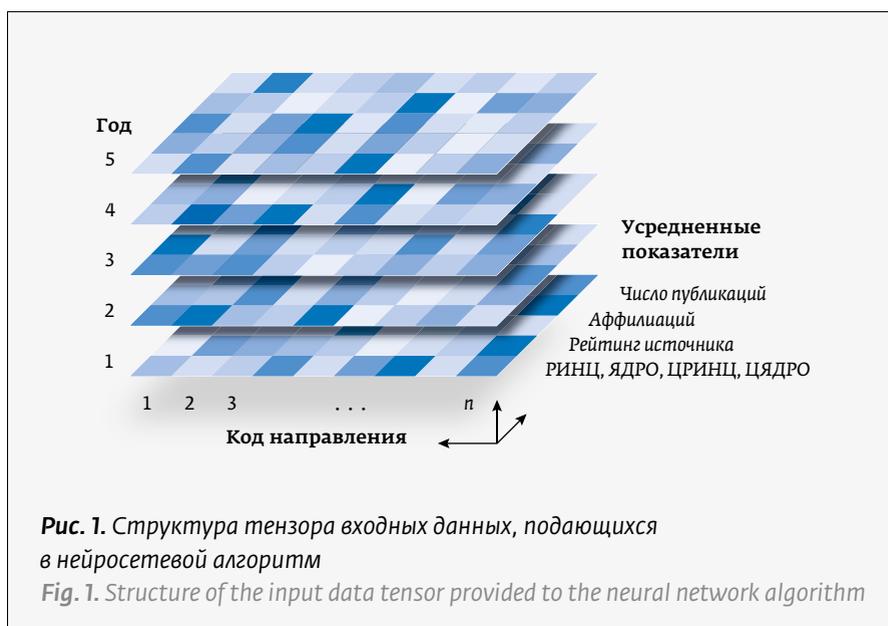


Рис. 1. Структура тензора входных данных, подающихся в нейросетевой алгоритм

Fig. 1. Structure of the input data tensor provided to the neural network algorithm

Таблица 2. Структура данных автора по определенной теме

Table 2. Structure of author's data in a specific area

Характеристика темы Topic specification	Анализируемый период Analyzed period						Прогно- зируемый период Forecast period
	1-й год Year 1	2-й год Year 2	3-й год Year 3	4-й год Year 4	5-й год Year 5	6-й год Year 6	
Число публикаций по теме Number of publications per topic	0	1	9	7	3	4	4
Среднее число авторов Average number of authors	0	2	4,56	5,86	4,67	4,5	4
Рейтинг Rating	0	1	5,1	5,36	5,37	8,71	20,86
Число аффилиаций Number of affiliations	0	1	2,22	3,71	2,67	3,25	3,25
РИНЦ (Российский индекс науч- ного цитирования) RSCI	0	1	1	1	0,33	0,75	1
ЦРИНЦ (количество цитирований публикаций из РИНЦ) C_{RSCI}	0	0	0,66	1	0,33	0,25	1
ЯДРО РИНЦ (собрания лучших оте- чественные журналов, составляю- щих основу российских изданий) KERNEL	0	0	3,33	7,71	4,33	0,5	0,25
ЦЯДРО РИНЦ (количество цитиро- ваний публикаций из ядра РИНЦ) C_{KERNEL}	0	0	2,67	5,71	3	2	0,25

Примечание: в строке «Число публикаций» представлено суммарное число публикаций за год по данной тематике; далее приведе- ны средние арифметические параметров за каждый год. В столбце «Прогнозируемый период» целевой переменной является только «Число публикаций», которая является составляющей целевого вектора - число публикаций по каждой тематике

зор входных данных за 5 лет и предсказывается вектор публикаций на следующий за указанным периодом год. Например, если для автора имеет информация за 7 лет (табл. 1), то в нейросеть можно подать данные за 1-5-й годы и предсказать характеристики за 6-й год. Затем на основе характеристик со 2-го по 6-й годы предсказать характеристики 7-го года. Подобный метод менее точен, чем если бы предсказание производилось на основе всей публикационной истории, однако он позволяет сохранять универсальность относительно количества лет для анализа и, учитывая малое количество авторов, увеличивает количество данных для обучения. В нашем случае набор данных состоял из 7 авторов и суммарно 98 тензоров. Из них 88 тензоров составляли обучающую выборку и 10 - валидационную. Всего публикационных тем 8, однако около 60% данных были представ-

specifications of the 2-6th years, it is possible to predict the features of the 7th year. Such a method is less accurate than prediction on the basis of the entire publication history, however, it allows to maintain versatility regarding the number of years to be analyzed. Given the small number of authors, it increases the amount of data for training. In our case, the data set consisted of 7 authors and a total of 98 tensors. Among them, 88 tensors were included in the training set and 10 tensors were included in the validation set. There were 8 publication topics in total, however, about 60% of data were provided in only 3 areas. However, this does not affect the algorithm, since the topics can be considered independent and their number leads only to the changes in the input and output tensor values.

Due to the input data specificity, namely, their tensor and matrix type, the first 2 input layers of the neural network architecture are convolutional. After each



лены только в 3 темах. Тем не менее, это не влияет на работу алгоритма, т.к. темы можно считать независимыми и от их количества изменятся только размеры входного и выходного тензоров.

В связи со спецификой входных данных, а именно их тензорно-матричного вида, первые 2 входных слоя нейросетевой архитектуры являются сверточными. После каждого слоя выполняется нормализация данных, в качестве активационной функции используется функция ReLU. Размер ядра свертки - 3, шаг - 1, паддинг - 1. После сверточных слоев следуют 3 полносвязных слоя с той же функцией активации. Размер выходного слоя соответствует числу тем, представленных в таблице данных. Так как по выходе из нейросети вектор состоит из дробных значений, для большего соответствия реальным данным его необходимо округлить. Нейросетевая архитектура была сформирована и обучена в программном пакете PyTorch. Количество эпох обучения - 30, оптимизатор - Adam с шагом оптимизации 0.005, ошибка считалась по формуле среднеквадратической ошибки MSE.

Средняя точность разработанного инструмента предиктивной аналитики публикационной активности автора составляет 71%, причем наблюдается тенденция занижения реального значения публикаций на 1-2 единицы. Для получения более высокого процента точности необходимо увеличить объем обучающей выборки, обратив внимание на авторов с более длительными периодами публикации по определенной теме.

Применение нашего инструмента возможно также для предсказания публикационной деятельности автора по выбранной теме и при оценке его возможностей.

ОБЛАСТИ ПРИМЕНЕНИЯ ПРОГНОЗА ПУБЛИКАЦИОННОЙ АКТИВНОСТИ АВТОРА

Предложенные нами i-индекс и инструмент предиктивной аналитики публикационной активности автора имеют широкий спектр применения.

Во-первых, индекс применим для подбора команды по выполнению НИОКТР или других видов научных работ. В этом случае формируется поисковый запрос с применением ключевых слов по теме проекта, а также исключаемых слов - автора, брендов и др. На основе поискового запроса формируется семантическое ядро, сравнивается с семантическими ядрами авторов из системы. При этом основными критериями для оценки будет общей i-индекс исходя из всей

layer, the data is normalized, and the ReLU function is used as an activation function. The convolution kernel size is 3, the stride is 1, and the padding is 1. After the convolution layers, there are 3 fully connected layers with the same activation function. The output layer size corresponds to the number of topics shown in the data table. Since the vector consists of fractional values upon exiting the neural network, it should be rounded to better match the actual data. The neural network architecture was generated and trained in the PyTorch software package. The number of training epochs was 30, the optimizer was Adam with an optimization step of 0.005. The error was calculated using the MSE mean square error formula.

The average accuracy of the developed tool for predictive analytics of the author's publication activity is 71%. Moreover, there is a tendency to underestimate the actual value of publications by 1-2 units. To obtain a higher accuracy percentage, it is necessary to increase the training set size by paying attention to the authors with the longer publication periods on a particular topic.

Our tool can also be used to predict the author's publication activity in the selected area and to assess his/her capabilities.

SCOPE OF THE AUTHOR'S PUBLICATION ACTIVITY FORECAST

The i-index and predictive analytics tool of the author's publication activity proposed by has have a wide range of applications.

Firstly, the index is applicable to the selection of a team for implementation of R&D or other types of academic work. In this case, a search query is generated using the keywords related to the project topic, as well as excluded words, such as the author, brands, etc. Based on the search query, a semantic kernel is formed and compared with the semantic kernels of authors from the system. Moreover, the main criteria for evaluation will be the general i-index based on the entire author's activities and i-index for the semantic kernel, i. e. the weight of competencies for a specific search query. In addition to analyzing own entity's employees for participation in the certain R&D, it is possible to search for employees of other entities with whom it is possible to establish a consortium for joint performance of such works. It is also possible to estimate changes in the author's interest vector in relation to various topics over time.

Secondly, it is possible to assess the degree of author's maturity for earning the next degree with the step of one year. Thus, on the basis of a set previously obtained from eLIBRARY.RU, the i-index proposed by

деятельности автора и *i*-индекс по семантическому ядру, т. е. вес компетенций по конкретному поисковому запросу. Кроме анализа сотрудников своей организации на предмет участия в тематических НИОКТР, возможен поиск сотрудников других организаций, с которыми возможно создание консорциума для совместного выполнения таких работ. Также возможна оценка изменения вектора интереса автора к различным тематикам с течением времени.

Во-вторых, для оценки степени зрелости автора к получению очередной ученой степени с шагом в один год. Так на основании полученной ранее выборки из eLIBRARY.RU выполнен расчет предложенного нами *i*-индекса. Оказалось, что для авторов с ученой степенью доктора технических наук среднее значение такого индекса равно 4,325, для кандидатов технических наук – 3,753. Также можно получить значения медианы или другого граничного значения, с определенной степенью вероятности разделяющего выборку на градации по ученым степеням. Применяя разработанный нами и описанный выше инструмент предиктивной аналитики, можно оценить степени зрелости автора к получению очередной ученой степени с шагом в один год.

Кроме того, сместив акцент с автора на издание, можно получить рейтинг изданий, в которых наиболее часто публикуются работы по заданной теме. Что интересно само по себе, потому что позволяет ранжировать издания по их доступности, кругу авторов, широте представляемых организаций.

ЗАКЛЮЧЕНИЕ

Предложен совокупный индекс публикационной активности – *i*-индекс, а также разработан инструмент предиктивной аналитики публикационной активности автора на основе машинного обучения глубоких нейронных сетей. Приведены кейсы применения инструмента для целевой работы с авторами-сотрудниками организации, привлечения в НИОКТР авторов-сотрудников других организаций при организации консорциумов для выполнения таких работ, оценке степени зрелости автора к участию в конкурсах на предоставление грантовой поддержки.

АВТОРЫ

Орлов В. О., vorlov@niime.ru, ORCID 0000-0003-0591-4664
 Карнуп Г. А., ORCID 0000-0001-6517-4712
 Тельминов О. А. к. т. н., ORCID 0000-0002-2358-3689
 Фоттеллер С. Л. к. ф. н., АО «НИИМЭ», Москва, г. Зеленоград, Россия,
 ORCID: 0000-0003-3346-5229

us was calculated. It turned out that for the authors with an academic degree of the doctor of technical sciences, the average value of such an index is 4.325, for Ph. Ds in technical sciences – 3.753. It is also possible to obtain the median values or other boundary values dividing the set into the groups by academic degrees with a certain degree of probability. Using the predictive analytics tool developed by us and described above, it is possible to assess the degree of author's maturity for earning the next academic degree with the step of one year.

In addition, by changing the focus from the author to the publication, it is possible to get a rating of publications where the works on a given topic are most often published. It is inherently interesting, since it allows to rank publications according to their availability, the pool of authors, the range of entities represented.

CONCLUSION

A cumulative index of publication activity, *i*-index, has been proposed. Moreover, we have developed a tool for predictive analytics of the author's publication activity based on the machine learning of deep neural networks. We have provided the cases of using this tool for targeted work with the entity's employees, involvement of the employees of other organizations in the R&D work when arranging the consortiums, assessment of the degree of author's maturity to participate in the grant support tenders.

REFERENCES

1. Krasnikov G. YA., Gornev E. S., Matyushkin I. V. Obshchaya teoriya tekhnologii i mikroelektronika: chast' 3. Uroven' tekhnologicheskoy operacii. *Elektronnaya tekhnika. Seriya 3: Mikroelektronika*. 2018; 3(171):63–93.
 Красников Г. Я., Горнев Е. С., Матюшкин И. В. Общая теория технологий и микроэлектроника: часть 3. Уровень технологической операции. *Электронная техника. Серия 3: Микроэлектроника*. 2018; 3(171):63–93.
2. Orlov O. et al. Features of the Improving of the HFO2 ReRAM Cells Characteristics Reproducibility. *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*. 2019. <https://doi.org/10.1109/eiconrus.2019.8657230>.
3. Tel'minov O. A., Gornev E. S., Teplov G. S., Moshkarova L. A. Varianty realizacii nejroprocessora dlya raspoznavaniya rechi. Sostoyanie i perspektivy razvitiya sovremennoj nauki po napravleniyu. *Informatika i vychislitel'naya tekhnika*. 2020; 57–68.
 Тельминов О. А., Горнев Е. С., Теплов Г. С., Мошкаророва Л. А. Варианты реализации нейропроцессора для распознавания речи. Состояние и перспективы развития современной науки по направлению. *Информатика и вычислительная техника*. 2020; 57–68.
4. eLIBRARY.RU. URL: <https://elibrary.ru> (data obrashcheniya 21.11.2022 g.).

AUTHORS

Orlov V. O., ORCID 0000-0003-0591-4664
 Karnup G. A., ORCID 0000-0001-6517-4712
 Telminov O. A., Ph.D. ORCID 0000-0002-2358-3689
 Fotteller S. L., Ph.D., Research Institute of Molecular Electronics SC, Zelenograd,
 Moscow, Russia.
 ORCID: 0000-0003-3346-5229