

Научная статья  
УДК 159.9:343.8

## ЭМБЕДДИНГ КАК МЕТОД ИССЛЕДОВАНИЯ В ЮРИДИЧЕСКОЙ ПСИХОЛОГИИ

Михаил Иванович Марьин<sup>1</sup>, Денис Алексеевич Курдин<sup>2</sup>

<sup>1</sup> Московский государственный психолого-педагогический университет, г. Москва, Россия, [marin\\_misha@mail.ru](mailto:marin_misha@mail.ru), <https://orcid.org/0000-0003-1142-8857>

<sup>2</sup> Академия ФСИН России, г. Рязань, Россия, [dnskurdin@gmail.com](mailto:dnskurdin@gmail.com), <https://orcid.org/0000-0002-9292-7342>

**Аннотация.** В статье рассматривается проблема применения психологами психодиагностических методов. Приводятся данные о возможности цифровизации методов психодиагностики и применения искусственного интеллекта. В качестве инновационной методики используется эмбединг – метод обработки естественного языка, в котором слова представляются в виде числовых векторов. Описываются результаты пилотажного исследования с целью проверки возможности эмбединга для применения в пенитенциарной психологии. Основная гипотеза исследования была представлена в виде тезиса о целесообразности использования эмбединга в качестве метода в психологических исследованиях, отвечающего требованиям методов изучения психических явлений, которые должны быть воспроизводимы, алгоритмичны и адекватны предмету науки психологии, а также должны раскрывать особенности исследуемого психологического процесса.

**Ключевые слова:** эмбединг, психологические исследования, метод, искусственный интеллект, искусственная нейронная сеть

### Для цитирования

Марьин М. И., Курдин Д. А. Эмбединг как метод исследования в юридической психологии // Векторы психолого-педагогических исследований. 2023. № 1(01). С. 118–125.

Original article

## EMBEDDING AS A RESEARCH METHOD IN LEGAL PSYCHOLOGY

Mihail Ivanovich Mar'in<sup>1</sup>, Denis Alekseevich Kurdin<sup>2</sup>

<sup>1</sup> Moscow State Psychological and Pedagogical University, Moscow, Russia, [marin\\_misha@mail.ru](mailto:marin_misha@mail.ru), <https://orcid.org/0000-0003-1142-8857>

<sup>2</sup> Academy of the FPS of Russia, Ryazan, Russia, [dnskurdin@gmail.com](mailto:dnskurdin@gmail.com), <https://orcid.org/0000-0002-9292-7342>

**Abstract.** The article deals with the problem of the use of psychodiagnostic methods by psychologists. The data on the possibility of digitalization of psychodiagnostic methods and the use of artificial intelligence are presented. Embedding, a natural language processing method in which words are represented as numeric vectors, is used as an innovative technique. The results of a pilot study are described in order to test the possibility of embedding for use in penitentiary psychology. The main hypothesis of the study was presented in the form of a thesis on the expediency of using embedding as a method in psychological research that meets the requirements of methods for studying mental phenomena, which must be reproducible, algorithmic and adequate to the subject of the science of psychology, as well as reveal the features of the studied psychological process.

**Keywords:** embedding, psychological research, method, artificial intelligence, artificial neural network

### For citation

Mar'in, M. I. & Kurdin, D. A. 2023, 'Embedding as a research method in legal psychology', *Vectors of psychological and pedagogical research*, iss. 1(01). pp. 118–125.

### Введение

Цифровая обработка данных позволяет открывать новые связи, которые обычные методы не в состоянии уловить. Слова «интеллектуальный анализ данных» (data mining), «большие данные» (big data), «нейронные сети» уже не принадлежат только математикам и программистам, а постепенно входят в психологический дискурс. Основанные на таких понятиях методы становятся частью психологических исследований в России и за рубежом. Это объективно назревшая ситуация, обусловленная накоплением больших объемов психологической информации и несовершенством традиционных методов исследования (тестирование, экспертная оценка и др.).

Следует обратить внимание на некорректное применение психологами психодиагностических методов, неактуальные стимулы в методиках. Данный тезис также подтверждается рядом ученых. Например, А. С. Эльзессер, Т. В. Капустина пишут: «Объективные причины широкого распространения ненадежных методик: использование устаревших методик; доступность методик; использование методик специалистами непсихологических направлений; механическое использование интерпретатора методики специалистами-психологами – все это снижает диагностические возможности психологических методик» [4].

Перечисленное подтверждает необходимость использования достижений цифровой реальности для развития методов психодиагностики. В силу этого обратимся к методам, основанным на искусственном интеллекте. Согласно приказу Министерства экономического развития Российской Федерации от 29 июня 2021 г. № 392 «Об утверждении критериев определения принадлежности проектов к проектам в сфере искусственного интеллекта» к технологиям искусственного интеллекта относятся технологии, основанные на использовании искусственного интеллекта, включая: а) компьютерное зрение; б) обработку естественного языка; в) распознавание и синтез речи; г) интеллектуальную поддержку принятия решений; д) перспективные методы искусственного интеллекта.

Одним из распространенных методов зачаточного искусственного интеллекта является искусственная нейронная сеть (ИНС). Это метод в искусственном интеллекте, математическая модель, алгоритм, а также программное воплощение, построенные по принципу организации и функционирования биологических нейронных сетей. ИНС способна к обучению и самоорганизации. Примеры использования ИНС довольно широки: от обучения вождению автомобилей до применения в голосовых помощниках. Описать более подробную систему построения и функционирования рассматриваемого математического алгоритма не позволяют рамки настоящей статьи, однако следует учесть, что данный метод обладает большой перспективой, а свойства ИНС при распознавании связей могут использоваться в психодиагностике [3].

Итак, как указывалось ранее, к технологиям искусственного интеллекта относится обработка естественного языка. Ее основным инструментом является так называемый эмбединг – метод, основанный на ИНС, результатом работы которого является цифровизация смысла некоего сообщения. Эмбединг – это представление слова в виде вектора, оцифровывание его смысла, который можно представить графически.

Единицей смысла в конкретном предложении выступает слово в составе синтаксемы. Смысл всего высказывания передается с помощью сложных конструкций, образуемых отдельными синтаксемами. Смысл текста, с точки зрения описываемого подхода, восстанавливается путем выделения в нем отдельных синтаксем, установления их значений и определения ролей для таких предикатов с помощью векторных представлений слов (эмбедингов) [1].

В русскоязычной литературе эмбедингами обычно называют числовые векторы, которые получены из слов или других языковых средств. Числовым вектором размерности  $k$  называют список из чисел  $k$ , в котором порядок чисел строго определен. Например, трехмерным вектором можно считать (2.3, 1.0, 7.35), а (1, 0, 0, 2, 0.1, 0, 0, 7.9) – восьмимерным числовым вектором.

Эмбединг – метод обработки естественного языка, в котором слова представляются в виде числовых векторов. Он представляет собой матрицу  $n$  на  $m$  чисел. Чисел в этой матрице может быть от нескольких сотен до нескольких тысяч, все зависит от модели для эмбединга.

Ранее анализ слова компьютером был невозможен без его распознавания в буквенном выражении, что предполагало подсчет знаков, пробелов и математический анализ этих данных. Однако в таком случае опускается смысл полученных данных, он теряется при их обработке. Например, смысл слов «петух», «курица», «цыпленок» человеку понятен и легко им объединяется смысловую группу. Однако начертание букв этих слов различно, например, омонимы создают определенную проблему для распознавания машиной (сушка как процесс и как кондитерское изделие). Данная проблема не решается простым кодированием букв, слов или словосочетаний в цифры, машина будет интерпретировать их одинаково.

Визуально объяснить разницу и продемонстрировать эмбединг возможно следующим образом. Если учесть, что эмбединг (от англ. embedding) – вложение, то возможно продемонстрировать данное вложение, то есть оцифрованный смысл слова. Например, слово «печь» обладает следующим вектором (рис. 1). В результате программного преобразования возможно спроецировать данные векторы на двумерную сетку координат (рис. 2).

У слов с одинаковым написанием могут быть разные смыслы. На рисунке 2 изображен пример проецирования смысла слова «печь» в разных контекстах: а) «Я буду печь пирожки»; б) «Я буду топить печь пирожками». Координаты слова «печь» находятся на противоположных полюсах. Другие слова, например, местоимения «Я» и глагол «буду», примерно в тех же областях, хотя последние не совпадают, поскольку из-за контекста также могут менять часть смысла. Слова «пирожки» и «пирожками» имеют разные, но относительно схожие смыслы, поэтому на координатной сетке они располагаются недалеко друг от друга.

[-0.01136075	-0.23160739	0.9454009	-0.0325595	-0.06851641	-0.01945336
0.8060865	-0.20493323	0.56193864	-0.55052423	-0.5562415	-0.3427965
-0.13062802	0.10840102	-0.34030777	0.5869993	0.32083762	0.6519262
0.44237193	0.13461974	-0.02100517	-0.3341292	-0.2556192	-0.53779936
-0.24333915	0.13010544	-0.81313336	0.55729985	0.6112134	0.21635778
0.28403178	0.8979656	0.42633978	0.43203092	-0.06938773	0.3434822
-0.29517195	0.377868	0.02959618	-0.10662676	-0.46310398	0.5064578
-0.21831648	0.29672262	-0.6638596	0.67649764	-0.19365735	0.83977824
-0.15203284	0.16496597	0.18969403	-0.24178664	-0.35380816	-0.02162013
-0.2216356	0.15699024	-0.20823078	-0.13563062	0.1496006	-0.03847416
-0.20554513	-0.646308	-0.627595	0.49312344	-0.06138189	0.1752066
-0.01280426	0.02406607	0.6251611	0.04228494	-0.24708371	-0.5171094
-0.1819439	-0.01144756	-0.4736475	-0.3675387	0.03941553	-0.21880088
0.13196704	0.22791919	-0.24038093	0.052205	0.01526532	0.52280945
0.11467616	0.30615494	-1.0485817	-0.01136648	0.31311217	-0.03598732
-0.37048918	-0.34068152	-0.7145787	-0.10016213	-0.02910835	0.14315982
-0.5800323	0.10635779	0.5868688	0.11425438	0.17850073	-0.17582823
-0.32198656	0.35593438	-0.19783798	-0.23833905	-0.05137948	0.27994165
-0.22603892	0.9018936	0.09616869	0.29477313	0.21424869	-0.08133722
-0.27219442	0.52154976	-0.14086929	-0.14238606	-0.74904436	0.09587999
0.54171103	0.13279387	-0.42780094	0.21083619	0.06117271	0.14243321
0.36555803	-0.45834267	-0.12591943	0.54371184	0.2242603	-0.17780629
-0.1518071	-0.07810455	-0.19518237	-0.17220351	0.5570392	-0.3996402
-0.15002681	0.15171698	0.36455926	0.26764816	-0.35490745	-0.62583005
-0.39296404	-0.3843748	0.3258819	-0.44798023	-0.07022218	0.5034997
-0.28445444	-0.18137892	-0.5113458	0.13640803	0.02058944	0.34832853
0.2502752	-0.52534944	-0.32761493	-1.0229346	0.26475555	0.49043855
-0.13620429	0.22802232	0.80917686	0.2318206	0.15149085	-0.09536232
-0.21408992	0.22943394	0.19578943	-0.23845348	0.20397569	-0.8505013
-0.00463229	0.05659954	0.4919043	0.57488394	0.6192224	0.25394958
0.26346436	-0.8641609	0.04372343	-0.02508703	0.59688526	-0.7244796
0.899172	-0.15629067	-0.46846405	0.05843081	-0.27881682	0.19161755
0.50260127	0.22581875	-0.25244972	-0.18729049	-0.5265474	-0.790254
-0.09349646	-0.09702313	-0.01423997	0.21044013	1.0370885	-0.03100438
0.38660342	0.11151879	0.09306628	-0.50090027	0.36814496	-0.08227072
-0.25314903	0.34779468	0.14598536	0.2531973	-0.47382805	-0.2539363
0.23289436	0.00794245	0.31617126	0.09212031	-0.08804523	-0.21115129
0.32714072	-0.6576086	-0.00309606	-0.11629348	0.33236754	-0.30142602
0.24236958	-0.17492138	0.79579556	-0.24926914	0.4617728	0.3672536
-0.30530575	0.07718785	-0.6303062	0.26881984	0.5614346	-0.37067947
-0.00600304	0.04352674	0.23551548	0.19042501	0.3371458	-0.311449
-0.42382553	0.73600256	-0.36386776	-0.18039742	0.5989648	-0.380900575
-0.20447552	0.06622235	-0.37475306	0.0280698	0.6039673	0.21896818
0.3661781	-0.10320444	-0.26140347	0.37961113	0.1944493	-0.01533087
-0.2297043	0.08951918	0.10416844	-0.3506819	0.36151022	0.35748914
-0.17569111	-0.4904412	0.5420365	0.41810396	-0.30843937	0.7592813
0.03478383	-0.20763385	0.02387038	-0.27994165	0.55906534	0.5368753
0.16584067	0.27482998	0.99505603	-0.40929642	-0.25490984	0.45280957
0.22910671	0.22870302	-0.19604152	0.01533751	0.25653428	0.17929718
0.6828513	-0.5203298	0.06492059	-0.09145571	0.03890252	-0.23337269]

Рис. 1. Вектор слова «печь» в трехсотмерном пространстве

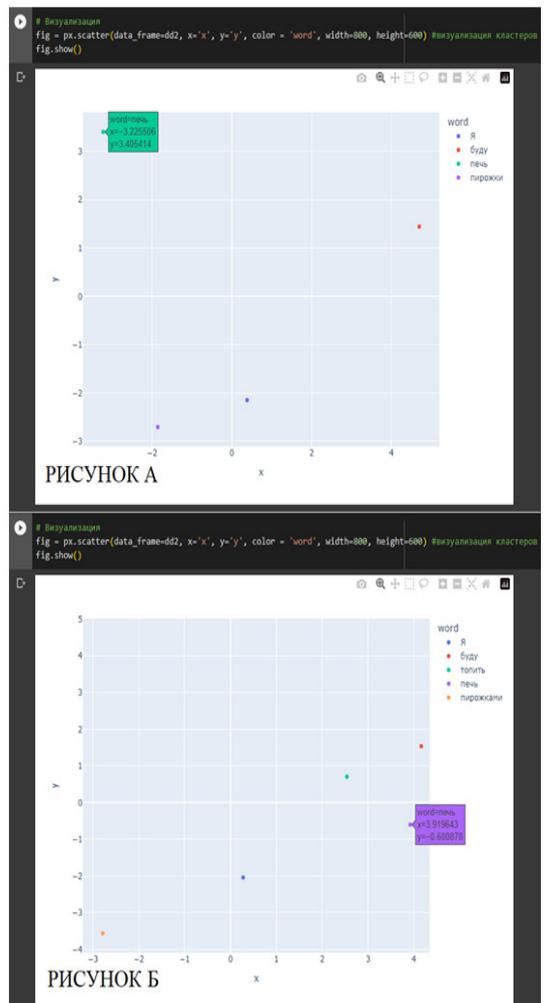


Рис. 2. Графическое представление эмбединга слова «печь»

### **Результаты исследования**

Мы провели пилотажное исследование с целью проверки возможностей эмбединга для применения в пенитенциарной психологии. Основная гипотеза была представлена в виде тезиса о целесообразности использования эмбединга в качестве метода в психологических исследованиях, который должен отвечать требованиям методов исследования психических явлений, то есть быть воспроизводимым, алгоритмичным и адекватным предмету науки психологии, а также раскрывать особенности исследуемого психологического процесса.

В качестве образцов текста взяты личные сообщения на сайте <https://pikabu.ru>. Сайт посещают 534 тыс. чел. в день. Сайт предоставляет возможность в том числе описывать различные ситуации из жизни и делить их по разной тематике. Для анализа выбраны следующие разделы сайта: «Истории из жизни» (33 352 поста, 71 752 подписчика); «Офисные будни» (4611 постов, 15 745 подписчиков); «Лига разбитых сердец» (3116 постов, 12 590 подписчиков). 10 самых популярных постов из этих разделов переведены в текстовый файл и программно проанализированы. Далее результаты были перенесены на сетки координат, представленные на рисунках 3, 4 и 5.

Анализируя представленные рисунки можно заключить, что истории из жизни, которыми люди делились на сайте, по смыслу, то есть графическому представлению эмбединга, отличаются от историй, рассказанных в разделах «Офисные будни» и «Лига разбитых сердец». Можно также отметить, что переживания, которыми люди делились в категории «Офисные будни», схожи с переживаниями из категории «Лига разбитых сердец». Следовательно, работа в офисе, по мнению написавших истории людей, больше связана с межличностными половыми переживаниями, чем с повседневной жизнью. Кроме того, описание так называемых офисных будней скуднее по использованию слов, чем истории в двух других группах.

### **Выводы**

По результатам проведенного пилотажного исследования можно заключить следующее: 1) графическое представление эмбединга в качестве метода психодиагностического обследования вполне возможно; 2) оно позволяет наглядно изобразить смысл используемых текстовых сообщений в различных исследуемых группах.

Где находятся точки соприкосновения использования данного метода в юридической психологии?

1) ситуации, связанные с нежеланием испытуемых участвовать в психодиагностическом обследовании (осужденные исправительных учреждений отрицательной направленности). Можно связать графическое представление эмбединга с психологическими характеристиками, на этой основе использовать его в качестве вспомогательного метода, тем более что примеры похожих психологических исследований уже существуют (А. К. Ковалев, Ю. М. Кузнецова, М. Ю. Пенкина). «С помощью разработанного в ФИЦ ИУ РАН инструмента автоматического анализа текста и методов машинного обучения были получены первые результаты в задаче выявления текстовых параметров, специфичных для людей с определенными психологическими особенностями. Инструмент корпусных лингвостатистических исследований, опирающийся на использование реляционно-ситуационного анализа, психолингвистических показателей и словарей, охватывающих лексику эмоциональной и рациональной оценки, позволили получить значения для 177 текстовых признаков эссе, написанных 486 испытуемыми» [2];

2) ситуации, связанные с изучением социально-психологического климата различных групп. По текстовым сообщениям членов групп, переведенным в графическое пред-

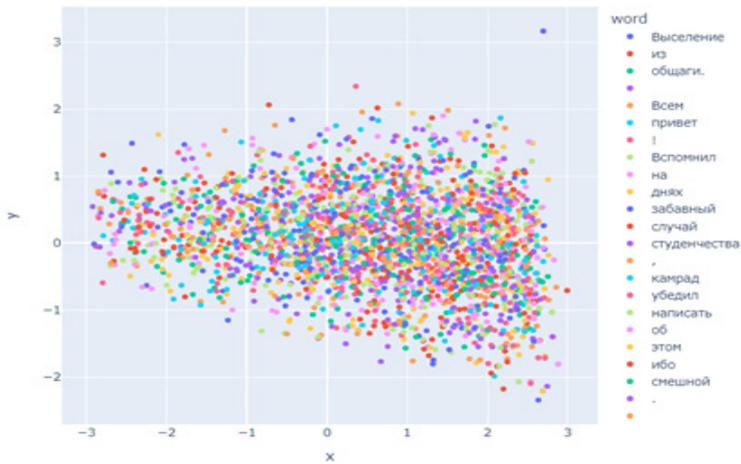


Рис. 3. Эмбеддинг постов из группы «Истории из жизни»

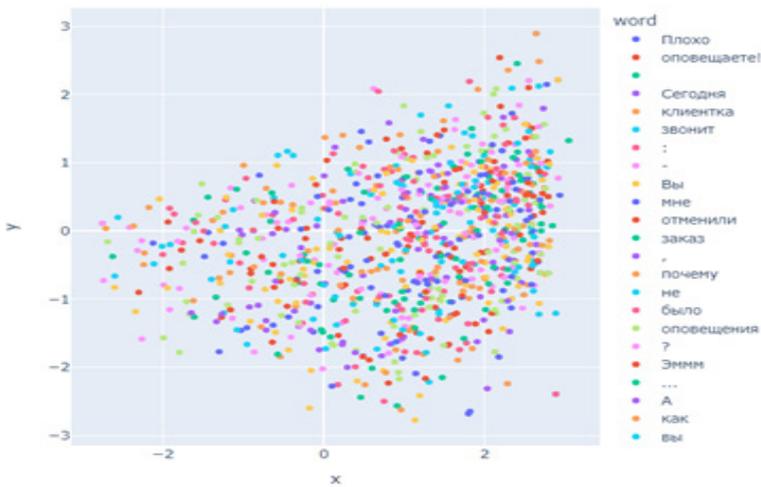


Рис. 4. Эмбеддинг постов из группы «Офисные будни»



Рис. 5. Эмбеддинг постов из группы «Лига разбитых сердец»

ставления смыслов, можно судить об основных тенденциях и интересах данной группы, что частично продемонстрировано в пилотажном исследовании, описанном в статье.

#### Список источников

1. Ениколопов С. Н., Кузнецова Ю. М., Осипов Г. С. Метод реляционно-ситуационного анализа текста в психологических исследованиях // Психология. Журнал Высшей школы экономики. 2021. Т. 18, № 4. С. 748–769.
2. Ковалев А. К., Кузнецова Ю. М., Пенкина М. Ю. Возможности автоматического анализа текста в задаче определения психологических особенностей автора // Экспериментальная психология. 2020. Т. 13, № 1. С. 149–158.
3. Сорокоумова С. Н., Курдин Д. А. Использование data mining в изучении динамики личностного роста курсантов ведомственного вуза Федеральной службы исполнения наказаний // Человек: преступление и наказание. 2022. Т. 30, № 1. С. 86–95.
4. Эльзессер А. С., Капустина Т. В. Проблема диагностических возможностей психологических методик: современное состояние // Психолог. 2019. № 3. DOI: 10.25136/2409-8701.2019.3.30013.

#### References

1. Enikolopov, S. N., Kuznecova, Ju. M. & Osipov G. S. 2021, 'The method of relational situational analysis of the text in psychological research', *Psychology, Journal of the Higher School of Economics*, vol. 18, iss. 4, pp. 748–769.
2. Kovalev, A. K., Kuznecova, Ju. M. & Penkina, M. Ju. 2020, 'The possibilities of automatic text analysis in the task of determining the psychological characteristics of the author', *Experimental psychology*, vol. 13, iss. 1, pp. 149–158.
3. Sorokoumova, S. N. & Kurdin, D. A. 2022, 'The use of data mining in the study of the dynamics of personal growth of cadets of the departmental University of the Federal Penitentiary Service', *Man: crime and Punishment*, vol. 30, iss. 1, pp. 86–95.
4. Jel'zesser, A. S. & Kapustina, T. V. 2019, 'The problem of diagnostic capabilities of psychological techniques: the current state', *Psychologist*, iss. 3. DOI: 10.25136/2409-8701.2019.3.30013.

#### Информация об авторах

**М. И. Марьин** – доктор психологических наук, профессор, профессор кафедры научных основ экстремальной психологии;

**Д. А. Курдин** – кандидат психологических наук, старший преподаватель кафедры общей и педагогической психологии.

#### Information about the authors

**M. I. Mar'in** – Doctor of Psychological Sciences, Professor, Professor of the Department of Scientific Foundations of Extreme Psychology;

**D. A. Kurdin** – Candidate of Psychological Sciences, Senior Lecturer at the Department of General and Pedagogical Psychology.

#### Примечание

Содержание статьи соответствует научной специальности 5.3.9. Юридическая психология и психология безопасности (психологические науки).

Статья поступила в редакцию 23.08.2023; одобрена после рецензирования 17.10.2023; принята к публикации 01.12.2023.

The article was submitted 23.08.2023; approved after reviewing 17.10.2023; accepted for publication 01.12.2023.