Научная статья

УДК 81’32+519.767.2
DOI 10.25205/1818-7935-2023-21-1-67-82

Высокоуровневая семантическая интерпретация структуры статических моделей для русского языка

Олег Алексеевич Сериков[1], Вероника Александровна Ганеева[2]
Анна Александровна Аксенова[3], Эдуард Станиславович Клышинский[4]

[1]Московский физико-технический институт
Москва, Россия

[1]Институт искусственного интеллекта AIRI
Москва, Россия

[1]Институт языкознания РАН
Москва, Россия

[1,2,4]Научно-исследовательский университет «Высшая школа экономики»
Москва, Россия

[3]ПАО «Сбербанк»
Москва, Россия

[1]srkvoa@gmail.com, https://orcid.org/0000-0002-3746-2642
[2]vaganeeva@edu.hse.ru, https:// 0000-0002-4020-488X
[3]aaaksenova2@gmail.com
[4]eklyshinsky@hse.ru, https:// 0000-0002-9569-9197

Аннотация

С момента своего появления векторное пространство Word2vec стало универсальным инструментом для научной и практической деятельности. С течением времени стало понятно, что необходима разработка новых методов интерпретации расположения слов в векторном пространстве. Существующие методы включали рассмотрение узкого круга аналогий либо кластеризацию пространства. В последние годы активно развивается подход на основе пробинга – анализа влияния небольших изменений в модели на результат. В этой работе мы предлагаем метод интерпретации расположения слов в векторном пространстве, применимый ко всему пространству в целом. Метод позволяет выявлять основные направления, вдоль которых выделяются наиболее крупные группы слов (около трети всех слов словаря), противопоставляемые друг другу по некоторым семантическим признакам, а также строить неглубокую иерархию таких признаков. Эксперименты были проведены на трех моделях, обученных на разных корпусах: Национальном корпусе русского языка, Araneum Russicum и коллекции научных статей из разных предметных областей. Для экспериментов использовались только имена существительные, входящие в словарь моделей. Рассмотрена экспертная интерпретация подобного разделения вплоть до третьего уровня. Набор и иерархия этих признаков отличаются для разных моделей, однако все они являются достаточно общими. Было обнаружено, что выделенные признаки разделения зависят от состава корпусов, на которых проводилось обучение моделей, их направленности и стиля. Полученное разделение не всегда коррелирует с принятым в области разработки онтологий. Так, совпадающим признаком является абстрактность или вещность объекта. Однако для моделей на верхнем уровне оказывается более важным разделение на повседневную/специальную лексику, архаичную лексику, разделение на имена собственные и нарицательные. В статье приведены примеры слов, входящих в полученные группы.

Ключевые слова

векторные модели, интерпретация модели, Word2vec, сингулярное разложение, построение онтологий

Благодарности

Авторы глубоко признательны Екатерине Владимировне Рахилиной за вдохновение, которое она дарила нам по мере написания этой статьи.

Для цитирования

Сериков О. А., *Ганеева В. А., Аксенова А. А., Клышинский Э. С.* Высокоуровневая семантическая интерпретация структуры статических моделей для русского языка // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2023. Т. 21, № 1. С. 67–82. DOI 10.25205/1818-7935-2023-21-1-67-82

High-Level Semantic Interpretation of the Russian Static Models Structure

Oleg A. Serikov[1], Veronika A. Geneeva[2]
Anna A. Aksenova[3], Eduard S. Klyshinskiy[4]

[1]Moscow Institute of Physics and Technology
Moscow, Russia

[1]Artificial Intelligence Research Institute
Moscow, Russia

[1]Institute of Linguistics RAS
Moscow, Russia

[1,2,4]HSE University
Moscow, Russia

[3]JSC Sberbank
Moscow, Russia

[1]srkvoa@gmail.com, https://orcid.org/0000-0002-3746-2642
[2]vaganeeva@edu.hse.ru, https:// 0000-0002-4020-488X
[3]aaaksenova2@gmail.com
[4]eklyshinsky@hse.ru, https:// 0000-0002-9569-9197

Abstract

Since its inception, the Word2vec vector space has become a universal tool both for scientific and practical activities. Over time, it became clear that there is a lack of new methods for interpreting the location of words in vector spaces. The existing methods included consideration of analogies or clustering of a vector space. In recent years, an approach based on probing—analysis of the impact of small changes in the model on the result—has been actively developed. In this paper, we propose a new method for interpreting the arrangement of words in a vector space, applicable for the high-level interpretation of the entire space as a whole. The method provides for identifying the main directions which are selecting large groups of words (about a third of all the words in the model's dictionary) and opposing them by some semantic features. The method allows us to build a shallow hierarchy of such features. We conducted our experiments on three models trained in different corpora: Russian National Corpus, Araneum Russicum and a collection of scientific articles from different subject domains. For our experiments, we used only nouns from the models' dictionaries. The article considers an expert interpretation of such division up to the third level. The set of selected features and their hierarchy differ from model to model, but they have a lot in common. We have found that the identified semantic features depend on the texts comprising a corpus used for the model training, their subject domain, and style. The resulting division of words does not always correlate with the common sense used for ontology development. For example, one of the coinciding features is the abstract or material nature of the object. However, at the upper level of models, words are divided into everyday/special lexis, archaic lexis, proper names and common nouns. The article provides examples of words included in the derived groups.

Keywords

мector models, interpretation of models, Word2vec, singular vector decomposition, ontology development

Acknowledgements

The authors are extremely grateful to Ekaterina V. Rakhilina for her support and inspiration all the way along the process.

For citation

Serikov O. A., Ganeeva V. A., Aksenova A. A., Klyshinskiy E. S. High-Level Semantic Interpretation of the Russian Static Models Structure. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2023, vol. 21, no. 1, pp. 67–82. (in Russ.) DOI 10.25205/1818-7935-2023-21-1-67-82

Введение

Векторные модели представляют собой удобный инструмент для решения практических задач – таких, для которых раньше не существовало решений на приемлемом уровне. При этом гораздо меньше внимания привлекает применение онтологий и тезаурусов, фреймовых моделей, а также некоторых других методологий. Подобное смещение связано с тем, что применение старых подходов требует бóльших затрат. Создание полной онтологии одной предметной области может занять многие годы работы большого коллектива[1], тогда как новые модели требуют относительно небольших затрат процессорного времени.

Первой успешной реализацией идей дистрибутивной семантики [Gallant, 1991] стала работа по Word2Vec [Mikolov, 2013]. В ней было также показано, что полученное векторное пространство обладает возможностью интерпретации, и была поставлена задача нахождения аналогий между понятиями, при этом таких, что пара, задающая аналогию, позволяет найти новые пары, отвечающие этой аналогии (см., например, [Ethayarajh, 2019; Korogodina, 2021]). Вместе с тем векторные представления плохо интерпретируются с точки зрения прагматики. Кроме того, в многомерном пространстве рядом оказываются семантически близкие слова, но слова из одной предметной области могут оказаться далеко друг от друга.

Подобные сложности с интерпретацией результатов приводят к возникновению задачи выделения осей, имеющих очевидную смысловую нагрузку. Для ее решения предложено два пути: разработка интерпретируемых моделей (см. [Faruqui, 2015; Subramanian, 2018]) либо интерпретация существующих. Поиску интерпретируемых моделей посвящена, например, работа [Kozlowski, 2017], где было предложено найти показатели для мужского/женского пола, статуса в обществе, достатка и некоторых других. Данные измерения выделялись на основе слов одной тематики, взятых из тематических словарей. В [Rubinstein, 2015] показано, что статические векторные модели лучше улавливают таксономические характеристики слов, чем другие их семантические атрибуты.

У векторных моделей, построенных по данным разной природы, могут обнаруживаться похожие признаки. В [Yao, 2019] показано сходство векторных представлений, полученных из текстовых описаний организаций и данных о биржевой торговле их акциями. В работе [Kutuzov, 2020] показано, что модели, построенные по разным корпусам, позволяют выявить регулярности в соотношениях между одними и теми же концептами, являющиеся информативными для сравнения корпусов, представляющих один и тот же домен в разные эпохи.

Аналогичный перенос работает и при сопоставлении векторных моделей, построенных для разных языков. В [Conneau, 2017] используется межъязыковая схожесть концептов, сопоставление моделей для разных языков проводится при помощи алгоритма прокрустова выравнивания. У [Rabinovich, 2020] строится мультиязычное семантическое поле для автоматизации сравнения средств лексического выражения концептов в разных языках. Анализ результатов позволяет авторам выделить филогенетические и другие экстралингвистические факторы, значимые для типологии полисемии.

Перечисленные методы хорошо подходят для анализа отдельных явлений, однако выделение всех измерений, существующих в векторном пространстве, оказывается труднее, поскольку оно состоит в локальной интерпретации слов. Более того, в векторном пространстве наблюдаются нелинейные искажения выделяемых измерений, препятствующие интерпретации [Kozlowski, 2017]. Авторы [Weeds, 2014] иллюстрируют тезис о трудности отделения синонимов и антонимов слов.

В данной работе мы ставим задачу исследовать интерпретационные свойства векторного пространства на верхнем уровне и выдвигаем гипотезу о том, что в текстовом массиве можно

[1] Так, онтология медицинской диагностики [Грибова, 2018] создается уже около пяти лет, в то время как самая крупная медицинская онтология Unified Medical Language System (UMLS) и ее тезаурус Metathesaurus создаются международным сообществом уже два десятилетия [Bodenreider, 2004].

выделить большие зоны (порядка четверти всего списка слов), которые объединены по какому-то лексико-семантическому признаку. Также мы проверяли возможность дальнейшего разделения этих зон на большие интерпретируемые фрагменты. Нахождение подобных интерпретируемых областей позволит понять природу трудностей в решении некоторых практических задач.

1. Контекстные модели и пробинг «черного ящика»

Пробинг – это направление, открывающее неявные, часто глубинные представления в моделях обработки текстов (дискурса, а не слов). Пробинговые исследования можно структурировать как реализующие три последовательных этапа анализа моделей [Lasri, 2022]: бихевиоральный, диагностический и инвазивный.

Бихевиоральный пробинг фокусируется на проверке того, имеет ли смысл подробное исследование модели с позиции лингвистических характеристик. Результаты работы модели подвергаются анализу с позиции произвольно выбранной лингвистической характеристики, например способности корректно дописывать текст.

Авторы [Linzen, 2016] оценили синтаксические возможности моделей, проанализировав их поведение на участках текстов, репрезентативных с точки зрения грамматической категории числа и соответствующих этой категории механизмов согласования. Авторы смогли отделить синтаксическую компетенцию модели от заложенных в нее простых эвристик (например, согласование прилагательного с ближайшим существительным). В работе [Loureiro, 2019] удалось построить алгоритм, относящий токены текста к тезаурусу WordNet, что позволяет установить наличие тезаурусоподобной подструктуры в векторных представлениях. Однако предложенный алгоритм указывал лишь на наличие/отсутствие признаков лингвистического знания в модели, а не на ответственные за него части модели.

Диагностический пробинг выявляет реакцию модели, рассматриваемой исследователем как «черный ящик». Пробинговые исследования, опирающиеся на модели линейной корреляции [Adi, 2016], устанавливают отношение между внутренними векторными представлениями и выбранной лингвистической характеристикой. Так, замеряется корреляция между представлениями слов в определенных слоях моделей и соотнесенностью этих слов с определенными частями речи. В работе [Conneau, 2018] автор исследует модели машинного перевода и понимание ими поверхностной семантики текста, его грамматической структуры и глубинной семантики по векторному представлению текста.

В работе [Voloshina, 2022] пробинг показал, что лингвистическая информация усваивается на ранних этапах обучения. При этом модели способны фиксировать различные свойства языка на уровнях морфологии, синтаксиса и дискурса, но могут не справляться с задачами, которые воспринимается как простые (итоговый уровень владения языком примерно соответствует уровню 11-летнего ребенка: следование общей теме диалога, дискурсивные особенности текста, богатство используемой лексики и др.).

Для поиска элементов модели, влияющих на точность решения лингвистической задачи, используют *инвазивный пробинг*. В [Ravfogel, 2021] области нейронной сети, выявленные диагностическими методами, искусственно зашумляются, чтобы можно было определить их влияние на анализируемое свойство. В [Vig, 2020] веса нейронов замораживаются для определения их влияния на «гендерную предвзятость» сети. В результате проводимого в работе [Chizhikova, 2022] *пробинга синтаксисом* делается вывод о том, что механизм внимания фокусируется не только на синтаксических отношениях, но и на семантике. Такой подход позволяет авторам указать на непоследовательность упорядочения объектов в энциклопедии Wikidata.

2. Задачи и гипотезы

На данный момент существует значительное количество работ, посвященных качественному анализу того, как векторным моделям удается отражать семантику предметной области. Методов определения семантической подструктуры в статических векторных моделях, способных упорядочить все пространство, пока не было представлено. Разработке такого метода и посвящена наша работа.

Исследуя векторное пространство, естественно задаться вопросом об устройстве базиса этого пространства (или базисов его подпространств) и о возможности выделения в нем интерпретируемого базиса. Если таковой существует, можно предположить существование измерений, каждому из которых сопоставлен семантический (или прагматический) признак (или набор признаков), внутри которого проводится противопоставление слов. Структура этих семантических и прагматических признаков может иметь вид дерева или онтологии. Таким образом, наша гипотеза состоит в том, что существуют интерпретируемые компоненты в статических векторных моделях и что возможно выделить графовую подструктуру этих компонент.

3. Использованные данные

Мы не знали заранее, как лингвистические свойства (семантические, прагматические или грамматические) распределятся по уровням ветвления выделенных компонент. Глаголы могли разделиться на верхнем уровне по категориям вида, возвратности и переходности, так как эти параметры также влияют на сочетаемость глагола с другими словами. В противоположность им у существительных категория рода выражает не столько различия между особями мужского и женского пола или одушевленными и неодушевленными предметами, сколько устоявшееся разделение слов по грамматическим признакам. Поэтому мы отказались от использования слов тех частей речи, грамматические параметры которых могут выражать семантику использования слова – глаголов, причастий и деепричастий. Количество служебных частей речи невелико, они чаще не несут собственной семантики, поэтому они также были отброшены.

В качестве кандидатов на проведение исследования были выбраны имена существительные (нарицательные и собственные), взятые из словарей использованных статических векторных моделей Word2Vec. Эти словари содержат неточности (ошибки лемматизации, аббревиатуры, опечатки и т. д.), которые мы отбросили при помощи метода, предложенного в [Bocharov, 2011], исключив также слова короче трех символов. Слово считалось существительным, если среди вариантов морфологического анализа имелось существительное в начальной форме.

Эксперименты проводились на трех моделях Word2vec. Первой была рассмотрена модель ruscorpora_upos_cbow_300_20_2019, взятая с сайта RusVectores[2]. Данная модель обучена на корпусе текстов НКРЯ объемом 270 млн слов и включает художественную литературу, коллекцию исторических текстов (начиная с XVII века), учебники, устную речь и др.[3] Второй моделью была araneum_upos_skipgram_300_2_2018, также взятая с сайта RusVectores. Она была обучена на корпусе Araneum Russicum объемом около 10 млрд слов, состоящем в основном из текстов, полученных из сети Интернет (новости, форумы, объявления), учебников, художественной литературы. Третью модель мы обучили самостоятельно на основе коллекции научных статей по нескольким отраслям науки: архитектура, искусствоведение, автоматизация, геология, история, культурология, лингвистика, филология. Общий объем данного корпуса ок. 550 млн слов.

[2] https://rusvectores.org/ru/models/.

[3] Более подробно о составе корпуса: https://ruscorpora.ru/. Однако состав корпуса значительно поменялся с момента обучения модели.

4. Метод поиска интерпретируемых компонент

Метод поиска интерпретируемых компонент заключается в следующем. Для списка слов проводится сингулярное разложение, выделяющее измерения с наибольшим разбросом в данных. Слова сортируются по координате вдоль выбранной оси и разделяются на три подгруппы (в простейшем случае – на три равные части). Мы предполагаем, что слова, попавшие в среднюю группу, должны быть семантически нейтральны по анализируемому признаку, поэтому исследуются только слова с периферии оси. Каждая выделенная пара групп слов анализировалась экспертом-лингвистом на предмет выявления признака, по которому они могли бы быть противопоставлены.

Для каждой из двух крайних выделенных групп слов мы рекурсивно повторяли SVD-разложение и разделение списка на три части. На практике выяснилось, что внутри подгрупп, выделенных на предыдущих шагах, первые измерения, возвращаемые SVD, могут совпадать. В связи с этим на шаге с номером n мы пропускаем первые выделенные оси и берем для анализа ось с номером *n*.

5. Результаты экспериментов

Общая работоспособность предложенного метода была проверена на специально составленном списке растений, включающем в себя травы, кустарники и деревья (с выделением плодоносящих и окультуренных). Эксперименты проводились на модели, обученной на текстах Национального корпуса русского языка (НКРЯ). Во время первой итерации растения, являющиеся привычными для российской культуры, отделились от нехарактерных: например, растения средней полосы России vs пальмы и кактусы. Одним из объяснений здесь может быть частотность употребления подобных слов в исходном корпусе текстов, однако у нас не было возможности проверить эту гипотезу. На втором уровне привычные растения (или их плоды) разделились по признаку употребления в пищу. Необычные растения разделились на те, что ассоциируются с пропагандой здорового образа жизни (*киноа*, *авокадо* и др.), и все остальные, чаще не употребляемые в пищу.[4]

Убедившись, что предложенный метод показывает интерпретируемые результаты, мы продолжили наши эксперименты уже на существительных, представленных в словаре выбранных моделей. Первой также рассматривалась модель, обученная на текстах НКРЯ.

На первом этапе предложенный метод выделил ось, к краям которой тяготели слова, которые могут быть проинтерпретированы как противопоставление конкретных и абстрактных существительных по Розенталю [1976]. Признаками абстрактных существительных в этом словаре являются суффиксы (например, *-ость-*, *-енщ-*) и отсутствие множественного числа у таких существительных, или иначе – множественное число, несочетаемое с количественными числительными: **пять правд*, **три абстрактности*. К краю оси физического мира (к конкретным) тяготеют существительные, обладающие грамматически выражаемой неотчуждаемостью.

Полученную ось следует рассматривать скорее как шкалу, спектр или континуум для выделенных признаков, где максимальными значениями этих признаков обладают слова, тяготеющие к краям оси. Однако разделение слов не является идеальным, а предложенная нами трактовка полученных результатов не всегда полностью описывает ситуацию.

Чтобы прояснить сказанное, приведем здесь списки из 20 слов, отнесенных нашим методом к концам оси, то есть получивших максимальные и минимальные значения координаты с приписанными нами признаками конкретности и абстрактности. Здесь и далее мы не стали

[4] Заметим, что НКРЯ содержит относительно небольшое количество современных текстов, посвященных здоровому образу жизни в его новом понимании, и потому полученное разделение было для нас сюрпризом, так как оно не полностью соответствует нашим представлениям об онтологии предметной области.

удалять слова из списков, чтобы продемонстрировать точность работы метода. Интерпретация осей проводилась по полным спискам слов, которые не приводятся в связи с ограниченным объемом работы.

- Конкретные: *ладонь, стол, шея, изба, шапка, рубаха, спина, комната, палец, щека, колено, грудь, губа, голова, платок, глаз, рука, крыльцо, дверь, нога.*
- Абстрактные: *развитие, деятельность, отношение, условие, система, задача, организация, действие, процесс, правительство, значение, изменение, государство, исследование, вопрос, решение, влияние, период, возможность, закон.*

На втором этапе абстрактные слова разделились по признакам «индустриальное» (прикладные науки и производство) vs «духовное» (внутренний мир человека, чувства, качества человека). Конкретные же термины разделились по признакам «архаичность» и «современность».

- Духовное, отнесенное ранее к абстрактному: *вера, истина, убеждение, чувство, любовь, добродетель, мысль, христианство, народ, жизнь, стремление, невежество, идеал, зло, поступок, страдание, воззрение, разум, религия, отрицание.*
- Индустриальное, отнесенное ранее к абстрактному: *топливо, транспорт, аэропорт, офис, скважина, стоимость, заявка, комбайн, аппаратура, температура, мощность, доставка, маршрут, тонна, доллар, зона, автомобиль, вертолет, компания, база.*
- Архаичное, отнесенное ранее к конкретному: к*озак, боярин, воевода, верста, стрелец, слобода, казак, купец, лях, государев, острог, барин, изба, поселянин, пан, деревня, разбойник, ямщик, лошадь, конь.*
- Современное, отнесенное ранее к конкретному: *холл, блокнот, сантиметр, колготки, комбинезон, футболка, майка, сумочка, шприц, галстук, витрина, очки, флакон, тумбочка, лампочка, холодильник, свитер, плитка, блузка, пепельница.*

Как видно из результатов, на втором этапе проявляются сложности, связанные с разделением на группы на первом этапе. Так, слова *деревня*, *село*, *милость*, *божий* отнесены к конкретному архаичному, хотя слова *село* и *деревня* могут обозначать как конкретные поселения, так и их собирательный образ. По сведениям НКРЯ, частота употребления слов *село* и *деревня* в архаичном контексте (подкорпусы 1682–1960 гг.) выше, чем в современном, где они уступают место *населенному пункту*. Аналогичная картина наблюдается в корпусе Google N-gramms, где популярность слов село и деревня пошла на спад после 2000 года. Аналогично *комбайн*, *автомобиль* и *вертолет* были отнесены к абстрактным понятиям. По всей видимости, здесь проявляется эффект разделения слов вдоль оси на три равные части. В итоге слова, расположившиеся ближе к середине списка, поднимаются к краям оси при следующем разделении.

Наконец, на третьем этапе разделение произошло следующим образом. Абстрактные слова из духовной сферы, разделились на «духовное» (на краю оси оказались следующие: *добродетель*, *токмо*[5], *чувствование*, *наставление*, *прилежание*, *житие*, *оный*, *слог*, *похвала*, *любовь*) и «общественно-социальное» (*лозунг*, *сторонник*, *конфликт*, *пропаганда*, *капитализм*, *фашизм*, *война*, *диктатура*, *кризис*, *социализм*). Абстрактные слова, относящиеся к индустриальной сфере, разделились на «сфера (индустрия) развлечений» (*фестиваль*, *шоу*, *чемпионат*, *фильм*, *чемпион*, *компьютер*, *экран*, *дизайнер*, *турнир*, *гонщик*) и «сфера управления предприятиями и хозяйственной деятельности» (*хозяйство*, *доход*, *ведение*, *распоряжение*, *увеличение*, *принятие*, *предотвращение*, *перевозка*, *подвоз*, *усиление*). Конкретные архаичные слова разделились на «книжные» (*сердце*, *взор*, *чертог*, *душа*, *меч*, *темница*, *око*, *уста*, *богиня*, *мрак*) и «повседневные» (*починка*, *трактир*, *огород*, *картуз*, *девчата*, *писарь*, *лавка*, *сельсовет*,

[5] Данное слово присутствует в словаре [Bocharov, 2011] как существительное, поэтому мы не стали убирать его из выдачи, стремясь показать не только достоинства, но и недостатки метода. Аналогично мы поступили с примерами ниже, которые метод неаккуратно выделил на предыдущих этапах.

фельдшер, *артельщик*). Конкретные современные слова разделились так же на «книжные»[6] (*волос*, *румянец*, *кудри*, *лицо*, *локон*, *платье*, *глаз*, *личико*, *бровь*, *бородка*) и «повседневные» (*фургон*, *электричка*, *метро*, *ангар*, *самосвал*, *камера*, *площадка*, *мусор*, *вездеход*, *кран*). Более подробные списки слов по разделам приведены в приложении к статье, вынесенном на внешний ресурс.[7]

Здесь мы сталкиваемся со сходным разделением двух групп на разных ветках разделения: как архаичные, так и современные слова разделяются на более возвышенную и редкоупотребимую в быту лексику (книжную) и на входящую в повседневную жизнь. Сходное разделение может быть связано с тем, что метод пропустил более высокоуровневое разделение, присущее двум ветвям, из-за деления слов на группы. Заметим также, что среди книжных слов присутствуют *сердце*, отнесенное к абстрактным, и *волос*, *лицо*, *глаз*, *бровь*, отнесенные к конкретным. Очевидно, что все эти слова имеют отношение к органам человека, но слово *сердце* употребляется в НКРЯ чаще в значении души, чем органа.

Результаты разбиения слов по тематикам для модели, построенной на основе НКРЯ, сведены в таблицу 1.

Таблица 1

Схема разделения модели на основе НКРЯ

Table 1

Hierarchy of the Model Trained over Russian National Corpus (ruscorpora_upos_cbow_300_20_2019)

Абстрактное, виртуальные референты				Конкретное, референты в физ. мире			
Духовное, внутренний мир человека, литература		Индустрия, прикладные науки		Архаичное		Современное	
Духовное	Общественное	Развлечения	Наука	Книжное	Повседневное	Книжное	Повседневное

Второй изучалась модель, обученная на корпусе Araneum Russicum. На первом и втором этапах для нее были выделены следующие группы:

- люди (социальное регулирование): *пашка, малец, сашка, гришка, шурка, санька, женька, гриша, валерка, ванька, колька, старуха, эдик, борька, васька, прохвост, машка, митя, володька, поганец;*
  – абстрактное: *иррационализм, догматизм, обскурантизм, отрицание, релятивизм, индивидуализм, рационализм, аморализм, мистицизм, интеллектуализм, безнравственность, агностицизм, спиритуализм, морализм, отступничество, пантеизм, односторонность, критицизм, этноцентризм, материализм;*
  – конкретное: *аленушка, санька, дубок, анжела, андрюшка, девчата, рябинка, регин, данил, нюся, есения, мишина, олешек, сережа, фунтик, светланка, артем, синичка, пчелка, чижик;*
- организации (технологии и правовое регулирование): *котирование, поставка, пакет, планир, неиспользование, длительность, долгосрочность, авизование, обновление, минимизация, учет, периодичность, объем, формирование, видеотерминал, обеспечь, фотопродукция, закупка, телерадиопрограмма, бизнес-сектор;*
  – технология производства: *смачивание, перемешивание, нагрев, вспенивание, натяжение, изгибание, прилипание, разбрызгивание, встряхивание, прижатие, смешива-*

[6] По сведениям НКРЯ, слово волос встречается 15 684 раза в 4749 нехудожественных текстах и 39 293 раза в 5652 художественных текстах. Подобное соотношение будет являться критерием для разделения лексики на «книжную» (чаще встречающуюся в художественной литературе), «повседневную» и «специальную» (чаще встречающуюся в других источниках). Заметим, что полученный список «книжных» слов соответствует предыдущим исследованиям, показавшим, что для женщин в книгах чаще описывается внешность, а для мужчин — характер.

[7] https://github.com/klyshinsky/interpretation_paper_2023.

ние, обжатие, склеивание, шлифование, термоэлемент, стекание, выглаживание, эмульгирование, засаливание, вдавливание;

– управление производством: *саратовэнерго, башавтотранс, бишкек, темиртау, ханты, караганда, октябрь, октябрьск, ставрополье, белогорск, сибакадемстрой, зеленодольск, краснодар, директор, армавир, жигулевск, пенза, гидрострой, братск, астана.*

Здесь мы видим, что со сменой корпуса для обучения модели меняется выделяемая лексика и ее группировка. НКРЯ содержит в себе больший объем литературы, написанной в прошлые века. Поэтому на верхнем уровне мы видим разделение на абстрактное и конкретное, а также на современное и архаичное. Корпус Araneum Russicum содержит в основном современные тексты, в том числе из сети Интернет. Как следствие, в нем можно увидеть, как человек и социум противопоставляется организации и ее функционированию. Интересно, что в гуманитарно-социальной сфере уже на втором уровне наблюдается разделение на абстрактное и конкретное, встретившееся и в модели НКРЯ. Следовательно, некоторые измерения обладают достаточно высоким уровнем абстракции для того, чтобы встречаться в разных моделях. Здесь мы также видим на краю оси кластер с именами собственными, причем имена людей отделяются от названий организаций и мест.

Разделение слов на трех уровнях показано в таблице 2, список слов приведен во внешнем приложении.

Таблица 2

Схема разделения модели на основе Araneum Russicum

Table 2

Hierarchy of the Model Trained over Araneum Russicum Corpus (araneum_upos_skipgram_300_2_2018)

Люди, социальное регулирование				Организация, технология, правовые отношения			
Абстрактное		Конкретное		Технологии производства		Управление и производство	
Оскорбления	Духовный мир	Диминутивы, имена	Природа	Химия и биология	Физика, автоматизация	Экономика	Политика и общество

Третьей изучалась модель, обученная на научных текстах. На первом этапе здесь выделились слова, имеющие повседневное и специальное употребление. Так, *стих* находится в группе общепринятого, а *дольник* или *акростих* – в группе более научного употребления, то есть сложность первых будет меньше, чем вторых. При этом в научные термины попали фамилии ученых и авторов статей, названия институтов, университетов и других учреждений, города их расположения.

Слова, которые метод распределил по группам после разделения на первом и на втором этапах, показаны ниже. Напомним, что здесь приводятся только слова, оказавшиеся на краю выделенной оси.

- Повседневное: *мочь, образ, место, вода, значение, лицо, день, сцена, ребёнок, действие, женщина, количество, комната, девушка, желание, линия, площадь, смерть, группа, эпизод.*

  – Абстрактное: *отношение, власть, интерес, идея, деятельность, позиция, поведение, убеждение, жизнь, признание, мысль, закон, характер, идеал, государство, смысл, норма, отказ, доверие, мочь.*

  – Конкретное: *бляшка, улитка, галька, лунка, рулон, коврик, палочка, кусочек, солонка, поднос, серьга, ящик, подол, войлок, доска, палатка, веточка, кувшин, трубка, яма.*

- Научное: *музыкознание, прагматик, дейк, рецепция, диахронии, реферирование, социолингвистика, кемерово, симпозиум, новосибирск, востоковедение, информ, лексикология, доцент, семиотика, методы, религиоведение, мгу, типология, семинара.*
  – Терминология: *подход, формирование, дискурс, функционирование, мышление, коммуникация, синтез, взаимосвязь, интеграция, актуальность, лингвистика, адаптация, становление, синкретизм, идентификация, усложнение, адекватность, идентичность, разработка, развёртывание.*
  – Имена собственные: *вилайет, петропавловск, глазго, маадыр, осло, георги, танзания, летие, онон, огайо, абакан, анадырь, калевала, зао, фио, сургут, тоо, йов, тегеран, элиста.*

Здесь снова проявляется разделение на абстрактное и конкретное для повседневной лексики и разделение на научное (книжное) и повседневное, которое появляется в модели НКРЯ на третьем уровне для предметных слов. Разделение слов на трех уровнях показано в таблице 3, список слов приведен во внешнем приложении.

Таблица 3

Схема разделения модели на основе научных статей

Table 3

Hierarchy of the Model Trained on Scientific Articles

Более повседневный дискурс				Более научный дискурс			
абстрактное		предметное		Научные термины		места и люди	
Общество и политика	Литература и духовный мир	Объекты – специальные термины	Бытовые предметы	Внешнее научное	Внутреннее научное	Фамилии и имена ученых	Места и организации

Показательным является разделение слов из некоторых тематических списков. Так, из списка слов, относящихся к информатике, используемый метод отнес к общеупотребительным слова *аргумент, почта, ядро, модуль, компьютер, буфер, сценарий, контейнер, субъект, протокол*, а к специальным – *кэширование, транслятор, репликация, октет, инкапсуляция, кодирование, трекер, эмуляция, битрейт, профайл*. Слова, относящиеся к философии, разделились следующим образом: общеупотребительные – *время, образ, случай, форма, действие, начало, момент, ситуация, душа, тело*; научные – *рефлексия, феномен, дихотомия, триада, трансцендентность, акциденция, относительность, дискурс, философия, онтология*. Заметим, что хотя общеупотребительные термины и являются более частотными, некоторые (но далеко не все) слова, оказавшиеся на разных концах осей, обладают сходной частотой употребления в НКРЯ. Для общеупотребительных и научных слов в подкорпусах художественных и нехудожественных текстов НКРЯ также не наблюдается зависимости от положения слов на осях и их частотности.

6. Интерпретация и обсуждение полученных результатов

На трех исследованных моделях видно, что разделение слов на самом верхнем уровне зависит от использованных текстовых корпусов, то есть разные векторные модели не создают единого разделения пространства на сходные области. Если НКРЯ содержит в себе большое (относительно других корпусов) количество старых текстов, то на верхнем уровне разделения входящих в его модель слов это оказывается важным параметром разделения (хотя и только на втором уровне) лексики на более старую и более современную. Модель на основе корпуса Araneum, содержащая в себе большое количество обсуждений тем на интернет-форумах, выделяет отношения внутри социума и производственные отношения. Наконец, для модели,

обученной на научных текстах, оказывается важным то, как употребляется слово: несет ли оно более общее значение или имеет узкое значение и употребляется в специальных текстах.

Несмотря на различия, в моделях можно выделить некоторые универсалии. Так, во всех трех моделях наблюдается следующее разделение: «абстрактное vs конкретное», «материальное vs идеальное», «общепринятое vs специальное», «имя собственное vs нарицательное». Исследование разных моделей может позволить выделить список таких универсалий и использовать их для анализа новых моделей. Заметим, что часть из них будет отвечать не столько за семантику, сколько за употребление слов в дискурсе. Таким образом, архаичность термина отражает скорее контекст употребления, чем прагматику. Однако на нижних уровнях иерархии проявляются свойства прагматики: пол, наличие плодов у растения, метод перемещения животного или механизма и проч.

Интересно, что логика разделения слов в векторном пространстве не похожа на логику построения онтологий и тезаурусов. К примеру, в модели, обученной на научных статьях, имена собственные расположились среди специальных терминов. Специальные термины часто выносятся в названия статей, а сами эти названия помещаются в списки литературы вместе с именами авторов и названиями организаций. То есть употребление названия организации или имени ученого носит примерно тот же характер, что и употребление термина. Вообще, судя по всему, модель учитывает не только предметную область, но и стиль текста, период его создания (то есть актуальная на тот момент лексика), частотность использования слова и некоторые другие параметры.

Заметим, что сходство слов подразумевает близость всех или почти всех параметров, а различие – несовпадение даже одного параметра. В итоге расположение близких по семантике слов в одной области пространства является очевидным, а вот измерения, в которых будут располагаться различающиеся слова, не задаются каким-то тривиальным образом. Логично высказать следующую гипотезу: если подобрать три группы слов так, чтобы слова внутри групп были сходны, а между группами наблюдалось отличие только по одному или двум параметрам (например, плодоносящие деревья, неплодоносящие деревья, плодоносящие кустарники), то и сами группы должны располагаться в векторном пространстве близко друг к другу. Однако на их взаимное расположение будет оказывать влияние расположение как семантически сходных, так и семантически далеких групп. При этом высокая размерность пространства дает большой простор для перемещений. Все это делает нетривиальной задачу определения связей между группами сходных слов, тогда как выделение самих сходных слов – задача по-прежнему относительно простая.

В целом, можно сказать, что векторные модели демонстрируют логику, отличающуюся от той, которой придерживаются разработчики тезаурусов и онтологий. Так, для онтологии WordNet также важно разделение на абстрактное и вещное, но для вещного на следующем уровне идет разделение на живое и неживое. В нашем случае мы видим совсем другие параметры: архаичность термина, его научность, социальные-производственные отношения и т. д. Как следствие, автоматически созданная онтология может быть отвергнута экспертом как не отражающая его представления о предметной области. Здесь мы также видим, что в зависимости от тематики текстов меняется представление слова и его окружение в векторном пространстве.

С другой стороны, наши эксперименты показывают, что создание единого интерпретируемого пространства – сложная задача, так как слова, относящиеся к разным подразделам онтологии, имеют разные наборы признаков. Иллюстрацией может послужить противопоставление материального в конкретных существительных и нематериального в абстрактных: материальное может быть противопоставлено по признакам размера, цвета, формы, одушевленности, а нематериальное этими признаками обладает далеко не всегда.

Вообще иерархия выделенных направлений обладает рядом интересных свойств и нуждается в тщательном изучении. Сами выделенные оси не могут составить метрического про-

странства, так как они не ортогональны. Сложно ввести какую-то единицу, относительно которой измерялось бы наличие или отсутствие того или иного свойства, что уже было замечено в работе [Kozlowski, 2017]. Даже само расположение осей необычно. Как уже говорилось выше, абстрактные непредметные сущности не обладают теми же свойствами, что конкретные предметные сущности. Это означает, что вместо привычного нам многомерного пространства, где каждая из точек имеет значение для каждой из координат, мы получаем пространство в форме дерева или графа, где переход в ту или иную сторону по одной из осей определяет наличие или отсутствие части координат. Подобное можно увидеть в сравнении векторов в работе [Faruqui, 2015], где показывается, что сходные слова обладают примерно одинаковым набором ненулевых значений, а слова отличающиеся имеют не так много общих ненулевых координат. И сам этот факт является вполне обычным для составителей онтологий.

Заключение

В данной работе был представлен метод семантической интерпретации пространств, заключающийся в выделении древовидной структуры в векторном пространстве с помощью многократного направленного применения метода разложения по собственным значениям.

Метод был применен для трех векторных моделей, отличавшихся использованными при обучении корпусами. Анализ результатов указывает как на общие закономерности, свойственные моделям, так и на индивидуальные свойства моделей. Так, прослеживается разделение на повседневное и квазиспециальное, материальное и нематериальное, абстрактное и конкретное. Описанные разделения, однако, возникают на различных по глубине уровнях дерева компонент, что не позволяет ввести четкую иерархию таких противопоставлений. Помимо смыслового сходства в разделении слов наблюдается зависимость от обучающего корпуса.

На первых этапах ветвления компонент обнаруживаются наиболее общие признаки, с большой вероятностью присущие всем или почти всем словам из списков. Первые разделения показывают стилистическую разницу: вероятно, потому что слова одного стиля чаще встречаются в текстах корпусов рядом друг с другом, чем со словами противоположного стиля.

Анализ указывает на представленность в векторном пространстве не только структур, соответствующих отдельным характеристикам слова, но и структур, соответствующих определенным типам дискурса. Более низкие уровни ветвления деревьев компонент, однако, не обнаруживают дискурсивных свойств.

Таким образом, предложенный подход к иерархической интерпретации статических векторных моделей позволяет установить закономерность в расположении векторов соответствующего пространства, и проинтерпретировать эту закономерность с качественной точки зрения. Предложенный метод также указывает на закономерности в самих корпусах, использованных для обучения моделей.

Список литературы

Грибова В. В., Петряева М. В., Окунь Д. Б., Шалфеева Е. А. Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решений // Онтология проектирования. 2018. Т. 8, № 1(27). С. 58–73.

Розенталь Д. Э., Теленкова М. А. Словарь-справочник лингвистических терминов. 2-е изд. М.: Просвещение, 1976, 543 с.

Adi Y. et al. Fine-grained analysis of sentence embeddings using auxiliary prediction tasks [Электронный ресурс]. URL: https://arxiv.org/abs/1608.04207 (дата обращения 01.09.2022).

Bocharov V., Bichineva S., Granovsky D., Ostapuk N., Stepanova M. Quality assurance tools in the OpenCorpora project // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 25–29 мая 2011 г.). Вып. 10(17). М.: РГГУ, 2011. С. 107–115.

Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology [Электронный ресурс]. Oxford University Press, 2004, pp. 267–270. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308795/ (дата обращения: 01.09.2022)

Chizhikova A., Murzakhmetov S., Serikov O., Shavrina T., Burtsev M. Attention Understands Semantic Relations // Proc. of the 13th Conference on Language Resources and Evaluation (LREC 2022), 2022, pp. 4040–4050.

Conneau A., Lample G., Ranzato M. A., Denoyer L., Jégou H. Word Translation Without Parallel Data. [Электронный ресурс]. URL: https://arxiv.org/abs/1710.04087 (дата обращения: 01.09.2022).

Conneau A. et al. What you can cram into a single vector: Probing sentence embeddings for linguistic properties [Электронный ресурс]. URL: https://arxiv.org/abs/1805.01070 (дата обращения: 01.09.2022).

Ethayarajh K. How contextual are contextualized word representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings // Proc. of Association for Computational Linguistics, Hong Kong, 2019, pp. 55–65.

Faruqui M., Tsvetkov Y., Yogatama D., Dyer C., Smith N. A. Sparse Overcomplete Word Vector Representations // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 1491–1500.

Gallant S. Context vector representations for document retrieval // Proc. of AAAI Workshop on Natural Language Text Retrieval, 1991.

Gustaf S. Meaning and change of meaning: with special reference to the English language. Indiana University Press, 1964, 490 p.

Korogodina, O., Karpik, O., Klyshinsky E. Evaluation of Vector Transformations for Russian Word2Vec and FastText Embeddings // Proc. of Graphicon-2020. DOI 10.51130/graphicon-2020-2-3-18

Kozlowski A., Taddy M., Evansa J. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings // American Sociological Review. 2017. Pp. 905–949.

Kutuzov A. Distributional word embeddings in modeling diachronic semantic change [Электронный ресурс] / Doctoral Thesis, University of Oslo, 2020. https://www.duo.uio.no/bitstream/handle/10852/81045/1/Kutuzov-Thesis.pdf.

Lasri K., Pimentel T., Lenci A., Poibeau T., Cotterell R. Probing for the Usage of Grammatical Number // Proc. of the 60th Annual Meeting of the Association for Computational Linguistics. 2022. Vol. 1, Pp. 8818–8831.

Linzen T., Dupoux E., Goldberg Y. Assessing the ability of LSTMs to learn syntax-sensitive dependencies // Transactions of the Association for Computational Linguistics. 2016. Vol. 4. Pp. 521–535.

Loureiro D., Alipio M. J. Language Modelling Makes Sense: Propagating Representations through WordNet for Full-Coverage Word Sense Disambiguation // Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. Pp. 5682–5691.

Representations. Virtual Event, Austria, May 3-7, 2021 [Электронный ресурс]. URL: https://openreview.net/forum?id=mNtmhaDkAr (дата обращения 01.09.2022)

Mikolov T., Chen K., Corrado G., Dean J. Efficient estimation of word representations in vector space // Proc. of International Conference on Learning Representations (ICLR), 2013 a.

Mikolov T., Chen K., Corrado G., Dean J. Distributed Representations of Words and Phrases and their Compositionality // Proc. of 27th Annual Conference on Neural Information Processing Systems. 2013. Pp. 3111–3119.

Rabinovich E., Xu Y., Stevenson S. The Typology of Polysemy: A Multilingual Distributional Framework, 2020 [Электронный ресурс]. URL: https://arxiv.org/abs/2006.01966v1 (дата обращения 01.09.2022).

Ravfogel S. et al. Counterfactual interventions reveal the causal effect of relative clause representations on agreement prediction [Электронный ресурс]. URL: https://arxiv.org/abs/2105.06965 (дата обращения 01.09.2022).

Rubinstein D., Levi E., Schwartz R., Rappoport A. How well do distributional models capture different types of semantic knowledge? [Электронный ресурс] // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Association for Computational Linguistics. 2015. Pp. 726–730. https://aclanthology.org/P15-2119.pdf.

Subramanian A., Pruthi D., Jhamtani H., Berg-Kirkpatrick T., Hovy E. SPINE: SParse Interpretable Neural Embeddings // The 32nd AAAI Conference on Artificial Intelligence (AAAI-18), 2018.

Tenney I., Das D., Pavlick E. BERT rediscovers the classical NLP pipeline. [Электронный ресурс]. URL: https://arxiv.org/abs/1905.05950 (дата обращения 01.09.2022).

Vig J. et al. Causal mediation analysis for interpreting neural nlp: The case of gender bias [Электронный ресурс]. URL: https://arxiv.org/abs/2004.12265 (дата обращения 01.09.2022).

Voloshina E., Serikov O., Shavrina T. Is neural language acquisition similar to natural? A chronological probing study // Proc. of Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2022”, 2022. Pp. 550–563.

Weeds J., Clarke D., Reffin J., Weir D., Keller B. Learning to distinguish hypernyms and co-hyponyms // Proceedings of COLING 2014. Dublin, the 25th International Conference on Computational Linguistics: Technical Papers, 2014. Pp. 2249–2259.

Yao S., Yu D., Xiao K. Enhancing Domain Word Embedding via Latent Semantic Imputation, 2019 [Электронный ресурс]. URL: https://arxiv.org/abs/1905.08900 (дата обращения 01.09.2022).

References

Gribova, V. V., Petryaeva, M. V., Okun, D. B., Shalfeeva, E. A. Medical Diagnosis Ontology for Intelligent Decision Support Systems. *Ontologiya Proektirovaniya* [Ontology of designing], 2018, vol. 8, no. 1(27), pp. 58–73. (in Russ.)

Rozental, D. E., Telenkova, M. A. Dictionary of Linguistic Terms [Slovar-Spravochnik Lingvisticheskikh Terminov]. 2nd ed. Moscow: Prosveschenie, 1976, 543 p. (in Russ.)

Adi, Y. et al. Fine-grained analysis of sentence embeddings using auxiliary prediction tasks [Online]. 2016 URL: https://arxiv.org/abs/1608.04207 (accessed on 01.09.2022).

Bocharov, V., Bichineva, S., Granovsky, D., Ostapuk, N., Stepanova, M. Quality assurance tools in the OpenCorpora project. In: Proc. of Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2011”. Moscow: RSUH, 2011, pp. 107-115

Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology [Online]. Oxford University Press, 2004, pp. 267–270. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC308795/ (accessed on: 01.09.2022)

Chizhikova, A., Murzakhmetov, S., Serikov, O., Shavrina, T., Burtsev, M. Attention Understands Semantic Relations. In: Proc. of the 13th Conference on Language Resources and Evaluation (LREC-2022), 2022, pp. 4040–4050.

Conneau, A., Lample, G., Ranzato, M. A., Denoyer, L., Jégou, H. Word Translation Without Parallel Data [Online]. URL: https://arxiv.org/abs/1710.04087 (accessed on: 01.09.2022).

Conneau, A. et al. What you can cram into a single vector: Probing sentence embeddings for linguistic properties [Online]. URL: https://arxiv.org/abs/1805.01070 (accessed on: 01.09.2022).

Ethayarajh, K. How contextual are contextualized word representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In Proc. of Association for Computational Linguistics, Hong Kong, 2019, pp. 55–65.

Faruqui, M., Tsvetkov, Y., Yogatama, D., Dyer, C., Smith, N. A. Sparse Overcomplete Word Vector Representations. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 1491–1500.

Gallant, S. Context vector representations for document retrieval. In: Proc. of AAAI Workshop on Natural Language Text Retrieval, 1991.

Gustaf, S. Meaning and change of meaning: with special reference to the English language. Indiana University Press, 1964, 490 p.

Korogodina , O., Karpik, O., Klyshinsky E. Evaluation of Vector Transformations for Russian Word2Vec and FastText Embeddings. In: Proc. of Graphicon 2020. DOI 10.51130/graphicon-2020-2-3-18

Kozlowski, A., Taddy, M., Evansa, J. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings. American Sociological Review, 2017, pp. 905–949.

Kutuzov, A. Distributional word embeddings in modeling diachronic semantic change [Online]. Doctoral Thesis, University of Oslo, 2020. URL: https://www.duo.uio.no/bitstream/handle/10852/81045/1/Kutuzov-Thesis.pdf.

Lasri, K., Pimentel, T., Lenci, A., Poibeau, T., Cotterell, R. Probing for the Usage of Grammatical Number. In: Proc. of the 60th Annual Meeting of the Association for Computational Linguistics, vol. 1, 2022, pp. 8818–8831.

Linzen, T., Dupoux, E., Goldberg, Y. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. Transactions of the Association for Computational Linguistics, 2016, vol. 4, pp. 521–535.

Loureiro, D., Alipio, M. J. Language Modelling Makes Sense: Propagating Representations through WordNet for Full-Coverage Word Sense Disambiguation. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019, pp. 5682–5691.

Representations. Virtual Event, Austria, May 3-7, 2021. URL: https://openreview.net/forum?id=mNtmhaDkAr (accessed on: 01.09.2022).

Mikolov, T., Chen, K., Corrado, G., Dean, J. Efficient estimation of word representations in vector space. Proc. of International Conference on Learning Representations (ICLR), 2013 a.

Mikolov, T., Chen, K., Corrado, G., Dean, J. Distributed Representations of Words and Phrases and their Compositionality. In: Proc. of 27th Annual Conference on Neural Information Processing Systems, 2013, pp. 3111–3119.

Rabinovich, E., Xu, Y., Stevenson, S. The Typology of Polysemy: A Multilingual Distributional Framework, 2020 [Online]. URL: https://arxiv.org/abs/2006.01966v1 (accessed on: 01.09.2022).

Ravfogel, S. et al. Counterfactual interventions reveal the causal effect of relative clause representations on agreement prediction [Online]. URL: https://arxiv.org/abs/2105.06965 (accessed on: 01.09.2022).

Rubinstein, D., Levi, E., Schwartz, R., Rappoport, A. How well do distributional models capture different types of semantic knowledge? In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Association for Computational Linguistics, 2015, pp. 726–730. https://aclanthology.org/P15-2119.pdf.

Subramanian, A., Pruthi, D., Jhamtani, H., Berg-Kirkpatrick, T., Hovy, E. SPINE: SParse Interpretable Neural Embeddings. The 32nd AAAI Conference on Artificial Intelligence (AAAI-18), 2018.

Tenney, I., Das, D., Pavlick, E. BERT rediscovers the classical NLP pipeline [Online]. URL: https://arxiv.org/abs/1905.05950 (accessed on: 01.09.2022).

Vig, J. et al. Causal mediation analysis for interpreting neural NLP: The case of gender bias [Online]. URL: https://arxiv.org/abs/2004.12265 (accessed on: 01.09.2022).

Voloshina, E., Serikov, O., Shavrina, T. Is neural language acquisition similar to natural? A chronological probing study. In Proc. of Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2022”. 2022. pp. 550-563

Weeds, J., Clarke, D., Reffin, J., Weir, D., Keller, B. Learning to distinguish hypernyms and co-hyponyms. In: Proceedings of COLING-2014. Dublin, the 25th International Conference on Computational Linguistics: Technical Papers, 2014, pp. 2249–2259.

Yao, S., Yu, D., Xiao, K. Enhancing Domain Word Embedding via Latent Semantic Imputation, 2019 [Online]: arXiv:1905.08900v1. URL: https://arxiv.org/abs/1905.08900 (accessed on: 01.09.2022).

Информация об авторах

Сериков Олег Алексеевич, исследователь, Школа Лингвистики НИУ ВШЭ; МФТИ; Институт искусственного интеллекта AIRI; Лаборатория исследования и сохранения малых языков ИЯЗ РАН

Ганеева Вероника Александровна, магистрант, НИУ ВШЭ

Аксенова Анна Александровна, исследователь данных, ПАО «Сбербанк»

Клышинский Эдуард Станиславович, доцент, канд. тех. наук, НИУ ВШЭ

Information about the Authors

Oleg A. Serikov, researcher at HSE University; MIPT; AIRI; Laboratory for Study and Preservation of Minority Languages of the Institute of Linguistics RAS

Veronika A. Geneeva, master student at HSE University

Anna A. Aksenova, data analyst, JSC Sberbank

Eduard S. Klyshinskiy, Assoc. Prof., PhD in CS, researcher at HSE University

Статья поступила в редакцию 06.09.2022;
одобрена после рецензирования 03.01.2023; принята к публикации 13.01.2023

The article was submitted 06.09.2022;
approved after reviewing 03.01.2023; accepted for publication 13.01.2023