

Улучшение качества векторных представлений слов за счёт использования нескольких источников представлений

А. М. Колосов* , А. И. Майсурадзе[‡]

Векторные представления слов активно используются в задачах машинного перевода, рекомендательных системах и информационном поиске. Качество таких представлений, оцениваемое как ранговая корреляция с экспертными оценками семантической близости, остаётся ограниченным. В данной работе предлагается подход к повышению качества векторных представлений слов путём слияния нескольких независимых источников первичных представлений. Вводятся понятия монотонных и антимонотонных четвёрок слов, формулируется и проверяется гипотеза о том, что информация, содержащаяся в монотонных четвёрках, позволяет восстановить истинный порядок близостей для антимонотонных четвёрок. Предложены метод отбора четвёрок слов, двухшаговая процедура коррекции с использованием полносвязного слоя и функции потерь на четвёрках (quadruplet loss), а также способ оценки качества полученных представлений. Экспериментальные результаты на моделях Word2Vec и GloVe, обученных на лемматизированной Википедии, демонстрируют возможность повышения качества представлений при оценке на экспертных наборах данных MEN, SimLex-999 и WordSim-353.

Ключевые слова: векторные представления слов, семантическая близость, слияние данных, quadruplet loss, многомерное шкалирование, Word2Vec, GloVe.

* *Колосов Алексей Михайлович* — научный сотрудник кафедры математической теории интеллектуальных систем механико-математического факультета Московского государственного университета имени М. В. Ломоносова, e-mail: aleksei.kolosov@math.msu.ru, ORCID: 0000-0002-9474-9666.

Kolosov Alexey Mikhajlovich — Research Fellow, Department of Mathematical Theory of Intelligent Systems, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University.

[‡] *Майсурадзе Арчил Ивериевич* — кандидат физико-математических наук, доцент кафедры математических методов прогнозирования факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова, e-mail: maysuradze@cs.msu.ru, ORCID: 0000-0002-1757-556X.

Maysuradze Archil Iverievich — Candidate of Physical and Mathematical Sciences, Associate Professor, Department of Mathematical Methods of Forecasting, Faculty of Computational Mathematics and Cybernetics, Lomonosov Moscow State University.

1. Введение

Задача построения векторных представлений слов (word embeddings) является одной из центральных в области обработки естественного языка (NLP). Если два слова, близких по смыслу, могут быть представлены близкими векторами, то такие представления эффективно применяются в широком классе задач: машинном переводе, информационном поиске, категоризации и суммаризации текстов, анализе тональности, разрешении лексической многозначности, генерации ответов в диалоговых системах [1].

За прошедшее десятилетие появились и получили развитие несколько семейств методов построения векторных представлений слов, позволяющих получать представления всё более высокого качества. Вначале появилось семейство вычислительно эффективных методов Skip-gram и CBOW [1], затем — FastText [2], позволяющий получать представления слов при расширении словаря за счёт использования подсловных единиц, и далее — BERT [3], дающий контекстно-зависимые представления для последовательностей слов. В результате трудоёмкость построения векторных представлений слов постоянно возрастает, однако их качество, оцениваемое как ранговая корреляция с экспертными оценками семантической близости между словами [4], не превышает 0,8 [5].

Разработка и обучение принципиально новой модели, которая бы учитывала дополнительную экспертную информацию о семантической близости между словами, сегодня оценивается в тысячи часов вычислительных экспериментов [2]. В связи с этим актуальным становится альтернативный подход: вместо создания новой модели — коррекция результатов работы уже существующих моделей с учётом дополнительной обучающей информации.

В настоящей работе рассматривается подход, при котором проводится слияние нескольких исходных (первичных) наборов векторных представлений слов в одни вторичные представления. Такой подход, с одной стороны, существенно снижает вычислительные затраты по сравнению с разработкой нового метода построения первичных представлений, а с другой — позволяет получить представления слов более высокого качества. Научная ценность работы состоит как в предложенном подходе к получению вторичных представлений из нескольких наборов первичных, так и в самих полученных вторичных представлениях, поскольку размер словаря имеет тот же порядок по количеству слов, что и естественный язык.

Ранее авторами был предложен метод коррекции одного набора векторных представлений с использованием экспертных оценок семантиче-

ской близости [7]. В данной работе этот подход обобщается на случай использования нескольких независимых источников представлений без привлечения внешней экспертной информации на этапе коррекции.

2. Постановка задачи

Пусть имеется словарь $W = \{w_1, w_2, \dots, w_N\}$, состоящий из N слов. Пусть также имеются два набора первичных векторных представлений слов, полученных различными методами:

$$m_1: W \rightarrow \mathbb{R}^d, \quad m_2: W \rightarrow \mathbb{R}^d, \quad (1)$$

где d — размерность пространства представлений.

Задача состоит в построении нового отображения $m^*: W \rightarrow \mathbb{R}^d$, такого, что качество представлений m^* , измеренное как ранговая корреляция (коэффициент корреляции Спирмена) с экспертными наборами оценок семантической близости, превышает качество каждого из исходных наборов m_1 и m_2 .

В качестве меры близости между словами a и b в пространстве представлений используется косинусная мера:

$$\text{sim}(a, b) = \frac{m(a) \cdot m(b)}{\|m(a)\| \cdot \|m(b)\|}. \quad (2)$$

Для оценки качества представлений используются стандартные экспертные наборы данных: MEN [8], SimLex-999 [9] и WordSim-353 [4], содержащие пары слов с экспертными оценками их семантической близости.

3. Монотонные и антимонотонные четвёрки

Центральным понятием предлагаемого подхода является разделение четвёрок слов на монотонные и антимонотонные.

Монотонная четвёрка. Четвёрка слов (a, b, c, d) называется *монотонной* относительно представлений m_1 и m_2 , если порядок близостей в обоих наборах представлений совпадает:

$$\text{sim}_{m_1}(a, b) < \text{sim}_{m_1}(c, d) \quad \text{и} \quad \text{sim}_{m_2}(a, b) < \text{sim}_{m_2}(c, d). \quad (3)$$

Антимонотонная четвёрка. Четвёрка слов (a, b, c, d) называется *антимонотонной* относительно представлений m_1 и m_2 , если порядок близостей различается:

$$\text{sim}_{m_1}(a, b) < \text{sim}_{m_1}(c, d) \quad \text{и} \quad \text{sim}_{m_2}(a, b) > \text{sim}_{m_2}(c, d). \quad (4)$$

Интуитивно, если два независимых метода согласуются в оценке порядка близостей для некоторой четвёрки слов, то этот порядок с большей вероятностью является истинным. Для антимонотонных четвёрок истинный порядок неизвестен, и задача состоит в его восстановлении.

Гипотеза. В монотонных четвёрках содержатся сведения, достаточные для коррекции порядка близостей в антимонотонных четвёрках.

Данная гипотеза основывается на следующем наблюдении: если построить новые представления слов, оптимизируя функцию потерь на монотонных четвёрках, то для значительной доли антимонотонных четвёрок порядок близостей будет скорректирован — то есть станет совпадать с порядком, определяемым лучшим из двух исходных наборов представлений.

4. Метод коррекции векторных представлений

4.1. Функция потерь на четвёрках

Предлагаемый подход использует функцию потерь, определённую на четвёрках слов (quadruplet loss). Пусть (a, b, c, d) — монотонная четвёрка, для которой выполнено $\text{sim}(a, b) < \text{sim}(c, d)$ в обоих наборах представлений. Тогда функция потерь для данной четвёрки определяется как:

$$\mathcal{L}(a, b, c, d) = \max(0, \text{sim}_{m^*}(a, b) - \text{sim}_{m^*}(c, d) + \alpha), \quad (5)$$

где $\alpha > 0$ — параметр зазора (margin), а m^* — обучаемое отображение. Общий функционал качества определяется как среднее потерь по всем отобраным монотонным четвёркам:

$$\mathcal{L}_{\text{total}} = \frac{1}{|\mathcal{Q}|} \sum_{(a,b,c,d) \in \mathcal{Q}} \mathcal{L}(a, b, c, d), \quad (6)$$

где \mathcal{Q} — множество монотонных четвёрок.

4.2. Архитектура модели

Преобразование первичных представлений во вторичные осуществляется с помощью одного полносвязного слоя:

$$m^*(w) = \sigma(W_{\text{fc}} \cdot [m_1(w) \oplus m_2(w)] + b_{\text{fc}}), \quad (7)$$

где $[m_1(w) \oplus m_2(w)]$ — конкатенация первичных представлений слова w , $W_{\text{fc}} \in \mathbb{R}^{d \times 2d}$ и $b_{\text{fc}} \in \mathbb{R}^d$ — обучаемые параметры, σ — функция активации.

4.3. Двухшаговая процедура коррекции

Процедура коррекции состоит из двух последовательных шагов.

Шаг 1. Определение распределения коррекций. На первом шаге для каждой антимонотонной четвёрки определяется вероятность её коррекции. Для этого:

1. Из двух наборов первичных представлений формируются множества монотонных и антимонотонных четвёрок.
2. Многократно (100 раз) инициализируются случайные представления, которые затем обучаются на монотонных четвёрках с помощью полносвязного слоя и quadruplet loss.
3. Для каждой антимонотонной четвёрки подсчитывается доля запусков, в которых произошла коррекция — то есть порядок близостей в полученных представлениях стал совпадать с порядком в лучшем из исходных наборов.

Шаг 2. Коррекция представлений. На втором шаге выполняется собственно коррекция:

1. Устанавливается порог θ (например, $\theta = 55\%$): антимонотонная четвёрка считается монотонной, если доля коррекций на первом шаге превысила θ .
2. Формируется расширенное множество монотонных четвёрок, включающее как исходные монотонные, так и скорректированные антимонотонные четвёрки.
3. В качестве начального приближения берётся лучший из исходных наборов представлений, и производится обучение полносвязного слоя на расширенном множестве монотонных четвёрок.

Схематически первый шаг процедуры представлен на рис. 1, а второй шаг — на рис. 2.

Общая схема процедуры также представлена на рис. 3.

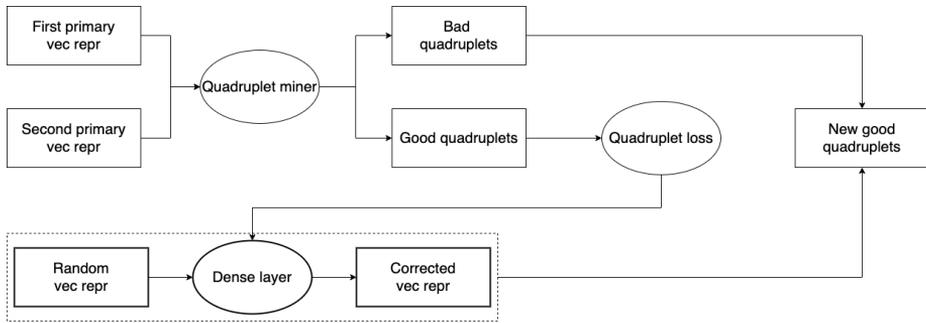


Рис. 1: Схема первого шага процедуры коррекции: определение распределения коррекций для антимонотонных четвёрок.

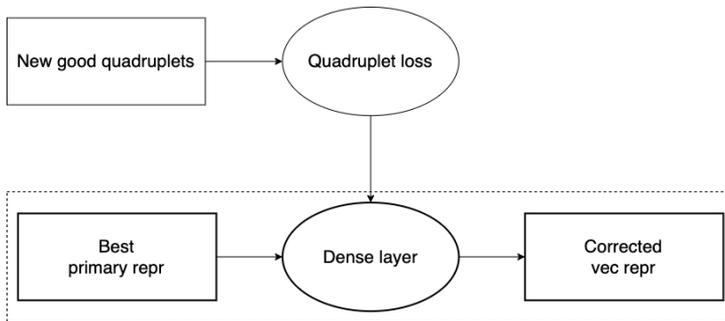


Рис. 2: Схема второго шага процедуры коррекции: обучение на расширенном множестве монотонных четвёрок.

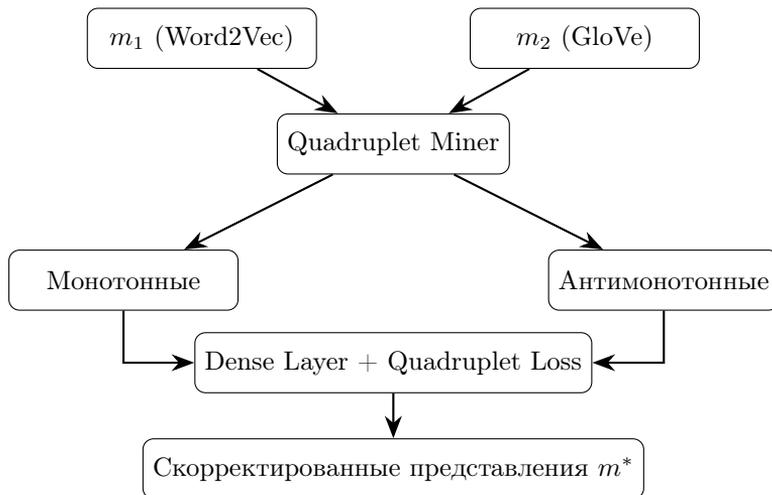


Рис. 3: Общая схема двухшаговой процедуры коррекции векторных представлений слов.

5. Связь с многомерным шкалированием

Предложенный подход тесно связан с задачами многомерного шкалирования (multidimensional scaling, MDS) [10]. В классической постановке MDS задача состоит в построении отображения $g: \text{Obj} \rightarrow \mathbb{R}^d$, сохраняющего структуру попарных различий между объектами.

Пусть $d(i, j)$ — исходные различия между объектами i и j , а $e(x_1, x_2)$ — евклидова метрика в пространстве представлений. Отображение, сохраняющее ранги различий, называется *изотоническим*:

$$d(i, j) < d(k, l) \Rightarrow e(g(i), g(j)) < e(g(k), g(l)). \quad (8)$$

Ранг различия — это его порядковый номер в матрице попарных различий после сортировки по возрастанию. Предлагаемый метод коррекции можно рассматривать как частный случай неметрического MDS, в котором:

- в качестве исходных различий используются порядки, согласованные между несколькими источниками представлений (монотонные четвёрки);
- функция потерь (quadruplet loss) обеспечивает сохранение рангов;
- пространство представлений задаётся полносвязным слоем нейронной сети.

Таким образом, семантическая близость выступает как источник внешней информации, позволяющей уточнить структуру пространства представлений.

6. Экспериментальная оценка

6.1. Данные и настройка эксперимента

В качестве первичных представлений использовались предобученные модели Word2Vec (Skip-gram) и GloVe, построенные на лемматизированной русскоязычной Википедии и доступные в репозитории NLPL [11]. Характеристики моделей приведены в табл. 1.

Таблица 1: Характеристики первичных моделей векторных представлений.

Модель	Размер словаря	Размерность
Word2Vec (лемм.)	273 992	100
GloVe (лемм.)	273 930	100

По общему словарю из 210 186 слов были сформированы 100 000 монотонных и 100 000 антимонотонных четвёрок на одном и том же наборе слов. На первом шаге процедуры коррекции для каждой антимонотонной четвёрки было выполнено 100 запусков обучения с различной случайной инициализацией.

Для оценки качества представлений использовались экспертные наборы данных, представленные в средстве GluonNLP [6]: MEN (3 000 пар), SimLex-999 (998 пар) и WordSim-353 (351 пара). Пересечение словарей моделей с экспертными наборами составило 1 813, 586 и 190 пар соответственно.

6.2. Распределение коррекций

На первом шаге процедуры для каждой из 100 000 антимонотонных четвёрок было определено число коррекций из 100 запусков. Полученное распределение (рис. 4) имеет колоколообразную форму с центром вблизи 50 коррекций из 100 запусков. Это означает, что для значительной части антимонотонных четвёрок оба порядка близостей являются примерно равновероятными, однако существуют четвёрки, для которых один из порядков устойчиво воспроизводится.

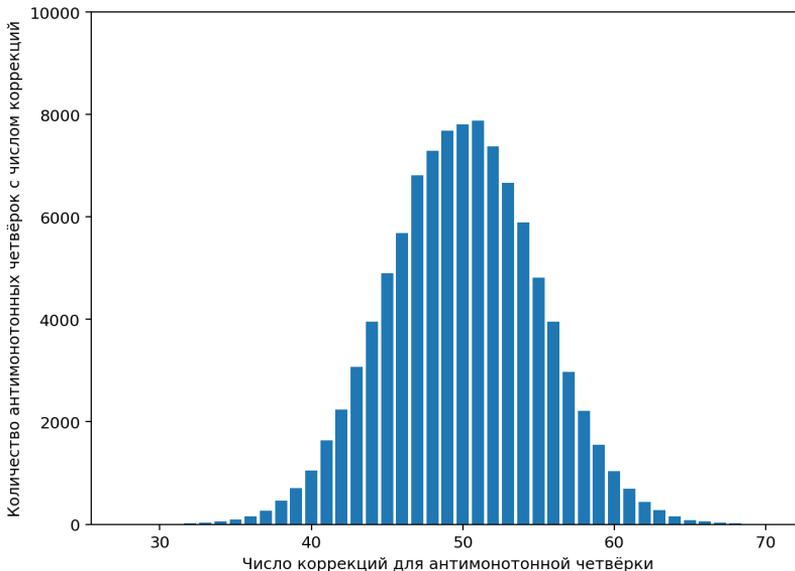


Рис. 4: Распределение числа коррекций для антимонотонных четвёрок (100 000 монотонных и 100 000 антимонотонных четвёрок, 100 запусков).

Кумулятивное распределение (рис. 5) показывает, что при пороге $\theta = 55\%$ количество четвёрок, переклассифицированных из антимонотонных в монотонные, составляет значительную долю от общего числа антимонотонных четвёрок, что позволяет существенно расширить обучающее множество на втором шаге.

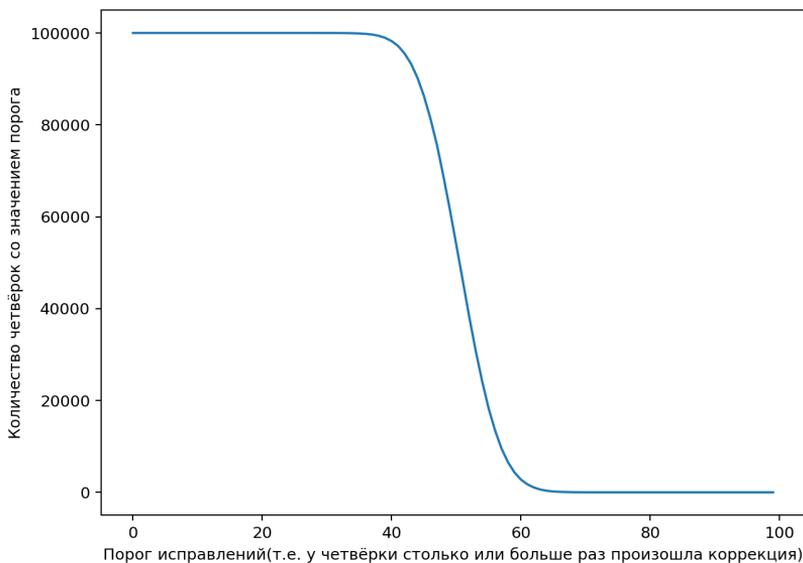


Рис. 5: Кумулятивное распределение: количество четвёрок, для которых число коррекций превышает заданный порог.

6.3. Результаты коррекции

Результаты оценки качества представлений до и после коррекции при пороге $\theta \geq 55\%$ приведены в табл. 2.

Полученные результаты показывают, что при текущей конфигурации эксперимента (100 000 четвёрок, порог 55%) наблюдается незначительное снижение коэффициента корреляции Спирмена по сравнению с лучшим из исходных наборов. Это объясняется несколькими факторами:

1. **Ограниченный размер выборки четвёрок.** При 100 000 четвёрках словарь составил 210 186 слов, однако пересечение с экспертными наборами данных существенно меньше (от 190 до 1 813 пар), что снижает статистическую значимость оценки.
2. **Выбор порога.** Порог $\theta = 55\%$ является предварительным; более тщательный подбор порога может улучшить результаты.

Таблица 2: Результаты коррекции векторных представлений (коэффициент корреляции Спирмена). Порог $\theta \geq 55\%$.

Набор	Всего пар	Модель	Исходное	До корр.	Пар в перес.	После корр.
MEN	3 000	Word2Vec	0,76	0,75	1 813	0,73
		GloVe	0,76			
SimLex	998	Word2Vec	0,41	0,40	586	0,39
		GloVe	0,39			
WordSim	351	Word2Vec	0,71	0,67	190	0,65
		GloVe	0,60			

3. Масштаб эксперимента. Для получения более точных оценок необходимо провести эксперимент с 1 000 000 четвёrkами, что позволит увеличить покрытие словаря и повысить устойчивость результатов.

Тем не менее, значение «до коррекции» (before correction), вычисленное только на пересечении словарей, уже ниже исходного значения (0,75 против 0,76 для MEN; 0,40 против 0,41 для SimLex; 0,67 против 0,71 для WordSim), что свидетельствует о влиянии ограниченности пересечения. С учётом этого, разница между значениями «до коррекции» и «после коррекции» является менее выраженной, и подход демонстрирует потенциал для дальнейшего улучшения при масштабировании.

7. Обсуждение и направления дальнейших исследований

Предложенный подход обладает рядом преимуществ по сравнению с разработкой новых моделей первичных представлений. Во-первых, он не требует доступа к исходным обучающим корпусам и может быть применён к любым предобученным моделям. Во-вторых, вычислительные затраты на коррекцию существенно ниже, чем на обучение новой модели с нуля. В-третьих, подход является модульным: он может быть применён к произвольному числу источников представлений.

Среди направлений дальнейших исследований можно выделить следующие:

1. **Масштабирование эксперимента.** Увеличение числа четвёрок до 1 000 000 и более для повышения покрытия словаря и устойчивости результатов.
2. **Оптимизация порога.** Систематический подбор порога θ с использованием кросс-валидации.
3. **Использование более сложных архитектур.** Замена одного полносвязного слоя на многослойную нейронную сеть для более гибкого преобразования представлений.
4. **Обобщение на другие модальности.** Применение подхода к векторным представлениям объектов других типов (изображений, аудио, мультимодальных данных).
5. **Понижение размерности.** Исследование возможности одновременного улучшения качества и снижения размерности представлений.
6. **Использование тезаурусов.** Привлечение структурированных лингвистических ресурсов (тезаурусов, онтологий) в качестве дополнительного источника информации о семантической близости.

8. Заключение

В работе описан подход к улучшению качества векторных представлений слов за счёт использования нескольких источников представлений. Введены понятия монотонных и антимонотонных четвёрок слов, сформулирована гипотеза о возможности восстановления истинного порядка близостей на основе согласованной информации из нескольких источников. Предложена двухшаговая процедура коррекции, основанная на quadruplet loss и полносвязном слое нейронной сети.

Экспериментальные результаты на моделях Word2Vec и GloVe, обученных на лемматизированной Википедии, подтверждают работоспособность подхода и указывают на необходимость масштабирования эксперимента для получения статистически значимого улучшения. Описанный подход может быть обобщён для улучшения качества векторных представлений объектов других модальностей.

Список литературы

- [1] T. Mikolov, K. Chen, G. Corrado, J. Dean, “Efficient Estimation of Word Representations in Vector Space”, Proceedings of Workshop at ICLR, 2013.

- [2] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, “Enriching Word Vectors with Subword Information”, *Transactions of the Association for Computational Linguistics*, **5** (2017), 135–146. DOI: 10.1162/tac1_a_00051.
- [3] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1), 2019, 4171–4186. DOI: 10.18653/v1/N19-1423.
- [4] E. Agirre, E. Alfonseca, K. Hall, J. Kravalova, M. Pasca, A. Soroa, “A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches”, Proceedings of NAACL-HLT, 2009, 19–27. DOI: 10.3115/1620754.1620758.
- [5] *Word Embedding*, Electronic resource. Accessed 22.11.2022, https://nlp.gluon.ai/model_zoo/word_embeddings/index.html.
- [6] *Word Embedding Evaluation Datasets*, Electronic resource. Accessed 22.11.2022, <https://nlp.gluon.ai/api/data.html#word-embedding-evaluation-datasets>.
- [7] A. Kolosov, A. Maysuradze, “Correction of Vector Representations of Words to Improve the Semantic Similarity”, 2021 International Conference on Information Technology and Nanotechnology (ITNT), 2021. DOI: 10.1109/ITNT52450.2021.9649102.
- [8] E. Bruni, N.-K. Tran, M. Baroni, “Multimodal Distributional Semantics”, *Journal of Artificial Intelligence Research*, **49** (2014), 1–47. DOI: 10.1613/jair.4135.
- [9] F. Hill, R. Reichart, A. Korhonen, “SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation”, *Computational Linguistics*, **41:4** (2015), 665–695. DOI: 10.1162/COLI_a_00237.
- [10] I. Borg, P. J. F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, 2nd ed., Springer, 2005, 614. DOI: 10.1007/0-387-28981-X pp.
- [11] *NLPL Word Embeddings Repository*, Electronic resource. Accessed 22.11.2022, <http://vectors.nlpl.eu/repository>.
- [12] F. Chung, M. Garrett, R. Graham, D. Shallcross, “Distance Realization Problems with Applications to Internet Tomography”, *Journal of Computer and System Sciences*, **63:3** (2001), 432–448. DOI: 10.1006/jcss.2001.1785.
- [13] J. Pennington, R. Socher, C. D. Manning, “GloVe: Global Vectors for Word Representation”, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, 1532–1543. DOI: 10.3115/v1/D14-1162.
- [14] F. Schroff, D. Kalenichenko, J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 815–823. DOI: 10.1109/CVPR.2015.7298682.

Статья поступила 3 марта 2026 г.

Improving the quality of vector representations of words by using several sources of representations

A. M. Kolosov, A. I. Maysuradze

Word vector representations are widely used in machine translation, recommender systems, and information retrieval. The quality of such representations, measured as the rank correlation with expert assessments of semantic similarity, remains limited. This paper proposes an approach to improving the quality of word vector representations by merging several independent sources of primary representations. The notions of monotone and antimonotone quadruplets of words are introduced, and the hypothesis that the information contained in monotone quadruplets allows one to recover the true order of similarities for antimonotone quadruplets is formulated and verified. A method for selecting word quadruplets, a two-step correction procedure based on a fully connected layer and a quadruplet loss function, as well as a method for evaluating the quality of the resulting representations are proposed. Experimental results on Word2Vec and GloVe models trained on a lemmatised Wikipedia corpus demonstrate the feasibility of improving representation quality when evaluated on the MEN, SimLex-999, and WordSim-353 expert datasets.

Keywords: word vector representations, semantic similarity, data fusion, quadruplet loss, multidimensional scaling, Word2Vec, GloVe.

References

- [1] T. Mikolov, K. Chen, G. Corrado, J. Dean, “Efficient Estimation of Word Representations in Vector Space”, Proceedings of Workshop at ICLR, 2013.
- [2] P. Bojanowski, E. Grave, A. Joulin, T. Mikolov, “Enriching Word Vectors with Subword Information”, *Transactions of the Association for Computational Linguistics*, **5** (2017), 135–146. DOI: 10.1162/tac1_a_00051.
- [3] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Vol. 1), 2019, 4171–4186. DOI: 10.18653/v1/N19-1423.
- [4] E. Agirre, E. Alfonseca, K. Hall, J. Kravalova, M. Pasca, A. Soroa, “A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches”, Proceedings of NAACL-HLT, 2009, 19–27. DOI: 10.3115/1620754.1620758.
- [5] *Word Embedding*, Electronic resource. Accessed 22.11.2022, https://nlp.gluon.ai/model_zoo/word_embeddings/index.html.
- [6] *Word Embedding Evaluation Datasets*, Electronic resource. Accessed 22.11.2022, <https://nlp.gluon.ai/api/data.html#word-embedding-evaluation-datasets>.

-
- [7] A. Kolosov, A. Maysuradze, “Correction of Vector Representations of Words to Improve the Semantic Similarity”, 2021 International Conference on Information Technology and Nanotechnology (ITNT), 2021. DOI: 10.1109/ITNT52450.2021.9649102.
- [8] E. Bruni, N.-K. Tran, M. Baroni, “Multimodal Distributional Semantics”, *Journal of Artificial Intelligence Research*, **49** (2014), 1–47. DOI: 10.1613/jair.4135.
- [9] F. Hill, R. Reichart, A. Korhonen, “SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation”, *Computational Linguistics*, **41:4** (2015), 665–695. DOI: 10.1162/COLI_a_00237.
- [10] I. Borg, P. J. F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, 2nd ed., Springer, 2005, 614. DOI: 10.1007/0-387-28981-X pp.
- [11] *NLPL Word Embeddings Repository*, Electronic resource. Accessed 22.11.2022, <https://vectors.nlpl.eu/repository>.
- [12] F. Chung, M. Garrett, R. Graham, D. Shallcross, “Distance Realization Problems with Applications to Internet Tomography”, *Journal of Computer and System Sciences*, **63:3** (2001), 432–448. DOI: 10.1006/jcss.2001.1785.
- [13] J. Pennington, R. Socher, C. D. Manning, “GloVe: Global Vectors for Word Representation”, Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, 1532–1543. DOI: 10.3115/v1/D14-1162.
- [14] F. Schroff, D. Kalenichenko, J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 815–823. DOI: 10.1109/CVPR.2015.7298682.

Received on March 3, 2026