НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
июль–август 2024 Том 24 № 4 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
July– August 2024 Vol. 24 No 4 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-4-588-593

Advanced methods for knowledge injection in large language models

Nikita I. Kulin[1]✉, Sergey B. Muravyov[2]

[1,2] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] kylin98@list.ru✉, https://orcid.org/0000-0002-3952-6080
[2] smuravyov@itmo.ru, https://orcid.org/0000-0002-4251-1744

Abstract

Transformer-based language models have revolutionized Natural Language Processing tasks, with advancements in language modeling techniques. Current transformer architectures utilize attention mechanisms to model text dependencies effectively. Studies have shown that these models embed syntactic structures and knowledge, explaining their performance in tasks involving syntactic and semantic elements. However, transformer-based models are prone to hallucination where incorporated knowledge is not utilized effectively. To address this, methods are emerging to mitigate hallucination and integrate external knowledge sources like knowledge graphs (e.g., Freebase, WordNet, ConceptNet, ATOMIC). Knowledge graphs represent real-world knowledge through entities and relationships offering a potential injection point to enhance model performance in inference tasks. Various injection approaches, including input, architectural, and output injections, aim to incorporate knowledge from graphs into transformer models. Input injections modify data preprocessing, architectural injections add layers for knowledge integration, and output injections adjust error functions to correct knowledge incorporation during training. Despite ongoing research, a universal solution to hallucination remains elusive, and a standardized benchmark for comparing injection methods is lacking. This study investigates knowledge graphs as one of the methods to mitigate hallucination and their possible integration into Large Language Models. Comparative experiments across General Language Understanding Evaluation benchmark tasks demonstrated that ERNIE 3.0 and XLNet outperform other injection methods with the average scores of 91.1 % and 90.1 %.

Keywords

LLM, knowledge graphs, knowledge injection methods, hallucination problem, BERT

For citation: Kulin N.I., Muravyov S.B. Advanced methods for knowledge injection in large language models. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 4, pp. 588–593. doi: 10.17586/2226-1494-2024-24-4-588-593

УДК 004.89

Продвинутые методы внедрения знаний в больших языковых моделях

Никита Игоревич Кулин[1]✉, Сергей Борисович Муравьев[2]

[1,2] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] kylin98@list.ru✉, https://orcid.org/0000-0002-3952-6080
[2] smuravyov@itmo.ru, https://orcid.org/0000-0002-4251-1744

Аннотация

Трансформерные языковые модели революционизировали Natural Language Processing задачи благодаря достижениям в методах моделирования языка. Текущие архитектуры трансформеров используют механизмы внимания для эффективного моделирования текстовых зависимостей. Исследования показали, что эти модели встраивают синтаксические структуры и знания, объясняя их эффективность в задачах, связанных с синтаксическими и семантическими элементами. Однако трансформаторные модели склонны к галлюцинациям, когда встроенные знания не используются эффективно. Для решения этой проблемы появляются методы, направленные на снижение галлюцинаций и интеграцию внешних источников знаний, таких как графы знаний (например, Freebase, WordNet, ConceptNet, ATOMIC). Графы знаний представляют реальные знания через сущности и отношения, предлагая потенциальную точку внедрения для повышения производительности модели

в задачах вывода. Различные подходы к внедрениям, включая внедрения входных и выходных данных, а также архитектурные, направлены на включение знаний из графов в трансформерные модели. Внедрения входных данных модифицируют предварительную обработку данных, архитектурные добавляют слои для интеграции знаний, а внедрения выходных данных корректируют функции ошибок для правильного включения знаний во время обучения. Несмотря на продолжающиеся исследования, универсальное решение проблемы галлюцинаций и стандартизированный бенчмарк для сравнения методов внедрения знаний отсутствуют. В данном исследовании рассматриваются графы знаний как один из методов решения галлюцинаций и их возможная интеграция в большие языковые модели. Сравнительные эксперименты на бенчмарке General Language Understanding Evaluation показали, что ERNIE 3.0 и XLNet превосходят другие методы внедрения со средними оценками 91,1 % и 90,1 %.

Ключевые слова

LLM, графы знаний, методы внедрения знаний, проблема галлюцинаций, BERT

Ссылка для цитирования: Кулин Н.И., Муравьев С.Б. Продвинутые методы внедрения знаний в больших языковых моделях // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 4. С. 588–593 (на англ. яз.) doi: 10.17586/2226-1494-2024-24-4-588-593

Introduction

Recently, the use of language models of transformer architecture [1] has greatly improved the quality of solving a huge range of Natural Language Processing (NLP) tasks. The key difference that has determined the success of transformers such as Bidirectional Encoder Representations from Transformers (BERT) [2] or XLNet [3] is the use of self-attention mechanisms. Self-attention helps to efficiently construct dependencies between words in the text, can be parallelized by computing attention for an individual token, and considers all tokens of the same text sequence under consideration, unlike previous RNN-based architectures.

Subsequently, it has been shown that transformational models have hallucination problems, which means generating text with false-semantic or false-factual assertions. One solution to this problem is to embed information from external sources into transformers. A promising approach is knowledge graph integration which is a data graph with nodes representing entities of interest and edges representing relationships between those entities. Among such knowledge graphs are ATOMIC, ConceptNet, or WordNet.

To date, there is no universal approach to implementing information through knowledge graphs. Among current solutions, three implementation options stand out: input injection, architectural injection, and output injection. Moreover, there are no unified benchmarks for comparing the effectiveness of knowledge injection approaches, and accordingly, no comparison of current frameworks has been made.

The objective of this study was to investigate potential methods with knowledge graph integration in order to advance the development of the subject area. This involved examining existing solutions, their technical attributes, and the underlying methods. Following this analysis, a comparative experiment was performed to assess the performance of the selected methods across various benchmark tasks.

Methods

This section discusses four types of knowledge injection with knowledge graphs: input, architecture, output, and hybrid. Input injection involves converting knowledge graph assertions into words and structuring input data to include semantic information from triplets in the knowledge graph. Architecture injections focus on architectural changes. Output injections change the output structure (layer or loss function) used in the language model in some way to incorporate knowledge. Hybrid injections include a combination of knowledge injection methods.

Some methods have been discussed in this scientific study [4]. Over time, these approaches have been refined and new ones have emerged, which motivates the construction of an experiment to compare their effectiveness. To begin, we will briefly discuss these methods. As an input injection method we focus on Align, Mask and Select (AMS) one [5]. AMS is a simple approach for incorporating knowledge into language models. This method involves creating a dataset by aligning a knowledge graph with plain text to generate questions and possible answers. The dataset is then used to pre-train a neural language representation model, such as BERT, to enhance its knowledge. AMS includes the following three steps. The knowledge graph is aligned with plain text, such as sentences from English Wikipedia. This alignment involves matching concepts from the knowledge graph with corresponding concepts in the plain text. Then, one concept from the aligned text is masked with a special token. This masked concept is used to generate plausible answers by considering neighboring concepts in the knowledge graph with the same masked token and relationship.

As an architecture injection, we discuss two approaches: OM-ADAPT [6] and KnowBERT [7]. OM-ADAPT reflects the impact of fine-tuning a BERT version equipped with adapters [8] on ConceptNet to incorporate knowledge [4]. By training models on sentences from the Open Mind Common Sense [9] corpus and traversing the ConceptNet graph, the researchers enhanced the encoded “World Knowledge” by using simple adapters and a small number of update steps on training sentences [4].

KnowBERT, developed as an enhancement to BERT, incorporates Knowledge Attention and Recontextualization (KAR) layers that utilize graph entity embeddings generated through Tucker Tensor Decompositions for knowledge graph completion [4, 10]. The KAR layers process embeddings through an attention mechanism to

create entity span embeddings which are added to BERT existing contextual representations [4]. This approach injects knowledge in later layers to promote training stability, though it could potentially overlook important knowledge due to the unfreezing of the BERT model after training the entity linker.

As an output injection method, we focus on SemBERT [11]. It utilizes a subsystem that generates embedding representations of the output of a semantic role labeling system. This representation is then combined with the contextualized representation from BERT to integrate relational knowledge [4].

As hybrid injection methods, we focus on various approaches, such as LIBERT [12], BERT-MK [13], K-BERT [14], KG-BERT [15], K-Adapter [16], ERNIE 3.0 [17], and XLNet [18].

LIBERT [12] processes batches of lexical constraints including synonyms and hyponyms/hypernyms represented as word tuples and negative examples to fit the BERT format for applications in natural language processing. The formatted input is then processed through BERT, with the [CLS] token serving as input for a softmax classifier to determine lexical relation validity.

BERT-MK [13] includes KG-transformer modules that merge transformer layers with entity representations learned from separate transformer layers trained on a KG converted into natural language sentences [4]. Additional layers incorporate an attention mask to replicate KG connections and integrate graph structure into embeddings [4].

K-BERT [14] inserts relevant triples for entities in a sentence and a knowledge graph (KG) with soft-position embeddings for ordering [4]. A masked self-attention mechanism similar to BERT-MK focuses on incorporating essential knowledge.

KG-BERT [15] fine-tunes the BERT model using text from KG triples for KG triple completion, binary classification for predicting triple authenticity, and multi-class classification for relationship type prediction.

K-Adapter [16] incorporates projection layers before and after specific transformer layers within a pre-trained RoBERTa model [4]. K-adapter utilizes the following external knowledge: factual knowledge from Wikipedia triples and linguistic knowledge from Stanford outputs [4].

ERNIE 3.0 [17] is a large-scale knowledge-enhanced pre-training model for language understanding and generation, part of a unified framework combining auto-regressive network and auto-encoding for customization in natural language tasks.

XLNet [18] is combined with a Graph Convolutional Network (GCN) for the consolidation of graph knowledge and contextual information, not for origin knowledge injection [4]. The researchers have incorporated XLNet embeddings into GCN to handle answering questions [4]. They extract relevant subgraphs from ConceptNet and Wikipedia, which contain relations among entities in a question-answer setup taken from ConceptNet, along with the top 10 most pertinent sentences from Wikipedia data through ElasticSearch [4].

The methods considered utilize different types of injections, and they were evaluated on different benchmarks and metrics in their articles and have not been previously compared with each other. In order to determine the most effective approach and most promised type of injection, we need to conduct our own experimentation.

Comparative evaluation of methods

We first briefly describe and discuss the experimental setup and comparison results.

Experimental Setup

In the current investigation, an uncased BERT-base model, composed of 12 transformer layers with 12 attention heads each, was employed to use knowledge injection techniques. The source knowledge graphs were procured for every knowledge injection method used.

These knowledge injection methodologies were evaluated using the tasks from the evaluation benchmark, commonly known as General Language Understanding Evaluation (GLUE) [19]. GLUE is a collection of datasets routinely used for training, evaluation, and comparative analysis of NLP models. The overarching goal is to propel progress in creating comprehensive and robust natural language understanding systems. The benchmark includes diverse and challenging task datasets that specifically assess a model competency in language comprehension. For our experiment, the following tasks were chosen: CoLA, SST-2, MRPC, STS-B, QQP, MNLI, QNLI and RTE.

Experimental results

Table provides a comparative performance analysis of various knowledge injection methods on the GLUE benchmark tasks. The experiment implemented a range of injection methods, both hybrid and non-hybrid: Input injection methods, such as AMS and COMET; Architectural injection methods like KnowBERT and OM-ADAPT; Output injection method namely SemBERT; Hybrid injection methods including LIBERT, BERT-MK, K-BERT, KG-BERT, K-Adapter, ERNIE 3.0, and XLNet. The foundational model adopted for the basis of comparison was the origin BERT-base.

In terms of performance, ERNIE 3.0 surpasses other methods by registering an average improvement of 12.05 %, thus showcasing the efficacy of combining input and output injections. ERNIE 3.0 excels in nearly all GLUE tasks, the only exception being QQP (Question Paraphrases), where KnowBERT outperforms ERNIE 3.0 by a margin of 21.14 %. XLNet emerges as the second-best performer, delivering promising results on average. Interestingly, methods employing non-hybrid injections (Align-Mask-Select, SemBERT, KnowBERT) are more effective than some that utilize hybrid injections, such as LIBERT and K-BERT. Methods producing lower average scores in relation to the origin BERT-base include OM-ADAPT, SemBERT, and KG-BERT.

Discussion

Experimental results delineate the efficacy of various knowledge injection methods in addressing the issue of hallucination. The results demonstrated that ERNIE 3.0 and XLNet, which are hybrid combination approaches, outperform others in the evaluation using GLUE

Table. Comparing Knowledge Injection Methods for GLUE tasks

Knowledge Injection Method	Knowledge Sources	Average Score	MNLI, m	MNLI, mm	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE
BERT-base	—	81.3	84.6	82.9	88.9	90.5	93.5	52.1	85.8	88.9	66.4
Align, Mask, Select	ConceptNet	82.1	84.7	83.9	72.1	91.2	93.6	54.3	86.4	85.9	69.5
KnowBERT	Wikipedia, WordNet	83.8	85.7	84.3	90.3	91.5	93.6	54.6	89.1	88.2	73.8
OM-ADAPT (50K)	OMCS	75.5	84.2	83.5	71.6	90.6	93.9	49.8	85.8	88.9	69.7
OM-ADAPT (100K)	OMCS	74.6	83.9	82.8	71.5	90.8	92.8	48.8	85.7	87.1	64.1
SemBERT	Semantic Role Labeling of Pre-Training data	80.9	84.4	84.0	71.8	90.9	93.5	57.8	87.3	88.2	69.3
LIBERT	Wordnet, Roget's Thesaurus	74.3	79.8	78.8	69.3	87.2	90.8	35.3	82.6	86.6	63.6
BERT-MK	Unified Medical Language System	83.0	88.5	87.6	74.1	91.2	93.2	63.4	88.4	89.1	73.2
K-BERT	TBD	76.1	82.8	81.4	70.2	90.4	92.5	48.2	82.7	86.9	69.2
KG-BERT	ConceptNet	80.1	85.2	83.7	71.4	91.3	93.6	55.8	83	88.5	69.1
K-Adapter	Wikipedia	82.2	84.1	83.3	71.3	91.1	93.4	52.1	87.1	85.2	70.1
ERNIE 3.0	Wikipedia, Reddit	91.1	92.3	91.7	75.2	97.3	97.8	75.5	93.0	93.9	92.6
XLNet	Wikipedia, ConceptNet	90.1	90.9	90.9	74.7	90.5	97.1	70.2	93.0	92.9	88.5

benchmark tasks, thereby substantiating the ongoing exploration of these selected methods and their possible modifications.

The results demonstrate that non-hybrid approaches are generally outperformed by hybrid combinations. However, some hybrid methods, such as LIBERT and K-BERT, are found to be less effective than non-hybrid methods, implying certain limitations in the compatibility of injections with each other.

Notably, the experiment revealed that approaches employing knowledge graphs are capable of structuring knowledge within language models and assisting in solving NLP tasks. Among knowledge graphs, WordNet and ConceptNet have established their potential as the tools for knowledge enrichment within language models, providing motivation to incorporate other knowledge graphs, such as ATOMIC and Freebase, in the injection methods mentioned in the beginning of this section.

In addition to seeking the best knowledge injection approach, the future research agenda may include identifying the most effective configuration of the language model and the knowledge injection approach, to maximize the benefits in addressing the problem of hallucination. This course of action could serve to stimulate further research in this field.

Conclusion

The paper explores various methods for incorporating structured knowledge into language models to enhance their ability to efficiently address NLP problems. The results have shown that ERNIE 3.0 and XLNet, with their combined knowledge injection techniques are the most promising for effectively tackling the hallucination issue.

References

1. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017, vol. 30.
2. Devlin J., Chang M.-W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding. *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. V. 1 (Long and Short Papers)*, 2019, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423
3. Yang Z., Dai Z., Yang Y., Carbonell J., Salakhutdinov R.R., Le Q.V. Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 2019, vol. 32.
4. Colon-Hernandez P., Havasi C., Alonso J., Huggins M., Breazeal C. Combining pre-trained language models and structured knowledge. *arXiv*, 2021, arXiv:2101.12294. https://doi.org/10.48550/arXiv.2101.12294

Литература

1. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., Polosukhin I. Attention is all you need // Advances in Neural Information Processing Systems. 2017. V. 30.
2. Devlin J., Chang M.-W., Lee K., Toutanova K. Bert: Pre-training of deep bidirectional transformers for language understanding // Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. V. 1 (Long and Short Papers). 2019. P. 4171–4186. https://doi.org/10.18653/v1/N19-1423
3. Yang Z., Dai Z., Yang Y., Carbonell J., Salakhutdinov R.R., Le Q.V. Xlnet: Generalized autoregressive pretraining for language understanding // Advances in Neural Information Processing Systems. 2019. V. 32.
4. Colon-Hernandez P., Havasi C., Alonso J., Huggins M., Breazeal C. Combining pre-trained language models and structured knowledge // arXiv. 2021. arXiv:2101.12294. https://doi.org/10.48550/arXiv.2101.12294

5. Ye Z.-X., Chen Q., Wang W., Ling Z.-H. Align, mask and select: A simple method for incorporating commonsense knowledge into language representation models. *arXiv*, 2019, arXiv:1908.06725. https://doi.org/10.48550/arXiv.1908.06725
6. Lauscher A., Majewska O., Ribeiro L.F.R., Gurevych I., Rozanov N., Glavaš G. Common sense or world knowledge? investigating adapter-based knowledge injection into pretrained transformers. *Proc. of Deep Learning Inside Out (DeeLIO): The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, 2020, pp. 43–49. https://doi.org/10.18653/v1/2020.deelio-1.5
7. Peters M.E., Neumann M., Logan R., Schwartz R., Joshi V., Singh S., Smith N.A. Knowledge enhanced contextual word representations. *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 43–54. https://doi.org/10.18653/v1/D19-1005
8. Houlsby N., Giurgiu A., Jastrzebski S., Morrone B., De Laroussilhe Q., Gesmundo A., Attariyan M., Gelly S. Parameter-efficient transfer learning for NLP. *International Conference on Machine Learning, PMLR*, 2019, vol. 97, pp. 2790–2799.
9. Singh P., Lin T., Mueller E.T., Lim G., Perkins T., Zhu W.L. Open mind common sense: Knowledge acquisition from the general public. *Lecture Notes in Computer Science*, 2002, vol. 2519, pp. 1223–1237. https://doi.org/10.1007/3-540-36124-3_77
10. Balažević I., Allen C., Hospedales T.M. TuckER: Tensor factorization for knowledge graph completion. *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 5185–5194. https://doi.org/10.18653/v1/D19-1522
11. Zhang Z., Wu Y., Zhao H., Li Z., Zhang S., Zhou X., Zhou X. Semantics-aware BERT for language understanding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 05, pp. 9628–9635. https://doi.org/10.1609/aaai.v34i05.6510
12. Lauscher A., Vulić I., Ponti E.M., Korhonen A., Glavaš G. Specializing Unsupervised Pretraining Models for Word-Level Semantic Similarity. *Proc. of the 28th International Conference on Computational Linguistics (COLING 2020)*, 2020, pp. 1371–1383. https://doi.org/10.18653/v1/2020.coling-main.118
13. He B., Zhou D., Xiao J., Jiang X., Liu Q., Yuan N.J., Xu T. BERT-MK: Integrating graph contextualized knowledge into pre-trained language models. *Proc. of the Findings of the Association for Computational Linguistics (EMNLP 2020)*, 2020, pp. 2281–2290. https://doi.org/10.18653/v1/2020.findings-emnlp.207
14. Liu W., Zhou P., Zhao Z., Wang Z., Ju Q., Deng H., Wang P. K-bert: Enabling language representation with knowledge graph. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34. N 03, pp. 2901–2908. https://doi.org/10.1609/aaai.v34i03.5681
15. Yao L., Mao C., Luo Y. KG-BERT: BERT for knowledge graph completion. *arXiv*, 2019, arXiv:1909.03193. https://doi.org/10.48550/arXiv.1909.03193
16. Wang R., Tang D., Duan N., Wei Z., Huang X., Ji J., Cao G., Jiang D., Zhou M. K-adapter: Infusing knowledge into pre-trained models with adapters. *Proc. of the Findings of the Association for Computational Linguistics (ACL-IJCNLP 2021)*, 2021, pp. 1405–1418. https://doi.org/10.18653/v1/2021.findings-acl.121
17. Sun Y., Wang S., Feng S., Ding S., Pang C., Shang J., Liu J., Chen X., Zhao Y., Lu Y., Liu W., Wu Z., Gong W., Liang J., Shang Z., Sun P., Liu W., Ouyang X., Yu D., Tian H., Wu H., Wang H. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv*, 2021, arXiv:2107.02137. https://doi.org/10.48550/arXiv.2107.02137
18. Lv S., Guo D., Xu J., Tang D., Duan N., Gong M., Shou L., Jiang D., Cao G., Hu S. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering. *Proceedings of the AAAI conference on artificial intelligence*, 2020, vol. 34, no. 05, pp. 8449–8456. https://doi.org/10.1609/aaai.v34i05.6364
19. Wang A., Singh A., Michael J., Hill F., Levy O., Bowman S. GLUE: A multi-task benchmark and analysis platform for natural language understanding. *Proc. of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, 2018, pp. 353–355. https://doi.org/10.18653/v1/W18-5446

5. Ye Z.-X., Chen Q., Wang W., Ling Z.-H. Align, mask and select: A simple method for incorporating commonsense knowledge into language representation models // arXiv. 2019. arXiv:1908.06725. https://doi.org/10.48550/arXiv.1908.06725
6. Lauscher A., Majewska O., Ribeiro L.F.R., Gurevych I., Rozanov N., Glavaš G. Common sense or world knowledge? investigating adapter-based knowledge injection into pretrained transformers // Proc. of Deep Learning Inside Out (DeeLIO): The First Workshop on Knowledge Extraction and Integration for Deep Learning Architectures. 2020. P. 43–49. https://doi.org/10.18653/v1/2020.deelio-1.5
7. Peters M.E., Neumann M., Logan R., Schwartz R., Joshi V., Singh S., Smith N.A. Knowledge enhanced contextual word representations // Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019. P. 43–54. https://doi.org/10.18653/v1/D19-1005
8. Houlsby N., Giurgiu A., Jastrzebski S., Morrone B., De Laroussilhe Q., Gesmundo A., Attariyan M., Gelly S. Parameter-efficient transfer learning for NLP // International Conference on Machine Learning, PMLR. 2019. V. 97. P. 2790–2799.
9. Singh P., Lin T., Mueller E.T., Lim G., Perkins T., Zhu W.L. Open mind common sense: Knowledge acquisition from the general public // Lecture Notes in Computer Science. 2002. V. 2519. P. 1223–1237. https://doi.org/10.1007/3-540-36124-3_77
10. Balažević I., Allen C., Hospedales T.M. TuckER: Tensor factorization for knowledge graph completion // Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019. P. 5185–5194. https://doi.org/10.18653/v1/D19-1522
11. Zhang Z., Wu Y., Zhao H., Li Z., Zhang S., Zhou X., Zhou X. Semantics-aware BERT for language understanding // Proceedings of the AAAI Conference on Artificial Intelligence. 2020. V. 34. N 05. P. 9628–9635. https://doi.org/10.1609/aaai.v34i05.6510
12. Lauscher A., Vulić I., Ponti E.M., Korhonen A., Glavaš G. Specializing Unsupervised Pretraining Models for Word-Level Semantic Similarity // Proc. of the 28th International Conference on Computational Linguistics (COLING 2020). 2020. P. 1371–1383. https://doi.org/10.18653/v1/2020.coling-main.118
13. He B., Zhou D., Xiao J., Jiang X., Liu Q., Yuan N.J., Xu T. BERT-MK: Integrating graph contextualized knowledge into pre-trained language models // Proc. of the Findings of the Association for Computational Linguistics (EMNLP 2020). 2020. P. 2281–2290. https://doi.org/10.18653/v1/2020.findings-emnlp.207
14. Liu W., Zhou P., Zhao Z., Wang Z., Ju Q., Deng H., Wang P. K-bert: Enabling language representation with knowledge graph // Proceedings of the AAAI Conference on Artificial Intelligence. 2020. V. 34. N 03. P. 2901–2908. https://doi.org/10.1609/aaai.v34i03.5681
15. Yao L., Mao C., Luo Y. KG-BERT: BERT for knowledge graph completion // arXiv. 2019. arXiv:1909.03193. https://doi.org/10.48550/arXiv.1909.03193
16. Wang R., Tang D., Duan N., Wei Z., Huang X., Ji J., Cao G., Jiang D., Zhou M. K-adapter: Infusing knowledge into pre-trained models with adapters // Proc. of the Findings of the Association for Computational Linguistics (ACL-IJCNLP 2021). 2021. P. 1405–1418. https://doi.org/10.18653/v1/2021.findings-acl.121
17. Sun Y., Wang S., Feng S., Ding S., Pang C., Shang J., Liu J., Chen X., Zhao Y., Lu Y., Liu W., Wu Z., Gong W., Liang J., Shang Z., Sun P., Liu W., Ouyang X., Yu D., Tian H., Wu H., Wang H. Ernie 3.0: Large-scale knowledge enhanced pre-training for language understanding and generation // arXiv. 2021. arXiv:2107.02137. https://doi.org/10.48550/arXiv.2107.02137
18. Lv S., Guo D., Xu J., Tang D., Duan N., Gong M., Shou L., Jiang D., Cao G., Hu S. Graph-based reasoning over heterogeneous external knowledge for commonsense question answering // Proceedings of the AAAI conference on artificial intelligence. 2020. V. 34. N 05. P. 8449–8456. https://doi.org/10.1609/aaai.v34i05.6364
19. Wang A., Singh A., Michael J., Hill F., Levy O., Bowman S. GLUE: A multi-task benchmark and analysis platform for natural language understanding // Proc. of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. 2018. P. 353–355. https://doi.org/10.18653/v1/W18-5446

Authors

Nikita I. Kulin — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57222386134, https://orcid.org/0000-0002-3952-6080, kylin98@list.ru
Sergey B. Muravyov — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57194035005, https://orcid.org/0000-0002-4251-1744, smuravyov@itmo.ru

Авторы

Кулин Никита Игоревич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация SC 57222386134, https://orcid.org/0000-0002-3952-6080, kylin98@list.ru
Муравьев Сергей Борисович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 57194035005, https://orcid.org/0000-0002-4251-1744, smuravyov@itmo.ru

Received 24.05.2024
Approved after reviewing 16.06.2024
Accepted 20.07.2024

Статья поступила в редакцию 24.05.2024
Одобрена после рецензирования 16.06.2024
Принята к печати 20.07.2024