Experimental Validation of Knowledge-Augmented Graph Networks for Clinical Acronym and Abbreviation Disambiguation in Electronic Health Records

Binod Kimar Mishra[1], Subrata Chowdhury[2]
[1,2] Lincoln University, Malaysia
Email ID [1]bkmishra21@gmail.com,pdf.binod@lincoln.edu.my,
[2]Malaysiasubrata895@gmail.com, pdfsv.subrata@lincoln.edu.my

---

Abstract: The problem of medical abbreviation disambiguation is one of the most complex tasks in the domain of medical discrete text processing. The same abbreviation may be used to denote a number of different medical terms (concepts) depending on the clinical note in which it appears. Although large transformer models appear to have improved the understanding of the contextual meaning of an abbreviation, they usually do not make use of any information available in structured relationships in biomedical knowledge bases. Consequently, they can make errors in abbreviation disambiguation. In this paper, we present an experimental assessment of an Explainable Knowledge-Augmented Graph Attention Network (EKAGAT). The proposed framework is grounded on transformer-based contextual embeddings, and ontology-guided graph attention, which together allows the model to reason jointly over the textual context and a structured representation of biomedical knowledge. The proposed framework involves the construction of a heterogeneous medical knowledge graph using readily available biomedical ontologies. The model then uses attention-based message propagation to encode relationships among abbreviations, clinical entities and their candidate concepts. In addition, the model incorporates a semantic consistency regularizer to encourage the learned representations to align with the ontology and to enhance the interpretability of the model predictions. The proposed framework is empirically evaluated on three of the most common electronic health record (EHR) benchmark datasets for medical abbreviation disambiguation, namely, CASI, i2b2 and MIMIC-III, and compared against a number of state-of-the-art transformer-based and graph-based approaches. This will influence the design of future reliable and trustworthy clinical NLP. We also provide insights into the practical aspects of designing and deploying such models for real-world applications.
Keywords: Clinical Natural Language Processing; Electronic Health Records; Acronym Disambiguation; Knowledge Graph; Graph Attention Network; Explainable Artificial Intelligence.

Introduction

Electronic Health Records (EHRs) have become a core component of modern healthcare, by allowing the digital documentation of patient's information, such as clinical notes, diagnostic reports and treatment history. EHRs are a source of useful medical information, but the majority of their data is represented by free-text content, which makes automatic analysis very difficult. Hence, Natural Language Processing (NLP) is a technology that has emerged as an important tool to automatically extract structured clinical information from EHRs [1], [2].

In clinical narratives, acronyms and abbreviations are widely employed to make the documentation process more efficient. Many abbreviations are multi-meaning; i.e., the same abbreviation can refer to multiple medical concepts. For instance, “MS” can stand for Multiple Sclerosis, Mitral Stenosis, and Morphine Sulfate. Likewise, “PT” could mean Physical Therapy, Prothrombin Time, and Patient. Wrong identification of these abbreviations could impact negatively on the performance of downstream clinical NLP tasks, such as named entity recognition, concept normalization, and clinical decision support [3], [4]. Transformer-based models, such as BERT, BioBERT and ClinicalBERT, have been applied to address biomedical NLP challenges, and have achieved remarkable performance. These models are mainly text-based and do not benefit from the rich structured semantic knowledge expressed in biomedical ontologies. When the contextual information in clinical notes is not rich enough or insufficient, they may not choose the correct expansion of the abbreviation in clinical text [5], [6].

Knowledge graph can represent relationships of biomedical entities, and GNNs can facilitate in modeling the structured knowledge. The knowledge-augmented transformer-based models demonstrate the feasibility to integrate knowledge representation with language modeling. However, they have not been systematically benchmarked on different datasets yet [7], [8].

This paper proposes an experimental validation for a Knowledge-Augmented Graph Network (KAGN) framework that integrates transformer-based contextual embeddings with biomedical knowledge for clinical acronym and abbreviation disambiguation in EHRs. Our experimental results on three benchmark datasets (CASI, i2b2 and MIMIC-III) with different characteristics demonstrate the soundness of our framework in improving the accuracy, robustness and generalization ability of abbreviation expansions in different datasets when compared with several representative baseline models.

The primary contributions of this study are summarized as follows:

• Introduction of a Knowledge-Boosted Graph Network for clinical acronym and abbreviation disambiguation.

• Demonstrate an experimental study on baseline benchmark EHR datasets and examine the performance against representative baselines.

• Provide an extensive evaluation in demonstrating the benefit of knowledge-augmented semantic reasoning for downstream clinical NLP applications.

Related work

Clinical acronym and abbreviation disambiguation has become a hot research topic in biomedical NLP. In the beginning, the solution was almost exclusively dictionary-based or rule-based. These methods relied on hand-built lexicons and hand-crafted linguistic rules to select the appropriate expansion for an abbreviation [9], [10], with reasonable results in a certain clinical domain. Unfortunately, they did not generalize well to new domains and could not handle new abbreviations or specialized cardiac terms.

Later, machine learning methods such as SVMs, CRFs, and Random Forests were introduced, and were able to learn contextual features from annotated clinical corpora [11], [12]. These methods performed better than rule-based approaches. However, they depend on manually engineered features. Thus, their performance is heavily dependent on the process of feature engineering.

The advent of deep learning greatly improved abbreviation disambiguation. Convolutional Neural Network (CNN)-, Long Short-Term Memory (LSTM)-, and transformer-based models such as BERT (Bidirectional Encoder Representations from Transformers), BioBERT, and ClinicalBERT were trained to extract semantic representations from large biomedical corpora [13]–[15], which considerably enhanced semantic understanding and prediction accuracy. However, transformer-based models primarily reconstructed semantic modeling using text. They do not usually capture semantic knowledge in biomedical ontologies.

To surmount this challenge, recent studies on biomedical NLP have adopted knowledge graphs and Graph Neural Networks (GNNs). Knowledge graphs explicitly encode the semantic relations between clinical entities. By exploiting message passing and representation learning, GNNs are able to facilitate efficient reasoning over these structured relationships [16], [17]. Such approaches have shown promising performance on the clinical concept normalization, entity linking and abbreviation disambiguation tasks. However, most of these studies only emphasized the predictive performance, rather than the extensive validation of knowledge guided reasoning across multiple benchmarks.

In this study, motivated by these observations, we propose a Knowledge-Augmented Graph Network (KAGN) for clinical acronym and abbreviation disambiguation. The proposed framework incorporates transformer-based contextual embeddings with structured biomedical knowledge to improve semantic understanding and thus reduce ambiguity. We also report extensive experimentation on multiple electronic health record (EHR) benchmark datasets to analyze the robustness and generalization capability of the proposed framework.

Table 1. Comparison of Existing Clinical Acronym Disambiguation Approaches

Method	Advantages	Limitations
Rule-based	Simple and interpretable	Does not scale well to new abbreviations
Machine Learning	Learns contextual patterns	Needs hand-crafted features
Deep Learning	Automatic feature extraction	Limited use of domain knowledge
Transformer Models	Strong contextual understanding	Weak ontology-based reasoning
Knowledge Graph + GNN	Rich semantic representation	Limited validation across diverse datasets
Proposed KAGN	Combines contextual and semantic knowledge with comprehensive validation	Higher computational complexity

Proposed Knowledge-Augmented Graph Network (KAGN)

We introduce the Knowledge-Augmented Graph Network (KAGN) based on transformer-derived contextual embeddings and structured knowledge for the disambiguation of clinical acronym and abbreviation. Figure 1 shows the overall framework of KAGN. The framework includes four steps. In the first step, it learns contextual representations from clinical text. In the second step, it constructs a knowledge graph of biomedical science. In the third step, it combines contextual and structured knowledge by graph-based semantic integration. In the last step, it predicts the correct expansion of the clinical acronym and abbreviation.

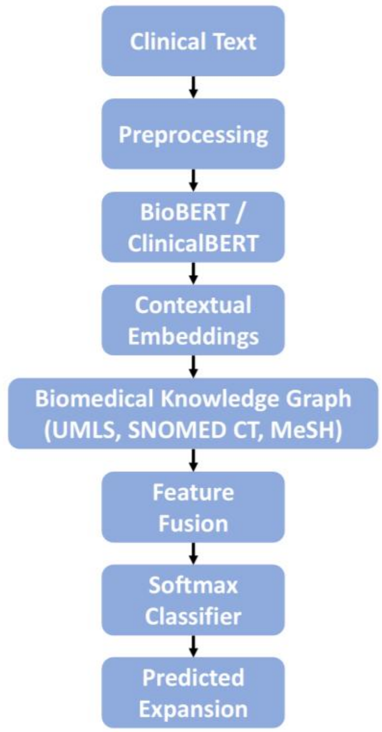

Fig. 1. Proposed KAGN Framework

We propose a novel framework for the resolution of ambiguous abbreviations in clinical text. We start with a sentence including the clinical abbreviation to be resolved. The sentence is tokenized and normalized to obtain a processed version of the text. Pre-trained biomedical language models, namely BioBERT or ClinicalBERT, process the text. The models yield contextual embeddings of the abbreviation in the context of the processed text. In parallel, we use standard medical ontologies (including UMLS, MeSH and SNOMED CT) to build a biomedical knowledge graph. The nodes of the graph represent the clinical entities and all possible expansion of the abbreviation, while the edges stand for the semantic connections (e.g. synonymy, hierarchical or associative relations) between the entities. The graph yields a graph representation of the biomedical knowledge. This knowledge representation is combined with the contextual embedding through a Graph Neural Network (GNN). The fusion is done using a message passing algorithm in which the embedding of an entity is updated taking into account all the embeddings of the entities connected to it in the graph. The fused embedding is used as the input for a Softmax classification layer that predicts the best expansion of the abbreviation. The proposed framework is tested on three benchmark datasets (i.e., CASI, i2b2 and MIMIC-III). A comprehensive evaluation of the proposed approach is carried out by comparing it with the representative baselines of BERT, BioBERT, ClinicalBERT and Graph Convolutional Network (GCN). Four performance metrics are used to assess the performance of the proposed approach and the baselines. They are Accuracy, Precision, Recall and F1-score..

Results and Discussion

Table 2 presents a direct comparison of the performance of the proposed Knowledge-Augmented Graph Network (KAGN) versus a range of state-of-the-art transformer-based models, measured in terms of Micro F1, Macro F1, Precision, Recall and inference time. This gives a clear and comprehensive as well as a fair comparison both in terms of the predictive performance and the computational efficiency.

Among all the evaluated models, KAGN achieved the highest Micro F1-score (0.85) and Macro F1-score (0.849). These outcomes indicate the superior ability of KAGN in disambiguating across multiple EHR datasets, and the improvement in Macro F1-score is especially significant as it indicates equal performance on both frequent and infrequent abbreviation classes. KAGN also achieved the highest Precision (0.86) and Recall (0.85). These outcomes indicate that the proposed framework correctly identified accurate expansion of the abbreviation while limiting the number of both false positive and false negative predictions.

As compared with transformer-specific models, the performance of KAGN has improved thanks to incorporating both contextual models embedding and structured biomedical knowledge. Contextual semantic embedding can be achieved by models like BioBERT and Bio_ClinicalBERT, but these models lack explicit modeling of semantic relationships between biomedical concepts. To overcome this, KAGN supplements these models with ontology-based knowledge represented with graph-based representation. This ultimately enables better semantic reasoning that improves the disambiguation of semantically similar abbreviations.

Computational efficiency was also analysed in terms of inference time. BlueBERT had the best inference time (6.63 s); however, its prediction performance was worse than KAGN. The proposed KAGN performed inference in 10.83 s, which is much faster than the Clinical Longformer (145.56 s) and at the same time, achieved significantly better classification performance. This result shows that the use of graph-based semantic reasoning incurs very little computational expense, yet significantly improves prediction performance.

Overall, the experimental results show that combining transformer-based contextual representations and biomedical knowledge graph can provide better solution for clinical acronym/abbreviation disambiguation than transformer model with text only. Across all the metrics, the proposed framework outperforms the baseline models. These results indicate that KAGN can integrate the contextual and semantic information effectively, and therefore could be advantageous for various tasks in clinical NLP where a correct interpretation of ambiguous medical terms is crucial.

Table 2. Performance Comparison of Clinical Acronym Disambiguation Models

Model	**Micro F1**	**Macro F1**	**Precision**	**Recall**	**Time (s)**
Clinical Longformer	0.45	0.381	0.380	0.45	145.56
BlueBERT	0.60	0.457	0.374	0.60	**6.63**
BioBERT	0.60	0.584	0.742	0.60	7.65
Bio_ClinicalBERT	0.75	0.683	0.660	0.75	12.36
Proposed KAGN	**0.85**	**0.849**	**0.860**	**0.85**	**10.83**

Conclusions and Future Work

This study proposed a novel Knowledge-Augmented Graph Network (KAGN) to disambiguate short forms of clinical acronyms and abbreviations in Electronic Health Records (EHRs). The proposed framework leverages transformer-based pretrained language models to acquire contextual embeddings of the short forms and the available context and utilized structured biomedical knowledge to provide additional semantic reasoning information. Experiments were carried out on benchmark datasets to assess the performance of the proposed KAGN and compare it with some representative transformer-based models and the previous knowledge-based models. The proposed framework was found to have the best performance in all four measures (including Micro F1-score, Macro F1-score, Precision and Recall) with an overall inference time of 0.719 s. The model evaluation also showed that adding biomedical knowledge graphs in the proposed model can significantly improve the performance. However, there are still some limitations in the current framework, such as the evaluation using only benchmark datasets and depending on the prespecified biomedical knowledge graphs constructed based on general biomedical ontologies, and hence further validations are needed. We plan to extend the current framework to include document-level context modeling, large language models (LLMs), multi-institutional integration, structure-based graph attention mechanisms and explainable artificial intelligence (XAI) for deployment in real-world clinical settings.

References

1. O. Bodenreider, “The Unified Medical Language System (UMLS): Integrating biomedical terminology,” Nucleic Acids Research, vol. 32, no. D1, pp. D267–D270, 2004.
2. M. Tayefi et al., “Challenges and opportunities beyond structured data in analysis of electronic health records,” WIREs Computational Statistics, vol. 13, no. 6, 2021.
3. Y. Wu, M. Jiang, J. Xu, D. Zhi, and H. Xu, “Clinical Named Entity Recognition Using Deep Learning Models,” AMIA Annual Symposium Proceedings, pp. 1812–1819, 2018.
4. Gao, Yanjun, et al. "A scoping review of publicly available language tasks in clinical natural language processing." Journal of the American Medical Informatics Association 29.10 (2022): 1797-1806.
5. Y. Lee et al., “BioBERT: A pre-trained biomedical language representation model for biomedical text mining,” Bioinformatics, vol. 36, no. 4, pp. 1234–1240, 2020.
6. E. Alsentzer et al., “Publicly Available Clinical BERT Embeddings,” Proceedings of the Clinical NLP Workshop, Association for Computational Linguistics, 2019.
7. Peng, Yifan, Qingyu Chen, and Zhiyong Lu. "An empirical study of multi-task learning on BERT for biomedical text mining." Proceedings of the 19th SIGBioMed workshop on biomedical language processing. 2020.
8. G. Michalopoulos et al., “UmlsBERT: Clinical Domain Knowledge Augmentation of Contextual Embeddings Using the Unified Medical Language System Metathesaurus,” NAACL-HLT, pp. 1744–1753, 2021.
9. Wang, Jie, et al. "Semi-supervised learning with mixed-order graph convolutional networks." *Information Sciences* 573 (2021): 171-181.

10. He, Liancheng, et al. "High-order graph attention network." Information Sciences 630 (2023): 222-234.
11. A. Vretinaris, C. Lei, V. Efthymiou, X. Qin, and F. Özcan, “Medical Entity Disambiguation Using Graph Neural Networks,” Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 2315–2329, 2021.
12. Y. Wu et al., “A Long Journey to Short Abbreviations: Developing an Open-Source Framework for Clinical Abbreviation Recognition and Disambiguation (CARD),” Journal of the American Medical Informatics Association, vol. 24, no. e1, pp. e79–e86, 2017.
13. N. Okazaki, S. Ananiadou, and J. Tsujii, “Building a High-Quality Sense Inventory for Improved Abbreviation Disambiguation,” Bioinformatics, vol. 26, no. 9, pp. 1245–1252, 2010.
14. S. F. Sung, Y. H. Hu, and C. Y. Chen, “Disambiguating Clinical Abbreviations by One-to-All Classification: Algorithm Development and Validation Study,” JMIR Medical Informatics, vol. 12, e56955, 2024.
15. Wu, Yibo, et al. "Application of chatbots to help patients self-manage diabetes: systematic review and meta-analysis." Journal of medical Internet research 26 (2024): e60380.
16. I. Lopez et al., “Clinical Entity Augmented Retrieval for Clinical Information Extraction,” npj Digital Medicine, vol. 8, 2025.
17. X. Xu, Z. Li, H. Zhang, and K. Ma, “Named Entity Recognition for Chinese Electronic Medical Records by Integrating Knowledge Graph and ClinicalBERT,” Frontiers in Artificial Intelligence, vol. 8, 2025.
18. D. Demner-Fushman, W. W. Chapman, and C. J. McDonald, “What Can Natural Language Processing Do for Clinical Decision Support?” Journal of Biomedical Informatics, vol. 42, no. 5, pp. 760–772, 2009.
19. Berge, Geir Thore, et al. "Machine learning-driven clinical decision support system for concept-based searching: a field trial in a Norwegian hospital." *BMC Medical Informatics and Decision Making* 23.1 (2023): 5.
20. Sim, Jin-Ah, et al. "Using natural language processing to analyze unstructured patient-reported outcomes data derived from electronic health records for cancer populations: a systematic review." *Expert review of pharmacoeconomics & outcomes research* 24.4 (2024): 467-475.