Advancing E-Commerce Search with KR-BERT and Entity Confidence

Alan Chuang
CS 274: Web Intelligence
Prof. Teng Moh
San Jose State University
alan.chuang@sjsu.edu

***Abstract*—E-commerce platforms, such as Amazon and Ebay, are an integral part of modern society. They bring the ease of browsing massive catalogs to the consumer, all while they are sitting comfortably in their home. However, e-commerce platforms are not without their shortcomings. Oftentimes, they find it difficult to accurately link products to relevant concepts due to ambiguous descriptions and noisy data. To solve this problem, this study explores Knowledge Relevance BERT (KR-BERT) to enhance language representations for improved product-concept matching. KR-BERT incorporates knowledge graph triplet embeddings and a dynamic relevance scoring mechanism, which has been shown to demonstrate significant improvements in product concept linking. Furthermore, we aim to elevate the performance of KR-BERT through the introduction of confidence scoring to improve precision, recall, and F1 scores over the baseline result.**

***Index Terms*—E-commerce, Knowledge Graphs, BERT, Triplet Embedding, Self-Attention Mechanism, , Entity Linking, Entity Confidence**

I. Introduction

Accurately linking products to relevant concepts is a significant challenge for e-commerce platforms due to ambiguous descriptions and noisy data. Traditional language models lack the structured, domain-specific knowledge needed to address these issues effectively. This study aims to enhance language representations and improve product-concept matching by enhancing Knowledge Relevance BERT (KR-BERT) with entity confidence. Thus, we integrate confidence scoring on top of knowledge graph triplet embeddings and a dynamic relevance scoring mechanism.

A. What is a Knowledge Graph?

A knowledge graph is a structured network of entities (nodes) and the relationships (edges) connecting them. It represents real-world information in a graph format, where each node typically represents a concept, and edges represent the relationships between these concepts.

B. Application in KR-BERT Model

Integrating knowledge graph data into the KR-BERT model allows BERT to leverage not only the contextual usage of words but also their relationships to real-world entities and concepts. This gives us several advantages:

- **Knowledge Graph Triplets**: In the KR-BERT model, knowledge graph triplets are used to represent relationships between entities, therefore it can understand and differentiate concepts based on their relationships. For example:
  - **[Product - categorized as - Category]**

This triplet indicates that a product belongs to a specific category, which helps the model understand and categorize product-related information more accurately. This in turn also improves how well the model generalizes to unseen data [1], [2].

II. Related Works

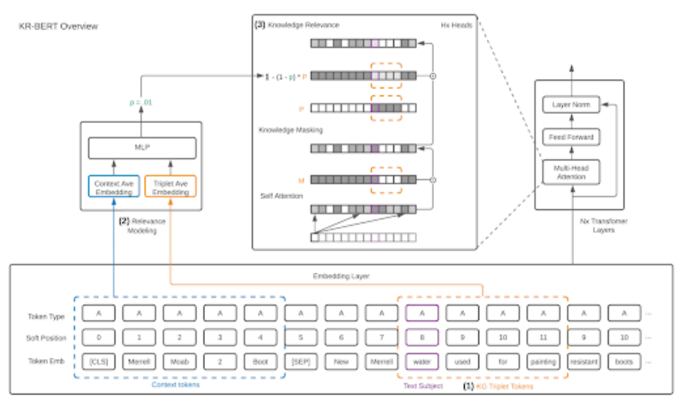

Fig. 1. The KR-BERT framework: (1) Insert triplet tokens into the text input. (2) Determine whether the inserted triplet is relevant to the context of the text. (3) The impact of a triplet's attention score is mitigated if deemed irrelevant. Reference original paper for enlarged diagram [1].

Existing approaches for integrating knowledge into language models include K-BERT and ERNIE, which fuse external knowledge into text representations. These models, however, often rely on thoroughly-processed knowledge graphs and do not address the noise inherent in newly mined domain-specific knowledge graphs [3], [4]. KR-BERT addresses this shortcoming by incorporating a relevance model to filter out irrelevant knowledge graph triplets.

III. Problem Definition

A. Context

E-commerce platforms often struggle with accurately linking products to relevant concepts due to:

- Ambiguous product descriptions by sellers
- Product data is often noisy
- Traditional language models lack structured, domain-specific knowledge

B. Objective

The objective of this research is to enhance language representations to improve product-concept matching by leveraging entity confidence scoring along with KR-BERT. This involves:

- **KG Triplet Embedding**: Integrating knowledge graph triplets to provide contextual clarity and enrich semantic representations [8].
- **Dynamic Relevance Scoring and Attention Mechanism**: Adjusting attention weights based on the reliability of knowledge graph entities to prioritize high-confidence information and reduce noise.
- **Entity Confidence**: Adjust attention weights based on the reliability of knowledge graph entities.

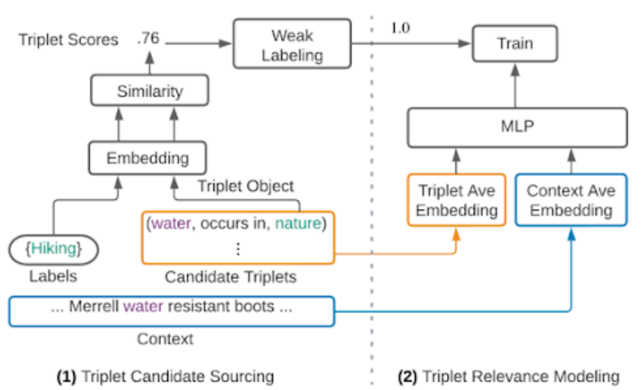

Fig. 2. Triplet sourcing is shown here, where the score of the triplet is based on its similarity with the embedding of other objects with the same label [1].

IV. Experimental Setup

In this study, we first seek to reproduce the results of KR-BERT as a baseline. We will then implement the proposed change to the KR-BERT architecture by adding a confidence scoring module to improve the final model performance.

A. Datasets

We used the Amazon Reviews dataset, specifically the All-Beauty category, which includes user reviews, item metadata, and user-item interaction graphs. The dataset was stratified by product categories and rating levels to ensure a representative sample [2].

B. Stratified Sampling

Stratified sampling is a method of sampling from a population which can be partitioned into subpopulations. In the context of the Amazon Reviews dataset, we divide the data into distinct subgroups or strata, such as different product categories and rating levels. This in turn reduces sampling bias.

V. Baseline Comparison

TABLE I
Baseline KR-BERT (10000 Samples, 5 Epochs)

Epoch	Avg Loss	Precision	Recall	F1 Score
1	0.3182	0.4851	0.3213	0.3866
2	0.1666	0.6905	0.5611	0.6191
3	0.1750	0.7097	0.6461	0.6764
4	0.1417	0.6717	0.3644	0.4725
5	0.1270	0.6967	0.6516	0.6891

TABLE II
Baseline KR-BERT (12000 Samples, 4 Epochs)

Epoch	Avg Loss	Precision	Recall	F1 Score
1	0.2882	0.4678	0.3519	0.4016
2	0.7840	0.2324	0.3602	0.2825
3	0.3068	0.4506	0.2167	0.2927
4	0.1694	0.6974	0.7018	0.6996

TABLE III
Baseline KR-BERT Results from Original Paper

Model	Relevance Model	F1	Precision	Recall
KR-BERT	yes (frozen)	.703	.673	.814
KR-BERT	yes (Lcls only)	.700	.696	.793
KR-BERT (proposed)	yes (Lcls + Lrel)	.717	.697	.826

A. Discrepancy in Performance

As we can see from the tables above, there exists a discrepancy between the performance of KR-BERT in the original paper and our implementation of KR-BERT. We can attribute this to several potential factors:

- **Sample Size**: The original paper may have used a larger and more diverse sample size, providing a more comprehensive training.
- **Architecture and Implementation**: Differences in model architecture or hyperparameter settings, as well as the methods of implementation (KG triplet generation and processing), could contribute to variations in performance.
- **Dataset**: The dataset for the original paper was not provided. Therefore, variations in the preprocessing steps and data quality could impact the final results.

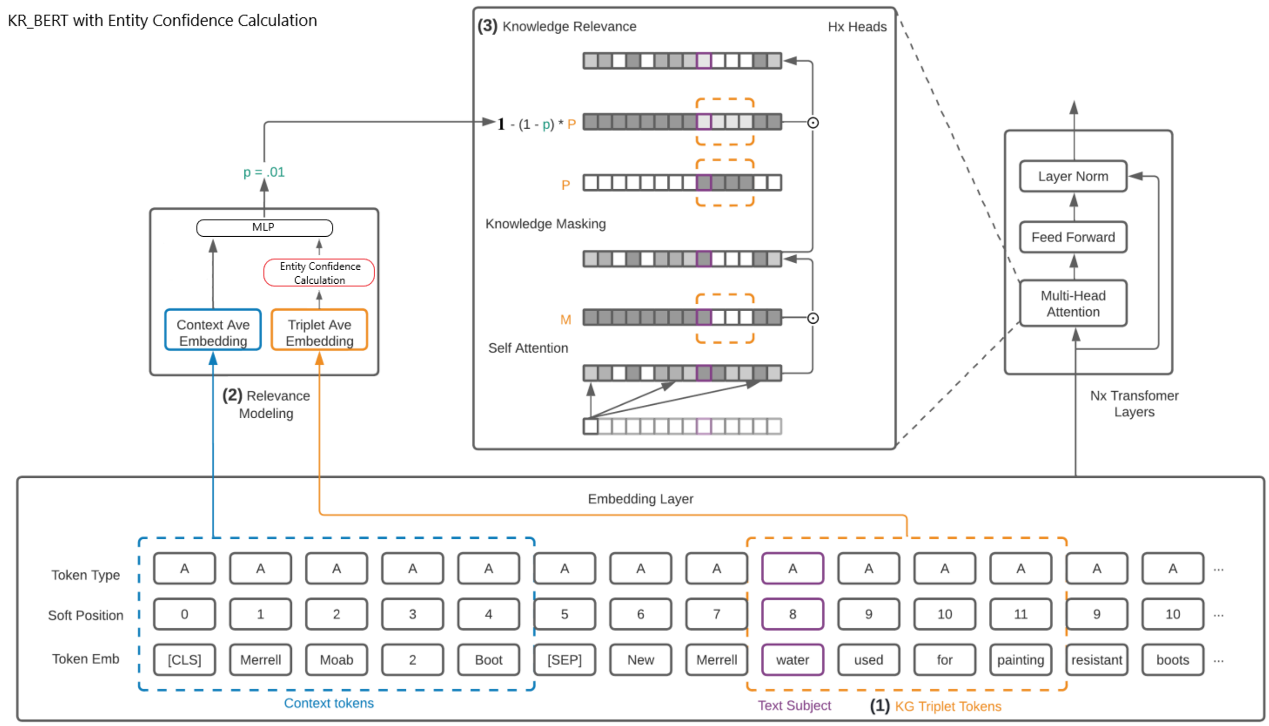

Fig. 3. Highlighted in red, the proposed implementation of entity confidence scoring would take the outputs of the triplet ave embedding as its input, seen in section (2). The entity confidence calculation module would then output its embeddings to the multilayer perceptron (MLP) before feeding to the Knowledge Relevance segment (3). [1]

VI. Proposed Changes

A. Entity Confidence

Entity Confidence adjusts attention weights based on the reliability of knowledge graph entities. The reasoning behind this approach is to prioritize more reliable or trustworthy information, thereby reducing noise in the data [12].

B. Components of Confidence Scoring

The confidence score for each entity is calculated based on three components:

- **Frequency-Based Scoring**: Entities that appear more frequently are considered more reliable.
- **Context Consistency**: This measures how consistently entities appear in similar contexts.
- **Source Reliability**: This evaluates the reliability of the entity source, such as whether the review is from a verified purchase.

The overall confidence score is calculated as:

$$\begin{aligned}\text{Confidence Score} = {} & \alpha \times \text{Frequency Score} \\ & + \beta \times \text{Context Score} \\ & + \gamma \times \text{Source Score}\end{aligned} \tag{1}$$

VII. Results

TABLE IV
KR-BERT with Confidence Scoring (10000 Samples, 5 Epochs)

Epoch	Avg Loss	Precision	Recall	F1 Score
1	0.2388	0.6071	0.2941	0.3963
2	0.2843	0.6226	0.3691	0.4634
3	0.2077	0.7149	0.5409	0.6159
4	0.1941	0.7361	0.4118	0.5281
5	0.1729	0.7195	0.6998	0.8369

TABLE V
KR-BERT with Confidence Scoring (12000 Samples, 4 Epochs)

Epoch	Avg Loss	Precision	Recall	F1 Score
1	0.2296	0.6116	0.3455	0.4416
2	0.3677	0.6036	0.4177	0.4937
3	0.2395	0.7251	0.2868	0.4110
4	0.1796	0.7221	0.8730	0.7904

A. Proposed Solution Insights

Entity Confidence and Targeted Attention

- **Frequency-Based Scoring**: KR-BERT will focus more attention on high-confidence entities, prioritizing reliable information over less relevant data, reducing overall noise. This was one of the key problems that the original

KR-BERT paper was tackling. in which they showed that a reduction in data noise led to significant performance [1].
- **Stability**: E-commerce data can vary greatly in quality and structure due to human input. Entity confidence helps maintain consistency by filtering out unreliable data points, so the model is less likely to propagate errors from the noisy data.

VIII. Analysis

TABLE VI
PERFORMANCE COMPARISON OF KR-BERT

Model	Avg Precision	Avg Recall	Avg F1
KR-BERT (Original)	0.697	0.826	0.717
KR-BERT (Baseline)	0.6971	0.6767	0.6944
KR-BERT (Confidence)	0.7208	0.7864	0.8137

The results demonstrate that KR-BERT with confidence scoring generally outperforms the version without confidence scoring, especially in terms of F1 score. The inclusion of confidence scores helps in better handling noisy data, thereby improving the model's precision and recall. The baseline KR-BERT results from the original paper showed an F1 score of 0.717, precision of 0.697, and recall of 0.826 [1]. Our model, with confidence scoring, achieved a higher F1 score, indicating better overall performance even on a limited dataset.

IX. Insights

A. Analysis of Improvements

The results demonstrate that KR-BERT with confidence scoring generally outperforms the version without confidence scoring, especially in terms of F1 score. The inclusion of confidence scores helps in better handling noisy data, thereby improving the model's precision and recall. The baseline KR-BERT results from the original paper showed an F1 score of 0.717, precision of 0.697, and recall of 0.826 [1]. Our model, with confidence scoring, achieved a higher F1 score, indicating better overall performance even on a limited dataset.

- **Enhanced Data Quality**: By incorporating entity confidence scoring, the model prioritized high-confidence entities, effectively filtering out noise and unreliable data. This led to more accurate and meaningful embeddings.
- **Better Contextual Understanding**: The dynamic relevance scoring mechanism allowed the model to adjust attention weights based on the reliability of entities, enhancing its ability to understand and disambiguate concepts within the text.
- **Stability and Consistency**: The use of confidence scoring helped maintain stability by choosing reliable entities, reducing the likelihood of propagating errors from noisy data. This consistency improved the overall robustness of the model.

B. Inhibiting Factors

We do note that our implementation of KR-BERT with confidence scoring did in fact perform worse when it came to recall compared to the original implementation of KR-BERT in the paper [1]. This could be due to:

- **Limited Sample Size**: The sample size used for training and evaluation might have been insufficient to fully capture the variability in the dataset. We also did not utilize the original dataset from the paper, which could lead to some inconsistencies during model evaluation.
- **Implementation Differences**: Differences in hyperparameter tuning could impact performance.

X. Future Work

While we successfully demonstrated the effectiveness of KR-BERT with entity confidence scoring, there are still some possible methods we could explore to improve performance in the future:

- **Increasing Sample Size**: Expanding the sample size for training and evaluation can help our model to generalize better.
- **Extended Dataset Evaluation**: Expanding the evaluation to other e-commerce categories and datasets to generalize the findings.
- **Refinement of Confidence Scoring**: Further refining the confidence scoring mechanism by incorporating additional factors such as user trustworthiness and review helpfulness scores.
- **Real-time Adaptation**: Implementing real-time adaptation of the model to continuously improve based on new data and user interactions.
- **Comparison with Other Models**: New advanced models, such as GPT-4 and T5, could be used to benchmark performance. We could also use them to improve our embedding generation.

XI. Conclusion

In this study, we proposed enhancements to the KR-BERT model for improving product-concept matching on e-commerce platforms. By integrating entity confidence scoring, we were able to improve precision and F1 scores compared to the baseline KR-BERT model. Our results demonstrate that the inclusion of confidence scoring helps in better handling noisy data and emphasizing reliable information.

The insights gained from this research highlight the potential of incorporating structured, graph-representation of knowledge to enhance the performance of language models in domain-specific applications. Future work will focus on expanding the scope of evaluation, refining the model further, and exploring real-world applications of KR-BERT in e-commerce.

References

[1] K. Samel et al., "Knowledge Relevance BERT: Integrating Noisy Knowledge into Language Representations," presented at the AAAI Conference on Artificial Intelligence, 2023. [Online]. Available: https://knowledge-nlp.github.io/aaai2023/papers/005-KRBERT-oral.pdf

[2] “McAuley-Lab/Amazon-Reviews-2023 · Datasets at Hugging Face,” Hugging Face, Mar. 31, 2024. [Online]. Available: https://huggingface.co/datasets/McAuley-Lab/Amazon-Reviews-2023
[3] S. Khalid, “BERT Explained: A Complete Guide with Theory and Tutorial,” Medium, Apr. 10, 2020. [Online]. Available: https://medium.com/@samia.khalid/bert-explained-a-complete-guide-with-theory-and-tutorial-3ac9ebc8fa7c
[4] arrrrrmin, “arrrrrmin/albert-guide,” GitHub, Dec. 15, 2023. [Online]. Available: https://github.com/arrrrrmin/albert-guide
[5] G. Singh, “Fine-Tune ERNIE 2.0 for Text Classification,” Medium, Aug. 2019. [Online]. Available: https://towardsdatascience.com/https-medium-com-gaganmanku96-fine-tune-ernie-2-0-for-text-classification-6f32bee9bf3c
[6] “How to Code BERT Using PyTorch - Tutorial With Examples,” neptune.ai, May 20, 2021. [Online]. Available: https://neptune.ai/blog/how-to-code-bert-using-pytorch-tutorial
[7] CheeKean, “Mastering BERT Model: A Complete Guide to Build it from Scratch,” Data And Beyond, Sep. 05, 2023. [Online]. Available: https://medium.com/data-and-beyond/complete-guide-to-building-bert-model-from-sratch-3e6562228891 (accessed May 23, 2024).
[8] A. Dadoun, “Knowledge Graph Embeddings 101,” Medium, May 23, 2023. [Online]. Available: https://towardsdatascience.com/knowledge-graph-embeddings-101-2cc1ca5db44f (accessed May 22, 2024).
[9] X. Ge, Y.-C. Wang, B. Wang, and C.-C. J. Kuo, “Knowledge Graph Embedding: An Overview,” arXiv.org, Sep. 21, 2023. [Online]. Available: https://arxiv.org/abs/2309.12501
[10] N. Kolitsas, O.-E. Ganea, and T. Hofmann, “End-to-end neural entity linking,” arXiv preprint arXiv:1808.07699, 2018.
[11] W. Liu, P. Zhou, Z. Zhao, Z. Wang, Q. Ju, H. Deng, and P. Wang, “K-bert: Enabling language representation with knowledge graph,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 03, 2020, pp. 2901–2908.
[12] Q. Gu, Y. Zhang, J. Cao, G. Xu, and A. Cuzzocrea, “A confidence-based entity resolution approach with incomplete information,” in *2014 International Conference on Data Science and Advanced Analytics (DSAA)*, 2014, pp. 97-103. doi: 10.1109/DSAA.2014.7058058.