Implicit Emotion Classification using BERTs with Knowledge Incorporation

Seohee Yoon, Nabin Kim, Feyi Ogunsanya, and Jimin Park

1 Introduction

Emotion classification in NLP is challenging due to the context-sensitive and complex nature of emotional expressions, particularly subtle ones (Seyeditabari, 2018). While most research focuses on detecting explicit emotions, identifying implicit emotions remains difficult in real-world applications. To address this limitation, this project introduces a methodology that integrates external knowledge bases to enhance language understanding. Knowledge bases, which store factual information relevant to NLP tasks, provide a broader context for understanding human emotions. These knowledge bases are typically built through manual effort and accessed via queries and keywords to retrieve relevant information (Balogh, 2020). Our approach aims to improve contextual understanding and model accuracy for implicit sentiment tasks by using knowledge bases.

Therefore, we propose a novel approach to improving implicit emotion classification by enhancing K-BERT-based models with structured knowledge graph integration. Specifically, we incorporate SenticNet (Cambria et al., 2014) to provide sentiment-specific knowledge, such as primary emotion, polarity values, and related semantic concepts. To reduce noise and maintain contextual integrity, we implement selective knowledge injection, ensuring only the most relevant SenticNet components are integrated into the model. Additionally, we refine the knowledge integration process using soft-position embeddings to preserve the original sentence structure and a visible matrix to control token interactions during the attention process (Liu, 2020). These techniques collectively enhance the contextual understanding and accuracy of implicit emotion classification models.

KnowBERT is a BERT-based transfer model that enriches contextual word representation by embedding a knowledge base, and provides the ability for the model to incorporate external knowledge based on knowledge attention and recontextualization (KAR) (Peters, 2019). It models entity coverage and forms entity embeddings by identifying entity coverage through entity links and retrieving information about those entities from the knowledge base. KnowBERT learns entity links from unlabeled data through self-supervision, allowing it to dynamically retrieve and integrate the latest published knowledge (Peters, 2019).

K-BERT evaluates knowledge graphs for models to achieve better text understanding. Unlike KnowBERT, K-BERT focuses on achieving domain-specific knowledge and loads a pre-trained model (BERT) and assigns triples from the knowledge graph in the knowledge hierarchy directly to the input text (Liu, 2020). This mechanism adds knowledge to the original sentence while still providing significant improvement over BERTs on domain-specific tasks.

This research advances implicit emotion classification by optimizing the integration of structured knowledge graphs into transformer models. Using SenticNet for sentiment-specific knowledge and employing selective knowledge injection, we reduced noise and enhanced contextual understanding (Cambria et al., 2014). By incorporating key techniques from K-BERT (Liu, 2020), we preserved sentence structure and improved token interactions. These innovations resulted in significant performance gains, demonstrating the value of domain-specific knowledge for sentiment tasks. Our approach enhances the reliability of knowledge-based models, broadening the potential for sentiment classification in NLP.

2 Related Work

2.1 BERT

As a transformer-based model, BERT excels at many general NLP tasks due to its tokenization process and attention mechanism (Devlin, 2019), which captures bidirectional context. However, when applied to implicit emotion detection, WordPiece tokenization dilutes emotional meaning by fragmenting emotionally charged words like "heartbroken" into separate tokens for "heart" and "broken." This fragmentation weakens BERT's ability to detect subtle emotions.

Because BERT relies on token-level attention mechanisms, it is limited in its ability to understand implicit emotions, which rely on sentence-level context. The lack of integration with external knowledge bases exacerbates this problem because implicit emotions require a broader understanding beyond training data.

2.2 T5

Raffel introduced T5, which treats all NLP tasks as text-to-text problems where both input and output are in the form of text with a specific prefix (Raffel, 2020). While T5 performs well on a variety of tasks, its effectiveness in implicit sentiment classification, which relies on external sentiment knowledge, is unclear, so we experimented with a pre-trained T5 model on HuggingFace and tested different prefixes. To fine-tune T5, we loaded the pre-trained T5 model, tokenized the dataset by adding task-specific prefixes, then tuned key hyperparameters such as learning rate, epoch, batch size, and maximum sequence length, and used T5's performance as a secondary criterion.

2.3 Know-BERT

Peters et al. propose KnowBERT (Peters, 2019), a method for integrating knowledge bases such as WordNet or Wikipedia into BERT to enrich word representations. However, this model becomes expensive when applied to large datasets due to its high computational complexity, and our project aims to address this limitation by integrating domain-specific knowledge bases tailored for sentiment datasets. KnowBERT is a model that combines the BERT model with a knowledge attention and recontextualization (KAR) mechanism to increase its efficiency, where KAR integrates structured knowledge by retrieving and linking candidate entities from the knowledge base. This entity embedding enhances word representation and improves performance on tasks such as named entity recognition and implicit sentiment classification.

In this project, we focused on implicit sentiment classification, which optimized KnowBERT through task-specific knowledge integration and addressed issues related to entity association accuracy and processing speed for large datasets.

2.4 K-BERT

K-BERT aims to avoid two challenges in knowledge integration: heterogeneous embedding spaces and knowledge noise (Liu, 2020).

The input to the model is a knowledge graph consisting of a collection of sentences and triples, and the architecture consists of four modules. In the knowledge layer, the knowledge graph embeds triples into sentences but is limited by the fixed depth of the branches at 1. The embedding layer receives the sentence tree and converts it into an embedding representation, the viewing layer receives the sentence tree and converts it into a visible matrix, and the mask transformer receives the visible matrix and embeddings and applies a mask self-attention mechanism. While the researchers state that the knowledge graph does not have a significant impact on sentiment analysis because "it is possible to determine the sentiment of emotional word-based sentences even without knowledge," the exclusion of implicit sentiment was a limitation for our project.

Both Know-BERT and K-BERT aim to use two different knowledge graphs: ConceptNet and SenticNet. Since EmotiNet stores "4-tuples" (actor, action, object, emotion) to represent action chains, it is incompatible with K-BERT, which relies on 3-tuple injection (Balahur et al., 2011), limiting the knowledge layer to only 3-tuples in the knowledge graph.

3 Proposed Methodology

We planned to experiment with four pre-trained models (BERT, KnowBERT, K-BERT, T5) using four datasets (ISEAR (Scherer and Wallbott, 1994), GoEmotion (Demszky et al., 2020), SemEval-2007 Task 14 (Strapparava and Mihalcea, 2007), Alm's Dataset (Herrmann and Lüdtke, 2023)) to determine which model performs best for implicit emotion detection tasks. We aim to understand under what conditions, such as hyperparameters or dataset size, each model yields the best results. For knowledge-enhanced models like KnowBERT, we will ensure the proper integration of two knowledge bases (ConceptNet (Cambria, 2009) and SenticNet (Cambria et al., 2014)) to enhance each model's contextual understanding.

3.1 Pre-Training

Both the Know-BERT (Peters, 2019) and K-BERT (Liu, 2020) will need to undergo pre-training from baseline BERT (Devlin, 2019), to learn the knowledge from our three knowledge graphs. We will experiment with altering the number and types (domain-general vs domain-agnostic) of knowledge graphs used in both types of models to determine which maximizes accuracy. To pretrain, we will prepare each knowledge base used by computing entity embeddings and creating a candidate generator for the entity linkers. Then we will pre-train the entity linker and fine-tune all parameters.

3.2 Fine-tuning

Fine-tuning each pre-trained model requires carefully adjusting specific hyperparameters to optimize performance for the task of implicit emotion detection. Being that all four models are transformer-based, hyperparameters like learning rate, batch size, dropout, and number of epochs will be tuned. Weight decay, where the learning rate is gradually increased at the beginning of training, helps stabilize training by preventing sudden, drastic changes in the model weights (Devlin, 2019). We will be fine-tuning hyperparameters to avoid overwhelming the contextual information, enhance generalization, and improve overall performance.

3.3 Tokenizer

RoBERTa's robust pretraining on a dataset 10 times larger than BERT (160GB vs. 16GB) and its use of dynamic masking improves generalization by applying different masks for each epoch, enhancing its robustness (Liu et al., 2020). Additionally, its Byte-Pair Encoding (BPE) tokenizer handles word variations and subwords more effectively than BERT's WordPiece tokenizer. These improvements make RoBERTa particularly well-suited for implicit emotion detection, where understanding subtle and nuanced language is essential.

4 Datasets and Knowledge Graphs

4.1 Knowledge Graphs

For both K-BERT and Know-BERT, the intended format of the knowledge graphs was supposed to be subject-predicate-object (SPO) format; thus, we converted them to maintain consistency. We investigated the use of five knowledge graphs: ConceptNet (Cambria, 2009), AffectiveSpace (Cambria, 2009), EmotiNet (Balahur et al., 2011), SenticNet (Cambria et al., 2014), and WordNet-Affect (Strapparava and Valitutti, 2004). We used ConceptNet and SenticNet due to both being large-scale graphs with diverse relationships between concepts, with ConceptNet being domain-general and SenticNet using commonsense knowledge of affective situations. The others struggled either from requiring manual construction (WordNet-Affect), not being publicly available (EmotiNet), or having only numerical embeddings for each concept (AffectiveSpace) and thus not being suitable for tasks requiring SPO format.

SenticNet is not in SPO format, but rather has nine properties (Introspection Value, Temper Value, Attitude Value, Sensitivity Value, Primary Emotion, Secondary Emotion, Polarity Label, Polarity Value, and Semantics) for each concept.

4.2 Datasets

We use two emotion classification datasets with various labels and contexts to train our models for better understanding their performance on implicit emotion detection across different types of emotional content.

The first dataset, the International Survey On Emotion Antecedents And Reactions (ISEAR), was originally used for implicit emotion detection. It includes approximately 7,600 records from around 3,000 respondents worldwide across various languages, covering seven emotions (Scherer and Wallbott, 1994). We have 5,500 entries in total and about 830 entries for each of the seven labels.

The second dataset, GoEmotion, consists of 58,000 Reddit comments annotated with 26 extensive emotion categories. This dataset is highly imbalanced; for instance, the neutral label has more than 58,000 entries, while grief has only 783 entries. Hence, we decided to choose subsets of the dataset with a relatively similar number of data entries. To better select the subsets of the emotion labels, we referred to the emotion label's official documentation in larger categories: positive, negative, and ambiguous. In sum, we selected 15 emotion labels with seven positive, five negative, and three ambiguous categories of emotion, ranging from 6,000 to 17,000 entries. The diversity of emotion labels and the size of the GoEmotion dataset allow us to test how well each model captures subtle emotional nuances (Demszky et al., 2020).

4.3 Data Preprocessing

The datasets require preprocessing to match the models' input format. This involves tokenization to break text into tokens, making it manageable for the models. For KnowBERT and K-BERT, entity linking enriches the text by connecting entities to knowledge bases (Peters, 2019). Special tokens like [CLS] and [SEP] for BERT or task-specific prefixes for T5 help the models understand the input structure (Devlin, 2019; Liu et al., 2020). Padding and truncation ensure uniform sequence lengths for efficient batch processing. Label encoding converts emotion labels into numerical formats for training. Additionally, knowledge graphs are converted to SPO format.

For KnowBERT inference, datasets and knowledge graphs are loaded, tokenized, enriched with knowledge, and batched to generate contextual embeddings for classification tasks (Peters, 2019).

5 Experiments and Findings

We aimed to improve our previous model's performance by adopting RoBERTa, a more optimized model, for faster and more efficient accuracy gains. We found that the K-BERT implementation from Progress Report 2 performed worse than the base BERT model. To address this, we identified the following key issues and implemented solutions to address them:

1. **Excessive SenticNet Knowledge Graph Injection:** Injecting all 7 components of the SenticNet as the input for every sentence caused increased noise, which hindered the model performance.

   Solution: Selective Knowledge Injection

   In the original implementation, all SenticNet components (Primary emotion, Secondary emotion, Polarity value, Polarity intensity, and semantics) were injected into every input sentence. Instead of using every component, we selected three components most relevant to sentiment classification tasks:

   - Primary emotion - primary emotion associated with the concept
   - Polarity value - numeric score for the polarity strength of the concept
   - Top 3 related concepts from semantics - 3 Most semantically relevant concepts to main concept

   This approach ensured that the model trained on more precise and task-relevant information. The primary emotion and polarity value directly align with sentiment-related tasks, providing essential emotional and polarity context. Limiting semantics to the top 3 concepts reduced noise and avoided overwhelming the model with redundant information.

2. **Improper Knowledge Graph Injection Method:** The existing code integrated SenticNet knowledge as plain text tokens, causing over-tokenization, truncation, or dilution of the original sentence context.

   Solution:
   Inspired by the K-BERT paper (Liu, 2020), we refined the knowledge graph injection method by adopting a "sentence tree" structure. This approach hierarchically organizes external knowledge to align with the syntactic structure of the input sentence, preserving token-level semantics. We incorporated SenticNet data (concepts, emotional context, polarity values) for each word, treating words as nodes with related concepts as sub-nodes. Words without relevant knowledge remained independent nodes. This design ensured that knowledge was contextually integrated, enhancing the model's understanding while maintaining sentence structure.

3. **Lack of Noise Handling and Sentence Structure Preservation**: The absence of mechanisms to handle noise and preserve the original sentence structure led to distorted sentence semantics and increased noise, ultimately reducing the model's accuracy.

   Solution:Noise Handling and Sentence Structure Preservation
   While improving the knowledge graph injection process, we faced challenges in handling noise generated by the injection. Since the original code lacked noise-handling mechanisms, we implemented two techniques proposed in the paper to enhance model performance:

   - **Soft-Position Embedding**: Instead of using hard positions for all tokens, we maintained the original sentence structure while preserving the relative position of the injected knowledge. By introducing a soft-position tensor, the enhanced model naturally aligned the injected knowledge with the sentence, mitigating positional mismatches and preserving sentence integrity.

- **Visible Matrix**: This controls token interactions during the attention process by defining which tokens can attend to others, ensuring only relevant relationships form. When knowledge is injected, the visible matrix restricts the injected tokens to attend only to related sentence tokens, preventing interactions with unrelated parts. This reduces noise, avoids irrelevant connections, and helps the model focus on meaningful interactions, improving predictive performance.

5.1 Evaluating model performance

Result:

Model	Accuracy
K-BERT with ConceptNet	0.644
K-BERT with SenticNet	0.845
K-BERT with SenticNet, improvements	0.881

Table 1: Accuracy Comparison of Various K-BERT

With these improvements, we could find that the accuracy of K-BERT increased comparing to our previous models.

6 Comparison with other SOTA

The K-BERT framework, as presented in the paper (Liu, 2020), integrates external knowledge into BERT by leveraging the sentence tree structure and mechanisms such as visible matrics. It utilizes ConceptNet as its knowledge graph to inject commonsense knowledge. Our implementation enhances this approach by:

1. Using **Roberta** instead of BERT as the baseline
   - Better: Improved contextual learning and faster convergence using optimized baseline model
   - Worse: RoBERTa requires more memory and computational power compared to BERT
2. Integerating **SenticNet** instead of Conceptnet
   - Better: Enabled model to focus on sentiment-related tasks which is our main goal
   - Worse: While SenticNet excels at sentiment tasks, it may not generalize as well as ConceptNet for broader NLP applications.
3. Selective injecting information from knowledge graph instead of using every component
   - Better: Reduced noise and equipped model with concise and domain-specific knowledge.
   - Worse: Potential loss of auxiliary information that could have enhanced the model's understanding in certain scenarios or tasks.
4. Adapting the Tokenizer and Dataset for English
   - Better: Unlike the original K-BERT, which uses a Chinese-focused tokenizer (pkuseg) and Chinese datasets, we adopted RoBERTa's Byte-Pair Encoding (BPE) tokenizer optimized for English. This improves the handling of English subwords and emotionally nuanced terms like "heartbroken" or "overwhelmed."
   - Worse: This required additional adaptation and pre-training, which increased implementation complexity.

Our implementation of K-BERT effectively adapted the framework for implicit emotion classification, achieving an accuracy of 88.0% on the implicit emotion dataset, close to the SOTA result of 88.8% on the Book Review Dataset. By leveraging SenticNet for sentiment-specific knowledge, our selective knowledge injection approach reduced noise and sharpened the model's focus on relevant emotional cues. Although broader commonsense knowledge was not fully utilized, our results underscore the effectiveness of domain-specific knowledge integration for improving implicit emotion classification.

7 Ablation Study

We performed ablation experiments by removing 2 methodologies for reducing noise and analyzing the resulting changes in performance.

Components Evaluated

- Soft-Position Embedding: Without this, the model used hard positions for all tokens, disrupting the alignment between injected knowledge and the original sentence structure.
- Visible Matrix: Removing the matrix allowed all tokens to interact freely, leading to noise and irrelevant connections.

Component Removed	Accuracy
None (Full Model)	0.881
Soft-Position Embedding	0.873
Visible Matrix	0.854

Table 2: Accuracy Comparison of Ablation Study

8 Conclusion

This research improves the K-BERT framework for implicit emotion classification by integrating SenticNet for sentiment-specific knowledge and adopting RoBERTa's BPE tokenizer to better handle English-language tasks. By applying targeted knowledge injection, we reduced noise by focusing on essential concepts. Additionally, techniques like soft-position embeddings and the visible matrix preserved sentence structure and optimized token interactions. These enhancements boosted accuracy to 88.1%, highlighting the effectiveness of domain-specific knowledge and advanced tokenization for sentiment analysis. Our approach advances transformer-based models enriched with external knowledge, expanding possibilities for NLP applications.

9 Limitations

K-BERT's use of soft-position embeddings can limit self-attention's ability to fully capture dynamic contextual relationships. Performance may benefit from targeted knowledge integration or attention-based models that focus on relevant information and minimize contextual noise. Targeted knowledge integration involves injecting knowledge into only contextually significant words, reducing the risk of overwhelming the model with irrelevant information and enhancing its ability to maintain coherence and accuracy (Liu, 2020).

KnowBERT demonstrates that combining multiple knowledge sources enhances performance (Peters, 2019). Integrating SenticNet (Cambria et al., 2014) for sentiment and ConceptNet (Cambria, 2009) for broader semantics could improve implicit emotion detection by capturing both affective and contextual nuances.

ConceptNet poses challenges for implicit emotion detection due to its ad-hoc weighting, where weights rely on subjective heuristics rather than systematic methods (Cambria, 2009). For example, an assertion from a single user may get a weight of 1.0, while a WordNet edge receives 2.0, reducing reliability. Accurate weights could be derived using metadata and confidence levels. Moreover, the dominance of the "RelatedTo" relation (58% of entries) overshadows more emotionally relevant relations like "Cause" and "Antonym" (Cambria, 2009). This ambiguity risks introducing noise into K-BERT by adding factually correct but contextually irrelevant associations. Systematic weighting and selective filtering can improve ConceptNet's reliability for implicit emotion detection.

On the basis of limited time and challenges associated with versioning and dependencies, our results were restricted to K-BERT models and we did not obtain results for KnowBERT.

10 Contribution Table

Member	Contribution
Nabin Kim	Preprocessed knowledge bases and datasets, analyzed K-BERT's lower performance by reviewing the paper and GitHub repo, conducted K-BERT experiments, implemented and identified issues with KnowBERT's implementation.
Seohee Yoon	Data Preprocessing, Implemented K-BERT, Integrating mechanisms (sentence tree, visible matrix, soft positions) into K-BERT based on paper, Experiments K-BERT, Ablation Study
Feyi	Implemented KnowBert codebase, performed K-BERT experiments on GoEmotion dataset and trained and ran inferences on BERT and RoBERTa baseline models.
Jimin Park	Implemented KnowBert codebase, performed K-BERT experiments on GoEmotion dataset and implemented.

Table 3: Contribution Table

References

Alexandra Balahur, José Manuel Hermida, and Andrés Montoyo. 2011. Detecting emotions in social affective situations using the emotinet knowledge base. In *International Conference on Affective Computing and Intelligent Interaction*, pages 536–545. Springer.

Berend G. Diochnos D. I. Turán G. Balogh, V. 2020. Understanding the semantic content of sparse word embeddings using a commonsense knowledge base. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Erik Cambria, David Olsher, and Dipankar Rajagopal. 2014. Senticnet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1):1515–1521.

Hussain A. Havasi C. Eckl C Cambria, E. 2009. Affectivespace: Blending common sense and affective knowledge to perform emotive reasoning. *OMSA at CAEPIA, Seville*.

Dorottya Demszky, Julian Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *ArXiv*.

Chang M. W. Lee K. Toutanova K. Devlin, J. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*.

Björn Herrmann and Jana Lüdtke. 2023. A fairy tale gold standard: Annotation and analysis of emotions in the children's and household tales by the brothers grimm. *Zeitschrift für digitale Geisteswissenschaften*, 8.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Ro{bert}a: A robustly optimized {bert} pretraining approach.

Zhou P. Zhao Z. Wang Z. Ju Q. Deng H. Wang P. Liu, W. 2020. K-bert: Enabling language representation with knowledge graph. *Proceedings of the AAAI Conference on Artificial Intelligence*.

Neumann M. Logan R. Schwartz R. Joshi V. Singh S. Smith N. A. Peters, M. E. 2019. Knowledge enhanced contextual word representations. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Shazeer N. Roberts A. Lee K. Narang S. Matena M. ... Liu P. J. Raffel, C. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*.

Klaus R Scherer and Harald G Wallbott. 1994. Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, 66(2):310–328.

Tabari N. Zadrozny W. Seyeditabari, A. 2018. Emotion detection in text: a review. *ArXiv*.

Carlo Strapparava and Rada Mihalcea. 2007. Semeval-2007 task 14: Affective text. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 70–74, Prague, Czech Republic. Association for Computational Linguistics.

Carlo Strapparava and Alessandro Valitutti. 2004. Wordnet-affect: An affective extension of wordnet. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, pages 1083–1086. European Language Resources Association (ELRA).