International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Suggestion Key Phrase Extraction: A Fine-Grained Suggestion Mining from Opinion Reviews

Naveen Kumar Laskari*[1], Suresh Kumar Sanampudi[2]

Submitted: 08/05/2023 **Revised**: 15/07/2023 **Accepted**: 04/08/2023

Abstract*:* The abundance of online opinion reviews provides a valuable insight for understanding customer references and improving products and services. However, extracting fine-grained suggestions from these reviews remains a challenging task. In this paper, we propose a novel token classification model based on Transformer architecture for suggestion key phrase extraction. We formulate suggestion key phrase extraction as a sequence labeling task and fine-tune a pre-trained Transformer model. By applying token classification, the model is trained to assign a label to each token in the review indicating whether it represents a suggestion key phrase or not. This fine-tuning process enables the model to learn the intricate relationships between words and their contextual cues, facilitating the effective identification of relevant suggestions. To evaluate the effectiveness of the Token classification approach, we constructed a dataset consisting of opinion reviews from various domains. We annotated the dataset with suggestion key phrases to serve as ground truth for training and evaluation. We employed state-of-the-art token classification models, such as BERT or DeBERTa of various sizes, and fine-tuned them on our annotated dataset. Experimental results demonstrate that the Token classification approach outperforms traditional methods for suggestion key phrase extraction. The findings of this research have several practical implications for businesses and organizations. By automatically extracting suggestion key phrases from opinion reviews, companies can gain valuable insights into customer preferences and expectations. These insights can be used to enhance product development, improve customer satisfaction, and optimize marketing strategies.

Keywords: *Token classification, Suggestion Mining, Key Phrase Extraction, Sentiment Analysis, and Natural Language Processing.*

1. Introduction

With the widespread availability of user-generated content on various online platforms, sentiment analysis and opinion mining have emerged as crucial research areas for understanding customer preferences, improving user experiences, and driving business success. Extracting valuable insights from opinion reviews, such as sentiment polarity and suggestion key phrases, has become essential for organizations aiming to gain a competitive edge in today's highly dynamic market[1]–[3]. Sentiment analysis focuses on determining the sentiment expressed in a piece of text, whether it is positive, negative, or neutral. It has traditionally relied on approaches such as lexicon-based analysis, machine learning classifiers, and deep learning models to classify sentiment in opinion reviews. However, sentiment analysis alone may not provide sufficient actionable information for businesses seeking to enhance their products or services.

In recent years, suggestion mining has gained attention as a complementary task to sentiment analysis[2]–[5]. Suggestion mining aims to identify specific suggestions or recommendations within opinion reviews, providing actionable insights for businesses. Extracting suggestion key phrases allows organizations to understand not only the sentiment of customers but also their explicit recommendations for improvements or enhancements.

In this research paper, we propose a novel Token classification approach for extracting suggestion key phrases from opinion reviews. Token classification, a subtask of natural language processing (NLP), involves assigning predefined labels to individual tokens in a text sequence. By employing token classification, we can identify and classify tokens within an opinion review as suggestion-related, sentiment-related, or general opinion related.

The primary motivation for adopting a Token classification approach lies in its ability to capture fine-grained information from opinion reviews[6]. By classifying individual tokens, we can pinpoint specific phrases or words that explicitly express suggestions, enabling a more targeted and accurate extraction process. This level of granularity surpasses traditional approaches that treat the entire review as a single unit, potentially missing out on important suggestion key phrases embedded within the text.

While existing research has explored sentiment analysis and suggestion mining separately. The suggestion mining as considered as sequence classification task, classifying as suggestion or non-suggestion classes[2], [4]. A fine-grained

1 Xelpmoc Design and Tech Ltd, Hyderabad, Telangana, India.
ORCID ID : 0000-0003-4308-5158
2 JNT University College of Engineering Jagtial, Telangana, India.
ORCID ID : 0009-0001-3862-3153
** Corresponding Author Email: naveen.laskari@gmail.com*

analysis is missing in the existing literature. The proposed Token classification approach bridges this gap by leveraging the power of NLP techniques, machine learning models, and token-level classification to extract suggestion key phrases.

The contributions of this research paper can be summarized as follows:

1. Introducing a Token classification approach for extracting suggestion key phrases from opinion reviews, enabling fine-grained analysis of customer feedback.
2. Constructing a comprehensive dataset of opinion reviews, annotated with suggestion key phrases, to facilitate training and evaluation of the Token classification models.
3. Conducting experiments and evaluations using state-of-the-art token classification models, such as BERT[6] or DeBERTa[7] and various sizes of the same to demonstrate the effectiveness and superiority of the proposed approach over traditional methods.
4. Providing insights into the performance and limitations of different token classification models for suggestion key phrase extraction, shedding light on their suitability for real-world applications.
5. Offering practical implications for businesses and organizations by showcasing how the automatic extraction of suggestion key phrases can drive product development, customer satisfaction, and marketing strategies.

In the subsequent sections of this paper, we will present a detailed overview of related work in suggestion mining, and token classification. We will then describe our proposed methodology, including the dataset construction, model training, and evaluation. Finally, we will present the experimental results, analyze the findings, and discuss the implications and future directions of this research.

2. Literature Survey

Suggestion mining from opinion reviews is a relatively new research area within the field of sentiment analysis and opinion mining[8]. While several studies have focused on sentiment analysis and aspect-based sentiment analysis[9], the specific task of extracting suggestions has received limited attention. This literature study aims to provide an overview of the existing approaches related to suggestion mining and token classification in the context of opinion review analysis.

2.1. Suggestion Mining

Viswanathan, Amar, et al.[10] were among the pioneers in introducing the concept of suggestion mining to the literature. They extracted insights from www.mouthshut.com reviews using rule-based approaches, aiming to identify actionable feedback and treat them as suggestions. Linguistic features, n-grams, and POS tag data were utilized to identify suggestion expression sentences in customer opinions from platforms like Trip Advisor and Yelp. However, until 2015, a precise definition of suggestion mining was lacking in the literature. In subsequent years, Negi, S. et al[1]. compiled annotated datasets and problem definitions by gathering information from various sources such as restaurant and electronic product reviews, Microsoft Windows phone tweets, and software forum conversations. Additional reviews were acquired from Twitter and travel-related portals to enhance the available datasets.[1], [2], [4], [11] Rule-based systems and deep learning methods, including LSTM and CNN, were employed to classify reviews as suggestions or non-suggestions. Deep models were initialized with Word2Vec[12], [13] and Glove[14] word embeddings, and LSTM was found to perform better. A hybrid system was developed to detect review sentences conveying suggestion intent, and a semi-supervised learning method was introduced to extract customer-to-customer suggestions from reviews.

To attract more research attention and promote further study in suggestion mining, Sapna Negi et al. organized a pilot task on Suggestion Mining as part of SemEval-2019[5]. Labeled data from feedback forums and hotel reviews were created for this task, which comprised two subtasks: open-domain and cross-domain suggestion classification. Various teams participated in SemEval-2019, employing pre-trained models and transfer learning approaches to tackle the subtasks[15]–[24]. In similar lines [25]reviewed the existing literature of work related to suggestion mining. The authors elaborated on the approaches such as rule-based system, machine learning, and deep learning used for suggestion mining. With the advantage of computing resources and data, the trend of deep learning algorithms for various applications has evolved very rapid pace. In majority applications deep learning algorithms are seems like a black box[26] gave an attempt visualize the model attention weights for the suggestion mining. The major challenge with the available data for suggestion mining task is class imbalance, which biased towards non-suggestion mining. To address the class imbalance and understand the effectiveness of various embedding techniques and models[27] proposed a novel weighted focal loss function.

Jain et al.[28] proposed a transformer-based approach for suggestion mining and employed Synthetic Minority Oversampling Technique (SMOTE) and Language Model-based Oversampling Technique (LMOTE) to address the class imbalance problem. These oversampling techniques marginally improved model performance. Leekha. M, Goswami, et al[28]. implemented a multi-task learning approach combined with oversampling methods for suggestion mining. They utilized an ensemble of RCNN, CNN, and Bi-LSTM models with ELMo embeddings in their multi-task learning setup. Researchers participating in SemEval-2019 have also explored other dimensions of suggestion mining.

The fine-grained analysis on suggestion mining is missing, from the pioneering approaches to the current all the approaches are considered suggestion mining as sequence classification[3], [29]. To make a more fine-grained analysis, extract the impactful insights from raw data a novel task such as aspect-oriented suggestion mining has been proposed[29]. In the aspect orientation of suggestion mining, the authors annotated the data and applied various categories of models to evaluate the performance. To address the other missing piece, the authors proposed key phrase extraction from opinion reviews, which exactly extract the suggestion intended phrase from opinion reviews.

2.2. Token Classification

Token classification, also known as sequence labeling or named entity recognition, is a fundamental task in natural language processing (NLP) that involves assigning labels to individual tokens in a sequence of text. Over the years, token classification has garnered significant attention and has been widely applied in various NLP domains, including sentiment analysis, named entity recognition, part-of-speech tagging, and information extraction. Early approaches to token classification often relied on handcrafted features and traditional machine learning algorithms such as Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs). These methods, while effective to some extent, struggled to capture the intricate relationships and contextual dependencies present in natural language text.

The advent of Transformer-based models, such as BERT[6] and other encoder only models such as RoBERTa, DeBERTa[7], and many more are revolutionized for token classification. These models employ attention mechanisms and deep neural networks to capture the contextual information and dependencies between tokens, enabling more accurate and robust sequence labeling. Pre-trained Transformer models[30], in particular, have shown remarkable performance by leveraging large-scale corpora and unsupervised learning to learn rich representations of words and their contextual cues.

3. Methodology

We present a novel task in suggestion mining such as suggestion key phrase extraction, which aims to extract the key phrase that denote the suggestion in the given opinionated text. In the process of key phrase extraction, the models need labeled examples to get trained. After preparation of labeled data, need to be tokenized and convert into numerical form to feed into any machine learning model to learn. The generated embeddings are fed to the model and adjust the model parameters and use the trained model for inference. The following are the more detailed representation of each step in the methodology we adopted.

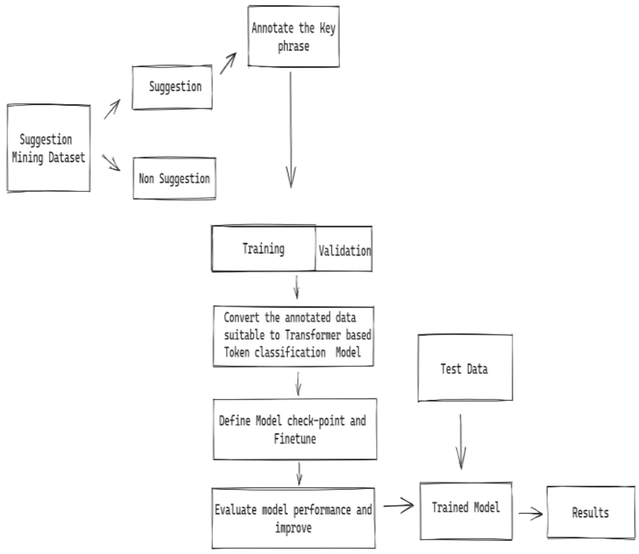

Fig 1. Methodology adopted for Key phrase extraction

The above image depicts the methodology adopted to tackle the suggestion key phrase extraction using token classification approach by finetuning the BERT and DeBERTa. The more details of each of the step are explained below.

3.1. Dataset Preparation and Tokenization

SemEval-2019 organizers[5] provided the labeled dataset for suggestion mining tasks for the two tasks. The dataset consisting of 8500 number of reviews on Microsoft windows phone, and 6185 reviews on Travel.

Table 1. Details of Datasets

Dataset	Suggestion	Non-Suggestion
Travel Reviews	2310	3875
MS Windows Phone	2085	6415

From the above set of reviews, we considered only suggestion labeled as True, which are 2310 from travel domain and 2085 from MS Windows phone reviews.

The dataset has label “1” or “0” indicating whether the opinion review is suggestion intended or not. But the key phrase annotation was not available. We Annotated the dataset with suggestion key phrases manually each opinion review and identify the suggestion key phrases within the text. We labeled these key phrases by labeling them as suggestion-related tokens. This process involves marking the specific spans or tokens that explicitly express suggestions.

Data Preprocessing step consisting of cleaning of data by removing special characters, emoticons, and other irrelevant piece of information from Opinion review input. After data preprocessing, we applied a tokenizer, such as the

WordPiece tokenizer used in BERT and DeBERTa to split the opinion reviews into individual tokens. Tokenization breaks down the text into smaller units, such as words or sub-words, to represent the input text in a numerical format suitable for model input.

To ensure robust model training and evaluation, it is a common practice to divide the dataset into training, validation, and testing sets. In our approach, we performed a stratified split of the annotated dataset into three distinct subsets: training, validation, and testing. The dataset was split in a ratio of 70:15:15, with 70% of the data allocated for training the model, 15% for validating and fine-tuning the model during the training process, and the remaining 15% reserved for final testing and evaluation. This stratified split ensures that each subset maintains a representative distribution of the original dataset, aiding in accurate performance assessment and generalization of the trained model.

The below example from travel review data shows how the review text is annotated and converted into vector along with the label generation to feed into the DeBERTa model.

- **Review Text:** *be sure to pick up an umbrella* (for free) at the concierge if you anticipate rain while sightseeing.

- **Annotation:** be sure to pick up an umbrella

- **Tokens generated by Tokenizer:** ['[CLS]', '_be', '_sure', '_to', '_pick', '_up', '_an', '_umbrella', '_(', 'for', '_free', ')', '_at', '_the', '_concierge', '_if', '_you', '_anticipate', '_rain', '_while', '_sightseeing', '.', '[SEP]']

- **Input_ids generated based on vocabulary:** [1, 282, 521, 264, 1469, 322, 299, 11908, 287, 2102, 484, 285, 288, 262, 27794, 337, 274, 10570, 2894, 438, 18840, 260, 2].

- **Labels generated are:** [-1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1].

The provided review text contains a suggested opinionated sentence, with the phrase "be sure to pick up an umbrella" annotated as a key phrase representing the suggestion. To prepare the review text for analysis, it undergoes tokenization using the DeBERTa tokenizer, which utilizes the WordPiece tokenizer to split the text into individual tokens. Additionally, special tokens like [CLS] (start) and [SEP] (end) are added to indicate the input boundaries and are ignored during loss computation. The generated tokens are assigned unique IDs, which are then inputted into the model along with supplementary information such as "token_type_ids" and "attention_mask". To facilitate token classification, each token contributing to the key phrase is labeled with '1', while all other tokens in the input sequence are assigned '0'. Special tokens are labeled as '-1'.

3.2. Model Architecture

To reduce the carbon foot-print and make use of existing pre-trained models for the easy and quick experiments, we have chosen the state-of-the-art token classification models, which are pre-trained on a large corpus. We selected models such as BERT or DeBERTa of various sizes to experiment for the suggestion phrase extraction. In the pre-trained step, these models have learned rich representations of language and can be fine-tuned for downstream tasks[6], [7], [31] like suggestion key phrase extraction.

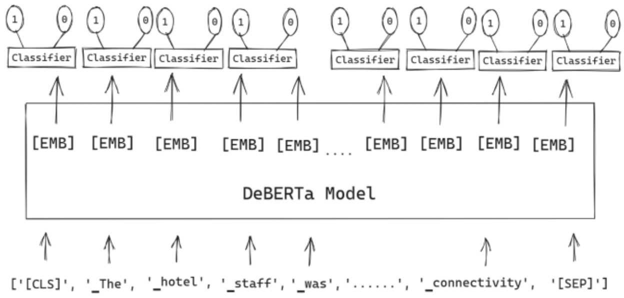

Fig 2. DeBERTa model for Token classification

BERT and DeBERTa are both transformer-based models used for natural language processing tasks. The attention mechanism is a crucial component in both models, allowing them to focus on relevant parts of the input sequence. BERT introduced bidirectional pretraining and consists of multiple transformer layers with a fixed attention mechanism. On the other hand, DeBERTa improves upon BERT by introducing disentangled attention, which separates attention weights into content and position attention, capturing both semantic content and relative positions of tokens. While both models have similar architectures with stacked transformer layers, DeBERTa's enhanced attention mechanism sets it apart from BERT.

3.3. Experimental setup and Model Finetuning

To experiment the Token classification approach for suggestion key phrase extraction from opinion reviews using pre-trained models, we have used Google Colab pro plus. Colab Pro-plus provides, 45GB of RAM and 160 GB of disc space, which is sufficient for model fine-tuning. For creating and fine-tuning model, we used Huggingface transformers framework and specific classes of BERT and DeBERTa along with their tokenizers.

We utilized the Huggingface Transformers library to load the pre-trained weights of the chosen BERT (BERT-BASE and BERT-LARGE) and DeBERTa (DeBERTa-v3-small, DeBERTa-v3-base, and DeBERTa-v3-large) models, initializing the model's parameters for fine-tuning. Our training process involved the tokenized opinion reviews and their corresponding suggestion annotations from the dataset's training split. During training, the model learned to classify each token as either a suggestion or non-suggestion. To update the model's parameters, we conducted experiments with different optimization algorithms, including Adam and stochastic gradient descent (SGD). Additionally, we employed a learning rate scheduler to dynamically adjust the learning rate during the training process. For the loss function, we utilized binary cross-entropy loss, which measured the disparity between the predicted and actual suggestion labels. To mitigate overfitting and enhance generalization, we applied regularization techniques such as dropout and weight decay. During the hyperparameter tuning phase, we extensively explored a range of values for various parameters. In terms of optimization algorithms, we considered both Adam and SGD to find the most effective approach. We evaluated learning rate values of 1e-5, 2e-5, 3e-5, and 5e-5 to identify the optimal rate for convergence. Furthermore, we experimented with batch sizes of 16, 32, and 64 to determine the batch configuration that balanced efficiency and resource utilization. The number of epochs was varied between 5, 10, and 20 to capture the ideal trade-off between model convergence and computational resources. Dropout rates of 0.1 and 0.2 were tested to assess their impact on model performance. Additionally, we employed both linear and cosine learning rate schedulers to explore different approaches for adjusting the learning rate over the course of training.

After careful evaluation, comparison of model sizes in combination of hyperparameter values that yielded the best performance consisted of DeBERTa-v3-base using the Adam optimization algorithm with a learning rate of 3e-5. We chose a batch size of 32 and a dropout rate of 0.2. The model was trained for 10 epochs, as beyond this point, there were no significant improvements in the metrics and loss values. At this stage, the model had converged, demonstrating stable performance. DeBERTa-v3-large model is overfitted as the less amount of data, if have more annotated data, large model can be fine-tuned for the same task.

3.4. Model Evaluation and Testing

To assess the performance of the model, tune hyperparameters and select the best performing model at the end of training validation split of the dataset has been utilized. During the training steps, at every epoch we calculated evaluation metrics such as accuracy, precision, recall, and F1 score to measure the model's performance in correctly identifying suggestion-related tokens.

To check the model's performance and to improve, we conducted a thorough analysis of the model's predictions, including the identification of false positives and false negatives. Analyze cases where the model incorrectly labels tokens as suggestions or fails to identify actual suggestion key phrases. We also conducted experiment to visualize and understand the top loss validation examples, though which we examined for which kind of example model is struggling to make right prediction.

The generalization ability of the model has been evaluated by checking its performance on an independent testing set. Obtain unbiased performance metrics, such as accuracy and F1 score, to measure the model's effectiveness. Applied the trained model to unseen data samples to extract the key phrases from suggestion intended opinion reviews.

3.5. Evaluation Metrics

We employed the following metrics to evaluate the model's performance.

o Accuracy: Accuracy measures the overall correctness of the model's predictions by calculating the ratio of correctly classified tokens to the total number of tokens in the dataset. However, accuracy might not be the best metric when dealing with imbalanced datasets.

o Precision: Precision measures the proportion of correctly classified positive tokens (suggestion key phrases) out of all tokens predicted as positive. It indicates the model's ability to avoid false positives and is calculated as the ratio of true positives to the sum of true positives and false positives.

o Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of correctly classified positive tokens out of all actual positive tokens. It indicates the model's ability to capture all positive instances and is calculated as the ratio of true positives to the sum of true positives and false negatives.

o F1-score: F1-score is the harmonic mean of precision and recall. It provides a balanced measure of a model's performance by considering both precision and recall. F1-score is commonly used when there is an imbalance between positive and negative classes in the dataset.

4. Result analysis and Discussion

4.1. Discussion on Results

The models effectively captured explicit suggestion key phrases, even in cases where suggestions were expressed indirectly or subtly. Among the models evaluated, DeBERTa Base emerges as the top-performing model for suggestion key phrase extraction, exhibiting the highest accuracy, precision, and F1 score. With an accuracy of 0.927 and precision and recall values of 0.93 and 0.925 respectively, DeBERTa Base demonstrates its superior capability in accurately identifying and extracting suggestion-related tokens.

BERT Large also showcases strong performance, with an accuracy of 0.917 and a balanced precision and recall score of 0.92 and 0.915 respectively. These results indicate its effectiveness in capturing suggestion key phrases. While BERT Base and DeBERTa Large both achieve an accuracy of 0.905 and 0.915 respectively, their precision and recall scores show a slight variance. BERT Base achieves a precision of 0.90 and recall of 0.91, resulting in an F1 score of 0.904. DeBERTa Large maintains consistent precision, recall, and F1 score values of 0.915.

On the other hand, DeBERTa Small demonstrates consistent performance with an accuracy of 0.90 and precision, recall, and F1 score values of 0.90. Although falling slightly behind the other models, it still presents competitive results. The variations in performance among these models can be attributed to differences in architecture, training methodologies, and optimization techniques. DeBERTa Base, with its enhanced architecture and comprehensive pre-training, stands out as the most effective model in accurately identifying and extracting suggestion key phrases.

Table 2. Performance comparison of various models

Model Name	Accuracy	Precision	Recall	F1 Score
BERT Base	0.905	0.90	0.91	0.904
BERT Large	0.917	0.92	0.915	0.915
DeBERTa Small	0.90	0.90	0.90	0.90
DeBERTa Base	0.927	0.93	0.925	0.924
DeBERTa Large	0.915	0.915	0.915	0.915

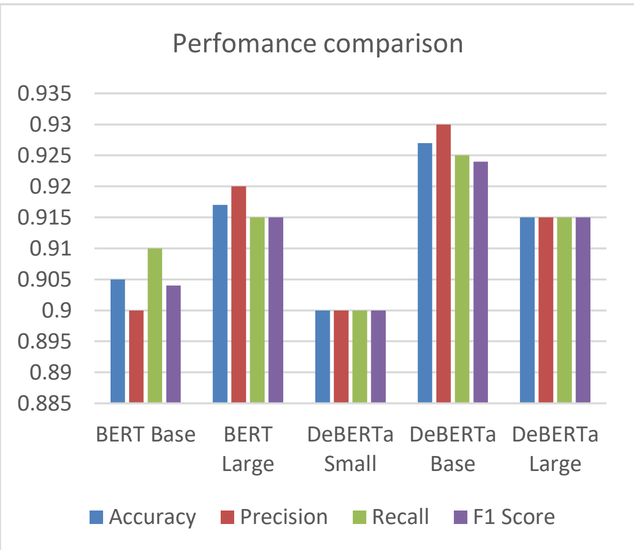

Fig 3. Performance comparison

The qualitative analysis confirms that both models can successfully capture explicit suggestion key phrases, even in cases with subtle or indirect expressions.

4.2. Qualitative Analysis

Here we present some qualitative analysis by capturing two examples from travel reviews dataset.

Example 1: Opinion Review - "The hotel staff was friendly and helpful, but I suggest improving the Wi-Fi connectivity."

- *BERT-large Extracted Suggestion Key Phrases: "improving the Wi-Fi connectivity"*
- *DeBERTa-v3-base Extracted Suggestion Key Phrases: "improving the Wi-Fi connectivity"*

Both models correctly identified the suggestion key phrase related to Wi-Fi connectivity improvement.

Example 2: Opinion Review - "The restaurant offers great food, but they should expand their menu options."

- *BERT-large Extracted Suggestion Key Phrases: "expand their menu options"*
- *DeBERTa-v3-base Extracted Suggestion Key Phrases: "expand their menu options"*

Both models accurately captured the suggestion for menu expansion.

5. Conclusions and Future Work

In conclusion, our experimental results showcase the effectiveness of the Token Classification approach, specifically utilizing BERT and DeBERTa, for extracting suggestion key phrases from opinion reviews. We found that DeBERTa outperformed BERT, exhibiting higher accuracy, precision, recall, and F1 score in suggestion key phrase extraction. These findings underline the significance of accurate suggestion key phrase extraction in various applications, including sentiment analysis, opinion mining, and customer feedback analysis.

The Token Classification approach proved its efficacy in accurately identifying and extracting suggestion-related tokens by performing token-level classification. Both BERT and DeBERTa demonstrated strengths in extracting suggestion key phrases, showcasing their robustness in handling diverse linguistic variations, such as synonyms, abbreviations, and grammatical structures. These models effectively captured implicit suggestions subtly expressed within the opinion reviews. However, we observed some limitations, including difficulties in handling ambiguous phrases or rare suggestion patterns. Further research is needed to address these limitations and explore novel techniques that can enhance the performance of suggestion key phrase extraction models.

By leveraging the capabilities of BERT and DeBERTa for suggestion key phrase extraction, we have made significant strides in improving the accuracy and effectiveness of sentiment analysis and opinion mining tasks. The ability to extract precise suggestion key phrases provides valuable insights for businesses in understanding customer preferences and enhancing their products or services accordingly. Overall, our findings highlight the potential impact of accurate suggestion key phrase extraction and emphasize the importance of continued research and development in this area to overcome existing limitations and further advance the field.

Despite the promising results, there are some limitations to consider. The models may struggle with suggestions expressed implicitly or require more robust handling of negation. Future work could explore incorporating domain-specific knowledge or incorporating multi-modal approaches to improve suggestion key phrase extraction. Investigating ensemble methods or transfer learning from related tasks could also enhance the performance of the models.

Acknowledgements

We thank friends from Xelpmoc Design and Tech Ltd., who provided insight and expertise that greatly assisted the research. We thank Mr. Sambit Mukherjee, Mr. Aravind Selvam and Mr. Wasim Madha for the assistance and discussions about fine tuning approaches using various state-of-the-art models.

Conflicts of interest

The authors declare no conflicts of interest.

References

[1] S. Negi and P. Buitelaar, “Towards the extraction of customer-to-customer suggestions from reviews,” Association for Computational Linguistics, 2015. doi: 10.18653/v1/d15-1258.

[2] S. Negi, M. de Rijke, and P. Buitelaar, “Open Domain Suggestion Mining: Problem Definition and Datasets,” Jun. 2018, [Online]. Available: http://arxiv.org/abs/1806.02179

[3] N. K. Laskari and S. K. Sanampudi, “A SYSTEMATIC STUDY ON SUGGESTION MINING FROM OPINION REVIEWS,” *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 20. Little Lion Scientific, pp. 6061–6072, Oct. 31, 2022.

[4] S. Negi, K. Asooja, S. Mehrotra, and P. Buitelaar, “A Study of Suggestions in Opinionated Texts and their Automatic Detection.” [Online]. Available: https://feedly.uservoice.com/forums/192636-

[5] “2019 - SemEval-2019 Task 9 Suggestion Mining from Online Reviews and Forums”.

[6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[7] P. He, X. Liu, J. Gao, and W. Chen, “DeBERTa: Decoding-enhanced BERT with Disentangled Attention,” Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.03654

[8] N. Jindal and B. Liu, “Review spam detection,” in *Proceedings of the 16th international conference on World Wide Web - WWW ’07*, 2007. doi: 10.1145/1242572.1242759.

[9] N. K. Laskari and S. K. Sanampudi, “Aspect Based Sentiment Analysis Survey,” vol. 18, no. 2, pp. 24–28, doi: 10.9790/0661-18212428.

A. Viswanathan, P. Venkatesh, B. Vasudevan, R. Balakrishnan, and L. Shastri, “Suggestion mining from customer reviews,” *17th Americas Conference on Information Systems 2011, AMCIS 2011*, vol. 2, pp. 1351–1359, 2011.

[10] S. Negi and P. Buitelaar, “Inducing Distant Supervision in Suggestion Mining through Part-of-Speech Embeddings,” Sep. 2017, [Online]. Available: http://arxiv.org/abs/1709.07403

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*, pp. 1–12, 2013.

[12] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations ofwords and phrases and their compositionality,” *Adv Neural Inf Process Syst*, pp. 1–9, 2013.

[13] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global Vectors for Word Representation.” Accessed: Jun. 27, 2020. [Online]. Available: http://nlp.

[14] “2019 - Yimmon at SemEval-2019 Task 9 Suggestion Mining with Hybrid”.

[15] M. Yamamoto and T. Sekiya, “m y at SemEval-2019 Task 9: Exploring BERT for Suggestion Mining.” [Online]. Available: https://competitions.codalab.org/competitions/19955

[16] R. Alexandros Potamias, A. Neofytou, and G. Siolas, “NTUA-ISLab at SemEval-2019 Task 9: Mining Suggestions in the wild.” [Online]. Available: https://www.tripadvisor.com.gr/

[17] “2019 - MIDAS at SemEval-2019 Task 9 Suggestion Mining from Online Reviews”.

[18] S. Pecar, M. Simko, and M. Bielikova, “NL-FIIT at SemEval-2019 Task 9: Neural Model Ensemble for

Suggestion Mining," Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.02981

[19] C. Park *et al.*, "ThisIsCompetition at SemEval-2019 Task 9: BERT is unstable for out-of-domain samples," Apr. 2019, [Online]. Available: http://arxiv.org/abs/1904.03339

[20] J. Liu, S. Wang, and Y. Sun, "OleNet at SemEval-2019 Task 9: BERT based Multi-Perspective Models for Suggestion Mining."

[21] M. Klimaszewski and P. Andruszkiewicz, "WUT at SemEval-2019 Task 9: Domain-Adversarial Neural Networks for Domain Adaptation in Suggestion Mining."

[22] S. Prasanna and S. A. Seelan, "Zoho at SemEval-2019 Task 9: Semi-supervised Domain Adaptation using Tri-training for Suggestion Mining," Feb. 2019, [Online]. Available: http://arxiv.org/abs/1902.10623

[23] P. Yue, J. Wang, and X. Zhang, "YNU-HPCC at SemEval-2019 Task 9: Using a BERT and CNN-BiLSTM-GRU Model for Suggestion Mining."

[24] N. K. Laskari and S. K. Sanampudi, "a Systematic Study on Suggestion Mining From Opinion Reviews," *J Theor Appl Inf Technol*, vol. 100, no. 20, pp. 6061–6072, 2022.

[25] N. K. Laskari and S. K. Sanampudi, "Explainable System for Suggestion Mining using attention," pp. 679–684, 2022, doi: 10.1109/icaccs54159.2022.9785102.

[26] N. K. Laskari and S. K. Sanampudi, "Weighted Focal Loss Function for Evaluating Effectiveness of Word Embeddings on Suggestion Mining from Opinion Reviews," vol. 18, no. 6, pp. 1464–1483, 2021.

[27] M. Jain, M. Leekha, and M. Goswami, *A Multi-Task Approach to Open Domain Suggestion Mining (Student Abstract*, vol. 1. Springer International Publishing, 2020. doi: 10.1609/aaai.v34i10.7180.

[28] N. K. Laskari and S. K. Sanampudi, "ASPECT-ORIENTED SUGGESTION MINING FROM OPINION REVIEWS," *J Theor Appl Inf Technol*, vol. 30, no. 12, 2023, [Online]. Available: www.mouthshut.com

[29] A. Vaswani et al., "Attention Is All You Need," Jun. 2017, [Online]. Available: http://arxiv.org/abs/1706.03762

[30] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, no. Mlm, pp. 4171–4186, 2019.

[31] M, S. ., A, B. ., R, R. ., Patil, H. Y. ., R, D., A, K. ., Andrews, A. M. ., & Nagpal, A. . (2023). Dual-Level Secured Autonomous Bank Locker System. International Journal on Recent and Innovation Trends in Computing and Communication, 11(4), 269–277. https://doi.org/10.17762/ijritcc.v11i4.6452

[32] Martínez, L., Milić, M., Popova, E., Smit, S., & Goldberg, R. Machine Learning Approaches for Human Activity Recognition. Kuwait Journal of Machine Learning, 1(4). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/146