O2S-SKE: Encoder-Decoder Model for Suggestion Key Phrase Extraction

Naveen Kumar Laskari[1](✉) and Suresh Kumar S[2]

[1] JNT University, Hyderabad, Telangana, India
naveen.laskari@gmail.com
[2] JNTUH College of Engineering, Hyderabad, Telangana, India

Abstract. Suggestion Mining refers to extracting suggestions from the opinionated text. It is a lately identified topic of interest by academia and industry. The pioneering research focused on various methods and approaches to classify opinion reviews into suggestion or non-suggestion classes. However, the methodologies in the literature missed the fine-grained analysis, such as the extraction of key phrases denoting the suggestion in an opinion review. In this paper, we present our experiments to implement a system to extract the key phrases from opinion reviews indicating suggestions. The methodology adopted includes the state-of-the-art sequence-to-sequence transformer-based models and rule-based systems followed by an ensemble approach to extract the familiar phrase from the above two methods. The performance of the hybrid model has been evaluated using variants of the ROUGE metric. To our knowledge, our hybrid approach is reasonably performing well.

Keywords: Suggestion · Key phrase · Embedding · Rule-Based Model

1 Introduction

Sentiment Analysis studies opinions, attitudes, and sentiments mentioned in written text on various social platforms. With Web 2.0 and technological advancements, the amount of textual data on social media is increasing. The opinionated text not only contains opinions but also includes suggestions. The pioneering work in Sentiment Analysis is carried out by [9] and [10]. In sentiment analysis, the significant contributions classified the sentiment analysis at different levels, such as document, sentence, and aspect levels. In document classification, the entire document is assigned either a positive or negative label, regardless of mixed opinions expressed in the whole document. At the sentence level, opinion extraction is computed, considering the sentence as the primary unit, contributing to the score. In Aspect Based Sentiment Analysis (ABSA), the sentiment is evaluated towards a particular aspect or subject mentioned in the re-view. In ABSA, the essential phases of operations such as Aspect Term Extraction (ATE), Aspect Category Detection (ACD), Aspect Term Polarity (ATP) Detection, and Aspect Category Polarity Detection (ACP) (Laskari & Sanampudi, 2016). In ATE, they identify and recognize the Key terms, which consist of polarity information. In ACD, the significant focus towards

B. Raj et al. (Eds.): ICETE 2023, AER 223, pp. 370–377, 2023.
https://doi.org/10.2991/978-94-6463-252-1_41

identifying the belongingness of terms detected towards a particular category of aspects. The ATP mainly focuses on identifying the polarity of positive, negative, or neutral words. Finally, in ACP, the main focus is on detecting the polarity of the aspect category.

Social platforms such as Twitter, Facebook, Instagram, and third-party web portals have become the first place to share opinion reviews about various things. Some opinion posts are concise and not much helpful and informative, but a few opinion sharings are a reasonably good length and more beneficial. Reading through the entire content on these platforms for a specific study and getting insights is cumbersome, especially for people seeking particular suggestions from unknown sources of information. Suggestion Mining is the lately identified subject of interest in Natural Language Processing applications, extracting suggestions from opinion reviews[2][3]. The pioneered suggestion mining work focused on classifying opinion reviews into suggestions or non-suggestion. The earlier work ignored the fine-grained analysis of extracting the key phrase that indicates the opinion review suggestion. This paper adopts a hybrid approach with Transformer architecture and a rule-based system to perform the Suggestion Key phrase Extraction (SKE) from opinion reviews. The model is named O2S-SKE (Opinion2Suggestion—Suggestion Key phrase Extraction).

The significant contributions of the paper are:

1. Converting and annotating the binary classification data suitable for seq2seq modelling.
2. Adopting the Transformer architecture to extract the key phrase from the opinion review
3. Devising a rule-based model for Suggestion Key phrase Extraction (SKE), and
4. Defining the proper evaluation metrics for seq2seq models.

The rest of the paper is organized as follows. In Sect. 2, we present the literature of work on suggestion mining. Section 3 explains the transformer-based model and rule-based system that we have adopted. Section 4 elaborates on evaluation metrics. The final section presents the results & discussions, followed by conclusions and future scope.

2 Literature Study

Suggestion Mining is a specific task in the field of NLP. The concept of suggestion mining was introduced by [2] for the first time. [4] They have experimented with similar work by collecting opinion reviews from various platforms such as Microsoft Windows Phone. The authors collected 3000 tweets from Twitter using the search keyword "Windows Phone 7" from September 2010 to April 2012. This tweet dataset contains sentences representing the product improvement suggestions mentioned about the Microsoft window-7 phones. The authors used various approaches to classify the annotated data as a suggestion or not a suggestion.

[5] formulated the suggestion mining problem, and the authors collected 10440 reviews about 14 different products ranging from mobile phones and digital cameras from Amazon's Mechanical Turk. The dataset has been manually annotated suggestion intended and non-suggestion-intended reviews. In which 1880 reviews of suggestion intended and 8560 are non-suggestion intended. The models used are decision trees,

support vector machines, and random forest with manually extracted linguistic features TF-IDF and other feature extraction approaches. They were later extended with neural network approaches such as Recurrent Neural Networks and Long Short-Term Memory networks.

In [4], the authors implemented a suggestion mining approach towards extracting customer-to-customer (CTC) suggestions. The authors presented their model for data preparation and annotated CTC suggestions by scraping data from TripAdvisor hotel reviews and electronics review datasets. The SVM model with various kernels and cross-validation values with linguistic and hand-crafted features such as unigram, bi-gram, POS, and reported results. In [6], the authors introduced a distant supervision approach for suggestion mining with Glove embeddings of various sizes by introducing three different domain datasets: hotel, travel, and software [7], The authors implemented deep learning approaches such as CNN and RNN with attention models toward CTC using hotel and electronics review datasets.

In (S. Negi & Buitelaar, 2019), they organized a shared task on Suggestion Mining as a part of SemEval-2019, a renowned international NLP workshop. The pilot task consists of two subtasks: open-domain and cross-domain suggestion classification. In subtask-1, open domain, the model for the sort of opinion reviews needs to be trained and evaluated on the same data field. In subtask 2, the cross-domain suggestion classification. The model has to be trained on one domain dataset, and evaluation must be done on another. In response to SemEval-2019, 33 teams submitted their solutions for subtasks. Most implementations are with the help of transfer learning and pre-trained models.

The suggestion mining dataset SemEval organizers provide is imbalanced in nature and biased towards non-suggestion. To handle class imbalance [9] experimented with focal loss function in training the neural network models for suggestion mining classification. In addition, the authors tested approaches such as a convolutional neural network for sequence classification and a Long Short-Term Memory network with various combinations of word embeddings.

Key phrase extraction has been studied in a different setting with other datasets. (Kamil et al., 2018) approached key phrase extraction using sentence embeddings in an unsupervised manner. The authors have embedded the phrases and the entire document separately from the DUC2001 dataset and fed them to the model for training. The model has been evaluated using maximal marginal relevance, adapted pre-trained large language models, and POS-Tags to extract the key phrases in unsupervised settings. The authors utilized the dataset from scholarly articles and tested the model.

In recent times most approaches are deep neural network models, and large pre-trained models are applied to crack the suggestion mining task [3]. To have trust in the deep models and understand the learning of the model and interpreting the model learning, [10] implemented an explainable system for suggestion mining using various neural network architectures and word embedding and visualized the attention weights as model interpretation.

3 Methodology

3.1 Suggestion Key Phrase Extraction (SKE)

In the opinion review, all the sentences and words may not express the suggestion. The pioneering research focused only on classifying the review as a suggestion or not a suggestion. In this work, we are experimenting first, extracting the critical phrase that expresses the suggestion. We employed a rule-based model with POS tagging and a transformer-based encoder-decoder model. The output of these two models is processed through the hybrid system to identify the most extended typical sequence to fetch the right part of the sentence as the suggested key phrase. The steps followed for suggestion extraction are depicted in the Fig. 1 as the algorithm.

Input: Opinion Review Statement(s).

Output: Key phrase that represents the suggestion.

`Step-1: Identify all the suggestions expressed in reviews Manually annotate all suggestion reviews with the suggestion key phrase, the critical phrase describing suggestion in opinion review using Prodi.gy.`

`Step 2: Generate the set of rules using POS tags to identify the key phrase, Apply all the rules to the input reviews Extract the key suggestion key phrase from the input review.`

`Step 3: Train a transformer-based encoder-decoder (Seq2Seq) model to generate the key phrase from the input opinion review.`

`Step 4: Pass the output of the rule-based system and transformer model to the hybrid model to fetch the longest common sequence in both outputs.`

`Step-5: Compare the individual and hybrid models' output with the ground truth and evaluate the performance using BLEU and ROUGE metrics.`

We need the labeled data to train the machine learning model and evaluate the result. But the data provided by the SemEval-2019 organizers is suitable for binary classification, classifying the review as suggestion or non-suggestion. Therefore, in the first step of the key phrase extraction process, we manually annotated the data, such as labeling the key phrase from the entire review.

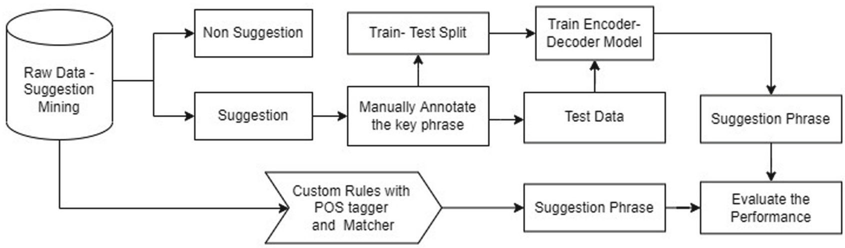

Fig. 1. A Hybrid Suggestion Key phrase Extraction (SKE) Model

In the following algorithm steps, step 2 is for the rule-based system toward key phrase extraction. The next are the steps followed in the pipeline for a rule-based SKE system.

Data Pre-processing and Tokenization: in this step, the suitable word tokenization approach in combination with stemming, lemmatization, and stop word removal has been used after converting each generated token into lowercase.

Removal of URLs, Hashtags, and Punctuation symbols: URLs, hashtags, and punctuations are mentioned in the review statement. To normalize the data, any URL has been replaced by [URL], any hashtag has been replaced by [HASHTAG], and punctuation symbols have been removed.

Preparation of POS-Tag-based rules for phrase extraction using spaCy: In this step, using the spaCy matcher, various rules are constructed, and added all the rules to the spacy model to extract the key phrases.

Extraction of key phrases using the model: after adding the rules to the model, the unseen data can be passed to the spacy model to extract the key phrase from the opinion review.

3.2 Encoder-Decoder Model for SKE

Transformers models had become state-of-the-art across the tasks in NLP, post-release of transformer model by [11] in 2017 with the title “Attention is all you need.“ Popular models like BERT, T5, BART, GPT, and many more originated from the same paper. The transformer architecture consists of a stack of encoder and decoder layers to process the contextual information generated by encoder layers. The more details encoder and decoder layers and their internal components are discussed below in detail (Fig. 2).

Encoder Layer: The essence of the encoder is to read the input sentence, try to make sense of it, and pass the summary as a context vector to the decoder. The encoder of the transformer consists of multiple blocks of the same layers. The bottom layer takes position embedding and the input embedding. Then, the concatenated representation is fed into the multi-head attention and normalization layers, followed by the feed-forward

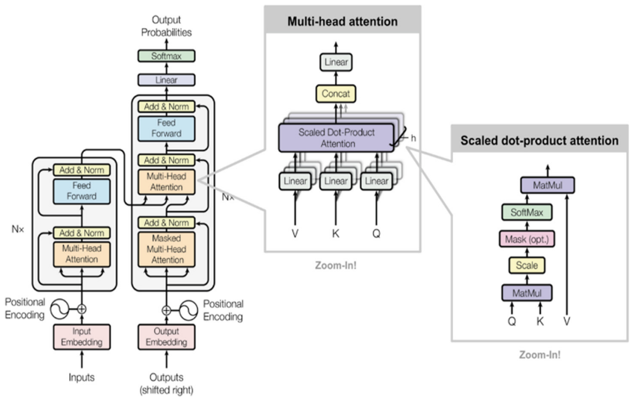

Fig. 2. Transformer Architecture

Table 1 Dataset Sample

Opinion review	Key phrase
Have a look at my topic Great weather site under travel advice	Have a look at my topic
Put all your clothes out and reduce to one third	Put all your clothes out
You can wear shorts over there, but I would wear them when you are cruising, mainly because of the sun	You can wear shorts

layer. A total of four stacked layers are included in the encoder of the resultant model. The output of the last transformer block is considered a context vector, inheriting the input sequence as a context vector.

Decoder Layer: Like the Encoder Block, the Decoder also Consists of Multiple Blocks of the Same Layer. Each Layer Has Masked Multi-Head Attention, Multi-Head Attention, and Feed-Forward Layers; the Output of the Last Decoder Layer is Fed into the SoftMax Layer to Produce the Target Output.

Multi-head Attention Mechanism: The multi-head attention mechanism improves the model performance by focusing on different positions. The input is split into fixed-size segments and computed scaled-dot product over each component in parallel.

3.3 Experimentation

We report the performance of our models for SKE on the modified version of the SemEval-2019 dataset. The data provided by the workshop organizers have labeled data for the binary classification, which has a suggestion or non-suggestion class. We considered only the data belonging to the suggestion class to extract the key phrase from suggestion reviews. Three expert annotators conducted the data annotation process to identify the key phrase from the opinion review. For the annotation process prodi.gy (https://prodi.gy/) tool has been utilized. In the annotation process, if the annotation is agreed upon among a minimum of two annotators as a suggestion, then only the phrase is considered a suggestion key phrase or ignored.

The sample of annotated data is presented in Table 1.

3.4 Evaluation Metrics

ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It is a set of metrics for evaluating text-to-text tasks such as text summarization and machine translations. SKE is also considered a seq2seq task, in which the input and output are also a sequence. The performance of models was evaluated using ROUGE metrics. ROUGE measures count the number of overlapping units such as n-gram, word sequences, and word pairs between generated output and ground truth summaries annotated by humans.

Table 2. Result Analysis

Model Name	**ROUGE-1**			**ROUGE-L**			**ROUGE-2**		
	N = 1	N = 3	N = 5	N = 1	N = 3	N = 5	N = 1	N = 3	N = 5
Seq2Seq Model	28.2	30.15	32.58	10.1	12.8	14.2	14.85	16.34	17.46
Rule-based Model	16.52	18.69	19.92	5.55	8.24	9.56	9.87	12.54	14.64
Hybrid Mechanism	14.25	16.38	17.64	5.15	7.24	8.58	8.67	10.57	12.57

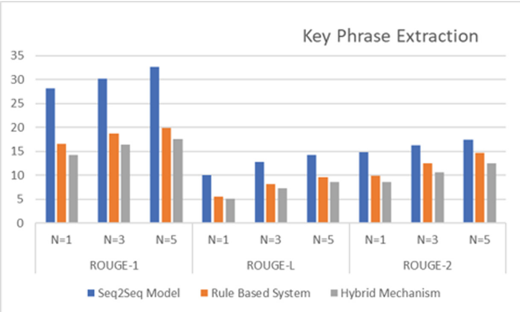

Fig. 3. Comparative study of various models

ROUGE has different variants such as ROUGE-N, ROUGE-L, ROUGE-Wand ROUGE-S. We are using here in this paper ROUGE-L, Longest Common Subsequence used as a metric, which can be defined as follows.

A sequence S = [s_1, s_2,….. s_n] is a subsequence of another sequence Y = [y_1, y_2,…… y_m], if there exists a strictly increasing sequence [i_1, i_2,…. i_k] of indices of S such that for all j = 1,2,…k, we have $S_{ij} = Z_j$.

4 Results and Discussion

We evaluate the performance of our models using two different metrics, namely BLEU and ROUGE scores.

The comparison is made in three ways:

1. Output of the rule-based system with ground truth.
2. Output of transformer model with ground truth.
3. Output of the hybrid model (LCS of two models with the ground truth) (Fig. 3).

The Table 2 presents the first comparison score.

5 Conclusion and Future Scope

This paper presents a hybrid model to extract critical phrases that denote opinion review suggestions. Key phrases capture the most salient topics of a document and facilitate extremely crucial information. Therefore, identifying them in an automated way from a text document can be helpful in NLP applications. In suggestion mining, the Suggestion Key phrase Extraction helps the users quickly identify the suggestions from the various opinion reviews mentioned on the platform. We experimented with three models to extract the suggestion key phrase from the opinion reviews. To the best of our knowledge, we are the first to explore the suggestion phrase extraction from opinion reviews. As a next step, we would like to fine-tune these models to perform better with cross-domain and other domain datasets.

References

1. N. K. Laskari and S. K. Sanampudi, "Aspect Based Sentiment Analysis Survey," vol. 18, no. 2, pp. 24–28. https://doi.org/10.9790/0661-18212428.
2. 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.
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, pp. 6061–6072, 2022.
4. S. Negi and P. Buitelaar, "Towards the extraction of customer-to-customer suggestions from reviews," Association for Computational Linguistics, 2015. https://doi.org/10.18653/v1/d15-1258.
5. 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.
6. 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.
7. H. Golchha, D. Gupta, A. Ekbal, and P. Bhattacharyya, "Helping each Other: A Framework for Customer-to-Customer Suggestion Mining using a Semi-supervised Deep Neural Network," Nov. 2018, [Online]. Available: http://arxiv.org/abs/1811.00379.
8. S. Negi and P. Buitelaar, "Suggestion Mining From Opinionated Text," in *Sentiment Analysis in Social Networks*, Elsevier Inc., 2017, pp. 129–139.
9. 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.
10. N. K. Laskari and S. K. Sanampudi, "Explainable System for Suggestion Mining using attention," pp. 679–684, 2022. https://doi.org/10.1109/icaccs54159.2022.9785102.
11. A. Vaswani *et al.*, "Attention Is All You Need," Jun. 2017, [Online].