Journal of Theoretical and Applied Information Technology
31st October 2022. Vol.100. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SYSTEMATIC STUDY ON SUGGESTION MINING FROM OPINION REVIEWS

NAVEEN KUMAR LASKARI, SURESH KUMAR SANAMPUDI

Data Scientist, Xelpmoc Design and Tech, Research Scholar, JNTUH Hyderabad, Telangana, India

Assistant Professor, Head, IT, JNTU College of Engineering, Jagitial, Telangana, India
E-mail: naveen.laskari@gmail.com, sureshsanampudi@jntuh.ac.in

ABSTRACT

Online product reviews have become eminent in the purchase decision-making process. With progress in web 2.0 technologies, huge volumes of unstructured text data are generated as reviews on e-commerce platforms and third-party web portals. Opinion review mining has become a critical area of research in language processing and applied machine learning. Opinion reviews available across various portals are perceived primarily to understand the sentiment polarity expressed by the reviewer at multiple granularities. The opinion review may also contain suggestions or tips for manufacturers and peer customers. Suggestion Mining refers to the automatic extraction of suggestions from opinionated text. The applications include product quality improvement, peer customer suggestions, summarizing collected surveys and feedback, a recommender system, and enhancing sentiment polarity classification. Suggestion mining is considered a sentence classification task, such as classifying a given review as suggestive intent or not. Various linguistic, syntactic, and semantic features with core machine learning and neural network approaches are used for suggestion mining. This paper presents a comprehensive and systematic review of suggestion mining from opinion reviews and their facets in the literature.

Keywords: Suggestion Mining, Word Embedding, Deep Learning, Opinion Reviews.

1. INTRODUCTION

Opinion reviews on various platforms like e-commerce portals, micro-blogs, and third-party web portals become part of human life for any decision-making process. These platforms had become the primary place to express feelings, and concerns and share feedback [1]. These also act as a primary source to acquire public opinions towards various entities, persons, events, and organizations. The study of the opinions from these social platforms is referred to as Opinion Mining[2]. Opinion mining is concerned with analyzing opinionated text in terms of sentiment, subjectivity, emotions, suggestions, arguments, etc. Therefore, sentiment analysis and opinion mining should not be used interchangeably since sentiment analysis can be considered a sub-task of opinion mining. The pioneering work in Opinion Mining focused on identifying or summarizing the polarity of the reviews as positive or negative[3]. Later points extended at different levels of granularity. We observe that opinion reviews also contain descriptive information like suggestions, tips, warnings, and recommendations[1]. These suggestions help the manufacturers, and service providers improve the quality of products or services. Peer customers also get benefited by looking into recommendations for the better utilization of services[3]. Peer customer suggestions are also mentioned as customer-to-customer (CTC) suggestions in the literature.

Suggestion mining mainly deals with extracting sentences containing suggestions from opinion reviews of text in nature. Most of the work done related to suggestion mining, classifying the sentences into suggestions and non-suggestions[4].

A standard definition followed in the literature on suggestion mining[1], [2], [4]:

Given a sentence S, predict a label l for S where l Є { suggestion, non-suggestion }

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The pioneering work on suggestion mining was done in [5] by taking the data from the opinion review platform. In [6] Wicaksono and Myaeng (2013), similar work was completed considering suggestion mining as advice. [7] Brun, C.D., & Hagège, C. (2013) also contributed to suggestion mining with a focus on improving product quality. Sapna Negi et al. conducted multiple experiments from 2015 to 2018 on suggestion mining and made available labeled datasets. In SemEval 2019, [4] introduced the pilot task Suggestion Mining to classify the given sentences into the suggestion and non-suggestion classes. In two subtasks, they evaluated the performance of various systems that work for domain-specific and cross-domain datasets, similar to the current progress in the suggestion mining task.

Most approaches used for suggestion mining tasks are traditional machine learning algorithms such as Random Forest (RF), Support Vector Machines(SVM), and statistical methods with linguistic and non-linguistic features [8]. Neural Network approaches such as Recurrent Neural Networks (RNN), Long Short-Term Memory networks (LSTM), in the combination of various word embedding techniques [9]–[11] pre-trained language models, and transformer models are used for Suggestion mining and reported state-of-the-art results.

The contribution of this survey paper is significant for several reasons. First, the report presents unique criteria to frame the research questions presented in section 2, the comprehensive survey of methods, and techniques used for suggestion mining as answers to the questions mentioned. Thus, this is the first of its kind review paper on suggestion mining. Second, we also discuss various challenges involved in suggestion mining when compared to the Sentiment Analysis task.

The rest of the paper is organized as follows: Section 2 presents the research method we followed and a list of research questions formulated. Then, the answers to the research questions are given in section 3. And the synthesis of all the research questions is addressed in section 4. Finally, section 5 offers the conclusion and future scope.

2. RESEARCH METHOD

The following are the parameters considered for framing research questions:

- Target users of produced results - peer-customer or service provider
- Kinds of suggestion extracted - Implicit and Explicit suggestions
- Types of solutions proposed – Open-domain and cross-domain
- Human Intervention - Making availability of annotated datasets for training models
- Comparison of various approaches built for suggestion mining

2.1. Research Questions

To provide a constructive survey on suggestion mining from the available research contributions, we formulated the following research questions

RQ1: What are the datasets made available for research on Suggestion Mining?
Answer to the question provided in section 3.1 with a detailed note on all the available datasets for suggestion mining
RQ2: What kind of methodologies are used for Suggestion Mining?
The answer to this question provided an elaborated explanation of various feature extraction methods, models, neural network approaches, and transfer learning methods in section 3.2
RQ3: Who are the end-users, and how does it help end-users?
The system's end-users are peer customers or service providers, a detailed note is presented in section 3.3.
RQ4: What are the metrics used to evaluate the performance of the models?
A clear message was shown related to evaluation metrics in section 3.4
RQ5: What are the challenges and research gaps identified in the current state-of-the-art method in the literature?

2.2. Search Process conducted

In conducting the comprehensive survey on suggestion mining, we collected research articles and papers from popular article repositories like Elsevier, Springer, Google Scholar, and other sources with the search keywords as Suggestion Mining or Suggestion Mining from opinion reviews. Furthermore, most of the research articles are found in submissions to the shared task organized by Sapna Negi et al.; in the year 2019 part of the SemEval-2019 workshop[4].

2.2.1 Selection criteria

A total of 95 research articles were collected

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

based on the search criteria defined. We considered the filtration process for identifying the relevant articles and completely irrelevant reports from the collection. The requirements that we specified made our job more accessible and more feasible for researching accurately and correctly.

2.2.2 Inclusion criteria

All the submissions made to participate in the SemEval-2019 shared task for suggestion mining are considered and included in the final list of papers for conducting the review. In addition, all the articles were published with different model architectures using the dataset provided by SemEval-2019 organizers, also considered part of the review process.

2.2.3 Exclusion criteria

We excluded the survey papers and papers that had keywords such as suggestion and were not relevant to the context of suggestion mining.

2.2.4 Quality Assurance

To make our paper more quality, we followed and included the work here for review with a good readability score and a good number of citations. Also, we made sure that the clean and clear papers that explained the methodology and results were only considered for the review. We, as authors, reviewed every article thoroughly and included it for study. The below graph (Figure 1) represents the number of articles considered for the review study from 2011 to 2022. Most of the research articles are published as part of the SemEval-2019 workshop.

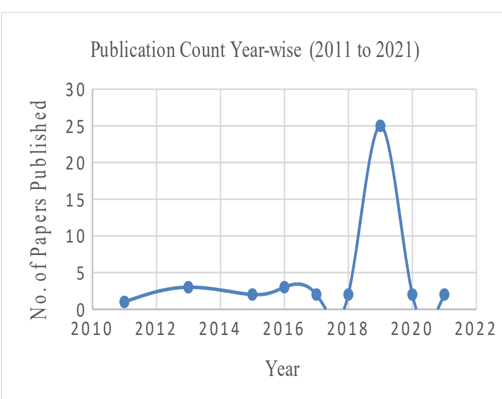

Figure 1: Year-wise publications

3. DETAILED ANSWERS TO THE RESEARCH QUESTIONS

3.1 What are the Datasets made available for research on Suggestion Mining?

Suggestions in opinion reviews have been identified by considering the opinionated text from various sources such as product review portals, discussion forums on travel, sociopolitical, and Twitter. [12] Collected 3000 reviews from Twitter with the search keyword "Windows Phone 7" from the timeline between September 2010 to April 2012. This tweet dataset contains sentences representing the product improvement suggestions mentioned about the Microsoft window-7 phones. Each tweet is manually annotated and finally included in the dataset, based on the agreement score between annotators. In a total of 3000 tweets in the dataset, only 238 tweets are of suggestion class, and the rest are non-suggestion type.

Jhamtani, H., Chhaya, N., Karwa, S., Varshney, D., Kedia, D., & Gupta, V. (2015, December) [13] Collected a dataset of 10440 product review sentences related to mobile phones and digital cameras. Each of the reviews was annotated by three experts and agreed upon with a kappa score of 0.91. In 10440 review sentences, 1880 sentences as a suggestions, and the remaining 8560 as a non-suggestion class.

Negi, S & Buitelaar P [2] scraped 6300 sentences from political discussions and 5000 sentences from travel discussion forums as a dataset in suggestion mining. In addition, SemEval-2019 organizers shared an annotated dataset collected from various social platforms like third-party web portals and e-commerce sites related to multiple domain categories[1]. Complete details of the datasets are summarized in Table 1.

3.2 RQ2 - What kind of methodologies are applied toward suggestion Mining?

Suggestion mining is a relatively young field of research in Natural language Processing (NLP) and lately got attention from academia and industry. Suggestion Mining is considered a text classification task, classifying the given opinion review into suggestion or non-suggestion. First, as represented in figure 2, the raw data collected from multiple sources are annotated and labeled as a suggestion or non-suggestion class. In the next step, the preprocessing of the data is conducted.

Table 1: Summary of Suggestion Mining Datasets

S.No	Dataset	Authors	Category or dataset information	Dataset Size
1	Tweet Dataset about Microsoft phones	Dong et al., 2013	Twitter dataset about windows phones about product improvement with keyword search Windows Phone 7.	3000
2	Amazon's Mechanical Turk, mobile phones, and digital camera	Jhamtani, H QT AL. 2012	Manually annotated sentences from the Amazon Turk dataset	10440
3	Travel Advice dataset and political discussions	Goldberg et al, Wicaksono and Myaeng, 2013	Sentences extracted from discussion threads.	11300
4	Electronic and hotel review dataset	Negi and Buitelaar, 2015b	Prepared from social networks, the sentences which convey suggestions to the fellow customers	7534
5	User voice suggestion forum	Negi and Buitelaar, 2015	Customer posts have been crawled and labeled a subset of the software suggestion forum.	6762

Data preprocessing is essential due to the noise and inconsistency in the data extracted from social platforms. Then, the suitable vectorization methods are applied and passed to the model based on the approaches used for suggestion classification. After building the models, performance evaluation and fine-tuning of the model are done in subsequent steps. Finally, the trained/best model is used to predict the class label for the unseen data. In the following subsection, we present all the steps applied in working with suggestion mining models in-detail.

3.2.1 Data Pre-processing

Preprocessing of unstructured raw data for any NLP task is essential due to noise and inconsistency in the data. For the suggestion mining task, different kinds of preprocessing methods were applied in the literature to improve the results of the classification task. A few critical preprocessing approaches are:

- **Text cleaning:** Removing non-alphanumeric characters, and special symbols from the reviews.
- **Character and word normalization:** Character level and word level normalization has been applied to various components of the reviews like apostrophe, punctuation, date and time, URL, and Hashtag.
- **Shorten phrase expanding:** The reviews in the dataset have many shortening words, which may not be present in the vocabulary of embedding models. So the majority of approaches used conversion from shortening phrases into appropriate long-term forms such as don't to do not.

3.2.2 Word Embeddings

A word embedding is a learned representation of text where contextually similar words have an equal representation. Word embedding methods learn a real-valued vector representation for a predefined fixed-sized vocabulary from a large corpus of unannotated text data. Numerous critical applications are built on top of word embeddings, such as POS tagging, syntactic parsing, named entity recognition, sentiment analysis, question answering, and machine translation. The popular word embedding approaches in the literature are Word2Vec[14], Glove[15], FastText[16], ELMo[17], ULMFiT[18], CoVe, OpenAI GPT[19], and BERT[20]. The word embedding techniques can be classified into static and dynamic embeddings. Static embeddings do not change with context as such after learning. The first stage develops sparse and high-dimensional vectors in static embedding, whereas the second phase involves dense and low-dimensional vectors. Though the low dimensional dense representation achieved better results, but could not solve the problem of polysemy. To address the issue of polysemy, a more effective dynamic representation of words is introduced called contextual embedding. Contextual embedding addresses these challenges and captures contextual information.

3.2.3 Approaches for Suggestion Mining

Suggestion mining is considered to be a text classification task. The majority of the approaches

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

used for this task reported state-of-the-art results using deep learning methods. Primarily, sequence modeling techniques such as LSTM, GRU in combination with BERT, and ULMFiT, Glove, and Word2Vec embeddings are used. The various implementations have been done for suggestion mining tasks considering Domain-specific and Cross-Domain datasets.

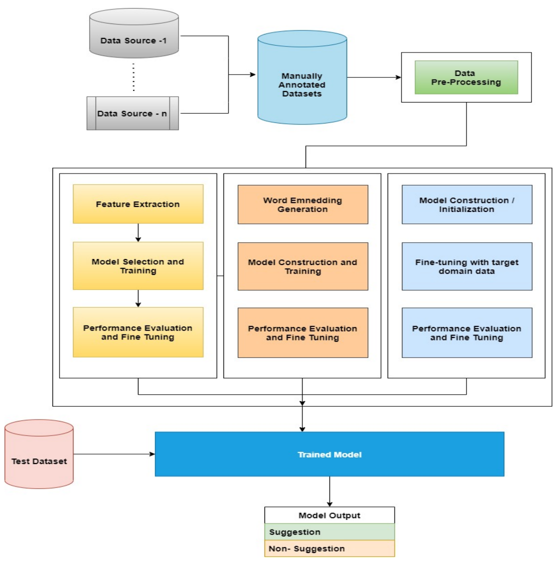

Figure 2: Model Architecture

To compare the analysis of work done for suggestion mining tasks, we considered the submitted models, such as open domain and Cross-domain. Open-domain models as such implemented and tested on the same domain dataset, whereas cross-domain models are trained on one domain dataset and validated and tested across other domain datasets. The comparison is made across the methodologies used, such as machine learning or deep learning: the F1-score, the performance measure used to compare the models for the task. Table 2 summarizes the various approaches and techniques used for suggestion mining tasks. The majority of the implementations are based on deep learning and neural network approaches. Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNNs), and Long Short-Term Memory networks (LSTM), are used for suggestion mining. In addition, transformer-based language modeling techniques such as BERT, and ULMFiT reported state-of-the-art results.

The figure mentions the details of various approaches used for suggestion mining. The techniques are categorized into three classes such as:

- Handcrafted features and Machine Learning models
- Neural Network approaches
- Transfer Learning approaches

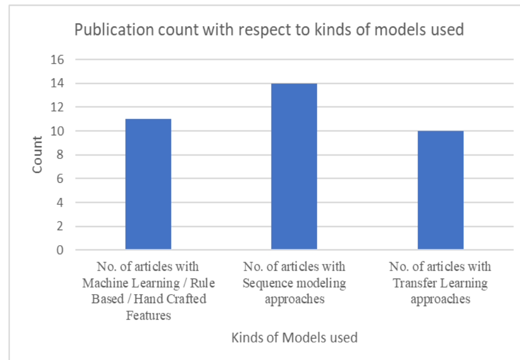

Figure 4: Publication Count to kinds of Models used

Handcrafted Features and Machine Learning Models

In [5] Viswanathan et al, collected opinion reviews from www.mouthshut.com about mobile phones. The model proposed separates the feedback phrases from the review dataset and presents them to the user in a suitable format. The models are built using linguistic features, rule look-up in the knowledge base, ontology-based approach, and general inferencing approaches to achieve better performance. The essential components of their model architecture are knowledge base, pre-processor, named entity extractor, syntactic engine, semantic engine, post-processor, and frame manager. The given review is processed through a pre-processor to check spelling, and sentence detection. The syntactic and semantic engine translates each sentence into a parse tree and maps the features based on POS tags. The post-processor maps the suggestion's intended keywords. Finally, the frame manager outputs the recommendations into a user-desired format.

Brun, C.D., & Hagège, C. (2013) [7] considered the reviews from www.epinion.com, a general website compiling millions of reviews about products, services, movies, and books. In this, the authors evaluated the 3500 reviews about printers and implemented a suggestion mining system. The model captures syntactic-semantic patterns from the reviews. Dong, L., Wei, F., Duan, Y., Liu, X., Zhou, M., & Xu, K. (2013) [7] implemented a suggestion mining system on tweets scraped from Twitter with the keyword search "Window Phone

Journal of Theoretical and Applied Information Technology
31st October 2022. Vol.100. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7". The authors built a suggestion mining classification framework based on Factorization Machines (FM). The authors built several Support Vector Machines model variants with different kernels and features like bag-of-words and hash-tags.

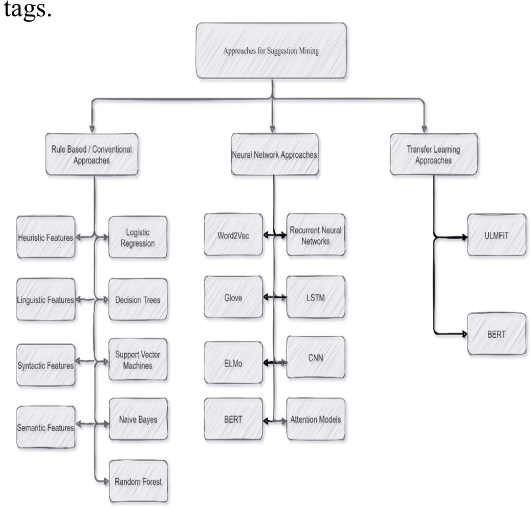

Figure 4: Approaches used for Suggestion Mining

Jhamtani H., Chhaya N., Karwa S., Varshney D., Kedia D., Gupta V. (2015) Considered the dataset of 10440 review sentences from Amazon's Mechanical Turk related to mobile phones and digital cameras. The various supervised machine learning models such as Randon Forest, Decision Trees, and Support Vector Machines are built on top of TF-IDF and TF-IDF with linguistic and syntactic features. However, the combination of all the mentioned features using the Random Forest model produced a better result.

Sapna Negi & Buitelaar, 2015 [1] formulated the concept of customer-to-customer suggestion mining and annotated the dataset from TripAdvisor for hotel and electronic products. Various supervised learning approaches such as Support Vector Machines with different kernels applied heuristic features, suggestion keywords, and POS tag features.

Fatyanosa et al. [21] participated in two subtasks of SemEval -2019 with machine learning approaches towards suggestion mining. Authors experimented with Random Forest (RF), Logistic Regression (LR), Multinomial Naïve Bayes (MNB), Linear Support Vector Machines (LSVM), Sublinear Support Vector Machines (SSVM), Convolutional Neural Network models (CNN), and Variable Length Chromosome Genetic Algorithm – Naive Bayes (VLCGA-NB). Out of all the methods implemented, RF, SSVM, and VLCGA-NB produced comparatively better performance.

In [21] Oostdijk & Van Halteren, They have participated in SemEval-2019 with an expert-rule-based system by using the task-specific lexicon and handcrafted rules. The authors designed rules in such a way that those target reviews only it is suggestion-oriented. In their lexicon base, a total of 1256 entries and 138 rules are created. The authors reported that the rule-based model produced better results for subtask B, and machine learning and combinational methods produced better results for subtask-A.

Neural Network Approaches

Negi, S., Asooja, K., Mehrotra, S., & Buitelaar, P. (2016) [7][1], [22] considered the five datasets presented in the list of datasets available and presented a qualitative and quantitative analysis on suggestions in opinion reviews. Authors applied SVM, LSTM, and CNN with domain-dependent and independent training processes and reported that Deep Learning mechanisms produced better results comparatively using Glove embeddings.

Golchha, H., Gupta, D., Ekbal, A., & Bhattacharyya, P. (2018) [6] Proposed a hybrid deep learning model with recurrent encoder, and convolutional encoder, along with linguistic features. The authors extracted the most frequent 300 unigrams, 100 bi-grams, and 100 tri-grams along with POG tags for each gram in linguistic features. The authors also experimented with a semi-supervised approach known as bootstrapping. A combination of hybrid and self-training methods produced better results.

Golchha et al., 2018 [6] participated in SemEval-2019 with the hybrid LSTM model with handcrafted features along with Glove word embeddings. Pecar, S., Simko, M., & Bielikova, M. (2019). Built an ensemble Bi-LSTM with an attention mechanism initialized using ELMo embeddings. (Alexandros Potamias et al., 2019) Participated in both the subtasks of SemEval-2019 with the rule-based system using heuristic, linguistic and syntactic features and also used deep learning models.

Naveen et.al, 2022 [24] implemented a explainable system for suggestion mining. The authors considered multiple neural network models in combination with word embeddings

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and build the models to visualize the learning parameters to understand the suggestion mining system. To deal with the class imbalance problem in suggestion mining the authors [25] utilized the weighted focal loss function in training the neural network for suggestion classification.

Table 2: Summary Of Various Approaches For Suggestion Mining

S.No	System	System Category	Approach / Algorithm Used
1	Viswanathan et al., (2011)	Open Domain	Rule-based / Machine Learning
2	Wicaksono, Alfan Farizki, and SungHyon Myaeng (2013)	Open Domain	Handcrafted features and Machine Learning / SVM, HMM Model
3	Brun, C.D., & Hagège, C. (2013)	Open Domain	Handcrafted features and Machine Learning / SVM
4	Dong, L., Wei, F., Duan, Y., Liu, X., Zhou, M., & Xu, K. (2013)	Open Domain	Factorization Methods / Machine Learning / SVM
5	Jhamtani H., Chhaya N., Karwa S., Varshney D., Kedia D., Gupta V. (2015)	Open Domain	Handcrafted features and Machine Learning / SVM with TF-IDF
6	Anand, Sarthak, et al.(2019)	Open Domain	Neural Networks / Word Embeddings / ULMFiT Language Modelling / Transfer Learning / LSTM FastText Embeddings.
7	Cabanski, Tobias. (2019)	Open and Cross-Domain	Ensemble model of CNN & LSTM architecture with BERT Embeddings as Input Features
8	Zhou, Qimin, et al.(2019)	Open Domain	CNN Model Architecture with BERT Embeddings as Input Features
9	Yue, Ping, Jin Wang, and Xuejie Zhang. (2019)	Open and Cross-Domain	Transfer Learning / BiLSTM, GRU with BERT Embeddings as features CNN Model for Task-B
10	Zhuang, Yimeng. (2019)	Open Domain	Neural Networks / Self-Attention Network with BERT Embeddings as Input Features
11	Klimaszewski, Mateusz, and Piotr Andruszkiewicz.(2019)	Open Domain	Transfer Learning / Ensemble Adversarial Neural Networks and ELMo Embeddings as Features
12	Park, Cheoneum, et al.(2019)	Open and Cross-Domain	Transfer Learning / Two neural-based encoders using multiple pre-trained word embeddings including BERT
13	Rajalakshmi, S., et al. (2019)	Open and Cross-Domain	A rule-based approach for feature selection, data augmentation and CNN for classification
14	Liu, Jiaxiang, Shuohuan Wang, and Yu Sun.(2019)	Open and Cross-Domain	Ensemble classifier (Logistic, GRU, FFA, CNN), with BERT
15	Yamamoto, Masahiro, and Toshiyuki Sekiya (2019)	Open and Cross-Domain	Transfer Learning / BERT and ULMFiT
16	Potamias RA, Neofytou A, Siolas G. (2019)	Open and Cross-Domain	Automatics rule learning, Lexical and syntactic patterns.
17	Ezen-Can, Aysu, and Ethem F. Can (2019)	Open Domain	A hybrid approach, Glove Embeddings
18	Pecar, Samuel, Marian Simko, and Maria Bielikova.(2019)	Open and Cross-Domain	Bi-LSTM with Self attention Mechanism and Elmo Embeddings
19	Prasanna, Sai, and Sri Ananda Seelan. (2019)	Open and Cross-Domain	Glove and BERT, CNN with Contextual Embeddings, Tri training, semi-supervised bootstrap
20	Jain et al., (2020)	Open Domain	Oversampling on data with Transformer based model
21	Leekha, M., Goswami, M., & Jain, M. (2020)	Open Domain	Oversampling on data with an ensemble of RCNN, CNN, and Bi-LSTM

Transfer Learning Approaches

Yamamoto & Sekiya, 2019 [21], experimented with a popular transfer learning using Bi-directional Encoder Representation from Transformers (BERT). The model training process is done in three steps: pre-training on general domain corpus, pre-training on target domain corpus, and fine-turning on the target task. Finally, an ensemble model makes the final label prediction by computing the average of output scores. The distant supervision approach has been used to generate the noisy annotated data that has been used as a training dataset. The authors claim that the distant supervision model outperformed all the models explored.

In [22] Liu et al., 2019, experimented with a multi-perspective architecture using various neural network models along with BERT. In multi-perspective architecture, various task-specific modules are integrated into the encoder to capture

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

specific features. For example, sentence perspective encoding, Time perspective encoding, and spatial perspective encoding are the different encoding modules added to capture the encoding representation of sentences, time-series perspective information, and spatial connections among the words, respectively. Finally, a majority voting was applied to combine the response of various encoders and generate the final label.

In [23] Anand et al., 2019 applied basic pre-processing on the raw dataset provided by the workshop organizers and applied Universal Language Model Fine-tuning for Text classification (ULMFiT) transfer learning approach. To deal with the imbalanced data, the authors used oversampling technique. The significant components of ULMFiT are the language model and classification model. The language model has been trained on Wiki text corpus to capture the general features of the language and fine-tuned to perform classification using Windows phone review statements.

In [24] Park et al., 2019, built a Bi-directional Encoder Representation from Transformer (BERT) based model for suggestion mining. The model consists of two sentence encoders by using GloVe and CoVe embedding initialization. The BERT could deliver better results for in-domain adversarial fine-tuning but poor performance with out-of-domain samples. The authors tested by integrating BERT with a non-BERT encoder and different combinations.

In [25] Jain et al., 2020, built a transformer-based based for the suggestion mining system. To deal with the class imbalance problem, the authors applied Synthetic Minority Oversampling Technique (SMOTE) and Language Model-based Oversampling Technique (LMOTE). The oversampling techniques could boost the performance of the model marginally better. In [25] Leekha. M, Goswami, M implemented a multi-task learning approach combined with an over-sampling method for suggestion mining. An ensemble model of RCNN, CNN, and Bi-LSTM with ELMo embeddings is used in multi-task learning

RQ3 - Who are the end-users, and how does it help the end-users?

Suggestions expressed in opinion reviews may be towards the manufacturer/service provider or peer customer. It is mentioned in the literature that close observation of linguistic features of reviews can differentiate between suggestions expressed towards the manufacturer or peer customers [1]–[5], [12].

Suggestions towards manufacturers: Applications of suggestion mining can be towards suggesting quality improvements and adding new features to the product or service. The intended receiver of the advice can be the manufacturer or brand owner. In [12] Brun, C.D., & Hagège, C. (2013) considered the opinion reviews and applied NLP approaches to identify the suggestions related to product quality improvement. Jhamtani, H., Chhaya, N., Karwa, S., Varshney, D., Kedia, D., & Gupta, V. (2015) [13] applied linguistic, syntactic, and semantic feature extraction approaches and Machine Learning algorithms to Amazon reviews to detect the quality improvement related suggestions from reviews.

Suggestions towards peer customers: The recommendations expressed on platforms may be targeted towards fellow customers for better utilization and proper planning before availing of services. [3] used heuristic features, specific keywords, syntactic features, and special features related to sequential patterns with Machine Learning algorithms to extract customer-customer suggestions from reviews. In [6] Golchha et al., applied a semi-supervised deep neural network to detect the suggestions from opinion reviews towards peer customers.

RQ4 - What are the metrics used to evaluate the performance of various models?

Suggestion mining is considered a binary text classification task; the performance of any binary classifier can be evaluated in terms of Precision, Recall, Accuracy, and F-score. The binary classifier for suggestion mining classifies all the data instances, either suggestion or non-suggestion. The classifier produces four different kinds of outcomes, two are correct classification, i.e., true, and two are incorrect classification, i.e., false. The correct classification is True Positives, and True Negative and false classifications are False Positives and False Negatives. Precision can be defined as the number of true positives divided by the total instances labeled as a positive class (True positives or False Positives). Recall can be defined as the number of true positives divided by the number of instances belonging to the positive class (True positives and False negatives). Accuracy is another widely used measure for classification

Journal of Theoretical and Applied Information Technology
31st October 2022. Vol.100. No 20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

performance; it is defined as the ratio between the correctly classified instances to the total number of instances. A step that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

$$\text{Precision} = \frac{TP}{TP+FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2}$$

$$\text{F1 Score} = \frac{2*Precision*Recal}{Precision+Recall} \tag{3}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{4}$$

Where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

RQ5 - What are the challenges and research gaps identified in the current state-of-the-art method in the literature?

Suggestion mining is a problem spread across various domains such as traveling, e-commerce, social media platforms, manufacturing, and many more. The main challenge in identifying the suggestions from opinionated reviews is extracting syntactic and semantic information related to suggestions that are not shared and are simple across various domains. The other difficulty is that we do not have large labeled datasets for training and building models with the best results. The available datasets are also not balanced in nature. The challenges mentioned above make the Suggestion Mining task more critical and challenging.

For example, consider the reviews from the hotel domain ***"Please remember to tip the staff"*** is an opinion review with suggestion intention towards peer visitors for offering the tip to the hotel's staff. Consider the other review example as ***"I think guests should get to park for free"***, this review has a suggestion to the hotel authority offering free parking for the customers or guests. Consider the other example from the electronics review dataset, "With the two stores combining, can we bring built-in trial mode support to Windows Phone?" which looks like a question kind of review that has an implicit suggestion for expecting a new feature to be added. Lastly, ***"Please arabic keyboard not found in Lumia 610 please support an Arabic"*** is a review statement having explicit suggestions requesting for the addition of new features in windows phones. The reviews mentioned in the above examples from various domains and datasets illustrate the complexity of suggestion mining. Below presented a categorization of types and models based on the nature of suggestions in opinion reviews and the kinds of models applied.

Implicit Suggestion: The implicit suggestions provide the information. The suggested action or entity is not mentioned directly in the opinion review; it needs to be inferred critically from the study.

Explicit Suggestion: Explicit suggestions are suggestive text that directly proposes, recommends, or advises an action or an entity. Extraction of explicit suggestions is comparatively easy with the help of keywords and linguistic features.

Open Domain Suggestion: In building a system for suggestion mining, the model training, validation, and testing happen on the same domain dataset. For example, if a suggestion mining system for hotel reviews, the model's training, validation, and testing must be done only on hotel domain reviews.

Cross-Domain Suggestion: In building a system for suggestion mining, the model training and validation are done on one domain data, and testing happens on the other domain dataset. For example, if a suggestion mining system is for restaurant reviews, the model's training, and validation may be done in restaurant reviews, and testing has to be done on hotel domain reviews.

Research gaps Identified

- As of now, this is the first of its kind systematic study on suggestion mining from opinion reviews. So far none has presented such a detailed summary of suggestion mining from opinion reviews.
- The state-of-the-art in suggestion mining deals with only text classification.
- Any specific methods are not available to deal with implicit suggestions, which are generally complex.
- Suggestions may be specific to a particular aspect of the product or features; so far, no mechanism has been defined to handle aspect-oriented suggestion mining.

4. SYNTHESIS AND NEED FOR THE STUDY

In our systematic review conducted on suggestion mining opinion reviews, we first collected 95 articles with keywords such as Suggestion Mining and Suggestion Mining from Opinion reviews from various scholarly databases. Then, after performing inclusion and exclusion criteria on the articles collected, we finally left with 35 articles that spread the timeline from 2011 to 2022.

The following are the observations derived from the literature study.

- All the contributions made in dealing with suggestion mining from opinion reviews considered it as a text classification problem, classifying the given opinion sentence into suggestion or non-suggestion.
- All the methods applied in predicting the class label for suggestion mining are categorized into three classes such as:
  - Rule-based models used with handcrafted features
  - Approaches with Neural Network approaches trained from scratch such as RNN, LSTM, GRU, and CNN.
  - Made use of transfer learning approaches to predict the labels, which deal with problems with less training data.
- In rule-based and handcrafted feature extraction approaches, [5] modeled an ontology-based knowledge representation system to mine suggestions from opinion reviews. [12] Implemented Hidden Markov Model (HMM), SVM, RF, Decision Trees, and Naïve Bayes models are used.
- LSTM, CNN, Bi-LSTM, and GRU are popular methods with word2vec and Glove embedding initialization in neural network approaches.
- Transfer learning approaches such as ULTFiT [23] and BERT [10], [21], [24], [26] are used to handle less labeled data.
- Recently, to deal with imbalanced data, various up-sampling approaches using language modeling techniques have been proposed.

The extraction of suggestions from opinions on social platforms is having a great value-added component in various business cases. The pioneered research considered a text classification problem and applied all possible approaches that exist in the literature. We noticed the fine-grained analysis and extension text classification methods of suggestion mining are missing. With this study, we would like to bring it back attention to suggestion mining and continue the research extensions. This study gives a starting point for the researchers to understand the research contributions on suggestion mining and helps to continue the research from here on.

5. PROBLEMS AND OPEN RESEARCH ISSUES

This study on suggestion mining from opinion reviews revealed the following open-ended problems and research issues in the literature on NLP.

i. The opinion reviews on social platforms are much noisier, which needs a proper mechanism to deal with noisy data.
ii. The suggestion mining classification has a class imbalance problem, the majority of the reviews are of non-suggestion.
iii. Classifying the review as a suggestion or non-suggestion does not add much business value. Having a fine-grained analysis at the aspect level creates a good business impact.

6. CONCLUSION AND FUTURE WORK

After thorough observation and analysis, this paper draws a systematic survey report on suggestion mining in about 45 research papers. Opinion reviews on social platforms tend to have suggestions that help service providers and peer customers. Though few datasets were made available for suggestion mining, those are imbalanced and biased towards the non-suggestion class. Overall, three approaches were used: handcrafted features with supervised learning (SVM, NB, LR, RF) and rule-based methods. Then, models famous for sequence prediction tasks such as RNN, LSTM, and GRU were used with different word embedding (Word2Vec, Glove) initialization. Finally, attention and transformer models (BERT, ULMFiT) for suggestion mining have shown a significant improvement. Though transfer learning and Transformer models could produce a better result, it has been applied to only text classification. Furthermore, no mechanism has been proven that works with implicit suggestions and aspect-level suggestions. In the future, we would like to implement models that deal with implicit suggestions and aspect-oriented suggestions from opinion reviews.

REFERENCES

[1] 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.

[2] S. Negi and P. Buitelaar, "Suggestion Mining From Opinionated Text," in *Sentiment Analysis in Social Networks*, Elsevier Inc., 2017, pp. 129–139.

[3] 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.

[4] "2019 - SemEval-2019 Task 9 Suggestion Mining from Online Reviews and Forums."

[5] 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.

[6] 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.

[7] 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-.

[8] I. Markov and E. V. De La Clergerie, "INRIA at SemEval-2019 Task 9: Suggestion Mining Using SVM with Handcrafted Features." [Online]. Available: https://www.uservoice.com.

[9] "2019 - SSN-SPARKS at SemEval-2019 Task 9 Mining Suggestions from Online Reviews using Deep Learning Techniques on Augmented Data."

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

[11] 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.

[12] C. Brun and C. Hagège, "Suggestion Mining: Detecting Suggestions for Improvement in Users' Comments," *Research in Computing Science*, vol. 70, no. 1, pp. 199–209, 2013, doi: 10.13053/rcs-70-1-15.

[13] M. F. Goodchild, "Social Informatics," in *Encyclopedia of Human Geography*, 2012, vol. 9471, doi: 10.4135/9781412952422.n265.

[14] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations ofwords and phrases and their compositionality," *Advances in Neural Information Processing Systems*, pp. 1–9, 2013.

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

[16] M. E. Peters *et al.*, "Deep contextualized word representations," *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, vol. 1, pp. 2227–2237, 2018, doi: 10.18653/v1/n18-1202.

[17] P. M. Brennan, J. J. M. Loan, N. Watson, P. M. Bhatt, and P. A. Bodkin, "Pre-operative obesity does not predict poorer symptom control and quality of life after lumbar disc surgery," *British Journal of Neurosurgery*, vol. 31, no. 6, pp. 682–687, 2017, doi: 10.1080/02688697.2017.1354122.

[18] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, vol. 1, pp. 328–339, 2018, doi: 10.18653/v1/p18-1031.

[19] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," 2018.

[20] 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.

[21] 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.

[22] "2019 - Suggestion Miner at SemEval-2019 Task 9 Suggestion Detection in."

[23] S. Anand *et al.*, "Suggestion Mining from Online Reviews using ULMFiT," Apr. 2019,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[Online]. Available: http://arxiv.org/abs/1904.09076.

[24] 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.

[25] M. Jain, M. Leekha, and M. Goswami, *A Multi-Task Approach to Open Domain Suggestion Mining (Student Abstract*, vol. 1. Springer International Publishing, 2020.

[26] 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.”

[20] 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.

[21] I. Markov and E. V. De La Clergerie, “INRIA at SemEval-2019 Task 9: Suggestion Mining Using SVM with Handcrafted Features.” [Online]. Available: https://www.uservoice.com.

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

[23] Q. Zhou, Z. Zhang, H. Wu, and L. Wang, “ZQM at SemEval-2019 Task9: A Single Layer CNN Based on Pre-trained Model for Suggestion Mining.” [Online]. Available: https://github.com/google-research/bert.

[24] Laskari, Naveen Kumar, and Suresh Kumar Sanampudi. "Explainable System for Suggestion Mining using attention." *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*. Vol. 1. IEEE, 2022.

[25]. Laskari Naveen Kumar, Suresh Sanampudi, “Weighted focal loss function for evaluating effectiveness of word embeddings on suggestion mining” Nov 2021, pp: 1464-1483. Volumns-18, webology.