

Ensemble Based Temporal Weighting and Pareto Ranking (ETP) Model for Effective Root Cause Analysis

Naveen Kumar Seerangan^{1,*} and S. Vijayaragavan Shanmugam²

¹Department of Computer Science (Category-B), Bharathiar University, Coimbatore, 641046, Tamil Nadu, India

²CSE, Muthayammal Engineering College, Rasipuram, 637408, Tamil Nadu, India

*Corresponding Author: Naveen kumar Seerangan. Email: naveen.seerangan@gmail.com

Received: 30 August 2020; Accepted: 24 March 2021

Abstract: Root-cause identification plays a vital role in business decision making by providing effective future directions for the organizations. Aspect extraction and sentiment extraction plays a vital role in identifying the root-causes. This paper proposes the Ensemble based temporal weighting and pareto ranking (ETP) model for Root-cause identification. Aspect extraction is performed based on rules and is followed by opinion identification using the proposed boosted ensemble model. The obtained aspects are validated and ranked using the proposed aspect weighing scheme. Pareto-rule based aspect selection is performed as the final selection mechanism and the results are presented for business decision making. Experiments were performed with the standard five product benchmark dataset. Performances on all five product reviews indicate the effective performance of the proposed model. Comparisons are performed using three standard state-of-the-art models and effectiveness is measured in terms of F-Measure and Detection rates. The results indicate improved performances exhibited by the proposed model with an increase in F-Measure levels at 1%–15% and detection rates at 4%–24% compared to the state-of-the-art models.

Keywords: Root cause analysis; sentiment analysis; aspect extraction; ensemble modelling temporal weighting; pareto ranking

1 Introduction

The current decade has witnessed a huge growth in the users of the internet. This global development has resulted in huge amounts of information available online, especially via social media applications. According to statista, the use of social media and messaging applications grew 203% for every year since 2013, with 115% increase in the overall application users. The number of social network users worldwide, as of 2017 was found to be 2.46 Billion, and it has been estimated to reach 3.02 Billion users by 2021 [1]. It has been observed that an average individual spends ~8h a day on digital media, including social networking applications. The high usage and dependency levels brought about a huge change in the decision-making perspective of the user. It was observed that millennial often read reviews and request opinions of their peers online prior



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

to making buying decisions [2]. Currently, not just individual decisions, business decisions are also taken based on the opinions available online. This often includes shopping portals, review blogs, tweets etc. Such reviews act as valid resources for organizations to identify their shortcomings and customer satisfaction levels. However, as a major downside, it should be noted that this information is huge and is not possible to be processed by humans manually. It requires an automated information processing model that identifies the prospective root-causes to provide an aggregated simplistic view of the user requirements [3].

Root-cause is defined as the most significant entity, based on which the users have provided their opinions. Identifying the root-causes, along with their associated opinions can provide an effective view on the product or service under scrutiny. Opinion or sentiment mining is the means to automate the process of knowledge extraction from users' reviews about the products or services offered by organizations. The major goal is to determine the sentiment associated with the product, which can be positive, negative or neutral. Opinions that have been extracted usually correspond to the user's view of the aspects. An aspect is a major concept or entity, upon which the opinion is based. Aspect extraction is the process of determining the entity from the document or review. Extracting aspects from reviews and relating them to opinions will provide a clear view of the user's opinion over the entity. Root-cause is the most significant entity or entities provided by the users. Business decision making requires evidence that supports the decisions. Technically, these evidences are termed as root-causes [4,5]. Analysing these root-causes provides an effective overview of the customer's view towards the products or services hence providing an aid for business decision making. Aspect extraction and sentiment identification are the significant components of a root-cause identification system [6]. This is referred to as Aspect based sentiment analysis (ABSA) [7]. Identifying aspects and extracting their corresponding sentiments provides a clear view of the customer's opinion towards the service. These aspects along with their opinions represent the root-causes.

This paper presents an ensemble-based root-cause identification model that aids in faster and more effective identification of aspects and sentiments. The major objectives of this work are

- To provide an effective root-cause identification model for business decision making
- To determine the most significant aspects from reviews
- To accurately identify the sentiment levels associated with the aspects
- To identify and filter significant aspects to enable for easier human consumption

To achieve these objectives, the proposed work is composed of three major components, an aspect extraction component, a sentiment identification component and temporal based aspect ranking component. Aspect extraction is performed on the reviews/documents, followed by aspect filtering, where the aspects that are closer to the entity in query are alone considered for the next phase. Opinions pertaining to aspects are determined using the proposed tree based boosted ensemble model. The resultant aspects are then weighted and based on their temporal significance and Pareto-rule based filtering is applied to trim down the number of aspects for human consumption.

The paper is structured as follows. Section 2 presents the related works, Section 3 presents the root cause identification model, Section 4 presents the results and discusses them and Section 5 concludes the work.

2 Related Works

Aspect-based sentiment analysis (ABSA) has gained huge prominence with the increase in the dependence of business decisions based on customer opinions. The process of aspect-based sentiment extraction is usually performed in two phases; aspect extraction and sentiment identification. Several works exist in literature to perform these phases individually and to perform them together to perform aspect based sentiment analysis. This section discusses some of the prominent works in the literature of this domain.

The OPINE model proposed by Rana et al. [7] is one of the state-of-the-art models used to compare the efficiency of an aspect extraction model. This is an unsupervised learning model that builds features by mining the actual reviews, review evaluation parameters and their relative quality across products. The opinion mining is performed using the relaxation-labeling technique, which is a rule-based model. Hence adaptations towards new patterns are not possible. A two-fold rule-based model for aspect extraction was proposed by Rana et al. [8]. This model extracts domain independent and domain dependent features individually. It further incorporates frequency and similarity-based approaches to enhance the prediction process. A standard rule-based model that performs opinion and opinion target extraction was proposed by Singh et al. [9]. This model is based on bootstrapping. The model is referred to as Double propagation (DP), as it propagates information between opinion words and their targets. Other popular rule-based models include RubE by Liu et al. [10], automated rule selection model by Zhang et al. [11] and syntactic pattern based aspect extraction by Poria et al. [12]. However, being rule-based models, the complex relationship between patterns were not effectively extracted from these models.

A convolutional neural network based aspect based sentiment analysis model was proposed by Rokach et al. [13]. This model utilizes a 7-layer deep convolutional neural network for opinion mining. This is a combination of rule and machine learning based model, hence creating an ensemble. A similar neural network based model was proposed by Wang et al. [14]. Fusion relation embedded representation learning framework for aspect extraction was proposed by Liu et al. [15]. This model performs a fusion of semantic structures and language expression features to perform aspect extraction. Heuristics have also become part of the aspect extraction process, due to their time-constrained nature [16]. Such models include PSO based model by Lin et al. [11] and heuristic based pattern recognition models by Asghar et al. [2] and Barsalou et al. [17]. A chunk-level rule extraction model was proposed by Zhang et al. [11–17]. This is a combination of rule and machine learning models. This model divides the processing tasks into aspect level and opinion level tasks. This is a two-phase model that extracts and filters aspects based on domain requirements. Other multiple model-based methods include works by Su et al. [18], Akthar et al. [19] and Xia et al. [20]. Although these models propose to exhibit good results, the computational complexity of this model is high and hence leads to computational throttle when used on large amounts of data. Further, the existing works result in retrieving all the available aspects, leading to information outburst, as they do not consider any filtering mechanism to aid manual consumption of data.

This work aims to reduce the downsides of these approaches and to provide a compute effective model for effective identification of root-causes, and to provide consumable amount of only the most significant information.

3 Ensemble Based Temporal Weighting and Pareto Ranking (ETP) Model for Root Cause Analysis

This work uses a combination model to provide an effective architecture for root-cause analysis. The proposed architecture utilizes rules for extracting aspects and machine learning to identify their corresponding sentiments. The proposed root-cause identification architecture is composed of three broad components namely, an aspect extraction component, a sentiment identification component and temporal based aspect ranking component. The first component analyzes the textual reviews to identify its aspects, the second module results in identifying its related sentiments, while the final module ranks the aspects, hence providing an effective filtering mechanism.

3.1 Aspect Extraction

Identifying aspects forms the first phase of the proposed architecture. This section involves utilizing rules to identify aspects and reducing the extracted aspects to obtain aspects that are most relevant to the search query.

3.1.1 Rule-Based Aspect Extraction

Textual reviews constitute the base data for the proposed model. Aspects are features of the text that represents the major component around which the review has been built upon. For example, in the review “Zoom quality of the camera is excellent” camera and lenses are considered to be the aspects. Aspects can be implicit aspects and explicit [21]. Explicit aspects are components that are obvious in the review, eg. Camera, while implicit aspects are to be inferred, eg. Lenses. The proposed model concentrates on identifying explicit aspects for analysis.

First step in the process of aspect extraction is to dissect the text into individual components. This process is called tokenization [22]. Tokenization divides the text into distinct components by considering space and other domain-based symbols as delimiters [21,22]. The Part of speech (POS) tag corresponding to each of the tokens is identified. Nouns and noun phrases usually correspond to aspects. However, the process of aspect extraction cannot be considered as highly simplistic, as multiple tokens might even be aggregated to form a single aspect. For example, in the text “The World Wide Web has a great article on Ruby on Rails”, “World Wide Web” and “Ruby on Rails” are the aspects, however, they are not single words. This work proposes a rule-based algorithm that can be used to effectively determine aspects irrespective of their size.

Algorithm: AspectExtraction (text)

1. $Tokens = tokenize(text)$
 2. $Pos = POS_Tags(Tokens)$
 3. For each Token t and its corresponding Pos p
 - a. If the POS p corresponding to t is a noun or a preposition
 1. Append t to the aspect list
 - b. If t is a noun and $t - 1$ is an adjective
 1. Append $t - 1$ to the aspect list ## $t - 1$ is part of the aspect
 4. For each aspect a in aspect list
 - a. If a is a noun and $a + 1$ is not a noun
 1. Combine a and $a + 1$
 - b. If a is a combined phrase and $a + 1$ is not a noun
 1. Combine a and $a + 1$
 5. Return aspect list
-

The final aspect list contains all the aspects composed of nouns and noun phrases (multiple words representing a single entity). This process is performed for each of the reviews and aspects pertaining to the reviews are obtained.

3.1.2 Aspect Filtering

A text can contain multiple aspects. However, root-cause can only be represented by a single aspect exhibiting the highest contribution. Considering too many aspects for analysis also reduces the overall accuracy of the prediction model. Hence it becomes mandatory to identify the head word pertaining to the text. Head word represents the major aspect of a text. Sentiment pertaining to the head-word is to be identified for effective root-cause analysis. Aspect filtering is performed to maintain a balance between the domain requirements and the user. Aspects semantically similar to entities or aspects with direct or high correlation with entities are selected. The entity is the major component based on which the entire analysis is performed. For example, in the review “The food is of great quality and is also affordable,” “Restaurant” is the entity that is being spoken about and “food” is the aspect.

3.2 Boosted Ensemble Based Opinion Identification

A boosting ensemble based machine learning model [7–23] is proposed to determine the opinions pertaining to the aspects. Opinion refers to the sentiment associated with the text being analyzed. Opinions can be positive, negative or neutral. Opinion target is the component on which the opinion is based on, which is the aspect. Hence determining the sentiment pertaining to the review corresponds to the opinion of the aspect. Sentiment analysis is usually rule-based or machine-learning based. The major advantage of using machine learning models for sentiment analysis is that it can effectively handle sarcasm if trained with sufficient samples. Highly complex rules are required to handle sarcasm for rule-based models. Handling sarcasm effectively improves the overall prediction accuracy levels.

Textual reviews are the major components of the training data. However, machine learning models do not operate on textual data. They require numerical data for predictions. Hence the available textual data is converted to a numerical format. The input reviews are tokenized and word embedding is applied on the tokens. Word embedding is the process of learning features from textual components, where the tokens are mapped to vectors of real numbers. The input text is converted from one dimension to a vector space with higher dimensions. It is one of the major components when performing NLP. This work uses Continuous bag of words (CBOW) [24] to obtain word embeddings. The ensemble machine learning model is trained using these vectors for sentiment prediction.

This work performs predictions using the Tree-based boosted ensemble model. Decision trees are used as base learners. Boosting is a type of ensemble that performs iterative enhancement of solutions for each iteration by incorporating the error levels at each iteration phase [11–24].

Let $P(x)$ be the predictions obtained by training the decision tree model ($DT(x)$), and let $f(x)$ be the actual results. Hence the error levels are given by Eq. (1)

$$e = P(x) - f(x) \quad (1)$$

Boosting trains the classifier to the next level by incorporating this error level in the decision-making process. This is given by Eq. (2)

$$P(x) = DT(x) + e \quad (2)$$

The next level error e' is given by Eq. (3)

$$e' = P'(x) - f(x) \quad (3)$$

This process is iteratively performed and the errors are incorporated into the training process at each level until the error reaches a minimal threshold, defined by the domain constraints [25].

The trained model is then applied to new data to obtain their sentiment levels. Sentiments identified from this level are combined with the aspects identified in the previous phase to obtain the aspect sentiment pair that corresponds to opinion and opinion target.

3.3 Temporal Based Aspect Ranking

3.3.1 Time Window Based Aspect Weighting

Although aspects and their corresponding opinions have been identified, they need to be ranked to provide a categorized view of the results. Temporal-based ranking is applied to highlight recent and most significant aspects of the current time. A time window (tw) is maintained by the aspect repository. The aspect repository is composed of all the aspects that had been in scope until the current date. The aspect repository maintains the time and frequency of aspects. Every encountered aspect is maintained in this repository, along with its frequency of occurrence and the recent time stamp. The contents of aspect repository are organized as shown in Tab. 1

Table 1: Aspect repository

Aspect	Frequency of occurrence	Time stamp
--------	-------------------------	------------

It also maintains a time window that highlights past m days of aspects. The size of the time window is defined by the domain constraints. Aspects within the time window are provided with highest weights, while other aspects are provided with reduced weights depending on their recency.

All aspects falling under the scope of the time window are considered to be recent and significant and are provided with highest weights. Weights for other aspects are assigned based on their temporal occurrence. The process of identifying aspect weights (w_i) is given by Eq. (4)

$$W_i = \begin{cases} 1 & \text{if } I \text{ falls in the scope of } tw \\ 10^{-(tw-t_0)} & \text{otherwise } \forall 1 \leq i \leq \eta \end{cases} \quad (4)$$

where tw is the current time window, t_0 is the time of occurrence of the aspect and i is the aspect for which the weight is to be assigned.

Rank pertaining to an aspect is given by the product of their occurrence frequency and their corresponding weight. This is given by Eq. (5)

$$\text{Rank}_i = \eta_i * w_i \quad (5)$$

where η_i is the frequency pertaining to aspect i .

3.3.2 Pareto–Rule Based Ranking

Pareto's [13–25] rule states that solving 20% of the issues is sufficient to solve 80% of problems. The analysis in the domain of root-cause analysis also reveals the same. It was observed

that solving 20% of the root-causes was sufficient to alleviate 80% of the complaints [26]. Further, it should also be understood that the recommendations from this architecture are to be analysed by a human to make decisions. Human cognition is limited and hence is difficult for them to perceive all the available information. Providing limited information can actually prove to be constructive.

The aspects are segregated according to their opinion magnitudes. This provides two sets of aspects positive and negative. Neutral aspects are eliminated. The groups are individually ranked and the top 20% aspects are provided to the user for business decision making.

4 Results and Discussion

This section presents the experimental analysis, performances and performance comparison of the proposed model with several well-known schemes.

4.1 Simulation Environment and Evaluation Corpus

The proposed model has been implemented using PySpark in the spark environment. This work uses the benchmark customer review dataset [1] for evaluation. The dataset is composed of customer reviews for five different products. Details about the dataset are shown in Tab. 2.

Table 2: Dataset description

Data	Represented as	Total no. of reviews	Reviews with polarity	Reviews without polarity	No. of aspects
Canon digital camera	D1	597	238	359	237
Nikon digital camera	D2	346	160	186	174
Nokia phone	D3	546	265	282	302
Creative mp3 player	D4	1716	720	996	674
Apex DVD player	D5	740	344	396	296

4.2 Simulation Results

Performance of the proposed model is measured in terms of precision, recall, and F-Measure, as reported by Rana et al. [8].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

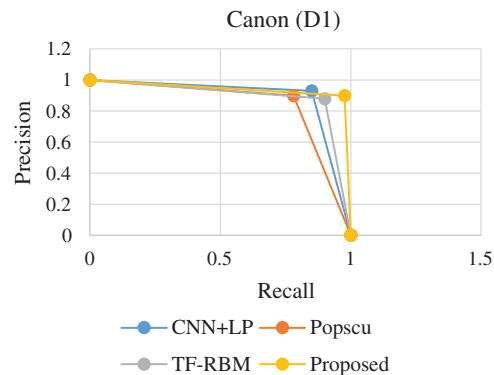
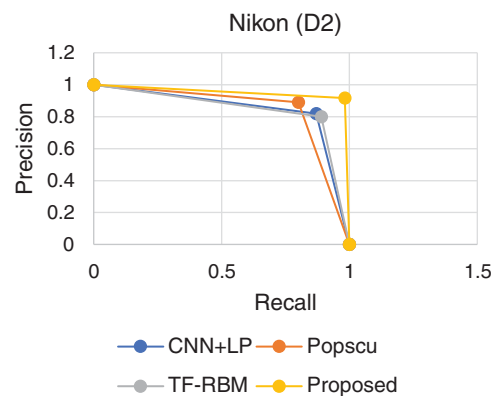
$$\text{FMeasure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

Performance of the proposed model in terms of precision, recall, and F-Measure is shown in Tab. 3. It could be observed that very high recall and F-Measure values are exhibited by the proposed model, depicting highly effective performances. The proposed model also exhibits high precision on all data except for D5, which represents reviews for Apex DVD player. This is attributed to the discrepancies existing in the reviews and insufficiency in the number of instances.

Table 3: Performance of proposed model

	Represented as	Recall	Precision	F-Measure
Canon	D1	0.98	0.90	0.92
Nikon	D2	0.98	0.92	0.95
Nokia	D3	0.99	0.88	0.93
Mp3	D4	0.99	0.84	0.91
Apex DVD	D5	0.94	0.76	0.84

PR Curve representing the performance of the proposed model is shown in Figs. 1–5. PR curve is plotted using the precision and recall values. The curve which represent the dominating curve in PR space is considered to exhibit the best performance. It could be observed that the curve corresponding to the proposed model exhibits domination in all the considered datasets. The high precision and recall levels are attributed to the effectiveness provided by the ensemble model. Boosted ensembles usually operate by iteratively reducing the error levels. This iterative reduction has resulted in effectively improving the prediction levels.

**Figure 1:** PR comparison (Canon)**Figure 2:** PR comparison (Nikon)

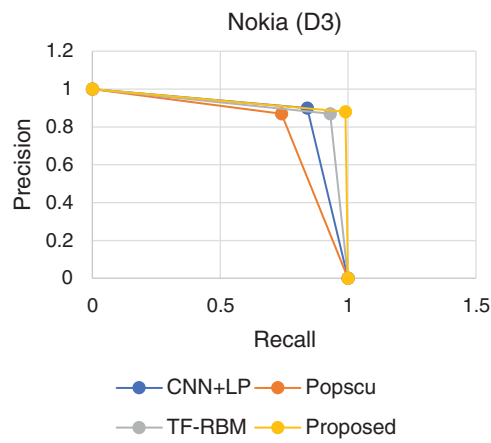


Figure 3: PR comparison (Nokia)

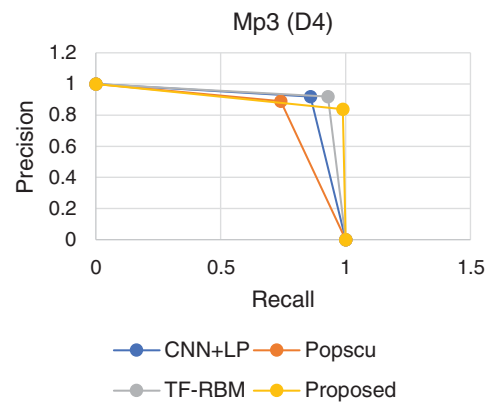


Figure 4: PR comparison (Mp3)

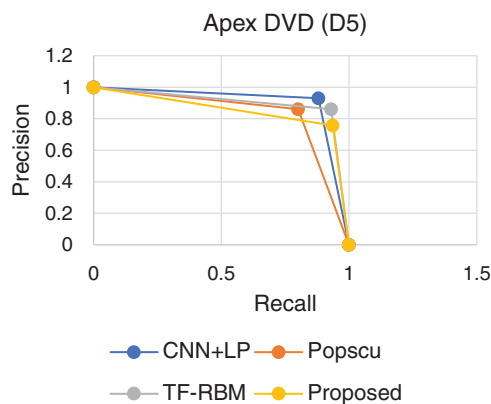


Figure 5: PR comparison (Apex DVD)

Comparisons are performed with three state-of-the-art models, proposed by Popscu, analyzed in work by Rana et al. [8] and two recent models CNN + LP proposed by Poria et al. [12] and TF-RBM proposed by Rana et al. [8].

A comparison of the F-Measure levels is shown in Fig. 6. It could be observed that the proposed model exhibits high F-Measure levels in datasets D1-D4 with increased performance levels between 1% and 15%. However, a slight reduction in performance with a reduced performance of 6% from the best result was observed in D5, which is attributed to the insufficiency of training attributes.

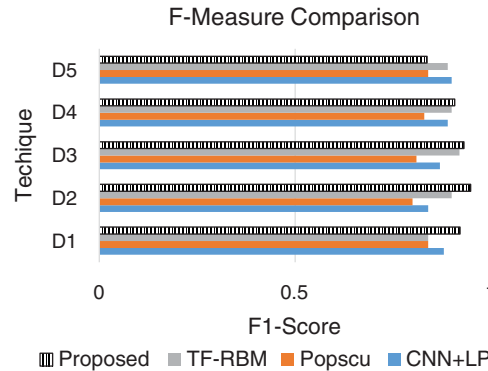


Figure 6: Comparison of F-measure levels

A comparison of the review detection rates is shown in Fig. 7. It could be observed that the proposed model exhibits high detection rates in all datasets with increased performance levels ranging between 4% and 24% when compared to the other state-of-the-art models [26–33].

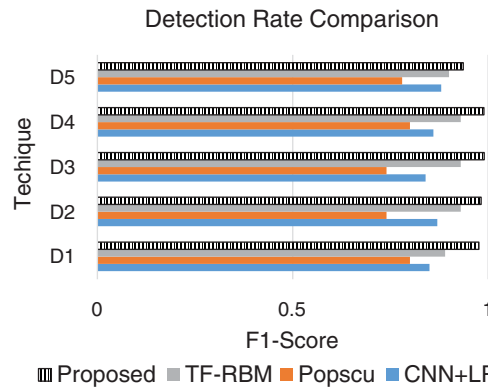


Figure 7: Comparison of the sentiment detection levels in reviews

5 Conclusion

This work proposes an Ensemble based temporal weighting and pareto ranking (ETP) for effective identification of root-causes corresponding to reviews pertaining to a product or service in an organization. The major advantage of this model is that it involves frequency and time window-based ranking of aspects, enabling recent and significant aspects to have higher ranks. This prioritizes recent items, hence causes that have been reported earlier in time are gradually pushed to lower ranks in the ranking process. This consideration of the temporal aspect enables better and timely business decision making. Experiments with state-of-the-art models in literature

shows that the proposed ETP model exhibits an increase in F-Measure levels at 1%–15% and an increase in detection rates at 4%–24%. This exhibits the efficiency of the prediction levels of ETP.

Limitations of this model include slightly reduced precision when operated upon aspects with discrepancies. Future extensions of this domain can be incorporated by creating and integrating a domain ontology into the prediction process. This can enable faster and more accurate identification of the aspects, their contexts, and the corresponding sentiments, thereby reducing discrepancies in the prediction process.

Acknowledgement: I would like to express my special thanks to parents, wife, son and friends who helped me a lot in finalizing this project within the limited time frame.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] <https://www.statista.com>.
- [2] M. Z. Asghar, A. Khan, S. R. Zahra, S. Ahmad and F. M. Kundi, "Aspect-based opinion mining framework using heuristic patterns," *Cluster Computing*, vol. 22, no. S3, pp. 7181–7199, 2017.
- [3] A. García-Pablos, M. Cuadros and G. Rigau, "W2VLDA: Almost unsupervised system for aspect based sentiment analysis," in *Expert Systems with Applications*. New York: Cornell University, 2017.
- [4] W. Maharani, D. H. Widyantoro and M. L. Khodra, "Aspect extraction in customer reviews using syntactic pattern," *Procedia Computer Science*, vol. 59, pp. 244–253, 2015.
- [5] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv: 1301.3781, vol. 1, pp. 1–12, 2013.
- [6] K. M. Vale, A. M. D. P. Canuto, F. L. Gorgônio, A. J. Lucena, C. T. Alves *et al.*, "A data stratification process for instances selection in semi-supervised learning," in *Int. Joint Conf. on Neural Networks*, Hungary, pp. 1–8, 2019.
- [7] T. A. Rana and Y. N. Cheah, "Aspect extraction in sentiment analysis: Comparative analysis and survey," *Artificial Intelligence Review*, vol. 46, no. 4, pp. 459–483, 2016.
- [8] T. A. Rana and Y. N. Cheah, "A two-fold rule-based model for aspect extraction," *Expert Systems with Applications*, vol. 89, no. 5, pp. 273–285, 2017.
- [9] V. K. Singh, R. Piryani, A. Uddin and P. Waila, "Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification," in *Int. Multi-Conf. on Automation, Computing, Communication, Control and Compressed Sensing*, India, IEEE, pp. 712–717, 2013.
- [10] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.
- [11] Z. Zhang, Y. Li, C. Wang, M. Wang, Y. Tu *et al.*, "An ensemble learning method for wireless multimedia device identification," *Security and Communication Networks*, vol. 2018, pp. 9, 2018.
- [12] S. Poria, E. Cambria and A. Gelbukh, "Aspect extraction for opinion mining with a deep convolutional neural network," *Knowledge-Based Systems*, vol. 108, no. 2, pp. 42–49, 2016.
- [13] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1, pp. 1–39, 2010.
- [14] X. Y. Wang, B. B. Zhang and H. Y. Yang, "Active SVM based relevance feedback using multiple classifiers ensemble and features reweighting," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 1, pp. 368–381, 2013.
- [15] P. Liu, S. Joty and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," in *Proc. of the Conf. on Empirical Methods in Natural Language Processing*, Portugal, pp. 1433–1443, 2015.

- [16] Y. Kang and L. Zhou, "Rule-based methods for extracting product features from [online] consumer reviews," *Information & Management*, vol. 54, no. 2, pp. 166–176, 2017.
- [17] M. A. Barsalou, *Root Cause Analysis a Step by Step Guide to Using the Right*. United States: Taylor & Francis CRC Press, 2015.
- [18] X. Su, X. Liu, J. Lin, S. He, Z. Fu *et al.*, "De-cloaking malicious activities in smartphones using http flow mining," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 6, pp. 3230–3253, 2017.
- [19] M. S. Akhtar, D. Gupta, A. Ekbal and P. Bhattacharyya, "Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis," *Knowledge-Based Systems*, vol. 125, no. 1–2, pp. 116–135, 2017.
- [20] R. Xia, C. Zong and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," *Information Sciences*, vol. 181, no. 6, pp. 1138–1152, 2011.
- [21] C. Wu, F. Wu, S. Wu, Z. Yuan and Y. Huang, "A hybrid unsupervised method for aspect term and opinion target extraction," *Knowledge-Based Systems*, vol. 148, no. 1, pp. 66–73, 2018.
- [22] J. Zhang, W. Wang, C. Lu, J. Wang and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7-8, pp. 369–379, 2019.
- [23] T. Wang, Y. Cai, H. Leung, R. Y. K. Lau, Q. Li *et al.*, "Product aspect extraction supervised with online domain knowledge," *Knowledge-Based Systems*, vol. 71, pp. 86–100, 2014.
- [24] G. Qiu, B. Liu, J. Bu and C. Chen, "Opinion word expansion and target extraction through double propagation," *Computational Linguistics*, vol. 37, no. 1, pp. 9–27, 2011.
- [25] A. M. Saeed, T. A. Rashid, A. M. Mustafa, R. A. Agha, A. S. Shamsaldin *et al.*, "An evaluation of Reber stemmer with longest match stemmer technique in Kurdish Sorani text classification," *Iran Journal of Computer Science*, vol. 1, no. 2, pp. 99–107, 2018.
- [26] N. Seerangan and V. Shanmugam, "Fast and effective root cause analysis of streaming data using in-memory processing techniques," *Indian Journal of Science and Technology*, vol. 10, no. 38, pp. 1–9, 2017.
- [27] A. Hassan, A. Abbasi and D. Zeng, "Twitter sentiment analysis: A bootstrap ensemble framework in social computing," in *Int. Conf. on IEEE*, pp. 357–364, 2013. [Online]. Available: <https://arizona.pure.elsevier.com/en/publications/twitter-sentiment-analysis-a-bootstrap-ensemble-framework>.
- [28] J. Liao, S. Wang, D. Li and X. Li, "FREERL: Fusion relation embedded representation learning framework for aspect extraction," *Knowledge-Based Systems*, vol. 135, no. 2, pp. 9–17, 2017.
- [29] B. Liu, "Web data mining," 2007. [Online]. Available: <http://sirius.cs.put.poznan.pl/~inf89721/Seminarium>.
- [30] Q. Liu, Z. Gao, B. Liu and Y. Zhang, "Automated rule selection for opinion target extraction," *Knowledge-Based Systems*, vol. 104, pp. 74–88, 2016.
- [31] Y. Yin, F. Wei, L. Dong, K. Xu, M. Zhang *et al.*, "Unsupervised word and dependency path embeddings for aspect term extraction," arXiv preprint arXiv: 1605.07843, vol. 1, pp. 1–7, 2016.
- [32] T. Craig, *The Art of Tokenization*. Canada: IBM Developerworks, 2013.
- [33] L. B. Zhang, F. Peng and M. Long, "Identifying source camera using guided image estimation and block weighted average," *Journal of Visual Communication and Image Representation*, vol. 48, pp. 471–479, 2017.