International Journal of Advanced Technology and Engineering Exploration, Vol 6(52)
ISSN (Print): 2394-5443 ISSN (Online): 2394-7454
http://dx.doi.org/10.19101/IJATEE.2019.650016

Review Article

A survey on sentimental cluster based opinion summarization in question answering community

Ankurpuri Jivapuri Goswami*
PhD Scholar, Development & Innovation Center, C. U. Shah University, Wadhwan, Surendranagar, Gujarat

Received: 07-January-2019; Revised: 19-March-2019; Accepted: 23-March-2019

Abstract
A sentiment analysis is a study which includes opinion mining, sentiment classification, and opinion summarization broadly. An opinion summarization plays an increasing research interest for automatically compressing the extensive information and generating a short summary with unlimited time. Opinion analysis is one of the emerging studies in computer domain which embrace of sentiment polarity, sentiment, opinion or semantic orientation. This paper presents the survey on sentiment analysis and summarization approaches with its challenges, methodology and pros and cons of the existing methodology. In this survey, we evaluated the research gaps of the existing technique for suggesting the new technique by the mean of applying the semi-supervised data undergo clustering; classification and summarization by means of convolutional neural network (CNN) network learning method which may use for the opinion summarization.

Keywords
Sentiment analysis, Opinion summarization, K-means clustering, Genetic algorithm, Sentiment analysis, Word embedding.

1.Introduction
Sentiment analysis is the art of presenting a short, exact, and coherent summary of an extensive text document. The sentiment analysis has emerged due to the increase in on-line publication, large internet users and the rapid growth of the electronic government (e-government). Also, the web has provided large packages of text in various topics. Basically, the text summarization is also classified into two type's namely extractive and abstractive text summarizations [1]. Extractive text summarization is selecting the words from an original text document. Abstractive text summarization (ATS) expresses to rephrase and generate the similar word, but that's not found in the original document [2]. It is improved for machine translation to summarization and also using the convolutional neural networks (CNN) and long short-term memory (LSTM) to develop the performance of text summarization [3−6]. The query-based text summarization is of great importance to analyze the various methods that are useful and provide a short summary. Two key points in query-based text summary: i.e. 1) How to choose essential content from a document that requires the question. 2) How to specify of the selected contents [7−13].

*Author for correspondence

The challenging tasks from the recent existing techniques are developed the performance on the unambiguous cases and also contain both positive sentiment and negative sentiment. In this paper, we have also focused on the technique use in existing research and try to find out the gap for developing the new emerging technique. Let us understand the terminology of summarization using the following block diagram. We have also discussed the advantage and disadvantage of the existing method. We have also found the research gaps by examining the current work and introduce a new methodology that give the solution in the area of sentimental cluster-based opinion summarization on the question answering community.

Figure 1 represents the conceptual diagram of opining mining. There are several phases in the opinion mining, which are clearer in the diagram. First, we try to extract the feature from the individual document corpus, which is known as information retrieval, including in the pre-processing phase by applying the clustering of the appropriate data. The approach of clustering can be applied for transforming data by standard data mining technique. Finally, we get some meaning full information by

extracting the pattern produce by the standard data mining.

Today there is an augmented explore awareness in opinion analysis seeing as it has turned into a model surrounded by those to give their sentiments on different characteristics of goods in blogs, check posts, and community networks. If we talk about the internet in today's world more than 51% of the user has been increases till 2018. As per recent study, more than 120 million online shopping users are increasing day by day. This much of a large corpus of internet user need the good review for the purchase of their relevant product, as well as supply and manufactures also want the accuracy and genuine feedback from the consumers to improve the product quality for the betterment of the business. This aim to the researcher to find the exact opinion from the multiple from the internet which gives the direction the Opinion summarization. For example, through Facebook review, politicians can reassess their image in the public for the upcoming election. So, developing such a system of application which can help today supplier and consumer to fulfil their demands.

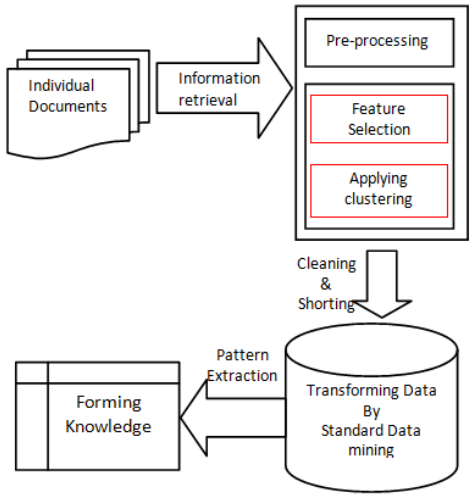

Figure 1 Basic terminology of summarization

2.Literature survey

Table 1 shows the advantage & disadvantage of the existing technique. *Table 2* shows the related literature presented with the comparison based on the features and outcome. *Table 3* shows the model outcome of the existing techniques.

Table 1 Advantage & Disadvantage of existing technique

Author details	Methods	Advantages	Disadvantages
Liu et al. [14] 2015	IncreSTS algorithm	It can incrementally update clustering results with latest incoming comments. It helps users easily and rapidly get an overview understanding of a comment stream.	It fails to target the efficiency issues Information overload problem.
Zhou et al. [15] 2016	CMiner, unsupervised label propagation algorithm, co-ranking algorithm	It does not require any manually labelled data. Low cost.	It is more challenging than micro blog sentiment classification in review texts. The trade-off between the benefit and the noise introduced by syntactic analysis is still difficult.
Jha et al. [16] 2017	Reputation system	It can be easily extended to any reviews in e-commerce domain by using language specific parser and tagger. It is more accurate in ranking the sellers.	Multi-language review mining, which itself is a challenging task.
Liu et al. [17] 2017	Recurrent neural network encoder–decoder probabilistic retrieval models	Effective for discovering meaningful questions of individual reviews. Utilizing sequence-to-sequence learning.	Problem in information retrieval human-generated questions are different.
AL-Sharuee et al. [18] 2018	ACAEC, K-means algorithm	It improves the clustering performance in term of accuracy, stability and generalizability.	It has multi-class problem based on the sentiment strength.
Huang et al. [19] 2018	Participant-based method with participant-centred social event summarization framework	It can capture all the important moments. It has a large impact in a wide range of applications.	Need more software development of applications.

Author details	Methods	Advantages	Disadvantages
Abdi et al. [20] 2018	QMOS method	It improves word coverage limit. It achieves better ARS value. It solves the problem of word mismatch.	More depth the problem of comparative sentences and sarcastic sentence handling is needed. It is not able to distinguish between an active sentence and passive sentence.
Kang et al. [21] 2018	New sentiment analysis method using ensemble TextHMM	It ample the availability and easy preparation of labelled text. It has comparative advantage over sentence without sentiment words.	It has some misclassified sentence by explicit and common sentiment words. Improve the model by sentiment lexicons.
Rudra et al. [22] 2018	ILP based summarization framework (MEDSUM)	Classify tweets into dissimilar sickness into associated group.	It faces the existence, absence, or indecision of a medical difficulty.
Zhang and Zhou [23] 2018	AQA	Low down time rate. Strong pertinence of research object. Long time span of research object massive numbers of users and reviews.	It has fake reviewer's limited information.

Table 2 Literature survey with feature and result outcome

Paper	Model used	Features	Outcome
Liu et al. [14], 2015	Two-stage approach	Query likelihood language model that retrieve the questions and recurrent neural network (RNN) encoder–decoder, a sequence-to-sequence, learning model designed to measure the answerability of questions to a product review.	Summarizing a review through questions.
Abdi et al. [20], 2018	QMOS	Identifies and extract user's query relevant sentences which contain an expression of opinion.	Extracts user's opinion from large review text.
Rudra et al. [22], 2018	low-level lexical based classifier	Uses low-level lexical class-specific features that categorize raw twitter messages.	Disease-category based summarization result.
Zhang and Zhou [23], 2017	Multi-view convolutional neural networks (MV-CNN) approach	Multi-view CNNs is developed to obtain the features of sentences and rank sentences jointly.	Generate a summary of a document.
Wu et al. [24], 2018	Knowledge graph-based approach	Statistically-based approach uses a human brain's perceptual judgment to extract text entities and relationships.	Personalized text summary is obtained.
Abdi et al. [25], 2018	Sentiment-oriented summarization of multi-documents (SOSML) approach	Employs sentiment knowledge to estimate the sentiment score for sentence classification of summarization task.	Text summarization task.
Ali et al. [26], 2018	Agglomerative clustering with hybrid TF-IDF approach	Summarized individually through sentiment-based word graph clustering.	Extract summaries from tweets.
Rautray and Balabantaray [27], 2018	Multi document summarization using Cuckoo search approach (MDSCSA)	Uses Cuckoo search meta-heuristic algorithm to summarize the document.	Extract relevant information from large documents.

Table 3 Model outcome of the existing techniques

Sr. no	Author details	Methods	Modal
1	Liu et al. [14] 2015	IncreSTS algorithm	Model a novel incremental clustering problem for comment stream summarization on SNS.
2	Zhou et al. [15] 2016	CMiner, unsupervised label propagation algorithm, co-ranking algorithm	The target opinion mining and the summarization.

Sr. no	Author details	Methods	Modal
3	Jha et al. [16] 2017	Reputation system	To solve "all good reputation" problem by evaluating reputation among all good sellers, as criteria-based reputation rating.
4	Liu et al. [17] 2017	Recurrent neural network encoder–decoder probabilistic retrieval models	To help customers to quickly capture the main idea of a lengthy product review before they read the details.
5	AL-Sharuee et al. [18] 2018	ACAEC, K means algorithm	To address the domain-dependency and the annotation cost problems in SA.
6	Huang et al. [19] 2018	Participant-based method with participant-centered social event summarization framework	Temporal mixture model to conduct sub-event detection for sports games.
7	Abdi et al. [20] 2018	QMOS method	Lexicon-based method to improve word coverage limit of the individual lexicon.
8	Kang et al. [21] 2018	New sentiment analysis method using Ensemble TextHMM	To analyze opinion and sentiment in texts based on text-based hidden Markov models.
9	Rudra et al. [22] 2018	MEDSUM	To build an automatic classification approach into different disease related.
10	Zhang and Zhou [23] 2018	AQA	To mine online users' attitudes from a huge pool of reviews.

Here we have compared with the related work in the related domain. The approaches are lacking in the use of proper data clustering like semi-supervised k-means based genetic algorithm (KGA). It is helpful in clustering the sentences into different class labels and a convolution neural network that may produce the opinion related to the rank summary of the document according to the input query which extract the opinion summary from the lengthy document based on the question to the review sentences.

3.Research challenges and analysis

In, daily lives there are lots of review on the tweeter or on Facebook which rise very crucial challenge. As we observe that there are no rules and regulation of the post on the social media on the and they are strongly shapeless, deafening, with an informal arrangement of language. We have also observed that they repeatedly surrounded by emoticons and non-dictionary words which are not in standard formation. They are also composed of number of grammatical mistaken and symbols without references, sentence structure and uppercases and lowercases, with endlessly varying conferencing. This means that it's not easy to apply a lexicon or knowledge base, which make difficult to find the relevant information from the knowledge corpus like Wikipedia. Which invokes the new challenges facing the data integration of the social media into the text analysis or into document summarization, such as alliance same meaning words features when summarizing products characteristics for opinion analysis in goods reviews. Also, for choice building statement, evaluation of huge amount of data is so difficult due to statement created by individual, and due to regional language participation on Twitter, redundancy, and irrelevant information caused by ambiguity in search keywords. There are other challenges are also identified in the language processing technology statements which are posted on the social media and other networks for the sentimental summarization. It is also important to resolve the issue of the finding the exact answer of the relevant question in the question answer communities from the give data corpus. So, there are multiple research challenges where we can identify the problem. We have listed some of the challenges which we have identified from the literature survey as the most important.

- There is the problem in the case of the efficiency issues Information overload problem.
- It is more challenging than micro blog sentiment classification in review texts.
- The trade-offs between the benefit and the noises introduced by syntactic analysis are still difficult.
- There are some other challenges in multi-language review mining, which itself is a challenging task.
- We have also found the problem in information retrieval, human-generated questions are different.

- The problem of comparative sentences and sarcastic sentence handling is needed.
- We also find the challenges that same technique is not able to distinguish between an active sentence and passive sentence.
- There are some challenges in some misclassified sentence with explicit and common sentiment words.
- We have also faced the existence, absence, or indecision of a medical difficulty.
- There is also some problem with fake reviewers' limited information.

The above mention challenges are discovered from the existing technique which many focuses on some fake reviewers, some focus on social blogs. Problem between active sentence and passive sentence. Question answer community on social networks and many more.

4.Suggested approach

The entire procedure of our suggested approach for the summarization task involves following phases: clustering, classification, and summarization. In the first phase clustering is performed for grouping the overall sentiment in the sentences with different attributes. Then ranking is executed to rank the answers according to the user's query. Finally, summarization is performed by the learning to make a comprehensive overview of the information related to the sentences expressed in the reviews. The semi-supervised learning utilizes both supervised and unsupervised data approach mostly employed in the classification task. This contains more unlabeled data during training that tends to enhance the accuracy of better machine learning model.

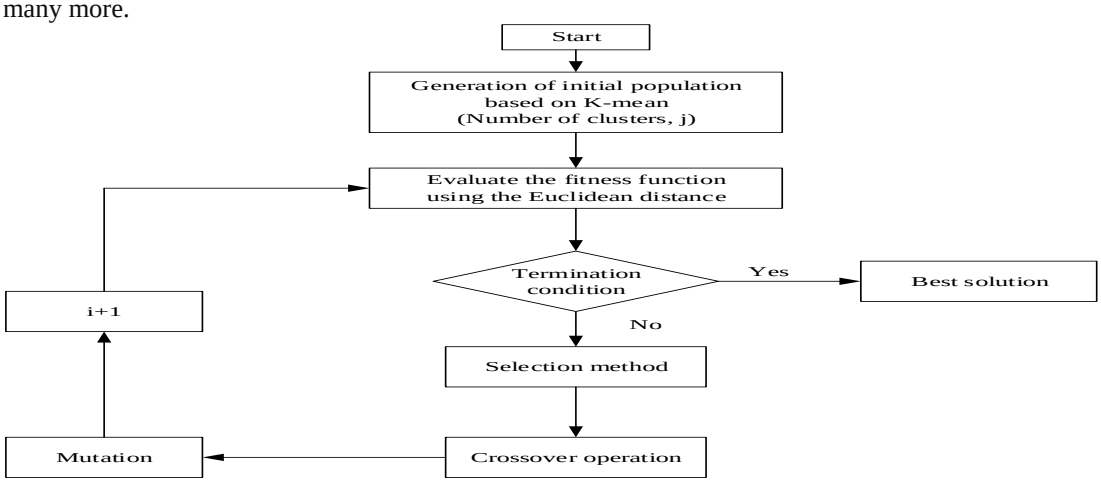

Figure 2 Flowchart for K- mean GA clustering algorithm

Figure 2 illustrates the suggested opinion summarization model. In this, the semi-supervised data undergo clustering; classification and summarization by means of CNN network learning methods are used for the opinion Summarization. It uses the following steps for the summaries of the document: Initialization, Evaluation, Selection, Crossover, Mutation, and Finally, an opinion-oriented summary is generated with different sentiment labels from the review sentences.

5.Conclusion and future work

In recent times data mining researcher is taking more interest in opinion analysis, particularly in abstractive summarization. The recent development among many researchers is the use of deep learning methodology for educating corpus of huge datasets. There are two points (deep learning and abstractive summarization) used for the future research in the area of opinion mining. In future we would like to suggest the efficient technique based on convolution CNN with the combination of GA for generating the text summarization in the area of opinion mining.

Acknowledgment

None.

Conflicts of interest

The author has no conflicts of interest to declare.

References

[1] Bhatia N, Jaiswal A. Trends in extractive and abstractive techniques in text summarization. International Journal of Computer Applications. 2015; 117(6):21-4.

[2] Moratanch N, Chitrakala S. A survey on abstractive text summarization. In international conference on

circuit, power and computing technologies 2016 (pp. 1-7). IEEE.
[3] Liu F, Flanigan J, Thomson S, Sadeh N, Smith NA. Toward abstractive summarization using semantic representations. Annual conference of the North American chapter of the ACL 2018 (pp. 1077–86). Association for Computational Linguistics.
[4] Wang L, Raghavan H, Castelli V, Florian R, Cardie C. A sentence compression based framework to query-focused multi-document summarization. Proceedings of the annual meeting of the ACL 2016 (pp. 1384–94). Association for Computational Linguistics.
[5] Wang L, Raghavan H, Cardie C, Castelli V. Query-focused opinion summarization for user-generated content. Proceedings of the international conference on computational linguistics 2014 (pp. 1660–9). Association for Computational Linguistics.
[6] Lloret E, Boldrini E, Vodolazova T, Martínez-Barco P, Muñoz R, Palomar M. A novel concept-level approach for ultra-concise opinion summarization. Expert Systems with Applications. 2015; 42(20):7148-56.
[7] Ali F, Kim EK, Kim YG. Type-2 fuzzy ontology-based opinion mining and information extraction: a proposal to automate the hotel reservation system. Applied Intelligence. 2015; 42(3):481-500.
[8] Ali F, Kwak KS, Kim YG. Opinion mining based on fuzzy domain ontology and support vector machine: a proposal to automate online review classification. Applied Soft Computing. 2016; 47:235-50.
[9] Wu H, Gu Y, Sun S, Gu X. Aspect-based opinion summarization with convolutional neural networks. In international joint conference on neural networks 2016 (pp. 3157-63). IEEE.
[10] Fang Q, Xu C, Sang J, Hossain MS, Muhammad G. Word-of-mouth understanding: entity-centric multimodal aspect-opinion mining in social media. IEEE Transactions on Multimedia. 2015; 17(12):2281-96.
[11] Somprasertsri G, Lalitrojwong P. Mining feature-opinion in online customer reviews for opinion summarization. Journal of Universal Computer Science. 2010; 16(6):938-55.
[12] Wang D, Liu Y. Opinion summarization on spontaneous conversations. Computer Speech & Language. 2015; 34(1):61-82.
[13] Yang G, Wen D, Chen NS, Sutinen E. A novel contextual topic model for multi-document summarization. Expert Systems with Applications. 2015; 42(3):1340-52.
[14] Liu CY, Chen MS, Tseng CY. IncreSTS: towards real-time incremental short text summarization on comment streams from social network services. IEEE Transactions on Knowledge and Data Engineering. 2015; 27(11):2986-3000.
[15] Zhou X, Wan X, Xiao J. CMiner: opinion extraction and summarization for Chinese microblogs. IEEE Transactions on Knowledge and Data Engineering. 2016; 28(7):1650-63.
[16] Jha V, Ramu S, Shenoy PD, Venugopal KR. Reputation systems: evaluating reputation among all good sellers. Data-Enabled Discovery and Applications. 2017; 1(8).
[17] Liu M, Fang Y, Choulos AG, Park DH, Hu X. Product review summarization through question retrieval and diversification. Information Retrieval Journal. 2017; 20(6):575-605.
[18] AL-Sharuee MT, Liu F, Pratama M. Sentiment analysis: an automatic contextual analysis and ensemble clustering approach and comparison. Data & Knowledge Engineering. 2018; 115:194-213.
[19] Huang Y, Shen C, Li T. Event summarization for sports games using twitter streams. World Wide Web. 2018; 21(3):609-27.
[20] Abdi A, Shamsuddin SM, Aliguliyev RM. QMOS: query-based multi-documents opinion-oriented summarization. Information Processing & Management. 2018; 54(2):318-38.
[21] Kang M, Ahn J, Lee K. Opinion mining using ensemble text hidden Markov models for text classification. Expert Systems with Applications. 2018; 94:218-27.
[22] Rudra K, Sharma A, Ganguly N, Imran M. Classifying and summarizing information from microblogs during epidemics. Information Systems Frontiers. 2018; 20(5):933-48.
[23] Zhang C, Zhou Q. Online investigation of users' attitudes using automatic question answering. Online Information Review. 2018; 42(3):419-35.
[24] Wu P, Zhou Q, Lei Z, Qiu W, Li X. Template oriented text summarization via knowledge graph. In international conference on audio, language and image processing (ICALIP) 2018 (pp. 79-83). IEEE.
[25] Abdi A, Shamsuddin SM, Hasan S, Piran J. Machine learning-based multi-documents sentiment-oriented summarization using linguistic treatment. Expert Systems with Applications. 2018; 109:66-85.
[26] Ali SM, Noorian Z, Bagheri E, Ding C, Al-Obeidat F. Topic and sentiment aware microblog summarization for twitter. Journal of Intelligent Information Systems. 2018:1-28.
[27] Rautray R, Balabantaray RC. An evolutionary framework for multi document summarization using Cuckoo search approach: MDSCSA. Applied Computing and Informatics. 2018; 14(2):134-44.

Mr. Ankur J Goswami started my academic carrier from 2008. Currently I am working as an Assistant Professor at Sankalchand Patel University in the Department of Computer Engineering. I have completed my masters from Ganpat University, Kherva, Gujarat and currently associated with the C.U. Shah University, Wadhwan, Surendranagar, Gujarat as a Research Scholar.
Email: ankurgoswami220882@gmail.com