ISSN No:-2456-2165

Aspect Based Sentiment Analysis: Approaches and Algorithms

SOWMYA B
ASSISTANT PROFESSOR,
DEPT OF COMPUTER SCIENCE & ENGINEERING,
EAST WEST COLLEGE OF ENGINEERING
YELAHANKA NEW TOWN, BANGALORE,
KARNATAKA 560064, INDIA

DR. T. JOHN PETER
PROFESSOR & HEAD,
DEPT OF COMPUTER SCIENCE
& ENGINEERING,SAMBHRAM INSTITUTE OF
TECHNOLOGY M.S.PALYA, BANGALORE,
KARNATAKA 560097, INDIA

Abstract:- The domain of Aspect-based Sentiment Analysis allows automatic extraction of the fine-grained sentiment information from text documents or sentences. Aspect based sentiment analysis is one among the three main types of sentiment analysis , where aspects are extracted , sentiments are analyzed and are evolved over time , is getting much attention with increasing reviews of customers and peoples on social media. The evolution of machine learning and deep learning algorithms has made noticeable mark towards aspect level sentiment analysis. This survey emphasizes a review of different works from the recent articles on aspect based sentiment analysis using several machine learning techniques.

Keywords:- *Aspect-Based Sentiment Analysis ; Fine-grained Sentiment Analysis ; Machine Learning ; Deep Learning.*

I. INTRODUCTION

Currently, sentiment analysis is a dynamic research area due to rapid growth of internet and swift participation for sharing, commenting and discussing over social media, forums, blogs and shopping websites. Sentiment Analysis plays an important role for government sectors, business platforms, manufacturers to know the impact of their products**.** The statistics of the year 2020 unleashed a jaw dropping figure of 700 million tweets in a year which can be considered as 8000 tweets per second by signifying Twitter as an active social platform. Customers and people share their feelings and opinions in the form of review for a product or service which results in a collection of huge amount of data on the Internet. This unstructured data contains lots of unused information which can be efficiently extracted through Sentiment Analysis . A piece of text can easily change the mind-set of a prospective buyer about the product or service. It is highly impossible and unwise idea to process every review comment posted by numerous customers manually. Sentiment Analysis is one of the fast-growing research areas in Natural Language Processing and is the perfect solution to analyze the trend purchase behavior. Sentiment analysis is the process of extracting opinions from a piece of text and classifying them based on the polarity like positive, negative or neutral. Sentiment analysis is exhibited at various granularities like sentence, aspect and the document as a whole.

Further Machine learning and Deep learning approaches are used to carry out the process of sentiment analysis. Hence, in this paper, we mainly focus on presenting the various tasks of aspect-based sentiment analysis that have used machine learning and deep learning approaches for the task. To carryout Aspect Based Sentiment Analysis , aspect identification is the crucial step. The aspects are classified into two types , implicit and explicit [1]. for example " The cost of room in hotel is high , but the rooms are very clean " . Here "cost" is the aspect for "hotel" and the polarity is negative towards hotel. In the above example, cost is an explicit aspect and clean is an implicit aspect. It's not easy to extract implicit and explicit aspects so some models use Rule Based Methods[9] based on machine learning concepts like semantic similarities[2] , SVM algorithm[3] , Conditional Random Fields .

II. ASPECT BASED SENTIMENT ANALYSIS (ABSA)

Aspect Based Sentiment Analysis is a fine grained technique that categorizes data by aspect and identifies the sentiment of various aspects of an entity within a textual data. An entity is a single identifiable object , it can be anything like a place, movie , hotel , an individual or a product or service. For example, let us consider a review : “I am using this laptop since one month , I love this laptop; the touch display is beautiful, but processing speed is not as expected .”, laptop is the entity and display and processing speed are the aspects. Beautiful , love , not as expected are the sentiments towards the entity and its aspect. An ABSA for a sentence of this type must give positive sentiment for the aspect ‘*display* ' , and negative sentiment for the aspect 'processing speed'.

Similar to sentiment Analysis , Aspect Based Sentiment Analysis also follows multistage analysis. Figure 1 shows the working of Aspect Based Sentiment Analysis , in which first step is to preprocess the text data to remove extraneous words.

ISSN No:-2456-2165

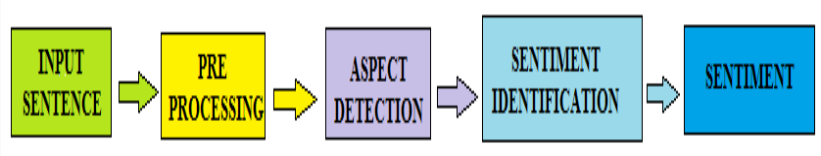

Fig. 1: Work flow of Aspect Based Sentiment Analysis

In pre-processing , the collected data must be made to a suitable format and then it can be further processed for the specific task. In Sentiment Analysis , the data will go through sequence of processing like tokenization , stop word removal , negation handling and so on to clean and convert the data into a suitable format. The process of converting the text into a series of tokens is called tokenization. It helps in vector formation and eliminating unwanted words from the text data. Negation handling is one of the important step in preprocessing , in which the actual polarity of the text will be opposite of the normal outcome , if it is not examined properly. Punctuations , special characters and numerical tokens are considered as useless and are removed , as they do not convey any sentiment or aspect in the textual data. After tokenization second step to be carried out is stemming , in which we tries to find out base word of the given word in textual data. Word embedding is carried out after the pre processing and is a very critical factor of sentiment analysis . Word embedding is the process of converting the token of words into a vector format. words like 'weather' and 'whether' sounds similar but there is a difference in meaning. word embedding's are used to make a machine understand the difference in meaning and will convert the text into some other dimension. further , the vectors will be sent to a machine learning model to extract aspect and sentiment from the word. In third level the aspects of corresponding entity of the given text are identified and further the words that define the sentiment of recognized aspects are identified. In the final step the polarity of word sentiment is identified accurately. Aspect Based Sentiment Analysis tasks are further divided into two categories , i.e., aspect category sentiment analysis and aspect term sentiment analysis [15].Aspect category sentiment analysis is the coarse-grained level of extraction and the second is the fine-grained level of extraction of the aspects. Scalability is one of the main advantages of Aspect Based Sentiment Analysis. Because at fine-grained level the textual data can be analyzed easily by Aspect Based Sentiment Analysis.

III. APPROACHES FOR ASPECT EXTRACTION

Many researchers have used different approaches for aspect extraction that includes supervised, unsupervised, semi-supervised and hybrid. supervised approach is most commonly used in wide variety of domains than unsupervised and semi supervised approaches.

A. Unsupervised rule-based approach

Poria.S [9] implemented a technique for both implicit and explicit aspect extraction from opinionated text. A rule based approach presented a detection of explicit aspects using common-sense knowledge and sentence dependency trees. Zainuddin, Nurulhuda, Ali Selamat, and Roliana Ibrahim [11] used dependency parser to identify grammatical relationship between aspects and opinions , that can be used to find implicit aspects.

B. Frequency-based or statistical approach

The implicit feature extraction can be taken place in three different steps. first is co-occurrence matrix , it will show the relationship between opinion words and product aspect. in second step the possible features set can be built and can check for presence of an implicit feature . In third step , correct implicit feature can be detected by considering the scores of possible implicit features which is calculated by checking opinion words and context information of implicit feature. Implicit aspects can be easily identified using a supervised approach[4]. They use review dataset for products and restaurant. The list and respective frequencies of implicit aspects and unique lemmas are generated first and then the score is generated i.e.,the ratio of co-occurrence of each word and the frequency of word for each implicit aspect is evaluated. The score of the aspect which is greater than the defined threshold is identified.

A classifier is used to predict the occurrence of multiple implicit features which are built using a score function. Basically the score function is based on the number of nouns , adjectives , commas and total number of 'and' words. The function parameters are estimated using logistic regression . By using the prediction of above classifier the feature detection part of an algorithm checks for one or more implicit features .

C. Pattern based approach

There is a close association between explicit and implicit aspects [10]. Instead of separately examining explicit and implicit aspect, all the correlated aspects can be represented in a network-like structure. A framework called Aspect Frame Net is proposed for sentiment analysis. The aspect pattern in review text can be read by the system and the patterns of the aspects and the sentiments for these aspects are aggregated . Further supervised and unsupervised techniques are proposed [16]. In unsupervised approach an association rule mining is applied on co-occurrence frequency data to find aspect categories.

D. Hybrid approach

Wordnet dictionary based and corpus-based methods for extraction of implicit aspect terms can be combined using Hybrid approach . Implicit Aspect Representation, learning model enhancement and implicit aspect identification are the different phases where this approach can be worked. In first phase all adjectives from the list of words are extracted and then all the relative adjectives for each aspect from training data are extracted. In second phase a Naïve Bayes classifier is trained for detecting implicit aspects which is taken using implicit aspect representation from phase 1 . In final phase

the naïve Bayes classification is tested for all terms of aspects.

IV. MACHINE LEARNING ALGORITHMS FOR ABSA

Machine Learning algorithms will work accurately on Aspect Based Sentiment Analysis approach. Following are the different machine learning algorithms and their suitability for Aspect Based Sentiment analysis.

A. Latent Dirichlet Allocation (LDA).

Latent Dirichlet Allocation is a generative statistical model that allows a set of observations to be explained as combinations of latent topics. LDA is a topic modeling technique which helps in automatically finding the fundamental topics from a given document. In LDA documents are considered as mixture of topics with words having specific probabilities. Kull-back-Leibler (KL) divergence [5] method is used for finding the association between theme model and the given paragraph. LDA model is used to the blogs theme on the other hand KL divergence method is used to identify the distance among the themes. Further a topic modeling tool named Mallet [21] is used in which LDA is considered to determine the aspect and latent information from a manually collected reviews. W2VLDA [22] is an unsupervised approach which mainly deals with multi-domain and multilingual ABSA. It uses a large amount of unlabelled data and the initial configuration can be made with minimum set of seed words.

For aspect identification and separation of opinion-words we can use combination of LDA topic modeling and an unsupervised pre-trained classification model. An Aspect Sentiment Unification Model (ASUM)[12] is designed which is a modified version of LDA that consolidates both aspect and its respective sentiment.

B. Conditional Random Field (CRF)

Conditional Random Field is a discriminative model and is used for predicting sequences. For more accurate predictions in CRF information obtained from the prior label is used. predictive framework is proposed for identifying the rating of non-rated reviews from squeal dataset [20]. A sentiCRF is a variant of CRF which is used for pair term generation and to find their sentiment scores. Cumulative logic model is introduced to predict the rating of a review which takes aspects and their respective sentiment values from the reviews. A heuristic re-sampling algorithm is proposed to solve the class imbalance problem during sentiment score estimation.

C. Support Vector Machine (SVM)

In machine learning Support Vector Machines are a promising supervised machine learning model that analyze data for classification and regression analysis. In SVM each and every data item can be plotted on an n-dimensional graph and based on the task of classification or regression a hyper-plane will be drawn.

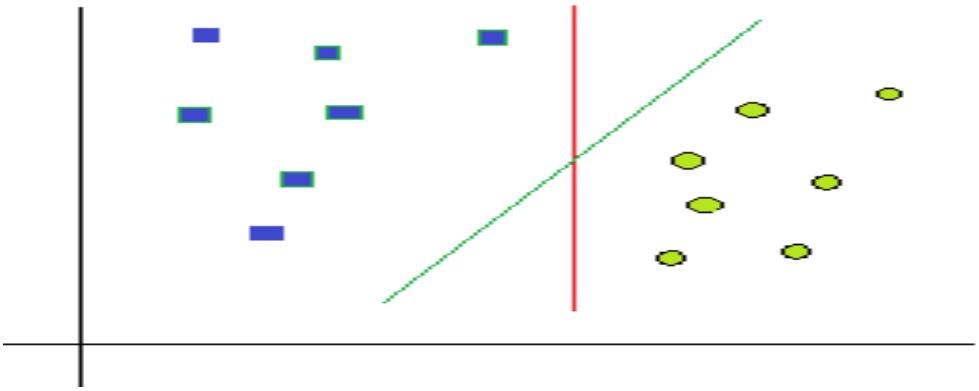

Fig. 2: Support Vector Machine Representation

For aspect term extraction on large movie reviews [7] five different feature selection algorithms that includes SVM ,Naive Bayes are compared and identified that SVM is giving the best result with the Gini index. The Gini SVM combines entropy and Kernel Based Model that normalizes classification margins and conditional probability is given directly , which is computationally less intensive . In Feature Based SVM instead of considering sentiment extraction it is better to show the importance of a feature. A hybrid sentiment classification approach[8] is proposed in which SVM is used in combination with an Association Rule Mining technique. Principal Component Analysis, Latent Sentiment Analysis etc , are feature selection methods which are applied with a combination of heuristic parts of speech and is also used for the extraction of aspects.

D. Convolutional Neural Network (CNN)

Convolutional Neural Network is a deep learning technique earlier it was used in image processing techniques and now which is used in almost all the areas. The given input to the CNN model will pass through many convolutional layers with filters in each layer called kernel . Basically kernel is smaller than image but is more in depth. Finally softmax function is used to conclude the final value to a probabilistic value, the value must be between 0 and 1. convolutional layer is the very first layer which extracts all the features from the given input. Non-Linearity problem in convolutional neural network (ConvNet) can be tackled using Rectified Linear Unit for a non-linear operation (ReLu). Compared to other classification algorithms the pre-processing required is much lower in a ConvNet. A Deep Learning method is combined with rule based method that

enhance the task of Aspect Based Sentiment Analysis and also the performance of sentiment scoring method is improved. A seven-Layer CNN has been used for the aspect identification from restaurant reviews. The Multichannel framework[17] using CNN is proposed in which every single input representation is controlled by a single CNN.

V. CONCLUSION

This survey aims to provide a comprehensive review of the Aspect Based Sentiment Analysis technique , including various tasks , approaches , algorithms and potential directions. we first set up the description of ABSA research with some elements , the definition and common modeling approaches. Then we describe each ABSA approaches with a recent advances of the compound ABSA. It is difficult to detect implicit aspect but it is very important . Due to advancement in technology , deep learning methods and other algorithms have given some promising results for ABSA. But sometimes results are not up to the expectation so more methods or algorithms are yet to be described.

REFERENCES

[1.] Schouten, K. and Frasincar, F. (2016) "Survey on aspect-level sentiment analysis" IEEE Transactions on Knowledge and Data Engineering, 28(3), pp.813-830

[2.] P. D. Turney, "Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews," in ACL, 2002, pp. 417–424.

[3.] M.Sivakumar, Dr.U.Srinivasulu Reddy, "Aspect Based Sentiment Analysis of Students Opinion using Machine Learning Techniques", Proceedings of the International Conference on Inventive Computing and Informatics (ICICI 2017) IEEE Xplore.

[4.] Schouten, K. and Frasincar, F. (2014) "Finding implicit features in consumer reviews for sentiment analysis" International Conference on Web Engineering (pp. 130-144). Springer, Cham

[5.] X. Fu, G. Liu, Y. Guo, and W. Guo, "Multi-aspect blog sentiment analysis based on LDA topic model and Hownet lexicon," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6988 LNCS, no. PART 2, pp. 131–138, 2011.

[6.] Nipuna Upeka Pannala, Chamira Priyamanthi Nawarathna, J.T.K.Jayakody, "Supervised Learning Based Approach to Aspect Based Sentiment Analysis", 2016 IEEE International Conference on Computer and Information Technology.

[7.] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. R. Venugopal, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," *World Wide Web*, vol. 20, no. 2, pp. 135–154, 2017.

[8.] N. Zainuddin, A. Selamat, and R. Ibrahim, "Hybrid sentiment classification on twitter aspect-based sentiment analysis," *Appl. Intell.*, vol. 48, no. 5, pp. 1218–1232, 2018.

[9.] Poria, S., Cambria, E., Ku, L.W., Gui, C. and Gelbukh, A. (2014) "A rule-based approach to aspect extraction from product reviews" In Proceedings of the second workshop on natural language processing for social media (SocialNLP) (pp. 28-37)

[10.] Chatterji, S., Varshney, N. and Rahul, R.K. (2017) "AspectFrameNet: a frameNet extension for analysis of sentiments around product aspects"The Journal of Supercomputing, 73(3), pp.961-972.

[11.] Zainuddin, Nurulhuda, Ali Selamat, and Roliana Ibrahim. "Hybrid sentiment classification on twitter aspect-based sentiment analysis." Applied Intelligence (2018): 1-15

[12.] Y. Jo and A. Oh, "Aspect and sentiment unification model for online review analysis," *Proc. 4th ACM Int. Conf. Web Search Data Mining, WSDM 2011*, pp. 815–824, 2011.

[13.] S. M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings," *Expert Syst. Appl.*, vol. 117, pp. 139–147, 2019.

[14.] B. Zhang, X. Xu, X. Li, X. Chen, Y. Ye, and Z. Wang, "Sentiment analysis through critic learning for optimizing convolutional neural networks with rules," *Neurocomputing*, vol. 356, pp. 21–30, 2019.

[15.] H. H. Do, P. W. C. Prasad, A. Maag, and A. Alsadoon, "Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review," *Expert Syst. Appl.*, vol. 118, pp. 272–299, 2019.

[16.] Schouten, Kim, et al. "Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data." IEEE transactions on cybernetics 48.4 (2018): 1263-1275.

[17.] D. H. Pham and A. C. Le, "Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis," *Int. J. Approx. Reason.*, vol. 103, pp. 1–10, 2018.

[18.] X. Ma, J. Zeng, L. Peng, G. Fortino, and Y. Zhang, "Modeling multi-aspects within one opinionated sentence simultaneously for aspect-level sentiment analysis," *Futur. Gener. Comput. Syst.*, vol. 93, pp. 304–311, 2019.

[19.] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Inf. Sci. (Ny).*, vol. 488, pp. 190–204, 2019.

[20.] J. Qiu, C. Liu, Y. Li, and Z. Lin, "Leveraging sentiment analysis at the aspects level to predict ratings of reviews," *Inf. Sci. (Ny).*, vol. 451–452, pp. 295–309, 2018.

[21.] N. Akhtar, N. Zubair, A. Kumar, and T. Ahmad, "Aspect based Sentiment Oriented Summarization of Hotel Reviews," *Procedia Comput. Sci.*, vol. 115, pp. 563–571, 2017.

[22.] A. García-Pablos, M. Cuadros, and G. Rigau, "W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis," *Expert Syst. Appl.*, vol. 91, pp. 127–137, 2018.