A Combined Approach for Effective Features Extraction from Online Product Reviews

D. Teja Santosh

Assistant Professor, GITAM University, Rudraram, Telangana, India

Abstract

Today E-commerce websites provide customers with the needed product information by giving various kinds of services to choose from. One such service is to allow the customer to read the end user online reviews. Online reviews contain features which are useful for the analysis in opinion mining. Converting these unstructured reviews into useful information require extracting the product features from them. Natural Language Processing (NLP) based technique extracts various kinds of product features including the low frequency features. Topic Modeling based approach also identifies specific product features from the online reviews. The effective number of product features is made available to the customer when these two approaches are combined. This results in the expanded product feature set so that the customer makes wise decisions without having to compromise on the feature set.

Index Terms: E-commerce, online reviews, opinion mining, NLP, Topic Modeling.

1 Introduction

World Wide Web contains many websites that solicit user information either on the product or person or organization. This information is useful to them to understand how people around the world have their opinion on the previously mentioned entities. This collective opinion [16] helps in improving the services offered by the websites to the customers thereby improving their businesses. Extracting useful pieces of data from these opinionated content actually called feature extraction is an important and complex task. This is because the vocabulary of the people around the world varies in different ways. Useful lexical resources namely WordNet [17] and SentiWordNet [18] and also the supervised learned corpuses help to complete this task.

* Corresponding author.
E-mail address:

There are three major ways [19] to extract product features from online reviews. They are making use of natural language rules learned from the input reviews, sequence modeling the reviews to identify the product features and to learn the topics from the underlying reviews collection. NLP based approaches often misses out low frequency features when frequency based approaches are used to extract product features. The work in [8] extracts the low frequent features using NLP feature extraction approach. Topic modeling based features extraction in [15] also identifies low frequency features but huge amount of reviews collection are to be used as input for the topic modeling algorithm. Sequence modeling the reviews additionally needs the support of lexical model to extract the product features. This helps in overcoming the frequency based limitations. Sequential modeling also requires huge amount of ground truth data for training in the feature extraction process. Leaving aside the sequential modeling approach which demands hours of time to come to a perfect training data, combining both the NLP and topic modeling based approaches improves the extracted features set as the obtained feature set from the two approaches contain frequent features, relevant features and the low frequency features. This feature set covers more number of product features hence maximizing the satisfaction of the customer with their feature set.

2   Related Work

Product features extraction from online reviews is considered as a major research work in identifying the exact product features that are useful for subsequent sentiment analysis. NLP based techniques are mainly of two kinds. These are namely frequency based methods and relation based methods. To start with frequency based methods, Minqing Hu and Bing Liu worked [1] on mining and summarizing customer reviews in which various product features were identified using Apriori Association Algorithm technique. Liu et al. used [2] NLPProcessor linguistic processor to parse the opinions and extracted product features. Association rule mining algorithm was applied to discover all frequent product features. Popescu et al. extracted [3] frequent noun phrases and identified potential features from them using OPINE system. Raju et al. extracted [4] product featured from the clustered noun phrases describing same feature. They used Attribute score formula to identify the product features. D. Teja Santosh and B. Vishnu Vardhan extracted [8] product features from online reviews using domain free approach following step by step approach. The research in relation based methods has gained its importance from the work of Liu et al. [5]. They performed Part of Speech (PoS) tagging on the online reviews, replaced the identified feature words with prototype notation and removed duplicate PoS tags. They have applied Apriori association rule mining on these pre processed reviews to identify the product features. Zhuang et al. used [6] Dependency parser called MINIPAR to identify the relation based words in a sentence. They have found out the dependency relation templates from training data and used them on the test data to extract features. Wu et al. identified [7] phrase segments in a sentence are dependent with each other. They have parsed these dependent phrases to identify product features.

Topic models like Latent Dirichlet Allocation (LDA) [20] help to extract product features from the learned topics using the underlying reviews collection. The research in this direction is progressed as the basic topic models are extended for joint modeling of product features and corresponding opinions. Intuitively topics from topic models cover features in reviews. To start with Mei at al. identified [9] the product features from online reviews by distributing the background words. Titov et al. extracted [10] product features by finding the correspondence between topics and product features. Lin et al. identified [11] product features from the topics by considering different feature distributions for each opinion word orientation. Zhan et al. identified [12] product features by chunking the reviews into opinion phrases. The topic model leverages on the relations among the words in the phrase. D. Teja Santosh and B. Vishnu Vardhan identified [15] the product features from online reviews using domain independent Feature Ontology Tree (FOT) applied on the learned topic clusters.

The combination of both NLP and topic models for extracting product features extraction has gained importance in the research community. To start with Ma et al. proposed [13] an integrative approach with LDA topic model and a synonym lexicon to extract product features. They have proposed a simple

"FeatureWordExtension" algorithm to include NLP mined features into LDA to expand the original feature set. Emitza Guzman and Walid Maalej extracted [14] product features from online reviews by supplying fine grained features obtained from NLP analysis with sentiment scores of each review to LDA model. This model with weighted average formula identifies high-level features with sentiment scores.

In the process of understanding product features extraction from online reviews, certain shortcomings were identified: Although NLP methods are very simple and effective, they produce some non-features matching the relation patterns in relation based approaches and misses out low frequency features in frequency based approaches. These techniques also require manual tuning of various parameters beforehand. Topic models perform both feature extraction and topics generation from online reviews collection at the same time in an unsupervised manner. The low frequency features are identified using the topics. To achieve this, these models need huge amount of reviews collection. The approaches discussed in [13, 14] integrated NLP obtained features with LDA model. These approaches cannot find the low frequency features with huge reviews collection available at hand as these techniques implemented synonym lexicon extension algorithm that adds synonym words to the feature list leading to repetition in [13] and used only fine grained features from collocated words to find out product features in [14]. The combined model that combine both the product features extracted from NLP based approach and LDA based approach removes the repeated features and expands the feature list so that precision of extracted features be improved. This leads to improved purchase decision on a particular product by the customer covering most of his or her product features set.

3   Combined Model for Effective Product Features Extraction

The unstructured dataset of the product reviews are to be first preprocessed for removing stop words. After preprocessing, the obtained dataset is left with only the latent product features and opinion words. NLP technique is applied on these pre processed reviews to extract product features. LDA with FOT technique is also applied on these pre processed reviews to extract product features. The obtained product features from these two techniques are merged to form distinct and expanded product features. In order to achieve this goal, a model is proposed that is composed of two main modules: NLP module and LDA+FOT module. The proposed model is presented in Fig. 1. below.
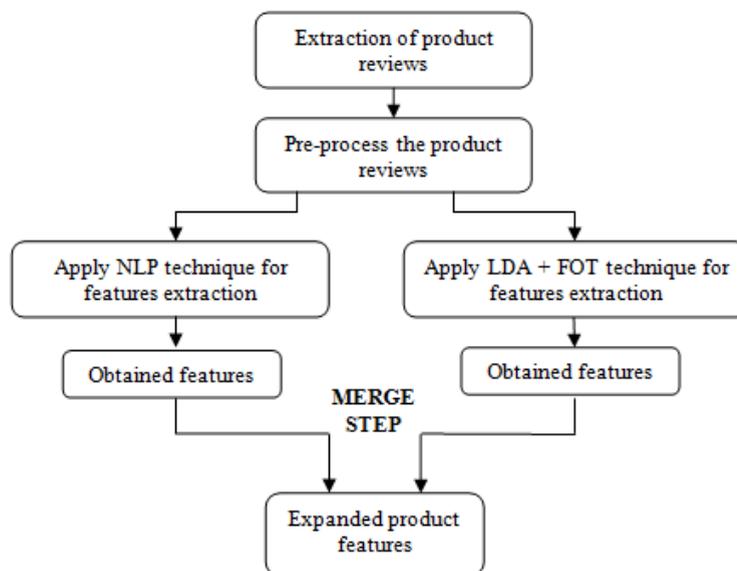
Fig.1. Proposed Model

3.1. NLP Module

Algorithm for extracting product features based on NLP approach

```
Input: Pre-processed Online Reviews {Rj}
Output: Product Feature Set f1
NLP_EXTRACT (Rj)
{
      ResFeaSet = { }, f {fr}={ }, f {rel}={ }, f {impl}={ }, f {infreq}={ }, fcf={ }, fcr={ }, fci={ }, fcinf={ }, f1={ };
      CandFeaSet = all nouns in Rj;
      CandOpSet = all adjectives in Rj;
      while (CandFeaSet ! = { })
          do for each candfea Є CandFeaSet
          f {fr} = FREQ_NOUNS(candfea)
          ResFeaSet = f {fr}
          f {rel} = RELEVANT_NOUNS(fcf)
          f {impl} = IMPLICIT_NOUNS(Rj)
          f {infreq} = INFREQUENT_NOUNS(candfea)
          ResFeaSet = f {fr} + f {rel} + f {impl} + f {infreq}
          f1 = ResFeaSet;
      write(f1);
}
String [] FREQ_NOUNS(candfea)
{
   count(candfea)= count of all nouns in Rj
  if (count(candfea) > = 3/sizeof(CandFeaSet))     /* To filter the frequent nouns */
   then  fcf = candfea;
   return fcf;
   else write("No frequent nouns were found");
}
String [] RELEVANT_NOUNS(fcf)
{
    candop = candidate opinion word present next to the identified frequent noun
  if (candfea ADJ candop)    /* ADJ is the adjacent operator that searches the nouns that are present next to the candop */
   then return candfea to fcr;
   else write("No relevant nouns were found");
}
String [] IMPLICIT_NOUNS(Rj)
{
  String iai[] = call CRF(token(Rj))    /* CRF(token(Rj)) is available as a open source tool */
          do for each token(Rj)
                if CRF_label(iai[k]) = = CRF_label(token(Rj))        /* k=0,1,2,… */
            then call SENTICNET_WEBSERVICE(Rj, iai[k])
                  return fci;
                else write("No implicit nouns were found");
}
String [] INFREQUENT_NOUNS(candfea)
{
   count(candfea)= count of all nouns in Rj
  if (count(candfea) < 3/sizeof(CandFeaSet))    /* To filter the infrequent nouns */
   then  fcinf = candfea;
   return fcinf;
   else write("No infrequent nouns were found");
}
```

The step by step procedure for extracting product features using NLP approach has been described in detail by the researchers in their work in [8]. An algorithm is developed for this approach in the current proposed approach. The algorithm for extracting NLP based product features works as follows: given the pre processed online reviews Rj, the product feature set f1 is to be generated. The sets ResFeaSet, f{fr}, f{rel}, f{impl},

f{infreq}, fcf, fcr, fci, fcinf and f1 are initialized to empty set at the beginning. The Candidate Feature Set CandFeaSet is assigned with all the nouns in Rj and the Candidate Opinion Set CandOpSet is assigned with all the adjectives in Rj. For each and every candidate feature available from the CandFeaSet, they are stored either in the frequent features set or in the relevant features set based on the statistical information given by the candidate feature from the reviews collection. The resultant feature set ResFeaSet is updated with frequent features and updated in fcf. It is also based on the feature word presence adjacent to the candidate opinion word available from the CandOpSet in the reviews collection and updated in fcr. The resultant feature set ResFeaSet is now updated with frequent features f{fr} and relevant features f{rel} respectively. The implicit nouns and the infrequent nouns are also identified. Finally, the feature set f1 contains the last updated ResFeaSet product features.

3.2. LDA and FOT Module

The procedure for extracting product features using LDA and FOT has been described in detail by the researchers in their work in [15]. An algorithm is developed for this approach in the current proposed approach. Algorithm for extracting product features based on LDA+FOT approach works as follows: given the pre processed online reviews Rj and the Feature Ontology Tree O, the product feature set f2 is to be generated. The sets ResTopics, f2, TopicProportions, ResLFeaSet are initialized to empty set at the beginning. The reviews collection is considered to generate topic proportions based on input specified topics using Dirichlet distribution. The resultant topics are learned from the topic proportions by applying multinomial distribution on the obtained topic proportions. The obtained topics are updated to ResTopics. The obtained topics contain potential product features. These are extracted from the topics by applying Feature Ontology Tree (FOT) on the generated topics. Before applying FOT, the topics are tagged with Stanford PoS Tagger to identify the noun words and these are stored in the array. The stored noun words are compared with the each and every node in the FOT tree to find a match. When a match occurs, the noun words are updated to the resultant LDA+FOT feature set ResLFeaSet. Finally, the ResLFeaSet is assigned to the product feature set f2.

Algorithm for extracting product features based on LDA+FOT approach

```
Input: Pre processed Online Reviews {Rj} and FOT tree {O}
Output: Product Feature Set f2
LDAFOT_EXTRACT (Rj, O)
{
 ResTopics = { }, f2 = { };
 TopicProportions = { };
 ResLFeaSet = { };
    for (readLine(Rj))
     {
        TopicProportions  = Dirichlet(Rj);
        ResTopics = Multinomial(TopicProportions);
     }
  for each node in Ontology O
   {
        Array[word_NN] = PoS(ResTopics);

     if (Array[word_NN] = = node[Ok]) /* kth node in ontology */
        then ResLFeaSet = word_NN;
        f2 = ResLFeaSet;
    write(f2);
    }
}
```

3.3. Merging the NLP and LDAFOT Product Features

The product features obtained from the NLP approach do not contain low frequency features and the product features obtained from the LDA+FOT approach uses huge reviews collection. The combination of product features by merging the two features set results in the availability of low frequency features from LDA+FOT approach with frequent and relevant product features obtained from NLP approach. The resulting feature set consists of distinct and expanded product feature set useful for the customer to make wise decisions. The detailed procedure expressed in algorithmic form for creating the merged feature set is presented below.

Algorithm for merging product features from both NLP and LDA+FOT approaches

```
Input: Product Feature Set fi  /* i takes values 1 and 2 */
Output: Merged Feature Set fm  /* fm = f1 U f2 */
MERGE_PRODUCT_FEATURES (f1 , f2)
{
    f1 = NLP_EXTRACT (Rj);
    f2 = LDAFOT_EXTRACT (Rj);
    fm = f1 U f2;  /* U denotes the union operator in the work. U merges the two product features sets obtained from NLP and
LDA+FOT approaches */
    write(fm);
}
```

Algorithm for merging product features works as follows: Given the product feature sets f_1 and f_2, the merged product feature set fm is to be generated. The feature set f_1 contains all the features obtained by applying NLP based algorithm on the reviews collection and the feature set f_2 contains all the features obtained by applying LDA with FOT based algorithm on the reviews collection. The merged feature set f_m contains the distinct union of both the feature set features. This provides the expanded feature set.

4   Experimental Results and Discussion

The data set that is used in the combined feature extraction approach is the collection of electronic device reviews from Amazon. The electronic device being Nokia 6610 cellular phone. The selection of the reviews was taken in such a way that each review contains explicit specification of product features. The number of selected reviews were 1500. The pre-processing of data is carried out by removing stop words and non English words. We have extracted features by using the Natural Language Processing approach specified earlier in section 3.1. The number of extracted features is 90. The Table 1 below is the population of extracted features using NLP based feature extraction model. The features populated against each kind of features set are the non-overlapped product features.

Table 1. Extracted Product Features using NLP Feature Extraction Model

Frequent Features	{phone, features, nokia, radio, battery, service, speakerphone, screen, signal reception, battery life, use, size, menu, volume, games, ringtones, headset, dialing, camera, keys, speaker, design, pc cable, weight, wap, pictures, network, calendar, buttons, bluetooth, applications, warranty, vibration, scroll, power, menus, java, jack, hardware, gsm, email, clock, call, alarm}
Relevant Features	{buy, reception, sound, sound quality, service, resolution, tones, colors, mms pictures, color screen, signal strength, battery power, information manager, visibility keys, pc software, ringer option}
Implicit Features	{look, operate}
Infrequent Features	{customer service, automatic key lock, backlight, interface, port, files, browsing, stopwatch, converter, keypad layout, gprs, gallery}

The evaluation of extracted product features from the reviews is carried out using the traditional LDA topic model. The dataset specific stop words are applied on the reviews at the time of pre-processing before learning the topics.

In order to extract maximum product features, the number of topics to be learned should be more. The number of topics 'k' to learn by the LDA model on the pre-processed reviews is specified as 50. This value for 'k' is empirically determined by analyzing the log likelihood per token on the held-out reviews collection. The log probability of held-out documents with different topics is shown in Fig. 2.
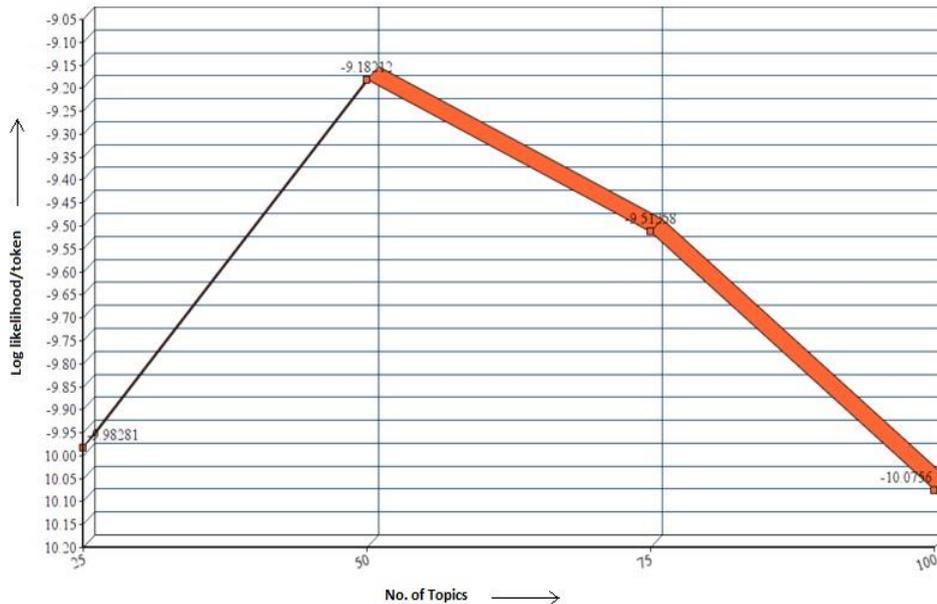
Fig.2. Line Graph of Log Likelihood on Held-out Reviews Collection

It is observed from the Fig.2 that the log likelihood per token for the number of topics has increased (in the negative axis) when the number of topics were increased from 25 to 50. Thereafter, the log likelihood per token has decreased (in the negative axis) drastically when the number of topics were increased with an increment of 25 topics from the current 50 topics to 100. This analysis specified that the number of topics to learn from the actual reviews collection is to be fixed at 50.

The parameter 'α' is set as 50/k. The numerator value of 'α' is the magnitude of the Dirichlet prior over the topic distribution of a review. The default value is 5. Whenever the number of topics to be learned is above the value of 10, the numerator value is increased. This is to learn stable topics after running on the reviews collection.

The parameter 'β' is set as 200/Wn where Wn is the total number of words in the reviews collection. The numerator value of 'β' is the magnitude of the topic assignment to the word in the reviews vocabulary. This helps in generating the coherent topics. These topics bring forward more number of product features from the reviews collection. The numerator value is based on the number of iterations the LDA model is implemented. The value varies based on the increase in the size of the dataset. These values of α and β are useful for better topic clusters generation.

It is worth noting that the extracted product features are based on the probabilities calculated by the LDA model on the features of the reviews collection. It is observed that the features found by the LDA are more general because of involving some of the non features of the product. A hierarchical FOT tree for mobile phone is developed using Apache Jena framework.

The FOT is a light weight and domain independent ontology tree developed for a class of mobile phones of different manufacturers. The concepts in the ontology are annotated with the actual product features mentioned by the manufacturer of the mobile phone. The annotated product features are considered as the instances of the concepts in FOT. The instances of FOT are compared with the words in the topic clusters. The matched product features are stored in a file for further analysis on opinions.

The number of product features extracted using the LDA with FOT approach specified earlier in section 3.2 is 63. The Table 2 below is the population of extracted features using FOT applied on topic clusters.

Table 2. Extracted Product Features using FOT Applied on Learned LDA Clusters

Extracted Features	{ringer, nokia, roaming, calling, apps, email, fm, backlight, menu, earpiece, games, infrared, gsm, speaker, hardware, network, vibrate, bluetooth, color, picture, appearance, voice, button, screen, phone, voice dialing, signal, graphics, camera, keys, headphones, battery, jack, light, sound, layout, stereo, upgrading, calendar, modem, tone, volume, speakerphone, messenger, resolution, gprs, wallpaper, music, size, keypad, plan, ringtone, zip, mms, browser, outlook, laptop, weather, scroll, weight, software, port}

The number of common product features extracted using NLP approach and LDA approach are 32. The Table 3 below is the common product features extracted using NLP and LDA+FOT approaches.

Table 3. Common Product Features from NLP and LDA+FOT Approaches

Extracted Features	{nokia, apps, email, fm, backlight, menu, games, gsm, speaker, network, vibrate, bluetooth, picture, appearance, button, screen, phone, camera, keys, battery, sound, calendar, volume, speakerphone, resolution, gprs, size, ringtone, mms, scroll, port, weight}

To improve the number of features the two product features are combined using the algorithm specified in section 3.3. The number of product features in the expanded feature set after applying the union operator is 104. The percentage of uncommon product features available before and after finding the common features from the two feature extraction approaches are tabulated in Table 4 below.

Table 4. Percentage of Uncommon Product Features

Approach	Percentage of uncommon product features	No. of uncommon product features
NLP feature extraction approach	64.5%	58 out of 90
LDA+FOT feature extraction approach	49.3%	31 out of 63
NLP & LDA+FOT extracted features	69.2%	72 out of 104

The percentages presented in the above table specifies clearly that by merging the NLP and LDA+FOT product feature extraction approaches, maximum number of uncommon product features are made available to the customer which improves the customer confidence in the reviews search process towards making final purchase decision. The comparison of information retrieval metrics among the works of Ma et al. in [13], Emitza Guzman and Walid Maalej in [14] and the work proposed by the researchers in this paper is tabulated in Table 5 below.

Table 5. Comparison of Information Retrieval Metrics

Approach	Precision	Recall	F-measure
F_{NS}	0.601	0.506	0.549
LDA_SYN	0.751	0.732	0.739
NLP+LDA+FOT	**0.865**	**0.775**	**0.818**

The corresponding precision-recall graph is plotted and illustrated in Fig 3. below.
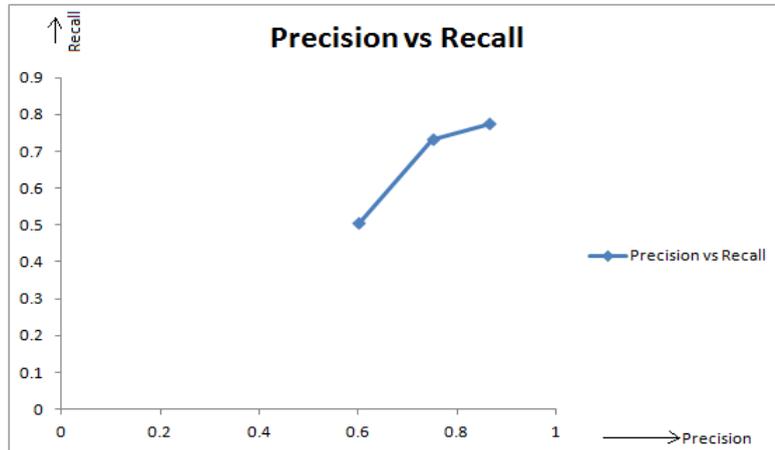
Fig.3. Precision-Recall Curve

The precision results from the above summarized table and from the figure indicate that our approach as compared with F_{NS} approach has increased drastically to 26%. This is because the frequent word collocations are considered for fine grained feature extraction. The proposed approach as compared with LDA_SYN approach has increased up to 11% as both the product features are combined after their extraction from the NLP and LDA+FOT approaches.

The recall results from the above summarized table indicate that our approach as compared with F_{NS} approach has increased drastically to 27%. This is because the frequent word collocations that are with the considered domain are only used for fine grained feature extraction. The proposed approach as compared with LDA_SYN approach has increased slightly to 4% due to the addition of synonym lexicon to LDA features in LDA_SYN approach.

By achieving the precision of 86.5% with respect to the features extraction on the considered dataset, it is concluded that the combined feature extraction approach performs better than the individual features extraction approach in the semantic environment.

5   Conclusion and Future Work

The combined feature extraction using the obtained features from both NLP and LDA+FOT approaches was carried out successfully. The algorithms for product features extraction using NLP based approach and LDA+FOT based approach were developed and implemented. The merging algorithm for the two product features also has been developed and implemented. The merging algorithm found to extract maximum product features when compared with the peer models.

In future, the obtained features from the current work and on the similar products are used as experiential case bases in the product search in E-Commerce. These case bases are useful in determining the similarity between the searched product and the stored case bases at the time of product recommendations using case based reasoning approach.

References

[1]   Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews."*Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.

[2]   Hu, Minqing, and Bing Liu. "Mining opinion features in customer reviews."*AAAI*. Vol. 4. No. 4. 2004.

[3]   Popescu, Ana-Maria, and Orena Etzioni. "Extracting product features and opinions from reviews." *Natural language processing and text mining*. Springer London, 2007. 9-28.

[4]   Raju, Santosh, Prasad Pingali, and Vasudeva Varma. "An unsupervised approach to product attribute extraction." *Advances in Information Retrieval*. Springer Berlin Heidelberg, 2009. 796-800.

[5]   Liu, Bing, Minqing Hu, and Junsheng Cheng. "Opinion observer: analyzing and comparing opinions on the web." *Proceedings of the 14th international conference on World Wide Web*. ACM, 2005.

[6]   Zhuang, L., F. Jing, and X. Zhu. Movie review mining and summarization. In Proceedings of ACM International Conference on Information and Knowledge Management (CIKM-2006), 2006.

[7]   Wu, Yuanbin, et al. "Phrase dependency parsing for opinion mining."*Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*. Association for Computational Linguistics, 2009.

[8]   Santosh, D. Teja, and B. Vishnu Vardhan. "OBTAINING FEATURE-AND SENTIMENT-BASED LINKED INSTANCE RDF DATA FROM UNSTRUCTURED REVIEWS USING ONTOLOGY-BASED MACHINE LEARNING." *International Journal of Technology (2015) 2: 198* 206.

[9]   Mei, Qiaozhu, et al. "Topic sentiment mixture: modeling facets and opinions in weblogs." *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007.

[10] Titov, Ivan, and Ryan McDonald. "Modeling online reviews with multi-grain topic models." *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008.

[11] Lin, Chenghua, and Yulan He. "Joint sentiment/topic model for sentiment analysis." *Proceedings of the 18th ACM conference on Information and knowledge management*. ACM, 2009.

[12] Zhan, Tian-Jie, and Chun-Hung Li. "Semantic dependent word pairs generative model for fine-grained product feature mining." *Advances in Knowledge Discovery and Data Mining*. Springer Berlin Heidelberg, 2011. 460-475.

[13] Ma, Baizhang, et al. "An LDA and Synonym Lexicon Based Approach to Product Feature Extraction from Online Consumer Product Reviews."*Journal of Electronic Commerce Research* 14.4 (2013): 304.

[14] Guzman, Emitza, and Wiem Maalej. "How do users like this feature? a fine grained sentiment analysis of app reviews." *Requirements Engineering Conference (RE), 2014 IEEE 22nd International*. IEEE, 2014.

[15] D. Teja, B. Vishnu Vardhan, and D. Ramesh. "Extracting Product Features from Reviews Using Feature Ontology Tree Applied on LDA Topic Clusters." *Advanced Computing (IACC), 2016 IEEE 6th International Conference on*. IEEE, 2016.

[16] Liu, Bing. "Opinion mining." *Encyclopedia of Database Systems*. Springer US, 2009. 1986-1990.

[17] Christiane Fellbaum (1998), *WordNet: An Electronic Lexical Database. Bradford Books*.

[18] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani, *SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining*, In *LREC*, vol. 10, pp. 2200-2204. 2010.

[19] Zhang, Lei, and Bing Liu. "Aspect and entity extraction for opinion mining."*Data mining and knowledge discovery for big data*. Springer Berlin Heidelberg, 2014. 1-40.

[20] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *the Journal of machine Learning research* 3 (2003): 993-1022.

Authors Profiles

D. Teja Santosh is presently working as Assistant Professor in CSE at GITAM University Hyderabad. He has submitted his research in the faculty of Computer Science and Engineering, JNTUK, Kakinada on 3rd June, 2017. His areas of interest are data mining, machine learning, natural language processing. He was the Additional Reviewer for 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI) in Kerala, India. He is the member of professional bodies of international repute namely IEEE, ACM, IAENG. He is one among the 16 research scholars of Dr Bulusu Vishnu Vardhan.