Aspect Category Detection Using Multi Label Multi Class Support Vector Machine With Semantic And Lexical Features

Mr. Mannava Yesu Babu, Research Scholar, Department of Computer Science and Engineering, JNTUAnanthapur, AP, India,
Dr. P. Vijaya Pal Reddy, Head & Professor, Department of Computer Science and Engineering,Matrusri Engineering College, Saidabad ,Hyderabad, Telangana, India,
Dr. C. Shoba Bindu, Professor, Department of Computer Science and Engineering, JNTU College of Engineering, Anantapur, AP, India, ,

Abstract---Aspect Based Sentiment Analysis (ABSA) is a type of a Sentiment Analysis (SA) technique which is used to determine the aspects of the entities and to identify the sentiment of each aspect. Aspect Category detection is one of the sub task in ABSA. For a given set of reviews and a set of predefined aspect categories, the task of Aspect Category detection is to determine the aspect categories which are discussed in each review. Aspect categories are typically coarser than the aspects and these are not necessarily occurring in the sentences as terms. In this paper, it is to investigate the effectiveness of linguistic, Lexicon, vector based and Semantic features for the task of Aspect Category Detection (ACD). Vector based features capture the sentiment and semantic information. In this experiment, the effectiveness of vector based features over text based features is addressed. The deep learning techniques are used to generate vector based features over word based features. The effectiveness of the proposed system is measured using recall, precision and F1-scores. The proposed system has achieved promising results compared with state-of-the-art techniques of ABSA.

Keywords--- Aspect category detection, ABSA, Semantic features, Vector based features.

Introduction

Sentiment Analysis systems provide suggestions for products that are likely to be liked by a particular user. In order to find the preferred products for a user, Sentiment Analysis systems predict or compare the utility of products before providing a list of ranks for the recommended items to users [1] . The reviews of products are a feedback form of a consumer where consumers express opinions about aspects of a product. In the context of product reviews and product recommendation, the term aspect denotes both components and the characteristics of a product. The reviewer expresses conflicting opinions about products aspects such as positive, negative, neutral or conflict sentiments. Such fine-grained opinions are crucial where in they mention the preferences of the consumers which drive their decisions on purchases and also influence the behavior of the Sentiment systems. However, in practice, not all users can be expected to rate products after a purchase and highly priced products tend to receive limited reviews from a single user [2,3]. To overcome this limitation, additional sources of social knowledge such as users purchase preferences are required to further improve recommendation performance.

There has been a steady increase in Sentiment Analysis systems research which is evidenced by the growth of research papers in this area in the last decade. Despite the existence of research work in exploiting product reviews and implicit feedback to recommend products, social recommender systems remain an open research field. Linguistic nuances that are caused by the informal nature of social media text make it challenging to extract aspects of products automatically from reviews and assess their sentiment value. Furthermore, the available forms of implicit feedback are abundant but not knowledge-rich. Therefore, strategies to utilize both knowledge sources to improve recommendation performance are needed.

In this paper, we are focused on a more granular approach for analyzing the sentiments which are captured in user generated restaurant reviews. Number of comments and reviews on an entity is much larger than our reading ability. Every product or restaurant has thousands of reviews where the customers specify their opinion on it. Several platforms such as Yelp and Amazon are trying to display opinions of customers to users by developing better ways. One of the famous technique for display opinions is summarizing the total information of entity by considering similar phrases with high frequency in the reviews over the high ratings of reviews. The limitation of these methods is the loss of important information by summarization techniques. Instead of applying ABSA on this large number of reviews, the representative models can develop easily for aspects of the entity and the sentiments of the aspects. In ABSA technique, the aspects are extracted automatically from the reviews of

DOI: 10.5373/JARDCS/V12I6/S20201180
**Corresponding Author: Mannava Yesu Babu, E-mail: mannavababu@gmail.com*
Article History: Received: April 02, 2020, Accepted: July 03, 2020

customers and define the thoughts of the reviewers about the sentiments of the aspects. Furthermore, we aggregate the information of specific aspect level from all reviews to develop an understanding of how people recognizing a particular entity. The application of ABSA techniques is extremely beneficial to the world of online opinion spaces in terms of extracting the useful information from those large datasets. The ABSA helps the user to quickly understand the aggregate sentiment about a particular entity and also helpful to understand the online perception of entities and the drivers behind them.

The SemEval 2014 competition addressed the ACD as a sub task of ABSA. In Aspect Category Detection, every sentence is tagged with aspect categories which are mentioned in the sentence. The FOOD, AMBIENCE, PRICE and SERVICE are examples of aspect categories. The sentences which are not contain the previous aspect categories belongs to the aspect category of ANECDOTES / MISCELLANEOUS. For example the sentence "The restaurant was expensive, but the menu was great" contains the aspect categories of FOOD and PRICE.

In section 2, the related work for sentiment identification, aspect terms extraction, identification of aspect term category and aspect term and aspect category sentiment polarity is presented. In section 3, various types of features used to train the machine learning algorithms to extract aspect categories for the reviews in the test dataset are explained. The dataset description, evaluation measures and comparison of results on various machine learning techniques and its analysis is presented in section 4. The section 5 presents the conclusions of this work and possible future direction to the proposed work.

Related Work

ABSA is a task that determines the orientation of a expressed sentiment on each aspect in a sentence [4]. Opinions can be expressed on any product, service or person. The target of an opinion is referred to as an entity. An entity can have a set of aspects. For example, iPhone is an entity that has a set of aspects such as battery and screen. In Liu, 2015, the aspects are often referred to as opinion targets or product features in the research area of sentiment analysis [5] . An opinion is a negative or positive view about an aspect or an entity expressed by an opinion holder. The positivity, negativity and neutrality characterises opinion orientation (or sentiment polarity in sentiment analysis literature) whereby no opinion is considered neutral sentiment. An aspect can specified explicitly in a review or implied through other expressions (implicitly). In contrast, implicit aspect expressions often identified through adjectives [6] . The ABSA is divided into two main tasks such as extraction of sentiments and sentiment classification.Here, the aspect extraction task is organized into three different classes such as dependency relations model, frequent noun approach and supervised learning. The frequent noun approach determines the aspects of a product which are expressed by noun phrases and nouns from a large dataset of product reviews.

Popescu and Etzioni (2007) extracted a noun phrases and nouns list from the product reviews dataset and thereafter prunes this list based on a threshold of frequency [7] . The remaining aspects in the list are evaluated by using the Pairwise Mutual Information (PMI) among the candidate aspect and associated extractor pattern. Rana and Cheah (2017) uses POS patterns to identify product aspects [8] . Noun adjective allows the identification and extraction of the associated noun. sentence analysis that does not rely on sentiment lexicon to find product aspects, is to use a dependency parser to determine the semantic relationships among words. This is more likely to produce accurately the opinion phrases than the methods that consider the proximity of words alone [9]. The propagation method proposed by Qiu et al. (2011) to determine all possible aspects and sentiments [10]. They used the sentiment lexicon of Hu and Liu (2004) [11] to start the extraction of small set of seed sentiment words. Kang and Zhou (2017) developed a method to extract many noun phrases or nouns that are not aspects and therefore does not scale well to large datasets [12]. This is because during propagation, adjectives that are not opinionated will be extracted as opinion words. Sentiment lexicons such as SentiWordNet and Hu and Liu's sentiment lexicon are commonly used to identify aspects. Poria et al. (2014) developed a rule-based aspect extraction algorithm called SenticNet aspect parser using SenticNet as the sentiment lexicon to extract aspects [13].

All methods of aspect detection method are divided into syntax-based (relation-based methods), unsupervised machine learning, supervised machine learning, frequency-based and hybrid approaches. Zhao et al. (2010)used tree kernel function to propose a generalization step for extracting the syntactic patterns [14] . These syntactic patterns extract all the annotated aspects for a given labeled dataset. The syntax trees are generated for all sentences of the unseen data. Jakob and Gurevych (2010) used a linear chain Conditional Random Field (CRF) to solve the common labeling problem of aspect detection [15]. CRF commonly used for processing the whole sequence of words in natural language processing. Lakkaraju et al. (2011) combined Hidden Markov Model (HMM) with Latent Dirichlet Allocation (LDA) to differentiate among background words and aspect words [16]. Hai et al. (2014) identified the helpfulness of reviews at aspect level by proposing a supervised joint sentiment and aspect model [17]. The proposed model is similar to supervised LDA. Marcheggiani et al. (2014) proposed a CRF model which is capable of dealing with multiple aspects in a sentence [18]. Furthermore, when the same sentence contain multiple aspects, it is likely that they control each other through certain discourse elements which influence the sentiment score of each aspect.

Sauper and Barzilay (2013) presented a probabilistic model that performs sentiment analysis and aspect detection jointly for the domain of reviews of restaurants and performs aspect detection only for the medical domain [19]. Mukherjee and Liu, 2012 combines MaxEnt classifier with LDA model [20]. The MaxEnt classifier is used to optimize the word priors which influence the process of generating the drawing words. Rana and Cheah (2017) uses POS patterns to identify product aspects [8] . A common pattern such as noun adjective allows the identification and extraction of the associated noun. Recent work using deep neural networks show performance improvement on aspect extraction [21, 22]. However, supervised approaches require annotated training data from social media text whichis often difficult to acquire. Sentiment Analysis algorithms targeting real-world datasets favour unsupervised approaches over supervised approaches. Besides supervised machine learning methods, topic modelling has been a popular approach in identifying implicit topics and sentiment words [5]. In [23] extracted aspects of a product by using a combination of shallow NLP and frequent noun approach to build a product representation.

Proposed System Description

Our approach to the aspect category detection task is based on supervised Machine Learning. Our system combines a large number of features to achieve competitive results. ACD is formulated as a multi- label multi-class classification task on the review-level, for every review we label with one or more of the categories. Sentences are first analyzed by the Stanford tokenizer, POS-tagger, and dependency tree extractor. Then, the combination of pre-processed data and representations of words are used to determine the task-specific features as presented in the Figure 1.

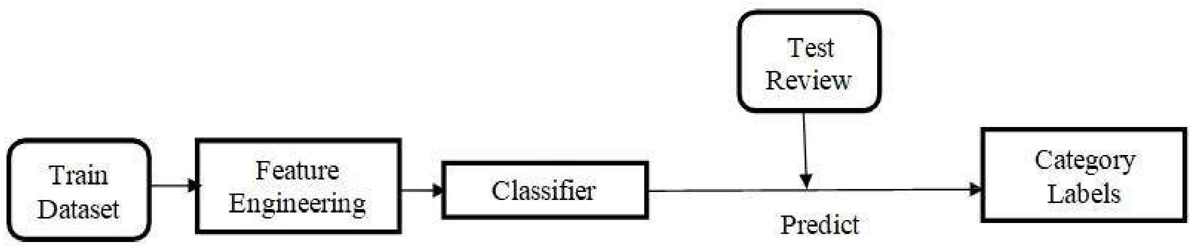


Figure 1: The Proposed System for Aspect Category Detection

The various types of features to represent the review sentence for learning the model by the classifier are presented below:

I. Linguistic Features

- Word N-grams: The count of word unigrams and bigrams in the sentence.
- TF-IDF: Term Frequency and Inverse Document Frequency of a word computed from the training data.

II. Semantic Features

To capture the semantic relations exist among the words present in the corpus, the following features are used.

- Topic modeling : The LDA could generate the document distribution among predefined topics.
- Word2Vec: We used the publicly available word vectors generated through word2vec model with dimensionality of 300, which uses Yelp restaurant dataset for training.
- WordNet: WordNet is used to match the hypernym tree of every word in a sentence with the four categories such as price, service, ambience and food. If the hypernym tree doesn't contain any of such categories of words, we check the next level words hypernym tree which are derived from the previous word hypernym. The frequency count of each category is listed as a feature. If these four categories are not matched in the hypernym tree, It is considered as anecdotes/ miscellaneous category.

Lexicon Features

These features are generated using Yelp Restaurant review corpus. The sentiment score of every word in the corpus is calculated using Point-wise Mutual Information (PMI).

- Frequency count: count of words with negative and positive scores.
- Polarity score: The overall score of a sentence (all the words sentiment scores are summed in the sentence).
- Word2Aspect: The association of each word in a sentence with one or more categories. The level of associate is

DOI: 10.5373/JARDCS/V12I6/S20201180
*Corresponding Author: Mannava Yesu Babu, E-mail: mannavababu@gmail.com
Article History: Received: April 02, 2020, Accepted: July 03, 2020

calculated using PMI.

Word Vector features

We used the skip-gram model of Word2Vec to compute the vector representations of words by using a Yelp dataset of restaurant reviews with 300-dimensional vector space. The following features are derived from the word vectors.

- Word Vector Average (WVA): It is attained by averaging the word vector representations in the sentence. It is computed as follows:

$$\text{WVA}=\frac{\frac{1}{N}\sum_{i=1}^{N} v_i}{|\frac{1}{N}\sum_{i-1}^{N} v_i|}$$

Where, vi is the vector representation of the ith word in the sentence, N is the words count. The adverbs, adjectives, verbs and nouns are used to calculate the WVA.

- Vec2Cat: Similarity between vector and category is computed. Initially, identify a set of words for each category. Then, determine the cosine similarity among the word vectors of the input sentence and the vectors of seed words. The maximum cosine similarity among word vectors of the sentence and seed words is considered as a feature.

Results and Discussions

Dataset Description

The training data of restaurants dataset contains 3041 English sentences which includes the annotations for aspect categories and overall sentence polarities. Additional 800 restaurant reviews were collected and annotated and used as test data. The dataset characteristics are presented in Table 1.

Table 1: Reviews in Restaurant Train and Test Dataset

Domain	Train	Test	Total
Restaurant	3041	800	3841

Table 2 presents the aspect categories distribution in both training and testing datasets. The FOOD category is the dominant aspect category in both test and training restaurant sentences and the major polarity class is ‘positive’

Table 2: Aspect categories distribution in Train and Test datasets

Category	Train	Test	Total
FOOD	1232	418	1650
PRICE	321	83	404
SERVICE	597	172	769
AMBIENCE	431	118	549
MISCELLANEOUS	1132	234	1366
Total	3713	1025	4738

In this dataset, the predefined aspect categories are SERVICE, FOOD, AMBIENCE, PRICE AND ANECDOTES/MISCELLANEOUS. The distributions of these categories over the dataset are displayed in table 3. Table 4 displays the percentage of the distribution of number of categories over reviews. Most of the reviews only discuss a single category of interest which makes sense.

Table 3: Number of reviews that contain multiple categories in each subset

#Cats	Train	Test
1	2465	611
2	486	155
3	84	32
4	6	2

Table 4: Percentage of reviews that contain multiple categories in each subset

#Cats	Train(%)	Test(%)
1	81%	76%
2	15%	19%
3	2%	4%
4	0.1%	0.2%

Evaluation Measures

The F1 measure is computed as

$$F1 = \frac{2 * P * R}{P + R}$$

Where, R is a Recall and P is a Precision. The P and R is defines as

$$R = \frac{|S \cap G|}{|G|}$$

$$P = \frac{|S \cap G|}{|S|}$$

Here, G is the set of correctly annotated aspect categories and S is the set of aspect category annotations that a system returned from all the test sentences.

Empirical Evaluations

Since a sentence can contain several categories, we employ multi-label one-vs-all Support Vector Machines (SVMs). We treat this subtask for the sentence. Each sentence can have multiple categories. A 10-fold cross validation is performed on training dataset with default C value.

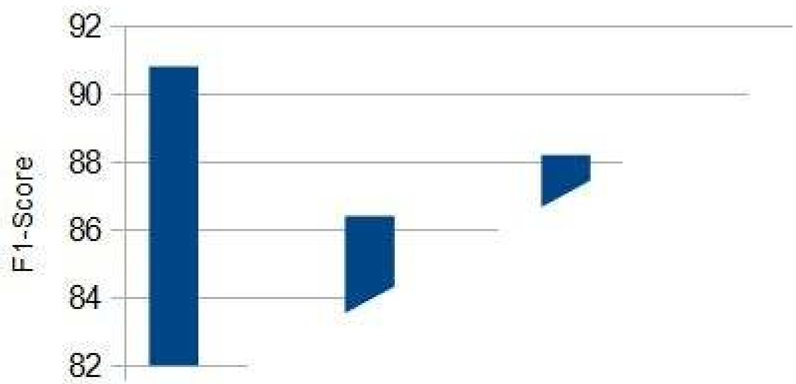
J48 is an implementation of decision tree algorithm which builds on the concept of Information entropy. The algorithm use information gain measure to choose the most informative features. Naive Bayes (NB) classifier works based on Bayes rule. IBk is an implementation of K-Nearest Neighbor (KNN) classifier. It computes the distance between the feature vectors of the unseen example and the training examples feature vectors. The label of the unseen example is based on the k nearest neighbors training examples. The Weka 2.7 implementation of J48, IBk and Naive Bayes were used with 5 as the k value. The performance of the classifiers with the linguistic, semantic, lexicon and word vector features are evaluated on restaurant dataset with 10-fold cross-validation and the results are displayed in Table 5. We allocate anecdotes/ miscellaneous category for the words which are not belonging to any of the four categories.

From the results, it is found that the semantic features improve the performance of the system7.5 % for multi-class multi label SVM implementation. The impact of lexicon based features increases the F1 score by 8.8 %. The vector based features impact is increases the performance of the system compared with text based features is 18.1%. The addition of vector based features to the lexicon features, Semantic based features and linguistic features increases the F1-score by 1.8%. The performance of the decision Tree implementation is also comparable with SVM implementation. The performance of K- Nearest Neighbor implementation is low compared with the remaining three algorithms.

Table 5: Results for the aspect category detection on Restaurant dataset

Classifier	Features	Precision	Recall	F1-Score
Support Vector Machine	Linguistic	78.3	67.8	72.7
	+ Semantic	84.5	76.4	80.2
	+ Lexicon	90.8	87.3	89.0
	+ Word Vector	93.2	88.6	90.8
Decision Tree	Linguistic	76.7	65.9	70.9
	+ Semantic	82.8	74.3	78.3
	+ Lexicon	88.5	83.8	86.1
	+ Word Vector	91.5	85.2	88.2
Naive Bayes	Linguistic	76.3	64.8	70.1
	+ Semantic	81.6	72.9	77.0
	+ Lexicon	86.2	79.6	82.7
	+ Word Vector	89.4	83.7	86.4
K-Nearest Neighbor	Linguistic	77.6	63.7	69.9
	+ Semantic	80.2	70.6	75.1
	+ Lexicon	84.7	77.2	80.7
	+ Word Vector	88.9	82.5	85.5

The comparison among four classifiers with the combination of all features is presented in the Figure 2. From the results, it found that the F1-score for the SVM classifier, NB classier, IBk classier and J48 classier are 90.8 %, 86.4%, 88.2% and 85.5% respectively.



The proposed system is compared with the baseline system given by the SemEval 2014 organizers for the Aspect Category Detection sub task of ASBA. NRC-Canada is the top performing system for Aspect Category Detection sub task. For a given a sentence, the baseline system identifies the K number of similar sentences from the training data. The similarity among two sentences is computed by determining the Dice coefficient among the sets of unique words in the two sentences (Pontiki et al., 2014). Finally, the most frequent aspect categories which are appeared in the K retrieved sentences are used to tag the input sequence. This approach has a limitation that it measures the semantic similarity among the sentences by employing the text-based similarity measure. The NRC-Canada system relied on five binary SVMs, one for every aspect category. The SVMs used features based on information from a lexicon learn from YELP data and various types of n-grams. The comparison of these two systems with proposed system is presented in the Table 6. The proposed system improves its performance by 2.2 % compared with top performing system in SemEval 2014 conference.

Table 6: Comparison with State-of-the-art methods

Method	Precision	Recall	F1-Score
Baseline	70.2	61.5	65.6
NRC-Canada	91.1	86.3	88.6
Proposed System	93.2	88.6	90.8

Conclusion

This paper addressed the task of Aspect Category Detection in ABSA. The empirical evaluations are performed on four classifiers namely SVM, DT, NB and KNN classifiers. The performance of the classifiers is measured using F1-Score. These classifiers are trained using linguistic features, Semantic features, Lexicon features and Vector based features. From the results it is observed that vector based features are highly influencing the performance of the system. The vector based features, semantic features and lexicons features captures the semantic relations among the words presented in the reviews. These features are trained on the Yelp restaurant dataset to capture the meaningful relations among the words. The test based features unable to establish the semantic relations among the words presented in the reviews.

In future, more suitable features will be included to improve accuracy of our system. For the future work, it would be interesting to explore sentiment lexicons specific to domain to improve the performance and examine more advanced ways of using sentiment lexicons and word embedding features.

References

[1] Priyanka Tyagi,R. C. Tripathi"A Review towards the Sentiment Analysis Techniques for the Analysis of

Twitter Data", Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) ,2019
[2] Jindal, N. and Liu, B. (2008). Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, pages 219–230. ACM.
[3] Xie, S., Wang, G., Lin, S., and Yu, P. S. (2012). Review spam detection via temporal pattern discovery. In Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 823–831. ACM.
[4] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1):1–167.
[5] Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.
[6] Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., Swen, B., and Su, Z. (2008). Hidden sentiment association in chinese web opinion mining. In Proceedings of the 17th international conference on World Wide Web, pages 959–968. ACM.
[7] Popescu, A. and Etzioni, O. (2007). Extracting product features and opinions from reviews. In Natural language processing and text mining, pages 9–28.
[8] Rana, T. A. and Cheah, Y.-N. (2017). A two-fold rule-based model for aspect extraction. Expert Systems with Applications.
[9] Moghaddam, S. and Ester, M. (2012). On the design of lda models for aspect-based opinion mining. In Proc. Inter. Conf. on Information and Knowledge Management, CIKM '12.
[10] Qiu, G., Liu, B., Bu, J., and Chen, C. (2011). Opinion word expansion and target extraction through double propagation. Computational linguistics, 37(1):9–27.
[11] Hu, M. and Liu, B. (2004). Mining and summarising customer reviews. In Proc. Of ACM SIGKDD Inter. Conf. on Knowledge Discovery and Data Mining, KDD '04, pages 168–177.
[12] Kang, Y. and Zhou, L. (2017). Rube: Rule-based methods for extracting product features from online consumer reviews. Information & Management, 54(2):166–176.
[13] 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), pages 28–37.
[14] Zhou, Guodong & Qian, Longhua & Fan, Jianxi. (2010). Tree kernel-based semantic relation extraction with rich syntactic and semantic information. Inf. Sci.. 180. 1313-1325. 10.1016/j.ins.2009.12.006.
[15] Jakob, Niklas & Gurevych, Iryna. (2010). Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields. EMNLP 2010 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference.
[16] H. Lakkaraju, C. Bhattacharyya, I. Bhattacharya, and S. Merugu. Exploiting Coherence for the Simultaneous Discovery of Latent Facets and Associated Sentiments. In SIAM International Conference on Data Mining 2011 (SDM 2011), pages 498–509. SIAM, 2011.
[17] Z. Hai, G. Cong, K. Chang, W. Liu, and P. Cheng. Coarse-to-fine Review Selection via Supervised Joint Aspect and Sentiment Model. In Proceedings of the 37th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2014), pages 617–626. ACM, 2014.
[18] D. Marcheggiani, O. Täckström, A. Esuli, and F. Sebastiani. Hierarchical Multi-label Conditional Random Fields for Aspect-Oriented Opinion Mining. In Proceedings of the 36th European Conference on Information Retrieval (ECIR 2014), pages 273–285. Springer, 2014.
[19] C. Sauper and R. Barzilay. Automatic Aggregation by Joint Modeling of Aspects and Values. Journal of Artificial Intelligence Research, 46(1):89–127, 2013. ISSN 1076-9757.
[20] A. Mukherjee and B. Liu. Modeling Review Comments. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL 2012), pages 320–329. Association for Computational Linguistics, 2012.
[21] Poria, Soujanya & Cambria, Erik & Gelbukh, Alexander. (2016). Aspect Extraction for Opinion Miningwith a Deep Convolutional Neural Network. Knowledge-Based Systems. 108. 10.1016/j.knosys.2016.06.009.
[22] Xue, Wei & Li, Tao. (2018). Aspect Based Sentiment Analysis with Gated Convolutional Networks. 2514-2523. 10.18653/v1/P18-1234.
[23] Dong, Ruihai & O'Mahony, Michael & Smyth, Barry. (2014). Further Experiments in Opinionated Product Recommendation. 10.13140/2.1.3639.1681.