A Proposal for Book Oriented Aspect Based Sentiment Analysis: Comparison over Domains

Tamara Álvarez-López, Milagros Fernández-Gavilanes, Enrique Costa-Montenegro, Patrice Bellot

A Proposal for Book Oriented Aspect Based Sentiment Analysis: Comparison Over Domains

Tamara Álvarez-López[1,2](✉), Milagros Fernández-Gavilanes[1], Enrique Costa-Montenegro[1], and Patrice Bellot[2]

[1] GTI Research Group, University of Vigo, Vigo, Spain
{talvarez,milagros.fernandez,kike}@gti.uvigo.es,
[2] Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France
patrice.bellot@univ-amu.fr

Abstract. Aspect-based sentiment analysis (ABSA) deals with extracting opinions at a fine-grained level from texts, providing a very useful information for companies which want to know what people think about them or their products. Most of the systems developed in this field are based on supervised machine learning techniques and need a high amount of annotated data, nevertheless not many resources can be found due to their high cost of preparation. In this paper we present an analysis of a recently published dataset, covering different subtasks, which are aspect extraction, category detection, and sentiment analysis. It contains book reviews published in Amazon, which is a new domain of application in ABSA literature. The annotation process and its characteristics are described, as well as a comparison with other datasets. This paper focuses on this comparison, addressing the different subtasks and analyzing their performance and properties.

Keywords: Aspect-based sentiment analysis, book reviews, datasets, annotation, evaluation

1 Introduction

People have access to a huge amount of information available online about a wide diversity of subjects and users can express their opinions about products or services by means of social networks or blogs. Analyzing the user-generated content on the Web became very interesting both for consumers and companies, helping them to better understand what people think about them and their products. For this reason, many studies arose in the fields of Opinion Mining and Sentiment Analysis [1], trying to extract the most relevant information in an automatic way. Most of these works aim at extracting the global sentiment from the whole text, by means of different approaches, which can be supervised [2], normally based on classifiers, or unsupervised, such as [3], where the authors aim at capturing and modeling linguistic knowledge by using rule-based techniques. In the last years many works can also be found applying deep learning techniques [4].

However, extracting the specific entities or aspects to which a sentiment is expressed provides a more specific insight of what people think about a particular product, and this is exactly the aim of the so-called ABSA. Several studies emerged in the ABSA field [5, 6], as well as competitions like SemEval [7], where researchers are encouraged to submit and evaluate their systems over common datasets. Again we can find supervised approaches based on classifiers and conditional random fields (CRFs) [8] and unsupervised ones based on frequency studies or syntax dependencies [9].

For developing ABSA systems, annotated datasets are essential for training and testing. However, manually annotating the user-generated reviews is a tough task, time and resource consuming, and not many datasets are available. We can find the following ones. The first one and widely used in the literature is on electronic products [10], tagged with aspects and the sentiment associated. Then, [11] works with restaurant reviews, annotated with categories from a predefined list and their sentiments. Similarly annotated with predefined categories, we find [12], which is about movie reviews. Finally we have to mention the datasets created for SemEval [7], providing datasets for 8 languages and 7 domains, including restaurants, laptops, mobile phones, digital cameras, hotels and museums. Particularly for the book domain, only two works can be found on annotated corpora, one for Arabic language [13] and the other one for Portuguese [14].

Motivated by the few amount of resources found, we developed a new dataset for English language in the book domain, which was firstly presented in [15]. The aim of this paper is to provide a new glance of this dataset, its structure and annotation process, as well as to provide an evaluation benchmark for different subtasks in ABSA and to compare its performance to other datasets available, analyzing the particularities of each one.

The remaining of this article is structured as follows. In Section 2 the datasets under analysis are described. Section 3 presents the new dataset we created, including inter-annotator agreement levels and different statistics. In Section 4 the evaluation and comparison of all the datasets are shown. Finally, Section 5 provides some conclusions and future work.

2 ABSA Datasets

In this section we present the different datasets considered, all of them manually annotated for ABSA. Not much work on annotated datasets for this task can be found, so we use those which are publicly available, in electronics, restaurant and laptop domains, and were widely used in the literature.

2.1 Electronic Product Reviews

The *Customer Reviews Dataset* [10] contains customer reviews of five electronic products, bringing a total number of 314 reviews, collected from Amazon.com and C|net.com and manually annotated. The features for which an opinion is

expressed are tagged, along with the sentiment, represented by a numerical value from +3 (most positive) to -3 (most negative). In this dataset [3] both explicit and implicit features (not explicitly appearing in the text) were annotated, however only the explicit ones were taken into account for this work. In Table 1 detailed information about this dataset is shown. For the next experiments, we will test each dataset separately, using the other four as training.

Table 1. Number of: reviews, sentences, aspects extracted and aspects tagged as positive or negative, for each product in the electronics dataset

Data	Product	#Revs.	#Sent.	#Aspects	#Pos.Asp.	#Neg.Asp.
D1	Digital camera	45	597	256	205	51
D2	Digital camera	34	346	185	155	30
D3	Cell phone	41	546	309	230	79
D4	MP3 player	95	1716	734	441	293
D5	DVD player	99	740	349	158	191
	Total	**314**	**3945**	**1833**	**1189**	**644**

2.2 SemEval Datasets

In the following, the two datasets provided by SemEval workshop [7] for English language are examined, in the domains of restaurants and laptops. They both have a similar structure, containing several reviews annotated at sentence level. In every sentence the specific aspects, called Opinion Target Expressions (OTE), are identified; the aspect category to which the OTE belongs, chosen from a predefined list; and the polarity (*positive*, *negative* or *neutral*).

The list of categories designed for restaurants is formed by a combination of *entity#attribute* and it is composed of 12 different ones, such as *restaurant#prices*, *food#quality* or *ambience#general*. However, for the laptop dataset the categories are much more specific, combining 22 different entities with 14 attributes, obtaining a great number of possibilities. For the aim of this paper we shorten this list by regrouping the entities and attributes, so we can obtain a similar number of categories for all the datasets under evaluation. We only keep the entities *laptop*, *software*, *support* and *company*, while the entities *hardware*, *os*, *warranty* and *shipping* are removed due to their low frequency of appearance. Finally, the rest of the entities are grouped in *components*, as they all refer to different components of a laptop. About the attributes, we keep all of them, but only associated to the *laptop* entity. For the rest of entities, the attribute is always *general*. Like that we obtain a list of 13 categories.

Both datasets are divided into training and test and more detailed information is displayed in Table 2. The information shown about the laptop dataset belongs to the new annotations we created by summarizing the categories.

[3] Available online at https://www.cs.uic.edu/ liub/FBS/sentiment-analysis.html

Table 2. Number of: reviews, sentences and categories, distinguishing between positive, negative and neutral ones, for restaurants and laptops

Domain		#Revs.	#Sent.	#Cat.	#Pos.Cat.	#Neg.Cat.	#Neu.Cat.
Restaurants	Train	350	2000	2506	1657	751	98
	Test	90	676	859	611	204	44
Laptops	Train	450	2500	2730	1547	1002	181
	Test	80	808	749	454	254	41

3 New Book Reviews Dataset

Performing ABSA on book reviews is very useful for different kind of users, including professional as well as non-expert readers, helping them when searching for a book which fits in certain requirements. It can be later applied to recommendation in digital libraries or electronic book shops.

In the following subsections we present the new dataset, the annotation process, its structure and some statistics. It is publicly available online[4] and it was previously introduced in [15], where some information about the annotation process and the dataset properties can be found. Throughout this work, we focus in providing additional information, as well as a baseline for its evaluation, comparing it to other datasets available for the same task.

3.1 Data Collection

For the construction of this dataset, 40 book records were selected randomly from the *Amazon/LibraryThing* corpus in English language provided by the Social Book Search Lab [16], a track which is part of the CLEF (Conference and Labs of the Evaluation Forum). Its main goal is to develop techniques to support readers in complex book search tasks, providing a dataset for the book recommendation task, consisting of 2.8 million book descriptions from Amazon, where each description contains metadata about the *booktitle*, *author* or *isbn*, as well as user generated content, like user ratings and reviews. However, it does not contain any annotations about the aspects or sentiments from the reviews.

Our dataset is composed by the textual content of the reviews associated to each of the 40 books selected, obtaining a total number of 2977 sentences from 300 reviews, which were annotated with aspects, categories and sentiment information.

3.2 Task Description

This new corpus is intended to cover different subtasks in the ABSA field:

[4] http://www.gti.uvigo.es/index.php/en/book-reviews-annotated-dataset-for-aspect-based-sentiment-analysis

1. Aspect Extraction. The aim is to extract the specific features mentioned in the text, which are related to the authors, characters, writing quality, etc. In our dataset we differentiate between explicit and implicit aspects. Implicit aspects are those which are not explicitly written in the analyzed sentence, but can be figured out by the context or the previous sentences:
   e.g.: Sent 1 → When Arthur is suffering from asthma,... (target = Arthur)
   Sent 2 → Then, he starts seeing a large house. (implicitTarget = Arthur)
2. Category Detection. Each target detected in a sentence is classified at a more coarse-grained level, assigning a category from a predefined list. The categories defined for the book domain try to cover most of the features that readers mention in their reviews and are divided into two groups, the ones related to the book itself and those related to its content. In the first group: *general*, *author*, *title*, *audience* (type of readers which the book was written for), *quality* (about the writing style), *structure* (related to the chapters, index, etc.), *period* (when the book was written or published), *length* and *price*. Then the categories included in the second group are: *characters*, *plot*, *genre* (related to the literary genre) and *period* (when the plot passes).
3. Sentiment Polarity. This last task consists of assigning a polarity to every aspect detected from three possible ones: *positive*, *negative* and *neutral*.

3.3 Annotation Process

In order to construct the annotated dataset and support all the tasks previously defined, different tags are attached at sentence level: *Out of Scope*, *Target*, *Occurrence*, *Implicit Target*, *Category* and *Polarity*. More information about them can be found in [15] and an example of an annotated sentence is shown in Figure 1.

```
−<sentence id="000_0007175000_79:6">
  −<text>
      The characters are likeable, the plot is complicated yet compelling and the writing superb
    </text>
  −<Opinions>
      <Opinion category="CONTENT#CHARACTERS" occurrence="1" polarity="positive" target="characters"/>
      <Opinion category="CONTENT#PLOT" occurrence="1" polarity="positive" target="plot"/>
      <Opinion category="BOOK#QUALITY" occurrence="1" polarity="positive" target="writing"/>
    </Opinions>
  </sentence>
```

Fig. 1. Example of an annotated sentence from the corpus

The reviews selected for the dataset were annotated by 3 different annotators, researchers in the field of NLP and, in particular, in Sentiment Analysis from the University of Vigo. In order to ensure the consistency of the final corpus, some guidelines were given to the annotators, as well as training sessions to solve the doubts arisen. Finally, only those annotations in which at least two of the three annotators agreed were taken into account.

When integrating the results from the three annotators, one of the main difficulties found was to determine the boundaries of a particular target. For

example, in the sentence *"I would recommend ages 8 and over"* one annotator can tag *ages 8 and over* as an aspect, whilst other can annotate just *ages 8*. For this situation we consider that both annotators agreed and the longest one is taken into account, as so they do the authors in [14] .

In order to calculate the inter-annotator agreement, the Dice coefficient (D_i) is used, instead of Cohen's Kappa coefficient, as the annotators may tag any word sequence from the sentences, which leads to a very large set of possible classes:

$$D_i = 2 \cdot \frac{|A_i \cap B_i|}{|A_i| + |B_i|} \tag{1}$$

where A_i, B_i are the aspects tagged by annotators A and B, respectively. In Table 3 the agreement between annotators A and B (A|B), A and C (A|C) and B and C (B|C) are shown, as well as the average of the three. The inter-annotator agreement study is also applied to the implicit aspects detected, the category identification, the polarity assigned to an aspect when there was already an agreement in annotating the aspect concerned and the pairs aspect-category, which means that both annotators extracted the same aspect and also assigned the same category to it.

Table 3. Detailed Dice coefficients for aspect, category, aspect+category, polarity and implicit aspect annotations

Annotators	**Aspect**	**Cat.**	**Asp.+Cat**	**Polarity**	**Implicit**
A|B	0.76	0.72	0.65	0.77	0.36
A|C	0.76	0.74	0.68	0.77	0.37
B|C	0.74	0.70	0.63	0.71	0.37
Avg.	**0.75**	**0.72**	**0.65**	**0.75**	**0.37**

As we can see in Table 3, for the task of identifying the specific aspects, the 75% of agreement is reached, which is a good result due to the difficulty of the task. Then similarly, the annotators tag the same categories for a particular sentence with an agreement of 72%, regardless of the aspects extracted. When taking into account the tuple <aspect, category>, the Dice coefficient decreases to 65%, which is normal as in this case they have to agree simultaeously on two parameters. Once the annotators agree on a particular aspect, they also coincide in the polarity annotation with an agreement of 75%. Finally, the Dice measure obtained for the implicit aspect extraction task is 37%, which means a very poor agreement. This can be explained due to the complexity of detecting implicit aspects, as they do not appear written in the sentence, so it makes more difficult the annotation task. There is still much room for improvement in this last task and for the aim of this paper we will not take them into account for the experiments.

3.4 Dataset Characteristics

The complete dataset is composed of 300 reviews, belonging to 40 different books, with a total of 2977 sentences. For the following experiments the dataset was divided into training and test, selecting around the 25% of the reviews for testing and the rest for training, making sure that the reviews included in each one belong to different books, in order to avoid biased results. In Table 4 some additional information is shown.

Table 4. Number of: sentences, annotated aspects (explicit and implicit), and sentences tagged as *out of scope* for the book dataset

Dataset	#Sent.	#Explicit Asp.				#Implicit	#OutOfScope
		#P	#N	#NEU	Total		
Train	2219	726	296	1663	2685	265	469
Test	758	230	88	501	819	64	182

In Table 4 we can also see the polarity distribution (*positive* (P), *negative* (N) and *neutral* (NEU)) of explicit aspects, which is similar for training and test. Moreover, it can also be observed that the number of neutral aspects is quite higher that the rest of the polarities. This is due to the annotation schema, as in this dataset not only the opinionated aspects are annotated, but also those with no opinion associated (*neutral*). We found very common to mention characteristics of the book related to the plot or the characters in an informative way and annotating also this kind of aspects will be useful for future tasks, such as recommendation or summarization.

Finally, in Figure 2 the distribution of the category annotations is shown for training and test, being similar across both of them. We find that the most common categories are *characters* and *plot*. However, categories like the *price* or the *length* of the book are not so usual in this kind of reviews.

4 Performance Evaluation: Comparative

In this section we present the evaluation results for the datasets previously described according to the three different tasks. A baseline system was designed for each task and applied to every dataset, analyzing and comparing the results.

For the evaluation of aspect term extraction and category detection we use precision, recall and F-measure, whilst for sentiment analysis we use the accuracy, following the same evaluation procedure as in the SemEval workshop [7].

4.1 Aspect Extraction Task

For this task the baseline system consists of CRFs, using the CRF++ tool [17]. We extract for each single word the following features: words, lemmas, bigrams,

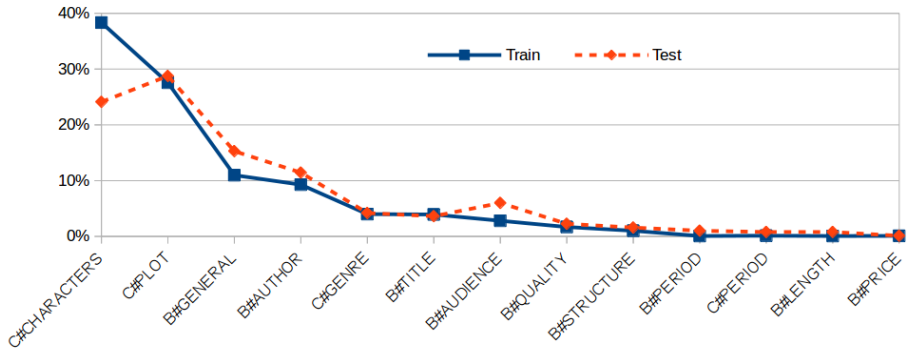

Fig. 2. Aspect category distribution in training and test sets for books

POS tags and entity recognition, obtained by means of Stanford CoreNLP tool [18]. We also extract the value of each two successive features in the the range -2,2 (the previous and subsequent two words of actual word) for the each feature. Moreover when the system identifies two or more consecutive words as aspects, we consider them as a single multiword aspect.

In Table 5 we can see the results obtained for the three datasets (the laptop dataset is not annotated with aspects, therefore it cannot be evaluated for this task), in terms of precision, recall and F-measure. It can be observed that applying the same baseline for all the datasets, the restaurant dataset obtains the highest F-measure. However, for electronics, and especially for books, we find quite low recall, what means that it is harder to extract all the aspects annotated, maybe due to the kind of texts which can be more complex in these two domains, in terms of the diversity of vocabulary.

Analyzing some properties of the datasets, we can see in Figure 3 the relation between the number of sentences, the vocabulary size and the target extraction performance. The vocabulary size of a dataset is determined by the number of unique word unigrams that it contains. As we can see in Figure 3, increasing the number of sentences does not always imply an increase in the vocabulary size. However, when the vocabulary size increases, the performance of the aspect extraction task becomes lower. With a higher amount of different words, there should also be more different aspects to detect.

If we inspect the list of aspects annotated in the test dataset for each domain, we find that for restaurants there are 312 different aspects, 278 different ones for electronics and 417 for books. Moreover, for the book dataset we find more terms which are considered as aspects just once in the whole dataset. For the electronics domain, even if there are less different aspects, the biggest difficulty is to differentiate when the same term should be considered as aspect or not. We find that the terms which are most frequently correctly detected (*true positive*) are usually the same as those which are most frequently not detected (*false*

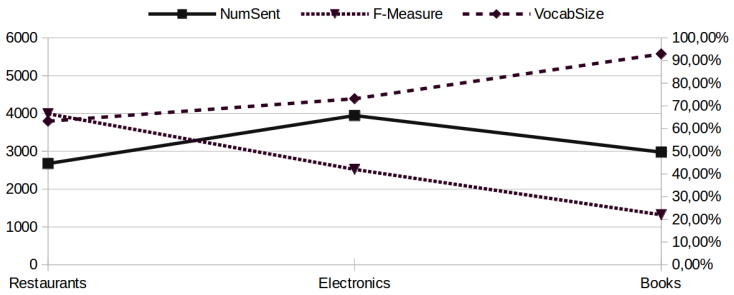

Fig. 3. Number of sentences, vocabulary size, and F-measure for the aspect extraction task for the different datasets

negative), as well as those that are extracted but should have not (*false positive*). This situation arises a very interesting challenge for aspect extraction, as it is not easy to decide when a particular word should be considered as aspect or not, according to the specific context or sentence.

4.2 Category Detection Task

The baseline developed here is based on an ensamble of binary SVM classifiers with a linear kernel, one for each category. The library *libsvm* [5] was used, with the following binary features at sentence level: words, lemmas, POS tags and bigrams. Then, for each sentence, one, several or no categories can be found.

The datasets evaluated for this task are restaurants, laptops and books, as for electronics there are no category annotations, and the performance results, in terms of precision, recall and F-measure, can be seen in Table 5.

We can see again that the restaurant dataset obtains the highest results, while the worst performance is obtained for the book dataset. When the evaluation is performed for each category separately, some differences can be found for the three domains. For the restaurant dataset the F-measure results obtained for the different categories are more similar to each other than those obtained for the categories in the book domain. For laptop dataset, and especially for book dataset, we can find some categories with really low performance. While the worst performance in the restaurant domain is 23% for the *restaurant#miscellaneous* category, in laptops we find four different categories whose F-measure is lower than 25%, rising to eight categories in the book domain. However, we also have to highlight the low representation of certain categories in the book dataset, making it more difficult for the classifier to learn when to annotate them for a particular sentence.

In addition to this, we find the categorization in general more difficult for the book domain. In this dataset the category which achieves the highest F-measure

[5] https://www.csie.ntu.edu.tw/ cjlin/libsvm/

is *characters*, as there are many sentences both in training and test annotated for it. However, the F-measure is still 50%, while in the restaurant domain the categories which are more accurately detected obtain results around 80%.

The categories which obtain the best performance are *restaurant#general*, *food#quality* and *ambience* for restaurants; *laptop#general*, *laptop#price* and *laptop#quality* for laptop dataset; and finally for the book domain they are *general*, *characters* and *author*. These categories are also the most common in both training and test sets for each domain and the classifiers tend to work better with bigger amount of data.

Table 5. Precision, recall and F-measure for aspect and category detection tasks

Dataset	**Aspect**			**Category**		
	Precision	Recall	F-measure	Precision	Recall	F-measure
Restaurants	0.69	0.64	0.66	0.72	0.64	0.68
Electronics	0.64	0.31	0.42	-	-	-
Laptops	-	-	-	0.66	0.40	0.50
Books	0.59	0.14	0.22	0.53	0.30	0.38

4.3 Aspect-Based Sentiment Analysis Task

For extracting the sentiment associated to each aspect, we consider a window of five words before and after the specific aspect. The window size was determined by performing several experiments varying this parameter, so the highest results were obtained when it was equal to five. Then, we add the polarities of all the words included, considered as a numerical value and extracted from a sentiment lexicon, SOCAL [19]. This dictionary is composed of a list of words with a polarity associated, expressed on a scale between -5 (most negative) and +5 (most positive). Then, if the addition of the polarities divided by the number of words with sentiment associated is higher than 1, the target is considered positive; if it is lower than -1 it is tagged as negative; and neutral otherwise.

This task is applied to electronics, restaurant and book datasets at aspect level. For laptop dataset, as there are no annotated aspects, we extract the sentiment of the whole sentence and assign it to the annotated categories. In Table 6 we can see the results obtained. It can be observed that the weighted average accuracy is quite similar for all the datasets. The highest results are obtained for positive aspects. One of the reasons is the high amount of positive aspects in relation to negative or neutral in most of the datasets. In the restaurant domain, the 71% of the aspects are positive, whilst only 5% of them are tagged as neutral. Similar percentages can be obtained from the laptop domain. In electronics, 65% of the aspects are positive and the other 35% are negative. Finally, for the book dataset it can be seen that the accuracy is similar for the three classes. In this case the 61% of the aspects are neutral, 28% are positive and only 10% are

negative. However, if we take into account the F-measure instead of accuracy, we observe that the highest results are obtained for neutral aspects (70%), followed by positive (58%) and negative (24%).

Table 6. Accuracy for aspect polarity extraction task

Dataset	Positive	Negative	Neutral	Weight.Avg.
Restaurants	0.72	0.52	0.32	0.66
Electronics	0.68	0.51	-	0.62
Laptops	0.6	0.5	0.35	0.55
Books	0.57	0.56	0.57	0.57

5 Conclusions

The aim of this paper was to provide a common evaluation benchmark for the ABSA task and its different subtasks: aspect extraction, category detection and sentiment analysis at the aspect level. Different datasets available in the literature were studied and compared, belonging to electronics, restaurant and laptop domains. Moreover, different characteristics and the annotation process were described for the new dataset in the domain of books, not yet explored in the state of the art. Then, different baselines were proposed for the evaluation and comparison. The aim here was not to provide a new approach for ABSA, but to analyze the distinctive features of reviews from different domains and how they affect to the ABSA performance, as well as to perform the evaluation of the new dataset developed.

As future work we plan to continue the research in the book domain and develop new baselines which fit better for this kind of reviews, which seem to present bigger challenges for ABSA. Moreover, we would like to work in the integration of aspect extraction from book reviews for improving book recommendation systems, introducing them as new inputs, so that the system could apply a reranking of the recommendation list according to this new information.

Acknowledgments This work is supported by Mineco grant TEC2016-C2-2-R, Xunta de Galicia grant GRC and the French ANR program "Investissements d'Avenir" EquipEx DILOH (ANR-11-EQPX-0013).

References

1. Liu, B.: Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies. 5, 1–167 (2012)
2. Pak, A., Paroubek, P.: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In: 7th International Conference on Language Resources and Evaluation, pp.1320–1326. LREC (2010)

3. Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., González-Castaño, F.J.: Unsupervised Method for Sentiment Analysis in Online Texts. Expert Systems with Applications. 58, 57–75 (2016)
4. Dos Santos, C. N., Gatti, M.: Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In: 25th International Conference on Computational Linguistics, pp.69–78. COLING (2014)
5. Schouten, K., Frasincar, F.: Survey on Aspect-Level Sentiment Analysis. IEEE Transactions on Knowledge and Data Engineering. 28, 813–830 (2016)
6. Akhtar, M. S., Gupta, D., Ekbal, A., Bhattacharyya, P.: Feature Selection and Ensemble Construction: A Two-step Method for Aspect Based Sentiment Analysis. Knowledge-Based Systems. 125, 116–135 (2017)
7. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., AL-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N., Kotelnikov, E., Bel, N., Jiménez-Zafra, S.M., Eryigit, G.: SemEval-2016 task 5: Aspect based sentiment analysis. In: Workshop on Semantic Evaluation (SemEval-2016), pp. 19–30. ACL (2016)
8. Jakob, N., Gurevych, I.: Extracting Opinion Targets in a Single and Cross Domain Setting With Conditional Random Fields. In: Conference on Empirical Methods in Natural Language Processing, pp.1035–1045. (2010)
9. Poria, S., Cambria, E., Ku, L., Gui, C., Gelbukh, A.: A Rule-Based Approach to Aspect Extraction from Product Reviews. In: 2nd Workshop on Natural Language Processing for Social Media, pp.28–37. (2014)
10. Hu, M., Liu, B.: Mining and Summarizing Customer Reviews. In: 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.168–177. ACM (2004)
11. Ganu, G., Elhadad, N., Marian, A.: Beyond the Stars: Improving Rating Predictions using Review Text Content. In: 12th International Workshop on the Web and Databases, pp.1–6. (2009)
12. Thet, T.T., Na, J.C., Khoo, C.SG.: Aspect-Based Sentiment Analysis of Movie Reviews on Discussion Boards. Journal of Information Science. 36, 823–848 (2010)
13. Al-Smadi, M., Qawasmeh, O., Talafha, B., Quwaider, M.: Human Annotated Arabic Dataset of Book Reviews for Aspect Based Sentiment Analysis. In: 3rd International Conference on Future Internet of Things and Cloud, pp.726–730. (2015)
14. Freitas, C., Motta, E., Milidiú, R., César, J.: Sparkling Vampire... lol! Annotating Opinions in a Book Review Corpus. New Language Technologies and Linguistic Research: A Two-Way Road. Cambridge Scholars Publishing. pp.128–146 (2014)
15. Álvarez-López, T., Fernández-Gavilanes, M., Costa-Montenegro, E., Juncal-Martínez, J., García-Méndez, S., Bellot, P.: A Book Reviews Dataset for Aspect Based Sentiment Analysis. In: 8th Language & Technology Conference. pp.49–53. (2017)
16. Koolen, M., Bogers, T., Gäde, M., Hall, M., Hendrickx, I., Huurdeman, H., Kamps, J., Skov, M., Verberne, S., Walsh, D.: Overview of the CLEF 2016 Social Book Search Lab. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp.351–370. (2016)
17. Kudo, T.: CRF++: Yet another CRF toolkit. http://crfpp.sourceforge.net/ (2005)
18. Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In: Association for Computational Linguistics (ACL) System Demonstrations. pp.55–60. (2014)
19. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational Linguistics. 37, 267–307 (2011)