Chapter 11
DomSent: Domain-Specific Aspect Term Extraction in Aspect-Based Sentiment Analysis

Ganpat Singh Chauhan and Yogesh Kumar Meena

Abstract In recent research aspect-based sentiment analysis has played a vital role in identifying user's opinions from the unstructured natural text. One of the most critical subtasks in aspect-based sentiment analysis is to extract the most prominent aspect terms. In this paper, we have studied previous research done on aspect term extraction and proposed an approach DomSent to identify aspects using domain-specific information while applying frequency and similarity pruning. Our experimental results show that the aspect extraction using domain-specific information contributes better as compared to the recent aspect term extraction approaches.

Keywords Opinion mining · Sentiment analysis · Domain-specific information · Aspect-based sentiment analysis · Aspect term extraction · Machine-learning

11.1 Introduction

In the last decade of e-commerce the online experiences, opinions, and reviews from users have influenced customers tremendously in making decisions. It is always important to retrieve opinions from the user text which increase purchasing of the product if users have recommended it and at the same time it gives an opportunity to do the required changes in the services and products if users have complaints and suggestions are given in the review. In the last few years, most of the popular internet websites (commercial, social, forums, blogs etc.) have started to use users experiences in place of using conventional approaches for the survey. These user's reviews add value for different application domains like the product, hotel, movie, education, government schemes etc. but the analysis and management of this much amount of text is very time-consuming [1, 2]. Aspect-based Sentiment Analysis

G. S. Chauhan (✉) · Y. K. Meena
MNIT, Jaipur, India
e-mail: ganpatchauhan@gmail.com

Y. K. Meena
e-mail: ymeena.cse@mnit.ac.in

A. K. Somani et al. (eds.), *Smart Systems and IoT: Innovations in Computing*, Smart Innovation, Systems and Technologies 141,
https://doi.org/10.1007/978-981-13-8406-6_11

(ABSA) extracts opinions expressed against multiple aspects (price, material etc.) of the entity (product) [3–5]. The most studied task in the ABSA is the aspect term extraction [6]. In the sentence "picture quality of this digital camera is very good but video making is average," the aspect terms are "picture" and "video making", and the entity is "digital camera".

The researchers focused on explicit aspects extraction that doesn't include the domain information [7, 8]. Once aspect terms are extracted then extra phases like pruning etc. are applied to remove unimportant and irrelevant aspect terms [9]. This affects the performance of aspect extraction approaches. In this paper, we have presented Domain-Specific aspect term extraction method (DomSent) which focuses on the approach of filtering prominent aspect terms using domain-specific information at the time of extraction of the aspect term [11, 12].

11.2 Related Work

In the early time, the unsupervised frequency-based approach has considered frequent nouns, noun phrases with association rule-miner to calculate the frequency of all noun terms from the document to explore aspect extraction subtask [3, 4, 10–13]. It had found that all the extracted aspect terms were not in the context of the domain being considered. It has been shown in research that unsupervised approaches such as double propagation (DP) [14] perform better than supervised CRF-based methods [15]. This frequency-based approach had further extended by introducing two problems: "most common but irrelevant aspects" and "non-common but most relevant aspects" [16, 17]. It has further been improved using aspect semantic similarity-based and frequency-based approach to detect irrelevant aspects [18, 19]. The extended rule-based TF-RBM uses sequential pattern based rules and also considers domain-independent and domain-dependent opinions [20]. To extend the coverage of aspect terms common-sense and contextual information has been considered for important feature selection [21]. An ontology-based technique has been used for twitter posts regarding a specific topic [22]. A learning-based approach context-based sentiment analysis (ConSent) [23, 24] and semi-supervised methods [25, 26] have been proposed for aspect extraction [27].

The above studies show that most of the aspect extraction approaches use a domain or context-dependent methods [7, 23]. In this paper, we have proposed DomSent to include domain-specific information while extracting explicit aspect terms.

11.3 Methodology

In proposed method, our major objective is to identify the most relevant aspects with the help of domain-specific information. We have used Opinion Lexicon prepared with the help of contextual information. This lexicon contains sentiment words with

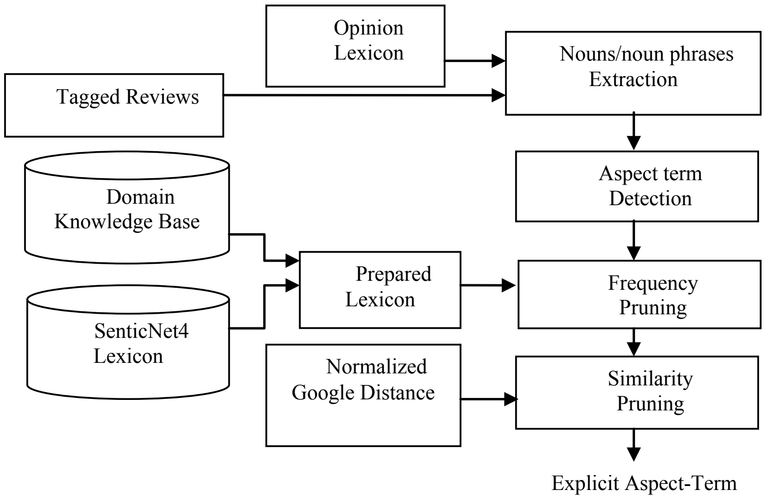

Fig. 11.1 DomSent model for aspect terms extraction using domain-specific information

their polarity scores. Polarity has been assigned considering the context of the aspect. We have adopted the methods given in [21] to prepare this opinion lexicon. With the use of this opinion lexicon we have extracted only contextual aspect terms. The aspect term "similarity" may be used in different domains like education, product etc. and at the same time its importance may be different in any other domain. Next, in the sentence "teaching quality and explanation quality is excellent but the content is very poor" although we have negative polarity on two aspects and positive comment only on one aspect still polarity will be positive because in "education" domain "content" is the most important aspect. Next, [21] have not applied contextual lexicon for extracting nouns/noun phrase rather they have used that only at the time of polarity detection. Further, we have prepared our lexicon with the domain-specific knowledge and SenticNet4 [28] as shown in Fig. 11.1. We have used this lexicon at the time of frequency pruning to extract domain-specific important and relevant aspect terms. The problem with frequency pruning is that some uncommon but relevant aspect terms from the reviews have also been removed and some common but irrelevant aspect terms has been considered as an explicit aspect. To overcome this problem similarity pruning is done [29]. The flow is presented in Fig. 11.1.

11.4 Evaluation and Comparison of Results

In this section, we have carried out experiments on dataset formed from [3, 4] on five different electronic products. If we have A extracted aspects in the given dataset then number of annotated aspects are N. Based on the assumption TP will be True Positive, FP will be False Positive and FN will be False Negative.

DataSet Description: In dataset statistics are as follows:

F1: Name of the entity (EP1, EP2, EP3, EP4, EP5)
F2: Product Name (Apex DVD player, Canon Digital Camera, Creative MP3 Player, Nikon Digital Camera, Nokia Cell Phone)
F3: Number of sentences (740, 597, 1716, 346, 546)
F4: Number of subjective sentences (344, 238, 720, 160, 265)
F5: Number of objective sentences (396, 359, 996, 186, 282)
F6: Number of explicit aspects (296, 237, 674, 174, 302)

The formula for precision is calculated as in Eq. 11.1:

$$P = TP/(TP + FP) \tag{11.1}$$

Similarly, recall and F-score has formulated as in Eqs. 11.2 and 11.3:

$$R = TP/(TP + FN) \tag{11.2}$$

$$F = 2 * P * R/(P + R) \tag{11.3}$$

Our results are compared with original results of different state-of-the-art and most recent aspect extraction approaches available in the literature such as: Hu and Liu [3, 4], DP, Popescu [30], and two-fold rule-based model (TF-RBM) [20]. To implement the DomSent approach, a manual inspection was performed on the dataset.

Table 11.1, shows the baseline values for used approaches. Table 11.1 also shows the precision and recall values when DomSent has been implemented for different products EP1–EP5 on the used dataset. The result shows that if we use the domain-specific lexicon to avoid important and relevant non-frequent aspects to be removed we got the precision 0.88. The comparison of results of proposed approaches with results from Table 11.1 has shown in graphs in Figs. 11.2 and 11.3.

Table 11.1 Precision and recall of DomSent on dataset and results of other recent approaches

Approach(s)	EP1		EP2		EP3		EP4		EP5	
	P	R	P	R	P	R	P	R	P	R
DomSent	0.89	0.76	0.83	0.86	0.83	0.92	0.89	0.87	0.94	0.86
Hu and Lui	0.75	0.82	0.71	0.79	0.72	0.76	0.69	0.82	0.74	0.80
DP	0.53	0.76	0.60	0.84	0.54	0.75	0.60	0.79	0.58	0.81
Popescu	0.89	0.80	0.87	0.74	0.89	0.74	0.86	0.80	0.90	0.78
TF-RBM	0.83	0.73	0.71	0.78	0.70	0.80	0.83	0.87	0.90	0.80

Precision (labeled as P) and Recall (labeled as R)

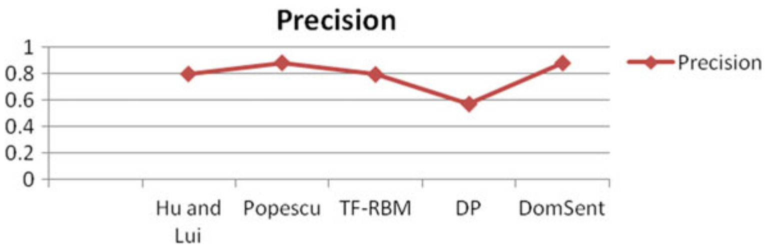

Fig. 11.2 Comparison of precision value of DomSent with related approaches

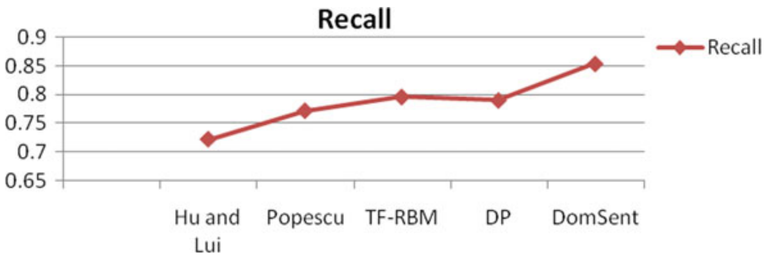

Fig. 11.3 Comparison of recall value of DomSent with related approaches

11.5 Conclusion and Future Directions

In this paper, we have used domain-specific information at the time of frequency pruning. None of the existing approaches have used domain information while pruning to identify domain-specific terms. It has also been analyzed with the results that DomSent produced better results than other state-of-art and recent approaches. In Future, deep learning using word embedding can also be considered for sentiment analysis using convolution neural nets.

Acknowledgements We would like to extend our acknowledgement to all the volunteers for carrying out sentiment analysis on a huge database in order to achieve survey based experimental results.

References

1. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, vol. 10, pp. 79–86 (2002)
2. Turney, P.D.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, pp. 417–424 (2002)
3. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM-SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177.

ACM (2004) http://dl.acm.org/citation.cfm?id=1014073
4. Hu, M., Liu, B.: Mining opinion features in customer reviews. AAAI **4**, 755–760 (2004)
5. Huang, S., Liu, X., Peng, X., Niu, Z.: Fine-grained product features extraction and categorization in reviews opinion mining. In: 12th IEEE International Conference on Data Mining Workshops (ICDMW), pp. 680–6 86 (2012)
6. Rana, T.A., Cheah, Yu.-N.: Aspect extraction in sentiment analysis: comparative analysis and survey. Artif. Intell. Rev. **46**, 459–483 (2016). https://doi.org/10.1007/s10462-016-9472-z (Springer Science + Business, Media, Dordrecht)
7. Wu, C., Huang, Y.: A hybrid unsupervised method for aspect term and opinion target extraction. Knowl.-Based Syst. **148**, 66–73 (2018) (Elsevier)
8. García-Pablos, A.,Rigau, G.: W2VLDA: almost unsupervised system for aspect based sentiment analysis. Expert Syst. Appl. **91**, 127–137 (2018) (Elsevier)
9. Poria, S.: Aspect extraction for opinion mining with a deep convolution neural network. Knowl.-Based Syst. **108**, 42–49 (2016) (Elsevier)
10. Liu, K., Xu, L., Zhao, J.: Opinion target extraction using word-based translation model. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1346–1356. ACL (2012)
11. Liu, B.: Sentiment analysis and subjectivity. Handb. Nat. Lang. Process. **2**, 627–666 (2010)
12. Liu, B.: Sentiment analysis and opinion mining. Synth. Lect. Hum. Lang. Technol. **5**(1), 1–167 (2012)
13. Sarawgi, K., Pathak, V.: Opinion mining: aspect level sentiment analysis using SentiWordNet and Amazon web services. Int. J. Comput. Appl. **158**(6), 0975–8887 (2017)
14. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. Comput. Linguist. **37**(1), 9–27 (2011). http://www.mitpressjournals.org/doi/abs/10.1162/coli_a_00034
15. Chen, L., Qi, L., Wang, F.: Comparison of feature-level learning methods for mining online consumer reviews. Expert Syst. Appl. **39**(10), 9588–9601 (2012)
16. Wang, B., Wang, H.: Bootstrapping both product features and opinion words from Chinese customer reviews with cross-inducing. In: IJCNLP, vol. 8, pp. 289–295 (2008)
17. Wei, C.-P., Chen, Y.-M., Yang, C.-S., Yang, C.C.: Understanding what concerns consumers: a semantic approach to product feature extraction from consumer reviews. Inf. Syst. E-Bus. Manag. **8**(2), 149–167 (2010)
18. Rana, T.A., Cheah, Y.-N.: Improving Aspect Extraction Using Aspect Frequency and Semantic Similarity-Based Approach for Aspect-Based Sentiment Analysis. Springer International Publishing AG, New York (2018)
19. Schouten, K., van der Weijde, O., Frasincar, F., Dekker, R.: Supervised and unsupervised aspect category detection for sentiment analysis with co-occurrence data. IEEE Trans. Cybern. 2168–2267 (2017). https://doi.org/10.1109/tcyb.2017.2688801 (IEEE)
20. Toqir, A.R., Cheah, Yu.-N.: A two-fold rule-based model for aspect extraction. Expert Syst. Appl. **89**, 273–285, 0957–4174 (2017) (Elsevier Ltd)
21. Agarwal, B., Mittal, N., Bansal, P., Garg, S.: Sentiment analysis using ConceptNet ontology and context information. Modern Computational Models of Semantic Discovery in Natural Language. IGI Global (2014)
22. Kontopolous, E., Berberidis, C., Dergiades, T., Bassiliades, N.: Ontology-based sentiment analysis on twitter posts. Expert Syst. Appl. **40**, 4065–4074 (2013) (Elsevier)
23. Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: tasks, approaches, and applications. Knowl.-Based Syst. **89**, 14–46 (2015)
24. Wang, T., Cai, Y., Leung, H.-F., Lau, R.Y.K., Li, Q., Min, H.: Product aspect extraction supervised with online domain knowledge. Knowl.-Based Syst. **71**, 86–100 (2014) (Elsevier)
25. Gao, Z., Liu, B., Zhang, Y.: Automated rule selection for aspect extraction in opinion mining. In: IJCAI, vol. 15, pp. 1291–1297 (2015)
26. Katz, G., Ofek, N., Shapira, B.: ConSent: context-based sentiment analysis. Knowl.-Based Syst. **84**, 162–178 (2015) (Elsevier)

27. Pavlopoulos, I.J.: Aspect based sentiment analysis. Ph.D. thesis, Department of Informatics Athens University of Economics and Business (2014)
28. Cambria, E., Poria, S., Bajpai, R., Schuller, B.: SenticNet4: a semantic resource for sentiment analysis based on conceptual primitives. In: COLING, pp. 2666–2677 (2016)
29. Cilibrasi, R.L., Vitanyi, P.M.: The Google similarity distance. IEEE Trans. Knowl. Data Eng. **19**(3), 370–338 (2007)
30. Popescu, A.-M., Etzioni, O.: Extracting product features and opinions from reviews. Nat. Lang. Process. Text Min. 9–28 (2007) (Springer, New York)