Sentiment Analysis Using Anaphoric Coreference Resolution and Ontology Inference

Thi Thuy Le(✉), Thanh Hung Vo, Duc Trung Mai,
Than Tho Quan, and Tuoi Thi Phan

Ho Chi Minh University of Technology, Ho Chi Minh City, Viet Nam
{thuylt, vthung, mdtrung, qttho, tuoi}@cse.hcmut.edu.vn

Abstract. Aspect-based sentiment analysis is an emerging trend in the NLP area nowadays. One of the major tasks in this work is to identify corresponding aspects for rating sentiment. Ontology is considered highly useful to cope with this issue, due to its capability of capturing and representing concepts in a certain domain. However, ontology-based sentiment analysis suffers from the difficulty when handling anaphoric coreference of mentioned entities, which commonly occurs in textual documents. This paper addresses this problem by introducing an approach combining coreference resolution with ontology inference. The initial results are quite promising.

Keywords: Aspect-level sentiment analysis · Aspect-oriented sentiment ontology · Anaphoric coreference resolution

1 Introduction

Sentiment analysis [1] is an emerging trend nowadays in the Natural Language Processing (NLP) field. There are three levels of sentiment analysis: (i) *document-based* level; (ii) *sentence-based* level; and (iii) *aspect-based* level. In this paper, we focus on the problem of *aspect-level* [2]. To be more precise, apart from rating positive/negative sense of a mention, the objects (or *aspects*) targeted by the mention must also be identified. *Ontology* is often used to represent the aspects in a machine-readable form. However, ontology-based approaches usually suffer from the difficulty of handling *anaphoric coreference* problem commonly occurring in natural languages. To illustrate it, let us consider the following example.

(S1) *I consider an iPhone 6S. Unlike Samsung S7, it is unfortunately not really affordable for students. However, the design looks nice and eye-catching.*

When analyzing the above text, Stanford CoreNLP toolkit [3], one of the most popular tools for natural language processing, returns the result shown in Fig. 1. Here one can observe an example of *anaphoric coreference*, when the pronoun *it* in the second sentence implies the aspect *iPhone6* in the first sentence. However, the problems which are still left in this example for further processing are as follows. In the second sentence, the negative term "*is not really affordable*" has already been detected by Stanford CoreNLP Toolkit. Similarity, the positive terms of "*nice and eye-catching*"

C. Sombattheera et al. (Eds.): MIWAI 2016, LNAI 10053, pp. 297–303, 2016.
DOI: 10.1007/978-3-319-49397-8_26

in the third sentence have also been recognized. However, these terms are not connected to the corresponding aspects, which are *iPhone 6* and *design* respectively. Moreover, as *design* is an *attribute* of *iPhone6* in this context, those positive terms should also be connected to *iPhone6* as well.

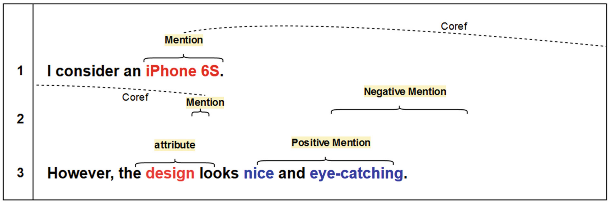

Fig. 1. The processing result of Stanford CoreNLP Toolkit for (S1)

In this paper, we address those problems by the following approach. Firstly, *conceptual graphs* (CGs) are used to represent individual sentences. Then, we perform *anaphoric conference resolution* to link the coreference nodes on the generated CGs together. *Ontology inference* is also used to make connection from the attributes with corresponding aspects. Eventually, we conduct sentiment analysis on the final resultant combined CGs.

2 Related Works

2.1 Ontology-Based Aspect-Level Sentiment Analysis

Ontology is often used more when conducting sentiment analysis at aspect level. Kontopoulos et al. [4] used ontology for analyzing the twitter posts. In [5], a fuzzy product ontology has been proposed. Hierarchical structures of ontology have been proved effective for handling online reviews [6–9]. In NETOWL [10] tool, the ontology-based problem for aspect sentiment and coreference resolution have also been mentioned. However, there is no concrete work reported to combine these two approaches.

2.2 Coreference Resolution

Research on the anaphoric coreference problem primarily focused on the resolution based on *noun*, *pronoun* (anaphora) and *name entities* (NEs). The approaches include using the supervised [11]; semi supervised or unsupervised [12] and machine learning techniques [11]. Work on semantic characteristic such as vocabulary and syntax [13, 14] are also reported. Other remarkable approaches include graph algorithm [15] and rule-based approaches [13]. In particular, *Stanford CoreNLP Toolkit* [3] has emerged recently as the most notable system for anaphoric coreference resolution.

3 Aspect-Oriented Sentiment Ontology

To adopt ontology for sentiment analysis, we firstly introduce formal definition of *Aspect-oriented Sentiment Ontology* as follows.

Definition 1 (Aspect-Oriented Sentiment Ontology). An aspect sentiment ontology SO is a pair of $\{C, R\}$; where $C = (C^A, C^S)$ represents a set of concepts, which consists of 2 elements: C^A is a set of aspect concepts, and C^S is a set of sentiment concepts; $R = (R^T, R^N, R^S)$ represents a set of relationships, which consists of 3 elements: R^N is a set of non-taxonomic relationships, R^T is a set of taxonomic relationships, R^S is a sentiment relationship. Each concept ci in C represents a set of objects, or instances, of the same kind, denoted as *instance-of* (c_i). Each relationship $r_i(c_p, c_q)$ in R represents a binary association between concepts c_p and c_q, and the instances of such a relationship, denoted as *instance-of* (r_i), are pairs of (c_p, c_q) concept objects. Especially, an instance $r_i^s(a, s)$ in R^S implies a relationship between an aspect a $\in A$ and a sentiment term s $\in S$.

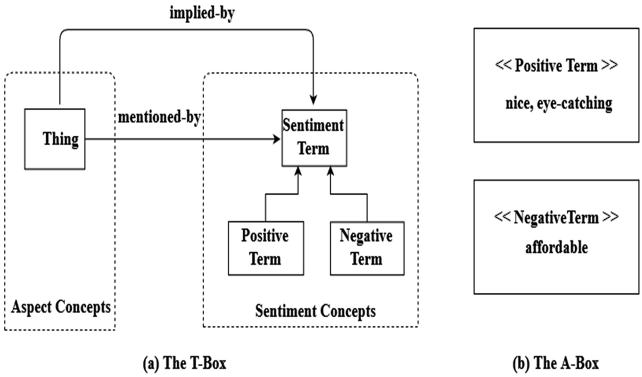

Fig. 2. An example of Generic Ontology

To graphically visualize an ontology, we rely on the idea of *T-Box* and *A-Box* [16]. Basically, a T-Box captures the relations between *concepts* and an A-Box describes *instances* of concepts. Figure 2 presents the T-Box and A-Box of Generic Ontology G_o. Generally speaking, G_o consists of one aspect concept of *Thing*, whose instances can be any real-life concepts. An instance of *Thing* can be *mentioned* or *implied* by a *Sentiment Term,* which can be either *Positive Term* or *Negative Term*. We also introduce two specific sentiment relationships for sentiment ontology, known as *mentioned-by* and *implied-by*. An aspect instance c can be *mentioned-by* a sentiment term s, meaning that c is being positive or negative rated, depending on whether s belongs to *Positive Term* or *Negative Term* classes, respectively. Moreover, *implied-by* is similar to *mentioned-by,* but carrying on more specific meaning. An aspect instance c can be *implied-by* a sentiment term s, that is s is only applicable for c, not for other aspects. Thus, when s occurs in a textual statement ϑ, one can infer that c is also implied in ϑ, without explicit mention.

4 Proposed Method

Our proposed method is carried out as follows. Firstly, we base on the work presented in [17] to construct conceptual graphs (CG) for the sentences in the given context. For example, the CGs of sentences in (S1) are generated as illustrated in Fig. 3.

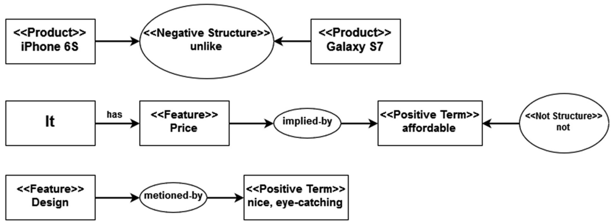

Fig. 3. Conceptual graphs are generated from sentences in (S1)

Then, we perform anaphoric coreference resolution and ontology inference to combine the corresponding nodes on the separate CGs. Ontology inference is performed in a heuristic-based manner. That is, once the system captures the occurrences of two related aspects in the same or two consecutive sentences, it infers that two aspects refer to the same entity. For instance, in the statement (S1), *S7 camera* and *design* are inferred as the same entity object where *design* is an attribute of *S7 camera.*

Finally, based on the discovered *mentioned-by* and *implied-by* as previously discussed, the system can establish relationships between aspects and sentiment terms. Depending on the opinion orientation of the sentiment terms, the rating (*positive/neutral/negative*) will be evaluated accordingly (Fig. 4).

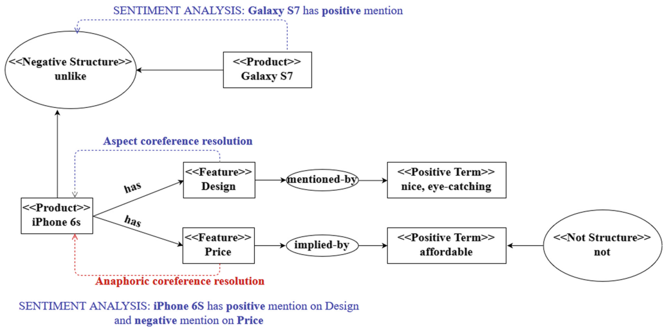

Fig. 4. An example of sentiment analysis on conceptual graph

5 Experimental Result

In order to conduct an experiment from real-life data, we obtain real datasets of user reviews on smartphone products from YouNet Media (YNM), a company dedicated for social listening and market intelligence[1]. The dataset covers 32 smartphone brands, 234 products. It consists of 2809, 3098 and 365 *negative*, *neutral* and *positive mentions*, respectively. The experts of YNM also helped us to define aspect (attribute) of the Smartphone domain, as depicted in Table 1.

Table 1. Some example of aspect and sentiment terms

Attribute	Aspect term	Sent. term (positive)	Sent. term (negative)
Design	Design, shape	attractive, eye-catching	cloddish, flat
Screen	inch, pixel	sharp, anti-glare	opaque, stained screen
Camera	lens, autofocus	wide, bright	blur, light-interference

We then measured the accuracy of our sentiment analysis approach. We did compare the performance of four sentiment analysis strategies as follows. The first strategy, SEN-FULL applied our full framework. SEN-NO-ONT and SEN-NO-RULES did not use Aspect-oriented Sentiment Ontology and Anaphoric Coreference Resolution respectively. Eventually, we used SVM for sentiment classification, as this technique was employed by various related works.

Figure 5 presents the accuracy percentage when we applied those analysis strategies on the collected datasets. It can be observed that in general our proposed method gains better performance.

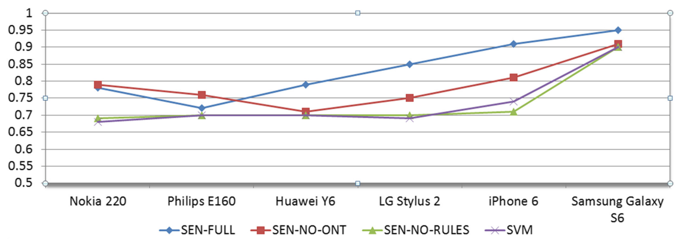

Fig. 5. Accuracy performance of sentiment analysis strategies

6 Conclusion

In this paper, we discuss an approach combining anaphoric coreference resolution with ontology inference for aspect-level sentiment analysis. As a result, the Aspect-oriented Sentiment Ontology is proposed, around which the coreference resolution and

[1] http://www.younetmedia.com/.

aspect-based sentiment analysis are centered. Our experiments on real datasets acquired from the actual discussion on social channels have achieved promising performance.

Acknowledgments. This research is funded by Vietnam National University Ho Chi Minh City (VNU-HCM) under grant number C2016-20-36. We are also grateful to YouNet Media for supporting real datasets for our experiment.

References

1. Padmaja, S., Fatima, S.: Opinion mining and sentiment analysis - an assessment of peoples' belief: a survey. Int. J. Ad hoc Sens. Ubiquitous Comput. **4**(1), 21–33 (2013)
2. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Kim, W., Kohavi, R., Gehrke, J., DuMouchel, W. (eds.) Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, Washington, USA, 22–25 August 2004, pp. 168–177. ACM (2004)
3. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: ACL (System Demonstrations), pp. 55–60, June 2014
4. Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N.: Ontology-based sentiment analysis of Twitter posts. Expert Syst. Appl. **40**(10), 4065–4074 (2013)
5. Lau, R.Y., Li, C., Liao, S.S.: Social analytics: learning fuzzy product ontologies for aspect-oriented sentiment analysis. Decis. Support Syst. **65**, 80–94 (2014)
6. Thet, T.T., Na, J.C., Khoo, C.S.: Aspect-based sentiment analysis of movie reviews on discussion boards. J. Inf. Sci. (2010). doi:10.1177/0165551510388123
7. Kim, S., Zhang, J., Chen, Z., Oh, A. H., Liu, S.: A hierarchical aspect-sentiment model for online reviews. In: AAAI, July 2013
8. Titov, I., McDonald, R.: A joint model of text and aspect ratings for sentiment summarization. In: McKeown, K., Moore, J.D., Teufel, S., Allan, J., Furui, S. (eds.), Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics, Columbus, Ohio, USA, 15–20 June 2008, ACL 2008, pp. 308–316. The Association for Computer Linguistics (2008)
9. Wei, W., Gulla, J.A.: Sentiment learning on product reviews via sentiment ontology tree. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 404–413. Association for Computational Linguistics, July 2010
10. Netowl About. Net Owl. N.p. (2016). https://www.netowl.com/our-story/. Accessed 7 Sept 2016
11. Kobdani, H., Schütze, H., Schiehlen, M., Kamp, H.: Bootstrapping coreference resolution using word associations. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, vol. 1, pp. 783–792. Association for Computational Linguistics, June 2011
12. Haghighi, A., Klein, D.: Unsupervised coreference resolution in a nonparametric Bayesian model. In: Annual meeting-Association for Computational Linguistics, vol. 45, no. 1, p. 848, June 2007
13. Haghighi, A., Klein, D.: Simple coreference resolution with rich syntactic and semantic features. In: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3, pp. 1152–1161. Association for Computational Linguistics, August 2009

14. Bengtson, E., Roth, D.: Understanding the value of features for coreference resolution. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp. 294–303. Association for Computational Linguistics, October 2008
15. Nicolae, C., Nicolae, G.: BESTCUT: a graph algorithm for coreference resolution. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 275–283. Association for Computational Linguistics, July 2006
16. Quan, T.T., Hui, S.C.: Ontology-based natural query retrieval using conceptual graphs. In: Ho, T.-B., Zhou, Z.-H. (eds.) PRICAI 2008. LNCS (LNAI), vol. 5351, pp. 309–320. Springer, Heidelberg (2008). doi:10.1007/978-3-540-89197-0_30
17. Soon, W.M., Ng, H.T., Lim, D.C.Y.: A machine learning approach to coreference resolution of noun phrases. Comput. Linguist. **27**(4), 521–544 (2001)