Sentic Demo: A Hybrid Concept-Level Aspect-Based Sentiment Analysis Toolkit

Soujanya Poria[1,4], Nir Ofek[2], Alexander Gelbukh[3], Amir Hussain[4], and Lior Rokach[2]

[1] School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore
[2] Department of Information Systems Engineering, Ben Gurion University, Israel
[3] Centro de Investigación en Computación, Instituto Politécnico Nacional, Mexico
[4] Department of Computing Science and Mathematics, University of Stirling, UK
`sporia@ntu.edu.sg,sp47@cs.stir.ac.uk,`
`nirofek@bgu.ac.il,liorrk@bgu.ac.il,www.gelbukh.com,ahu@cs.stir.ac.uk`

Abstract. The ways people express their opinions and sentiments have radically changed in the past few years thanks to the advent of social networks, web communities, blogs, wikis, and other online collaborative media. Ideally, automatic analysis of online opinions should involve deep understanding of natural language text by machines, from which we are still very far. In this work, we introduce a novel paradigm for concept-level sentiment analysis that merges linguistics, common-sense computing, and machine learning for improving the accuracy of tasks such as polarity detection. By allowing sentiments to flow from concept to concept based on the dependency relation of the input sentence, in particular, we achieve a better understanding of the contextual role of each concept within the sentence. With this, our polarity detection engine outperforms the state-of-the-art statistical methods.

1 Concept Parser and Aspect Parser

1.1 Concept Parser

Semantic parsing is crucial for such tasks as concept-based opinion mining [1], big social data analysis [2], and crowd validation [3]. The proposed concept parser deconstructs input sentences into multi-word expressions. For this, it extracts concepts from the dependency parse tree of the sentence,[5] basing on hand-crafted rules. Examples of such rules are given below; see more details in [4,5].

Subject Noun Rule If the active token *h* is in a subject noun relationship with a verb *t*, then the concept *t-h* is extracted. E.g., in (1), *movie* is the subject of *boring*; the concept `boring-movie` is extracted.

(1) The movie is boring.

[5] We used the Stanford Dependency parser, `http://nlp.stanford.edu/software/lex-parser.shtml`.

Joint Subject Noun and Adjective Complement Rule If the active token *h* is in a subject noun relationship with a verb *t* and *t* is in adjective complement relationship with a adverb *w*, then the concept *w-h* is extracted. E.g., in (2), *flower* is the subject of *smells*, which is in adjective complement relationship with *bad*; the concept `bad-flower` is extracted.

(2) The flower smells bad.

Experiments and Results To calculate the performance, we selected 300 sentences from the *Stanford Sentiment Dataset* [6] and extracted the concepts manually, which gave 3204 concepts. On these sentences, our parser achieved 92.01% accuracy.

1.2 Aspect Parser

Aspect-based opinion mining aims to model relations between the polarity of a document and its opinion targets, or aspects. Our system is able to extract both implicit and explicit aspects.

Compilation of an implicit aspect lexicon We used the product review dataset described in [7] to create the implicit aspect lexicon.

We selected from it the sentences that had implicit aspects. From those sentences, we extracted the implicit aspect terms and manually labeled them with suitable categories. For example, from the sentence *The car is expensive* we extracted the implicit aspect term *expensive* and labeled it with the category *price*. We identified in this corpus the following categories: *functionality*, *weight*, *price*, *appearance*, *behavior*, *performance*, *quality*, *service*, *size*.

For each identified implicit aspect term, we retrieved its synonyms from the WordNet. This gave us a lexicon of 1128 implicit aspect terms labeled by aspect categories listed above.

Opinion Lexicon We used SenticNet 3.0 [8,9,10,11] as an opinion lexicon. It contains 14,000 common sense-knowledge concepts labeled by their polarity scores.

Algorithm We used the Stanford parser to obtain the dependency parse structure of each sentence. Then we employed a complex system of hand-crafted rules on these parse trees to extract the aspects. Examples of such rules are given below. Note that some of our rules can block the application of other rules, so the rules given below are not always applied.

Subject Noun Rule If the active token *h* is in a subject noun relationship with a word *t*, then:

1. if *t* has an adverbial or adjective modifier that exists in the SenticNet, then we extract *t* as an aspect. E.g., in (3), according to Stanford parser, *it* is in a subject noun relationship with *camera*, which has an adjective modifier *nice*, so *camera* is extracted.

(3) It is a nice camera.

2. if the sentence has no auxiliary verb (*is*, *was*, *would*, *should*, *could*, etc.), then:
   - if *t* is a verb modified by an adjective or adverb or it is in *adverbial clause modifier* relation with another token, then both *h* and *t* are extracted as aspects. E.g., in (4), *battery* is in a subject relation with *lasts*, so the aspects *last* and *battery* are extracted.

     (4) The battery lasts little.

   - if *t* has a noun *n* as a direct object and *n* is not in SenticNet, then *n* is extracted as an aspect. E.g., in (5), *like* is in direct object relation with *lens*, so the aspect *lens* is extracted.

     (5) I like the lens of this camera.

   - if *t* has a noun *n* as a direct object, *n* is in SenticNet, and *n* is in a prepositional relation with another noun *m*, then both *n* and *m* are extracted as aspects. E.g., in (6), *like* is in direct object relation with *beauty*, which is connected to *screen* via a preposition relation. So the aspects *screen* and *beauty* are extracted.

     (6) I like the beauty of the screen.

3. Copula is the relation between the complement of a copular verb and the copular verb. If the token *h* existing in the implicit aspect lexicon is in a copula relation with a couplar verb, then we extract *h* as an aspect. E.g., in (7) *expensive* is extracted as an aspect.

   (7) The car is expensive.

Sentences with no subject noun relation in the parse tree We extracted the aspects from such sentences using the following rules:

1. if an adjective or adverb *h* is in infinitival or open clausal complement relation with a token *t* and *h* exists in the implicit aspect lexicon, then we extract *h* as an aspect. E.g., in (8) we extract *big* as an aspect, since it is connected to *hold* via a clausal complement relation.

   (8) Very big to hold.

2. if a token *h* is connected to a noun *t* via a prepositional relation, then we extract both *h* and *t* as aspects. E.g., in (9), *sleekness* is extracted as an aspect.

   (9) Love the sleekness of the player.

Obtaining implicit aspect categories After obtaining the aspects using these rules, we retrieved the categories of the implicit aspects from the implicit aspect lexicon.

Experiments and Results We experimented on the Semeval 2014 aspect-based sentiment analysis data.[6] On this dataset, we obtained 91.25% precision and 88.12% recall.

2 Common Sense Knowledge-based Sentiment Analysis

This section describes the algorithm we used to compute the polarity score of a sentence. We used an ensemble algorithm for detecting sentiment in the sentences. Below we describe some rules we used and the ensemble classification process.

2.1 Dependency Rules

We used rules based on specific dependency patterns to drive the way concepts were searched in SenticNet. Below are some examples of such rules.

Subject nouns This rule is applied when the active token *h* is the syntactic subject of a word *t*. If the complex concept *t-h* was found in SenticNet, then it was used to calculate the polarity of the relation (otherwise, other rules are activated later). E.g., in (10), *movie* is in a subject relation with *boring* and `(boring-movie)` is in SenticNet, so its corresponding polarity was used.

(10) The movie is boring.

Adjective and clausal complements These rules deal with verbs having as complements either an adjective or a closed clause (i.e., a clause, usually finite, with its own subject).

Modifiers Modifiers, by definition, affect the interpretation of the head they modify, so in most of our rules the dependent is the guiding element for computing the polarity.

2.2 Machine Learning Technique

For each sentence, we extracted concepts from it and looked them up in SenticNet. If we found at least one concept in SenticNet, then we used our knowledge-based method to detect sentiment. Otherwise, we relied on our machine learning-based technique. The Machine Learning module was trained on the Blitzer dataset. Below we describe some of the features we used for training.

Sentic feature The polarity scores of each concept extracted from the sentence were obtained from the SenticNet and summed up to produce a single scalar feature. This feature was used for training, but as described above, was not available in testing.

Part-of-speech features This feature was defined by the numbers of adjectives, adverbs, and nouns in the sentence, which gave three distinct features.

[6] `http://alt.qcri.org/semeval2014/task4/index.php?id=data-and-tools`

Modification feature This binary feature was set to 1 if we found any modification relation in the sentence; otherwise it was set to 0.

2.3 Results on the Blitzer-derived Dataset

At the sentence level, on the Blitzer dataset 87.00% accuracy was achieved.

3 Demos

Online demos of aspect and concept parsing and the sentiment analysis are available on `http://www.soujanyaporia.com/software/`.

References

1. Poria, S., Gelbukh, A., Hussain, A., Howard, N., Das, D., Bandyopadhyay, S.: Enhanced SenticNet with affective labels for concept-based opinion mining. Intelligent Systems, IEEE **28**(2) (2013) 31–38
2. Cambria, E., Rajagopal, D., Olsher, D., Das, D.: Big social data analysis. Big Data Computing (2013) 401–414
3. Cambria, E., Hussain, A., Havasi, C., Eckl, C., Munro, J.: Towards crowd validation of the UK national health service. WebSci10 (2010)
4. Poria, S., Agarwal, B., Gelbukh, A., Hussain, A., Howard, N.: Dependency-based semantic parsing for concept-level text analysis. In: Proceedings of the 15th International Conference on Intelligent Text Processing and Computational Linguistics, CICLing 2014, Part I. Volume 8403 of Lecture Notes in Computer Science., Springer (2014) 113–127
5. Cambria, E., Poria, S., Gelbukh, A., Kwok, K.: Sentic API: A common-sense based API for concept-level sentiment analysis. In Rowe, M., Stankovic, M., Dadzie, A.S., eds.: Proceedings of the 4th Workshop on Making Sense of Microposts (#Microposts2014), co-located with the 23rd International World Wide Web Conference (WWW 2014), Seoul, Korea, April 7th, 2014. Volume 1141 of CEUR Workshop Proceedings (CEUR-WS.org)., CEUR (2014) 19–24
6. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Conference on Empirical Methods in Natural Language Processing (EMNLP). (2013)
7. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. Computational linguistics **37**(1) (2011) 9–27
8. Cambria, E., Olsher, D., Rajagopal, D.: Senticnet 3: A common and common-sense knowledge base for cognition-driven sentiment analysis. AAAI (2014)
9. Poria, S., Gelbukh, A., Cambria, E., Yang, P., Hussain, A., Durrani, T.: Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis. In: 2012 IEEE 11th International Conference on Signal Processing (ICSP). Volume 2., IEEE (2012) 1251–1255
10. Poria, S., Gelbukh, A., Cambria, E., Das, D., Bandyopadhyay, S.: Enriching SenticNet polarity scores through semi-supervised fuzzy clustering. In: 2012 IEEE 12th International Conference on Data Mining Workshops (ICDMW), IEEE (2012) 709–716
11. Poria, S., Gelbukh, A., Das, D., Bandyopadhyay, S.: Fuzzy clustering for semi-supervised learning—case study: Construction of an emotion lexicon. In: Advances in Computational Intelligence. Proceedings of MICAI 2012. Volume 7629 of Lecture Notes in Artificial Intelligence., Springer (2013) 73–86