NILC_USP: Aspect Extraction using Semantic Labels

Pedro P. Balage Filho and **Thiago A. S. Pardo**
Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo
São Carlos - SP, Brazil
`{balage, taspardo}@icmc.usp.br`

Abstract

This paper details the system NILC_USP that participated in the Semeval 2014: Aspect Based Sentiment Analysis task. This system uses a Conditional Random Field (CRF) algorithm for extracting the aspects mentioned in the text. Our work added semantic labels into a basic feature set for measuring the efficiency of those for aspect extraction. We used the semantic roles and the highest verb frame as features for the machine learning. Overall, our results demonstrated that the system could not improve with the use of this semantic information, but its precision was increased.

1 Introduction

Sentiment analysis, or opinion mining, has gained lots of attention lately. The importance of this field of study is linked with the grown of information in the internet and the commercial attention it brought.

According to Liu et al. (2010), there are two kinds of information available in the internet: facts and opinions. Facts are objective statements about entities and events in the world. Opinions are subjective statements that reflect people's sentiments or perceptions about the entities and events. According to Liu, by that time, there was a lot of attention on the processing of facts but little work had been done on the processing of opinions.

Three levels of analysis for sentiment analysis are known (Liu, 2012): document level, sentence level and aspect level. The aspect-based sentiment analysis is the name of the research topic that aims to extract the sentiments about the aspects present in the text.

This work presents a system evaluated in the *SemEval Task4: Aspect Based Sentiment Analysis* shared task (Pontiki et al., 2014). Our system participated only in subtask 1: Aspect Term Extraction. In this subtask, given a text, the system should extract all aspects that are present. There were two different domains for this task: restaurants and laptops.

The goal of our system was to verify how semantic labels used in machine learning classification would improve the aspect extraction task. For this goal, we used two kinds of semantic labels: the semantic roles (Palmer et al., 2010) and the semantic frames (Baker et al., 1998).

Liu et al. (2012) categorizes the works for aspect extraction in four types, regarding the approach they follow, using: frequent terms, infrequent terms, machine learning, and topic modeling. This work uses a machine learning approach that consists in training a sequential labeling algorithm for aspect detection and extraction.

In what follows, we present some related work in Section 2. Section 3 and 4 introduce our system and report the achieved results. Some conclusions are presented in Section 5.

2 Related work

Jin and Hovy (2009) reported one the first works using sequential labeling for aspect extraction. In this work, the authors used a Lexicalized Hidden Markov Model to learn patterns to extract aspects and opinions. Jakob and Gurevych (2010) trained

a Conditional Random Field for aspect extraction. In this work, the authors report the results for a single domain and a cross domain experiment. They show that even in other domains the method could be good.

Kim and Hovy (2006) explored the semantic structure of a sentence, anchored to an opinion bearing verb or adjective. Their method uses semantic role labeling as an intermediate step to label an opinion holder and topic using data from FrameNet.

Houen (2011) presented a system for opinion mining with semantic analysis. The author explores the use of the semantic frame-based analyzer FrameNet (Baker et al., 1998) for modeling features in a machine learning approach. The author found that the FrameNet information was not helpful in this classifier.

3 System Description

Our system uses a sequential labeling algorithm. In our work, we use the Conditional Random Field (Lafferty et al., 2001) algorithm provided by the CRF++ tool[1].

For training the sequential labeling algorithm, we give as input features for each word in the corpus. The algorithm will then learn how to classify those words. In our approach, the possible classes are: True, representing an aspect word; and False, representing the remaining words.

The goal of our system was to evaluate the performance of the semantic labels for the task. In order to model our system, we built a feature set consisting of 6 features.

1. the word

2. the part-of-speech

3. the chunk

4. the named-entity category

5. the semantic role label (SRL)

6. the most generic frame in FrameNet

The use of the first four features is consistent with the best approaches in aspect-based sentiment analysis. The last two features are the ones we are testing in our work.

In order to extract the features, we used two important tools: the Senna (Collobert et al., 2011), a

---

[1]Available at http://crfpp.googlecode.com/

semantic role labeling system, and the ARK SE-MAFOR, a Semantic Analyzer of Frame Representations (Das et al., 2010).

The Senna system uses a deep learning neural network (Collobert, 2011) to provide several predictions for natural language processing. The system output is represented in the CONLL format, the same used in CRF++.

Our first 5 features were directly provided by the Senna output. In these features, we decided to keep the IOBE information since the initial experiments showed the results were better with it than without.

Our fifth feature, the semantic role label, was retrieved from Senna as well. In the corresponding paper, they reported Senna could achieve a F1 of 75% for the SRL task.

The example below shows how the features were represented. In this example, we are only showing four features: the word, the part-of-speech, the chunk and the SRL. The classes are in the last column.

```
WORD      POS   CHUNK   SRL   IS_ASPECT?
Great     JJ    B-NP    B-A0     False
laptop    NN    E-NP    E-A0     False
that      WDT   S-NP    S-R-A0   False
offers    VBZ   S-VP    S-V      False
many      JJ    B-NP    B-A1     False
great     JJ    I-NP    I-A1     False
features  NNS   E-NP    E-A1     True
!         .     O       -        False
```

The last feature was retrieved by ARK SE-MAFOR tool. ARK SEMAFOR uses a probabilistic frame-semantic parsing using the FrameNet resource. The ARK SEMAFOR output is the analysis of the frames present in the text for a given verb. As our feature set has only word related features, we decided to use the most upper level structure in the frame. In case of multiple verbs in the sentence, we used the structure for the verb that is closest to the word of interest.

The following example shows how the frames were added into the training model. We limit to show only the word, frame and the class. For training, we used the full training set with the six features plus the class.

```
WORD       FRAME                IS_ASPECT?
I          Shopping                  False
shopped    Shopping                  False
around     Relational_quantity  False
before     Relational_quantity  False
```

```
buying  Relational_quantity False
.       O                     False
```

The organization from *SemEval-2014 Task 4: Aspect Based Sentiment Analysis* provided two domains for evaluation: restaurants and laptops. For each domain, the organization provided three datasets: a trainset, a devset and a testset.

We executed our algorithm with C parameter equal to 4.0. The experiment code is fully available at the weblink https://github.com/pedrobalage/

4   Results

Tables 1 and 2 show our system results for the restaurants and laptops domains respectively. In these tables, the results are discriminated by the feature sets that were used. The reader may see that a "+ Frame" system, for example, stands for all the features discriminated above (Word, POS, Chunk, NR, SRL) plus the Frame feature. The last line shows the results scored by our system in the SemEval shared task with all the features. We also show the results for the baseline system provided by the shared task (Pontiki et al., 2014).

Table 1: Results for restaurants domain

System	Precision	Recall	F1-mesaure
Baseline	52.54	42.76	47.15
Word + POS	83.76	68.69	75.48
+ Chunk	83.38	68.16	75.01
+ NE	83.45	68.07	74.98
+ SRL	82.79	67.46	74.34
+ Frame	**87.72**	**34.03**	**49.04**

Table 2: Results for laptops domain

System	Precision	Recall	F1-mesaure
Baseline	44.31	29.81	35.64
Word + POS	80.87	39.44	53.03
+ Chunk	78.83	39.29	52.44
+ NE	79.93	39.60	52.96
+ SRL	78.22	38.99	52.04
+ Frame	**83.62**	**14.83**	**25.19**

Comparing with the baseline, we may noticed that our submitted system (+Frame) outperformed the baseline for the restaurants domain but it did not outperformed the baseline for the laptops domain (considering F1 mesaure).

When we look in detail for the inclusion of features in our feature set, we may notice that, at every new feature, the precision goes up, but the recall goes down. We believe this is due to the behaviour of the conditional random field algorithm for compensating for a sparser feature set.

In general, the semantic labels (SRL and Frame) could not improve the results. However, if we are interested only on precision, these features are helpful. This may be the case in scenarios where a lot of information is available, as in the web, and we want to be sure about the retrieved information. Certainly, there is a conflict between precision and computational complexity, since the semantic features are more expensive to be achieved (in relation to the usual simpler features that may be used).

Despite of that, we judge to be necessary to conduct more experiments in order to better evaluate the impact of semantic labels in the aspect extraction task.

5   Conclusion

We presented an aspect extraction system built on a conditional random field algorithm. We used a rich feature set with the semantic roles and the FrameNet upper frames for each word. We have showed that the semantic labels may help to achieve a more precise classifier, but it did not help to improve the overall F-measure of the system.

Regarding the shared task, our system achieved the second best precision value among the competing systems, but the lowest recall value. Future work should investigate ways of also improving recall without penalty for the achieved precision.

Acknowledgments

References

Collin F Baker, Charles J Fillmore, and John B Lowe. 1998. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics-Volume 1*, pages 86–90. Association for Computational Linguistics.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from

435

scratch. *The Journal of Machine Learning Research*, 12:2493–2537.

Ronan Collobert. 2011. Deep learning for efficient discriminative parsing. In Geoffrey J. Gordon and David B. Dunson, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS-11)*, volume 15, pages 224–232. Journal of Machine Learning Research - Workshop and Conference Proceedings.

Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A Smith. 2010. Semafor 1.0: A probabilistic frame-semantic parser. Technical report, Language Technologies Institute, School of Computer Science, Carnegie Mellon University.

Søren Houen. 2011. *Opinion Mining with Semantic Analysis*. Ph.D. thesis, Department of Computer Science, University of Copenhagen.

Niklas Jakob and Iryna Gurevych. 2010. Extracting opinion targets in a single-and cross-domain setting with conditional random fields. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1035–1045.

Wei Jin and Hung Hay Ho. 2009. A Novel Lexicalized HMM-based Learning Framework for Web Opinion Mining. In Léon Bottou and Michael Littman, editors, *Proceedings of International Conference on Machine Learning (ICML-2009)*, ICML '09, pages 1–8. ACM, ACM Press.

Soo-Min Kim and Eduard Hovy. 2006. Extracting opinions, opinion holders, and topics expressed in online news media text. In *Proceedings of the Workshop on Sentiment and Subjectivity in Text*, SST '06, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, ICML '01, pages 282–289, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Bing Liu. 2010. Sentiment Analysis and Subjectivity. In N Indurkhya and F J Damerau, editors, *Handbook of Natural Language Processing*, number 1, chapter 28, pages 627–666. CRC Press, Taylor and Francis Group, Boca Raton.

Bing Liu. 2012. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167.

Martha Palmer, Daniel Gildea, and Nianwen Xue. 2010. Semantic role labeling. *Synthesis Lectures on Human Language Technologies*, 3(1):1–103.

Maria Pontiki, Dimitrios Galanis, John Pavlopoulos, Haris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation*, SemEval 2014, Dublin, Ireland.