Target Focused Sentiment Extraction Framework

Gregorios A Katsios*[1], Meenakshi Alagesan†[1], Arun Sharma‡[1] and Tomek Strzalkowski§[1]

[1]ILS Institute and Department of Computer Science, SUNY at Albany

Abstract

Albany's *Target Focused Sentiment Extraction Framework* consists of a step-wise approach to analyze the sentiment components incorporated into the social aspects of dialogue. We build an extensible modular framework that extracts sentiment from natural language text, along with the sentiment holder and the sentiment target. It improves our previous work [DWW+16] regarding sentiment extraction by tackling the problem of generating proper sentiment holder and target pairs, thus leading to an increased Recall measure.

1 Introduction

Sentiment is observable in language through interaction dynamics and semantic role modeling, where it manifests as feelings, attitudes, emotions or opinions. These manifestations capture the subjective impressions which usually assumes a binary opposition in opinions such as *good/bad*, *like/dislike*, *for/against*, etc.

Our goal is to determine the attitude a speaker has towards another person, an object, an event or other appropriate entities. Being able to detect and extract sentiment from a discussion, enables us to predict and determine the participants behaviour. Additionally, it enables us to conduct an in depth study of idea propagation through groups, since opinions and reactions to ideas are relevant to adoption of new ideas. Analyzing sentiment reactions on blogs can give insight to this process [KKJ+07]. It also posses the potential to augment psychological investigations/experiments with data extracted from natural language text, such as dream sentiment analysis [NSDK+06]. In general, humans are subjective creatures and opinions are important.

2 Data

During this evaluation we used the 2016 Sentiment training and test data as our current training data, assuming that they can be treated equally. We focused on *English* documents, so the combined documents have the following properties, as shown in Table 1. Table 2 includes the properties of the 2017 test data.

All data sets are provided by the *Linguistic Data Consortium* [1] (LDC) and are annotated with entities, relations and events (ERE). The *ground truth*, sentiment polarity among annotated pairs, was provided for the training data.

Type/#	Files	EREs	Mentions	Positive	Negative
DF	283	19521	42517	2283	4992
NW	118	10403	17793	408	535
Total	401	29924	60310	2691	5527

Table 1: Training data properties.

*gkatsios@albany.edu
†malagesan@albany.edu
‡asharma5@albany.edu
§tomek@albany.edu

[1]https://www.ldc.upenn.edu/

Type/#	Files	EREs	Mentions
DF	84	5649	10821
NW	83	5715	9725
Total	167	11364	20546

Table 2: Test data properties.

3 Method

Our approach automatically extracts sentiment from natural language text, along with the sentiment holder and the sentiment target. Since we focus on extraction of individual instances of sentiment, our first step is to identify sentiment triples of the form $\langle source, relation, target \rangle$. Then, as our second step we identify the triples that contain entities already annotated in the data. As a third step, we estimate the polarity and intensity of sentiment from the source towards the target.

3.1 Identifying Sentiment Triples

Sentiment triples contain the necessary information required to extract target focused sentiment, without the noise introduced by having multiple conflicting opinions per sentence. In essence, we require a shallow semantic representation of larger amounts of text in the form of verbs or verbal phrases and their arguments. This task falls under the scope of *Open Information Extraction* (OIE), which is a well studied area and multiple successful approaches exist.

In our framework, we use ClausIE [DCG13], which given an input sentence, generates relation triples of the form: $\langle subject, relation, arguments \rangle$. Specifically, relation triples are extracted from clauses, which are identified based on the results from the dependency parser that helps to reveal the syntactic structure of the input sentence. In particular, this tool uses Stanford unlexicalized dependency parser [KM03]. ClausIE is similar to Ollie [SBS+12] and ReVerb [FSE11], and it was preferred over those due to it's increased performance.

We made several modifications to the original implementation of ClausIE in order to adapt it to the task of identifying sentiment triples. A *subject* in each relation triple is a potential source for reported sentiment and the *object* is a potential target. The *relation* is the verb with target as an argument. Also, the *speaker/writer* is always a potential source. In such case, any entity, relation or event in the text attributed to this *speaker/writer* is a potential target, unless it is in a segment expressly attributed to another source.

Three possible types of sentiment relations are defined (*agentive*, *patientive* and *propertive*), which are determined based on the role of the target. This is determined by syntactic information obtained from the same dependency parse used to identify the clauses.

- Agentive: the way target acts or affects other things.
  - Examples: crushing, helps, adapts, etc.
- Patientive: the way to deal with Target or to affect it.
  - Examples: navigate, fight, donate to, etc.
- Propertive: the way Target appears, looks, smells, sounds, feels, etc.
  - Examples: heavy, harmful to, affordable, etc.

The relation is *agentive* when the target entity is in the agent role (typically the subject); conversely, the relation is *patientive* when the target is in a patient role. A *propertive* relation is when a property of the target is described, typically in a unary relation often anchored at an adjective.

3.2 Selection of Sentiment Triples

The goal of our system is to extract all instances of sentiment between any potential source and a potential target mentioned in a text document. The source and target of the sentiment are restricted to *ERE* entities (or relations or events), and may come from *Gold* or automatically generated annotations. Towards that goal, we select the triples that have *ERE* entities fulfilling the roles described above.

	Agentive		Patientive		Propertive
Rel. Polarity	positive(x)	negative(x)	positive(x)	negative(x)	-
Positive	positive	negative	positive	negative	positive
Negative	negative	positive	neutral	positive	negative
Neutral	neutral	negative	neutral	negative	neutral

Table 3: Affect Calculus Algorithm: A simple affect calculus specifies affect polarity towards a target as an argument of a affect carrying relation, where 'x' is the argument of the relation.

3.3 Sentiment Estimation

We combine the information about syntactic and semantic structure of a sentence with base polarity values of words and phrases in order to estimate polarity and intensity of sentiment from the source towards the target.

The information about the syntactic and semantic structure of a sentence is captured by the identification of sentiment triples step described at Section 3.1. The instantiated sentiment triple is passed to *Affect Calculus Algorithm* (ACA) [SSC+14] for sentiment determination. To achieve this, we adapted the ACA which was originally designed to compute affect in metaphors.

The base polarity and strength values of words and phrases that are required to calculate the affect score towards a target entity are assigned based on the expanded *Affective Norms in English* (ANEW+) lexicon [SCS+16].

These scores are combined based on the type of relation with respect to the target, using the *Affect Calculus* as shown in Table 3. This way we obtain the value of sentiment towards the target in the $\langle source, relation, target \rangle$ triple.

3.3.1 Machine Learning Classification

To further improve the quality of the ACA output, we implement a machine learning approach to classify the sentiment the source has towards the target. Specifically, in our submitted systems we use *Support Vector Machine* (SVM) [Pla98, KSBM01] and *Naive Bayes* [JL95] classifiers as well as no classifier, avoiding this last step. The WEKA [WFHP16] toolkit was used to implement the SVM and Naive Bayes classifiers, considering as features:

- Relation polarity value provided by ANEW+,
- Source and Target type provided by Ere annotations,
- Part-of-Speech and word n-grams (n = {2, 3, 4}),
- ACA output.

4 Results

In this section we present our experimental results and they include some observations on our current training set. Specifically, 10-fold cross validations on our data, yield *F-Measure* values of 0.83 with SVM and 0.68 with Naive Bayes classifiers respectively, shown on Figure 4.

The following results are based on the 2017 test set, focusing on English texts and Gold-ERE files. When the trained classifiers are applied on the 2017 test set, we obtain the results presented in Table 5. Additionally, during the evaluation we observed differences in the output confidence of our classifiers. Figure 1 illustrate the confidence levels, where y axis is the confidence level and x represents each classification.

Compared with the 2016 BeSt evaluation on English texts, our newer system outperforms the previous in terms of recall. However, that comes with a sacrifice in our accuracy. Table 6 shows the evaluated performance of our previous (2016 evaluation) and current (2017 evaluation) systems.

Classifier	Precision	Recall	F-Measure
SVM	0.839	0.834	0.834
Naive Bayes	0.694	0.684	0.686

Table 4: Evaluated performance of classifiers.

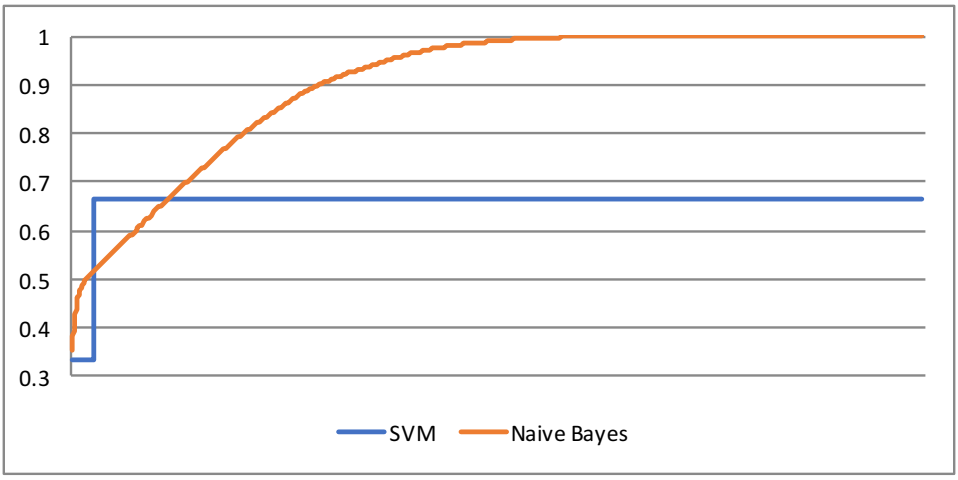

Figure 1: 2017 evaluation results: Confidence for SVM and Naive Bayes classifiers.

	SVM		**Naive Bayes**		**No Classifier**	
	Positive	Negative	Positive	Negative	Positive	Negative
DF	1725	4461	3730	8250	4387	1327
NW	545	1187	781	1411	4210	805
Total	2270	5648	4511	9661	8597	2132

Table 5: 2017 evaluation results: Count of estimated polarity.

	Precision	**Recall**	**F-Measure**
2016(DF)	0.138	0.165	0.151
2016(NW)	0.046	0.018	0.026
2017(DF)	0.094	0.4	0.153
2017(NW)	0.042	0.184	0.068

Table 6: 2016-2017 evaluation results.

5 Acknowledgements

This work was supported in part by the Defense Advanced Research Projects Agency (DARPA) as part of the DEFT Program.

6 Conclusion

Albany's current system improved since the previous evaluation in terms of recall, as we put effort to curb the very high miss rate of the previous version. However, this came at the cost of somewhat lower precision, which we believe is related to the marginal polarity values (i.e., the values near the "neutral zone") in the polarity lexicon.

One clear piece of future work is to determine the best range of values to consider in the neutral zone from the range of valence scores in ANEW lexicon. Using an optimized range will maximize performance. Furthermore, this value appears to vary by context and genre of text, and data driven optimization may be appropriate. We have not used machine learning to adjust ANEW polarity values to the text genre, which would have likely improved precision, but resulted in an unrealistic assessment of system capabilities, since such results would not transfer to other genres, or even other types of topics. However, making the polarity lexicon more flexible and adaptable is clearly an avenue to explore.

References

[DCG13] Luciano Del Corro and Rainer Gemulla. Clausie: clause-based open information extraction. In *Proceedings of the 22nd international conference on World Wide Web*, pages 355–366. ACM, 2013.

[DWW+16] Adam Dalton, Morgan Wixted, Yorick Wilks, Meenakshi Alagesan, Gregorios Katsios, Ananya Subburathinam, and Tomek Strzalkowski. Target-focused sentiment and belief extraction and classification using cubism. 2016.

[FSE11] Anthony Fader, Stephen Soderland, and Oren Etzioni. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics, 2011.

[JL95] George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers. In *Eleventh Conference on Uncertainty in Artificial Intelligence*, pages 338–345, San Mateo, 1995. Morgan Kaufmann.

[KKJ+07] Anubhav Kale, Pranam Kolari, Akshay Java, Tim Finin, and Anupam Joshi. On modeling trust in social media using link polarity1. Technical report, University of Maryland, Baltimore County, 2007.

[KM03] Dan Klein and Christopher D Manning. Accurate unlexicalized parsing. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pages 423–430. Association for Computational Linguistics, 2003.

[KSBM01] S.S. Keerthi, S.K. Shevade, C. Bhattacharyya, and K.R.K. Murthy. Improvements to platt's smo algorithm for svm classifier design. *Neural Computation*, 13(3):637–649, 2001.

[NSDK+06] David Nadeau, Catherine Sabourin, Joseph De Koninck, Stan Matwin, Peter D Turney, et al. Automatic dream sentiment analysis. In *In: Proceedings of the Workshop on Computational Aesthetics at the Twenty-First National Conference on Artificial Intelligence, Boston, Massachussetts, USA*, 2006.

[Pla98] J. Platt. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press, 1998.

[SBS+12] Michael Schmitz, Robert Bart, Stephen Soderland, Oren Etzioni, et al. Open language learning for information extraction. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 523–534. Association for Computational Linguistics, 2012.

[SCS+16] Samira Shaikh, Kit Cho, Tomek Strzalkowski, Laurie Feldman, John Lien, Ting Liu, and George Aaron Broadwell. Anew+: Automatic expansion and validation of affective norms of words lexicons in multiple languages. In *LREC*, 2016.

[SSC+14] Tomek Strzalkowski, Samira Shaikh, Kit Cho, George Aaron Broadwell, Laurie Feldman, Sarah Taylor, Boris Yamrom, Ting Liu, Ignacio Cases, Yuliya Peshkova, et al. Computing affect in metaphors. In *Proceedings of the Second Workshop on Metaphor in NLP*, pages 42–51, 2014.

[WFHP16] Ian H Witten, Eibe Frank, Mark A Hall, and Christopher J Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.