Project Proposal : Learning Fine-grained Opinions

Arzoo Katiyar   AK979@CORNELL.EDU
Cornell University

Fine-grained opinion analysis task is about identifying opinions in the text at the expression level which includes – identifying the opinion expression itself along with the opinion holder and the target (Wiebe & Cardie, 2005). This task is useful for any application which may be concerned about what does an entity think about the target entity. Kim et al.(Kim & Hovy, 2006) solve a similar problem where they extract opinions, opinion holders, and topics expressed in online news media text. However, they decompose this problem into three phases: identifying an opinion-bearing word, labeling semantic roles related to the word in a sentence, and then finding the holder and topic of the opinion word among the labeled semantic roles. However, we would like to be able to jointly infer the opinion expression, the holder and the targets.

Another related task is that of Semantic Role Labeling. (Màrquez et al., 2008) state that the sentence-level semantic analysis of text is concerned with the characterization of events, such as determining "who" did "what" to "whom", "where", "when" and "how". The predicate of a clause establishes "what" took place and other sentence constituents express the participants in the event(such as "who" and "where") as well as further event properties. In practice, semantic role labeling task is generally carried out in two phases : Semantic Frame Identification and the argument identification task. Noticeably, (O'Connor, 2013) have used an unsupervised latent variable model to discover such semantic frames. And in the past, there have been attempts on learning fine-grained opinions based on the semantic role labeling techniques which perhaps indicate that an appropriate(similar?) latent variable model might be useful for our task.

Another related task to learning fine-grained opinions is to jointly extract aspects and sentiments for online reviews. It is direct to see that this task is similar to our task as the reviewer is the holder of the opinion and the aspects are the targets. (Kim et al., 2013) have recently shown that using a bayesian nonparametric model, recursive Chinese Restaurant Process (rCRP) as the prior they can jointly infer the aspect-sentiment tree from the review texts.

We would just briefly like to mention another interesting but probably not-so related work in Entity Topic Models. (Kim et al., 2012) build more expressive, entity-based topic models, which can capture the term distributions for each topic, each entity as well as each topic-entity pair.

The task we proposed is hard due to the way it is defined because it requires to model fine-grained syntactic as well as semantic relations between the entities. Also, most of the previous work do not jointly extract the opinion expression and the entities together, instead this task is carried out in phases. However, (Yang & Cardie) provide an ILP-based approach for the joint-inference of opinion extraction task. Most of these approaches use features from dependency-parse trees and these features have been proved to be useful. However, it seems worthwhile to explore this task using latent variable models since the state-of-the-art of this task especially for finding opinion targets (opinion target : 43.07%; opinion expression : 59.79%; opinion holder : 62.47%) can be improved. We did not find any related work for this problem that uses latent variable models except for the ones mentioned before. Since those tasks are related and the models used were applicable for those tasks, it seems worthwhile to apply similar models for this problem as well.

In order to evaluate our method, we can use the MPQA data used by (Yang & Cardie) and compare against their results. Since this problem and other related problem has been of interest to the research community, we can hope that the results will be helpful for the progress of the field with respect to this task.

References

Kim, Hyungsul, Sun, Yizhou, Hockenmaier, Julia, and Han, Jiawei. Etm: Entity topic models for mining documents associated with entities. In *Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, ICDM '12, pp. 349–358, Washington,

DC, USA, 2012. IEEE Computer Society. ISBN 978-0-7695-4905-7. doi: 10.1109/ICDM.2012.107. URL http://dx.doi.org/10.1109/ICDM.2012.107.

Kim, Soo-Min and Hovy, Eduard. 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, pp. 1–8, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics. ISBN 1-932432-75-2. URL http://dl.acm.org/citation.cfm?id=1654641.1654642.

Kim, Suin, Zhang, Jianwen, Chen, Zheng, Oh, Alice H., and Liu, Shixia. A hierarchical aspect-sentiment model for online reviews. In *AAAI*, 2013.

Màrquez, Lluís, Carreras, Xavier, Litkowski, Kenneth C., and Stevenson, Suzanne. Semantic role labeling: An introduction to the special issue. *Comput. Linguist.*, 34(2):145–159, June 2008. ISSN 0891-2017. doi: 10.1162/coli.2008.34.2.145. URL http://dx.doi.org/10.1162/coli.2008.34.2.145.

O'Connor, Brendan. Learning frames from text with an unsupervised latent variable model. *CoRR*, abs/1307.7382, 2013. URL http://arxiv.org/abs/1307.7382.

Wiebe, Janyce and Cardie, Claire. Annotating expressions of opinions and emotions in language. language resources and evaluation. In *Language Resources and Evaluation (formerly Computers and the Humanities*, pp. 2005, 2005.

Yang, Bishan and Cardie, Claire. Joint inference for fine-grained opinion extraction.