Visual Active Learning with Distant Supervision for Relation Extraction

Kairong Jiang
University of Arizona
jiangkairong@email.arizona.edu

Mihai Surdeanu
University of Arizona

1 Introduction

Relation extraction is an import part of information extraction (IE) where a classifier is trained to label the *relation* between two entity mentions in one sentence. For example, in the sentence "**Obama** *was born in* the **United States** just as he has always said.", the classifier should lable relation *"BornIn"* with entity mention pair "Barack Obama" and "United States".

While supervised learning methods have been developed for relation extraxtion tasks, they typiaclly require large amount of annotated training data to perform competitively. It is likely in real world problems that annotated data is limited or expensive to acquire. Therefore, it is benifitial to look for active learning methods that exploit a few informative annotated data and achieve resonable performance while greatly reduce the work needed for human annotators.

On the other hand, existing active learning methods (Angeli et al., 2014; Fu and Grishman, 2013; Sun and Grishman, 2012) for relation extraction mainly focus on selecting the sampling strategies and improving the active learning model. While they have achieved notable improvements, the effectiveness and efficiency of human interactions are largely neglected in the aforementioned works. Human annotation is often simulated with fully-labeled data (Fu and Grishman, 2013; Sun and Grishman, 2012), or conducted using a listed multiple-choice view (Angeli et al., 2014).

In this paper, we present a relation extraction system implementing a distantly supervised model from (Surdeanu et al., 2012) and active learning strategy from (Angeli et al., 2014) with a 2D scatterplot visual interface similar to (Berger et al., 2018) for human annotators, and we conduct user studies to show that carefully designed and implemented visual interface can further improve the effectiveness and efficiency of active learning methods for relation extraction, primarily thanks to greater number of annotations that can be done with the same human effort. We also experiment with several sampling methods outlined in (Angeli et al., 2014) and (Berger et al., 2018) and explore the best sampling strategy suitable for the 2D scatterplot interface.

We are able to achieve following contributions through our studies:

- We present a 2D scatterplot visual interface for the human annotation process of relation extraction active learning model, which, despite lower accuracy, increases the number of annotations that can be made with the same human effort, hence imporves the overall performance of the system.
- We experiment with different sampling methods and acquire understandings on the strong sides as well as draw backs of those methods, providing insights on sampling method selection with active learning using 2D scatterplot interface.

References

Gabor Angeli, Julie Tibshirani, Jean Wu, and Christopher D Manning. 2014. Combining distant and partial supervision for relation extraction. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1556–1567.

Matt Berger, Ajay Nagesh, Joshua Levine, Mihai Surdeanu, and Hao Helen Zhang. 2018. Visual supervision in bootstrapped information extraction. In *To Be Published*.

Lisheng Fu and Ralph Grishman. 2013. An efficient active learning framework for new relation types. In

Proceedings of the Sixth International Joint Conference on Natural Language Processing, pages 692–698.

Ang Sun and Ralph Grishman. 2012. Active learning for relation type extension with local and global data views. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1105–1112. ACM.

Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D Manning. 2012. Multi-instance multi-label learning for relation extraction. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, pages 455–465. Association for Computational Linguistics.