

Visual Active Learning for Relation Extraction

Kairong Jiang

University of Arizona

jiangkairong@email.arizona.edu

Mihai Surdeanu

University of Arizona

1 Introduction

Relation extraction is an important part of information extraction (IE) where a classifier is trained to label the *relation* between a set of entity mentions in some text. It provides crucial information for later IE processes like disambiguation. For example, in the sentence “**Obama** was born in the **United States** just as he has always said.”, the classifier should label relation “*BornIn*” with entity mention pair “Barack Obama” and “United States”.

While supervised learning methods have been developed for relation extraction tasks, they typically 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 beneficial to look for active learning methods that exploit a few informative annotated data and achieve reasonable 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) with a 2D scatter plot 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 scatter plot interface.

We are able to achieve following contributions through our studies:

- We present a 2D scatter plot visual interface for human annotations in relation extraction, which, despite lower accuracy, increases the number of annotations that can be made with the same human effort, hence improves 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 scatter plot interface.

2 Related Work

Distant supervision has become a popular branch of approaches in relation extraction. It automatically generates labeled data by looking for the argument pair in the relation tables in a knowledge base (like Wikidata). Surdeanu et al. (2012) presents a multi-instance multi-label (MIML) model for distant supervision in relation extraction, addressing two major challenges in distant supervised models, namely incorrect labeling and multiple possible relations for one pair of entities. Still another problem exists where the knowledge base is often incomplete. Treating the missing knowledge as negative data will result in large

amount of false negative labels, which will hinder the model’s performance. [Min et al. \(2013\)](#) addresses this problem by only generating positive and unlabeled data.

Active learning methods have also been developed during the years. [Sun and Grishman \(2012\)](#) presents a *co-testing* active learning model with local and global views. While local view is based on the context of a pair of entity, the global view classifies new examples based on the distributional similarity of the relation phrases. [Fu and Grishman \(2013\)](#) implements several improvements to the *co-testing* model including better initial setting and balancing imbalanced classifiers. Both of these works simulates user annotated data using fully-labeled data, thus neglecting the human interactions in the classification process. Also, these papers only test their models on a single dataset, providing limited experimental results.

[Angeli et al. \(2014\)](#) combines the MIML model presented by [Surdeanu et al. \(2012\)](#) with active learning. The paper present two criteria for selecting examples to annotate based on disagreement provided by QBC. Their experiments showed that combining the MIML model with annotated data yields improved results.

All of the previous active learning methods incorporate a list-based view presented to the human annotators. [Bernard et al. \(2018\)](#) conducts a thorough set of experiments comparing the effectiveness of traditional active learning strategies with visual-interface labeling methods, showing that visual-interface labeling can compete with traditional active learning methods in various tasks. Though in the paper uniformed sampling of example data is used as baseline while other sampling method may have better performance. [Berger et al. \(2018\)](#) presents a 2D-scatter-plot visual interface for human annotations of name-entity classification and conducts user-studies showing their visual interface out-performs traditional list-based active learning methods despite of noisy data. However, the paper doesn’t compare different sampling methods and their visual interface approach can be extended to other information extraction tasks.

3 Baseline

We compare our model with the MIML_RE model presented by [Surdeanu et al. \(2012\)](#). The model is trained and tested on KBP 2010 and 2011 data ([Ji](#)

Table 1: Experimental results.

Model	Precision	Recall	F1
MIML_RE	19.5	30.7	23.8

[et al., 2010, 2011](#)), aligning the relations with both the knowledge based provided by the shared tasks and a snapshot of the English Wikipedia from June 2010. They use Stanford’s CoreNLP package to identify entity mentions in text and they only consider entity mention candidates occurred in the same sentence. Table 3.1 shows the results of the MIML_RE model.

The data and source code can be found at <https://github.com/Inlinebool/CSC585-Project>.

3.1 Error Analysis

As mentioned before, one major problem for distant-supervised models such as the MIML_RE model is that the knowledge base is often incomplete, resulting in a large number of false-negative examples. The model subsamples the negative examples at a rate of 5%, but the result can yet be improved by providing better negative examples, using human annotations. Another cause of error is that some relation pairs do not exist in the knowledge base, making it impossible for the model to learn those relations. This can also be mitigated by providing human annotations for the unknown relation pairs.

Other errors include system error inherited from CoreNLP and imperfect initialization.

4 Relation Labeling Tool

To achieve efficient and accurate labeling, we provide a Relation Labeling Tool (Figure 1), allowing the human annotators to select and label instances of their choice, either one at a time or multiple instances at a time.

The Relation Labeling Tool and human labeling are incorporated into the MIML_RE relation extraction model in the following steps:

- For each EM iteration, after the E-step the model will select certain number of candidates for user to label based on some active learning criterion.
- The Relation Labeling Tool will then process the candidates and project them into 2D. The candidates will be plotted onto the interface

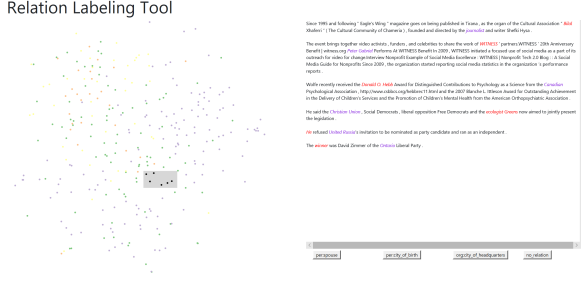


Figure 1: Overview of the Relation Labeling Tool. The left view shows the 2D projection of the sentences. The right view lists the sentences selected and allows the user to label them.

using a pre-defined color map that will help the user to select instances. i.e The 2D projection and the color map will ideally help the user identify classes and patterns.

- The user can click on data points to select a single instance or use a selection tool (rectangle or lasso) to select multiple instances. The original sentences of selected instances will be shown in the right view, with the underlying entity pair highlighted, and the user can select a proper relation label or “no relation” for the instances.
- The results of user selection will be sent back to the model and be viewed as hard labeled data that cannot be changed by later E-steps.
- The process repeats for a few times until the model is trained.

Several active learning strategies for relation extraction are considered in [Angeli et al. \(2014\)](#). However, the results of their experiment may not hold in our case because the number of candidates are reduced and the candidates may not have a good 2D projection, i.e. the classes might not be sufficiently distinguishable in 2D projections. In their experiments their active learning models selected over 30,000 instances for crowdsourced annotation, while in our study the number of instances that the user can annotate is limited due to time and resources available.

A key component of the Relation Labeling Tool is projecting the data instances to 2D. This consists of two steps. The first is to construct a vector representation for each sentence i.e. sentence embedding. Then we can apply a dimensionality reduction method such as PCA, MDS, or T-SNE

to project the instances. Selecting the right dimensionality reduction method is crucial to our tool because class separability of the plot greatly affects the users’ selection ([Bernard et al., 2018](#)).

Unlike the MNIST dataset tested in ([Bernard et al., 2018](#)), our data is textual and much more complicated. It will take longer for the user to label data and the accuracy of the annotations will be lower.

5 Implementation Details and Experimental Results

Due to time limitations, only a primitive prototype of the Relation Labeling Tool is implemented. The tool is not yet connected and tested with the MIML-RE model. Further experiments and user studies need to be done and improvements of the sentence embedding and data projection algorithms can be applied.

5.1 Development Data

Ideally the input data for Relation Labeling Tool should be the output of the active learning strategy applied to the E-step of the training model. However, the MIML-RE model in [Surdeanu et al. \(2012\)](#) used sophisticated features extracted from a custom corpora, and the original sentences are not provided with their training data. Therefore, for development purpose, we used labeled sentences from [Angeli et al. \(2014\)](#). There are in total 33,814 labeled sentences and 42 relation categories including “no relation”. To simplify the problem, we manually select 3 categories “per:spouse”, “per:city_of_birth”, “org:city_of_headquarters” plus “no relation” and randomly sample 300 instances from the labeled sentences as candidate instances. The number of “no relation” instances sampled is proportional to the ratio of negative samples in the whole data.

We acknowledge that this development dataset could be problematic, in that the real candidates from the active learning strategy could be vastly different from our random samples. The ratio of negative samples could also be different. Moreover, We didn’t consider all the categories due to complexity, but the usefulness of the Relation Labeling Tool designed for small number of categories could decrease dramatically when the data is more complicated.

5.2 Sentence Embedding and 2D Projection

To present the candidate instances as 2D plots we need to encode them into vectors in order to apply a dimensionality reduction algorithm. As discussed in the previous section, this is the key part of Relation Labeling Tool.

There are multiple ways to construct sentence embedding. A good starting point is the weighted average bag-of-words algorithm from [Arora et al. \(2016\)](#), although a supervised method might have better results. It is also possible to construct sentence embedding using the features used by [Surdeanu et al. \(2012\)](#) which has rich information on the syntactic structures of the sentences. In our experiments, we implemented the bag-of-words algorithm from [Arora et al. \(2016\)](#), and we estimated word frequencies using the development data.

For dimensionality reduction, T-SNE generally yields better results in terms of class separability. It's the default method implemented in our tool.

Figure 2 shows the results of the 2D projection for 4 different samples. The data points are colored by their true labels. From the plots we can see that although data instances from the same categories tend to be grouped together, the classes are hardly separated, and instances with “no relation” tend to scatter all over the space. In reality, we can only color the data points by their projected label from the model, and it could make the colormap even worse.

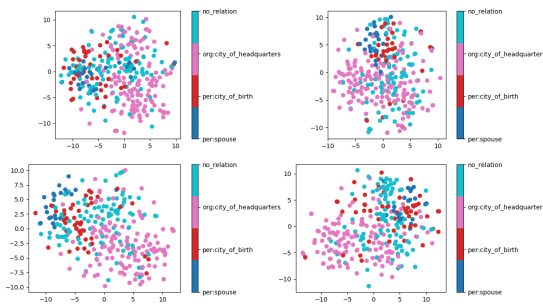


Figure 2: Results of 2D projection for 4 different samples, colormapped by the true label of the instances.

5.3 Visual Labeling Interface

Finally the projected candidate instances can be viewed in the Relation Labeling Tool interface showed in Figure 1 and the user can select single or multiple instances using a rectangle tool.

The original sentence of the selected instances are shown with the entity pair highlighted.

The Relation Labeling Tool is a web-based interface developed using `D3.js`. Due to the complexity of the textual data, the interface combines the scatter-plot view and the list view for easier annotation. The list view from [Angeli et al. \(2014\)](#) has the most probable relation labels predicted by the model as choices for the user, which is not implemented in our tool. This can be implemented as future work. However, for multi-instance labeling, this could cause the interface to be crowded.

In our experiment, the data instances are colored by their true label, which is not possible in real training scenarios where the true labels are unknown. We could color the data points by the predicted label from the model, or we could use a 2D positional colormap independent of the model. User study is needed to verify which colormap scheme is the best for the task.

6 Conclusion

This project aims to design and implement a visual interactive labeling tool for relation extraction, hoping to improve the efficiency and accuracy of human annotation for relation extraction. As seen in the previous sections, many future work is needed for this purpose. However, a basic framework for Relation Labeling Tool is implemented and some clear improvement paths are discovered from the experiments.

Firstly, an active learning strategy can be combined with the visual labeling tool to better select candidate instances. Which active learning strategy is best suited with the tool remains a future study project. Secondly, vector representation of the sentences needs to be carefully constructed so that the 2D projection can be more distinctive. More sophisticated features and sentence embedding models could be used instead of simple weighted bag-of-words. Different dimensionality reduction methods may also have an impact on the quality of 2D projection. Moreover, improvements can be done to the labeling interface regarding the colormap and sentence list view. Lastly, the experiments are only done on a small subsample of the dataset, with only 3 categories. The interface may need some major over-haul when more categories are considered.

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.
- Sanjeev Arora, Yingyu Liang, and Tengyu Ma. 2016. A simple but tough-to-beat baseline for sentence embeddings. In *ICLR 2016*.
- Matt Berger, Ajay Nagesh, Joshua Levine, Mihai Surdeanu, and Hao Helen Zhang. 2018. Visual supervision in bootstrapped information extraction. In *To Be Published*.
- Jürgen Bernard, Marco Hutter, Matthias Zeppelzauer, Dieter Fellner, and Michael Sedlmair. 2018. Comparing visual-interactive labeling with active learning: An experimental study. *IEEE transactions on visualization and computer graphics*, 24(1):298–308.
- 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.
- Heng Ji, Ralph Grishman, and Hoa Trang Dang. 2010. Overview of the tac 2010 knowledge base population track. In *Proceedings of the Text Analytics Conference*.
- Heng Ji, Ralph Grishman, and Hoa Trang Dang. 2011. Overview of the tac 2011 knowledge base population track. In *Proceedings of the Text Analytics Conference*.
- Bonan Min, Ralph Grishman, Li Wan, Chang Wang, and David Gondek. 2013. Distant supervision for relation extraction with an incomplete knowledge base. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 777–782.
- 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.