Relation Extraction Using Deep Learning

Ankur Bapna
Stanford University
ankurbpn@stanford.edu

Melvin Johnson Premkumar
Stanford University
melvinj@stanford.edu

Mikhail Sushkov
Stanford University
msushkov@stanford.edu

Abstract

Relation extraction is an important problem in natural language processing (NLP) and is directly applicable in downstream tasks such as knowledge base population (KBP) and question answering. The current state of the art in relation extraction involves distant supervision, which produces large amounts of noisy training data. Recently, deep learning methods have achieved competitive results in many NLP tasks. These approaches require large quantities of training data, which makes them well-suited to be combined with current distant supervision approaches. In this work we explore applying deep learning methods to relation extraction on a large dataset. We aim to investigate the performance of a simple baseline that involves averaging word vectors, as well as more sophisticated methods based on recursive neural networks (RNN).

1 Introduction

Modern technological advances have made it feasible to develop intelligent systems that can analyze vast quantities of gene-disease relationships, understand scenes in images, and answer free-form questions. Despite their diverse applications, these systems all have a common goal: structure the world‘s information in order to make predictions and infer facts. We generally think of this information as relationships between entities, which often come from knowledge bases such as Wikipedia. The process of finding these relationships in text is referred to as relation extraction. For example, given the sentence ”Barack Obama was a student at Columbia University,” relation extraction would find the relation *per:schools_attended* between the entities *BarackObama* and *ColumbiaUniversity*.

Traditional relation extraction methods have involved manually defining lexical rules to identify certain patterns in text [2], or learning these rules from labeled data in a supervised [15] manner. These approaches require extensive feature engineering and, while impressive results have been demonstrated in domain-specific applications [14], currently no high-quality general-purpose relation extraction system exists.

With the recent explosion of neural network-based methods known as deep learning, state-of-the-art results have been achieved on benchmark tasks in computer vision [5], speech recognition [3], and natural language processing [11]. While deep learning generally requires larger amounts of training data than traditional feature-based approaches, it carries the appealing promise of little to no feature engineering. Given the recent success of these methods for modeling the semantics of individual words [8, 4] as well as longer pieces of text [6], these methods lend themselves nicely for application to linguistic tasks such as relation extraction.

The current state of the art of relation extraction employs distant supervision to generate vast quantities of noisy labeled training data from an unlabeled text corpus [9]. With this technique it is possible to generate millions of labeled examples, which makes it possible to train expressive deep neural network-based models for relation extraction.

2 Related Work

Existing work on applying deep learning to relation extraction was pioneered by Socher et. al to achieve semantic relationship labeling [13]. The goal of the task is to find the semantic relationship between a pair of nominals. For instance, in the sentence

> My [apartment]e1 has a pretty large [kitchen]e2,

we want to predict that *kitchen* and *apartment* are in a component-whole relationship. The authors show that building a single compositional semantics for the minimal constituent including both terms achieves very competitive performance. The model is a Matrix-Vector Recursive Neural Network (MV-RNN) which can do compositionality over variable-length sequences. The authors in [1] extend this work further by using the shortest path between two entities in the sentence dependency graph. Converting this path into a chain and using an RNN to perform relation classification is shown to beat the previous approach using MV-RNNs.

The authors in [12] use neural tensor networks to perform knowledge base completion - extraction of new *(subject, predicate, object)* triples from unlabeled text. This approach achieves high accuracy on extracting new instances of known relationships from text by modeling the text within each entity mention as the average of their word vectors. We aim to extend this work by considering not only the words inside of the entity mentions, but also the words between them in the sentence.

3 Objective and Data

We use the dataset and evaluations from the slot filling track of the TAC-KBP annual competition held by NIST. The competition defines a set of 42 relations and provides a publicly-released distantly-supervised training corpus of 6 million documents (approx. 200 million sentences) obtained from NewsWire articles. For a given sentence, the corpus gives the position of both entity mentions (subject and object), as well as one or several correct relations involving them. An example sentence that contains the relations *per:employee_of* and *per:schools_attended* is:

> NASA announced the imminent descent on June 27, and principal investigator Steven Squyres of Cornell University said the wind and dust began blowing soon after.

The Stanford NLP group has preprocessed the data using the Stanford NLP toolkit [7]: for each sentence, named-entity (NER) tags, part-of-speech (POS) tags, and dependency parse trees are stored.

4 Approach

Our initial approach involves a simple baseline model. Given the words in the sentence and the two entity mentions, obtain a sequence of words between the mentions, including the words that are part of the mentions themselves. For each word, take its n-dimensional vector representation from a pretrained GloVe model [4], and compute the vector representation of the sequence as the simple average of the word vectors in the sequence. This produces an n-dimensional representation for the sequence of words between the mentions. This n-dimensional vector serves as the input layer to a neural network with 1 hidden layer of 500 units, which produces a probability distribution over the 42 relations using a softmax after the hidden layer.

5 Evaluation

Since the KBP competition provides an evaluation set each year, we will ultimately use the 2010 evaluation set as development data and use the 2013 evaluation set as test data to evaluate our models. Additionally, the Stanford NLP group has developed a pipeline to test the performance of relation extraction systems that are part of the KBP framework. We will use this pipeline for evaluating our models and comparing their performance against existing results.

6 Results

For our baseline method we train on a subset of 1 million examples and evaluate on a development (dev) set of 1557 examples. The current state of the art system, MIML-RE, achieves an accuracy of 27.9% on this particular dev set. A random prediction will achieve an accuracy of 1/42, or 2.4%, and simply predicting the most frequently-occurring relation in the dev set (*org:city_of_headquarters*) achieves an accuracy of 8.7%. After tuning the learning rate and regularization, our baseline model achieves an accuracy of 12.9% on the dev set.

7 Future Work

We will extend this baseline by applying more complex models. The first of these is a basic recursive neural network [10]. Given the dependency tree of a sentence and two entity mentions that are nodes in that tree, it is possible to train a recursive neural network on the dependency tree of the sentence. The model will start with a vector representation for the individual words that are the leaves of the dependency tree, and obtain a combined representation for these word vectors, as specified by the dependency tree.

References

[1] Javid Ebrahimi and Dejing Dou. Chain based rnn for relation classification. 2015.

[2] Oren Etzioni, Michael Cafarella, Doug Downey, Stanley Kok, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Web-scale information extraction in knowitall: (preliminary results). In *Proceedings of the 13th International Conference on World Wide Web*, WWW '04, pages 100–110, New York, NY, USA, 2004. ACM.

[3] Awni Y. Hannun, Carl Case, Jared Casper, Bryan C. Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, and Andrew Y. Ng. Deep speech: Scaling up end-to-end speech recognition. *CoRR*, abs/1412.5567, 2014.

[4] Christopher D. Manning Jeffrey Pennington, Richard Socher. Glove: Global vectors for word representation. 2014.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. pages 1097–1105, 2012.

[6] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. *CoRR*, abs/1405.4053, 2014.

[7] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, 2014.

[8] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.

[9] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2 - Volume 2*, ACL '09, pages 1003–1011, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics.

[10] Andrew Y. Ng Richard Socher, Christopher D. Manning. Learning continuous phrase representations and syntactic parsing with recursive neural networks. 2011.

[11] Jean Y. Wu Jason Chuang Christopher D. Manning Andrew Y. Ng Christopher Potts Richard Socher, Alex Perelygin. Recursive deep models for semantic compositionality over a sentiment treebank. 2013.

[12] Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. Reasoning With Neural Tensor Networks For Knowledge Base Completion. In *Advances in Neural Information Processing Systems 26*. 2013.

[13] Richard Socher, Brody Huval, Christopher D. Manning, and Andrew Y. Ng. Semantic Compositionality Through Recursive Matrix-Vector Spaces. In *Proceedings of the 2012 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2012.

[14] Ce Zhang, Vidhya Govindaraju, Jackson Borchardt, Tim Foltz, Christopher Ré, and Shanan Peters. Geodeepdive: Statistical inference using familiar data-processing languages. In *Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data*, SIGMOD '13, pages 993–996, New York, NY, USA, 2013. ACM.

[15] Ce Zhang, Christopher Ré, Amir Abbas Sadeghian, Zifei Shan, Jaeho Shin, Feiran Wang, and Sen Wu. Feature engineering for knowledge base construction. *CoRR*, abs/1407.6439, 2014.