Yes/No Question Answering

Ajinkya Indulkar
CICS
University of Massachusetts Amherst
aindulkar@cs.umass.edu

Divyendra Mikkilineni
CICS
University of Massachusetts Amherst
dmikkilineni@cs.umass.edu

Abstract

One of the hardest tasks in NLP is question answering. In this project we investigate the problem of answering questions using deep neural networks with memory. We also use a basic classifier that acts as a baseline for our final performance. We mainly focus our attentions on the Facebook bAbI QA tasks dataset that has recently gained a lot of traction. It is a well-structured simulated Question Answer dataset. We primarily focus on Yes/No type questions i.e. questions which can be answered by either "yes" or "no".

1 Introduction

Addressing the QA challenge subsumes addressing many fundamental tasks like understanding the context, entity recognition, relationship between entities, word sense disambiguation, fact chaining etc. which are all crucial to the whole field of Natural language understanding.

(Jurafsky, 2000) describes two approaches traditionally used for question answering: Information Retrieval approach and semantic parsing using a structured Knowledge base approach. Both of the approaches involve long processing pipelines. The hybrid approach pipeline used in IBM Watson corroborates this fact(Ferrucci, 2012). There has been a lot of work recently, using deep learning to address QA.(Kumar et al., 2016; Sukhbaatar et al., 2015) Although these systems generally involve a smaller learning pipeline, they require a significant amount of training. On the brighter side, they have been said to shown huge accuracy improvements. An array of techniques in Neural networks, for example *Attention mechanisms* which help focus on the most relevant facts, further the state of the art performance in QA.

In this project, we implement a traditional Machine Learning model for QA and compare it with a deep neural network model using some of the recent techniques involving attention mechanisms.

2 Related Work

(Jurafsky, 2000) has discussed QA in his book. He goes over the two traditional QA approaches: Information Retrieval approach and Knowledge based approach. (Das et al., 2017) tried to combine the KB approach with unstructured text. They used a Universal Schema to align the KB with unstructured text in a common space.

Recently, people have demonstrated the effectiveness of Neural networks for Language Modeling.(Bengio et al., 2003; Sundermeyer et al., 2012) In the past few years there has been an increase in research on Question Answering techniques. A lot of attention is being paid to Visual Question Answering and a huge amount of work is being done in that area.(Xiong et al., 2016; Santoro et al., 2017)

There have been several advancements in textual question answering recently in the form of Dynamic Memory Networks(Kumar et al., 2016) and weakly supervised end-to-end memory networks(Sukhbaatar et al., 2015).

We aim to investigate some of these techniques and compare their performance with some of the classical models used in NLP and Machine Learning.

3 Datasets

Our literature review showed that they are many specific branches under QA: Factoid QA, Reasoning QA, Science and Math QA, Visual QA, each having their own class of datasets and demanding their own specific technical approaches to confront with. To stay within the scope of an academic project, we look towards targeting Factoid QA. Factoid QA generally involves identifying some information related to a person, place or thing, given some description of the entity.

We could find another classification of QA datasets based on their scope. The datasets are either:

- Open-domain (TREC) in nature, where for a given question the answer depends on any general world domain. The QuizBowl dataset is one such dataset.
- Focused-domain in nature, where the knowledge for answering the given question is provided. We target the focused-domain datasets to bypass implementing the IR setup needed to tackle open-domain datasets.

Facebook AI Research has released a simulated dataset for Question Answering called bAbI (Weston et al., 2015). They have provided datasets for various QA tasks such as Yes/No questions, single supporting fact questions, two supporting fact questions etc. We have chosen to work on the Yes/No Questions task. All the data is generated by simulations and the data is human-readable. We have chosen this dataset because the questions have been clearly divided into tasks. Furthermore, it uses limited vocabulary and therefore it is well-suited for academic projects.

Another dataset we came across was The Stanford Question Answering Dataset (SQuAD) (Rajpurkar et al., 2016) which is a reading comprehension dataset that was manually created by crowdworkers. It consists of passages with questions posed at the end.

There is another popular QA dataset called MCTest dataset (Richardson et al., 2013) which consists of a set of stories and questions associated with these stories.

4 Approach

We plan to start constructing a standard NLP classification system with hand-built feature engineering like semantic role labelling and POS tagging. Then, we will go forward implementing deep models, known for their feature selection, starting from simple Neural network models to complex ones.

We will convert the words in the support sentences and questions into vector embeddings and input them as features to a fully connected Neural network and use a classification layer as an output layer to get a probability distribution over the answer class.

We will then try RNN models, specifically inclined to implement GRU units first followed by Deep Averaging Networks(Iyyer et al., 2015) and Dependency Tree RNNs(Iyyer et al., 2014). While Dynamic Memory Networks have been said to give state of the art results beating even hand-engineered traditional models(Weston et al., 2015), they require significant amount of time which we might not have during the course of the term.

The bAbI dataset for the task 'Yes/No questions' has 1000 training examples and 1000 test entries. Since the training set size is not huge, we are planning to adopt a 5-fold cross validation approach where we will split the training set into 5 equal folds, four for training and one for performance validation. We would then iteratively use different folds for validation and select the parameters with the best average performance across cross-validation.

At this time, our scope is limited to Yes/No type questions, but the actual dataset contains 20 different kinds of questions. The aforementioned techniques could very well be generalized to other question types.

4.1 Preliminary Experiment

We plan on using an N-gram model with a standard NLP classifier such as SVM to predict the answer as either yes or no. One way of doing this is creating a bag of N-grams that contain at least one word from the question and using that to classify the answer. We can also use a structured SVM classifier for the final classification. Another baseline model we are considering is a simple fully-connected neural network with GloVe[1] word vectors as input.

5 Softwares

We plan on using NLTK for basic NLP tasks such as tokenization, lemmatization etc. For our preliminary experiments we plan to use the Stanford GloVe library for word vectorization. We will implement our neural nets in Tensorflow or PyTorch.

[1] https://nlp.stanford.edu/projects/glove/

References

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. 2003. A neural probabilistic language model. *Journal of machine learning research* 3(Feb):1137–1155.

Rajarshi Das, Manzil Zaheer, Siva Reddy, and Andrew McCallum. 2017. Question answering on knowledge bases and text using universal schema and memory networks. *arXiv preprint arXiv:1704.08384* .

David A Ferrucci. 2012. Introduction to this is watson. *IBM Journal of Research and Development* 56(3.4):1–1.

Mohit Iyyer, Jordan L Boyd-Graber, Leonardo Max Batista Claudino, Richard Socher, and Hal Daumé III. 2014. A neural network for factoid question answering over paragraphs. In *EMNLP*. pages 633–644.

Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daumé III. 2015. Deep unordered composition rivals syntactic methods for text classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. volume 1, pages 1681–1691.

Dan Jurafsky. 2000. *Speech & language processing*. Pearson Education India.

Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. 2016. Ask me anything: Dynamic memory networks for natural language processing. In *International Conference on Machine Learning*. pages 1378–1387.

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250* .

Matthew Richardson, Christopher JC Burges, and Erin Renshaw. 2013. Mctest: A challenge dataset for the open-domain machine comprehension of text. In *EMNLP*. volume 3, page 4.

Adam Santoro, David Raposo, David GT Barrett, Mateusz Malinowski, Razvan Pascanu, Peter Battaglia, and Timothy Lillicrap. 2017. A simple neural network module for relational reasoning. *arXiv preprint arXiv:1706.01427* .

Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. 2015. End-to-end memory networks. In *Advances in neural information processing systems*. pages 2440–2448.

Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. 2012. Lstm neural networks for language modeling. In *Thirteenth Annual Conference of the International Speech Communication Association*.

TREC. ???? TREC qa description. trec.nist.gov/data/qa.html.

Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. 2015. Towards ai-complete question answering: A set of prerequisite toy tasks. *arXiv preprint arXiv:1502.05698* .

Caiming Xiong, Stephen Merity, and Richard Socher. 2016. Dynamic memory networks for visual and textual question answering. In *International Conference on Machine Learning*. pages 2397–2406.