CSCI 5622
Project Definition Document (PDD)

RNN for Factoid Question Answering

Akshay Singh, Siddharth Singh, Scott Sheahan, Camden Elliott-Williams

November 11, 2016

1   Team Members

Name	Email Address
Akshay Singh	Akshay.Singh@colorado.edu
Siddarth Singh	Siddharth.Singh@colorado.edu
Scott Sheahan	scott.sheahan@colorado.edu
Camden Elliott-Williams	Camden.ElliottWilliams@colorado.edu

Table 1: Team Member Contact Information

2   Project Basics

2.1   Task

We will be porting and trying to improve a model that answers factoid questions i.e. given a description of an entity, identify the per- son, place, or thing discussed.

2.2   Data

We will be working on the dataset used in the paper [1]. The dataset has 20,3407 questions and about 2,347 answers. The data is in the following information:

1. Question ID

2. Fold - The fold defines whether the data is used in the training set or the test set.

3. Answer - The answer can either be a single word answer or a multiple word answer

4. Category - Literature, Science, etc.

5. Text - A series of sentences which provide the context for answering the question.

   A sample question-answer set from the dataset is as follows:

   3009,train,Henry IV of France,History,"With the assistence of his chief minister, the Duc de Sully, he lowered taxes on peasantry, promoted economic recovery, and instituted a tax on the Paulette. ——— Victor at Ivry and Arquet, he was excluded from succession by the Treaty of Nemours, but won a great victory at Coutras. ——— His excommunication was lifted by Clement VIII, but that

Model	History			Literature		
	Pos 1	Pos 2	Full	Pos 1	Pos 2	Full
BOW	27.5	51.3	53.1	19.3	43.4	46.7
BOW-DT	35.4	57.7	60.2	24.4	51.8	55.7
IR-QB	37.5	65.9	71.4	27.4	54.0	61.9
FIXED-QANTA	38.3	64.4	66.2	28.9	57.7	62.3
QANTA	**47.1**	**72.1**	**73.7**	**36.4**	**68.2**	**69.1**
IR-WIKI	53.7	76.6	77.5	41.8	74.0	73.3
QANTA+IR-WIKI	**59.8**	**81.8**	**82.3**	**44.7**	**78.7**	**76.6**

Figure 1: Baseline results with RNN (Percent Correct)

pope later claimed to be crucified when this monarch promulgated the Edict of Nantes. ——— For 10 points, name this French king, the first Bourbon who admitted that ""Paris is worth a mass"" when he converted following the War of the Three Henrys."

2.3 Baseline

As a team we will implement the RNN algorithm for factoid question answering that is presented in [1]. We will develop this on Tensor Flow. Our goal will be to achieve results similar to those in the paper (see Figure 1).

3 Techniques to be explored

1. Feature engineering:

   (a) LSA: Latent Semantic Analysis
   Word and passage meaning representations derived by LSA have been found capable of simulating a variety of human cognitive phenomena, ranging from developmental acquisition of recognition vocabulary to word-categorization, sentence-word semantic priming, discourse comprehension, and judgments of essay quality. LSA can be used in our model for generating new features for our model.

   (b) The paper[1] uses a weight matrix corresponding to the depdendency relations in the sentence. We were thinking of exploring temporal relations and Propbank labels as additional features.

2. Model Adjustments:

   (a) Alterations in neuron components.
   (b) Freeze word vectors from backpropagation as in [2].

4 Timeline

November 18: Learn Tensor Flow and complete baseline

December 2: Finish exploration

December 9: Project Write-Up and Video Presentation

References

[1] M. Iyyer, J. Boyd-Graber, L. Claudino, R. Socher, H. Daumé III, H. Daum, and H. III, "A Neural Network for Factoid Question Answering over Paragraphs," *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 633–644, 2014. [Online]. Available: http://www.cs.colorado.edu/{~}jbg/docs/2014{_}emnlp{_}qb{_}rnn.pdf$\backslash$nhttp://www.aclweb.org/anthology/D14-1070

[2] B. Wang and M. Liu, "Deep Learning for Aspect-Based Sentiment Analysis."