Proposal 2 : Question Answering From Semi-structured Tables

Divya Sree Korimi
Arizona State University
Tempe, AZ, USA
dkorimi@asu.edu

Adam Chmurzynski
Arizona State University
Tempe, AZ, USA
achmurzy@asu.edu

Nitesh Thali
Arizona State University
Tempe, AZ, USA
nthali@asu.edu

Problem Definition

In the field of natural language processing several techniques have emerged for processing data by learning through supervised and un-supervised learning in order to develop more effective and communicative applications. When we say data it is either available as structured data in the form of tables, ontologies, rules or unstructured data as natural language text. In this project we are proposing to develop a Question Answering (QA) model from knowledge available in the form of semi-structured tables. In Our approach we aim at developing a feature-driven model that uses human annotated MCQs to perform QA and reasoning over tables.

1 Motivation

Knowledge in the form of natural language text is easy to acquire, but difficult for automated reasoning. Highly-structured knowledge bases can facilitate reasoning, but are difficult to acquire. In this project we are using the tables as the semi-structured knowledge representation of data. The table store consists of AI2s tables [Jauhar et.al.] which are organized based on topic pertaining to a limited number of topics provided. The tables format is semi-structured where the rows of the table are a list of sentences, but with well-defined recurring filler patterns. Joining them along with the header these patterns divide the rows into meaningful columns. Now here we implement the table search algorithm by defining the set of questions Q = $\{q1, q2, q3..qn\}$ denote a set of MCQs, and An = $\left\{a_n^1, ..., a_n^k\right\}$ be the set of candidate answer choices for a given question qn. Let the set of tables be defined as T = $\{T_1, ..., T_M\}$.With these as initialization of the data sets, training using [Jauhar-et.al] FRETS log -linear model on the tables is performed.

2 Related Work

The semi-structured knowledge work with tables relates to several other research work on the question answering pattern. In the data set of crowd-sourcing, Aydin et al. (2014) harvest MCQs via a gamified app, although their work is not involved with tables. Other research work done by Pasupat and Liang (2015) have used tables from Wikipedia to construct the Question Answering pairs, but their setup does not involve tables with structural constraints. Sun et al. (2016) perform cell search over web tables via relational chains, but are more generally interested in web queries. We evaluate our model on MCQ answering for three benchmark datasets which have shown the benefits of semi-structured data and models over unstructured or highly-structured counterparts.

3 Examples

Following table shows the example of a small portion of the knowledge base. Given a question as shown below we will be choosing the answer as the one which gets the maximum score.

Earth science term	Subtype of term	Condition for event	Example event
runoff	stormwater	rain falls during a rainstorm	it can flow over the earth's surface
precipitation	freezing rain	the temperature is below freezing	rain freezes on contact with cold surfaces
precipitation	sleet	in the winter	a mix of rain and snow falls

Q: What is it called when rain freezes on the objects it falls upon?
A) Stormwater B) Freezing rain C)Rain D)Cirrus

4 Knowledge/Data resources

- Table Source consists of 65 handcrafted tables. (http://ai2-website.s3.amazonaws.com/data/TabMCQ_v_1.0.zip)
- Elementary School Science Questions Dataset. (http://aristo-public-

data.s3.amazonaws.com/AI2-Elementary-NDMC-Feb2016.zip)

5 Software Resources/Tools

- Log Linear Model
- Adaptive gradient descent with an L2 penalty
- Amazon's Mechanical Turk (Mturk)
- Python Scikit-learn

6 Evaluation Parameters

We will train our model with 4th grade science exam MCQ datasets: using Regents dataset and dataset of Elementary School Science Questions (ESSQ). We are planning to split these data sets into standard 70:30 training and testing data sets. Finally we will evaluate our Model based on testing dataset and a single value evaluation using accuracy. Accuracy will represent the percent of questions correctly answered by our model.

References

SK Jauhar, PD Turney, E Hovy 2016. *Tables as Semi-structured Knowledge for Question Answering.* FRETS-Association for Computational Linguistics (ACL) -2016

Peter Clark, Oren Etzioni, Tushar Khot, Ashish Sabharwal, Oyvind Tafjord, Peter Turney, and Daniel Khashabi 2016. *Combining retrieval, statistics, and inference to answer elementary science questions.*, Proceedings of the 30th AAAI Conference on Artificial Intelligence, AAAI-2016.

Anthony Fader, Luke Zettlemoyer, and Oren Etzioni. 2016. *Open question answering over curated and extracted knowledge bases..* In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 11561165. ACM.