

Relation Extraction for Matrices(type) entities in Introductory programing problems

Himanshu Shukla(13309)¹
Kumar Gaurav(12368)¹

¹Computer Science and Engineering Department
IIT Kanpur

31 March, 2015

Summary

- 1 Introduction
- 2 Related Work
- 3 Methodology
- 4 References

What we want ?

- Extract relations among matrix related entities

Example

Matrix

- Attributes : size, symmetric, complex entries, binary matrix, row matrix, column matrix
- Operations : Add, Subtract, Multiply, trace, determinant

What we want ?

In English

Write a program to input two matrices A and B of size $M \times N$ each , add them and find the trace of resultant matrix.

In Metalanguage

two matrices A size $M \times N$ B size $M \times N$. add them. find trace resultant matrix.

Related Work

- In the field of relation extraction in daily English and specifically in Biological Field.
- Semantic mapping between Natural Language Questions and SQL queries via Syntactic Pairing

This can help a lot in improving the IDE's available for elementary programming courses like SigPact.

Metalanguage Mapping

- Mapping using Tree Kernels and Functions.
- Feature space is defined using Joint space $STK+STK$ or Cartesian Product $STK \times STK$

Metalanguage Mapping

$$TK(T_1, T_2) = \sum_{n_1 \in N_{T_1}} \sum_{n_2 \in N_{T_2}} \Delta(n_1, n_2)$$

FIGURE: Tree Kernel function over T_1 and T_2

$$\Delta(n_1, n_2) = \lambda \prod_{j=1}^{l(n_1)} (1 + \Delta(c_{n_1}(j), c_{n_2}(j)))$$

Metalanguage Mapping

Engineer Kernel which generates best results. Eg.

- $STK^2 + POLY^2$
- $(1 + LIN^2)^2$
- $(1 + LIN^2)^2 + STK^2$

Semantic Analyzer

We use the Lex-Yacc to generate 2 sets :

- Diff set
- Context Set

Handling Anaphora

Instances like *its* and *these* are looked up in the *Context* and *Diff* sets to get resolution.

- *The object mentioned earlier will appear in the Context set. Getting the first instances will give the required object.*
- *Objects mention in the same sentences will appear in the Diff set. Getting the newly constructed item will give the result.*

Corpus Collection

- Generalizing question and mapped instances : substitute the involved concepts in questions and their related field values mapped by means of variables (expressing the category of such values).
- Dataset annotation : consider all possible pairs, i.e. cartesian product between all the questions and results of the dataset, and annotate them as negatives if they have not been annotated as positives in the previous step.

References

- 1 Kewalramani, Pankaj Prateek, et al. "Anaphora without syntax-A Multi-lingual Approach for Geometry Constructions."
- 2 Giordani, Alessandra, and Alessandro Moschitti. "Semantic mapping between natural language questions and SQL queries via syntactic pairing." *Natural Language Processing and Information Systems. Springer Berlin Heidelberg, 2010. 207-221.*
- 3 Collins, Michael, and Nigel Duffy. "Convolution kernels for natural language." *Advances in neural information processing systems. 2001.*