

The University of Sheffield
Department of Computer Science

Strucutred prediction for arithmetic math problems



Giorgos Myrianthous

Supervisor: Dr. Andreas Vlachos

This report is submitted in partial fulfilment of the requirement
for the degree of MSc in Data Analytics by Giorgos Myrianthous.

July 18, 2017

Declaration

All sentences or passages quoted in this report from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s). Any illustrations that are not the work of the author of this report have been used with the explicit permission of the originator and are specifically acknowledged. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure in this project and the degree examination as a whole.

Name:

Signature:

Date:

Abstract

Understanding human language is a difficult task, with varied fields of study which aim at explaining and researching the human language faculty. Linguistics, Psychology and Computer Science all use domain specific tools to describe and model language. Natural Language Processing is the field which aims at using computational mechanisms to process naturally occurring human language. Modelling syntax gives language structure, but how do we model meaning? Using general sense classes, or “supersenses” we can potentially enrich texts with semantic information.

Given a simple arithmetic math problem in text form, is it possible to construct the right equation and derive the right answer?

The goal of this project, was to create a system that is able to predict the right equation and answer for a given arithmetic math problem. Given the textual description of the problem and the alignments that are used to locate the extracted quantities, the system is able to compose an equation that will eventually lead to the right solution to the problem.

The goal of this project is to create a Multiword Expression boundary and Supersense labelled sentence by training with Word, Part-Of-Speech, Multiword Expression and Supersense tagged training data. The semantically tagged sentences can be used for many tasks such as Question Answering systems, Information Retrieval, Discourse and Sentiment analysis.

Contents

List of Figures

List of Tables