

## Abstract

We plan to investigate methods to create a system that automatically learns to solve algebra word problems, inspired by the paper from Kushman et al., 2014. Specifically, we plan to focus on solving system of linear equation problems of middle school level difficulties. After successful implementation of such system, we believe we can extend this to solving general mathematical word problems.

We think this project will heavily involve natural language processing and information extraction.

## Why?

We thought it would be interesting to create a system that automatically solves mathematical problems. We also thought this would be a great opportunity to learn areas of natural language processing. Moreover, algebra problem text has more limited style (following some sort of patterns), making it easier to solve than other challenging NLP problems. Finally, the main reference paper has been only recently published (2014), so we believe there are many areas of improvements we can make.

## What is the learning component?

We plan to solve this problem in three steps.

First, we determine the number of variables involved in the word problem. We believe this can be easily determined using basic features such as number of sentences and number of constants that appear in the text. Since most middle school algebra problems involve one or two variables (rarely three), we think this would be pretty simple.

After the number of variables are determined, we need to map the word tokens (nouns and numerical values) in the problem text to the corresponding variables in the template. We first preprocess the text to extract nouns. Then, our learning model calculates the probability of specific matching of each noun to an unknown variable in the template.

After the mapping, what is left is just solving the system of linear equations using Gaussian elimination.

## What we already know

We are continuously collecting data sets from various sources, such as SAT Math and online algebra problems. The paper by Kushman et al used Algebra.com as the main source of data.

We are reading various algebra word problems to capture important features that we would be able to use to create our model.

We have not yet thought of detailed methods to solve this problem. While our approach is based on the reference paper by Kushman et al, we plan to investigate novel methods to solve this problem.