

Maths Word Problem Solver in Hindi

Aditya Holikatti
holikatt@usc.edu
1125-7868-01

Anjan Shah
anjansha@usc.edu
7680-3598-80

Harshita Kasera
hkasera@usc.edu
3790-0173-05

Priyesh Lakar
lakar@usc.edu
2631-1745-58

Introduction:

The ability to solve word problems is an integral part of our mathematical education because real-life problems almost never present themselves to us in mathematical form. The process of solving word problems involves understanding the components of the problem written in natural language in conjunction with mathematical reasoning.

The aim of our project is to construct a very basic Hindi word problem solver which solves elementary school math problems involving only a single mathematical operation.

Solving word problems is an interesting problem because of the inability to formulate a mathematical representation of the text automatically as it involves additional mathematical reasoning apart from comprehending natural language.

All the prior research work in this domain focuses on English language. There is significant work done by Cognitive Computation Group at Illinois and Institute of Information Science at Taiwan which we will use as a reference for our project.

Method:

Materials:

Since there is no readily available corpus of Hindi word problems, we have collected problems from publicly available websites and school textbooks. The word problems that we have included in our corpus are mainly elementary mathematical problems which include general mathematical operations.

We have obtained about 300 examples for now and we will continue to look for more sources. Most of the examples we have only provide us with the solution to the word problem and no other annotations. We will classify and annotate them based on the type of problem.

At the end of the project we aim to open source our corpus for other projects in the future.

Data links:

- NCERT Textbook for Maths in Hindi (Class 1-7)
- <https://onlinetyari.com/hindi/question-bank/>
- <http://www.allexamgurublog.com/2016/01/mathematics-questions-in-hindi.html>
- <https://onlinetyari.com/hindi/tags/algebra-i152.html?page=1>
- http://www.cggkquiz.in/p/blog-page_24.html
- <http://www.jagranjosh.com/>

Procedure:

We are using [1] paper as our reference for implementation. We plan to break down our implementation into following stages -

- **Text Preprocessing** - We will curate stop words in Hindi as general stop words are strongly associated with math word problems. We will use various techniques like tokenization, sentence splitting, POS tagging, lemmatization, named entity recognition for identifying critical components of the problem .
 - Tool - [NLTK](#).
- **Feature Creation** - We will use features based on sentences, verbs and keywords to define our model for solution type. A sentence level feature could be question mark in the sentence which will help identify the entity in the sentence as the unknown variable.
- **Tuple Generation & Inference** - We will generate tuples based on above features to convert the sentences into logic forms. We will create an equation as an expression tree using the tuples generated.
- **Equation Solving** - We will use [SymPy](#) to solve the equations.
- **Explanation Generation** - Based on the feasibility, we will convert the above into explanation using specific templates.

Evaluation:

To evaluate performance, the equations generated by our system are compared to the annotated equation labels for the respective problems. When comparing equations, we avoid spurious differences by canonicalizing the equation system.

Because we do not estimate a more abundant corpus, we plan to use k-fold cross-validation to both train and evaluate our system's performance. We will judge our system based on the metric of equation accuracy, i.e how often our system generates the correct equation. In the absence of a similar system in hindi, we will compare the performance of our system to state-of-the-art English word problem resolvers.

References:

- [1] <https://www.aclweb.org/anthology/N/N16/N16-3014.pdf>
- [2] <https://aclweb.org/anthology/N/N16/N16-3011.pdf>
- [3] https://cogcomp.cs.illinois.edu/page/publication_view/804
- [4] <https://www.wolframalpha.com/>
- [5] <http://www.aaai.org/ocs/index.php/FLAIRS/2009/paper/viewFile/33/239>
- [6] <http://www.aclweb.org/anthology/Y14-1015>

Division of Labor:

Every member will work on all components of the project but we are assigning each component to a specific person who would oversee that component.

Aditya Holikatti: Data collection and annotation

Anjan Shah: Data preprocessing

Harshita Kasera: Logic Implementation

Priyesh Lakar: Evaluation

Word Count of the Project Proposal (excluding links and title): 580 words