HAWP: a Dataset for Hindi Arithmetic Word Problem Solving

Harshita Sharma, Pruthwik Mishra, Dipti Misra Sharma
{harshita.sharma, pruthwik.mishra}@research.iiit.ac.in, dipti@iiit.ac.in

LREC 2022 Marseille

SUMMARY

- Most of the publicly available datasets and work for Word Problem Solving has been carried out for English.
- We have attempted to address this issue for any Indian Language, especially Hindi.
- We have developed HAWP (Hindi Arithmetic Word Problems), a dataset consisting of 2336 arithmetic word problems in Hindi and baseline systems for solving these word problems.

DATASET

- **2336** Word Problems in Hindi
- **Arithmetic Operations:** Addition, subtraction, multiplication and division operations having one unknown.
- **Number of Operations:** One and two operation word problems.
- **Annotation**: Equations, Number of operations and indices of relevant quantities in the word problems (used annotation tool)

```
{
  "pIndex": 48,
  "Problem": "सिद्धांत के पास 86 पीले और 20 हरे कंचे थे। अरजीत ने
सिद्धांत के पीले कंचों में से 25 ले लिए। सिद्धांत के पास अब कितने पीले
कंचे हैं?",
  "Equation": "X = ( 86 – 25 )",
  "Relevant Indices": [
    0,
    2
  ],
  "Number of Operators": 1
},
```

Fig. Snapshot of the dataset

MANUALLY CRAFTED PROBLEMS

- Hindi-medium Math Textbooks
- Hindi-medium Math Teachers

AUGMENTATION USING TRANSLATION

- Translating word problems from benchmark English datasets - AI2 [1], Unbiased [2] and ASDiv [3].
- Issues with Translated Problems:
  - Translation errors
  - Missing one to one mapping
  - Cultural Differences
- Post Edit Guidelines (using post editing tool):
  - Localisation: Units, Currency, Proper Nouns, food items, sport names etc.
  - Borrowing
  - Naturalness
  - Correction of ungrammatical sentences etc.

DATASET EVALUATION

The developed dataset was evaluated on two parameters:
- **Naturalness and Complexity:** We asked Hindi-medium school students to solve some of our translated word problems. Students were able to form correct equations for 90% of the MWPs, out of which 88.33% of them were solved with correct answers.
- **Diversity:** Metrics proposed by different researches:

Metric	MAWPS	ASDiv	HAWP
MAWPS Lexical Diversity*[4]	6.52%	5.84%	5.92%
Corpus Lexicon Diversity**[3]	49%	42%	73%

*lower value indicates more diverse corpus
**higher value indicates more diverse corpus

BASELINE MODEL

- **Embeddings**: We used the publicly available pre-trained subword embedding [5] for encoding our Hindi input data. These embeddings are learnt by training on Hindi Wikipedia data using byte pair encoding. We used the same subword embeddings to encode both the word problems and the target equations.
- **Model:** We used a **2 layer Bi-LSTM encoder decoder network with global attention** for predicting the **prefix equations** given a word problem.

MODEL EVALUATION and RESULTS

- We performed **10 fold cross validation** on the whole dataset. The results are shown below:

Metric	Accuracy	Metric	Accuracy
Full Set	34.82	Full Set w/ No Implicit	39.92
One Operation	40.04	One Op. w/ No Implicit	44.43
Two Operation	17.81	Two Op. w/ No Implicit	19.03

- **Equation Equivalence:** We also observed that equation equivalence improves the performance of the model by 2% on an average.
- Most of the errors is attributed to incorrect operator identification.

  Role of Implicit Quantities:
- The solver also struggles to identify the implicit quantities and could not make correct association with the actual quantity.
- The accuracies of the solver increases by approximately 5% on an average (across the dataset improving both the one operator and two operator equation) after removing problems with implicit quantities.
- This is due to very low frequency of such numbers in the word problems. Only 2% (64 out of 2336) of the word problems contained implicit quantities.

EXAMPLES

Translation Errors:
- crayons and *rulers*: kreyon aur *shaasak*
- chicken *wings*: murge ke *pankh*
- how many shells did he *end up with*?: vah kitne shankhon *ke saath samapt* hua?

Localisation:
- Ounces, pounds: litre, gram
- Sam, Meredith: Ram, Geeta
- Candies, cookies: toffee, biscuit

Naturalness:
- Molly had 14 candles on her birthday cake. She grew older and *got 6 more* on her birthday cake. *How old is Molly* now?: maulee ke barthade kek par 14 momabattiyaan theen. vah badee ho gaee aur usake barthade kek par *use 6 aur mileen*. maulee ab *kitanee badee hai*?

Implicit Quantities:
- p *tikonon* mein kul milaakar kitane kone ho jaenge? [tikonon = 3 corners; triangle]
  - Gloss: How many corners would p *triangles* have?
- ek bageeche mein paudhon kee t *panktiyaan* aur r *kolam* hain. kul kitane paudhe hain? [area = row x col]
  - Gloss: b garden has t *rows* and r *columns* of plants. How many plants are there in total?

REFERENCES

1. Hosseini, M. J., Hajishirzi, H., Etzioni, O., and Kushman, N. (2014). Learning to solve arithmetic word problems with verb categorization.
2. Roy, S. and Roth, D. (2018). Mapping to declarative knowledge for word problem solving.
3. Miao, S.y., Liang, C.C., and Su, K.Y. (2020). A diverse corpus for evaluating and developing English math word problem solvers.
4. Koncel-Kedziorski, R., Roy, S., Amini, A., Kushman, N., and Hajishirzi, H. (2016). Mawps: A math word problem repository.
5. Heinzerling, B. and Strube, M. (2018). BPEmb: Tokenization-free Pre-trained Subword Embeddings in 275 Languages.