Spatial Text Labelling

Prachi Gupta (12485)
Satyam Kumar Shivam (11917658)

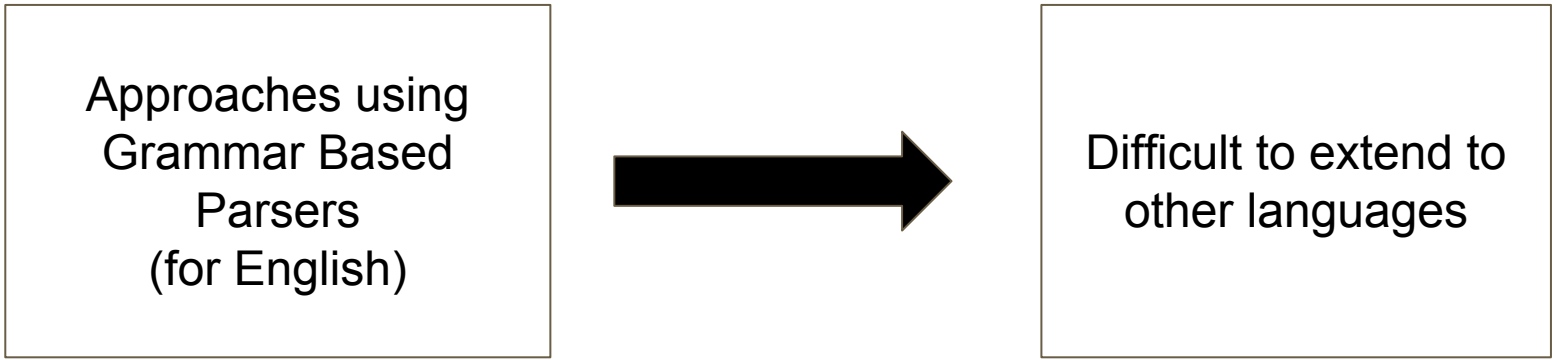
Problem Description

1. Identify and label spatial keywords in the text using Word2Vec and Neural Networks.
2. Find relationships among spatial keywords in English and Hindi corpus.

Motivation

Spatial Role labelling is useful in: [3]

1. Route recognition by identification of places and geo locations, as reported in a travel blog.
2. Interpretation of directions for moving towards a specific goal.
3. Integrating verbal spatial description with metric data derived from identification of landmarks.



Data Set

Obtained from Spatial Role Labelling shared task at SemEval-2012 and SemEval-2013 [6]

1. SpRL 2012
    a. static in nature
    b. For example: a lake in the forest
2. SpRL 2013
    a. dynamic in nature
    b. For example: I stepped into a small forest

Data Set

“A lake in the forest”

TRAJECTOR: lake; LANDMARK: forest; SPATIAL_INDICATOR: in

“I stepped into a small forest”

TRAJECTOR: I; LANDMARK: forest; SPATIAL_INDICATOR: into

Approach

1. Create Word Embeddings
    a. Build word vocabulary by training Google's Word2Vec on sufficiently large dataset.

2. Train Neural Network
    a. Train the neural network with word embeddings for trajectors and landmarks as input and embedding for spatial indicators as output.

Challenges

1. Requires many ‘TRAJECTOR-LANDMARK-SPATIAL_INDICATOR’ tuples to train a backpropagation neural network
    a. Used Wikipedia text as our data set

2. Spatial indicators, in the dataset, have more than one words eg. ‘in front of’, ‘on top of’, etc.
    a. Tried combining these multiple words into a single word eg. ‘in_front_of’, ‘on_top_of’, etc

Work So Far...

Parsing the Data Set

1. XML file where each child is a sentence description containing CONTENT, TRAJECTOR, LANDMARK, SPATIAL_INDICATOR and RELATION as tags.
2. Extracted trajectors, landmarks and spatial indicators.
3. Cleaned the data to use it with Word2Vec.

Creating Word Vectors

1. A 284140 x 50 size matrix created using Word2Vec on Wikipedia data, appended with the extracted landmarks, trajectors and spatial indicators.
2. Contains all trajectors and landmarks from SpRL 2012.

Future Work

1. Extend the Wikipedia data set to incorporate prepositional phrases.
2. Determine the parameters to be used in Neural Network and train it to predict prepositions, given a trajector and a landmark.
3. Test if the proposed model can adopt to the dynamic data set (SpRL 2013)
4. Extend the work for Hindi.

References

1. Kolomiyets, Oleksandr, et al. "Semeval-2013 task 3: Spatial role labeling." *Second joint conference on lexical and computational semantics (* SEM), Volume 2: Proceedings of the seventh international workshop on semantic evaluation (SemEval 2013)*. 2013.
2. Kordjamshidi, Parisa, and Marie-Francine Moens. "Global machine learning for spatial ontology population." *Web Semantics: Science, Services and Agents on the World Wide Web* 30 (2015): 3-21.
3. Pustejovsky, James, Jessica Moszkowicz, and Marc Verhagen. "A linguistically grounded annotation language for spatial information." (2013).
4. Clough, Paul. "Extracting metadata for spatially-aware information retrieval on the internet." *Proceedings of the 2005 workshop on Geographic information retrieval.* ACM, 2005.
5. Ramisa, Arnau, et al. "Combining geometric, textual and visual features for predicting prepositions in image descriptions." *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), Lisboa, Portugal, September. Association for Computational Linguistics.* 2015.
6. "Kolomiyets/sprl2013." GitHub. N.p., n.d. Web data repository.

Thank You !!

Any Questions ???