Preposition Disambiguation

Final Project Presentation by Team Mitron

The Problem

- Ambiguity is a central problem in NLP.
- Much of the work has been concentrated on preposition disambiguation because it is the most difficult.
- Eg: We graduate in May (temporal locality)
  Childhood Malnutrition is a major problem in India
  (spatial locality)

Importance of the Problem

- Prepositions are relational, therefore disambiguating prepositions helps in disambiguating all-word sense
- Different preposition senses might have different translations in a foreign language, therefore it helps in machine translations
- Preposition senses can be used as information extraction labels

Challenges

"You should book a room for 2 nights."
"For some reason, he is not here yet."
"I went there to get a present for my mother".

- Small Annotated DataSet.

- SemVal and WebReviews Corpus

.

Supervised Learning

- Collocation Features.
- Syntactic Features.
- Semantic Role based Features.

$$y = \underset{j}{\operatorname{argmax}}\, MLP_{sense}(\phi(s, i))[j]$$

Semi Supervised Learning

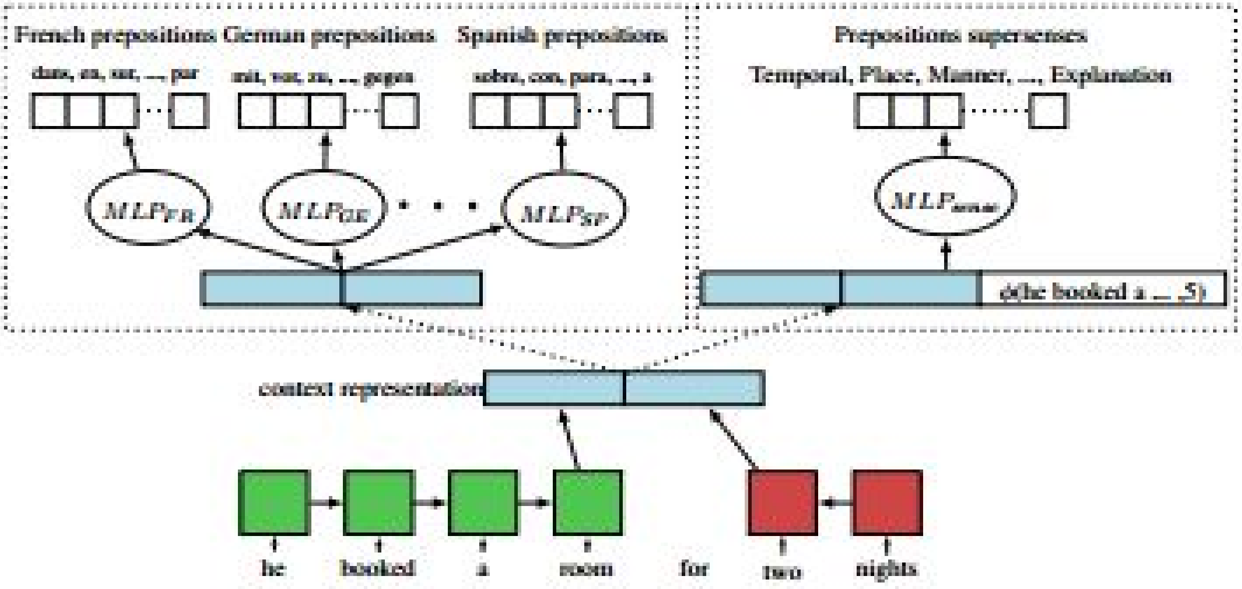
“What action will it take to defuse the crisis and tension ***in*** the region?”

“These are only available ***in*** English, which is totally unacceptable”.

- In the first sentence, the preposition “in” is translated into the French preposition “dans”, whereas in the second one, it is translated into the French preposition “en”.

$$ctx(s, i) = LSTM_f(w_{1:i-1}) \circ LSTM_b(w_{n:i+1}) \qquad \hat{p} = \operatorname*{argmax}_j MLP_L(ctx(s, i))[j]$$

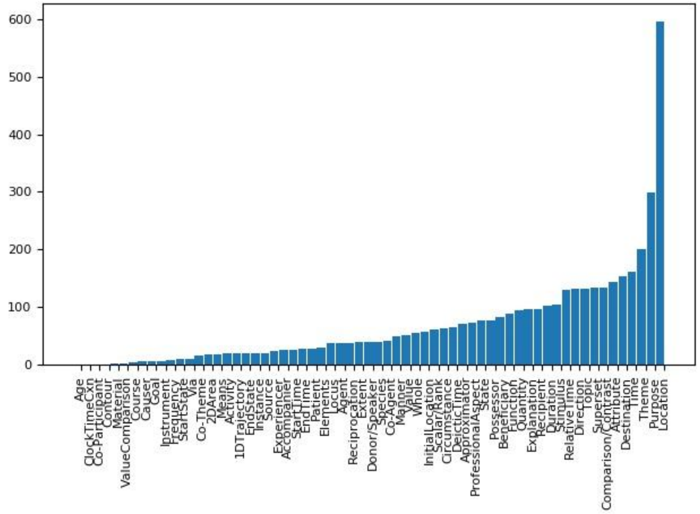
$$y = \operatorname{argmax} MLP_{sense}(ctx(s, i) \circ \phi(s, i))[j]$$



Datasets

1. <u>SemEval 2007 Corpus</u>-
   No. of prepositions: 34
   Training Samples: 16557
   Test Samples: 8096
   No. of senses for each preposition: 2-25

2. <u>Web Reviews Corpus</u>-
   No. of prepositions: 114
   No. of samples: 4250
   Training Samples: 3803 (3353+450)
   Test Samples: 447
   Unified sense inventory: 63 supersenses



Datasets-Continued!

3. MASC Word Sense Sentence Corpus-

   No. of distinct lemmas: 116

   No. of samples per lemma: around 1000

   Annotation done on WordNet senses.

4. Penn Tree Bank-

   2,499 stories taken from three year Wall Street Journal(WSJ) collection.

Pre-processing

- We have used Web Reviews Corpus.

- Each sample contains (TSV) the ID, Preposition, Class and Sentence.

- Each word in the preposition is surrounded by two ‘|’ symbols.

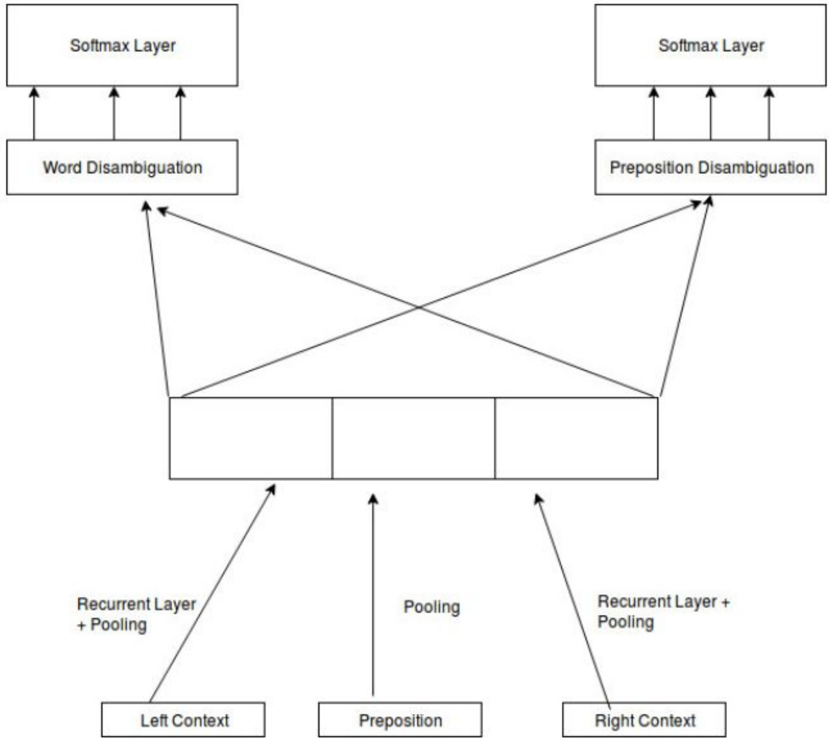
- The sentence is splitted on ‘|’ symbols to obtain the:
    - Left Context (LC)
    - Preposition
    - Right Context (RC)

Experiment Design

I. <u>Baselines</u>: MaxClass, MaxClassPrep, RandClass, RandClassPrep, UniformRandClass, UniformRandClassPrep
II. <u>Library Models</u>: SVC, KNeighborsClassifier, RandomForestClassifier, MLPClassifier
    Average and Max Pooling
III. <u>Feed Forward Neural Network</u>: Average and Max Pooling
IV. <u>Recurrent Neural Network</u>: Max Pooling
V. <u>Proposed Models</u>: Transfer Learning approaches

Approach 1-Combining with Word Disambiguation

- Consider training dataset for each word in WD dataset separately.
- Train the model, reusing recurrent layer parameters and throwing away rest of the parameters.
- Train and evaluate the model on preposition disambiguation dataset.



Approach 2-Using Language Modelling

- Predicting the next word in a sentence given a previous set of words!
- Train the recurrent layer on a language modelling task.
- Use the learned parameters to initialize the recurrent model for preposition disambiguation task.

Initial Results on 63 class problem

The following tables show the results obtained for the various models that we run:

Baselines

Model Name	Dev. Set	Test Set
MaxClass	0.135	0.136
MaxClassPrep	**0.411**	**0.425**
RandClass	0.040	0.044
RandClassPrep	0.275	0.277
UniformRandClass	0.020	0.024
UniformRandClassPrep	0.135	0.176

Feed-forward

Model Name	Pooling	Dev. Set	Test Set
Feed Forward NN	Average	0.495	0.508
Feed Forward NN	Max	**0.526**	**0.510**

Library

Model Name	Pooling	Dev. Set	Test Set
SVC	Average	**0.542**	**0.537**
KNeighborsClassifier	Average	0.424	0.425
RandomForestClassifier	Average	0.500	0.521
MLPClassifier	Average	0.535	0.492
SVC	Max	0.446	0.472
KNeighborsClassifier	Max	0.395	0.416
RandomForestClassifier	Max	0.511	0.525
MLPClassifier	Max	0.440	0.481

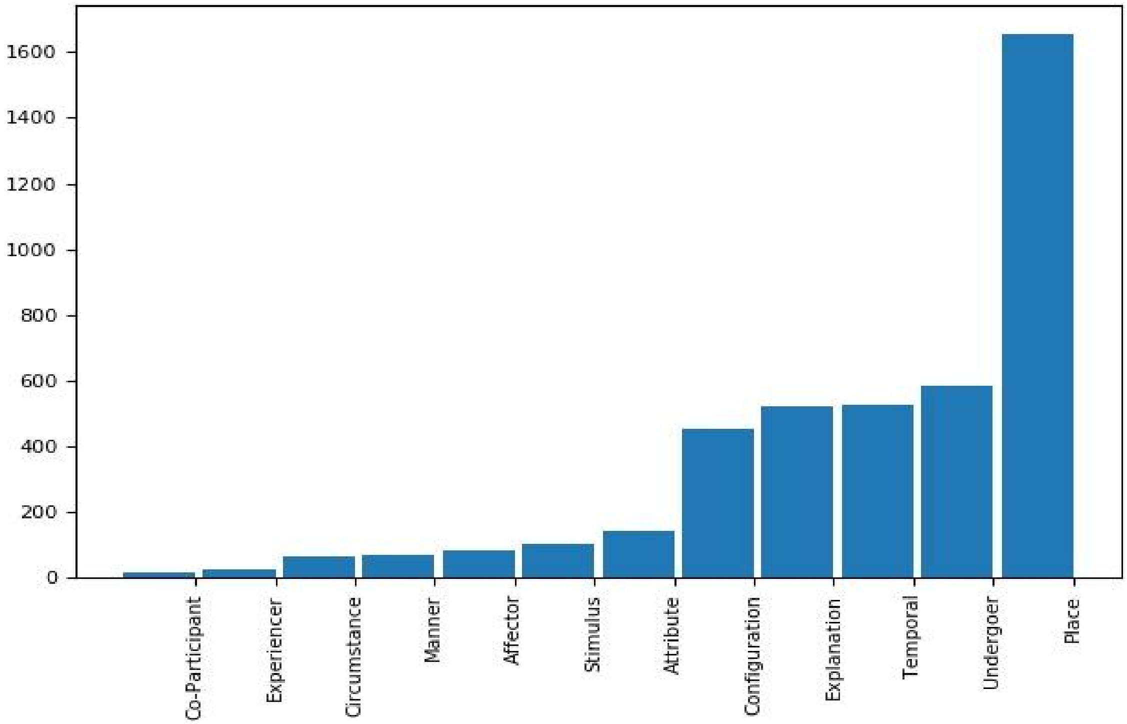
Recurrent models

Model Name	Dev. Set	Test Set
RNN	0.546	0.489
LSTM	0.557	0.523
GRU	**0.584**	**0.528**

Using 12 supersenses instead

Normalized Entropy Values:

- 0.22(12 classes)
- 0.43(63 classes)



Results on 12 supersense problem

Transfer learning approach(LSTM)

Model Name	Test Set
LSTM	0.684
LSTM+WD approach	0.708
LSTM+WD approach+ensemble	**0.719**

Recurrent models

Model Name	Dev. Set	Test Set
RNN	**0.727**	**0.684**
LSTM	0.711	0.669
GRU	0.707	0.667

Analysis

- Intuition behind transfer learning - Knowledge gained by learning word contexts from one scenario can be applied to another similar tasks.
- Effects on performance due to skewness in the dataset.
- Quality of Word disambiguation dataset doubtable - Varying accuracies for individual words.
- Improvement is similar, overall lower accuracies - Need for a more optimized implementation along with a more systematic hyperparameter tuning and trying similar word embedding.

Conclusion and Future Work

- Two proposed approaches using transfer learning.
- Observed improvement in performance using our method.
- Aim to devote more time to optimizations and word embedding.
- Efficiently implementing language modelling.
- Using transfer learning approaches on SemEval corpus for better comparisons with previous methods.

Thank You!