Department of Computer Science and Engineering
Indian Institute of Technology , Kanpur

CS 365 Artificial Intelligence
Project Report

Spatial Role Labeling

Submitted by
Satvik Gupta (12633) and Garvit Pahal (12264)

Under the guidance of
Dr. Amitabha Mukerjee

Abstract

One of the key aspects today in Natural Language processing is to describe the spatial relations between objects and their locations. As describes by the SemEval challenge, Spatial Role labelling involves extraction of spatial roles and their relations. In this project, prepositions being used in a spatial sense are found and the related landmarks and trajectors are extracted. Additionally, all the words are classified into categories of location, organization, person, date, time or none. Stanford Parser and Named Entity Recognition along with other learning techniques are used to produce good results on the CLEF corpus.

Contents

1 Introduction

1.1 Problem Definition

Spatial role labelling is the task of automatic labelling of words and phrases in natural language text with a set of spatial roles. These roles can be trajector, landmark, spatial indicator, distance, direction etc. In this task, we would be classifying spatial prepositions as spatial indicators and the use it to extract the trajectors and landmarks.

As defined by the SemEval Challenge 2012, these are the definitions of the various spatial roles:

- **Trajector**: Trajector is a spatial role label assigned to a word or a phrase that denotes a central object of a spatial scene or an object that moves.
  Example: ... [**the book**] on the desk ...

- **Landmark**: Landmark is a spatial role label assigned to a word or a phrase that denotes a secondary object of a spatial scene, to which a possible spatial relation (as between two objects in space) can be established.
  Example: ... the book on [**the desk**] ...

- **Spatial Indicator**: Spatial indicator is a spatial role label assigned to a word or a phrase that signals a spatial relation between objects (trajectors and landmarks) of a spatial scene.
  Example: ... the book [**on**] the desk ...

Generally, a spacial indicator can either be a preposition, verb, noun, adjective and so on. In this work, we are only focusing on preposition as a form of spatial indicators.

Additionally, all the words in a sentence are classified into different categories like person, organization, person, date, time or none using Stanford Named Entity Recognition. These may be used to improve spatial role labelling by restricting the search for trajectors and landmarks.

The task can thus be divided in three parts:

- The first part considers identifying the spatial indicators. This involves finding preposition that are used in a spatial sense.

- In the second part, we try to extract the trajectors and landmarks for each of the prepositions identified as a spatial indicator.

- The last part involves classifying all the words in a sentence using Stanford NER.

2 Motivation

Spatial role labelling is a key task for many natural language processing applications which require information about locations of objects referenced in text, or relations between them in space. For example, the phrase 'the book is on the desk' contains information about the location of the object book with respect to another object desk and this information may be used by a robot that is trying to pick up the book or by a software trying to do text-to-scene conversion.

Various applications of spatial role labelling includes controlling a robot by audio or textual instructions, navigation, traffic management, performing text-to-scene conversion and query answering systems.

3 Previous Work

- **SpatialML (Mani et al.), 2008:** This work focuses on geographical information rather than general spatial relations.

- **Pustejovsky and Moszkowicz, 2009:** Here, the verb is classified as a spatial verb or a temporal verb and both these classes have many subclasses.

- **O. Stock, 1997:** This work also involved both spatial and temporal Reasoning.

All these previous methods are either very specific and application-dependent or they they try to extract both spatial and temporal information at the same time. Also, none of these methods have focused on prepositions which are the most expressive in terms of the spatial information they can provide.

4 Dataset

These are the various datasets used in this project:

- **SemEval 2012 dataset:** The data we got SemEval 2012 challenge is a subset of the IAPR TC-12 Benchmark. This contains images taken by tourists with descriptions in different languages.

- **TPP dataset:** This dataset is from a SemEval-2007 task. The TPP dataset is used for preposition disambiguation.

- **Stanford CoreNLP:** The tools used in this project include the Stanford Parser and Named Entity Recognition. The datasets used in these tools are - CoNLL, MUC-7 and ACE datasets.

5 Methodology

There task is divided into 3 parts:

5.1 Part 1: Spatial Indicator Extraction

The steps involved in this part are:

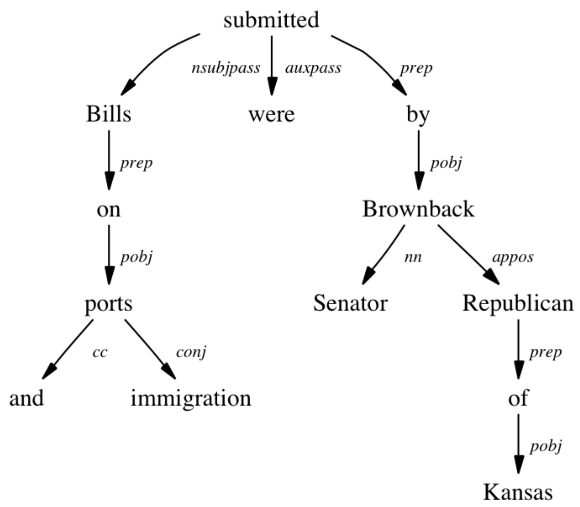
- **Part of Speech Tagging:** Each word in a sentence is tagged with its part of speech. Python's nltk library is used for carrying out POS tagging.
- **Dependency Parsing:** Based on the structure of the sentence and the POS tag attached to each word, a dependency graph is constructed (using the Stanford Parser).

  For example:(Example from ̈http://nlp.stanford.edu/software/stanford-dependencies.shtml ̈)
  Bills on ports and immigration were submitted by Senator Brownback, Republican of Kansas

  Dependencies are:

  nsubjpass(submitted, Bills)
  auxpass(submitted, were)
  agent(submitted, Brownback)
  nn(Brownback, Senator)
  appos(Brownback, Republican)
  prep_of(Republican, Kansas)
  prep_on(Bills, ports)
  conj_and(ports, immigration)
  prep_on(Bills, immigration)

Figure 1: Dependency graph



- **Preposition Disambiguation:** In this step, the feature set of the preposition is extracted using the dependency graph. The entities on which the preposition depends are named 'HEAD_1' and the entities which depend on the preposition are named 'HEAD_2'.

- **Spatial Sense Calculation:**Here, the probability of a preposition being a spatial indicator is calculated. Based on the number of occurrences of a preposition in a spatial sense and in a non-spatial sense in the TPP dataset, the probability of the preposition being a spatial indicator is calculated.

- **Identification of Spatial Indicators:** The feature set of the preposition (which includes the POS tags and the features extracted from the previous steps) is then used to classify whether the preposition is a spatial indicator or not using a Naive Bayes Classifier.

5.2 Part 2: Trajector and Landmark Extraction

For every spatial indicator in the sentence we consider every pair of word as landmark and trajector and find features for both landmark and trajector in the sentence and also tag each set of these features as 1 if it is correctly labelled as trajector or landmark else we tag it to 0

- **Trajector Extraction:** The feature set used to identify trajector in each sentence consisted of word, root word, part of speech tag related to the same.

- **Landmark Extraction:**The feature set used to identify trajector in each sentence consisted of word, its part of speech tag and wether that word lies on the left of the spatial indicator to the right of the spatial indicator

5.3 Part 3: Named Entity Recognition

This part involved classification of all the words in a sentence in the following categories:

- Location
- Organization
- Person
- Date
- Time
- None of the above

This classification can later be used to refine and improve the extraction of trajectors and landmarks.

Stanford NER library and CRF

The Named Entity Recognition was carried out using the Stanford NER library. It uses Conditional Random Field (CRF) models for classification. Conditional random fields as described in [1] are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for structured prediction. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account. It is an undirected graphical model or a Markov random field.

Let $x_{1:N}$ be the observations (e.g., words in a document), and $z_{1:N}$ be the labels (e.g., tags), then a linear chain Conditional Random Field defines a conditional probability
$p(z_{1:N}|x_{1:N}) = \frac{exp(\sum_{n=1}^{N}\sum_{i=1}^{F}\lambda_i f_i(z_{n-1},z_n,x_{1:N}))}{Z}$
where Z is the normalization factor so that the sum of probability is 1, we sum over n=1 $\ldots N$ word positions in the sequence. For each position, we

sum over i $= 1 \dots F$ weighted features. The scalar λ_i is the weight for feature $f_i()$. Example of feature function can be

$$fi(z_{n-1}, z_n, x_{1:N}) = \begin{cases} 1, z_n = \text{PERSON and } x_n = \text{John} \\ 0, \text{ o/w} \end{cases}$$

$$fi(z_{n-1}, z_n, x_{1:N}) = \begin{cases} 1, z_n = \text{PERSON and } x_{n+1} = \text{said} \\ 0, \text{ o/w} \end{cases}$$

The motivation behind chosing CRF over HMM for classification is that in HMM inorder to predict the z_n we only know $z_{1:n-1}$ and $x_{1:n}$ but in CRF we have the knowledge of $z_{1:n-1}$ and $x_{1:N}$ like in the above feature we have used the value of x_{n+1} to predict z_n . Also in CRF the we can also add the weights of the features like for a sentence 'John said' probabilty of John being labelled as Person increases

6 Results

6.1 Spatial Role Labelling

Some example results:

- The [**cat**]$_{\text{TRAJECTOR}}$ is [**under**]$_{\text{SPATIAL INDICATOR}}$ the [**table**]$_{\text{LANDMARK}}$.
- The [**pencil**]$_{\text{TRAJECTOR 1}}$ is [**on**]$_{\text{SPATIAL INDICATOR 1}}$ the [**bed**]$_{\text{LANDMARK 1}}$ and the [**bed**]$_{\text{TRAJECTOR 2}}$ is [**under**]$_{\text{SPATIAL INDICATOR 2}}$ the [**roof**]$_{\text{LANDMARK 2}}$.
- [**Clothes**]$_{\text{TRAJECTOR}}$ are hanging [**on**]$_{\text{SPATIAL INDICATOR}}$ the [**wire**]$_{\text{LANDMARK}}$.

Some example with incorrect results:

- [**He**]$_{\text{TRAJECTOR}}$ is [**on**]$_{\text{SPATIAL INDICATOR}}$ [**time**]$_{\text{LANDMARK}}$.
- [**Cake**]$_{\text{TRAJECTOR}}$ is not [**on**]$_{\text{SPATIAL INDICATOR}}$ the [**list**]$_{\text{LANDMARK}}$.

The accuracy of the extraction of the spatial roles is as follows:

Spatial Role	Accuracy	Current Best
Spatial Indicator	83%	88%
Trajector	76%	84%
Landmark	79%	91%

6.2 Named Entity Recognition

Some example results:

- [**Google**]$_{\text{ORGANIZATION}}$ [**Inc.**]$_{\text{ORGANIZATION}}$ is located in [**California**]$_{\text{LOCATION}}$.
- I went to [**IIT**]$_{\text{ORGANIZATION}}$ in the [**morning**]$_{\text{TIME}}$.
- Today is 22nd [**January**]$_{\text{DATE}}$.

The performance statistics are available on the Stanford NER website[2]

7 Limitations

These are some of the limitations of the proposed method:

- Only prepositions can be labelled as spatial indicators which is not always the case. A verb or a noun can also provide spatial information.
- Some words like 'time' or 'letter' can never be landmarks. Similarly, there are words that can never be trajectors. This information is not utilized and in sentences like 'Be on time', 'time' is classified as a landmark.
- Many prepositions are multi-word. They have not been handled in our method.
- While extracting trajectors and landmarks, context is not taken into account. Only their relation to the spatial indicator is a factor. Techniques taking context into account have performed better in these cases (for example: Conditional Random Field models).

8 Conclusion

In this project, we have presented a method for spatial language understanding and Named Entity Recognition. These techniques can be combined together to further refine and improve the task of Spatial Role Labelling. Much of this method's success stems from the fact that prepositions provide a substantial amount of spatial information. These prepositions along with words related to them give us spatial information. Several ways of extending and improving the method have also been documented. Improvement in this area would greatly help systems requiring spatial information extraction like robotic systems.

9 Future Work

Some of the improvements that can be made in this project are as follows

- The feature set for the spacial indicator classification can be improved by using word2vec instead of the current model which only only gives a static probability of a preposition being a spatial indicator based on the TPP dataset.
- Currently, only prepositions are being used as spatial indicators. The model can be extended to include other parts of speech as spatial indicators.
- The Named Entity Recognition can be taken into account to improve the prediction of trajectors and landmarks.
- Instead of just extracting the spatial roles like trajector, landmark and spatial indicator, the relations between them like path, direction, distance, motion etc. can also be extracted out.
- A method can be developed to incorporate multi-word prepositions as spatial indicators. For example - in front of, at the back of etc.
- The trajector and landmark extraction can be improved by incorporating linear chain conditional random field as some of the recent work in this area suggests.

10 Software used

- Python nltk library
- Stanford Parser
- Stanford NER library
- Code adapted from: `https://code.google.com/p/pln-pmt-pract/`
- Dataset used: TPP, CoNLL, MUC-7, ACE and SemEval 2012 dataset

References

[1] Spatial role Labeling: Task definition and annotation scheme, P. Kordjamshidi et.al, In Proceedings of the 7th Conference on International Language Resources and Evaluation(LREC) 2010, http://www.researchgate.net/publication/220745904_Spatial_Role_Labeling_Task_Definition_and_Annotation_Scheme

[2] Learning to Interpret Spatial Natural Language in terms of Qualitative Spatial Relations Parisa Kordjamshidi1, Representing space in cognition: interrelations of behavior, language, and formal models, Series Explorations in Language and Space, Oxford University Press, Oxford University Press, 2013, pp. 115146 https://lirias.kuleuven.be/bitstream/123456789/351891/1/KordjamshidiHoisOtterloMoens.pdf

[3] SemEval-2013 Task 3: Spatial Role Labeling , http://www.cs.york.ac.uk/semeval-2013/task3/

[4] Spatial Role Labeling: Towards Extraction of Spatial Relations from Natural Language By P. Kordjamshidi et.al, ACM Transactions on Speech and Language Processing, 8(3):article 4, 36 p, 2011 https://lirias.kuleuven.be/bitstream/123456789/321540/1/Kordjamshidietal-ACM-TSLP-2011.pdf

[5] http://nlp.stanford.edu/software/CRF-NER.shtml