**89680 Introductions to NLP – Relation Extraction**

1. We focused on the relation extraction of “Live\_In”  
   Our system is based on a machine-learning approach   
   External resources:

* spacy – we used spacy for Name Entity Recognition
* countries.json – a json file with a list of countries
* sklearn – for svm supervised learning technique

Model:

* SVM LinearSVC
* Given more examples our model could train better.
* We saw improvement when training the data with all connection types, since the classification was less hard.

Features:

* Entity type (for first and second entities)
* Word entity string (for first and second entities)
* A concatenation of entity types
* If entity is a location based on .json file
* Word string before first entity
* Word string after second entity
* Entity head based on dependency tree (for first and second entities)
* Bag of words between chunks
* Dependency tree - POS tag list
* Dependency tree - word string list
* Dependency tree – dependency type list

We also tried additional features which didn’t Improve / hurt accuracy:

* Distance between entities
* Entity structure (upper case / has numbers)
* Words that contain information about elections
* And more…

1. Error Analysis

Example of common recall errors:  
  
*None Live\_In / Kravchuk / Ukraine*

Kravchuk won over 60 percent of the votes and Chornovil , nearly 28 percent in Ukraine 's first presidential elections in December 1991 .  
  
In this error we tagged None where the answer was Live In. A prior information that winning an election in a country means you have to live there. Adding an election feature didn’t help to solve this problem.

Example of common precision errors:

*Live\_In None / William Miller / U.S*

( Text ) Crimean Supreme Soviet Chairman Mykola Bahrov met today with U.S. Ambassador to Ukraine William Miller .

In this error we tagged “Live\_In” where the answer was None. This error has previous knowledge that an ambassador of a country and to a country has different semantic meaning.

Example of very common errors which hurt both:  
*None Live\_In / Mrs. Cicippio / Beirut*

` ` Please don ' t execute him , please , ' ' said Mrs. Cicippio , who works at the American Embassy in east Beirut .  
  
*Live\_In None / Cicippio / Beirut*

` ` Please don ' t execute him , please , ' ' said Mrs. Cicippio , who works at the American Embassy in east Beirut .  
  
In this error our NER and the Golden NEW has extracted different entities (Cicippio vs Mrs. Cicippio ) which hurts both the precision and recall.

1. Results

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Relation | Train Recall | Train Perc | Train F1 | Dev Recall | Dev Perc | Dev F1 |
| Live-In | 0.85 | 1.0 | 0.92 | 0.42 | 0.61 | 0.50 |