Joint Named Entity Recognition and Relationship Extraction

Akanksha
Indian Institue of Technology, Guwahati
164101016
akanksha.chaudhary8

Bhawana Gupta
Indian Institue of Technology, Guwahati
164101036
bhawana2016

Shruti Ganapathy Subramanian
Indian Institue of Technology, Guwahati
164101006
shruti.ganesh

Smriti Bahuguna
Indian Institue of Technology, Guwahati
164101019
smriti.bahuguna

1. ABSTRACT

Now-a-days data is increasing by leaps and bounds in the Internet. In availability of such abundant amount of data, Information Extraction (IE) has emerged as a hot topic. IE targets on converting unstructured data into structured form. This converted structured data can be recognized by machine. This helps in optimization of Web Search Engine, Question-Answering Systems and many other application. IE consists of many tasks such as Segmentation, POS tagging, Recognition of Named Entities, labeling of Semantic Roles, Relation Extraction and Classification. Named Entity Recognition (NER) and Relation Extraction (RE) are closely related as RE is done between two named entities. Approaches for doing NER and RE are "Pipelined" and "Joint Approach". Pipelined approach first performs named NER on the text and then over found named entities, RE is done. But there are several shortcomings of pipelined approach :

(i) The errors are propagated from NER to RE which affect the overall performance of relation extraction.

(ii) It facilitates only one-way information flow. In this approach, the knowledge about entities is used for RE but not vice-versa. However, the knowledge about relations can help in correcting some NER errors.

To overcome these shortcomings of pipelined approach, joint approach is used. Joint approach can be adopted if a strong relation exists between the tasks. In many state-of-the-art works, joint approach has been adopted to combine two IE tasks. Joint approach can also improve the performance of both the tasks, i.e. NER and RE. For RE, named entities are required. If relation classification is accurate, it helps in recognition of named entities. Here, we have implemented a joint approach where feedback from RE task improves learned named entities which in turn help to eliminate noisy relations.

2. INTRODUCTION

Named entity recognition and relationship extraction are the most crucial steps in information extraction. A pipelined approach of first recognizing named entities and then finding relations between them loses the information for entity recognition task where a relation can help improve the type of entities. The task of improving recognized named entities with the help of relations and then using it to refine relations can be done using a joint model with feedback. The final goal is to compare the named entities recognized by the independent approaches using Conditional Random Field and Hidden Markov Model and the joint approach using feedback from Maximum Entropy to Conditional Random Field.

The main challenge when it came to improving performance was selecting the appropriate dataset specific to the model. Since CRF and ME are trained on separate datasets, any discrepancies between them can drastically affect performance. Therefore, an important design challenge to consider is the makeup of the datasets on which these models are trained on. A dearth of properly annotated free datasets is another considerable challenge in designing specific models. Further the feature selection is a very specific process for which multiple iterations are required to obtain competitive results . Both CRF and ME heavily depend on features for tagging and relation extraction.

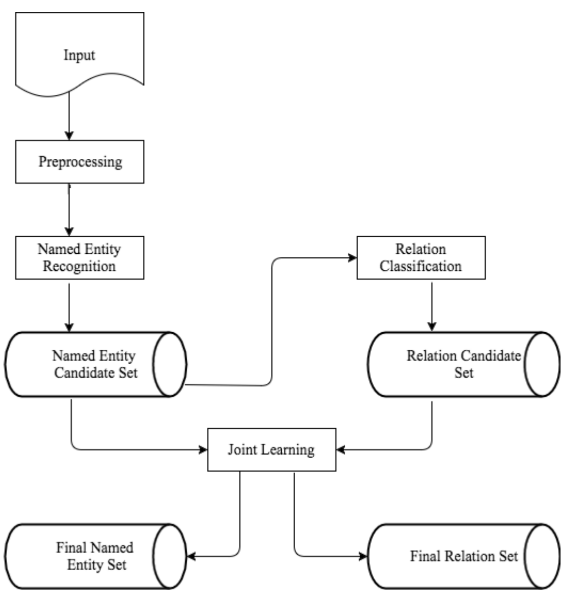
In this direction, several models have been employed like Hidden Markov Models, structured Perceptron with beam search, table filling methods using RNN, card pyramid parsing,etc.A card-pyramid compactly encodes the entities and relations in a sentence thus reducing the joint extraction task to jointly labeling its nodes. Also, during extraction from a part of the sentence it also allows use of features based on the extraction from its sub-parts, thus leading to a more integrated extraction.This approach considers relations between two entities.

Thus in-order to generalize relations containing more than two entities, a way to binarize these relations must be available. Performance of this model can be further improved by using a classifier which uses more linguistic information such as including word-categories and their POS tags in subsequences. The table filling method using RNN for this joint task detects both entity mention pairs and the corresponding relations in a single framework with an entity-relation table representation. But it does not make use of improving learning of named entities with the help of relations learned or vice-versa. While CNN has been effectively used for named entity recognition and relation extraction, but joint learn-

ing of these two tasks has not been achieved as generalized as other methods, attributed to the fixed length sized sentences. So we are looking into a joint method that has scope for improvement by changing the ways in which named entities can be learned and then used to learn relations between entities.
The joint model using feedback from extracted relations helps to improve the recognized entities because knowing a relation type between two entities in a sentence gives the information about the types of entities also. An independent approach for entity recognition task loses this information and hence such entities are missed out which can be extracted with the help of a relation identified between them. Similarly a better set of learned entities would help in learning appropriate relations between them and hence discard garbage relations due to wrongly identified entities.

3. METHOD

Our method has the following three sub-tasks:
A. Named Entity Recognition,
B. Relation Classification and
C. Joint Learning.
Framework for joint approach is shown in figure.



A. Named Entity Recognition

In our experiment, two types of named entities are considered: names of persons and organizations. Organization includes companies, schools, universities, governments and so on. We implemented two approaches for named entity recognition:
(i) Hidden Markov Model(HMM)
(ii) Conditional Random Field(CRF) Model

Description of approaches:
(i) HMM: Hidden Markov Model is generative model, it specifies a joint probability distribution over observation and label sequences. The parameters are then trained to maximize the joint likelihood of training sets. This method is easy to use since it requires minimum efforts for Named Entity Recognition in any natural language. We need to annotate the corpus and test the system for any sentence.

Three main steps that we followed for HMM model :
a. Data Preparation for experiment:
Converting the dataset into acceptable format for HMM.

b. Parameter Estimation (Training): Steps for parameter estimation are:
1: Find states: State is a vector contains all the named entity tags candidate interested. For each tag in annotated text file, if it is not already present in state, Add to state vector.
2: Calculate Start probability (Π): For each starting tag, Find starting probability that a sentence start with a particular tag.

$$\text{start probabilities} = \frac{\text{No.of sentences starting with specific tag}}{\text{total number of sentence in the corpus}}$$

3: Calculate transition probability (A) : For each tag in state vector (T_i), for each other tag (T_j) in state vector, transmission probability is to be calculated.

$$\text{transition probabilities} = \frac{\text{Number of sequences from } T_i \text{ to } T_j}{\text{Number of } T_i}$$

4: Calculate emission probability (B) : For each unique word W_i in annotated corpus, calculate emission probability which is the probability of assigning particular tag to the word in the corpus.

$$\text{emission probability} = \frac{\text{Total occurrence of word as a tag}}{\text{total occurrence of that tag}}$$

c. Test the system:
After calculating all parameters, these parameters are applied in Viterbi algorithm and test sentence as an observation to find named entities.
HMM result for Named Entity Recognition shown in result section.

(ii) CRF: A CRF is a discriminative sequence tagging model. The parameters are learned using maximum likelihood approach. Our model uses CRF to label NER tags specifically PER and ORG tags. We have used two CRFs according to the different types of features for person and organization names. The following are the common features for recognizing PER and ORG entities:
1. tokens at [-2,+2] window
2. POS tags at [-2,+2] window
3. Chunk(phrase) tags at [-2,+2] window
4. previous token's output (NER) tag
These features were stored into template file that is used at the time of training to generate model. In addition to the above features for PER, we have also included a feature that indicates whether a token contains punctuation or digits which a name would generally not contain. But we have not included this feature for ORG because an organization might consist of digits or punctuations. Similarly adding gazetteer features for persons and organization will help in recognizing them properly. Thus we have used two CRFs for creating model files for recognizing names and organizations.
CRF accuracy for Named Entity Recognition is better compare to HMM, so we made our joint model with CRF as NER model and Maximum Entropy as RE model.

Relations (Between Person and Organization)
Head
Director
Analyst
Lawyer
Professor
Founder
Adviser
Leave
Staff
Spokesperson
Others

Table 1: Relations

B. Maximum Entropy Model:
The maximum entropy model which is used for relation classification. It is a statistical model which tries to find a uniform distribution of probability given a set of constraints. Constraints are designed based on the data and its context. Thus, this does not make any independence assumptions about features. To guide the development of the maximum entropy model, first a set of features were selected. The features selected for this process include two words before the first entity occurrence, two words after the second entity occurrence and words in between. These are know as contexts and different contexts can be created by using more or less information pertaining to the entities. Once the features are selected, they are fed into the model which trains itself based on LBFGS approach. LBFGS is an iterative method for solving unconstrained nonlinear optimization.
The dataset presented a diverse number of relations between person and organisation named entities. The context is represented with tags in Maximum entropy toolkit. They are given in table 1 and 2.

Position	Feature	Description
BEFORE	SM1F	First word before M1
BEFORE	SM1L	Second word before M1
BETWEEN	WBNULL	No word is in between
BETWEEN	WBFL	One word exists between
BETWEEN	WBF	First word in between
BETWEEN	WBL	Last word in between
BETWEEN	WBO	Other words in between
AFTER	SM2F	First word after M2
AFTER	SM2L	Second word after M2

Table 2: Tags corresponding to context

C. Joint Learning:
In the joint learning part of the model, the CRF provides best-3 person entities and best-5 organization entities which we will consider as our candidate set. The CRF also provides best-1 person entity and best-1 organization entity which is our initial set. The relation classification module classifies the type of relation between every pair of persons and organizations given in the candidate set and adds these relations to a relation set R for every sentence. Thus R contains the set of best relations existing between every pair of entity in a sentence and each relation in this set is between a person-organisation pair. Amongst these, a relation is chosen having the highest probability and whose relation type is not 'other' . Thus, the final set of entities is updated by resolving conflicts between the initial set and the person-organisation pair of the best relation chosen using the following threshold measures:
1. if the prob(relation) > 0.9, then name is updated
2. if the prob(relation) > 0.8, the organisation is updated

4. EXPERIMENT

Purpose of the experiment:
The purpose of this experiment is to prove the correlation between NER and RE and use that to improve the performance of NER.

Experiment settings:
Our training dataset for CRF contains eng.train from CoNLL databank. The ME tool was trained on IJCNLP2011 NewYork Times dataset which contains around 500 lines of news sentences taken between the years 2009 and 2011. The test data for our model consisted of eng.testb from CoNLL databank. For CRF, the open source CRF project CRF++ is used. Likewise, for ME, the Maximum Entropy modeling toolkit was used. Our HMM model is using CoNLL dataset. For training purpose, "eng.train" file is used and for testing "eng.testb" is used. Training Data: eng.train contains data column wise. First column is sentence, second column is POS tags, third column is chunking information and the fourth column is NER tag. HMM takes data for training in the following form:
– for each sentence, first row contains sentence itself.
– second row contains POS tags for tokens(words) in the sentence.
– third row contains NER tags.
Testing data: eng.testb contains data column wise. First column is sentence, second column is POS tags, third column is chunking information and the fourth column is NER tag. HMM takes data for testing in the following form:
– for each sentence, first row contains sentence itself.
– second row contains POS tags for tokens(words) in the sentence.
– third row contains indices.

5. RESULTS

The performance of named entity recognition task for PER and ORG entities using independent approach and joint approach is summarized in table 3. The joint learning of named entities and relation between them, using a feedback approach gave results very close to the named entity recognition task using independent approach. The joint model could not give better performance than the independent approach because of fewer types of relations learnt by the relation extraction model between PER and ORG entities. Hence all relations between these entities could not be extracted and thus the relation extraction model failed to generate these entities into expected entity sets for names and organization which further discarded these entities from final entity sets. If the types of relations learnt by the relation extraction model can be increased, all such relations will be extracted by the model and hence their named entities will be generated into final named entity sets.

Table 3: Results

Models	Performance Measure	PER	ORG
Independent Approach (using CRF)	Precision	79.17%	77.10 %
	Recall	83.43 %	70.74 %
	FB1	81.24 %	73.78 %
Independent Approach (using HMM)	Precision	72.40 %	70.95 %
	Recall	74.54 %	65.68 %
	FB1	78.25 %	70.33%
Joint approach (using feedback approach)	Precision	83.69 %	76.63 %
	Recall	79.04 %	64.96 %
	FB1	81.30 %	70.32 %

6. FUTURE WORK

Future work should focus on better feature selection for both CRF and ME to improve their prediction performance which in turn would improve the performance of the joint model. Also, joint learning could also be improved by using different threshold values for accepting names or organizations in the feedback set and including better methods for conflict resolution. Future work for HMM could include finding n-best results for NER and using it along with the joint learning model.

7. CONTRIBUTION

Akanksha - NER using HMM
Bhawana - RE using MEM
Shruti - Joint Learning of NER and RE using MEM
Smriti - NER using CRF

8. REFERENCES

1. Qiyuan Xu, Fang Li - Joint Learning of Named Entity Recognition and Relation Extraction - 2011
2. Adam L. Berger, Stephen A. Della Pietra, Vincent J. Della Pietra - A Maximum Entropy Approach to Natural Language Processing - 1996
3. Sudha Morwal, Nusrat Jahan, Deepti Chopra - Named Entity Recognition using Hidden Markov Model (HMM) - 2012