TRDDC @ FIRE 2013: System for Classification of Legal Propositions

Report for Legal Track at FIRE 2013

Nitin Ramrakhiyani and Sachin Pawar

Tata Research Development and Design Centre (TRDDC),
Tata Consultancy Services Ltd., Pune
{nitin.ramrakhiyani, sachin7.p}@tcs.com

Abstract. In this work, we describe a system for classification of legal propositions using a multi-class maximum entropy classifier. We designed various features to capture the underlying characteristics of legal propositions. The best performing feature set was chosen by 10-fold cross-validation over the training set. The system achieved the best F-measure of 62.7%.

1 Introduction

This paper presents our Legal Proposition Classification System, which is task 2(b) in the Legal Track at FIRE 2013. We experimented with various different classifiers like Naïve Bayes, Decision Trees, Support Vector Machines (SVM) and Maximum Entropy classifier. We also designed several features for this task, some which are standard features like bag-of-words and some are legal domain specific features. Maximum Entropy classifier was found to be working better than the other classifiers. Here, we will describe our experiments, features and evaluation for this particular classifier. This paper is arranged as follows: Section 2 describes our overall methodology along with description of various features used and Section 3 describes our evaluation strategy along with the results.

2 Methodology

Our initial attempt at building the legal propositions classifier was to try the most widely used text classification method – Naïve Bayes. It is a generative model which uses just the bag of words as the features. Error analysis of the output of this classifier revealed many characteristics of the data which were very specific to this problem and especially to the Legal domain. To capture all these characteristics, we had to design a set of additional features.

The generative models require independent and non-overlapping features, hence feature designing is generally considered to be a difficult task for them. Maximum Entropy models proposed by Berger et.al.[1] are discriminative models which don't require any independence assumptions for features. Hence, we modeled the legal propositions classification problem using a maximum entropy model, as it handles multiple dependent and overlapping features well. Maximum Entropy classifier assigns weights to each feature and probability of class y, given input x is calculated as follows:

$$P(y|x) = \frac{\exp\left(\sum_{i=1}^{N} \lambda_i f_i(x,y)\right)}{\sum_{y'} \exp\left(\sum_{i=1}^{N} \lambda_i f_i(x,y')\right)}$$

Here,

- y : Class label, in this case "type of proposition"
- x : Instance to be classified, in this case "proposition"
- $f_i(x,y)$: i^{th} feature. Feature for Maximum Entropy model is combination of some characteristic and class label. For example,
  $f(x,y) = 1$ if x contains at least one PERSON name and y = "FI"
  $f(x,y) = 0$ otherwise
- λ_i : Weight of the i^{th} feature
- N : Total number of features

Table 1 describes various features that we used to model this problem of legal propositions classification.

Sr. No.	Feature	Feature Description
1	Bag of words	Here, we consider the proposition as an unordered set of words. Each word itself acts as a feature. We remove some obvious stop words like *the, a, an, in, of, and, or,* etc.
2	PERSON	This feature indicates that at least one instance of PERSON Named Entity is present in the proposition
3	ORGANIZATION	This feature indicates that at least one instance of ORGANIZATION Named Entity is present in the proposition
4	LOCATION	This feature indicates that at least one instance of LOCATION Named Entity is present in the proposition
5	PROP_LENGTH	It was observed that the length of proportion is one of the important factors deciding its type. This feature encodes the length of the proposition as follows: • SHORT – Proposition length less than 30 words • MEDIUM – Proposition length less than 50 words but more than 30 words • LONG – Proposition length more than 50 words
6	MONTH_PRESENT	This feature indicates that at least one month name is present in the proposition (E.g. *Jan, February, etc.*)
7	YEAR_PRESENT	This feature indicates that at least one year is present in the proposition. (E.g. *1970*)
8	MONETARY_	This feature is captures the presence of any monetary

	VALUE	value in the proposition. (E.g. *Rs.27,260*)
9	POS_PROP	One of the important factors for predicting the type of proposition was observed to be the position of the proposition in the file. File was divided in k sections containing equal number of propositions. This feature indicates the section number of the proposition in the file. Different values of k were tried and value of k=*6* was found to be the best by 10-fold cross-validation. This is motivated from the work of Hachey et.al.[2]
10	FIRST_PERSON	This feature indicates that at least one first person pronoun is present in the proposition. (E.g. *I, we, us, our, my*)
11	SECOND_PERSON	This feature indicates that at least one second person pronoun is present in the proposition. (E.g. *you, your*)
12	THIRD_PERSON	This feature indicates that at least one third person pronoun is present in the proposition. (E.g. *he, him, his, she, her, they, their, etc.*)
13	PAST_TENSE	If at least one verb in the past tense form is present in the proposition, this feature is fired.
14	FUTURE_TENSE	If at least one modal verb indicating future tense is present in the proposition, then this feature is fired.
15	SECTION	This feature indicates that some section/sub-section is mentioned in the proposition. (E.g. *section 2(14)*)
16	PMI-based features	We used one feature for each proposition type, which indicates whether at least one cue word for that particular proposition type occurs within the proposition. The set of cue words for each of the proposition type was created using Point-wise Mutual Information (PMI) based scores. For each pair of word (*w*) and proposition type (*t*), following score was computed: $PMI(w,t) \propto Score(w,t) = \frac{No.of\ propositions\ of\ type\ t\ containing\ w}{No.of\ propositions\ of\ type\ t \times Frequency\ of\ w}$ For each proposition type t, top k words according to the above score are selected as a set of cue words for the proposition type t.

Table 1: Description of various features used

3 Evaluation

We experimented with various classifiers like Decision Trees, Support Vector Machines (SVM), Naïve Bayes and Maximum Entropy. The best performance was obtained by the Multi-class Maximum Entropy classifier using the features described in Table 1. The performance of the classifiers was measured in terms of the accuracy obtained in 10-fold cross-validation over the training data. The accuracy was micro-averaged over various proposition types, i.e.

$$Accuracy = \frac{No.of\ correctly\ classified\ propositions\ in\ the\ test\ data}{Total\ no.of\ propositions\ in\ the\ test\ data}$$

	No. of files	No. of propositions
Training data set	10	712
Test data set	5	390

Table 2: Data sets of Legal propositions used in the experiments

Table 2 shows details about the data sets used for our experiments. We created data folds at the file level instead of the proposition level, because file level folds more closely resemble the testing scenario. As there were only 10 files in the training data set, in each fold we used 9 files for training the classifier and a remaining file for testing.

We obtained the best accuracy of **62.78%** using a Maximum Entropy model based on the features in table 1. The corresponding confusion matrix is as shown below in table 3.

		Assigned Labels								
		I	FI	LR	R	SS	A	SP	FE	SO
Actual Labels	I	1	5	0	4	1	0	0	0	0
	FI	1	189	11	41	1	1	1	0	1
	LR	0	26	51	17	2	0	0	0	0
	R	0	25	12	186	3	0	2	0	1
	SS	0	13	6	13	17	0	0	0	0
	A	0	9	6	4	1	1	0	0	0
	SP	0	4	1	16	1	0	2	0	1
	FE	0	4	0	7	0	0	0	0	0
	SO	0	1	0	23	0	0	0	0	0

Table 3:Confusion Matrix obtained by 10-fold cross-validation on the training data

For producing results on the test data, we used the same feature set which gave us the best 10-fold cross-validation accuracy and used complete training data (10 files, 712 propositions) for training the classifier.

References

[1] Berger, Adam L., Vincent J. Della Pietra, and Stephen A. Della Pietra. *"A maximum entropy approach to natural language processing."* Computational linguistics 22.1 (1996): 39-71.

[2] Hachey, Ben, and Claire Grover. "Sentence classification experiments for legal text summarisation." In Proc. 17th Annual Conference on Legal Knowledge and Information Systems (Jurix-2004), pp. 29-38. 2004.