Classification of Clauses in Non-Disclosure Agreements (NDAs)

Rida Hijab Basit

20th January, 2015 | Center for Language Engineering (CLE)

Overview

- Non-Disclosure Agreements (NDAs)
- Examples of Clauses in NDAs
- Pre-processing
- Feature Extraction
- Dataset
- Classification
- Results

Non-Disclosure Agreements (NDAs)

- Non-Disclosure Agreement is a legal contract between at least two parties that outlines confidential material, knowledge, or information that the parties wish to share with one another for certain purposes, but wish to restrict access to or by third parties.

Examples of Clauses

- THIS AGREEMENT (the 'Agreement') made as of the 1st day of December, 2013 BETWEEN: Bank of Montreal, a Canadian chartered bank, with an office at 100 King Street West, Toronto, Ontario, Canada, M5X 1A1 (called 'BMO') - and - Vaultive Inc., having an office at 489 Fifth Avenue, 31st Floor, New York, NY, U.S.A, 10017 (called \"Supplier\")

- 2.6 Notwithstanding the foregoing, BMO may disclose Confidential Information of the Supplier to any member of the BMO Financial Group for any purpose without a written confidentiality agreement in place between BMO and such member of BMO Financial Group.

Data Format

- Legal contracts in the form of text files.
- Contracts consist of various clauses/sentences that need to be classified

Data Pre-Processing

- Has be divided into three phases
  - Tokenization (Sentence Segmentation)
    - Based on full stop & question mark
    - Full Stop can also come at some place other than the end of the sentence like Dr., Mr., John F. James etc.
    - To handle this, an exception list has been generated
  - Cleaning (Removal of stop words)
    - Words like "the", "of" etc.
  - Stemming (Reduction of words to their stems)
    - Receiving, received, receives all stemmed to receive

Feature Extraction

- Lexical level features have been used. These are:
  - Bag of Words (Window Size = 3 – 5)
  - N-grams (N = 1-3)
- For each feature, its TF-IDF values have been computed
- TF-IDF stands for Term Frequency – Inverse Document Frequency

Dataset

- Total labels = 29
- Total sentences = 7926 (Marked as clauses and assigned labels manually)
- Selection of Training and Testing Dataset
  - Training Instances = 6342
  - Testing Instances = 1584

Classification of Clauses in Non-Disclosure Agreements (NDAs)

Classes	No. of Sentences
Parties Bound	567
Inclusion of affiliates	60
Unilateral agreement	185
Mutual Agreement	210
Business Purpose	243
Definition of confidential information	421
Publicly available information carveout	232
Already in possession carveout	167
Received from a third party not obligated carveout	164
Independently developed without use of confidential information	145
Disclosure required by law carveout	407
Trade Secrets covered	97
Includes information indirectly disclosed	11
Use restrictions	273
Record keeping obligation	20
Return or Destroy Information	292
Certification obligation	102
Non-Solicitation	771
Non-Contact	31
Exception for ordinary course	7
Indemnification	623
Survival of obligations	323
Period specified	124
Terminates when definitive agreement signed	48
Remedies	453
Including equitable relief	950
Governing Law	946
Residuals	45
Gramm-Leach-Biley	9
Total	**7926**

Classification

- Various classification algorithms have been tested using Weka (Ian H. Witten, 2000) data mining software.
- Classification Algorithms include:
  - Support Vector Machine (SVM)
  - Decision Tree
  - Random Forest
  - Naïve Bayes
  - Bagging

Flat-Structure Classification

- First, flat-structure classification was adopted
- Tested each feature vector with different classification algorithm

Features	SVM	Decision Tree	Naïve Bayes	Bagging	Random Forest
N-grams (Unigram Cutoff = 50 and Bigram Cutoff = 30)	63.64%	55.0505 %	41.0354 %	54.4192 %	57.3864 %
Bag of Words (Window Size = 3, Unigram Cutoff = 100)	58.59%	55.303%	54.9874 %	53.5354 %	56.5025 %
Bigrams (Cutoff = 40)	56.57%	51.7677 %	36.4899 %	50.947 %	51.1364 %
Unigrams	63.57%	57.2601 %	42.6136 %	53.5985 %	58.5859 %

Table 1: Flat-Structure Classification Result Analysis

Two-Level Classification

- Based on experiment results and confusion matrix analysis, two-level classification has been used.
- Classes with higher confusion are merged resulting into 13 classes at Level 1
- Level 2 classification is then performed on merged classes
- At level 2, 8 different classifiers have been developed with local features

Level 1 Classification

Classification Algorithms	Accuracy
Decision Tree	79.143%
Random Forest	82.9868%
Naïve Bayes	67.1708%
Bagging	80.5293%
SVM	**87.21%**

Table 2: Level 1 Classification Result Analysis

Level 2 Classification

Classification Algorithms	Average Accuracy
Decision Tree	73.66%
Random Forest	79.94%
Naïve Bayes	72.56%
Bagging	**79.95%**
SVM	69.10%

Table 3: Level 2 Classification Result Analysis

Overall System Performance

- Based on detailed analysis and experimental results, SVM for Level 1 and Bagging for Level 2 has been selected
- Using these algorithms, the overall system accuracy turns out to be **78.60%**

Related Issues

- Some labels had less data thus decreasing its accuracy.
- Some clauses in the training data were given multiple labels.
- Tokenization issues.

Possible Solution

- Some of the issues can be resolved by using Rule Based Systems (RBS) before the process of classification

References

- Ian H. Witten, E. F. (2000). *Data mining: practical machine learning tools and techniques with Java implementations.* San Francisco: Morgan Kaufmann .

Thank you