

Automatic Text Annotating: Application on First Samples of Comments

**Automatic Text Annotating:
Application on First Samples of Comments
Regarding Conventions 87 and 98**

Djamel A. Zighed, Hakim Hacid, Vincent Pisetta

Laboratoire ERIC, Université de Lyon2

September 2005

1. Introduction

In this document, we describe the performed analysis in order to predict violations for unlabelled texts. First, we remind some important details about the analysis (corpus and construction of a representation space). Finally, we show the obtained results from machine learning. We will remind the description of violations to be predicted.

2. Reminders

2.1 Corpus

We have two different conventions, n°87 and n°98. Each of them is considered as a corpus of documents. Each corpus is divided into two groups as illustrated on fig.1.

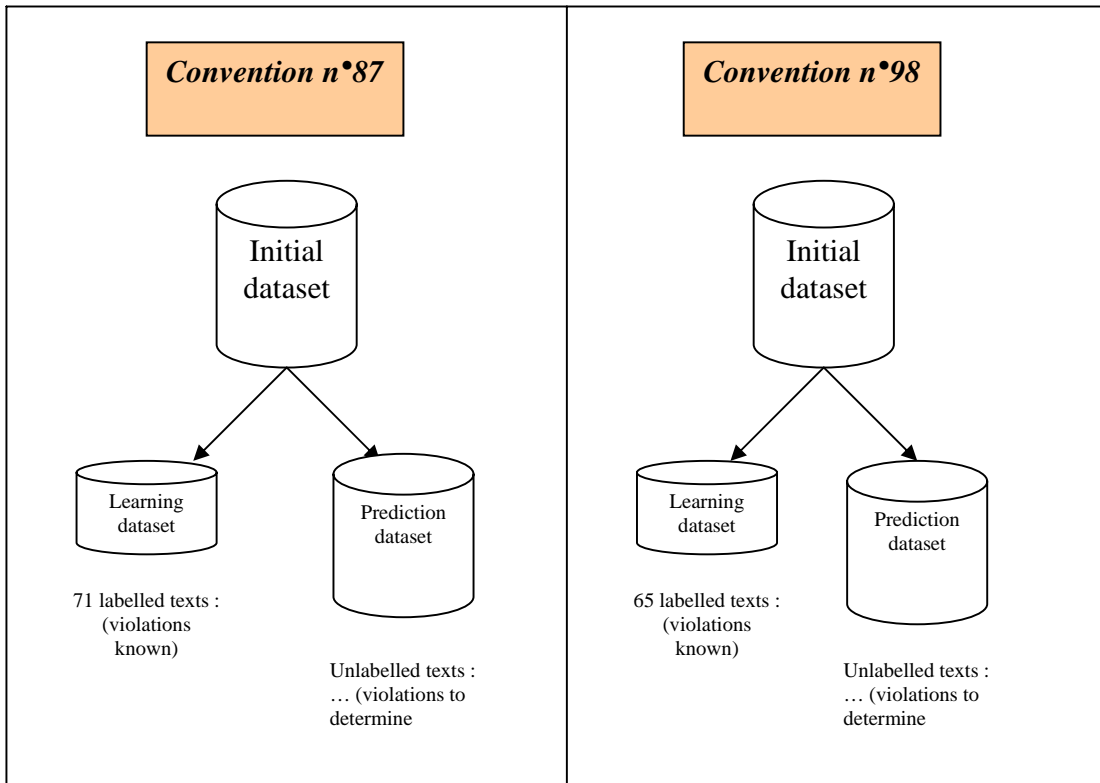


FIG.1 Illustration of the corpus

2.2 Construction of a representation space

A representation space is a mathematical structure able to represent texts in order to analyze them by datamining tools. As described in our previous report, the analysis phases are:

- Terminology extraction (find words and association of words while keeping their meaning)
- Concepts creation (gathering of words and association of words extracted by terminology extraction).
- Use of prediction models (*proximity graphs* and *decision trees*) in order to find key-concepts (violations) in unlabelled texts.

3. Preliminary studies

3.1 Concepts extraction

Several concept lists have been created. A final list which integrates the main candidate-terms (words and association of words) seems to capt the essential information contained in the corpus.

3.2 Key concepts

A key-concept is a “violation”. It is a feature to be predicted. At the beginning of the study, the Convention n°87 had 17 key-concepts and the Convention n°98 had 10 key-concepts. Some key-concepts have been grouped in order to have a better identification of them. We describe the gatherings here after:

Convention n°87	
Group 1: Civil Liberties Pertinent for Freedom of Association	Key Concepts 1-3: Right to Organise and Civil Liberties
Right to life and physical integrity	Right to life and physical integrity
Right to liberty and security of person / Right to a fair trial	Right to liberty and security of person / Right to a fair trial
Protection of property	Protection of property/Financial independence
Group 2: Right to Establish Trade Unions and All Concomitant Rights	Key Concepts 4-9: Right to Organise
Administrative/Financial Independence	Exclusion from the right to establish or join workers' organisations
Right to establish and join workers' organisations	Trade union pluralism
Trade union pluralism	Approval and registration of workers' organisations
Establishment and registration of workers' organisations	Dissolution or suspension of workers' organisations
Establishment of federations and confederations	Election of representatives / Eligibility criteria
Approval and registration of Constitutions and by-laws	Organisation of activities
Dissolution or suspension of workers' organisations	Key Concepts 10-11: Right to Industrial Action
Election of representatives / Eligibility criteria	Restriction of the right to industrial action
Organisation of activities	Minimum service
Group 3: Right to Industrial Action	Compulsory arbitration in the context of industrial action
Restrictions on the right to industrial action	
Conditions for lawful industrial action	
Penalties for instigation of, or participation in, industrial action	
Compulsory arbitration in the context of industrial action	
Minimum service	

Key-concepts merged

TAB. Key-concepts gathering (Convention n°87)

Convention n°98	
Key Concepts: Protection against Anti-Union Discrimination and Interference Acts	Key Concepts : Protection against Anti-Union Discrimination and Interference Acts
Anti-union discrimination	Anti-union discrimination
Acts of interference	Acts of interference
Solidarist associations	Solidarist associations
Key Concepts : Collective Bargaining	Key Concepts : Collective Bargaining
Promotion of free and voluntary collective bargaining	Promotion of free and voluntary collective bargaining
Exclusion from the right to collective bargaining	Exclusion from the right to collective bargaining
Designation of the bargaining partner / Most representative trade union	Designation of the bargaining partner / Most representative trade union
Restrictions on the level and scope of collective bargaining	Restrictions on the level and scope of collective bargaining
Compulsory arbitration in the context of collective bargaining	Compulsory arbitration in the context of collective bargaining
Restrictions of the negotiable issues and substantive outcomes of collective bargaining	Restrictions of the negotiable issues and substantive outcomes of collective bargaining / Approval and registration of collective agreements
Approval and registration of collective agreements	

TAB. Key-concepts gathering (Convention n°98)

We have thus defined a new list of key-concept. Key-concepts situated on the right side of the table are violations to be predicted. The next step of the study is the construction of prediction models. After this construction, it is necessary to test the identification's power of the models. Results are described in the next section.

4. Prediction on a sample test

We have extracted 20 unlabelled texts from each corpus and we applied prediction models. The results are presented here after:

- *Convention n°87*

Decision trees

Country	Violation	Prob.
Mali 2001	Right to establish and join workers' organisations	96%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%
	Election of representatives / Eligibility criteria	22%
Slovakia 2001	Trade union pluralism	100%
	Organisation of activities	100%
	Approval and registration of workers' organisations	100%
	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Election of representatives / Eligibility criteria	81%
	Protection of property / Financial independence	31%
Switzerland 2001	Organisation of activities	100%
	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Approval and registration of workers' organisations	56%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%
	Election of representatives / Eligibility criteria	22%
Norway 2002	Trade union pluralism	100%
	Approval and registration of workers' organisations	100%
	Right to establish and join workers' organisations	96%
	Election of representatives / Eligibility criteria	81%
	Protection of property / Financial independence	31%
Panama 2002	Minimum service	100%
	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Organisation of activities	89%
	Election of representatives / Eligibility criteria	81%
	Compulsory arbitration in the context of industrial action	81%
	Approval and registration of workers' organisations	56%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%
Guyana 2003	Compulsory arbitration in the context of industrial action	81%
	Restrictions on the right to industrial action	75%
	Right to establish and join workers' organisations	63%
	Trade union pluralism	33%

	Protection of property / Financial independence	31%
	Election of representatives / Eligibility criteria	22%
Portugal 2003	Restrictions on the right to industrial action	96%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%
	Election of representatives / Eligibility criteria	22%
Turkey 2004	Trade union pluralism	100%
	Organisation of activities	100%
	Approval and registration of workers' organisations	100%
	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Compulsory arbitration in the context of industrial action	81%
	Minimum service	60%
	Protection of property / Financial independence	31%
Cameroon 2000	Election of representatives / Eligibility criteria	22%
	Approval and registration of workers' organisations	100%
	Right to establish and join workers' organisations	96%
	Trade union pluralism	33%
Myanmar 2001	Protection of property / Financial independence	31%
	Right to establish and join workers' organisations	63%
	Trade union pluralism	33%
Saint Lucia 1997	Protection of property / Financial independence	70%
	Right to establish and join workers' organisations	63%
	Trade union pluralism	33%
Mauritania 1997	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Election of representatives / Eligibility criteria	81%
	Compulsory arbitration in the context of industrial action	81%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%
Djibouti 2003	Protection of property / Financial independence	31%
	Approval and registration of workers' organisations	100%
	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Trade union pluralism	33%
Antigua and Barbuda 2001	Protection of property / Financial independence	31%
	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Trade union pluralism	33%
Kuweit 1997	Protection of property / Financial independence	31%
	Right to establish and join workers' organisations	63%
	Trade union pluralism	33%
Ghana 1996	Trade union pluralism	100%
	Approval and registration of workers' organisations	100%
	Right to establish and join workers' organisations	63%

	Protection of property / Financial independence	33%
Hong-Kong 2004	Right to establish and join workers' organisations	63%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%
Japan 95	Organisation of activities	100%
	Right to establish and join workers' organisations	96%
	Restrictions on the right to industrial action	96%
	Approval and registration of workers' organisations	56%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%
Jamaica 95	Restrictions on the right to industrial action	96%
	Compulsory arbitration in the context of industrial action	81%
	Right to establish and join workers' organisations	63%
	Trade union pluralism	33%
	Protection of property / Financial independence	31%

TAB. Identified violations on Conv.87 using Decision trees

Relative neighborhood graph

Country	Violation	Prob.
Mali 2001	trade union pluralism	50,00%
	election of representatives / eligibility criteria	50,00%
	organisation of activities	50,00%
	minimum service	50,00%
	compulsory arbitration in the context of industrial action	50,00%
	right to life and physical integrity	50,00%
	right to liberty and security of person / right to a fair trial	50,00%
	right to establish and join workers' organisations	50,00%
	dissolution or suspension of workers' organisations	50,00%
	protection of property / financial independence	50,00%
Slovakia 2001	right to establish and join workers' organisations	100,00%
	trade union pluralism	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	protection of property / financial independence	100,00%
	approval and registration of workers' organisations	100,00%
	restrictions on the right to industrial action	100,00%
Switzerland 2001	right to establish and join workers' organisations	100,00%
	trade union pluralism	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	protection of property / financial independence	100,00%
Norway 2002	restrictions on the right to industrial action	100,00%
	trade union pluralism	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	right to establish and join workers' organisations	50,00%
	protection of property / financial independence	50,00%
	approval and registration of workers' organisations	50,00%
	restrictions on the right to industrial action	50,00%
	minimum service	50,00%
compulsory arbitration in the context of industrial action	50,00%	
Panama 2002	trade union pluralism	100,00%
	minimum service	100,00%
	compulsory arbitration in the context of industrial action	100,00%
	restrictions on the right to industrial action	100,00%
Guyana 2003	trade union pluralism	100,00%
	compulsory arbitration in the context of industrial action	100,00%
	election of representatives / eligibility criteria	50,00%
	organisation of activities	50,00%
	minimum service	50,00%
	right to establish and join workers' organisations	50,00%

	restrictions on the right to industrial action	50,00%
Portugal 2003	right to establish and join workers' organisations	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	protection of property / financial independence	100,00%
	approval and registration of workers' organisations	100,00%
	restrictions on the right to industrial action	100,00%
Turkey 2004	right to establish and join workers' organisations	100,00%
	trade union pluralism	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	protection of property / financial independence	100,00%
	approval and registration of workers' organisations	100,00%
	restrictions on the right to industrial action	100,00%
Cameroon 2000	right to establish and join workers' organisations	100,00%
	restrictions on the right to industrial action	33,33%
Myanmar 2001	right to establish and join workers' organisations	100,00%
	trade union pluralism	100,00%
	protection of property / financial independence	100,00%
	organisation of activities	50,00%
	restrictions on the right to industrial action	50,00%
Saint Lucia 1997	right to establish and join workers' organisations	100,00%
	protection of property / financial independence	75,00%
	restrictions on the right to industrial action	75,00%
	approval and registration of workers' organisations	50,00%
	trade union pluralism	25,00%
	organisation of activities	25,00%
	dissolution or suspension of workers' organisations	25,00%
	election of representatives / eligibility criteria	25,00%
Mauritania 1997	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	compulsory arbitration in the context of industrial action	100,00%
	trade union pluralism	75,00%
	right to establish and join workers' organisations	25,00%
	protection of property / financial independence	25,00%
	approval and registration of workers' organisations	25,00%
	restrictions on the right to industrial action	25,00%
Djibouti 2003	right to establish and join workers' organisations	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	compulsory arbitration in the context of industrial action	100,00%
	protection of property / financial independence	100,00%
	approval and registration of workers' organisations	100,00%
	restrictions on the right to industrial action	100,00%
Antigua and Barbuda 2001	right to life and physical integrity	100,00%

	right to liberty and security of person / right to a fair trial	100,00%
	right to establish and join workers' organisations	100,00%
	dissolution or suspension of workers' organisations	100,00%
	protection of property / financial independence	100,00%
Kuweit 1997	right to establish and join workers' organisations	75,00%
	trade union pluralism	75,00%
	protection of property / financial independence	50,00%
	minimum service	25,00%
	compulsory arbitration in the context of industrial action	25,00%
Ghana 1996	right to establish and join workers' organisations	100,00%
	protection of property / financial independence	100,00%
	restrictions on the right to industrial action	50,00%
	minimum service	50,00%
Hong-Kong 2004	right to establish and join workers' organisations	100,00%
	trade union pluralism	100,00%
	protection of property / financial independence	100,00%
Japan 95	right to establish and join workers' organisations	100,00%
	trade union pluralism	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	minimum service	100,00%
	compulsory arbitration in the context of industrial action	100,00%
	restrictions on the right to industrial action	100,00%
Jamaica 95	trade union pluralism	100,00%
	election of representatives / eligibility criteria	100,00%
	organisation of activities	100,00%
	minimum service	100,00%
	compulsory arbitration in the context of industrial action	100,00%

TAB. Identified violations on Conv.87 using RNG

- *Convention n°98*

Decision trees

Country	Violation	Prob.
Turkey 1997	Anti-union discrimination	93%
	Right to collective bargaining	89%
	Designation of the bargaining partner / Most representative trade union	100%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Sweden 1997	Designation of the bargaining partner / Most representative trade union	100%
	Negotiable issues-substantive outcomes / approval-refusal	50%
Haiti 1997	Anti-union discrimination	93%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Jamaica 1999	Designation of the bargaining partner / Most representative trade union	100%
	Negotiable issues-substantive outcomes / approval-refusal	50%
Sri Lanka 1999	Anti-union discrimination	93%
	Acts of interference	100%
	Designation of the bargaining partner / Most representative trade union	100%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Morocco 1999	Anti-union discrimination	93%
	Acts of interference	100%
	Right to collective bargaining	89%
	Designation of the bargaining partner / Most representative trade union	60%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Iceland 1998	Right to collective bargaining	89%
	Level and scope of collective bargaining	80%
	Negotiable issues-substantive outcomes / approval-refusal	50%
Chad 1998	Negotiable issues-substantive outcomes / approval-refusal	50%
Dominican Republic 1998	Right to collective bargaining	89%
	Negotiable issues-substantive outcomes / approval-refusal	50%
Venezuela 1998	Anti-union discrimination	93%
	Designation of the bargaining partner / Most representative trade union	100%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Czech Republic 2003	Anti-union discrimination	93%
	Acts of interference	100%
	Right to collective bargaining	89%
	Designation of the bargaining partner / Most representative trade union	60%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Belize 2002	Anti-union discrimination	93%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Trinidad and Tobago 2001	Right to collective bargaining	89%
	Designation of the bargaining partner / Most representative trade union	100%
	Level and scope of collective bargaining	80%
	Negotiable issues-substantive outcomes / approval-refusal	50%
Guinea 2001	Anti-union discrimination	93%

	Acts of interference	100%
	Designation of the bargaining partner / Most representative trade union	60%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Malaysia 2000	Anti-union discrimination	93%
	Right to collective bargaining	89%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Iraq 2000	Anti-union discrimination	93%
	Right to collective bargaining	89%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Fiji 2000	Acts of interference	100%
	Designation of the bargaining partner / Most representative trade union	60%
	Negotiable issues-substantive outcomes / approval-refusal	50%
Brazil 2000	Designation of the bargaining partner / Most representative trade union	89%
	Negotiable issues-substantive outcomes / approval-refusal	50%
Ukraine 1999	Anti-union discrimination	93%
	Acts of interference	100%
	Designation of the bargaining partner / Most representative trade union	100%
	Negotiable issues-substantive outcomes / approval-refusal	24%
Honduras 1999	Anti-union discrimination	93%
	Acts of interference	100%
	Designation of the bargaining partner / Most representative trade union	60%
	Negotiable issues-substantive outcomes / approval-refusal	24%

TAB. Identified violations on Conv.98 using Decision trees

Relative neighborhood graph

Country	Violation	Prob.
Turkey 1997	anti-union discrimination	100%
	designation of the bargaining partner / most representative trade union	100%
	negotiable issues and substantive outcomes of collective bargaining	100%
Sweden 1997	designation of the bargaining partner / most representative trade union	100%
	anti-union discrimination	33%
	acts of interference	33%
	negotiable issues and substantive outcomes of collective bargaining	33%
Haiti 1997	anti-union discrimination	100%
	acts of interference	100%
	designation of the bargaining partner / most representative trade union	100%
Jamaica 1999	acts of interference	100%
	designation of the bargaining partner / most representative trade union	100%
Sri Lanka 1999	acts of interference	100%
	designation of the bargaining partner / most representative trade union	100%
	right to collective bargaining	50%
	negotiable issues and substantive outcomes of collective bargaining	50%
	anti-union discrimination	50%
Morocco 1999	negotiable issues and substantive outcomes of collective bargaining	100%
	anti-union discrimination	50%
	acts of interference	50%
	promotion of free and voluntary collective bargaining	50%
	right to collective bargaining	50%
	compulsory arbitration in the context of collective bargaining	50%
Iceland 1998	promotion of free and voluntary collective bargaining	100%
	right to collective bargaining	100%
	level and scope of collective bargaining	100%
	anti-union discrimination	50%
	acts of interference	50%
	compulsory arbitration in the context of collective bargaining	50%
Chad 1998	promotion of free and voluntary collective bargaining	100%
	right to collective bargaining	100%
	level and scope of collective bargaining	100%
Dominican Republic 1998	acts of interference	100%
	right to collective bargaining	100%
	designation of the bargaining partner / most representative trade union	100%
	negotiable issues and substantive outcomes of collective bargaining	100%
Venezuela 1998	designation of the bargaining partner / most representative trade union	100%
Czech Republic 2003	anti-union discrimination	100%
	acts of interference	100%
	right to collective bargaining	100%
	designation of the bargaining partner / most representative trade union	100%
Belize 2002	anti-union discrimination	100%

	acts of interference	50%
	designation of the bargaining partner / most representative trade union	50%
	right to collective bargaining	50%
Trinidad and Tobago 2001	designation of the bargaining partner / most representative trade union	100%
	negotiable issues and substantive outcomes of collective bargaining	100%
Guinea 2001	anti-union discrimination	100%
	acts of interference	100%
	designation of the bargaining partner / most representative trade union	50%
Malaysia 2000	negotiable issues and substantive outcomes of collective bargaining	100%
	promotion of free and voluntary collective bargaining	33%
	right to collective bargaining	33%
Iraq 2000	negotiable issues and substantive outcomes of collective bargaining	100%
Fiji 2000	acts of interference	100%
	designation of the bargaining partner / most representative trade union	100%
	negotiable issues and substantive outcomes of collective bargaining	100%
	right to collective bargaining	50%
	anti-union discrimination	50%
Brazil 2000	promotion of free and voluntary collective bargaining	100%
	right to collective bargaining	100%
	negotiable issues and substantive outcomes of collective bargaining	100%
Ukraine 1999	anti-union discrimination	100%
	acts of interference	50%
	right to collective bargaining	50%
	designation of the bargaining partner / most representative trade union	50%
	negotiable issues and substantive outcomes of collective bargaining	50%
Honduras 1999	anti-union discrimination	100%
	acts of interference	100%
	designation of the bargaining partner / most representative trade union	100%

TAB. Identified violations on Conv.98 using Decision trees

5. Predictions' improvement

The representation space (extracted concepts) seems to own good characteristics. The main way in order to improve the accuracy of the models is to increase the number of labelled texts. For that, we will give our classification of key-concepts founded by machine learning methods. After the experts correction, we will re-integrate each incorrect labelled text and re-design machine learning. The result of this operation should be an improvement of the prediction.

6. Conclusion and future work

The concepts seems to be able to determine the main part of violations contained in the texts. The key of a model improvement is an increase of the learning database, i.e, the number of labelled texts. For that, we need to know the real violations of texts included in the sample test. We will then re-integrate these texts in the model construction and, thanks to this operation, we hope improve the accuracy.