Transfer of Predictive Models for Classification of Statutory Texts in Multi-jurisdictional Settings

Jaromir Savelka
Kevin D. Ashley

Intelligent Systems Program
University of Pittsburgh

jas438@pitt.edu

ICAIL 2015, San Diego, CA
June 9, 2015

Acknowledgements: Elizabeth F. Bjerke, Hasan Guclu, Margaret A. Potter, Patricia M. Sweeney, Matthias Grabmair, Rebecca Hwa

We are grateful to the University of Pittsburgh’s University Research Council Multidisciplinary Small Grant Program for funding this project. This publication was also supported in part by the Cooperative Agreement 5P01TP000304 from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the CDC.
Presentation Overview

- Motivation
- Task Description
- Related and Prior Work
- Data from Multiple Jurisdictions
- Framework
- Experimental Setup
- Evaluation and Results
- Future Work
- Conclusions
Presentation Overview

Motivation

Task Description

Related and Prior Work

Data from Multiple Jurisdictions

Framework

Experimental Setup

Evaluation and Results

Future Work

Conclusions
Motivation
Description of Manual Task

1. Set of candidate statutory texts is retrieved on basis of predefined set of search queries from legal IR system.
2. Expert human annotators go through texts and identify relevant spans, i.e. parts containing relevant legal norms.
3. Each relevant span is represented as numeric code following guidelines provided in codebook (citation and 9 descriptors).

NOTE: 95% confidence interval for average inter-annotator agreement for all tasks was reported as (63.1%, 74.9%).

[28] PHASYS Codebook [online]
Coding Scheme Elements

- Citation
- Relevance
- Acting PHS agent (Who is acting?)
- Prescription
- Action (Which action is being taken?)
- Goal
- Purpose (For what purpose is action being taken?)
- Type of Emergency Disaster
- Receiving PHS agent
- Timeframe (In what timeframe can/must action be taken?)
- Condition

In our work we perform described tasks automatically, i.e.:
1. We transform textual data corresponding to relevant provisions into feature vectors.
2. We classify vectors in terms of each of nine code categories.

In prior work data sparsity was recognized as key element limiting performance.

We decided to focus on use of data from other jurisdictions as one possible way to mitigate problem of data sparsity.

Simple pooling of data tends to improve performance but with very small margin.

Currently, we have developed a framework for transfer of text classification models among different jurisdictions.
Presentation Overview

Motivation

Task Description

Related and Prior Work

Data from Multiple Jurisdictions

Framework

Experimental Setup

Evaluation and Results

Future Work

Conclusions
1. Classification of legal norms in terms of type. [3], [8], [10], [11], [13]
   We classify texts as containing, e.g., obligation (‘must’), permission (‘may’) or prohibition (‘must not’).

2. Classification of legal literature and legislative texts with hierarchically organized topics. [12], [18]
   Closely related to classification of the texts in terms of relevance.

3. Rule-based techniques for extraction of specific elements. [3], [10], [11], [13], [24], [25]
   We mine texts for presence of similar elements.

4. Classification of EU documents with terms from EuroVoc. [4], [7], [20], [21]
   Close to mining texts for specific topical and functional information.

Presentation Overview

Motivation
Task Description
Related and Prior Work
Data from Multiple Jurisdictions
Framework
Experimental Setup
Evaluation and Results
Future Work
Conclusions
## Comparison of Similar MD and FL Provisions

<table>
<thead>
<tr>
<th>COMAR 01.01.2003.18(D)(2)</th>
<th>Fla. Stat. § 943.0312(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CODE OF MARYLAND REGULATIONS</strong></td>
<td><strong>Florida Annotated Statutes</strong></td>
</tr>
<tr>
<td><strong>TITLE 01. EXECUTIVE DEPARTMENT</strong></td>
<td><strong>TITLE 47. CRIMINAL PROCEDURE AND CORRECTIONS</strong></td>
</tr>
<tr>
<td><strong>SUBTITLE 01. EXECUTIVE ORDERS</strong></td>
<td><strong>CHAPTER 943. DEPARTMENT OF LAW ENFORCEMENT</strong></td>
</tr>
<tr>
<td>Establishement of the Governor’s Office Of Homeland Security</td>
<td>Regional domestic security task forces</td>
</tr>
<tr>
<td>The Director shall be responsible for the following activities:</td>
<td>The Chief of Domestic Security, in conjunction with the Division of Emergency Management,</td>
</tr>
<tr>
<td>Advise the Governor on policies, strategies, and measures to enhance and improve the ability to detect, prevent, prepare for, protect against, respond to, and recover from, man-made emergencies or disasters, including terrorist attacks;</td>
<td>the regional domestic security task forces, and the various state entities responsible for</td>
</tr>
<tr>
<td></td>
<td>establishing training standards applicable to state law enforcement officers and fire, emergency, and first-responder personnel shall identify appropriate equipment and training needs, curricula, and materials related to the effective response to suspected or actual acts of terrorism or incidents involving real or hoax weapons of mass destruction […]</td>
</tr>
</tbody>
</table>
## Data Sets Properties

The individual provisions are stored in XML files (one for each state).

These files are the starting point for all of our experiments.

There are 18,998 unique terms/lemmas (i.e., features) after stop-words removal.

<table>
<thead>
<tr>
<th>state</th>
<th># statutes</th>
<th># provisions</th>
<th># relevant</th>
<th># codes</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK</td>
<td>135</td>
<td>1965</td>
<td>331</td>
<td>386</td>
</tr>
<tr>
<td>CA</td>
<td>1174</td>
<td>19857</td>
<td>2296</td>
<td>2712</td>
</tr>
<tr>
<td>FL</td>
<td>464</td>
<td>16618</td>
<td>1033</td>
<td>1476</td>
</tr>
<tr>
<td>KS</td>
<td>304</td>
<td>5003</td>
<td>713</td>
<td>1190</td>
</tr>
<tr>
<td>MD</td>
<td>248</td>
<td>7593</td>
<td>687</td>
<td>760</td>
</tr>
<tr>
<td>ND</td>
<td>208</td>
<td>3114</td>
<td>458</td>
<td>656</td>
</tr>
<tr>
<td>PA</td>
<td>808</td>
<td>10882</td>
<td>1665</td>
<td>1873</td>
</tr>
<tr>
<td>TX</td>
<td>811</td>
<td>30474</td>
<td>1462</td>
<td>1712</td>
</tr>
</tbody>
</table>
Presentation Overview

Motivation

Task Description

Related and Prior Work

Data from Multiple Jurisdictions

Framework

Experimental Setup

Evaluation and Results

Future Work

Conclusions
Framework: Intuition

In traditional ML scheme, we use $D_{\text{train}}$ to train $f(\cdot)$ that performs as well as possible on $D_{\text{test}}$.

Here, in addition to $D_{\text{train}}$ we have number of $D_{\text{aux}}^{(i)}$ coming from different domain (different jurisdiction or different context).

Because $D_{\text{train}}$ and $D_{\text{aux}}^{(i)}$ come from different domains simple pooling of data does not improve performance reliably.

Presented framework uses $D_{\text{aux}}^{(i)}$ to train $f(\cdot)$ which performs better than predictive function trained on $D_{\text{train}}$ only.

Underlying idea is to train a number of different $f_i(\cdot)$ on different $D_i$ and decide about their usefulness in particular contexts.
Framework: Training

1. **Train** $f_i(\cdot)$ for each available data set (ensemble of experts).
2. **Evaluate** performance of each $f_i(\cdot)$ with respect to each label in label space.
3. Record the accuracy of each $f_i(\cdot)$ for each possible class in accuracy matrix $A$.

$$A = \begin{pmatrix}
  a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\
  a_{2,1} & a_{i,j} & \cdots & a_{2,n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{m,1} & a_{m,2} & \cdots & a_{m,n}
\end{pmatrix}$$

where $a_{i,j}$ stands for accuracy of $f_i(\cdot)$ on class $j$ with respect to $D_{\text{train}}$.

Output of training phase is set of $f_i(\cdot)$ and $A$. 
Framework: Prediction

To predict $y^{(i)}$ for unseen $x^{(i)}$ we

1. Use each available $f(\cdot)$ to provide probability distribution over label space and, thus, obtain prediction matrix $P$.

$$P(x^{(k)}) = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,n} \\ p_{2,1} & p_{i,j} & \cdots & p_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m,1} & p_{m,2} & \cdots & p_{m,n} \end{pmatrix}$$

2. Multiply ("correct") each value in $P$ with corresponding accuracy value from $A$ and obtain confidence matrix $C$.

$$C(x^{(k)}) = A \odot P(x^{(k)}) = \begin{pmatrix} a_{1,1} \times p_{1,1} & \cdots & a_{1,n} \times p_{1,n} \\ \vdots & \ddots & \vdots \\ a_{m,1} \times p_{m,1} & \cdots & a_{m,n} \times p_{m,n} \end{pmatrix}$$

3. Sum over columns of $C$ and predict any label with confidence passing set threshold.
Presentation Overview

Motivation

Task Description

Related and Prior Work

Data from Multiple Jurisdictions

Framework

Experimental Setup

Evaluation and Results

Future Work

Conclusions
Experimental Setup

We wish to evaluate performance of the system as we use increasing number of models trained on data from other states.

We generate one hundred $D^{(i)}_{train}$ (training sets) and one hundred corresponding $D^{(i)}_{test}$ (test sets).

For each task (e.g., acting agent or time frame) we monitor performance of the system as we increase number of models.

+0 (no model trained on data from other states is used)
+1 (one model trained on data from one other state is used)
+2, +3, +4, +5, +6, +7 (etc.)
Presentation Overview

Motivation
Task Description
Related and Prior Work
Data from Multiple Jurisdictions
Framework
Experimental Setup
Evaluation and Results
Future Work
Conclusions
Evaluation Metrics

**Precision**
Ratio of correctly retrieved instances over all instances that were retrieved.

\[ P(f(\cdot), D) = \sum_{i=1}^{n} \frac{|f(x^{(i)}) \cap y^{(i)}|}{|f(x^{(i)})|} \]

**Recall**
Ratio of correctly retrieved instances over all instances that should have been retrieved.

\[ R(f(\cdot), D) = \sum_{i=1}^{n} \frac{|f(x^{(i)}) \cap y^{(i)}|}{|y^{(i)}|} \]

**F\textsubscript{1} Measure**
Harmonic mean of precision and recall where both measures are treated as equally important.

\[ F_1(P(f(\cdot), D), R(f(\cdot), D)) = \frac{2 \cdot P(\cdot) \cdot R(\cdot)}{P(\cdot) + R(\cdot)} \]
Results ($F_1$-measure)

**Acting agent (AA)**
- Prescription (PR)
- Action (AC)
- Goal (GL)
- Purpose (PP)
- Emergency type (ET)
- Receiving agent (RA)
- Condition (CN)
- Timeframe (TF)

**Florida**
Comparison of Pooled Data vs. Transfer Framework

Acting agent

Purpose
COMAR 01.01.2003.18(D)(2)

**Establishment of the Governor’s Office Of Homeland Security**

The Director shall be responsible for the following activities:

Advise the Governor on policies, strategies, and measures section enhance and improve the ability to detect, prevent, prepare for, protect against, respond to, and recover from, man-made emergencies or disasters, including terrorist attacks;

---

### Improvement Example

<table>
<thead>
<tr>
<th>task</th>
<th>Manual</th>
<th>+0</th>
<th>+1</th>
<th>+2...+5</th>
<th>+6</th>
<th>+7</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>1</td>
<td>.56</td>
<td>.59</td>
<td>.78</td>
<td>.83</td>
<td>.83</td>
</tr>
<tr>
<td>R</td>
<td>1</td>
<td>.5</td>
<td>.78</td>
<td>.89</td>
<td>.89</td>
<td>.89</td>
</tr>
<tr>
<td>F₁</td>
<td>1</td>
<td>.53</td>
<td>.67</td>
<td>.83</td>
<td>.86</td>
<td>.86</td>
</tr>
</tbody>
</table>

---

**Establishment of the Governor’s Office Of Homeland Security**

The Director shall be responsible for the following activities:

Advise the Governor on policies, strategies, and measures section enhance and improve the ability to detect, prevent, prepare for, protect against, respond to, and recover from, man-made emergencies or disasters, including terrorist attacks;

---

### Establishment of the Governor’s Office Of Homeland Security

The Director shall be responsible for the following activities:

Advise the Governor on policies, strategies, and measures section enhance and improve the ability to detect, prevent, prepare for, protect against, respond to, and recover from, man-made emergencies or disasters, including terrorist attacks;
Presentation Overview

Motivation
Task Description
Related and Prior Work
Data from Multiple Jurisdictions
Framework
Experimental Setup
Evaluation and Results
Future Work
Conclusions
Future Work

▶ Implement similar framework for the relevance task.
▶ Experiment with techniques to handle imbalanced and sparse data sets, e.g. SMOTE. [6] Chawla et al. 2002
▶ Generate richer text representation (automatic annotation).
▶ Experiment with other transfer learning techniques. [19] Pan & Yang 2010
▶ Utilize existing knowledge:
  ▶ codebook [28] Codebook [online]
  ▶ tables of corresponding agents from different states
  ▶ data generated by network analysis
Presentation Overview

Motivation
Task Description
Related and Prior Work
Data from Multiple Jurisdictions
Framework
Experimental Setup
Evaluation and Results
Future Work
Conclusions
Conclusions

▶ We have presented framework for transfer of text categorization models.
▶ Performance of most classifiers gradually improves as we use increasing number of models.
▶ Relatedness of domains as well as tasks we deal with was confirmed.
▶ Possible way to deal with data sparsity was further explored and confirmed as promising.
▶ The framework’s benefits are not limited to context of United States (e.g., multiple EU jurisdictions similarly-purposed laws).


References II


Thank you!

Questions, comments and suggestions are welcome now or any time at jas438@pitt.edu.