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

ISP AI Forum
January 23, 2015
Presentation Overview

Motivation

Task Description

Related and Prior Work

Data from Multiple Jurisdictions

Data Processing

Framework

Experimental Setup

Evaluation and Results

Future Work

Conclusions
Presentation Overview

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

Motivation

Task Description

Related and Prior Work

Data from Multiple Jurisdictions

Data Processing

Framework

Experimental Setup

Evaluation and Results

Future Work

Conclusions
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).\[^{28}\]

NOTE: 95% confidence interval for average inter-annotator agreement for all tasks was reported as (63.1%, 74.9%).
The number of patients admitted to any area of the hospital shall not exceed the number for which the area is designed, equipped, and staffed except in cases of emergency, and then only in accordance with the emergency or disaster plan of the hospital. (28 Pa. Code para 101.172)

28 Pa. Code § 101.172; Hospital (14); Must Do (2); Suspend (29); Rule/Regulations/Restrictions (4); For Emergency Response (2); Non-specified Disaster/Emergency (5); Public/Individuals (27); Silent (0); Silent (0)
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

Task Description (Automated)

- In our work we perform described tasks automatically, i.e.:
  1. We transform textual data into feature vectors.
  2. We classify vectors in terms of relevance for PHS analysis.
  3. We classify vectors in terms of each of nine code categories.
  4. We evaluate performance of our system with respect to labels created by expert annotators (treated as gold standard).

- 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.

- 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
Data Processing
Framework
Experimental Setup
Evaluation and Results
Future Work
Conclusions
Related Work (AI & Law)

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.

Transfer learning, in contrast to traditional ML framework, allows the domains, tasks, and distributions used in training and testing to be different.

Transfer learning aims to extract the knowledge from one or more source tasks and applies the knowledge to a target task.

Prior Work (Results)

<table>
<thead>
<tr>
<th></th>
<th>PA→PA</th>
<th>FL→PA</th>
<th>FL+PA→PA</th>
<th>FL→FL</th>
<th>PA→FL</th>
<th>FL+PA→FL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relevance</td>
<td>F: 0.72</td>
<td>F: 0.54</td>
<td>F: 0.73</td>
<td>F: 0.52</td>
<td>F: 0.35</td>
<td>F: 0.54</td>
</tr>
<tr>
<td></td>
<td>P: 0.75</td>
<td>P: 0.62</td>
<td>P: 0.77</td>
<td>P: 0.62</td>
<td>P: 0.27</td>
<td>P: 0.55</td>
</tr>
<tr>
<td></td>
<td>R: 0.70</td>
<td>R: 0.47</td>
<td>R: 0.69</td>
<td>R: 0.45</td>
<td>R: 0.50</td>
<td>R: 0.52</td>
</tr>
<tr>
<td>Act. agent</td>
<td>A: 0.49</td>
<td>A: 0.30</td>
<td>A: 0.52</td>
<td>A: 0.36</td>
<td>A: 0.25</td>
<td>A: 0.44</td>
</tr>
<tr>
<td></td>
<td>A: 0.76</td>
<td>A: 0.72</td>
<td>A: 0.77</td>
<td>A: 0.77</td>
<td>A: 0.75</td>
<td>A: 0.75</td>
</tr>
<tr>
<td>Prescription</td>
<td>A: 0.29</td>
<td>A: 0.23</td>
<td>A: 0.30</td>
<td>A: 0.23</td>
<td>A: 0.18</td>
<td>A: 0.24</td>
</tr>
<tr>
<td>Action</td>
<td>A: 0.32</td>
<td>A: 0.17</td>
<td>A: 0.32</td>
<td>A: 0.20</td>
<td>A: 0.16</td>
<td>A: 0.25</td>
</tr>
<tr>
<td>Goal</td>
<td>A: 0.59</td>
<td>A: 0.53</td>
<td>A: 0.61</td>
<td>A: 0.58</td>
<td>A: 0.61</td>
<td>A: 0.62</td>
</tr>
<tr>
<td>Purpose</td>
<td>A: 0.78</td>
<td>A: 0.69</td>
<td>A: 0.80</td>
<td>A: 0.76</td>
<td>A: 0.72</td>
<td>A: 0.77</td>
</tr>
<tr>
<td>Emg. Type</td>
<td>A: 0.36</td>
<td>A: 0.25</td>
<td>A: 0.35</td>
<td>A: 0.25</td>
<td>A: 0.25</td>
<td>A: 0.28</td>
</tr>
<tr>
<td>Rec. agent</td>
<td>A: 0.84</td>
<td>A: 0.81</td>
<td>A: 0.85</td>
<td>A: 0.80</td>
<td>A: 0.78</td>
<td>A: 0.80</td>
</tr>
<tr>
<td>Time frame</td>
<td>A: 0.77</td>
<td>A: 0.68</td>
<td>A: 0.75</td>
<td>A: 0.65</td>
<td>A: 0.65</td>
<td>A: 0.67</td>
</tr>
</tbody>
</table>

Prior Work (Similar Traits in Both Jurisdictions)

Intra-jurisdictional classifiers trained for Florida (yellow) and Pennsylvania (blue) show that they both share similar traits.
Presentation Overview

Motivation
Task Description
Related and Prior Work
**Data from Multiple Jurisdictions**
Data Processing
Framework
Experimental Setup
Evaluation and Results
Future Work
Conclusions
## Comparison of Similar PA 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>&lt;br&gt;Title 01. Executive Department&lt;br&gt;Subtitle 01. Executive Orders**&lt;br&gt;Establishment of the Governor's Office of Homeland Security&lt;br&gt;The Director shall be responsible for the following activities:&lt;br&gt;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 <em>terrorist</em> attacks;</td>
<td>Florida Annotated Statutes&lt;br&gt;Title 47. Criminal Procedure and Corrections&lt;br&gt;Chapter 943. Department of Law Enforcement&lt;br&gt;Regional Domestic Security Task Forces&lt;br&gt;The Chief of Domestic Security, in conjunction with the Division of Emergency Management, the regional domestic security task forces, and the various state entities responsible for 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 <em>terrorism</em> or incidents involving real or hoax weapons of mass destruction [...]</td>
</tr>
</tbody>
</table>


Presentation Overview

Motivation
Task Description
Related and Prior Work
Data from Multiple Jurisdictions
Data Processing
Framework
Experimental Setup
Evaluation and Results
Future Work
Conclusions
§ 328.307. Licensing of eligible organizations to conduct games of chance

(a) LICENSE REQUIRED. — No eligible organization shall conduct or operate any games of chance unless such eligible organization has obtained and maintains a valid license or limited occasion license issued pursuant to this section. An auxiliary group of a licensed eligible organization shall be eligible to conduct games of chance using the license issued to the eligible organization provided that the auxiliary group or groups are listed on the application and license of the eligible organization. An auxiliary group is not eligible to obtain a license or a limited occasion license. No additional licensing fee shall be charged for an auxiliary group’s eligibility under this chapter. Auxiliary groups shall not include branches, lodges or chapters of a Statewide organization.

(b) ISSUANCE AND FEES. — The licensing authority shall license, upon application, within 30 days any eligible organization meeting the requirements for licensure contained in this chapter to conduct and operate games of chance at such locations within the county or in such manner as stated on the application as limited by subsection (b.1). The license fee to be charged to each eligible organization shall be $100, except for limited occasion licenses which shall be $10. Licenses shall be renewable annually upon the anniversary of the date of issue. The license fee shall be used by the licensing authority to administer this act.

(B.1) LOCATION OF GAMES OF CHANCE. —

(1) Except as otherwise provided in this section, a licensed eligible organization, except a limited occasion licensee, may conduct small games of chance at a licensed premises. The licensed premises shall be...
Statutory documents are (in comparison to other types of documents) well structured.

Document can be viewed as a tree graph with given spans of text as nodes and sub-part relations as edges.

We need to divide each statutory text into smaller parts that could be referred via citations.

(Root) ..............................................
  (1) ..............................................
    (a) ...........................................
    (b) ...........................................
    (c) ...........................................
  (2) ..............................................
§101.62. Request for absentee ballots

(1) (a) The supervisor shall accept a request for an absentee ballot from an elector in person or in writing. One request shall be deemed sufficient to receive an absentee ballot for all elections through the next regularly scheduled general election, unless the elector or the elector’s designee indicates at the time the request is made the elections for which the elector desires to receive an absentee ballot. Such request may be considered canceled when any first-class mail sent by the supervisor to the elector is returned as undeliverable.

(b) The supervisor may accept a written or telephonic request for an absentee ballot from the elector, or, if directly instructed by the elector, a member of the elector’s immediate family, or the elector’s legal guardian. For purposes of this section, the term ”immediate family” has the same meaning as specified in paragraph (4)(b).

The person making the request must disclose:
1. The name of the elector for whom the ballot is requested.
2. The elector’s address.
3. The elector’s date of birth.
4. The requester’s name.
5. The requester’s address.
6. The requester’s driver’s license number, if available.
7. The requester’s relationship to the elector.
Request for absentee ballots

(1) (a) The supervisor shall accept a request for an absentee ballot from an elector in person or in writing. One request shall be deemed sufficient to receive an absentee ballot for all elections through the next regularly scheduled general election, unless the elector or the elector’s designee indicates at the time the request is made the elections for which the elector desires to receive an absentee ballot. Such request may be considered canceled when any first-class mail sent by the supervisor to the elector is returned as undeliverable.

(b) The supervisor may accept a written or telephonic request for an absentee ballot from the elector, or, if directly instructed by the elector, a member of the elector’s immediate family, or the elector’s legal guardian. For purposes of this section, the term “immediate family” has the same meaning as specified in paragraph (4)(b).

The person making the request must disclose:
1. The name of the elector for whom the ballot is requested.
2. The elector’s address.
3. The elector’s date of birth.
4. The requester’s name.
5. The requester’s address.
6. The requester’s driver’s license number, if available.
7. The requester’s relationship to the elector.
ELECTORS AND ELECTIONS
VOTING METHODS AND
PROCEDURE

Request for absentee ballots
(a) The supervisor shall accept a request for an absentee ballot from an elector in person or in writing. One request shall be deemed sufficient to receive an absentee ballot for all elections through the next regularly scheduled general election, unless the elector or the elector’s designee indicates at the time the request is made the elections for which the elector desires to receive an absentee ballot. Such request may be considered canceled when any first-class mail sent by the supervisor to the elector is returned as undeliverable.

(b) The supervisor may accept a written or telephonic request for an absentee ballot from the elector, or, if directly instructed by the elector, a member of the elector’s immediate family, or the elector’s legal guardian. For purposes of this section, the term "immediate family" has the same meaning as specified in paragraph (4)(b).

The person making the request must disclose:
1. The name of the elector for whom the ballot is requested.
2. The elector’s address.
3. The elector’s date of birth.
4. The requester’s name.
5. The requester’s address.
6. The requester’s driver’s license number, if available.
7. The requester’s relationship to the elector.
The supervisor may accept a written or telephonic request for an absentee ballot from the elector, or, if directly instructed by the elector, a member of the elector’s immediate family, or the elector’s legal guardian. For purposes of this section, the term "immediate family" has the same meaning as specified in paragraph (4)(b).

The person making the request must disclose:
1. The name of the elector for whom the ballot is requested.
2. The elector’s address.
3. The elector’s date of birth.
4. The requester’s name.
5. The requester’s address.
6. The requester’s driver’s license number, if available.
7. The requester’s relationship to the elector.
Partitioning into Subtrees

ELECTORS AND ELECTIONS
VOTING METHODS AND PROCEDURE

Request for absentee ballots

The supervisor may accept a written or telephonic request for an absentee ballot from the elector, or, if directly instructed by the elector, a member of the elector’s immediate family, or the elector’s legal guardian. For purposes of this section, the term "immediate family" has the same meaning as specified in paragraph (4)(b). The person making the request must disclose:

1. The name of the elector for whom the ballot is requested.
2. The elector’s address.
3. The elector’s date of birth.
4. The requester’s name.
5. The requester’s address.
6. The requester’s driver’s license number, if available.
7. The requester’s relationship to the elector.
8. The requester’s address.

The requester’s address.
The individual text units 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.
Acting agent  Emergency type  Prescription
Framework: Data Sets

At minimum, framework assumes existence of labeled dataset
\[ D_{\text{train}} = \langle X_{\text{train}}, Y_{\text{train}} \rangle \in D_{\text{target}} \]

In addition, there may be an arbitrary number of labeled datasets
\[ D_{\text{aux}} = \langle X_{\text{aux}}, Y_{\text{aux}} \rangle \in D_{\text{aux}} \sim D_{\text{target}} \]

Goal is to train \( f(\cdot) \) which performs well on unseen \( x^{(i)}_{\text{test}} \in D_{\text{target}} \).

Framework uses \( D_{\text{aux}} \) 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: Predictive Models

Framework does not rely on a specific model of $f(\cdot)$. For different datasets different models or combination of models may be used.

Instead of actual prediction for $x_{test}^{(i)}$ probability distribution over label space is used.

Therefore, $f(\cdot)$ should be capable of providing probability distribution (or at least some score for each possible $y_j$).

$$ f(x_{test}^{(i)}) \rightarrow \langle p(y_1), p(y_2), \ldots, p(y_m) \rangle $$
Framework: Training

We train a predictive function $f_{train}(\cdot)$ on $D_{train}$.

In addition, we train $f_{aux}^{(i)}(\cdot)$ for each available $D_{aux}^{(i)}$.

Next we generate accuracy matrix:

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

where

$$a_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \left[ f^{(i)}(x^{(k)}) = j \right]$$
First, we generate a prediction matrix:

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

We can perform element-wise multiplication of \( A \) and \( P(x^{(k)}) \) to obtain confidence matrix for \( x^{(k)} \):

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

Each \( a_{i,j} \times p_{i,j} \) can be understood as our confidence that \( x^{(k)} \) should be labeled with class \( j \) emulated by \( f_i(\cdot) \).
Presentation Overview

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

We generate following data sets:

\[
D^{(i)}_{\text{train}} = \langle X^{(i)}_{\text{train}}, Y^{(i)}_{\text{train}} \rangle \quad (100 \text{ times})
\]

\[
D^{(i)}_{\text{test}} = \langle X^{(i)}_{\text{test}}, Y^{(i)}_{\text{test}} \rangle \quad (100 \text{ times})
\]

\[
D^{(i)}_{\text{aux}} = \langle X^{(i)}_{\text{aux}}, Y^{(i)}_{\text{aux}} \rangle \quad (\# \text{ of auxiliary states})
\]

For each task we conduct 8 related experiments:

(AK, MD, TX, KS, CA, ND, PA)
(KS, PA, AK, ND, CA, TX, MD)
(PA, CA, ND, MD, AK, TX, KS)

In related experiments there are 100 runs for first and eighth experiments and 300 runs for other experiments.

Experiments show how performance changes as we use more \( D^{(i)}_{\text{aux}} \).
We create vectorized data sets $X^{n \times m}$ with rows as documents and columns as terms by setting each entry of matrix to:

$$weight(t, d, D) = tf(t, d) \times \log(idf(t, D))$$

$t$: term  
$d$: document  
$D$: document collection  
$tf(t, d)$: number of occurrences of $t$ in $d$  
$idf(t, D)$: number of $d \in D$ over number of $d \in D$ containing $t$

Each $x^{(i)} \in X^{n \times m}$ is vector with $m$ dimensions, where $m$ is number of unique terms that occur in document collection.

Each $x^{(i)} \in X^{n \times m}$ is referenced with unique citation connecting vector to text unit from which it originates.
Presentation Overview

Motivation

Task Description

Related and Prior Work

Data from Multiple Jurisdictions

Data Processing

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.

\[ F_1(P(f(\cdot), D), R(f(\cdot), D)) = \frac{2 \times P(\cdot) \times R(\cdot)}{P(\cdot) + R(\cdot)} \]
Results (F₁-measure)

<table>
<thead>
<tr>
<th>task</th>
<th>0aux</th>
<th>1aux</th>
<th>2aux</th>
<th>3aux</th>
<th>4aux</th>
<th>5aux</th>
<th>6aux</th>
<th>7aux</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.43</td>
<td>0.45</td>
<td>0.45</td>
<td>0.46</td>
<td>0.47</td>
<td>0.48</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>0.78</td>
<td>0.80</td>
<td>0.81</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>AC</td>
<td>0.21</td>
<td>0.22</td>
<td>0.23</td>
<td>0.24</td>
<td>0.25</td>
<td>0.26</td>
<td>0.26</td>
<td></td>
</tr>
<tr>
<td>GL</td>
<td>0.25</td>
<td>0.27</td>
<td>0.28</td>
<td>0.28</td>
<td>0.29</td>
<td>0.30</td>
<td>0.30</td>
<td></td>
</tr>
<tr>
<td>PP</td>
<td>0.67</td>
<td>0.70</td>
<td>0.71</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td>0.72</td>
<td></td>
</tr>
<tr>
<td>ET</td>
<td>0.78</td>
<td>0.79</td>
<td>0.79</td>
<td>0.80</td>
<td>0.80</td>
<td>0.79</td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td>0.30</td>
<td>0.30</td>
<td>0.31</td>
<td>0.31</td>
<td>0.32</td>
<td>0.33</td>
<td>0.33</td>
<td></td>
</tr>
<tr>
<td>CN</td>
<td>0.62</td>
<td>0.66</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>TF</td>
<td>0.80</td>
<td>0.81</td>
<td>0.82</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>task</th>
<th>0aux</th>
<th>1aux</th>
<th>2aux</th>
<th>3aux</th>
<th>4aux</th>
<th>5aux</th>
<th>6aux</th>
<th>7aux</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>0.42</td>
<td>0.44</td>
<td>0.45</td>
<td>0.47</td>
<td>0.48</td>
<td>0.50</td>
<td>0.50</td>
<td>0.51</td>
</tr>
<tr>
<td>PR</td>
<td>0.86</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.89</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>AC</td>
<td>0.24</td>
<td>0.25</td>
<td>0.26</td>
<td>0.26</td>
<td>0.26</td>
<td>0.27</td>
<td>0.27</td>
<td>0.28</td>
</tr>
<tr>
<td>GL</td>
<td>0.27</td>
<td>0.29</td>
<td>0.30</td>
<td>0.31</td>
<td>0.32</td>
<td>0.32</td>
<td>0.33</td>
<td>0.33</td>
</tr>
<tr>
<td>PP</td>
<td>0.74</td>
<td>0.77</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
<td>0.78</td>
</tr>
<tr>
<td>ET</td>
<td>0.73</td>
<td>0.76</td>
<td>0.76</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.77</td>
<td>0.78</td>
</tr>
<tr>
<td>RA</td>
<td>0.30</td>
<td>0.30</td>
<td>0.31</td>
<td>0.31</td>
<td>0.31</td>
<td>0.32</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>CN</td>
<td>0.58</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
<td>0.64</td>
<td>0.63</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>TF</td>
<td>0.81</td>
<td>0.83</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
</tr>
</tbody>
</table>
## Comparison to Prior Work

<table>
<thead>
<tr>
<th>task</th>
<th>+0</th>
<th>+1</th>
<th>+2</th>
<th>+3</th>
<th>+4</th>
<th>+5</th>
<th>+6</th>
<th>+7</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.43</td>
<td>.43</td>
<td>.43</td>
<td>.43</td>
</tr>
<tr>
<td>R</td>
<td>.45</td>
<td>.44</td>
<td>.44</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
<td>.45</td>
</tr>
<tr>
<td>F</td>
<td>.43</td>
<td>.43</td>
<td>.44</td>
<td>.44</td>
<td>.44</td>
<td>.44</td>
<td>.44</td>
<td>.44</td>
</tr>
<tr>
<td>PP</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
<td>.66</td>
</tr>
<tr>
<td>R</td>
<td>.70</td>
<td>.70</td>
<td>.70</td>
<td>.70</td>
<td>.70</td>
<td>.70</td>
<td>.70</td>
<td>.70</td>
</tr>
<tr>
<td>F</td>
<td>.67</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
<td>.68</td>
</tr>
<tr>
<td>ET</td>
<td>.78</td>
<td>.79</td>
<td>.79</td>
<td>.80</td>
<td>.80</td>
<td>.80</td>
<td>.80</td>
<td>.80</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>task</th>
<th>+0</th>
<th>+1</th>
<th>+2</th>
<th>+3</th>
<th>+4</th>
<th>+5</th>
<th>+6</th>
<th>+7</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.41</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
</tr>
<tr>
<td>R</td>
<td>.45</td>
<td>.48</td>
<td>.50</td>
<td>.52</td>
<td>.54</td>
<td>.55</td>
<td>.56</td>
<td>.57</td>
</tr>
<tr>
<td>F</td>
<td>.43</td>
<td>.45</td>
<td>.45</td>
<td>.46</td>
<td>.47</td>
<td>.47</td>
<td>.48</td>
<td>.48</td>
</tr>
<tr>
<td>PP</td>
<td>.66</td>
<td>.65</td>
<td>.64</td>
<td>.65</td>
<td>.65</td>
<td>.64</td>
<td>.64</td>
<td>.64</td>
</tr>
<tr>
<td>R</td>
<td>.70</td>
<td>.75</td>
<td>.78</td>
<td>.80</td>
<td>.81</td>
<td>.82</td>
<td>.83</td>
<td>.84</td>
</tr>
<tr>
<td>F</td>
<td>.67</td>
<td>.70</td>
<td>.71</td>
<td>.71</td>
<td>.72</td>
<td>.72</td>
<td>.72</td>
<td>.72</td>
</tr>
<tr>
<td>ET</td>
<td>.78</td>
<td>.76</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>task</th>
<th>+0</th>
<th>+1</th>
<th>+2</th>
<th>+3</th>
<th>+4</th>
<th>+5</th>
<th>+6</th>
<th>+7</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
<td>.42</td>
</tr>
<tr>
<td>R</td>
<td>.45</td>
<td>.48</td>
<td>.50</td>
<td>.52</td>
<td>.54</td>
<td>.55</td>
<td>.56</td>
<td>.57</td>
</tr>
<tr>
<td>F</td>
<td>.43</td>
<td>.45</td>
<td>.45</td>
<td>.46</td>
<td>.47</td>
<td>.47</td>
<td>.48</td>
<td>.48</td>
</tr>
<tr>
<td>PP</td>
<td>.66</td>
<td>.65</td>
<td>.64</td>
<td>.65</td>
<td>.65</td>
<td>.64</td>
<td>.64</td>
<td>.64</td>
</tr>
<tr>
<td>R</td>
<td>.70</td>
<td>.75</td>
<td>.78</td>
<td>.80</td>
<td>.81</td>
<td>.82</td>
<td>.83</td>
<td>.84</td>
</tr>
<tr>
<td>F</td>
<td>.67</td>
<td>.70</td>
<td>.71</td>
<td>.71</td>
<td>.72</td>
<td>.72</td>
<td>.72</td>
<td>.72</td>
</tr>
<tr>
<td>ET</td>
<td>.78</td>
<td>.76</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
<td>.75</td>
</tr>
</tbody>
</table>
## Improvement Example

<table>
<thead>
<tr>
<th>task</th>
<th>Manual</th>
<th>0aux</th>
<th>1aux</th>
<th>2-5aux</th>
<th>6aux</th>
<th>7aux</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>26</td>
<td>20</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>PR</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>AC</td>
<td>21</td>
<td>43</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>GL</td>
<td>1</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>PP</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>ET</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>RA</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>CN</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>29</td>
<td>29</td>
<td>30</td>
<td>30</td>
<td>40</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>TF</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| P    | 1      | .56  | .59  | .78    | .83  | .83  |
| R    | 1      | .5   | .78  | .89    | .89  | .89  |
| F<sub>1</sub> | 1 | .53 | .67 | .83 | **.86** | **.86** |
Presentation Overview

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

- We have presented framework for transfer of text categorization models among different US state jurisdictions.
- Performance of most classifiers gradually improve as we use models from increasing number of states.
- 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 potential benefits are not limited to context of United States.
References I


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