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

Jaromír Šavelka
Intelligent Systems Program
Learning Research and Development Center
jas438@pitt.edu

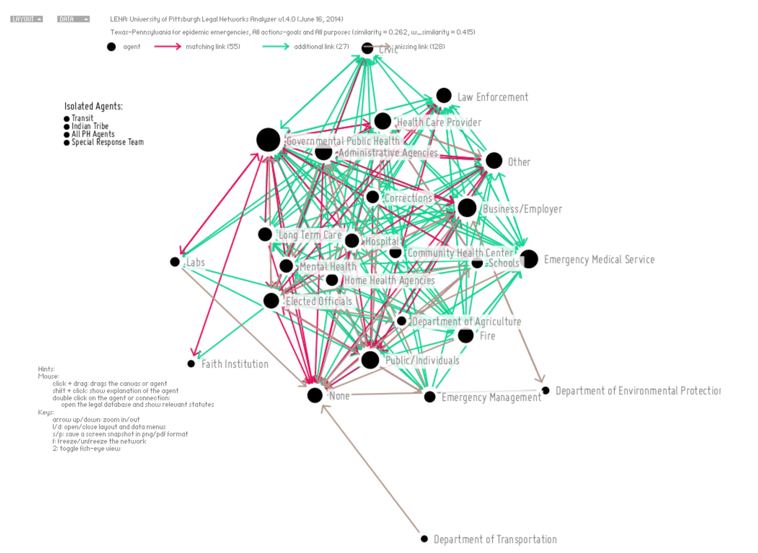
Kevin D. Ashley
Intelligent Systems Program
Learning Research and Development Center
School of Law

PHASYS PUBLIC HEALTH ADAPTIVE SYSTEMS STUDIES

PHASYS is committed to developing practical and useable instruments aimed at improving **public health preparedness and response** practice and policy making through the incorporation of science-based knowledge.

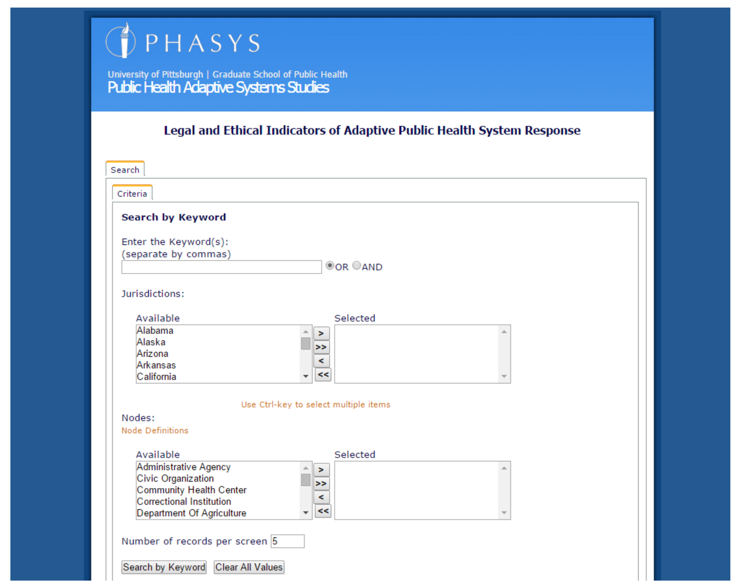
LENA

To improve emergency preparedness and response capacity in state and local public health systems, the **Legal Network Analyzer** (LENA) was created. LENA allows policy-makers to visualize legally directed relationships between PHS agents.



PHS Legal Database

The coded sections of state statutes and regulations, as well as all coded sections of the National Response Framework Nuclear Incident Annex, were entered into a **publicly accessible database** that is searchable by state and/or PHS agent.



Coding Scheme

- 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

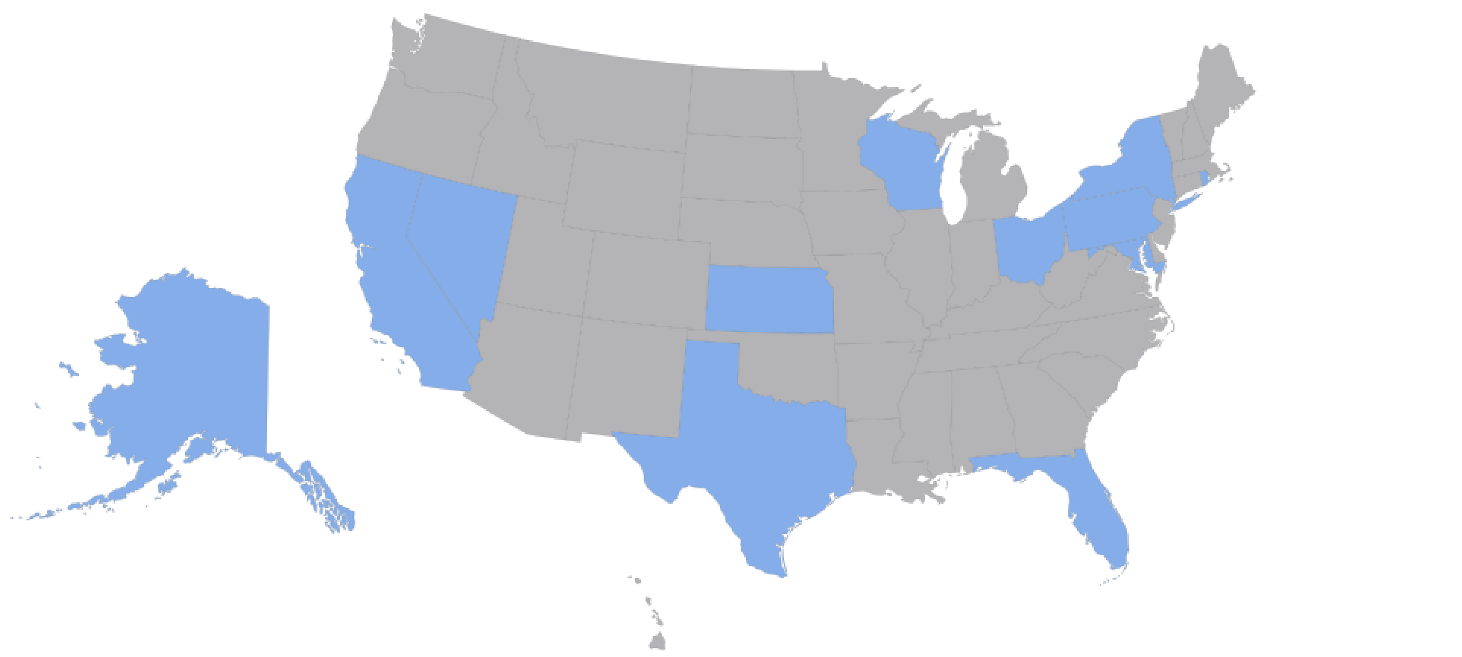
Example Code Assignment

Example Statutory Provision

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.

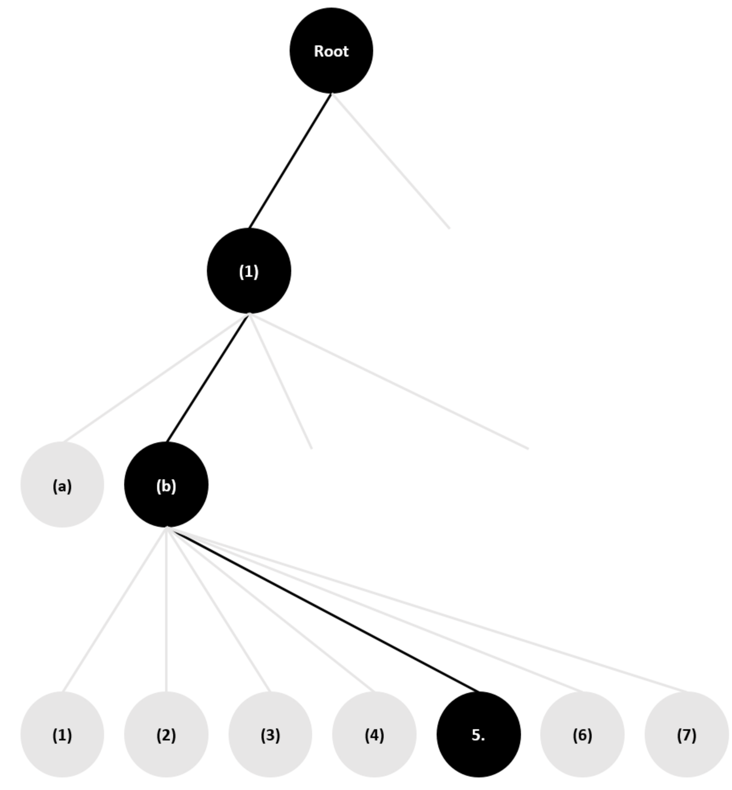
Corresponding Code

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)



Data Processing

Fla. Stat. §101.62
Florida Annotated Statutes
TITLE 9. ELECTORS AND ELECTIONS (Chs. 97-107)
CHAPTER 101. VOTING METHODS AND PROCEDURE
Fla. Stat. §101.62 (2010)
§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.



state	# statutes	# text u.	# relevant	# codes	state	# statutes	# text u.	# relevant	# codes
AK	135	1965	331	386	MD	248	7593	687	760
CA	1174	19857	2296	2712	ND	208	3114	458	656
FL	464	16618	1033	1476	PA	808	10882	1665	1873
KS	304	5003	713	1190	TX	811	30474	1462	1712

Framework for Transfer of Predictive Models

At minimum, framework assumes existence of labeled dataset:

$$\boldsymbol{D}_{train} = \langle \boldsymbol{X}_{train}, \boldsymbol{Y}_{train} \rangle \in \mathcal{D}_{target}$$

In addition, there may be an arbitrary number of labeled datasets:

$$\boldsymbol{D}_{aux} = \langle \boldsymbol{X}_{aux}, \boldsymbol{Y}_{aux} \rangle \in \mathcal{D}_{aux} \sim \mathcal{D}_{target}$$

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

Instead of actual prediction for $\boldsymbol{x}^{(i)}_{test}$ **probability distribution** over label space is used.

$$f(\boldsymbol{x}^{(i)}_{test}) \rightarrow \langle p(y_1), p(y_2), \dots, p(y_m) \rangle$$

We train a predictive function $f_{train}(\cdot)$ on $\boldsymbol{D}_{train}$.

In addition, we train $f^{(i)}_{aux}(\cdot)$ for each available $\boldsymbol{D}^{(i)}_{aux}$.

We generate **accuracy matrix**:

$$\boldsymbol{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{i,j} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix} \quad \text{where } a_{i,j} = \frac{1}{n} \sum_{k=1}^{n} \left[f^{(i)}(\boldsymbol{x}^{(k)}) = j \right]$$

Next, we generate a **prediction matrix**:

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

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

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

Each $a_{i,j} \times p_{i,j}$ represents a confidence that $\boldsymbol{x}^{(k)}$ should be labeled with class j emulated by $f_i(\cdot)$.

Experiments

We generate following data sets:

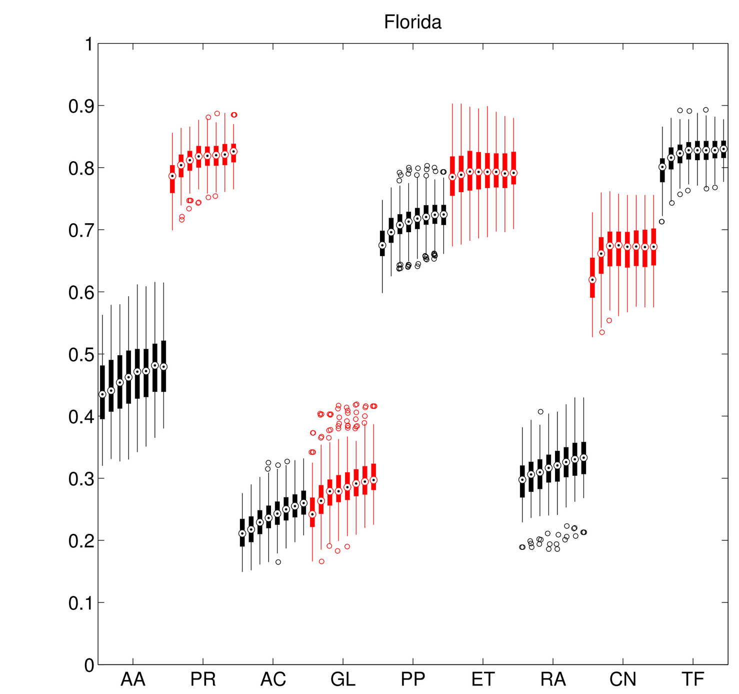
$$\begin{aligned} \boldsymbol{D}^{(i)}_{train} &= \langle \boldsymbol{X}^{(i)}_{train}, \boldsymbol{Y}^{(i)}_{train} \rangle \ \text{(100 times)} \\ \boldsymbol{D}^{(i)}_{test} &= \langle \boldsymbol{X}^{(i)}_{test}, \boldsymbol{Y}^{(i)}_{test} \rangle \ \text{(100 times)} \\ \boldsymbol{D}^{(i)}_{aux} &= \langle \boldsymbol{X}^{(i)}_{aux}, \boldsymbol{Y}^{(i)}_{aux} \rangle \ \text{(\# of aux states)} \end{aligned}$$

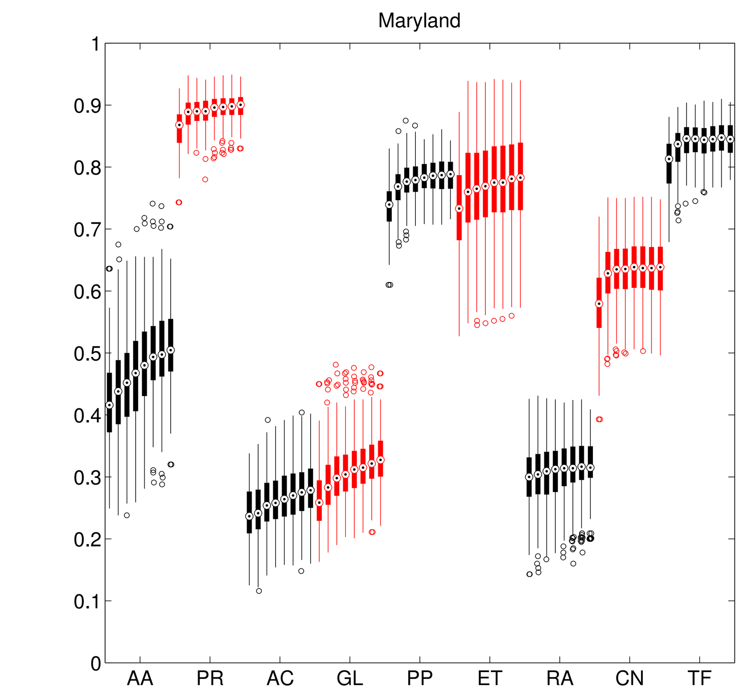
For each task we conduct **8 related experiments**:
(AK,MD,TX,KS,CA,ND,PA) (KS,PA,AK,ND,CA,TX,MD) (PA,CA,ND,MD,AK,TX,KS)

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

Results

	Florida								Maryland							
task	0aux	1aux	2aux	3aux	4aux	5aux	6aux	7aux	0aux	1aux	2aux	3aux	4aux	5aux	6aux	7aux
AA	**.43**	.45	.45	.46	.47	.47	**.48**	**.48**	**.42**	.44	.45	.47	.48	.50	.50	**.51**
PR	**.78**	.80	.81	**.82**	**.82**	**.82**	**.82**	**.82**	**.86**	.89	.89	.89	.89	**.90**	**.90**	**.90**
AC	**.21**	.22	.23	.24	.24	.25	**.26**	**.26**	**.24**	.25	.26	.26	.27	.27	**.28**	**.28**
GL	**.25**	.27	.28	.28	.29	.29	**.30**	**.30**	**.27**	.29	.30	.31	.32	.32	**.33**	**.33**
PP	**.67**	.70	.71	.71	**.72**	**.72**	**.72**	**.72**	**.74**	.77	.78	.78	.78	.78	**.79**	**.79**
ET	**.78**	.79	.79	.79	.80	.80	.79	**.80**	**.73**	.76	.76	.77	.77	.77	**.78**	**.78**
RA	**.30**	**.30**	.31	.31	.32	.32	**.33**	**.33**	**.30**	**.30**	.31	.31	.31	**.32**	**.32**	**.32**
CN	**.62**	.66	**.67**	**.67**	**.67**	**.67**	**.67**	**.67**	**.58**	.63	.63	.63	**.64**	.63	**.64**	**.64**
TF	**.80**	.81	.82	**.83**	**.83**	**.83**	**.83**	**.83**	**.81**	.83	.84	.84	.84	.84	**.85**	**.85**





Example

COMAR 01.01.2003.18(D)(2)

CODE OF MARYLAND REGULATIONS
TITLE 01. EXECUTIVE DEPARTMENT
SUBTITLE 01. EXECUTIVE ORDERS

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 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;

task	Manual	0aux	1aux	2aux	3aux	4aux	5aux	6aux	7aux
AA	26	20	26	26	26	26	26	26	26
		26							
PR	2	1	1	1	1	1	1	2	2
			2	2	2	2	2		
AC	21	43	21	21	21	21	21	21	21
		48	43	43	43	43	43	48	48
				48	48	48	48		
GL	1	50	50	50	50	50	50	50	50
PP	1	1	1	1	1	1	1	1	1
	2	2	2	2	2	2	2	2	2
	4	4	4	4	4	4	4	4	4
ET	5	5	5	5	5	5	5	5	5
	19		19	19	19	19	19	19	19
RA	20	20	20	20	20	20	20	20	20
CN	0	7	7	0	0	0	0	0	0
		29	29						
		30	30						
		40	40						
		41	41						
TF	0	0	0	0	0	0	0	0	0
P	1	.56	.59	.78	.78	.78	.78	.83	.83
R	1	.5	.78	.89	.89	.89	.89	.89	.89
F_1	1	.53	.67	.83	.83	.83	.83	**.86**	**.86**

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.