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

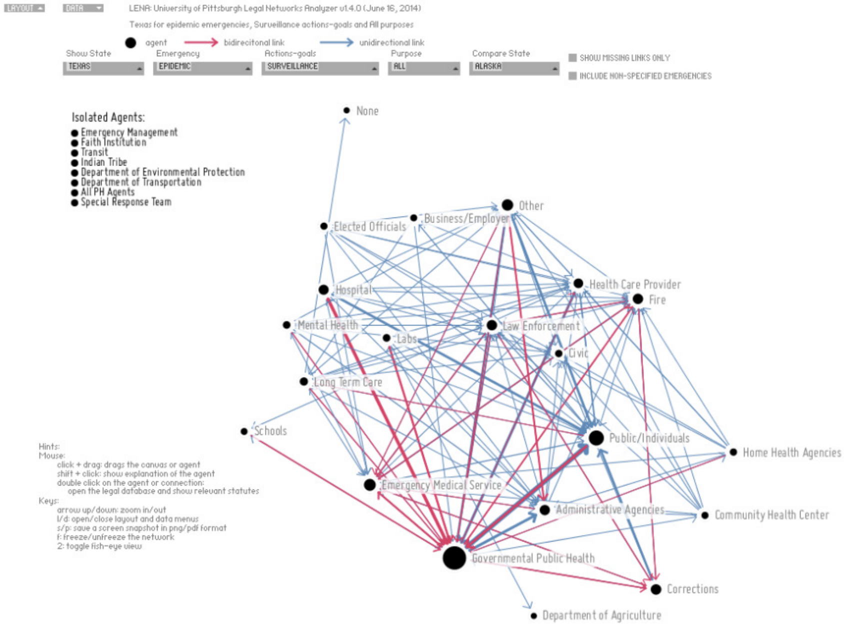
Acknowledgements: Elizabeth F. Bjerke, Hasan Guclu, Margaret A. Potter, Patricia M. Sweeney, Matthias Grabmair, Rebecca Hwa

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

Presentation Overview

Motivation

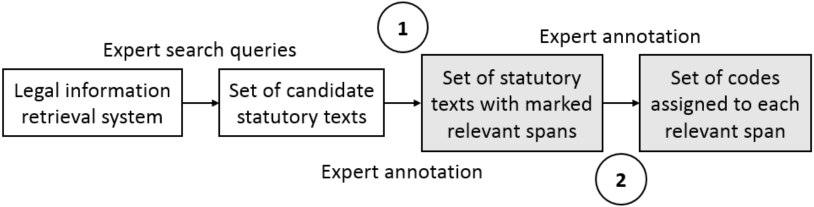


LAYOUT
DATA
LENA: University of Pittsburgh Legal Networks Analyzer v1.4.0 (June 16, 2014)
Texas for epidemic emergencies, Surveillance actions-goals and All purposes
agent
bidirecitonal link
unidirectional link
Show State
TEXAS
Emergency
EPIDEMIC
Actions-goals
SURVEILLANCE
Purpose
ALL
Compare State
ALASKA
SHOW MISSING LINKS ONLY
INCLUDE NON-SPECIFIED EMERGENCIES
Isolated Agents:
Emergency Management
Faith Institution
Transit
Indian Tribe
Department of Environmental Protection
Department of Transportation
All PH Agents
Special Response Team
None
Other
Business/Employer
Elected Officials
Health Care Provider
Fire
Hospital
Mental Health
Law Enforcement
Labs
Civic
Long Term Care
Schools
Public/Individuals
Home Health Agencies
Emergency Medical Service
Administrative Agencies
Community Health Center
Governmental Public Health
Corrections
Department of Agriculture
Hints:
Mouse:
click + drag: drags the canvas or agent
shift + click: show explanation of the agent
double click on the agent or connection:
open the legal database and show relevant statutes
Keys:
arrow up/down: zoom in/out
l/d: open/close layout and data menus
s/p: save a screen snapshot in png/pdf format
f: freeze/unfreeze the network
2: toggle fish-eye view

Presentation Overview

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).[28]



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

[15] Grabmair et al. 2011, [22] Sweeney et al. 2014

Description of Automated Task

- 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

Related Work

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.

[3] Biagioli et al. 2005, [4] Boella 2012, [7] Daudaravicius 2012, [8] de Maat & Winkels 2007, [10] Francesconi et al. 2010, [11] Francesconi 2009, [12] Francesconi & Peruginelli 2008, [13] Francesconi & Passerini 2007, [18] Opsomer et al. 2009, [20] Pouliquen 2003, [21] Steinberger 2012, [24] Winkels & Hoekstra 2012, [25] Wyner & Peters 2011

Presentation Overview

Comparison of Similar MD and FL Provisions

COMAR 01.01.2003.18(D)(2)	Fla. Stat. § 943.0312(3)
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;	Florida Annotated Statutes TITLE 47. CRIMINAL PROCEDURE AND CORRECTIONS CHAPTER 943. DEPARTMENT OF LAW ENFORCEMENT Regional domestic security task forces 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 terrorism or incidents involving real or hoax weapons of mass destruction [...]
Administrative agency [Active agent: 26] of the State [Active agent subset: 2] (homeland security) [Active agent footnote: 502] must [Prescription: 2] advise [Action: 21] the elected officials [Receiving agent: 20] on a plan [Goal: 1] for emergency preparedness, response, and recovery [Purpose: 1, 2 and 4] for an event of terrorist/bioterrorist/biohazardous emergency [Emergency type: 5, 19].	Law enforcement agency [Active agent: 16] of the State [Active agent subset: 2] must [Prescription: 2] advise [Action: 21] the elected officials [Receiving agent: 20] on a training program, equipment and personnel [Goal: 5, 7, 16] for emergency preparedness, response, and recovery [Purpose: 1, 2 and 4] for an event of terrorist/bioterrorist/biohazardous emergency [Emergency type: 5, 19].

Data Sets Properties

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

- 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.

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 $\boldsymbol{D}_{train}$ to train $f(\cdot)$ that performs as well as possible on $\boldsymbol{D}_{test}$.

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

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

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

Underlying idea is to train a number of different $f_i(\cdot)$ on different $\boldsymbol{D}_i$ 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 $\boldsymbol{A}$.

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

where $a_{i,j}$ stands for accuracy of $f_i(\cdot)$ on class j with respect to $\boldsymbol{D}_{train}$

Output of training phase is set of $f_i(\cdot)$ and $\boldsymbol{A}$.

Framework: Prediction

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

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

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

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

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

3. Sum over columns of $\boldsymbol{C}$ and predict any label with confidence passing set threshold.

Presentation Overview

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 $\boldsymbol{D}_{train}^{(i)}$ (training sets) and one hundred corresponding $\boldsymbol{D}_{test}^{(i)}$ (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

Evaluation Metrics

Precision

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

$$P(f(\cdot), \boldsymbol{D}) = \sum_{i=1}^{n} \frac{\left| f(\boldsymbol{x}^{(i)}) \cap \boldsymbol{y}^{(i)} \right|}{\left| f(\boldsymbol{x}^{(i)}) \right|}$$

Recall

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

$$R(f(\cdot), \boldsymbol{D}) = \sum_{i=1}^{n} \frac{\left| f(\boldsymbol{x}^{(i)}) \cap \boldsymbol{y}^{(i)} \right|}{\left| \boldsymbol{y}^{(i)} \right|}$$

F_1 Measure

Harmonic mean of precision and recall where both measures are treated as equally important.

$$F_1(P(f(\cdot), \boldsymbol{D}), R(f(\cdot), \boldsymbol{D})) = \frac{2 * P(\cdot) * R(\cdot)}{P(\cdot) + R(\cdot)}$$

Results (F_1-measure)

Acting agent (AA)

Prescription (PR)

Action (AC)

Goal (GL)

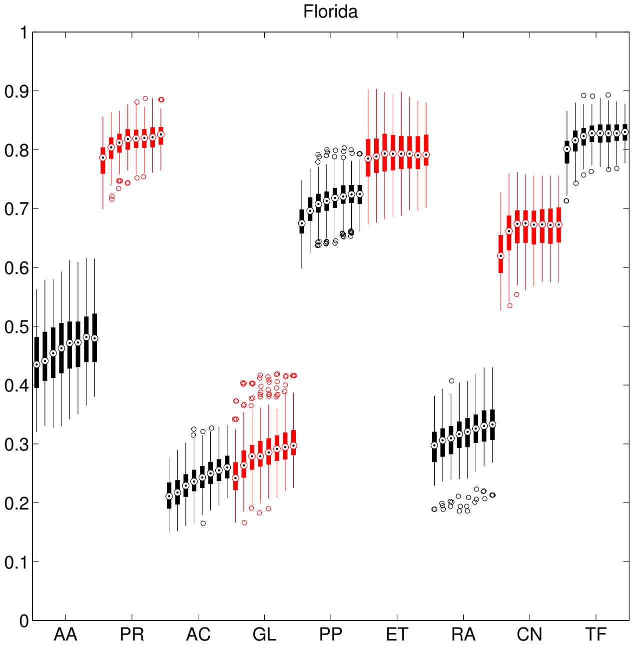
Purpose (PP)

Emergency type (ET)

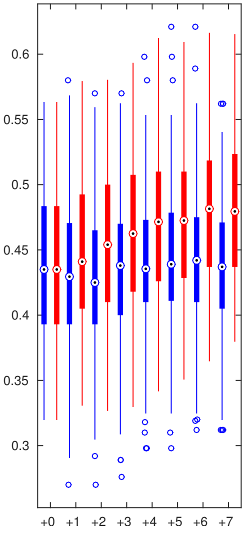
Receiving agent (RA)

Condition (CN)

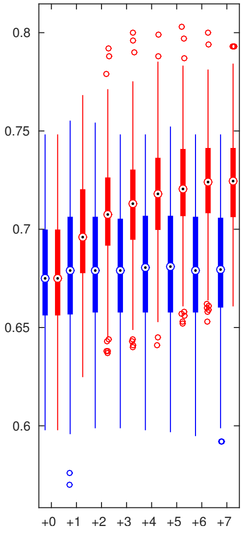
Timeframe (TF)



Comparison of Pooled Data vs. Transfer Framework



Acting agent



Purpose

Improvement Example

AA	20	Elected Officials
	26	Administrative Agencies
PR	1	May/Can
	2	Must Do
AC	21	Advise
	43	Ensure/Coordinate
	48	Amend
GL	1	Plan
	50	Collaboration
PP	1	For Emergency Preparedness
	2	For Emergency Response
	4	For Emergency Recovery
ET	5	Non-Specified Disaster/Emergency
	19	Terrorist/Bioterrorist/ Biohazardous Emergency
RA	20	Petroleum/Energy/ Utility Emergency
CN	0	Silent
	7	When there is an Emergency Resulting from an Enemy Attack
	29	When a Disaster, Emergency, or Riot Occurs or is Imminent
	30	When a Disaster is Declared in Another State
TF	0	Silent

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

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;

Presentation Overview

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
- Experiment with overlay framework for multi-dimensional classification.[2] Batal et al. 2013
- Generate richer text representation (automatic annotation).
- Experiment with learning tasks simultaneously (multi-task learning).[9] Evgeniou & Pontil 2004
- 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

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 I

Aggarwal, Charu & Zhai, ChengXiang (eds.). *Mining Text Data*. Springer, 2012.

Batal, I., Hong, C., and Hauskrecht, M., An Efficient Probabilistic Framework for Multi-Dimensional Classification. ACM Conference on Information and Knowledge Management. San Fransisco (2013).

Biagioli, C., Francesconi, E., Passerini, A., Montemagni, S., and Soria, C., Automatic Semantics Extraction in Law Documents, ICAIL 2005 Proceedings, 133–140, ACM Press (2005).

Boella, G., Di Caro, L., Lesmo, L., Rispoli, D., and Robaldo, L., Multi-label Classification of Legislative Text into EuroVoc. JURIX 2012 Proceedings, pp. 21–30, B. Schafer (Ed.), IOS Press (2012).

Breiman, L., J. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Boca Raton, FL: CRC Press, 1984.

Chawla, N.V., Bowyer, K.W., Hall, L.O., and Kegelmeyer, W.P., SMOTE: synthetic minority over-sampling technique. *J. Artif. Int. Res.* 16:321–357 (2002).

Daudaravicius, V., Automatic multilingual annotation of EU legislation with Eurovoc descriptors. EEOP2012 Workshop Proceedings (2012).

References II

- de Maat, E., Winkels, R., Categorisation of norms. JURIX 2007, pp.79–88, IOS Press (2007).
- Evgeniou, Theodoros & Pontil, Massimiliano. Regularized Multi–Task Learning. *KDD'04*. Seattle, WSH, USA, 2004.
- Francesconi, E., Montemagni, S., Peters, W., and Tiscornia, D., Integrating a Bottom-Up and Top-Down Methodology for Building Semantic Resources for the Multilingual Legal Domain. In *Semantic Processing of Legal Texts*. LNAI 6036, pp. 95–121. Springer: Berlin (2010).
- Francesconi, E., An Approach to Legal Rules Modelling and Automatic Learning. JURIX 2009 Proceedings (G. Governatori, Ed.), 59–68, IOS Press (2009).
- Francesconi, E., and Peruginelli, G., Integrated Access to Legal Literature through Automated Semantic Classification. *Artificial Intelligence and Law* 17:31–49 (2008).
- Francesconi, E., and Passerini, A., Automatic Classification of Provisions in Legislative Texts. *Artificial Intelligence and Law* 15:1–17 (2007).
- Jones, Karen Sparck. "A statistical interpretation of term specificity and its application in retrieval." *Journal of documentation* 28.1 (1972): 11-21.

References III

- Grabmair, M., Ashley, K.D., Hwa, R., and Sweeney, P.M., Toward Extracting Information from Public Health Statutes using Text Classification and Machine Learning. JURIX 2011 Proceedings, pp. 73-82 (Katie M. Atkinson ed.) IOS Press 2011.
- Kakwani, N., On a class of poverty measures. *Econometrica*, pp. 437-446 (1980).
- Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., and McClosky, D., The Stanford CoreNLP Toolkit. In *52nd Annual Meeting of the ACL: System Demonstrations*, pp. 55-60.
- Opsomer, R., De Meyer, G., Cornelis, C., van Eetvelde, G., Exploiting Properties of Legislative Texts to Improve Classification Accuracy. JURIX 2009 (G. Governatori, Ed.), 136–145, IOS Press (2009).
- Pan, Sinno Jialin & Yang, Qiang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.
- Pouliquen, B., Steinberger, R., and Ignat, C., Automatic annotation of multilingual text collections with a conceptual thesaurus. *arXiv preprint* cs/0609059 (2006).
- Steinberger, R., Ebrahim, M., and Turchi, M., JRC EuroVoc Indexer JEX-A freely available multi-label categorisation tool. *arXiv preprint* arXiv:1309.5223 (2013).

References IV

- Sweeney, P.M., Bjerke, E.F., Potter, M.A., Guclu, H., Keane, C.R., Ashley, K.D., Grabmair, M., Hwa, R., Network Analysis of Manually-Encoded State Laws and Prospects for Automation. In Winkels, R., Lettieri, N., Faro, S., (Eds.) *Network Analysis in Law*. Diritto Scienza Tecnologia (2014).
- Šavelka, J., Ashley, K.D., Grabmair, M., Mining Information from Statutory Texts in Multi-jurisdictional Settings. In Hoekstra, R. (Ed.) *JURIX 2014*. IOS Press (2014).
- Winkels, R., and Hoekstra, R., Automatic Extraction of Legal Concepts and Definitions. JURIX 2012, pp. 157–166, IOS Press (2012).
- Wyner, A., and Peters, W., On Rule Extraction from Regulations. JURIX 2011, pp. 113–122, IOS Press (2011).
- Zhang, M., and Zhou, Z., ML-KNN: A lazy learning approach to multi-label learning. *Pattern recognition* 40.7:2038-2048 (2007).
- *Lucene* [online]. 2012 [cit. 08/27/2014]. Accessed at: http://lucene.apache.org/core/
- *PHASYS ARM 2 - LEIP codebook* [online]. Revised 11/18/2012 [cit. 08/27/2014]. Accessed at: http://www.phasys.pitt.edu/

Thank you!

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

This work was supported by the University of Pittsburgh's University Research Council Multidisciplinary Small Grant Program. 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.