VoidWiz: Resolving Incompleteness Using Network Effects

Christina Christodoulakis* Christos Faloutsos** Renée J Miller*

*Department of Computer Science
University of Toronto

**School of Computer Science
Carnegie Mellon University

"What sorcery is this?"

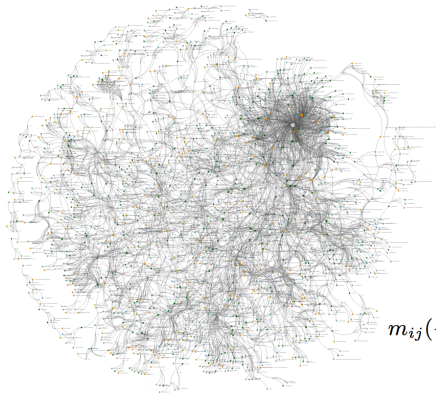
Incomplete data is one of the plagues of real data (categorical data, sensor networks, data integration, signal processing, image retrieval). Many approaches exist for recommending substitutes for missing values.
Non expert users need explanation of recommended values [1].

Patient Visit

#	Patient Name	Doctor Name	Diagnosis	Treatment
1	Lisa	Dr. Brown	Emphysema	A
2	Alice	Dr. Smith	Acne	B
3	Lisa	Dr. Brown	Emphysema	
4	Tom	Dr. Smith	Rash	C
5	Alice	Dr. Brown	Asthma	D

Value	Prob	Accept
A	0.874	
D	0.768	
B	0.629	
C	0.544	

Incompleteness Resolution



Intuition

Probable values will be found in tuples most similar to the incomplete tuple

Method

Treat data as a graph. Use Belief Propagation [2] as a fast belief diffusion mechanism, to find most similar tuples in the neighborhood.

$$m_{ij}(\{x_j\}) \leftarrow \sum_{x_i} \phi_i(x_i)\psi_{ij}(x_i,x_j) \prod_{k\in N(i)\setminus j} m_{ki}(x_i)$$

$$b_i(\{x_i\}) = k\phi(x_i) \prod_{j\in N(i)} m_{ji}(x_i)$$

Provenance, Insight, Trust

Guide the user with information for understanding contributing factors for recommended values.

Warn for possible constraint/business rule violations.

Display domain specific information on the data that might add to the trust of a recommended value:
- *combinations*
- *synonyms*
- *type of*

Provide interface for exploration, understanding, and verification of recommended values.

Contributions

- *Principled value imputation using network effects*
- *Visualization of provenance of recommended values*
- *Visualization and interaction with tabular and graph data*
- *Application on real world data (U.S. National Library of Medicine Clinical Trials)*
- *Insight and verification with domain knowledge (UMLS)*

[1] N. Tintarev and J. Masthoff. A survey of explanations in recommender systems. In Data Engineering Workshop, 2007 IEEE 23rd International Conference on, pages 801–810. IEEE, 2007.
[2] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," Exploring artificial intelligence in the new millennium, vol. 8, pp. 236–239, 2003.