Empirical Comparison of Approximate Inference Algorithms for Networked Data

Prithviraj Sen   Lise Getoor

Department of Computer Science
University of Maryland, College Park.

Workshop on Open Problems in Statistical Relational Learning, 2006

Introduction

- Recent, widespread interest in structured classification.
- Numerous approximate inference algorithms for networked data exist.
- We empirically compare three of the most popular ones:
  - Iterative Classification Algorithm
  - Mean Field Relaxation Labeling
  - Loopy Belief Propagation

Parameters of Interest

- Performance on random graph data.
- Effects of noise in attribute values.
- Effects of noise in correlations across links.
- Effects of varying link density.
- Effects of different link patterns.

Iterative Classification Algorithm (ICA)

- Simple, greedy, iterative algorithm.
- Introduced by Besag [Besag, 1986].
- In each iteration, for each node, looks at neighbourhood class labels.

$$b_i(y) \leftarrow \alpha\phi_i(y) \exp\{\sum_{Y_j \in \mathcal{N}(Y_i)} w_{y,y_j}\}$$

Mean Field Relaxation Labeling (MF)

- Soft-version of ICA. Many other versions exist.
- Discovered by vision community [Hummel & Zucker, 1983].
- In each iteration, for each node, looks at neighbour's label distribution.

$$b_i(y) \leftarrow \alpha\phi_i(y)\exp\{\sum_{Y_j \in \mathcal{N}(Y_i), y'} w_{y,y'} b_j(y')\}$$

Loopy Belief Propagation (LBP)

- Message-passing algorithm. Attempts to stop sending messages in loops.
- Discovered by iterative decoding community [Kschischang & Frey, 1998, McEliece et al, 1998, Kschischang et al, 2001].
- Messages computed without considering destination node's message.

$$m_{i\rightarrow j}(y) \leftarrow \alpha \sum_{y'} \phi_i(y') e^{w_{y,y'}} \prod_{Y_k \in \mathcal{N}(Y_i) \setminus Y_j} m_{k\rightarrow i}(y')$$

$$b_i(y) \leftarrow \alpha \phi_i(y) \prod_{Y_j \in \mathcal{N}(Y_i)} m_{j\rightarrow i}(y)$$

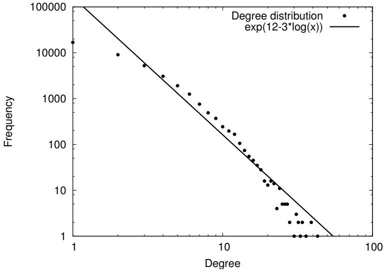
Synthetic Graph Generation Algorithm

Based on power-law graph generation algorithm [Bollobas et al, 2003]

```
Begin with a single node graph G.
repeat
    With probability α introduce an edge in G
    With probability 1 − α introduce a new node with a randomly sampled
    label, connect new node to G
until size of G = n
generate attributes for all nodes.
```

Preferential attachment scheme used

- When choosing node to link node ν to:
  - With probability ρ choose node with same label.
  - With probability $1-\rho$ choose node with different label.
  - Preference given to nodes with high degree.



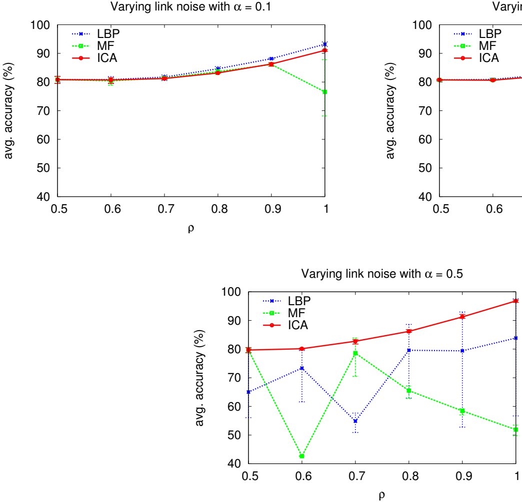
Experimental setup

- Performed 3-fold cross validation.
- Metric used: avg. classification accuracy.
- Compared three models: *ICA*, *MF*, *LBP*
- Performed experiments on binary class data.

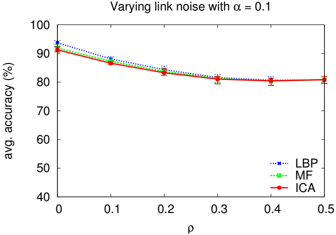
Parameters of interest

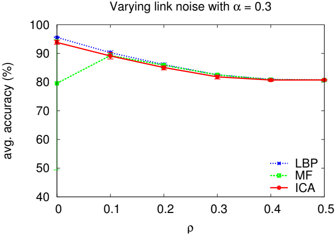
- α: controls number of edges
- ρ: controls degree of correlation across edges.
- ω: controls noise in attribute values.

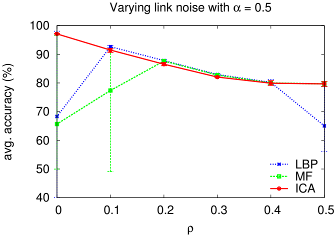
Varying Correlations across Links



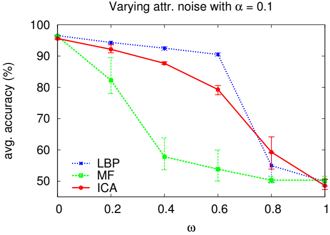
Varying Correlations across Links – contd.

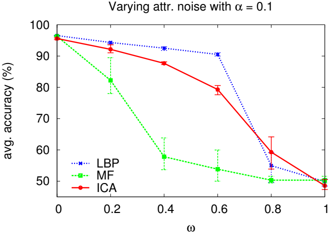


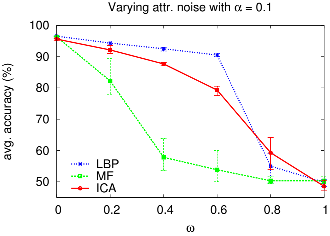




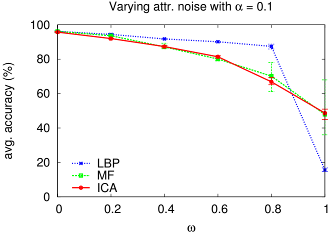
Varying Attribute Noise

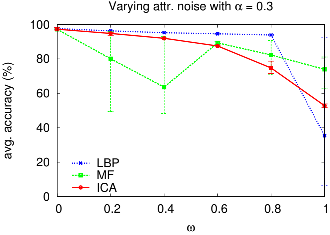


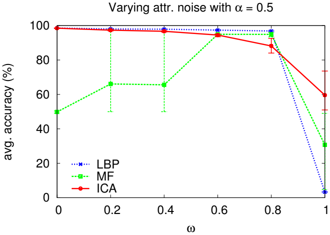




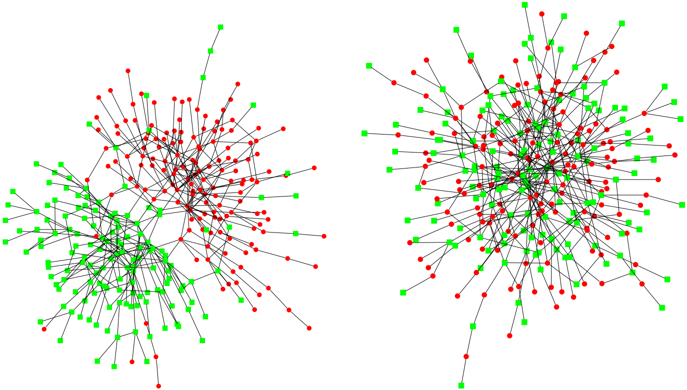
Varying Attribute Noise – contd.







Effect of different link patterns



- In the case of *Homophily* or *Perfect Assortative Mixing* (figure on left), the generated graphs form densely connected clusters introducing closed loops hampering LBP and MF.

Conclusion

- We empirically compared three of the most popular approximate inference techniques for networked data.
- MF tends to get stuck at local minima in a variety of cases, e.g., high link correlation, high link density.
- LBP tends to face issues in the presence of high link density and a specific type of link pattern known as *Homophily* or *Perfect Assortative Mixing* but otherwise performs well.
- We found that LBP's convergence does *not* necessarily indicate good results.
- ICA is the most consistent of the three approaches considered, returning reasonable results in a wide variety of conditions.

References

- J. Besag, *On the statistical analysis of dirty pictures*", Journal of the Royal Statistical Society, 1986.
- R. Hummel and S. Zucker, *On the foundations of relaxation labeling processes*, IEEE Trans. on Pattern Analysis and Machine Intelligence, 1983.
- F. R. Kschischang and B. J. Frey, *Iterative decoding of compound codes by probability progation in graphical models*, IEEE Journal on Selected Areas in Communication, 1998.
- F. R. Kschischang and B. J. Frey and H. A. Loeliger, *Factor graphs and the sum-product algorithm*, IEEE Trans. on Information Theory, 2001.
- R. J. McEliece and D. J. C. MacKay and J. F. Cheng, *Turbo decoding as an instance of Pearl's belief propagation algorithm*, IEEE Journal on Selected Areas in Communication, 1998.
- B. Bollobas, C. Borgs, J. T. Chayes and O.Riordan, *Directed scale-free graphs*, In Proceedings of ACM-SIAM Symposium on Discrete Algorithms, 2003.