Towards Context-Enriched Trust Prediction: A Proposal

Marcin Sydow

Web Mining Lab,
Polish-Japanese Institute of Information Technology,
Koszykowa 86, 02-008, Warszawa, Poland
msyd@pjwstk.edu.pl

16.06.08

Motivation

There is an explosion of interest in on-line communities where users express their personal preferences

- Trust management plays an important role in such systems
- The amount of information in social networks grows exponentially

A strong need for automatic techniques that support users in making their on-line decisions – in particular: trust prediction

Recommender Systems

An example of automatic trust-related support of user decision is a recommender system:

Given the preferences of the users *and* trust values among them automatically recommend some items to particular users.

Examples of such systems: `epinions`[1] or `FilmTrust`[2].

[1] `http://www.epninions.com`

[2] `http://trust.mindswap.org/FilmTrust/`

Trust Prediction

‘**Cold start**’ problem: new users to the system do not have enough information to express their trust to others

How to use the additional information available in the system to predict unknown trust between a pair of users?

Contextual Information

Many meanings of the term [9, 8]. The most general could be:

All the information available in the social network system except the trust information itself

User Similarity as the Context

In the initial phase of our research we narrow the definition of the contextual information to the recommendations (ratings) given by the users.

This is a limited interpretation, however such a data is currently available [1].

General Idea

Recently Ziegler et al. [11] measured:
significant linear correlation between trust and user similarity

Building on the above result: similarity-based trust prediction
(with more general models than linear correlation)

Combination of Trust and Context-based Attributes

Comparing the predictive power of the following three groups of attributes:

1. the attributes based solely on the pure topology of the trust network
2. the attributes based on the contextual information. (begin with the recommendation-based similarity)
3. the combined trust-based and contextual-based attributes

Hypothesis

Combining trust topology information with context, such as similarity between the users will improve the prediction accuracy

Proposed General Experimental Scheme

1. **Dataset**: real dataset representing a trust network enhanced with contextual information
2. **Sampling**: reducing computations
3. **Training-testing split**: a common technique in supervised learning to avoid the *over-fitting* problem
4. **Pure trust-topology-based prediction baseline**: predict the value of trust (or distrust) between the nodes, based only on the trust-network information
5. **Pure contextual-based prediction baseline**: Compute contextual features, based on the additional information available in the dataset
6. **Contextually-enriched predictor**: combine the trust-based and context-based features
7. **Evaluation**: compare the performance of the three prediction schemes

Details of Proposed Experiments

- **Dataset**: `epinions dataset` [1], containing trust network enriched with user ratings
- **Sampling**: take a random sample of user pairs (seems to be necessary, since there are potentially $O(n^2)$ pairs for a n-node trust graph)
- **Training-testing split**: split the sample into the *training* and *testing* subsets to avoid over-fitting, with some experimetally tuned training/testing set size ratio

Details of Proposed Experiments, cont.

- **Pure trust-topology prediction baseline**: apply graph-theoretic *and* special trust-propagation techniques [3, 10]
- **Pure contextual-based prediction baseline**: compute features based on user ratings, similar to those described in [11, 2] and predict the trust using only these features
- **Contextually-enriched predictor**: combine the trust-based and context-based features and finally train the combined trust predictor on the same sample as above.
- **Evaluation**: some standard prediction-performance measures, such as prediction accuracy, the F-measure, etc.

Related Work

- The problem of pure link-based classification and appropriate link-based features for *undirected* co-authorship network are discussed by Liben-Nowell et al. [6]
- Karamon et al. discusses similar issues [4], however for the *node classification* problem
- Correlation between trust and user profile similarity is discussed in [11] and further studied with survey-based experiments in [2]
- An algorithm based on Bayesian network for trust inference is discussed in [5]
- Matsuo et al. [7] concerns research similar to the research presented here, however, only a draft was available in the time of writing, which does not seem to be completed.

Related Work, cont.

Remarkable overlap of proposed ideas exists with Matsuo et al. [7]. However, the mentioned draft (in contrast to ours):

- proposes general topology-based characteristics for trust computation, which, for example, never propagate further than through two links. We propose special trust propagation techniques which and are not constrained in such a way
- unfortunately does not address the problem of the extreme rating-data sparsity. We take this issue into account
- does not seem to take into account the distrust concept

It would be very interesting to compare the final results of the mentioned paper, and the research envisaged here, when they are ready.

Elements of Novelty

1. full supervised machine-learning approach, instead of simple linear correlation measurement (as in [11]).
2. introduce features based on special trust-propagation models [3, 10]. Previously, only general link-topology characteristics [6, 4] were used in this context.
3. due to ratings data sparsity consider *clustering* the rated items (some bi-partiteness-aware method) or other dimensionality-reduction techniques such as SVD.
4. consider *distrust* prediction (if such data is available)
5. regard the true meaning of the rating values instead of sentiment-unaware approach (such as dot-product)
6. if review texts are available, consider their textual similarity to compute additional user-similarity features

Summary

We formulated a hypothesis that additional information available in social-network systems can improve the performance of trust-prediction.

We proposed a general experimental research to be done on real datasets to test the hypothesis

Finally, we described particular envisaged experiments on the `epinion` dataset and explained how the outcome would relate to the existing results in the area

Acknowledgements

The work is supported by the Polish Ministry of Science grant N N516 4307 33

Thank you for attention

http://www.trustlet.org.

Jennifer Golbeck.

Trust and nuanced profile similarity in online social networks.
Technical report.

R. Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins.

Propagation of trust and distrust.
In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 403–412, New York, NY, USA, 2004. ACM.

J. Karamon, Y. Matsuo, H. Yamamoto, and M. Ishizuka.

Generating social network features for link-based classification.
In *PKDD*, pages 127–139, 2007.

U. Kuter and J. Golbeck.

Sunny: A new algorithm for trust inference in social networks using probabilistic confidence models.
In *AAAI*, pages 1377–1382, 2007.

David Liben-Nowell and Jon Kleinberg.

The link prediction problem for social networks.
In *Proceedings of CIKM '03*, pages 556–559, NY, USA, 2003. ACM.

Yutaka Matsuo and Hikaru Yamamoto.

Measuring bidirectional effects on trust and rating on online social networks.
Unpublished.

R. Neisse, M. Wegdam, P. Dockhorn Costa, and M. J. van Sinderen.

Context-aware management domains.
In B. Hulsebosch, G. Lenzini, and M. Wegdam, editors, *Proceedings of Context Awareness and Trust 2007 (CAT07), First International Workshop on Combining Context with Trust, Security and Privacy, Moncton, Canada*, volume 269 of *CEUR Workshop Proceedings*, pages 42–47, Nederlands, July 2007.

Sini Ruohomaa and Lea Kutvonen.

Making multi-dimensional trust decisions on inter-enterprise collaborations.
In *Proceedings of the Third International Conference on Availability, Security and Reliability (ARES 2008)*, pages 873–880, Barcelona, Spain, March 2008. IEEE Computer Society.

Baoning Wu, Vinay Goel, and Brian D. Davison.

Propagating trust and distrust to demote web spam.
In *Workshop on Models of Trust for the Web*, Edinburgh, Scotland, May 2006.

Cai-Nicolas Ziegler and Jennifer Golbeck.

Investigating interactions of trust and interest similarity.
Decis. Support Syst., 43(2):460–475, 2007.