

User Role Prediction in Online Discussion Forums using Probabilistic Soft Logic

Extended Abstract

Arti Ramesh¹, Jaebong Yoo², Lise Getoor¹, Jihie Kim² ¹Department of
Computer Science, University of Maryland, College Park, USA
²Information Sciences Institute, University of Southern California 1artir,
getoorl@cs.umd.edu, 1jihie, jaebongl@isi.edu

Online discussions have become a popular tool for supporting student communication in college-level courses. Participating in online discussions is an important part of student activities in both distance education as well as classroom education [2]. We identify two major roles for a student participating in online discussions - information seeker and information provider. One of the key challenges for modeling student conversation lies in identifying the role played by the student in the discussions. Identifying the student's role can help in inferring interesting facts such as her understanding of the subject, her willingness to participate in discussions and credibility of her answers.

In this paper, we propose the use of Probabilistic Soft Logic (PSL) [1] as a framework for modeling the probabilistic relationships and similarities between entities in online forums, and use probabilistic reasoning to infer users' roles in online discussion threads. PSL, like many other statistical relational learning methods [3], uses first order logic rules annotated with weights to model the dependencies. However PSL uses soft truth values, relaxing boolean truth values to the interval $[0,1]$. Triangular norms, which are continuous relaxations of logical connectives AND and OR, are used to combine the atoms in the first order clauses. Soft truth values enable integration of similarity functions in the same interval. With the help of these similarity functions, PSL supports reasoning about similarity between entities, relations and also sets of entities which are related by the same relation. And as a result of the soft formulation and the triangular norms, inference in PSL is a convex optimization problem, which makes inference in PSL tractable in comparison to other formalisms.

Modeling users' roles in online discussions is challenging due to the following reasons: 1) online discussion threads often contain noisy data which makes it difficult to distinguish between information sought and information provided; 2) questions can be similar, not essentially same and they can be asked in different ways with different content words; and 3) the questions and answers in online discussion forums do not necessarily follow the syntactic rules of grammar and identifying them correctly is key to inferring the roles. PSL is well-suited towards modeling these dependencies. For example, finding and modeling the similarity between the questions/answers probabilistically can help infer users' roles in related questions. With the convenient means to represent similarities and connect them using weighted first order clauses, PSL is capable of representing the relations existing between various entities in the discussion threads and propagate similarity values across the network. We are in the process of finding and formalizing the dependencies that exist in this domain as first-order logic rules in PSL. Using the weighted first order clauses, we propose to infer users' roles in the discussion threads on an existing collection of online discussions.

References

- [1] Broecheler, M., Mihalkova, L. & Getoor, L., Probabilistic Soft Logic. In Conference on Uncertainty in Artificial Intelligence, 2010.
- [2] Kang, J. H. & Kim, J., Analyzing Answers in Threaded Discussions using a Role-Based Information

Network. In proceedings of the IEEE International Conference on Social Computing, 2011.

[3] Getoor, L. & Tasker, B., Introduction to Statistical Relational Learning. The MIT Press, 2007.