User Role Prediction in Online Discussion Forums using Probabilistic Soft Logic

Arti Ramesh¹, Jaebong Yoo², Shitian Shen², Lise Getoor¹ and Jihie Kim²
¹Department of Computer Science, University of Maryland, College Park
²Information Sciences Institute, University of Southern California

Motivation

Online discussions: popular tool for supporting student communication in distance and classroom education

Identifying **user role** is key to modeling online student discussion

Goal

Identify two major roles for users

- Information Seeker (sink)
- Information Provider (source)

Probabilistic Soft Logic

Declarative language for probabilistic models using first-order logic (FOL) syntax

Truth values are relaxed to soft truth in [0,1]

Mechanisms for incorporating similarity functions and reasoning about sets

Efficient inference: convex optimization in continuous space

Maximum Likelihood weight learning via voted perceptron

Continuous probability distribution over truth values:

\[ P_r(x) = \frac{1}{Z} \exp \left( - \sum_{r \in G(r)} \sum_{g \in G(r)} w_r(1 - t_g(x)) \right) \]

- P: the PSL program
- G(r): set of all groundings of rule r
- \( w_r(1 - t_g(x)) \): truth value of grounding g

Software: http://psl.umiacs.umd.edu

Online Discussion Data

- **Local Features**
  - features: question, answer, acknowledgement, task, issue, message position

- **Network Features**
  - source, sink

Online discussion data organized into **threads**: threads broken down into **messages**

- "reply" connects messages in a thread
- "user" connects messages posted by same user within and across threads

PSL Rules

**Messages** = \( M_1, M_2 \)

**Local Rules**

\[ \text{question}(M_1) \land \text{firstMessage}(M_1) \Rightarrow \text{sink}(M_1) \]
\[ \text{answer}(M_1) \land \text{lastMessage}(M_1) \Rightarrow \text{source}(M_1) \]
\[ \text{task}(M_1) \Rightarrow \text{source}(M_1) \]

**Network Rules**

\[ \text{sameThread}(M_1, M_2) \land \text{sameUser}(M_1, M_2) \land \text{sink}(M_1) \Rightarrow \text{sink}(M_2) \]
\[ \neg \text{sameThread}(M_1, M_2) \lor \neg \text{sameUser}(M_1, M_2) \land \text{sink}(M_1) \Rightarrow \text{sink}(M_2) \]
\[ \text{reply}(M_1, M_2) \land \text{sink}(M_2) \land \neg \text{sameUser}(M_1, M_2) \Rightarrow \text{source}(M_2) \]

PSL Online Discussion Models

- **Model 1**
  - Combines local and network rules
  - Maximum likelihood weight learning via voted perceptron
  - Weight values indicate importance of local and network features

- **Model 2**
  - SVM model using bigram/trigram word features in messages
  - PSL network rules act on SVM label outputs to improve predictions

Evaluation

Model 2 outperforms Model 1 and SVM

PSL network rules take advantage of predictions generated by SVM model, propagate beliefs across network better

Discussion/Future Work

- PSL network rules provide convenient and effective means to model relationships in online discussion domain
- Model can be extended to infer student's understanding of subject, participation, credibility of answers in online forums