

Multidimensional User Expertise in Open Forums

The case of Machine Learning and Software Engineering on Stack Overflow

Summary notes for each section:

Introduction:

- Stack overflow is a software engineering Q & A platform
- Experts provide the best answers
- Each question asked demands one particular expertise

Related Work:

- Models of expertise evaluation in four major groups
 - Methods using graphing (graph) theory
 - identify the network of user connection and predict the expertise
 - Methods using content generated by user
 - Methods using users' activity
 - Hybrid models that combine the above methods

Graph Models

- Yang & Manandar: probability model for proposing users
 - users answers (relevant)
 - Based on expertise and what answer is provided
- Raj et al.: compared user behavior in Q & A platform
 - used voting system and social network graph to predict user reputation.

- PO-20-22-02
- "defined"
- Zhang et al.: graphs of user activities predicting flow of information experts & non-experts
 - This method is almost as good as human rating when determining experts.

Content-based Models

- Adamic et al.: users with a higher depth knowledge increases the chances of their responses results in higher answer ratings
- Balog et al.: generative probabilistic model
 - evaluated expertise in document and considered the expertise of user
 - extracted topics using topic modeling method.
- Mimno et al.: used topic modeling to match expert users
 - Rexa database + Author-Persona-Topic, where each author can write under multiple personas
- Budalakoti et al.: Aim to find best person to respond to question based on historical content made by users
- Cheng et al.: used listwise learning & tagword topic model
 - used COUPE (competition-based user expertise extraction) to find most relevant answer.
- Graph based pattern has higher accuracy
- Jenders et al.: Random Forest classifier on TF-IDF
 - Similarities between and the respective response

Activity-based Models.

- Bosu et al: analyzed reputation value on Stack Overflow
 - tags with lower expertise density, answering questions quickly, being the first responder, being active during off peak hours, contributing to diverse areas: gives higher correlation with user reputation.
- Anderson et al: reputation of questioners, # of questions asked by users, # of page views, # of positive & negative votes
 - positive correlation between user reputation and # of tag contributions

Random Forest classifiers



Hybrid Models

- Dalip et al: used user, graph (interaction), review, content structure, length, style, readability, and relevance for ranking most relevant response pattern
- Liu et al: topic, user power, reputation, user interaction, and subject relevance to find experts
- Was successful in finding expert for proposed question.
- extended to competition-based TS & SVM based methods
- Campbell et al: used content of email combining graph based communication to identify expertise.
- Hu et al: convolutional Neural Network (CNN) + deep belief network (DBN) which trained with tf-idf with surface linguistic features
 - DBN is better than Bag of word (BoG) model using TF-IDF
 - trivial improvement, higher computational cost

Hilroy

Conclusion:

- Article discusses variety of methodologies and models to "formulate" best answer

2022-05-09