

Enhanced User Interest and Expertise Modeling for Expert Recommendation

Tongze He, Caili Guo, Yunfei Chu

Beijing University of Posts and Telecommunications

Introduction

With the increasing demand of knowledge sharing services, Community Question Answering (CQA) websites, such as Stack Exchange and Zhihu, have already obtained popularization use in reality. An important challenge in CQA systems is to effectively match questions with the potential good answerers. Under this situation, expert recommendation has become a promising technique to tackle the challenge.

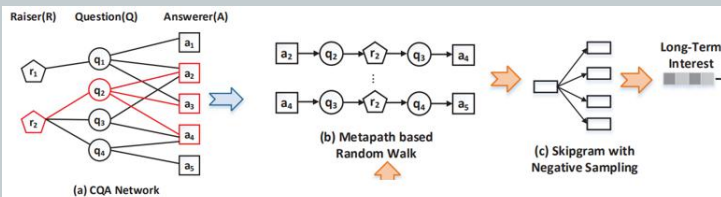
The existing methods learn one embedding for each user, which preserve both interest information and expertise information. However, the real-world CQA websites are dynamic, with users' interest changing over time. A user may not answer the question because his current interest is not in this question domain even if he has high expertise on it. Under the circumstances, user interest and user expertise are inconsistent, making it hard to capture both of them by one embedding. In addition, these methods design ranking loss according to the vote information of the answer to guide the learning of user expertise. However, users' historical answering behavior and feedback information are not fully explored in this process, which are important auxiliary information for modeling user expertise. In order to overcome the above limitations, we propose EUIEM to solve these problems.

Methods

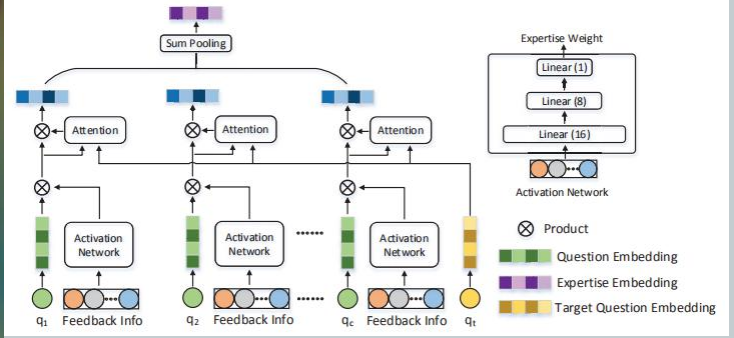
User interest can be divided into long-term and short-term interest. To capture long-term interest, we utilize metapath based random walk for its superiority in exploring graph structure and preserving proximities in heterogeneous network. We use metapath "AQRQA" to generate node sequence on CQA network. Then the neighbor prediction method with negative sampling is applied to learn embeddings.

Users may have dynamic short-term interests, which to a great extent determine whether they are willing to answer the given questions. Consequently, to capture the sequential patterns, we use LSTM due to its remarkable ability in processing sequential data. Specifically, we define the questions answered in the last n days as users' recent behavior.

In order to combine the long-term interest and short-term interest more naturally, we use the long-term interest to initialize the first hidden state of the LSTM, so that the last hidden state is a combination of short-term and long-term interest and we use it as users' interest embedding.



The feedback information is the key to measure users' expertise. Thus we design the user expertise network to model these information explicitly. In order to fully mine it, we construct several statistics and feed it into an activation network to calculate the expertise weight, which characterizes the users' expertise level of the question. We also use attention mechanisms to pay more attention to the answers which are related to the target question.



We introduce the relative rank to model the score. The ranking loss:

$$Obj_{rank} = \sum_{(i,j,q_t,r) \in D_1} S(p_i^a, e_i^a, q_t, p^r) - S(p_j^a, e_j^a, q_t, p^r) + \sum_{(y,z,q_t,r) \in D_2} S(p_y^a, e_y^a, q_t, p^r) - S(p_z^a, e_z^a, q_t, p^r)$$

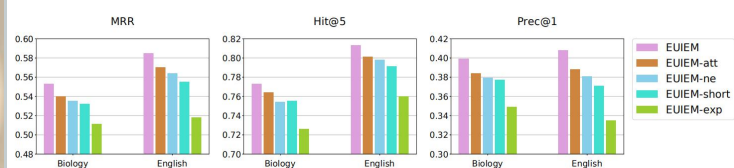
Experiment

We choose two sub-dataset of Stack Exchange Dataset: Biology and English.

Compared with other state-of-the-art methods, the proposed method performs much better. Compared to NeRank, the relative performance improvements achieved by EUIEM on the English are 5.7%, 3.8% and 6.5%, respectively. The reason is that compared with the baseline methods which only learn one single embedding for each user, our model considers the inconsistency between user interest and user expertise and learns the embedding for each of them. The addition of short-term interest helps predict users' willingness more accurately. In addition, the designed user expertise network explicitly exploits the feedback information to capture user expertise on the specific question domain. The attention mechanism helps learn more critical information, boosting the performance.

Method	T=2019-01-01						T=2018-10-01					
	Biology			English			Biology			English		
	MRR	Hit@5	Prec@1	MRR	Hit@5	Prec@1	MRR	Hit@5	Prec@1	MRR	Hit@5	Prec@1
Similarity	0.211	0.387	0.102	0.225	0.406	0.097	0.216	0.374	0.100	0.218	0.398	0.099
GBDT	0.329	0.491	0.203	0.372	0.612	0.188	0.301	0.483	0.185	0.362	0.600	0.192
ConvNCF	0.392	0.611	0.274	0.417	0.668	0.245	0.366	0.601	0.277	0.422	0.669	0.246
RMNL	0.483	0.697	0.315	0.491	0.717	0.336	0.488	0.691	0.317	0.492	0.708	0.331
NeRank	0.518	0.737	0.351	0.528	0.775	0.343	0.501	0.730	0.337	0.515	0.766	0.341
EUIEM	0.553	0.773	0.399	0.585	0.813	0.408	0.541	0.765	0.389	0.572	0.797	0.403

We also provide further investigations to better understand the contributions of model components to the proposed framework. Compared with its variants, the proposed method is better. The user expertise network can precisely capture user expertise on specific question domain with the help of attention network, and modeling users' short-term interest is useful to predict users'



Future Works

The profile of user such as user personal description should be considered to solve cold-start problem. The dynamic modeling of expertise can be applied since user expertise keeps improving over time.