Recommending Forum Posts to Designated Experts

Jason H.D. Cho†, Yanen Li‡, Roxana Girju†, Chengxiang Zhai†
†*University of Illinois at Urbana-Champaign*
‡*LinkedIn Inc.*
hcho33@illinois.edu, yali@linkedin.com, girju@illinois.edu, czhai@illinois.edu

***Abstract*—There are users who generate significant amounts of domain knowledge in online forums or community question and answer (CQA) websites. Existing literature defines them as 'experts.' These users attain such statuses by providing multiple relevant answers to the question askers. Past works have focused on recommending relevant posts to these users. With the rise of web forums where certified experts answer questions, strategies that are tailored towards addressing the new type of experts will be beneficial. In this paper, we identify a new type of user called 'designated experts' (i.e., users designated as domain experts by the web administrators). These are the experts who are guaranteed by web administrators to be an expert in a given domain. Our focus is on how we can capture the unique behavior of designated experts in an online domain.**

We have noticed designated experts have different behaviors compared to CQA experts. In particular, unlike existing CQAs, only one designated expert responds to any given thread. To capture this intuition, we introduce a matrix factorization algorithm with regularization to capture the behavior. Our results show that the regularization method improves the performance significantly compared to the baseline approach.

***Keywords*-Matrix factorization; Designated Experts; Recommender Systems; Online Forums**

I. Introduction

Recommender systems are developed to help many different types of users. Many of the existing systems focus on helping average users [21], [16] while others help experienced users [13], [19], [8]. Our work focuses on a specific type of user called *designated experts*. These users are defined as those who have significant domain knowledge and are formally designated by the forum administrators to aid average users. For instance, medical doctors are designated experts on medical help websites such as MedHelp[1] , while instructors and course staffs are designated experts on Massive Online Open Course (MOOC) sites like Coursera. These users, along with many other roles they have, are tasked to answer questions average users post on question-answer forums.

Designated experts should respond to many people at once. In the case of medical forums, medical experts have to answer questions submitted by users. For example, MedHelp, which is one of the leading medical websites, has over 12 million users every month. However, there are less than 300 designated medical experts who respond to patients' questions.

Furthermore, only a small portion of forum posts are answered by designated medical experts. These are people who have medical expertise that are certified by the website that hosts health forums. One work analyzed the role of designated medical experts on online health forums [15] The results show that designated medical experts play a critical role in responding to patients' information needs. In particular, 62.1% of online forum posts may benefit from clinical expertise. However, only 4.7% had responses from medical experts.

MOOCs similarly suffer from imbalance between the number of designated experts and users as well. Coursera has 5.2 million students and 532 courses[2], which corresponds to an average of approximately $10,000$ students per course. By building an effective recommender systems for designated experts, it would be possible to mitigate such imbalance problems, in particular by minimizing the time experts spend on looking for questions to answer. The system should utilize current approaches used by available recommender systems as well as exploit the unique behavior that designated experts might have.

With this goal in mind, we analyzed the behavior of designated experts. We analyzed 160 thousand threads from 'Ask a Doctor' forum (which consists of more than 60 sub-forums related to health) from MedHelp. This forum consists of designated experts (medical doctors) who answer average users' (patients') questions. Other non-designated users can also respond to the patients' questions as well. We found that less than 1% of threads have more than one expert who responded to a patient's query. This is a unique behavior of designated experts which does not seem to exist on CQA websites. We believe this is because the designated experts are very good at giving good answers to question askers. Other bystander experts may feel it is unnecessary to give an addendum to these answers. This contrasts to the behavior seen on 'Medical Support Communities' from MedHelp. These communities consist of average patients who seek support from fellow users. About 30% of the threads on these communities have more than one user who responded

[1] http://www.medhelp.org

[2] http://blog.coursera.org/post/64907189712/a-triple-milestone-107-partners-532-courses-5-2

to the patient seeking help.

In this paper we focus on this unique behavior designated experts have. In particular, we propose to capture it by using matrix factorization with a regularization framework and show that, by capturing how designated experts behave, the retrieval performance of the recommendation system improves.

Our contributions are as follows:

1. To the best of our knowledge, our work is the first in addressing the problem of recommending forum threads to designated experts. Although some previous research has focused on recommending questions and answers to users in CQA websites [9], [30], [34], [8], [19], [6], we were unable to find other works that addresses recommending forum threads to designated experts. Other related research has focused on e-mail routing commercial software such as IM an Expert[3]. However, these allow users to provide expertise and takes place in a corporate environment while our work focuses on online sources.
2. We propose a collective matrix factorization framework [28] to incorporate 1) forum posts, 2) user profiles and 3) user similarity.
3. We propose a matrix regularization framework [1] to capture the behavior of designated experts. In particular, we note that only one expert is likely to answer a given forum post. We propose a model that captures such characteristics.

The rest part of this paper is organized as follows: Our approach is described in detail in Section III. Section IV describes and shows results of our experiments. In particular, we first analyze the impact of each semantic types we introduce in this paper, and then the impact of the parameters. We describe related works in Section II and provide a conclusion in Section V V.

II. Related Works

Recommending or finding experts [13], [26] is an ongoing research field that helps question askers find the best domain expert. The goal here is to return a ranked list of best answerers based on the similarity between the query and the answerer's posting behavior. Guo et al. [13] proposed a probabilistic generative model for the QA community and employed a user-question-answer model. They evaluated their method on Yahoo! Answers. Another line of work [26] proposed combining both the user's profile and the questions that they have answered. They utilized a variant of LDA called segmented topic models [11] to recommend posts to users. Similarly, CQARank [30] made use of the Topic Expertise Model to learn about users topical interests, and inferred their model using Gaussian Mixture Model. Such research is especially helpful for CQA websites like Quora[4] where users can ask specified users to answer in exchange for on-website credits. The main goal of these works, however, is different from ours. While they focus on recommending experts to question askers (which helps question askers), our aim is to recommend forum posts to experts (which in turn helps experts save time).

Some research has investigated the expertise of such users. These are known as expertise retrieval [3] and widely used systems. Some popular examples are ArnetMiner[5] and Microsoft Academic Search[6]. These can be either profile based [2], where the works leverage existing user profiles to further build expertise of a given user, or document-based [25], which ranks candidates based on a combination of the documents relevance score and the degree to which the person is associated with that document Our approach builds users' expertise by analyzing documents that they have responded previously, and hence, is an example of document-based user modeling.

A more closely related line of research to our work is that of question recommender systems. These can be divided into two main areas: 1) those that utilize only the forum text, and 2) those that also leverage user's behavior. In the first line of research, there are works that recommend questions based on users' queries [7], forum threads [29] or posts [12]. Other works utilize answerer behavior to recommend users to posts [34], [8], [21], [33]. These methods are based on collaborative filtering [8], probabilistic models [34], [21], [6], or classification models [33]. Similar to our work, these works recommend which questions users should respond to. The difference, however, is that we focus on designated experts while they address only general question answerers, i.e., non-designated experts. To the best of our knowledge, our work is the first to tackle the problem of recommending questions to designated experts in online forum environments. We have noticed that, when it comes to designated experts, only one expert is likely to respond to a given post. On the other-hand, multiple users may respond to the question in CQA websites that do not have designated experts.

Our proposed solution has multiple objective functions to capture the characteristics of the forum posts and designated experts. Other works have phrased this as social collaborative filtering [23], [28] which captures various social aspects. We further extend this framework by adding constraints, or regularization [22] terms based on the our intuitions and data observation.

III. Our Approach

A. Base Formulation

Our formulation is defined using low-rank matrix formulation with implicit user feedback. In low-rank matrix

[3]http://research.microsoft.com/en-us/projects/imanexpert/

[4]http://www.quora.com/

[5]http://www.arnetminer.org/

[6]http://academic.research.microsoft.com/

factorization formulation, the goal is to approximate the feedback matrix R with two low-rank matrices. The general formulation is given as follows:

$$\mathcal{L}_{base}(U,P) \;=\; ||C \odot (R - U \cdot P^{\top})||_F \tag{1}$$

where $U \in \mathbb{R}^{n \times k}$ and $P \in \mathbb{R}^{m \times k}$ are low-rank matrices that corresponds to experts and posts, respectively, and $|| \cdot ||_F$ is Frobenius norm. The element-wise product is represented by $\odot$. The matrix C is relative importance of a given entry.

The feedback matrix R is captured by using binary user feedback, in particular, by determining whether an expert has responded to the given post or not. More formally, if a user i has responded to a j-th thread, then $R_{ij} = 1$, 0 otherwise. Furthermore, not all responses are equal. A user may not have responded to a particular thread because he was not interested in the thread, or because he has not seen it. We utilize a well known weighting metric [24] to mitigate the problem. The method proposes to frame the problem as weighted user feedback. The weight C is defined as follows:

$$C_{ij} = \begin{cases} 1, & \text{if user } i \text{ responded on } j\text{-th thread} \\ 1 - \mathrm{sim}(\vec{U}_i, \vec{P}_j), & \text{otherwise} \end{cases} \tag{2}$$

The vectors $\vec{U}_i$ and $\vec{P}_j$ are row vectors of matrices U and P, respectively. Our similarity metric $\mathrm{sim}(\vec{U}_i, \vec{P}_j)$ is cosine similarity between the word vector of the i-th user used and that of the j-th thread.

B. Modeling Post Content

Pure collaborative filtering approaches are not sufficient to handle the semantics. Past literature has suggested incorporating the context of the items [27], [4] or user profiles [20] to improve performance. We follow a similar formulation and model document content into the matrix factorization framework. Our formulation is as follows:

$$\mathcal{L}_{posts}(P,W) \;=\; ||D - P \cdot W^{\top}||_F \tag{3}$$

where $P \in \mathbb{R}^{m \times k}$ and $W \in \mathbb{R}^{|\vec{w}| \times k}$ are low-rank matrices for posts and words, respectively. The matrix $D \in \mathbb{R}^{m \times |\vec{w}|}$ is modeled from $TF\text{-}IDF$ weighting across the documents.

Modeling post content is an improvement over the base model as it explicitly states the document formulation. However, as will be evident in the experiments section, this approach does not model experts' interests.

C. Modeling Expert Interest

Similar to what we have done to model post contents, we wish to capture experts' interest as well. The challenge here is the user's profile page is often blank, or near blank. Instead of modeling these profile pages, we leverage the threads that the expert has responded to, and use these as the feature set. Our formulation is given as follows:

$$\mathcal{L}_{users}(U,W) \;=\; ||E - U \cdot W^{\top}||_F \tag{4}$$

where $U \in \mathbb{R}^{n \times k}$ and $W \in \mathbb{R}^{|\vec{w}| \times k}$ are low-rank matrices for experts and words, respectively. Similar to the document matrix D representation, the matrix $E \in \mathbb{R}^{n \times |\vec{w}|}$ is modeled using $TF\text{-}IDF$ weighting.

D. Modeling Expert Similarities

We cannot rely on interaction between two experts in our problem setting since in most cases, only one expert responds to a given thread. More precisely, since only one expert is likely to answer a given forum post, we cannot rely only on collaborative filtering to recommend posts for them to answer. Instead, we measure the similarities between the two experts based on the threads that they have responded to. If two experts are similar, they would be interested in similar forum threads, and vice versa for two experts that are not similar to each other. These are called social regularization framework in the context of social collaborative filtering [31], [10]. Our formulation to capture this intuition is given as follows:

$$\mathcal{L}_{sim}(U,V) \;=\; ||S - U \cdot V^{\top}||_F \tag{5}$$

where $U \in \mathbb{R}^{n \times k}$, $V \in \mathbb{R}^{n \times k}$ are expert low-rank matrices and their weights, respectively. $S \in \mathbb{R}^{n \times n}$ is a user similarity matrix.

Similarity matrix S is modeled using cosine similarity [14] between the two experts. For a given user i and i', we have

$$S_{i,i'} = \cos(\vec{E}_i, \vec{E}_{i'}) \tag{6}$$

where $S_{i,i'} \in S$, $\vec{E}_i$ is the word vector for the user i, and $\cos(\cdot,\cdot)$ is the cosine similarity between the two vectors.

E. Modeling Designated Expert Constraints

1) One response per post.: We have noted only one expert is likely to respond to a given forum post. This formulation is written as follows:

$$\sum_i R_{i,j} = 1, \;\forall j \in \{1, \ldots, m\} \tag{7}$$

This equation means some expert would answer the post j. Based on our formulation, $R_{i,j} \sim \vec{U}_i \cdot \vec{P}_j$, we now have, for each posts P_j

$$\sum_i \vec{U}_i \cdot \vec{P}_j = 1, \;\vec{U}_i \cdot \vec{P}_j = \{0,1\}, \;\forall j \in \{1, \ldots, m\} \tag{8}$$

as the constraints.

We note this is an NP-complete problem [17] since it is an instance of 0-1 integer linear programming and is not easily solvable. Instead, we relax the binary optimization problem to an optimization problem as the following

$$\sum_i \vec{U}_i \cdot \vec{P}_j = 1, \;\forall j \in \{1, \ldots, m\}, \vec{U}_i, \vec{P}_j \in \mathbb{R}^k \tag{9}$$

This can be reformulated, by taking its square, as

$$\mathcal{C}_{1j}(U,P) = \left(\sum_i \vec{U}_i \cdot \vec{P}_j - 1 \right)^2 = 0,\ \forall j \in \{1,\ldots,m\} \tag{10}$$

2) Propensity to answer.: Each experts has a different propensity to answer questions. Some may be prolific at answering questions, while others may not be. To capture this intuition, we introduce a constraint to capture the expected number of posts each expert may answer in the future. We note that $\sum_j R_{i,j}$ is the number of forum threads the expert i has responded to, or expected number of posts expert U_i is thought to have answered. We denote this as $E[U_i]$. Following similar logic as before, we propose the following formulation:

$$\begin{array}{l} \sum_j R_{i,j} = E[U_i] \\ \sum_j \vec{U}_i \cdot \vec{P}_j \approx E[U_i],\ \forall i \in \{1,\cdots,n\} \end{array} \tag{11}$$

Intuitively, this equation forces how often the expert responds to a given post. In the case the real value of $E[u]$ is not known, $E[u]$ can computed by calculating $m \times p(U=u|P)$, where m is the number of posts in forum, and $p(U=u|P)$ is the probability that the expert u would respond to a random post based on historic data. The constraints representation of this formula is given as follows:

$$\mathcal{C}_{2i}(U,P) = \left(\sum_j \vec{U}_i \cdot \vec{P}_j - E[U_i] \right)^2 = 0,\ \forall i \in \{1,\cdots,n\} \tag{12}$$

F. Learning Algorithm

We note that all of our constraints are convex which allows us to leverage KKT condition. We can express the constraints into part of the objective function as the following:

$$\mathcal{C}(U,P) = \sum_{j=1}^{m} \lambda_{c1j}\ \mathcal{C}_{1j}(U,P) + \sum_{i=1}^{n} \lambda_{c2i}\ \mathcal{C}_{2i}(U,P) \tag{13}$$

This allows us to encode both the constraints and the objective functions that we introduced in the previous section into an unconstrained optimization problem as the following:

$$\begin{aligned} \underset{U,P,V,W}{\arg\min}\ & \mathcal{L}_{base}(U,P) + \lambda_p \cdot \mathcal{L}_{posts}(P,W) \\ & +\lambda_u \cdot \mathcal{L}_{users}(U,W) + \lambda_s \cdot \mathcal{L}_{sim}(U,V) \\ & +\lambda_c \cdot \mathcal{C}(U,P) \\ & +\lambda_\beta \cdot (||U||_F + ||P||_F + ||W||_F + ||V||_F) \end{aligned} \tag{14}$$

In optimizing Equation 14, we applied a *stochastic gradient descent* (SGD) approach. SGDs are widely used in approximating low-rank matrices and have been shown to show good convergence and scalability properties [5]. We further note that computing equation 10 and equation 12 is rather expensive since it requires summation across rows and columns, respectively. Rather than computing this every iteration of SGD, we precompute these values every nm iterations. This proved to be a good compromise between the run time and the performance. Pseudo-code for our algorithm is shown in Algorithm 1.

Algorithm 1 Thread Recommender System for Experts

Input: $R \in \mathbb{R}^{n\times m}$, $D \in \mathbb{R}^{m\times|w|}$, $E \in \mathbb{R}^{n\times|w|}$, $S \in \mathbb{R}^{m\times m}$
Output: $U \in \mathbb{R}^{n\times k}$, $V \in \mathbb{R}^{n\times k}$, $P \in \mathbb{R}^{m\times k}$, $W \in \mathbb{R}^{|w|\times k}$

```
currIdx = 0
while Not converged do
    if currIdx % nm = 0 then
        Precompute Σ_i U_i · P_j  ∀j ∈ {1, ··· , m}
        Precompute Σ_j U_i · P_j  ∀i ∈ {1, ··· , n}
    end if
    Sample row index (i, j, l, i') from U, P, W, and V
    Update U_{i,*}, P_{j,*}, W_{l,*}, V_{i,i'} from ∂L(U, P, W, V)
    currIdx + +
end while
```

IV. Evaluations

A. Dataset and Experimental Settings

MedHelp (http://www.medhelp.org/) is a website where users can satisfy their medical information needs. The website consists of medical news, health diaries, support community forums and an 'Ask a Doctor' forum. Of special interest to us is the 'Ask a Doctor' forum. Average users post question on this forum and designated experts answer them if particular question suits their interest. There are 67 different forum categories, ranging from 'addiction' to 'undiagnosed symptoms.' For the purpose of our system, we only evaluated a forum if it had at least 5 active designated experts. In total, our dataset had 168 experts, 56,194 threads, and 18 forum categories.

For all of our experiments, we set $k = 100$ as the dimension of our latent features. We calculated all of the performance results by running the algorithm 10 times on each of the parameter settings. Unless otherwise mentioned, for each runs of evaluation, we randomly chose 80% of the posts as training data and the rest as testing data. The performance results we report here are based on taking the mean performance value of all 18 categories based on the best performing parameter combination for the given forum with the exception of the relative weights of the two constraints, λ_{c1j} and λ_{c2i}. Both of these parameters were set to 1 throughout all the forums. Based on the sensitivity analysis that we have run, the relative weights do not cause significant changes in performance. We set $\lambda_\beta = 0.001$, which prevents the algorithm from overfitting. Finally, our algorithm converges if the objective function changes less than 0.001 compared to the previous iteration.

B. Evaluation Metrics

Performance of each method is measured using Precision at 1, 5, 10 ($P@1$, $P@5$, $P@10$), Mean Average Precision (MAP) and Mean Reciprocal Ranking (MRR). $P@1$, $P@5$ and $P@10$ measures the proportion of recommended items that are ground-truth items amongst the top k retrieved results. This represents how many relevant threads an expert would see if they were presented with 1, 5 and 10 recommended threads. MAP is the arithmetic mean of average precision values over a set of all the retrieved results, and represents a generalization of what experts would see. MRR on the other hand, measures the inverse rank at which the relevant ground-truth appears. This measures how many recommended threads a user would need to browse to find a relevant thread.

C. Competing Algorithms

In order to evaluate our algorithm, we compare our algorithm against various other methods.

NMF : Non-negative Matrix Factorization [18] is a widely used algorithm in recommender system literature. These capture how users and items interact with each other. This is our first baseline approach.

MF-D : Matrix factorization with post content model. This approach combines $\mathcal{L}_{base}(U, P)$ and $\mathcal{L}_{posts}(P, W)$ to recommend posts to designated experts. Notice that this method does not take expert's interest into the model. This is our second baseline approach.

MF-ED : Matrix factorization with **MF-D** and expert interest model $\mathcal{L}_{users}(U, W)$. This is our third baseline approach.

MF-SED : Matrix factorization with **MF-ED** and expert similarity $\mathcal{L}_{sim}(U, V)$. This captures how similar experts are to each other.

MF-DEC : Matrix factorization with **MF-ED** and constraints $\mathcal{C}(U, P)$.

MF-all : Matrix factorization with **MF-ED**, expert similarity $\mathcal{L}_{sim}(U, V)$ and constraints $\mathcal{C}(U, P)$. This is our proposed method.

D. Performance Comparison

To evaluate the comparative performance of the algorithms, we ran all competing algorithm on all 18 forum categories. The performance comparison for all our methods is shown in Table I.

We notice that NMF does not perform very well. This is because NMF only captures the interaction between users (designated experts) and forum posts. In particular, NMF does not capture the content of the forum itself. We notice a huge performance difference between MF-D and MF-ED. This shows that, in recommending forum posts to experts, incorporating both the content of the post and that of the expert is both important. Adding expert similarity gives

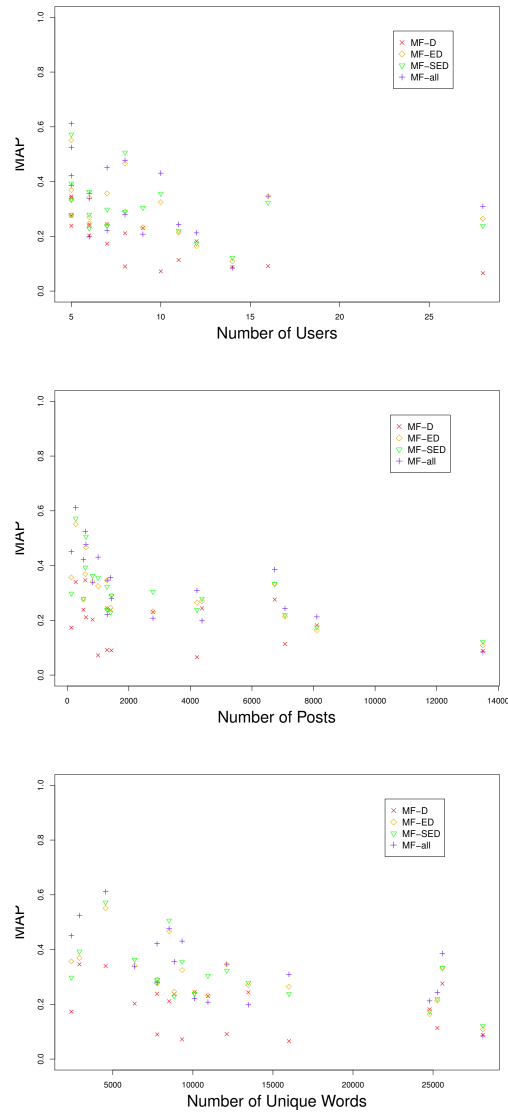

Figure 1. Impact of number of users, number of posts and number of words on the performance across all 18 forum categories.

some performance improvements over MF-ED. In a given forum, there are experts whose interests intersect. Similarity characteristics are captured by adding expert similarity factor, which is consistent with existing literature in social collaboration [32], [22], [23].

A significant boost is seen for MF-DEC over MF-SED and MF-ED. This indicates the constraints are suitable at capturing the macro level behavior in the existence of multiple designated experts. As we have argued in the preceeding sections, only one expert is likely to respond to a given post. The performance gain indicates that, indeed, constraints are well-modeled and improve the performance. Finally, the combined method performs the best which implies that all factors are complementary.

Algorithm	$P1$	$P5$	$P10$	MAP	MRR
NMF	0.1959	0.2005	0.1953	0.2263	0.1975
MF-D	0.0736	0.0587	0.0703	0.1914	0.0907
MF-ED	0.3114	0.2708	0.2639	0.2999	0.3066
MF-SED	0.3117	0.2827	0.2750	0.3068	0.3067
MF-DEC	0.3408*	0.3139*	0.3062*	0.3389*	0.3200
MF-all	**0.3698***	**0.3281***	**0.3172***	**0.3438***	**0.3390***

Table I

AVERAGE PERFORMANCE ON 18 DIFFERENT FORUM CATEGORIES. * DENOTES STATISTICAL SIGNIFICANCE OVER BASELINE METHODS AT $\alpha = 0.05$.

E. Impact of the Size of the Dataset

The next question we asked was, are there any differences in performance as we increase the size of the data set? In particular, how does the number of posts, unique words and users affect the performance of each of the algorithms? In this experiment, we plot the performance of all 18 forum categories into scatter plots. The x axis indicates number of users, posts or words, and y axis denotes the value for MAP. Our results are shown in Figure 1. In almost all forum categories, regardless of the size of the vocabulary, number of posts or users, MF-all consistently outperformed all the other comparison methods. This shows that our results are not biased by subset of forums on which our methods perform very well, or not penalized because of a few forums on which it performs poorly.

F. Sensitivity Analysis

We also conducted a sensitivity analysis of each parameter. There were four parameters that we tuned: These are λ_p which assigns how much weight to give to post contents, λ_u, which captures expert's interest, λ_s, which assigns relative weights to expert similarities, and λ_c which dictates the importance weight of the constraints. The results that we present are based on averaging the performance across all 18 forums for each of the parameters that we have ran.

We first tuned λ_p from MD-D method. We see here that this method is not very sensitive to the parameter. On the other hand, λ_u from MD-ED seems to more sensitive to the parameter. The two remaining parameters, λ_s and λ_c from MD-SED and MD-DEC, respectively, were not very sensitive to the parameters. This experiment shows that the two constraints which capture experts' behaviors are fairly robust. Furthermore, we notice that encoding expert similarities based on contents they write is robust as well.

We were also curious on how sensitive assigning different weights to the two constraints are. In particular, these weights are λ_{c2i} and λ_{c1j} as shown in Equation 13. Tuning each individual λ_{c1j} and λ_{c2i} for all of i and j would be rather expensive. Instead, we give uniform weights to all of λ_{c1j} and λ_{c2i} and vary the parameters. We note from this experiment that both λ_{c1j} and λ_{c2i} are not very sensitive to the parameters. All of our results are in Figure 2. Based on the two sensitivity analysis, we learn that our algorithm is robust to varying parameters. In particular, once λ_c is set, differing relative weights of the two types of constraints do not really affect performance.

V. DISCUSSION AND CONCLUSION

In this paper, we introduced the problem of recommending forum posts to designated experts. In particular, designated experts exhibit behavior that is not seen in other websites such as CQAs or web forums. For most of the questions that average users post, only one designated expert will answer them. To capture this intuition, we introduced constraints such that only one expert will answer a given post. The proposed constraints have improved performance based on 18 different forum categories that we have run the experiments on. We solved the recommendation problem using a matrix factorization framework. This was chosen because factorization framework allows easy extension to add additional constraints.

For future work, one can envision modeling how long an expert takes to answer a question, and incorporate this into the model. Furthermore, we only focused on the behavior of designated experts in our work. Moreover, can we capture interesting expert-user interactions? We also did not explore how semantics may affect whether an expert responds to a post or not. Studying the impact of language semantics may further help us understand designated experts' behaviors.

REFERENCES

[1] Francis Bach, Jacob Abernethy, Jean-Philippe Vert, and Theodoros Evgeniou. A new approach to collaborative filtering: Operator estimation with spectral regularization. *CoRR*, abs/0802.1430, 2008.

[2] Krisztian Balog, Leif Azzopardi, and Maarten de Rijke. Formal models for expert finding in enterprise corpora. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 43–50, New York, NY, USA, 2006. ACM.

[3] Krisztian Balog, Yi Fang, Maarten de Rijke, Pavel Serdyukov, and Luo Si. Expertise retrieval. *Found. Trends Inf. Retr.*, 6(2–3):127–256, February 2012.

[4] Linas Baltrunas, Bernd Ludwig, and Francesco Ricci. Matrix factorization techniques for context aware recommendation. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 301–304, New York, NY, USA, 2011. ACM.

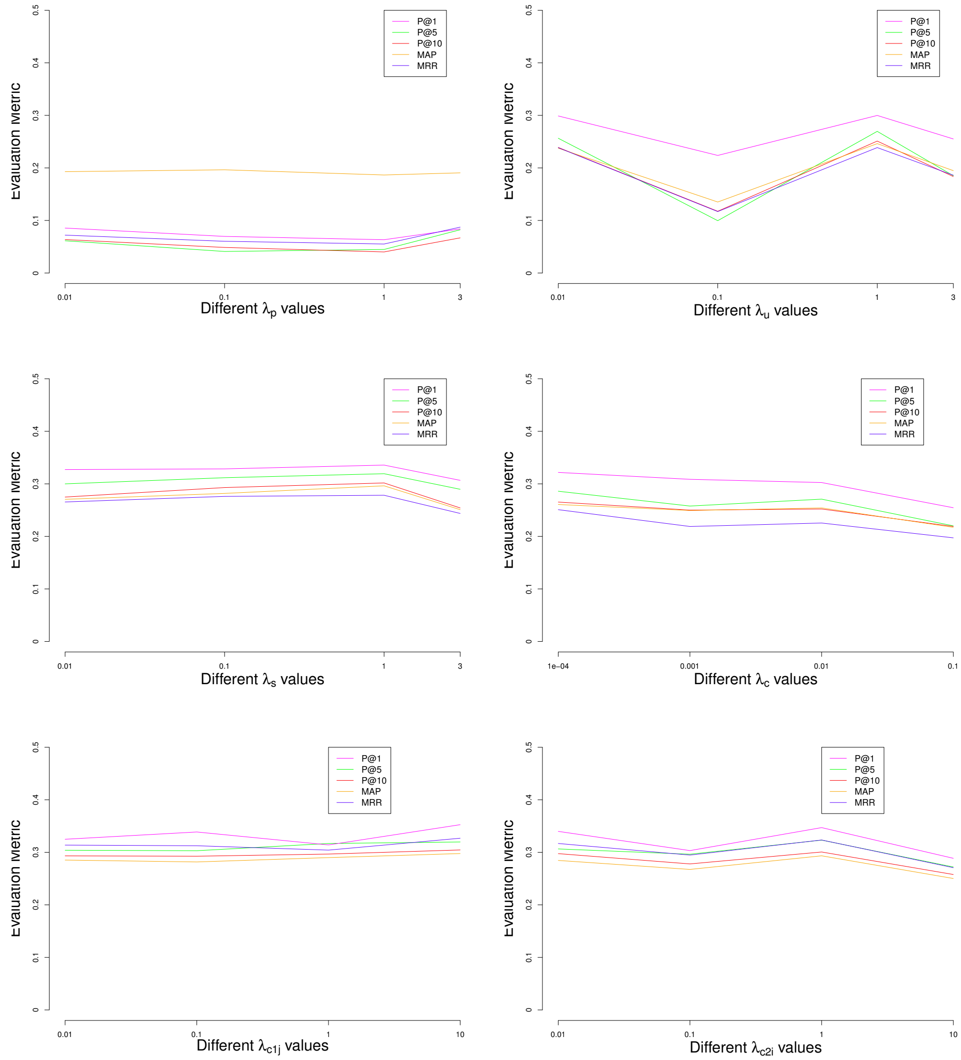

Figure 2. Parameter sensitivity on λ_p, λ_u, λ_s, λ_c, λ_{c1j} and λ_{c2i}

[5] Lon Bottou. Large-scale machine learning with stochastic gradient descent. In Yves Lechevallier and Gilbert Saporta, editors, *Proceedings of COMPSTAT'2010*, pages 177–186. Physica-Verlag HD, 2010.

[6] Grégoire Burel, Paul Mulholland, Yulan He, and Harith Alani. Predicting answering behaviour in online question answering communities. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, HT '15, pages 201–210, New York, NY, USA, 2015. ACM.

[7] Yunbo Cao, Huizhong Duan, Chin-Yew Lin, Yong Yu, and Hsiao-Wuen Hon. Recommending questions using the mdl-based tree cut model. In *Proceedings of the 17th International Conference on World Wide Web*, WWW '08, pages 81–90, New York, NY, USA, 2008. ACM.

[8] Shuo Chang and Aditya Pal. Routing questions for collaborative answering in community question answering. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM '13, pages 494–501, New York, NY, USA, 2013. ACM.

[9] Jason H. D. Cho, Parikshit Sondhi, Chengxiang Zhai, and Bruce R. Schatz. Resolving healthcare forum posts via similar thread retrieval. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, BCB '14, pages 33–42, New York, NY, USA, 2014. ACM.

[10] Peng Cui, Fei Wang, Shaowei Liu, Mingdong Ou, Shiqiang Yang, and Lifeng Sun. Who should share what?: Item-level social influence prediction for users and posts ranking. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '11, pages 185–194, New York, NY, USA, 2011. ACM.

[11] Lan Du, Wray L. Buntine, and Huidong Jin. A segmented topic model based on the two-parameter poisson-dirichlet process. *Machine Learning*, 81(1):5–19, 2010.

[12] Huizhong Duan and Chengxiang Zhai. Exploiting thread structures to improve smoothing of language models for forum post retrieval. In *Proceedings of the 33rd European Conference on Advances in Information Retrieval*, ECIR'11, pages 350–361, Berlin, Heidelberg, 2011. Springer-Verlag.

[13] Jinwen Guo, Shengliang Xu, Shenghua Bao, and Yong Yu. Tapping on the potential of Q&A community by recommending answer providers. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 921–930, New York, NY, USA, 2008. ACM.

[14] Jiawei Han. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2005.

[15] J. Huh, D. W. McDonald, A. Hartzler, and W. Pratt. Patient moderator interaction in online health communities. *AMIA Annu Symp Proc*, 2013:627–636, 2013.

[16] Jiwoon Jeon, W. Bruce Croft, Joon Ho Lee, and Soyeon Park. A framework to predict the quality of answers with non-textual features. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 228–235, New York, NY, USA, 2006. ACM.

[17] R. Karp. Reducibility among combinatorial problems. In R. Miller and J. Thatcher, editors, *Complexity of Computer Computations*, pages 85–103. Plenum Press, 1972.

[18] Daniel D. Lee and H. Sebastian Seung. Algorithms for non-negative matrix factorization. In *NIPS*, pages 556–562, 2000.

[19] Baichuan Li and Irwin King. Routing questions to appropriate answerers in community question answering services. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 1585–1588, New York, NY, USA, 2010. ACM.

[20] Yanen Li, Jia Hu, ChengXiang Zhai, and Ye Chen. Improving one-class collaborative filtering by incorporating rich user information. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, pages 959–968, New York, NY, USA, 2010. ACM.

[21] Mingrong Liu, Yicen Liu, and Qing Yang. Predicting best answerers for new questions in community question answering. In *Proceedings of the 11th International Conference on Web-age Information Management*, WAIM'10, pages 127–138, Berlin, Heidelberg, 2010. Springer-Verlag.

[22] Hao Ma, Dengyong Zhou, Chao Liu, Michael R. Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 287–296, New York, NY, USA, 2011. ACM.

[23] Joseph Noel, Scott Sanner, Khoi-Nguyen Tran, Peter Christen, Lexing Xie, Edwin V. Bonilla, Ehsan Abbasnejad, and Nicolas Della Penna. New objective functions for social collaborative filtering. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 859–868, New York, NY, USA, 2012. ACM.

[24] Rong Pan, Yunhong Zhou, Bin Cao, N.N. Liu, R. Lukose, M. Scholz, and Qiang Yang. One-class collaborative filtering. In *Data Mining, 2008. ICDM '08. Eighth IEEE International Conference on*, pages 502–511, Dec 2008.

[25] D. Petkova and W.B. Croft. Hierarchical language models for expert finding in enterprise corpora. In *Tools with Artificial Intelligence, 2006. ICTAI '06. 18th IEEE International Conference on*, pages 599–608, Nov 2006.

[26] Fatemeh Riahi, Zainab Zolaktaf, Mahdi Shafiei, and Evangelos Milios. Finding expert users in community question answering. In *Proceedings of the 21st International Conference Companion on World Wide Web*, WWW '12 Companion, pages 791–798, New York, NY, USA, 2012. ACM.

[27] Yue Shi, Martha Larson, and Alan Hanjalic. Mining mood-specific movie similarity with matrix factorization for context-aware recommendation. In *Proceedings of the Workshop on Context-Aware Movie Recommendation*, CAMRa '10, pages 34–40, New York, NY, USA, 2010. ACM.

[28] Ajit P. Singh and Geoffrey J. Gordon. Relational learning via collective matrix factorization. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, pages 650–658, New York, NY, USA, 2008. ACM.

[29] Amit Singh, Deepak P, and Dinesh Raghu. Retrieving similar discussion forum threads: A structure based approach. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 135–144, New York, NY, USA, 2012. ACM.

[30] Liu Yang, Minghui Qiu, Swapna Gottipati, Feida Zhu, Jing Jiang, Huiping Sun, and Zhong Chen. Cqarank: Jointly model topics and expertise in community question answering. In *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM '13, pages 99–108, New York, NY, USA, 2013. ACM.

[31] Shuang-Hong Yang, Bo Long, Alex Smola, Narayanan Sadagopan, Zhaohui Zheng, and Hongyuan Zha. Like like alike: Joint friendship and interest propagation in social networks. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 537–546, New York, NY, USA, 2011. ACM.

[32] Quan Yuan, Li Chen, and Shiwan Zhao. Factorization vs. regularization: Fusing heterogeneous social relationships in top-n recommendation. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 245–252, New York, NY, USA, 2011. ACM.

[33] Tom Chao Zhou, Michael R. Lyu, and Irwin King. A classification-based approach to question routing in community question answering. In *Proceedings of the 21st International Conference Companion on World Wide Web*, WWW '12 Companion, pages 783–790, New York, NY, USA, 2012. ACM.

[34] Yanhong Zhou, Gao Cong, Bin Cui, Christian S. Jensen, and Junjie Yao. Routing questions to the right users in online communities. In *Data Engineering, 2009. ICDE '09. IEEE 25th International Conference on*, pages 700–711, March 2009.