



Collaborative Recommendation of Temporally-Discounted Tag-Based Expertise for Community Question Answering

Gianni Costa and Riccardo Ortale^(✉)

ICAR-CNR, Via P. Bucci 8/9C, Rende, CS, Italy
{costa,ortale}@icar.cnr.it

Abstract. We propose an innovative approach to finding experts for community question answering (CQA). The idea is to recommend answerers, who are credited the highest expertise under question tags at routing time. The expertise of answerers under already replied question tags is intuitively discounted by accounting for the observed tags, votes and temporal information of their answers. Instead, the discounted expertise under not yet replied tags is predicted via a latent-factor representation of both answerers and tags. These representations are inferred by means of Gibbs sampling under a new Bayesian probabilistic model of discounted user expertise and asking-answering behavior. The devised model unprecedentedly explains the latter two CQA aspects as the result of a generative process, that seamlessly integrates probabilistic matrix factorization and network behavior characterization. An extensive comparative experimentation over real-world CQA data demonstrates that our approach outperforms several-state-of-the-art competitors in recommendation effectiveness.

1 Introduction

Expert recommendation [17] enables the timely sharing of high-quality knowledge for community question answering (CQA) [16]. Unfortunately, despite several previous research efforts, expert recommendation still remains problematic for various reasons. Firstly, question-answering (QA) communities are inherently time-evolving [21], with new users (both askers and answerers) joining daily and the existing users changing their interests and behavior (such as, e.g., long-term inactive users, who turn into active participants). Hitherto, expert recommendation has been studied, mostly by ignoring the temporal information of posts. Accordingly, the devised approaches are suitable neither to deal with the natural drift of users' interests over time, nor to promote short-term answerers (i.e., users with a limited recent answering history). This is a severe limitation, that lowers the effectiveness of expert recommendation, since the recommended experts may no more reply to questions matching initially-relevant and already-outdated

interests. In addition, short-term answerers would likely not be recommended at all, being their expertise gained in a far too short time period for such users to build a solid reputation as actual experts. Secondly, there is no common agreement on the choice of the discriminative content features to capture answerers' expertise. In most studies, the latter is inferred from the raw text of their answers, suitably weighted by the respective votes from the QA community. Tags are mainly ignored or, alternatively, incorporated into post contents as in [20]. However, tags are more insightful, concise and explicit user-generated explanations of both post meaning and topical expertise, with respect to the general topics inferrable from the textual post contents [19]. Thirdly, supplementing content features with further auxiliary data (e.g., networks of user interactions) for more effective expert recommendation involves devising a plausible joint processing of such information. Hitherto, both sources of information have been combined mainly through simplistic schemes such as, e.g., linear interpolation [17].

In this paper, we propose a new collaborative approach to recommending question-specific experts in QA communities. The expertise of answerers is determined from the tags, votes and temporal information of their answers as well as the asking-answering relationships in the targeted QA community. More precisely, answer tags are employed to capture and represent the topical expertise of answerers. Votes indicate the degree to which answerers are publicly acknowledged within the QA community as experts under the tags of the respective answers. Posting time allows for discounting [8] earlier answers of responders, so that to account for the natural drift of their interests over time, without penalizing short-term answerers. Besides, asking-answering interactions inform the identification of experts, since repliers to expert askers are likely to be expert as well. Essentially, for each posted question, the intuition behind the presented approach consists in recommending answerers, who are credited the highest degree of expertise under the tags of the particular question at routing time. In particular, the expertise of answerers under already replied tags is intuitively determined by means of the votes and temporal information of their answers to the questions labelled with such tags. Instead, the unknown expertise of answerers under not yet replied tags is predicted through a latent-factor generative model of temporally-discounted user expertise and asking-answering behavior. Under such a model, Bayesian probabilistic matrix factorization [15] and the statistical formalization of asking-answering are seamlessly integrated. This allows for explaining both the expertise of users and their behavioral patterns as the result of a generative process, that is governed by a certain number of latent factors. These are estimated via a MCMC algorithm, that implements the derived mathematical details of Gibbs sampling inference under the devised model.

Extensive tests over real-world CQA data show that our approach overcomes several state-of-the-art competitors in recommendation effectiveness.

This paper proceeds as follows. Notation and preliminaries are introduced in Sect. 2. The devised model is developed in Sect. 3. Expertise prediction for recommendation and posterior inference are covered in Sect. 4. The experimental evaluation of our approach is presented in Sect. 5. Finally, conclusions are drawn in Sect. 6, where future research is also previewed.

2 Preliminaries

A question-answering (QA) community \mathcal{D} can be formalized as a triple $\mathcal{D} \triangleq \langle \mathbf{U}, \mathbf{T}, \mathbf{G} \rangle$, where

- $\mathbf{U} = \{u_1, \dots, u_N\}$ is a set of N users;
- $\mathbf{T} = \{t_1, \dots, t_M\}$ is a set of M tags;
- $\mathbf{G} = \langle \mathbf{V}, \mathbf{A} \rangle$ is a directed communication network shaped by user interaction behavior, with $\mathbf{V} \subseteq \mathbf{U}$ and $\mathbf{A} \subseteq \mathbf{U} \times \mathbf{U}$ being the set of nodes and edges, respectively.

The generic user $u \in \mathbf{U}$ can ask questions and/or provide answers. In order to capture the expertise of u , we focus on her answering history $\mathbf{a}_u = \{a_{u,1}, \dots, a_{u,N_u}\}$. \mathbf{a}_u is the time sequence of N_u replies from u to as many questions posted by other users of \mathcal{D} . The arbitrary answer $a_{u,h} \in \mathbf{a}_u$ (with $h = 1, \dots, N_u$) is associated with a respective timestamp $ts_{u,h}$, an explicative set of tags $\mathbf{t}_{u,h} \subseteq \mathbf{T}$ and a vote score $s_{u,h}$. $ts_{u,h}$ indicates when $a_{u,h}$ was posted. For any two answers $a_{u,h_i}, a_{u,h_j} \in \mathbf{a}_u$, $h_i < h_j$ iff $ts_{u,h_i} < ts_{u,h_j}$. Timestamps are useful for reasonably dealing with the drift of the interests and skills of u , across the respective answering history \mathbf{a}_u , by means of gradual forgetting [8]. The latter consists in estimating the expertise of u from the whole answering history \mathbf{a}_u , so that the earlier answers are realistically considered to be outdated and, thus, less informative of her current interests and skills. The tags in $\mathbf{t}_{u,h}$ are an insightful description of the actual themes covered by $a_{u,h}$ ¹. In principle, $\mathbf{t}_{u,h}$ is a more accurate representation of the both the intended meaning of $a_{u,h}$ and the topical expertise of u , in comparison with the more general topics inferrable from the textual content of $a_{u,h}$ [19]. For this reason, the wording of $a_{u,h}$ is disregarded and, consequently, the computational burden of processing very large amounts of raw text is avoided. $s_{u,h}$ indicates the acknowledged degree of expertise gained by u with regard to the question answered through $a_{u,h}$ and, by extension, under each tag within $\mathbf{t}_{u,h}$.

At the current timestamp now , the expertise of all users in \mathbf{U} under the tags of \mathbf{T} is summarized by matrix $\mathbf{E}^{(now)}$. Its generic entry $E_{ut}^{(now)}$ quantifies the expertise of user u under tag t at time now as the below weighted average

$$E_{ut}^{(now)} = \frac{\sum_{a_{u,h} \in \mathbf{a}_u} s_{u,h} \cdot \delta_{t,\mathbf{t}_{u,h}} \cdot w_{u,h}^{(now)}}{\sum_{a_{u,h} \in \mathbf{a}_u} \delta_{t,\mathbf{t}_{u,h}} \cdot w_{u,h}^{(now)}} \quad (1)$$

In Eq. 1, $\delta_{t,\mathbf{t}_{u,h}}$ is 1 iff $t \in \mathbf{t}_{u,h}$ for some h (with $h = 1, \dots, N_u$), and 0 otherwise. If $\delta_{t,\mathbf{t}_{u,h}} = 0$ for each $h = 1, \dots, N_u$, $E_{ut}^{(now)}$ is assumed to be 0, which corresponds to an unknown or missing value. Besides, $w_{u,h}^{(now)} = e^{-\lambda(now-ts_{u,h})}$ is a weighting scheme, that implements gradual forgetting by exponential ageing. Intuitively, the earlier answers are not ignored in the estimation of the current expertise of u under t . Rather, their contribution to $E_{ut}^{(now)}$ exponentially decays

¹ Answers retain the tags attached to the respective questions.

according to the respective timestamps. Remarkably, such a modeling choice does not penalize the expertise of those users with a short replying history (such as new users or mostly inactive users with a recent answering history), without discarding the old answers of long-term answerers. Notice that $w_{u,h}^{(now)}$ is parameterized by the decay rate λ . The latter determines how rapidly the contribution of answers to user expertise decays over time. Essentially, larger values of λ imply a faster decay of earlier answers.

As a supplement to the information from the answering history of users, their asking-answering interactions are also captured as edges of \mathbf{G} . More precisely, an edge $u_i \rightarrow u_j$ from a responder u_i to an asker u_j belongs to \mathbf{A} , if u_i answered at least one question posted by u_j . By an abuse of notation, we also write \mathbf{G} to denote the adjacency matrix associated with the asking-answering graph. The generic entry G_{ij} is 1 iff $u_i \rightarrow u_j \in \mathbf{A}$ and 0 otherwise.

2.1 Problem Statement

Given a question q , let \mathbf{t}_q be the set of tags attached to q by the asker. Also, assume that *now* is the time, when q is routed to the answerers. We aim to recommend q to targeted users, with the highest acknowledged expertise in the tags of \mathbf{t}_q at time *now*, who are most likely to reply with high-quality answers.

Unfortunately, in the context of the generic QA community \mathbf{D} , $\mathbf{E}^{(now)}$ and \mathbf{G} are generally very sparse. Consequently, the expertise of users under specific tags within \mathbf{t}_q may not be known. In this paper, we exploit latent-factor modeling to predict the unknown values of $\mathbf{E}^{(now)}$, that correspond to the current expertise of answerers under the various adopted tags. Thus, experts can be simply recommended from a list of answerers, ranked by their average expertise under the tags attached to q .

Hereinafter, to avoid cluttering notation, we will write \mathbf{E} to mean $\mathbf{E}^{(now)}$.

3 The ENGAGE Model

ENGAGE (*timE-evolviNG tAG-based Expertise*) is a Bayesian generative latent-factor model of temporally-discounted expertise and asking-answering behavior in QA communities. Under ENGAGE, the matrices \mathbf{E} and \mathbf{G} of a QA community \mathbf{D} are the result of a probabilistic generative process, that is ruled by K latent factors. These are captured by embedding users and tags in a K -dimensional latent space, through the seamless integration of Bayesian probabilistic matrix factorization [15] and the statistical modeling of the asking-answering behavior.

Formally, each user $u \in \mathbf{U}$ is associated with a column vector $\mathbf{L}_u \in \mathbb{R}^K$. The k -th entry of \mathbf{L}_u (with $k = 1, \dots, K$) is a random variable representing the unknown degree to which the k -th latent factor explains the expertise of u . Analogously, each tag $t \in \mathbf{T}$ is associated with a column vector $\mathbf{H}_t \in \mathbb{R}^K$. The k -th entry of \mathbf{H}_t (with $k = 1, \dots, K$) is a random variable representing the unknown extent to which the k -th latent factor is inherently characteristic of t . The latent-factor representation of all users and tags is collectively denoted

as $\mathbf{L} \in \mathbb{R}^{K \times N}$ and $\mathbf{H} \in \mathbb{R}^{K \times M}$, respectively. The data likelihood (i.e., the conditional distribution over the entries of \mathbf{E} and \mathbf{G}) is

$$\Pr(\mathbf{E}|\mathbf{L}, \mathbf{H}, \alpha) = \prod_{u \in \mathbf{U}} \prod_{t \in \mathbf{T}} \mathcal{N}(E_{ut}; \mu_{u,t}, \alpha^{-1})^{\delta_{ut}} \quad (2)$$

$$\Pr(\mathbf{G}|\mathbf{L}, \beta) = \prod_{u_i \rightarrow u_j \in \mathbf{A}} \mathcal{N}(G_{ij}; \mu_{u_i, u_j}, \beta^{-1}) \quad (3)$$

with

$$\mu_{u,t} = \mathbf{L}_u^T \cdot \mathbf{H}_t \text{ and } \mu_{u_i, u_j} = \mathbf{L}_{u_i}^T \mathbf{L}_{u_j}$$

In the above Eq. 2 and Eq. 3, $\mathcal{N}(\cdot|\mu, \alpha^{-1})$ is the Gaussian distribution having mean μ and precision α . In particular, according to Eq. 2, the current expertise of answerers under the adopted tags is centered around the intuitive explanation provided by the dot product of the respective latent-factor representations. δ_{ut} is 1 iff $E_{ut} > 0$ (i.e., if the expertise of u under t is actually acknowledged) and 0 otherwise. Equation 3 seamlessly incorporates the supplementary information regarding the asking-answering interactions of users. Specifically, according to Eq. 3, the asking-answering interactions are centered around the degree of agreement between the involved users. This provides a valuable contribution to the identification of experts, since those users, who answer questions from other users with a high expertise, are also likely to have gained a high expertise.

The latent-factor representations of users and tags stem from multivariate Gaussian prior distributions parameterized, respectively, by $\Theta_{\mathbf{L}} = \{\boldsymbol{\mu}_{\mathbf{L}}, \Lambda_{\mathbf{L}}\}$ and $\Theta_{\mathbf{H}} = \{\boldsymbol{\mu}_{\mathbf{H}}, \Lambda_{\mathbf{H}}\}$. In turn, such parameters are drawn from the below Gaussian-Wishart prior distributions (hereinafter indicated as \mathcal{NW}) [2]

$$\Pr(\Theta_{\mathbf{X}}|\Theta_0) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}; \boldsymbol{\mu}_0, [\beta_0 \Lambda_{\mathbf{X}}]^{-1}) \cdot \mathcal{W}(\Lambda_{\mathbf{X}}; \nu_0, \mathbf{W}_0)$$

where $\mathbf{X} \in \{\mathbf{L}, \mathbf{H}\}$, $\mathcal{W}(\Lambda_{\mathbf{X}}; \nu_0, \mathbf{W}_0)$ denotes the Wishart distribution [2] and $\Theta_0 = \{\boldsymbol{\mu}_0, \beta_0, \nu_0, \mathbf{W}_0\}$ is a set of hyperparameters.

The conditional (in)dependencies among the random variables under ENGAGE are shown by means of plate notation in Fig. 1a. Notice that unshaded nodes correspond to latent factors, whereas shaded nodes correspond to observed magnitudes. The generative process modeled by ENGAGE performs the realization of the observed random variables (i.e., the individual entries of \mathbf{E} and \mathbf{G}) according to the conditional (in)dependencies of Fig. 1a as detailed in Fig. 1b.

4 Model Inference

Under ENGAGE, the experts for a given question q are found by ranking users based on a recommendation score, that involves the latent-factor representations of users and tags. The recommendation score is introduced in Sect. 4.1. The inference of the latent-factor representations is discussed in Sect. 4.2.

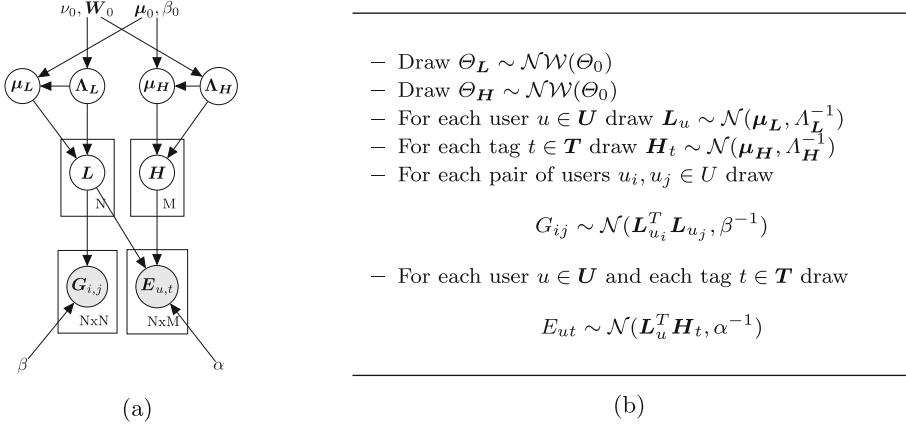


Fig. 1. Graphical representation of ENGAGE (a) and its generative process (b).

4.1 Answer Ranking for Recommendation

The rank of an answerer $u \in \mathbf{U}$ in the list of experts for q is determined by the score P_{uq} of her acknowledged/predicted expertise. P_{uq} is computed by averaging the current expertise of u under the individual tags of \mathbf{t}_q . This requires to distinguish between two alternative cases. Let t be an adopted tag of \mathbf{t}_q . If the expertise of u under t is acknowledged, E_{ut} can be directly used in the definition of P_{uq} . Otherwise, if the expertise of u under t is an unknown entry of \mathbf{E} , then E_{ut} is suitably predicted by resorting to the latent-factor representations of u and t under ENGAGE. Accordingly, $P_{uq} = \frac{1}{|\mathbf{t}_q|} \sum_{t \in \mathbf{t}_q} \hat{E}_{ut}$, where \hat{E}_{ut} is defined in the below Eq. 4, so that to incorporate the current expertise of u under t according to the two above cases.

$$\hat{E}_{ut} = \begin{cases} E_{ut} & \text{if } E_{ut} > 0 \\ \frac{1}{S} \sum_{s=1}^S (\mathbf{L}_u^{(s)})^T \cdot \mathbf{H}_t^{(s)} & \text{if } E_{ut} = 0 \end{cases} \quad (4)$$

In Eq. 4, S is the number of samples of both \mathbf{L}_u and \mathbf{H}_t , which are respectively referred to as $\mathbf{L}_u^{(s)}$ and $\mathbf{H}_t^{(s)}$ (with $s = 1, \dots, S$). Assume that $\Theta = \{\mathbf{L}, \mathbf{H}\} \cup \Theta_L \cup \Theta_H$. In principle, all samples $\mathbf{L}_u^{(s)}$ and $\mathbf{H}_t^{(s)}$ are to be drawn from the posterior distribution $\Pr(\Theta | \mathbf{E}, \mathbf{G}, \alpha, \beta, \Theta_0)$. However, the latter is analytically intractable. Therefore, the generic $\mathbf{L}_u^{(s)}$ and $\mathbf{H}_t^{(s)}$ are drawn through approximate posterior inference, as described in Sect. 4.2.

4.2 Approximate Posterior Inference

A well-known technique for approximate stochastic inference [14] is Gibbs sampling. The latter defines a (first-order) Markov chain, whose stationary distribution eventually approaches the true posterior distribution $\Pr(\Theta | \mathbf{E}, \mathbf{G}, \alpha, \beta, \Theta_0)$.

This is accomplished by means of reiterated transitions from the current sample of the model parameters Θ to a new one. More precisely, at the generic transition, each parameter $\theta \in \Theta$ is sequentially sampled from the respective full conditional $\Pr(\theta|\Theta - \theta, \mathbf{E}, \mathbf{G}, \alpha, \beta, \Theta_0)$. This is the conditional distribution over θ , given all other parameters $\Theta - \theta$, the (hyper)parameters Θ_0 as well as the observations \mathbf{E} and \mathbf{G} .

The derived full conditional distributions over the individual parameters of ENGAGE are reported next, along with the algorithm designed to perform Gibbs sampling inference.

Parameters \mathbf{L}_u and \mathbf{H}_t . Due to the conjugacy between the multivariate Gaussian distribution on \mathbf{L}_u (with unknown parameters Θ_L) and the Gaussian-Wishart prior on Θ_L , the full conditional on \mathbf{L}_u is a multivariate Gaussian distribution, i.e.,

$$\mathbf{L}_u \sim \mathcal{N}\left(\boldsymbol{\mu}_L^{*(u)}, \left[\Lambda_L^{*(u)}\right]^{-1}\right) \quad (5)$$

where

$$\begin{aligned} \Lambda_L^{*(u)} &= \Lambda_L + \alpha \sum_{t \in \mathcal{T}} \delta_{ut} \mathbf{H}_t \mathbf{H}_t^T + \beta \sum_{u \in \mathcal{U}} \mathbf{L}_u \mathbf{L}_u^T \\ \boldsymbol{\mu}_L^{*(u)} &= \left[\Lambda_L^{*(u)}\right]^{-1} \left[\alpha \sum_{t \in \mathcal{T}} \delta_{ut} \mathbf{H}_t E_{ut} + \beta \sum_{v \in \mathcal{U}} \mathbf{L}_v G_{uv} + \Lambda_L \boldsymbol{\mu}_L \right] \end{aligned}$$

Likewise, because of the conjugacy between the multivariate Gaussian distribution on \mathbf{H}_t (with unknown parameters Θ_H) and the Gaussian-Wishart prior on Θ_H , the full conditional on \mathbf{H}_t is a multivariate Gaussian distribution, i.e.,

$$\mathbf{H}_t \sim \mathcal{N}\left(\boldsymbol{\mu}_H^{*(t)}, \left[\Lambda_H^{*(t)}\right]^{-1}\right) \quad (6)$$

with

$$\begin{aligned} \Lambda_H^{*(t)} &= \Lambda_H + \alpha \sum_{u \in \mathcal{U}} \delta_{ut} \mathbf{L}_u \mathbf{L}_u^T \\ \boldsymbol{\mu}_H^{*(t)} &= \left[\Lambda_H^{*(t)}\right]^{-1} \left[\alpha \sum_{u \in \mathcal{U}} \mathbf{L}_u \delta_{ut} E_{ut} + \Lambda_H \boldsymbol{\mu}_H \right] \end{aligned}$$

Parameters Θ_L and Θ_H . For each $\mathbf{X} \in \{\mathbf{L}, \mathbf{H}\}$, the conditional distribution over $\Theta_X = \{\boldsymbol{\mu}_X, \Lambda_X\}$ is the below Gaussian-Wishart distribution [6, pp. 178]

$$\begin{aligned} \Pr(\boldsymbol{\mu}_X, \Lambda_X | \mathbf{X}, \Theta_0) &= \mathcal{N}(\boldsymbol{\mu}_X | \boldsymbol{\mu}_X^*, [(\beta_0 + c)\Lambda_X]^{-1}) \\ &\quad \cdot \mathcal{W}(\Lambda_X | \nu_0 + c, \mathbf{W}_X^*) \end{aligned} \quad (7)$$

where c is the number of columns within matrix \mathbf{X} and

$$\begin{aligned} \boldsymbol{\mu}_X^* &= \frac{\beta_0 \boldsymbol{\mu}_0 + c \bar{\mathbf{X}}}{\beta_0 + c}; \quad \mathbf{S}_X = \frac{1}{c} \sum_{i=1}^c (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^T; \quad \bar{\mathbf{X}} = \frac{1}{c} \sum_{i=1}^c \mathbf{X}_i \\ [\mathbf{W}_X^*]^{-1} &= \mathbf{W}_0 c^{-1} + c \mathbf{S}_X + \frac{\beta_0 c}{\beta_0 + c} (\boldsymbol{\mu}_0 - \bar{\mathbf{X}})(\boldsymbol{\mu}_0 - \bar{\mathbf{X}})^T \end{aligned}$$

Algorithm 1: Pseudo code of the Gibbs sampling algorithm

```

Input:  $\alpha, \beta, \Theta_0 = \{\mu_0, \beta_0, \nu_0, W_0\}$  and  $H$ ;
Output: samples  $L_u^{(s)}, H_t^{(s)}$  with  $s = 1, \dots, S$ ;
1 Initialize  $L^{(0)}$  and  $H^{(0)}$ ;
2 for  $h = 1, \dots, H$  do
3   Draw  $\Theta_L^{(h)} \sim \text{Pr}(\Theta_L^{(h)} | L, \Theta_0)$  through Eq. 7;
4   Draw  $\Theta_H^{(h)} \sim \text{Pr}(\Theta_H^{(h)} | H, \Theta_0)$  through Eq. 7;
5   for each  $u \in U$  do
6     Draw  $L_u^{(h)} \sim \mathcal{N}(\mu_P^{*(u)}, [A_P^{*(u)}]^{-1})$  through Eq. 5;
7   end
8   for each  $t \in T$  do
9     Draw  $H_t^{(h)} \sim \mathcal{N}(\mu_H^{*(t)}, [A_H^{*(t)}]^{-1})$  through Eq. 6;
10  end
11 end

```

Gibbs Sampling. Algorithm 1 sketches the pseudo code of the sampler, designed to implement approximate posterior inference under ENGAGE. After a preliminary initialization (line 1), the sampler enters a loop (lines 2–11), whose generic iteration h embraces two steps. $\Theta_L^{(h)}$ and $\Theta_H^{(h)}$ are drawn at the first step (lines 3–4), being functional to draw $L_u^{(h)}$ and $H_t^{(h)}$ at the second step (lines 5–10).

The maximum number H of iterations is established, by following the widely-adopted convergence-criterion in [12]. This allows the Markov chain behind the Gibbs sampler to reach its equilibrium after an initial burn-in period. As a consequence, the S samples used in Eq. 4, can be drawn when convergence is met (i.e., after the burn-in period, in which samples are instead still sensible to the preliminary initialization).

5 Experimental Evaluation

We comparatively investigated the recommendation effectiveness of ENGAGE.

5.1 Data Set

All experiments were conducted on *Stack Overflow*² [1, 16], i.e., a real-world QA community for sharing knowledge on computer programming. More precisely, we formed our training and test sets from an anonymized and quarterly dump³ of all *Stack Overflow* data, produced by its users within a time interval ranging from Jan 1, 2015 to July 31, 2015. Such a dump is publicly released by the *Stack Exchange* network under the *Creative Commons BY-SA 4.0* licence. More precisely, as far as the training set is concerned, we retained all those tags that were adopted at least 50 times in the time interval from Jan 1, 2015 to June 31, 2015. Further, we considered all those users, who provided more than 80

² <https://stackoverflow.com/>.

³ <https://archive.org/download/stackexchange>.

posts [20] in the same period. The selected tags and users, along with their answers, the respective questions, timestamps and votes were included into the training set. Overall, the latter consists of 3,376 users, 40,382 questions, 60,968 answers. Regarding the test set, we focused on a collection \mathbf{Q} of questions (with $|\mathbf{Q}| = 1,357$), that were posted by the users in the training set in a later time interval from July 1, 2015 to July 31, 2015. These questions are labelled with tags and answered by answerers in the training set. We chose such users, their answers to the questions of \mathbf{Q} , the respective timestamps and votes as the test set. As a whole, the latter consists of 1,357 questions, 3,376 users, 3,771 answers.

5.2 Competitors

We contrasted ENGAGE against a selection of various competitors.

Votes [17] ranks answerers based on the mean of the difference between the positive and negative votes of their answers as well as the average percentage of the positive votes.

InDegree [3] ranks answerers by their respective numbers of best answers.

The state-of-the-art model in [19], hereinafter called TER (*Tag-based Expert Recommendation*), infers user expertise from the factorization of the user-tag matrix. The latter is built, so that the generic entry reflects the expertise of an answerer under a tag, as captured by averaging the votes of her answers marked by that tag. Unlike ENGAGE, TER ignores both the drift of users' interests over time and their asking-answering behavior.

TEM [20] is a state-of-the-art joint model of topics and expertise. Essentially, under TEM, tags are incorporated into the textual content of posts, in order to infer the topical interests of users. The specific expertise of users under the different topics is explicitly captured.

CQARank [20] combines the user topical interests and expertise under TEM with the link analysis of the asking-answering interaction graph, in order to enhance the inference of user topical expertise.

Both TEM and CQARank disregard the drift of users' interests over time.

5.3 Recommendation Effectiveness

We comparatively assessed the recommendation performance of ENGAGE through several evaluation metrics. Let $q \in \mathbf{Q}$ be a generic question of the test set. Assume that $\overline{\mathbf{R}}^{(q)}$ and $\mathbf{R}^{(q)}$ are, respectively, the ground-truth and the recommended list of experts for q . Essentially, $\overline{\mathbf{R}}^{(q)}$ is the list of users, who actually answered q , ranked by the known scores of their answers. Instead, the users in $\mathbf{R}^{(q)}$ are ranked by the recommendation score of Sect. 4.1. $R^{(q)} = |\mathbf{R}^{(q)}|$ is the size of $\mathbf{R}^{(q)}$. $R_i^{(q)}$ denotes the user at position i of $\mathbf{R}^{(q)}$. $R_{best}^{(q)}$ indicates the rank of the best answerer. The adopted evaluation metrics are enumerated next.

- *Precision at top $R^{(q)}$ ($Precision^{(q)}@R^{(q)}$)* [9] is the correctness of $\mathbf{R}^{(q)}$, i.e., the fraction of top- $R^{(q)}$ recommended experts, who are ground-truth answerers. More precisely,

$$Precision^{(q)}@R^{(q)} = \frac{|\mathbf{R}^{(q)} \cap \overline{\mathbf{R}}^{(q)}|}{R^{(q)}}$$

- *Recall at top $R^{(q)}$* ($Recall^{(q)}@R^{(q)}$) [9] is the coverage of $\mathbf{R}^{(q)}$, i.e., the fraction of ground-truth answers in the top- $R^{(q)}$ recommended experts. Specifically,

$$Recall^{(q)}@R^{(q)} = \frac{|\mathbf{R}^{(q)} \cap \overline{\mathbf{R}}^{(q)}|}{|\overline{\mathbf{R}}^{(q)}|}$$

- *nDCG* [10] (normalized Discounted Cumulative Gain) measures the goodness of the ranking of the recommended experts, based on their position in $\mathbf{R}^{(q)}$. This is accomplished by accumulating expert relevance to question q along $\mathbf{R}^{(q)}$, so that the relevance of higher-ranked experts is suitably discounted. Formally, $nDCG(q) = \frac{DCG(q)}{IDCG(q)}$, where

$$DCG(q) = s_1^{(q)} + \sum_{i=2}^{R^{(q)}} \frac{s_i^{(q)}}{\log_2 i}$$

In the above equation, $s_i^{(q)}$ represents the relevance of $R_i^{(q)}$ to q (according to thumbs-up/down). $IDCG(q)$ is the $DCG(q)$ value of the ideal ranking.

- *Accuracy* ($Acc^{(q)}$) [22] measures the quality of the best-answer’s rank, i.e.,

$$Acc^{(q)} = \frac{R^{(q)} - R_{best}^{(q)}}{R^{(q)} - 1}$$

Larger values of the above measures denote a higher recommendation effectiveness. Table 1 summarizes the average values of such measures over the whole set \mathbf{Q} of questions for all competitors. The reported results were found by adopting the following empirical settings. In all tests, the time decay factor λ was fixed to 0.2. The number K of latent factors was set to 15. The number S of samples used in Eq. 4 was set to 200. The overall number H of iterations for Algorithm 1 was fixed to 1,000, in compliance with the convergence-criterion in [12]. Additionally, for each $q \in \mathbf{Q}$, the number $R^{(q)}$ of recommended answers for q was set to 10.

By looking at Table 1, it is evident that ENGAGE overcomes all tested competitors. In particular, the lower effectiveness of Votes and InDegree is due to the fact that both focus only on the importance of users, without accounting for their specific discounted expertise. TER is a state-of-the-art competitor, that captures the tag-based expertise of answerers. Nonetheless, TER is still less effective than ENGAGE for two main reasons. Firstly, TER does not account for the drift of users’ interests over time. Secondly, TER does not exploit any auxiliary information from the communication network, that is shaped by the asking-answering behavior. The latter is instead conveniently used, under ENGAGE, in order to more accurately inform the latent factor representation of users and tags. TEM and CQARank are two state-of-the-art competitors, that use tags to capture topical expertise. However, their effectiveness is penalized with respect to ENGAGE,

Table 1. Recommendation effectiveness of the compared approaches

| <i>Competitor</i> | <i>Precision@10</i> | <i>Recall@10</i> | <i>nDCG</i> | <i>Accuracy</i> |
|-------------------|---------------------|------------------|-------------|-----------------|
| Votes | 0.2658 | 0.3880 | 0.7849 | 0.5618 |
| InDegree | 0.3379 | 0.4656 | 0.8235 | 0.5788 |
| TER | 0.3639 | 0.4803 | 0.8157 | 0.6161 |
| TEM | 0.3567 | 0.5066 | 0.8291 | 0.6548 |
| CQARank | 0.3896 | 0.5153 | 0.8348 | 0.6803 |
| ENGAGE | 0.4113 | 0.5379 | 0.8561 | 0.6952 |

since tags are mixed up with the textual content of posts, rather than being used as user-generated explanations of their topical expertise. Moreover, neither TEM nor CQARank discount the expertise of users, in order to account for the drift of their interests over time.

6 Conclusions and Further Research

We proposed a new latent-factor approach to expert recommendation in QA communities. The idea is to infer the time-evolving expertise of users from the tags of the answered questions, the votes and posting time of the respective answers as well as the asking-answering behavior of the CQA users. A thorough experimentation on real-world CQA data showed the overcoming recommendation effectiveness of our approach with respect to several state-of-the-art competitors.

It is interesting to explore the impact of alternative implementations of gradual forgetting on recommendation effectiveness. In this regard, temporal hyperbolic discounting [21] is a viable choice. Finally, three further lines of innovative research involve studying the incorporation of, respectively, user roles [5, 7, 13, 18], exposure [11] to posted questions as well as the recent generative models of text corpora (such as, e.g., [4]) for more effective expert recommendation.

References

1. Anderson, A., et al.: Discovering value from community activity on focused question answering sites: a case study of stack overwo. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 850–858 (2012)
2. Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)
3. Bouguessa, M., Dumoulin, B., Wang, S.: Identifying authoritative actors in question-answering forums: the case of Yahoo! answers. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 866–874 (2008)

4. Costa, G., Ortale, R.: Document clustering meets topic modeling with word embeddings. In: Proceedings of SIAM International Conference on Data Mining (2020)
5. Costa, G., Ortale, R.: Mining overlapping communities and inner role assignments through Bayesian mixed-membership models of networks with context-dependent interactions. *ACM Trans. Knowl. Disc. Data* **12**(2), 18:1–18:32 (2018)
6. DeGroot, M.: *Optimal Statistical Decisions*. McGraw-Hill, New York (1970)
7. Fu, C.: Tracking user-role evolution via topic modeling in community question answering. *Inf. Process. Manage.* **56**(6), 102075 (2019)
8. Gama, J., et al.: A survey on concept drift adaptation. *ACM Comput. Surv.* **46**(4), 44:1–44:37 (2014)
9. Herlocker, J.L., et al.: Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **22**(1), 5–53 (2004)
10. Järvelin, K., Kekäläinen, J.: Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* **20**(4), 422–446 (2002)
11. Liang, D., et al.: Modeling user exposure in recommendation. In: Proceedings of International Conference on World Wide Web, pp. 951–961 (2016)
12. Liu, J.S.: *Monte Carlo Strategies in Scientific Computing*. Springer, New York (2001). <https://doi.org/10.1007/978-0-387-76371-2>
13. Ma, Z., et al.: A tri-role topic model for domain-specific question answering. In: Proceedings of AAAI Conference on Artificial Intelligence, pp. 224–230 (2015)
14. Robert, C., Casella, G.: *Monte Carlo Statistical Methods*. Springer, New York (2004). <https://doi.org/10.1007/978-1-4757-4145-2>
15. Salakhutdinov, R., Mnih, A.: Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In: Proceedings of International Conference on Machine Learning, pp. 880–887 (2008)
16. Srba, I., Bielikova, M.: A comprehensive survey and classification of approaches for community question answering. *ACM Trans. Web* **10**(3), 18:1–18:63 (2016)
17. Wang, X., Huang, C., Yao, L., Benatallah, B., Dong, M.: A survey on expert recommendation in community question answering. *J. Comput. Sci. Technol.* **33**(4), 625–653 (2018). <https://doi.org/10.1007/s11390-018-1845-0>
18. Xu, F., Ji, Z., Wang, B.: Dual role model for question recommendation in community question answering. In: Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 771–780 (2012)
19. Yang, B., Manandhar, S.: Tag-based expert recommendation in community question answering. In: Proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 960–963 (2014)
20. Yang, L. et al.: CQArank: jointly model topics and expertise in community question answering. In: Proceedings of ACM International Conference on Information and Knowledge Management, pp. 99–108 (2013)
21. Yeniterzi, R., Callan, J.: Moving from static to dynamic modeling of expertise for question routing in CQA sites. In: Proceedings of International AAAI Conference on Web and Social Media, pp. 702–705 (2015)
22. Zhao, Z., et al.: Expert finding for community-based question answering via ranking metric network learning. In: Proceedings of the International Joint Conference on Artificial Intelligence, pp. 3000–3006 (2016)