Health Forum Thread Recommendation Using an Interest Aware Topic Model

Kishaloy Halder
School of Computing
National University of Singapore
kishaloy@comp.nus.edu.sg

Min-Yen Kan
School of Computing
National University of Singapore
kanmy@comp.nus.edu.sg

Kazunari Sugiyama
School of Computing
National University of Singapore
sugiyama@comp.nus.edu.sg

ABSTRACT

We introduce a general, interest-aware topic model (IATM), in which known higher-level interests on topics expressed by each user can be modeled. We then specialize the IATM for use in consumer health forum thread recommendation by equating each user's self-reported medical conditions as interests and topics as symptoms of treatments for recommendation. The IATM additionally models the implicit interests embodied by users' textual descriptions in their profiles. To further enhance the personalized nature of the recommendations, we introduce jointly normalized collaborative topic regression (JNCTR) which captures how users interact with the various symptoms belonging to the same clinical condition.

In our experiments on two real-world consumer health forums, our proposed model significantly outperforms competitive state-of-the-art baselines by over 10% in recall. Importantly, we show that our IATM+JNCTR pipeline also imbues the recommendation process with added transparency, allowing a recommendation system to justify its recommendation with respect to each user's interest in certain health conditions.

KEYWORDS

Recommender Systems; Collaborative Filtering; Graphical Model; Topic Models

1 INTRODUCTION

We aim to improve the visibility of the discussion threads in this chapter. People participate in online health forums in part to discuss their symptoms and clinical conditions with others. They post health related questions to learn from the experience of the community. The majority of users participate in online health communities with the goal of meeting a medical information need [10, 16, 18]. This is the problem we address in this work. We acknowledge that patients also participate for emotional support and social reasons [3, 7, 34], but this is beyond the scope of our work. Finding relevant information can be difficult, and recommendation systems can help bridge this gap by providing users with discussion threads relevant to their condition- and symptom-specific interests.

ALS Threads	Parkinson's Disease Threads
is there anyone experiencing lower back pain while standing up after sitting for a while?...	… My lower back and hips pains a lot. It even hurts to walk some days …
	Back pain has been bothering me for 4 years and it is getting worse. The lower back gets …

Table 1: A query ALS thread (left), and lexically similar but unrelated posts for Parkinson's Disease (right).

We observe that the symptoms experienced by patients with different clinical conditions are often similar. However, the proper treatment crucially depends on the underlying cause (*i.e.*, the clinical condition or disease). This leads to many lexically similar user queries which require different answers as shown in Table 1. Many traditional approaches — such as topic models — struggle to identify the correct underlying condition, as they mainly use word co-occurrence to determine relevant answers.

We observe that context is a key factor to identify the appropriate latent conditions and symptoms. In this scenario, the key contextual evidence is the participation that a user manifests with respect to a specific medical condition, either by subscribing to a subforum related to a condition or by authoring a post in the forum inferrably related to a condition[1]. We believe that such context must be accounted for in order to recommend relevant discussion threads in health forums. We introduce a two-stage approach that captures such context.

We introduce a general, interest-aware topic model (IATM), in which known higher-level interests on topics expressed by each user can be modeled. We then specialize the IATM for use in consumer health forum thread recommendation by equating each user's self-reported medical conditions as interests and topics as symptoms of treatments for recommendation. The IATM additionally models the implicit interests embodied by users' textual descriptions in their profiles. To further enhance the personalized nature of the recommendations, we introduce jointly normalized collaborative topic regression (JNCTR) which captures how users interact with the various symptoms belonging to the same clinical condition.

Our solution leverages the topic model framework to properly incorporate the contextual information. Our topic model — which we term the *interest-aware topic model* (IATM) — is a general model that encompasses both the evidence of each user's thread and word interactions, but crucially, also the user's self-reported (and thus observed) interests. A key characteristic of the IATM is that even

---

[1]In our scenario, we require actions that leave a traceable correlation with interest. This allows our framework to be applied even in cases where the recommendation is done by a third party (as done in our evaluation) and not necessarily done by the service provider.

though it can model explicit user interests (*i.e.*, a patient is subscribed to a Parkinson's disease subforum), in the absence of such explicitly indicated interests, the IATM treats users' interests as a partially-observed random variable and attempts to infer the full and latent value. As users may not explicitly type themselves, yet actively participate, this is important to account for.

The IATM also natively models the side information of user profiles. User profiles are ubiquitous to many Web 2.0 sites, inclusive of health forums. In IATM, user profiles are treated as normal documents during the training process, used in determining the interests of the user. In our health forum recommendation scenario, user descriptions do give useful information about the user, which significantly aids the recommendation process, especially for users that have little interaction history — a form of cold start. In further analyses of our datasets, we note further modeling difficulties. Even when confined to a single condition, discussion on different symptoms also often appear similar due to commonly affected parts of the body. Consider the following posts:

(a) [about back pain] "*Have been suffering from back pain for couple of years now and it's getting worse -now I can barely put **weight** on the right **leg**...*"

(b) [about leg cramps] "*I get a lot of cramps in the **leg**. The only remedy is to stand up and put the body **weight** on it...*"

These posts have common words (bolded) but are about different symptoms of Parkinson's disease. Although the topic distributions in two posts (documents) are similar, each user's preferences are clearly directed towards different particular topics. We observe similar distribution bias with users' participation in clinical treatment discussions and other condition-specific topics.

To address this second, fine-grained disambiguation problem, we develop a novel graphical model, jointly normalized collaborative topic regression (JNCTR). JNCTR is a logical adaptation of the original collaborative topic regression (CTR) model [32], itself motivated to handle such divergences in each user's interests in documents with similar topic distributions. JNCTR extends CTR by taking both the user–topic and thread–topic distributions coming from IATM as input, but additionally accounts for the user–thread interaction history in the form of ratings. This model allows us to understand the differences between symptoms that originate from a single condition. JNCTR updates both the user–topic and thread–topic distributions based on the past user–thread interactions. We compute thread recommendations for each user using the resultant user–topic and thread–topic distributions.

Online health forum users often use their own words and phrases to describe their experiences [12, 20]. Standard medical ontologies and thesauri (*e.g.*, UMLS[2]) struggle to cover the medical terms found in user–generated medical content [9]. We believe that our specialized IATM+JNCTR model is the first attempt to understand how clinical conditions and their symptoms and treatments explain the interaction of users in a health forum.

In our experiments on two real-world consumer health forums, our proposed model significantly outperforms competitive state-of-the-art baselines by over 10% in recall. Importantly, we show that our IATM+JNCTR pipeline also imbues the recommendation process with added transparency, allowing a recommendation system to justify its recommendation with respect to each user's interest in certain health conditions.

The contributions of our work are summarized as follows:

• We formalize the problem of interest-aware recommendation, of which health forum thread recommendation is a specific instantiation of *conditions-as-interests*. We investigate how to best utilize user participation in the forum, formulating this as an implicit feedback-based recommendation problem.

• We apply our framework to a real-world dataset obtained from HealthBoards[3], demonstrating significant improvement over state-of-the-art baselines.

• We extend our experiments to demonstrate how our proposed IATM+JNCTR model deals gracefully with cold-start items ("threads" in our work). The model can explain a recommendation due to its modeling of latent variables. We further investigate how our model performs in recommendation justification by analyzing its recommendations to specific users.

2 RELATED WORK

The motivations behind user participation in consumer medical information discussion forums have been systematically studied by the medical informatics community in recent years [3, 14, 34]. However, to the best of our knowledge, such work has been limited to large-scale surveys of self-reported behavior, and the community has not seen much development of practical recommendation techniques for health forums as of current.

While acknowledging the varied societal and emotional support needs of users, we find value in addressing the primary information need for consumer health information. As such, our task falls into the guise of recommendation systems, an area which has seen much recent interest with the popularity of Web 2.0 systems that integrate users and items into Web applications. For brevity, we limit our discussion to relevant prior work in the areas of topic modeling, content-based and context-aware recommendations, and community question answering.

Topic Models. Topic models regard documents as mixtures of latent topics with certain distributional properties. For textual documents, several works have focused on modeling latent factors of the content using latent Dirichet allocation (LDA) [4] and its variants [17, 26, 28]. For example, the author–topic model [28] learns the topic distribution of authors for a set of documents. On the other hand, labeled LDA relies on annotated tags to constrain the possible topics for each document [26].

While these models are useful on their own for modeling either users or items, they do not capture the dynamics between both. LDA can be used as a starting point for refinement to account these factors. Agarwal *et al.* [1] leveraged LDA-discovered latent topic distributions for matrix factorization-based collaborative filtering (CF). They report modest improvement over other methods — the reason being that often the topic distributions of different items look similar, even though they appeal to different sets of people. There also exists a set of focused topic models that cater to specific use cases [6, 35]. Chen *et al.* [6] proposed a Contextual Focused Topic Model, where they assume a word to be generated from either

the author or the venue or the document characteristics – not from a joint combination of them, as in our case.

Content-based and Context-aware Recommendation. Certain content-based recommendation systems further account for information associated with content associated with users. Wang *et al.* [32] proposed collaborative topic regression (CTR) and showed the effectiveness of adjusting the LDA-discovered topic distributions by doing a regression from the observed user–item ratings in scholarly documents. In the same context, Charlin *et al.* [5] showed that the cold-start performance of a similar model can be improved if there is bootstrapping information available in the form of document content associated with users. Although it is possible to use a textual content agnostic off-the-shelf CF method to recommend articles to a user to comment on [30], considering associated textual content improves performance significantly in other platforms, *e.g.,* news articles or blogs [2], demonstrating the efficacy of modeling such side information.

Along with the past user–item interaction history, *Context Aware Recommendation* methods consider the interaction contexts which can be equated to the medical conditions of a user in our scenario. Tensor Factorization [11] and Factorization Machines [27] are two promising methods, primarily designed to predict ratings in an explicit feedback-based system. Nguyen *et al.* [19] demonstrate that such techniques can also be profitably applied in implicit feedback scenarios such as ours.

Recommendation in Community Question Answering. In community question answering systems, prior work has addressed recommending semantically related question threads that reflect different aspects of the user's query and provide supplementary information. Wang *et al.* [33] recommend more relevant threads by extending a language model with the popularity of a question. Pedro and Karatzoglou [24] extend Learning to Rank to supervised LDA applied specifically to recommend relevant question threads. Zhou *et al.* [39] propose a translation model-based thread recommendation by incorporating answer information. In recent work, Omari *et al.* [22] and Palotti *et al.* [23] improve ranking of relevant discussion threads in health forums. However, both works do not address the recommendation of relevant threads to specific user's interests.

While the previous work can handle recommendations in health forums, there is important evidence that needs to be modeled to achieve better recommendation accuracy [5, 15, 37]. In particular, in health forums, each user can often express explicit interests in different conditions. We note that, in most of existing recommendation systems, construction of user profiles has been independent of the recommendation process itself.

We propose a unified framework of both user profiles and user participation in a health forum. This leads to not only significant improvement over other state-of-the-art methods but also enhancement of transparency in the recommendation task.

3 METHODS

Our recommendation methodology takes full advantage of the different sources of evidence that influence recommendations of items in a generic context. It is a three-stage methodology comprising of:

1) a topic model (IATM), 2) topic regression (JNCTR), and 3) ranking combination, as shown in Figure 1. We first give a short overview of the first two key models before describing how we instantiate them for the health forum recommendation task. We then describe the three stages in technical detail, and finally discuss our instantiation of the model for health forum thread recommendation to create a condition-aware topic model.

Method Overview. Our proposed interest-aware topic model is a generic topic model that can be used in many recommendation scenarios involving users. Without loss of generality, IATM assumes that *users* interact with *documents* (or *items*, as in the literature). The interactions generate some textual evidence that ties users and documents together — such as contributing a post within a larger, multi-user thread (document), commenting or authoring the entire document, such as one's own user profile. Like the standard topic model, documents are modeled as mixtures of *topics*; however, a key distinction in IATM is that it assumes that topics are related to certain higher-level *interests* in a generative relationship. IATM captures explicit expressions of user *interests*, but crucially maintains the observations of these interests only as partially observed. This distinction allows the IATM to infer other interests of the user that are suggested by the contextual evidence of the user's other interactions.

JNCTR advances this step further, taking in the output of IATM's user–topic and document–topic distributions and further accounting for user–document interactions. As IATM already accounts for interests, we can instantiate standard collaborative topic regression for each interest separately, and jointly normalize them to output refined user–topic and document–topic distributions. These are then fused to generate recommendations.

Instantiating the Model for Health Forum Recommendations. IATM+JNCTR can be applied to various Web 2.0 contexts — recommendation tasks such as ones for movies, products, and discussion forums. These contexts all have document–user interactions, where user interests are partially observed through forum subscriptions or folksonomy tags, among other means.

For clarity, we now instantiate IATM+JNCTR for the health forum thread recommendation problem. In our scenario, users express their interests by subscribing to forums at health websites, which are largely specific to a medical condition. As in the general case, we do not expect users to necessarily subscribe to all the condition-specific forums that are relevant to them; we model such subscriptions as being partially observed.

The goal of our recommendation system is to recommend relevant health forum threads to users. Users can participate in forum threads by contributing posts, which forms the user–document interactions in our IATM+JNCTR framework. A user can report her clinical conditions as part of her user profile's free text description (*e.g.*, "About Me"). Such user documents are only used as evidence during training; to be clear, we do not recommend user profiles. Finally, individual threads on a particular condition discuss different symptoms and treatments in differing proportions. We assume that users are interested in certain symptoms that they experience, and treatments that they are undergoing.

In our health forum thread recommendation, we equate the following IATM terms with ones specific to our scenario: interest
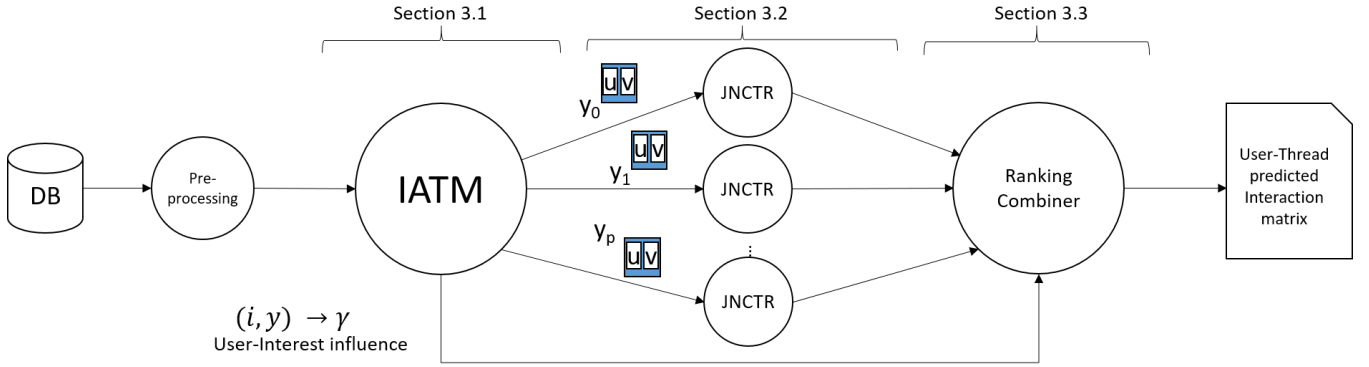
Figure 1: The pipeline for our three-stage recommendation framework. IATM first provides user–topic and document–topic distributions for all the interests, and then JNCTR further optimizes those distributions. Finally, the ranking combiner merges the ranked documents depending on the user–interest alignment estimated from the first stage.
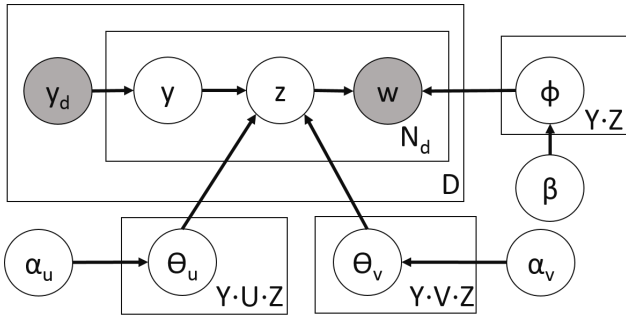

Figure 2: Plate notation for our interest-aware topic model (IATM). We observe the document words as well as partially observe the interests that select the topics for the words in the document (gray nodes). Topics are dependent on the interest (y), user (θ_u), and thread document (θ_v).

$\rightarrow$ condition; topic $\rightarrow$ (symptom, treatment); and document $\rightarrow$ thread.

3.1 Interest-Aware Topic Model (IATM)

We use the standard plate notation for the graphical model as shown in Figure 2. There are U users and V thread documents. Since each user has a user document (*i.e.,* a user profile), there are U user documents; hence we have altogether $D = U + V$ documents. Y denotes the set of all possible interests. In Figure 2, an interest y is sampled from a uniform distribution from the set of interests $y_d \subset Y$, where y_d is the union of all interests reported by the users participating in document d. Each interest y has Z latent topics which denote the fine-grained sub-topics of an interest (*e.g.,* in our instantiated IATM for medical conditions, they would be different symptoms or medications for a condition).

For each word in a document, a latent topic z is sampled from an interest y according to the topic distribution of the user θ_u, as well as the topic distribution of the thread θ_v. The reason behind this approach is intuitive: when a user contributes to a thread document,

the topic of the user's words are dependent on the overall thread topic as well as the user's own set of interests. However, in the case of a user document (*i.e.,* the user profile), the choice of topic is only dependent on the user's own interests. A topic z is sampled only from the interest y and θ_u, for such user documents.

A word w is sampled from z and the word–topic distribution ϕ. We invoke blocked Gibbs sampling as the exact inference of the full posterior is intractable. The inference process is similar to the author–topic model [28]; but in IATM, the author of a word is observed. We have two sets of latent variables, z and y. We draw each (z, y) pair as a block, conditioned on all other variables:

$$P(z_x = h, y_x = k | w_x = m, \mathbf{z_{-x}}, \mathbf{y_{-x}}, \mathbf{w_{-x}}, \mathbf{y_d}) \propto$$

$$(q_1 \frac{n_{hk}^i + \alpha_u}{\sum_z n_{zk}^i + Z\alpha_u} + q_2 \frac{n_{hk}^d + \alpha_v}{\sum_z n_{zk}^d + Z\alpha_v}) * \frac{n_{mh}^k + \beta}{\sum_w n_{wh}^k + W\beta}, \quad (1)$$

where $z_x = h$ and $y_x = k$ denote that the xth word in d^{th} document is assigned to topic h under interest k; $w_x = m$ represents that x^{th} word is the mth word in the vocabulary; $\mathbf{z_{-x}}$ and $\mathbf{y_{-x}}$ represent all topic and interest assignments not including the xth word; n_{hk}^i is the number of times topic h is assigned with interest k for user i, not including the instance under consideration; and W is the total number of unique words in the vocabulary. Similarly, n_{hk}^d represents the number of times topic h has appeared under interest k in the dth document; and n_{mh}^k denotes the number of times the mth word in the vocabulary has appeared in topic h under interest k — excluding the current instances in all the cases. The three factors in Equation (1) represent the random variables θ_u (probability of topic given interest and user), θ_v (probability of topic given interest and thread), and ϕ (probability of a word given interest and topic). The Dirichlet priors for these three distributions are α_u, α_v, and β, respectively. We use a Dirichlet mixture of the two individual Dirichlet densities (θ_u, θ_v) as the prior [25, 31], giving equal weights to the mixture coefficients (*i.e.,* $q_1 = q_2 = 0.5$). We also learn the user–interest distribution γ. These distributions are estimated from

the samples using the following equations:

$$\theta_u^{hik} = \frac{n_{hk}^i + \alpha_u}{\sum_z n_{zk}^i + Z\alpha_u}, \theta_v^{hdk} = \frac{n_{hk}^d + \alpha_v}{\sum_z n_{zk}^d + Z\alpha_v}, \quad (2)$$

$$\phi^{mhk} = \frac{n_{mh}^k + \beta}{\sum_w n_{wh}^k + W\beta}, \quad (3)$$

$$\gamma^{ik} = \frac{n_k^i}{\sum_y n_y^i}, \quad (4)$$

where n_k^i is the number of times interest k is sampled for user i.

Once the distributions are learned, we create sub-spaces of the entire user–thread interaction matrix based on each interest. The interaction matrix R^k for interest k is defined by:

$$r_{ij}^k = \begin{cases} 1 & \text{if } R_{ij} = 1, k \in Y_i \\ 0 & \text{otherwise,} \end{cases}$$

where Y_i is the set of interests for user i, $R_{ij} = 1$ if user i participated in thread j; 0 otherwise. Similarly, we define the user–, thread–, and word–topic distributions for this sub-space as θ_u^k, θ_v^k, and ϕ^k, respectively:

$$\theta_u^k = \theta_u\{k\}, \theta_v^k = \theta_v\{k\}, \phi^k = \phi\{k\}.$$

As an example, given the three threads in Table 1, IATM places the left one in the *ALS* sub-space, and the right ones inside the *Parkinson's disease* sub-space.

3.2 Joint Normalized Collaborative Topic Regression (JNCTR)

We treat each of the resultant interest-specific user–thread sub-spaces originating from IATM as a separate problem instance and optimize them individually using JNCTR as shown in Figure 1.

Figure 3 shows the plate model for each individual instance of JNCTR. Here, we use the notations θ_u, θ_v, ϕ, and R without the interest-specific superscript k. I and J denote the set of users and threads within this sub-space, respectively. Note that we omit the plate for word generation as we do not assume any particular generative process for them, and CTR does not re-sample topics once θ_v is obtained from the topic model as discussed in [32]. As in CTR, we introduce a latent variable ϵ_u^i that offsets the topic proportions θ_u^i for i^{th} user when modeling the user's ratings. JNCTR assumes that there are Z topics both in user content and thread content $\beta = \beta_{1:Z}$. The generative process of JNCTR consists of the following steps:

(Step 1) For each user i, draw user latent offset $\epsilon_u^i \sim \mathcal{N}(0, \lambda_u^{-1}\mathbb{I}_Z)$ and set the user latent vector as: $u_i = \epsilon_u^i + \theta_u^i$,

(Step 2) For each thread j, draw thread latent offset $\epsilon_v^j \sim \mathcal{N}(0, \lambda_v^{-1}\mathbb{I}_Z)$ and set the thread latent vector as: $v_j = \epsilon_v^j + \theta_v^j$,

(Step 3) For each user–thread pair (i, j), draw the rating as: $r_{ij} \sim \mathcal{N}(u_i^T v_j, c_{ij}^{-1})$.

where $\mathbb{I}_Z$ is Z-dimensional identity matrix; λ_u and λ_v are the regularization parameters; c_{ij} is the precision parameter for r_{ij}, a confidence parameter for rating r_{ij}, where larger values denote higher trustworthiness. This is important in the case of implicit feedback-based systems like ours (note that $r_{ij} = 0$ denotes either
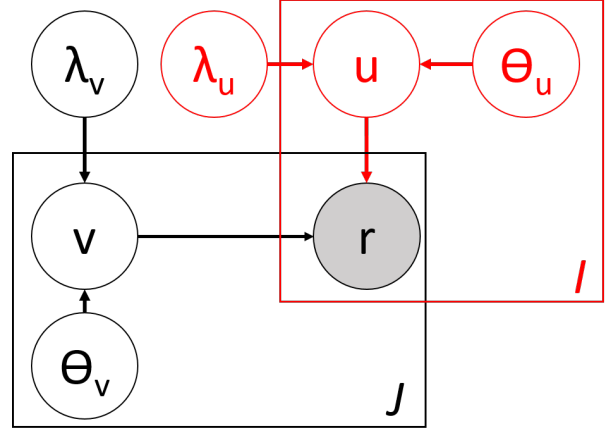

Figure 3: Plate notation for jointly normalized collaborative topic regression (JNCTR). Components in black are from collaborative topic regression (CTR [32]). Components in red are introduced for user modeling. Note that both plates for user and thread are form-identical.

that the ith user is not interested in the jth thread or the user is unaware of it). We set $c_{ij} = a$, if $r_{ij} = 1$, otherwise we set it to b, where a and b are tuning parameters satisfying $a > b > 0$. We discuss parameter tuning in Section 4.

Learning the Parameters for JNCTR. Given topic parameter β, computing the full posterior of u_i, v_j, θ_u, θ_v is intractable. We need to develop an EM-style algorithm to learn these parameters. Extending the posterior mentioned in [32], given λ_u, λ_v, and β, the complete log likelihood $\mathcal{L}$ of U, V, $\theta_u^{1:I}$, $\theta_v^{1:J}$, and R is defined as follows:

$$\mathcal{L} = -\frac{\lambda_u}{2}\sum_i (u_i - \theta_{u^i})^T(u_i - \theta_{u^i}) - \frac{\lambda_v}{2}\sum_j (v_j - \theta_{v^j})^T(v_j - \theta_{v^j})$$
$$+ \sum_i \sum_m \log(\sum_k \theta_{u^{ik}}\beta_{k,w_{im}}) + \sum_j \sum_n \log(\sum_k \theta_{v^{jk}}\beta_{k,w_{jn}})$$
$$- \sum_{i,j} \frac{c_{i,j}}{2}(r_{ij} - u_i^T v_j)^2. \quad (5)$$

We optimize this likelihood function by coordinate ascent, optimizing the CF variables u_i, v_j iteratively. To update u_i and v_j, we take the gradient of $\mathcal{L}$ with respect to u_i and v_j and set it to zero. This yields:

$$u_i \leftarrow (VC_iV^T + \lambda_u I_K)^{-1}(VC_iR_i + \lambda_u\theta_u^iR_i), \quad (6)$$

$$v_j \leftarrow (UC_jU^T + \lambda_v I_K)^{-1}(UC_jR_j + \lambda_v\theta_v^jR_j), \quad (7)$$

where $U = (u_i)_{i=1}^I$, $V = (v_j)_{j=1}^J$, C_i is a diagonal matrix with c_{ij} $(j = 1, \ldots, J)$ as its diagonal elements and $R_i = (r_{ij})_{j=1}^J$ for user i. C_j and R_j are similarly defined for thread j.

Prediction. Once the locally optimal parameters U^*, V^*, θ_u^*, θ_v^* are learned, JNCTR can predict ratings. Given that D is the observed data, the prediction is estimated as:

$$\mathbb{E}[r_{ij}|D] \approx (\mathbb{E}[\theta_{U_i}|D] + \mathbb{E}[\epsilon_{u_i}|D])^T \cdot (\mathbb{E}[\theta_{V_j}|D] + \mathbb{E}[\epsilon_{v_j}|D]). \quad (8)$$

As shown in Figure 4, we address the prediction of ratings as follows:

(Fig. 4a) In-matrix prediction: All the users and threads have at least one interaction. We use the point estimate of $\theta_{u^i}, \theta_{v^j}, \epsilon_u^i$, and ϵ_v^j to approximate the expectations:

$$r_{ij}^* \approx (\theta_u^{i*} + \epsilon_u^{i*})^T (\theta_v^{j*} + \epsilon_v^{j*}) = (u_i^*)^T v_j^*. \qquad (9)$$

(Fig. 4b) Out-of-matrix prediction (for a *thread*): Some threads do not have interaction history; *i.e.*,

$$r_{ij}^* \approx (\theta_u^{i*} + \epsilon_u^{i*})^T (\theta_v^{j*}) = (u_i^*)^T \theta_v^{j*}. \qquad (10)$$

(Fig. 4c) Out-of-matrix prediction (for a *user*): Some users do not have interaction history; *i.e.*,

$$r_{ij}^* \approx (\theta_u^{i*})^T (\theta_v^{j*} + \epsilon_v^{j*}) = (\theta_u^{i*})^T v_j^*, \qquad (11)$$

substituting u_i and v_j from Steps 1 and 2 in the JNCTR generative process. We thus obtain a ranked list of interest-specific threads that is recommended to a user.

Efficiency. Note that in Figure 1, multiple instances of JNCTR are run, but each instance is run on a partition of the full matrix. The computational complexity of JNCTR is comparable to the original CTR algorithm; the number of updates in both are identical.

3.3 Fusing the Final Ranked List

Once we obtain the condition–specific optimized user- and thread-topic distributions, we combine them into a single ranking using the γ distribution defined by Equation (4). We explore three methods to fuse the individual, interest-specific lists:

1. Proportional Selection. For every interest that user i is interested in, we prepare a list of threads (in descending order of predicted score) that the user might find interesting. We select the top-M threads from each condition sub-space according to user i's γ distribution. For illustration, say John has a γ distribution of {Multiple Sclerosis : 0.8, Asthma : 0.2}. Then when recommending threads to John, 80% are chosen from the top threads in Multiple Sclerosis, and the remaining 20% come from Asthma.

2. Combined Score-Based Selection. Here, we obtain a single ordered list of threads from their combined score for each user. First, we normalize scores in each interest sub-space between [0, 1]. For user i, the combined score of a thread j is defined by Equation 12 which ranks the items in descending order of their total score:

$$R_{ij}^* = \sum_y \gamma^{iy} \times r_y^{ij}, (y \in Y_i). \qquad (12)$$

3. Maximum Split Preference-Based Selection. This is similar to the binary preference based merging in [36]. For each user i, we only consider the condition with the highest preference γ score.

We note that although the proportional selection strategy makes the recommended list of threads more diverse in nature, the combination score-based selection obtains superior results on our datasets.

4 EXPERIMENTS

To answer important questions about our model, we consider specific experimental settings. In the following, we describe the main results of our study after detailing the datasets, evaluation metrics,

Dataset	# Users	# Threads	# Posts	Avg P:T	# Distinct Conditions	Avg C:U
HBD	127,903	155,863	716,744	4.6	235	4.01

Table 2: Statistics on our health forum dataset. "Avg P:T" and "Avg C:U" denote the average number of posts in a thread and conditions reported by a user, respectively.

and baselines.

Dataset. We constructed following large real-world consumer health forum dataset to validate our model:

HealthBoards (HBD)[2]: We use the publicly available Health-Boards dataset[4]. We collate all the posts made by a user and treat them as the user document. We ignore generic categories of threads such as 'Family', 'Support', 'Healthcare' and 'General'.

We remove all stop words and select the top 8,000 words based on TF-IDF scores. The TF-IDF computation was done only on the training data for all the experiments. Similar to other recommendation works, we remove users with few interactions, namely, with less than three thread interactions. Table 2 shows some statistics on our datasets and their user reported conditions.

Metrics. Similar to [32], we do not rely on precision, as our ground truth is only implicit feedback. Samples with negative values could be threads that the user had not seen (but would have been interested in), as well as those where the user explicitly did not interact with. As such, we use three metrics to assess recommendation quality:

Recall@M considers how many top-M threads were actually interacted by the user (higher is better). The recall for the entire system can be summarized as the average recall value for all users.

Mean Reciprocal Rank (MRR) indicates where in the ranking the first relevant thread is returned by the system, averaged over all users. This measures the ability of the system to return a relevant thread at the top of the ranking. Let r_i be the rank of the highest ranking relevant thread for a target user i, then MRR is just the reciprocal rank, averaged over all target users, N_U:

$$MRR = \frac{1}{N_U} \sum_{i=1}^{N_U} \frac{1}{r_i}.$$

Normalized Discounted Cumulative Gain (nDCG) is well suited for evaluation of recommendation system, as it rewards relevant threads in the top ranked results more heavily than those ranked lower. nDCG is computed as:

$$nDCG_i = Z_i \sum_{j=1}^{M} \frac{2^{r(j)} - 1}{log(1 + j)},$$

where Z_i is a normalization constant calculated so that a perfect ordering would obtain nDCG of 1; and each $r(j)$ is an integer relevance level (for our case, $r(j) = 1$ and $r(j) = 0$ for relevant and irrelevant recommendations, respectively) of result returned at the rank j ($j = 1, \cdots, M$). Then, $nDCG_i$ is averaged over all our target users. in this work, we use $nDCG@M$ ($M = 5, 10$) for evaluation where
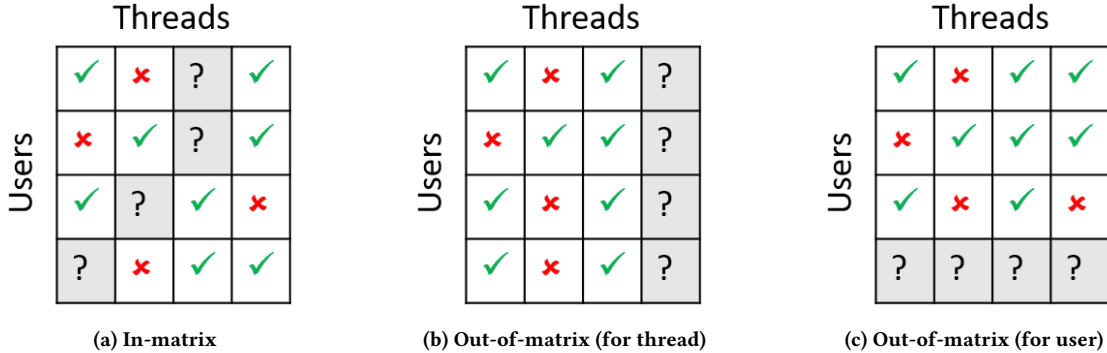
---

[4]http://resources.mpi-inf.mpg.de/impact/peopleondrugs

(a) In-matrix   (b) Out-of-matrix (for thread)   (c) Out-of-matrix (for user)

Figure 4: Illustration of three prediction tasks for our thread recommendation system. "$\checkmark$," "$\times$," and "?" denote "like," "dislike," and "unknown," respectively.

M is the number of top-M threads recommended by our approaches.

Baselines. We compare our instantiated IATM+JNCTR with six baselines. Our complete model uses four signals overall: the user–thread interaction history, textual content of threads, user profiles, and the user–reported conditions. We chose baselines for their competitiveness, recency, and use of particular signals common to our model. Comparing among the various models can also be seen as assessing how important each form of evidence is in achieving quality recommendation. Table 3 summarizes how the baselines account for some subsets of the evidence in our model.

1. *Collaborative Filtering (CF)*: This is the non-negative matrix factorization-based (NMF) method for collaborative filtering of [13].

2. The *Author-Topic Model (AT)* learns the author–topic distributions [28]. We estimate the thread–topic distributions from the learned word–topic distributions. These can be interpreted as the user and thread latent topic matrices for our task, respectively. We empirically set the hyperparameters $\alpha = 0.1$, $\beta = 0.01$, and the number of iterations and latent topics to 2,000 and 50, respectively.

3. *Collaborative Topic Regression (CTR)* [32] is the basis for JNCTR, but without the individual interest- (condition-) specific instances, and with LDA (instead of IATM) as the input. We tune the parameter settings, $b = 0.01$, $\lambda_u = 0.01$, $\lambda_v = 0.1$ to yield its optimal results.

4. *IATM alone* can also generate recommendations. Unlike the configuration described in Section 3.1, we consider the entire user- and thread-topic matrices obtained from IATM, and use them for the recommendation task. We empirically set the hyperparameters β=0.1, and α_u=α_v=5, and set the number of topics for each medical condition to 3 and the number of iterations to 2,000.

5. *Context Aware Recommendation (CAR)* uses Factorization Machines for the recommendation [27]. We use the libFM package (http://www.libfm.org). We create the test set following the sampling policy described in [19]. This models user–reported conditions as the context for each interaction.

6. *AT + JNCTR*: In this experiment, we replace the first stage of our pipeline with Author–Topic model (AT) [28]. This model

Method	User-Thread Interaction	User Docs	Thread Docs	User Reported Conditions
1. CF (NMF) [13]	$\checkmark$			
2. AT [28]		$\checkmark$	$\checkmark$	
3. CTR [32]	$\checkmark$		$\checkmark$	
4. IATM		$\checkmark$	$\checkmark$	$\checkmark$
5. CAR [27]	$\checkmark$			$\checkmark$
6. AT + JNCTR	$\checkmark$	$\checkmark$	$\checkmark$	
7. IATM + JNCTR	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$

Table 3: Signals considered by comparative methods.

Method	MRR	nDCG	
		@5	@10
1. CF (NMF) [13]	0.179	0.180	0.194
2. AT [28]	0.023	0.033	0.036
3. CTR [32]	0.186	0.178	0.193
4. IATM	0.113	0.059	0.064
5. CAR [27]	0.092	0.081	0.087
6. AT+JNCTR	0.213*	0.221*	0.254*
7. **IATM+JNCTR**	**0.327***	**0.329***	**0.361***

Table 4: MRR and nDCG scores obtained by in-matrix prediction. "*" denotes the difference between the best baseline ("3. CTR") and our methods ("6. AT+JNCTR") and ("7. IATM+JNCTR") is significant for $p < 0.005$.

directly contrasts with our full model to see the comparative difference when using IATM over AT.

For our IATM+JNCTR model, we obtain the optimized matrices from the second stage of the pipeline and obtain the final prediction after combining the ratings as described in Section 3.3. We keep the optimal settings for IATM. For JNCTR, we empirically optimize the hyperparameters, $\lambda_u = 0.01$, $\lambda_v = 0.1$, and $b = 0.1$ which are estimated from grid search.

In-Matrix Setting: We report results using 5-fold cross validation. We split users with 5+ threads into a training set (80%) and a test
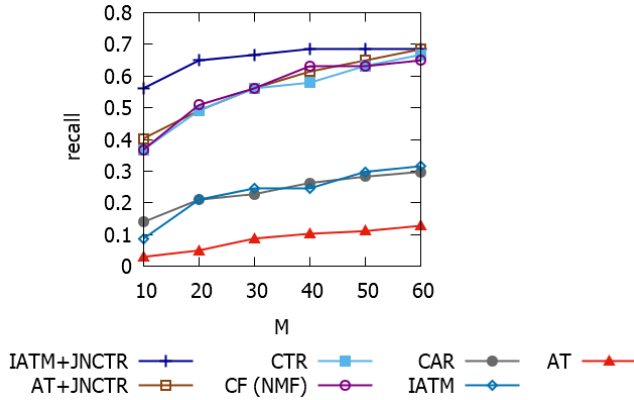
Figure 5: Recall scores at various M top ranks for the HBD dataset.

Method	Recall	MRR	nDCG	
	@60		@5	@10
1. CF (NMF) [13]	0.483	0.166	0.136	0.160
2. AT [28]	0.114	0.021	0.027	0.032
3. CTR [28]	0.517	0.211	0.178	0.205
4. IATM	0.286	0.142	0.055	0.064
5. CAR [27]	0.348	0.183	0.132	0.152
6. AT+JNCTR	0.549*	0.256*	0.176	0.202
7. IATM+JNCTR	**0.674***	**0.340***	**0.289***	**0.318***

Table 5: Recall@60, MRR, and nDCG@{5, 10} scores for temporal prediction. "*" denotes the difference between the best baseline (Row 3) and our methods (Rows 6–7) are significant for $p < 0.005$.

set (the remaining 20%). Users with fewer than five interactions always appear in the training set. For these main results, we use a *warm-start* (*i.e.,* in-matrix) setting, ensuring that each user or thread in the test set has been observed at least once in the training set.

Temporal Setting: Apart from the 5-fold cross validation, where a randomly selected 20% split is used as test, we also report results for a temporal experiment. In this setting, for each user, the last 20% of her interacted threads are kept for test. Similar to the in-matrix setting, we make sure that all the user and threads appear at least once in the training data.

Results. Figure 5 shows the recall@M ($M = 10, 20, \ldots, 60$) for comparative methods for the in-matrix setting. In both datasets, the IATM+JNCTR pairing achieves the highest recall. CTR and CF (NMF) give comparable performance, which is consistent with [32]. We observe that, in the HBD dataset, while the recall scores converge when $M \geq 60$, our IATM+JNCTR method outperforms the others with a significant margin for lower values of M (*i.e.,* more important ranks). This indicates that our pipeline can rank relevant items high in the recommendation list. This phenomena is also depicted in Table 4 where we present the MRR, nDCG@{5, 10} scores for all methods. The AT model alone works poorly in both datasets, indicating that it is insufficient to consider only user and thread documents. However, when paired with JNCTR (namely, AT+JNCTR), AT significantly improves recommendation accuracy, which factors in the user–item interaction. With IATM+JNCTR performing best, we conclude that considering the user documents and the user reported interests enhances the user–item interaction history for best recommendation accuracy. We perform a paired t-test to verify whether the obtained results are statistically significant or not. As shown in Table 4, we observe that, in the HBD dataset, our full IATM+JNCTR outperforms all the other baselines in both MRR and nDCG@{5, 10}.

Table 5 shows the recall@60, MRR, and nDCG@{5, 10} scores in the temporal setting. We find a similar trend in recall scores at various top ranks compared to the in-matrix setting. We observe that

IATM+JNCTR outperforms the others with statistical significance in this setting. Unlike the in-matrix setting, AT+JNCTR achieves statistically significant improvement only in the HBD dataset in MRR and recall@60. This indicates the robustness of our model in both of the randomized 5-fold and temporal settings.

5 DISCUSSION

Aside from the main in-matrix results, there are several important research questions that merit deeper investigation. As shown in Figure 4, our IATM+JNCTR pipeline handles cold start by incorporating prior knowledge. Our research questions (RQ) are:

RQ1: How does it perform with cold-start *documents* (*i.e.,* newly-introduced threads)?

RQ2: How well can the IATM+JNCTR pairing explain its recommendations?

RQ3: How well does the IATM+JNCTR pairing capture users' interests for specific symptoms and treatments?

RQ4: Does it actually recover the users' implicit interest in specific conditions?

In the following, we answer each of these RQs.

RQ1: Out-of-matrix Thread Recommendation. It is important for a newly-posted thread (usually some form of question) to receive quality answers. To simulate this, we partition all threads evenly among five folds. For each fold, we form a submatrix from the threads which are not within this fold and the corresponding users. We treat this submatrix as training data and learn user–topic and thread-topic distributions from the same. We ensure that none of the in-fold threads occurs in the training data. In the test phase, for all the in-fold threads, we consider the textual content of the query and user profile (*i.e.,* "About Me" text and user-reported conditions) of the user to estimate the topic distributions from the model learned during training. The task is to correctly predict the set of in-fold threads each user will actually interact with.

Table 6 shows the MRR and nDCG@{5, 10} obtained by the relevant comparative methods for this out-of-matrix thread recommendation task. IATM+JNCTR again achieves the best scores. In general, performance degrades compared to the in-matrix setting shown in Table 4, due to the harder nature of the task. It is

Method	MRR	nDCG	
		@5	@10
2. AT [28]	0.025	0.029	0.036
3. CTR [32]	0.131	0.098	0.114
4. IATM	0.094	0.068	0.081
6. AT+JNCTR	0.164*	0.146*	0.172*
7. **IATM+JNCTR**	**0.221***	**0.234***	**0.263***

Table 6: MRR and nDCG scores obtained for out-of-matrix (thread) recommendation. "*" denotes statistical signficance between the best baseline (Row 3) and our methods (Rows 6 and 7) at $p < 0.005$. CF and CAR do not work in this setting.

Recommended Thread	Candidate Conditions
I have been suffering from lupus lately having red dots all over my face. Anyone else having the same?	**1. Systemic Lupus Erythematosus** 2. Spinal Stenosis
Suffering from degenerative spine, spinal stenosis, severe scoliosis..	**1. Spinal Stenosis** 2. Systemic Lupus Erythematosus
Is there anyone with ALS who doesn't catch colds anymore?	**1. ALS** 2. Dysautonomia
has anyone stopped rytary and gone back to stalevo or something like that?	**1. Parkinson's Disease** 2. Vitamin B12 Deficiency

Table 7: Recommended threads for sample users. The explaining condition chosen by IATM+JNCTR is bolded.

interesting that prior evidence in the form of user profiles and the user-reported conditions significantly help, in the absence of observed user–item interactions. Examining the relative performance of IATM, CTR, and AT+JNCTR, we argue that the user-reported conditions further improve recommendation accuracy, compared against the evidence from user profiles.

RQ2: Transparency of Recommended Threads. While CF-based recommendation algorithms work well in terms of prediction accuracy, their latent factors make it difficult to justify the recommendation to the user [8, 38]. User participation (measured through metrics like clickthrough rate) alleviates this difficulty if items are recommended to a user with semantic explanation. Our IATM+JNCTR adds transparency by providing users with the context when recommending a thread. We learn the user–condition distribution (γ) in the first phase of our model, which is used downstream to combine recommended threads from different condition-specific sub-spaces. While presenting the combined list of threads to the user, the recommendation system can disclose which sub-space a particular thread belongs to. When a thread exists in multiple sub-spaces, we can select condition $c = \arg\max_c \gamma_{ic}$ and present it as the context recommending it to a user i, as "recommended due to your interests in c". Table 7 presents sample threads recommended by our pipeline. Note that it can identify the context (*i.e.*, the condition) for recommending the threads.

RQ3: Significance of Discovered Symptoms and Treatments. A challenge in achieving quality recommendation is to appropriately learn the topics even when overlapping words appear among several conditions. Our IATM+JNCTR leverages the user-reported conditions and learns the appropriate word distribution. Table 8 shows the top words discovered by our model for several conditions. Note that, while there are few common words across conditions, — *ALS*, *Epilepsy*, and *Multiple Sclerosis* all list *brain* among top keywords — our method can distinguish among these conditions.

It is important to analyze the condition-specific topics learned by our approach. Since different users express various levels of interest towards particular aspects (symptoms or treatments) of a condition, it is necessary to capture these aspects to achieve quality recommendation. Table 9 presents some condition-specific topics discovered by our pipeline. From simple observation, one can see correspondences for *Diabetes*, Topic 0 lists affected body parts and associated difficulties, Topic 1 discusses diets, and Topic 2 relates to human physiology — having words such as *blood, insulin*. In the scenario where a user has *Diabetes* and is interested in managing the condition through her diet, our model can recommend threads that would match her interests at this topical level.

In the case of serious terminal diseases, such as *cancer*, psychological and spiritual words, such as *god* and *luck*, appear in the top words as topics. Consider the following posts by cancer patients:

"*Would love to talk to anyone with ovarian cancer..really believe faith can play a huge role in recovering and also positive attitude..I wish this disease didn't exist*"

"*Sending all my positive energy to you..Feel free to reach out if you need anything*".

Users with similar conditions often participate in health forums for such emotional support rather than informational need [21, 29, 34]. Our model can capture this phenomenon as a topic for certain conditions.

RQ4: Predicting Implicit Conditions. In IATM, recall that we sample both interest (condition) and topic (symptom or treatment) for each word as described in Section 3.1. As a result, along with word- and thread-topic distribution, the model also learns the user–condition distribution γ. Although it is used later on for recommendation in our pipeline, it can also serve to predict implicit conditions. For an example culled from our dataset, a user reports *Multiple Sclerosis* as a condition he is afflicted with in his profile. However, from all of the posts that he interacts with, our model estimates the γ distribution to be {Multiple Sclerosis : 0.8, Asthma : 0.2}. In this case, the unreported, implicit condition "Asthma" is predicted by our model. We argue that this is a desirable nature of our model.

To quantitatively evaluate the capability of our model to predict the missing condition, in a separate experiment, we omit 1 to 3 conditions for each user for 1/5 of the users during training. We train our model and obtain the γ distribution for all users. We then evaluate how many cases our model can recover all of the missing conditions, *i.e.*, whether it achieves perfect recall. Table 10 reports our findings, indicating that our model can correctly predict over 60% of the cases in the single missing conditions. Unsurprisingly, performance degrades as the number of missing condition increases.

Eye & Vision	ALS	Parkinson's Disease	Diabetes	Cancer	Epilepsy	Multiple Sclerosis
eye	als	neurologist	carb	cancer	seizure	copaxone
vision	reflexes	pd	sugar	chemo	seizures	lesions
drops	muscle	nervous	insulin	radiation	keppra	mri
cataracts	amyotrophic	tremors	glucose	cells	hope	immune
red	nervous	scan	levels	kidney	meds	brain
reduce	irregular	shaking	eat	scan	brain	help
laser	brain	facial	diet	prayers	care	celebrex
opthalmologist	feel	control	exercise	god	pain	scoliosis
omeprazole	weight	tissue	test	luck	alcohol	breathing

Table 8: Example of the top words for certain medical conditions learned by our IATM+JNCTR model.

Table 9: Example of condition-specific topics (*i.e.,* symptoms and treatments) discovered by our IATM+JNCTR model.

Diabetes			Parkinson's Disease			ALS			Multiple Sclerosis			Cancer		
Topic 0	Topic 1	Topic 2	Topic 0	Topic 1	Topic 2	Topic 0	Topic 1	Topic 2	Topic 0	Topic 1	Topic 2	Topic 0	Topic 1	Topic 2
legs	carbs	blood	neurologist	feel	shaking	leg	pain	disc	brain	warm	instruction	cancer	treatment	prayer
shake	sugar	insulin	brain	help	cold	heart	feel	cervical	feeling	burning	breathing	lump	chemo	god
feet	eat	high	mri	hope	tension	muscles	issues	spine	painful	lesions	remembering	lymph	radiation	afraid
walking	diet	glucose	disorders	people	dizziness	body	help	brain	walker	harder	recall	growth	stage	doctor

# held-out conditions	Perfect recall
1	0.64
2	0.45
3	0.39

Table 10: Unreported conditions recovered by the IATM. Perfect recall denotes to the fraction of cases where it can recover all the held-out conditions.

However, gradually, even in the three missing conditions, our model can predict 39% of the cases.

6 CONCLUSION

We have systematically investigated how to best utilize each user's participation in online health forums to recommend relevant threads. Our IATM+JNCTR model leverages the user-reported clinical conditions to distinguish lexically similar yet different threads, additionally accounting for each user's specific, latent preferences for particular treatments and symptoms. In our experiments on warm- and cold-start scenarios, involving both users and threads, our framework demonstrated significant improvements over the current state-of-the-art methods. Deeper analysis reveals that IATM+JNCTR's modeling of latent conditions and user profiles are key to achieve competitive performance.

As our framework is general and language independent, we believe that it could be useful in other domains, including community question answering and scholarly paper recommendation. We hope the research community will apply our model to other scenarios to validate its modeling capabilities.

REFERENCES
[1] D. Agarwal and B.-C. Chen. fLDA: Matrix Factorization through Latent Dirichlet Allocation. In *Proc. of WSDM'10*, pages 91–100, 2010.
[2] T. Bansal, M. Das, and C. Bhattacharyya. Content Driven User Profiling for Comment-Worthy Recommendations of News and Blog Articles. In *Proc. of RecSys'15*, pages 195–202, 2015.
[3] A. Batenburg and E. Das. Emotional Coping Differences Among Breast cancer Patients from an Online Support Group: A Cross-Sectional Study. *Journal of Medical Internet Research (JMIR)*, 16(2):e28, 2014.
[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research (JMLR)*, 3:993–1022, 2003.
[5] L. Charlin, R. S. Zemel, and H. Larochelle. Leveraging User Libraries to Bootstrap Collaborative Filtering. In *Proc. of KDD'14*, pages 173–182, 2014.
[6] X. Chen, M. Zhou, and L. Carin. The Contextual Focused Topic Model. In *Proc. of KDD '12*, pages 96–104, 2012.
[7] K. Halder, L. Poddar, and M.-Y. Kan. Modeling Temporal Progression of Emotional Status in Mental Health Forum: A Recurrent Neural Net Approach. In *Proc. of WASSA '17*, pages 127–135, 2017.
[8] X. He, T. Chen, M.-Y. Kan, and X. Chen. TriRank: Review-aware Explainable Recommendation by Modeling Aspects. In *Proc. of CIKM'15*, pages 1661–1670, 2015.
[9] Z. He, M. S. Park, and Z. Chen. UMLS-Based Analysis of Medical Terminology Coverage for Tags in Diabetes-Related Blogs. *IConference 2016 Proceedings*, 2016.
[10] S. Kanthawala, A. Vermeesch, B. Given, and J. Huh. Answers to Health Questions: Internet Search Results Versus Online Health Community Responses. *Journal of Medical Internet Research (JMIR)*, 18(4):e95, 2016.
[11] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver. Multiverse Recommendation: N-dimensional Tensor Factorization for Context-aware Collaborative Filtering. In *Proc. of RecSys'10*, pages 79–86, 2010.
[12] S. Kim. An Exploratory Study of User-centered Indexing of Published Biomedical Images. *Journal of the Medical Library Association (JMLA)*, 101(1):73–76, 2013.
[13] D. D. Lee and H. S. Seung. Algorithms for Non-negative Matrix Factorization. In *Proc. of NIPS'01*, pages 556–562, 2001.
[14] J. Li, Y.-L. Theng, and S. Foo. Predictors of Online Health Information Seeking Behavior: Changes between 2002 and 2012. *Health Informatics Journal*, 22(4):804–814, 2016.
[15] J. Lin, K. Sugiyama, M.-Y. Kan, and T.-S. Chua. Addressing Cold-Start in App Recommendation: Latent User Models Constructed from Twitter Followers. In *Proc. of SIGIR'13*, pages 283–292, 2013.
[16] R. Magnezi, D. Grosberg, I. Novikov, A. Ziv, M. Shani, and L. S. Freedman. Characteristics of Patients Seeking Health Information Online via Social Health Networks versus General Internet Sites: A Comparative Study. *Informatics for Health and Social Care*, 40(2):125–138, 2015.
[17] J. D. Mcauliffe and D. M. Blei. Supervised Topic Models. In *Proc. of NIPS'07*, pages 121–128, 2008.
[18] A. Meier, J. E. Lyons, G. Frydman, M. Forlenza, and K. B. Rimer. How Cancer Survivors Provide Support on Cancer-Related Internet Mailing Lists. *Journal of Medical Internet Research (JMIR)*, 9(2):e12, 2007.
[19] T. V. Nguyen, A. Karatzoglou, and L. Baltrunas. Gaussian Process Factorization Machines for Context-aware Recommendations. In *Proc. of SIGIR'14*, pages 63–72, 2014.
[20] L. Nie, Y.-L. Zhao, M. Akbari, J. Shen, and T.-S. Chua. Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge. *IEEE Transactions on*

Knowledge and Data Engineering (TKDE), 27(2):396–409, 2015.

[21] H. J. Oh, C. Lauckner, J. Boehmer, R. Fewins-Bliss, and K. Li. Facebooking for Health: An Examination into the Solicitation and Effects of Health-Related Social Support on Social Networking Sites. *Computers in Human Behavior*, 29(5):2072–2080, 2013.

[22] A. Omari, D. Carmel, O. Rokhlenko, and I. Szpektor. Novelty based Ranking of Human Answers for Community Questions. In *Proc. of SIGIR'16*, pages 215–224, 2016.

[23] J. Palotti, L. Goeuriot, G. Zuccon, and A. Hanbury. Ranking Health Web Pages with Relevance and Understandability. In *Proc. of SIGIR'16*, pages 965–968, 2016.

[24] J. S. Pedro and A. Karatzoglou. Question Recommendation for Collaborative Question Answering Systems with RankSLDA. In *Proc. of RecSys'14*, pages 193–200, 2014.

[25] L. Poddar, W. Hsu, and M. L. Lee. Author-aware Aspect Topic Sentiment Model to Retrieve Supporting Opinions from Reviews. In *Proc. of EMNLP '17*, pages 483–492, 2017.

[26] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning. Labeled LDA: A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora. In *Proc. of EMNLP'09*, pages 248–256, 2009.

[27] S. Rendle. Factorization Machines with libFM. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):Article No. 57, 2012.

[28] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth. The Author-Topic Model for Authors and Documents. In *Proc. of UAI'04*, pages 487–494, 2004.

[29] L. Schnitzler, S. K. Smith, H. L. Shepherd, J. Shaw, S. Dong, D. M. Carpenter, F. Nguyen, and H. M. Dhillon. Communication During Radiation Therapy Education Sessions: The Role of Medical Jargon and Emotional Support in Clarifying Patient Confusion. *Patient Education and Counseling*, 100(1):112–120, 2017.

[30] E. Shmueli, A. Kagian, Y. Koren, and R. Lempel. Care to Comment?: Recommendations for Commenting on News Stories. In *Proc. of WWW 2012*, pages 429–438, 2012.

[31] M. D. Smucker, D. Kulp, and J. Allan. Dirichlet Mixtures for Query Estimation in Information Retrieval. Technical Report IR-445, Center for Intelligent Information Retrieval, University of Massachusetts, Amherst, 2005.

[32] C. Wang and D. M. Blei. Collaborative Topic Modeling for Recommending Scientific Articles. In *Proc. of KDD'11*, pages 448–456, 2011.

[33] J. Wang, X. Hu, Z. Li, W. Chao, and B. Hu. Learning to Recommend Questions Based on Public Interest. In *Proc. of CIKM'11*, pages 2029–2032, 2011.

[34] Y.-C. Wang, R. Kraut, and J. M. Levine. To Stay or Leave?: The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups. In *Proc. of CSCW '12*, pages 833–842, 2012.

[35] S. Williamson, C. Wang, K. A. Heller, and D. M. Blei. The IBP Compound Dirichlet Process and its Application to Focused Topic Modeling. In *Proc. of ICML '10*, pages 1151–1158, 2010.

[36] B. Xu, J. Bu, C. Chen, and D. Cai. An Exploration of Improving Collaborative Recommender Systems via User-Item Subgroups. In *Proc. of WWW'12*, pages 21–30, 2012.

[37] M. Zhang, J. Tang, X. Zhang, and X. Xue. Addressing Cold Start in Recommender Systems: A Semi-supervised Co-training Algorithm. In *Proc. of SIGIR'14*, pages 73–82, 2014.

[38] Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis. In *Proc. of SIGIR'14*, pages 83–92, 2014.

[39] T. C. Zhou, M. R.-T. Lyu, I. King, and J. Lou. Learning to Suggest Questions in Social Media. *Knowledge and Information Systems (KAIS)*, 43(2):389–416, 2015.