A Topic Modeling Approach for Traditional Chinese Medicine Prescriptions

Liang Yao, Yin Zhang, Baogang Wei, Wenjin Zhang, and Zhe Jin

Abstract—In traditional Chinese medicine (TCM), prescriptions are the daughters of doctors' clinical experiences, which have been the main way to cure diseases in China for several thousand years. In the long Chinese history, a large number of prescriptions have been invented based on TCM theories. Regularities in the prescriptions are important for both clinical practice and novel prescription development. Previous works used many methods to discover regularities in prescriptions, but rarely described how a prescription is generated using TCM theories. In this work, we propose a topic model which characterizes the generative process of prescriptions in TCM theories and further incorporate domain knowledge into the topic model. Using 33,765 prescriptions in TCM prescription books, the model can reflect the prescribing patterns in TCM. Our method can outperform several previous topic models and group recommendation methods on generalization performance, herbs recommendation, symptoms suggestion, and prescribing patterns discovery.

Index Terms—Traditional chinese medicine, prescriptions, topic model, domain knowledge

1 Introduction

As a system of ancient medical practice that differs in substance, methodology and philosophy to modern medicine, traditional Chinese medicine (TCM) plays an indispensable role in health care for Chinese people for several thousand years, and is becoming more frequently used in countries in the West [1].

In TCM, a prescription is a group of herbal medicines (mineral medicines and animal medicines are also used, we will use the word "herb" to refer to medicinal materials in prescriptions), which is the main way to cure diseases for thousands of years. In the long Chinese history, a lot of prescriptions have been invented to treat diseases and more than 100,000 have been recorded [2]. An example TCM prescription in *Dictionary of Traditional Chinese Medicine Prescriptions* [3] is given in Fig. 1. It has a source book, composition herbs, usage and indication symptoms.

Regularities on the herbs composition in prescriptions and corresponding symptoms play a significant role for clinical treatment and novel prescription development. For instance, common herb combinations are important for efficient clinical prescriptions [4], and the necessity of prescription patterns discovery for new drug research and development in TCM has been shown in [5].

Previous works proposed many methods that could discover regularities in prescriptions [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], but they failed to comprehensively describe how a prescription is generated using TCM theories or utilize TCM domain knowledge well. The detailed discussions of these works are in Section 2.2.

The therapeutic process in traditional Chinese medicine can be called as *li-fa-fang-yao* which is of critical importance in clinical practices [11], [16], [17]. *li-fa-fang-yao*, which means principles, methods, prescriptions and Chinese herbs respectively. It indicates the four basic steps of diagnosis and treatment: determining the cause, mechanism (syndromes) of the disease according to symptoms, then deciding the treatment methods based on the mechanism, and finally selecting a prescription as well as proper herbs. Fig. 2 shows the general process of *li-fa-fang-yao*. We refer readers to Fig. 1 in [17] which shows an intuitive example process of *li-fa-fang-yao* when TCM practitioners treat diabetes mellitus.

Regarding the basic composition of TCM prescriptions, one of the most influential theories is the principle of *jun-chen-zuo-shi* [2], [16] (also known as "emperor-minister-assistant-courier"). It means different herbs play different roles in a prescription. The *jun* (emperor) herbs treat the main cause or primary symptoms of a disease. The *chen* (minister) herbs serve to augment or broaden the effects of *jun*, and relieve secondary symptoms. The *zuo* (assistant) herbs are used to improve the effects of *jun* and *chen*, and to counteract the toxic or side effects of these herbs. The *shi* (courier) herbs are included in many prescriptions to ensure that all components in the prescription cooperate well, or to help deliver or guide them to the target organs. Taking the famous prescription "Ephedra Decoction" in Fig. 1 as an example, Ephedra (marked red) is the *jun* herb, which is used to induce sweating and treat the main symptoms aversion to cold with fever and asthma without sweat. Cassia Twig (marked blue) is the *chen* herb which helps Ephedra to induce sweating and treat secondary symptom headache and body pain. Apricot Seed (marked green) is the *zuo* herb which helps Ephedra to treat asthma. Liquorice Root (marked orange) is the *shi* herb which makes the other three herbs to work well together. The same

- *L. Yao, Y. Zhang, B. Wei, and Z. Jin are with the College of Computer Science and Technology, Zhejiang University, No. 38, Zheda Road, Hangzhou, Zhejiang 310027, China. E-mail: {yaoliang, yinzh, wbg, shrineshine}@zju.edu.cn.*
- *W. Zhang is with the First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310027, P. R. China. E-mail: drzwj2002@163.com.*

Manuscript received 26 July 2016; revised 2 Nov. 2017; accepted 11 Dec. 2017. Date of publication 1 Jan. 2018; date of current version 27 Apr. 2018. (Corresponding author: Yin Zhang.)
Recommended for acceptance by Q. He.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TKDE.2017.2787158

13140 麻黄汤 (Ephedra Decoction)

【方源】《伤寒论》。
(Source: "Treatise on Cold Damage Diseases".)

【组成】麻黄 9g 桂枝 6g 杏仁 6g 甘草 3g
(Composition herbs: Ephedra 9g, Cassia Twig 6g, Apricot Seed 6g, Liquorice Root 3g.)

【用法】水煎服。
(Usage: decocted in water for oral dose.)

【主治】外感风寒，恶寒发热，头身疼痛，无汗而喘，口不渴，舌苔薄白，脉浮紧。
(Indication symptoms: exogenous wind cold, aversion to cold with fever, headache and body pain, asthma without sweat, no thirst, white and thin coating of the tongue, floating and tense pulse.)

Fig. 1. An example TCM prescription. The herbs marked red, blue, green, and orange are the *jun* (emperor) herb, *chen* (minister) herb, *zuo* (assistant) herb, and *shi* (courier) herb, respectively. They play different roles in the prescription.

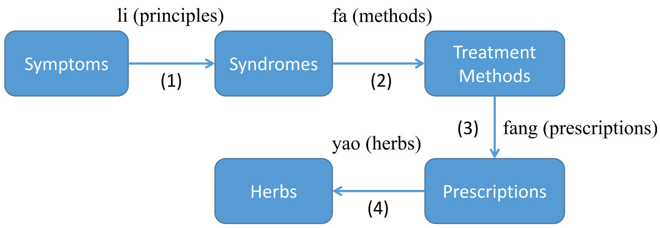


Fig. 2. The general process of *li-fa-fang-yao*.

herb can have distinct roles in different prescriptions. Thus, *jun* and *chen* herbs in one prescription may serve as *zuo* and *shi* herbs in another prescription.

Another important concept for prescribing is herb compatibility [16], [18], which means the combination of two or more herbs based on the clinical settings and the properties of herbs. The combination of herbs can improve the treatment and avoid adverse reactions. Herb pairs, the unique combinations of two relatively fixed herbs, are the most fundamental and the simplest form of herb compatibility. In the procedure of forming a prescription, herb pairs are always used as the basic units. For instance, in Fig. 1, the *jun* (emperor) herb Ephedra and the *chen* (minister) herb Cassia Twig can cooperate to induce sweating, if we only use one of them, the sweating inducing effect would be much weaker. Similarly, Radix Aconiti Lateralis Preparata and Dried Ginger are always used together in many prescriptions for dispelling cold [16].

To model the complex TCM domain, we resort to topic models [19] which are widely used in exploratory data analysis. Topic models are mainly used to uncover latent “topics” in a collection of documents. The topics are distributions over words which shows semantic patterns in the documents. Each document exhibits those topics with different degrees (topic proportions). One advantage is that topic models can be adapted to other kinds of data when we make a direct analogy from a kind of data to documents [19]. For instance, in computer vision, researchers have made a direct analogy from images to documents [20], [21]. They assume each image is a group of “visual words” and shows a combination of visual patterns (topics). In medical domain, one can regard a medical record as a “document”, view treatment activities and patient features as “words” and treatment patterns as “topics” [22]. Similarly, we can view a prescription as a “document” (a group of “herbs words” or “symptom words”) and treatment patterns in prescriptions as “topics”. Another advantage of topic models is that they can easily express relations among elements of a complex domain, and explain how the modeled data is generated and incorporate domain knowledge. Taking Fig. 1 as an example, we can only see the herbs and symptoms but could not see other elements in Fig. 2. Topic models can characterize this by regarding herbs and symptoms as observed variables, and treating syndromes, treatment methods as hidden variables. The relations among the herbs, symptoms, syndromes, treatment methods and herb roles are complicated. Topic models can easily express relations among these elements by putting edges among closely related variables. The directions of edges show how a variable is generated given other variables, thus topic models can tell how a prescription is generated by tracking these edges among variables. Topic models can be readily extended if we have prior knowledge about specific elements. In text mining tasks, a number of works have incorporated linguistic knowledge about words into topic modeling [23], [24], [25]. Similarly, we can incorporate knowledge about herbs and symptoms into prescription topic modeling.

In this work, we propose a topic model which characterizes the generative process of prescriptions in TCM theories and further incorporate domain knowledge into the topic model. Using 33,765 prescriptions in TCM prescription books, this model can reflect the prescribing patterns in TCM. The method can help TCM practitioners prescribe and pharmaceutical companies decide what combination of herbs to test.

The contributions of the paper are summarized as follows:

- It proposes a novel prescription topic model which characterizes the generative process of prescriptions based on TCM theories.
- To the best of our knowledge, our work is one of the earliest works to study the problem of herbs recommendation and symptoms suggestion. Our work is also among the earliest works to introduce a knowledge-based topic model in medical data mining.
- Our experimental results demonstrate that our method outperforms several baselines on generalization performance, herbs recommendation, symptoms suggestion and prescribing patterns discovery.
- We provide a benchmark TCM prescription dataset.[1]

2 Related Work

2.1 Topic Models

Topic models lie in a more general framework called *probabilistic graphical models* [26] which provide an elegant and principled approach to developing novel methods for data analysis and knowledge discovery. Probabilistic graphical models give us a visual language for expressing assumptions about data and its hidden structure.

Probabilistic topic models [19] such as Latent Dirichlet Allocation (LDA) [27] are commonly used machine learning methods that could find latent topics in documents. Topic models can be adapted to model many other data forms as long as we can treat target data samples as documents (groups of words). Apart from the medical record example in the introduction section, topic models could also be used

1. We released the data set, source code, and domain knowledge files of this paper at the first author's GitHub: https://github.com/yao8839836/PTM.

in many health care and biomedicine tasks. For instance, in population genetics, one can treat each individual's genotype as a "document" and genetic patterns are "topics" of those documents [28]. Chen et al. [29] showed that the configuration of functional groups in meta-genome samples can be inferred by probabilistic topic modeling. Van Esbroeck et al. [30] explored the application of topic models on heart rate time series to identify functional sets of heart rate sequences and to concisely describe patients. Recently, latent treatment patterns for clinical pathways [31] were discovered with topic modeling.

Some knowledge-based topic models [23], [24], [25] have been proposed. These models mainly use different forms of external linguistic knowledge for better text mining, but knowledge-based topic models have not been extensively explored for other kinds of data, especially for medical data.

2.2 TCM Knowledge Discovery

Knowledge discovering and data mining have become hot topics in health care and biomedicine [32], [33]. Compared with data mining research in modern biomedicine, TCM data mining just becomes popular in recent years. The efforts of TCM data mining have been reviewed by Feng et al. [34], Lukman et al. [35], Liu et al. [36] and Li and Liu [37].

A number of works have been devoted to studying the component patterns in TCM prescriptions. For example, Li et al. [6] constructed herb network using a method called Distance-based Mutual Information Model to identify useful relationships among herbs in numerous prescriptions. Zhang et al. [8] discovered interesting regularities using latent tree models [38], these regularities are of interest to students of TCM as well as pharmaceutical companies that manufacture medicine using Chinese herbs. He et al. [9] proposed an approach that could discover herbal functional groups from a large set of prescriptions recorded in TCM books. Poon et al. [7] proposed an approach that could systematically generate combinations of interacting herbs that might lead to good outcome. Zheng et al. [11] constructed prescription associated networks by mining literature data sets. Yao et al. [10] introduced a system which mines the evolutionary relationship among TCM prescriptions from prescription books.

The closest works to ours are [12], [13], [14], [15] which have explored topic modeling on TCM clinical data. Zhang et al. [12] proposed a hierarchical symptom-herb topic model which uses Link latent Dirichlet allocation (LinkLDA) [39] model and nested Chinese restaurant process to automatically extract hierarchical latent topic structures with both symptoms and their corresponding herbs in TCM clinical records. The number of hierarchical topics is automatically determined. Zhang et al. [13] proposed the Symptom-Herb-Diagnosis topic model which uses Author-topic model (ATM) [40] and diagnoses information to discover the common relationships among symptoms, herb combinations and diagnoses in clinical cases. Jiang et al. [14] applied LinkLDA directly to the same problem. Our model is an extension to LinkLDA model. In our previous work [15], we presented a framework to mine medicine usage patterns in clinical cases. We first mapped symptoms to treatment methods defined in TCM domain ontology, then viewed treatment methods as labels of a prescription and employed a supervised topic model to learn herb usage patterns under each topic (label). The method could reflect treatment methods-herbs relations. However, it could not learn direct symptom-herb relations and perform recommendation or suggestion tasks in this study because a label could correspond to different combinations of symptoms.

Although these topic models described the prescribing process, they failed to characterize the two important principles *jun-chen-zuo-shi* and herb compatibility, or could not utilize domain knowledge well, while our topic model is more consistent with TCM theories and domain knowledge.

3 Data

We collect 98,334 prescriptions from *Dictionary of Traditional Chinese Medicine Prescriptions* [3] which contains almost all (about 100,000) prescriptions recorded in China. We focus on herbs and symptoms in this work.

We filter indication symptoms by using 603 standard symptoms in *Traditional Chinese Medicine Symptoms differential diagnosis* [41], and filter herbs by using 970 herbs in *Traditional Chinese Medical Subject Headings* (TCM MeSH) [42][2] which is compatible with Medical Subject Headings (MeSH). Each symptom has a syndrome category and each herb has efficacy description text. Among all 98,334 prescriptions, 33,765 of them have both symptoms and herbs in two filters. $S = 390$ symptoms and $H = 811$ herbs appear in $P = 33{,}765$ prescriptions. We run our experiments on the 33,765 prescriptions. We randomly divided the $P = 33{,}765$ prescriptions into a training set of 28,746 prescriptions and a test set of 5,019 prescriptions.

4 Prescription Topic Model (PTM)

Guided by *li-fa-fang-yao*, TCM practitioners usually synthesise disease manifestations (symptoms) and determine syndromes of a patient first. Then treatment methods are easily determined according to syndromes. In general, a particular treatment method corresponds to a syndrome. For example, in Fig. 1, TCM practitioners first determine the syndrome "depressed nutrient and defense" which means the nutrient in blood is not well absorbed and immunity is weak and the syndrome "failure of lung qi in dispersion" which means respiratory movement is depressed, then the treatment methods "inducing sweating to releasing exterior" (which means inducing sweating and move qi (the fundamental substance which constitutes the human body) to skin) corresponding to "depressed nutrient and defense" and "diffuse the lung to calm panting" (which means regulating respiratory movement to calm panting) corresponding to "failure of lung qi in dispersion" are decided. Finally, practitioners form a prescription based on the treatment methods. In the prescription, each treatment method is implemented by some herbs (e.g., the two treatment methods mentioned above are mainly implemented by Ephedra), and each herb has a *jun-chen-zuo-shi* role (e.g., Ephedra is the *jun* herb).

Based on this process, here we introduce the details of our Prescription Topic Model (PTM). Let P be the number of prescriptions where each prescription p has N_{h_p} herbs and N_{s_p} symptoms, h_{pn} is the nth herb in p and s_{pm} is the mth symptom in p. The prescription in Fig. 1 has $N_{h_p} = 4$ herbs and $N_{s_p} = 7$ symptoms. z_{pn} is the latent treatment method assignment for h_{pn}, z'_{pm} is the latent syndrome assignment for s_{pm}, x_{pn} is the latent *jun-chen-zuo-shi* role assignment for h_{pn} (The prescriptions with known *jun-chen-*

2. Available at http://zcy.ckcest.cn/tcm/dic/home

TABLE 1
Mathematical Notations

Symbol	Description
P	The number of prescriptions.
K	The number of topics (syndromes/treatment methods).
H	The number of unique herbs.
S	The number of unique symptoms.
X	The number of unique *jun-chen-zuo-shi* roles, $X = 4$.
N_{h_p}	The number of herbs in prescription p.
N_{s_p}	The number of symptoms for prescription p.
N_{l_p}	The number of herb pairs for prescription p.
h_{pn}	The nth herb in prescription p.
$\bar{h}_{pl}$	The lth herb pair in prescription p.
h_{pl1}	The first herb of the lth herb pair in prescription p.
h_{pl2}	The second herb of the lth herb pair in prescription p.
s_{pm}	The mth symptom for prescription p.
z_{pn}	The latent treatment method assignment for h_{pn}.
z_{pl}	The latent treatment method assignment for $\bar{h}_{pl}$.
x_{pn}	The latent *jun-chen-zuo-shi* role assignment for h_{pn}.
x_{pl1}	The latent *jun-chen-zuo-shi* role assignment for h_{pl1}.
x_{pl2}	The latent *jun-chen-zuo-shi* role assignment for h_{pl2}.
z'_{pm}	The latent syndrome assignment for s_{pm}.
θ_p	The prescription-topic multinomial for prescription p.
π_{pk}	The prescription-treatment method-role multinomial for prescription p and treatment method k.
ϕ_{kx}	The treatment method-role-herb multinomial for treatment method k and role x.
ϕ'_k	The syndrome-symptom multinomial for syndrome k.
α	Hyperparameter of the Dirichlet prior on θ_p.
β	Hyperparameter of the Dirichlet prior on ϕ_{kx}.
β'	Hyperparameter of the Dirichlet prior on ϕ'_k.
η	Hyperparameter of the Dirichlet prior on π_{pk}.

zuo-shi herb roles are limited, there are only several hundred prescriptions in textbooks like [16] to our knowledge). In Fig. 1, the latent treatment method assignment for Ephedra should be “inducing sweating to releasing exterior” or “diffuse the lung to calm panting”, and the latent syndrome assignment for the symptom “asthma without sweat” should be “depressed nutrient and defense”. Let K be the number of topics, a topic $k \in 1 \ldots K$ is a syndrome and the syndrome's corresponding treatment method, (e.g., “depressed nutrient and defense” and its corresponding “inducing sweating to releasing exterior”), ϕ'_k is the S-dimensional syndrome-symptom multinomial for syndrome $k \in 1 \ldots K$, where S is the number of unique symptoms. ϕ_{kx} is the H-dimensional treatment method-role-herb multinomial for treatment method k and *jun-chen-zuo-shi* role x, where H is the number of unique herbs. θ_p is the K-dimensional prescription-topic multinomial for p. π_{pk} is the X-dimensional prescription-treatment method-role multinomial for prescription p and treatment method k, $X = 4$, which means an herb is a *jun*, *chen*, *zuo* or *shi* herb. α, β, β' and η are hyperparameters of the Dirichlet prior on θ_p, ϕ_{kx}, ϕ'_k and π_{pk} respectively. We illustrate the mathematical notations in Table 1. The generative story of our prescription topic model is given as follows:

(1) For each prescription p draw $\theta_p \sim \mathrm{Dir}(\alpha)$.
(2) For each syndrome k in $1 \ldots K$, draw $\phi'_k \sim \mathrm{Dir}(\beta')$.
(3) For each prescription p and treatment method k in $1 \ldots K$, draw $\pi_{pk} \sim \mathrm{Dir}(\eta)$.
(4) For each treatment method k in $1 \ldots K$ and *jun-chen-zuo-shi* role x in $1 \ldots X$, draw $\phi_{kx} \sim \mathrm{Dir}(\beta)$.
(5) For each of the N_{s_p} symptoms in prescription p:
    a) Draw a syndrome $z'_{pm} \sim \mathrm{Mult}(\theta_p)$.
    b) Draw a symptom $s_{pm} \sim \mathrm{Mult}(\phi'_{z'_{pm}})$.
(6) For each of the N_{h_p} herbs in prescription p:
    a) Draw a treatment method $z_{pn} \sim \mathrm{Mult}(\theta_p)$.
    b) Draw a role $x_{pn} \sim \mathrm{Mult}(\pi_{pz_{pn}})$.
    c) Draw an herb $h_{pn} \sim \mathrm{Mult}(\phi_{z_{pn}x_{pn}})$.

This generative story is shown in the probabilistic graphical models representation of Fig. 3a. It is similar to Link latent Dirichlet allocation (LinkLDA) model [39]. The distinction is that we encode herb role x into our model. We name it PTM(a).

4.1 Model Inference and Learning

We use Gibbs sampling to infer latent assignments z'_{pm}, z_{pn} and x_{pn}. The Gibbs sampling equation for z'_{pm} is defined as

$$p(z'_{pm} = k | s_{pm}, \mathbf{s}_{-pm}, \mathbf{z}'_{-pm}, \mathbf{z}, \alpha, \beta') \propto \frac{n_{pk} + \alpha}{N_{s_p} + N_{h_p} + K\alpha} \times \frac{n_{ks_{pm}} + \beta'}{n_k + S\beta'}, \quad (1)$$

where k is a syndrome, $\mathbf{s}_{-pm}$ are all symptoms except s_{pm}, $\mathbf{z}'_{-pm}$ are syndrome assignments for all symptoms except s_{pm}, $\mathbf{z}$ are treatment method assignments for all herbs. n_{pk} is the number of times topic (syndrome or treatment method) k is assigned to a symptom or an herb in prescription p, $n_{ks_{pm}}$ is the number of times s_{pm} is assigned to syndrome k, n_k is the number of times any symptom is assigned to syndrome k.

The sampling equation for z_{pn} and x_{pn} is defined as

$$p(z_{pn} = k, x_{pn} = x | h_{pn}, \mathbf{z}_{-pn}, \mathbf{x}_{-pn}, \mathbf{h}_{-pn}, \mathbf{z}', \alpha, \beta, \eta) \propto \frac{n_{pk} + \alpha}{N_{s_p} + N_{h_p} + K\alpha} \times \frac{n_{pkx} + \eta}{n'_{pk} + X\eta} \times \frac{n_{kxh_{pn}} + \beta}{n_{kx} + H\beta}, \quad (2)$$

where k is a treatment method, x is a *jun-chen-zuo-shi* role, $\mathbf{z}_{-pn}$ are treatment method assignments for all herbs except h_{pn}, $\mathbf{x}_{-pn}$ are role assignments for all herbs except h_{pn}, $\mathbf{h}_{-pn}$ are all herbs except h_{pn}, $\mathbf{z}'$ are syndrome assignments for all symptoms. n_{pkx} is the number of times treatment method k and role x are assigned to an herb in prescription p, n'_{pk} is the number of times treatment method k is assigned to an herb in prescription p, $n_{kxh_{pn}}$ is the number of times k and x are assigned to h_{pn}.

With Gibbs sampling, we can make the following parameter estimation

$$\theta_{pk} = \frac{n_{pk} + \alpha}{N_{s_p} + N_{h_p} + K\alpha} \quad (3)$$

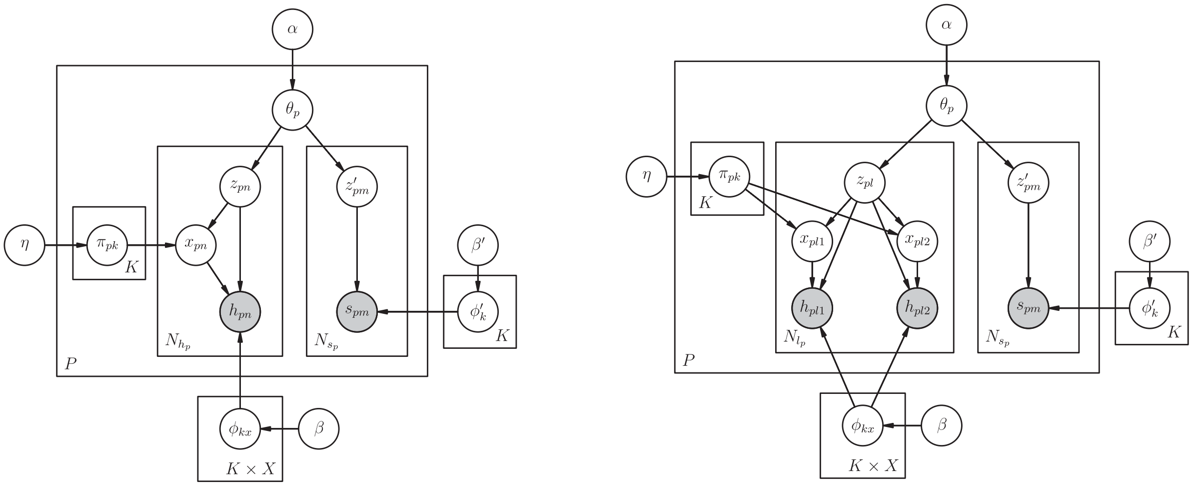
$$\phi'_{ks_{pm}} = \frac{n_{ks_{pm}} + \beta'}{n_k + S\beta'} \quad (4)$$

$$\pi_{pkx} = \frac{n_{pkx} + \eta}{n'_{pk} + X\eta} \quad (5)$$

$$\phi_{kxh_{pn}} = \frac{n_{kxh_{pn}} + \beta}{n_{kx} + H\beta} \quad (6)$$

4.2 Herb Compatibility

Since herb pairs are always used as the basic units, and each herb pair often implements a certain treatment method [16], we extract herb pairs from each training prescription, i.e., if any two herbs co-occur in a prescription p of the training set (e.g., the two herbs Ephedra and Cassia Twig in Fig. 1), we add the herb pair to the herb pair set of p. There are



(a) PTM(a) (b) PTM(b)

Fig. 3. The probabilistic graphical models representation of PTM. (a) PTM(a): The prescription topic model with herb role only. (b) PTM(b): The prescription topic model with herb role and herb compatibility.

$N_{l_p} = C^2_{N_{h_p}} = N_{h_p}(N_{h_p} - 1)/2$ herb pairs in p when $N_{h_p} > 1$, if there is only one herb in p, we assume that p has one herb pair, but the pair consists of two identical herbs. As shown in the left part of Fig. 3b, the generative story of the link set of prescription p is as follows:

(1) For each herb pair $\vec{h}_{pl}$ of the N_{l_p} herb pairs in prescription p:
   a) Draw a treatment method $z_{pl} \sim \text{Mult}(\theta_p)$.
   b) Draw two roles $x_{pl1}, x_{pl2} \sim \text{Mult}(\pi_{pz_{pl}})$.
   c) Draw two herbs $h_{pl1} \sim \text{Mult}(\phi_{z_{pl}x_{pl1}})$, $h_{pl2} \sim \text{Mult}(\phi_{z_{pl}x_{pl2}})$.

We name this model with herb compatibility PTM(b). The inference equation for z_{pl}, x_{pl1} and x_{pl2} is defined as

$$
\begin{aligned}
&p(z_{pl} = k, x_{pl1} = x_1, x_{pl2} = x_2 | \vec{h}_{pl}, \mathbf{z}_{-pl}, \mathbf{x}_{-pl}, \alpha, \beta, \eta) \\
&\propto \frac{n''_{pk} + \alpha}{N_{s_p} + N_{l_p} + K\alpha} \times \frac{n_{pkx_1} + \eta}{n'_{pk} + X\eta} \times \frac{n_{kx_1 h_{pl1}} + \beta}{n_{kx1} + H\beta} \\
&\times \frac{n_{pkx_2} + \eta}{n'_{pk} + X\eta} \times \frac{n_{kx_2 h_{pl2}} + \beta}{n_{kx2} + H\beta},
\end{aligned}
\tag{7}
$$

where $\vec{h}_{pl}$ is the lth herb pair in prescription p, $\mathbf{z}_{-pl}$ are treatment method assignments for all herb pairs except $\vec{h}_{pl}$, $\mathbf{x}_{-pl}$ are role assignments for all herb pairs except $\vec{h}_{pl}$, n''_{pk} is the number of times any herb pair or symptom in p is assigned to topic k. The inference equation for z'_{pm} in PTM(b) is similar to Equation (1), but we need to replace n_{pk} and N_{h_p} with n''_{pk} and N_{l_p}

$$
\begin{aligned}
&p(z'_{pm} = k | s_{pm}, \mathbf{s}_{-pm}, \mathbf{z}'_{-pm}, \mathbf{z}, \alpha, \beta') \\
&\propto \frac{n''_{pk} + \alpha}{N_{s_p} + N_{l_p} + K\alpha} \times \frac{n_{ks_{pm}} + \beta'}{n_k + S\beta'}
\end{aligned}
\tag{8}
$$

The parameter estimation equations for $\phi'_{ks_{pm}}$, π_{pkx} and ϕ_{kxh} in PTM(b) are the same as in PTM(a), the only difference is

$$
\theta_{pk} = \frac{n''_{pk} + \alpha}{N_{s_p} + N_{l_p} + K\alpha} \tag{9}
$$

4.3 Incorporating Herb Efficacy Knowledge

In this section, we use TCM prior knowledge to improve the prescription topic model. We extract the symptom-herb correspondences from the training prescriptions. Specifically, we use the 390 symptoms to filter efficacy descriptions of 811 herbs in TCM MeSH, and obtain the symptom-herb correspondences, then for each prescription in the training set, if an herb h in a prescription p can treat a symptom s of p's indication, we add h and s (e.g., the herb Ephedra and the symptom aversion to cold with fever in Fig. 1) to the symptom-herb corresponding set of prescription p. Since their correspondence in TCM knowledge, we assume a symptom s in the corresponding set can only be assigned to the topics of s's corresponding herbs in prescription p.

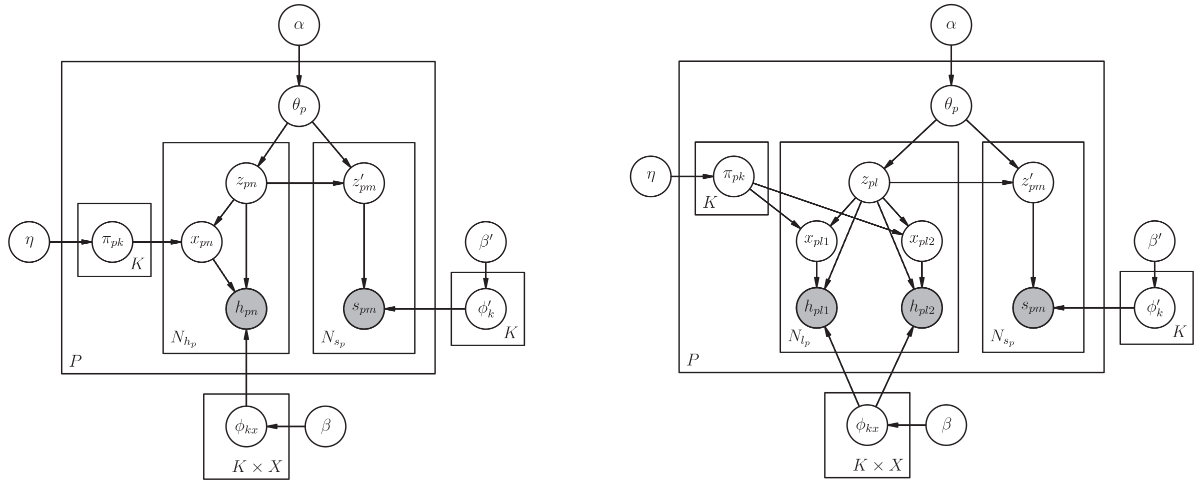
We name the prescription topic model with herb role and herb efficacy knowledge PTM(c) which is illustrated in Fig. 4a. If s_{pm} has no corresponding herb in prescription p, the inference equation for z'_{pm} is the same as Equation (1); otherwise, z'_{pm} can only be sampled from the topic assignment set $\{z_{pn} | h_{pn} \text{ treats } s_{pm}\}$ of s_{pm}'s corresponding herbs $\{h_{pn} | h_{pn} \text{ treats } s_{pm}\}$ in p, the inference equation for z'_{pm} is

$$
\begin{aligned}
&p(z'_{pm} = k | s_{pm}, \mathbf{s}_{-pm}, \mathbf{z}'_{-pm}, \mathbf{z}, \alpha, \beta') \propto \\
&I[k \in \{z_{pn} | h_{pn} \text{ treats } s_{pm}\}] \times \\
&\frac{n_{pk} + \alpha}{N_{s_p} + N_{h_p} + K\alpha} \times \frac{n_{ks_{pm}} + \beta'}{n_k + S\beta'},
\end{aligned}
\tag{10}
$$

where $I[y] = 1$ when y is true and $I[y] = 0$ when y is false. The inference equation for z_{pn} and x_{pn} in PTM(c) is the same as Equation (2). The parameter estimation equations for PTM(c) are the same as PTM(a).

We name our prescription topic model with herb role, herb compatibility and herb efficacy knowledge PTM(d) which is shown in Fig. 4b. If s_{pm} has no corresponding herb in prescription p, the inference equation for z'_{pm} is the same as Equation (8); otherwise, z'_{pm} can only be sampled from the topic assignment set $\{z_{pl} | h_{pl1} \text{ treats } s_{pm} \text{ or } h_{pl2} \text{ treats } s_{pm}\}$ of s_{pm}'s corresponding herbs in p, the inference equation for z'_{pm} is

$$
\begin{aligned}
&p(z'_{pm} = k | s_{pm}, \mathbf{s}_{-pm}, \mathbf{z}'_{-pm}, \mathbf{z}, \alpha, \beta') \\
&\propto I[k \in \{z_{pl} | h_{pl1} \text{ treats } s_{pm} \text{ or } h_{pl2} \text{ treats } s_{pm}\}] \\
&\times \frac{n''_{pk} + \alpha}{N_{s_p} + N_{l_p} + K\alpha} \times \frac{n_{ks_{pm}} + \beta'}{n_k + S\beta'}
\end{aligned}
\tag{11}
$$



(a) PTM(c) (b) PTM(d)

Fig. 4. The probabilistic graphical models representation of PTM with herb efficacy knowledge. (a) PTM(c): The prescription topic model with herb role and herb efficacy knowledge. (b) PTM(d): The prescription topic model with herb role, herb compatibility and herb efficacy knowledge.

The inference equation for z_{pl}, x_{pl1} and x_{pl2} in PTM(d) is the same as Equation (7). The parameter estimation equations for PTM(d) are the same as PTM(b).

5 Experiment

In this section we evaluate our prescription topic model on four experimental tasks. Specifically we want to determine:

- Can our model achieve better generalization performance than other topic models?
- Can our model recommend herbs for a list of symptoms?
- Can our model suggest symptoms for a list of herbs?
- Can our model reflect the prescribing patterns in TCM?

We compare our prescription topic model (PTM) with eight baselines. Among them, six baselines are topic models, three baselines are group recommendation methods. We compare our model with group recommendation methods because recommending herbs (symptoms) for a list of symptoms (herbs) is analogous to recommending items to a group of users.

- *Author-topic model (ATM)* [40] employed by previous work [13] which treats herbs as authors and symptoms as words.
- *LinkLDA* [39] used in previous works [12] and [14] which views herbs and symptoms as words and references.
- *Block-LDA* [43], a topic model that extends Link-LDA. It can model links between certain type of entities. We treat herb-pairs set extracted from all training prescriptions as the external links.
- *Link-PLSA-LDA* [44], a topic model that extends Link-LDA. It can model links between different types of entities. We treat symptom-herb correspondence set extracted from all training prescriptions as the external links.
- *User-based collaborative filtering with averaging strategy (CF-AVG)* [45], a widely used group recommendation method. CF-AVG first estimates the rating score of each user in a user group by user-based collaborative filtering, then uses the average of these scores as the recommendation score for the group. We compute the conditional probability of items (herbs/symptoms) given users (symptoms/herbs) in training prescriptions as the rating score.
- *User-based collaborative filtering with least-misery strategy (CF-LM)* [45], a widely used group recommendation method which uses the smallest rating score of group users as the recommendation score for the group. We also compute the conditional probability of items given users in training prescriptions as the rating score.
- *COnsensus Model (COM)* [46], a group recommendation method which simulates the generative process of group events and make recommendations for a group of users. We treat herbs in a prescription as a group of users when recommending symptoms, and view symptoms as a group of users when recommending herbs.
- *Bilingual Biterm Topic Model (BiBTM)* [47], a topic model describing the generation process of a paired bilingual document corpus. We treat herbs in a prescription as the words in a document and symptoms as the words in the translated version of the document.

We set the following hyperparameters: for PTM: $\alpha = 1$, $\beta = 0.1, \beta' = 0.1, \eta = 1$; for LinkLDA: $\alpha = 1, \beta = 0.1, \beta' = 0.1$; for Block-LDA: $\alpha_D = \alpha_L = 1, \gamma = 0.1$; for Link-PLSA-LDA: $\alpha_\theta = \alpha_L(\text{hyperparameter of } \pi) = 1,\ \beta'(\text{hyperparameter of } \Omega) = \gamma\ (\text{hyperparameter of } \beta) = 0.1$; for BiBTM: $\alpha = 1, \beta = 0.1$; for ATM: $\alpha = 50/K, \beta = 0.01$ as suggested in [40]; for COM: $\alpha = 50/K, \beta = \eta = 0.01, \gamma = \gamma_t = 0.5$ and $\rho = 0.01$ as suggested in [46]. For CF-AVG and CF-LM, we use Pearson correlation similarity and top 10 similar users. We find that small changes of hyperparameters do not change the results much. All topic models are trained using 1,000 Gibbs iterations.

5.1 Generalization Performance

5.1.1 Herbs Predictive Perplexity

We use the predictive perplexity to evaluate the herbs predictive power of topic models. Perplexity is a standard measure for estimating the performance of a probabilistic model which has been used to evaluate predictive capability of topic models in previous works [13], [14], [40]. The predictive perplexity of a set of test herbs given symptoms is

$$\begin{aligned} perplexity(h_{test}|s_{test}) &= \exp\left(-\frac{\sum_{p=1}^{P_{test}} \log p(\vec{h}_p|\vec{s}_p)}{\sum_{p=1}^{P_{test}} N_{h_p}}\right) \\ p(\vec{h}_p|\vec{s}_p) &= \prod_{h_{pn}\in\vec{h}_p} p(h_{pn}|\vec{s}_p) = \prod_{h_{pn}\in\vec{h}_p} \frac{1}{N_{s_p}} \sum_{s_{pm}\in\vec{s}_p} p(h_{pn}|s_{pm}), \end{aligned} \tag{12}$$

where s_{test} are the symptoms in test prescriptions, h_{test} are the herbs in test prescriptions, $\vec{s}_p$ are symptoms in prescription p of the test set, $\vec{h}_p$ are herbs in prescription p of the test set, $P_{test} = 5,019$ is the number of prescriptions in the test set. Better predictive performance is indicated by a lower perplexity over test prescriptions.

The probability of an herb h given a symptom s for PTM is

$$\begin{aligned} p(h|s) &= \sum_{p,k,x} p(h|k,x)p(x|p,k)p(p|k)p(k|s) \\ &= \sum_{p,k,x} p(h|k,x)p(x|p,k)\frac{p(k|p)}{\sum_{p'} p(k|p')}\frac{p(s|k)}{\sum_{k'} p(s|k')} \\ &= \sum_{p,k,x} \phi_{kxh}\pi_{pkx}\frac{\theta_{pk}}{\sum_{p'}\theta_{p'k}}\frac{\phi'_{ks}}{\sum_{k'}\phi'_{k's}}, \end{aligned} \tag{13}$$

Fig. 5 shows the herbs predictive perplexity of several topic models with different number of topics. We do not compute predictive perplexity for COM and BiBTM because they only describe generative story of group events (a group of symptoms and a selected herb for herbs recommendations) or herb/symptoms pairs set without modeling prescriptions explicitly. We can see that ATM does not perform well, which implies treating herbs as authors and symptoms as words is not consistent with the generative story of prescriptions. LinkLDA performs better than ATM, which shows the correctness of modeling herbs and symptoms as two parts of a prescription. Block-LDA performs better than LinkLDA, which demonstrates using herb links can improve herb predictive capabilities. Link-PLSA-LDA outperforms LinkLDA, which shows extracting symptom-herb correspondences from prescriptions can help herb prediction. PTM(a) performs better than LinkLDA and similarly to Link-PLSA-LDA, because considering herb roles can highlight most relevant herbs (*jun* (emperor) and *chen* (minister) herbs) of given symptoms and ignore less relevant herbs. PTM(b) has lower perplexity scores than PTM(a) and LinkLDA ($p < 10^{-3}$), which means considering herb compatibility in each prescription can significantly improve the herb predictive power. This is intuitive because when seeing a symptom, practitioners not only use an herb that can treat the symptom, but also use a compatible herb to augment the effect or counteract the toxic [16]. PTM(c) also significantly outperforms PTM(a) ($p < 10^{-6}$), which demonstrates restricting symptom topic assignments using herb efficacy knowledge is also an efficient way to help herbs prediction, this is also intuitive because the knowledge makes an herb and its indication symptoms tend to be under the same topic. PTM(d) has the lowest perplexity scores, and significantly outperforms PTM(c) ($p < 10^{-6}$), which means considering both herb compatibility and herb efficacy knowledge leads to the best herb predictive power. However, compared to PTM(b), PTM(d) only improves a little, as connecting herb

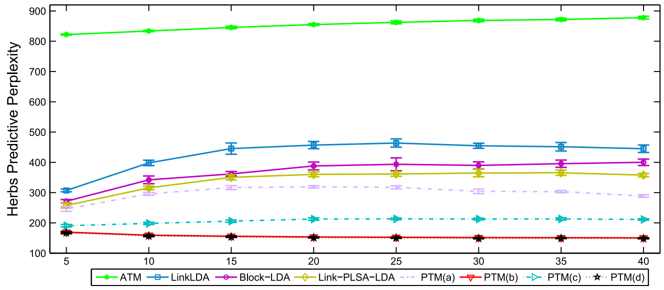


Fig. 5. Herbs predictive perplexity of each model with different number of topics K. A lower perplexity means the predictive power is better. We run all models 10 times and report the mean $\pm$ standard deviation. Improvements of PTM(a), PTM(b), PTM(c), and PTM(d) over LinkLDA are all significant ($p < 10^{-3}$) based on 2-tailed paired t-test.

pairs and symptoms can make a symptom and its corresponding herb appear in the same topic, but meanwhile highlight some less related herbs for the symptom.

5.1.2 Symptoms Predictive Perplexity

The predictive perplexity of a set of test symptoms given herbs is

$$\begin{aligned} perplexity(s_{test}|h_{test}) &= \exp\left(-\frac{\sum_{p=1}^{P_{test}} \log p(\vec{s}_p|\vec{h}_p)}{\sum_{p=1}^{P_{test}} N_{s_p}}\right) \\ p(\vec{s}_p|\vec{h}_p) &= \prod_{s_{pm}\in\vec{s}_p} p(s_{pm}|\vec{h}_p) = \prod_{s_{pm}\in\vec{s}_p} \frac{1}{N_{h_p}} \sum_{h_{pn}\in\vec{h}_p} p(s_{pm}|h_{pn}). \end{aligned} \tag{14}$$

The probability of a symptom s given an herb h for PTM is

$$\begin{aligned} p(s|h) &= \sum_k p(s|k) \sum_x p(k,x|h) \\ &= \sum_k \phi'_{ks} \sum_x \frac{\phi_{kxh}}{\sum_{k',x'} \phi_{k'x'h}} \end{aligned} \tag{15}$$

Fig. 6 gives the symptoms predictive perplexity of each model with different number of topics. From Table 4, we can see that ATM also does not perform well on symptoms prediction, and LinkLDA performs better than ATM again, which shows modelling herbs and symptoms as two types of words of a document is a better choice. Block-LDA performs similarly to LinkLDA, which means using extracted herb pairs as external links outside the training prescriptions could not help symptom prediction much. Link-PLSA-LDA significantly outperforms Link-LDA ($p < 10^{-4}$), which means herb-symptom links can also help symptom prediction. PTM(a) has lower perplexity than LinkLDA ($p < 0.01$), which means considering herb roles can significantly improve the symptoms predictive power. This is because when seeing a list of herbs, the *jun-chen-zuo-shi* labels can highlight *jun* (emperor) herbs and *chen* (minister) herbs, and the corresponding symptoms are mainly treated by *jun* herbs and *chen* herbs. PTM(b) performs slightly better than PTM(a), which shows considering compatible herb may highlight *chen* (minister) herbs or *zuo* (assistant) herbs, which are also used to treat the corresponding symptoms. PTM(c) also slightly outperforms PTM(a), which shows restricting symptom topic assignment can also improve symptom predictive capability, but the improvement is not obvious as the improvement in herb prediction task, the reason could be that corresponding symptoms are fewer than

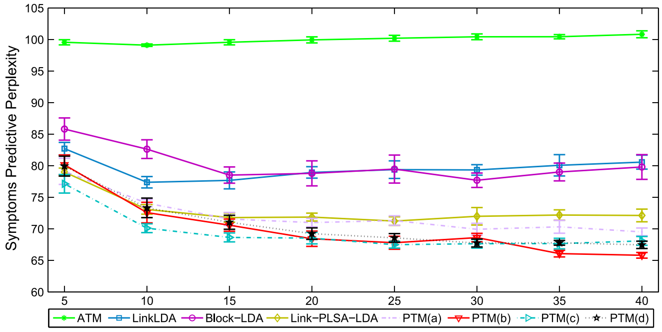


Fig. 6. Symptoms predictive perplexity of the topic models with different number of topics K. A lower perplexity means the predictive power is better. We run all models 10 times and report the mean $\pm$ standard deviation. Improvements of PTM(a), PTM(b), PTM(c), and PTM(d) over Link-LDA are all significant ($p < 0.01$) based on 2-tailed paired t-test.

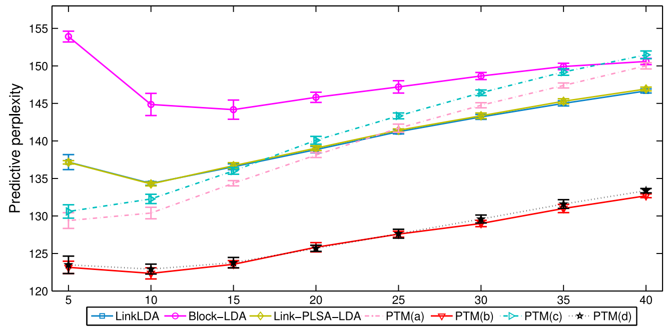


Fig. 7. Prescription predictive perplexity of the topic models with different number of topics K. A lower perplexity means the predictive power is better. We run all models 10 times and report the mean $\pm$ standard deviation. Improvements of PTM(b) and PTM(d) over LinkLDA are significant ($p < 10^{-10}$) based on 2-tailed paired t-test.

herbs in a prescription, which makes the data more sufficient for the symptom prediction task, but incorporating external knowledge is still useful. PTM(d) performs similarly to PTM(b), because connecting herb pairs and symptoms can highlight some less related herbs for a symptom and weaken the effect of herb-symptom correspondences.

5.1.3 *Prescription Predictive Perplexity*

We also compute prescription predictive perplexity to evaluate the generalization performance of the whole prescription, including both herbs and symptoms. Following [48], [49], [50], we add the first half of each test prescription to the training data, while retaining the second half for evaluation. We estimate prescription level parameters θ_p and π_p on the first half of the test prescriptions, then use the learned parameters to calculate the perplexity. We randomly split the test prescriptions (including both herbs and symptoms) into the first half and the second half. The prescription predictive perplexity is defined as

$$perplexity(pre_{test}|pre_{train}) = \exp\left(-\frac{\sum_{p=1}^{P_{test}} \log p(\vec{s}_p, \vec{h}_p)}{\sum_{p=1}^{P_{test}} (N_{s_p} + N_{h_p})}\right), \quad (16)$$

where pre_{test} are the test prescriptions, pre_{train} are the training prescriptions.

Fig. 7 presents the prescription predictive perplexity of several topic models. We do not compute prescription predictive perplexity for ATM because we could not compute the probability of herbs (authors) in test set given parameters learned in training set. We can note that Block-LDA performs worse than Link-LDA. And Link-PLSA-LDA, PTM(a) and PTM(c) also do not improve Link-LDA. This is because the symptoms in a prescription are usually fewer than herbs in the same prescription. Most "words" among the first half and the second half of test prescriptions are herbs. Thus outside herb-pairs set, herb-symptom correspondences and herb roles may not help the herbs prediction given the other herbs in the same prescription. PTM(b) and PTM(d) significantly outperform Link-LDA, the reason is that extracting all herb pairs in each prescription can highlight herb co-occurrence and help the herbs prediction given other herbs.

5.2 Herbs Recommendation

We compute the following conditional probability of an herb given a set of test symptoms

$$p(h|\vec{s}_p) = \frac{1}{N_{s_p}} \sum_{s_{pm} \in \vec{s}_p} p(h|s_{pm}). \quad (17)$$

We then use the top N herbs with the largest probabilities as the recommendation herbs of our PTM model. Following previous work [46], We use average $Precision@N$ ($P@N$) as the recommendations effectiveness metric. $Precision@N$ is the proportion of the top N herb recommendations that are in the real prescription p. Formally, $Precision@N$ is defined as

$$Precision@N = \frac{|\{\text{top } N \text{ herbs}\} \cap \{\text{true herbs}\}|}{|\{\text{top } N \text{ herbs}\}|}. \quad (18)$$

We average the $precision@N$ of all testing prescriptions as the final $P@N$.

Table 2 presents herbs $Precision@N$ of each model with different K and N values. Note that CF-AVG and CF-LM results are always the same because they do not have the parameter. We can observe that for topic models, the $Precision@N$ scores are generally consistent with perplexity scores. A topic model with lower perplexity scores also tends to achieve higher $Precision@N$. PTM(a) performs slightly better than LinkLDA on average, and it can significantly outperform LinkLDA when K increases ($p < 0.01$ at $K = 40$), which shows using herb roles makes sense and it is more suitable to distinguish herbs roles when there are more treatment methods. All topic models tend to perform better when K increases, and the highest $Precision@N$ are generally achieved by PTM(b) and PTM(d). CF-AVG and CF-LM perform well, because computing symptom (user) similarities using conditional probability $p(h|s)$ can highlight most relevant herbs of a symptom and filter some noise, while topic models use bag of words (herbs/symptoms) representation, they may consider less related herbs. We can further consider using weighted representation of prescriptions for topic modeling. COM could not produce satisfactory results. It treats symptoms in a prescription as a group of users, and views a group of users and an herb selected by the users as an independent group event, this can neglect relations among herbs. BiBTM performs worse than LinkLDA, considering herb pairs, symptom pairs and symptom-herb pairs respectively may ignore structure information of original prescriptions.

TABLE 2
Herbs $Precision@N$ ($P@N$) of Each Model with Different K (the Number of Topics) and N

Model \ K	20			30			40		
	$P@5$	$P@10$	$P@20$	$P@5$	$P@10$	$P@20$	$P@5$	$P@10$	$P@20$
ATM	0.0088 ± 0.0021	0.0091 ± 0.0026	0.0087 ± 0.0005	0.0089 ± 0.0023	0.0093 ± 0.0016	0.0092 ± 0.0004	0.0081 ± 0.0023	0.0089 ± 0.0014	0.0083 ± 0.0006
LinkLDA	0.2301 ± 0.0067	0.1851 ± 0.0015	0.1336 ± 0.0010	0.2277 ± 0.0036	0.1789 ± 0.0023	0.1298 ± 0.0014	0.2188 ± 0.0031	0.1786 ± 0.0014	0.1276 ± 0.0005
Block-LDA	0.2269 ± 0.0030	0.1817 ± 0.0015	0.1321 ± 0.0016	0.2286 ± 0.0029	0.1803 ± 0.0020	0.1300 ± 0.0014	0.2192 ± 0.0052	0.1770 ± 0.0019	0.1283 ± 0.0006
Link-PLSA-LDA	0.2320 ± 0.0037	0.1858 ± 0.0016	0.1356 ± 0.0013	0.2284 ± 0.0036	0.1813 ± 0.0016	0.1392 ± 0.0010	0.2236 ± 0.0034	0.1793 ± 0.0021	0.1297 ± 0.0015
BiBTM	0.2143 ± 0.0000	0.1604 ± 0.0000	0.1216 ± 0.0000	0.2143 ± 0.0000	0.1604 ± 0.0000	0.1216 ± 0.0000	0.2143 ± 0.0000	0.1604 ± 0.0000	0.1216 ± 0.0000
CF-AVG	0.2324 ± 0.0000	0.1933 ± 0.0000	0.1476 ± 0.0000	0.2324 ± 0.0000	0.1933 ± 0.0000	0.1476 ± 0.0000	0.2324 ± 0.0000	0.1933 ± 0.0000	0.1476 ± 0.0000
CF-LM	0.2320 ± 0.0000	0.1936 ± 0.0000	0.1481 ± 0.0000	0.2320 ± 0.0000	0.1936 ± 0.0000	0.1481 ± 0.0000	0.2320 ± 0.0000	0.1936 ± 0.0000	0.1481 ± 0.0000
COM	0.2197 ± 0.0008	0.1731 ± 0.0010	0.1289 ± 0.0008	0.2194 ± 0.0011	0.1746 ± 0.0012	0.1295 ± 0.0007	0.2197 ± 0.0011	0.1745 ± 0.0008	0.1316 ± 0.0005
PTM(a)	0.2320 ± 0.0032	0.1835 ± 0.0027	0.1346 ± 0.0013	0.2299 ± 0.0039	0.1819 ± 0.0019	0.1348 ± 0.0006	0.2241 ± 0.0032	0.1810 ± 0.0033	0.1326 ± 0.0007
PTM(b)	0.2475 ± 0.0029	0.1998 ± 0.0027	0.1497 ± 0.0009	0.2507 ± 0.0029	0.2039 ± 0.0020	$\mathbf{0.1525 \pm 0.0008}$	$\mathbf{0.2533 \pm 0.0024}$	$\mathbf{0.2056 \pm 0.0011}$	0.1528 ± 0.0009
PTM(c)	0.2385 ± 0.0041	0.1920 ± 0.0016	0.1414 ± 0.0008	0.2376 ± 0.0037	0.1880 ± 0.0020	0.1326 ± 0.0007	0.2313 ± 0.0039	0.1846 ± 0.0024	0.1398 ± 0.0006
PTM(d)	$\mathbf{0.2486 \pm 0.0023}$	$\mathbf{0.2009 \pm 0.0019}$	$\mathbf{0.1497 \pm 0.0006}$	$\mathbf{0.2522 \pm 0.0029}$	$\mathbf{0.2040 \pm 0.0022}$	0.1512 ± 0.0016	0.2528 ± 0.0027	0.2053 ± 0.0011	$\mathbf{0.1531 \pm 0.0008}$

We run all models 10 times and report the mean $\pm$ standard deviation. Improvements of PTM(b), PTM(c), and PTM(d) over LinkLDA are all significant ($p < 0.01$) based on 2-tailed paired t-test.

TABLE 3
Symptoms $Precision@N$ ($P@N$) of Each Model with Different K (the Number of Topics) and N

Model \ K	20			30			40		
	$P@5$	$P@10$	$P@20$	$P@5$	$P@10$	$P@20$	$P@5$	$P@10$	$P@20$
ATM	0.0742 ± 0.0006	0.0522 ± 0.0004	0.0375 ± 0.0002	0.0738 ± 0.0009	0.0520 ± 0.0005	0.0367 ± 0.0001	0.0738 ± 0.0008	0.0519 ± 0.0006	0.0363 ± 0.0001
LinkLDA	0.1062 ± 0.0009	0.0719 ± 0.0006	0.0464 ± 0.0002	0.1063 ± 0.0008	0.0715 ± 0.0004	0.0460 ± 0.0003	0.1063 ± 0.0011	0.0709 ± 0.0006	0.0458 ± 0.0001
Block-LDA	0.1027 ± 0.0019	0.0692 ± 0.0009	0.0453 ± 0.0002	0.1040 ± 0.0008	0.0693 ± 0.0008	0.0452 ± 0.0003	0.1038 ± 0.0014	0.0694 ± 0.0006	0.0456 ± 0.0005
Link-PLSA-LDA	$\mathbf{0.1081 \pm 0.0008}$	0.0723 ± 0.0005	0.0468 ± 0.0003	$\mathbf{0.1080 \pm 0.0015}$	0.0728 ± 0.0005	0.0469 ± 0.0002	$\mathbf{0.1085 \pm 0.0010}$	0.0725 ± 0.0005	0.0469 ± 0.0002
BiBTM	0.0750 ± 0.0000	0.0528 ± 0.0000	0.0371 ± 0.0000	0.0749 ± 0.0000	0.0528 ± 0.0000	0.0371 ± 0.0000	0.0749 ± 0.0000	0.0528 ± 0.0000	0.0371 ± 0.0000
CF-AVG	0.1050 ± 0.0000	$\mathbf{0.0769 \pm 0.0000}$	$\mathbf{0.0514 \pm 0.0000}$	0.1050 ± 0.0000	$\mathbf{0.0769 \pm 0.0000}$	$\mathbf{0.0514 \pm 0.0000}$	0.1050 ± 0.0000	$\mathbf{0.0769 \pm 0.0000}$	$\mathbf{0.0514 \pm 0.0000}$
CF-LM	0.0977 ± 0.0000	0.0716 ± 0.0000	0.0478 ± 0.0000	0.0977 ± 0.0000	0.0716 ± 0.0000	0.0478 ± 0.0000	0.0977 ± 0.0000	0.0716 ± 0.0000	0.0478 ± 0.0000
COM	0.0775 ± 0.0009	0.0597 ± 0.0005	0.0413 ± 0.0003	0.0849 ± 0.0013	0.0649 ± 0.0008	0.0437 ± 0.0004	0.0918 ± 0.0013	0.0681 ± 0.0008	0.0449 ± 0.0001
PTM(a)	0.1064 ± 0.0010	0.0717 ± 0.0006	0.0459 ± 0.0003	0.1071 ± 0.0016	0.0714 ± 0.0006	0.0463 ± 0.0003	0.1078 ± 0.0008	0.717 ± 0.0006	0.0469 ± 0.0002
PTM(b)	0.0996 ± 0.0016	0.0697 ± 0.0006	0.0460 ± 0.0002	0.1026 ± 0.0011	0.0713 ± 0.0008	0.0471 ± 0.0004	0.1036 ± 0.0011	0.0722 ± 0.0008	0.0475 ± 0.0002
PTM(c)	0.1018 ± 0.0015	0.0705 ± 0.0005	0.0464 ± 0.0001	0.1038 ± 0.0011	0.0705 ± 0.0005	0.0467 ± 0.0003	0.1029 ± 0.0008	0.0707 ± 0.0003	0.0467 ± 0.0002
PTM(d)	0.0981 ± 0.0012	0.0694 ± 0.0005	0.0453 ± 0.0003	0.1005 ± 0.0013	0.0709 ± 0.0009	0.0460 ± 0.0003	0.1011 ± 0.0008	0.0718 ± 0.0007	0.0469 ± 0.0002

We run all models 10 times and report the mean $\pm$ standard deviation.

5.3 Symptoms Suggestion

We compute the following conditional probability of a symptom given a set of test herbs

$$p(s|\vec{h}_p) = \frac{1}{N_{h_p}} \sum_{h_{pn} \in \vec{h}_p} p(s|h_{pn}) \tag{19}$$

The $Precision@N$ for symptom recommendation is defined as

$$Precision@N = \frac{|\{\text{top } N \text{ symptoms}\} \cap \{\text{true symptoms}\}|}{|\{\text{top } N \text{ symptoms}\}|} \tag{20}$$

We also average the $precision@N$ of all testing prescriptions as the final $P@N$.

Table 3 presents symptoms $Precision@N$ of each model with different K and N values. We note that ATM and BiBTM do not perform well as in herbs recommendation, the reasons are also similar. COM also neglects symptoms correlations, so it cannot produce satisfactory results. CF-AVG and CF-LM perform well when N is large, the conditional probability $p(s|h)$ can also highlight most relevant symptoms of an herb. Link-PLSA-LDA performs better than LinkLDA which shows the effect of extracted herb-symptom correspondences using herb efficacy knowledge. PTM(a) can perform better than LinkLDA when K increases ($p < 0.002$ at $K = 40$ and $N = 5$), which shows herbs roles are more helpful for larger topic number in recommendation tasks. We notice that PTM(b), PTM(c) and PTM(d) have slightly lower perplexity than Link-LDA but cannot achieve higher symptoms $Precision@5$. But they can achieve higher $Precision@N$ than LinkLDA when N increases, which means they can rank the true symptoms higher on average, but may not rank true symptoms to top 5. Moreover, the $Precision@N$ scores are low because symptoms are often few in a prescription.

5.4 Prescribing Patterns Discovery

We now evaluate topics learned from all 33,765 prescriptions by our model. We first qualitatively show some topics. Then we quantitatively evaluate learned topics by comparing to TCM prior knowledge.

5.4.1 Qualitative Results

Table 4 presents three topics generated by several topic models with $K = 25$. We show top 10 symptoms on the left

TABLE 4
Example Topics Learned by Several Topic Models with $K = 25$

Blood-regulating		Nourishing heart and tranquilizing mind		Harmonizing intestines and stomach	
ATM					
oppression in the chest	*Semen Trichosanthis*	*abdominal pain*	*Longtube Groundivy Herb*	*spontaneous sweating*	*Folium Hibisci Mutabilis*
aversion to cold	*Mercury Oxidum*	amnesia	Fluorite	abdominal fullness	*Amber*
stomachache	*Calculus Equi*	*measles*	Nardostachys Root	*chronic shank ulcer*	*Ardisia Japonica*
profuse spittle	*Terminalia chebula Retz*	palpitation	Bamboo Shavings	*bloody stool*	*Snakegourd Root*
hyperopia	*Folium Phyllostach Lophatheri*	*vomiting*	*Emblic Leafflower Fruit*	stomach reflux	*Coffea Arabica*
waggling tongue	*Serissa Serissoides*	*infantile malnutrition*	*Radix Boehmeriae*	borborigmus	*Officinal Magnolia Flower*
hypermenorrhea	*Pharbitis Seed*	*arthralgia*	Motherwort Herb	*dizziness*	*Chives*
palpitations below the heart	*Hibiscus Mutabilis*	*metrorrhagia*	*Lotus Leaf*	*retention of the lochia*	*Fermented Soybean*
postpartum metrorrhagia	*Air Potato*	*indigestion*	Radix Aconiti Kusnezoffii	*greenish complexion*	*Rumex Japonicus*
vomiting	*Pilose Antler*	*tremor of feet*	Foeniculum Vulgare	*rigidity of limbs*	*Fruit of Sharpleaf Calangal*
LinkLDA					
epistaxis	Chinese Angelica	palpitation	*Common Yam Rhizome*	vomiting	Common Aucklandia Root
bloody stool	*Paeonia Veitchii*	amnesia	Dodder Seed	nausea	Clove
hemafecia	Red Peony Root	deafness	Eucommia Bark	borborigmus	Fructus Amomi Rotundus
hemoptysis	*Liquorice Root*	lumbago	Chinese Magnoliavine Fruit	stomach reflux	Chinese Eaglewood Wood
hematuria	Paeonia Suffruticosa	frequent urination	Asiatic Cornelian Cherry Fruit	acid swallow	Foeniculum Vulgare
dizziness	Unprocessed Rehmannia Root	night sweating	*Achyranthes Bidentata*	*tenesmus*	Nutmeg
heaviness of head	*Debark Peony Root*	enuresis	Desertliving Cistanche	abdomen cold	*Medicine Terminalia Fruit*
hypermenorrhea	Tree Peony Root Bark	dreamfulness	Prepared Rehmannia Root	dysphagia	Villous Amomum Fruit
hematemesis	Cattail Pollen	*infertility*	Barbary Wolfberry Fruit	abdominal pain	Cablin Patchouli Herb
infertility	Colla Corii Asini	dizziness	Pilose Antler	*spasm*	Cardamon Fruit
Block-LDA					
hematemesis	*Paeonia Veitchii*	amnesia	Milkwort Root	vomiting	*Dried Tangerine Peel*
bloody stool	*Chinese Angelica*	lumbago	*Achyranthes Bidentata*	acid swallow	Officinal Magnolia Bark
epistaxis	Red Peony Root	dizziness	Eucommia Bark	nausea	Villous Amomum Fruit
limbs pain	Unprocessed Rehmannia Root	palpitation	*Common Yam Rhizome*	epigastric upset	*Massa Medicata Fermentata*
hematuria	*Liquorice Root*	night sweating	Chinese Magnoliavine Fruit	belching	Atractylodes Rhizome
hemoptysis	*Debark Peony Root*	frequent urination	Dodder Seed	dysphagia	Nutgrass Galingale Rhizome
retention of the lochia	Paeonia Suffruticosa	enuresis	Prepared Rehmannia Root	diarrhea	Cablin Patchouli Herb
hemafecia	Tree Peony Root Bark	dreamfulness	Asiatic Cornelian Cherry Fruit	anorexia	*Hawthorn Fruit*
retention of placenta	Cattail Pollen	fatigue	Desertliving Cistanche	hiccup	*Green Tangerine peel*
yellow sweat	*Sichuan Lovage Rhizome*	deafness	*Dendrobium*	stomach reflux	Pinellia Tuber
Link-PLSA-LDA					
white vaginal discharge	Chinese Angelica	dizziness	Dwarf Lilyturf Tuber	vomiting	Common Aucklandia Root
red and white vaginal discharge	Debark Peony Root	palpitation	Milkwort Root	abdominal pain	Clove
hematemesis	*Sichuan Lovage Rhizome*	amnesia	*Common Yam Rhizome*	nausea	Fructus Amomi Rotundus
threatened abortion	*Paeonia Veitchii*	dreaminess	Salvia Root	borborygmus	Chinese Eaglewood Wood
tidal fever	Paeonia Suffruticosa	vertigo	Tangshen	regurgitation	Nutmeg
infertility	Tree Peony Root Bark	oppression in chest	Chinese Angelica	acid regurgitation	Villous Amomum Fruit
vaginal bleeding during pregnancy	Nutgrass Galingale Rhizome	vexation	Chinese Magnoliavine Fruit	dysphagia	Cablin Patchouli Herb
hypochondriac pain	Unprocessed Rehmannia Root	insomnia	Grassleaf Sweetflag Rhizome	hiccup	Foeniculum Vulgare
flooding and spotting	Prepared Rehmannia Root	fatigue	Spine Date Seed	abdomen cold	*Medicine Terminalia Fruit*
bloody stool	Colla Corii Asini	night sweating	Debark Peony Root	stomachache	Cardamon Fruit
PTM(a)					
hematemesis	Chinese Angelica	dizziness	Milkwort Root	abdominal pain	Common Aucklandia Root
epistaxis	*Paeonia Veitchii*	palpitation	Chinese Magnoliavine Fruit	vomiting	Clove
hemafecia	*Liquorice Root*	amnesia	*Common Yam Rhizome*	nausea	Fructus Amomi Rotundus
hematuria	Paeonia Suffruticosa	lumbago	Eucommia Bark	borborygmus	Chinese Eaglewood Wood
bloody stool	Unprocessed Rehmannia Root	deafness	*Achyranthes Bidentata*	*spasm*	Cablin Patchouli Herb
hemoptysis	Debark Peony Root	dreaminess	*Dodder Seed*	diarrhea	Foeniculum Vulgare
flooding and spotting	Colla Corii Asini	anorexia	Cornus Officinalis	vomiting and diarrhea	Nutmeg
menorrhagia	*Radix Ophiopogonis*	fatigue	Grassleaf Sweetflag Rhizome	regurgitation	Villous Amomum Fruit
shortage of qi	Eriobotrya Japonica	vertigo	Desertliving Cistanche	abdomen cold	*Officinal Magnolia Bark*
glossorrhagia	Tree Peony Root Bark	frequent urination	Chinese Arborvitae kernel	acid regurgitation	Common Floweringqince Fruit
PTM(b)					
hematemesis	Golden Thread	amnesia	Milkwort Root	vomiting	Fructus Amomi Rotundus
tidal fever	Liquorice Root	blurred vision	*Lightyellow Sophora Root*	abdominal pain	Clove
night sweating	Radix Bupleuri	dizziness	Liquorice Root	nausea	Common Aucklandia Root
bloody stool	Turtle Carapace	palpitation	Poria	borborigmus	Liquorice Root
infantile malnutrition	Figwortflower Picrorhiza Rhizome	vexation	Chinese Angelica	acid regurgitation	*Ginseng*
epistaxis	Chinese Angelica	insomnia	Divaricate Saposhnikovia Root	spasm	*Officinal Magnolia Bark*

TABLE 4
(*Continued*)

Blood-regulating		Nourishing heart and tranquilizing mind		Harmonizing intestines and stomach	
emaciation	Areca Seed	dreaminess	Ginseng	abdomen cold	White Atractylodes Rhizome
abdominal pain	Rangooncreeper Fruit	dysphoria	Spine Date Seed	regurgitation	Fresh Ginger
indigestion	Common Aucklandia Root	weep	Fleeceflower Root	abdominal fullness	Radix Aconiti Lateralis Preparata
flooding and spotting	*Massa Medicata Fermentata*	headache	Chrysanthemum Flower	bitter taste in mouth	Dried Ginger
		PTM(c)			
abdominal pain	Chinese Angelica	lumbago	Chinese Magnoliavine Fruit	abdominal pain	Common Aucklandia Root
hematemesis	Debark Peony Root	deafness	Milkwort Root	borborigmus	Fructus Amomi Rotundus
red and white vaginal discharge	Sichuan Lovage Rhizome	amnesia	Eucommia Bark	vomiting	Clove
flooding and spotting	Colla Corii Asini	night sweating	*Achyranthes Bidentata*	nausea	Foeniculum Vulgare
dystocia	Nutgrass Galingale Rhizome	shortness of breath	Dodder Seed	lumbago	*Common Buried Tuber*
metrostaxis	*Paeonia Veitchii*	dizziness	*Common Yam Rhizome*	abdomen cold	Nutmeg
white vaginal discharge	Argy Wormwood Leaf	blurred vision	Asiatic Cornelian Cherry Fruit	hiccup	*Zedoray Rhizome*
metrorrhagia	Prepared Rehmannia Root	frequent urination	Grassleaf Sweetflag Rhizome	tenesmus	Chinese Eaglewood Wood
threatened abortion	Cattail Pollen	spontaneous sweating	Desertliving Cistanche	acid regurgitation	Areca Seed
vaginal bleeding during pregnancy	Motherwort Herb	infertility	*Dendrobium*	halitosis	Green Tangerine peel
		PTM(d)			
hematemesis	Oyster Shell	amnesia	Dodder Seed	vomiting	Fructus Amomi Rotundus
abdominal pain	Chinese Angelica	lumbago	Poria	abdominal pain	*Ginseng*
bloody stool	Bone Fossil of Big Mammals	night sweating	Milkwort Root	borborygmus	Dried Ginger
white vaginal discharge	Garden Burnet Root	dizziness	*Achyranthes Bidentata*	nausea	White Atractylodes Rhizome
night sweating	Liquorice Root	deafness	Chinese Magnoliavine Fruit	acid regurgitation	Liquorice Root
metrorrhagia	Red Halloysite	palpitation	Desertliving Cistanche	reversal cold of hands and feet	Radix Aconiti Lateralis Preparata
red and white vaginal discharge	Colla Corii Asini	white vaginal discharge	Chinese Angelica	spasm	*Officinal Magnolia Bark*
metrostaxis	Golden Thread	blurred vision	Pilose Antler	abdomen cold	Fresh Ginger
tenesmus	*Dried Ginger*	infertility	Eucommia Bark	abdominal fullness	Common Aucklandia Root
epistaxis	Debark Peony Root	frequent urination	*Ginseng*	*hiccup*	Poria

We show top 10 symptoms (left) and top 10 herbs (right). Symptoms italicized and marked in red do not appear in other symptoms' syndrome categories. Herbs italicized and marked in red could not treat the top 10 symptoms. We manually labeled the topic names.

and top 10 herbs on the right (The probability of an herb given a topic for PTM is $p(h|k) = \sum_{p,x} p(h|k,x)p(x|p,k)p(p|k) = \phi_{kxh}\pi_{pkx}\frac{\theta_{pk}}{\sum_{p'}\theta_{p'k}}$). We do not present topics of COM and BiBTM because all the K topics in each model are basically the same. Symptoms italicized and marked in red do not appear in other topical symptoms' syndrome categories in [41]. Herbs italicized and marked in red could not treat the top 10 symptoms (validated by TCM MeSH symptom-herb correspondences in Section 4.3). Note that for PTM's topics that are not discovered by the baseline models, we try to find the best possible matches from the topics of the baseline models.

The first topic is about blood-related symptoms and their corresponding blood-regulating herbs. We can see that: (1). ATM could not find good topic. On the left, only hypermenorrhea and postpartum metrorrhagia are blood-related symptoms. On the right, none of the ten herbs can treat the ten symptoms on the left. (2). LinkLDA finds much better topic. On the left, seven symptoms are blood-related symptoms. On the right, Chinese Angelica can treat hypermenorrhea. Red Peony Root, Paeonia Suffruticosa, Unprocessed Rehmannia Root and Tree Peony Root Bark can treat hematemesis. Cattail Pollen can treat hematemesis and bloody stool. Colla Corii Asini can treat hematemesis, bloody stool and hematuria. (3). Block-LDA finds six blood-related symptoms and five correct herbs. (4). Link-PLSA-LDA can find coherent symptoms and accurate herbs. Debark Peony Root can treat hypochondriac pain and flooding and spotting. Nutgrass Galingale Rhizome can treat threatened abortion and flooding and spotting. Prepared Rehmannia Root can treat tidal fever, infertility and flooding and spotting. (5). PTM(a) finds coherent symptoms, Eriobotrya Japonica can treat hematemesis and hemoptysis. (6). PTM (b) finds coherent symptoms and nine correct herbs. Golden Thread and Figwortflower Picrorhiza Rhizome can treat hematemesis, tidal fever and night sweating. Liquorice Root and Common Aucklandia Root can treat abdominal pain. Radix Bupleuri can treat infantile malnutrition. Turtle Carapace can treat tidal fever. Areca Seed can treat abdominal pain and indigestion. Rangooncreeper Fruit can treat infantile malnutrition and abdominal pain. (7). PTM(c) finds nine blood-related symptoms and nine herbs are correct. Sichuan Lovage Rhizome can treat abdominal pain. Argy Wormwood Leaf can treat flooding and spotting, hematemesis and threatened abortion. Cattail Pollen can treat flooding and spotting, hematemesis and abdominal pain. Motherwort Herb can treat vaginal bleeding during pregnancy, dystocia and abdominal pain. (8). PTM(d) finds nine correct herbs. Oyster Shell and Bone Fossil of Big Mammals can treat night sweating. Garden Burnet Root can treat red and white vaginal discharge, white vaginal discharge, hematemesis and bloody stool. Red Halloysite can treat bloody stool.

The second topic is about "nourishing heart and tranquilizing mind". We can find that: (1). ATM could not find good topic again. On the left, only amnesia and palpitation are mental symptoms. On the right, only Fluorite can treat the mental symptoms palpitation. (2). LinkLDA finds much better topic

TABLE 5
Example Topic Roles Learned by PTM(a) with $K = 25$

Blood-regulating				
Symptoms	Role 0	Role 1	Role 2	Role 3
hematemesis	Chinese Angelica	Paeonia Suffruticosa	*Liquorice Root*	Colla Corii Asini
epistaxis	*Paeonia Veitchii*	Tree Peony Root Bark	Eriobotrya Japonica	Chinese Angelica
hemafecia	*Liquorice Root*	Chinese Angelica	Loquat Leaf	Unprocessed Rehmannia Root
hematuria	Debark Peony Root	*Paeonia Veitchii*	*Radix Ophiopogonis*	Cattail Pollen
bloody stool	Red Peony Root	Unprocessed Rehmannia Root	*Ginseng*	Debark Peony Root
hemoptysis	*Caulis Akebiae*	Prepared Rehmannia Root	Bamboo Shavings	Garden Burnet Root
flooding and spotting	*Radix Ophiopogonis*	Golden Thread	*Reed Rhizome*	Sophora Flower
menorrhagia	Beautiful Sweetgum Resin	*Radix Ophiopogonis*	Unprocessed Rehmannia Root	Panax Notoginseng
shortage of qi	Lotus Rhizome Node	Red Peony Root	*Pyrus Bretschneideri*	Chinese Arborvitae Twig and Leaf
glossorrhagia	*Orange Fruit*	*Liquorice Root*	Egg	India Madder Root
Nourishing heart and tranquilizing mind				
Symptoms	Role 0	Role 1	Role 2	Role 3
dizziness	Desertliving Cistanche	*Common Yam Rhizome*	Milkwort Root	*Milkvetch Root*
palpitation	*Dodder Seed*	Eucommia Bark	Chinese Magnoliavine Fruit	Deer horm
amnesia	*Achyranthes Bidentata*	Asiatic Cornelian Cherry Fruit	Grassleaf Sweetflag Rhizome	Fleeceflower Root
lumbago	Pilose Antler	*Achyranthes Bidentata*	Chinese Arborvitae kernel	Tangshen
deafness	*Dendrobium*	*Oriental Waterplantain Rhizome*	Spine Date Seed	Prepared Rehmannia Root
dreaminess	Eucommia Bark	Chinese Magnoliavine Fruit	*Salvia Root*	Barbary Wolfberry Fruit
anorexia	*Morinda Root*	Prepared Rehmannia Root	*Dwarf Lilyturf Tuber*	*Dodder Seed*
fatigue	Asiatic Cornelian Cherry Fruit	*Gordon Euryale Seed*	Poria	Ligustrum Lucidum
vertigo	Chinese Magnoliavine Fruit	Radix Codonopsis	Dimocarpus Longan	Deer-Horm Glue
frequent urination	Palmleaf Raspbery Fruit	*Malaytea Scurfpea Fruit*	Arillus Longan	Glossy Privet Fruit

We show top 10 symptoms (left) and top 10 herbs of each role (right). Herbs italicized and marked in red could not treat the top 10 symptoms.

again. On the left, nine symptoms are mental symptoms except infertility. On the right, Dodder Seed can treat enuresis. Eucommia Bark can treat dizziness and lumbago. Chinese Magnoliavine Fruit can treat palpitation, night sweating and enuresis. Asiatic Cornelian Cherry Fruit can treat dizziness, deafness, frequent urination and enuresis. Desertliving Cistanche can treat lumbago and infertility. Prepared Rehmannia Root can treat palpitation, deafness, night sweating and infertility. Barbary Wolfberry Fruit can treat dizziness. Pilose Antler can treat deafness and infertility. (3). Block-LDA finds coherent symptoms and seven correct herbs. Milkwort Root can treat amnesia and dreamfulness. (4). Link-PLSA-LDA finds good topic. Dwarf Lilyturf Tuber and Salvia Root can treat vexation. Tangshen and Chinese Angelica can treat palpitation. Milkwort Root can treat amnesia and dreaminess. Grassleaf Sweetflag Rhizome can treat amnesia. Spine Date Seed can treat dreaminess and night sweating. Debark Peony Root can treat night sweating. (5). PTM(a) finds coherent symptoms and seven correct herbs. Cornus Officinalis can treat dizziness, deafness and frequent urination. Chinese Arborvitae kernel can treat palpitation and amnesia. (6). PTM (b) finds nine correct herbs. Liquorice Root and Fleeceflower Root can treat palpitation. Poria can treat amnesia and palpitation. Divaricate Saposhnikovia Root can treat headache. Ginseng can treat dysphoria. Chrysanthemum Flower can treat blurred vision and headache. (7). All the ten symptoms found by PTM(c) are mental symptoms and seven herbs are correct. (8). PTM(d) finds ten mental symptoms and eight correct herbs. Pilose Antler can treat deafness and infertility.

The third topic presents intestines and stomach-related symptoms and herbs for "Harmonizing intestines and stomach". We can note that: (1). ATM still finds poor topic. On the left, only abdominal fullness, stomach reflux and borborigmus are intestines and stomach-related symptoms. On the right, none of the ten herbs can treat the ten symptoms on the left. (2). LinkLDA still shows its superiority to ATM. On the left, eight symptoms are intestines and stomach-related symptoms. On the right, Common Aucklandia Root can treat abdominal pain, vomiting, borborigmus and tenesmus. Clove, Fructus Amomi Rotundus and Chinese Eaglewood Wood can treat vomiting and abdomen cold. Foeniculum Vulgare can treat abdominal pain, abdomem cold and vomiting. Nutmeg and Cardamon Fruit can treat vomiting. Villous Amomum Fruit can treat abdominal pain, vomiting and nausea. Cablin Patchouli Herb can treat abdominal pain and vomiting. (3). Block-LDA finds coherent symptoms and six correct herbs. Officinal Magnolia Bark and Atractylodes Rhizome can treat anorexia. Nutgrass Galingale Rhizome can treat acid regurgitation and belching. Pinellia Tuber can treat vomiting and stomach reflux. (4). Link-PLSA-LDA performs very well on symptoms and makes one mistake on herbs. (5). PTM(a) finds both nine correct symptoms and herbs. Common Floweringqince Fruit can treat spasm. (6). PTM(b) also performs well on symptoms. On the right, Fresh Ginger can treat vomiting. Radix Aconiti Lateralis Preparata can treat spasm. Dried Ginger can treat vomiting and abdomen cold. (7). PTM(c) finds ten intestines and stomach-related symptoms and eight correct herbs. Green Tangerine peel can treat abdominal pain. (8). PTM(d) finds nine intestines and stomach-related symptoms and eight correct herbs. Poria can treat vomiting.

From the three topics, we observe that our prescription topic model could find topics that reflect TCM prescribing patterns well. After incorporating herb compatibility and herb efficacy knowledge, the patterns discovery capability can be improved as shown in PTM(b), PTM(c), PTM(d) and Link-PLSA-LDA topics.

Table 5 shows four roles' top herbs of two topics generated by PTM(a). In the "Blood-regulating" topic, we can see that all ten herbs of Role 3 can treat at least one of the symptoms, and we find seven of the ten herbs can treat at least 3 symptoms of the top ten symptoms. Because Role 3 treats main symptoms of the syndrome, we can label it as *jun*

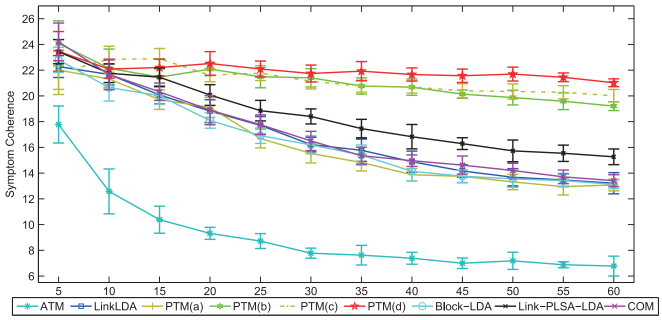


Fig. 8. Average topic symptom coherence of several topic models with different K (the number of topics) and top 10 symptoms. We run all models 10 times and report the mean $\pm$ standard deviation of all topics' average coherence for each model. PTM(d) significantly outperforms LinkLDA ($p < 0.05$) based on 2-tailed paired t-test.

(emperor). We may also label Role 1 as *chen* (minister) and Role 0 as *zuo* (assistant). Because Role 2 has the least corresponding herbs and Liquorice Root is the most widely used *shi* (courier) herb, we can label Role 2 as *shi* (courier). For "Nourishing heart and tranquilizing mind" topic, we may label Role 2 as *jun* (emperor), Role 3 as *chen* (minister), Role 0 as *zuo* (assistant) and Role 1 as *shi* (courier) in a similar way.

5.4.2 Quantitative Results

We now quantitatively evaluate learned topics by comparing to TCM prior knowledge. We want to determine: (1) whether symptoms under a topic are closely related and can represent a certain syndrome? (2) how many herbs in a topic can treat the corresponding symptoms?

Topic Symptom Coherence. We define topic symptom coherence to measure topics' symptom quality. Let N be the number of top symptoms of a topic k, s_{ki} be a symptom in top N symptoms, c be any syndrome category in [41], and S_c be symptoms in c. For each pair of top symptoms s_{ki} and s_{kj}, we check if they co-occur in any category c. Formally, the symptom coherence of topic k is given as

$$coherence(k) = \sum_{j=2}^{N}\sum_{i=1}^{j-1} \mathbf{1}_{\{\exists c:\, s_{ki}\in S_c \wedge s_{kj}\in S_c\}} \qquad (21)$$

Fig. 8 shows average topic symptom coherence of several topic models with different number of topics and $N = 10$. For COM, we treat symptoms as users and herbs as items. We can observe that ATM could not find coherent symptoms, and the coherence scores are much lower than others. Block-LDA and COM perform similarly to LinkLDA, and they can find reasonable symptom topics. PTM (a) also performs similarly to LinkLDA, the difference is not obvious. Despite of significant predictive perplexity difference, the symptoms extracted by two models remain almost the same, at least for the top 10 symptoms. This result is similar to the discovery in [51] that the lower perplexity may not enhance interpretability of inferred topics. Link-PLSA-LDA, PTM(b), PTM(c) and PTM(d) find more coherent symptoms than others, which shows using herb-herb links and herb efficacy knowledge can improve the coherence of symptoms.

Topic Herb Precision. We compute the herb precision to measure topics' herb quality, the herb precision is defined as: if an herb h_{ki} in the top N herbs of topic k can treat a symptom s_{kj} in the top N symptoms S_k of k (validated by TCM MeSH symptom-herb correspondences in Section 4.3), we label the herb h_{ki} as a correct herb, and the herb precision is the proportion of correct herbs in the top N herbs. Formally, the herb precision of topic k is given as

$$precision(k) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\exists s_{kj}:\, s_{kj}\in S_k \wedge h_{ki} \text{ treats } s_{kj}\}} \qquad (22)$$

Table 6 presents average topic herb precision of several topic models with different number of topics and $N = 10$. We can see that most herbs learned by LinkLDA, Block-LDA, Link-PLSA-LDA and PTM can be validated, while ATM and COM could not find accurate herbs. We think the reasons are similar as that in predictive tasks, and we also note that COM tends to find most commonly used herbs in all topics because common herbs are selected by most symptom groups. PTM(a) performs similarly to LinkLDA, which is also similar to the discovery in [51]. PTM(b), PTM(c) and PTM(d) find more accurate herbs than others and the improvements are significant ($p < 0.01$), which shows using herb compatibility and herb efficacy knowledge can also improve the precision of herbs.

5.5 Discussion

From experimental results, we can conclude that ATM is not suitable for TCM prescriptions modeling, and it shows the worst performances on generalization performance, herbs/symptoms recommendation and treatment patterns discovery. Group recommendation methods such as CF-AVG, CF-LM and COM perform well on herbs/symptoms recommendation tasks, but could not discover meaningful

TABLE 6
Average Topic Herb Precision of Several Topic Models with Different K (the Number of Topics) and Top 10 Symptoms/Herbs

Model \ K	5	10	15	20	25	30	35	40
ATM	0.236 ± 0.058	0.252 ± 0.042	0.234 ± 0.032	0.228 ± 0.024	0.217 ± 0.025	0.233 ± 0.015	0.237 ± 0.014	0.231 ± 0.034
LinkLDA	0.750 ± 0.044	0.693 ± 0.211	0.660 ± 0.032	0.647 ± 0.025	0.622 ± 0.031	0.639 ± 0.023	0.624 ± 0.026	0.606 ± 0.025
Block-LDA	0.710 ± 0.050	0.652 ± 0.034	0.621 ± 0.034	0.575 ± 0.019	0.589 ± 0.045	0.582 ± 0.027	0.595 ± 0.027	0.584 ± 0.027
Link-PLSA-LDA	0.778 ± 0.048	0.711 ± 0.031	0.677 ± 0.029	0.690 ± 0.027	0.696 ± 0.019	0.701 ± 0.021	0.693 ± 0.018	0.678 ± 0.013
COM	0.574 ± 0.038	0.461 ± 0.042	0.419 ± 0.028	0.421 ± 0.032	0.406 ± 0.015	0.393 ± 0.023	0.378 ± 0.022	0.382 ± 0.017
PTM(a)	0.774 ± 0.040	0.710 ± 0.049	0.647 ± 0.026	0.618 ± 0.030	0.597 ± 0.026	0.615 ± 0.025	0.593 ± 0.019	0.579 ± 0.024
PTM(b)	0.836 ± 0.042	0.781 ± 0.034	0.749 ± 0.031	0.713 ± 0.009	0.699 ± 0.026	0.701 ± 0.022	0.684 ± 0.019	0.670 ± 0.017
PTM(c)	0.864 ± 0.034	0.817 ± 0.026	$\mathbf{0.807 \pm 0.025}$	$\mathbf{0.820 \pm 0.013}$	$\mathbf{0.808 \pm 0.023}$	$\mathbf{0.808 \pm 0.012}$	$\mathbf{0.801 \pm 0.010}$	$\mathbf{0.800 \pm 0.013}$
PTM(d)	$\mathbf{0.866 \pm 0.044}$	$\mathbf{0.817 \pm 0.028}$	0.803 ± 0.018	0.770 ± 0.015	0.790 ± 0.021	0.780 ± 0.025	0.763 ± 0.011	0.770 ± 0.022

We run all models 10 times and report the mean $\pm$ standard deviation of all topics' average precision for each model. PTM(b), PTM(c), and PTM(d) significantly outperform others ($p < 0.01$) based on 2-tailed paired t-test.

treatment patterns. LinkLDA performs relatively well on all four tasks. Block-LDA and BiBTM generally do not improve LinkLDA because they model herb/symptoms pairs outside training prescriptions and may ignore the original prescriptions structures. By considering herb roles, PTM(a) can obtain better generalization and herbs/symptoms recommendation performance, but the treatment patterns discovery capabilities are not improved. Nevertheless, PTM(a) could infer herb roles in a prescription, and herb roles inference in each prescription is another interesting problem to explore. By incorporating herb compatibility, PTM(b) further gains better performances on all four tasks. By incorporating herb efficacy knowledge, Link-PLSA-LDA and PTM(c) also gain better performances on all four tasks. PTM(d) generally achieves the best results on all tasks because it considers both herb compatibility and herb efficacy knowledge. These results demonstrate it is necessary to consider TCM background in TCM data analysis, and this work can be a promising start for incorporating domain knowledge into the prescription topic modeling.

6 Conclusion and Future Work

This paper presented a novel topic model for TCM prescriptions. It characterizes the generative process of prescriptions in TCM theories. Using 33,765 prescriptions, this model can discover the prescribing patterns in TCM. Furthermore, it can outperform several previous methods on recommending herbs for a list of symptoms and predicting symptoms for a prescription. The method is helpful for clinical research and practice.

In future work, we plan to incorporate more prescription information such as usage, form and herbal dosage, and more domain knowledge such as symptoms' syndrome category as prior knowledge into our model. And evaluating herb roles inferred by our model is an interesting problem we are going to investigate.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 61572434), the China Knowledge Centre for Engineering Sciences and Technology (No. CKCEST-2017-1-3), the Zhejiang Provincial Natural Science Foundation of China (No. LY14F020027), and the Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) (20130101110136).

References

[1] F. Cheung, "TCM: Made in china," *Nature*, vol. 480, no. 7378, pp. S82–S83, 2011.

[2] J. Qiu, "Traditional medicine: A culture in the balance," *Nature*, vol. 448, no. 7150, pp. 126–128, 2007.

[3] H. Peng, *Dictionary of Traditional Chinese Medicine Prescriptions*. Beijing: People Health Press, 1996, [in Chinese, ISBN: 7117018879].

[4] X. Zhou, et al., "Development of traditional chinese medicine clinical data warehouse for medical knowledge discovery and decision support," *Artif. Intell. Med.*, vol. 48, no. 2, pp. 139–152, 2010.

[5] H. Yang, et al., "New drug R&D of traditional chinese medicine: Role of data mining approaches," *J. Biol. Syst.*, vol. 17, no. 03, pp. 329–347, 2009.

[6] S. Li, B. Zhang, D. Jiang, Y. Wei, and N. Zhang, "Herb network construction and co-module analysis for uncovering the combination rule of traditional chinese herbal formulae," *BMC Bioinf.*, vol. 11, no. 11, 2010, Art. no. 1.

[7] S. K. Poon, et al., "A novel approach in discovering significant interactions from TCM patient prescription data," *Int. J. Data Mining Bioinf.*, vol. 5, no. 4, pp. 353–368, 2011.

[8] N. L. Zhang, R. Zhang, and T. Chen, "Discovery of regularities in the use of herbs in traditional chinese medicine prescriptions," in *New Frontiers in Applied Data Mining*. Berlin, Germany: Springer, 2012, pp. 353–360.

[9] P. He, K. Deng, Z. Liu, D. Liu, J. S. Liu, and Z. Geng, "Discovering herbal functional groups of traditional chinese medicine," *Statist. Med.*, vol. 31, no. 7, pp. 636–642, 2012.

[10] L. Yao, Y. Zhang, and B. Wei, "An evolution system for traditional chinese medicine prescription," in *Knowledge Engineering and Management*. Berlin, Germany: Springer, 2014, pp. 95–106.

[11] G. Zheng, M. Jiang, C. Lu, and A. Lu, "Prescription analysis and mining," in *Data Analytics for Traditional Chinese Medicine Research*. Berlin, Germany: Springer, 2014, pp. 97–109.

[12] X. Zhang, X. Zhou, H. Huang, S. Chen, and B. Liu, "A hierarchical symptom-herb topic model for analyzing traditional chinese medicine clinical diabetic data," in *Proc. 3rd Int. Conf. Biomed. Eng. Inf.*, 2010, pp. 2246–2249.

[13] X. Zhang, X. Zhou, H. Huang, Q. Feng, S. Chen, and B. Liu, "Topic model for chinese medicine diagnosis and prescription regularities analysis: Case on diabetes," *Chin. J. Integrative Med.*, vol. 17, pp. 307–313, 2011.

[14] Z. Jiang, X. Zhou, X. Zhang, and S. Chen, "Using link topic model to analyze traditional chinese medicine clinical symptom-herb regularities," in *Proc. IEEE 14th Int. Conf. E-Health Netw., Appl.. Serv.*, 2012, pp. 15–18.

[15] L. Yao, et al."Discovering treatment pattern in traditional chinese medicine clinical cases by exploiting supervised topic model and domain knowledge," *J. Biomed. Inf.*, vol. 58, pp. 260–267, 2015.

[16] Z. Deng, *Formulae of Chinese Medicine*. Hong Kong: China Press of Traditional Chinese Medicine, 2008, [in Chinese, ISBN: 9787532384761].

[17] Y. Huang, et al., "Exploring the rules of li-fa-fang-yao on diabetes mellitus within traditional chinese medicine through text mining," in *Proc. 7th Int. Conf. Comput. Convergence Technol.*, 2012, pp. 1369–1373.

[18] S. Wang et al., "Compatibility art of traditional chinese medicine: From the perspective of herb pairs," *J. Ethnopharmacology*, vol. 143, no. 2, pp. 412–423, 2012.

[19] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[20] L. Fei-Fei and P. Perona, "A Bayesian hierarchical model for learning natural scene categories," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 524–531.

[21] E. Bart, M. Welling, and P. Perona, "Unsupervised organization of image collections: Taxonomies and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2302–2315, Nov. 2011.

[22] Z. Huang, W. Dong, P. Bath, L. Ji, and H. Duan, "On mining latent treatment patterns from electronic medical records," *Data Mining Knowl. Discovery*, vol. 29, no. 4, pp. 914–949, 2015.

[23] D. Andrzejewski, X. Zhu, and M. Craven, "Incorporating domain knowledge into topic modeling via dirichlet forest priors," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 25–32.

[24] D. Andrzejewski, X. Zhu, M. Craven, and B. Recht, "A framework for incorporating general domain knowledge into latent dirichlet allocation using first-order logic," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, vol. 22, no. 1, 2011, Art. no. 1171.

[25] R. Balasubramanyan, B. Dalvi, and W. W. Cohen, "From topic models to semi-supervised learning: Biasing mixed-membership models to exploit topic-indicative features in entity clustering," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2013, pp. 628–642.

[26] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT press, 2009.

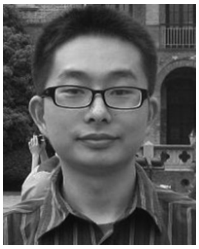
[27] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.

[28] J. K. Pritchard, M. Stephens, and P. Donnelly, "Inference of population structure using multilocus genotype data," *Genetics*, vol. 155, no. 2, pp. 945–959, 2000.

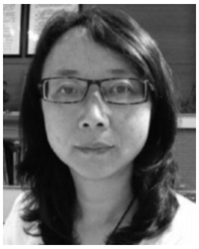
[29] X. Chen, T. He, X. Hu, Y. An, and X. Wu, "Inferring functional groups from microbial gene catalogue with probabilistic topic models," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2011, pp. 3–9.

[30] A. Van Esbroeck, C.-C. Chia, and Z. Syed, "Heart rate topic models," in *Proc. 26th AAAI Conf. Artif. Intell.*, 2012, pp. 1635–1641.

[31] Z. Huang, W. Dong, L. Ji, C. He, and H. Duan, "Incorporating comorbidities into latent treatment pattern mining for clinical pathways," *J. Biomed. Inform.*, vol. 59, pp. 227–239, 2016.
[32] I. Yoo, et al., "Data mining in healthcare and biomedicine: A survey of the literature," *J. Med. Syst.*, vol. 36, no. 4, pp. 2431–2448, 2012.
[33] N. Esfandiari, M. R. Babavalian, A.-M. E. Moghadam, and V. K. Tabar, "Knowledge discovery in medicine: Current issue and future trend," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4434–4463, 2014.
[34] Y. Feng, Z. Wu, X. Zhou, Z. Zhou, and W. Fan, "Knowledge discovery in traditional chinese medicine: State of the art and perspectives," *Artif. Intell. Med.*, vol. 38, no. 3, pp. 219–236, 2006.
[35] S. Lukman, Y. He, and S.-C. Hui, "Computational methods for traditional chinese medicine: A survey," *Comput. Methods Programs Biomed.*, vol. 88, no. 3, pp. 283–294, 2007.
[36] B. Liu, et al., "Data processing and analysis in real-world traditional chinese medicine clinical data: Challenges and approaches," *Statist. Med.*, vol. 31, no. 7, pp. 653–660, 2012.
[37] G.-Z. Li and B.-Y. Liu, "Big data is essential for further development of integrative medicine," *Chin. J. Integrative Med.*, vol. 21, pp. 323–331, 2015.
[38] N. L. Zhang, S. Yuan, T. Chen, and Y. Wang, "Latent tree models and diagnosis in traditional chinese medicine," *Artifi. Intell. Med.*, vol. 42, no. 3, pp. 229–245, 2008.
[39] E. Erosheva, S. Fienberg, and J. Lafferty, "Mixed-membership models of scientific publications," *Proc. Nat. Acad. Sci. USA*, vol. 101, no. suppl 1, pp. 5220–5227, 2004.
[40] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proc. 20th Conf. Uncertainty Artif. Intell.*, 2004, pp. 487–494.
[41] N. Yao, J. Zhu, and R. Gao, *Traditional Chinese Medicine Symptoms Differential Diagnosis*. Beijing: People Health Press, 2013, [in Chinese, ISBN: 9787117036139].
[42] L. Wu, *Chin. Traditional Medicine and Materia Medical Subject Headings*. Beijing: Chinese Medical Ancient Books Publishing, 1996, [in Chinese, ISBN: 9787801743596].
[43] R. Balasubramanyan and W. W. Cohen, "Block-LDA: Jointly modeling entity-annotated text and entity-entity links," in *Proc. SIAM Int. Conf. Data Mining*, 2011, pp. 450–461.
[44] R. Nallapati and W. W. Cohen, "Link-plsa-LDA: A new unsupervised model for topics and influence of blogs," in *Proc. Int. Conf. Weblogs Social Media*, 2008, pp. 84–92.
[45] S. Amer-Yahia, S. B. Roy, A. Chawla, G. Das, and C. Yu, "Group recommendation: Semantics and efficiency," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 754–765, 2009.
[46] Q. Yuan, G. Cong, and C.-Y. Lin, "Com: A sgenerative model for group recommendation," in *Proc. 20th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2014, pp. 163–172.
[47] T. Wu, G. Qi, H. Wang, K. Xu, and X. Cui, "Cross-lingual taxonomy alignment with bilingual biterm topic model," in *Proc. AAAI Conf. Artif. Intell.*, 2016, pp. 287–293.
[48] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno, "Evaluation methods for topic models," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 1105–1112.
[49] A. Asuncion, M. Welling, P. Smyth, and Y. W. Teh, "On smoothing and inference for topic models," in *Proc. 25th Conf. Uncertainty Artif. Intell.*, 2009, pp. 27–34.
[50] C. Archambeau, B. Lakshminarayanan, and G. Bouchard, "Latent IBP compound dirichlet allocation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 2, pp. 321–333, Feb. 2015.
[51] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-Graber, and D. M. Blei, "Reading tea leaves: How humans interpret topic models," in *Advances Neural Inf. Process. Syst.*, 2009, pp. 288–296.



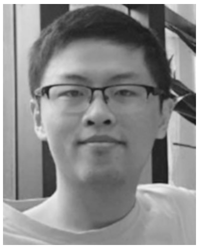
Liang Yao received the BE degree from the College of Computer Science, Sichuan University, Chengdu, China, in 2012. He is currently working toward the PhD degree from the College of Computer Science and Technology, Zhejiang University. His current research interests include data mining, medical informatics, natural language processing, topic models, and probabilistic graphical models.



Yin Zhang received the PhD degree in computer science from Zhejiang University, in 1999. Currently, she is an associate professor with the College of Computer Science, Zhejiang University. Her research interests mainly include data mining and knowledge discovery, medical informatics, multimedia information processing, pattern recognition, and knowledge engineering.

Baogang Wei received the PhD degree from Northwestern Polytechnical University, China, in 1997. He is currently a professor with Zhejiang University, China. His main research interests include artificial intelligence, pattern recognition, image processing, machine learning, digital library, and information & knowledge management. Since 1999, he has been a member of the Chinese Association for Artificial Intelligence. So far, he has published more than 50 papers in international journals including the *IEEE Transactions on Knowledge and Data Engineering* and *IEEE Transactions on Visualization and Computer Graphics*, and conference proceedings including AAAI, CIKM, and PAKDD.

Wenjin Zhang received the graduate degree from the School of Medicine, Zhejiang University, in 1990 and received the PhD degree, in 2005. Now, he is an associate chief physician in the Hepatopancreatobiliary Surgery Department of the first affiliated Hospital of Zhejiang University. In addition to clinical work, he is also engaged in medical research involving experimental science and data analysis by IT technology.



Zhe Jin received the BE degree from the College of Computer Science and Technology, Zhejiang University, Hangzhou, China, in 2015, where he is currently a PhD candidate at the same college. His current research interests include text mining and semantic network.