MtiRec: A Medical Test Recommender System based on the Analysis of Treatment Programs

Nengjun Zhu*, Jieyun Huang
School of Computer Engineering and Science
Shanghai University, China
{zhu_nj, huang0615}@shu.edu.cn

Jian Cao
Department of Computer Science and Engineering
Shanghai Jiao Tong University, China
cao-jian@sjtu.edu.cn

Xinjiang Lu
Business Intelligence Lab
Baidu Research, China
luxinjiang@baidu.com

Hao Liu, Hui Xiong
The Hong Kong University of Science and Technology (Guangzhou), China
{liuh,xionghui}@ust.hk

***Abstract*—Medical tests are crucial for treatment decision making. However, over-testing can often occur in any medical speciality or level of expertise. Since over-testing usually results in a financial burden for patients and is also a waste of medical resources, this naturally leads to the question: which medical test items (MTIs) are necessary and should be prioritized for the target patients? It is a nontrivial task to identify the right MTIs due to the diversified health status of patients and the complicated prerequisites of therapies. To this end, in this paper, we propose a data-driven approach to evaluate the priority which should be given to MTIs by modeling the relationships between MTIs and therapies. Specifically, we first develop a dual hierarchical topic model (DHTM), which views the adopted hierarchical therapies as labeled topics and the MTI reports, i.e., the set of hierarchical attribute-value pairs (AVPs), as documents. Then, with the therapy-AVP distribution and the partial MTI reports of the target patient, we can scope the candidate therapies, which are further utilized to evaluate the accumulated gain of MTIs to be tested. Moreover, the next MTI recommendation is conducted based on the gains. Finally, extensive experiments on real-world medical data validate the effectiveness of our approach, and some interesting observations are also provided. The code is available at https://github.com/mtirec/MtiRec.**

***Index Terms*—Medical test, recommender system, hierarchical topic model, attribute analysis**

I. Introduction

Medical tests, which are conducted in various medical facilities to evaluate a patient's health status, are crucial in assisting medical professionals with therapy decisions. With the development of medical technology, on the one hand, various medical test items (MTIs), such as gene detection, can be utilized. On the other hand, over-testing can often occur in any medical speciality or level of expertise. Since over-testing usually results in a financial burden for patients and is also a waste of medical resources, the adopted MTIs should be customized and deliberated according to a patient's situation.

MTI decision-making is a nontrivial task since it depends on many factors, such as the diversified health status of patients, the complicated prerequisites of therapies, and even the patient's willingness and financial ability. Therefore, the decision can only be determined with medical professionals involved. But we can utilize machine learning technologies to learn valuable experience from historical cases and then recommend MTIs to medical professionals.

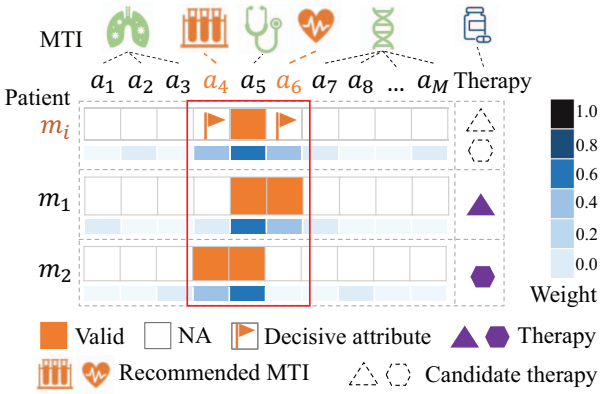


Fig. 1: An example to illustrate MTI recommendations.

In this paper, we aim our task at the next MTI recommendation. It usually occurs when medical professionals require more evidence from medical tests for determining therapies. At this time, patients already have partial MTI reports, which are insufficient. Thus, the other MTIs are needed, and their priorities are highly related to candidate therapies. The partial MTI reports, which can be parsed into several attribute-value pairs (AVPs), are utilized to evaluate possible therapies. Based on the results and obtained AVPs, we can recommend the next MTIs. For example, in Figure 1, one MTI is related to multiple attributes, and one attribute comprises several values, i.e., AVPs. The valid attribute means the corresponding MTI has already been conducted. Then, patient m_i's situation is similar to m_1 and m_2, and thus will be offered similar therapies. To confirm the decision, the missing but important AVPs are identified and utilized to recommend MTIs.

* Corresponding author

2374-8486/23/$31.00 ©2023 IEEE
DOI 10.1109/ICDM58522.2023.00099

However, the partial and inconsistent MTI reports cause conventional data-driven approaches hard to achieve our task. First, existing approaches require all attributes to be as non-null as possible [1], [2]. But one non-empty attribute may imply a high-cost MTI. Thus it is sparse compared to the massive volume of optional attributes. Second, there are no satisfactory *feature-label* samples for conventional supervised learning since the target patient has no complete features or labeled therapies. Third, the weights of attributes are usually the same for all examples in these approaches. Thus, we cannot customize MTIs by analyzing the weights of attributes.

Furthermore, existing methods usually consider the correlation between labels in a flat way, ignoring the hierarchical structure and interactions between treatments. Therapies (i.e., labels) usually have a hierarchical structure, for example, *Need Chemotherapy* is a coarse-grained category while *Need CMF Chemotherapy* is a fine-grained therapy. Therapies in the same category usually have exclusive effects. Conversely, therapies in different categories might have complementary effects. However, the flattened approaches neglect these mutual effects.

To this end, we propose MtiRec (i.e., the **m**edical **t**est **i**tem **rec**ommender system), to evaluate the priority given to MTIs. The approach consists of two steps. First, we model the distribution between therapies and AVPs using a proposed dual hierarchical topic model (DHTM). DHTM views the AVP as a word, the set of AVPs as a document, and the determined therapies as labels for a patient. Second, according to the therapy-AVP distribution and the already-known AVPs of the target patient, we scope the candidate therapy combinations, which are further utilized to evaluate the weight of the empty attributes by backward reasoning. We thus can recommend MTIs according to the accumulated weights of these attributes.

MtiRec has several advantages. First, it can handle sparse and randomly combined attributes since it packs AVPs as a document. MtiRec doesn't fill out empty attributes or align features. It thus can avoid artificial noise. Second, MtiRec models the co-occurrence relation between the therapies, and models the hierarchical information of AVP and therapies simultaneously. This property is vital since medication dependencies and adverse drug interactions always exist in medical domains. Third, MtiRec has great interpretability, which is important in medical domains [3] since it can reveal the principles behind each step in the decision making.

The main contributions are summarized as follows:

1) We propose a next medical test item recommendation approach, i.e., MtiRec, in the perspective of assisting therapy decision making.

2) We propose DHTM based on the analysis of large-scale medical treatment data, which effectively models the relations between hierarchical AVPs and therapies.

3) We evaluate MtiRec with real-world medical data. The experimental results prove the effectiveness of the approach. We also reveal some interesting discoveries.

II. Preliminaries

A. Definitions

*Definition 1 (**Attribute**):* An attribute l is one dimension recording a patient's information. Each attribute has a name a_l and some possible values $v_{l,*}$. For example, BLOOD TYPE is an attribute, and the value of this attribute can be TYPE O or TYPE B. Patients may have different values for these attributes. When the attribute is non-empty, it is called **valid attribute** for a patient.

*Definition 2 (**Attribute-Value Pair (AVP)**):* One attribute l can comprise multiple AVPs $[a_l, v_{l,*}]$, such as [BLOOD TYPE, O] and [BLOOD TYPE, B]. Obviously, each valid attribute of a patient can only generate one AVP.

*Definition 3 (**Medical Test Item (MTI)**):* By using medical test facilities, an MTI t can assess the patient's physical status and disclose the value of attributes, i.e., $t \xrightarrow{check} \{l_1^t, l_2^t \cdots, l_i^t \cdots\}$. Thus, the report of t can be parsed into a set of AVPs $r_t = \{[a_{l_1}^t, v_{l_1,*}^t], [a_{l_2}^t, v_{l_2,*}^t] \cdots, [a_{l_i}^t, v_{l_i,*}^t] \cdots\}$.

*Definition 4 (**Therapy**):* Therapy k is a treatment program for a patient according to MTI reports $\{r_1, r_2 \cdots, r_t \cdots\} \xrightarrow{determine} k$. If the evidence from current reports is sufficient, then k is confirmed, else more MTIs are required. The treatment program can be a set of therapies, which is called a **therapy combination**. Medical tests serve in therapy decision making in this paper.

B. The Next MTI Recommendation

Our task can be viewed as a next item (i.e., MTI) recommendation problem [4]–[6]. Given a set of obtained AVPs $r_1 \cup r_2 \cdots \cup r_t \cdots$ for a patient after several medical tests, our task is to recommend the next MTI t^*, which contributes most to determine therapies.

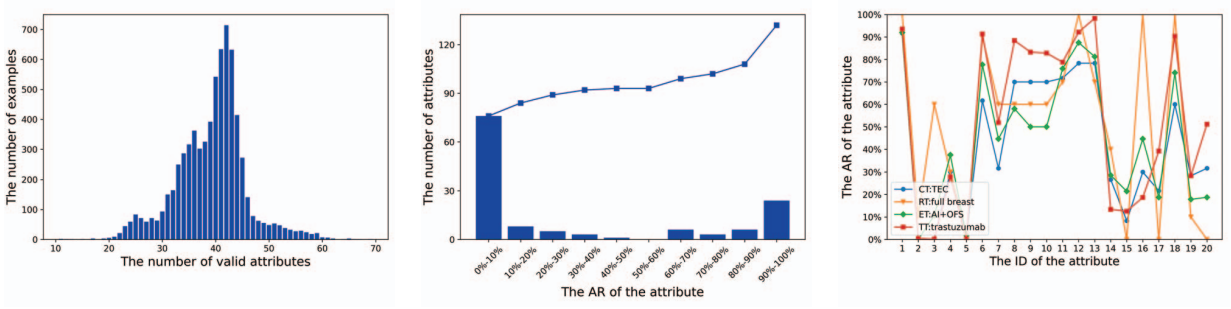
However, to simplify the problem, we assume one MTI checks one attribute, i.e., $t^* \xrightarrow{check} l^*$. Then, our task is equivalent to recommending an attribute l^* with no value yet. But it is easy to extend to one-to-many situation as the relation map is obtained. In the following, if not especially mentioned, the definition of attribute and MTI has no difference.

C. BCDB Dataset

BCDB[1] records the MTI reports and treatment therapies of patients with breast disease [7]. It contains 7,040 samples (i.e., patients). Each sample has 132 attributes (which may be null) and four types of therapies. The average number of valid attributes per example is 39.32, and the sparsity of the data is 68.14%, which is very high. Some examples of attributes and therapies are presented in Table I.

To further illustrate the characteristics of medical data, Figure 2(a) shows the distribution of examples with respect to the number of valid attributes. We can see that most examples have few valid attributes (i.e., the average is 39.32) compared to the total number of attributes (i.e., 132). This means a few customized MTIs are sufficient for making therapy decisions in

[1]It can be requested from its website, i.e., http://bcdb.mdt.team:8080



(a) The sparsity of attributes per sample. (b) The sparsity of each attribute over examples. (c) The attribute distributions per therapy.

Fig. 2: The statistical characteristics of the attributes in BCDB.

TABLE I: Example of attributes and therapies in BCDB

Attributes	Categories	Therapies		
surface area	**CT** (22)*	EC(AC)	···	CMF
aln	**RT** (7)	APBI	···	GPBI
bmi	**ET** (15)	SERM	···	AI+OFS
···	**TT** (8)	T-DM1	···	PD-1

* The number of therapies in different categories.

practice. Figure 2(b) shows the distribution of attributes with respect to the percentage of samples which have a value of such an attribute, i.e., the appearance rate (AR) of the attributes in examples. For example, there are 76 sparse attributes (i.e., AR is less than 10%). Only 24 attributes have an AR over 90% and these usually correspond to the mandatory or basic MTIs, such as *height* and *weight*. In Figure 2(c), we randomly select four therapies and twenty attributes and calculate the AR of each attribute given the therapy. The figure shows that the ARs of attributes for different therapies vary, hence it is challenging to determine which attributes are more important when considering a combination of different therapies.

III. The Proposed MtiRec Framework

Our proposed MtiRec is inspired by doctors' backward reasoning process, i.e., first evaluating the most possible candidate therapies, and then determining the MTIs to further verify the candidate therapies. Two key issues should be addressed:

1) *How to determine the candidate therapies or therapy combinations according to inconsistent MTI reports.*
2) *How to evaluate the available MTIs' priorities according to the most possible candidate therapies.*

A. Overview

In MtiRec, as shown in Figure 3, the relations between therapies and AVPs is a cornerstone to address the aforementioned issues. Thus, we first propose a dual hierarchical topic model (DHTM), which takes historical treatment data as a corpus, to learn the distribution of AVPs over different therapies.

Then, to determine the candidate therapy combinations instead of individual therapies, we combine therapies in different categories to form therapy tuples. With the relations between therapies and AVPs obtained, we can further acquire the distribution of AVPs over different therapy tuples. By doing so, the mutual interactions between different types of therapies in tuples are considered, which is vital in the medical field.

So far, we can determine the most possible candidate therapy tuples for a patient using two strategies. First, we link the patient's AVPs to all tuples and rank the tuples according to the weighted accumulation of the AVPs. The weight of each AVP is from the distribution of AVPs over different therapy tuples. Second, DHTM directly infers the distribution of therapies for a patient. We thus can rank tuples according to the combination of their therapies' probability values. These two strategies can be fused to determine more accurate and robust candidate tuples.

Finally, with the distribution of AVPs over candidate tuples and the tuple's probabilities, we can evaluate the importance of each AVP for the patient and the recommendation score of an attribute is calculated as the accumulated weights of its all possible AVPs. The following subsections introduce each part in detail and Table II lists the related notations.

TABLE II: Mathematical Notations

Symbol	Description
M	The number of patients
K	The number of therapies
C	The number of the therapy categories
N	The number of AVPs
N_m	The number of AVPs of patient m's document
γ	Bernoulli prior for generating Λ
ζ	Bernoulli prior for generating ε
α	Dirichlet prior for generating θ
β	Dirichlet prior for generating ϕ
δ	Dirichlet prior for generating π
Λ_k^m	Present/absent of the therapy k of patient m
ε_c^m	Present/absent of the therapy category c of patient m
θ_m	Distribution over therapies for patient m
ϕ_k	Distribution over AVPs for therapy k
π_m	Distribution over attributions for patient m

B. The Dual Hierarchical Topic Model (DHTM)

Inspired by a labeled topic model [8], we develop DHTM, which considers the hierarchical structure of both attributes

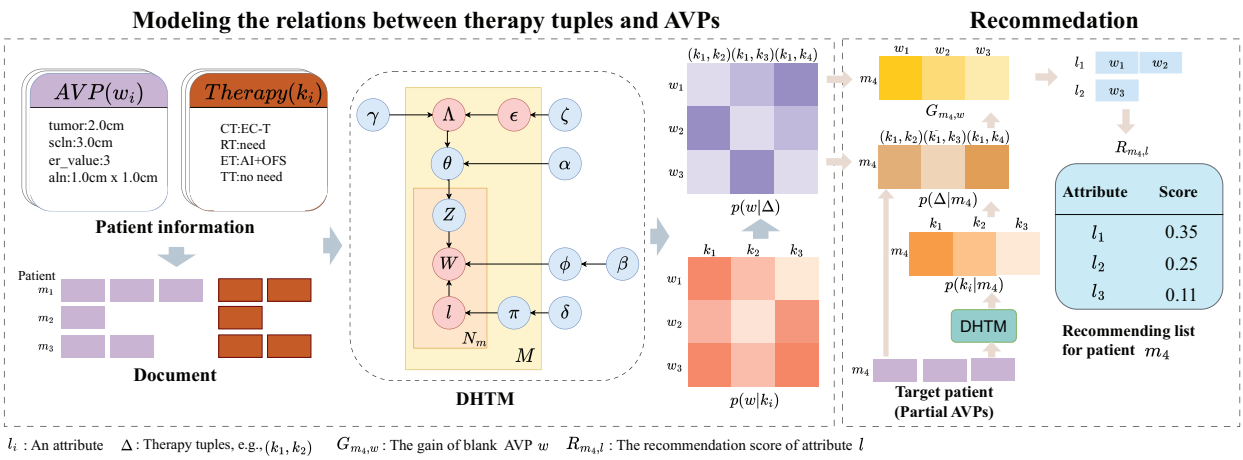


Fig. 3: An overview of our approach for MTI recommendation.

and therapies. We view the AVP as a word and the set of AVPs as a document. Naturally, each patient has a "document" describing their health status. Accordingly, the therapies of a patient can be viewed as the explicit topics (i.e., labels) of the "document". Thus, the topic model can be applied to learn two types of distributions: the patient-therapy (i.e., document-topic) distribution and the therapy-AVP (i.e., topic-word) distribution. Similar to the work [8], we assume both of the distributions follow a multinomial distribution.

We utilize $\phi_k \in \mathcal{R}^N$ to represent the distribution over AVPs for therapy k, and $\theta_m \in \mathcal{R}^K$ to represent the distribution over therapies for patient m, where N is the number of AVPs and K is the number of therapies. Then, the selection of AVPs for a patient is not only affected by the therapy but also the attribute, thus we assume the i-th AVP of the m-th patient obeys $w_i^m \sim Multi\left(\cdot|z_i^m, l_i^m\right)$, where z_i^m and l_i^m are the therapy and attribute of the i-th AVP, respectively.

Furthermore, there is a map matrix between therapy categories and therapies $\mathcal{M} \in \{0,1\}^{C\times K}$, where C is the number of the categories. Only when the element $\mathcal{M}_{c,k} = 1$, the therapy k belongs to the category c. When prescribing therapies for a patient m, DHTM first generates therapy categories $\varepsilon^m \in \{0,1\}^C$, and only when $\varepsilon_c^m = 1$, the category c is selected, which affects the selection of therapies. We have another binary vector $\Lambda^m \in \{0,1\}^K$, in which the element Λ_k^m indicates whether therapy k should be prescribed to patient m without considering the therapy categories. Finally, we limit the scope of therapies and derive the distribution over therapies for patient m by $Dir\left(\cdot|\alpha \times \varepsilon^m \mathcal{M} \times \Lambda^m\right)$, where α is the prior distribution of therapies and set as $50/K$ for simplicity. Algorithm 1 illustrates the details of DHTM.

We use collapsed Gibbs sampling for training [9], where the sampling probability of z_i^m, which indicates the therapy of the i-th AVP of the m-th patient, is given by:

$$p\left(z_i^m = k \mid z_{-i}^m, w_i^m, l_i^m, c_{-i}^m\right) \propto p\left(z_i^m = k \mid z_{-i}^m, c_{-i}^m\right) \cdot p\left(w_i^m, l_i^m \mid z_i^m = k, z_{-i}^m, w_{-i}^m, l_{-i}^m\right) \quad (1)$$

Algorithm 1 The Generation Process of DHTM

Input: Patients' AVPs and their therapies
Output: The therapy-AVP distribution ϕ and patient-therapy distribution θ

for each therapy $k \in \{1, 2, \cdots, K\}$ **do**
  Generate $\phi_k \sim Dir\left(\cdot|\beta\right)$
end for
for each patient $m \in \{1, 2, \cdots, M\}$ **do**
  for each therapy category $c \in \{1, 2, \cdots, C\}$ **do**
    Generate $\varepsilon_c^m \in \{0, 1\} \sim Bernoulli\left(\cdot|\zeta\right)$
  end for
  for each therapy $k \in \{1, 2, \cdots, K\}$ **do**
    Generate $\Lambda_k^m \in \{0, 1\} \sim Bernoulli\left(\cdot|\gamma\right)$
  end for
  for each therapy k belongs to therapy category c **do**
    $\Lambda_k^m = \Lambda_k^m * \varepsilon_c^m$
  end for
  Generate $\theta_m \sim Dir\left(\cdot|\alpha \times \Lambda_m\right)$
  Generate $\pi_m \sim Dir\left(\cdot|\delta\right)$
  for each AVP $i \in \{1, 2, \cdots, N_m\}$ **do**
    Generate $z_i^m \sim Multi\left(\cdot|\theta_m\right)$
    Generate $l_i^m \sim Multi\left(\cdot|\pi_m\right)$
    Generate $w_i^m \sim Multi\left(\cdot|z_i^m, l_i^m\right)$
  end for
end for

where w_i^m denotes the i-th AVP of patient m, w_{-i}^m denotes all AVPs that patient m has except the i-th one; z_i^m denotes the therapy assigned to AVP w_i^m; z_{-i}^m means all therapies belonging to patient m except the therapy of AVP w_i^m; c_{-i}^m indicates the therapy categories assigned to patient m except the therapy category assigned to AVP w_i.

The first term in the right part of Equation (1) represents the conditional probability of therapy k given the therapies and therapy categories of other AVPs. It can be further calculated

as follows:

$$p\left(z_i^m = k \mid z_{-i}^m, c_{-i}^m\right) = \frac{\left(n_{m,k,-(m,i)} + \alpha_k\right) \cdot \Lambda_k^m \cdot \varepsilon_c^m}{N_m - 1 + \sum_{j=1}^{K} \alpha_j \cdot \Lambda_j^m \cdot \varepsilon_c^m} \quad (2)$$

where $n_{m,k,-(m,i)}$ is the number of therapy k of patient m, but eliminating the current assigned therapy; N_m is the total number of AVPs of patient m.

When therapy k is assigned to z_i^m, the probability of AVP w_i^m along with its attribute l_i^m can be estimated. This is the second term in the right part of Equation (1) and it can be realized as:

$$p\left(w_i^m, l_i^m \mid z_i^m = k, z_{-i}^m, w_{-i}^m, l_{-i}^m\right) = \frac{n_{w_i^m, l_i^m, k, -(m,i)} + \beta_{w_i}}{\sum_{s=1}^{N}\left(n_{w_s, l_s, k, -(m,i)} + \beta_{w_s}\right)} \cdot \frac{n_{m, l_i^m, -(m,i)} + \delta_{l_i^m}}{\sum_{l'=1}^{L}\left(n_{m,l',-(m,i)} + \delta_{l'}\right)} \quad (3)$$

where l_s represents the attribute l to which AVP s belongs; $n_{w_s,l_s,k,-(m,i)}$ is the number of AVP w_s under attribute l_s and therapy k by eliminating the i-th AVP of patient m; $n_{m,l_i^m,-(m,i)}$ is the number of attribute l_i^m belonging to the patient m by eliminating the current attribute assigned to AVP w_i; β and δ are hyper-parameters.

When the training process of DHTM has been completed, we can obtain two primary distributions, i.e., ϕ_k and θ_m, related to therapy-AVP pairs and patient-therapy pairs, respectively. Compared to conventional approaches, DHTM has the ability to handle the hierarchical structure of AVPs and therapies, simultaneously. These structures commonly appear in medical data. Recall the definition of AVP, it represents the specific value of an attribute, and thus two AVPs belonging to the same attribute cannot appear in the same "document" according to our settings. Furthermore, the combination of therapies in the same category is packaged as a joint therapy. Thus, doctors cannot prescribe multiple therapies belonging to the same category at the same time either. Therefore, the consideration of the constraint behind these hierarchical structures has a positive effect on distribution learning.

C. AVP Distribution over Therapy Tuples

There are mutual interactions between different types of therapies. Thus, it is necessary to consider their packages. Since the same type of therapies cannot be packed as previously discussed, we select therapies in different categories to form the combinations, i.e., therapy tuples $\Delta = \{k_i \mid i = 1, 2, 3 \ldots\}$, where k_i is a specific therapy and belongs to different categories.

The weight of AVP w given a therapy tuple is calculated:

$$p(w|\Delta) = \sum_{k_i \in \Delta} p(w|k_i)p(k_i) \quad (4)$$

where $p(w|k_i)$ is the distribution over AVPs of each therapy k_i, which can be estimated by the trainable parameter of DHTM, i.e. ϕ_k, and $p(k_i)$ is the probability that we select k_i. We estimate $p(k_i)$ by calculating the proportion of therapy k_i in the samples:

$$p(k_i) = \frac{n_{k_i}}{M} \quad (5)$$

where n_{k_i} is the number of occurrences of therapy k_i and M is the number of all patients in the samples. Then, Equation (4) can be reformulated as:

$$p(w|\Delta) = \frac{1}{M} \sum_{k_i \in \Delta} n_{k_i} \phi_{k_i, w} \quad (6)$$

We rank AVPs according to the value of $p(w|\Delta)$. Then, the top-N AVPs, which are denoted by symbol W_Δ, are selected as principal components (PCs) of Δ.

D. The Choice of Therapy Tuples

We can calculate the relations between patient m and therapy tuple Δ according to their co-occurrence AVPs as:

$$p(\Delta|m) = \frac{1}{Z} \sum_{w_i \in W_\Delta} \sum_{w_j \in W_m} \mathbb{I}(w_i = w_j) p(w_i|\Delta) \quad (7)$$

where W_m represents the set of AVPs of patient m, $\mathbb{I}(x)$ is an indicator function: $\mathbb{I}(x) = 1$ when x is true, otherwise $\mathbb{I}(x) = 0$, and Z is a normalization term as follows:

$$Z = \sum_{w_i \in W_\Delta} p(w_i|\Delta) \quad (8)$$

According to Equation (7), if patients have more important AVPs in W_Δ, then they have a tighter connection with the therapy tuple. Some AVPs in W_m and W_Δ may not be identical, but they could belong to the same attributes. For these cases, we can have a trade-off factor $\omega \in [-1, 1]$ to further identify their influence. When $\omega < 0$, these cases would decrease the connection of the therapy combination to the patient, and otherwise the connection is enhanced. Thus, the relation $p(\Delta|m)$ can be updated:

$$p(\Delta|m) = \frac{1}{Z} \sum_{w_i \in W_\Delta} \sum_{w_j \in W_m} \mathbb{I}(w_i = w_j) p(w_i|\Delta) + \omega * \mathbb{I}(l_{w_i} = l_{w_j} \wedge w_i \neq w_j) p(w_i|\Delta) \quad (9)$$

The relation between therapies in a tuple is considered in the above equations. If assuming the independence between therapies for a patient, we can also calculate the relations between the patient and the therapy tuple from another perspective:

$$p'(\Delta|m) = \sum_{k_i \in \Delta} p(k_i|m)p(k_i) \quad = \frac{1}{M} \sum_{k_i \in \Delta} n_{k_i} \theta_{m,k_i} \quad (10)$$

Recall that we have a trainable parameter θ_m of DHTM. We thus can utilize it to evaluate $p(k_i|m)$. We utilize the global $p(k_i)$ as a prior distribution to smooth the distribution $p(k_i|m)$.

We further combine these two kinds of connections by a weighted fusion of Equation (9) and Equation (10) as follows:

$$p(\Delta|m) = \eta p(\Delta|m) + (1 - \eta) p'(\Delta|m) \quad (11)$$

where $\eta \in [0, 1]$ is another trade-off parameter and is tuned according to the experiments. Finally, we can determine the

candidate therapy tuples for a patient based on their value of $p(\Delta|m)$.

E. Recommendation

Doctors usually make a decision using backward reasoning, i.e., determining MTIs based on candidate therapies. Inspired by this process, we design the following method to assess the recommendation score of an MTI for a patient. We first calculate the gain $G_{m,w}$ of each blank AVP w by considering the relations between the AVP and the patient through different therapy tuples:

$$G_{m,w} = \sum_{\Delta_i} p(\Delta_i|m) \sum_{w \in W_{\Delta_i} - W_m} p(w|\Delta_i) \tag{12}$$

Since one attribute only accepts one AVP and excludes the others belonging to it, we calculate the accumulated gain of null attribute l according to how many AVPs it can reveal and eliminate. The definition of this process is given by:

$$R_{m,l} = \sum_{w \in l} G_{m,w} \tag{13}$$

Finally, we recommend the unadopted MTIs l^* (i.e., null attributes) with the highest scores of $R_{m,l}$ to patient m.

IV. Experiment

A. Experimental Setup

To better organize the AVPs for patients, we first partition several continuous numerical values into various scopes as discrete variables. Then, we randomly select 20% of patients to form a test set, and the remaining are used as a training set. For each patient in the test set, we further randomly remove **one** AVP and all therapy information. In our setting, the attributes of the removed AVPs are the MTIs that should be recommended.

Evaluation Metrics: We apply two common metrics, namely Accuracy@K and mean reciprocal rank (MRR), to evaluate the performance of approaches. They are introduced as follows, where the larger the value of the metrics, the better the performance of the approaches.

Accuracy@K is calculated as the proportion of the hit MTIs to all patients in the test set. The case is hit if the removed MTI is in the top-K recommendations.

MRR evaluates the position of ground truth in the recommendation list. It is calculated as follows: for patient m in the test set, m_i represents the ground truth position in the recommendation list, and $\frac{1}{m_i}$ is the score. Then, we sum the scores for all patients and divide them by the number of patients in the test set.

B. Baselines

We compare our MtiRec with several conventional supervised approaches, i.e., Adaboost, LR, DPCNN, and EANet, which can reveal the attributes' importance using different strategies. However, to better utilize them, we treat each therapy as a supervised label and construct a multi-hot vector for the set of AVPs. Each element in the vector indicates whether the patient has the corresponding AVP. By doing so, the feature of samples can be aligned. Besides, to further test the effect of DHTM in MtiRec, we replace DHTM with its prototype, i.e., LLDA, to obtain a variant of MtiRec.

Adaboost [10] integrates a set of Decision Trees (DTs) in a weighted manner. Iterative Dichotomiser 3 (ID3), which selects the attributes with the maximum information gains (IG), is a standard algorithm for creating DTs. In Adaboost, IGs are the weights of AVPs. Then, the weight of an attribute is calculated as the sum of its AVPs' weights. We finally recommend the attributes with the largest weights to patients.

LR [11] views the non-negative coefficient as a learnable parameter, and we utilize the Karush-Kuhn-Tucker conditions [12] for the non-negative least squares problem. The coefficients of LR are the weights of the AVPs, and we then calculate the weights of attributes and recommend attributes according to the method in Adaboost.

DPCNN [13] is a deep but low-complexity network architecture, as the computation time per layer decreases exponentially in a *pyramid shape*. DPCNN has a learnable weight matrix between the multi-hot vector of AVPs and different therapies. By summing the weights of all therapies for each AVP, we can obtain the final weights of AVPs. Then, we calculate the weights of attributes and recommend attributes according to the method in Adaboost.

EANet [1] is a recently introduced attention-based neural network, which can be implemented by simply using two cascaded linear layers and two normalization layers. We feed the learnable weight matrix of DPCNN to EANet through two linear layers and supervised by therapies. The weights of the corresponding linear layers are viewed as the distribution of AVPs over therapies. Then, like DPCNN, we first sum the weights of all therapies for each AVP, and then calculate the weights of attributes and recommend attributes according to the method in Adaboost.

MtiRec$_{LL}$ is a variant following the framework of MtiRec. But when learning the relations between therapies and AVPs, it replaces DHTM with LLDA [14]. DHTM not only inherits all properties of LLDA, but also further introduces the hierarchical information both of the attributes and therapies as aforementioned.

C. Performance Comparison

The hyper-parameters of the approaches are fine tuned to compare their best performance. For example, in Adaboost, the learning rate is set to 0.1 and the number of base estimators is set to 5. In MtiRec$_{DHTM}$ and MtiRec$_{LL}$, the default values of parameters are set as: $\omega = 0.2$, $\eta = 0.1$. Then, the comparative results are shown in Figure 4. To save figure space, MtiRec$_{DHTM}$ and MtiRec$_{LL}$ are labeled as Mti$_{DHTM}$ and Mti$_{LL}$, respectively. There are three main conclusions:

First, our MtiRec, including MtiRec$_{DHTM}$ and MtiRec$_{LL}$, is superior to the others on three metrics, i.e., Accuracy@5, Accuracy@10, and MRR. For example, MtiRec$_{DHTM}$ increases the value of Accuracy@5 by 0.47%, 7.95%, 31.09%, 47.21%, and 52.75% over MtiRec$_{LL}$, Adaboost, LR, EANet,

TABLE III: The influence of the length of W_Δ: χ

χ	1	2	3	4	5	6	7	8	9	10	13
Accuracy@5	91.62	+0.16%	+0.47%	+0.47%	+0.70%	+0.78%	+0.85%	**+0.93** %	+0.70%	+0.78%	+0.62%
ACCuracy@10	97.30	+0.15%	**+0.22**%	+0.15%	+0.15%	+0.07%	+0.00%	+0.00%	+0.00%	+0.00%	+0.07%
MRR	60.57	**+25.28**%	+24.02%	+23.20%	+19.92%	+19.69%	+12.29%	+3.50%	-2.47%	-4.98%	-11.57%

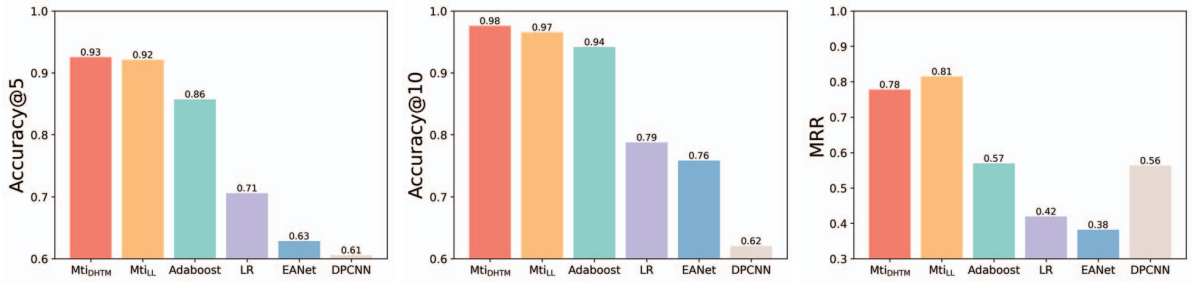


Fig. 4: Performance comparison of different approaches.

and DPCNN, respectively. The value of Accuracy@10 and MRR of MtiRec$_{DHTM}$ is also higher than the other approaches. With the exception of MtiRec$_{DHTM}$, MtiRec$_{LL}$ achieves the best performance. This observation proves the effectiveness of MtiRec.

Second, the NN-based approaches, i.e., DPCNN and EANet, do not perform better than expected. The reasons for this might be as follows: 1) The dimension of our dataset is relatively high, so a larger scale of data is needed for these approaches to achieve better performance. 2) As discussed in the introduction, the sparsity and missing values make it difficult for conventional approaches, including LR, to make MTI recommendations. 3) Neural networks are end-to-end architecture. The supervision layer of the architecture is the prediction loss of therapy. However, we aim to recommend important attributes of samples, and can only utilize the intermediate outcomes of the neural networks rather than the whole model. Thus, this inconsistent task may have a negative impact on MTI recommendations.

Third, the DHTM component can better promote the performance of MtiRec compared to the LLDA component in Accuracy@5 and Accuracy@10. It can be concluded that the hierarchical information of therapies is important when building the connection between AVPs and therapies, and the design of DHTM is effective.

D. *Parameter Analysis*

MtiRec has three main parameters (i.e., ω in Equation (9), η in Equation (11), and $\chi = |W_\Delta|$) to be tuned: ω and η are trade-off parameters adjusting the connection strength of the therapy tuple Δ to the patient; χ represents the length of W_Δ. When exploring one of three parameters, the others are set to their default values.

The upper half of Figure 5 explores the performance of MtiRec when setting ω to different values. Although we claim that the value of ω is in the range $[-1, 1]$, the experiments show that the negative values significantly impair performance. This means that different AVPs with the same attribute cannot reveal the divergence between the patient and the therapy tuple, and thus the case should not be punished. Thus, we only drew the results when $\omega \in [0, 1]$. Then, we find that the approach achieves the best performance on Accuracy@5 and Accuracy@10 when $\omega = 0.6$. However, the value of MRR decreases as the value of ω increases. The rank of recommended MTIs is very sensitive to the value of ω. A positive value of ω might bring more information to evaluate the connection strength between the patient and the therapy tuple. But the information might be noisy sometimes.

The lower half of Figure 5 shows the changes in the three metrics as the value of η increases. We find that the accuracy metrics achieve the highest performance when $\eta = 0.6$. Observing the trend of bars, Accuracy@10 is smoother than Accuracy@5, while the value of MRR decreases when η increases. Like the influence of ω, the ranks of recommended MTIs are also sensitive to the value of η.

Table III shows the impact of χ on different metrics. The table's first column is the metrics' specific value with $\eta = 1$. The remaining columns show the percentage increase compared to the first column. For example, Accuracy@5 achieves the maximum value when $\chi = 8$, and the increase in the value is 47% compared with $\chi = 1$. Besides, we find that an appropriate increase of χ can improve the performance, while an overlarge value of χ will degrade the performance.

E. *Case Study*

We randomly select several therapy tuples to visualize their most important attributes and AVPs. We calculate the popularity of each attribute and AVP according to their appearance frequency in samples. The method is called Pop for short. When the dataset is small, the statistic of attributes or AVPs related to some therapy tuples might fail as their samples need to be included. Thus, the performance of Pop is unsatisfactory. However, we can take it as an example to show the difference between our approach and the statistical approach.

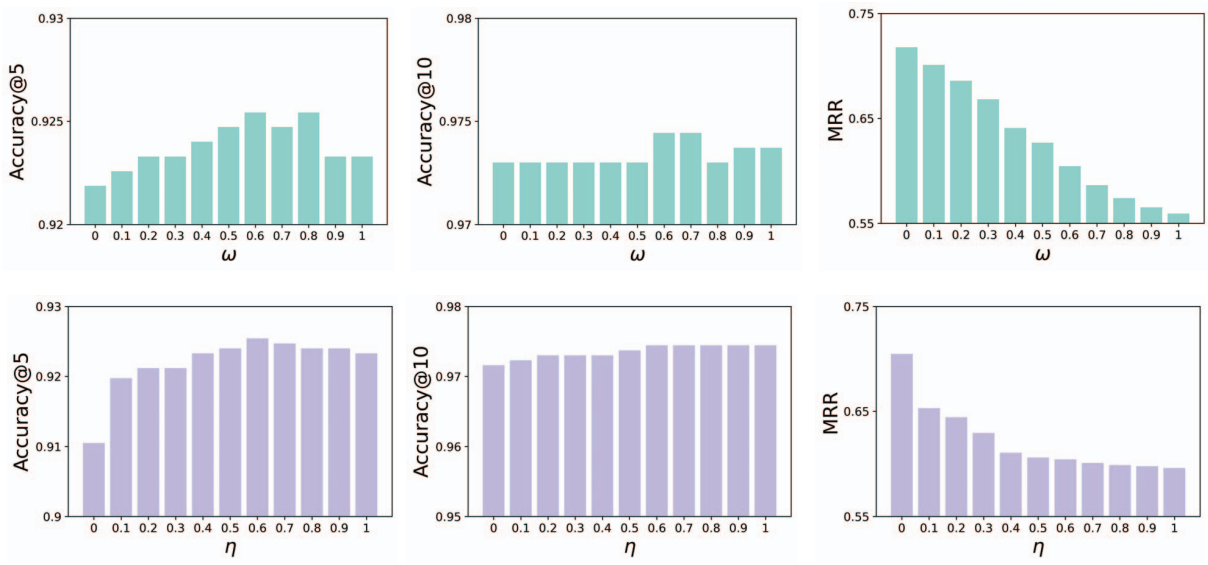
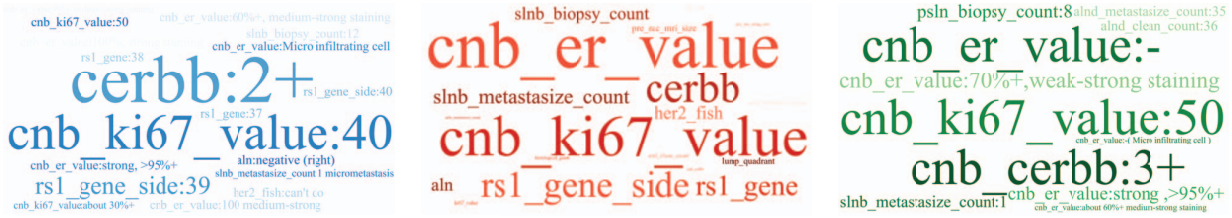


Fig. 5: The influence of parameters ω and η.



(a) Top-N AVPs of *EC(AC)* + *APBI*. (b) Top-N attributes of *EC(AC)* + *APBI*. (c) Top-N attributes of *EC(AC)*.

Fig. 6: Visualization of the most important attributes and AVPs in MtiRec given the therapy tuple.

(a) Top-N AVPs of *EC(AC)* + *APBI*. (b) Top-N attributes of *EC(AC)* + *APBI*. (c) Top-N attributes of *EC(AC)*.

Fig. 7: Visualization of the most important attributes and AVPs in Pop given the therapy tuple.

Given a therapy tuple, the importance of an AVP, i.e., $p(w|\Delta)$ in Equation (4), is produced by MtiRec, while the importance of an attribute can be expressed as $p(l|\Delta) = \sum_{w \in l} p(w|\Delta)$.

As shown in Figure 6 and 7, we select $\Delta = (EC(AC), APBI)$ as the therapy tuple. The figures visualize the most critical AVPs or attributes in MtiRec and Pop, respectively. The larger the font in the word clouds, the more critical the AVP or attribute. We make the following main observations:

First, Figure 6(a) and (b) show that *cnb_k167_value* is a key attribute when considering the therapy combination of *EC(AC)* and *APBI*. The decisive value is *40*, i.e., the AVP is *cnb_k167_value:40*. However, when exploring the individual therapy *EC(AC)*, the decisive value becomes *50*. The AVP *cnb_k167_value:50* attains the largest size, as shown in Figure 6(c). Furthermore, the top-N attributes of *EC(AC)* + *APBI* are inconsistent with those of *EC(AC)*. The observation reveals the difference in decision making for different therapy combinations. Thus, it is necessary to distinguish the attribute

checklist for different therapy tuples to make a better MTI recommendation.

Second, the difference in top-N AVPs/attributes also exists, as shown in Figure 7(a), (b) and (c). However, compared to MtiRec, the basic information, such as *age, height* and *weight*, acquire a higher rank in Pop. The reason for this is that these attributes can be easily obtained and thus this information is available for almost every patient. However, it is not necessary to recommend a required attribute. Thus, these basic but commonly used attributes contribute less when making recommendations. That is another limitation of Pop. In contrast, Figure 6(b) shows that MtiRec can rank the attributes which are more correlated with the therapy tuple to a higher position.

Third, in both MtiRec and Pop, the attributes with the most important AVPs, as shown in Figure 6(a) and Figure 7(a), may be inconsistent with the attributes which have the highest ranks in Figure 6(b) and Figure 7(b), respectively. That is reasonable since the importance of an attribute is not only related to the weights of its AVPs, it is also affected by the number of involved AVPs.

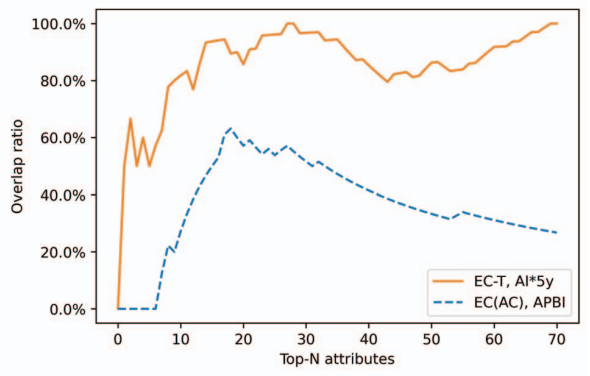


Fig. 8: The overlap ratio in the Top-N attributes.

We further evaluate the ratio of overlapped top-N attributes between MtiRec and Pop. The orange and blue broken lines in Figure 8 show the overlap ratios given the therapy tuple *(EC-T, AI*5y)* and *(EC(AC), APBI)*, respectively. In the dataset, the samples and AVPs related to *(EC-T, AI*5y)* are much more than that of *(EC(AC), APBI)*. This difference may result in larger performance fluctuations in Pop. Limited samples narrow the available attributes in Pop when calculating the appearance rate of attributes because some attributes might not appear in the scope of the statistics, i.e., AR = 0. In contrast, since MtiRec is trained on an entire training set, it can explore all attributes. The size of the samples has a minimal effect on the performance given different therapy tuples. Thus, as the number of samples increases, the diversity of available attributes in Pop approaches that in MtiRec, increasing the ratio of the overlapped top-N attributes. That is why the orange broken line is above the blue one in Figure 8.

F. *Complexity Analysis*

Our training process's time complexity mainly comprises the DHTM sampling and the calculation of relation matrices. The former is similar to traditional topical models and costs $O(M\overline{N}_m K)$ time, where M and K are the number of patients and therapies respectively, and $\overline{N}_m$ is the average number of AVPs of a patient. The latter costs $O(N + KM + K_2 + n_\Delta N log N)$ time, where N is the number of all AVPs, and n_Δ is the number of all therapy tuples.

The practical training time on an Ubuntu server having 64 cores and 128G memories is about 15 minutes for convergence, which goes through 150 iteration times.

V. Related Work

Medical Recommendation: Many works have applied personalized recommendations to healthcare. There are also several efforts like ours which focus on medical test item recommendations. For example, the work in [15] built an e-medical test recommendation system based on analyzing patients' symptoms and anamneses. Several machine learning algorithms are utilized on symptom data and demographic information to classify MTIs. Our solution is different from this work. We conduct the next MTI recommendation based on partial test reports and reveal the relations between patients' attributes and therapy tuples. Thus, the recommended results are directly related to candidate therapy combinations which is a backward reasoning approach.

Topic Model: Conventional topic models, such as LDA [16], generate words in documents according to latent topics. Many variants of LDA have been proposed [17]–[21]. Recently, modeling hierarchical information on topics has become a hot spot in the research on topic models. For example, HV-HTM [22] enables the topical tree to expand horizontally. The work exploits the Chinese restaurant process to incorporate label information into the topic-generation process. Furthermore, the work in [23] used a hierarchical Dirichlet process for topic models and proposed a stochastic variational inference algorithm. HAT [24] integrates a hierarchical attention mechanism into topic models.

The conventional topic models are primarily unsupervised. Thus, to incorporate the label information, the LLDA [14] is devised. SPTM [8] extends LLDA by incorporating the hierarchical information of words. Inspired by LLDA and SPTM, our DHTM treats the therapies of each patient as the labeled topics. However, the difference is that we design a duel hierarchical structure to model the relations between different types of therapies and the relations between attributes and AVPs, but SPTM only handles the hierarchy of attributes.

Attribute Importance Analysis: Feature selection technologies comprise three branches: filter approaches, wrapper approaches, and embedded approaches. The first ones usually evaluate the correlations between attributes and the label, and select the best attributes that can yield maximum IG for splitting the samples, such as Adaboost [10]. The second ones, such as genetic algorithms, tentatively assign weights to attributes by optimizing the objective function. Then, they

determine which is the most conducive to prediction. The embedded approaches, e.g., attention mechanisms, adjust the weights of attributes by adaptive learning. Then, the importance of attributes is analyzed according to the weights, such as DPCNN [13] and EANet [1]. An increasing number of researchers are combining the different methods by considering their benefits [25], [26].

In contrast to the aforementioned methods, our MtiRec evaluate the importance of attributes using a topic model. According to our analysis and the convincing experiment results, MtiRec could be more suitable for sparse and hierarchy medical data in our task.

VI. Conclusion

MTI recommendation is valuable in assisting medical decisions. However, medical data has high sparsity and hierarchy characteristics, making conventional approaches challenging for MTI recommendations. Thus, we proposed MtiRec, which consists of a dual hierarchical topic model (i.e., DHTM) and a backward reasoning mechanism. DHTM first produces the therapy-AVP distribution. Then, the backward reasoning mechanism recommends MTIs by analyzing the relations between most candidate therapies and MTIs. This makes our MtiRec greedily pursue the optimal MTIs, which are more helpful in therapy decision making. The experiments on a real-world medical dataset validated the effectiveness of MtiRec and also revealed some interesting observations.

There are several future directions. First, we can consider the maps between MTIs and attributes instead of treating each attribute as an independent MTI. Second, MtiRec can recommend therapies directly if the candidates have a high probability of being determined. However, how to determine the threshold probability should be well-designed in the future. Third, this work was conducted for the next MTI recommendation. We can further explore how to recommend the next package of MTIs by considering the relations between MTIs.

Acknowledgment

This work is supported by National Natural Science Foundation of China (No. 62202282) and Shanghai Youth Science and Technology Talents Sailing Program (No. 22YF1413700).

References

[1] M. Guo, Z. Liu, T. Mu, and S. Hu, "Beyond self-attention: External attention using two linear layers for visual tasks," *CoRR*, vol. abs/2105.02358, 2021. [Online]. Available: https://arxiv.org/abs/2105.02358

[2] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: Past, present, and future," *Multimedia Tools and Applications*, vol. 80, pp. 8091–8126, 2021.

[3] A. Vellido, "The importance of interpretability and visualization in machine learning for applications in medicine and health care," *Neural computing and applications*, vol. 32, no. 24, pp. 18 069–18 083, 2020.

[4] N. Zhu, J. Cao, X. Lu, and H. Xiong, "Learning a hierarchical intent model for next-item recommendation," *ACM Transactions on Information Systems*, vol. 40, no. 2, pp. 1–28, 2022.

[5] N. Zhu, J. Cao, Y. Liu, Y. Yang, H. Ying, and H. Xiong, "Sequential modeling of hierarchical user intention and preference for next-item recommendation," in *Proceedings of the ACM International Conference on Web Search and Data Mining*. ACM, 2020, pp. 807–815.

[6] C. Ma, P. Kang, and X. Liu, "Hierarchical gating networks for sequential recommendation," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 825–833.

[7] N. Zhu, J. Cao, K. Shen, X. Chen, and S. Zhu, "A decision support system with intelligent recommendation for multi-disciplinary medical treatment," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 1s, pp. 1–23, 2020.

[8] T. Xu, H. Zhu, C. Zhu, P. Li, and H. Xiong, "Measuring the popularity of job skills in recruitment market: A multi-criteria approach," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[9] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences*, vol. 101, no. suppl_1, pp. 5228–5235, 2004.

[10] T. Chengsheng, L. Huacheng, and X. Bing, "Adaboost typical algorithm and its application research," in *MATEC Web of Conferences*, vol. 139. EDP Sciences, 2017, p. 00222.

[11] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*. John Wiley & Sons, 2021.

[12] A. B. Zemkoho and S. Zhou, "Theoretical and numerical comparison of the karush–kuhn–tucker and value function reformulations in bilevel optimization," *Computational Optimization and Applications*, vol. 78, no. 2, pp. 625–674, 2021.

[13] R. Johnson and T. Zhang, "Deep pyramid convolutional neural networks for text categorization," in *Proceedings of the Association for Computational Linguistics*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 562–570.

[14] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2009, pp. 248–256.

[15] M. Ceyhan, Z. Orhan, and E. Domnori, "e-medical test recommendation system based on the analysis of patients' symptoms and anamneses," in *Proceedings of the International Conference on Medical and Biological Engineering (CMBEBIH)*. Singapore: Springer Singapore, 2017, pp. 654–659.

[16] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[17] X. Li, J. Chi, C. Li, J. Ouyang, and B. Fu, "Integrating topic modeling with word embeddings by mixtures of vmfs," in *Proceedings of the International Conference on Computational Linguistics: Technical Papers*, 2016, pp. 151–160.

[18] L. Gui, J. Leng, J. Zhou, R. Xu, and Y. He, "Multi task mutual learning for joint sentiment classification and topic detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 4, pp. 1915–1927, 2022.

[19] Y. Yang, B. Pan, D. Cai, and H. Sun, "Topnet: Learning from neural topic model to generate long stories," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2021, pp. 1997–2005.

[20] H. Choi and Y. Ko, "Using adversarial learning and biterm topic model for an effective fake news video detection system on heterogeneous topics and short texts," *IEEE Access*, vol. 9, pp. 164 846–164 853, 2021.

[21] Z. Cao, S. Li, Y. Liu, W. Li, and H. Ji, "A novel neural topic model and its supervised extension," in *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI Press, 2015, pp. 2210–2216.

[22] X. Zou, Y. Zhu, J. Feng, J. Lu, and X. Li, "A novel hierarchical topic model for horizontal topic expansion with observed label information," *IEEE Access*, vol. 7, pp. 184 242–184 253, 2019.

[23] K. Batmanghelich, A. Saeedi, K. Narasimhan, and S. Gershman, "Nonparametric spherical topic modeling with word embeddings," in *Proceedings of the Conference Association for Computational Linguistics Meeting*, vol. 2016. NIH Public Access, 2016, p. 537.

[24] X. Sun and B. Ding, "Neural network with hierarchical attention mechanism for contextual topic dialogue generation," *IEEE Access*, vol. 10, pp. 4628–4639, 2022.

[25] G. S. Thejas, S. R. Joshi, S. S. Iyengar, N. R. Sunitha, and P. Badrinath, "Mini-batch normalized mutual information: A hybrid feature selection method," *IEEE Access*, vol. 7, pp. 116 875–116 885, 2019.

[26] J. Zhang, Y. Xiong, and S. Min, "A new hybrid filter/wrapper algorithm for feature selection in classification," *Analytica Chimica Acta*, vol. 1080, pp. 43–54, 2019.