



• 不确定性知识发现 •

## 面向在线问诊平台的三支推荐方法

梁建树<sup>1</sup>, 叶晓庆<sup>2</sup>, 刘盾<sup>1</sup>

(1. 西南交通大学 经济管理学院 四川 成都 610031; 2. 西南交通大学 计算机与人工智能学院 四川 成都 611756)

**摘要:** 近年来,随着互联网和在线医疗服务的飞速发展,在线问诊平台已逐渐成为满足民众基本医疗问诊的重要渠道。与此同时,随着医患数据的飞速增长,推荐系统被逐渐应用于在线问诊平台中解决其信息过载问题。然而,现有推荐方法存在以下问题:一方面,现有推荐方法主要基于医生热度进行推荐,导致医生资源分配不均;另一方面,现有医生推荐研究主要从单粒度的视角出发,忽略了医生特征的多粒度性和多源性。基于此,该文通过融合多源数据,挖掘医生多粒度特征,提出一种面向多源数据的多步骤三支医生推荐方法(multi-stage doctor recommendation based on three-way decision, 3WD-msDRec)。首先,利用 Word2Vec 和 LDA 等技术挖掘医生特征,构建多粒度医患推荐信息;其次,引入三支决策的思想,提出多步骤三支推荐方法;最后,通过从好大夫网站中爬取的医患数据验证所提方法的有效性。实验表明,文中方法在提高推荐精度的同时,还能提高推荐系统的覆盖率。

**关键词:** 在线问诊平台; 推荐系统; 三支决策; 粒计算; 多源数据

**中图分类号:** TP391 **DOI:** 10.16152/j.cnki.xdxbr.2022-05-006

## Three – way recommendations for online medical consultation platform

LIANG Jianshu<sup>1</sup>, YE Xiaoqing<sup>2</sup>, LIU Dun<sup>1</sup>

(1. School of Economics and Management, Southwest Jiaotong University, Chengdu 610031, China;

2. School of Computer and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China)

**Abstract:** In recent years, with the rapid development of the Internet and online medical services, online consultation platform has gradually become an important channel to meet the basic medical consultation of the public. Meanwhile, with the increase of doctor and patient information, the recommendation systems have been used in the online consultation platform to solve the problem of information overload. Nevertheless, there are still some problems with the existing recommendation methods. On the one hand, existing methods generally recommend the doctors with higher popularity for the users, which leads to the uneven distribution of doctor resources. On the other hand, existing researches primarily make recommendation from the perspective of single granularity, ignoring the multi-granularity and multi-source of doctor characteristics. Based on this, this paper proposes a multi-step three-way recommendation method for online consultation platform by fusing multi-

收稿日期: 2022-07-15

基金项目: 国家自然科学基金(62276217, 61876157, 71571148, 62136002); 四川省科技计划项目(2022JDJQ0034); 重庆市计算智能重点实验室项目(2020FF03); 中央高校基本科研业务费项目(2682022ZTPY057)

第一作者: 梁建树,男,广东湛江人,从事数据挖掘、推荐系统、三支决策、粒计算研究。

通信作者: 叶晓庆,女,四川成都人,博士,从事数据挖掘、推荐系统、三支决策、粒计算研究, E-mail: xqxiaoqingye@163.com。

刘盾,男,重庆人,博士,教授,从事数据挖掘、知识发现、粗糙集、粒计算、决策支持系统研究, E-mail: new-ton83@163.com。

source data and mining the multi-granularity characteristics of doctors. Firstly, Word2Vec and LDA are used to mine the characteristics of doctors, and build multi-granularity doctor recommendation information. Secondly, the idea of three-way decisions is introduced to build a multi-step three-way recommendation. Finally, the experiments verify our proposed method can not only improve the accuracy of recommendation, but also improve the coverage of the recommendation.

**Key words:** online medical consultation platform; recommendation system; three-way decision; granular computing; multi-source data

随着“互联网+”的推广和快速发展,就医模式逐渐开始从线下往线上模式倾斜<sup>[1-2]</sup>。同时,自2020年新冠疫情爆发以来,线下就医的渠道受到极大的阻碍,“足不出户”的线上问诊模式得到更多人青睐。好大夫、丁香医生、春雨医生、阿里健康、京东健康等在线医疗平台能够为用户提供线上疾病咨询服务。作为线下医疗的补充,在线医疗一方面能合理利用全国各城市闲置的医疗资源;另一方面,能够缓解线下医院的诊疗负担,提高医生的诊疗效率<sup>[3]</sup>。在线医疗主要有在线问诊、医药零售、在线挂号等功能,其中在线问诊是解决民众就医需求的重要渠道。Ranard等研究发现在线问诊平台可以弥补传统线下就医模式带来的反馈信息匮乏问题,平台提供的信息对用户了解和评估医院的质量提供了帮助<sup>[4]</sup>。不仅如此,在线问诊平台还可以指导用户就医,让用户和医生、用户和用户之间得以互动,进而提高用户体验<sup>[5-7]</sup>。

然而,随着在线问诊平台上用户规模和医疗资源不断增长,用户想要从海量的医疗数据中寻找合适的医生变得越发困难。一方面,在线问诊平台为用户提供大量的医生信息,用户逐一通过科室、疾病等多层级的属性筛选合适的医生极其耗时;另一方面,用户不一定具有专业的医学知识,清楚自身的疾病状况以及分辨自身疾病所属的科室。基于此,作为一种解决数据过载问题的有效方法,个性化推荐技术能够为用户推荐符合其实际需求的医生,不仅能够提高在线问诊平台中用户的就诊效率,而且可以改善用户的在线体验<sup>[8]</sup>。一般而言,现有平台主要为用户推荐当前热度高、较多人访问的医生。然而,基于热度的推荐方法会造成马太效应,使得热门医生的工作超负荷,而新医生却难以被关注和发现。

为了满足用户需求,相关研究已开始将个性化推荐技术应用到医生推荐问题中,进而为用户推荐满足其需求的医生。徐守坤和吴伟根根据用

户咨询文本,通过协同过滤的方法为用户推荐满足其个性化需求的医生<sup>[9]</sup>。刘通将 Word2Vec 和 Kmeans 方法融合来解决用户咨询内容与医生聚类中心的相似度比较问题,并为用户推荐高匹配度的专业医生<sup>[10]</sup>。叶佳鑫等主要考虑了用户与医生的文本向量的相似度来推荐符合用户实际需求的医生<sup>[11]</sup>。潘有能和倪秀丽通过 Labeled-LDA 模型训练咨询文本数据,为用户推荐关键词相似度较高的医生<sup>[12]</sup>。Huang 等提出了一种基于协作的医药知识推荐方法,根据医生在历史就诊文本中的医药评分来生成信任因子,并引入协同过滤的推荐算法进行医生推荐<sup>[13]</sup>。Zhang 等基于在线评论,利用矩阵分解技术,预测医生排名得分为用户提供个性化的医生推荐<sup>[14]</sup>。杨晓夫和秦函书则根据医生所在科室进行精准推荐<sup>[15]</sup>。

通过对文献的分析和整理,个性化推荐方法能够有效地解决医患信息过载问题。然而,现有研究方法主要从单一数据的视角对用户进行医生推荐,忽略了医患信息的多源性和多粒度性。事实上,用户在实际的决策过程中往往会结合不同数据,从多粒度、多维度的视角来筛选适合的医生。由于在线问诊平台中医生信息来源众多(包括用户反馈、医生简介、医院信息等),这些数据从不同视角、不同维度、不同粒层描述医生特征。医生特征的多粒度划分可以有效帮助在线问诊平台构建医生和用户之间的分级匹配,是平台实现分级诊断的前提。其中,用户咨询文本从较细的粒度描述了医生擅长治疗的病患特征,用户评价从较粗的粒度描述了医生擅长的疾病领域特征,而医生简介从更粗的粒度描述了医生所在科室等方面特征。因此,平台可以通过筛选相似疾病特征的咨询文本,为用户推荐合适的医生;还可以通过比较用户对医生问诊服务的评价和主页简介中医生的专业领域,为用户推荐其他合适的医生。

基于上述分析,本文从医患数据的多源性和多粒度两个视角出发,着重考虑推荐过程的不确定性,提出一种面向在线问诊平台的多步骤三支推荐方法。首先,针对现有研究中医生推荐数据单一性问题,本文考虑医患信息的多源性,通过 Word2Vec 和 LDA 等技术对多源数据进行学习,挖掘医生特征,构建多粒度的医患推荐信息。其次,考虑到推荐过程中的不确定性问题,引入三支决策思想,构建三支医生推荐方法。最后,将多粒度的医患推荐信息和三支决策思想融合,实现面向多源数据的多步骤三支医生推荐。

## 1 相关工作

### 1.1 在线医生推荐

已有关于在线医生推荐的研究主要关注两个方面:医患特征挖掘以及医生推荐。在医患特征挖掘上,现有研究主要通过文本挖掘技术探究医生用户的潜在特征信息。Bekhui 等利用 NLP 相关技术,从在线医疗平台论坛中提取医生领域的关键词<sup>[16]</sup>。Chen 通过对医患双方进行聚类来获取医生和用户特征<sup>[17]</sup>。Liu 等通过注意力机制模型,构建一种新的深度网络,用以学习中文医疗平台中咨询文本的语义特征<sup>[18]</sup>。在医生推荐研究上,Abacha 和 Zweigenbaum 基于医疗命名体识别方法,对医学问题和文档进行深入分析,给出了一种基于相似用户特征的医生推荐方法<sup>[19]</sup>。Naderi 等通过使用统一的医学语言系统和自然语言处理技术,提出一种基于医生领域匹配的医生推荐方法<sup>[20]</sup>。Yang 等探讨了一个基于用户医疗信息需求、问答记录内容和医生背景信息的深度学习问答推荐框架,来提高医生推荐的精确度<sup>[21]</sup>。

基于上述分析,当前研究存在一定的局限性。一方面,现有文献大多通过单一的数据源进行特征提取,忽略了医患数据的多源性和多粒度性;另一方面,现有方法主要关注推荐结果的精度和质量,忽略了推荐的多样性和覆盖率。从用户的角度来看,推荐系统应推荐用户需要的医生;从平台的角度来看,推荐系统也需要实现平台资源的合理分配。因此,本文将基于医患数据的多源性和多粒度性,以提高推荐质量和推荐覆盖率为研究目标,构建在线医生推荐模型。

### 1.2 文本特征挖掘

文本特征挖掘是从自然语言提取特征的过

程,是各种语言处理任务的基础。根据不同的研究方法,可以分为传统的词袋模型、主题模型和深度学习模型 3 类。相对于传统的词袋模型,基于主题模型更关注语言处理过程,它通过从文档中发现的潜在主题并为其提供低维的特征表示,在文本特征提取中取得成功的应用。LDA 是一种经典的层次概率主题模型<sup>[22]</sup>,其核心思想是把主题中的词作为文本的部分特征,每个主题都由这些词组成。由于 LDA 模型能够有效提高文本语义分析的准确性,因而被广泛运用在提取医疗平台中文本特征的工程中<sup>[23-24]</sup>。基于深度学习的文本表示模型主要聚焦于人工神经网络的文本表示,其代表性方法为 Word2Vec 技术。Word2Vec 生成词向量的基本思想来自 NNLM (神经网络语言模型, neural network language model),它通过使用稠密、低维的实数向量来表示 1 个单词,每个维度代表了词语所包含的某个方面的潜在特征,能够表征有价值的句法和语义特征。此外,Word2Vec<sup>[25]</sup>使用具有单个隐藏层的神经网络来训练加权参数,并将采样窗口中单词的贡献率设置为目标函数。由于在线问诊平台中,医患数据都以文本数据为主,因此,Word2Vec 被广泛地应用于医患特征的挖掘中<sup>[10-11]</sup>。

### 1.3 三支决策

三支决策<sup>[26]</sup>理论是在决策粗糙集理论的基础上发展而来的,主要被用于处理延迟决策的不确定性问题。由于传统的二支决策可能会导致较高的误分类成本。因此,三支决策在此基础上提出了延迟决策策略<sup>[27]</sup>,以规避在信息不足的情形下做出错误决策带来的严重后果。在现实的决策过程中,信息的多粒度性会增加决策处理过程。基于此,本文引入序贯三支决策方法,从粒计算的视角处理和实现多粒度信息下的多步骤决策过程。具体而言,在每一步骤中,当现有信息充分时,可以直接做出接受或拒绝的决策;而当前信息无法支持其做出决策时,可以将对象划分到边界域中,在下一个步骤中获得其他信息后再对其进行决策。

一般而言,在传统推荐问题上,推荐模型都是基于二支决策模型,即推荐和不推荐。为了处理推荐中的不确定性问题,三支决策的思想被引入推荐系统中,以解决传统推荐过程“非好即坏”的不足。Zhang 等将三支决策引入到协同过滤的推荐系统中,提出三支推荐规则制定方法,并将其应

用在智慧城市的推荐当中<sup>[28]</sup>。进一步地,考虑到多粒度信息下的多步骤推荐,Ye和Liu将序贯三支决策引入到推荐系统中,构建了一种新的动态三支推荐系统,以解决静态的二支推荐的局限性<sup>[29]</sup>。可以看到,针对在线问诊平台中的医生推荐,多步骤的三支决策推荐能够解决多粒度信息下的推荐过程,并有效解决医生推荐过程中面临的实际问题。

## 2 面向在线问诊平台的三支推荐方法

下面,开始着手对多粒度的医患推荐数据的

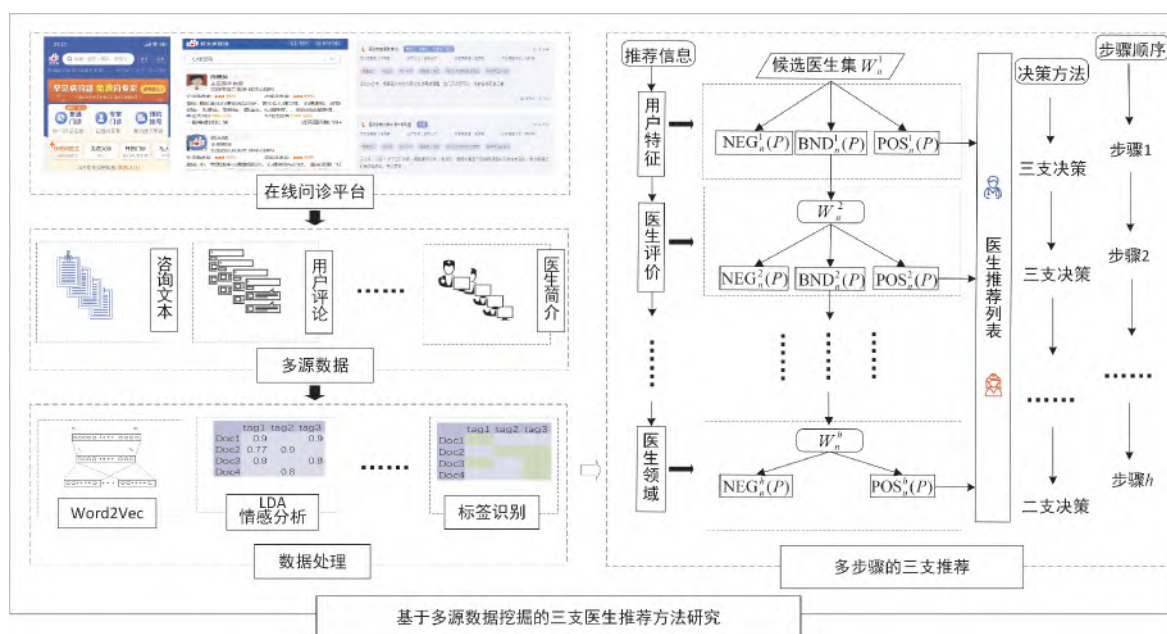


图1 面向在线问诊平台的三支推荐方法的构架图

Fig. 1 The framework of three-way recommendation methods for online consultation platform

医生进行咨询。给定一个推荐系统中,令  $U = \{u_1, u_2, \dots, u_n, \dots, u_N\}$  ( $1 \leq n \leq N$ ) 为用户集;  $V = \{v_1, v_2, \dots, v_m, \dots, v_M\}$  ( $1 \leq m \leq M$ ) 为医生集。对于用户  $u_n$ ,其咨询文本分为两个部分,第一部分是用户对自己疾病的大致归类,第二部分是对自身病情的描述。为了获取用户需求,通过 Word2Vec 提取用户咨询文本中的信息。令  $D_n^j$  表示用户病情描述的文本  $D_n^j = \{d_n^1, d_n^2, d_n^3, \dots, d_n^j\}$ 。其中  $j$  表示文本中词的个数。通过 Word2Vec 训练得到每个词的词向量  $b_j$ ,可以计算句子中的词向量并取平均值表示句向量  $d_n$ 。

$$d_n = \frac{1}{j} \sum_{j=1}^j b_j \quad (1)$$

2.1.2 基于用户评论的医生特征挖掘 由于在线问诊平台允许用户对已完成的医疗服务进行评

清洗、整理和分析。考虑到在线问诊平台的数据整合过程需要爬取用户咨询文本和用户评论、挖掘用户需求、搜集医生简介信息,以及获取医生多粒度特征。因而,本小节从医患数据的多源性和多粒度性出发,提出一种面向在线问诊平台的多步骤三支医生推荐方法,其总体研究框架如图1所示。

### 2.1 基于多源数据的在线医患特征挖掘

2.1.1 基于咨询文本的用户需求挖掘 用户咨询文本是用户在进行咨询服务时对自身病情的描述。用户通常会根据咨询文本,选择与其具有相似需求且得到良好服务的某一治愈患者所选择的

价并形成历史记录,用户通常会在众多历史评论中寻找自己关心的话题,进而选择好评率高、专业能力强、服务态度好的医生。因此,首先需要对用户评论进行语言处理和主题提取。本文主要采用 LDA 方法来分析用户评论的主题数,并挖掘评论中的主题信息。为了更详细地说明 LDA,图2给出了 LDA 的图模型。假设有  $K$  个主题  $\zeta = \zeta_{1:k}$ ,每个主题都满足 Dirichlet 分布,则每个用户评论的文档  $w_{n*}$  的生成过程可表示为:

- 1) 每个主题的分布  $\theta_n$  满足  $\theta_n \sim \text{Dirichlet}(\epsilon)$ 。
- 2) 对于文档  $w_{n*}$  中的每个单词  $w_{nv}$ ,
  - a) 主题分配  $z_{jv}$  满足  $z_{jv} \sim \text{Multinomial}(\theta_n)$ ;
  - b) 词满足  $w_{nv} \sim \text{Multinomial}(\zeta_{z_{jv}})$ 。

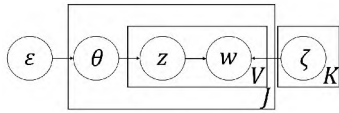


图 2 LDA 的图模型

Fig. 2 The graph model of LDA

可以看到, LDA 模型能够获取每个用户评论不同主题的概率分布  $P_n^K = \{p_n^1, p_n^2, p_n^3, \dots, p_n^K\}$ 。其次, 运用情感分析技术获得用户评论的情感极性。用户  $u_n$  对医生  $v_m$  的评论可以通过基于词典的情感分析获得情感得分  $r_{nm}$  ( $r_{nm} \in [0, 1]$ )。为了方便起见, 假设  $r_{nm} > 0.5$  表示用户对医生的评价为好评; 反之  $r_{nm} \leq 0.5$  则为差评。此外, 可以计算每个医生下每条评论的  $k$  个主题情感得分, 记

$$R_{nm}^K = \{r_{nm}^1, r_{nm}^2, \dots, r_{nm}^K\}, \quad (2)$$

其中  $r_{nm}^k = p_n^k \cdot r_{nm}$ 。

除此之外, 问诊平台还会给用户评论贴上就诊的疾病标签  $t$ , 则每条评论的情感得分可以进一步细化为  $R_{nm}^{K,t} = \{r_{nm}^{1,t}, r_{nm}^{2,t}, \dots, r_{nm}^{K,t}\}$ 。其中  $r_{nm}^{K,t}$  代表在疾病标签  $t$  和主题  $k$  上用户对医生的情感得分。因此, 医生  $v_m$  在该疾病标签下的用户总体满意度可以表示为  $S_m^t$ , 其计算公式为

$$S_m^t = \frac{1}{n^*} \sum_n \sum_k r_{nm}^{k,t}. \quad (3)$$

**2.1.3 基于医生简介的医生特征提取** 除了用户评论, 用户还能够通过浏览医生主页, 查看医生简介信息。医生简介信息中包含了医生对自己专业领域的介绍, 以及对工作经历、教育经历、获得荣誉、科研成果等信息的描述。推荐系统可以根据医生简介中医生擅长的疾病与用户实际确诊疾病进行匹配并推荐。假设总共有  $H$  种疾病标签 ( $T = \{t_1, t_2, \dots, t_H\}$ )。通过标签识别的方法将医生简介中存在擅长的疾病标签标为 1, 即  $Q_m^T = \{q_m^{t_1}, q_m^{t_2}, \dots, q_m^{t_H}\}$ , 其中  $q_m^{t_H} = \{0, 1\}$ 。若  $q_m^{t_H} = 1$ , 则表示医生  $v_m$  擅长治疗疾病  $t_H$ , 否则为不擅长。

## 2.2 基于三支决策的多步骤推荐方法

传统的医生推荐是二支决策模型, 采用推荐和不推荐两种策略。然而, 在二支推荐过程中, 错误的推荐可能会产生高昂的决策成本。基于此, 本文引入三支决策的思想, 对不确定的事件采取延迟推荐策略。给定两个阈值  $\alpha$  和  $\beta$ , 三支推荐系统基本结构如表 1 所示。

相比于二支推荐模型, 三支推荐模型对不确定的事件采用延迟推荐手段。具体而言, 当已有

推荐信息不充分, 难以做出接受或拒绝判断时, 采用延迟推荐策略, 等待收集更多有用的数据信息再进行推荐。这种通过加入更多的推荐信息进行延迟推荐的方法, 使边界域中的医生逐渐被进一步划分和推荐, 进而实现了多步骤推荐。

表 1 三支推荐系统

Tab. 1 Three-way recommendation systems

数学判定条件	决策区域	决策规则
$\text{Score}(P   (u_n, v_m)) \geq \alpha$	正域	推荐
$\beta < \text{Score}(P   (u_n, v_m)) < \alpha$	边界域	延迟推荐
$\text{Score}(P   (u_n, v_m)) \leq \beta$	负域	不推荐

基于上述分析, 需要从在线问诊平台中获取多源数据, 进而通过数据清洗和处理后形成有用的推荐信息。假设  $G_1, G_2, \dots, G_l, \dots, G_h$  是一组多源医患推荐信息 ( $1 \leq l \leq h$ )。针对目标用户  $u_n$ , 令  $W_n$  代表可以推荐给该用户的医生集合,  $W_n^l$  表示在推荐信息  $G_l$  上的候选医生集合。对于医生  $v_m \in W_n^l$ ,  $\text{Score}(p^l | (u_n, v_m))$  表示将医生  $v_m$  推荐给用户  $u_n$  的推荐得分。在每一步骤中, 根据医生推荐得分  $\text{Score}(P^l | (u_n, v_m))$  与阈值 ( $\alpha^l, \beta^l$ ) 的大小关系可以将待推荐的医生划分为正域、负域和边界域, 其数学表达式为

$$\begin{aligned} \text{POS}_n^l(P) &= \{v_m \in W_n^l | \text{Score}(p^l | (u_n, v_m)) \geq \alpha^l\}, \\ \text{BND}_n^l(P) &= \{v_m \in W_n^l | \beta^l < \text{Score}(P^l | (u_n, v_m)) < \alpha^l\}, \\ \text{NEG}_n^l(P) &= \{v_m \in W_n^l | \text{Score}(P^l | (u_n, v_m)) \leq \beta^l\}. \end{aligned} \quad (4)$$

在每个步骤中, 对于任何一个医生  $v_m \in W_n^l$ , 如果  $v_m \in \text{POS}_n^l(P)$ , 则将其推荐给该用户; 如果  $v_m \in \text{NEG}_n^l(P)$ , 则不推荐给用户; 而对  $v_m \in \text{BND}_n^l(P)$  的医生, 则放入下一个步骤进行推荐。随着更多数据和更多推荐信息被获取, 边界域  $\text{BND}_n^l(P)$  会逐渐减少, 边界域中的医生也会逐渐被划分到正域和负域。特别地, 如果当推荐过程到达最后一个步骤  $h$  时, 推荐的医生未达到所需要的推荐数  $\mathcal{N}$ , 则根据医生推荐得分, 从高到低选择得分较高的医生, 直到满足需求为止。总而言之, 多步骤三支推荐过程直到推荐列表中的医生达到所需数量  $\mathcal{N}$  时终止。换言之, 当正域累计的数量

达到或超过  $\mathcal{N}$  时, 推荐终止。令  $G_{\text{term}_n}$  表示为用户  $u_n$  终止推荐医生所在的推荐信息阶段, 其推荐遵循以下规则:

$$\text{term}_n = \begin{cases} 1, & |\cup_{l=1}^1 \text{POS}_n^l(P)| \geq \mathcal{N}; \\ (1, h], & |\cup_{l=1}^{\text{term}_n-1} \text{POS}_n^l(P)| < \mathcal{N} \text{ 且} \\ & |\cup_{l=1}^{\text{term}_n} \text{POS}_n^l(P)| \geq \mathcal{N} \end{cases} \quad (5)$$

通过上述分析可以看到, 面向多源数据的多步骤三支推荐方法适用于医患信息来源众多的在线问诊平台, 从多粒度视角来看, 随着推荐信息的增加, 医生会被进一步划分。推荐系统从多源的

数据信息为用户进行医生推荐, 随着推荐信息的逐步加入, 能够推荐给用户的医生数量也会变得更多。

### 2.3 面向在线问诊平台的三支推荐方法

针对在线问诊平台中医患信息的多源性和多粒度性及推荐过程中的不确定性, 本文拟融合用户咨询文本、用户评论和医生简介, 挖掘医生特征等数据, 构建一种基于三支决策思想的多步骤医生推荐方法(3WD-msDRec), 其核心技术路线如图3所示。

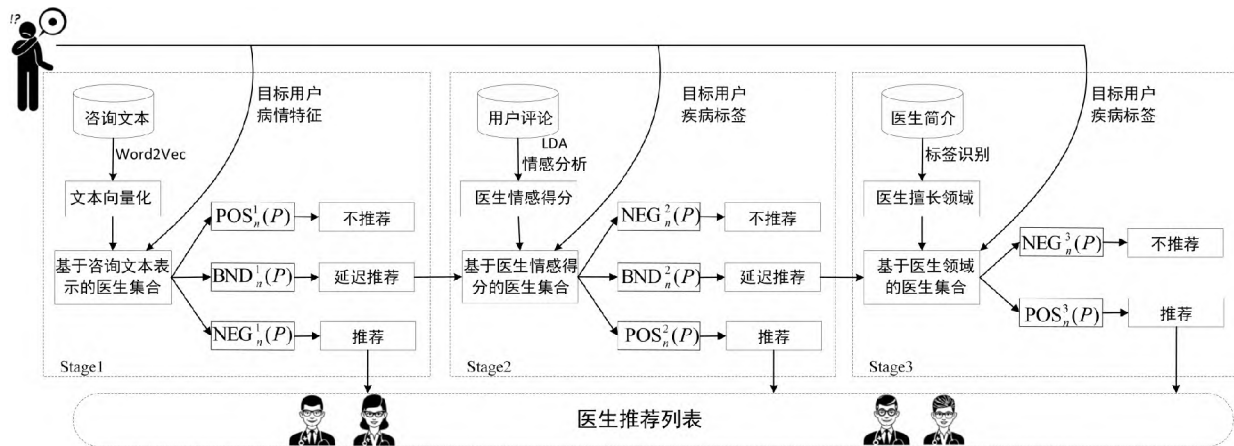


图3 基于三支决策思想的多步骤医生推荐方法(3WD-msDRec)

Fig. 3 Multi-stage doctor recommendation based on three-way decision(3WD-msDRec)

图3考虑了用户咨询文本、用户评论和医生简介3种不同信息源。由于每个阶段考虑的信息源不同, 随着信息源的不断增加, 为用户推荐的医生也更为丰富。基于此, 多步骤的三支医生推荐可分为3个阶段, 其对应的推荐信息集合分别记为  $G_1$ ,  $G_2$  和  $G_3$ 。

在第1阶段中, 通过对咨询文本的语义挖掘获取医生特征记为  $G_1$ 。目标用户  $u_n$  的咨询文本  $d_n$  和以往就诊医生  $v_m$  的用户文本  $d_m$  通过 Word2Vec 向量化表示。这里, 通过余弦相似度计算目标用户和咨询文本之间的语义相似度, 将相似的以往用户推荐给目标用户, 即

$$\text{sim}(d_n, d_m) = \frac{d_n \cdot d_m}{\|d_n\| \times \|d_m\|} \quad (6)$$

若两个文本之间相似度较高, 则可以为目标用户推荐其就诊过的医生。进一步地, 定义第1阶段的医生推荐得分为

$$\text{Score}(P^1 | (u_n, v_m)) = \text{sim}(d_n, d_m) \quad (7)$$

在第2阶段中, 通过对用户评论挖掘获取基

于医生情感得分的医生特征记为  $G_2$ , 并由式(3)计算医生在疾病  $t$  上的用户满意度  $S_m^t$ 。由于用户患有的疾病可能有多种, 因此, 需要对疾病标签进行加权平均, 获得医生在多个领域上的平均用户满意度, 进而计算第2阶段医生的推荐得分为

$$S_m^{T_n} = \frac{1}{|T_n|} \sum_t S_m^t = \frac{1}{n^* k^* |T_n|} \sum_t \sum_n \sum_k r_{nm}^{k,t}, \quad (8)$$

$$\text{Score}(P^2 | (u_n, v_m)) = S_m^{T_n}. \quad (9)$$

在第3阶段中, 通过医生简介挖掘获取基于医生领域的医生特征记为  $G_3$ , 这里需要考虑  $G_3$  与用户疾病标签  $T_n$  的关系, 并据此得到第3阶段医生推荐得分  $\text{Score}(P^3 | (u_n, v_m)) = q_m^{T_n}$ 。显然地, 推荐系统最终会给目标用户推荐  $\text{Score}(P^3 | (u_n, v_m))$  排序较高的医生。

基于上述分析, 面向在线问诊平台的多阶段三支推荐方法可以根据每个步骤中医生推荐得分, 并在给定的阈值 ( $\alpha^l, \beta^l$ ) 下对候选医生集合进

行如式(5)的推荐规则,最终形成对不同目标用户的医生推荐列表,具体的算法伪代码如算法1所示。第1阶段注重于从相似文本的角度进行考量,寻找与目标用户有相似病征的以往用户,并将他们咨询过的医生推荐给目标用户。当第1阶段推荐的医生数量小于 $\mathcal{N}$ 时,将继续通过第2阶段为用户推荐医生。而第2阶段侧重于从用户评价、用户满意度的角度进行考量,向目标用户推荐满意度较高的医生。第3阶段则是在前面两个阶段推荐的医生数不足的情况下,根据目标用户的疾病标签从医生简介中进行推荐。

**算法 1** 多步骤的三支医生推荐方法 (3WD-msDRec)

**Input:** 用户集  $U = \{u_1, u_2, \dots, u_n, \dots, u_N\}$ , 医生集  $V = \{v_1, v_2, \dots, v_m, \dots, v_M\}$ , 推荐信息  $G_1, G_2, \dots, G_l, \dots, G_h$ , 医生推荐得分  $\text{Score}(P^l | (u_n, v_m))$ , 所需医生推荐数量  $\mathcal{N}$ , 各阶段阈值  $(\alpha^l, \beta^l)$  且  $1 \leq l \leq h-1$ ;

**Output:** 医生推荐列表  $\text{Rec}_{\text{list}}^n$ 。

1: // 步骤 1. 初始化参数

2:  $\text{Rec}_{\text{list}}^n = \emptyset$

3:  $W_n^l = W_n$

4:  $l = 1$

5: // 步骤 2. 多步骤三支推荐过程

6: **While**  $l < h$  and  $|\text{Rec}_{\text{list}}^n| < \mathcal{N}$  **do**

7:  $\text{POS}_n^l(P) = \{v_m \in W_n^l \mid \text{Score}(P^l | (u_n, v_m)) \geq \alpha^l\}$

8:  $\text{BND}_n^l(P) = \{v_m \in W_n^l \mid \beta^l < \text{Score}(P^l | (u_n, v_m)) < \alpha^l\}$

9:  $\text{NEG}_n^l(P) = \{v_m \in W_n^l \mid \text{Score}(P^l | (u_n, v_m)) \leq \beta^l\}$

10:  $y = \mathcal{N} - |\text{Rec}_{\text{list}}^n|$

11: **if**  $|\text{POS}_n^l(P)| \leq y$  **then**

12:  $\text{Rec}_{\text{list}}^n = \text{Rec}_{\text{list}}^n \cup \text{POS}_n^l(P)$

13: **else**

14:  $\text{Rec}_{\text{list}}^n = \text{Rec}_{\text{list}}^n \cup \text{POS}_n^l(P_y)$

15: **end if**

16:  $W_n^{l+1} = \text{BND}_n^l(P)$

17:  $l = l + 1$

18: **end while**

19: // 步骤 3. 二支推荐过程

20: **if**  $|\text{Rec}_{\text{list}}^n| < \mathcal{N}$  **then**

21:  $y = \mathcal{N} - |\text{Rec}_{\text{list}}^n|$

22:  $\text{Rec}_{\text{list}}^n = \text{Rec}_{\text{list}}^n \cup W_n^h(P_y)$

23: **end if**

24: **Return**  $\text{Rec}_{\text{list}}^n$

### 3 实验结果及分析

#### 3.1 数据收集

为了验证本文所提推荐方法的有效性,本文从好大夫在线网站(haodf.com)上爬取了从2016年到2020年心理咨询科和内科2个科室的用户咨询文本、用户评论、医生简介等原始数据,其具体信息如下:

1) 心理咨询科室数据集: 包括咨询文本共11万条,320位医生及其主页简介,医生评论2.1万条。经过数据整理和清洗,删除评论不足10条的医生,且文本单词数少于10个字的评论、简介和咨询文本,保留了89000条咨询文本,250位医生以及10699条评论。此外,通过统计分析得到心理咨询的疾病标签46个。

2) 内科数据集: 与心理咨询科室不同,内科科室之中门类繁多,它由消化内科、心血管科、内分泌科、变态反应科等子科室组成。因此,在爬取内科科室数据时重点爬取了上述4个子科室中的咨询文本、用户评论、医生简介等数据。经过数据清洗后,保留了113255条咨询文本,499位医生以及19888条评论。此外,通过统计分析得到内科的疾病标签88个。

接下来,在实验设计中将80%的咨询文本作为训练集,用于训练Word2Vec模型和用于匹配文本的相似度,剩下20%的样本作为测试集进行实验分析。

#### 3.2 实验评价指标

这里,采用推荐精度、推荐平均精度、归一化折损累计增益和推荐覆盖率4个评价指标来验证本文所提出推荐方法的有效性。

##### 1) 推荐精度

医生推荐的推荐精度主要反映了推荐系统是否推荐了用户实际就诊的医生。推荐精度由式(10)计算得到。其中,  $|U^{\text{test}}|$  表示测试集中用户个数,  $\text{real}^i$  表示测试集用户  $i$  实际就诊的医生,推荐系统为其推荐的医生列表为  $\text{Rec}_{\text{list}}^i$ 。

$$\text{Accuracy} = \frac{\sum_{i=1}^{|U^{\text{test}}|} |\text{real}^i \cap \text{Rec}_{\text{list}}^i|}{|U^{\text{test}}|} \quad (10)$$

##### 2) 推荐平均精度



平均精度反映了推荐系统中给用户推荐相关项目排名的能力,平均精度越高,最终推荐的列表中医生的排名也就越靠前。平均精度计算公式如式(11)~式(12)所示。

$$AP(i) = \frac{1}{X} \sum_{x=1}^X \frac{x}{\text{position}(x)} \quad (11)$$

$$MAP = \frac{\sum_{i=1}^{|U^{\text{test}}|} AP(i)}{|U^{\text{test}}|} \quad (12)$$

考虑到每个用户最终只与一个医生成功匹配,因此  $x = X = 1$ 。 $\text{position}(x)$  表示推荐成功的医生在推荐列表中的位置,而没有推荐成功的用户  $AP(i) = 0$ 。

### 3) 归一化折损累计增益

折损累计增益(DCG)主要考虑了推荐系统中排序的结果,排名靠前的项目增益更高,对排名靠后的项目进行折损。归一化折损累计增益(nDCG)则是对不同用户的指标进行归一化,最后,对每个用户取平均得到的最终分值,其计算公式如式(13)所示。

$$\text{nDCG}@p = \frac{1}{\text{IDCG}} \sum_{i=1}^p \frac{2^{\text{rel}_i} - 1}{\log_2(i+1)} \quad (13)$$

其中: $p$  为推荐列表中医生的排序位置; $\text{rel}_i$  表示第  $i$  个医生为某用户的问诊医生。在医生推荐的情境下,1 个医生就诊 1 个用户,即  $\text{rel}_i = \{0, 1\}$ 。因此,理想的排序结果应为就诊医生排在推荐列表的第 1 位,即  $\text{IDCG} = 1$ 。

### 4) 推荐覆盖率

覆盖率反映了推荐系统覆盖项目的能力,覆盖率越高,表明最终推荐的列表中能够覆盖的医生越多,它由式(14)计算,其中  $V$  为医生集合。

$$\text{Coverge} = \frac{|U_{i \in |U^{\text{test}}|} \text{Rec}_{\text{list}}^i|}{|V|} \quad (14)$$

## 3.3 基准方法

进一步地,为了验证本文提出的医生推荐系统(3WD-msDRec)的有效性,本文以下面 3 种方法作为基准算法。

1) DP-bRec (doctor popularity based recommendation): 基于医生热度的推荐方法是现存医疗网站上主要使用的推荐方法,该方法是将热度排序高的医生推荐给用户。

2) W2V-TextRec: 该方法是以用户咨询文本为数据源,经过 Word2Vec 训练,将与以往用户文本咨询过文本相似度最高的医生推荐给目标用户的推荐方法<sup>[11]</sup>。

3) Labeled-LDADRec: 该方法是以用户咨询文本为数据源,通过 LDA 主题模型把训练得出的关键词作为医生专业领域的标签,并将标签相似度较高的医生推荐给目标用户的推荐方法<sup>[30]</sup>。

## 3.4 实验结果分析

3.4.1 算法比较分析 首先,为了验证本文所提算法在 4 个评价指标的有效性,利用 Word2Vec 和 LDA 模型对在线问诊平台中的数据进行处理。一方面,在训练 Word2Vec 的过程中,将词向量的维度设置为 250,迭代次数设置为 10 次。另一方面,借鉴文献[31]对医生评论进行主题提取的结果,对用户评论进行 LDA 主题模型训练,设定  $K = 4$ 。进一步地,假设所需医生推荐数  $\mathcal{N} = 10$ ,表 2 和表 3 展示了阈值  $\alpha$  和  $\beta$  在不同组合变化下,推荐精度、覆盖率、MAP 和 nDCG 的实验结果。

通过表 2 和表 3 可以看出,推荐精度、MAP 和 nDCG 随着  $\alpha$  和  $\beta$  的取值变小而表现更为有益。对于覆盖度来说, $\alpha$  的取值区间位于 0.65 ~ 0.85 表现较好。为了兼顾 4 个指标的表现,选取  $(\alpha^1, \beta^1) = (0.65, 0.3)$ 。在第 1 阶段推荐中,当文本的相似度  $\geq 0.65$  将其归到正域,并且向用户推荐该文本咨询的医生;当文本相似度介于 0.3 ~ 0.65 时,将该医生归到边界域,即延迟推荐;剩下的医生便进入负域,不向用户推荐。类似地,第 2 阶段选取  $(\alpha^2, \beta^2) = (0.8, 0.4)$ 。

进一步地,在确定好阈值后,将本文提出的 3WD-msDRec 方法与 3 种基准方法进行实验分析,相关比较结果如图 4 和图 5 所示。

一方面,由图 4A 可知,在心理咨询数据集上,3WD-msDRec 的推荐精度都优于其他基准方法。由图 4B 和图 4C 可以看出,3WD-msDRec 方法在 MAP 和 nDCG 两个指标上的表现优于其他基准方法。通过图 4D 可以得到,当  $\mathcal{N} \leq 25$  时,DP-bRec 方法的覆盖率最低,3WD-msDRec 方法的覆盖率最优。当  $\mathcal{N} \geq 30$  时,Labeled-LDADRec 方法的推荐覆盖率比 3WD-msDRec 方法表现更好,不过 3WD-msDRec 方法的覆盖率也达到了 90% 以上,与最优方法表现相差甚微。

另一方面,通过图 5 可以发现,3WD-msDRec 方法在两个数据集上的表现是一致的。此外,综合图 4 和图 5 发现,3WD-msDRec 方法在内科数据集的推荐精度、MAP 和 nDCG 的结果表现比心理咨询数据集的表现要高出 10% 左右。总而言



之,本文所提方法在推荐精度和覆盖度上总体都 个指标来看,3WD-msDRec 方法具有良好的推荐  
优于其他基准方法,而且从平均精度和 nDCG 两 质量。

表 2  $\alpha$  和  $\beta$  组合变化下推荐精度和覆盖度的实验结果( $N=10$ )

Tab. 2 The experiment results on accuracy and coverage with different values of  $\alpha$  and  $\beta$  ( $N=10$ )

$\alpha$	$\beta$													
	0.60		0.55		0.50		0.45		0.40		0.35		0.30	
	Acc.	Cov.	Acc.	Cov.	Acc.	Cov.	Acc.	Cov.	Acc.	Cov.	Acc.	Cov.	Acc.	Cov.
0.95	0.070	0.750	0.076	0.723	0.074	0.711	0.074	0.699	0.074	0.699	0.074	0.703	0.074	0.703
0.90	0.076	0.789	0.080	0.781	0.080	0.781	0.078	0.774	0.078	0.774	0.078	0.774	0.078	0.774
0.85	0.124	0.887	0.130	0.887	0.130	0.895	0.128	0.891	0.128	0.891	0.128	0.891	0.128	0.891
0.80	0.152	0.902	0.156	0.898	0.156	0.910	0.156	0.910	0.156	0.910	0.156	0.910	0.156	0.910
0.75	0.170	0.891	0.174	0.891	0.174	0.891	0.174	0.891	0.174	0.895	0.174	0.895	0.174	0.895
0.70	0.176	0.867	0.176	0.880	0.176	0.880	0.174	0.883	0.174	0.887	0.174	0.887	0.174	0.887
0.65	0.186	0.867	0.188	0.867	0.186	0.871	0.186	0.880	0.186	0.880	0.186	0.883	0.186	0.883
0.60	0.188	0.867	0.188	0.867	0.188	0.867	0.188	0.867	0.188	0.867	0.188	0.871	0.188	0.871

表 3  $\alpha$  和  $\beta$  组合变化下 MAP 和 nDCG 的实验结果( $N=10$ )

Tab. 3 The experiment results on MAP and nDCG with different values of  $\alpha$  and  $\beta$  ( $N=10$ )

$\alpha$	$\beta$													
	0.60		0.55		0.50		0.45		0.40		0.35		0.30	
	MAP	nDCG	MAP	nDCG	MAP	nDCG	MAP	nDCG	MAP	nDCG	MAP	nDCG	MAP	nDCG
0.95	0.019	0.026	0.021	0.029	0.020	0.029	0.020	0.028	0.020	0.028	0.020	0.028	0.020	0.028
0.90	0.020	0.028	0.024	0.030	0.025	0.032	0.024	0.031	0.024	0.031	0.024	0.031	0.024	0.031
0.85	0.024	0.031	0.041	0.049	0.042	0.050	0.042	0.050	0.042	0.050	0.042	0.050	0.042	0.050
0.80	0.050	0.060	0.051	0.062	0.051	0.062	0.051	0.062	0.051	0.062	0.051	0.062	0.051	0.062
0.75	0.053	0.066	0.054	0.067	0.054	0.067	0.054	0.067	0.054	0.067	0.054	0.067	0.054	0.067
0.70	0.054	0.069	0.056	0.069	0.056	0.069	0.056	0.068	0.056	0.068	0.056	0.068	0.056	0.068
0.65	0.056	0.068	0.060	0.073	0.060	0.073	0.060	0.073	0.060	0.073	0.060	0.073	0.060	0.073
0.60	0.060	0.074	0.060	0.074	0.060	0.074	0.060	0.074	0.060	0.074	0.060	0.074	0.060	0.074

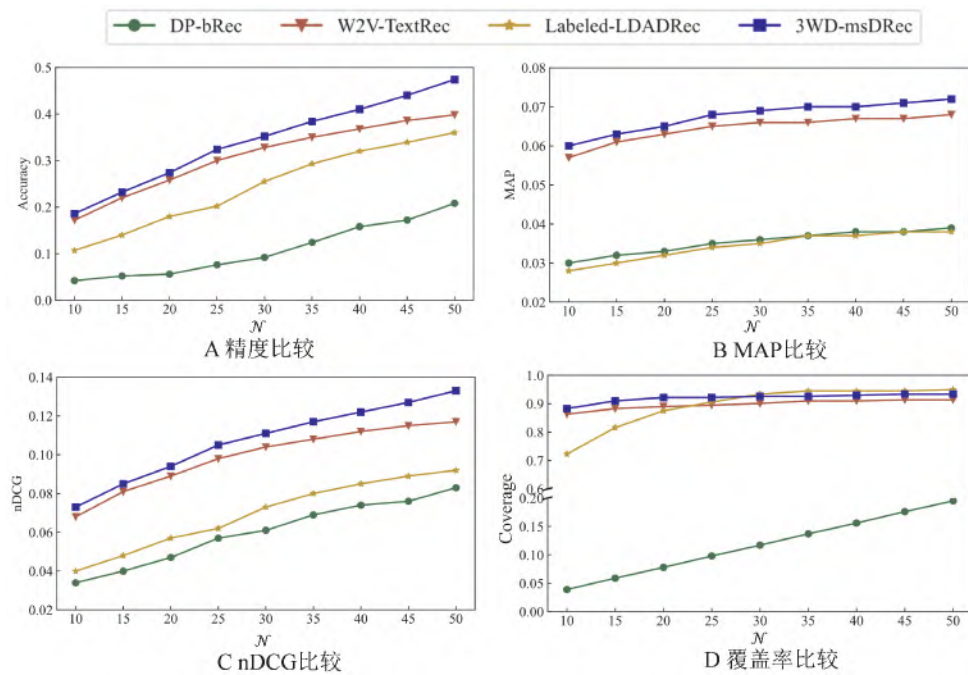


图 4 与基准方法结果对比图(心理咨询科室)

Fig. 4 Comparison of the results of the baseline method( psychological counseling department)

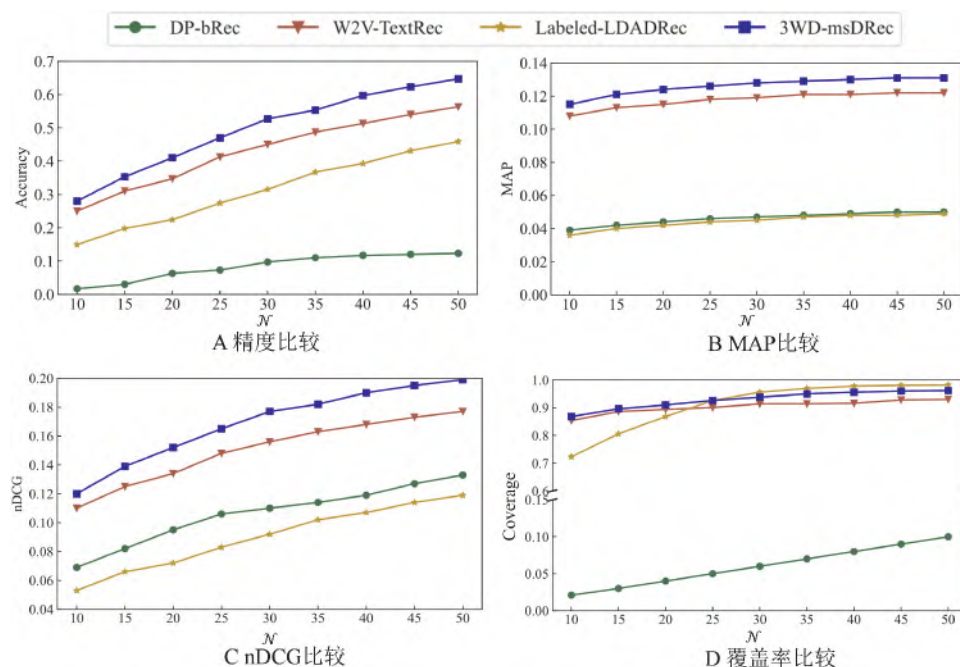


图5 与基准方法结果对比图( 内科科室)

Fig. 5 Comparison of the results of the baseline method( internal medicine department)

3.4.2 不同数据源组合比较分析 其次,为了探索所提方法使用多种数据源进行多粒度医生推荐的优越性,进一步将本文所提方法与不同数据源组合推荐方法做对比分析,使用单一数据源进行

的医生推荐为单粒度医生推荐,使用两个数据源则为两个医生粒度特征的医生推荐,相关实验结果如图6和图7所示。

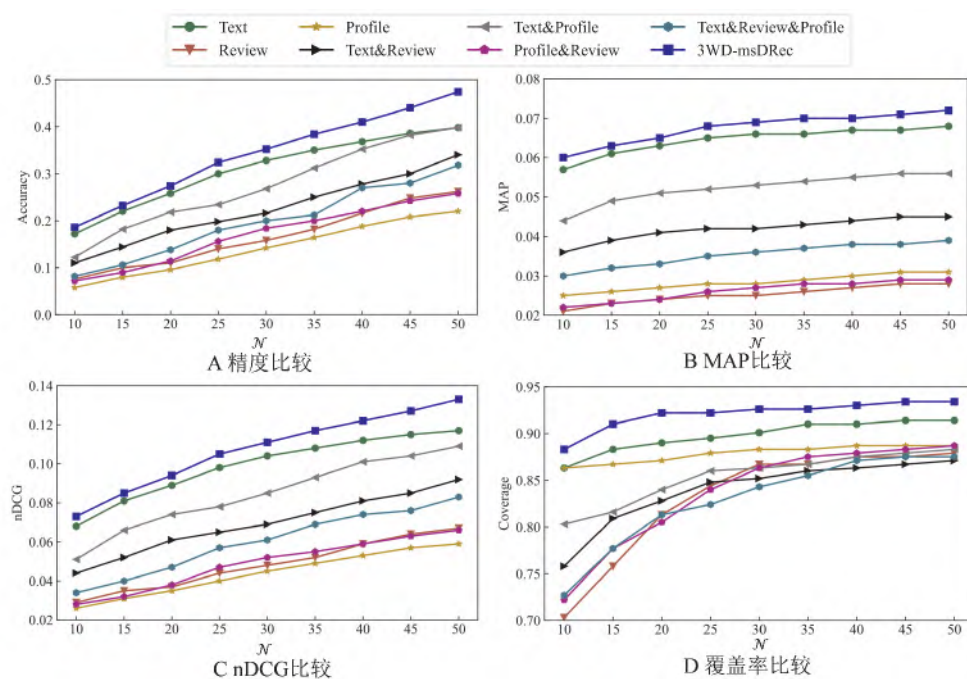


图6 不同数据源结果比较图( 心理咨询科室)

Fig. 6 Comparison of results from different data sources ( psychological counseling department)

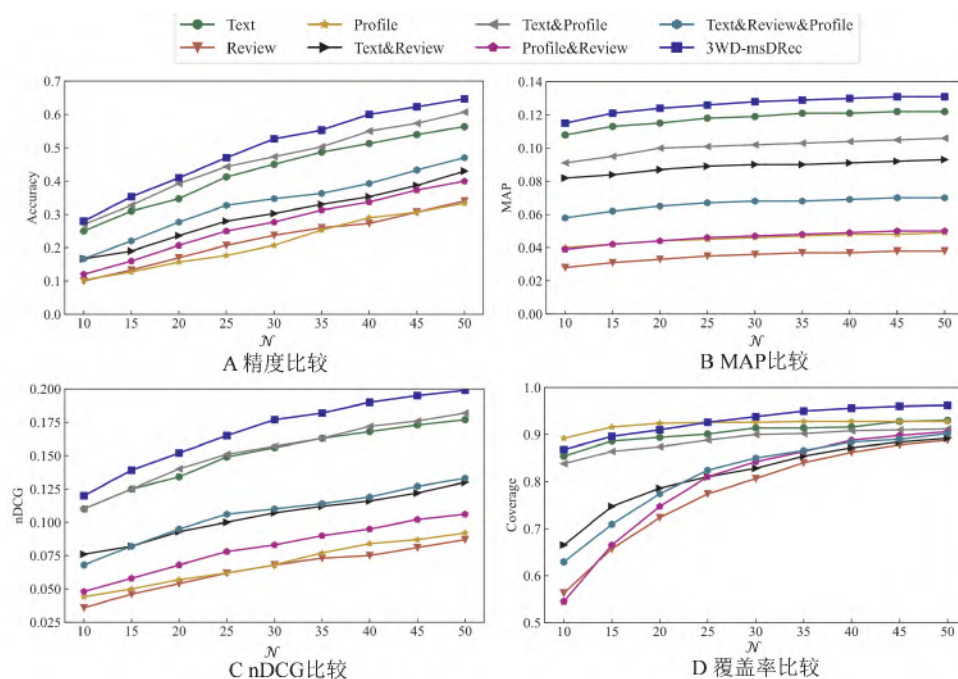


图 7 不同数据源结果比较图(内科室)

Fig. 7 Comparison of results from different data sources (internal medicine department)

通过图 6 和图 7 可以发现 3WD-msDRec 方法在各指标的表现都优于其他数据源组合推荐的方法。相比其他数据源组合推荐进行的不同粒度医生推荐方法,采取了 3 个不同数据源进行多粒度的 3WD-msDRec 方法在推荐精度、MAP 和 nDCG 上优势明显。此外 3WD-msDRec 方法在覆盖率的表现也优于其他方法,这在一定程度上验证了本文所提出方法的有效性。综上所述,通过对推荐精度、MAP、nDCG 和覆盖率 4 个指标的分析,相比其他基准方法及其他数据源组合推荐方法,

本文提出的方法在提高推荐精度的同时,还提高了推荐的覆盖率。

3.4.3 结果可视化分析 最后,为了清晰刻画面向多源数据的多阶段可视化推荐过程,分析采用多源数据进行推荐的合理性和有用性,本文从心理咨询科室数据集中随机选取 10 名用户,为他们推荐得分排名前 50 位的医生 ( $N=50$ )。对于不同用户,为其推荐医生的信息来源是不同的。图 8 统计并展示了 10 位用户在咨询文本、用户评论和医生简介 3 种不同信息源的医生推荐数量。

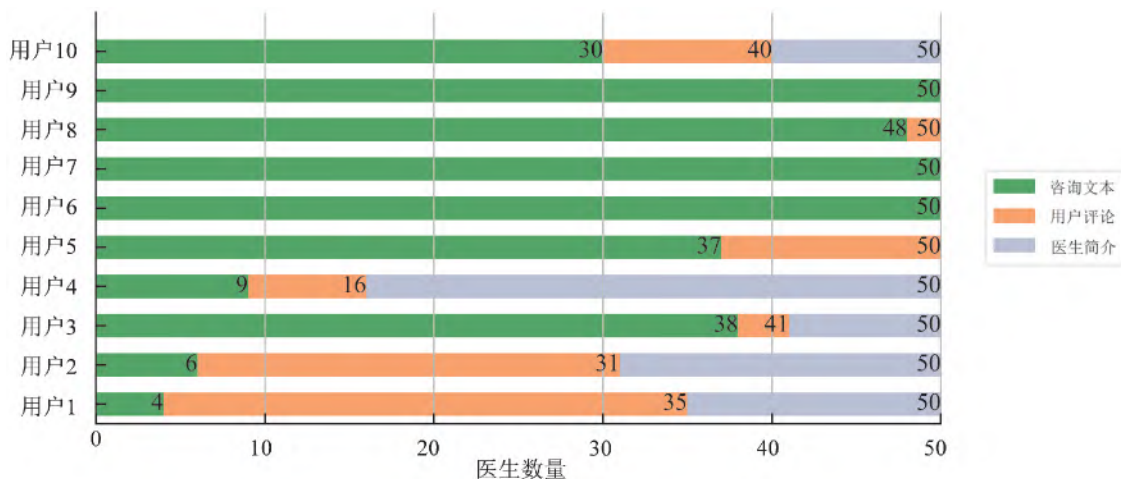


图 8 10 名用户可视化推荐过程

Fig. 8 The visual recommendation process of 10 users

图8中,绿色柱状代表通过咨询文本进行医生推荐的数量,红色柱状代表通过用户评论进行医生推荐的数量,而蓝色柱状代表通过医生简介进行医生推荐的数量。由图8可以发现,用户6、用户7和用户9只需要利用咨询文本来推荐就能满足其需求,即经过第1阶段就终止推荐。用户5和用户8需要咨询文本和用户评论,第2阶段就可以结束推荐。其他用户则同时需要3个信息源的数据,进行3个阶段的推荐。图8的结果证实了本文所提出研究方法的合理性和有用性,同时说明在问诊平台推荐中多信息源的重要性。

## 4 结语

考虑到在线问诊平台中医患信息的多源性和多粒度性,以及推荐过程中的不确定性问题,本文提出了一种面向在线问诊平台的多步骤三支推荐方法,通过挖掘多源数据中的医患特征,利用多粒度的推荐信息,构建了一种多步骤三支医生推荐方法。本文通过爬取好大夫网站上的医患数据对所提方法进行有效性验证。实验结果表明,本文所提方法不仅能够提高医生推荐精度,同时还提高了推荐的覆盖率,具有良好的应用价值。而且,本文方法作为在线问诊平台中医生推荐的方法,它从多个维度、多个角度以及多个粒度为用户进行的推荐,有着良好的语义解释性。此外,本文方法还能在一定程度上缓解医生配置不均衡的问题,具有较好的理论价值和实际意义。在未来研究中,笔者会进一步探索将深度学习的技术引入到在线问诊平台中的相关模型与方法,通过数据处理和分析,将多源医生特征进行融合,挖掘医生的多粒度特征,实现医患的精准匹配。

## 参考文献:

- [1] LI J, ZHANG Y, MA L, et al. The impact of the Internet on health consultation market concentration: An econometric analysis of secondary data [J]. *Journal of Medical Internet Research*, 2016, 18(10): e276.
- [2] YANG H L, ZHANG X F. Investigating the effect of paid and free feedback about physicians' telemedicine services on patients' and physicians' behaviors: Panel data analysis [J]. *Journal of Medical Internet Research*, 2019, 21(3): e12156.
- [3] LU H Y, SHAW B R, GUSTAFSON D H. Online health consultation: Examining uses of an interactive cancer communication tool by low-income women with breast cancer [J]. *International Journal of Medical Informatics*, 2011, 80(7): 518-528.
- [4] RANARD B L, WERNER R M, ANTANAVICIUS T, et al. Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care [J]. *Health Affairs*, 2016, 35(4): 697-705.
- [5] KONG L Q. A study on the AI-based online triage model for hospitals in sustainable smart city [J]. *Future Generation Computer Systems*, 2021, 125: 59-70.
- [6] YANG H L, GUO X T, WU T S, et al. Exploring the effects of patient-generated and system-generated information on patients' online search, evaluation and decision [J]. *Electronic Commerce Research and Applications*, 2015, 14(3): 192-203.
- [7] YANG H L, GUO X T, WU T S. Exploring the influence of the online physician service delivery process on patient satisfaction [J]. *Decision Support Systems*, 2015, 78: 113-121.
- [8] CHEN M, YANG J H, ZHOU J H, et al. 5G-smart diabetes: Toward personalized diabetes diagnosis with health-care big data clouds [J]. *IEEE Communications Magazine*, 2018, 56(4): 16-23.
- [9] 徐守坤, 吴伟伟. 基于语义的医疗资源均衡推荐算法 [J]. *计算机工程*, 2015, 41(9): 74-79.
- [10] 刘通. 基于在线咨询记录的医生自动匹配算法应用研究 [J]. *情报理论与实践*, 2018, 41(6): 143-148, 123.
- [11] LIU T. Research on the application of automatic matching algorithm for doctors based on online consultation records [J]. *Intelligence Theory and Practice*, 2018, 41(6): 143-148, 123.
- [12] 叶佳鑫, 熊回香, 蒋武轩. 一种融合患者咨询文本与决策机理的医生推荐算法 [J]. *数据分析与知识发现*, 2020, 4(S1): 153-164.
- [13] YE J X, XIONG H X, JIANG W X. A doctor recommendation algorithm combining patient consultation text and decision-making mechanism [J]. *Data Analysis and Knowledge Discovery*, 2020, 4(S1): 153-164.
- [14] 潘有能, 倪秀丽. 基于 Labeled-LDA 模型的在线医疗专家推荐研究 [J]. *数据分析与知识发现*, 2020, 4(4): 34-43.
- [15] PAN Y N, NI X L. Research on online medical expert recommendation based on Labeled-LDA model [J]. *Data-*

- ta Analysis and Knowledge Discovery, 2020, 4 (4): 34-43.
- [13] HUANG Z X, LU X D, DUAN H L, et al. Collaboration-based medical knowledge recommendation [J]. Artificial Intelligence in Medicine, 2012, 55(1): 13-24.
- [14] ZHANG Y, CHEN M, HUANG D J, et al. iDoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization [J]. Future Generation Computer Systems, 2017, 66: 30-35.
- [15] 杨晓夫, 秦函书. 基于电子病历利用矩阵乘法构建医生推荐模型 [J]. 计算机与现代化, 2019(6): 81-86, 97.
- YANG X F, QIN H S. A doctor recommendation model is constructed by matrix multiplication based on electronic medical records [J]. Computer and Modernization, 2019 (6): 81-86, 97.
- [16] BEKHUIS T, KREINACKER M, SPALLEK H, et al. Using natural language processing to enable in-depth analysis of clinical messages posted to an Internet mailing list: A feasibility study [J]. Journal of Medical Internet Research, 2011, 13(4): e98.
- [17] CHEN A T. Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups [J]. Patient Education and Counseling, 2012, 87(2): 250-257.
- [18] LIU X, ZHOU Y J, WANG Z R. Deep neural network-based recognition of entities in Chinese online medical inquiry texts [J]. Future Generation Computer Systems, 2021, 114: 581-604.
- [19] BEN ABACHA A, ZWEIGENBAUM P. MEANS: A medical question-answering system combining NLP techniques and semantic Web technologies [J]. Information Processing & Management, 2015, 51(5): 570-594.
- [20] NADERI H, KIANI B, MADANI S, et al. Concept based auto-assignment of healthcare questions to domain experts in online Q&A communities [J]. International Journal of Medical Informatics, 2020, 137: 104108.
- [21] YANG Z, XU W, CHEN R Y. A deep learning-based multi-turn conversation modeling for diagnostic Q&A document recommendation [J]. Information Processing & Management, 2021, 58(3): 102485.
- [22] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. The Journal of Machine Learning Research, 2003, 3: 993-1022.
- [23] ZENG Y Q, LIU X S, WANG Y S, et al. Recommending education materials for diabetic questions using information retrieval approaches [J]. Journal of Medical Internet Research, 2017, 19(10): e342.
- [24] HAO H J, ZHANG K P. The voice of Chinese health consumers: A text mining approach to web-based physician reviews [J]. Journal of Medical Internet Research, 2016, 18(5): e108.
- [25] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in Vector space [EB/OL]. 2013: arXiv: 1301.3781. <http://arxiv.org/abs/1301.3781>.
- [26] YAO Y Y. Three-way decision: An interpretation of rules in rough set theory [M]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, 5589: 642-649.
- [27] YAO Y Y. Three-way granular computing, rough sets, and formal concept analysis [J]. International Journal of Approximate Reasoning, 2020, 116: 106-125.
- [28] ZHANG H R, MIN F, SHI B. Regression-based three-way recommendation [J]. Information Sciences, 2017, 378: 444-461.
- [29] YE X Q, LIU D. An interpretable sequential three-way recommendation based on collaborative topic regression [J]. Expert Systems with Applications, 2021, 168: 114454.
- [30] LIU X, ZHOU Y J, WANG Z R. Preference access of users' cancer risk perception using disease-specific online medical inquiry texts [J]. Information Processing & Management, 2022, 59(1): 102737.
- [31] HU J H, ZHANG X H, YANG Y, et al. New doctors ranking system based on VIKOR method [J]. International Transactions in Operational Research, 2020, 27(2): 1236-1261.

(编辑 张欢)