

## 基于诊疗知识库和电子病历的整合推荐方法研究

胡德华,方浩,周宇葵,罗爱静

中南大学生命科学学院医药信息系,湖南长沙 410083

**【摘要】目的:**探索一种基于临床诊疗知识库和电子病历为患者推荐诊疗相关知识的有效方法。**方法:**引入诊疗知识库的内容过滤结合电子病历建立整合评分矩阵,融合基于电子病历的患者协同过滤(PbCF)和疾病诊断的协同过滤(DbCF)结果,构建整合推荐方法,减少预测结果的误差。以平均绝对百分误差(MAPE)作为评价指标,对整合推荐方法与独立PbCF、DbCF预测结果进行评估比较。**结果:**整合推荐方法的MAPE要小于PbCF和DbCF。局限于所选诊疗知识库的高血压疾病种类覆盖面不及电子病历中高血压患者所涉及的疾病种类,扩大诊疗知识库疾病覆盖面对推荐结果的影响仍需进一步探讨。**结论:**整合推荐方法比独立的PbCF、DbCF效果更好,为患者推荐疾病诊疗知识,具有良好的应用前景。

**【关键词】**临床知识库;电子病历;协同过滤;内容过滤

**【中图分类号】**R319;TP391

**【文献标志码】**A

**【文章编号】**1005-202X(2019)04-0492-05

### A hybrid recommendation method based on medical knowledge base and electronic medical records

HU Dehua, FANG Hao, ZHOU Yukui, LUO Aijing

Department of Medical Information, School of Life Sciences, Central South University, Changsha 410083, China

**Abstract: Objective** To explore an effective method for providing recommendation knowledge related to diagnosis and treatment for patients based on medical knowledge base and electronic medical records. **Methods** The content filtering method of medical knowledge base was combined with electronic medical records to establish an integrated scoring matrix. The fusion of the results obtained with patient-based collaborative filtering and diagnosis-based collaborative filtering was used to construct a hybrid recommendation method, so as to reduce the error of prediction results. With mean absolute percentage error as the evaluation index, the prediction results obtained with hybrid recommendation method, patient-based collaborative filtering and diagnosis-based diagnostic collaborative filtering were evaluated. **Results** The average absolute percent error of hybrid recommendation method was lower than that of patient-based collaborative filtering and diagnosis-based collaborative filtering. The limitation of this study is that the types of hypertensive diseases in medical knowledge base were less than those in electronic medical records, and that the effects of the disease coverage on the recommendation results remained to be further explored. **Conclusion** The hybrid recommendation method which is superior to patient-based collaborative filtering and diagnosis-based diagnostic collaborative filtering can be used to provide medical knowledge for patients, having a good application prospect.

**Keywords:** medical knowledge base; electronic medical record; collaborative filtering; content filtering

### 前言

随着互联网技术的迅速发展和大数据时代的到来,电子医疗信息的规模也日益庞大<sup>[1-2]</sup>。受限于精

力和记忆能力,人类的信息存储量有限,无法对所有信息进行有效的存储和利用,信息过载所带来的压力也越来越明显。同时人们对自身健康的关注度不断提高,对医疗信息的质量和准确性不断提出新的要求。

近年来医疗信息化在政府和市场的推动下,已经取得了一定的研究成果。但诊疗知识库领域的研究方向多为诊疗知识库的构建和临床辅助决策支持,研究如何为患者提供准确且权威的疾病相关信息的文献资料较少<sup>[3-5]</sup>。目前,患者自主了解自身疾病情况多采用互联网检索、询问他人的方式,该方式虽然相对便捷,但无法保证获取的疾病相关信息的

**【收稿日期】**2019-01-15

**【基金项目】**国家社科基金重点项目(17ZDA037);教育部-中国移动联合实验室移动医疗项目;万方数据-中南大学医学知识服务创新联合实验室规划项目;中南大学中央高校基本科研业务费专项资金(2018zzts303)

**【作者简介】**胡德华,教授,博士生导师,研究方向:医药信息管理、医学数据挖掘、信息检索、信息查询行为等, E-mail: hudehua2008@qq.com

**【通信作者】**罗爱静,教授,博士生导师, E-mail: luojia@mail.csu.edu.cn

准确性和权威性,导致患者对自身健康情况的认识产生偏差。因此,为保证患者获取准确的疾病知识,开展基于临床诊疗知识库和电子病历的整合推荐方法研究具有重要的现实意义。

本文探讨根据临床诊疗知识库和患者电子病历信息,通过对症状分词,探讨基于内容推荐和协同过滤的整合推荐方法,识别其诊疗信息,并映射到疾病诊疗知识库中相应的疾病、检查、药品信息上,为患者提供诊疗过程中疾病相关信息,满足其对自身健康管理的需求。

## 1 国内外研究现状

疾病、药物、检查、症状等医疗信息作为诊疗知识库的核心内容,贯穿患者的整个健康档案,是医生、患者最为关心的内容。临床诊疗知识库可以为患者提供全面且权威的临床指南、医疗知识,从而提高医患整体的认识水平。目前,国内已构建了许多诊疗知识库,且各个知识库的特点和内容存在或多或少的差异,如中国医院知识总库、中国疾病知识总库主要收录临床文献;万方临床诊疗知识库、中国医学科学院医学信息研究所临床医学知识库则以疾病、检查、药品、指南为基础<sup>[6]</sup>。相比之下,国外诊疗知识库,如UpToDa等,在临床证据来源、规模、更新频率、评价等方面更加完善,更注重临床证据资源,加以整合分析,并对临床诊疗证据的可信度进行评估。UpToDa等知识库主要由知名出版机构或者学术团体创建和维护,多为循证总结类的临床知识库,对知识库内容的编辑和审核都较为严格,这得益于其以医生作者和专业编辑团队为核心,依靠临床经验,并且根据临床证据不断对诊疗知识库进行更新和完善。虽然国外临床知识库的质量优于国内,但是碍于语言差异等因素,选择国内相对高质量的知识库作为临床诊疗知识推荐的基础,实用性更强。因此,本文选择同样经过临床医生编辑、审核构建的万方临床诊疗知识库,其无论在知识数据支持还是实用性上,都更为合适,能更好地为患者提供疾病、药品、检查、症状等方面的信息。

在CNKI和WOS两大数据库中检索发现,“临床知识库”和“电子病历”相关研究文献以探讨如何实现有效的临床辅助和决策支持为主,依靠临床知识库或电子病历中的规则约束医疗过程中医生和患者的行为,为临床诊断和治疗提供建议,从而降低医疗意外的发生率,提高整体医疗质量<sup>[7-9]</sup>。聂丽雨等<sup>[10]</sup>利用临床药学知识库加强抗菌药物的管理,规范激素类药物与抗肿瘤药物的使用,并为临床用药提供辅

助作用,以降低患者用药损害,从而提高医生用药水平,显著降低用药错误率,实现精细化管理临床用药。尹梓名等<sup>[11]</sup>基于国际头痛诊断标准构建知识库,并以此为基础,开发了覆盖头痛诊断整个流程的原发性头痛辅助决策系统,对常见的原发性头痛诊断具有较高的准确率。Ohno等<sup>[12]</sup>构建临床知识库,用于开发临床决策支持系统和临床数据的摘要,提出记录的适当参数,并验证疾病、药物和实验室结果。总之,临床知识库和电子病历的相关研究从20世纪90年代开始呈快速增长趋势,并取得了一定的成果,主要面向医疗工作者,而针对患者群体,提供相应医疗知识的研究较少。

随着人们生活水平的提高以及健康意识不断加强,对自我健康管理的需求也愈发强烈,而基于权威的临床知识库为患者推荐与其健康状况相关的疾病知识,有助于提高患者对自身状况的认识,避免从其他渠道获取错误知识而导致患者认知的偏差。另一方面,医患信息不对称、患者缺乏医学知识也是导致医患关系紧张的重要因素。为患者推荐相关的医学知识,有助于缓解医患之间的紧张关系,推动医患关系的健康良性发展<sup>[13]</sup>。因此,探究个性化的疾病知识推荐,对满足患者健康管理的需求,了解自身健康状况,改善医患关系具有重要的理论意义和实践价值。

## 2 整合推荐方法

本文在基于诊疗知识库的内容过滤推荐方法以及基于电子病历的患者和疾病诊断的协同过滤推荐方法基础上,提出一种整合推荐方法。整合推荐方法模型如图1所示。

### 2.1 基于诊疗知识库的内容过滤

基于诊疗知识库的内容过滤采用中文分词、TF-IDF、倒排索引等技术算法来实现。以临床诊疗知识库内容为疾病诊断元数据,以患者电子病历中主述为主相关字段表达患者特征,计算两者之间的相似度,得到患者与相应疾病相关度的评分。针对疾病诊断元数据和电子病历中的文本信息,以Python中的Jieba包为主要工具,加载基于临床诊疗知识库构建的专业术语词典,以优化分词效果,对文本信息进行分析处理。TF-IDF是“词频”与“逆文档频率”的乘积,其与词在某一记录中出现的次数成正比,与所有记录中出现的次数成反比<sup>[14]</sup>。

本文引入了疾病诊断元数据的基于诊疗知识库内容的推荐,提升了推荐结果的相关性,使得推荐结果可解释,也更易于被患者所感知,但同时,该评分结果缺乏个性化,且依赖于对疾病诊断元数据的深度分析。在此基础上,引入患者属性,可以构建相应

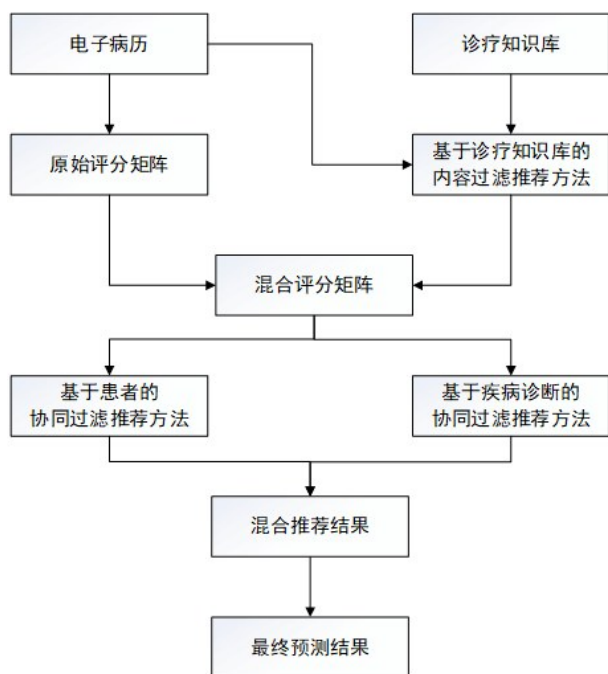


图1 整合推荐方法模型图

Fig.1 Diagram of hybrid recommendation model

的患者模型,分析其兴趣需求,分析结果具有个性化,但推荐精度低,存在马太效应和因为患者F行为稀疏而导致覆盖率低等问题<sup>[15]</sup>。

## 2.2 基于电子病历的患者协同过滤 (Patient-based Collaborative Filtering, PbCF)

基于电子病历的PbCF是电子病历中具有类似疾病诊断的其他患者给予测试疾病诊断的相似度来估计测试患者对测试疾病诊断的相似度的一种过滤方法,其相似度计算公式如式(1)所示:

$$SU_{ij} = \frac{(\sum_{i \in I(i,j)} R_{ui} \cdot R_{uj}) \cdot (|I(i,j)| - 1)}{\sqrt{\sum_{i \in I(i,j)} R_{ui}^2 \cdot \sum_{i \in I(i,j)} R_{uj}^2 \cdot (|I(i,j)| - 1 + \lambda)}} \quad (1)$$

其中,  $SU_{ij}$  表示标号为  $i$ 、 $j$  的两个患者的相似度,  $I(i,j)$  表示患者  $i$ 、 $j$  同时患有的疾病诊断集合,  $R_{ui}$  表示患者  $u$  对疾病诊断  $i$  的评分,  $\lambda$  为惩罚(平滑)参数,避免出现除零的情况。PbCF的实现基于两个假设:当前患者对于其相似患者所涉及的疾病存在一定的患病风险;具有相似疾病的患者在未来也具有相似性的疾病。与基于内容的推荐相比,协同过滤的优势表现在充分利用群体的智慧,其推荐精度要高于基于内容的推荐算法,更加利于挖掘患者或疾病诊断间的隐含相关性,但相应的,协同过滤的推荐结果解释性较差,对时效性较强的疾病诊断不适用,同时面临冷启动问题<sup>[16]</sup>。冷启动主要涉及患者冷启动、疾病诊断冷启动、系统冷启动3个方面<sup>[17]</sup>。解决患者冷启动可利用患者提供的年龄、性别等数据做粗粒

度个性化处理,要求患者对一些疾病诊断进行反馈,然后给患者推荐相似的疾病诊断。疾病诊断冷启动,多利用内容信息,将疾病诊断推荐给具有相似疾病诊断的患者,同时可以采用基于内容的推荐来解决该问题,频繁更新相关数据。再者,系统冷启动需引入专家知识,通过高效的方式迅速建立起疾病诊断相关性矩阵。

## 2.3 基于电子病历的疾病诊断协同过滤 (Diagnostic-based Collaborative Filtering, DbCF)

基于电子病历的DbCF是电子病历中被同一患者患有的其他疾病诊断对应测试患者的相似度来估计测试患者对测试疾病诊断的相似度的一种过滤方法。其相似度计算公式如式(2)所示:

$$SI_{ij} = \frac{(\sum_{u \in U(i,j)} R_{ui} \cdot R_{uj}) \cdot (|U(i,j)| - 1)}{\sqrt{\sum_{u \in U(i,j)} R_{ui}^2 \cdot \sum_{u \in U(i,j)} R_{uj}^2 \cdot (|U(i,j)| - 1 + \lambda)}} \quad (2)$$

其中,  $SI_{ij}$  表示标号为  $i$ 、 $j$  的两个疾病诊断的相似度,  $U(i,j)$  表示同时对  $i$ 、 $j$  有评分的患者集合,  $R_{ui}$  表示患者  $u$  对疾病诊断  $i$  的评分,  $\lambda$  为惩罚(平滑)参数。DbCF是建立在患者可能患有与该疾病相关的其他疾病以及疾病之间的相关性较为稳定两个假设之上。与基于内容的协同过滤相比,DbCF适合疾病诊断丰富且患者个性化需求强烈的领域,当患者有新疾病,可以推荐与该疾病诊断相关的其他疾病,但必须在离线更新疾病诊断相似度表的情况下将新疾病诊断推荐给患者。

## 2.4 整合推荐算法

本文以患者-疾病诊断的交互评分矩阵为原始评分矩阵,结合基于诊疗知识库内容过滤所产生的归一化后结果,形成整合评分矩阵,可防止评分矩阵过于稀疏,也将患者属性所表达的与疾病诊断相关的因素引入评分矩阵,使得评分矩阵更加客观。整合评分矩阵分别由基于患者和基于疾病诊断的协同过滤预测,再依据式(3)结合两者评分得到最终预测结果:

$$S_{ij} = \alpha \cdot SU_{ij} + \beta \cdot SI_{ij} \quad (3)$$

其中,  $S_{ij}$  为最终预测结果,  $SU_{ij}$  和  $SI_{ij}$  分别为PbCF和DbCF的预测值,  $\alpha$  和  $\beta$  分别表示PbCF和DbCF所对应的权重值。  $\alpha$  和  $\beta$  值采用线性规划来确定最优值<sup>[18]</sup>。本文采用的线性规划模型如式(4)所示:

$$\text{Minimize } \frac{\sum_{k=1}^n |\alpha \cdot SU_{ij} + \beta \cdot SI_{ij} - y_{ij}|}{n} \quad (4)$$

$$\text{Subject to } \alpha + \beta = 1; \alpha, \beta \geq 0$$

该线性规划模型为了求预测最终结果  $\alpha \cdot SU_{ij} + \beta \cdot SI_{ij}$  和实际结果  $y$  之间差异最小时所对应



的 $\alpha$ 、 $\beta$ 的值。其中 $y_{ij}$ 代表患者 $i$ 对疾病诊断 $j$ 的实际评分值。本文采用十折交叉验证,将整合评分矩阵分为10组训练集和测试集,分别计算PbCF和DbCF的预测值,代入线性规划模型,求得各组 $\alpha$ 、 $\beta$ 的最优值,最后取10组数据结果的平均值,作为 $\alpha$ 、 $\beta$ 的最终最优值。

### 3 实验评估

#### 3.1 数据源

本文所选取的数据源主要分为两部分,分别是万方临床诊疗知识库的高血压相关数据和某医院高血压疾病患者的电子病历数据。

万方临床诊疗知识库包含疾病库、检查库、药品库、指南规范、循证文献和病例报告等数据库,由国内上千名临床资深专家团队共同创作,确保了诊疗知识库的可靠性,为本文研究基于诊疗知识库实现知识推荐提供有力的数据支持。其中,本文所实现的智能推荐功能,以疾病为结果,以患者为研究对象,进行实验评估。因此,在上述的6个数据库中,选取与患者关系更为紧密的疾病数据库,作为知识推荐的基础,且以高血压相关的疾病数据作为示范疾病诊断元数据。抽取某医院10 000名高血压患者的电子病历信息,以患者的诊断信息作为原始评分矩阵构建的依据,设定若患者具有某一疾病诊断结果,则该患者对该疾病诊断评分为1。虽然临床诊断中有主要诊断和次要诊断之分,但两者的划分主要依据当前住院的主要原因,且就患者个体而言,主要诊断与次要诊断所对应的疾病与患者密切相关程度不一定具有较大差异,所以诊断结果中出现的诊断结果都将评分设为1。在此基础上,引入基于诊疗知识库内容过滤的评分,可解决原始评分矩阵中评分的无差异化和稀疏问题。

#### 3.2 实验结果和评估

十折交叉验证后,取 $\alpha$ 、 $\beta$ 各组数据结果的均值,得到 $\alpha=0.61$ , $\beta=0.39$ ,因此,最终预测结果 $S_{ij}=0.61 \cdot SU_{ij}+0.39 \cdot SI_{ij}$ 。本文选取平均绝对百分误差(MAPE)作为预测结果的评价指标,对PbCF、DbCF和整合推荐方法3种方法进行评估对比,每个方法都针对6组 $K$ 近邻( $K$ -Nearest Neighbor, KNN)进行试验,由于 $K$ 值一般不超过50,所以 $K$ 选择10、20、30、40、50、60,其中 $K=60$ 作为观察结果,与其他结果做对比,分析MAPE值是否有明显的优化<sup>[19]</sup>。结果如表1所示。

上述3种方法的最小MAPE值对比结果表明整合推荐方法的MAPE最小值为0.143,比PbCF和

表1 PbCF、DbCF和整合推荐方法的平均绝对百分误差值  
Tab.1 Mean absolute deviation of PbCF, DbCF and hybrid recommendation method

近邻数	PbCF	DbCF	整合推荐方法
10	0.195	0.193	0.156
20	0.187	0.191	0.143
30	0.181	0.186	0.145
40	0.176	0.187	0.148
50	0.193	0.197	0.150
60	0.187	0.193	0.149

PbCF:基于电子病历的患者协同过滤;DbCF:基于电子病历的疾病诊断协同过滤

DbCF所对应的最小值0.176和0.186要小,表明整合推荐方法的预测效果明显优于PbCF和DbCF。

### 4 结论

本文提出一种在诊疗知识库内容过滤的基础上,融合PbCF和DbCF整合推荐方法。以临床诊疗知识库和电子病历中高血压数据为实验数据,对整合推荐方法进行评估,为高血压患者推荐诊疗知识库中相关的疾病知识。结果表明,整合推荐方法比独立的PbCF、DbCF效果更好,为患者推荐疾病诊疗知识,具有广泛的应用前景。但由于所选诊疗知识库的高血压疾病种类覆盖面不及电子病历中高血压患者所涉及的疾病种类,所以扩大诊疗知识库疾病覆盖面对于推荐结果的影响仍需进一步探讨。

### 【参考文献】

- [1] 苏玉强,刘颖. 基于消息的嵌入式系统应用框架研究[J]. 计算机测量与控制, 2011, 19(5): 1142-1144.  
SU Y Q, LIU Y. Study on Message based application framework for embedded system[J]. Computer Measurement and Control, 2011, 19(5): 1142-1144.
- [2] 夏一行,曲琳,陈耀武. 嵌入式数据分析类仪器软件应用框架的开发[J]. 江南大学学报(自然科学版), 2008, 7(3): 258-262.  
XIA Y X, QU L, CHEN Y W. Application framework development for embedded data analysis instrumentation software [J]. Journal of Jiangnan University (Natural Science Edition), 2008, 7(3): 258-262.
- [3] KASHFI H. An open EHR-based clinical decision support system: a case study[J]. Stud Health Technol Inform, 2009, 150: 48.
- [4] YANG J, KANG U, LEE Y. Clinical decision support system in medical knowledge literature review[J]. Inform Technol Manag, 2016, 17(1): 1-10.
- [5] DOUALI N, ABDENNOUR M, SASSO M, et al. Noninvasive diagnosis of nonalcoholic steatohepatitis disease based on clinical decision support system.[J]. Stud Health Technol Inform, 2013, 192: 1178.
- [6] 邓盼盼,李军莲. 国外临床知识库发展现状及启示[J]. 医学信息学

- 杂志, 2017, 38(3): 77-83.
- DENG P P, LI J L. Development status and enlightenment of foreign clinical knowledge base[J]. Journal of Medical Intelligence, 2017, 38(3): 77-83.
- [7] STEWART K L. Reducing medical liability risk. Live decision support-enabled EMR reduces obstetric medical professional liability with best-practice protocol[J]. Health Manag Technol, 2007, 28(12): 30.
- [8] KING A J, HOCHHEISER H, VISWESWARAN S, et al. Eye-tracking for clinical decision support: a method to capture automatically what physicians are viewing in the EMR[J]. AMIA Jt Summits Transl Sci Proc, 2017: 512-521.
- [9] SETO R, WAKABAYASHI S. Clinical decision support using electronic medical records: for the improvement of diabetes care and proper use of insulin for inpatients[J]. Stud Health Technol Inform, 2015, 216: 961.
- [10] 聂丽丽, 陈妍妍, 廖家智, 等. 基于临床药知识的全过程临床用药管理系统构建[J]. 中国医院管理, 2016, 36(1): 66-67.
- NIE L L, CHEN Y Y, LIAO J Z, et al. Construction and evaluation of clinical pharmacy management system based on the clinical knowledge base[J]. Chinese Hospital Management, 2016, 36(1): 66-67.
- [11] 尹梓名, 董钊, 孔祥勇. 基于国际头痛诊断标准的原发性头痛辅助决策系统[J]. 计算机应用研究, 2019, 36(2): 1-2.
- YIN Z M, DONG Z, KONG X Y. Assistant decision-making system based on international diagnostic criteria for primary headache disorders[J]. Application Research of Computers, 2019, 36(2): 1-2.
- [12] OHNO K, NAGASAWA I, UMEDA M, et al. Development of medical knowledge base for clinical decision support [C]. Proceedings: Applications of Informatics and Cybernetics in Science and Engineering, 2004: 193-198.
- [13] 张莉, 陶骏贤. 医患双方对医患关系的认知差异研究现状及展望[J]. 江苏卫生事业管理, 2018, 29(1): 22-24.
- ZHANG L, TAO J X. Current situation and prospect of cognitive differences between doctors and patients on doctor-patient relationship [J]. Jiangsu Healthcare Administration, 2018, 29(1): 22-24.
- [14] GUO A, YANG T. Research and improvement of feature words weight based on TFIDF algorithm[C]//2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference. IEEE, 2016: 415-419.
- [15] AYGÜN C, YILDIZ O. Development of content based book recommendation system using genetic algorithm [C]//Signal Processing and Communication Application Conference. IEEE, 2016: 1025-1028.
- [16] PALAK R, NGUYEN N T. An effective collaborative filtering based method for movie recommendation [M]//Multimedia and Network Information Systems. Springer International Publishing, 2017.
- [17] KBAIER M E, MASRI H, KRICHEN S. A personalized hybrid tourism recommender system [C]//IEEE/ACS International Conference on Computer Systems and Applications. IEEE, 2018: 244-250.
- [18] GEETHA G, SAFA M, FANCY C, et al. A hybrid approach using collaborative filtering and content based filtering for recommender system [M]. Bristol: IOP Publishing Ltd, 2018.
- [19] THAKKAR P, VARMA K, UKANI V. Outcome fusion- based approaches for user-based and item-based collaborative filtering [C]. International Conference on Information and Communication Technology for Intelligent Systems, 2017: 127-135.

(编辑: 薛泽玲)