

控制与决策

Control and Decision

融合粗糙集与GRA的异构信息多准则三支推荐及其在医疗推荐中的应用

张萌, 孙秉珍, 王婷, 楚晓丽, 同思蓉

引用本文:

张萌, 孙秉珍, 王婷, 楚晓丽, 同思蓉. 融合粗糙集与GRA的异构信息多准则三支推荐及其在医疗推荐中的应用[J]. 控制与决策, 2022, 37(7): 1883–1893.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1631>

您可能感兴趣的其他文章

Articles you may be interested in

[基于矩阵的双论域模糊概率粗糙集增量更新算法](#)

Incremental updating of fuzzy probability rough sets over two universes based on matrix method

控制与决策. 2021, 36(3): 553–564 <https://doi.org/10.13195/j.kzyjc.2019.0692>

[基于知识粒度特征的多目标粗糙集属性约简算法](#)

Multi objective rough set attribute reduction algorithm based on characteristics of knowledge granularity

控制与决策. 2021, 36(1): 196–205 <https://doi.org/10.13195/j.kzyjc.2019.0490>

[基于地标特征和元学习方法推荐最适用优化算法](#)

Recommending best suitable metaheuristic based on landmarking feature and meta-learning approach

控制与决策. 2021, 36(5): 1223–1231 <https://doi.org/10.13195/j.kzyjc.2019.0993>

[混合决策下考虑第三方偏好的远程医疗服务匹配方法](#)

Matching method for telemedicine service considering third-party preferences in context of mixed decision-making

控制与决策. 2021, 36(11): 2803–2811 <https://doi.org/10.13195/j.kzyjc.2020.0447>

[大群体应急决策中考虑属性关联的偏好信息融合方法](#)

Preference information fusion method of large groups emergency decision-making based on attributes association

控制与决策. 2021, 36(10): 2537–2546 <https://doi.org/10.13195/j.kzyjc.2020.0117>

融合粗糙集与 GRA 的异构信息多准则三支推荐 及其在医疗推荐中的应用

张 萌¹, 孙秉珍^{1†}, 王 婷¹, 楚晓丽^{1,2}, 同思蓉¹

(1. 西安电子科技大学 经济与管理学院, 西安 710126;

2. 广州中医药大学 第二附属医院中医药大数据研究团队, 广州 510260)

摘要: 在临床实践中, 医疗推荐可能存在数据多源异构和推荐项目多准则的问题, 考虑到医疗推荐的这些特征, 定义了异构信息系统中不同数据类型的距离测度, 实现多源异构数据的有效处理. 首先, 根据两个对象之间的混合距离得到异构信息系统中的二元关系, 并构建异构信息粗糙集模型; 然后, 将多准则推荐与多准则决策方法 (MCDM) 相结合, 运用灰色关联分析 (GRA) 聚合每个项目下多准则评分将其转化为单评分推荐; 最后, 在异构信息粗糙集模型的基础上引入三支决策, 同时基于协同过滤方法实现三支推荐, 考虑了推荐过程中的决策成本. 在医疗应用部分采用临床数据实验, 验证了所提出的模型能够为临床诊断提供知识支持, 有效降低推荐决策成本, 提高推荐的准确性.

关键词: 多准则推荐; 医疗推荐; 三支推荐; 灰色关联分析; 粗糙集; 异构信息

中图分类号: TP18

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1631

开放科学(资源服务)标识码(OSID):



引用格式: 张萌, 孙秉珍, 王婷, 等. 融合粗糙集与 GRA 的异构信息多准则三支推荐及其在医疗推荐中的应用 [J]. 控制与决策, 2022, 37(7): 1883-1893.

Multi-criteria three-way recommendation of heterogeneous information based on rough set and GRA and its application in medical recommendation

ZHANG Meng¹, SUN Bing-zhen^{1†}, WANG Ting¹, CHU Xiao-li^{1,2}, TONG Si-rong¹

(1. School of Economics and Management, Xidian University, Xi'an 710126, China; 2. Traditional Chinese Medicine Big Data Research Team, The Second Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou 510260, China)

Abstract: Medical recommendation may have the problems of multi-source heterogeneous data and multi-criteria of recommendation items in clinical practice. Considering the characteristics of medical recommendation, this paper defines the distance measure of different types of data in heterogeneous information systems and the effective processing of multi-source heterogeneous data is realized. Firstly, meanwhile the binary relationship in heterogeneous information systems is obtained according to the hybrid distance between two objects and the heterogeneous information rough set model is constructed subsequently. Then the multi-criteria recommendation is combined with the multi-criteria decision-making method (MCDM). And the multi-criteria rating of each item is aggregated by grey relational analysis (GRA) to transform multi-criteria recommendation into single-rating recommendation. Finally, three-way decision-making is presented on the basis of the heterogeneous information rough set model, and three-way recommendation is achieved based on the collaborative filtering method, which considers the cost of decision-making in the process of recommendation. Clinical practice data in the part of medical application is used to verify that the proposed model can provide a knowledge support for clinical diagnosis, which effectively reduces the cost of recommendation and improves the accuracy of recommendation.

Keywords: multi-criteria recommendation; medical recommendation; three-way recommendation; grey relational analysis; rough set; heterogeneous information

收稿日期: 2020-11-25; 录用日期: 2021-04-21.

基金项目: 国家自然科学基金项目 (72071152, 71571090, 81774218); 西安市科技计划软科学研究项目 (XA2020-RKXY-0086); 陕西省高校青年创新团队项目 (2019); 广东省教育厅基金项目 (2018GWQNCX050).

†通讯作者. E-mail: bzsun@xidian.edu.cn.

0 引言

推荐系统根据用户偏好信息,利用智能算法为用户提供个性化推荐,是解决信息过载的有效方法之一.近年来“互联网+”战略的提出以及互联网技术向传统行业的下沉,推荐系统已广泛应用在多个领域^[1-3].目前推荐算法在医疗领域的应用研究主要集中在在线医疗社区中的医生推荐^[2]和内容推荐^[3].而侯梦薇等^[4]提出未来推荐系统在医疗行业的应用应主要集中在辅助诊疗、就诊引导、合理用药等业务.其中辅助诊疗是指推荐系统利用临床数据,将患有同种疾病的病例就诊数据从不同的角度建模,从而推荐个性化的检验、药品和治疗方案等.由于患者诊断数据和电子病历信息的特殊性和隐私性,目前关于推荐系统在辅助诊疗模块的应用研究尚不够广泛和深入,导致大量的医疗数据没有得到有效的利用和开发.

西医诊疗通过医疗仪器和临床实验对疾病进行诊断,是一套规范化和流程化的体系.医生在诊断过程中通常会让患者做多项检查辅助确诊疾病类型.而实际中多数检查对于确诊病因所发挥的作用并不显著,只有一些异常的检查项目扮演着至关重要的角色.因此,在保证辅助医生诊疗效果不变的基础上,利用推荐算法和数据挖掘理论挖掘患者与历史数据库中的相似病例,实现个性化检查项目的精准推荐能够缓解“过度检查”造成的医疗资源的浪费,同时“按需检查”和“适度检查”能够避免患者检查不必要的医疗项目,减少患者的就医费用.在基于检查项目的医疗推荐背景下(后文所说的医疗推荐均是基于此背景),传统的单评分推荐不能很好地适应医疗推荐决策,这是由于每个医疗检查项目包含多个指标衡量项目的正常与否.在医疗推荐中其输入模型所涉及到的数据往往具有多源性^[4],这导致患者的检查信息包括多种数据类型:实数值、符号值、区间值和语义值等.因此医疗推荐具有数据多源异构和推荐项目多准则的鲜明特征,如何解决上述问题获得项目的总体评分值并对不同的患者作出个性化推荐,是本文的主要研究内容.

粗糙集理论是Pawlak^[5]提出的一种处理不确定性信息的有效工具,传统的粗糙集模型建立在离散型信息系统的基础上,但医疗推荐中存在数据异构的问题,这导致单一数据类型的信息系统和传统的粗糙集模型无法满足医疗推荐中不同类型的数据测度和衡量,不利于多准则项目总体评分的计算.因此,本文在已有文献的基础上研究具有实数值、符号值、区间值

和语义值的异构信息系统,通过定义信息系统中两个对象在不同数据类型属性集上的距离测度来处理复杂医疗数据,同时基于异构信息系统上两个对象之间的混合距离得到对象集上的二元关系^[6],构建了基于异构信息的粗糙集模型^[7-8],奠定了后续三支医疗推荐的理论基础.

异构信息粗糙集模型解决了医疗推荐中数据异构的问题,但医疗推荐还具有项目多准则的特点.多准则推荐是在推荐过程中既考虑用户对项目的总体评分,又考虑用户对项目不同准则的评分,能够更好地捕捉和挖掘用户的偏好信息.近年来,推荐领域的研究焦点也逐渐转移到多准则推荐,涌现出很多相关的研究^[9-11].Adomavicius等在文献[12]中阐述了多准则推荐的两种基本研究思路:基于相似度的方法和基于聚合函数的方法,奠定了多准则推荐的研究范式.而医疗推荐中的项目准则具有差异性,这导致已有的方法运用在医疗多准则推荐中存在较大的局限性.MCDM是决策科学领域研究热点之一,其与多准则推荐本质上都是对多维信息加工做出优选.Manouselis等^[13]从决策对象、决策准则、全局偏好模型、决策支持过程等4个方面论述了MCDM运用在多准则推荐中的巨大潜能.灰色关联分析(grey relational analysis, GRA)是灰色系统理论的分支,是解决不确定性问题的有效工具.本文考虑到医疗系统的“灰色性”和GRA^[14]的应用场景,将其引入医疗多准则推荐模型,充分考虑了医疗推荐场景中的不确定性,且利用异构信息粗糙集模型中定义的距离公式衡量指标序列与参考序列的相关程度,实现并反映项目的总体水平,将多准则推荐退化为单评分推荐.

在解决了医疗推荐中的数据异构和项目多准则的问题后,医疗多准则推荐退化为传统的单评分推荐.而医疗推荐过程基于代价敏感,并且对推荐的准确度要求较高,传统“非此即彼”的三支决策在信息不充分的条件下会造成较高的误判率和决策风险.三支决策^[15]能够根据人们对信息的掌握程度,作出3种决策:接受、拒绝和延迟,其对于延迟决策可以等待信息充分时再作出进一步的判断.由于决策过程考虑了决策成本,同时能够有效提高决策准确性,三支决策也逐渐应用到了推荐领域^[16-18],其推荐效率也得到有效验证.因此,本文通过异构信息粗糙集模型引入三支决策,并在协同过滤的基础上构建了三支推荐,考虑了医疗推荐中的误推荐成本和学习成本,提

高了推荐的准确度。

粗糙集和灰色理论都是解决不确定性问题的有效工具,虽然它们理论基础不同,但都具有无需先验信息的共同优势. 本文将GRA方法引入粗糙集并构建了多准则三支推荐,既能利用GRA方法处理灰色不确定性问题,同时又能利用其计算项目的总体评分评估医疗检查项目,使计算结果更具有可解释性,能够更好地挖掘医疗潜在知识. 在实验部分,运用临床医疗数据对本文提出的模型进行验证,结果表明该方法能够有效降低推荐成本,提高推荐的准确度,解决了医疗推荐中存在的项目多准则、数据多源异构等问题,并充分结合了推荐技术、多准则决策和三支决策的优势. 尤其在我国医疗资源分布不均、水平差异较大和医疗数据过载的背景下,医疗推荐作为解决信息过载的有效工具在一定程度上能够优化医疗资源的配置实现精准医疗,实现数字医疗向智慧医疗的过度转型.

1 基于异构信息的多准则医疗推荐

1.1 单评分推荐

推荐算法能够根据用户偏好推荐其可能感兴趣的条目. 目前主流的推荐方法包括协同过滤推荐、基于模型的推荐、基于深度学习^[19-20]和神经网络^[21]的推荐等,这些方法都取得了非常好的推荐效果. 本文的研究采用基于用户的协同过滤方法,其能从用户历史评分记录中筛选出具有相似特征的最近邻用户群,并预测目标用户对未评分项目的评分,将较高评分的项目推荐给目标用户. 表1为用户-项目单评分矩阵表,其中 $U = \{u_1, u_2, \dots, u_n\}$ 是用户集, $I = \{i_1, i_2, \dots, i_m\}$ 是项目集, $R \times I$ 则代表了用户对项目的总体评分集合. 对于所有目标用户来说,推荐算法旨在寻找推荐度最大的项目 I^* ,其中 $I^* = \arg \max_{i \in I} r(u, i)$. 而衡量用户之间相似性的方法有余弦相似性和相关性相似等多种方法^[1].

表1 用户-项目单评分矩阵表

user	item			
	i_1	i_2	...	i_m
u_1	r_{11}	r_{12}	...	r_{1m}
\vdots	\vdots	\vdots	\ddots	\vdots
u_n	r_{n1}	r_{n2}	...	r_{nm}

通过计算用户之间的相似度可以得到目标用户 u 的最近邻域用户群 N_u ,并根据用户 u 对项目 i 的预

测评分

$$P_{u,i} = \bar{R}_u + \frac{\sum_{n \in N_u} \text{sim}(u, n) \times (R_{n,i} - \bar{R}_n)}{\sum_{n \in N_u} |\text{sim}(u, n)|} \quad (1)$$

作出相应的推荐,其中 $\text{sim}(u, n)$ 为用户 u 和用户 n 的相似度^[1].

1.2 多准则推荐

古语所言“横看成岭侧成峰,远近高低各不同”,关于事物的评价需要从多个维度去衡量才能有全面的认知. 关于多维度场景的研究,学者们从不同领域进行了探索:在决策科学领域对MCDM进行广泛地研究;在粒计算领域其根据属性具有不同粒度的特征构建多尺度决策信息系统等^[22]. 同样地,在推荐系统中仅仅有总体评分信息会限制用户的偏好表达且不利于挖掘用户潜在偏好信息,因此在推荐领域也展开了多准则推荐的研究^[9-11]. 表2为多准则推荐中用户-项目评分表,可以用 $R : \text{User} \times \text{Items} \rightarrow R_0 \times R_1 \times \dots \times R_k$ 来表示,其中 R_0 表示总体评分值集合, $R_t (t = 0, 1, \dots, k)$ 表示第 t 个准则的评分值集合. 在表2中第1行和第1列中所对应的条目表示用户 u_1 对项目 i_1 的多准则评分,对于5(2, 2, 8, 8),5代表总体评分,(2, 2, 8, 8)分别表示用户 u_1 对项目 i_1 下4个准则的评分. 因此在多准则推荐中除项目总体评分外,还能提供更多用户在不同维度上的偏好信息,所以在推荐过程中能够更加准确地预测用户的偏好. 与单评分推荐一样,多准则推荐的关键问题仍然是预测用户未评分项目的总体评分并向用户推荐最佳项目.

表2 用户-项目多准则评分表

user	item			
	i_1	i_2	...	i_m
u_1	5(2, 2, 8, 8)	7(5, 5, 9, 9)	...	6(3, 3, 9, 9)
\vdots	\vdots	\vdots	\ddots	\vdots
u_n	7(9, 9, 5, 5)	5(2, 2, 8, 8)	...	6(9, 9, 3, 3)

目前关于多准则推荐的研究主要有以下两种研究思路^[12]:一种是基于相似性的方法,其建立在协同过滤的基础上. 对于用户 u_i 和 u_j ,其第 t 个准则的相似性可用 $\text{sim}_t(u_i, u_j) (t = 0, 1, \dots, k)$ 来表示,那么其总体相似性的计算可通过聚合每个准则(包括总体评分)的相似性,比如平均相似性和最差原则相似性

$$\text{sim}(u_i, u_j) = \frac{1}{k+1} \sum_{t=0}^k \text{sim}_t(u_i, u_j), \quad (2)$$

$$\text{sim}_{\min}(u_i, u_j) = \min_{t=0, \dots, k} \text{sim}_t(u_i, u_j). \quad (3)$$

在得到用户之间总体相似性后,可以采用协同过滤的方法进行推荐.

另一种是基于聚合函数的方法,该方法不限于特定的推荐方法,其假设前提是项目的总体评分不仅仅独立于其他准则的评分,而且还是其他多准则评分的聚合函数,即 $f: r_0 = f(r_1, \dots, r_k)$. 也就是说这种方法假设项目的总体评分与多个准则的评分具有一定的关系. 该方法的计算思路和步骤如下:

1) 预测多准则评分. 首先将 $k + 1$ 维多准则评分空间分解成 $k + 1$ 个单评分推荐问题,即可以用传统单评分 $\text{User} \times \text{Item}$ 矩阵表示(如表1所示),此时多准则推荐问题由 $R: \text{User} \times \text{Items} \rightarrow R_0 \times R_1 \times \dots \times R_k$ 转换为 $k + 1$ 个单评分推荐问题 $R: \text{Users} \times \text{Items} \rightarrow R_t (t = 0, 1, \dots, k)$. 此时可以采用现有的单评分推荐方法预测每个准则未知的评分.

2) 学习聚合函数. 这一步主要是评估和学习项目总体评分与多准则评分之间的关系,可以通过领域内专家知识、统计技术和机器学习^[9-11]等方法利用已知的评分来获得该函数 f .

3) 预测总体评分. 这一步使用步骤1)中预测的多准则评分和步骤2)中学习的聚合函数 f 直接计算每个项目未知的总体评分 $r'_0 = f(r'_1, \dots, r'_k)$.

1.3 基于异构信息的多准则医疗推荐

已有的推荐研究多数建立在单一数字评分的基础上,而在医疗领域其数据获取和来源是多渠道的,并且涉及到的数据类型往往具有多样性,常包括以下几类: 1) 患者的基本属性(人口自然属性和统计学信息); 2) 患者的症状描述信息(症状描述、病史、医生诊断史等); 3) 患者的检查检验信息(肝功、肾功、血常规、影像等结果). 例如在患者基本属性中,性别属于符号值,而年龄则属于自然数,患者的症状信息又属于语义值,故患者的临床信息包括实数值、语义值、符号值和区间值等多种信息,因此医疗推荐背景中的数据类型具有异构性.

在医疗推荐中,每个检查项目下包含多个指标衡量每个检查项目的正常与否,例如在血脂项目中包含 TC、TG、HDL-C、LDL-C 等多项指标,因此基于检查项目的医疗推荐是一个典型的多准则推荐. 此外,对于不同的项目其包含的指标都是不同的,这是与以往多准则推荐研究相比最显著的差异. 在多准则推荐常用的 Yahoo!Movies 数据集中,用户从故事情节、导

演执导、演员表演和视觉效果4个维度对电影进行评分,也就是说评价不同项目的维度是相同的,而在医疗推荐中不同项目的评价维度却是不同的.

表3展示了医疗推荐中具有异构信息的多准则推荐评分表,可以用 $R: \text{User} \times \text{Items} \rightarrow R_1 \times \dots \times R_{|k_p|}$ 来表示,其中 $i = \{i_1, i_2, \dots, i_m\}$ 表示 m 个检查项目, $u = \{u_1, u_2, \dots, u_n\}$ 表示 n 个患者, $|k_p|$ 表示第 $p (p \in \{1, 2, \dots, m\})$ 个检查项目包含的准则的数量, $R_t (t = 1, 2, \dots, |k_p|)$ 表示第 t 个准则的值集合. 由表3可以看出,不同的准则其数据类型不同,同时每个项目下所包含的准则还具有差异性. 在1.2节中已经系统地阐述了多准则推荐的研究思路,无论是基于相似度的方法还是基于聚合函数的方法,其核心都要进行维度的拆解,即将不同项目的 k 维多准则拆解成 k 个单准则推荐问题. 但是本文研究的医疗多准则推荐存在准则差异的情况,这会导致其运用在医疗多准则推荐中具有一定的局限性. 另外,在多准则推荐中比较关键的一步是项目总体评分的计算和预测,其在很大程度上影响着推荐的效率,因此如何有效获得项目的总体评分并提升推荐效率是本文主要考虑的问题之一.

表3 基于异构信息的多准则推荐评分表

项目	i_1			...	i_p			...	i_m		
	e_1	...	$e_{ k_1 }$		e_1	...	$e_{ k_p }$		e_1	...	$e_{ k_m }$
u_1	s_0	...	1	3	...	[5, 8]	2	...	F		
u_2	s_1	...	5	4	...	[4, 7]	4	...	M		
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots		
u_n	s_3	...	3	3	...	[2, 6]	3	...	F		

2 基于异构信息的粗糙集模型

针对医疗推荐中存在的异构数据和项目多准则问题,以及医疗多准则推荐与传统多准则推荐在数据结构上的差异,本文抽象出异构信息下差异多准则推荐的一般研究框架,如图1所示.

第1.3节阐述了信息异构是医疗推荐面临的问题之一,因此本节定义了异构信息系统上两个对象之间的距离测度来解决数据异构问题. 粗糙集作为粒计算的有效工具之一,能将复杂问题粒化,更加符合人类认知的过程.

定义1 设 U 为对象集, A 为属性集,且 U 和 A 是有限集合. 将二元组 (U, A) 称为信息系统,对于每个属性 $a \in A$ 都有其对应的信息函数 $a: U \rightarrow V_a$,其中 V_a 是属性 a 的信息函数值集合.

定义2 设 (U, A) 是一个信息系统,对于 $P \subseteq A$, $x, y \in U$, 属性 P 的二元等价关系 $\text{IND}(P)$ 定义为 $\text{IND}(P) = \{(x, y) \in U \times U \mid a(x) = a(y), \forall a \in P\}$.

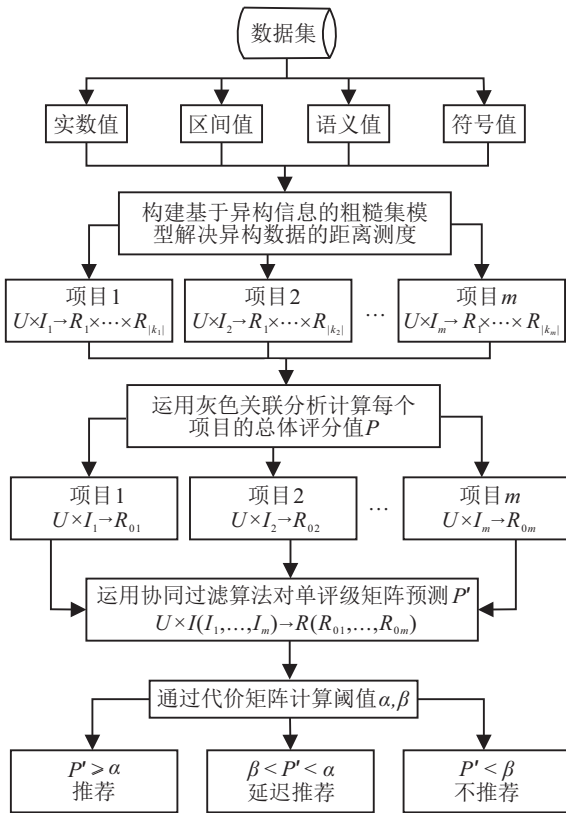


图1 具有异构信息的多准则三支推荐研究框架

定义3 设 (U, A) 是一个信息系统, 对于 $P \subseteq A$, $x, y \in U$, x 关于属性集 P 的等价类 $[x]_P$ 可定义为 $[x]_P = \{y \mid (x, y) \in \text{IND}(P)\}$.

定义4 设 (U, A) 是一个信息系统, 如果 $A = C^i \cup C^b \cup C^r \cup C^s$, 其中 C^i, C^b, C^r, C^s 分别代表区间值、符号值、实数值和语义值, 则称 (U, A) 为异构信息系统.

由于数据的复杂性, $a(x) = a(y)$ 这个条件较难达到, 将导致无法满足信息系统 (U, A) 中的二元等价关系 $\text{IND}(P)$. 事实上, 可以从两个方面考虑 $a(x)$ 与 $a(y)$ 之间的关系: 一方面是相似性, 上述的二元关系就是从这个角度出发的; 另一方面是 $a(x)$ 与 $a(y)$ 之间的差异性, 即用距离函数度量其差异, 当 $a(x) = a(y)$ 时可认为 $a(x)$ 与 $a(y)$ 之间的距离为 0.

接下来定义信息系统中两个对象在属性集上的距离测度, 在此基础上给出异构信息系统上两个对象之间的混合距离. 同时, 构建了基于异构信息系统的粗糙集模型, 并定义了该模型的上近似和下近似^[6].

定义5 设 (U, A) 是一个异构信息系统, 假定 $a \in C^b$, 对于 $x, y \in U$, $a(x)$ 与 $a(y)$ 之间的距离定义为

$$d(a(x), a(y)) = \begin{cases} 0, & a(x) = a(y); \\ 1, & a(x) \neq a(y). \end{cases} \quad (4)$$

定义6 设 (U, A) 是一个异构信息系统, 假定 $a \in$

C^r , 对于 $x, y \in U$, $a(x)$ 与 $a(y)$ 之间的距离定义为

$$d(a(x), a(y)) = \frac{|a(x) - a(y)|}{\theta_a}, \quad (5)$$

其中 $\theta_a = \max\{a(x)\} - \min\{a(x)\}, x \in U$.

定义7 设 (U, A) 是一个异构信息系统, 假定 $a \in C^s, S = \{s_t \mid t = -\tau, \dots, -1, 0, 1, \dots, \tau\}$ 是语义值集合, 对于 $x, y \in U$, $a(x)$ 与 $a(y)$ 之间的距离定义为

$$d(a(x), a(y)) = d(s_\alpha, s_\beta) = \frac{|\alpha - \beta|}{2\tau + 1}. \quad (6)$$

定义8 设 (U, A) 是一个异构信息系统, 假定 $a \in C^i$, 对于 $x, y \in U$, $a(x)$ 和 $a(y)$ 之间的距离定义为

$$d(a(x), a(y)) = \frac{\sqrt{2}}{2} \sqrt{(a(x)^l - a(y)^l)^2 + (a(x)^r - a(y)^r)^2}. \quad (7)$$

其中: $a(x)^l, a(x)^r$ 分别为区间值 $a(x)$ 的下限值和上限值; $a(y)^l, a(y)^r$ 分别为区间值 $a(y)$ 的下限值和上限值.

定义9 设 (U, A) 是一个异构信息系统, 对于 $a \in A, x, y \in U$, x 和 y 基于异构信息系统整体的混合距离定义为

$$d_A(x, y) = \sqrt{\frac{1}{|A|} \sum_{a \in A} d^2(a(x), a(y))}. \quad (8)$$

定义10 设 (U, A) 是一个异构信息系统, 对于 $\theta \in [0, 1]$, 论域 U 上的二元关系记为 T_A^θ , 定义 $T_A^\theta = \{(x, y) \in U \times U \mid d_A(x, y) \leq \theta\}$, 则对于任意 $x(x \in A)$, 其等价类定义为 $[x]_A^\theta = \{y \mid (x, y) \in T_A^\theta\}$.

定义11 设 (U, A) 是一个异构信息系统, 对于 $X \subseteq U$, 基于异构信息系统的粗糙集模型的上、下近似的定义分别为

$$\overline{T}_A^\theta X = \{x \in U \mid T_A^\theta \cap X \neq \emptyset\},$$

$$T_A^\theta X = \{x \in U \mid T_A^\theta \subseteq X\}.$$

对于任意 $X \subseteq U$, 其对应的上、下近似将论域 U 划分为 3 个互不相交的区域, 分别是正域 $\text{POS}(X)$ 、边界域 $\text{BND}(X)$ 和负域 $\text{NEG}(X)$, 有

$$\text{POS}(X) = T_A^\theta X, \quad \text{NEG}(X) = U - \overline{T}_A^\theta X,$$

$$\text{BND}(X) = \overline{T}_A^\theta X - T_A^\theta X.$$

而在基于异构信息的概率粗糙集中, 对于一个对象 x 在属于等价类 $[x]_A^\theta$ 的条件下属于 X 的条件概率为 $P(X \mid [x]_A^\theta) = |X \cap [x]_A^\theta| / |[x]_A^\theta|$, 可以得到如下的等价条件:

$$P(X \mid [x]_A^\theta) = 1 \Leftrightarrow [x]_A^\theta \subseteq X,$$

$$P(X \mid [x]_A^\theta) = 0 \Leftrightarrow [x]_A^\theta \cap X = \emptyset,$$

$$0 < P(X | [x]_A^\theta) < 1 \Leftrightarrow [x]_A^\theta \cap X \neq \emptyset \wedge \neg([x] \subseteq X).$$

根据正域、边界域和负域的定义可以得到如下的等价表示:

$$\begin{aligned} \text{POS}(X) &= \{x \in U \mid P(X | [x]_A^\theta) = 1\}, \\ \text{BND}(X) &= \{x \in U \mid 0 < P(X | [x]_A^\theta) < 1\}, \\ \text{NEG}(X) &= \{x \in U \mid P(X | [x]_A^\theta) = 0\}. \end{aligned}$$

基于异构信息的决策粗糙集模型^[15]是在概率粗糙集的基础上引入阈值对 (α, β) 对论域 U 进行划分,其中 $0 \leq \beta < \alpha \leq 1$,并通过贝叶斯决策理论给出阈值的计算和解释.因此正域、负域和边界域的定义又可以得到如下表示:

$$\begin{aligned} \text{POS}(X) &= \{x \in U \mid P(X | [x]_A^\theta) \geq \alpha\}, \\ \text{NEG}(X) &= \{x \in U \mid P(X | [x]_A^\theta) \leq \beta\}, \\ \text{BND}(X) &= \{x \in U \mid \beta < P(X | [x]_A^\theta) < \alpha\}. \end{aligned}$$

3 基于GRA的多准则推荐

第1.3节阐述了医疗多准则推荐和传统多准则推荐在数据结构方面的差异,这些差异导致其运用传统多准则推荐的方法会有一定的局限性. MCDM是决策科学的重要理论方法之一,其能有效评价事物的多维信息并得到事物的综合评估.多准则推荐和MCDM本质都是对多维信息加工并作出优选.因此,本文从MCDM的角度来研究多准则推荐,在多准则推荐中如何获取项目的总体评分值是首要解决的问题.从MCDM的角度来说,就是寻找合适的聚合函数聚合不同准则的评分值,得到用户某个项目的一致性(总体)评分并作出推荐.近年来,决策信息的不确定性导致MCDM的研究重心转移到了不确定性决策,灰色理论就是主要研究方法之一.灰色系统主要研究系统因素、系统结构、系统作用原理不明确的系统.在医疗推荐中,指标内部之间的作用机理是不完全明确的,因此这种“有所知,有所不知”的医疗推荐系统是个典型的灰色系统.

GRA是灰色理论的主要研究内容之一^[14],它主要面向受多种因素影响的事物,通过运用灰色关联度系数测度指标序列的关联程度,从整体观念出发对事物进行综合评价.本文运用GRA方法计算不同用户关于项目 i 的灰色关联度并将其作为项目 i 的总体评分,其计算过程中参照了用户的数据,使得到的总体评分更加准确有效.在医疗推荐中,检查项目 i 中的

指标有最优的值序列,也就是指检查项目所包含指标的正常值序列,故可以利用GRA计算患者的实际数据序列与最优值序列的关联程度来判断其与最优值序列的接近程度,关联度越小,说明该项目 i 越偏离正常状态.

下面以项目 i_1 的评分矩阵(如表4所示)为例给出GRA方法计算的一般性步骤.设 $u_q(q = 1, 2, \dots, n)$ 为第 q 个用户,序列 $U_q = \{u_q(1), \dots, u_q(h), \dots, u_q(|k_1|)\}$ 则表示用户 u_k 对评价准则 $e_1, \dots, e_h, \dots, e_{|k_1|}$ 的评价序列.

表4 用户-项目 i_1 的评分矩阵表

	e_1	...	e_h	...	$e_{ k_1 }$	总体评分值
u_1	$u_1(1)$...	$u_1(h)$...	$u_1(k_1)$	$P(U_1, U_0)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
u_q	$u_q(1)$...	$u_q(h)$...	$u_q(k_1)$	$P(U_q, U_0)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
u_n	$u_n(1)$...	$u_n(h)$...	$u_n(k_1)$	$P(U_n, U_0)$

1) 确定分析序列. 根据项目 i_1 的评分矩阵可以用 $U_q = \{u_q(1), \dots, u_q(h), \dots, u_q(|k_1|)\}(q = 1, 2, \dots, n)$ 表示比较序列,而 $U_0 = \{x_{01}, \dots, x_{0h}, \dots, x_{0|k_1|}\}$ 则表示评价准则 $e_1, \dots, e_h, \dots, e_{|k_1|}$ 的参考序列.

2) 序列无量纲化. 由于不同准则量纲的差异性和不可公度性,计算过程中需要进行无量纲化操作.常见的序列无量纲化方法有初值化和均值化.

3) 计算关联系数

$$\varphi(U_0(h), U_q(h)) = \frac{\min_q \min_h \Delta + \xi \max_q \max_h \Delta}{\Delta + \xi \max_q \max_h \Delta}. \quad (9)$$

其中: $\varphi(U_0(h), U_q(h))$ 为序列 U_q 和参考序列 U_0 关于第 h 个准则的关联系数; $\xi(\xi \in [0, 1])$ 为分辨系数, ξ 一般小于等于0.5,主要是为了减少极值对计算的影响; $\Delta = d(U_0(h), U_q(h))$,其为序列 U_q 和参考序列 U_0 关于第 h 个准则的距离,其中距离的计算可参照式(4)~(7).

4) 计算关联度. 为了便于比较,将不同准则下的关联系数综合为灰色关联度($h = 1, 2, \dots, |k_1|$),即

$$P(U_0, U_q) = \frac{1}{|k_1|} \sum_1^{|k_1|} \varphi(U_q(h), U_0(h)). \quad (10)$$

通过GRA方法计算不同用户所有项目的总体评分值,此时基于异构信息的多准则推荐可转化为单评分推荐,即推荐评分表由表3转化为表1,此时可采用单评分推荐方法进行推荐预测.通过运用GRA方法对项目的不同准则聚合,可以突破传统多准则推荐方

法运用在差异多准则推荐上的局限性,同时具有更好的可解释性.

4 三支推荐

传统二支推荐包括推荐或者不推荐项目,实际中并不是总可以作出双向决策.尤其是在代价敏感的医疗推荐中,若将不需要检查的项目推荐给患者,则会给其带来更多的时间和金钱成本;而将患者亟待检查的项目未推荐给患者,则会导致潜在的疾病未被及时筛查出来,给患者造成不可估量的潜在危害.因此,“非此即彼”的二支推荐如果误判则会造成较大的决策风险.而三支决策^[15]在决策中考虑了决策成本,同时引入延迟决策降低了二支决策的风险.三支决策由于决策过程的优越性在推荐领域也得到很好的验证和应用^[16-18].三支决策基于最小风险贝叶斯决策和粗糙集理论,其核心思想是将整体划分成3个互不相交的区域,对不同的区域实行不同的决策行为.本文在基于异构信息的粗糙集模型基础上构建了三支推荐,即在推荐中根据预测评分和成本决定推荐、不推荐还是延迟推荐.

在推荐系统中, $P(R|(u, i))$ 表示用户 u 喜欢项目 i 的可能性.在二支推荐中,对于每一个 (u, i) 可将其划分为推荐或者不推荐,也就是正域 $POS(R)$ 或者负域 $NEG(R)$,则有

$$POS_{\xi}(R) = \{(u, i) \mid P(R \mid (u, i)) \geq \varphi\};$$

$$NEG_{\xi}(R) = \{(u, i) \mid P(R \mid (u, i)) < \varphi\}.$$

在三支推荐中,对于每一个 (u, i) 可以将其划分为正域 $POS(R)$ 、负域 $NEG(R)$ 和边界域 $BND(R)$. $BND(R)$ 表示不确定是否将项目推荐给用户.在三支推荐模型^[18]中假定集合 $A = \{P, N, B\}$,其表示3种决策行为,分别是 $(u, i) \in POS(R)$, $(u, i) \in NEG(R)$ 和 $(u, i) \in BND(R)$.表5描述了不同推荐行为下所付出的代价, $\lambda_{PP}, \lambda_{BP}, \lambda_{NP}$ 分别表示当用户 u 喜欢项目 i 时采取行动 P, N, B 所付出的代价, $\lambda_{PN}, \lambda_{BN}, \lambda_{NN}$ 分别表示当用户 u 不喜欢项目 i 时采取行动 P, N, B 所付出的代价.其中 $\lambda_{PP}, \lambda_{NN}$ 为推荐正确的成本; $\lambda_{BP}, \lambda_{BN}$ 表示推荐过程中的学习成本; $\lambda_{NP},$

λ_{PN} 则代表错误推荐的成本,显然错误推荐的成本要大于其他推荐成本,一般来说 $\lambda_{PP} < \lambda_{BP} < \lambda_{NP}, \lambda_{NN} < \lambda_{BN} < \lambda_{PN}$.

在已知决策代价后,可以得到如下不同决策行为的期望代价:

$$C(P \mid (u, i)) =$$

$$\lambda_{PP}P(R \mid (u, i)) + \lambda_{PN}P(\bar{R} \mid (u, i)),$$

$$C(B \mid (u, i)) =$$

$$\lambda_{BP}P(R \mid (u, i)) + \lambda_{BN}P(\bar{R} \mid (u, i)),$$

$$C(N \mid (u, i)) =$$

$$\lambda_{NP}P(R \mid (u, i)) + \lambda_{NN}P(\bar{R} \mid (u, i)).$$

根据贝叶斯最小代价原则,可以得到如下规则:

1) 如果 $C(P \mid (u, i)) < C(B \mid (u, i)), C(P \mid (u, i)) < C(N \mid (u, i))$, 则 $(u, i) \in POS(R)$;

2) 如果 $C(B \mid (u, i)) < C(P \mid (u, i)), C(B \mid (u, i)) < C(N \mid (u, i))$, 则 $(u, i) \in BND(R)$;

3) 如果 $C(N \mid (u, i)) < C(B \mid (u, i)), C(N \mid (u, i)) < C(P \mid (u, i))$, 则 $(u, i) \in NEG(R)$.

通过一系列推导,可以根据代价函数得到阈值的计算公式如下:

$$\alpha = \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \quad (11)$$

$$\beta = \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \quad (12)$$

根据阈值 α, β 可以得到最终决策规则,其中 $0 < \beta < \alpha < 1$,具体的推导内容详见文献[18].

$$POS_{(\alpha, \beta)}(R)|_u = \{i \mid P(R \mid (u, i)) \geq \alpha\};$$

$$BND_{(\alpha, \beta)}(R)|_u = \{i \mid \beta < P(R \mid (u, i)) < \alpha\};$$

$$NEG_{(\alpha, \beta)}(R)|_u = \{i \mid P(R \mid (u, i)) \leq \beta\}.$$

根据上述规则和预测评分 P' , 可以作出相应的决策行为.当 $P' \geq \alpha$ 时进行推荐;当 $P' \leq \beta$ 时不进行推荐;当 P' 介于 α 与 β 之间时,对其进行延迟决策,可以有效降低决策成本,提高推荐的准确性.

5 融合粗糙集与GRA的异构信息医疗多准则推荐

粗糙集能够在不需要先验知识的情况下对信息进行有效的处理,在多准则推荐中利用粗糙集理论能够量化测度患者之间不同数据类型信息的距离,但它无法有效确定患者在多个准则下的总体评估和排序,而GRA能够在不需要先验知识的条件下对信息不完备的不确定系统进行要素间量化分析.在异

表5 推荐代价矩阵表

决策行为	R	\bar{R}
推荐 P	λ_{PP}	λ_{PN}
延迟推荐 B	λ_{BP}	λ_{BN}
不推荐 N	λ_{NP}	λ_{NN}

构信息粗糙集基础上引入GRA方法既考虑到了医疗信息系统的灰色性,又能通过比较患者指标序列与最优指标序列的相似程度评估检查项目,更加符合医疗实际决策场景.因此,通过结合粗糙集和GRA方法可以解决医疗推荐过程中存在的数据多源异构和推荐项目多准则的问题,更能适应医疗决策中的不确定性特征.

5.1 问题描述

基于本文提出的医疗推荐背景,本节采用临床数据进行实验验证本文方法的有效性.本文所采用的数据包含2041位患者的痛风诊断数据.目前我国痛风患者已超8000万,是我国第2大慢性疾病.痛风在诊断过程中除了观察关节外在肿胀程度外还需要进行血脂、肾功等多项检查辅助确诊^[23],这些项目又细分为多项指标,具体的数据描述如表6所示,每个指标的正常范围可参考文献^[24].其中:症状属于语义值,性别属于符号值,其他指标属于实数值.由表6可以看出,痛风诊断所涉及的检查项目包含的指标具有差异性,而不同的指标其衡量准则亦有差异性.本实验通过融合粗糙集和GRA实现异构信息下医疗多准则三支推荐,为患者推荐个性化的关键检查项目以辅助医生诊断.

表6 数据集描述表

项目	指标	项目	指标
肝功-白蛋白	ALB	血脂	TC
	GLB		TG
	PA		HDL-C
	TP		LDL-C
肝功-胆红素	DBIL	肾功	UA
	TBIL		UREA
	IBIL		CR
自然属性	Gender	症状	肿胀程度
	AGE	血糖	GLU
肝功-酶类	AST	肝功-酶类	ALT
	ADA		GGT
	ALP		

首先,取所有指标最优范围的上限值和下限值的平均值作为其最优值序列;接下来,基于GRA方法对每个项目下的不同指标进行聚合,并将其作为该项目的总体评分值.其实际意义是指:患者的指标序列与最优序列的关联度越小,说明该项目偏离正常状态的程度越大.因此,对于总体评分较小的项目会优先推荐,这是与传统推荐问题存在的差异之一.在

医疗推荐场景中, $\lambda_{PP}, \lambda_{NN}$ 为推荐结果正确的代价; $\lambda_{BP}, \lambda_{BN}$ 分别表示对于患者需要或者不需要的检查项目延迟推荐的代价; $\lambda_{NP}, \lambda_{PN}$ 则表示推荐结果错误的代价.本文结合实际推荐场景和专家意见得到推荐代价,其中 $\lambda_{PP} = \lambda_{NN} = 0, \lambda_{BP} = 3.6, \lambda_{NP} = 9.0, \lambda_{PN} = 9.0, \lambda_{BN} = 3.6$.根据式(11)和(12)计算出决策阈值 $\alpha = 0.6, \beta = 0.4$,然后利用协同过滤算法计算用户未检查的项目总体评分预测值,并结合上述求出的阈值作出推荐.

5.2 结果分析

在运用GRA方法计算总体评分过程中,本文还从指标的角度计算出每个项目下不同指标之间的灰色关联度,部分结果如图2所示,从中可以挖掘出项目内部指标的关联性.例如在酶类项目下,可以得到指标ALT与AST之间具有极强的关联性,而在临床实践中ALT和AST都是转氨酶,且常用AST/ALT的比值来反应肝细胞的损伤情况;同样,白蛋白中ALB和GLB也高度关联,临床中也常用ALB/GLB的比值反映肝功能代谢情况,进一步验证了该方法的有效性,并从数据挖掘的角度为临床实践提供了数据解释.

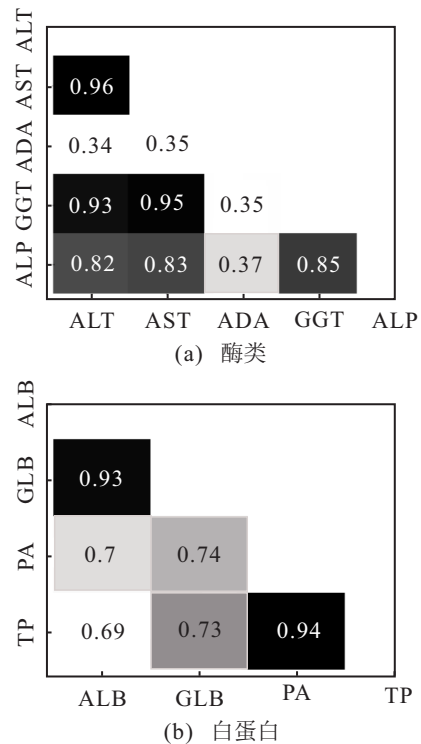


图2 不同项目中各指标之间的灰色关联度

实验选用平均绝对误差(MAE)、准确率(ACC)、均方误差(MSE)、精确率(PRE)和召回率(REC)来评估推荐效率^[1].图3为不同近邻用户数N所对应的指标评估效果折线图,可以看出随着N的增加,ACC

和PRE值逐渐增大,在 N 达到30后趋于稳定.实际推荐过程中,在一定范围内参考的 N 值越大,其对未评分项目的预测越准确.图4给出了阈值 α 和 β 在不同组合下所对应的ACC和PRE,可以看出阈值 α 和 β

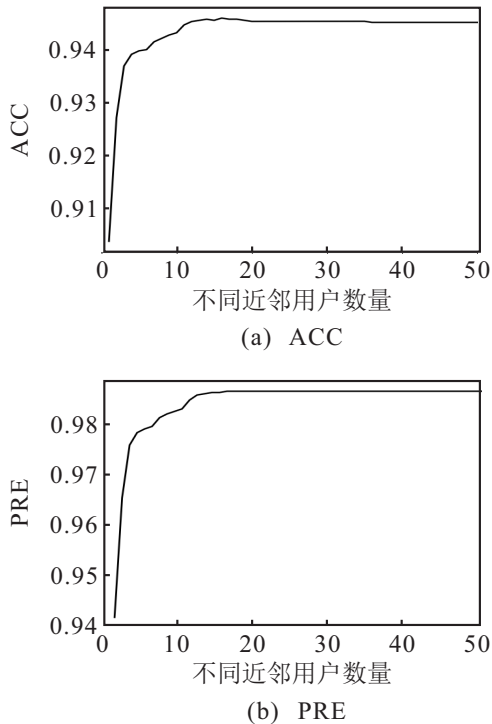


图3 不同近邻用户数量对应的评估效果折线图

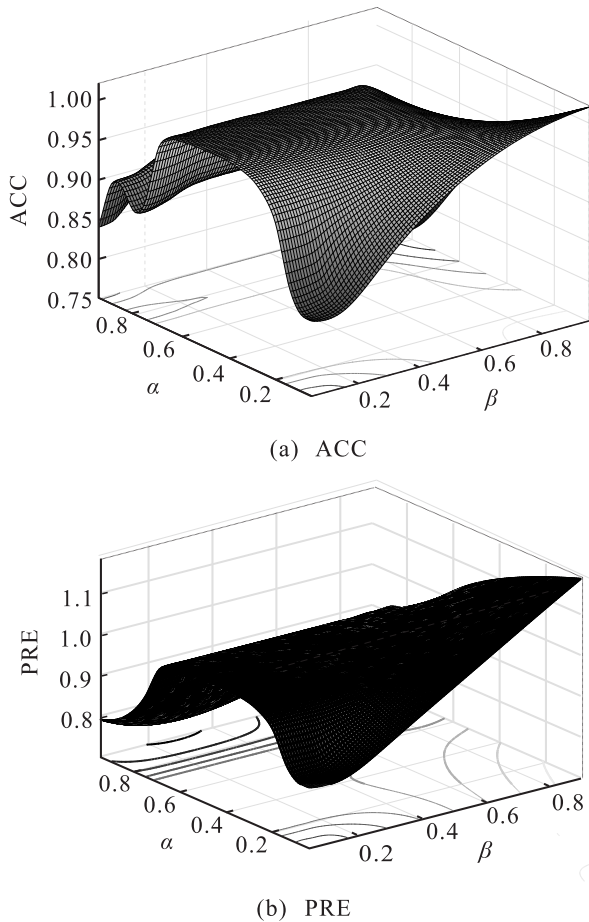


图4 不同阈值组合对应的评估效果

在0.4~0.7之间的交叉组合所对应的ACC和PRE较高.因此,在考虑决策成本的基础上,可选择合适的阈值实现更加精准的推荐.

第1.2节阐述了多准则推荐的研究思路,其中基于聚合函数的研究方法主要通过运用统计学、机器学习等^[9-11]方法学习多准则与总体评分的关系,因此本文采用多种方法计算项目的总体评分,例如 K 近邻回归(KNN)、BP神经网络(BPNN)、支持向量机(SVM)、逻辑回归(LR)、余弦相似性(Cos)和皮尔逊相关系数(Pearson),并将这些方法的推荐结果与GRA方法的推荐结果进行对比,如表7所示.通过观察表7可以发现,采用GRA方法计算项目的总体评分得到的推荐结果优于其他方法,进而验证了采用GRA方法融合项目多准则的有效性.

表7 不同多准则推荐方法的推荐结果对比

指标	GRA	KNN	BPNN	SVM	LR	Cos	Pearson
MAE	0.0630	0.0932	0.1053	0.1354	0.0941	0.0661	0.2639
MSE	0.0077	0.0185	0.0191	0.0211	0.0178	0.0365	0.2048

深度学习近年来得到了众多专家的广泛关注和研究,其具有很强的适应性,因此被广泛应用在推荐领域^[19-21].本实验增加了经典的深度学习推荐方法(如DeepFM和Wide&Deep等)在MAE、MSE、ACC、PRE、REC和Time等多个维度方面的比较,如表8所示.

表8 不同推荐算法的推荐结果对比

算法	MAE	MSE	ACC	PRE	REC	Time/s
本文方法	0.0630	0.0077	0.9742	0.9837	0.8770	35.7067
DeepFM	0.0846	0.1285	0.8745	0.9301	0.6978	39.8569
Wide&Deep	0.0929	0.1312	0.8923	0.9908	0.6926	31.6783

通过观察表8可以看出,本文提出的方法在MAE、MSE、ACC、PRE和REC等指标的表现结果都比较好,具有比较高的精度和准确度,能够满足医疗诊断对推荐准确度的高要求.从计算时间上看,本文提出的方法优于DeepFM,略逊色于Wide&Deep方法,但整体差异不是很显著.尽管准确性和时效性对于医疗诊断来说都是很重要的衡量指标,但考虑到医疗诊断领域的特殊性,低准确度带来的错误代价高于低时效性带来的代价,因此优先考虑诊断结果的准确性.此外,医疗领域需要高度的可解释性,即辅助诊断系统提供的建议一定符合医理,这需要解释算法以及证明其可靠性.而深度学习模型具有多个参数,算法

结果的优劣很大程度上取决于参数的选择,并且其具有“黑匣子”性质,这导致深度学习运用在医疗诊断领域的可解释性较弱.本文采用的基于灰色关联和三支协同过滤的推荐方法在计算过程中语义明确,具有较好的可解释性和推荐结果.

根据上述不同参数推荐结果的表现,选取近邻用户数 $N = 50$, 阈值 $\alpha = 0.6, \beta = 0.4$, 预测空缺值评分并进行二支推荐和三支推荐,得到推荐结果如表9所示,其中括号内表示的是推荐中的成本.由表9中可以计算得到二支推荐和三支推荐的决策总成本分别为1107和1065.6,平均推荐成本分别约为0.23和0.22.通过表9可以看出,三支推荐中增加了延迟推荐,非常符合医疗推荐中信息不充分时的决策场景,即不作出明确的推荐还是不推荐,而待症状等信息充分时再进行相应的决策,既可以提高推荐的准确性,又可以避免果断决策带来的高代价.综上可得:该方法在推荐成本和推荐准确性上都要优于二支推荐,具有更加适用于医疗推荐场景中代价敏感以及对准确度要求高等特点.

表9 二支推荐和三支推荐结果对比

决策行为	三支推荐		二支推荐	
	需要	不需要	需要	不需要
推荐 P	536(0)	8(9.0)	784(0)	13(9.0)
延迟推荐 B	237(3.6)	4(3.6)	—	—
不推荐 N	14(9.0)	3974(0)	110(9.0)	3866(0)

6 结论

相较于传统推荐场景,医疗推荐中项目数量的有限性决定了医疗推荐数据的稀疏程度较小,因此能够更精准地衡量患者之间的相似性;同时,医疗问诊过程中存在的显式反馈,能较好地解决冷启动等问题,因此将推荐方法运用在医疗推荐领域具有巨大的潜能和优势.本文充分结合推荐技术、MCDM和三支决策等方法,提出了融合粗糙集和GRA的多准则三支推荐方法,提高了该推荐方法在数据异构和项目多准则背景下的适应性.考虑到推荐过程的决策代价,降低了决策成本.最后将该方法运用于医疗推荐,向患者推荐个性化的检查项目以辅助医生诊疗,并从数据挖掘角度丰富了临床试验中经验科学的方法.

参考文献(References)

[1] 王国霞,刘贺平.个性化推荐系统综述[J].计算机工程与应用,2012,48(7):66-76.
(Wang G X, Liu H P. Survey of personalized

recommendation system[J]. Computer Engineering and Applications, 2012, 48(7): 66-76.)

[2] 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(Z1):153-164.
(Ye J X, Xiong H X, Jiang W X. A physician recommendation algorithm integrating inquiries and decisions of patients[J]. Data Analysis and Knowledge Discovery, 2020, 4(Z1): 153-164.)

[3] 翟姗姗,胡畔,潘英增,等.融合知识图谱与用户病情画像的在线医疗社区场景化信息推荐研究[J].情报科学,2021,39(5):97-105.
(Zhai S S, Hu P, Pan Y Z, et al. Scenario-based information recommendation of online medical community based on knowledge graph and disease portrait[J]. Information Science, 2021, 39(5): 97-105.)

[4] 侯梦薇,卫荣,樊林,等.推荐系统在医疗领域的处理模型及应用研究[J].中国数字医学,2019,14(1):80-82.
(Hou M W, Wei R, Fan L, et al. Research on processing model and applications of recommendation system in medical field[J]. China Digital Medicine, 2019, 14(1): 80-82.)

[5] Pawlak Z. Rough sets[J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341-356.

[6] Li Z W, Zhang P F, Xie N X, et al. A novel three-way decision method in a hybrid information system with images and its application in medical diagnosis[J]. Engineering Applications of Artificial Intelligence, 2020, 92: 103651.

[7] 黄恒秋,曾玲,黎利辉.混合值不完备系统的双邻域粗糙集分类方法[J].控制与决策,2018,33(7):1207-1214.
(Huang H Q, Zeng L, Li L H. Double-neighborhood rough set classification method in incomplete decision system with hybrid value[J]. Control and Decision, 2018, 33(7): 1207-1214.)

[8] 杨臻,邱保志.混合信息系统的动态变精度粗糙集模型[J].控制与决策,2020,35(2):297-308.
(Yang Z, Qiu B Z. Dynamic variable precision rough set model of mixed information system[J]. Control and Decision, 2020, 35(2): 297-308.)

[9] Gupta S, Kant V. Credibility score based multi-criteria recommender system[J]. Knowledge-Based Systems, 2020, 196: 105756.

[10] Nilashi M, Jannach D, Ibrahim O B, et al. Clustering- and regression-based multi-criteria collaborative filtering with incremental updates[J]. Information Sciences, 2015,

- 293: 235-250.
- [11] Nassar N, Jafar A, Rahhal Y. A novel deep multi-criteria collaborative filtering model for recommendation system[J]. Knowledge-Based Systems, 2020, 187: 104811.
- [12] Adomavicius G, Kwon Y. New recommendation techniques for multicriteria rating systems[J]. IEEE Intelligent Systems, 2007, 22(3): 48-55.
- [13] Manouselis N, Costopoulou C. Analysis and classification of multi-criteria recommender systems[J]. World Wide Web, 2007, 10(4): 415-441.
- [14] 贺利军, 李文锋, 张煜. 基于灰色综合关联分析的多目标优化方法[J]. 控制与决策, 2020, 35(5): 1134-1142. (He L J, Li W F, Zhang Y. Multi-objective optimization method based on grey synthetic incidence analysis[J]. Control and Decision, 2020, 35(5): 1134-1142.)
- [15] Yao Y Y. Three-way decisions with probabilistic rough sets[J]. Information Sciences, 2010, 180(3): 341-353.
- [16] Huang J J, Wang J, Yao Y Y, et al. Cost-sensitive three-way recommendations by learning pair-wise preferences[J]. International Journal of Approximate Reasoning, 2017, 86: 28-40.
- [17] Liu D, Ye X Q. A matrix factorization based dynamic granularity recommendation with three-way decisions[J]. Knowledge-Based Systems, 2020, 191: 105243.
- [18] Ye X Q, Liu D. An interpretable sequential three-way recommendation based on collaborative topic regression[J]. Expert Systems with Applications, 2021, 168: 114454.
- [19] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems[C]. Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. Boston, 2016: 7-10.
- [20] Guo H F, Tang R M, Ye Y M, et al. DeepFM: A factorization-machine based neural network for CTR prediction[C]. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, 2017: 1725-1731.
- [21] Hu Z B, Wang J C, Yan Y, et al. Neural graph personalized ranking for top- N recommendation[J]. Knowledge-Based Systems, 2021, 213: 106426.
- [22] Wu W Z, Leung Y. Theory and applications of granular labelled partitions in multi-scale decision tables[J]. Information Sciences, 2011, 181(18): 3878-3897.
- [23] Chu X L, Sun B Z, Li X, et al. Neighborhood rough set-based three-way clustering considering attribute correlations: An approach to classification of potential gout groups[J]. Information Sciences, 2020, 535: 28-41.
- [24] 张萌, 孙秉珍, 楚晓丽. 基于邻域代价敏感三支决策的痛风诊断模型[J]. 计算机工程与应用, 2020, 56(16): 218-225. (Zhang M, Sun B Z, Chu X L. Gout diagnosis model based on neighborhood cost sensitive three-way decision[J]. Computer Engineering and Applications, 2020, 56(16): 218-225.)

作者简介

张萌(1996—),女,硕士生,从事数据挖掘、智能决策的研究, E-mail: 982417242@qq.com;

孙秉珍(1979—),男,教授,博士生导师,从事粒计算与不确定性决策、应急管理决策等研究, E-mail: bzsun@xidian.edu.cn;

王婷(1986—),女,博士生,从事数据科学、预测决策的研究, E-mail: wangting1986@stu.xidian.edu.cn;

楚晓丽(1985—),女,博士生,从事中医药数据挖掘、智能决策的研究, E-mail: 1053305721@qq.com;

同思蓉(1996—),女,硕士生,从事粒计算、不确定决策的研究, E-mail: sr65798007@163.com.

(责任编辑:孙艺红)