

基于云模型的指挥信息多重协同过滤算法研究

杜 波 俞 岩 戴 刚
(武警工程大学 乌鲁木齐 830049)

摘 要 针对传统协同过滤算法在构建指挥员与指挥要素间的协同过滤关系时面临着数据稀疏、冷启动等问题,提出了一种面向反恐任务的指挥信息多重协同过滤算法。该算法首先通过作战类型对指挥要素进行基于云模型的预协同过滤,然后将凝聚子集分析融入基于用户的协同过滤中,挖掘特定战斗类型下的指挥员与指挥要素间的相似性关系,进而实现精准推荐。实验表明,该算法贴合面向作战任务的指挥信息系统应用实践,有效提高了系统的推荐效率和准确度。

关键词 指挥系统,协同过滤,云模型,凝聚子集
中图法分类号 TP391.3 **文献标识码** A

Study on Multi-collaborative Filtering Algorithm of Command Information Based on Cloud Models

DU Bo YU Yan DAI Gang
(Engineering University of PAP, Urumqi 830049, China)

Abstract To overcome the problems of data sparse and cold start based on traditional collaborative filtering algorithm, when collaborative filtering relationship is established between commanders and command elements, a mission-oriented multi-collaborative filtering algorithm of command information was proposed. Firstly, the algorithm performs cloud model-based pro-collaborative filtering by operation type on the command elements, and then integrates the cohesion subset analysis into user-based collaborative filtering to mine the similarities between the commanders and command elements under specific operation types, thus to achieve accurate recommendations. The experimental results show that the proposed algorithm can be applied to the command information system of operational mission effectively and improves the recommendation efficiency and accuracy of the system.

Keywords Command system, Collaborative filtering, Cloud models, Cohesion subset

1 引言

近年来,包括伊斯兰国(ISIS)、基地组织在内的极端恐怖势力不断在世界范围内掀起腥风血雨。美国2011年发布的《反恐报告》称,恐怖主义在未来10年至20年仍将存在,国际反恐形势依然严峻^[1]。反恐指控平台是各指挥节点实施反恐行动的数据交互枢纽,随着数据量的激增,信息超载问题愈演愈烈,大量的冗余信息导致指挥员无法及时获取准确、有效的指挥要素信息,从而影响行动决策效率。推荐系统能够通过捕捉指挥员的偏好特征,从海量信息中高效、准确地挑选满足指挥员实际作战需求的指挥信息,从而缓解“信息海洋”查找问题,提高指挥效能。

推荐系统(Recommendation System)是为了解决由于网络数据过载引起的用户获取所需信息难、商家寻找潜在客户的问题而被提出的^[2]。推荐的本质是一种信息过滤机制,

通过该机制过滤掉用户不需要的冗余信息,将用户感兴趣的“商品”进行主动推送,进而减少用户信息搜索过程中的附加成本。作为推荐系统的主要算法,协同过滤的基本原理是根据“邻居用户”的兴趣来预测“目标用户”的“兴趣”,根据用户间的相似性特征,将相应的“商品”推荐给“目标用户”。目前常用的协同过滤推荐技术主要包括:基于用户的协同过滤推荐^[3]、基于项目的协同过滤推荐^[4]、基于模型的协同过滤推荐^[5]、基于时间加权的协同过滤推荐^[6]以及混合协同过滤推荐^[7]等。这些推送技术应用领域不同,但从推荐效果上看,学术界认为协同过滤推荐普遍存在数据稀疏性、冷启动、实时性等问题,直接影响了数据推荐的效率和准确性。许多国内外学者就此展开研究,提出了一些改进和优化策略^[8-12]。这些方法从相似性判定、初始数据的降维处理、改进评分方法等方面进行改进^[13-15],从而提高了推荐精度。

反恐指挥信息系统作为特殊的“推荐系统”,具有数据类

本文受国家自然科学基金(U1603261)、新疆自治区自然科学基金(2016D01A080)资助。

杜 波(1984—),男,讲师,主要研究方向为反恐数据挖掘与分析;俞 岩(1976—),女,副教授,主要研究方向为指挥信息系统开发;戴 刚(1974—),男,副教授,主要研究方向为反恐数据挖掘与分析。

型多样、信息编组复杂、实时性和精确性要求高等特点。因此,针对反恐行动多任务、多信息要素、多指挥层级、多地形样式等特点开展协同过滤算法研究,对解决上述问题、指导指挥信息系统开发的工程实践具有一定的理论和实践价值。

本文所述基于信息系统的反恐怖行动包含 3 个核心要素,即指挥员、指挥要素信息和战斗类型(即反恐任务类型)。历次反恐行动实践表明,不同战斗类型所对应的指挥要素信息有所不同;同一战斗类型中,不同层级指挥员所需的指挥信息也是千差万别。随着反恐怖行动的日趋多样化,战斗所需的要素信息激增,直接构建指挥员与指挥要素信息间的协同过滤关系面临着数据稀疏和冷启动等困境,不仅效率低下,准确性也难以保证。为此,本文提出一种指挥信息多重协同过滤算法(Multi-Collaborative Filtering Algorithm of Command

Information, MC-CF)。考虑到专家评分中存在的模糊性和随机性问题,首先通过基于云模型的预协同过滤(Pro-Collaborative Filtering Based on Cloud Model, P-CF-CM)挖掘战斗类型与指挥信息要素的协同关系;然后将社会网络分析(Social Network Analysis, SNA)的凝聚子群分析方法融入传统协同过滤算法中,用于改进相似性计算方法,有效解决了数据稀疏和冷启动问题,为实现反恐行动指挥要素信息的精准、实时推送及提高基于信息系统的反恐作战能力奠定理论基础。

2 指挥信息多重协同过滤算法

2.1 算法的基本原理

指挥信息多重协同过滤算法的流程如图 1 所示。

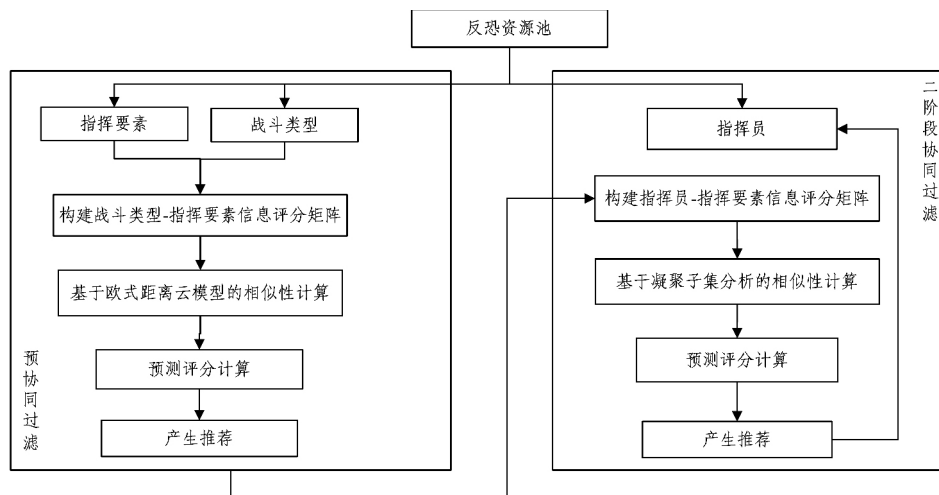


图 1 反恐指挥信息多重协同过滤算法流程

算法流程分为预协同过滤和二阶段协同过滤两部分。其基本原理是,首先从反恐资源池中提取战斗类型和指挥要素信息,并进行预协同过滤,目的是在不预设战斗类型的前提下登录资源池,然后系统自动挖掘出不同战斗类型当前最急需的 Top-N 个指挥要素信息。鉴于专家评分的随机性和模糊性特征,在该阶段采用基于欧氏距离云模型的相似性计算方法。然后,当指挥员(或上级指挥机关)发出“指令”时,根据指挥员基本特征信息(指挥层级-Hie、定位信息-Area、请求时间-Time 等)及数据“指令”信息(战斗类型-Task),构建特定战斗类型下的“指挥员-指挥信息要素”评分矩阵,并结合凝聚子集分析方法给出指挥员相似性评判,根据综合关系矩阵得出更为准确的“邻居集”。最后进行预测评分,为该指挥员精确推送所需的战斗指挥要素信息。为了得到较为精准的推送结果,其核心在于评分矩阵的构建与相似性的计算。

2.2 基于云模型的预协同过滤相似性分析

2.2.1 云模型的概念与数字特征

云模型是自然语言表示的某个定性概念与定量概念间的双向认知模型,其定义如下^[16-17]。

定义 1 设 C 是论域 U 上的概念,若 $x \in U$ 是 C 的一次随机实现, x 对 C 的隶属度 $\mu(x) \in [0, 1]$ 是有稳定倾向的随

机数,则 x 在论域 U 上的分布称为云,每一个 x 及其隶属度 $\mu(x)$ 称为一个云滴。

由定义可知,云滴是定性概念与论域空间定量描述的随机映射,云模型是由众多云滴构成的,反映了定性概念的随机性特征,其示意图如图 2 所示。云模型可用期望 Ex (Expected Value)、熵 En (Entropy)、超熵 He (Hyper Entropy) 3 个数字特征来表征。其中, Ex 表示云滴在论域空间分布的期望值,是表征定性概念的最典型样本,距离 Ex 越近,云滴越集中; En 反映云滴的离散程度,是定性概念的不确定性度量,熵越大,云滴的离散程度越强,相应概念的不确定性越强; He 即熵的熵,是 En 的不确定性度量,由 En 的随机性和模糊性共同决定。

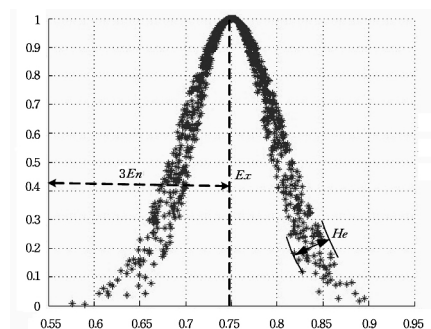


图 2 云模型示意图

2.2.2 逆向云发生器

云发生器是生成云滴的算法的统称。如前所述,云模型是定性概念与其定量值间的双向认知模型,与正向云发生器相反,逆向云发生器(Backward Cloud Generator)能够从服从同一类分布特征的样本(即云滴)中抽象出以期望值、熵和超熵为特征的定性概念。其逻辑结构如图3所示。



图3 逆向云发生器的逻辑结构

正态逆向云发生器算法流程如下。

Input: N 个 m 维云滴 $x_i = (x_{i1}, x_{i2}, \dots, x_{im}), i \in [1, n]$

Output: 定性概念的云模型特征值 $C(Ex, En, He)$

Step1: 计算样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^N x_i$, 一阶样本绝对中心距为 $\frac{1}{n} \sum_{i=1}^N |\bar{X} - x_i|$, 样本方差为 $S^2 = \frac{1}{N-1} \sum_{i=1}^N (\bar{X} - x_i)^2$;

Step2: 计算期望值 $Ex = \bar{X}$;

Step3: 计算熵值 $En = \hat{En} = \sqrt{\frac{\pi}{2}} \times \frac{1}{N} \sum_{i=1}^N |x_i - Ex|$;

Step4: 计算超熵 $He = \hat{He} = \sqrt{|S^2 - En^2|}$ 。

其中 $\hat{Ex}, \hat{En}, \hat{He}$ 分别表示 Ex, En, He 的参数估计值。

2.2.3 基于云模型的预协同过滤的基本原理

基于云模型的预协同过滤的主要思想是:将云模型中的每个“云滴”视为一种战斗类型对各指挥要素的单独评分,每一朵“云”则代表该战斗类型下所有指挥要素的评分集合。由于指挥要素数量已知,即云滴数已知,根据逆向云发生器可得云模型的数字特征值 Ex, En, He ,进而得到该战斗类型下的评分集合的云模型 $C(Ex, En, He)$ 。其中, Ex 表示该作战类型对指挥要素的平均需求程度, En 表示需求的集中程度, He 表示需求的稳定度。以此为依据,然后采用基于欧氏距离云模型的相似度计算方法,即可得出该战斗类型的云间相似度和最近邻居集,从而为分析构建相似性矩阵提供依据。

2.2.4 基于欧氏距离云模型的相似性计算

通过分析发现,在云模型的数字特征中时常出现期望值 Ex 远大于熵 En 和超熵 He 的情况,这时使用余弦相似性计算方法就往往容易忽视 En 和 He 的贡献,进而造成相似性度量出现偏差,甚至与事实相违背,严重影响推荐精度^[18]。为此,本文采用基于欧氏距离的云模型相似性计算方法,即将 $C(Ex, En, He)$ 的 Ex, En, He 3 个特征值映射到欧氏空间中,分别对应三维坐标系中的 x, y, z 轴坐标,则空间中两点 $S(Ex_s, En_s, He_s)$ 和 $K(Ex_k, En_k, He_k)$ 的欧氏距离 $dis(SK)$ 定义为:

$$dis(SK) = \sqrt{(Ex_s - Ex_k)^2 + (En_s - En_k)^2 + (He_s - He_k)^2} \quad (1)$$

由式(1)可知,作战类型云模型 $C(Ex, En, He)$ 特征向量间的欧氏距离与其相似度成反比,即距离越小,两种作战类型的相似度越大;反之,相似度越小。为了便于相似性度量和计算,引入指数函数进行标准化处理,则基于欧氏距离云模型的

相似性计算公式定义为:

$$Sim(s, k) = f(dis(SK)) = \exp(-dis(SK)) = \frac{1}{e^{\sqrt{(Ex_s - Ex_k)^2 + (En_s - En_k)^2 + (He_s - He_k)^2}}} \quad (2)$$

其中, $Sim(s, k)$ 表示 s, k 两种战斗类型的相似性。

2.3 基于凝聚子集分析的二阶段协同过滤相似性分析

社会网络分析(Social Network Analysis, SNA)方法是一种结构化分析方法,它通过提取社会成员中的关系模式来挖掘社会网络的内在结构^[19-20]。凝聚子群分析是社会网络分析的重要研究方向。反恐怖行动的主体是指挥员,不同指挥员间指挥关系明确,指挥流程清晰,呈现出明显的层级化结构,但随着战斗类型的日趋多样,指挥关系扁平化、网络化和模块组合化的结构趋势日趋明显。本文利用凝聚子群对指挥员进行分类,目的在于构建基于用户相似关系的凝聚子群。

2.3.1 相似关系的确定

本文应用凝聚子集分析方法进行相似性计算的依据在于,处于相同或临近指挥层级的指挥员,对于同一类型的反恐怖行动,其所需的指挥要素信息具有较高的相似性。为此,对指挥员(Commander)间的相似关系做如下定义:

$$Csim = \kappa Hie(u, v) + \delta Area(u, v) + \xi Time(u, v) \quad (3)$$

其中, u, v 分别表示指挥员 u 和 v ; 变量 κ, δ, ξ 为权重系数,满足 $0 \leq \kappa, \xi, \delta < 1, \kappa + \xi + \delta = 1$, 分别影响指挥层级 $Hie(u, v)$ 、指挥的地域 $Area(u, v)$ 、战斗时间 $Time(u, v)$ 在相似性计算中的比重,实际计算中可根据战斗类型确定,如山地捕歼战斗指挥的地域 $Area(u, v)$ 比重应当较大,反劫持战斗中更应强调战斗时间 $Time(u, v)$ 的重要性。本文为了便于计算,取 $\kappa = \xi = \delta = 1/3$ 。

相关约束条件定义如下:

(1)在反恐行动中,我们将指挥层级 $Hie(u, v)$ 定为 7 级,分别对应总部、总队(军)、支队(旅、团)、大队(营)、中队(连)、排、班,不同指挥层级的指挥关系用式(4)表征,用 (0.1, 0.2, 0.3, 0.4, 0.6, 0.8, 1) 表征 7 级隶属度。

$$Hie(u, v) = \begin{cases} 1.0, & \text{一级隶属} \\ 0.8, & \text{二级隶属} \\ 0.6, & \text{三级隶属} \\ 0.4, & \text{四级隶属} \\ 0.3, & \text{五级隶属} \\ 0.2, & \text{六级隶属} \\ 0.1, & \text{七级隶属} \end{cases} \quad (4)$$

(2)指挥地域 $Area(u, v)$ 表示指挥员 u, v 所处的战斗地域关系,1 表示处于同一战斗地域,0 表示处于不同战斗地域。 $Area_u$ 表示指挥员 u 所处的战斗地域, $Area_v$ 表示指挥员 v 所处的战斗地域。

$$Area(u, v) = \begin{cases} 0, & (Area_u \cap Area_v) \notin \text{同一地域} \\ 1, & (Area_u \cap Area_v) \in \text{同一地域} \end{cases} \quad (5)$$

(3)战斗时间 $Time(u, v)$ 表示指挥员 u, v 发出请求的战斗时间关系,1 表示发出请求时处于同一战斗时段,0 表示处

于不同战斗时段。 $Time_u$ 表示指挥员 u 的请求时段, $Time_v$ 表示指挥员 v 的请求时段。

$$Time(u, v) = \begin{cases} 0, & (Time_u \cap Time_v) \notin \text{同一时段} \\ 1, & (Time_u \cap Time_v) \in \text{同一时段} \end{cases} \quad (6)$$

2.3.2 相似性矩阵构建

通过式(3)可以构建起指挥员的相似性矩阵 UU_A 。在图中,节点间的可达性一般采用矩阵乘法表示。由指挥员间的相似关系可知,目标指挥员与其他指挥员间的相似性有多条可达路径,即存在相似性传递关系,包括直接相似性与间接相似性,即在反恐行动中存在层级指挥与越级指挥关系,如图4所示。

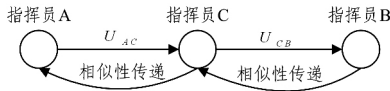


图4 相似性关系传递

在基于凝聚子群分析的相似性计算中,指挥员间的相似性关系传递可以用相似性矩阵乘法进行度量,即 $UU_A \times UU_A$ 表示二步可达间接相似性关系。基于 SNA 的 2-派系间接关系研究理论,本文只考虑指挥员间的二步可达间接相似性关系,二步可达间接相似性在本文中表示越级指挥关系。将矩阵运算后的二步可达间接相似性矩阵记作 UU' ,其中各元素 UU'_{ij} 表示指挥员 i 与指挥员 j 的二步可达相似性。因此,指挥员间的相似性应为直接相似性与二步可达相似性之和,记为 UUA 。为了便于计算,需要对相似性矩阵去量纲化,即:

$$UUA'_{ij} = \frac{UUA_{ij} - \overline{UUA}}{S_j} \quad (7)$$

其中:

$$\overline{UUA} = \frac{1}{n} \sum_{i=1}^n UUA_{ij}, S_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (UUA_{ij} - \overline{UUA})^2}$$

UUA'_{ij} 表示经过标准差变换后的相似性,其均值为 0,标准差为 1。为了确保 UUA'_{ij} 在区间 $[0, 1]$ 内波动,还需进行标准差变换,公式如下:

$$UR_{ij} = \frac{UUA'_{ij} - \min_{1 \leq i \leq m} (UUA')}{\max_{1 \leq i \leq m} (UUA') - \min_{1 \leq i \leq m} (UUA')} \quad (8)$$

其中, UR_{ij} 表示指挥员 i 与指挥员 j 的最终相似性, UR 为最终相似性矩阵。

2.3.3 相似性计算

相似性计算在协同过滤算法中至关重要,它决定了推荐结果的准确性,通常用余弦相似性、Pearson 相关系数和修正余弦相似性来表征。考虑到余弦相似性对评分尺度的一致性考虑存在缺陷,以及 Pearson 相关系数需要考虑样本间的相关系数等因素,本文在二阶段协同过滤中采用改进的修正余弦,如式(9)所示。

$$Sim(u_i, u_j) = UR_{ij} \times \frac{\sum_{c \in U_{ij}} (R_{i,c} - \overline{R_i})(R_{j,c} - \overline{R_j})}{\sqrt{\sum_{c \in U_i} (R_{i,c} - \overline{R_i})^2} \sqrt{\sum_{c \in U_j} (R_{j,c} - \overline{R_j})^2}} \quad (9)$$

$$R_{i,c} = \sqrt{\frac{\sum_k x_{i,c}^k}{\sum_k x_{i,c}^k}} \cdot y_{i,c}^{k_2} \quad (10)$$

其中, $R_{i,c}$ 和 $R_{j,c}$ 分别表示指挥员 i 与指挥员 j 对于指挥要素信息 C 的评分,计算方法如式(10)所示。 $x_{i,c}$ 表示该类指挥要素的历史请求次数。 $y_{i,c}$ 为专家打分值。 ϵ_1, ϵ_2 为权重值,为了便于计算本文取 $\epsilon_1 = \epsilon_2 = \frac{1}{2}$ 。 UR_{ij} 表示指挥员 i 与指挥员

j 的最终相似性。 $\overline{R_i}, \overline{R_j}$ 表示指挥员 i 与指挥员 j 对指挥要素信息的评分均值。

2.4 预测评分计算

确定指挥员(战斗类型)邻居集的方法有两种,在预协同过滤阶段,本文按照基于欧氏空间云模型的相似性大小进行排序,并选取前 N_1 个任务类型作为目标任务的邻居任务;在二阶段协同过滤中,给定阈值 β ,当指挥员间的相似性大于 β 时,认为该指挥员为目标指挥员的邻居,根据相似性计算式(7)可知 $\beta \in (0, 1)$ 。由此可得到两个阶段的邻居集 N_1 和 N_2 。其预测评分方法如式(11)所示:

$$P_{u,c} = \overline{R_u} + \frac{\sum_{i=1}^n [sim(u, i)(R_{i,c} - \overline{R_i})]}{\sum_{i=1}^n sim(u, i)} \quad (11)$$

通过式(11)可以给出指挥员对未使用指挥要素信息的评分,进而将评分靠前的 M 个指挥要素信息推送给该指挥员。

2.5 误差分析

信息推送质量需要定量的评判标准,本文以平均绝对偏差(Mean Absolute Error, MAE)作为度量标准。平均绝对偏差 MAE 通过计算预测评分与实际评分之间的偏差来度量预测的准确性,MAE 越小,表示推荐质量越高,其计算公式如式(12)所示:

$$MAE = \frac{\sum_{i=1}^n |p_i - q_i|}{N} \quad (12)$$

通过预测得到的评分集合表示为 $\{p_1, p_2, \dots, p_n\}$, 实际评分集合为 $\{q_1, q_2, \dots, q_n\}$ 。

3 仿真实验

3.1 基本想定

运用本文所述方法进行仿真实验。基本想定为:两名恐怖分子在 W 市持枪劫持 3 名人质,驾车企图冲撞 ZS 路临时检查点,造成市区部分路段瘫痪。该伙恐怖分子在受到交警围堵后逃窜至纺织厂大桥处,弃车进入附近厂房,企图借助地形优势与警方对峙。联指命令武警 W 市支队迅速向任务地区开进,并完成封控部署、现场处置和人质解救任务。

支队根据情况通报,拟定由 1,3 中队各派出 20 名官兵处置此次任务,由副支队长担任总指挥。到达任务地域后,副支队长 C、1 中队长 E 和 3 中队长 F 分别通过指挥终端向指挥中心请求任务相关数据,以便进一步拟定详尽的作战计划。

3.2 数据准备

本文所用相关数据来源于某数据工程建立的反恐数据资源池。在反恐数据资源池中记录了劫持人质、个体袭击、暴(骚)乱等 15 个反恐战斗样式,恐怖分子数量、装备、社情、天候、地形等 120 个指挥要素,2300 个指挥员对各指挥要素的请求 12000 次,以及专家根据请求次数和相关关系对指挥要

素给出的评分记录 20000 条,评分分布在 $[0,5]$ 区间之内,评分越高,说明指挥员在某种战斗样式下对特定指挥要素的需求越强。以上数据按照总部 A、总队 B、支队 C、大队 D、中队 E、排 F、班 G 7 个指挥层级分级存储。本想定所述算例中协

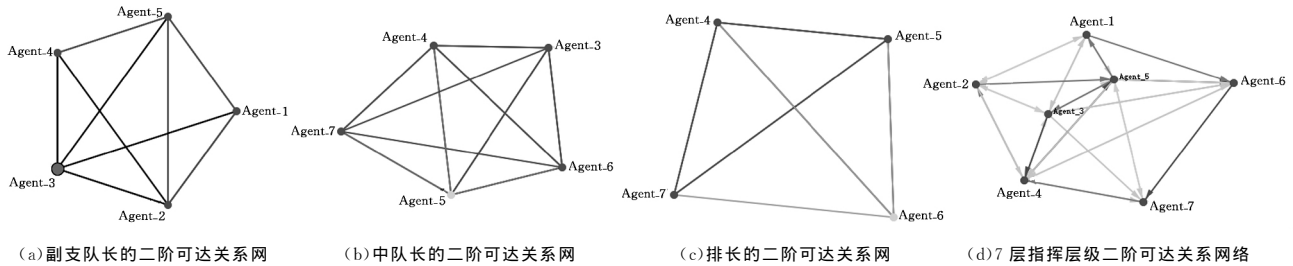


图 5 目标指挥员的二阶可达关系网络

其中, Agent1—Agent7 分别表示总部 A、总队 B、支队 C、大队 D、中队 E、排 F、班 G 7 个指挥层级,各层级间采用双向箭头连接,表示其信息流是交互性的;其相似性指标 C_{sim} 通过式(3)进行计算;相似关系的二阶可达关系则通过矩阵相乘得到,最后根据式(7)、式(8)进行去量纲化处理得到结果。

3.3.1 参数 β 的影响实验

本实验运用 MATLAB 工具箱,通过改变二阶段协同过滤中相似性判定参数 β ,来观察该参数对本文算法的影响,从而找到参数 β 的最佳值。根据反恐作战数据分析的一般规律,取协同过滤指挥要素数为 $N_1 = 20$ 来考查二阶段协同过滤中不同指挥要素数下 β 对推荐精度的影响,其中 N_2 分别取 10, 15, 20, 25, 30, 60, 仿真计算结果如图 6 所示。

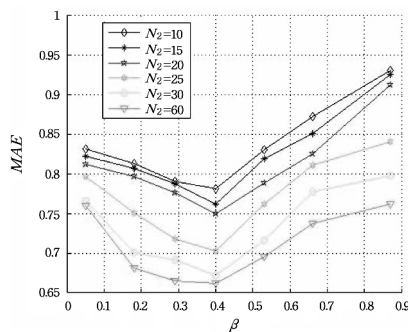


图 6 不同邻居集下参数 β 对推荐精度的影响

由图 6 可得出以下结论:

(1) 在不同邻居个数 N_2 下,随着 β 值的增大,推荐误差 MAE 的变化趋势基本一致;

(2) $\beta = 0.4$ 为最佳阈值,当 $\beta < 0.4$ 时,误差随 β 值的增大而减小,当 $\beta > 0.4$ 时,误差与 β 正相关;

(3) 相同 β 值下,误差 MAE 与 N_2 负相关,当 $N_2 > 30$ 时误差增幅减小。

实际上,随着推送指挥要素的增加,推荐效率会降低,指挥员的决策效率也会因要素的激增而受到影响,即全要素推荐对于指挥决策是没有意义的。为此,根据分析结果,在实际应用中可取 $\beta = 0.4$, $N_1 = 20$, $N_2 = 30$,从而保证在满足作战需求的条件下达到最高的推荐精度和推荐效率。

3.3.2 MCFM 与 Item Based-CF 推荐精度的对比实验

为了进一步检验本文所提出算法的有效性,在 3.3.1 节

同过滤矩阵的建立、相关关系的确定等数据均以此为基础。

3.3 实验设计与分析

基于凝聚子集分析构建的指挥层级间的相似性关系网络如图 5 所示。

的实验基础上(所涉及参数取最佳实验参数),以传统基于项目的协同过滤算法(Item Based Collaborative Filtering, IBCF)作为对照组进行推荐精度对比实验,以考查最近邻居集变化对算法推荐精度的影响,结果如图 7 所示。

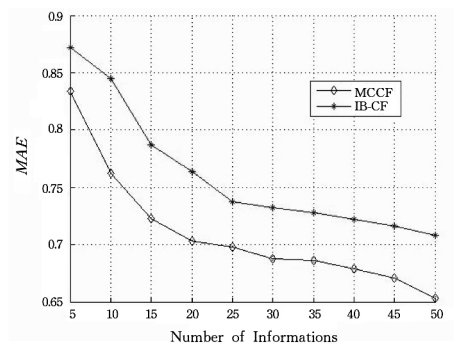


图 7 算法精度比较

由图 7 可知,随着最近邻居集即指挥要素数目的增加, MAE 值减小,推荐精度逐渐提高。本文所提出的面向任务的多重协同过滤算法相较于基于项目的协同过滤算法,在不同的指挥信息要素数目下,推荐误差 MAE 都相对较小,可见本文所提算法有效提高了推荐精度及指控系统的信息传输和推送精度。

3.3.3 MCFM 与 Item Based-CF 的推荐效率对比实验

为了验证本文所述方法的实际运行效率,选取不同指挥要素数目,以传统基于项目的协同过滤算法作为对照组进行推荐效率对比实验,结果如图 8 所示。

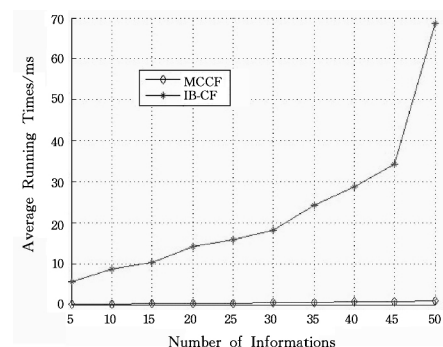


图 8 算法效率比较

由图 8 可知,随着“邻居项目”即指挥信息数量的增加,传

统基于项目的协同过滤算法的平均运行时间急剧增加,而本文所述面向任务的多重协同过滤算法的运行时间相对稳定,且远少于传统算法。这说明本文提出的多重协同过滤算法高效稳定,能够适应不同规模数据下的检索、过滤、推荐,对于提高指控系统的推荐效率作用显著。

结束语 针对指控系统中存在的数据量大、数据类型多样、数据推送实时性要求高等问题,以及反恐行动中多任务、多要素、多层次的特点,提出一种反恐指挥信息多重协同过滤算法。仿真实验表明,该算法能够较好地适应指控系统中指挥要素信息的推送要求,能够一定程度上提高系统的自主推荐效率和推荐准确性。

下一步研究方向包括:1)着重对不同规模数据,特别是大数据条件下算法的适应性进行研究,拓展算法的适用范围;2)针对反恐行动通常由多警种、多军种联合实施的客观实际,结合不同反恐力量指挥员间的相似性关系分析,进行协同过滤分析,提高算法的泛化性,以期在联合反恐中发挥作用;3)评分矩阵的初始化与矩阵填充是协同过滤算法的难点,也是克服协同过滤稀疏性问题的关键。本例虽然通过改进相似性计算方法在一定程度上提高了评分的容错率,但效果十分有限。下一步研究中可以将机器学习和人工智能的相关理论应用到评分矩阵的初始化计算中,以降低评分的主观性,同时进一步完善指挥信息要素体系,细分战斗时节,提高专家评估的准确性,在执行任务中不断提高算法的精准度。

参 考 文 献

- [1] National strategy for counter terrorism[R]. RAND,2011:1-10.
- [2] 高娜. 基于标签和评分联合学习的协同过滤推荐算法研究[D]. 南京:南京师范大学,2015:64-73.
- [3] LUO X,OUYANG Y X,XIONG Z,et al. The effect of similarity support in K-Nearest-Neighborhood based collaborative filtering [J]. Journal of Computer,2010,8(33):1339-1445.
- [4] DESHPANDE M,KARPIS G. Item-based top-n recommendation algorithms[J]. ACM Transactions on Information Systems,2004,22(1):143-177.
- [5] ZHANG X J. Personalized online sales using web usage data mining [J]. Computers in Industry,2007,58(3):772-782.
- [6] LIU J,WANG Q P,FANG K. An optimized collaborative filtering approach combining with item based prediction[C]//Proceeding of the 2007,11th International Conference on Computer Supported Cooperative Work in Design. 2007:157-161.
- [7] 黎明,徐德智. 一种结合基于项目和用户的个性化推荐算法[J]. 小型微型计算机系统,2011(4):611-613.
- [8] JALIRE M,TOSCHER A,LEGENSTEIN R. Combing predictions for accurate recommend systems [C]//Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM,2010:693-702.
- [9] GONG S. A collaborative filtering recommendation algorithm based on user clustering and item clustering[J]. Journal of Software,2010,5(7):745-752.
- [10] 李红梅,郝文宁,陈刚. 基于改进 LSH 的协同过滤推荐算法[J]. 计算机科学,2015,42(10):256-261.
- [11] 王付强,彭甫镭,丁小焕,等. 基于位置的非对称相似性度量的协同过滤推荐算法[J]. 计算机应用,2016(1):171-174,180.
- [12] GOLDBERG K,ROEDER T,GUPTA D,et al. Eigenstate:a constant time collaborative filtering algorithm [J]. Information Retrieval,2001,4(2):133-151.
- [13] 徐红艳,杜文刚,冯勇,等. 一种基于多属性评分的协同过滤算法[J]. 辽宁大学学报(自然科学版),2015(2):136-142.
- [14] 柳先辉,徐梦锦. 基于多级图划分的协同过滤算法研究[J]. 机械设计与制造工程,2015(12):14-17.
- [15] 硕良勋,柴变芳,张新东. 基于改进最近邻的协同过滤推荐算法[J]. 计算机工程与应用,2015(5):137-141.
- [16] 李德毅,孟海军,史雪梅. 隶属云和隶属云发生器[J]. 计算机研究与发展,1995,3(6):15-20.
- [17] 宋远骏,李德毅,杨孝宗. 电子产品可靠性的云模型评价方法[J]. 电子学报,2000,28(12):74-76.
- [18] 廖列法,黎晨,孟祥茂. 基于欧氏空间相似度的云模型协同过滤算法[J]. 计算机工程与科学,2015,10(37):1977-1982.
- [19] WASSERMAN S,FANST K. Social network analysis:Methods and applications[M]. Cambridge University Press. 1994.
- [20] WELLMAN B. Structural analysis:From method and metaphor to theory and substance[M]//Social structures:A network approach. Cambridge University Press;Cambridge,1988.