

基于项目的三支决策推荐系统

熊文丹, 马建敏*

长安大学, 理学院, 陕西 西安

收稿日期: 2021年12月25日; 录用日期: 2022年1月21日; 发布日期: 2022年1月28日

摘要

为了用最小的推荐损失给用户推荐他们感兴趣的项目, 解决基于项目的推荐系统只根据项目相似度进行推荐的单一标准, 本文提出一种基于项目的三支决策推荐系统。该系统在基于项目的推荐系统的基础上引入三支决策, 根据受欢迎程度将项目划分为推荐、延迟推荐和不推荐三类, 进而根据用户对推荐项目的兴趣度, 优先为用户推荐兴趣度最大的项目。实验结果表明, 基于项目的三支决策推荐系统与基于项目的推荐系统、基于用户的三支决策推荐系统和基于项目的二支决策推荐系统相比, 推荐精度提高且损失较小。

关键词

三支决策推荐系统, 三支决策, 阈值, 损失矩阵, 兴趣度

Item-Based Three-Way Decision Recommendation System

Wendan Xiong, Jianmin Ma*

School of Science, Chang'an University, Xi'an Shaanxi

Received: Dec. 25th, 2021; accepted: Jan. 21st, 2022; published: Jan. 28th, 2022

Abstract

In order to recommend items of interest to users with minimal loss of recommendation, and avoid the single recommendation standard based on the similarity of items in traditional recommendation system, this paper proposes a three-way decision recommendation system. On the basis of the traditional recommendation system, the idea of three-way decision is introduced into it. Any item

*通讯作者。

can be made the categories of recommended, delayed or not recommended according to the popularity of them. Finally, items with the maximum interestingness are recommended to a user. The experimental results show that, compared with the traditional recommendation system, three decision recommendation systems based on users and two decision recommendation systems based on item, the item-based three-way decision recommendation system can improve the quality of recommendation and minimize the total loss of the recommendation.

Keywords

Three-Way Decisions Recommendation System, Three-Way Decisions, Threshold, Loss Matrix, Interestingness

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

21 世纪处于大数据时代, 人们经常遇到自己所需要的信息与被推荐的信息不匹配的情况, 因此推荐系统的逐步完善成为了当今世界亟待解决的问题, 本文提供了一种基于项目的推荐系统, 能够提高推荐准确率, 使得其更人性化。在过去的十几年里, 催生了非常多的推荐技术, 涌现了许多的推荐系统, 例如: Amazon 的个性化物品推荐系统、Netflix 的视频推荐、Facebook 的好友推荐, 今日头条的时事新闻推荐, 最近几年, 抖音、快手、淘宝也是国内非常火热的推荐系统。基于项目的推荐系统是一种根据项目之间的相似度, 再根据用户对相关项目的兴趣度, 为用户提供有用的建议。然而对于信息消费者来说从大量信息中找到自己感兴趣的信息是一件非常困难的事情; 对于信息生产者来说让自己生产的信息脱颖而出也是一件非常困难的事情, 而推荐系统让事情变得简单[1]。

常用的推荐系统有基于内容的推荐, 就是推荐与用户过去喜欢的内容相似的项目[2]; 基于知识的推荐, 该类推荐系统利用用户的显示需求和特定的领域产品的深度来推荐项目[3]; 混合推荐系统, 该类推荐系统往往综合了上面提到的多种技术, 利用优势互补来提高推荐系统的总体性能[4]; 基于协同过滤的推荐系统[5], 该推荐系统有两种类型: 基于用户的协同过滤推荐系统[6]和基于项目的协同过滤推荐系统[7]。本文将从基于项目的推荐系统进行研究分析。

为了解决传统推荐系统仅考虑相似度进行项目推荐, 缺乏推荐精度, 本文将三支决策思想引入推荐系统中, 提出基于项目相似度和受欢迎程度的三支决策推荐系统。作为粗糙集理论的一种扩张[8], 三支决策主要被用于处理延迟决策的不确定性问题。目前三支决策已在很多领域得到成功应用, 如垃圾邮件的过滤[9]、政策制定[10]、聚类系统[11]等。叶晓庆等人[12]将三支粒融入基于用户的推荐系统中, 得到了推荐精度提高且损失减小的结果; 张恒汝等人[13]提出了基于回归的三支推荐系统, 通过调整不同行为阈值使总的推荐损失达到最小。

本文在基于项目的推荐系统的基础上, 将三支决策与推荐系统相结合, 充分考虑项目间的相似度和受欢迎程度, 通过调整阈值得到最小推荐损失, 实现对项目的推荐、延迟推荐和不推荐的分类准则。进而根据用户对项目的兴趣度, 为用户精准推荐项目, 提高为用户推荐项目的准确性, 减少误推荐, 使推荐损失达到最小。最后利用 MovieLens 数据集中用户对电影的评分数据对推荐系统进行验证。

2. 相关概念

2.1. 基于项目的推荐系统

三元组 (U, I, F) 称为用户 - 项目评分信息系统[12], 其中 $U = \{u_1, u_2, \dots, u_n\}$ 是所有用户的集合, $I = \{t_1, t_2, \dots, t_m\}$ 是所有项目的集合, $F = \{f: U \times I \rightarrow V\}$ 是 $U \times I$ 上信息函数构成的集合, 则对任意的 $u_k \in U$, $t_j \in I$, $r_{kj} = f(u_k, t_j) \in V$ 表示用户 u_k 对项目 t_j 的评分值, $V = \{0, 1, 2, 3, 4, 5\}$ 为评分值域. $r_{kj} = 0$ 表示用户 u_k 未给项目评分, $r_{kj} \neq 0$ 表示用户 u_k 给项目评过. 表 1 给出了用户 - 项目评分信息系统.

Table 1. User-Itemrating information system (U, I, F)

表 1. 用户 - 项目评分信息系统 (U, I, F)

	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
u_1	5	3	0	4	4	4	0	0
u_2	0	1	0	3	0	1	3	2
u_3	0	3	4	0	0	4	5	0
u_4	0	0	2	0	0	0	3	0
u_5	3	2	1	2	0	2	0	1

如表 1 所示, $r_{12} = f(u_1, t_2) = 3$ 表示用户 u_1 对项目 t_2 的评分为 3 分, $r_{21} = f(u_2, t_1) = 0$ 表示用户 u_2 对项目 t_1 未评分. 若判断是否为用户 u_2 推荐项目 t_1 , 文献[1]提出了基于项目的推荐系统, 其核心思想是, 对于未评分项目, 通过计算用户对未评分项目与用户对评分项目的相似度, 进而引入用户对未评分项目的兴趣度, 决定是否给用户推荐该项目.

对任意两个项目 t_i 与 t_j , 相似度[1]定义为:

$$s_{ij} = \frac{|N(t_i) \cap N(t_j)|}{\sqrt{|N(t_i)| |N(t_j)|}}, \quad (1)$$

其中, $N(t_i) = \{u_k \in U | r_{ki} > 0\}$ 表示给项目 t_i 评分的用户集合, $|N(t_i)|$ 表示集合 $N(t_i)$ 的基数.

用户 u_k 对项目 t_j 的兴趣度[1]为:

$$q_{kj} = \sum_{i \in P(u_k) \cap V(t_i, n)} s_{ij} r_{ki}, \quad (2)$$

$P(u_k) = \{t_i \in I | r_{ki} > 0\}$ 是用户 u_k 评价的项目集合, $V(t_i, n)$ 表示和项目 t_i 最相似的 n 个近邻项目的集合.

算法 1: 基于项目的推荐系统算法如下[14]:

输入: 用户 - 项目评分信息系统 (U, I, F)

输出: 为用户 u_k 推荐的项目

步骤 1: 根据公式(1), 计算两个项目 t_i 和 t_j 的相似度 s_{ij} , 将最相似的 n 个项目提出;

步骤 2: 根据公式(2), 计算用户 u_k 对推荐项目 t_j 的兴趣度 q_{kj} ;

步骤 3: 将兴趣度从大到小排列, 依次将项目推荐给用户 u_k .

2.2. 三支决策

定义 1 [15] [16] 设论域 U 表示非空有限对象构成的集合, $R \subseteq U \times U$ 是定义在 U 上的等价关系. 记 $apr = (U, R)$ 为 Pawlak 近似空间. 对任意 $x \in U$, 记 $[x]_R = \{y \in U | (x, y) \in R\}$ 为 x 关于等价关系 R 的等价

类。对 $X \subseteq U$, $\alpha, \beta \in (0,1)$ 且 $0 \leq \beta < \alpha \leq 1$, X 的 (α, β) -下、上近似集分别定义为:

$$\begin{aligned} \underline{apr}_{(\alpha, \beta)}(X) &= \{x \in U \mid P(X \mid [x]_R) \geq \alpha\}, \\ \overline{apr}_{(\alpha, \beta)}(X) &= \{x \in U \mid P(X \mid [x]_R) > \beta\}, \end{aligned} \quad (3)$$

其中, $P(X \mid [x]_R) = \frac{|X \cap [x]_R|}{|[x]_R|}$ 为条件概率。 (α, β) -下、上近似集将论域分为三个部分[17] [18] [19] [20]:

$$\begin{aligned} POS_{(\alpha, \beta)}(X) &= \{x \in U \mid P(X \mid [x]_R) \geq \alpha\}, \\ BND_{(\alpha, \beta)}(X) &= \{x \in U \mid \beta < P(X \mid [x]_R) < \alpha\}, \\ NEG_{(\alpha, \beta)}(X) &= \{x \in U \mid P(X \mid [x]_R) \leq \beta\}. \end{aligned} \quad (4)$$

文献[19]提出了利用损失函数构造决策总风险最小时的三支决策规则, 从而得到阈值 α 和 β 。

设 $\Omega = \{\omega_1, \dots, \omega_s\}$ 表示 s 个状态构成的集合, $A = \{a_1, \dots, a_m\}$ 表示 m 个行为构成的集合, $\lambda(a_i \mid \omega_j)$ 表示状态为 ω_j 时采取行为 a_i 的损失, $P(\omega_j \mid x)$ 表示对象 x 在状态 ω_j 下的条件概率。则采取行为 a_i 的期望损失为[19]:

$$R(a_i \mid x) = \sum_{j=1}^s \lambda(a_i \mid \omega_j) P(\omega_j \mid x). \quad (5)$$

在三支决策理论中[21], $\Omega = \{X, X^c\}$ 表示状态集合(属于 X 和不属于 X), $A = \{a_p, a_N, a_B\}$ 表示行为集合, 即对于不同状态, 采取接受, 延迟和拒绝三种行为策略。由于采取不同行为会产生不同损失, 分别用 λ_{PP} 、 λ_{BP} 、 λ_{NP} 表示当对象 x 处于状态 X ($x \in X$) 时, 采取行为 a_p 、 a_B 、 a_N 的损失; 分别用 λ_{PN} 、 λ_{BN} 、 λ_{NN} 表示当对象 x 处于状态 X^c ($x \notin X$) 时, 采取行为 a_p 、 a_B 、 a_N 的损失。表 2 给出了对应的损失矩阵:

Table 2. Loss matrix
表 2. 损失矩阵

行为	状态	
	X	X^c
a_p	λ_{PP}	λ_{PN}
a_N	λ_{BP}	λ_{BN}
a_B	λ_{NP}	λ_{NN}

于是采取 a_p 、 a_B 、 a_N 三种行为对应的期望损失分别为:

$$\begin{aligned} R(a_p \mid [x]_R) &= \lambda_{PP} P(X \mid [x]_R) + \lambda_{PN} P(X^c \mid [x]_R), \\ R(a_B \mid [x]_R) &= \lambda_{BP} P(X \mid [x]_R) + \lambda_{BN} P(X^c \mid [x]_R), \\ R(a_N \mid [x]_R) &= \lambda_{NP} P(X \mid [x]_R) + \lambda_{NN} P(X^c \mid [x]_R). \end{aligned} \quad (6)$$

根据最小风险决策规则, 需选择损失最小的行为集作为最佳行动方案, 可得如下决策规则:

- (P): 若 $R(a_p \mid [x]_R) \leq R(a_B \mid [x]_R)$ 和 $R(a_p \mid [x]_R) \leq R(a_N \mid [x]_R)$, 则 $x \in POS(X)$,
- (B): 若 $R(a_B \mid [x]_R) \leq R(a_p \mid [x]_R)$ 和 $R(a_B \mid [x]_R) \leq R(a_N \mid [x]_R)$, 则 $x \in BND(X)$,
- (N): 若 $R(a_N \mid [x]_R) \leq R(a_p \mid [x]_R)$ 和 $R(a_N \mid [x]_R) \leq R(a_B \mid [x]_R)$, 则 $x \in NEG(X)$ 。

由于 $P(X|[x]_R) + P(X^c|[x]_R) = 1$, 上述规则只与概率 $P(X|[x]_R)$ 和损失函数 λ_{ij} ($i, j = P, B, N$) 有关。此外, 考虑到接受正确事物的损失不大于延迟接受正确事物的损失, 且这两者都小于拒绝正确事物的损失; 拒绝错误事物的损失不大于延迟拒绝错误事物的损失, 且这两者都小于接受错误事物的损失, 因此, 损失函数的大小关系满足[18]:

$$\begin{aligned} \lambda_{PP} &\leq \lambda_{BP} < \lambda_{NP}, \\ \lambda_{NN} &\leq \lambda_{BN} < \lambda_{PN}. \end{aligned} \quad (7)$$

故上述决策规划可化简为以下决策规则:

(P): 若 $P(X|[x]_R) \geq \alpha$ 和 $P(X|[x]_R) \geq \gamma$, 则 $x \in POS(X)$;

(B): 若 $P(X|[x]_R) \leq \alpha$ 和 $P(X|[x]_R) \geq \beta$, 则 $x \in BND(X)$;

(N): 若 $P(X|[x]_R) \leq \beta$ 和 $P(X|[x]_R) \leq \gamma$, 则 $x \in NEG(X)$,

其中:

$$\begin{aligned} \alpha &= \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \\ \beta &= \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}, \\ \gamma &= \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}. \end{aligned} \quad (8)$$

由决策规则(B)可知 $\beta < \alpha$, 进而可得:

$$\frac{\lambda_{BP} - \lambda_{PP}}{\lambda_{PN} - \lambda_{BN}} < \frac{\lambda_{NP} - \lambda_{BP}}{\lambda_{BN} - \lambda_{NN}}. \quad (9)$$

由公式(7)和公式(8)即得 $0 \leq \beta < \gamma < \alpha \leq 1$. 于是决策规则(P) (B) (N)可进一步简化为:

(P1): 若 $P(X|[x]_R) \geq \alpha$, 则 $x \in POS(X)$;

(B1): 若 $\beta < P(X|[x]_R) < \alpha$, 则 $x \in BND(X)$;

(N1): 若 $P(X|[x]_R) \leq \beta$, 则 $x \in NEG(X)$ 。

3. 基于项目的三支决策推荐系统

本文在基于项目的推荐系统中引入三支决策思想, 构建基于项目的三支决策推荐系统, 其目的是给目标用户推荐项目时提高推荐精度并降低推荐过程中的误推荐损失。 $\Omega = \{W, W^c\}$ 表示项目受欢迎和不受欢迎的状态集, $A = \{a_P, a_B, a_N\}$ 表示推荐、延迟推荐和不推荐的推荐行为集。表 3 给出了基于项目的三支决策推荐系统的损失矩阵。

Table 3. Three-way decision recommendation system cost matrices

表 3. 三支决策推荐系统损失矩阵

行为	项目状态	
	项目受欢迎(W)	项目不受欢迎(W^c)
推荐(a_P)	λ_{PP}	λ_{PN}
延迟推荐(a_B)	λ_{BP}	λ_{BN}
不推荐(a_N)	λ_{NP}	λ_{NN}

表中 λ_{PP} , λ_{BP} , λ_{NP} 分别表示项目受欢迎且被推荐, 延迟推荐和不推荐的损失; λ_{PN} , λ_{BN} , λ_{NN} 分别表示项目不受欢迎且被推荐, 延迟推荐和不推荐的损失。

当 $r_{kj} = 0$ 时, 给用户 u_k 推荐(a_P)、延迟推荐(a_B)或不推荐(a_N)项目 t_j 对应的损失分别为:

$$\begin{aligned} R(a_P | T_k(t_j)) &= \lambda_{PP}P(W | T_k(t_j)) + \lambda_{PN}P(W^c | T_k(t_j)), \\ R(a_B | T_k(t_j)) &= \lambda_{BP}P(W | T_k(t_j)) + \lambda_{BN}P(W^c | T_k(t_j)), \\ R(a_N | T_k(t_j)) &= \lambda_{NP}P(W | T_k(t_j)) + \lambda_{NN}P(W^c | T_k(t_j)). \end{aligned} \quad (10)$$

其中, $T_k(t_j) = \{t_i : r_{kj} = 0, r_{ki} > 0 \text{ 且 } s_{ij} > \bar{S}\}$ 表示与未评分项目 t_j 相似又被用户 u_k 评分的项目集合, \bar{S} 表示所有项目的平均相似度, 其公式为:

$$\bar{S} = \frac{2 \cdot \sum_{i=1}^{m-1} \sum_{j>i}^m s_{ij}}{m(m-1)}. \quad (11)$$

$P(W | T_k(t_j))$ 表示与未评分项目 t_j 相似且被用户 u_k 评分的项目受欢迎的概率。

根据最小风险决策规则(P)、(B)、(N)可得如下规则:

(P2): 若 $P(W | T_k(t_j)) \geq \alpha$, 给用户 u_k 推荐项目 t_j ;

(B2): 若 $\beta < P(W | T_k(t_j)) < \alpha$, 给用户 u_k 延迟推荐项目 t_j ;

(N2): 若 $P(W | T_k(t_j)) \leq \beta$, 给用户 u_k 不推荐项目 t_j 。

其中阈值 α 和 β 取值为:

$$\begin{aligned} \alpha &= \frac{\lambda_{PN} - \lambda_{BN}}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})}, \\ \beta &= \frac{\lambda_{BN} - \lambda_{NN}}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})}. \end{aligned} \quad (12)$$

特别的, 二支决策推荐过程只考虑推荐与不推荐两种行为, $A = \{a_P, a_N\}$, $\Omega = \{W, W^c\}$ 表示受欢迎的和不受欢迎的状态集。表 4 给出了基于项目的二支决策推荐损失矩阵:

Table 4. Two-way decision recommendation system cost matrices
表 4. 基于项目的二支决策推荐系统

行为	项目状态	
	项目受欢迎(W)	项目不受欢迎(W^c)
推荐(a_P)	λ_{PP}	λ_{PN}
不推荐(a_N)	λ_{NP}	λ_{NN}

根据最小风险决策规则(P) (B) (N)可得如下规则:

(P3): 若 $P(W | T_k(t_j)) \geq \gamma$, 则为用户 u_k 推荐项目 t_j ,

(N3): 若 $P(W | T_k(t_j)) < \gamma$, 则为用户 u_k 不推荐项目 t_j ,

阈值 γ 取值为:

$$\gamma = \frac{\lambda_{PN} - \lambda_{NN}}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})}. \quad (13)$$

根据决策规则, 获取推荐项目后, 根据公式(2), 计算出用户 u_k 对推荐项目的兴趣度 q_{kj} , 并根据兴趣度从大到小排列, 选取排名靠前的项目 s 优先推荐给用户 u_k 。

因此, 得到表 5 基于项目的三支决策推荐系统算法。

Table 5. Item-based three-way decisions recommend system algorithms

表 5. 基于项目的三支决策推荐系统算法

算法 2: 基于项目的三支决策推荐系统算法	
输入:	用户 - 项目评分信息系统 (U, I, F)
输出:	为用户 u_k 推荐项目 t_j
步骤 1:	根据公式(1)计算项目间相似度 s_{ij} 。
步骤 2:	根据公式(12), 利用损失函数计算出阈值 α 和 β 。
步骤 3:	根据三支决策风险决策规则将用户 u_k 评分项目 t_i 划分为三类, 推荐, 延迟推荐和不推荐。
步骤 4:	与项目 t_i 最相似的 n 个未评分项目 t_j 也划分为三类。
步骤 5:	根据公式(2), 计算用户 u_k 对划分到推荐项目 t_j 的兴趣度 q_{kj} 。
步骤 6:	将兴趣度从大到小排列, 依次推荐给用户 u_k 。

4. 实验结果及分析

4.1. 数据集

实验数据采用 <http://www.grouplens.org> 网站提供的公开 MovieLens 数据集(用户对电影评分数据集)。评分是 1~5 的离散值, 由 6040 名用户, 3706 部电影和约 100 万条评分组成, 本文随机选取 87.5% 的数据进行训练, 其余的数据进行测试, 本文在采集数据后, 将评分量过少, 且不具有意义的数据进行数据清洗[22], 得到表 6 所示数据。

Table 6. Dataset

表 6. 数据集

数据集	用户数	项目数	评分范围	最少评分数	稀疏等级
MovieLens	6040	3706	1~5	20	0.8958

4.2. 误推荐损失计算

本文在训练集中计算误推荐损失, 比较基于项目的三支决策推荐系统与基于项目的二支决策推荐系统的损失。当 $r_{kj} \geq \bar{r}_k$ 时, 表明用户 u_k 喜欢项目 t_j , 故项目 t_j 应该被推荐; 当 $r_{kj} < \bar{r}_k$ 时, 表明用户 u_k 不喜欢项目 t_j , 故项目 t_j 应该不被推荐, 其中, \bar{r}_k 为用户 u_k 对项目的喜好阈值:

$$\bar{r}_k = \frac{\sum_{j=1}^n r_{kj}}{|P(u_k)|}, \quad (14)$$

其中 $P(u_k) = \{t_j \in I : r_{kj} > 0\}$ 表示用户 u_k 评过分的的项目集合。

对于三支决策的三种决策: 推荐、不推荐和延迟推荐, 其所产生的误推荐损失分为四类[23]: 推荐 \rightarrow 不推荐 ($a_p \rightarrow a_N$)、推荐 \rightarrow 延迟推荐 ($a_p \rightarrow a_B$)、不推荐 \rightarrow 推荐 ($a_N \rightarrow a_p$)、不推荐 \rightarrow 延迟推荐 ($a_N \rightarrow a_B$)。

也即, 若 $r_{kj} \geq \bar{r}_k$ 时, 则项目 t_j 应该被推荐给用户 u_k , 却做出对项目 t_j 不推荐的决策, 从而造成损失, 记为 $a_p \rightarrow a_N$ 。用 $n_{a_p \rightarrow a_N}$ 表示将推荐项目归类为不推荐项目的误推荐项目的数量。故根据三支决策推荐系统的损失矩阵可得其误推荐损失为:

$$E_3 = \lambda_{NP}n_{a_p \rightarrow a_N} + \lambda_{BP}n_{a_p \rightarrow a_B} + \lambda_{PN}n_{a_N \rightarrow a_p} + \lambda_{BN}n_{a_N \rightarrow a_B}. \quad (15)$$

特别的, 对于二支决策, 其误推荐损失仅含两类: 推荐 \rightarrow 不推荐 ($a_p \rightarrow a_N$), 不推荐 \rightarrow 推荐 ($a_N \rightarrow a_p$)。根据二支决策推荐系统的损失矩阵可得其误推荐损失为:

$$E_2 = \lambda_{NP}n_{a_p \rightarrow a_N} + \lambda_{PN}n_{a_N \rightarrow a_p}. \quad (16)$$

4.3. 评价指标

实验采用覆盖率(COVERAGE)、准确率(PRECISION)、召回率(RECALL)和 F1 值 4 个评价指标来比较系统的推荐性能。

1) 覆盖率(COVERAGE)计算能够预测项目占所有待测项目的比率, 从而评价预测的全面性。COVERAGE 越大, 预测质量越高。假设有推荐结果的项目为 h 个, 则 COVERAGE 的计算公式为[1]:

$$\text{COVERAGE} = \frac{h}{n}. \quad (17)$$

2) 准确率(PRECISION), 它表示用户对一个被推荐项目感兴趣的可能性; 召回率(RECALL)表示一个用户喜欢的项目被推荐的概率, F1 值表示得分值, 准确率, 召回率, F1 值计算公式如表 7 所示:

Table 7. Accuracy, Recall Rate, and F1 Value

表 7. 准确率、召回率和 F1 值

准确率(P)	召回率(R)	F1
$\text{PRECISION} = \frac{TP}{TP + FP}$	$\text{RECALL} = \frac{TP}{TP + FN}$	$\text{F1} = \frac{2PR}{P + R}$

其中相关指标说明如下表 8 所示:

Table 8. Description of relevant metrics

表 8. 相关指标说明

		预测		合计
		推荐项目数	未被推荐项目数	
实际	正确推荐(正)	TP	FN	TP + FN
	错误推荐(负)	FP	TN	FP + TN
合计		TP + FP	TN + FN	

4.4. 结果及分析

4.4.1. 基于项目的二支决策推荐系统与基于项目的三支决策推荐系统损失对比

计算推荐系统产生的损失时, 主要考虑错误推荐所产生的误分类损失以及延迟推荐所产生的学习损失, 因此可认为正确推荐不产生推荐损失, 即 $\lambda_{pp} = \lambda_{NN} = 0$, 且延迟推荐损失相同 $\lambda_{BP} = \lambda_{BN}$ 。

通过实验, 可以得出 $\bar{S} = 0.0764$, 取 $\lambda_{PP} = \lambda_{NN} = 0$, 固定 $\lambda_{PN} = 0.9$ 得出项目推荐损失和延迟推荐数与误推荐数, 本文将基于项目的三支决策推荐系统与基于项目的二支决策推荐系统所产生的损失对比试验如表 9 所示:

Table 9. Fixed $\lambda_{PN} = 0.9$ recommendation loss

表 9. 固定 $\lambda_{PN} = 0.9$ 推荐损失

$(\lambda_{PN}, \lambda_{NP})$	E_3					E_2
	$\lambda_{BP} = \lambda_{BN} = 0.18$	$\lambda_{BP} = \lambda_{BN} = 0.225$	$\lambda_{BP} = \lambda_{BN} = 0.27$	$\lambda_{BP} = \lambda_{BN} = 0.315$	$\lambda_{BP} = \lambda_{BN} = 0.36$	
(0.9, 0.6)	730.26	915.525	1037.52	1134.45	1178.46	2231
(0.9, 0.5)	750.06	916.65	1037.97	1134.225	1177.56	2049
(0.9, 0.45)	757.62	922.5	1042.11	1136.655	1178.28	2049
(0.9, 0.4)	765.9	925.875	1040.58	1130.22	1166.94	2047
(0.9, 0.36)	771.3	924.615	1032.66	1115.64	1145.7	2046

Table 10. Fixed $\lambda_{PN} = 0.9$ the number of error recommendations (errors) and delayed recommendations (delays) that result from Item recommendations based on three decisions

表 10. 固定 $\lambda_{PN} = 0.9$ 时基于三支决策的项目推荐产生的误推荐(误)和延迟推荐(延迟)项目数

$(\lambda_{PN}, \lambda_{NP})$	$\lambda_{BP} = \lambda_{BN} = 0.18$		$\lambda_{BP} = \lambda_{BN} = 0.225$		$\lambda_{BP} = \lambda_{BN} = 0.27$		$\lambda_{BP} = \lambda_{BN} = 0.315$		$\lambda_{BP} = \lambda_{BN} = 0.36$	
	误	延迟	误	延迟	误	延迟	误	延迟	误	延迟
(0.9, 0.6)	206	3027	360	2629	536	2056	753	1450	923	966
(0.9, 0.5)	215	3112	369	2614	545	2041	762	1435	932	951
(0.9, 0.45)	248	3074	402	2576	578	2003	795	1397	965	913
(0.9, 0.4)	323	2965	477	2467	653	1894	870	1288	1040	804
(0.9, 0.36)	425	2817	579	2319	755	1746	972	1140	1142	656

通过表 9 的对比试验可以看出, 基于二支决策推荐系统所产生的损失远远高于三支决策推荐系统所产生的损失, 通过三支决策推荐系统所产生的损失矩阵可以得出, 当 $\alpha = 0.8$, $\beta = 0.3$ 时, 推荐损失达到最小。表 10 数据可以得出, 基于项目的三支决策推荐系统的误推荐项目数量和延迟推荐项目数量在 $\alpha = 0.8$, $\beta = 0.3$ 时, 数量最少。因此本文将用此阈值为用户推荐项目。

4.4.2. 推荐精度对比

将提出的基于三支决策的推荐系统与传统的推荐系统进行准确度、召回率、F1 和覆盖率的对比, 在这里将传统基于项目的推荐系统简称为 IT-CF-SN, 将基于项目的二支决策推荐系统简称为 IT-CF-ST, 将基于用户的三支决策推荐系统简称为 IT-CF-SH, 将基于项目的三支决策推荐系统简称为 IT-CF-SL, n 表示近邻个数。

如图 1 所示, 基于项目的三支决策推荐系统(IT-CF-SL)与传统基于项目的推荐系统(IT-CF-SN)、基于用户的三支决策推荐系统(IT-CF-SH), 基于项目的二支决策推荐系统(IT-SF-ST)相比, 基于项目的三支

决策推荐系统的准确率高于传统推荐系统。从整体来看, 提出的基于项目的三支决策推荐系统在推荐精度上具有明显优势。

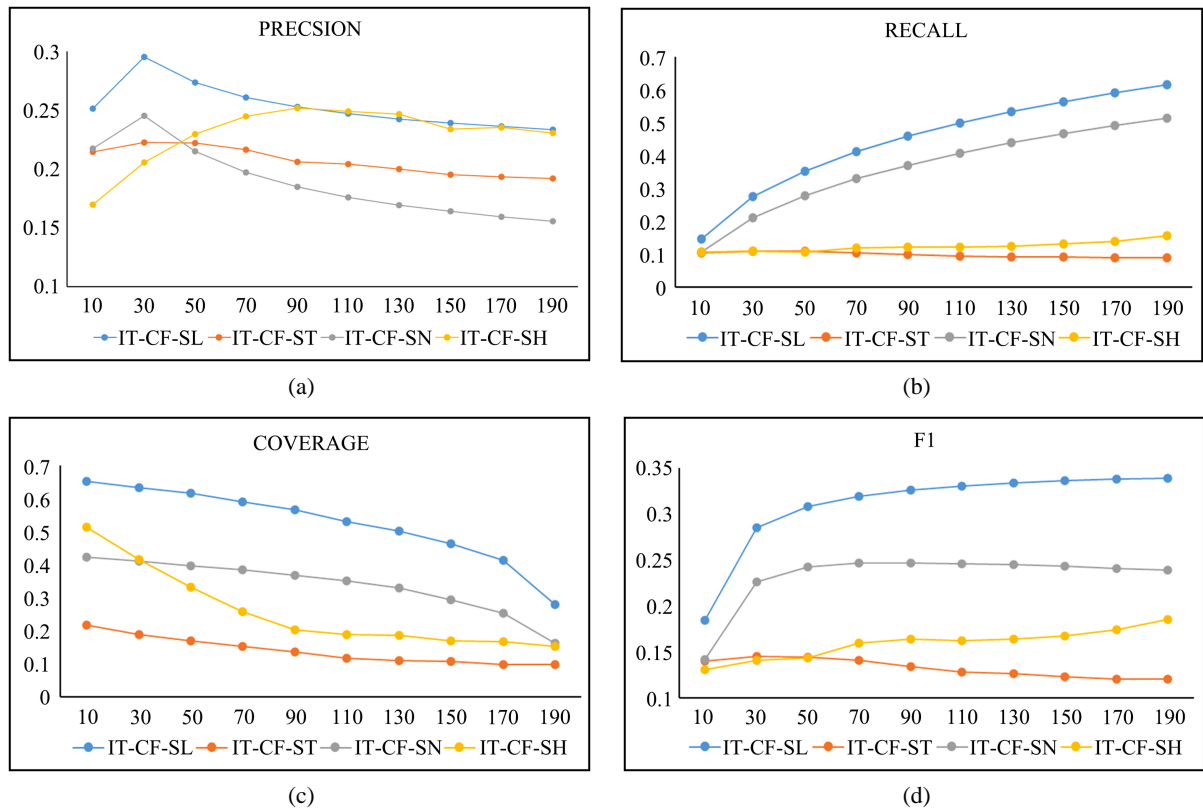


Figure 1. Recommended system algorithm comparison. (a) Accuracy; (b) Recall; (c) Coverage; (d) F1 value
图 1. 推荐系统算法比较。(a) 精确度; (b) 召回率; (c) 覆盖率; (d) F1 值

4.4.3. 推荐结果

本文拟给 id 为“1”的用户推荐电影, 因篇幅限制, 只为其推荐 5 部电影, 结果如表 11:

Table 11. Movie recommended by a user with id as “1”
表 11. 为 id 为“1”的用户推荐的电影

电影 id	电影名	决策规则	用户兴趣度
1196	Star Wars: Episode V-The Empire Strikes Back (1980)	推荐	3.4336
318	Shawshank Redemption, The (1994)	推荐	2.855
364	Lion King, The (1994)	推荐	2.770
595	Beauty and the Beast (1991)	推荐	2.608
1265	Groundhog Day (1993)	推荐	2.334

5. 结论

基于项目的三支决策推荐系统与传统推荐系统、基于项目的二支决策推荐系统和基于用户的三支决

策推荐系统相比, 推荐精度更高, 且基于项目的三支决策推荐系统能够将相似度高且受欢迎的项目推荐给用户。基于项目的三支决策推荐系统在提高推荐质量的同时, 减少了误推荐的损失, 从而降低了总的推荐损失, 使用户的推荐体验更好。

由于各方面因素限制, 本文只考虑了项目间相似度和项目的受欢迎程度, 在进一步工作中将考虑项目的类别来进行三支决策的划分, 讨论推荐结果的精确性, 达到因人而异的推荐效果。

基金项目

国家自然科学基金资助项目(61772019, 61603278)。

参考文献

- [1] 项亮. 推荐系统实践——利用用户行为数据[M]. 北京: 人民邮电出版社, 2012.
- [2] Aciar, S., Zhang, D., Simoff, S. and Debenham, J. (2007) Informed Recommender: Basing Recommendations on Consumer Product Reviews. *IEEE Intelligent Systems*, **22**, 39-47. <https://doi.org/10.1109/MIS.2007.55>
- [3] 刘平峰, 聂规划, 陈冬林. 基于知识的电子商务智能推荐系统平台设计[J]. 计算机工程与应用, 2007, 43(19): 199-201+216.
- [4] de Campos, L.M., Fernández-Luna, J.M., Huete, J.F. and Rueda-Morales, M.A. (2010) Combining Content-Based and Collaborative Recommendations: A Hybrid Approach Based on Bayesian Networks. *International Journal of Approximate Reasoning*, **51**, 785-799. <https://doi.org/10.1016/j.ijar.2010.04.001>
- [5] 崔梓凝. 基于协同过滤的推荐算法研究[J]. 数字化用户, 2017, 23(45): 160+42.
- [6] 王成, 朱志刚, 张玉侠, 等. 基于用户的协同过滤算法的推荐效率和个性化改进[J]. 小型微型计算机系统, 2016, 37(3): 428-432.
- [7] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2001) Item-Based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th International Conference on World Wide Web*, Hong Kong, 1-5 May 2001, 285-295. <https://doi.org/10.1145/371920.372071>
- [8] Yao, Y.Y. (2013) Granular Computing and Sequential Three-Way Decisions. *International Conference on Rough Sets and Knowledge Technology*, Halifax, 11-14 October 2013, 16-27. https://doi.org/10.1007/978-3-642-41299-8_3
- [9] Zhou, B., Yao, Y.Y. and Luo, J.G. (2014) Cost-Sensitive Three-Ways Email Spam Filtering. *Journal of Intelligent Information Systems*, **42**, 19-45. <https://doi.org/10.1007/s10844-013-0254-7>
- [10] Liu, D., Li, T.R. and Liang, D.C. (2012) Three-Way Government Decision Analysis with Decision-Theoretic Rough Sets. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, **20**, 119-132. <https://doi.org/10.1142/S0218488512400090>
- [11] Yu, H., Zhang, C. and Wang, G.Y. (2015) A Tree-Based Incremental Overlapping Clustering Method Using the Three-Way Decision Theory. *Knowledge-Based Systems*, **91**, 189-203. <https://doi.org/10.1016/j.knosys.2015.05.028>
- [12] 叶晓庆, 刘盾, 梁德翠. 基于协同过滤的三支粒推荐算法研究[J]. 计算机科学, 2018, 45(1): 90-96.
- [13] Zhang, H.R., Min, F. and Shi, B. (2017) Regression-Based Three-Way Recommendation. *Information Sciences*, **378**, 444-461. <https://doi.org/10.1016/j.ins.2016.03.019>
- [14] 郑荔平, 胡敏杰, 杨红和, 林耀进. 基于粗糙集的协同过滤算法研究[J]. 山东大学学报(理学版), 2019, 54(2): 41-50.
- [15] Yao, Y.Y. (2009) Three-Way Decision: An Interpretation of Rules in Rough Set Theory. *International Conference on Rough Sets and Knowledge Technology*, Gold Coast, 14-16 July 2009, 642-649. https://doi.org/10.1007/978-3-642-02962-2_81
- [16] Pawlak, Z. (1982) Rough Sets. *International Journal of Computer and Information Sciences*, **11**, 341-356. <https://doi.org/10.1007/BF01001956>
- [17] 刘盾, 姚一豫, 李天瑞. 三支决策粗糙集[J]. 计算机科学, 2011, 38(1): 246-250.
- [18] Yao, Y.Y. (2010) Three-Way Decisions with Probabilistic Rough Sets. *Information Sciences*, **180**, 341-353. <https://doi.org/10.1016/j.ins.2009.09.021>
- [19] Yao, Y.Y. (2011) The Superiority of Three-Way Decision in Probabilistic Rough Set Models. *Information Sciences*, **181**, 1080-1096. <https://doi.org/10.1016/j.ins.2010.11.019>

- [20] Yao, Y.Y. (2012) Three-Way Decisions Using Rough Sets. In: Peters, G., Lingras, P., Ślęzak, D. and Yao, Y., Eds., *Rough Sets: Selected Methods and Applications in Management and Engineering*, Springer, London, 79-93.
https://doi.org/10.1007/978-1-4471-2760-4_5
- [21] Yao, Y.Y. (2007) Decision-Theoretic Rough Set Models. *International Conference on Rough Sets and Knowledge Technology*, Toronto, 14-16 May 2007, 1-12. https://doi.org/10.1007/978-3-540-72458-2_1
- [22] 王璇璇, 陈宁江, 练林明, 郭芷柔. 基于谱聚类 and 矩阵分解的改进协同过滤推荐算法[J]. 广西大学学报(自然科学版), 2020, 45(2): 313-320.
- [23] Zhou, B., Yao, Y.Y. and Luo, J. (2010) A Three-Way Decision Approach to Email Spam Filtering. *23th Canadian Conference on Artificial Intelligence*, Ottawa, 31 May-2 June 2010, 28-39.
https://doi.org/10.1007/978-3-642-13059-5_6