

doi: 10.12068/j.issn.1005-3026.2019.11.002

一种基于旁信息增强的协同过滤自动编码器模型

刘海博^{1,2}, 冯 时¹, 于 戈¹

(1. 东北大学 计算机科学与工程学院, 辽宁 沈阳 110169; 2. 河北大学 网络空间安全与计算机学院, 河北 保定 071002)

摘 要: 针对用户评分预测不准确的问题,提出了一种基于旁信息(side-information)对用户兴趣进行预测的协同过滤自动编码器推荐模型,给出了模型的设计原理、损失函数以及具体结构.模型使用单隐藏层自动编码器实现,用户评分与旁信息同为模型的输入/输出数据,旁信息也直接参加模型的训练,这种设计不仅降低了模型的规模和复杂度,而且旁信息可以直接对用户兴趣进行修正.同时,通过对训练数据集合的合理划分与扩充,使得训练的网络模型增加了表达能力.在真实数据集上的对比实验表明,本文提出的方法提高了评分预测的准确度,具有一定的实用价值.

关 键 词: 自动编码器;协同过滤;旁信息;推荐系统;评分预测

中图分类号: TP 301.6 **文献标志码:** A **文章编号:** 1005-3026(2019)11-1527-06

Auto-Encoder for Collaborative Filtering Based on Side-Information Enhancement

LIU Hai-bo^{1,2}, FENG Shi¹, YU Ge¹

(1. School of Computer Science & Engineering, Northeastern University, Shenyang 110169, China; 2. School of Cyber Security and Computer, Hebei University, Baoding 071002, China. Corresponding author: LIU Hai-bo, E-mail: liuhaibo@stumail.neu.edu.cn)

Abstract: In view of inaccurate rating prediction of traditional recommendation model, a model based on side-information enhanced collaborative filtering auto-encoder was proposed, and the design principle, loss function and structure of the model were discussed in detail. The model was implemented by a single hidden layer auto-encoder network with ratings and side-information as input/output data. The model was directly trained by ratings and side-information. The proposed framework not only reduces the scale and complexity of the model, but also amends the user's interest with side-information. At the same time, the representation ability of the trained model is increased by reasonable division and expansion of training datasets. The comparisons on real datasets show that the proposed model with side-information can effectively ameliorate the user's interest predicted by the model without side-information, and improve the accuracy of rating prediction, which has a certain practicality.

Key words: auto-encoder; collaborative filtering; side-information; recommender system; rating prediction

推荐系统根据用户的历史行为特征,对用户进行分类,并预测用户未来的兴趣与行为趋势,进而向用户推荐其可能感兴趣的商品或信息,提高用户的使用体验.神经网络是机器学习的一个分支研究领域.它通过单层或者多层次的神经网络学习表达数据的抽象特征,可以解决有监督和无

监督学习任务.伴随着深度神经网络的飞跃式发展,它逐渐取代了传统的基于过滤技术的推荐方法.

在很多实际应用中,系统不仅仅包含像样本评分或标签这样的监督信息,还包含很多附加信息,这些附加信息,称之为旁信息(side-

information)^[1]. 以电影推荐为例, 用户的性别、年龄、职业信息, 电影的类型、年份、演员、内容描述以及海报等都是旁信息. 旁信息可以辅助系统对样本标签或评分数据进行处理, 在推荐技术中得到了比较广泛的使用.

评分预测^[2]问题是推荐系统需要解决的主要问题之一. 系统通过大量用户的历史评分记录, 预测其对某个项目的可能评分. 这里主要讨论如何利用项目的旁信息来增强神经网络中的自动编码器模型对用户评分的预测以及模型规模简化问题.

目前, 科技工作者对基于神经网络的推荐模型的研究已经取得了一些成果, 提出了多种解决方案. 其中, 多层感知机与自动编码器^[3]在基于神经网络模型的推荐系统中应用较多.

文献[4]提出了一种神经协同过滤(neural collaborative filtering, NCF)网络, NCF 是一种基于多层感知机, 并利用用户和项目的旁信息作为用户和项目的潜在特征进行建模的推荐模型. NCF 的预测规则为 $\hat{r}_{ui} = f(\mathbf{U}^T \cdot \mathbf{s}_u^{\text{user}}, \mathbf{V}^T \cdot \mathbf{s}_i^{\text{item}} | \mathbf{U}, \mathbf{V}, \theta)$, $f(\cdot)$ 是多层感知器, θ 是该网络的参数, $\mathbf{s}_u^{\text{user}}$ 和 $\mathbf{s}_i^{\text{item}}$ 分别是用户与项目的旁信息. 损失函数为加权平方误差: $l = \sum_{(u,i) \in O \cup O^-} w_{ui} (r_{ui} - \hat{r}_{ui})$. 文献[5]提出了神经社会协同排序推荐系统, 将 NCF 模型推广到跨领域的社会推荐中. 文献[6]提出了跨域内容增强协同过滤模型, 网络的最后一层使用点积, 混合了基于旁信息的用户与项目特征. 文献[7]提出了一种宽度与深度学习模型, 宽度学习部分是单层感知器, 可以从历史数据中获得直接特征, 通过产生更一般和抽象的特征表示来获得泛化能力, 深度学习部分是多层感知器.

在基于自动编码器的推荐方法中, 提取用户与项目的抽象特征以及填充用户-项目评分矩阵中的缺失值, 是通过自动编码器完成推荐的两个主要手段. 文献[8]提出了基于用户与基于项目的自动编码器推荐模型, 作者通过编码器将用户评分向量映射到抽象低维特征表示空间, 并通过解码器将其恢复到与输入特征空间相同的输出特征空间, 从而预测用户缺失的项目评分. 文献[9]提出了基于协同过滤的自动编码器模型, 并且结合了用户和项目旁信息, 以减轻数据的稀疏性和冷启动问题. 文献[10]提出了协同降噪自动编码器, 使用单隐含层自动编码器构建推荐模型, 在输入层用户偏好向量被格式化为二进制向量, 并且加入高斯噪声. 文献[11]将变分自动编码器推广

到协同过滤中. 文献[12]通过有效的全反馈训练机制训练了深层自动编码器, 提高了评分预测的精度.

上述研究对于神经网络以及旁信息在推荐系统中的应用取得了一定的进展. 但是其研究成果, 大部分都是只使用用户-项目评分信息进行模型训练, 在小部分引入旁信息的方法中, 基本上都是利用单独的深层网络来挖掘旁信息的抽象特征, 进而辅助用户-项目评分矩阵的挖掘. 其方法增加了网络的规模与复杂度, 提高了推荐系统的代价. 针对上述问题, 本文提出了一种基于旁信息对用户兴趣进行修正和预测的协同过滤自动编码器模型. 真实数据集上的对比实验表明, 本文设计的模型降低了计算规模与复杂度, 提高了评分预测的准确度.

1 推荐模型

1.1 传统自动编码器模型

自动编码器是一种数据压缩算法, 是一种从样本中自动学习的自监督学习算法. 模型主要分为编码器与解码器. 其中, 数据的编码部分和解码部分是数据相关的, 编码部分通过神经网络将高维的输入数据映射到低维度的抽象特征空间, 解码部分同样通过神经网络将低维度的抽象特征数据还原为高维度的原始数据. 因此, 自动编码器是有损失的. 其终极目标是实现输出与输入无限接近. 自动编码器可表示为 $\mathbf{h} = f(\mathbf{x}) = \mathbf{W} \cdot \mathbf{x} + \mathbf{b}_x$, $f(\cdot)$ 为编码映射, \mathbf{W} 为编码权重矩阵, \mathbf{x} 为样本原始输入信息向量, \mathbf{b}_x 是编码偏置向量, \mathbf{h} 为输入信息经过编码后得到的抽象特征向量. 解码器可表示为 $\mathbf{y} = g(\mathbf{h}) = \mathbf{W}^T \cdot \mathbf{h} + \mathbf{b}_y$, $g(\cdot)$ 为解码映射, \mathbf{W}^T 为解码权重矩阵, \mathbf{b}_y 为解码偏置向量, \mathbf{y} 为经过解码后得到的重构信息向量. 因此, 其目标函数为

$$L_{\min} = \sum_{m=0}^M l(\mathbf{x}^m, g(f(\mathbf{x}^m))),$$
 \mathbf{x}^m 表示样本集合中第 m 个样本的原始信息, l 表示损失函数. 传统单隐含层自动编码器结构如图 1 所示.

1.2 基于旁信息增强的协同过滤自动编码器模型

传统的基于自动编码器的推荐模型计算复杂度高, 并且只是简单地要求目标输出与输入尽量相同, 模型尽量拟合用户的现有兴趣, 而很少考虑用户的未来兴趣.

本文提出的基于旁信息对用户兴趣进行增强的协同过滤自动编码器模型(auto-encoder model

for collaborative filtering based on side-information enhancement) 将项目的旁信息抽象为隐藏特征, 并与用户的抽象兴趣特征直接融合, 从而增强了自动编码器对用户兴趣的预测能力. 对于自动编码器网络, 网络层次越深, 则特征抽取越抽象. 为了降低模型的规模, 并且增强项目旁信息对用户兴趣的直接修正作用, 模型采用单隐含层神经网络实现. 把数据集中每个用户与项目的交互数据按照时间排序, 并将其划分为若干份. 通过对数据集的合理划分, 使得训练的模型可以反映用户的未来兴趣. 假设用户集合为 $U = \{u_1, \dots, u_M\}$, M 为用户数量, 项目集合为 $I = \{i_1, \dots, i_N\}$, N 为项目数量, 部分可观测的用户-项目评分矩阵为 $R \in \mathbf{R}^{M \times N}$, 每一个用户 u_m 的可观测评分向量为 $r^m = (R_{m,1}, \dots, R_{m,N}) \in \mathbf{R}^N$. 设项目拥有 K 种旁信息, 则所有种类旁信息的集合为 $S = \{S_1, \dots, S_K\}$, 每一类旁信息集合大小不同, 设第 k 类旁信息集合大小为 L_k , 则项目的第 k 类旁信息集合表示为 $S_k = \{s_{k,1}, \dots, s_{k,L_k}\}$. 用户 u_m 的第 k 类旁信息向量表示为 $x_k^m = (w_{m,k,1}, \dots, w_{m,k,L_k}) \in \mathbf{R}^{L_k}$, 对于任意的 $w_{m,k,L_k} = \frac{\sum_{n=1}^N R_{m,n} \times w_{k,L_k}^n}{|I_m|}$, I_m 为用户 u_m 可观测评分的项目集合, 如果 i_n 拥有旁信息 s_{k,L_k} , 则

$$w_{k,L_k}^n = 1, \text{ 否则 } w_{k,L_k}^n = 0.$$

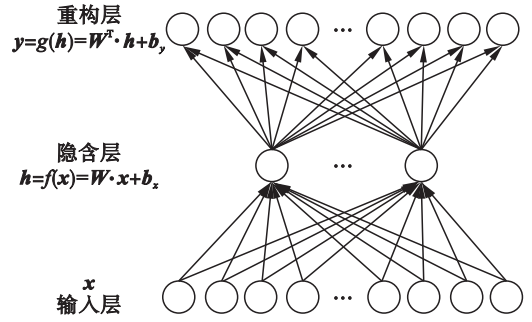


图1 传统单隐含层自动编码器结构图

Fig. 1 Traditional single hidden layer auto-encoder

基于旁信息增强的协同过滤自动编码器模型结构如图2所示. 对于用户 u_m , 用户-项目评分向量 r^m , 以及所有旁信息向量 x_k^m 同时为模型的输入, 通过式(1)完成编码器抽象并在隐藏层融合.

$$h^m = f(r^m, x_1^m, \dots, x_K^m) = W_r^h \cdot r^m + \sum_{k=1}^K W_k^h \cdot x_k^m + b_r^h. \quad (1)$$

其中: h^m 为用户 u_m 的抽象特征表示; W_r^h 为评分编码权重; W_k^h 为旁信息编码权重; b_r^h 为隐藏层偏置项.

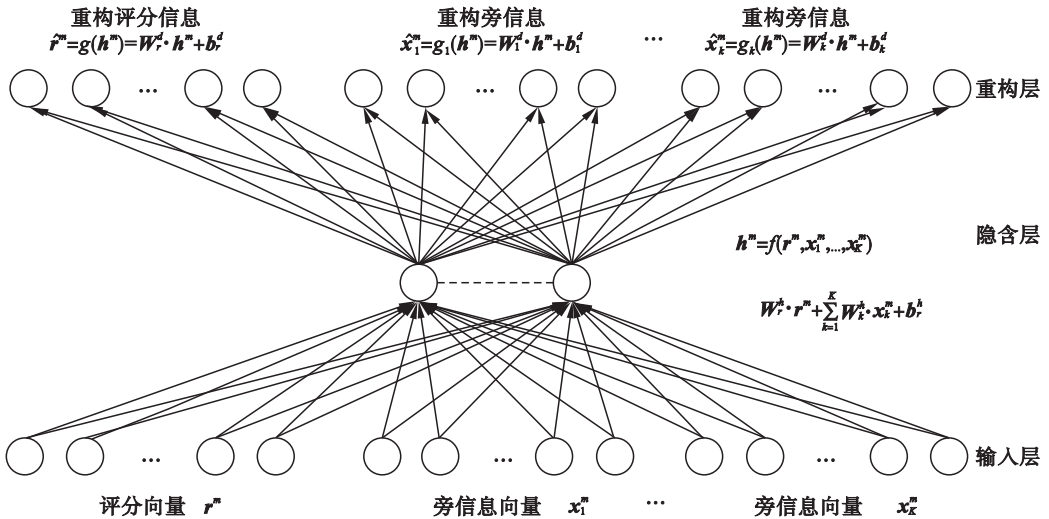


图2 基于旁信息增强的协同过滤自动编码器模型

Fig. 2 Auto-encoder model for collaborative filtering based on side-information enhancement

对于重构层, 分别通过式(2)与式(3)完成对评分信息与旁信息重构.

$$\hat{r}^m = g(h^m) = W_r^d \cdot h^m + b_r^d, \quad (2)$$

$$\hat{x}_k^m = g_k(h^m) = W_k^d \cdot h^m + b_k^d. \quad (3)$$

其中: \hat{r}^m 表示经过解码得到的重构评分信息; W_r^d 为评分解码权重; b_r^d 为评分解码偏置项; \hat{x}_k^m 表示经过解码得到的旁信息; W_k^d 为旁信息解码权重;

b_k^d 为旁信息解码偏置项.

2 损失函数

损失函数用于评价模型的预测值与真实值的不一致程度, 因此, 需要依靠损失函数来选取最优化模型. 采用可观测值的平方误差 (masked

squared error, MSE) 作为损失函数. 与平方误差损失函数不同, 可观测的平方误差只考虑目标集合中用户已评分项目的平方和误差. 损失函数定义为

$$L_{\min} = \sum_{m=1}^M (\| \vec{r}^m - \hat{r}^m \|_o^2 + \sum_{k=1}^K \lambda_k \| \vec{x}_k^m - \hat{x}_k^m \|_o^2) + L_{\text{reg}} . \quad (4)$$

其中: \vec{r}^m 为目标评分; \vec{x}_k^m 为目标旁信息; λ_k 为第 k 类旁信息的损失权重超参数; $\|\cdot\|_o^2$ 表示可观测值的平方和误差. 模型在训练时, 目标重构数据信息与输入数据信息不一定相同(输入数据集合与目标数据集合不一定相等), 因此, r^m 与 \vec{r}^m , x_k^m 与 \vec{x}_k^m 不一定相等. 为了防止模型过拟合, 损失函数中加入正则化项 L_{reg} , 定义为

$$L_{\text{reg}} = \frac{\lambda}{2(K+1)} (\| W_r^h \|_F^2 + \| W_r^d \|_F^2 + \sum_{k=1}^K \lambda_k (\| W_k^h \|_F^2 + \| W_k^d \|_F^2)) . \quad (5)$$

其中: λ 为损失函数的正则化项超参数; $\|\cdot\|_F^2$ 表示矩阵各元素绝对值平方和(即不进行开平方运算的弗洛贝尼乌斯范数).

3 实验及分析

3.1 数据集与数据划分

使用 MovieLens 10M (ML - 10M) 与 MovieLens 20M (ML - 20M) 两个数据集, 进行实验. 每个用户的评分数据按时间顺序排列, 在最后的 20% 数据中, 随机选择 50% 作为验证集, 另外 50% 作为测试集, 前面的 80% 数据作为训练集. 在测试集与验证集中删除了没有在训练集中出现的电影. 为了增强模型对用户未来兴趣的表达能力, 训练集按时间顺序等分 4 份: T_1, T_2, T_3, T_4 . 训练时输入数据集合为 $\{(T_1), (T_1, T_2), (T_1, T_2, T_3), (T_1, T_2, T_3, T_4)\}$, 对应的目标集合为 $\{(T_1), (T_1, T_2), (T_1, T_2, T_3), (T_1, T_2, T_3, T_4)\}$. 即, 采用的是目标集合大于输入集合的增量方式进行训练. 由于电影的旁信息是附加在电影中的, 因此, 对于电影的旁信息也采用同样的划分方式.

旁信息的选择. 影响用户对电影评分的主要因素包含两类: 电影内容以及电影演员阵容(明星效应). 由于电影内容信息的特征提取涉及到文本内容分析, 必须使用单独的网络进行特征提取, 这样增加了模型的复杂度, 因此采用电影分类信息来代替内容信息. 使用词袋法分别对电影分

类信息和演员信息进行处理. 数据集详细信息见表 1.

表 1 数据集信息

Table 1 Datasets information

集合属性	旁信息	ML - 10M	ML - 20M
训练集合	用户	69 878	138 493
	电影	10 352	22 085
	评分	7 972 582	15 945 812
	分类	20	20
	演员	12 290	27 549
验证集合	用户	34 939	69 246
	电影	10 117	19 462
	评分	1 010 185	2 029 698
	分类	20	20
	演员	11 921	23 661
测试集合	用户	34 939	69 246
	电影	10 073	18 685
	评分	1 016 227	2 016 633
	分类	20	20
	演员	11 872	22 931
稀疏度	—	0. 013 8	0. 006 5

3.2 评价指标及实验参数

经过实验得出模型最佳参数, 如表 2 所示.

表 2 模型参数值

Table 2 Model parameter values

参数	值
损失函数正则化项的超参数 λ	0. 1
电影分类损失权重超参数 λ_0	0. 1
电影演员损失权重超参数 λ_1	0. 1
模型学习率	0. 001
神经元丢弃率	0. 5
网络参数优化方法	Adam
神经元激活函数	Relu
隐藏层神经元数量	128

本文采用可观测的均方根误差 (masked root mean squared error, MRMSE) 作为评价方法的性能指标. MRMSE 是一种常用的测量评分之间差异的量度. 这里的 \vec{r}^m 是目标集上的真实值, \hat{r}^m 是预测值, L_M 是目标集合中所有可观测的真实评分数量.

$$\text{MRMSE} = \sqrt{\frac{1}{L_M} \sum_{m=1}^M \| \vec{r}^m - \hat{r}^m \|_o^2} . \quad (6)$$

3.3 实验结果

为了验证本文方法以及旁信息与数据划分对评分预测的影响, 同时设计了 4 个不同的方案进

行对比实验:①AE, 使用原始自动编码器模型, 不使用旁信息与数据划分;②AE - d, 使用本文模型,使用数据划分,但不使用旁信息;③AE - s, 使用本文模型,使用旁信息,但不使用数据划分;④AE - sd,使用本文模型,同时使用旁信息与数据划分. 为了更加直观表达实验结果,展示训练数据集以及其对应的验证数据集上的 MRMSE 值. 实验结果如图 3、图 4 所示.

由图 3a 可以观察到使用原始自动编码器模型的 AE 的实验误差要高于 AE - d, AE - s 与 AE - sd,而且 AE 与 AE - d 比较接近,但是与 AE - s 以及 AE - sd 相差较大. 由图 3b 可以发现,AE 与 AE - d 在验证集合上的误差差距要远高于其在训练集上的误差差距. 由于数据划分方

法以及增量式的训练模式,提高了模型对用户未来兴趣的预测能力. AE - s 的误差也远小于 AE, 这是因为,项目的旁信息可以映射用户的兴趣特征,因此,旁信息的使用有效地修正了仅仅使用评分信息进行训练得到的用户兴趣特征. 此外,还可以发现,在训练集上 AE - sd 的性能指标低于 AE - s 的性能指标,而在验证集上 AE - sd 的性能指标略高于 AE - s 的性能指标. 这是因为,对于旁信息,使用了同样的数据划分,这样增强了旁信息对用户未来兴趣的预测能力. 由图 3b 还可以发现,在模型的预测能力的平稳性上,AE - sd 最好,AE - s 次之,AE - d 的波动比较大. 这是因为,旁信息对用户的兴趣进行修正的同时也增强了模型的稳定性.

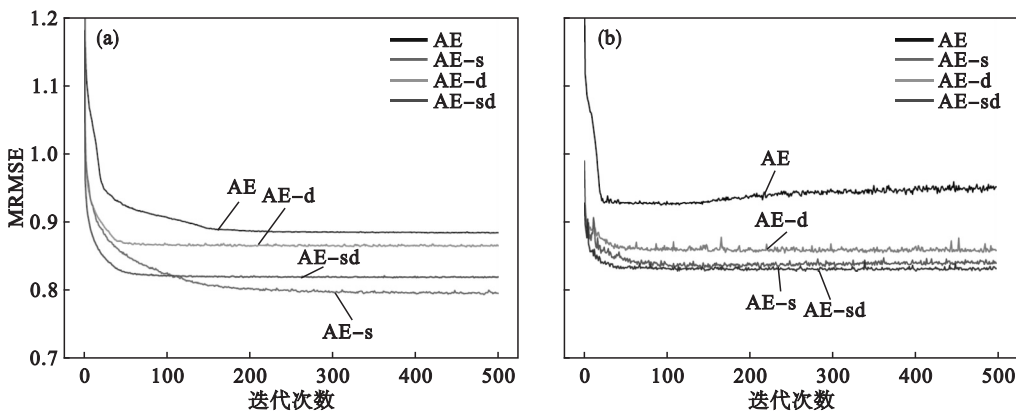


图 3 ML - 10M 数据集实验结果
Fig. 3 Experimental results of ML-10M dataset
(a)—ML - 20M 训练结果; (b)—ML - 20M 估计结果.

图 4 描述了 4 个方法在数据集 ML - 20M 上的实验结果. 实验结果与其在数据集 ML - 10M 上的结果类似. 但是,需要注意的是在 ML - 20M 上的训练结果可以看出 AE 与 AE - d 的性能很接近,它们的实验差距要小于其在 ML - 10M 训练集上的实验差距. 同样,AE 与 AE - d 在 ML -

20M 验证集上的实验结果的差距也小于在 ML - 10M 验证集上的差距. 这主要是因为 ML - 20M 要比 ML - 10M 稀疏. 稀疏的数据减弱了数据划分的作用. 但是,总体上看,训练数据集的划分都起到了提高评分预测准确度的作用.

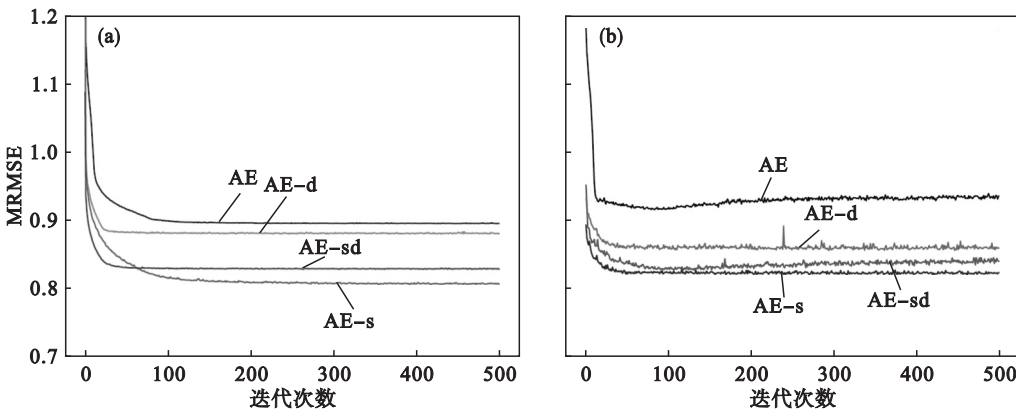


图 4 ML - 20M 数据集实验结果
Fig. 4 Experimental results of ML-20M dataset
(a)—ML - 20M 训练结果; (b)—ML - 20M 估计结果.

在测试集上的实验结果如表 3 所示. 可以发现 AE - sd 取得了较好的评分预测结果.

表 3 测试集实验结果
Table 3 Testing results

数据集	AE	AE - d	AE - s	AE - sd
ML - 10M	0. 921 8	0. 852 2	0. 829 2	0. 824 7
ML - 20M	0. 917 2	0. 856 2	0. 826 9	0. 819 6

4 结 语

本文针对如何利用旁信息增强自动编码器推荐模型的预测能力进行了研究. 首先基于用户或者项目的旁信息可以映射用户兴趣的假设, 设计了基于旁信息增强的协同过滤自动编码器推荐模型. 为了使旁信息可以直接修正用户的兴趣, 使用单隐含层自动编码器网络实现推荐模型, 并且旁信息与评分信息共享中间隐含层. 同时, 将训练数据集按时间等分, 训练时采用目标集合大于输入集合的增量训练的方法. 实验表明本文模型可以降低模型的复杂度, 并且提高评分预测的准确度.

参考文献:

[1] Elhamifar E, Wang Y. High rank matrix completion with side information [C]// AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018 :4252 - 4259.
[2] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. *IEEE Transactions on Knowledge*

and Data Engineering, 2005, 17(6):734 - 749.
[3] Zhang S, Yao L, Sun A, et al. Deep learning based recommender system: a survey and new perspectives [J]. *ACM Computing Surveys*, 2018, 1(1):1 - 35.
[4] He X, Liao L, Zhang H, et al. Neural collaborative filtering [C]//International Conference on World Wide Web. Perth: ACM, 2017:173 - 182.
[5] Wang X, He X, Nie L, et al. Itemsilk road: recommending items from information domains to social users [C]// International ACM SIGIR Conference on Research and Development in Information Retrieval. Tokyo: ACM, 2017: 185 - 194.
[6] Lian J, Zhang F, Xie X, et al. CCCF net: a content-boosted collaborative filtering neural network for cross domain recommender systems [C]//International Conference on World Wide Web Companion. Perth: ACM, 2017:817 - 818.
[7] Cheng H T, Koc L, Harmsen J, et al. Wide & deep learning for recommender systems [C]// The Workshop on Deep Learning for Recommender Systems. Boston: ACM, 2016: 7 - 10.
[8] Sedhain S, Menon A K, Sanner S, et al. AutoRec: autoencoders meet collaborative filtering [C]// International Conference on World Wide Web. Florence: ACM, 2015: 111 - 112.
[9] Strub F, Mary J. Collaborative filtering with stacked denoising autoencoders and sparse inputs [C]// Neural Information Processing Systems. Montreal: NIPS, 2015: 136 - 144.
[10] Wu Y, Dubois C, Zheng A X, et al. Collaborative denoising auto-encoders for top-*n* recommender systems [C]// International Conference on Web Search and Data Mining. San Francisco: ACM, 2016:153 - 162.
[11] Li X, She J. Collaborative variational autoencoder for recommender systems [C]// ACM SIGKDD International Conference. Halifax: ACM, 2017:305 - 314.
[12] Kuchaiev O, Ginsburg B. Training deep autoencoders for collaborative filtering [J]. *arXiv: Machine Learning*, 2017, 1(1):1 - 5.