

基于自编码机和聚类的混合推荐算法

赵阳^{1, 2}, 朱全银^{1, 2}, 胡荣林², 瞿学新²

(1 南京工业大学 计算机科学与技术学院, 江苏 南京 211800; 2 淮阴工学院 计算机与软件工程系, 江苏 淮安 223003)

摘要: 针对传统的协同过滤推荐算法在稀疏数据集上表现不佳的情况, 提出了一种将自编码器(Autoencoder)和聚类结合的混合推荐算法. 首先将用户项目评分数据和用户人口统计学数据作为自编码器的输入, 提取用户特征. 然后利用提取到的用户特征对用户聚类得到用户类别, 从而使近邻搜索范围减小. 接着通过计算平均绝对误差(MAE)寻找到适用于同一类别用户的推荐算法, 最后组合各类别上适用的推荐算法, 得到混合推荐模型. 推荐阶段, 计算目标用户类别, 并使用混合推荐模型得到推荐结果. 实验结果表明, 该算法可以有效的提高推荐质量.

关键词: 混合推荐; 协同过滤; 自编码器; 聚类; 平均绝对误差

Hybrid Recommendation Algorithm Based on Autoencoder and Clustering

ZHAO Yang^{1,2}, ZHU Quan-Yin^{1,2}, HU Rong-Lin², QU Xue-Xin²

(1 School of Computer Science and Technology, Nanjing Tech University, Nanjing 211800, China; 2 Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223001, China)

Abstract: With the problem that traditional collaborative filtering recommendation algorithm does not recommend high quality on sparse data sets, a hybrid recommendation algorithm that based on autoencoder and clustering is proposed. First, the user's project rating data and user demographic data as the input data from the autoencoder to extract user characteristics. Then use the extracted user characteristics to cluster the users to obtain the user categories, so that the search range of the neighbor decreases. Then search for a recommendation algorithm for the same type of users by calculating mean absolute error (MAE), and get a hybrid recommendation model. Finally calculate the target user category and use the hybrid recommendation model to get the recommended results. Experimental results show that this algorithm can effectively improve the quality of recommendation.

Key words: hybrid recommender; collaborative filtering; autoencoder; clustering; mean absolute error

作者简介:

赵阳男, (1992-), 硕士研究生. 研究方向为数据挖掘、推荐系统.

朱全银 (通讯作者) 男, (1966-), 教授, 硕士生导师. 研究方向为智能信息处理、接口与通信、数据挖掘等. E-mail: hyitzqy@126.com.

胡荣林男, (1978-), 博士, 副教授, 硕士生导师. 研究方向为物联网应用技术.

瞿学新男, (1992-), 硕士研究生. 研究方向为数据挖掘、机器学习.