

文章编号: 1007-5321(2009)06-0042-05

基于 BP 神经网络的协作过滤推荐算法

张 磊, 陈俊亮, 孟祥武, 沈筱彦, 段 铠
(北京邮电大学 网络与交换技术国家重点实验室, 北京 100876)

摘要: 研究、探讨了协同推荐问题, 提出了一种基于两层面的多个后向传播(BP)神经网络的协作过滤推荐算法(TMNN-CFRA)。两层面的多个 BP 神经网络协同工作, 高层面 BP 网反向误差传播直至低层面多个人工神经网络(ANN)进行网络权值修正, 以此为基础, 借助用户评价等特征前向给出项目推荐。标准评测集 MovieLens 上的实验评测表明了 TMNN-CFRA 的可行性和有效性。
关 键 词: BP 神经网络; 项目推荐; 协作过滤
中图分类号: TP301 **文献标志码:** A

BP Neural Networks-Based Collaborative Filtering Recommendation Algorithm

ZHANG Lei, CHEN Jun-liang, MENG Xiang-wu, SHEN Xiao-yan, DUAN Kun
(State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications Beijing 100876, China)

Abstract: A novel two-level multiple neural networks-based collaborative filtering recommendation algorithm (TMNN-CFA) for rating prediction is presented. By cooperating the multiple back propagation (BP) networks together, the higher layer neural network propagates conversely the output deviation until to the lower layer neural networks to amend the network weights and based on which, item recommendation is accomplished in the forward process of two layers networks relying on the factors such as ratings, etc. Experiment results on the standard MovieLens show that TMNN-CFA method is effective and feasible for item recommendation.
Key words: back propagation neural networks; item recommendation; collaborative filtering

推荐系统能有效解决信息过载问题。基于内容的推荐^[1]和协作过滤式推荐(CF, collaborative filtering)^[2-3]作为 2 种主要的推荐技术, 被 Netflix 和 MovieLens 等商业、科研网站广泛采用。在非“冷启动”情况下, CF 方法作为最具应用前景的推荐技术, 比基于内容的推荐具有更高的推荐精度。笔者工作的基本出发点是: 在项目评价相对充足的情况下, 在基本协同的基础上, 探讨新的有效推荐的研究思路。因此, 提出了一种新的 TMNN-CFRA, 采用两层面

多个并行人工神经网络(ANN)完成项目推荐的全过程。两层面的多个 BP 网络协同工作, 高层面 BP 网反向误差传播直至低层面多 ANN 进行网络权值修正, 以此为基础借助用户评价等特征前向给出项目推荐。

1 数据预处理

记项目内容属性为 $\langle \alpha_1, \alpha_2, \dots, \alpha_m \rangle$, 协作过滤评分要素为 β_1 , 人口统计学要素(包括性别、年龄

收稿日期: 2009-04-07
基金项目: 国家高技术研究发展计划项目(2007CB307100); 国家自然科学基金项目(60872051); 国家科技支撑计划重大项目(2006BAH02A11); 北京市教委产学研项目(zh100130525)
作者简介: 张 磊(1980—), 男, 博士生, E-mail: zlei565@gmail.com; 陈俊亮(1933—), 男, 教授, 博士生导师, 中国科学院院士, 中国工程院院士。
©1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

等)为 $\langle \gamma_1, \gamma_2, \dots, \gamma_n \rangle$, 记项目集和用户集分别为 $\mathcal{V} = \{ \vartheta_1, \vartheta_2, \dots \}$, $U = \{ u_1, u_2, \dots \}$ 则 $\forall \vartheta_k \in \mathcal{V}, u_i \in U, u_j \in U, u_i$ 和 u_j 相对于 ϑ_k 的相似度向量为 $\vec{S}_{ij,k} = \langle s_{\alpha_1}^{ij,k}, \dots, s_{\alpha_m}^{ij,k}, s_{\beta_1}^{ij,k}, s_{\beta_2}^{ij,k}, \dots, s_{\beta_n}^{ij,k} \rangle$.

Winnow 算法^[4]对文本分类能产生良好的效果, 综合 Rocchio^[5]和 Winnow 算法, 采用迭代修正方式有效挖掘项目特征偏好; 采用余弦向量方法基于用户项目特征喜好度进行用户的内容属性(喜好)相似度运算, 得到任意 2 个用户针对任一特定项目内容属性 $\alpha_l \in \{ \alpha_1, \alpha_2, \dots, \alpha_m \}$ 的喜好相似度; 采用 Pearson 方法^[6]获得最近邻居; 基于不同要素间相对距离计算项目评分要素及年龄要素之间相似度, 以上述相似度运算为基础求得 $\vec{S}_{ij,k}$ 作为 ANN 输入.

2 基于 BP 网络的协作过滤推荐算法

设目标用户为 u_i , u_i 的最近邻居集为 $U_i = \{ u_1^i, u_2^i, \dots, u_\tau^i \}$, τ 为预设邻居规模. 基于协作过滤方法的评价过程, TMNN-CFRA 神经网络拓扑如图 1 所示. 采用 $\tau+1$ 个神经网络, 划分 2 个层面分别表征相似度和评价预测, 并关联低、高层面为一个整体以完成反向误差和前向评价的全过程. 设目标用户 u_i 给出的所有项目评价所对应的项目集为 $\theta^i = \{ \vartheta_1^i, \vartheta_2^i, \dots, \vartheta_t^i \}$, t 为评价规模. 从 θ^i 中随机选择 M 个项目 (M 个项目构成的集合记为 θ_M^i) 连同相应的项目评价构成用户 u_i 规模为 M 的训练集. 设第 1 层面的全拓扑 BP 神经网络集为 $\xi_1 = \{ \zeta_1^1, \dots, \zeta_\tau^1 \}$, 第 2 层面的全拓扑 BP 网络集为 $\xi_2 = \{ \zeta_1^2 \}$.

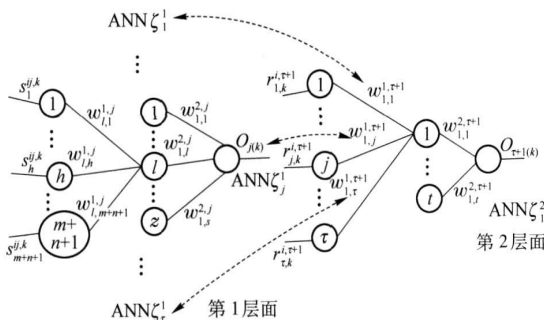


图 1 神经网络拓扑

$\forall \vartheta_k^i \in \theta_M^i, \zeta_j^1 \in \xi_1, s_{ij,k}^{ij,k} \in \vec{S}_{ij,k} = \langle s_{\alpha_1}^{ij,k}, \dots, s_{\alpha_m}^{ij,k}, s_{\beta_1}^{ij,k}, s_{\beta_2}^{ij,k}, \dots, s_{\beta_n}^{ij,k} \rangle = \langle s_{\alpha_1}^{ij,k}, \dots, s_{\alpha_m}^{ij,k}, s_{\beta_1}^{ij,k}, \dots, s_{\beta_n}^{ij,k} \rangle$ 为 ζ_j^1 的第 h 个输入, 由前向传播得到 u_i 和 $u_j^i \in U_i$ 关于项目 ϑ_k^i 的相似度 $s_{ij,k}$, ζ_j^1 输出为

$$O_{j(k)} =$$

$$f \left(\sum_{l=1}^z w_{l,j}^{l,j} g \left(\sum_{h=1}^{m+n+1} w_{l,h}^{l,h} s_h^{ij,k} + b_{l,j}^{l,j} \right) + b_{l,j}^{l,j} \right) = s_{ij,k} \quad (1)$$

其中, z 为隐藏层节点数; 权重 w 和偏移 b 的上标(左)表示节点层次, 设隐藏层、输出层的节点层次分别为 1 和 2, 上标(右)表示第 j 个网络, 下标(左)表示隐藏层/输出层的相应节点, $w_{l,h}^{l,h}$ 和 $w_{l,j}^{l,j}$ 下标(右)分别表示相应节点的第 h 和第 l 个输入; $f(\cdot)$ 和 $g(\cdot)$ 为传输函数. 为了清晰地阐释算法, 式(1)省略了 w, b 等符号下标中的 i 标号, 式(2)~式(4)同样如此.

下面给出高层面神经网络的评价预测过程. 设 ζ_1^2 的输入节点数为 τ , 用户 u_j^i 对项目 ϑ_k^i 的评价为 $r_{j,k}^{i,\tau+1}$, 即第 $\tau+1$ 个网络的第 j 个输入. 设 $w_{l,j}^{l,j,\tau+1}$ 是 ζ_1^2 隐藏层第 1 个节点的第 j 个输入的权重, 对于项目 ϑ_k^i , 令 $w_{l,j}^{l,j,\tau+1} = O_{j(k)}$, 关联第 1 层面各输出至高层面网络 ζ_1^2 , 得到推荐预测为

$$O_{\tau+1(k)} = f \left(w_{1,1}^{2,1,\tau+1} g \left(\sum_{j=1}^{\tau} O_{j(k)} r_{j,k}^{i,\tau+1} + b_{1,1}^{2,1,\tau+1} \right) + \sum_{e=2}^t w_{e,1}^{2,e,\tau+1} g \left(\sum_{j=1}^{\tau} w_{e,j}^{1,j,\tau+1} r_{j,k}^{i,\tau+1} + b_{e,1}^{1,\tau+1} \right) + b_{1,1}^{2,\tau+1} \right) \quad (2)$$

其中 t 为隐藏层节点个数. 记 u_i 对 ϑ_k^i 的评价为 $r_{i,k}$, 传输函数 $f(\cdot)$ 和 $g(\cdot)$ 取线性函数和 sigmoid 函数, 由式(1)和式(2), 使用梯度下降方法, 得到网络的后向传播修正规则, 如式(3)和式(4)所示.

$$\Delta w_{l,j}^{1,j,\tau+1} = -\eta \frac{\partial F}{\partial w_{l,j}^{1,j,\tau+1}} = \eta \left(r_{i,k} - \left(\sum_{e=1}^t w_{e,1}^{2,e,\tau+1} g \left(\sum_{j=1}^{\tau} w_{e,j}^{1,j,\tau+1} r_{j,k}^{i,\tau+1} + b_{e,1}^{1,\tau+1} \right) + b_{1,1}^{2,\tau+1} \right) (w_{1,1}^{2,1,\tau+1} g' r_{j,k}^{i,\tau+1}) \right) \quad (3)$$

令 $\Delta O_{j(k)} = \Delta w_{l,j}^{1,j,\tau+1} (\eta = 1)$, 得到

$$\Delta w_{l,h}^{1,h} = -\eta \frac{\partial \left(\frac{1}{2} (\Delta O_{j(k)})^2 \right)}{\partial w_{l,h}^{1,h}} =$$

$$\eta (\Delta O_{j(k)}) \frac{\partial O_{j(k)}}{\partial w_{l,h}^{1,h}} w_{l,1}^{2,1,\tau+1} (r_{j,k}^{i,\tau+1})^2 \left((g')^2 w_{1,1}^{2,1,\tau+1} + \left(\sum_{e=1}^t w_{e,1}^{2,e,\tau+1} g \left(\sum_{j=1}^{\tau} w_{e,j}^{1,j,\tau+1} r_{j,k}^{i,\tau+1} + b_{e,1}^{1,\tau+1} \right) + b_{1,1}^{2,\tau+1} \right) - r_{i,k} \right) \times$$

$$\left\{ \frac{-\exp\left[-\left(\sum_{j=1}^{\tau} w_{l,j}^{1,\tau+1} r_{j,k}^{i,\tau+1} + b_l^{1,\tau+1}\right)\right]}{1 + \exp\left[-\left(\sum_{j=1}^{\tau} w_{l,j}^{1,\tau+1} r_{j,k}^{i,\tau+1} + b_l^{1,\tau+1}\right)\right]} \right\}^2 + \left\{ \frac{2\exp\left[-\left(\sum_{j=1}^{\tau} w_{l,j}^{1,\tau+1} r_{j,k}^{i,\tau+1} + b_l^{1,\tau+1}\right)\right]}{1 + \exp\left[-\left(\sum_{j=1}^{\tau} w_{l,j}^{1,\tau+1} r_{j,k}^{i,\tau+1} + b_l^{1,\tau+1}\right)\right]} \right\}^3 \quad (4)$$

3 实验评测

3.1 数据集和试验过程

实验部分采用推荐系统广泛使用的 MovieLens^[7] 标准数据集作为评测数据集, 其中每部电影的评价规模为 1~5. 另外, 采用 IMDb, The Internet Movie Database 网站 <http://www.imdb.com> 获取数据集中任意电影的导演、类型等资源特征以及推荐系统广泛使用的 mean absolute error (MAE)^[7] 维度作为评测指标. 设 $\langle p_{\vartheta_k}, q_{\vartheta_k} \rangle$ 是项目 ϑ_k 的评价预测对, MAE 维度采用 $|p_{\vartheta_k} - q_{\vartheta_k}|$ 表征绝对误差. 针对任一用户, 若评价-预测对个数为 N , 则 MAE 为 $\sum_{e=1}^N |p_{\vartheta_e} - q_{\vartheta_e}| / N$.

实验采用交叉验证方式, 从 943 个用户的数据集中完全随机地划分出若干互不相交的子集/切片, 每次选取 1 个子集作为测试用户集, 其余用户为训练用户集. 考虑到实际工作量 (从网页信息预处理到神经网络生成等基本功能编码实现耗时 5 个月, 代码过程复用等未做优化, 5 个切片的实验结果生成耗时 21 d), 在确保完全随机的前提下生成 5 个子集/切片后完成实验. 具体而言, 每次选取 1 个切片用作测试用户集, 整个数据集中剩下的 $943 - 50 = 893$ 个用户即为本切片对应的训练用户集 (历史数据), 并用于从中提取目标用户的最相似邻居等. 对于每个切片独立计算 MAE 后将各切片 MAE 算术平均, 得到最终 MAE 性能值. 对于指定的任一切片, 其中的用户即推荐系统目标用户; 对于任一目标用户 u_i , 从其项目评价集中随机选择 M 个评价用作训练, 其余数据作测试集. 针对目标用户某个邻居未给出训练集某部电影的评价等情形, 使用经典 CF 值给出评价值. 评测过程引入 α, β 2 个因子. 对于任一目标用户 u_i , α 表示 u_i 相似邻居的尺寸; β 为 u_i 任一邻居评价集对应的电影集与 u_i 训练集对

应的电影集的交集中的电影数需要满足的最小值, 在实验中令 β 取值不小于 10.

3.2 实验结果

TMNN-CFRA 的 MAE 如图 2 所示, 横坐标为目标用户的已知电影评价数 (M); 试验评测给出参数 α 取值为 5、10、20, β 由 10 变为 20 时的 6 条性能曲线可以刻画 MAE 误差水平. 随着训练样本的增多, 6 条曲线的 MAE 误差值均具有不断下降的趋势, M 值的增大使得 ANN 中的训练信息逐步增大, 故网络可以更好地学习拟合以趋近于目标用户的真实评价. 取定 α, β 的变化对系统没有明显影响, 就现有的评价集规模, β 取值不应太大. ANN 的复杂性使系统对训练样本规模具有一定要求, 固定 β 而增大 α, M 值较小时不同曲线的 MAE 误差差异相对明显; M 值较大时误差间距相对较小. 这一变化特征同样被认为是受制于 ANN 自身特性. 按图 1 所示的 ANN 拓扑, 邻居尺寸 α 直接决定整个网络的规模, 在 M 值较小时, 过于复杂的网络反而有不易训练的特性, 故系统误差值维持在相对较高的水平. 随着训练样本的增加, 网络对邻居规模的承载能力加强, 故 M 值较大时 α 的变化对系统的影响相对趋缓. 从实验结果可见, 基于现有评测集中的评价数和 ANN 对于训练样本数的要求, α 取值不易过大 (见图 2), 当 $\alpha = 5, \beta = 20$ 时, 系统获得最佳性能.

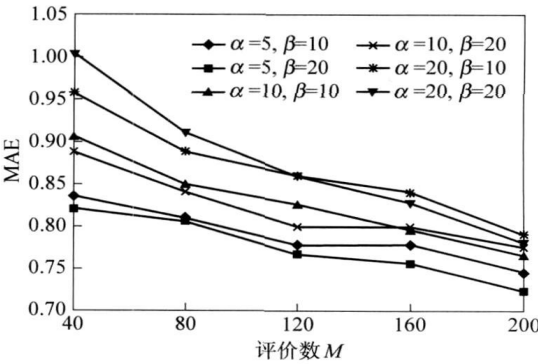


图 2 TMNN-CFRA 算法的 MAE 值

图 3 给出 TMNN-CFRA 方法同经典的基于用户 CF 方法的 MAE 对比评测. 就现有评测集而言, CF 具有比 TMNN-CFRA 更高的性能水平. 经典的基于内容的预测和协同预测是项目推荐中最成功的推荐策略. 一般在样本数达到 100 以后, 经典的基于内容分析预测方法的 MAE 误差控制在 0.95~0.9, TMNN-CFRA 的 MAE 较基于内容的分析策略具有相对明显的优势, 但性能略不及经典的 CF

方法. 如果将 TMNN-CFRA 和 CF 在 $M=100, 120, 140, 160, 180$ 和 200 时各自的 MAE 误差进行算术平均, TMNN-CFRA 的平均误差将比 CF 高出 0.016 的水平.

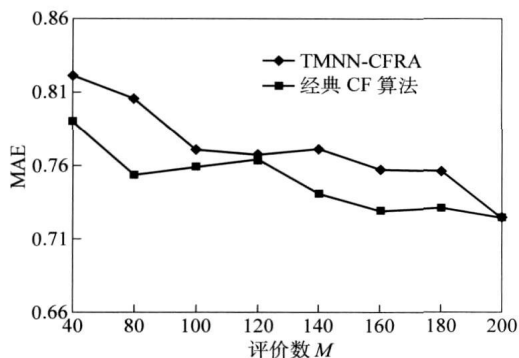


图3 TMNN-CFRA 同经典 CF 算法的 MAE 对比

如果邻居 u_j 同目标用户 u_i 的相似度非常高, 但该邻居并未给出项目 ϑ_r 的评价, 则经典 CF 算法将不选择该邻居参与目标用户 u_i 的针对 ϑ_r 的评价预测; CF 的策略是选择同目标用户的相似度略低于 u_j 但评价过 ϑ_r 的用户 u_k 参与 ϑ_r 的评价预测, CF 算法不存在下面的“模拟值”问题. 在 TMNN-CFRA 中, 针对目标用户 u_i , 同样首先得到同其最相似的若干个邻居, 这些邻居将被关联至目标用户 u_i 的神经网络. 针对项目 ϑ_r , 如果加入到 ANN 中的某个最近邻 u_j 未给出针对该项目的评分, TMNN-CFRA 网络本身需要使用邻居 u_j 的针对项目 ϑ_r 的评分预测来模拟 u_j 针对项目 ϑ_r 的真实评分, 大量模拟值的引入为系统带来了很大的“干扰”, 这是 TMNN-CFRA 的精度水平(略)低于经典 CF 算法的主要原因. 在 Movielens 评测集中, 平均每个用户给出评价的电影数占其总数的不足 10% ($100\,000 / (943 \times 1682) = 6.3\%$). 换言之, 平均到任一电影 ϑ_r , TMNN-CFRA 使用的 90% 以上的邻居评价都是非真实的, 相对于借助相似邻居对 ϑ_r 的真实评价来预测目标用户 u_i 对 ϑ_r 的评价, 如果使用的数据源并不是最近邻居的真实评价, 而是采用协同(过滤)算法得到的邻居的评价预测——预测误差的绝对值一般在 0.7 以上, 即大量的模拟值需要被引入至 TMNN-CFRA 的工作过程以给出目标用户 u_i 针对项目 ϑ_r 的评价预测, 模拟值带来的“误差”将直接影响 ANN 训练等的有效性, 从而导致推荐精度出现比较大的下滑.

根据经典 CF 算法推荐预测的整个过程, 在 TMNN-CFRA 的“模拟值”大量减少的环境中, CF 算法性能将基本保持不变(CF 不受模拟值影响); 就 TMNN-CFRA 而言, “模拟值”是影响其性能比较重要的部分, 如果撇开用户实际观看电影和时耗复杂度等因素, 可以满足 TMNN-CFRA 完全理想的“工作环境”——网络的训练等阶段全部使用真实值, 此时由于模拟值的大量减少 TMNN-CFRA 的性能将获得有效提升. 当然, 必须考虑实际应用层面的限制. 为了在相对合理的应用中充分发挥 TMNN-CFRA 的效能, 应对实际问题追加约束.

针对 $1\,682$ 部电影, Movielens 只需用户给出部分电影的评价即可, 称该方式为评价方式 1; 在评价方式 2 中, 将电影集分为 2 个互不相交的子集 S_1 和 S_2 . 具体而言, 从 $1\,682$ 部电影中选择一般用户比较熟悉的 m 部电影构成 S_1 , 用户评价时尽可能多地针对这些电影进行打分; 剩下的 $1\,682 - m$ 部电影——对应于 S_2 , 用户可随机选择进行打分. 撇开 TMNN-CFRA, 这样做具有非常直观的实际意义. 用户 u_i 和 u_j 针对较多的相同电影进行评价, 2 个用户的“异同”可以比较充分地表现出来, 若 u_i 和 u_j 评价过的电影中相同的电影很少, 此时用户 u_i 和 u_j 的“异同”很难通过评价表现出来, 因为他们评价的电影本身就是不同的. 换言之, 同样是针对电影进行打分, 采用评价方式 2, 用户给出相同的评价数目, 但其中的信息却是“更有意义”的.

从 TMNN-CFRA 的角度来看, 针对 S_1 中电影的评分是 ANN 训练时较基础且重要的样本数据, 用户较多地针对 S_1 中电影进行评价可以非常有效地减少训练时模拟值的出现(而对于 Movielens 评测集, 在网络训练等工作过程中, TMNN-CFRA 需要引入高达 90% 以上的模拟值), 从而可以更为有效地完成 ANN 训练. 将目标用户 u_i 未评价过的电影分为 m 个子集, 如果系统用户足够多(这里需要这个假设, Movielens 的设计初衷是给出 1 个小型的基本试验集, 其中包含 943 个用户的评价信息, 这样的用户规模是相对较为有限的; 而真实商用系统的用户规模可以达到百万级甚至更多, 是 Movielens 的千倍以上), 则针对每个子集而言, 都会有若干个同 u_i 比较相似的用户, 这些用户评价过该子集中较多的电影, 而且这些用户针对 S_1 中 u_i 评价过的

多数、甚至全部电影给出了评价^{*}。相应地,在 TMNN-CFRA 的训练阶段,网络中相似邻居模拟评价的减少可有效提升 ANN 训练的效果。以此为基础,在使用 TMNN-CFRA 针对目标用户进行的预测阶段,良好的网络训练质量和相似邻居模拟评价的减少使得 ANN 在更多的“真实值”环境下更为可靠、有效地进行项目的评价预测。以上是 TMNN-CFRA 比较有效而且符合实际的应用场景,现有的 Movielens 评测集已不可能支持上述应用。TMNN-CFRA 在模拟值较少情况下真实性能指标的获得需要专门评价集的支持,这是今后的工作方向之一。实际上, TMNN-CFRA 从编码实现到在 Movielens 评测集上产生基本实验结果已经历半年多的时间,其间的网络预处理、神经网络实现等工作量非常大。

实验采用 1 台双核 PC 服务器, 2.19 GHz, 2 GB 内存, 7 200 RPM, 网络编码实现于 VS 平台之上。通过目标用户的 ANN 迭代 1 000 次可直观表征时耗(见表 1)。网络时耗具有较好的线性特性,系统效率可很好地满足实际需求,因为整个训练过程可离线完成。另外, TMNN-CFRA 低层面的 τ 个网络满足并行,如果将低、高层面的 $\tau+1$ 个网络分布式部署在 $\tau+1$ 个服务器上,并忽略通信费用等,系统平均训练时耗将仅为单机串行执行时耗的 $\frac{1}{\tau+1}$, 系统效率将极大提升。

表 1 神经网络训练时耗 ms

训练样本数 (目标用户)	网络规模(低层面)		
	5	10	20
40	1 402	2 271	5 473
80	2 818	5 542	10 885
160	5 599	11 073	21 766

4 结束语

研究、探讨了协作推荐问题,提出了 TMNN-CFRA。实验集上的评测验证了 TMNN-CFRA 的有效和可行性。今后将在力所能及的范围内构建出满足需求的中文电影推荐系统,以此为基础更全面、有效地给出 TMNN-CFRA 的性能指标;另外,将 TMNN-CFRA 部署到多服务器群构成的分布式系统,充分发挥其“天然的”并行优势,进一步提升系统的运算执行效率。

参考文献:

[1] Lang K. Newsweeder: learning to filter netnews[C]//

Proceedings of the 12nd Int’l Conference on Machine Learning. [S. l.]: Morgan Kaufmann, 1995: 331-339.

[2] Nakamura A, Abe N. Collaborative filtering using weighted majority prediction algorithm[C]//Proceedings of the 15th Int’l Conference on Machine Learning. [S. l.]: Morgan Kaufmann, 1998: 395-403.

[3] Breese J S, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[C]//Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. [S. l.]: Morgan Kaufmann, 1998: 43-52.

[4] Blum A, Hellerstein L, Littlestone N. Learning in the presence of finitely or infinitely many irrelevant attributes [J]. Journal of Computer and System Sciences, 1995, 50 (1): 32-40.

[5] Rocchio J J. “Relevance feedback in information retrieval,” SMART retrieval system-experiments in automatic document processing[M]. [S. l.]: Prentice Hall, 1971. 313-323.

[6] Resnick P, Iacovou N, Suchak M, et al. GroupLens: an open architecture for collaborative filtering of netnews[C]//Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work. [S. l.]: Chapel Hill, 1994. 175-186.

[7] GroupLens Research Group. GroupLens [EB/OL]. [2008-10-09]. <http://movielens.umn.edu>.

^{*} 实际上,在评价方式 2 和足够的用户规模下(如果是百万用户级商用推荐系统,其用户数将是 Movielens 评测集的千倍以上,当然实际未必需要这样的用户规模),针对任一子集,满足上述条件的同 u_i (比较)相似的用户是相对容易得到的,基于此,影响 TMNN-CFRA 性能的关键的模拟值问题也就基本被消除(至少很大程度上被削弱了)。评价方式 2 使得用户评价的项目相对集中,商用系统的规模可以确保系统中用户数充足,据此,在模拟值大幅减少(Movielens 模拟值达到 90% 以上)的情况下, TMNN-CFRA 可以更有效地完成训练(而有效的训练是 TMNN-CFRA 整个工作过程中最基础的部分;有效的评价预测离不开 TMNN-CFRA 的有效训练),并更有效地进行评价预测。