ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 33,No. 3,May 2018,pp. 496-503
DOI:10.16337/j.1004-9037.2018.03.013
© 2018 by Journal of Data Acquisition and Processing
http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:+86-025-84892742

基于多核学习的协同滤波算法

宋恺涛　彭甫镕　陆建峰

(南京理工大学计算机科学与工程学院，南京，210094)

摘　要：协同滤波是当前推荐系统中一种主流的个性化推荐算法，通过近似用户对商品的评价进行推荐。核函数是解决非线性模式问题的一种方法。协同滤波通常会选用不同的核函数来分析用户之间的影响关系。由于单核函数无法适应于复杂多变场景。因此，结合多个核函数成为一种解决方法。多核学习能够针对场景来组合各个核函数以获取更好的结果。本文提出了一种基于多核学习的协同滤波算法。该算法在现有核函数的基础上，优化各个核函数的权重以匹配数据的分布。在大众点评数据集和Foursquare数据集上的实验结果表明：基于多核学习的协同滤波算法比经验给定的相似函数的性能要高，具有更好的普适性。

关键词：协同滤波；多核学习；随机梯度；个性化推荐

中图分类号：TP311　　**文献标志码**：A

Collaborative Filtering Algorithm Based on Multiple Kernel Learning

Song Kaitao, Peng Furong, Lu Jianfeng

(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China)

Abstract: As a frequently personalized recommendation algorithm of the currently recommendation system, collaborative filtering uses the item evaluation by the approximate users to recommend. Kernel function is an approach for non-linear pattern analysis problems. Ordinarily, collaborative filtering will choose some different kernel functions to analyse the influence between the users. Since the single kernel function can not be adapted to the complicated and various scene, the combination of multiply kernel function becomes a solution. In terms of scenes, multiply kernel learning can combine every kernel function for a better result. This paper proposes a collaborative filtering algorithm based on multiple kernel learning. Based on the available kernel function, this algorithm optimizes the weights of every kernel function to match the data distribution. The experimental result on dianping dataset and foursquare dataset shows that compared with the collaborative filtering algorithm based on common similarity, the collaborative filtering algorithm based on multiple kernel learning achieves better performance. That is, multiple kernel learning has a better common adaptation.

Key words: collaborative filtering; multiple kernel learning; stochastic gradient; personalized recommendation

---

基金项目：江苏省"六大人才高峰"资助项目。
收稿日期：2016-05-30；**修订日期**：2016-06-20

引 言

随着互联网的普及,中国的互联网用户数量已经在 2015 年超过了 9.7 亿。互联网的快速发展,促使了移动端电子商务的快速发展。越来越多快捷方便的基于移动端的 O2O(Online to offline)电子商务应用出现。个性化推荐系统的出现为移动互联网的信息过载提供了帮助。个性化推荐系统帮助移动端用户快速寻找并向其推荐感兴趣的商品、服务。因此,对个性化推荐技术的研究,也成为了当前的一个研究热点。

协同滤波[1]是目前推荐系统中主流的一种个性化推荐算法。1992 年,Xerox 公司在针对 Palo Alto 研究中心的信息重载问题中设计了 Tapestry,该系统首次引入了协同滤波的概念。1994 年,GroupLens 提出后,协同滤波得到大幅度发展,许多电子商务网站都开始采用协同滤波算法为用户提供个性化推荐服务。基于协同滤波的推荐算法,相比于常规的基于关联规则、基于内容等推荐算法而言,拥有自动分析兴趣的优点,能够体现个性化推荐的优点,结果直观,易解释。协同滤波可以分为基于用户(User-based)的协同滤波和基于物品(Item-based)的协同滤波。基于用户的协同滤波通过分析不同用户来进行推荐,该算法能够挖掘用户的潜在兴趣。但是在目前的应用场景下,用户的数量规模日益庞大,使得基于用户分析兴趣的性能逐渐下降。Item-based 协同滤波算法是对用户的评分数据进行分析。该算法通过采用常规相似性度量来分析项目的最近邻居,向预测评分值较高的用户推荐相似项。传统的协同滤波推荐算法,只需要用户对项目的评分表就可以进行推荐。在基于移动端的电子商务系统中,通常还能够利用到用户的地理位置信息。Item-based 协同滤波算法能够不需要其他用户的行为特征就可以进行个性化推荐。而且,对于多数数据集,用户数量都是远大于项目的数量,这使得 Item-based 协同滤波具有时间上的优势。但该方法同样存在采用传统相似性度量的缺陷,其只能获取传统方法的特点,并且伴随数据规模的变大,会面临着数据稀疏性问题、冷启动问题、系统扩展性差等问题。因此,寻找一种适应性更好的相似性度量也成为了协同滤波的一个研究热点。

核函数是模式识别中常用的一种非线性分类技巧。其中,最出名的应用是在 1995 年由 Vapnik 等提出的支持向量机(Support vector machine, SVM)[2]。核函数通过一个非线性变换将输入空间映射到高维特征空间,然后在线性空间使用线性计算算法。因此,核函数的出现使得线性不可分问题变成某些高维特征空间的线性可分问题,从而降低了问题的复杂性。但面对不同的场合,核函数的性能差异很大,并且核函数的选择也没有很好的理论依据。因此,将多个单核函数组合成灵活性更强的多核学习(Multiple kernel Learning, MKL)[3,4],成为了一个新的研究热点。近年来的理论学习,例如 Boosting 的多核组合模型学习方法,基于半定规划(Semi definite programming, SDP)的多核学习方法,简单多核学习方法(Simple KML)等,都证明了多核学习模型能够相比于单核模型能够提升分类的精准度。在多核学习中,比较常用的是多个单核函数的凸组合,其形式如:$K=\sum_{j=1}^{M}\beta_j K_j, \beta\geqslant 0, \sum_{j=1}^{M}\beta_j=1$。除此之外,还包含了直接求和核[3],加权求和核[3]等。因此,多核学习模型的问题转化成了如何去寻找核函数最优系数的问题。

1 核函数原理与设计

1.1 核函数原理

假设 X 是一个非空集合,H 为一个内积空间,φ 为 X 到 X 的映射。如果函数 $K:X\times X\rightarrow \mathbf{R}$,满足对于 $\forall x,x'\in X$ 有 $K(x,x')=<\varphi(x),\varphi(x')>$,则称 K 为核函数。因此,核函数可以看作是内积概念的一个推广。根据 Hilbert-Schmidt 原理[5],任何满足 Mercer 条件[6]的运算,都可以作为核函数的内积。

Mercer 条件 对于任意的对称函数 $K(x,x')$，它是某个特征空间中的内积运算的充分必要条件是，对于任意的 $\varphi(x)\neq 0$ 且 $\int \varphi^2(x)\mathrm{d}x<\infty$，满足

$$\iint K(x,x')\varphi(x)\varphi(x')\mathrm{d}x\mathrm{d}x' > 0 \tag{1}$$

Mercer 条件是帮助核函数检验其是否定义了一个特征空间的充分条件。其中，满足 Mercer 条件的核函数为容许核。容许核函数满足部分闭包性质。容许核的正系数线性组合同样也是容许核。

1.2 核函数设计

在基于项目的协同过滤推荐算法中，首先要计算项与项之间的相似度来寻找目标项的最相似的邻居集合。通常，相似性度量方法[7]，需要保证其度量值越大，相似程度越高的性质。核函数通常可以作为样本在特征空间的相似性度量，因此，可以作为协同滤波算法的相似性度量方法。利用协同滤波算法中常用的传统相似性度量方法和项目的地理位置信息，分别设计了皮尔逊核、余弦核，Jaccard 核和径向基核。

皮尔逊相关系数(Pearson correlation coefficient，PCC)[8]常用于度量两个向量之间的线性相关性。假设项 x 和项 y 的共同评分项为 I，使用皮尔逊相关系数分析项之间的相似度公式为

$$\mathrm{PCC}(x,y)=\frac{\sum_{p\in I}(r_{x,p}-\bar{r}_x)(r_{y,p}-\bar{r}_y)}{\sqrt{\sum_{p\in I}(r_{x,p}-\bar{r}_x)^2}\sqrt{\sum_{p\in I}(r_{y,p}-\bar{r}_y)^2}} \tag{2}$$

式中 $r_{x,p}$，$r_{y,p}$ 表示用户 p 对项 x，项 y 的评分。$\bar{r}_x$ 和 $\bar{r}_y$ 表示项 x 和项 y 上用户集评分的平均值。

余弦相关性(Cosine similarity)[8]将两个向量的夹角余弦值来衡量向量之间的相似性。把 n 用户对于项的评分看做一个 n 维向量，使用余弦相似度的公式

$$\mathrm{COS}(x,y)=\frac{\boldsymbol{r}_x\cdot\boldsymbol{r}_y}{\|\boldsymbol{r}_x\|\cdot\|\boldsymbol{r}_y\|} \tag{3}$$

式中 $\boldsymbol{r}_x$ 和 $\boldsymbol{r}_y$ 表示项 x 与项 y 的评分向量。$\|\cdot\|$ 表示向量的模。

Jaccard[8]相关系数是两项的交集与并集的比值，相似度公式为

$$\mathrm{JACCARD}(x,y)=\frac{X\cap Y}{X\cup Y} \tag{4}$$

式中 X 与 Y 分别表示项 x 与项 y 的评分集。

径向基核[9,10](Radial basis function kernel)，又称为 RBF 核，是一种常用核函数，通常定义为空间中任一点 x 到某一点到某一中心 x_c 之间欧氏距离的单调函数

$$K(\boldsymbol{x},\boldsymbol{x}')=\exp\left(-\frac{\|x-x'\|^2}{2\sigma^2}\right) \tag{5}$$

式中 $\|\boldsymbol{x}-\boldsymbol{x}'\|^2$ 是两个特征之间的欧拉距离平方根，σ 是自由参数。

在带有地理位置信息的数据集上，通常可以获取项的经纬度，项 x 与项 y 的直接地理位置距离[10]的公式如下

$$\begin{aligned}\mathrm{dis}(x,y)=R\times a\cos(&\sin x_{\mathrm{lat}}\sin y_{\mathrm{lat}}+\\&\cos x_{\mathrm{lat}}\cos y_{\mathrm{lat}}\cos(x_{\mathrm{lng}}-y_{\mathrm{lng}}))\end{aligned} \tag{6}$$

式中 x_{lat}，x_{lng}，y_{lat}，y_{lng} 分别表示项 x 与项 y 的纬度与经度。由于不满足其度量值越大，相似度越高的性质，将项与项之间的直接地理位置距离以径向基核的形式表示

$$\mathrm{LOC}(x,y)=\exp\left(-\frac{\mathrm{dis}(x,y)}{l}\right) \tag{7}$$

式中 $\mathrm{dis}(x,y)$ 是式(4)所描述的项 x 与项 y 的直接地理位置距离，l 是自由参数。

2 多核学习方法

2.1 合成核方法

在多核学习中,最优核通常是采用多个基本核函数的线性组合方式。图 1 为一个多核线性组合的示意图。

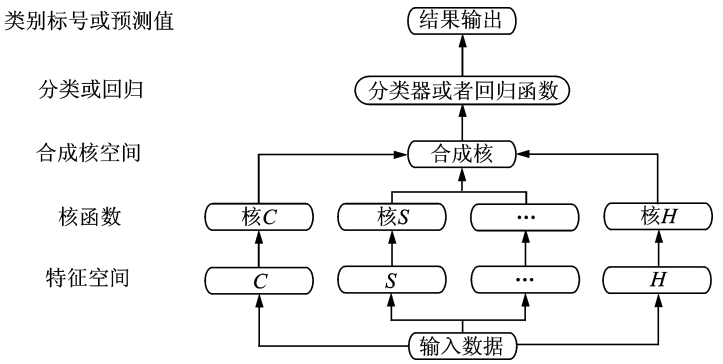

图 1 多核函数线性组合过程

Fig. 1 Linear combination of multiple kernel function

本文采用了加权求和核[11,12]的方法,将皮尔逊核、余弦核、Jaccard 核和径向基核等进行线性组合。其形式如下

$$\begin{aligned}&\mathrm{sim}(x,y)=\sum_{i=0}^{3}w_i\mathrm{sim}_i(x,y)\\&\text{s.t.}\quad\sum_{i=0}^{3}w_i=1,w_i\geqslant 0\end{aligned}\tag{8}$$

式中 sim_i 对应某种类型的核函数,w_i 分别对应各个核函数的核系数,并确保各个核系数之和等于 1,且每个核系数都大于等于 0。

2.2 合成核的学习方法

度量用户偏好商品 i 的概率采用评分加权推荐公式[13]来计算。该方法考虑了相似邻居集的评分尺度的影响。公式如下

$$P(i,u)=\frac{\sum_{j\in SN_i}\mathrm{sim}(i,j)\times r_{j,u}}{\sum_{j\in SN_i}\mathrm{sim}(i,j)}\tag{9}$$

式中:SN_i 表示项 i 的 K 个最近邻居集合;$\mathrm{sim}(i,j)$是式(8)所设计的多核线性组合模型;$r_{j,u}$是用户 u 在项目 j 上的评分值。

本文采用随机梯度下降[14]作为学习方法。随机梯度下降是一种常用的最小化损失函数方法。随机梯度下降通过对单样本的损失误差求解梯度,从而更新参数。其一次梯度迭代下降的时间复杂度较低。评估协同滤波算法的损失函数可以使用最小二乘误差[15]。在本文中,其最小二乘误差表示如下

$$\mathrm{loss}(v,u)=\frac{1}{2}(P(v,u)-y(v,u))^2\tag{10}$$

式中:$P(v,u)$为式(9)中的评分方式;$y(v,u)$为观察值。

损失函数的梯度求解主要是对式(10)进行梯度求解,其求解过程如下:

(1) 损失函数的梯度形式：$(P(v,u)-y(v,u))\dfrac{\partial P(v,u)}{\partial w}$，由于 $y(v,u)$ 是常数，$P(v,u)$ 可以通过式(9)求出，因此，只需要考虑 $\partial P(v,u)/\partial w$ 的情况。

(2)形式化解：$\dfrac{\partial P(v,u)}{\partial \boldsymbol{w}}=\partial\left\{\dfrac{\sum\limits_{i=0}^{3}w_i\sum\limits_{j\in SN_v}r_{j,u}\mathrm{sim}_i(v,j)}{\sum\limits_{i=0}^{3}w_i\sum\limits_{j\in SN_v}\mathrm{sim}_i(v,j)}\right\}\Big/\partial \boldsymbol{w}$

(3)简化过程，令 $a_i=\sum\limits_{j\in SN_v}r_{j,u}\mathrm{sim}_i(v,j)$，$b_i=\sum\limits_{j\in SN_v}\mathrm{sim}_i(v,j)$，因此，整体过程可以简化为向量的形式。梯度公式为 $\dfrac{\partial P(v,u)}{\partial \boldsymbol{w}}=\partial\left\{\dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{w}}{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{w}}\right\}\Big/\partial \boldsymbol{w}$，其中 $\boldsymbol{a},\boldsymbol{b},\boldsymbol{w}$ 均为列向量。

(4)对于向量 $\boldsymbol{w}$ 的求导结果为

$$\frac{\partial P(v,u)}{\partial \boldsymbol{w}}=\frac{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{w}\boldsymbol{a}-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{w}\boldsymbol{b}}{(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{w})^2}$$

(5)因此整体损失函数的梯度值为

$$\frac{\partial \mathrm{loss}(v,u)}{\partial \boldsymbol{w}}=(P(v,u)-y(v,u))\frac{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{w}\boldsymbol{a}-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{w}\boldsymbol{b}}{(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{w})^2}$$

本文所采用的随机梯度下降的学习方法，梯度的求解采用了损失函数梯度求解方法。因此基于多核学习的推荐系统算法过程如下：

输入：评分表，测试集和初始多核系数 $\boldsymbol{w}$

输出：推荐用户-项目对

(1) 遍历项目-用户对(v,u)，对于当前项目 v，确定其邻居集 SN

(2) 对于当前的项目-用户对。采用章节 3.2.1 中的方法，求解梯度 η，采用随机梯度的方法更新多核系数 w。随机梯度的更新公式为：$w^{(n+1)}=w^n-\alpha\eta$，其中 α 为步长。

(3) 重复迭代(1)～(2)，直到损失函数变化趋于稳定，获取多核系数 $\boldsymbol{w}$，换到第(4)步

(4)遍历训练集，采用训练得到的多核方程，重新计算每个项目的邻居集。

(5)对于当前项目 v，采用式(9)计算一定范围内的每个用户的期望评分，并选出评分前 N 高个用户，进行推荐。

(6)将推荐集合与测试集合进行比较，评测推荐系统的性能。

由于每次更新核系数后，项目的邻居集都会变化，使得直接在整体集上求单个项目的邻居集的时间复杂度大。因此，本文预先处理单个项目在多个常规相似性度量上的邻居集，从而在新参数 w 下寻找项目的邻居集时只需要在常规相似性度量上的邻居集合中搜索。而计算各个核函数的过程同样需要一定时间复杂度。因此，在搜索常规度量邻居集合的同时，将项目与邻居集的相似性度量值保存，方便在多核学习过程中直接使用，避免重复计算。

3 实验结果分析

3.1 数据集

为了验证本文所提出的方法，实验中使用大众点评数据集和 Foursquare 两个真实数据集进行验证。

大众点评数据集选用了某地区店铺的用户评价信息以及店铺地理位置。用户对店铺的喜好采用评分形式。数据集仅有 0.148 9%的用户-店铺对有评分项。

表 1 大众点评数据集信息

Tab. 1 Dianping dataset information

用户数量	商品数量	评分量
82 475	8 318	1 021 509

Foursquare 数据集选用了新加坡地区 2010 年 8 月到 2011 年 7 月用户的签到数据。将用户是否有签到行为作为评分项。由于 Foursquare 数据集是属于二进制数据集,因此,各个核的效果在 Foursquare 数据集上会出现较大波动。

表 2 Foursquare 数据集信息

Tab. 2 Foursquare dataset information

用户数量	商品数量	评分量
2 310	5 528	105 459

本文将数据集按照 4∶1 的比例划分训练集和测试集,进行 5-折交叉验证。

3.2 推荐评分标准

在获取最优多核系数 w 后,将采用式(9)计算评分,进行 Top-N 推荐。对于推荐性能的评价,在推荐系统中,通常会选用 F_1 值(F_1 Score)[16]作为标准,因为 F_1 值可以同时顾及二分类推荐模型中准确率(Precision)和召回率(Recall)。

准确率:推荐命中的个数占推荐商品总的个数比率。

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{11}$$

召回率:用户所喜欢的商品最终被推荐出来的比率。

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{12}$$

F_1 值是准确率和召回率的调和平均数,公式为

$$F_1=2\times\frac{\text{Precsion}\times\text{Recall}}{\text{Precsion}+\text{Recall}} \tag{13}$$

式中:TP 表示推荐集合中正类的个数,FP 表示推荐集合中负类的个数;FN 表示没有在推荐集合中的正类个数。

3.3 实验结果

在实验中,采用平均绝对误差(Mean absolute difference, MAE)[17]来评估多核学习过程中的损失情况。平均绝对误差是指所有单样本的观察值与算术预测值的绝对偏差之和的平均值,其公式为

$$\text{MAE}=\frac{1}{n}\sum_{i=1}^{n}|f_i-y_i| \tag{14}$$

式中:f_i 表示预测值,y_i 代表观察值。

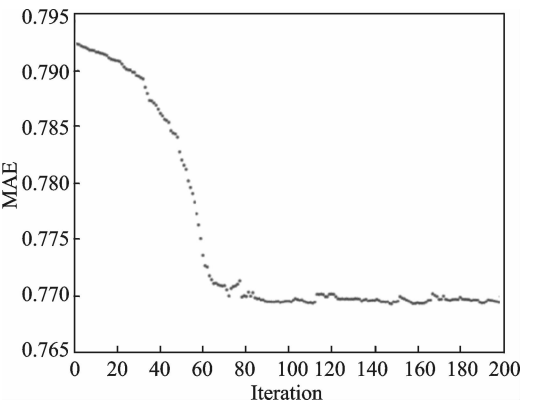

图 2 核系数目标损失函数下降

Fig. 2 Descent of kernel factor cost function

由于数据集的规模较大,计算整体损失函数的时间较多。因此,在核系数学习过程中,采用批次处理的方法,进行梯度上的更新。在本文的实验中,设置迭代计算间隔次数为 1 000 次,步长 α 为 0.00 002。观察图 2,在 60 次左右的损失函数计算后,即大约 60 000

次左右的学习迭代，整体损失函数下降逐渐趋于稳定。将最后趋于稳定的核系数作为多核方程进行协同滤波推荐。

本文将基于多核学习的协同滤波算法与采用常规相似性度量方法，包括了 Jaccard 核、余弦核、径向基核这 3 种方法，以及平均加权常规相似性度量方法的协同滤波算法进行比较。进行 Top-N 推荐时，选择不同的 Top-N 系数，Top-N 的范围设置在 10%～25%，采用式(11)的 F_1 值作为评分标准。

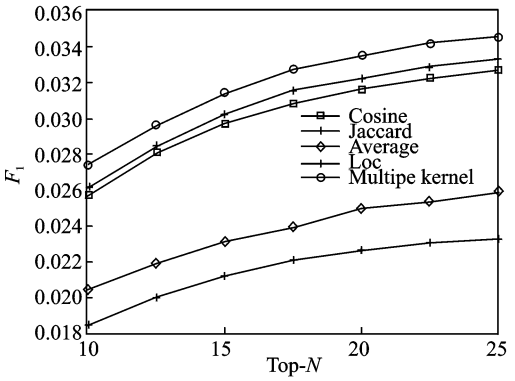

图 3 大众点评数据集实验结果

Fig. 3 Experimental results of dianping dataset

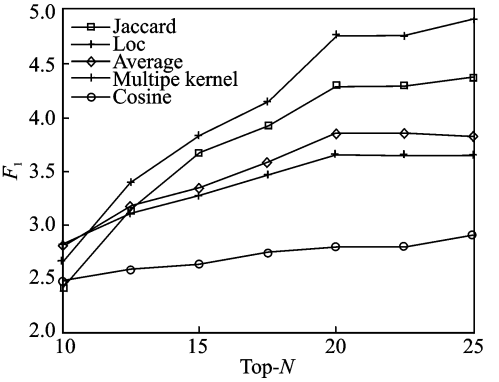

图 4 Foursquare 实验结果

Fig. 4 Experimental results of foursquare dataset

大众点评的数据集的最近邻个数设置为 30，径向基核的参数 l 设置为 1。通过观察图 3，在大众点评数据集上，基于多核学习的协同滤波算法相比于采用传统相似性度量方法的协同滤波算法，F_1 值提升了 3.973%。观察图 4，受限于 Foursquare 数据集的二进制数据，使得在 Top-N 较小时，效果不是很明显，在 Top-N 大于 20%的时候，基于多核学习的协同滤波算法相比于采用传统相似性度量方法的协同滤波算法的推荐性能有了整体的提升，平均 F_1 值相比于采用传统相似性度量方法的协同滤波算法提升了 6.523%。说明了基于多核学习的协同滤波算法拥有更好的推荐性能。

4 结束语

本文提出了一种基于多核学习的协同滤波算法，并给出学习方法的推导。该方法相比于采用传统相似性度量的协同滤波算法而言，具有选择最优核的能力，能够自动学习核系数来提升性能，并体现各个核函数对实验结果的影响。在大众点评数据集和 Foursquare 数据集上的实验表明，本文所提出的基于多核学习的协同滤波算法提升了推荐性能，具有有效性。但是，核系数的初始化选择以及核函数的组合，对实验结果会有很大的影响。如何针对于数据集，选择核系数与核函数，仍是多核学习中的难点。

参考文献：

[1] Badrul Sarwar. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the 10th International Conference on World Wide Web. New York: WWW10, 2001: 285-295.

[2] Suykens J K. Least squares support vector machine classifiers[J]. Neural Processing Letters, 1993, 9(3): 293-300.

[3] 汪洪桥. 多核学习方法[J]. 自动化学报, 2010, 36(8): 1037-1050.
Wang Hongqiao. On multiple kernel learning methods[J]. Acta Automatic Sinica, 2010, 36(8): 1037-1050.

[4] Meirom E A. Nuc-MKL: A convex approach to non linear multiple kernel learning[J]. AISTATS, 2016, 51: 610-619.

[5] Li Liyan. Gene networks identification using independence measurement based on the Hilbert space[D]. Hangzhou: Hangzhou Dianzi University, 2014.

[6] Lyu Siwei. Mercer kernels for object recognition with local features[J]. CVPR, 2005, 2: 1063-6919.

[7] Liu Haifeng. A new user similarity model toimprove the accuracy of collaborative filtring[J]. Knowledge-Based Systems, 2014, 56: 156-166.

[8] Ekstrand M D, Riedl J T, Konstan J A. Collaborative filtering recommender system [J]. Foundations and Trends in Human-Computer Interaction,2010,4(2):81-173.

[9] Chung Kai-Min. Radius margin bounds for support vector machines with the RBF kernel[J]. Neural Computation,2003,15(11):2643-2681.

[10] 奚吉,赵力,左加阔.基于改进多核学习的语音情感识别算法[J].数据采集与处理,2014,29(5):730-734.
Xi Ji, Zhao Li, Zuo Jiahuo. Speech emotion recognition based on modified multiple kernel learning algorithm[J]. Journal of Data Acquisition and Processing, 2014, 29(5):730-734.

[11] 王国胜.核函数的性质及其构造方法[J].计算机科学,2006,33(6):172-178.
Wang Guosheng. Properties and construction methods of kernel in support vector machine[J]. Computer Science,2006,33(6):172-178.

[12] Lu Yanting. Research and application of clustering and hierarchical classification algorithms based on multiple kernel learning [D]. Nanjing:Nanjing University of Science and Technology, 2013.

[13] 王付强,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,2016,36(1):171-174.
Wang Fuqiang. Location-based asymmetric similarity for collaborative filtering recommendation algorithm[J]. Journal of Computer Applications, 2016, 36(1):171-174.

[14] Léon Bottou. Large-scale machine learning with stochastic gradient descent[C]// 19th International Conference on Computational Statistics. Paris: Proceedings of COMPSTAT,2010:177-186.

[15] York D. Least squares fitting of a straight line with correlated errors [J]. Earth and Planetary Science Letters,1968,5:320-324.

[16] Yang Yiming. A re-examination of text categorization methods[C]// the 22nd Annual International ACM SIGIR Conference. USA:SIGIR,1999:42-49.

[17] Vassiliadis S. The sum-absolute-difference motion estimation accelerator[C]// Proceedings of the 24th Euromicro Conference. Germany: Euromicro Conference,1998:559-566.

作者简介:

宋恺涛(1993-),男,硕士研究生,研究方向:数据挖掘、推荐系统,E-mail:sktsxy@gmail.com。

彭甫镕(1987-),男,博士,研究方向:数据挖掘、推荐系统。

陆建峰(1969-),男,教授,研究方向:模式识别。

(编辑:张 彤)