

# 一种面向非线性回归的迁移学习模型

杨沛<sup>1</sup> 谭琦<sup>1,2</sup> 丁月华<sup>1</sup>

(华南理工大学计算机学院 广州 510640)<sup>1</sup> (华南师范大学计算机学院 广州 510631)<sup>2</sup>

**摘要** 迁移学习能够有效地在相似任务之间进行信息的共享和迁移。之前针对多任务回归的迁移学习研究大多集中在线性系统上。针对非线性回归问题,提出了一种新的多任务回归模型——HiRBF。HiRBF基于层次贝叶斯模型,采用RBF神经网络进行回归学习,假设各个任务的输出层参数服从某种共同的先验分布。根据各个任务是否共享隐藏层,在构造HiRBF模型时有两种可选方案。在实验部分,将两种方案进行了对比,也将HiRBF与两种非迁移学习算法进行了对比,实验结果表明,HiRBF的预测性能大大优于其它两个算法。

**关键词** 迁移学习,层次贝叶斯,回归,RBF神经网络

**中图分类号** TP181 **文献标识码** A

## Non-linear Transfer Learning Model

YANG Pei<sup>1</sup> TAN Qi<sup>1,2</sup> DING Yue-hua<sup>1</sup>

(Institute of Computer Science, South China University of Technology, Guangzhou 510640, China)<sup>1</sup>

(Institute of Computer Science, South China Normal University, Guangzhou 510631, China)<sup>2</sup>

**Abstract** Multi-task learning utilizes labeled data from other "similar" tasks and can achieve efficient knowledge-sharing between tasks. Previous research mainly focused on multi-task learning for linear regression. A novel Bayesian multi-task learning model for non-linear regression, i. e. HiRBF, was proposed. HiRBF is constructed under a hierarchical Bayesian framework. According to whether the input-to-hidden is shared by all tasks or not, we have two options to build the HiRBF model. There is a comparison between them in the experiment section. The HiRBF algorithm is also compared with two transfer-unaware approaches. The experiments demonstrate that HiRBF significantly outperforms the others.

**Keywords** Transfer learning, Bayesian hierarchical model, Regression, RBF network

## 1 概述

训练样本不足是许多机器学习应用面临的主要问题之一。例如:在超市商品销量预测系统中,节假日(如圣诞节)的数据样本每年只能统计一次;在内容推荐系统中,新用户的历史数据非常稀缺,但老用户的历史数据非常多。利用稀缺数据往往不足以训练出一个具有良好泛化能力的模型。而多任务学习则能有效地在相关但不相同任务之间进行信息的共享和迁移。

多任务学习是基于这样一个假设,即多个任务之间共享某种隐含结构,例如,神经网络的隐层单元<sup>[1]</sup>;相同的特征映射<sup>[2,3]</sup>;多层贝叶斯模型中的协方差<sup>[4,5]</sup>。

多层贝叶斯模型是对来自相关场景的数据进行建模的有力工具<sup>[6]</sup>。Caruana<sup>[1]</sup>利用多个任务同时训练一个神经网络,通过共享隐藏层节点来进行多任务学习。Baxter<sup>[7]</sup>提出采用多层贝叶斯模型来进行多任务学习,不同学习任务之间共享的参数被看作是超参数。针对线性模型, Heskes<sup>[8]</sup>将神经网络

的输入层-隐藏层的权值看作超参数,由多个学习任务共享;而将隐藏层-输出层的权值看作对应学习任务的模型参数。Yu<sup>[5]</sup>从非参数化的角度提出了多任务学习的多层贝叶斯模型。Liao<sup>[9]</sup>利用辅助的数据训练 Logistic 回归模型。Zhang<sup>[10]</sup>利用多层贝叶斯模型应用于内容协同推荐系统。Roy等<sup>[11]</sup>提出了基于 Dirichlet 过程的多任务学习模型,在该模型中,多个朴素贝叶斯分类器的模型参数服从 Dirichlet 先验分布。Yu<sup>[12]</sup>提出了基于 t-过程的多任务学习算法,该算法能有效地从噪声任务中挑选出好的任务。

在多任务回归问题上,由于似然函数积分计算上的困难,基于层次贝叶斯的多任务学习的研究目前大多集中在线性模型上。但是,现实中大多数的回归问题都是非线性的,因此,研究非线性问题的多任务回归具有重要的理论和应用价值。针对非线性回归问题,本文提出了一种新的基于层次贝叶斯的多任务回归模型。该模型采用 RBF 神经网络,多个任务之间通过两种方式进行信息的共享和迁移。其一,多个任务共享 RBF 网络的输入层-隐藏层。其二,假设各个任务对应的权值

到稿日期:2008-10-14 返修日期:2009-02-03 本文受国家自然科学基金项目(60574078)资助。

杨沛(1974-),男,博士,主要研究方向为机器学习、数据挖掘等,E-mail:yangpei@scut.edu.cn;谭琦(1977-),女,博士生,主要研究方向为情感计算等;丁月华(1950-),男,教授,主要研究方向为人工智能等。

向量服从某种共同的先验分布。

## 2 HiRBF 多任务回归模型

HiRBF 采用 RBF 神经网络构建回归模型。RBF 网络包括 3 层:输入层、隐藏层和输出层。HiRBF 多任务回归模型如图 1 所示。其中,每个虚线框内部分表示一个学习模型, $x$  为输入, $y$  为隐藏层输出, $z$  是输出层输出, $\phi$  是隐藏层的径向基函数, $w$  是输出层的权值向量,假设各个任务的输出层权值向量服从某种共同的先验分布,该分布的参数用  $\theta$  表示。

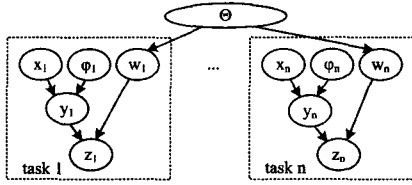


图 1 HiRBF 多任务回归模型

根据各个任务是否共享相同的  $\phi$ , 在构造 HiRBF 多任务回归模型时,有两种可选方案。方案一如图 2 所示,各个任务共享相同的  $\phi$ ,即各个任务共享相同的隐藏层。方案二如图 3 所示,各个任务的  $\phi$  是不同的,即各个任务的隐藏层是各不相同的。在实验分析部分,我们会对两个方案进行对比。

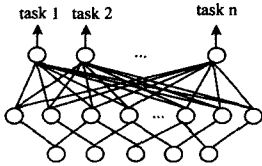


图 2 各个任务共享相同的隐藏层

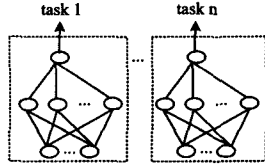


图 3 各个任务的隐藏层各不相同

假设待学习任务共有  $n$  个。设  $w_i$  ( $1 \leq i \leq n$ ) 为对应第  $i$  学习任务的输出层权值向量。设  $x_{i,j}$  和  $z_{i,j}$  表示第  $i$  个学习任务的第  $j$  个数据样本的输入和输出,  $y_{i,j}$  是隐藏层输出,第  $i$  个学习任务的数据集用  $D_i = \{(x_{i,j}, z_{i,j})\}$  表示。假设输入  $x$  是  $s$  维向量,隐藏层输出  $y$  是  $m$  维向量。为了讨论方便,假设输出  $z$  是一维向量。

具体而言,HiRBF 多任务回归模型包括以下几个要点:

a) 对于每个学习模型,假设权值向量  $w_i$  服从  $m$  维正态分布,即:  $w_i \sim N(\mu, C)$ , 其中  $1 \leq i \leq n$ ,  $\mu$  和  $C$  为先验参数。

b) 输入层-隐藏层采用非线性模型。对于每个数据样本,根据径向基函数计算隐藏层输出  $y_{i,j} = \phi_i(x_{i,j})$ , 其中  $1 \leq i \leq n, 1 \leq j \leq |D_i|$ 。径向基函数  $\phi$  采用高斯函数。

c) 隐藏层-输出层采用线性模型。对于每个数据样本,假设输出层输出  $z_{i,j} \sim N(w_i^T y_{i,j}, \sigma^2)$ , 其中  $1 \leq i \leq n, 1 \leq j \leq |D_i|$ 。

从中可以看出,HiRBF 多任务回归模型的实现可分为两个阶段:1) 确定隐藏层参数;2) 确定输出层参数。下面分别进行介绍。

### 2.1 计算隐藏层参数

在 RBF 神经网络中,径向基函数通常采用高斯函数。确定隐藏层参数就是要确定每个隐藏层单元的径向基函数的期望值和方差。

在 HiRBF 模型中,利用  $k$  均值聚类算法确定 RBF 网络隐藏层参数。 $k$  是用户指定的核函数的数量。在第一种方案中,所有学习任务共享一个隐藏层。利用  $k$  均值聚类算法对

所有任务的数据样本进行聚类,确定  $k$  个核函数。在第二种方案中,各个任务的隐藏层各不相同。每个模型的训练单独进行,利用  $k$  均值聚类算法对任务内部的数据样本进行聚类,确定  $k$  个核函数。

### 2.2 计算输出层参数

由于先验参数  $\theta_k$  是未知的,所以不能直接计算出权值向量  $w_i$ 。但是,可以将权值向量  $w_i$  ( $1 \leq i \leq n$ ) 看成隐变量,采用 EM 算法求解之。给定各个学习任务的数据集  $D_i$  ( $1 \leq i \leq n$ ) 和先验参数  $\theta_k$ ,完整数据的似然函数可表示为:

$$Q(\theta_{k+1}) = \sum_{i=1}^n \int w_i p(w_i | D_i, \theta_k) \ln p(D_i, w_i | \theta_k) dw_i \quad (1)$$

其中,  $\theta_k$  表示第  $k$  轮迭代时得到的对整个分布的最好的估计,  $\theta_{k+1}$  表示改进后的参数估计。权值向量  $w_i$  的后验概率可利用贝叶斯公式计算:

$$p(w_i | D_i, \theta_k) = \frac{p(D_i | w_i) p(w_i | \theta_k)}{\int_{w_i} p(D_i | w_i) p(w_i | \theta_k) dw_i} \quad (2)$$

其中,

$$p(D_i | w_i) = \prod_{j=1}^{|D_i|} p(z_{i,j} | w_i, y_{i,j}) \quad (3)$$

在式(1)中,  $D_i$  和  $w_i$  的联合概率可通过式(4)计算:

$$p(D_i, w_i | \theta_k) = p(D_i | w_i) p(w_i | \theta_k) \quad (4)$$

而由于  $z_{i,j} \sim N(w_i^T y_{i,j}, \sigma^2)$ , 所以有:

$$p(z_{i,j} | w_i, y_{i,j}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(z_{i,j} - w_i^T y_{i,j})^2\right\} \quad (5)$$

由于  $w_i \sim N(\mu, C)$ , 所以有:

$$p(w_i | \mu, C) = \frac{1}{(2\pi)^{n/2} |C|^{1/2}} \exp\left\{-\frac{1}{2}(w_i - \mu)^T C(w_i - \mu)\right\} \quad (6)$$

求得  $Q$  后,可通过下式求得优化后的参数向量:

$$\theta_{k+1} = \operatorname{argmax}_{\theta} Q(\theta_{k+1}) \quad (7)$$

具体算法如下:

算法(确定输出层参数的 EM 算法)

1) initialized, 给先验参数  $\theta_0$ 、均方差  $\sigma$  设置初始值,  $k=0$ , 误差阈值  $\epsilon$ 。

2) do {  $k++$ ;

3) E 步: 首先计算完整数据的似然函数  $Q(\theta_{k+1})$ 。然后,对每一个学习模型进行下述处理:根据数据集  $D_i$  和当前的先验参数  $\theta_k$  计算权值向量  $w_i$  的期望值  $\bar{w}_i$  和协方差  $C_i$ 。

$$C_i = D(w_i) = \left(C^{-1} + \frac{\sum_{j=1}^{|D_i|} y_{i,j} y_{i,j}^T}{\sigma^2}\right)^{-1} \quad (8)$$

$$\bar{w}_i = E(w_i) = C_i \left(C^{-1} \mu + \frac{\sum_{j=1}^{|D_i|} y_{i,j} z_{i,j}}{\sigma^2}\right) \quad (9)$$

4) M 步: 对  $Q(\theta_{k+1})$  求导, 得到优化后的  $\theta_{k+1}$ :

$$\mu = \frac{1}{n} \sum_{i=1}^n \bar{w}_i \quad (10)$$

$$C = \frac{1}{n} \sum_{i=1}^n (C_i + (\bar{w}_i - \mu)(\bar{w}_i - \mu)^T) \quad (11)$$

5) while ( $\theta_{k+1} - \theta_k > \epsilon$ )

6) return 权值向量  $w_i$  的期望值  $\bar{w}_i$  和协方差  $C_i$ , 其中  $1 \leq i \leq n$ 。

## 3 实验分析

根据文献[1]提到的方法,按以下方式构造不同任务的数

据集:先选取基函数,各个任务具有同样的基函数;然后构造互不相关的噪声函数,通过将基函数和噪声函数叠加,产生对不同任务的函数。构造公式如下:

$$y_i(x) = y(x) + \delta_i(x)$$

其中, $y(x)$ 是基函数, $\delta_i(x)$ 是第*i*个任务对应的噪声函数, $y_i(x)$ 是第*i*个任务对应的函数。

在本实验中,选取如下两个函数作为基函数:

1)FRIEDMAN 函数:

$y = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5$  变量范围: $x_i \sim U[0, 1]$

2)MULTI 函数:

$y = 0.79 + 1.27x_1 x_2 + 1.56x_1 x_4 + 3.42x_2 x_5 + 2.06x_3 x_4 x_5$  变量范围: $x_i \sim U[0, 1]$

设任务数  $n=5$ ,噪声函数符合正态分布,且两两互不相关:

$$\delta_i \sim N(\alpha_i \bar{y}, \sigma_i^2)$$

其中, $\bar{y}$ 是  $y$  的期望值, $\alpha_i \sim U[0, 0.2]$ , $\sigma_i \sim U[0.5, 1.0]$ ,其中  $1 \leq i \leq n$ 。

实验中采用了 4 种不同算法进行对比:1)Baseline:普通的 RBF 网络,只利用主任务本身的数据集;2)Merge:普通的 RBF 网络,将所有任务的数据集进行简单的合并;3)HiRBF\_a:多个任务共享 RBF 的隐藏层;4)HiRBF\_b:各个任务对应的隐藏层各不相同。

实验结果如图 4—图 7 所示。图 4 和图 5 分别是上述 4 种算法针对 MULTI 的平均相对误差和均方差。图 6 和图 7 中分别是 4 种算法针对 FRIEDMAN 的平均相对误差和均方差。其中,横坐标表示单个任务的数据集的大小,纵坐标表示预测误差。从 4 个图中都可以看到,Baseline 方法在数据集较小时预测误差很大,随着数据集的增大,其预测误差逐渐减少。从图 6 和图 7 中可以看到,在 FRIEDMAN 数据集大小从 200 增加到 400 时,Baseline 方法的预测误差曲线下降的梯度较大;在数据集从 400 增加到 1600 时,其误差曲线下降较为平缓。图 4 和图 5 的情况也与此类似。这说明,根据稀疏的训练样本往往不足以训练出具有较高预测精度的模型。

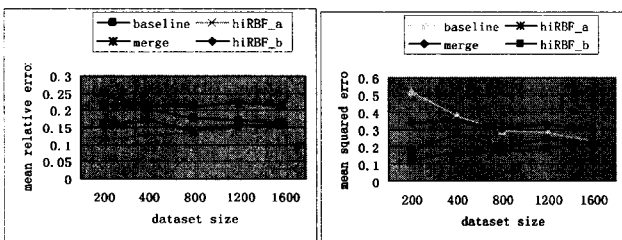


图 4 平均相对误差(multi)

图 5 均方差(multi)

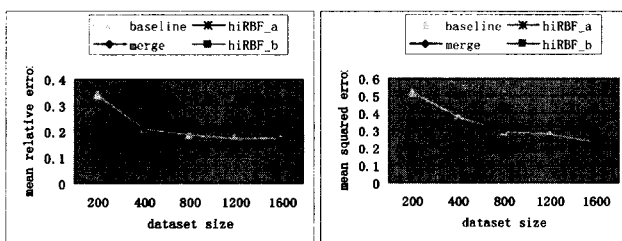


图 6 平均相对误差(friedman#1)

图 7 均方差(friedman#1)

从 4 个图中可以看出,在 MUTLI 和 FRIEDMAN 这两

个数据集上,HiRBF\_a 方法都取得了最好的泛化性能。当数据集大小从 200 增大到 1600 时,在 MULTI 数据集上,HiRBF\_a 的预测精度比 Baseline 分别提高了 55.1%,36.5%,25.0%,13.1%,3.9%;在 FRIEDMAN 数据集上,HiRBF\_a 的预测精度比 Baseline 分别提高了 63.7%,35.7%,24.1%,19.6%,12.2%。数据集越小,预测精度提高得越大。当数据集逐渐增大时,HiRBF\_a 和 Baseline 的性能差距逐渐缩小。此外,还有一点需要关注的是,随着数据集的增大,HiRBF\_a 的预测精度有缓慢上升的趋势,这可能是由于大量辅助任务的数据对预测模型造成了一定程度的干扰,具体原因还有待进一步的研究。

从 4 个图中还可以看出,在 MUTLI 和 FRIEDMAN 这两个数据集上,HiRBF\_b 方法的预测性能与 HiRBF\_a 方法较为接近,但是稍逊于 HiRBF\_a 方法。这说明多个学习任务共享隐藏层能够在更大程度上进行信息的共享和迁移,从而进一步提高学习系统的泛化能力。

而在 MUTLI 和 FRIEDMAN 这两个数据集上,Merge 方法都表现平平。从图 4 和图 6 中可以看到,当数据集增大到 400 时,Merge 的预测性能已经不如 Baseline。这说明简单地将各个任务的数据进行合并,往往并不能提高模型的泛化性能,反而可能由于大量噪声数据的加入,导致模型的预测精度降低。

**结束语** 多任务学习能够挖掘各个相关任务之间的隐含关联关系,从而通过某种方式的共享结构,在任务之间进行信息的共享和迁移。以上实验也表明,多任务学习能有效提高预测系统的泛化性能。尤其是在训练样本不足和存在多个相关任务的情况下,多任务学习要优于单任务学习。

但是,如何判断任务之间的相关呢?即使是相关的任务,也存在“好”任务和“坏”任务之分。“好”任务能提高多任务学习模型的泛化性能,而“坏”任务则相反。如何筛选出“好”任务,剔除“坏”任务,这都将是我们下一步的研究目标。

## 参考文献

- [1] Caruana R. Multi-task learning [J]. Machine Learning, 1997, 28 (1):41-75
- [2] Ando R K, Zhang T. A framework for learning predictive structures from multiple tasks and unlabeled data [J]. The Journal of Machine Learning Research, 2005, 6(1):1817-1853
- [3] Argyriou A, Evgeniou T, Pontil M. Multi-task feature learning [C]//NIPS'06. New York: ACM Press, 2006:41-48
- [4] Bakker B, Heskes T. Task clustering and gating for Bayesian multitask learning [J]. The Journal of Machine Learning Research, 2003, 4(12):83-89
- [5] Yu K, Tresp V, Schwaighofer A. Learning Gaussian processes from multiple tasks [C]//Proceedings of the 22nd international conference on machine learning. New York: ACM Press, 2005: 1012-1019
- [6] Gelman A, Carlin J, Stern H, et al. Bayesian data analysis [M]. Texts in Statistical Science. New York: Chapman & Hall, 2000
- [7] Baxter J. A model of inductive bias learning [J]. Journal of Artificial Intelligence Research, 2000, 12:149-198
- [8] Heskes T. Empirical bayes for learning to learn [C]//Proceedings of the Seventeenth International Conference on Machine Learning. New York: ACM Press, 2000:367-374

(下转第 242 页)

出其与影响因素之间的关联特性。

分析数据采用某矿瓦斯突出监测数据库。数据库记录了该矿某段时间瓦斯突出监测的 3526 条记录,选取其中的 6 个字段:开采深度、瓦斯压力、放散速度、地质构造、坚固系数、突出类别,期望找出前 5 个字段和最后一个字段之间的关联关系。瓦斯监控数据库部分数据如表 2 所列。

表 2 某矿瓦斯监控数据库数据

开采深度 (m)	瓦斯压力 (MPa)	放散速度 ( $\Delta P$ )	地质构造 (D)	坚固系数 (f)	突出类别 (T)
455	0.95	6	3	0.24	一般
428	1.4	4	3	0.58	无
426	1.4	6	3	0.42	严重
460	2	7	1	0.48	无
425	2.8	8	3	0.59	一般
395	1.17	5	1	0.61	无
552	3.95	14	3	0.24	一般
...	...	...	...	...	...

首先对数据库进行预处理,把数据库中的连续属性值离散化和进行编码。预处理和部分编码结果如表 3 所列。

表 3 离散化及编码结果表

属性名	原属性值范围	对应的离散值	编码值
开采深度	400m 以下	浅	1
	400~499m	较浅	2
	500~599m	较深	3
	600m 以上	深	4
瓦斯压力	<1	小	1
	1~2	较小	2
	2~3	较大	3
	$\geq 3$	大	4
放散速度	<5	慢	1
	5~10	较慢	2
	10~15	较快	3
	$\geq 15$	快	4
坚固系数	<0.4	松软	1
	0.4~0.5	一般	2
	>0.5	坚固	3
突出强度	无	无	1
	一般	一般	2
	严重	严重	3

然后应用算法对关联规则进行挖掘,取种群规模 900,免疫选择阈值为 90%,迭代次数为 1000 代。对支持度阈值为 30%、置信度阈值为 90%进行测试。测试时挖掘出关联规则 276 条,挖掘时间为 21.2s。挖掘的部分关联规则如下:

开采深度=2  $\wedge$  放散速度=2  $\Rightarrow$  瓦斯突出强度=2 [support=15%, confidence=65%];

开采深度=3  $\wedge$  瓦斯压力=4  $\Rightarrow$  瓦斯突出强度=3 [support=12%, confidence=69%];

放散速度=1  $\wedge$  坚固系数=3  $\Rightarrow$  瓦斯突出强度=1 [support=10%, confidence=63%];

瓦斯压力=4  $\wedge$  放散速度=3  $\wedge$  坚固系数=1  $\Rightarrow$  瓦斯突出强度=3 [support=10%, confidence=65%];

瓦斯压力=1  $\wedge$  放散速度=2  $\wedge$  坚固系数=1  $\Rightarrow$  瓦斯突出强度=2 [support=15%, confidence=85%];

瓦斯压力=2  $\wedge$  放散速度=1  $\wedge$  地质构造=3  $\wedge$  坚固系数=3  $\Rightarrow$  瓦斯突出强度=1 [support=10%, confidence=80%].

通过以上挖掘出来的关联规则可以看出,瓦斯突出的强度与开采深度、瓦斯压力、瓦斯放散初速度、煤的坚固系数以及地质破坏程度密切相关,且开采深度越深,煤层中瓦斯压力越高,瓦斯放散初速度越大,煤层强度越低,受地质破坏程度越大,瓦斯突出的危险性也就越大,因此在这样的煤矿开采条件下采取一些防突措施是非常必要的。

**结束语** 免疫算法是受生物免疫系统启发而来,在多个领域的应用已显示出强大的信息处理和求解问题的能力。本文把免疫算法应用到数据挖掘领域,提出了一种基于免疫算法的多维关联规则挖掘算法。本算法充分利用了免疫算法的免疫记忆特性,通过免疫学习把挖掘的关联规则保存在记忆库中,加快了关联规则的挖掘速度。同时算法具有较强的鲁棒性和快速、有效的全局搜索能力。将其应用到煤与瓦斯突出预测中,结果表明在多维关联规则的挖掘中具有可行性和高效性。

## 参考文献

- [1] Han J, Kamber M. Data Mining: Concepts and Techniques [M]. Beijing: Higher Education Press, 2001
- [2] Hipp J, Guntzer U, Nakhaeizadeh G. Algorithms for association rule mining—a general survey and comparison [J]. SIGKDD Explorations, 2000, 2(1): 58-64
- [3] 肖人彬, 王磊. 人工免疫系统: 原理、模型、分析及展望 [J]. 计算机学报, 2002, 25(12): 1281-1293
- [4] Hunt J E, Cooke DE. Learning using an artificial immune system [J]. Journal of Network and Computer Applications, 1996, 19(2): 189-212
- [5] Timmis J. Artificial Immune System: A novel data analysis technique inspired by the immune network theory [D]. Department of Computer Science, University of Wales, 2001
- [6] De Castro L N, Von Zuben F J. An evolutionary immune network for data clustering [A] // Proc. 6th Brazilian Symposium on Neural Networks [C]. Rio de Janeiro, Brazil, 2000: 84-89
- [7] 刘韬, 蔡淑琴, 石双元. 人工免疫数据挖掘方法的分析与研究展望 [J]. 计算机工程与设计, 2005, 26(12): 3170-3174
- [8] 蔡伟杰, 张晓军, 朱建秋, 等. 关联规则挖掘综述 [J]. 计算机工程, 2001, 27(5): 31-33
- [9] 王磊, 潘进, 焦李成. 免疫算法 [J]. 电子学报, 2000, 28(7): 74-78
- [10] 李成武, 何学秋. 工作面煤与瓦斯突出危险程度预测技术研究 [J]. 中国矿业大学学报, 2005, 34(1): 71-76

(上接第 214 页)

- [9] Liao X J, Xue Y, Carin L. Logistic regression with an auxiliary data source [C] // Proceedings of the 22nd international conference on Machine learning. New York: ACM Press, 2005: 505-512
- [10] Zhang Y, Koren J. Efficient Bayesian hierarchical user modeling for recommendation systems [C] // Proceedings of the 30th annual international ACM SIGIR conference on Research and de-

velopment in information retrieval. New York: ACM Press, 2007: 47-54

- [11] Roy D M, Kaelbling L P. Efficient Bayesian task-level transfer learning [C] // Proc. of the 20th Joint Conference on Artificial Intelligence. New York: ACM Press, 2007: 2599-2604
- [12] Yu S P, Tresp V, Yu K. Robust multi-task learning with t-processes [C] // Proc. of the 24th international conference on Machine learning. New York: ACM Press, 2007: 1103-1110