

文章编号:1001-9081(2011)12-3302-03

doi:10.3724/SP.J.1087.2011.03302

# 基于粒子群优化模式搜索的支持向量机参数优化及应用

王喜宾<sup>1</sup>, 张小平<sup>2</sup>, 王翰虎<sup>1</sup>

(1. 贵州大学 计算机科学与信息学院, 贵阳 550025; 2. 贵州省科学技术情报研究所, 贵阳 550004)

(binxiwang@126.com)

**摘要:** 针对核函数参数选择的重要性, 提出了粒子群(PSO)模式搜索算法来搜索最优参数, 该算法结合了 PSO 算法的全局搜索能力强和模式搜索的局部收敛性好的优点, 使 PSO 模式搜索算法表现出了较高的性能, 并将其应用到农业科技项目分类中。实验结果表明, 该算法不仅效率高, 收敛速度快, 而且搜索到的最优参数达到了较高的准确率。

**关键词:** 支持向量机; 核参数选取; 粒子群优化; 模式搜索

**中图分类号:** TP181    **文献标志码:**A

## Parameter optimization of support vector machine and application based on particle swarm optimization mode search

WANG Xi-bin<sup>1</sup>, ZHANG Xiao-ping<sup>2</sup>, WANG Han-hu<sup>1</sup>

(1. School of Computer Science and Information, Guiyang University, Guiyang Guizhou 550025, China;

2. Guizhou Institute of Scientific and Technical Information, Guiyang Guizhou 550004, China)

**Abstract:** Considering the importance of selecting Kernel parameters, the Particle Swarm Optimization (PSO) model search algorithm was proposed to search optimal parameters. This method combined the global search capability of PSO algorithm and the good local convergence of mode search, that making PSO model search algorithm displays higher performance, and applied to an the practice of agricultural technological project classification. The results of experiment show that this method is not only efficient, but also catches the optimal parameters that have achieved higher accuracy.

**Key words:** Support Vector Machine (SVM); kernel parameters selection; Particle Swarm Optimization (PSO); mode search

## 0 引言

支持向量机(Support Vector Machine, SVM)是 Vapnik 等人根据统计学习理论中结构风险最小化原则, 为解决小样本学习问题提供的统一框架<sup>[1]</sup>。它能较好地解决非线性、过学习、高纬度分类问题, 具有良好的推广能力; 它能克服神经网络局部最优解、大样本、收敛速度慢等难题, 具有较强的实用性<sup>[2]</sup>。此外, 支持向量机是一个凸二次规划, 能够保证求得的极值就是全局最优解。

支持向量机的核心是核函数, 而核函数的参数选择决定了其分类性能, 所以, 参数的选择至关重要。以往支持向量机的参数选择都是凭借经验法或实验法, 这样不仅计算量大、效率低, 而且选取的参数往往不是全局最优参数, 限制了支持向量机的应用。为了克服这些缺陷, 一些学者提出了基于梯度下降的参数选择方法<sup>[3]</sup>、基于蚁群算法的参数选择方法<sup>[4]</sup>、基于网格的参数优化方法<sup>[5]</sup>、基于粒子群优化算法的参数选择方法<sup>[6]</sup>、基于遗传算法(Genetic Algorithm, GA)的参数优化方法<sup>[7]</sup>、基于基因表达式编程(Gene Expression Programming, GEP)的参数优化方法<sup>[8]</sup>。这些方法在对应文献的实验中证明了各自的有效性, 但是基于梯度下降的方法是一种线性搜索法, 在初始值选择不当时, 易陷入局部最优; 基于蚁群的方法存在初始信息素匮乏, 求解速度慢的缺陷; 网格方法存在计算量大、学习精度低的缺点; 基于粒子群优化(Particle Swarm

Optimization, PSO)算法的方法存在局部搜索能力差、过早收敛而陷入局部极值点的问题; 基于 GA 的方法存在操作复杂, 对不同的优化问题需要设计不同的交叉或变异方式; 基于 GEP 的方法存在收敛速度慢, 甚至不收敛的缺点。此外, 这些方法在样本数量较大、支持向量的维度较高时, 将导致计算量大, 训练时间长而不能达到实际需求。

考虑到以上参数优化方法存在的问题以及常用分类方法(例如神经网络、遗传算法、C4.5 等)的分类性能, 同时结合农业科技项目数据集的特点和支持向量机的优势, 本文提出了基于 PSO 模式的 SVM 参数优化并将其应用到农业科技项目分类实验中。该算法不但保持了基本 PSO 算法的强全局搜索能力、易实现等优点, 而且也保持了模式搜索的强局部搜索能力, 避免了算法陷入局部最小值的可能; 同时, 在迭代过程中利用模式搜索算法不断调整全局最优值, 提高了算法的收敛速度, 在较少的迭代次数内达到了较好的寻优效果。实验结果表明, 该算法在效率和性能方面都有了较大的提高, 且达到了较高的准确率, 为农业科技项目投入提供了很好的参考依据。

## 1 PSO 模式搜索算法

### 1.1 PSO 算法

在 PSO 算法中, 群体的每个成员被称为粒子, 每个粒子在多维搜索空间中飞行, 并不断根据粒子自己的经验、粒子邻

收稿日期:2011-05-08;修回日期:2011-07-13。    基金项目:贵州省科技计划项目(黔科合 NY 字[2010]3078 号)。

作者简介:王喜宾(1985-), 男, 河南洛阳人, 硕士研究生, CCF 会员, 主要研究方向:数据仓库、数据挖掘、智能计算; 张小平(1956-), 男, 贵州贵阳人, 研究员, 主要研究方向:数据挖掘、软件工程; 王翰虎(1946-), 男, 贵州遵义人, 教授, 主要研究方向:数据挖掘。

居的经验或整群体的经验更新自己的速度和位置,从而找到问题的最优解。对应的迭代公式是:

$$\begin{aligned} v_i^l &= w * v_i + c_1 * rd_1^l * (pBest - p_i^l) + \\ &\quad c_2 * rd_2^l * (gBest - p_i^l) \end{aligned} \quad (1)$$

$$p_i^l = p_i^d + v_i^l \quad (2)$$

其中,  $w$  是惯性权重, 相对大的  $w$  有更多全局搜索能力, 相对小的  $w$  会导致快速收敛;  $c_1$  和  $c_2$  是学习因子;  $rd_1^l$  和  $rd_2^l$  是区间  $[0,1]$  上的随机数;  $p$  是粒子的当前位置表示支持向量机参数  $c$  和  $\sigma$  的当前值;  $v \in [v_{\max}, v_{\min}]$  是粒子的速度, 决定下一代的  $c$  和  $\sigma$  更新方向和大小。

为了改善算法的收敛速度, 采用带有收缩因子  $\chi$  的速度进化方程(式(3)):

$$\begin{aligned} v_i^l &= \chi \{ w * v_i + c_1 * rd_1^l * (pBest - p_i^l) + \\ &\quad c_2 * rd_2^l * (gBest - p_i^l) \} \end{aligned} \quad (3)$$

PSO 算法具有大范围全局搜索能力; 搜索从群体出发, 具有隐并行性; 搜索使用评价函数值启发; 收敛速度快, 参数调整简单; 具有扩展性, 容易与其他算法结合等优点<sup>[9]</sup>。但是算法后期的局部搜索能力差, 反馈信息利用不充分, 使算法过早收敛而没有达到预期效果。

## 1.2 模式搜索算法

模式搜索算法又叫 Hooke-Jeeves 方法<sup>[10]</sup>, 算法主要由两个移动过程组成: 探测移动和模式移动。探测移动是沿着坐标轴的方向移动; 模式移动则是沿着相邻两个探测点连线方向上移动。两个过程也可以理解为: 1) 确定有利的搜索方向; 2) 加速过程, 也就是在有利方向上加速搜索。

模式搜索算法不要求目标函数可导, 是一种直接法, 学习精度高且迭代比较简单, 尤其是局部搜索能力强。但是存在迭代次数多、计算量大、容易使目标函数陷入局部最小点, 并且全局搜索能力差等缺点。

因此, 为了充分发挥模式搜索算法的优势, 结合 PSO 算法的优点, 提出了 PSO 模式算法。

## 1.3 PSO 模式搜索算法

从全局角度考虑, PSO 算法是一种可行的、鲁棒性很好的优化算法<sup>[11]</sup>。而从局部收敛角度考虑, 模式搜索算法具有局部收敛性好, 并且还有对初始值敏感等特点, 所以, 在 PSO 算法中引入模式搜索可以使 PSO 算法具有较强全局搜索能力的同时, 也可以摆脱陷入局部极小值的缺陷, 提高了算法的收敛速度和精度。

PSO 模式搜索算法的核心思想是: 先用 PSO 算法在全局范围内搜索, 找到一个较优参数, 再利用模式搜索算法在这个较优参数附近进行局部搜索, 若搜索结果没有达到精度要求, 则进行全局搜索, 依次循环直到找到最优参数。

# 2 PSO 模式在农业科技项目分类中的应用

## 2.1 数据预处理

农业科技项目数据集具有高维度、非线性和小样本的特点, 而高维度及其包含的噪声会降低分类器的性能。针对农业科技项目数据集的特点, 本文建立了基于支持向量机的农业科技项目分类模型。在建立分类模型前需要对数据进行预处理操作, 本文采用小波变换对数据集进行滤波去噪。

小波在信号去噪领域已得到越来越广泛的应用。阈值去噪方法是一种实现简单、效果较好的小波去噪方法<sup>[12]</sup>。先对农业科技项目数据集中缺失数据和错误数据进行处理, 再对数据集进行小波去噪。在小波去噪中选用 Battle-Lemarie 正

交小波和启发式阈值进行 3 层分解。

## 2.2 特征提取

为了避免高维数据给支持向量机带来的大运算量而导致维数灾难, 采用局部线性嵌入法对高维数据进行降维, 并且在某种意义上它也是一种新的特征提取方法<sup>[13]</sup>。

步骤 1 将农业科技项目数据输入式(4), 并设定  $k = 5$ , 降维后的维数  $p = 10$ , 样本数  $n = 500$ 。

$$\min_{w_i} \| \mathbf{x}_i - \sum_{j=0}^n w_{ij} \mathbf{x}_j \| \quad (4)$$

当  $x_j \notin Q_i$  时,  $w_{ij} = 0$ , 得解  $w_i^*$ 。其中  $Q_i$  为  $k$  个邻近点集合。

步骤 2 构造并求解以  $(n+1)$  个  $p$  维向量  $\bar{\mathbf{x}}_0, \bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_n$  为变量的最优问题:

$$\min_{\bar{\mathbf{x}}} \| (\mathbf{I} - \mathbf{W}^{*\top}) \bar{\mathbf{x}} \|^2 \quad (5)$$

所得到的解  $\hat{\mathbf{x}} = \bar{\mathbf{x}}_0^*$  和  $\bar{\mathbf{x}}_1^*, \bar{\mathbf{x}}_2^*, \dots, \bar{\mathbf{x}}_n^*$  为输入  $\mathbf{x}$  和  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  降维后的向量。

经特征提取后, 由 20 维降到 10 维, 降低了输入数据的维数, 提高了 SVM 的运算效率。

## 2.3 PSO 模式优化 SVM 参数

选择 Guess 核函数作为支持向量机的核函数, 利用 PSO 模式进行核参数优化, 并选用分类误差作为评价函数进行核参数选择。

PSO 模式搜索算法的步骤如下:

步骤 1 读取农业科技项目数据样本, 并随机产生  $(c, \sigma)$  作为粒子的初始位置  $\mathbf{P}_i$ 。

步骤 2 计算各粒子适应度函数值, 以分类误差(式(6))作为粒子的适应度评价函数。

分类误差公式:

$$1 - \frac{1}{n} \sum_{i=1}^n \left| \frac{T_i}{(T_i + F_i)} \right| \quad (6)$$

其中:  $T_i$  表示正确分类样本数,  $F_i$  表示错分样本数。

步骤 3 记录最小适应度函数值并根据式(2)和式(3)更新粒子的位置和速度。

步骤 4 令步骤 3 得到的  $\mathbf{P}_g = \mathbf{P}_g^k \in (c, \sigma)$  作为模式搜索的初始值, 并给定单位向量  $\mathbf{e}^j (j = 1, 2, \dots, n)$ 。初始步长  $\sigma^0 = (\sigma_1^0, \sigma_2^0, \dots, \sigma_n^0)^T > 0$ , 加速度系数  $\gamma > 0$ , 收缩系数  $\theta \in (0, 1)$  及精度  $\varepsilon > 0$ , 置  $k = 0, \mathbf{y} = \mathbf{P}_g^k, j = 1$ 。

步骤 5 从  $\mathbf{y}$  出发, 依次做平行于单位矢量  $\mathbf{e}^j (j = 1, 2, \dots, n)$  的轴向探测移动:

1) 若  $f(\mathbf{y} + \sigma_j^k \mathbf{e}^j) < f(\mathbf{y})$ , 则令  $\mathbf{y} = \mathbf{y} + \sigma_j^k \mathbf{e}^j$ , 否则转 2);

2) 若  $f(\mathbf{y} - \sigma_j^k \mathbf{e}^j) \geq f(\mathbf{y})$ , 则令  $\mathbf{y} = \mathbf{y} - \sigma_j^k \mathbf{e}^j$ , 否则令  $\mathbf{y} = \mathbf{y}$ ;

步骤 6 令  $\mathbf{P}_g^{k+1} = \mathbf{y}$ , 若  $f(\mathbf{P}_g^{k+1}) < f(\mathbf{P}_g^k)$ , 则对  $\mathbf{P}_g^{k+1}$  沿着加速方向  $\mathbf{p}^k = \mathbf{P}_g^{k+1} - \mathbf{P}_g^k$  做模式移动, 令  $\mathbf{y} = \mathbf{P}_g^{k+1} + \gamma \mathbf{p}^k$ ,  $\sigma^{k+1} = \sigma^k, k = k + 1$ , 转步骤 5, 否则转步骤 7。

步骤 7 若  $|\sigma^k| \leq \varepsilon$ , 则停止迭代, 输出结果。否则当  $\mathbf{P}_g^{k+1} \neq \mathbf{P}_g^k$  时, 令  $\mathbf{y} = \mathbf{P}_g^{k+1}, \sigma^{k+1} = \sigma^k, k = k + 1$ , 转步骤 5; 当  $\mathbf{P}_g^{k+1} = \mathbf{P}_g^k$  时, 令  $\mathbf{y} = \mathbf{P}_g^{k+1}, \sigma^{k+1} = \theta \sigma^k, k = k + 1$  转步骤 5; 若达到最大迭代次数时  $|\sigma^k| > \varepsilon$ , 则转步骤 2。

优化结果为误差惩罚参数  $c = 200, \sigma = 0.2$ 。

## 2.4 SVM 分类模型

将处理过的训练数据集和得到的最优参数输入 SVM 进行训练并得到训练模型, 即分类模型:

$$f(\mathbf{x}) = \operatorname{sgn} \left( \sum_{i=1}^l \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \right) \quad (7)$$

其中,核函数  $K(\mathbf{x}, \mathbf{x}') = \exp(-\|\mathbf{x} - \mathbf{x}'\|^2/\sigma^2)$ , 核参数  $\sigma$  为宽度系数。

把经过处理的测试数据集作为分类模型的输入进行分类预测。根据  $f(\mathbf{x})$  的符号可判别测试样本  $\mathbf{x}_i$  属于哪一类, 即可以判定哪些农业科技项目是收益项目, 进而对农业科技项目的投入做出科学决策。

### 3 实验结果及分析

从贵州省农业科技项目数据库<sup>[14]</sup>中选取数据 500 条, 其中综合收益项目有 400 条。将这 500 条数据分为两组: 一组为训练集, 共 300 条, 其中综合收益项目 240 条; 另一组为测试集, 其中综合收益项目有 160 条。通过对第一组数据进行训练, 并建立一个训练模型; 第二组是测试集, 用于检测训练模型的分类能力。

在实验中分别采用 PSO 算法、模式搜索算法和 PSO 模式搜索算法进行参数优化并建立分类模型进行分类检测。三种算法均采用 Matlab 7.1 实现, 系统平台为 Pentium4 (2.66 GHz) 处理器, Windows XP SP3, 512 MB RAM。三种算法对应的学习时间和准确率如表 1 所示。

表 1 三种算法的检测结果

算法	学习时间/min	准确率/%
PSO 算法	22.35	95.37
模式搜索算法	23.68	96.23
PSO 模式搜索算法	20.27	98.76

从表 1 可知, PSO 算法的训练时间比模式搜索算法的训练时间短, 但学习精度没有模式搜索算法高。而 PSO 模式搜索算法在训练时间和精确度方面都优于 PSO 算法和模式搜索算法, 该算法不但考虑到了群体历史最好位置对当前位置的影响, 增强了算法的局部搜索能力, 而且每次迭代都利用模式搜索算法对  $P_g$  进行优化和更新, 使更好的  $P_g$  引导整个种群向更好的领域搜索, 提高了搜索效率, 从而在保证 PSO 算法快速收敛的前提下也不影响其较强的全局搜索能力。总之, PSO 模式搜索算法把 PSO 算法和模式搜索算法的优点结合起来, 避其各自的缺点, 使 PSO 模式搜索算法表现出了较高的性能。

为了说明参数选择的重要性, 先利用 PSO 模式搜索算法找到最优参数组合, 再在最优参数附近取其他参数组合进行准确率比较, 不同参数组合对应的准确率如表 2 所示。

从表 2 中可以看出, 在最优参数组合附近的参数组合虽然也达到了较高的分类准确率, 但是没有利用 PSO 模式搜索算法找到的最优参数达到的分类准确率高。结果表明, 参数的选取对分类器的性能影响很大。其原因是  $c$  的作用控制了对错分样本的惩罚程度,  $c$  的值越小表示对错分样本的惩罚程度越低, 这时经验误差较大;  $c$  的取值趋向无穷大, 则所有的约束条件都必须满足, 这时意味着所有训练样本都要正确分类, 这样将导致分类超平面复杂、计算量大、耗时多。所以, 对  $c$  值的选择要结合具体的应用, 在满足分类准确率的同时, 尽可能取小的  $c$  值, 从而使决策(分类)函数简单。若  $\sigma$  取值过小, 则导致所有的测试样本成为支持向量, 产生“过度拟合

现象”; 若  $\sigma$  取值过大, 支持向量机的分类性能将大幅下降, 对测试样本的分类能力几乎为零, 将把所有的样本都判为一类; 若  $\sigma$  取值合适, 支持向量个数明显下降, 且分类器对测试样本的正确分类能力会大大提高。因此, 选择合适的参数组合才能构造出性能较高的分类器。

表 2  $c$  和  $\sigma$  不同组合下的准确率 %

$\sigma$	$c$			
	1	100	200	300
0.00	55.0	56.0	59.0	70.1
0.02	61.0	64.1	66.1	73.0
0.04	72.0	75.0	75.0	77.3
0.06	80.0	76.0	85.0	85.0
0.08	73.0	88.0	90.0	88.0
0.20	75.6	94.0	98.4	90.0
0.40	74.0	91.0	94.0	90.0
0.70	73.0	89.0	87.0	87.2
1.00	72.0	87.2	83.0	85.0
1.50	71.0	86.0	82.3	84.3
2.00	70.1	84.2	82.0	83.1

总之, 利用 PSO 模式搜索算法对具有高纬度、小样本、非线性特点的农业科技项目数据集进行参数优化是有效、可行的, 从而为支持向量机的参数选择提供了一种新的、有效的方法。

### 4 结语

PSO 模式搜索算法是将 PSO 算法和模式搜索算法的优点结合起来, 使 PSO 模式搜索算法具有较强的全局能力和局部搜索能力, 克服了 PSO 算法的局部搜索能力差和模式搜索算法易陷入局部最优的可能, 从而使 PSO 模式搜索算法更加完善。本文利用 PSO 模式搜索算法来优化核参数, 并将其运用到农业科技项目分类中, 不仅避免了依靠经验选取参数带来的局限性, 而且缩短了支持向量机的训练时间, 同时也提高了分类的准确率。所以, PSO 模式搜索算法是一种高效可行的参数优化方法。

#### 参考文献:

- [1] 白鹏, 张喜斌, 张斌, 等. 支持向量机理论及工程应用实例 [M]. 西安: 西安电子科技大学出版社, 2008: 13.
- [2] OSUNA E, FREUND R, FIROSI F. Training support vector machines: an application to face[C]// IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 1997: 130 – 136.
- [3] CHAPELLE O, VAPNIK V. Choosing multiple parameters for support vector machines[J]. Machine Learning, 2002, 46(1): 131 – 160.
- [4] 刘春波, 王鲜芳, 潘丰. 基于蚁群优化算法的支持向量机参数选择及仿真[J]. 中南大学学报: 自然科学版, 2008, 39(6): 1309 – 1313.
- [5] 李兵, 姚全珠, 罗作民, 等. 基于网格模式搜索的支持向量机模型选择[J]. 计算机工程与应用, 2008, 44(15): 136 – 138.
- [6] 任洪娥, 霍满冬. 基于 PSO 优化的 SVM 预测应用研究[J]. 计算机应用研究, 2009, 26(5): 741 – 744.
- [7] CHEN P W, JUNGYING W, HAHNMING L. Model selection of SVMs using GA approach[C]// Proceedings of 2004 IEEE International Joint Conference on Neural Networks. Piscataway: IEEE, 2004: 2035 – 2040.

(下转第 3326 页)

```

flag = false;
if(flag)
    context.write(_, key);

```

Reducer 若发现 value\_list 中任意一个三元组是原始数据, 则无输出, 因为新推导的数据和原始数据重复, 没必要存储。若 value\_list 中没有原始数据, 表明它们是推导出的重复数据, 仅输出一份。

整个推理过程只进行了一次迭代, 相对于 Jacopo Urbani 等的 WebPie<sup>[9]</sup>, 算法很简洁。

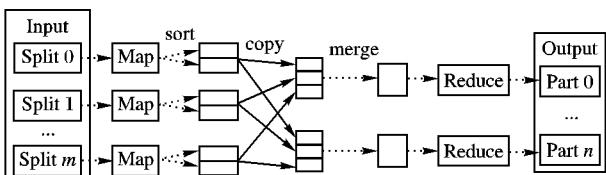


图 2 推理过程

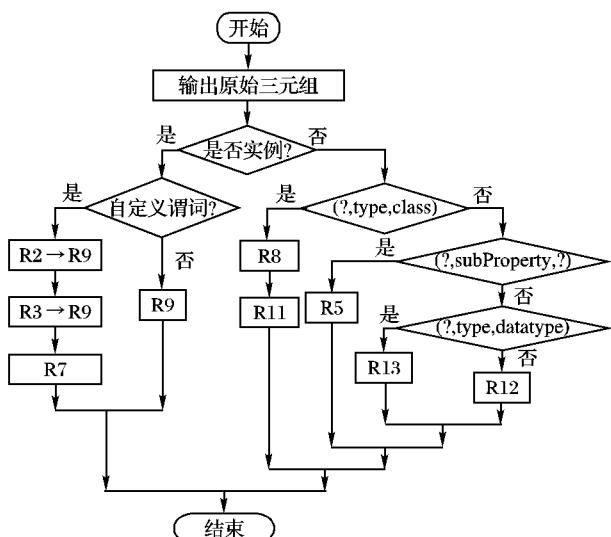


图 3 Map 处理过程

## 4 实验

实验环境由 30 台个人电脑组成, 中央处理器是双核 1.8 GHz, 2 GB 内存, 其中 DataNode 有 28 台, 使用 Hadoop-0.20.2 搭建实验平台。测试数据采用利哈伊大学基准 (Lehigh University Benchmark, LUBM) 生成, 是关于大学的语义数据。对不同数量的数据进行处理时, WebPie<sup>[4]</sup> 和本文算法的实验数据对比如表 2 所示, 表明本文算法相对 WebPie 性能上有很大提高。实验受到硬件环境的限制, 随着硬件环境的改善, 推理时间会进一步缩短。

## 5 结语

本文研究算法能够对数量以亿计的三元组进行推理, 解决了海量本体在若干分钟、甚至更短时间内可推理的问题。若配以适当的计算环境, 推理时间还可以进一步改善。但算

法支持的语义仅限于 RDFs 层次, 今后还可以进一步针对万维网本体语言 (Web Ontology Language, OWL) 本体展开研究。另外, 还可以针对 Hadoop 的 Map/Reduce 实现进行优化, 进一步提高计算效率。

表 2 WebPie 实验数据对比表

大学 数量	三元组数量	WebPie 推导时间/s	本文算法 推导时间/s
100	13 830 000	154.426	77.331
250	34 490 000	220.329	110.231
400	55 040 000	265.892	145.888
600	81 360 000	309.273	189.969
800	105 790 000	334.801	223.297

### 参考文献:

- [1] MCBRIDE B. Jena: A semantic Web toolkit [J]. IEEE Internet Computing, 2002, 6(6): 55–59.
- [2] SIRIN E, PARSIA B, BERNARDO E C, et al. Pellet: A practical OWL-DL reasoner [J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2007, 5(2): 51–53.
- [3] KIRYAKOV A, OGNYANOV D, MANOV D. Owllim: A pragmatic semantic repository for OWL [C]// Web Information Systems Engineering 2005. Berlin: Springer, 2005: 182–192.
- [4] WBITE T. Hadoop权威指南 [M]. 曾大响, 周微英, 译. 北京: 清华大学出版社, 2010: 9–12.
- [5] 刘鹏. 云计算 [M]. 北京: 电子工业出版社, 2010: 14–18.
- [6] SEIDENBERG J, RECTOR A. Web ontology segmentation: Analysis, classification and use [C]// Proceedings of the World Wide Web Conference. New York: ACM, 2006: 13–22.
- [7] SOMA R, PRASANNA V. Parallel inferencing for OWL knowledge bases [C]// Proceedings of the 37th International Conference on Parallel Processing. Washington, DC: IEEE Computer Society, 2008: 75–82.
- [8] OREN E, KOTOULAS S, ANADOTIS G, et al. Marvin: A platform for large-scale analysis of semantic Web data [EB/OL]. [2011-04-15]. <http://www.larkc.eu/marvin/bte2008.pdf>.
- [9] URBANI J, KOTOULAS S, OREN E, et al. Scalable distributed reasoning using Map/Reduce [C]// Proceedings of 8th International Semantic Web Conference. Berlin: Springer Press, 2009: 634–649.
- [10] QU ZHENXIN, TANG SHENGQUN. Deep semantic integration for information system [C]// CINC2009: International Conference on Computational Intelligence and Natural Computing. Washington, DC: IEEE Computer Society, 2009: 292–295.
- [11] HAYES P. RDF semantics [EB/OL]. [2011-02-15]. <http://www.w3.org/TR/rdf-mt/>.
- [12] HEFLIN J. SWAT projects: The Lehigh University benchmark (LUBM) [EB/OL]. [2011-02-15]. <http://swat.cse.lehigh.edu/projects/lubm/>.

(上接第 3304 页)

- [8] 王文栋, 钟智, 元昌安. 基于 GEP 的支持向量机参数优化 [J]. 广西师范学院学报, 2010, 27(2): 66–70.
- [9] 张小云, 刘允才. 高斯核支撑向量机的性能分析 [J]. 计算机工程, 2003, 29(8): 22–25.
- [10] 龚纯, 王正林. Matlab 最优化计算 [M]. 北京: 电子工业出版社, 2009: 133–134.
- [11] 王佳, 徐蔚鸿. 基于动量粒子群的混合核 SVM 参数优化方法 [J]. 计算机应用, 2011, 31(2): 501–503.
- [12] 程正兴. 小波与小波变换导论 [M]. 北京: 机械工业出版社, 2008: 61–62.
- [13] PANG YANWEI, LIU ZHENGKAI, YU NENGHAI. A new nonlinear feature extraction method for face recognition [J]. Neurocomputing, 2006, 69(7/9): 949–951.
- [14] 贵州省科学技术厅农村处. 贵州省星火计划管理办法 [DB/OL]. [2010-12-28]. <http://www.nec.gzst.gov.cn>.