

基于动态多种群粒子群支持向量机的短期负荷预测^{*}

李丹¹ 高立群¹ 王珂² 黄越¹

(东北大学信息科学与工程学院 沈阳 110004)¹ (辽宁省电力有限公司 沈阳 110000)²

摘要 针对标准粒子群优化(PSO)算法存在易陷入局部极值点的缺点,提出了一种基于物种概念的动态多种群粒子群优化算法(DMPPO)。在DMPPO中引入了物种概念,在进化过程中动态确定物种,利用种群多样性信息动态调整物种半径,通过物种对解空间的不同区域进行搜索,最终确定出各极值点。将DMPPO算法和支持向量机(SVM)相结合,形成了解决电力系统短期负荷预测问题的新方法(DMPPO-SVM)。在该方法中利用DMPPO算法来优化SVM中的参数,利用快速傅立叶变换(FFT)进行频谱分析并确定SVM的输入量。电力系统短期负荷预测的实际算例表明,与传统预测方法相比,该方法具有更高的预测精度和鲁棒性。

关键词 粒子群优化,动态多种群,物种,支持向量机,负荷预测

Short-term Load Forecasting Approach Based on Support Vector Machine with Dynamic Multi-population Particle Swarm Optimization Algorithm

LI Dan¹ GAO Li-qun¹ WANG Ke² HUANG Yue¹

(School of Information Science & Engineering, Northeastern University, Shenyang 110004, China)¹

(Liaoning Electric Power Company Limited, Shenyang 110000, China)²

Abstract Aiming at the precocious convergence problem of particle swarm optimization algorithm, a dynamic multi-population particle swarm optimization (DMPPO) algorithm is presented. In DMPPO algorithm, the notion of species is used to determine its neighborhood best values and the swarm population is divided into species subpopulations according to their similarity. Each of these species subpopulations is built around a dominating individual called the species seed. At each iteration step, species seeds are identified from the entire population and then adopted as neighborhood bests for these species groups separately. A strategy for adaptively changing the species radius based on population diversity information is proposed. During iterations, species subpopulations are able to simultaneously optimize toward potentially regions containing multiple optima. A new short-term load forecasting model based on SVM with DMPPO algorithm (DMPPO-SVM) is proposed in which the SVM's parameters are optimized by DMPPO algorithm and the input variables of the SVM are determined by fast Fourier transform(FFT). The example of electricity load data from California power market is used to illustrate the proposed DMPPO-SVM approach. The empirical results reveal that the DMPPO-SVM approach outperforms the other traditional model.

Keywords Particle swarm optimization, Dynamic multi-population, Species, Support vector machine, Load forecasting

随着电力系统的商品化和市场化,电力负荷预测的准确性对电力系统安全经济运行和国民经济发展具有越来越重要的意义。近年来,电力负荷的预测理论和方法有了很大的进展,各种新的理论和方法不断涌现,如时间序列法、回归分析法^[1]、专家系统和神经网络^[1]等。时间序列法、回归分析法计算量小,速度较快,但由于模型过于简单而无法模拟复杂多变的电力负荷。专家系统虽然在负荷预测中取得了一定成果,但其通用性较差,缺乏学习能力。神经网络模型应用最为广泛。但神经网络过分强调克服学习错误而泛化性能不强、网络的最终权重受初始值影响大等,其预测效果还不是很令人满意。随着预测理论不断发展,基于统计学习理论的支持向量机(support vector machine, SVM)方法^[2,3]被用于负荷预测中。但是, SVM在具体应用中存在一个突出问题,那就是如何设置影响算法的一些关键参数。文献^[4]提出使用

模拟退火算法选择 SVM 的参数,但该方法由于模拟退火算法本身的缺点,在搜索过程中容易陷入局部最优。

针对上述的问题,本文提出了动态多种群粒子群优化(DMPPO)算法,并将DMPPO算法用于SVM的参数优化,结合加州电力市场的实际负荷数据进行短期电力负荷预测,编制了基于Matlab语言的预测计算程序。算例仿真表明DMPPO-SVM方法用于短期电力负荷预测是有效可行的。

1 动态多种群粒子群优化算法

1.1 标准粒子群优化算法

由文献^[5]可知, PSO算法是一类基于群智能的随机优化算法,由Kennedy和Eberhart在1995年提出。每个粒子根据如下的公式来更新自己的速度和在解空间的位置

$$v_{id}^{(t+1)} = wv_{id}^{(t)} + c_1 r_1 (p_{id}^{(t)} - x_{id}^{(t)}) + c_2 r_2 (p_{gd}^{(t)} - x_{id}^{(t)}) \quad (1)$$

^{*} 基金项目:国家自然科学基金资助项目(60274009);教育部博士点基金资助项目(20020145007)。李丹 博士研究生,主要研究方向为智能优化及其在电力系统中的应用;高立群 教授,博士生导师,主要研究方向为智能优化、模式识别;王珂 博士,高级工程师;黄越 硕士研究生。

$$x_{id}^{(t+1)} = x_{id}^{(t)} + v_{id}^{(t+1)} \quad (2)$$

其中 $d=1, 2, \dots, n, i=1, 2, \dots, m, m$ 为种群规模, t 为当前进化代数, r_1 和 r_2 为均匀分布于 $[0, 1]$ 的随机数, c_1 和 c_2 为加速常数, 是调整粒子的自身经验与社会(群体)经验在其运动中所起的作用的权重。 ω 为惯性权重, 其值由下式来确定^[6]:

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} \times t \quad (3)$$

其中, $\omega_{\max}, \omega_{\min}$ 分别是 ω 的最大值和最小值; t, t_{\max} 分别是当前迭代次数和最大迭代次数。

1.2 动态多种群粒子群优化算法

PSO 算法在应用当中碰到的一个主要问题是算法运行过程中, 如果某粒子发现一个当前最优位置, 其他粒子将迅速向其靠拢。如果该最优位置为一局部最优点, 粒子群就无法在解空间内重新搜索, 从而陷入局部最优, 出现所谓的早熟收敛现象。针对这一问题, 本文提出 DMPSO 算法来解决 PSO 算法的早熟收敛问题。该算法可以以较高的概率搜索到解空间中的多个全局最优值。算法的相关概念及具体操作将在下面各节中详细介绍。

1.2.1 物种及物种的核

确定物种及物种的核的步骤如下:

①计算粒子群中的所有粒子的适应值, 按照适应值的大小做降序排列。例如种群包含 m 个粒子, 可表示为 $X = [x_1, x_2, \dots, x_m]$, 则他们的适应值为 $F = [f(x_1), f(x_2), \dots, f(x_m)]$, 对适应值进行排序, 不妨设 $f(x_1) \geq f(x_2) \geq \dots \geq f(x_m)$ 。设集合 S 为物种集合, 在开始迭代时集合 S 为空集。

②选取集合 F 中的最大值所对应的粒子作为物种的核记为 s_i^* , 从集合 X 中取出与物种 s_i^* 之间的距离小于物种半径 r_s 的粒子放入集合 S_i 中, S_i 为由核 s_i^* 所形成的物种。用公式表示如下: $s_i^* \in S_i$, 若 $\forall y \in S_i, \exists f(s_i^*) \geq f(y)$, 则称 s_i^*

是物种 S_i 的核。若 $\forall y \in S_i, \exists d(s_i^*, y) \leq \frac{r_s}{2}$, 则称物种 S_i 以核 s_i^* 为中心。其中 $d(\cdot, \cdot)$ 为欧氏距离, r_s 为物种半径。当集合 X 中的某个粒子与集合 S 中的所有核之间的距离都大于 r_s 时, 则将此粒子作为新的物种的核 s_{i+1}^* 放入集合 S 中。

③判断集合 X 是否为空集, 是则说明物种的核的确定及物种生成完毕, 否则回到第②步继续。

通过以上步骤就将初始的种群初步划分成了若干物种。

1.2.2 物种半径的自适应调整

如果在进化过程中将物种半径 r_s 设为固定值, 若 r_s 较大则划分的物种较少, 一些重要个体由于没有成为核而不能被保留, 从而有可能在进化过程中被淘汰掉; 另一方面, 若 r_s 较小则划分的物种较多, 在进化过程中就可能形成大量的不必要的种群, 大大增加了算法的计算量。本文提出物种半径的自适应调整策略, 其基本思想为: 在进化过程的初期物种半径较大, 对整个粒子群进行粗分; 随着种群的不断进化物种半径不断减小, 逐步将整个粒子群进行细分。调整时充分考虑整个种群的多样性信息, 当种群多样性较差时, 种群中的粒子分布比较密集, 应减小种群距离, 使每个物种尽量只包含一个极值点; 而当种群多样性较好时, 种群中的各粒子比较分散, 应增大物种半径, 减少计算量。本文提出的物种半径调整公式为:

$$r_s(t) = [1 - e^{-D(t)}] r_s(0)$$

$$D(t) = \frac{1}{k} \sum_{i=1}^k D_i(t)$$

$$D_i(t) = \frac{1}{m_i} \sum_{j=1}^{m_i} \sqrt{\sum_{d=1}^n (p_{jd} - \bar{p}_d)^2} \quad (4)$$

式中, $r_s(0)$ 为初始物种半径, $D(t)$ 为整个种群的多样性指数, 多样性指数值较小时, 说明种群多样性较差, 多样性指数较大时, 说明种群多样性较好。 $D_i(t)$ 为第 i 个物种的多样性指数, m 为整个种群的粒子个数, $m_i (1 \leq i \leq k)$ 为第 i 个物种的粒子个数, k 为物种的个数, p_{jd} 为第 j 个粒子位置向量的第 d 个分量, \bar{p}_d 为物种内各粒子位置向量的第 d 个分量的平均值。

1.2.3 物种成员调整

在物种核的动态辨识与物种形成的过程中, 与核之间的距离小于物种半径 r_s 的个体并不一定都属于该核所形成的物种。如图 1 所示, 图中表示的是当进化到某一代时种群中个体的分布状态, 横坐标表示个体所处的位置, 纵坐标表示的是个体的适应值。由图中可以看出, 个体 a, b, c 位于同一峰上, 而个体 d 位于另一峰上。这里的峰被认为是包含了极值点(局部极值点或全局极值点)的区域。如果个体 b 被确定为核, 则若只按照距离标准来衡量, 则个体 d 应和 b, c 属于同一物种, 而个体 a 有可能被划分到另一物种中。个体 d 对 A 峰的搜索几乎不起作用, 但是对于搜索 B 峰却能起到很大的作用, 所以这种划分物种的方法有可能使 B 峰错过搜索, 而造成“漏峰”的后果。所以有必要对 1.2.2 节形成的物种作出必要的调整, 尽量使每个物种中的粒子都位于同一峰上。

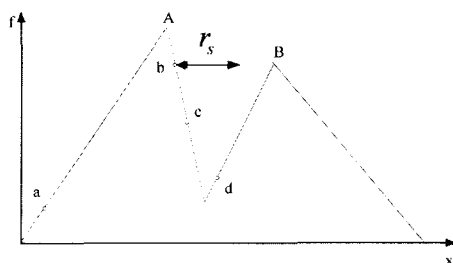


图 1 物种分布示意图

本文使用同峰判别法来对物种进行调整, 其具体步骤如下:

①以求解函数的全局最大值为例, 在进化的某一代中, 假设个体 s^* 被确定为物种的核, 则在以 s^* 为中心, 以 r_s 为半径的区域内共用 m 个个体, 分别记为 x_1, x_2, \dots, x_m , 每个个体分别用两个量来描述, 分别为与物种的核之间的距离 r_i 及个体的适应值 $f_i, 1 \leq i \leq m$ 。通常情况下, 如果个体与物种的核位于同一个峰上, 则个体与核之间的距离变大时, 其适应值应该变小, 反之, 此个体记为可疑个体;

②分析核与可疑个体分别趋向于高适应值的方向, 对可疑个体做进一步的检测。核与可疑个体唯一地确定了一条直线, 沿着这条直线的方向, 核与可疑个体的趋向于高适应值的方向只能有一个。核与可疑个体的趋向于高适应值方向的组合有三种情况: 同向、反向和相向, 如图 2 所示。如果此时核与可疑个体的趋向于高适应值方向的组合情况为反向, 则可以确定此可疑个体和物种的核不位于同一峰上, 应从物种内去掉, 否则转至③;

③如果核与可疑个体的趋向于高适应值方向的组合情况为同向或者相向则利用峰谷函数^[7]做进一步的处理, 如图 2 所示。简言之, 从核与可疑个体所确定的直线上按照一定的步长选择一定数量的采样点, 记为 $s_i, 1 \leq i \leq L$, 令 s_f 为可疑个

体的位置, $s_i = s_f + i(s^* - s_f)/L$, 则采样点的适应值为 $f(s_i)$, 则 $f_{\max} = \max(f(s^*), f(s_f))$, $f_{\min} = \min(f(s^*), f(s_f))$ 若 $\forall f(s_i) < f_{\max}$ 或 $\forall f(s_i) > f_{\min}$, 则物种的核与可疑个体位于同一峰上, 否则, 位于不同峰上, 将该个体从此物种内去除。

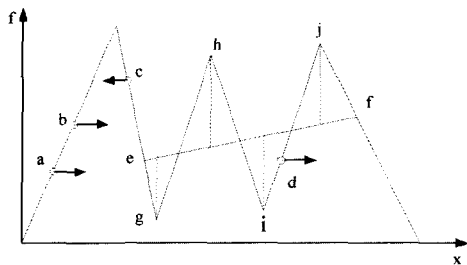


图2 同峰判别示意图

图2中个体a、b、c和d上的箭头表示趋向于高适应值的方向, 由图可知个体a和b的趋向于高适应值的方向为同向的, 个体a和c为相向的, 个体c和d为反向的。在图2中若令个体e为核, 个体f为可疑个体, 则 $f_{\max} = \max(f(e), f(f))$, $f_{\min} = \min(f(e), f(f))$, 若采样点g、h、i和j的适应值同时小于 f_{\max} 或同时大于 f_{\min} , 则核与可疑个体位于同一峰上, 否则, 位于不同峰上。

④处理完全部可疑个体后, 对物种边缘地区的个体进行同峰判别。令物种的核与种群内个体之间的距离为 r_i , 则将满足 $ar_s < r_i < r_s, a \in [0, 1]$ 的个体定义为位于物种的边缘地区的个体。将边缘地区的个体列为边缘个体, 对所有边缘个体重复步骤②和③;

⑤如发现边缘个体中存在与核不属同一峰者, 转至⑥, 否则转至⑦;

⑥令 $r_s = ar_s, a \in [0, 1]$ 对满足 $ar_s < r_i < r_s, a \in [0, 1]$ 的个体重复④;

⑦对下一个物种的核重复步骤①~⑥的操作, 直至完成全部物种的调整。

1.2.4 物种规模调整及物种合并

由上面的叙述可知, 物种形成后, 各物种的规模有时会相差很大, 当某物种的规模很大时势必会造成计算量的增大也会影响搜索效果, 而当物种的规模很小时也会对粒子群算法的性能产生很大的影响, 为此我们将对每个物种的规模 m 加以限定, 设 m_{\max} 和 m_{\min} 分别为物种规模的上限和下限, 即 $m_{\min} \leq m \leq m_{\max}$ 。当物种的规模大于最大值时, 则计算物种内各个体的当前适应值和个体最优适应值之和, 保留适应值之和较大的 m_{\max} 个个体, 其余的舍弃; 当物种的规模小于最小值时, 在此物种的范围内随机生成粒子, 使物种规模等于最小值。

对物种规模调整好的物种分别按照公式(1)-(3)进行进化, 此时物种的核的位置即为此物种的最优位置 p_g 。此时各个物种在其核的引导下将对不同的区域进行搜索, 从而收敛到多个不同极值点(包括全局极值点和局部极值点)的附近, 所期望得到的结果是在每个物种的范围内都存在一个潜在的极值。此时就需要判断各物种是否位于同一峰上, 对多余的物种进行合并或删除。具体的做法是对不同物种的核利用前面所提的同峰判别法进行判断。若两个物种的核位于同一峰上, 则认为这两个物种位于同一峰上, 此时保留适应值更优的核所在的物种, 删除适应值较差的物种。这样处理为后面要进行的精细搜索扫除了不必要的搜索区域, 提高了寻优速度。

按以上原则设计的算法不仅可以在可行域内寻找全局最优解, 而且还可以搜索到多个全局最优解和有意义的局部最优解, 从而为决策者提供多种选择或者多方面的信息。

2 持向量机回归模型

支持向量机^[8,9]近年来推广到了回归问题, 这使得将支持向量机引入短期负荷预测领域成为可能。

对于训练样本集 $\{(x_i, y_i)\}_{i=1}^n, x_i \in R^n$ 为输入变量的值, $y_i \in R$ 为相应的输出值, n 为训练样本个数, 函数回归问题就是寻找一个从输入空间到输出空间的映射 $f: R^n \rightarrow R$, 使得 $f(x) \approx y$ 。SVM回归方法就是先用一个非线性映射 $\phi: R^n \rightarrow R^m (m \geq n)$, 将输入空间映射到高维的特征空间, 再在特征空间中用下述线性函数来拟合数据:

$$y = f(x) = \langle W, \phi(x) \rangle + b \quad (5)$$

式中 $W, \phi(x)$ 维 m 维向量; $\langle \cdot, \cdot \rangle$ 表示特征空间中的点积; b 为阈值。

根据统计学理论, SVM回归方法通过极小化目标函数来确定系数 W, b , 即确定回归函数式(6)。

$$\min W, b: \frac{1}{2} \|W\|^2 + C \sum_{i=1}^n |y_i - \langle W, \phi(x_i) \rangle - b|_\epsilon \quad (6)$$

在式(6)给出的正则化风险泛函中, 第一部分 $\frac{1}{2} \|W\|^2$

是正则化部分。第二部分 $C \sum_{i=1}^n |y_i - \langle W, \phi(x_i) \rangle - b|_\epsilon$ 是经验风险, 其中 C 是正常数, 它决定着经验风险与正则化部分之间的平衡, 称为平衡系数(或惩罚因子); $\phi(x)$ 是从输入空间到高维特征空间的非线性映射; $|y - \langle W, \phi(x) \rangle - b|_\epsilon$ 称为 ϵ 不敏感损失函数。

为了寻找系数 W 和 b , 需要引入松弛变量 ξ 和 ξ^* , 式(6)变为:

$$\frac{1}{2} \|W\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (7)$$

其约束条件为:

$$\begin{cases} W\phi(x_i) + b_i - y_i \leq \epsilon + \xi_i^*, & \xi_i^* \geq 0 \\ y_i - W\phi(x_i) - b_i \leq \epsilon + \xi_i, & \xi_i \geq 0 \end{cases}$$

使式(7)最小, 便能获得最佳回归函数。其中 C 是一个先确定的正常数, ξ_i 和 ξ_i^* 是表征系统输出上下限的松弛变量。

SVM通过引入点积核函数 $K(x_i, x_j)$ 和利用 Wolfe 对偶技巧来求解式(6)。本文使用的核函数为 RBF 核函数, 定义如下:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$$

引入拉格朗日乘子 α_i 和 α_i^* , 对任意 $i=1, 2, \dots, n$, 都有 $\alpha_i \alpha_i^* = 0, \alpha_i \geq 0, \alpha_i^* \geq 0$ 成立。这样就把式(6)简化为对一个二次优化寻找向量 W 的问题。要找到所求的 $W = \sum_{i=1}^n (\alpha_i^* - \alpha_i)x_i$, 就要找到最大化二次型:

$$\max_{\{\alpha_i, \alpha_i^*\}}: -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*)$$

$$\text{约束条件: } \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases}$$

式中 α_i, α_i^* 为对偶参数。

将解出的 W 代入式(5)得:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

根据 KKT 条件, 在最优处满足拉格朗日乘子与约束式得乘积为 0, 得:

$$\begin{cases} \alpha_i (\epsilon + \xi_i - y_i + \langle w, x_i \rangle + b) = 0 \\ \alpha_i^* (\epsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) = 0 \\ (C - \alpha_i) \xi_i = 0 \\ (C - \alpha_i^*) \xi_i^* = 0 \end{cases}$$

由此可得出 b 的计算公式。

SVM 中需要确定的参数有平衡参数 C , 不敏感性参数 ϵ 及核函数参数 σ 等。研究中发现, 参数 C, σ 和 ϵ 的选择, 对于 SVM 回归估计的精度影响很大。其中参数 C 太小会对训练数据造成欠学习现象, C 值太大容易对训练数据造成过学习现象而导致泛化性能恶化。参数 ϵ 表明了系统对估计函数在样本数据上误差的期望, 其值影响了构造回归函数的支持向量数目。 ϵ 值过大, 回归估计精度低, 支持向量数量少, 会导致过于平滑的估计。 σ 的值太小, 会对训练集造成过学习现象, σ 的值太大, 会对训练集造成欠学习现象。由此可见 σ 对 SVM 的泛化性能起着关键作用。如何合理选择 SVM 的参数, 目前尚未有有效的方法, 一般通过交叉验证试算或梯度下降法求解, 这些方法或者人为因素太多, 或者要求函数连续可导, 易陷入局部极小, 严重影响了 SVM 的实际应用。为此, 本文利用上节所提的 DMPSO 算法来优化 SVM 的参数。

3 基于 DMPSO-SVM 算法的短期负荷预测

3.1 DMPSO-SVM 算法流程

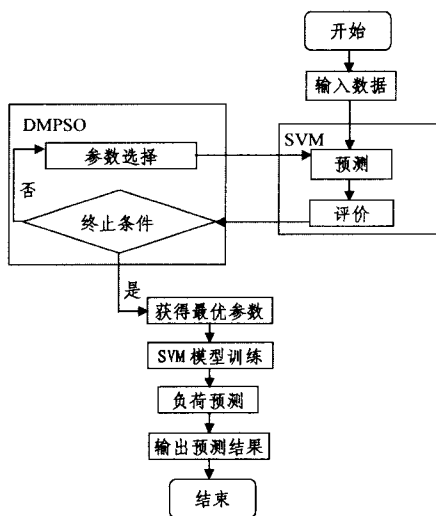


图 3 算法流程

3.2 历史负荷数据预处理

电力负荷预测是依赖于大量历史资料及其相关因素资料的被动型预测, 预报者占有资料的翔实可靠程度, 对基于任何方法的预报结果都会有很大影响。在电力系统实际运行时, 数据采集系统中的量测、记录、转换、传输过程的任意环节都可能引起故障而导致观测数据的反常态势, 以致与大多数观测值不一致; 另一方面, 当数据采集系统正常, 由于特殊事件引起负荷的异常变化, 也会导致观测数据的违背常规。所有这些非正常的观测数据统称为异常值。如果这些异常数据得不到有效的校正, 它们将以伪信息、伪变化规律的方式提供给负荷预测作为参考, 必然误导负荷预测模型的建立, 影响预测结果的精

确度及可靠性。本文采用文献[10]中提出的方法对历史负荷数据进行预处理, 使其更能体现电力负荷的运行趋势。

3.3 历史负荷数据的频谱分析

文献[11]的研究表明负荷变化将存在波动性, 所以各类用电负荷信号一般都表现为连续频谱。文献[11]通过详细研究表明, 电力负荷特性同样可以采用时频分析工具进行分解并分析, 从而得到预测结果。本文选取加州电力市场提供的 2003 年的负荷数据为例进行负荷的频谱分析。快速傅立叶变换(FFT)已被广泛用于电力系统的频谱分析中。本文将 2003 年的所有小时负荷数据构成的一个数据序列作为 FFT 的输入, 得到隐含与负荷数据中的周期性特性。由分析可知, 2003 年的负荷数据中存在 8 个主导频率, 其周期分别为: 1008h(6 周), 504h(3 周), 336h(2 周), 168h(1 周), 24h, 12h, 6h, 3h。

3.4 基于频谱特性的输入变量选取

由上述的频谱分析结果可以确定 SVM 的输入变量。将第 i 时刻的负荷记为 $L(i)$, 则本文所选择的 SVM 负荷预测模型的输入变量如下: $L(i-3), L(i-6), L(i-12), L(i-24), L(i-168), L(i-336), L(i-504), L(i-1008)$ 和气温 T 。

4 实例分析

利用本文所提方法(DMPSO-SVM)和文献[1]中的方法(BPNN)和(RM)分别对加州电力市场 2003 年的历史负荷数据进行分析, 其中, 将 2003 年 5-7 月的负荷数据作为训练数据, 2003 年 8 月 1 日至 2003 年 8 月 14 日的负荷数据作为测试数据, 2003 年 8 月 15 日的电力负荷预测结果如表 1 所示, 三种方法的预测曲线与实际负荷曲线如图 4 所示。

本文中的 DMPSO 算法的参数如下: 初始种群规模为 50; 最大种群规模为 100; $m_{max} = 10$; $m_{min} = 3$; $\omega_{max} = 0.9$; $\omega_{min} = 0.4$; $c_1 = c_2 = 2$ 。SVM 中的三个参数的取值范围为: $\sigma \in (0, 10]$; $\epsilon \in (0, 0.8]$; $C \in (0, 1000]$ 。

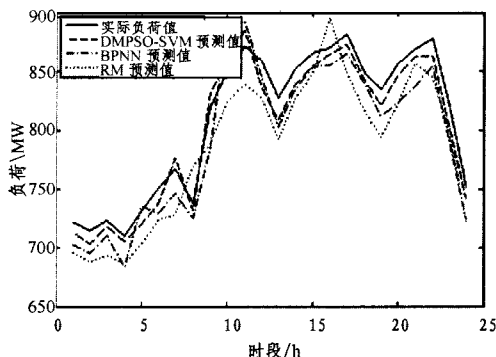


图 4 不同方法的预测结果

文中取相对误差(RPE)、平均相对误差(MAPE)和均方根相对误差(RMSE)作为各种方法预测效果的主要评价指标, 分别定义如下:

$$RPE = \left| \frac{L(i) - \hat{L}(i)}{L(i)} \right| \times 100$$

$$MAPE = \frac{1}{M} \sum_{i=1}^M \left| \frac{L(i) - \hat{L}(i)}{L(i)} \right| \times 100$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M \left| \frac{L(i) - \hat{L}(i)}{L(i)} \right|^2} \times 100$$

(下转第 172 页)

Apriori 算法。因此,本文提出的 CBA 算法不仅有效减少数据库的重复扫描次数,还减少了事务的比较运算量,节省时间与存储空间,在挖掘大量的事务数据库时将发挥很大的优势。

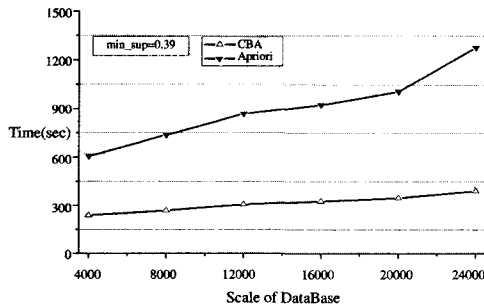


图 5 不同规模事务数据库的效率比较 (min_sup=0.39)

参考文献

[1] Agrawal R, Imilienski T, Swami A. Mining Association Rules between Sets of Items in Large Databases // Proc. of the ACM SIGMOD Int'l Conf. on Management of Data, 1993:207-216
 [2] Agrawal R, Srikant R. Fast Algorithm for Mining Association

Rules in Large Databases // Proc. 1994 Int'l Conf. VLDB. Santiago, Chile, Sep. 1994:487-499
 [3] Agrawal R, Srikant R. Mining Sequential Patterns // Proc. of the Int'l Conference on Data Engineering (ICDE). Taipei, Taiwan, March 1995. Expanded version available as IBM Research Report RJ9910, Oct. 1994
 [4] Berzal F, Cubero J C, Marin N, et al. TBAR: An efficient method for associatin rule mining in relational databases. Elsevier, Data & Knowledge, Engineering, 2001, 37:47-64
 [5] Brin S, Motwani R, Ullman J D, et al. Dynamic Itemset Counting and Implication Rules for Market Basket Data // ACM SIGMOD Conference on Management of Data, 1997:255-264
 [6] Brin S, Motwani R, Silverstein C. Beyond Market Baskets: Generalizing Association Rules to Correlations // 1997 ACM SIGMOD Conference on Management of Data, 1997:265-276
 [7] Cabena P, Hadjinian P, Stadler R, et al. Discovering Data Mining From Concept to Implementation. Prentice-Hall Inc., 1997
 [8] Carter C, Hamilton H, Cercone N. Share Based Measures for Itemsets, Principles of Data Mining and Knowledge Discovery // J. Komorowski and J. Zytkow, eds, 1997:14-24
 [9] Chen M S, Han J, Yu P S. Data Mining: An Overview from a Database Perspective. IEEE Transactions on Knowledge and Data Engineering, 1996, 8(6)
 [10] Chen M S, Park J S, Yu P S. Efficient Data Mining for Path Traversal Patterns. IEEE Transactions on Knowledge and Data Engineering, 1998, 10(2):209-221

(上接第 136 页)

以上公式中, $L(i)$ 表示实际负荷, $\hat{L}(i)$ 表示负荷预测值, M

表示一天内负荷数据由 M 个数据点组成, 这里为 24 点。

表 1 不同方法的预测误差比较

时段	DMPSO-SVM			BPNN	RM	时段	DMPSO-SVM			BPNN	RM		
	σ	ϵ	C				RPE _D	RPE _B	RPE _R			σ	ϵ
1	2.8025	0.0835	279.6	1.25	2.67	3.60	14	9.4189	0.0493	548.2	1.52	1.97	2.93
2	9.9981	0.0711	961.1	1.53	2.70	3.67	15	4.2546	0.0586	248.5	1.25	1.31	1.91
3	7.1816	0.0529	639.8	0.74	1.83	4.08	16	9.9997	0.0299	1.2	0.72	1.77	2.95
4	9.0026	0.0619	495.8	0.67	3.70	3.35	17	10.000	0.0214	18.5	1.01	1.86	3.66
5	4.1808	0.0617	405.1	1.52	0.59	3.87	18	7.5609	0.0703	537.2	0.77	1.11	3.76
6	3.8819	0.0101	19.2	1.77	2.75	3.63	19	9.9086	0.0392	84.6	1.60	2.65	4.85
7	6.7168	0.0545	333.6	1.20	2.71	5.06	20	5.2874	0.1044	503.8	1.12	3.86	3.92
8	3.5551	0.0511	1000	0.90	1.74	4.22	21	3.3533	0.0972	679.8	0.70	3.59	1.33
9	5.1204	0.0281	799.1	1.46	4.25	3.82	22	1.2635	0.0618	237.7	1.66	2.66	3.78
10	4.4169	0.0272	547.3	1.21	4.49	2.98	23	4.4897	0.0531	949.3	2.33	3.83	2.84
11	5.7473	0.0505	344.2	1.60	2.48	3.66	24	1.2368	0.0647	951.8	1.42	3.69	3.98
12	10.000	0.0361	594.2	2.61	1.59	3.70	MAPE				1.37	2.61	3.57
13	4.7601	0.0686	556.9	2.26	3.00	4.31	RMSE				1.46	2.81	3.66

由表 1 可知, DMPSO-SVM 方法的最大相对误差为 2.61%, 此结果明显好于 BPNN 方法和 RM 方法的 4.49% 和 5.06%。DMPSO-SVM 方法的平均相对误差和均方根相对误差也明显优于另外两种方法, 这说明 DMPSO-SVM 方法在预测短期电力负荷时, 预测的精度和预测的稳定性均强于另外两种方法。可见, 将本文提出的方法应用于电力系统短期负荷预测中是有效可行的。

结束语 在对支持向量机的参数性能进行分析的基础上, 本文提出了一种基于动态多种群粒子群支持向量机的电力系统短期负荷预测模型, 其中利用动态多种群粒子群优化算法来优化支持向量机的参数, 克服了 SVM 参数选择的盲目性; 利用快速傅立叶变换得到隐含于负荷数据中的周期性特性, 从而基于频谱特性确定预测模型的输入变量。电力系统短期负荷预测的实际算例表明, 与人工神经网络及回归分析模型相比, 本文所提出的方法具有更高的预测精度和鲁棒性。

参考文献

[1] Hsu C C, Chen C Y. Regional load forecasting in Taiwan: applications of artificial neural networks [J]. Energy Convers. Manag., 2003, 44(12):1941-1949

[2] 唐亮贵, 程代杰. 基于小波的支持向量机预测模型及应用 [J]. 计算机科学, 2006, 33(3):202-204
 [3] 李元斌, 方廷健, 于尔铿. 短期负荷预测的支持向量机方法研究 [J]. 中国电机工程学报, 2003, 23(6):55-59
 [4] Feng Pai Ping, Chang Hong Wei. Support vector machines with simulated annealing algorithms in electricity load forecasting [J]. Energy Conversion and Management, 2005, 46(3):2669-2688
 [5] Kennedy J, Eberhart R. Particle swarm optimization [C] // Proceedings of IEEE International Conference on Neural Networks. Perth: IEEE Press, 1995:1942-1948
 [6] Eberhart R C, Shi Y. Comparing inertia weights and constriction factors in particle swarm optimization [C] // Proceedings of the Congress on Evolutionary Computing. IEEE Service Center, California USA, 2000:84-88
 [7] Ursem R K. Multinational evolutionary algorithms [C] // Proceedings of Congress of Evolutionary Computation. Washington, DC, USA, 1999(3):1633-1640
 [8] 张浩然, 韩正之, 李昌刚. 支持向量机 [J]. 计算机科学, 2002, 29(12):135-137
 [9] 李华, 张简政. 基于模糊支持向量机的网络入侵检测研究 [J]. 计算机科学, 2005, 32(11):77-80
 [10] Yang Jingfei, Stenzel J. Historical load curve correction for short-term load forecasting [C] // Proceedings of The 7th International Pwer Engineering Conference. Singapore, IEEE Press, 2005:1-6
 [11] 邵能灵. 小波分析在电力系统中的应用及相关问题研究 [D]. 上海: 上海交通大学, 2002