Power load forecasting using support vector machine and ant colony optimization

2010

本文利用支持向量机和蚁群优化建立了电力负荷预测系统。集群优化方法用于处理大量数据并消除冗余信息。该系统挖掘历史日负荷，其具有与预报日相同的气象类别，以便组成具有高度相似气象特征的数据序列。通过这种方法，我们减少了SVM训练数据，克服了构建SVM模型时数据量大，处理速度慢的缺点。本文提出了一种基于蚁群优化的新特征选择机制，试图克服上述困难。然后应用该方法在模糊粗略数据简化过程中找到最佳特征子集。目前的工作应用于复杂系统监测，蚁群优化可以比原始模糊粗糙方法，基于熵的特征选择器和基于变换的简化方法PCA更全面，更准确地挖掘数据。与单个SVM和BP神经网络在短期负荷预测中的比较，这种新方法可以实现更高的预测精度。它表示当信息预处理基于数据挖掘技术时，SVM学习系统具有优势。

功率负荷预测引起了实践和学术界的极大关注。短期电力负荷预测对于电网的可靠性和经济发展非常重要。由于短期电力负荷预测对电网的可靠性和经济利用至关重要，因此它越来越受到实践和学术界的关注。负荷预测的目的是充分利用电能，缓解供需矛盾。

负荷预测已成为电力系统运行规划者和研究人员的一个关键问题。对于电力负荷依赖，电力供应商面临需求市场日益激烈的竞争，必须更加关注电力质量，包括单位承诺，热液协调，短期维护，交换和交易评估，网络电力流量调度优化和安全策略。

对电力负荷的不准确预测将导致电力公司的巨大损失。 Bunn和Farmer指出，预测误差增加1％意味着运营成本增加了1000万（Pai＆Hong，2005）。短期预测是指电力负荷需求的每小时预测，提前1小时至数天。在某些情况下，预测每日峰值负荷是短期负荷预测的目标，因为它是任何一天中最重要的负荷。短期小时负荷预测的质量对电力公用事业的经济运行具有重大影响，因为许多决策都是基于这些预测。这些决策包括发电容量的经济调度，燃料采购的安排，系统安全评估和能源交易计划。由于放松管制和竞争导致公用事业行业结构发生巨大变化，因此未来准确负荷预测的重要性将会增加。这种环境迫使公用事业公司以尽可能高的效率运行，如前所述，这需要准确的负荷预测。负载具有复杂和非线性的关系，有几个因素，如天气因素，气候条件，社会活动和季节性因素，过去的使用模式，星期几和一天中的时间。

一些研究人员在负荷预测模型中考虑了季节温度和日间类型等相关因素。 Christianse（1971）和Park，Park和Lee（1991）通过傅里叶级数变换设计了指数平滑模型，用于电力负荷预测。 Mbamalu和El-Hawary（1993）提出了多重自回归（AR）模型，该模型考虑了负荷​​预测中的季节性因素。分析结果表明，所提模型的预测精度优于不变模型。为了实现精确的负荷预测，Brown（1983）和Gelb（1974）建立了负荷预测模型，采用状态空间和卡尔曼滤波技术，这些技术旨在减少实际负荷和预测负荷之间的差异。状态空间和卡尔曼滤波技术将负载的周期性分量作为随机过程引入，并且需要覆盖超过3 - 10年的历史数据来建立周期性负载变化，以估计电力系统的因变量（负载或温度）。 。 Moghram和Rahman（1989）设计了一个基于状态空间和卡尔曼滤波技术的模型，并验证了所提出的模型优于其他四种预测方法。然而，这种方法流总是不能避免预测中观察噪声的影响。为了解决这个问题，Douglas，Breipohl，Lee和Adapa（1998）考虑验证温度对预测模型的影响。 Sadownik和Barbosa（1999）提出了用于负荷预测的动态非线性模型。这些方法的主要缺点是随着变量数量的增加，它们在计算上耗费时间。已经进行了非常有限的研究，以提供有效的方法来考虑时间复杂性和负荷预测的各种影响因素。因此，在负荷预测任务中面临大规模计算时，特征选择是重要的。

特征选择（FS）的主要目的是确定问题域中的最小特征子集，同时保持表示原始特征的适当高精度（Dash和Liu，1997）。群体智能（SI）是一种系统的属性，简单代理与其环境在本地交互的集体行为导致出现连贯的功能性全局模式（Bonabeau，Dorigo，＆Theraulez，1999）。 SI提供了一种解决集体（或分布式）问题的工具，无需集中控制或提供全局模型。例如，蚂蚁能够在不使用视觉信息的情况下找到食物来源与其巢穴之间的最短路径，因此不具备适应环境变化的全球世界模型。基于用于解决离散优化问题的真实蚁群行为的SI技术是蚁群优化（ACO）技术（Bonabeau等，1999）。这些已成功应用于大量难以解决的组合问题，如二次分配（Maniezzo＆Colo-mi，1999）和旅行推销员（Dorigo，Maniezzo，＆Colomi，1996）。该方法对于特征选择特别有吸引力，因为非常有限的启发式可以每次都将搜索引导到最佳的最小子集。此外，情况可能是蚂蚁在整个搜索空间中发现最佳特征组合。本文研究如何应用ACO来产生最优特征子集。

元启发式优化理论的发展一直蓬勃发展。许多元启发式范例，例如遗传算法（Goldberg，1997），模拟退火（Kirkpatrick，Gel-att，＆Vecchi，1983）和禁忌搜索（Glover，1989）已经显示出它们在解决计算密集问题方面的功效。其中，Abo-Sinna（Osman，Abo-Sinna，＆Mousa，2005）提出了一种用于多目标资源分配问题的遗传算法。染色体是资源分配的工作，目标是以最小的成本最大化效率。 Lin和Gen（2007）针对同一问题提出了一种混合遗传算法。可以看出，与数学规划方法相比，仅进行了基于元启发式算法的少数多目标资源分配问题研究。

支持向量机（SVM）基于Vapnik提出的统计理论。它是一种机器学习算法，开始在九十年代中期使用。统计理论采用最小化结构风险的标准;同时，它可以降低模型的全局误差。它提高了模型的泛化能力，这在小样本学习中更为突出。

通过以上分析，提出了一种可以提高负荷预测精度的新思路。为电力负荷预测创建了一种基于蚁群优化的支持向量机的新方法。我们案例研究中的计算表明，预测准确性高于BP神经网络和单个SVM。

本文的其余部分安排如下。在第2节中，我们讨论了电力负荷数据的特征。第3节和第4节讨论了蚁群优化（ACO）的特征选择和SVM回归。通过第5节中的计算和分析证明了一个案例研究。最后，结论在第6节中给出。

特征选择：

我们将在本节讨论短负荷预测中的各种影响因素。要构建负载预测模型，选择合适的输入变量是最重要的任务之一。在现有文献中，很少有一般规则可以用来执行这项任务。根据重复试验的工程判断和经验等经验法则通常很有用。其他定量技术，如负荷曲线分析和统计分析，可以帮助选择关键变量来构建这样的负荷预测模型。众所周知，短期负荷受到诸如天气，经济发展等诸多因素的影响。为了获得有影响的因素和负载摆动之间的直观关系，我们从原始时间序列数据中描绘出的曲线模式。我们的数据分析基于中国陕西省一个非常典型的电力市场。我们注意到影响负荷预测的时间因素包括一年中的时间，一周中的某天和一天中的小时。工作日和周末之间的负载存在很大差异。天气条件也会影响负载。事实上，预测的天气参数是短期负荷预测中最重要的因素。温度是最常用的负载预测器。我们从图2中显示了中国陕西省2007年的电力负荷曲线。 1-5。

2.1。 小时预测功能

每日负荷从一小时变化到另一小时。 图1显示了2007年7月15日中国电力市场 - 陕西省电力市场典型日的负荷数据。 负载峰值每天发生两次，一次是大约上午9:00，另一次是大约下午5:30，负载大小存在显着差异。 因此，小时指标（时间从1变为24）对短期负荷预测非常有帮助。

2.2。 周预测功能

一周内负荷也会从一天变为另一天。 图2显示了陕西省2007年7月23日至2007年7月29日第29周的典型周平均日负荷。 很明显，周末的电力消耗远低于所有季节的工作日，并且还可以看出周一和周二的电力消耗远远超过其他工作日。 因此，日期指标（其中n从1变为7）可能有助于负荷预测。

2.3。天气预报功能

为了研究天气变量对省系统负荷的影响，图3-5展示了负荷预测与温度，风和阴天覆盖之间的关系。天气数量在该省的5个以上不同地点进行测量。很明显，温度是最重要的天气变量，图3证实了温度和负荷之间的这种强相关性。如图4所示，风速对负荷的影响小于温度。可以看出云盖与图5中的负载之间的关系。

图图1-5还示出了负载曲线的可变趋势。从图。如图1和图2所示，负载曲线的规律性非常明显，这些可以作为负荷预测的重要特征;图3表明负荷曲线随着一年内的温度变化明显，这个因素将是影响负荷预测的重要因素。但是，从图在图4和图5中，不能看出负荷曲线与风，云层，降雨等之间的关系。需要一些方法来选择这些特征并计算与这些影响因素的关系程度。在本文中，我们使用蚁群优化模型来选择影响因素并分析这些因素的关系。

3. ACO用于特征选择

蚁群优化算法（ACO）提供了一种替代的特征选择工具，其灵感来自蚂蚁在寻找从群体到食物的路径时的行为。真正的蚂蚁表现出强大的能力，通过在旅行时沉积信息素，找到从殖民地到食物的最短路线。 ACO模仿这种寻找食物现象的蚂蚁产生最短的路径（这意味着利益的“系统”已经融合到单一的解决方案中）。可以存在不同的同样最短的路径。只要可以定义，ACO算法通常可以应用于任何组合问题：

首先，可以在一组节点和节点之间的边缘中描述该问题以形成图。因此，可以很容易地看到问题，找到主要问题并克服它。

其次，路径的启发式可取性：它是一种合适的启发式度量，可以找到从一个节点到每个其他连接节点的更好路径，并且可以在图形细节中描述。

第三，解决方案的构建：制定的单元间布局问题的可行且完整的解决方案被认为是制造单元的排列。该解决方案的每个部分都称为状态。在最佳过程中，每个蚂蚁最初将一个小区分配给位置1，然后将另一个小区分配给位置2，依此类推，直到获得完整的解决方案。

第四，信息素更新规则：首先，它是区域信息更新规则，其效果是使边缘的合意性动态变化以便改组游览。一个蚂蚁之旅中的节点将以较低的概率选择建立其他蚂蚁之旅。因此，蚂蚁将倾向于探索尚未访问的边缘并防止会聚到共同路径。接下来，转到全局更新规则，此过程在所有蚂蚁完成其旅程后执行。因此，只允许在算法的当前迭代中找到最佳解决方案的全局最佳蚂蚁存放信息素。

第五，概率转换规则：该规则确定蚂蚁从图中的一个节点穿越到下一个节点的概率。遍历和边缘pheromone水平的启发式合意性结合起来形成所谓的概率转换规则。它可以表示在图6中。