基于离散Frechete距离和K-means

聚类分析的供热锅炉出水温度评价模型

孙志伟**1**，冯海波1，单渊博1，董亮亮1

（1. 天津科技大学计算机科学与信息工程学院，天津 300457）

**摘　要**：供热锅炉系统是一个非线性复杂系统，在实际应用中存在诸多优化问题。通过分析锅炉系统在运行过程中的数据特点，结合实时采集的气象条件，建立了一个基于Frechete距离的评价模型，并增加延时变量从而优化Frechete对于时序数据处理的不足，最后与聚类分析结合进行验证，结果表明此评价模型对于锅炉系统的输出参数（出水温度）有较好的评价结果，可以为锅炉管控的相关人员提供决策支持和分析依据。

**关键词：**锅炉控制；数据挖掘；Frechete；聚类分析

**中图分类号：              文献标志码：                   文章编号：1672-6510（0000）00-0000-00**

**Topic in English**（一般为短语形式，不用陈述句）

Sun Zhiwei1，Feng Haibo1，Shan Yuanbo1，Dong Liangliang1

(1. College of Computer Science and Information Engineering, Tianjin University of Science & Technology, Tianjin 300457, China)

**Abstract:** 1.英文摘要应是中文摘要的转译；2.常用过去时叙述作者工作，用现在时叙述结论；3.取消不必要的语句，如＂It is reported＂等，其他要求同中文摘要。

**Key words:** keyword1；keyword2；keyword3；keyword4；…

我国是一个能源消费大国，同时也在能源匮乏的国家之列，我国北方地区冬季主要以燃煤燃气供热为主，且我国供热效率偏低，单位面积供热能耗是同纬度发达国家的3~4倍，而锅炉供热的操作直接影响了能源转换的效率，所以如何优化供热锅炉系统，减少过度的能源浪费，从而达到节能减排的目的，称为现在一个亟待解决的问题。

国内外诸多专家学者已经对锅炉运行参数进行数据分析，利用数据挖掘技术展开了对锅炉燃烧优化的一系列工作。目前较为先进的优化技术都是与自动控制系统结合形成优化的闭环操作。如美国的Ultramax公司的Ultramax燃烧优化技术、Pegasus公司的Power Pefecter燃烧优化技术，国内对于锅炉系统的优化也有很多成果，文献1通过建立宝钢能源数据仓库对时序数据挖掘的离群挖掘、相似性挖掘、规则挖掘和趋势挖掘等算法进行了系统深入的研究；文献2通过对时间序列进行相空间重构，建立了基于支持向量回归的时序数据预测模型，从实现对锅炉输出参数的预测。

由于中国国情所限，供热锅炉的智能控制在我国普及程度较低，且其自动性在复杂气象条件的情况下表现较差，且绝大部分在供热锅炉方面的研究与应用利用数据挖掘技术只进行了简单数据的预测和分析，并未考虑供热锅炉系统的时间延迟和气象条件分析。综上所述，利用供热锅炉系统的历史数据并利用实时气象条件，结合延迟参数运用到提高供热锅炉系统的效率上，可以更有效的为供热锅炉自动控制化提供理论依据。

本文基于Frechete距离，增加出水温度的延迟参数，结合复杂的气象条件和时间段管控机制并利用聚类分析，建立供热锅炉输出参数的评价模型。通过仿真实验，得出了多种控制模式的差异化对比分析，为锅炉系统的相关管控人员提供了决策依据。

1. **锅炉控制系统结构**

燃气供热锅炉系统，是通过燃气锅炉加热水循环，输出热水至换热站，经过换热处理向各单位供暖。在燃气供暖锅炉控制现场，主要的输出参数有锅炉出水温度、出口总管温度、出水压力、出口总管压力、回水温度、入口总管温度、回水压力、入口总管压力等，而依据室外温度、风速、日照、活动时间等因素确定系统的出水温度，当室外气象条件变化时，依据不同偏移值，出水温度将由自动或者手工方式进行相应调控，以达到用户的供暖需求，因此评判一个供热锅炉系统的优劣的表中可以通过出水温度是否与气象复合因素的变化关系来确定。

1. 评价曲线定义

供热锅炉系统的出水温度是依据室外温度等复合因素综合考虑的，为了更好的判断出水温度的优劣，我们将定义一条评价曲线，将室外温度、风速、日照、活动时间段作为评价曲线的影响因子，定义不同的偏移量得出我们的评价函数：

Y = x + y + z;

3.曲线相似性比较

由于出水温度的合适与否是依靠室外温度等气象条件来判断的，所以我们可以描绘出两条曲线，分别是一天中的气象条件也就是我们的评价曲线，另外一条是一天的出水温度，通过判断这两条曲线的相似程度来确定供热系统的好坏。

目前 H ausdorff 距离作为距离的测度被广泛应用于判断两个点集间的相似性 ,而( 连续) Fré chet 距离[ 2] 则被用来研究两个连续曲线相似性 . 文献 [ 3 ～ 5] 都曾将Fréchet 距离应用到判别两条曲线的相似性上 ,并且文献[ 6] 还将 Fré chet 距离应用到判断蛋白质结构排列的相似性上 ,文献[ 7] 将其应用到了签名认证中 ; Eiter 和 Mannila[ 8] 在连续 Fréchet 距离的基础上提出了离散 Fré chet 距离的定义 , 而江明辉等人[ 9] 将其运用到了判断蛋白质的结构排列中去 , 收到了一定的效果 , 但是离散 Fré chet 距离还没有得到全面的应用 .

1. 仿真实验

本文利用天津XX供热公司提供的历史数据，对2014年1-3月份的数据进行了提取。由于自动采集系统的稳定性等原因会导致缺失值，错误值等离群点的出现，这里我们先进行数据预处理，通常的缺失值补充方式有，均值法，极值法，拉格朗日插值法，基于时序数据的特点，这里我们采用拉格朗日插值法来进行缺失值的补充。对二十天的出水温度及评价模型的计算，结果如下图所示：

XXX

利用聚类分析，kmeans

**技术介绍**

* 1. **曲线相似**
  2. **Frechete距离**
  3. **Hdoff距离的对比**
  4. **增加延时**

1. **评价模型**
2. **数据预处理实验**
   1. **聚类分析**
3. **结论**
4. **参考文献**