基于离散Fréchet距离的供热效果评价模型

孙志伟**1**，冯海波**1**，王福全**2**，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Fréchet距离的评价模型，并增加延时度量从而优化Fréchet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：**曲线相似度；数据挖掘；Fréchet；聚类分析

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

**Based on discrete Fréchet**

**Evaluation model of water outlet temperature of heat supply boiler**

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:** Heating boiler system is a nonlinear complex system, there are many problems in the practical application. Through the analysis of the boiler system in the running process of the characteristics of the data, combined with real-time acquisition of the meteorological conditions, established a evaluation model based on fréChet distance and increased delay metric to optimize fréChet for lack of time series data processing. Finally, and cluster analysis were combined to verify. The results show that the evaluation model for boiler system output parameters (water temperature) have better evaluation results can provide decision support and basis for the relevant personnel of the boiler control.

**Key words:** boiler control；data mining；Fréchet；cluster analysis；

我国是一个能源消费大国，同时也在能源匮乏的国家之列，我国北方地区的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，成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，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结合实时气象条件，加入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Fréchet距离，增加出水温度的延迟度量，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结果，为锅炉系统的相关管控人员提供了决策依据。

# 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。在燃气供热锅炉控制现场，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，控制不同偏移值，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标准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。

# 参考曲线定义

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，本文将定义一条参考曲线，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，并依据供热公司多年的实践经验值采用不同的偏移量得出我们的参考曲线，其表达式如下:

R(t) = aTt + bWt + cSt + dPt  （1）

其各参数解释如下：

t: time，一天当中的某个时刻

Tt: Temperature，t时刻的室外温度

Wt: Wind speed, t时刻的风速

St: Sunshine, t时刻的光照

Pt: Plan，供热计划，依据日常活动调整的供热变化时间段

a, b, c, d分别为四种条件因素的偏移值，将结合具体环境因素给出

# 评价模型的构建

## 3.1 供热系统输出参数的应变程度

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，即上文的参考曲线，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输出参数随参考曲线的应变程度，因此此问题可以转化成比较两条曲线的相似度问题。

## 3.2 曲线相似度比较

目前 Hausdorff 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，(连续) Fréchet 距离则被用来研究两个连续曲线相似性。文献都曾将Fréchet 距离应用到判别两条曲线的相似性上，并且文献还将Fréchet 距离应用到判断蛋白质结构排列的相似性上，文献将其应用到了签名认证中；Eiter 和 Mannila在连续 Fréchet 距离的基础上提出了离散 Fréchet 距离的定义 , 而江明辉等人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，但离散Frechet距离过滤了两条曲线的延时条件，所以并不能完全体现供热锅炉系统的曲线的特点，我们将在下文中进行相应的改进形成一个新的复合度量。

## 3.3 Fréchet距离及离散Frechet距离

Fréchet 距离由 M. Fréchet 于1906年提出，描述了两质点分别沿着2条给定曲线以任意速度单向运动时，二者之间的最短距离。Fréchet 距离的直观解释是：给定距离空间的2条曲线A和B，一个人牵着一条狗，分别沿着曲线A和B以任意的自由速度从起点移动到终点，但不得后退，那么 Fréchet 距离就是人与狗之间最短拴狗绳的长度，所以又称狗绳距离，其具体定义如下：







其中， 表示 Fréchet 距离， 表示 L2 范数，α和 β 表示［0,1］ ［a,b］的任意连续非递减函数。

Eiter和Mannila在连续Frechet距离的基础上提出了离散Frechet距离，本文将离散Frechet距离作为曲线中的关键特征之一来研究曲线的相似性，其定义如下：

（1） 给定1个有n个至高点的多边形链P=<P1,P2,P3,...,Pn>，1个沿着P的k步，分割P的至高点成为k个不相交的非空子集{Pi}i=1,…,k，使得pi=<Pni-1+1,…,Pni>和1=n0<n1<…nk=n

（2） 给定2个多边形链A=<a1,…,am>，B=<b1,…,bn>，1个沿着A和B的组合步是1个沿着A的k步{Ai}i=1,…,k和1个沿着B的k步{Bi}i=1,…,k组成，使得对于1<=i<=k，且Ai,Bi中有1个敲好包含1个至高点。

（3） 1个沿着链A和B的组合步W={(Ai,Bi)}的花费(cost)就是

Dfw

其中dist()为a,b间的欧氏距离，则链A和B间的离散Frechet距离就是

Df（AB）

## 3.4 带有延时度量的离散Fréchet距离

由于Fréchet距离对曲线的时序透明，而延时恰恰是供热曲线中恒量供热锅炉系统是否与气象条件反馈及时的表现，所以我们为Fréchet距离增加延时度量，构成我们评判供热锅炉系统的复合度量标准。这里采取将曲线的至高点与另一条曲线至高点的差值作为两条曲线的延时元素，然后取次高点的差值作为下一个延时元素，一次类推，得到一个时间段内的延时并取均值，以此作为我们的延时度量：



*d*i为第i个至高点

现在我们将两个度量合并为一个复合度量来作为我们出水温度曲线和参考曲线的相似度:



α: 离散Frechet距离度量的权重

β: 延时度量的权重

ε1: 离散Frechet距离的阈值

ε2: 延时度量的阈值

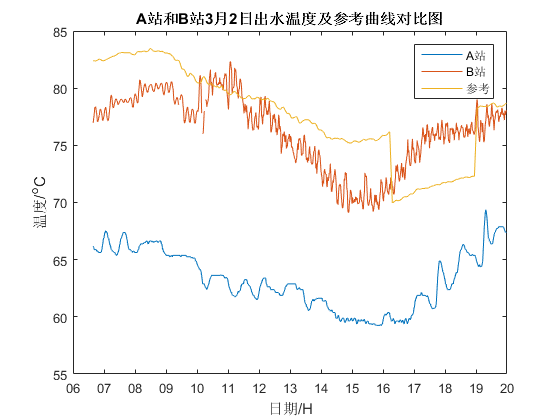
至此，本文在离散Frechet距离的基础上加入延时度量形成评价供热锅炉出水温度和参考曲线的复合度量，将作为我们评价模型的理论依据。

# 仿真实验

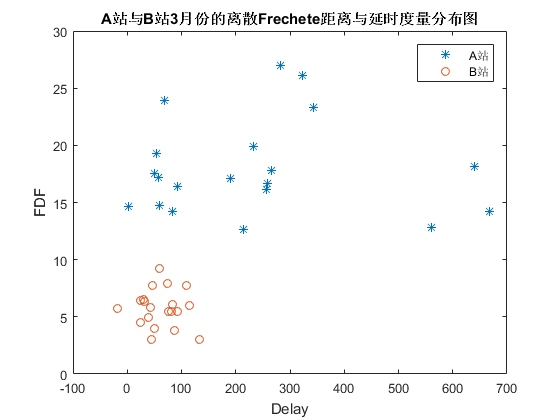
本文利用天津某供热公司提供的历史数据，对2014年3月份共20天的数据进行了提取。由于自动采集系统的稳定性等原因会导致缺失值，错误值等离群点的出现，这里我们先进行数据预处理，通常的缺失值补充方式有，均值法，极值法，拉格朗日插值法，基于时序数据的特点，这里我们采用拉格朗日插值法来进行缺失值的补充，以此作为我们实验的初始数据集，部分数据如表一所示：

|  |
| --- |
| 出水温度/°C 回水温度/°C 气温/°C 日照/Lux |
| 65.88 40.87 6.00 522 66.13 40.86 6.00 522 66.25 40.88 5.90 516 66.38 40.88 5.80 519 66.63 40.88 5.70 516 66.75 40.88 5.60 521 66.88 40.88 5.60 526 66.86 40.87 5.60 522 66.76 40.88 5.60 521 66.88 40.88 5.70 516 |

现依据（1）式将20天的气象数据计算出相应的参考曲线，3月2日的A供热站，B供热站及参考曲线的对比图如下所示：



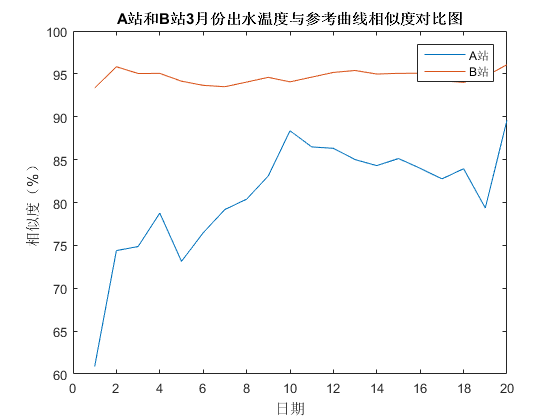
然后我们将A站的出水温度曲线与B站的出水温度曲线分别于参考曲线进行相似度对比，首先依据（2）式、（3）式计算出相应的离散Frechet距离度量与延时度量，其结果如下图所示：



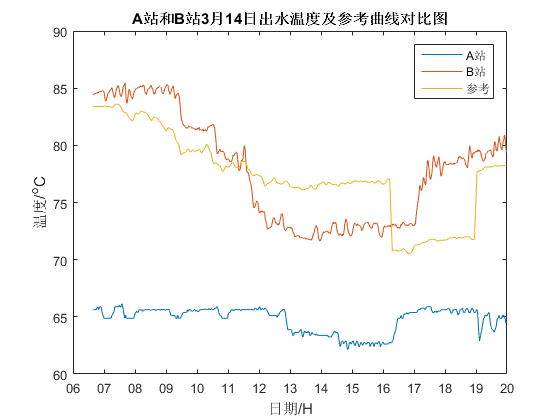
此图已大致能看出B站较A站更为相似，在离散Frechet距离和延时度量上有更好的表现，在对数据库中的所有数据进行分析，并通过大量的实验验证以及锅炉供热系统相关管控人员历史经验，确定（4）式中的复合度量参数值分别为：

α = 0.8, β = 0.2, ε1 = 200, ε2 = 400

现依据（4）式将上文所述离散Frechet距离和延时度量复合成为一个度量，如下图所示：



此图可以看出B站整体上较A站相比在我们的评价模型上表现更为相似，现取出一天的相应数据作为特征点加以验证，如下图所示



回溯数据可知，B站采取了自动的智能控制系统来调节出水温度，在考虑室外温度的同时还按照相应的人类活动时间段设置了偏移量，而A站只是按照室外温度进行了相应的人工控制，所以在与参考曲线的相似度及延时度量上更为接近，而其实际意义是B站只是根据室外温度调节出水温度，而A站可以根据室外温度、日照等复杂气象条件及时调节出水温度，表现为灵敏度高且反馈及时并能按照人类活动时间段下调出水温度，极大的减少了资源浪费，节约了能源，这将对锅炉供热系统的相关管控人员提供极大的参考经验，对节约资源浪费具有重大意义。

# 结论

本文给出了一条参考曲线，利用离散Frechet距离及延时度量作为两条曲线的复合度量，构建了一个合适的供热锅炉系统出水温度评价模型，可以很好将不同的供热锅炉系统按照度量区分出来。利用对天津某供热公司提供的历史数据，对2014年3月份的数据进行分析，通过本文提出的供热锅炉出水温度评价模型将A站和B站进行了评价，结果表明，本评价模型通过复合度量可以很好的将操作方式不同的几个站进行区分，这为供热锅炉系统的相关人员提供了很好的决策依据，对相关系统节能减排具有较大意义。由于出水温度受多个复杂的气象条件影响，本文的复合气象条件评价曲线上还有提高的空间，这将在接下来的工作中继续完善。

**参考文献**