

基于降噪自编码器降维的汽车行驶工况分析^①



顾珉, 施华君

(华东计算技术研究所, 上海 201808)

通讯作者: 顾珉, E-mail: gmforstudy@outlook.com

摘要: 近年, 随着环境治理要求的不断提升, 汽车尾气作为污染的主要来源之一. 汽车工业是我国如今重点关注的对象, 而汽车行驶工况在是汽车工业界内被看做是反映汽车性能的主要标准, 因此对于工况的研究就成了时下急需的研究项目之一. 本文根据汽车行驶工况的各项行业参数, 研究出一种通用的反映我国不同地区汽车行驶工况的方法. 同时在对复杂数据降维时, 利用了深度学习中的降噪自编码器降维方式, 取得了较优且贴合实际意义的实验效果. 本文数据来源于对上海市嘉定区汽车实验所得, 在对数据处理后, 经过本征模态分析、特征提取、动态时间规划、小波分解等不同手段和方式实现了对通用工况图的构建, 为以后我国其余城市和总体大致的工况图构建提供参考.

关键词: 汽车行驶工况; 降噪自编码器; 动态时间规整; 本征模态分析; 小波分解; 近似熵

引用格式: 顾珉, 施华君. 基于降噪自编码器降维的汽车行驶工况分析. 计算机系统应用, 2021, 30(1): 38-44. <http://www.c-s-a.org.cn/1003-3254/7745.html>

Analysis of Vehicle Driving Condition Based on De-Noise Autoencoder

GU Min, SHI Hua-Jun

(East China Institute of Computing Technology, Shanghai 201808, China)

Abstract: In recent years, with the continuous improvement of environmental governance requirements, automobile exhaust becomes one of the main sources of pollution. The automobile industry is the focus of attention in China nowadays, and the driving conditions of automobiles are considered as reflections in the automobile industry. Therefore, the study of working conditions has become one of the urgently needed research projects. Based on various industry parameters of automobile driving conditions, this study develops a general method for reflecting automobile driving conditions in different regions of China. At the same time, we use the dimensionality reduction method of the de-noise encoder used in deep learning when reducing dimensionality of complex data, and has achieved good and practical experimental results. The data in this paper is derived from the automobile experiment in Jiading District, Shanghai. After processing the data, the construction of the general working condition map has been realized through different means and methods such as EMD, feature extraction, dynamic time planning, wavelet decomposition, etc., providing a reference for the construction of the general working conditions map of the city and the overall.

Key words: vehicle driving conditions; de-noise encoder; dynamic time warping; eigenmode analysis; wavelet decomposition; approximate entropy

对汽车行驶工况的构建可以有效地反映一个城市甚至一个国家的汽车行业的发展水平, 对于汽车行业、环境保护都有着深刻的影响^[1-3]. 工况的变化对于汽车燃料消耗、污染物排放都有着重要的影响, 因此

① 收稿时间: 2020-05-11; 修改时间: 2020-06-10; 采用时间: 2020-07-03; csa 在线出版时间: 2020-12-31

本文主要针对汽车行驶工况进行科学有效的研究. 作为国内知名的汽车之城, 上海市嘉定区拥有着丰富的汽车制造资源, 我国上汽集团正是坐落于此. 由于目前该地区还没有关于汽车行驶工况的成熟的研究, 因此本文选择上海市嘉定区作为实验区域. 实验数据是通过为期2周的实验测量获得, 实验分别对3辆同款车型汽车从早中晚三个不同的时段, 通过传感器反馈收集所得, 以能充分满足实验所需的一般性.

研究工况的主要参数涉及到加速时间比例、怠速时间比例、减速时间比例、平均速度、平均行驶速度、加速段平均加速度、减速段平均减速度、最大加速度、最小减速度和平均加速度等. 因此为了衡量工况的各项性能, 本文在获取数据的行车实验时, 记录下包含且不只包含以上参数在内的多项参数, 利用这些潜在的影响因素进行后续的实验.

同时针对可能存在的怠速问题, 本文进行了一个全面的讨论. 因为汽车在堵车状态或者失去信号的状态下, 信号可能会不完整, 因此需要讨论汽车在一定范围的时间缺失值下, 是否仍视为处于同一个连续运动区间的时间片段. 这样不但满足对数据完整性的要求, 而且也符合实际生活中的意义, 处理过后对数据进行运动学片段划分. 对于不同路况行驶的测量数据, 其中上述涉及到的影响因素都通过每秒记录一次的方式保存下来. 由于数据集本身的一些特殊性, 根据实际情况分析, 根据GPS信号消失的时间长短, 对于消失时间较长的数据段进行切割和删除. 而对于较短和残缺的数据段, 主要采用了线性插值法、拉格朗日插值法、样条插值等不同的方法进行实验. 考虑到与实际情况可能有所偏差, 然后在得到的新数据集中, 进行异常值的过滤、筛选和清洗, 通过计算瞬时加速度来检测加速度异常的情况, 并用限幅算法进一步调整到合理的区间.

最终在构建工况的问题里, 本文对之前数据预处理后所得的新数据集, 进行了以下流程: 提取特征—降维—聚类—提取典型片段—合成—评价. 由于时间序列本身的前后倚赖关系和复杂性, 我们进一步对每个片段不同维度的特征进行了细粒度更高的提取, 从而得到了一个影响因素在29维的运动片段特征数据集. 维度过大是因为部分属性存在一定的冗余, 出于对计算效率以及准确性的考量, 需要对数据进行降维. 基于这种思想, 本文提出了一个改进的降维方法, 首先利用

T-SNE 算法提取主成分以获得非线性信息, 接着使用PCA主成分分析发现实验效果并不理想, 因此最终选择了深度学习中比较流行的降噪自编码器进行降维处理, 取得了不俗的实验效果, 且非常贴近实际意义. 在聚类过程中, 经过实验发现K-Means的欧氏距离在数据集上效果不是很理想, 因此本文考虑结合时间序列相关性, 将度量替换成DTW(动态时间规划)和Pearson相关性, 取得贴近现实意义的结果.

同时我们根据收集数据和研究发现, 工况与速度频率信息可能还有着密切的关系, 而传统的工况研究里并没有涉及到此方面的研究, 因此本文还决定将速度因素加入到新的考量中. 因此本文采用了基于EMD和平均近似熵的时间序列复杂度度量方法^[4]. 先对所有运动片段进行了EMD(本征模态分解), 得到了自适应个数的模态. 熵是一种能够反映序列复杂度的指标, 为了衡量这些模态的复杂度信息, 设计模态计算近似熵并求平均值作为一个运动片段的复杂度描述, 再根据复杂度对属于同一类的运动片段进行排序, 加入随机种子, 得到最终的候选片段. 图1为本实验总体所采取的流程.

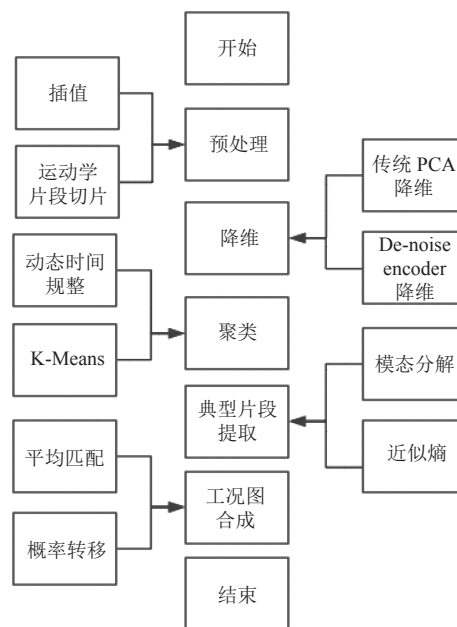


图1 实验总体流程图

1 数据预处理阶段

数据采集阶段, 对汽车上安置的传感器所收集的数据采用一秒一次的方式进行记录, 虽然传感器已经有

了相当的精度,但是可能还会出现一些不良的数据^[5,6]. 经过后期调研和数据检查发现,不良数据的原因主要分为3类,第1类是缺省值,这个通常是传感器在经过高楼或者隧道时丢失信号造成;第2类是汽车加、减速度的异常,这个通常是传感器自身出现的问题;第3类是怠速情况下的异常,其中通常认为汽车在长时间处于10 km/h的速度以下视为怠速,怠速的异常最主要原因通常是长时间停车未关闭检测器,也可能是堵车等交通路况复杂导致.

由于这些原因造成的数据不连续情况,经过仔细考虑,本文对3类异常出现的原因进行总结,针对第1类情况和第2类情况,可以首先对异常的数据进行清零,然后再使用插值的方法进行补充.而针对怠速问题,本文采用限幅算法对车速进行约束,如果汽车在180 s内一直小于10 km/h的速度视为异常状态,对这段时间就采用忽略的方法,而对180 s内能恢复速度的采用插值的方法.

本文实验对收集到的数据处理的方法如下:首先以时间序列作为划分度量,分别得到汽车行驶的时间区间,汽车从上路到熄火的完整过程应视为同一个切片,实验划分以相邻两个记录的时间点间隔为依据,如果相距5分钟以上的片段则进行划分和切片,对这些时间跨度小于5分钟但不连续的运动区间进行数据补充的处理,然后主要是进行插值补充的方法.针对插值的方法,本文对线性插值法和拉格朗日插值法两种不同的方法分别进行讨论,为了得到更优的结果需要分别进行实验.

(1) 线性插值法,即在两个片段之间按照两个区间边界的情况进行插值,呈现出一条线性递增或递减的曲线.

(2) 拉格朗日插值法,即在已知的连续片段下,按照过这 $n+1$ 个点,且次数不超过 n 的多项式 $y = P_n(X)$,要估计任何一点 ξ , $\xi \neq X_i$, $i = 0, 1, 2, \dots, n$,用该多项式计算结果进行插值.在这种情况下,插值曲线呈现出类抛物线的形式.

在这两种插值法的比较下,本文通过实验分别编程分析,最终发现线性插值法的效果更优.通过分析原因,主要是由于在信号丢失的情况下,汽车往往还能保持匀速或者其他非怠速状态的运动方式,即怠速时间比会小于正常数据的比例,因此,在时间段插值的补充结果里,我们都采用了线性插值法.

预处理实验阶段,本文利用Python和Matlab作为

主要编程工具,首先将收集得到的数据导入,检查是否存在缺省值.接着,以时间作为唯一的index进行筛选,第一步对时间中的小时进行划分,目的是将文件以时间为单位,将原始文件分为了13个行间隔为1小时的新文件,接着利用相似的方法,继续将文件细分成最大间隔为5分钟的不同数据段.在此基础上,本文对加减速异常进行筛选后,利用线性插值法,将所有划分的时间段进行补充,最终得到处理后所有新文件记录时间的行间隔均为1 s.

运动学片段是指从一个怠速阶段开始到另一个怠速阶段开启的车速区间,这是构建汽车工况最主要的依据来源,因此需要针对怠速的各项特征进行分割运动学片段.针对之前经过预处理得到的数据集,本文对5 s以内能够恢复到正常速度的值视为连续运动片段,因而进行保留,而对5 s内不能恢复到非怠速状态的值视为异常,因而进行清零.最终根据怠速的定义(即小于每小时10 km的速度)进行运动学片段的分割,将初始数据集中进一步细化成一个独立的运动学片段.

2 运动学片段的提取

根据怠速的定义,设定在真怠速状态下的速度为0,因此经过预处理后得到一组0或者10以上的数据,本文的算法是以相邻不连续的两个0作为怠速的划分.这具体是通过构建的函数`cut_slice(data,level)`来实现,其中具体参数`data`表示数据,`level`表示划分的标准,在本文中设为两段小于10.

算法流程是通过遍历在预处理后的数据集后,以时间片划分作为主要基准,对加速度、怠速等异常清理后得到,特别是针对怠速的处理,本文实验编写了`data_search(data,level,time)`函数,编程思想是将怠速的最大值设为10 km/h,如果5 s以内仍然处于怠速状态,那么我们就将两个大于10的值之间以0填充,视为真怠速,否则即为假怠速,假怠速物理意义是车速在进入小于10 km/h后能够很快恢复到10 km/h以上,这样仍应保留数据视为行驶连续过程.

运动学片段提取的实验流程是通过遍历在预处理过程下,根据时间为主要基准,加速度、怠速等异常清理后的时间切片,对每段切片再分别进行运动学片段的提取,最终汇总成实验结果.核心算法是:考虑不同的怠速定义之下对实验结果的影响,本文分别设定能够从正常速度(大于10 km/h的速度)恢复到下一个正

常速度的最小时间间隔为 3 s 和 5 s, 根据两者的实际情况及物理含义, 分别代表汽车减速→怠速→加速的最短时间, 完成运动学片段的切割. 根据实验结果, 本文最终采用了假设中的 5 s 作为最大阈值, 在此基础上, 得到了运动学片段的划分结果. 图 2 为实验下运动学片段的部分展示.

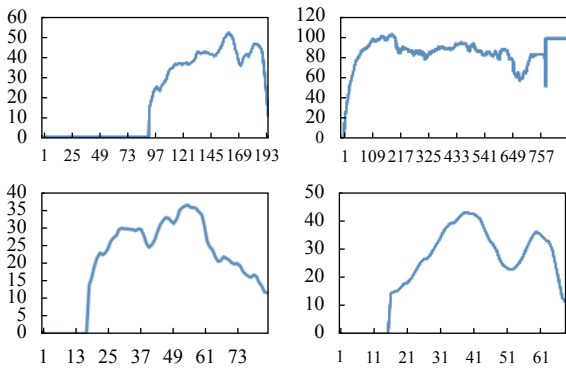


图 2 运动学片段的部分展示

考虑到汽车合理的汽车运动特征评估体系, 本文工况对于工况本身涉及的参数, 主要包括平均速度 (km/h)、平均行驶速度 (km/h)、平均加速度 (m/s²)、平均减速度 (m/s²)、怠速时间比 (%)、加速时间比 (%)、减速时间比 (%)、速度标准差 (km/h)、加速度标准差 (m/s²) 等, 最后根据实验得出的工况图以及汽车特征评估体系计算出汽车行驶工况与城市所采集数据源 (经处理后的数据) 的各指标 (运动特征) 值, 并且说明本文结果的合理性.

3 降维分析及讨论

为了更有效的描述运动学片段并对其进行分类, 通常采用特征值来表示运动学片段所具有的特点. 而特征值个数的选择是此次实验的一大难点, 如果只选用少量的特征, 如加速度、平均速度等作为特征值, 可能包含的信息并不是很全面, 这样得到的分类一般不是很好. 本实验中, 大量的特征值其实存在着巨大的冗余, 还会导致运算时间过长, 计算开销太大. 因此经过综合考虑和实验验证, 本文对收集数据的 29 个特征值结合运动学片段进行描述, 并在此基础上采用降维思想, 提升运算速度, 减少冗余. 本文实验采取了多种降维的思想, 根据实验结果的好坏选择结果.

首先利用 Pearson 相关性来辨识特征的线性相关性, 但是实验发现这样的效果并不是很好, 如图 3 所示,

因为实验结果除了对角线是自相关以外, 别的关联度不是很高, 因此决定采取第 2 种非线性降维的方法.

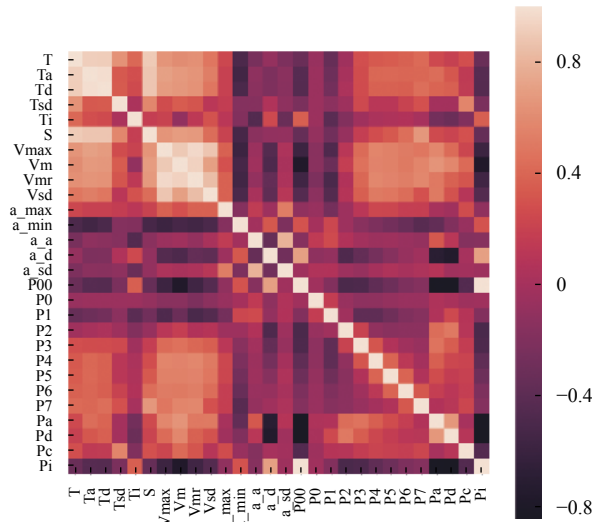


图 3 协方差矩阵

T-SNE 是一种常用的非线性降维方法, 首先对这些明显的线性相关项直接去除, 并插入这些项的 TSNE 压缩项, 主要是利用 T 分布和随机近邻嵌入的思想, 随机近邻嵌入通过将高维空间映射到低维空间转化为概率来实现, 采用传统的欧氏距离, 将距离关系通过概率的相似性来表示. 如高维空间的两个数据点 x_i 和 x_j , x_i 以条件概率 $P_{j|i}$ 选择 x_j 作为它的邻近点. 考虑到 x_i 为中心点的高斯分布, 若 x_j 越靠近 x_i , 则 $P_{j|i}$ 越大. 定义 $P_{j|i}$ 如下:

$$P_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)} \quad (1)$$

但是 T-SNE 是一种非线性降维, 这就导致其降维结果具有一定的随机性, 因此为了保证此次实验的通用性, 本文决定采用另一种降维方法——主成分分析, 这是一种常用的线性降维方法. 因为目前从综合表现来看, PCA 降维是丢失原始信息相对较少的一种线性降维方式, 而根据线性投影的方式, 将高维空间里的数据映射到低维空间中, 以投影的方差大小来衡量优劣. 根据 PCA 优化目标, 需要设定 n 维向量 W 为目标子空间的一个坐标轴方向, 以此获得最大化数据映射后的方差, 如下为 PCA 的一般通式:

$$\max_w \frac{1}{m-1} \sum_{i=1}^m (w^T (x_i - \bar{x}))^2 \quad (2)$$

其中, m 是数据实例的个数, x_i 是数据实例 i 的向量表达, \bar{x} 是所有数据实例的平均向量. 式中 W 为所有映射向量为列向量的矩阵, 经过线性代数变换, 得到如下优化目标:

$$\max_w \text{tr}(W^T A W), \text{ s.t. } W^T W = 1 \quad (3)$$

其中, tr 表示矩阵的迹, 化简过后, 得到如下数据协方差矩阵:

$$A = \frac{1}{m-1} \sum_{i=1}^m (x_i - \bar{x})^T (x_i - \bar{x}) \quad (4)$$

该式最优解是由协方差矩阵的前 k 个最大特征值, 把其所对的特征向量作为新的列向量所形成的, 正交基可以清楚地表示了数据中原有的特征向量. PCA 最终的输出是 $Y = W^T x$, 因此 X 的原始维度降低到了 K 维.

针对上述 PCA 需要处理的数据, 由于原数据涉及到多种不同形式, 因此需要采取归一化或标准化的方法, 使数据消除量纲的影响. 因此本文实验里分别采用了标准化和归一化的方法, 实验结果如图 4 所示.

```
In [6]: reduced_x1      In [4]: reduced_x1
Out[6]:                Out[4]:
array([[ -0.09415777],   array([[ -1.11264412],
      [-0.08442375],     [-0.99352444],
      [-0.09771905],     [-1.15581854],
      ...,
      [-0.03897841],     [-0.45050389],
      [ 0.71984876],     [ 8.62927243],
      [ 0.03730493]])])  [ 0.42162395]])
```

图 4 标准化归一化结果

本文在对传统 PCA 降维方法进行实验的同时, 考虑到 PCA 可能不一定适用于如此高维度的实验分析, 因此借鉴了如今深度学习中非常流行的降噪自动编码器 (de-noise Autoencoder, dA)^[7] 降维, dA 是一种类似 PCA 的无监督机器学习算法, 目的是与传统的 PCA 降维进行对比, 从而选择更优的降维方式.

而根据自编码器的定义, dA 是一种利用反向传播算法使得输出值等于输入值的神经网络, 通过将输入压缩成潜在的空间表征, 然后将这种表征重构为输出来实现. 相比于传统 PCA 算法, 自编码器在适当的维度和系数约束下可以学习到比 PCA 更有意义的映射.

同时, 降噪自编码器作为一种无监督的机器学习算法, 相较于传统 PCA 算法, 增加了非线性补充, 这就降噪自编码器比 PCA 具有更加灵活的特点, 不但能进行线性变换, 还可以进行非线性变换, 因此可以学习到比 PCA 更加有意义的数据分析. 针对数据模糊情况,

降噪自编码器还可以通过设置噪声, 恢复以前可能“损坏”的原始数据. 因此, 将通过实验来对这两种不同的降维方法进行实验验证.

4 降维实验流程

本文首先使用 SPSS 软件对数据进行主成分分析, 最终 PCA 实验验证保留了最有特征的 4 项. 具体实验流程如下, 首先对于原始文件中的 18 万条数据, 通过预处理过程, 将其分割成 1230 个运动学片段. 利用 Python 建立函数将计算得到的各运动片段的特征值作为 SPSS 的输入, 最终, 得到一个 1230×29 的特征矩阵. 接着, 对数据进行标准化处理, 用以消除量纲对特征分析的影响. 使用 SPSS 将得到的特征文件的标准化处理, 并使用因子分析, 选择主成分分析, 得到 4 个主要的特征, 主成分表用来列出所有的主成分, 且按照特征根的从大到小排列如图 5 所示, 前 4 项分别主要包括着运动片段时间、平均速度、加速时间和减速时间.

成份	初始特征值			提取平方和载入		
	合计	方差的 %	累积 %	合计	方差的 %	累积 %
1	5.007	38.516	38.516	5.007	38.516	38.516
2	2.214	17.028	55.544	2.214	17.028	55.544
3	1.185	9.115	64.659	1.185	9.115	64.659
4	1.011	7.777	72.436	1.011	7.777	72.436
5	.969	7.454	79.889			
6	.935	7.195	87.084			
7	.661	5.083	92.167			

图 5 主成分分析结果

可以从图 5 中直观地看出, 经过主成分分析提取后得到了 7 个相关成分的方差、贡献率以及累计贡献率. 然而从 PCA 分类结果来看, 传统 PCA 并没有将该 29 项特征值进行很好的降维, 结果中至少前 7 个成分包含了 29 个特征值的超过 90% 的信息, 然而主成分分析仅仅列出其中 4 项作为降维后的主成分, 且实验数据上前 4 项的累积占比也仅占 72.436%, 这就与主成分分析所达到的效果不相吻合, 分析原因可能是因为 PCA 降维是线性降维, 本实验可能在初始数据存在一定的偏差, 同时在运动学片段的分割上也不是均匀且连续, 因此 PCA 对于该实验的降维效果不是很理想.

因此, 结合之前的分析, 本文采取深度学习中的 dA 降维方法重新对该实验进行降维, 具体实验流程如下, 首先设置初始网络的节点数为数据维度 29, 由于维度不是很高, 因此设置网络深度为 4, 同时对中间隐层的节点数逐渐收敛到 15 和 10, 激活函数选择了 tanh.

针对可能存在一定的原始数据损害情况,因此将噪声系数设置为0.2。

实验发现,在设置噪声系数后,对初始数据的不连续性以及可能存在的损坏进行了很好的修复。同时在经过反复地实验后,在得到了当输出节点数为7作为最终的降维结果时,如图6所示,实验效果是最优的,这也与之前传统PCA分类所体现出的实验效果高度吻合,而且这个降维后的主成分也十分贴近实际意义。

FACTOR_1	FACTOR_2	FACTOR_3	FACTOR_4	FACTOR_5	FACTOR_6	FACTOR_7
-50.160	-4.1621	-0.959	1.1757	-1.1323	1.1727	2.4433
+0.8439	+0.1927	0.0013	1.1029	-0.0011	-0.0011	7.0116
1.04920	4.1611	1.7190	1.5556	3.0712	4.6504	4.1935
0.0813	1.0270	0.0013	-1.1887	0.0017	0.0017	2.1113
0.0000	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
-1.01796	1.18155	-0.1979	-0.1979	0.2053	0.2053	-2.0215
-0.0000	-0.0011	-0.0007	-0.0000	0.0011	0.0007	-1.15173
-0.0467	-0.1451	-0.2312	-0.2785	-0.3933	-0.4683	-1.15173
-0.0000	-1.15173	0.0000	0.0000	-0.0000	0.0000	-0.0000
1.15173	0.0000	1.15173	0.0000	0.0000	0.0000	0.0000
2.1113	0.0000	-0.1979	0.2053	-0.3933	0.2053	0.0000
0.0011	1.15173	0.0000	0.0000	0.0000	0.0000	0.0000
-0.0000	-0.1451	0.0000	0.0000	-0.0000	0.0000	0.0000

图6 主成分分析结果

经过实验证明,本文选择前7个作为主成分进行分析,大大减少了冗余情况。分析结果发现,第1主成分主要反映运动片段时间、加速段时间、减速段时间、匀速段时间、平均速度、行驶时平均速度;第2主成分主要反映速度标准差、总速时间和行驶距离;第3主成分主要反映加速段平均加速度、加速度标准差、加速度最大值,即可以理解主要反映行驶过程中加速度相关信息;第4主成分主要反映行驶时间标准差、速度为10到20 m/s的信息;第5主成分主要反映加速段的最大加速度、减速段的最大加速度以及匀速段的占比;第6主成分主要反映运行片段时间标准偏差、中等行驶速度(速度为30到50 m/s);第7主成分主要反映速度为零的时间、加速段平均加速度和减速段平均加速度。

5 聚类实验流程

在取得了影响因子的最主要成分后,本文对之前数据集中分离出的1230个样本进行聚类^[8,9],主要采用了动态聚类的思想方法,具体实验步骤如下。

首先需要定义以变量或指标的个数为维度的空间里的一种距离,根据K-Means算法,需要计算出1230个样本两两之间的距离,本文在计算传统层次聚类方式时,考虑到传统欧式距离是一般层次聚类所选的主要度量方式。但由于本文实验数据是以时间序列为基准的,拥有很强的间断性和不等长的特点,且时间序列还存在着伸缩或者平移的可能性,因此选用传统的欧氏距离不能反映两段序列之间的相似性。即使序列有等

长的可能,欧氏距离在对时间序列进行距离的度量时,由于仅仅以时间轴对齐方式来计算各个序列值的距离,而没有考虑序列的平移,因此极易导致距离计算结果的不准确,最终算法准确性也会下降。

基于上述所说的欧氏距离在进行时间序列上度量计算的诸多不足,本文采用DTW算法作为改进,采用DTW作为距离计算的度量。具体流程如下:

(1) 初始阶段,将每个分段序列单独视为一个簇,同时计算每个分段序列之间的DTW距离,将得到一个初始距离矩阵;

(2) 对距离按升序方式进行排序,将升序后数组中按索引存入二维数组中;

(3) 将数组中第1个值(最小值)对应的子序列合并到一个新簇中;

(4) 从数组第2个值开始,先判断其对应的两个子序列有无合并。若未合并,则将两个序列合并为一簇;若其中一个已合并,则将另一个子序列合并到该簇;若两个子序列分别合并到不同簇中,则将两个粗合并为一个簇。每次合并时簇的个数减1;

(5) 取数组下一个值,并将簇的个数减1;

(6) 重复步骤(4)和步骤(5),直至数组中所有元素均处理完。

最终本文主要通过组间连接的聚类方法,使用DTW作为距离度量得到图7示意的聚类结果。

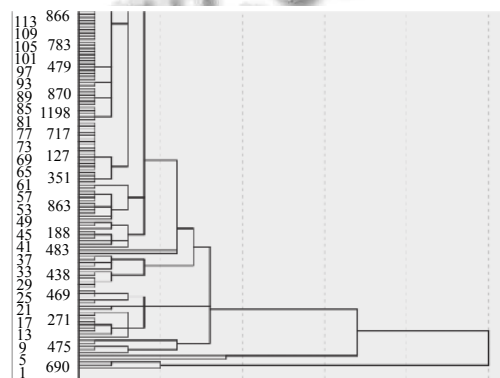


图7 聚类过程

6 EMD本征模态分解

以往的研究对从各个子工况中提取典型运动片段,基本上采用的平均速度匹配,近似一种随机挑选的方法。实际上,工况与汽车的速度波动有着密切的联系,运动片段的频域信号往往隐藏着速度变化的信息。

EMD 本征模态分解可以理解成一种自适应的小波分解, 它的优点在于并不需要预先设定分解的频段, 所以可以从运动片段中自然分解出很多潜在的行为。

如图 8, 一个运动片段中分解出了 5 个模态。第 3 个模态与减速有关, 第 1 个第 2 个模态与趋势有关, 第 4 个模态与速度突变有关。当然并不是所有的运动片段都有相同的对应, 所以我们需要一种方法去评价这个运动片段所包含的信息。相对地认为, 信息越多, 越能代表这一类的信息, 而熵经常被用于衡量一个序列的信息复杂度。我们采用运动片段所有模态的平均近似熵来描述这个运动片段的特征。

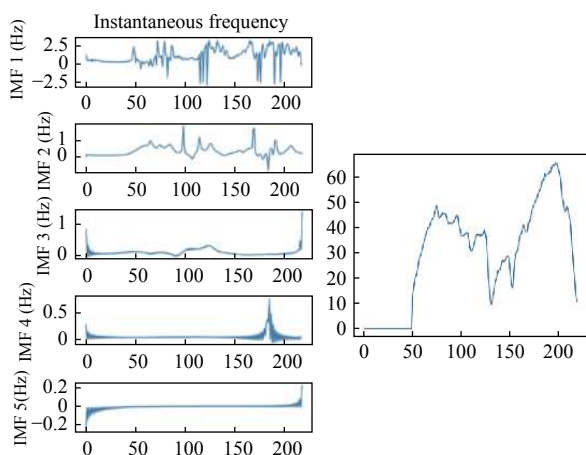


图 8 本征模态分解

按比例加入随机种子, 得到最终的候选片段, 从各分类中选择相应数量的运动片段组成 1200 到 1300 s 的运动工况, 校验标准参照与总体平均值的偏差不超过 5%。最终构建出的嘉定区汽车行驶工况如图 9 所示, 从整体来看, 可以看出嘉定区汽车行驶的工况在连续行驶的大部分时间还是处于一个相对比较连续的过程, 证明堵车情况较好, 总体行驶工况结论为: 嘉定区汽车行驶速度中等, 加减速度较为频繁, 怠速时间比例总体偏低。

7 结论与展望

本文运用机器学习和人工智能的方法, 结合实际应用需求, 创新性地提供了一种利用降噪自编码器降维方式对汽车工况实验数据进行降维的方法, 为创建

和构造汽车行驶工况的研究提供了一些可行性的方法, 也希望为我国未来其余城市的汽车行驶工况研究提供参考。同时, 提出了对于汽车行驶工况的主要影响因素, 当然不同城市的工况可能会存在着差异, 但是本文将一种通用的研究方法提出以供参考, 希望为我国环境保护和汽车工业的发展提供一定的帮助。

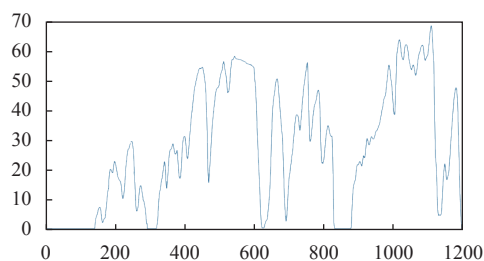


图 9 工况构建图

参考文献

- 姜平, 石琴, 陈无畏, 等. 基于小波分析的城市道路行驶工况构建的研究. 汽车工程, 2011, 33(1): 70-73, 51.
- 王鑫, 韩秀坤, 葛蕴珊, 等. 北京市公交车典型行驶工况的构建. 汽车工程, 2010, 32(8): 703-706.
- 杜爱民, 步曦. 上海市公交车行驶工况的调查和研究. 同济大学学报(自然科学版), 2006, 34(7): 943-946, 959. [doi: 10.3321/j.issn:0253-374X.2006.07.019]
- 丁一峰, 李君, 盖洪超, 等. 小波变换在行驶工况构建用车速数据处理中的应用. 科学技术与工程, 2017, 17(28): 274-279. [doi: 10.3969/j.issn.1671-1815.2017.28.048]
- 吴云波, 王惠中, 刘军, 等. 南京市轻型车行驶工况的研究. 汽车实用技术, 2011, 5: 61-64.
- 王海鲲, 陈长虹, 黄成, 等. 上海市城区典型道路行驶特征研究. 交通环保, 2005, 26(3): 35-39.
- 曾安, 张艺楠, 潘丹, 等. 基于稀疏降噪自编码器的深度置信网络. 计算机应用, 2017, 37(9): 2585-2589. [doi: 10.11772/j.issn.1001-9081.2017.09.2585]
- 李友文, 石琴, 姜平. 基于马尔科夫过程的行驶工况构建中数据处理与分析. 合肥工业大学学报(自然科学版), 2010, 33(4): 491-494.
- 马志雄, 朱西产, 李孟良, 等. 动态聚类法在车辆实际行驶工况开发中的应用. 武汉理工大学学报, 2005, 27(11): 69-71, 81. [doi: 10.3321/j.issn:1671-4431.2005.11.021]