

基于 SOM 聚类的典型车况和动力 匹配分析与决策

高涛^{1,3} 丁伟东² 张敏² 谭魏盟³ 赵祥模³

(陕西汽车集团有限责任公司博士后科研工作站¹,西安 710200;

陕西重型汽车有限公司汽车工程研究院²,西安 710200; 长安大学信息工程学院³,西安 710072)

摘要 在残酷的重型卡车市场竞争环境下,如何客观分析用户的动力配置需求,并对其进行最优配置是车企始终追求的目标。基于车速与车辆工况最强相关的观点,首先对车载数据终端所采集的样本数据进行预处理,采用自组织特征映射网络(self-organizing feature map, SOM)聚类分析方法进行速度分布聚类。然后针对不同类型车辆真实数据的聚类结果,总结出各类车型的典型工况。最后结合典型工况与车辆具体配置进行动力匹配分析,从而挖掘出各类车辆的工况需求和动力配置要求。数据化的需求分析为车企在研发、生产、销售等环节消除了歧义,降低了质量投诉以及售后支出,这表明该方案在实践中有着重要的实用价值。

关键词 典型车况 动力匹配分析 自组织特征映射网络
中图分类号 TP391.7; **文献标志码** A

伴随着上半年国内重卡市场持续下跌的窘境,重型卡车的市场竞争不断加剧。为了走出市场的“寒冬”,各个重卡企业都在苦练内功,优化车辆配置。车辆配置的核心是动力匹配^[1],它的匹配与否直接决定了整车动力性和经济性的优劣。传统的动力匹配分析多以微分方程和传递函数为基础,在多个软件中进行联合仿真,这种方法调试周期长、可移植性差,不能反应系统整体的动力匹配^[2]。东风汽车有限公司提出了利用 CRUSE 软件建立整车仿真模型的方法,此方法建立在模型的基础之上,难以真实客观地反应实际状况。

近年来,随着计算机技术的发展,基于大量运算的人工神经网络模型作为一种新的信息处理模式被运用于数据挖掘中。杨黎刚^[3]提出将 SOM 聚类应用在电信客户细分中并取得了良好的效果;刘艳杰等人^[4]利用 SOM 神经网络实现了城市土地覆盖遥

感分类。基于以上分析,提出了一种新的动力匹配分析方法,利用汽车主厂车载终端系统采集的真实重卡运行数据,采用 SOM 聚类方法,总结出各类车型的典型工况,最后结合典型工况与车辆具体配置进行动力匹配分析,从而挖掘出各类车辆的工况需求和动力配置要求。

1 自组织神经网络(SOM)聚类原理

聚类就是按照某个特定标准(如距离准则)把数据集分割成不同的类或簇,使得同一个簇内的数据对象的相似性尽可能大,不在同一个簇中的数据对象的差异性也尽可能地大。即聚类后同一类的数据尽可能聚集到一起,不同数据尽量分离。

SOM 是一种人工神经网络聚类方法^[5],利用其自组织映射功能,结合人眼对低维模式的识别能力,能够在短时间内挖掘出高维数据中所包含的聚簇分布特征,从而提高了整个聚类过程的效率(图1)。较之传统的聚类过程,SOM 方法更直观、快捷,而且一定程度上避免了传统类别选择过程中的盲目性。SOM 网络的学习过程是以无监督的训练方式进行的,在学习过程中只需向网络提供一些学习样本,而无需提供理想的目标输出。网络根据输入样本的特征进行自组织映射,从而对样本进行自动排序和分类。

SOM 网络学习算法如下^[6]。

(1) 初始化。对输出层各权值赋小随机数并进行归一化处理,得到 $W_j, j = 1, 2, \dots, m$; 建立初始优胜邻域 $N_j * (0)$; 学习率 η 赋初始值。

2016年7月16日收到 国家自然科学基金(61302150)、
中国博士后科学基金(2014M562356)和陕西省协同
创新计划(2014XT-03)资助

第一作者简介:高涛(1980—),男,博士,副教授。研究方向:智能
信号处理、模式识别。

引用格式:高涛,丁伟东,张敏,等.基于 SOM 聚类的典型车况和
动力匹配分析与决策[J].科学技术与工程,2017,17(3):
109—113

Gao Tao, Ding Weidong, Zhang Min, et al. Analysis and deci-
sion for rypical automobile condition and power match base on
SOM[J]. Science Technology and Engineering, 2017, 17
(3): 109—113

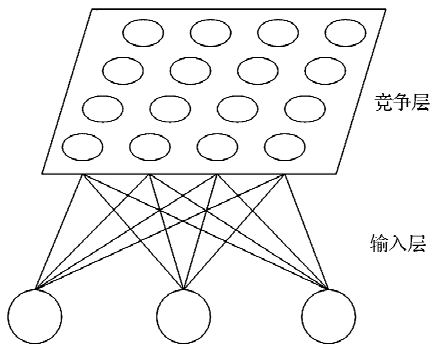


图1 SOM网络结构模型

Fig. 1 Structure model of SOM network

(2) 接受输入。从训练集中随机选取一个输入模式并进行归一化处理, 得到 $X^p, p \in \{1, 2, \dots, P\}$ 。

(3) 寻找获胜节点。计算 X^p 与 W_j 的点积, $j = 1, 2, \dots, m$, 从中选出点积最大的获胜节点 j^* ; 如果输入模式未经归一化, 应按式(1)计算欧式距离, 从中找出距离最小的获胜节点。

$$\|X - W_j\| = \min_{j \in \{1, 2, \dots, m\}} \{\|X - W_j\|\} \quad (1)$$

(4) 定义优势邻域 $N_{j^*}(t)$ 。以 j^* 为中心确定 t 时刻的权值调整域, 一般初始邻域 $N_{j^*}(0)$ 较大, 训练过程中 $N_{j^*}(t)$ 随训练时间逐渐收缩。

(5) 调整权值。对优势邻域 $N_{j^*}(t)$ 内的所有节点调整权值

$$W_{ij}(t+1) = W_{ij}(t) + \eta(t, N) [\chi_i^p - W_{ij}(t)];$$

$$i = 1, 2, \dots, n; j \in N_{j^*}(t) \quad (2)$$

式(2)中, $\eta(t, N)$ 是训练时间 t 和邻域内第 j 个神经元 j^* 之间的拓扑距离 N 的函数, 该函数一般有以下规律

$$t \uparrow \rightarrow \eta \downarrow, N \uparrow \rightarrow \eta \downarrow \quad (3)$$

很多函数都能满足以上规律, 例如可构造如下函数: $\eta(t, N) = \eta(t) e^{-N}$ 。式中, $\eta(t)$ 可采用 t 的单调下降函数。这种随时间单调下降的函数也称为退火函数。SOM 网的训练不存在类似 BP 网中的输出误差概念, 训练何时结束是以学习率 $\eta(t)$ 是否衰减到零或某个预定的正小数为条件, 不满足结束条件则回到步骤(2)。

2 基于 SOM 神经网络的车况聚类过程

车况聚类过程如图2所示, 首先对原始车速数据进行清洗, 然后统计其分布, 之后进行聚类分析与动力匹配分析。

2.1 数据来源

车载终端系统是汽车主厂为用户全新打造的一

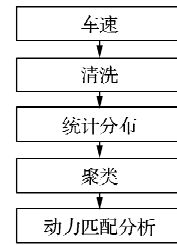


图2 动力匹配分析总体框图

Fig. 2 Framework of power match analysis

款智能服务系统。其主要分为车载智能终端和管理网站两大部分。车载智能终端能够实时采集车辆的运行数据, 管理网站则是向客户、监管单位等呈现全面的车辆运行数据。

使用数据集是从车载终端系统 2015 年 7 月份采集的部分数据中抽取而成。原始数据包含车速、转速、里程、油耗、GPS 等, 这些数据和用户的工况及驾驶行为密切相关。基于车速与车辆工况最强相关的观点, 从车速数据中挖掘出车辆工况。

2.2 数据清洗

在实际应用中, 由于存在诸如设备老化、恶劣环境影响、维护检修不当等原因, 车载终端系统采集的部分车速数据会产生质量问题。因此在应用这些数据前, 首先应进行数据清洗工作^[7]。

本研究将清洗目标分为 3 类。

(1) 错误数据。数据记录存在不在规定范围内的现象。这是出现频率最高的一种异常数据。

(2) 冗余数据。冗余数据是指在检测时间段内存在相似或相同数据。

(3) 时间点飘移^[8]。由于车载智能终端或网络不稳定造成数据记录中的时间点间隔存在误差, 导致实际时间体系与标准时间体系无法匹配。

依据以上原则, 对原始数据进行清洗, 共得到符合条件的 9 146 辆车的车速数据, 这些数据构成了本文的数据集。

2.3 SOM 网络模型

设置的 SOM 网络模型^[9], 输入层为 1 个节点, 代表不同的车速分布。

$$P_k = (p_1^k) \quad (4)$$

式(4)中, $k = 1, 2, \dots, q (q = 9\ 164)$, 构成样本数为 9 146, 指标数为 1 的输入矩阵。选取数据集中自卸车的指标数据作为 SOM 神经网络的网络输入 p 。竞争层组织结构根据分类方法将 SOM 神经网络的竞争层次选为 $[15, 8]$ 类型, 即将 9 164 个样本最终划分为 120 类进行学习, 使用 newsom 函数创建一个 SOM 网络

$$\text{net} = \text{newsom}(\min \max(p), [15\ 8]) \quad (5)$$

式(5)中, $\min \max(p)$ 为网络输入 p 的最大值和最小值; [15 8] 为竞争层的网络结构。SOM 网络的参数设置: 拓扑函数默认为 'hextop', 距离函数为 'linkdist', 排列阶段邻域半径为两个神经元的最大可能距离, 排列阶段学习速率设定为 1, 排列阶段学习次数为 1 000, 调整阶段邻域半径为 1, 调整阶段学习速率为 0.02。

对建立好的 SOM 网络训练 1 000 次, 并用训练好的 SOM 神经网络对输入数据进行自组织竞争选择, 自动将样本分为 120 级。流程图见图 3。

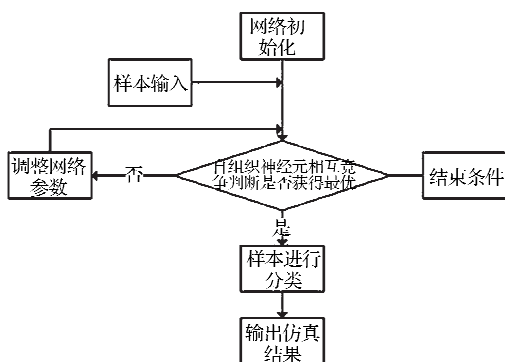


图 3 SOM 神经网络的训练流程图

Fig. 3 Training flowchart of SOM neural network

3 实验结果与分析

3.1 SOM 聚类结果分析

3.1.1 全体车聚类结果分析

图 4 是速度分布统计图, 从图中可以看出, 左上角车辆车速在 50 ~ 70 km/h 占比很高; 左下角车辆车速在 60 ~ 90 km/h 占比高; 右上角车辆怠速很高; 右下角车辆车速在 0 ~ 40 km/h 占比很高。可以初

步断定, 左侧车辆多跑高速, 右侧车辆多跑城市或重载短途运输。

图 5 为全体车车速 SOM 聚类图。图 5 说明了使用 SOM 聚类的方法, 来寻找典型工况是可行的。对数据集中分布的区域选取其核心节点并用圆圈标记, 这些核心节点就代表了一种典型工况。

在此基础之上, 对原始数据集根据车型的不同进行简单分类, 以下给出自卸车的聚类分析结果。

3.1.2 自卸车聚类结果分析

其中, 自卸车被分为了 6 个典型工况, 这 6 个典型工况对应的速度分布如图 6(b) 所示。

3.2 典型工况与动力匹配分析

陕汽 F3000 自卸车的部分配置如表 1。

表 1 陕汽 F3000 自卸车配置表

Table 1 Dump truck (F3000) configuration table

项目	配置	项目	配置
型号	F3000	发动机型号	WP10.336E40
驱动形式	6 × 4	发动机功率	336
版本	矿用加强级	喷油系统	共轨
设计库型号	SX3256DR354	车桥速比	MAN9.5T
变速器	12JSD160T	车架	850 × 300

依据图 6(a) 的典型工况, 和表 1 车辆配置, 分析出其核心需求如下。

(1) 速度主要集中于 20 km/h 内, 最高车速为 60 km/h, 怠速时长占比超过 50%。

(2) 运输过程中多在非铺装道路。

(3) 货物种类: 矿物运输、基础建设、园区内运输。

牵引车和搅拌车的分析过程类似。

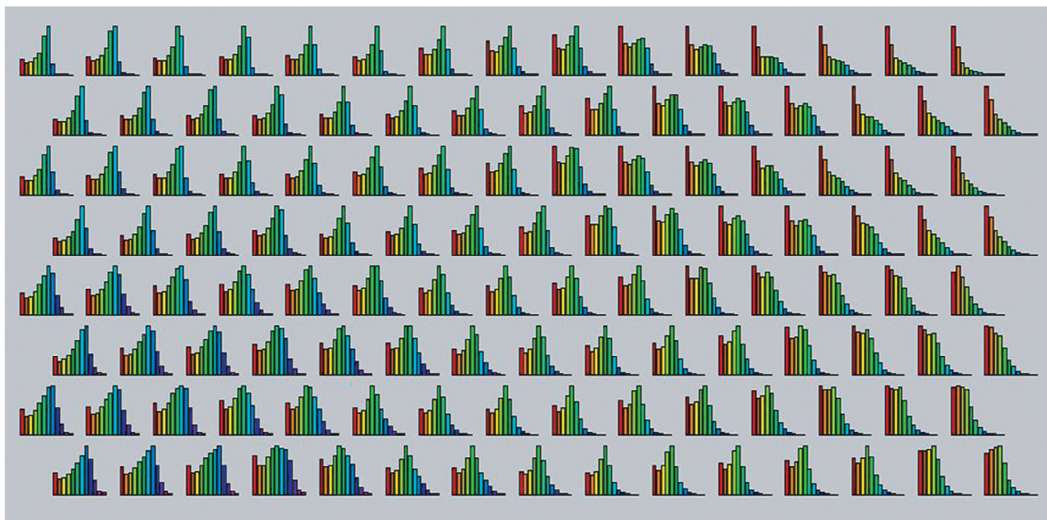


图 4 速度分布统计图

Fig. 4 Statistical chart of speed distribution

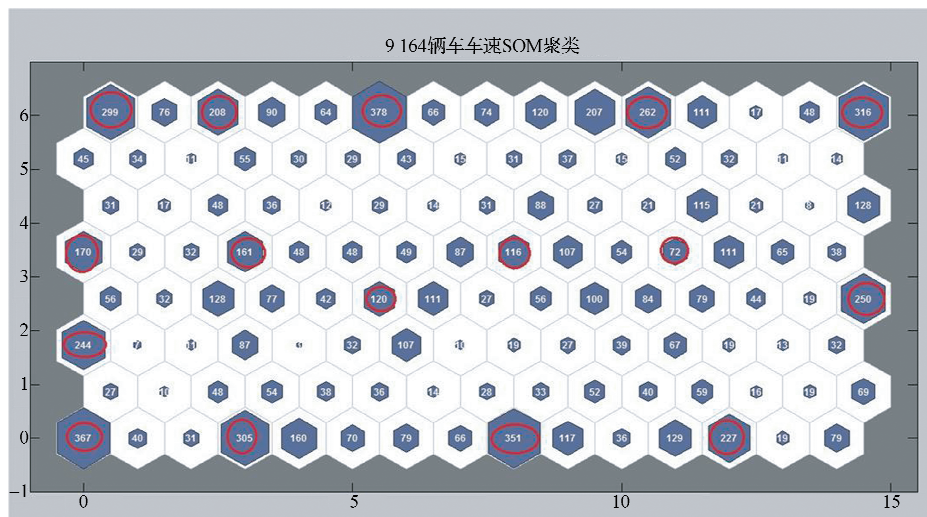


图5 全体车车速 SOM 聚类图

Fig. 5 SOM clustering graph of automobile speed

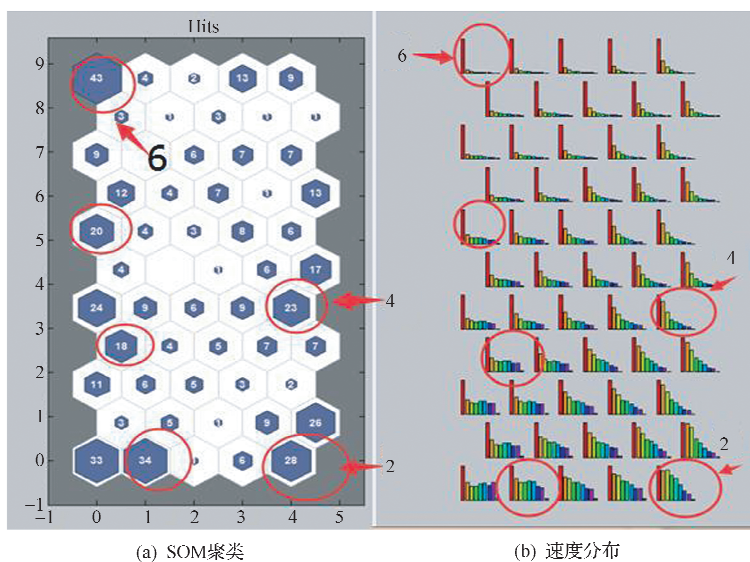


图6 自卸车的车速 SOM 聚类结果

Fig. 6 SOM clustering of dump truck speed

4 总结

针对目前缺乏解决现实环境下整车动力匹配分析方法的现状,提出了基于 SOM 聚类的典型车况和动力匹配分析方法并且进行了相关试验,试验总结如下。

(1)通过分析汽车主厂车辆运行数据建立了工况与速度分布评价体系,使用 Matlab 工具箱的 SOM 神经网络对重卡工况进行分类。结果发现,使用神经网络的分类方法是可行的,并且具有计算速度快,结果客观准确的优点。

(2)根据动力匹配分析的结果,在对牵引车进

行配置时,应考虑多跑高速的特点;对自卸车进行配置时,应偏向其重载短途运输的特点;对搅拌车进行配置时,应针对其多跑城市道路的工况进行设计。

结果表明该方法在整车配置中的动力匹配分析与决策中取得了良好的结果。

参考文献

- 1 陈建华,穆希辉,杜峰坡,等. 基于经济线的柴油机节能控制方法研究田. 机电工程,2014;31(1):129—132
Chen J H, Mu X H, Du F P, et al. Power saving control theory of diesel engine based on economical curve. Journal of Mechanical & Electrical Engineering, 2014;31(1):129—132
- 2 吴文海,黄亮,柯坚,等. 基于 AMESim 的混合动力挖掘机动

- 力匹配研究. 机械科学与技术, 2013; 32(2): 263—268
- Wu W H, Huang L, Ke J, *et al.* Hybrid excavator power matching based on AMESim. Mechanical Science & Technology for Aerospace Engineering, 2013; 32(2): 263—268
- 3 杨黎刚. 基于 SOM 聚类的数据挖掘方法及其应用研究. 杭州: 浙江大学, 2006
 - Yang L G. A method of data mining based on SOM clustering and its application. Hangzhou: Zhejiang University, 2006
 - 4 刘艳杰, 曾永年. 基于 SOM 神经网络的城市土地覆盖遥感分类研究. 测绘与空间地理信息, 2012; (6): 42—45
 - Liu Y J, Zeng Y N. Research on remote sensing classification of urban land coverage based on SOM neural network. Geomatics & Spatial Information Technology, 2012; (6): 42—45
 - 5 杨黎刚, 苏宏业, 张 英, 等. 基于 SOM 聚类的数据挖掘方法及其应用研究. 计算机工程与科学, 2007; 29(1): 133—136
 - Yang L G, Su H Y, Zhang Y, *et al.* A method of data mining based on SOM clustering and its application. Computer Engineering & Science, 2007; 29(1): 133—136
 - 6 刘 博, 肖长来, 梁秀娟. SOM-RBF 神经网络模型在地下水位预测中的应用. 吉林大学学报(地球科学版), 2015; 45(1): 225—231
 - Liu B, Xiao C L, Liang X J. Application of combining SOM and RBF neural network model for groundwater levels prediction. Journal of Jilin University (Earth Science Edition), 2015; 45(1): 225—231
 - 7 王晓原, 张敬磊, 吴 芳. 交通流数据清洗规则. 计算机工程, 2011; 37(20): 191—193
 - Wang X Y, Zhang J L, Wu F. Research on traffic flow data cleaning rules. Computer Engineering, 2011; 37(20): 191—193
 - 8 宋金玉, 陈 爽, 郭大鹏, 等. 数据质量及数据清洗方法. 指挥信息系统与技术, 2013; 10(4): 63—64
 - Song J Y, Chen S, Guo D P, *et al.* Data quality and data cleaning methods. Command Information System and Technology, 2013; 10(4): 63—64
 - 9 Yen C L, Lu M C, Chen J L. Applying the self-organization feature map (SOM) algorithm to AE-based tool wear monitoring in micro-cutting. Mechanical Systems and Signal Processing, 2013; 341—342.

Analysis and Decision for Typical Automobile Condition and Power Match based on SOM

GAO Tao^{1,3}, DING Wei-dong², ZHANG Min², TAN Wei-meng³, ZHAO Xiang-mo³

(Postdoctoral Programme¹, Shaanxi Automobile Group CO., LTD, Xi'an 710200, P. R. China;

Automobile Engineering Research Institute², Shaanxi Heavy Duty Automobile CO., LTD, Xi'an 710200, P. R. China;

School of Information Engineering³, Chang'an University, Xi'an 710072, P. R. China)

[**Abstract**] In the cruel competitive environment of heavy duty automobile market, how to objectively analyze the requirements of power configuration and the optimal configuration is a final goal for car companies. Based on the view of strong related with speed and the working condition of vehicle, a novel analysis and decision method was proposed. First, the sample form the vehicle terminal system were pre-processed and the artificial neural network SOM was utilized to cluster velocity distribution. Next, the typical automobile condition is summarized from typical working conditions of all kinds of models. Finally, the demands of conditions and power configuration for each type automobile were explored based on typical automobile condition and automobile specific configuration. A numerical demand analysis reduces the quality complaints and eliminates ambiguity for car companies in the research, development, production and sales, which indicates that the scheme has important practical value in practice.

[**Key words**] typical automobile condition power match analysis SOM