

· 争鸣与探索 ·

汽油车 B A S M 排放检测方法关联规则研究

彭美春¹, 王文涛², 林怡青¹, 赵铨泽³, 许志刚³, 彭汉锐¹

(1. 广东工业大学机电工程学院, 广东 广州 510090 2 广东工贸职业技术学院机电工程系, 广东 广州 510510 3. 深圳市汇银实业开发有限公司, 广东 深圳 518019)

摘要: 应用数据挖掘法对 3 720 台次汽油车的 B A S M 排放检测数据进行关联性分析, 对汽车 B A S M 排放检测方法中检测工况及排放污染物检测项目 CO、HC、NO_x 间关联性进行研究, 得出在 B A S M 5024 工况下没有通过检测的车辆在 B A S M 2540 工况检测的通过率只有 5 7%, CO 检测合格对 HC 的检测合格有 76 5% 的支持度, HC 检测合格对 NO_x 检测合格的支持度为 58 9%, 说明 B A S M 5024 与 B A S M 2540 检测工况具有较高的关联性, HC 与 CO 存在较高的关联度, 而 HC 与 NO_x 的关联度不高。

关键词: 汽车排气; 污染物; 加速模拟工况法; 关联性; 数据挖掘

中图分类号: X 830. 2 **文献标识码:** B **文章编号:** 1006-2009(2008)01-0042-03

Study of Association Rule of B A S M Test Method for Gasoline Vehicle Exhaust Emission

PENG M ei-chun¹, WANG W en-tao², LIN Y i-Q ing¹, ZHAO X in-ze³, XU Zh i-gang³, PENG H an-ru¹

(1. *Electromechanical Engineering Faculty, Guangdong University of Technology, Guangzhou, Guangdong 510090, China;* 2. *The Department of Mechanical Engineering, Guangdong Vocational College of Industry & Commerce, Guangzhou, Guangdong 510510, China;* 3. *Shenzhen Hu yin Industrial Economic Co., Ltd., Shenzhen, Guangdong 518019, China*)

Abstract The association rule of B A S M exhaust emission inspection method was studied. It was analyzed 3 720 groups of data about gasoline vehicle exhaust emission under B A S M test by the way of data mining. The results indicated that there were only about 5 7% vehicles which could passed B A S M 2540 test if the vehicles did not passed B A S M 5024 test. The support was 76 5% between the confirmation of CO and HC, and 58 9% between HC and NO_x. It can be concluded that there was high association between B A S M 5024 test and B A S M 2540 test. CO and HC, but HC and NO_x not.

Key words Automobile engineering; Pollutants; A S M; Association rule; Data mining

随着汽车保有量急剧增加, 在用车排放的尾气已成为中国中心城市大气污染的主要来源。研究 B A S M 排放检测法中两个检测工况关联性, 以及排放污染物检测项 CO、HC、NO_x 三者间的关联性, 应用数据挖掘法对所测取的 3 720 台次在用汽油车 B A S M 尾气排放检测数据进行关联分析, 得出检测工况 B A S M 5024 和 B A S M 2540 的关联规则及排气污染物 CO、HC、NO_x 检测的关联规则, 为在用车尾气排放检测方法的选择提供参考。

1 汽车尾气测试方法

为了控制汽车尾气污染, 我国的尾气排放标准逐年严格, 表现在排放限值越来越低, 尾气检测方法越来越接近车辆实际使用工况^[1-4]。北京市较早采用简易工况法测量在用车尾气, 采用《汽油车

收稿日期: 2007-05-29 修订日期: 2007-10-17

基金项目: 国家“十五”科技攻关基金资助项目 (2003BA408B29)

作者简介: 彭美春 (1963-), 女, 湖南衡阳人, 教授, 博士, 从事汽车排放研究。

稳态加载污染物排放标准》(DB 11/122-2003), 用加速模拟工况法 (Acceleration Simulation Mode 简称 BASM) 测量在用车尾气排放。BASM 有两种汽车尾气测试工况, 即 BASM 5024 与 BASM 2540 检测车辆先进行 BASM 5024 工况的检测, 测试车速为 24 km/h, 测得的 CO、HC、NO_x 污染物均低于限值, 则合格; 否则车辆继续进行 BASM 2540 工况的检测, 测试车速为 40 km/h, 若测得的全部污染物均低于限值, 则合格, 否则判定受检车辆测试结果不合格^[5]。

2005 年颁布的国家汽车排放标准《点燃式发动机汽车排气污染物排放限值及测量方法》(GB 18285-2005), 规定可以采用包括加速模拟工况法 (Acceleration Simulation Mode 简称 ASM) 等来测量在用车排放^[6]。ASM 包括两种测试工况, 即 ASM 5025 与 ASM 2540, ASM 5025 工况测试车速为 25 km/h, ASM 2540 工况测试车速为 40 km/h, 测试方法基本同 BASM 方法。相比于怠速法, BASM 与 ASM 检测由于模拟了在用车道路行驶工况, 且增加了 NO_x 的排放测量, 更能反映车辆的尾气排放对大气污染的状况。

2 数据挖掘法

数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的实际数据中发现隐含的、规律性的、人们事先未知的, 但是有潜在有用的并且最终可以理解的信息和知识的非平凡过程^[7]。数据挖掘中常用的方法有关联规则的挖掘、分类挖掘、预测回归分析、聚类挖掘等。关联挖掘是从大量数据中发现描述数据项之间有价值的相互关系。

设 $I = \{i_1, i_2, \dots, i_m\}$ 是项的集合, D 为事务处理集合, 每个事务 T 是项的集合, 于是 $T \subseteq I$ 。设 A 是一个项集, 当且仅当 $A \subseteq T$ 时, 事务 T 包含 A 。

关联规则是形如 $A \Rightarrow B$ 的蕴涵式, 其中 $A \subset I, B \subset I$ 并且 $A \cap B = \emptyset$ 。

关联规则 $A \Rightarrow B$ 在事务数据库 D 中的支持度 s 是指事务数据库 D 中包含 $A \cup B$ 事务的概率, 即概率 $P(A \cup B)$, 计算式如下:

$$\text{support}(A \Rightarrow B) = P(A \cup B) = \frac{\text{support_count}(A \cup B)}{\text{support_count}(D)} \quad (1)$$

上式中 $\text{support_count}(A \cup B)$ 是指事务数据库 D 中包含 $A \cup B$ 的事务数, $\text{support_count}(D)$ 是指整

个事务数。

关联规则 $A \Rightarrow B$ 在事务数据库 D 中的置信度是指如果 D 中包含 A 事务同时也包含 B 事物的概率。这是条件概率, 计算式如下:

$$\text{confidence}(A \Rightarrow B) = P(A/B) = \frac{\text{support_count}(A \cup B)}{\text{support_count}(A)} \quad (2)$$

上式中, $\text{support_count}(A \cup B)$ 是事务数据库 D 中包含 $A \cup B$ 的事务数, $\text{support_count}(A)$ 是指包含 A 的事务数。

研究采用 SGI 公司和美国 Stanford 大学联合开发的多任务数据挖掘系统软件 MineSet。MineSet 集成多种数据挖掘算法和可视化工具, 帮助用户直观地、实时地发掘、理解大量数据背后的知识。MineSet 使用了 6 种可视化工具来表现数据和知识, 提供多种数据挖掘模式包括分类器、回归模式、关联规则、聚类、判断列重要度, 支持多种关系数据库, 可以直接从 Oracle、Informix、Sybase 的表读取数据, 也可以通过 SQL 命令执行查询, 具有多种数据转换功能的特点。

3 关联规则的挖掘

3.1 检测装置及数据来源

研究采用的排放检测装置为深圳某公司生产的 QWJ-3000 轻型汽油车 BASM 排放检测装置, 数据来源为 2004 年—2005 年北京市某车辆检测场排放检测线, 测量标准符合 DB 11/122-2003 均为轻型汽油车, 共 3 720 组车次排放检测数据。记录的数据有 BASM 5024 与 BASM 2540 两种工况下尾气放污染物 CO、HC、NO_x 测量值, 以及排放限值 (即判断值) 等, 分别用 CO 测量值 24、HC 测量值 24、NO 测量值 24、CO 判定 24、HC 判定 24、NO 判定 24、CO 测量值 40、HC 测量值 40、NO 测量值 40、CO 判定 40、HC 判定 40、NO 判定 40 来表示。

3.2 数据清洗处理

(1) 检测数据项的冗余消除。消除与该次挖掘相关性不大的属性项, 保留测量值项与判断项。

(2) 数据归约。将总的检测结果 Passed 归结为 0 或 1, 0 代表不通过检测, 1 代表通过检测。每种工况下或每种污染物测量结果归结为 N 或 Y, N 代表不通过检测或超过判断值, Y 为通过检测或不高于判断值。

(3) 噪声消除。清除检测空值项。

3.3 ASM 5024 与 ASM 2540 关联度挖掘

关联挖掘的最小支持度设置为 50%，最小置信度设置为 1%，没有对权重进行限制。采用多路关联规则。

(1) 总的检测合格率为 60.4%，说明北京在当时的车辆技术状况下按照北京地方标准 (DB 11/122-2003) 对轻型汽油车进行 BASM 排气检测，约 60% 的车辆能通过检测。

(2) $\text{support}(\text{Passed} = 1 \Rightarrow \text{BASM } 5024 = Y) = 54.4\%$

$\text{confidence}(\text{BASM } 5024 = Y \Rightarrow \text{Passed} = 1) = 90.5\%$ ，从置信度说明 BASM 5024 检测结果对总的检测结果的控制力较强。

(3) $\text{support}(\text{Passed} = 1 \Rightarrow \text{BASM } 2540 = Y) = 5.7\%$

$\text{confidence}(\text{BASM } 5024 = N \Rightarrow \text{Passed} = 1) = 9.5\%$ ，表明绝大部分车辆如果在 BASM 5024 工况下检测不合格，也很难在 BASM 2540 工况下检测通过，BASM 5024 与 BASM 2540 两种检测工况具有很高的关联性。

3.4 CO、HC 及 NO_x 关联规则的挖掘

同 3.3 为了得到更多的支持度结果，关联挖掘的最小支持度设置为 50%，最小置信度设置为 1%，没有对权重进行限制。采用多路关联规则。

输出结果如下：

(1) $\text{support}(\text{CO 判定 } 24 = Y \Rightarrow \text{passed} = 1) = 59.9\%$

$\text{confidence}(\text{CO 判定 } 24 = Y \Rightarrow \text{passed} = 1) = 69.0\%$

(2) $\text{support}(\text{CO 判定 } 24 = Y \Rightarrow \text{HC 判定 } 24 = Y) = 76.5\%$

$\text{confidence}(\text{CO 判定 } 24 = Y \Rightarrow \text{HC 判定 } 24 = Y) = 89.2\%$

(3) $\text{support}(\text{HC 判定 } 24 = Y \Rightarrow \text{NO}_x \text{ 判定 } 24 = Y) = 58.9\%$

$\text{confidence}(\text{HC 判定 } 24 = Y \Rightarrow \text{NO}_x \text{ 判定 } 24 = Y) = 71.4\%$

由 (1) 可见，CO 检测结果对总体检测结果判定的支持度为 59.9%，低于检测车辆总体检测合格率 60.4%，置信度为 69.0%。说明如果仅用一个排气检测物 CO 来代表整体的检测合格率不够全面。

由 (2) 可见，CO 的检测结果对 HC 的检测结果判定有 76.5% 的支持度，置信度也有 89.2%。

由 (3) 可见 HC 检测合格对 NO_x 检测合格的支持度是 58.9%，置信度是 71.4%。

可见 CO 对 HC 的检测置信度比 HC 对 NO_x 的检测支持度高出近 20 个百分点，说明 HC 与 CO 存在较高的关联度，而 HC 与 NO_x 关联度不高。主要原因是由于 CO 与 HC 形成条件类似，CO 排放大，HC 排放往往也不低，而 HC 与 NO_x 生成条件相差较大，HC 排放大时，NO_x 排放并不高。

4 结语

采用数据挖掘法对汽油车 BASM 排放检测方法中的检测工况之间及排放污染物检测项之间的关联性进行研究。利用 MineSet 数据挖掘软件对 3720 台次汽油车 BASM 排放检测数据进行了关联挖掘，得出以下结论。

(1) 在 BASM 5024 检测工况下没有通过检测的车辆在 BASM 2540 检测中的通过率只有 5.7%，说明两种检测工况具有很高的关联性。

(2) CO 检测合格对 HC 检测合格有 76.5% 的支持度，HC 检测合格对 NO_x 检测合格的支持度为 58.9%，说明 HC 与 CO 存在较高的关联度，而 HC 与 NO_x 关联度不高。

[参考文献]

- [1] 彭美春, 赵铎杰, 许志刚, 等. 轻型汽车 ASM 排放特性 [J]. 交通运输工程学报, 2005, 5(2): 65-67.
- [2] 吴义虎, 周育才, 张利军, 等. 车用汽油机污染物排放特性模糊分析方法及应用研究 [J]. 中国公路学报, 2001, 14(3): 101-105.
- [3] 彭美春, 周桂添, 王文涛, 等. 轻型汽油车工况法排放检测方法相关性研究 [J]. 汽车技术, 2005, 363(12): 31-34.
- [4] 王文涛, 彭美春, 周桂添, 等. 在用汽油车瞬态排放测试工况的相关性 [J]. 环境监测管理与技术, 2005, 17(6): 36-39.
- [5] 北京市环境保护局, 北京市技术监督局. DB 11/122-2003 汽车稳态加载污染物排放标准 [S]. 北京: 中国标准出版社, 2003.
- [6] 国家环境保护总局, 国家质量监督检验检疫总局. GB 18285-2005 点燃式发动机汽车排气污染物排放限值及测量方法 [S]. 北京: 中国环境科学出版社, 2005.
- [7] DUNHAM H M. 数据挖掘教程 [M]. 郭崇慧, 田凤占, 靳晓明, 等译. 北京: 清华大学出版社, 2005.