

基于知识发现的城市交通流状态快速辨别方法研究

汪凌¹, 查伟雄²

(1. 安徽理工大学 经济管理学院 淮南 232001; 2. 华东交通大学 交通运输与经济研究所 南昌 330013)

摘要: 针对智能交通系统开发和城市交通流特性, 提出了一种基于知识发现的城市交通流状态快速辨别方法。该方法采用粗糙集理论从采集的大量交通流数据中获取“交通流参数—交通状态”间的映射关系知识, 构建了相应的交通流状态辨别知识库。在交通流状态知识发现模型和知识库的支持下, 给出了实时交通流状态的辨别算法。将本文算法与其他传统算法进行对比分析, 结果表明本文算法具有更好的实时交通流状态辨别性能。

关键词: 粗糙集; 交通流状态辨别; 知识发现模型; 知识库

中图法分类号: TP182 文献标识码: A

0 引言

交通流状态辨别是智能交通系统开发中的一个重要内容。实时准确地判别城市交通流状态是正确实施交通控制和交通诱导的前提, 对于提高城市动态交通管理水平, 解决城市道路交通拥挤问题具有十分重要的理论和现实意义。

为了准确的辨别城市交通流状态, 近年来国内外很多学者提出了多种交通流理论模型^[1-3], 以及模糊推理^[4,5]、模糊聚类^[6]、模糊神经网络^[7,8]、支持向量机^[9]等交通流状态辨别方法, 并取得了一定的效果。但这些研究均存在诸多局限, 如模糊推理方法需要完备的专家知识确定推理规则, 且缺乏学习能力, 对交通流变化适应性较差; 模糊聚类方法是一种局部搜索技术, 具有初始值敏感性、易陷入局部最小值; 神经网络存在训练耗时、网络结构不确定性及易陷入局部最优等问题; 支持向量机具有泛化能力较强, 克服了对训练数据的依赖性的优势, 但存在计算时间较长, 影响城市交通流状态的在线分类等缺点。

鉴于上述分析, 本文提出一种基于知识发现的城市交通流状态快速辨别方法。该方法的核心是将城市交通流的特征数据和领域专家的先验知识提炼成一系列产生式规则, 构造城市交通流状态辨别知识库, 研制与这些规则适应的推理引擎对交通流状态辨别知识库与交通流参数进行匹配处理, 从而实现交通流状态的实时快速辨别。

1 城市交通流特征数据

城市交通流状态辨别是以城市交通系统交通流特征数据为基础的。交通流是指由车流量、平均车速、车辆密度、道路占有率等参数描述的状态量, 反映了城市道路交通管理系统所处的交通状况。交通流量在时空和方向上的分布不均匀性造成不同时间、路段、方向上交通状况也不同; 平均车速主要集中在 10-60km/h 之间, 不同车速呈现出不同的交通状态; 道路占有率在同一个交通流量下也可能表现不同的交通状态等。

交通流状态变化规律在不同的路段或交叉口会呈现不同的特点。对典型的路段或交叉口而言, 主要有畅通模式、正常模式、拥挤模式以及堵塞模式四种交通流状况。其中畅通模式的服务水平最高, 其他等级服务水平逐次降低。

在城市交通流状态辨别中, 通常采集到的交通流特征数据包括: 交通流量、平均车速、车辆密度、道路占有率、排队长度、以及 OD 矩阵等。而在交通信号控制、交通流分配以及路径诱导过程中, 通常选取交通流量、平均速度以及道路占有率作为交通流状态辨别的特征参数。因此本文也以交通流量、平均车速及道路占有率等为基础对城市交通状态辨别问题进行分析与研究。

2 粗集交通流状态知识发现模型

利用粗糙集方法来获取交通流状态知识，其核心是建立“交通流参数—交通状态”之间的映射关系模型。粗糙集规则提取的矩阵方法是通过两两比较所有对象在某一属性上取值的构造所有属性的 0-1 矩阵，进而求出条件属性矩阵与决策属性矩阵等用于从不完备交通流状态决策表中提取规则。该方法比常规的粗糙集规则获取方法更为简单，无需粗糙集的约简步骤，能够直接从不完备决策系统中提取所有的决策规则集，进而发现交通流参数与交通流状态之间的变化规律知识。同时，我们将规则的支持度和确信度评价指标引入到本知识发现方法中，有效地提高了规则的抗噪能力和解释评价能力。

2.1 粗集规则获取的矩阵方法

对于不完备信息决策系统 $DS = \langle U, C \cup D, V, f \rangle$ ，其中 U 为论域， C 、 D 分别为条件属性集和决策属性， $C \cap D = \emptyset$ ， $V = \bigcup_{a \in C \cup D} V_a$ ， V_a 是属性 a 的值域， $f: U \times A \rightarrow V$ 是一个信息函数， $\forall a \in C$ ，

$x \in U$ ， $f(x, a) \in V_a$ 。由属性子集 $\forall \emptyset \subset B \subseteq C$ 定义相容关系

$T(B) = \{(x, y) \mid \forall a \in B, f(x, a) = f(y, a) \vee f(x, a) = * \vee f(y, a) = *\}$ 及

$T_B(x) = \{y \in U \mid (x, y) \in T(B)\}$ 。定义 B 的相容条件属性矩阵 $M_B = (m_{ij}^B)_{n \times n} = \begin{cases} 1 & x_j \in T_B(x_i) \\ 0 & otherwise \end{cases}$ ，

$i, j = 1, 2, \dots, n$ ， m_{ij} 为第 i 行 j 列元素。记 $\partial_B(x)$ 表示 $T_B(x)$ 中元素在 d 上取值集合，

$\partial_B(x) = \{i \mid i = f(y, d), y \in T_B(x)\}$ ，则由 $\partial_B(x)$ 定义决策属性矩阵

$M_D = (m_{ij}^D)_{n \times n} = \begin{cases} 1 & f(x_j, d) \in \partial_C(x_i) \\ 0 & otherwise \end{cases}$ ， $i, j = 1, 2, \dots, n$ ， m_{ij} 为第 i 行 j 列元素。显然， M_D 是主对角线

为 1 的 n 阶对称矩阵。

2.2 交通流状态辨别规则获取的矩阵算法

不完备信息决策系统 $DS = \langle U, C \cup D, V, f \rangle$ 规则获取的矩阵算法步骤如下：

Step1 计算 DS 中条件属性集 C 对应的相容条件属性矩阵 M_C 和决策属性矩阵 M_D ；

Step2 对任意条件属性 $c \in C$ ，计算一阶相容条件矩阵 $M_{\{c\}} = (M_{\{c\}}^1, M_{\{c\}}^2, \dots, M_{\{c\}}^n)$ ；

Step3 for $i=1$ to n ；

① 将一阶相容条件矩阵 $M_{\{c\}}$ 的第 i 行与决策属性矩阵 M_D 的第 i 行相交；

② 若存在 $M_{\{c\}}$ 的第 i 行与 M_D 的第 i 行不一致，next i ，否则转 Step4；

Step4 根据对象 x_i 的属性第 i 行对应的属性 $c_m \in C$ ， $d_n \in D$ 的取值生成决策规则：

$\wedge (c_m = v_{c_m}) \rightarrow \wedge (d_n = v_{d_n})$ ，同时将第 i 行置为零，提取出所有的一阶决策规则；

Step5 将 Step2 得到的所有非零一阶相容条件矩阵两两相交，得到满足条件的所有二阶相容矩阵 $M_{\{c_s, c_t\}}^2 (s \neq t, M_{\{c_s, c_t\}}^2 \neq M_{\{c_s\}}, M_{\{c_s, c_t\}}^2 \neq M_{\{c_t\}})$ ，再对所有非零二阶相容矩阵按 Step3-Step4 的处理方法提

取所有的二阶决策规则；

Step6 用类似 Step5 的处理方法提取所有三阶以上的决策规则，直到没有非零的同阶相容矩阵为止，最后输出全部决策规则集。

2.3 交通流状态辨别知识库的建立

交通流状态辨别知识库是用来存储交通流特征数据。通过特征提取，对不同交通流状态信息获取相关的平均速度、车流量和道路占有率等特征参数，并存入数据库中进行实时辨别。知识库主要包括样本库和规则库，其中样本库是由样本的独特特征和训练好的神经网络权值组成，而规则库是由一定辨别条件推出确定的交通流状态知识。交通流状态辨别知识大多源于领域专家对交通流信息描述的经验、常识，这些知识在表达上具有模糊性和不确定性，因此交通流状态辨别主要通过模糊产生式规则形式来表达知识，其一般形式为：IF (Condition) THEN (Conclusion) AND CF (Certainty Factor)

其中，CF 是对应知识规则的确信度系数，且 $CF \in [0,1]$ ，若 $CF=0$ ，则规则对应的命题完全为假；若 $CF=1$ ，则规则对应的命题完全为真，否则它介于两者之间。用笛卡尔乘积的形式表示为： $\langle \text{Condition}, \text{Conclusion}, \text{CF} \rangle$ 。如“非常接近”等具有不确定性的概率表达可采用如下形式：IF $((30 < \text{Average Speed} < 45) \wedge (100 < \text{Traffic Flow} < 150))$ THEN Pattern is Normal, AND CF is 0.67；

通常情况下，知识库中包含的知识越多，辨别结果就越准确，但也要防止知识规则的相互冲突带来的负面影响。采用 KDD 中关联规则挖掘算法即可实现交通流状态信息特征的提取并获取知识。将这些完整规则存入临时库，然后请领域专家逐一检查这些规则，通过存储有价值的规则知识，使得交通流状态辨别知识库不断更新和完善。

3 基于知识发现的城市交通流状态辨别

3.1 城市交通流状态划分

一般而言，由于城市各路段的交通条件和交通诱导系统不同，交通状态的划分是不同的。例如，有些交通诱导系统将交通流分为三类，采用红、黄、绿色表示其拥挤程度；国内将高等级公路服务水平分为四级，采用车速和密度作为交通状态特征参数；美国 HCM 将城市主干道服务水平分为 A 至 F 六级，采用平均车速作为划分指标。本文按照交通诱导系统，结合国内对公路服务水平划分和美国 HCM 对主干道服务水平划分标准，将城市道路交通流划分为畅通、正常、拥挤和堵塞四个等级，既可以真实的反映交通流微观运行状态，也可避免因分类过细造成各状态间差异性降低。

3.2 基于知识发现的交通流状态辨别算法

交通流状态辨别主要思路是，从多个待定的交通流参数中选取交通流状态判别指标，即以当前采集的交通流特征数据作为待辨别的数据，训练样本作为分类样本，通过计算待辨别交通流特征与知识库中各分类规则的匹配强度，实现对交通流状态的辨别。根据上述思路，设计基于知识发现的交通流状态辨别算法如下：

Step1 根据交通数据采集条件及交通流参数之间的关联关系，从车流量、平均车速、车流密度、道路占有率、路段行程等参数中，选取 n 个相互独立的交通流特征属性，计算待测样本的交通流特征与知识库各规则的匹配强度因子 $match(k)$ ；

$$match(k) = \frac{1}{n} \sum_{i=1}^n I_{i,k}, \quad I_{i,k} = \begin{cases} 1 & a_i(e) \in [x_{2,i-1}, x_{2,i}] \\ 0 & \text{else} \end{cases}, k = 1, 2, \dots, m$$

其中， m 表示分类规则数， $I_{i,k}$ 表示若辨别样本的第 i 个属性值位于规则前件第 i 项的取值范围内，则

$I_{i,k} = 1$ ，否则 $I_{i,k} = 0$ ；

Step2 选取匹配强度因子 $match(k)$ 最大的分类规则进行决策, 若 $match(k)$ 最大的分类规则仅有一条, 则按此规则决策, 判定其交通流状态类别, 否则转 Step3;

Step3 若有两条及以上的分类规则具有相同且最大的匹配强度因子 $match(k)$, 则对于决策类别相同的分类规则, 计算规则评价函数 $EvalFunc=Cer(X,Y)*Sup(X,Y)$, 选取 $EvalFunc$ 和值最大的决策属性值作为待识样本的决策属性值判定其模式; 若有多个模式满足 $EvalFunc$ 和值最大, 则以所属个体数最多的作为识别样本的决策属性值判别其模式;

Step4 若待识样本的交通流特征属性不与任何规则相匹配, 我们采用 K-近邻算法进行交通流状态辨别。

4 应用实例

为检验本文提出的交通流状态辨别算法的有效性, 以某典型的城市交叉口采集到的交通流数据为例。图 1 为该交叉口相位示意图, 其中序号 1、2、3、4 表示检测器, a_1 、 b_1 、……、 d_2 分别表示各车道车流方向。

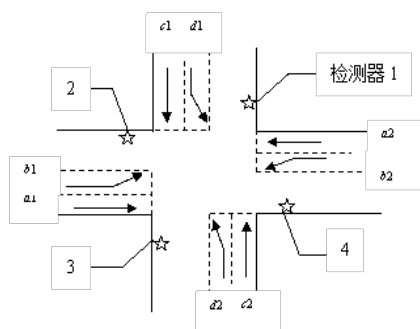


图 1 某典型交叉口相位示意图

4.1 数据准备

本文从该交叉口的交通历史数据库中选取连续三个工作日内时间跨度为 6:00-18:00, 采样间隔为 5min、采样周期为 20s 的交通流数据集, 并将中、小车流量折算成标准车流量(f), 平均车速(s)和道路占有率(d)相应变化, 剔除明显异常数据, 整理成 3428 条有效样本数据。从有效样本数据中随机抽取 70% 的数据作为样本训练集, 其余的 30% 作为测试集。部分样本的交通流特征参数及交通状态如表 1 所示。

表 1 部分样本特征参数及交通状态

检测数据	样本编号										
	1	2	3	4	5	6	7	8	9	10	...
流量(pcu/5mins)	152.5	140.7	139.1	93.2	85	95.8	107.6	116.5	120.4	115.4	...
速度(km/h)	51.6	28.8	57.2	47.5	65.8	70.2	68.6	69.8	62.7	67	...
占有率(%)	23	36.5	19.7	16.5	8.6	9.5	9.2	13.5	12.9	14.8	...
状态(M)	拥挤	拥挤	拥挤	畅通	畅通	畅通	畅通	正常	正常	正常	...

4.2 分类与辨别

首先运用 Rosetta 软件对随机抽取的样本训练集进行离散化和特征约简处理, 然后依据不完备信息系统的规则提取矩阵算法从训练样本集中获取交通流状态分类规则, 并对测试样本的交通流状态进行辨别。表 2 第一列给出了部分训练样本的交通流状态分类规则集。

分别应用变精度粗集(VPRS)算法^[10]、基于支持度、一致度及覆盖度的粗集(SCC)算法^[11]对相同的训练样本进行规则获取, 得决策规则如表 2 中相应列所示。为说明本文算法中交通流状态决策规则比其他两种算法的决策规则更有效, 现随机抽取有效数据集中 15 组数据作为测试样本, 用三种算法预测交通流状态与真实交通流状态进行对比, 其结果如表 3 所示。

表2 部分样本的交通流状态分类规则集

本文算法规则	确信度	支持度	VPRS 算法规则	SCC 算法规则
$(s_2=2) \wedge (f_3=2) \wedge (f_4=2) \rightarrow M=N$	1	0.10	$(s_2=4) \wedge (f_3=2) \wedge (f_4=2) \rightarrow M=C$	$(s_2=4) \wedge (f_3=3) \wedge (f_4=2) \rightarrow M=N$
$(s_2=4) \wedge (f_3=3) \wedge (f_4=3) \rightarrow M=C$	1	0.10	$(s_2=3) \wedge (f_3=4) \wedge (f_4=2) \rightarrow M=C$	$(s_2=3) \wedge (f_3=4) \wedge (f_4=4) \rightarrow M=C$
$(s_2=3) \wedge (f_3=4) \wedge (f_4=2) \rightarrow M=C$	1	0.10	$(s_2=3) \wedge (f_3=2) \wedge (f_4=2) \rightarrow M=C$	$(s_2=3) \wedge (f_3=2) \wedge (f_4=3) \rightarrow M=C$
$(s_2=2) \wedge (f_4=3) \rightarrow M=N$	1	0.20	$(s_2=2) \wedge (f_4=2) \rightarrow M=N$	$(s_2=2) \wedge (f_3=3) \wedge (f_4=2) \rightarrow M=N$
$(s_2=4) \wedge (f_3=3) \wedge (f_4=2) \rightarrow M=C$	1	0.10	$(s_2=2) \wedge (f_3=4) \wedge (f_4=3) \rightarrow M=C$	$(d_1=2) \wedge (f_2=3) \rightarrow M=N$
$(s_2=2) \wedge (f_3=3) \wedge (f_4=2) \rightarrow M=N$	1	0.10	$(s_2=2) \wedge (f_3=4) \wedge (f_4=4) \rightarrow M=N$	
			$(s_2=2) \wedge (f_3=3) \wedge (f_4=2) \rightarrow M=C$	

表3 三种辨别算法的预测结果对比

交通流状态真实值	交通流状态预测值	案例数		
		本文算法	VPRS 算法	SCC 算法
S(畅通)	无匹配规则	0	2	5
N(正常)	N(正常)	2	1	1
C(拥挤)	C(拥挤)	9	9	9
J(堵塞)	J(堵塞)	4	3	0

从表3可以看出,与常规的VPRS算法和SCC算法进行规则分类的预测结果相比,本文算法由于引入0-1矩阵进行规则获取,使得预测准确率大幅提高;同时通过引入矩阵算法进行交通流状态辨别,也可以使得预测结果更加完备精细。将有效的规则知识存储于交通流状态知识库中,促进交通流状态知识库不断得到补充完善,从而为不完全信息下的交通流状态辨别提供了一条全新的思路和途径。

5 结论

交通流状态辨别对于交通控制和诱导具有非常重要的作用。本文提出了一种基于知识发现的城市交通流状态快速辨别方法,在积累大量交通流历史特征数据及交通状态的基础上,建立了“交通流特性—交通状态”的知识发现模型,构建了相应的交通流状态识别知识库,并给出了待测对象的交通流状态辨别算法。以某典型交叉口的交通流状态辨别为例分析,验证了该算法具有较好的交通流状态辨别性能,从而为采取城市交通控制管理措施和交通诱导策略提供科学依据。

参考文献:

- [1] Helbing D, et al. Micror and macro-simulation of freeway traffic[J]. Mathematical and Computer Modeling, 2002,35: 547
- [2] Kerner B S. Three-phase traffic theory and highway capacity[J]. Physica A: Statistical Mechanics and its Applications, 2004, 333: 379-440
- [3] Treiber M A, Kesting and Helbing D. Delays, inaccuracies and anticipation in microscopic traffic models[J]. Physica A: Statistical Mechanics and its Applications, 2006,360(1): 71-88
- [4] 戢晓峰, 刘澜, 何增辉. 基于模糊推理的区域路网交通状态分析方法[J]. 交通运输工程与信息学报, 2009, 7(3): 27-32
- [5] 蒲世林, 李瑞敏, 史其信. 基于粗糙集-模糊识别技术的交通流状态识别算法研究[J]. 武汉理工大学学报(交通科学与工程版), 2010, 34(6): 1154-1158
- [6] 陈德望. 基于模糊聚类的快速路交通流状况分类[J]. 交通运输系统工程与信息, 2005, 5(1): 62-67
- [7] 王辉. 基于自适应模糊神经网络的交通流状态预测[J]. 交通与计算机, 2007(4): 46-49
- [8] Yang H, Bell M G H. Models and Algorithms for road network design: A Review and some new developments[J]. Transport Review, 1998, 18(3): 257-278
- [9] Adel W Sadek, Michael J Demetsky, Brain L Smith. Case-Based Reasoning for Real-time traffic flow management[J]. Computer-Aided Civil and Infrastructure Engineering, 1999(14): 347-356
- [10] Ziarko W. Variable precision rough set model[J]. Journal of Computer and System Sciences, 1993, 46(1): 39-59
- [11] 菅利荣, 达庆利, 陈伟达. 基于粗糙集的不一致信息系统规则获取方法[J]. 中国管理科学, 2003, 11(4): 91-95