

粗糙集理论和遗传算法在预防城市 道路交通拥堵中的应用

马吉明, 黄宪芳, 蒋亚平, 王秉政

(郑州轻工业学院 计算机与通信工程学院, 河南 郑州 450002)

摘要:建立了反映道路拥堵原因与道路拥堵类型之间的关系决策表,利用粗糙集理论和遗传算法相结合的属性约简方法,去除决策表中冗余的规则,从而找到了造成道路交通拥堵的主要因素.试验结果表明,此算法可高效分析交通拥堵原因,为预防城市道路交通拥堵提供科学的决策依据.

关键词:交通拥堵;粗糙集理论;遗传算法

中图分类号:TP399 **文献标志码:**A

Application in prevent urban road traffic congestion based on rough set theory and genetic algorithm

MA Ji-ming, HUANG Xian-fang, JIANG Ya-ping, WANG Bing-zheng
(College of Comp. and Com. Eng., Zhengzhou Univ. of Light Ind., Zhengzhou 450002, China)

Abstract: The decision table of the relationship between reason and type of reflecting road congestion was established. The attributes reduction method based on rough set theory combined with genetic algorithm was presented. This method removes redundant rules from decision table and finally the main factors causing road traffic congestion are found. The test result showed that this algorithm in analysis of causes for traffic congestion is efficient, and that it provides scientific basis for decision making of preventing urban road traffic congestion.

Key words: traffic congestion; rough set theory; genetic algorithm

0 引言

交通拥堵是城市普遍存在的一个问题,不仅影响城市运营的效率和质量,而且带来了环境污染等一系列社会问题.目前许多城市通过建设立交桥、地铁或高架公路等交通硬件来解决城市交通拥堵问题,这仅是解决交通拥堵有效手段之一,而建设一个基于信息技术的智能交通系统才是根本.智能交通系统是一个复杂的综合性系统,而找出交通拥

堵问题产生的原因是智能交通系统建立的主要目的之一.城市道路交通拥堵往往是由天气状况、道路设施质量、上下班时间段、节假日等^[1-2]多种因素综合作用产生的,但这些因素在不同情况下对交通拥堵的影响有大有小,有时甚至可以忽略不计,因此存在许多冗余信息.

粗糙集理论能够去除数据中的冗余信息而保持原有的分类能力不变;而遗传算法是一种自适应全局优化搜索算法,是模拟生物在自然环境中的遗

传和进化过程而形成的,具有广泛的适用性.因此本文拟利用粗糙集理论和遗传算法,从海量数据中挖掘出有利于分析城市道路交通拥堵的信息,以期预防城市道路交通拥堵提供科学的决策依据.

1 粗糙集和遗传算法的基本理论

1.1 粗糙集相关概念

定义1 四元组 $S = (U, A, V, f)$ 是一个决策表.其中, U 是对象的有限集合,也称为论域; $A = C \cup D, C \cap D = \emptyset, A$ 表示属性集, C 表示条件属性集, D 表示决策属性集; $V = \bigcup_{a \in A} V_a$ 是属性的值域集, V_a 表示属性 $a \in A$ 的属性值范围,即属性 a 的值域;信息函数 $f: U \times A \rightarrow V$, 指定 U 中每一个对象 a 的属性值. **定义2** 约简:令 R 为一族等价关系,记作 $\text{ind}(R)$. $r \in R$, 若 $\text{ind}(R) = \text{ind}(R - \{r\})$ 则称 r 为 R 中不必要的;否则称 r 为 R 中必要的.如果每一个 $r \in R$ 都为 R 中必要的,则称 R 为独立的;反之称 R 为依赖的.设 $Q \subseteq P$, 若 Q 是独立的,且 $\text{ind}(Q) = \text{ind}(P)$, 则称 Q 为 P 的一个约简.

定义3 决策属性 d 对条件属性 c 的依赖度: $\gamma_c(d) = \frac{|\text{pos}_c(d)|}{|U|}$, 其中, $|U|$ 表示 U 的基数, $\text{pos}_c(d) = \bigcup_{x \in U/d} c(x)$ 是决策属性 d 关于条件属性 c 的正域^[3].

1.2 遗传算法相关理论

1.2.1 染色体编码 采用长度为 m (m 表示条件属性的个数)的二进制字符串 $\{0, 1\}$ 来表示每个个体编码,编码的每一位对应于1个条件属性.例如: (C_1, C_2, \dots, C_m) , 如果个体中包含第 i ($i = 1, 2, \dots, m$) 个属性,则 $C_i = 1$; 否则, $C_i = 0$.

1.2.2 适应度函数 $F(r) = \frac{m - m_r}{m} + \gamma_c(d)$. 其中, m_r 表示染色体 r 中基因为1的个数, m 为染色体 r 的长度, $\gamma_c(d)$ 表示决策属性 d 对条件属性 c 的依赖度.

1.2.3 遗传算子^[4] 1) 选择. 通过轮盘赌轮方式来实现,首先计算每个染色体 r 的适合度值 $F(r_i)$, 找出群体的适合度之和 $\sum_{i=1}^m F(r_i)$, 之后计算每个染色体的选择概率,即 $P(r_i) = F(r_i) / \sum_{i=1}^m F(r_i)$, 得出每个被包括的染色体 r_i 的累积概率 $q_i = \sum_{i=1}^m P(r_i)$.

如果 $r < q_i$, 选择第1个染色体 r_i ; 否则选择第 i 个染色体 r_i , 使得 $q_{i-1} < r \leq q_i$. 2) 交叉. 采用单点交叉算子,根据遗传代码随机选择一个交叉点,之后以交叉概率 p_c 在此交叉点处相互交换2个个体的部分染色体,从而产生2个新的个体. 3) 变异. 采用均匀变异算子,根据变异概率 p_m 随即反转某位等位基因的二进制字符值. 4) 停止规则. 当迭代的步数达到最大迭代数 T 或者 $\gamma_{\text{reduct}(c)}(d) = \gamma_c(d)$ 时停止算法运算.

2 基于粗糙集和遗传算法的城市道路交通拥堵分析

2.1 城市道路交通拥堵信息表的确定

以从郑州市交通管理部门获取的某路段的980组数据作为论域,导致城市道路交通拥堵的因素作为条件属性集,城市道路交通拥堵类型作为决策属性集.将车辆发生交通事故、车辆违法并线、公交车进出站频繁、出租车不规范停车、行人是否违章、道路施工、天气状况、交通流量、高峰时期9个因素作为条件属性,分别用 C_1, C_2, \dots, C_9 表示;交通拥堵类型 D 为决策属性.通过对原始交通拥堵的数据预处理,即数据除噪、离散化、归一化,建立交通拥堵决策表,如表1所示.其中, $C_1: 0$ 表示车辆未发生交通事故,否则用1表示; $C_2: 0$ 表示车辆未违法并线,否则用1表示; $C_3: 0$ 表示公交车进出站不频繁,否则用1表示; $C_4: 0$ 表示出租车规范停车,否则用1表示; $C_5: 0$ 表示行人不违章,否则用1表示; $C_6: 0$ 表示没有道路施工,否则用1表示; $C_7: 0$ 表示晴,1表示雨,2表示雾; $C_8: 0$ 表示交通流量密度小,否则用1表示; $C_9: 0$ 表示不是高峰时期,否则用1表示.此处考虑的9个因素是造成交通拥堵的主要原因,而其他的一些次要因素被忽略. $D: 0$ 表示轻度拥堵,1表示拥堵,2表示严重拥堵.

2.2 属性约简算法

基于遗传算法约简删除冗余条件属性得到简化决策表的具体算法步骤如下:

输入:决策表 $S = (U, A, V, f)$.

输出: S 的所有约简.

1) 计算出决策属性 D 关于条件属性 C 的支持度 $\gamma_c(D)$.

2) 令 $\text{reduct}(C) = \emptyset$, 逐一去掉个属性 $C_i \in C$, 若 $\gamma_{C - \{C_i\}}(D) \neq \gamma_c(D)$, 则 $\text{reduct}(C) = \text{reduct}(C) \cup$

$\{C_i\}$;若 $\gamma_{\text{reduct}(C)}(D) = \gamma_C(D)$,则终止计算, $\text{reduct}(C)$ 表示属性 C 的约简,否则进行3).

表1 交通拥堵决策

U	C									D
	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	
1	0	1	0	0	0	1	0	0	0	0
2	0	1	1	0	0	1	1	1	1	2
3	1	1	1	0	0	1	2	1	1	2
4	1	1	1	0	1	1	2	1	1	2
5	1	1	0	1	1	1	1	1	0	2
6	1	0	1	1	1	1	2	1	1	2
7	0	0	0	1	0	0	1	0	1	0
8	0	1	0	1	0	0	2	0	1	1
9	1	1	1	1	0	0	0	1	1	1
10	1	0	0	0	1	1	0	1	0	2
11	0	0	1	0	1	1	1	1	0	1
12	0	0	1	0	1	1	0	1	0	1
13	0	1	1	1	0	1	2	1	1	0
14	1	1	1	1	0	0	2	0	0	1
15	0	1	1	1	0	1	2	0	0	1
16	0	1	0	0	1	1	1	0	1	2
17	1	0	0	0	1	1	0	0	1	1
18	1	0	0	1	1	0	1	0	1	1
19	0	0	1	0	1	0	1	1	0	1
20	0	1	1	1	0	0	2	1	1	1
21	1	1	1	0	0	1	1	1	1	2

3)由随机产生的 p 个长度为 m 的二进制串所代表的个体组成初始种群:对应位随机取0或1,并计算初始群体中每个个体的适应度.

4)依轮盘赌方法选择个体,通过交叉概率 p_c 和变异概率 p_m 产生新一代的群体,并且在变异时保持该属性对应的基因位不发生变异.

5)计算新一代群体中每个个体的适应度.

6)依据最优保存策略将最优个体保留到下一代群体中,并且依最优保存策略保存最优个体.

7)判断是否连续 t 代的最优个体适应值不再提高:若是则终止计算并输出最优个体;否则转至4).

2.3 道路交通拥堵分析决策规则生成

交通拥堵的预警规则是由属性约简去除冗余条件属性及重复信息得到简化信息表而生成.

3 结果分析

选取 $p_c = 0.7, p_m = 0.01, T = 100$,得最优解为100001111,即 C_1, C_6, C_7, C_8, C_9 属性被保留,从而得

到约简后的决策规则如表2所示.

表2 决策规则

C ₁	C ₆	C ₇	C ₈	C ₉	D
0	0	0	*	0	0
0	0	2	1	1	0
1	0	0	1	*	0
0	1	1	0	1	0
1	1	*	0	0	0
1	0	1	0	1	1
1	1	2	0	1	1
1	0	1	1	0	1
0	1	0	1	1	1
1	1	1	1	1	2

从决策规则中可以看出,造成交通拥堵的因素主要是交通事故、道路施工、天气状况、交通流量和高峰时期.在雨雾天气、高峰时期、无交通事故发生但交通流量密度大时,城市交通易发生轻度拥堵;在高峰时期又发生交通事故,且是雨雾天气时,则易发生拥堵;在雨雾天气、有交通事故发生、道路施工且交通流量密度大时,则易发生严重的城市交通拥堵.因此,无论拥堵的程度大小如何,天气状况和高峰期都是造成城市交通道路拥堵的重要因素,因此人们在出行前应多关注天气状况和出行时刻是否处于上下班高峰期或假期,从而避免交通拥堵;城市交通管理部门则需在这些时刻增加交通指挥员,做好道路疏通工作,不断强化交通参与者的交通法规意识,从而降低交通拥堵的发生率.

4 结论

本文根据交通历史数据利用粗糙集理论和遗传算法提取出有效规则,找出了造成交通拥堵的主要因素,为建立智能交通系统提供科学依据,从而有效防止交通拥堵的发生.

参考文献:

[1] 管满泉.从道路因素分析交通拥堵的成因和对策[J].中国人民公安大学学报:自然科学版,2005(4):62.
 [2] 房丽侠,魏连雨,闫伟阳.基于决策树的交通拥挤状态预测[J].河北工业大学学报,2010,39(2):105.
 [3] 张文修,吴伟志,梁吉业,等.粗糙集理论与方法[M].北京:科学出版社,2005.
 [4] 刘洪伟,石红瑞.遗传算法在汽车巡航控制系统中的应用[J].自动化仪表,2009,30(11):48.