

基于决策树的消费行为因素建模与实现

黎 旭^{1 2} 李国和^{1 2 3} 吴卫江^{1 2 3} 洪云峰³ 刘智渊³ 程 远³

¹(中国石油大学(北京) 地球物理与信息工程学院 北京 102249)

²(中国石油大学(北京) 油气数据挖掘北京市重点实验室 北京 102249)

³(石大兆信数字身份管理与物联网技术研究院 北京 100029)

摘 要 消费行为因素分析对产品生产和销售具有重要指导作用。为了利用消费者的消费数据进行消费行为建模和分析, 首先进行消费数据形式化表示, 形成消费客户交易数据集和交易统计信息表达。然后在消费客户交易数据集上定义信息增益率, 反映消费因素的分类能力。在 C4.5 算法基础上, 改进二分法为多分法, 对连续型属性(因素) 进行离散化, 并建立决策树。决策树每一分支构成决策规则, 反映消费者的消费因素之间的依赖关系。每条规则的统计信息表示决策规则的不确定性。采用 Web 体系架构, 以 Oracle 为数据库, 实现了消费行为建模与分析系统, 该系统不仅消费行为模型分析精度高, 而且具有高效性和友好性。

关键词 决策树 C4.5 算法 信息增益率 连续型属性 因素分析

中图分类号 TP3 文献标识码 A DOI: 10.3969/j.issn.1000-386x.2015.05.045

MODELLING AND IMPLEMENTATION OF DECISION TREE-BASED CONSUMPTION BEHAVIOUR FACTORS

Li Xu^{1 2} Li Guohe^{1 2 3} Wu Weijiang^{1 2 3} Hong Yunfeng³ Liu Zhiyuan³ Cheng Yuan³

¹(College of Geophysics and Information Engineering, China University of Petroleum, Beijing 102249, China)

²(Beijing Key Lab of Data Mining for Petroleum Data, China University of Petroleum, Beijing 102249, China)

³(PanPass Institute of Digital Identification Management and Internet of Things, Beijing 100029, China)

Abstract The analysis on consumption behaviour factors plays an important guiding role on production and sales of products. In order to use consumers' consumption data to model and analyse the consumption behaviours, first the formalised presentation of consumption data is made to form the consumer transaction data sets and the transaction statistics expression. Then, on consumer transaction data sets the information gain-ratio is defined to reflect the classification ability of the consumption factors. On the basis of C4.5 algorithm, the bi-segmentation is improved to multi-segmentation, the discretisation is applied to continuous attributes (namely factors), and the decision tree is constructed as well. Each branch of the decision tree forms a decision rule which reflects the dependency relationship between the consumption factors of consumer. Statistical information of each rule expresses the uncertainty of the decision rule. By means of WEB architecture and using Oracle as database, the modelling and analysis system of consumption behaviour is implemented, which not only has high accuracy in consumption behaviour model analysis, but is also high efficient and friendly.

Keywords Decision tree C4.5 algorithm Information gain-ratio Continuous attributes Factors analysis

0 引 言

消费行为分析对企业的产品生产和销售具有重要指导意义。根据消费者的消费数据, 建立消费行为因素分析模型, 发现不同因素间的关系。消费行为因素分析大多依据经济学、社会学和心理学的理论^[1-3], 具有很强的主观性。数据挖掘为消费行为因素分析建模提供了新的手段^[4-6], 如结合聚类和关联规则挖掘算法, 为每个消费者提供购物指南; 利用 Clementine 软件导出决策树和规则集, 指导销售人员根据消费者的外表进行营销。决策树反映条件因素和决策因素的关系, 其每一分支构成一条决策规则。对于 C4.5 算法实现决策树挖掘, 对连续型条件

属性(条件因素, 如年龄) 只进行一次最优二分, 无法实现多次二分(多区间化), 导致决策规则准确率降低。

针对消费建模分析的需要, 改进 C4.5 算法, 对连续型条件属性进行多次二分区化, 以提高消费因素建模分析的精度, 并采用 Web 体系结构, 以 Oracle 为数据库, 实现了消费行为分析建模系统, 集成到北京兆信信息技术股份有限公司的消费行为分析系统中, 取得很好的应用效果。

收稿日期: 2013-10-22。国家自然科学基金项目(60473125); 国家高新技术研究发展计划项目(2009AA062802); 中国石油(CNPC) 石油科技中青年创新基金项目(05E7013); 国家重大专项子课题(G5800-08-ZS-WX)。黎旭, 硕士生, 主研领域: 知识发现。李国和, 教授。吴卫江, 副教授。洪云峰, 工程师。刘智渊, 总工程师。程远, 工程师。

1 决策规则挖掘

1.1 数据集形式化表示

定义1 客户数据集为 $L = (ID, A, VID)$ 。 $ID = \{id_1, id_2, \dots, id_{|ID|}\}$ 为客户集合 ($|ID|$ 为集合中元素的个数); $A = \{a_i \mid i = 1, 2, \dots, k\}$ 为客户属性集; $VID = \{V_{a_i} \mid i = 1, 2, \dots, k\}$, V_{a_i} 为客户属性 a_i 的值域, 表示客户 $id \in ID$ 在 a_i 上的投影, 即 $a_i: ID \rightarrow V_{a_i}$, $a_i(id) \in V_{a_i}$ 。

定义2 交易数据集为 $K = (T, ID \cup S, VS)$ 。其中 $T = \{t_1, t_2, \dots, t_{|T|}\}$ 为交易集合 ($|T|$ 为集合中元素的个数); $S = \{s_i \mid i = 1, 2, \dots, k\}$ 为商品集合; $ID \neq \Phi, S \neq \Phi, ID \cap S = \Phi$; $VS = \{V_{s_i} \mid i = 1, 2, \dots, k\}$, V_{s_i} 为商品 s_i 的交易量, 表示交易 $t \in T$ 在 s_i 上的投影, 即 $s_i: T \rightarrow V_{s_i}$, $s_i(t) \in V_{s_i}$ 。

定义3 客户交易数据集 $CK = (T, I, N)$ 。其中 $I = ID \cup A \cup S, N = VID \cup VS$ 。令 $C \subseteq I, D \subseteq I$ 并且 $C \cap D = \Phi, C \cup D = I$ 称 C 为条件属性集, D 为决策属性集。 C 和 D 也分别为条件因素集和决策因素集。

定义4 $CK' = (T, C \cup D, N)$ 为客户交易数据子集。其中 $C \subseteq I, D \subseteq I$ 并且 $C \cap D = \Phi$, 决策规则 $DecRule$ 为: $\{ \langle c, p_c \rangle \mid c \in C, p_c \in V_c, V_c \in N \} \Rightarrow \{ \langle d, p_d \rangle \mid d \in D, p_d \in V_d, V_d \in N \}$ (f) 其中 f 为不确定性因子:

$$f = \frac{|\{t \mid t \in T, f \in C, f(t) = v_c, \langle c, p_c \rangle \in DecRule\}|}{|\{t \mid t \in T, d \in D, d(t) = v_d, \langle d, p_d \rangle \in DecRule\}|} \quad (1)$$

且 $f \in [0, 1]$ 表示决策属性集依赖于条件属性集的程度, 或条件属性集的分类能力。决策树 $DecTree = \{Path \mid Path = DecRule\}$, 决策树路径为决策规则, 决策树为决策规则的集合。

定义5 客户交易数据集 CK 的划分 $DivT(T) = \{T_i \mid T_i \subseteq T\}$ 。其满足: 对于 $\forall T_1, T_2 \in DivT(T), T_1 \cap T_2 = \Phi, T = \bigcup_{\forall T_i \in DivT(T)} T_i$ 。

对于属性集 $B \subseteq I, T/B$ 为 CK 关于 B 的一个划分, 其满足: 对于 $\forall T_i \in T/B, \forall u, v \in T_i, \forall b \in B, b(u) = b(v)$, 称 $T_i \in T/B$ 为关于 B 的一个等价类, T/B 称为关于 B 的等价类族, $p_i = p(T_i/B) = |T_i|/|T|$ 为该等价类的概率, $P(T/B) = \{p(T_i/B) \mid \forall T_i \in T/B\}$ 为客户数据集关于属性集 B 的概率分布。

1.2 C4.5 算法

对于客户交易数据集 CK, T 关于 D 的期望信息:

$$Info(T, D) = - \sum_{\forall p_k \in P(T, D)} p_k \log_2(p_k) \quad (2)$$

对于属性 $a \in C, T$ 关于 a 的期望信息:

$$Info_a(T, D) = \sum_{\forall T' \in T/\{a\}} \frac{|T'|}{|T|} Info(T', D) \quad (3)$$

T 关于属性 a 的信息增益:

$$Gain(a) = Info(T, D) - Info_a(T, D) \quad (4)$$

信息增益刻画属性 a 的分类能力。 $Gain(a)$ 越大, a 的分类能力越强。

T 关于属性 a 的分裂信息:

$$SplitInfo_a(CK) = - \sum_{\forall p \in P(T, \{a\})} p \log_2(p) \quad (5)$$

分裂信息描述属性 a 划分 T 均匀程度。 $SplitInfo_a(CK)$ 越大, 在 a 上分布的交易 T 越均匀。

信息增益率:

$$GainRatio(a) = \frac{Gain(a)}{SplitInfo_a(CK)} \quad (6)$$

信息增益率为规范化的信息增益。

C4.5 算法是在 ID3 算法^[7] 基础上发展而来的, 其主要针对连续型条件属性优化二分, 并采用信息增益率进行分类能力评估, 选取信息增益率大的条件属性作为决策树分支节点。其主要过程描述为: 设条件属性集 $C = C_1 \cup C_2, C_1 \cap C_2 = \Phi, C_1$ 表示连续型属性集, C_2 表示离散属性集。设 $a \in C_1$, 对 a 的取值排序为 $Sort(a) = \langle a(t'_1), a(t'_2), \dots, a(t'_{|T|}) \rangle$, 且 $t'_i, t'_{i+1} \in T, i = 1, 2, \dots, |T| - 1$ 。断点集 $cuts(a) = \{cut_i \mid cut_i = [a(t'_i) + a(t'_{i+1})]/2, a(t'_i), a(t'_{i+1}) \in Sort(a)\}$ 。对连续型属性 a 定义信息增益率为:

$$GainRatio(a, cut) = \frac{Gain(a, cut)}{SplitInfo_a(CK, cut)} \quad (7)$$

其中, $cut \in cuts(a)$, 且 $\forall t' \in T$, 如果 $a(t') \leq cut$, 则 $a(t')$ 取值 0; 否则 $a(t')$ 取值 1, 即根据断点 cut 对属性值进行 0-1 归一化。 $GainRatio(a, cut)$ 为连续型属性 a 关于断点 cut 的信息增益率。

$$opt_cut(a) = \arg \max_{\forall cut \in cuts(a)} \{GainRatio(a, cut)\} \quad (8)$$

即获得属性 a 的最优断点 opt_cut 。根据连续型属性 a 和断点 opt_cut, T 可被划分为 T^0, T^1 , 即 $T/\{a\}_{opt_cut} = \{T^0, T^1\}$, 其中 $\forall t \in T^0, a(t) \leq opt_cut, \forall t \in T^1, a(t) > opt_cut$ 。

决策树的分支为:

$$\begin{aligned} \langle a, p_a \rangle &= \arg \max_{\forall b \in C} \{SplitInfo_b(CK, opt_cut(b)) \mid b \in C_1, \\ &\quad \langle b, p_{opt_cut(b)} \rangle \} \cup \{SplitInfo_b(CK) \mid b \in C_2, \\ &\quad \langle b, p(b) \rangle \} \end{aligned} \quad (9)$$

1.3 连续型属性多次二分的决策树生成算法

为了提高含有连续型属性的决策树精确性, 改进 C4.5 算法的不足^[8], 引入 $CN = \{n(a) \mid n(a) \in N, a \in C\}$ 表示 a 在决策树中最多可选为节点的次数集合。如果 $a \in C_1$, 则 $n(a) \geq 1$; 如果 $a \in C_2$, 则 $n(a) = 1$ 。对于离散型属性在决策树中只出现一次, 而连续型属性在决策树中可多次出现, 即多次对连续型属性二分。基于连续型属性多次二分的决策树生成算法如下:

```
DecTree(T, C1 ∪ C2 ∪ D, CN)
gr = 0 //最大信息增益率
If |T/D| = 1 or C1 ∪ C2 = Φ then return Φ //类别相同或没有条件属性
For each a ∈ C //任意属性
    if a ∈ C1 and n(a) ∈ CN and n(a) > 0 then //连续型属性
        grc = max_{∀ cut ∈ cuts(a)} {GainRatio(a, cut)}
    else //离散属性
        grc = GainRatio(a)
    Endif
    If gr < grc then //增益率大
        gr = grc
        b = a
    Endif
Endfor
If gr = 0 then return Φ //信息没增益
If b ∈ C1 then //连续型属性
    cut = opt_cut(b) //最优断点
    Ts = T/{b} cut //0-1 等价类
    Node = { [b, <= cut], [b, > cut] } //决策树节点
    cn(b) = cn(b) - 1 //二分次数
    if cn(b) = 0 then C1 = C1 - {b} //该属性不能再分
Else //离散型属性
```

```
Ts = T/{ b} //等价类
cn( b) =0 //只选择 1 次
Node = { [b , = b( b) ] | t ∈ Ti , Ti ∈ Ts} //决策树节点
C2 = C2 - { b} //该属性不能再分
Endif
Nodes = { Node} //决策树上级节点
For each Ti ∈ Ts
Nodesi = DecTree( Ti , C1 ∪ C2 ∪ D , CN)
Nodes = Nodes ∪ { Nodesi} //所有下级节点
Endfor
Retrun Nodes
```

2 消费行为因素分析系统的实现

2.1 系统架构

采用 Web 体系结构^[9-11] ,以 Oracle 为数据库服务器 ,JIS 为应用服务器 ,Web 浏览器为客户端。逻辑上为表现层(UI) 、业务逻辑层(BLL) 和数据访问层(DAL) 。如图 1 所示。

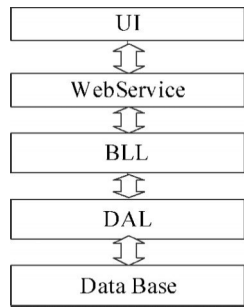


图 1 系统架构

其中 ,UI 层接受用户的输入并传给 BLL 层 ,接收 BLL 层返回的结果并显示于界面 ,该层通过 aspx 页面实现界面控件 ,在对应的 aspx. cs 页面实现控件事件; BLL 层实现具体的业务 ,可完成复杂的逻辑运算 ,实现决策树生成; DAL 层通过数据库访问类(DbHelperOra) 连接数据库 ,调用数据库中的存储过程 ,将存储过程执行结果传给 BLL 层处理。此外 ,将 BLL 层的方法包装成 Webservice ,供 UI 层调用 ,并方便其他系统与本系统的对接。

2.2 系统设计与实现

消费行为因素分析按流程分为用户输入、数据库操作、决策树生成、结果处理四大块(如图 2 所示) 。

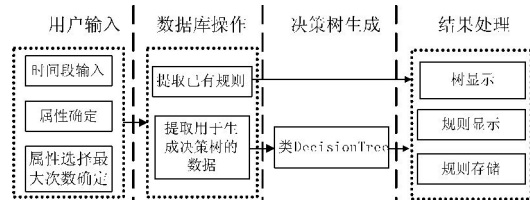


图 2 整体设计图

(1) 用户输入: 用户输入包括时间段输入、属性确定和属性选择最大次数。时间段输入缺省设置为所有时间 ,属性选择最大次数缺省设置为一次。条件属性和决策属性必须选定。该模块在 UI 层。

① 时间段输入: 采用 TextBox 控件绑定 My97DatePicker 日期控件 ,用户可直接在 TextBox 中输入日期 ,也可在 My97DatePicker 上勾选日期。My97DatePicker 自带日期格式纠

错功能。

② 属性确定: 用两个 ListBox 控件显示所有消费行为属性。左侧的 ListBox 为可用属性 ,右侧 ListBox 为选定属性。添加四个按钮 ,分别实现全选、单选、删除、全部删除的功能。

③ 属性选择最大次数确定: 用 TextBox 接收输入的数字 ,并建立方法 toIntByString() 对 TextBox 中的内容进行检测 ,确保其为数字并转换成 Int 类型。

时间和属性经过 BLL 层传入 DAL 层的数据库访问类 DbHelperOra ,属性选择最大次数传入决策树生成类 DecisionTree。

(2) 数据库操作: 包括提取已有规则和提取用于生成决策树的数据。该模块在 DAL 层 ,通过 DbHelperOra 类的 RunProcedure 方法调用 Oracle 数据库中的存储过程。

① 提取已有规则: 如果决策树已生成并保存于数据库中 ,则从数据库中提取规则。DbHelperOra 接收到输入后 ,通过 RunProcedure 方法执行存储过程 sp_getRules。该存储过程根据用户输入条件取出规则 ,返回结果并存入 DataSet。如果数据库中有相应规则 ,则 DataSet 不为空 ,进入结果处理中的规则显示; 否则提取用于生成决策树的数据。

② 提取用于生成决策树的数据: 如果决策树是首次生成 ,则提取用于生成决策树的数据。DbHelperOra 类通过 RunProcedure 方法调用存储过程 sp_selectData。该存储过程完成两个功能: 数据的预处理^[12] ,包括含缺值记录的剔除和数据格式的转换(如生日转换成年龄 ,由 DateTime 转换成 Number 类型) 以及部分连续型属性的离散化(如按照企业的需求对月收入进行离散化) ;另一功能是提取数据集 CK ,按照属性集 I 的取值进行分组 ,提取统计结果(而非每一条记录) 以减少参与构建决策树的记录数目。在存储过程中通过分组查询(group by) 和行数统计(count(*)) 实现。

(3) 决策树生成: 采用连续型属性多次二分的决策树生成算法生成决策树。该功能集成为 BLL 层的类 DecisionTree ,其方法如图 3 所示。

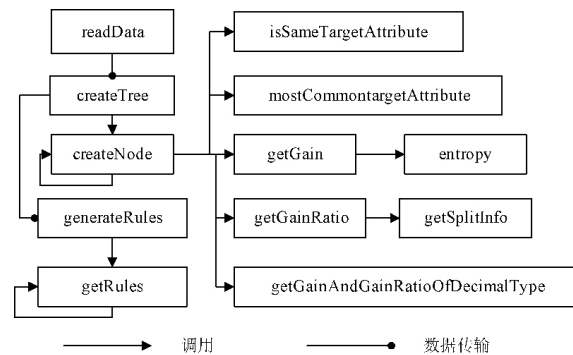


图 3 DecisionTree 类

① readData 方法接收 CK ,转换成样本集合 samples 并传给 createTree 方法。

② createTree 方法接收 samples ,把所有属性最大选择次数存入集合 CN 中。把 samples 和 CN 作为参数调用 createNode 方法 ,得到决策树的根节点并传给 generateRules 方法 ,或者传给树显示模块用于界面显示。

③ createNode 方法使用系统提供的 TreeNode 类实例化节点 ,递归创建节点生成决策树。该方法调用 isSameTargetAttribute 方法判断决策属性值是否相同。当所有条件属性都不可分割时 ,调用 mostCommonTargetAttribute 方法求最大可能的决策属性值; 当条件属性可分割时 ,如果是离散属性 ,则通过 getGain

方法计算信息增益,通过 `getGainRatio` 方法计算增益率。其中,`getGain` 方法调用 `entropy` 方法计算期望信息,`getGainRatio` 方法调用 `getSplitInfo` 方法计算分裂信息。如果是连续型属性,则通过 `getGainAndGainRatioOfDecimalType` 方法求最佳分割点并计算信息增益和增益率。

④ `generateRules` 方法接收决策树的根节点,然后遍历整棵决策树形成从决策树根节点到任一叶子节点的 if-then 规则。其通过 `getRules` 方法形成一条规则,递归调用 `getRules` 遍历决策树形成所有规则。对于连续型属性分裂 n 次,生成规则时对连续型属性的取值进行合并。

(4) 结果处理:包括树显示、规则显示和规则存储。

① 树显示:在 UI 层接收 `createTree` 方法产生的根节点,使用 `TreeView` 控件加载根节点。`TreeView` 控件会遍历显示完整的决策树。

② 规则显示:在 UI 层接收 `generateRules` 方法传来的规则,使用 `Table` 控件以列表的形式显示规则。

③ 规则存储:在 DAL 层接收 `generateRules` 方法产生的规则,把分析时间、属性及规则传入 `DbHelperOra` 类,并通过 `Run-Procedure` 方法调用存储过程 `sp_insertRules` 存储规则。

3 运行效果及分析

通过浏览器登录北京兆信消费行为分析系统,进入消费行为因素分析页面,显示消费者关联的所有数据表字段(如图 4 所示)。数据库中共有 516 799 条购买记录。



图 4 消费行为因素分析界面

在界面上填入分析时间段、属性选择最大次数并选择因素。点击分析按钮,得到决策树(部分子树如图 5、图 6 所示),

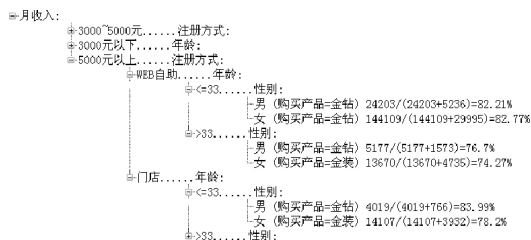


图 5 决策树(CN={1,1,1,1})



图 6 决策树(CN={1,2,1,1})

其中收入已预处理为离散化数据。决策树上每条路径对应一条决策规则。

从决策树上可获取如下信息:

(1) 从根节点到叶节点构成了消费行为的属性及其取值的联系,也就是决定消费行为的因素之间联系。如:月收入 > 5 000,注册方式为 Web 自助,年龄 ≤ 33,性别为男性的消费者购买金钻奶粉有 24 203 人,购买其他有 5 236 人,且购买金钻奶粉者占 82.21%。

(2) 节点位置反映属性(因素)的重要性。越靠近根节点,属性重要性越大。如在选择购买金钻或金装奶粉时,消费者月收入是最重要的影响因素,而消费者性别的重要性最低。

(3) 连续型属性最大可分割次数影响决策规则的精度。随着连续型属性二分数增加,决策规则确定性提高。如图 6 所示,年龄最多进行 2 次分割,可得决策规则:月收入 5 000 元以上,注册方式为门店,年龄 ≤ 25 岁,购买金钻奶粉有 3 082 人,购买其他有 0 人,且购买金钻奶粉者占 100%。月收入 5 000 元以上,注册方式为门店,年龄在 25 ~ 33 岁,性别为男性的消费者购买金钻奶粉有 3 338 人,购买其他有 766 人,且购买金钻奶粉者占 81.34%。月收入 5 000 元以上,注册方式为门店,年龄在 25 ~ 33 岁,性别为女性的消费者购买金装奶粉有 14 107 人,购买其他有 1 531 人,且购买金装奶粉者占 90.21%。综合分析,月收入 5 000 元以上,门店注册,年龄 ≤ 33 岁,规则精度为 $(3\ 028 + 3\ 338 + 14\ 107) / (3\ 082 + 0 + 3\ 338 + 766 + 14\ 107 + 1\ 531) = 89.70\%$ 。图 5 中月收入 5 000 元以上,门店注册,年龄 ≤ 33 岁,规则精度为 $(4\ 019 + 14\ 107) / (4\ 019 + 766 + 14\ 107 + 3\ 932) = 79.42\%$ 。图 6 规则的精度高于图 5。

4 结 语

针对北京兆信信息技术有限公司消费行为分析系统的需要,在 Web 三层体系结构基础上,以 Oracle 为数据库,采用 .Net 和数据库存储过程编程实现了基于决策规则的消费行为建模,分析消费者因素(属性)与购买产品,或其他因素(决策属性)之间的关系,为企业的生产、促销提供辅助决策依据。在决策规则生成算法上,C4.5 决策树算法能够依据连续型属性数据集进行一次二分离散化,通过信息增益率选定分类能力强的属性,并按属性的各个离散值形成决策树的分支节点,直到所有记录都是同一类别或没有可再分支的属性,前者形成决策规则精确度为 100%,后者为小于 100%。对于连续型属性而言,记录类别可能分布不均匀,采用 C4.5 算法进行一次二分离散化(寻找一个分割点),决策规则不确定性往往比较高。通过改进 C4.5 算法,实现基于连续型条件属性多次最优二分的决策生成算法,也就是增加连续型属性最大可分割次数 N ,形成决策树中连续型属性在决策树任一分支上出现的次数均不超过 N ,而且多数分支的确定性有所增强,即使确定性有所降低也是由于该连续型属性的记录类别分布不均匀所致,但另一分支确定性则肯定增强。总之,通过连续型属性多次最优二分,提高了决策规则的准确率,在消费行为因素分析中取得了令人满意的效果。由于连续型属性多次二分算法以记录类别为启发信息,通过“枚举+测试”获取最优二分分割点,算法计算量大,后续研究拟采用增量算法进行分割点的确定,或先进行离散化后再进行决策树生成的方法,以提高决策树生成的性能。

(下转第 211 页)

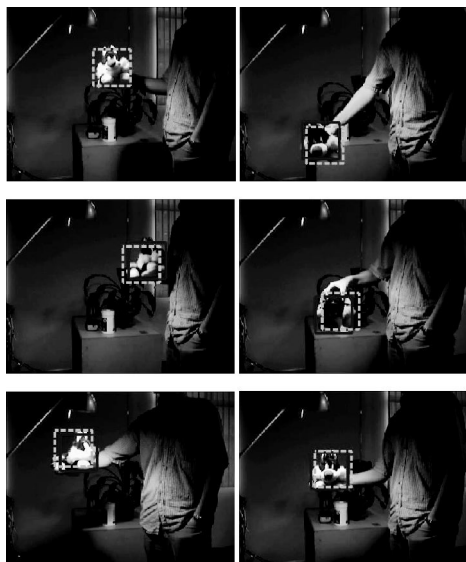


图8 对 Sylvester 视频序列的跟踪结果

通过对上述多个实验结果的分析,可以充分地证明本文算法相比于CT算法的优势。CT算法对压缩后的数据进行分类,在提高速度的同时必然在一定程度上降低分类的准确性。本文以CT算法为基础,针对CT容易发生目标跟丢的情况,将跟踪与检测相融合。在跟踪失败或发生漂移的情况下,通过检测模块来进行目标的重定位,虽然检测部分本质上采用的是模版匹配,但是在跟踪的过程中,本文的匹配模版是动态更新的。也就是说,即使被跟踪目标在跟踪的过程中发生了形变、姿态、光照等一系列的线性变化,检测模块依然能够再次定位到目标的准确位置。

4 结 语

针对文献[5]中CT算法的不足,本文提出了将跟踪与检测相融合的长时间实时目标跟踪算法,通过引入级联搜索的策略,在保证跟踪效果的同时可以达到25 fps以上的跟踪速度。对不同视频序列中目标物体跟踪的结果表明,本文所提出的算法在目标物体移动速度较快和目标纹理以及光照变化时,跟踪的准确性和稳定性都得到了较大的提升,有效地减少了漂移发生的次数。同时,在长时间跟踪的过程中,本文提出的算法在目标发生部分或完全遮挡、离开摄像头视野、瞬间移动以及外观、姿态等一系列变化的情况下,也都能够再次准确检测到目标物体,并继续跟踪。但是当目标尺度发生大的变化时,本文所提的方法还不能达到很好的效果,所以,如何解决尺度问题将是下一步的研究重点。

参 考 文 献

- [1] 田峥,徐成,杨志邦,等.一种基于视觉注意机制的改进粒子滤波跟踪算法[J].计算机应用与软件,2011,28(11):85-88.
- [2] Du Yong Kim, Moongu Jeon. Spatio-temporal auxiliary particle filtering with l_1 -norm-based appearance model learning for robust visual tracking[J]. IEEE Transaction on Image Processing, 2013, 22(2): 511-522.
- [3] Kalal Z, Matas J, Mikolajczyk K. Online learning of robust object detectors during unstable tracking[C]//Proceedings of the 12th IEEE International Conference on Computer Vision Workshops. Los Alamitos: IEEE

Computer Society Press 2009: 1417-1424.

- [4] Nejhum S M S, Ho J, Yang M H. Visual tracking with histograms and articulating blocks[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press 2008: 1-8.
- [5] Zhang K H, Zhang L, Yang M H. Real-time compressive tracking[C]//Proceedings of the 12th European Conference on Computer Vision. London: Springer-Verlag 2012: 866-879.
- [6] Donoho D L. Compressed sensing[J]. IEEE Transactions on Information Theory 2006, 52(4): 1289-1306.
- [7] Papadourakis V, Argyros A. Multiple objects tracking in the presence of long-term occlusions[J]. Computer Vision and Image Understanding, 2010, 114(7): 835-846.
- [8] Zdenek Kalal, Krystian Mikolajczyk, Jiri Matas. Tracking-Learning-Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence 2012, 34(7): 1409-1422.
- [9] Tomas Blumensath. Compressed sensing with nonlinear observations and related nonlinear optimization problems[J]. IEEE Transactions on Information Theory 2013, 59(6): 3466-3474.
- [10] 朱秋平, 颜佳, 张虎, 等. 基于压缩感知的多特征实时跟踪[J]. 光学精密工程 2013, 21(2): 437-444.
- [11] Jae Chern Yoo, Tae Hee Han. Fast Normalized Cross-Correlation[J]. Circuits Systems and Signal Processing 2009, 28(6): 819-843.
- [12] 刘德春. 对偶算法的有效改进[J]. 中国图象图形学报, 2006, 11(4): 465-468.
- [13] 徐德江, 史泽林, 罗海波. 零均值归一化互相关跟踪算法特性研究[J]. 红外与激光工程 2010, 39(S): 485-489.

(上接第188页)

参 考 文 献

- [1] 张亚明, 刘海鹏, 赵培卿. 中美网络消费者消费行为分析[J]. 企业经济 2012(6): 123-127.
- [2] 陈晓芬. 温州大学生服装消费行为分析[J]. 中国商贸 2011(34): 246-247.
- [3] 韩占兵. 我国城镇消费者有机农产品消费行为分析[J]. 商业研究 2013(8): 184-190.
- [4] 刘蓉, 陈晓红. 基于数据挖掘的移动通信客户消费行为分析[J]. 计算机应用与软件 2006, 23(2): 60-62.
- [5] 朱建平, 来升强. 时态数据挖掘在手机用户消费行为分析中的应用[J]. 数理统计与管理 2008, 27(1): 42-53.
- [6] 张革伙, 欧阳浩男, 徐琪. 决策树在基于消费者外表的服装营销中的应用[J]. 计算机应用 2010, 30(2): 1919-1929.
- [7] 范明, 孟小峰. 数据挖掘概念与技术[M]. 北京: 机械工业出版社 2012.
- [8] 姚亚夫, 荆留涛. 决策树C4.5连续属性分割阈值算法改进及其应用[J]. 中南大学学报: 自然科学版 2011, 42(12): 3772-3776.
- [9] 代乾坤, 仲梁维, 屈年凯. 基于.NET架构的产品结构及物料管理系统[J]. 计算机系统应用 2012, 21(8): 27-31.
- [10] 肖文. .net平台开发中表现层的常用设计模式[J]. 计算机应用与软件 2012, 29(5): 231-233.
- [11] 呼和, 张钦, 吕克强. SynchroFLOW工作流在.NET架构系统下的典型应用[J]. 计算机应用 2011, 31(1): 209-211.
- [12] 巩固, 张虹. 决策树算法中属性缺失值的研究[J]. 计算机应用与软件 2008, 25(9): 242-244.