

ID3 算法在包装机器人自动分拣中的应用

黄国栋^{1,2}, 陈杰²

(1. 宿迁学院, 宿迁 223800; 2. 南京理工大学, 南京 210094)

摘要: 研究了包装机器人(机械手)多属性识别问题,结合数据挖掘技术的 ID3 算法,建立了一种适合混流包装线机器人进行多属性产品自动分拣装箱的方法。讨论了构建机器人多属性分拣系统所面临的问题,以及 ID3 算法的实现。研究表明,该算法能较好实现机器人在混流包装线上的分拣功能。

关键词: 数据挖掘; 现代物流技术; 自动识别; 机器人; ID3 算法

中图分类号: TB486+.03 **文献标识码:** A **文章编号:** 1001-3563(2012)11-0084-04

Application of ID3 Algorithm in Automatic Sorting of Packaging Robot

HUANG Guo-dong^{1,2}, CHEN Jie²

(1. Suqian College, Suqian 223800, China; 2. Nanjing University of Science & Technology, Nanjing 210094, China)

Abstract: Multi-attribute identification of packaging robot or manipulator was studied. A method of automatic multi-attribute product sorting by robot in mixed packaging line was proposed, which combined with ID3 algorithm in data mining technology. Problems faced when building multi-attribute sorting system and using ID3 algorithm were discussed. It was proved that ID3 algorithm can easily realize sorting function of packaging robot in mixed packaging line.

Key words: data mining; modern logistics technology; automatic identification; robot; ID3 algorithm

机器人技术运用于物流分拣作业,是提高生产效率的一种有效手段,许多自动化包装线中大量使用机器人辅助完成自动分拣^[1-2]、包装等相关作业。目前,包装机器人(机械手)主要应用在重物转移,产品装箱、捆扎、封边,识别检测、挑选,有害气、液体灌装,食品包装等方面。随着企业在各种推拉式生产系统^[3-4]中广泛采用混流生产方式,要求自动化机器人也能参与到混流过程中,尤其是包装机器人不仅要能准确地将产品装箱,还必须能根据产品的属性自动识别多种产品并准确装入对应的包装容器。对机器人自动识别能力的研究是自动化领域、计算机领域以及生产控制领域研究的焦点之一。文中主要集中讨论包装机器人实现多属性产品自动分拣的可行方法。

货期、客户定制信息等),为了完成特定的包装作业,传统的包装机器人一般分为袋装机器人、装箱机器人、堆码机器人、灌装机器人、包装输送机器人和识别检测机器人。随着机器人硬件技术的升级,机器人的自动化水平越来越高,能够处理的作业任务难度也随之提高。

包装机器人的多属性分拣过程实际上属于机器人的多目标(属性)分类(聚类)过程,见图 1。多属性产

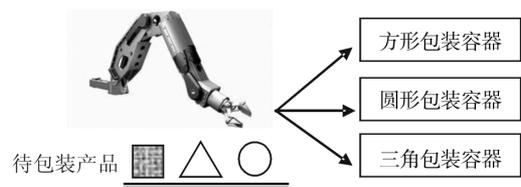


图 1 多属性分拣过程

Fig. 1 Multi-attribute sorting process

1 多属性包装机器人自动分拣功能分析

根据产品属性不同(如质量、形状、品质、尺寸、交

收稿日期: 2012-03-28

作者简介: 黄国栋(1982-),男,四川岳池人,宿迁学院讲师,主要研究方向为制造系统优化、工业工程。

品识别指的是不同产品按属性实现聚类,同时相同产品按照不同属性进行分类,即要考虑到相同产品属性是可能不同的。所以在构建多属性包装机器人自动分拣功能时,需考虑以下问题:

1) 包装对象信息特征。包装对象(产品)的信息特征一般包括质量、尺寸大小、几何形状,甚至包括颜色、温度等。随着传感器技术的发展,可以通过称重传感器、颜色传感器、色温传感器,较好解决机器对于重量、温度、颜色的信息采集。而且最近几年,随着图形图像技术的发展,机器对于尺寸和几何形状的信息提取能力也获得了一定程度的提高。因此,多数工业用机器人通过加装各种传感器来实现信息的采集。在实际生产现场,考虑到现场环境因素影响较大,信息一次采集成功率并不高,尤其在速度较快的传输带上,一次成功识别率无法达到工业生产要求,所以初始信息采集时,主要是将一维码或二维码贴附在产品或托盘上来实现高速信息采集,这种方式不仅技术成熟,而且成本低廉。

2) 机器学习与识别。信息采集只是实现自动分拣功能的初始环节,在确保信息采集正确的基础上,需对大量数据进行分析。模式识别的一个重要目标就是从大量看似杂乱的数据中,建立一些较为有规律且便于区分的模式,然后通过对此种模式进行学习,建立机器识别规则。这一过程中,如何建立易于辨识模式的方法是一个重点,其次建立的规则优劣也决定着机器识别程度的高低。

3) 包装机器人的系统架构。包装机器人实现多属性自动分拣功能要涉及机器人控制系统、传感器、条形码、数据库 4 个模块,再结合识别算法建立识别规则进行机器学习、模式识别,从而向机器人计算机系统发出指令,实现自动分拣与分类包装。系统架构见图 2。

2 分类与 ID3 算法

2.1 分类方法

分类实质上就是让机器通过学习获得一个目标函数 ψ ,将需要分类的对象的属性集 χ 映射到一个定义为 Π 的类中。分类过程中先提供一组训练属性集合 $\tau \in \chi$,进行归纳,建立分类模型 $\Pi(\tau)$,两者关系可用元组 (τ, Π) 表示。再通过一组测试属性集合 $\varphi \in \chi$ 对 Π 进行测试,建立元组 (τ, X) ,若测试结果 $X = \Pi$,则说明分类成功。本质上,这是一种归纳推理方法,归纳推理模式符合人类学习过程,从对象一部分或整体的特定观察中获得一个完备且正确的描述,通过比较、总结、概括而得出规律性认知,然后通过推理实现对特定目标的分类识别。分类识别方法包括决策树分类法、基于规则的分类法、神经网络、支持向量机、朴素贝叶斯分类法等。

ID3 算法属于决策树分类法,由 Quinlan^[5-6] 提出,并在 1983 年和 1986 年进行了总结和简化,成为决策树学习算法的典型。ID3 算法的核心是:在决策树的各级节点上,使用信息增益方法作为属性的选择标准,来帮助确定生成每个节点时所应采用的合适属性。这样就可以选择具有最高信息增益的属性作为当前节点的测试属性,以便使用该属性所划分获得的训练样本子集进行分类所需信息量最小。ID3 算法的理论清晰,学习能力较强,尤其可以处理非数值型的属性,在处理不同类型多属性的机器学习及模式识别研究中得到关注^[7-10]。

2.2 ID3 原理

设 θ 为 N 维向量空间, $\Gamma \in \theta$ 为其元素,又称为样例。将 θ 分成 2 个样例集合,分别称为 $\Gamma_1 \in \theta$ 正例集合及 $\Gamma_2 \in \theta$ 反例集合,其大小分别用 $m(\Gamma_1), n(\Gamma_2)$ 表示,简写为 m, n 。ID3 有如下 2 种假设:(1)对任意 $\Gamma \in \theta$ 正确分类概率为 p ,则满足 $p(\Gamma_1) = p(\Gamma_2)$;(2)决策树对某样例正确分类用信息量 $H(m, n)$ 度量。

$$H(m, n) = -\frac{m}{m+n} \log_2 \frac{m}{m+n} - \frac{n}{m+n} \log_2 \frac{n}{m+n}$$

若某属性 A 具有 V 个值 $\{V_1, V_2, \dots, V_V\}$,可将 θ 分成 V 个子集 $\{\theta_1, \theta_2, \dots, \theta_V\}$ 。若 θ_i 中含有 m_i 个正例和 n_i 个反例,则子集 θ_i 的信息量^[6]为 $H(m_i, n_i)$ 。

从而 A 的信息熵:

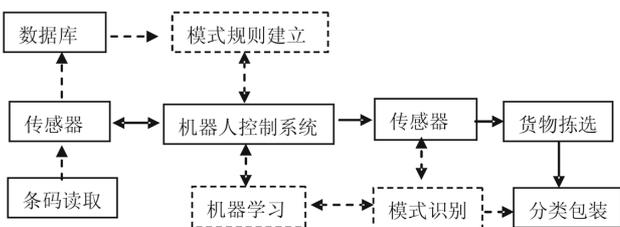


图 2 包装机器人的系统架构

Fig. 2 System architecture of packaging robot

$$E(A) = \sum_{i=1}^v \frac{m_i + n_i}{m + n} H(m_i, n_i)$$

A 的信息增益 $g(A) = H_s(m, n) - E(A)$, $H_s(m, n)$ 为所研究目标变量的总信息量。其中, 最终 ID3 算法选择 $\max(g(A))$ 的属性为决策树根节点, 再从根节点采用上述方式递归得到各分支节点 B_1, B_2, \dots, B_k , 从而达到分类的目的。

2.3 ID3 算法编程实现流程图

ID3 算法编程实现的流程图见图 3。

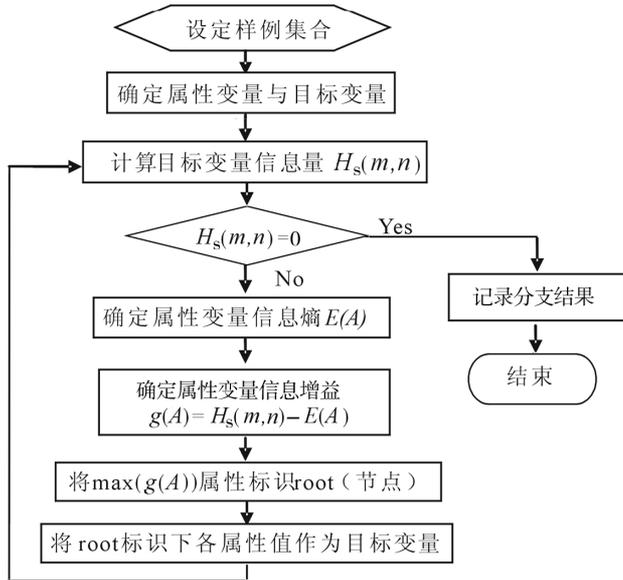


图 3 ID3 流程

Fig. 3 ID3 flow chart

3 基于 ID3 算法的分拣规则确定

根据某企业实际属性数据样例作为训练集合(见表 1), 说明如何确定包装机器人自动分拣装箱规则。

表 1 装箱训练集合

Tab.1 Training set of encasement

RFID	货物	瓦楞	产品	RFID	货物	瓦楞	产品
标签	质量/g	纸箱	等级	标签	质量/g	纸箱	等级
YES	52	双层	II	NO	20	单层	II
NO	42	双层	I	YES	73	双层	II
NO	61	双层	II	NO	37	单层	III
YES	30	单层	II	YES	33	双层	I
YES	26	双层	I	YES	48	单层	III
NO	29	单层	II				

首先确定分类根节点, 由于目标为选定箱型, 所以目标变量为瓦楞纸箱。

$$H_s(\text{纸箱}) = H(\text{单层}(6), \text{双层}(6)) = -\frac{6}{12} \log_2 \frac{6}{12} - \frac{6}{12} \log_2 \frac{6}{12} = 1$$

由于瓦楞纸箱有一个极限承载量, 为了防止搬运过程中造成货物损坏, 企业规定 40 kg 以上的货物要用承载力更强的双层瓦楞纸箱进行包装。以瓦楞纸箱为目标变量的各属性信息计算结果见表 2。

表 2 以瓦楞纸箱为目标变量的属性信息表

Tab.2 Attribute information using corrugated

case as target variable

属性变量	属性值	产品等级 II		E(A)	g(A)	max(g(A))	Hs
		m单层	n双层				
RFID	NO	4	2	0.918	0.082		
标签	YES	2	4				
货物	40 kg 以下	4	2	0.918	0.082		
质量	40 kg 以上	2	4				
产品等级	I	0	3	0.5	0.5	"root"	0
	II	3	3				1
	III	3	0				0

可知产品等级 = {I, II, III} 为分支节点 (root 节点), 所以目标变量分别设定为 I, II, III。 $H_s(I) = 0$, 产品等级 I 无分支, 选择双层瓦楞纸。 $H_s(III) = 0$, 产品等级 III 无分支, 选择单层瓦楞纸。 $H_s(II) = 1$, 产品等级 II 可继续分支。以产品等级 II 为目标变量的各属性信息计算结果见表 3。

表 3 以产品等级 II 为目标变量的属性信息表

Tab.3 Attribute information using product

grade II as target variable

属性变量	属性值	产品等级 II		E(A)	g(A)	max(g(A))	Hs
		m单层	n双层				
RFID	NO	2	1	0.918	0.082		
标签	YES	1	2				
货物	40 kg 以下	3	0	0	1	"root"	0
质量	40 kg 以上	0	3				

所以 40 kg 以下无分支, 选择单层瓦楞纸; 40 kg 以上均无分支, 选择双层瓦楞纸。

最终形成决策树规则, 见图 4。包装机器人将根据产品等级属性将其分别装入不同型号的瓦楞纸包装箱中, 其识别规则为: 当机器人夹取产品等级为 I 的产品时, 装入双层瓦楞纸包装箱, 等级为 III 的产品装入单层瓦楞包装箱; 等级为 II 的产品需根据质量进一步判别, 超过 40 kg 的装入双层瓦楞包装箱, 不到 40 kg 的装入单层瓦楞包装箱中。而包装箱中是否嵌

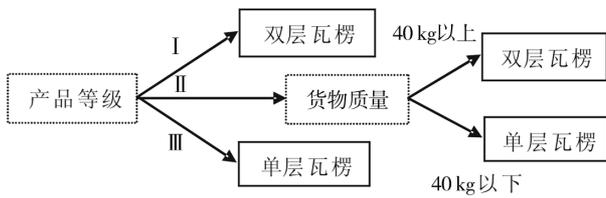


图 4 规则决策树
Fig. 4 Decision tree of rules

入 RFID 标签,不作为自动分拣的依据。

4 结论

该自动识别分拣系统中采用 C/C++ 编写相应 ID3 算法代码, ID3 算法建立的决策树判别具有很好的两元性,这种判定机制可以较好转换成数字逻辑系统,可进一步开发成功能芯片或功能卡嵌入到机器人硬件中,使其具有很好的工业应用价值。在实践过程中,所使用的数据样例可以根据企业需求预先定制,然后借助 ID3 算法构建识别规则,从而增加包装机器人对特定情况的识别处理功能。

参考文献:

[1] 王保升,汪木兰. 基于货单的多品种产品自动分拣系统[J]. 包装工程,2010,31(19):109-112.
WANG Bao-sheng, WANG Mu-lan. Automatic Sorting System for Multi-variety Product Based on Manifest[J]. Packaging Engineering, 2010, 31(19):109-112.

[2] 晏祖根,王立权,孙智慧,等. 面向食品生产的高速自动分拣系统的研究[J]. 包装工程,2009,30(7):16-18.
YAN Zu-gen, WANG Li-quan, SUN Zhi-hui, et al. Study

of High-speed Auto-sorting System for Food Production [J]. Packaging Engineering, 2009, 30(7):16-18.

[3] 刘琼,刘炜琪,张超勇. 基于 GA-PSO 的多目标混流装配线排序研究[J]. 华中科技大学学报(自然科学版),2011(10):1-5.
LIU Qiong, LIU Wei-qi, ZHANG Chao-yong. Hybrid GA-PSO Algorithm for Sequencing Multi Objective Mixed-model Assembly Lines[J]. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2011(10):1-5.

[4] 黄国栋,陈杰. CONWIP 拉式机制在自动包装线中的应用设计[J]. 包装工程,2011,32(11):58-68.
HUANG Guo-dong, CHEN Jie. Design of CONWIP Pull Mechanism Implementation in Automatic Packaging Line[J]. Packaging Engineering, 2011, 32(11):58-68.

[5] QUINLAN J R. Induction of Decision Tree[J]. Mach Learn, 1986, 1(1):81-106.

[6] SHANNON C E. A Mathematical Theory of Communication[J]. Bell System Technical Journal, 1948, 27:379-423.

[7] OSEIBRYSON K. Post-Pruning In Decision Tree Induction Using Multiple Performance Measures[J]. Computers & Operations Research, 2007, 11(34):3331-3345.

[8] LOMAX S, VADERA S. An Empirical Comparison of Cost-Sensitive Decision Tree Induction Algorithms[J]. Expert Systems the Journal of Knowledge Engineering, 2011, 28(3):227-268.

[9] ATRAMENTOV A, LEIVA H, HONAVAR V. Learning Decision Trees from Multi-Relational Data[C]//Proceedings of the 13th International Conference on Inductive Logic Programming, 2003.

[10] SELVAMANI B, KHEMANI D. Decision Tree Induction with CBR[J]. Pattern Recognition and Machine Intelligence, 2005, 3776:786-791.

(上接第 65 页)

[4] KENT Wahlen. Sheet Material Export Packaging[Z]. NEFAB, 2003.

[5] DING Yi, XIANG Ying. The Transient Response Analysis for the Damped Free Vibration of Plywood Box[J]. Advanced Material Research, 2011, 338:513-517.

[6] 王新荣. 有限元基础及 ANSYS 应用[M]. 北京:科学出版社,2008.
WANG Xin-rong. The Finite Element-based and ANSYS Application[M]. Beijing: Science Press, 2008.

[7] 邓凡平. ANSYS10.0 有限元分析自学手册[K]. 北京:人民邮电出版社,2007.
DENG Fan-ping. ANSYS10.0 Self-Study Manual[M]. Beijing: People Post Press, 2007.

[8] 洪庆章. ANSYS 教学范例[M]. 北京:中国铁道出版社,2002.
HONG Qing-zhang. ANSYS Teaching Example [M]. Beijing: China Railway Press, 2002.