

草图理解技术研究进展^{*}

周若鸿 孙正兴 张莉莎 徐晓刚

(南京大学计算机软件新技术国家重点实验室 南京210093)

摘要 本文概述了草图理解技术(包括草图识别和语义理解)的现状。首先,草图识别包括预处理和图形识别,根据各种图形处理机制,介绍了四种图形识别方法:基于笔划的、基于图元的、基于几何特征的、基于组合图形的。其次,深入分析了几个语义理解方法,一般包括三个环节:语义获取关系分析和高层应用。最后,本文提出了几何模型、用户模型和领域模型。讨论了草图理解中的主要问题并给出初步解决方案。

关键词 草图理解,草图识别,语义理解

Sketch Understanding: Review and Research

ZHOU Rou-Hong SUN Zheng-Xing ZHANG Li-Sha XU Xiao-Gang

(State Key Laboratory for Novel Software, Nanjing University, Nanjing 210093)

Abstract This paper summarizes the state of the art of sketch understanding in aspects of sketch recognition and semantic understanding. Firstly, sketch recognition is divided into preprocessing and graphics recognition, and according to various graphics processing mechanisms, a few graphics recognition methods based on strokes, primitive shapes, geometric features or composite shapes are introduced. Secondly, some typical methods for semantic understanding are analyzed in detail by classifying them into three classes: semantics acquisition, relation analysis and applications. Finally, geometry model, user model and domain model are proposed in this paper, we discuss the key issues in sketch understanding and give preliminary solutions.

Keywords Sketch understanding, Sketch recognition, Semantic understanding

1 引言

图形用户界面技术使得计算机应用的普及成为现实,但随着计算机及网络技术的不断发展及其向人类工作和生活的不断渗透,人们对用户界面的要求也越来越高,尤其需要实现以计算机系统为中心向以用户为中心的交互模式转变,自然化、人性化、个性化、智能化成为这一领域的主导。笔式交互(Pen-based Human Computer Interaction)^[1,2]应运而生,它以模拟现实的纸和笔为目标,具有易用、自然、高效和利于创造性思维等特点,成为感知用户界面(Perceptual User Interface)^[3]中不可缺少的部分,引起了学术界和产业界越来越多的兴趣。但是,由于笔式交互具有输入随意性和语义模糊性,尤其是手绘草图(Sketch)没有固定结构和模板,计算机的草图理解能力成为笔式交互走向实用化的主要瓶颈,因而成为用户界面领域的一个十分重要的研究课题。

关于草图理解技术,学术界目前还没有明确的定义。本文认为,草图理解技术是一种以笔式交互为信息获取手段,以捕捉和理解用户的输入意图为目的,融合用户交互技术、图形图像技术及视觉形象思维技术,并面向人类思维过程的技术。由于草图理解技术以视觉形象表达和传递概念,易于理解和记忆,且自然、便捷、具有能够表达不完备模糊概念的能力,在计算机辅助设计、计算机动画与虚拟现实、可视化对象建模、计算机支持的协同工作、网络及多媒体应用、远程教学及数字化办公和教育等诸多领域都有广泛的应用需求,尤其适合于设

计早期的概念设计等创造性活动^[4~8]。同时,草图理解技术又是一个极具挑战性的课题,它需要根据人对形状的视觉认知规则、思维习惯和信息模型对用户的输入草图及其过程进行推理和猜测,为用户勾画思路、深化思维和萌生新意提供支持。与手写字符相比,手绘草图的笔划数目和笔划顺序具有更大的自由度,其图形对象的构成比字符复杂,语义也比字符更具有模糊性和不确定性。

20世纪90年代以来,随着小型化和无线化设备的出现和发展,尤其是“数字墨水(Digital Ink/Electronic Ink)”等显示技术和平板计算机(Tablet Computer)的出现^[9,10],笔式交互技术引起工业界和学术界广泛重视,草图理解技术也进入新的发展阶段,成为诸如 SIGCHI、IFIP TC13等 HCI 学术团体和学术会议的重要研究内容。2002年春天 AAAI 召开了“Sketch Understanding”会议,反映了草图理解领域的最新进展。国内研究机构,如中科院软件所戴国忠教授项目组^[11~13]、微软亚洲研究院^[14]及南京大学计算机软件新技术国家重点实验室孙正兴博士项目组^[7,8,15~24]也对此开展了相关研究,取得了一系列的成果。

作为笔式交互技术的核心内容,草图理解的困难是如何根据用户的笔输入来理解其意图,这主要包括草图识别和语义理解两个方面(如图1所示)。其中,草图识别指的是将通过笔式交互获取的模糊的草图表达映射至精确的图形表达,它更侧重于图形形状构成,而语义理解则是将用户输入的草图图形映射为某个领域内具有特定含义的对象,它相对更注重

^{*} 本文得到国家自然科学基金项目(编号:69903006、60373065)资助。周若鸿 硕士生,研究方向为智能化人机界面;孙正兴 副教授,博士后,研究方向为智能用户接口、多媒体挖掘和图形图像技术;张莉莎 硕士生,研究方向为智能用户接口;徐晓刚 硕士,研究方向为智能用户接口。

于草图图形所具有的含义或草图图形对象间的关系。由于用户意图通常与其应用领域和目标密切相关,而且草图识别和语义理解可以看成是用户笔划输入在不同应用层面或从应用视角的映射,因此,它们的界限不是绝对的。此外,作为一种输入方式,草图理解还存在着二维和三维的区分,但由于草图理解更多的是关注从模糊草图图形向用户意图的转化,因此,本文将侧重于二维草图理解技术。本文的第2、3部分将分别介绍草图识别和语义理解的技术现状(表1概括了本文所涉及的草图理解技术及相关研究成果),第4部分探讨草图理解技术中的关键问题,并提出初步的解决思路。

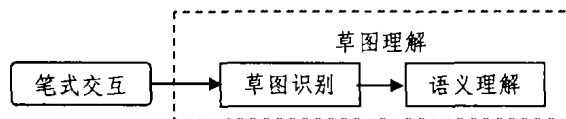


图1 草图理解系统结构图

表1 草图理解技术研究现状总表

主要问题	子问题	技术	成果
草图识别	预处理	消除冗余点及噪声	[25][26]
	图形识别	基于笔划表示的方法	[18][27]~[31]
		基于图元表示的方法	[15][17][25][32]~[35]
		基于几何特征的方法	[36][37]
		基于组合图形的方法	[22][36]
语义理解	语义获取	上下文信息利用	[36]
		多通道语义获取	[38]
		多领域描述	[39]
	关系分析	关系的描述及约束	[40][41]~[44]
	高层应用	与具体领域相关	[31][36][41][45]~[52]

2 草图识别

草图识别是注重于草图图形构成的草图理解,它需要从

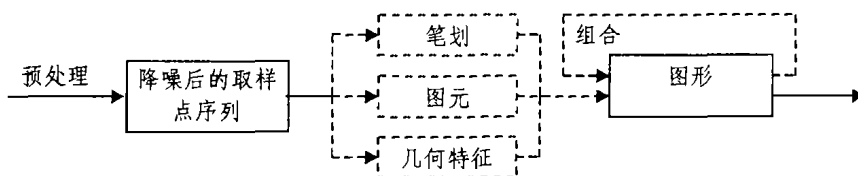


图2 草图识别技术现状框架图

2.2 基于图元表示的方法

同一图形存在不同的绘制方式,基于笔划表示的识别方法往往要求用户以固定模式的笔划绘制图形。基于图元表示的方法则可以解除这种限制,它将图形表示为直线、弧、曲线等图元在一定空间关系下的组合。基于图元的方法通常包含分割、拟合、规整、识别等步骤。

•分割 早期的草图理解系统直接将笔划的曲率作为分割依据^[32~34],但是用户勾勒草图时往往比较随意,使得一些有意或无意的拐角常常难以区分。基于当人们勾画有意拐角时速度明显变慢的认识,Sezgin^[35],Calhoun等^[25]都提出结合绘制速度和曲线转角两个因素进行拐角判断的方法,取得了较好的分割效果。

•拟合、规整 拟合则常常是设定原曲线和拟合后图元的距离公式,当相似度在容忍度范围内时接受^[17,35]。规整假定用户输入的是规整图形,如平行线规整^[35]、多边形及椭圆规

用户勾画草图的图形构造及表现中挖掘图形约束,将自由不规则的草图图形映射成规则精确的几何图形,其难易程度与加给绘制的约束直接相关。包括预处理与图形识别两部分工作。其中,预处理的目的是消除因输入条件的限制和输入习惯的差异而存在于笔划中的噪声,如重复取样点^[25]、无意识的勾角^[26]等。而图形识别则将预处理后的草图归到相应的图形类别中,是草图识别的主要工作。目前主要存在四种方式的图形识别方法:基于笔划表示的方法,基于直线、弧、曲线等图元表示的方法,基于几何特征的图形识别方法和基于组合图形的方法(如图2所示)。

2.1 基于笔划表示的方法

通常情况下,草图理解系统通过笔式交互方式获取的草图表现为若干由用户在落笔与提笔间的取样点序列构成的笔划。笔划是图形重要的构成信息。早期的系统视笔划为符号,赋予它特定的含义,将笔划的识别等同于图形的识别。随着人们对用户界面人性化要求的提高,这种助记符式的方法逐渐隐退,但因其识别简单、有效,一些系统则在单笔划识别的基础上添加了简单的空间位置信息作为图形识别的基础,或用来支持较为简单的 gesture (gesture 在草图理解系统中特指用来辅助编辑的笔划)识别。

单笔划识别^[18,27~29]中,较为经典的是 Rubine^[27]方法。它使用初始角的正余弦值、闭包盒对角线的长度及倾斜度、起点与终点的距离等11个几何特征和笔划的最大速度值、起点到终点的时间值等2个动态特征组成的13维向量表示笔划。通过训练获取各分量的权重,而后,根据一个线性评估函数来区分不同的手写符号。由于 Rubine 采用统计方法对 gesture 进行分类,只需要很少的训练样本,识别效果也较理想。因此,Rubine 方法被成功地应用到一些系统中,如 gesture 设计工具 Quill^[30],SILK (Sketching Interfaces Like Krazy)^[31]。但是,Rubine 算法的局限在于只能用于单笔划识别,虽然 SILK 对它进行了改进,引入笔划间的空间位置,以识别由多笔划构成的图形,但它要求用户必须以固定模式的笔划来勾勒图形,而且图形构成比较简单。

整^[17]等。

•识别 识别主要有两种方式:第一种是以 Sezgin 为代表的基于规则方式,手工定制模板的属性,在识别过程中依次测试图形是否与模板的属性一致。这种方法计算简单,适合于识别结构简单且数量不多的图形。第二种是训练模板的方式^[19,25],系统事先定义好若干种属性的图形表现,而后从用户提交的样本中自动生成模板的属性。这种方式适合于结构相对复杂、模板数量较多的图形的识别。Calhoun^[25]定义了图元之间的位置关系:相交、交点相对位置、相交直线间的夹角、平行等,及图形元素的属性:类型(线或弧)、长度、相对比例、斜率(线)和半径(弧),而后为使用语义网络组织图形模板。进行图形识别时,计算输入与图形模板的相似度。由于用户绘制的草图可以自动拆分成图元及其关系,图形模板的训练不需要用户干预,模板库容易扩展,适应性强。但不可避免地,训练模板以及对图形测试所有已定义的属性增加了系统开销。

2.3 基于几何特征的方法

基于几何特征的方法直接将图形当成识别单元,直接提取草图的几何特征用于分类。通常它要以一定的方式先分割好不同的图形。目前主要的分割方法有两种,第一种设置提笔的时间阈值(timeout),当用户提笔超过此阈值时,系统认为是新的一个图形的开始,这种方法较为简单,但它以牺牲用户输入的灵活性为代价^[36,37]。第二种方法则将当前输入图形与模板库的所有图形计算相似度,并记录最大相似值,当随着输入的进行,最大相似值骤减时,对图形分组^[24]。这种方式的不足在于用户输入的草图必须是模板库中已有的图形,否则不能正确分组,而且系统的开销较大。

Electronic Cocktail Napkin Project (ECNP)^[36]中使用的多笔划识别算法是基于几何特征方法的一个例子。ECNP 使用笔路径(pen path)、闭包盒的大小、外形比例(aspect ratio)和顶点数等信息描述字形(glyph),并按照一定的匹配策略从候选图形中选择最佳匹配图形。其中,字形指的是相邻两次提笔超过一定时间绘制的笔划组合,笔路径是把字形归一化到3×3的网格中形成的网格位置1到9的数字编号序列。几何特征的抽取实质上是将笔划信息降维,但由于图形的复杂性,难以固定维数的几何特征保留笔划的完整信息,部分信息在降维过程中丢失了。因此,几何特征的方法识别的精确性不高,一些视觉不相似的图形的几何特征表示可能相近。

2.4 基于组合图形的方法

基于组合图形的识别方法^[22,36]提高了图形识别粒度,它的存在主要是为了复杂图形识别的需要。虽然上述三种方法具有较强的简单图形的描述能力,但是对于复杂图形则不具有适应性。基于组合图形的方法,目前基本上用的是图匹配策略:将子图形抽象为结点,其相应关系抽象为边。

ECNP^[36]为了描述具有特定意义的结构,在字形的基础上又构建了配置层,其中配置:=字形|配置+配置(+表示一定空间关系)。它预先定义多个二元空间关系谓词,如垂直、正上方、T型相交等,按照不同的细化粒度分层,并建立不同关系的属性表,如交换性、传递性,及约束表,如T型相交仅作用于线性图形元素等。当用户提交的配置识别成若干字形后,系统根据细化者优先的原则测试其二元关系谓词,将所有值为真的二元关系显示给用户,同时允许用户显式地编辑这些关系。

基于图形内部的图元存在一定的层次性,如果直接根据图形内部图元及两两间位置关系识别复杂图形,存在不符合视觉相似性和计算复杂等不足的认识,Xu在图元和复杂图形之间,加入一些基本图形以提高识别效率。其中,基本图形具体表现为出现频繁且结构简单的图形,如三角形、四边形等。进而将不同层次的图形元素构成抽象为属性关系图,并利用约束的部分枚举算法预测后继的有效状态,跳过无效的中间匹配状态,进行状态空间剪枝,得到较大的匹配算法优化率^[22]。虽然类似于ECNP中的配置,但基本图形不具有特定的意义,它由系统选定,用于提高识别效率。但是,这也带来基本图形的选择问题,因为对于不同的图形集,其基本图形可能不相同。

3 语义理解

用户输入的草图在没有被赋予语义信息时,图形构成和对其的理解具有不确定性。语义理解是在图形识别的基础上,应用一定的背景知识,包括领域知识、用户信息等等,或借助

其它手段,赋予草图真实的含义,从而实现用户绘制意图的捕捉及辅助设计的目的。语义理解的领域依赖性强,不同领域间的方式差别大。但归结起来,语义理解一般含有三个环节:语义获取、关系分析和高层应用。

3.1 语义获取

语义获取指的是输入草图经过图形识别后,系统获取草图中的图形在特定用户及应用背景下的含义。语义获取的主要困难来自:相同的图形表现的草图对象在不同的领域、不同的用户、不同的上下文环境下,可能存在不同的解释。

ECNP^[36]保存已知的上下文列表,包括“当前上下文”和“当前上下文链”。其中,上下文相当于某个特定的应用,如电路图,它包括:字形模板、配置识别器、空间关系、其它上下文对字形的解释等四个组成部分。识别器会先在当前上下文中寻找与输入匹配的字形和配置,而当前上下文链中包括了一系列会影响识别的其它的上下文。当属于某领域的字形或配置匹配成功时,ECNP将设置或重置当前上下文和当前上下文链,以保证识别处在最佳的搜索路径上。

不同于ECNP,nuSketch^[38]提供了多通道的草图输入方式,它通过分析语音输入和笔划输入之间的关系直接给相应的笔划标记语义信息。它不精确地识别对象的图形表现,而是分析语音和笔划的语法,使用对话管理和知识推理方法标记笔划的语义以消除笔划的语义模糊性。nuSketch实质上是语义标记工具。

Hammond等^[39]则更进一步,考虑到语义解释对应用领域的依赖性,提出在领域无关的草图识别基础上用于描述领域相关知识的一种领域描述语言。它主要包含形状信息、约束信息、编辑行为、领域描述语法等,还可以是任何有助于识别过程的信息,如笔划顺序、方向等。这种领域描述语言通过统一的描述框架将图形在不同领域内的表示规范化,分离图形的几何特性和语义特性增强了系统在多领域内的适应性,提高系统的扩展能力。

3.2 关系分析

草图理解系统中,关系在不同的概念层次上都存在。此处的关系指的是草图的空间布局(Spatial Arrangement)在特定领域背景下的语义解释。关系分析的核心问题是对空间布局建模。Guan^[40]指出空间布局方法常常在一些显式或隐式定义的要求和原则指导下进行,但这些要求和原则却可能相互矛盾。这种现象在许多领域都存在,尤其是建筑设计和机械设计领域。

因此,大多数系统试图利用领域的特殊性,将定量的位置信息转化为定性表达,并使用更为自然的语言在较高的抽象层次上建立便捷的关系描述机制^[41~44]。值得注意的是,关系分析是草图在语义对象层次上的空间关系分析方法,与草图识别部分中空间关系分析方法没有本质的区别,两者相互间有许多方法可以借用。分析现有的关系分析技术,主要包括了以下环节:第一,预先定义基本关系,确定关系的属性、作用对象及其约束条件等;第二,人为定制或系统半自动定制复合关系,复合关系由多个基本关系构成,它代表某种确定的关系语义;第三,系统根据已有的关系定义发现草图中存在的关系。

虽然上述技术在特定的领域能够达到一定的效果,但是从整体上看,存在两点不足:首先,系统的空间关系必须由用户或系统设计者显式地定义或编辑,系统不具有自动的学习能力;其次,将关系定性化后,由于丢失了量的信息,其提供的空间关系无法继续深化,使得一些要求更细粒度关系的应用

无法开展。但是,这些不足与空间布局问题的复杂性有关,目前还没有好的解决方案。

3.3 高层应用

高层应用是在确定了基本对象及其关系的语义,捕捉到用户意图,服务于特定领域的应用。从某种意义上,高层应用是草图理解技术的延伸,它使草图理解技术的效能最大化,体现了草图理解技术的真正价值,同时,应用需求会对低层的草图理解产生很大的影响,因此,高层应用被视作草图理解技术的一部分。其中,领域知识的表示与推理方法是关键。由于高层应用服务于特定的领域,不同应用间的实现方法差别很大,不具有可比性。因此,本文仅对典型的几个原型系统做概要介绍:

SketchIT^[45,46]是由 Carnegie Mellon 大学机械工程的 Tom Stahovich 设计开发的支持概念机械设计的工具。当前流行的 CAD 软件很大程度上是精确图形绘制和展现的工具,它实际上不能够真正理解设计的含义,并且也不能够直接使用设计师在精确设计前为了辅助思维而在纸上勾画的草图。SketchIT 正是为弥补现有 CAD 软件的不足,使概念设计可计算,将机械设计的草图转化为几何描述,能够理解、支持草图设计而创建的。它能够向设计者提供多个基于此草图的设计方案。

Landay 设计实现的 SILK^[31,47,48]是一个支持使用草图进行图形用户界面设计的系统。这个系统是基于 CMU 的 Garnet 系统。该工具可以让用户自由地绘制粗略的设计思想,还可以对设计思想进行简单的测试——将用户的界面草图设计自动转化为可运行的界面实现,在最初的粗略设计之后可以加入颜色和字体等细节的信息。相比于传统的商用界面设计软件,SILK 提供给用户基于笔的自然的交互方式,使得用户界面设计者更接近于最初的设计习惯。SILK 的另一特点是它采用了实时识别的机制,这就使得一个设计不必全部完成之后才可测试,设计者可以在设计中任何时刻进行测试并作进一步改进。

加州大学伯克利分校的 GUIR 实验室设计实现了基于草图的网站设计软件 DENIM^[49],它允许网站设计者可以在不考虑具体细节的情况下对网站设计过程中的重要的问题进

行思考,如网站的结构等。该工具使用了可伸缩用户界面 ZUI (Zoomable User Interfaces)^[50]的思想,通过提供多个不同的视图,使得在缩放的不同层次为设计者提供不同的信息,并允许设计者在每个层次进行相应的交互动作。DENIM 可以把设计的结果输出为 HTML 格式的文件进行测试。

Tivoli^[51,52]是由 Xerox PARC 开发的用于支持非正式会议的电子白板系统,它运行在 Xerox 的电子白板 Liveboard 之上。Tivoli 的目标是将传统白板的工作方式和计算机的处理能力结合起来,使得人们更有效地进行思想交流、信息共享,同时能够对会议的内容进行有效的管理。它作为基于笔的用户界面研究中的一个划时代的交互系统,原创性地提出了一些基于笔的用户界面的基本概念,如手势、笔划等。Tivoli 首先提供给用户的是一个能够完成草图任务的白板,保留手绘信息的原有外观,并进行会议场景分类。

Electronic Cocktail Napkin 是由 Mark D. Gross 领导的研究小组研发的^[36,41],其目的是设计一个允许使用者进行各种类型设计、建模和仿真的自由绘制系统,对使用者瞬间的思想火花进行捕捉和验证。Electronic Cocktail Napkin 除了提供基本的基于笔的交互功能之外,还提供了可训练的符号识别器,能够分析复杂符号的空间关系以及进行相似草图的匹配。

4 研究思路与方向

本文在前面部分较为系统地介绍草图理解技术在各个方面取得的成果。为了更充分地分析草图理解系统,本文以下将从系统模型的角度分析草图理解系统中的关键问题。回顾图1中的草图理解系统结构图,用户以笔式交互方式输入草图,草图识别模块提取输入草图的图形构成信息,而后语义理解模块根据草图的构成信息在特定领域及用户的背景下获取草图对象及其关系的语义解释,从而完成草图理解基本过程。分析这个过程:首先,在这个过程中存在三个相对独立的实体:用户、几何构成和应用领域,如图3(a);其次,从功能角度,用户习惯影响着草图识别模块与语义理解模块的工作结果,如图3(b)。因此,本文将草图理解系统模型拆分成用户模型、几何模型及领域模型三个子模型,如图3(c)。

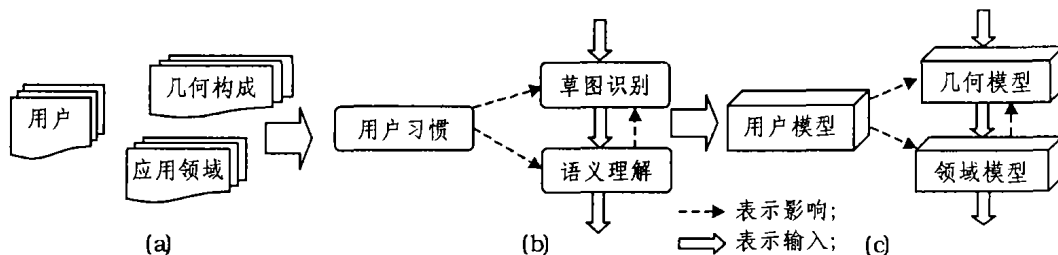


图3 草图理解系统建模

4.1 几何模型

几何模型服务于草图识别,它注重草图的构造及展现。理想状态下,系统应存储所有需识别图形的模板,而后与用户输入的草图进行匹配。但图形形态的不可枚举性,用户绘制的随意性及图形展现的个性化都给草图识别带来很大的困难。事实上,草图理解不必强求识别后图形的精确表现(完全消除模糊性),而且,对于特定领域其图形数量相对少得多。因此,为了捕捉和理解用户意图,在草图识别阶段,只需提取足够的信息以区分不同的图形,并保证有一定的扩展性。图2中基于笔划表示和几何特征的草图识别方法就是明显的例子。

除了在第2部分介绍的基于不同图形表示的草图识别方法外,几何模型中还存在两个关键问题:

• 模板的动态管理机制 当系统无法枚举所有的有效图形模板时,允许用户在使用过程中动态添加模板是一种较为可行的做法,即模板库的动态管理机制。目前主要有两种方法:第一,使用某种模板描述语言,允许用户使用该语言显式地添加、编辑模板^[39];第二,用户提交图形模板的样例,系统生成模板属性并返回给用户,由用户编辑,并最终确认^[36,41]。另外,用户事先不定义模板,系统从用户绘制过程的反馈(如对识别的纠正、增加模板的 gesture 等)中,自动地添加模板

也是一种方法,它能保证用户更大的灵活性,目前还没有研究者这样做。

•模糊性的适时消除 模糊性提高了用户绘制的灵活性,同时也增加了识别的难度。现有的技术过于注重自下向上的精确识别,而忽视了自顶向下的模糊消除。当低层图形的信息量不足以消除模糊性的时候,如果强制转化为精确图形,可能导致上层图形的识别错误。人的识别习惯往往先保留模糊性,自底而上地识别,当到达某个层次信息量足以确定图形时,再使用自顶而下的策略,消除低层图形的模糊性。基于此,Alvarado 等^[53]提出综合低层的形状信息与高层的与领域相关的上下文信息来提高识别精确度的方法。它有四个步骤:1)使用领域无关的识别工具生成复合对象模板,模板中可能有些形状未被识别;2)与已知模板比较,对未识别的形状重新解释(多种解释);3)系统根据其它一些已确定解释的模板确定时间和空间上下文,并对2)中的多种解释做出评价;4)去除不可能的解释。在领域知识较充分的情况下这种方法能有效地消除模糊性。

4.2 领域模型

领域模型服务于语义理解,它关心概念设计时草图图形及其空间关系表达的领域内存在对象及其关系的语义解释,以及这些对象及关系组合而构成的方案特征。一方面,它在草图识别的基础上获取对象语义,并挖掘相互间存在的关系语义;另一方面,根据对象及关系组合的方案特征为用户提供设

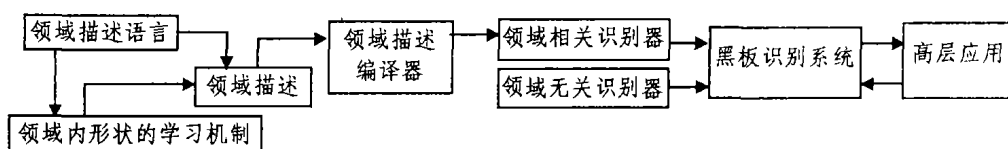


图4 草图语义映射框架

Hammond 等提出的草图语义映射框架,提供了一个有效解决领域适应性的方向,但它的实现还依赖于框架中的一些关键技术,如领域描述语言、领域内图形的学习机制、领域描述编译器、领域相关识别器、领域无关识别器和黑板识别系统等。

4.3 用户模型

用户模型解决的是用户适应性问题。草图理解系统强调“以人为中心”,但是目前的草图理解机制,无论是草图识别还是语义理解,对人的因素考虑得还不够充分。在草图绘制过程中,不同的用户具有不同的特性或习惯,例如笔划的随意性、笔序、语义表达习惯(草图理解技术主要应用于早期的概念设计阶段,用户表达图形并不一定遵循领域规范,因此,对于领域内的概念和关系,不同的用户可能会有不同的表达机制)。这些习惯都为正确的意图预测带来困难,因此,需要建立一种用户适应机制,能够通过用户的反馈,使得系统逐步适应用户特性或习惯,从而准确地预测用户的意图。其中,用户适应性的解决关键在于用户特征的抽取及利用两个方面。

虽然现有草图理解技术关于用户适应性的问题考虑得较少,但是用户适应性作为智能化人机接口的重要部分,在许多其它领域已经得到重视。Anthony^[57]指出用户适应系统(User Adaptive Systems)是指以用户模式的获取和应用为基础,适应个体用户习惯的交互式系统,其中包含某种形式的学习、推理或者决策。目前解决用户适应性的方法主要有三类:基于数据训练的方法、基于理论推理的方法、决策制导的方法。

计环境支持。由于方案特征的分析更侧重于对应用领域方面的技术内容,本文只关注第一方面。

草图设计行为总是在特定的领域背景下进行,因此,领域模型关键解决的还是草图识别对不同应用领域的适应性问题,即领域适应性问题。Hammond 等^[54]认为目前大部分草图识别工作领域依赖性强,领域细节内嵌在草图识别系统中,是很难在已有应用基础上扩展新的草图应用的主要原因,为此,他们提出多领域草图识别框架(如图4所示),以建立一个领域无关的草图识别系统,简化开发过程。其中,“领域描述”是一个文本文件,它遵循“领域描述语言”的语法规则,描述需要识别系统支持的必要信息,包括预定义的图形、约束、编辑行为及任何其它的有助于识别的信息。“领域描述”可由手工书写或由系统从若干手绘的样例及问题解答中学习而得^[55]。“领域描述”经过“领域描述编译器”^[56]编译生成“领域相关识别器”。识别器表现为可读代码及相关的数据结构。“领域无关识别器”使用如线、弧或更复杂的图形等图形识别的结果^[35],甚至是更自然的语音识别器。“领域相关识别器”和“领域无关识别器”被置于“黑板识别系统”的黑板结构^[53]中。“黑板识别系统”集成自顶向下和自底向上的策略,综合了上下文及图形信息,通过贝叶斯网络处理模糊性问题。该框架较为全面地分析了草图图形与语义的关系,较为有效地解决不同领域草图理解系统模块重用的问题。

在草图理解系统中,用户习惯主要表现在笔划、笔序及语义表达习惯等几个方面。由于用户习惯与用户的领域背景、心理以及认知联系较为紧密,目前的草图理解系统无法有效地模拟用户习惯,因此使用基于数据训练的方法较为合适。在数据训练的方法中,根据应用需求的不同,又存在两种做法:一种是系统收集一定量的当前用户数据,然后使用一种或者多种学习方法来训练模型^[58],是系统对特定用户的适应;另一种是系统使用多个用户的数据,这样学习的模型具有普遍的用户适应性,至少对同一个组里的用户具有很强的适应性^[59]是系统对多用户的适应。这些为草图理解系统中的用户适应性问题提供了较好的解决思路。

总结 草图理解系统的目的是捕捉和理解用户的输入意图。围绕着如何将带有模糊信息的草图转化成计算机能够理解的用户意图这一问题,本文从草图识别和语义理解两方面,总结了草图理解技术的研究现状,主要包括关键的处理步骤、代表性方法和相关评价等。

草图理解技术因其为用户勾画思路、深化思维和萌生新意提供支持,受到众多研究者的关注,而成为一个重要的研究课题。同时,由于它需要根据人对形状的视觉认知规则、思维习惯和信息模型对用户的输入草图及其过程进行推理和猜测,极富挑战性。在当前人工智能的发展无法对用户绘制的草图及其意图表达建立统一模型的情况下,有效利用草图的几何构成、用户、应用领域三方面信息不失为提高草图理解系统理解能力的一种有效途径。

参考文献

- Moran T P, Chiu P, van Melle W. Pen-based interaction techniques for organizing material on an electronic hiteboard. In UIST 1997, Symposium on User Interface Software and Technology, ACM, 1997. 45~54
- Jorge J A, Glinert E P. Calligraphic Interfaces: towards a new generation of interactive systems. Jorge J A, Glinert E P (Guest editors), Computers & Graphics, 2000, 24(6): 817~823
- Turk M. Moving from GUIs to PUIs: [Technical Report, MSR-TR-98-69]. Microsoft Research, Microsoft Corporation, Symposium on Intelligent Information Media, Tokyo, Japan, Dec. 1998
- Maher M L, Boulanger S, Poon J, et al. Assessing computational methods with a framework for creative design processes. In: Third Intl. Round-Table Conf. on Computational Models of Creative Design, University of Sydney, 1995. 233~265
- Gero J S. Recent advances in computational models of creative design. In: Pahl P, Werner H, eds. Computing in Civil and Building Engineering, Balkema, 1995. 21~30
- Elsas P A, Vergeest J S M. New functionality for computer-aided conceptual design: The displacement feature. Design Studies, 1998, 19(1): 81~102
- 孙正兴, 徐晓刚, 孙建勇, 金翔宇. 一种面向方案设计的图形输入工具. 计算机辅助设计与图形学学报, 已录用
- Xu X G, Liu W Y, Jin X Y. Sketch-based User Interface for Creative Tasks. In: 5th Asia Pacific Conf. on Computer Human Interaction, APCHI 2002, Beijing, Nov. 2002. 560~570
- Microsoft Presspass, Digital Ink, Breakthrough Technology in Tablet PC, Brings the Power of the Pen to the Desktop. <http://www.microsoft.com/presspass/features/2002/>, 2002
- Microsoft Presspass, With Launch of Tablet PCs, Pen-Based Computing Is a Reality. <http://www.microsoft.com/presspass/features/2002/>, 2002
- Li Y, Guan Z W, Dai G Z. Modeling Post-WIMP User Interfaces Based on Hybrid Automata. Journal of Software (China), 2001, 11(5)
- Li Y, Guan Z W, Chen Y D, Dai G Z. Research on Gesture-based Human-Computer Interaction. In: Proc. of China-graph, Hangzhou, China, 2000
- Li Y, Guan Z W, Wang H, Dai G Z, Ren X. Structuralizing Freeform Notes by Implicit Sketch Understanding. In: Proc. of AAAI Symposium on Sketch Understanding, Stanford University in Palo Alto, CA, 2002
- 王坚. <http://research.microsoft.com/asia/group/research/motimodalUI.asp>, 2003
- Xu X G, Sun Z X, Peng B B, Jin X Y. A SRG-Based Online Composite Graphic Recognition Strategy For Sketch-based User Interface. In: IEEE The First Intl. Conf. on Machine Learning and Cybernetics 2002, Beijing, China, Nov. 2002. 723~728
- Peng B B, Sun Z X, Xu X G. SVM-based Incremental Active Learning for User Adaptation For Online Graphics Recognition system. In: IEEE The First Intl. Conf. on Machine Learning and Cybernetics 2002, Beijing, China, Nov. 2002. 1379~1386
- 孙建勇, 金翔宇, 彭彬彬, 孙正兴. 一种快速在线图形识别与规整化方法. 计算机科学, 2003, 30(2): 172~176
- 彭彬彬, 金翔宇, 徐晓刚, 孙正兴. 基于 SVM 增量学习的用户适应性研究. 计算机科学, 2003, 30(3): 75~79
- 邱庆华, 张斌, 孙正兴. 一个面向构思的手绘草图识别系统. 计算机科学, 已录用
- Peng B B, Sun Z X, Liu W Y, Cong L L. A Dynamic User Modeling for Sketchy Shape Recognition. In: Proc. of 5th IAPR Intl. Workshop on Graphics Recognition, Catalonia, Spain, July, 2003, 已录用
- Peng B B, Sun Z X, Jin X Y. Study of SVM-Based Incremental Learning for User Adaptation in Multi-Class Classification Environment. Journal of Nanjing University (Natural Science Edition), 已录用
- Xu X G, Sun Z X, Liu W Y. Matching Spatial Relation Graphs Using a Constrained Partial Permutation Strategy. Journal of South East University (Natural Science Edition), 已录用
- 彭彬彬. 草图理解中的用户适应性研究: [南京大学研究生院硕士学位论文]. 2003
- 徐晓刚. 草图理解系统及关键技术研究: [南京大学研究生院硕士学位论文]. 2003
- Calhoun C, Stahovich T F, Kurtoglu T, Kara L B. Recognizing multi-stroke symbols. In: Proc. of AAAI 2002 Spring Symposium on Sketch Understanding, 2002
- Liu W Y, Qian W, Xiao R, Jin X Y. Smart Sketchpad-An On-line Graphics Recognition System. In: Proc. of ICDAR2001, Seattle, 2001. 1050~1054
- Rubine D. Specifying gestures by example. Computer Graphics, 1991, 25: 329~337
- Esther M A. An Efficient Computable Metric for Comparing Polygonal Shapes. IEEE Trans. on PAMI, 1999, 13(3): 209~216
- Sajjapong A, Vatanawood W, Covavisaruch N. On-line Handwritten Thai Character Recognition Using A Neural Network: [Master Thesis]. Department of Computer Engineering, Faculty of Engineering, Culalongkorn University, 1999
- Allan Christian Long Jr, Landay J A, Rowe L A. Implications For a Gesture Design Tool. In Human Factors in Computing Systems (SIGCHI Proceedings). ACM, ACM Press, May 1999. 40~47
- Landay J A. SILK: Sketching Interfaces Like Crazy. In: Proc. of Human Factors in Computing Systems, ACM CHI'96, Vancouver, Canada, 1996. 398~399
- Chen C L P, Xie S. Freehand drawing system using a fuzzy logic concept. Computer-Aided Design, 1996, 28(2): 77~89
- Jenkins D L, Martin R R. Applying constraints to enforce user's intentions in free-hand 2D sketches. Intelligent systems engineering, 1992(1. part 1.)
- Liu W Y, Jin X Y, Sun Z X. Sketch-Based User Interface for Inputting Graphic Objects on Small Screen Devices. Lecture Notes in Computer Science, vol. 2390 (Graphics Recognition: Algorithms and Applications, Selected and revised papers from Proc. Of GREC2001, September 7~8, 2001, Kingston, Canada), Springer, 2002. 67~80
- Sezgin T M, Stahovich T, Davis R. Sketch based interfaces: Early processing for sketch understanding. In: Proc. of 2001 Perceptive User Interfaces Workshop (PUI'01)
- Gross M D, Do E Y. Ambiguous intentions: a paper-like interface for creative design. In: Proc. of UIST'96, 1996. 183~192
- Stahovich T F. SketchIT: a Sketch Interpretation Tool for Conceptual Mechanical Design: [PhD Dissertation of MASSACHUSETTS Institute of Technology]. (Technical Report No 1573), March 2002
- Ferguson R W, Forbus K D. A cognitive approach to sketch understanding. In: Proc. of AAAI 2002 Spring Symposium on Sketch Understanding, 2002
- Hammond T. A domain description language for sketch recognition. MIT Student Oxygen Workshop. 2002
- Guan X. Computational support for early geometric design, in CAD center, Dept. of DMEM. Strathclyde, Glasgow. 1994. 275
- Gross M D. The Electronic Cocktail Napkin-A Computational Environment for Working with Design Diagrams. Design Studies, 1996, 17: 53~69
- Forbus K, Usher J. Sketching for knowledge capture: A progress report. In: Proc. of IUI'02, San Francisco, California, 2002
- Baykan C A, Fox M S. WRIGHT: A Constraint Based Spatial layout system. Artificial Intelligence in Engineering Design, Edited by Christopher Tong & Duvvuru Sriram., 1992, 1: 395~432
- Guan X, Duffy A H B, MacCallum K J. Prototype system for supporting the incremental modeling of vague geometric configurations. Artificial Intelligence for Engineering Design, Analysis and Manufacturing, 1997, 11: 287~310
- Stahovich T F. SketchIT: a Sketch Interpretation Tool for Conceptual Mechanical Design: [PhD Dissertation of MASSACHUSETTS Institute of Technology]. (Technical Report No 1573), March 2002
- Stahovich T F, Davis R, Shrobe H. Generating Multiple New Designs from a Sketch. In: Proc. of 13th National Conf. on Artificial Intelligence, 1996

- 47 Landay J A, Myers B A. Sketching Interfaces: Toward More Human Interface Design. IEEE Computer, 2001, 34(3): 56~64
- 48 Hearst M A, Gross M D, Landay J A, Stahovich T F. Sketching Intelligent Systems. IEEE Intelligent Systems, 1998, 13(3): 10~19
- 49 Lin J, Newman M, Hong J, Landay J A. DENIM: Finding a Tighter Fit between Tools and Practice for Web Site Design. In: Proc. of ACM CHI'2000, 2(1): 510~517
- 50 Bederson B B, Meyer J, Good L. Jazz: an Extensible Zoomable User Interface Graphics Toolkit in Java. In: Proc. of UIST'2000, San Diego, CA, 2000. 171~180
- 51 Pedersen E R, McCall K, Moran T P, Halasz F G. Tivoli: An Electronic Whiteboard for Informal Workgroup Meetings. In: Proc. of the ACM INTERCHI'93 Conf on Human in Computing Systems, 1993. 391~398
- 52 Moran T P, Melle W V, Chiu P. Spatial Interpretation of Domain Objects Integrated into a Freeform Electronic Whiteboard. In: Proc. of UIST'98, San Francisco, CA, 1998. 175~184
- 53 Alvarado C, Oltmans M, Davis R. A framework for multi-domain sketch recognition. Sketch Understanding, Papers from the 2002 AAAI Spring Symposium, 1~8
- 54 Hammond T, et al. Multi-domain sketch recognition. MIT Student Oxygen Workshop, 2002
- 55 Veselova O. Perceptually based learning of shape descriptions from one example. MIT Student Oxygen Workshop, 2002
- 56 Sezgin M. Generating domain specific sketch recognizers from object descriptions. In: MIT Student Oxygen Workshop, 2002
- 57 Anthony J. Adaptive interfaces and agents. In: Julie A. Jacko, Andrew Sears, eds. Handbook of Human-Computer Interaction in Interactive Systems. Erlbaum, Mahwah, NJ, 2002. In press
- 58 Geoff W, Michael J P, Daniel B. Machine learning for user modeling. User Modeling and User-Adapted Interaction, 2001, 11: 19~29
- 59 Bruce K. Lifestyle Finder: Intelligent user profiling using large-scale demographic data. AI Magazine, 1997, 18(2): 37~45
- 60 Anthony J. Numerical uncertainty management in user and student modeling: An overview of systems and issues. User Modeling and User-Adapted interaction, 1996, 5: 193~251
- 61 Alfred K, Wolfgang P. The user modeling shell system BGP-MS. User Modeling and User Adapted Interaction, 1995, 4: 59~106
- 62 Anthony J, Barbara Grobmann-Hutter, Leonie March, Ralf Rummner, Thorsten Bohnenberger, and Frank wittig. When actions have consequences: Empirically based decision making for intelligent user interfaces. Knowledge-Based Systems, 2001, 1: 75~92

(上接第 119 页)

表 5

效果	一步预测		二步预测	
	次数	百分比	次数	百分比
成功	57	83.83%	60	89.55%
失效	5	7.35%	0	0%
失败	6	8.82%	7	10.45%
平均收益率	1.578%		2.815%	

误差	一步预测		二步预测	
	小于1%	1%-3%	小于1%	1%-3%
次数	53	14	15	25
百分比	77.94%	20.59%	22.39%	37.31%
平均误差	0.639%		1.385%	
备注:平均误差大小可反映预测风险大小				

可见通过异常挖掘后再对序列挖掘算法加以改进, 预测效果明显好转, 尤其对于多步预测。实际中多步预测的效果往往更受投资决策者青睐。

我们用同样的挖掘算法测试其它的股票(部分结果可见表 6), 结果是一致的。

通过异常挖掘可以有效地改进序列挖掘算法以提高预测的精度, 尤其对于多步预测。同时平均收益率也会有相应的增加。由此可认为这种改进设计是切实可行的, 而且效果也不错。当然, 我们的挖掘算法改进还有潜力, 可以利用第二次建立的模型再次进行异常挖掘, 如此循环逐步提高预测精度。当然这种提高是有限的, 这是由于股票走势本身的不可预测性和受各种因素的干扰所致, 通过有限次的循环, 一般可以最大限度地剔除历史数据中的外部干扰因素对预测精度的影响。

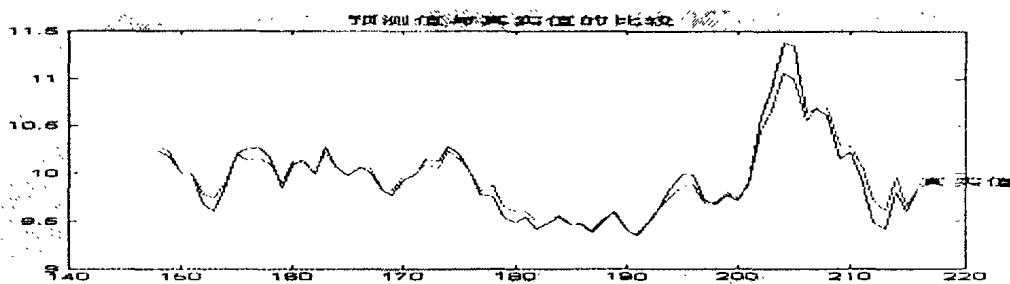


图 10

表 6

股票名称	初步模型				改进模型			
	一步预测		二步预测		一步预测		二步预测	
	成功率	收益率	成功率	收益率	成功率	收益率	成功率	收益率
青岛啤酒	83.82%	2.141%	89.55%	2.815%	89.71%	2.094%	92.54%	3.244%
深圳泰丰	88.24%	2.485%	89.55%	4.439%	88.06%	2.650%	92.54%	4.743%
涌鼎光辉	92.65%	1.759%	92.54%	3.562%	95.59%	2.020%	94.03%	3.806%

参考文献

- 1 Han J, Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers, Inc., 2001
- 2 John G H, Miller P. Building long/short portfolios using rule induction. In: Computational Intelligence for Financial Engineering, Piscataway NJ, IEEE Press, 1996
- 3 John G H. Stock selection using rule Induction. IEEE Expert, 1996. 52~58
- 4 Glymour C, Madigan D, Pregibon D, Smyth P. Statistical themes and lessons for data mining. Data mining and knowledge discovery, 1997, 1: 11~28
- 5 Liu L-M, et al. Data mining on time series: an illustration using fast-food restaurant franchise data. Computational Statistics & Data Analysis, 2001, 37: 455~476
- 6 王振龙. 时间序列分析. 中国统计出版社, 2000
- 7 易丹辉. 统计预测—方法与应用. 中国统计出版社, 2001
- 8 韦博成, 等. 统计诊断引论. 东南大学出版社