

基于贝叶斯网络的草图符号识别研究

李 路 周 良 丁秋林

(南京航空航天大学信息科学与技术学院 南京 210016)

摘 要 针对草图识别算法大多通过限制用户绘制习惯来提高识别精确度的问题,提出一种动态构造贝叶斯网络模型的草图符号识别方法。该方法采用了从下而上与从上而下相结合的识别算法。从下而上实现笔画的分割,根据后验概率产生假设模板,继而产生图形模板。在从上而下的处理中,通过假设模板重构实现笔画重组、根据图形模板的空槽实现笔画识别的纠错处理。通过对 UML 领域中草图符号的识别,表明算法能在不限制用户绘制习惯的基础上取得较好的识别效果。

关键词 贝叶斯网络,先验概率,后验概率,符号识别,假设模板

Research of Sketch Symbol Recognition Based on Bayesian Network

LI Lu ZHOU Liang DING Qiu-lin

(College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract To solve the current algorithm's limitation of restricting the users' drawing style, this article introduced a method of dynamically constructing Bayes net to sketch symbol recognition system. This paper adopted a identifying algorithm which is a combination of bottom-top and top-bottom. From bottom to top it realizes the segmentation of strokes, generating hypothesis templates according to posterior probability then generating graphics templates. From top to bottom it realizes regrouping strokes through reconfiguring hypothesis templates and handling noisy input according to the empty slot of templates. Through being applied to the domain of UML, we can get better recognition effect without restricting users' freely input.

Keywords Bayesian network, Prior probability, Posterior probability, Symbol recognition, Hypothesis templates

1 前言

草图符号识别一直是学者关注的研究领域之一,人们同时希望将其应用多个领域,比如建筑业、医药业、工程设计等。但是建立这样一个相关领域的草图符号识别系统却存在较多的困难。困难主要来源于人们可以以不同的笔序、任意的笔画数和笔画方向来刻画一组符号,同时人们刻画的符号通常又不是规整的而是凌乱的,再者相同的符号在不同的领域将会有不同的含义。因此,如何建立一个合理而有效的领域符号识别系统,将会是一项比较复杂的工作。

草图符号与一定领域相结合时,就具有一定语义上的含义。已经有很多学者在语义层着手来识别草图,并提出了许多有效的识别框架。CAIL^[1]建立在图形全局几何属性的基础之上,通过将模糊逻辑与识别符号的确信度联系起来,从而消除识别歧义。AgentSketch 是以代理为基础的草图符号解释框架^[2],主要是采用语境信息来消除歧义。在这种识别方法中,代理管理底层的符号识别,代理之间通过交换语境信息相互协调,从而对草图进行有效而精确的解释。G. Sahoo 和 B. Singh 提出了一种启发式框架^[3],它提高了绘画的自由度,降低了识别的复杂性。它取消局部查询策略,引入了一种全

局的启发式查询策略,从而提高了草图结构识别的精确度。LADDER 是最早用于草图符号描述的语言,支持领域草图的绘制、显示和编辑等^[4]。它由 Hammond 提出,语言包括一系列预定义的图形、约束、编辑行为和展现方法。麻省理工学院的草图研究项目组在 LADDER 语言的基础上设计实现了 SketchREAD,这是一种多领域的草图识别系统^[5]。当系统识别某一领域草图时,只需要用 LADDER 语言对该领域图形进行结构性描述。之后,系统使用语境来引导查询可能的解释,使用动态构造贝叶斯网络结构来评价这些解释。在识别过程中,可同时纠正底层的识别错,比如拐点的错误识别、线段或弧的错误划分等。Danilo Avola 等提出了另一种与 LADDER 相似的草图模型语言 SketchML^[6]。SketchML 语言建立在 XML 标准之上,具有更好的概括性,并且适合于各种应用领域。其特点是用户可以通过手画的方式建立领域图形基,从而建立草图接口。

识别草图符号的最大困难就是提高识别的精确度。由于草图符号是由基本图形(直线、弧、圆)组合而成,以一定的领域知识为应用背景,具备一定领域上的语义,因此要想提高草图符号识别的精确性,就需要将识别上升到语义层,通过符号的语境来有效提高识别的精确度。

到稿日期:2010-07-01 返修日期:2010-12-08

李 路(1986—),女,硕士生,主要研究方向为人机交互,E-mail:spring_dew@126.com;周 良(1966—),男,副教授,主要研究方向为人机交互技术、信息系统与信息安全、知识工程;丁秋林(1935—),男,教授,主要研究方向为 CIM, DSS, MIS。

2 基于贝叶斯网络的草图符号识别系统框架

本文建立的基于贝叶斯网络的草图符号识别系统,对各个领域的草图符号进行相关领域的语言描述,建立领域图形模板和领域符号的先验概率;在识别的过程中通过贝叶斯推理产生假设模板,动态产生贝叶斯网络模型;利用贝叶斯网络模型的不确定性模拟草图识别中由于噪声而造成的不确定性;通过假设模板重组以及图形模板的空槽处理来实现纠错处理。

基于贝叶斯网络的草图符号识别系统如图 1 所示。分为 4 个部分:笔画信息采集模块、训练模块、草图领域定义模块以及草图符号识别模块。鉴于原始笔画采样过程的不稳定性,笔画信息采集模块实现对原始笔画的采样值进行预处理,给其它 3 个模块提供处理后的笔画信息。在训练模块中,通过采集相关领域的图形样本,得到该领域图形的先验概率,为识别系统建立模块提供 CPT(condition probability table)。草图领域定义模块实现通过一种结构性描述语言来定义领域图形的几何基和符号,从而为识别系统建立模块提供领域模板。

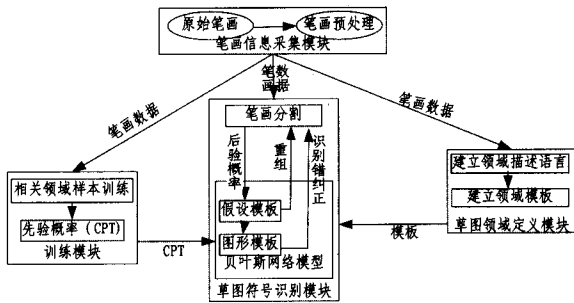


图 1 基于贝叶斯网络的草图符号识别框架

草图符号识别模块是草图识别框架的核心,对笔画数据实现从下而上和从上而下相结合的处理。在从下而上的处理中,首先笔画依据拐点分割成基本几何基,然后依据对基本几何基的观测值和领域的 CPT 形成后验概率,据此产生假设模板来构造贝叶斯网络模型。假设模板再依据其后验概率和产生时间生成图形模板。最后图形模板将作为最终识别的结果以结构性描述语言的形式展现给用户。在从上而下的处理中,主要是假设模板的重新产生实现笔画的重组,以及对图形模板中空槽的处理实现几何基识别错的纠正。

针对用户绘制草图的随意性和草图的凌乱性提出的基于贝叶斯网络识别系统,实现了允许用户自主性绘制和草图凌乱性处理的功能。用户自主性绘制,即允许用户交叉绘制草图符号。本文采用允许产生多个假设模板,并可在最近 3 个模板中查找几何基可入的空槽进行解释。

草图凌乱性处理,即针对用户绘制草图过程中出现的抖动现象,通过图形模板中的空槽查找未识别出的几何基和约束,完成识别纠错处理。

3 算法实现

3.1 贝叶斯网络模型设计

贝叶斯网络是一种将先验知识和周围环境观测结合起来推测现实中不确定性的一种方法。它由两部分组成:一

个有向无环图,实现对现实中相互影响的因素进行编码;另一个是一系列条件概率分布,说明这些因素相互影响的程度。网络中每一个节点代表一个假设,每个节点都包含一个条件概率表(CPT),用来说明父节点对其影响的方式^[5]。为说明网络中的概率信息,必须给予每个根节点先验概率,非根节点是条件概率。顶层图形先验概率的设置是根据当前领域该图形出现的相对概率得到的。通过子节点的观测值可得到非根节点的条件概率^[8]。

图 3 是在绘制图 2 笔画 1,2,3,4 后产生的贝叶斯网络结构。第一层实际上不属于贝叶斯网络,列出它是为了更好地解释假设评价产生的方法。在第一层上升到第二层中,使用底层的识别方法,识别出了几何基和约束。几何基和约束入槽时,将计算出第二层的 $O_i - O_j$ 和各约束的观测值(0, 1, 2)。 O_i 为 0 时,底层数据支持观测点; O_i 为 1 时,底层数据的支持是可选的; O_i 为 2 时,底层数据不支持观测点。本文采用最优平方错的计算方法作为形状偏移度量,即绘制的草图的形状或约束的相关度量值与标准模板库设置的最佳值之间的差值平方错。比如几何基中的直线,其值是绘制过程的角度累计值和标准值之间的差值平方。很显然,值越大,假设强度越低。表 1 列出了实现基本几何基(直线、弧、圆)形状偏移度量的连续值到观测点分离值之间的投影。在识别中,根据观测值得到条件概率,并根据条件概率得到后验概率,根据后验概率产生假设模板,进而产生识别出的图形模板。如图 3 所示,所建的模型是双向的,由下而上是识别的过程,由上而下是识别错纠正的过程。由上而下的箭头表明高层的假设为真,将直接使底层的假设也为真,实现识别出尚未入槽的笔画,从而实现对噪声的处理^[7]。

表 1 基本图形先验概率表

几何基(S)	基本图形			
	观测点(O)		CPT	
	观测值(v)	偏移值	$P(O=v S)$	$P(O=v \sim S)$
直线	0	$(L_L - S)^2 / S < 0.5$	0.98	0.111
	1	$0.5 \leq (L_L - S)^2 / S < 2.5$	0.0199997	0.258
	2	其它	0.0000003	0.631
弧	0	$(L_a - S)^2 / S < 0.167$	0.97	0.13
	1	$0.167 \leq (L_a - S)^2 / S < 1.5$	0.023	0.248
	2	其它	0.007	0.622
圆	0	$(L_e - S)^2 / S < 0.12$	0.8883	0.14
	1	$0.12 \leq (L_e - S)^2 / S < 1.3$	0.1006	0.136
	2	其它	0.0111	0.724

说明: $L_L, L_a, L_e, |S|$ 是根据当前所画图形的特征值得到的标准直线、弧、圆的弧长、笔画的实际长度。

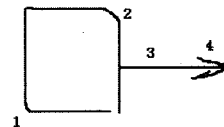


图 2 用户手绘草图

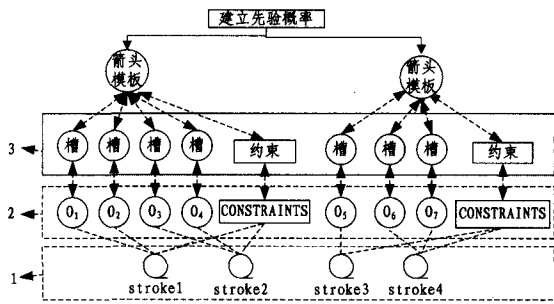


图3 贝叶斯网络模型

3.2 后验概率计算

先验概率是根据经验获得的领域图形模板的概率,后验概率是根据几何基的观测值推测出的领域图形的概率。在识别过程中,我们根据后验概率产生假设模板。

以下列出的是在一个模板中为计算后验概率所做的数据结构设计: n 个相互排斥的符号假设 $H_1 \sim H_n$; m 个观察属性 $O_1 \sim O_m$; 数组 $num[1 \dots m]$ 代表 m 个观测值的取值范围; $obs[a]$ 是第 a 个观测点的度量偏移值。先验概率 $Prior[1 \dots n]$ 为 n 个相互排斥的假设在该领域出现的概率。 $val[a]$ 是第 a 个观测点的观测值。 $Chance[1 \dots n][1 \dots m]$ 是 n 个假设下这 m 个观测点的概率分布函数。 $Chance[i][a]$ 即在第 i 种假设下第 a 个观测点观测值分别为 0, 1, 2 的概率。 $val[a]$ 的取值为 0, 1, 2。

边缘概率

$$likelihood(val[a]) = chance[1][a][val[a]] \times prior[1] + \dots + chance[n][a][val[a]] \times prior[n]$$

后验概率

$$postprob[i][a] = \frac{prob(H_i | val[a])}{likelihood(val[a])} = \frac{chance[i][a][val[a]] \times prior[i]}{likelihood(val[a])}$$

从式中可以看出,后验概率的结果值是依赖于 $chance[i][a][val[a]]$, 即第 i 个假设下第 a 个观测点取具体某个值的概率,也就是前面提到的条件概率表(CPT)。最后计算出观测点 a 的 $post[1 \dots n]$, 也就是根据 $val[a]$ 具体值得到的 n 个假设的后验概率。其值是根据 $val[a]$ 进行计算的,可写成函数 $postprob(val[a])$ 。后验概率的计算是在识别过程中动态进行,先验概率和 CPT 是根据该领域相关经验获得。我们根据后验概率进行假设评价和最优部分假设的选择。

以上的 $postprob(val[a])$ 是根据单个观测值计算得到的,在实际的识别过程中需要随着观测点入槽来实现假设模板的后验概率重新计算^[9]。当 $val[a+1]$ 入槽时,模板后验概率更新为

$$postprob(H_i = true / val[a], val[a+1]) = \frac{p(H_i, val[a], val[a+1])}{p(val[a], val[a+1])} = \frac{p(H_i) \times p(val[a], val[a+1] / H_i = true)}{\sum_{i=1}^n p(H_i) \times p(val[a], val[a+1] / H_i = true)}$$

式中, $p(val[a], val[a+1] / H_i = true)$ 也是由 CPT 决定的。

3.3 基于贝叶斯网络的草图符号识别过程

基于贝叶斯网络的草图符号识别主要包括笔画过滤、假设模板的产生、假设模板的判定、假设模板的重解和几何基纠错 5 个步骤。

步骤 1 笔画过滤

用户每画一个笔画,进行笔画分割,识别出几何基。每个几何基有一标记 flag,记录其状态,0 为初始态,1,2,3 分别为进行了 1,2,3 次假设模板的重解,-1 为不可识别态。每个笔画有个人槽时间 $stime$ 。假设已产生的假设模板(hypotemplate)数为 hn ,每个假设模板有个产生时间 $htime$,hypotemplate 中假设模板以产生时间的先后排序;每个假设模板还有个生存时间 $ltime$,过了这个时间就需要对该模板的入槽笔画进行重新解释,该时间的确定是根据经验获得的;每个假设模板还有个后验概率 $postprob$,它随着笔画的入槽动态变化。

步骤 2 假设模板的产生

依次处理识别的几何基,依据查询最近产生的假设模板中是否有与当前几何基观测值相匹配的空槽进行不同的处理。

- 取 $hypotemplate[hn-1]$ (最后产生的假设模板),查看该假设模板中是否有该几何基可入的空槽。如果有,则入槽,同时置该几何基 $flag=1$,并且对该假设模板进行后验概率的重新计算。

- 如果 $hypotemplate[hn-1]$ 中没有空槽可入,则依次查询 $hypotemplate[hn-2]$ 和 $hypotemplate[hn-3]$ 是否有空槽可入。如果有,则入槽,同时置该几何基的 $flag=1$,并且对该假设模板进行后验概率的重新计算。假设用户最多交叉绘制 3 个图形。

- 如果查询 $hypotemplate[hn-1]$, $hypotemplate[hn-2]$ 和 $hypotemplate[hn-3]$ 均无空槽可入,则根据该几何基的观测值,查询 n 个相互排斥的符号假设 $H_1 \sim H_n$,计算存在该几何基的符号假设的后验概率。选择最高后验概率产生新的 $hypotemplate$,该几何基入槽, $flag=1$ 。

步骤 3 假设模板的判定

依次扫描 hn 个 $hypotemplate$,计算其存在时间, $life = 当前时间 - hypotemplate[n].htime$ 。

- 如果 $life > hypotemplate[n].ltime$,则对该 $hypotemplate[n]$ 中入槽的笔画进行重新解释,进入步骤 4 处理。

- 如果 $life < hypotemplate[n].ltime$ 且 $hypotemplate[n].postprob >$ 模板产生门限,则产生模板 $template$,进入步骤 5 处理。

步骤 4 假设模板的重解

取出需重新解释模板 $hypotemplate[n]$ 的所有入槽的几何基,依据几何基产生的先后顺序分别做以下处理:

- 对于每个几何基,先看本次重新解释的过程中是否产生新的假设模板。如果产生,则看是否有空槽可入。如果有,则入槽,处理下个几何基;否则,找到先于该几何基产生时间最近的 $hypotemplate$,查看是否有空槽可入。如果有,则入槽,同时该几何基 $flag$ 加 1,继续处理下个笔画;如果没有空槽,则进入下一步处理。

- 根据该几何基的观测值,计算存在该几何基的图形的后验概率。选择后验概率是 $flag$ 高的图形,产生新的 $hypotemplate$ 。将该 $stroke[i]$ 入槽,该几何基 $flag$ 加 1。如果该 $stroke[i].flag > 3$,则 $stroke[i].flag = -1$,继续处理下个几何基。最多进行深度为 3 的后验概率试探。

步骤 5 几何基纠错

对于每个产生的 $template$ 中的空槽,找到离其源 $hypotemplate$ 产生时间最近的几何基且 $flag = -1$ 的笔画,对其通

过扩展直线或者弧的门限值来重新解释该笔画,看是否可成功入槽。

步骤6 读取下一个笔画返回步骤1。

在识别过程中,当假设模板中有多个槽可供几何基匹配时,算法中并不检查每个匹配的空槽,而是选择其中一个空槽进入,允许以后根据几何基之间的约束值进行假设模板中几何基的位置调整。

4 实验和分析

本文选用 UML 领域的常见符号作为识别样本,如图 4 所示。我们一共收集了 300 多幅 UML 草图、1800 个符号,并根据收集的草图符号设定每个符号的先验概率,建立 UML 图中每个符号的 CPT。实验中,让 15 位用户绘制 UML 图中的常见符号,系统实现对这些手绘的草图符号进行识别。

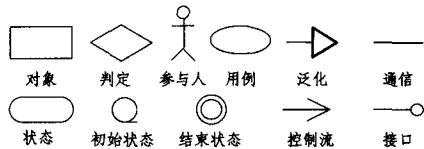


图 4 常见的 UML 语义符号

针对建立的基于贝叶斯网络的草图符号识别系统的分层处理结构,对收集的 UML 草图符号分别统计了其经过基于贝叶斯网络模型由上而下处理前后的平均识别率,如表 2 所列。

表 2 由上而下处理前后识别精确度比较

基本图形符号	由下而上处理后识别率	由上而下处理后识别率
对象	69%	95%
判定	70%	94%
参与人	52.3%	75%
用例	70.2%	95%
泛化	60%	69%
通信	80.2%	96%
控制流	75%	89%
初始状态	80%	92%
结束状态	73.2%	83%
接口	74%	80%

在做了从下而上的识别后,绘制符号的平均识别率最低为 52.3%,最高达到 80.2%。在做了从上而下的根据假设模板中基本图形的重新组合和根据模板空槽进行纠错处理后,识别的精确度得到极大的提高。符号的平均识别率达到 86.8%,最低为 75%,最高达到 96%。算法很好地将语义、时间以及空间信息结合起来,具有较好的识别效果。语义上,通过建立领域图形模板,将几何基和图形模板的空槽匹配,从而考虑图形的领域语义信息。时间上,通过赋予几何基和假设模板时间信息,从而考虑了笔画之间的时间关系。空间上,通过计算几何基之间的约束关系,从而考虑了空间关系。

结束语 本文提出了通过动态构造贝叶斯网络模型来模拟用户草图绘制的动态过程的一种基于领域语义的草图符号识别系统。它允许在草图绘制的过程中适时得到识别结果。系统首先通过建立领域符号描述形成图形模板库,并根据经验知识获得领域图形的先验概率,进而产生几何基的 CPT。

其次,在草图符号的识别过程中,系统将笔画分割成几何基,根据几何基的观测值和 CPT 产生后验概率,并根据后验概率进行假设评价,产生假设模板。最后,根据假设模板和成功识别后产生的模板进行从上而下的纠错处理,从而实现纠正错误的笔画分组和基本图形的错误识别。

本方法已经初步实现了对 UML 领域的草图符号识别,并且得到了较好的识别效果。在未来的工作中将从以下 3 个方面对系统进行改进。

1)虽然本识别算法能够很好地利用草图符号的领域语义、时间和空间信息,但是在具体识别某个领域的草图符号时,必须先手动建立领域图形库和 CPT。在未来的工作中,将通过学习来自动生成该领域符号的先验概率和条件概率,从而简化系统,提高先验概率的精确度。

2)为促进系统同时识别多个领域的符号,将引入公共图形库和领域专有图形库实现对多领域图形模板的管理。

3)为提高识别精确度和降低假设模板重组所带来的时间和空间复杂度,将提出软约束和硬约束的设计思想。即给几何基之间的约束关系设置一定的权重,具有较高权重的约束称之为硬约束,即假设模板产生时必须满足的约束条件。反之为软约束,即可选的约束条件。

参考文献

- [1] Fonseca M J, Pimentel C, Jorge J A. CALI: an online scribble recognizer for calligraphic interface[C]//Proceedings of AAAI Spring Symposium on Sketch Understanding. 2002;51-58
- [2] Casella V M G C G, Deufemia V, Martelli M. An agent-based framework for sketched symbol interpretation[J]. Journal of Visual Languages and Computing, 2008, 19(2): 225-257
- [3] Sahoo G, Singh B. A New Approach to Sketch Recognition Using Heuristic[J]. International Journal of Computer Science and Network Security, 2008, 8(2): 102-108
- [4] Hammond T, Davis R. LADDER: A language to describe drawing, display, and editing in sketch recognition[C]//Proc. of IJ-CAI'03. 2003;461-467
- [5] Alvarado C, Davis R. SketchREAD: a multi-domain sketch recognition engine[C]//Proceedings of User Interface and Software Technology(UIST'04). 2004;23-32
- [6] Avola D, Buono A D, Gianforme G, et al. SketchML a representation language for novel sketch recognition approach[C]//Proceeding of the 2nd International Conference on PErvasive Technologies Related to Assistive Environments. 2009
- [7] Alvarado C. Multi-domain Sketch Understanding [D]. United States, Massachusetts Institute of Technology, 2004
- [8] Charniak E. Bayesian networks without tears: making Bayesian networks more accessible to the probabilistically unsophisticated [J]. Artificial Intelligence, 1991, 12(4): 50-63
- [9] Liao Shi-zhong, Wang Liao, Lu Jin-liang, et al. An incremental Bayesian approach to sketch recognition[C]//Proceedings of the Fourth International Conference on Machine Learning and Cybernetics. 2005;18-21