



45-47

KDD 中基于感觉数据特性抽取的概念提升*

Model of Lift Concept in KDD Based on Mapping of Perceptual Character Data Abstraction

周永权¹ 冯嘉礼² TP18

(广西民族学院数学与计算机科学系 南宁 530006)¹

(广西师范大学计算机应用研究所 桂林 541004)²

摘要 感觉识别和概念的判断分别是直觉思维和逻辑思维的基础,对感觉数据特性的抽取可归纳为一定性映射,本文以感觉数据特性抽取的定性映射为模型,提出一种概念提升的新方法和神经网络学习算法,相比传统的提升方法,该方法更符合人们思维推理的习惯,对确定每一模糊概念隶属度问题似乎变得简洁,实用直观,通过对基准族学习,实现对感觉概念属性的检测。

关键词 KDD,感觉特征抽取,概念提升,基准族,神经网络学习算法

人脑神经网络

一、引言

感觉识别是人类思维的基础,知觉是对事物的属性整合的反映,为了识别事物的属性,各感觉和知觉神经元应针对其属性有一识别和检测的基准,否则,无法操作^[1]。感觉抽取的操作是将感觉属性 $a(x)$ 的数据 $d_a(x)$,转换为其特征 $p_a(x)$,而感觉特性 $p_a(x)$ 又随定性基准 $c_i(x)$ 而变,故感觉数据特性抽取数学模型为二元函数 $J(d(x), c_i(x)) = p_i(x)$ 的形式,因 $c_i(x)$ 可用区间 (α, β) 表示,这样,感觉抽取可归结为一定性映射^[2],且 $J(d(x), c_i(x))$ 随基准 $c_i(x)$ 改变而改变。

概念的提升是利用属性概念树进行爬升,使属性一个较具体的值被概念树上的父概念替代,并且对定性基准 $c_i(x)$ 利用基准变换放大或缩小(通常采用坐标平移,旋转)进行划分归并,利用划分后的基准来实施概念的提升,用感觉数据抽取特性模型,通过对基准区间的调整来实现学习,使得属性值抽象度愈提高,从而完成概念的提升,其中,概念的基准可随具体的对象和问题以及人们的认知结构而变,特别是基准区间随人的大脑而变的现象也是思维科学特有的,不仅概念的判断、推理,而且思维操作所产生的各种不确定性和矛盾现象(如:模糊性,歧义性)都可以从基准变换找到它的逻辑根源^[3]。

目前对概念的提升,主要建立在正态隶属云基础上,正态隶属云是指用来刻画模糊概念所涵盖论域中元素隶属于该概念的程度,每一模糊概念所涵

盖的论域中元素都是以模糊概念的期望值对应的元素为汇聚中心,其它的元素各自隶属于该模糊概念的隶属度值,依此来接近汇聚中心,值愈大,愈接近。这种方法对于概念上有种属关系的精确知识来说非常有效,但对 KDD 中普遍存在模糊概念进行提升显得力不从心^[4],本文另辟新径,以感觉数据特性抽取的定性映射为模型,借助于概念提升的思想,设计出一种基于感觉数据特性抽取定性映射模型的概念提升的新方法及概念提升的神经网络学习方法。

二、感觉数据特性抽取的数学模型

为了建立感觉数据特性抽取的概念提升的数学模型,首先给出几个定义。

定义 1 设 $N(p) = \{N_i | i \in N\}$ 为命题 p 的判断基准 $N_i(p)$ 的族,称映射 $T: N(p) \rightarrow N(p)$ 为命题的基准变换,如果对任意 $N_i \in N(p)$,存在一个 $N_j \in N(p)$ 使得:

$$T(N_i) = N_j, \text{ 记为: } T_i(N_i) = N_j$$

一般地,当基准族为区间,采用基准变换为平移,完成区间放大、缩小,如,设命题 $p(x)$: "x 是未成年人",判断基准取: $N_0 = (0, 18)$,随着年龄,经平移变换缩小为: $N_1 = (0, 8)$, $N_2 = (8, 12)$, $N_3 = (12, 18)$ 分别表示"儿童", "少年", "青少年"等基准区间。

定义 2^[5] 设 $a(x) = \bigwedge_{k=1}^m a_k(x)$ 为事物 x 的 m 个属性 $a_k(x)$ 的整合属性, $d(x) = (d_1(x), d_2(x), \dots, d_m(x))$ 是 $a(x)$ 的数据向量, $d_k(x)$ 是 $a_k(x)$ 的数据,若

*) 国家 863 高新技术资助项目。

存在性质 $p_i(x)$ 的一个定性基准 $c_i(x)$, 使断言“ x 具有性质 $p_i(x)$ ”成立, 则称 $p_i(x)$ 为感觉抽取 $J(d(x))$ 在基准 $c_i(x)$ 下的结论, 记: $J(d(x), c_i(x)) = p_i(x)$, 当 $m=1, J(d(x), c_i(x)) = p_i(x)$ 为简单感觉识别问题。

定义 2 表明, 感觉识别可视为数据 $d(x)$ 和定性基准 $c_i(x)$ 的二元函数, 当基准 $c_i(x)$ 变化, 导致性质 $p_i(x)$ 也发生相应的变化, 对感觉属性而言, $c_i(x)$ 可用一区间 (α, β) 来表示, 故感觉特性抽取可归结为一定性映射。

设 $D_1^+ = \{d(x)\} \subseteq R^+, R^+$ 为非负实数集, $P_1 = \{p_i(x)\}$ 为属性 $a(x)$ 的数值集, 则 $p_i(x)$ 的定性基准 $c_i(x)$ 可表示为一区间 $(\alpha, \beta) \subseteq D_1^+$, 又设 Δ 为提升算子, 记:

$$\delta_{d(x)\Delta(\alpha, \beta)} = \begin{cases} 1 & \text{iff } d(x) \in (\alpha, \beta) \\ 0 & \text{否则} \end{cases}$$

于是我们来定义概念的提升:

定义 3 由概念属性值 $d(x)$ 在感觉抽取定性映射的模型下, 确定概念的提升映射:

$$L: D_n \times \Gamma = P_n \text{ 使得 } L(d(x), (\alpha, \beta)) = d(x) \Delta (\alpha, \beta) = \delta_{d(x)\Delta(\alpha, \beta)} p_i(x) \quad (1)$$

其中, $\Gamma = \{(\alpha, \beta) | i \in N\}$ 为概念的定性基准族。

定义 4 设 $D_1^+ = \{d(x)\}, (\alpha, \beta)$ 是待提升概念性质 $p_i(x)$ 的定性基准, $\Gamma = \{(\alpha, \beta) | i \in N\}$ 为定性基准集, 则定义概念提升(1)式可写成:

$$L(d(x), (\alpha, \beta)) = \begin{cases} L_1(x) = L_{1(d)}(x) \text{ iff } d(x) \in (\alpha_1, \beta_1) \\ L_2(x) = L_{2(d)}(x) \text{ iff } d(x) \in (\alpha_2, \beta_2) \\ L_3(x) = L_{3(d)}(x) \text{ iff } d(x) \in (\alpha_3, \beta_3) \\ \vdots \\ L_n(x) = L_{n(d)}(x) \text{ iff } d(x) \in (\alpha_n, \beta_n) \end{cases}$$

若令 $\Gamma_n = \Gamma_{n-1}$, 则 $L: D_n \times \Gamma_{n-1} \rightarrow P_n$, 写成向量的形式:

$$L \left[d(x), \begin{pmatrix} (\alpha_1, \beta_1) \\ (\alpha_2, \beta_2) \\ \vdots \\ (\alpha_n, \beta_n) \end{pmatrix} \right] = \begin{pmatrix} d(x) \Delta (\alpha_1, \beta_1) \\ d(x) \Delta (\alpha_2, \beta_2) \\ \vdots \\ d(x) \Delta (\alpha_n, \beta_n) \end{pmatrix} = \begin{pmatrix} \delta_{d(x)\Delta(\alpha_1, \beta_1)} p_1(x) \\ \delta_{d(x)\Delta(\alpha_2, \beta_2)} p_2(x) \\ \vdots \\ \delta_{d(x)\Delta(\alpha_n, \beta_n)} p_n(x) \end{pmatrix} \quad (2)$$

这样, 在定义 1 的基础上, 引进定义 4, 为概念提升提供了理论根据, 我们可从下面的实例看出, 更符合人们的日常思维、推理的习惯。

三、基于感觉数据特性抽取定性映射的模型——模糊概念“年龄”提升过程

为了便于说明, 我们以“年龄”这个普遍使用的概念为例来介绍, 首先, 设 $D_1^+ = \{0, 1, 2, \dots, 100\}$ 为“年龄”概念的属性的数值集, 按照人们表示习惯和日常生活中常用的概念, 则性质集 $P_1^{(1)} = \{\text{儿童, 少年, 青年, 中年, 老年}\}$, 若调整基准族为: $\Gamma^{(1)} = \{(0, 10), (10, 18), (18, 35), (35, 45), (45, 100)\}$, 对年龄的定性映射可表示为:

$$L \left[\text{age}(\text{man}), \begin{pmatrix} (0, 10] \\ (10, 18] \\ (18, 35] \\ (35, 45] \\ (45, 100] \end{pmatrix} \right] = \text{age}(\text{man})$$

$$\Delta \begin{pmatrix} (0, 10] \\ (10, 18] \\ (18, 35] \\ (35, 45] \\ (45, 100] \end{pmatrix} = \begin{pmatrix} \delta_{\text{man}\Delta(0, 10]}(\text{儿童}) \\ \delta_{\text{man}\Delta(10, 18]}(\text{少年}) \\ \delta_{\text{man}\Delta(18, 35]}(\text{青年}) \\ \delta_{\text{man}\Delta(35, 45]}(\text{中年}) \\ \delta_{\text{man}\Delta(45, 100]}(\text{老年}) \end{pmatrix}$$

这样, 可视不同的场合调整基准区间, 实现对“年龄”概念的提升, 我们称第一层提升。

其次, 若调整基准族 $\Gamma^{(2)} = \{(0, 15), (10, 25), (20, 45), (35, 45), (35, 100)\}$ 按照人们日常生活中所期望的概念, 对年龄提升为:

$$L \left[\text{age}(\text{man}), \begin{pmatrix} (0, 15] \\ (10, 25] \\ (20, 45] \\ (35, 100] \end{pmatrix} \right] = \text{age}(\text{man})$$

$$\Delta \begin{pmatrix} (0, 15] \\ (10, 25] \\ (20, 45] \\ (35, 100] \end{pmatrix} = \begin{pmatrix} \delta_{\text{man}\Delta(0, 15]}(\text{少儿}) \\ \delta_{\text{man}\Delta(10, 25]}(\text{青少年}) \\ \delta_{\text{man}\Delta(20, 45]}(\text{中青年}) \\ \delta_{\text{man}\Delta(35, 100]}(\text{中老年}) \end{pmatrix}$$

这样, “少儿”看作“少年”、“儿童”的父概念, “青少年”看作“少年”、“青年”的父概念, 其余类推, 相当于在第一层提升的基础上实施第二层提升, 得到提升后年龄的性质集 $P_2^{(2)} = \{\text{少儿, 青少年, 中青年, 中老年}\}$ 。

再次, 按照人们期望的概念, 若调整基准族为: $\Gamma^{(3)} = \{(0, 18), (18, 100)\}$, 对年龄概念提升:

$$L \left[\text{age}(\text{man}), \begin{pmatrix} (0, 18] \\ (18, 100] \end{pmatrix} \right] = \text{age}(\text{man})$$

$$\Delta \begin{pmatrix} (0, 18] \\ (18, 100] \end{pmatrix} = \begin{pmatrix} \delta_{\text{man}\Delta(0, 18]}(\text{未成年}) \\ \delta_{\text{man}\Delta(18, 100]}(\text{成年}) \end{pmatrix}$$

同样, “少儿”、“青少年”被父概念“未成年”代替, “青少年”、“中青年”、“中老年”被父概念“成年”替代, 完

