

形象智能, 多媒体处理技术

13 形象智能与多媒体处理技术的综合研究^{*)}

鲁东明 潘云鹤

(浙江大学人工智能研究所 杭州 310027)

摘要 Based on the analysis of relationship between human thinking and information processing, this paper will discover the internal relationship between imaginary intelligence and multimedia technology. Imaginary intelligence will be the theoretical basis of multimedia techniques, while multimedia technology can work as the technical platform for imaginary intelligence research. Finally, this paper proposes several approaches for integrated research of imaginary intelligence and multimedia technology.

关键词 Imaginary Intelligence, Multimedia Processing, Mental Image, Multimedia Feature

一、引言

用计算机来模拟人类智能一直是人工智能研究者所追求的目标。其研究的范畴已由模拟人类逻辑思维的智能拓展到模拟人类形象思维的智能。

认知心理学的研究发现,人类智能包括感知觉、注意、表象、学习、记忆、思维和语言等方面,而这些方面均可理解为信息的获得、存储、加工和使用的过程,即人类智能的活动过程是信息处理的过程。计算机信息处理技术是人工智能研究的技术基础之一。同时,用计算机信息处理技术实现对人类智能活动的模拟,将以人工智能研究成果为理论基础。

认知科学的研究结果表明,人类所接受的90%以上信息是形象性的,模拟形象思维智能的研究是人工智能研究发展有巨大潜力的新方向。为此,浙江大学的潘云鹤教授首次提出了人工形象智能的概念(与之相对应,本文称原有人工智能为人工符号智能),提出了一系列新的智能方法、智能形式和研究内容,如:新的智能方法:综合推理、分析推理;新的智能形式:在无限空间的连续智能;新的研究内容:涉及基于各种媒体的知识表达、推理、问题求解和思维模型等方面,要解决不同媒体间的转换、视觉推理、分解与综合等难题,其研究的计算机技术平台将降落到多媒体处理技术之上。

为此,本文将分析形象智能与多媒体处理的关系,并提出两者综合研究的方法和途径。

二、人类思维与信息处理

2.1 逻辑思维与形象思维^[1]

人类思维可以分为逻辑思维和形象思维两种基本形式^[2]。逻辑思维解决一种形式上的真理,通常通过概念、判断和推理来实现。其形式将最终归结为一些纯粹的符号,但每一个符号仍对应一定的意义;形式作用本身虽然独立于其所指的内容,却仍然是一种带有清晰意义的系统;逻辑结构表示着某种思维的规律,是一种思维运算形式的理论。为此,逻辑思维具有如下等各方面的特点:

□ 逻辑思维是一种符号性表达体系,完全忽视了人的另一种重要思维形式即形象思维;

□ 逻辑思维对于对象或是或非的二值性,使得模糊性对象的表达和操作陷于困境;

□ 逻辑思维的推理领域的字符表一经确定,问题答案的搜索空间即被界定,推理机制没有能力也不可能到给定的字符表以外去搜索、产生新的内容;

□ 逻辑思维是一种判定性思维方式,由同一律保证前后概念一致;由矛盾律保证真实判断唯一;排中律保证真实性存在;充足理由律从形式上保证结论的可论证性。

形象思维是使用形象的思维,形象即是心象,它是人脑长期记忆中的事物形象。思维的主要方式,可以有形象-形象、形象-逻辑、逻辑-形象等多种形式;

*)本文工作得到国家青年自然科学基金资助

同逻辑思维相比,形象思维具有如下等各方面的特点:

□ 形象思维具有较多的非理性成分,主观色彩较浓,常常带有深刻世界观烙印并常以经验为转移;

□ 形象思维涉及到的许多客观属性难以精确表达,使得对象空间从二态变成为连续,搜索范围极度地开阔了;

□ 形象思维通过构成结构的方法组合生成新形象对象,通过类比结构方法获得转移经验,通过时空结构方法在高层次上获得新的心象结构;

□ 形象思维具有较高的形象度、丰富的内涵,思维能够在高维空间进行,导致了思维的高度并行和发散。

2.2 思维方式与信息处理的内在联系

人类的思维方式与计算机信息处理存在着密切的内在联系,如下图所示:

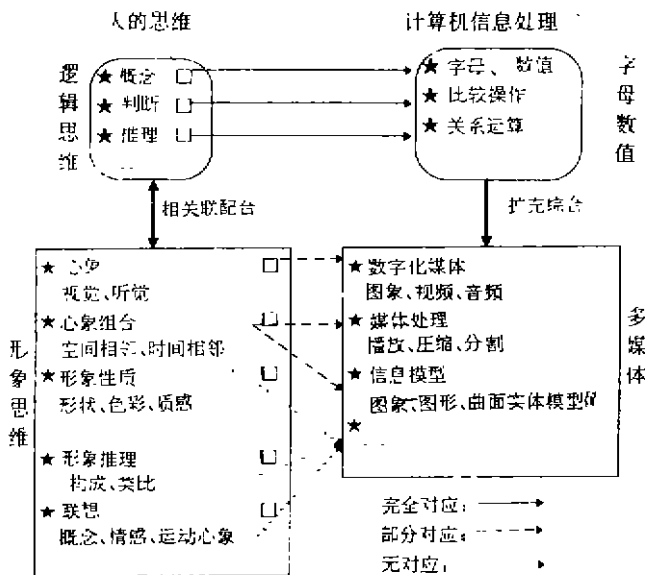


图1 人类思维与计算机信息处理的关系

图中,逻辑思维与字母数值型信息处理的关系比较对称;应用领域对象的字母数值信息,是人们对客观对象的理性和精确的反映,计算机对这些信息的处理和比较,与人脑中的理解和判断是完全一致的。如:概念:可以用字母、数值及其组合来表达;判断:能由确定结果的比较操作来体现;推理:能通过各种关系运算及其复合关系得到实现。因此,字母数

值型信息处理技术对于逻辑思维而言,是完备的。

形象思维与现有的多媒体信息处理的关系却不是完全对称的;应用领域对象的多媒体信息,是人们对客观对象的感性和模糊综合的反映;而计算机只对多媒体数字化的信息进行处理和比较,与人们对客观对象的多媒体信息的理解和判断不一致,如:心象:可以用数字化媒体来部分地表达,能刻画心象原始的内容;心象组合:能通过信息模型与媒体处理结合来部分地体现;形象的性质、推理和联想:基本上没有对应的技术或方法来实现。

形象思维与多媒体信息处理的不对称性,其根源在于两大方面的基本原因:一是,形象思维与逻辑思维相比有着本质上的区别,尽管形象思维的形式中存在着形象-逻辑、逻辑-形象两种形式,但形象思维本身不可能用逻辑思维来完全地替代;二是,现有的多媒体信息处理技术,是对传统的字母数值型信息处理技术的扩充与延伸;能对数字化的原始多媒体信息进行表达和处理,但对体现在原始信息之上的蕴含的特征却难以表达和处理。

显然,从概念模型来看,逻辑思维、人工符号智能和字母数值型信息处理之间的关系与形象思维、人工形象智能和多媒体信息处理之间的关系有着很好的对称性;而从现有的计算模型来看,却并不对称,即模拟逻辑思维智能的人工符号智能可以降落到计算机对字母数值型信息处理的技术平台上,而人工形象智能研究没有相应的完整的多媒体处理技术来支持,同时多媒体处理技术开发,没有有效利用人工形象智能研究的理论成果。上述现象已对两者的研究造成了严重的瓶颈。因此,一种解决上述计算模型不对称性的有效途径是,针对形象智能中的形象性质、推理和联想等概念和方法,研究多媒体信息处理技术,为人工形象智能理论研究提供可降落的计算技术平台。本文在第三、四节将讨论的形象智能与多媒体处理综合研究方法。

三、形象智能与多媒体处理

3.1 人工形象智能

思维科学的研究^[2],在著名学者钱学森教授的倡导下,自八十年代以来进入了研究高潮。其中,形象思维被认为是思维科学研究的突破口,并且如何

将研究成果降落到可计算的计算机技术平台上,成为应用发展的关键。

同时,认知科学中关于感知、记忆与思维的研究成果构成了人工智能发展的基础。如人工智能中的各种知识表达,产生式、语义网络和框架等都直接取材于认知心理学中的记忆结构分析模型,从而形成迄今的人工智能理论的内核。但人工符号智能研究主要是针对符号处理机制,对形象信息难以处理。而人工形象智能是思维科学和认知科学综合研究的最新方向,不仅能为思维科学的研究成果寻找计算上的着落点,并且将拓展人工智能的研究内涵。

思维科学/认知科学的研究表明,人的创造性思维与心象是密切相关的。因此,人工形象智能的研究将围绕心象这一核心问题而展开。目前比较有代表性的理论有如下几类^[3]:

□ 心象计算理论。依据心象与知觉在功能上的类似,将心象的表征分为两类:表层表征和深层表征,前者是指出现在视觉短时记忆中的类似图画表征;后者为存储在长时记忆中的信息,它又被分为两:本义表征和命题表征。

□ 心象的付立叶模式。付立叶模式认为:人脑中长时记忆存储的是景物形状的付立叶变换形式,而不是景物形状的原形,心象在人脑中是以原景物的付立叶变换后的信息存储的。

□ 区域生长法。区域生长论者把视觉性质相同的点集看作区域,心象在人脑中存储的是这些分割成区域的信息。

□ 心象的纹理表示理论。纹理是一种表面特性,由紧密交织在一起的单元组成,人们在进行认知活动时,倾向于把大小变化的纹理与三维物体的表面联系起来,用它来表示三维深度等,纹理表示论正是建立在这一事实基础上,认为心象是通过纹理分析来获得和加以表征的。

□ 心象的形态学理论。其设想是,视觉心象是由各种形态及扩张、生长、侵蚀等操作构成,视觉感知信息是以形态学方法加工,然后在大脑中按形态学方法认知。

□ 形象信息模型。将视觉心象分成如下三部分结构:1)构成结构,指一个表象由若干小的表象结合而成的;2)类比结构,指一个表象由另外的表象通过类比而形成的;3)时空结构,指一个表象与另外一些在时间上和空间上相接近的其他表象之间是有密切联系的。

3.2 多媒体信息处理

目前,对多媒体信息处理的研究与应用,均围绕多媒体信息特征这一核心问题而展开。从研究现状

来看,有代表性的工作涉及如下等几个方面:

□ 特征表达研究。特征表达一般是多媒体数据模型研究的一部分,往往与多媒体信息的检索、查询结合在一起考虑。如,H. Ishikawa 等^[4]提出了一种集成化的多媒体数据模型,包含结构、时序、空间和控制上的特征;G. Yihong 等^[5]针对通用多媒体信息对象的复杂性,采用“色彩分布”、“形象”特征来捕获内在的图象性质,以便减少搜索空间;C. H. C. Leung 等^[7]提出了一种以“FACTS”为基本特征元素的数据模型,以便对图象的内容进行系统化的表达,支持基于内容的检索和查询。

□ 特征获取研究。特征获取是一个瓶颈问题,很难有一种通用的解决方法,往往均是领域或个性相关的。T. Kurita 等^[6]针对多媒体信息形象特征的个性差异,研究了基于多变量数据分析的个性印象学习算法,使获取的多媒体信息的特征能较好地反映个性;S. Deborah 等^[8]提出了基于知识处理的特征获取模型,利用领域相关知识来获取多媒体信息特征。

□ 特征操作研究。特征操作一般是多媒体查询语言研究的一部分,其目的是为了解决任何应用系统中多媒体信息操作问题,仍是一个至今没有完整解决方案的问题。P. Bose 等^[10]提出的模糊 DB 查询语言 SQLF,引入了在模糊集合理论下可以解释的灵活谓词;Qi Yang 等^[11]提出基于加权的欧氏距离相似度量操作的查询语言;Hirzalla 等^[12]提出了一种多媒体查询语言,利用特征所刻画的时序及空间关系操作。

□ 特征管理研究。特征管理是特征存储管理技术开发的基础。当多媒体信息达到一定规模之后,选择合适的访问结构和方式将显得非常的重要和迫切。这种访问机制使与用户的检索请求最相关的多媒体信息从二级或三级空间中查询出来。访问机制将充当一种过滤器,能够消除不相关的媒体信息同时保证相关信息不被遗失,使交互信息检索模块中的查询处理算法只对通过过滤器的信息实施检索操作。现存的大多数访问机制都基于空间访问方式,且假设多媒体的特征均是数值的并将多媒体信息看作为多维空间中的一点。事实上,并不是所有特征都是数值的,为此需要新的访问结构和方法。

□ 特征学习研究。特征学习是开发有效的交互信息检索技术的基础。用户主观性和印象往往与确定和解释主观的属性联系在一起,因此,检索处理器必须设计成为能交互地处理上述问题。查询和检索接口将设计成一个基于知识的系统,通过检索描述说明过程来指导用户,支持用户相关的反馈和增

量式的检索重设计。

此外,目前多媒体信息特征研究的一种重要趋势是与知识处理和模糊理论相结合^[4]。基于知识处理模型的研究,主要围绕基于领域知识的多媒体信息处理、特征抽取,多媒体领域对象的表达以及辅助判断多媒体信息对象的相似性;基于模糊数据理论的研究工作将模糊数据理论应用于多媒体数据查询,T. Gevers 等在称作为 MSTERDAM 的系统中定义模糊操作实现存在细微差异的图像之间的匹配,E. Binaghi 等在色彩图像的处理中,应用模糊集合理论来表达人脑在判断色彩相似性时的不确定性。

四、综合研究的途径与方法

4.1 心象和多媒体特征

所谓心象,是一个富有特色的心理过程,包括通常所说的记忆心象和想象心象。它是在对同一事物或同类事物多次感知的基础上形成的,具有一定的间接性和概括性。因为它能够以自身的不断改变来模拟外在对象的连续变化,因而通常将它看成外在对象表征的抽象类似物。

而所谓多媒体特征,是对应用领域的多媒体信息特性的综合反映,体现了人们对多媒体信息的一种理解和认识,其目标将是描述刻画多媒体信息内在语义并定义对应的操作,以解决多媒体信息检索系统开发中的瓶颈问题。

心象和多媒体特征是从两个不同层次和角度对同一客观事物的反映。因而,从广义上而言,是属于同一概念范畴。两者分别是形象智能和多媒体处理技术的核心问题,迄今为止,两者是独立研究发展的,心象研究起源于抽象的理论,正逐步向可计算实现技术方向发展,而多媒体特征研究,却从具体处理技术开始,逐步寻找一种通用的理论模型。两者尽管在研究途径上有很大的差异,但有比较一致的研究目标,理论技术上有着很大的兼容和互补性。因此,心象和多媒体特征的集成是人工智能和多媒体处理技术综合的基础和核心。

4.2 综合研究的途径

形象智能和多媒体处理技术的综合研究,将以心象和多媒体特征集成为基础,从如下等几个层次基础上展开。

□ 理论基础层。人工形象智能尽管没有完全成熟的理论,但以心象为基础已有较多工作,如心象计算理论、纹理表示理论和形态学理论等,而多媒体信息处理却缺少理论基础。因此,后者可以直接利用或借鉴前者的理论,作为自身的理论基础,指导特征

表达、操作和管理等方面的研究开发工作,并在研究方法上,除了结合人工符号智能的方法之外,还将结合新的智能方法。

□ 模型表达层。形象智能和多媒体信息处理均展开了许多模型表达的研究工作,一些概念在名称上尽管不一致,但在本质上表现的是同一回事,可以彼此补充和借鉴。如心象计算理论中的心象表征方法,表层表征、深层本义表征和命题表征类似于多媒体特征研究中的多层次模型;基本原始特征、对象特征和领域特征,又如心象表达中的形象信息模型,将心象分成三部分,构成结构、类比结构和时空结构,类似于 H. Ishikawa 所提出的集成化多媒体信息特征模型,包含结构、时序、空间和控制上的特征等等。

□ 技术平台层。多媒体信息处理有较好的技术基础,从媒体的播放、压缩、分割到特征的获取、操作、管理和学习均有较成熟的算法和技术。鉴于在模型表达层上共同的形式,形象智能的理论研究成果可以在多媒体处理技术的基础上,降落到可计算的技术平台之上,得到应用和实用化。

总结 人类的思维方式与计算机信息处理技术存在着密切的内在联系,模拟人类思维是人工智能研究者所追求的目标,而现有的形象智能研究和多媒体处理技术的关系却并不对称,成为研究和开发中的瓶颈问题。

形象智能与多媒体处理技术的综合研究是促进两者研究发展的良好途径。形象智能是多媒体处理技术开发的理论基础,而多媒体处理技术能成为形象智能理论研究成果的可降落的技术平台。两者的综合研究可以从理论基础、模型表达和技术平台等多个不同层次上展开。

本文工作,是形象智能与多媒体处理技术的综合研究的开端,进一步的工作将围绕上述三个方面,深入研究综合的理论模型和技术开发平台。

主要参考文献

- [1]徐冬溶,设计中形象类比生成的研究,浙江大学博士论文,1995
- [2]钱学森,《关于思维科学》,上海人民出版社,1996
- [3]耿卫东、潘云鹤,形象思维中的心象及表征理论,计算机科学 21(1)1994
- [4]H. Ishikawa et al., A next-generation database system for advanced multimedia applications, IE-ICE Trans. on Comm. Vol. E78-B Iss. 7 1995
- [5]Ramesh Jam, MM, Large Multimedia Database, SIGGRAPH93

不确定推理模型,

14

54-56

几种不确定推理模型研究*

信度网络,

张师超 覃秋梅

(广西师范大学数学与计算机科学系 桂林 541004)

专家系统

A

摘要 The two uncertain reasoning models, the possibility theory and belief network, are first discussed in this paper. Next, a point-value causality reasoning model is established. In final, these three models are compared.

关键词 Uncertain reasoning, Fuzzy relation matrix, Belief network, Conditional probability matrix, Causality.

在现实世界中,事物通常以不确定的形式存在,事物之间的关系也具有不确定性。另一方面,人们从自然现象中总结出的规律或知识是经验的,或不确定的。因此,在推理判断和决策过程中,均不可避免地涉及不确定的数据和知识的处理。现在,已建立了许多不确定的推理模型,其主要的有主观 Bayes 方法、确定性理论、可能性理论等,而最理想的推理模型应首推可能性理论。然而,由于可能性理论的基本操作是不充分的,所以尽管其计算复杂度低,但它在专家系统中的运用仅有 2% 左右。J. Pearl 的信度网络方法是主观 Bayes 方法的一个扩展,它以其坚实的概率理论基础以及其有效性被认为是目前最好的不确定推理方法^[1]。不过,信度网络方法有一个致命的弱点,它以点值的先验概率为基础,而先验概率值的获取是昂贵的。因此,可以说信度网络方法的结果是漂亮的,但实用价值不大。

本文先介绍了可能性理论和信度网络这两种方

法和我们在这方面的一些研究工作。然后对这些模型的优缺点进行述评,并提出一些观点和看法,为推进不确定推理方法的发展提供参考。

1 可能性理论

可能性理论是采用模糊集来描述不确定性的,并提供了操作模糊命题的几个基本算子。在可能性理论中,对涉及的对象或变量定义一个可能性分布。如:命题“X 是 A”,考虑的论域是 $U = (u_1, u_2, \dots, u_n)$, 则可能性分布 $\mu(A) = (a_1, a_2, \dots, a_n)$, 其中 a_i 表示 u_i 隶属于 A 的程度。

若有规则“若 X is A, 则 Y is B”, X 和 Y 分别在 $U = (u_1, u_2, \dots, u_m)$ 和 $V = (v_1, v_2, \dots, v_n)$ 中取值, 则说 X 与 Y 存在因果关系, 这样的一种限制关系可以用一模糊关系矩阵 $R = (r_{ij} / (u_i, v_j), i \leq m, j \leq n)$ 来表示。 r_{ij} 为 (u_i, v_j) 隶属于关系 R 的程度。

Zadeh 提供了两种方法来确定模糊关系矩阵 R,

*) 本文得到国家自然科学基金及广西跨世纪人才基金资助。

- [6] G. Yihong et al., An image database system with fast image indexing capability based on color histograms, Proc. of 1994 Region 10's 9th Annual Intl. Conf. Vol. 1, Singapore
- [7] C. H. C. Leung et al., Image data modeling for efficient content indexing, Proc. of intl. Workshop on Multi-Media Database Management Systems, CA, USA
- [8] T. Kurita et al., Learning of personal visual impression for image database systems, Proc. of the 2nd Intl. Conf. on Document Analysis and Recognition, CA, USA, 1993
- [9] Deborah Swanberg et al., Knowledge Guided Parsing in Video Databases, SIGGRAPH 93
- [10] P. Bose., Subqueries in SQLF, a fuzzy database query language, Proc. of 1995 IEEE Intl. Conf. on Systems, Man and Cybernetics, Vol. 4, NY, USA, 1995
- [11] Qi Yang et al., MB/sup +/- tree: a new index structure for multimedia database, Same to [7]
- [12] N. Hirzalla et al., A multimedia query specification language, Same to [7]