

# 面向视听跨媒体检索的神经认知计算模型研究

刘 扬<sup>1,2,4</sup> 涂春龙<sup>3</sup> 郑逢斌<sup>1,4</sup>

(河南大学空间信息处理实验室 开封 475004)<sup>1</sup> (河南大学环境与规划学院 开封 475004)<sup>2</sup>

(浙江大学生物医学工程系 杭州 310027)<sup>3</sup> (河南大学计算机与信息工程学院 开封 475004)<sup>4</sup>

**摘 要** 跨媒体语义映射和跨媒体语义检索是跨媒体搜索引擎的核心技术问题。对视听神经认知的功能、层次和结构进行了分析,借鉴深度信念网络和时空层次记忆模型的设计思想,建立了一种仿脑的面向视听跨媒体应用的神经认知计算模型。依据神经系统的信息处理机制和认知理论的功能架构来设计可计算模型,系统地讨论了皮层柱的视听信息整合机制和协同认知的处理流程。本模型可为解决跨媒体语义映射和跨媒体语义检索的相关应用提供借鉴和参考,对实现非冯·诺依曼结构的仿脑认知计算进行了一次有意义的探索。

**关键词** 媒体神经认知计算,跨媒体语义检索,跨媒体语义映射,认知计算,仿脑计算

**中图法分类号** TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.3.004

## Research of Neural Cognitive Computing Model for Visual and Auditory Cross-media Retrieval

LIU Yang<sup>1,2,4</sup> TU Chun-long<sup>3</sup> ZHENG Feng-bin<sup>1,4</sup>

(Laboratory of Spatial Information Processing, Henan University, Kaifeng 475004, China)<sup>1</sup>

(College of Environment and Planning, Henan University, Kaifeng 475004, China)<sup>2</sup>

(Department of Biomedical Engineering, Zhejiang University, Hangzhou 310027, China)<sup>3</sup>

(College of Computer Science and Information Engineering, Henan University, Kaifeng 475004, China)<sup>4</sup>

**Abstract** Cross-media semantic mapping and cross-media semantic retrieval are key problems of the search engine. In this paper, we analyzed the functionality, the hierarchy and the structure of the brain's neurocognitive, and established a brain-like neural cognitive computing model for visual and auditory cross-media application after taking into account the idea of deep belief network and hierarchical temporal memory. According to the mechanism of information processing in central nervous system and framework of functional approach in cognitive theories, we designed computational model, and discussed systematically information integration mechanism in cortical column and cooperative cognitive processing of visual and auditory. This model provides a reference to resolve application of cross-media semantic mapping and retrieval, and is significant exploration for brain-like cognitive computation of non-von Neumann structure.

**Keywords** Multimedia neural cognitive computing, Cross-media semantic retrieval, Cross-media semantic mapping, Cognitive computing, Brain-like computing

## 1 引言

认知科学和神经科学相关的研究证实,人脑的认知过程和多感觉神经整合具备跨媒体的特性。人脑是自然界中最复杂的系统之一,仿脑计算是人类对大脑在功能和结构上进行的建模和仿真。近年来许多国家在认识脑、保护脑和创造脑的科学研究方面做了重大投入,相关学者从微观分子水平到宏观系统水平上分别对脑进行系统的研究。2010年美国制定人脑连接组计划(Human Connection Project, HCP),2013年美国又启动大脑活动图(Brain Activity Map, BAM)计划。我国也在脑与认知科学上加强了部署。2001年中国正式成

为参与人类脑计划(Human Brain Project, HBP)与神经信息学研究的第二十个国家。2008年NSFC正式启动了“视听觉信息的认知计算”重大研究计划,该项目以无人驾驶车辆为集成验证平台,研究非结构化视听觉感知的特征提取、表达、整合、理解和海量异构多模态信息的协同计算。近期,紧跟欧美日先后启动的投资巨大的脑科学研究计划,我国的“脑科学计划”也即将启动。但从整体上说,目前脑科学关于脑的高级功能研究尚未取得突破性进展,这无疑对以仿生计算为目的的人工脑(Artificial Brain)信息处理模型的研究造成巨大的困难。

跨媒体搜索引擎的核心技术是特征提取与降维、跨媒体

到稿日期:2014-04-03 返修日期:2014-07-20 本文受国家自然科学基金(61305042,61202098),国防科技工业民用专项科研技术研究项目(2012A03A0939),河南省教育厅科学技术研究重点项目(13A520071)资助。

刘 扬(1971—),男,博士生,副教授,硕士生导师,CCF会员,主要研究领域为媒体神经认知计算、时空信息高性能计算, E-mail: ly\_sci\_art@gmail.com;涂春龙(1988—),男,博士生,主要研究领域为神经信息学、生物医学工程;郑逢斌(1963—),男,博士,教授,博士生导师,主要研究领域为空间信息处理、自然语言处理。

语义映射(Cross-media Semantic mapping, CSM)和跨媒体语义检索(Cross-media Semantic Retrieval, CSR)。跨媒体语义计算面临的“语义鸿沟”和“维度灾难”问题是目前亟需解决的核心问题。由于 CSM 和 CSR 问题本身的复杂性,认知科学的相关功能结构和神经科学的相关处理机制是媒体计算研究的重要参考,对媒体信息智能的分析和检索也同样具有重要启发作用。但二者与计算科学的研究方法和实施方案存在着本质上的不同。根据现有已知的认知信息处理机制和神经信息处理方法,建立仿脑的可计算模型(computational model),将有助于解决跨媒体语义的计算问题。

目前,对于仿脑计算的研究,主要从两个方面进行:其一是从认知功能的整体性模拟,其二是从局部微观的神经元、突触或局部网络的神经结构进行探索,而对于以基本的局部网络如何组装成复杂的高级系统功能尚缺乏有效的研究。本文在脑的神经认知的功能和结构的基础上,以深度信念网络(Deep Belief Network, DBN)、层次时间记忆(Hierarchical Temporal Memory, HTM)和概率图模型(Probabilistic Graphical Model, PGM)为蓝本,对视听跨媒体检索的媒体神经认知计算(Multimedia Neural Cognitive Computing, MNCC)模型进行了初步研究。

## 2 相关工作

### 2.1 媒体神经认知计算

在引入 MNCC 定义之前,有必要阐述媒体计算、认知计算和神经计算的相关概念。

**定义 1**(媒体计算, Multimedia Computing, MC) 指以研究视觉、听觉、触觉、味觉、嗅觉等各类感觉媒体信息的采集、表示和展示的方法和理论,探索以文本、图形、图像、音频、MIDI、视频、动画等各种表示媒体的传输、处理、内容分析与识别算法等共性的基础计算理论和应用技术。

**定义 2**(认知计算, Cognitive Computing, CC) 旨在探索大脑信息加工的机理,包括感觉、知觉、注意、记忆、语言、思维、表象、意识等认知过程,并建立可实施的仿脑计算模型和相关功能的算法。模仿大脑从经验中学习,发现不同事物之间的联系,进而实现逻辑推理和记忆等功能。

**定义 3**(神经计算, Neural Computing, NC) 指研究神经网络机制,模拟生物神经系统的结构特征和生物神经信息的处理机理,构造人工神经网络的计算模型和相关算法。

**定义 4** MNCC 旨在面向非结构化的、海量多模态的、多时空分布的多媒体信息处理的语义鸿沟和维度灾难问题,结合宏观的系统行为层面的 CC 和微观的生理机理层面的 NC,建立新一代的多媒体信息处理模型及算法。

在 CC 和 NC 基础上产生的 MNCC 可为“语义鸿沟”和“维度灾难”问题的解决提供有效的算法思想。长期以来,科研工作者在类脑机制的研究方面进行了不懈的探索。文献[1]通过抽象智能、概念代数、系统代数等建立认知信息学(Cognitive informatics, CI)理论,模拟大脑计算推理和因果分析,表达人脑的认知以及大脑中抽象信息之间的数据、信息、知识和行为之间的关系。文献[2]对大量神经科学数据集进行分析、计算和模拟,尝试发现和证明大脑的核心算法,获得对心智如何感知、思考和行动的深刻的科学解释,进一步提出

全新的 CC 方法,并企图建立一种全新的认知系统、计算体系结构、编程范式和实用的智能设备。相关研究中几个重要的方向分别是:人工神经网络(Artificial Neural Network, ANN)、HTM、DBN 和 PGM。

#### (1) ANN

自 1943 年 McCulloch 和 Pitts 提出神经元模型以来,ANN 发展几经波折。1958 年 Rosenblatt 提出的感知机本质上是一个线性分类器;采用 back propagation 算法的多层网络(MultiLayer Perceptron, MLP)在可求解线性不可分问题时,随着层数增多容易陷入局部最优,难以解决过拟合问题;后期出现的反馈神经网络(如 1982 年的 Hopfield、SOM 网络、1988 年 Grossberg 的 ART 网络等),很多复杂的认知现象(如联想、记忆等)都可以用反馈神经网络进行模拟和解释;其后还有 Neocognitron<sup>[3]</sup>、Convolutional 网络、HMAX 模型等新型 ANN 出现;1992 年 Vapnik 等提出了支持向量机(Support Vector Machines, SVM),其利用核方法把非线性问题转换成线性问题。SVM 本质上也是一特殊的双层神经网络,具有高效的学习算法,无局部最优问题,但对非线性问题没有通用解决方案,必须谨慎选择核函数来处理。ANN 与 SVM 都是统计学习的代表方法。

生物神经网络在结构上具备反馈和时态特性,是一种有向的回路网络。ANN 是生物神经网络的一种简化,其训练方法可分为监督学习、非监督学习、半监督学习和增强学习等。基于系统的角度理解 ANN 学习本质是动力学问题,而基于功能的角理解 ANN 学习本质又是一种优化问题。

#### (2) DBN

Hintont 认为简单的有监督学习无法解决复杂机器学习问题,提出建立通过神经网络抽取模式结构为目标,外在环境在神经网络中形成内在表示的机理作为发展神经网络的基础。2006 年 Hinton 用一种具有双层结构的受限玻尔兹曼机(Restricted Boltzmann Machine, RBM)和贝叶斯信念网络构造 DBN。文献[4,5]认为无论是皮层结构还是认知过程都是具备深度架构的,进而提出一种深度学习(deep learning)的快速算法。

DBN 在做有监督学习前要先做非监督学习,然后将非监督学习学到的权值当作有监督学习的初值进行 BP 算法训练。目前深度学习算法在众多应用领域取得了前所未有的成果,微软研究院发现深度网络在 Switchboard 标准数据集的语音识别错误率比最低错误率降低了 33%<sup>[6]</sup>;谷歌实验室发现深度网络在 ImageNet 数据集上得到的图像分类准确率比当前最好的结果提高了 70%<sup>[7]</sup>。

#### (3) HTM

鉴于中枢神经系统的复杂性,通常企图寻找一种通用的模型往往是不现实的。神经网络不仅需在功能上模拟,而且要在机理上模拟。2006 年 Hawkins 和 George 在研究皮层微电路的基础上提出 HTM 模型。文献[8,9]描述了在 HTM 中贝叶斯置信度的传播理论,并以此作为设计大脑功能柱的皮质微电路(cortical micro-circuits)的数学模型。

类似于 RBM,构造 HTM 的节点也采用双层结构,不同的是 HTM 节点同时考虑了空间和时间局部性和层次性。HTM 计算思想来源于 Pearl 的贝叶斯信念网所采用的近似

推理的信念传递(belief propagation)算法,但目前尚无有效的模式学习算法。为理解大脑的行为复杂性、动态性和可塑性,了解信息在大脑传递和处理方式,文献[10]成功地发现并绘制了猕猴大脑白质的长距离神经通路(Long-Distance Pathways, LDP)。文献[11]采用扩散光谱成像技术构造了人类大脑皮层的 66 个脑区的解剖分区网络和 998 个感兴趣区域(ROI)的解剖分区网络。作者发现脑皮质解剖网络并不符合无尺度网络的小世界分布特征,而符合指数分布特征,并发现在左右脑半球中间的中后部存在一个高度密集连接的网络核心。文献[12]用内存模拟突触,用核心芯片内通信总线模拟轴突,用内核的集成和计算模拟神经元,构造一种非冯·诺依曼的认知计算架构来模拟大脑计算功能。文献[13]用 Nengo 构建一个大规模的语义指针架构统一网络(Semantic Pointer Architecture Unified Network, Spaun)的模拟大脑的绘画、计数、回答问题和推理任务。

#### (4) PGM

基于概率的角度理解 ANN 的学习本质上又是个图模型问题。PGM 有其完善的理论体系,PGM 可将概率统计应用于复杂领域进行不确定性推理和数据分析<sup>[14,15]</sup>。常用的 PGM 分为无向图结构的马尔可夫随机场(Markov random field)和有向无环图结构的贝叶斯网络(Bayesian network)两大类。PGM 每个节点只与数量有限的其它节点发送直接的联系,局部结构的稀疏链接具备小世界特性。PGM 近似推理常见方法有随机抽样(MCMC Gibbs sampling)算法和信念传递算法<sup>[16]</sup>。

Hintont 用 RBM 构造 DBN 的提供无监督模式学习算法与 Hawkins 用 HTM 构造皮层算法相比,二者都可构造分类器进行无监督学习,且可堆叠形成带反馈层次结构。但 RBM 没有充分利用时空的局部性,而 HTM 更加模块化地利用了时空局部性和层次性,HTM 的模式学习是基于 PGM 的信念传递算法。从数学形式上说,DBN 和 RBM 都可看成是 PGM 的特例。

### 2.2 CSM 与 CSR

跨媒体搜索引擎是利用 CSM 与 CSR 技术,在网络多媒体数据库中实现不同模态媒体的基于语义相似性和相关性的一种信息检索技术。

设多媒体数据库定义为  $MD = \{M_{S_1}, \dots, M_{S_i}, \dots, M_{S_n}\}$ , 其中  $M_{S_i}$  是语义为  $S_i$  相关的不同模态媒体集合。 $M_{S_i}$  定义为  $M_{S_i} = \{T_{S_i}, I_{S_i}, V_{S_i}, A_{S_i}\}$ , 其中  $T_{S_i}, I_{S_i}, V_{S_i}, A_{S_i}$  分别是语义为  $S_i$  相似的相同模态的文本、图像、视频和音频的媒体集合, 定义为:  $T_{S_i} = \{T_{S_{i,1}}, \dots, T_{S_{i,X}}\}$ ,  $I_{S_i} = \{I_{S_{i,1}}, \dots, I_{S_{i,Y}}\}$ ,  $V_{S_i} = \{V_{S_{i,1}}, \dots, V_{S_{i,Z}}\}$ ,  $A_{S_i} = \{A_{S_{i,1}}, \dots, A_{S_{i,P}}\}$ ,  $X, Y, Z, P$  为语义  $S_i$  相似的媒体个数。另有语义对象  $SO$  定义为  $SO = \{C_{S_1}, \dots, C_{S_i}, \dots, C_{S_J}\}$ , 其中  $C_{S_i}$  为语义对象集合, 定义为  $C_{S_i} = \{W_{S_{i,1}}, \dots, W_{S_{i,j}}, \dots, W_{S_{i,V}}\}$ ,  $W_{S_{i,j}}$  为语义概念。

**定义 5(CSM)** 在语义相似的同模态的媒体和语义相关的跨模态媒体的语义特征中寻找对应语义对象的一个过程。CSM 即实现给定媒体  $M_{S_i}$ , 从  $SO$  中找出最合适描述  $M_{S_i}$  的语义对象集合  $C_{S_i}$ , CSM 的形式化定义为:

$$CSM: ArgC_{S_i} \text{Max}(M_{S_i} \rightarrow SO) \quad (1)$$

其中,  $M_{S_i}$  是包含语义概念  $W_{S_{i,j}}$  所有的文本  $T_{S_i}$ 、图像  $I_{S_i}$ 、视

频  $V_{S_i}$  和音频  $A_{S_i}$  的媒体集合。例如将图灵奖相关的文章、图像或视音频等媒体映射到“图灵奖”这个语义对象中。

**定义 6(CSR)** 利用多媒体信息的数据挖掘技术, 从非结构化的多媒体信息中检索出语义相似的同模态的媒体和语义相关的跨模态媒体的一种信息检索技术。CSR 过程即给定语义概念  $W_{S_{i,j}}$ , 从  $MD$  中找出包含概念  $W_{S_{i,j}}$  的媒体集合  $M_{S_i}$ , CSR 可由下面三元组描述:

$$CSR = \langle MD, C_{S_i}, Rank(W_{S_{i,j}}, M_{S_i}) \rangle \quad (2)$$

其中,  $M_{S_i}$  是包含语义概念  $W_{S_{i,j}}$  所有的文本  $T_{S_i}$ 、图像  $I_{S_i}$ 、视频  $V_{S_i}$  和音频  $A_{S_i}$  的媒体集合,  $Rank$  是一个排序函数。例如, 提交关键词“图灵奖”, 系统能检出所有关于“图灵奖”的文章、图像、视音频, 甚至动画等多模态的相关信息。实现上述两个过程需要构造一个合理的媒体语义分类模型, 并在一个训练集合  $TS = \{\{W_{1,1}, M_1\}, \dots, \{W_{1,v}, M_1\}, \dots, \{W_{k,1}, M_k\}, \{W_{k,u}, M_k\}\}$  中进行学习和推理。

CSR 的发展主要经历基于关键词的文本信息检索、基于内容的单模态媒体相似性检索、基于语义的跨媒体相关性检索 3 个阶段。有关 CSR 方法基本上采用文本共生与文本注释、媒体交叉关联图、语义概念等模型。相关研究内容主要集中在媒体低层信息的高维索引描述、高层信息的语义挖掘、跨模态异维信息的相关性比较以及基于人机交互的相关反馈对检索性能的改进等几个方向。在基于语义的 CSR 中, 深度学习和统计学习是近期相关研究热点, 文献[17,18]提出多模态深度学习方法实现跨模态视音频分类; 文献[19,20]采用多模态概率主题模型, 从概率统计的角度来进行多模态数据的语义分析和检索。

## 3 视听信息神经处理架构

### 3.1 长距离神经通路

人类中枢神经系统由白质(white matter)、灰质(grey matter)、黑质(substantia nigra)等组织组成。灰质的新皮层(neocortex)结构类似于一种混合的数模电子处理单元, 同时具备线性和非线性特征。白质的 LDP 构造了复杂的神经网络信息处理的布线图。

脑的功能与结构是目前最为复杂的系统之一, 一般认为大脑的新皮层是逻辑智能处理的重要部分, 丘脑是信息进出及选择性注意控制开关, 海马体及边缘系统是记忆及情绪控制器。不同学科采用不同方法探索大脑的工作机制: 神经科学采用类似于白盒方法, 在皮层结构及神经通路层次上研究神经信息的处理机制; 认知科学采用类似于黑盒方法, 利用认知功能及现象分析大脑的信息处理模型; 计算科学则在图灵机模型基础上实现有限状态机的数学逻辑运算。然而图灵机的结构和大脑有着本质的不同。自 MP 模型提出至近期的 HTM 及 DBN, 人们一直在探索如何利用神经系统的认知处理机制实现复杂信息的计算功能。一般认为大脑的神经系统结构与认知功能是一种同构关系, 二者反映了大脑的不同角度的信息。

**假设 1** 通过建立相关计算模型  $C$ , 可实现神经结构(过程)  $\Phi$  和认知操作(功能)  $\Psi$  之间的映射关系, 即:

$$C: \Psi \Leftrightarrow \Phi \quad (3)$$

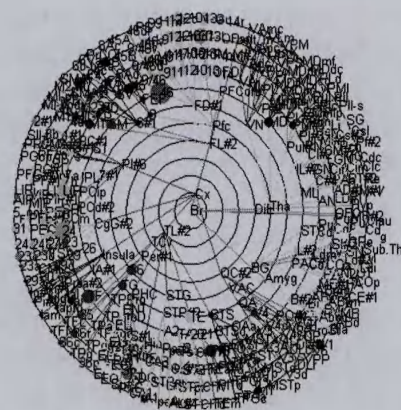


图1 大脑神经通路和环路层次结构

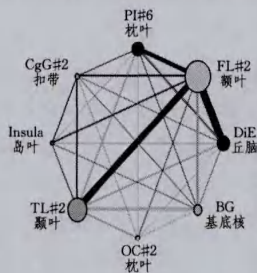


图2 全脑8区连接简图

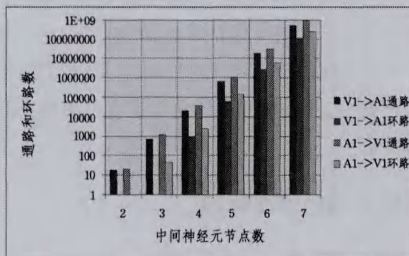


图3 A1, V1 神经通路和环路数与中间神经元数量关系

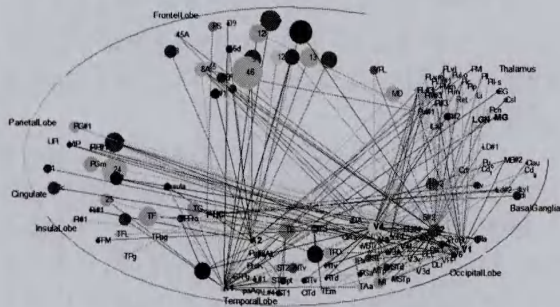


图4 大脑视觉和听觉中枢神经网络情况

根据 CoCoMac 数据库<sup>[21]</sup>对大脑长距离神经通路数据进行数据挖掘和分析,图1显示了神经通路连接情况。图2显示了脑的8个分区皮层通路关系。图中连线粗细表示连接数量,节点大小表示神经数据库中位于本区节点的数量及度。由图可见,大脑区域之间基本都有通路,但通路数量和节点的度有着明显区别。统计发现听觉基本区 A1 和视觉基本区 V1 之间存在大量的神经通路和环路,且其数量的对数与中间神经元个数呈线性关系,如图3所示。环路结构是相关反馈、随机共振,甚至是递归迭代的重要物质基础。皮质功能的层次结构及其之间的神经通路和环路结构可为构造神经认知模

型提供重要探索依据。相关研究指出,中枢神经网络是一种无尺度和小世界特性的复杂网络。图4显示了视觉 V1 区和听觉 A1 区间隔2个区域、间隔节点不超过6个神经元的神经通路连接情况。视听功能区之间的这些神经通路是大脑进行视听协同认知的物质基础。

### 3.2 皮层模型

大脑皮层有古皮层、旧皮层和新皮层3种类型。古皮层分层不清晰,旧皮层由3层神经元构成,而占90%面积的新皮层由6层神经元组成的。在功能上新皮层又可分为初级区、次级区和联络区3个级区,各区层级结构如图5所示<sup>[22]</sup>。新皮层的6个层次承担的功能各不相同:L4接收传入信息,L2和L3层构成局部回路进行信息加工,L1层实现内部神经元投射信息的交汇抑制,L5和L6则解决信息的输出。

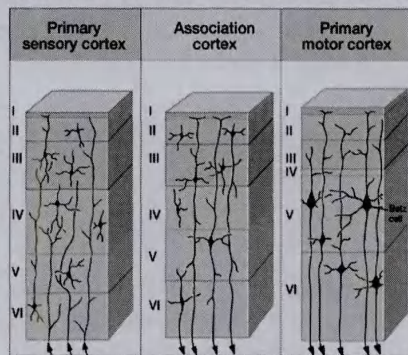


图5 皮层6层结构示意图<sup>[22]</sup>

大量研究表明,在视觉、听觉以及联合区的皮质结构是基本相似的。皮层柱是新皮层信息处理的基本单位,且皮层柱之间存在分层处理和侧抑制现象。皮层的局部回路(local circuit)构成微柱(micro-column)。视听觉感受细胞的基本功能是对特定物理信息感知并将其编码为神经冲动,而微柱的基本功能是实现特征检测,由微柱构成巨柱(macro-column)或超柱(super-column)来完成特定信息处理功能。同一层次的微柱之间传递发放确定性信息的概率,不同层级的微柱聚合下层邻近信息,向上层邻近辐射传递置信度,同时接受远距离通路的确定性反馈和上层邻近的指导性的反馈预测。一个皮层微柱信息处理和传递功能可用图6描述。

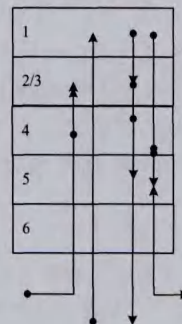


图6 皮层微柱结构示意图<sup>[22]</sup>

### 3.3 微柱的时空节点结构

假设2 鉴于新皮层结构的相似性,大脑各区的新皮层可假定其信息处理机制是通用的,视听信息可采用统一的皮质功能建模,并可将其运用到学习、推理、预测等多种问题中。

为简化模型设计,我们首先将图7(a)所示皮层的微柱6层结构合并成如图7(b)所示的3层结构,以L4层构成的中

层实现信息输入,以 L1-L3 层构成的上层实现信息处理,以 L5 和 L6 层构成的下层实现信息输出。考虑到皮层结构的时空信息处理的事实,将其进一步处理简化为如图 7(c)所示的 2 层结构。需要指出的是,这种简化不会丧失仿生学优势。事实上,HTM 节点、RBM 节点和 SVM 都采用双层结构。其中 S 实现皮层 L1-L4 层的空间模式的存储和处理,T 实现皮层 L5-L6 层时间模式的存储和处理,S 和 T 节点内部分别存储来自本节点和其他节点传递的置信度。

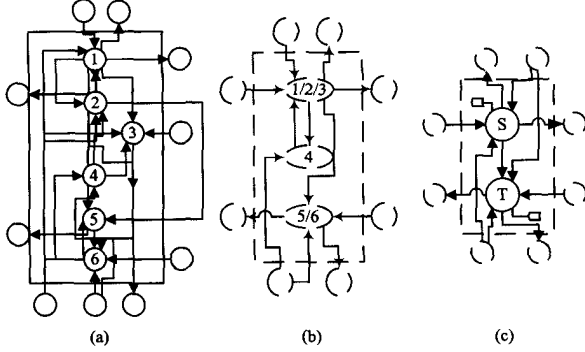


图 7 微柱的分层结构模型

### 3.4 超柱的层次网络结构

根据神经认知系统的时空层级结构和时空局部性,超柱结构采用多级分层、逐级双向映射结构。由微柱复合组成的超柱采用“同层协同”和“分层加工”原则进行构造,超柱结构如图 8 所示。

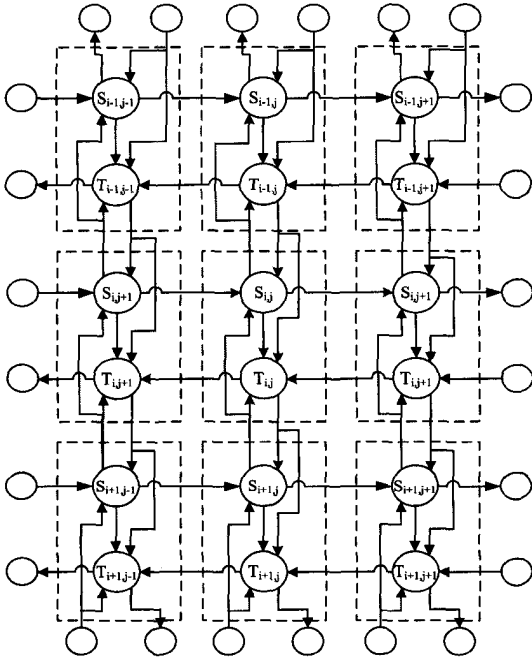


图 8 超柱结构模型

此外,丘脑与皮层存在大量连接,丘脑中继核团实现感觉信息的选择和注意,丘脑联络核团实现皮层处理信息交换。根据神经认知功能,图 9 示出了皮层-丘脑结构的视听协同认知的神经信息处理通路。其中输入层为视觉和听觉特征检查器,如亮度、边缘、音调、声强等。中间两层组成超柱模拟初级视听感觉区功能,上面两层模拟视听次级区功能,在视听次级之间加入 3 层模拟丘脑和联络区多模式感觉区功能。

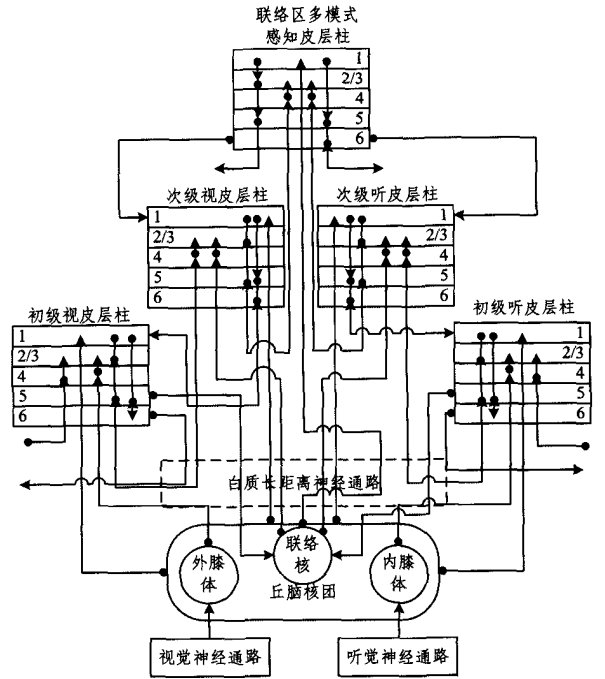


图 9 皮层-丘脑架构的视听协同认知神经信息处理通路示意

## 4 视听信息认知处理流程

### 4.1 皮层柱的置信度传递

皮层柱接收和传递信息内容如图 10 所示。定义皮层柱接收来自下层和同层的上行置信度为  $\lambda_{yx}$ ,接收来自上层和同层的下行置信度为  $\pi_{ux}$ ,自身确定性信息为  $Bel(X)$ ,发送下层或同层的上行置信度为  $\lambda_{xu}(X)$ ,发送来自上层或同层的下行置信度为  $\pi_{xy}(X)$ 。

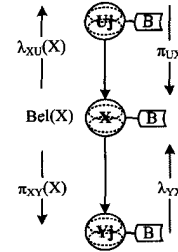


图 10 皮层柱接收及传递信息定义

皮层柱时空传递信息流程如图 11 所示。一个皮层柱节点的时空置信度  $Bel(X)$ 由同一时刻不同区域的空间位置相关的置信度  $Bel(S)$ 和先后传入时间相关的置信度  $Bel(T)$ 两部分决定,时空置信度可用函数  $Bel(X) = Bel(Bel(S), Bel(T))$ 形式化描述。

其中皮层柱空间置信度  $Bel(S)$ 接收皮层柱空间节点  $S_{i,j}$ 的来自下级皮层柱和邻近柱传入的信息  $\lambda(S)$ 和皮层柱空间节点的来自上级皮层柱和邻近皮层柱传入的信息  $\pi(T)$ ,同时向邻近节点发送向上级皮层柱和邻近柱传出的信息  $\lambda_s(u)$ 和皮层柱空间节点向下级皮层柱和邻近柱传出的信息  $\pi_s(y)$ ,设  $NodeS = \langle \lambda(S), \pi(T) \rangle$ ,  $BelS = \langle Bel(S), \lambda_s(u), \pi_s(y) \rangle$ ,则空间置信度  $Bel(S)$ 可用映射  $F: nodeS \rightarrow BelS$ 形式化描述。

相对应的皮层柱时间置信度  $Bel(T)$ 接收皮层柱时间节点  $T_{i,j}$ 的来自下级皮层柱和邻近柱传入的信息  $\lambda(T)$ 和皮层柱时间节点的来自上级皮层柱和邻近柱传入的信息  $\pi(T)$ 。同时向邻近节点传递向上级皮层柱和邻近皮层柱传出的信息

自底向上逐级进行无监督的输入信息  $A$  和  $V$  处理和预加工, 获得 MNCC 初始参数。

定义媒体的时空模式 Pattern、皮层柱对媒体对象的置信度 BeliefO、视听对象的注意度 Attention、视听对象情绪控制值 Emotion、视听对象的时态存储值 Memory, MNCC 模型可由下面的六元组描述:

$$MNCC = \langle Pattern, Attention, BeliefO, Emotion, Memory, Bel(X) \rangle \quad (4)$$

其中,  $Bel(X)$  是皮层柱时空置信度, 满足映射  $\langle Pattern, BeliefO, Memory \rangle \rightarrow Bel(X)$ 。基于 MNCC 的视听信息处理流程具体分解为皮层柱置信度计算、丘脑注意控制、边缘系统情绪控制、海马体时空缓存 4 部分, 如图 13 所示。

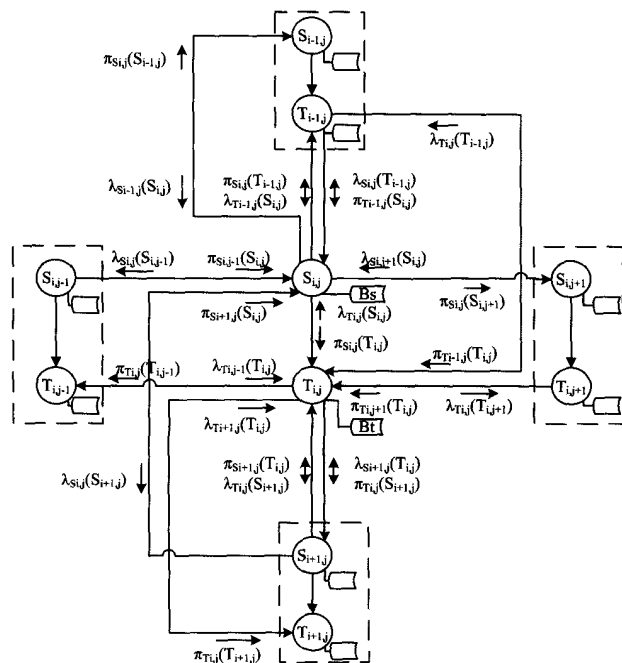


图 11 3 层结构皮层柱之间信息接收及传递计算

具体而言,可参照有向图的 Belief Propagation 算法、无向图的 Sum Product 算法,以及 Junction Tree 等 PGM 算法思想推导计算。

#### 4.2 基于 MNCC 的视听信息处理流程

根据神经系统结构及认知处理过程,大脑的视听协同认知信息处理流程可用图 12 进行整体描述。图中矩形节点是用于记忆和推理逻辑控制的层级概率网络(probability network),六边形节点是用于情绪、情感控制和优化控制的回报网络(payoff network)。

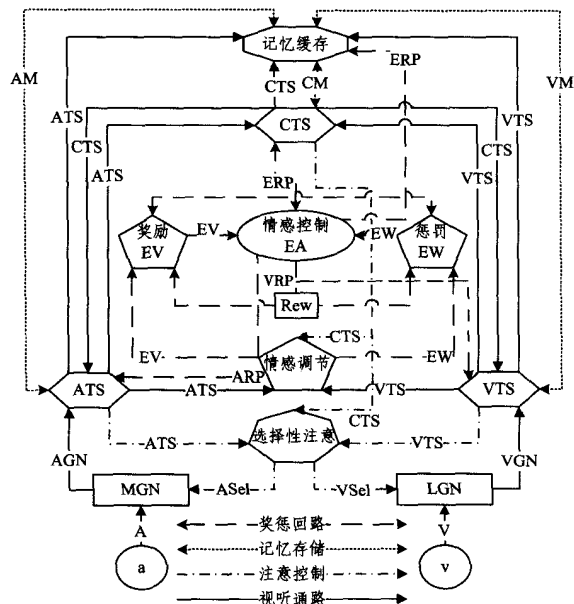


图 13 视听神经认知跨媒体语义计算流程

### (1) 皮层柱置信度计算

视觉媒体  $v$  和听觉媒体  $a$  的蕴含语义信息的归一化形式为  $V$  和  $A$ 。定义视觉媒体  $V$  的时空模式  $VGN$ 、听觉媒体  $A$  的时空模式  $AGN$ 。设  $Media = \langle A, V, Attention, s, t \rangle$ , 在经过视听神经通路的  $MGN$  和  $LGN$  初级处理后, 媒体的时空模式  $Pattern \in \{AGN, VGN\}$  由映射  $F: Media \rightarrow Pattern$  形式化确定。

设视皮层置信度  $VTS$ 、听皮层置信度  $ATS$ 、视听融合置信度  $CTS$  分别表示皮层柱对媒体对象的置信度, 则  $Cortical = \langle Emotion, Bel(X) \rangle$ , 媒体对象的置信度  $BeliefO \in \{ATS, VTS, CTS\}$  由映射  $F: Cortical \rightarrow BeliefO$  形式化确定。

## (2) 丘脑注意控制

通过邻近媒体的时空紧致性和时空上下文设计丘脑皮层架构对视听对象的注意力,对象显著度与选中概率成正比。设视觉对象的注意力  $VSel$  和听觉对象的注意力  $ASel$  由  $Thalamus = \langle Pattern, s, t \rangle$  决定,视听对象的注意力  $Attention \in \{VSel, ASel\}$  满足映射  $F: Thalamus \rightarrow Attention$ 。

### (3) 边缘系统情绪控制

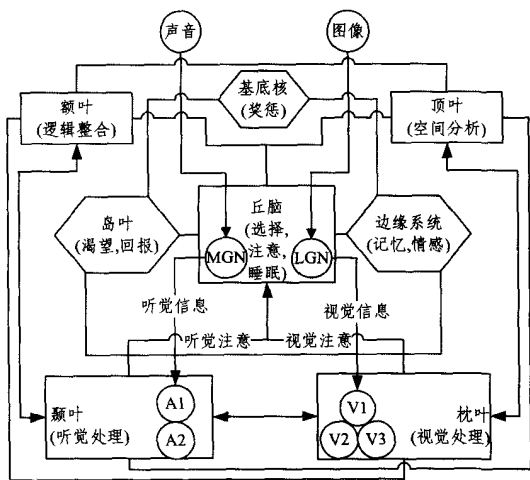


图 12 视听协同认知信息处理流程示意图

基于 MNCC 的视听信息处理流程整体策略采用逐级训练, 分层强化。处理具体采用清醒态预学习 (Pre-Learning in Waking state, PLW) 算法和睡眠态精调整 (Precisely Adjust in Sleeping state, PAS) 算法两步进行:

(1) PLW 算法模拟清晰状态,在情绪和记忆控制下,实现

情绪控制值描述了边缘系统杏仁核、眶额叶皮层柱对情感的导向奖惩控制信息,且三者相互制约,共同组成边缘系统的杏仁核和眶额叶皮层的情感控制回路,并控制海马体暂存信息。定义视觉对象情绪控制值  $VRP$ 、听觉对象情绪控制值  $ARP$ 、视听复合对象情绪控制值  $ERP$ 。设边缘系统由七元组  $Limbic = \langle Belief, t, Rew, Pattern, EV, EW, EA \rangle$  构成,其中  $EV, EW, EA$  分别为奖励值、惩罚值和情感控制值,  $Rew$  为激励信号,对象情绪控制值  $Emotion \in \{ARP, VRP, ERP\}$  满足映射  $F: Limbic \rightarrow Emotion$ 。

#### (4) 海马体的时空计算

长期记忆中,根据海马体突触前和突触后活动情况随时间变化的长时程增强(Long Term Potentiation, LTP)和长时程抑制(Long Term Depression, LTD)效应,海马体对视觉对象时态存储值  $VM$ 、听觉对象时态存储值  $AM$ 、视听复合对象时态存储值  $CM$  根据 STDP (Spike-Timing Dependent Plasticity) 模型进行计算。设  $Hippocampus = \langle Emotion, s, t \rangle$ ,  $Memory \in \{AM, VM, CM\}$ , 二者满足映射  $F: Hippocampus \rightarrow Memory$ 。

### 5 MNCC 在跨媒体搜索引擎中的应用

一个基于 MNCC 的跨媒体搜索引擎的原型结构如图 14 所示。本搜索引擎是高分辨率遥感数据检索和多模态研学知识检索平台的一个重要功能。跨媒体搜索引擎主要由搜索器、分析器、索引器、检索器、媒体库、特征库、索引库和用户接口等组成,其核心是分析器、索引器和检索器。

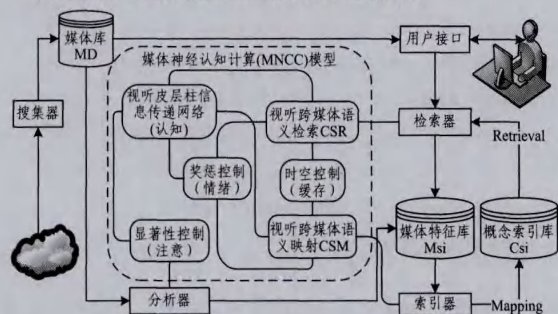


图 14 基于 MNCC 的跨媒体搜索引擎结构

其中搜索器利用网络爬虫抓取网络多媒体信息,并将其存放于媒体库 MD。分析器利用 MNCC 模型对媒体库提取媒体特征并入库媒体特征库  $M_{si}$ 。索引器采用视听跨媒体语义映射 CSM 建立语义概念的索引和语义概念倒排文件库  $C_{si}$ , 实现媒体的语义标注,解决查询关键词的快速检索。检索器接受用户交互请求,利用视听跨媒体语义检索 CSR 实现跨媒体的扩展查询,并通过用户接口对检索结果排序展示。

在显著性控制下,分析器选择注意度高的媒体对象,训练视听皮层信息传递网络,实现媒体语义特征的提取工作。如图 15 所示,根据视听通路神经元个数及皮层信息处理特征,设计由皮层柱构成的视听信息处理网络结构。在视觉处理和听觉处理通道底层微柱个数分别采用逐级递减的 3 层和 4 层,视听协同处理采用 3 层微柱构成。此外文本采用概率主题模型(Probabilistic Topic Models, PTM)进行处理。

在皮层柱网络的媒体结构化分析的基础上,利用情绪控制下的增强学习和语义时空上下文,及索引器和检索器分别实现语义相似性标注和语义相关性匹配功能。图 16 显示了基于 MNCC 跨媒体搜索引擎原型系统在多模态研学知识检索平台的应用情况。

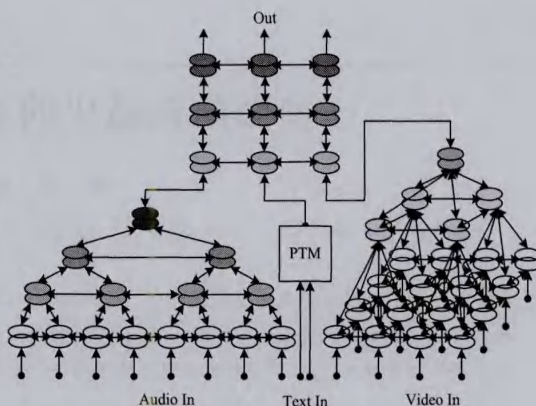


图 15 由皮层柱构成的视听信息传递网络结构

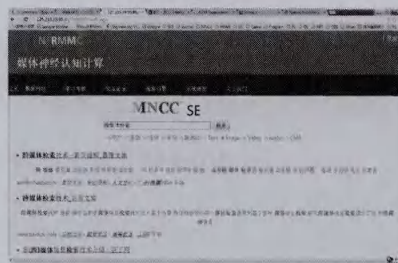


图 16 基于 MNCC 跨媒体搜索引擎原型系统

**结束语** 本文对中枢神经系统有关视听皮层神经认知、皮层-丘脑选择性注意控制结构、边缘系统情绪控制及海马体记忆增强机制进行系统的建模,设计了面向视听跨媒体检索的 MNCC 计算模型。该模型以 DBN、HTM、PGM 及分层强化学习理论为理论基础,可为 CSM 和 CSR 相关应用提供可行的建设方案。由于大脑的神经认知机制的复杂性,本文仅依据神经科学和认知科学相关理论和数据设计模型,对 MNCC 进行了初步的探索。下一步需结合深度学习理论、概率认知理论及现代神经科学的发现,进一步完善和优化基于 MNCC 的跨媒体语义搜索引擎原型相关算法。

### 参考文献

- [1] Wang Ying-xu. On cognitive informatics [J]. Brain and Mind, 2003, 4(2): 151-167
- [2] Dharmendra S. Modha, Rajagopal Ananthanarayanan, et al. Cognitive Computing [J]. Communications of the ACM, 2011, 54(8): 62-71
- [3] Fukushima K. Neocognitron trained with winner-kill-loser rule [J]. Neural Networks, 2010, 23(7): 926-938
- [4] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554
- [5] Arel I, Rose D C, et al. Deep Machine Learning-A New Frontier in Artificial Intelligence Research [J]. Computational Intelligence Magazine, IEEE, 2010, 5(4): 13-18
- [6] Dahl G, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition [J]. IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1): 30-42
- [7] Le Q, Ranzato M, Monga R, et al. Building high-level features using large scale unsupervised learning [C]// The 29th International Conference on Machine Learning (ICML), 2012. Edinburgh, 2012: 81-88

(下转第 30 页)

- [8] Hardekopf B, Lin C. Semi-Sparse Flow-Sensitive Pointer Analysis[C]// Proceedings of the 36<sup>th</sup> Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL'09). ACM, 2009; 226-238
- [9] Yan Da-cong, Xu Guo-qing, Rountev A. Demand-Driven Context-Sensitive Alias Analysis for Java[C]// Proceedings of the 2011 International Symposium on Software Testing and Analysis (ISSTA'11). ACM, 2011; 155-165
- [10] Li Xin, Ogawa M. An ahead-of-time yet context-sensitive points-to analysis for Java[J]. Electronic Notes in Theoretical Computer Science, 2009, 253(5): 31-46
- [11] Chris L, Lenharth A, Adve V. Making context-sensitive points-to analysis with heap cloning practical for the real world[J]. ACM SIGPLAN Notices, ACM, 2007, 42(6): 278-289
- [12] Pearce D J, Kelly P H J, Hankin C. Efficient Field-Sensitive Pointer Analysis of C[J]. ACM Transactions on Programming Languages and Systems (TOPLAS), 2007, 30(1): 4
- [13] Yu Hong-tao, Zhang Zhao-qing. An Aggressively Field-Sensitive Unification-Based Pointer Analysis[J]. Chinese Journal of Computers, 2009, 32(9): 1722
- [14] Experiences coding Non-Uniform Parallel using the CUDA GPGPU Architecture[C]// NJPLS2. August 2008
- [15] Randy A, Kennedy K. Optimizing Compilers For Modern Architectures: A Dependence-based Approach [M]. Morgan Kaufmann, 2001; 790
- [16] Samuel L, Amarasinghe S. Exploiting superword level parallelism with multimedia instruction sets[J]. Sigplan Notices-SIGPLAN, 2000, 35(5): 145-156
- [17] Transforming and parallelizing ANSI C programs using pattern recognition[C]// 7th International Conference, HPCN Europe 1999 Amsterdam, 1999; 673-682
- [18] Ralf K, Hack S. Whole-function vectorization. Code Generation and Optimization[C]// 2011 9th Annual IEEE/ACM International Symposium on Code Generation and Optimization (CGO). IEEE, 2011; 141-150
- [19] Rajkishore B, Zhao Ji-sheng, et al. Efficient selection of vector instructions using dynamic programming[C]// 2010 43rd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO). IEEE, 2010
- [20] Markus W, Luk W. Pipeline vectorization. Computer-Aided Design of Integrated Circuits and Systems[J]. IEEE Transactions on Computer-aided Design of Integrated Circuits and System, 2001, 20(2): 234-248
- [21] Markus M, et al. Dynamic points-to sets: A comparison with static analyses and potential applications in program understanding and optimization[C]// Proceedings of the 2001 ACM SIGPLAN-SIGSOFT workshop on Program analysis for software tools and engineering. ACM, 2001
- [22] Mock M, Berryman M, Chambers C, et al. Calpa: A tool for automating dynamic compilation[C]// 2nd Workshop on Feedback-Directed Optimization, 1999; 291-302
- [23] Manuvir D. Unification-based pointer analysis with directional assignments[J]. Acm Sigplan Notices, 2000, 35(5): 35-46
- [24] Axel G. Evaluation of dynamic Points-To Analysis[M]. 2004
- [25] Wu Jing-yue, et al. Effective dynamic detection of alias analysis errors[C]// Proceedings of the 2013 9th Joint Meeting on Foundations of Software Engineering. ACM, 2013
- [26] Zhang Kun. Dynamic pointer tracking and its applications[M]. 2010
- [27] Maydan, Dror E, Amarasinghe S P, et al. Array-data flow analysis and its use in array privatization[C]// Proceedings of the 20th ACM SIGPLAN-SIGACT symposium on Principles of programming languages. ACM, 1993
- [28] Hao Yun-long. Research on Profile-Guided SIMD Vectorization Identification and Optimization [D]. Information Engineering University, 2011

(上接第 25 页)

- [8] Modha D S, Singh R. Network architecture of the long-distance pathways in the macaque brain[J]. Proceedings of the National Academy of Sciences, 2010, 107(30): 13485-13490
- [9] Hagmann P, Cammoun L, et al. Mapping the Structural Core of Human Cerebral Cortex[J]. PLOS Biology, 2008, 6(7): 1479-1493
- [10] Preissl R, Wong T M, et al. Compass: A scalable simulator for an architecture for Cognitive Computing [C]// High Performance Computing, Networking, Storage and Analysis (SC) 2012. Almaden, San Jose, CA, USA, 2012; 1-11
- [11] Eliasmith C, Stewart T C, et al. A large-scale model of the functioning brain[J]. Science, 2012, 338(6111): 1202-1205
- [12] George D, Hawkins J. Towards a Mathematical Theory of Cortical Micro-circuits[J]. PLoS computational biology, 2009, 5(10): 1-26
- [13] Jaros R G, George D. Hierarchical temporal memory system with enhanced inference capability[J]. Google Patents, 2012(7)
- [14] 姚宏亮, 王秀芳, 王浩. 一种基于结构分解和因子分析的贝叶斯网络隐变量发现算法[J]. 计算机科学, 2012, 39(2): 244-249
- [15] Liu Lu, Zhu Fei-da, Zhang Lei, et al. A probabilistic graphical model for topic and preference discovery on social media[J]. Neurocomputing, 2012, 95: 78-88
- [16] Frey B J, Jojic, Nebojsa. A comparison of algorithms for inference and learning in probabilistic graphical models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(9): 1392-1416
- [17] Ngiam Ji-quan, Khosla A, et al. Multimodal deep learning[C]// Proceedings of the 28th International Conference on Machine Learning (ICML-11), 2011. Bellevue: ACM, 2011; 689-696
- [18] Srivastava N, Salakhutdinov R. Multimodal learning with deep Boltzmann machines[C]// Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems (NIPS), 2012. Lake Tahoe: Curran Associates, 2012; 2231-2239
- [19] Blei D M. Probabilistic topic models[J]. Communications of the ACM, 2012, 55(4): 77-84
- [20] 李志欣, 施智平, 李志清, 等. 融合语义主题的图像自动标注[J]. 软件学报, 2011, 22(4): 801-812
- [21] Modha D S, Singh R. Network architecture of the long-distance pathways in the macaque brain[J]. Proceedings of the National Academy of Sciences of the United States of America, 2010, USA, 2010, 107(30): 13485-13490
- [22] Joseph R. Neuropsychiatry, Neuropsychology, Clinical Neuroscience[M/OL]. <http://brainmind.com/BrainOverview.html>, 2000