

神经网络

ECSN系统

解释部件

(14)

57-61

ECSN 神经网络系统的解释部件^{*} TP18

杨莉 葛建新

(浙江大学计算机系 杭州310027)

摘 要 This paper discusses the ability of neural networks to generate explanations, and then describes the explanation component (NEC) for ECSN neural network system with an explicit conceptual hierarchy. NEC can give meaningful responses to a limited class of "HOW" questions. NEC explanation component is presented in detail, including a formal specification, the sketch of a proof that inference patterns can be reproduced during explanation and so on.

1. 前言

解释在传统人工智能(AI)系统中是一个关键的功能部件,但在非结构化的神经网络系统里却难以实现,因为NN没有显式、陈述式(说明式)知识结构来允许解释结构的表示和产生。

本文以实例来讨论在神经网络系统里产生解释的能力,并介绍扩展了CSN功能的一个系统ECSN。对于一个有限的"HOW"问题类能给予一个有意义的回答。例如,对"Porsche-94有一个发动机,你是如何得到的"给予一个有意义的回答:"Porsche-94是一个Porsche, Porsche是小汽车,而所有的小汽车均有发动机"。这个解释在系统已经推得Porsche-94有发动机后产生。在解释阶段,ECSN系统的中间激活模式被重新生成,且神经网络知识库里对象之间的隐式关系也被显式化。

2. ECSN 中的解释部件

CSN是大规模并行系统,它能在概念知识上以很高的效率执行一定类型的推理^[10]。虽然当前CSN的表达能力有限,但允许有效的推理过程,因为所有计算通过通信来完成,处理能够最好地被展现为知识集成或知识聚集。另外,由于所有通讯是通过带权的连接完成,因而CSN是用于约束满足和证据推理的天然计算模型。总之,CSN的效能允许约束满足、证据推理、能量最小化和熵的最大值等信息处理

流程的实现和集成,并且其效能已明显超出了传统专家系统处理能力的范围。

此外,CSN的显式结构(如层次继承)使它适合用户解释的实现,与文[11]中描述的模型类似,因为在刺激向前传播系统中,处理过程的每个状态是有意义的且中间状态能用于解释,这使它具有一定的优越性,NN系统里显式结构的使用在此是关键的。

本文所描述的解释部件NEC中,为了要得到推理过程的一个解释,允许用户提出"HOW"问题。这同专家系统里的解释部件相似。ECSN系统给出的这种解释由CSN网络所编码关系的类型来决定,例如,在使用概念包含关系的标准CSN系统中,在继承期间属性和值的激活完全由概念层次结构所决定。

3. ECSN 系统

ECSN系统以推理和解释这两种不同的模式运行。在推理期间,刺激传播过程的中间激活模式被记录 and 保存;在解释期间,ECSN上的中间激活模式被重新产生,且关系被显式化,即解释不是激活模式的简单重显,而是推理状态的局部或全部重新生成,这些状态使推理期间已被使用的关系显式化。

3.1 ECSN 的结构

网络中相邻层节点间有相互激活连接(双向),而同一层节点间有相互抑制连接。这样,每层是一个特殊的WTA网络("winner-take-all")^[7]。同层节点

^{*} 国家自然科学基金和浙江省自然科学基金资助项目,并得到浙江大学CAD&CG重点实验室支持。杨莉 工学博士,博士后,主要从事人工智能、神经网络、智能CAD/CAPP等领域的研究工作。葛建新 工学博士,主要从事人工智能、CAD、图形学、CIMS、用户界面等领域的研究工作。

间的相互竞争导致“胜者”节点(只有一个)有很强的激活值,值得注意的是,如果几个节点接受到很强的刺激,可能同时被激发。

整个 ECSN 有4个空间:一个空间用于结构化对象的表示,一个空间用于属性的表示,一个空间用于属性值的表示,这三个网络模块有上述的结构;最后一个为实例空间,是一个特殊的单层空间,所包含的节点没有内部的组织。上面所有的表示均为本地的,对于每个对象、属性、属性值和实例的表示均对应网络空间中的一个节点。

从网络级看,每个空间有两类节点:概念节点和延迟节点。概念节点可有两种状态:“自由”、“约束”。在自由状态的概念节点称为自由节点,约束节点建立了一个层次结构,即被约束的对象、属性和值节点被编码在一个多层网络里并形成刺激向前传播的网络,这里仅仅使用约束节点的网络是可用的(自由节点用于学习新的知识,并把新知识集成到 ECSN 中)。约束节点不仅与同一空间中的节点相连,也同其它不同空间中的节点相连。对象、属性和值节点通过约束节点相连,对每个对象、属性和值之间的连接仅有唯一的一个约束节点。延迟节点和这些概念节点之间的连接不是直接相连的,它们至少用一层进行分开。

3.2 ECSN 的形式化描述

我们以 Shastri 相似的方法定义这个系统^[10],系统分为两个相关层次:①在概念级,一般的概念指示有公共特征的个体类,而个体表示这些概念的实例;②在网络级,节点以称为“空间”的集合形式进行组织,且节点间的相互作用通过带权的连接线进行。

本节中将使用下面这些标记:大写字母(D, A, ..., F)表示集合,小写字母表示集合中的个体元素,如 $o \in O, a \in A$ 。

3.2.1 概念级 在概念级,模型是一个八元组: $ECSN = \{O, A, V, I, F, \lambda, \alpha, \leftarrow\}$, 这里 O 是某一领域里所有对象的集合, A 是所有属性的集合, V 是 A 中所有元素可能取值的集合, I 是所有实例组成的集合, F 是在 ECSN 的概念层次的迭代学习期间可以被约束但目前尚未约束的所有自由实体的集合, λ 是从 O 到 A 的幂集的映射函数, α 是从 A 到 V 的幂集的映射, $\leftarrow$ 是定义在 O, A 和 V 上的一个偏序。注意到,这里不仅 O 中的元素是一般的概念, A 和 V 中的元素也是一般的概念,自由实体没有属性和值。

3.2.2 网络级 在网络级,模型是一个二元组 $M = \{U, W\}$, U 是网络中所有节点的集合, W 是网

络中所有权值组成的集合, U 由六种不同类型的元素组成, U^o 表示所有对象节点组成的集合, U^a 表示 $a \in O, U^a$ 是所有属性节点集, U^v 是所有特征节点集, U^i 是所有实例节点集, U^f 由所有自由节点组成且 U^f 表示 $f \in F$, 第六个也是最后一类节点 U^b 表示限制节点集。

3.2.3 节点类型和功能函数 节点是简单的计算单元,且节点之间通过带权的连接线相连,节点由下列相关元素来定义:

状态:可能两种取值:“推理”和“解释”

势能: $[0, 1000]$ 区间的某一连续值

输出:在 $[0, 1000]$ 区间取某一连续值,且输出值能被传播到其它节点

输入:输入 $i_1, i_2, \dots, i_n$ 组成的一个向量

一个节点有多个输入单元,每个输入单元由一连接线和节点相联,每个单元有一相应的单元函数,它能对相应单元上的输入值执行本地计算,且计算结果被单元函数用于算出单元的势能、输出和单元的状态,我们给出单元概念是使得能把一些输入处理为“能动”输入,“能动”输入的存在或缺席决定节点是否注意到(保留)此输入或忽略此输入的效果。

在 ECSN 中有两类节点(除实例空间之外):概念节点和延迟节点,后者是种特殊的约束节点,它有一特殊的结构,有三个“单元”表示输入到达“up”、“down”和“inhibition”上分开计算的地方:“up”单元的输入来自网络中的高层节点,“down”的输入来自网络中的低层结点,而“inhibition”的输入来自网络中同一层的节点。延迟节点对“inhibition”单元有一特殊的单元函数。

概念节点、自由节点与延迟节点有相同的结构,即两个单元“up”、“down”的输入来自高层或低层节点,“inhibition”单元,它的输入来自同层节点(但概念节点和自由节点的“inhibition”单元无特殊的单元函数)。另外,还有一“影响”单元,它的输入来自其它空间。

单元值 x 更形式化地表达为:

$$\text{单元值}_x = \sum_{i=1}^{N_x-1} w_{xi} \quad (1)$$

N_x 是单元 x 的输入数目,在时刻 $t+2$ 的“推理”期间,节点的势能为(注意 ECSN 中的节点与延迟节点相连):

$$\text{势能}_{t+2} = \begin{cases} \sum_{i=1}^{N-1} \text{单元值}_i & \text{若势能}_t > \text{下降级} \\ \text{下降级} & \text{若势能}_t < \text{下降级} \\ 0 & \text{若势能}_t = \text{下降级} \end{cases} \quad (2)$$

在解释阶段:

$$\text{势能}_i = \sum_{j=0}^{n_i} \text{单元值}_j \quad (3)$$

如果系统处于推理状态,那么系统的门阈值为0;处于解释状态时,系统的门阈值为1000;

$$\text{门阈值} = \begin{cases} 0 & \text{若处理模式为推理} \\ 1000 & \text{若处理模式为解释} \end{cases} \quad (4)$$

如果门阈值为0,节点处于推理状态;门阈值为1000,节点处于解释状态。

在推理期间,ECSN 系统节点的输出为:

$$\text{输出}_i = \begin{cases} \text{势能}_i & \text{若势能}_i > \text{门阈值} \\ 0 & \text{若势能}_i < \text{门阈值} \end{cases} \quad (5)$$

由于在推理期间,门阈值为0,所以输出_i = 势能_i。在解释期间:

$$\text{输出}_i = 0 \quad (6)$$

3.3 ECSN 系统中的推理

限定推理(如继承)在 ECSN 系统中能被有效地实现。CSN 系统中所描述的继承在本文中的形式如下:刺激网络中表达对象的节点,并且把刺激传播到表示更一般对象的节点,激活该节点的属性和属性值,随着刺激传播过程继续进行,越来越一般的对象被激活且越来越多的属性和特征节点被激发。

允许推理过程具有非单调风格的例外情形在 ECSN 系统中用附加的抑制链来实现,抑制链修改从对象类到特征的激活流,从而它也修改了继承过程,更特殊的类能修改来自更一般概念在特征上的影响。因此,更特殊概念和它们属性之间的连接将变得更为重要,这也实现了语义网络的一个基本假设:更特殊的比更一般的更为重要,且占据主导地位。

3.4 ECSN 中的简单解释

上面简单地解释了在刺激传播的 ECSN 系统中为什么中间状态是有意义的,输入引起的刺激沿着继承层次链进行流动,这也与超类/子类关系和特征恢复的连续应用相对应,每个中间状态是这种关系使用的结果,而且,借助于“恢复学习”的使用,在处理期间能记录和重放中间状态,即激活模式存储中表示一特殊时间步的单一节点的权值中(在相应时间步,权值被设置成输入值)。

解释部件使用具有快速权值变化的恢复学习,来保存推理过程中的中间状态,并用中间状态进行简单的解释。这些面向用户的解释可以有不同程度的解释粒度。本文工作的目标是使得 ECSN 系统具有回答“HOW”问题的能力。因此,ECSN 的解释部件具有下列功能:

(1)推理过程的全部重新生成,甚至当 ECSN 系

统已经发生变化;

(2)推理过程的部分重新生成;

(3)继承过程具有延迟性的重新生成;

(4)在一个单一空间里,几个推理过程的联合;

(5)在解释期间,借助于取消两个或更多的“历史节点”来复合多个中间激活模式;

(6)没有活跃的辅助的固定中间激活模式。

3.4.1 解释部件的结构 解释部件借助于 ECSN 系统上的两个功能模块——历史空间和关系空间来实现。历史空间是一个层次网络,它使用具有快速权值改变的特殊形式的恢复学习算法。恢复学习算法的详细描述清见文[7]、[1,2,3,4,5]。Fanty (88)首先提出把恢复学习算法用于 NN 的顺序处理环境里。历史空间组织成层次结构的原因是允许以不同级别的粒度进行解释。关系空间是一个节点集,其中关系空间中的节点与 ECSN 系统中的节点相连(关系空间和历史空间中的节点例外),并同历史空间中第一层的所有节点相连,关系空间中的每个节点明确地表示了 ECSN 系统中所编码的一个关系。例如,标准型的节点可能表达“子类/超类”、“概念/特征”和“实例/特征”关系。

当然必须限制作为 ECSN 系统一个组成部分的刺激传播过程的形式。对每个 ECSN 系统的节点而言,在此节点已到达最大激活值附近不久,再执行一个非常强的延迟过程是很有效的,这个延迟过程能减少单元的激活值,从而保证在 NN 继承层次里至多只有一个节点被完全地激活。使用很强的延迟过程的另一个原因是避免在解释模式中出现失真情况。这里“失真”指被两个或更多个概念所拥有的特征的混淆。

3.4.2 历史空间 历史空间由一个层次网络构成,此网络第一层中的每个结点与 ECSN 系统中的所有节点(ECSN 系统中的辅助节点,如延迟节点以及历史空间和关系空间中的节点除外)之间有一对称链接,历史空间第一层的节点被组织成一条链,即第一个节点有一个激活链连接到第二个节点,第二个节点有一激活链和第三个节点相连……。在第一层的每个节点上有一个激活函数,它仅对单个修改步输出一高的激活值,并在具有一很强的正输出的修改步之后关闭节点本身(即节点处于不活跃状态)。进一步,来自或到达第一层中节点的输入链执行“快速权值的改变”,即一旦节点有一高的输出值时,权值被设置为输入值。

在本节的最后,将讨论解释的两种不同的粒度。

这使得历史空间除第一层外还需另外三层节点。历史空间中第二层的节点和第一层中的节点相连,并用于“细粒度”的解释。第二层的节点象第一层的节点那样组织成一条链,但每个节点和下一个节点之间有一传播延迟。另外,在第二层和第一层节点之间有一对一的连接,即第二层的第一个节点与第一层的第一个节点相连,第二层的第二个节点与第一层的第二个节点相连,以此类推。

历史空间的第三层节点用于“更细粒度”的解释,如第二层节点那样组织起来,但它同第一层中的节点有一稀疏(交叉)连接。例如,第三层的第一个节点和第一层的第一个节点相连,第三层的第二个节点与第一层的第四个节点相连,以此类推,每隔三个节点相连。第四层也是最后一层,由多个独立节点组成,用于表示解释的粒度,在本文的结构里,第四层仅有二个节点,第一个节点与第二层的第一个节点相连;第二个节点与第三层的第一个节点相连。

3.4.3 关系空间 如上所述,关系空间由 ECSN 中表示关系的节点组成,例如对象空间表示了概念的一个分类,即它编码了“子类/超类”关系,因此在关系空间里的一个单一节点必须表达这样一种关系。从对象节点到属性和特征节点的连接线表示了“概念/特性”关系,且用关系空间里的一个单一节点表示这种关系。ECSN 系统中的节点同关系空间里的相应节点之间有一附加的连接线,即无论什么时候在对象空间的“子类/超类”链上(或中间延迟单元)有一刺激流,关系空间中的相应节点也将被激发。关系空间里的节点和历史空间里第一层的节点之间有一全互联关系,反之亦然。

3.4.4 解释部件的形式化描述 3.2节描述的 ECSN 系统用下列方法具有一个良好的解释部件:解释部件是一个二元组 $EC = \{R, H\}$, 这里 R 是 ECSN 中的关系集, H 是 ECSN 里所有状态的一个子集(所有可能的激活模式),即 H 是一继承过程中所有中间状态的集合。在网络级, R 由一个节点集合来表示,它表达了 ECSN 里所编码的每个关系, H 由一 k 层网络表示。在 H 中,从第 k_0 层到第 k_{n-1} 层有一分枝因子为 b (由用户确定)的联接,且从 $k_{n,m}$ 到 $k_{n,m+1}$ 有一连接线,这里 m 为第 k_0 层中节点位置的标记。第 k_i 层(历史空间中的第一层)的所有节点和 ECSN 中所有约束节点以及 R 中所有节点相连; $CCU = \{U^0, U^1, U^2, U^3, U^4\}$, 反之亦然。另外, R 中的所有节点共同地被连接到 CCU 中的元素。

时刻 t , k_i 层中第 m 个节点,也就是历史空间中

第一层第 m 个节点,它的输出和势能可由下式计算:

$$\begin{aligned} \text{输出}_{i,1,m} &= \text{势能}_{i,1,m} \\ &= \begin{cases} 1000 & \text{若输入}_{i,1,m-1} = 1000 \\ & \text{且输出}_{i,1,m-1} = 0 \\ 0 & \text{否则} \end{cases} \quad (7) \end{aligned}$$

这里输入 $u_{i,1,m-1}$ 是来自历史空间第一层中当前节点前驱的输入。

在推理期间,CCU 节点执行来自历史空间第一层节点发出链上权值的快速变化:

$$w_{ij} = \begin{cases} \text{势能}_i & \text{若势能}_i > \text{门限值} \\ & \text{且输入}_i > \text{门限值} \\ 0 & \text{否则} \end{cases} \quad (8)$$

即:如果接受节点的势能和它所收到的输入很高,那么权值设置为接受节点的势能,否则权值为0。另外,在推理期间,当历史空间第一层的节点有很高的势能时,它们将执行来自 CCU 节点发出链上权值的快速变化:

$$w_{ij} = \begin{cases} \text{输入}_i & \text{如果势能}_i = 1000 \text{ 且输入}_i > \text{门限值} \\ 0 & \text{否则} \end{cases} \quad (9)$$

本节的其余部分将展示,历史空间第一层中的节点能保留一个推理过程的所有步骤,且这个推理过程在解释期间能够被重新产生。

假设, X_i 是在时刻 t (步骤 t) 时 CCU 里所有节点上的激活模式,即 X_i 是在时刻 t , 导出 CCU 中所有节点输出值的一个激活向量, $w_{i,1,m}$ 是历史空间第一层节点 $k_{i,1,m}$ 的权值向量。

引理1 若在推理期间,势能 $u_{i,1,m} = 1000$, 那么 $w_{i,1,m} = X_i$ 。

证明:由公式(9)可显然地得出。

引理2 在时刻 t_0 , 节点 $k_{i,0}$ 和输入节点 U_i^0 同时成为活跃状态并且整个系统使用同步修改策略。那么在推理期间,按 k_i 层节点的次序顺序地存储所有的 $X_{i,0}, \dots, X_{i,n}$ 。

证明:从 $k_{i,m}$ 到 $k_{i,m+1}$ 存在一“能够”链,另外,若 $O_i \leftarrow O_{i+1} \leftarrow O_{i+2}$ 成立,也存在从 U_i^0 到 U_{i+1}^0 的链,因为公式(8)、(9)成立,所以在时刻 t_0 , 若输出 $u_{i,1,0} = \text{输出}(U_i^0) = 1000$, 那么 k_i 层中的节点顺序地保存所有的 X_i 。公式(7)保证了 $k_{i,m}$ 中的节点顺序地保存所有的 X_i 。

引理3 在解释期间,若输出 $u_{i,1,0} = 1000$ 且在时刻 t_0 , CCU 中所有节点的输出为0,并且系统使用同步修改,那么在所有 CCU 节点上,激活模式 $X_{i,0}, \dots, X_{i,n}$ 按顺序被重新生成。

证明:解释期间,在时刻 t_0 , 节点 $k_{i,0}$ 是活跃的。

并根据等式(7),在 k_1 层所有节点被顺序地激活,而根据等式(6),不存在激活值的传播过程,节点 CCU_1 的势能等于输出 $u_{1,m}$ 乘以到达接受节点的权值(或最大势能),即重新生成的势能和推理期间的势能一致。

换句话说,CCU节点有高的势能,但如果在解释期间输出 $u_{1,m}=1000$,那么CCU节点的输出为0,不存在激活值的传播过程且激活模式处于静态。

3.4.5 解释部件是如何工作的 再一次,对每个时间步,历史空间第一层的每个节点是活跃的。在这个单一时间步期间,节点执行快速权值改变过程并传送一个信号到当前节点(历史空间第一层中)的下一个节点,且启动此节点,这要求在历史空间中同时处理。快速权值改变允许在单一节点的权值里存储暂时的激活模式,换句话说,ECSN系统中的激活模式被存储在节点的连接上。历史空间第一层的节点数目必须大,足以能够存储在最终结果被到达为止的所有激活模式。如果需要执行“100步规则”^[1],那么第一层必须使用100个节点。

已存储的激活模式用于解释,这样做是可行的,因为作为ECSN系统组成部分的继承层次在语义上

是有意义的。因此,推理中的每步也是有意义的,即概念层次中激活值传播的每步和特征的结果激活值中的每步都是有意义的。另外,历史空间的结构允许存储限制推理的过程,它的目的是用于解释。启动历史空间第四层的第一个节点将引起推理过程所有中间状态的顺序重新生成,并使这一过程可为解释服务。启动历史空间第四层中的第二个节点将引起推理过程某些中间状态的顺序重新生成,因此这样得到的解释有更细的粒度。简言之,用户有可能在几种不同程度的解释中进行选择,但如果小粒度模式被使用,有可能遗失有价值的信息。

4. 结论

本文用实例说明神经网络系统里解释机制实现的方法,由此可知,神经网络系统受益于关系的显式编码和高度结构化网络的使用,使用刺激传播的结构化神经网络系统有下面的优点:处理过程中的中间状态在语义上是有意义的,且这些中间状态能用于解释,这也为神经网络系统行为的合理解释提供了一条可行的途径。(参考文献共11篇略)

(上接第56页)

进入到更高一级集成的计划——计算机集成制造系统(CIMS)。CIMS的一种定义为:将计算机的自动化和决策支持系统集成地用于管理制造系统的全部操作,这些操作覆盖了从产品和库存管理以及财务、资源管理的全过程。CIMS汇集了智能机器人、FMS、局域网(LAN)及互接技术、智能CAD/CAM、工业控制装置、信息管理决策、工程专家系统等现代化技术,受到学术界和工业界的高度重视。CIMS作为一种全新的制造哲理正从体系结构、设计与制造方法论、信息处理模型等角度影响并决定着以多品种、中小批量生产占主导地位的现代化机械制造的模式,是制造业发展的方向和未来。

参考文献

[1]殷国富,余卫东,陆春进,90年代CAD/CAM技

术展望,机械,1993(3)

[2]殷国富,专家系统技术概况与工程应用,机械,1989,16(1):21~27

[3]M. L. Maher, D. Sriram and S. T. Fenres, 用于工程设计的基于知识的专家的工具和技术,计算机科学,1985(6):24~37

[4]G. M. Hare, Designing Intelligent Manufacturing System: A Distributed Artificial Intelligence Approach, Computer in Industry, 1990(15):17~25

[5]殷国富,工程专家系统技术及其应用,成都科技大学出版社,1993

[6]V. Arman et al., A Fundamental and Theoretical Framework for an Intelligent CAD system, Computer Aided Design, 1990, Vol. 22, No. 6