28卷 1期 心 理 学 报 Vol.28, No.1
1996年2月 ACTA PSYCHOLOGICA SINICA Feb., 1996

汉语句子格角色分配的一种神经网络方法[1)]

张东松 陈永明 喻柏林
(中国科学院心理研究所，北京，100012)

摘 要

提出了一个基于分布式表征的计算模型，通过并行分布加工方式完成六类汉语句子的格角色分配任务。模型是一个四层的前传网络，包括输入层(词的分布式表征层)，两个隐层，输出层(格角色层)；其中第一隐层的一部分反馈到输入层。模型采用误差反传算法，通过提供学习样本和目标输出，不断调整三个权值矩阵，使得网络稳定时能得到正确的结果。经过训练后的网络具有一定的稳定性和鲁棒性。还对这种方法与传统的符号处理方法作了比较和分析。

关键词 神经网络，权值，BP算法，格角色。

1 引 言

语言理解是指从语言的表层结构中提取出深层的命题结构的一种推理过程，即从语言的表层结构去建构意义的过程。人们理解语言首先要接受由外部输入的语言刺激，然后在心理词典中进行搜索，获得词的知识，再经过句法分析和语义分析，得到句子的意义。因而，语言理解不仅依赖于外部输入的信息，并且依赖于人内部的知识组织、认知结构。计算机处理自然语言，也必须具备有关的知识，包括语言学知识和非语言学知识。对自然语言理解而言，句子的分析和理解是最关键的一环，它是课文理解的基础[1,2]。本文涉及的汉语句子的格角色分配，则是句子理解的一个重要环节。

2 人工神经元网络方法的崛起

传统的人工智能(AI)主要采用了串行的符号处理方法来进行语言的理解和加工。它需要一个庞大的知识库和一个复杂的规则集，根据条件利用已定的规则对语言进行操作。在这方面的研究虽然已取得了一些成就，但也遇到了一些棘手的问题。例如由于语言，尤其汉语本身的模糊性，多变性及不精确性，使串行符号处理方法在处理自然语言时，会遇到以下问题[3]。

(1) 知识的表征和获取，这一直是符号处理方法面临的“瓶颈”问题。

(2) 系统的脆弱性。系统只能利用已有的知识和规则来处理问题，当它遇到新的语言现象时，由于无对应规则，或知识库内无有关的知识，它就表现得无能为力了。

(3) 在语言理解，比如句子的理解过程中，句法分析和语义分析应同时进行，两者相互作用，相互补充。经验告诉我们，任何先句法后语义或先语义后句法的分析方法都有其

1) 本文修改稿于1995年9月8日收到。

局限性。而串行符号处理方法无法做到句法、语义和语境信息的并行加工。

(4) 在串行符号处理方法中，推理策略单一，学习能力差，知识库管理困难。著名的人工智能学者 Moor 曾说过："搜索和匹配是人工智能的核心"。传统 AI 技术中的推理过程，主要是在求解空间内进行符号的匹配、搜索和回溯的过程，所涉及的数学计算量很少。当规则集很庞杂时，会产生组合爆炸，从而大大增加推理的复杂性。

在人类理解语言的过程中，句法分析和语义分析是紧密联系的。当然，对于两者在句子理解过程中的具体作用方式，认知心理学中也存在着两种不同的观点。一种是句法自主理论，认为句法分析不依赖于语义而自主地进行，语义分析部分只是对已作出的结构分析进行语义匹配，如果匹配失败，则转回句法分析部分。另一种观点认为句法分析和语义分析之间存在着很强的，即时的信息交换，两者是相互作用，相互影响的。两种理论都得到了一些心理学实验的证明。总的说来，这两种理论都认为，虽然句法分析和语义分析相互间存在着信息交换，但在句子理解过程中它们仍是相对独立的模块[1]。

McClelland 等人根据并行分布加工理论提出，在句子理解中不存在独立的句法分析和语义分析模块，词汇信息、句法信息、语义信息是以并行方式同时加工的，并由此得到一个综合了各种信息的句子表征[4]。随着神经网络理论的发展，这种并行分布加工的思想在处理语义和句法分析的关系上显示了一定的优越性。

人工神经元网络是模拟人脑并行加工机制的一种模型，起源于一九四三年，由McCulloch 和 Pitts 提出了第一个神经元的计算模型，而应用于人工智能的研究，则是从五十年代末，六十年代初开始的。人工神经元网络系统是由大量的形式神经元连接而成的高度复杂的非线性系统。它采用自下而上的方法，研究大量简单的神经元的集团信息处理能力及其动态行为，重点在于模拟和实现人的认知过程中的感知觉过程，形象思维，分布式记忆和自学习、自组织过程。它采用分布式存储方法表示知识，并把通过训练和学习获得的知识存储在连接各神经元的权值中。人工神经元网络主要进行的是数字计算，从而避免了在问题求解空间内进行求解的"组合爆炸"问题，并具有潜在的并行计算能力。节点特性、连接拓朴结构、学习规律是确定一个神经网络的三个要素。目前虽然它的知识处理能力与符号处理方法相比还表现出不足，但在处理一些不精确的、不完全的和环境信息不明确的信息，以及在知识表征，获取和学习等方面仍不乏其吸引力。总的说来，神经网络模型具有几个优点。

(1) 内在的自学习、自组织能力

网络依赖于学习样本而非规则。当环境变化时，网络可自动调节内部状态，使其能适应环境的变化。

(2) 内在的泛化能力

网络把从训练集里抽取到的隐含规则和知识分布存储于权值里，然后再用来加工以前并未见过的样本。

(3) 误差和噪音的容忍性

网络对带噪音的或不完全的输入模式表现出一定的辨别和再认能力。

(4) 网络的行为表现出一定的预期性

由此可见，神经网络在知识获取等方面有其独到的地方，它应与传统的人工智能方法

相结合，取长补短。到目前为止，已提出了若干神经网络模型用于自然语言的处理，如 Boltzman 机，Kohonen 自组织网络，ART 网络，BP 网络等。本文是用神经网络处理汉语句子的一个尝试。

3　本模型描述

3.1　模型的基本结构

本模型为一个四层带反馈的前传网络（如图 1 所示），它的任务是对一个完整的汉语句子进行格角色分析。'格'的概念是由美国著名学者 Fillmore[5] 提出的，他认为一个句子是以动词为核心的，其它词都与动词有着某种关系，这种关系称为格关系。格分析体现了对句子中词的语义关系的一定层次上的理解。在建造本模型时，考虑了如下几个方面。(1) 输入层采取局部表征还是全局表征。若是后者，各输入节点代表哪些信息。(2)如何在网络中表示上下文关系。(3)网络连结的拓朴结构。(4)采用何种学习规则。

根据并行分布加工的思想，本模型在输入层采用全局表征，每个词都由一组特征神经元表征。这种全局表征方法有如下优点：

(1) 能直接体现出对被加工对象的并行处理。

(2) 当要处理新词时，只需将该词的特征集加入到词典中去即可，不用改变网络结构，从而保证了网络系统的完整性。

(3) 由于所有的词在输入层都由同样的一组神经元来表征，因而可以体现出同类词之间的类似和联系以及异类词之间的差别。

网络利用反馈和测试部分实现句子内上下文的内在联系，并采用误差反传算法（BP 算法）学习。

模型构造如下：

第一层是输入层，采用分布式表征方法，共有 116 个单元。具体由五部分构成：

(1) 16 个句法属性单元；

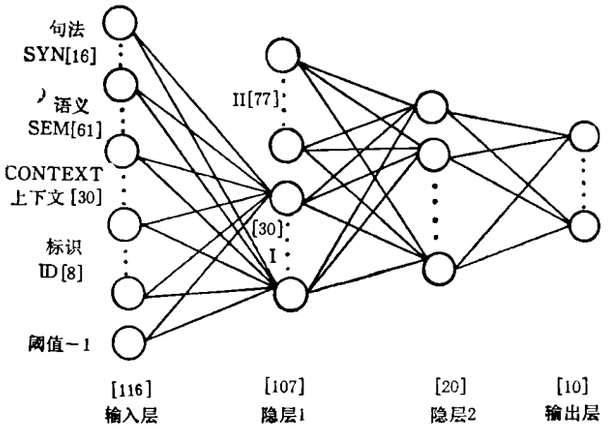

图 1　模型结构

(2) 61 个语义信息单元；

(3) 8个ID单元;

(4) 1个总保持-1的偏值单元;

(5) context部分:30个记录上下文信息的单元。

汉语句子中的每个词都由(1)—(4)部分来表征。词是一个一个地加入到网络中的。当它具有某个句法和语义属性时,则输入层中对应的该单元置1,否则置0;ID部分相当于每个词的标识符,因而是各不相同的;在句子的第一个词加入网络前,context部分的单元全部置0。随着词的不断加入,context部分将不断改变,为句子的下一个词提供当前的上下文线索。

第二层,即第一隐层,包含两部分:Ⅰ和Ⅱ。第Ⅰ部分有30个单元,它们与输入层和第三层采取全连接的形式,相当于是通过不断抽取输入层词的信息而得到的句子信息部分。第Ⅱ部分是一个有77个单元的测试部分,也包含16个句法属性单元和61个语义信息单元。这部分不与输入层相连,只与第三层全连接。它的设计目的在于:由于对句子的加工是通过对各个词的加工总合而成,词和词之间是相互联系而又相互制约的。人在加工句子时,前面的词会对后续词的出现产生一定的预期,而后续词的加入也会影响对前面词的理解和判断。测试部分的存在正是为训练和测试网络提供这种机制。

第三层,即第二隐层,有20个单元,它们起着从前一层的句子信息和测试信息中抽取特征的作用。

第四层是输出层,用10个单元分别表示10个格角色(参见附录1),其中施事、受事、时间、处所、动作、客体格等的选择依据Fillmore的格语法理论[5]。当10个单元中的某一个输出近似为1,而其余皆近似为0时,说明当前词在句子中担当这个输出单元所代表的格角色。第二层和第三层,第三层和输出层之间为全连接,而隔层之间及同一层的单元之间无连接。

3.2 网络训练的方式和过程

网络模型要学习对汉语句子进行正确的格分析,即当一个完整的汉语句子输入网络后,应在网络输出层得到该句中各个词所担当的格角色。

3.2.1 训练样本集,目标集和词典的建立

用来训练网络的汉语句子共六类,108个单句(依据[6]),这六类汉语句型为:

(1) 主语+谓语 或 宾语+谓语结构

例:我去学校。 昨天水沟已经填了。

(2) 主语+谓语+宾语结构 例:她今天买了许多菜。

(3) 宾语+主语+谓语结构,即宾语前置句 例:菜我还没有吃。

(4) 双主句 例:小张和陈叔明天去北京。

(5) 被动句 例:弟弟被大家批评一顿。

(6) 含介词短语结构句 例:学生在教室看书。

这些训练句分成12组,放在12个文件中,且每组中各类型句子至少有一个。目标集同训练集一样,分别放在对应的另12个文件中。当从训练集中取出一个样本句,并将词一个一个地加到输入层时,同时从目标集中取出该句各词的目标输出依次加到输出层上。若词在句中担任某个格角色,则对应输出单元的目标输出为1,其余9个的目标输出为

0。以下述句子为例：　　小张被大家批评一顿。

这句在训练时的目标模式为：

小张[0 0 0 0 0 1 0 0 0 0]

被　[0 0 1 0 0 0 0 0 0 0]

大家[0 0 0 0 1 0 0 0 0 0]

批评[0 0 0 0 0 0 1 0 0 0]

一顿[0 0 1 0 0 0 0 0 0 0]

由此看出，输出层的十个单元在同一时刻只能有一个被激活为 1，其余为 0。

词典存储了 200 个左右的汉字和词，其中包括 62 个名词，57 个动词，及一些代词，形容词，副词，方位词，量词等(依据[7])。因模型主要是研究网络能否对汉语句子进行正确的格分析，故句子的分词不在本研究之内。为方便起见，在训练句加入网络前，已在样本集里预先进行了人工分词，词与词之间由空格隔开。

3.2.2　网络采用误差反传学习算法(BP)。流程图见图 2。

3.2.3　有关参数的设置

η为学习参数。BP 学习算法的效率和收敛在很大程度上取决于学习参数 η 的值。大

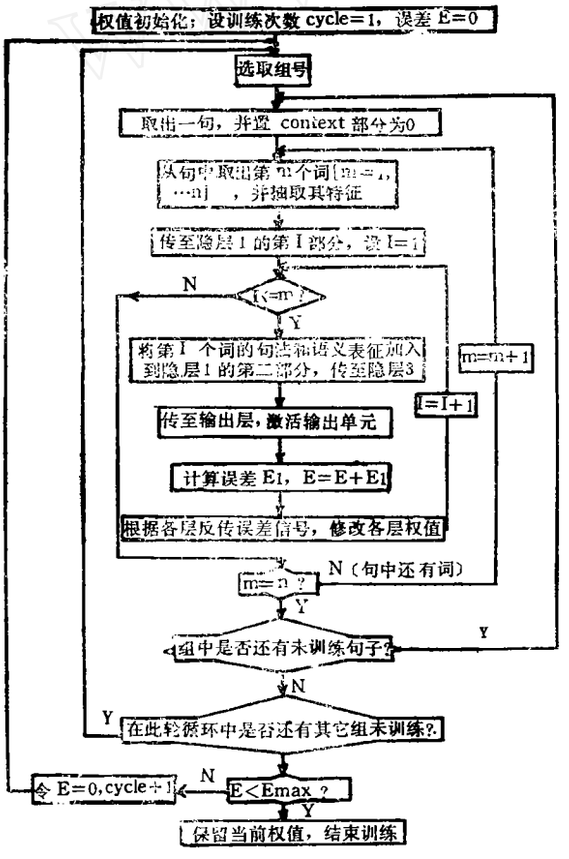

图 2　学习流程图

的η值对应一个快速的学习过程,但可能产生振荡而不收敛;小的η值则会使学习缓慢,效率下降。也就是说一个小的η值能保证真正的梯度下降,然而却要付出增加更多学习次数的代价。所以η的确定应根据不同问题,不同情况而定。本模型经过多次试验,最后选定在学习过程中η的值介于区间[0.1,0.4]之间:当学习刚开始时,令η=0.4,随着学习的进行和误差的逐渐减小,η值也由0.4逐渐下降到0.1,这种调整既可由规则控制,也可由人在训练中任何需要的时候加以控制。

mom为动量项,加入它是为了加快算法的收敛。

$$\triangle W_{ij}(n+1) = -\eta \times E(t) + mom \times \triangle W_{ij}(n)$$

这个式子规定在第(n+1)步W的变化还与第(n)步的权值变化有关。mom试图使第n步权值的改变,或多或少地与第n-1步在相同方向上进行。非无限小的mom将能抑制振荡的发生,但有可能使学习速度降低。本模型中的动量项mom的大小选择类似于η,也随着误差的下降由0.2→0.05。

在网络学习开始时要建立初始权值,而且权值的初始化对训练的最终结果影响很大。如果所有权值都初始化成相同的值,很可能使训练不正确或不收敛,所以一般都将权值初始化成小的随机值。在本模型的学习开始时,由一随机函数random( )给三个权值矩阵中的权值赋以初值,其值在[-0.5,+0.5]之间。

Emax的值将决定网络何时停止训练及网络的精度。Emax越大,则学习次数越少,网络的精度越低,网络对测试样本的判断正确率也越差;反之,则增加学习次数,提高网络的精度。然而Emax并非越小越好,因为当网络学习到一定程度时,收敛已极其缓慢,虽误差的变化趋势仍有可能下降,但这种变化微乎其微,且经常有升降反复的现象。故兼顾到学习效率和网络精度的要求,本模型将Emax定为15(总误差)。

为方便学习,建立了交互界面,可在需要的时候随时进入交互状态,既可修改η,mom,及各层单元数,也可以单独训练某个句组或者句子。

3.3 训练结果

训练122个循环后,网络的总误差(即一轮学习108个句子的总误差)为14.994,小于最大允许误差Emax,学习停止,此时已向网络呈现了108×122=13176个句子,平均单句误差error=0.138。

学习次数	1	2	3	4	5	10
误 差	113.53	63.00	51.41	45.55	40.20	37.73
学习次数	15	20	25	30	40	50
误 差	34.34	31.69	29.34	27.10	25.94	23.33
学习次数	60	70	……	122		
误 差	21.95	20.54	……	14.994		

在学习开始时,误差下降很快。随着循环次数的增加,这种下降趋势虽仍然保持,但逐渐变缓,甚至最后近似于停滞状态。

将当前三个权值矩阵存储下来。用测试程序测试已训练过的句子,结果100%正确。

用15个未曾训练过的句子测试网络,结果13句正确,2句有个别词判断错误,正确

率87%，表现出一定的泛化能力。测试的例子见附录 2 。

编程语言为C语言。网络训练在 AST 486/33 微机上进行，训练时间为二个半小时左右，平均 88 句/分钟；当测试时，处理速度为每句 1～2 秒。

4 结　论

本实验表明，神经网络用于处理汉语句子有一定的优点。

(1) 人类对语言的理解，是综合运用句法、语义和语境知识的结果。本模型将这三种知识融为一体，对句子进行分析时，同时运用了这三种知识，网络的行为表现出一定的预期性。因此，在某种程度上较接近于人处理句子的过程。

(2) 本模型对字词的分布式表征作了一定尝试。这种分布式全局表征或许比局部表征更近似于人脑存储信息的方式。

(3) 网络学习完毕后，有关的知识全部存储在神经元的连接权值中；网络的输出结果是所有神经元及其连接权值的整体行为。这种并行分布式的加工，在某种意义上说，更类似于人类对信息的加工机制。

(4) 经过训练后的网络表现出一定的稳定性和鲁棒性。它不仅能对学过的句子作出正确的分析，而且也能对未学过的、类似的句子表现出一定的认知能力。在这一点上，与人的语言学习是类似的。

(5) 当传统的符号处理方法在自然语言处理的某些方面遇到困难而显得能力不足时，神经网络可以作为一个有力的工具来补偿这种不足。因此，神经网络和传统的 AI 之间是一种互补的关系。把神经网络信息处理技术与传统的 AI 有机地结合起来，应是未来研究和建构较完善的自然语言理解系统的方向。

[附录 1] 输出层的格

处所格(Location)	时间格(Time)	修饰格-1(Modifier-1)
非施事主格	施事格(Agent)	受事格(Patient)
动作格(Action)	修饰格-2(Modifier-2)	
客体格(Object)	工具格(Instrument)	

[附录 2] 测试结果举例

```
例：(分析正确)  在家他没有认真看书。
    The word (or phrase) 在
    The maximum output unit is NO. 3 (参见附录 1)
    value is 0.934751(可能性), MODIFIER-1 (格)
    The word (or phrase) 家
    The maximum output unit is NO. 1
    value is 0.927407, LOCATION
    The word (or phrase) 他
    The maximum output unit is NO. 5
    value is 0.960721, AGENT
    The word (or phrase) 没有
    The maximum output unit is NO. 3
    value is 0.956303, MODIFIER-1
    The word (or phrase) 认真
    The maximum output unit is NO. 3
    value is 0.941852, MODIFIER-1
    The word (or phrase)看
```

The maximum output unit is NO. 7
value is 0.966792, ACTION
The word (or phrase)书
The maximum output unit is NO. 9
value is 0.976504, OBJECT

参 考 文 献

1 彭聃龄等。语言心理学。北京：北京师范大学出版社，1991：141—179。
2 刘开瑛，郭炳炎。自然语言处理。北京：科学出版社，1991：1—14。
3 曹焕光。人工神经元网络原理。北京：气象出版社，1992：38—50。
4 John F S, McClelland J L. Learning and applying contextual constraints in sentence comprehension. Artificial Intelligence, 1993; 62: 217—252。
5 杨成凯。Fillmore 的格语法理论。国外语言学，1986；(1)：37—42。
6 李英哲等。实用汉语参考语法。
7 吕淑湘。现代汉语八百词。北京：商务印书馆，1991：1—41。

ASSIGNING CASE ROLE OF CHINESE SENTENCES WITH A NEURAL NETWORK

Zhang Dongsong, Chen Yongming, Yu Bolin

(Institute of Psychology, Chinese Academy of Sciences Beijing, 100012)

Abstract

According to PDP theory, we tried to use a calculating model based on distributed representation to complete the task of case role assignment of Chinese sentences by parallel processing. There were six types of sentences. The model was a four-layer forward neural network. In input layer, we used distributed representaton to collect syntactic, semantic information (of the word in Chinese sentence) and context information, and there were two hidden layers, and output layer (case role layer). One part of the first hidden layer was feedback to the input layer. Error back propagation learning algorithm was used to adjust three weight matrices sequently according to learning samples and target output in order to get correct answers when the network was stable. After training, the network was somewhat robust. In addition, the neural network method and the traditional symbol processing method used in natural language understanding was compared and analyzed.

Key words neural network, weight, error back-propagation algorithm, case role.