基于神经网络的汉语口语多义选择*

王海峰 高 文 李 生

(哈尔滨工业大学计算机科学与工程系 哈尔滨 150001)

E-mail: i-haiwan@microsoft.com

摘要 汉语口语分析是交互式话语处理中的重要环节.在汉语中,有意义的最小单位是词,因此多义选择是口语分析系统必须首先解决的问题.该文提出了一种基于精简循环网络的汉语口语多义选择方法,并从词汇的语法、语义分类所固有的内在联系出发,给出了语法、语义的一致化处理策略.通过使用会面安排领域的口语语料进行实验,多义选择的开放测试的正确率为96.9%.

关键词 神经网络,精简循环网络,口语分析,汉语口语,多义选择.

中图法分类号 TP18

口语分析是口语的计算机处理中的重要环节之一.口语具有许多不同于书面语的特点,比如,口语中的不连贯现象(包括迟疑、文吾、重复、更正、语气词插入等)、语法约束相对较弱、没有明确的句子边界等[1,2].另外,口语分析还必须容忍和处理语音识别错误.因此,口语分析方法必须具有更强的容错能力和鲁棒性.

同一词汇在不同的上下文环境中可能具有不同的语法和语义属性,这就是词的多义现象.一个多义词的语法、语义属性在一定的上下文环境中是唯一确定的,多义选择任务就是在特定语境中为多义词选择唯一正确的义项.词的语法和语义属性是构成语言结构和意义的基础,词汇级多义选择的准确程度直接关系到整个口语分析的成败.

为了满足口语分析的需要,各国研究者在借鉴书面语分析方法的同时,也对其进行了必要的改造[3,4].神经网络方法因其较强的学习能力和良好的鲁棒性而受到众多口语分析研究者的重视[4].精简循环网络(simple recurrent network,简记为SRN)[5]通过上下文单元(context units)的引入而使网络具备了记忆和利用上下文的能力.Wermter等人将SRN用于德语口语分析,取得了良好的结果[4].但Wermter等人对语法和语义类的选择各自独立,忽视了语法和语义的内在联系,没有考虑它们的一致性问题.

本文首先介绍SRN,然后给出基于SRN的汉语口语多义选择方法,并提出一种语法和语义的一致化理策略,最后分析实验结果并得出结论.

1 精简循环网络

SRN由Elman首先提出,并用于时间序列预测问题[5].之后,很多学者对SRN的能力和应用做过研究[4,6,7].此网络结构如图1所示.

在网络运算和训练中,输入层和上下文层组成网络的联合输入层.这样,就可以把网络作为一个$k+m$个输入神经元的3层前馈网络来处理了,只是其输入层的前k个单元把网络隐藏层的前一组输出作为自己的输入.通常,SRN采用BP算法进行网络参数训练.

* 本文研究得到国家自然科学基金、国家863高科技项目基金、国家教育部跨世纪人才基金和中国科学院"百人计划"基金资助.作者**王海峰**,1971年生,博士生,主要研究领域为机器翻译,计算语言学,人工神经网络.**高文**,1956年生,博士,教授,博士生导师,主要研究领域为人工智能,多媒体技术.**李生**,1943年生,教授,博士生导师,主要研究领域为机器翻译,计算语言学,人工智能.

本文通讯联系人:王海峰,北京 100080,北京市海淀区知春路49号希格玛中心五层

本文1998-09-11收到原稿,1998-12-01收到修改稿

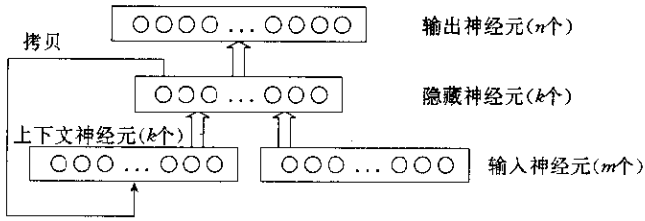

图1　精简循环网络

通过上下文层的引入,SRN可以使用较大范围的上下文作为当前运算的依据[5].

2　基于精简循环网络的多义选择

2.1　基本知识集

鉴于话语处理存在突出的困难,目前各国学者在从事这方面的研究时,都将其限定在一个明确的应用领域内,这样既降低了难度,又有明确的应用价值.本文所述方法用于面向会面安排(meeting schedule)的汉英口语翻译[8],共录制了关于会面安排的对话150段,然后人工逐音、逐字地听写、誊录获得与语音材料完全对应的文本语料库约20 000字(包含对话中各种人为错误和非人为噪声).经过分词处理,得到839个词的汉语词表.将通常语法书上的词类进行细化,确定了与领域无关的语法类26个.语义描述的是语言的意义,而话语的意义与其领域背景有关,因此,在限定领域的口语分析中,语义分类的确定也都是与领域相关的.本文从会面安排领域口语对话的特点出发,建立了包括37个语义类的领域相关语义集.最后,为词表中的词逐个添加语法语义分类,得到口语分析用词典.

一个词条可能有多个义项,每个义项都有自己的语法和语义类.同一词条的任意两个义项的语法和语义类不同时相等.义项数大于1的为多义词.多义现象的静态和动态分布特征见表1.

表1　多义现象的分布特征

	静态特征(对词典统计结果)		动态特征(对语料统计结果)	
	个　数	百分比(%)	个　数	百分比(%)
总词数	839	—	10 838	—
多义词	62	7.4	2 580	23.8
语法兼类词	40	4.8	1 721	15.9
语义兼类词	54	6.4	2 234	20.6

由表1可以看出,多义词的比例虽然不大,但使用频度却较高,多义选择很有必要.值得指出的是,这一比例低于通用词典和普通文本语料库是因为我们所使用的词典是限定领域的.

2.2　语法类选择

口语中没有逗号、句号等明显的分隔标志,因而无法像书面语那样以句子为单位进行分析.为此,我们引入了话轮(dialog turn)[9]的概念.一个话轮是指一个说话人不被打断地、连续说出的一段话.本文所实现的选择算法以话轮为单位,选择时只使用话轮内上下文信息.

令网络在接收第p个输入$A^p=[a_1^p,a_2^p,\ldots,a_m^p]$时,相应各层的输出为:隐藏层$H^p=[h_1^p,h_2^p,\ldots,h_k^p]$,上下文层$B^p=[b_1^p,b_2^p,\ldots,b_k^p]$,输出层$O^p=[o_1^p,o_2^p,\ldots,o_n^p]$.若网络处在学习阶段,与$A^p$相应的理想输出$T^p=[t_1^p,t_2^p,\ldots,t_n^p]$.

上下文层的值等于网络处理前面一个样本时隐藏层的值,所以,当$p\geqslant 2$时,有$B^p=H^{p-1}$.而当$p=1$时,b_i^1被赋予$(-1,+1)$区间内的初始值.

网络的输入层和上下文层组成有$k+m$个神经元的联合输入层,当接收第p个输入时,联合输入层的网络总输入为$C^p=[B^p,A^p]=[b_1^p,b_2^p,\ldots,b_k^p,a_1^p,a_2^p,\ldots,a_m^p]=[c_1^p,c_2^p,\ldots,c_{k+m}^p]$.

对于f个元素的语法类集合$W=\{w_1,w_2,\ldots,w_f\}$(本系统中$f=26$),有$m=f,n=m$,输入层和输出层神经

元分别与语法类集合 W 中的元素一一对应. 隐藏层和上下文层神经元数目 k 根据经验确定.

对于有 q 个词的话轮 $S=[s_1,s_2,\ldots,s_q]$，设 S 的正确标记串为 $R=[r_1,r_2,\ldots,r_q]$，其中 $r_p\in W(p=1,2,\ldots,q)$. 对于词 $s_p(1\leqslant p\leqslant q)$，其所有可能标记的集合 $W^p=\{W_{d_1},W_{d_2},\ldots,W_{d_h}\}$，$W^p\in 2^W$ 且 $W^p\neq\varnothing$，W^p 中的元素是由词典中该词的所有词类确定的. 在话轮中，词 s_p 的正确标记为 r_p.

当系统接收 s_p 时，与之对应的网络输入向量 $A^p=[a_1^p,a_2^p,\ldots,a_m^p]$ 的分量 a_i^p 按如下方法构造. 若词典中 s_p 有语法类 w_i(即 $w_i\in W^p$)，则 $a_i^p=1$，否则 $a_i^p=0$. 此时，网络理想输出 $T^p=[t_1^p,t_2^p,\ldots,t_n^p]$ 的分量 t_i^p，按如下方法构造. 若 s_p 的正确标记为 w_i(即 $w_i=r_p$)，则 $t_i^p=1$，否则 $t_i^p=0$.

系统的运行分为学习过程和选择过程. 在学习过程，采用引入动量项的 BP 算法[10]作为基本学习算法，训练集由经过人工标注和校对的汉语口语对话组成. 在选择过程，网络使用学习过程得到的网络参数. 第 p 个词 s_p 产生网络输入 A^p，经过计算得到相应的网络输出 $X^p=[x_1^p,x_2^p,\ldots,x_g^p]$. 若只单独考虑语法类选择，令 $i=\operatorname{argmax}_i(x_i^p)$，则 s_p 的语法类为 w_i.

2.3 语义类选择

语义类选择的过程与语法类选择类似，但其中各层神经元的个数需作相应调整. 对于 g 个语义类的集合 $V=\{v_1,v_2,\ldots,v_g\}$(本系统中的 $g-37$)，有 $m=g,n=m$.

采用训练好的网络进行语义类选择时，第 p 个词 s_p 输入时得到网络输出 $Y^p=[y_1^p,y_2^p,\ldots,y_g^p]$. 若只单独考虑语义类选择，令 $j=\operatorname{argmax}_j(y_j^p)$，则 s_p 的语义类为 v_j.

2.4 语法、语义的一致化处理

分别用两个 SRN 对语法、语义独立进行选择，难免会出现不一致问题. 由于同一词条的任意两个义项的语法和语义分类值不同时相等，所以可由语法语义对 (w_i,v_j) 来唯一标志词条的每一个义项. 词 s_p 的 u 个义项组成义项集：$Z_{s_p}=\{(w_{d_1},v_{e_1}),(w_{d_2},v_{e_2}),\ldots,(w_{d_u},v_{e_u})\}$. 其中对于任意 $1\leqslant i\leqslant u$，有 $1\leqslant d_i\leqslant f,1\leqslant e_i\leqslant g$. 且对于任意 $i\neq j,1\leqslant i,j\leqslant u$，有 $w_{d_i}=w_{d_j}\Rightarrow v_{e_i}\neq v_{e_j}$. 对任意 $1\leqslant i\leqslant f,1\leqslant j\leqslant g$，若语法语义对 $(w_i,v_j)\in Z_{s_p}$，则 (w_i,v_j) 对于词 s_p 是一致的，若 $(w_i,v_j)\notin Z_{s_p}$，则 (w_i,v_j) 对于词 s_p 是不一致的. 在词典填写无误时，若为词 s_p 选择的语法语义对 (w_i,v_j) 对于 s_p 是不一致的，则意味着语法和语义的选择至少有一个是错误的，需一致化处理.

一个正确选择，必须首先保证语法语义对是一致的. 为确保一致性，决定综合语法选择输出向量 X^p 和语义选择输出向量 Y^p 来为词 s_p 的每一义项 $(w_{d_i},v_{e_i})\in Z_{s_p}$ 构造评价函数. 与 w_{d_i} 对应的语法选择的网络输出为 $x_{d_i}^p$，与 v_{e_i} 对应的语法选择的网络输出为 $y_{e_i}^p$，求它们的加权和，得

$$o_i^p=\alpha\cdot x_{d_i}^p+(1-\alpha)y_{e_i}^p. \tag{1}$$

加权系数 α 根据经验确定. 令

$$i=\operatorname{argmax}_i(o_i^p), \tag{2}$$

则选定 (w_{d_i},v_{e_i}) 为 s_p 的义项. 由于 $(w_{d_i},v_{e_i})\in Z_{s_p}$，所以它对于词 s_p 必然是一致的.

2.5 实　验

对第 2.1 节所述的口语对话进行人工标注，得到训练和测试用语料. 从中随机抽取 80 段作为训练集，其余 70 段作为开放测试集. 先用训练集分别训练用于语法和语义选择的两个网络. 然后分别用训练好的两个网络对语法选择和语义选择进行测试，并在确定加权系数 α 后，对多义选择的综合结果进行测试. 测试时，分别用训练集和开放测试集得到封闭和开放测试结果. 测试结果包括消歧率和准确率.

$$\text{消歧率}=\frac{\text{作出正确选择的词数(不含无多义词)}}{\text{语料中需要选择的总词数(不含无多义词)}}\times 100\%$$

$$\text{准确率}=\frac{\text{作出正确选择的词数(包括无多义词)}}{\text{语料中的总词数}}\times 100\%$$

其中，“语料中需要选择的总词数”在语法、语义和多义选择中分别指语料中语法兼类的总词数、语义兼类的总词数和多义词总数. 表 2 列出了单独语法、语义选择及最终的多义选择 3 种结果.

表2 多义选择实验结果

	封闭测试		开放测试	
	消歧率(%)	准确率(%)	消歧率(%)	准确率(%)
语法选择	89.4	98.3	84.0	97.5
语义选择	91.5	98.2	85.7	97.1
多义选择	93.0	98.3	87.3	96.9

在实验中,用于语法类选择的网络隐藏层神经元数目为39,用于语义类选择的网络隐藏层神经元数目为50.在奔腾MMX166(32M内存)机器上,多义选择速度约为650词/s.

在进行多义选择时,加权系数α的确定过程如下:

① α将的初始值置为0.2,置当前加权系数$\beta=\alpha$,当前准确率$\gamma=0$,α的变化步长$\delta=0.01$;

② 根据当前的α值进行多义选择,并计算准确率μ;

③ 若$\mu>\gamma$,则$\beta=\alpha$,$\gamma=\mu$;

④ $\alpha=\alpha+\delta$;

⑤ 如果$\alpha\leqslant 0.8$,转②;

⑥ $\alpha=\beta$,结束.

将上述过程用于封闭集和开放集,分别得到α值为0.42和0.41,相应的准确率见表2.

3 结 论

通过观察和分析实验结果,发现口语分析有如下几个特点.

(1) 在表2中,鉴于需要选择的总词数不同,准确率可比性不强,而消歧率则可较好地反映算法能力.语义选择的消歧率高于语法选择,说明口语分析的特有困难对语义分析的影响较小,语法上不完整的话语在语义上往往是完整的.多义选择整体消歧率好于语法、语义各自独立选择的消歧率,这说明一致化处理策略切实有效,它不但保证了语法和语义的一致,还使消歧率有所提高.

(2) α值小于0.5表明,当语法语义选择结果不一致,需加权求和时,语义权重更大,这也从另一侧面说明了口语中语义分析的结果比语法分析更可靠.

(3) 使用本文的方法取得了良好的效果可以归结为如下原因:具有良好鲁棒性和容错能力的神经网络方法确实能较好地处理包含大量不连贯现象和不符合语法现象的口语;限定明确的应用领域效果显著.

(4) 本文的研究是在限定应用领域的条件下展开的.目前,无限制的、任意内容的口语分析还无法实现[1,2,4].但神经网络较强的学习能力保证了本文的方法很容易向其他领域扩展,扩展时只需收集、加工相应的口语语料,重新定义与领域相关的语义集并填写词典即可.

关于本文的方法,尚有如下几个问题需要说明:

(1) 标准的SRN只记忆和使用上文信息(当前输入之前的所有输入信息),而未使用下文信息,为了能使用下文信息,一些学者提出了改进措施[7].本文没有采用改进措施是基于如下原因:利用下文信息将使时空开销大大增加,如文献[7]中的方法使输入层神经元个数增加了7倍;利用下文信息将影响实时性.实验结果表明,即使未使用下文信息,准确率相当高,可以满足口语分析的需要.

(2) 在网络训练和运算中不但要处理多义词,而且无需选择的非多义词也要逐一处理,增加了系统开销.但由于上文的非多义词构成了多义选择的根据,而上文信息记忆于神经网络内部状态(上下文层).因此,非多义词参加训练和运算是必要的.好在系统的多义选择速度是完全可以接受的.

本文提出的基于SRN的汉语口语多义选择方法已应用于面向会面安排的汉英口语翻译系统,效果令人满意.

参考文献

1 Waibel Alex. Interactive translation of conversational speech. IEEE Computer, 1996,29(7):41~48

2 Kitano Hiroaki. Speech-to-Speech Translation: A Massively Parallel Memory-based Approach. Hingham, MA, Kluwer Academic Publishers, 1994

3 Lavie A. GLR*: a robust grammar-focused parser for spontaneously spoken language [Ph. D. Thesis]. Pittsburgh, PA: Carnegie Mellon University, 1996

4 Wermter S, Weber V. SCREEN: learning a flat syntactic and semantic spontaneous language analysis using artificial neural networks. Journal of Artificial Intelligence Research, 1997,6(1):35～85

5 Elman Jeffery L. Finding structure in time. Cognitive Science, 1990,14(2):179～211

6 Servan-Schrieber D, Cleeremans A, McClelland L. Graded state machines: the representation of temporal contingencies in simple recurrent networks. Machine Learning, 1991,7(2～3):161～194

7 刘伟权，钟义信. 基于 SRNN 神经网络的汉语文本词类标注方法. 计算机研究与发展，1997,34(6):421～426
(Liu Wei-quan, Zhong Yi-xin. Part-of-speech tagging with simple recurrent neural network. Computer Research and Development, 1997,34(6):421～426)

8 王海峰，高文，李生. 面向受限领域的汉英口语翻译. 见：陈力为，袁琦编. 语言工程. 北京：清华大学出版社，1997. 219～224
(Wang Hai-feng, Gao Wen, Li Sheng. Chinese-English spoken language translation in limited domain. In: Chen Li-wei, Yuan Qi eds. Language Engineering. Beijing: Tsinghua University Press, 1997. 219～224)

9 何自然. 语用学概论. 长沙：湖南教育出版社，1988
(He Zi-ran. A Survey of Pragmatics. Changsha: Hu'nan Education Press, 1988)

10 Rumelhart D E, McClelland J L. Parallel Distributed Processing, 1: Foundations. Cambridge, MA: MIT Press, 1986

Word Sense Disambiguation of Spoken Chinese Using Neural Network

WANG Hai-feng　GAO Wen　LI Sheng

(*Department of Computer Science and Engineering　Harbin Institute of Technology　Harbin*　150001)

Abstract　Spoken Chinese analysis lies in the center of interactive speech processing system. The smallest meaningful unit in Chinese language is the word, so word sense disambiguation is the basis of spoken Chinese analysis. In this paper, the authors propose a novel method for spoken Chinese word sense disambiguation based on a simple recurrent network. This method provides a consistent processing strategy for syntax and semantics according to the internal logic between the word syntactic classification and semantic classification. Applied in the corpus for meeting schedule, this method achieves an accuracy of 96.9% in an open testing of word sense disambiguation.

Key words　Neural network, simple recurrent network, spoken language analysis, spoken Chinese, word sense disambiguation.