维普资讯 http://www.cqvip.com

语料 词类分析 语言学模型 数据驱动 (18)

计算机科学1999Vol. 26№. 1

69-74

基于数据驱动技术在语料词类分析中的应用*)

The Application of Technic Based on Data-Driven in Part-of-Speech Analysis of Coprus

张玥杰 姚天顺

TP391

(东北大学信息科学与工程学院 沈阳 110006)

Abstract A very important aspect of corpus linguistics is performing the part-of-speech analysis of the corpus and tagging it. This is the key for making the corpus machine-readable and improving the using value. In this paper, we will present the part-of-speech analysis method based on data-driven technic in more detail. We hope the introduction will be helpful to the other researchers in this field.

Keywords Corpus linguistics, Data-driven, Part-of-speech analysis, Part-of-speech disambiguation

语料库语言学的主要研究内容之一是对所建立语料库中的语料，进行不同层次的语法分析。语法分析一般由两个层次组成，即词类分析(Part-of-Speech Analyzing)与句法分析(Parsing)。词类分析通常包括两个过程：(i)引入歧义(即，词法分析过程)；(ii)消除歧义(即，排除非法选择的过程)。经过词类分析之后，按语料库中每一个单词在句子中的句法功能，利用已建立的词类标记，将其加标。词类分析构成进一步对语料库进行句法分析的重要基础，其精确率对此后句法分析的正确率具有很大影响。词类分析工作多由词类分析器(Part-of-Speech Analyzer)自动完成。

在语料自动词类分析方面已有大量研究，主要存在两种方法：

1)基于语言学的方法。是劳力密集的，在所给定的上下文中，需使用大量手工制作的语言学约束，为所给出的单词消除不可能的词类标记或词法分析。

2)基于数据驱动的方法。使用大量语料库来训练概率模型，然后将其用于分析新文本，在所给出的上下文中为每一个单词分配最可能的词类标记。在这种方法中，不需要人工努力来书写约束规则。然而，需要相当程度的努力来决定所使用的标记集合，并注释训练语料库。

基于语言学的方法看起来是显然的选择，但存在的主要问题是：大规模解决词类歧义很困难，在一定程度上没有引入错误限制，基于数据驱动的系统依赖于关于词或标记的短的序列的统计概括，尽管这些系统通常不利用关于长距离现象或语言学家抽取能力的信息(例如，关于上下文什么是相关的知识)。根据已知系统的精确率，该方法提供词类分析的最佳模型。该方法在词类分析中的成功，已经对基于语言学方法的可行性形成一种反对。

本文对利用基于数据驱动技术的词类分析方法做较为详细的介绍。希望通过这些介绍，使我国进行汉语及其它外国语语料库词类分析的研究者有所借鉴。

1. 基于数据驱动技术的词类分析方法的分类

基于数据驱动的词类分析方法很多，就是否需要预先经过人工标注的语料库进行训练的词类分析器而言，可分为无监督学习型与有监督学习型。就语料库信息的表示形式而言，可分为局部规则、神经元网络及隐马尔可夫模型三大类。

1.1 无监督学习型与有监督学习型

无监督学习型词类分析方法不依赖于预先已分析的语料库作为参照物，而是通过一些复杂的计算自动归纳出标记集合，并以此为基础计算出概率分析方法所需的概率数据。有监督学习型词类分析方法需要依靠预先经过人工分析的语料库作为参照

*) 本文受国家自然科学基金和国家教委博士点专项基金资助。张玥杰 博士研究生。姚天顺 教授、博士导师。

维普资讯 http://www.cqvip.com

物，生成用于词类分析操作全过程的工具，如标记集合，已分析的具有各种可能的词汇表(Lexicon)，词频表、标记序列概率表，等等。

这两种词类分析方法各具有优缺点，一般情况下，经过训练的有监督学习型词类分析器，在对其受训语域类型相同的语料分析时正确率很高。但是，如果需要标注语料的语域不同，则正确率就会降低。无监督学习型标注器没有这些限制，它不需要实现对作为参照物的语料库做费时费力的人工标注操作，便可以对语料进行词类分析。但是其缺点是，与有监督学习型的分析结果相比，其自动归纳的词类分析分类较为粗糙。

目前所实现的基于数据驱动技术的词类分析方法，大多数是采用有监督学习。因此，本文主要介绍有监督学习型的基于数据驱动技术的词类分析方法。

1.2 三种类型的词类分析方法

基于局部规则的方法是一种可训练的基于规则的词类分析器，具有类似于随机词类分析器的性能。这种分析器的训练是完全自动的，但又与可训练的随机标注器不同，在简单的非随机规则中可直接编码语言学信息。

基于神经元网络的方法是一种相当新的方法，可利用人工智能的神经元网络进行词类分析。其中根据前面及后面单词的词类标记，作为输入预测所分析单词的词类标记，利用一种类推驱动的学习过程训练分析器。

基于隐马尔可夫模型的方法使用统计的概率模型，也可以将其称为随机模型，在词类分析过程中，将上下文模型化为统计的，根据可标记的单词的词频或概率等大量统计数字进行概率计算，从而最终决定该单词的所属词类。训练过程是在大量的手工分析的语料库中进行。

2. 词类分析的任务描述

无论哪一类词类分析方法，一般都需要经过四个步骤：源文本的预处理、词类加标、未登录词处理和词类消歧。

2.1 预处理

首先，将文本中对词类分析无用的标记去除，根据段落起始标志及标点符号，将其分割为句子。其次，按照特定的规范，对每一个句子按相应语种分词单位进行划分，即分词。

2.2 词类加标

将经过预处理的文本作为下一步词类加标的输入，由词法分析器对该文本中每一个单词进行词类加标操作。在这期间，至少需要使用一种知识源：基本词典。在该词典中，每一个单词后面都标有该单词所对应的可能词类标记，可为一个或多个，例如，英语单词"below"，通常可以作为名词、介词及副词，则在词典中，它所对应的标记集合为：

"below"[NOUN(名词)，ADV(副词)，PREP(介词)]

在基于数据驱动技术的词类分析中，对于标记数目大于1的单词，基本词典所记录的标记顺序按照概率大小排列。单词所具有的各种词类标记的概率，是通过在人工标注的语料库中统计该标记使用情况来获得的。

2.3 未登录词处理

未登录词是指不能由词典识别的词汇。目前一般是采用"猜测"策略处理未登录词，猜测器利用对于单词特征的分析，例如，单词的领头及跟随字符，以指出其最可能的词类类别。最流行的猜测策略是所谓的"结尾猜测"，即，仅仅基于一个单词的跟随字符，猜测其可能的词类标记集合。未登录词处理策略主要分为两种：

- 基于猜测规则的方法。利用简单的非随机猜测规则直接编码语言学信息，基于未登录词词法特征将其激发，猜测其最可能的词类标记；
- 基于概率统计的方法。预先给定，如果一个未登录词具有一个具体的词类标记，其大写字母书写及结尾等重要特征的概率，通过已建立的随机数学模型，计算每一个未登录词的传播概率，来猜测其词类标记。

2.4 词类消歧

经过以上三个步骤处理之后的语料，其中每一个单词都具有一个或多个词类标记，第四个步骤就是对具有多个词类标记的单词进行词类消歧，并纠正带有错误的唯一标记的单词。这是进行词类分析的核心过程，也是前面所提及的三种词类分析方法的区别之处，下面将详细介绍三种词类分析方法的处理策略。

3. 词类消歧的实现

3.1 基于局部规则的方法

这种方法最典型的代表是[Brill，1995]中所介绍的基于变换的词类分析方法。在实现过程中，应用一种基于变换的错误驱动学习算法，如图1所示。

维普资讯 http://www.cqvip.com

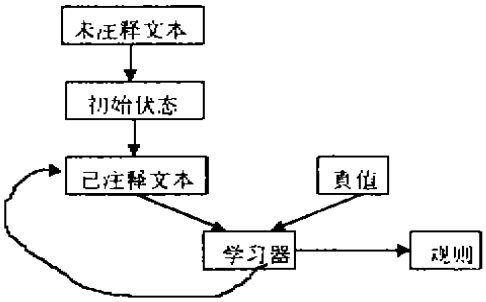

图1 基于变换的错误驱动学习

首先，将未标注的文本通过初始状态注释器。一旦文本已经过初始状态注释器，则将其与真值比较，并且学习可应用于初始状态注释器输出的变换，以使其更相似于真值。

3.1.1 一种早期的基于变换的词类分析器。最初的基于变换的词类分析器[Brill，1992]，工作如下。初始状态转换器将训练语料库中所说明的每一个单词最可能的标记分配给单词。所允许的变换模板为：

当：
1. 前面（后面）的单词标注为 z；
2. 之前（之后）第二个单词标注为 z；
3. 两个前面单词之一标注为 z；
4. 三个前面（后面）单词之一标注为 z；
5. 前面单词标注为 z，后面单词标注为 w；
6. 前面（后面）单词标注为 z，之前（之后）单词标注为 w

时，将标记 a 变为 b。其中 a，b，z 及 w 是词类集合中的变量。

对于学习变换，实质上学习器应用每一种可能的变换，计数在应用变换之后分析错误的数目，并选择导致错误最大限度降低的变换。当超出预先规定的阈值，没有发现其应用可减少错误的变换时，学习终止。学习变换的一个样本是：如果分析前面单词作为情态动词，则将当前单词词类标记从名词改变为动词。每当训练系统时，就通过应用初始状态转换器，之后依次应用每一种变换至新句子，进行词类分析。

3.1.2 词汇化词类分析器。为有效地利用许多有用的关系，如单词与前面单词之间，或一个标记与后面单词之间的关系，通过加入可参考单词及词类标记的上下文信息，将早期实现的基于变换的标注器扩展为一种新型的词汇化标注器。所加入的变换模板为：

当：
1. 前面（后面）单词为 w；
2. 之前（之后）第二个单词为 w；
3. 两个前面（后面）单词之一为 w；
4. 当前单词是 w，且前面（后面）单词是 x；
5. 当前单词是 w，且标注前面（后面）单词为 z

时，将标记 a 变为标记 b。其中 w 与 x 是训练语料库中所有单词中的变量，z 是所有词类中的一个变量。

以下是所学习的两种词汇化的变换：

1. 如果单词右侧第二个位置中是 as，将标记从介词（preposition）变为副词（adverb）；
2. 如果前面两个单词之一是 n't，将标记从第三人称单数现在时动词（non-3rd person singular present verb）变为原形动词（base form verb）。

Penn Treebank 标注风格人工规定，在排列 as…as 中，标注第一个 as 作为副词，标注第二个 as 作为介词。因为在训练语料库中经常将 as 标注作为介词，则初始状态标注器将错误地标注 as tall as 如下：as/preposition/tall/adjective as/preposition。

第一种词汇化变换纠正这种错误标注。第二种变换是从以下事实中产生，当一个动词出现于上下文如 We don't…或 We didn't usually…中时，动词应该是原形。

3.1.3 实验测试与结果。在 Penn Treebank 已进行词类分析的 Wall Street Journal 语料库中进行训练与测试，当在600,000个单词中通过早期变换模板集合训练分析器时，获得96.9%的精确率，其中这些模板不包含涉及单词的变换。当在后期改进的分析器中进行相同测试时，获得97.2%的精确率。可见，词汇信息对于词类分析精确率有很大影响。

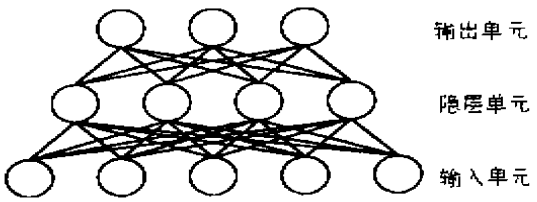

图2 三层感知器网络

3.2 利用神经元网络的方法

Schmid[1994]词类分析器是以人工智能的多层感知器网络（Multilayer perceptron networks，MLP-网络）为基础建立的，由 MLP-网络与词典组成。

3.2.1 MLP-网络。是最流行的网络类型，处理单元垂直排列在几层中。网络结构如图2所示。连接只存在于邻接层中的单元之间，底层作为输入层—其中单元激活表示网络输入。顶层作为输出层，输入层与输出层之间的任何层为隐层。它们的激活在外部不可见。

维普资讯 http://www.cqvip.com

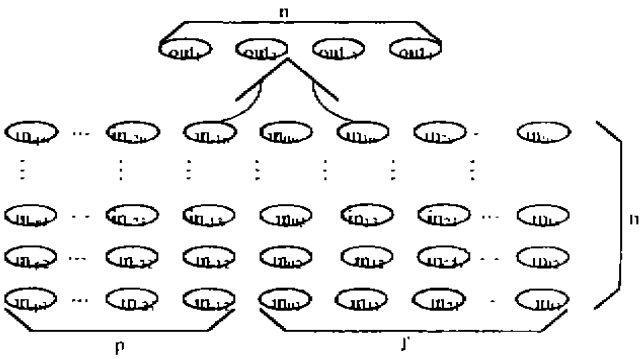

图3 Net-标注器的结构(箭头表示层之间连接)

在词类分析器中所采用的 MLP-网络结构，其中不具有隐层，如图3所示。在 MLP-网络的输出层中，每一个单元与词类标记集合中的标记之一相对应。在已训练的网络中，具有最高激活的输出单元指示，应该将哪一标记附加在当前所处理的单词上。

网络的输入由系统关于当前单词词类的所有信息组成，这些信息包括 p 个前面单词与 f 个后面单词。更为精确地，对于上下文中 p+1+f 个单词与每一个的词类标记 pos_j，存在一个输入单元，其激活 in_{ij} 表示单词 $word_i$ 具有词类 pos_j 的概率。

对于当前分析的及后面的单词，词汇词类概率 $P(pos_j|word_i)$ 是已知关于词类的所有信息。因此，对于这些单词，获得如下输入表示：

$$in_{ij}=P(pos_j|word_i)，如果 i\geqslant 0 \quad (1)$$

对于前面的单词，因为已将其分析，所以存在较多的可用信息，处理过程中使用输出单元的激活值替代词汇词类概率：

$$in_{ij}(t)=out_j(t+i)，如果 i<0 \quad (2)$$

复制输出激活至输入单元，将循环引入网络。这使训练过程变得复杂，因为当训练开始时，网络输出是不正确的。因此当训练开始时，不能直接反馈网络输出，代之以使用实际输出与目标输出的加权平均值。在训练开始时，目标输出的权重是高的，在训练期间降至0。

在已标注语料库中训练网络。与正确标记相一致的单元目标激活为1，除此之外的所有输出单元目标激活为0。如果输出单元 j 的激活与相应目标输出的激活之间的差别低于预先规定的阈值(例如，0.1)，将设置错误信号 δ_{pj} 为0。这样，强迫网络更为注意较大的错误信号，结果精确率比原值提高超过1%。

完成单词词类分析，是通过将当前单词及其邻居的概率拷贝至输入单元，通过网络将激活传播至输入单元，并决定具有最高激活的输出单元。然后将与该单元相对应的标记附加在当前单词上。

如果输出层中第二个最强壮激活接近于最强壮激活，给出与第二个最强壮激活相对应的激活，并将其作为一个可选择的输出，对此不要求额外计算。而且，给出所有标记的得分列表作为输出也是可能的。

3.2.2 词典，所使用的词典具有每一个单词的优先标记概率，它分为三部分：全形词典、后缀词典及缺省条目。利用网络-词类分析器在词典中进行查找期间，首先查找全形词典。如果找到单词，则返回相应的标记概率向量。否则将单词大写字母变为小写，在全形词典中重复查找。如果再次失败，继续查找后缀词典。如果两个步骤均未成功，则返回词典中的缺省条目。

全形词典是从训练语料库中产生的。首先计数每一个单词/标记对出现的次数。之后，去除每一个单词估计概率小于1的那些标记，因为它们在多数情形中是原始语料库中分析错误的结果。

词典的第二个组成一后缀词典，形成一棵树。利用字符标记树中每个结点(除根结点之外)。在叶结点处，附加标记概率向量。在查找期间，从根结点查找后缀树。在每一步骤中，都是从单词后缀的结尾开始利用下一字符标记的分支而行。例如，在后缀词典中查找单词 tagging，以根结点为起点，沿着导致标记为 g 的结点的分支。由此移至标记为 n 的结点，最终结束于标记为 i 的结点。该结点是一个叶结点，并返回所附加的标记概率向量。

3.2.3 实验测试与结果。在 Penn Treebank 语料库的2百万单词中训练两层网络-词类分析器，并测试其性能。网络参数设置如下：上下文中前面单词数目 p 为3，后面单词数目 f 为2，训练循环次数为4百万。分析器在 Sparc 10工作站上花费一天时间进行训练，最终获得96.22%的精确率。

3.3 基于隐马尔可夫模型的方法

隐马尔可夫模型(HMM)是由马尔可夫过程扩充而来的一种随机模型。本节将利用(Kempe，1994)中所描述的一种基于 HMM 模型的概率词类分析器，说明该类词类分析方法的实现过程。Kempe 采用词类、性别及数目等等特征的结构作为标记，通过连接标记的唯一特征值(fv，feature value)对的上下文概率，估计所使用的各种概率，由这些概率统计与

维普资讯 http://www.cqvip.com

标记完成相应词类分析过程。

3.3.1 数学背景。为将词类标记分配给单词序列，可将词类分析器在所有可能的标记序列中选择最可能序列的地方使用 HMM。给定一个单词序列 $w=w_0\cdots w_{N-1}$，一个标记序列 $t=t_0\cdots t_{N-1}$ 的连接概率是：

$$p(\bar{t},\bar{\omega})=\pi(t_{001})\cdot p(\omega_0|t_2)\cdot p(\omega_1|t_1)\cdot \prod_{i=2}^{N-1}(p(\omega_i|t_i)\cdot p(t_i|t_{i-2}t_{i-1})) \quad (1)$$

项 π_{001} 代表初始状态概率，例如，序列由前两个标记开始的概率。N 是序列中单词的数目，例如，语料库大小。项 $p(\omega_i|t_i)$ 是在已分配标记 t_i 的上下文中单词 ω_i 的概率。将其称为观察符概率（词汇概率），可由以下公式估计：$p(\omega t|tt)=\frac{f(\omega_i t_i)}{f(t_i)}$ (2)

公式(1)中的上下文概率 $p(t_i|t_{i-2}t_{i-1})$ 表示标记 t_i 出现于其前面两个标记 t_{i-2} 与 t_{i-1} 的上下文中的可能性。通常将其估计作为所给定训练语料库中三元组 $\langle t_{i-2},t_{i-1},t_i\rangle$ 频率，与相同语料库中二元组 $\langle t_{i-2},t_{i-1}\rangle$ 频率的比率：

$$p(t_i|t_{i-2}t_{i-1})=\frac{f(t_{i-2}t_{i-1}t_i)}{f(t_{i-2}t_{i-1})} \quad (3)$$

利用大的标记集合及一个相对小的手工制作的训练语料库，公式(3)具有一个明显的缺点：不能确切地估计多数转换概率，因为多数可能的三元组（三个连续标记的序列）根本不会出现或仅出现少数几次。

利用386个不同的标记集合标注10,000个单词的法语训练语料库，可形成$386^3=57,512,456$个三元组，但因为语料库尺寸仅为10,000-2个三元组。因为其中一些出现多次，实际上数目仅4,815，占所有可能的0.008%。当根据公式(3)由二元组频率2分割三元组频率1时，获得概率 p=0.5，但由于作为基础的频率特别小，则它并不精确。可利用386个标记由词类、性别、数量等等57个不同的 fv-对构造的事实。如果考虑唯一 fv-对之间的概率关系，则获得更高的频率，并且结果概率是更为确切的。

从等式：$\{t_i\}=\{e_{i0}\cap e_{i1}\cdots e_{i,n-1}\}=\{\bigcap_{k=0}^{n-1}e_{ik}\}$ (4)

其中 t_i 表示一个标记，e_{ik} 表示其 fv-对，及：

$$p(\bigcap_{k=0}^{n-1}e_{ik}|C_i)\cdot p(C_i)=p((\bigcap_{k=0}^{n-1}e_{ik})\cap C_i)$$
$$=p(C_i)\cdot p(e_{i0}|C_i)\cdot p(e_{i1}|C_i\cap e_{i0})\cdots\cdots p(e_{i,n-1}|C_i\cap\bigcap_{k=0}^{n-2}e_{ik}) \quad (5)$$

其中 C_i 表示 t_i 的上下文，并且包含遵循以下公式的 t_{i-2} 与 t_{i-1}：

$$p(t_i|C_i)=p(e_{i0}|C_i)\cdot\prod_{k=1}^{n-1}p(e_{ik}|C_i\cap\bigcap_{j=0}^{k-1}e_{ij}) \quad (6)$$

后者描述一个标记的上下文概率与其 fv-对的上下文概率之间的关系。

间接完成名词短语内部词法特征的合一。在所给定 fv-对的上下文中，正确的 fv-对包含概率 p=1，因此将不会影响它所属标记的概率。一个错误的 fv-对将包含 p=0，使整个标记变为不可能。

3.3.2 训练算法。在训练过程中，未着重分析并存储整个标记的上下文概率（状态转换概率），但着重于唯一 fv-对的上下文概率。根据概率特征关系(PFR)注意到这些概率：

$$PFR:\langle e_i|C_i^{sub};p(e_i|C_i^{sub})\rangle \quad (7)$$

后来在标注过程中，将其组合，以获得上下文标记概率。

公式(7)中的项 e_i 是 fv-对，C_i^{sub} 是减少的上下文，它仅包含确实出现的上下文 C_i 的 fv-对的一个子集。C_i^{sub} 是通过消除所有不影响 e_i 相关频率的 fv-对，从 C_i 中获得，根据条件：

$$p(e_i|C_i^{sub})/p(e_i|C_i)\in[1-\varepsilon,1+\varepsilon] \quad (8)$$

所考虑的 fv-对在完整及在减少的上下文中几乎具有相同的概率，例如，C_i 不提供比 C_i^{sub} 更多的关于 e_i 的信息。

PFR 生成—在训练过程中，首先从训练语料库中抽取一系列三元组，其中将标记分割为它们的 fv-对，由这些三元组，分别为每一个 fv-对 e_i 生成一系列 PFR。对于每一个三元组，我们生成其 fv-对的所有可能子集。如果三元组仅不同于一个 fv-对，一般具有多数它们 fv-对的子集。完整的三元组与子集一起组成上下文与子上下文的集合（C_i 与 C_i^{sub}）。其中一个 fv-对可能出现。为一个所给定的 fv-对生成 PFR，预先选择并标记那些假设对于 fv-对上下文概率有影响的（子）上下文，如果一个（子）上下文频率小于一个已定义的阈值，将不预先选择。接下来从每一个预先选择的上下文中消除所有 fv-对（式8）。由从训练语料库中所抽取的三元组集合中，分别为每一个 fv-对生成一棵二进制分支的决策树，它能够描述该 fv-对的各种上下文概率。

3.3.3 标注算法。特征结构词类分析器的实现基于 Viterbi 算法，始于第二排序的 HMM 词类分析器（三元组）一种修正版本，早期采用 C 语言实现该算法。其中将通过由二元组频率分割三元组概率（式3）估计一个标记上下文概率的函数（状态转换概率），替换为使用上面所描述的方法（式6,7），或者通过查询决策树实现该计算函数。为估计标记的上下

维普资讯 http://www.cqvip.com

文概率，必须知道其fv-对的上下文概率，以求它们之间的乘积(式6)。

3.3.4 测试结果。在训练与测试过程中，利用不同的值实验，针对参数如：对于预先选择最小公认的频率，考虑作为相等概率之间的百分比距离ε，当在6,000个单词语料库中进行测试时，其中平均歧义为每个单词2.63个词类标记(在词典查找之后)，最佳情形是88.89%的精确率。

结束语 目前，比较好的词类分析系统，其精确率一般都能够达到97%左右，但是由于自然语言非常复杂，不需人工参与而完全依靠词类分析器获得100%的精确率是不可能实现的。对于大型语料库来说，很小的错误率就会造成大量单词的词类分析是错误的。这对以其为基础的各项处理具有很大副作用，这需要耗费大量时间与人力解决这些错误。因此，最大限度地降低词类分析错误率是词类分析研究中的重要课题之一。

由上面介绍，可以看到，基于数据驱动技术的三类方法各有所长。

● 基于局部规则方法的一个优点是其变换规则很直观地反映了语言的结构特点，可以处理相距较远的单词或标记之间的关系。但对于自然语言中所存在的大量不规则现象，有些不能利用规则清晰概括，即使可以通过加入新规则加以概括，但又有可能与原有概括相矛盾，发生冲突，产生不良影响。

● 基于神经元网络方法所具有的优点是，容易探测到标记之间存在问题的决策，可在输出中给出附加的标记，也可将最终决策延迟至后来的处理阶段，例如，句法分析器，在小的训练语料库中利用这种方法实现，效果比较显著，其精确率同其它方法相比，相同或稍高一些。但存在一个明显劣势，处理速度较低。

● 基于HMM的方法，其优势为通过大规模分析语料库获得各种概率统计，为进行正确词类分析提供充足可靠的数据，而且在这种方法中不存在规范化的特别要求，但也存在一个缺点，即不能从大量概率统计中直接获取对语言学现象的总结，不能够处理相距较远的标记之间的关系。

今后的词类分析方法可能会向着各种方法互相综合的方向发展，从而取长补短，优劣互补，例如，基于局部规则的方法与基于HMM的方法相结合，基于局部规则的方法与基于神经元网络的方法相结合，等等。(参考文献共18篇略)

(上接第54页)

能再现高光，由于插值是基于描述物体表面朝向的法向量，所以生成的图形具有很强的真实感。在具体计算中(图3)，我们没有用cosα，而使用点积 $N \cdot H$，其中 N 是物体表面法向量，H 是 L 和 E 的平均向量再单位化，即：

$$H=(L+E)/|L+E|$$

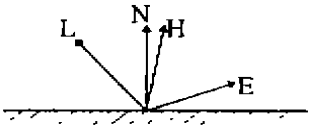

图3 Phong模型所涉及的几何量

图4为虚拟人的四个运动状态。

结论 本文就虚拟人的复杂几何模型和运动模型作了阐述，此模型目前已应用于863课题多功能感知机，实现了实时人机交互，包括语言、人体语言(表情、手势等)的理解和合成，由于现实生活中的人虽有其共性，但人的体形、相貌、动作是千差万别的，所以虚拟人要真正模拟人的动作，还需要解决虚拟人的几何模型和运动模型个性化问题，这也正是我们下一步要做的研究工作。

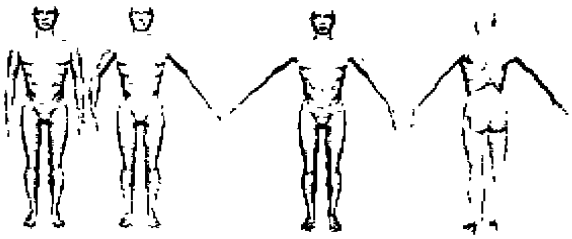

图4 虚拟人的几种运动状态

参考文献

1 Miller J V, Breen D E. Geometrically deformed models: A method for extracting closed geometric models from volume data. Computer Graphics (SIGGRAPH' 91 Proceedings), 1991, 25(4): 217～226

2 Michael L. Parametric surface interpolation. IEEE Computer Graphics and Applications, 1992, 12(5): 45～52

3 宋顺林，等. 三维计算机动画中的人体建模方法的研究. 软件学报, 1995, 6(5): 311～315

4 Hodgins J K, et al. Animating Human Athletics. In: Computer Graphics Proceedings, Annual Conference Series. Los Angeles, California, August, 1995. 71～78