

常用统计翻译模型在口语汉英翻译中的比较研究^{*}

李 俊, 薛永增, 赵铁军

(哈尔滨工业大学 计算机科学与技术学院 语音语言教育部微软重点实验室, 黑龙江 哈尔滨 150001)

摘 要: 通过汉语到英语的翻译实验以及对结果译文的分析, 对基于词的模型、基于短语的模型和基于句法的模型的翻译性能进行了比较。结果表明基于短语的模型性能优于其他两个模型, 但是使用的参数较多; 基于句法的模型虽然翻译性能不理想, 但可以用较少的参数表达更丰富的信息, 值得深入研究。

关键词: 自然语言处理; 统计机器翻译; 翻译模型; 句法分析器

中图分类号: TP391. 2 文献标志码: A 文章编号: 1001-3695(2007) 06-0069-03

Comparative Study of Statistical Translation Models on Chinese-English Speech Translation

LI Jun, XUE Yong-zeng, ZHAO Tie-jun

(MOE-MS Key Laboratory of Natural Language Processing & Speech, School of Computer Science & Technology, Harbin Institute of Technology, Harbin Heilongjiang 150001, China)

Abstract: According to the linguistics information, there are primary word-basd, phrased-based and syntax-based translation models: By analyzing and comparing the translation results, it 's found that the performance of the phrase-based translation model is the best. The syntax-based model is the worst, but it uses less parameters than the other two. It encodes rich information with a few parameters, so it 's deserved further research.

Key words: natural language processing; statistical machine translation; translation model; parser

机器翻译的目标就是将给定的一个源语言文本翻译成目标语言文本。对汉英翻译来说, 输入一个汉语句子 $c(c_1^m, m$ 为句子长度), 可能会有很多英语译文 $e(e_1^n, n$ 为句子长度), 统计机器翻译的任务是在所有可能的译文中, 找到最佳译文。根据 Bayes 公式可得到

$$e^* = \arg \max_e P(e|c) = \arg \max_e P(c|e)P(e) \tag{1}$$

式(1)包含了两方面的问题, 即建模和解码。其中, $P(e)$ 是语言模型(Language Model, LM); $P(c|e)$ 表示翻译模型(Translation Model, TM); 这里的 $\arg \max$ 表示解码问题。

早在 1949 年, Weaver 就提出利用统计方法研究机器翻译问题。其基本思想是把外语看成是对本地语言的一种编码, 而翻译过程就是对外语文章进行解码, 用本地语言表达同样的意思。20 世纪 90 年代初, IBM T. J. Watson 研究中心的 Brown 等人开创性地提出了词对词的统计翻译模型, 并以此为基础构建了 Candide 系统^[1]。此后对 IBM 模型比较重要的改进包括在 IBM 模型 2 的基础上提出基于隐马尔可夫模型的对齐模型(HMM-based Alignment Model), 以及基于 IBM 模型 4 和基于 HMM 词对齐模型的对数线性模型。在基于词的统计翻译模型基础上, 又相继提出了基于短语和基于句法的统计翻译模型。基于短语的统计翻译模型是目前研究的一个热点, 主要包括基于浅层短语结构的翻译模型、对齐模板(Alignment Template)模型、Koehn 的短语翻译模型、基于双语语块(Co-Chunk)的翻译模型等。基于句法的统计翻译模型由于引入了层次结构信

息, 有望处理长距离依赖和调序问题, 正逐渐成为新的研究热点。这类模型大致可以分为语言学驱动(Linguistically-motivated)的模型和非语言学驱动驱动的模型。前者依赖于句法分析树的指导, 如 Yamada 的树—串统计翻译模型、概率树替换文法(Probabilistic Tree Substitution Grammar)模型以及多文本文法模型(Multi-Text Gramar, MITG); 后者是无指导的, 在翻译过程中建立层次结构, 主要包括反向转换文法(Inversion Transduction Grammar, ITG)模型、中心词转录机(Head Transducer)模型、层次化短语翻译模型(Hierarchical Phrase-based Model)等。此外还有一类模型, 利用句法信息来抽取非层次化的短语翻译等价对, 可以看做是介于短语和句法翻译模型之间的一类统计翻译模型。

1 翻译模型概述

1.1 基于词的翻译模型

IBM 翻译模型是目前统计翻译模型研究的基础, 包括模型 1~5。其中模型 1、2 是基于对齐的模型; 模型 3~5 是基于繁殖数的模型。

模型 1、2 是假设英语句子中的每个单词都与汉语句子的一个或多个词存在着对应关系, 具体描述为^[2,3]

$$P(c|e) = \sum_a \prod_{(e,c)} P(c,a|e) \tag{2}$$

其中, $\Lambda(e, c)$ 代表汉语句子 $c = c_1^m = c_1, c_2, \dots, c_m$ 与英语句子

$e = e_1^n = e_1, e_2, \dots, e_n$ 所有可能的对齐关系的集合(m 和 n 分别为汉语和英语句子的长度); a 为对齐序列, $a_1^m = a_1 a_2 \dots a_m$, a_i 的取值范围是 $0 \sim n$ 。当汉语句子中第 i 个位置的词 c_i 不与任何一个英语句子中的单词对应时, $a_i = 0$ 。

在翻译时, 先在给定英语句子的前提下确定对应汉语句子的长度, 再确定英语句子 e 中哪个位置对应到汉语句子中 c_1 的位置, 并且进一步确定对应的汉语词; 接着确定英语句子 e 哪个位置与 c_2 相对应, 依此类推。为了不失一般性, $P(c, \alpha|e)$ 可精确地表示为

$$P(c, \alpha|e) = P(m|e) \prod_{j=1}^m P(a_j | a_1^{j-1}, c_1^{j-1}, m, e) P(c_j | a_1^j, c_1^{j-1}, m, e) \quad (3)$$

其中, $P(m|e)$ 是句长概率, 表示英语句子长度对汉语句子长度的影响; $P(a_j | a_1^{j-1}, c_1^{j-1}, m, e)$ 是对齐概率, 表示英语句子中对应的单词在句子中可能的位置对汉语句子的影响; $P(c_j | a_1^j, c_1^{j-1}, m, e)$ 是词汇翻译概率, 表示英语单词本身对汉语句子的影响。

模型 3 ~5 中引入了如下几个概念^[11]:

(1) 繁殖(Fertility) Φ 。它是英语句子 e 中单词 e_i 所对应的汉语词的个数。 $\Phi = \{ \phi_i | i = 1, \dots, l \}$ 。其中, ϕ_i 表示第 i 个英语单词的繁殖(包括对空的情况, 即 $\phi_i = 0$)。

(2) 语言片(Tablet)。它是每个英语单词所对应的汉语词串(包括空集合)。

(3) 语言片的集合(Tableau) T 。它是英语句子对应的中文语言集合。 $T = \{ \tau_i | i = 1, \dots, n \}$ 。 τ_i 表示第 i 个英语单词对应的汉语语言片, 第 i 个语言片中的第 k 个汉语词可记为 τ_{ik} 。

(4) 位置集(Permutation) Π 。它是语言片集合中的单词在英语句子中位置的集合。 $\Pi = \{ \pi_{ik} | i = 1, \dots, n; k = 1, \dots, n_i \}$ 。 π_{ik} 表示第 i 个语言片中的第 k 个单词在英语句子中的位置; n_i 为第 i 个语言片的长度。

基于上述假设, 式(3) 可重写为如下的形式:

$$P(c, \alpha|e) = \sum_{(\tau, \pi)} P(\tau, \pi|e) \quad (4)$$

其中, 语言片和位置的联合概率为

$$P(\tau, \pi|e) = \prod_{i=1}^n P(\phi_i | \phi_1^{i-1}, e) P(\phi_0 | \phi_1^i, e) \times \prod_{i=0}^n \prod_{k=1}^i P(\tau_{ik} | \tau_{i1}^{k-1}, \tau_0^{k-1}, \phi_0^n, e) \times \prod_{i=1}^n \prod_{k=1}^i P(\pi_{ik} | \pi_{i1}^{k-1}, \pi_1^{k-1}, \pi_0^n, \phi_0^n, e) \times \prod_{k=1}^0 P(\pi_{0k} | \pi_{01}^{k-1}, \pi_1^n, \tau_0^n, \phi_0^n, e)$$

其中, $\tau_1^{k-1} = \tau_{i1}, \dots, \tau_{ik-1}$, $\pi_1^{k-1} = \pi_{i1}, \dots, \pi_{ik-1}$ 。确定了参数 τ 和 π 就基本确定了词串与它们的对应关系。

模型 4 不仅考虑了繁殖概率, 还将语言片作为一个整体进行考虑。模型 5 在模型 4 基础上进一步扩展, 不仅考虑了当前对位状况, 还考虑了对位历史情况, 因此是一个无缺陷的模型; 但模型过于复杂, 对齐的效果与模型 4 相差不多。在实际的翻译应用中采用模型 4 就可以了。

1.2 基于短语的翻译模型

基于短语的模型在基于词的模型的基础上引入了上下文信息, 其核心是短语的抽取与评分。通常情况下, 短语的抽取是基于词对齐的结果; 而在 IBM 的模型中, 对于每一个汉语词, 只允许最多一个英语单词与之相对应。但是在实际翻译中存在多对多的情况, 需要进行一些转换。短语抽取的启发式处理过程如下^[4, 5]:

(1) 从中英文的平行语料中获得两种词对齐表, 即汉语到英语的对齐表和英语到汉语的对齐表:

(2) 从两个对齐表交集的词对齐开始。选择一个对齐 (e, c) , 从中英文词的邻节点 $(e\text{-new}, c\text{-new})$ 开始扩展。如果这两个词都没有对齐的目标, 并且 $(e\text{-new}, c\text{-new})$ 出现在并集中, 就扩展到短语中; 接着第二个, 依此类推, 直到没有可以扩展的词为止。由此可以获得短语的翻译概率。

对于抽取的短语片断, 要满足以下原则:

(1) 汉语短语片段中的每一个词对应的英语词都应该不能出现在与汉语短语对应的英语短语片段之外, 反之亦然;

(2) 对于汉语短语片段中的每个词, 对应的英语短语中不能没有英语词与之相对应, 反之亦然。

获得短语片断后, 可以用概率分布 $(c_i | e_i)$ 进行建模。为了简化, 可以用如下的公式来计算短语的翻译概率:

$$\phi(\overline{c_i} | \overline{e_i}) = \text{count}(\overline{c_i}, \overline{e_i}) / \sum \overline{c_i} \text{count}(\overline{c_i}, \overline{e_i})$$

对输出的英语短语需要进行排序, 可以使用相对位置概率分布 $d(a_i - b_{i-1})$ 进行建模。其中 a_i 代表被翻译成英语短语的汉语短语的开始位置; b_{i-1} 表示被翻译成第 $i-1$ 个英语短语的外文短语的结束位置。在应用中可以使用一个简化的扭曲模型 $d(a_i - b_{i-1}) = \alpha^{|a_i - b_{i-1} - 1|}$ 。这里要选择一个合适的参数来进行简化。为了校正输出句子的长度, 除了 Trigram 的语言模型 P_{LM} 外, 还引入了一个参数 ω 这个参数是大于 1 的。这样做是为了优化性能。

在解码时, 输入的汉语句子 c 被分割成 I 个短语的序列 c_1^I , 对于 c_1^I 中的每个短语 c_i 都被翻译成英语的短语 e_i , 这些对应的英语短语的顺序再使用扭曲模型进行调整。这样模型可以描述如下:

$$e^* = \arg \max_e P(e|e) = \arg \max_e P(c|e) P_{LM}(e) \omega^{\text{len}(e)} \quad (5)$$

其中, $P(c|e)$ 被分解成

$$P(\overline{c_1^I} | \overline{e_1^I}) = \prod_{i=1}^I \phi(\overline{c_i} | \overline{e_i}) d(a_i - b_{i-1})$$

1.3 基于句法的解释模型

基于句法的翻译模型有很多种。在实验中采用的模型是由 Melamed 提出并在 2005 年由 Johns Hopkins University 的 Workshop 实现的基于泛化的句法分析方法的统计机器翻译系统。

在自然语言处理中, 句法分析是一个推导语言学结构的算法。它可以分为几个部分: 文法、逻辑、搜索策略、终止条件等。对于输入的字符串集合, 句法分析器使用文法在字符串上进行句法分析的处理, 得到树型句法机构。各种句法分析器都有自己不同的方式和侧重点。Melamed 采用了语法来递增的推倒结构方法^[6, 7]。Melamed 在他提出的泛化的句法分析中, 引入了多维文法的概念 D-GMTG(Generalized Multitext Grammar, $D > 0$)。它是对 CFG 的一种泛化, 其文法为

$$G = (V_N, V_T, P, S)$$

其中, V_N 和 V_T 分别是不相交的终结符和非终结符的集合; $S \in V_N$ 是开始符; P 是产生式的有限集合。每一个产生式都具有这样的形式: $\alpha \rightarrow \beta$ 其中 α 和 β 都是 D 维的。

为了简化, 可以将 GMTG 限制成 CNF 的形式, 被称为 GCNF(Generalized Chomsky Normal Form)。在 GCNF 下, 每一个 GMTG 的产生式或者是终结产生式或者是非终结产生式。其

终结产生式和非终结产生式可分别表示如下:

$$\begin{array}{ccc} \dots & \dots & \\ \phi & \phi & \\ X \Rightarrow t & & \\ \phi & \phi & \\ \dots & \dots & \end{array} \quad \begin{array}{c} X_1 \\ \dots \\ X_D \end{array} \Rightarrow \left[\begin{array}{c} \pi_1 \\ \dots \\ \pi_D \end{array} \right] \left(\begin{array}{c} Y_1 \ Z_1 \\ \dots \ \dots \\ Y_D \ Z_D \end{array} \right)$$

其中, 每一维中的产生式都满足 CNF; ϕ 表示一个占位符; π_i 是一个优先数组, 表示其后的 $Y_i Z_i$ 间的优先关系。二维文法的获取可描述如下:

- (1) 对中英文的平行语料分别进行句法分析, 获取相应的句法分析树结构(也可以仅对英语(即目标)进行句法分析)。
- (2) 对中英文的平行语料进行词的对齐。
- (3) 用词对齐的结果, 利用句法分析后的树型结构进行层次化对齐, 进而可以抽取出二维文法结构。

泛化的句法分析就可以同时处理多个句对。输入多种语言的句对, 用维度 d 来表示。当 $d=1$ 时就是一般的句法分析器; 当 $d>1$ 时, 被称为多维句法分析器, 可以对输入句对中每一维的句子同时进行句法分析, 推导出树型结构(即句法分析树)。泛化的句法分析器不仅是一个多维的句法分析器, 而且当输入多维语句对的维数少于其文法的维数时, 对输入进行解码翻译; 反之则进行层次化的对齐, 训练获取多维文法, 为提取多维语法规则作准备。可以看出, 一般的句法分析只对一种语言的字符串序列进行处理, 是泛化句法分析的一个特例^[8, 9]。

2 对比实验及分析

2.1 对比实验

本文对基于词、基于短语和基于句法三类统计翻译模型的统计机器翻译系统的翻译性能进行比较, 通过实验来对比其翻译效果的差异。其中基于词的模型采用了 ISI 的 Rewrite-Decoder, 使用 IBM 模型 4 作为翻译模型; 基于短语的模型选用的是 Philipp Koehn 的 Pharaoh, 是在词对齐的基础上进行短语的抽取; 基于句法的模型采用的 GenPar 是由 Melamed 提出并在 2005 年由 Johns Hopkins University 的夏季 Workshop 中实现的泛化句法分析方法器。

语料选用了 IWSLT2004 的训练语料和测试语料。其中, 训练语料为 20 000 个中英双语句对; 测试语料为 500 句。英文语料共出现 185 713 个字, 平均字长为 9.3 个, 如表 1 所示。

评分方法采用了 BLEU 和 NIST。它们都是基于 n 元语法的机器翻译自动评测方法。其基本思想是, 将机器翻译产生的候选译文与人类翻译者提供的多个参考译文相比较, 越接近, 则候选译文的正确率越高。实验的目的是为了比较研究, 未进行最小误差率训练, 采用了系统的默认进行实验。通过实验, 评分结果如表 2 所示。

表 1 训练语料和测试语料的统计数据			
语料		中文	英文
训练语料	句子数	20 K	20 K
	词数	176 195	185 713
测试预料	句子数	500	
	词数	3 515	

表 2 三种统计模型的评分结果			
模型	BLEU	NIST	
基于词的模型	0.222 1	5.734	5
基于短语的模型	0.255 5	6.100	5
基于句法的模型	0.085 2	3.579	8

从表 2 中的数据可以得到: 基于短语的模型 BLEU 的分值为 0.255 5, 明显高于基于词模型的 0.222 1 和基于句法模型的 0.085 2。NIST 的评分结果也显示出相似的结果。

2.2 实验结果分析

通过对三种模型翻译结果的分析, 整体来看可以发现以下一些特点:

(1) 对于在训练语料的词表中未出现的词, 基于词模型和短语的模型都采用了输出源语言词(汉语词)的方法; 而基于句法的模型则是将这些词全部丢弃。这种现象导致了评分下降, 其原因是训练语料的规模较小所致。

(2) 基于词的模型和基于句法的模型都存在较多句子结构不完整的情况, 在基于句法的模型中表现更为明显。基于词的模型是逐词翻译, 不能发现句子的省略情况, 如对“可以试穿吗?”等句子的翻译就不能加上主语之类的成分; 基于句法的模型对于未出现的词, 无法用文法进行推导, 所以出现句子不完整的情况。

对于译文的质量, 通过对三种模型翻译结果的分析可以发现, 相对于基于词的模型, 基于短语的模型由于包含了词的上下文信息, 而不只是单个词, 有以下几个方面的优势:

(1) 词在短语中进行翻译时, 可以较好地处理词形态变化、词义等情况, 使翻译结果的正确性更高。例如: “两个晚上”在词模型中会被逐词翻译, 结果成了 “two a night”; 而基于短语的模型由于带有上下文信息就可以发现这种单复数的信息, 翻译结果为 “two nights”。“看看”在句子“请给我看看那个”的翻译时, 在词模型中翻译成概率较高的 “see”; 基于短语的模型就能提取出短语“给我看看那个”, 将之翻译成 “show”, 最终翻译成 “please show me that”。基于短语的模型的这种特点可以使译文更加准确。

(2) 短语能够固定词序, 得到正确的词序。例如“给父亲的礼物”, 基于短语的模型可以抽取出短语 “a gift for my father”; 而基于词的模型就出现了 “take father gift”这样的翻译结果。当训练数据较多时, 能获得的短语长度越长, 就更有助于解码。

与基于词的模型和基于短语的模型不同的是, 基于句法的模型能够包含更多的语言学信息; 在训练时获取句式特点, 如疑问句、被动句、“把”字句等, 能更好地处理句子的结构问题。例如“我觉得冷”, 在短语的模型中, 在进行短语抽取时将“我觉得”翻译为 “I feel”、“I think”等, 相应的“冷”的译文为 “it’s cold”的概率较高, 翻译结果就成了 “I think it’s cold”; 而基于句法的模型能在文法中体现句子的结构信息, 从而避免产生这种情况, 翻译结果为 “I feel chilly”。再如“有会讲日语的医生吗?”和“给您安排座位”在基于句法的模型中进行翻译时, 由于文法中能体现句式特点, 可以正确地翻译成 “do you have a Japanese speaking doctor?”和 “arrange a seat for you”; 而基于短语的模型却不能解决这种句子结构问题, 其结果为 “I’ll have a Japanese speaking doctor?”和 “give you arrange a seat”。

相对于基于词的模型和基于短语的模型, 基于句法的模型由于翻译时对出现的未登录词都采取了丢弃的方式, 造成翻译结果句子的长度过短, 在评分时受到的惩罚较大, 所以分值较低。在翻译较短的句子时, 基于句法的模型体现不出其优势。

3 结束语

通过对三类统计翻译模型在口语汉英翻译领域的对比实验可以得到: 基于短语的模型翻译性能最好, 基于(下转第 74 页)

交叉率 = 0.6, 变异率 = 0.15。图 5 给出了算法运行得到的最优结果, 运行时间 39.7 s, 目标函数值 = 24 974.4。实验结果表明, 本文所提出的模型及算法是有效的。

表 1 $n = 30$ 问题的几何数据

单元		纵 横 比		单元		纵 横 比	
ID	面积	下 限	上 限	ID	面积	下 限	上 限
1	10	1.0	1.43	16	2	1.0	1.5
2	8	1.0	1.0	17	4	1.0	2.5
3	5	1.0	1.43	18	2	1.0	1.9
4	6	1.0	2.0	19	8	1.0	1.0
5	12	1.0	1.1	20	10	1.0	1.15
6	4	1.0	1.67	21	4	1.0	2.0
7	2	1.0	1.43	22	5	1.0	1.1
8	4	1.0	1.0	23	8	1.0	1.67
9	15	1.0	1.25	24	1	1.0	1.1
10	12	1.0	2.0	25	4	1.0	1.25
11	5	1.0	1.43	26	1	1.0	2.0
12	1	1.0	1.25	27	4	1.0	1.25
13	2	1.0	1.5	28	1	1.0	2.0
14	3	1.0	1.33	29	8	1.0	1.05
15	5	1.0	1.1	30	4	1.0	1.1
1	10	1.0	1.43	16	2	1.0	1.5

地板尺寸: (47, 36)

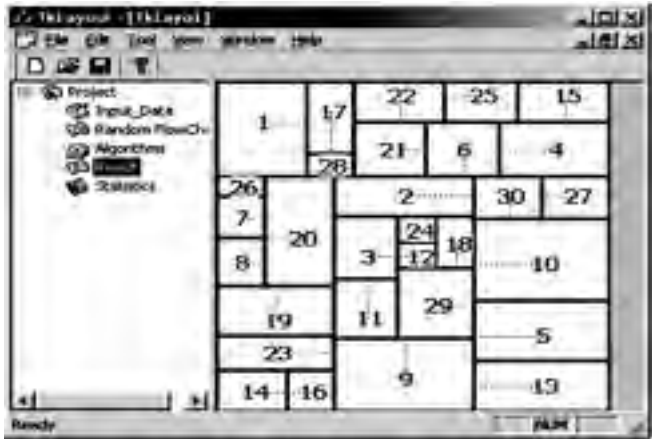


图 5 $n = 30$ 问题布局结果

(上接第 71 页)词的模型次之, 基于句法模型最差。相对于基于词的模型, 基于短语的模型由于带有词的上下文信息, 使单个词的翻译正确率更高; 短语结构由于可以固定词的顺序, 使译文的正确性提高。基于句法的模型相比其他两个模型来说虽然性能较差, 但是能很好地处理句子的结构问题。

基于句法的模型时间空间开销较大, 因为树型结构信息比较复杂; 而基于词和基于短语的模型相对较简单。可能在语料较小的情况下, 基于句法的模型训练获得的信息较少从而影响了翻译性能。从模型使用的参数个数来看, 基于词的模型和基于短语的模型使用的参数较多, 两者的模型都较复杂。实际应用中能够被可靠估计的参数个数要么受限于可用于训练的数据大小; 要么受限于可利用的计算资源。对全世界所有的语言来说, 适当的训练数据都是有限的, 即使对可利用的训练数据是无限的资源丰富的语言来说, 能够适合计算机内存的模型参数的个数也是有限的。相比之下, 基于句法的模型使用的参数最少, 能够用较少的参数表达丰富的句子信息, 值得进一步深入研究。

参考文献:

[1] 刘群. 统计机器翻译综述[J]. 中文信息学报, 2003, 17(4): 1-12.
[2] BROWN P, PIETRA S D, PIETRA V D, *et al.* The mathematics of statistical machine translation: parameter estimation[J]. Computa-

4 结束语

与以往文献相比, 本文编码方案易于设计遗传算子, 并保证遗传操作不产生非法子串, 且使进化搜索可以覆盖整个解空间。实验结果表明, 该模型及算法是有效的。

参考文献:

[1] ASSAN M M D. Layout design in group technology manufacturing [J]. International Journal of Production Economics, 1995, 38(2): 173-188.
[2] JAJODIA S, MINIS I, HARHALAKIS G, *et al.* CLASS: computerized layout solution using simulated annealing [J]. International Journal of Production Research, 1992, 30(1): 95-108.
[3] TAM K Y. Genetic algorithms, function optimization, and facility layout design [J]. European Journal of Operational Research, 1992, 63(2): 322-346.
[4] TAM K Y, CHAN S K. Solving facility layout problems with geometric constraints using parallel genetic algorithms: experimentation and findings[J]. International Journal of Production Research, 1998, 36(12): 3253-3272.
[5] SHAYAN E, CHITTILAPPILLY A. Genetic algorithm for facilities layout problems based on slicing tree structure [J]. International Journal of Production Research, 2004, 42(19): 4055-4067.
[6] AZADIVIR F, WANG J. Facility layout design using simulation and genetic algorithms [J]. International Journal of Production Research, 2000, 38(17): 4369-4383.
[7] NUGENT C E, VOLLMAN T E, RUML J. An experimental comparison of techniques for the assignment of facilities to locations [J]. Operations Research, 1968, 16(1): 150-173.

tional Linguistics, 1993, 19(2): 263-311.

[3] 程葳. 限定领域内汉英口语的统计翻译方法研究[D]. 北京: 中国科学院研究生院, 2003: 20-27.
[4] PHILIPP K, OCH F J, MARCU D. Statistical phrase-based translation: proc. of the Human Language Technology Conference and the North American Association for Computational Linguistics (HLT-NAACL) [C]. Edmonton: [s. n.], 2003: 127-133.
[5] KNIGHT K, PHILIPP K. What 's new in statistical machine translation: proc. of the Human Language Technology Conference (HLT) [C]. [S. l.]: [s. n.], 2003: 1-89.
[6] MELAMED I D. Multitext grammars and synchronous parsers: proc. of the Human Language Technology Conference and the North American Association for Computational Linguistics (HLT-NAACL) [C]. Edmonton: [s. n.], 2003: 158-165.
[7] MELAMED I D. Statistical machine translation by parsing: proc. of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL) [C]. Barcelona: [s. n.], 2004.
[8] YAMADA K, KNIGHT K. A syntax-based statistical translation model: proc. of the 39th Annual Conference of the Association for Computational Linguistics[C]. Toulouse: [s. n.], 2001.
[9] YAMADA K, KNIGHT K. A decoder for syntax-based statistical MT: proc. of the 40th Annual Conference of the Association for Computational Linguistics[C]. Philadelphia: [s. n.], 2003: 303-310.