一种基于词对齐的中文深层语义解析模型

郑晓东[1],胡汉辉[1],赵林度[1],吕永涛[2]
ZHENG Xiaodong[1], HU Hanhui[1], ZHAO Lindu[1], LV Yongtao[2]

1.东南大学 经济管理学院,南京 211189
2.东南大学 计算机科学与工程学院,南京 211189
1.School of Economics and Management, Southeast University, Nanjing 211189, China
2.School of Computer Science and Engineering, Southeast University, Nanjing 211189, China

ZHENG Xiaodong, HU Hanhui, ZHAO Lindu, et al. Word alignment-based Chinese deep semantic parsing. Computer Engineering and Applications, 2017, 53(20):8-13.

Abstract: Semantic parsing is the task of transforming natural-language sentences into complete, formal, symbolic Meaning Representations (MR) suitable for reasoning or machine-understanding. In recent years, the research of semantic parsing in English has made great progress. However, little work has been done in Chinese semantic parsing. There are inherent differences between Chinese and English, therefore one cannot simply apply methods, which are feasible for English, to Chinese. This paper proposes a statistical approach called WACSP aiming at Chinese semantic parsing, which considers the process of converting Chinese sentence into its corresponding meaning as a machine translation procedure. At first, it turns the frequently-used dataset GEOQUERY into Chinese dataset, in which each data contains a Chinese sentence and its accurate meaning. Then it uses the word alignment model to acquire the bilingual dictionary made up by the Chinese natural language string and its meaning. In the end, it determines the ultimate semantic analysis by learning a statistical model. Experimental results show that WACSP performs well with higher precision and coverage.

Key words: natural language processing; semantic parsing; word alignment model

摘 要:语义解析是指将自然语言句子转化成便于机器理解和推理的意义形式。近年来英文语义解析的研究取得了很大进展。然而,中文语义解析的相关工作则相对较少。中文和英文之间存在一定的差异,适用于英文的语义解析方法不一定适合中文。因此,针对中文的语言特点,提出一种基于词对齐的中文语义解析方法,将中文句子转化成其相应的意义表示看作是一个机器翻译的过程。首先将英文语义解析方法中常用的训练数据集GEOQUERY转化成中文数据集,数据集中每条训练数据包括一个中文句子及其正确的意义表示。然后利用词对齐模型来获取由中文自然语言字符串及其相应的意义表示所组成的双语词典。最后通过学习一个概率估计模型来确定最终的语义解析模型。实验结果表明,WACSP有较高的精确度和覆盖率。

关键词:自然语言处理;语义解析;词对齐模型
文献标志码:A **中图分类号:**TP39 doi:10.3778/j.issn.1002-8331.1707-0132

1 引言

语义解析是将自然语言句子转化成便于机器理解和推理的意义表示(MR),它从线性的词语序列中获取潜在的语义结构。意义表示语言(MRL)是一种形式化表示语言,可确保每一个意义表示(MR)有唯一的解析树。随着分词、词性标注和句法解析等自然语言处理技术的逐步成熟,浅层语义解析已得到广泛研究和应用。由于浅层语义解析的局限性,以及问答系统、信息抽取、机器翻译和机器人控制等领域的应用需求,使得深层语义解析越来越受到重视。

深层语义解析技术当前处于探索研究阶段,且大多数针对英文。如Ge等2005年提出基于句法的语义解

基金项目:国家自然科学基金面上项目(No.70673010)。
作者简介:郑晓东(1976—),男,博士,高级工程师,研究领域为信息处理、知识管理、系统工程,E-mail:51847986@163.com;胡汉辉(1956—),教授,博导;赵林度(1965—),教授,博导;吕永涛(1991—),硕士。
收稿日期:2017-07-10 **修回日期:**2017-08-25 **文章编号:**1002-8331(2017)20-0008-06

析方法SCISSOR[1]，缺点是需手动构造带有语义标签的句法解析树作为训练语料，代价很大。李等人在2015年尝试用组合范畴文法(Combinatory Categorical Grammar，CCG)进行语义解析[2]，该模型使用词典归纳过程归纳CCG词典，缺点是需要人工手写规则。Kate等后来提出基于字符串核函数的语义解析算法KRISP[3]，当输入的自然语言句子有噪音时，KRISP比其他语义解析器的鲁棒性更强。

中英文差异主要有两点：一是中文与英文的语法结构有较大差异，适用于英文的语义解析模型并不一定适用于中文；二是英文重结构，中文重语义，英文语义解析方法没有较好地考虑中文语言特点。因此本文针对中文语法结构和中文语言特点提出一种基于词对齐的中文语义解析模型(Word Alignment-based Chinese Semantic Parsing，WACSP)，图1是其流程图，结合中文语言的特点在数据预处理算法中，对数据集进行中文分词、数据清理和数据重构等，使得中文语义解析算法性能有较大的提升。

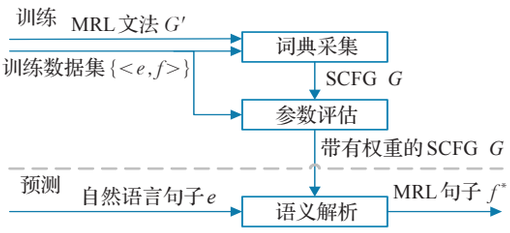

图1　语义解析模型流程图

WACSP用嵌套结构处理MRs，通过Kate基于转换规则的语义解析方法做语义解析[4]。本文提出的算法通过给定的数据是中文句子以及其正确的MRs。算法不需要中文句法先验知识，且假设上下文无关文法是明确的。本文主要创新点是用统计机器翻译技术做中文语义解析。具体来说，用统计词对齐模型[5]来获取双语词典，该词典包含自然语言字符串及MRL表示。在解析框架中通过结合这些自然语言字符串以及它们的MRL翻译来最终形成完整的MRs，这个解析框架就是同步上下文无关文法SCFG[6]，该文法是大部分现有的基于句法的统计翻译模型的基础[7-8]。

2　基于词对齐的中文语义解析模型WACSP

从图2中可以看出，WACSP的任务就是将中文句子翻译成用形式化语言CLANG表示的MR格式。为了完成这一任务，首先需要用语义语法[9]解析中文句子的句法结构，语义语法中的非终结符与CLANG语法的非终结符相同。通过语义解析器获得字符串的意义表示，然后通过结合字符串的意义表示来获取整个中文句子的意义表示。图3(a)是例句的语义解析树中的一种可能，其中的非终结符是基于CLANG文法的非终结符。图3(b)表示对应的MR结构的CLANG解析树。

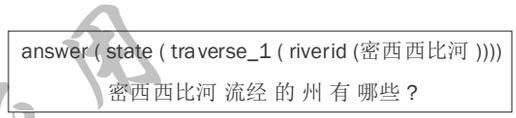

图2　中文句子对应MRL的意义表示

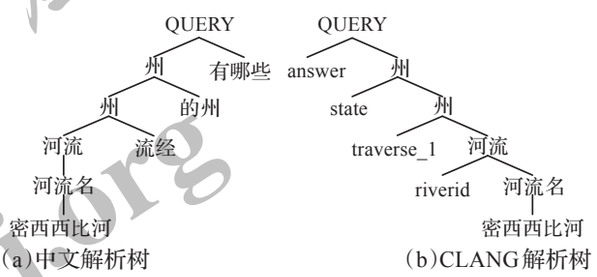

图3　图2中字符串对的部分解析树

上述过程可以看作同步解析的一个实例[6]，最终推导出两个字符串，一个是源语言的字符串，另一个是目标语言的字符串。输入是中文句子 e，然后语义解析器的任务就是找出一种推导，它可以推导出字符串对 $<e,f>$，这里的 f 就是句子 e 的MRL翻译。为了防止字符串对的集合是无限多个，本文用加权的SCFG生成字符串对，它的定义如下：

$$G=<N,T_e,T_f,\mathcal{L},S,\lambda> \tag{1}$$

式中 N 代表有限的非终结符集合，T_e 表示有限的自然语言的终结符集合。T_f 表示有限的MRL语言的终结符集合。$\mathcal{L}$ 表示词典，词典包括有限的规则集合。S 属于 N，S 是一个开始符号。λ 是参数评估的集合，其定义了推导的概率分布。$\mathcal{L}$ 中的每一个产生式都是如下形式：

$$A\rightarrow<\alpha,\beta> \tag{2}$$

其中 $A\in N,\alpha\in(N\cup T_e),\beta\in(N\cup T_f)$。非终结符 A 称为产生式左部(Left-Hand Side，LHS)，产生式右部(Right-Hand Side，RHS)是一对字符串 $<\alpha,\beta>$。对于 α 中的每一个非终结符在 β 中都有一个与之关联的完全相同的非终结符。换句话说 α 中的非终结符是 β 中非终结符的排列。

下面是一些可以用来产生图3中的解析树的SCFG规则：

QUERY $\rightarrow<$州[1]有哪些,(answer 州[1])$>$

州 $\rightarrow<$州[1]的州,(state 州[1])$>$

州 $\rightarrow<$河流[1]流经,(traverse$_1$ 河流[1])$>$

河流 $\rightarrow<$河流名[1],(riverid 河流名[1])$>$

河流名 $\rightarrow<$密西西比河,(密西西比河)$>$

每一个SCFG规则 $A\rightarrow<\alpha,\beta>$ 可看成两部分结合而成，$A\rightarrow\alpha$ 是自然语言句子的句法解析产生式、$A\rightarrow\beta$ 是MRL语法产生式。本文将字符串 α 称为自然语言(NL)字符串，字符串 β 称为MR字符串。NL和MR字符串中的非终结符用 [1],[2],… 来进行索引显示它

们之间的关联。所有的推导都是由相关联的开始符号对 $<S_{[1]},S_{[2]}>$ 每一步推导都需要重写上一步相互关联的非终极符对。本文给出生成简单的中文句子及其CLANG表示的一种推导,如下所示:

$<\text{QUERY}_{[1]},\text{QUERY}_{[1]}>\Rightarrow$

$<$州$_{[1]}$有哪些,(answer 州$_{[1]}$)$>\Rightarrow$

$<$州$_{[1]}$的州有哪些,(answer(state 州$_{[1]}$))$>\Rightarrow$

$<$河流$_{[1]}$流经的州有哪些,

(answer(state(traverse_1 河流$_{[1]}$)))$>\Rightarrow$

$<$河流名$_{[1]}$流经的州有哪些,

(answer(state(raverse_1(rverid 河流名$_{[1]}$))))$>\Rightarrow$

$<$密西西比河流经的州有哪些,

(answer(state(traverse_1(riverid 密西西比河))))$>$

可简单理解为CLANG的表示就是中文句子的一种翻译。因此对于输入句子 e,会有多种可能的推导(如:非终结符 州 有多种推导)。为了找出正确的推导,本文设计了一个对于推导 d 的概率模型,概率模型的参数为 λ,返回值为 d 正确的概率。对于中文句子的翻译结果 f^* 有如下定义:

$$f^*=f\left(\arg\max_{d\in D(G|e)}P_\lambda(d|e)\right)$$

式中 $f(d)$ 是推导 d 中的MR串,并且 $D(G|e)$ 是一个集合,集合包含了 e 所有可能的推导。简单来说,最终输出的MRL翻译是推导 d 中的MR串,而且 d 是自然语言句子 e 概率最大的推导。f^* 可以通过动态规划算法有效地计算出来。

由于在给定NL和MRL时 N,T_e,T_f,S 就会相应获取到,所以本文语义解析的学习算法只需要学习一个词典 $\mathcal{L}$ 和带有参数 λ 的概率模型即可。因为词典是所有可能推导的集合,所以要想生成概率模型,需要首先学习得到词典。因此学习任务可以分为以下两个子任务:

(1)学习一个词典 $\mathcal{L}$,词典隐式地定义了一个集合,这个集合包含所有可能的推导,$D(G)$。

(2)学习一个参数 λ 的集合,这个集合定义了 $D(G)$ 中推导的概率分布。

两个子任务都需要训练数据集 $\{<e_i,f_i>\}$,每个训练样例 $<e_i,f_i>$ 都是成对的,即自然语言句子 e_i,以及其正确的MR串 f_i。词典的生成同样需要明确的MRL上下文无关文法。因此开始训练数据集时若没有词典则无法生成正确的推导。本文将这些推导作为隐藏变量,通过EM算法对其进行极大似然估计。

3　WACSP关键技术:词典采集

在WACSP中,本文用词对齐模型来进行词典的采集。最基本的思想是在训练集上训练一个统计词对齐模型,然后找出每个训练样例的最有可能的词对齐。通过从这些词对齐中提取SCFG规则来生成词典[8]。

本文举例说明上述算法。假设训练数据集和图2中的字符串对一样,那么词对齐模型就是找到这对字符串的词对齐。图4是一个简单的词对齐例子,在这个词对齐中每个CLANG符号都被当作一个单词处理。这样会带来两个问题:第一个,并不是所有的MR字符都有特定的含义。举例来说,在CLANG 中括号((,))和花括号({,})并没有实际的语义含义。这样的符号并不会对齐任何自然语言单词,如果训练集中包含这些字符会很有可能混淆词对齐模型。第二个,MR符号表示可能会产生歧义。例如CLANG谓词pt,它根据给定的论元类型可能会有三种含义,它可能代表着坐标(e.g.(pt 0 0)),或者某物体所在的位置((pt our 4))。如果单独判断谓词pt,词对齐模型无法正确判断出其含义。

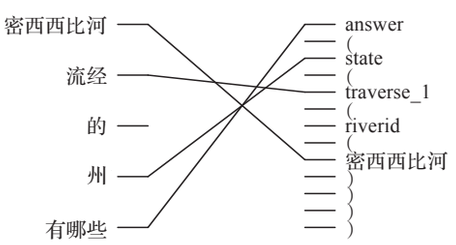

图4　一个中文与CLANG字符的词对齐

为了避免上述问题,本文用MRL产生式序列表示MR。MRL产生式序列对应MR的自顶向下最左推导。每一个MRL产生式相当于一个单词。图5中文句子与其CLANG表示的线性化解析的词对齐。

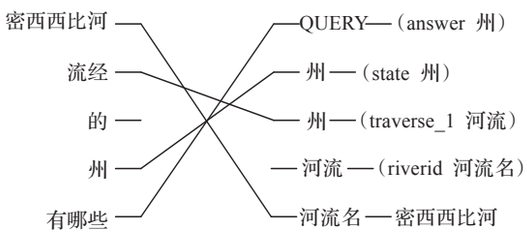

图5　一个中文与CLANG产生式的词对齐

如上例所给出的第二个产生式,州→(state 州),它就是来重写第一个产生式QUERY→(answer 州)中的非终结符“州”,同理可知其他产生式。特别需要提醒的是解析树的结构是通过线性化保存的,并且对于每一个MR都有一个唯一的线性化解析,这是因为MRL语法是明确的。在后续的提取SCFG规则中,MR解析树的结构扮演重要角色。

通过已有的词对齐模型来获取词对齐。本文使用的是GIZA++[10]实现的IBM MODEL 5[5]。

假设每个自然语言单词最多对应一个MRL产生式,SCFG规则采用自下而上方式提取。这个提取过程首先从右部(LHS)都是终结符的产生式开始,比如

河流名→密西西比河 。对于每个产生式 $X\rightarrow\beta$ ，规则 $X\rightarrow<\alpha,\beta>$ 便提取出来，其中 α 包含了所有的产生式 $X\rightarrow<\alpha,\beta>$ 所对齐的单词，例如河流名→<密西西比河，(密西西比河)> 。

然后再考虑产生式右部(LHS)包含非终结符的情况，比如带有论元的谓词。在这种情况下，NL字符串 α 包含了单词以及非终结符，其中单词与产生式相对应，非终结符表示了论元实现的位置。例如，谓词state提取的规则是：州→<州[1](1)州，(state州[1])>，公式中(1)代表着字间距为1，因为“的州”中“的”是属于没有对齐的字。字间距(g)可以看作特殊的非终结符，在数据流中最多可展开 g 个NL字，这样一来对于模式匹配会增加一定的灵活性。规则的提取过程是在线性化MR后进行的(因此谓词的提取过程是在其所有论元都提取完毕后进行)，最后便可提取出所有产生式规则。

河流→<河流名[1]，(riverid 河流名[1])>

州→<河流[1]流经，(traverse$_1$ 河流[1])>

QUERY→<州[1]有哪些，(answer 州[1])>

WACSP词典采集算法如算法1所示：首先用训练数据集 $T=\{<e_i,\ f_i>\}$ 训练词对齐模型 M ，然后从词对齐模型中获得每个训练样例最有可能的词对齐，本文取前十个最有可能的词对齐(k=10)。SCFG规则便可从每一个词对齐中提取出来。因为提取过程采用自下而上的方式，所以谓词的提取过程是在其所有论元都提取完毕后进行。字典 $\mathcal{L}$ 包含所有的规则，这些规则是从训练样例 k 个最好的词对齐中提取出来的。

算法1 词典采集算法

输入 训练 $T=\{<e_i,f_i>\}$，明确的MRL文法 G'

输出 词典 $\mathcal{L}$

LEXICON-ACQUIRE(T，G')

1：$\mathcal{L}\leftarrow\varnothing$

2：for　i←1 to |T|

3：do　$f_i^{'}$←利用G'线性化解析f_i

4：{<e_i，$f_i^{'}$>}作为训练数据集，训练一个词对齐模型M

5：for　i←1 to |T|

6：do　$a^{*}_{1,\cdots,k}$←从词对齐模型M中获得{<e_i，$f_i^{'}$>}的k个最好的词对齐

7：for　k′←1 to k

8：　dofor j←$|f_i^{'}|$ downto1

9：　　do A←lhs($f_{ij}^{'}$)

10：　α←在a_k^{*}中$f_{ij}^{'}$及其论元所对应的字

11：　β←rhs($f_{ij}^{'}$)

12：　$\mathcal{L}\leftarrow\mathcal{L}\cup\{A\rightarrow<\alpha,\beta>\}$

13：　在$a_{k'}^{*}$用A替换α

14：return $\mathcal{L}$

4　WACSP关键技术：概率估计

一旦词典获取到，下一步任务就是学习语义解析的概率评估模型。对于推导 d 用极大熵模型定义一个条件概率分布：

$$Pr_\lambda(d|e)=\frac{1}{Z_\lambda(e)}\exp\sum_i\lambda_i f_i(d)$$

式中 f_i 是特征函数，并且 $Z_\lambda(e)$ 是归一化因子。对于词典中的每一条规则 γ 都有一个特征函数，这个特征函数返回的是 γ 在推导中所用到的次数。同样对于每个单词 ω 也有一个特征函数，它返回的是字间距 ω 的数量。

模型中无法看到的单词被作为额外的特征，这一特征值是字间距中单词的总数。由于SCFG的输出文法是MRL文法，MRL产生式具有较好的结构，因此概率模型相对简单。对数线性模型使用的特征数量相对较少。本文用到与Zettlemoyer[11]相似的特征集。

用Vertibi算法解码模型，需要句子长度的立方时间。用Earley图保持所有的推导与输入的一致。

用极大似然准则评估参数 λ_i 。用高斯先验来正则化模型[12]。由于黄金准则推导在训练集中并不适用，故将正确的推导作为隐藏变量。本文用改进迭代算法与EM算法来找到最佳参数。与全监督相比，条件似然对于参数 λ 不敏感，EM算法对于 λ 是敏感的。为了假设最小可能，WACSP将 λ 初始化为0。EM算法需要统计对于句子或者句子MR对所有可能的推导。然而枚举所有可能的推导并不是好的方法，因此本文采用内向外向算法来提高统计效率[13]。根据Zettlemoyer和Collins[11]的研究思想，最终的词典只返回最好的那个规则，其他所有规则都舍弃。这样的做法也是Viterbi逼近算法，以此来提高精确度。

5　WACSP的实验结果与分析

GEOQUERY[14]是语义解析领域著名的评测数据库，目前还没有中文语义解析评测数据库，本文的主要工作之一是将此评测数据库人工翻译为中文，对于GEOQUERY的翻译遵循不改变句子语义的情况下，使用符合中文语法结构的翻译原则，由于中文分词会影响后续的解析结果，因此对数据集中的自然语言句子进行了人工分词。共包含880个样例，807条规则，13个非终结符，QUERY是开始符，含义表示如表1所示。

本文在GEOQUERY上用十折交叉验证进行实验。在测试实验中本文统计输出MRL翻译的句子的个数。当解析器没有覆盖某个句子的结构时，这个句子将会翻译失败。实验中同样需要统计生成正确MRL翻译的句子的个数。如果一个句子的MRL翻译与数据集中的MRL翻译相同，则认为这个句子的MRL翻译是正确的。本文采用精确度 *Precision*、召回率 *Recall* 以及 *F*1-*Measure* 作为评价标准。

表1　非终结符含义说明

实体名	非终结符	产生式举例
城市名	CityName	CityName→奥斯汀
国家名	CountryName	CountryName→美国
地方名	PlaceName	PlaceName→塔霍湖
河流名	RiverName	RiverName→密西西比河
州名缩写	StateAbbrev	StateAbbrev→德州
州名	StateName	StateName→德克萨斯州
数量	Num	Num→0
城市	City	City→cityid(CityName, StateAbbrev)
国家	Country	Country→countryid(CountryName)
地方	Place	Palce→placeid(PlaceName)
河流	River	River→riverid(RiverName)
州	State	State→stateid(StateName)

实验1　改变训练样例个数,测试WACSP的精确度和召回率

本实验的主要目的是在$K=0$条件下,测试训练样例个数与WACSP的精确度和召回率的关系。本次实验共分为八组,训练样例个数分别是10、20、40、80、160、320、640、792。图6(a)是WACSP的精确度与训练样例个数的关系,图6(b)是WACSP的召回率与训练样例个数的关系。

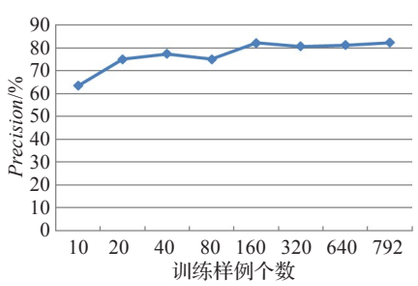

图6(a)　精确度与训练样例个数的关系

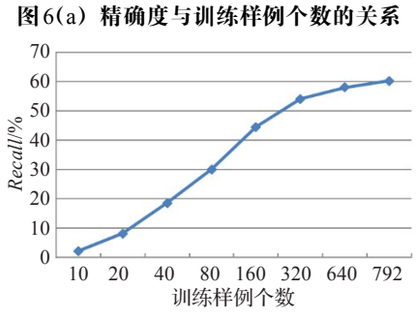

图6(b)　召回率与训练样例个数的关系图

图6(a)表明WACSP的精确度随着训练样例个数的增加而提高。图6(b)表明WACSP的召回率同样随着训练样例个数的增加而提高。实验表明在训练样例较大的情况下WACSP表现出较好的性能。实验结果分析:随着训练样例的增加,WACSP训练得到的词对齐模型更准确而且评估SCFG概率时误差更小,所以精确度和召回率会有相应的提升。

实验2　改变*K*-Best值,测试WACSP的精确度和召回率

本实验的主要目的是在训练样例个数(792句)固定条件下,测试*K*-Best值与WACSP的精确度和召回率的关系。进而找出精确度和召回率最高的情况下,*K*的最小值。本次实验共分为五组,*K*取值分别是2、4、6、8、10。图7(a)、(b)分别是WACSP的精确度、召回率与*K*的关系。

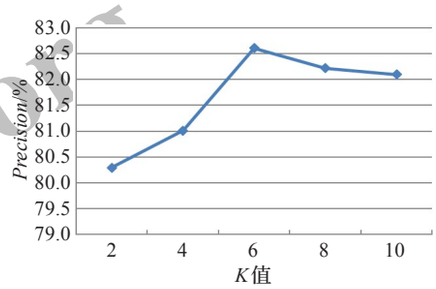

图7(a)　WACSP的精确度与*K*的关系图

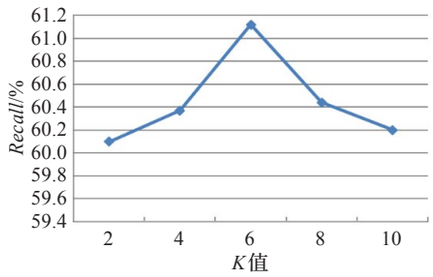

图7(b)　WACSP的召回率与*K*的关系

图7(a)表明在训练样例为792的条件下,$K=6$时WACSP的精确度最高。图7(b)表明在训练样例为792的条件下,同样$K=6$时WACSP召回率最高。实验表明,在训练样例为792的条件下,最小的$K=6$,此时WACSP的精确度和召回率最高。实验结果分析:每个训练样例的MRL产生式平均个数是6.3个,对于本文的训练样例来说,可能前六个词对齐是最佳对齐。因此$K=6$时准确度和召回率是最高的。

实验3　改变GIZA++每个模型的迭代次数,测试WACSP的精确度和召回率

本实验的主要目的是在训练样例(792句)和*K*(*K*=10)值固定条件下,测试IBM模型迭代次数与WACSP的精确度和召回率的关系。

表2表明随着每个IBM模型迭代次数的增加,WACSP的精确度和召回率都相应地增加。实验结果分

表2　GIZA++每个模型的迭代次数与WACSP精确度和召回率的关系

IBM模型迭代次数	$M1=3$, $M2=3$, $M3=1$, $M4=1$, $M5=1$	$M1=4$, $M2=4$, $M3=2$, $M4=2$, $M5=2$	$M1=5$, $M2=5$, $M3=3$, $M4=3$, $M5=3$	$M1=6$, $M2=6$, $M3=4$, $M4=4$, $M5=4$
精确度/%	80.21	81.13	82.11	82.34
召回率/%	55.54	57.21	60.30	60.33

析：GIZA++实现了IBM公司提出的5个模型[5]和隐马尔科夫模型[10]，其主要思想是利用EM算法对双语语料库进行迭代训练，由句子对齐得到词语对齐。因此随着每个模型迭代训练次数的增加，得到的词对齐模型就越准确，WACSP的精确度和召回率也随之提高。

6　总结

语义解析是生成意义表示并将这些意义表示指派给语言输入的一种处理[15]。机器翻译是将一个源语言句子转化为对应的目标语言句子。本文研究并提出了一种新颖的基于词对齐模型的语义解析模型，可将编译原理和机器翻译技术应用到语义解析领域，即可以用机器翻译技术做语义解析以解决自然语言理解。该方法用统计词对齐模型来获取双语词典，解析模型本身可以看作是基于句法的翻译模型。

本文介绍了使用同步解析技术的WACSP语义解析的学习算法，同步解析已经被广泛地应用在基于句法的统计机器翻译领域中。WACSP与其他基于短语的翻译模型相似，这些模型都需要一个简单的词对齐模型来获取短语词典。本文对WACSP进行了大量的评估实验，实验表明WACSP有较好的精确度和召回率。

参考文献：

[1] Ge R，Mooney R J.A statistical semantic parser that integrates syntax and semantics[C]//Proceedings of the Ninth Conference on Computational Natural Language Learning，2005：9-16.

[2] 李金淼.基于组合范畴文法的中文语义解析[D].南京：东南大学，2015.

[3] Kate R J，Mooney R J.Using string-kernels for learning semantic parsers[C]//Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics，2006：913-920.

[4] Kate R J，Wong Y W，Mooney R J.Learning to transform natural to formal languages[C]//Proceedings of the Twentieth National Conference on Artificial Intelligence（AAAI-2005），Pittsburgh，PA，2005：1062-1068.

[5] Brown P F，Pietra V J D，Pietra S A D，et al.The mathematics of statistical machine translation：parameter estimation[J].Computational Linguistics，1993，19(2)：263-311.

[6] Aho A V，Ullman J D.Properties of syntax directed translations[J].Journal of Computer and System Sciences，1969，3(3)：319-334.

[7] Yamada K，Knight K.A syntax-based statistical translation model[C]//Proceedings of the 39th Annual Meeting on Association for Computational Linguistics，2001：523-530.

[8] Chiang D.A hierarchical phrase-based model for statistical machine translation[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics，2005：263-270.

[9] Allen J.Natural language understanding[M].2nd ed.[S.l.]：Pearson，1995.

[10] Och F J，Ney H.A systematic comparison of various statistical alignment models[J].Computational Linguistics，2003，29(1)：19-51.

[11] Zettlemoyer L S，Collins M.Learning to map sentences to logical form：structured classification with probabilistic categorial grammars[J].Eprint Arxiv，2012：658-666.

[12] Chen S F，Rosenfeld R.A Gaussian prior for smoothing maximum entropy models[R].[S.l.]：Carnegie Mellon University，1999.

[13] Yusuke M，Jun'Ichi T.Maximum entropy estimation for feature forests[C]//HLT，2002.

[14] Seneff S.TINA：a natural language system for spoken language applications[J].Computational Linguistics，1992，18(1)：61-86.

[15] 詹志建，杨小平.基于语言网络和语义信息的文本相似度计算[J].计算机工程与应用，2014，50(5)：33-38.