统计机器翻译中的非连续短语模板抽取及其应用

孙越恒 段 楠 侯越先
(天津大学计算机科学与技术学院 天津 300072)

摘 要 目前基于短语的统计机器翻译模型很少将非连续短语的情况考虑在内,由此造成翻译结果在目标语言中的意义变化或缺失。以非连续介词短语为例,提供了一种短语模板抽取算法。首先采用基于规则的方法,抽取出中文非连续介词短语模板,而后借助双语对齐语料和介词_方位词翻译表,获得模板对应的英文翻译。最终形成的双语模板被加入短语翻译表中。在标准测试语料上的对比实验表明,加入非连续短语模板后,译文更加符合语法规范,而翻译结果也取得了相对稳定的提高。

关键词 统计机器翻译,短语模板,非连续介词短语,模板抽取
中图法分类号 TP391 **文献标识码** A

Extraction and Application of Discontinuous Phrase Templates in Statistical Machine Translation

SUN Yue-heng DUAN Nan HOU Yue-xian
(School of Computer Science and Technology, Tianjin University, Tianjin 300072, China)

Abstract The discontinuous phrases are seldom taken into account in present phrase-based statistical machine translation models, which leads to the distortion or omission of translation results. This paper took discontinuous preposition phrases for example and proposed a phrase template extraction algorithm. It first extracted phrase templates from chinese corpus based on some specified rules, and then got their english translations with a bilingual alignment corpus and a preposition and location-word translation table. The generated bilingual templates were then added into the translation table. Comparative experiments in standard test corpus indicate that when these discontinuous phrase templates are applied in the translation system, the resulted translations are well comply with grammar specifications, and the translation quality is also improved.

Keywords Statistical machine translation, Phrase template, Discontinuous preposition phrases, Template extraction

基于统计的机器翻译近年来获得了迅速发展,从最初的词翻译模型[1]逐步过渡到短语翻译模型[2-5],即翻译的基本单元从词过渡到了短语。这样,短语所蕴含的词语序列在翻译过程中被当作一个整体,而不必要求其必须满足对应语言的语法要求。同时,以短语为单元进行翻译,可以在一定的范围内保持语言内在的调序信息,而不用后续建立相应的模块进行专门处理。

但是,在对翻译结果进行仔细观察和统计之后,发现基于短语的传统翻译系统所存在的一个重要问题是,在句子切分过程中获得的短语全都是连续的序列,而中文里存在一些不连续的特定短语结构,对这种结构的每个部分单独翻译都会使原来的意义发生变化或者缺失。为此需要考虑,如果单独将那些结构固定的非连续短语抽取出来进行处理,是否可以提高翻译质量呢? Simard 等人首先将非连续短语方法用于统计机器翻译模型,翻译质量得到了一定程度的改善[6]。在他所使用的模型中,要求短语内部的间隔部分必须是严格的词,因此短语长度是固定的。文献[7]认为,这种短语形式不够灵活,且数量巨大,由此带来计算时间和空间上的开销,并提出了一种可变长度的非连续短语定义和动态的短语抽取过程。但上述工作均未考虑非连续短语的边界修正问题,尤其是对非连续的介词短语来说,这一问题尤为突出。同时,这些工作也缺乏从训练语料中选择和抽取非连续短语的标准。

本文以非连续介词短语为例,提供了一种短语模板的抽取算法,其中的短语模板实际上是这样的固定语言结构:这些结构本身是不连续的,通过根据上下文添加进去的句子成分可以形成一系列意义相似的句子,而这些结构对应的英文翻译却是连续的。如"在 # # # 的 基础上",其对应翻译为"on the basis of # # #",其中,# # # 表示模板在实际的句子中可加入的内容。在汉语中,满足上述模板要求的非连续短语可以分成很多的种类,上例的介词短语仅仅是其中一种。本文的目标就是要从原始的双语对齐语料中抽取出这样的短语模板和它们对应的英文翻译,并添加进短语翻译表中,以考查其对翻译质量的影响。

1 短语模板抽取在整个机器翻译中所处的位置

基于短语的统计机器翻译系统主要涉及如下过程:(1)双

到稿日期:2008-11-11 返修日期:2009-02-02 本文受国家自然科学基金项目(60603027)和微软亚洲研究院(MSRA)资助。
孙越恒(1974-),男,讲师,主要研究方向为自然语言处理、Web信息检索与挖掘等,E-mail:yhs@tju.edu.cn;段 楠(1981-),男,博士生,主要研究方向为机器翻译;侯越先(1972-),男,副教授,主要研究方向为机器学习、人工智能。

语语料的预处理(包括中英文分词、语料去重、数字、时间等实体替换);(2)双语对齐文件的生成(主要应用GIZA++);(3)在双语对齐语料的基础上进行短语抽取,生成短语翻译表,并计算对应短语翻译的翻译概率,词汇权重等;(4)语言模型的生成(主要采用开源工具 SRILM);(5)解码过程;(6)对翻译结果进行评测。

短语模板抽取工作主要是在第2步双语对齐文件生成之后进行,并将生成的短语模板转化为短语翻译表的形式,加入到第三步生成的短语翻译表中。最终的翻译表被用于机器翻译的解码过程。

2 非连续介词短语模板抽取

2.1 中文非连续介词短语模板抽取

首先定义了如下规则:将中文句子中以介词开头,而以方位词结束的短语作为一个模板候选项单独抽取出来。对于中文来说,介词和方位词基本上都是封闭的,也就是说,可以制作这样的一个词表,表项的内容是对中文介词和方位词的穷举。但同时也要注意到,并不是任意一个介词和任意一个方位词的搭配都是一个合法的介词短语结构,所以在抽取出所有可能包含短语模板的候选项之后,还要根据已有的固定搭配,对该候选序列进行过滤,只保留下那些搭配合法的候选序列作为结果。表1给出了介词_方位词搭配表中的部分实例。

表1 介词_方位词搭配表举例

中文	出现次数
在_上	724
在_中	690
Z_中	588
Z_上	581

其中,包含 Z_的选项是忽略介词仅仅包含方位词的选项,后面的数字表明了该搭配在抽取语料中出现的频数。

在初步获得了包含合法介词_方位词搭配的序列之后,还要对该中文序列做进一步的约束,有3个约束条件:(1)保证该序列不以"和"、"与"开头,这条规则的制定是建立在对语料的大量观察基础上的:以这些词开头的单词序列数量较多,往往又只是完整短语序列的一个并列子部分;(2)保证该序列中不包含任何短语标点符号,当然,有可能包含诸如"{"、"}"、"《"、"》"这样的符号,这种情况是被允许的,但是在这种情况下,需要判断这些符号是否总是在序列中成对的出现;(3)对于现有的长度大于等于4的中文序列,首尾单词已经为合法的搭配,则考查实意单词前面是否为"的",如果是,也将其保留。这是因为"的"字与模板中实意单词搭配出现的次数非常频繁,而其翻译往往又是模板英文翻译中的一个连接部分,将其与实意单词一起抽取出来是符合统计规律的。

对满足上述约束条件的中文序列的起始位置进行记录保存,并进一步确认中文序列包含的单词信息。除了对介词_方位词进行抽取和保留之外,还要对方位词之前的实意单词进行考察。通过统计可以得出这样的结论:方位词短语前面的单词往往可以结合很多不同的修饰成分,构成各种不同意义的介词短语结构。因此,抽取出的模板需要保留这样的成分,这样的短语模板比单纯的介词_方位词搭配会更加充实和丰富,同时又不至于产生太多的变化而导致无法进行统计抽取。

2.2 中文短语模板的对应英文翻译抽取

在初步确定了中文模板之后,需要进一步确定该中文模板在英文中的对应翻译,此时,系统第二步生成的双语词对齐矩阵就可发挥重要的作用。之前已经记录了抽取出来的中文短语模板中单词在句子中的位置信息,在对齐矩阵中,以中文模板的介词和方位词作为源语言的边界,寻找与该边界内的中文单词存在对齐的英文单词的位置,并取所有位置中的最小和最大位置作为目标语言的边界。由此得到了一个区域块,在这个区域里,包含了该短语及其对应英文翻译,表2说明了这样一种情况。

表2 短语模板对齐示例("+"表示对应行列的中英文互为翻译)

	从	去年	底	开始
since	+			
the				
end			+	
of				
last		+		
year		+		

在上面例子中,抽取出来的中文模板是"从 # # # 底 开始"。如果仅仅根据词对齐的结果来寻找这个模板的对齐,会得到这样的一个中英文对译模板:

从 # # # 底 开始《=》since # # # end

这个英文翻译的抽取质量很差,与希望的结果"since the end of"相差很远。这个结果主要是受对齐质量的影响所导致的。对齐结果对于实词能够保持比较高的准确性,但是,像the,of,in,with等一些并没有实际意义而出现次数又频繁的词来说,统计方法就显得有些无能为力了。所以为了抽取出期望的结果,需要对对齐结果做进一步修正工作。

由于直接利用对齐的方法不能有效地挖掘出介词_方位词的搭配关系,本文另采用了自定义的一些规则。由于有效的介词与方位词搭配数量有限,从 Chinese Tree Bank 中将这样的搭配及其对应翻译抽取出来生成一个特殊的"介词_方位词翻译表"。图1是从翻译表中截取的一个示例。

除-外 ||| apart from ||| 0.18623482 0.03017837 0.00693084 0.00340520 2.718
除-外 ||| in addition ||| 0.17004049 0.02302020 0.00375873 0.00173095 2.718
除-外 ||| in addition to ||| 0.12955466 0.02302020 0.01013300 0.00173095 2.718
除-外 ||| besides ||| 0.08502024 0.07586210 0.00381056 0.00346240 2.718
除-外 ||| apart from the ||| 0.05263158 0.03017837 0.00778910 0.00340520 2.718

图1 介词_方位词翻译表示例

对每个中文介词短语,逐个查看抽取出的对应英文序列在上表英文中是否有匹配。如存在匹配,则说明英文句子中存在该对应的翻译,进而根据对应的翻译,修正局部的对齐结果。本文抽取出来的短语可能如上述情况所示,并不是一个连续的结构,这不符合初始定义要求。因为需要抽取的非连续的中文短语的对应英文翻译必须是连续的。经过观察发现,在非连续单词之间的,往往是那些被称为"停用词"的英文单词或者是一些人称代词,于是可作如下限定:对于抽取出来的中文模板短语的英文翻译,如果存在非连续的情况,则考查这些导致非连续的单词,如果它们是一些停用词或是一些具有宽泛意义的人称代词,则将其也作为英文短语的一部分抽取出来,由此得到比原来结果好很多的连续英文短语。

2.3 非连续介词短语模板抽取过程

上述思想可表示为图2所示的抽取过程。

加载介词表、方位词表、介词-方位词翻译表
加载英文停用词表
依次读取每个中文句子
抽取出以正确的介词-方位词搭配为首尾的中文序列并判断其有效性
对判断为有效的序列，抽取出候选中文非连续模板并记录起始位置
根据对齐算法，生成包含该中文模板的区域块
根据介词-方位词翻译表判断该搭配翻译是否存在模板对应翻译中
若存在，根据指定的规则，抽取出该中文模板对应的
翻译，并记录其出现的次数统计信息
对所有的句子都进行如上处理后，将保留的合法中文短语模板及其英文翻译
改写成短语翻译表的形式。

图 2　非连续介词短语模板抽取算法

基于上述的模板介词搭配抽取流程，本文在大规模双语语料上进行了初步的抽取工作，其中双语对齐语料包含4715546对中英文对译的句子，抽取出的文件格式如图3和图4所示。在图3中，英文左边的数值代表抽取出该英文翻译的次数，右面是产生这个英文翻译的所有可能的中文短语模板及其出现次数，所有次数的相加和应该与英文出现总次数相等。图4的解释与之类似。

```
634 ‖ at the beginning of
        488  在 ### 开始 时
         45  在 ### 之 初
         32  在 ### 开始 之际
         14  于 ### 开始 时
```

图 3　英-中短语模板

```
2136 ‖ 在 ### 的 范围 内
        891  in the context of
        412  within the framework of
        272  within the context of
        126  in the framework of
```

图 4　中-英短语模板

2.4　短语翻译概率与词汇权重的计算

通过统计出来的出现次数，可以计算出一个中文非连续短语模板翻译成每个英文短语的概率，见式(1)以及由一个英文短语获得某个中文短语模板的概率，见式(2)。

$$Pr(engTranslation\,|\,chiTemplate)=\frac{Count(engTranslation)}{Count(chiTemplate)} \tag{1}$$

$$Pr(chiTemplate\,|\,engPhrase)=\frac{Count(chiTemplate)}{Count(engPhrase)} \tag{2}$$

假设 $f_1^J=f_1,\cdots,f_j,\cdots,f_J$ 和 $e_1^I=e_1,\cdots,e_i,\cdots,e_I$ 分别是平行语料源语言和目标语言中的句子。定义 $A\subseteq\{(j,i):j=1,\cdots,J;i=1,\cdots,I\}$。则词汇权重可按式(3)和式(4)计算。

$$lex(f\,|\,e)=\prod_{j=1}^{J}\frac{1}{|\{i|(i,j)\in a\}|}\sum_{\forall(i,j)\in a}p(f_j\,|\,e_i) \tag{3}$$

$$lex(e\,|\,f)=\prod_{i=1}^{I}\frac{1}{|\{j|(i,j)\in a\}|}\sum_{\forall(i,j)\in a}p(e_i\,|\,f_j) \tag{4}$$

注意，短语翻译概率的计算仅仅涉及到对齐短语共现的次数，与双语对齐结果无关，而词汇权重的计算则是直接建立在词对齐结果基础之上的，在计算其数值的时候需要遵循以下两条具体规则：(1)对于完全相同的中英文短语翻译，选择对齐矩阵出现次数最多情况对应的结果作为最终的词汇权重；(2)在计算过程中会遇到这样一种情况：一个短语中的单词可能在其翻译短语中并没有单词与之对应，这涉及到 w(word|NULL)的计算情况，现有的计算公式中默认此类情况的结果均为1。因为通过实验发现，有意识地赋予这种情况一个概率，得到的结果都会比直接忽略它差，具体原因还有待于进一步的实验。

最终得到的中-英短语翻译表如图5所示。从左到右分别对应了中文(源语言)短语、英文(目标语言)短语、$Pr(engTranslation|chiTemplate)$、$lex(e|f)$、$Pr(chiTemplate|engPhrase)$、$lex(f|e)$和 exp。选取那些出现过至少一次以上的中文短语做为短语翻译表中的项，并对抽取结果进行“修剪”，保留翻译概率最高的前10项。英-中短语翻译表的情况与图5类似。将所有通过这种方法获得的非连续短语模板加入到系统第三步生成的短语翻译表中，并添加进词汇翻译正确数量特征。

```
在 ### 问题 上 ‖ on the issue of ‖ 0.18202633 0.09621290 0.77941176 0.66157370 2.718
在 ### 问题 上 ‖ on the question of ‖ 0.13222667 0.08663490 0.77000000 0.52587190 2.718
在 ### 问题 上 ‖ on the issue ‖ 0.03720663 0.09621290 0.87837837 0.66157370 2.718
在 ### 问题 上 ‖ on issues ‖ 0.02747567 0.13613070 0.41739130 0.81610080 2.718
在 ### 问题 上 ‖ on the matter ‖ 0.02633085 0.01915840 0.83636363 0.22700390 2.718
在 ### 问题 上 ‖ on issues of ‖ 0.02175157 0.13613070 0.62295081 0.81610080 2.718
```

图 5　中-英短语翻译表示例

3　实验结果与分析

首先介绍一下实验的 Baseline：(1)采用2107673句的中英文双语对齐语料，其中包括 FBIS 语料、联合国语料和其他一些翻译质量较高且双语对齐较准的语料；(2)由 GIZA++生成双语对齐文件，并用 IG[8] 方法进行修正，修正后的对齐包含了词语多对多的情况；(3)在对齐修正结果的基础上进行短语抽取，并且限定中文短语最大长度为3，英文短语最大长度为5，并对每个中文短语所对应的所有英文短语翻译进行“修剪”，即对翻译个数多于10个的短语，只保留概率值最高的前10项；(4)在已有的短语翻译表中加入“词汇特征”；(5)在已有的短语翻译表中加入中文单词不产生任何翻译，即翻译为空的情况，并统计记录这些可能对空的单词及其对空的概率，结果将用于解码器；(6)语言模型由开源软件 SRILM 在大规模的英语语料上获得，最大支持“5元语法”；(7)采用统一的解码器，解码器的编写采用最大熵方法，不同特征的权重通过在特定的 Mert 程序上对不同的语料进行训练而获得；(8)采用 BLEU 作为评测标准，在2002年－2005年的 NIST 汉-英翻译测试语料上进行翻译对比实验。

本文的翻译模型主要包含下述特征：(1)中文翻译为英文的翻译概率 TmF2E；(2)中文到英文的词汇权重 TmF2ELex；(3)英文翻译为中文的翻译概率 TmE2F；(4)英文到中文的词汇权重 TmE2FLex；(5)短语惩罚特征 PenaltyPhrase；(6)单词惩罚特征 PenaltyWord；(7)调序模型 DM；(8)语言模型 LM；(9)双语词汇对应翻译正确词对个数 BiLexPairs；(10)词汇对空特征 PenaltyNull；(11)非连续短语特征 PenaltyDP；(12)短语合并特征 PenaltyMP。对不同的测试语料，应用 Mert 程序对这些特征的权重进行训练。表3为 Baseline 实验结果，分别采用2002到2005年4年的测试语料做训练数据，获得4组不同的权重，再分别在每年的数据上进行解码后，评测得到分数结果。

表 3　Baseline 测试结果

训练权重	评测标准	2002	2003	2004	2005
2002 Training Weight	BLEU	0.3792	0.3569	0.3917	0.3449
2002 Training Weight	NIST	9.5296	8.7481	9.7673	8.8316
2003 Training Weight	BLEU	0.3546	0.3547	0.3677	0.3404
2003 Training Weight	NIST	9.6099	9.0999	10.0232	9.0729
2004 Training Weight	BLEU	0.3687	0.3689	0.3960	0.3556
2004 Training Weight	NIST	9.6946	9.1377	10.1928	9.1595
2005 Training Weight	BLEU	0.3588	0.3607	0.3805	0.3581
2005 Training Weight	NIST	9.6819	9.2168	10.2112	9.2829

在 Baseline 的基础上，在2002年到2005年的训练语料上，按照本文中介绍的方法分别抽取出短语模板，并将其加入到各自的翻译表中。对于训练语料中不存在的短语模板，将其过滤，这样留下的新增项均为训练语料中存在的合法项。接下来，在其它所有实验条件均相同的条件下，重复上面的实验得到应用了改进短语翻译表后的实验结果，如表4所列。

表 4　加入短语模板后的测试结果

训练权重	评测标准	2002	2003	2004	2005
2002 Training Weight	BLEU	0.3783	0.3598	0.3936	0.3462
2002 Training Weight	NIST	9.5529	8.7955	9.8074	8.8392
2003 Training Weight	BLEU	0.3665	0.3696	0.3899	0.3554
2003 Training Weight	NIST	9.7513	9.2892	10.2501	9.2579
2004 Training Weight	BLEU	0.3666	0.3704	0.3961	0.3552
2004 Training Weight	NIST	9.6893	9.2001	10.2068	9.1694
2005 Training Weight	BLEU	0.3667	0.3655	0.3884	0.3599
2005 Training Weight	NIST	9.7969	9.3618	10.3359	9.3639

可以看到，在加入了短语模板翻译和短语合并的特征之后，BLEU 和 NIST 评测分数在绝大多数年份上都有了一定水平的提升。

为了更加明显地观测包含短语模板句子的翻译效果，在第三个实验中将之前所有年份的测试语料合并在一起，并从其中抽取出 1145 句包含合理非连续短语模板的句对，进行对比实验。在此制作了两个翻译表，一个不包含短语模板，一个包含所有测试语料中含有的短语模板，这样，各自训练参数后进行翻译，结果如表 5 所列。加上短语模板的翻译项后，翻译效果可以有 12.02%的提高，而且更适于训练出合适的特征参数。表 6 列出了部分句子的翻译结果。

表 5　加入介词短语模板前后的对比评测结果

	BLEU	NIST
不带短语模板	0.3526	9.5901
带短语模板	0.3950	9.8722

表 6　中文句子翻译示例

中文句子 1	现年 $number 岁 的 帕拉马翰斯 是 在 执政 印度 人民党 的 压力 下
加入短语模板前	the 93-year-old pressure on the ruling party
加入短语模板后	the 93-year-old is under pressure from the ruling party
解释	"在＃＃的压力下"译为"under pressure from"更加合理
中文句子 2	在 经济 全球化 趋势 的 推动 下
加入短语模板前	in the trend of economic globalization
加入短语模板后	driven by the trend of economic globalization
解释	"在＃＃＃的推动下"译为"driven by"更加合理

在加入了非连续短语模板后，很多原有的由于短语长度限制而不能正确翻译的选项都能够得到比较合理的结果，而且，在一些特定的句子中还能够帮助纠正一些语法上的翻译错误。通过对数据和翻译结果的观察，我们发现，相比较 Baseline 系统，加入模板后的增益主要来自于两个方面：(1) 模板中的通配符可以在翻译解码过程中被替换成任何单元，这样可以抓住绝大部分习惯的语言结构，使得翻译结果更加流畅。调序问题是机器翻译最根本的问题，也可以说是最难以解决的问题。通过本文的方法，可以将调序信息蕴含于模板之中，使得结果中翻译片段的相对位置更为准确；(2) 功能词对结构的影响。现有的词对齐工具(如 GIZA++)对具有实际意义的单词(如名词、动词等)对齐效果很好，但对功能词(如介词、冠词、连词等)的对齐效果并未获得令人满意的效果。虽然功能词本身并不具有过多的具体含义，但是其对整体句意和句子结构起着至关重要的作用。翻译模板通过统计加上规则的方法将语言学中常用搭配中的功能词保留下来并做翻译，一定程度上缓解了词对齐对功能词翻译不足的缺陷，使得翻译结果包含更多有效的语言单元。另外，本文所提供的方法也具有良好的可扩展性和通用性。汉语中的许多短语结构，如"动词 ＋ ＃＃＃ ＋ 名词(宾语)"、"介词 ＋ ＃＃＃ ＋ 动词"、"名词从句 ＋ 的 ＋ 名词"等都可以按照类似的情况进行处理。

上述实验所用到的训练数据是非常巨大的，这样将导致生成结果的短语和模板的数量非常庞大。为此需进行"修剪"，即：仅仅保留每条源语言短语或模板所对应的概率最高的 10 条目标语言短语或模板作为最终的结果。由此既保留了最常见的翻译对，又可除去因其他原因造成的错误翻译噪音。

结束语　在基于短语的统计机器翻译系统中，短语翻译表的质量和内容起到了至关重要的作用。原有的短语翻译表由于只考虑了连续的短语序列，中文中很多固定的非连续结构都不能得到正确的翻译。另外，连续短语很好地处理了局部的翻译情况，但对于较长距离的依存关系根本无法处理。短语模板有效地解决了连续短语无法处理的结构问题和功能词的翻译问题，这就是非连续模板方法与连续短语方法间最根本的区别。本文介绍的方法在短语翻译表中加入了非连续短语模板，这样就可以直接对中文的非连续短语结构进行处理。通过在 2002 至 2005 不同年份标准测试语料上的对比试验，得出如下结论：加入非连续短语模板后，翻译得到的译文更符合语法规范，中文特定语法结构的正确英文翻译能够在结果中得到较为准确的重现，最终的翻译结果取得了相对稳定的提高。并且，通过模板的通用性和该方法对不同类型非连续短语的普适性可以推断，在接下来对其他类型短语的扩展中，短语模板的应用也必将期望对翻译的结果进行改进。

随着统计机器翻译的发展和基于短语机器翻译的日趋成熟，人们已经把研究重点逐步转移到对翻译结果结构的研究上来。非连续短语模板作为一种初步的尝试，已经可以看到值得鼓舞的结果。未来的研究工作将放在更广义的蕴含短语结构的模板抽取上，如动宾模板、从句模板等。通过对语法信息和句子中所蕴含的结构信息的有效利用，必定可以得到更好的机器翻译结果。

参 考 文 献

[1] Brown P F, Pietra S A D, Pietra V J D, et al. The Mathematics of Statistical Machine Translation: Parameter Estimation [J]. Computational Linguistics. Computational Linguistics, 1993, 19 (2): 263-311

[2] Oeh F J, Tillmann C, Ney H. Improved Alignment Models for Statistical Machine Translation [C]// Proceedings of the Joint Conference of Empirical Methods in Natural Language Processing and Very Large Corpora. University of Maryland, College Park, 1999: 20-28

[3] Yamada K, Knight K. A Syntax - based Statistical Translation Model. Proceedings [C]// the 39th Annual Meeting of the Association of Computational Linguistics. New Brunswick, NJ, 2001: 132-139

[4] Marcu D, Wong W. A Phrase-Based, Joint Probability Model for Statistical Machine Translation [C] // Proceedings the Conference on Empirical Methods in Natural Language Processing. Philadelphia, USA, 2002: 133-139

[5] Koehn P, Och F J, Marcu D. Statistical Phrase-Based Translation [C] // Proceedings the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. Edmonton, Canada, 2003: 127-133

[6] Simard M, Cancedda N, Cavestro B, et al. Translating with non-contiguous phrases[C]// Proceedings Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP). Vancouver, British Columbia, Canada, 2005: 755-762

[7] 张大鲲，张玮，冯元勇，等. 基于非连续短语的统计翻译模型研究[J]. 中文信息学报，2007，21(1)：101-108

[8] Och F J, Ney H. A Comparison of Alignment Models for Statistical Machine Translation[C]// Proceedings The 18th Conf. on Computational Linguistics. 2000: 1086-1090

(上接第 182 页)

一致的，所以访问次数的差距是造成新用户在服务组合的可靠性较差的主要原因。同时，访问次数的差别也造成了新用户在耗费时间上长于老用户。因此，在这个服务组合中，改进 Login in 和 Sign up 两个服务的耗费时间和可靠性，不仅可以有效地提升新用户的服务组合的可靠性，同时可以有效地减少服务组合的耗费时间，这样也间接地提高了老用户使用该服务组合时的可靠性和耗费时间。因为对于老用户来说，这两个服务在可靠性和耗费时间上对整体的影响也是最大。由于新用户在可靠性较低的 Sign up 上访问次数比老用户多，老用户比新用户访问次数多的 Login in 服务可靠性相对较高，所以同样的改进措施，对于老用户来说，效果将不如新用户明显。

结束语 如何评估服务组合的整体指标，是服务组合领域研究的重要方向。基于 DTMC 的分析是目前最有效的组合指标评判手段之一。随着服务组合的日益流行、场景类别的不断增加，如何根据不同情况分析服务组合的可靠性及相关指标，成为一个亟待解决的问题。

针对服务组合的性能分析，目前已经有不少运用不同的方法从不同层面上进行评估的工作。W. Abramowicz 等[12]考察了同一服务组合在单元服务和组合服务、提供者和消费者的不同角度下可靠性评价的区别，存在的不足是缺乏严谨、完备的形式化体系。S. Yacoub 等[8]运用 CDG 图良好地刻画了软件系统的运行场景，并依据不同运行场景分析了软件系统的可靠性，得到系统的可靠性。该方法的不足是，对采用该模型的系统约束较多，难以适用于结构更加复杂、状态更多的情况，比如无法良好地处理循环结构。S. S. Gokhale 等[13]基于 DTMC 对商务网站进行了多执行场景的可靠性和性能分析，也可以进行敏感度分析。但该文仅适用于网页浏览这样较为简单的结构，并要求抽象出的模型必须是完全符合 DTMC 性质的，难以兼容服务组合这样具有大量不符合 DTMC 性质的结构的情况。

本文运用 DTMC 理论的方法解决服务组合的可靠性和性能分析问题。通过研究服务组合的一些特点，首先使用 SDG 图描述用户对该服务组合的执行行为，接着对 SDG 图做相关处理，然后从场景的角度区分不同的执行情况，再通过相关映射方法得出服务组合的状态图以便进行最终分析。综合计算图中相关指标之后，可以定量地刻画和分析服务组合的可靠性和性能指标，并指出该服务组合的性能瓶颈，以便更好地对服务组合进行理解、评估、预测、控制和改进。最后通过举例论证了本方法的可行性。

下一步的工作是研究服务组合的整体指标在各个服务的性能指标随时间变化的情况下的变化情况。同时，将开发基于设计文档和网络日志等内容，自动形成 SDG 图并收集可靠性和其他性能指标数据的系统。

参 考 文 献

[1] 岳昆，王晓玲，周傲英. Web 服务核心支撑技术：研究综述[J]. 软件学报，2004，15(3)：428-442

[2] Menasce D A, Almeida V A F. Capacity planning for web services: metrics, models and methods[M]. Prentice Hall, Upper Saddle River, NJ, 2002

[3] Menasce D A. QoS Issues in Web Services[M]. IEEE Internet Computing. IEEE Press, November-December 2002: 72-75

[4] Zeng L, Benatallah B, Ngu A H, et al. QoS-aware Middleware for Web Services Composition[J]. IEEE Transactions on Software Transactions, 2004, 30(5): 311-327

[5] Jaeger M C, Ladner H. Improving the QoS of WS Compositions Based on Redundant Services[C]// Proceedings of the International Conference on Next Generation Web Services Practices. 2005

[6] 李海泉，李刚. 系统可靠性分析与设计[M]. 北京：科学出版社，2003

[7] 蔡开元，白成刚，钟小军. 构件软件系统的可靠性评估模型简介[J]. 西安交通大学学报，2003，37(6)：551-554

[8] Yacoub S, Cukic B, Ammar H H. A Scenario-Based Reliability Analysis Approach for Component-based Software[J]. IEEE Transactions on Reliability, 2004, 53(4)

[9] Yacoub S, Ammar H. A methodology for architectural-level risk analysis[J]. IEEE Trans. Software Eng., 2002, 28: 529-547

[10] Whittaker A, Thomason M. A Markov chain model for statistical software testing[J]. IEEE Trans. Software Eng., 1994, 20(10): 812-824

[11] Whittaker A, Rekab K, Thomason M. A Markov chain model for predicting the reliability of multi-build software[J]J. Inform. Software Technol., 2000, 42(12): 889-894

[12] Abramowicz W, Kaczmarek M, Zyskowski D. Duality in web service reliability[C]// Proceedings of the Advanced International Conference on Telecommunications and International Conference on Internet and Web Applications and Services. 2006

[13] Gokhale S S, Lu Jijun. Performance and availability analysis of an e-commerce site[C]// Proceedings of the 30th Annual International Computer Software and Applications Conference. 2006

[14] Papoulis A, Pillai S U. Probability, Random Variables and Stochastic Processes[M]. McGraw-Hill, 2002

[15] Jaeger M C, Rojec-Goldmann G, Mühl G. QoS Aggregation for Service Composition Using Workflow Patterns[C]// Proceedings of the 8th International Enterprise Distributed Object Computing Conference (EDOC'04). Monterey, California: IEEE Press, 2004: 149-159