

一种基于分类的平行语料选择方法*

王星¹, 涂兆鹏², 谢军², 吕雅娟², 姚建民¹

(1.苏州大学 计算机科学与技术学院, 江苏 苏州 215006;

2.中国科学院 计算技术研究所 智能信息处理重点实验室, 北京 100190)

摘要: 大规模高质量双语平行语料库是构造高质量统计机器翻译系统的重要基础, 但语料库中错误和噪音影响着统计机器翻译系统的性能, 因此有必要对大规模语料库中语料进行筛选。区别于传统的语料选择排序模型, 本文提出一种基于分类的平行语料选择方法。通过少数句对特征构造差异较大的分类器训练句对, 在该训练句对上使用更多的句对特征对分类器进行训练, 然后对其他未分类句对进行分类。相比于基准系统, 我们的方法不仅缩减 40% 训练语料规模, 同时在 NIST 测试数据集上取得 0.87 BLEU 点提高。

关键词: 统计机器翻译; 平行语料选择;

中图分类号: TP391

文献标识码: A

Selection of Parallel Corpus Based on Classification

Wang Xing¹, Xie Jun², Tu Zhaopeng², Lv Yajuan², Yao Jianmin¹

(1.School of Computer Science & Technology, Soochow University, Suzhou, Jiangsu 215006;

2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

Abstract: Large-scale bilingual corpus is an fundamental resource to build a high-quality statistical machine translation system. However, there are usually a large number of noises in the corpus, which would affect the performance of translation system. Therefore, it is essential to filter noisy sentences. In this paper, we propose a classification based selection approach distinguish high-quality bilingual sentences from the noisy ones. We first exploit several metrics to find the best and worst sentences in the corpus. Then we classify the rest sentences with the classifier, which is trained with more features on these sentences. Experimental results show that our approach not only eliminates 40% less promising sentences, but also significantly improves translation performance by 0.87 BLEU points over using all sentences.

Keywords: statistical machine translation; bilingual corpus selection.

1. 引言

平行语料在机器翻译系统中起着重要的作用。大部分机器翻译系统, 无论是基于短语的系统^[1], 还是基于句法的系统^[2,3,4], 都是从经过词语对齐的双语语料中抽取翻译规则。Och 等人^[5]的工作表明词语对齐的平行语料的质量直接决定了翻译性能。一般来说, 平行语料规模越大, 质量越高, 则词语对齐的质量越高。

然而, 大规模高质量的平行语料并不容易获取。由于统计机器翻译所需要的平行语料通常都是百万句对级的, 人工构建显然是不可能的。随着网络的发展, 大量网络双语语料的存

* 收稿日期: 定稿日期:

基金项目: 863 重大项目课题(No.2011AA01A207), 国家自然科学基金(No. 61003152,61272259, 61272260, 90920004), 教育部博士学科点专项基金(No. 20103201110021), 江苏省自然科学基金(No. BK2011282), 江苏省高校自然科学基金重大项目(No. 11KJ520003)以及苏州市自然科学基金(No. SYG201030, SH201212)资助。

作者简介: 王星(1988—), 男, 硕士研究生, 主要研究方向为自然语言处理、机器翻译; 涂兆鹏(1988—), 男, 博士研究生, 主要研究方向为自然语言处理、机器翻译; 谢军(1978—), 男, 博士后, 主要研究方向为自然语言处理、机器翻译。

在使得构建大规模平行语料成为了可能。但是网络双语语料中存在着大量的噪声语料，质量较低，直接使用会噪声引入到翻译系统中，影响系统性能。

如何从包含噪音的双语句对中选择高质量的平行语料受到了越来越多研究者的关注。比如，陈毅东等人^[6]曾研究面向处理平行语料库的筛选排序模型。姚树杰等人^[7]在组织训练语料时考虑语料覆盖度问题。但是这些方法都经验性较强，需要人工干预，陈毅东等人^[6]指出特征权重和语料选择的分数阈值一般都是通过人工经验给出。

针对该类问题，本文提出一种自动的基于分类的平行语料选择方法。首先，我们使用少数特征对平行语料进行初步打分，选择差异较大的少量句对构建训练集（如最好的 n 个句对和最差的 n 个句对）。然后，在该训练集上我们使用更多的特征（包括词汇特征，句法特征等）训练一个分类器，从而对其他平行语料进行分类，以选择高质量的句对（正例句对）。实验表明，使用过滤后的平行语料规模仅为原始语料的 60%，在 NIST 测试数据集上取得了 0.87 BLEU 百分点的提高。

后续章节组织如下：第 2 节介绍相关工作，第 3 节介绍基于分类的训练语料选择方法，第 4 节给出实验和结果分析，最后一节给出结论和未来工作。

2. 相关工作

陈毅东等人^[6]曾研究面向处理平行语料库的筛选排序模型，这个模型利用预先设定的特征将已有的平行语料进行打分排序，之后选取分数靠前的部分组织成为训练语料。姚树杰等人^[7]在组织训练语料时考虑语料覆盖度问题。但是特征权重和得分阈值的选择需要人工经验。

吕雅娟等人^[8,9]曾提出一种基于信息检索模型的统计机器翻译训练数据选择与优化方法，选择现有训练数据资源中与待翻译文本相似的句子组成训练子集，在不增加计算资源的情况下获得与使用全部数据相当甚至更好的机器翻译结果。此方法需要提前知道测试文本的内容。

Han 等人^[10]在基于训练语料句对可以分为字面互译和意译的前提下，提出一种基于词典和词性的方法判断句对是否字面互译，调整字面互译和意译句对在词对齐训练段时权重，达到翻译性能的提升。此方法是调整权重更好的利用语料，减小意译句对的影响，此处的权值也需要人工经验给出。

Munteanu 等人^[11,12]给出了大量的平行句对特征，通过利用少量的高质量平行语料构建出平行句对正反例句对，训练分类器从大规模的非平行语料选择出平行语料。但是此方法需要用到少量高质量句对作为正例句对资源。

3. 基于分类的平行语料选择方法

语料可以划分为完全平行句对、部分平行句对和完全不平行句对（噪声句对）。我们的任务是从大规模训练平行语料中选择高质量的平行语料，希望获得的是那些完全平行句对，即高质量句对。

通过观察发现，高质量平行句对一般会表现出很多共性：比如源语句和目标语句的互译准确、句对中源语句和目标语句都是比较流畅等。基于此，我们提出使用句对特征评价平行句对质量，利用分类器自动的判别句对质量的好坏。下面分别介绍分类器的构建和特征的选择。

3.1 分类器的构建

传统的监督式学习需要标记样本数据来训练分类器，然而在现实中很少有标注好的平行语料库。在此我们需要构建足够的正反例句对供分类器学习。如何构造训练分类器的正反例句对是语料选择关键的一步。

文章^[7]给出一种简单有效的排序模型对平行语料库句对进行排序。在设定权重后，句对得分成为衡量句对平行与否一个重要的参考指标。句对得分越大，句对被判定为平行句对的可能性越大。虽然对单个句对来看，无法依据其得分判定句对质量好坏，但实验论证分数高的句对集合比分数低的句对集合质量更好。我们关心的是什么样的句对更有可能成为高质量平行句对？设想如果一个句对在各个特征都比另一个句对表现优异，是否说明该句对成为高质量平行句对有着更大的可能性？答案是肯定的，因为一个句对的综合表现是由其各个特征所决定。所以我们能够通过寻找在特征上表现好或者表现不好的句对构造分类器训练的正负例句对。

文章^[11]指出句对长度比例特征和基于双语翻译词典的翻译质量特征可以简单高效的评价句对的质量过滤掉一些非平行句对。实验证明翻译模型概率也是十分有效的区分特征。同时，部分特征对（比如语言模型得分特征）对高质量句对的区分度不强。所以我们使用上述几个特征作为构造分类器训练正负例句对的特征。在这几个特征上全部得分较优的句对选为正例句对，全部得分较差的句对选为负例句对。这样，分类器训练所需要的正负例句对被构造出来。

本实验采用 ZhangLe 的最大熵模型工具包 MaxEnt^{*}，该分类器实现了包含高斯平滑的最大熵算法，采用 LBFGS 参数估计方法。

3.2 句对质量评价特征

基于双语翻译词典的翻译质量

$$P_{dic}(sen) = \frac{\sum_{w_{sen}} Translate(w_{sen})}{Length(sen)}$$

Length(s)表示句子的长度，即句子含有词的个数。Translate(w)表示语句中所有能在另一端对应句子能找到译文的词的总数。对于 Translate(w)，如果在对应的目标句子中存在词 w 翻译项则为 1，否则为 0。

文章^[7,10,11]都指出基于双语词典的特征是一种简单有效的评价特征。直观上看，P 值大，表明句子的很多词能够翻译到对应的句子上，说明句对成为平行句对的可能性更大。所以，我们分别选取源端目标端句子基于双语翻译词典的翻译质量作为句对特征。

翻译模型概率

$$P_{TM}(e|f) = \sqrt[m]{\arg \max_a \left(\prod_{j=1}^m t(e_j | f_{a_j}) \right)}$$

*http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html

$$P_{TM}(f|e) = \sqrt[n]{\arg \max_a \left(\prod_{j=1}^n t(f_j | e_{a_j}) \right)}$$

对齐得分被证明^[12] 1是一个简单有效并具有辨别力的特征。在此我们用此公式表示源端句子 f 与目标端句子 e 相互的翻译概率。 $\arg \max_a$ 表示寻找出具有最大对齐概率作为最终对齐使用。其中, $t(e_j | f_{a_j})$ 表示句子在对齐源端在对齐 a 中与目标端单词 e 对齐的单词 f 与该单词的翻译概率。在此我们选取源端到目标端翻译概率、目标端到源端的翻译概率作为句对特征。

语言模型

$$P_{LM}(sen) = \frac{1}{\sqrt{\text{Length}(sen)}} \text{LanguageModel}(sen)$$

$\text{Length}(s)$ 表示句子的长度, 即句子含有词的个数。 $\text{LanguageModel}(sen)$ 表示句子的语言模型得分。

语言模型得分能够衡量句对的单语部分是否流畅。但是根据语言模型公式, 我们清楚一个句子长度对句子模型得分有影响。为了减缓这种影响, 利于不同句子间的得分比较, 我们对语言模型得分按句子长度进行归一化处理。

实验中语言模型参数在大规模单语语料上训练得到。实验中我们分别句对中的中英文句子计算其四元语言模型。

句子长度

$$L_{diff} = \text{Length}(s) - \text{Length}(t)$$

$$L_{ratio} = \left| \frac{\text{Length}(s)}{\text{Length}(t)} - \theta \right|$$

$\text{Length}(s)$ 表示源端句子的长度, 即源端句子含有词的个数。 $\text{Length}(t)$ 表示目标端句子的长度, 即目标端句子含有词的个数。 L_{diff} 值表示源端目标端句子长度差值, L_{ratio} 值表示源端目标端归一化后的商值。

文章^[7,11]都指出, 句子长度是一个非常重要的特征。文章^[7]指出给定的语言对里的翻译对的句长度应符合一定的比例。并且给出汉英平行句对长度比经验范围 0.5-1.2 之间。通过对我们语料库中语料的统计 (如图 1), 发现语料库中大部分汉英平行句对的长度比落入该经验范围。实验中我们取经验值 $\theta = 0.85$ 对汉英长度比进行归一化处理。

文章^[11]进一步指出句子长度的差值也是一个衡量平行句对好坏的重要标志。所以我们分别选取源端目标端句子长度、两者的差值以及两者的商均作为句对特征。

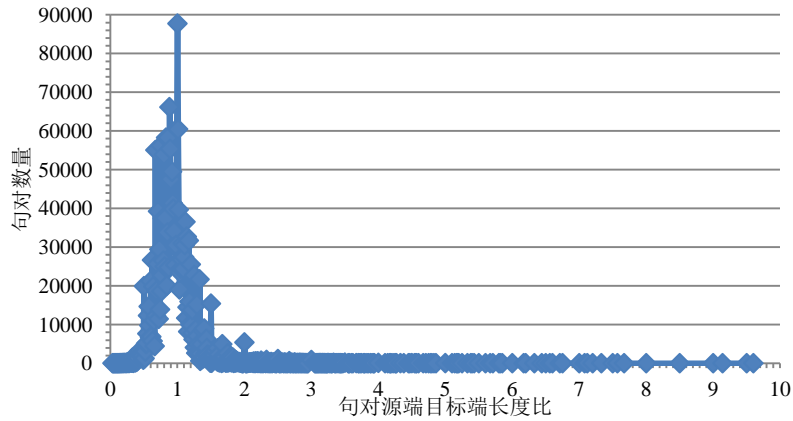


图 1 语料库中汉英句对不同长度比所含句对数量分布

未对齐词数量

$$P_{unaligned(sen)} = \frac{\sum_{w_{sen}} Unaligned(w_{sen})}{Length(sen)}$$

$Length(sen)$ 表示句子的长度，即句子含有词的个数。 $Unalign(w)$ 表示语句中所有不能在另一端对应句子能找到译文的词的总数。对于 $Unalign(w)$ ，如果存在词 w 对齐链链（对齐链连接源端词和目标端词，表明这两个词语存在互译关系）接则为 0，否则为 1。

文章^[11]指出，在经过词语对齐后平行句对间产生对齐链。一般来讲，对齐链越多，说明句对间互译的单词越多，句对成为平行句对的概率也就越大。所以未对齐词的数量也能够说明平行句对质量的好坏。在此，我们把源端目标端未对齐词数量和未对齐词所占百分比作为句对特征。

最长对齐一致性片段和最长连续未对齐片段的长度

文章^[11]指出最长对齐一致性片段的长度和最长连续未对齐片段的长度对句对是否平行有着很强的提示信息。这也和我们的直观感觉相符合。所以本文中也将这两者选择作为句对特征。

4. 实验

实验训练中英双语语料约 150 万句对，由实验室内部语料 100 万句对和实验室网络挖掘语料 50 万句对混合构成。中文平均句长 15，英文平均句长 17。对于训练语料，我们使用 GIZA++^[13]工具包进行双向对齐，然后采用“grow-diag-final-and”策略获得多到多的词语对齐。语言模型特征中的语言模型使用的单语语料中中文使用搜狗新闻语料，英文使用 Gigaword 语料中新华部分，使用 SRILM 工具^[14]训练的 4 元语言模型，使用 KN 方法进行平滑。双语词典使用 LDC 汉英双语词典，含有中英文互译词汇 54170 对。

4.1 语料筛选实验

语料筛选流程:

1. 对所有的句对计算所有（共 18 个）特征得分。
2. 按基于双语翻译词典的翻译质量得分（双向）由高到低，翻译模型得分（双向）由高到低，句子长度商值归一化得分（单向）由高到低分别对句对进行排序，总共得到五个排序结果。
3. 根据五个排序，分别取排序其前 $m\%$ 和后 $n\%$ 判别为伪正例句对与伪负例句对。如果句对在五个排序中都被判别为伪正例句对，我们取其为正例句对。类似的，如果句对在五个排序中都被判别为伪负例句对，我们取其为负例句对。其他句对作为待分类句对待用。（实验中取 $m=30$ $n=30$ ，关于 m, n 的讨论见后文）
4. 利用所有正例句对和负例句对构造的训练集训练最大熵分类器。
5. 使用训练好的最大熵分类器对待分类句对进行分类。得到分类结果。

在本实验中，我们在步骤 3 取得 473,249 句对作为训练集，其中包括 193,445 个正例，279,804 个反例。使用从该数据集上训练得到的分类器对 1,026,751 个待分类句对进行分类，我们得到了 682,145 个正例句对。最后，我们使用所有的 875,590 个正例句对作为我们选出的新训练集，以进行翻译实验。分类过程中训练句对和待分类句对中正负例句对数量如表 1 所示。

表 1 分类过程中训练句对和待分类句对中正负例句对数量

	训练句对	待分类句对	总计
正例句对数	193, 445	682, 145	875, 590
负例句对数	279, 804	344, 606	624, 410
总计	473, 249	1, 026, 751	1, 500, 000

4.2 机器翻译实验

4.2.1 实验设置

实验评测语料使用 NIST2002 年的评测语料（NIST02）作为开发集。NIST2005,20008 年的评测语料（NIST05, NIST08）作为测试集。语言模型使用语料筛选试验中的英文 4 元语言模型。实验使用开源的基于短语的统计机器翻译 moses 系统^[15]，短语抽取限制长度为 7，采用 msd-bidirectional-fe 调序模型。机器翻译实验中使用最小错误训练方法^[16]优化线性模型的参数，采用大小写不敏感的 IBM BLEU-4^[17]作为评测指标。为了更合理的评测我们的方法，文本设置如下 7 个翻译系统：

All: 使用所有训练句对训练翻译系统。

Pos: 使用所有正例句对训练翻译系统。

Neg: 使用所有负例句对训练翻译系统。

Rand1: 从所有训练句对中随机选取与正例句对数量相同的句对组成句对集合，训练翻译系统。

Rand2: 从所有训练句对中随机选取与负例句对数量相同的句对组成句对集合，训练翻译系统。

RandPos: 从正例句对集合随机选取与负例句对数量相同的正例句对组成正例句对子集, 训练翻译系统。

Pos+recallNeg: 在 Pos 系统的已有正例句对语料基础上, 对每个负例句对进行检测, 若该句对源端含有 Pos 系统训练集未覆盖到的新词, 则将该句对加入到训练集中, 否则跳过。使用最终得到的训练集合训练翻译系统。

4.2.2 实验结果及分析

表 2. 各个系统在开发集测试集的结果。*和**分别表示显著性测试中 $\rho < 0.05$ 和 $\rho < 0.01$

	训练语料句对数量	开发集	测试集			
		NIST02	NIST05	NIST08	AVE ¹	未翻译词数量
All	1, 500, 000	29.31	25.88	20.41	23.145	1, 981
Rand1	875, 590	28.86	25.28	21.11	23.195	2, 367
Rand2	624, 410	28.39	25.55	20.20	22.875	2, 612
RandPos	624, 410	28.35	25.87	20.98	23.425	2, 819
Neg	624, 410	28.01	25.33	19.97	22.650	2, 691
Pos+recallNeg	923, 355	29.74	26.39	21.70	24.045	2, 315
Pos	875, 590	29.81	26.62*	21.28**	23.950	2, 439

¹AVE 是指两个测试集的平均 BLEU 值

通过表 2 的实验结果数据我们可以观察到各个系统在开发集和测试集上的实验结果。Pos 系统对比 All 系统在两个测试集上均取得更好的 BLEU 值结果 (NIST05:+0.74 NIST08:+0.87), 测试集平均 BLEU 值取得了 0.805 百分点的涨幅。Pos 系统不仅成功在在 BLEU 值的取得涨幅, 还大大减小了训练语料的规模——训练语料句对的数量减少近 40%, 缩减短语表规模, 加快翻译速度。但是在删掉部分语料后, 语料的覆盖度降低, 测试集翻译结果中未翻译词的数量对比 All 系统增加了 23%。

对比 Pos 系统与 Rand1 系统的结果, 在训练句对数量相同的情况下, 两个测试集都取得更好的结果。这也说明我们选择的正例句对有着更高的质量。现在的疑问是, 未被选择的句对即 Pos 系统中未使用的负例句对质量如何? 我们可以观察 Rand2 系统、RandPos 系统、Neg 系统的结果, 我们可以看出在训练语料句对数量相同, 未翻译词数量相当的情况下, RandPos 系统 BLEU 值、Rand2 系统 BLEU 值、Neg 系统的 BLEU 值在两个测试集上依次递减。这也和我们的预期所相符。RandPos 系统与 Neg 系统结果的差异也表明正负例句对质量的差异较大。

Rand2 系统、Rand1 系统、All 系统的训练语料的句对数依次增加, 但两个测试集合的 BLEU 值并非依次上涨。这与语料的选择有关, 因为我们是随机选择的部分语料, 新加入的语料中可能参杂着非平行句对, 导致测试结果的 BLEU 值没有增长。

为了提高语料覆盖度, 针对未翻译词处理, 我们加入了 Pos+recallNeg 系统对训练语料进行如下处理: 检查每个负例句对, 如果负例句对的源端含有正例句对源端单词集合未出现的单词, 我们就将此负例句对加入训练集合。在轻微提高训练集规模的情况下, 极大减小了未翻译词的数量, 及进一步提高翻译质量。对比 Pos 系统, Pos+recallNeg 系统增加了 47,765

个句对。其翻译结果中的未翻译词对比 Pos 系统有所减少，但是比 All 系统仍然多出很多，这与我们预估有所不同。考虑未翻译词出现的原因，一是测试集中存在未登录词。二是测试集中存在的非未登录词但是包含该词翻译的译文没有被选为最优译文而被抛弃。对比 All 系统与 Pos+recallNeg 系统的未翻译词，发现 All 系统未翻译词并非 Pos+recallNeg 系统的未翻译的子集，证明是第二点原因导致 Pos+recallNeg 系统的未翻译词数量还是很多。在测试集 BLEU 值方面，对比 Pos 系统，虽然在 NIST08 测试集提升 0.42。但是另外一个测试集 NIST05 上却降低（NIST05: -0.23）。其原因是添加负例句对，虽然在语料覆盖度问题上有所改善，但是负例句对的融入导致训练语料的整体质量降低，翻译性能无法取得提升。这也说明在选取语料的情况下不仅要保证语料覆盖度，同时要保证语料质量。

5. 结论和未来工作

本文提出一种自动的基于分类的平行语料选择方法。利用少数特征选取差异较大的少量句对构建训练集，在该训练集上我们使用更多的特征训练一个分类器，从而对余下平行语料进行分类，以选择高质量的句对。实验表明，使用过滤后的平行语料在规模仅为原始语料的 60% 的情况下翻译 BLEU 值能有所提高。

值得注意的是，我们所构造的分类器训练正负例句对是通过取交集的方式构建得到，我们根据每个特征排序取前 $m\%$ 和后 $n\%$ 构造伪正负例句对然后取交集得到分类器训练正负例句对。这里选取的正负例句对并不是真实的。在本文中减少实验复杂度，我们取 $m=30$, $n=30$ 进行试验。但我们可以引入已有信息对语料质量进行判断，从而帮助我们更加合理地设置 m , n 的取值。换言之， m , n 取值可以依赖于我们自身对语料质量的评估。比如，对从网络的获取的语料，通过对 m 取值小于 n ，构造出数量少的训练正例句对。对人工构造的平行语料，我们可以调整 m 大于 n 从而获得较多的训练正例句对。引入更多信息，能够帮助我们更好进行语料选择。

未来工作从以下方面展开。如何构造训练分类器的训练句对是语料选择关键的一步。我们在将来会进一步探索其他句对特征，构造更具区分性的分类器训练句对。

参 考 文 献

- [1]. Koehn P, Och F J, Marcu D. Statistical phrase-based translation[C]//Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1. Association for Computational Linguistics, 2003: 48-54.
- [2]. Chiang D. A hierarchical phrase-based model for statistical machine translation[C]//Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2005: 263-270.
- [3]. Yang Liu, Qun Liu, Shouxun Lin. Tree-to-string alignment template for statistical machine translation. In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics. 2006: 609-616
- [4]. Jun Xie, Haitao Mi, Qun Liu. A novel dependency-to-string model for statistical machine translation[C]. in Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011: 216-226.
- [5]. Och F J, Ney H. The alignment template approach to statistical machine translation[J]. Computational linguistics, 2004, 30(4): 417-449.
- [6]. 陈毅东, 史晓东, 周昌乐. 平行语料库处理初探: 一种排序模型[J]. 中文信息学报增刊, 2006: 66-70.
- [7]. 姚树杰, 肖桐, 朱靖波. 基于句对质量和覆盖度的统计机器翻译训练语料选取[J]. 中文信息学报, 2011, 25(002): 72-77.

- [8]. 黄瑾, 吕雅娟, 刘群. 基于信息检索方法的统计翻译系统训练数据选择与优化[J]. 中文信息学报, 2008, 22(2): 40-46.
- [9]. Lü Y, Huang J, Liu Q. Improving statistical machine translation performance by training data selection and optimization[C]//Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). 2007: 343-350.
- [10]. Han X, Li H, Zhao T. Train the machine with what it can learn: corpus selection for SMT[C]//Proceedings of the 2nd Workshop on Building and Using Comparable Corpora: from Parallel to Non-parallel Corpora. Association for Computational Linguistics, 2009: 27-33.
- [11]. Munteanu D S, Marcu D. Improving machine translation performance by exploiting non-parallel corpora[J]. Computational Linguistics, 2005, 31(4): 477-504.
- [12]. Munteanu D S, Fraser A, Marcu D. Improved machine translation performance via parallel sentence extraction from comparable corpora[C]//HLT-NAACL 2004: Main Proceedings. 2004: 265-272.
- [13]. Franz Josef Och, Hermann Ney. Improved Statistical Alignment Models [C]. in Proceedings of the 38th ACL, 2000.
- [14]. Andreas Stolcke. SRILM-an extensible language modeling toolkit [C]. in Proceedings of the 7th International Conference on Spoken Language Processing 2002 : 901-905.
- [15]. Koehn P, Hoang H, Birch A, et al. Moses: Open source toolkit for statistical machine translation[C]//Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions. Association for Computational Linguistics, 2007: 177-180.
- [16]. Och F J. Minimum error rate training in statistical machine translation[C]//Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003: 160-167.
- [17]. Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th annual meeting on association for computational linguistics. Association for Computational Linguistics, 2002: 311-318.

作者联系方式: 王星

地址 北京市海淀区科学院南路 6 号中国科学院计算所智能信息处理实验室

邮编 100089

电话 15606132201

电子邮箱 wangxing@ict.ac.cn