

中国科学技术信息研究所CWMT'2013 技术报告 *

何彦青, 石崇德, 张均胜, 王惠临

中国科学技术信息研究所, 北京 100038

摘要: 本文介绍了中国科学技术信息研究所 (ISTIC) 参加第九届全国机器翻译研讨会机器翻译评测的情况。本单位参加了英汉科技、英汉新闻两个领域的机器翻译评测项目。本文阐述了本单位机器翻译系统的实现框架以及实施细节, 并分析了它们在评测数据上的性能表现。

关键词: 机器翻译; 自然语言处理; 系统融合

ISTIC Evaluation Technical Report for CWMT'2013

Yanqing He, Chongde Shi, Junsheng Zhang, Huilin Wang

Institute of Scientific and Technical Information of China, Beijing 100038

Abstract: This is an overview of ISTIC evaluation technical report for the 9th China workshop on machine translation. ISTIC participated in the English-to-Chinese machine translation task in scientific and technical domain and news domain. This paper describes the implement framework of our machine translation system. We also give the key techniques and analyze the experimental results over the evaluation data.

Keywords: machine translation; natural language processing; system combination

1 引言

中国科学技术信息研究所 (Institute of Scientific and Technical Information of China, ISTIC) 参加了 2013 年第九届全国机器翻译研讨会 (CWMT' 2013) 组织机器翻译评测活动。在所有的评测项目中, ISTIC 参加了英汉科技、英汉新闻两个领域的机器翻译项目。在这两个项目的评测中, ISTIC 采用了规则和统计两类机器翻译多引擎相结合进行系统融合的策略, 提交了多机器翻译融合的结果。

本文的结构安排如下: 第二节给出 ISTIC 机器翻译系统的总体框架和系统融合策略; 第三节介绍数据的使用和处理; 实验结果及相关分析在第四节; 最后在第五节给出结论和展望。

2 系统描述

ISTIC 提交的英汉科技领域和英汉新闻领域的翻译结果都是在两类机器翻译多引擎生成的多机器翻译系统的翻译输出基础上通过系统融合生成的。两类机器翻译多引擎包括基于统计的机器翻译多引擎 (SMT-ME) 和基于规则的机器翻译多引擎 (RBMT-ME), 前者包含 2 个单引擎, 后者包含 3 个单引擎, 每一个单引擎使用不同的参数得到多机器翻译系统来生成 1-Best, 汇集所有的 1-Best 组成 1-Best List, 进而进行系统融合 [1]。系统的总体框架见图 1。

2.1 多机器翻译系统

基于统计的机器翻译多引擎包括基于短语的统计机器翻译单引擎 (Phrase Based

*

基金项目: 本文受中国科学技术信息研究所学科建设“自然语言处理”课题 (XK2012-6)、中国科学技术信息研究所重点工作“多语言科技信息语义关联网络构建及其应用”课题 (ZD2012-3-3) 和中国科学技术信息研究所科研项目预研资金 (YY-201220) 支持。

作者简介: 何彦青 (1974—), 女, 博士, 主要研究方向机器翻译和自然语言处理; 石崇德 (1979—), 男, 博士, 主要研究方向为机器翻译和自然语言处理; 张均胜 (1980—), 男, 博士, 主要研究方向为多语言信息处理和语义计算; 王惠临 (1947—), 博士, 主要研究方向为多语言信息服务、机器翻译和自然语言处理。

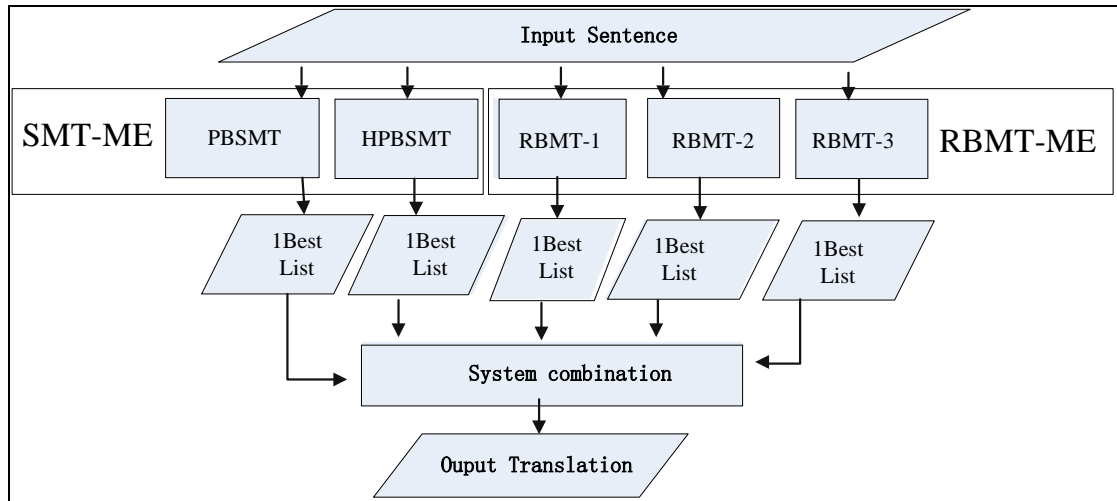


图 1. ISTIC 机器翻译系统框架

statistical machine translation, PBSMT) 和基于层次短语的统计机器翻译单引擎(Hierarchical Phrase Based statistical machine translation, HPBSMT)。PBSMT [2,3]使用连续的短语对为翻译单元, HPBSMT[4]为基于层次短语的统计机器翻译引擎, 采用了非连续的短语。这两个引擎都采用了对数线性模型来进行翻译结果的遴选, 采用的特征同 Moses。为了更好地利用统计机器翻译引擎获取尽可能多的翻译结果来为后续的系统融合服务, 我们采用了如下措施: 1) 不同的源语言和目标语言的预处理方式: 这里分别采用了东北大学的预处理方式和中信所的预处理方式。二者的不同在于分词以及特殊字符的处理。2) 不同的语言模型: 除了使用评测方提供的训练语料的目标语言端用来训练语言模型, 还加入了搜狗语料。3) 不同的开发集: 我们使用不同规模的开发集来得到不同的翻译模型的参数获取多个翻译结果。基于规则的机器翻译多引擎包含三个单引擎: 规则引擎 1、2 和 3 (Rule Based machine translation, RBMT-1、RBMT-2 和 RBMT-3)。这三个引擎都是采用规则和模板相结合的技术, 在传统的基于规则的机器翻译中融入了模板技术、统计技术, 属于基于转换的机器翻译引擎。

2.2 系统融合

基于统计的机器翻译能够利用与待翻译句子所处的领域更为接近的训练语料, 能够获取更多的专业词汇的翻译, 其劣势在于该语料的规模有限。基于规则的机器翻译对于通用词汇的翻译更为擅长, 对于专业词汇的翻译能力有限。针对两类机器翻译多引擎的特点, 我们将两类翻译结果有效地进行系统融合。

常用的多机器翻译系统融合可以从句子、短语和词三个级别上独立进行[5,6,7,8,9]。ISTIC 采用词级的系统融合技术来构建混淆网络, 将该混淆网络转换为短语表。然后使用该短语表利用短语级的系统融合技术中的重解码技术来进行解码, 生成最后的融合结果。整个系统融合框架如图 2 所示。

对于每一个基于统计的单引擎, 采用了不同策略模型来生成 1-best, 组成 1-best List。合并每一个单引擎的 1-Best List 为 1-Best Lists 来进行系统融合。我们采用了最小贝叶斯风险解码技术[10]来动态地进行选择对齐参考作为骨架翻译。在将每个翻译假设与骨架翻译进行词对齐时选用了 GIZA++ [3]工具生成骨架-假设和假设-骨架双向的词对齐。GIZA++词对齐需要将骨架翻译和其余的每一个翻译假设组成平行句对, 为了使得 GIZA++的词对齐质量尽可能少地受测试集的大小的限制, 我们将所有的翻译假设中的单个词和它自身也组合成平行句对, 两种平行句对合并在一起后使用 GIZA++ 工具包进行训练, 同样也采用 Grow-Diag-Final 式启发函数进行词对齐扩展。

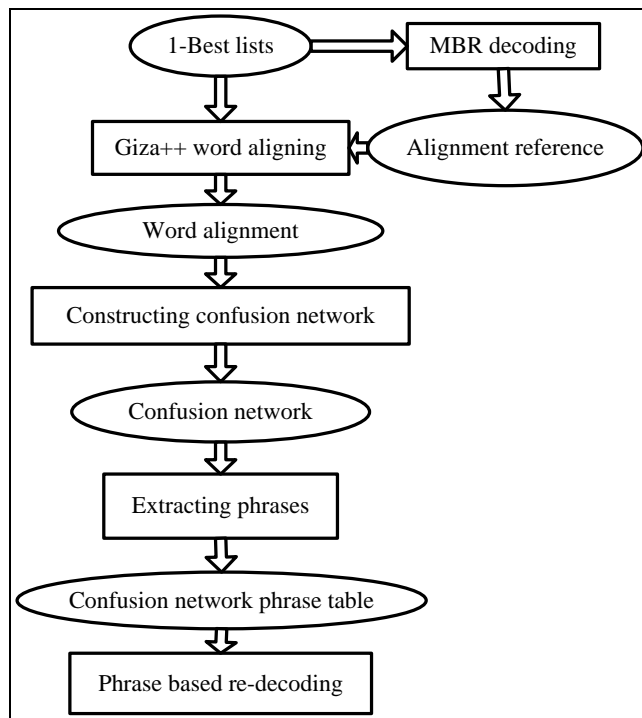


图 2. ISTIC 机器翻译系统融合框架

利用词对齐构建混淆网络时，为了使搜索路径中的节点尽量减少不可靠的词汇，没有使用空词来扩展混淆网络。直接将对齐参考的每个词作为对照点，利用其余翻译假设与它的词对齐信息，收集在每一个翻译假设中与该词对齐的词汇作为这个词的候选翻译。这样骨架翻译的每个词会有 0 个或者多个候选词汇，重复的词要记录重复次数，这样就形成一个词包。将每个词包放在混淆网络的每条弧上，之后通过投票来计算每个弧上词汇的后验概率，混淆网络就构建完成。在将混淆网络转化为混淆网络短语表时，参考翻译的每个词号看作是一个短语对的源短语，它的目标短语为该词号所对应的混淆网络上每条弧上的词，短语对的概率为该目标词的后验概率。该方法对于混淆网络上每个节点的候选词汇全部来自于多机器翻译系统的翻译结果，为了增加每个词的候选词汇，我们在英汉科技的 2013 测试集上尝试了增加科技术语词典的方法。首先使用规则翻译引擎将双语术语的英文端翻译成中文得到两端都是中文的单语术语对，然后加入到骨架翻译和假设翻译的句对中去一起生成词对齐，这样就得到了单语术语之间的词对齐。然后利用单语术语对的词对齐，抽取词包一起加入到混淆网络中去一起计算后验概率。

利用混淆网络短语表和一个基于短语的统计机器翻译系统来生成最后的目标翻译，这个过程类似于短语级的系统融合的再解码过程。但是这里的短语表不是源语言句子和翻译假设重新进行词对齐生成的，而是利用词级系统融合的混淆网络生成的。给定源语言句子 f ，融合的过程就是搜索具有最大概率的目标翻译 e ：

$$e^* = \arg \max_e \sum_{m=1}^M \lambda_m h_m(e, f)$$

其中， $h_m(e, f)$ 为特征函数， λ_m 为特征权重。基于短语的重解码过程中，仍然采用对数线性模型来完成最终的翻译融合，使用的特征函数有：

- 1: 短语后验概率；
- 2: 语言模型特征；
- 3: 基于距离的重排序特征；

4: 词惩罚;

解码的搜索策略为柱状搜索算法,最后生成的 1-Best 作为融合结果输出。每个特征函数的权重由最小错误训练算法训练得出。

3 数据的使用和处理

训练语料采用了英汉科技领域和英汉新闻领域项目发布的所有训练语料,语言模型训练数据采用了训练语料的中文部分和搜狗全网新闻语料库(SogouCA)¹,所有的基于统计的机器翻译的单引擎的参数在该项目发布的开发集上训练。

对中文数据进行的处理有:中文的分词和全角变半角;对英文数据进行的处理为:大小写转换和标点符号的分离处理。采用的Stanford的中文分词²工具和Moses³的英文Tokenization工具。词对齐工具采用了GIZA++⁴(全部使用默认的参数)并对该对齐结果进行扩展对齐(grow-diag-final)[2]。语言模型工具采用了Srlm⁵[11]工具包来获取5元语法概率信息。

4 实验

我们分别在英汉科技领域和英汉新闻领域的机器翻译项目发布的开发集和测试集上来验证翻译效果的优劣。在开发集上的打分使用了评测组织方发布的打分工具。在测试集上的打分使用了评测组织方开放的在线评测平台。

我们使用100个词来限制训练语料的最大长度,表1列出了使用的所有语料的详细统计量。

表 1. 实验语料的统计量

| 数据集 | | 语言 | 句子个数 | 词汇表 | 平均句长 |
|------------------|-----|------|---------|--------|------|
| 英 汉 新 闻 | 训练集 | 中文 | 4681740 | 466122 | 17.6 |
| | | 英文 | 4681740 | 431971 | 18.9 |
| | 开发集 | 中文 | 12484 | 18351 | 25.2 |
| | | 英文 | 3121 | 10591 | 25.7 |
| | 测试集 | 2009 | 1002 | 4951 | 25.6 |
| | | 2011 | 1001 | 5237 | 26.8 |
| 英 汉 科 技 | 训练集 | 中文 | 896410 | 189255 | 26.0 |
| | | 英文 | 896410 | 181007 | 26.5 |
| | 开发集 | 中文 | 4464 | 4658 | 20.9 |
| | | 英文 | 1116 | 3427 | 21.8 |
| | 测试集 | 2011 | 1497 | 4996 | 41.5 |
| | | 2013 | 1000 | 3953 | 22.9 |

4.1 英汉科技的翻译结果

在英汉科技领域任务上,我们采用了7个翻译系统。表2列出了在此任务上参与系统融合的单個翻译结果以及融合结果在开发集上的打分。PBSMT-1对训练语料使用了东北大学提供的预处理方式,只使用了训练语料的中文部分来训练5元的语言模型,采用了基于短語的统计机器翻译模型;PBSMT-2对训练语料使用了在第3节的预处理方式,其余同PBSMT-1。

¹ <http://www.sogou.com/labs/dl/ca.html>

² <http://www-nlp.stanford.edu/downloads/segmenter.shtml>

³ <http://www.statmt.org/moses/>

⁴ <http://giza-pp.googlecode.com/>

⁵ <http://www.speech.sri.com/projects/srlm/download.html>

HPBSMT-1 与 PBSMT-1 的训练语料和语言模型相同, 采用了基于层次短语的统计机器翻译模型; HPBSMT-2 与 PBSMT-2 的训练语料和语言模型相同, 采用了基于层次短语的统计机器翻译模型; 我们还采用了 3 个规则翻译引擎。我们采用了基于词和短语的系统融合(WPSC)方法。表 3 和表 4 分别列出了在英汉科技领域任务上参与系统融合的单系统翻译结果以及融合结果在 2011 年以及 2013 年测试集上的打分。从表 2-表 4 可以看出, 基于统计的翻译多引擎的翻译表现要优于基于规则的翻译多引擎, 基于层次短语的翻译单引擎的表现要优于基于短语的翻译单引擎。在预处理方式的比较上, 两种预处理方法在开发集上没有明显的翻译效果的差距, 但是在两个测试集上, 东北大学的预处理方式要优于第 3 节的预处理方式, 其根本原因在于对特殊字符的处理要更为细致和广泛。无论是在开发集还是在测试集 2011 上, 系统融合 WPSC 在 BLEU-SBP 上都得到了比单系统更优越的表现。在测试集 2013 上, 添加了科技术语词典的系统融合方法得到的翻译效果比较显著, 超过了不添加词典的融合表现。但是遗憾的是两个融合的结果都没有超过最好的翻译系统--规则系统 RULE-1。RULE-1 在测试集 2013 上表现一反在开发集和测试集 2011 的表现, 竟然超越了统计系统, 这进一步说明了翻译系统的表现和测试集的设置有很大的关系。

表 2: 英汉科技任务在开发集上的比较

| 引擎 | NIST | BLEU | BLEU_SBP | GTM | mWER | mPER | ICT |
|----------|---------|--------|----------|--------|--------|--------|--------|
| RULE-1 | 9.9934 | 0.4504 | 0.4469 | 0.8197 | 0.5331 | 0.3437 | 0.4145 |
| RULE-2 | 8.3867 | 0.3351 | 0.3341 | 0.7639 | 0.6116 | 0.4109 | 0.3024 |
| RULE-3 | 8.7435 | 0.3588 | 0.3553 | 0.7523 | 0.6001 | 0.4030 | 0.3376 |
| PBSMT-1 | 10.5259 | 0.5188 | 0.5015 | 0.8209 | 0.5532 | 0.3202 | 0.4897 |
| PBSMT-2 | 10.3691 | 0.5028 | 0.4895 | 0.8149 | 0.5666 | 0.3330 | 0.4653 |
| HPBSMT-1 | 10.6527 | 0.5183 | 0.5057 | 0.8308 | 0.5500 | 0.3134 | 0.4931 |
| HPBSMT-2 | 10.6610 | 0.5143 | 0.5040 | 0.8358 | 0.5454 | 0.3162 | 0.4847 |
| WPSC | 10.5457 | 0.5487 | 0.5188 | 0.8191 | 0.5037 | 0.3320 | 0.5000 |

表 3: 英汉科技任务在 2011 测试集上的比较

| 引擎 | BLEU5_SBP | BLEU-5 | BLEU-6 | NIST6 | NIST7 | GTM | mWER | mPER | ICT |
|--------|-----------|--------|--------|--------|--------|--------|--------|--------|--------|
| RULE-1 | 0.3037 | 0.3072 | 0.2476 | 8.9906 | 9.0039 | 0.8096 | 0.6096 | 0.3600 | 0.2758 |
| RULE-2 | 0.2413 | 0.2482 | 0.1936 | 8.1213 | 8.1302 | 0.7596 | 0.6733 | 0.4143 | 0.2367 |
| RULE-3 | 0.2251 | 0.2269 | 0.1742 | 7.7782 | 7.7849 | 0.7798 | 0.6644 | 0.4125 | 0.2050 |

| | | | | | | | | | |
|----------|--------|--------|--------|---------|---------|--------|--------|--------|--------|
| PBSMT-1 | 0.3905 | 0.4125 | 0.3450 | 10.4190 | 10.4424 | 0.8259 | 0.6015 | 0.3085 | 0.4051 |
| PBSMT-2 | 0.3657 | 0.3783 | 0.3126 | 10.0925 | 10.1130 | 0.8181 | 0.6284 | 0.3271 | 0.3626 |
| HPBSMT-1 | 0.3928 | 0.4105 | 0.3435 | 10.5485 | 10.5727 | 0.8334 | 0.6048 | 0.3025 | 0.4049 |
| HPBSMT-2 | 0.3814 | 0.3924 | 0.3259 | 10.3932 | 10.4160 | 0.8409 | 0.6079 | 0.3050 | 0.3822 |
| WPSC | 0.4032 | 0.4298 | 0.3626 | 10.4206 | 10.4473 | 0.8349 | 0.5787 | 0.3188 | 0.4056 |

表 4: 英汉科技任务在 2013 测试集上的比较

| 引擎 | BLEU5_SBP | BLEU-5 | BLEU-6 | NIST6 | NIST7 | GTM | mWER | mPER | ICT |
|-----------|-----------|--------|--------|---------|---------|--------|--------|--------|--------|
| RULE-1 | 0.4650 | 0.4728 | 0.4167 | 10.6492 | 10.6756 | 0.8439 | 0.4526 | 0.2650 | 0.4377 |
| RULE-2 | 0.3501 | 0.3605 | 0.2976 | 9.4843 | 9.4987 | 0.7926 | 0.5963 | 0.3139 | 0.4338 |
| RULE-3 | 0.3196 | 0.3365 | 0.2736 | 9.0910 | 9.1027 | 0.7707 | 0.6367 | 0.3395 | 0.3790 |
| PBSMT-1 | 0.3417 | 0.3501 | 0.2877 | 9.2007 | 9.2140 | 0.7826 | 0.5942 | 0.3276 | 0.4314 |
| PBSMT-2 | 0.3196 | 0.3365 | 0.2736 | 9.0910 | 9.1027 | 0.7707 | 0.6367 | 0.3395 | 0.3790 |
| HPBSMT-1 | 0.3501 | 0.3605 | 0.2976 | 9.4843 | 9.4987 | 0.7926 | 0.5963 | 0.3139 | 0.4338 |
| HPBSMT-2 | 0.3428 | 0.3581 | 0.2932 | 9.5541 | 9.5677 | 0.8029 | 0.6021 | 0.3066 | 0.4165 |
| WPSC | 0.3894 | 0.4000 | 0.3424 | 9.4101 | 9.4299 | 0.7958 | 0.5288 | 0.3381 | 0.4447 |
| WPSC-DICT | 0.4244 | 0.4354 | 0.3766 | 9.9488 | 9.9729 | 0.8187 | 0.5019 | 0.3099 | 0.4825 |

4.2 英汉新闻的翻译结果

在英汉新闻领域任务上,我们采用了 7 个翻译系统。表 5 列出了在此任务上参与系统融合的单个翻译结果以及融合结果在开发集上的打分。PBSMT-1 和 PBSMT-2 都采用了基于短语的统计机器翻译模型,前者采用了 3121 句的开发集,后者采用了 1000 句的开发集;HPBSMT-1、HPBSMT-2 和 HPBSMT-3 都采用了基于层次短语的统计机器翻译模型,这三个系统的开发集规模分别为 3121 句、2121 句和 1000 句。这 5 个基于统计的翻译系统都不但使用了训练语料的中文部分来训练 5 元的语言模型,还采用了搜狗语料。我们还采用了 2 个规则翻译引擎。系统融合仍然采用了基于词和短语的系统融合(WPSC)方法。表 5 和表 6 分别列出了在英汉新闻领域任务上参与系统融合的单个翻译结果以及融合结果在 2009 年以及 2011 年测试集上的打分。从表 5-表 7 可以看出基于统计的翻译多引擎的翻译表现大体上要优于基于规则的翻译多引擎。在表 5 中基于层次短语的翻译单引擎的表现要优于基于短语的翻译单引擎,这和英汉科技任务上的结论是一致的,但是在表 6 中, PBSMT-2 的翻译

表现明显不足，分析原因在于过训练，在开发集上训练再在开发集上打分，得分肯定会高。表 7 中，HPBSMT-1、HPBSMT-2 和 HPBSMT-3 的表现反而不如 PBSMT，经过反复查找原因，这是因为系统运行中参数出现了错误，所以导致后续的系统融合也受到了影响。虽然在开发集上，系统融合 WPSC 在 BLEU-SBP 上比单系统并没有更为优越的表现，但是在测试集 2009 和测试集 2011 上，系统融合方法得到的翻译效果仍然高于单一系统。

表 5：英汉新闻任务在开发集上的比较

| 引擎 | NIST | BLEU | BLEU_SBP | GTM | mWER | mPER | ICT |
|----------|--------|--------|----------|--------|--------|--------|--------|
| RULE-1 | 8.8633 | 0.3493 | 0.3414 | 0.7343 | 0.6665 | 0.4338 | 0.3136 |
| RULE-2 | 8.1622 | 0.2999 | 0.2902 | 0.6909 | 0.7222 | 0.4756 | 0.2869 |
| PBSMT-1 | 8.8541 | 0.3830 | 0.3639 | 0.7315 | 0.6759 | 0.4271 | 0.3688 |
| PBSMT-2 | 8.9440 | 0.3848 | 0.3630 | 0.7342 | 0.6829 | 0.4252 | 0.3633 |
| HPBSMT-1 | 8.8773 | 0.3858 | 0.3672 | 0.7340 | 0.6697 | 0.4252 | 0.3700 |
| HPBSMT-2 | 8.7685 | 0.3802 | 0.3632 | 0.7285 | 0.6722 | 0.4257 | 0.3687 |
| HPBSMT-3 | 9.0922 | 0.3929 | 0.3691 | 0.7404 | 0.6798 | 0.4236 | 0.3570 |
| WPSC | 8.5125 | 0.3823 | 0.3652 | 0.7245 | 0.6603 | 0.4414 | 0.3692 |

表 6：英汉新闻任务在 2009 测试集上的比较

| 引擎 | BLEU5_SBP | BLEU-5 | BLEU-6 | NIST6 | NIST7 | GTM | mWER | mPER | ICT |
|----------|-----------|--------|--------|--------|--------|--------|--------|--------|--------|
| RULE-1 | 0.3158 | 0.3226 | 0.2608 | 9.5917 | 9.6018 | 0.7830 | 0.6359 | 0.3924 | 0.3569 |
| RULE-2 | 0.2842 | 0.2940 | 0.2333 | 9.2336 | 9.2422 | 0.7619 | 0.6647 | 0.4069 | 0.3476 |
| PBSMT-1 | 0.3202 | 0.3330 | 0.2686 | 9.3944 | 9.4040 | 0.7675 | 0.6505 | 0.3875 | 0.4008 |
| PBSMT-2 | 0.2289 | 0.2337 | 0.1780 | 8.3663 | 8.3708 | 0.7442 | 0.7001 | 0.4459 | 0.2859 |
| HPBSMT-1 | 0.3107 | 0.3229 | 0.2591 | 9.1935 | 9.2030 | 0.7635 | 0.6531 | 0.3953 | 0.3988 |
| HPBSMT-2 | 0.3056 | 0.3151 | 0.2522 | 8.9368 | 8.9458 | 0.7609 | 0.6473 | 0.4009 | 0.4022 |
| HPBSMT-3 | 0.3190 | 0.3378 | 0.2721 | 9.5855 | 9.5955 | 0.7760 | 0.6587 | 0.3841 | 0.3841 |
| WPSC | 0.3259 | 0.3351 | 0.2746 | 8.7505 | 8.7614 | 0.7633 | 0.6269 | 0.4139 | 0.4223 |

表 7: 英汉新闻任务在 2011 测试集上的比较

| 引擎 | BLEU5_SBP | BLEU-5 | BLEU-6 | NIST6 | NIST7 | GTM | mWER | mPER | ICT |
|----------|-----------|--------|--------|--------|--------|--------|--------|--------|--------|
| RULE-1 | 0.2769 | 0.2865 | 0.2286 | 8.9580 | 8.9668 | 0.7400 | 0.6606 | 0.4161 | 0.3093 |
| RULE-2 | 0.2628 | 0.2763 | 0.2187 | 8.8123 | 8.8196 | 0.7320 | 0.6655 | 0.4180 | 0.3175 |
| PBSMT-1 | 0.3044 | 0.3143 | 0.2542 | 9.0448 | 9.0536 | 0.7508 | 0.6430 | 0.3947 | 0.3737 |
| PBSMT-2 | 0.3036 | 0.3150 | 0.2544 | 9.1392 | 9.1479 | 0.7527 | 0.6449 | 0.3914 | 0.3668 |
| HPBSMT-1 | 0.2977 | 0.3068 | 0.2474 | 8.8049 | 8.8137 | 0.7434 | 0.6479 | 0.4025 | 0.3714 |
| HPBSMT-2 | 0.2899 | 0.2977 | 0.2398 | 8.4946 | 8.5029 | 0.7387 | 0.6445 | 0.4116 | 0.3723 |
| HPBSMT-3 | 0.3031 | 0.3171 | 0.2562 | 9.1861 | 9.1957 | 0.7527 | 0.6525 | 0.3932 | 0.3581 |
| WPSC | 0.3116 | 0.3187 | 0.2625 | 8.3316 | 8.3420 | 0.7456 | 0.6218 | 0.4202 | 0.3925 |

5 结论

ISTIC 参加了 CWMT'2013 英汉科技领域和英汉新闻领域两个任务的机器翻译项目, 在英汉科技领域上取得了较好的成绩, 但是在英汉新闻领域上与最好的翻译系统仍存在一定的差距。经过上述实验分析以及整个参评过程, 我们积累了下述经验:

1) 数据的预处理问题。从实验中可以得知不同的预处理会影响到统计机器翻译系统的翻译效果。我们对数据的预处理仍然较为粗糙, 如果更为细致深入地研究语料的特点, 根据不同的语料设定个性化的预处理方案, 特殊字符制定统一的标准, 得到更符合统计机器翻译处理方式的标准语料值得进一步研究的工作。

2) 机器翻译研究如果想获得大规模应用性发展, 系统融合研究是必不可少的。但是我们的系统融合在英汉科技领域的翻译效果并没有在英汉新闻领域得到同样的体现, 这进一步说明了任何翻译方法的制定都要认真研究该领域语言的特征以及翻译的规律, 因此, 在新闻领域, 我们致力于采取进一步改善词对齐和增加新闻领域的通用词典等方式来增强系统融合的翻译效果。

参考文献

- [1]何彦青、张均胜、王惠临, 基于词和短语的多机器翻译系统融合方法研究[J], 情报学报, 2011, 30(12), 1168-1173。
- [2] Philipp Koehn, Franz Josef Och, and Daniel Marcu. Statistical Phrase-Based Translation[C]. In Proceedings of the Human Language Technology conference / North American Chapter of the Association for Computational Linguistics (HLT/NAACL-2003), 2003, pages 127-133.
- [3]Franz Josef Och, Hermann Ney. A Systematic Comparison of Various Statistical Alignment Models[J], Computational Linguistics, 2003, volume 29, number 1, pp. 19-51.
- [4]Chiang, 2005]David Chiang. 2005. A Hierarchical Phrase-based model for Statistical Machine Translation. In Proceedings of ACL 2005, pages 263-270.

- [5]Srinivas Bangalore, German Bordel, and Giuseppe Riccardi. 2001. Computing consensus translation from multiple machine translation systems. In Proc. ASRU, pages 351–354.
- [6] J.G. Fiscus. A Post-Processing System to Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction (ROVER). IEEE Workshop on Automatic Speech Recognition and Understanding, 1997.
- [7]Xiaodong He, Mei Yang, Jianfeng Gao, Patrick Nguyen, and Robert Moore, Indirect-HMM-based Hypothesis Alignment for Combining Outputs from Machine Translation Systems. In Proceedings of EMNLP 2008.
- [8] Antti-Veikko I.Rosti, Necip Fazil Ayan, Bing Xiang, Spyros Matsoukas, and Richard Schwartz, Bonnie J.Dorr. 2007a. Combining Outputs from Multiple Machine Translation Systems. In Proceedings of NAACL HLT, pages 228-235, Rochester, NY, April 2007.
- [9] Antti-Veikko I.Rosti, Bing Zhang, Spyros Matsoukas, and Richard Schwartz. 2007b. Improved Word-level System Combination for Machine Translation. In Proceedings of ACL 2007.
- [10]K.C. Sim, W. Byrne, M. Gales, H. Sahbi and P. Woodland. Consensus Network Decoding For Statistical Machine Translation System [A]. In: ICASSP, 2007.
- [11]Andreas Stolcke, 2002. SRILM-An extensible language modeling toolkit. In Proceedings of International Conference on spoken language processing, volumn 2, pages 901-904.