

doi:10.3772/j.issn.2095-915x.2016.05.007

# 2015 全国机器翻译研讨会 ISTIC 评测报告

何彦青, 孟令恩, 丁亮, 石崇德, 李颖, 王莉军, 孙瑶

(中国科学技术信息研究所, 北京 100038)

**摘要:** 中国科学技术信息研究所 (ISTIC) 机器翻译研发工程历史悠久, 几经变迁。本文在回顾 ISTIC 机器翻译发展历史的基础上, 介绍了最近参加第十一届全国机器翻译研讨会机器翻译评测 (简称 CWMT' 2015) 系统的具体表现。CWMT' 2015 评比中, ISTIC 参加了维汉、藏汉、蒙汉三个机器翻译评测项目。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:** ISTIC, 机器翻译, 系统融合, 假设对齐

**中图分类号:** G355

## Evaluation Technical Report of ISTIC from CWMT'2015

HE YanQing, MENG LingEn, DING Liang, SHI ChongDe, LI Ying, WANG LiJun, SUN Yao  
(Institute of Scientific and Technical Information of China, Beijing 100038, China)

**基金项目:** 本文受国家自然科学基金项目“面向科技文献的统计机器翻译语境分析”(61303152), 中国科学技术信息研究所重点工作项目“日汉机器翻译双语资源建设与翻译引擎研发”(ZD2016-05)资助。

**作者简介:** 何彦青 (1974-), 博士, 副研究员, 研究方向: 机器翻译、语义角色标注、自然语言处理, heyq@istic.ac.cn; 孟令恩 (1989-), 硕士, 研究方向: 语义角色标注、自然语言处理; 丁亮 (1994-), 硕士研究生, 研究方向: 机器翻译、自然语言处理。石崇德 (1979-), 博士, 副研究员, 研究方向: 机器翻译、自然语言处理。

**Abstract:** This article presented the history of machine translation R&D in Institute of Scientific and Technical Information of China (ISTIC), and showed the evaluation technical report from ISTIC for the 11th China Workshop on Machine Translation (CWMT'2015). During the CWMT'2015, ISTIC participated in the Uyghur-Chinese, Mongolian-Chinese and Tibetan-Chinese machine translation tasks. This article describes the implement framework of machine translation system of ISTIC. We also described the key techniques and analyzed the experimental results over the evaluation data.

**Keywords:** ISTIC, machine translation, system combination, hypothesis alignment

## 引言

机器翻译研发工程,同计算机情报检索、汉字信息处理系统、复印技术、全息缩微技术、科技声像、信息资源与数据库建设等重大工程一样,是中国科学技术信息研究所(以下简称中信所)特色与引领性服务项目。中信所自1956年10月15日成立近60年以来,机器翻译研发工作跌宕起伏,曾经历过严厉的质疑。然而,一代又一代的中信所研发人员承受住了风雨的考验,取得了不可替代的成果,为中国的机器翻译事业,特别是科技领域机器翻译的发展做出了重要贡献。

进入“十三五”,在国家创新驱动、“一带一路”全球经济战略方针之下,机器翻译研发工程受到了众人的瞩目,着实迎来了大规模实用化、面向技术转移的春天。

借此春风,本文以中信所机器翻译研发工程暨CWMT'2015技术报告为着眼点,全面梳理中信所机器翻译研发工作的历程,同时,挑选一个案例重点介绍最新的研究成果,暨CWMT'2015评测报告。以期有面有点的再现中信所的相关工作。借用本刊架构的平台,宣传中信所的公益性科技情报服务。期望在国际大工程、大科学的环境下,更多的人能充分了解与使用国家资金支持下的研究成果。强化同行业者之间的交流,更重要的是,实现支撑各行各业的科技创

新活动的目标。

本文内容构成如下:

前部分,第1节。介绍ISTIC机器翻译研发历程;

后部分,第2~5节。介绍CWMT'2015机器翻译ISTIC评测结果。ISTIC参加了维汉、藏汉和蒙汉三个机器翻译项目的测评,采用了统计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。在此,给出ISTIC机器翻译系统的总体框架和系统融合策略、介绍采用的数据和处理过程、实验结果及相关分析,以及所获启示;

最后是本文的展望。

## 1 ISTIC 机器翻译研发历程

中信所在成立的三年后,就尝试了机器翻译的实验,直到今日,这种探索从未间断。以下按照时间顺序,列出值得一提的重要事件。

1959年,进行了我国最早的机器翻译实验。机器翻译研究团队设计出一套俄汉机器翻译实验型规则系统,经过编程后,在国产第一台104型电子计算机上成功进行了翻译试验。

1974年,“机器翻译”、“情报检索”一并纳入国家“748”工程,开始情报检索系统研制工作。同年,中情所(当时的机构名)、北京图书馆、

国防科委情报所3家签订协议书，设立了情报检索组，主要负责领导情报检索系统的研制工作，由中情所牵头，“748”工程办公室挂靠在中情所。

1975年，计算机室正式成立，下设软件组、硬件组、词表组和机器翻译组，承担“748”工程的部分组织协调任务。

1975年，由中情所、社科院语言所、中科院计算所、上海情报所、冶金部情报所及化工研究院等单位组织建成了机器翻译协作研究团队。

1978年，首次进行我国“英汉题录机器翻译”的试验，并研试法汉、德汉、俄汉、日汉等机器翻译，由一对一的机译，发展为多种语言的互译。

1992年，与中国科学院计算所等单位协作完成“863”计划中的《英汉机器翻译系统IMT/EC—863》的攻关任务。该系统1993年获得中国科学院科学进步一等奖，1995年获得国家科技进步一等奖。

2007—2009年，参与中科院宗成庆研究员牵头的国家支撑计划项目“多语言信息服务环境关键技术研究及应用”。

2011年，中信所开发的“ISTIC英汉科技自动翻译系统”在第七届全国机器翻译研讨会（CWMT 2011）系统评测中获得科技领域自动评测第一名。

2014年，完成了“中英日多语言科技文本分析基础软件工具与平台建设”重点工作。

2015年，完成了“日汉机器翻译双语资源建设与翻译引擎研发”重点工作。

2014—2016年4月，与日本科学技术振兴机构JST合作，分别作为两国的牵头机构，组织中日两国最著名的科研机构，实施了国家科技合作专项，“面向科技文献的日汉双向实用型机器翻译合作研究”。后续合作处于审批之中。

另外，正在开展的国家自然科学基金项目“面向科技文献的统计机器翻译语境分析”，进展很

顺利。

如上所述，六十年的科研积累，全面夯实了中信所机器翻译领域的研发基础，CWMT’ 2015评测中系统的表现足以说明了这一事实。参加本次评测的ISTIC系统如后文所述，主要内容有系统框架、实施细节，在评测数据上的性能表现等。

## 2 系统框架

ISTIC研究团队提交的维汉、藏汉和蒙汉三个测试任务的翻译结果都采用了统一的翻译策略：利用多个机器翻译系统来生成多个翻译输出，之后通过系统融合来生成融合的翻译结果，根据开发集上的评价结果来进行选择，如果融合系统的翻译结果最优，则采用融合系统来翻译测试集；如果融合系统没有达到最优，则采用最优的单系统来翻译测试集。在机器翻译多系统中，ISTIC团队使用了基于短语的统计机器翻译方法和基于层次短语的统计机器翻译方法，每一个方法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<sup>[1]</sup>。系统的总体框架见图1。

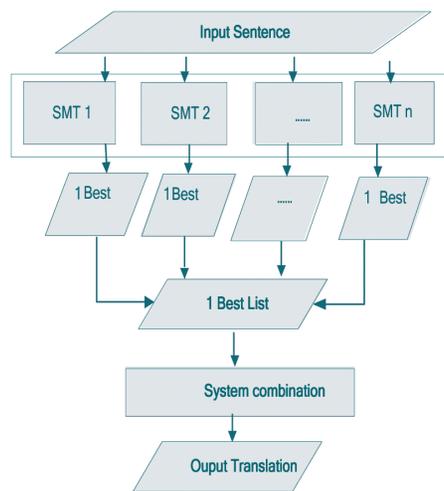


图1 ISTIC机器翻译系统框架

## 2.1 多机器翻译系统

研究团队采用的统计机器翻译方法包括基于短语的统计机器翻译<sup>[2,3]</sup>和基于层次短语的统计机器翻译<sup>[4]</sup>，前者使用连续的短语作为翻译单元，后者采用了非连续的短语，都采用了对数线性模型来进行翻译结果的遴选。为了获取尽可能多的翻译结果来为后续的系统融合服务，研究团队采用了如下措施：1) 不同的词对齐，除了常用的 GIZA++ 双向词对齐之后进行 grow-diag-final-and 扩展得到多对多的词对齐之外，还采用了 Berkeley<sup>①</sup>的词对齐。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) 不同的翻译引擎：研究团队使用了 Moses<sup>②</sup>的基于短语的统计机器翻译引擎和基于层次短语的统计机器翻译引擎，还使用了 NiuTrans<sup>③</sup>的基于短语的统计机器翻译引擎。通过不同措施的选择配置就可以得到多个翻译系统。

## 2.2 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<sup>[1, 6-9]</sup>。ISTIC 采用词级的基于混淆网络 (Confusion-network) 的系统融合算法<sup>[1]</sup>，融合多个翻译系统的输出结果来形成一致性输出 (Consensus Output)，作为最后的翻译结果。混淆网络包含一序列替换词集合，其中可能包括 null 词，并附带有相应的打分。通过从每一个替换词集合中选择一个词产生得分最好的序列。研究团队的整个系统融合框架如图 2 所示。对于每一个基于统计机器翻译的单引擎，采用了不同策略来生成 1-best，合并每一个单引擎的 1-Best 为 1-Best Lists 来进行系统融合。研究团队

采用了最小贝叶斯风险解码技术<sup>[10]</sup>来动态地选择骨架 (Backbone)。在将每个假设与骨架进行词对齐时选用了 IHMM (Indirect Hidden Markov)<sup>[11]</sup>假设对齐方法，利用相似度模型和扭曲模型，应用 Viterbi 算法获取最好的假设对齐，之后进行对齐归一化得到一对一对齐，生成混淆网络。

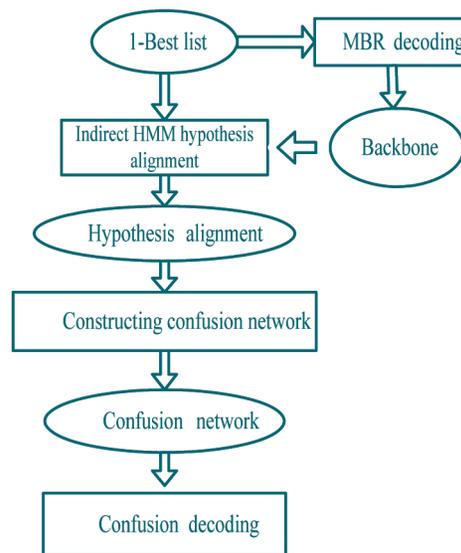


图 2 ISTIC 机器翻译系统融合框架

给定源语言句子  $f$ ，混淆网络解码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， $h_m$  为特征函数， $\lambda_m$  为特征权重。混淆网络解码过程中，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) 词后验概率；
- 2) 语言模型特征；
- 3) 基于距离的重排序特征；
- 4) 词惩罚；

每个特征函数的权重由最小错误训练算法训练得出。

① <http://code.google.com/p/berkeleyaligner/downloads/list>

② <http://www.statmt.org/ Moses/>

③ <http://www.niplab.com/NiuPlan/NiuTrans.ch.html>

### 3 数据的使用和处理

在蒙汉、藏汉、维汉三个翻译任务中，训练语料均采用了评测方发布的针对每个任务的所有训练语料，语言模型训练数据采用了相应任务的训练语料的中文部分和评测方发布的搜狗新闻语料库，所有的基于统计的机器翻译单系统的参数都是在该项目发布的开发集上进行训练。

对中文数据进行的处理有：中文的分词和全角变半角，采用了 Urhen2.2<sup>④</sup>的中文分词工具。对藏语数据进行了基于最大熵的分词处理，对蒙语和维语数据分别进行了 Tokenization 的处理。翻译工具采用了 Moses 和 NiuTrans，词对齐工具采用了 GIZA++<sup>⑤</sup>（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）<sup>②</sup>和 Berkeley 对齐工具，语言模型工具采用了 Srilmm<sup>⑥</sup> 工具包来获取 5 元语法概率信息。中文的后处理只是去掉了空格。Berkeley 对齐在藏汉任务上采用了 15 次迭代，在蒙汉任务上选择了 5 次迭代。

### 4 实验

本研究分别在藏汉、维汉、蒙汉三个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 CWMT2013 评测组织方发布的打分工具。在测试集上的打分使用了 CWMT2015 评测组织方开放的在线评测平台。

研究团队使用 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。

#### 4.1 维汉翻译结果

在维汉翻译任务上，研究团队在开发集上测试了 7 个翻译系统。表 2 列出了在此任务上参与系统融合的单翻译结果以及融合结果在开发集上的打分。表中的系统命名采用了“翻译工具-翻译模型-语言模型”的格式。翻译工具分别使用了 Moses 和 NiuTrans，Moses1 和 Moses2 表示采用了不同版本的两个 Moses，前者为分步运行 Moses，后者采用 Moses 的“Experiment management system”。词对齐研究团队采用了 GIZA++ 的词对齐。翻译模型采用了基于短语的统计机器翻译模型（Phrase）、基于层次短语的统计机器翻译模型（Hiero）。语言模型中，smalllm 表示只使用了当前任务的训练语料的中文部分来训练 5 元的语言模型，biglm 表示使用了当前任务的训练语料的中文部分加上 Sogou 语料来训练 5 元的语言模型。

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	
藏汉	训练集	藏语	126530	35843	13.40
		汉语	126530	45301	9.74
	开发集	藏语	650	2258	19.73
		汉语	2600	2877	13.65
	测试集	藏语	729	1520	12.41
	维汉	训练集	维语	139801	137921
汉语			139801	81656	19.71
开发集		维语	700	5946	26.05
		汉语	2800	7006	24.70
测试集		维语	1000	8110	21.45
蒙汉		训练集	蒙语	124240	81357
	汉语		124240	64277	14.71
	开发集	蒙语	1000	2246	7.40
		汉语	4000	2831	7.31
	测试集	蒙语	1001	3000	5.92

④ <http://www.openpr.org.cn/index.php/NLP-Toolkit-For-Natural-Language-Processing/68-Urhen-A-Chinese/English-Lexical-Analysis-Toolkit/View-details.html>

⑤ <http://giza-pp.googlecode.com/>

⑥ <http://www.speech.sri.com/projects/srilmm/download.html>

表 2 维汉任务在开发集上的比较

系统	NIST	BLEU	BLEU_SBP	GTM	mWER	mPER	ICT
Moses1-phrase-smalllm	10.5785	0.5395	0.5122	0.8123	0.5074	0.3329	0.4642
Moses1-hiero-biglm	11.0808	0.5705	0.5482	0.8303	0.4917	0.3148	0.4842
Moses2-phrase-smalllm	10.5343	0.5124	0.4953	0.8094	0.5337	0.3391	0.4330
Moses2-phrase-biglm	10.9255	0.5533	0.5314	0.8253	0.5077	0.3227	0.4671
Moses2-hiero-smalllm	10.5322	0.5124	0.4953	0.8091	0.5341	0.3396	0.4327
NiuTrans-phrase-smalllm	10.7617	0.5412	0.5204	0.8136	0.5179	0.3293	0.4608
NiuTrans-phrase-biglm	11.2141	0.5735	0.5546	0.8326	0.5012	0.3139	0.4810
Inhmm system combination	10.2833	0.5551	0.5276	0.8011	0.4840	0.3434	0.4886

表 3 维汉任务在测试集上的比较

系统	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT	METEOR	TER
NiuTrans-phrase-biglm	0.4190	0.4390	0.3956	10.07	10.09	0.7645	0.5009	0.3504	0.4473	0.5782	0.4415
Moses1-hiero-biglm	0.3839	0.4004	0.3565	9.47	9.49	0.7481	0.5016	0.3767	0.4286	0.5519	0.4537

从表 2 中可以看出，同样的设置参数下，大的语言模型要优于小的语言模型，NiuTrans 的翻译效果要优于 Moses<sup>⑦</sup>。对这 7 个系统进行系统融合，这里使用了 1best 的翻译假设。表 2 显示出 1best 的融合结果没有达到预期的融合目的，虽然比最差的翻译结果要好，但是比最好的翻译结果要差。因此在测试集中研究团队选择了最优的单系统 NiuTrans-phrase-biglm 的翻译结果作为 primary，第二优的单系统 Moses1-hiero-biglm 的翻译结果作为对比进行提交，表 3 给出测试集上的翻译效果对比结果，从中可以看出在开发集上的最优单系统仍然要优于第二优的单系统。需要说明的是：由于当时参加评测的时间比较紧张，在维汉上研究团队没有使用 Berkeley 对齐进行实验。

## 4.2 藏汉翻译结果

在藏汉翻译任务上，研究团队在开发集上测试了 6 个翻译系统。表 4 列出了在此任务上参与系统融合的单系统翻译结果，以及融合结果在开发集上的打分。表中的系统命名采用了“词对齐\_语言模型\_翻译模型\_翻译工具”的格式。词对齐研究团队采用了 Berkeley 的词对齐和 GIZA++ 的词对齐两种，语言模型中，smalllm 表示只使用了当前任务的训练语料的中文部分来训练 5 元的语言模型，biglm 表示使用了当前任务的训练语料的中文部分加上 Sogou 语料来训练 5 元的语言模型。翻译模型采用了基于短语的统计机器翻译模型 (Phrase) 基于层次短语的统计机器翻译模型 (Hiero)。翻译工具分别使用了 Moses 和

⑦ 这个主要的原因是 Moses 翻译中采用了二进制。

NiuTrans。

从表4中可以看出，同样的设置参数下，大的语言模型要优于小的语言模型，Berkeley的对齐要优于GIZA的对齐，Hiero的翻译效果要优于Phrase的翻译效果，NiuTrans的翻译效果要优于Moses。从这6个系统中研究团队选择了前4个系统进行系统融合，这里分别使用了1best的翻译假设和100best的翻译假设，表4中最后两行显示，1best的融合结果要优于100best。表4也显示出1best的融合结果达到了预期的融合目的，优于每个单系统的翻译，因此在测试集中研究团队选择了1best的融合。

结果作为primary进行提交，同时采用开发集上最优单系统Berkeley-biglm-Phrase-NiuTrans翻译测试集作为对比结果。表5给出测试集翻译效果对比结果，可以看出，在测试集上系统融合也达到了比单系统更好的结果。

### 4.3 蒙汉翻译结果

在蒙汉翻译任务上，研究团队在开发集上测试了5个翻译系统。表6列出了在此任务上参与系统融合的单个翻译结果在开发集上的打分。表中的系统命名采用了“词对齐\_语言模型\_翻译模型\_翻译工具”的格式。

表4 藏汉任务在开发集上的比较

系统	NIST	BLEU	BLEU_SBP	GTM	mWER	mPER	ICT
Berkeley_biglm_Hiero_moses	10.8134	0.6584	0.6378	0.8504	0.3522	0.2177	0.6388
Berkeley_biglm_Phrase_moses	10.7301	0.6398	0.6145	0.8442	0.3778	0.2337	0.6253
Berkeley-biglm-Phrase_NiuTrans	10.9445	0.6735	0.6567	0.8561	0.3336	0.2081	0.6428
Berkeley-smalllm-Phrase_NiuTrans	10.6057	0.6387	0.6196	0.8392	0.3650	0.2291	0.6114
Giza-biglm-Phrase-NiuTrans	10.1647	0.6176	0.6031	0.8253	0.3794	0.2541	0.5982
Giza-smalllm-Phrase-NiuTrans	9.9586	0.5994	0.5843	0.8132	0.3916	0.2713	0.5752
1best inhmm-system-combination	11.0447	0.6830	0.6630	0.8583	0.3260	0.2068	0.6580
100 best inhmm-system-combination	10.7737	0.6504	0.6382	0.8622	0.3430	0.2094	0.6181

表5 藏汉任务在测试集上的比较

系统	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT	METEOR	TER
1best inhmm-system-combination	0.8809	0.8973	0.8773	12.90	13.03	0.9442	0.1384	0.0839	0.9218	0.9052	0.1014
Berkeley-biglm-Phrase_NiuTrans	0.8751	0.8895	0.8684	12.86	12.99	0.9438	0.1441	0.0885	0.9137	0.9001	0.1069

词对齐研究团队采用了Berkeley的词对齐和GIZA++的词对齐两种，语言模型中，smalllm表示只使用了当前任务的训练语料的中文部分来训练5元的语言模型，biglm表示使用了当前任务的训练语料的中文部分加上Sogou语料来训练5元的语言模型。翻译模型采用了基于短语的统计机

器翻译模型(Phrase)基于层次短语的统计机器翻译模型(Hiero)。翻译工具分别使用了Moses和NiuTrans。从表6中可以看出，同样的设置参数下，大的语言模型要差于小的语言模型，Berkeley的对齐要优于GIZA的对齐，NiuTrans的翻译效果要优于Moses。

表 6 蒙汉任务在开发集上的比较

系统	NIST	BLEU	BLEU_SBP	GTM	mWER	mPER	ICT
Giza-smalllm-phrase-NiuTrans	4.8869	0.2404	0.2258	0.5361	0.6069	0.5403	0.3953
Berkeley-smalllm-phrase-NiuTrans	9.1365	0.5645	0.5234	0.7617	0.3827	0.3173	0.6179
Giza-biglm-phrase-NiuTrans	4.9575	0.2406	0.2242	0.5297	0.6232	0.5471	0.3847
Berkeley-biglm-phrase-NiuTrans	8.9369	0.5451	0.5031	0.7537	0.3963	0.3273	0.6118
Berkeley-biglm-phrase-Moses	8.8186	0.5383	0.4907	0.7526	0.3992	0.3329	0.6153

表 7 蒙汉任务在测试集上的比较

系统	BLEU5-SBP	BLEU5	BLEU6	NIST6	NIST7	GTM	mWER	mPER	ICT	METEOR	TER
Berkeley-smalllm-phrase-NiuTrans	0.3642	0.4061	0.3749	7.1364	7.1589	0.6060	0.5429	0.4831	0.4895	0.5038	0.5066
Berkeley-biglm-phrase-NiuTrans	0.3618	0.3933	0.3642	6.6981	6.7233	0.6007	0.5431	0.4830	0.5034	0.4916	0.5044

由于时间紧张以及各个单系统的翻译效果差别特别大，蒙汉任务上研究团队没有进行系统融合。在测试集中研究团队选择了最优单系统 Berkeley-smalllm-phrase-NiuTrans 在测试集上的翻译结果做为 primary，第二优单系统 Berkeley-biglm-phrase-NiuTrans 的翻译结果作为 contrast 进行提交。表 7 给出测试集翻译效果对比结果，可以看出，在开发集上的最优单系统仍然要优于第二优的单系统。

## 5 评测后的启发

中信所参加了 CWMT' 2015 维汉、藏汉和蒙汉三个少数民族语言的翻译任务的机器翻译项目，在维汉和藏汉翻译任务上取得了较好的成绩，但是在蒙汉翻译任务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，研究团队积累了下述经验：

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。由于缺少精通少数民族语言的人才，对于源

语言端语料出现的问题不能准确洞察，只能给以简单的分词和标点隔离。尤其是蒙汉翻译任务的训练语料质量比较差，蒙语中夹杂汉语，汉语中夹杂蒙语，很难对语料进行高质量的清洗，导致翻译的效果也大打折扣。因此需要加强对源语言端的语料的清洗和预处理研究。

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。但是研究团队的系统融合在藏汉翻译效果较好，但是在维汉上并没有同样的体现，这主要是因为对多个单系统各自的优势和缺陷分析不够细致，仍需要进一步进行筛选以期得到互补性融合，今后，研究团队致力于采取进一步改善单系统遴选方案来增强系统融合的翻译效果。

## 展望

本文重点介绍了中信所机器翻译研发工程暨 CWMT' 2015 机器翻译评测结果，客观地反映了中信所相关工作。中信所的案例，从一个侧面反应了中国机器翻译的情况。未来，就机器翻译而

言, 产学研的一体化联动将进入前所未有的活跃期, 本文作者计划今后持续带给大家相关文章, 不仅局限于中信所, 还要聚焦于国内外重点会议、重要任务, 以及重要产品等, 期待读者的关注。

#### 参考文献

- [1] He X, Yang M, Gao J, et al. Indirect+HMM-based Hypothesis Alignment for Combining Outputs from Machine Translation Systems[C]// Conference on Empirical Methods in Natural Language Processing. 2008:25-27.
- [2] Koehn P, Och F J, Marcu D. Statistical Phrase-based Translation[C]// Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology. Association for Computational Linguistics, 2003:127-133.
- [3] Och F J, Ney H. A Systematic Comparison of Various Statistical Alignment Models[J]. Computational Linguistics, 2003, 29(1):19-51.
- [4] Chiang D. A Hierarchical Phrase-Based Model for Statistical Machine Translation[C]// ACL 2005, Meeting of the Association for Computational Linguistics. Proceedings of the Conference, 25-30 June 2005, University of Michigan, USA. 2005:263-270.
- [5] Xiao T, Zhu J, Zhang H, et al. NiuTrans: an Open Source Toolkit for Phrase-based and Syntax-based Machine Translation[C]// ACL 2012 System Demonstrations. 2012:19-24.
- [6] Bangalore S, Bordel G, Riccardi G. Computing Consensus Translation from Multiple Machine Translation Systems[C]// Automatic Speech Recognition and Understanding, 2001. ASRU '01. IEEE Workshop on. IEEE, 2002:351-354.
- [7] Fiskus J G. A Post-processing System to Yield Reduced Word Error Rates: Recognizer Output Voting Error Reduction[C]// IEEE Automatic Speech Recognition and Understanding Workshop. 1997:347-354.
- [8] Rosti A V I, Ayan N F, Xiang B, et al. Combining Outputs from Multiple Machine Translation Systems[C]// Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics, Proceedings, April 22-27, 2007, Rochester, New York, USA. 2007:228-235.
- [9] Rosti A V I, Matsoukas S, Schwartz R M. Improved Word-Level System Combination for Machine Translation[C]// ACL 2007, Proceedings of the Meeting of the Association for Computational Linguistics, June 23-30, 2007, Prague, Czech Republic. 2007.
- [10] Sim K C, Byrne W J, Gales M J F, et al. Consensus Network Decoding for Statistical Machine Translation System Combination[C]// International Conference on Acoustics, Speech and Signal Processing. 2007:IV105-IV108.
- [11] Andreas Stolcke. SRILM-An Extensible Language Modeling Toolkit[C]// In Proceedings of International Conference on Spoken Language Processing, Denver, Colorado. 2002:901-904.