文章编号：1003-0077(2018)11-0001-07

基于统计和神经网络的蒙汉机器翻译研究

任 众，侯宏旭，武 静，王洪彬，李金廷，樊文婷，申志鹏

（内蒙古大学 计算机学院，内蒙古 呼和浩特 010021）

摘 要：该文对基于传统统计模型的蒙汉机器翻译模型和基于神经网络机器翻译模型进行了研究。其中，神经网络翻译模型分别为基于 CNN、RNN 的翻译模型，并通过将所有翻译模型结果进行句子级融合得到一个融合模型。面对蒙汉翻译面临资源稀少、蒙古文形态复杂等困难，该文提出多种翻译技术，对各个模型进行改进，并对蒙古文进行形态分析与处理。在翻译效果最好的 CNN 模型上，采用字和短语融合训练方法；基于 RNN 的翻译模型除用上述方法外，还采用 Giza＋＋指导对齐技术调整 RNN 注意力机制；针对 SMT 采用了实验室提出的重对齐技术。该文对实验结果进行了对比和分析，这三种技术方法对相应系统翻译效果有显著提升。此外，蒙古文形态分析与处理对缓解数据稀疏、提升译文质量也有重要作用。

关键词：汉蒙机器翻译；蒙古文形态分析；融合训练方法

中图分类号：TP391　　**文献标识码**：A

Research on Mongolian-Chinese MT Based on Statistical and Neural Network

REN Zhong, HOU Hongxu, WU Jing, WANG Hongbin, LI Jinting, FAN Wenting, SHEN Zhipeng

(College of Computer Science, Inner Mongolia University, Hohhot, Inner Mongolia 010021, China)

Abstract: This paper investigates the statistical Mongolian-Chinese machine translation model and the neural networkbased machine translation model, including CNN and RNN translation models. To address the low-resource and rich morphology issue, this paper proposes multiple methods to improve the three translation models. For the top-performed CNN model, we apply the character and phrase joint-training method. We further augmented the improvement to RNN model with a Giza＋＋ guided alignment to attention. We designed a realignment method to the SMT model. Experiments indicate the above approaches improve the Mongolian-Chinese translation performance significantly.

Keywords: Mongolian-Chinese machine translation; Mongolian morphology process; jointtraining method

0 引言

蒙汉机器翻译属于稀少资源及少数民族语言翻译领域任务，对于促进语言、文字和文化交流，以及民族团结进步具有重要意义。然而，此类翻译任务普遍面临双语对齐语料不足，资源稀少，蒙古文形态复杂，翻译研究时间短，成果少等困难。

本文中共提及四个系统分别为：CNN(Convolutional Neural Network)系统、RNN(Recurrent Neural Network)系统、SMT (Statistical Machine Translation)系统和以上三个系统的句子级融合系统。其中，CNN 系统取得最好的翻译效果(BLEU5－SBP＝0.702 4)，其次分别是融合系统、RNN 系统以及 SMT 系统。蒙汉翻译任务主要面临的困难是资源稀少和蒙古文形态复杂。针对这两个问题，本文的 CNN 系统和 RNN 系统采用短语和字融合训练方法以获取多粒度特征，RNN 还采用了 Giza＋＋对齐指导模型来调整注意力机制产生的对齐；本文的 SMT 系统采用了从小粒度到大粒度的重对齐算法。通过蒙古文形态分析，对格的附加成分进行预处理。这几种模型和语言学方面的处理技术都使得蒙汉翻译实验结果有显著提升。本文将 CWMT2017 去重后的蒙汉测试集作为测试集合进行测试，并对

收稿日期：2018-01-16　**定稿日期**：2018-03-21

基金项目：国家自然科学基金(61362028)

实验结果进行了比较和说明。

1 蒙古文形态分析

蒙古文词汇丰富,形态构成复杂。因此,蒙古文的形态分析是语料预处理的关键步骤,对翻译结果有着重要影响[1]。本文对蒙古文语料进行了格的附加成分的预处理。

蒙古文有多种词缀。其中一种称为格的附加成分,通常为蒙古文词缀中的最后一个附加成分,包含部分句法信息。例如,名词在句子中作为主语还是宾语,就通过其后缀接的格的附加成分来说明。对格的附加成分的处理对词本身的词性、词义影响不大。蒙古文中格的数量庞大,因此对其进行预处理可以充分缓解数据稀疏。有别于一般处理蒙古文词缀的方法,我们只对蒙古文进行格的附加成分预处理[2]。

图 1 列出蒙古文中的七类格的附加成分的样例。我们通过编码方式识别出格,进行处理。

1	2	3	4	5	6	7
定格	向位格	宾格	凭借格	从比格	和同格	联合格
						

图 1 格的附加成分

格的附加成分处理方法主要有三种[3]。第一种是将控制符去除,然后将格的附加成分与前面的词干进行连接,形成一个新的词,这种处理方法意义不大。第二种方法是将控制符与格的附加成分一同去除,只留下词干部分。第三种方法是将格的附加成分保留,但与词干分开,作为新的处理单元。经过实验证明第二种方法在缓解数据稀疏方面最为有效,在 SMT 上表现最好。本文在 SMT 实验中采用去除格的附加成分的方法(第二种方法)。然而,这种方法在 NMT 系统上却不如切分并保留格(第三种方法)效果更好。因为,NMT 擅长处理更多特征,去除格会丢失一部分句法特征,不利于 NMT 网络的学习。

表 1 分析了格的处理对语料数据稀疏性的影响。没有进行任何处理的蒙古文语料单频词数量巨大,处理后的蒙古文数据稀疏问题明显改善。格处理对翻译结果的提升也比较显著。

表 1 蒙古文预处理数据稀疏分析

蒙古文处理方法	单频词	总词频
无处理	86 495	158 623
第二种格处理	46 754	94 472
第三种格处理	48 841	98 617

2 系统描述

在实验中本文搭建了 CNN、RNN 以及 SMT 三个系统。在基于神经网络的 CNN 和 RNN 模型中采用了短语和字融合训练方法;对 RNN 模型中还采用了 Giza++对齐指导模型;在 SMT 上采用了基于小粒度向大粒度重对齐的 SMT 模型[4]。SMT 的译文结果 BLEU 得分低于 CNN 和 RNN 系统译文结果 BLEU 的得分,但该重对齐模型为 SMT 基线系统带来明显的翻译效果提升。本文采用了基于双语 N-gram 词嵌入的相似度重排序方法对三个各异的系统进行句子级重排序,但由于语料规模较小,双语词嵌入相似度准确率受限,融合系统结果并没有高于最优的单系统。

2.1 深度神经网络 NMT 系统

2.1.1 CNN 系统概述

本文的 CNN 系统采用以 Facebook AI Research 开源系统 fairseq 为基础构建的蒙汉翻译系统[5-6]。该系统实现序列到序列翻译,系统架构分为编码器和解码器两部分。两部分均利用 CNN 卷积核获取短距离依赖信息,并通过增加 CNN 深度来获取远距离依赖信息。此外,每层解码器配备一个注意力机制。下面对该架构进行介绍。

1. 位置向量。因 CNN 无法获取输入词在句子中的位置信息,故需要在输入词时为词添加位置信息。CNN 输入为词向量与位置向量相加形成,这里位置向量为词在句子中的绝对位置向量,如式(1)~式(4)所示。

$$x = (x_1, \cdots, x_m) \tag{1}$$

$$w = (w_1, \cdots, w_m) \tag{2}$$

$$p = (p_1, \cdots, p_m) \tag{3}$$

$$e = (w_1 + p_1, \cdots, w_m + p_m) \tag{4}$$

其中，x 表示输入序列，w 表示输入序列对应的词向量，p 表示位置向量，e 表示词向量和位置向量相加而成的 CNN 的输入向量。

2. 卷积层结构。编码器和解码器使用相同的卷积层结构，每一层均由一个一维的卷积网络加一个非线性层组合而成，如式(5)、式(6)所示。

$$v([A\ B]) = A \otimes \sigma(B) \tag{5}$$

$$h_i^l = v(W^l[h_{i-\frac{k}{2}}^{l-1},\cdots,h_{i+\frac{k}{2}}^{l-1}] + b_w^l) + h_i^{l-1} \tag{6}$$

其中，v 表示非线性层，$\otimes$表示逐元素相乘，h_i^l 表示第 l 层第 i 个输出，W^l 表示第 l 层的卷积核矩阵，b_w^l 表示第 l 层的卷积核矩阵的偏置，h_i^{l-1} 表示第 $l-1$ 层第 i 个输出。

3. 多步注意力。解码器部分为多层深度 CNN 结构，系统为解码器每一层都配备一个注意力机制，如式(7)～式(9)所示。

$$d_i^l = W_d^l h_i^l + b_d^l + g_i \tag{7}$$

$$a_{ij}^l = \frac{\exp(d_i^l \cdot z_j^u)}{\sum_{t=1}^{m}\exp(d_i^l \cdot z_t^u)} \tag{8}$$

$$c_i^l = \sum_{j=1}^{m} a_{ij}^l (z_j^u + e_j) \tag{9}$$

其中，g_i 表示上一个生成的目标词的词向量，h_i^l 表示解码器部分第 l 层第 i 个隐层状态，z_j^u 表示编码器部分最后一层第 j 个输出，a_{ij}^l 表示第 l 层第 i 个隐层状态和源句子中第 j 个元素的对齐系数，e_j 表示编码器部分第 j 个输入，c_i^l 表示解码器部分第 l 层第 i 个注意力向量。

2.1.2　RNN 系统概述

RNN 在基于注意力机制 GroundHog 开源 RNN 系统上重现神经网络的编码、解码翻译模型[7]。系统由双向 RNN 编码器[8]和解码器构成。注意力机制基于解码器隐层状态 z_{i-1} 以及相应编码器隐层状态 h_j，针对源语言句子每个处理单元获得对齐 a_j，由此得到 c_i 作为编码器输出的摘要。在有了前一个编码器生成的词 w_{i-1}，以及上一个编码器隐层状态 z_{i-1} 和向量 c_i 的基础上，解码器更新隐层状态 z_i。根据 z_i 对目标词典中所有词计算概率分布，如式(10)～式(14)所示。

$$a_{ij} = \alpha(z_{i-1}, h_j) \tag{10}$$

$$c_i = \sum_{j=1}^{T} a_{ij} h_j \tag{11}$$

$$z_i = \phi_\theta(w_{i-1}, z_{i-1}, c_i) \tag{12}$$

$$e(k) = w_k^T z_i + b_k \tag{13}$$

$$p(w_i = k \mid w, c_i) = \frac{\exp(e(k))}{\sum_t \exp(e(t))} \tag{14}$$

2.1.3　对齐融合

本文使用的是含有注意力机制神经网络翻译系统。注意力建模预测某个目标语言单词时，对于源语言句子中单词的依赖关系，称作对齐权重。基于注意力机制的神经网络翻译模型中，对齐权重表示当前产生的目标词与源语言单词对齐的概率。而对齐权重的产生依赖于上一个隐层状态、当前的输出及前一步产生的目标单词。神经网络翻译模型对齐权重的产生依赖的信息过多导致独立性较差。当前一个单词出现翻译错误会影响神经网络下一步对齐结果的准确性。与注意力机制模型的对齐相比，IBM 翻译模型产生的对齐结果更为准确。在 IBM 翻译模型中目标语言句子中的第 i 个单词与源语言句子中第 j 个单词对齐的概率仅依赖于源语言的第 j 个词以及第 j 个词的位置。每个对齐位置的产生过程均有很好的独立性。所以本文使用 Giza++的对齐结果[9]指导神经网络训练从而提高翻译结果。

神经网络翻译模型的代价函数是利用系统的翻译结果与标准译文的相似度衡量的。为了提高注意力机制产生的对齐结果，本文把 Giza++产生的对齐结果作为标准的对齐加入到神经网络翻译模型的训练中，使神经网络注意力对齐权重的训练过程变为有监督的训练。通过修改代价函数将 Giza++的对齐加入到神经网络翻译模型中，神经网络翻译模型的代价函数包含两部分：一部分是原本的翻译代价 $L_{de}(y,x)$，另一部分是新加入的对齐代价，如式(15)所示。

$$L_{al}(A,a) = \frac{1}{T}\sum_t\sum_s (A_{st} - a_{st})^2 \tag{15}$$

其中，A 是统计方法的对齐矩阵，a 是神经网络翻译模型的注意力权重。T 是目标语言句子长度，s 和 t 是源语言句子和目标语言句子单词下标。

神经网络翻译模型与基于统计方法的对齐结果意义不同，若要融合两种对齐，需把统计方法的对齐结果转换为可以融合到神经网络的训练中的对齐。加入对齐后重新定义的代价函数，如式(16)所示。

$$\text{Loss} = w_{al} \cdot L_{al}(A,a) + (1 - w_{al}) \cdot L_{de}(y,x) \tag{16}$$

其中，w_{al} 是对齐代价和翻译代价的线性组合权重。

2.1.4　多粒度融合

目前的神经网络机器翻译(NMT)模型倾向于为解码器采用较小粒度的翻译单元，来减少由有限的目标词典引起的未登录词(OOV)问题[10-11]。然

而,句子级别翻译任务本身需要更大的语义单位。因此,基于短语的 SMT 模型也优于基于词的 SMT 模型。此外,短语和词包含比字或字符更完整和精确的语义特征[12]。通过句法分析或共现概率等方法获得的短语包含更多的句法特征。尤其对于缺乏局部特征表示的 RNN,短语能够为解码提供更多的局部语义和句法特征选择。

如何解决这种矛盾呢? 本文提出了一种短语和字符多粒度融合方法。我们引入短语和字词混合解码器,在训练中使短语和字词共享向量空间和概率空间,且不设置选择粒度的优先级,解码器根据共享概率空间中的概率分布选择粒度。这有别于一般的一个解码器对应一个解码器的模型,一个源语言句子映射到解码器中的两个不同切分粒度的目标句子,即短语级和字级的汉文句子序列。2.1.3 中所述的加入对齐信息的有监督训练的注意力机制,将不同粒度的目标语言单词与源语言单词对齐,同时得到短语和字词在同一空间的概率分布。如图 2 所示,左边为汉文字级的解码器,右边为汉文短语级的解码器。本文提及的短语是从语料库中得到的连续出现的相邻词和字,可以通过基于浅层句法分析[13]或共现词概率分析[14]等方法获得。

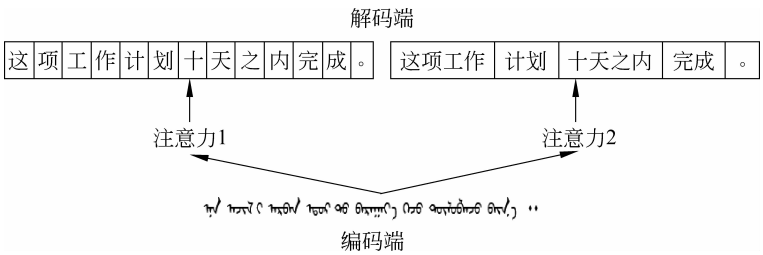

图 2 多粒度融合的注意力循环神经网络解码器模型

多粒度融合方法,是由解码器来选择粒度较大还是较小的单元作为翻译候选。短语和字符的联合训练尤其适合于稀少资源翻译任务,一方面丰富了 NMT 解码器的多个粒度特征,另一方面扩大了训练语料。短语和字按照词频排序进入目标词典。中文常用词汇和短语的数目庞大,但常用汉字不超过一万字。因此,即使限制目标词典大小,绝大部分中文字也都在词典中。联合训练在解码端共享字和短语概率分布,因而即使目标词典中没有充足的短语表示,也能够通过选择概率相近的字来避免未登录词问题。此外,由于稀少资源翻译中一般语料和目标词典都很小,训练时间和复杂度有限,可以在训练中不限定目标词典规模。这种情况下联合训练方法充分扩大了词典和多粒度特征,可以取得更理想的效果。我们提出的联合训练方法在蒙汉机器翻译 NMT 系统上提升明显。

2.2 SMT 系统

2.2.1 系统描述

我们使用 Moses 作为统计机器翻译的基础系统[15],采用基于短语的统计机器翻译解码方式进行解码[16],GIZA++进行词对齐,采用 Och 等提出的 MERT 进行权值的优化[17],使用 3-gram 语言模型。

2.2.2 重对齐模型

本文在 SMT 系统上也采用了多粒度的融合。基于蒙古文和汉文都存在多粒度切分的可能,而统计机器翻译的对齐模型和解码模型,分别在小粒度和大粒度上会取得更好的效果。因此,我们提出了一个基于小粒度融合为大粒度的转换算法。对齐阶段采用汉字和蒙古文词干的小粒度,在解码阶段无误差合成大粒度的汉文词和蒙古文词进行解码,以通过在保证解码大粒度优势不变的前提下,优化对齐模型。从而在整体上优化统计机器翻译模型,获得更好的译文质量。该方法流程图如图 3 所示。

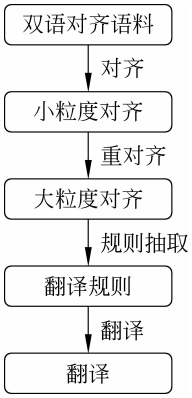

图 3 多粒度对齐融合模型流程图

2.3　系统融合

本文使用基于双语 N-gram 词嵌入的相似度计算方法来对多个系统译文进行句子级重排序。该重排序方法由于没有利用各系统内部产生的概率分布，因而可以对多个完全不同的模型译文进行重排序。也正因为如此，其排序算法全部依赖于双语向量表示。然而在蒙汉双语语料规模较小、单语语料领域庞杂的情况下，在重排序上的表现不稳定。参与融合的三个系统为第 3 节描述的 CNN、RNN 和 SMT 系统字级别译文。

3　实验

3.1　实验数据

实验数据均基于 CWMT2017 提供的蒙汉评测提供的训练集(26.16 万句)、开发集(1 000 句)和测试集(1 001 句)。由于测试集重复问题，本文实验中所采用测试集为去重后的 678 句。评测指标为 BLEU5-SBP。实际训练数据包含训练集和开发集经过长度处理后的全部数据。中文语料处理方面采用中科院计算所开发的 ICTCLAS 中文分词系统按字进行切分。

为了分别分析多粒度融合方法和对齐指导方法对 RNN 的影响，本文从 CWMT2017 提供的蒙汉双语语料中随机抽取了五组包含 1 000 句的语料作为测试集。

3.2　实验配置

本文的 CNN 系统使用 Facebook AI Research 开源系统 fairseq。CNN 翻译系统，系统参数配置如下：编码器层数 5 层，解码器层数 9 层，解码器的每一层均配备一个注意力机制。编码器和解码器的核宽度均为 3，词向量维度 500，每个隐层单元数量 500。batchsize 32，训练算法 Nesterov's accelerated gradient (NAG)。

RNN 系统使用基于注意力机制的开源实现 dl4mt-tutorial 作为 RNN 翻译系统。系统参数配置如下：词向量维度 500，隐层单元数量 500，batch-size 32，训练算法 SGD，优化算法采用 adadelta。

SMT 系统使用 Moses 基础系统。通过 Och 等提出的 GIZA++进行词对齐，采用 Koehn 等提出的基于短语的统计机器翻译解码方式进行解码。采用 Och 等提出的 MERT 进行权值的优化，使用 3-gram 语言模型。采用训练集和开发集为目标语言训练语言模型，并未采用评测主办方提供的中文单语语料。

因蒙汉翻译训练语料小，词汇规模小，训练时间短，所以采用较大词典对计算复杂度、计算时间和内存占比不仅不会造成过大压力，还能使取得的翻译结果明显提升。因此，CNN 和 RNN 源和目标语言词典均采用全部训练语料中获取的词(实验中字和短语融合系统词典最大，蒙文和汉文词典均为 8 万词左右)。

3.3　实验结果与分析

表 2 所示为四个系统以及各模型对比系统。提交的单个系统中 CNN 系统取得的效果最好，BLEU5-SBP 值达到 0.702 4；其次是 RNN 系统为 0.599 8，SMT 系统为 0.568 6。融合系统在对以上三个系统进行句子级别融合时，由于 2.3 节提到的原因，在开发集上融合系统表现高于单个最好成绩。但在评测中融合系统 BLEU5-SBP 的值为 0.668 0，并没有超过单个 BLEU5-SBP 值最高的系统。

表 2　实验结果

	系统	BLEU5-SBP
CNN	Contrast b(多粒度融合)	0.702 4
	CNN 基线系统	0.579 6
RNN	Contrast c(多粒度融合＋对齐指导)	0.599 8
	RNN 基线系统	0.507 5
SMT	Contrast d (重对齐)	0.568 6
	SMT 基线系统	0.514 7
	Primary a(句子级融合系统)	0.668 0

本文针对字和短语融合方法、对齐指导方法和重对齐方法在各系统基线系统上进行了比较试验。所有基线系统配置按系统描述所述配置，蒙文进行了格的附加成分预处理，汉语部分均为分字处理。

实验可见，基于 NMT 的机器翻译，无论是 RNN 和 CNN，其在蒙汉机器翻译上的表现已经超过了 SMT。然而，简单利用神经网络基础模型来进行蒙汉翻译实验效果并不理想。我们在不对语料进行任何处理，也没有采用相关技术的情况下，得到的蒙汉 NMT 和 SMT 翻译结果并不理想。而本文将多种粒度融合方法及对齐技术运用于基本的 SMT

和NMT上，却取得了非常好的效果。

SMT重对齐方法，通过提高蒙汉双语对齐准确率来提升模型性能，其实验结果较基线SMT提升0.053 9个点。

本文使用的RNN是单层循环神经网络。RNN因其自身的网络结构，特别适合于序列的建模，尤其适合序列化的自然语言处理问题。从表2中看到，RNN的基线系统与目前较为成熟的SMT基线系统基本持平。而加入多粒度融合和对齐指导技术，RNN网络已经完全超过了传统的SMT。

机器翻译不是简单的序列处理。句子中的词和短语包含大量重要的局部信息，且对于蒙汉两种句法差异较大的语言，局部信息显得更加重要。但是RNN缺乏对局部信息的把握，相比CNN的优势就在于学习局部信息，且CNN可以在有限的实验条件下采用多层网络，获取全局信息。因此，从实验中也取得了更好的效果。

CNN和RNN中，均使用了字词和短语融合方法，且都使得系统有较为显著的提升。

如表3所示，RNN中单独加入Giza++对齐指导方法，提升相对较小。加入多粒度融合的方法，在缺乏局部信息学习的RNN网络中加入短语级的局部信息，因此，BLEU值均有显著提升，且Contrast c对比RNN基线系统BLEU5-SBP提升了0.092 3个点。

表3 RNN对比实验结果

系统	BLEU5				
	Test1	Test2	Test3	Test4	Test5
RNN baseline	0.6074	0.6109	0.6313	0.2331	0.6097
RNN+对齐指导	0.6879	0.6979	0.7097	0.2542	0.6635
RNN+多粒度融合	0.8208	0.8413	0.8305	0.3125	0.9219
RNN+多粒度融合+对齐指导	0.8508	0.8748	0.8688	0.3034	0.9282

在CNN中加入多粒度融合方法后提升0.122 8个点，提升非常明显。CNN本身就是通过抽象局部信息来得到全局信息，因此多粒度融合方法对于CNN来讲，主要是通过扩大训练集语料大小以及丰富目标词典中的语言现象来提升翻译效果。

本文还参考了CWMT2015最佳系统以及CWMT2017厦门大学和呼和浩特民族学院的测试结果，如表4所示。在完全相同的语料条件下，本文的CNN和融合系统均已赶超了CWMT2015最佳系统(best in 2015)；厦门大学的参赛系统(S12-primary-a)主要运用了蒙古文转拉丁以及多系统融合等方法，可以看到的CNN以及多系统融合系统效果要好于厦门大学的最好结果；呼和浩特民族学院(S8-primary-a)主要在SMT使用了与本文类似的方法对蒙古文词缀进行了处理，且在蒙古文端使用了BPE算法。可以看出，SMT中加入重对齐方法的效果与该系统基本持平，且本文其他三个系统明显好于该系统。

表4 多单位对比实验结果

系统	BLEU5-SBP	BLEU5
best in 2015	0.655 9	0.707 7
Contrast b(多粒度融合)	0.702 4	0.717 8
Primary a(句子级融合系统)	0.668 0	0.675 5
S12-primary-a	0.649 1	0.667 3
S8-primary-a	0.565 0	0.602 8

4 总结

本文构建了三个蒙汉翻译系统，以及一个融合系统，分别使用多种技术提升翻译效果。并在实验中对多种系统及技术进行了比较研究。

在蒙汉机器翻译任务中，最为严峻的两个问题就是资源较少，不足以训练出好的SMT或者NMT系统，以及蒙古文本身形态复杂，难以处理。二者也是困扰资源稀少翻译任务，或者说少数民族语言翻译任务的主要问题。在对蒙汉翻译进行研究的过程中，我们发现：一方面蒙古文的处理对其翻译质量有较大的影响，尤其是形态分析与处理，但是仅仅依靠形态分析难以进一步提升；另一方面在大语种任务上取得成果的一些技术方法如何因地制宜地用到蒙汉翻译上，如何开发适用于蒙汉，乃至整个稀少资源语言翻译任务特殊性的翻译技术方法，则更为重要。

本文在蒙汉翻译研究中，重视蒙古文的形态分析和处理，进行了大量实验来分析多种处理方法对不同翻译模型的影响。同时及时学习国内外机器翻译研究优秀成果，研发多种蒙汉翻译系统及其融合技术。此外，针对蒙汉翻译实际研究中遇到的困难和不足，提出多种具有创新性的技术方法，这些技术方法对蒙汉翻译有显著提升。

参考文献

[1] 侯宏旭，等. 基于统计语言模型的蒙古文词切分[J]. 模式识别与人工智能，2009，22(1)：108-112.

[2] 明玉. 基于词典、规则与统计的蒙古文词切分系统的研究[D]. 呼和浩特：内蒙古大学硕士学位论文，2011.

[3] 李金廷等. 语料预处理对蒙古文—汉文统计机器翻译的影响[J]. 计算机科学，2017，44(10)：259-264.

[4] Wu J，Hou H X，Xie C J. Realignment from Finer-grained Alignment to Coarser-grained Alignment to Enhance Mongolian-Chinese SMT[C]//Proceedings of 28th Pacfic，2015.

[5] Gehring J，et al. Convolutional Sequence to Sequence Learning [C]//Proceedings of the 34th International Conference on Machine Learning，Sydney，Australia，PMLR 70，2017.

[6] Gehring J，et al. A Convolutional Encoder Model for Neural Machine Translation [C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics，Vancouver，Canada，July 30-August 4，2017：123-135.

[7] bahdanau cho K H，Bengio Y. Neural machine translation by jointly learning to align and translate. [C]//Proceedings of ACL-IJCNLP，Volume 1：Long Papers，2015.

[8] Schuster M，Kuldip K Paliwal. Bidirectional recurrent neural networks[J]. Signal Processing IEEE Transaction，1997，45(11)：2673-2681.

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

[10] Marta R，Costa-jussà，José A R，Fonollosa. Character-based Neural Machine Translation[J]. arXiv preprint arXiv，2015：1511.04586.

[11] Luong M T，Manning C D. Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics，Berlin，Germany，August 7-12，2016：1054-1063.

[12] Mikolov T，et al. Distributed Repre-sentations of Words and Phrases and their Compositionality [J]. arXiv preprint arXiv，2013：1310.4546.

[13] Sang E F T K. Memory-Based Shallow Parsing[J]. Journal of Machine Learning Research，2002，2(4)：559-594.

[14] 钟伟金. 共现关键词—叙词同义关系自动识别研究——基于互信息法、概率法的对比分析[J]. 图书情报工作，2012，56(18)：122-126.

[15] Koehn P，et al. Moses：Open source toolkit for statistical machine translation[C]//Proceedings of the 45th Annual Meeting of the All on Interactive Poster and Demonstration Sessions. 2007：177-180.

[16] Koehn P，Och F J，Marcu D，Statistical phrase-based translation[C]//Proceeding of NAACL，2003.

[17] Och F J. Minimum error rate training in statistical machine translation[C]//Proceedings of the 41st Annual Meeting on Association for Computational Linguistics，2003，32(17)：160-167.

任众(1994—)，硕士研究生，主要研究领域为自然语言处理。

E-mail：18586037896@163.com

武静(1989—)，博士研究生，主要研究领域为自然语言处理。

E-mail：wujingyaya@163.com

侯宏旭(1972—)，通信作者，博士，教授，主要研究领域为自然语言处理、信息检索。

E-mail：cshhx@imu.edu.cn