

北京航空航天大学 CWMT2015 评测技术报告

巢文涵, 徐庆星

(1. 北京航空航天大学, 北京市海淀区 100191)

摘要: 本文介绍了北京航空航天大学自然语言处理实验室 (BUAA-NLP) 参加 CWMT-2015 评测情况。本次测评, 本课题组一共参加了汉英新闻和英汉新闻两个机器翻译翻译评测任务。本文介绍了本组参加各个评测任务的系统框架、主要方法和评测结果。

关键词: 统计机器翻译; 预调序; CWMT2015

BUAA Evaluation Technical Report for CWMT'2015

Wenhan Cao, Qingxing Xu

(1. BeiHang University ,Beijing 100191 , China)

Abstract: This is an overview of BUAA evaluation technology report for the 11th China workshop on machine translation. BUAA participated in the English-to-Chinese and Chinese-to-English machine translation task in news domain. This paper describes the implement framework of our machine translation system. We also give the key techniques , methods and experimental results.

Keyword: statistical machine translation, pre-Reorder ,Moses

1 引言

本文对北京航空航天大学自然语言处理研究组 (BUAA-NLP) 参加 CWMT2015 机器翻译评测情况进行描述。本课题组一共参加了汉英新闻和英汉新闻两个机器翻译翻译评测任务。

2 参评系统描述

BUAA-NLP 在本次评测中利用开源工具 Moses 搭建基于短语统计机器翻译系统。

在参与的两个评测任务中, 本课题组在英汉新闻领域机器翻译评测任务和汉英新闻领域机器翻译评测任务中, 使用了基于短语的统计机器翻译。

在两个翻译评测任务中, 本课题组尝试使用预调序的方法提高翻译质量。

本章将对系统和方法做简单介绍。

2.1 基于短语的统计机器翻译

本次评测中, 本课题组使用开源工具 Moses 构建基于短语的统计机器翻译系统。该系统基于对数线性模型将句子得分描述

为若干个特征的线性组合, 如下公式(1)所示:

$$\hat{e} = \arg \max_e \frac{\exp(\sum_{m=1}^M \lambda_m h_m(e, f))}{\sum_{e'} \exp(\sum_{m=1}^M \lambda_m h_m(e', f))} \quad (2.1)$$

其中, e 表示目标语言句子, f 表示源语言句子, $h_m(e, f)$ 表示第 m 个特征函数,

λ_m 表示第 m 个特征的特征函数对应的权重。

Moses 采用的特征参数包括正反向短语翻译概率, 正反向词汇翻译概率 [Koehn et al, 2004], 调序模式惩罚, 语言模型和句子长度惩罚等等。

Moses 使用 GIZA++ [Och and Ney, 2005] 训练词对齐模型, 并抽取短语对。基于最小错误率训练 (MERT) [Och et al, 2003] 通过开发集对对数线性模型参数进行调整。

2.2 源语言预调序模型

我们提出用神经网络的方法学习更好的调序特征, 该模型是基于线性排序的预调序神经网络模型。神经网络将源语言句子 f

以及一对调序基本单元 (f_i, f_j) 作为输入, 输出对于这一对调序基本单元的调序为 π_i, π_j 的分数。神经网络 S_{nm} 由两部分构成: 一部分 S_{local} 从两个调序基本单元的窗口中抽取局部分特征; 另一部分 S_{global} 从整个序列抽取整体特征。

为了学习更好的调序模型, 我们打算提出用神经网络的方法学习更好的调序特征表示。在本节中, 我们描述基于线性排序的预调序神经网络。

对于线性排序, 我们用一个神经网络替换其中的基于稀疏特征的线性评价函数

$$B: B[\pi_i, \pi_j] = S_{nm}(f, i, j, \pi_i, \pi_j) \quad (2.2)$$

此神经网络将源语言句子 f 以及一对调序基本单元 (f_i, f_j) 作为输入, 输出对于这一对调序基本单元的调序为 π_i, π_j 的分数。神经网络 S_{nm} 由两部分构成: 一部分 S_{local} 从两个调序基本单元的窗口中抽取局部分特征; 另一部分 S_{global} 从整个序列抽取整体特征。

局部特征: 给定一个调序单元 f_i 我们首先以它为中心的上下文 $f_{i-k} \dots f_i \dots f_{i+k}$ 窗口抽取局部特征。这些特征包括词汇特征、词汇的 n 元特征和语法信息等。

全局特征: 除了局部特征, 我们还希望对于需要排序的序列整体给出一个描述。对于两个调序单元 (f_i, f_j) , 我们分别从它们的左边 $(f_1 \dots f_{i-1})$, 中间 $(f_i \dots f_{j-1})$ 以及右边

$(f_{j+1} \dots f_{|f|})$ 抽取整体特征。对于每个部分, 我们分别使用它们包含的词汇的词袋 (bag-of-words) 作为特征。除此之外, 我们还额外使用二值特征表示每个部分是否含有标点等对调序指示作用很强的词。

2.3 参数调优

本文的工作是基于 Moses 的训练短语统计机器翻译系统。该模型基于线性对数模型, 除了词汇化的调序模型特征外还有一般短语翻译系统的主要特征, 比如正反向短语, 词汇翻译概率, 3-gram 语言模型, 词汇和短语惩罚。本工作的基准系统是: 一个是未经预调序的翻译系统, 它用于验证使用预调序技术对翻译系统的影响。

首先对源语言和目标语言进行词对齐, 得分词对齐结果完成对源语言调序 (即调序目标)。对于每一个源语言单词, 定义一个整数为词对齐中最左目标语言单词位置, 没有对齐定义为 0, 然后基于该数字对源语言句子进行稳定排序, 得到训练集合。

得到训练样本 (f, π^*) 后, 我们需要依据训练集进行参数调优。所以, 我们定义优化的目标是最小化以下排序损失。

$$l(\theta) = \sum_f \max(0, 1 + \text{score}_\theta(f, \pi^-) - \text{score}_\theta(f, \pi^*)) \quad (2.3)$$

我们训练时采用基于 CYK 的所有训练集中求出该模型反向转录文法下的最优方案 π_p , 并基于这个方案最优化参数:

$$l(\theta) = \sum_f \max(0, 1 + \text{score}_\theta(f, \pi_p) - \text{score}_\theta(f, \pi^*)) \quad (2.4)$$

计算出该函数损失后, 基于随机梯度下降方法优化参数。

3 实验

3.1 系统硬件配置

在本次评测中使用的计算机配置与操

作系统如表 1 所示

CPU	内存	操作系统
Inter(R) Core i7-3770	24G	Ubuntu 14.04

3.2 数据使用

本次使用数据均为评测方所提供的训练数据。汉英新闻领域机器翻译中，语言模型训练数据为全部英文部分语料得到的三元语言模型；英汉新闻领域机器翻译中，语言模型是利用全部训练数据汉语部分合并部分新闻爬取新闻语料训练的三元语言模型。预处理后的训练数据如表 2 所示。

表 2. 评测系统使用数据情况

评测项目	训练集	开发集	语言模型
英 汉 新闻	6,021,778 句对	1000 句对, 是个参考译文	14,003, 232 句对
汉 英 新闻	6,021,778 句对	1006 句, 四个参考译文	6,201,778 句对

平行训练预料针对中文的预处理使用斯坦福分词^[10]，英文预处理 PorterStemmer。语言模型在分词后基于 SRILM 训练的 3-gram 语言模型；词对齐工具使用 GIZA++[Och and Ney, 2005]。

3.3 实验结果分析

3.3.1 英汉新闻领域机器翻译评测结果与分词

预调序前后翻译结果对比如表 3 所示：

表 3 英汉新闻翻译对比试验

翻译系统	测试集	BLEU-4
Baseline	2013-ec-news	0.3112
NN 预调序	2013-ec-news	0.3134

3.3.2 汉英新闻领域机器翻译评测结果与分词

预调序前后翻译结果对比如表 4 所示：

表 4 汉英新闻翻译对比试验

翻译系统	测试集	BLEU-4
Baseline	2013-ce-news	0.1884
NN 预调序	2013-ce-news	0.1924

从实验结果可以看出，预调序后翻译质量相对于 baseline 有一定的提升。说明该方法在解决中英翻译中的长距离调序问题有一定的作用。下一步可以分析更好的调序特征加大训练调序模型的预料，预期结果会有更大的提升。

4 总结

本文主要介绍了北京航空航天大学计算机学院自然语言处理实验室 BUAA-NLP 小组参加全国机器翻译翻译研讨会的测评情况。本小组基于开源统计机器翻译系统 Moses 训练短语的统计机器翻译，加入预调序模块。实验表明，本研究组的系统性能在汉英新闻领域以及英汉英汉新闻领域相对于基线系统，有一定的提升。

5 致谢

本论文工作由“国家科技资源共享服务工程技术研究中心开放课题”资助。

参考文献

- [1] Och F J, Ney H. A systematic comparison of various statistical alignment models[J]. Computational linguistics, 2003, 29(1): 19-51.
- [2] 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.
- [3] 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.
- [4] Koehn P. Pharaoh: a beam search decoder for phrasebased statistical machine translation models[M] 2004: 115-124.

[5] 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.

[6] Tromble R, Eisner J. Learning linear ordering problems for better translation. Proceedings of Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2. Association for Computational Linguistics, 2009. 1007-1016.

[7] Li P,Liu Y, Sun M. Recursive Autoencoders for ITG-based Translation. 2013.

[8] Visweswariah K, Rajkumar R, Gandhe A, et al. A word reordering model for improved machine translation. Proceedings of Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011. 486-496.

[9] Sudoh K,Wu X,Duh K,et al. Post-ordering in statistical machine translation. Proceedings of Proc. MT Summit, 2011.

[10] Pi-Chuan Chang, Michel Galley and Chris Manning. 2008. Optimizing Chinese Word Segmentation for Machine Translation Performance. In WMT