

基于融合策略的短语表剪枝模型

薛征山 张大鲲 张倩 郝杰

东芝(中国)研究开发中心 北京 100738

E-mail: {xuezhengshan, zhangdakun, zhangqian, haojie}@toshiba.com.cn

摘要: 本文提出了一种基于融合策略的短语表剪枝模型, 将基于单一准则的相关性剪枝, 重要性剪枝及熵剪枝得到的翻译模型进行组合, 采用投票方式选择最终的短语对。这种方法能够避免使用单一准则进行模型剪枝时由于剪枝模型的缺陷导致的有效短语去除。实验结果表明, 融合剪枝方法在仅使用 3.8% 的翻译模型时, 翻译质量能够达到甚至超过基准系统。和单一准则的剪枝方法相比, 这种方法在翻译质量保持不变的前提下, 能够进一步缩减短语表规模。

关键词: 统计机器翻译; 短语模型; 相关性剪枝; 重要性剪枝; 熵剪枝; 融合剪枝

A phrase table pruning technique based on fusion strategy

Zhengshan Xue, Dakun Zhang, Qian Zhang and Jie Hao

Toshiba (China) R&D Center, Beijing, 100738

E-mail: {xuezhengshan, zhangdakun, zhangqian, haojie}@toshiba.com.cn

Abstract: *This paper proposes a mixed phrase table pruning method based on fusion strategy, which combines relevance pruning method, significance pruning method and entropy-based pruning method. The proposed method prunes phrase table based on phrase hit counts, which could reduce the risk of useful phrases being removed from phrase table by using any single pruning criterion. The experimental results show that our method can achieve as high as or even better result than baseline system, while keeping only 3.8% of the phrase table. Compared with single pruning method, this method can achieve the quality in the same level with fewer phrase table usage.*

KeyWords: *Statistical Machine Translation; Phrase-based Model; Relevance Pruning; Significance Pruning; Entropy-based Pruning; Mixed Pruning*

1 引言

经过近年来的研究与发展, 目前基于统计的机器翻译方法[Koehn et al., 2003; Chiang et al., 2007; Knight et al., 2001; Liu et al., 2006; 刘群, 2003; Xiao et al., 2012; 宗成庆等, 2012]逐渐占据了主流。在统计机器翻译系统中, 不管是基于短语的模型, 还是基于句法的模型, 都面临一个很重要的问题, 即训练后的翻译模型规模十分庞大。庞大的翻译模型不仅提高了运行成本(需要大量的存储空间, 甚至是分布式的集群, 以及大量的内存来加载翻译模型), 而且影响翻译的解码时间, 甚至是影响翻译质量和用户体验。此外, 庞大的翻译模型, 在一定程度上制约了机器翻译系统的应用场景, 比如在有限内存和存储空间的手持设备上的翻译引擎。另一方面, 如果我们大略地浏览一下翻译模型, 可以发现, 在翻译模型中有很大一部分短语对或者规则对¹并不含有实际的意义, 是无用的。这说明在一般情况下翻译模型中的短语对或者规则对服从长尾定理[刘凯等, 2013], 删除那些无用的短语对, 并不会对翻译系统的质量有很大影响。所以, 如何在不明显降低翻译质量的前提下, 从翻译模型中

¹这里我们称基于句法的模型(包括层次型短语模型)中使用的非词汇化短语对为规则对。

剔除无用短语对或者规则对，从而有效缩减翻译模型的规模（翻译模型剪枝），成为比较重要的研究课题之一。

目前，主要的剪枝技术可以分为两类，一类是在翻译解码时，为了提高解码效率，对候选短语进行过滤，包括：（1）阈值剪枝（Threshold Pruning）[Zens et al., 2012]：给定一源语言短语 f ， $\max\{p(\tilde{e}|f)\}$ 是 f 对应的所有目标短语的最大概率。阈值剪枝就是当某个短语对的翻译概率和 $\max\{p(\tilde{e}|f)\}$ 的比值小于某个比率或者某个固定值，就将该短语剪掉。（2）直方图剪枝或者柱剪枝（Histogram Pruning）[Zens et al., 2012]：在解码过程中，对于每个源语言短语对或者候选翻译项，仅保留概率最大的前 K 个，作为短语翻译可能的候选。

另一类技术是在翻译解码前，在生成短语表之后对生成的短语表进行裁剪，包括（1）基于短语对频率的剪枝（Count-based Pruning）[Zens et al., 2012]：将短语对频率低于某个阈值的短语对过滤掉，短语对频率可以在短语抽取过程中统计得到，这种方法能够去除出现频率较低的不可靠短语。（2）基于正向翻译概率的剪枝（Probability-based Pruning）[Zens et al., 2012]：将正向翻译概率低于某个阈值的短语对过滤掉。（3）相关性剪枝（Relevance Pruning）[ECK et al., 2007]：Eck 等通过统计解码过程中最佳翻译（n-best）结果中使用的短语对的频数得分，然后计算每个短语对的得分。按照高低排序，保留前 K 个短语对作为最终的翻译模型。（4）重要性剪枝（Significance Pruning）[Johnson et al., 2007]：Johnson 等人利用显著性检验去除一些不可靠的短语对，达到缩减短语表的目的。这种剪枝方法考虑的是短语对的质量，效果比较显著。（5）基于熵的剪枝（Entropy-based Pruning）[Zens et al., 2012; Ling et al., 2012]：Zens, Ling 等针对短语表中的所有短语对，判断较长的短语对是否可以用较小单位的短语对来进行组合，并且概率损失很小。这种方法与其它方法的不同之处在于，它判断的是短语对的冗余，剪枝效果也比较显著。（6）基于联合方法的剪枝[刘凯等, 2013]：刘凯等利用翻译日志并联合朴素贝叶斯、模型参数、规则频度等剪枝方法，在保留原有模型仅 1%-3% 的翻译规则的前提下，翻译质量没有显著下降。

本文主要关注的是第二类剪枝技术。在第二类剪枝技术中，各种剪枝方法都可以有效缩减翻译模型规模，但是方法（1）-（5）都是基于单一准则进行的模型剪枝，可能将许多不符合该准则，但实际上却非常有益的短语对给剪掉。刘凯等提出的方法（6），虽然使用了多种剪枝方法，但是它仅仅是将其它剪枝方法作为辅助补充，而且这些剪枝方法之间的组合类似于并集（将符合朴素贝叶斯，模型参数，规则频度的短语对或者规则对，直接并入经过翻译日志剪枝后的模型中），在本质上还是基于单一标准的剪枝。

针对以上问题，本文提出了一种融合剪枝技术，将剪枝效果比较明显的相关性剪枝（3）、重要性剪枝（4）和熵剪枝（5）技术进行融合。该技术将每种剪枝技术生成的翻译短语表进行比较，采用投票的策略，在仅使用原来模型约 3.8% 左右的短语表前提下，翻译质量不仅没有下降，反而得到提升。这种基于多准则的模型融合剪枝，对每个短语对的判断将不再依赖单一的准则，从而降低在剪枝过程中去除有效短语的风险。具体来说，我们将短语表剪枝问题看成是一个分类投票问题，相关性剪枝、重要性剪枝和熵剪枝相当于三个不同的分类器（三种不同的判断准则）。对于翻译模型中的每个短语对，只有获得半数以上投票时，我们才认为这个短语对确实是有用的，低于半数投票的短语对，可以直接从翻译模型中去除。在大规模数据上的实验表明，这种方法能够有效的去除短语表中的无效短语对，同时并不降低系统的翻译质量。

后续章节组织结构如下：第 2 节介绍相关的剪枝技术，详细介绍本文采用的相关性剪枝，重要性剪枝和熵剪枝技术，第 3 节介绍本文提出的融合策略剪枝方法，包括融合策略的思想来源及选择有效短语的策略，第 4 节给出实验结果及分析，最后一节结束语。

2 剪枝技术介绍

2.1 相关性剪枝 (Relevance Pruning)

相关性剪枝[Eck et al., 2007]的依据是短语对在翻译系统实际进行翻译时被使用的频度。首先利用训练语料训练翻译系统，再对训练语料进行翻译解码，然后根据如下公式 (1) 或者公式 (2)：

$$\text{Score (pp)} = \sum_{i=1}^n \frac{\text{count(pp in } i\text{-best)}}{i} \quad (1)$$

$$\text{Score (pp)} = \sum_{i=1}^n \frac{\text{count(pp in } i\text{-best)}}{i^2} \quad (2)$$

计算最佳翻译 (n-best) 结果中的短语对的得分，按照得分从高到底的顺序保留前 K 个短语对作为最终的剪枝后翻译模型。相关性剪枝优点是考虑了短语对或者规则对在实际翻译时的使用频度，使用频率越高的短语越倾向于有用。缺点是由于解码系统的偏向性，通常较长短语会出现在最佳翻译结果中，因此较长的短语更容易被保留。本文中使用的相关性剪枝方法与[Eck et al., 2007]略有不同。Eck 取前 K 个得分最高的短语对作为最终的翻译模型，本文因为采用多种剪枝技术，所以保留出现在最好翻译 (n-best) 中的所有短语对，作为后续融合短语表剪枝的候选。

2.2 重要性剪枝 (Significance Pruning)

重要性剪枝判断的是短语对内部源短语和目标短语的相互独立性[Johnson et al., 2007; Ling et al., 2012]。其依据是，对于固定的数据集 (训练集) 而言，并不是所有的短语对都被数据集同等的支持，那些获得强支持的短语对比弱支持的短语对更加重要，所以，去除被弱支持的短语对，对翻译模型的翻译效果影响较小。这里定义了 4 个重要性的相关变量，分别是：

(1): $C(\tilde{f})$: 源短语 \tilde{f} 的数量。

(2): $C(\tilde{e})$: 目标短语 \tilde{e} 的数量。

(3): $C(\tilde{f}, \tilde{e})$: 源短语 \tilde{f} 和目标短语 \tilde{e} 共现的数量。

(4): N : 双语短语对的数量。

判断短语对重要性的依据是 p-value 值，定义如下：

$$p_h(C(\tilde{f}, \tilde{e})) = \frac{\binom{C(\tilde{f})}{C(\tilde{f}, \tilde{e})} \binom{N-C(\tilde{f})}{C(\tilde{e})-C(\tilde{f}, \tilde{e})}}{\binom{N}{C(\tilde{e})}} \quad (3)$$

$$p\text{-value}(C(\tilde{f}, \tilde{e})) = \sum_{k=C(\tilde{f}, \tilde{e})}^{\infty} p_h(k) \quad (4)$$

设定阈值 α ，当短语对 p-value 值小于 α 时，说明短语对协同出现的可能性大，也就是意味着该短语对应该被保留。同理，当 p-value 大于 α 时，说明短语对协同出现的可能性小，说明该短语对应该被剪除。重要性剪枝利用短语对内部的信息估计该短语的可靠性。

2.3 熵剪枝 (Entropy-based Pruning)

基于熵剪枝方法[Zens et al., 2012; Ling et al., 2012]判断的是短语对之间的相互独立性。根据熵的值变化判断短语对是否冗余,其准则是翻译模型剪枝前后的相对熵不发生明显的变化,则该短语可以被去除。定义如下:

$$D(P_p \parallel P) = -\sum_{\tilde{f}, \tilde{e}} P(\tilde{f}, \tilde{e}) \log \frac{P_p(\tilde{e}|\tilde{f})}{P(\tilde{e}|\tilde{f})} \quad (5)$$

公式(5)中, $P_p(\tilde{e}|\tilde{f})$ 表示剪枝后的模型, $P(\tilde{e}|\tilde{f})$ 表示未剪枝的模型。设定阈值 δ , 当 $D(P_p \parallel P) > \delta$ 时,表示剪枝前后,翻译模型的相对熵差异较大,所以该短语对不可剪掉。同理,当 $D(P_p \parallel P) < \delta$ 时,表示剪枝前后,翻译模型的相对熵差异很少,则认为是可以安全的将该短语对去除。

熵剪枝方法也可以理解为,当某个长短语对可能由多个较小单位的短语对组合而成时,组合概率如果接近于长短语对的概率时,就可以将长短语删除了。因为在这种情况下,删除了长短语,同样的源语言语句,用较小单位的短语对同样可以得到相同的翻译结果,并且损失很小的概率。

3. 基于融合策略的短语表剪枝技术

3.1 思想来源

在利用支持向量机 SVM (Support Vector Machine) 做多分类时,有一种方法是将多分类问题分解为一对一分类,在每两类之间训练一个分类器,因此对于 k 类问题,需要训练 $k(k-1)/2$ 个分类器。当对一个未知样本进行分类时,每个分类器都对其类别进行判断,并为相应的类别投上一票,最后得票最多的类别即作为该未知样本的最终类别。这种分类方法的一种特殊情况可能出现在投票决策阶段,即存在多个类别的票数相同情况,从而影响对未知样本的类别判定,影响分类精度。

和以上思想类似,如果我们将翻译模型剪枝问题看作是一个分类问题,那么可以看作是一个二分类的问题。仅仅依靠单一的剪枝技术可能得不到理想的结果,原因是单一的剪枝技术通常只考虑了一方面问题,要么是短语对的频度,要么是短语对的质量,或者短语对的冗余等,不能够兼顾多个因素。所以本文考虑训练多个分类器,每个分类器的分类标准不同,各有侧重,这些不同的分类器联合成一个强的分类器。实际上,本文的方法采用三种分类器(三种不同的单一剪枝方法,分别考虑短语对的频度,质量和冗余),采用类似 SVM 的投票策略来决定分类结果。同时,因为是三分类器二分类问题,所以不存在 SVM 的特殊情况。

3.2 融合剪枝技术

本文选择了相关性剪枝、重要性剪枝和熵剪枝作为判断准则,针对每个短语对或者规则对,如果能得到这三种判断准则半数以上(2/3)的投票支持,我们就认为该短语对是有用的,需要保留,否则直接丢弃。之所以选择这三种方法作为基本的剪枝方法,其理由是:首先单一剪枝技术的剪枝效果明显,这三种剪枝方法都能够有效的对短语表进行剪枝[Zens et al., 2012; Eck et al., 2007; Johnson et al., 2007; Ling et al., 2012]。其次,所选择的这三种剪枝技术的准则各自不同,对短语表进行不同角度的判断。相关性剪枝使用的判断准则是短语对在实际翻译过程中是否被使用,该准则判断的是短语对的频度。重要性剪枝的判断准则是短语对内部源短语与目标短语的相互独立性,该准则判断的是短语对的质量。而熵剪枝的判断准则是短语对之间的相互独立性,该准则判断的是短语对的冗余。

假设初始翻译模型为 M ，相关性剪枝模型¹为 M_1 ，重要性剪枝模型为 M_2 ，熵剪枝模型为 M_3 ，显然 $M_1 \subset M$ ， $M_2 \subset M$ ， $M_3 \subset M$ 。且根据 M_1 、 M_2 、 M_3 的物理意义易知， $M_1 \cap M_2 \cap M_3$ 为常用的高频短语对，不为空。

对于任意的短语对 Pr ， $Pr \in M$ ， $i \in (1, 2, 3)$ ，定义函数：

$$f(Pr, M_i) = \begin{cases} 1, & \text{if } Pr \in M_i \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

公式(6)表示的是任意的短语对和单一剪枝模型的关系。融合策略剪枝模型定义如下：

$$c(Pr) = \frac{\sum_{i=1}^3 f(Pr, M_i)}{3} \quad (7)$$

公式(7)表示的是任意的短语对和融合剪枝模型的关系，它的取值有4种，分别是0，1/3，2/3和1。在本文中，当任意短语对的 c 值大于1/2时，我们就将该短语对保留，否则直接丢弃。

3.3 技术实现

本文采用了三种基本的剪枝技术。在相关性剪枝中，我们使用翻译模型对训练语料进行翻译，得到每个训练句子的多个最好(n-best)翻译结果，在n-best结果中，保留相应的位置对齐信息，根据对齐信息，可以得到短语对的对应关系，从而获得在翻译实际语料中使用的短语列表。根据这一短语列表对原始短语表进行过滤，得到最终的基于相关性剪枝的模型 M_1 。

对于重要性剪枝和熵剪枝，我们使用Moses[Kohlen et al., 2007]提供的相应的工具包。基于熵的剪枝需要计算短语对之间的相互独立性，剪枝过程计算量较大，我们改进了实际的计算过程，利用多线程技术，提高了计算速度。在得到每个短语对的重要性得分和熵得分之后，根据此得分，可以截取任意比例的短语对作为最终的剪枝翻译模型 M_2 和 M_3 。

在融合剪枝时，为了实现公式(7)，本文采用了等价的方法，即将相关性剪枝模型，重要性剪枝模型和熵剪枝模型，两两取交集，然后对结果取并集并去重。这样做的理由是，当某一个短语对被任意两种剪枝技术所支持时，不管它是否被第三种剪枝技术支持，它都符合我们的融合剪枝准则，应该被保留。同时，同一短语对可能被不同种剪枝技术支持，所以最后需要去重，相同的短语对仅需要保留一份。

4. 实验

本小节，我们将进行融合剪枝技术的有效性验证。实验中使用开源的Moses短语模型系统作为实验翻译解码器，短语抽取长度设定为7，语言模型使用5元语言模型。模型训练数据使用WMT06的法语-英语(Fr-En)语料。语料基本信息如表1所示：

表1 语料基本信息

Corpus type	# of Sentences
Training corpus	688,031
En corpus for LM	1,003,349
Dev corpus	2,000

测试语料使用WMT06-WMT13的测试集，其中去除了重复的WMT07测试集，共7个测

¹ 利用相关性剪枝方法得到的短语翻译模型，简称相关性剪枝模型，以下同。

试集。对测试集进行评价的参考译文为1个参考译文。各数据集的信息如表2所示：

表2 测试语料信息

Test set	# of Sentences
WMT06	3,064
WMT08	2,066
WMT09	3,027
WMT10	2,489
WMT11	3,003
WMT12	3,003
WMT13	3,000

本文设计了5组实验。(1) 相关性剪枝实验，即利用训练语料训练得到的翻译模型，再对训练语料进行翻译，得到1-best或20-best的翻译结果，根据多翻译结果中的短语对齐信息对原始短语表进行过滤，得到最终的翻译模型。(2) 重要性剪枝实验，首先计算每个短语对的重要性得分，然后取得分最高的10%、20%、...、90%的短语对得到剪枝后的模型。(3) 熵剪枝实验，保留熵得分最高的10%、20%、...、90%的短语对得到剪枝模型。(4) 融合剪枝实验一。将相关性剪枝实验中的1个最好结果(1-best)分别和比例对应的重要性剪枝和熵剪枝得到的短语模型进行融合，测试融合剪枝模型。(5) 融合剪枝实验二。将相关性剪枝实验中的20个最好结果(20-best)分别和比例对应的重要性剪枝和熵剪枝得到的短语模型进行融合，测试融合剪枝模型。

在本节实验中，一些实验方法的命名(如Sig10, Rel20等)原则如下：编号后面的数字表示百分比，这个百分比是针对重要性和相对熵的得分而言，并不是指短语表的规模。比如Sig10，意思是先将重要性得分按照倒序排列，找到第10%大的那个得分作为阈值，然后用这个阈值对原始短语表进行过滤，保留得分大于等于该阈值的短语对，去除得分低于该阈值的短语对。此外，实验结果中的percentage，表示剪枝后短语表文件大小与原始短语表文件大小的比值。表格中的average为在WMT 7个不同的测试集上得到的BLEU值的平均。在实验(2)重要性剪枝实验中，存在短语对得分相同的情况，因此Sig20、Sig30和Sig40得到的剪枝模型相同。各实验结果数据表和图如下所示：

表3 相关性剪枝实验结果

method	percentage	average
baseline	100.0%	19.38
1-best	5.8%	19.21
20-best	14.1%	19.36

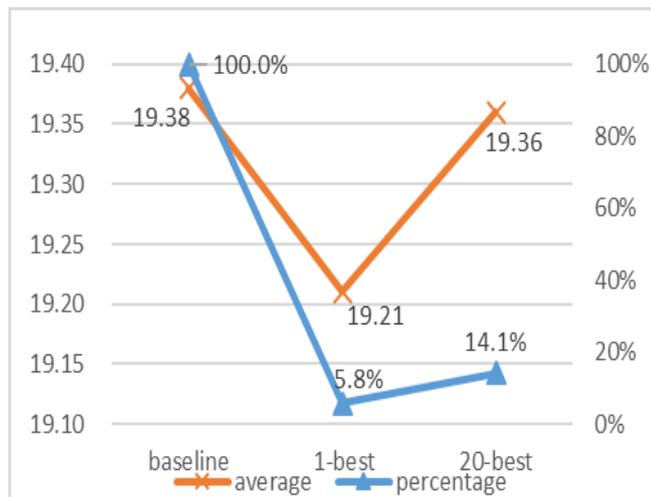


图1 相关性剪枝折线图

表4 重要性剪枝实验结果

method	percentage	average
baseline	100.0%	19.38
Sig10	9.1%	19.47
Sig20	45.3%	19.44
Sig30	45.3%	19.44
Sig40	45.3%	19.44
Sig50	54.4%	19.41
Sig60	62.6%	19.38
Sig70	72.1%	19.36
Sig80	81.5%	19.36
Sig90	90.5%	19.35

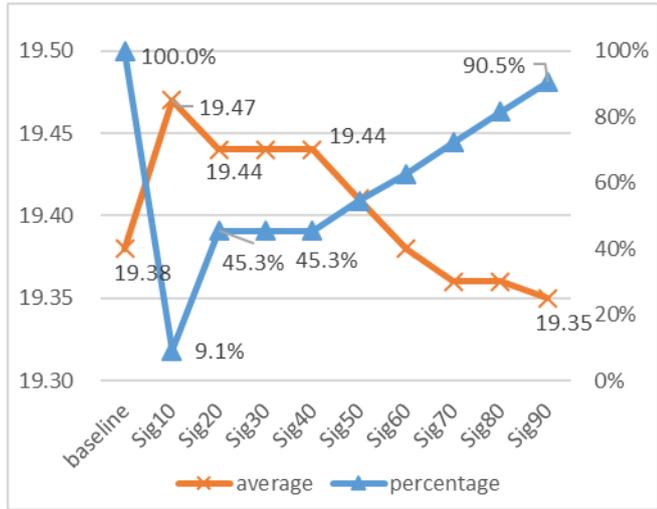


图2 重要性剪枝折线图

表5 熵剪枝实验结果

method	percentage	average
baseline	100.0%	19.38
Rel10	8.3%	19.05
Rel20	17.3%	19.23
Rel30	26.9%	19.54
Rel40	36.7%	19.57
Rel50	46.7%	19.50
Rel60	56.9%	19.47
Rel70	67.2%	19.42
Rel80	77.8%	19.42
Rel90	98.9%	19.38

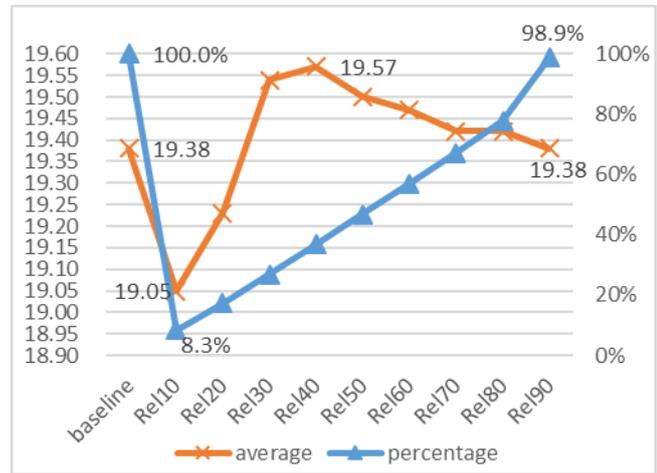


图3 熵剪枝折线图

表6 融合剪枝一实验结果

method	percentage	average
baseline	100.0%	19.38
comb10	3.8%	19.51
comb20	7.8%	19.49
comb30	11.1%	19.49
comb40	14.7%	19.49
comb50	22.7%	19.46
comb60	33.1%	19.43
comb70	47.1%	19.40
comb80	63.5%	19.40
comb90	90.6%	19.35

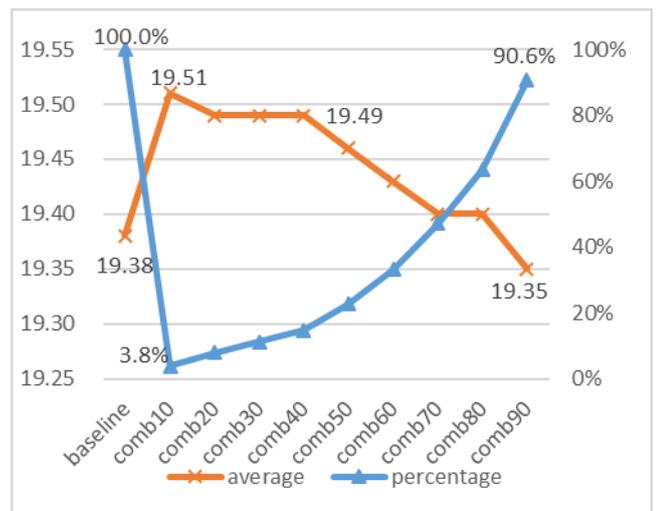


图4 融合剪枝一折线图

表7 融合剪枝实验二完整实验数据

method	Size (M)	Percentage (%)	WMT 06	WMT 08	WMT 09	WMT 10	WMT 11	WMT 12	WMT 13
baseline	2433	100.0	28.69	15.48	18.03	17.20	18.91	18.78	18.60
comb10	123	5.1	28.65	15.70	18.21	17.33	19.14	18.96	18.78
comb20	315	13.0	28.65	15.62	18.19	17.31	19.13	18.94	18.77
comb30	394	16.2	28.67	15.56	18.22	17.33	19.10	18.89	18.73
comb40	481	19.8	28.69	15.49	18.18	17.31	19.10	18.88	18.73
comb50	686	28.2	28.68	15.44	18.18	17.25	19.07	18.87	18.73
comb60	933	38.4	28.70	15.42	18.12	17.22	19.03	18.78	18.69
comb70	1251	51.4	28.68	15.50	18.11	17.24	18.92	18.78	18.59
comb80	1615	66.4	28.67	15.52	18.12	17.22	18.89	18.80	18.55
comb90	2206	90.7	28.66	15.50	18.03	17.17	18.85	18.76	18.57

表8 融合剪枝二实验结果

method	percentage	average
baseline	100.0%	19.38
comb10	5.1%	19.54
comb20	13.0%	19.52
comb30	16.2%	19.50
comb40	19.8%	19.48
comb50	28.2%	19.46
comb60	38.4%	19.42
comb70	51.4%	19.40
comb80	66.4%	19.40
comb90	90.7%	19.36

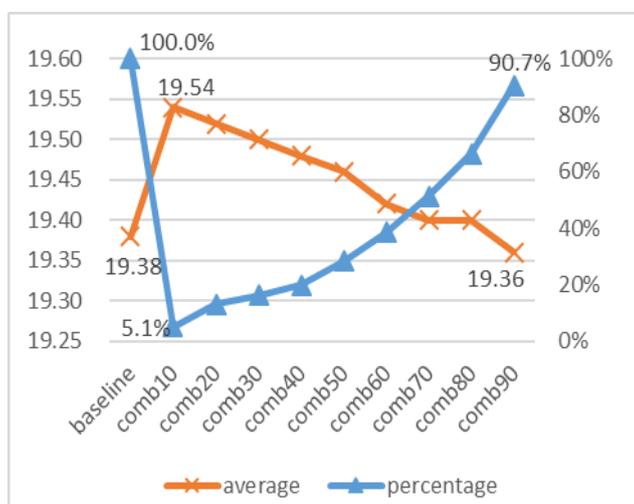


图5 融合剪枝二折线图

上述实验结果中可以看出，基线系统在WMT 7个测试集上的翻译得分（BLEU值）均值为19.38。除了相关性剪枝（表3）外，重要性剪枝（表4）和熵剪枝（表5）技术都使用较小的翻译模型，获得了超过基线系统的测试结果。相关性剪枝在使用原有翻译模型14.1%的情况下，仅比基线系统低0.02个BLEU值。重要性剪枝最好的效果是使用9.1%的翻译模型，翻译质量却上升到19.47（表4）。熵剪枝的实验结果呈现抛物线状，最好的剪枝结果是使用36.7%的翻译模型，获得了19.57的BLEU值（表5，图3）。

重要性剪枝和融合剪枝实验结果都呈现递减曲线（图2、图4、图5）。递减曲线说明，在使用一定比例（10%）的翻译模型时，已经达到了翻译质量的最优，随后增加的短语对并不能进一步提高翻译质量。从实验结果可以发现，融合剪枝效果比重要性剪枝效果好，分别使用3.8%和5.1%的翻译模型，翻译质量却上升到19.51和19.54（表4、表6、表8）。融合剪枝相比重要性剪枝的优势是：翻译模型规模更小，但是翻译质量更高。和熵剪枝方法相比，融合剪枝虽然在翻译质量上稍差，但是在翻译模型规模对比中，却占明显优势（表5、表6、表8）。例如，在仅比熵剪枝低0.06个BLEU值的情况下，翻译模型规模从36.7%缩小到3.8%，或者低0.03个BLEU值的情况下，翻译模型规模从36.7%缩小到5.1%。

上述实验表中的数据是WMT06-WMT13的平均结果，在实际针对单独测试集的测试实

验中实验结论不变，表7给出了融合剪枝实验二的完整数据。除了上述实验外，本文还进行了任意两种剪枝技术的交叉实验，实验结果是缩减后的翻译模型规模更小，但是翻译质量也急剧下降。究其原因可能是两种剪枝的投票策略太过苛刻（每个短语对都需要得2票才可以保留下来），导致会将有用的短语对裁剪掉。对于使用多于三种的单一剪枝技术来进行融合的方法，如果使用偶数种剪枝技术，会面临SVM多分类的问题，如果是奇数种单一剪枝技术的融合，从理论上来说，将可能得到比本文实验更好的结果，相关实验将在未来工作中进行。

5. 结束语

本文针对机器翻译领域比较热门的模型剪枝技术，提出了一种多剪枝技术融合的剪枝方法，基于三种剪枝效果较好的相关性剪枝，重要性剪枝以及熵剪枝，采用投票的策略进行融合，降低了使用单一剪枝方法对有效短语的去除风险。实验表明，在仅用3.8%的翻译模型的情况下，翻译质量能够达到甚至超过基准系统。和单一准则的剪枝方法相比，这种方法在翻译质量保持不变的前提下，能够进一步缩减短语表规模。

基于统计的机器翻译方法随着训练语料规模的不断增加，训练得到的翻译模型也将不断增大，相应的无用短语对比例也会逐渐上升。本文中提出的融合剪枝模型在应用到更大规模的翻译模型时，翻译模型裁剪的比例越高，因此能够进一步提高模型加载速度，减小模型存储空间，使得翻译系统在移动设备上部署更加容易。此外，本文中使用的三种基本剪枝方法仅使用了短语的基本信息，除了可以应用于基于短语的翻译模型外，也适用于基于句法的翻译模型（树到串模型等）。

参考文献:

- David Chiang. 2007. Hierarchical phrase-based translation. *Computational Linguistics*, June. Pages 201-208
- Matthias Eck, Stephan Vogel, and Alex Waibel. Estimating Phrase Pair Relevance for Machine Translation Pruning, *Proceedings of MT Summit XI*, Copenhagen, Denmark, September 2007.
- Howard Johnson, Joel Martin, George Foster, and Roland Kuhn. 2007. Improving translation quality by discarding most of the phrasetable. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, , Prague, Czech Republic, June. Pages 967-975.
- Philipp Koehn, Franz Joseph Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Human Language Technology Conf. / North American Chapter of the Assoc. for Computational Linguistics Annual Meeting (HLT-NAACL)*. Edmonton. Canada, May/June. Page 127-133
- Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst, Moses: Open Source Toolkit for Statistical Machine Translation, *Annual Meeting of the Association for Computational Linguistics (ACL)*, demonstration session, Prague, Czech Republic, June 2007.
- Yamada, K. and K. Knight. 2001. A syntax-based statistical translation model. In *Proceedings of ACL*.
- Wang Ling, Joao Graca, Isabel Trancoso, Alan Black. 2012. Entropy-based Pruning for

- Phrase-based Machine Translation. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island, Korea. Pages 962-971.
- Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-String Alignment Template for Statistical Machine Translation. In Proceedings of COLING/ACL 2006, Sydney, Australia, July. Pages 609-616,
- 刘群. 2003. 统计机器翻译综述[J]. 中文信息学报, 17(4).
- 刘凯, 吕雅娟, 姜文斌, 刘群. 2013. 基于翻译日志的统计机器翻译模型剪枝[J]. 北京大学学报 (自然科学版).
- Tong Xiao, Jingbo Zhu, Hao Zhang and Qiang Li. 2012. NiuTrans: An Open Source Toolkit for Phrase-based and Syntax-based Machine Translation. In Proc. of ACL 2012 System Demonstrations.
- 宗成庆, 张宵军. 2012. 统计机器翻译[M]. 电子工业出版社.
- Richard Zens and Daisy Stanton and Peng Xu. 2012. A Systematic Comparison of Phrase Table Pruning Techniques. Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Jeju Island, Korea. Pages 962-971.