

基于双向线性变换的可比语料双语词典抽取重排序方法

张春越¹, 赵铁军¹

(1.哈尔滨工业大学 计算机科学与技术学院 机器智能与翻译研究室, 黑龙江省哈尔滨市 邮编 150001)

摘要: 近几年来, 一种基于词嵌入的线性变换方法在可比语料双语词典抽取任务上取得了显著的效果。这种方法假定双语词嵌入空间在翻译时满足线性关系, 然而, 在实际中, 双语词嵌入空间的线性假设在单语词嵌入的训练中并不能得到很好的保证, 这使得在源语言到目标语言和目标语言到源语言两个方向上学习到的线性变换并不一致。为此, 对给定的待翻译词的翻译候选 N-best 列表, 本文综合了上述两个方向的线性变换得分, 提出了一种基于双向线性得分的重排序方法。在汉英方向的双语词典抽取任务上的实验表明, 相比单向线性基准方法, 这种重排序方法能够显著提升抽取到的字典的准确度, 在 Top-1 排名上能够提升 69%。

关键词: 双语词典抽取; 可比较语料; 线性变换; 重排序; 双向

Re-Ranking for Bilingual Lexicon Extraction with Bi-directional Linear Transformation from Comparable Corpora

Chunyue Zhang¹, Tiejun Zhao¹

(1. Machine Intelligence and Translation Lab, School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang, 15001, China)

Abstract: Recently a simple linear transformation with word embedding has been found to be highly effective to extract a bilingual lexicon from comparable corpora. However, the assumption that the pairs of bilingual word embedding for training this transformation satisfies a linear relationship automatically actually can't be guaranteed absolutely in practice. So the transformation of the source to the target is not consistent with the one of the target to the source. Given the translation candidate n-best list of a source word, we propose a bi-directional linear based re-ranking method by combining the two direction linear score. The experimental results confirm that the proposed solution can achieve a significant improvement of 69% in the precision at Top-1 over the unidirectional baseline approach on the English-to-Chinese bilingual lexicon extraction task.

Key words: Bilingual lexicon extraction; comparable corpora; linear transformation; re-ranking; bi-directional

1 引言

双语字典 (Bilingual Lexicons) 在许多跨语言处理任务中都扮演着十分重要的角色, 如机器翻译^{[1][19]}、跨语言信息检索^[2]等。一般地, 双语字典的获取可以通过以下两种方式: 其一是通过专家手工编制而成, 但这样做需要花费较大的人力和物力; 其二是使用平行语料 (Parallel Corpus) 通过统计词对齐方法获得^[3]。这种方式较第一种方式而言, 不需要人工代价, 但平行语料的获取同样代价较大, 尤其是在某些资源缺乏的语言对上或特定领域中, 平行语料是非常稀少的甚至可能根本就不存在。为此, 基于可比语料 (Comparable Corpus) 的双语字典自动抽取技术成为解决这一问题的有效途径^{[4][5][6]}。

可比语料主要是指在主题上相近或相关, 内容上有重叠 (overlapping) 的双语文本, 其中蕴藏着很多互译的词语、短语和文本等。和平行语料相比, 获取大量的双语可比语料是非常容易的。因此, 如何从可比语料中获取双语字典以及其他类型的翻译知识成为近年来一个热点问题。

近年来, 词的分布表示 (Distributed Representation)^[7], 也经常被叫做词嵌入 (Word Embedding), 被研究人员进行了广泛的研究。词嵌入把离散的词表示转变为一种连续的、低维度的、实值的向量表示, 使得语义上相似的词在词嵌入空间上分布位置较近。词嵌入有着很好的性质, 在自然语言处理任务上有了很多应用^{[8][20]}。Mikolov 等人^[9]发现词嵌入在双语场景上的线性关

系, 参见图 1, 可以看出在图 1 中, 互译的英语词与西班牙语词在上图与下图中的 2 维空间上有一定的线性关系。受此启发, Mikolov 等人假设双语词嵌入空间满足线性关系, 提出了一种基于线性变换的方法^[9], 将源语言空间词嵌入使用一个双语变换矩阵投影到目标语言词嵌入空间中。尽管这种方法相对简单, 但实验表明在双语词典抽取任务上十分有效。

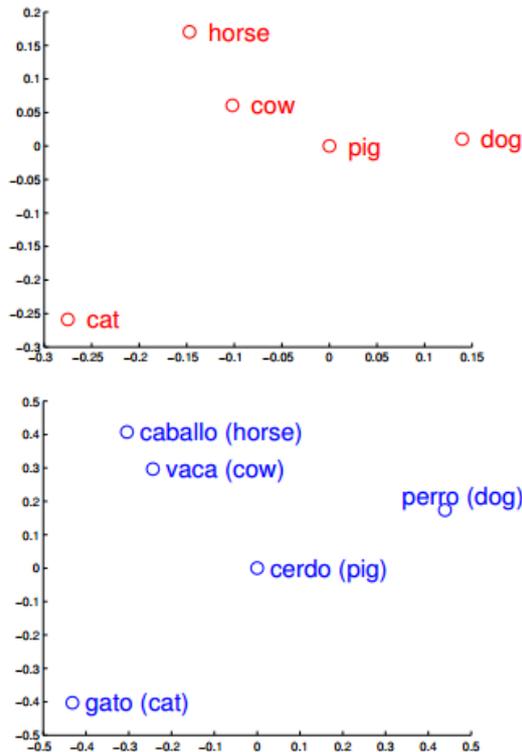


图 1 英语和西班牙语部分互译词的双语词嵌入通过 PCA 降维后的 2 维空间图

实际中, 在可比较语料上训练得到的互译的两个词的词嵌入是很难自动地满足线性关系的。原因可以解释如下:

1. 在 Mikolov 等人的工作中, 双语词嵌入是从 WMT11 平行语料上学习到的, 平行语料上的语义空间是相当对称的, 因此训练到双语词嵌入满足线性关系是有可能的, 而可比语料则是高度不对称的, 语义空间相差非常大;

2. 每种语言的词嵌入是独立地从某个语料上自动学习到的, 学习到的词嵌入受语料规模、语料领域的影响较大, 因此双语词嵌入的线性关系也没法得到显式的保证。

在双语词嵌入的线性关系没法得到充分保证的条件下, 在源语言到目标语言和从目标语言到源语言两个方向上学习到的线性变换并不一致, 这会使得对想要翻译的源语言词, 以及目标语言中该词的正确翻译组成的词对, 在两个方向的翻译候选集中的排序不一致。为此, 在本文中, 我们提出了一种基于双向线性变换得分的重排序方法, 这种方法综合考虑了这个词对在源语言到目标语言以及目标语言到源语言两个方向的线性变换后的相似度得分, 用它们的加权和作为最终的翻译得分, 从而选择两个方向都表现良好的翻译候选作为最终的翻译候选输出, 实验表明这种方法的性能能够显著超过基线系统的性能。

具体地, 在本文中, 我们:

1. 首先利用可比较语料独立地学习两个语言的词嵌入;
2. 根据预先已有的双语种子词典以及步骤 1 中学习到的词嵌入分别学习源语言到目标语言的线性变换和目标语言到源语言的线性变换;
3. 对给定待翻译的源语言词, 根据源语言到目标语言的线性变换获取其在目标语言的翻译候选集以及相应的翻译得分;
4. 对翻译候选集中的翻译候选, 使用目标语言到源语言的线性变换计算出这个方向的翻译得分, 接下来使用本文提出的双向线性得分模型进行重排序输出最终的排序结果。

2 词嵌入

Mikolov 等人^[10]使用了 Skip-gram 模型来学习词嵌入。Skip-gram 模型训练的目标是想学习到一种词的实值向量表示, 这种表示能够较好的预测同一个句子中当前词的上下文词。这个模型的时间复杂度非常低, 可以高效地在大规模单语数据上进行训练。在实践中, 当单语数据比较小的时候, Skip-gram 能给出很好的词嵌入表示。

形式上, 给定一个训练词序列 w_1, w_2, \dots, w_T , Skip-gram 模型的训练目标是最大化下述目标函数:

$$\frac{1}{T} \sum_{t=1}^T \sum_{-c \leq j \leq c, j \neq 0} \log P(w_{t+j} | w_t)$$

(1)

其中, $2c + 1$ 为训练中设定的窗口的大小, 里层的求和是计算给定当前的词 w_t , 正确地预测其上下文词 w_{t+j} 的对数概率, 外层的求和则是遍历训练数据中的所有词。

在 Skip-gram 模型中, 每个词 w 被关联到两个参数向量 u_w 和 v_w , 均为一个实值行向量。相应地, 它们分别代表了词 w 做为网络输入和网络输出的两种表示向量。最终, 给定词 w_i , 预测其上下文词 w_j 的概率定义如下:

$$P(w_j | w_i) = \frac{\exp(u_{w_j}^T v_{w_i})}{\sum_{j'=1}^W \exp(u_{w_{j'}}^T v_{w_i})} \quad (2)$$

其中, W 是词汇集的总词数, v_{w_i} 为词 w_i 的词嵌入。

3 双向线性重排序方法

3.1 基于单向线性变换的双语词典抽取

Mikolov 等人^[9]提出了在两个可比较(单语)语料上解决双语词典抽取的一种方法。这种方法假定双语词嵌入空间在翻译上满足线性关系, 因此将双语词典抽取看做一个多变量线性回归问题。给定记号 $\Sigma \in R^{n_1 \times d_1}$ 和 $\Omega \in R^{n_2 \times d_2}$, 分别代表源语言和目标语言的词汇集合以及对应的词嵌入空间, n_1 和 n_2 代表源语言和目标语言的词表规模大小, d_1 和 d_2 分别代表源语言和目标语言词嵌入空间的维度。在训练阶段, 利用一个预先已有的双语种子字典 $BV = \{(x, y) | x \in \Sigma', y \in \Omega', x \text{与} y \text{互译}\}$, $\Sigma' \subset \Sigma$ 和 $\Omega' \subset \Omega$ 分别为双语种子字典的源语言词汇集和目标语言端词汇集, 其中 $\Sigma' \in R^{n \times d_1}$, $\Omega' \in R^{n \times d_2}$, 来学习从源语言词嵌入空间到目标语言词嵌入空间的线性变换。具体地, 将此视为一个多变量最小二乘问题, 优化目标如下:

$$\hat{W} = \operatorname{argmax}_{W \in R^{d_2 \times d_1}} \sum_{i=1}^n \|Wx_i^T - y_i^T\|^2 \quad (3)$$

其中, $x_i \in \Sigma'$, $y_i \in \Omega'$ 。学习得到的线性变换 \hat{W} 称作双语变换矩阵。问题的解为

$$\hat{W} = (X^T X)^{-1} X^T Y \quad (4)$$

在预测阶段, 给定一个源语言待翻译的词嵌入 x , 首先利用训练阶段学习到的双语变换矩阵 \hat{W} 将 x 映射到目标语言词嵌入空间中, 即 $y^T = \hat{W}x^T$, 然后在目标词嵌入空

间 Ω 找到与 y 最近邻(本文使用余弦距离度量)的 N 个词作为翻译候选的 N best 输出, 即

$$\text{cand}_{N\text{best}}(x) = \{y | \operatorname{argmax}_{y \in \Omega} \text{score}(\hat{W}x^T, y^T)\} \quad (5)$$

其中 $\text{score}(a, b) = \text{cosine}(a, b)$ 。

3.2 基于双向线性得分加权和的重排序方法

正如在引言中介绍的, 上节中提到的线性方法假定学习到的双语词嵌入满足线性关系, 然而, 双语词嵌入空间的线性假设在单语词嵌入的训练中并不能得到很好的保证, 这使得学习到的线性变换在源语言到目标语言和从目标语言到源语言两个方向并不一致, 这会导致互译的两个词组成的词对在两个方向上的翻译得分和在翻译候选集上的排序有较大的不同。为此, 本文综合了上述两个方向的得分, 提出了一种基于双向线性得分加权和的重排序方法, 希望使得那些在两个方向都表现好的翻译候选的排名更高。

具体地, 我们首先使用双语种子字典学习两个方向的双语变换矩阵,

$$\hat{W}_{s2t} = \operatorname{argmax}_{W \in R^{d_2 \times d_1}} \sum_{i=1}^n \|Wx_i^T - y_i^T\|^2 \quad (6)$$

$$\hat{W}_{t2s} = \operatorname{argmax}_{W \in R^{d_1 \times d_2}} \sum_{i=1}^n \|Wy_i^T - x_i^T\|^2 \quad (7)$$

在预测阶段, 对给定一个待翻译的词的词嵌入 x , 我们首先使用 \hat{W}_{s2t} 获取 x 的翻译候选集 $\text{cand}_{N\text{best}}(x)$, 然后对此翻译候选集进行重排序计算, 计算方法如下:

$$\text{bi_score}(x, y) = \alpha \cdot \text{score}_{s2t}(\hat{W}_{s2t}x^T, y) + \beta \cdot$$

$$\text{score}_{t2s}(\hat{W}_{t2s}y^T, x) \quad (8)$$

其中 $y \in \text{cand}_{N\text{best}}(x)$, $0 < \alpha, \beta < 1, \alpha + \beta = 1$ 。

4 实验

4.1 实验数据和配置

本文在英汉方向上进行双语词典抽取的任务。对于可比较语料, 我们使用英语 Gigaword 语料 (LDC2009T13) 和汉语 Gigaword 语料 (LDC2009T27)。为了让语料有更好的可比较性, 我们只使用了这两个语料中的新华社发布的部分 (Xinhua News),

时间跨度为从1995年1月到2008年12月。

对于语料，我们进行了如下的预处理步骤：

1. 删除重复句子，将数字词用一个特殊的符号替换；

2. 对于英语语料，我们对文本进行了tokenize和词根还原；

3. 对于中文语料，我们使用了斯坦福中文分词工具¹进行了分词。

处理好的语料的细节见表1。我们仅考虑了词频不少于10词的词汇。

表1 英语和汉语的语料规模

	英语	汉语
训练词数 (Tokens)	326M	346M
词汇表集 (Words)	136K	205K

表2 训练集、开发集、测试集的统计信息

	训练集	开发集	测试集
总条数	64692	2062	10576
词数	14413	500	2500
平均值	4.48	4.12	4.23

对于每个语言，我们使用同样的配置来训练skip-model来学习词嵌入，我们使用word2vec²工具包。我们仅考虑的词频不少于10的词，窗口大小为7，迭代次数为5次，负采样样本数为5，sample阈值设为 10^{-4} ，其他参数为工具包默认配置。

为了得到用于训练双语变换矩阵的双语种子字典，我们使用了一个实验室内部维护的英汉词典，这个词典有55668个英文词，137420个汉语词。我们使用英语和汉语的单语语料的词汇表对这个种子字典进行过滤。从过滤后的种子字典中，我们随机的选择500个英文词做开发集来调整超参数，随机选择出2500个英文词做测试集，其余的作为训练集。具体的训练集和测试集的数据参见表2。

我们采用单向线性变换作为基准方法，采用Top-N Precision和MRR (mean rank reciprocal)两个评价指标来衡量本文方法的有效性。

Top-N Precision的定义为：

$$P_N = \frac{\sum_t m(t)}{T}$$

其中， T 为测试样例个数， t 为测试样例。若 t 的某一个正确翻译出现在前 N 个翻译候选中，则 $m(t)=1$ 。 P_N 值越大越接近于1表明系统的性能越好。

MRR的定义为：

$$MRR = \frac{1}{T} \times \sum_{i=1}^T \frac{1}{rank_i}$$

其中， T 为测试样例个数， $rank_i$ 表示在第 i 个测试样例的翻译候选中，正确候选所在的排序，如果不在，则为0。本实验中，每个待翻译的词会有多个正确翻译，平均值为4.5个左右，当出现候选集中有多个正确翻译的情况时，我们选择排名最靠前的正确翻译的排序。MRR值越大越接近于1表明系统的性能越好。

4.2 实验结果

我们首先使用word2vec工具包在英语语料和汉语语料上获取词嵌入。为了验证我们的重排序方法在不同维度词嵌入上的有效性，我们在两个语言上都分别训练了200, 400, 600, 800, 1000维的词嵌入，以此来测试基准方法和重排序方法在开发集的性能。我们首先固定了超参数 $\alpha = \beta = 0.5$ ，翻译候选集的长度为100，两种方法的Top-1 Precision性能参见图2。图2中，虚线为双向重排序方法的性能，实线为基准方法的性能。从图2中可以看出，在所有的维度上，相比仅使用单向线性方法的基线系统，使用双向重排序方法可以显著提升系统的性能。其中词嵌入维度在600维和800维时表现最好。综合考虑，我们在接下来的实验中都选择800维的词嵌入。

接下来，我们考察不同长度的翻译候选集对双向线性重排序方法性能的影响。我们设置了翻译候选集的长度为 $N = \{1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100, 150, 200\}$ 。具体地性能曲线如图3所示。

从图3中可以看出，双向重排序方法在初始增加翻译候选集大小时，性能增长很快。

¹ <http://nlp.stanford.edu/software/segmenter.shtml>

² <https://code.google.com/p/word2vec>

当N增加到一定程度（50）时，系统的性能达到饱和。从图3中可以看到，增加翻译候选集的大小，可以显著提升双向线性重排序方法的性能。综合考虑，我们在接下来的实验中都选择翻译候选集的大小为100。

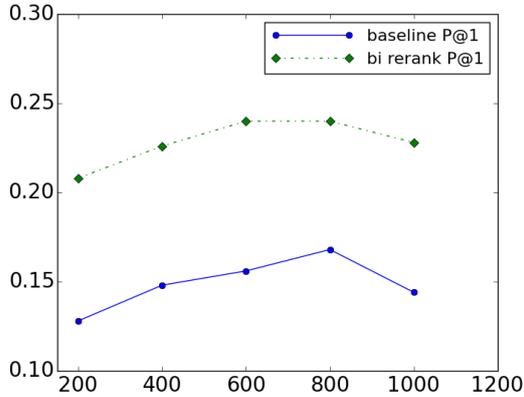


图2 双向重排序方法在开发集上不同维度词嵌入上的 Top-1 Precision 性能曲线

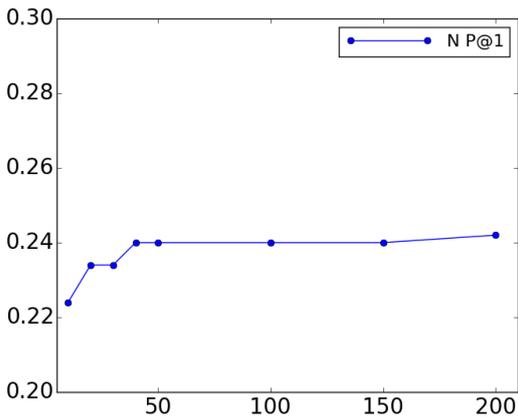


图3 双向重排序方法在开发集上不同规模翻译候选集上的性能曲线

最后，我们固定词嵌入的维度为800，翻译候选集的长度为100，考察不同的 α 的值对双向重排序方法的影响，我们设置了 $\alpha = \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ 。具体的性能曲线参见图4。

表3 双向重排序方法在测试集上的 Top-N Precision

	P@1	P@10	P@50
单向方法	0.134	0.328	0.472
双向方法	0.227	0.424	0.520

根据不同超参数在开发集上的性能的影响，最终我们选择了如下的超参数配置：

$$d = 800, \alpha = 0.4, N = 100$$

最终，双向线性重排序的方法和基准方法的在测试集上的性能对比见表3和表4。

表4 双向重排序方法在测试集上的 MRR 值

	MRR
单向方法	0.199
双向方法	0.295

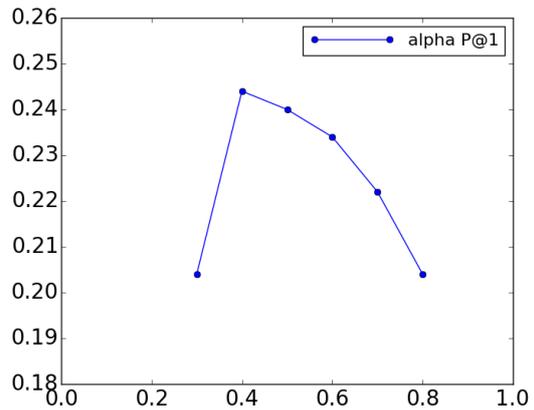


图4 双向重排序方法在开发集上不同 α 值时的性能曲线

从表3和表4中可以看出，基于双向重排序的方法可以显著提升基线系统的性能，Top-1的精度可以从0.134增长到0.227，提升可以达到69%，在Top-10和Top-50上的增长也分别达到了29%和10%。在MRR上，也有显著的增长。我们在开发集上选择3个样例，定性地观察双向重排序方法对翻译结果的影响，如表5所示。从表5可以看出，由于“召回”译为“recall”的得分很高，当使用双向重排序后，“recall”的正确翻译“召回”的得分变得最高；由于“打扰”译为“beware”的得分很低，当使用双向重排序后，将“beware”的错误翻译“打扰”的排序降低，正确翻译“提防”和“当心”的排序提升；对于“netscape”使用双向重排序方法后，虽然正确翻译“网景”未能排序到第一，但仍然比之前的排序有所提高。

最后，为了验证双向线性重排序方法在其他数据集上的有效性，我们使用了一个由其他研究先前在Google新闻数据集(Google News dataset)上事先训练好的英语词嵌入³进行了实验。这个Google新闻数据集有近1000亿个词形(Tokens)，训练好的词嵌入

³ GoogleNews-vectors-negative300.bin.gz

包含 300 百万个词和短语，维度为 300 维。我们保持中文词嵌入不变，维度为 800 维，用同样的测试集进行测试，性能对比参见表

6。尽管 Google 的词嵌入词汇集非常大，但仍然有约 5% 的测试词不能被覆盖。但在能覆盖的测试词上，双向重排序方法仍然能够获

表 5 双向重排序方法的部分样例输出

样例	翻译候选	双向重排序得分	S2T 得分	T2S 得分	正确翻译
recall	召回	0.418	0.406	0.427	召回、提醒、记得、罢免、回忆、收回、回想
	撤下	0.363	0.368	0.360	
	退回	0.360	0.424	0.318	
	扣留	0.352	0.434	0.298	
	开除	0.345	0.348	0.344	
beware	提防	0.470	0.482	0.465	提防、注意 当心、小心
	当心	0.430	0.453	0.415	
	打扰	0.425	0.498	0.377	
	提醒	0.425	0.440	0.415	
	小心	0.419	0.485	0.374	
netscape	因特网	0.427	0.476	0.379	网景
	互联网	0.423	0.464	0.382	
	网景	0.412	0.428	0.395	
	浏览器	0.403	0.473	0.332	
	服务器	0.394	0.417	0.371	

较为明显的性能提升，表明基于双向线性得分加权重排序的方法对在其他语料上（Google 英文新闻语料与中文训练语料的可比性更差）获得的词嵌入也有十分明显的效果。

表 6 双向重排序方法在 Google 数据集上的性能

	P@1	P@10	P@50	Coverage
单向	0.104	0.262	0.400	95.2%
双向	0.186	0.361	0.438	95.2%

5 相关工作

先前，基于可比较语料的双语种子字典抽取的方法一般是基于分布假设（**Distributional Hypothesis**）^[11]。这种假设认为一个词和它的翻译词在语义空间上应该共享类似的上下文。基于这样的假设，通常一个无监督的标准实现^{[6][12]}会被采用，在这种实现中，使用与目标词共同出现的词作为其的表示向量，抽取目标语言端与这个表示向量最相似的词作为翻译候选输出。这种方法受词频影响较大，通常只有高频词的上下文向量表示才不太稀疏，对中低频词的性能非常差。

另外一种无监督的实现方法是使用双语主题模型^[13]来表示目标词。这种方法希望能避免种子字典的使用，但是不足是这种方法需要实现篇章对齐的文档。

近年来有监督的双语词典抽取方法吸引了研究者的注意。一个线性分类器^[14]和随机森林分类器^[15]被用于了这个任务，这种方法需要手工地来设计分类器的特征，方法的适应性和鲁棒性不好。最近，Mikolov 等人^[9]利用词嵌入来学习双语变换矩阵，这种方法获取了较高的精度和覆盖率。基于这种方法，很多研究者进行相关的扩展。文献^[16]中提出了一个正则化的目标函数来加强词嵌入。文献^[17]中使用典型相关分析（CCA）技术来对提高双语词嵌入空间的线性程度，进而提高了双语字典抽取的性能。文献^[18]采用 EMD 方法来改进双语词典抽取中在预测阶段时用近邻方法检索目标翻译的不足。

6 结论

本文针对基于词嵌入的线性变换可比语料中双语词典抽取方法的线性假设不足，提出了一种基于双向线性得分加权和的翻译候

选重排序的方法。这种方法综合考虑和源语言到目标语言和从目标语言到源语言两个方向的线性变换得分,在英汉新闻领域的可比语料英汉双语词典任务上的实验表明,基于双向线性的重排序方法能够显著提升基线系统的性能,Top-1的精度相对增长69%,在 Google 英文数据集上重排序同样也获得较为明显的增长。未来工作将在其他语言对、其他领域数据集进行双向线性重排序方法的测试。同时,基于机器学习的重排序方法也将被考虑。

致谢

本文受国家自然科学基金项目:面向汉语文本理解的语义计算方法(项目号:91520204)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] Ballesteros L, Croft W B. Phrasal translation and query expansion techniques for cross-language information retrieval[C]//ACM SIGIR Forum. ACM, 1997, 31(SI): 84-91.
- [3] Munteanu, Dragos Stefan, and Daniel Marcu. "Extracting parallel sub-sentential fragments from non-parallel corpora." *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2006.
- [4] Fung P. A statistical view on bilingual lexicon extraction: from parallel corpora to non-parallel corpora[M]//*Machine Translation and the Information Soup*. Springer Berlin Heidelberg, 1998: 1-17.
- [5] Rapp R. Identifying word translations in non-parallel texts[C]//*Proceedings of the 33rd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1995: 320-322.
- [6] Laroche A, Langlais P. Revisiting context-based projection methods for term-translation spotting in comparable corpora[C]//*Proceedings of the 23rd international conference on computational linguistics*. Association for Computational Linguistics, 2010: 617-625.
- [7] Bengio Y, Schwenk H, Senécal J S, et al. *Neural probabilistic language models*[M]//*Innovations in Machine Learning*. Springer Berlin Heidelberg, 2006: 137-186.
- [8] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. *J. Mach. Learn. Res.* 12, 2493–2537 (2011)
- [9] Mikolov T, Le Q V, Sutskever I. Exploiting Similarities among Languages for Machine Translation[J]. *arXiv preprint arXiv:1309.4168*, 2013.
- [10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of NIPS*, 2013.
- [11] Harris Z S. *Distributional structure*[M]. Springer Netherlands, 1981.
- [12] Bouamor D, Semmar N, France C, et al. Context Vector Disambiguation for Bilingual Lexicon Extraction from Comparable Corpora[J]. *ACL*. Sofia, Bulgaria (Cité page 35), 2013.
- [13] Vulic, I., Moens, M.F.: Detecting highly confident word translations from comparable corpora without any prior knowledge. In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 449–459. Association for Computational Linguistics (2012)
- [14] Irvine, A., Callison-Burch, C.: Supervised bilingual lexicon induction with multiple monolingual signals. In: *HLT-NAACL*, pp. 518–523. Citeseer (2013)

- [15] Kontonatsios, G., Korkontzelos, I., Tsujii, J., Ananiadou, S.: Using a random forest classifier to compile bilingual dictionaries of technical terms from comparable corpora. In: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics: Short Papers, vol. 2, pp. 111–116 (2014)
- [16] Xing, C., Wang, D., Liu, C., Lin, Y.: Normalized word embedding and orthogonal transform for bilingual word translation. In: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1006–1011. Association for Computational Linguistics
- [17] Zhang, Chunyue, and Tiejun Zhao. "Bilingual Lexicon Extraction with Forced Correlation from Comparable Corpora." International Conference on Neural Information Processing. Springer International Publishing, 2015.
- [18] Zhang, Meng, et al. "Building Earth Mover's Distance on Bilingual Word Embeddings for Machine Translation." Thirtieth AAAI Conference on Artificial Intelligence. 2016.
- [19] Daumé III H, Jagarlamudi J. Domain Adaptation for Machine Translation by Mining Un-seen Words[C]//ACL (Short Papers). 2011: 407-412.
- [20] Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pp. 384–394. Association for Computational Linguistics (2010)