

# 基于双语合成语义的翻译相似度模型

王超超 熊德意<sup>†</sup> 张民

苏州大学计算机科学与技术学院, 苏州 215006; <sup>†</sup> 通信作者, E-mail: dyxiong@suda.edu.cn

**摘要** 提出基于双语合成语义的翻译相似度模型, 通过在翻译过程中引入双语语义相似度特征提高翻译性能。该模型首先利用分布式方法分别在源端和目标端获取短语的单词合成语义向量, 然后利用神经网络将它们映射到同一语义空间, 以获得双语合成语义向量, 在该语义空间, 计算源语言短语和对应的目标语言短语之间基于合成语义向量的翻译相似度, 将其作为一个新特征加入解码器。在汉英翻译 NIST06 和 NIST08 测试数据集上, 基于双语合成语义的翻译相似度模型相较于基准系统获得 0.56 和 0.42 BLEU 值的显著性提高。

**关键词** 语义合成; 机器翻译; 分布式表示; 神经网络

**中图分类号** TP391

## Translation Similarity Model Based on Bilingual Compositional Semantics

WANG Chaochao, XIONG Deyi<sup>†</sup>, ZHANG Min

School of Computer Science and Technology, Soochow University, Suzhou 215006; <sup>†</sup>Corresponding author,  
E-mail: dyxiong@suda.edu.cn

**Abstract** The authors propose a translation similarity model based on bilingual compositional semantics to integrate the bilingual semantic similarity feature into decoding process to improve translation quality. In the proposed model, monolingual compositional vectors for phrases are obtained at the source and target side respectively using distributional approach. These monolingual vectors are then projected onto the same semantic space and therefore transformed into bilingual compositional vectors. Base on this semantic space, translation similarity between source phrases and their corresponding target phrases is calculated. The similarities are integrated into the decoder as a new feature. Experiments on Chinese-to-English NIST06 and NIST08 test sets show that the proposed model significantly outperforms the baseline by 0.56 and 0.42 BLEU points respectively.

**Key words** semantic compositionality; machine translation; distributed representations; neural network

统计机器翻译作为人工智能领域的一个重要分支, 经历了基于词、短语和句法的发展阶段, 目前越来越多的研究人员开始探索基于语义信息的统计机器翻译方法。Zou 等<sup>[1]</sup>提出基于联合优化模型, 在双语语料库中同时训练, 得到双语的单词语义向量。Mikolov 等<sup>[2]</sup>提出基于神经网络的方法, 分别获取源端和目标端的单词语义向量, 然后采用线性映射方法, 将源端单词语义向量映射到目标端所在的语义空间。

上述工作虽然一定程度提升了翻译性能, 但是仍然存在着如下 3 个问题: 1) 对于语义信息的研究仍然局限于单词层面; 2) 采用联合优化模型(即在双语语料库中同时训练语义向量), 过程繁琐且训练时间较长; 3) 相对于非线性映射方法, 线性映射方法在向量值传递等方面的准确性相对较弱, 在实际映射中可能会丢失源端语义向量所包含的语义信息<sup>[3]</sup>。

本文提出基于双语合成语义的翻译相似度模

型, 不仅将语义信息从单词层延伸到短语层, 而且采用非线性映射方法, 将源端语义向量映射到目标端语义空间, 有效克服了上述问题。合成语义的概念最早在 1892 年由德国逻辑学家 Frege<sup>[4]</sup>正式提出, 是允许语言使用者理解他们从未听过的短语和句子语义的一个基本原则, 即一个合成表达的语义由它的各组成部分的语义以及把它们连接在一起的语法规则共同决定。合成语义信息能够有效扩大语义信息的表示范围, 基于合成语义信息的统计机器翻译方法对于翻译性能的提升具有重要的意义<sup>[5]</sup>。

基于双语合成语义的翻译相似度模型训练与集成过程可以概括为如下 3 个步骤: 首先, 在大规模语料库中, 通过分布式表示方法中的向量混合的方法, 在源端和目标端分别获取短语的单词合成语义向量; 其次, 利用神经网络作为向量映射模型, 将双语语义空间下的合成语义向量映射到同一语义空间; 最后, 利用翻译相似度模型计算它们的翻译相似度, 在解码过程中将该翻译相似度作为特征加入其中。在实验部分本文以无映射的双语翻译相似度、线性映射方法的翻译相似度以及带隐含层的非线性映射方法的翻译相似度作为特征。实验结果表明, 集成了带隐含层的非线性映射方法的翻译相似度特征在解码过程中取得最优的结果。以 BLEU 值作为评价尺度, 在测试数据集 NIST 06 和 08 上分别提升了 0.56 和 0.42 个点。

## 1 相关工作

近年来, 随着自然语言处理以及统计机器翻译技术的发展, 国内外研究人员在合成语义表示以及基于语义信息的统计机器翻译方法方面做了相关研究<sup>[6-7]</sup>。

基于分布式方法中向量混合方法的语义表示方面, Islam 等<sup>[8]</sup>提出使用逐点互信息 PMI 来表示单词的语义信息。在单词语义表示的基础上, Mitchell 等<sup>[9]</sup>提出通过向量乘法和带权向量加法表示短语的合成语义信息。基于分布式方法中向量混合方法的语义表示能够反映上下文窗口中合成语义的相关信息, 且训练速度较快。

基于语义信息的统计机器翻译方法方面, Zou 等<sup>[1]</sup>提出利用联合优化模型同时获取双语的语义信息。首先通过一个基于神经网络的模型获取源端的单词语义表示, 然后利用源端的语义信息以及词对齐信息初始化目标端的语义表示, 选择损失最小的

语义信息作为目标端的语义表示, 在解码的过程中取得一定的效果。Mikolov 等<sup>[2]</sup>提出利用神经网络和线性映射的方法获取单词语义信息, 并集成到解码器中。首先在源端和目标端大规模语料库中利用基于神经网络的方法训练两个单词的单词语义表示模型, 然后通过一个状态转移矩阵将源端的单词语义信息映射到目标端, 最后集成到解码器中, 取得良好的翻译效果。Gao 等<sup>[10]</sup>提出利用短语语义相似度改善机器翻译的质量。首先在源端和目标端分别抽取短语, 然后通过多层神经网络模型, 将源端和目标端中的短语映射到一个低维的浅层语义空间。于是, 对于源端目标端短语, 都可以用一个与语言环境无关的特征向量表示它们的语义信息, 最后通过计算各特征向量之间的翻译得分, 并将其作为一个特征加入到训练模型中。

尽管对合成语义表示以及基于语义信息的统计机器翻译方法有了一定的研究, 但是基于合成语义信息的统计机器翻译方法的研究仍然十分有限。针对此问题, 本文在双语合成语义的基础上, 提出翻译相似度模型<sup>[11]</sup>, 研究合成语义信息对于统计机器翻译质量的影响。

## 2 翻译相似度模型

翻译相似度模型刻画了同一个语义向量空间中两个多维向量  $\mathbf{u}$  和  $\mathbf{v}$  的相似程度。向量  $\bar{u}$  和  $\bar{v}$  的相似程度越高, 说明  $\mathbf{u}$  和  $\mathbf{v}$  所表示的语义越接近。在解码的过程中, 选择与源端短语语义最接近的目标短语作为翻译结果。翻译相似度可以由多种函数计算得到, 例如标量积、向量距离等, 但是这些方法在计算多维空间向量的相似程度时, 不能有效区分相似度的高低<sup>[10]</sup>。因此, 本文选择余弦相似度函数, 计算两个语义向量之间的翻译相似度, 通过计算语义向量  $\mathbf{u}$  和  $\mathbf{v}$  之间夹角的余弦值近似表示两个语义向量之间的相似程度, 如式(1)所示:

$$\text{Sim}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \times \|\mathbf{v}\|} = \frac{\sum_i (a_i \times b_i)}{\sqrt{\sum_i a_i^2 \times \sum_i b_i^2}} \quad (1)$$

式中,  $\text{Sim}(\mathbf{u}, \mathbf{v})$  表示语义向量  $\bar{u}$  和  $\bar{v}$  的翻译相似度,  $a_i$  和  $b_i$  分别表示  $\mathbf{u}$  和  $\mathbf{v}$  每一维的值。夹角越小, 相似性越高, 其余弦值越大; 夹角越大, 相似性越低, 其余弦值越小。

如表 1 所示, 在同一个语义空间下, 有一个源端短语和两个目标端短语的合成语义向量, 通过式

表 1 计算合成短语的翻译相似度实例

Table 1 Examples of calculating the translation similarity between compositional phrases

合成短语	语义向量	翻译相似度
举行 选举	30, 20, 20, 10, 0	$\text{Sim}_1(D_1, D_2) = 0.86$
hold election	40, 0, 30, 20, 10	
举行选举	30, 20, 20, 10, 0	$\text{Sim}_2(D_1, D_3) = 0.53$
celebrate election	10, 0, 20, 20, 5	

注: 表中合成短语的语义向量只是为了说明翻译相似度所用的测试数据。实验中我们训练 50 维的语义向量表示合成短语的语义信息。

(1) 得到它们的翻译相似度分别为  $\text{Sim}_1(D_1, D_2) = 0.86$  和  $\text{Sim}_2(D_1, D_3) = 0.53$ 。  $D_1$  表示短语“举行 选举”;  $D_2$  表示短语“hold election”;  $D_3$  表示短语“celebrate election”。比较发现,  $D_1$  与  $D_2$  之间的翻译相似度更高, 因此解码时, 选择  $D_2$  作为源语言的翻译结果。

## 2.1 合成语义向量获取

逐点互信息(PMI)提供了一种有效的方法, 它通过向量表示单词的语义信息, 向量中的每一个元素反映了目标单词在预先定义的上下文窗口中与单词表中其它单词同现的频率。PMI 可以通过式(2)计算得到:

$$\text{pmi}(c, t) = \log \frac{p(c, t)}{p(c)p(t)} = \frac{\text{freq}_{c,t} \times \text{freq}_{\text{total}}}{\text{freq}_{c_i} \times \text{freq}_t} \quad (2)$$

式中,  $t$  表示目标单词;  $c_i$  表示在上下文窗口中与  $t$  相邻的关联词, 本文预定义上下文窗口的长度为 5;  $\text{freq}_{c_i,t}$  表示关联词  $c_i$  与目标单词  $t$  共现的次数;  $\text{freq}_{\text{total}}$  表示所有单词出现的次数;  $\text{freq}_{c_i}$  表示上下文单词出现的次数;  $\text{freq}_t$  表示目标词出现的次数。

为了避免负数信息对真实语义信息的干扰, 我

们选取正逐点互信息(PPMI)计算单词的语义信息, 计算过程如式(3):

$$\text{ppmi}(c, t) = \begin{cases} \text{pmi}(c, t) & \text{当 } \text{pmi}(c, t) \geq 0, \\ 0 & \text{其它。} \end{cases} \quad (3)$$

获取到用 PPMI 值表示的单词语义向量表示后, 借鉴 Mitchell 等<sup>[9]</sup>提出的方法获取合成语义向量, 合成函数如式(4):

$$p = f(u, v, R, K) \quad (4)$$

式中,  $p$  表示合成语义向量,  $u$  和  $v$  分别表示用于合成的单词语义向量,  $R$  表示单词之间的语法关系,  $K$  表示所需的语言学背景知识, 即构造单词的合成语义时所需要的附加信息或知识。在同一个语料库中, 词与词之间、短语与短语之间具有相同的  $R$  和  $K$ 。因此, 对上述模型简化后, 可以使用带权加法函数计算短语的合成语义向量, 如式(5)所示:

$$p = \alpha u + \beta v \quad (5)$$

其中,  $\alpha$  和  $\beta$  表示两个可以通过训练调整的特征权值。本文通过 DISSECT<sup>①</sup>训练得到  $\alpha$  和  $\beta$  的值分别为 0.6 和 0.4。

## 2.2 双语语义向量映射

在源端和目标端分别获取短语的单语合成语义向量后, 采用如图 1 所示的神经网络作为向量映射模型, 将源端语义向量映射到目标端所在的语义向量空间。神经网络是一种模仿生物神经网络的结构和功能的数学模型, 它能够在外界信息的基础上改变内部结构, 常用来对输入和输出间复杂的关系进行建模。本节使用的单层神经网络通常被称作 BP 神经网络, 它是一种按误差逆传播算法训练的前馈网络, 能够学习和存储大量的输入-输出模式映射关系, 无需事前揭示描述这种映射关系的数学方程, 是目前应用最广泛的神经网络模型之一。

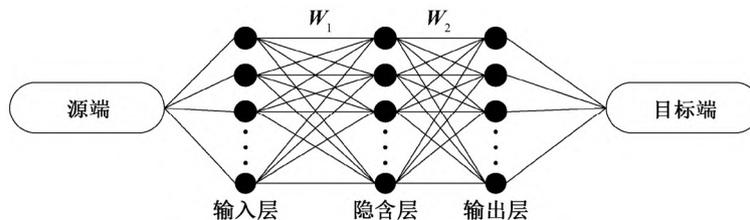


图 1 基于神经网络的双语语义向量映射框架

Fig. 1 Architecture of the bilingual projection neural network that projects vectors of source phrases to target semantic space

① <http://clic.cimec.unitn.it/composes/toolkit/index.html#>

本文采用的神经网络为只含有一个隐含层的单层神经网络, 具体的映射过程如式(6)所示:

$$p = g(W_2(g(W_1x + b_1)) + b_2) \quad (6)$$

其中,  $W_1$  是从输入层到隐藏层的映射矩阵;  $W_2$  是从隐藏层到输出层的映射矩阵;  $b_1$  和  $b_2$  为偏置值。 $g(x)$  通常选用非线性函数如 sigmoid 和 tanh。为了计算神经网络的权值, 我们采用式(7)计算平方差损失函数:

$$e = \frac{1}{2} \sum_i (t_i^m - p_i^m)^2 \quad (7)$$

其中,  $p_i^m$  是通过式(6)计算得到的映射语义向量,  $t_i^m$  是与输入的源语言短语语义向量对应的目标语言短语语义向量。向量映射的权值可以通过后向传播算法训练得到。对于收集到的训练实例集  $\{(\bar{s}, \bar{t})\}_1^n$  ( $p$  和  $t$  分别表示短语对  $(s, t)$  中源语言和目标语言短语向量), 通过不断修正平方差损失函数  $e$ , 最终得到最合适的权值矩阵。

### 3 实验

实验在汉语到英语方向上的翻译进行, 组织结构如图 2 所示。首先, 对语料库进行预处理, 通过分布式表示方法中的向量混合方法分别获取源端以及目标端短语的单语合成语义向量。然后, 使用向量映射模型将源端的语义向量映射到目标端所在的语义空间; 其次, 通过翻译相似度模型, 计算同一语义空间下, 源端和目标端短语语义向量的翻译相似度。最后, 将该翻译相似度作为一个特征加入到解码过程中, 获取相应的翻译结果。

实验中对比了非线性映射方法与线性映射方法

得到的双语语义向量对基于短语的机器翻译质量的影响。线性映射方法即为此处神经网络中去掉隐藏层的结果。通过该双语语义向量映射方法, 将源语言端与目标语言端合成语义向量映射到同一个语义空间, 在该语义空间使用翻译相似度模型计算源语言端与目标语言端短语的翻译相似度。

#### 3.1 实验数据

实验所采用的双语语料约含有 400 万平行句对。这些句对来自 LDC 语料的部分子集, 包括: LDC2003E07, LDC2004T07, LDC2004T08 和 LDC2005T06。首先, 使用 GIZA++<sup>[12]</sup> 工具获得汉英、英汉两个方向的词语对齐, 然后采用 growdiag-final-and<sup>[13]</sup> 的启发式方法获得多对多的词语对齐。使用 SRILM<sup>[14]</sup> 工具训练的四元语言模型, 并使用 KN 方法进行平滑。

实验评测语料使用 2005 年 NIST 的评测语料 NIST05 作为开发集, 使用 2006 和 2008 年 NIST 的评测语料 NIST06 和 NIST08 作为测试集。使用的解码器是层次短语解码器的 C++ 重实现版本<sup>[15-16]</sup>, 该解码器采用 CKY 方式进行解码, 并使用 Cube-Pruning 的方法进行剪枝以减少搜索空间, 实验使用的栈大小为 100。机器翻译实验中使用最小错误训练方法<sup>[17]</sup>, 优化线性模型的参数, 采用大小写不敏感的 BLEU-4<sup>[18]</sup> 作为评测指标。

#### 3.2 实验预处理

实验预处理主要包括合成语义向量获取和双语语义向量映射两个工作, 可以分为以下两个步骤。

1) 获取合成语义向量。对获取到的短语语料库进一步处理, 统计各个短语出现的次数, 将该数

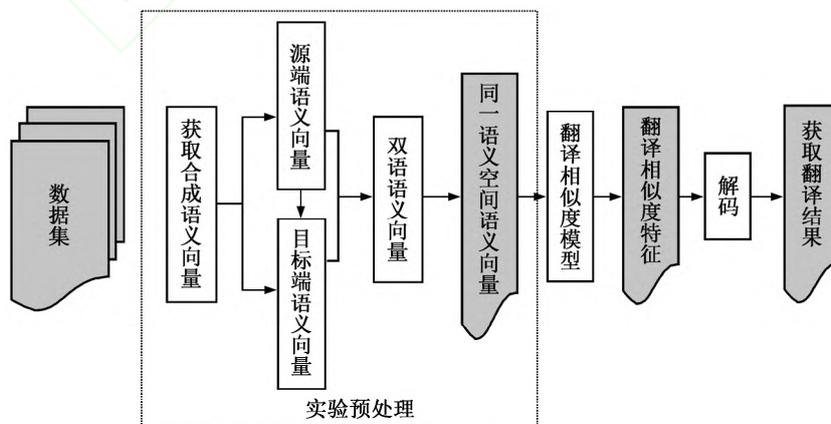


图 2 基于语义表示的翻译相似度模型框架

Fig. 2 Framework of the semantic-based translation similarity model

据作为输入利用开源的构建向量模型工具 DISSECT, 在中文端和英文端分别获取短语的合成语义向量<sup>[19]</sup>。

2) 双语语义向量映射。经过第一步, 在中文端和英文端分别获取约一千万个 50 维的合成语义向量。此时, 使用基于神经网络的向量映射方法, 将中文端的语义向量映射到英文端的语义空间。首先, 将中文端的语义向量作为输入, 利用双语对齐信息将与之对应的英文端的语义向量作为输出, 训练生成具有一个隐含层包含 50 个隐含层单元的神经网络模型。然后, 仍然以中文端的语义向量作为输入, 定义输出格式为 50 维, 利用该训练好的神经网络模型, 将中文端的语义向量映射到英文端语义空间。

### 3.3 模型训练和解码

#### 3.3.1 模型训练

我们要解决在层次短语模型解码器中如何计算短语的双语语义相似度特征, 如何将该双语语义相似度作为对数线性模型中新增的特征  $h_{f_e}(f, e)$ 。开发集 NIST05 通过 MERT<sup>[14]</sup>迭代多轮后, 选择最优的权重  $\lambda_{f_e}$  作为测试集 NIST06 和 NIST08 的权重。该新增特征的计算过程如式(8)所示:

$$h_{f_e}(f, e) = \sum_{(f, e) \in A} \text{sim}(u_f, v_e) \quad (8)$$

其中,  $f$  和  $e$  分别表示源语言和目标语言;  $A$  为对齐规则;  $u_f$  和  $v_e$  分别表示 3.2 节实验预处理中获取的源语言和目标语言语义向量表中查找到的短语语义向量。模型的训练过程如算法 1 所示。

**算法 1** 双语合成语义的翻译相似度模型特征提取算法

输入: 1) 同步语法规则  $X \rightarrow \langle \gamma, \alpha, \sim \rangle$ , 其中  $X$  是非终结符,  $\gamma, \alpha$  是由终结符和非终结符构成的字符串,  $\sim$  是非终结符对应关系

2) 源语言短语语义向量表 f-table

3) 目标语言短语语义向量表 e-table

输出: 对数线性模型中的翻译相似度特征  $h_{f_e}(e, f)$

算法流程:

1 for 每一个终结符 f\_word in  $\gamma$  do

2 对于每一个终结符记录下标号  $i =$

index(f\_word)

3 if  $(\gamma[i], \gamma[i+1])$  在 f-table 中 //在源语言语义向量表中依序查找规则表中相应短语

4 记录下它的向量  $v_f$

5 SumVec $_{\gamma} += v_f$  //求出翻译假设中

所有短语语义向量的总和

```

6          AvgVec $_{\gamma} = \text{SumVec}_{\gamma} / n$  //求出整个
翻译假设的语义向量均值
7          i+=2
8          else
9          i++
10         Continue
11        end if
12    end for
13    for 每一个终结符 e_word in  $\alpha$  do
14        与源端的处理过程相似, 得到 AvgVec $_{\alpha}$ 
15    end for
16    计算翻译相似度特征
 $h_{f_e}(f, e) += \text{Sim}(\text{AvgVec}_{\gamma}, \text{AvgVec}_{\alpha})$ 
17    for 每一个非终结符 X in  $\alpha$  do
18        把 X 中的特征提取出来加载到  $h_{f_e}(f, e)$ 
19    end for
20    return  $h_{f_e}(f, e)$ 

```

本文使用的短语定义为步长为 2 的二词短语。

对于源语言句子, 首先在层次短语模型解码器的对齐规则表中找到相应的短语; 若不存在则跳过, 然后在源语言语义向量表中, 查找到翻译假设中所有二词短语对应的短语语义向量, 并求出该翻译假设的均值。对于目标语言句子, 通过类似的过程查找到对应的短语语义向量。最后, 通过式(8)计算得到该翻译相似度特征, 并将它作为对数线性模型的一个特征集成到解码器。

#### 3.3.2 解码

实验使用的解码器是层次短语解码器的 C++ 重实现版本, 并使用 Cube-Pruning 方法进行剪枝, 加快了解码的速度。直接翻译概率可以用式(9)形式化地描述:

$$\Pr(e|f) \approx p_{\lambda_1 \dots \lambda_M}(e|f) = \frac{\exp[\sum_{m=1}^M \lambda_m h_m(e, f)]}{\sum_e \exp[\sum_{m=1}^M \lambda_m h_m(e', f)]} \quad (9)$$

其中  $e$  和  $f$  是机器翻译的目标端和源端句子;  $h_1(e, f), \dots, h_M(e, f)$  分别是  $e$  和  $f$  上的  $M$  个特征;  $\lambda_1, \dots, \lambda_M$  是与这些特征分别对应的  $M$  个参数。在解码过程中, 直接将本文提出的翻译相似度模型作为一个新的特征  $h_{f_e}(f, e)$ , 集成到对数线性模型中。

### 3.4 实验结果与分析

表 2 是实验结果, 包含开发集和测试集两个部分。其中, 第 3 行是基准系统的结果, 第 4 行是加入无映射的翻译相似度特征的结果。第 5 行是加入线性映射方法, 将双语语义向量映射到同一语义空

间下翻译相似度特征的结果。最后一行是加入带有隐含层的神经网络映射方法,将双语语义向量映射到同一语义空间下翻译相似度特征的结果。

以 BLEU-4 作为评价指标,从表 2 可以清楚地看到: 1) 与基准系统相比,加入无映射的翻译相似度特征的方法在测试集 NIST06 和 NIST08 分别下降 0.28 和 0.20; 2) 线性映射方法,在同一个语义空间下,计算翻译相似度并将它作为一个特征,相比于基准系统,该方法的结果上升并不明显,在测试集 NIST06 和 NIST08 分别只提升 0.11 和 0.24; 3) 带有单个隐含层的神经网络的向量映射方法,选择映射到同一语义空间下的翻译相似度作为特征,该方法的结果取得明显提升,在测试集 NIST06 和 NIST08 分别提升 0.56 和 0.42。

在翻译的过程中,本文以 NIST06 中的“谈判和妥协是民主的重要手段!”为例,比较说明加入语义向量翻译相似度特征对翻译结果的影响,对比结果如表 3 所示。从表 3 可以得到以下结论。

表 2 开发集与测试集上的实验结果

Table 2 Experimental results on develop set and test sets

方法	开发集 NIST05	测试集	
		NIST06	NIST08
基准系统	33.53	30.66	23.32
+无映射的翻译相似度特征	32.76	30.38	23.12
+线性映射翻译相似度特征	34.62	30.77*	23.56*
+非线性映射翻译相似度特征	34.53	31.22*	23.74*

注: \*表示显著性测试中 $p < 0.05$ 。

表 3 翻译结果对比实例

Table 3 Examples of translation results comparison

方法	翻译实例
待翻译句子	谈判和妥协是民主的重要手段!
参考译文	Negotiation and Compromise are the Important Means of Democracy!
基准系统	talks with Compromise is the to hold of elections
+无映射的翻译相似度特征	they and summer go to the main that
+线性映射翻译相似度特征	talks and Compromise is the main tools of democracy
+非线性映射翻译相似度特征	talks and Compromise are the important tools of democracy

注: 表中加粗斜体表示与参考译文一致的单词。

1) 加入无映射的翻译相似度特征时,其翻译结果与参考译文以及基准系统相比,都比较差。究其原因,两个不同语义空间下训练得到的语义向量的语义信息不同,简单计算两者之间的翻译相似度,没有实质意义,反而会导致错误的调优。

2) 双语语义向量映射方法得到的翻译结果,明显比基准系统和无映射的翻译相似度方法更接近参考译文。原因是,基于合成语义信息的统计机器翻译方法相比于传统的层次短语翻译方法,携带了部分上下文信息,有利于解码器选择更优的翻译结果。

3) 相比于线性映射方法,采用非线性映射方法的翻译结果显然更为优越。原因是线性映射方法在语义向量之间映射时,可能丢失了部分重要的语义信息。

## 4 总结与展望

本文提出一种基于双语合成语义向量表示的翻译相似度模型,利用双语合成语义信息,有效提高了翻译的质量。本文的创新点主要包括两个方面:首先,成功地将单词语义信息扩展到了短语的层次,利用分布式表示方法中的向量混合方法,分别在源端和目标端获取短语的单语合成语义向量;其次,采用带隐含层的神经网络模型将双语语义向量映射到同一个语义向量空间,再利用翻译相似度模型,计算语义向量的翻译相似度。以 BLEU 值作为评价尺度,在测试集 NIST06 和 NIST08 的实验表明,加入语义相似度特征的 BLEU 值比基准系统平均上升 0.49 个点。

本文提出的翻译相似度模型对于合成语义向量的质量有较高要求,获取更高质量的合成语义向量将是我们的未来的研究方向,包括如下 3 个方面。

1) 虽然分布式向量混合模型的方法具有计算简单、训练速度快等特点,但是深度学习的方法能够更大程度地考虑句子的语义。因此,未来我们将考虑利用深度学习的方法获取单词以及合成短语的语义信息。

2) 本文利用只含有一个隐含层的神经网络,将中文端语义向量映射到英文端,虽然取得了良好效果,但是仍然具有一定的局限性,在映射的过程中可能丢失了部分语义信息。因此,未来我们将结合目前相关学者提出的深度学习方法,探究更好的向量映射模型。

3) 对于本文所使用的解码器, 我们将进一步优化对齐规则与语言模型, 以获得更快的速度和更优的结果。

### 参考文献

- [1] Zou W J, Socher R, Cer D, et al. Bilingual word embeddings for phrase-based machine translation // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, 2013: 1393–1398
- [2] Mikolov T, Le Q V, Sutskever I. Exploiting similarities among language for machine translation. arXiv preprint arXiv, 2013, 1309.4168
- [3] Bengio Y, Lamblin P, Popovici D, et al. Greedy layer-wise training of deep networks // Advances in Neural Information Processing Systems 19 (NIPS'06). Cambridge: MIT Press, 2007: 153–160
- [4] Frege G. *Über Sinn und Bedeutung* // Geach P, Black M. Translation from the Philosophical Writings of Gottlob Frege. Oxford: Blackwell, 1892: 56–78
- [5] Kartsaklis D. Compositional operators in distributed semantics. arXiv preprint arXiv, 2014, 1401.5327
- [6] Xiong D, Zhang M, Li H. Modeling the translation of predicate-argument structure for SMT // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Jeju: Association for Computational Linguistics, 2012: 902–911
- [7] Xiong D, Zhang M. A sense-based translation model for statistical machine translation // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Baltimore: Association for Computational Linguistics, 2014: 1459–1469
- [8] Islam A, Inkpen D. Second order co-occurrence PMI for determining the semantic similarity of words // Proceeding of the International Conference on Language Resource and Evaluation (LREC 2006). Genoa, Italy, 2006: 1033–1038
- [9] Mitchell J, Lapata M. Vector-based models of semantic composition // Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics. Columbus, Ohio, 2008: 236–244
- [10] Gao J, He X, Yih W, et al. Learning semantic representations for the phrase translation model // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore: Association for Computational Linguistics. Baltimore, 2014: 699–709
- [11] Xiao X, Xiong D, Zhang M, et al. A topic similarity model for hierarchical phrase-based translation // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1. Jeju: Association for Computational Linguistics, 2012: 750–758
- [12] Och F J, Ney H. A systematic comparison of various statistical alignment models. Computational Linguistics, 2003, 29(1): 19–51
- [13] Koehn P, Och F J, Marcu D. Statistical phrase-based translation // Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics. Edmonton, 2003: 54–58
- [14] Stolcke A. SRILM-an extensible language modeling toolkit // Proceedings of the 7th International Conference on Spoken Language Processing. Denver, 2002: 901–905
- [15] 肖欣延, 刘洋, 刘群, 等. 面向层次短语翻译的词汇化调序方法研究. 中文信息学报, 2012, 26(1): 37–50
- [16] Chiang D. Hierarchical phrase-based translation. Computational linguistics, 2007, 33(2): 201–228
- [17] Och F J. Minimum error rate training in statistical machine translation // Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics. Sapporo: Sapporo Convention Center, 2003: 160–167
- [18] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, 2002: 311–318
- [19] Turney P D, Pantel P. From frequency to meaning: vector space models of semantics. Journal of artificial intelligence research, 2010, 37(1): 141–188