

# 基于深层神经网络(DNN)的 汉-越双语词语对齐方法

莫媛媛<sup>1</sup>, 郭剑毅<sup>1,2\*</sup>, 余正涛<sup>1,2</sup>, 毛存礼<sup>1,2</sup>, 牛翊童<sup>1</sup>

(1. 昆明理工大学信息工程与自动化学院, 云南 昆明 650051;

2. 昆明理工大学智能信息处理重点实验室, 云南 昆明 650051)

**摘要:**针对汉-越双语因语言特点差异较大而导致难以实现词语自动对齐的问题,提出了一种基于深层神经网络(deep neural network, DNN)的汉-越双语词语对齐方法。该方法先将汉-越双语词语转化成词向量,作为DNN模型的输入,再通过调整和扩展HMM模型,并融入上下文信息,构建DNN-HMM词语对齐模型。实验以HMM模型和IBM4模型为基础模型,通过大规模的汉-越双语词语对齐任务表明,该方法的准确率、召回率较两个基础模型都有明显的提高,而词语对齐错误率大大降低。

**关键词:**汉语;越南语;词语对齐;DNN

**中图分类号:**TP391 **文献标志码:**A

**引用格式:**莫媛媛, 郭剑毅, 余正涛, 等. 基于深层神经网络(DNN)的汉-越双语词语对齐方法[J]. 山东大学学报(理学版), 2016, 51(1): 77-83.

## A bilingual word alignment method of Vietnamese-Chinese based on deep neural network

MO Yuan-yuan<sup>1</sup>, GUO Jian-yi<sup>1,2\*</sup>, YU Zheng-tao<sup>1,2</sup>, MAO Cun-li<sup>1,2</sup>, NIU Yi-tong<sup>1</sup>

(1. School of Information Engineering and Automation, Kunming University of Science and Technology,

Kunming 650051, Yunnan, China; 2. Intelligent Information Processing Key Laboratory, Kunming University of Science and Technology, Kunming 650051, Yunnan, China)

**Abstract:** It is difficult to achieve auto-alignment between Vietnamese and Chinese, because their syntax and structure are quite different. In this case, we present a novel method for the Vietnamese-Chinese word alignment based on DNN (deep neural network). Firstly, we should convert Vietnamese-Chinese bilingual word into word embedding, and as the input within DNN. Secondly, DNN-HMM word alignment model is constructed by expanding HMM model, which also integrating the context information. The basic model of the experiments are HMM and IBM4. The results of large-scale Vietnamese-Chinese bilingual word alignment task show that this method not only significantly improved its accuracy and recall rate than the two basic models, but also greatly reduced word alignment error rate.

**Key words:** Chinese; Vietnamese; word alignment; DNN

## 0 引言

汉-越双语词语对齐方法是自动构建大规模双语平行语料库的关键技术,对机器翻译、跨语言检索、双语词典编纂等研究都有很大的价值。但针对处于被全世界广泛使用第14位的越南语的研究还很少。目前

收稿日期:2015-03-03; 网络出版时间:2015-12-29 16:07:29

网络出版地址:<http://www.cnki.net/kcms/detail/37.1389.N.20151229.1607.004.html>

基金项目:国家自然科学基金资助项目(61262041);云南省教育厅基金重大专项资助项目(2013FA030)

作者简介:莫媛媛(1989—),女,硕士研究生,研究方向为自然语言处理、信息抽取. E-mail:yuanyuan2013ly@163.com

\* 通讯作者:郭剑毅(1964—),女,硕士,教授,研究方向为自然语言处理、信息抽取. E-mail:gjade86@hotmail.com

已有较成熟的越南语词法分析及词性标记的研究<sup>[1-3]</sup>。双语词语对齐最早是由 Brown 等<sup>[4]</sup>作为机器翻译的中间隐含过程提出来的,并采用最大期望算法估计词语对齐概率;Och 等学者<sup>[5]</sup>提出了对齐模板的方法,开发出基于 IBM 的 5 个模型的词语对齐开源工具 GIZA++ 实现双语词语自动对齐;Blunsom 等<sup>[6]</sup>利用条件随机场模型,通过二叉决策和改善对称启发式搜索,实现一对多、多对多的词对齐;Liu 等<sup>[7]</sup>等将统计机器翻译中非常有效的对数线性模型应用于词语对齐,建议在框架下用短语和规则对上下文信息建模。上述的双语词语对齐方法都是建议用手工选取的特征来提高词语对齐的效率。2006 年 Hinton 等<sup>[8]</sup>提出深度学习(deep learning)方法。一些学者将深层神经网络(deep neural network, DNN)用于解决 NLP 中的一些问题,如:Collobert 等<sup>[9]</sup>利用 DNN 解决命名实体识别、语义标注和句法分析等问题,取得与先前工作相似甚至更好的效果;Niehues 等<sup>[10]</sup>通过结合神经语言模型和传统的 n-gram 语言模型来改善机器翻译结果。在词语对齐中使用 DNN 的方法开始于词向量阶段,并利用 DNN 自动学习词语的特征来提高词语对齐的效率, Mikolov 等<sup>[11-12]</sup>利用 Word2vec 构建了一个简单地线性分类网络;Titov 等<sup>[13]</sup>学习上下文无关的双语词向量来促进跨语言信息检索;Zheng 等<sup>[14]</sup>利用 word2vec 工具训练产生的词向量来进行中文分词和词性标注;Dahl 等<sup>[15]</sup>将上下文相关的 DNN-HMM 方法应用到语音识别领域,并提出该方法明显优于上下文相关的 GMM-HMM 模型;Yang 等<sup>[16]</sup>利用类似于 Dahl 等上下文相关的 DNN-HMM 方法实现汉-英双语词语对齐,取得较好的词语对齐结果。

本文以实现自动从原词中进行“特征学习”来提高词语对齐的效率为目的,借鉴深度学习<sup>[16]</sup>的思想,通过查找词向量矩阵  $W_v$ ,将汉-越双语词语映射成词向量作为 DNN 输入,深层信念网络(deep belief network, DBN)根据汉-越双语语言特点,考虑句法结构、词汇翻译信息、周围词汇信息等丰富的上下文信息,自动地从汉-越词语中学习合适的特征,利用 DNN 计算词汇翻译得分,最后以 DNN-HMM 位变模型实现汉-越词语对齐。

## 1 深层神经网络(DNN)

受限波尔兹曼机(restricted boltzmann machine, RBM)可以被视为一个由可见层和隐藏层组成的无向图模型<sup>[17]</sup>,如图 1 所示。 $v$  为可见层,用于表示观测数据, $h$  为隐层,可视层为一些特征提取器, $W$  为两层之间的连接权重。假设一个 RBM 有  $n$  个可见单元和  $m$  个隐含单元,用向量  $v_i$  和  $h_j$  分别表示可见单元和隐含单元的状态,其中, $v_i$  表示第  $i$  个可见单元的状态, $h_j$  表示第  $j$  个隐含单元的状态。对于一组给定的状态  $(v_i, h_j)$ ,RBM 作为一个系统所具备的能量函数定义为

$$E(v, h | \theta) = - \sum_{i=1}^n a_i v_i - \sum_{j=1}^m b_j h_j - \sum_{i=1}^n \sum_{j=1}^m v_i W_{ij} h_j, \quad (1)$$

其中, $\theta = \{W_{ij}, a_i, b_j\}$  是 RBM 的参数,均为实数, $W_{ij}$  表示可见单元  $i$  与隐含单元  $j$  之间的连接权重, $a_i$  表示可见单元  $i$  的偏置, $b_j$  表示隐含单元  $j$  的偏置。当参数确定时,基于该能量函数可以得到  $(v, h)$  的联合概率分布为

$$P(v, h | \theta) = \frac{e^{-E(v, h | \theta)}}{Z(\theta)}, \quad Z(\theta) = \sum_{v, h} e^{-E(v, h | \theta)}. \quad (2)$$

其中, $Z(\theta)$  为归一化因子或称为配分函数(partition function),它的作用是保证  $P(v, h | \theta)$  是一个概率值。模型赋予可见单元  $v$  的边缘分布概率等于对  $P(v, h | \theta)$  的所有隐藏单元求和(也称为似然函数):

$$P(v | \theta) = \frac{1}{Z(\theta)} \sum_h e^{-E(v, h | \theta)}. \quad (3)$$

为了确定该分布,需要计算归一化因子  $Z(\theta)$ ,这需要  $2^{n+m}$  次计算。因此,即使通过训练可以得到模型的参数  $W_{ij}$ 、 $a_i$  和  $b_j$ ,但仍旧无法有效地计算由这些参数所确定的分布。由 RBM 的特殊结构(即层间有连接,层内无连接)可知:当给定可见单元的状态时,各隐含单元的激活状态之间是条件独立的。此时,第  $j$  个隐含单

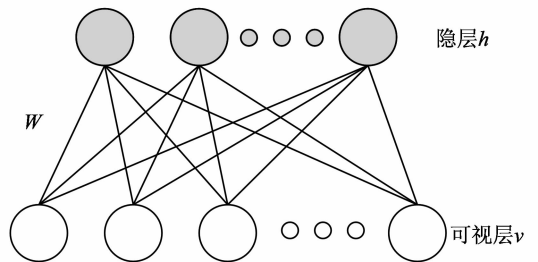


图1 RBM 示意图

Fig. 1 The schematic diagram of RBM

元的激活概率为

$$P(h_j = 1 | v, \theta) = \sigma\left(\sum_i v_i W_{ij} + b_j\right), \quad (4)$$

其中,  $\sigma(x) = \frac{1}{1 + e^{-x}}$  为 sigmoid 激活函数。

由于 RBM 的结构是对称的,当给定隐含单元的状态时,各可见单元的激活状态之间也是条件独立的,即第  $i$  个可见单元的激活概率为

$$P(v_i = 1 | h, \theta) = \sigma\left(\sum_j W_{ij} h_j + a_i\right). \quad (5)$$

学习 RBM 的任务是求出参数  $\theta$ , 以拟合给定的训练数据。参数  $\theta$  可通过采用梯度下降法最大化  $\log p(v | \theta)$  得到

$$\frac{\partial \log p(v | \theta)}{\partial w_{ij}} = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}}, \quad (6)$$

其中,  $\langle v_i h_j \rangle_{\text{data}}$  表示由输入数据所确定的期望,  $\langle v_i h_j \rangle_{\text{model}}$  表示所有可能的  $\langle v, h \rangle$  模型的期望。

采用对比散度(contrastive divergence, CD) 算法训练 RBM, 该算法首先将可见单元状态设置为当前的训练样本集, 然后通过执行一步吉布斯采样获得一个可见层的重构  $\langle v_i \rangle_{\text{reconstruction}}$ , 最后根据式(4) 更新隐含单元的状态。因此, 各参数的更新准则为:

$$\Delta W_{ij} = \varepsilon(\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{reconstruction}}); \quad (7)$$

$$\Delta a_i = \varepsilon(\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{reconstruction}}); \quad (8)$$

$$\Delta b_j = \varepsilon(\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{reconstruction}}). \quad (9)$$

其中,  $\varepsilon$  是学习率(Learning rate),  $\langle \cdot \rangle_{\text{reconstruction}}$  表示重构后模型定义的分佈。

将多个 RBM 自底向上依次堆叠起来得到 DBN, 通过由低到高逐层训练这些 RBM 来实现: 首先将底部 RBM 作为原始输入数据训练; 然后将底部 RBM 抽取的特征作为顶部 RBM 的输入训练; 重复以上过程训练所需尽可能多的层数。获得 DBN 模型之后, 在其顶层之上增加一层经典神经网络计算两种语言句对的词汇翻译概率, 输出层的每个节点对应 HMM 中的一个状态。DNN 结构图如图 2 所示。

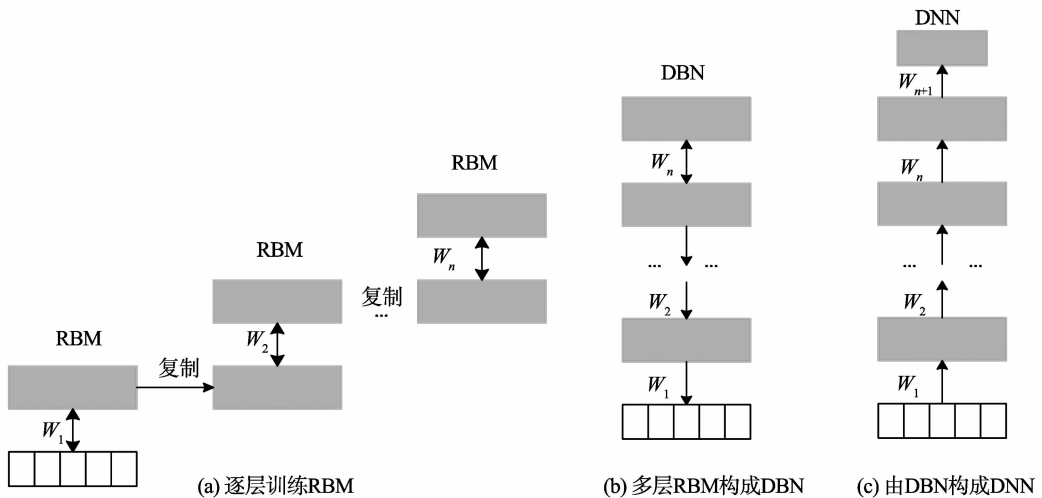


图 2 DNN 结构示意图  
Fig. 2 The schematic structure of DNN

## 2 词语对齐模型

### 2.1 语料的预处理

在 NLP 中最重要和最普遍的特点是词语本身, 把 DNN 的应用于 NLP 任务, 关键是把汉语和越南语中离散的词语转化为低维、高密度、实值的词向量, 词向量隐含编码的语义或句法等信息。假设一个有限大小的词汇集  $V$ , 其词向量化后形成  $(L \times |V|)$  维向量矩阵  $W_v$ , 词语向量化结果可通过简单地查找向量矩阵  $W_v$  各自

的行,每个查找过程被称为一个查找层 LT,其中  $L$  是一个预先确定的向量长度。本文使用 word2vector 工具包对汉语和越南语语料库中的词语进行词向量化,汉-越单语语料库分别用来预训练汉语和越南语的词向量,从单语语料库训练得到的词向量含有一定的语法知识,但这与汉-越双语之间的语法知识存在较大的差异。例如,许多中国词语可以同时作为一个动词、名词和形容词,没有任何变化,而在越南语中其不同的句法结构将产生完全不同的词向量,因此,在双语语料库训练过程中需要根据汉-越双语特点对单语语料库中的词向量进行微调。DBN 将考虑句法结构、词汇翻译信息、周围词汇信息等丰富的上下文信息,自动地从原词中进行“特征学习”。如表 1 中的句对 a 中,“Cô ấy(她) hát(唱歌) rất(很) hay(好听)”,汉语和越语句法成分最基本的排列顺序都是 SVO,在语序排列中,立题居前,解释居后是一般的语言规律,这种上下文的模式“X 唱歌很好听” $\Rightarrow$ “X hát rất hay”可能是有助于基于 DNN 词对齐填充变量  $X$ 。表 1 中句对 b 和句对 c 所示:汉语里修饰名词性成分的正定语一律排在中心语之前,而越南语里定语一般位于中心语之后,汉译越的翻译顺序为:中心语+分类性定语+描绘性定语+限定性定语,根据这样的词汇翻译信息,可方便在词语对齐过程中调整词序。对于表 1 句对 d 中的“vĩ đại”这个词不常见,很难直接对齐汉语词“大”,“to”、“lớn”与“vĩ đại”意思相近,在训练词向量时有一定相似度的词语的欧式距离较近,这样处理会易于找到对应的汉语词语。

表 1 汉-越词语结构特点  
Table 1 The structural features of Vietnamese and Chinese

汉语	越南语
a. 她唱歌很好听	Cô ấy(她) hát(唱歌) rất(很) hay(好听)
b. 她是我见过的最美丽的女孩	Cô là(她是) Cô gái(女孩) xinh đẹp nhất(最美丽的) mà tôi từng thấy(我见过的)
c. 孩子们很喜欢那种上面有字的白色的心状糖果	Trẻ con(孩子们) rất(很) thích(喜欢) loại(那种) kẹo(心状糖果) màu trắng(白色) bên trên có chữ(上面有字)
d. 这个饼挺大	Cái(个) bánh(饼) này(这) rất(挺) vĩ đại(大)

## 2.2 基于 DNN-HMM 的词语对齐

深层神经网络的词语对齐模型扩展了经典的 HMM 词对齐模型。给定一个句子对  $(e, f)$ , HMM 模型采用如下的词对齐模型:

$$P(a, e | f) = \prod_{i=1}^{|e|} P_{\text{lex}}(e_i | f_{a_i}) P_d(a_i - a_{i-1}), \quad (10)$$

其中,  $P_{\text{lex}}$  的是词汇翻译概率,  $P_d$  是跳转距离位变概率。直接将 DNN 融入 HMM 模型,需要增加一个 softmax 层将源语言中的词语归一化,由于自然语言中的词汇量非常庞大,很难归一化,故不采用概率表示,而采用如下形式的词语对齐模型:

$$S_{\text{NN}}(a | e, f) = \prod_{i=1}^{|e|} S_{\text{lex}}(e_i, f_{a_i} | e, f) S_d(a_i, a_{i-1} | e, f), \quad (11)$$

其中  $S_{\text{lex}}$  是通过 DNN 计算的词汇翻译得分,  $S_d$  是位变概率。在经典的 HMM 词对齐模型中不考虑上下文信息。本文把  $P_{\text{lex}}(e_i | f_{a_i})$  重写成  $P_{\text{lex}}(e_i | \text{context of } f_{a_i})$ , 考虑上下文信息,并没有由于数据稀疏而导致严重欠拟合问题。图 2 为 DNN 计算汉-越双语词汇翻译得分  $S_{\text{lex}}$ 。

词汇集  $V$  由源词汇  $V_e$  和目标词汇  $V_f$  组成,对于词语对  $(e_i, f_j)$  构造滑动窗口时,取固定长度的窗口  $e_i$  和  $f_j$  作为输入:  $(e_i - \frac{s_w}{2}, \dots, e_i + \frac{s_w}{2}, f_j - \frac{t_w}{2}, \dots, f_j + \frac{t_w}{2})$ , 其中  $s_w, t_w$  分别表示汉语和越南语的窗口的大小,使用查找表 LT 把词语转换为向量,输入到具有 3 个隐藏层的 DNN,该网络输出为词汇翻译得分:

$$S_{\text{lex}}(e_i, f_j | e, f) = \sigma_3 \circ \dots \circ \sigma_1 \circ \sigma_0(W^0 \times \text{LT}(e_i, f_j) + b^0), \quad (12)$$

其中  $\text{LT}(e_i, f_j) = \text{LT}(\text{window}(e_i), \text{window}(f_j))$ ,  $\sigma_0 \dots \sigma_2$  均使用 S 函数作为激活函数,  $\sigma_3$  是一个线性变换。对于位变概率  $S_d$  建立如下模型:

$$S_d(a_i, a_{i-1} | e, f) = S_d(a_i - a_{i-1} | \text{window}(f_{a_{i-1}})). \quad (13)$$

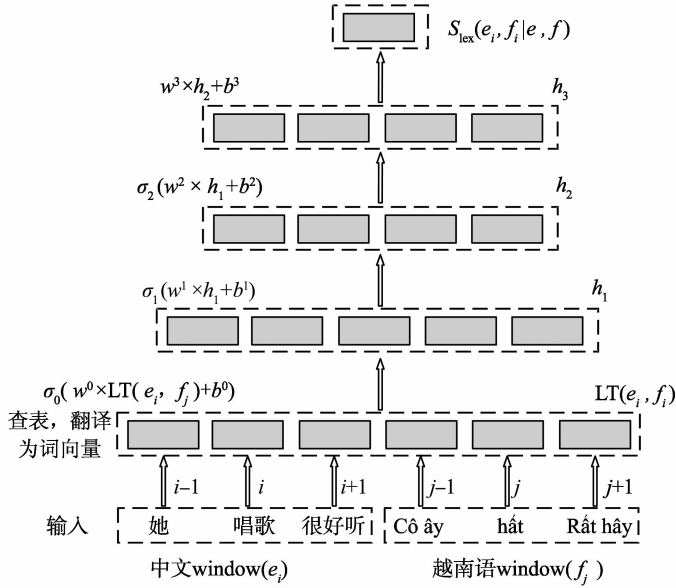


图 3 DNN 计算汉-越双语词汇翻译的得分

Fig. 3 DNN for computing lexical translation scores of Vietnamese and Chinese

式(13)中,词汇化位变参数相对于简单的距离位变参数来说,实验达不到预期的效果,因此,仍采用简单的位变参数:

$$S_d(a_i, a_{i-1} | e, f) = S_d(a_i - a_{i-1})。 \tag{14}$$

在词汇集中,除了真正的词语,每个词语含有一种特殊的未知字符 $\langle \text{unk} \rangle$ 处理隐藏的词语,两个句子的边界符号 $\langle \text{s} \rangle$ 和 $\langle \text{/s} \rangle$ ,在必要时将其填充到周围的窗口。为了处理空对齐,我们还必须包括一个特殊的空符号 $\langle \text{null} \rangle$ ,如果 $f_j$ 是空词,只需填写周边的窗口,使用相同的空符号。在模型的解码过程中,需要 DNN 为句中每个汉-越词对计算 $(|e| \times |f|)$ 次词汇翻译概率,使用类似于 HMM 模型中的反向传播算法查找 Viterbi 路径。

### 3 参数训练

由于采用无监督训练方法从句对中计算词汇翻译得分的训练过程较复杂,本文采用有监督方法训练 DNN 模型。DNN-HMM 词对齐模型中需要调整的参数包括:查找表 LT 中的词向量、权重矩阵  $\mathbf{W}$ 、偏置  $a_i, b_j$  和位变概率  $S_d$ 。采取如下所示的损失函数作为训练准则:

$$\text{loss}(\theta) = \sum_{\text{every}(e, f)} \max \{0, 1 - s_\theta(a^+ | e, f) + s_\theta(a^- | e, f)\}, \tag{15}$$

其中,  $\theta$  表示需要调整的超参数,  $a^+$  表示正确的对齐路径,  $a^-$  表示得分最高的不正确的对齐路径,  $s_\theta$  是在式(11)中定义的对齐路径的概率。一个汉语词语可能会对多个越南语词语,一个越南语词语也可能对应多个汉语词语,使用 grow-diagram-final 启发式策略来融合两个单向的词语对齐结果,并随机选择其中的一对作为最佳对齐路径。由于多层神经网络本质上是非线性的和非凸的,直接依据式(15)训练 DNN-HMM 模型很难产生良好的效果,因此忽略位变概率,利用局部句对准则训练 DNN-HMM 模型:

$$\max \{0, 1 - S_\theta((e, f)^+ | e, f) + S_\theta((e, f)^- | e, f)\}, \tag{16}$$

其中,  $(e, f)^+$  是一个正确的词对,  $(e, f)^-$  是同一句子中的一个不正确的词对。当 DNN 对正确词对的打分高于对随机词对的打分时, DNN 的输出才有意义,随机循环遍历训练数据的所有句对,对每一个正确的词语对(包括空对齐),得到一个正例,并通过随机更改该词语在句子中的前后词语产生两个负例,重复上述训练过程,寻找最佳对齐路径。由于训练没有明确的终止标准,在平行语料库中迭代  $N$  次对词对进行打分,  $N$  被设置为 50;为具体化 DNN 模型中的一些超参数,根据经验设置  $s_w, t_w$  为 11,隐藏层单元数分别为 300、120、10。

## 4 实验与结果分析

本文实验数据的获取来源于两个方面:从一些汉越双语网站上采用网络爬虫技术获取,本实验室人工收集的双语语料。我们使用20万条汉语-越南语句子对的平行语料作为训练集,使用人工对齐的5000条汉越双语句子对作为测试集。表2给出了它们的一些统计数据。

表2 训练集和测试集中包含词语数量  
Table 2 The number of words of training and testing set

语种	训练集			测试集		
	句子数	词语数	词汇量	句子数	词语数	词汇量
汉语	200 000	5 654 276	72 134	950	23 673	37.85
越南语	200 000	5 927 635	79 658	950	25 691	39.62

实验以HMM模型和IBM4模型为基础比较模型,并给出在汉-越双语平行语料词语对齐的实验结果和分析。在汉-越双语词语对齐过程中,采用准确率( $P$ )、召回率( $R$ )、 $F$ 值( $F$ )和对齐错误率(AER)作为评估指标,测试结果如表3所示。

表3 基于HMM模型和IBM4模型的词语对齐结果  
Table 3 The result of words alignments based on HMM and IBM4 model

测试类型	$P/\%$	$R/\%$	$F/\%$	AER/ $\%$
HMM	49.32	47.25	48.65	50.45
HMM + NN	53.83	51.13	52.84	46.18
IBM4	52.32	50.25	51.39	47.69
IBM4 + NN	58.58	56.13	57.34	41.45

实验结果表明,将DNN融入基础模型后,其准确率、召回率和 $F$ 值均有所提高,对齐错误率有较为显著的降低,这说明在基础HMM模型和IBM4模型加上DNN-HMM模型后能够取得较好的词语对齐效果。在实验过程中发现以下几个方面的问题:

(1) 如表4所示,在IBM4模型的基础上,用不同的窗口大小来训练模型,当窗口数增加到5时, $F$ 值基本保持不变,直至增加到13,也没有明显的过拟合问题。若实验中窗口的大小设置为1,即不考虑上下文信息,相对于设置一个大的窗口值,词语对齐效果欠佳。

(2) 本文模型中的DNN在查找层上只有三层隐层,在IBM4模型的基础上,分别将DNN中的隐层数设置1、2、3、4,然后训练DNN-HMM模型。表5为不同的隐藏层个数对应的 $F$ 值,当隐藏层数为3时, $F$ 值最大。

表5 不同的隐藏层个数对应的 $F$ 值  
Table 5  $F$ -value with different hidden layer number

IBM + NN	$F/\%$
1个隐层	53.76
2个隐层	56.72
3个隐层	57.34
4个隐层	57.28

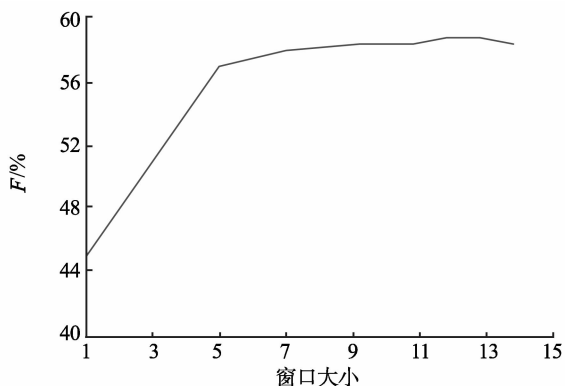


图4 不同窗口大小对应的 $F$ 值  
Fig. 4  $F$ -value with different window sizes

## 5 结论

为了实现汉-越双语词语自动对齐,本文提出了DNN-HMM词语对齐模型,将DNN融合到HMM模型中,利用DNN计算上下文信息的词汇翻译得分,通过人工构建的词语对齐语料开展相应实验。实验结果表明,该方法能有效提高汉-越词语对齐质量。对于下一步工作是在模型中设置不同的隐藏层参数进行训练,并进一步研究深度学习理论在词对齐领域的应用。

## 参考文献:

- [1] LE H P, HO T V. A maximum entropy approach to sentence boundary detection of Vietnamese texts[C]//IEEE International Conference on Research, Innovation and Vision for the Future-RIVF 2008. New York: IEEE, 2008:1-6.
- [2] HUYÊN N T M, ROUSSANALY A, VINH H T. A hybrid approach to word segmentation of Vietnamese texts[J]. Language and Automata Theory and Applications, 2008:240-249.
- [3] 越南语词法分析系统 [EB/OL]. [2014-11-12]. <http://www.loria.fr/~lehong/tools/vn-Tokenizer.php>.
- [4] BROWN P F, PIETRA V J D, PIETRA S A D, et al. The mathematics of statistical machine translation: parameter estimation [J]. Computational Linguistics, 1993, 19(2):263-311.
- [5] Franz Josef Och, Hermann Ney. A systematic comparison of various statistical alignment models[J]. Computational Linguistics, 2003, 29(1):19-51.
- [6] BLUNSOM P, COHN T. Discriminative word alignment with conditional random fields[C]//Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics. Philadelphia: Association for Computational Linguistics, 2006:65-72.
- [7] LIU Y, LIU Q, LIN S. Discriminative word alignment by linear modeling[J]. Computational Linguistics, 2010, 36(3):303-339.
- [8] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7):1527-1554.
- [9] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch[J]. The Journal of Machine Learning Research, 2011, 12:2493-2537.
- [10] NIEHUES J, WAIBEL A. Continuous space language models using restricted boltzmann machines[C]//Proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT). [S. l.]: [s. n.], 2012:1-48.
- [11] GOLDBERG Y, LEVY O. word2vec Explained: deriving Mikolov et al. s negative-sampling word-embedding method[EB/OL]. [2014-10-24]. <http://arxiv.org/pdf/1402.3722v1.pdf>.
- [12] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. [2014-10-24]. <http://arxiv.org/pdf/1301.3781.pdf>.
- [13] KLEMENTIEV A, TITOV I, BHATTARAI B. Nducing crosslingual distributed representations of words [C]//Proceedings of the International Conference on Computational Linguistics. Bombay: 2012 Organizing Committee, 2012.
- [14] ZHENG Xiaoqing, CHEN Haiyang, XU Tianyu. Deep learning for Chinese word segmentation and POS tagging[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2013:647-657.
- [15] DAHL G E, YU D, DENG L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition[J]., IEEE Transactions on Audio, Speech, and Language Processing, 2012, 20(1):30-42.
- [16] YANG Nan, LIU Shujie, LI Mu, et al. Word alignment modeling with context dependent deep neural network[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics. Somerset: ACL, 2013: 166-175.
- [17] BENGIO Y. Learning deep architectures for AI[M]. Now Publishers Inc Hanover, 2009.

(编辑:许力琴)