

文章编号: 1003-0077 (2017) 00-0000-00

## 基于隐含主题协同注意力网络的领域分类方法

黄培松 黄沛杰 丁健德 艾文程 章锦川

(华南农业大学 数学与信息学院, 广东省 广州市 510642)

**摘要:** 基于注意力机制的神经网络模型在文本分类任务中显示出了很好的效果。然而当训练数据的规模有限, 或者测试数据的分布与训练数据有较大差异时, 一些有价值的信息词很难在训练中被模型捕捉到。为此, 该文提出了一种新的基于协同注意力 (co-attention) 网络的领域分类方法。该文利用隐含主题模型学习隐含主题注意力, 并将其引入到文本分类常用的双向长短期记忆网络 (BiLSTM) 中, 与研究进展中的软或硬注意力 (soft- or hard-attention) 机制一起构成协同注意力。中文话语领域分类基准语料 SMP-ECDT 的实验结果表明, 隐含主题协同注意力网络取得了显著优于注意力机制研究进展的领域分类效果, 比基线注意力机制 Soft att、Hard att 以及单独的隐含主题注意力机制 BTM att 分别提高了 2.85%、1.85% 和 1.74% 的分类正确率。此外, 实验结果还验证了, 在额外的未标记数据上训练隐含主题, 可以进一步提高该文方法的领域分类性能。

**关键词:** 领域分类; 协同注意力; 隐含主题; BiLSTM; BTM

中图分类号: TP391

文献标识码: A

## Latent Topic-enriched Co-attention Networks for Domain Classification

HUANG Peisong, HUANG Peijie, DING Jiande, AI Wencheng and ZHANG Jinchuan

(College of Mathematics and Informatics, South China Agricultural University, Guangzhou, Guangdong 510642, China)

**Abstract:** Attention-based bidirectional long short-term network models have recently shown promising results in text classification tasks. However, when the amount of training data is restricted, or the distribution of the test data is quite different from the training data, some potential informative words maybe hard to be captured in training. In this work, we propose a new method to learn co-attention for domain classification. Unlike the past attention mechanisms only guided by domain tags of training data, we explore using the latent topics in the data set to learn topic attention mechanism, and employ it for bidirectional long short-term memory network (BiLSTM). Then the co-attention is obtained by combining the topic attention and the network attention. Experiments on the SMP-ECDT benchmark corpus show that the proposed co-attention attention mechanism outperforms the state-of-the-art soft mechanism, hard attention mechanism and topic attention mechanism in domain classification, with 2.85%, 1.85% and 1.74% classification accuracy improvement, respectively. Moreover, experiment result shows that the proposed method can be trained on additional unlabeled data and further improve the domain classification performance.

**Key words:** domain classification; co-attention; latent topic; BiLSTM; BTM

收稿日期: ; 定稿日期: =

基金项目: 国家自然科学基金(71472068); 广东省大学生创新训练计划项目(201810564094)

## 0 引言

口语语言理解(spoken language understanding, SLU)的应用在我们的日常生活中变得越来越重要<sup>[1]</sup>。在使用基于人机对话<sup>[2]</sup>的应用程序时,人可能有各种各样的意图,例如聊天、提问、预订机票、询问天气等。因此,在接收到来自用户的输入消息(文本输入或语音识别结果)后,系统第一步要将用户的话语分类到特定的领域中,以便进行进一步处理。例如,用户话语“今天福州的天气怎么样?”应该首先分类到“询问天气”领域。领域分类可以看作是一个话语语义分类问题,可以采用支持向量机(support vector machines, SVM)<sup>[3]</sup>和深度神经网络(deep neural network)方法<sup>[4]</sup>等常用分类器。尤其是递归神经网络(recurrent neural network, RNN)模型<sup>[5-6]</sup>由于易于处理变长输入话语,已成功地应用于许多序列学习问题中。

最近的研究方法引入了注意力机制(attention mechanisms)<sup>[7-11]</sup>,将模型的注意力集中在信息词上。然而,据我们所知,在所有现有的注意力方法中,注意力的参数向量都是在领域标签的引导下训练的。当训练数据的规模有限,或者测试数据的分布与训练数据有较大的差异时,一些潜在有价值的信息词在训练中很难被模型捕捉到。

将文本建模为若干隐含主题的构成是一种以无监督方式推断语义的方法。利用 LSA (latent semantic analysis)<sup>[13]</sup>、pLSA (probabilistic LSA)、LDA (latent Dirichlet allocation)<sup>[14]</sup>、BTM(biterm topic model)<sup>[15]</sup>等主题建模技术,可以通过一组具有代表性的词(主题)<sup>[16]</sup>成功地推断出了文本的高层含义。在文本分类中,许多研究采用主题模型挖掘隐藏主题作为分类特征<sup>[17-18]</sup>。与他们不同的是,本文的前期工作<sup>[12]</sup>利用隐含主题设计一种新的注意力机制,取得了优于主流的软注意力<sup>[7-9]</sup>和硬注意力<sup>[11]</sup>的性能。在进一步的研究中,我们发现隐含主题注意力与主流的软、硬注意力机制的关注点存在一定的互补性。因此,本文将主题模型获取隐含主题注意力与带注意力机制的神经网络中的网络注意力一起构成协同注意力,帮助模型挖掘和关注更多的有价值的信息词。在中文话语领域分类基准语料 SMP-ECDT 的实验表

明,本文提出的新的协同注意力机制取得了优于研究进展注意力机制的性能。

本文的其余部分组织如下。在第 2 节,我们介绍了相关的基础工作,包括注意力机制和 BTM 主题模型。在第 3 节,我们介绍了所提出的方法。第 4 节讨论了验证本文方法效果的 SMP-ECDT 基准测试任务的实验设置和实验结果。第 5 节总结了本文的工作。

## 1 基础工作

### 1.1 注意力机制

近年来,各种注意机制在文本分类问题中得到了广泛的应用。这里,我们考虑两种流行的选择,软注意力(soft attention)<sup>[7-9]</sup>和硬注意力(hard attention)<sup>[11]</sup>机制。

软注意力在文本分类中得到了广泛的应用,如意图检测<sup>[7]</sup>、关系分类<sup>[8]</sup>和文档分类<sup>[9]</sup>。虽然这些工作的得分函数的计算存在一定的差异,但其机制都是由编码器的输出来计算一个“软”对齐的注意力权重,然后对编码器的输出根据注意力的权重进行放缩后叠加:

$$u_i = \text{score}(h_i), \quad (1)$$

$$\alpha_i = \frac{\exp(u_i)}{\sum_l \exp(u_l)}, \quad (2)$$

$$h^* = \sum_l \alpha_l h_l, \quad (3)$$

其中 score 为前馈神经网络。我们首先通过一个单层感知机根据 LSTM 输出  $h_i$  来计算得到  $u_i$ , 其中  $u_i$  作为  $h_i$  的得分表示。然后对  $u$  进行归一化得到权重,根据权重  $(\alpha_1, \alpha_2, \dots, \alpha_L)$  对 LSTM 输出  $(h_1, h_2, \dots, h_L)$  进行加权求和得到  $h^*$ 。

另一种代表性的注意力机制的硬注意力<sup>[11]</sup>,它将每个输出精确地对齐到一个输入状态,但需要复杂的训练来教会网络选择该状态:

$$p(s_i = 1 | \alpha) = \alpha_i, \quad (4)$$

$$h^* = \sum_l s_l \alpha_l h_l, \quad (5)$$

其中  $\alpha_i$  是第  $i$  个 LSTM 时间步的输出被选中的概率。 $s$  是一个 one-hot 变量,当第  $i$  个 LSTM 时间步的输出被选中时,它对应的  $s$  就被设置为 1,其它 LSTM 时间步对应的  $s$  就被设置为 0。

虽然上述的代表性注意力机制在已有的文本分类任务中都取得较好的性能,但它们都是通过分类标签引导训练的。当带标注的训练数据规模有限时,现有的注意机制可能难以捕捉到一些潜在有价值的信息词。

## 1.2 BTM 主题模型

Biterm 主题模型(BTM)是 Yan 等人<sup>[15]</sup>提出的。Yan 等人利用 BTM 模型训练了一些文本特征,并利用文本特征进行文本分类。图 1 显示了 BTM 的图形表示。

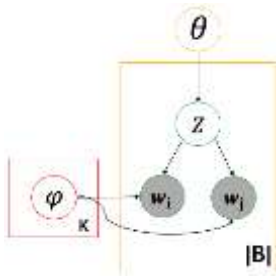


图 1 BTM 图示

BTM 的核心思想是通过短文本学习主题。狄利克雷先验假设 $\beta, \gamma$ 。BTM 中语料库的具体生成过程可以描述为:

1. 对每个主题  $z$ , 都会拥有一个特定主题词分布  $\phi_z \sim \text{Dir}(\beta)$ 。
2. 计算所有文档的主题分布  $\theta \sim \text{Dir}(\gamma)$ 。
3. 对于词对集合  $B$  中的每个词对  $b$ :
  - (a) 根据  $\theta$  的概率抽取出主题  $z \sim \text{Multi}(\theta)$ ,
  - (b) 再根据这个主题  $z$  下  $\phi_z \sim \text{Dir}(\beta)$  中得到各个词对的概率下  $\phi_z$ 。抽取词对的两个词:  $w_i, w_j \sim \text{Multi}(\phi_z)$ 。

按照上述步骤, 一个词对  $b = (w_i, w_j)$  的联合概率为:

$$P(b) = \sum_z P(z)P(w_i|z)P(w_j|z) = \sum_z \theta_z \phi_{i|z} \phi_{j|z}. \quad (6)$$

因此, 整个语料库的可能性为:

$$P(B) = \prod_{(i,j)} \sum_z \theta_z \phi_{i|z} \phi_{j|z}. \quad (7)$$

通过使用 BTM, 我们抽样可以得到主题下单词的概率。此外, BTM 可以充分利用没有标记的

数据, 因为它是一个无监督模型。在我们之后的方法中我们将充分利用 BTM 的特点, 将 BTM 引入注意力机制。

## 2 本文的方法

在本节中, 我们将介绍如何将主题模型学到的注意力权重和神经网络的注意力权重相融合, 构成系统注意力网络。

### 2.1 隐含主题注意力

主题模型目的在于构建隐主题-词和隐主题-话语的分布。它将隐主题作为词和话语的联系的隐变量。在我们的前期工作中<sup>[12]</sup>, 我们利用主题模型构建出的主题-词关系来计算得到每个词的权重。

首先, 根据下面公式可以容易的在训练好的词对主题模型中通过统计估计出隐主题-词分布<sup>[15]</sup>:

$$\phi_{w|z} = \frac{n_{w|z} + \delta}{\sum_w n_{w|z} + M\delta}, \quad (8)$$

其中  $M$  是词的总数量,  $n_{w|z}$  代表词  $w$  被分配到主题  $z$  的次数,  $\delta$  是一个超参数。

接着, 我们通过每个单词在不同主题下的概率来衡量这个单词在当前语料中的重要性。比如当一个单词在不同主题下的概率相差不大时, 我们就认为这个单词在语料中的重要性是比较低的。而当一个单词在不同主题下概率相差比较大时, 我们就认为这个单词在语料中的重要性是比较高的。由于词表太大, 会导致各个单词的概率相差过小, 于是我们改用该单词的概率在主题下的排名来表示。我们用  $c_{ij}$  来表示词  $i$  在主题  $j$  下排名。接着我们使用极差来表示单词在语料中的重要性, 也就是词的权重  $d_i$ :

$$d_i = \max(c_{i1}, c_{i2}, \dots, c_{ij}) - \min(c_{i1}, c_{i2}, \dots, c_{ij}). \quad (9)$$

我们使用 softmax 对输入的词权重进行归一化:

$$\beta = \text{softmax}(d_i, \dots, d_L). \quad (10)$$

记归一化后的词权重为  $\beta = \{\beta_1, \dots, \beta_L\}$ , 作为隐含主题注意力, 其中  $L$  是句子的长度。

## 2.2 协同注意力网络结构

本文的模型是在基于注意力的 BiLSTM<sup>[8]</sup>基础上融入主题注意力构成的协同注意力网络, 由词向量编码层、主题模型层、BiLSTM 网络层、注意力层和 softmax 分类层组成, 如图 2 所示。

(1) 在词向量编码层, 我们用预先训练好的词向量对词进行编码。给一个固定长度为  $L$  的短句  $U = \{x_1, x_2, \dots, x_L\}$ , 每个单词  $x_i$  会在词向量编码层被编码成词向量  $e_i$ 。这样一个固定长度为  $L$  的短句  $U = \{x_1, x_2, \dots, x_L\}$  就被编码成了  $\text{emb}_u = \{u_1, u_2, \dots, u_L\}$ 。

(2) 网络编码层, 我们采用在语音识别<sup>[19]</sup>和口语语言理解<sup>[1]</sup>领域有着非常成功的应用的 BiLSTM<sup>[20]</sup>。在我们的实验中, 我们将  $\text{emb}_u = \{u_1, u_2, \dots, u_L\}$  输入到双向的长短时记忆网络, 然后分别从网络中得到正向的输出

$\vec{h} = \{\vec{h}_1, \dots, \vec{h}_L\}$  和反向的输出  $\overleftarrow{h} = \{\overleftarrow{h}_1, \dots, \overleftarrow{h}_L\}$ 。我们将正向和反向的结果拼接起来  $h_i = [\vec{h}_i, \overleftarrow{h}_i]$  这样  $\text{emb}_u = \{u_1, u_2, \dots, u_L\}$  经过这一层网络就得到输出  $h = \{h_1, \dots, h_L\}$ 。

(3) 主题模型层中, 我们通过采用 BTM 主题模型, 计算出隐主题-词分布。

(4) 在注意力层中, 我们分别计算来自于 BiLSTM 的网络注意力以及基于隐含主题的网络注意力。网络注意力可以采用不同的机制得到, 如 1.1 小节介绍的软注意力和硬注意力机制, 得到的注意力权重记为  $\alpha = \{\alpha_1, \dots, \alpha_L\}$ 。

另一方面, 我们在隐主题-词分布的基础上得到输入话语的词权重  $d = \{d_1, \dots, d_L\}$ , 并进一步使用 softmax 将其归一化得到隐含主题注意力  $\beta = \{\beta_1, \dots, \beta_L\}$  (详见前面 2.1 小节)。

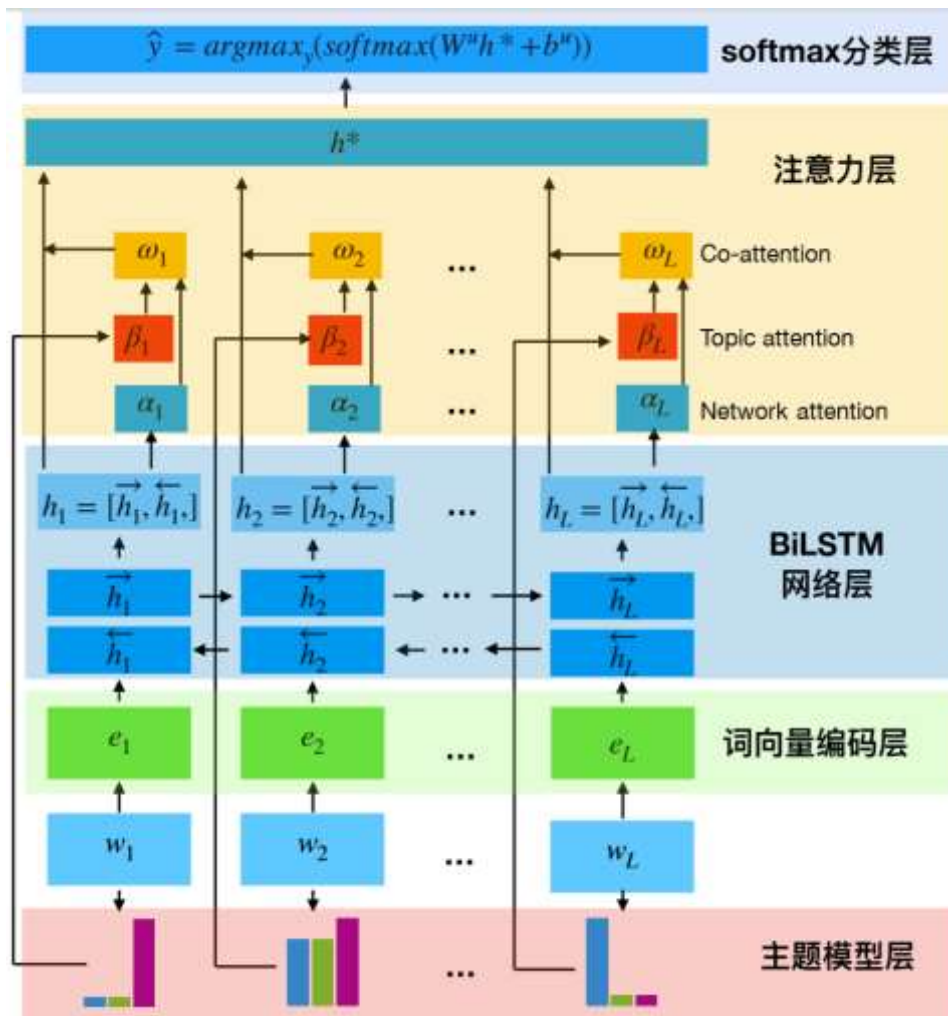


图 2 协同注意力网络

最后使用加法将这两种注意力权重进行融合, 记为 $\omega_i$ :

$$\omega_i = \alpha_i + \beta_i. \quad (11)$$

接着我们按照下面公式

$$h^* = \sum_{i=1}^L \omega_i * h_i. \quad (12)$$

利用它们对应的注意力权重对 BiLSTM 的输出进行放缩, 并将所有放缩后的结果累加得到 $h^*$ 作为注意力层的输出。

(5) 在 softmax 分类层, 我们使用 softmax 分类器来输出各个类别的概率。其中这层网络的公式是:

$$Y = \text{softmax}(W^U * h^* + b^U). \quad (13)$$

预测的时候我们输出概率最大的类别:

$$y^* = \text{argmax}(Y). \quad (14)$$

另外在训练过程中我们使用的是交叉熵作为代价函数。

### 3 实验

#### 3.1 数据集

领域分类可以看作是一个短文本分类问题, 本文的语料是对单个话语的领域分类, 采用的是中文口语领域分类基准语料 SMP-ECDT<sup>[21]</sup>。该语料由哈尔滨工业大学社会计算与信息检索研究中心(哈工大 SCIR)和科大讯飞股份有限公司(iFLYTEK)提供。具体上采用的是 2018 年第二届中文人机对话技术评测任务一“中文话语领域分类”中的数据集, 分为面向闲聊和面向任务两大类。同时, 面向任务的对话还包含 30 个子类, 构成了 31 类的分类任务。该语料库包含 3736 项训练数据和 4528 项测试数据项。

#### 3.2 实验设置

我们将训练数据集切分成 10 折进行交叉验证, 并通过网格搜索(grid search)方式对模型进行训练, 使交叉熵损失函数最小化, 选择最优超参数。LSTM num\_units 的范围是[50, 100, 150, 300], 输

入 LSTM 的 dropout 输入 keep prob 的范围是[0.2, 0.3], 状态 keep prob 的范围是[0.2, 0.3], 输出 keep prob 被设置为 0.3。BTM 的主题数选择包括 5、10、20、30 和 50。批次训练的大小为 32。此外, 我们使用 CBOW 预训练词向量<sup>[22]</sup>, 并按照文献[23]的参数设置应用 Adam 优化方法。实验结果取了 5 次独立实验的平均值。

本文提出的协同注意力网络模将于以下代表性的基线方案进行比较, 包括:

- BiLSTM: 该方法是文本分类的经典基线<sup>[20, 24]</sup>, 非常适合于序列问题。

- Soft att: Liu 等人<sup>[7]</sup>在 ATIS 数据集的意图识别任务采用了软注意力。我们在领域分类的 BiLSTM 模型上应用了他们的软注意力机制。

- Hard att: 我们在领域分类的 BiLSTM 模型上应用了文献<sup>[11]</sup>的硬注意力机制。

- BTM att: 在领域分类的 BiLSTM 模型上单独应用隐含主题注意力机制<sup>[12]</sup>。

本文的方案包括:

- Soft+BTM att: 融合隐含主题注意力和软注意力的协同注意力机制。

- Hard+BTM att: 融合隐含主题注意力和硬注意力的协同注意力机制。

### 3.3 结果和分析

#### 3.3.1 主题选择数量

我们比较了协同注意力网络模型 Soft+BTM att 和 Hard+BTM att 在不同主题数量下的交叉验证情况, 如图 3 所示。

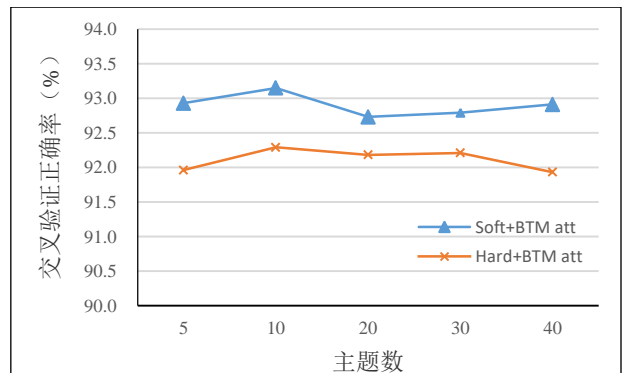


图 3 不同主题数的验证正确率

在实验中,两个协同注意力网络模型均在主题数 10 获得了最佳的交叉验证性能。

### 3.3.2 整体性能

然后,我们将提出的潜在主题 att 模型与几种基线方法进行比较,包括 BiLSTM 和带软、硬注意力机制的 BiLSTM。实验结果如表 1 所示。

从表 1 可以看到,两种当前主流的注意力机制都能在 BiLSTM 模型的基础上有效地提升分类性能,其中 Hard att 在这个数据集上的效果优于 Soft att。而单独的隐含主题注意力机制 BTM att 也可以取得略微优于 Hard att 的效果。两种协同注意力方案均取得优于研究进展注意力机制的领域分类效果,其中 Hard+BTM att 方案取得了最高分类正确率,达到 80.83%,比基线注意力机制 Soft att、Hard att 和 BTM att 分别提高了 2.85%、1.85%和 1.74%的分类正确率。

表 1 不同模型的性能对比

模型	正确率
BiLSTM <sup>[20, 24]</sup>	76.97%
Soft att <sup>[7]</sup>	77.98%
Hard att <sup>[11]</sup>	78.97%
BTM att <sup>[12]</sup>	79.09%
Soft+BTM att	80.36%
Hard+BTM att	80.83%

### 3.3.3 注意力的可视化

前面的实验结果表明了本文提出的模型的有效性。为了更为直观地观察我们的方案如何借助主题模型挖掘更多的有利于分类任务的信息词,我们通过案例进一步观察不同方案的注意力层的情况。我们从测试数据中选择一个标记为 calc (计算领域)的示例话语,不同方案的注意力情况如图 4 所示。



图 4 不同模型的注意力可视化对比例子

从图 4 可以看到,Soft att 和 Hard att 机制分别在一定程度上关注了“等于”和“多少”,Soft att 还微弱地关注了“平方”。此外,它们都很强烈地关注了“20”。20 作为一个数字,在 calc 领域有一定信息价值,但语料中同时也有不少领域含有数字,过于关注数字有时会导致领域分类错误。而我们的 BTM att 重点关注了“平方”和“等于”,还稍微关注了“算”。这使得协同注意力机制 Soft+BTM att 和 Hard+BTM att 都较为全面地和强烈地关注了“算”、“平方”、和“等于”等 calc 领域相对特有的信息词,从而有助于模型更稳定地进行正确的领域分类。对 calc 领域的训练语料的进一步分析我们发现,造成 Soft att 和 Hard att 机制对“算”和“平方”等信息词关注不够的原因确实来自于训练数据的规模有限,这些词出现的频率并不是很高。由于是口语话语,存在大量的省略和口语化表述,在 calc 领域,有许多话语如“54 乘上 45”、“根号 3 是多少”等,并没有像图 3 选取的例子这样有着较为完整的表述。而我们的模型采用了有别于分类标签引导的视角,从而挖掘出额外的一些具有领域分类价值的信息词。并且,通过这个例子,我们还可以看到,训练得到的协同注意力网络,确实达到了融合两方面注意力的效果。

### 3.3.4 额外未标记数据训练隐含主题的效果

由于 BTM 在隐含主题计算方面是一个无监督模型,我们研究该方法是否可以用额外的未标记数据进行隐含主题的计算,进一步改善隐含主题注意力的质量,进而提高领域分类性能。为此,我们削减了 4/5 的测试集,并使用它作为额外的未标记的隐含主题训练数据,剩下的 1/5 的测试

集用于测试。实验结果如表 2 所示。

表 2 采用额外未标记数据 (4/5 测试集) 计算隐含主题的效果 (测试用 1/5 测试集)

训练数据	BTM att	Soft + BTM att	Hard+ BTM att
训练集	79.29%	80.22%	80.02%
训练集+4/5 测试集	79.90%	80.45%	80.52%

从表 2 可以看到, 只采用训练集计算隐含主题时, BTM att、Soft+BTM att、Hard+BTM att 三个方案在 1/5 测试集上的分类正确率分别是 79.29%、80.22% 和 80.02%。采用额外的未标记数据(4/5 测试集)之后, 三个方案在 1/5 测试集上的分类正确率都有了一定的提升, 平均提升了 0.5% 左右。

## 4 结束语

本文提出了一种新的基于协同注意力网络的领域分类方法。利用隐含主题模型学习隐含主题注意力并与神经网络的注意力共同构成协同注意力, 以挖掘更多对领域分类有价值的信息词。中文话语领域分类基准语料 SMP-ECDT 的实验结果表明, 单独的隐含主题注意力机制就可以取得略微优于研究进展注意力机制的效果, 而隐含主题协同注意力网络则取得了优于研究进展注意力机制的领域分类效果。此外, 我们还通过实验验证了, 在额外的未标记数据上训练隐含主题, 可以进一步提高本文方法的领域分类性能。未来工作主要包括在更多预训练方法 (如 BERT、ELMO 等) 得到的词 (或者字) 向量上验证本文方法的效果, 以及将本文的方法应用到更多的不同的语料和任务上。

## 参考文献

- [1] P. Xu and R. Sarikaya. Contextual domain classification in spoken language understanding systems using recurrent neural network [C]. In Proceedings of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014), pp. 136–140.
- [2] 俞凯, 陈露, 陈博等. 任务型人机对话系统中的认知技术——概念、进展及其未来 [J]. 计算机学报, 2015, 38 (12):2333-2348.
- [3] P. Haffner, G. Tur, and J. H. Wright. Optimizing SVMs for complex call classification [C]. In Proceedings of the 28th IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP 2003), pp. I-632-I-635.
- [4] R. Sarikaya, G. E. Hinton, and B. Ramabhadran. Deep belief nets for natural language call-routing [C]. In Proceedings of the 36th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011), Prague, Czech Republic, pp. 5680–5683.
- [5] S. Ravuri and A. Stoicke. A comparative study of neural network models for lexical intent classification [C]. In Proceedings of the 5th IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2015), pp. 368–374.
- [6] 柯子烜, 黄沛杰, 曾真. 基于优化“未定义”类话语检测的话语领域分类 [J]. 中文信息学报, 2018, 32(4): 105-113.
- [7] B. Liu and I. Lane. Attention-based recurrent neural network models for joint intent detection and slot filling [C]. In Proceedings of the 17th Annual Conference of the International Speech Communication Association (INTERSPEECH 2016), pp. 685–689.
- [8] P. Zhou, W. Shi, J. Tian, et al. Attention-based bidirectional long short-term memory networks for relation classification [C]. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016), pp. 207–212.
- [9] J. Ive, G. Gkotsis, R. Dutta, et al. Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health [C]. In Proceedings of the 5th Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic (CLPsych 2018), pp. 69–77.
- [10] Y. Kim, D. Kim, A. Kumar. Efficient large-scale neural domain classification with personalized attention [C]. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018), pp 2214–2224.
- [11] S. Shankar, S. Garg, and S. Sarawagi. Surprisingly easy hard-attention for sequence to sequence learning [C]. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP 2018), pp. 640–645.
- [12] P. S. Huang, P. J. Huang, W. C. Ai, et al. Latent topic attention for domain classification. [C]. In Proceedings of the 20th Annual Conference of the International Speech Communication Association (INTERSPEECH 2019).
- [13] S. Deerwester, S. T. Dumais, G. W. Furnas, et al. Indexing by latent semantic analysis [J] Journal of the American Society for Information Science, 1990, 41(6): 391–407.
- [14] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993–1022.
- [15] X. Yan, J. Guo, Y. Lan, et al. A biterm topic model for short texts [C]. In Proceedings of the 22nd International Conference on World Wide Web (WWW 2013), pp. 1445–1456.
- [16] V. K. R. Sridhar. Unsupervised topic modeling for short texts using distributed representations of words [C]. In

Proceedings of the 1st workshop on vector space modeling for natural language processing (VSM-NLP 2015), pp. 192–200.

- [17] C. Li, H. Wang, Z. Zhang, et al. Topic modeling for short texts with auxiliary word embeddings [C]. In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2016), pp. 165–174.
- [18] A. Budhkar and F. Rudzicz. Augmenting word2vec with Latent Dirichlet Allocation within a Clinical Application [C]. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), pp. 4095–4099.
- [19] A. Graves, N. Jaitly, and A. Mohamed. Hybrid speech recognition with deep bidirectional LSTM [C]. In Proceedings of the 4th IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU 2013), pp. 273–278.
- [20] M. Schuster and K. K. Paliwal. Bidirectional recurrent neural networks [J]. *Signal Processing IEEE Transactions on Signal Processing*, 1997, 45(11): 2673–2681.
- [21] W.-N. Zhang, Z. Chen, W. Che, et al. The first evaluation of Chinese human-computer dialogue technology. arXiv:1709.10217 [cs], Sep. 2017.
- [22] T. Mikolov, K. Chen, G. Corrado, et al. Efficient estimation of word representations in vector space [C]. In Proceedings of the 1st International Conference on Learning Representations (ICLR 2013), pp. 1–12.
- [23] D. P. Kingma and J. Ba. Adam: a method for stochastic optimization [C]. In Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015), pp. 1–13.
- [24] N.T. Vu, P. Gupta, H. Adel, et al. Bi-directional recurrent neural network with ranking loss for spoken language understanding [C]. In Proceedings of the 41th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), pp. 6060–6064.



丁健德 (1999—), 本科, 主要研究领域为自然语言处理。

Email: bighead@stu.scau.edu.cn



黄培松 (1996—), 本科, 主要研究领域为自然语言处理。

Email: bringtree@stu.scau.edu.cn



黄沛杰 (1980—), 通信作者, 博士, 副教授, 主要研究领域为人工智能、自然语言处理、口语对话系统。

Email: pjhuang@scau.edu.cn