

文章编号:1007-5321(2019)03-0021-08

DOI:10.13190/j.jbupt.2018-289

基于胶囊的英文文本蕴含识别方法

朱 皓, 谭咏梅

(北京邮电大学 智能科学与技术中心, 北京 100876)

摘要: 提出了一种基于胶囊的英文文本蕴含识别方法. 分别为每一种蕴含关系构建一个胶囊, 用于模拟此蕴含关系的识别, 并将其指定为该胶囊的属性. 给定两段文本, 经过 highway 编码层和序列编码层获取语义表示, 分别输入胶囊中, 依次通过其内部的交互模块、比较模块和聚合模块. 交互模块利用交互注意力机制提取文本间的局部交互特征, 比较模块和聚合模块使用前馈神经网络进行语义信息比较和聚合. 最后对所有胶囊的输出归一化, 得到两段文本的蕴含关系. 该方法在 SNLI 测试集上的准确率为 89.2%, 在 MultiNLI 匹配测试集和不匹配测试集上的准确率分别为 77.4% 和 76.4%. 对交互模块中注意力关系矩阵的可视化分析结果验证了胶囊在英文文本蕴含识别任务中的有效性.

关键词: 文本蕴含识别; 胶囊; 交互注意力机制

中图分类号: TN929.53

文献标志码: A

English Textual Entailment Recognition Using Capsules

ZHU Hao, TAN Yong-mei

(Intelligence Science and Technology Center, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: An English textual entailment recognition method using capsules is presented. This method builds a capsule for each relationship to model the recognition of this relationship and assigns it as the attribute of the capsule. Given two texts, they are first encoded by highway encoding layer and sequence encoding layer to obtain the semantic representations, and then fed into each capsule, passing through its internal interaction module, comparison module and aggregation module in turn. The interaction module uses the inter-attention mechanism to extract local interactive features between texts. The comparison module and the aggregation module use the feedforward network to compare and aggregate the semantic information. Finally, the output of all capsules is normalized to obtain the relationship between the two texts. The accuracy on SNLI test dataset is 89.2%, the accuracy on MultiNLI matched and mismatched test dataset is 77.4% and 76.4%. The visual analysis of the attentional relationship matrix of interaction module also verifies the effectiveness of capsules in the English textual entailment recognition task.

Key words: textual entailment recognition; capsules; inter-attention mechanism

文本蕴含识别(RTE, recognizing textual entailment)是自然语言处理中一项基础而又极具挑战性的任务,其目标在于判断给定的两段文本前提 P 和假设 H 之间是否具有蕴含关系. 文本蕴含关系的识别在自然语言处理中有很多重要的应用,如关系抽

取、问答系统以及文档自动摘要等.

目前研究人员在英文文本蕴含识别任务上做了大量有益的工作,其中包括大规模数据集 SNLI (Stanford natural language inference)^[1] 和 MultiNLI (multi-genre natural language inference)^[2] 的创建,这

使得训练深度神经网络模型变得可行. 许多基于神经网络的英文文本蕴含识别方法也因此取得比较好的效果. 随着注意力机制成功引入自然语言处理领域,它能够通过对文本表示进行软对齐,构建文本表示之间的注意力关系矩阵,而不需要考虑文本的长度. 现有大多数性能良好的文本蕴含识别方法也是基于该核心表示学习技术^[3-4],它通过将两段文本中的单词进行软对齐,并比较它们之间的关系来学习软对齐表示. 但是,在判断蕴含关系时,不同类别对该表示的侧重点不同,仅用一个注意力关系矩阵来表示对齐结果会受到很大限制.

为了解决单个软对齐表示所带来的局限性问题,将胶囊引入英文文本蕴含识别任务中,结合交互注意力机制提出一种基于胶囊的英文文本蕴含识别方法. 胶囊是一组具有丰富意义的神经元^[5-6],不同的胶囊能够针对不同的蕴含关系来学习相应的对齐表示,通过比较和聚合各个注意力关系矩阵的特征信息来判断最终的蕴含关系.

1 相关工作

在早期,由于英文文本蕴含识别数据集的规模都比较小,大部分研究工作是基于传统的统计学习方法. 近年来,随着大规模人工标注数据集 SNLI 的发布和深度学习技术的发展,大量基于神经网络的英文文本蕴含识别方法涌现,这些方法大致可以归纳为 2 类:基于文本编码的方法和基于交互特征的方法.

基于文本编码的文本蕴含识别方法通常使用编

码器将文本转换为固定长度的向量表示,然后通过拼接、内积和作差等操作对向量进行比较,判断它们之间的蕴含关系^[7-8]. 该类方法有 2 个明显的缺陷,一方面由于文本长度不尽相同,固定长度的向量对不同文本的表达能力也不同,对于某些长度较长的文本,可能无法将文本完整的语义信息编码到向量中;另一方面,该类方法忽略了文本之间的局部交互,而这些交互特征信息在文本蕴含识别过程中是非常有必要的.

基于交互特征的文本蕴含识别方法利用注意力机制提取文本之间的局部交互信息,然后通过聚合交互特征信息来预测最终的蕴含关系^[9-11]. 通常该类方法将两段文本中的单词进行软对齐,并学习它们之间的对齐表示,但是仅学习一个对齐表示对于文本蕴含识别这种多分类任务是有限的,无法从每种蕴含关系的角度提取相关的特征信息.

综上所述,在英文文本蕴含识别任务中,针对不同类别的蕴含关系分别构建胶囊,使每个胶囊能够刻画其对应蕴含关系的特征信息,进而更准确地判断蕴含关系.

2 基于胶囊的英文文本蕴含识别方法

基于胶囊的英文文本蕴含识别模型由嵌入层、编码层、胶囊层和输出层 4 个部分组成,如图 1 所示.

模型的输入为两段文本,即前提 $P = (w_1, w_2, \dots, w_p)$ 和假设 $H = (w_1, w_2, \dots, w_h)$,其中: p 和 h 分别为前提和假设的长度.

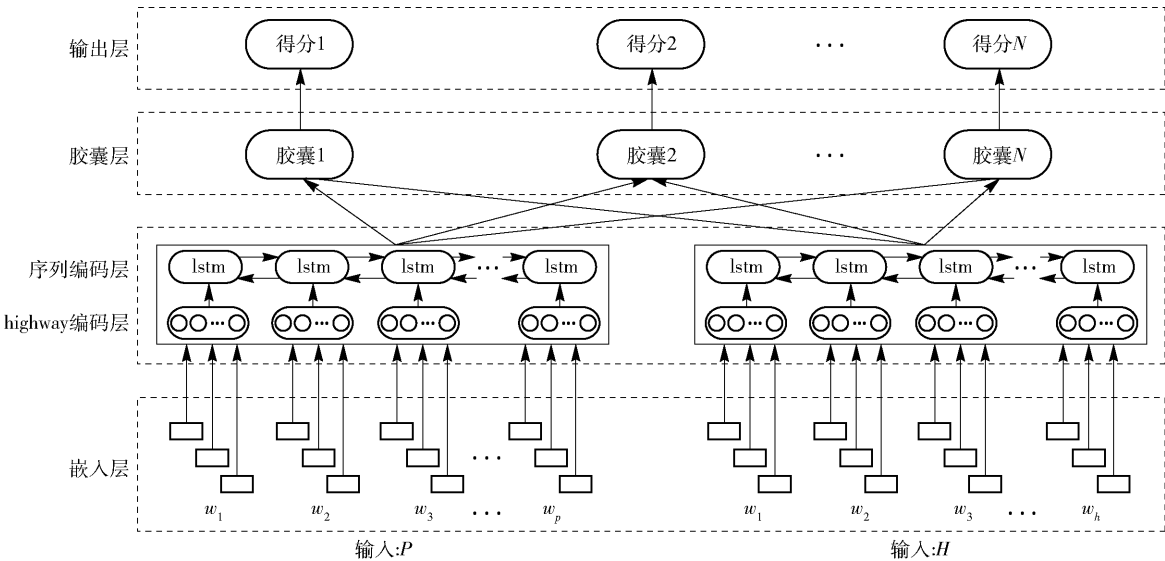


图 1 基于胶囊的英文文本蕴含识别方法框架

2.1 嵌入层

为了获取丰富的特征信息,嵌入层分别从单词和字符层面对单词进行表示^[12]. 单词层面的信息包括2部分,首先利用预训练的词嵌入模型,将单词映射到高维向量空间,获取每个单词的词向量;然后将每个单词的词性表示为固定维度的独热向量. 在字符层面,嵌入层将随机初始化的字符向量作为卷积神经网络的输入,结合最大池化来获取单词的字符特征. 通过拼接词向量、字符特征和词性向量得到每个单词的向量表示,进而可以获得前提的向量表示 $P \in \mathbf{R}^{p \times d}$ 和假设的向量表示 $H \in \mathbf{R}^{h \times d}$, 其中: d 表示单词向量表示的维度.

2.2 编码层

1) Highway 编码层

随着神经网络的深度不断增加,在训练过程中梯度信息回流受阻,导致模型的训练难度越来越大. 为了解决模型训练困难的问题,编码层首先将 P 和 H 分别通过一个2层的 highway 网络^[13], 依次对其进行编码. 该网络基于一种可学习的门限机制,可以自适应地控制信息流的传输,定义如下:

$$y = H(x, W_H)T(x, W_T) + xC(x, W_C) \quad (1)$$

其中: x 为 highway 网络的输入; $H(\cdot)$ 为使用 ReLU 激活函数的仿射变换; $T(\cdot)$ 和 $C(\cdot)$ 为2个使用 Sigmoid 激活函数的非线性变换,分别称为转换门和携带门. 为了计算方便,通常将 $C(\cdot)$ 定义为 $1 - T(\cdot)$, 即

$$y = H(x, W_H)T(x, W_T) + x(1 - T(x, W_T)) \quad (2)$$

W_H 、 W_T 和 $W_C \in \mathbf{R}^{d \times r}$ 为仿射变换和非线性变换对应的权重. highway 编码层的输出 $P^{hw} \in \mathbf{R}^{p \times r}$ 和 $H^{hw} \in \mathbf{R}^{h \times r}$ 即为前提和假设新的向量表示, r 为 highway 网络隐藏层的单元数.

2) 序列编码层

循环神经网络(RNN, recurrent neural network)是一种人工神经网络,其内部的记忆单元对于处理序列数据有着天然的优势,能够捕捉序列数据的长期依赖信息,被广泛应用到各种自然语言处理任务. 但是标准的循环神经网络存在梯度消失和梯度爆炸的缺点,为了克服该问题,充分获取文本的上下文信息,在序列编码层中使用双向长短时记忆神经网络(BiLSTM, bidirectional long short-term memory)分别对 P^{hw} 和 H^{hw} 进行语义信息提取.

$$\vec{h}_i, \vec{c}_i = \vec{L}(x_i, \vec{h}_{i-1}, \vec{c}_{i-1}) \quad (3)$$

$$\overleftarrow{h}_i, \overleftarrow{c}_i = \overleftarrow{L}(x_i, \overleftarrow{h}_{i-1}, \overleftarrow{c}_{i-1}) \quad (4)$$

其中: $L(\cdot)$ 表示单向长短时记忆网络(LSTM, long short-term memory). 在 t 时刻, BiLSTM 以 x_t 为输入,从2个方向计算当前的隐藏层状态 h_t 和记忆单元状态 c_t ,通过拼接这2个方向的隐藏层状态 \vec{h}_t 和 \overleftarrow{h}_t 得到其输出向量,即 $h_t = [\vec{h}_t; \overleftarrow{h}_t]$. 前提和假设的语义信息可表示为 $P^{LSTM} \in \mathbf{R}^{p \times 2u}$ 和 $H^{LSTM} \in \mathbf{R}^{h \times 2u}$, u 为单向长短时记忆神经网络的隐藏层单元数.

2.3 胶囊层

由于该任务中包含多种蕴含关系,对于利用交互注意力机制学习得到的软对齐表示,不同的蕴含关系对其的侧重点不同,仅从单个软对齐表示中提取文本间的交互特征会受到很大限制. 通过在胶囊层中构建不同的胶囊,对应不同的蕴含关系,进而每个胶囊能够学习其对应蕴含关系的交互特征信息.

在胶囊层中,胶囊的数量与所要识别的蕴含关系数量相同,每个胶囊对应一种特定的蕴含关系,并将此蕴含关系指定为相应胶囊的属性.

每个胶囊内部包括3个模块:交互模块、比较模块和聚合模块,胶囊的网络结构如图2所示.

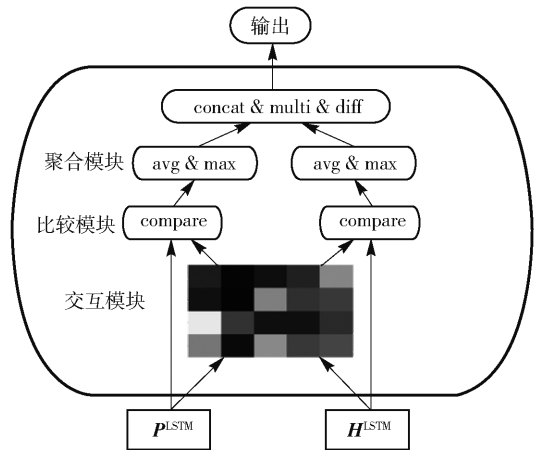


图2 胶囊网络结构

1) 交互模块

文本蕴含关系的识别通常依赖于对文本间语义关系的分析. 由于英文中语法结构的多样性和复杂性,传统的基于硬对齐的语义关系分析存在很多缺陷,例如句子成分分析错误,句型结构过于复杂难以对齐等. 所以该模块基于交互注意力机制对 P^{LSTM} 和 H^{LSTM} 进行软对齐,一定程度上避免了上述问题. 首先构建1个 P^{LSTM} 和 H^{LSTM} 之间的注意力关系矩阵 $E \in \mathbf{R}^{p \times h}$:

$$E_{ij} = F_{\text{inter}}(P_i^{LSTM})^T F_{\text{inter}}(H_j^{LSTM}) \quad (5)$$

其中 $F_{\text{inter}}(\cdot)$ 为前馈神经网络.

分别计算软对齐结果,有

$$\beta_i = \sum_{j=1}^h \frac{\exp(E_{ij})}{\sum_{k=1}^h \exp(E_{ik})} H_j^{\text{LSTM}} \quad (6)$$

$$\alpha_j = \sum_{i=1}^p \frac{\exp(E_{ij})}{\sum_{k=1}^p \exp(E_{kj})} P_i^{\text{LSTM}} \quad (7)$$

其中 β_i 为 P_i^{LSTM} 与 H^{LSTM} 中各短语的软对齐结果, α_j 为 H_j^{LSTM} 与 P^{LSTM} 中各短语的软对齐结果.

2) 比较模块

经过交互模块的软对齐后,利用前馈神经网络 $G(\cdot)$ 依次进行对齐结果的比较^[4]. 具体过程定义为

$$v_i^p = G([P_i^{\text{LSTM}}; \beta_i]), \forall i \in [1, 2, \dots, p] \quad (8)$$

$$v_j^h = G([H_j^{\text{LSTM}}; \alpha_j]), \forall j \in [1, 2, \dots, h] \quad (9)$$

其中 $[\cdot]$ 为拼接操作.

3) 聚合模块

聚合模块采用平均池化和最大池化分别对比较结果 $\{v_i^p\}_{i=1}^p$ 和 $\{v_j^h\}_{j=1}^h$ 进行聚合,以前提 P 为例,有

$$v_{\text{avg}}^p = \text{avg}([v_1^p, v_2^p, \dots, v_p^p]) \quad (10)$$

$$v_{\text{max}}^p = \text{max}([v_1^p, v_2^p, \dots, v_p^p]) \quad (11)$$

对平均池化结果和最大池化结果进行拼接,得到 $x^p = [v_{\text{avg}}^p; v_{\text{max}}^p]$. 类似地, $x^h = [v_{\text{avg}}^h; v_{\text{max}}^h]$.

为了保证特征信息的完整性和显式地对特征进行变换,采用了 3 种启发式的操作:向量的拼接、内积和作差,对 x^p 和 x^h 进行融合.

$$v^{\text{agg}} = F([x^p; x^h; x^p x^h; x^p - x^h]) \quad (12)$$

$$o = W_a v^{\text{agg}} + b_a \quad (13)$$

其中 W_a 和 b_a 为线性层的权重和偏差, o 为胶囊的输出.

2.4 输出层

将所有胶囊的输出经过拼接后通过 softmax 层进行归一化,得到模型的输出结果 D . 该向量表示经过胶囊网络结构预测得到的 N 种蕴含关系的概率分布.

$$D = \text{softmax}([o_1; o_2; \dots; o_N]) \quad (14)$$

在此概率分布中选择概率最大的胶囊,该胶囊随即被激活,它所对应的蕴含关系就作为前提 P 和假设 H 最终的分类结果.

3 实验

3.1 数据集

SNLI 数据集包含 570×10^3 个由人工书写的句

对,全部由人工标注并且分类平衡. 原始的 SNLI 数据集中提供“蕴含”、“中立”、“矛盾”和“-”4 种标签,其中“-”表示标注者之间无法对该句对达成一致意见,因此在训练和测试过程删除该标签的句对.

MultiNLI 数据集包含 433×10^3 个句对,并且使用文本蕴含信息进行标注. 其创建过程模拟了 SNLI 数据集,但不同的是该数据集涵盖了一系列口语和书面文本体裁,并且支持独特的跨体裁的泛化评价. 由于 MultiNLI 未提供匹配测试集和不匹配测试集的标签,所以需要将测试结果提交到 Kaggle 开放评估平台来获取在测试集上的性能.

SNLI 和 MultiNLI 数据集的分布如表 1 所示.

表 1 SNLI 和 MultiNLI 数据集分布

数据集	训练集/句对	验证集/句对	测试集/句对
SNLI	550 152	10 000	10 000
MultiNLI	392 702	20 000	20 000

3.2 评价指标

实验中在 SNLI 和 MultiNLI 数据集上采用的评价指标均为准确率,定义为^[1]

$$A = \frac{N_c}{N_p} \quad (15)$$

其中 N_c 为蕴含关系预测正确的句对数量, N_p 为预测的句对数量.

3.3 参数设置

在数据预处理过程中,删除标签为“-”的样本,并将 SNLI 和 MultiNLI 数据集的文本长度分别设置为 32 和 48 个单词.

在模型训练过程中,嵌入层中的词向量由 300 维的 GloVe 词向量进行初始化,该词向量是在大型的、未标记的语料库上预先训练的. 对于未登录词,模型使用其在 GloVe 中的近义词进行替代. 字符向量和其他所有参数以默认方式进行随机初始化,并且保持更新. 模型采用 Adam 优化算法对参数进行优化,初始学习率在 SNLI 中设置为 0.000 1,在 MultiNLI 中设置为 0.000 2,其中一阶矩估计的指数衰减率 β_1 为 0.9,二阶矩估计的指数衰减率 β_2 为 0.999. 隐藏层的维度设置为 300 维. 为了控制网络的复杂度,防止过拟合,在模型中加入了 Dropout 层,比率初始值设置为 0.8,并在训练过程中保持指数衰减. 对于 SNLI 和 MultiNLI 数据集,批尺寸均设置为 48.

3.4 实验结果

对基于胶囊的英文文本蕴含识别方法在 SNLI 数据集上的实验结果与其他已公开发表的方法进行了对比,如表 2 所示. 表格中第 1 部分是由 Bowman 等提出的基准分类器模型,该方法融合了重叠词、BLEU 得分等人工特征^[1]. 第 2 部分是基于文本编码的模型. 第 3 部分的模型利用了文本之间的交互特征进行蕴含关系识别. 由表 2 可知,基于胶囊的英文文本蕴含识别方法在 SNLI 数据集上的准确率达到了 88.5%,超过所有基于文本编码的方法和基于交互特征的方法. 通过集成学习策略整合多个模型的输出结果之后,该方法的性能又得到了进一步提升,准确率达到了 89.2%.

表 2 各方法在 SNLI 数据集上的实验结果对比

方法	测试集/%
Unlexicalized features ^[1]	50.4
+ Unigram and bigram features ^[1]	78.2
300D LSTM encoders ^[14]	80.6
300D Tree-based CNN encoders ^[15]	82.1
600D BiLSTM encoders + intra-att ^[7]	84.2
300D NSE encoders ^[8]	84.6
300D DiSAN ^[16]	85.6
100D LSTM with attention ^[3]	83.5
300D mLSTM ^[9]	86.1
200D DecompAtt + intra-att ^[4]	86.8
300D NTI-SLSTM-LSTM ^[10]	87.3
300D re-read LSTM ^[11]	87.5
448D DIIN ^[12]	88.0
600D ESIM ^[17]	88.0
所提方法	88.5
600D ESIM (集成) ^[17]	88.6
448D DIIN (集成) ^[12]	88.9
所提方法 (集成)	89.2

在 MultiNLI 数据集上的实验结果如表 3 所示. 除 ESIM 和 DIIN 模型以外,其他均为 2017 年 RepEval 会议中的方法. 由表 3 可知,基于胶囊的英文文本蕴含识别方法在匹配和不匹配测试集上的准确率分别为 77.4% 和 76.4%,超过 2017 年 RepEval 会议中所有的方法. 与目前最好结果相比还有一定的差距,在匹配测试集和不匹配测试集上分别比 DIIN 模型低 2.6% 和 2.3%. 但是相比 DIIN 模型,该方法中的胶囊结构能够从不同的角度对样本进行分

析,并根据各个胶囊学习得到的软对齐表示,提取每一种蕴含关系的特征信息.

表 3 各方法在 MultiNLI 数据集上的实验结果对比

方法	match/%	mismatch/%
BiLSTM ^[2]	67.0	67.6
Rivercorners ^[18]	72.1	72.1
ESIM ^[17]	72.3	72.1
Shortcut-stacked encoders ^[19]	74.6	73.6
DIIN ^[12]	78.8	77.8
所提方法	76.7	76.1
Rivercorners (集成) ^[18]	72.2	72.8
DIIN (集成) ^[12]	80.0	78.7
所提方法 (集成)	77.4	76.4

3.5 分析

为了更好地理解模型的性能,首先对模型中每个关键组件的作用进行分析,实验结果如表 4 所示. 其中第 1 部分是对嵌入层中不同层面的特征信息进行分析的结果. 由此可以看出,在对文本进行向量转换的过程中,丰富的特征信息在一定程度上可以提升模型的性能. 第 2 部分的结果显示,将序列编码层的 BiLSTM 替换为 LSTM 后,模型在 SNLI 测试集上的准确率下降至 87.88%. 为了进一步验证 highway 网络和胶囊结构的有效性,在原有模型的基础上去除 highway 编码层后,准确率下降至 88.26%. 在只去除胶囊结构,保留其他组件的实验中,模型在 SNLI 测试集上的准确率仅有 82.81%. 消融实验的结果验证了模型中的 BiLSTM、highway 网络和胶囊结构对最终的性能都做出了积极的贡献.

表 4 模型在 SNLI 数据集上的消融实验结果

方法	测试集/%
词向量	88.23
词向量 + 字符特征	88.31
词向量 + 字符特征 + 词性特征	88.50
LSTM	87.88
去除 highway 网络	88.26
去除胶囊结构	82.81

为了更直观地验证不同属性的胶囊能够学习到相应蕴含关系独有的特征信息,接下来对各个胶囊中的注意力机制进行可视化分析. 首先,遵循文献

[9]的分析原则,观察由词与词之间的软对齐权重组成的注意力关系矩阵,检验基于注意力机制的软对齐是否起到了作用. 然后观察对于同一个样本在不同胶囊中的注意力关系矩阵,检验各个胶囊是否已经学习到对应蕴含关系的特征信息.

从 SNLI 测试集中随机选择 3 个样本进行分析. 如表 5 所示,这 3 个样本有相同的前提 P ,均为“A man playing an electric guitar on stage.”,但是假设 H 各不相同,所以它们有不同的蕴含关系标签. 将这 3 个样本的注意力关系矩阵进行可视化,如图 3 所示.

图 3 中的子图表示前提 P 和假设 H 之间的对齐权重分布,其中颜色越深表示该项受关注的程度越高. 可以观察到,基于交互注意力机制的软对齐

表 5 实验分析样本

样本	文本	标签
1	A man playing guitar on stage.	蕴含
2	A man is performing for cash.	中立
3	A man playing banjo on the floor.	矛盾

可以重点关注到两段文本中语义相近或者语义矛盾的部分. 例如,在样本 1 的注意力关系矩阵可视化图中,“playing”、“guitar”和“stage”的权重值高于其他单词,并且这些词对于蕴含关系的推断也起到了关键作用.

每一行中的 3 个子图表示同一个样本在不同胶囊中的注意力关系矩阵可视化图. 可以看出,属性为“蕴含”和“矛盾”的胶囊对样本的关注点集中在

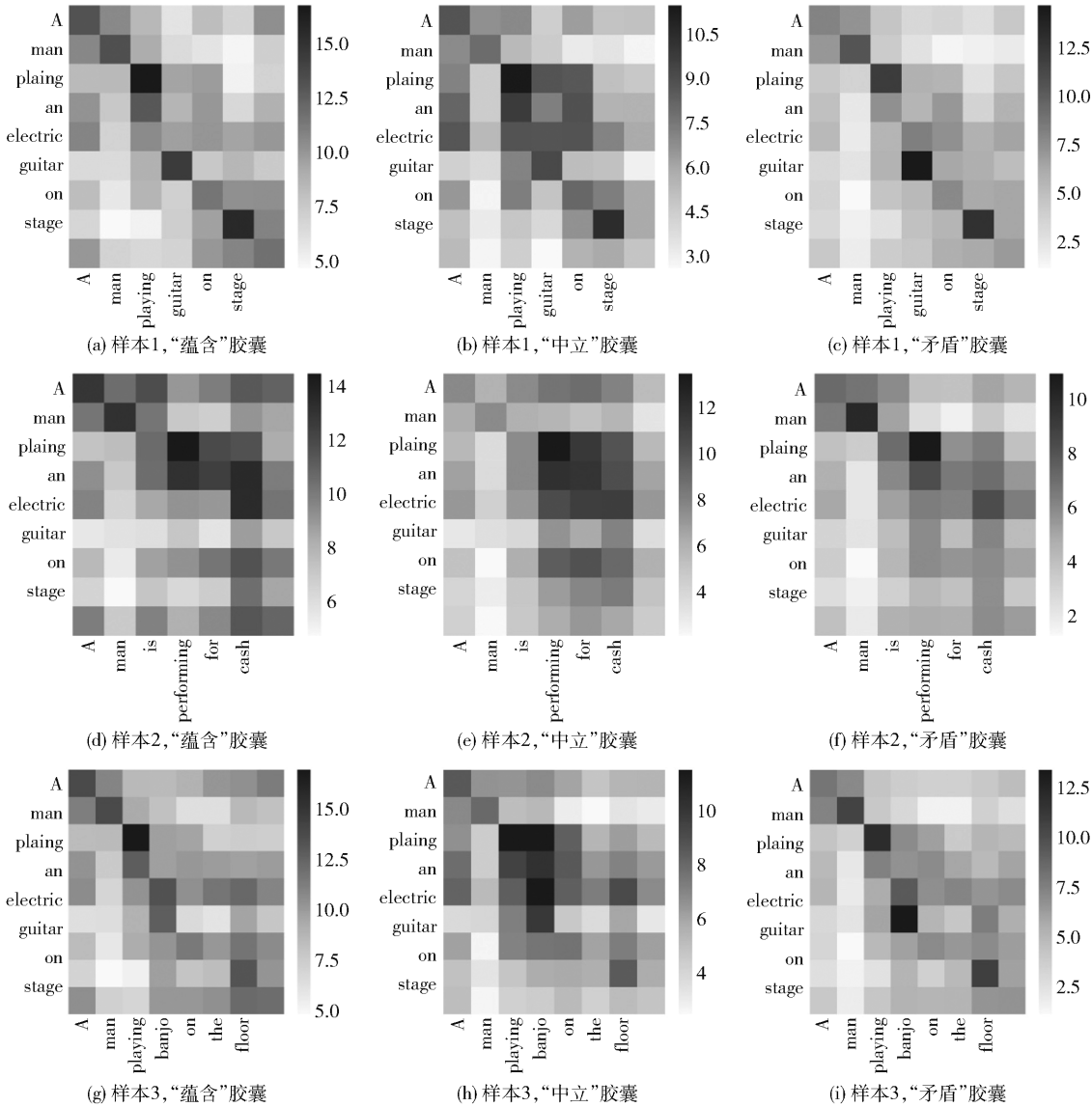


图 3 注意力关系矩阵可视化

主语、谓语和宾语等主要成分,如“man”、“playing”、“performing”、“guitar”和“banjo”,但是由于属性不同,侧重点会有所区别。当样本属于矛盾这一类别时,属性为“矛盾”的胶囊相对其他胶囊更注重文本之间具有矛盾关系的单词。例如样本3,“guitar”与“banjo”、“stage”与“floor”的权重在属性为“矛盾”的胶囊中明显高于另外两个,因此在判断最终蕴含关系时,“矛盾”的概率也高于它们。对于标签为“中立”的样本,它们不具备很强的语义关系,为了准确判断两段文本的蕴含关系,属性为“中立”的胶囊相比其他胶囊关注的范围更广泛。通过以上的可视化分析,验证了胶囊的有效性,不同属性的胶囊能够学习到相应类别的特征信息。

4 结束语

将胶囊引入英文文本蕴含识别任务中,并提出一种基于胶囊的英文文本蕴含识别方法。胶囊使模型具备更强的关系推理能力,并且可以针对不同的蕴含关系学习特定的特征信息。该方法在 SNLI 数据集上的准确率为 89.2%,在 MultiNLI 数据集上的准确率为 77.4% 和 76.4%,实验结果表明了该方法的有效性。这是在英文文本蕴含任务上对胶囊进行的初步探索,胶囊仍具有很大的开发潜力,例如,针对不同类别设计不同内部结构的胶囊等,这有待对其进一步研究。

参考文献:

- [1] Bowman S R, Angeli G, Potts C, et al. A large annotated corpus for learning natural language inference[C]//Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL Press, 2015: 632-642.
- [2] Williams A, Nangia N, Bowman S R. A broad-coverage challenge corpus for sentence understanding through inference[C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Volume 1, Long Papers. New Orleans: ACL Press, 2018: 1112-1122.
- [3] Rocktäschel T, Grefenstette E, Hermann K M, et al. Reasoning about entailment with neural attention[EB/OL]. 2016(2016-03-01) [2018-11-16]. <https://arxiv.org/abs/1509.06664>.
- [4] Parikh A P, Täckström O, Das D, et al. A decomposable attention model for natural language inference[C]//Conference on Empirical Methods in Natural Language Processing. Austin: ACL Press, 2016: 2249-2255.
- [5] Sabour S, Frosst N, Hinton G E. Dynamic routing between capsules[C]//Advances in Neural Information Processing Systems. Long Beach: MIT Press, 2017: 3859-3869.
- [6] Wang Y, Sun A, Han J, et al. Sentiment analysis by capsules[C]//Proceedings of the 2018 World Wide Web Conference. Lyon: International World Wide Web Conferences Steering Committee, 2018: 1165-1174.
- [7] Liu Y, Sun C, Lin L, et al. Learning natural language inference using bidirectional LSTM model and inner-attention[EB/OL]. 2016(2016-05-30) [2018-11-16]. <https://arxiv.org/abs/1605.09090>.
- [8] Munkhdalai T, Yu H. Neural semantic encoders[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia: ACL Press, 2017: 397-407.
- [9] Wang S, Jiang J. Learning natural language inference with LSTM[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego: ACL Press, 2015: 1442-1451.
- [10] Munkhdalai T, Yu H. Neural tree indexers for text understanding[C]//Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Valencia: ACL Press, 2017: 11-21.
- [11] Sha L, Chang B, Sui Z, et al. Reading and thinking: re-read LSTM unit for textual entailment recognition[C]//International Conference on Computational Linguistics: Technical Papers. Osaka: The COLING 2016 Organizing Committee, 2016: 2870-2879.
- [12] Gong Y, Luo H, Zhang J. Natural language inference over interaction space[EB/OL]. 2018(2018-05-26) [2018-11-16]. <https://arxiv.org/abs/1709.04348>.
- [13] Srivastava R K, Greff K, Schmidhuber J. Highway networks[EB/OL]. 2015(2015-11-03) [2018-11-16]. <https://arxiv.org/abs/1505.00387>.
- [14] Bowman S R, Gauthier J, Rastogi A, et al. A fast unified model for parsing and sentence understanding[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACL Press, 2016: 1466-1477.
- [15] Mou L, Men R, Li G, et al. Natural language inference by tree-based convolution and heuristic matching[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: ACL Press, 2016: 130-136.
- [16] Shen T, Zhou T, Long G, et al. DiSAN: directional

- self-attention network for RNN/CNN-free language understanding[C]//The 32th AAAI Conference on Artificial Intelligence. New Orleans: AAAI Press, 2018: 5446-5455.
- [17] Chen Q, Zhu X, Ling Z, et al. Enhanced LSTM for natural language inference[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: ACL Press, 2016: 1657-1668.
- [18] Balazs J A, Marrese-Taylor E, Loyola P, et al. Refining raw sentence representations for textual entailment recognition via attention[C]//Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP. Copenhagen: ACL Press, 2017: 51-55.
- [19] Nie Y, Bansal M. Shortcut-stacked sentence encoders for multi-domain inference[C]//Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP. Copenhagen: ACL Press, 2017: 41-45.

(上接第 6 页)

- [11] 杜洋, 董彬虹, 王显俊, 等. 基于串行策略的 SCMA 多用户检测算法[J]. 电子与信息学报, 2016, 38(7): 1888-1893.
- Du Yang, Dong Binhong, Wang Xianjun, et al. Multi-user detection scheme for SCMA systems based on serial strategy[J]. Journal of Electronics & Information Technology, 2016, 38(7): 1888-1893.
- [12] Beek J V D, Popovic B M. Multiple access with low-density signatures[C]//IEEE Global Telecommunications Conference. New York: IEEE Press, 2009:1-6.
- [13] Proakis J G, Salehi M. Digital communication [M]. New York: McGraw-Hill Book Co., 2007:182-188.
- [14] 吴昊, 张建秋, 宋汉斌. 一种基于格理论构造高维星座图的方法[J]. 电子学报, 2014, 42(8): 1672-1679.
- Wu Hao, Zhang Jianqiu, Song Hanbin. A lattice based approach to the construction of multi-dimensional signal constellations[J]. Acta Electronica Sinica, 2014, 42(8): 1672-1679.
- [15] 邵小桃, 郭鸣坤, 杨维. 基于资源块星座图的稀疏码多址接入码本设计[J]. 通信学报, 2018, 39(9): 1-9.
- Shao Xiaotao, Guo Mingkun, Yang Wei. SCMA codebook design based on resource block constellation[J]. Journal on Communications, 2018, 39(9): 1-9.