

文章编号: 1008-1542(2022)06-0594-08

基于度量学习和层级推理网络的抽取式摘要方法

成悦¹, 赵康¹, 勾智楠², 高凯¹

(1. 河北科技大学信息科学与工程学院, 河北石家庄 050018; 2. 河北经贸大学信息技术学院, 河北石家庄 050061)

摘要:当前大部分的抽取式摘要方法主要关注对摘要句的表示和抽取, 容易忽略对文本特征表示的充分性。为了解决这一问题, 提出一种基于度量学习和层级推理网络的抽取式摘要方法。首先, 在抽取式任务基础上提出基于度量学习和层级推理的抽取式摘要模型(MLHIN); 其次, 在 CNN/DailyMail 数据集上进行模型评估, 并在英文摘要数据集 CNN/DailyMail 上进行测试; 最后, 对测试结果进行验证。结果显示, 所提方法模型在 Rouge-1, Rouge-2, Rouge-L 上的得分明显优于其他模型, 比 Lead-3 模型分别高出 0.84%, 1.29% 和 2.43%; 通过将提出的度量损失 metric 和层级推理模型中的句子编码器替换掉, 可以看出模型性能均有不同程度的下降, 证明了提出的层级推理网络和度量损失的有效性。新算法能够提高模型捕捉长距离依赖的能力, 增强模型对摘要句与非摘要句的分辨力, 有效改善了抽取式摘要方法的性能。

关键词: 自然语言处理; 句子编码器; 文档编码器; 度量学习; 层级推理; 抽取式文本摘要

中图分类号: TN958.98

文献标识码: A

DOI: 10.7535/hbkd.2022yx06004

Metric learning based hierarchical inference network for extractive text summarization

CHENG Yue¹, ZHAO Kang¹, GOU Zhinan², GAO Kai¹

(1. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, Hebei 050018, China; 2. School of Information Technology, Hebei University of Economics and Business, Shijiazhuang, Hebei 050061, China)

Abstract: Most of the current extractive summarization methods mainly focus on the representation and extraction of summary sentences, and tend to ignore the adequacy of text feature representation. In order to solve this problem, an extractive summarization method was proposed. Firstly, on the basis of abstract tasks, an extractive summarization model (MLHIN) based on metric learning and hierarchical inference was proposed. Secondly, the model was evaluated and tested on

收稿日期: 2021-11-21; 修回日期: 2022-09-18; 责任编辑: 王淑霞

基金项目: 河北省自然科学基金(F2022208006); 河北省高等学校科学技术研究项目(QN2020198)

第一作者简介: 成悦(1997—), 女, 河北邯郸人, 硕士研究生, 主要从事自然语言处理及文本摘要方面的研究。

通讯作者: 高凯教授。E-mail: gaokai@hebust.edu.cn

成悦, 赵康, 勾智楠, 等. 基于度量学习和层级推理网络的抽取式摘要方法[J]. 河北科技大学学报, 2022, 43(6): 594-601.

CHENG Yue, ZHAO Kang, GOU Zhinan, et al. Metric learning based hierarchical inference network for extractive text summarization[J].

Journal of Hebei University of Science and Technology, 2022, 43(6): 594-601.

the English CNN/DailyMail dataset. Finally, the test results of the model on the dataset are verified. The results show that the proposed model has significantly higher scores than other models on Rouge-1, Rouge-2 and Rouge-L, which are 0.84%, 1.29% and 2.43% higher than the Lead-3 model respectively. After replacing the metric loss metric and the sentence encoder with other modules, it can be seen that the performance of the model has declined to varying degrees, which proves the effectiveness of the proposed hierarchical inference network and metric loss. The algorithm can improve the ability of model to capture long-distance dependency, enhance the ability of model to distinguish summary sentences from non-summary sentences, and effectively improve the performance of the extractive summarization methods.

Keywords: natural language processing; sentence encoder; document encoder; metric learning; hierarchical inference; extractive text summarization

抽取式摘要是从源文本中选择突出句或其他语义单位生成摘要,保证摘要的语法正确性以及与相应文档的语义相关性。传统方法主要采用基于图的方法^[1]、子模函数^[2]或整数线性规划^[3]对句子进行评分,再加上手工特征。由于深度学习方法在捕获文本语义和语法信息方面表现出了突出的能力^[4-5],因此,最近出现了使用神经网络提取突出句子的方法^[6]。这些方法虽然可以从文档中识别出重要的句子,但因为模型会选择语义上相互独立的句子来组成摘要,缺乏确保摘要连贯的能力,所以使人对文档的理解造成困难。

基于自注意力的体系结构,尤其是 Transformer,已成为自然语言处理(NLP)任务中的常用模型。主要方法是在大型文本语料库上进行预训练,然后在较小的特定任务的数据集上进行微调。Transformer 在机器翻译中是有效的,但对文本摘要来说,由于其在编码时不包含句子的位置信息,对句子级上下文的建模能力较差,因此直接应用到文本摘要任务有不足之处,为文本摘要建模长距离依赖的上下文仍然具有挑战^[7]。

为了解决这个问题,大部分工作试图探索不同的神经组件或组合来构建端到端的学习模型。CHOWDHURY 等^[8]提出了一个基于结构注意的层次编码器,来建模句子间和文档间的依赖关系。JIA 等^[9]提出了分层注意异构图模型,能很好地模拟不同层次的信息,包括单词和句子,并强调句子间的冗余依赖关系,通过消息传递标签依赖关系。YUAN 等^[10]提出了提取事实级语义单位,基于 BERT^[11]模型结合使用一个层次图掩码,将文本信息的多层次粒度合并到模型中,以便更好地抽取摘要。WANG 等^[12]提出了一种基于异构图的抽取式摘要神经网络,该神经网络包含除句子外不同粒度级别的语义节点,丰富了句子间的关系。这些方法的性能在很大程度上取决于机器对文档的理解,也就是文档表示。为了使文本能够学习到更深层次的数据之间的关系以及更好的文档表示,有学者对度量学习方法开展了相关研究。

和传统分类目标损失函数不同,度量学习目标损失函数能够使同一类别的数据在几何空间分布更加紧凑,不同类别的数据彼此远离,因此特征向量具有较强的区分能力。深度度量学习方法在人脸识别领域取得了突出效果^[13],证明了方法的有效性。对比损失(contrastive loss)^[14]是深度度量学习的开篇之作,首次将神经网络引入度量学习。对比损失仅约束类内对的特征尽量近,而类间对的特征尽量远,三元组损失(triplet loss)^[15]在对比损失的基础上,进一步考虑了类内对与类间对之间的相对关系。由于三元组损失一次采样 3 个样本,虽然能够同时考虑类内距、类间距以及二者的相对关系,但该损失没有充分利用训练时每个 batch 内的所有样本。因此,SONG 等^[16]提出在一个 batch 内建立稠密的成对(pair-wise)的连接关系。由于三元组损失在锚点选取时具有任意性,因此,有些不满足类间距 $>$ (类内距+间隔)的样本,可能并没有被挖掘到。MISHCHUK 等^[17]提出成对聚类损失函数(coupled clusters loss, CCL)为同类样本估计了一个类内中心,从而希望所有的正样本到聚类中心的距离间隔小于其他类间样本到聚类中心的距离。度量学习为衡量句间相似性提供了新的解决思路。

本文针对抽取式文本摘要任务提出一种基于度量学习的层级推理模型(metric learning based hierarchical inference network, MLHIN),由句子编码器、文档编码器和度量损失构成。模型使用 Bi-GRU 网络结合上下文信息来编码文档中每一个句子为一个向量,并将所有句子向量和一个可训练的文档向量输入到 Transformer 中,输出文档表示和句子表示。然后,将度量学习概念迁移到文本摘要任务中,计算文档表示和句子表示的相似性度量,利用度量得分选择摘要句。最后,在英文摘要数据集 CNN/DailyMail^[18]上对其进行测试,并进行消融实验对测试结果进行验证。

1 基于度量学习的层级推理模型

1.1 抽取式摘要任务

给定 $D = \{s_1, s_2, \dots, s_N\}$ 表示包含 N 个句子的源文档序列,其中 s_i 为文档的第 i 个句子。另外,每个句子包含 $|s_i|$ 个单词 $s_i = \{x_1, \dots, x_{|s_i|}\}$ 。令 T 表示人工总结的摘要。抽取式摘要的目的是从 D 中选择 M 个句子生成摘要 $D^* = \{s_1^*, s_2^*, \dots, s_M^*\}$,其中 $M \leq N$ 。标签 $Y = \{y_1, y_2, \dots, y_N\}$ 由 T 派生而来,其中 $y_i \in \{0, 1\}$ 表示抽取的摘要中是否应该包含 s_i 句。

MLHIN 模型主要由句子编码器、文档编码器和度量学习 3 大模块组成。首先,利用句子编码器将每个句子 s_i 转换为一个句子表示 s_i 。然后,这些句子表示将由文档编码器编码。最后,通过计算句子与文档表示之间的相似度来选择摘要句。

1.2 层级推理编码器

由于抽取摘要模型的目的是从文档中选择句子,因此对文档进行良好的建模至关重要。文档是分层结构的,一个文档是由多个句子组成,每个句子都由单词组成。因此利用层次结构,设计一个层级神经模型,原理如图 1 所示。层级推理结构由句子编码器、文档编码器构成。句子编码器的目的是学习文档中的每个句子的表示,文档编码器的目的是学习文档的表示。本文将详细描述模型的每个组件。

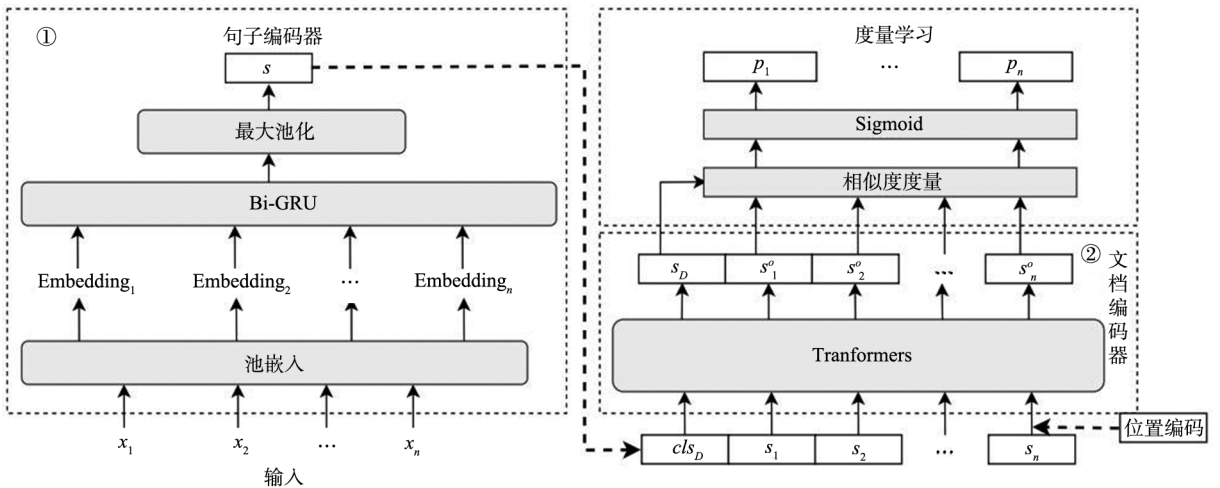


图 1 基于度量学习和层级推理网络的原理图

Fig.1 Schematic diagram of MLHIN

1) 句子编码器

句子编码器的作用是读取句子中的每一个单词并对每个词产生一个向量表示,经过最大池化后组成句子表示。为了能高效地对输入序列进行编码,模型使用了基于双向门控循环单元(bidirectional gated recurrent unit,简称 Bi-GRU)的编码器,其相对循环神经网络,不需要逐词处理输入序列,而是通过门控机制同时计算每个词的上下文向量,因此有着良好的并行能力,计算复杂度较低,对长距离依赖友好。

给定一个句子 $s_i = \{x_1, \dots, x_{|s_i|}\}$, x_i 为序列其中的一个词。首先,将序列词依次进行词嵌入得到词向量;然后,使用门控循环单元(gated recurrent unit,简称 GRU)网络编码句子,GRU 定义如下:

$$z_i = \sigma(\mathbf{W}_z[x_i, h_{i-1}]), \tag{1}$$

$$r_i = \sigma(\mathbf{W}_r[x_i, h_{i-1}]), \tag{2}$$

$$\tilde{h} = \tanh(\mathbf{W}_h[x_i, r_i \odot h_{i-1}]), \tag{3}$$

$$h_i = (1 - z_i) \odot h_{i-1} + z_i \odot \tilde{h}. \tag{4}$$

式中: \mathbf{W}_z , \mathbf{W}_r 和 \mathbf{W}_h 都是权重矩阵。Bi-GRU 由前向 GRU 和后向 GRU 组成,从左到右和从右到左 2 个方向处理一个句子,并产生 2 个隐藏序列 $(h_1^f, h_2^f, \dots, h_N^f)$ 和 $(h_1^b, h_2^b, \dots, h_N^b)$ 。将 h_j^f 和 h_j^b 进行拼接,经过最大池化得到句子的最终表示 s_i 。

2) 文档编码器

由句子编码器编码得到句子表示 s_i , 将文档中每一个句子向量拼接起来, 并在开头加上可学习的初始化的文档向量表示 cls_D 。然后, 与可学习的句子位置编码相加。

$$H = E_w + E_p, \quad (5)$$

式中: $E_w = [s_1, s_2, \dots, s_n]$, 表示所有的句子向量; $E_p = [\text{pos}_1, \text{pos}_2, \dots, \text{pos}_N]$ 是一个可学习的位置嵌入矩阵, 表示句子在文档中的位置。最后, 将其输入到 Transformer 中进行编码, 更新句子向量和文档向量。

文档编码器的目的是从输入文本中提取文档语义特征, 并映射到特征空间中。在 Transformer 模型中, 编码器由 L 个编码层堆叠而成, 每个编码层包括多头自注意力子层 (MultiHeadAtt) 以及按位置的前馈网络子层组成。另外, 在每 2 个子层之间, 使用残差连接 (Residual connection), 最后归一化 (Layer normalization)。多头注意力通过将输入文本特征映射到不同的特征空间, 可以提取到不同的语义特征信息。每一个编码层的最后的隐藏状态都会通过前馈神经网络。第 l 层编码层的计算公式如下:

$$h_s^{(l)} = \text{Norm}(\text{MultiHeadAtt}(Q^{(l)}, K^{(l)}, V^{(l)}) + Q^{(l)}), \quad (6)$$

$$h_{\text{out}}^{(l)} = \text{Norm}(\text{PFF}(h_s^{(l)}) + h_s^{(l)}), \quad (7)$$

$$\text{PFF} = (\text{ReLU}(h_s^{(l)} \omega_s^{(l)}) h_s^{(l2)} + h_s^{(l)}). \quad (8)$$

式中: $h_s^{(l)}$ 表示经过多头注意力和残差连接的输出; $\text{Norm}(\cdot)$ 表示层归一化操作, 得到第 l 个编码层的输出 $h_{\text{out}}^{(l)}$ 。编码器的第 1 个编码层接收的输入 $Q^{(1)} = K^{(1)} = V^{(1)} = H$, 其余层的输入为 $Q^{(l)} = K^{(l)} = V^{(l)} = h_{\text{el}}^{(l-1)}$ 。 $\omega_s^{(l)}$ 为模型的训练参数, PFF 为前馈网络子层。最终, 模型取最后一层的输出 $h_{\text{out}}^{(L)} [s_D, s_1^o, \dots, s_N^o]$, 得到文档向量表示 s_D 和更新后的句子表示 s_i^o 。

1.3 基于度量学习的损失函数

本文提出使用度量学习的方法来选择摘要。首先, 通过余弦距离计算文档表示和句子表示的相似度:

$$p_i = 1 / (1 + \exp(-\cos \langle s_D, s_i^o \rangle)), \quad p_i \in (0, 1). \quad (9)$$

以交叉熵损失作为整个模型的训练目标, 为了使摘要句向量与文档向量更近, 非摘要句与文档更远, 摘要句和文档向量的相似性更大, 模型通过优化以下损失函数进行学习:

$$L = \frac{1}{N} \sum_i L_i = \frac{1}{N} \sum_i - [y_i \log p_i + (1 - y_i) \log (1 - p_i)]. \quad (10)$$

式中: $y_i \in \{0, 1\}$ 表示参考摘要中是否包括第 i 句; N 为文档中包含的句子个数。在推理阶段, 模型通过计算句子与文档表示之间的相似度选择摘要句。

2 实验部分

2.1 环境及参数设置

模型在一个 GPU (GeForce GTX TITAN X, 12G) 上训练。根据验证集上的评估, 选择结果最好的模型参数, 并报告测试集上的结果。使用 Adam^[19] 作为优化器, epoch 设置为 20, 学习率时间表遵循前 1 000 步热身的策略。

参数方面, 基于 CNN/DailyMail 训练集的源文本和摘要文本建立词表, 大小为 5 000, 使用维度大小为 50 的 GloVe 词向量。文档编码器的层数设置为 4, 每层的注意力头数设置为 4, 隐藏层神经元个数为 100。使用位置编码, 以利用序列的顺序信息, 编码维度和句向量维度相同, 设置为 200。学习率设置为 0.000 5, 批大小设为 32。

2.2 实验测度

Trigram blocking 启发于 PAULUS 等^[20] 和 LIU 等^[21] 所提出的方法, 模型使用 Trigram Blocking 进行解码, 这是一个简单但功能强大的 Maximal Marginal Relevance^[22] 方法。具体地说, 在选取摘要句时, 根据句子与文档的相似性分数来排列句子, 并且丢弃那些与前句有 3 个词重叠的句子。

按照惯例, 在接下来的实验中使用了 Rouge-1 (uni-gram), Rouge-2 (bi-gram) 和 Rouge-L (最长公共子序列) 评估本文的模型。通过计算人类创建的候选摘要和参考摘要之间的重叠单元 (如 n -gram、单词序列和单

词对)的数量,自动评估摘要的质量。Rouge-N^[23]的计算公式为

$$\text{Rouge-N} = \frac{\sum_{S \in \text{Reference Summaries}} \sum_{\text{gram}_n \in S} \text{Count}_{\text{match}}(\text{gram}_n)}{\sum_{S \in \text{Reference Summaries}} \sum_{\text{gram}_n \in S} \text{Count}(\text{gram}_n)}, \quad (11)$$

式中:分子为模型生成的摘要与参考摘要之间共有的 n -gram 的个数;分母为参考摘要 n -gram 的总个数。

2.3 基线模型

Lead-3 即抽取文档的前 3 句作为文章的摘要。Lead-3 方法虽然简单直接,但却是非常有效的方法。

RAMESH 等^[24]提出了一个基于序列分类器的循环神经网络模型,采用 GRU-RNN 结构,由单词级别的循环神经网络得到句子表示,又经过句子级别的循环神经网络得到文档表示。该模型表述简单,可解释性强,并提出新的训练机制,使用生成式摘要的模式来训练抽取式任务。

DONG^[25]提出了一种新的方法来训练神经网络执行单文档抽取摘要,而不需要启发式生成提取标签。采用策略梯度强化学习算法对模型进行训练,以选择最大 Rouge 分数的句子序列。

NEUSUM^[26]模型将选择策略集成到打分模型中,解决了此前抽取式文本摘要中句子打分和句子选择这 2 部分割裂的问题。通过 MLP 网络学习 Rouge 和 F1 的评分,把打分和句子选择整合在同一个网络中,最终训练出来的模型通过选择不同的句子得到尽可能高的 Rouge 评分。

2.4 主体实验

本文在抽取式任务上提出了基于度量学习和层级推理的抽取式摘要模型(metric learning based hierarchical inference network, MLHIN),该模型由 3 部分组成,分别是句子编码器、文档编码器和度量损失。句子编码器由 Bi-GRU 构成,将文档中每个句子编码为一个向量,将这些向量组合起来输入到文档编码器中,文档编码器由 Transformer 构成,输出文档表示和句子表示。度量损失计算句子和文档的相似度,抽取摘要句。

本文使用 CNN/DailyMail 数据集进行模型评估。该数据集包含新闻文档及其对应的摘要,通常用于抽取式摘要和生成式摘要。本文使用 WANG 等^[12]提供的脚本来下载数据集。其中训练集、验证集和测试集中分别包含 287 226,13 368 和 11 490 份文档。

模型在英文摘要数据集 CNN/DailyMail 上进行测试,并报告各个模型对应的 Rouge-1, Rouge-2, Rouge-L 的得分,使用封装了 Rouge 脚本的 pyRouge 工具计算得分。

2.5 消融实验

本文提出的模型(MLHIN 模型)中的句子编码器使用 GRU,文档编码器使用 Transformer,句子选择部分使用度量损失。为了证明本文所提方法的有效性,进行消融实验,将 3 部分组件分别进行替换研究。

MLHIN 模型 为 GRU-Transformer-metric 结构,即句子编码器使用 GRU,文档编码器使用 Transformer,句子选择部分使用度量损失。

GGM 模型 为 GRU-GRU-metric 结构,即句子编码器和文档编码器均使用 GRU,句子选择使用度量损失。

CTM 模型 为 CNN-Transformer-metric 结构,即句子编码器使用 CNN,文档编码器使用 Transformer,句子选择部分使用度量损失。

CTS 模型 为 CNN-Transformer-sigmoid 结构,即句子编码器使用 CNN,文档编码器使用 Transformer,句子选择部分使用 sigmoid 函数。

GTS 模型 为 GRU-Transformer-sigmoid 结构,即句子编码器使用 GRU,文档编码器使用 Transformer,句子选择部分使用 sigmoid 函数。

TTM 模型 为 Transformer-Transformer-metric 结构,即句子编码器使用 Transformer,文档编码器使用 Transformer,句子选择部分使用度量损失。

首先,分析其他抽取式摘要模型中不同句子编码器的性能。其次,为了进一步分析其他抽取式摘要模型中不同文档编码器的性能,将本文的文档编码器开展替换研究。最后,本文对度量损失进行 2 个替换研究实验,替换实验如下。

- 1)仅将本文提出的度量损失 metric 替换为其他模型中的 sigmoid。
- 2)在第 1)步的基础上,将层级推理模型中的句子编码器替换为 CNN。

3 实验结果与分析

3.1 主体实验结果与分析

表 1 展示了各个模型在 CNN/DailyMail 数据集上的 Rouge 指标得分。

由表 1 可以看出,本文提出的 MLHIN 在 Rouge-1, Rouge-2, Rouge-L 3 个指标上的得分都明显优于其他模型,相对于已有的模型 Lead-3,得分分别提升了 0.84%, 1.29%, 2.43%;而相对于 SummaRuNNer 模型,得分分别提升了 1.66%, 2.71%, 3.8%;相比 GRU 模型 NEUSUM,在 Rouge-L 的得分提升了 1.12%;对比 BANDITSUM 模型,在 Rouge-2 和 Rouge-L 上分别提升了 0.21% 和 1.5%。MLHIN 模型相对 Lead-3 和 SummaRuNNer,在 Rouge-2 和 Rouge-L 得分上有明显提升。相对于 BANDITSUM 和 NEUSUM 模型,所提模型的效果在 Rouge-L 的评分更高。以上结果表明,本文所提方法能够有效抽取摘要。

由上述结果可以看出,模型在 CNN/DailyMail 数据集上关于 Rouge-1 和 Rouge-2 的分值提升较少,在 Rouge-L 上的得分提升较多。Rouge-L 表示模型生成的摘要和参考摘要之间相似的最长公共子序列, Rouge-L 分值越高,最长公共子序列越长,表明模型生成的摘要与参考摘要越接近。度量学习用来衡量当前句子为摘要句的概率,说明度量学习能够较好地选取摘要句,证明 MLHIN 模型能够提升摘要抽取的效果。同时,模型在 Rouge-1 和 Rouge-2 上的分数相比基线模型也有一定的提高,进一步证明了所提方法在抽取式摘要任务上的效果有整体提升。

3.2 消融实验结果与分析

表 2 展示了不同模型在 CNN/DailyMail 数据集上的得分。

从表 2 可以看出,CTM 使用 CNN 作为句子编码器,将 CNN 替换为本文提出的层级推理网络中的 GRU 句子编码器后,模型性能有了很大提升。CTM 中使用 CNN 作为句子编码器,将 CNN 替换为本文提出的层级推理网络中的 GRU 句子编码器后,模型性能有了很大提升,表明本文提出的层级推理网络中的句子编码器 GRU 性能优于其他模型中的句子编码器 CNN。GGM 模型中使用了 GRU 作为文档编码器,将该文档编码器替换为本文提出的层级推理网络中的 Transformer,从结果可以看出,层级推理网络中的文档编码器 Transformer 大大提高了模型的性能。通过替换实验,将本文提出的度量损失 metric 替换为其他模型中的 sigmoid,可以看出 3 个指标分别有不同程度的下降,说明了本文提出的度量损失的有效性。将层级推理模型中的句子编码器替换为 CNN,可以看出模型性能也有不同程度的下降,证明了本文提出的层级推理网络和度量损失的有效性。

4 结 语

本文提出了一种基于度量学习和层级推理网络的抽取式文本摘要方法(MLHIN),层级推理网络由句子编码器和文档编码器构成,文档编码器的输入为句子编码器的输出,结合句子的位置信息,编码器不仅能编码句子信息,也能够对包含上下文的文档进行编码,增强了对编码器的训练。同时引入度量学习,对编码后的文档和句子信息进行筛选,保留关键信息,使得模型能根据原文的关键信息来抽取摘要。在摘要数据集 CNN/DailyMail 上的实验表明,提出的 MLHIN 方法能明显提升模型抽取摘要的质量。

表 1 MLHIN 与其他模型的实验结果对比
Tab.1 Comparison of experimental results between MLHIN and other models

模型名称	Rouge-1	Rouge-2	Rouge-L
Lead-3	40.42	17.62	36.67
SummaRuNNer ^[24]	39.60	16.20	35.30
BANDITSUM ^[25]	41.50	18.70	37.60
NEUSUM ^[26]	41.59	19.01	37.98
MLHIN	41.26	18.91	39.10

表 2 消融实验结果对比

Tab.2 Comparison of ablation experiment results

模型名称	Rouge-1	Rouge-2	Rouge-L
GGM	39.91	17.81	37.87
CTM	40.92	18.59	38.78
CTS	40.61	18.23	38.51
GTS	40.47	18.35	38.40
TTM	32.50	16.70	36.50
MLHIN	41.26	18.91	39.1

本研究在探索算法有效性的过程中,发现文档编码器存在一定的局限性,即文档编码器的输入依赖于句子编码,若句子编码不够准确,会导致后续的训练出现误差。在学习过程中发现,更强大且更有效的算法框架能够帮助模型获得更好的抽取结果。强化学习能够将参考摘要作为信号来激励句子向量编码。未来,将继续探索将抽取式摘要方法与强化学习相结合,以此来提升模型性能,得出更好的抽取式摘要模型。

参考文献/References:

- [1] ERKAN G,RADEV D R.LexRank:Graph-based lexical centrality as salience in text summarization[J].Journal of Artificial Intelligence Research,2004,22(1):457-479.
- [2] LIN H,BILMES J.A class of submodular functions for document summarization[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics;Human Language Technologies-Volume 1.[S.l.]: Association for Computational Linguistics,2011:510-520.
- [3] BERG-KIRKPATRICK T,GILLICK D,KLEIN D.Jointly learning to extract and compress[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies-Volume 1. [S. l.]: Association for Computational Linguistics,2011:481-490.
- [4] MIKOLOV T,CHEN K,CORRADO G,et al.Efficient Estimation of Word Representations in Vector Space[EB/OL].<https://arxiv.org/abs/1301.3781>,2021-10-01.
- [5] HU Baotian,TANG Buzhou,CHEN Qingcai,et al.A novel word embedding learning model using the dissociation between nouns and verbs [J].Neurocomputing,2016,171:1108-1117.
- [6] CHENG J P,LAPATA M.Neural summarization by extracting sentences and words[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics.Berlin:Association for Computational Linguistics,2016:484-494.
- [7] XU Jiacheng,CHEN Danlu,QIU Xipeng,et al.Cached long short-term memory neural networks for document-level sentiment classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing,Austin:Association for Computational Linguistics,2016:1660-1669.
- [8] CHOWDHURY T,KUMAR S,CHAKRABORTY T.Neural abstractive summarization with structural attention[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence.Yokohama:IJCAI'20,2021:3716-3722.
- [9] JIA Ruipeng,CAO Yanan,TANG Hengzhu,et al.Neural extractive summarization with hierarchical attentive heterogeneous graph network[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing.[S.l.]: Association for Computational Linguistics,2020:3622-3631.
- [10] YUAN Ruifeng,WANG Zili,LI Wenjie.Fact-level extractive summarization with hierarchical graph mask on BERT[C]//Proceedings of the 28th International Conference on Computational Linguistics,Barcelona:International Committee on Computational Linguistics,2020:5629-5639.
- [11] DEVLIN J,CHANG M W,LEE K,et al.BERT:Pre-training of deep bidirectional transformers for language understanding[J].Computation and Language (cs.CL),2018.DOI:10.48550/arXiv.1810.04805.
- [12] WANG Danqing,LIU Pengfei,ZHENG Yining,et al.Heterogeneous graph neural networks for extractive document summarization[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.[S.l.]:Association for Computational Linguistics,2020:6209-6219.
- [13] DENG Jiankang,GUO Jia,XUE Niannan,et al.ArcFace:Additive angular margin loss for deep face recognition[J].IEEE,2022,44(10):5962-5979.
- [14] HADSELL R,CHOPRA S,LECUN Y.Dimensionality reduction by learning an invariant mapping[C]//2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.New York:IEEE,2006:1735-1742.
- [15] SCHROFF F,KALENICHENKO D,PHILBIN J.FaceNet:A unified embedding for face recognition and clustering[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition.Beijing:IEEE,2015:815-823.
- [16] SONG H O,XIANG Y,JEGELKA S,et al.Deep metric learning via lifted structured feature embedding[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition.Las Vegas:IEEE,2016:4004-4012.
- [17] MISHCHUK A,MISHKIN D,RADENOVIC F,et al.Working Hard to Know Your Neighbor's Margins:Local Descriptor Learning Loss [EB/OL].<https://arxiv.org/abs/1705.10872v1>,2018-01-12.
- [18] HERMANN K M,KOCISKY T,GREFENSTETTE E,et al.Teaching machines to read and comprehend[C]// Proceedings of the 28th International Conference on Neural Information Processing Systems.MA:MIT Press,2015:1693-1701.
- [19] KINGMA D P,BA L J.Adam:A method for stochastic optimization[J].Computer Science,2015.DOI:10.48550/arXiv.1412.6980.
- [20] PAULUS R,XIONG Caiming,SOCHER R.A Deep Reinforced Model for Abstractive Summarization[EB/OL].<https://arxiv.org/abs/1705.04304>,2021-10-01.

- [21] LIU Y,LAPATA M.Text summarization with pretrained encoders[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).Hong Kong: Association for Computational Linguistics,2019:3730-3740.
- [22] CARBONELL J,GOLDSTEIN J.The use of MMR,diversity-based reranking for reordering documents and producing summaries[C]// Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.New York: Association for computing machinery,1998:335-336.
- [23] LIN C Y.ROUGE:A package for automatic evaluation of summaries[C]//Text Summarization Branches Out.Barcelona:Association for Computational Linguistics,2004:74-81.
- [24] NALLAPATI R,ZHAI F F,ZHOU B W.SummaRuNNer:A recurrent neural network based sequence model for extractive summarization of documents[C]//Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence,San Francisco:AAAI Press,2017:3075-3081.
- [25] DONG Y,SHEN Y K,CRAWFORD E,et al.Banditsum:Extractive summarization as a contextual bandit[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing.Brussels:Association for Computational Linguistics,2018:3739-3748.
- [26] ZHOU Qingyu,YANG Nan,WEI Furu,et al.Neural document summarization by jointly learning to score and select sentences[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1:Long Papers).Melbourne:Association for Computational Linguistics,2018:654-663.

向本期载文的审稿专家致谢

本期《河北科技大学学报》共发表论文 11 篇,这些论文的发表是与有关专家的认真审读、细查资料、推敲分析、中肯评价分不开的。对此,本刊编辑部特向这些专家表示敬意,对他们的辛勤劳动表示感谢。

本期载文的审稿专家名单如下(按姓名的汉语拼音字母顺序排列):

毕秀丽 程振波 戴玮辰 姜保军 蒋 斌 乐启炽
 李国庭 李建武 李天平 李元凯 廖 勇 刘金琨
 马履中 马晓军 皮大伟 孙晓东 王洪亮 魏霖静
 杨金显 张 宇 赵 京 赵延治 赵智忠 郑宏宇
 郑建勇

(本刊编辑部)