



计算机科学

COMPUTER SCIENCE

基于预训练模型的无监督剧本摘要

苏琦, 王红玲, 王中卿

引用本文

苏琦, 王红玲, 王中卿. 基于预训练模型的无监督剧本摘要[J]. 计算机科学, 2023, 50(2): 310-316.

SU Qi, WANG Hongling, WANG Zhongqing. [Unsupervised Script Summarization Based on Pre-trained Model](#) [J]. Computer Science, 2023, 50(2): 310-316.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[亮度自调节的无监督图像去雾与低光图像增强算法研究](#)

Study on Unsupervised Image Dehazing and Low-light Image Enhancement Algorithms Based on Luminance Adjustment

计算机科学, 2023, 50(1): 123-130. <https://doi.org/10.11896/jsjcx.211100058>

[基于多模态注意力的噪声事件分类模型](#)

Noise Event Classification Model Based on Multimodal Attention

计算机科学, 2022, 49(11A): 211000161-7. <https://doi.org/10.11896/jsjcx.211000161>

[EGOS-DST:对话现象感知和模式引导的一步对话状态追踪算法](#)

EGOS-DST:Efficient Schema-guided Approach to One-step Dialogue State Tracking for Diverse Expressions

计算机科学, 2022, 49(11A): 210900246-7. <https://doi.org/10.11896/jsjcx.210900246>

[基于论据边界识别的立场分类研究](#)

Stance Detection Based on Argument Boundary Recognition

计算机科学, 2022, 49(11A): 210800180-5. <https://doi.org/10.11896/jsjcx.210800180>

[结合情感信息的个性化对话生成](#)

Personalized Dialogue Generation Integrating Sentimental Information

计算机科学, 2022, 49(11A): 211100019-6. <https://doi.org/10.11896/jsjcx.211100019>

基于预训练模型的无监督剧本摘要

苏琦 王红玲 王中卿

苏州大学计算机科学与技术学院 江苏苏州 215006

(20205227102@stu.suda.edu.cn)

摘要 剧本是一种特殊的文本结构,以人物的对话和对场景的描述信息组成文本。无监督剧本摘要是指对篇幅很长的剧本进行压缩、提取,形成能够概括剧本信息的短文本。提出了一种基于预训练模型的无监督剧本摘要方法,首先在预训练过程中通过增加对文本序列处理的预训练任务,使得预训练生成的模型能够充分考虑剧本中对话的场景描述及人物说话的情感特点,然后使用该预训练模型作为训练器计算剧本中的句间相似度,结合 TextRank 算法对关键句进行打分、排序,最终抽取得分最高的句子作为摘要。实验结果表明,该方法相比基准模型方法取得了更好的效果,系统性能在 ROUGE 评价上有显著的提高。

关键词: 训练模型;预训练任务;剧本摘要;无监督;句间相似度;对话

中图法分类号 TP391

Unsupervised Script Summarization Based on Pre-trained Model

SU Qi, WANG Hongling and WANG Zhongqing

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Abstract The script is a special text structure, which is composed of the dialogue between characters and the description of the scene. Unsupervised script summary refers to compressing and extracting a long script to form a short text that can summarize the information of the script. Therefore, this paper proposes an unsupervised script summary method based on a pre-training model. By adding pre-training tasks for text sequence processing in pre-training, the generated pre-training model fully takes into account the description of the dialogue in the script and the emotional characteristics of the characters, then the model is used as a trainer to calculate the similarity between sentences and combined with the TextRank algorithm to score and sort the key sentences. Finally, the sentence with the highest score is selected as the summary. Experimental results show that the proposed method has better performance than the base model, and the performance is significantly improved in the ROUGE evaluation.

Keywords Pre-trained model, Pre-training task, Script summary, Unsupervised, Sentence similarity, Dialogue

剧本是一部影视作品最完整的书面形式,除了对话以外,还包括对剧中人物、场景等的详细描述。通常人们根据影视剧本来导演影视作品,因此剧本是合作电影作品的一个重要组成部分。但是剧本通常包含的信息量庞大,篇幅过长,人们如果通过阅读整个剧本来了解关键剧情内容就会相当费时。快速且准确地从剧本中提炼关键内容,可帮助人们快速了解剧本内容,这个任务称为剧本摘要,其实质是文本摘要。但剧本摘要有别于一般的文本摘要任务,它们处理的文本是不同的。以图 1 为例,从原剧本中抽取出的最具有代表性的句子里可以看出剧本摘要的主要信息极可能存在于场景描述和主要人物对话上,所以摘要需要考虑场景以及在不同场景下每个发言者的身份、目的、语言风格等,而传统的抽取式文本摘要则不考虑这些信息,导致其具有局限性。现在的预训练模型是一种迁移学习的应用,利用几乎无限的文本,学习输入句子的每一个成员的上下文相关的表示,目前它在几乎所有自然语言处理任务中都取得了最佳的效果。BERT^[1]模型预

训练过程中,一个预训练任务为 NSP(下一句预测),用于判断输入的句子对是否拥有上下文关系。比如,对于 A-B 句子对,它通过一定规则对句子对的关系做分类,预测 B 到底是不是 A 句子的下一个句子,使模型具有句子级别的识别能力。

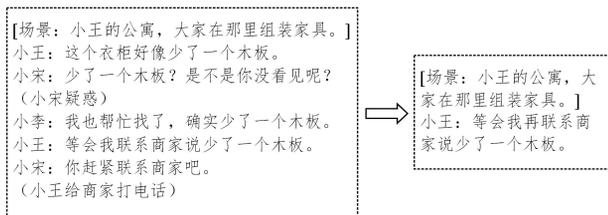


图 1 无监督剧本摘要实例

Fig. 1 Example of unsupervised script summary

受到 NSP 任务的启发,为了更好地利用剧本的特殊结构,本文提出了一种基于多个分类任务的预训练模型的无监督剧本摘要方法。具体来讲,就是在预训练模型形成过程中增加两个预训练任务,在预训练时以一定的概率对每个输入

到稿日期:2021-11-03 返修日期:2022-06-28

基金项目:国家自然科学基金(61976146)

This work was supported by the National Natural Science Foundation of China(61976146).

通信作者:王红玲(hlwang@suda.edu.cn)

的第二句话进行预测:1)是否与第一句话来自同一场景;2)是否来自同一人物。进行摘要前,首先通过预训练模型进行上述预测,根据预测的结果挑选句子作为第二句话,挑选出的句子与输入拼接形成句子对,将该句子对和预测结果(标签)一起送入模型进行预训练;然后用预训练获得的模型来计算句间相似度;根据句间相似度,结合 TextRank^[2]模型对句子进行打分、排序,最后抽取最代表剧本主要信息的句子作为摘要。

本文第1节介绍相关工作;第2节介绍预训练模型的构建和训练,以及句间相似度计算和无监督剧本摘要;第3节介绍具体实验及实验结果分析;最后总结全文。

1 相关工作

剧本摘要,本质就是文本摘要的特殊体。剧本是以对话形式组成的文本,所以剧本摘要主要就是对话形成摘要。本文摘要采用的是基于预训练模型的无监督抽取式的自动文摘方法。近几年 Liu^[3]提出的 BERTSUM 使用两层的 Transformer^[4],被应用于抽取式摘要工作。Ryuji 等^[5]提出的一种隐式引用抽取器,是能够无监督端到端抽取对话摘要的神经网络模型。Pinelopi 等^[6]提出了一种用于剧本抽取式摘要的结构,其根据关键叙事事件形式化叙述结构,并将其视为潜在内容,以便提取最佳场景序列。Zhou 等^[7]针对抽取整句摘要方法存在冗余性的问题,提出了一种以子句作为抽取单元的抽取式摘要方法。2021年以来,Feng 等^[8]提出将预训练对话回复生成模型作为一种无监督的对话标注器,借助其在预训练阶段编码的背景知识来获得辅助信息。Zou 等^[9]提出了一个领域不可知的低资源对话摘要的多源预训练范式,除去传统的预训练任务,采用对抗信号来学习领域未知信息。Chen 等^[10]提出了一组会话数据增强(CODA)方法,用于半监督生成会话摘要,例如随机交换/删除会话中的句子,把对话行为标注到对话文本中,将话语替换为基于对话上下文生成的释义;使用伪摘要对未标记对话的摘要模型进行预训练,然后在标记的对话上对其进行微调。Zhao 等^[11]设计了一个具有槽级掩码交叉注意的异构语义图,以增强特征,获得更正确的摘要。Zou 等^[12]提出了一种新颖的主题增强两阶段对话摘要器和感知神经主题模型,用于面向主题的客户对话摘要。

数据集以场景分块,每个场景主要由人物的对话组成。

【场景一:中央公园:罗斯和莫妮卡正在看着菲比睡觉。】
莫妮卡:她是怎么做到的?
菲比:(惊醒他们)哦!什么什么什么!...你好。
罗斯:没关系,你知道,你刚刚又打瞌睡了。
莫妮卡:你怎么了?
菲比:我昨晚没睡!
(钱德勒和乔伊进来。乔伊在数他的步数。)
乔伊:……九十五,九十六,九十七。看,我告诉你了!从我们的地方到这里不到一百步。
……
【场景二:麦迪逊广场花园,伙计们正在寻找他们的座位。】
罗斯:(挤过去的人)对不起,对不起……
钱德勒:什么?那天晚上和卡罗尔在一起有冰吗?塑料座椅?
罗斯:不,实际上我只是说看起来我们没有坐在一起。但是现在你说起来,那天晚上那里结了冰,那是第一场霜。
乔伊:来吧,坐下。
……
【场景三:莫妮卡和瑞秋都在客厅里闲逛。】
……

随着时代的发展,预训练模型得到了越来越多的关注,Dai 等^[13]使用语言建模和序列自编码来改善循环神经网络的序列学习,被认为是现代预训练模型的开端。Devlin 等提出的 BERT,是一个预训练的语言表征模型,它采用新的掩蔽语言模型,能生成深度的双向语言表征。Zhang 等^[14]提出了基于预训练的编码解码框架。Liu 等^[15]提出了基于 BERT 的文档级编码器,该编码器能够表达文档的语义,并获得文档的句子表示以获取摘要。Li 等^[16]基于 LSTM^[17]神经网络适用于序列数据的特点,设计了一个用于对句子语义建模的神经网络模型。

目前关于剧本这种特殊文本的无监督抽取式摘要工作较少,基于对上述研究的分析,为了提高无监督剧本对话摘要的准确率,更好地构建对话、场景和人物的关联,本文提出了利用目前主流的预训练模型 BERT,通过增加预训练任务,构建适合剧本的模型;然后计算剧本句间相似度,用于无监督抽取式模型获得摘要。

2 基于预训练模型的无监督剧本摘要

本节将详细介绍如何增加预训练任务来构建模型。首先在预训练时通过增加两个预训练任务来判断输入是否来自同一场景和人物,使用最终生成的预训练模型对数据集进行训练:输入的每条数据为两句话,经过三层嵌入层形成句子向量送入模型进行训练,使用 Sigmoid 激活函数对输出进行处理,输出一个 0~1 的概率值作为句间相似度;结合无监督抽取式摘要模型 TextRank 对句子进行打分、排序,将得分高的几句作为最终的摘要。

2.1 预训练模型的构建

BERT 模型预训练过程中原本有两个预训练任务:MLM (Masked Language Models) 和 NSP (Next Sentence Prediction)。MLM 是指完形填空任务,即遮住句子中的某些单词,让模型去预测遮住的单词是什么。在 NSP 中,模型只考虑了相邻句子之间的关系,但从整个具有多个场景的剧本来看,场景模块是由具有不同说话特点的人物的对话组成的,这就需要模型不仅仅只关注相邻句子间的关系,因为同一场景,同一人物说的话之间也具有很强的关联性。我们截取剧本中一集的部分内容进行说明,如图 2 所示。

图 2 数据集内容展示

Fig. 2 Data set content display

在此,在相同数据集下,为预测下一句的关系添加了两个预训练任务:APTS(Add Pre-training Tasks-Secene)和 APTP(Add Pre-training Tasks-People)。在预训练过程中,数据中有一半 B 是跟随 A 的实际相邻下一句,一半是非实际相邻下一句。当 50% 的概率为 B 是跟随 A 的实际下一句子的同时,要区分 B 和 A 是否来自同一场景, A 与 B 属于同一场景时,要区分是否是同一人所说;同样地,有 50% 的概率为来自非相邻的随机句子的同时,同样需要进行场景和人物的区分。我们会为此标上标签: IsSecene 或 NotSecene, IsPeople 或 NotPeople,取值 1 或 0。

本文预训练模型的构成过程如图 3 所示。首先对原始文本进行解析,形成标准化输入;原始数据本是无标签的数据,经过 NSP, APTS 和 APTP 任务形成句子的拼接和 NSP、

场景、人物关系的标签,将拼接后的句子和标签、Mask 后的句子以及对句子对进行 0,1 位置表示的结果(A 句的词表示 0, B 句的词表示 1)组成的实例作为输入。在预训练时,针对以上 4 种标签,分别利用 BERT 模型的不同输出部件计算 loss, 然后进行梯度下降优化。根据它们的 loss,通过交叉熵求解预测的损失,四者相加,即为模型的 loss,如式(1)所示:

$$L(\theta, \theta_1, \theta_2, \theta_3, \theta_4) = L_1(\theta, \theta_1) + L_2(\theta, \theta_2) + L_3(\theta, \theta_3) + L_4(\theta, \theta_4) \quad (1)$$

其中, θ 为 BERT 中 Encoder 部分的参数; θ_1 为 MLM 任务中在 Encoder 上所接的输出层中的参数; θ_2 为句子相邻预测任务中 Encoder 接上的分类器参数; θ_3 和 θ_4 为新增的参数,代表句子场景预测任务中 Encoder 接上的分类器参数和句子人物预测任务中 Encoder 接上的分类器参数。

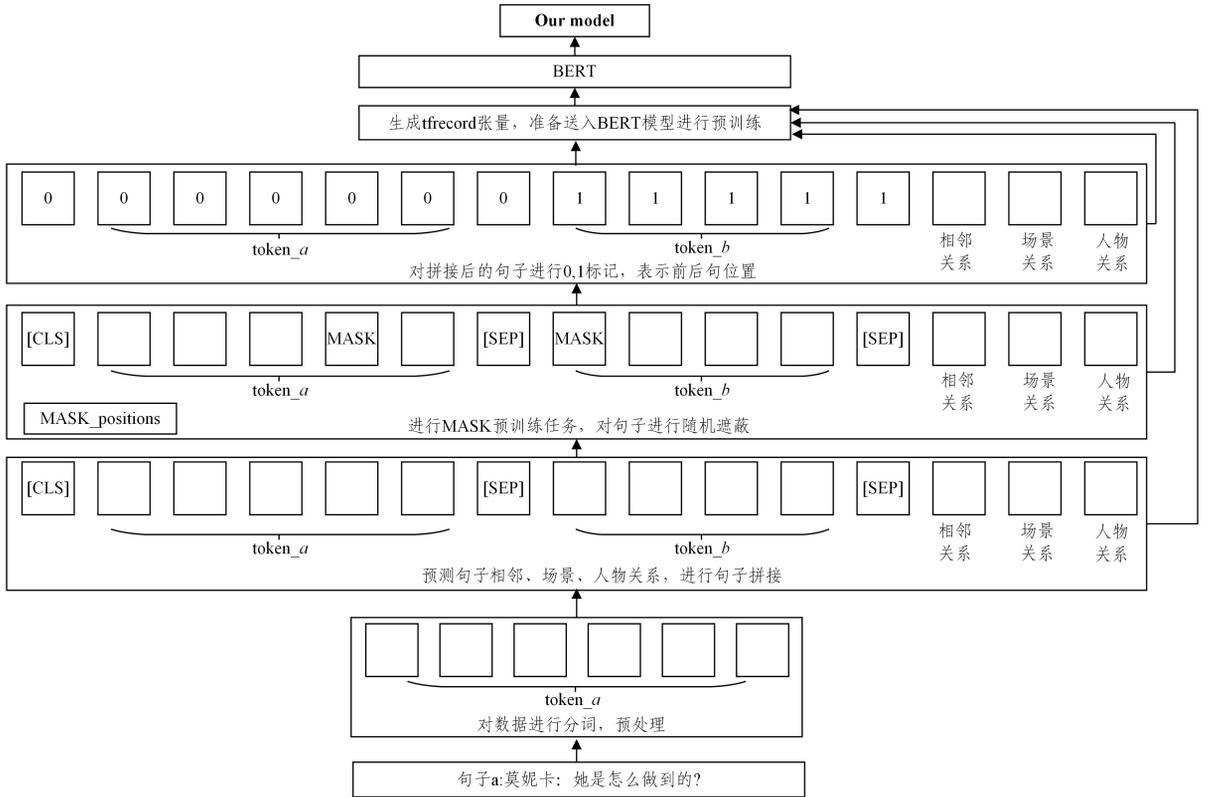


图 3 预训练模型生成流程图

Fig. 3 Flow chart of pre-training model generation

在第一部分的损失函数中,设 Mask 的词集合为 M ,因为它是一个 $|V|$ 大小词典上的多分类问题,所用的损失函数为负对数似然函数(且是最小化,等价于最大化对数似然函数),第二、三、四部分的损失函数也属于分类问题的损失函数,分别对应句子、场景、人物的预测。相关损失函数的计算如式(2)一式(5)所示:

$$L_1(\theta, \theta_1) = -\sum_{i=1}^M \log p(m = m_i | \theta, \theta_1) \quad (2)$$

其中, $m_i \in [1, 2, \dots, |V|]$ 。

$$L_2(\theta, \theta_2) = -\sum_{i=1}^N \log p(n = n_i | \theta, \theta_2) \quad (3)$$

其中, $n_i \in [IsNext, NotNext]$ 。

$$L_3(\theta, \theta_3) = -\sum_{i=1}^N \log p(n = n_i | \theta, \theta_3) \quad (4)$$

其中, $n_i \in [IsSecene, NotSecene]$ 。

$$L_4(\theta, \theta_4) = -\sum_{i=1}^N \log p(n = n_i | \theta, \theta_4) \quad (5)$$

其中, $n_i \in [IsPeople, NotPeople]$ 。

2.2 预训练模型的训练与句间相似度的计算

由 2.1 节可以得到我们的预训练模型,使用此模型计算句间相似度。在这一步,训练输入每条数据为两句话,每个输入对应一个标签,标记两句话是否相邻。相邻则标记为 1,不相邻则为 0。模型首先将输入转换成词向量、文本向量和位置向量,文本向量的取值在模型训练过程中自动学习,用于刻画文本的全局语义信息,并与单词的语义信息相融合。位置向量用于区分不同位置的词。将字向量、文本向量和位置向量的加和作为模型的输入,之后通过 Sigmoid 激活函数对输出向量进行处理,输出一个 $0 \sim 1$ 的概率值作为句间相似度,这是之后无监督抽取式摘要模型性能提高的

关键。具体模型图如图4所示。

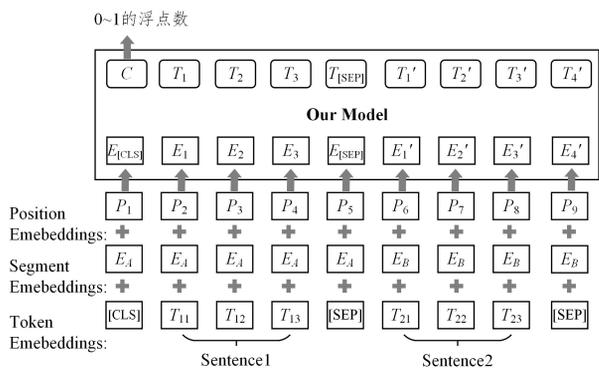


图4 句间相似度计算模型图

Fig. 4 Diagram of calculation model of similarity between sentences

2.3 无监督剧本摘要

将2.2节得到的每对句子的句间相似度作为抽取摘要中两个句子构成的边的权值 W_{ji} 。本文使用TextRank模型提取摘要,TextRank算法是由网页重要性排序算法PageRank^[18]迁移而来的。TextRank算法将词视为“万维网上的节点”,根据词之间的共现关系计算每个词的重要性,并将PageRank中的有向无权边变为无向有权边。首先将文本分割成 $T=[S_1, S_2, \dots, S_m]$,对于每个句子 $S_i \in T$,对其进行分词和词性标注;然后根据停用词表剔除停用词,只保留指定词性的词,如名词、动词、形容词等,即 $S_i=[t_{i,1}, t_{i,2}, \dots, t_{i,n}]$,其中 $t_{i,j}$ 为句子 i 中保留的候选关键词,构建词图 $G=(V, E)$, V 为节点集合,由以上步骤生成的词组成;然后采用共现关系构造任意两个节点之间的边。将权值 W_{ji} 代入式(6)(迭代传播权重)计算各句子的得分,直到收敛。

$$W_s(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{W_{ji}}{\sum_{V_k \in Out(V_j)} W_{jk}} W_s(V_j) \quad (6)$$

将得到的句子得分进行倒序排序,根据字数要求,抽取重要度最高的 M 个句子作为候选摘要句。

在本文中对比的标准摘要字数平均为100,所以我们限制实验获取的摘要不少于100个关键词。从候选摘要句中选取句子,直到组成的摘要句子的词数不少于100个。

3 实验及评价

3.1 实验数据集

本文所用的实验数据集是剧本语料《老友记》,共有228集,即有228篇原文本,每个原文本大概有800句话,每一集包含多个以人物对话组成的场景模块,对应的标准摘要是从Wikipedia网站¹⁾收集得到,文本拼写等错误已被人工修改。

数据集共有3个部分,分别是预训练数据集、训练集和测试集。其中,预训练数据集是对除去测试集的所有句子进行预测,选择合适的下一句组成句子对,是送入Bert进行预训练的数据,具体作用在2.1节已详细介绍。训练集是在100集(98-170,201-228)中均匀随机挑选的85111个句子对,是本文提出的模型计算句间相似度时所需要的训练数据(每

一个输入是以[SEP]进行连接的句子对)。测试集是171-200集的所有句子对。本文提出的模型经过训练后计算出测试集每个句子对的句间相似度,将其用于无监督摘要模型中进行实验,以获得171-200集的摘要,具体已在2.2节和2.3节详细介绍。表1列出了数据集中所用句子对个数的统计信息。数据集是剧本中包含场景描述信息或者体现人物情感特点的对话句子,场景描述是例如“[场景:中央公园,罗斯和莫妮卡正在看着菲比睡觉。]”的句子,人物对话是例如“莫妮卡:她是怎么做到的?”的句子,本文中会对数据进行进一步处理,具体形式将在实验设置中展示。组成的句子对有以下几种情况:1)相邻人物的对话;2)不相邻人物的对话;3)场景描述及临接的人物话语;4)场景描述及不临接的人物话语。数据集选取的句子长度(包含词的个数)大于7,舍去了过短的句子,因为有些句子如“哇哇哇哇!”“哈哈哈哈哈…”等短句子包含的信息过少,不足以提供有用的信息,无法构成摘要句。在本文提出的实验以及对比实验中,所用数据集均一样。

表1 数据集中句子的统计信息

Table 1 Sentence statistics in data sets

| | 篇数/篇 | 选取句子对个数/对 |
|------|------|-----------|
| 预训练集 | 198 | 227387 |
| 训练集 | 100 | 85111 |
| 测试集 | 30 | 2624500 |

3.2 评价方法

本文使用ROUGE作为无监督抽取式摘要的评价标准。针对人工评价成本高、耗时长和主观性强等缺点,Lin于2004年提出了ROUGE评价准则^[19]。其基本思想为:首先通过多个专家分别获得人工摘要,由摘要构成标准摘要集,然后将系统抽取的摘要与专家的标准摘要进行对比,通过统计二者之间重叠基本单元(n 元语法、词序列和词对)的数目,来评价摘要的质量。ROUGE评价方法现已成为摘要评价技术的通用标准之一。

本文主要通过ROUGE-1(单字组)、ROUGE-2(双字组)和ROUGE-L(最长公共子序列)来评测实验效果。

3.3 对照实验

本文共设立了10组对照实验。TextRank是本文的Baseline模型。对比模型中有监督(Supervised Learning, SL)摘要模型包括LSTM, BERT, UniLM^[20]和Seq2seq^[21],无监督(Unsupervised Learning, USL)模型包括LSTM, BERT, BERT+APTS, BERT+AFTP和BERT+APTS+AFTP(最后3个为本文提出的模型)。无监督实验中,先使用模型计算句间相似度,然后结合TextRank(TR)模型抽取得分高的句子作为摘要。

TextRank: Baseline模型。该模型是基于图排序的算法,即将文本中每一句话都视为图中的一个节点,通过计算句间相似度的方式给句子打分,并通过迭代的方式使句间相似度矩阵趋于平稳,最后选择得分最高的几句作为生成的摘要句。

LSTM: 抽取式摘要模型,由两个双向LSTM组成。分别对待测的两个句子进行编码,得到待测句子的表示向量,最后

¹⁾ <https://en.jinzhao.wiki/wiki/Friends>

基于得到的两个向量用 Softmax 函数计算两个分布之间的距离表示相似度, 值在 0~1 之间。

BERT:抽取式摘要模型, 模型没有考虑剧本间的场景信息以及人物信息。利用 BERT 作为训练器, 最后得到一个二分类概率分布。结果为一个 0~1 之间的浮点数, 表示句间相似度。

LSTM(SL):有监督抽取式摘要模型。利用 LSTM 网络作为训练器, 将训练好的神经元经过 Dense 层后得到一个二分类概率分布, 最后选取概率最高的几句话作为摘要句。

BERT(SL):有监督抽取式摘要模型。利用 BERT 网络作为训练器, 使用 Transformer 作为算法的主要框架, 通过一个二分类的线性层得到预测的值, 最后选取概率最高的几句话作为摘要句。

Unilm:有监督摘要模型。和 BERT 类似, 而且新增了 Seq2Seq 训练方式, 将接收原文本作为输入序列, 将摘要作为输出序列。

Seq2Seq:有监督摘要模型。属于 Encoder-Decoder 的一种, 思想就是利用两个 RNN: 一个 RNN 作为 Encoder, 另一个 RNN 作为 Decoder。

BERT+APTS+APTP:本文所提出的效果最好的用于无监督抽取式摘要模型。在预训练中增加了两个预训练任务来预测下文是否和第一句来自同一个场景和人物, 然后将预训练生成的模型作为训练器, 最后得到一个二分类概率分布。结果为一个 0~1 之间的浮点数, 表示句间相似度。

BERT+APTP:本文所提出的用于无监督抽取式摘要模型。仅增加一个预训练任务, 预测两句话是否来自同一人物。

BERT+APTS:本文所提出的用于无监督抽取式摘要模型。仅增加一个预训练任务, 预测两句话是否来自同一场景。

3.4 实验设置

3.4.1 文本预处理

在预处理时, 为了使人物所占长度一致, 本文将每句话对应的集数和人物名字通过字符串连接后放在了每句话的前面。集数占 3 个字符, 人物名字占 10 个字符, 如果字符数不够则用 '*' 来代替。舍弃了文本中长度过短的句子, 处理后的输入如图 5 所示。

```
228*****雷切尔雷切尔: 好的。 你和我, 好吗? 就是这个。
228*****罗斯罗斯: 就是这样。 除非我们休息。
228*****菲菲菲菲: 等等, 他吃粉笔吗?
228*****罗斯罗斯: 现在不要开玩笑了。
```

图 5 处理后的数据展示

Fig. 5 Display of processed data

3.4.2 实验参数设置

实验的参数设置如表 2 所列。

表 2 实验参数设置

Table 2 Experimental parameter setting

| 参数 | 取值 |
|-------------------------|-----|
| HIDDEN_SIZE | 768 |
| MAX_SEQ_LENGTH | 60 |
| BATCH_SIZE | 32 |
| MAX_POSITION_EMBEDDINGS | 128 |
| EPOCHS | 3 |

3.5 实验结果及分析

本节中首先将本文提出的基于预训练模型结构的模型与之前所述的模型结果进行对比; 然后在保证参数、数据不变的情况下, 分析所提模型中不同侧重部分以及整体对结果的影响; 最后通过摘要的实例样例来对比和分析加入模型各部分对抽取摘要质量的影响。

3.5.1 与现有系统比较的结果

本文模型与其他基准模型对比的结果如表 3 和表 4 所列。

表 3 无监督模型实验结果的对比

Table 3 Comparison of experimental results of unsupervised models

| Model | ROUGE-1 | ROUGE-2 | ROUGE-L |
|--------------------|---------|---------|---------|
| TextRank | 22.2 | 2.5 | 12.2 |
| LSTM+ TR | 22.4 | 1.8 | 10.9 |
| BERT+ TR | 23.2 | 1.8 | 11.5 |
| BERT+APTS+APTP+ TR | 26.2 | 3.9 | 14.0 |

表 4 有监督模型实验结果的对比

Table 4 Comparison of experimental results of supervised models

| Model | ROUGE-1 | ROUGE-2 | ROUGE-L |
|-------------------|---------|---------|---------|
| Seq2Seq | 15.4 | 0.9 | 10.4 |
| Unilm | 25.7 | 2.7 | 13.9 |
| BERT+APTS+APTP+TR | 26.2 | 3.9 | 14.0 |
| LSTM | 26.5 | 2.1 | 12.5 |
| BERT | 27.4 | 3.3 | 13.8 |

从表 3 可以看出, 本文提出的基于预训练模型的无监督抽取式摘要模型的实验结果远好于其他无监督学习模型, 在各项 ROUGE 评测指标上都有明显提升; 相比于无监督对比实验中性能最好的模型(没有新增预训练任务的 BERT), BERT+APTS+APTP+TR 在 ROUGE-1, ROUGE-2, ROUGE-L 上分别提高了 2.959%, 2.083% 和 2.49%。实验结果证实本文提出的模型是有效的。

从表 4 可以看出, 本文提出的基于预训练模型的无监督抽取式摘要模型的实验结果虽然在 ROUGE-1 上与 BERT 有监督抽取摘要模型有一定差距, 但与 Unilm, LSTM 已十分接近, 且本文提出的模型在 ROUGE-2 和 ROUGE-L 上均取得了最高值, 而且在各项 ROUGE 评测指标上都显著超过了 Seq2Seq 模型。从这个结果来看, 本文提出的无监督抽取式摘要模型是有效的。

3.5.2 模型中不同因素的影响比较

在模型相应参数、数据保持不变的情况下, 考虑不同方面的预训练任务对评估结果的影响。本节分析两个预训练任务分别加入以及一起加入时模型与 Baseline(TextRank)的对比结果, 如表 5 所列。可以看出, 模型在同时加入两个预训练任务时取得了最优的效果。

表 5 不同预训练任务时模型的实验结果

Table 5 Experimental results of different pre-training task models

| Model | ROUGE-1 | ROUGE-2 | ROUGE-L |
|----------------|---------|---------|---------|
| Baseline | 22.2 | 2.5 | 12.2 |
| BERT+APTP | 23.6 | 2.5 | 12.2 |
| BERT+APTS | 25.3 | 3.6 | 12.6 |
| BERT+APTS+APTP | 26.2 | 3.9 | 14.0 |

从表 5 中可以得知: 本文所提模型的实验结果均超过了 Baseline 的结果, 这说明本文采用的基于预训练模型的无

监督剧本摘要方法均是有效的。首先,我们考虑到了同一场景中的对话之间会有比较强的关联性,所以在预训练模型中加入了输入的两句话的场景预测。其次,考虑到不同人物说话所带的情感特点不同,我们加入了人物预测。实验结果表明:相比于只加人物预测的 BERT+APTP,只加场景预测的模型 BERT+APTS 在 ROUGE-1, ROUGE-2, ROUGE-L 上分别提高了 1.651%, 1.081% 和 0.393%。经分析,这主要是因为剧本中,场景中的每个人说话都是围绕一个主题来讨论,对话的关联性更强,捕获到的重要信息就会更多。

将二者结合的模型 BERT+APTS+APTP 相比于 Baseline,在 ROUGE-1, ROUGE-2, ROUGE-L 上分别提高了 3.961%, 1.477% 和 1.477%;相比于只加人物预测的 BERT+APTP,在 ROUGE-1, ROUGE-2, ROUGE-L 上分别提高了 2.566%, 1.383% 和 1.784%;相比于只加场景预测的 BERT+APTS,在 ROUGE-1, ROUGE-2, ROUGE-L 上分别提高了 0.915%, 0.302% 和 1.391%。这说明将句子之间的场景关系以及人物关系加入到预训练任务中有很大的帮助,通过本文模型计算出来的句间相似度对无监督抽取式剧本摘要的质量有一定的提升。

3.5.3 抽取摘要分析

为了更好地展示抽取摘要的结果,本节将具体展示各个模型所得到的摘要并对其进行比较和分析。将利用本文模型计算句间相似度后获得的摘要与基准模型做对比,每个模型获得的摘要如表 6 所列。为了便于对得到的摘要进行比较,将抽取摘要中的重要信息使用粗体突出显示。

表 6 不同模型摘要结果的对比

Table 6 Comparison of summary results of different models

| 模型 | 摘要 |
|-----------|---|
| 标准 | 菲比和莫妮卡为瑞秋的妈妈洗澡做准备,忘记邀请瑞秋的妈妈。洗澡时,莫妮卡拼命想寻求桑德拉的原谅。瑞秋意识到她不知道如何照顾孩子出生后,她的母亲坚持要陪她八周,这让她和罗斯很烦恼。罗斯最终让瑞秋有信心成为一名母亲。乔伊为一个名为《Bamboozled》的新游戏节目的主持人试镜,钱德勒和罗斯帮助他练习。 |
| Baseline | 格林夫人:瑞秋!夫人。格林:你好,莫妮卡。格林:嗯,瑞秋需要帮助照顾婴儿。[场景:莫妮卡和钱德勒家,婴儿洗澡结束了,除了格林夫人,其他人都离开了][场景:莫妮卡和钱德勒,莫妮卡和非比正在为瑞秋的妈妈洗澡做准备。][场景:莫妮卡和钱德勒,瑞秋正在给婴儿洗澡。] |
| APTP | [场景:罗斯和瑞秋家,罗斯正在帮助瑞秋研究宝宝什么时候会出生。]雷切尔:什么?你的意思是他们不会参加一个没有男人也没有酒的社交活动?等一下!如果你负责邀请,为什么我要打电话给她妈妈——格林夫人你好!嗨,我是莫妮卡·盖勒。乔伊:是的,如果我在白天得到它,我会。 |
| APTS | [场景:莫妮卡和钱德勒,莫妮卡和非比正在为瑞秋的妈妈洗澡做准备。][场景:乔伊的公寓,乔伊正在阅读剧本,罗斯和钱德勒拿着篮球走进来。]罗斯:你一定疯了!你会成为一个可怕的母亲!格林夫人:不,你留下来。我会让自己出去。就好像我不在这里。罗斯:每个人第一次当妈妈都有这种感觉。你会学会的。你知道你第一次来这座城市是什么时候吗? |
| APTS+APTP | [场景:乔伊公寓,乔伊正在阅读剧本,罗斯和钱德勒拿着篮球走进来。][场景:莫妮卡和钱德勒,瑞秋正在给婴儿洗澡。][场景:乔伊在试镜,正在准备节目。]罗斯:每个人第一次当妈妈都有这种感觉。你会学会的。[场景:乔伊的公寓,伙计们在玩游戏,只是每个人都投入。][场景:格林夫人正在告诉罗斯需要做些什么来保证他的公寓的婴儿安全。] |

从表 6 中可以看出,Baseline 模型抽取的句子组成的摘要只包括剧本中的部分信息,有很多冗余信息,无法涵盖整体的重要部分;在加入预测人物的预训练任务后,因为模型考虑到了人物说话的情感特征,人物的丰富度得到了明显提高,但是关于人物的冗余信息增多;在单独加入场景预训练任务后,主要信息涵盖增多,主题鲜明,基本可以看出主要信息;当同时加上两个预训练任务时,结果明显优于 Baseline,并且有更多的重要信息出现,抽取出的句子质量更高,可读性更强,这证明本文提出的方法是有效的。

结束语 本文针对剧本抽取句子作为摘要的问题,考虑到文本中不同人物说话的情感特性以及不同场景对话都会有密切的联系,提出了一种基于预训练模型的无监督剧本摘要。通过增加预训练任务,基于一定的规则预测每个对话的场景关系和人物关系,结合 BERT 模型进行预训练,最终得到需要的模型。利用模型作为训练器计算句间相似度,进一步提高了句子间相似度的精确性。实验证明,相比其他基准模型,所提模型效果提升显著,能提高无监督剧本摘要的质量。

参考文献

- [1] DEVLIN J, CHANG M W, LEE K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv:1810.04805, 2018.
- [2] MIHALCEA R, TARAU P. Textrank: Bringing order into text [C]// Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing. 2004:404-411.
- [3] LIU Y. Fine-tune BERT for extractivesummarization[J]. arXiv:1903.10318, 2019.
- [4] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Advances in Neural Information Processing Systems. 2017:5998-6008.
- [5] KANO R, MIURA Y, TANIGUCHI T, et al. Identifying Implicit Quotes for Unsupervised Extractive Summarization of Conversations[C]// Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing. 2020:291-302.
- [6] PAPALAMPIDI P, KELLER F, FRERMANN L, et al. Screenplay Summarization Using Latent Narrative Structure[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020:1920-1933.
- [7] ZHOU Q, WEI F, ZHOU M. At Which Level Should We Extract? An Empirical Analysis on Extractive Document Summarization[C]// Proceedings of the 28th International Conference on Computational Linguistics. 2020:5617-5628.
- [8] FENG X, FENG X, QIN L, et al. Language model as an annotator: Exploring dialogpt for dialogue summarization[J]. arXiv:2105.12544, 2021.
- [9] ZOU Y, ZHU B, HU X, et al. Low-Resource Dialogue Summarization with Domain-Agnostic Multi-Source Pretraining[J]. arXiv:2109.04080, 2021.

- [10] CHEN J, YANG D. Simple Conversational Data Augmentation for Semi-supervised Abstractive Dialogue Summarization[C]// Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 2021;6605-6616.
- [11] ZHAO L, ZENG W, XU W, et al. Give the Truth: Incorporate Semantic Slot into Abstractive Dialogue Summarization[C]// Findings of the Association for Computational Linguistics: EMNLP. 2021;2435-2446.
- [12] ZOU Y, ZHAO L, KANG Y, et al. Topic-oriented spoken dialogue summarization for customer service with saliency-aware topic modeling[J]. arXiv;2012. 07311, 2020.
- [13] DAI A M, LE Q V. Semi-supervised sequence learning[J]. Advances in Neural Information Processing Systems, 2015, 28: 3079-3087.
- [14] ZHANG H, CAI J, XU J, et al. Pretraining-Based Natural Language Generation for Text Summarization[C]// Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL). 2019;789-797.
- [15] 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). 2019;3730-3740.
- [16] LI R N. Research on Semantic-based Text Similarity Calculation Method[D]. Beijing: Beijing University of Technology, 2018.
- [17] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [18] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web[R]. Stanford InfoLab, 1999.
- [19] LIN C Y. Rouge: A package for automatic evaluation of summaries[C]// Text Summarization Branches Out. 2004;74-81.
- [20] DONG L, YANG N, WANG W, et al. Unified language model pre-training for natural language understanding and generation [C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019;13063-13075.
- [21] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]// Advances in Neural Information Processing Systems. 2014;3104-3112.



SU Qi, born in 1997, postgraduate, is a member of China Computer Federation. Her main research interests include natural language processing and text summarization.



WANG Hongling, born in 1975, assistant professor, is a member of China Computer Federation. Her main research interests include natural language processing and information retrieval.

(责任编辑:柯颖)