

基于抽取的高考作文生成

冯晓骋 龚 恒 冷海涛 秦 兵 孙承杰 刘 挺

(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

摘 要 机器人自动写作是人工智能和自然语言处理领域重要的研究方向,然而传统的自动写作方法主要针对体育新闻、天气预报等较短的段落级文本进行研究,并没有对篇章级文本自动生成技术进行深入地建模.针对这一问题,我们着重研究面向高考作文的篇章级文本生成任务.具体而言我们提出了一种基于抽取式的高考作文生成模型,即先进行抽取再利用深度学习排序方法进行段落内部的文本组合生成.通过实际专家评测,我们所生成的作文能够达到北京高考二类卷平均分数,具有一定的实际应用价值.

关键词 文本生成;文本抽取;句子排序;作文生成;自然语言处理

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2020.00315

Extractive Essay Generation for College Entrance Examination

FENG Xiao-Cheng GONG Heng LENG Hai-Tao QIN Bing SUN Cheng-Jie LIU Ting

(Department of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001)

Abstract Automatic writing is an important research direction in the field of Artificial Intelligence and Natural Language Processing. However, the traditional automatic writing methods mainly focused on generating short text, such as sports news and weather forecast, and lack deep modeling of the automatic generation of discourse-level text. In this paper, we focus on the discourse-level text generation task oriented to the essay generation in College Entrance Examination. In particular, we present an extractive essay generation model for the College Entrance Examination. We formulate the task as essay generation from mind, namely taking the input as many topic words in mind and outputting an organized article (a document) with several paragraphs under the theme of the topic. The task is challenging as it requires the generator to deeply understand the way human beings write articles. In addition, after understanding the meaning of a topic word, the following challenge is how to generate a topic focused article, e. g. how to collect topic-specific “fuel” (e. g. sentences) and how to organize them to form an organized article. This is of great importance as an article is not a set of sentences chaotically. Natural language is structured and the coherence/discourse relationship between sentences is a crucial element to improve the readability of a document and to guarantee the structured nature of a document in terms of lexicalization and semantic. Hopefully, solving this problem contributes to making progress towards Artificial Intelligence. For the issues mentioned above, our proposed model consists of two major modules including sentence extraction module and paragraph generation module. First, in order to generate a high-quality essay, the extractive essay generation model needs to determine the focus of each paragraph. Therefore, we first expand the given topic with more related topic words. Then we

收稿日期:2018-10-17;在线出版日期:2019-07-16. 本课题得到国家重点研发计划(2018YFB1005103)和国家自然科学基金(61632011, 61772156)资助. 冯晓骋, 博士, 助理研究员, 中国计算机学会(CCF)会员, 主要研究方向为文本生成与信息抽取. E-mail: xcfeng@ir.hit.edu.cn. 龚 恒, 博士, 中国计算机学会(CCF)学生会会员, 主要研究方向为文本生成. 冷海涛, 硕士, 主要研究方向为文本生成. 秦 兵(通信作者), 博士, 教授, 中国计算机学会(CCF)会员, 主要研究领域为自然语言处理. E-mail: qinb@ir.hit.edu.cn. 孙承杰, 博士, 副教授, 中国计算机学会(CCF)会员, 主要研究方向为自然语言处理. 刘 挺, 博士, 教授, 国家“万人计划”领军人才, 中国计算机学会(CCF)会员, 主要研究领域为社会计算、信息检索和自然语言处理.

cluster them into multiple sets. Each of them represents the focus of each paragraph. Second, the model needs to find candidate sentences that are related to each paragraph's topic. Therefore, it first finds sentences that include the given topic words. Then, we propose two methods to expand the candidate sentences set with more diverse sentences. After obtaining sentences candidate sets for each paragraph, the model needs to choose and arrange sentences to be coherent paragraphs. In this paper, we explore three methods to achieve this. Based on experiments, paragraph generation via pointer network and paragraph generation via pair-wise LSTM network outperforms paragraph generation via learning to rank. In addition, paragraph generation via pointer network achieves best result among those three methods. Through experts' evaluation, the essay our proposed model produced can reach the average scores of level two of Beijing college entrance examination, which indicates that our proposed methods have certain practical application value. For each component in the framework, we explore several strategies and empirically compare between them in terms of qualitative or quantitative analysis. We also analyze the pros and cons of each approach. Although we run experiments on Chinese corpus, the method is language independent and can be easily adapted to other language.

Keywords text generation; text extraction; sentence ordering; essay generation; natural language processing

1 引 言

自然语言生成(NLG)是自然语言领域中重要并具有挑战性的任务^[1-3]. 相比自然语言生成领域里古诗文生成、回复生成等比较热门的任务, 高考作文的生成无疑更具有挑战性. 在古诗文生成以及回复生成任务中, 生成的信息相对较少, 而在高考作文生成任务中, 我们需要生成较多的信息. 不仅如此, 相比于古诗文生成, 高考作文生成中的语句对连贯度、语义清晰度的要求更高, 类似于诗歌中生成比较朦胧的语言是不可行的, 而相比于回复生成任务, 高考作文中的语句无疑更加丰富, 形式更加多变, 更加富有文采.

随着深度学习的发展, 很多深度学习方法被引入到了自然语言生成领域中, Wang 等人^[4]在 2016 年提出了基于 RNN(Recurrent Neural Network)和主题词信息的古诗文生成模型, Yin 等人^[5]在 2016 年提出了以 Encoder-Decoder 为基础并融入背景知识的回复生成模型. Feng 等人^[6]在 2018 年提出了根据话题词生成一段话的任务. Yang 等人^[7]通过引入外部的常识知识, 增强模型根据话题生成文本的能力. 但是对于作文生成任务, 直接采用端到端的深度学习方法目前来看并不是最好的选择, 由于高考作文的信息量太大, 直接使用端到端的深度学习方法

会出现严重的语义漂移问题, 而且在大篇幅的生成中, 会出现比较多的重复以及句内不通顺的问题, 而这些问题对于一篇高考作文而言, 无疑是严重的, 而我们提出了一种基于抽取式的高考作文生成模型来避免这样的问题.

在基于抽取式的文本生成研究领域中, Zhang 等人^[8]在 2016 年提出了以抽取为基础的体育新闻生成模型, 但其任务在给定的实况评论中抽取完成, 有时序作为依据, 且生成的内容风格较为固定, 因而生成难度相比高考作文要低. Zhang 等人^[8]在文本生成的任务中采用了 learning to rank 的排序方法. 本文则在此方法基础上进行了拓展, 提出了基于深度学习的文本组合生成方法.

高考作文生成领域的工作较少, 与我们相关的工作有 Li 等人^[9]提出的利用词汇链进行高考作文生成. 但是这篇工作主要侧重于构造词汇链表示子话题, 而我们的工作更加侧重于段落生成部分, 目的是生成连贯的文本.

在我们的模型中, 我们通过句子抽取以及段落生成两个步骤来得到一篇完整的高考作文. 抽取步骤中, 我们不仅进行了基于语义的抽取, 同时还在此基础上做了基于相关度的扩展, 以此来满足作文生成的需求, 而在段落生成的模块, 我们以 learning to rank 作为 baseline 方法, 尝试了基于深度学习的组合生成方法. 模型的输入, 是针对作文材料进行分析

后得到的主题词. 这部分工作我们借鉴了 Liu 等人^[10]在 2017 年的研究方法. 他们使用了两层的 RNN 网络来进行训练, 并使用知乎这样的外部数据进行迁移学习来解决作文语料本身不够的问

题. 整个抽取式作文生成模型框架如图 1 所示, 主要由句子抽取和段落生成两部分组成, 本文的方法将对抽取式作文生成模型中的各部分进行详细介绍.

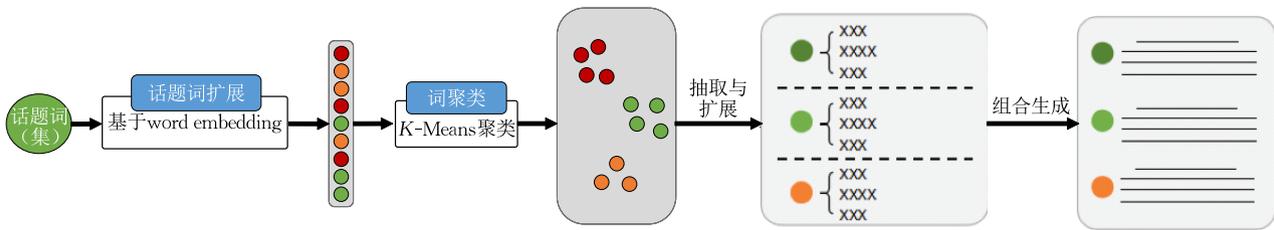


图 1 基于抽取的作文生成模型框架

本文做出的主要贡献如下:

(1) 本文提出了一种基于抽取组合的篇章级文本生成框架.

(2) 在抽取模块中, 提出了扩展相关句的方法来解决仅仅使用语义抽取带来的语义狭窄的问题.

(3) 将深度学习句子排序方法迁移到了段落内文本的组合生成.

2 问题描述

抽取式的高考作文生成模型包含了两个模块, 即句子抽取模块和段落生成模块. 在句子抽取模块中我们需要利用给定的主题抽取出合适的作文候选句, 而在段落生成模块中则通过使用句子排序的方法对句子抽取模块得到的结果进行组合, 以此得到高考作文.

对于作文的主题, 我们利用 Liu 等人^[10]在 2017 年的研究方法, 通过分析作文题目材料直接得到一篇作文的主题 t . 在句子抽取模块中, 从作文语料库中对给定的主题 t 进行抽取, 得到若干个段落的候选句集合 $S = \{S_1, S_2, \dots, S_n\}$, 其中 S_i 表示生成的作文中第 i 个段落的候选句子集合, 而 S 也是段落生成模块的输入.

在段落生成模块中, 对于给定的若干段落候选句子集合 S , 通过基于深度学习的句子排序方法得到生成的作文 $S' = \{S'_1, S'_2, \dots, S'_n\}$, 其中 $|S'|$ 代表生成的作文的段落个数, S'_i 则表示生成的作文中的第 i 个段落.

3 方法

3.1 句子抽取

在句子抽取模块中, 单纯使用语义表示(比如利

用 skip-gram 模型预训练的 word embedding)抽取作文候选句是不够的. 对于一篇高考作文, 其段落内部的语句虽然都是围绕着某一个主题来展开, 但是这些语句有丰富的语义层次感, 并不仅仅局限在主题上, 而普通的基于语义的抽取方法, 会使得抽取的句子集中在某一个狭窄的语义上, 使得这些句子本身语义不够丰富. 如果使用这样的句子集合来生成作文, 会显得作文是同一个语义的句子的堆砌. 基于此, 我们不仅需要得到相似的句子还需要得到相关的句子, 因而在句子抽取模块提出先基于语义抽取再扩展的方法, 从而使得抽取的作文候选句在语义层次上较为丰富. 对于扩展方法, 我们提出了基于关联规则以及基于 LDA (Latent Dirichlet Allocation)^[11]的句子扩展方法.

3.1.1 基于语义的句子抽取方法

利用 Liu 等人^[10]在 2017 年提出的方法, 分析作文题目材料得到主题词集合 t , 我们需要把它扩展成各段落的主题词集合 $T = \{T_1, T_2, \dots, T_n\}$. 我们使用 word embedding 以及余弦相似度截取与主题词 t 在语义上最相似的词语集合, 然后利用 K -Means 方法聚类, 便得到了 T , 每一个类对应作文的一个段落.

在一篇高考作文中, 开头段落与结尾段落和文章中的其他段落的语言形式有较大不同, 因而对于开头段落或结尾段落我们进行针对性处理, 即抽取来源的语句在原文本也只能是开头段落或者结尾段落. 而其他段落的抽取没有这样的限制. 而其他段落的抽取没有这样的限制. 由于各个段落的抽取是用同样的方法并行的, 这里我们对某一个段落的抽取进行展开介绍. 在第 i 个段落中, 对于得到的主题词集合 T_i , 我们利用语料库的倒排索引表, 对于任意词语 $w \in T_i$, 把包含 w 的所有句子均找出放入集合 U_i 中, 对于集合 U_i 中的每个句子, 我们使用 word embedding average

的方法表示其向量,同理对于主题词集合 T_i ,我们也使用相同的方式求出该集合的向量表示,最后利用余弦相似度,截取与主题词集合在语义上最相似的部分句子,便得到第 i 个段落的语义抽取结果 U'_i .

3.1.2 基于关联规则的句子扩展方法

在语料库中,某一句中出现的词语集合记为 $w = \{w_1, w_2, \dots, w_n\}$,它的后一句中出现的词语集

合记为 $w' = \{w'_1, w'_2, \dots, w'_n\}$,当语料库中概率 $P(w' | w, Corpus)$ 较大时,那么根据数据挖掘领域中的关联规则,我们相信词语集合 w 和 w' 在语义和逻辑上有一定关联.基于这一点,对于存在于 U'_i 的任意词语集合 w ,如果当 $P(w' | w, Corpus)$ 较大时,我们便将包含词语集合 w' 但是不在集合 U'_i 中的句子作为扩展得到的结果,由此最终得到 S_i . 具体示意图如图 2 所示.

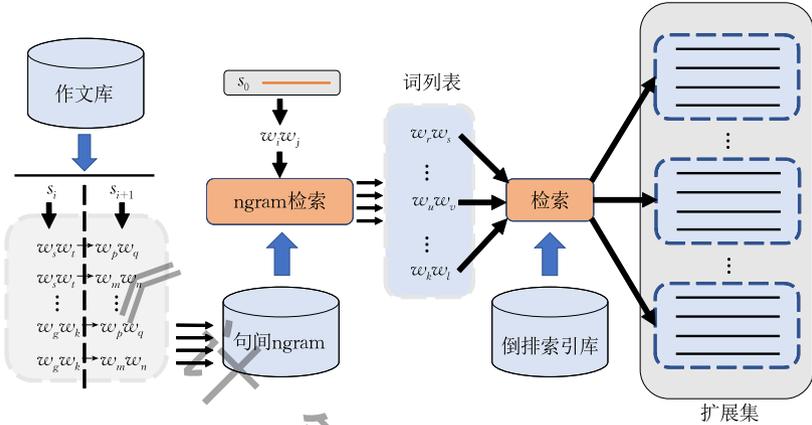


图 2 关联规则扩展示意图

3.1.3 基于 LDA 的句子扩展方法

为了使得段落中的语义信息足够丰富,我们考虑引入主题信息来帮助我们基于语义的句子抽取方法得到的句子进行扩展.基于此,我们引入了 LDA 模型,LDA 主题模型是“文本-主题-词”的三层贝叶斯模型,是一种非监督的学习技术.该模型在大规模语料库中能够有效地识别其潜在的主题分布.主题信息的存在可以令我们扩展得到的句子是在主题上密切相关的.

对第 i 段中基于语义抽取得到的句子集合 U'_i 中的任意一个句子 s_0 ,利用训练好的 LDA 模型,我们可以得到其在各个主题模型下的概率分布 $topic_{s_0}$,类似地,我们可以对于语料库中的其他句子进行类似地处理,从而得到它们在主题模型下的概率分布

$topic = \{topic_{s_1}, topic_{s_2}, \dots, topic_{s_n}\}$. JS (Jensen-Shannon) 散度可以用于度量两个概率分布的相似度,对于任意 $topic_{s_i} \in topic$,利用 JS 散度计算出和句子 s_0 在主题概率分布上的相似程度,并选最接近的一些句子作为扩展,从而得到 S_i ,具体示意图如图 3 所示.其中任意两个句子的相似性计算公式如式(1)所示.

$$D_{JS}(topic_{s_0}, topic_{s_i}) = \frac{1}{2} \left[D_{KL} \left(topic_{s_0}, \frac{topic_{s_0} + topic_{s_i}}{2} \right) + D_{KL} \left(topic_{s_i}, \frac{topic_{s_0} + topic_{s_i}}{2} \right) \right] \quad (1)$$

其中, $D_{KL}(\cdot)$ 是 KL (Kullback-Leibler) 散度计算公式.

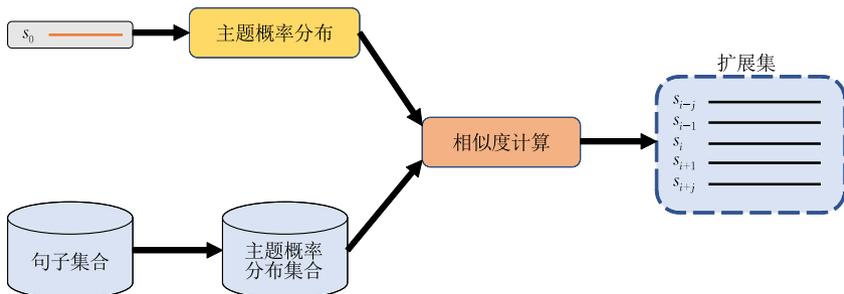


图 3 LDA 扩展示意图

3.2 段落生成

对于任意段落 i , 通过句子抽取模块我们可以得到一系列候选句子集合 S_i , S_i 中的句子来源于语料库中不同的篇章与段落, 因而我们很难简单地衡量这些句子之间的连贯性与逻辑性. 而对于衡量句子间的连贯性以及逻辑关系这样的问题, 基于深度学习的句子排序技术取得了不错的效果, 因而我们提出了基于深度学习句子排序的段落生成方法, 在连贯并具有逻辑性的范文上利用句子排序方法训练得到可以度量连贯性以及逻辑性的模型, 并将该模型应用到句子集合 S_i , 以此来衡量这些来源于不同篇章不同段落的语句之间的连贯性以及逻辑性. 在段落生成任务上, 我们使用了基于特征提取的 learning to rank^[12] 方法, 基于神经网络的 pair wise 方法^[13] 以及 Ptr-Net(Pointer Network)方法^[14-16], 其中 learning to rank 方法作为我们的 baseline 方法. 相比于 baseline 方法, 两种深度学习方法都取得了不错的结果, 其中 Ptr-Net 效果最好.

3.2.1 基于 learning to rank 方法的段落生成

在该模型中, 我们通过分析高考作文语料, 提出了一些在高考作文写作时比较重要的特征, 作为我们考虑的关键. 对于每个高考作文句, 我们在其每个特征上进行打分, 然后选择在当前情况下综合分数最高的句子作为下一句. 由于每个句子都考察多个特征, 因而需要考虑各个特征分数的权重分配. 句子的综合分数计算公式如式(2)所示.

$$score(s) = \sum_{i=1}^{|\alpha|} \alpha_i f_i(s) \quad (2)$$

其中 $score(s)$ 代表了该句子综合各个特征的分数的, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ 代表了各个特征分数之间的权重分配, 而 $f = [f_1, f_2, \dots, f_n]$ 则代表了在各个特征上的打分函数. 我们使用 learning to rank 的方法来学习各个特征分数之间的权重分配, 即向量 α .

在该模型中, 我们考虑如下特征:

(1) 作文句的篇章角色^[17]. 在高考作文中, 每个句子在篇章结构中都担任着一定的角色, 比如论点句、分论点句、事例句等等, 而这些结构特征在排序的时候无疑起着非常重要的作用.

(2) 与作文主题的相似度. 利用 word embedding average 的方法向量化作文句以及作文主题然后计算余弦相似度.

(3) 与当前句子的相似度. 利用 word embedding average 方法向量化高考作文句, 然后计算与当前句子之间的余弦相似度.

(4) 与作文主题的相关程度. 利用 LDA 模型向量化作文句以及作文主题, 利用 JS 散度公式计算两者的距离.

(5) 与当前句子的相关程度. 利用 LDA 模型向量化高考作文句, 然后利用 JS 散度公式计算与当前句子的距离.

(6) 与当前句子的公用词. 去除停用词以后, 考量和当前句子之间公共词语所占的比例.

(7) 与当前句子的连贯度. 参考了 Li 和 Hovy^[18] 在 2014 年提出的连贯度模型, 在模型中, 先利用 Recursive Neural Networks 作为句子 encoder, 接着把窗口内的句子向量进行拼接, 再通过两层的神经网络来预测窗口内的句子是否是连贯的, 预测的值是一个 0~1 之间的浮点数, 数值越大代表连贯度越高.

通过 learning to rank 方法, 可以学习到上述特征之间的权重分配. 而排序过程则是一个贪心的过程, 每次都以前一个句子作为基础, 选择分数最高的句子作为下一个句子.

3.2.2 基于 pair wise 的深度学习模型

如图 4 所示, 该模型是一个基于深度学习的 pair wise 模型, 即在该模型中, 我们的主要工作是预测一个句子对 (s_i, s_j) 的相对顺序. 我们使用

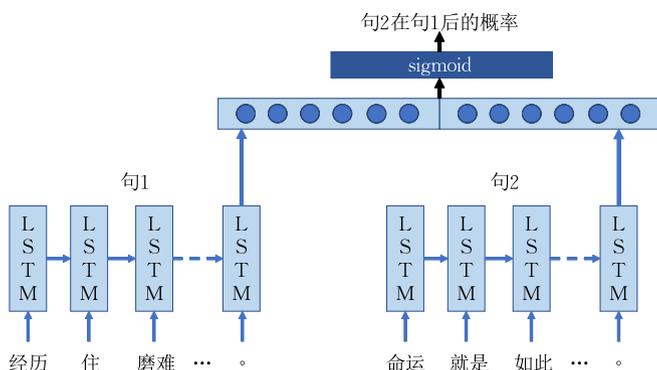


图 4 pair wise 深度学习方法

LSTM(Long Short-Term Memory)神经网络作为句子 encoder,将融入了前文信息的每句话最后一个词对应的 LSTM 神经网络的隐含层表示作为整句话的表示,最后使用 sigmoid 函数来预测句子对 (s_i, s_j) 相对顺序. 使用 LSTM 神经网络作为句子 encoder 的细节,在 3.2.3 节有详细地介绍. 对于句子对 (s_i, s_j) ,我们在句子 encoder 部分,得到句子 s_i 以及 s_j 的向量表示,将两个句子向量进行拼接,利用新得到的向量预测句子对的相对顺序. 模型可以用如下式(3)和式(4)表示.

$$\mathbf{h}_{ij} = \varnothing(\mathbf{W}_h^T(\mathbf{e}_i \oplus \mathbf{e}_j) + \mathbf{b}_h) \quad (3)$$

$$p_{ij} = \sigma(\mathbf{W}_p^T \mathbf{h}_{ij} + b_p) \quad (4)$$

其中 \mathbf{e}_i 和 \mathbf{e}_j 代表了句子对中 s_i 和 s_j 的句子向量. $\varnothing(\cdot)$ 是 tanh 函数, σ 是 sigmoid 函数, $\mathbf{W}_h, \mathbf{b}_h, \mathbf{W}_p, b_p$ 均为可训练的参数.

p_{ij} 表示了句子对 (s_i, s_j) 中 s_i 在 s_j 前的概率. 对

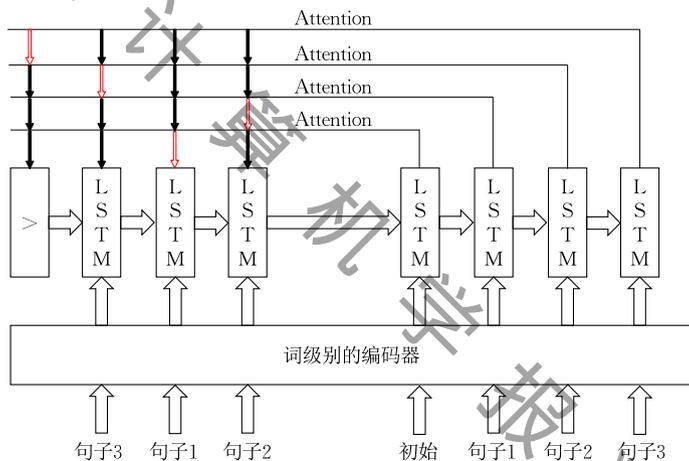


图 5 Ptr-Net 方法

Ptr-Net 主要由句子 encoder、encoder、decoder 三部分构成,下面将会详细介绍.

句子 Encoder

在句子 encoder 部分,我们使用 LSTM 神经网络对高考作文句进行向量化. 句子 encoder 把句子 s 所包含词语按序输入,利用 LSTM 神经网络计算出整个句子的表示. 因而如果用 $|s|$ 表示句子 s 中词语的个数,那么显然句子 s 的表示为 $\mathbf{h}_{|s|}$.

Encoder

encoder 部分同样采用了 LSTM 神经网络,和句子 encoder 不同,句子 encoder 在每个节点上的输入是词语的 embedding,而 encoder 部分,每个节点的输入是句子的表示,因而在 encoder 部分,我们最后得到是文段的表示. 如果一个文段总共有 n 个句子,那么文段的表示如式(7)所示.

$$\mathbf{e}_i = \text{encoder}(s_i, \mathbf{e}_{i-1}), i = (1, \dots, n) \quad (7)$$

于排序分数,可以用式(5)和式(6)表示.

$$\text{score}(s, o, i, j) = \log p_{o_i, o_j} \quad (5)$$

$$\text{score}(s, o) = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{score}(s, o, i, j) \quad (6)$$

其中 s 代表句子集合, o 代表一个指定的顺序,而 o_i 则代表了在这个顺序中的第 i 个句子, $\text{score}(s, o, i, j)$ 代表句子对 (s_{o_i}, s_{o_j}) 的得分, $\text{score}(s, o)$ 则代表了对指定顺序 o 的打分. 在预测的过程中,我们使用 beam search 的方法来得到预测的句子顺序.

3.2.3 基于 Ptr-Net 的深度学习模型

前面的工作,无论是 learning to rank 方法,还是 pair wise 模型,都有比较大的限制, learning to rank 方法受限于所采用的特征,而 pair wise 受限于缺少文段的前后信息. 所以在前两者的基础上,我们使用了如图 5 的基于 Ptr-Net 模型的段落生成方法.

其中 s_i 是第 i 个句子 embedding,同时,我们将 encoder 的初始值 \mathbf{e}_0 设为 0.

Decoder

对于 Ptr-Net 的 decoder,我们同样使用了 LSTM. 不同之处在于,decoder 中我们利用了 attention 的思想,每一步我们都需要用当前隐层向量与 encoder 部分各个节点的向量进行计算,计算过程如式(8)和式(9)所示. 其中, \mathbf{e}_j 代表 encoder 得到的第 j 个句子的表示, \mathbf{d}_i 代表 decoder 选择第 i 个句子时, decoder 的隐含层表示, \mathbf{W}^T 为待训练的权值矩阵, \mathbf{v}^T 为待训练的权值向量.

$$\mathbf{u}_j^i = \mathbf{v}^T \tanh \left(\mathbf{W}^T \begin{bmatrix} \mathbf{e}_j \\ \mathbf{d}_i \end{bmatrix} \right), j = (1, \dots, n) \quad (8)$$

$$P(o_i | o_{i-1}, \dots, o_1, \mathbf{s}) = \text{softmax}(\mathbf{u}^i)_{o_i} \quad (9)$$

softmax 计算后得到的概率分布为各个句子成为下一句的概率.

4 实验

4.1 句子抽取

对于这一模块的实验我们没有找到非常合适的自动化评价指标,因而对于这一部分的实验结果更多的是人为的分析比较.如图 6 和图 7 所示,是使用我们的方法分别针对话题词“诚信”和“爱国”在作文语料库中进行抽取的结果.

仅采用基于语义的句子抽取方法得到的句子虽

然语义紧扣话题,但是抽取出来的句子句式单一、语义重复,不足以构成一个段落.基于关联规则的句子拓展方法获得的结果语义虽有拓展,但以“诚信”为话题词抽取的结果出现了偏移的情况,“诚信”和“爱国”为话题抽取的结果均缺少具体的事例.而基于 LDA 的句子扩展方法获得的句子语义和句式较为丰富,还引入了一些事例.这部分句子由于句内有很多叙述性的词语或者甚至没有明确出现和“诚信”或“爱国”相关的词语,很难直接通过基于语义的句子抽取方法得到.

<p>诚信,顾名思义就是做人要诚实守信. 诚信是立身处世的准则,是人格的体现,是衡量个人品行优劣的道德标准之一. 诚信是中华民族的传统美德.</p> <p style="text-align: center;">基于语义的句子抽取方法</p>
<p>道德:社会意识形态之一,是人们共同生及其行为的准则和规范,道德通过社会的或一定阶级的舆论,对社会生活起的约束作用. 古人有“慎独”之说,意思是说当一个人独处的时候,最容易做出违反他做人准则的事情. 无论天荒地老,无论沧海桑田,是忠孝之人的赤子之心,见证了中华文明五千年血脉相随的道义,而我们也会将它这样传承下去,因为我们都是炎黄子孙.</p> <p style="text-align: center;">基于关联规则的句子扩展方法</p>
<p>汉朝的季步以真诚守信而为世人所赞扬,当时的人都传颂着这一句话. 想必“狼来了”的故事对大家来说并不陌生,一个小孩,单纯的放牧小孩,为了图一时开心三番两次欺骗山下的村民狼来了,从而博取人们的同情,以同情来开玩笑,导致人们对这小孩产生不信任的态度. 晏殊的诚实品质值得我们学习,他以诚实的珍贵品质取得人们的信任.</p> <p style="text-align: center;">基于 LDA 的句子扩展方法</p>

图 6 “诚信”抽取结果

<p>这时,许多爱国人士挺身而出,为了保护这条养育自己的土地,奋勇抗敌! 将爱国进行到底! 有多少英雄儿女,爱国志士,为了祖国,为了正义,宁死不屈.</p> <p style="text-align: center;">基于语义的句子抽取方法</p>
<p>他们抛头颅、洒热血,他们不畏强暴、不屈不挠、同仇敌忾、浴血奋战把我们伟大的祖国从水深火热中解放出来,他们是永垂不朽的人们. 他们将自己的满腔热血洒在了这片土地上,他们将自己的赤胆忠心献给了祖国,他们用自己的生命报效祖国,他们是那样的无怨无悔,他们的事迹是那样的可歌可泣. 许多正值青春的人,来到祖国的边疆,保卫着祖国,他们把青春献给了祖国.</p> <p style="text-align: center;">基于关联规则的句子扩展方法</p>
<p>刘胡兰的英勇就义,黄继光的前仆后继,邱少云的舍身取义…… 宋有挂帅征战抛洒一腔热血保家卫国的穆桂英; 现代有勇于献出生命拯救祖国的英雄人物,邓世昌献身大海,为祖国浴血奋战,在红军长征万里时,这些无名战士为祖国贡献自己的生命,来完成使命.</p> <p style="text-align: center;">基于 LDA 的句子扩展方法</p>

图 7 “爱国”抽取结果

4.2 段落生成

4.2.1 实验数据

我们从互联网上爬取了 6686 篇高中议论文作为我们的实验数据.而所有的实验都是以段落为单位进行的,即在段落上考察模型的效果,同时在前期处理数据的时候,已经过滤了段落中句子数量小于 3 句的段落.数据集的详细信息见表 1.

表 1 作文数据情况

类别	段落数量	句子数量	各段句子平均数量
训练集	27351	145583	5.32
开发集	719	4443	6.17
测试集	721	3994	5.54

4.2.2 实验方法

learning to rank 方法

在该方法中我们使用了 Ranklib 工具包来完成

我们的实验,对于优化算法我们选择了 Coordinate Ascent,而优化指标我们选了 NDCG (Normalized Discounted Cumulative Gain)^[19].

Pair Wise 的深度学习方法

在该方法中,对于段落中任意连续的句子对,我们把它看作一个正例,并把它的反序看作一个反例,使用式(10)所示的损失函数来进行训练.

$$J = -\frac{1}{2m} \sum_{i=1}^{2m} y^i \log p_{x^i} + (1-y^i) \log(1-p_{x^i}) \quad (10)$$

其中 x^i 代表 $2m$ 个训练数据中的第 i 个句子对,而 p_{x^i} 则表示模型对该句子对计算出的概率.

Ptr-Net 方法

在该方法中,我们对于每一个段落打乱其原本的顺序作为训练输入,而使用正确的顺序作为训练目标,在训练的时候使用了式(11)所示的最大似然的方法来优化.

$$\max_{y \in D} \sum_{t=1}^{|y|} \log p(y^t | y^{t-1}, \dots, y^1) \quad (11)$$

其中, D 代表了所有训练语料集, y 则代表了每个训练实例中的正确顺序.

参数设置方面, batch 的大小设置为 128, LSTM 的隐层大小为 200, 句子 encoder 部分, 采用两层 LSTM.

4.2.3 实验指标

在实验指标上,我们参考了 pair wise 深度学习方法的提出者在论文中使用的评价指标,并结合我们的具体任务,使用了 *Rouge-S*、*Rouge-N*、完全正确率以及位置正确率.

在实验中,使用 *Rouge-S* 这个指标来考察各个模型在 skip-bigram 上的处理情况.

$$Rouge-S = \frac{1}{M} \sum_{m=1}^M \frac{|S(s^m, \hat{o}^m) \cap S(s^m, o^{m*})|}{|S(s^m, o^{m*})|} \quad (12)$$

$$Rouge-N = \frac{1}{M} \sum_{m=1}^M \frac{|N(s^m, \hat{o}^m) \cap N(s^m, o^{m*})|}{|N(s^m, o^{m*})|} \quad (13)$$

其中 $S(\cdot)$ 代表了所有的 skip-bigram 句子对, $N(\cdot)$ 代表了给出正确顺序中的所有 N 个连续句子组合, \hat{o}^m 和 o^{m*} 分别代表预测的顺序和真实的顺序. 而完全正确率则是统计完全正确的排序占的比例, 如式(14)所示.

$$P = \frac{1}{M} \sum_{m=1}^M 1_{\{\hat{o}^m = o^{m*}\}} \quad (14)$$

对于位置正确率 (*Pos*), 如果某个句子在正确的顺序中所处的位置和预测结果中所处的位置一样, 那么就认为该句子是位置正确的, 而位置正确率

则是位置正确的句子数目除以总的句子数目.

对于 *Rouge-N*, 我们在实验中, 分别取了 $N=2$ 以及 $N=3$ 来评价.

4.2.4 实验结果

使用以上实验方法和实验测试指标, 分别在高考议论文数据集上进行实验, 结果如表 2 和图 8 所示. 其中 Random 一行表示随机打乱顺序, 没有经过任何排序算法的顺序在以上指标下所得到的分数.

表 2 句子排序结果

方法	<i>Rouge-S</i>	<i>Rouge-2</i>	<i>Rouge-3</i>	<i>P</i>	<i>Pos</i>
Ptr-Net	0.67	0.34	0.21	0.19	0.40
pair wise	0.65	0.31	0.17	0.14	0.36
learning to rank	0.63	0.29	0.16	0.14	0.35
Random	0.49	0.24	0.08	0.06	0.22

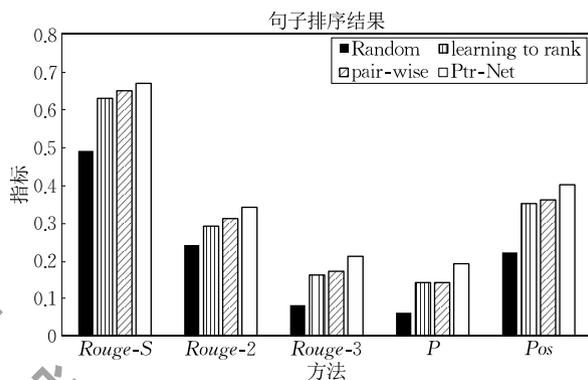


图 8 句子排序结果对比图

从表 2 中可以看到, Ptr-Net、pair wise 模型、learning to rank 模型相比 Random 的结果都有明显的提升, 这也说明, 这 3 个模型, 在作文数据集上都是有明显效果的. 同时通过比较这三种方法的结果, 可以发现 Ptr-Net 在各个指标上都是表现最好的, 其中 *Rouge-3* 和 *P* 这两个对全局信息要求更高的指标上优势更明显. 相比 pair wise 模型, Ptr-Net 在选择的时候考虑更多的全局信息, 而 pair wise 模型则是基于句子对的; Ptr-Net 和 learning to rank 相比较, learning to rank 所使用的特征数量也比较少, 所以也会丢失一些潜在的重要信息, 同时 learning to rank 的一些特征, 比如篇章角色也是使用机器学习方法预测的, 所以难免会有一些的噪声, 并且 LDA 得到的一些特征, 粒度上较粗, 所以也可以从原理上解释 Ptr-Net 的结果要明显好于 learning to rank 方法.

4.2.5 实验样例

根据话题词“诚信”和话题词“爱国”分别通过三种句子排序方法生成的段落的结果如图 9 和图 10

所示。通过比较可以发现,针对话题词“诚信”,基于 learning to rank 的句子排序方法生成的结果,句子与句子之间不够连贯。根据话题词“爱国”生成的段落中出现了语义重复的句子。基于 pair wise 的句子排序方法生成的段落较为连贯,但是根据话题词“诚信”生成的段落中提到的“执政为民”与段落中的前、后句子缺乏较为紧密的联系。根据话题词“爱国”

生成的段落最后一句与段落内其他内容也缺乏较为紧密的联系。基于 Ptr-Net 的句子排序方法较好。根据话题词“诚信”生成的段落也提到了“执政为民”的句子,并在最后通过“取信于民”呼应“执政为民”,突出了段落的整体主题“诚信”。根据话题词“爱国”生成的段落围绕着“国旗”进行论述,并在最后呼吁大家“贡献出自己的力量”。

<p>诚信,是一言既出驷马难追的君子胸怀,能不能落实立党为公、执政为民这个本质,是衡量有没有真正学懂、是不是真心实践“三个代表”重要思想最重要的标志。“诚信”“诚信”信与诚是分不开的啊!所谓诚信,就是要诚实守信,说话算数不骗人。诚信为政,可以取信于民,从而政通人和。有了诚信怎样做到言而有信了。一谈起诚信,谁也不陌生,可又有谁真正做到言而有信呢。中华民族自古就有“言必信,行必果”的教训,作为现代人的中国人民,我们更必须做到诚信,诚信待人也是真心待己,诚信无价!</p> <p>基于 learning to rank 的句子排序方法</p>
<p>诚信,是一言既出驷马难追的君子胸怀,一个“立木取信”,一诺千金;足见在德国,人无信不立。有了它,才有了“君子一言、驷马难追”的承诺;能不能落实立党为公、执政为民这个本质,是衡量有没有真正学懂、是不是真心实践“三个代表”重要思想最重要的标志。诚信是一个道德范畴,即待人处事真诚、老实、讲信誉,言必信、行必果,一言九鼎,一诺千金。一谈起诚信,谁也不陌生,可又有谁真正做到言而有信呢。有了诚信怎样做到言而有信了。</p> <p>基于 pair wise 的句子排序方法</p>
<p>诚信,是一言既出驷马难追的君子胸怀,一谈起诚信,谁也不陌生,可又有谁真正做到言而有信呢。有了诚信怎样做到言而有信了。有人说啊,“言必行,行必果。”足见在德国,人无信不立。能不能落实立党为公、执政为民这个本质,是衡量有没有真正学懂、是不是真心实践“三个代表”重要思想最重要的标志。所谓诚信,就是要诚实守信,说话算数不骗人。诚信为政,可以取信于民,从而政通人和。</p> <p>基于 Ptr-Net 的句子排序方法</p>

图 9 “诚信”段落生成结果

<p>曾听老师说,他看见自己的一个女学生,在一次集体搞卫生的时候,用自己的手摘了一颗所有人看见而没有搞的杂草。“下面升国旗奏国歌。”升国旗时,肃立,向国旗行注目礼;例如:在升国旗时,要行注目礼;“下面进行升国旗仪式。”而今,珠峰上插遍了各国国旗;然而做为中国人,当国旗一次次地在竞技场上升起的时候,当国歌一遍遍在耳畔回荡的时候,五星红旗在南极上空飘扬;</p> <p>基于 learning to rank 的句子排序方法</p>
<p>金秋十月,国旗招展,举国同庆,国旗飘扬,扬世纪豪情。“下面进行升国旗仪式。”然而做为中国人,当国旗一次次地在竞技场上升起的时候,当国歌一遍遍在耳畔回荡的时候,他身披国旗,代表一个民族站在世界的前列。中国梦是民族的梦,是泱泱中华的复兴梦,是炎黄子孙的强国梦,也是每个中国人、每个家庭充满渴望和理想的五彩斑斓的梦。</p> <p>基于 pair wise 的句子排序方法</p>
<p>国旗飘扬,扬世纪豪情。然而做为中国人,当国旗一次次地在竞技场上升起的时候,当国歌一遍遍在耳畔回荡的时候,例如:在升国旗时,要行注目礼;而今,珠峰上插遍了各国国旗;“下面进行升国旗仪式。”金秋十月,国旗招展,举国同庆。就让我们对着课堂里庄严神圣的国旗立下不朽誓言:作为中华儿女的我们,应全力响应党和国家的号召,全面开展、接受文明素质教育,坚决按“爱国为核心、守法为重点、诚信为关键、知礼为基础”的文明宗旨做,使自己真真正正做一个文明人,不求所图的为祖国的繁荣昌盛和文明强大贡献出自己的力量。</p> <p>基于 Ptr-Net 的句子排序方法</p>

图 10 “爱国”段落生成结果

4.3 整体评价

4.3.1 作文样例

图 11 是使用我们的方法就话题词“挫折”生成的作文样例。在第一段中,点明了主题,且主题较为明确。第二段和第三段引用事实论据,令议论更具说服力。文章最后对全文进行总结,升华主题。

4.3.2 人工评价

对于整个系统生成的高考作文,我们也进行了

专业的评测,我们使用了由北京师范大学命题的两套高考模拟试卷进行封闭测试,在测试现场公布作文题目,并且由专业的语文老师进行现场打分。对于满分 50 分的作文,我们得到的平均分为 37.5 分。北京高考的评分细则如表 3 所示。可见我们生成的作文是基本达到了高考评阅中的二类卷水平,是比较好的作文。

逆境磨练意志。懦弱者遇到困难选择了退缩；人生总是坎坎坷坷，荆棘密布，难免会受到困难和挫折的困扰。人生的道路中，它便是战胜一切挫折和困难的不竭动力。重要的是在挫折中能坚持到底，永不言弃，直至击败挫折。只有能面对困难和挫折而毫无惧色的人，才能到达成功的顶峰。又或许我们会选择迎难而上，不屈向困难低头而奋起跨越。面对挫折，要乐观、自信。

遇到挫折就垂头丧气，一蹶不振；失败是不气馁，坦然面对，再接再厉敢再向困难挑战是；面对失败与挫折，我们沮丧，不知所措；稍遇困难、挫折就会一蹶不振，准备放弃。敢于直面人生挫折，勇于接受困难的挑战，即使眼看山穷水尽，仍会峰回路转，柳暗花明。并非自满而不知进取的消极态度。高大的姚明摔倒在地，被对手牵起来，他总是以笑相对，当教练指导谈话，他总是微笑聆听，这个优秀的中国小伙子，正是以其谦和、进取的人生态度潜移默化地影响着中国人乃至世界人民。原打算要去北京而后又去了上海的那个乡下人，抱着积极进取，顽强拼搏的生活态度，树立了乐观向上的世界观，所以他认为去上海就是选择了发财致富的好路子。

勾践卧薪尝胆，忍辱负重，逆境中毫不退缩，挫折中未曾止步，而是于逆境中崛起，重创后屹立。孝庄太后和张学良将军在人生遇到挫折时都选择了坚强。但他并没有绝望，而是积极向上，不怕困难，不怕挫折，克服了阻碍他人生的绊脚石。经受住磨难，勇敢面对挫折，成功之路就在自己脚下。”命运就是如此，不让人经历苦难，是不会让人成功的，但是，要记住命运如棋，掌握在自己手中，那么便会看到彩虹。同样，人也是这样，不经历风雨，怎见彩虹。命运为我们设置重重障碍，苦难失意，挫折坎坷便应运而生。以魄人的乐曲面对挫折，萦绕挫折，直至将挫折覆灭。

不怕困难勇往直前。因为大海的雄奇伟力造就了水手们敢于冒险，敢于拼搏，勇往直前的进取性格，磨炼水手们不怕吃苦，不怕牺牲的坚强意志。的杜甫，他不畏艰险，勇于攀登；因为男儿不畏艰险，不怕困难，不惧伤痛，所以优先于信念！战胜挫折，需要永不言弃的顽强品格。面对挫折，勇往直前；那经受挫折永不消沉，坚强勇敢抗击灾难，那面对挫折不屈不挠艰苦奋斗，奋发向上的精神。强者不畏艰难险阻，涉水翻山。

让我们坚强永不放弃，勇敢面对困境；战胜挫折，需要坚韧不拔的奋斗精神。遇到挫折就垂头丧气，一蹶不振；选择奋进，不甘失败、挫折和坎坷吓倒，是明智的选择。勇敢面对挫折吧！自信是拥有克服困难的毅力。逆境是磨刀石，逆境是试金石，逆境是助推器，只要我们在逆境面前迎难而上，直面挫折，定能成就完美的人生！战胜挫折，需要积极乐观的人生态度。

图 11 “挫折”生成结果

表 3 北京高考作文评分细则

类别	评分要求	评分说明
一类卷 (42~50分)	符合题意、论点明确 论据充实、论证合理 语言流畅、表达得体 结构严谨、层次分明	以46分为基准分。 符合一类卷的基本要求，有创意、有文采的作文可得48分以上
二类卷 (33~41分)	符合题意、论点明确 论据较充实、论证合理 语言通顺、表达基本得体 结构完整、条理清楚	以37分为基准分。 符合二类卷的基本要求，其中某一方面较为突出的作文可得39分以上
三类卷 (25~32分)	基本符合题意、论点基本明确 论据较充实、论证基本合理 语言基本通顺、有少量语病 结构基本完整、条理基本清楚	以29分为基准分。 符合三类文的基本要求，其中某一方面较好的作文可得31分以上
四类卷 (0~24分)	偏离题意、立意不当 论点不明确、内容空洞 语言不通顺、语病多 结构不完整、条理紊乱	以24分为基准分

5 结 论

本文的主要贡献是提出了抽取式作文生成模型，在句子抽取过程中提出了先抽取再扩展的方法，同时将句子排序方法迁移到了段落生成任务上。通过结果我们可以看到，在句子抽取模块中，我们解决了仅仅使用语义抽取带来的语义狭窄的问题，而句子排序技术也确实能在段落生成中学习到底段落内部的连贯性与逻辑性，最终生成的作文可以达到北京高考的二类卷平均水平。

参 考 文 献

[1] Manning C D, Schütze H. Foundations of Statistical Natural

Language Processing. Cambridge: MIT Press, 1999

[2] Jurafsky D, Martin J H. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River: Prentice Hall PTR, 2000

[3] Reiter E, Dale R, Feng Z. Building Natural Language Generation Systems. Cambridge: Cambridge University Press, 2000

[4] Wang Z, He W, Wu H, et al. Chinese poetry generation with planning based neural network//Proceedings of the COLING 2016, the 26th International Conference on Computational Linguistics. Osaka, Japan, 2016: 1051-1060

[5] Yin J, Jiang X, Lu Z, et al. Neural generative question answering//Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, USA, 2016: 2972-2978

[6] Feng X, Liu M, Liu J, et al. Topic-to-essay generation with

- neural networks//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm, Sweden, 2018; 4078-4084
- [7] Yang P, Li L, Luo F, et al. Enhance topic-to-essay generation with external commonsense knowledge//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019; 2002-2012
- [8] Zhang J, Yao J G, Wan X. Towards constructing sports news from live text commentary//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany, 2016; 1361-1371
- [9] Li L, Wan X, Yao J, et al. Leveraging diverse lexical chains to construct essays for Chinese college entrance examination//Proceedings of the 8th International Joint Conference on Natural Language Processing. Taipei, China, 2017; 355-360
- [10] Liu J, Sun C, Qin B. Deep learning based document theme analysis for composition generation//Proceedings of the Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Nanjing, China, 2017; 333-342
- [11] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 2003, 3(1): 993-1022
- [12] Liu T Y. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 2009, 3(3): 225-331
- [13] Chen X, Qiu X, Huang X. Neural sentence ordering. *arXiv preprint arXiv:1607.06952*, 2016
- [14] Vinyals O, Fortunato M, Jaitly N. Pointer networks//Proceedings of the Advances in Neural Information Processing Systems. Montréal, Canada, 2015; 2692-2700
- [15] Logeswaran L, Lee H, Radev D. Sentence ordering using recurrent neural networks. *arXiv preprint arXiv:1611.02654*, 2016
- [16] Gong J, Chen X, Qiu X, et al. End-to-end neural sentence ordering using pointer network. *arXiv preprint arXiv:1611.04953*, 2016
- [17] Song W, Fu R, Liu L, et al. Discourse element identification in student essays based on global and local cohesion//Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal, 2015; 2255-2261
- [18] Li J, Hovy E. A model of coherence based on distributed sentence representation//Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, Qatar, 2014; 2039-2048
- [19] Valizadegan H, Jin R, Zhang R, et al. Learning to rank by optimizing NDCG measure//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2009; 1883-1891



FENG Xiao-Cheng, Ph. D., associate researcher. His research interests include text generation and information extraction.

GONG Heng, Ph. D. His research interest is text generation.

LENG Hai-Tao, M. S. His research interest is text generation.

QIN Bing, Ph. D., professor. Her research interest is natural language processing.

SUN Cheng-Jie, Ph. D., assistant professor. His research interest is natural language processing.

LIU Ting, Ph. D., professor. His research interest is social computing, information retrieval and natural language processing.

Background

The task we have studied in this paper is essay generation which is in the area of text generation. It takes topic words as input and output a long coherent essay under the theme of those topic words, consisting of several paragraphs. Similar work on this task mainly focuses on extracting sentences while using heuristic rules on sentence ordering. We explore three methods on sentence extraction and paragraph generation in this paper. Our proposed model's generated essay can achieve average score of level two of Beijing college entrance examination, evaluated by experts. This paper is supported by the National Key R&D Program of China under Grant No.2018YFB1005103 and the National Natural Science

Foundation of China under Grant 61632011 and 61772156. It focuses on essay generation in National College Examination, which is a part of Key Technologies and Systems for Natural Language Question Solving and Answer Generation. This research project aims at studying on semantic-based deep question answer system with respect to the complexity and diversity of the natural language of questions in education. The project proposes a system with the ability on extracting knowledge, understanding and reasoning according to the texts, and answering questions. It also aims at putting forward a smart question answer system for elementary education.