

复述平行语料构建及其应用方法研究

王雅松 刘明童 张玉洁[†] 徐金安 陈钰枫

北京交通大学计算机与信息技术学院, 北京 100044; [†] 通信作者, E-mail: yjzhang@bjtu.edu.cn

摘要 以汉语为研究对象, 提出构建大规模高质量汉语复述平行语料的方法。基于翻译引擎进行复述数据增强, 将英语复述平行语料迁移到汉语中, 同时人工构建汉语复述评测数据集。基于构建的汉语复述数据, 在复述识别和自然语言推理任务中验证复述数据构建及其应用方法的有效性。首先基于复述语料生成复述识别数据集, 预训练基于注意力机制的神经网络句子匹配模型, 训练模型捕获复述信息, 然后将预训练的模型用于自然语言推理任务, 改进其性能。在自然语言推理公开数据集上的评测结果表明, 所构建的复述语料可有效地应用在复述识别任务中, 模型可以学习复述知识。应用在自然语言推理任务中时, 复述知识能有效地提升自然语言推理模型的精度, 从而验证了复述知识对下游语义理解任务的有效性。所提出的复述语料构建方法不依赖语种, 可为其他语言和领域提供更多的训练数据, 生成高质量的复述数据, 改进其他任务的性能。

关键词 复述语料构建; 数据增强; 迁移学习; 复述识别; 自然语言推理

Research on the Construction and Application of Paraphrase Parallel Corpus

WANG Yasong, LIU Mingtong, ZHANG Yujie[†], XU Jin'an, CHEN Yufeng

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044;

[†] Corresponding author, E-mail: yjzhang@bjtu.edu.cn

Abstract Taking Chinese as the research object, the authors put forward the method to construct large-scale and high-quality paraphrase parallel corpora. The paraphrase data augmentation method include transferring English paraphrase corpus to Chinese, by using the method of translation engines, and manually annotating evaluation data set. Based on the constructed Chinese paraphrase data, the validity of the paraphrase data construction application method is verified in the paraphrase recognition task and natural language inference task. Firstly, the paraphrase recognition data is generated based on the constructed paraphrase corpus, and the attention-based neural network model of sentence matching is pre-trained to capture the paraphrase information. Then, the pre-trained model is applied to the natural language inference task to improve the performance. The experimental results on the open set show that the constructed paraphrase corpus can be effectively applied to the paraphrase recognition task, and the model can learn paraphrase knowledge. When applied to natural language inference task, paraphrase knowledge can effectively improve the accuracy of natural language inference models and verify the effectiveness of paraphrase knowledge for downstream semantic understanding tasks. Meanwhile, the proposed construction method for the paraphrase corpus is language-independent, which can provide more training data for other languages and fields, generate high-quality paraphrase data, and further improve the performance of other tasks.

Key words paraphrase corpus construction; data augmentation; transfer learning; paraphrase recognition; natural language inference

自然语言处理的关键在于理解语言的歧义性和多样性。语言多样性指“异形同义”现象, 即复述。复述指同一语言对同一意思的不同表达。思想表达

的复杂性导致语言现象多种多样, 机器处理自然语言任务时, 面对同一语义的多样化表达方式, 难以准确地理解语义并做出相同的反应, 因此难以达到

机器理解自然语言的目标。

复述已广泛应用于复述生成^[1]、机器翻译^[2]和信息检索^[3]等自然语言处理任务中,目前多采用神经网络模型框架,包含语言多样性的数据可以增强模型对语言现象的学习能力。因此,构建复述语料可以覆盖更多的语言现象,从而提升模型的鲁棒性。同时,利用复述数据以及验证复述对提升自然语言处理的性能具有重要意义。

不同的语种具有不同的语言现象和特点。英语的语法结构有较严格的约束,日语的语法比英语灵活,但也有格助词等直接的表层特征可用于机器识别。与英语和日语相比,汉语的形态更复杂,也缺乏可以帮助机器识别的标志,因此计算机理解汉语语义的难度较大。

目前,与复述相关的研究及资源多数集中于英语和日语,基于汉语的复述平行语料十分匮乏,阻碍了汉语复述研究的发展。本文提出利用多翻译引擎构建汉语复述平行语料的方法,将现有的英文复述语料迁移到汉语中,首次构建大规模高质量的汉语复述平行语料,为汉语复述研究提供数据基础。

为了验证所构建的复述数据在不同自然语言处理任务中的有效性,我们分别将复述数据应用于复述识别任务和自然语言推理任务,在基于注意力机制^[4]的句子匹配模型上进行实验和评测。根据迁移学习的思想,将预训练复述识别任务中学习到的复述知识迁移到自然语言推理任务中,并在公开的推理评测数据集上进行评测,验证复述数据对语义理解任务的有效性。作为一种数据增强手段,本文构建的复述语料可以弥补汉语其他任务训练样本不足的问题,增强自然语言处理任务模型的泛化能力。

1 相关工作

复述可以用于扩展信息检索和自动问答^[5],也可以帮助扩展机器翻译的训练数据^[2],还可以应用于计算机辅助阅读^[6]和自动摘要^[7],是自然语言处理领域的研究热点。在基于神经网络模型训练的技术主流下,复述数据的获取和构建对复述研究十分重要。随着机器翻译技术的成熟,已有研究者提出基于多种翻译引擎的多枢轴方法^[8],将源语言翻译成多个语言的译句,再通过反向翻译成源句,得到多个候选复述,并利用不同的方法选取最佳的复述句对。这种方法能够获取更多的句子表达形式,增加复述的多样化。此外,Wieting等^[9]提出反向翻译

的方法,结合神经机器翻译的编码-解码框架,将经过二次翻译的源句译文作为输入句子的复述,同时与自然语言处理应用相结合,将构建的复述句对应于句子语义表示的学习。利用机器翻译技术,能够在一定程度上将丰富的英语复述平行资源迁移到汉语复述语料中,对推动汉语复述研究的发展有重要意义。

对于其他自然语言处理任务中训练样本不足的问题,有研究人员利用同义词替换、随机插入删除或随机替换^[10]等添加噪声的方法来进行文本数据增强,这些方法能够促进生成对抗网络模型^[11]的发展,但需要更多的数据资源,也会增加模型训练的难度。本文提出的方法能够快速生成高质量的复述数据,并且复述知识可以增强其他任务的语义理解能力,应用于不同自然语言处理任务中能够提升性能,具备有效的数据增强能力。

2 汉语复述平行语料构建方法

2.1 基于多翻译引擎构建汉语复述平行语料

汉语复述研究资源的匮乏导致神经网络模型的训练样本数量不足,难以利用复杂的深度学习模型来学习句子的语义特征。受已有的数据增强方案启发,基于英语复述资源较为充足、神经机器翻译技术已有成熟的应用以及主流的翻译引擎取得良好翻译性能的现状,本文提出基于多翻译引擎的方法,利用英语复述资源增强汉语复述数据。

采用由问句复述句对组成的 Quora 英语复述平行语料。由于不同的翻译引擎在不同的领域表现出不同的翻译质量,我们首先在谷歌翻译、必应翻译、百度翻译、搜狗翻译和有道翻译这 5 个主流翻译引擎中,选择更适合 Quora 数据集的翻译引擎。为探究不同的翻译引擎翻译不同长度句子的效果,我们从 Quora 数据集中挑选 40 对句子用于翻译引擎筛选,包括长度为 5, 10, 15 和 20 词的复述句对各 10 对,分别经 5 个翻译引擎进行翻译。然后,人工制定判断两个句子是否互为复述及其多样性的评分标准。由于人工标注存在标注者主观性强的问题,因此我们制定的评分标注尽可能通过量化的方法,将句子的样式变化做出定义,以便减小由于标注者不同而引起的统计结果偏差。

评分标准如表 1 所示,其中 1 分和 2 分判定为不可用的翻译结果,即非复述句对; 3 分、4 分和 5 分判定为复述句对。并且,表达方式越多样,则人

表1 评测汉语译文句对的评分标准
Table 1 Scoring criteria on Chinese translation pairs

分数	评测标准
1	两个汉语译文翻译不准确,用词和语法错误超过2处
2	两个汉语译文翻译部分不准确,存在1~2处用词和语法错误
3	两个汉语译文翻译基本上正确,无语法错误,语法结构基本上不变,存在1~2处词汇变化
4	两个汉语译文翻译准确且语法正确,语法结构基本上不变,词汇变化超过2处;或语法结构存在简单的变换
5	两个汉语译文翻译准确,语法正确,语法结构存在较大的变换;或词汇变化超过3处

工评测得分越高,即复述的质量越高。

采用表1的评分标准,对上述40对汉语译文进行人工评分,统计评分结果为3~5分的不同长度句对的译文个数,结果如图1所示,其中纵轴表示不同句长的句子由5个翻译引擎得到的译文统计数。可以看出,搜狗翻译和有道翻译引擎的翻译性能在Quora数据集上优势最明显,二者均获得较多得分为3~5分的复述句对。此外,对于15和20词的长句,二者也生成个数远超其他引擎的复述句对。因此,我们选取搜狗翻译和有道翻译来翻译Quora训练集中的每一对英文复述句,用来构建汉语复述训练集。过滤掉编辑距离小于2(即表现形式差异小)的翻译句对,得到搜狗翻译和有道翻译的汉语复述句对分别为13.0万和13.3万。如果将二者合并,可以得到26.3万对的汉语复述训练集。采用同样的方式,对验证集和测试集分别翻译和过滤,各自得到1万对汉语句对。

本文提出的基于翻译引擎的复述语料构建方法在迁移数据方面具有通用性,不仅可以应用于Quora的英文问句数据集,而且可以应用于陈述句或其他复述数据集,当机器翻译引擎具备英语至其他语言的翻译条件时,可以为其他语言构建复述数

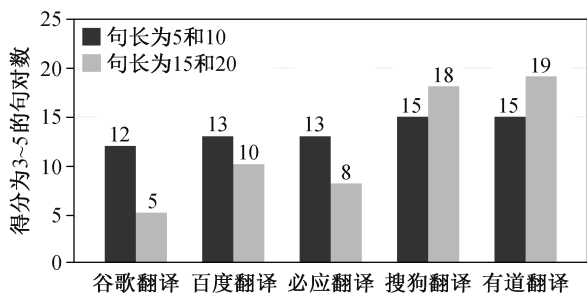


图1 5个翻译引擎不同长度译文的评分统计结果

Fig. 1 Statistical results of scoring different-length Chinese translation pairs on the five translation engines

据集提供可行的思路。

2.2 人工标注辅助方法构建复述评测集

机器翻译引擎方法得到的文本会有一定程度的损失,可能存在汉语译文句对语义不一致的情况。为了构建一个高质量的汉语复述评测集,我们对译文进行人工标注。数据集标注人数为2人,按照表1的评分标准进行标注。对1万对汉语复述测试集进行注释,筛选出符合4分或5分标准的翻译句对,作为复述评测数据。通过这种方式过滤掉翻译错误、语义不同以及表达形式差异小的句对。最终,将筛选后得到的高质量汉语复述句对作为汉语复述评测集,应用于复述生成及相关任务的评测。利用本文方法构建的汉语复述数据集及复述识别数据集在开源网站(<https://github.com/Wys997/Chinese-Paraphrase-from-Quora>)发布。

3 神经网络句子匹配模型

将构建的复述数据分别应用于复述识别与自然语言推理任务中,进行复述数据的有效性验证。在复述识别任务中直接将构建的复述数据作为训练集;在自然语言推理任务中采用迁移学习^[12]的思想,将复述信息迁移到基于注意力机制的句子匹配深度网络模型中,训练复述识别任务的模型参数。对自然语言推理任务模型进行微调后,进行自然语言推理任务的训练和评测,通过公开评测集的验证来评估本文构建的复述数据的质量。

3.1 基于注意力机制的句子匹配模型

复述识别任务和自然语言推理任务都可以视为一种文本分类任务:输入两个句子,输出两个句子的相关性标签。二者均将平行句对用于模型的训练,不同任务输出的标签类型不同。复述识别任务的输出标签包括两类:非复述0和复述1;自然语言推理任务的输出标签包括三类:蕴含(entailment)、矛盾(contradiction)和中立(neutral)。我们采用Duan等^[13]提出的基于注意力机制的句子匹配神经网络模型(AF-DMN),在两个任务中进行分类实验。

AF-DMN句子匹配模型的结构主要包括三层:编码层、匹配层和预测层。在编码层,输入的两个句子分别经过一层Bi-LSTM网络,得到句子的上下文语义表示。在匹配层,基于注意力机制,将两个句子表示进行联系,经过cross attention层、fusion层、self-attention层和fusion层,得到新的句子表示。注意力机制可以计算句子中每一个词和另一个

句子中词汇的相关性,即注意力权重。匹配层中的 cross attention 表示跨句子注意力机制,是计算两个不同句子的注意力权重。同理, self-attention 表示自注意力机制,是计算一个句子内单词之间的注意力权重。fusion 层是将句子的上一层表示和带有注意力权重的句子表示进行拼接融合。最后,预测层采用池化^[14]的方法来融合语义信息,得到固定长度的句子表示,将提取的语义信息输入全连接网络层进行标签预测。

3.2 模型细节设计

我们采用腾讯 AI Lab 开源的大规模高质量中文词向量(<https://ai.tencent.com/ailab/nlp/embedding.html>)为预训练的词向量,维度为 200,在模型训练过程中不更新词向量。模型参数设置如下: Batch 大小为 16,在层之间采用 Dropout 正则化技术, drop 率为 0.35。我们采用 Adam 优化算法^[15]训练模型,设置初始学习率为 0.0004,在复述识别和自然语言推理两个任务中都采用准确率作为评测指标。

4 在复述识别任务中的有效性验证

4.1 构建复述识别数据

基于 Quora 英文数据中标记为 1 的复述句对,获取 26.3 万对汉语复述数据,其中搜索引擎和有道引擎翻译结果各有约 13 万对汉语复述句对,我们在此基础上构建汉语复述识别数据。训练集由两部分组成,以 13 万搜狗翻译结果作为复述数据,即标记为 1 的正例样本;另一部分由非复述数据构成,标记为 0,作为负例样本。采用两种方式构建非复述数据,一种是随机采样的方法,另一种是本文提出的基于翻译引擎的方法。最终,利用两种方法得到的训练集,在复述识别任务中进行对比实验。

随机采样的方法是将有道翻译结果分为 13 个集合,为每个集合中每个原句在集合内随机匹配出目标句,将组合的句对作为非复述数据。用完全随机的匹配方式得到的非复述句对,其句子相似度几乎为 0,训练集简单的分布会导致句子匹配模型对字面相似但语义不同的非复述句对的学习能力不足。为提高模型的学习精度,我们在匹配目标句时,计算句对的 BLEU 值^[16],选取其值大于 0 的句对作为非复述句对,增加模型数据训练的难度,令模型学习更多的非复述句特征,共获得 13 万对非复述句对。同时,利用本文方法,基于搜狗翻译引擎,将 Quora 英文训练集中非复述数据句对翻译成汉

语,得到 5 万对汉语非复述句对。训练集中保持正例和负例的数量相等,且复述识别验证集和测试集的构建方式和训练集相同。最终构成的复述识别数据集的统计信息如表 2 所示。

4.2 实验过程及结果分析

我们利用不同的复述识别训练集,分别训练复述识别任务,模型参数均采用随机初始化的方法,将实验结果与 AF-DMN 模型在 Quora 英文复述识别数据上的评测结果(来自模型复现的实验结果)进行比较,结果如表 3 所示。可以看出,将本文构建的汉语随机数据用于复述识别任务的训练,所得结果与英文数据评测结果有较大的差距,并且测试集只达到 50% 的准确率。这是由于我们构建的训练集的负例样本来自机器的计算和随机匹配,导致模型对汉语数据中负例样本的学习难度较低,对正例样本的学习难度较高,因此模型难以泛化地学习到复述的表示。将利用翻译引擎方法获得的汉语数据用于复述识别任务的训练,虽然实验结果的准确率与英文数据测试集有 7.95 个百分点的差距,但比随机数据的实验结果有明显的提升,证明本文数据构建方法的有效性。利用本文方法构建汉语复述平行语料,能够有效地迁移复述知识,在复述识别任务中获得较高的性能。另一方面,在非复述数据的知识迁移方面,本文方法也比随机匹配方法更有效。我们提出的基于翻译引擎的方法能够将复述和非复述中不同类型的数据分布信息迁移到汉语数据中,增强模型的学习表示和泛化能力,在复述识别任务中得

表 2 复述识别任务和自然语言推理任务数据统计信息
Table 2 Data statistical information for paraphrase recognition task and natural language inference

数据集	训练集/ 万对	验证集/ 万对	测试集/ 万对	词汇表/ 万条
复述识别(随机数据)	26	1	1	2
复述识别(翻译数据)	10	1	1	2
自然语言推理	9	1	1	0.5

表 3 复述识别任务评测结果

Table 3 Experiment results on paraphrase recognition task

实验数据	测试集准确率/%
Quora 英文数据	87.20
Quora 汉语随机数据	50.00
Quora 汉语翻译数据	79.25

说明: 粗体数字表示本文构建汉语数据的实验结果。

到有效的验证。

我们发现汉语复述识别任务的实验结果中,神经网络模型对句子结构改变较多,而对于语义确实为复述的句对,会给出更高的非复述概率,难以正确地识别为复述。对于句子结构和词序放入相同或相似,却因部分单词的不同导致语义不同的非复述句对,神经网络模型会在复述类别上得到更高概率的结果。

在机器翻译的评测方法中, BLEU 值是通过基于文本语序匹配的句子相似度计算得到,因此从4.1节得到的汉语复述测试集中抽取在复述识别任务中被预测错误的句对来计算 BLEU 值。我们发现复述识别任务中难以识别的两类现象:高 BLEU 值的非复述句对和低 BLEU 值的复述句对,如表4所示。可以看出,示例1的两个句子结构和语序相同,由于时间和实体(“1963年千岛群岛”和“1957年安德里亚诺夫群岛”)两个部分不同,导致两个句子的语义不一致,因此不能认为是复述,但模型预测为复述。示例2的两个句子结构差异较大,但是都表达“询问学习编程的方法”的语义,二者可以作为复述句,但模型预测为非复述。神经网络复述识别模型对上述两种类型的文本特征学习不足,难以正确地识别。因此,构建更多具有多样性复述现象的数据十分重要。

5 在自然语言推理任务中的有效性验证

受 Wieting 等^[9]的启发,我们首先利用第4节中复述识别任务的模型,通过模型微调的方式,将学习到的模型参数迁移到自然语言推理任务中,在公开的推理数据集上进行评测,证明本文构建的复述数据对改进自然语言推理性能的有效性。

表4 神经网络复述识别现象
Table 4 Phenomena on paraphrase recognition task

示例	现象	示例	BLEU
1	高 BLEU 非复述	S: 柬埔寨地震的主要影响是什么, 这些影响与 1963年千岛群岛 地震相比如何?	0.84
		T: 柬埔寨地震的主要影响是什么, 这些影响与 1957年安德里亚诺夫群岛 地震相比如何?	
2	低 BLEU 复述	S: 学习编程的 方法是什么 ? T: 如何 学习编程?	0.20

说明: S 和 T 代表输入数据对中的两个句子; 示例1中粗体字表示两个句子中相对应的语义不同的词组, 即非复述现象; 示例2中粗体字表示两个句子中相对应的语义相同的词组, 即复述现象。

5.1 实验数据

实验中使用 CCL2018 评测任务 (<https://github.com/blcunlp/CNLI>) 中的汉语自然语言推理公开评测集。该评测集来自 SNLI 和 MultNLI 两个数据集, 经过机器翻译后再进行人工整理。CNLI 自然语言推理评测集共有训练集 9 万对, 验证集和测试集各 1 万对(表2)。

5.2 实验过程及结果分析

用基于翻译引擎方法构建的复述识别数据来训练 AF-DMN 模型的参数, 除最后一层外, 其他层的参数完全一致。我们在自然语言推理数据上微调模型, 实验结果如表5所示。其中, 基线模型对所有参数均采用随机初始化的方法, 利用 AF-DMN 模型训练以及评测 CNLI 数据的实验结果; 微调模型是将迁移 AF-DMN 模型在复述识别任务中的模型参数用于自然语言推理任务数据集 CNLI 的训练和评测; 可分解注意力模型为 CCL2018 评测任务比赛中公开的模型。

从表5看出, 与基线模型相比, 微调模型在验证集和测试集上的准确率分别提升 3.45% 和 2.64%。利用本文构建的复述数据训练模型, 并经过微调后训练自然语言推理任务, 精度比随机初始化参数进行模型学习有一定程度的提升。微调模型的验证集准确率超过公开评测实验的结果, 提升 1.58 个百分点, 说明本文构建的数据使实验能够得到在一定程度上超过公开语料库基准的结果。

上述实验结果说明我们构建的数据可以有效地用于复述识别模型的训练学习, 并将复述知识迁移到自然语言推理任务中, 增强语义理解能力, 提高模型学习精度, 增加有监督学习方法的鲁棒性, 证明了我们构建的复述数据的质量以及多翻译引擎构建方法的作用, 同时说明复述在自然语言处理任务

表5 自然语言推理任务评测结果
Table 5 Experiment results on natural language inference

模型	验证集准确率/%	测试集准确率/%
可分解注意力模型 ^[17]	69.35	—
基线模型	67.48	66.67
微调模型	70.93 (3.45↑)	69.31 (2.64↑)

说明: 粗体数字突出显示本文预训练模型的实验结果, 括号内数字表示本文模型与基线模型实验结果的差值, 向上的箭头表示性能提升, 下同。

中具有数据增强的作用，能加强模型对语义的理解和学习能力，提升不同自然语言处理任务的性能。

对于一些利用神经网络进行自然语言处理的任务，由于数据资源匮乏，会导致模型学习性能较低。为验证本文构建数据集用于小样本数据任务的质量，我们降低 CNLI 数据集的训练集和验证集数据量，分别随机抽取原数据集中 50% 和 30% 的数据，构建小样本数据集。在基线模型和微调模型上分别利用两个小样本数据集进行训练和评测，对比实验结果如表 6 所示。可以看出，两个数据集在微调模型上训练和评测的结果比基线模型分别有 2.23 和 11.82 个百分点的提升，说明本文利用翻译引擎方法构建的复述数据可以有效地应用于复述识别任务，并且，预训练模型参数后，在自然语言推理任务中进行微调，对推理任务的性能也有较大幅度的提升。

表 6 小样本数据评测结果

Table 6 Experiment results on small sample data

数据量/万	实验数据量在 CNLI 数据集集中的占比/%	模型	测试集准确率/%
4.5	50	基线模型	60.71
		微调模型	62.94 (2.23↑)
3.0	30	基线模型	50.69
		微调模型	62.51 (11.82↑)

实验结果显示，当模型训练数据急速下降(由原始的 9 万降到 4.5 万，最后降至 3 万)时，基线模型和微调模型的性能都降低。随着数据从 4.5 万降到 3 万，基线模型的性能下降约 10 个百分点，说明模型性能受训练数据数量的影响很大，数据过少会大大地降低模型性能。然而，在 3 万数据量时，微调模型的性能比基线模型有 11.82 个百分点的提升，约是 4.5 万数据量时提升值 2.23 个百分点的 5 倍。实验结果证明，我们基于翻译引擎的方法构建的复述数据的质量能够在自然语言推理任务中得到验证，利用该数据训练的微调模型能够提高低资源任务的性能，该数据对自然语言处理下游任务中性能的提升也可以起到较大的作用。

6 结语

本文基于自然语言理解中的瓶颈问题——语言表达的多样性，即复述现象，探索复述语料对自然

语言处理任务的重要性。提出一种基于多翻译引擎构建复述平行语料的方法，将 Quora 英文复述数据迁移到汉语复述中，构建大规模的汉语复述平行语料，其中训练集包括 26 万对复述句，验证集包括 1 万对复述句，并根据获得的翻译结果，人工筛选得到高质量复述句的复述评测集。将构建的复述数据分别应用于复述识别任务和自然语言推理任务中，进行模型的训练和微调。实验结果表明，本文构建的数据可以有效地用于复述识别模型的训练，表现出与英文数据训练接近的性能。将复述知识经过迁移到自然语言推理任务中，能够增强模型的学习能力，提高模型的鲁棒性，表明所构建的复述数据具有较高的质量。在低资源的领域和任务中，本文构建的复述数据能够大大地缓解语料不足导致模型性能低的问题，对自然语言处理中其他任务性能的提升有积极作用。

参考文献

- [1] 赵世奇. 基于统计的复述获取与生成技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2009
- [2] 姚振宇. 基于复述的机器翻译系统融合方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2015
- [3] Zukerman I, Raskutti B. Lexical query paraphrasing for document retrieval // Proceedings of the International Conference on Computational Linguistics. Morristown: Association for Computational Linguistics, 2002: 1-7
- [4] Bahdanau D, Cho K, Bengio Y, et al. Neural machine translation by jointly learning to align and translate // Proceedings of 3rd International Conference on Learning Representations. San Diego: International Conference on Learning Representations [EB/OL]. (2015) [2020-11-24]. <https://arxiv.org/pdf/1409.0473.pdf>
- [5] Mckeown K R. Paraphrasing using given and new information in a question-answer system // Proceedings of Meeting on Association for Computational Linguistics. Morristown: Association for Computational Linguistics, 1979: 67-72
- [6] Carroll J, Minnen G, Pearce D, et al. Simplifying text for language-impaired readers // Proceedings of EACL. Morristown: Association for Computational Linguistics, 1999: 269-270
- [7] Gambhir M, Gupta V. Recent automatic text summarization techniques: a survey. Artificial Intelligence Review. 2017, 47(1): 1-66

- [8] Kok S, Brockett C. Hitting the right paraphrases in good time // *Proceedings of Human Language Technologies: Conference of the North American Chapter of the Association for Computational Linguistics*. Los Angeles: Association for Computational Linguistics, 2010: 45–153
- [9] Wieting J, Mallinson J, Gimpel K. Learning paraphrastic sentence embeddings from back-translated bitext // *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Morristown: Association for Computational Linguistics, 2017: 274–285
- [10] Wei J, Zou K. EDA: easy data augmentation techniques for boosting performance on text classification tasks // *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2019: 6382–6388
- [11] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks // *Proceedings of Neural Information Processing Systems*. Cambridge, 2014: 2672–2680
- [12] Tan C, Sun F, Kong T, et al. A survey on deep transfer learning // *Artificial Neural Networks and Machine Learning — Proceedings of the 27th International Conference on Artificial Neural Networks*. Berlin: Springer Verlag, 2018: 270–279
- [13] Duan C, Cui L, Chen X, et al. Attention-fused deep matching network for natural language inference // *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm: International Joint Conferences on Artificial Intelligence, 2018: 4033–4040
- [14] Weng J, Ahuja N, Huang T S. Cresceptron: a self-organizing neural network which grows adaptively // *International Joint Conference on IEEE Xplore*. Baltimore, 1992: 232–238
- [15] Kingma D P, Ba J. Adam: a method for stochastic optimization // *Proceedings of 3rd International Conference on Learning Representations*. San Diego: International Conference on Learning Representations [EB/OL]. (2015) [2020–11–24]. <https://arxiv.org/pdf/1412.6980.pdf>
- [16] Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation // *Proceedings of the 40th annual meeting on association for computational linguistics*. Stroudsburg: Association for Computational Linguistics, 2002: 311–318
- [17] Parikh A P, Tckstrm O, Das D, et al. A decomposable attention model for natural language inference // *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing and the 6th International Joint Conference on Natural Language Processing*. Stroudsburg: Association for Computational Linguistics, 2016: 2249–2255