第 38 卷 第 5 期
2023 年 10 月

天津科技大学学报
Journal of Tianjin University of Science & Technology

Vol. 38 No. 5
Oct. 2023

DOI:10.13364/j.issn.1672-6510.20230040

基于回译和分层对齐的医疗专家–外行风格迁移的并行句子增强方法

吕焯宸，李孝忠
（天津科技大学人工智能学院，天津 300457）

摘　要：医疗专家–外行风格迁移任务旨在将医疗专家的知识和语言转化为外行易于理解的语言风格，解决医疗专业知识传播中的语言障碍，使医疗知识能够更加广泛地被外行所理解和应用，帮助医疗工作者更好地与患者和家属进行沟通，提高医疗工作的效率和质量。目前医疗专家–外行风格迁移任务可用的数据集匮乏，并且没有可用的并行数据集。基于回译和文本分层对齐方法，提出一种简单的无监督方法，这种方法可以从两种不同文本风格（专家风格和外行风格）的可比语料中提取伪并行句子对。将使用此方法获得的并行句子对与 MSD 数据集、SimpWiki 数据集进行对比，验证了方法的有效性。结果表明，本文方法提取的并行句子对的效果优于 SimpWiki 数据集，可以对目前匮乏的并行数据集进行有效补充。

关键词：回译；分层对齐；文本风格迁移；专家–外行风格迁移
中图分类号：TP391.1　　**文献标志码**：A　　**文章编号**：1672-6510（2023）05-0074-07

Parallel Data Enhancement Method Based on Back Translation and Hierarchical Alignment for Medical Expert-Layman Text Style Transfer

LÜ Zhuochen，LI Xiaozhong
（College of Artificial Intelligence，Tianjin University of Science & Technology，Tianjin 300457，China）

Abstract：The task of medical expert-layman style transfer aims to convert the knowledge and language of medical experts into a language style that is easily understood by the general public，in order to address the language barriers in the dissemination of medical professional knowledge，so that medical knowledge can be more widely understood and applied by the general public，and to help medical workers better communicate with patients and their families，thereby improving the efficiency and quality of medical work. Currently，there is however a paucity of available corpora for medical expert-layman text style transfer tasks，and no parallel corpus is currently available. In this article we propose a simple unsupervised method based on reverse translation and text hierarchical alignment methods. This method can extract pseudo-parallel single-speech pairs from a comparable corpus of two different text styles（expert style and layman style）. The effectiveness of the method was verified by comparing the parallel sentence pairs obtained using this method with the MSD dataset and the SimpWiki dataset. The results showed that the parallel sentence pairs extracted by the method were better than the SimpWiki dataset，and could effectively complement the currently scarce parallel data.

Key words：back translation；hierarchical alignment；text style transfer；expert-layman style transfer

随着网络的发展，在线的医疗服务已经日趋丰富，尤其是近十年来，人们开始以网络平台为媒介，在网上查询医疗信息，寻求医疗咨询，甚至凭借其获取的残缺信息进行自我诊断。专家的建议和外行对

收稿日期：2023–02–26；**修回日期**：2023–05–19
作者简介：吕焯宸（1996—），男，广西人，硕士研究生；**通信作者**：李孝忠，教授，lixz@tust.edu.cn

它的理解之间往往存在着一定的差异[1]。以医疗咨询为例，患者使用的搜索词可能过于模糊，以至于找不到对应的医疗资源[2]。

为了消除医疗专家（expert）与外行（layman）间的沟通障碍，Cao 等[3]基于文本风格迁移，定义了一个全新的子任务，即医疗专家–外行风格迁移。这一任务旨在通过文本风格迁移的相关技术，在专业语言与外行语言之间进行文本风格迁移，在保留语句内容的基础上，将医疗专家的专业文本转化为外行也能够快速理解的通俗化语言文本，或将外行对于医疗咨询、病情症状相关的描述性语言文本转化为医学专业语言。Cao 等[3]还贡献了一个人工注释的非并行数据集——MSD 数据集，这个数据集是目前医学专家–外行风格迁移领域唯一可用的公开数据集。Irbaz 等[4]和李慧等[5]对文本风格迁移模型进行改进，使用不同的无监督方法和不同的数据进行研究，并取得了一定的成果。Xu 等[6]使用基于边缘的数据挖掘[7]对 MSD 数据集进行并行化拓展，从 MSD 数据集中提取了 11 500 个并行句子对，将其用于预训练之后的微调。

作为文本风格迁移的一个新的方向，目前仅有一个非并行的数据集 MSD。由于专家风格和外行风格在句子结构上存在较大的差异，传统的基于词对齐的方法效果并不好，如最小编辑距离、TF-IDF（term frequency-inverse document frequency）等。本文针对医疗专家–外行风格迁移任务缺乏并行数据集的问题，提出一种基于回译和大规模分层对齐[8]（large-scale hierarchical alignment，LHA）的并行句子增强方法。

1　回译与句子对齐

1.1　回译

回译通常被用于机器翻译。由于机器翻译依赖大量双语的并行语料，而实际可用的并行语料有限，但是单语语料十分充足，因此可以对单语语料使用回译生成并行数据集，然后将并行数据集添加到原有的数据集中。使用回译进行数据扩充，可以有效提高机器翻译的效果[9]。使用回译进行数据增强，依赖于将文本数据翻译成另一种语言，然后再将其翻译回原始语言。这种技术可以生成与原始文本不同的文本数据，同时保留原始的上下文和含义。

1.2　句子对齐方法

1.2.1　机器翻译中的句子对齐

句子对齐是一种文本到文本的重写方法，是一种从原始语料中提取句子对的方法。目前关于句子对齐的大部分工作研究，主要集中在机器翻译的任务中，被用于从不同语言的大型并行语料库中提取合适的句子对。这种添加伪并行句子对的方法可以提高机器翻译模型的性能[10]。句子对齐方法主要用于从双语句子中提取伪并行句子对，对于单语语料库的句子对齐方法的研究非常少。Barzilay 等[11]提出了一种针对单语语料库的层次对齐方法，即先将相似主题的段落聚类，然后再进行句子对齐。Marie 等[12]使用预训练的单词嵌入和句子嵌入提取两种语言的粗略翻译句子对，随后再使用通过并行翻译数据训练的分类器过滤掉低质量的翻译句子对。

1.2.2　LHA 方法

LHA 方法将句子对齐方法用于文本风格迁移任务[8,11-12]。该方法使用一种基于层次结构的对齐算法，先进行短语对齐，再进行句子对齐。从不同语言风格的单语语料中提取语义相似的伪并行句子对。

LHA 方法是一种使用文档对齐和句子对齐的分层对齐方法。它先在文档级别上进行对齐，根据源数据集中的文档，从目标数据集中检索与其匹配的文档进行文档对齐，产生文档对；然后，再在对齐的文档对中检索并提取语义高度相似的句子对。

文档对齐：首先将源语言文本和目标语言文本分别划分为短语，然后在两种语言中查找相似的短语。如果两个短语相似，则将它们对齐。对齐的短语被组织成树形结构，形成源语言文本和目标语言文本的文档对齐结构。

句子对齐：使用一种基于最长公共子序列（longest common subsequence，LCS）算法对齐两个层次结构之间的句子。将源语言和目标语言的句子都表示成一个单词序列，然后计算它们之间的 LCS。通过 LCS 的匹配，找到源语言文本和目标语言文本之间的对应关系，从而实现句子对齐。

2　本文方法

本文将回译和 LHA 方法相结合，针对医疗专家–外行风格迁移任务缺少并行数据集的现状，提出一种从单语可比语料中获取并行数据集的方法。首先，对两种风格的可比语料库中的原始文本进行预处理，再借助回译生成增强文本；然后，对增强文本使用 LHA 方法，从中提取并行句子对和句子对应的序号；最终，根据句子序号和人工方法，从原始文本中提取对应的句子对。本文方法的总体流程如图 1 所示。

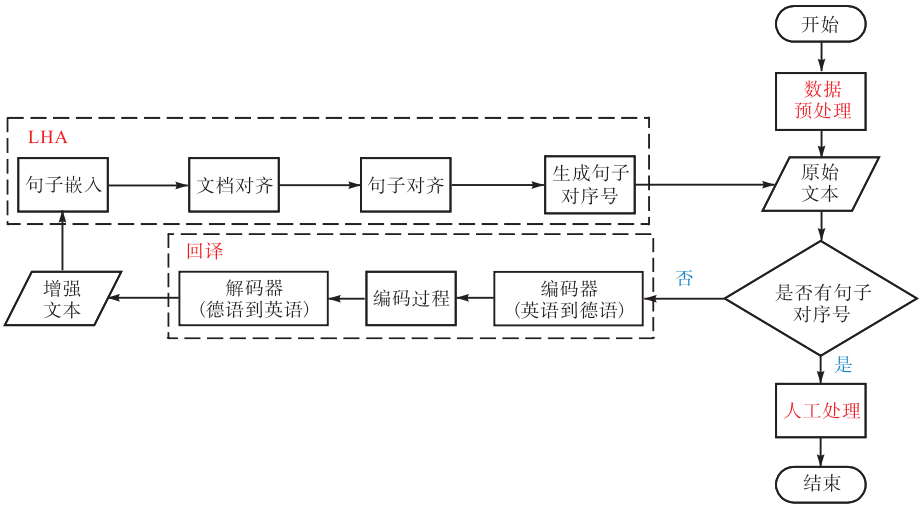

图 1　本文方法的总体流程

Fig. 1　Overall flow chart of the method in this article

2.1　数据来源与预处理

2.1.1　数据来源

原始文本来源于默沙东医疗手册（MSD MANUAL，https://www.msdmanuals.com）。默沙东医疗手册的每篇文章均由数百名医学专家合作编写，包括两个版本：一个是为外行量身定制的消费者版，另一个是为专业医学生和医学专家量身定制的专业版。本文从默沙东医疗手册的消费者版获取外行风格的文章，从专业版获取专家风格的文章。

2.1.2　对文本的预处理

对文本的预处理主要是对两种风格的文章进行匹配以及对每篇文章进行分句处理：根据每篇文章对应的医疗主题（一共有 23 个医疗主题）和文章标题，对两种风格的文章进行简单匹配，从多组医疗主题中获得对应的文章对；使用 Python 的 NLTK 库对每篇文章进行分句处理，给每个句子额外标注一个序号，用于提取相应序号的句子。

2.2　使用回译获得增强文本

2.2.1　基本原理

一般的机器翻译模型无法显式地捕捉数据中出现的各种风格，也不能以不同的、可控的风格生成新数据，从而导致回译前后的文本风格发生变化[13]。这种变化会导致回译产生的增强文本发生风格上的趋同：专家句子的结构发生了一定变化，从而降低了句子对的编辑距离（editing distance，ED）；外行句子的部分非专业词汇被回译成了专业词汇（如高血压 high blood pressure 变成了 hypertension），从而提高句子对的文本相似度。将回译与分层对齐相结合，通过回译带来的风格趋同，消除医疗专家风格和外行风格的部分差异，从而缩小句子对的编辑距离，提高二者的文本相似度。

2.2.2　相关模型

使用 Edunov 等[9]的英德翻译模型，该模型相对成熟，已被用于大规模的翻译，并取得了良好的效果。它基于预训练的 transformer 模型[14]，使用 6 层相同的编码器–解码器结构（具体结构如图 1 所示）。该模型使用来自 WMT'18 英德新闻翻译任务[15]的数据进行训练，并使用 BLEU（bilingual evaluation understudy）衡量回译前后的相似度。

本文方法先将专家文本和外行文本分别输入英德翻译模型中，借助德语作为中间语言，将原文本翻译成德语，再翻译回英语，由此生成增强后的专家句子和外行句子，之后再将回译生成的增强文本作为分层对齐模型的输入，参与后续的分层对齐工作。

2.3　文本分层对齐

2.3.1　句子相似度的计算标准

在 LHA 方法中，对文本的相似度或相关度进行匹配通常涉及使用特定的计算公式或原则标准。具体的方法和标准可以根据具体的应用场景和任务而有所不同。本文的相似度是基于词向量的相似度计算。使用预训练的词向量模型可以将词语表示为词向量，然后通过计算词向量之间的相似度评估句子或文档之间的相似度。

本文所使用的词嵌入模型是由 Facebook AI Research 团队开发的 wiki.simple.bin，它是一个基于维基百科语料库训练得到的词嵌入模型。它将单词表示为在连续向量空间中的稠密向量，通过向量表示捕捉词语之间的语义关联性，从而可用于衡量词语的相

似度、句子的语义相似度。

2.3.2　文档对齐

分层对齐方法需要先对同一次级主题内的两种不同风格的文档 S 和 T 中的文章进行对齐，其中 S 表示专家风格的源文档集，T 表示外行风格的目标文档集，S 和 T 分别由多篇文章组成。

对于目标文档，LHA 方法可以计算它与源文档集合中每个文章的相似度，然后将它们按照相似度进行排序。通过设置参数（阈值），使用类似二分查找的方法快速确定相似度大于等于阈值的文档个数。

原始 LHA 方法的 S 和 T 中的文章数量不同，即 $S=\{s_1,s_2,\cdots,s_n\}$，$T=\{t_1,t_2,\cdots,t_m\}$，可能会为每篇文章 s_i 产生多组文章对 (s_i,t_j)、(s_i,t_k) 等。

由于本文方法在预处理时根据文章的标题对文章进行了简单匹配，所以 S 与 T 由等量文章组成，即 $S=\{s_1,s_2,\cdots,s_n\}$，$T=\{t_1,t_2,\cdots,t_n\}$。本文方法为 S 中的每篇文章在 T 中找到唯一一篇与之匹配的文章即可，即通过设置参数，只为每篇源文档保留相似度最高的一篇目标文档，产生一组文章对 (s_i,t_i)。

2.3.3　句子对齐

经过文档对齐后，从两种风格的语料库中获得了匹配的文章对 (s_i,t_i)，其中 s_i 和 t_i 由数量不同的句子组成，记作 $s_i=\{s_i^1,s_i^2,\cdots,s_i^n\}$，$t_i=\{t_i^1,t_i^2,\cdots,t_i^m\}$。通过句子对齐，在每个文章对中找到匹配的句子对。

对 t_i 的每个句子，都计算它与 s_i 中每一个句子的相似度，并根据相似度进行排序。与文档匹配类似，可以通过设定阈值控制句子匹配的个数。

由于专家句属于复杂句，一句专家句的内容，往往包含了多句外行句的内容，所以在句子对齐时，不能像原始 LHA 方法那样将两种风格的句子一一匹配，而是要将一个专家句子与多个外行句子进行匹配。

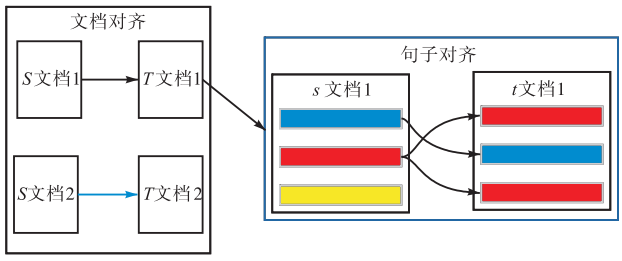

图 2　改进后的 LHA 方法

Fig. 2　Improved LHA method

经过实验观察，为每个 s_i 中的句子最多匹配 2 个 t_i 中的句子，即为 s_i^a 匹配 2 个句子 $\{t_i^j,t_i^k\}$，最终产生句子对 $(s_i^a,\{t_i^j,t_i^k\})$。

2.4　并行句子对的提取

2.4.1　提取并行句子对

回译所生成的增强文本缩小了专家文本和外行文本的风格差异，同时也意味着增强文本中的专家文本和外行文本都发生了风格的改变，因此不能使用增强文本的输出作为最终对齐的句子对。

为了最终获得“原汁原味”的专家风格、外行风格的句子对，还需要根据 LHA 方法获得句子对的序号，从没有经过回译的原文本中分别提取对应序号的句子。

2.4.2　人工过滤句子对

分层对齐方法是通过计算句子相似度进行并行句子的提取，但由于医疗专家句子和外行句子的风格差异过大，而且医疗术语和外行用语之间的单词嵌入不匹配，使最终提取的句子对中存在与实际内容不相符的句子对。

经过观察发现，可以通过简单的单词替换轻易骗过自动评估指标，比如只需要在外行句子中添加 1 个专业词汇，就可以大幅度提高 BLEU 分数，从而影响实验效果。因此，不能使用传统的自动过滤方法，而是必须对提取的句子对进行人工验证，判断两个句子的内容是否相同，并将不合格的句子对过滤掉。由于本文为每个专业句子最多匹配 2 个外行句子，所以进一步增加了人工过滤的必要性。

2.4.3　组合外行句

由于专家句子的复杂性，经过人工过滤后，存在一个专家句匹配多个外行句的情况，即 $(s_i^a,\{t_i^j,t_i^k\})$，因此需要对其进行人工处理。本文参考 MSD 测试集中的例子，将多个外行句作为一组数据组合在同一行中，即对于最终的句子对 $(s_i^a,t_i^{a'})$，其中 $t_i^{a'}$ 可能是由多个完整句子组成的一个段落。

3　实　验

3.1　实验数据集

3.1.1　MSD 数据集

MSD 数据集由一个非并行的训练集和一个并行的测试集组成。非并行训练集中包含 245 000 个医学句子，其中有 130 000 个专家句子和 114 000 个外行句子。并行测试集中包含 675 对具有相同含义的专家–外行句子对，其中测试集经过专家的人工标注。

3.1.2　SimpWiki 数据集

为了检验并行句子对的质量，除了 MSD 数据集之外，本文还使用另一个数据集 SimpWiki 进行比

较。SimpWiki 数据集来源于简单维基百科和普通维基百科之间的链接文章，它专注于医学领域，通过计算文章的 BLEU 分数自动提取并行句子。

3.2　实验设计

本文的实验目的主要有两个，其一是比较本文提出的回译 + LHA 方法与其他句子对齐方法以及 LHA 方法之间的优劣，其二是验证本文方法所提取的伪并行句子对的质量。

为了对比实验效果，设计了两组实验。首先，为了测试本文算法并行句子的提取效果，从 MSD 测试集中抽取 500 对句子，分别使用 LHA 方法和基于词对齐的方法（使用 GIZA++实现）进行句子对齐，并比较这些算法的准确率。除了 LHA 方法之外，其他方法都只测试在原文本上的句子对齐效果。为了比较回译对 LHA 方法的提升效果，LHA 方法在原文本和回译产生的增强文本上分别进行测试。其次，为了验证伪并行句子对的质量，使用本文方法和 LHA 方法从默沙东医疗手册的文章中提取并行句子对。LHA 方法作为对照方法，从经过预处理的原文本中获取并行句子对。本文方法则是先对专家、外行两种风格的原文本分别进行回译，再对回译后的增强文本使用 LHA 方法。将使用原文本生成的数据集记作 LHA-MSD，将使用回译产生的增强文本生成的数据集记作 BT-MSD（back translation MSD）。

3.3　自动评价指标

3.3.1　算法性能

评价句子对齐算法的常用指标包括准确率（precision）、召回率（recall）、F_1 值和对齐错误率（AER）。但由于用于测试的总句子数已定，此时准确率=召回率=1 – 对齐错误率=F_1 值，故只使用准确率作为相关评价指标。

3.3.2　句子简易度

参考 Cao 等[3]的方法，使用 3 个标准可读性指标衡量数据集中句子的简易度水平。FleshKincaid[16]是用来衡量文本可读性的指标，Gunning[17]是计算文本复杂度的指标，Coleman[18]是用于划分文本阅读等级的指标。指标的分数越低，表示句子越简单，可读性越好，反之则句子越复杂，可读性越差。

对每个数据集中两种风格的句子分别测定其单个句子的化简水平，从而判断每个数据集整体的句子复杂度。

3.3.3　并行句质量

在常用的评价指标中，BLEU 分数越高，则内容保存度越高。对于专家和外行两种风格，由于句子结构、专业术语之间的差距太大，无法单纯使用 BLEU 衡量并行句子的质量，因此本文参考 Cao 等[3]的方法，使用 ED 与 4-gram BLEU 共同判断并行句的质量。ED 越高，表明两个句子的结构差异越大，风格差异越明显。

4-gram BLEU 和 ED 的计算公式见式（1）和式（2）。

$$\mathrm{BLEU}=\min\left(1,\mathrm{e}^{\left(1-\frac{r}{c}\right)}\right)\times\mathrm{e}^{\left(\frac{1}{4}\times\sum(\ln p_i)\right)} \tag{1}$$

式中：r 表示目标文本的长度，本文中为外行句子的长度；c 表示参考文本的长度，本文中为专家句子的长度；p_i 表示目标文本中的 4 个连续的单词（4-gram）在参考翻译中出现的概率。

$$\mathrm{ED}(X,Y)=\boldsymbol{D}[m][n] \tag{2}$$

式中：m 和 n 分别为字符串 X 和字符串 Y 的长度；$\boldsymbol{D}[i][j]$是一个矩阵，表示 X 的前 i 个字符与 Y 的前 j 个字符之间的编辑距离。

本文方法的评价标准并非 BLEU 越高越好，ED 越低越好，而是以人工标注的 MSD 测试集的对应分数为黄金参考，认为越接近此分数的并行数据集的质量越好。

3.4　实验结果与分析

3.4.1　算法准确率

以专家风格作为源语言，以外行风格作为目标语言，对比 LHA 方法在原文本上的表现及其在回译的增强文本上的表现，将后者记作 BT-LHA。用于比较基于词对齐的算法包括基于 TF-IDF、最小编辑距离（Least Edit Distance）、词义相似度匹配这 3 种词对齐算法，其中词义相似度匹配使用的是期望最大化算法（EM）。算法准确率见表 1。

表 1　算法准确率

Tab. 1　Algorithm accuracy

算法	准确率/%
TF-IDF	19.85
EM	23.40
Least Edit Distance	18.20
LHA	49.20
BT-LHA	58.67

由表 1 可知，LHA 方法的准确率远远高于基于词对齐的 3 种方法，而 LHA 方法在回译产生的增强文本上的表现又高于其在原文本上的表现。回译 + LHA 方法的效果约为 LHA 的 1.2 倍。但是，即使使用了回译，此方法的准确率也只有 58.67%，这也表明了对伪并行句子对进行人工过滤的必要性。

3.4.2　句子简易度分数

表 2 展示了公共数据集和本文生成的两个数据集的专业句子和外行句子的流畅度指标。公共数据集的相关数据来自 Cao 等[3]的报道，LHA-MSD 和 BT-MSD 的数据是分别从两种风格的句子中随机抽取的 350 个句子计算而得。

表 2　句子简易度分数

Tab. 2　Sentence simplicity score

可读性指标	MSD 训练集		MSD 测试集		SimpWiki		LHA-MSD		BT-MSD	
	专家	外行	专家	外行	专家	外行	专家	外行	专家	外行
FleshKincaid	12.61	9.97	12.05	9.53	12.10	9.63	12.95	11.40	12.95	11.18
Gunning	18.43	15.29	17.89	15.07	17.66	14.86	15.52	13.47	15.43	13.30
Coleman	12.66	10.41	12.26	9.74	10.89	9.70	15.68	13.11	15.64	12.74
Avg	14.57	11.89	14.07	11.45	13.55	11.40	14.71	12.66	14.68	12.41

由表 2 可知，本文的两组实验所提取数据集的总体简易度与 MSD 训练集相比没有太大差距，与人工标注的 MSD 测试集相比，依旧存在一定的差距。而外行句子的复杂度明显高于 MSD 和 SimpWiki 中的外行句子。

与 SimpWiki 相比，本文的两组数据集的专家风格和外行风格的句子都表现出了更为明显的差异，这种差异可能主要与原文本的来源有关。

MSD 测试集中的专家句子相对简单，这可能是由于它对专家句子进行了分割，而本文数据集中的专家句子都没有经过拆分，因此其复杂度高于 MSD 测试集。由于本文方法对部分外行句子进行了拼接，因此 LHA-MSD 和 BT-MSD 的外行句子都更为复杂。

与 LHA-MSD 相比，BT-MSD 专家句子的复杂度更高，外行句子的复杂度更低。两种风格的句子复杂度都更接近 MSD 训练集的。这表明经过回译之后，LHA 方法提取到质量更好的并行句子，可以获得与原文本句子风格差异更大的句子对。

3.4.3　并行句子对的质量分数

表 3 中记录了本文生成的两个数据集和 SimpWiki、MSD 的测试集的并行句子的相关指标。SimpWiki、MSD 测试集的分数源于 Cao 等[3]的数据，本文从本文的两个数据集中分别抽取 350 对句子，测量其 BLEU 和 ED 分数。由于 MSD 训练集是非并行的语料库，故无法计算相关指标。

表 3　并行句子对的质量分数

Tab. 3　Evaluation score for parallel sentences

数据集	BLEU	ED
SimpWiki	49.98	64.16
MSD-Test	14.01	139.73
LHA-MSD	15.22	82.82
BT-MSD	13.15	86.83

从表 3 来看，与人工标注的 MSD 测试集相比，本文方法提取的并行句子对的质量仍然具有相当大的改进空间。两组数据集的句子风格差异都与人工标注的并行句子对有一定差距。相对于 SimpWiki 数据集，本文方法的两组并行句子对的风格差异更为明显，句子结构变化也更加复杂，更接近 MSD 测试集的数据。这说明本文方法提取的并行句子对的质量更高。

对比两组实验，BT-MSD 的并行句子具有更低的 BLEU 分数和更高的 ED。这表明使用回译可以获取风格差异更大、句子结构变化更大的句子对。这也符合本文方法的预期和前文的观点。

实验中还发现，对回译后的增强文本使用 LHA 方法，经过人工过滤之后，BT-MSD 的合格句子对数约为 LHA-MSD 的 1.3 倍。这一结果在实践中验证了表 1 中算法的准确率，进一步表明本文方法比原始 LHA 方法具有更好的性能。

4　结　语

本文针对医疗专家–外行风格迁移工作，提出了一个基于回译和句子对齐的并行数据增强方法，并对其结果进行了比较。本文使用基于序列到序列的回译模型和基于预训练的词（句子）嵌入的分层对齐方法，从两种风格的文本中提取并行句子对，实现了对 MSD 数据集的并行化拓展。与 MSD 人工标注的句子对比，本文方法在质量上仍存在差距，但考虑到人工标注的成本，本文方法具有较好的可行性。与自动获取的 SimpWiki 数据集相比，本文方法获得并行句子的质量更高。未来可以考虑从提高人工过滤效率的角度出发，进一步改进本文方法，同时尝试寻找更为有效的自动评价指标。

参考文献：

［1］ TAN S S L，GOONAWARDENE N. Internet health in-

formation seeking and the patient physician relationship: a systematic review[J]. Journal of medical internet research, 2017, 19(1): e9.

[2] AU A. Why physician guidance matters: a night of neuralgia, meningitis, and WebMD[J]. The annals of family medicine, 2019, 17(5): 462–464.

[3] CAO Y, SHUI R, PAN L, et al. Expertise style transfer: a new task towards better communication between experts and laymen[C]//ACL. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. New York: Association for Computational Linguistics, 2020: 1061–1071.

[4] IRBAZ M S, AZAD A, PREOTY A T, et al. Medical expertise style transfer using denoising auto encoder[M]. Bangladesh: Islamic University of Technology, 2021.

[5] 李慧, 宁康林, 冯志斌. 一种基于文本风格转换模型的医疗辅助咨询平台设计[J]. 科技与创新, 2022(8): 114–117.

[6] XU W, SAXON M, SRA M, et al. Self-supervised knowledge assimilation for expert-layman text style transfer[J]. Proceedings of the AAAI conference on artificial intelligence, 2022, 36(10): 11566–11574.

[7] SCHWENK H. Filtering and mining parallel data in a joint multilingual space[C]//ACL. Proceedings of the 56 th Annual Meeting of the Association for Computational Linguistics. Melbourne: Association for Computational Linguistics, 2018: 228–234.

[8] NIKOLA N I, HAHNLOSER R H R. Large-scale hierarchical alignment for data-driven text rewriting[C]//RANLP. Proceedings of the International Conference on Recent Advances in Natural Language Processin. Varna: INCOMA Ltd., 2019: 844–853.

[9] EDUNOV S, OTT M, AULI M, et al. Understanding back-translation at scale[C]//ACL. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018: 489–500.

[10] USZKOREIT J, PONTE J, POPAT A, et al. Large scale parallel document mining for machine translation[C]//ACL. Proceedings of the 23rd International Conference on Computational Linguistics. New York: Association for Computational Linguistics, 2010: 1101–1109.

[11] BARZILAY R, ELHADAD N. Sentence alignment for monolingual comparable corpora[C]//ACL. Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. NEW York: Association for Computational Linguistics, 2003: 25–32.

[12] MARIE B, FUJITA A. Efficient extraction of pseudo-parallel sentences from raw monolingual data using word embeddings[C]//ACL. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver: Association for Computational Linguistics, 2017: 392–398.

[13] WANG Y, HOANG C, FEDERICO M. Towards modeling the style of translators in neural machine translation[C]//ACL. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics. New York: Association for Computational Linguistics, 2021: 1193–1199.

[14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//ACM. Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 5998–6008.

[15] SENNRICH R, HADDOW B, BIRCH A. Improving neural machine translation models with monolingual data[C]//ACL. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin: Association for Computational Linguistics, 2016: 86–96.

[16] KINCAID J P, FISHBURNE R P, RICHARD L R, et al. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel[R]. Springfield: National Technical Information Service, 1975.

[17] GUNNING R. The technique of clear writing[M]. New York: McGraw-Hill, 1968.

[18] COLEMAN M, LIAU T L. A computer readability formula designed for machine scoring[J]. Journal of applied psychology, 1975, 60(2): 283–284.

责任编辑：郎婧