基于风格化嵌入的中文文本风格迁移

王晨光
大连理工大学/ 大连
wk1997@mail.dlut.edu.cn

林鸿飞†
大连理工大学/ 大连
hflin@dlut.edu.cn

杨亮
大连理工大学/ 大连
liang@dlut.edu.cn

摘要

对话风格能够反映对话者的属性，例如情感、性别和教育背景等。在对话系统中，通过理解用户的对话风格，能够更好地对用户进行建模。同样的，面对不同背景的用户，对话机器人也应该使用不同的语言风格与之交流。语言表达风格是文本的内在属性，然而现有的大多数文本风格迁移研究，集中在英文领域，在中文领域则研究较少。本文构建了三个可用于中文文本风格迁移研究的数据集，并将多种已有的文本风格迁移方法应用于该数据集。同时，本文提出了基于DeepStyle算法与Transformer的风格迁移模型，通过预训练可以获得不同风格的隐层向量表示。并基于Transformer构建生成端模型，在解码阶段，通过重建源文本的方式，保留生成文本的内容信息，并且引入对立风格的嵌入表示，使得模型能够生成不同风格的文本。实验结果表明，本文提出的模型在构建的中文数据集上均优于现有模型。

关键词： 文本风格迁移 ； 对话生成 ； 深度学习

Chinese text style transfer based on stylized embedding

Wang Chenguang
DLUT / DaLian
wk1997@mail.dlut.edu.cn

Lin Hongfei†
DLUT / DaLian
hflin@dlut.edu.cn

Yang Liang
DLUT / DaLian
liang@dlut.edu.cn

Abstract

Dialogue style can reflect the attributes of the interlocutors, such as emotion, gender and educational background. In the dialogue system, the user can be modeled better by understanding the user's dialogue style. Similarly, in the face of users of different backgrounds, the dialogue robot should also communicate with it in different language styles. Language expression style is the intrinsic attribute of text, but most of the existing researches on text style transfer focus on English and less in Chinese. This paper constructs three data sets that can be used in the research of Chinese text style migration, and applies many existing methods to the data set. At the same time, this paper proposes a style transfer model based on DeepStyle algorithm and transformer. Different styles of implicit vector representation can be obtained by pre-training. At the decoding stage, the content information of the generated text is retained by reconstructing the source text, and the embedded representation of opposite styles is introduced to make the model generate different styles of text. The experimental results show that the model proposed in this paper is superior to the existing model in the Chinese data set.

第二十届中国计算语言学大会论文集，第269页-第279页，呼和浩特，中国，2021年8月13日至15日。

Keywords: Text style transfer , Dialogue generation , Deep learning

1 引言

文本风格是文本表达的隐式特征，与文本情感等特征不同的是，文本风格的分析囊括了表达方式分析、情感分析、主题分析等多方面的分析方法，是多种特征分析的综合体。因此，深度模型成为了获取文本风格特征的最佳方法。文本风格迁移任务的任务目标，是在保存文本主要内容不变的前提下，将文本的表达迁移到另一风格．目前大部分的研究集中在英文领域，但在中文领域，相关的研究较少，因此本文构建了三个中文文本风格迁移数据集，用以进一步研究中文中的风格迁移效果。

风格迁移任务起源于图像处理领域，(Shen et al., 2017)将其引入到自然语言上，并且第一次使用了强化学习的方法进行文本风格迁移，这一框架被许多研究者沿用。使用强化学习的一个重要原因，是因为文本风格迁移任务缺少平行语料，为了缓解这一问题，(Li et al., 2018)提出了基于删除与检索的方法，利用TFIDF算法，根据词语的共现情况，对文本进行建模，进而生成伪平行语料。同样使用伪对齐数据生成方法的还有(Pryzant et al., 2020)。

强化学习方法对于判别器有较高要求，但因为缺少平行数据而无往往无法验证判别器的效果。伪对齐生成方法也十分依赖于对齐数据的生成质量，并且二者都将文本主题重建与风格化迁移作为两个较为独立的任务，无法更好的共享两个任务的共性。因此使用了基于StyleEmbedding的方法，将风格标签与文本在同一空间下进行向量表示，故而对风格与文本同时建模。为此，本文引入图神经网络中的DeepWalk算法，通过预训练生成文本风格化嵌入，进而利用Transformer构建生成端模型，使用Seq2Seq结构产生对立风格文本。使用基于StyleEmbedding的DeepWalk算法，解决了对于判别器质量和对齐文本的依赖。同时，常见英文文本风格迁移模型中的Seq2Seq结构，是以RNN模型作为编码结构，对于语义表达更复杂的中文文本，学习效果较差，因此本文使用了Transformer作为编码器基础结构，进一步提升模型对文本的建模能力。

与强化学习和伪对齐数据生成的方法不同，风格化嵌入方法的目标，是在同一隐层空间中，对文本风格与文本内容进行嵌入表示，利用对立风格的嵌入，指导生成对立风格的文本。同样利用该方法的还有(Kim and Sohn, 2020)，与本文不同的是，Kim利用两个encoder计算输入文本与风格嵌入的相似度，进而获取风格化嵌入。然而本文利用词语与风格标签构建关联图，使用DeepWalk算法，获取风格化嵌入表示，与(Kim and Sohn, 2020)相比，能够综合相同类别文本的共有特性，得到更有效的风格化标签。总的来说，本文的主要贡献如下

- 构建了三个可用于中文文本风格迁移研究的数据集。
- 基于改进的DeepWalk算法，通过预训练获取文本风格嵌入表示，并利用Transformer构建生成端模型。

以下将按照四个章节来介绍论文的工作。在第二部分，本文回顾了英文文本风格迁移的相关工作，并详细介绍了几种经典的文本风格迁移方法。在第三部分，本文构建了三个可用于中文文本风格迁移研究的数据集，并提出了基于风格嵌入的DeepStyle算法。在第四部分，构建了对比实验与消融实验，验证了数据集的可用性以及DeepStyle算法的优越性。

2 相关工作

风格迁移研究诞生于图像领域，由(Shen et al., 2017)引入自然语言处理。文本风格迁移的任务目标是，在保持内容不变的基础上，将文本的表达方式、用词习惯等迁移到另一种风格。文本风格迁移研究方法可以分为三种，分别是基于有监督学习的方法、基于半监督学习的方法与基于无监督学习的方法。

2.1 基于有监督学习与半监督学习的文本风格迁移

(Pryzant et al., 2020)将有偏见的文本与无偏见的文本视为两种对立风格的文本，并且利用维基百科构建了18W行的平行数据集。并利用Bert作为编码器，构建Encoder-Decoder结构。同时，与本文一样，也使用基于Bert的分类器，判别风格迁移效果。(Wang et al., 2020)认为编码器能够学习到文本内容的隐层表示，而这种表示是与风格无关的，由此，Wang通过隐层的参数共享，使得两种文本共用同一个编码器，令模型能够学习风格无关的文本表达，结果表明，这种方式有较好的迁移效果。

(Wang et al., 2019)认为在数据预处理阶段，使用传统的基于规则的文本迁移方法，会给深度模型引入噪声，因此Wang将规则方法使用在神经模型的建模过程中，大大提升了原有深度模型的文本风格迁移效果。(Zhang et al., 2020)在利用平行数据集建模的同时，引入了非平行数据，并且提出了多任务的数据增强方法，将增强后的数据作为先验知识输入到模型中，并提出了基于预训练-微调的两阶段风格迁移模型。有监督的文本风格迁移方法依赖于平行数据集，然而目前可用的平行数据集较少，数据质量也参差不一，自主构建又需要极大的资源消耗，因此，如何使用更加丰富的非平行语料，成为研究者们关注的焦点。

2.2 基于无监督学习的文本风格迁移

相较于有监督方法，无监督方法能够有效利用海量的情感分析语料，也有更广阔的应用前景。

(Shen et al., 2017)最先提出了交叉对齐方法，利用风格化判别器，结合强化学习生成对立风格文本。(Sancheti et al., 2020)在传统的Seq2Seq模型中，引入了copy机制，利用风格迁移效果、内容保存程度等指标作为网络的反馈激励，在多个数据集上取得了不错得迁移效果。利用强化学习是解决缺少平行语料的有效方法，但是仅仅靠深度模型对风格化文本建模，缺少了风格相关的先验知识的引入，利用伪对齐数据生成以及风格化嵌入方法，能够更好的解决这一问题。

基于伪对齐数据生成的方法，近来被较多研究者所使用，(Li et al., 2018)通过删除、检索以及生成关键词的方法，生成伪对齐数据。例如句子“我很快乐”，其中“快乐”是表达积极情感的关键词，将“快乐”替换成“悲伤”，即可以将句子迁移为消极的风格。*Li*提出了一种基于ngram与词频的方法，用于标记句子中的关键词。将语料中的关键词剔除后，并且利用TFIDF以及编辑距离，确定两种风格语句的内容相似度，并将相似度最高的一对语句作为目标句子对，将二者的关键词互换并且回填，即完成了伪对齐数据的生成．(Lai et al., 2019)利用神经翻译系统，减少语句的风格化表达，提取句子的内容表示。Lai首先将英文语言，翻译为中间语言法文，然后利用翻译系统中的编码器，获取法文的编码向量，将其作为无风格化表达，与目标风格的语句构成了平行数据集，进一步利用端到端的文本生成模型，即可完成文本的风格迁移．(Zhou et al., 2020)从词语入手，分析词语与语句风格之间的关系，提出了一种新的基于词级别attention的seq2seq模型，在迁移的准确率以及文本意义的保存度上取得了SOTA的结果。(Madaan et al., 2020)将标注方法引入文本风格迁移，利用标注模型生成风格无关的文本，再基于无风格文本，利用生成模型完成风格迁移任务。(Malmi et al., 2020)引入MLM模型，通过对两种风格的文本分别预训练，通过比较预训练的结果，得到风格化词语，与(Li et al., 2018)不同的是，Malmi并未使用生成模型，而是直接进行删除以及替换来完成文本风格迁移。

在其他的一些研究中，(Cao et al., 2020)将文本风格迁移技术与实际相结合，提出了一个新的任务，利用风格迁移技术，帮助专家与普通人更方便的交流。(Jhamtani et al., 2017)利用风格迁移模型，实现了现代语言到莎士比亚风格语言的相互转换。

基于伪对齐数据生成的方法，十分依赖于对齐文本的生成质量，然而也没有有效的测试方法，对伪对齐数据的正确性进行验证。而强化学习的架构，十分依赖于判别器模型的质量，难以更好的对风格进行融合。本文使用风格化嵌入的方法，能够将文本与风格标签，在同一个空间下进行表示，进而可以直接将文本内容与风格相结合。

3 数据集与方法

3.1 中文文本风格迁移数据集

现有的文本风格迁移研究集中于英文领域，而中文领域的研究较少，相应数据集也较为匮乏，本文通过标注与组合等方式，构建了三个中文文本风格迁移数据集。

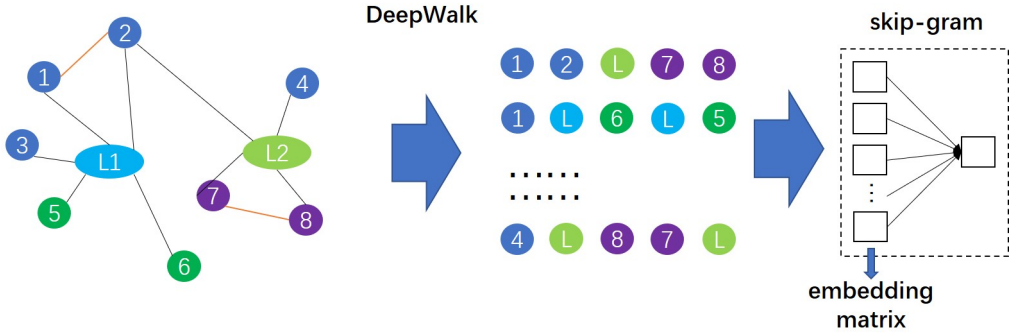

Figure 1: 基于DeepWalk的风格化嵌入方法

Weibo数据集：本文利用爬虫技术，从微博平台上爬取了5万条微博。利用字数、表达规范等规则进行初步筛选。通过多人多次标注的方式，每个人从数据集中进行随机抽样1000个样本，从文本情感极性，以及表达清晰度两个方面，进行交叉标注，其中情感极性分为积极与消极，表达清晰度为1至5的数值，越高表达越清晰。同一条数据会经过多人的标注，利用投票的方式，决定文本的最终情感极性，数据格式与示例如表 1所示。

内容	情感极性	清晰度
我真是太喜欢草莓味的冰淇淋了！！！	积极	5
他来了他来了真是太好了	积极	3
这家餐厅我再也不会去了，差评！	消极	5
我不喜欢这些，我怀疑也不会有人喜欢。	消极	4

Table 1: Weibo数据集标注示例

通过将不同标注者的结果进行合并与排序，选择其中得分最高的1万余条作为正向情感样本，另外一万余条作为负向情感样本，并构建相应的训练集以及测试集。

NLPCC数据集：除自主标注外，本文利用已有的情感分类数据集，通过人工标注，对数据的表达清晰度进行筛选，排序后选择了一部分子集构建正负情感文本集，并且划分了训练集以及测试集。

YELPCN数据集：本文基于现有的英文文本风格迁移数据集YELP，利用翻译软件获得其中文翻译，利用人工筛选，进一步选择翻译质量较好，表达较清晰的文本，并截取其中1万条训练集与测试集构建中文数据集。同时，针对测试集，利用常见的中文表达方式，给出了对立风格的中文文本，构建了500余条平行测试集。本文利用三种方式，构建了中文文本风格迁移数据集，可以更好的对模型的性能进行验证。

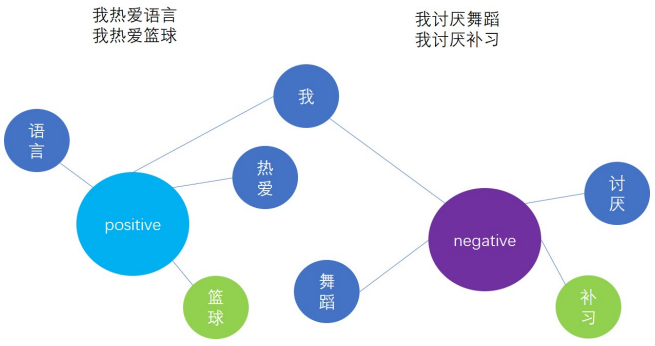

Figure 2: DeepWalk图结构

3.2 DeepWalk 风格化嵌入方法

(Cavalin et al., 2020)提出了基于Word Graph的深度游走算法，通过将Label进行隐层表达，进而解决意图识别中的的无效语义问题。本文对其算法进行了进一步改进提出了DeepStyle算法，将文本风格标签与文本内容在同一隐层进行表达。

与(Cavalin et al., 2020)不同的是，意图分类任务有多个类别标签，而本文面临的风格迁移任务只有两个类别，也因此导致Word Graph中的路径深度过短，因此，本文在图构建方法上添加了内连接率，用以提高生成图的稠密度。

同时与原有算法不同的是，本文利用深度模型，通过先预先训练再联合训练的方式，在训练风格迁移模型的同时，能够不断优化风格化向量表示。其核心算法流程如算法 1所示。

Algorithm 1 引入内连接率的DeepWalk算法

Require:
Token-Label图结构G=(V,E, α)
风格嵌入维度d
随机游走深度w
随机采样次数s

Require:
文本与标签的嵌入表示E

```
set i = 1
set Emb = Random()
while (i ≤ s) do
    O = Shuffle(V)
    for v in V do
        W = RandomWalk(G,v,w, α )
        SkipGram(W,Emb)
    end for
    i = i + 1
end while
return Emb
```

本文首先定义了如图2的图结构G，令$G=(V,E,\alpha)$，其中V为图中的节点，其构建方法如式(1)所示，节点集合由字典词语W以及类别标签L构成，二者在同一个图结构中，因此可以在同一隐层空间中进行表示学习 。

$$V=\{W_1,\dots W_n\}\cup\{L_1\dots L_m\} \tag{1}$$

边集合E的构建方法由式(2)所示，其中C表示将两个节点相连。边集合由两部分构成，首先对于训练集中每个实例，将其词语节点与对应的类别标签节点向量，但由于类别数较少，图中最长的路径受限于类别个数，因此本文以一定概率α连接同一个句子中的词语，利用函数P进行采样连接。

$$E=\sum_k C\left(W_{k1\dots kl},L_k\right)\cup P\left(C\left(W_k,W_k\right),\alpha\right) \tag{2}$$

通过以上方法构建的图结构，将风格标签以及字典中的词语置于同一空间下，利用图1所示的DeepWalk算法同时学习风格化嵌入与词嵌入表示。

$$G\overset{RandonWalk}{\longrightarrow}R^{w\times s}\overset{SkipGram}{\longrightarrow}E^{(|V|+|L|)\times d} \tag{3}$$

DeepWalk算法分为两个步骤，首先通过RandomWalk算法，从图结构中随机抽取节点，构成预训练所需的数据集$R^{w\times s}$ ，其中w是随机游走的最大深度，s是随机采样的次数。获取数据集后，通过Word2Vec中的SkipGram算法，进行预训练，即可获得词语与风格标签的嵌入表示$E(|V|+|L|)\times d$ ，其中$|V|$与$|L|$分别为字典词语数与风格标签个数，d为嵌入表示的向量维度，算法流程如表1所示。

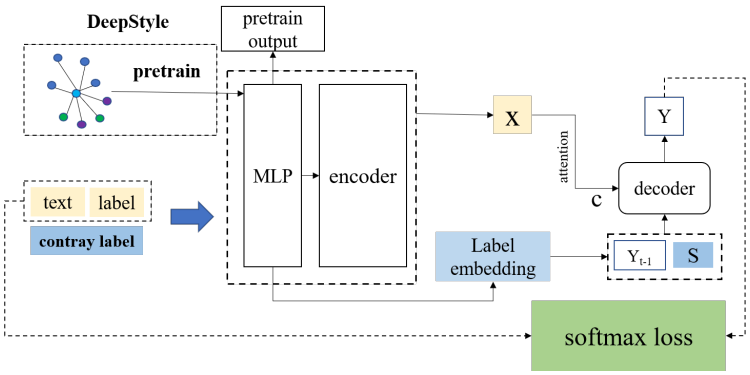

Figure 3: 基于DeepStyle的文本风格迁移模型

3.3 基于内连接DeepWalk的文本风格迁移算法

(Cho et al., 2014)提出了Seq2seq模型用于神经翻译领域，而相同的encoder-decode架构及其变种，也广泛地被应用于文本生成领域。本文利用Transformer构建编码器，利用LSTM-attention构建解码结构，结合DeepStyle算法所得的嵌入表示，构建生成端模型。其结构如图3所示。

本文使用的风格化嵌入模型DeepStyle与(Perozzi et al., 2014)的结构相同，使用双隐藏层的全连接神经网络，结合RandomWalk算法以及SkipGram进行预训练，不过与之不同的是，本文进一步将风格嵌入表示与风格迁移模型进行联合训练。

$$P\left(x_2 \mid E\left(D\left(x_1\right)\right), D\left(y_2\right)\right) \tag{4}$$

文本风格迁移任务的定义可用式(4)来表达。x_1代表输入的风格化文本，y_2为目标风格标签，D为嵌入算法，E表示Endoder结构。输入文本与对立风格标签，通过DeepStyle算法，在同一表示空间下，得到其嵌入向量$D(x_1)$与$D(y_2)$。其中编码器是由4层Transformer Layer构成，输入文本x_1的风格标签为y_1，x_1通过嵌入模块后，与所有风格化标签在同一空间进行向量化表示，再经过Transformer编码器后，可获得文本句级别的嵌入向量$E(D(x_1))$，而对立风格标签y_2经过同一嵌入模块进行表示后，可以得到对立风格向量$D(y_2)$。由此，可以将文本与风格标签在同一隐层空间进行表达。

在Decoder端，本文使用LSTM作为基模型构建解码结构，同时在解码阶段，引入attention机制，利用Encoder的输出，加强记忆状态的表示，其原理如式(5)所示。

$$LSTM\left(x_t \mid mlp\left(x_{t-1} : l\right)\right) \tag{5}$$

x_t为当前时间步的预测值，$x_(t-1)$为上一时间步的预测结果，l为对立风格嵌入表示。由图3亦可知，在解码的每一时间步，上一时间步的预测值与风格化嵌入向量进行拼接后，通过一个全连接网络进行维度转换输入到下一个时间步，通过对立风格化向量，指导解码器生成的文本更倾向于对立风格。

4 实验与结果

4.1 数据集

现有的文本风格迁移研究集中于英文领域，而中文领域数据集也较为匮乏，本文构建了三个中文文本风格迁移数据集，中文词表包含约2.2万个中文词语，并且在英文数据集上进行了实验。

Weibo：该数据集包含1万余条正向情感文本以及1万余条负向情感文本，文本的平均长度为69.4.

YelpCN：本数据集包含1万余条正向情感样本与1万余条负向情感样本，文本平均长度为20.1。

NLPCC：NLPCC数据集是NLPCC2016比赛的任务二中的文本分类数据集，该数据集包含1.6万的正向文本以及1.7万的负向文本，文本平均长度为19.1。

模型	ACC/Weibo	ACC/NLPCC	ACC/YELPCN
DRG-D	0.58	0.95	0.824
DRG-DR	0.419	0.7875	0.418
CA	0.43	0.788	0.856
ASE	0.67	0.821	0.83
DeepStyle	0.852	0.91	0.88

Table 2: 不同文本风格迁移模型在多个数据集上的效果

4.2 验证指标

风格化ACC：如公式(6)所示，本文使用fine-tuned Bert作为验证模型，利用其预测得到的ACC分数，作为风格迁移效果的指标。

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

内容保存度BLEU指标：本文利用常见中文表达方式，构建了500余条风格迁移平行测试集，因此可以利用传统的文本生成指标，检验生成文本与目标文本的一致性，该指标越高，模型性能越好。

$$PPL = P\left(w_1, w_2 \ldots w_n\right)^{-\frac{1}{N}} \tag{7}$$

流畅度PPL：PPL根据历史词，预测当前词出现的概率，以此来衡量生成文本的流畅度，该指标越小越好，计算公式如公式(7)所示。

4.3 基线模型

ASE(Kim and Sohn, 2020)：Kim基于自编码器结构，通过学习Style Embedding，进而生成对立风格文本，Kim并未给模型命名，因此本文称之为ASE模型。

DRG-D与DRG(Li et al., 2018)：Li提出了基于伪对齐数据生成的风格迁移模型，通过删除、检索与生成三个步骤，达到迁移风格的效果，其中DRG-D是仅仅使用删除方法的模型，DRG是使用完整步骤的模型。

Cross-Align(Shen et al., 2017)：Shen基于强化学习的范式，提出了基于强化学习以及交叉对齐的风格迁移模型。

模型	BLEU	PPL
DRG-D	4.23	325
DRG-DR	4.13	289
CA	3.67	303
ASE	8.32	138
DeepStyle	10.4	108

Table 3: 不同文本风格迁移模型在数据集YELPCN上的效果

4.4 实验方法

参数设置：本文取词嵌入与风格化嵌入的维度$d = 300$，取Random Walk的步长$w = 10$，采样次数$s = 2 \times |V| \div w$，其中采样次数根据数据集大小的不同而有所变化。同一句子内词语连接的概率α=0.3。

实验对比：本文将提出的DeepStyle模型与以上三个Baseline进行了对比，并且进一步验证了在不同的采样步长、采样次数以及内连接率下，模型的效果。同时为了验证风格化嵌入与词

嵌入的关联性，本文对嵌入向量进行降维，在低维度将之表示为坐标点，进一步观察二者之间的距离关系。

Teacher Forcing：在训练阶段，Decoder的输出并不准确，因此本文在训练时，使用Teacher Forcing技巧，将训练集中的上一时间的词语作为模型的输入，用以预测下一个单词，而在测试阶段，则使用上一步输出的隐层状态作为输入。

Beam Search：在解码阶段，为了进一步提高解码预测的准确率，本文使用束搜索方法，同时保留多个预测结果，并且在下一步预测时，结合之前的多个结果进行预测，进一步提高了模型的生成效果。

Warmup：本文采用联合训练的方式，将预训练的风格嵌入模型与生成模型同时训练，在训练初期，为了防止预训练后的模型效果受到噪声干扰，采用了Warmup的方式，随着训练轮次增加，训练的学习率也不断提高，直到一定阈值后，再以指数速率下降。通过该方法可以有效防止风格嵌入向量被噪声污染。

4.5 实验结果

由表2可知，现有的英文模型针对中文文本风格迁移问题大多是有效的，但对于文本较长的Weibo数据集，原有模型的迁移效果较差，本文使用的Transformer编码器能够更好的对长文本进行建模与嵌入表示。DeepStyle模型在两个数据集上取得了最好的迁移效果，在NLPCC数据集上也优于大多数模型。

DeepWalk参数	**ACC/Weibo**	**ACC/NLPCC**	**ACC/YELPCN**
$\alpha = 0.1$	0.67	0.53	0.482
$\alpha = 0.2$	0.71	0.852	0.68
$\alpha = 0.3$	0.852	0.91	0.88
w=5	0.76	0.65	0.691
w=10	0.852	0.91	0.88
w=15	0.73	0.72	0.641

Table 4: DeepWalk方法不同参数下模型表现

在第2个实验中，我们利用YELPCN数据集所提供的平行测试集，利用文本生成当中的BLEU指标对不同模型的效果进行测试。如表3所示，本文提出的DeepStyle模型获得了最佳的结果。本文进一步提供了不同数据集中，模型的风格迁移实例，结果如表5所示，可以更加具体的比对不同模型的迁移效果。

4.6 不同参数下DeepStyle算法效果

本文进一步验证了DeepWalk算法中不同参数，对于模型性能的影响。主要验证了两个关键参数，分别是采样步长w，以及内连接概率 α 。采用控制变量的方式，改变一个参数时，会控制其他参数为效果最好的参数值。

Weibo积极-消极	原句	这家餐厅我好喜欢，服务态度和厨师水平都是一流。
	ASE	这家餐厅位置不好，所以生意也没起色。
	DRG	这间餐馆还不错，可惜已经关门了。
	CA	这家餐厅我很喜欢。
	Deep	这间餐厅服务态度很差，食物很难吃，我再也不想来了。
YELPCN消极-积极	原句	自从乔接手后，这家公司的表现越来越糟。
	ASE	自从乔来到这里，一切都在变好。
	DRG	自从乔来了以后，我就特别开心。
	CA	自从他来到这里，情况就越来越糟糕。
	Deep	自从乔来了，公司越来越好了。

Table 5: 风格迁移实例

从表4中可以看出，随着内连接率 α 的不断增加，模型效果呈现增加的趋势，在内连接率较低的实验中，NLPCC数据集以及YELPCN数据集表现较差，而Weibo数据集表现相对较好，原因在于Weibo数据的平均句子长度较长，因此在相同的内连接率的情况下，内连接的词语节点较多，所构成的图更加密集，因此能够获得更好的预训练效果。本文通过PCA降维方法，将不同风格文本的词嵌入向量进行压缩表示，进而在二维空间中进行绘制，在内连接率 α 不断增加的情况下，效果如图4所示。可见，更大的内连接率有利更有利于风格嵌入模型的训练，能够使模型更好的对两种对立风格进行不同的隐层表达。

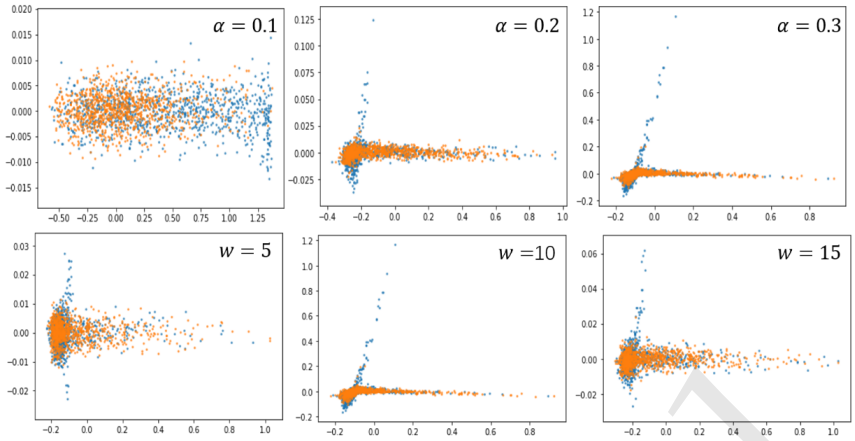

Figure 4: 不同参数下DeepStyle模型嵌入表达效果

本文同样对采样深度参数w进行了对比，由表5可知，随着采样步长的增加，模型效果呈现先增加后降低的趋势。其原因可由图5分析得出，当步长较短，采样的预训练数据中，标签节点出现的较少，因此模型难以更好地学习到风格化方面的词语表达，因此导致生成模型效果较差。同样，当采样步长过大时，数据中的词语节点远多于风格标签节点，因此风格化标签数据量被稀释，导致模型学习到的风格化表达变弱，进一步影响生成模型的效果。

4.7 Transformer Encoder 实验对比

本文使用Transformer搭建Encoder结构，与RNN模型相比，能够更好的解决长期依赖不足的问题。

编码器结构	ACC/Weibo	ACC/NLPCC	ACC/YELPCN
RNN	0.67	0.53	0.482
LSTM	0.71	0.852	0.68
GRU	0.76	0.65	0.691
Transformer	0.852	0.91	0.88

Table 6: 使用不同Encoder消融实验

LSTM通过引入记忆细胞，一定程度上缓解了长期依赖问题，然而与使用Attention的Transformer相比，其表达能力依然受限于距离，通过构建消融实验，在相同参数设置下，对比二者的效果，实验结果如表6可知，在所有数据集上，使用Transformer Encoder明显优于LSTM Encoder，尤其在文本较长的weibo数据集上，效果差距更加明显。

Transformer中的层数以及隐层维度对于模型性能有较大影响，本文进一步测试了不同参数设置下，Transformer Encoder的效果，结果如表7所示，其中L为Transformer Layer的层数，H为Self Attention的维度。可见随着层数的增加模型的，编码效果不断地增强，而且层数参数的影响要强于Attention维度的影响。

通过图5的attention可视化热力图可见，使用Transformer Encoder作为编码器，在编码过程中，风格化更加明显的词语，例如“喜欢”，在不同的得分中，普遍拥有更高的权重，进而对于整个文本的表达贡献更大，因此能够很好的加强模型对风格化文本建模的能力。

Transformer参数	ACC/Weibo	ACC/NLPCC	ACC/YELPCN
L4H300	0.852	0.91	0.88
L4H200	0.849	0.9	0.878
L4H100	0.83	0.884	0.861
L2H300	0.841	0.874	0.877
L2H200	0.83	0.865	0.87
L2H100	0.824	0.86	0.856

Table 7: Transformer Encoder不同参数效果

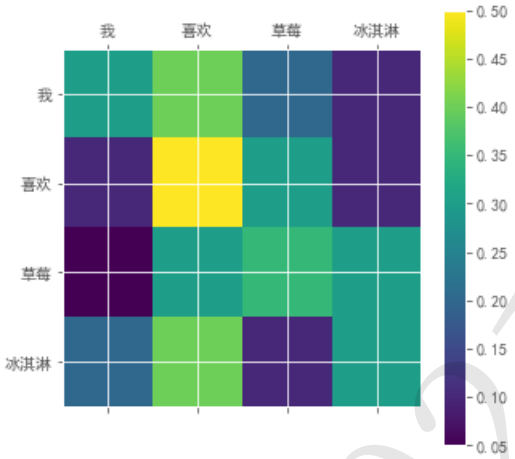

Figure 5: attention可视化热力图

5 总结

目前的文本风格迁移研究多集中于英文领域，而中文领域研究较少，数据集也较匮乏。本文构建了三个可用于中文文本风格迁移的数据集，并且将现有的模型与方法应用其上，实验可知，现有的文本风格迁移算法在中文领域是有效的，本文构建的数据集是具有可用性的。同时，本文提出了基于风格化嵌入的中文文本风格迁移模型DeepStyle，利用改进的DeepWalk算法，在同一个隐层空间，将风格化标签与文本同时进行嵌入表达，与现有方法相比，得到了更有效的风格嵌入向量。与原有算法不同的是，本文通过联合学习方式，进一步学习文本风格嵌入表达与对立风格文本生成，在中文数据集上，取得了更好的风格迁移效果。

参考文献

Yixin Cao, Ruihao Shui, Liangming Pan, Min-Yen Kan, Zhiyuan Liu, and Tat-Seng Chua. 2020. Expertise style transfer: A new task towards better communication between experts and laymen. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1061–1071.

Paulo Cavalin, Victor Henrique Alves Ribeiro, Ana Appel, and Claudio Pinhanez. 2020. Improving out-of-scope detection in intent classification by using embeddings of the word graph space of the classes. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3952–3961.

Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734.

Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg. 2017. Shakespearizing modern language using copy-enriched sequence to sequence models. In *Proceedings of the Workshop on Stylistic Variation*, pages 10–19.

Heejin Kim and Kyung-Ah Sohn. 2020. How positive are you: Text style transfer using adaptive style embedding. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2115–2125.

Chih-Te Lai, Yi-Te Hong, Hong-You Chen, Chi-Jen Lu, and Shou-De Lin. 2019. Multiple text style transfer by using word-level conditional generative adversarial network with two-phase training. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3570–3575.

Juncen Li, Robin Jia, He He, and Percy Liang. 2018. Delete, retrieve, generate: A simple approach to sentiment and style transfer. In *2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2018*, pages 1865–1874. Association for Computational Linguistics (ACL).

Aman Madaan, Amrith Setlur, Tanmay Parekh, Barnabas Poczos, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan W Black, and Shrimai Prabhumoye. 2020. Politeness transfer: A tag and generate approach. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1869–1881.

Eric Malmi, Aliaksei Severyn, and Sascha Rothe. 2020. Unsupervised text style transfer with masked language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8671–8680.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 701–710.

Reid Pryzant, Richard Diehl Martinez, Nathan Dass, Sadao Kurohashi, Dan Jurafsky, and Diyi Yang. 2020. Automatically neutralizing subjective bias in text. In *Proceedings of the aaai conference on artificial intelligence*, volume 34, pages 480–489.

A Sancheti, K Krishna, BV Srinivasan, and A Natarajan. 2020. Reinforced rewards framework for text style transfer. *Advances in Information Retrieval*, 12035:545–560.

Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 6833–6844.

Yunli Wang, Yu Wu, Lili Mou, Zhoujun Li, and Wenhan Chao. 2019. Harnessing pre-trained neural networks with rules for formality style transfer. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3564–3569.

Yunli Wang, Yu Wu, Lili Mou, Zhoujun Li, and Wenhan Chao. 2020. Formality style transfer with shared latent space. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 2236–2249.

Yi Zhang, Tao Ge, and SUN Xu. 2020. Parallel data augmentation for formality style transfer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3221–3228.

Chulun Zhou, Liang-Yu Chen, Jiachen Liu, Xinyan Xiao, Jinsong Su, Sheng Guo, and Hua Wu. 2020. Exploring contextual word-level style relevance for unsupervised style transfer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7135–7144.