

面向小样本意图识别的分步式阶段性数据增强^①



李玉茹, 张晓滨

(西安工程大学 计算机科学学院, 西安 710048)

通信作者: 张晓滨, E-mail: xiaobinzhngcn@126.com

摘要: 文本意图识别任务中常面临训练数据不足的问题, 且由于文本数据离散性导致在标签不变的条件下进行数据增强并提高模型性能具有一定困难, 为解决小样本意图识别任务中的上述问题, 提出一种分步式数据增强与阶段性训练策略相结合的方法. 该方法从全局和局部两个角度将原始数据在全体语句和同类别中的样本对上进行递进式增强, 并在模型训练期间根据递进层次的不同划分阶段进行学习, 最后在多个意图识别数据集上进行实验以评估其有效性. 实验结果表明, 该方法可以有效提高小样本环境中意图识别模型的准确率, 同时模型的稳定性也得到了提升.

关键词: 小样本; 意图识别; 数据增强; 分步式; 阶段性训练

引用格式: 李玉茹, 张晓滨. 面向小样本意图识别的分步式阶段性数据增强. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/8891.html>

Stepwise and Phased Data Augmentation for Few-shot Intent Detection

LI Yu-Ru, ZHANG Xiao-Bin

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: Insufficient training data is often faced in the task of text intent detection, and due to the discreteness of text data, it is difficult to perform data augmentation and improve the performance of the original model with the unchanged label. This study proposes a method combining stepwise data augmentation with a phased training strategy to solve the above problems in the few-shot intent detection. The method progressively augments the original data on whole statements and sample pairs in the same category from both global and local perspectives. During model training, the original data is learned according to different partition stages of the progressive level. Finally, experiments are performed on multiple intent detection datasets to evaluate the validity of the method. The experimental results show that the proposed method can effectively improve the accuracy and the stability of the few-shot intent detection model.

Key words: few-shot; intent detection; data augmentation; stepwise; phased training

基于深度神经网络的模型往往需要大量标记数据才能够成功, 然而在意图识别等诸多仅包含少量标注数据的小样本任务中训练数据的有限使得模型容易过拟合, 导致泛化能力的丧失. 为解决这一问题, 采用数据增强的方式在不显式地收集新数据的情况下增加训练实例的多样性是一个有效的方法. 虽然在计算机视觉和语音领域应用数据增强已经取得了一定成功, 但是由于文本数据的离散性, 图像和语音领域一些通用

的数据增强技术并不能确保在自然语言处理领域中中文本数据的语义不变性和标签正确性. 近年来, 人们开始关注在低资源、小样本情境下如何在自然语言处理领域进行高效的数据增强^[1-3], 以此缓解由于资源稀缺导致诸多任务无法高质量完成的现象, 其中就包含对话系统中的诸多任务. 由于某些领域的对话数据资源稀缺, 如何在训练数据稀少的情况下提高对话系统的性能显得尤为重要, 又因为意图识别的准确性直接影响

^① 基金项目: 陕西省自然科学基金 (2019JQ-849)

收稿时间: 2022-05-15; 修改时间: 2022-06-15; 采用时间: 2022-06-29; csa 在线出版时间: 2022-08-26

着对话系统的性能,因此利用数据增强技术提升小样本情境下的意图识别准确性具有很高的研究价值。

数据增强方法大致分为基于生成和基于编辑两类。利用受限变分自编码器^[4]进行文本数据增强的方式属于基于生成的方法。随着对抗神经网络^[5]的发展以及不断进步,对抗生成技术也被应用于数据增强方面,如利用一阶扰动进行多次迭代寻找最优扰动的对抗训练算法 PGD^[6]以及针对生成对抗示例所需的时间开销进行改进的 FreeAT^[7]都在图像领域取得了较好的成果,在文本的词嵌入上增加扰动便可以将 PGD、FreeAT 等数据增强方法应用于自然语言处理领域。Zhu 等人^[8]提出了对抗性训练方法 FreeLB,通过在词嵌入中添加对抗性扰动并最小化输入样本周围区域内的对抗性风险来提高嵌入空间的不变性,从而提高模型的泛化能力和自然语言理解能力。Zhang 等人^[9]提出了一种基于模型的对抗性文本生成学习框架,引入引导网络对生成环境进行建模,从而为下一个词的选择提供指导,以此缓解奖励稀疏问题,进一步优化了生成器。随着预训练语言模型的发展,条件生成方法也相继产生。Wu 等人^[10]提出了条件生成模型 CBERT,在 BERT^[11]微调阶段引入类标签信息预测提前设定的单词,以此得到广泛的替换词修改原始语句,生成新的文本数据。Anaby-Tavor 等人^[12]提出了一种 LAMBADA 的方法,通过利用现有的标记数据进行模型初始训练,将语言生成器微调到特定任务,使用微调模型和类标签来生成新的句子,其中通过提供的类别信息作为微调模型的输入来为给定类生成示例。

在上述基于生成的方法中,模型预训练和解码的成本太高,并且只适用于特定任务,一旦任务改变,则需要重新微调、训练。相比于此,基于编辑的文本数据增强方法成本低,可操作性较强。Wei 等人^[13]提出了一种称为简单数据增强算法,利用一组词级的随机扰动操作如同义词替换、随机插入和随机交换实现数据增强。Xie 等人^[14]提出无监督数据增强方法,该技术的核心思想是将无标签文本数据中一定比例的不重要的词替换为字典中的词,从而产生新的文本。Haralabopoulos 等人^[15]使用句子排列来扩充初始数据集,同时保留数据集的关键词统计属性。

上述方法大多只对句子进行一次操作,扩充数量有限,若对同一句子的操作次数过多,则会改变原有的标签类别,从而导致更差的结果。此外,在上述方法

的训练过程中,原始数据和利用数据增强技术获得的数据均被平等对待,并以完全随机的顺序呈现,并未充分考虑二者之间的差异,在一定程度上限制了性能的增益。

本文针对基于编辑的数据增强方法受标签类别约束而对原始语句扩充次数受限以及在训练过程中忽视了增强数据和原始数据差异的问题,提出了一种应用于小样本意图识别任务的分步式数据增强和阶段性训练策略相结合的方法。该方法首先从全局角度对原始数据进行初步扩充,然后以类为单位,依次在每个类的内部选取成对样本进行数据增强,以此解决扩充次数过多使得标签发生迁移从而导致最终结果下降的问题。其次,在模型训练过程中,因利用数据增强技术获得的数据相较于原始数据含有更多的噪声,本文将原始数据和分布式阶段性增强的数据按照批次划分,逐步输入模型,使得模型在噪声的波动下表现得更为稳健,从而提高了模型性能。

1 分步式阶段性数据增强

1.1 分步式阶段性数据增强处理流程

首先从全局考虑,利用常规的基于编辑的数据增强技术进行限量增强。本文选取 EDA 技术并按照 1:4 的原始数据与增强数据比例对原始数据进行同义词替换、随机交换、插入和删除操作,从而得到初步增强的数据。随后从局部角度出发,在每个类的内部多次随机选择两个样本,在嵌入层利用样本对进行线性插值 MixEmb 操作,从而得到进一步的增强数据。

由于原始数据和增强数据之间存在的差异性,模型的训练过程分 3 个阶段进行。首先将原始数据输入意图识别模型进行训练,当初始迭代次数达到指定数值时加入初步增强的数据进行第 2 阶段训练,直至达到指定的二次迭代次数后加入 MixEmb 增强的数据开始模型的第 3 阶段训练,当模型完成指定迭代次数后,整个训练过程结束。

整体流程图如图 1 所示。

1.2 分步式数据增强

由于常规的基于编辑的数据增强方法是在语义层面对原始数据进行扩充,若操作次数过多则会直接改变原有语义,并且有研究表明^[16]更高级别的表示扩展了特征空间中合理数据点的相对体积,从而缩小了分配给不太可能的数据点的空间,因此与输入空间相比,

数据增强更可能在特征空间中获得更真实的样本. 由于在小样本情境下过渡缺乏原始数据, 因此本文提出的分步式数据增强在对原始语句进行限量的基于编辑的数据增强后, 再利用数据在特征空间中的区域性进一步得到标签不变的增强数据.

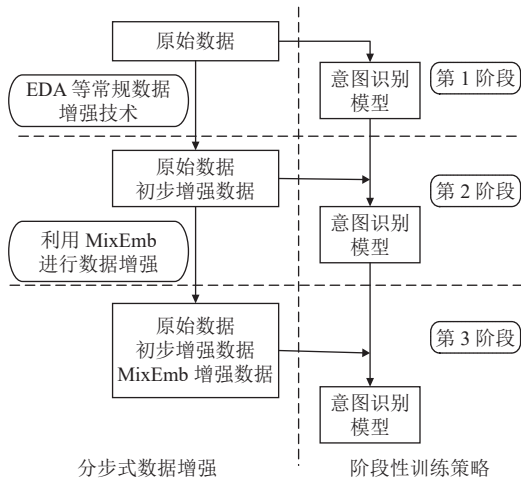


图1 整体流程图

分步式数据增强首先从全局角度利用常规的基于编辑的数据增强方法对原始数据进行初步扩充, 以得到部分既保留原有语义又具有正确标签的增强数据, 随后为了保证标签准确性的同时增加训练样本的数量, 本文在每个类的内部, 利用样本对进行 MixEmb 操作来合成新数据. MixEmb 操作是对图像分类中的 mixup^[17]方法的扩展. Mixup 的主要思想是通过线性插入一对训练样本来生成合成样本. 具体来说就是给定两个已标记的数据点 (x_i, y_i) 和 (x_j, y_j) , 其中 x 可以是一幅图像, y 是标签的表示, 算法通过线性插值创建虚拟训练样本训练神经网络模型:

$$\begin{cases} x' = \text{mix}(x_i, x_j) = \lambda x_i + (1 - \lambda)x_j \\ y' = \text{mix}(y_i, y_j) = \lambda y_i + (1 - \lambda)y_j \end{cases} \quad (1)$$

其中, $\lambda \in [0, 1]$.

直接在原始数据上进行的数据增强操作具有高度的领域相关性, 而在特征空间上进行增强操作则是与领域无关的. Mixup 操作就是在特征空间上减少了部分数据盲区, 因此本文将文本数据映射到特征空间之后为保证增强数据的标签准确性, 选取同类别中的样本对在特征空间中进行 MixEmb 操作, 这样便扩展了特征空间内合理数据点的相对数量, 与在原始空间中的数据增强形成了互补. MixEmb 操作如下:

$$\begin{cases} \tilde{x} = \lambda E(x_i) + (1 - \lambda)E(x_j) \\ \tilde{y} = y_i = y_j \end{cases} \quad (2)$$

其中, $E(x_i)$ 表示 x_i 通过预训练语言模型映射到特征空间后的表示, λ 是样本对的混合比率, 并且 λ 是从具有超参数 α 的 $Beta(\alpha, \alpha)$ 分布中采样的. 因为此时是在同类别中选取样本对进行线性插值, 所以得到的训练样本的标签类别保持不变.

1.3 阶段性训练策略

现有研究表明^[18], 当示例不是随机呈现时, 而是以有意义的顺序组织起来, 逐渐显示出更多的概念和复杂性, 可以有效引导训练过程在参数空间中向更好的局部最小值学习, 从而使得神经网络训练得更好. 一般假设数据中存在一系列示例难度, 然后根据难度系数进行排序, 逐步增加困难数据训练模型.

在本文提出的分步式数据增强方法中, 将示例难度的评判指标设定为训练数据中所含有的噪声量. 由于初步增强的数据是按照一定的扰动比例获得的, 所以相比于原始数据而言具有一定的噪声, 属于较为困难的训练示例. 利用 MixEmb 获得的增强数据是在初步增强的基础上进行扩充, 噪声含量存在叠加现象, 从而进一步加大了模型学习的难度. 由此可见, 在整个分步式数据增强的过程中, 训练数据的学习难度随着数据增强深度的增加而增加.

因此为了从简单训练数据开始, 逐步过渡到困难数据, 本文根据数据增强的递进程度将训练数据划分成原始数据、初步增强数据和利用 MixEmb 增强的数据 3 个层次. 在模型训练初期, 选择难度系数较低的原始数据进行训练, 当迭代次数达到指定数值后, 加入较难学习的初步增强数据进行第 2 阶段的训练, 在第 2 阶段训练完成后加入学习难度最高的 MixEmb 增强数据进行第 3 阶段的训练, 从而完成整个阶段性训练的过程.

1.4 分步式数据增强与阶段性训练策略相结合的实现步骤

分步式数据增强和阶段性训练策略相结合则是从数据本身以及模型训练两个角度进行了数据增强. 二者结合的具体步骤如下.

步骤 1. 输入原始数据 $data$, 初始迭代次数 n_1 , 二次迭代次数 n_2 , 三次迭代次数 n_3 .

步骤 2. 分步式数据增强.

(1) 利用 EDA 技术^[13]对原始数据进行初步增强,

得到初步增强数据 $data_1$, 初步增强数据和原始数据的数量比是 4:1.

(2) 利用预训练语言模型 BERT-base 对原始数据 $data$ 和初步增强数据 $data_1$ 进行处理, 从而得到相应的词向量.

(3) 随机选取同类别中源自不同原始数据的词向量对进行 MixEmb 操作, 从而得到最终的增强数据 $data_2$, 其中操作次数是原始数据的类别数乘以初步增强数据和原始数据的比例.

步骤 3. 搭建三元组网络作为意图识别模型, 三元组网络结构图如图 2 所示.

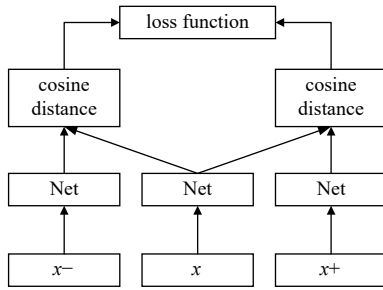


图 2 三元组网络结构图

三元组网络由 3 个相同的子网络组成, 每次输入 3 个样本 x 、 x^+ 和 x^- , 其中 x^+ 和 x 属于同一类的样本, 而 x 和 x^- 属于不同类的样本. 三元组网络输出两个中间值, 分别是相同类别样本 x 、 x^+ 通过网络表示后二者的距离和两个不同类样本 x 和 x^- 通过网络表示后的距离. 将子网络表示为 Net , 则三元组网络在倒数第二层输出的将是两个向量:

$$TripletNet(x, x^+, x^-) = \begin{bmatrix} \cos(Net(x), Net(x^+)) \\ \cos(Net(x), Net(x^-)) \end{bmatrix} \quad (3)$$

损失函数为:

$$Loss(d^+, d^-) = d^+ - d^- + margin \quad (4)$$

其中,

$$d^+ = \cos(Net(x), Net(x^+)) \quad (5)$$

$$d^- = \cos(Net(x), Net(x^-)) \quad (6)$$

通过训练, 让同类样本之间的距离尽可能小, 不同类样本之间的距离尽可能大, 并且要让上述两个距离之间有一个最小的间隔 $margin$.

步骤 4. 阶段性训练策略.

(1) 输入原始数据 $data$ 对模型进行训练, 优化器采

用 Adam 优化器.

(2) 当迭代次数达到 n_1 时, 将初步增强的数据 $data_1$ 加入模型进行训练.

(3) 当迭代次数达到 n_2 时, 将利用 MixEmb 增强的数据 $data_2$ 加入模型进行训练, 直到迭代次数达到 n_3 时, 模型的整个训练过程结束.

2 实验分析

2.1 数据集

本文首先在 SNIPS、NLUE、ATIS 三个意图识别任务的基准数据集上评估本文的方法, 其中 SNIPS 和 NLUE 是在人机交互过程中收集的数据, 分别含有 7 个和 64 个类别的意图. ATIS 是有关航班信息查询的数据, 含有 17 个意图类别信息, 特点是数据分布不平衡. 此外, 为了进一步验证所提出的方法是否适用于关系分类和问题分类, 本文附加使用了关系分类数据集 FewRel 和问题分类数据集 TREC.

2.2 实验设置

本文的实验是基于 PyTorch 实现, 使用标准的 BERT-base 对句子进行编码, 选取三元组网络为分类器, 三元组网络的架构参考文献 [19] 实现, 采用余弦距离表示样本之间的关系, 阈值设定为 0.4, 学习率设定为 2×10^{-5} , 批量大小设定为 64. 在验证分步式数据增强有效性时, 初步的常规数据增强本文采用的是 EDA^[14] 数据增强技术, 扰动比例设置为 0.2, MixEmb 操作中混合比例 λ 来自 $Beta(\alpha, \alpha)$ 分布, α 设定为 0.2. 在进行阶段性训练时, 3 个阶段的迭代次数比是 1:1:1. 此外, 本文每个类仅选取 10 个训练示例来构建小样本场景, 其中由于 ATIS 数据集中有的类别所含有的示例不足 5 个, 故每个类选取的是 4 个示例.

2.3 实验结果与分析

2.3.1 分步式数据增强和阶段性训练策略的有效性

本文对比了在小样本意图识别情境下, 无数据增强、常规的数据增强、分步式数据增强以及分步式数据增强和阶段性训练策略相结合时模型效果的不同, 并且本文还额外验证了在小样本情境下的关系分类和问题分类任务中所提出的方法对模型的性能是否有增益.

表 1 总结了在多种常规的数据增强情况下同一模型在不同数据集上的分类准确率. 其中, EDA 技术^[13] 是对原始语句进行同义词替换、随机交换、插入和删除操作, SwitchOut^[20] 是从词汇表中均匀采样的随机单

词替换原始语句中的单词, Permutation^[15]操作则是在保留原始数据集的关键统计属性的前提下对每句话进行重新排列。

表1 不同增强策略下的模型准确率

增强策略	SNIPS	NLUE	ATIS	TREC	FewRel
无数据增强	0.843	0.691	0.432	0.622	0.421
EDA	0.872	0.631	0.451	0.639	0.437
SwitchOut	0.894	0.600	0.391	0.640	0.438
Permutation	0.887	0.621	0.392	0.637	0.433
本文方法	0.891	0.632	0.477	0.666	0.441

首先, 在小样本情境下, 与未使用任何数据增强技术相比, 意图识别模型在使用本文提出的分步式数据增强和阶段性训练策略相结合的方法后, 准确率最高提升了 4.8%, 并且模型在关系分类数据集 FewRel 和问题分类数据集 TREC 上的分类准确率也得到了提升, 这说明本文提出的方法可以明显提高模型的分分类准确率。其次, 相较于 EDA、SwitchOut 和 Permutation 这类常规的数据增强策略, 本文提出的数据增强方法在 5 个数据上的平均准确率分别高出了 1.54%、2.88%、2.74%, 这说明本文提出的数据增强方法在模型分类准确率上普遍优于 EDA、SwitchOut 和 Permutation 这类常规的数据增强策略。

表 2 总结了本文方法中的分步式数据增强和阶段性训练两种策略在同一模型不同数据集中的分类准确率。与未使用任何数据增强技术相比, 分步式数据增强在 5 个数据集上的平均准确率高出了 1.56%, 当分步式数据增强策略和阶段性训练策略相结合时, 模型的平均准确率再次提升了 0.4%。这表明, 分步式数据增强和阶段性训练策略对模型性能的提升都有一定的促进作用, 并且相比较阶段性训练策略, 分步式数据增强方法对模型性能的增益更为显著。

表2 分步式数据增强和阶段性训练不同策略的模型准确率

增强策略	SNIPS	NLUE	ATIS	TREC	FewRel
无数据增强	0.843	0.691	0.432	0.622	0.421
分步式数据增强	0.888	0.629	0.474	0.658	0.438
分步式数据增强 +阶段性训练策略	0.891	0.632	0.477	0.666	0.441

此外, 对于 ATIS 数据集, 本文提出的方法与没有使用数据增强相比模型的分分类准确率提高了 4.5%, 与常规的数据增强策略相比模型的分分类准确率增长了 2.6%。然而 ATIS 数据集的类别信息很不平衡, 与飞行

航班相关的意图具有充足的训练数据, 但是其他个别种类只有几个标注数据。这说明本文的数据增强方法对于处理数据不平衡的问题有一定的帮助, 缓解了数据信息失衡的问题。

2.3.2 数据集大小对实验结果的影响

本文研究了在不同数据集大小的情况下所提出的方法对于模型意图识别准确率的影响情况。表 3 显示了在 SNIPS 数据集上进行实验的结果, 其中 N 的数值代表每个意图类别中含有的训练示例数量。

表3 本文方法在不同数据集大小下的准确率

增强策略	$N=3$	$N=5$	$N=10$	$N=20$	$N=50$	$N=100$
无数据增强	0.737	0.812	0.843	0.914	0.933	0.943
本文方法	0.745	0.841	0.891	0.923	0.938	0.947

从表 3 可以看出, 随着训练数据的增多, 本文的数据增强方法对于模型性能的提升作用也会越来越弱。这说明在小样本情境下, 通过本文的数据增强方法可以有效的增加模型所需学习的数据信息, 当有效的训练数据相对充分时, 所提供的信息也较为完整, 数据增强所带来的效益也就随之减弱。

2.3.3 增强数据的扰动程度对实验结果的影响

利用数据增强技术获得的数据具有一定的噪声, 在本文的设置中扰动系数的大小代表了最终获得的增强数据的学习难度。因此, 本节研究了增强数据的扰动程度对于实验结果的影响, 比较了不同扰动情况下, 模型在常规数据增强 EDA 方法与分步式数据增强和阶段性训练策略相结合的方法作用下分类准确率的变化情况。图 3 显示了在不同扰动情况下, 模型在 5 个数据集上分类准确率的平均值。

在扰动较低时, 常规的数据增强方法 EDA 和本文的方法所获取到的数据和原始数据都非常相似, 因此在模型效果的改进上差异并不是特别明显。但是随着扰动的增加, 本文提出的方法对于模型的分分类准确率影响要明显高于常规的数据增强方法。从整体来看, 本文提出的方法随着扰动系数的增加性能的波动也更为平缓, 能够使模型保持较为稳健的改进。

2.3.4 数据增强的迭代轮次对实验结果的影响

本文研究了数据增强的迭代轮次对于本文方法的影响。表 4 总结了在不同数据集上进行实验的结果, 其中 $I=1$ 代表只进行一次分步式数据增强。从表 4 可以看出迭代轮次的增加并不能提高意图识别模型的分分类准确率。这说明迭代多轮次的分步式数据增强, 虽然样

本总量会得到大幅度的增加,但是得到的样本所代表的信息会有大量重叠,从而使得模型出现过拟合现象,最终导致意图识别模型的分​​类准确率降低。

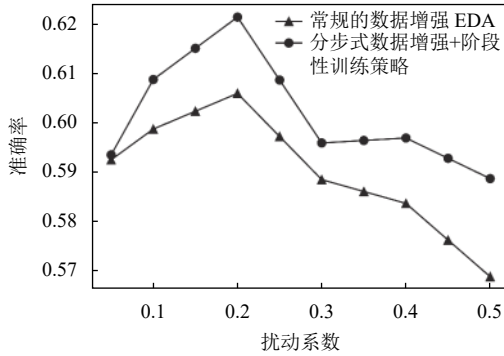


图3 不同扰动下模型准确率

表4 本文方法在不同迭代轮次下的准确率

增强策略	$l=1$	$l=5$	$l=10$
SNIPS	0.891	0.783	0.422
NLUE	0.632	0.562	0.311
ATIS	0.477	0.401	0.219

2.3.5 数据可视化

为了清晰的展示利用分步式数据增强获得的数据,本文利用 t-SNE 对分步式增强的数据进行可视化处理。图4展示了 SNIPS 数据集在分步式数据增强过程中的可视化效果。图中的聚类代表着不同的样本类别,从图中可以看出,初步的数据增强增加了一定的样本量,当完成整个分步式数据增强后,样本在保持原有聚类的基础上得到了进一步扩充。

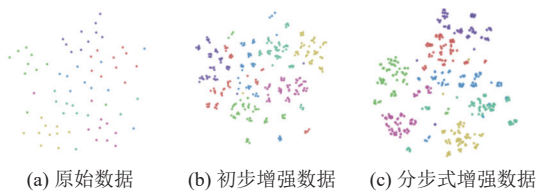


图4 分步式增强数据的可视化效果图

3 总结

本文针对在文本领域的小样本意图识别任务中训练数据不足的问题提出了一种分步式数据增强与阶段性训练策略相结合方法,首先从全局角度对整体数据进行统一的数据增强获得初步的增强数据,之后利用同类别样本对之间标签的一致性,在嵌入层进行数据增强,从而在保留原有标签的基础上达到数据增强效

果。最后,将最终获得的所有数据依据分步式数据增强的不同阶段进行划分,逐步输入意图识别模型进行训练。实验结果表明,在小样本意图识别任务中,本文提出的方法可以有效提高意图识别模型的分​​类准确率,并且模型的稳定性也有一定的提升。

参考文献

- Hou YT, Lai YK, Wu YS, *et al.* Few-shot learning for multi-label intent detection. Proceedings of the 35th AAAI Conference on Artificial Intelligence. Online: AAAI Press, 2021. 13036–13044.
- Dopierre T, Gravier C, Logerais W. Protaugment: Unsupervised diverse short-texts paraphrasing for intent detection meta-learning. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021. 2454–2466. [doi: 10.18653/v1/2021.acl-long.191]
- 王新哲, 于泽沛, 时斌, 等. 基于元学习的小样本数据生成算法. 计算机系统应用, 2021, 30(9): 161–170. [doi: 10.11888/j.cnki.csa.008063]
- Malandrakis N, Shen MM, Goyal A, *et al.* Controlled text generation for data augmentation in intelligent artificial agents. Proceedings of the 3rd Workshop on Neural Generation and Translation. Hong Kong: Association for Computational Linguistics, 2019. 90–98. [doi: 10.18653/v1/D19-5609]
- Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2014. 2672–2680.
- Madry A, Makelov A, Schmidt L, *et al.* Towards deep learning models resistant to adversarial attacks. Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018. 1–28.
- Shafahi A, Najibi M, Ghiasi A, *et al.* Adversarial training for free! Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 3358–3369.
- Zhu C, Cheng Y, Gan Z, *et al.* FreeLB: Enhanced adversarial training for natural language understanding. Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: OpenReview.net, 2020. 1–14.
- Zhang RY, Chen CY, Gan Z, *et al.* Improving adversarial

- text generation by modeling the distant future. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020. 2516–2531. [doi: [10.18653/v1/2020.acl-main.227](https://doi.org/10.18653/v1/2020.acl-main.227)]
- 10 Wu X, Lv SW, Zang LJ, *et al.* Conditional BERT contextual augmentation. Proceedings of the 19th International Conference on Computational Science. Faro: Springer, 2019. 84–95.
- 11 Devlin J, Chang MW, Lee K, *et al.* BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019. 4171–4186. [doi: [10.18653/v1/N19-1423](https://doi.org/10.18653/v1/N19-1423)]
- 12 Anaby-Tavor A, Carmeli B, Goldbraich E, *et al.* Do not have enough data? Deep learning to the rescue! Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020. 7383–7390.
- 13 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. Hong Kong: Association for Computational Linguistics, 2019. 6382–6388. [doi: [10.18653/v1/D19-1670](https://doi.org/10.18653/v1/D19-1670)]
- 14 Xie QZ, Dai ZH, Hovy E, *et al.* Unsupervised data augmentation for consistency training. Proceedings of the 34th Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 6256–6268.
- 15 Haralabopoulos G, Torres MT, Anagnostopoulos I, *et al.* Text data augmentations: Permutation, antonyms and negation. Expert Systems with Applications, 2021, 177: 114769. [doi: [10.1016/j.eswa.2021.114769](https://doi.org/10.1016/j.eswa.2021.114769)]
- 16 DeVries T, Taylor GW. Dataset augmentation in feature space. Proceedings of the 5th International Conference on Learning Representations. Toulon: OpenReview.net, 2017. 1–12.
- 17 Zhang HY, Cissé M, Dauphin YN, *et al.* Mixup: Beyond empirical risk minimization. Proceedings of the 6th International Conference on Learning Representations. Vancouver: OpenReview.net, 2018. 1–13.
- 18 Xu BF, Zhang LC, Mao ZD, *et al.* Curriculum learning for natural language understanding. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: Association for Computational Linguistics, 2020. 6095–6104. [doi: [10.18653/v1/2020.acl-main.542](https://doi.org/10.18653/v1/2020.acl-main.542)]
- 19 Wei J, Huang CY, Vosoughi S, *et al.* Few-shot text classification with triplet networks, data augmentation, and curriculum learning. Proceedings of 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg: Association for Computational Linguistics, 2021. 5493–5500. [doi: [10.18653/v1/2021.naacl-main.434](https://doi.org/10.18653/v1/2021.naacl-main.434)]
- 20 Wang XY, Pham H, Dai ZH, *et al.* SwitchOut: An efficient data augmentation algorithm for neural machine translation. Proceedings of 2018 Conference on Empirical Methods in Natural Language Processing. Brussels: Association for Computational Linguistics, 2018. 856–861. [doi: [10.18653/v1/D18-1100](https://doi.org/10.18653/v1/D18-1100)]

(校对责编: 牛欣悦)