

西安电子科技大学

硕士学位论文



小样本算法及其在关系抽取中的应用研究

作者姓名_____郝小慧_____

学校导师姓名、职称_____李雁妮 教授_____

企业导师姓名、职称_____刘志鹏 高工_____

申请学位类别_____工程硕士_____

学校代码 10701
分 类 号 TP39

学 号 18031211513
密 级 公开

西安电子科技大学

硕士学位论文

小样本算法及其在关系抽取中的应用研究

作者姓名：郝小慧

领 域：计算机技术

学位类别：工程硕士

学校导师姓名、职称：李雁妮 教授

企业导师姓名、职称：刘志鹏 高工

学 院：计算机科学与技术学院

提交日期：2021 年 6 月

Research on Few-Shot Learning Algorithm and Its Application of Relation Extraction

A thesis submitted to
XIDIAN UNIVERSITY
in partial fulfillment of the requirements
for the degree of Master
in Engineering

By

Hao Xiaohui

Supervisor: Li Yanni Title: Professor

Supervisor: Liu Zhipeng Title: Senior Engineer

June 2021

西安电子科技大学 学位论文独创性（或创新性）声明

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文若有不实之处，本人承担一切法律责任。

本人签名： 郝小慧

日 期： 2021.6.1

西安电子科技大学 关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留送交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名为西安电子科技大学。

保密的学位论文在____年解密后适用本授权书。

本人签名： 郝小慧

导师签名： 李雁冰

日 期： 2021.6.1

日 期： 2021.6.1

摘要

随着神经网络的发展,深度学习得益于从大量有标签数据中学习与拟合,今天它已被广泛应用于计算机视觉、语音以及自然语言处理等应用领域。但如何从少量样本中进行有效地深度学习?即**小样本问题(Few Shot Samples)**,是目前深度学习以及应用领域里亟待解决的关键难题。**关系抽取(Relation Extraction)**是指根据一对给定实体的相关上下文对它们之间的语义关系进行分类,它是自然语言处理中一个基础关键问题。然而关系抽取常常遭受长尾问题的困扰与挑战,即大量关系类型的实例样本非常少。因此,解决深度学习中的小样本问题,以及在关系抽取应用中的小样本问题具有重要的理论意义与应用价值。

尽管已提出了一些深度学习以及应用领域的小样本问题的解决方案,但目前已有方案还存在着以下缺陷:(1)训练样本过少导致的过拟合问题;(2)已有算法的精度有待于进一步提高;(3)由于文本不能直接计算,现有的小样本学习算法不能直接应用于关系抽取任务。

论文选题来源于国家自然科学基金项目,针对深度学习中的小样本问题,以及在关系抽取中的小样本问题,本文进行了较深入的探索研究。论文的主要工作及其创新如下:

(1)针对已有小样本学习算法精度较低的问题,在深度网络模型层面,通过引入残差注意力机制的双路膨胀卷积网络,以更充分地提取数据中不同尺寸的特征,以及对重要特征的关注度从而有效地提高算法的精度;在算法层面,针对小样本过拟合问题,通过提出加入 l_2 正则化项的损失函数,从而限制模型复杂度以缓解了过拟合问题;同时,针对小样本学习中没有有效利用自监督信息的问题,本文提出一种**基于自监督辅助任务的预训练方法**,通过构建的自监督辅助任务增强了监督信息。基于上述所提出的机制,并结合目前最新最好的**元学习(Meta Learning)+迁移学习(Transfer Learning)**的思想,本文提出了一种新的高效的**基于自监督的元迁移小样本学习算法 SS-MTL (Self-Supervised based Meta Transfer Learning, SS-MTL)**。

(2)为了解决在关系抽取中的小样本问题,并保证所提出的小样本关系抽取学习算法高效性与准确性,本文将元学习思想引入到小样本的关系抽取任务中。针对小样本学习算法无法有效利用文本数据的语义信息的问题,基于对文本数据的预处理,本文提出一种**基于匹配度的实例重表示方法**,突出对测试样本中有益的训练样本;并通过添加任务过滤策略,以有效提高算法精度;针对元学习中共享学习率导致参数适应效率低的问题,本文提出一种类名与实例信息融合的模型参数快速适应方法,为每个实例计算不同的学习率,从而更好地对模型参数进行快速适应。基于上述所提出的各

种策略/机制，本文提出了一种新的高效的信息融合的元学习小样本关系抽取算法 IFML-FSRE (**I**nformation **F**used **M**eta-**L**earning for **F**ew-**s**hot **R**elation **E**xtraction, IFML-FSRE)。

(3)作者基于权威的公开数据集，对本文提出的上述两个算法与最新相关算法进行了对比，并对本文提出的小样本学习算法 SS-MTL 进行了鸟类识别与手写字符识别的应用研究。上述实验结果表明：本文所提出的算法取得了最新最好的精度。

由于目前大部分小样本学习算法及本文提出的 SS-MTL 算法都是针对计算机视觉领域，而文本不能直接计算，因此这些算法不能直接应用于关系抽取领域。作者今后的努力方向是进一步改进本文提出的小样本学习算法使其能同时应用于计算机视觉与文本领域，以及在真实的符合长尾分布的数据集上对本文提出的小样本关系抽取算法的性能进行进一步的性能实验验证。

关 键 词：小样本学习，关系抽取，自监督，膨胀卷积，匹配信息

ABSTRACT

With the development of deep neural networks, deep learning technology is widely used in various fields such as computer vision, speech, and natural language processing. The success of this technology benefits from the acquisition of a large amount of labeled data. How to learn from few numbers of samples is a key issue to be solved in the field of deep learning. Therefore, few-shot learning arises at the historic moment. Few-shot learning is also widespread in the field of relation extraction. The relation extraction task aims to classify the relationship between two given entities according to their relevant context, which is a very important task in natural language processing. Relation extraction has received great attention in recent years. However, relation extraction often suffers from the long-tail problem, that is, there are very few instances of a large of relationships. How to learn from a few examples is also a very important issue in relation extraction. Therefore, few-shot learning has become a new challenge in the field of deep learning and relation extraction.

Many scholars have proposed a series of methods to solve the few-shot learning problem of deep learning, and conducted research on the problem of few-shot relation extraction. However, the current few-shot learning algorithms still cannot solve the following problems: (1) Overfitting caused by too few training samples. (2) Compared with traditional deep learning, the performance of these algorithms needs to be further improved. (3) Since text cannot be directly calculated, traditional few-shot learning algorithms cannot be directly applied to relation extraction tasks.

The topic of the thesis is from a project of the National Natural Science Foundation of China. Regarding the issues above, a new few-shot learning algorithm and a new few-shot relation extraction algorithm are proposed.

(1) Aiming at the problem of low accuracy of few-shot learning algorithms, a two-way dilated convolutional network with residual attention mechanism is proposed in this thesis. It more fully extracts features of different sizes in the data and improves the network's attention to important features. The accuracy of the algorithm is effectively improved. In view of the overfitting of current few-shot learning algorithms, a loss function that adds a l_2 regularization term is proposed, which alleviates the problem of over-fitting by limiting the complexity of the model. Aiming at the problem of lack of supervision information for

few-shot learning, a self-supervised auxiliary task is added to pre-train the model, which augments the supervision information. Based on the above ideas, a self-supervised based meta transfer few-shot learning algorithm SS-MTL is proposed in this thesis.

(2) The model of SS-MTL few-shot learning algorithm proposed in this thesis is too complicated and low in efficiency. In order to ensure the efficiency and accuracy of applying few-shot learning algorithms to the relation extraction tasks, the simple meta-learning algorithm MAML is applied to the relation extraction tasks. Aiming at the problem that the semantic information of text data cannot be used effectively in the few-shot learning algorithms, an instance re-representation method based on matching degree is proposed, which highlights the training samples that are beneficial to the test samples. A task filtering strategy is added to effectively improve the accuracy of the algorithm. Aiming at the problem of low parameter adaptation efficiency caused by the shared learning rate of MAML, a rapid adaptation method of model parameters based on the fusion of class name and instance information is proposed in this thesis. Different learning rates for each instance are calculated to make the model parameters be better quickly adapted. Based on the above ideas, an information fusion meta-learning few-shot relation extraction algorithm IFML-FSRE is proposed.

(3) The two algorithms proposed in this thesis are compared respectively with the latest related algorithms based on authoritative public data sets. The few-shot learning algorithm SS-MTL proposed in this thesis is used for bird recognition and character recognition. Experimental results shows that the algorithms proposed in this thesis have achieved the latest performance.

Since the SS-MTL algorithm proposed in this thesis and most of the few-shot learning algorithms are aimed at the field of computer vision, and text cannot be directly calculated, these algorithms cannot be directly applied to the field of relation extraction. The author's future efforts are to further improve the few-shot learning algorithm proposed in this article so that it can be applied to both computer vision and text fields. And we will verify the performance of the few-shot relationship extraction algorithm proposed in this thesis on the real long-tailed data set.

Keywords: Few-shot learning, few-shot relation extraction, self-supervision, dilated convolution, matching information

插图索引

图 2.1 残差单元示意图.....	8
图 2.2 Resnet12 结构图	9
图 2.3 Transformer 架构图	9
图 2.4 Transformer Encoder 与 Decoder 示意图	10
图 2.5 BERT 模型架构图	10
图 2.6 MAML 算法框架示意图.....	12
图 2.7 缩放和移位示意图.....	13
图 3.1 SS-MTL 算法模型图(红色框内为预训练阶段添加的自监督分类器)	20
图 3.2 SS-MTL 算法框架示意图	21
图 3.3 传统卷积与膨胀卷积对比示意图.....	22
图 3.4 改进前的 Resnet12 网络残差块示意图.....	23
图 3.5 改进后的残差块(红色框为本文添加的膨胀卷积层).....	23
图 3.6 残差注意力机制示意图(红色框为本文设计的注意力分支).....	24
图 3.7 线性插值示意图.....	25
图 3.8 基于自监督的预训练示意图(红色框内为本文自监督辅助任务).....	27
图 3.9 SS-MTL 算法伪码示意图	29
图 3.10 SS-MTL 算法在一个任务上的学习步骤示意图	30
图 3.11 小样本鸟类识别流程图.....	35
图 3.12 手写字符识别流程图.....	37
图 4.1 IFML-FSRE 算法框架示意图(红色部分为本文所提出的机制).....	40
图 4.2 实例编码模块示意图(红色框为本文提出的策略).....	41
图 4.3 参数初始化模块示意图.....	43
图 4.4 参数快速适应模块示意图(红色部分为动态的学习率计算器).....	44
图 4.5 学习率计算器设计示意图.....	45
图 4.6 元优化模块示意图.....	47
图 4.7 IFML-FSRE 算法伪码示意图.....	48

表格索引

表 2.1	最相关小样本学习算法对比表	14
表 2.2	最新小样本关系抽取算法性能分析及对比表	17
表 3.1	SS-MTL 算法参数表	32
表 3.2	自监督辅助任务实验精度对比表	33
表 3.3	l_2 正则化项对过拟合有效性对比表	33
表 3.4	SS-MTL 算法与对比算法在 miniImageNet 上精度对比表	34
表 3.5	SS-MTL 算法与对比算法在 FC100 上精度对比表	34
表 3.6	SS-MTL 算法在 CUB-200-2011 数据集上性能对比表	36
表 3.7	算法在增强的 Omniglot 数据集上性能对比表	37
表 4.1	IFML-FSRE 算法超参数表	49
表 4.2	IFML-FSRE 算法与其他对比算法在 FewRel 上的精度对比表	50
表 4.3	各个算法在不同噪声比例下的分类精度	50

符号对照表

符号	符号名称
N	样本类别数
K	每类样本数
D_{train}	训练集
D_{test}	测试集
x_Q	查询集 Q 中的实例
y	关系标签
h	头实体
t	尾实体
S	支持集
Q	查询集
X	模型输入
X'	模型输出
W_i	输入句子的每个单词
W_i'	编码后的每个单词
$P(T)$	任务分布
T_i	第 i 个任务
Φ_{S_1}	缩放参数
Φ_{S_2}	移位参数
$T(x)$	残差注意力机制主分支
$M(x)$	残差注意力机制掩码分支
Θ	特征提取器
ϕ_a	类向量表示参数
sim_{ij}	实例 i 与实例 j 的相似度
ϕ_e	编码器参数
ϕ_n	初始化器参数
Φ	元参数
$p(C)$	类分布
θ_{ss}	自监督训练过程中的旋转角度分类器

缩略语对照表

缩略语	英文全称	中文对照
FSL	Few-Shot Learning	小样本学习
FSRE	Few-Shot Relation Extraction	小样本关系抽取
HT	Hard Task	困难任务
IFML-FERE	Information Fused Meta-learning for Few-shot Relation Extraction	信息融合的元学习 小样本关系抽取
ML	Machine Learning	机器学习
MLP	Multi-layer perceptron	多层感知机
MTL	Meta-Transfer Learning	元迁移学习
RE	Relation Extraction	关系抽取
SS	Scaling and Shifting	缩放与移位
SS-MTL	Self-Supervised based MTL	基于自监督的 MTL

目录

摘要	I
ABSTRACT	III
插图索引	V
表格索引	VII
符号对照表	IX
缩略语对照表	XI
第一章 绪论	1
1.1 选题背景与研究意义	1
1.2 国内外研究现状	2
1.2.1 小样本学习算法的研究现状	2
1.2.2 小样本关系抽取算法的研究现状	4
1.3 论文主要工作与组织结构	5
1.3.1 论文主要工作与创新	5
1.3.2 论文组织结构	6
第二章 问题定义与相关工作综述	7
2.1 问题定义	7
2.1.1 小样本学习问题定义	7
2.1.2 小样本关系抽取问题定义	7
2.2 深度神经网络模型	8
2.2.1 残差神经网络	8
2.2.2 BERT	9
2.3 小样本学习算法综述	11
2.3.1 基于元学习的小样本学习算法	11
2.3.2 混合型的小样本学习算法	12
2.4 小样本关系抽取算法综述	15
2.5 本章小结	17
第三章 SS-MTL: 基于自监督的元迁移小样本学习算法	19
3.1 算法设计动机	19
3.2 算法设计思想及框架	19
3.3 算法主要设计策略	22
3.3.1 特征提取器的设计	22

3.3.2	模型损失函数.....	26
3.3.3	基于自监督辅助任务的预训练阶段.....	26
3.3.4	元迁移学习阶段.....	28
3.4	算法实现	29
3.5	算法实验及实际应用	31
3.5.1	实验环境及设置.....	31
3.5.2	实验参数设置.....	32
3.5.3	对比算法介绍.....	32
3.5.4	算法精度比较及分析.....	33
3.5.5	SS-MTL 算法实际应用 1: 鸟类识别.....	35
3.5.6	SS-MTL 算法实际应用 2: 手写字符识别.....	36
3.6	本章小结	38
第四章	IFML-FSRE: 信息融合的元学习小样本关系抽取算法	39
4.1	算法设计动机	39
4.2	算法设计思想及框架	40
4.3	算法主要设计策略	41
4.3.1	实例编码模块的设计.....	41
4.3.2	参数初始化模块的设计.....	43
4.3.3	参数快速适应模块的设计.....	44
4.3.4	元优化模块的设计.....	46
4.4	算法实现	47
4.5	算法实验与分析	48
4.5.1	实验环境及设置.....	48
4.5.2	实验参数设置.....	49
4.5.3	对比算法介绍.....	49
4.5.4	算法性能比较及分析.....	50
4.6	本章小结	51
第五章	总结与展望	53
5.1	论文工作总结	53
5.2	后续工作展望	53
参考文献	55
致谢	61
作者简介	63

第一章 绪论

本章首先对小样本学习算法以及小样本关系抽取算法的选题背景与研究意义进行了介绍。之后，对其国内外研究现状进行了深入分析，最后，简述了本文的主要工作与论文组织结构。

1.1 选题背景与研究意义

近年来深度神经网络已经在监督识别任务上取得了令人振奋的突破，如视觉^[1]，语音^[2]，自然语言处理^[3]等各个不同的领域。然而，深度学习是一种数据饥渴的技术，它取得出色性能往往依赖于大量可用的有标签数据。但是现实世界中，很多领域没有足够的有标签数据，或者获得有标签数据的成本非常高，例如医疗^[4]，安全^[5]等领域。为了从有限的监督信息中学习以达到任务的需求，出现了一个新的深度学习问题，称为小样本学习^[6-7]。在人类快速学习能力的启发下，研究人员希望机器学习模型在学习了一定类别的大量数据后，对于新的类别，只需要少量的样本就能快速学习，这就是小样本学习(Few Shot Learning)。

小样本学习领域一个典型的应用就是字符识别^[8]，它要求计算机程序在给定少量样本的情况下进行分类，解析并生成少量新的手写字符。小样本学习还可以帮助减少数据密集型应用程序的数据收集工作，如图像分类，图像检索，对象跟踪，手势识别，语言建模等^[9]。小样本学习既有真正的学术研究价值，又有廉价学习的工程需求，因此备受关注成为热门话题。近年来，涌现出许多优秀的算法试图解决这个问题，但目前大部分算法仍然存在精度低，过拟合等问题，且这些算法所表现出的精度相对于传统的深度学习算法有待于进一步的提高。

关系抽取是自然语言处理中一个非常重要的任务，旨在根据两个给定实体的相关上下文对它们之间的语义关系进行分类^[10]。由于能够提取文本信息，并使得许多自然语言处理任务受益，如信息检索，对话生成和智能问答等，近年来关系抽取获得了极大关注。传统的有监督模型的效果在很大程度上取决于训练数据的规模和质量，这意味着这些有监督的模型通常遭受数据稀缺的困扰，因此很难很好地针对语义信息进行概括^[11-13]。小样本关系抽取(Few Shot Relationship Extraction)任务的目标即是在测试期间，在只给定很少的训练实例的情况下，对给定的测试实例进行分类。

然而，当前的小样本学习模型主要关注低噪声的视觉任务，由于文本是多种多样的，且不能直接计算，所以，传统的小样本学习算法不能直接应用于关系抽取任务。且由于语义信息的多样性和复杂性，小样本学习更容易受到数据噪声的影响，这对于小样本关系抽取问题的研究产生了极大的阻碍。

本文选题来源于国家自然科学基金面上项目,小样本学习是深度学习以及关系抽取领域又一新的挑战,也是近几年深度学习领域的热门研究课题。诚然,众多专家学者已经提出了效果较好的算法,但是仍然留给后人很大的改进空间。基于深度学习的小样本学习算法以及小样本关系抽取算法的研究具有重要的理论意义与工程价值。

1.2 国内外研究现状

本文针对小样本学习算法以及小样本关系抽取算法进行研究,因此本文分两小节分别对小样本学习算法的研究现状以及小样本关系抽取的研究现状进行详细阐述。

1.2.1 小样本学习算法的研究现状

近几年小样本学习算法一直是深度学习领域的热门话题,涌现出大量优秀的算法。目前小样本学习算法主要分为以下几类:

(1) 基于数据增强的方法。数据增强旨在增加训练数据的个数,以提高深度模型的泛化性能,数据增强已被证明有利于深层体系结构中机器学习模型的训练。在以图像分类为主要任务的小样本学习算法中,许多工作通过使用手工规则对训练数据集中的原始示例进行变换,作为数据预处理过程来增强训练数据集,例如平移^[14]、翻转^[15]、错切^[16]、缩放^[17]、反射^[18],裁剪^[19]以及旋转^[20]等。这种基于手工规则的数据增强方法广泛应用于深度模型中,以降低过拟合的风险。然而,这些规则的设计严重依赖于领域知识并且需要昂贵的劳动力成本,而且它们是特定于数据集的,很难应用到其他数据集,不具有普适性。因此,这些方法只能缓解而不能真正解决小样本问题,并且仅用作图像数据的预处理步骤。

目前最新的方法是**通过生成网络来学习一个数据生成器**,例如**生成对抗网络**(Generative Adversarial Net, GAN)^[21], Goodfellow 等人提出 GAN 方法,生成器通过随机噪声构造假的数据,判别器对传输过来的数据真假性进行判别,通过生成网络和判别网络的动态“博弈”过程来生成新数据。文献[22]提出的数据增强生成对抗网络(Data Augmentation-GAN, DAGAN)为特定任务生成少量假的例子,然后将增强数据添加到原始示例中。在 Meta-GAN^[23]中进一步为特定任务生成虚假示例以补充训练示例,并将**元学习**(Meta Learning)与 GAN 结合以解决小样本问题。Delta^[24]提出使用编码器-解码器架构来增强特征空间中的数据,其中编码器学习提取相同的实例对之间的可转换变形,而解码器学习如何将这些变形应用于来自新类别的其他实例。通常,从增强训练数据的角度来解决小样本学习问题是非常直观且易于理解的,但是真实数据的分布是未知的,因此不可能有完美的先验知识。这意味着先验知识指导的增强过程不精确。估计的数据分布与真实数据分布之间的差距会极大地干扰数据质量,甚至会导致概念

漂移。

(2) **基于迁移学习的小样本学习方法**。迁移学习即从一个或多个源领域任务中提取有用的知识并将其用在新的目标任务上。文献[25]显示，基于迁移学习的方法与基于元学习的方法相比，显示出具有竞争力的优势。这类方法旨在智能地应用以前学到的知识来更快地解决新问题。迁移什么和如何迁移是迁移学习中要解决的关键问题。对于深层模型，一种强大的迁移方法是将预训练的模型用于新任务，通常称为微调(Fine-tune, FT)。通过微调预训练网络权重是主要的迁移学习方法之一，但是对模型所有参数进行微调，学习的参数过多，容易导致过拟合。为了解决这个问题，文献[26]通过冻结预训练模型的参数，并给冻结后的预训练参数附加相应的缩放参数来达到适应新任务的目的，不仅大量减少了需要优化的参数的数量，而且很好地避免了灾难性遗忘问题。另一种流行的迁移方法是将预训练的网络作为骨干网络，并添加高级功能，例如用于物体检测和识别^[27-29]以及图像分割^[30-31]等。基于迁移学习的方法可以显著减少模型对训练数据量的要求，但围绕这些参数进行的繁杂的调试过程，就足够让人望而生畏，其次如何选择合适的预训练模型以及预训练模型难以更新是迁移学习难以推广的重要原因。

(3) **基于元学习的方法**。元学习，亦称为学会学习(Learning to Learn)，即利用以往的经验知识来指导新任务的学习，具有学会学习的能力。基于元学习的小样本学习算法可分为两类：分别为基于度量(Metric-Based)的方法和基于优化(Optimization-Based)的方法。基于度量的方法主要侧重于学习一种良好的度量以测量支持集数据和查询集数据之间的距离或相似性。例如，原型网络^[32](Prototype Network, PN)将分类问题转化为嵌入空间中的最近邻问题，它认为每个类别都存在一个原型表示，该类的原型是支持集在嵌入空间中的均值。而关系网络^[33](Relation Network, RN)实现了一个网络，用于测量支持实例和查询实例之间的关系相似性。基于度量学习的方法是学习一个嵌入空间，很容易导致过拟合问题。基于优化的方法旨在设计一种在元训练阶段到元测试阶段适应信息的优化算法。MetaLSTM^[34]将问题表述为基于 LSTM 的元学习算法，以在多次学习中更新优化算法。模型无关的元学习算法(Model Agnostic Meta Learning, MAML)^[35]是最为经典的元学习算法，它学习了一种优化方法，该方法可以沿着快速梯度方向来快速学习新颖类的分类器，通过学习模型参数的更新函数或规则，在少量迭代步骤的情况下就能获得很好的泛化性能。MAML 为后来许多工作提供了指导，如 MAML++^[36]针对 MAML 所存在的缺陷进行了一系列改进，增强了 MAML 算法的稳定性。TADAM(Task Dependent Adaptive Metric, TADAM)^[37]算法提出了任务相关的自适应度量方法，并提出基于辅助任务协同训练的优化过程。基于元学习的算法模拟人脑学习的过程，但是缺点是选择性遗忘，在学习新任务后容易忘记学习到的关于旧任务的知识。

(4) **混合型算法**。单纯的基于元学习和基于迁移学习的算法具有局限性，目前大部分最新的小样本学习算法都是将不同的相关技术添加到元学习或迁移学习的思想框架中对算法性能进行提升。如文献[38]提出一种元学习和迁移学习相结合的算法 MTL(**M**eta-**T**ransfer **L**earning, MTL)，Yu 等人^[39]提出一种新的基于半监督的小样本迁移学习框架 TransMatch，T-NAS(**T**ransferable **N**eural **A**rchitecture **S**earch, T-NAS)^[40]算法将元学习与神经体系结构搜索相结合，Mangla^[41]等人通过将自监督与流形混合的方法引入小样本学习算法中。混合型算法能够结合不同技术的优点而显著提升算法的性能，如何将不同技术应用于小样本学习是此类算法的难点，且不同的技术仍然存在其局限性与缺陷，充分利用不同技术的优点并且最大程度地降低其缺陷是此类算法未来的研究方向。

1.2.2 小样本关系抽取算法的研究现状

关系抽取旨在识别句子中实体之间的关系，由于神经网络强大的特征提取能力，深度学习在关系抽取方面已取得了显著成功。自从文献[42]第一次成功运用卷积神经网络在关系提取中提取词汇和句子的特征，之后一些基于神经网络^[43-45]的有监督方法在关系抽取任务中取得了最新的性能。尽管与传统方法相比，这些工作取得了显著改善，但它们仍然缺乏有标签的训练数据。为了获得大量的有标签数据，文献[46]提出了远程监督方法。但是远程监督生成的句子并不都代表相同的关系，由此带来了错误标签问题。为了解决这个问题，[47-50]等文献采用多实例学习来减轻错误实例的影响，在远程监督的基础上添加句子级别的注意力机制，动态地降低错误实例的权重。总的来说，关系抽取的主要挑战是仅依靠少量的监督数据来训练模型，然而小样本学习算法大多集中在计算机视觉上，在自然语言处理领域，特别是关系抽取领域缺乏系统的研究，且由于语义信息的多样性和复杂性，传统的小样本学习算法很难运用到关系抽取领域，因此，针对小样本关系抽取算法的研究具有深远的意义。

由于原型网络的简单性，目前最新的关于小样本关系抽取的工作大部分都采用了基于原型网络的方法，如 Yang 等人^[51]设计了一个协作注意力模块，以分别提取句子和实体的有益和指导信息并通过门机制来动态融合这两种信息，从而获得知识感知的实例。文献[52]提出一种混合注意力的原型网络(**H**ybrid **A**ttention-Based Prototypical Networks, HATT-Proto)，通过实例级别和特征级别的注意力机制，凸显关键实例和特征，从而在有噪声的小样本学习场景中显著增强了关系抽取模型的性能和鲁棒性。文献[53]提出了一种新型的多原型嵌入网络，该网络将文本表示学习和知识约束联系在一起。此外，作者还提出了一种原型感知正则化方法，以了解更多具有代表性的原型，使得关系抽取算法性能得到进一步的提升。虽然这些算法取得了不错的性能，但是当前常规的有监督和远程监督的关系抽取模型始终是基于封闭世界的假设，它忽略了开

放环境中新关系的出现。

为了识别出现的新关系, Ren 等人^[54]提出了一个具有原型注意力对准和三重态损失的两阶段原型网络 BERT-IncreProtoNet, 以动态地识别具有少量支持实例的新关系。但是算法 BERT-IncreProtoNet 的精度很低。所有这些基于原型网络的算法都是独立计算查询实例的嵌入向量和每个支持集的原型向量, Ye 等人^[55]提出的 MLMAN(Multi-Level Matching and Aggregation Network, MLMAN)模型通过考虑实例级别的匹配信息以交互方式对查询实例和每个支持集进行编码。每个支持集的最终类原型是通过对其支持实例的表示进行注意力汇总而获得的, 该方法实现了最新的性能。

除了基于原型网络的方法, 近两年也涌现出其他一些优秀的算法, 如文献[56]构建了一种基于小样本关系抽取的异构图神经网络 HGNN, 该网络通过图中实体节点和句子节点之间传递的信息捕获图的丰富邻域信息。为了解决实例不可靠和可解释性差的问题, 文献[57]提出了一种新颖的元信息引导元学习(Meta-Information Guided Meta-Learning, MIML)框架, 通过类名的语义信息进行模型参数的快速初始化以及快速的适应, 在基于实例的信息和基于语义的信息之间建立连接, 从而实现更有效的初始化和更快的适应性。

虽然目前涌现出许多优秀的算法, 但是大部分算法主要关注于实例的重表示和算法精度的提升, 算法对于噪声实例的鲁棒性有待于进一步提高。基于匹配信息的实例重表示方法是一种新的趋势, 所以本文针对这些问题提出一种新的小样本关系抽取算法。

1.3 论文主要工作与组织结构

1.3.1 论文主要工作与创新

为了解决小样本学习算法以及小样本关系抽取算法目前所存在的问题, 本文对小样本学习算法以及小样本算法在关系抽取领域的应用进行了深入研究。本文主要工作与创新点如下:

(1) 在小样本学习算法的研究中, 针对目前深度网络模型不能提取不同尺寸特征以及缺乏对重要特征关注度的缺陷, 本文提出了一种添加膨胀卷积以及残差注意力机制的网络模型, 通过膨胀卷积扩大了模型视野, 残差注意力机制捕获了重要特征; 针对目前小样本学习算法不能解决过拟合的缺陷, 本文提出一种加入 l_2 正则化项的损失函数, 并添加自监督辅助任务对模型进行预训练, 基于上述机制, 本文提出了一种新的高效的基于自监督的元迁移小样本学习算法 SS-MTL(Self-Supervised based Meta Transfer Learning, SS-MTL)。

(2) 由于本文提出的 SS-MTL 小样本学习算法模型过于复杂, 效率较低, 为了保

证将小样本学习算法应用到关系抽取任务中的效率与准确性, 本文将简单的元学习算法 MAML 应用到关系抽取任务中。针对小样本学习算法无法有效利用文本数据的语义信息的问题, 本文提出一种基于匹配度的实例重表示方法, 突出了对测试样本有益的训练样本, 并添加任务过滤策略, 有效提高了算法精度; 针对 MAML 共享学习率导致参数适应效率低的问题, 本文提出一种类名与实例信息融合的模式参数快速适应方法, 为每个实例计算不同的学习率, 从而更好地对模型参数进行快速适应。基于以上思想, 本文提出了一种新的高效的信息融合的元学习小样本关系抽取算法 IFML-FSRE (**I**nformation **F**used **M**eta-**l**earning for **F**ew-**s**hot **R**elation **E**xtraction, IFML-FSRE)。

(3) 作者基于权威的公开数据集, 对本文所提出的两个算法与最新相关算法进行了性能对比, 并对本文提出的小样本学习算法 SS-MTL 进行了鸟类识别与手写字符识别的应用研究, 实验结果表明: 本文提出的算法取得了较高的精度。

1.3.2 论文组织结构

本文的组织结构如下所示:

第一章 绪论。本章论述了小样本学习算法以及关系抽取小样本学习算法的选题背景及意义, 然后分别对两个课题国内外相关工作进行了详细论述分析, 最后对本文的主要工作及创新点进行总结, 并阐述了本文的组织结构。

第二章 基础理论与相关工作综述。本章首先介绍了小样本学习算法与小样本关系抽取相关概念以及本文所涉及的两个深度神经网络模型; 然后对最新的与本文最相关的小样本学习算法以及小样本关系抽取算法进行分析, 并详细阐述各个算法的优点与不足。

第三章 SS-MTL 基于自监督的元迁移小样本学习算法。本章首先介绍了目前最新的小样本学习算法所存在的缺陷与不足, 然后详细介绍了本文针对这些缺陷所提出的改进策略及具体实现步骤, 最后通过实验对本章提出的算法与目前最新的算法进行比较分析, 并将本文提出的算法应用到鸟类识别与手写字符识别领域。

第四章 IFML-FSRE 信息融合的元学习小样本关系抽取算法。本章首先介绍了现有小样本学习算法应用到关系抽取领域的难点和缺陷, 然后详细介绍了本文针对这些缺陷所提出的改进策略及具体实现步骤, 最后通过实验对本章所提出的算法与目前最新的其他算法进行比较, 并对实验结果进行分析。

第五章 总结与展望。对本文主要工作进行总结, 对本文工作的贡献及优缺点进行了深入的分析, 以及本文对小样本学习算法以及小样本关系抽取接下来的工作进行展望。

第二章 问题定义与相关工作综述

本章首先介绍基于小样本学习算法及小样本关系抽取算法所涉及的相关概念及基础理论。之后，对本文涉及的两个重要的深度神经网络模型进行简介，最后，对与本文最相关的小样本学习算法以及小样本关系抽取算法进行了分析综述。

2.1 问题定义

2.1.1 小样本学习问题定义

由于小样本学习是机器学习的一个子领域，因此在给出小样本学习的定义之前，先给出机器学习的定义，本文采用 Mitchell^[58]的定义，如定义 2.1 所示。

定义 2.1 机器学习 (Machine Learning, ML): 如果计算机程序的性能可以通过性能测度 P (Performance measure) 测得的任务 T (Task) 上的经验 E (Experience) 来提高，则可以说计算机程序可以通过从与任务 T 和性能测度 P 相关的经验 E 中学习。

如定义 2.1 所示，机器学习问题由 E , T 和 P 指定。例如， T 为图像分类任务，机器学习程序可以通过训练大规模有标签图像获得的经验 E 来提高分类精度 P 。小样本学习是机器学习的一种特殊情况，它的确切目标是在数据集提供的有限监督信息下获得良好的学习性能。小样本学习的正式定义在定义 2.2 中给出。

定义 2.2 小样本学习 (Few-Shot Learning, FSL)^[9]: 小样本学习是一类由 E , T 和 P 指定的机器学习问题，其中 E 包含非常少的针对目标任务 T 的监督信息。

小样本分类算法是典型的小样本学习算法，旨在学习一种仅需少数几个可用训练实例即可处理分类任务的模型。具体而言，假设 D 是整个有标签数据集，通常根据类别将其分为两部分：训练集 D_{train} 和测试集 D_{test} 。换言之，这两个数据集具有不同的标签空间，并且彼此不相交。通常，当前大多数 FSL 算法都采用“episode”训练策略。为此， D_{train} 和 D_{test} 分别进一步分为两部分：支持集 $D_{train-support}$ 和 $D_{test-support}$ 以及查询集 $D_{train-query}$ 和 $D_{test-query}$ 。具体而言，在每个 episode 中，首先从 D_{train} 中随机抽取 N 个类别，然后对这 N 个类别中的每个类别的 K 个实例进行采样构成 $D_{train-support}$ 对模型进行训练，在 FSL 任务中，通常称其为 **N-way-K-shot** 设置；然后从这 N 个类别的其余样本中随机抽取 R 个实例构造 $D_{train-query}$ 对模型修正。同样，也可以以类似的方式获取 $D_{test-support}$ 使模型去适应新的类别，使用 $D_{test-query}$ 对模型性能进行测试。

2.1.2 小样本关系抽取问题定义

关系抽取任务是信息提取的基本任务之一，旨在检测非结构化的自然语言文本中

的实体对之间的关系信息，并以实体关系三元组的形式生成结构化数据。关系抽取的形式化定义如下所示：

定义 2.3 关系抽取 (Relation Extraction, RE)^[56]：一个关系抽取任务可以表示为 $T:(x, h, t) \rightarrow y$ ，即给定一个句子 x ，以及该句中的一对实体 (h, t) ，得出两个实体之间的语义关系 y ，其中 h 表示头实体(head)， t 表示尾实体(tail)。

例如有一句话：“<e1> Beijing</ e1>is the capital of<e2> China</ e2>.”，它通过一对 XML 标签标记实体“Beijing”和“China”。从该句子中，RE 模型可以输出一个三元组表示(Beijing, China, capital_of)，该表示表明实体“Beijing”和“China”之间的关系为“capital_of”。

定义 2.4 小样本关系抽取 (Few-Shot Relation Extraction, FSRE)^[56]：由 RE 的定义，可以将 FSRE 定义为给定关系集 $R = \{r_i, i = 1, \dots, N\}$ 和支持集 S ，预测查询集 Q 中的实例 x_q 的关系 y 的任务，其中类的数量 N 和实例的数量 K 通常是很小的。FSRE 必须从支持集中很少的几个实例中训练模型，然后再使用训练好的模型以预测任何给定查询实例的关系，此外，数据集集中的每个实例都包含一个实体对 (h, t) 。

2.2 深度神经网络模型

本文基于最新的小样本学习算法 MTL 以及最新的小样本关系抽取算法 MIML 进行了改进，其中 MTL 算法采用 Resnet12 网络模型作为特征提取器，MIML 算法采用 BERT 模型作为句子编码器，因此，本节主要对这两个模型进行介绍。

2.2.1 残差神经网络

Resnet (Residual Neural Network)残差神经网络模型自 2015 年在 ImageNet 图像分类比赛中获得第一名之后，影响了之后深度学习的发展方向，很多工作都建立在其之上。Resnet 网络以“简单与实用”著称，它采用了给卷积层添加直连边的方式，将前一个网络层的中间特征直接传输到后一层，克服了传统深度网络模型中随着模型层数加深而出现的梯度消失以及梯度爆炸的问题，Resnet 结构如图 2.2 所示：

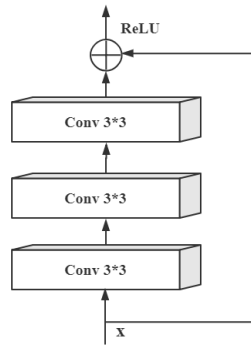


图2.1 残差单元示意图

图 2.1 所示为 Resnet 一个典型的残差单元，也叫残差块，其中， X 表示输入，左边是三个非线性的卷积层，右边是一个直连边。Resnet12 则是由 4 个三层的残差块构成，如图 2.2 所示：

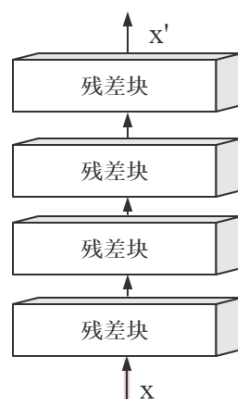


图2.2 Resnet12 结构图

其中 X 表示输入， X' 表示模型的输出，MTL 算法采用 Resnet12 网络模型作为特征提取器，尽管 Resnet12 模型已经取得了非常不错的效果，本文基于 Resnet12 网络模型针对小样本学习算法进行了进一步的改进，取得了更好的效果。

2.2.2 BERT

BERT^[59](Bidirectional Encoder Representation from Transformers)是谷歌公司于 2018 年推出的一个非常强大的自然语言处理框架，在小样本关系抽取算法中应用非常广泛。它是利用 Transformer 构造的一个多层双向编码器网络，首先对 Transformer 进行介绍，其网络结构如图 2.3 所示。

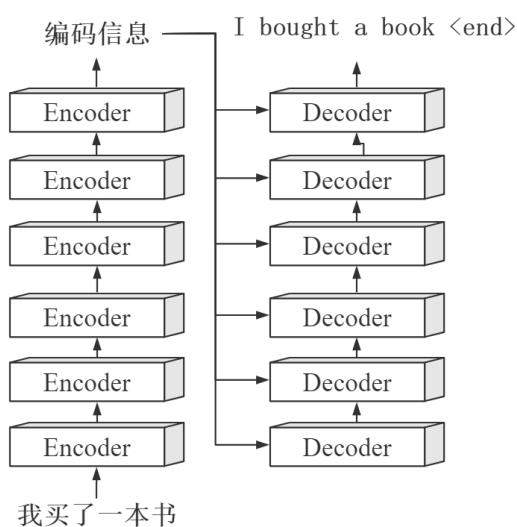


图2.3 Transformer 架构图

由图 2.3 可知, Transformer 由 6 个堆叠的编码器(Encoder)与 6 个堆叠的解码器(Decoder)这两个部分组成。Transformer 最早是被设计用于机器翻译的一种模型, Transformer 首先通过 Word2Vec 等词嵌入方法获得输入句子的每一个单词的嵌入位置信息的向量表示,然后将得到的句子向量表示矩阵输入整个网络,通过 6 个 Encoder Block 之后可以得到句子中所有单词的编码信息矩阵;将 Encoder 输出的编码信息矩阵传到 Decoder 中,每一个 Decoder 会依次根据当前翻译过的单词与编码信息矩阵翻译下一个单词,从而得到整个句子的翻译。其中 Encoder 与 Decoder 的具体结构如图 2.4 所示:

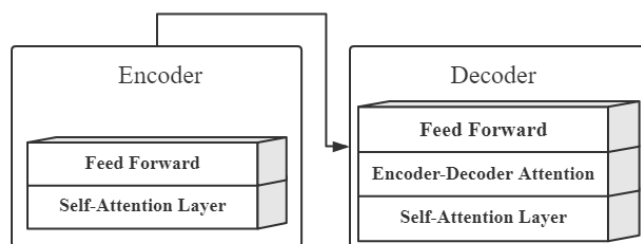


图2.4 Transformer Encoder 与 Decoder 示意图

在图 2.4 中, Encoder 将得到的句子编码信息输入 Decoder 进行解码。每个 Encoder block 包括一个 Self-Attention Layer 层和一个 Feed Forward Neural Network 层, Self-Attention 能帮助当前节点不仅仅关注当前的词,从而能获取到上下文的语义, Self-Attention 处理完数据后把数据送给 Feed Forward Neural Network, 每一个 Encoder 得到的输出输入到下一个 Encoder, 最后一个 Encoder 的输出输入到 Decoder 层中。每个 Decoder 也具有这两层,然后在这两层中间插入一个 Encoder-Decoder Attention 层,帮助当前节点获取到当前需要关注的重点。

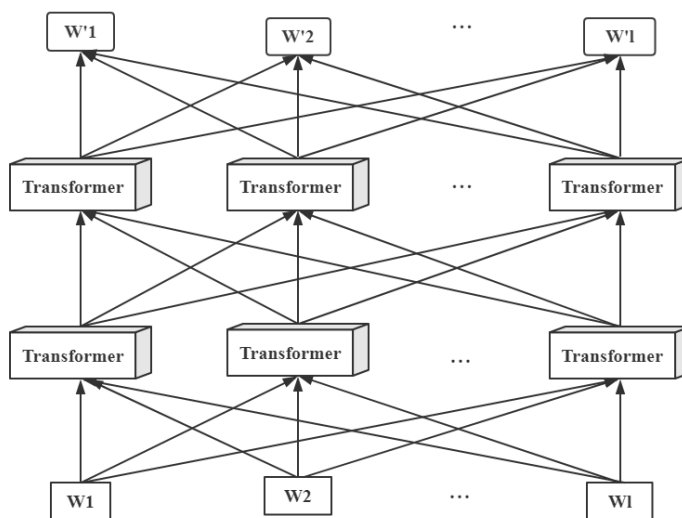


图2.5 BERT 模型架构图

BERT 模型架构如图 2.5 所示，其中 $W_i, i=1, \dots, l$ 表示输入句子的每个单词， $W'_i, i=1, \dots, l$ 表示编码后的每个单词， l 表示一个句子中单词的数量。由图可见，BERT 基于 Transformer 进行多层双向解码，即同时利用每个词前面和后面的词的信息，在使用 BERT 模型时，只需要根据具体任务在其之后添加一层输出层，不需要执行大量任务，也不需要特定的体系结构的修改，正是因为 BERT 如此强大的功能才使其广泛应用于自然语言处理领域。

2.3 小样本学习算法综述

当前小样本学习算法主要有基于数据增强，基于元学习，基于迁移学习的方法以及混合型方法。基于数据增强^[14-20]的小样本学习算法旨在增大训练数据量以缓解过拟合，这类方法只能缓解而不能解决小样本问题，并且增强数据的质量难以保证^[21-24]；而基于迁移学习^[26-31]的算法利用从源域任务中提取的有用知识来解决目标任务，这类方法难以克服复杂的调试过程，且预训练模型的选择和更新的困难是迁移学习难以推广的重要阻碍。基于元学习的方法^[32,33,35-37]利用以往的经验知识来指导新任务的学习，混合型方法^[38-41]将不同的相关技术添加到元学习或迁移学习的思想框架中对算法性能进行提升，这两种算法是目前主流的小样本学习算法，并取得了较好的性能，本节对具有代表性的基于元学习的以及混合型小样本学习算法进行具体阐述。

2.3.1 基于元学习的小样本学习算法

基于元学习的小样本学习算法主要分为基于度量的方法和基于优化的方法。基于度量的方法将输入数据映射到嵌入空间，然后通过嵌入空间的相似性度量来区分不同的类。Snell^[32]等人提出的原型网络^[32] (**Prototype Network, PN**)将分类问题转化为嵌入空间中的最近邻问题，Sung^[33]等人提出的关系网络(**Relation Network, RN**)是一种典型基于度量的方法，该方法学会学习一种在少量样本下的深度距离度量标准，通过训练可以计算测试数据与每个新类别的几个实例之间的关系得分对新类别的数据进行分类，而无需进一步更新网络。但是基于度量的元学习方法学习一个嵌入空间，很容易导致过拟合问题。

Finn^[35]等人提出的模型无关的元学习算法 MAML 是非常经典的基于优化的元学习算法。完整的 MAML 算法框架如图 2.6 所示，MAML 可以考虑为一个由参数 θ 参数化的函数 f_θ 表示的模型，其中，任务分布为 $P(T)$ ，当适应新任务 T_i 时，模型的参数 θ 使用任务 T_i 上的一个或多个梯度下降来计算更新的参数矢量 θ'_i ，通过优化从 $P(T)$ 采样的任务相对于 θ 的 $f_{\theta'_i}$ 的性能来训练模型参数。

MAML 算法框架

- 步骤 1: 随机初始化模型参数 θ ;
 步骤 2: 从任务分布中采样一个批次的任务;
 步骤 3: 计算一个任务中每个实例的梯度;
 步骤 4: 根据梯度计算相对于每个任务更新后的参数 θ'_i ;
 步骤 5: 重复步骤 3-4, 直到遍历完所有实例;
 步骤 6: 通过在每个任务上的更新后的模型性能来更新模型参数 θ ;
 步骤 7: 重复步骤 3-6, 直到遍历完所有任务。

图2.6 MAML 算法框架示意图

MAML 是非常重要的简单通用的元学习技术, 该技术可以应用于任何问题和模型, 但是 MAML 算法需要大量的类似的任务来进行元训练, 这是代价非常高的; 反向传播导致 MAML 算法训练过程存在梯度不稳定的问题; 由于每个任务通常是由浅层神经网络进行建模, 如果使用更深, 更强大的体系结构就会出现过拟合的问题。Antoniou^[36]等人提出的 MAML++算法是基于 MAML 改进的一种元学习框架, 它针对 MAML 所存在的缺陷提出了一系列改进策略。MAML++算法改善了 MAML 的泛化误差, 稳定和加快了 MAML 算法的训练过程, 但是 MAML++仍然需要大量的任务进行训练且不适用于深度神经网络, 容易出现过拟合问题。

为了克服元学习算法的过拟合问题, Oreshkin^[38]等人提出一种基于度量的元学习算法 TADAM, TADAM 算法通过辅助任务协同训练以根据特定任务调整特征提取器, 使得度量空间与任务相关, 在这种情况下的特征提取器具有良好的泛化能力。但是, 任务相关的度量空间很容易受到噪声的影响; 仅对相似性度量进行缩放会导致完全不同的参数更新机制; 学习任务相关的度量空间比学习静态空间更具有挑战性。

基于元学习的小样本学习算法模拟人脑学习的过程, 旨在学习模型参数的更新函数或规则, 是非常简单易于实施的小样本学习算法, 但是缺点是选择性遗忘, 在学习新任务后容易忘记学习到的关于旧任务的知识。

2.3.2 混合型的小样本学习算法

TransMatch^[39]是 Yu 等人于 2020 年提出的一种新的混合型小样本学习算法, 该算法将迁移学习与半监督学习相结合, 以充分利用来自有标签的基类数据和无标签的新颖类数据的辅助信息。该框架包含三个组件: (1)对基于基类数据的特征提取器进行预训练; (2)使用特征提取器来初始化新颖类的分类器权重, 具体而言, 它通过权重压印的方法, 利用 N-way-K-shot 设置下的实例的特征向量均值直接设置分类器权重, 其中特征向量是通过模型从预训练阶段获得的; (3)进一步采用半监督学习方法对模型进行更新。TransMatch 将迁移学习与半监督学习相结合, 有效利用了未标记的数据, 尽管 TransMatch 取得了最新的性能, 但是 TransMatch 采取微调的策略进行权重的迁

移，微调策略所需要调试的参数量过大，容易产生过拟合；在典型的半监督学习中，标记图像的数量仍大于小样本学习，半监督方法的技术可能无法直接用于小样本学习，直接将半监督技术应用于小样本学习会导致精度降低。

相比于 TransMatch 微调策略所产生的大量需要更新的参数，2019 年，Sun^[38]等人提出一种元迁移小样本学习算法 MTL，该算法将迁移学习与元学习相结合，首先通过大规模数据上的预训练获得良好的初始权重，然后将其“冻结”(如图 2.7 黄色部分)，针对小样本学习场景利用元学习的两个轻量级神经元操作缩放与移位(Scaling and Shifting, SS)：给每个“冻结”的卷积核添加两个缩放与移位参数 Φ_{s_1} 和 Φ_{s_2} (如图 2.7 绿色部分)，去适应新的小样本学习任务。SS 操作不仅大大降低了需要优化的参数的数量，在一定程度上减少了过拟合的风险，而且作者提出一种困难任务(Hard Task, HT)再训练的方式，从训练精度较低的类中随机采样进行再训练，进一步提高了算法的性能。

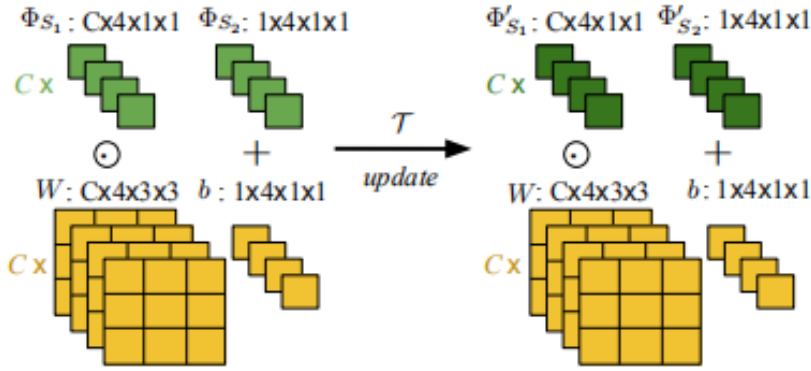


图2.7 缩放和移位示意图

如图 2.7 所示，针对 FSL 场景，模型不再需要更新所有参数(如图中黄色部分)而只需要更新添加的两个缩放与移位参数 Φ_{s_1} 和 Φ_{s_2} 。其中，SS 操作如公式(2-1)所示：

$$SS(X; W, b; \Phi_{s_{1,2}}) = (W \otimes \Phi_{s_1})X + (b + \Phi_{s_2}) \quad (2-1)$$

其中， X 表示输入， \otimes 表示逐元素相乘， Φ_{s_1} 和 Φ_{s_2} 分别表示缩放和移位参数。MTL 算法相比于 MAML 不再需要高代价的大量的类似的类进行元训练，MTL 虽然减少了需要更新的参数量，但是对于小样本问题，需要更新的参数量还是太多，仍然不能完全解决过拟合问题；简单的大规模预训练缺乏对数据本身结构信息的挖掘，仍然依赖于大量类似任务的训练。

受 MAML 算法的启发，Lian^[40]等人提出了 T-NAS 算法，针对 FSL 场景，T-NAS

创新性地将元学习的思想与神经体系结构搜索相结合,提出了一种新颖的可迁移神经体系结构搜索方法。T-NAS 学习了一种元体系结构,该体系结构可以通过几个梯度步骤快速适应新任务,使得迁移的体系结构能够适合于特定任务。T-NAS 算法在 FSL 场景下显示出与其他基准算法相当的性能,其参数量却大大低于最新的小样本学习算法。但是相对于目前最新的小样本学习算法, T-NAS 算法的精度不够高。

Mangla^[41]等人提出的 S2M2 算法将自监督与流形混合的方法引入小样本学习算法中,显著提高了小样本学习算法的性能。具体地来说, S2M2 算法在训练过程中利用旋转等自监督任务添加自监督的损失,可以实现强大的语义特征学习,从而显著提高小样本分类算法的精度;流形混合的正则化技术增强了模型对数据分布微小变化的鲁棒性,显著提高了小样本学习任务的性能。

混合型算法能够结合不同技术的优点而显著提升算法的性能,如何将不同技术应用于小样本学习是此类算法的难点,且不同的技术仍然存在其局限性与缺陷,充分利用不同技术的优点并且最大程度地降低其缺陷是此类算法未来的研究方向。

本文针对目前最新的关于小样本学习算法在算法类型,过拟合解决能力,是否引入辅助任务,是否提取不同尺寸特征这四个方面进行性能比较,如表 2.1 所示。

表2.1 最相关小样本学习算法性能对比表

算法名称	算法类型	过拟合解决能力	引入辅助任务	提取不同尺寸特征
RN ^[34] (2018)	元学习	一般	否	否
MAML ^[35] (2017)	元学习	一般	否	否
MAML++ ^[36] (2019)	元学习	一般	否	否
TADAM ^[37] (2018)	元学习	较好	是	否
MTL ^[38] (2019)	混合型	较好	否	否
TransMatch ^[39] (2020)	混合型	一般	是	否
T-NAS ^[40] (2020)	混合型	一般	否	否
S2M2 ^[41] (2020)	混合型	较好	是	否

表 2.1 展示了目前具有代表性的基于元学习和混合型的小样本学习算法, MAML^[35]是非常经典的元学习算法,很多后来的工作都是基于 MAML 算法的思想。MAML++^[36]针对 MAML 算法的缺点进行进一步改进,但仍然没有解决过拟合等问题。过拟合是小样本学习算法目前面临的严重挑战, S2M2^[41]和 TADAM^[37]通过引入辅助任务,对模型过拟合问题有了一定的改进, MTL^[38]将迁移学习与元学习相结合,是一个非常优秀的算法,但是在提取不同尺寸特征及添加辅助任务等方面仍然存在待改进的空间。从上述算法中可以发现,没有算法能够提取不同尺寸的特征,因此有效利用

辅助任务，通过提取不同尺寸特征提升算法精度，解决小样本学习算法过拟合问题，是小样本学习算法的下一步研究目标。

2.4 小样本关系抽取算法综述

语义信息的复杂性和多样性使得传统的小样本学习算法难以直接应用于关系抽取领域。由于原型网络的简单性，目前大部分小样本学习算法^[51,52,54,55]都是基于原型网络的。基于原型网络的方法在嵌入空间通过计算实例间距离对句子关系进行分类。针对原型网络对于噪声的脆弱性，算法^[52]通过注意力机制降低噪声实例的权重，然而，基于原型网络的方法难以避免地产生原型偏差的问题，因此一些学者提出了新的方法以避免陷入原型偏差，如文献^[56]提出了一种异构图神经网络模型来解决小样本关系抽取问题，一种基于 MAML^[35]框架所提出的小样本关系抽取算法 MIML 取得了最新的性能^[57]，本节对这些具有代表性的小样本关系抽取算法进行详细阐述。

Yang 等人^[51]提出了一个带有关系和实体的文本描述的增强型原型网络 TD-Proto，具体地，该原型网络设计了一个协作注意模块，该模块利用关系描述来突出句子的上下文特征信息，并提取描述实体关系的特定信息。为了充分利用外部信息，Yang 等人采用了门机制将这两种表示形式进行动态融合。TD-Proto 引入外部信息以避免传统原型网络做出不可靠预测，但是这种方法没有考虑到噪声实例的影响，且难以避免地出现原型偏差的问题。

为了尽可能降低噪声实例的影响，Gao 等人^[52]提出一种基于混合注意力的原型网络 HATT-Proto，该混合注意力机制包括两个模块：实例级别的注意力模块更加突出那些与查询实例相关的支持实例，特征级别的注意力模块旨在缓解特征的稀疏性问题。原始的原型网络将实例的平均向量作为该类的原型，由于 FSL 场景中缺乏支持数据，当含有噪声实例或一个实例包含多种语义时将导致巨大的原型偏差，因此 HATT-Proto 模型通过实例级别的注意力机制在支持集中选择更多的信息实例，并在训练过程中对实例进行去噪。距离函数的选择会显著影响原型网络的性能，而简单的欧几里得距离会遭受特征稀疏性的困扰，因此，对于一些特殊关系在特征空间的分类，某些维度会更具有区分性，针对不同的关系制定特定的距离函数，缓解了特征稀疏性的困扰。但是，HATT-Proto 算法仍然存在以下不足：只能降低噪声实例的比重从而降低噪声实例的影响，不能有效规避可能存在的噪声实例；相比于目前最新的成果，算法精度不够高。

虽然原型网络在深度有监督学习和小样本学习两个领域均具有更好的性能，但是，当新关系出现时，它仍然遭受不兼容的特征嵌入问题。为了解决这个问题，Ren 等人^[54]提出了一种增量式的小样本关系抽取原型网络 BERT-IncreProtoNet，该网络分为两

个阶段,第一阶段以监督学习的方式学习基本关系的特征嵌入空间,在第二阶段中提出具有新颖原型生成器的增量原型网络,通过比较查询实例与每个原型之间的距离进行分类。尽管 BERT-IncreProtoNet 可以有效地识别具有少量支持实例的新型关系,但是它取得的性能相对于目前最新的算法比较低。

传统的原型网络独立计算查询实例的嵌入向量和每个支持集的原型向量, Ye^[55]等人提出的 MLMAN 算法是目前最先进的基于原型网络的小样本关系抽取网络模型。MLMAN 的特别之处在于它通过考虑实例级别的匹配信息以交互方式对查询实例和支持实例进行编码。其次,MLMAN 设计了一种辅助损失函数,用于度量支持实例之间的一致性。具体而言,MLMAN 算法采用卷积神经网络对句子进行上下文重表示,然后采用注意力机制获得查询实例和支持实例之间的局部匹配信息然后进行聚合,得到每个实例的重表示;通过使用多层感知机(Multi-layer perceptron, MLP)计算查询实例与支持实例之间的匹配信息,以匹配度为权重,对支持实例的重表示进行加权求和以获得类原型;最后通过 MLP 计算查询实例的重表示与类原型的匹配度从而获得查询实例的关系类型。MLMAN 取得了基于原型网络方法的最新性能,并通过匹配度获得了对查询实例有益的支持实例,但是 MLMAN 也具有局限性:原型网络对于噪声(即不可靠实例)敏感,噪声实例会影响每个类别原型的计算,从而导致算法精度的下降。

基于原型网络的方法由于其局限性,难以避免地产生原型偏差问题,精度难以进一步提高。为了避免陷入原型偏差的问题,一些新的用于小样本关系抽取任务的网络模型被提出, Xie 等人^[56]提出了一种新颖的异构图神经网络 HGNN 以对含有噪声的数据进行关系抽取,具体地, Xie 等人基于句子节点和实体节点构建异构图网络,并将关系分类任务转化为节点分类任务。句子节点可以将标签信息传递到其相邻的实体节点,并且实体节点可以充当句子和关系之间的桥梁。HGNN 可以从两种类型的节点中捕获丰富的邻域信息。此外,为了缓解小样本学习对过拟合和噪声数据的脆弱性, Xie 等人引入对抗学习来训练健壮模型。HGNN 避免了原型偏差并对噪声问题提出了一定的解决策略,但是缺乏针对小样本场景下对查询实例有益的支持实例的选择。

受 MAML^[35]等先进元学习思想启发, Dong 等人于 2020 年提出一种最新的元信息引导元学习框架 MIML^[57],其中类名的语义概念为元学习过程中参数的初始化以及参数的快速更新提供了有力的指导。模型能够在实例信息和语义信息之间建立连接,从而实现更为有效的初始化和更快的适应性。MIML 学习可以有效地为每个类自定义初始化参数的元参数,并选择信息丰富的支持实例以进行快速适应,从而在查询集上产生良好的分类结果。MIML 算法通过类名的语义信息解决了元学习可解释性问题,并通过注意力机制学习一个动态的学习率,解决了实例不可靠的问题,但是 MIML 算法仍然存在以下问题:在无法有效扩充训练样本的情况下, MIML 忽略了对查询实例

有益的支持实例的选择；类名信息的单一性可能导致的偏差，进一步导致对每个实例权重计算的偏差；无法规避可能存在的噪声实例的影响。

本文针对这些具有代表性的小样本关系抽取算法以及最经典的 MAML 算法，在关于算法类型，解决噪声问题，是否引入外部信息，是否选择有益支持实例，是否产生原型偏差等五个方面进行了性能比较，如表 2.2 所示。

表2.2 最新小样本关系抽取算法性能分析及对比表

算法名称	算法类型	噪声	引入外部信息	选择有益实例	原型偏差
MAML ^[35] (2017)	基于优化	一般	否	否	否
TD-Proto ^[51] (2020)	原型网络	较差	是	否	是
BERT-IncreProtoNet ^[54] (2020)	原型网络	一般	否	否	是
MLMAN ^[55] (2019)	原型网络	一般	否	是	是
HATT-Proto ^[52] (2019)	原型网络	较好	否	是	是
HGNN ^[56] (2020)	图神经网络	一般	否	否	否
MIML ^[57] (2020)	基于优化	较好	是	否	否

表 2.2 中展示了一些具有代表性的小样本学习算法，MAML^[35]是最经典的基于优化的元学习框架；TD-Proto^[51]引入外部信息以避免传统原型网络做出不可靠预测，HATT-Proto^[52]是非常优秀的基于原型网络的方法，针对噪声问题有很好的解决能力；MLMAN^[55]通过匹配信息选择对查询实例有益的支持实例取得了基于原型网络最新的性能；但是基于原型网络的方法难以避免地陷入原型偏差问题，MIML^[57]针对噪声问题有较好的性能，但是缺乏针对有益实例的选择且由于类名信息的单一性导致其对实例权重计算的偏差。对这些算法进行研究发现，这些方法都具有局限性，因此，有效利用外部信息，提出一种能够选择对查询实例有益的支持实例且不产生原型偏差的高效的小样本关系抽取算法是本文的关键。

2.5 本章小结

本章首先对小样本学习以及小样本关系抽取相关概念进行介绍，然后介绍了本文所涉及的两个神经网络模型，最后对与本文工作最相关，当前最好的小样本学习算法以及小样本关系抽取算法进行研究并对其优缺点进行分析，这些优秀的算法为本文工作提供了重要的指导意义。

第三章 SS-MTL: 基于自监督的元迁移小样本学习算法

本章首先论述了现有小样本算法的缺陷。之后, 针对现有算法缺陷, 详细论述了本文所提出新的高效的基于自监督的元迁移小样本学习算法 SS-MTL(Self-Supervised based Meta Transfer Learning, SS-MTL)。最后, 在小样本数据集 miniImageNet 和 Fewshot-CIFAR100(FC100)上, 将本文所提出的算法与目前最新最好的小样本学习算法进行精度比较, 以证明本文算法的高效性。

3.1 算法设计动机

深度学习小样本问题的本质是训练过程中的监督信息过少而很难拟合模型中大量参数, 导致其产生过拟合以及精度低等问题。正如前文所述, 现有的普通数据增强方法只能缓解而不能解决小样本问题, 利用生成网络进行数据增强由于先验知识的不完美, 生成的数据与真实数据之间的差异会导致概念偏移。因此如何利用模型从有限少量数据中充分挖掘其信息是解决小样本问题的关键。

元迁移学习是目前最新效果最好的小样本学习算法, 然而仍存在以下缺陷: (1)现有的模型缺乏对不同尺寸特征的挖掘, 卷积神经网络的感受视野都取决于卷积核的大小, 因此只能挖掘固定尺寸的特征, 而图像中特征的尺寸并不是固定的, 因此不能很好地挖掘不同尺寸的特征限制了模型的精度; (2)缺乏对重要特征的关注, 通过对人类对图像的感知的分析表明, 人们主要关注显眼的物体, 而不是图像中的整个场景。而现有网络针对图像或者中间特征的处理都是无差别的, 缺乏对重要特征的关注势必会影响模型的精度; (3)现有算法缺乏针对数据本身结构信息的监督信息的利用, 目前大部分小样本学习算法都是针对计算机视觉任务, 对图像本身结构信息的忽略导致丢失了大量的监督信息。

为了解决现有算法的缺陷, 本文提出了一种基于自监督的元迁移小样本学习算法 SS-MTL(Self-Supervised based Meta Transfer Learning, SS-MTL)。具体地, 本章设计了一种模型不仅能够提取不同尺寸的特征而且增强了对重要特征的关注度; 在此模型的基础上基于图像数据本身的结构信息建立监督信息进行训练以提高模型精度; 并设计了一种新的损失函数, 缓解了算法过拟合的问题。

3.2 算法设计思想及框架

为了解决现有算法的上述缺陷(1), 本文提出了一种双路膨胀卷积网络, 通过膨胀卷积增大了模型的感受视野, 双分支的设计使模型能提取不同尺寸的特征; 为了解决

上述缺陷(2), 本文提出了一种残差注意力网络, 增强了模型对于重要信息的关注度; 为了解决上述缺陷(3), 本文提出了一种基于自监督辅助任务的预训练方法, 通过对图像数据进行旋转, 建立了关于图像结构信息的标签, 增加了训练任务中的监督信息; 此外, 本文提出了一种加入 l_2 正则化项的损失函数, 缓解了小样本学习算法的过拟合问题。基于这些思想, 本文提出了一种基于自监督的元迁移小样本学习算法(Self-Supervised based Meta Transfer Learning, SS-MTL)。下面首先介绍 SS-MTL 算法所使用的网络模型。

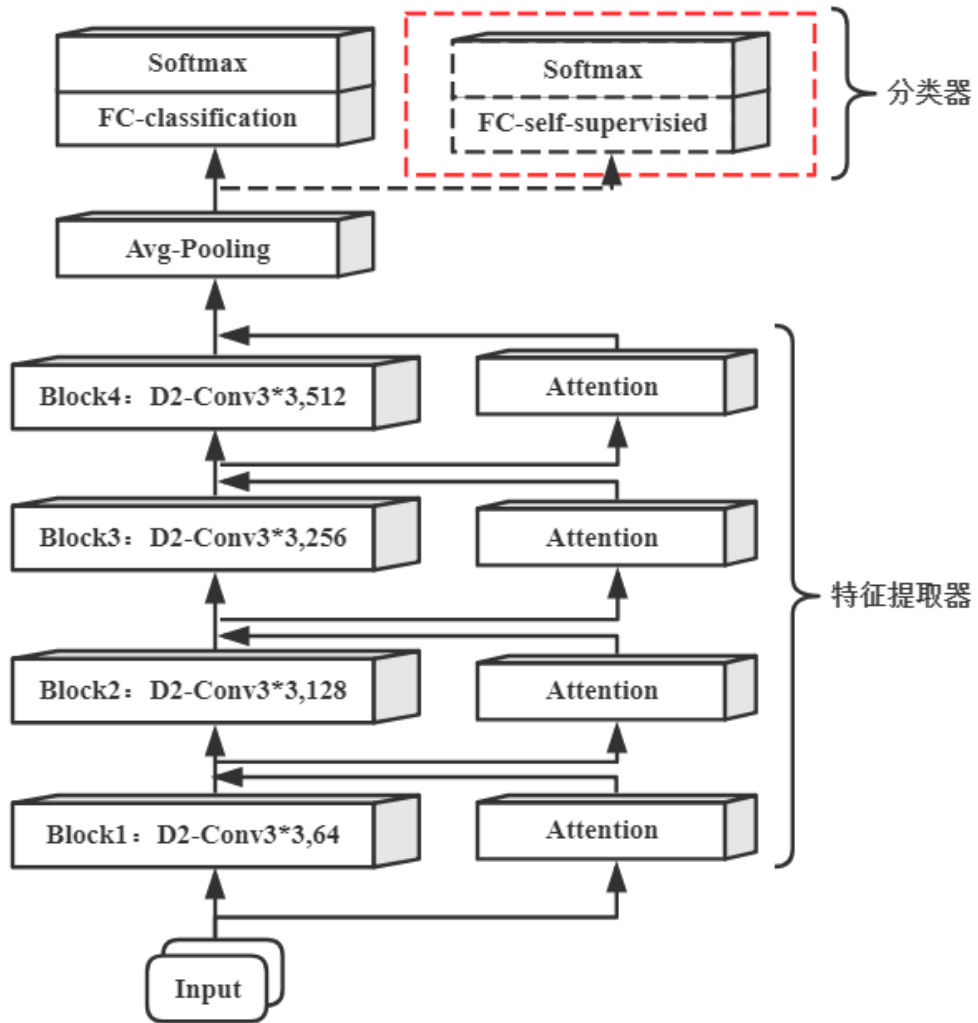


图3.1 SS-MTL 算法模型图(红色框内为预训练阶段添加的自监督分类器)

本文所使用的网络模型如图 3.1 所示, SS-MTL 算法主要由特征提取器以及分类器两个模块构成, 各模块的功能如下:

特征提取器: 为了克服现有算法无法提取不同尺寸特征以及缺乏对重要特征关注度的缺陷, 本文设计了一种添加了膨胀卷积与残差注意力机制的 Resnet12 网络模型作为特征提取器, 对输入网络模型中的图像数据进行处理, 以提取其不同尺寸的特征;

其中 D2-Conv3*3 表示添加了膨胀卷积的, 卷积核尺寸为 3*3 的卷积操作。

分类器: 本文采用一个全连接层 FC-classification 与 Softmax 层作为类别分类器, 全连接层对特征提取器的输出进行处理, 并采用 Softmax 对输出进行归一化处理, 从而获得输入数据属于每个类别的概率。

特别地, 如图 3.1 中红色框所示分类器为在预训练阶段本文设计的自监督辅助任务的自监督分类器(一个全连接层 FC-self-supervised 与 Softmax 层, 与预训练类别分类器相似), 以解决现有算法忽略图像数据本身结构信息的缺陷。具体来说, 本文构建了一种基于旋转的自监督任务, 通过旋转输入数据图像并构建新的旋转角度标签, 从而获得更多的监督信息对模型进行训练, 以训练更强大的特征提取器。预训练阶段的分类器在预训练结束后被丢弃。

下面对本文算法框架进行介绍, SS-MTL 算法框架如图 3.2 所示:

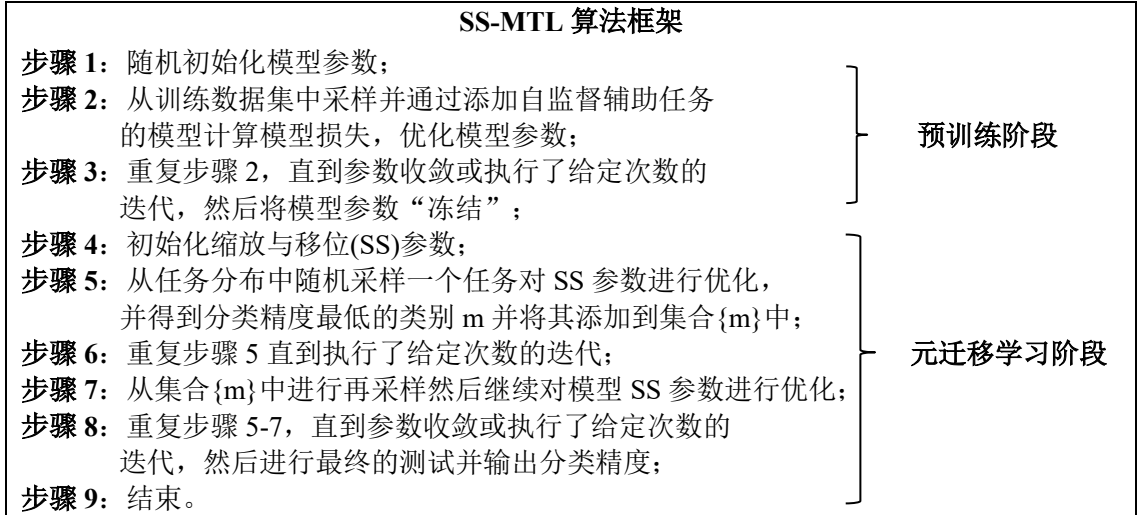


图3.2 SS-MTL 算法框架示意图

如图 3.2 所示, SS-MTL 算法主要分为两个阶段: 预训练阶段和元学习阶段, 其中, 步骤 1-3 为预训练阶段, 步骤 4-9 为元迁移学习阶段。步骤 1 对模型参数进行随机初始化, 步骤 2 对训练集数据进行采样, 然后采用添加自监督辅助任务的方法对模型参数进行优化。步骤 4 对元学习阶段要学习的 SS 参数 Φ_{s_1} 和 Φ_{s_2} 进行初始化, Φ_{s_1} 初始化为全 1, Φ_{s_2} 初始化为全 0。步骤 5-6 通过不断随机采样任务对 SS 参数进行优化, 并得到分类精度最低的类别集合 {m}。步骤 6 采取 Hard Task(HT)机制从精度低的困难任务中进行采样再训练, 进一步优化 SS 参数。

本文将在下一节详细介绍 SS-MTL 算法的主要设计策略。

3.3 算法主要设计策略

3.3.1 特征提取器的设计

为了克服现有算法不能提取不同特征以及缺乏对重要特征的关注度从而导致模型性能受限的缺陷,本文设计了一种添加了膨胀卷积与残差注意力机制的特征抽取器。其中,膨胀卷积增大了模型的感受视野,从而能够提取更大尺寸的特征,残差注意力机制增强了模型对于重要特征的关注度。

增大卷积网络模型感受视野的方法有两种,一是增大卷积核尺寸,二是增大卷积核相邻两步间隔 dp , dp 也称为膨胀系数。固定的卷积核尺寸只能获取固定尺寸的特征,过大的卷积核尺寸无法捕获局部细腻特征,而且大尺寸的卷积核意味着更多的参数,更易产生过拟合,因此本文通过增大卷积间隔来扩大视野。本文中将引入膨胀系数的卷积网络称为膨胀卷积。传统卷积与膨胀系数为 2 的膨胀卷积对比如图 3.3 所示:

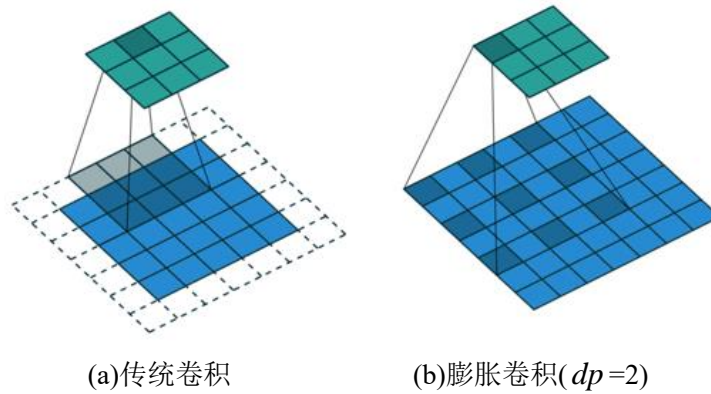


图3.3 传统卷积与膨胀卷积对比示意图

如图 3.3 所示为传统卷积与膨胀系数为 2 的膨胀卷积对比示意图,在卷积核尺寸相同的情况下,膨胀卷积能够提取更大尺寸的特征。具体地,本文基于 Resnet12 模型的结构,在特征提取器的每一个残差块(Block)添加了膨胀卷积,如图 3.1 中所示的特征提取器网络中共包括 4 个残差块(Block),其中 Conv3*3 代表每个残差块中卷积核尺寸为 3*3,每个残差块中卷积核个数分别为: 64, 128, 256 和 512。四个残差块的结构相同,只是卷积核个数不同,为了便于描述,本文中用 Num 表示各个残差块中的卷积核个数。改进前的 Resnet12 网络模型中每个残差块的结构如图 3.4 所示:

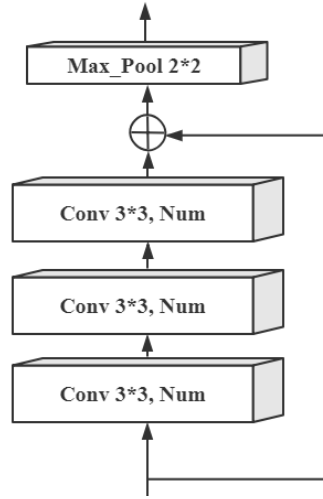


图3.4 改进前的 Resnet12 网络残差块示意图

改进后的残差块的结构，如图 3.5 所示：

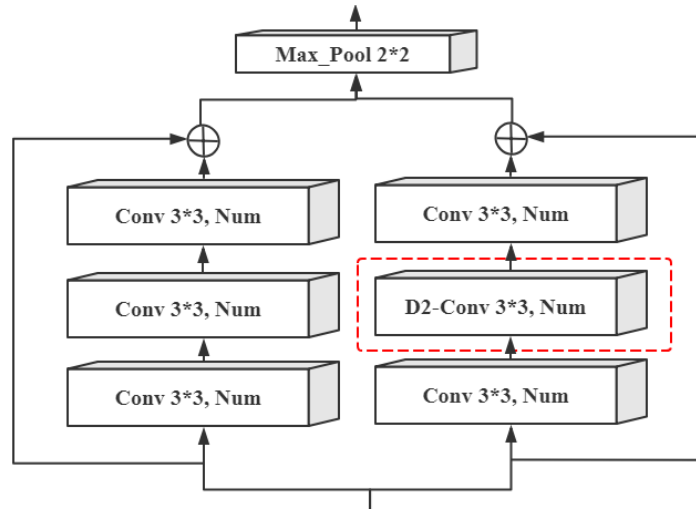


图3.5 改进后的残差块(红色框为本文添加的膨胀卷积层)

如图 3.5 所示，在保持原始网络模型(如图 3.5 左分支)的基础上，添加一个膨胀卷积分支(如图 3.5 右分支)，其中红色框中 D2-Conv3*3 表示膨胀系数为 2 的卷积操作。最后通过逐元素取最大的操作将两个分支所得特征向量进行整合，作为下一步的输入。

膨胀卷积克服了传统卷积网络模型只能提取固定尺寸特征的缺陷，通过对膨胀系数 dp 的调节能够获得不同尺寸的特征。本文采用双分支的设计，原始卷积分支提取尺寸为 $3*3$ 的特征，本文提出的膨胀卷积在卷积核为 $3*3$ 的情况下采用膨胀系数为 2 的卷积可以提取尺寸为 $5*5$ 的特征，实现了同时提取不同尺寸特征的目标，进一步增强了模型的性能。双路膨胀卷积网络通过膨胀卷积能够捕获不同尺寸的特征，且双路

设计使得每个分支独立更新自己的参数，进一步增强了模型的适应性。

为了克服目前算法缺乏对重要特征关注度的缺陷，本文基于提出的膨胀卷积网络对每一个 Block 添加了残差注意力机制，以提高对重要特征的关注度。

与大部分卷积网络相同，Resnet12 网络作为特征提取器，对于输入网络模型中的图像或网络中的中间特征的处理都是无差别的。但是人类智能与深度学习不同，人类偏向于关注显眼的物体，深度学习中的注意力机制从本质上来说和人类的选择性视觉注意力机制类似，核心目标也是从众多信息中选择出对当前任务目标更关键的信息，它会增强有意义的特征，而抑制无意义的信息。本文在提出的膨胀卷积网络模型中加入一种残差注意力机制，该注意力网络是通过堆叠在网络的每个残差块上的注意力模块而构建的，该机制的主要思想如图 3.6 所示：

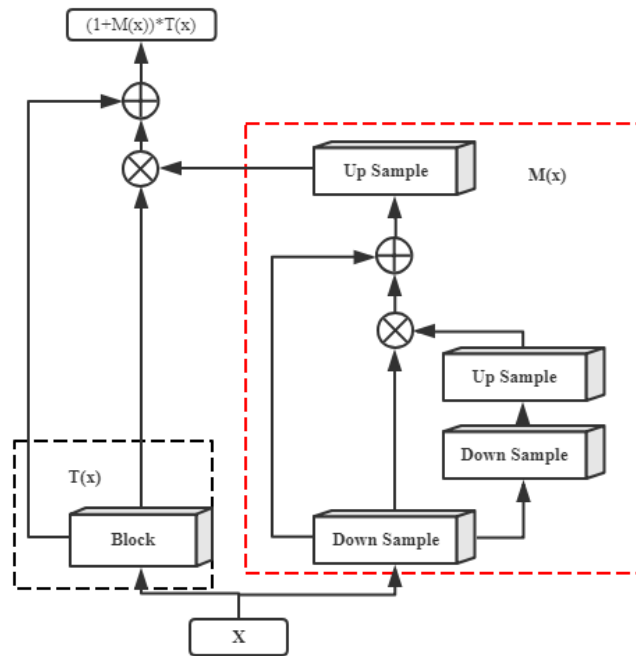


图3.6 残差注意力机制示意图(红色框为本文设计的注意力分支)

在图 3.6 中，每个注意力模块都分为两个分支：主干分支和掩码分支。在图 3.6 中，X 表示输入，黑色分支表示原网络残差块中的卷积主分支(Trunk branch)， $T(x)$ 表示主分支输出，红色分支表示对输入进行两次下采样和两次上采样操作后，输出与 $T(x)$ 尺寸相同的掩码分支(Mask branch)，掩码分支输出 $M(x)$ 则作为主分支的 Attention Map(注意力向量)，Attention Map 中的每一个值相当于对原始特征向量上每一个元素的权重，将 $(1+M(x))*T(x)$ 作为最终输出，其中掩码分支 $M(x)$ 的计算与整个模型的输出计算方式相似，当 $M(x)=0$ 时，该层的输出就等于 $T(x)$ ，同时这样的相加，使得显著的特征更加显著，增加了特征的判别性。

其中，下采样操作采用最大池化方法，上采样则采用双线性插值法，双线性插值算法是一种很好的图像缩放算法。它充分利用原始图像中虚拟点周围的四个实际像素值来共同确定目标图像中的像素值。如图 3.7 所示为在二维空间内，通过 $P_{11}, P_{12}, P_{21}, P_{22}$ 4 个点的像素值利用双线性插值方法确定点 Q 的像素值，其计算过程如下：

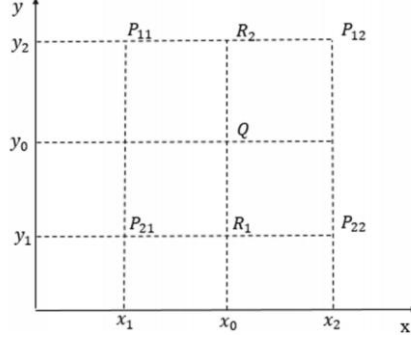


图3.7 线性插值示意图

首先，如公式(3-1)与 (3-2) 所示，分别在 x 方向上执行两次插值计算，以计算 R_1 和 R_2 的像素值。

$$f(x_0, y_1) \approx \frac{x_2 - x_0}{x_2 - x_1} f(P_{21}) + \frac{x_0 - x_1}{x_2 - x_1} f(P_{22}) \quad (3-1)$$

$$f(x_0, y_2) \approx \frac{x_2 - x_0}{x_2 - x_1} f(P_{11}) + \frac{x_0 - x_1}{x_2 - x_1} f(P_{12}) \quad (3-2)$$

其中， $f(x_0, y_1)$ 表示 R_1 的像素值， $f(x_0, y_2)$ 表示 R_2 的像素值， $f(P_{ij}), i, j \in \{1, 2\}$ 表示点 P_{ij} 的像素值。其次，如公式(3-3)所示，执行在 y 方向上的线性内插，通过 R_1 和 R_2 计算 Q 点的像素值 $f(x_0, y_0)$ 。

$$f(x_0, y_0) \approx \frac{y_2 - y_0}{y_2 - y_1} f(x_0, y_1) + \frac{y_0 - y_1}{y_2 - y_1} f(x_0, y_2) \quad (3-3)$$

最大池化即只取对应位置的最大值，并且不经过反向传播的修改，池化层对特征进行压缩时只是过滤掉无关的信息，而留下来的信息则具有尺度不变性的特征，是最能表达图像内容的特征，这一过程加强了重要的特征。由双线性插值的具体过程可知，对获得的重要特征进行双线性插值能使得原来不重要特征产生退化，这一过程减弱了不重要的特征。通过两次上采样与下采样的操作，模型更加突出了重要的特征而抑制了不重要的特征。本文提出的残差注意力机制的添加使得提出的膨胀卷积网络模型能够专注于更重要的特征，堆叠结构的应用使得模型可以在不同的注意力模块中各自捕获不同的注意力向量，进一步提高了算法的精度。

综合以上两种机制，本文通过设计一种双路膨胀卷积网络，使得模型能够提取到不同尺寸的特征，克服了传统小样本学习模型不能提取不同尺寸特征的缺陷；本文通过设计一种残差注意力机制，增强了模型对于重要特征的关注度。

3.3.2 模型损失函数

深度神经网络模型中的损失函数用来衡量模型预测的好坏，并对模型参数进行更新。一个好的损失函数对于训练一个优秀的模型至关重要。当训练样本过少时，深度模型很容易产生过拟合问题，即模型在训练样本上性能好，而在测试样本上性能差。为了缓解小样本学习算法中过拟合的问题，本文提出了一种添加 l_2 正则化项的损失函数，如公式(3-4)所示：

$$loss = -\frac{\sum_{c=1}^N y_c \log(p_c)}{K} + \|w\|^2 \quad (3-4)$$

其中， $-\sum_{c=1}^N y_c \log(p_c)/K$ 表示模型的分类损失项， N 代表类别数， y_c 表示变量 0 或 1，如果预测类别和样本的实际类别相同则为 1，否则为 0， p_c 是样本属于类别 c 的概率，即模型的输出值， K 代表每类样本数量； $\|w\|^2$ 表示 l_2 正则化项， w 是权重参数。

深度网络模型产生过拟合的原因是深度网络模型中的参数太多导致模型复杂度过高，而小样本学习过程中训练样本过少，很容易导致过拟合现象。 l_2 正则化就是模型参数的 l_2 范数，将参数的 l_2 范数加入损失函数中求最优， $\|w\|^2$ 项最小化，从而使得 w 的每一个元素都是一个接近于 0 的非常小的值。因为一般认为参数值小的模型比较简单，能够适应不同的数据集，也在一定程度上避免了过拟合现象。若模型参数大，只要数据有微小的偏移，就会对结果产生非常大的影响，但模型参数足够小，即使数据偏移大，对结果的影响也会很小，模型的抗扰动能力就会很强。

3.3.3 基于自监督辅助任务的预训练阶段

为了解决小样本学习算法缺乏监督信息的问题，本文提出了一种基于自监督辅助任务的预训练方法，自监督主要是利用辅助任务从大规模的无监督数据中挖掘自身的监督信息，通过这种构造的监督信息对网络进行训练，从而可以学习到丰富的，通用的对下游任务有价值的表征。也就是说，可以仅使用数据中存在的结构信息来提取监督信息。目前大多数预训练方法以监督的方式训练，所以导致模型的迁移能力是有限的。如果使用自监督方式，设置一个辅助任务，就可以从数据本身学习到健壮的代表，提高模型的泛化能力。目前最常用的是基于旋转的方法，并且对许多模型都是比较合适的。本文提出的基于自监督辅助任务的预训练方法如图 3.8 所示：

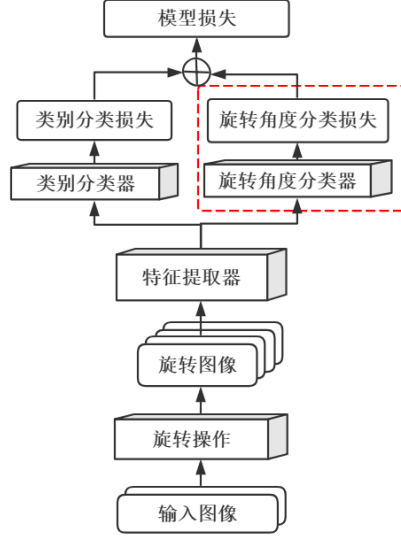


图3.8 基于自监督的预训练示意图(红色框内为本文自监督辅助任务)

如图 3.8 所示, 本文提出了一种基于旋转的自监督预训练方法, 对于每一个批次的图片分别进行 0° , 90° , 180° , 270° 旋转, 并将旋转角度作为监督信息建立图像旋转角度标签, 从而获得了更多的监督信息。由此本文构建了一个旋转角度分类器, 旋转角度分类损失与标准类别分类损失作为最终的损失函数对模型进行预训练。

与经典的预训练方法不同, 本文采用大规模数据集进行基于自监督辅助任务的预训练, 如对于 miniImageNet 数据集(如 3.5.1 节所示), 训练集中总共有 64 个类, 每个类包含 600 个用于预训练 64 类分类器的样本。在本阶段, 首先随机初始化特征提取器 Θ , 传统类别分类器 θ 以及自监督辅助任务的旋转角度分类器 θ_{ss} , 然后通过梯度下降法利用公式(3-5)对其进行优化:

$$[\Theta, \theta, \theta_{ss}] = [\Theta, \theta, \theta_{ss}] - \alpha \nabla L_D([\Theta, \theta, \theta_{ss}]) \quad (3-5)$$

其中 L_D 代表损失函数, 此处采用交叉熵损失函数, α 代表学习率, 如上一小节所叙述, 本文采取了添加自监督辅助任务的预训练方式, 因此 L_D 包含传统的分类损失与旋转角度分类损失, 具体计算方式如下所示:

$$L_D = \lambda L_{\text{Classification}} + (1 - \lambda) L_{\text{Self-supervised}} \quad (3-6)$$

$$L_{\text{Classification}} = L_D([\Theta, \theta]) = \frac{1}{|D|} \sum_{(x, y) \in D} l(f_{[\Theta, \theta]}(x), y) \quad (3-7)$$

$$L_{\text{Self-supervised}} = L_D([\Theta, \theta_{ss}]) = \frac{1}{|D|} \sum_{(x,y) \in D} l(f_{[\Theta, \theta_{ss}]}(x), y) \quad (3-8)$$

其中, λ 代表类别分类损失与旋转分类损失之间的平衡参数, 在本文的实验中选取为 0.7。在预训练阶段主要学习特征提取器 Θ , 它将在接下来的元迁移学习阶段被冻结。分类器 θ 以及 θ_{ss} 将被丢弃, 因为随后的小样本学习任务包含不同的分类目标, 例如 5-way-1-shot 的分类是 5 个类别而不是预训练阶段的 64 类, 且在元迁移学习阶段没有添加关于旋转的自监督分类目标。

3.3.4 元迁移学习阶段

元学习由两个阶段组成: 元训练和元测试。整个训练集 D_{train} 是从分布 $P(T)$ 采样的分类任务 T , 对整个模型进行训练, 测试集 D_{test} 构建模型从未见过的任务 T_{unseen} , T 被称为一个 episode, 在元训练阶段, 任务 T 中包括用于优化特征提取 Θ 的训练划分 $T^{(tr)}$ 即支持集 $D_{train-support}$ 和用于优化分类器 θ' 的测试划分 $T^{(te)}$ 即查询集 $D_{train-query}$ 。元训练旨在从 $P(T)$ 采样的多个 episode, 即 $\{T\}$ 中学习。元测试阶段旨在从未曾见过的任务 T_{unseen} 上基于特征提取 Θ 的经验去适应分类器 θ' , 通过测试 $T_{unseen}^{(te)}$ 即 $D_{test-query}$ 在模型上的效果来完成最终评估。

正如 2.3.2 节中所述, 本文采用的 HT 元批处理训练优化了 SS 参数, SS 参数表示为 Φ_{s_1} 和 Φ_{s_2} , 图 2.7 展示了 SS 操作的具体过程, 在元迁移学习期间不会更改冻结的特征提取器 Θ 神经元权重。下面本文将详细介绍元迁移学习的过程:

给定一个任务 T , 与其相关的训练损失用于通过梯度下降来优化当前分类器 θ' , 如公式^[38](3-10)所示:

$$\theta' \leftarrow \theta - \beta \nabla_{\theta} L_{T^{(tr)}}([\Theta, \theta], \Phi_{s_{\{1,2\}}}) \quad (3-9)$$

与公式(3-5)不同, 这里不更新特征提取器 Θ , 需要注意的是, 这里的 θ 与上一阶段(公式(3-5)中的大型分类器 θ)不同。这个 θ 仅涉及少量几个类别, 每次都对新的小样本设置进行分类。 θ' 对应于仅在当前任务中起作用的, 有时效性的分类器, 通过针对上一个任务优化过的 θ 进行初始化(请参见公式(3-11))。

对 SS 参数 Φ_{s_1} 进行行全 1 初始化, Φ_{s_2} 进行全 0 初始化, 然后通过任务 $T^{(te)}$ 的测试损失对他们进行优化, 如公式^[38](3-10)所示:

$$\Phi_{s_i} = \Phi_{s_i} - \gamma \nabla_{\Phi_{s_i}} L_{T^{(te)}}([\Theta, \theta'], \Phi_{s_{\{1,2\}}}) \quad (3-10)$$

其中, γ 表示学习率, 在这一步骤, θ 以与公式^[38](3-10)中相同的学习率 γ 来更新。

$$\theta = \theta - \gamma \nabla_{\theta} L_{T^{(te)}}([\Theta, \theta], \Phi_{S_{\{1,2\}}}) \quad (3-11)$$

在公式^[38](3-11)中, θ' 来自于任务 $T^{(tr)}$ 的最后一个 epoch。关于 SS 参数的具体原理已在 2.3.2 节进行介绍, 这里不再赘述。

3.4 算法实现

本文提出的 SS-MTL 算法伪码如图 3.9 所示。其中算法 1 宏观地介绍了算法的整个流程, 包括预训练与元迁移学习阶段; 算法 2 介绍了在一个任务上的学习步骤。

算法 1: SS-MTL
输入: 任务分布 $P(T)$ 及相关的数据集 D , 学习率 α, β 以及 γ
输出: 特征提取器 Θ , 分类器 θ , SS 参数 Φ_{S_1} 和 Φ_{S_2}
<pre> 1: 随机初始化 Θ 和 θ 2: for samples in D do 3: 分别旋转每张图像并构建自监督的标签 4: 通过公式(3-6)计算 L_D 5: 通过公式(3-5)优化 Θ 以及 θ 6: end for 7: 对 Φ_{S_1} 进行全 1 初始化, 对 Φ_{S_2} 进行全 0 初始化 8: 针对小样本任务重新设置并初始化分类器 θ 9: for meta-batches do 10: 从任务分布 $P(T)$ 随机采样一组任务 $\{T\}$ 11: while not done do 12: 从这组任务中采样一个任务 $T_i \in \{T\}$ 13: 利用任务 T_i 通过算法 2 优化 $\Phi_{S_{\{1,2\}}}$ 以及 θ 14: 得到返回的精度最低的类 $class-m$ 并将其添加到 $\{m\}$ 15: end 16: 从 $p(T \{m\})$ 中采样困难任务 $\{T_{hard}\}$ 17: while not done do 18: 采样一个困难任务 $T_j^{hard} \in \{T^{hard}\}$ 19: 利用 T_j^{hard} 通过算法 2 优化 $\Phi_{S_{\{1,2\}}}$ 以及 θ 20: end 21: 清空 $\{m\}$ 22: end for </pre>

图3.9 SS-MTL 算法伪码示意图

如图 3.9 所示, 输入: 任务分布 $P(T)$ 以及与其相关的数据集 D , 学习率 α , β 和 γ ; 输出: 特征提取器 Θ , 分类器 θ , SS 参数 Φ_{S_1} 和 Φ_{S_2} 。第 1-6 行表示大规模数据

集上的预训练阶段；其中第 1 行，随机初始化特征提取器 Θ 和大规模分类器 θ ；第 2-6 行遍历数据集 D 中的样本，通过公式(3-6)计算损失，由于添加了自监督辅助任务，这里的损失既包括原始的分类损失，也包括添加自监督辅助任务后的旋转角度分类损失，然后第 4 行通过公式(3-5)对特征提取器 Θ 和大规模分类器 θ 进行优化。第 7-22 行表示元迁移学习阶段；其中第 7 行表示 SS 参数的初始化，在元迁移阶段特征提取器的参数被冻结，通过 SS 操作给特征提取器 Θ 的卷积核添加了缩放与移位参数 Φ_{s_1} 和 Φ_{s_2} ， Φ_{s_1} 初始化为全 1， Φ_{s_2} 初始化为全 0；第 8 行针对小样本任务重新设置并初始化分类器 θ ；第 9-22 行表示通过多个批次对模型进行训练；第 10 行从任务分布 $P(T)$ 中采样一批任务 $\{T\}$ ；第 11-15 行通过算法 2 利用这些任务对模型进行训练，并得出精度最低的类 $class-m$ 加入到集合 $\{m\}$ 中，其中第 13 行采用算法 2 对 Φ_{s_1} 和 Φ_{s_2} 以及分类器 θ 进行优化；第 16 行从集合 $\{m\}$ 的分布中采样一批困难任务 $T_j^{hard} \in \{T^{hard}\}$ ；第 17-20 行通过算法 2 利用 HT 元批处理训练对 Φ_{s_1} 和 Φ_{s_2} 以及分类器 θ 进行进一步优化；第 21 行表示每一个批次训练结束后将集合 $\{m\}$ 清空。

算法 2: 在一个任务上的 SS-MTL	
输入: 任务 T ，学习率 β 以及 γ ，特征提取器 Θ ，分类器 θ ，SS 参数 Φ_{s_1} 以及 Φ_{s_2}	
输出: 更新后的 θ 以及 Φ_{s_1} 和 Φ_{s_2} ，在任务 T 中分类结果最差的类 $class-m$	
1:	从 T 中采样 $T^{(tr)}$ 和 $T^{(te)}$
2:	for samples in $T^{(tr)}$ do
3:	计算 $L_{T^{(tr)}}$
4:	通过公式(3-9)优化 θ'
5:	end for
6:	通过公式(3-10)以及(3-11)优化 θ , Φ_{s_1} 以及 Φ_{s_2}
7:	while not done do
8:	从 $T^{(te)}$ 中采样一个类 $class-k$
9:	计算 $T^{(te)}$ 中的分类精度 Acc_k
10:	end
11:	返回具有最差分类精度 Acc_m 的类 $class-m$

图3.10 SS-MTL 算法在一个任务上的学习步骤示意图

图 3.10 中对算法 1 中针对每个任务的具体学习步骤进行阐述。如图 3.10 所示，输入：任务 T ，学习率 β 和 γ ，特征提取器 Θ ，小样本分类器 θ ，SS 参数 Φ_{s_1} 和 Φ_{s_2} ；输出：更新之后的小样本分类器 θ ，SS 参数 Φ_{s_1} 和 Φ_{s_2} ，以及在任务 T 中分类结果最差的类 $class-m$ ；第 1 行从任务 T 中划分出支持集 $T^{(tr)}$ 以及查询集 $T^{(te)}$ ；第 2-5 行遍历支持集 $T^{(tr)}$ 的样本并计算分类损失，采用公式(3-9)优化 θ' ；第 6 行通过公式(3-10)和(3-11)优化小样本分类器 θ ，SS 参数 Φ_{s_1} 和 Φ_{s_2} ；第 7-10 行从查询集 $T^{(te)}$ 中采样并

计算分类精度；第 11 行返回精度最低的 *class-m*。

3.5 算法实验及实际应用

本节在两个标准数据集上对所提出的算法 SS-MTL 进行实验，并与其他最新最好的小样本学习算法进行比较。本节对算法的实验平台，实验数据集，对比标准，实验结果以及对实验结果的分析进行了介绍。最后

3.5.1 实验环境及设置

为了对比的公平性，本文所有实验均在 Linux 环境下运行，采用 Intel(R) Xeon(R) Gold 5115 CPU @ 2.40GHz 的 CPU 以及 NVIDIA Tesla P40 的 GPU。其中 SS-MTL 算法是使用 Python 3.5 和 Tensorflow 1.3.0 实现和进行测试的。

本文针对两个小样本标准数据集 miniImageNet 和 FC100 进行实验。

miniImageNet 数据集由 Vinyals 等人于 2016 年提出，用于小样本学习算法的评估，miniImageNet 节选自 ImageNet 数据集，ImageNet 是为促进视觉识别的研究而建立的一个大型的数据库，它包含 20000 多个类别，如“汽车”，“水母”，“气球”等，每个类别均包含 500 张以上的图片。miniImageNet 节选了其中 100 个类别，每个类别包含 600 个 84×84 彩色图像样本。这 100 个类别中有 64 个类别用于元训练，16 个类别用于元验证，20 个类别用于元测试的采样任务。

FC100 数据集是基于流行的对象分类数据集 CIFAR100 构建的。它提供了更具挑战性的方案，它具有较低的图像分辨率和更具挑战性的元训练/测试划分。它包含 100 个对象类，每个类有 600 个 32×32 彩色图像样本。这 100 个类别属于 20 个超类。元训练数据来自属于 12 个超类的 60 个类，元验证和元测试集分别包含 20 个类，分别属于 4 个超类。这些划分是基于超类的，从而最大程度地减少了训练和验证以及测试任务之间的信息重叠。

本文基于 miniImageNet 数据集进行了 2 种分类实验设置，分别为 5-way-1-shot, 5-way-5-shot；基于 FC100 数据集进行了 3 种分类实验设置，分别为 5-way-1-shot, 5-way-5-shot, 以及 5-way-10-shot, N-way-K-shot 设置表示每个测试批次具有 N 个未出现在训练集中的类别，每个类别有 K 个图像样本。本文采用图像分类准确率 Acc 作为评估指标，具体计算方法如下：

$$Acc = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} * 100\% \quad (3-12)$$

3.5.2 实验参数设置

下面对算法 SS-MTL 所需的一些重要超参数进行介绍，本文所做的关于 SS-MTL 的实验均遵循表 3.1 所示的参数值：

表3.1 SS-MTL 算法参数表

参数名称	参数意义	参数值
PRE_TRA_ITER_MAX	预训练阶段迭代次数	20000
PRE_LR	预训练学习率	0.001
MAX_ITER	元训练迭代次数	20000
β, γ	元学习率	0.001
α	基础学习率	0.001

3.5.3 对比算法介绍

本文将提出的算法 SS-MTL 算法与以下几个较为优秀的小样本学习算法进行比较：

(1) MAML 算法：该方法是经典基于优化的元学习框架，很多元学习算法均基于此优化框架进行改进，具体可见 2.3.1 小节。

(2) MAML++算法：基于 MAML 算法进行改进，不仅提高了 MAML 训练的稳定性，而且可以极大地提高 MAML 的泛化性能，收敛速度并降低计算开销。

(3) TADAM 算法：基于度量的小样本学习算法，提出一种辅助任务协同训练的方法。

(4) TransMatch 算法：目前最新的基于迁移学习的混合型小样本学习算法，该算法采用半监督学习方法对模型进行更新。

(5) MTL 算法：一种元学习与迁移学习相结合的小样本学习算法，该算法通过冻结预训练权重并添加缩放参数的机制以适应新的任务。

(6) T-NAS：提出一种基于元学习的可迁移的神经体系结构搜索方法，在小样本学习算法中有效减少了参数数量。

这 6 个算法都是非常优秀的小样本学习算法，其中 MAML 是 2017 年提出的经典的元学习框架，TADAM 是 2018 年提出的一种辅助任务训练的小样本学习算法，2019 年提出的 MAML++算法针对 MAML 进行了一系列的改进，取得了更好的性能，本文基于 2019 年提出的 MTL 算法进行改进，MTL 算法是本文的重点对比算法，TransMatch 和 T-NAS 是 2020 年最新基于迁移学习和基于元学习的混合型小样本学习算法，也是目前效果最好的算法。

3.5.4 算法精度比较及分析

(1) 自监督辅助任务的有效性

本节对本文所提出基于自监督的预训练方法的有效性进行验证, 分别对采用 2 分类自监督辅助任务, 4 分类自监督辅助任务以及不采用自监督辅助任务的本文算法在 miniImageNet 数据集上进行了 5-way-1-shot 以及 5-way-5-shot 实验, 如表 3.2 所示。

表3.2 自监督辅助任务实验精度对比表

实验 设置	Pretrain(%)		2-SS-Pretrain(%)		4-SS-Pretrain(%)	
	Meta-Tr	Meta-Te	Meta-Tr	Meta-Te	Meta-Tr	Meta-Te
5-way-1-shot	59.7	57.1	60.2	57.9	62.4	58.4
5-way-5-shot	73.6	72.8	74.7	73.5	75.8	74.3

在表 3.2 中, 本文分别基于 2 分类自监督辅助任务与 4 分类自监督辅助任务, 以及未添加自监督辅助任务的算法进行实验, 其中 2-SS-Pretrain(2-self supervised-Pretrain)表示添加 2 分类自监督辅助任务的 MTL 算法, 分别对输入图像进行 0° , 180° 旋转; 4-SS-Pretrain(4-self supervised-Pretrain)表示添加 4 分类自监督辅助任务的 MTL 算法, 分别对输入图像进行 0° , 90° , 180° , 270° 旋转。Meta-Tr 代表在元训练阶段的分类精度, Meta-Te 表示在元测试阶段的分类精度。由实验结果可知:

相比于传统的预训练方法, 添加自监督辅助任务的预训练不仅在预训练的过程中取得较好的精度, 在元迁移学习阶段也能取得具有竞争力的结果, 充分说明添加自监督辅助任务的有效性; 由于构造了更多的监督信息, 相对于 2 分类自监督, 4 分类自监督辅助任务对于算法精度提高取得的效果更为显著; 预训练阶段自监督辅助任务的添加对于算法的精度有了进一步的提升, 并且是在保持原始需要优化的参数量的情况下。

(2) l_2 正则化项缓解过拟合的有效性

表3.3 l_2 正则化项对过拟合有效性对比表

实验设置	5-way-1-shot(%)		5-way-5-shot(%)	
	train	test	train	test
未添加 l_2 正则化项	45.17	38.69	60.94	52.63
添加 l_2 正则化项	45.55	43.20	59.68	54.45

为了验证本文提出的添加 l_2 正则化项对缓解模型过拟合问题的有效性, 本文分别

使用添加 l_2 正则化项的损失函数与未添加的损失函数在 miniImageNet 数据集上进行 5-way-1-shot 以及 5-way-5-shot 实验(5000 次迭代), 实验结果如表 3.3 所示:

表 3.3 所示为添加 l_2 正则化项的损失函数与未添加的损失函数在 miniImageNet 数据集上进行 5-way-1-shot 以及 5-way-5-shot 实验分别在训练阶段与测试阶段的精度。由表中数据可知, 未添加 l_2 正则化项的实验中相比于训练过程中的精度, 测试精度下降严重, 产生了过拟合现象。而添加了 l_2 正则化项的实验中, 精度下降幅度较小, 实验结果表明, 添加 l_2 正则化项的损失函数能够有效缓解小样本学习算法的过拟合问题。

(3) SS-MTL 算法与对比算法的精度比较

本小节对本文提出的 SS-MTL 算法与目前最好的小样本学习算法在 miniImageNet 以及 FC100 上进行了实验, 结果如下所示。

表3.4 SS-MTL 算法与对比算法在 miniImageNet 上精度对比表

算法名称	5-way-1-shot(%)	5-way-5-shot(%)
MAML ^[35] (2017)	48.70±1.75	63.11±0.92
MAML++ ^[36] (2019)	52.15±0.26	68.32±0.44
TADAM ^[37] (2018)	57.8±0.3	75.90±0.30
MTL ^[38] (2019)	57.1±1.4	72.8±0.7
TransMatch ^[39] (2020)	58.43±0.93	76.43±0.61
T-NAS ^[40] (2020)	54.11±1.3	69.59±0.8
SS-MTL(ours)	59.8±1.6	76.6±0.9

表3.5 SS-MTL 算法与对比算法在 FC100 上精度对比表

算法名称	5-way-1-shot(%)	5-way-5-shot(%)	5-way-10-shot(%)
MAML ^[35] (2017)	38.1±1.7	50.4±1.0	56.2±0.8
MAML++ ^[36] (2019)	38.7±0.4	52.9±0.4	58.8±0.4
TADAM ^[37] (2018)	40.1±0.4	56.1±0.4	61.6±0.5
MTL ^[38] (2019)	43.8±1.4	55.4±1.1	61.2±0.6
TransMatch ^[39] (2020)	44.7±1.1	55.8±0.7	63.9±0.7
T-NAS ^[40] (2020)	40.4±1.2	54.6±0.9	60.2±0.7
SS-MTL(ours)	45.7±1.3	56.9±1.1	64.2±0.9

表 3.4 和表 3.5 展示了本文提出的小样本学习算法与目前最新算法在不同的小样本场景下的分类精度结果。通过对表 3.4 及表 3.5 实验结果进行分析, 可得出以下结

论:

本文提出的 SS-MTL 算法在两个实验数据集上不同实验设置下均优于目前最新的小样本学习算法,充分验证了本文提出算法的有效性。本文提出的 SS-MTL 算法精度相对于 MTL 算法有了明显的提升,特别是在 miniImageNet 数据集上,这是因为 miniImageNet 数据集中的图像尺寸相对于 FC100 数据集比较大,本文通过膨胀卷积增大了模型的感受野,所以对于固定尺寸的卷积核,尺寸较大的图像能挖掘其更多大尺寸的特征,因此针对于算法精度的提升更为显著。

3.5.5 SS-MTL 算法实际应用 1: 鸟类识别

传统的深度神经网络模型通过大规模数据的训练已经可以实现简单的图像分类任务,然而对于复杂的图像,特别是数据量不足的情况下,这些模型难以达到令人满意的结果。图像分类是经典的基于深度学习的小样本学习任务,基于小样本学习的鸟类识别旨在将少量不同的鸟类图像输入深度神经网络模型,深度神经网络模型就能对这些图像特征进行学习,进而根据其特征识别出其属于的类别,当再给定一些深度神经网络模型从未学习过的鸟类图像,深度神经网络也能很好地对其进行分类。基于本文提出的 SS-MTL 算法,通过自监督的预训练方式已经获得了一个良好的特征提取器,通过迁移学习的方法很容易将其应用到小样本学习场景中。

CUB-200-2011 数据集是由加利福尼亚理工学院所提供的用于小样本鸟类细粒度识别的数据集,其中包括 200 个鸟类的超过 10000 张图片,本文采用其中 100 个类别作为元训练集,剩余的 100 个类别作为元测试集。本章所采用的鸟类识别应用的流程如下图所示:

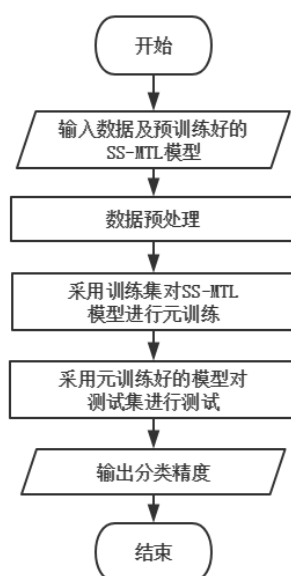


图3.11 小样本鸟类识别流程图

如图 3.11 所示为整个鸟类识别的流程图，首先对所有鸟类图片进行了预处理，将每一张图片采用 Python 的 PIL 工具通过裁剪压缩处理为 $84 \times 84 \times 3$ 的图片，以方便模型的处理。然后通过训练数据对 SS-MTL 模型进行元训练，最后将训练好的模型对测试数据进行分类。本文基于 CUB-200-2011 数据集进行了 100-way-1-shot 以及 100-way-5-shot 的分类实验，并与最新的 MTL 算法以及 TransMatch 算法进行比较，所得精度如下表所示：

表3.6 SS-MTL 算法在 CUB-200-2011 数据集上性能对比表

算法	100-way-5-shot(%)	100-way-10-shot(%)
TransMatch ^[28] (2020)	52.25	63.79
MTL ^[26] (2019)	51.97	63.54
SS-MTL	52.78	65.03

由表 3.6 中数据可知，本文提出的算法相对于小样本情况下的鸟类识别精度高于目前最新的算法，进一步证明了本文提出的 SS-MTL 算法针对鸟类识别的有效性，以及相对于其他算法在鸟类识别方面的优越性。

3.5.6 SS-MTL 算法实际应用 2：手写字符识别

人类可以很容易理解一张图片所表达的信息，但是对于计算机来说，这是非常困难的任务。手写字符识别作为智能计算机接口的重要组成部分，在现实生活中的应用非常广泛，如银行票据的处理，文献检索，邮政系统及自动化办公等方面。但是，不同语言间不同意义的字符可能存在非常相近的特征，而同一字符由不同的人手写可能表现出非常迥异的特征，这是手写字符识别面临的非常严峻的挑战。

本文基于 Omniglot 手写字符数据集进行了手写字符识别的应用研究，该数据集包含了 50 种不同语言的 1623 个手写字符，每个字符包含 20 种不同的笔迹，本文选取其中 1150 个类作为训练集，50 个类作为验证集，剩下的 423 个类作为测试集。本文基于手写字符识别的应用的基本流程如图 3.12 所示：

如图 3.12 所示为整个手写字符识别流程，由于数据集中的每张图像已经被裁剪到统一尺寸为 28×28 ，这里不再需要对其进行预处理，由于每个类别的样本量过少，本文采用旋转的方式进行数据增强，采用增强的数据集对本文提出的 SS-MTL 模型进行元训练，然后将测试集输入模型，对模型的性能进行测试，本文基于 Omniglot 数据集分别进行 20-way-1-shot 以及 20-way-5-shot 实验，并与其他对比算法进行比较，实验结果如下表 3.7 所示：

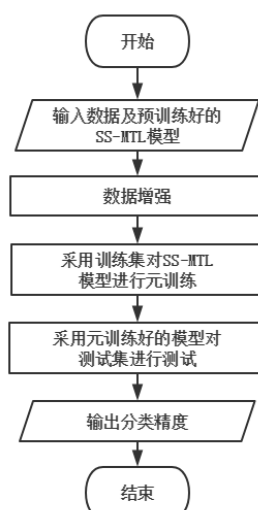


图3.12 手写字符识别流程图

表3.7算法在增强的 Omniglot 数据集上性能对比表

算法	20-way-1-shot(%)	20-way-5-shot(%)
MAML ^[37] (2017)	95.8±0.3	98.9±0.2
MAML++ ^[38] (2019)	97.65±0.05	98.9±0.2
TADAM ^[40] (2018)	98.16±0.12	99.45±0.09
MTL ^[26] (2019)	97.73±0.24	99.34±0.06
SS-MTL	98.42±0.13	99.63±0.07

由于手写字符图片是灰度图且图像特征相对于传统图片比较简单,在传统分类问题中其精度都能达到 **100%**。在小样本场景下的分类精度虽然很高(都能达到 95%以上),但是与传统分类精度来比,这样的性能还是相形见绌。

与最新的小样本学习算法相比,如表 3.7 所示,SS-MTL 算法在两种实验设置下的性能均高于其他算法,证明了本文提出的 SS-MTL 算法对于手写字符识别任务的有效性,更体现了 SS-MTL 算法对于其他算法的优越性。

为了对比小样本学习算法针对具有不同复杂度特征的图像数据的性能,将本文提出的小样本学习算法 SS-MTL 应用于具有复杂特征的图像数据(鸟类)和具有简单特征的图像数据(手写字符)可以发现,目前小样本学习算法虽然对于具有简单特征的图像数据取得了非常可观的性能,但是对于具有复杂特征的图像数据性能仍有待于进一步提升。

3.6 本章小结

本章首先介绍了目前最新的小样本学习算法所存在的缺陷, 然后针对这些缺陷, 本文进行改进, 提出了 SS-MTL 算法。在算法方面, 本文提出了基于自监督辅助任务的预训练方式, 通过挖掘图像自身的结构信息以提高算法精度, 通过添加 l_2 正则化项的损失函数对网络进行训练, 缓解了算法的过拟合问题; 在模型方面, 针对 Resnet12 网络模型, 提出了一种双路膨胀卷积网络, 通过膨胀卷积增大模型的感受视野, 获取更大尺寸的特征, 同时添加残差注意力机制增强了有意义的特征, 提高了算法的精度。最后将本文提出的算法与目前最新的小样本学习算法在基准数据集 miniImageNet 以及 FC100 上进行对比实验, 并在真实的鸟类数据集 CUB-200-2011 以及手写字符数据集 Omniglot 上对本文提出的 SS-MTL 算法进行鸟类识别与手写字符识别的应用研究, 充分证明了本文所提出策略的有效性以及本文算法相对于目前比较优秀的算法的优越性。

第四章 IFML-FSRE: 信息融合的元学习小样本关系抽取算法

本章首先对目前小样本学习算法应用到关系抽取领域的难点以及缺陷进行了阐述,之后对本文的解决方案,即本章所提出的一种新的高效信息融合的元学习小样本关系抽取算法 IFML-FSRE(**I**nformation **F**used **M**eta-**l**earning for **F**ew-**s**hot **R**elation **E**xtraction, IFML-FSRE)进行了详细论述。最后,在基准测试数据集 FewRel 上对本文算法及相关最新最好算法进行性能测试。

4.1 算法设计动机

本文第三章提出的小样本学习算法 SS-MTL 是基于元学习与迁移学习的混合型算法,算法较为复杂,不适应于小样本的关系抽取问题。且由于文本不能直接计算,目前大部分小样本学习算法难以直接应用于关系抽取领域。为了保证小样本学习算法应用到关系抽取任务中的效率与准确性,所以本文选取较为简单的元学习算法进行小样本学习算法在关系抽取领域的应用研究,并提出了一种新的有效的小样本关系抽取算法。

关系抽取是自然语言处理领域的重要任务,旨在根据两个给定实体的相关上下文对它们之间的语义关系进行分类。然而,小样本关系抽取任务仍然面临着以下问题:(1)关系抽取任务中的数据为文本数据,为了提升小样本学习算法应用到关系抽取领域的性能,如何有效利用文本数据的语义信息是小样本关系抽取算法的难点。(2)简单的小样本学习算法——元学习 MAML 框架学习如何进行参数的初始化和在有限实例下进行参数的快速适应。但是 MAML 算法基于统计信息进行模型参数的适应,即对于参数适应而言,每个实例共享相同的学习率,参数适应效率低,算法精度有待于进一步提高。

针对以上问题,本文基于经典的小样本学习框架 MAML 提出一种信息融合的元学习小样本关系抽取算法 IFML-FSRE(**I**nformation **F**used **M**eta-**l**earning for **F**ew-**s**hot **R**elation **E**xtraction, IFML-FSRE)。具体而言,本文提出一种基于匹配信息的实例重表示方法,利用句子实例之间的匹配度以增强对查询集(测试样本)有益的支持实例(训练样本),并基于实例间的相似度进行任务过滤,从而提高算法精度;本文提出一种融合类名信息与实例信息的参数快速适应方法,充分利用文本数据的语义信息以提高有益实例的学习率,从而更好地对模型参数进行快速适应。

4.2 算法设计思想及框架

针对元学习算法应用到关系抽取领域面临的问题(1)，本章设计了一种基于匹配信息的实例重表示方法，增强了对查询实例有益的支持实例的权重，小样本分类问题的实质是需要模型首先学习支持集上的分类器然后处理查询集上的分类任务，通过突出对查询集有益的支持实例，能增强查询集在模型上的性能；同时本文提出一种任务过滤机制对有效任务进行筛选以提高算法精度；针对上述问题(2)，本文提出一种实例信息与类名信息融合的参数快速适应方法，利用类名与句子实例的语义信息，为每个实例计算相应的学习率，从而更好地对模型参数进行快速适应。基于以上这些思想，本文针对小样本关系抽取任务，提出了一种新的有效的信息融合的元学习小样本关系抽取算法 IFML-FSRE(Information Fused Meta-learning for Few-shot Relation Extraction, IFML-FSRE)。IFML-FSRE 的算法框架如图 4.1 所示：

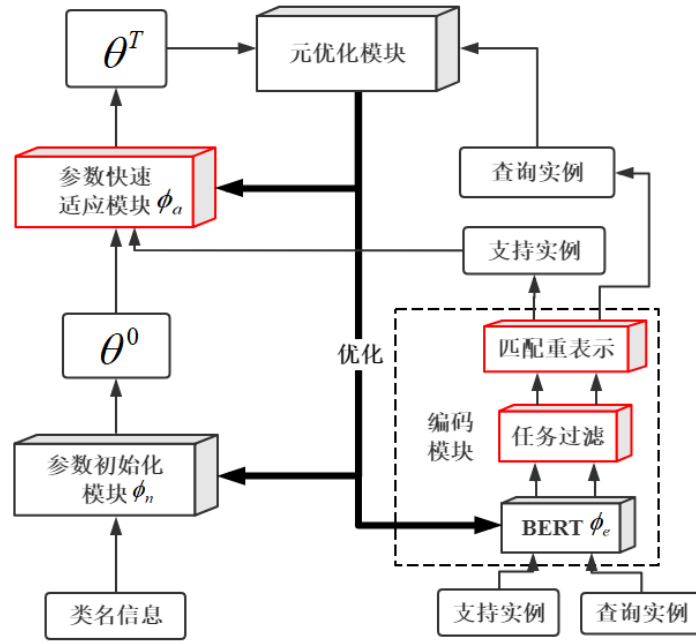


图4.1 IFML-FSRE 算法框架示意图(红色部分为本文所提出的机制)

IFML-FSRE 算法主要由实例编码模块，参数初始化模块，参数快速适应模块，及元优化模块四个部分组成，其中每个部分的功能如下：

实例编码模块：将句子实例编码为向量。为了充分利用文本数据的语义信息，本文在这一模块中采用任务过滤策略对每个任务中采用 BERT 编码后的实例进行计算，以筛选有效任务。然后本文采用一种基于匹配度的重表示方法，突出了对查询实例有益的支持实例。

参数初始化模块：带有元参数 ϕ_n 的参数初始化模块根据类名信息动态地初始化

与类相关的参数 θ^0 。

参数快速适应模块: 针对 MAML 框架参数适应过程中共享学习率导致参数适应效率低的问题, 本文提出一种融合了支持实例与类名信息的模型参数快速适应方法, 带有元参数 ϕ_a 的参数快速适应模块对初始化参数 θ^0 进行快速适应, 得到适应好的模型参数 θ^T 。

元优化模块: 通过查询集在适应后的模型上的性能对元参数 $\Phi = \{\phi_e, \phi_n, \phi_a\}$ 进行更新。

本章提出的 IFML-FSRE 算法旨在基于改进的元学习框架 MAML 下进行模型参数的快速适应以进行小样本关系抽取任务。本文将在下一节对本文提出的算法的主要设计策略进行详细阐述。

4.3 算法主要设计策略

本节主要针对 IFML-FSRE 算法中的主要设计策略进行详细的介绍。

4.3.1 实例编码模块的设计

针对元学习对文本数据的语义信息难以有效利用的问题, 在实例编码模块本文设计了一种基于相似度的任务过滤策略, 对有效任务进行筛选; 然后通过一种基于匹配度的重表示方法, 增强了对查询实例有益的支持实例的权重以提高查询集在模型上的性能, 具体来说, 本文通过计算查询实例与支持实例的匹配度, 利用注意力机制为每个实例分配权重。

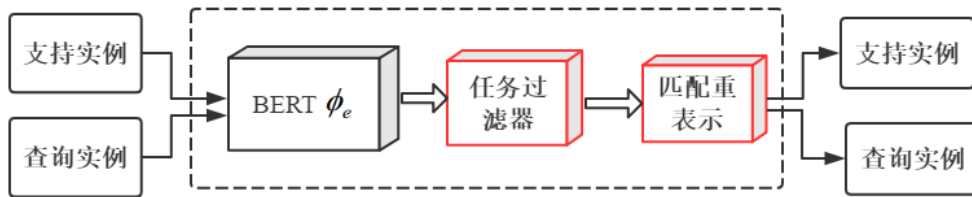


图4.2 实例编码模块示意图(红色框为本文提出的策略)

实例编码模块的具体过程如图 4.2 所示, 给定一个任务, 由于 BERT 对于 NLP 任务的有效性, 本文首先采用 BERT 对任务中的每个实例进行编码(BERT 模型具体结构可见本文 2.2.2 节), 然后添加了任务过滤策略对任务进行过滤, 针对过滤后的任务中的实例进行基于匹配信息的重表示。本节将对这三个部分的具体操作进行详细介绍。

(1) BERT 编码

具体来说, 在 BERT 编码阶段, 给定一个句子和相应的目标实体对(即头实体 h

和尾实体 t)，采用 BERT 对实例进行编码，BERT 编码器的形式化表示如公式^[57](4-1)所示：

$$x_j = g(x_j, h, t; \phi_e) \quad (4-1)$$

其中， x_j 表示一个句子， h 和 t 分别表示句子中的头实体和尾实体， $g(\cdot)$ 是编码器， ϕ_e 是编码器的参数， $x_j \in R^{d_s}$ 是句子实例的编码， d_s 为编码长度。

(2) 任务过滤器

目前大部分小样本关系抽取算法采用注意力机制突出重要实例，然而一个任务中存在相似实例过多会严重影响模型的训练，因此，选择含有丰富实例的任务进行训练能够有效增强模型性能。

针对上述问题，本文提出一种基于相似度的任务过滤策略，具体来说，对于小样本关系抽取中的 N-way-K-shot(即抽取 N 个类，每类抽取 K 个样本)的设置，一个任务中的支持集共包含 $N*K$ 个实例，如果支持实例中相似实例的个数大于 $N*K^2$ 个(包含实例与自己本身相似)，则说明这个任务是无效的。具体地，本文采取余弦相似度对两个实例相似度进行衡量，如公式(4-2)所示：

$$sim_{ij} = \cos(x_i, x_j) = \frac{x_i \bullet x_j}{\|x_i\| \times \|x_j\|} \quad (4-2)$$

其中， x_i 与 x_j 分别表示支持集中两个不同的实例， $sim_{ij} \in (0,1)$ 表示这两个实例间的余弦相似度，当 sim_{ij} 大于阈值 0.5，相似实例个数大于阈值 $N*K^2+K$ 时，则表示相似实例过多，这个任务是失败的，将这个任务舍弃。同时，这种任务过滤机制使得模型尽可能选择不相似的实例进行训练，进一步增强了模型的性能。

(3) 基于匹配度的实例重表示

小样本分类问题的实质是需要模型首先学习支持集(包含少量训练实例)上的分类器然后处理查询集(测试实例)上的分类任务，因此选择对查询实例有益的支持实例能有效提高查询集在分类器上的性能，为了在一个任务中选出对查询实例有益的支持实例，本文提出了一种基于匹配信息的重表示方法，具体方法如公式(4-3)所示：

$$\begin{cases} \alpha_{mn} = q_m^T s_n \\ \tilde{s}_n = \sum_{m=1}^{T_q} \frac{\exp(\alpha_{mn})}{\sum_{m'=1}^{T_q} \exp(\alpha_{m'n})} q_m \\ \tilde{q}_m = \sum_{n=1}^{T_s} \frac{\exp(\alpha_{mn})}{\sum_{n'=1}^{T_s} \exp(\alpha_{mn'})} s_n \end{cases} \quad (4-3)$$

公式(4-3)描述了每个支持实例与查询实例的重表示方法, 其中, α_{mn} 表示第 n 个支持集 s_n 与第 m 个查询集 q_m 之间的匹配度, $m \in \{1, \dots, T_q\}$, $n \in \{1, \dots, T_s\}$, T_q 及 T_s 分别表示查询实例与支持实例的个数, \tilde{q}_m 和 q_m 表示重表示后的查询集 \tilde{Q} 以及原始查询集 Q 的第 m 行, \tilde{s}_n 和 s_n 表示重表示后的支持集 \tilde{S} 以及原始支持集 S 的第 n 行。本文利用注意力机制通过获得的查询实例与支持实例的匹配度分别计算出每个实例的权重, 从而获得含有注意力信息的支持实例与查询实例的重表示。基于匹配信息的查询实例与支持实例的重表示进一步增强了对查询集有益的支持实例权重, 由于 FSL 场景中缺少支持数据, 因此选择对查询集有益的支持实例能够有效提升算法的性能。

4.3.2 参数初始化模块的设计

从类分布 $p(C)$ 中抽取一组类 $\{C_1, C_2, \dots, C_N\}$ 构建一个任务, MAML 算法学习与类无关的初始化, 这种初始化可以通过少量梯度更新步骤而适应新类。受 MIML^[57] 启发, 本文通过参数初始化模块以生成方式利用类名信息进行类感知的参数初始化。具体过程如图 4.3 所示:

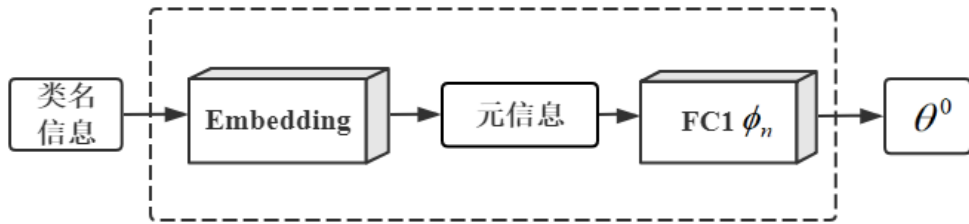


图4.3 参数初始化模块示意图

参数初始化模块使用从语义概念派生的元信息来初始化每个类的参数, 本文将类名用作元信息。具体而言, 如图 4.3 所示, 给定类 C_i 的名称, 通过词嵌入(Embedding)模块得到该类名信息的词嵌入作为元信息 $c_i \in R^{dw}$, 然后, 通过参数初始化模块按公式^[57](4-4)所示初始化该类的参数:

$$\theta_i^0 = \Psi(c_i; \phi_n) \quad (4-4)$$

其中, θ_i^0 是类 C_i 的类感知的初始化参数, $\Psi(\cdot)$ 是参数初始化器, 通过一个全连接层 FC1(Fully Connect)实现, ϕ_n 是相应的参数。初始化参数 θ_i^0 可以用来计算实例的分类得分, 如公式^[57](4-5)所示:

$$s_{i,j} = \theta_i^{0T} x_j \quad (4-5)$$

其中, $s_{i,j}$ 是实例 x_j 属于类 c_i 的得分, 快速初始化后的模型可以表示为 $f_{\theta^0, \{\phi_c, \phi_n\}}$, 其中 $\theta^0 = \{\theta_1^0, \theta_2^0, \dots, \theta_N^0\}$ 表示所有类的初始化参数。

4.3.3 参数快速适应模块的设计

针对 MAML 算法所有实例共享学习率进行参数适应效率低的问题, 本文提出一种信息融合的参数快速适应方法。具体而言, 本文提出了一种基于类名与实例信息融合类向量的计算方法, 通过计算实例与类向量的匹配度得到实例属于该类的得分, 根据得分利用选择性注意力机制为每个实例分配不同的学习率, 为含有重要信息的实例分配较高的学习率从而更好地进行参数适应。

本文根据支持集上实例的分类性能, 通过梯度下降步骤来自适应初始化参数 θ^0 , 具体过程如图 4.4 所示:

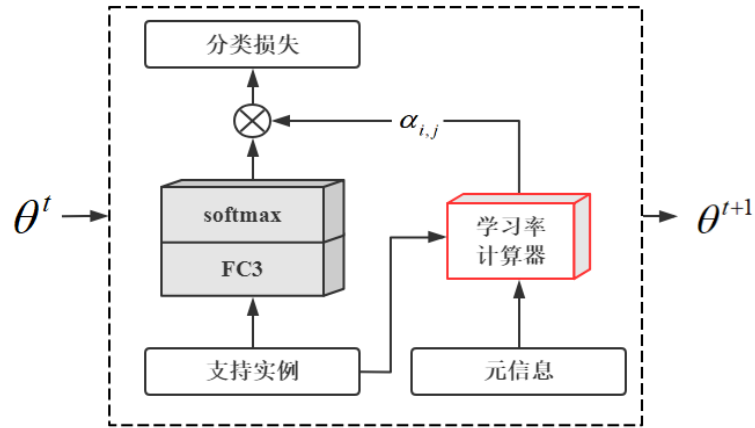


图4.4 参数快速适应模块示意图(红色部分为动态的学习率计算器)

如图 4.4 所示为一个时间步内的参数更新机制, 在快速适应模块, 模型参数根据给定的 T 个时间步内进行动态地迭代, 在每个时间步 t , 参数 θ^t 通过梯度下降的方式进行更新, 如公式(4-6)^[57]所示:

$$\theta^{t+1} = \theta^t - \sum_{i,j} \alpha_{i,j} \nabla_{\theta^t} L(f_{\theta^t, \{\phi_e, \phi_n\}}, x_j, y_j) \quad (4-6)$$

其中, L 表示通过模型 $f_{\theta^t, \{\phi_e, \phi_n\}}$ 计算的支持实例 (x_j, y_j) 的交叉熵损失, 该损失由支持实例经过图 4.4 中的全连接层 FC3(Fully Connect)以及 softmax 层而得到, x_j 表示一个句子实例, y_j 表示该句子的类别标签, $\alpha_{i,j}$ 是支持实例 (x_j, y_j) 在 θ_i 上的学习率, L 的具体计算方法如公式(4-7)所示:

$$L_{cross-entropy} = - \sum_{c=1}^N y_c \log(p_c) \quad (4-7)$$

其中, N 代表类别的数量, y_c 表示变量 0 或 1, 如果该类别和样本的类别相同则为 1, 否则为 0, p_c 是对于观测样本的预测概率。

为了解决 MAML 框架参数适应效率低的问题, 不同于 MAML 算法假设支持集中的所有实例同等重要, 从而共享相同的学习率以进行参数调整。考虑到实际情况文本实例通常是多种多样的, 并且有益实例可以在快速适应中主导模型参数, 本文分别对每个实例赋予不同的学习率 $\alpha_{i,j}$ 进行适应, 每个实例的学习率的计算方法如图 4.5 所示:

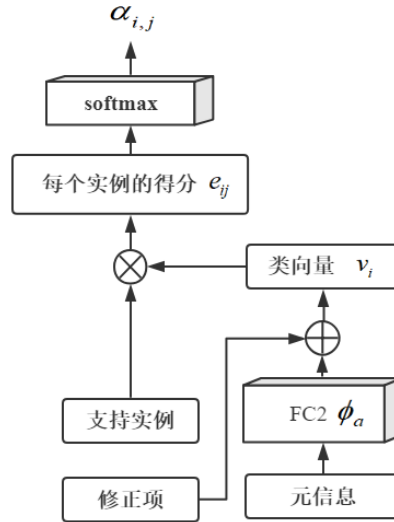


图4.5 学习率计算器设计示意图

如图 4.5 所示, 首先, 本文提出了一种融合了从类名信息得到元信息与实例信息的类向量计算方法, 通过计算每个支持实例相对于类向量的匹配度作为支持实例属于该类的得分, 然后通过选择性注意力机制动态确定每个实例的学习率。类向量的计算方法如公式(4-8)所示:

$$v_i = \lambda \Psi(c_i, \phi_a) + (1 - \lambda) \sum_{j=1}^K x_j / K \quad (4-8)$$

其中, v_i 是类 C_i 的类向量表示, 其中 $\Psi(\cdot)$ 由一个具有参数 ϕ_a 的全连接层所实现, c_i 是第 i 类的类名重表示, 即元信息, $c_i \in R^{d_v}$ 是通过类名的词嵌入所获得, 嵌入空间向量长度为 d_w 。K 表示在一个任务中每个类的支持实例个数, x_j 表示支持实例, $j \in \{1, \dots, K\}$ 。其中, 为了克服类名信息的单一性可能导致的偏差, 通过实例均值 $\sum_{j=1}^K x_j / K$ 对类名信息在嵌入空间进行修正, λ 为修正系数, 这里选择为 0.7。

然后, 本文采用实例 x_j 与类向量 v_i 的匹配度作为支持实例 x_j 属于类 C_i 的得分 e_{ij} , 如公式(4-9)所示:

$$e_{ij} = v_i^T x_j \quad (4-9)$$

其中, v_i 是类 C_i 的类向量表示, 如果 x_j 是类 C_i 的有益实例, 则 e_{ij} 应该较大, 因此 x_j 应该对参数的适应做出更大的贡献。为了选择有益实例以在模型中快速适应, 每个实例的学习率由选择性注意力机制动态确定, 如公式(4-10)所示:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_j \exp(e_{ij})} \quad (4-10)$$

其中, e_{ij} 表示支持实例 x_j 属于类 C_i 的得分, 本文通过分别对每个实例针对不同的类别的得分赋予不同的学习率来进行参数适应, 经过 T 步自适应后的参数表示为 θ^T 。

针对 MAML 框架参数适应过程共享学习率导致效率低的问题, 本模块设计了一种类名信息与实例信息相融合的类向量计算方法, 从而得到每个支持实例的学习率, 有效突出了每个类中的有益实例, 实现了模型参数的快速适应。

4.3.4 元优化模块的设计

元优化模块的模型图如图 4.6 所示:

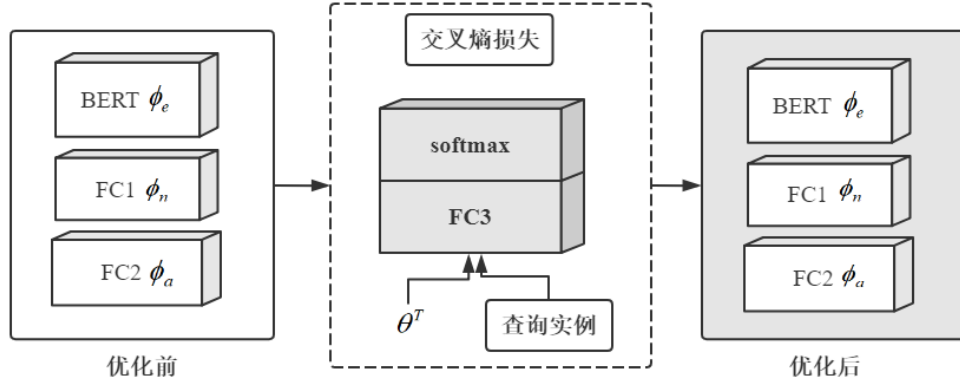


图4.6 元优化模块示意图

如图 4.6 所示，在对支持实例的快速适应后，元参数 $\Phi = \{\phi_e, \phi_n, \phi_a\}$ 根据查询集 Q 在适应后的模型上的性能进行优化，如公式^[57](4-11)所示：

$$\Phi = \Phi - \beta \nabla_{\Phi} L(f_{\theta^T, \Phi}, x_j, y_j) \quad (4-11)$$

其中， β 是元参数的学习率。 L 表示查询集数据在训练好的模型上的损失函数。这样，IFML-FSRE 算法可以学习到可以有效地为每个实例进行编码以及为每个类初始化参数的元参数，并选择有效支持实例以进行快速适应，从而在查询集上产生良好的分类结果。

4.4 算法实现

为了更详细地介绍 IFML-FSRE 算法的具体过程，本文给出了 IFML-FSRE 算法的伪码，如图 4.7 所示：

算法 3 IFML-FSRE	
输入: 类分布 $p(C)$, 元学习率 β	
输出: 更新后的模型参数	
1:	随机初始化元参数: $\Phi = \{\phi_e, \phi_a, \phi_n\}$
2:	while not done do
3:	从类分布采样一组类 $C_i \sim p(C)$
4:	采样支持集 S 及查询集 Q 构成一个任务并通过公式(4-1)对所有实例进行编码
5:	通过公式(4-2)计算支持集 S 中相似实例的个数 Num_{sim}
6:	if $Num_{sim} > N * K^2 + K$ then
7:	将这个任务过滤掉
8:	end if
9:	采用公式(4-3)对过滤后的查询实例及支持实例进行重表示
10:	for all C_i do

```

11:      通过公式(4-4)对类  $C_i$  的参数进行初始化
12:      end for
13:      for  $t = 1, \dots, T$  do
14:          使用支持实例集  $S$  计算梯度和学习速率以实现快速适应
15:          通过公式(4-6)利用梯度下降法计算自适应参数
16:      end for
17:      通过公式(4-11)使用查询集  $Q$  进行元优化
18:  end

```

图4.7 IFML-FSRE 算法伪码示意图

在图 4.7 中, 输入: 类分布 $p(C)$, 元学习率 β ; 第 1 行: 随机初始化元参数, 第 2-18 行表示整个算法的迭代过程, 其中第 3 行表示从数据分布中采样一个批次的类别, 第 4-9 行表示实例编码阶段, 其中第 4 行表示从这个批次中构建查询集 Q 与支持集 S 并采用 BERT 对每个实例进行编码, 第 5-8 行表示任务过滤策略对每一个任务进行判断, 如果这个任务所含相似实例过多, 就将这个任务过滤掉, 第 9 行对未被过滤的任务中的每个实例进行基于匹配信息的重表示; 第 10-12 行表示对每个类进行基于元信息的初始化阶段, 在这个阶段, 类名作为元信息引导针对于每个类的参数的初始化; 第 13-16 行表示类名信息与实例信息融合的参数快速适应阶段, 在这一阶段, 针对每个任务通过实例修正的类向量计算每个支持实例相对应的学习率与梯度, 采用公式(4-6)来计算适应后的模型参数; 第 17 行表示元优化过程, 采用公式(4-11)通过查询集在适应后的模型上的性能对元参数进行优化。

4.5 算法实验与分析

本节对所提出的 IFML-FSRE 算法进行了大量的实验验证, 本节主要介绍了实验平台, 实验数据, 实验设置及参数设置, 对比算法以及实验结果及分析。

4.5.1 实验环境及设置

为了对比的公平性, 本文所有实验均在 Linux 环境下运行, 采用 Intel(R) Xeon(R) Gold 5115 CPU @ 2.40GHz 的 CPU 以及 NVIDIA Tesla P40 的 GPU。本文所提出的 IFML-FSRE 算法使用 Python 3.6 以及 Pytorch 1.0 实现。

本文在 FewRel 数据集上对算法性能进行了实验验证, FewRel 数据集是由清华大学自然语言处理实验室所提出的用于小样本关系抽取的数据集。FewRel 中包含 100 个关系类别, 每个关系分别有 700 个实例。其中训练集包含 64 个关系类别, 验证集包含 16 个关系类别, 测试集包含 20 个关系类别。

本文设计了 4 种类型的小样本关系分类实验设置, 分别为 5-way-1-shot, 5-way-5shot, 10-way-1shot 以及 10-way-5-shot, N-way-K-shot 设置表示每个测试批次具有 N

个未出现在训练集中的类别, 每个类别有 K 个支持实例。本文采用查询实例的分类准确率 Acc 作为评估指标, 具体计算方法如公式(4-12)所示:

$$Acc = \frac{\text{Number of correct prediction s}}{\text{Total number of prediction s}} * 100\% \quad (4-12)$$

4.5.2 实验参数设置

本文采用 Adam 优化器对元参数进行优化, 下面对本文的一些超参数进行介绍:

表4.1 IFML-FSRE 算法超参数表

参数名称	参数意义	参数值
$\beta - \phi_n$	参数初始化模块学习率	1
$\beta - \phi_a$	类向量编码模块学习率	1
$\beta - \phi_e$	实例编码模块学习率	5e-5
d_s	实例编码长度	1536
d_w	元信息编码长度	50
T	适应步骤数	150

4.5.3 对比算法介绍

本文将 IFML-FSRE 算法与以下几个具有强大性能的算法进行比较:

(1) MAML 算法: MAML 是经典的与模型无关的小样本学习算法, MIML 算法与本文提出的 IFML-FSRE 算法均基于 MAML 算法的思想进行改进。

(2) MLMAN 算法: MLMAN 算法通过多级匹配与聚合的方法获得类原型, 是目前较为优秀的基于原型网络的方法。

(3) HATT-Proto 算法: HATT-Proto 是目前最为优秀的基于原型网络的算法, 通过实例与特征级别的双重注意力机制显著提高算法精度和模型鲁棒性。

(4) MIML 算法: MIML 算法基于 MAML 算法框架, 提出了元信息引导的快速初始化和适应的方法, 且取得了目前最具有竞争力的性能。

其中, MAML 是 2017 年提出的非常经典的小样本学习算法, MLMAN 以及 HATT-Proto 是 2019 年比较优秀的算法, MIML 是 2020 年最新的算法, 是本文重点对比算法。

4.5.4 算法性能比较及分析

(1) 算法分类精度比较

本节对所提出的 IFML-FSRE 算法在精度方面与 4.5.3 节中算法进行比较, 其中 N-K 表示 N-way-K-shot 实验设置, 实验结果如表 4.2 所示。

表 4.2 为各个算法在 5-way-1-shot, 5-way-5shot, 10-way-1shot 以及 10-way-5-shot 这四种实验设置下所取得的分类精度, IFML-FSRE 算法均取得了优于目前最新算法的效果。本文提出的算法突出了所有实例中的有益实例, 因此本算法在小样本情境下也能通过有益实例主导模型参数快速适应, 从而取得更好的效果。因此, 在这四种实验设置下, 本算法的效果更为显著。

表4.2 IFML-FSRE 算法与其他对比算法在 FewRel 上的精度对比表

算法	5-1(%)	5-5(%)	10-1(%)	10-5(%)
MAML ^[35] (2017)	80.32±0.11	90.39±0.13	73.46±0.23	84.25±0.18
MLMAN ^[55] (2019)	82.98±0.20	92.66±0.09	75.59±0.27	87.29±0.19
HATT-Proto ^[52] (2019)	81.50±0.09	90.12±0.04	74.28±0.09	83.05±0.05
MIML ^[57] (2020)	84.55±0.12	92.96±0.17	84.91±0.13	87.02±0.12
IFML-FSRE(ours)	87.40±0.15	94.72±0.08	87.28±0.21	89.13±0.11

(2) 针对存在噪声情况下的算法精度比较

由于文本的多样性和复杂性, 现实中的文本数据大多是含有噪声的。为了研究本文提出的 IFML-FSRE 算法在存在噪声的情况下的分类精度, 本文分别将 0%, 10%, 20%, 30%的支持实例替换为从 FewRel 数据集中不同关系中随机采样的噪声实例, 进行 5-way-5-shot 以及 10-way-5-shot 实验(在表中以 5-5, 及 10-5 进行表示), 并与所有对比算法进行比较, 实验结果如表 4.3 所示:

表4.3 各个算法在不同噪声比例下的分类精度

算法	噪声率	5-5(%)	10-5(%)	噪声率	5-5(%)	10-5(%)
MAML ^[35] (2017)	0%	90.39±0.13	84.25±0.18	10%	88.60±0.12	82.78±0.13
HATT-Proto ^[52] (2019)		90.12±0.04	83.05±0.05		88.74±0.06	81.09±0.08
MIML ^[57] (2020)		92.96±0.17	87.02±0.12		91.89±0.08	84.93±0.22
IFML-FSRE(ours)		94.72±0.08	89.13±0.11		93.86±0.12	87.36±0.18
MAML ^[35] (2017)	20%	86.22±0.10	80.24±0.13	30%	83.96±0.20	77.86±0.11
HATT-Proto ^[52] (2019)		87.04±0.13	78.46±0.18		84.71±0.07	75.68±0.11

MIML ^[57] (2020)		90.04 \pm 0.10	82.81 \pm 0.23		88.78 \pm 0.18	80.13 \pm 0.27
IFML-FSRE(ours)		92.35\pm0.12	84.93\pm0.16		90.03\pm0.21	82.55\pm0.20

由表 4.3 中数据可知, MAML 算法的性能会随着噪声比例的增加而显著降低, 因为它在参数快速适应过程中会严重受到噪声实例的影响。Prot-HATT 算法采用基于原型网络的方法, 原型网络使用所有实例的平均值构建原型, 显示出较小的性能下降。相比于基于原型网络的方法, 基于优化的元学习方法在处理噪声实例时性能下降明显, 从中可以得出基于优化的元学习方法在处理噪声实例时的脆弱性。而 IFML-FSRE 显示出了与原型网络相当的抗噪声性能, 分类精度下降较少, 充分说明本文提出的 MAML 框架下的参数快速适应方法能有效在含有噪声的数据中突出重要的有益实例并实现在存在噪声的情况下取得更强大的性能。

4.6 本章小结

本章首先介绍了目前小样本学习算法应用到关系抽取领域所面临的难点, 然后针对这些问题, 本文提出了 IFML-FSRE 算法。首先为了有效利用文本数据的语义信息, 本文提出一种基于匹配信息的实例重表示方法, 利用句子实例之间的匹配度以增强对查询集(测试样本)有益的支持实例(训练样本), 并基于实例间的相似度进行任务过滤, 从而提高算法精度; 针对简单的元学习算法 MAML 参数适应效率低的问题, 本文提出一种融合类名信息与实例信息的参数快速适应方法, 充分利用文本数据的语义信息以更好的进行参数适应。本文将所提出的算法与目前最新的小样本关系抽取算法在 FewRel 数据集上进行了大量的实验, 充分验证了本文所提出的算法优越性。

第五章 总结与展望

5.1 论文工作总结

随着深度学习的发展,通过少量有标签样本进行学习是深度学习亟待解决的关键问题,小样本学习以及小样本关系抽取是目前深度学习非常热门的研究领域。然而目前最新的小样本学习算法还存在过拟合,精度低等问题,小样本关系抽取算法深受数据噪声的困扰,针对这些关键问题,本文开展了以下的研究工作:

首先,本文对小样本学习算法及小样本关系抽取算法的发展与研究现状进行深入研究,并对目前算法的优点与不足进行了详细的分析。

然后本文针对目前最新的小样本学习算法所存在的精度低和过拟合等问题,设计了一种新的小样本学习算法 **SS-MTL**。在算法方面,本文提出了一种基于自监督辅助任务的预训练方法,通过自监督任务挖掘数据内部的结构信息,进一步提高了算法的精度。针对小样本学习算法存在的过拟合问题,提出了添加 l_2 正则化项的损失函数,使得模型倾向于选择数值更小的参数,从而能适应更多的任务。在模型方面,本文基于非常优秀的 **Resnet12** 网络模型进行改进,提出了一种双路膨胀卷积网络,通过膨胀卷积增大了模型的感受视野,双分支的设计使模型能提取不同尺寸的特征;此外,本文提出一种残差注意力机制来增强有意义的特征,而抑制无意义的特征,进一步增强了模型的性能。同时本文基于通用的数据集对本文提出的算法进行实验,并将本文提出的小样本学习算法进行鸟类识别以及手写字符识别的应用研究,实验结果进一步表明本文提出的小样本学习算法的有效性与优越性。

由于文本不能直接计算,且目前大部分小样本学习算法及本文提出的小样本学习算法 **SS-MTL** 不能直接应用于关系抽取任务,所以本文选取简单的 **MAML** 元学习算法应用到关系抽取任务中。并针对其面临的难点以及存在的缺陷,本文提出了一种新的信息融合的小样本关系抽取 **IFML-FSRE**。针对小样本学习算法无法有效利用文本数据的语义信息的问题,本文提出一种基于匹配度的实例重表示方法,突出了对测试样本有益的训练样本,并添加任务过滤策略,有效提高了算法精度;针对 **MAML** 共享学习率导致参数适应效率低的问题,本文提出一种类名与实例信息融合的模型参数快速适应方法,为每个实例计算不同的学习率,从而更好地对模型参数进行快速适应。实验证明,本文提出的算法在精度和对噪声的鲁棒性方面优于最新的算法。

5.2 后续工作展望

小样本学习算法以及小样本关系抽取算法都是目前深度学习领域热门研究问题。

如何通过少量的有标签数据进行学习是小样本学习的关键,本文针对现有的小样本学习算法以及小样本关系抽取算法进行深入研究,并取得了一定的成果,完成了预期内容,但是与本课题相关的以下内容还有待于进一步的研究:

(1) 由于文本不能直接计算,目前大部分小样本学习算法及本文提出的小样本学习算法 SS-MTL 不能直接应用于关系抽取任务,本文后续工作是通过对本文提出的小样本学习算法进行改进,将其应用于关系抽取领域。

(2) 小样本关系抽取问题最初来源于关系抽取遭受长尾数据的困扰,因此,使用真实的符合长尾分布的数据集对本文所提出的及目前最新的小样本关系抽取算法进行进一步的验证是本文下一步的研究目标。

参考文献

- [1] Xiong W, He Y, Zhang Y, et al. Fine-grained image-to-image transformation towards visual recognition[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 5840-5849.
- [2] Wang D, Chen J. Supervised Speech Separation Based on Deep Learning: An Overview[J]. IEEE Transactions on Audio, Speech, and Language Processing, 2018, 26(10): 1702-1726.
- [3] Zhou M, Duan N, Liu S, et al. Progress in neural NLP: modeling, learning, and reasoning[J]. Engineering, 2020, 6(3): 275-290.
- [4] Litjens G, Kooi T, Bejnordi B E, et al. A survey on deep learning in medical image analysis[J]. Medical image analysis, 2017, 42: 60-88.
- [5] Huang X, Kwiatkowska M, Wang S, et al. Safety verification of deep neural networks[C]. International conference on computer aided verification. Springer, Cham, 2017: 3-29.
- [6] Fei-Fei L, Fergus R, Perona P. One-shot learning of object categories[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(4): 594-611.
- [7] Elsken T, Staffler B, Metzen J H, et al. Meta-learning of neural architectures for few-shot learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12365-12375.
- [8] Lake B M, Salakhutdinov R, Tenenbaum J B. Human-level concept learning through probabilistic program induction[J]. Science, 2015, 350(6266): 1332-1338.
- [9] Wang Y, Yao Q, Kwok J T, et al. Generalizing from a few examples: A survey on few-shot learning[J]. ACM Computing Surveys (CSUR), 2020, 53(3): 1-34.
- [10] Trisedya B D, Weikum G , Qi J , et al. Neural Relation Extraction for Knowledge Base Enrichment[C]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 229-240.
- [11] Zelenko D, Aone C, Richardella A. Kernel methods for relation extraction[J]. Journal of machine learning research, 2003, 3(Feb): 1083-1106.
- [12] Bach N, Badaskar S. A review of relation extraction[J]. Literature review for Language and Statistics II, 2007, 2: 1-15.
- [13] Gormley, M. R, Yu, M and Dredze, M. Improved relation extraction with feature-rich compositional embedding models. In Proceedings of EMNLP, 2015:1774–1784.
- [14] Santoro A, Bartunov S, Botvinick M, et al. Meta-learning with memory-augmented neural networks[C]. International conference on machine learning. PMLR, 2016: 1842-1850.

-
- [15] Qi H, Brown M, Lowe D G. Low-shot learning with imprinted weights[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 5822-5830.
 - [16] Shyam P, Gupta S, Dukkipati A. Attentive recurrent comparators[C]. International conference on machine learning. PMLR, 2017: 3173-3181.
 - [17] Lake B M, Salakhutdinov R, Tenenbaum J B. Human-level concept learning through probabilistic program induction[J]. Science, 2015, 350(6266): 1332-1338.
 - [18] Kozerawski J, Turk M. Clear: Cumulative learning for one-shot one-class image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3446-3455.
 - [19] Zhang Y, Tang H, Jia K. Fine-grained visual categorization using meta-learning optimization with sample selection of auxiliary data[C]. Proceedings of the european conference on computer vision (ECCV), 2018: 233-248.
 - [20] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[C]. In Advances in Neural Information Processing Systems, 2016: 3630–3638.
 - [21] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J]. Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
 - [22] Antoniou A, Storkey A, Edwards H. Data augmentation generative adversarial networks[J]. arXiv preprint arXiv:1711.04340, 2017.
 - [23] Zhang R, Che T, Ghahramani Z, et al. MetaGAN: An Adversarial Approach to Few-Shot Learning[J]. Conference and Workshop on Neural Information Processing Systems, 2018: 2371-2380.
 - [24] Schwartz E, Karlinsky L, Shtok J, et al. Delta-encoder: an effective sample synthesis method for few-shot object recognition[J]. Conference and Workshop on Neural Information Processing Systems, 2018: 2850-2860.
 - [25] Chen W Y, Liu Y C, Kira Z, et al. A closer look at few-shot classification[J]. arXiv preprint arXiv:1904.04232, 2019.
 - [26] Keshari R, Vatsa M, Singh R, et al. Learning structure and strength of CNN filters for small sample size training[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 9349-9358.
 - [27] Huang J, Rathod V, Sun C, et al. Speed/accuracy trade-offs for modern convolutional object detectors[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 7310-7311.
 - [28] Zhu R, Zhang S, Wang X, et al. ScratchDet: Training single-shot object detectors from scratch[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019:

- 2268-2277.
- [29] Sun Q, Ma L, Oh S J, et al. Natural and effective obfuscation by head inpainting[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 5050-5059.
 - [30] Liu X, Deng Z, Yang Y. Recent progress in semantic image segmentation[J]. Artificial Intelligence Review, 2019, 52(2): 1089-1106.
 - [31] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4):834-848.
 - [32] Snell J, Swersky K, Zemel R S. Prototypical networks for few-shot learning[J]. arXiv preprint arXiv:1703.05175, 2017.
 - [33] Sung F, Yang Y, Zhang L, et al. Learning to compare: Relation network for few-shot learning[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 1199-1208.
 - [34] Ravi S, Larochelle H. Optimization as a Model for Few-Shot Learning[C]. Proceedings of the International Conference on Learning Representations, 2017.
 - [35] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]. International Conference on Machine Learning. PMLR, 2017: 1126-1135.
 - [36] Antoniou A, Edwards H, Storkey A. How to train your MAML[J]. Proceedings of the International Conference on Learning Representations, (Poster) 2019.
 - [37] Oreshkin B N, Rodriguez P, Lacoste A. TADAM: Task dependent adaptive metric for improved few-shot learning. Conference and Workshop on Neural Information Processing Systems, 2018: 719-729.
 - [38] Sun Q, Liu Y, Chua T S, et al. Meta-transfer learning for few-shot learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 403-412.
 - [39] Yu Z, Chen L, Cheng Z, et al. Transmatch: A transfer-learning scheme for semi-supervised few-shot learning[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 12856-12864.
 - [40] Lian D, Zheng Y, Xu Y, et al. Towards fast adaptation of neural architectures with meta learning[C]. Proceedings of the International Conference on Learning Representations, 2020.
 - [41] Mangla P, Kumari N, Sinha A, et al. Charting the right manifold: Manifold mixup for few-shot learning[C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision., 2020: 2218-2227.
 - [42] Zeng D, Liu K, Lai S, et al. Relation classification via convolutional deep neural network[C]. Proceedings of COLING 2014, the 25th international conference on computational linguistics: technical papers, 2014: 2335-2344.

-
- [43] Nii M, Tsuchida Y, Kato Y, et al. A Convolution Neural Network Based Nursing-Care Text Classification Model with a New Filter for Expressing Dependency Relations of Words[C]. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2018: 875-880.
 - [44] Miwa M, Bansal M. End-to-end relation extraction using lstms on sequences and tree structures[J]. arXiv preprint arXiv:1601.00770, 2016.
 - [45] Yang Y, Tong Y, Ma S, et al. A position encoding convolutional neural network based on dependency tree for relation classification[C]. Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 65-74.
 - [46] Mintz M, Bills S, Snow R, et al. Distant supervision for relation extraction without labeled data[C]. Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009: 1003-1011.
 - [47] Riedel S, Yao L, McCallum A. Modeling relations and their mentions without labeled text[C]. Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2010: 148-163.
 - [48] Hoffmann R, Zhang C, Ling X, et al. Knowledge-based weak supervision for information extraction of overlapping relations[C]. Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, 2011: 541-550.
 - [49] Surdeanu M, Tibshirani J, Nallapati R, et al. Multi-instance multi-label learning for relation extraction[C]. Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, 2012: 455-465.
 - [50] Zeng D, Liu K, Chen Y, et al. Distant supervision for relation extraction via piecewise convolutional neural networks[C]. Proceedings of the 2015 conference on empirical methods in natural language processing, 2015: 1753-1762.
 - [51] Yang K, Zheng N, Dai X, et al. Enhance Prototypical Network with Text Descriptions for Few-shot Relation Classification[C]. Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020: 2273-2276.
 - [52] Gao T, Han X, Liu Z, et al. Hybrid attention-based prototypical networks for noisy few-shot relation classification[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(01): 6407-6414.
 - [53] Yu H, Zhang N, Deng S, et al. Bridging Text and Knowledge with Multi-Prototype Embedding for Few-Shot Relational Triple Extraction[C]. Proceedings of the 28th International Conference on Computational Linguistics, 2020:6399-6410.
 - [54] Ren H, Cai Y, Chen X, et al. A Two-phase Prototypical Network Model for Incremental Few-shot Relation Classification[C]. Proceedings of the 28th International Conference on Computational

- Linguistics, 2020: 1618-1629.
- [55] Ye Z X, Ling Z H. Multi-Level Matching and Aggregation Network for Few-Shot Relation Classification[J]. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019: 2872-2881.
- [56] Xie Y, Xu H, Li J, et al. Heterogeneous graph neural networks for noisy few-shot relation classification[J]. Knowledge-Based Systems, 2020, 194: 105548.
- [57] Dong B, Yao Y, Xie R, et al. Meta-Information Guided Meta-Learning for Few-Shot Relation Classification[C]. Proceedings of the 28th International Conference on Computational Linguistics, 2020: 1594-1605.
- [58] Mitchell T M. Machine Learning. McGraw-Hill International Editions[M]. 1997.
- [59] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

致谢

光阴似箭，研究生三年的时光就这么匆匆结束了，也即将要离开我曾经生活与学习了7年的母校——西安电子科技大学，不舍之情油然而生。在西电读研的三年，我的生活过得非常充实，很多人给予了我无私的帮助，在这里真诚地向他们表示感谢！

首先，最感谢我的导师李雁妮教授，是她带领我走进了研究生的学习生涯，也很感谢她给了我这次读研的机会。第一次见李老师，她和蔼温暖的笑容就让我倍感亲切，也让我对研究生生涯充满了期待。入学以后，李老师在我的科研与生活中给予了非常多的指导。她以身作则，为我们树立了一个非常优秀的科研工作者的典范，同时为我们做人做事也树立了非常好的榜样。开始毕业论文工作后，李老师给予了我极其悉心的指导，从论文选题、文献筛选、算法设计、仿真实验，直至最后论文的撰写工作，都给予了我莫大的指导与帮助。在整个论文工作中，是她教我求真务实、深入细致、严格审视与把关工作的每一个环节，终使我能顺利较圆满地完成毕业论文工作。

在此，也特别感谢实验室的博士王陟与焦一源师兄！他们在我读研的三年中给予了我很多无私的关怀与帮助，无论是学习、科研，还是生活琐事，他们都能耐心细致给我解答帮助，并且给我传授了很多科研方面的工作经验。王陟师兄是我们实验室科研工作的榜样，焦一源师兄在我科研与生活中给予了我很多细致的指导，帮我解决了很多难题，在我毕业论文中给了我很多建设性的建议。同时，也深深感谢实验室的张文辉、杨文成、费航同学，我们一同入学，入学以后在学习科研中一起讨论学术问题，共同探讨在科研中遇到的难题。每当我遇到困难，他们为我提供了帮助，给我提出很多建设性的意见。感谢姚凯程、刘佳伟、史家辉、胡代旺师弟，感谢李朝霞、吕鹏帆师妹，在我平时的科研与生活中的帮助，以及感谢史家辉师弟对我论文的审查与诸多有益的建议。

感谢我的舍友，在我平时科研与生活中对我的帮助！每当我遇到困难是她们给予我非常多的鼓励，在最后的论文撰写阶段，我们互相督促，互相鼓励。感谢她们三年的陪伴，让我度过了非常快乐充实的三年，也让我收获了非常珍贵的友谊。

感谢我的父母，感谢他们的养育之恩，是他们无私的付出才让我顺利走完这么多年的学习生涯，非常感谢他们在我读研期间对我的支持和理解。

最后，感谢各位评审老师百忙之中抽出时间对我论文进行审阅并提出了非常宝贵意见与建议，让我的论文更加趋于完善。

作者简介

1. 基本情况

郝小慧，女，陕西府谷人，1995 年 8 月出生，西安电子科技大学计算机科学技术学院计算机技术专业 2018 级硕士研究生。

2. 教育背景

2014.09~2018.06 西安电子科技大学，本科，专业：计算机科学与技术

2018.09~至今 西安电子科技大学，硕士研究生，专业：计算机技术

3. 攻读硕士学位期间的研究成果

3.1 发表学术论文

无

3.2 参与科研项目及获奖

[1] 西安中科微精光子制造科技有限公司合作项目，对外轮廓图形内部填充算法研究，2019.5~2019.12，已完成，项目核心参与者。

[2] 获校二等奖学金，2018~2019 年度。

[3] 获校二等奖学金，2019~2020 年度。

[4] 获校二等奖学金，2020~2021 年度。



西安电子科技大学
XIDIAN UNIVERSITY

地址：西安市太白南路2号

邮编：710071

网址：www.xidian.edu.cn