领飞科创中期答辩

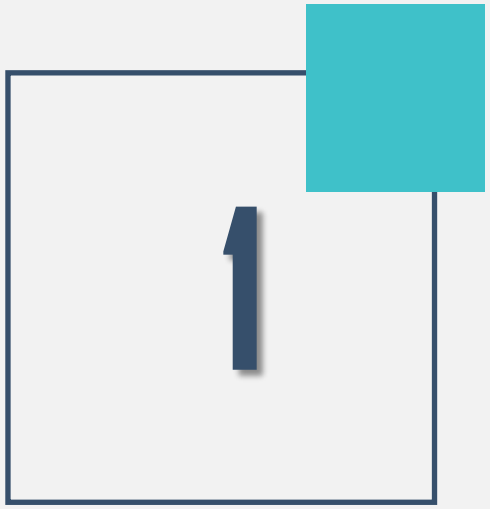
智慧典藏：面向动态互联网环境的持续识别系统

前言

经典的学习系统往往被部署在封闭环境中，利用预收集的数据集对固定类别的数据进行建模。然而在开放互联网环境中，这种假设难以满足，例如电商平台每天都会新增多种产品，社交媒体上的新热点话题层出不穷。因为新的类别会随时间不断增长，模型需要在数据流中持续地学习新类，持续实体识别中面临的灾难性遗忘问题亟待解决。

- 2024年1月初　　启动项目，开始筹备
- 2024年1月底-3月中　　文献调研，理论学习，形成技术路线
- 2024年3月中-5月底　　技术验证及实现，同步准备项目中期答辩

CONTENT



背景及发展

Background and Development

为使模型形成自主演化的实体识别能力，本项目基于类增量学习的方法，结合真实数据和合成数据，使模型通过分别学习任务不变和任务相关的提示信息，解决持续实体识别中面临的灾难性遗忘问题，保持对至少8种类型实体的识别准确率。

启动背景

连续数据流

传统模型假设数据分布是固定或平稳的，训练样本是独立同分布的，所以模型可以一遍又一遍地看到所有任务相同的数据。

但当数据变为连续的数据流时，训练数据的分布就是非平稳的，模型从非平稳的数据分布中持续不断地获取知识时，新知识会干扰旧知识，从而导致模型性能的快速下降 。

模型性能

静态A1模型不足以应对复杂多变的真实世界环境。如今的A1系统越来越需要释放机器智能体的终身学习能力。

神经网络缺陷

终身学习是参考人类的学习方式，使机器通过保存和积累过去所学的知识并用于未来的学习中。但神经网络模型在适应新任务之后，几乎完全忘记之前学习过的任务。

问题普遍性

如LLM微调和训练中的灾难性遗忘问题，以及深度学习的样本遗忘等。

完善人工智能，需要缓解模型在学习新任务过程中的灾难性遗忘问题，即当连续学习的任务越多时，学习下一个任务的速度就越快。

研究目标

基于**类增量学习**的方法，结合真实数据和合成数据，使模型通过分别学习任务不变和任务相关的提示信息，解决持续实体识别中面临的灾难性遗忘问题，保持对至少8种类型实体的识别准确率，使模型形成自主演化的实体识别能力。

通过深入拆解两篇论文，运用相关代码解决深度学习模型**固有的灾难性遗忘问题**，探讨得出解决**类增量NER任务中无法有效识别未标记实体问题**的有效方案，依次单个模块最终实现任务耦合。

积累论文资料和数据，将对特征表示学习法的研究成果初步转化在解决方案和论文初稿中。

核心技术分析

知识蒸馏

知识蒸馏通过**将一个大型、复杂的教师模型的知识转移到一个小型、简单的学生模型中**，以减少模型复杂性并提高泛化能力。知识蒸馏可以在学习新任务时，通过保留先前任务的知识，减轻先前任务的遗忘，并且可以在学习新任务时提供额外的信息和指导。

原型学习

原型学习通过学习任务特定的原型来表示类别，从而实现对任务特征的抽象和泛化。在解决灾难性遗忘问题中，原型学习可以帮助模型学习到通用的任务特征表示，减少对任务特定信息的依赖，从而提高模型的泛化能力，并且可以通过动态更新原型来适应新任务的要求。

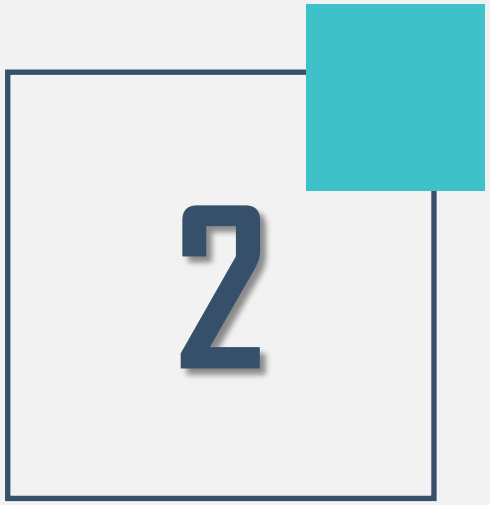
对抗性学习

对抗性训练通过引入对抗性扰动来增强模型的鲁棒性和泛化能力，可以使模型更加稳健，抵抗对新任务的干扰，同时可以通过在训练过程中引入对抗性样本，促使模型学习到更加通用的特征表示，从而减轻灾难性遗忘的影响。

当前技术成果概况

Overview of Current technological achievements

基于对两篇论文《Learning “O” Helps for Learning More: Handling the Unlabeled Entity Problem for Class-incremental NER》（下简称“Learning ‘O’ ”）和《Few-Shot Class-Incremental Learning for Named Entity Recognition》（下简称“Few-shot”）的重点阅读，形成项目开展的技术路线。同步开展对深度学习、NER等相关基础知识和模型的学习，以及Pytorch、Zotero、GitLab、Hugging face等软件和平台的学习实践，并依次突破单个模块最终实现任务耦合。同时，稳步跟进技术的文本记录和文献管理。



基于Learning “0”的仿真验证

Learning “0”-based simulation verification

技术路线

数据集制作

使用Few-NERD数据集和OntoNotes 5.0数据集，其中各包含多种细粒度实体模型，将其按任务划分，每个任务对应1个步骤，每个任务包含相同数量且相互不重复的实体类和一个“O”类。
任务的训练集和开发集：包含仅标记当前的任务类别的句子。
任务的测试集：包含标记有任务种所有学习类别的句子。

模型建立

1.加载bert-base-chinese模型或上一轮模型
2.加载预训练的参数配置、模型、分词器和训练集
3.获取每个batch的预测分数和输出标签，计算原型重新标记阈值后对旧实体类重标记。

模型训练

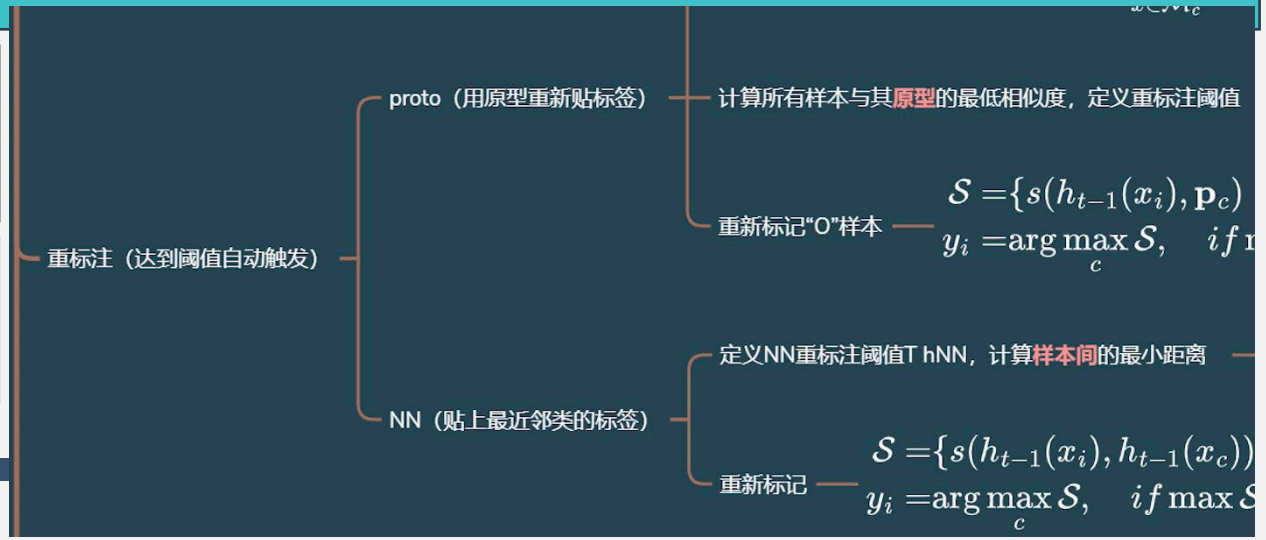
1.计算训练步数和训练轮数
2.配置AdamW优化器，使用权重衰减、学习率调节器
3.更新每个类别的类别相似度，获取样本的logits和标签，损失函数的确定，分析前k个训练周期，更新参数。

模型评估

在开发集上评估模型

模型测试

在测试集上进行预测



改进方案

模型重构

将原来的"bert-base-uncase"模型改为"bert-base-chinese"模型，以适用于中文命名实体识别。

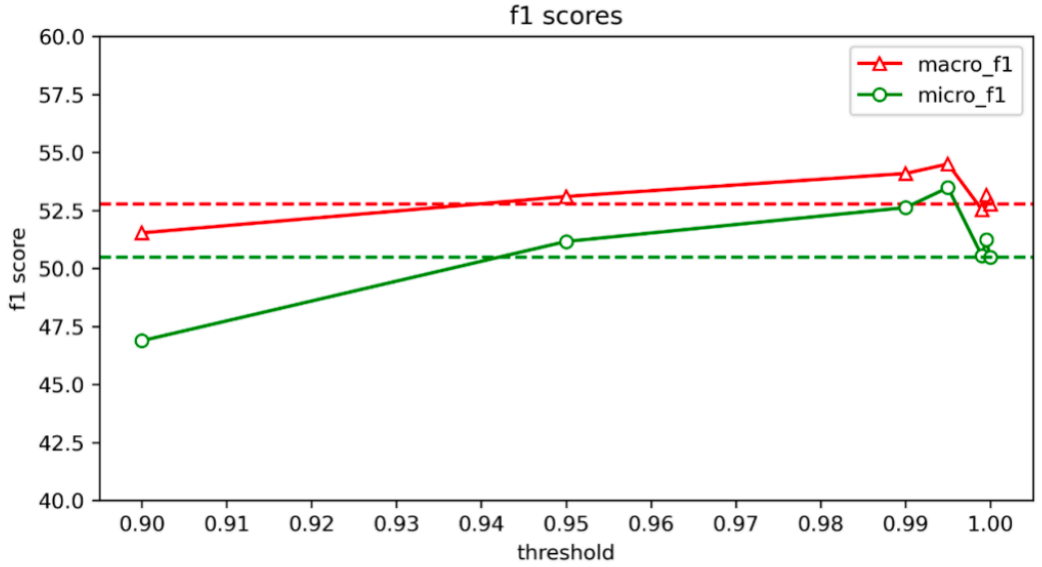
1、下载适用于中文任务的 BERT **预训练模型**。
2、**模型微调**，包括加载预训练模型、添加适当的输出层（如 CRF 层）等。
3、**模型评估**，使用评估指标（如准确率、F1 值等）评估微调后的模型在命名实体识别任务上的性能。
4、**优化和调整**，根据评估结果对模型进行优化和调整，以进一步提高性能。

标签意义扩充

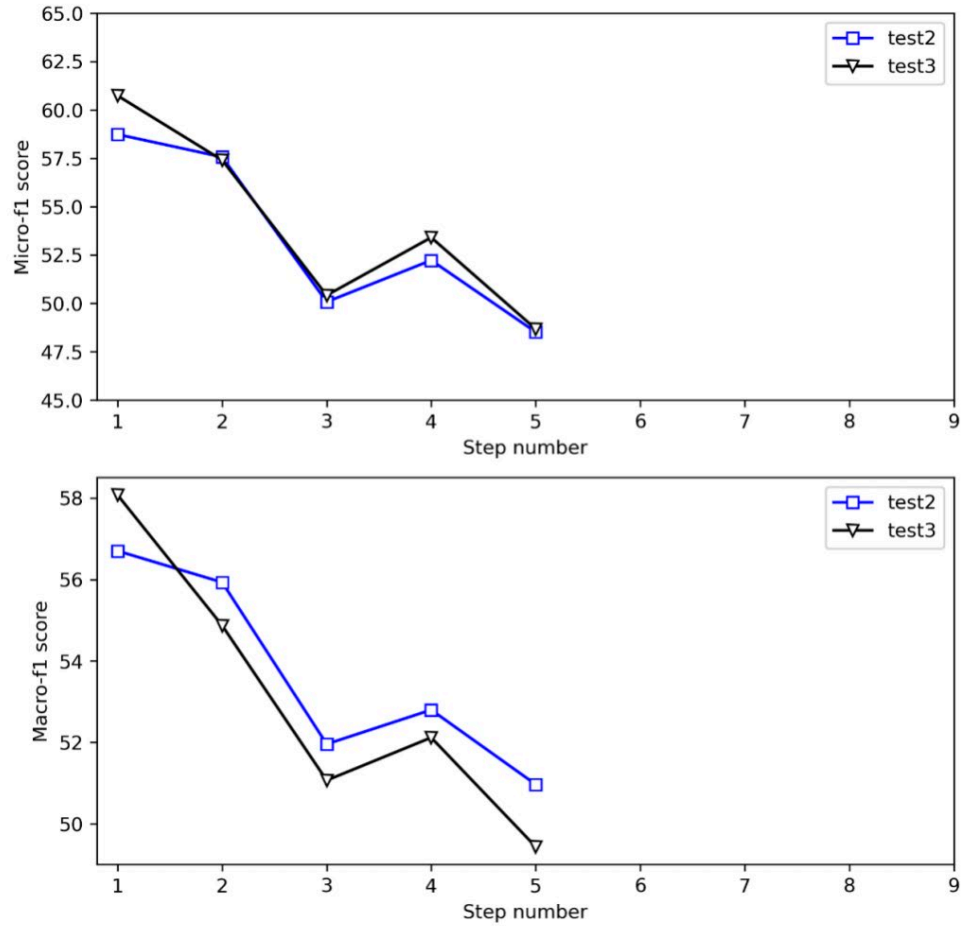
加入label-interpretation-learning，充分利用标签信息，以提升模型在少样本上的性能。

1、**标签解释表示学习**，通过构建标签之间的关系图或表示空间，学习标签的嵌入表示。
2、**标签信息融合**，将学习到的标签表示与输入数据进行融合，以提供额外的监督信息。
3、**模型训练与评估**，将融合了标签信息的输入数据输入到模型中进行训练。
4、模型优化，根据性能指标进行调优。

技术成果



threshold	0.9	0.95	0.99	0.995	0.999	0.9995	1
Micro f1	46.88	46.88	52.62	53.48	53.48	51.23	50.47
Macro f1	51.53	53.10	54.09	54.50	54.50	53.17	52.78

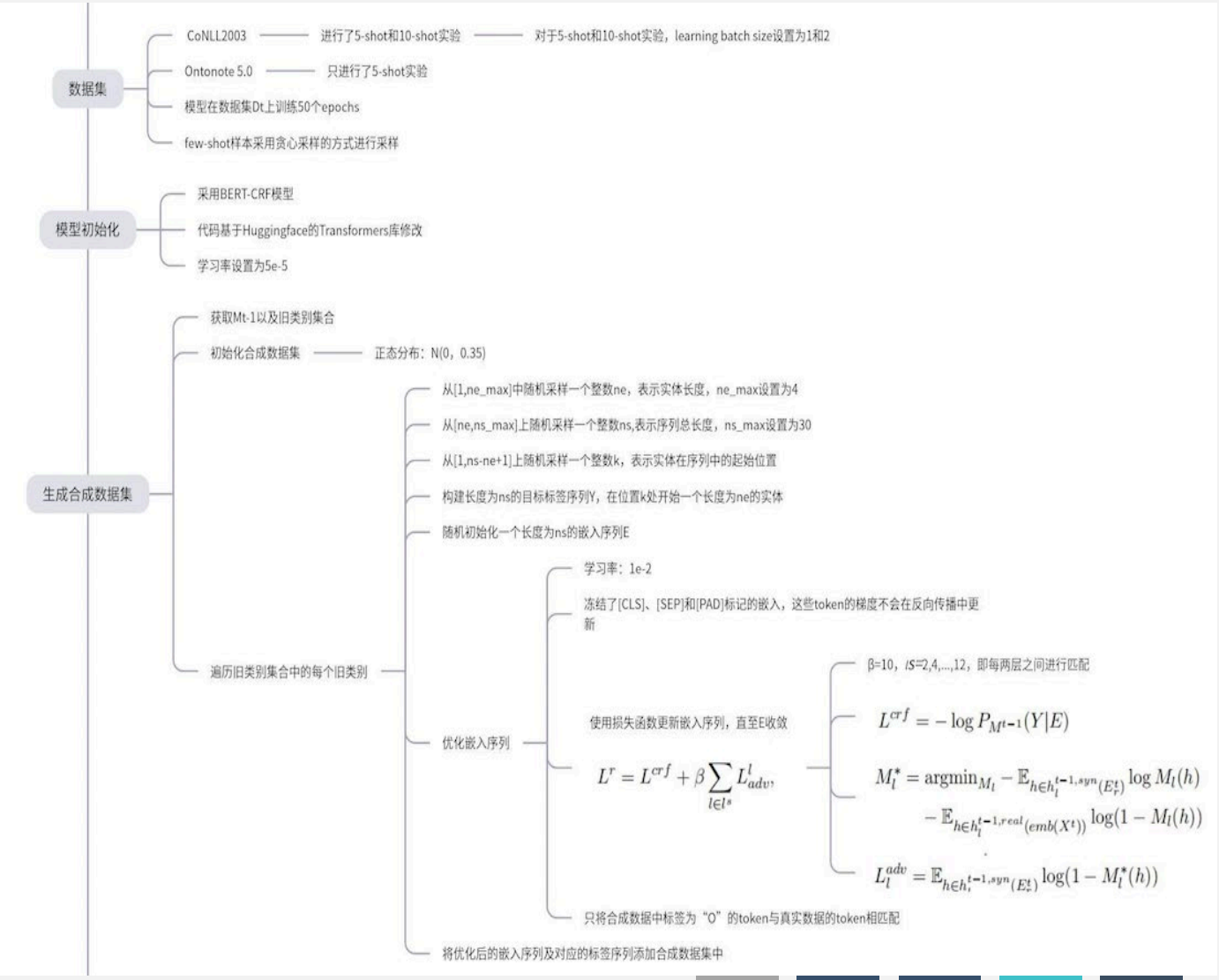


Steps		1	2	3	4	5
Micro f1	test2	58.74	57.57	50.08	52.23	48.52
	test3	60.74	57.41	50.42	53.41	48.68
Micro f1	test2	56.70	55.93	51.96	52.80	50.97
	test3	58.07	54.87	51.07	52.12	49.44

基于Few-shot的仿真验证

Few-shot -based simulation verification

技术路线



数据集

模型建立

模型训练

模型评估

模型测试

改进方案

模型重构

使用uie_nase_pytorch模型代替BERT-CRF作为预训练模型，并使用对抗性匹配方法增强模型的抗干扰以及对数据进行扩充的能力。

1、下载uie_nase_pytorch 模型作为基础的预训练模型，并使用 PyTorch 框架加载。
2、引入对抗性匹配方法，使用对抗性生成网络（GAN）生成对抗性样本，或者在训练过程中引入对抗性扰动，使模型更鲁棒地应对输入数据的干扰。
3、模型训练与优化，评估模型性能。

LSTM的数据优化

使用LSTM模型合成了较为真实的数据，训练效果有一定提升（F1分数高了0.01-0.05）。我们使用合成的数据集，不存在人为针对数据集的调整，结果更优秀，更加有广泛的适用性。

1、构建LSTM模型，用未标记的真实数据按类给模型做训练，定义生成函数，生成合成的数据。
2、对合成数据进行处理，处理成原模型（UIE）的的训练格式。
3、将数据加入原数据集中，获得增强数据集。
4、模型优化与评估。

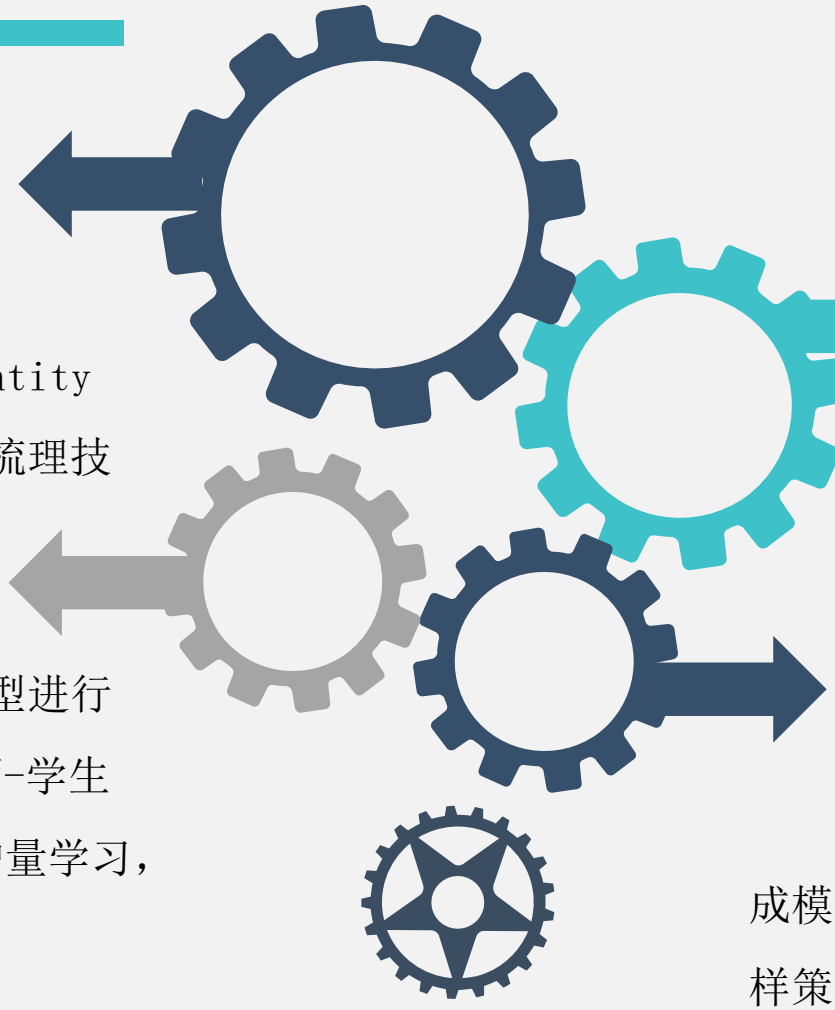
技术成果

文献学习

学习《Few-Shot Class-Incremental Learning for Named Entity Recognition》，了解NER技术原理，梳理技术路线并复现细节。

模型训练

训练uie_base_pytorch模型进行命名实体识别任务的训练，利用教师-学生模型，通过知识蒸馏的方法进行类增量学习，减轻灾难性遗忘问题。



模型优化

FGM对抗训练引入对抗扰动来提升模型的鲁棒性；EMA指数移动平均，对模型参数的指数加权平均进行平滑，减少了训练过程中模型参数的波动。

LSTM合成数据

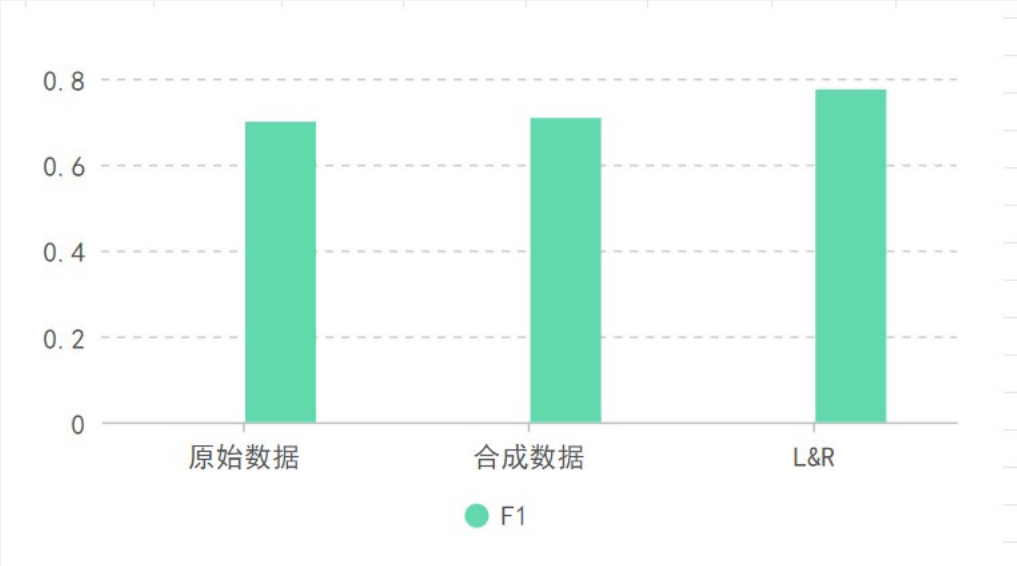
为解决数据集过少的问题，使用文本生成模型创建不打标签的旧有知识样本。通过采样策略批量合成负样本，提升模型训练效果，经实践，模型学习效果有提升。

技术成果

		T5	T6	T2	T1	T3	T4
	T5	0.74602	0.59497	0.69565	0.64	0.66947	0.61422
原始数据	T6		0.80169	0.52492	0.51573	0.69489	0.65933
	T2			0.74603	0.70164	0.6606	0.58917
	T1				0.85383	0.74299	0.76673
	T3					0.86732	0.66425
	T4						0.85201
		0.74602	0.69833	0.65553	0.6778	0.72705	0.69095
			F1	0.69928			

		T5	T6	T2	T1	T3	T4
	T5	0.74602	0.64812	0.70226	0.68274	0.67707	0.62391
合成数据	T6		0.78785	0.52952	0.55259	0.66328	0.66904
	T2			0.77835	0.6549	0.64957	0.61786
	T1				0.85006	0.73513	0.79404
	T3					0.85714	0.72115
	T4						0.85608
		0.74602	0.71799	0.67004	0.68507	0.71644	0.71368
			F1	0.70821			

		T5	T6	T2	T1	T3	T4
	T5	0.74602	0.76028	0.72984	0.73716	0.70007	0.68776
Learn	T6		0.79351	0.78894	0.80171	0.78785	0.7812
And	T2			0.79677	0.75446	0.72388	0.70331
Review	T1				0.84795	0.83895	0.83077
	T3					0.86347	0.85305
	T4						0.85534
		0.74602	0.7769	0.77185	0.78532	0.78284	0.78524
			F1	0.77469			



4

项目总结

Project Summary

目前，我们的项目进展已经进行过半，对于数据集的学习能力达到预期效果，基本达到对灾难性遗忘问题的解决预期，目前正在针对伪标签的创新点和通过文本模型增强副样本进行攻克。同时，整理所得技术成果和文本材料，形成论文，申请专利。

当前成果小结

基于类增量学习的灾难性希望问题解决

为了缓解灾难性遗忘问题，我们分别采用**bert-base-chinese**和**uie_base_pytorch**模型进行命名实体识别任务的训练，**通过知识蒸馏的方法进行类增量学习**，同时，基于bert模型，通过实体感知对比学习和原型重新标记策略来减轻灾难性遗忘问题。在此基础上，我们加入对标签语义对扩充和LSTM的合成数据，优化模型性能。并在后期寻找到使用伪标签的创新点，对命名实体识别中的模型进行优化，赋予其更丰富的语义，达到进一步减轻机器学习中的灾难性遗忘问题的效果。

风险评估

风险分析

- **过拟合问题：**在处理新旧数据集和进行数据合成的过程中，可能会过度记忆过去的数据，导致模型过拟合，对新数据的泛化能力下降。
- **资源需求问题：**实现持续学习和记忆重放技术需要大量的计算和存储资源，对硬件资源也有着更高的需求，现阶段对研究可能导致成本上升和性能下降。
- **算法偏见：**比如对于Bert和Ernie模型的代码有着不同的适配性，优化对象的不同往往导致指标不适配，使模型性能不佳。可能会加剧系统中的算法偏见，导致不理想或歧视性的结果。
- **技术成熟度不足：**持续学习和记忆重放技术可能还处于技术成熟度不足的阶段。可能会面临技术实施难度大、算法不稳定或者不可预测的问题，影响系统的可靠性和效果。

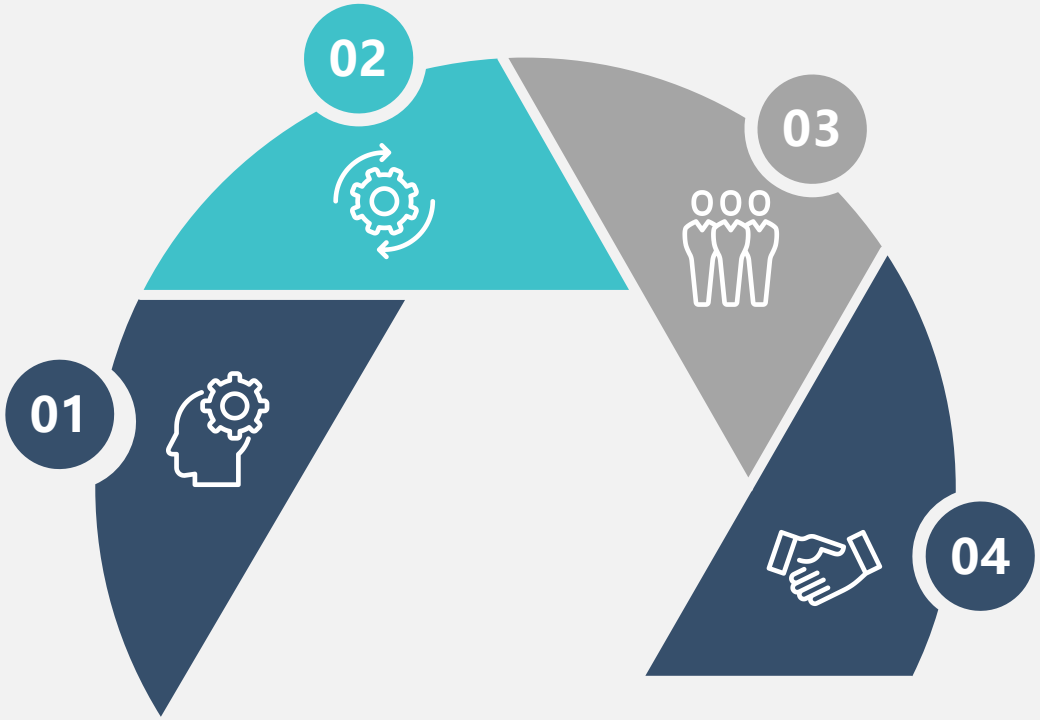
风险控制

我们通过使用LSTM生成合成数据，获取增强数据集，以提高训练效果，并测试label interpretation learning方法在少样本问题中的作用，通过充分利用标签信息提升模型性能，并加强对学习过程对监督，以及模型对以及对新数据的适应性和泛化能力。

依托指导老师的帮助，我们通过云端服务器测试和验证了大量模型，由于数据集庞大和转换等问题，也会延迟验证进度。

由于使用新模型覆盖代码对工作量过大，我们采取扩充伪标签的方法优化模型性能，避免不成熟的技术问题影响整体效果。

目标问题及应用场景



缓解灾难性遗忘问题

增强机器学习模型的适应性，使其能够持续学习和适应新的数据和情境，从而更好地应对不断变化的环境和需求。将推动自适应智能系统的普及，在智能交通、智能家居和智能医疗等方面带来更多应用。

大数据分析

引入多模态学习，结合多种信息来源，如文本、图像、语音等，以增强学习的稳健性和鲁棒性。并且通过不断更新和扩展模型，助力大数据分析提供更个性化、精准的推荐和服务，满足用户的个性化需求和偏好。

人机协作

持续学习系统的发展将促进人机协作的深化。智能系统不仅能够从人类那里学习新知识，还能够与人类共同解决问题，并且不断改进和优化自身的性能。在此过程中，也可以动态加强人类对人工智能的认知和监督。

跨领域知识整合

通过多模态学习和知识迁移，促进不同领域之间的知识整合和交叉创新，推动跨领域研究和创新。同时，基于不断积累的多领域知识和经验，实现更智能、更有效的决策和优化，提高系统的效率和性能。

发展规划及策略

01

将语义扩充和LSTM数据合成的方法代入bert模型中，优化模型性能，并基于伪标签的整理文档，开始系统性的验证。

02

汇总程序和技术成果，形成技术文档报告，针对伪标签创新点的进行系统的材料整理论证。

03

基于现有模型，进一步完善解决灾难性遗忘问题的解决模型，将所使用的方法进行评估，代入实际场景中进行验证。

04

在模型优化和创新点实现的基础上，有序开展论证撰写和发表，进行成果展示。

05

成果转化，将技术成果和文档进一步挖掘和运用，投入到更多的比赛和应用场景。