

基于联邦知识蒸馏的跨语言社交媒体事件检测

周帅帅^{1,2}, 朱恩昌^{1,2}, 高盛祥^{*1,2}, 余正涛^{1,2}, 线岩团^{1,2}, 赵子霄^{1,2}, 陈霖^{1,2}

1.昆明理工大学, 信息工程与自动化学院, 昆明, 650500

2.昆明理工大学, 云南省人工智能重点实验室, 昆明, 650500

20222204121@stu.kust.edu.cn, {zhuenchang.kmust, gaoshengxiang.yn}@foxmail.com, ztyu@hotmail.com, xianyantuan@qq.com, {zhaozixiao, 20232204066}@stu.kust.edu.cn

摘要

社交媒体事件检测是指在从各类社交媒体的内容中挖掘热点事件。在实际情况中, 由于数据稀缺, 社交媒体事件检测在低资源的情况下表现较差。现有的方法主要通过跨语言知识迁移等方式来缓解低资源问题, 但忽略了数据隐私问题。因此, 本文提出了基于联邦知识蒸馏的跨语言社交媒体事件检测框架 (FedEvent), 旨在将富资源客户端知识蒸馏到低资源客户端。该框架通过结合参数高效微调技术和三组对比损失, 实现非英文语义空间到英文语义空间的有效映射, 并采用联邦蒸馏策略, 保障数据隐私的前提下实现知识的迁移。此外, 我们还设计了一套四阶段生命周期机制以适应增量场景。最后, 我们在真实数据集上进行实验以证明该框架的有效性。

关键词: 社交媒体事件检测; 低资源; 联邦知识蒸馏; 跨语言知识迁移

Cross-Lingual Social Event Detection Based on Federated Knowledge Distillation

Shuashuai Zhou^{1,2}, Enchang Zhu^{1,2}, Shengxiang Gao^{*1,2}, Zhengtao Yu^{1,2}, Yantuan Xian^{1,2}, Zixiao Zhao^{1,2}, Lin Chen^{1,2}

1. Faculty of Information Engineering and Automation,

Kunming University of Science and Technology, Kunming 650500, China

2. Yunnan Key Laboratory of Artificial Intelligence,

Kunming University of Science and Technology, Kunming 650500, China

20222204121@stu.kust.edu.cn, {zhuenchang.kmust, gaoshengxiang.yn}@foxmail.com, ztyu@hotmail.com, xianyantuan@qq.com, {zhaozixiao, 20232204066}@stu.kust.edu.cn

Abstract

Social event detection involves identifying trending events from various social event content. In practice, due to data scarcity, social event detection performs poorly in low-resource settings. Existing methods primarily utilize cross-lingual knowledge transfer to mitigate low-resource issues; however, these approaches often overlook data privacy concerns. Consequently, this paper introduces a cross-lingual social media event detection framework, named FedEvent, which is based on federated knowledge distillation and is aimed at distilling knowledge from high-resource clients to those with low resources. The framework employs parameter-efficient fine-tuning techniques and triple contrastive losses to effectively map non-English semantic spaces to English ones, while using a federated distillation strategy to ensure data privacy. Additionally, we have developed a four-stage lifecycle mechanism to accommodate incremental scenarios. Finally, experiments on real datasets demonstrate the effectiveness of our framework.

Keywords: Social event detection, Low-resource, Federated knowledge distillation, Cross-lingual knowledge transfer

1 引言

社交媒体事件检测是指从社交媒体上的内容中，挖掘引发公众广泛关注和讨论的热点事件。近年来，各类社交媒体平台迅速发展，对社交媒体事件进行检测，在控制信息传播和正确引导舆论等方面具有重要意义。然而，在实际情况中，富资源语言数据量充足，模型训练效果好。但在低资源场景下，由于资源稀缺等问题，模型无法得到充足的训练数据，导致表现不达预期。因此，如何解决社交媒体事件检测存在的数据分布不均衡问题引发了广泛关注。同时，有效的社交媒体检测也为事件演化分析(Atefeh and Khreich, 2015)、舆情分析(Peng et al., 2021)、情绪分析(Gaspar et al., 2016)、企业风险管理(Nisar and Yeung, 2018)以及政治选举预测(Marozzo and Bessi, 2018)等下游任务提供了坚实的基础。

在社交媒体事件检测方面，大多数相关研究将其转化为增量聚类问题(Aggarwal and Subbian, 2012; Hu et al., 2017; Ozdakis et al., 2017; Zhang et al., 2007)、社区检测问题(Nguyen and Jung, 2015; Liu et al., 2020a; Liu et al., 2020b; Liu et al., 2020c; Yu et al., 2017)或主题建模问题(Becker et al., 2011; Cheng et al., 2014; Cordeiro, 2012; Zhou and Chen, 2014)来处理。然而，上述方法忽略了社交消息文本中丰富的语义和结构信息。为了解决以上问题，相关研究者提出了基于图神经网络(GNN)(Chung et al., 2020; Su et al., 2022)的社交媒体事件检测方法(Cui et al., 2021; Peng et al., 2019; Peng et al., 2021; Peng et al., 2023; Ren et al., 2022a; Ren et al., 2022b)，以有效捕获社交消息文本中深层语义信息和结构信息。此外，为了实现社交媒体增量事件检测，Cao et al. (2021)提出了名为KPGNN的方法，该方法设计了一套生命周期机制，有效解决了流式场景下事件知识保留和更新难题。上述基于GNN的社交媒体事件检测方法在富资源语言(如英语)场景下取得了显著的效果，但是在低资源语言(如越南语)场景下模型性能还有很大的提升空间。因此，如何有效利用富资源语言知识来提升低资源语言的社交媒体事件检测模型性能成为了最近的研究热点。基于此，Peng et al. (2023)提出了基于跨语言模型参数知识迁移的社交媒体事件检测方法，该方法借助预训练的跨语言模块-LNMAP(Mohiuddin et al., 2020)，实现将富资源语言(英语)的模型参数知识迁移到低资源语言中。Ren et al. (2023)提出了基于跨语言知识蒸馏的社交媒体事件检测方法，其核心思想是，利用知识蒸馏将在富资源语言上训练的教师模型中的特征知识和关系知识蒸馏到低资源语言的学生模型中，该方法在低资源场景下取得了显著的效果。上述研究尝试通过跨语言知识迁移等方式来缓解低资源问题，利用富资源语言数据来帮助低资源语言模型提升性能。然而，在实际应用中，富资源数据与低资源数据通常分布于不同社交媒体平台，现有的方法采用一种数据中心化的方式来进行富资源语言到低资源语言的迁移，显然忽略了这一特点。由于隐私和法律法规的限制，大量的高质量数据无法在不同平台之间流动，联邦学习作为一种数据去中心化的学习架构，为这一情况提供了可能的解决方法。值得一提的是，面对大量的社交媒体事件数据，联邦学习还可以降低数据传输和存储的成本，更符合显示需求。

然而，现有的基于联邦蒸馏的框架还面临诸多挑战：**i) 低资源数据缺失**：由于数据稀缺等问题的影响，单独构建低资源语言模型难度较大，在低资源场景下的社交媒体事件检测方面表现达不到预期。**ii) 语言迁移难度大**：由于不同语言之间的语法和表达方式不同，分布在不同客户端的数据差异较大，导致模型漂移，从而影响知识蒸馏的效果。与此同时，富资源语言与低资源语言之间的数据共享受到诸多限制也是亟待解决的问题。**iii) 社会事件持续涌现**：在现实场景中，社交媒体事件具有数据量大和增长速度快的特点，对模型接收和处理新信息的能力提出更高的要求。

针对以上问题，本文提出了一种基于联邦知识蒸馏的跨语言社交媒体事件检测方法，实现了在保护数据隐私的同时，将富资源客户端知识有效地蒸馏到低资源客户端，提高低资源语言模型的性能。具体而言，针对低资源数据稀缺的问题，我们提出了在联邦学习框架下，采用知识蒸馏的策略，使得富资源客户端的数据能够有效地迁移到低资源客户端，并指导低资源客户端的模型进行训练。同时，我们训练了一个跨语言词嵌入模型，采用参数高效微调技术和三组

*高盛祥(通信作者): gaoshengxiang.yn@foxmail.com

基金项目: 国家自然科学基金(U23A20388, U21B2027, 62376111, 62266028, 62266027); 云南省重点研发计划(202303AP140008, 202401BC070021, 202103AA080015); 云南省科技人才与平台计划(202105AC160018); 云南省基础研究项目(202301AT070393)。

©2024 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

对比损失，捕捉不同语言之间的语义差异，克服了语言之间的语义空间不一致的困难。此外，面对持续涌现的社交媒体事件数据量，本文在联邦框架下引入增量的生命周期，设计了一套四阶段的生命周期机制，使得模型能够不断适应传入的新数据，持续检测新的事件，从而达到数据增量的效果。实验结果表明，在低资源场景下，我们提出的框架在社交媒体事件检测任务上取得了显著的效果。通过实验分析，我们证明了FedEven框架不仅在性能上超越了传统的单机与多机环境下的基线方法，而且在确保不共享数据的前提下，实现了跨语言知识的有效迁移。

2 方法

本框架包括跨语言词嵌入模块、消息表征图学习模块和联邦网络消息蒸馏模块。首先，介绍联邦跨语言社交媒体事件检测的任务定义。其次，描述跨语言词嵌入模块。接着，介绍消息表征图学习模块。最后，详述联邦网络消息蒸馏模块，模型框架图如图1所示。

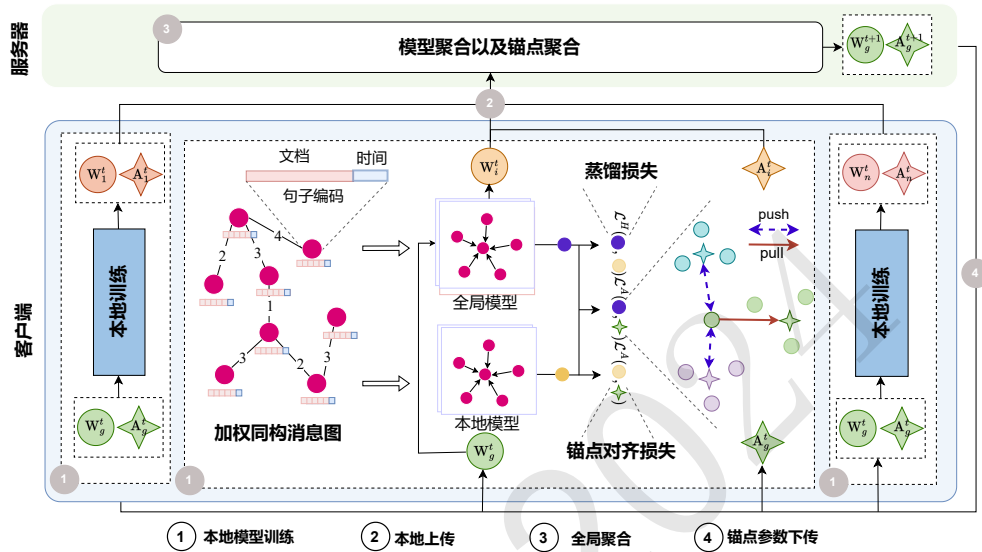


图 1. FedEvent框架图

2.1 任务定义

本小节将介绍本文中所使用的主要符号。我们扩展了Cao et al. (2021)中的社会消息流和社交媒体事件的定义，并给出了联邦跨语言社交事件检测的定义以及符号说明。

定义2.1: 社会消息流 $M_0, M_1, M_2, \dots, M_i, \dots$ 是随时间演进的消息序列块，其中 M_i 代表在 $[t_i, t_{i+1}]$ 期间所到达的全部社会事件 $\{m_0, m_1, m_2, \dots\}$ ，即 $M_i = \{m_j | 1 \leq j \leq |M_i|\}$ ， $|M_i|$ 是 M_i 中消息的个数。 $M_0, \{M_i | i \neq 0\}$ 分别是离线场景和在线场景下的消息块。

定义2.2: 对于**联邦跨语言**社交媒体事件检测，低资源客户端模型可以从富资源客户端中学习有益知识从而提升模型性能。假设， H 为富资源客户端集合， L 为低资源客户端集合， $M_{k,i}$ 为第 k 个客户端上第 i 个消息块，其中每个客户端上的消息块均是单种语言。富资源客户端上第 i 个消息块中的事件集合 $C(M_{k,i} | k \in H)$ 包含低资源客户端事件集合 $C(M_{k,i} | k \in L)$ ，即 $C(M_{k,i} | k \in L) \subseteq C(M_{k,i} | k \in H)$ 。低资源客户端 k 上第 t 轮次，本地模型 $f_{k,l}(\mathbf{W}_{k,l}^{(t)}, x | x \in M_{k,i}, k \in L)$ 通过本地客户端数据集微调以及间接获取富资源客户端本地模型 $f_{k,l}(\mathbf{W}_{k,l}^{(t)}, x | x \in M_{k,i}, k \in H)$ 的知识进行学习。

本框架的重点是在资源有限的客户端上训练一个鲁棒的模型，而非专注于训练一个能够支持多种语言的检测模型。在第 i 个消息块中，整个框架的目标是最小化所有低资源客户端的总误差，即：

$$\min_{\mathbf{w}} \Phi(\mathbf{w}) = \min_{\mathbf{w}} \sum_{k=1}^K p_k \Phi_k(\mathbf{w}) \quad (1)$$

其中， $\Phi_k(\mathbf{w})$ 是第 k 个客户端所要优化的目标， p_k 表示第 k 个客户端的预定义聚合权重。

2.2 跨语言词嵌入

我们利用非英文到英文的平行语料训练了一个跨语言词嵌入模型，旨在实现非英文文本到英文文本的语义空间映射。通过借助参数高效微调技术LoRA(Hu et al., 2021)的低秩自适应策略，捕捉不同语言之间的语义差异。本研究所设计的跨语言词嵌入模块 (\mathcal{M}) 是在多语言预训练模型中的注意力计算矩阵 \mathbf{W}_q 和 \mathbf{W}_v 上加一个旁支。具体而言，模型改进如下：

$$\mathbf{Q}_{\text{origin}} = \mathbf{X} \cdot \mathbf{W}_q, \mathbf{V}_{\text{origin}} = \mathbf{X} \cdot \mathbf{W}_v \quad (2)$$

$$\mathbf{Q}_{\text{lorea}} = \mathbf{X} \cdot (\mathbf{W}_q + \text{scale} \cdot \mathbf{BA}), \mathbf{V}_{\text{lorea}} = \mathbf{X} \cdot (\mathbf{W}_v + \text{scale} \cdot \mathbf{BA}) \quad (3)$$

其中， $\mathbf{Q}_{\text{origin}}, \mathbf{V}_{\text{origin}}$ 代表改进前 \mathbf{Q} 以及 \mathbf{V} 的计算方式， $\mathbf{Q}_{\text{lorea}}, \mathbf{V}_{\text{lorea}}$ 代表改进后 \mathbf{Q} 以及 \mathbf{V} 的计算方式。 $\mathbf{B} \in \mathbb{R}^{d \times r}, \mathbf{A} \in \mathbb{R}^{r \times k}, \text{scale} = \alpha/r$ ，仅仅训练 \mathbf{BA} 以达到适配效果。

为了使得所引入的LoRA模块不会损失英文语义，我们引入了一个指导模型(\mathcal{M}')，即被冻结的多语言预训练模型。利用对比学习来实现不同语言语义空间的映射，语义对齐损失 $\mathcal{L}_{\text{total}}$ 为：

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{c1} + \mathcal{L}_{c2} + \mathcal{L}_{c3} \quad (4)$$

其中， \mathcal{L}_{c1} ， \mathcal{L}_{c2} ，和 \mathcal{L}_{c3} 三组对比学习损失，具体形式为：

$$\mathcal{L}_{c1} = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(\mathcal{M}(x_i), \mathcal{M}'(y_i))/\tau)}{\sum_{j=1}^N \exp(\text{sim}(\mathcal{M}(x_i), \mathcal{M}'(y_j))/\tau)} \quad (5)$$

$$\mathcal{L}_{c2} = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(\mathcal{M}(y_i), \mathcal{M}'(y_i))/\tau)}{\sum_{j=1}^N \exp(\text{sim}(\mathcal{M}(y_i), \mathcal{M}'(y_j))/\tau)} \quad (6)$$

$$\mathcal{L}_{c3} = - \sum_{i=1}^N \log \frac{\exp(\text{sim}(\mathcal{M}(x_i), \mathcal{M}(y_i))/\tau)}{\sum_{j=1}^N \exp(\text{sim}(\mathcal{M}(x_i), \mathcal{M}(y_j))/\tau)} \quad (7)$$

其中， N 是一个batch中样本的数量， $\text{sim}(a, b)$ 表示 a 和 b 之间的相似度， x 代表非英文文本， y 代表相应的英文文本， τ 是温度参数，用于调节相似度的尺度。与其他自然语言处理任务一致，我们使用模型所输出的第一个TOKEN的表征（即CLS）作为整个句子的表征。该损失函数的目标是最大化正样本对之间的相似度，同时最小化与其他非对应样本之间的相似度。

2.3 消息表征图学习模块

2.3.1 加权同构消息图构建

为了充分利用社交数据中的关键信息，我们依据用户、实体和hashtag这三种元素，构建加权同构消息图。具体而言，当两条消息共享上述三种元素中的任意一种时，就在图中为它们添加一条边。直观而言，共享的元素数量越多，两条消息之间的关系就越紧密。因此，以共现元素的个数来作为边的权重，构建加权同构消息图。在初始消息的编码方面，我们遵循了Cao et al. (2021)中结合语义和时间信息的方法。对于时间信息，我们将时间戳转换为OLE日期，进而得到二维时间特征。对于语义信息，我们采用2.2节训练的跨语言词嵌入模块对消息进行统一编码，得到语义表征。最终消息表征是消息的语义表征以及时间特征在最后一个维度上的拼接。

2.3.2 GNN编码器

模型选择方面，我们选择GNN模型来作为特征提取器，主要原因有以下两点：1) GNN保留了强大的结构信息以及语义信息，并可以充分利用社交媒体上的用户、实体、hashtag元素信息。2) 图神经网络较少的参数，可以更灵活的适应各个客户端，减少联邦通信时间。为了深度融合丰富的语义和关系信息，GNN编码器通过迭代地从其一跳邻居处汇总信息，学习结点表示。形式上，第 l 层消息 m_i 表示为 \mathbf{h}_{m_i} ，其在 $l+1$ 层的更新可表示为：

$$\mathbf{h}_{m_i}^{(l+1)} \leftarrow \parallel^{\text{heads}} \left(\mathbf{h}_{m_i}^{(l)} \oplus \underset{\forall m_j \in \mathcal{N}(m_i)}{\text{Aggregator}} \left(\text{Extractor} \left(\mathbf{h}_{m_j}^{(l)} \right) \right) \right) \quad (8)$$

其中, $\mathcal{N}(m_i)$ 代表消息 m_i 的一跳邻居。 \oplus 代表特征聚合方式, \parallel^{heads} 代表多个头的拼接。聚合器 (Aggregator) 以及抽取器 (Extractor) 在不同的GNN模型中是不同的。不同的客户端可以根据本地数据设计不同的GNN, 进行本地模型的私有化定制。

2.4 联邦网络消息蒸馏模块

在整体架构上, 我们遵循了传统的基于知识蒸馏的服务器-客户端设计。服务器端, 我们设计了一个全局模型 \mathbf{W}_g 和全局锚点 \mathbf{A}_g 来进行富资源客户端以及低资源客户端之间的通信。每个客户端包含一个本地模型 $\mathbf{W}_{k,l}$ 用于适应本地数据、一个全局模型 $\mathbf{W}_{k,g}$ 用于特征蒸馏、一个全局锚点 $\mathbf{A}_{k,g}$ 用于对齐各个客户端特征空间。我们利用全局模型作为过程监督, 全局锚点作为结果监督来实现知识的有效迁移。其中, 本地模型根据本地数据进行更新以适应本地数据特点。全局模型在所有客户端之间共享, 并进行协作学习。我们利用生命周期机制来适应在线以及离线场景。接下来, 我们将详细介绍联邦聚合方式以及生命周期机制。

2.4.1 联邦聚合

在生命周期不同阶段, 联邦聚合方式略有不同, 但也有一定的共性。我们假设当前块为 M_i , 以此块为例, 介绍全局锚点聚合以及全局模型聚合。

本地锚点计算: 本地锚点由当前客户端上消息块 M_i 中每个簇内所有样本表征的均值来表示。因此, 本地锚点是当前客户端上源自同一事件消息簇特性的综合体现。 t 轮次时, 客户端 k 中事件簇 c 的局部锚点表征 $a_{k,c}^{(t)}$ 计算如下:

$$a_{k,c}^{(t)} = \frac{1}{|\mathcal{B}_{k,c}|} \sum_{(\mathbf{x},c) \in \mathcal{B}_{k,c}} f_{k,l}(\mathbf{W}_{k,l}^{(t)}, \mathbf{x}) \quad (9)$$

其中, $|\mathcal{B}_{k,c}|$ 表示客户端 k 上事件簇 c 的样本数量。 $f_{k,l}$ 代表本地模型的输出。

全局锚点聚合: 中央服务器汇总来自参与聚合的客户端上本地锚点表征, 聚合方式根据各个客户端消息块中消息个数进行加权平均, 即:

$$a_c^{(t)} := \frac{1}{\sum_{k \in S} |\mathcal{B}_{k,c}|} \sum_{k \in S} |\mathcal{B}_{k,c}| a_{k,c}^{(t)} \quad (10)$$

其中, S 为满足聚合条件的客户端集合, 在生命周期机制中不同阶段, 集合 S 有所不同, 我们将在2.4.2节进行说明。通过公式(10)得到全局锚点 $A_g^{(t+1)} = \{a_c^{(t)}\}_{c=1}^C$ 将会被广播到所共享的各个客户端。此时, 所获得的全局锚点能够综合反映同一事件簇的特征, 可以用来指导每个客户端的模型训练。

全局模型聚合: 中央服务器接收并汇总来自所有客户端上的全局模型, 通过聚合处理生成下一轮更新的全局模型 $\mathbf{W}_g^{(t+1)}$, 即:

$$\mathbf{W}_g^{(t+1)} := \frac{1}{\sum_{k \in S} |\mathcal{B}_k|} \sum_{k \in S} |\mathcal{B}_k| \mathbf{W}_{k,g}^{(t)} \quad (11)$$

其中, $|\mathcal{B}_k|$ 表示客户端 k 中样本个数, $\mathbf{W}_{k,g}^{(t)}$ 为第 k 个客户端中第 t 轮时全局模型的参数。

在每一轮通信中, 当本地训练完后, 客户端 k 将 $\{\mathbf{A}_k^{(t)}, \mathbf{W}_{k,g}^{(t)}\}$ 上传至中央服务器, 通过公式(10),(11)进行聚合, 然后将聚合结果 $\{\mathbf{A}_g^{(t+1)}, \mathbf{W}_g^{(t+1)}\}$ 下拉至各个客户端, 进行新一轮的本地训练。

2.4.2 生命周期机制

生命周期如图2所示。FedEvent的生命周期包括富资源客户端预训练(Stage I)、低资源客户端预训练(Stage II)、各个客户端检测(Stage III)以及维护(Stage IV)四个阶段。在第一阶段, 我们利用富资源客户端消息块上 M_0 构建一个初始的加权同构消息图并利用特定的算法训练初始模型。在第二阶段, 我们利用阶段一所训练的模型以中央服务器作为知识传输的媒介来辅助低资源客户端训练初始模型。在第三阶段, 为每个到来的块构建一个新的图, 然后使用已训练好的模型直接检测每个到来的消息块。在第四阶段, 持续用最新的消息块训练模型, 并利用联邦

共蒸馏机制来实现各个客户端在线学习，使模型能够学习新知识。其中，阶段一以及阶段二用于离线学习，阶段四用于在线学习。阶段三及阶段四交替进行，通过这种方式，模型不断适应传入的数据，可以检测新事件并更新模型的知识。通过生命周期机制，FedEvent能够更好地适应流媒体特性。

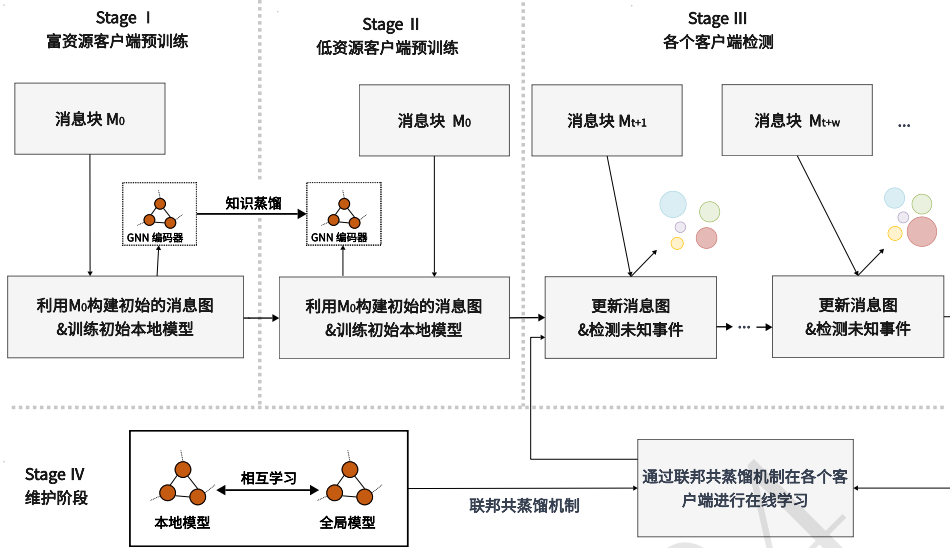


图 2. FedEvent生命周期机制

Stage I: 富资源客户端预训练。该阶段的目标是在富资源客户端利用初始消息块 M_0 训练一个鲁棒的本地模型 $\mathbf{W}_{k,l}$ 。随着新事件的不断涌现，事件的总数很难提前知道。考虑到交叉熵损失只适用于事件总数是已知和固定的情况，难以处理流式数据。因此，我们转向利用成对学习方(Ren et al., 2022a)，保证簇间紧密性和簇内分离性。由于这种特性，使得在后续第二、三阶段，可以将富资源客户端清晰的簇间界限信息蒸馏至低资源客户端。该阶段任务特定损失以及总损失分别为 \mathcal{L}^P , \mathcal{L}_1 :

$$\mathcal{L}_1 = \mathcal{L}^P = \sum_{\substack{(m_i, m_{i+}) \in \{Pos\} \\ (m_j, m_{j-}) \in \{Neg\}}} \max\{\mathcal{D}(\mathbf{h}_{m_i}, \mathbf{h}_{m_{i+}}) - \mathcal{D}(\mathbf{h}_{m_j}, \mathbf{h}_{m_{j-}}) + a, 0\} \quad (12)$$

其中， \mathcal{L}_1 是第一阶段损失， $\{Pos\}$ 是来自相同事件簇的样本对， $\{Neg\}$ 是来自不同事件簇的样本对。 a 是边界距离超参数， $\mathcal{D}(\cdot)$ 计算欧氏距离， \mathbf{h} 是由GNN编码器所学习到的消息表征。

Stage II: 低资源客户端预训练。该阶段目标是利用阶段一所训练出的本地模型来协助低资源客户端训练初始模型。我们提出了以全局模型为导向的过程监督及以全局锚点为导向的结果监督来实现知识的迁移。具体来说，对于富资源客户端，冻结本地模型，并以本地模型作为教师模型，全局模型作为学生模型。学生模型学习富资源客户端数据特征的同时，也将学习教师模型的隐藏层输出（Local-To-Global策略）。在低资源客户端，冻结全局模型，以全局模型作为教师模型，本地模型作为学生模型，学生模型从教师模型中学习通用知识的同时，利用本地数据学习适合本地数据特征的知识（Global-To-Local策略）。我们将以上两种策略作为以过程为导向的监督。并引入以全局锚点为导向的结果监督来统一各个客户端的簇心。从结果导向和过程导向两个方面进行知识迁移。以下将介绍这两种监督方式及该阶段总损失。

在以过程为导向的监督学习中，学生模型可以根据本地数据特点，灵活地学习教师模型中特定或若干隐藏层的输出。客户端 k 上，各个客户端学生模型蒸馏损失 \mathcal{L}_k^{KD} 可表示为:

$$\mathcal{L}_k^{KD} = \frac{1}{I} \sum_{i=1}^I \text{MSE}(\mathbf{H}_k^{t,i}; \mathbf{H}_k^{s,i}) \quad (13)$$

其中，MSE表示均方误差， I 代表学生模型要向教师模型学习的隐藏层个数， $\mathbf{H}_k^{t,i}, \mathbf{H}_k^{s,i}$ 分别代表教师模型、学生模型第 i 个隐藏层的输出。

在本地客户端训练时，我们期待每个特征可以与相应全局锚点更接近，同时远离非相应全局锚点，我们利用对比学习来协调所有客户端的本地中心，以便在客户端之间交换信息。全局锚点以及客户端上模型可分别表示为 $\mathbf{A}_k^{(t)} = \{\mathbf{a}_c^{(t)}\}_{c=1}^C$ ， $\mathbf{W}_k^{(t)}$ ，对于第 i 块的数据采样 x ，锚点对齐损失 \mathcal{L}_k^A 为：

$$\mathcal{L}_k^A = \mathcal{L}_{CE}(s, c) \quad (14)$$

$$s_n = \frac{\exp\left(\langle \mathbf{a}_n^{(t)}, f(W_k^{(t)}, x) / \alpha \rangle\right)}{\sum_{i=1}^C \exp\left(\langle \mathbf{a}_i^{(t)}, f(W_k^{(t)}, x) / \alpha \rangle\right)} \quad (15)$$

其中 \mathcal{L}_{CE} 为交叉熵损失， $s = [s_1, s_2, \dots, s_c]$ 是一个相似度向量， s_n 可表示当前特征与各个锚点的相似度， α 决定浓度的水平， $\langle \cdot, \cdot \rangle$ 是内积。最小化交叉熵损失既最大限度地提高特征与其对应的锚点 $\mathbf{a}_c^{(t)}$ 之间的相似性，又最小化该特征与 $\{\mathbf{a}_n^{(t)}\}_{n \neq c}$ 中每个非对应锚点之间的相似性。

阶段二中参与聚合的客户端集合 $S=H$ （公式(10),(11)），即仅仅富资源客户端参与聚合。客户端 k 上，该阶段仅仅训练各个客户端上学生模型，该阶段总损失 \mathcal{L}_2 可表示为：

$$\mathcal{L}_2 = \mathcal{L}_k^P + \mathcal{L}_k^{KD} + \mathcal{L}_k^A \quad (16)$$

Stage III: 各个客户端检测。在检测阶段，我们根据所训练出来的模型对 $[M_i, M_{i+w}]$ 窗口内的消息块进行聚类。基于距离的聚类算法（如K-Means）或基于密度的算法（如HDBSCAN(Campello et al., 2013)）都可以对消息进行聚类。其中，HDBSCAN不需要指定类别的总数，非常适合增量的检测需求。因此，我们利用HDBSCAN算法进行消息检测。

Stage IV: 各个客户端联合训练（维护阶段）。该阶段，受Wu et al. (2022)工作的启发，我们设计了基于任务特定损失引导的共蒸馏机制来适应在线的场景。在此过程中，各个客户端的本地模型以及全局模型均参与模型训练，相互学习并适应在线场景下的数据分布和模式变化。由于本地模型比全局模型更加适应本地数据，本地模型编码的有用知识可以帮助指导全局模型。此外，本地模型只能从本地数据中学习，而全局模型融合了所有客户端的数据特性。因此，本地模型还可以从全局模型蒸馏的知识中受益。所以，本地模型和全局模型之间可以相互学习。基于此，该阶段我们设计了基于任务特定损失引导的共蒸馏机制，使本地模型及全局模型相互学习对方有用的知识。针对于全局模型和本地模型，蒸馏损失可表示为：

$$\mathcal{L}_{k,l}^H = \mathcal{L}_{k,g}^H = \frac{\frac{1}{I} \sum_{i=1}^I \text{MSE}(\mathbf{H}_k^{t,i}; \mathbf{H}_k^{s,i})}{\mathcal{L}_{k,l}^P + \mathcal{L}_{k,g}^P} \quad (17)$$

该阶段中参与聚合的客户端集合 $S = \{L \cup H\}$ （公式(10),(11)），即所有客户端均参与聚合。该阶段总损失包含任务特定损失、基于任务特定损失的共蒸馏损失以及锚点对齐损失。该阶段本地模型、全局模型的损失分别可表示为 $\mathcal{L}_{l,3}$ 、 $\mathcal{L}_{g,3}$ 。

$$\mathcal{L}_{l,3} = \mathcal{L}_{k,l}^P + \mathcal{L}_{k,l}^H + \mathcal{L}_{k,l}^A \quad (18)$$

$$\mathcal{L}_{g,3} = \mathcal{L}_{k,g}^P + \mathcal{L}_{k,g}^H + \mathcal{L}_{k,g}^A \quad (19)$$

其中， $\mathcal{L}_{l,3}$ 、 $\mathcal{L}_{g,3}$ 分别代表该阶段本地模型及全局模型的总训练损失。 $\mathcal{L}_{k,l}^P$ 、 $\mathcal{L}_{k,g}^P$ 分别代表该阶段本地模型和全局模型的任务特定损失（公式(12)）。 $\mathcal{L}_{k,l}^H$ 、 $\mathcal{L}_{k,g}^H$ 分别代表该阶段本地模型和全局模型的共蒸馏损失（公式(17)）。 $\mathcal{L}_{k,l}^A$ 、 $\mathcal{L}_{k,g}^A$ 分别代表该阶段本地模型和全局模型的锚点对齐损失（公式(14)）。

3 实验

3.1 数据集及评价指标

(1) 实验数据集：

为了评估FedEvent框架的性能，本研究在一个大型公开的英文数据集Events2012(McMinn et al., 2013)的基础上，根据其主题描述补充了中文和越南语事件样本，构建了低资源客户端上私有数据集。这项工作旨在创建一个多客户端、跨语言的实验环境。首先，我们依据twitter高

质量筛选器指南¹，通过关键词、时间范围和语言限制进行初步的数据筛选。例如，使用查询语句“(莫言AND 奖) since:2012-10-12 until:2012-10-13 lang:zh”可以获得2012年10月12日至13日期间，包含关键词“莫言”和“奖”的中文推文。接下来，我们通过人工审核和数据清洗过程来确保数据的质量。对于文本实体的抽取，我们分别采用了spacy²(对于中文和英文)和Underthesea³(对于越南语)。该数据集各个消息块统计信息如表1所示。

消息块	M_0	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}
中文消息个数	3765	1151	348	468	316	405	300	1022	108	132	238
越南语消息个数	2716	1270	207	239	244	249	145	743	181	146	113
消息块	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	M_{21}
中文消息个数	234	297	204	714	248	175	757	346	290	101	267
越南语消息个数	140	426	264	387	325	113	350	246	155	80	292

表 1. 数据集统计情况

(2)评价指标:

本研究采用标准化互信息 (NMI) 和调整互信息 (AMI) 作为评价指标，以确保与以往社交媒体事件检测工作的可比性。

3.2 实验参数设置

在本研究中，跨语言模块训练基于Hugging Face平台提供的平行语料库⁴，其中，LoRA的秩 r 设置为4。为保证实验的公正性，除mBERT外的所有基线方法均使用同一跨语言模块(\mathcal{M})进行编码。全局模型采用双层图注意力网络 (GAT)，注意力头设置为4，隐藏层和输出层维度均设为32。第一阶段，各个客户端本地迭代次数设置为2，轮次设置为10。第二阶段，轮次减至5。第三阶段中窗口大小 w 设为3。第四阶段，轮次设置为10。各个阶段中模型学习率设置为0.001，批次大小设置为1000，成对损失的系数 α 取10，优化器选用Adam(Kingma and Ba, 2014)。所有实验重复5次，来记录结果的均值。聚类算法HDBSCAN中核心样本的密度阈值 (min_samples) 设置为3，最小聚类大小(min_cluster_size)设置为10。

3.3 基线方法

本文将所提出的FedEvent方法与下述社交事件检测方法进行离线和在线场景下的比较。

- mBERT(Pires et al., 2019): 利用mBERT(Pires et al., 2019)得到的768维句子嵌入作为消息的表示，以捕捉深层次语义信息。
- KPGNN(Cao et al., 2021): 通过图注意力网络 (GAT) 结合三元损失函数获取语义及结构上的消息表示。
- QSGNN(Ren et al., 2022a):通过最佳的已知样本和可靠的知识转移，将知识从已知扩展到未知。
- FedKP: 我们将FedAvg(McMahan et al., 2017)和KPGNN(Cao et al., 2021)进行结合，让每个客户端都拥有一个KPGNN主干模型。
- FedQS: 我们将FedAvg(McMahan et al., 2017)中各个客户端主干模型设为QSGNN(Ren et al., 2022a)。
- CLKD(Ren et al., 2023): 利用中心化的训练方式将富资源语言上训练的教师模型中的特征知识以及关系知识蒸馏到低资源语言的学生模型中。

3.4 主实验

为了评估本文提出的模型的性能，我们在每个客户端上进行了离线评估以及在线评估。

¹<https://developer.twitter.com/zh-cn/docs/tutorials/building-high-quality-filters>

²<https://spacy.io/api/annotation#section-named-entities>

³<https://underthesea.readthedocs.io/en/latest/readme.html>

⁴越英平行语料:<https://huggingface.co/datasets/larryvrh/WikiMatrix-v1-En-Zh-filtered>，中英平行语料:https://huggingface.co/datasets/harouzie/vi_en-translation

我们使用该客户端数据集中第一周的消息构建初始消息块 M_0 ，并将随后21天的数据划分成消息块 M_1, M_2, \dots, M_{21} 。我们在这两种场景下做了对比实验以及消融实验。实验结果记录在表2,3,4,5中。

Metrics	$mBERT_c$		$KPGNN_c$		$QSGNN_c$		$FedKP_c$		$FedQS_c$		\mathcal{M}_a		$w/o NC_a$		$w/o NA_a$		$FedEvent$	
	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi
NMI	.23	.01	.59	.42	.62	.44	.63	.47	.64	.53	.41	.13	.66	.45	.68	.60	.70	.63
AMI	.14	.01	.47	.31	.51	.35	.52	.37	.53	.36	.30	.12	.53	.34	.58	.45	.62	.45

表 2. 低资源客户端离线评估结果。其中，zh代表中文客户端，vi代表越南语客户端，带脚标 c 的方法用于对比实验，带脚标 a 的方法用于消融实验。 $w/o NC_a$ 、 $w/o NA_a$ 分别代表本方法中去除跨语言词嵌入模块、锚点对齐损失。

离线评估：对于每个客户端，我们将该客户端上的初始消息块 M_0 按照70%、20%和10%划分为训练集、测试集以及验证集。从表2中带有脚标 c 的实验结果可以看出，在离线场景下，多机环境（FedEvent, FedKP等）优于单机环境下的事件检测方法，证明了多机环境可以提升单机环境下模型的性能。究其原因，是因为联邦框架可以整合来自各个客户端上的知识。具体而言，在中文客户端上，多机环境相比于单机环境下的方法在NMI以及AMI上分别表现出了1.6%-204.3%，2.0%-342.9%的显著提升，在越南语客户端上，分别提升了2.3%-6200%，0%-4400%。引入联邦后，FedKP、FedQS方法虽然效果也有提升，但并不明显，是因为各个客户端上的特征空间不一致以及不同语言之间的差异。而FedEvent利用全局锚点来统一各个客户端上的特征空间，利用本地模型来适应本地私有数据的特性，并利用全局模型来传递全局的知识。所以，FedEvent在处理分布式社交事件检测任务时，更适合来实现富资源客户端知识迁移至低资源客户端。具体而言，与FedKP、FedQS相比，FedEvent在NMI和AMI指标上分别提升了11.1%-34.0%，17.0%-28.6%，展现了我们框架的有效性。

Blocks	M_1		M_2		M_3		M_4		M_5		M_6		M_7															
	NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI															
	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi														
$mBERT_c$.15	.01	.10	.01	.22	.56	.19	.46	.10	.10	.08	.01	.09	.02	.01	.06	.07	.03	.05	.01	.40	.42	.34	.33	.02	.02	.01	.01
$KPGNN_c$.54	.49	.49	.47	.75	.79	.71	.74	.55	.71	.53	.65	.64	.62	.58	.56	.69	.65	.74	.56	.78	.69	.75	.60	.46	.46	.43	.42
$QSGNN_c$.57	.51	.53	.47	.73	.74	.68	.69	.54	.72	.51	.69	.65	.59	.59	.50	.68	.66	.64	.58	.70	.67	.64	.65	.47	.40	.45	.36
$FedKP_c$.58	.55	.54	.53	.74	.77	.69	.73	.58	.80	.55	.78	.65	.65	.58	.59	.74	.69	.71	.61	.84	.73	.75	.72	.53	.55	.52	.51
$FedQS_c$.60	.56	.53	.54	.70	.76	.61	.71	.59	.79	.56	.71	.57	.51	.48	.41	.71	.53	.68	.42	.79	.72	.77	.66	.54	.39	.51	.32
\mathcal{M}_a	.39	.14	.33	.13	.57	.57	.49	.47	.39	.21	.31	.16	.50	.47	.43	.42	.40	.34	.31	.28	.62	.56	.58	.47	.01	.15	.01	.12
$w/o NC_a$.51	.45	.49	.40	.70	.68	.66	.63	.63	.78	.60	.73	.63	.67	.60	.61	.69	.63	.66	.63	.82	.71	.80	.64	.50	.48	.47	.42
$w/o NA_a$.56	.55	.50	.54	.79	.75	.75	.70	.66	.82	.61	.78	.67	.62	.61	.55	.75	.67	.72	.59	.89	.77	.86	.71	.53	.55	.50	.51
FedEvent	.61	.62	.55	.57	.81	.82	.76	.79	.70	.87	.66	.84	.70	.65	.65	.58	.74	.69	.69	.60	.91	.80	.89	.73	.58	.58	.55	.54

表 3. 低资源客户端在线评估结果 ($M_1 - M_7$)

在线评估：在线场景中，实验结果如表3,4,5所示，我们的方法在在线场景下大多都超过了单机环境和多机环境下的所有方法。在线评估中，我们观察到不同块之间性能差异明显，可能是因为不同消息块中消息数量有着较大差异以及消息块中数据的长尾分布。

3.5 对照实验

为了探究我们提出的去中心化蒸馏方法FedEvent与目前最先进的中心化蒸馏方法（CLKD）之间的性能差异，我们设计了一组对照实验。实验结果如表6所示。表格中的数据代表低资源客户端（中越）模型在消息块 M_i 上评估结果（NMI、AMI）的加权平均。实验结果表明，本文所提出的方法与最先进的中心化方法（CLKD）在性能上非常接近，总体上仅下降了0.02，这进一步证明了我们方法的显著优势。主要涉及两个方面的原因，一方面，FedEvent通过锚点对齐损失来统一各个客户端上的簇心，使得全局模型拥有全局一致的知识。在此基础上，本地模型会结合全局知识以及本地数据进行微调，从而使其更加符合本地数

Blocks	M_8		M_9		M_{10}		M_{11}		M_{12}		M_{13}		M_{14}															
	NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI															
	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi														
mBERT_c	.28	.18	.23	.13	.10	.09	.03	.03	.33	.18	.27	.13	.29	.47	.23	.41	.33	.02	.29	.01	.42	.44	.37	.39	.27	.38	.20	.31
KPGNN_c	.72	.75	.65	.66	.74	.71	.69	.64	.80	.82	.86	.79	.74	.52	.70	.50	.75	.61	.70	.53	.75	.75	.74	.72	.41	.58	.37	.52
QSGNN_c	.74	.71	.68	.64	.79	.71	.76	.63	.81	.79	.83	.76	.71	.53	.67	.49	.71	.44	.65	.37	.66	.55	.63	.49	.42	.62	.38	.56
FedKP_c	.77	.78	.72	.71	.79	.78	.74	.72	.83	.83	.81	.80	.71	.54	.66	.45	.81	.68	.77	.61	.79	.80	.77	.78	.44	.64	.36	.57
FedQS_c	.78	.75	.71	.70	.79	.76	.78	.70	.84	.80	.80	.74	.72	.57	.68	.48	.73	.66	.67	.60	.70	.64	.68	.60	.40	.61	.31	.55
\mathcal{M}_a	.54	.41	.49	.34	.42	.51	.36	.41	.52	.51	.49	.43	.46	.45	.39	.38	.45	.33	.39	.26	.62	.36	.58	.29	.33	.42	.26	.35
w/o NC _a	.65	.75	.60	.66	.83	.77	.80	.72	.79	.80	.77	.78	.70	.59	.65	.51	.76	.62	.72	.52	.77	.78	.75	.75	.43	.57	.37	.53
w/o NA _a	.79	.78	.72	.71	.85	.80	.82	.75	.84	.82	.83	.79	.74	.55	.70	.46	.77	.66	.72	.60	.89	.78	.86	.75	.43	.62	.36	.57
FedEvent	.85	.84	.79	.78	.85	.85	.82	.81	.87	.93	.86	.92	.79	.68	.72	.64	.78	.72	.73	.65	.91	.87	.90	.82	.47	.65	.41	.58

表 4. 低资源客户端在线评估结果 ($M_8 - M_{14}$)

Blocks	M_{15}		M_{16}		M_{17}		M_{18}		M_{19}		M_{20}		M_{21}															
	NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI		NMI AMI															
	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi	zh	vi														
mBERT_c	.34	.32	.29	.28	.32	.29	.28	.24	.25	.03	.19	.01	.36	.21	.33	.16	.04	.06	.02	.03	.42	.32	.34	.26	.03	.39	.01	.33
KPGNN_c	.53	.60	.47	.54	.61	.74	.58	.71	.53	.73	.51	.70	.51	.37	.44	.32	.77	.75	.75	.79	.73	.72	.69	.66	.48	.56	.42	.50
QSGNN_c	.54	.61	.51	.56	.62	.74	.61	.69	.44	.71	.42	.68	.45	.37	.42	.33	.73	.61	.71	.53	.73	.53	.69	.47	.43	.47	.40	.40
FedKP_c	.59	.65	.56	.59	.89	.85	.88	.82	.52	.73	.47	.70	.66	.49	.62	.44	.79	.68	.77	.63	.78	.78	.75	.75	.52	.63	.47	.57
FedQS_c	.55	.53	.46	.44	.80	.81	.76	.79	.48	.58	.46	.52	.62	.34	.57	.29	.75	.69	.70	.63	.67	.57	.58	.50	.52	.54	.50	.50
\mathcal{M}_a	.47	.41	.39	.32	.59	.43	.56	.35	.36	.50	.31	.44	.32	.28	.28	.23	.51	.29	.46	.23	.51	.29	.42	.31	.03	.29	.01	.24
w/o NC _a	.55	.61	.55	.55	.65	.79	.62	.76	.56	.63	.52	.60	.55	.47	.48	.42	.74	.73	.72	.68	.74	.70	.68	.67	.57	.57	.50	.52
w/o NA _a	.61	.64	.46	.58	.89	.77	.87	.73	.56	.75	.53	.70	.70	.58	.67	.49	.82	.75	.81	.70	.78	.81	.72	.78	.59	.59	.51	.52
FedEvent	.70	.67	.63	.60	.83	.80	.80	.76	.63	.83	.58	.78	.73	.60	.69	.51	.75	.83	.73	.80	.80	.88	.77	.83	.62	.66	.55	.63

表 5. 低资源客户端在线评估结果 ($M_{15} - M_{21}$)

据的特性。另一方面，本方法采用了自适应蒸馏策略，来选择性的获取对方有益的知识，从而避免对无益的知识进行传输。

Blcoks	M_0	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}	M_{16}	M_{17}	M_{18}	M_{19}	M_{20}	M_{21}
CLKD	.57	.57	.79	.78	.65	.69	.82	.58	.84	.85	.86	.72	.72	.86	.55	.67	.82	.72	.62	.78	.81	.61
FedEvent	.60	.59	.80	.77	.65	.68	.83	.56	.81	.83	.90	.71	.72	.87	.53	.65	.80	.71	.63	.78	.82	.62

表 6. 去中心化与中心化蒸馏方法对照实验结果 ($M_0 - M_{21}$)

3.6 消融实验

为了验证本文方法中各个模块的有效性，本文设置了以下两组消融实验：跨语言词嵌入模块和锚点对齐损失对模型性能的影响。通过对比表2,3,4,5中带有脚标 a 的实验结果，我们观察到，无论是在离线还是在线场景下，跨语言词嵌入模块和锚点对齐损失都显著提升了模型性能。跨语言词嵌入模块 \mathcal{M} 与mBERT相比， \mathcal{M} 明显缓解了mBERT在低资源语言文本表征上的不足，具体而言，在中文客户端上，跨语言词嵌入模块 \mathcal{M} 相比于mBERT在NMI和AMI指标上平均提升了149.4%，在越南语客户端上，平均提升了322.3%。进一步地，FedEvent通过引入跨语言词嵌入，与未加入该模块（即w/o NC）的方法相比，在中文客户端上，平均提升了2.4%，在越南语客户端上，平均提升了4.3%。由于引入跨语言词嵌入模块后，FedEvent确保了不同客户端上输入特征空间的一致性，从而增强了蒸馏效果。可以发现，跨语言词嵌入模块（ \mathcal{M} ）在越南语客户端上表现尤为突出，可能是mBERT对越南语的表征能力不如对中文那样强大，利用 \mathcal{M} 对消息进行表征后，效果更加明显。

此外，在各个客户端上，相较于未加入锚点对齐损失（即w/o NA）的方法，FedEvent在NMI和AMI上平均下降了5.3%，进一步印证了锚点对齐损失的重要性。在缺

少锚点对齐损失的情况下，低资源客户端不易捕获到锚点对齐的信息，导致蒸馏效果较差。

3.7 案例分析

为了更直观地评估本文所提出方法的蒸馏效果，我们在英文和越南语的数据集 M_{17} 块上进行了案例分析。其中，英文客户端 M_{17} 块上共有2676条消息，分散在35个簇中。越南语客户端 M_{17} 块上共有350条消息，分散在14个簇中。具体来说，我们利用混淆矩阵来探究我们的方法以及KPGNN方法（单机）的效果，从而更直观地分析FedEvent获取了多少先验知识。

混淆矩阵是一种特定的表格布局，用于可视化监督学习算法的性能。在混淆矩阵中，每一行代表实际类别，每一列代表预测类别。矩阵的每个单元格包含了在该实际类别和预测类别下的样本数量，可以更直观的了解预测值与真实值之间的差异。本实验中，混淆矩阵的行数和列数与越南语客户端的类簇个数一致。考虑到基于密度的聚类方法所获得的事件数量与实际事件数量之间的差异，我们统一采用K-Means方法来获得预测标签。本实验共构建了三个混淆矩阵：第一个混淆矩阵 ($FedEvent_{en}$) 代表本文方法中富资源客户端上的真实结果与预测结果之间的关系；第二个混淆矩阵 ($FedEvent_{vi}$) 代表低资源客户端上的真实结果与预测结果之间的关系；第三个混淆矩阵 ($KPGNN_{vi}$) 代表KPGNN算法在越南语消息块 M_{17} 块上的预测结果与真实结果之间的关系。

通过三个混淆矩阵，我们发现 $FedEvent_{vi}$ 的聚类结果在对角线上大多显示出比 $KPGNN_{vi}$ 更高的值，意味着其预测结果更为集中，准确度更高。这一发现表明，即便在低资源环境下，FedEvent也能有效地帮助模型获取到一定的先验知识，从而提升性能。相反，KPGNN由于数据的限制，其模型性能受到了显著影响。进一步地，我们观察到当富资源客户端在某一类簇上的表现会直接影响低资源客户端在这一类簇上性能表现，这说明富资源客户端会将边界知识蒸馏至低资源客户端。而对于KPGNN而言，由于缺乏足够的先验知识，它在面对相似文本或共享相同元素的文本时，往往会出现判别能力不足的情况，从而导致模型识别错误。

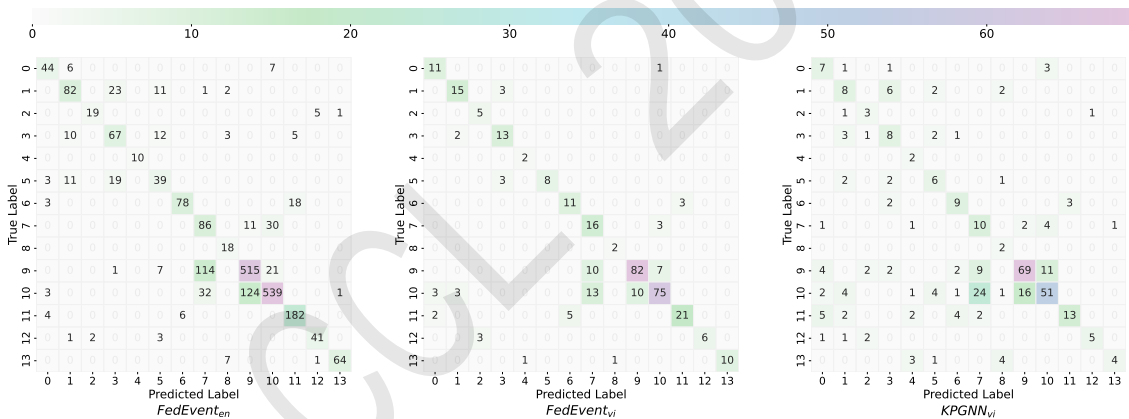


图 3. 消息块 M_{17} 混淆矩阵实验结果

3.8 可视化

为了直观地展示我们所提出的FedEvent框架的聚类效果，我们采用t-SNE(Van der Maaten and Hinton, 2008)对消息块 M_7 的表示进行了降维可视化。可视化结果如图4所示。与单机以及多机环境下的方法相比，FedEvent能够学习到更加可区分的表示。同时，观察到FedEvent比 w/o NA聚集的更加紧密，验证了锚点对齐损失的重要性。此外，跨语言词嵌入模块的引入同样也提升了模型性能。

4 结论

本文提出了基于联邦蒸馏的跨语言社交媒体事件检测框架。该框架采用生命周期机制以适应在线及离线场景下的事件检测需求，并通过过程监督以及结果监督实现知识的有效迁移。本方法通过知识蒸馏技术，有效缓解了面对低资源语言时的挑战，并利用跨语言词嵌入模块，实现了不同语言间的语义空间映射。通过实验证明了我们所提出的方法的有效性。据我们所知，

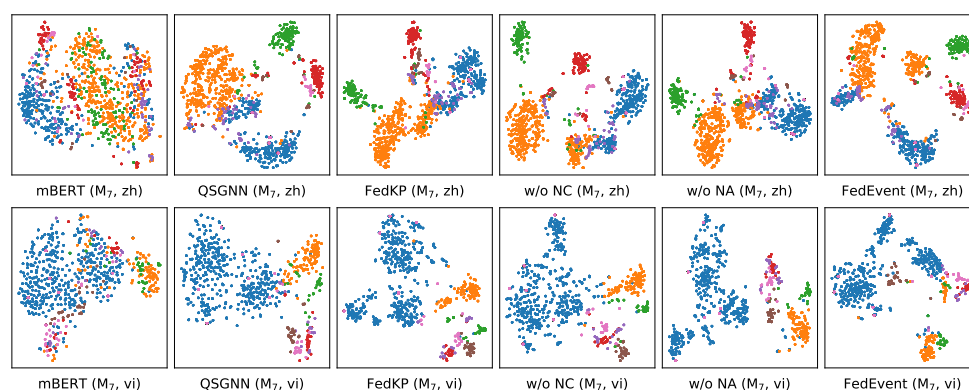


图 4. FedEvent可视化图

这是首个在联邦学习环境下，针对低资源语言环境进行社交媒体事件检测的研究。在下一步的研究工作中，我们计划进一步探索针对低资源场景下长尾数据分布的社交媒体事件检测策略。

参考文献

- Charu C Aggarwal and Karthik Subbian. 2012. Event detection in social streams. In *Proceedings of the 2012 SIAM international conference on data mining*, pages 624–635. SIAM.
- Farzindar Atefeh and Wael Khreich. 2015. A survey of techniques for event detection in twitter. *Computational Intelligence*, 31(1):132–164.
- Hila Becker, Mor Naaman, and Luis Gravano. 2011. Beyond trending topics: Real-world event identification on twitter. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 438–441.
- Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. 2013. Density-based clustering based on hierarchical density estimates. In Jian Pei, Vincent S. Tseng, Longbing Cao, Hiroshi Motoda, and Guandong Xu, editors, *Advances in Knowledge Discovery and Data Mining*, pages 160–172, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Yuwei Cao, Hao Peng, Jia Wu, Yingtong Dou, Jianxin Li, and Philip S Yu. 2021. Knowledge-preserving incremental social event detection via heterogeneous gnns. In *Proceedings of the Web Conference 2021*, pages 3383–3395.
- Xueqi Cheng, Xiaohui Yan, Yanyan Lan, and Jiafeng Guo. 2014. Btm: Topic modeling over short texts. *IEEE Transactions on Knowledge and Data Engineering*, 26(12):2928–2941.
- Hyung Won Chung, Dan Garrette, Kiat Chuan Tan, and Jason Riesa. 2020. Improving multilingual models with language-clustered vocabularies. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4536–4546, Online, November. Association for Computational Linguistics.
- Mário Cordeiro. 2012. Twitter event detection: combining wavelet analysis and topic inference summarization. In *Doctoral symposium on informatics engineering*, volume 1, pages 11–16.
- Wanqiu Cui, Junping Du, Dawei Wang, Feifei Kou, and Zhe Xue. 2021. Mvgan: Multi-view graph attention network for social event detection. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(3):1–24.
- Rui Gaspar, Cláudia Pedro, Panos Panagiotopoulos, and Beate Seibt. 2016. Beyond positive or negative: Qualitative sentiment analysis of social media reactions to unexpected stressful events. *Computers in Human Behavior*, 56:179–191.
- Linmei Hu, Bin Zhang, Lei Hou, and Juanzi Li. 2017. Adaptive online event detection in news streams. *Knowledge-Based Systems*, 138:105–112.
- J. Edward Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *ArXiv*, abs/2106.09685.

- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. [arXiv preprint arXiv:1412.6980](#).
- Bang Liu, Fred X Han, Di Niu, Linglong Kong, Kunfeng Lai, and Yu Xu. 2020a. Story forest: Extracting events and telling stories from breaking news. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(3):1–28.
- Fanzhen Liu, Shan Xue, Jia Wu, Chuan Zhou, Wenbin Hu, Cecile Paris, Surya Nepal, Jian Yang, and Philip S Yu. 2020b. Deep learning for community detection: progress, challenges and opportunities. [arXiv preprint arXiv:2005.08225](#).
- Yaopeng Liu, Hao Peng, Jianxin Li, Yangqiu Song, and Xiong Li. 2020c. Event detection and evolution in multi-lingual social streams. *Frontiers of Computer Science*, 14:1–15.
- Fabrizio Marozzo and Alessandro Bessi. 2018. Analyzing polarization of social media users and news sites during political campaigns. *Social Network Analysis and Mining*, 8:1–13.
- Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR.
- Andrew J McMinn, Yashar Moshfeghi, and Joemon M Jose. 2013. Building a large-scale corpus for evaluating event detection on twitter. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 409–418.
- Tasnim Mohiuddin, M Saiful Bari, and Shafiq Joty. 2020. LNMap: Departures from isomorphic assumption in bilingual lexicon induction through non-linear mapping in latent space. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2712–2723, Online, November. Association for Computational Linguistics.
- Duc T Nguyen and Jason J Jung. 2015. Real-time event detection on social data stream. *Mobile Networks and Applications*, 20:475–486.
- Tahir M Nisar and Man Yeung. 2018. Twitter as a tool for forecasting stock market movements: A short-window event study. *The journal of finance and data science*, 4(2):101–119.
- Ozer Ozdakis, Pinar Karagoz, and Halit Oğuztüzin. 2017. Incremental clustering with vector expansion for online event detection in microblogs. *Social Network Analysis and Mining*, 7:1–17.
- Hao Peng, Jianxin Li, Qiran Gong, Yangqiu Song, Yuanxing Ning, Kunfeng Lai, and Philip S Yu. 2019. Fine-grained event categorization with heterogeneous graph convolutional networks. [arXiv preprint arXiv:1906.04580](#).
- Hao Peng, Jianxin Li, Yangqiu Song, Renyu Yang, Rajiv Ranjan, Philip S Yu, and Lifang He. 2021. Streaming social event detection and evolution discovery in heterogeneous information networks. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 15(5):1–33.
- Hao Peng, Ruitong Zhang, Shaoning Li, Yuwei Cao, Shirui Pan, and Philip S. Yu. 2023. Reinforced, incremental and cross-lingual event detection from social messages. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1):980–998.
- Telmo Pires, Eva Schlinger, and Dan Garrette. 2019. How multilingual is multilingual BERT? In Anna Korhonen, David Traum, and Lluís Màrquez, editors, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4996–5001, Florence, Italy, July. Association for Computational Linguistics.
- Jiaqian Ren, Lei Jiang, Hao Peng, Yuwei Cao, Jia Wu, Philip S Yu, and Lifang He. 2022a. From known to unknown: quality-aware self-improving graph neural network for open set social event detection. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1696–1705.
- Jiaqian Ren, Lei Jiang, Hao Peng, Zhiwei Liu, Jia Wu, and S Yu Philip. 2022b. Evidential temporal-aware graph-based social event detection via dempster-shafer theory. In *2022 IEEE International Conference on Web Services (ICWS)*, pages 331–336. IEEE.
- Jiaqian Ren, Hao Peng, Lei Jiang, Jia Wu, Yongxin Tong, Lihong Wang, Xu Bai, Bo Wang, and Qiang Yang. 2023. Transferring knowledge distillation for multilingual social event detection.

- Xing Su, Shan Xue, Fanzhen Liu, Jia Wu, Jian Yang, Chuan Zhou, Wenbin Hu, Cecile Paris, Surya Nepal, Di Jin, et al. 2022. A comprehensive survey on community detection with deep learning. IEEE Transactions on Neural Networks and Learning Systems.
- Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. Journal of machine learning research, 9(11).
- Chuhan Wu, Fangzhao Wu, Lingjuan Lyu, Yongfeng Huang, and Xing Xie. 2022. Communication-efficient federated learning via knowledge distillation. Nature communications, 13(1):2032.
- Weiren Yu, Jianxin Li, Md Zakirul Alam Bhuiyan, Richong Zhang, and Jinpeng Huai. 2017. Ring: Real-time emerging anomaly monitoring system over text streams. IEEE Transactions on Big Data, 5(4):506–519.
- Kuo Zhang, Juan Zi, and Li Gang Wu. 2007. New event detection based on indexing-tree and named entity. In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, pages 215–222.
- Xiangmin Zhou and Lei Chen. 2014. Event detection over twitter social media streams. The VLDB journal, 23(3):381–400.