A Contrastive Learning and Prompt-Driven Framework for Low-Resource User Geolocation

Menglin Li
menglin_li@mymail.sutd.edu.sg
Singapore University of Technology and Design
Singapore, Singapore

Kwan Hui Lim
kwanhui@acm.org
Singapore University of Technology and Design
Singapore, Singapore

Abstract

Social media user geolocation is a fundamental yet challenging problem due to the scarcity of geotagged data and the heterogeneity of online user information. To address these challenges, we propose **FewUser**, a contrastive learning and prompt-driven framework for few-shot social media user geolocation. FewUser aligns user and location representations through a dual-objective framework that jointly optimizes contrastive and matching losses with hard negative mining, enabling robust geolocation under limited supervision. The model comprises a user representation module that fuses heterogeneous social media inputs via a pre-trained language model (PLM) and a lightweight user encoder, and a geographical prompting module that employs hard, soft, and semi-soft prompts to bridge PLM semantics with location-specific knowledge. To facilitate few-shot and cross-platform evaluation, we construct two new datasets, TwiU and FliU, featuring rich and standardized user- and post-level metadata. Extensive experiments on TwiU, FliU, and two public benchmarks demonstrate that FewUser consistently outperforms competitive baselines in various few-shot settings.

CCS Concepts

• **Computing methodologies** → **Machine learning**; **Representation learning**; • **Information systems** → *Geographic information systems*; *Information extraction.*

Keywords

User Geolocation, Contrastive Learning, Prompt Learning, Few-Shot Learning, Deep Learning, Social Media

ACM Reference Format:
Menglin Li and Kwan Hui Lim. 2026. A Contrastive Learning and Prompt-Driven Framework for Low-Resource User Geolocation. In *Companion Proceedings of the ACM Web Conference 2026 (WWW Companion '26), April 13–17, 2026, Dubai, United Arab Emirates.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3774905.3795858

1 Introduction

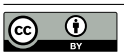
Social media user location is a critical signal for many Web applications, including web search [9, 19], personalized recommendations [5, 11], and targeted advertising [7]. Geographical information also provides valuable context for analyzing regional opinions and social trends on the Web, such as discussions of COVID-19 and presidential elections [17]. However, most users do not explicitly disclose their locations, and fine-grained signals such as GPS tags or IP addresses are typically unavailable to third-party consumers [23]. As a result, user locations must be inferred from publicly available Web data, such as social media profiles, posts, and interactions.

Social media user geolocation aims to infer a user's residence city from online traces. Although this task has been widely studied [1, 16, 20], several limitations hinder its applicability in real-world, few-shot settings. Most existing methods rely heavily on large-scale labeled data and suffer sharp performance degradation when training samples are scarce [36]. In practice, user activity follows a long-tail distribution: many users are passive or provide limited information, while only a small number of cities are well represented. This naturally gives rise to few-shot scenarios that remain underexplored. To address data sparsity and improve generalization, we argue that geolocation models should learn transferable representations rather than depend solely on supervised classification. Accordingly, we adopt a contrastive learning objective between users and cities to capture transferable similarity structures, combined with hard negative mining to improve discrimination among geographically similar locations.

Another challenge lies in how user information is represented. Prior methods often concatenate multiple posts into a single sequence, overlooking heterogeneous metadata such as profile descriptions and posting timestamps. This simplification limits the model's ability to capture informative geolocation cues. To address this issue, we design a user representation module that systematically selects, integrates, and fuses heterogeneous social media features from user profiles, posts, and metadata, enabling richer representations even under limited supervision.

A further limitation is the weak utilization and alignment of geographical semantics. Classification-based approaches treat city names as discrete labels, ignoring the rich semantic information they convey, such as linguistic, cultural, and regional cues. Moreover, the semantic space of PLMs is not inherently aligned with geographical concepts, resulting in a domain gap between textual representations and location semantics. To mitigate this, we incorporate city names directly into contrastive learning and introduce a geographical prompting module with hard, soft, and semi-soft templates to align PLM representations with location semantics, improving generalization to rare locations.

Finally, existing benchmark datasets, such as GeoText and TwitterUS, are limited in scale and metadata diversity, restricting the study of user-level representation learning and cross-platform generalization. Most prior work relies on uniform input formats [16, 22] or text-only data that simply concatenate posts, while metadata-aware approaches [18, 37] are often ad hoc and difficult to reproduce.



WWW Companion '26, Dubai, United Arab Emirates

ACM ISBN 979-8-4007-2308-7/2026/04
https://doi.org/10.1145/3774905.3795858

To address these gaps, we construct two new user-level datasets[1], **TwiU** (Twitter-based) and **FliU** (Flickr-based), which provide rich and standardized metadata from both user profiles and posts. These datasets enable systematic investigation of heterogeneous user representations and cross-platform evaluation between text-centric and image-centric social platforms.

Building on these insights, we propose **FewUser**, a contrastive and prompt-driven framework for **Few**-shot social **User** geolocation. FewUser integrates user representation learning, prompt-based semantic alignment, and metadata-aware input modeling. By combining prompt-driven contrastive learning with hard negative mining, FewUser enables fine-grained user–location alignment and robust generalization under few-shot settings.

Our contributions are as follows:

- We formulate the few-shot social media user geolocation problem and introduce two new datasets, TwiU and FliU, that support user-level and cross-platform evaluation.
- We propose FewUser, a contrastive and prompt-driven framework that integrates heterogeneous user representations and aligns PLM-based embeddings with geographical semantics.
- We conduct extensive experiments across four datasets and multiple supervision levels, demonstrating that FewUser consistently outperforms strong baselines under few-shot scenarios.

2 Related Work

2.1 Social Media User Geolocation

Social media user geolocation, which aims to predict users' locations from social media data, has attracted substantial attention from both academia and industry, resulting in numerous studies and shared tasks [2, 24, 32]. Most prior work is developed on widely used benchmark datasets such as GeoText, TwitterUS, and TwitterWorld, where users are represented by concatenating their tweets without incorporating additional metadata [6]. As a result, the input design of these approaches is often unnecessarily uniform [16, 22]. Although some studies explore richer input representations using external corpora [37], partial metadata [6], or even weather information [18], these designs are frequently ad hoc and exhibit limited generalizability. In contrast, our work addresses this gap by constructing datasets with rich user- and post-level metadata and systematically examining the impact of input design choices on geolocation performance (Section 4.4).

User geolocation approaches have evolved from relatively simple neural models [1, 23, 28] to more complex architectures [15, 20, 36]. MetaGeo [36], for example, adopts a meta-learning strategy via an ensemble of sub-tasks. More recently, PLM-based methods have become increasingly popular [13, 17], yet they continue to frame geolocation as a conventional classification problem. To date, contrastive learning has not been applied to social media user geolocation, with its use limited to related tasks such as post geolocation, including ba*p*tti [10] and ContrastGeo [12]. Importantly, user geolocation differs fundamentally from post geolocation, as it involves more heterogeneous inputs and typically targets broader geographic scopes [34]. Our work departs from prior approaches by introducing an end-to-end contrastive framework for user geolocation, enabling robust few-shot generalization.

[1]Codes and data are publicly available at https://github.com/lazylml/FewUser-public.

2.2 Prompting Learning

Prompt learning frames downstream tasks using natural language prompts, allowing models to leverage pre-trained knowledge effectively [14]. Early approaches relied on manually designed prompts crafted by human experts [8], demonstrating strong performance in few-shot settings [21, 33]. To reduce reliance on human expertise, subsequent work proposed automated prompt optimization methods such as AutoPrompt [27] and OptiPrompt [35], which aim to discover effective prompts in a scalable manner. In this work, we combine manually designed and automatically generated prompts to guide PLM-based encoders and bridge the semantic gap between textual representations and geographical knowledge, enabling effective user geolocation with minimal training data.

3 Our Approach: FewUser

The overall architecture of FewUser is shown in Figure 1, comprising two shared text encoders and a user encoder. For effective representation learning, FewUser integrates a user representation module for encoding social media inputs, alongside a geographical prompting module for enhancing location representations. The model is then trained using a dual-objective strategy that combines a contrastive loss with a matching loss, further refined through hard negative mining to improve its discriminative ability.

3.1 User Representation

The user representation module encodes heterogeneous social media information into a unified embedding for each user. It consists of three steps: selecting informative inputs, integrating user and post fields into sentence-level representations, and fusing them through a user encoder. This design enables FewUser to effectively utilize both textual and metadata features while adapting to different input richness and data scales.

Input Selection. Before encoding, we selectively sample user information to control input richness. Each user contains profile-level metadata (e.g., description, location, language) and a sequence of historical tweets with post-level metadata (e.g., source, timestamps).We vary both the number of tweets T and the subset of metadata fields to examine their influence on user representation quality. This selection design enables FewUser to flexibly scale from text-only to metadata-rich settings and supports our analysis of how different user signals affect geolocation performance, as detailed in Section 4.3.

Integration. For each user, even with a small number of tweets (e.g., $T = 6$), the total number of input fields can exceed thirty. To manage such high dimensionality while maintaining efficiency, we thus design six integration strategies, as illustrated in Figure 2.

- **In1**: Concatenate all input fields into a single sentence.
- **In2**: Concatenate user profile fields into one sentence and tweet-related fields into another, resulting in two sentences.
- **InT**: Concatenate all user profile fields and combine fields for each tweet into T sentences, yielding $T + 1$ sentences.
- **InUser+1**: Concatenate tweet-related fields into a single sentence while keeping user-related fields separate.

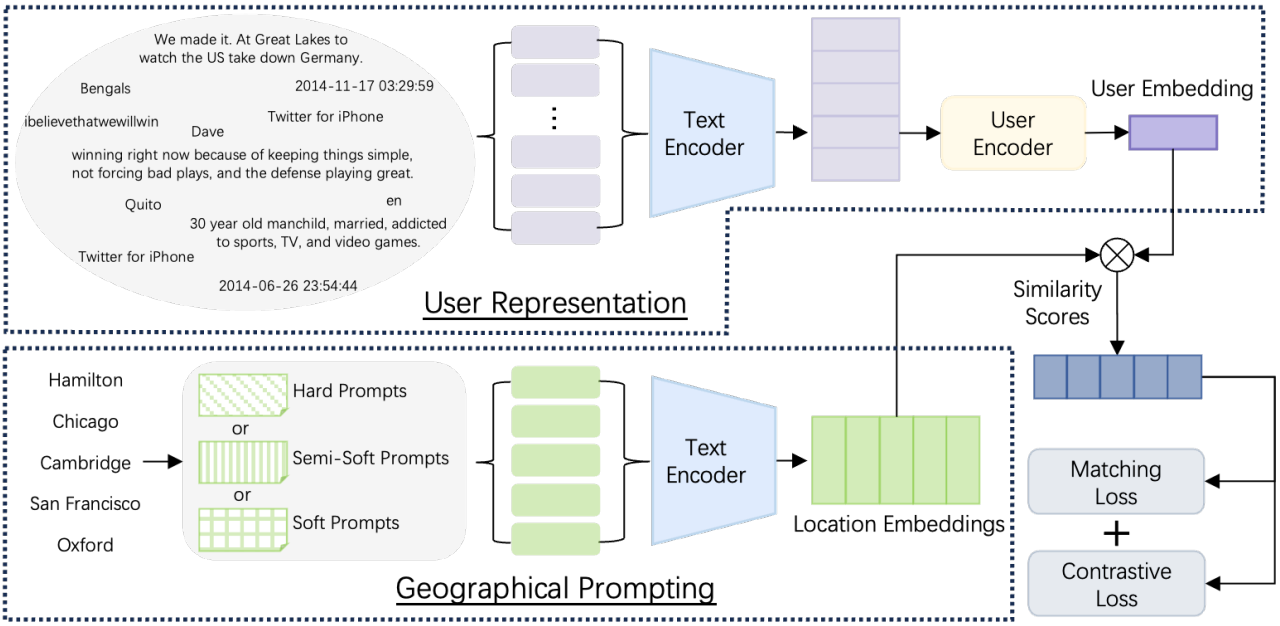


Figure 1: The Overall Architecture of FewUser for Few-Shot User Geolocation.

- **InUser+T**: Keep the original user-related fields and combine fields related to each tweet into separate sentences.
- **NoIn**: No concatenation of fields.

After integration, we obtain N input sentences, where N depends on the integration strategy. These sentences are encoded by a PLM, and their `[CLS]` tokens form a feature matrix $F \in \mathbb{R}^{N \times H}$, where H is the hidden size of the encoder.

Fusion. To capture feature-wise dependencies among heterogeneous inputs, the feature matrix F is further processed by a user encoder followed by mean pooling, producing a unified user embedding $u \in \mathbb{R}^H$. We explore three types of user encoders. First are time-sensitive models, like BiLSTM, which capture temporal dependencies across posts and time-related metadata. Second are time-insensitive models, including Transformer, Adapter, and MLP layers, which focus on feature-level correlations independent of time. Third is a parameter-free model, mean pooling, a simple yet effective fusion strategy that maintains efficiency and stability under few-shot settings.

This systematic pipeline, covering input selection, integration, and feature fusion, enables FewUser to flexibly utilize metadata-rich user information while maintaining scalability across platforms and resource settings.

3.2 Geographical Prompting

To enhance the semantic alignment between user and location representations, we introduce a geographical prompting module with hard, soft, and semi-soft prompt templates. This module injects geographical context into the text encoder, which enables it to better capture city-level semantics and align the embedding space of users and locations.

Hard Prompts. We define hard prompts based on fixed task-specific textual templates. The original prompt follows the task description of user geolocation, such as: "A user resides in [CLASS]." To enhance lexical diversity while maintaining fidelity to the original prompt, we generate prompts through paraphrasing, producing semantically similar or identical expressions, such as "A user from [CLASS]." which is paraphrased from the original prompt. The target location names are inserted into the [CLASS] token slot to apply the prompt. We also extend the prompt set with a Question-Answering (QA) format to further increase variation: "Question: Where does this user reside in? Answer: [CLASS]." All hard prompts are fixed during training and are not updated.

Soft Prompts. Hard prompts require manual design and rely on human intuition to determine suitable templates. To reduce this dependence and improve flexibility, we introduce a soft prompt mechanism, in which the prompt tokens are learnable vectors directly optimized in the continuous embedding space. Formally, a learnable prompt is defined as: $t_{prompt} = [V]_1\ [V]_2\ \cdots\ [V]_m\ [CLASS]$, where each $[V]_i \in \mathbb{R}^H$ is a trainable dense vector with the same dimension as the hidden size of the text encoder, and m denotes the number of learnable tokens, a predefined hyperparameter. These prompt tokens are randomly initialized and jointly optimized with model parameters during training. By allowing gradient-based updates, soft prompts enable the model to automatically learn latent representations that better capture geographical semantics, without requiring any manually crafted template.

Semi-Soft Prompts. While soft prompts provide flexibility, their random initialization may lead to unstable training and suboptimal convergence. To leverage linguistic priors from natural language, we further design a semi-soft prompt strategy that combines the interpretability of hard prompts with the adaptability of soft prompts. Specifically, the number and layout of learnable tokens $[V]_i$ are

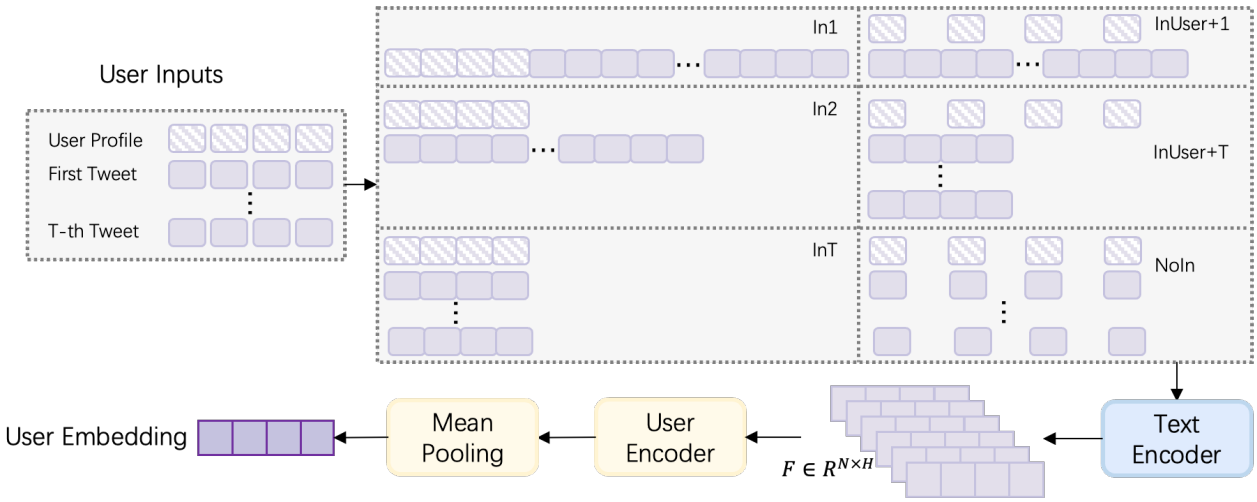


Figure 2: User Representation Module.

aligned with the word tokens in a hard prompt template (e.g., "A user resides in [CLASS]."), and each $[V]_i$ is initialized using the embedding of its corresponding token. These vectors are then fine-tuned during training, enabling the model to start from meaningful lexical representations while still benefiting from continuous optimization. In practice, the learnable tokens are registered as additional embeddings in the tokenizer and concatenated with the class token $[CLASS]$ to form the label representation used in the contrastive and matching objectives.

3.3 Training Objectives

Contrastive learning is implemented through a contrastive loss and a matching loss based on user and location representations.

Contrastive Loss is designed to minimize the distance between positive pairs and maximize the distance between negative pairs. Let u denote a user, and its embedding be obtained via the user representation module $g_u(u)$. For a dataset of K locations, each location l_j is represented by its name and encoded using a transformation function $g_l(l_j)$ to generate dense embeddings as location representations. We define the ground-truth label of the user u as l^+, with a positive user-location pair as (u, l^+) and negative pairs as $(u, l)\ \forall\ l \neq l^+$. Following InfoNCE [4], we utilize the dot product as a measure of similarity and define contrastive loss as:

$$\mathcal{L}_{contrast} = -\log \frac{exp(g_u(u) \cdot g_l(l^+)/\tau)}{\sum_{j=1}^{K} exp(g_u(u) \cdot g_l(l_j)/\tau)}, \tag{1}$$

where τ is a temperature hyperparameter. This loss can be interpreted as the log loss of a K-way softmax-based classifier that aims to classify u as l^+.

Matching Loss is designed to determine whether a user-location pair is positive (matched) or negative (unmatched). To compute this loss, we first construct a joint representation by a feature-wise concatenation of the user embedding and location embedding, followed by passing this concatenated vector through a bottleneck adapter to obtain a fused representation. A classification head, comprising a fully connected layer followed by a softmax function, then computes a probability distribution $\mathbf{p}$ over $(k+1)$ classes, where k represents the number of hard negatives. The matching loss is defined as the cross entropy loss between the predicted probability and the true distribution:

$$\mathcal{L}_{match} = \mathbb{E}_{(u,l)\sim\mathcal{D}}[F_{cross-entropy}(\mathbf{y}(u,l), \mathbf{p}(u,l))], \tag{2}$$

where $\mathbf{y}$ is a $(k+1)$-dimensional one-hot vector representing the ground-truth label and $F_{cross-entropy}$ is the cross entropy loss function.

Hard Negative Mining. To select k hard negatives from the $K-1$ locations (excluding the ground-truth location l^+), we present two approaches: **multinomial** and **top**. The multinomial approach samples negative locations based on a multinomial distribution weighted by user-location similarity, while the top approach selects the top-k most similar but incorrect locations, providing the most challenging negative examples for robust representation learning.

The final training objective combines both losses:

$$\mathcal{L} = \mathcal{L}_{contrast} + \mathcal{L}_{match}. \tag{3}$$

Inference. During inference, we utilize contrastive similarity scores between the target user and all unique prompted locations in the dataset to identify the most probable (user, location) pair, thereby determining the predicted location according to FewUser.

4 Experiment

4.1 Experiment Setting

Datasets. We conducted experiments on four social media datasets, comprising public benchmarks, TwitterUS [25] and TwitterWorld [3], as well as two newly constructed datasets, TwiU and FliU, which we publicly release. Dataset construction follows three design criteria. (i) Benchmark Continuity: We extend existing datasets while preserving user and post identities to ensure comparability with prior work; (ii) Metadata Enrichment: We retrieve all available user and post metadata via official APIs to enable multimodal and user-level modeling; (iii) Few-Shot and Cross-Platform Compatibility: We organize data at user level with standardized city labels to support

Table 1: 1-shot performance comparison across four datasets.

Model	TwiU				TwitterUS				TwitterWorld				FliU
	acc	acc@161	meanD	medD	acc	acc@161	meanD	medD	acc	acc@161	meanD	medD	acc
GeoBERT	7.02	10.21	8045	7825	3.85	6.68	1918	1714	3.67	4.97	6664	3730	6.50
GeoDare	11.32	16.43	6260	4444	6.99	8.62	1497	1036	3.93	4.91	5301	2543	7.92
HLPNN	23.29	24.24	4263	898	2.52	8.46	1567	1273	3.83	5.40	7223	6307	9.35
transTagger	8.45	15.47	7046	5886	3.22	7.40	1646	1577	2.53	4.21	6890	5286	9.46
SetFit	16.27	25.68	5138	1344	10.82	17.50	1583	1218	4.10	5.40	5190	2394	11.46
LLM	45.77	54.23	3411	60	4.46	7.65	2013	1932	4.21	5.88	5744	3197	45.75
ClassUser	12.92	16.43	5997	3513	2.86	2.86	1740	1652	4.64	5.85	4646	1706	10.04
FewUser	**58.21**	**66.67**	**1303**	**0**	**31.81**	**45.65**	**879**	**322**	**17.79**	**21.84**	**2947**	**1304**	**51.35**

Table 2: 8-shot performance comparison across four datasets.

Model	TwiU				TwitterUS				TwitterWorld				FliU
	acc	acc@161	meanD	medD	acc	acc@161	meanD	medD	acc	acc@161	meanD	medD	acc
GeoBERT	19.94	28.71	3386	906	16.56	22.82	1513	1154	6.72	9.65	4719	1707	7.18
GeoDare	15.63	22.81	5321	2727	8.71	16.81	1653	1273	3.49	4.94	4987	2379	7.61
HLPNN	38.12	43.70	3417	358	4.08	4.08	1482	1203	1.93	1.93	4543	1721	3.96
transTagger	28.07	37.16	3318	657	9.57	11.81	1776	1533	4.89	7.33	6345	4491	19.81
SetFit	51.04	61.88	887	11	31.36	45.79	832	328	16.30	20.42	2600	1264	51.93
LLM	49.44	57.26	2908	11	4.58	7.79	1950	1760	4.86	6.64	5702	3034	46.75
ClassUser	51.83	63.80	1005	13	5.99	6.74	1449	829	3.11	4.87	4874	2271	45.01
FewUser	**69.86**	**78.47**	**582**	**0**	**36.08**	**50.17**	**718**	**114**	**28.06**	**32.26**	**2317**	**1009**	**58.58**

few-shot adaptation and cross-platform evaluation. Specifically, TwiU is constructed based on WNUT16 [2], containing 7,789 users from 1,261 cities worldwide. User metadata include user name, description, account creation time, self-declared location, language, and time zone. Post metadata include creation time, posting source, and hashtags. FliU is derived from YFCC-100M [29], comprising 11,395 users from 1,688 cities across U.S. User metadata consist of user name, description, joined time, occupation, hometown, and country, while post metadata cover creation time, posting source, user tag, machine tag, and upload time.

Metrics. We evaluate user geolocation performance using three commonly used metrics in geolocation prediction: accuracy (*acc*), accuracy@161 (*acc@161*), mean distance error (*MeanD*), and median distance error (*MedD*). *acc@161* measures the proportion of correctly predicted users whose predicted locations fall within 161 km (approximately 100 miles) of their true locations. *MeanD* and *MedD* measure the average and median distances between the predicted and true locations, respectively, providing a complementary view of model accuracy, which are reported in kilometers. Note that only the FliU dataset contains no coordinate information of labels, making distance-related metric computation infeasible. However, as the primary evaluation metric, *acc* provides sufficient insight into geolocation performance.

Baselines. We benchmark FewUser against several established social media user geolocation models, including HLPNN [6] and GeoDare [23]. In addition, we include two PLM-based geolocation models, transTagger [13] and GeoBERT [26], as well as SetFit [30], a representative few-shot method that fine-tunes sentence transformers with contrastive learning. We implement a Large Language Model (LLM) baseline, following the prompting framework proposed by Xiao et al. [31], who analyzed the capability of instruction-tuned LLMs (e.g., FLAN-T5) for location prediction. We replace their models with the more recent LLaMA-3-8B-Instruct while keeping the overall few-shot prompting setup consistent. Considering the unpredictability of LLM outputs (e.g., explanations or inconsistent formats), we map model responses to the predefined label set through exact and fuzzy string matching before computing evaluation metrics. Moreover, to enable a direct comparison between contrastive learning–based and classification-based approaches in user geolocation, we introduce a variant of FewUser named ClassUser, which replaces the contrastive learning objective with cross-entropy loss for location classification while keeping the rest of the framework unchanged.

4.2 Few-Shot User Geolocation

Main Results. To evaluate model performance under limited supervision, we conduct experiments under both 1-shot and 8-shot settings, as summarized in Tables 1 and 2. Across all datasets, FewUser consistently achieves the best performance in both accuracy and localization precision.

In the 1-shot scenario, FewUser reaches 58.21% accuracy on TwiU and 51.35% on FliU, surpassing all baselines by a large margin. On TwitterUS and TwitterWorld, FewUser maintains strong results, while achieving the lowest mean and median distances, indicating superior spatial consistency even with minimal training data. Compared with large-scale pretrained models like LLaMA3, FewUser demonstrates significantly smaller distance errors (e.g., 1,303 km vs. 3,411 km on TwiU), highlighting its robustness and better alignment between semantic and geographic representations. While instruction-tuned LLMs benefit from extensive world knowledge, they require substantial computational resources and may produce unpredictable outputs that require post-processing. In contrast, FewUser is a lightweight, task-specific model that can be trained efficiently with limited labeled data, offers stable inference behavior, and explicitly optimizes user–location alignment through contrastive objectives. This makes FewUser more suitable for practical, resource-constrained geolocation scenarios.

When supervision increases to 8-shot, FewUser further improves across all datasets, achieving 69.86% on TwiU and 58.58% on FliU. The consistent improvement over 1-shot results shows strong scalability with additional samples. On TwitterUS and TwitterWorld, FewUser again achieves the best results, substantially lower than all other baselines, including SetFit and ClassUser.

Table 3: Cross-dataset performance under different label sets.

	Label Set	TwiU	FliU	CrossT2F	CrossF2T
1-shot	1	59.46	62.82	57.48	61.26
	2	71.74	60.46	54.83	59.42
	3	68.39	56.85	54.71	60.92
8-shot	1	75.68	65.81	66.45	64.87
	2	79.71	61.49	57.59	73.91
	3	66.67	59.38	53.45	64.94

Cross Datasets. To further evaluate FewUser's generalization ability across platforms, we conduct cross-dataset experiments between the two representative datasets, TwiU and FliU. We identify their common city labels and construct three different label sets with varying sizes. As shown in Table 3, we report results for both within-dataset (TwiU and FliU) and cross-dataset settings, where CrossT2F denotes training on TwiU and testing on FliU, and CrossF2T represents the opposite direction.

Across all label sets and shot configurations, FewUser demonstrates strong generalization across domains. In the 1-shot setting, cross-dataset performance remains comparable to within-dataset training, showing the model's robustness even under limited supervision. For example, under Label Set 1, FewUser achieves 59.46% on TwiU and 57.48% on CrossT2F, indicating minimal degradation when transferring between platforms. In the 8-shot setting, performance improves across all scenarios, with the highest cross-dataset accuracy reaching 73.91% in CrossF2T (Label Set 2). Notably, the small gap between within- and cross-dataset results highlights FewUser's ability to transfer learned representations between heterogeneous social media domains.

Shot Sensitivity. We further analyze the effect of the number of shots s on FewUser's few-shot geolocation performance and take

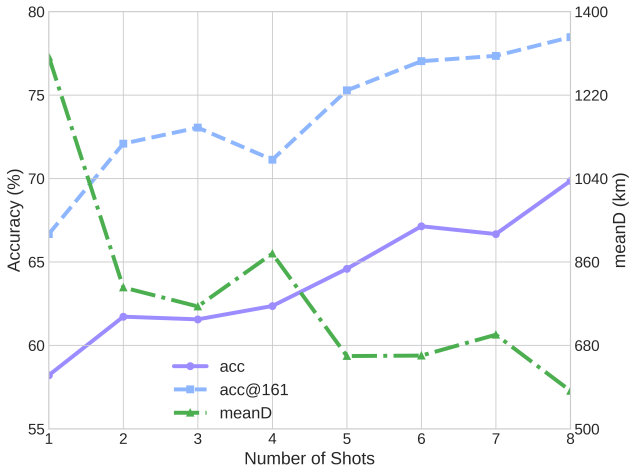


Figure 3: Effect of the number of shots on FewUser's geolocation performance on TwiU.

TwiU as an illustrative example, with similar trends observed on other datasets. As shown in Figure 3, with the number of training samples per class increasing from 1 to 8, both acc and acc@161 consistently improve, while meanD gradually decreases. The largest improvement occurs between 1-shot and 2-shot, indicating that FewUser can rapidly adapt even with a few additional examples. Beyond 4-shot, performance gains become smoother, suggesting that FewUser efficiently leverages small-scale supervision and quickly reaches a stable representation space. These results demonstrate that FewUser is data-efficient and robust to variations in the number of shots, maintaining stable geolocation performance under few-shot conditions.

Table 4: Effect of different numbers of tweets on geolocation.

#Tweets	TwiU		FliU	
	FewUser	ClassUser	FewUser	ClassUser
1	65.23	**54.07**	54.41	40.52
2	65.71	51.04	55.63	42.16
3	66.35	52.95	56.68	42.63
4	68.42	29.35	56.52	43.85
5	67.94	40.03	56.74	**45.01**
6	**69.86**	51.36	**58.58**	44.32
7	67.15	34.45	57.74	44.32
8	66.03	38.44	58.53	44.43
9	68.42	27.91	58.48	43.11
10	66.99	27.43	**58.58**	44.80

4.3 User Representation

In this section, we conduct comprehensive experiments with FewUser and ClassUser on the TwiU and FliU datasets, to examine the impact of user representation on the performance of few-shot social media user geolocation, for both contrastive learning-based and

classification-based models across different social media platforms. Specifically, we perform these experiments under an 8-shot setting.

What information is useful? We investigate this problem from two perspectives: the number of tweets and field selection. Older tweets may provide limited relevance to a user's current location, and including too many tweets increases computational cost. The results in Table 4 show that increasing the number of tweets generally improves geolocation accuracy for both models. FewUser demonstrates more stable gains as tweet count increases, reaching its best performance at six tweets, while ClassUser fluctuates more significantly across different counts. This suggests that FewUser effectively aggregates multi-tweet information through its contrastive learning framework, enabling robust user representation even with limited input data.

Table 5: Effect of various input sets on geolocation.

Input Set	TwiU		FliU	
	FewUser	ClassUser	FewUser	ClassUser
All	**69.86**	54.07	**58.58**	**45.01**
NoMeta	45.61	10.05	24.78	8.08
NoUserMeta	44.34	16.59	33.76	17.54
NoPostMeta	66.19	**56.30**	51.56	39.41
NoUserLocation	49.44	35.41	41.73	21.45
NoUserTime	68.42	55.50	57.16	42.53
NoUserTimezone	66.35	48.17	–	–
NoUserLanguage	67.31	54.23	–	–
NoUserDescription	65.23	48.64	54.78	42.00
NoUserName	65.39	55.34	57.79	43.32
NoUserOccupation	–	–	57.63	43.85
NoPostSource	66.99	55.98	58.43	**45.01**
NoPostTime	66.83	56.14	57.69	43.11
NoPostTag	67.46	53.91	51.19	38.56

To understand which specific metadata elements contribute positively or potentially interfere with model learning, we construct multiple input sets by removing certain metadata fields, as presented in Table 5. Removing all metadata leads to the most severe performance degradation across both datasets, confirming that metadata plays a crucial role in user geolocation. On TwiU, FewUser's accuracy drops by 32.22%, while ClassUser suffers an even larger 46.41% decrease. This pattern holds on FliU, indicating that FewUser maintains stronger robustness when metadata is limited. Among individual metadata fields, NoUserLocation causes the most substantial decline, highlighting that self-declared location is the most informative signal for city prediction. In contrast, removing time, timezone, or language has relatively minor effects, suggesting that temporal and linguistic cues are less decisive. Post-level metadata such as posting time and source produce only small variations in accuracy, implying that textual content remains a strong indicator for location inference. Finally, platform-specific features exhibit different levels of importance: Twitter-specific attributes (timezone, language) show moderate impact, whereas Flickr-specific attributes (occupation) contribute minimally.

How to integrate inputs? As shown in Table 6, we evaluate various integration strategies for user geolocation. For FewUser, the overall best performance is achieved with In1, which simply concatenates all inputs into a single sentence. This supports the idea that contrastive learning benefits from unified representations that preserve alignment between user and location embeddings. Interestingly, In2 performs slightly worse than In1 for FewUser but achieves the best result in the ClassUser setting. This indicates that separating tweet content and user profile features is more effective for classification tasks. More complex integration strategies such as InT and InUser+T result in a significant performance drop across both datasets. These methods introduce fragmented input structures, which increases sparsity and reduces the effectiveness of representation learning. The consistent trends across TwiU and FliU suggest that these observations are not dataset-specific but instead reflect the sensitivity of geolocation models to input design.

Table 6: Effect of different integration methods on geolocation.

	TwiU		FliU	
	FewUser	ClassUser	FewUser	ClassUser
In1	**69.86**	51.83	58.58	45.01
In2	65.07	**54.07**	**58.95**	**45.27**
InT	50.72	51.04	42.16	19.02
InUser+1	51.99	39.55	58.21	33.70
InUser+T	47.53	39.23	49.71	24.99
NoIn	33.01	39.23	17.64	11.36

Table 7: Effect of various user encoders on geolocation.

User Encoder	TwiU		FliU	
	FewUser	ClassUser	FewUser	ClassUser
BiLSTM	68.10	22.17	57.95	**45.17**
Transformer	67.62	49.28	57.58	35.08
Adapter	69.06	40.35	56.37	26.99
MLP	66.99	**51.36**	58.06	40.62
MP	**69.86**	45.77	**58.58**	30.53

How to fuse user features? We evaluate different user encoders of input representation, as shown in Table 7. Among all methods, MP achieves the best performance for FewUser on both TwiU and FliU datasets, highlighting the strength of this simple, parameter-free approach in few-shot scenarios. However, for ClassUser, MLP outperforms others on TwiU, while BiLSTM achieves the highest accuracy on FliU. These results suggest that more expressive models like MLP and BiLSTM can be beneficial for traditional classification-based geolocation models. Interestingly, Transformer-based encoders and Adapter modules do not consistently outperform simpler baselines, indicating that added complexity does not always translate into better geolocation performance.

4.4 Geographical Prompting

To examine the impact of geographical prompting, we conduct detailed experiments.

Hard Prompts. We use "[CLASS]" as the baseline for hard prompts. Prompts that state residence and include the word "city" are robust. On FliU, "A user resides in the city [CLASS]." is best, and on TwiU it also improves performance. In addition, declarative and colloquial phrasing shows domain effects. "I'm in [CLASS]." is the top template on TwiU, but drops below baseline on FliU. Interestingly, the slangy "[CLASS] in the house!" ranks near the top on FliU yet brings only a small gain on TwiU. This suggests that the best wording depends on the platform/domain. Furthermore, QA-style prompts help, but the exact question matters. On TwiU, "Question: where does this user reside in? Answer: [CLASS]." reaches second only to "I'm in [CLASS]." On FliU, the best QA variant is "Question: which city does this user live in? Answer: [CLASS].", while other QA wordings are close to baseline. Some phrasings even reduce accuracy (e.g., "Representing [CLASS]." on TwiU). Overall, hard prompts provide steady gains but are dataset-sensitive.

Table 8: Performance with hard and semi-soft prompts.

Prompt	Hard		Semi-soft	
	TwiU	FliU	TwiU	FliU
`[CLASS]`	66.51	57.21	67.94	57.26
A user from the city `[CLASS]`.	65.71	57.63	68.10	**58.06**
Someone from the city `[CLASS]`.	67.15	57.32	67.94	57.05
A user resides in the city `[CLASS]`.	67.46	**58.16**	66.19	57.00
This user resides in the city `[CLASS]`.	67.31	56.84	67.62	57.32
Question: which city does this user live in? Answer: `[CLASS]`.	66.83	57.79	66.03	58.00
Question: which city does this user reside in? Answer: `[CLASS]`.	66.67	57.11	66.83	57.32
Question: where does this user reside in? Answer: `[CLASS]`.	67.78	57.26	66.35	57.95
I'm in `[CLASS]`.	**68.26**	56.73	**69.86**	57.85
A guy from the city `[CLASS]`.	66.83	57.11	68.26	57.58
`[CLASS]` in the house!	66.67	58.11	68.58	57.26
Representing `[CLASS]`.	66.03	57.11	66.83	57.26
Avg.	66.78	57.37	67.97	57.62

Semi-soft Prompts. Compared with hard prompts, semi-soft prompts consistently improve performance across both datasets, with most templates showing small but stable gains. This indicates that FewUser benefits from learnable prompt representations that preserve linguistic priors from hard templates while refining them for geolocation. A closer examination reveals several trends. First, the best-performing template remains "I'm in [CLASS]." on both datasets. This shows that the semi-soft mechanism can further strengthen concise and natural patterns. Second, templates that were suboptimal in the hard setting, such as "A user from the city [CLASS].", improve notably, implying that learnable prompts can mitigate the brittleness of fixed wording. Third, QA-style templates also benefit from semi-soft tuning. For example, "Question: which city does this user live in? Answer: [CLASS]." lifts performance, suggesting that adaptive representations help the model understand QA structure.

Table 9: Performance of FewUser with soft prompts.

m	1	2	3	4	5	6	7
TwiU	65.87	64.75	65.87	**68.26**	66.19	67.46	65.87
FliU	56.84	57.48	57.90	57.53	58.00	58.27	57.00
m	8	9	10	11	12	13	14
TwiU	66.19	66.19	66.51	65.39	67.62	63.80	65.55
FliU	58.06	**58.58**	57.11	57.63	57.79	58.11	57.53

Soft Prompts. As shown in Table 9, geolocation performance varies with the number of soft tokens m. On TwiU, the accuracy peaks with 4 soft tokens. This suggests that a moderate number of learnable tokens is sufficient to represent location-related semantics without overfitting to noisy features. When the token count exceeds 10, performance begins to decline, implying that overly long prompts may disrupt the alignment between text and label representations. On FliU, the trend is smoother but still clear: performance steadily improves up to 9 tokens, after which additional tokens bring no benefit or cause slight drops. This moderate growth pattern indicates that photo-sharing platforms, such as Flickr, require more prompt capacity to encode heterogeneous user information.

These strategies exhibit complementary characteristics. Hard prompts offer interpretability, semi-soft prompts balance between human knowledge and adaptability, and soft prompts deliver maximum flexibility and performance with sufficient prompt capacity.

5 Conclusion

In this work, we introduced FewUser, a contrastive and prompt-driven framework for few-shot social media user geolocation. FewUser jointly models heterogeneous user information and geographical semantics through a user representation module and a geographical prompting module. Trained under a dual-objective scheme combining contrastive and matching losses with hard negative mining, FewUser demonstrates robust generalization across limited-supervision and cross-domain settings. Comprehensive experiments on four datasets show that FewUser consistently outperforms strong baselines in both accuracy and localization precision. Our analyses further reveal that simple yet unified representations, combined with prompt-based semantic alignment, can yield substantial improvements in few-shot user geolocation. In future work, we plan to extend FewUser toward multimodal social media geolocation, by incorporating visual and temporal cues from images and videos alongside textual and metadata signals. Such an extension would bridge the gap between text-centric and multimodal user modeling, enabling more comprehensive understanding of social users across platforms and modalities.

Acknowledgments

This research is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (Award No. MOE-T2EP20123-0015). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the Ministry of Education, Singapore.

References

[1] Tien Huu Do, Duc Minh Nguyen, Evaggelia Tsiligianni, Bruno Cornelis, and Nikos Deligiannis. 2018. Twitter user geolocation using deep multiview learning. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 6304–6308.

[2] Bo Han, Afshin Rahimi, Leon Derczynski, and Timothy Baldwin. 2016. Twitter Geolocation Prediction Shared Task of the 2016 Workshop on Noisy User-generated Text. In *Proceedings of WNUT'16*. 213–217.

[3] Bo Han, Afshin Rahimi, Leon Derczynski, and Timothy Baldwin. 2016. Twitter geolocation prediction shared task of the 2016 workshop on noisy user-generated text. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*. 213–217.

[4] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the CVPR'20*. 9729–9738.

[5] Ngai Lam Ho and Kwan Hui Lim. 2022. POIBERT: A Transformer-based Model for the Tour Recommendation Problem. In *Proceedings of IEEE BigData*.

[6] Binxuan Huang and Kathleen Carley. 2019. A Hierarchical Location Prediction Neural Network for Twitter User Geolocation. In *Proceedings of EMNLP-IJCNLP'19*. 4732–4742.

[7] Haosheng Huang, Georg Gartner, Jukka M Krisp, Martin Raubal, and Nico Van de Weghe. 2018. Location based services: ongoing evolution and research agenda. *Journal of Location Based Services* 12, 2 (2018), 63–93.

[8] Zhengbao Jiang, Frank F. Xu, Jun Araki, and Graham Neubig. 2020. How Can We Know What Language Models Know? *Transactions of the Association for Computational Linguistics* 8 (2020), 423–438.

[9] Chloe Kliman-Silver, Aniko Hannak, David Lazer, Christo Wilson, and Alan Mislove. 2015. Location, location, location: The impact of geolocation on web search personalization. In *Proceedings of IMC'15*. 121–127.

[10] Alkis Koudounas, Flavio Giobergia, Irene Benedetto, Simone Monaco, Luca Cagliero, Daniele Apiletti, Elena Baralis, et al. 2023. baρtti at GeoLingIt: Beyond Boundaries, Enhancing Geolocation Prediction and Dialect Classification on Social Media in Italy. In *CEUR Workshop Proceedings*.

[11] Kunrong Li, Zhu Sun, and Kwan Hui Lim. 2026. HyMoERec: Hybrid Mixture-of-Experts for Sequential Recommendation (Student Abstract). In *Proceedings of the 40th Annual AAAI Conference on Artificial Intelligence (AAAI'26)*.

[12] Menglin Li and Kwan Hui Lim. 2024. Leveraging Contrastive Learning for Few-shot Geolocation of Social Posts. arXiv:2403.00786 [cs.IR]

[13] Menglin Li, Kwan Hui Lim, Teng Guo, and Junhua Liu. 2023. A transformer-based framework for poi-level social post geolocation. In *Proceedings of ECIR*. 588–604.

[14] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2023. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *Comput. Surveys* 55, 9 (2023), 1–35.

[15] Yimin Liu, Xiangyang Luo, Zhiyuan Tao, Meng Zhang, and Shaoyong Du. 2023. UGCC: Social Media User Geolocation via Cyclic Coupling. *IEEE Transactions on Big Data* (2023).

[16] Ismini Lourentzou, Alex Morales, and ChengXiang Zhai. 2017. Text-based geolocation prediction of social media users with neural networks. In *Proceedings of IEEE BigData*. 696–705.

[17] Kateryna Lutsai and Christoph H. Lampert. 2023. Geolocation Predicting of Tweets Using BERT-Based Models. arXiv:2303.07865 [cs.CL]

[18] Masahiro Matsumoto and Kazuaki Ando. 2022. A Deep Learning Model of Estimating User's Place of Residence Using Tweets and Weather Information. In *Proceedings of CSDE'22*. 1–6.

[19] Wenchuan Mu, Menglin Li, and Kwan Hui Lim. 2025. A Social Data-Driven System for Identifying Estate-related Events and Topics. In *International Conference on Advances in Social Networks Analysis and Mining*. Springer, 443–449.

[20] Yaqiong Qiao, Xiangyang Luo, Jiangtao Ma, Meng Zhang, and Chenliang Li. 2023. Twitter user geolocation based on heterogeneous relationship modeling and representation learning. *Information Sciences* 647 (2023), 119427.

[21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *Proceedings of ICML'21*. 8748–8763.

[22] Afshin Rahimi, Timothy Baldwin, and Trevor Cohn. 2017. Continuous Representation of Location for Geolocation and Lexical Dialectology using Mixture Density Networks. In *Proceedings of EMNLP'17*. 167–176.

[23] Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2017. A Neural Model for User Geolocation and Lexical Dialectology. In *Proceedings of ACL*. 209–216.

[24] Alan Ramponi and Camilla Casula. 2023. GeoLingIt at EVALITA 2023: Overview of the Geolocation of Linguistic Variation in Italy Task. In *International Workshop on Evaluation of Natural Language and Speech Tools for Italian*.

[25] Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing, and Jason Baldridge. 2012. Supervised text-based geolocation using language models on an adaptive grid. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*. 1500–1510.

[26] Yves Scherrer and Nikola Ljubešić. 2021. Social media variety geolocation with geobert. In *Proceedings of the Eighth Workshop on NLP for Similar Languages, Varieties and Dialects*.

[27] Taylor Shin, Yasaman Razeghi, Robert L Logan IV, Eric Wallace, and Sameer Singh. 2020. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. In *Proceedings of EMNLP'20*. 4222–4235.

[28] Philippe Thomas and Leonhard Hennig. 2018. Twitter Geolocation Prediction Using Neural Networks. In *Language Technologies for the Challenges of the Digital Age*, Georg Rehm and Thierry Declerck (Eds.). 248–255.

[29] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. 2016. YFCC100M: The new data in multimedia research. *Commun. ACM* 59, 2 (2016), 64–73.

[30] Lewis Tunstall, Nils Reimers, Unso Eun Seo Jo, Luke Bates, Daniel Korat, Moshe Wasserblat, and Oren Pereg. 2022. Efficient few-shot learning without prompts. *arXiv preprint arXiv:2209.11055* (2022).

[31] Zhaomin Xiao, Yan Huang, and Eduardo Blanco. 2024. Analyzing Large Language Models' Capability in Location Prediction. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue (Eds.). ELRA and ICCL, Torino, Italia, 951–958. https://aclanthology.org/2024.lrec-main.85/

[32] Jingyu Zhang, Alexandra DeLucia, and Mark Dredze. 2022. Changes in Tweet Geolocation over Time: A Study with Carmen 2.0. In *Proceedings of WNUT'22*.

[33] Renrui Zhang, Ziyu Guo, Wei Zhang, Kunchang Li, Xupeng Miao, Bin Cui, Yu Qiao, Peng Gao, and Hongsheng Li. 2022. Pointclip: Point cloud understanding by clip. In *Proceedings of CVPR'22*. 8552–8562.

[34] Xin Zheng, Jialong Han, and Aixin Sun. 2018. A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering* 30, 9 (2018), 1652–1671.

[35] Zexuan Zhong, Dan Friedman, and Danqi Chen. 2021. Factual Probing Is [*MASK*]: Learning vs. Learning to Recall. In *Proceedings of NAACL'21*. 5017–5033.

[36] Fan Zhou, Xiuxiu Qi, Kunpeng Zhang, Goce Trajcevski, and Ting Zhong. 2022. MetaGeo: a general framework for social user geolocation identification with few-shot learning. *IEEE Transactions on Neural Networks and Learning Systems* (2022).

[37] Paola Zola, Costantino Ragno, and Paulo Cortez. 2020. A Google Trends spatial clustering approach for a worldwide Twitter user geolocation. *Information Processing & Management* 57, 6 (2020), 102312.