# A Unified Neural Network Model for Geolocating Twitter Users

**Mohammad Ebrahimi,  Elaheh ShafieiBavani,  Raymond Wong,  Fang Chen**

University of New South Wales, Sydney, Australia

Data61 CSIRO, Sydney, Australia

{mohammade,elahehs,wong,fang}@cse.unsw.edu.au

## Abstract

Locations of social media users are important to many applications such as rapid disaster response, targeted advertisement, and news recommendation. However, many users do not share their exact geographical coordinates due to reasons such as privacy concerns. The lack of explicit location information has motivated a growing body of research in recent years looking at different automatic ways of determining the user's primary location. In this paper, we propose a unified user geolocation method which relies on a fusion of neural networks. Our joint model incorporates different types of available information including tweet text, user network, and metadata to predict users' locations. Moreover, we utilize a bidirectional LSTM network augmented with an attention mechanism to identify the most location indicative words in textual content of tweets. The experiments demonstrate that our approach achieves state-of-the-art performance over two Twitter benchmark geolocation datasets. We also conduct an ablation study to evaluate the contribution of each type of information in user geolocation performance.

## 1 Introduction

Knowing physical locations involved in social media data helps us to understand what is happening in real life, to bridge the online and offline worlds, and to develop applications for supporting real-life demands. For example, we can monitor public health of residents (Cheng et al., 2010), recommend local events (Yuan et al., 2013) or attractive places (Noulas et al., 2012) to tourists, identify locations of emergency (Ao et al., 2014) or even disasters (Lingad et al., 2013), and summarize regional topics (Rakesh et al., 2013). Even though platforms such as Twitter allow users to geolocate their posts to reveal their locations either manually

or with the help of GPS, it is reported that less than 1% of Twitter data has geo-coordinates provided (Jurgens, 2013). Moreover, location information on Twitter is far from being complete and accurate. For instance, self-declared home information in many user profiles is inaccurate or even invalid (Hecht et al., 2011). The lack of explicit location information in the majority of tweets has motivated a growing body of research in recent years looking at different automatic ways of determining the user's primary location (i.e.,*user geolocation*) and/or - as a proxy for the former - the location from which tweets have been posted (Ajao et al., 2015).

Geolocation methods usually train a model on a small set of users whose locations are known (e.g., through GPS-based geotagging), and predict locations of other users using the resulting model. These models broadly fall into three categories: text-based (Eisenstein et al., 2010; Wing and Baldridge, 2011; Roller et al., 2012), network-based (Jurgens, 2013; Compton et al., 2014; Jurgens et al., 2015), and hybrid methods that combine text, user network, and metadata information (Rahimi et al., 2015b,a; Jayasinghe et al., 2016; Miura et al., 2016) with the aim of achieving state-of-the-art performance.

In this paper, we present a neural network-based system that we developed for user geolocation in Twitter. Our model combines different sources of information including tweet text, metadata, and user network. We employ a neural network model to generate a dense vector representation for each field and then use the concatenation of these representations as the feature for classification. Our main contributions can be summarized as follows:

1. We propose a unified user geolocation method that relies on a fusion of neural networks, incorporating different types of avail-

able information: tweet message, users' social relationships, and metadata fields embedded in tweets and profiles.

2. For modeling the tweet text (and textual metadata fields), we use bidirectional Long Short-Term Memory (LSTM) networks augmented with a context-aware attention mechanism (Yang et al., 2016), which helps to identify the most location indicative words.

3. Through the empirical studies on two standard Twitter datasets, we demonstrate that the proposed method outperforms other state-of-the-art approaches in addressing the problem of user geolocation.

4. We train an individual model for each information field, and analyze the contribution of each component in the geolocation process.

The rest of the paper is organized as follows. We review the related work in Section 2. Utilized data is described in Section 3. Section 4 explains the proposed approach. The experimental results are given in Section 5, and finally, we conclude the paper and outline possible future work in Section 6.

## 2 Related Work

### 2.1 Text-based Methods

Text-based methods utilize the geographical bias of language use in social media for geolocation. These methods have widely used probability distributions of words over locations. Maximum likelihood estimation approaches (Cheng et al., 2010, 2013) and language modeling approaches minimizing KL-divergence (Roller et al., 2012) have succeeded in predicting user locations using word distributions. Topic modeling approaches to extract latent topics with geographical regions (Eisenstein et al., 2010; Hong et al., 2012; Ahmed et al., 2013; Yuan et al., 2013) have also been explored considering word distributions.

Supervised learning methods with word features are also popular in text-based geoinference. Multinomial Naïve Bayes (Han et al., 2012, 2014; Wing and Baldridge, 2011), logistic regression (Wing and Baldridge, 2014; Han et al., 2014), hierarchical logistic regression (Wing and Baldridge, 2014), and multi-layer neural network with stacked denoising autoencoder (Liu and Inkpen, 2015) have realized geolocation predic-

tion from text. A semi-supervised learning approach has been proposed by Cha et al. (2015) using a sparse-coding and dictionary learning. Hulden et al. (2015) have used a kernel-based method to smooth linguistic features over very small grid sizes and consequently alleviate data sparseness. Chi et al. (2016) have employed Multinomial Naïve Bayes and focused on the use of textual features (i.e., location indicative words, GeoNames gazetteers, user mentions, and hashtags) for geolocation inference. More recently, Rahimi et al. (2017b) have proposed a neural network-based geolocation approach. They used the parameters of the hidden layer of the neural network as word and phrase embeddings, and performed a nearest neighbor search on a sample of city names and dialect terms.

While having good results, text-based approaches are often limited to those users who generate text that contains geographic references (Jurgens, 2013).

### 2.2 Network-based Methods

Network-based methods rely on the geospatial homophily of interactions (of several kinds) between users. An early work by Davis Jr et al. (2011) proposed an approach in which the location of a given user is inferred by simply taking the most-frequently seen location among its social network. Jurgens (2013) have extended the idea of location inference as label propagation over some form of friendship graph by interpreting location labels spatially. Locations are then inferred using an iterative, multi-pass procedure. This method has been further extended by Compton et al. (2014) to take into account edge weights in the social network, and to limit the propagation of noisy locations. They weigh locations as a function of how many times users interacted there, hence favoring locations of friends with evidence of a close relationship. Jurgens et al. (2015) have released a framework for nine network-based geolocation methods targeting Twitter.

The main limitation of network-based models is that they completely fail to geolocate users who are not connected to geolocated components of the graph (i.e., isolated users).

### 2.3 Hybrid Methods

Several attempts have been made to combine different sources of information for geolocating social media users. Li et al. (2012) have proposed a

geolocation method by integrating both friendship and content information in a probabilistic model.

Rahimi et al. (2015b) showed that geolocation predictions from text can effectively be used as a back-off for disconnected users in a network-based approach. In another work by Rahimi et al. (2015a), a hybrid approach has been proposed by propagating information on a graph built from user mentions in Twitter messages, together with dongle nodes corresponding to the results of a text-based geolocation method. Ebrahimi et al. (2017, 2018b) have presented a hybrid approach by incorporating both text and network information, and shown that the filtering of highly mentioned users in the social graph can improve the geolocation performance. Rahimi et al. (2017b) have proposed a text geolocation method based on neural network and incorporated it into their network-based approach (Rahimi et al., 2015a). Wang et al. (2017) have introduced a collective geographical embedding algorithm to embed multiple information sources into a low dimensional space, such that the distance in the embedding space reflects the physical distance in the real world.

Metadata such as location fields have also been used as effective clues to predict the user's location (Hecht et al., 2011). Different geoinference approaches have been proposed to consider text and metadata information simultaneously, such as dynamically weighted ensemble method (Mahmud et al., 2012), and stacking approach (Han et al., 2014). Jayasinghe et al. (2016) have proposed a cascade ensemble approach by combining text-based, metadata-based, and network-based geolocation methods. Additionally, their approach makes use of several dedicated services, such as GeoNames gazetteers, time zone to GeoName mappings, IP country resolver and customized scrapers for social media websites.

Miura et al. (2016) have trained a neural network utilizing the fastText n-gram model (Joulin et al., 2016) on tweet text, user location, user description, and user timezone. They have utilized several mapping services using external resources, such as GeoNames and time zone boundaries for feature preprocessing. This model has been further extended by Miura et al. (2017) to also consider user network information for geolocation.

Thomas and Hennig (2017) have proposed a geolocation method that relies on the combination of individual neural networks trained on text and

metadata fields. Ebrahimi et al. (2018a) have proposed a word embedding-based approach to predict the geographic proximity of connected users in the social graph based on their linguistic similarities. The calculated similarity scores have been used for weighting edges between users in the graph. Tweet content and metadata is also combined with an ensemble learning method to geolocate isolated users in the graph.

## 3 Data

We have used two benchmark Twitter geolocation datasets in our experiments:

- **TWITTERUS** is a dataset compiled by (Roller et al., 2012), which contains 38M tweets from 450K users in the United States. Out of 450K users, 10K are reserved for the development set and another 10K for the test set. The ground truth location of each user is set to its first geotag in the dataset. To make city prediction possible in this dataset, we additionally assigned city centers to ground truth geotags using the city category of Han et al. (2012).

- **WNUT** is a user-level dataset from the geolocation prediction shared task of WNUT 2016 (Han et al., 2016). The dataset covers 13M tweets from 3362 cities worldwide, and consists of 1M training users, 10K development users, and 10K test users. The ground truth location of a user is decided by majority voting of the closest city center.

Note that the metadata of a tweet includes not only the tweet message (text) but also a variety of information such as tweet publication time, and user account data such as location and timezone. The organizers have provided full metadata for the test sets but only the tweet IDs for training and development sets. We collect metadata for training/development tweets using the Twitter API[1].

## 4 The Proposed Approach

Figure 1 illustrates an overview of the proposed model for user geolocation. We make use of the following sources of information to train our model: 1) Tweet text; 2) User network; and 3)

---

[1]We were able to obtain approximately 71-75% of the full datasets. The remaining tweets are no longer available, mainly because users deleted these messages and/or accessibility changes in Twitter.

Metadata including user-declared location, user description, user name, timezone, user language, tweet creation time, user UTC offset, links (URL domains), and application source.

Each field is processed by a separate sub-network to generate a feature vector representation $R_j$. These feature vectors are then concatenated to build a final user representation $\hat{R}$ which is fed into a linear classification layer:

$$\hat{R} = R_1 \parallel ... \parallel R_N \tag{1}$$

$$r = softmax(W_r \hat{R} + b_r) \tag{2}$$

where N is the number of features (11 in total), $r \in \mathbb{R}^R$ is the hidden representation at the penultimate layer. $W_r$ is a weight matrix and $b_r$ is a bias vector. $r$ is fully connected to the output layer and activated by softmax to generate a probability distribution over the classes. We employ the cross-entropy loss as the objective function. Let $M$ be the number of examples (i.e., users) and $c$ be the number of classes (i.e., regions), then the cross-entropy loss is defined by:

$$L = -\sum_{i=1}^{M} \sum_{j=1}^{c} y_i^j \log(\tilde{y}_i^j) \tag{3}$$

where $y_i, i = 1, ..., M$ is the ground-truth vector, $\tilde{y}_i$ is the predicted probability vector, and $\tilde{y}_i^j$ is the probability that user $i$ resides in region $j$. We minimize the objective function through Stochastic Gradient Descent (SGD) over shuffled mini-batches with Adam (Kingma and Ba, 2014).

We design several sub-networks to provide vectorized representation for each raw field. For processing the tweet text, we utilize word embeddings (Mikolov et al., 2013) and bidirectional Long Short-Term Memory (LSTM) unit (Hochreiter and Schmidhuber, 1997) augmented with a context-aware attention mechanism (Yang et al., 2016) (Section 4.1). We construct a @-mention graph as a representation of users' interactions, and utilize this graph to extract the user network. We then use an embedding layer with attention mechanism to create the final user network representation (Section 4.2).

We divide metadata fields into two classes: textual, and categorical. For representing textual metadata fields (i.e., location, description, user name, and timezone), we use word embeddings and bidirectional LSTM networks with attention

mechanism. We treat other metadata fields (language, tweet time, UTC offset, links, and source) as categorical features, and convert them to one-hot encodings which are then fed forward to a dense layer (Section 4.3). In the following subsections, we describe details of each component.

## 4.1 Text Component

Figure 2(a) demonstrates the architecture of text sub-network. It takes the sequence of words in the tweet $T = \{w_1, w_2, ..., w_n\}$ as input. An embedding layer is used to project the words to a low-dimensional vector space $\mathbb{R}^E$, where $E$ is the size of the embedding layer. We initialize the weights of the embedding layer using our pre-trained word embeddings (Section 5.1). The embeddings of tweet words are then forwarded to an LSTM layer. An LSTM takes as input the words of a tweet and produces the word annotations $H = (h_1, h_2, ..., h_n)$, where $h_i$ is the hidden state of the LSTM at time-step $i$, summarizing all the information of the sentence up to $w_i$. We use bidirectional LSTM (BiLSTM) in order to get annotations for each word that summarize the information from both directions of the message. A bidirectional LSTM consists of a forward LSTM, $\overrightarrow{f}$, that reads the sentence from $w_1$ to $w_T$, and a backward LSTM, $\overleftarrow{f}$, that reads the sentence from $w_T$ to $w_1$. We obtain the final annotation for each word $w_i$, by concatenating the annotations from both directions:

$$h_i = \overrightarrow{h_i} \parallel \overleftarrow{h_i}, h_i \in \mathbb{R}^{2L} \tag{4}$$

where $\parallel$ denotes the concatenation operation and $L$ the size of each LSTM. In order to amplify the contribution of important words in the final representation, we use a context-aware attention mechanism (Yang et al., 2016), that aggregates all the intermediate hidden states using their relative importance. An attention mechanism assigns a weight $a_i$ to each word annotation, which reflects its importance. We compute the representation of the tweet text, $R_{text}$, as the weighted sum of all the word annotations using the attention weights. This attention mechanism introduces a context vector $u_h$ that helps to identify the informative words and it is randomly initialized and jointly learned with the rest of the attention layer weights. Formally, $R_{text}$ is defined as:

$$R_{text} = \sum_{i=1}^{|T|} a_i h_i, \quad R_{text} \in \mathbb{R}^{2L} \tag{5}$$
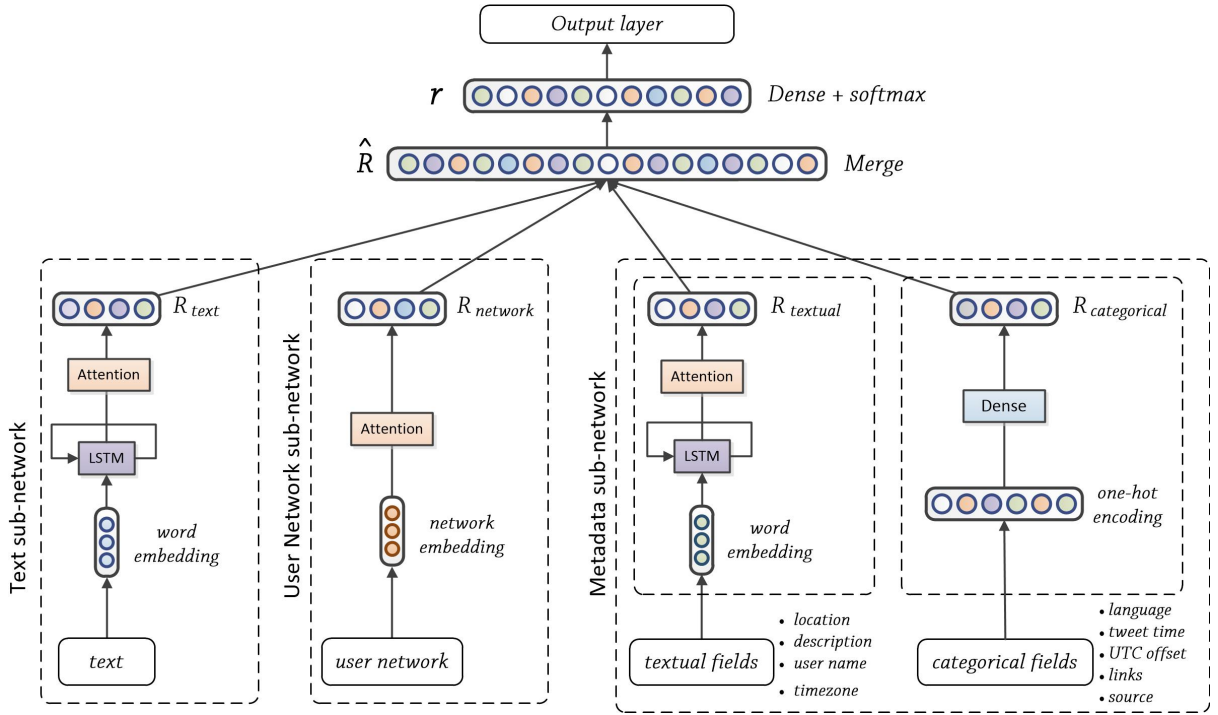
Figure 1: Overview of the proposed approach

$$a_i = \frac{exp(e_i{}^\top u_h)}{\sum_{t=1}^{|T|} exp(e_t{}^\top u_h)}, \quad \sum_{i=1}^{|T|} a_i = 1 \qquad (6)$$

$$e_i = tanh(W_h h_i + b_h), \quad e_i \in [-1, 1] \qquad (7)$$

where $W_h$, $b_h$ and $u_h$ are the layer's weights.

We use batch normalization (Ioffe and Szegedy, 2015) for normalizing inputs in order to reduce internal covariate shift. The risk of overfitting by co-adapting units is reduced by implementing dropout (Srivastava et al., 2014) between individual neural network layers.

### 4.2 User Network Component

As a representation of users' social relationships, we construct an undirected graph from interactions among Twitter users based on @-mentions in their tweets (Rahimi et al., 2015b). In this graph, nodes are all users in the dataset (train and test), as well as other external users mentioned in their tweets, and undirected edges are created between two users if either user mentioned the other. This unidirectional setting results in large numbers of edges. To make the process more tractable, we remove all nodes corresponding to external users with degree less than 3 (i.e., external users who have been mentioned by less than 3 different users in a training set).

Figure 2(b) illustrates an overview of the user network component. After filtering the graph, we consider the adjacent nodes (i.e., immediate linked users) of each training user as its network. The user network $N = \{u_1, u_2, ..., u_n\}$ is given as input to an embedding layer. Embedding of user network $E_N = (e_1, e_2, ..., e_n)$ is then fed to an attention layer to compute the final representation of user network, $R_{network}$:

$$R_{network} = \sum_{i=1}^{T} a_i e_i, \quad R_{network} \in \mathbb{R}^{2L} \qquad (8)$$

where $a_i$ is the weight assigned to embedding $e_i$ by the attention mechanism (Equation 6).

### 4.3 Metadata Component

According to (Guo and Berkhahn, 2016), the embeddings of categorical variables can reduce the network size while capturing the intrinsic properties of the categorical variables. Hence, we convert metadata fields with a finite set of elements (UTC offset, links, user language, tweet publication time, and application source) to one-hot encodings, which are then forwarded to a dense layer with Rectified Linear Units (ReLU) activation function.

The user UTC offset is an integer in seconds representation (e.g., $-18000$), and the tweet publication time is given in UTC time , e.g., Fri Mar 02
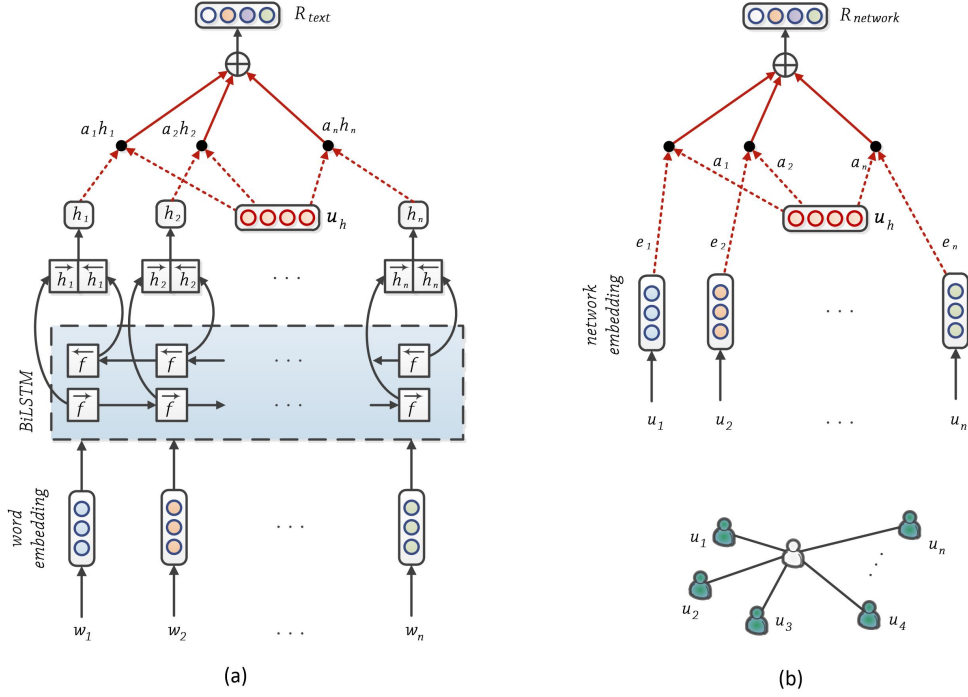
46

Figure 2: Architectures of the proposed sub-networks for (a) Tweet text, and (b) User network . $\oplus$ shows element-wise addition.

12:19:40 +0000 2012. We convert user UTC offset into hours representation (e.g., $-18000/3600 = -5$). For tweet publication time, we use only time of the day information (e.g., 12:19) and split it into multiple bins. Specifically, we interpret every 10 minutes as a bin (144 bins in total). The intuition is that tweets originated from a particular location (e.g., Germany) favor certain bins, and this preference of bins should be different to tweets from a distant location (e.g., Japan) (Lau et al., 2017).

For metadata fields containing texts (i.e., user description, user location, user name, and timezone), we use an embedding layer and consequently forward the results to an LSTM layer. The attention mechanism is also employed to provide the final representation of textual metadata fields. Again batch normalization and dropout is applied between individual layers to avoid overfitting.

## 5 Experimental Results

### 5.1 Experiment Settings

In the text sub-network, words are input to the model as n-dimensional word embeddings. We pre-trained word embeddings using word2vec (Mikolov et al., 2013) over tweet text of the full training data. The model was trained using the Skip-gram architecture and negative sampling (k = 5) for five iterations, with a context window of 5 and subsampling factor of 0.001. It is noteworthy that to be part of the vocabulary, words should occur at least five times in the corpus. We chose word embeddings of size 200/300 for TWIT-TERUS/WNUT datasets because smaller embeddings experimentally showed to capture not as much detail and resulted in a lower accuracy. Larger word embeddings, on the other hand, made the model too complex to train. In the preprocessing step, we used replacement tokens for URLs, mentions and numbers. However, we did not replace hashtags as doing so experimentally demonstrated to decrease the accuracy.

The layers and the embeddings in our sub-networks have parameters like embedding dimension, LSTM unit size, and attention context vector size. We chose optimal values for these parameters in terms of accuracy with a grid search using the development sets of TwitterUS and WNUT. The selected parameters values are reported in Table 1. It should be noted that the main reason for selecting smaller values for the TWITTERUS dataset is its larger size (in terms of tweet number) comparing to the WNUT dataset. We set the hyper-parameters of our final model as follows: batch size = 256, learning rate = 0.001, epochs = 5. The dropout rate between layers is set to 0.2.

| Model | Embedding size | | LSTM unit size | | Attention vector size | |
|---|---|---|---|---|---|---|
| | TwitterUS | WNUT | TwitterUS | WNUT | TwitterUS | WNUT |
| Tweet text | 200 | 300 | 100 | 200 | 200 | 400 |
| User network | 200 | 400 | 100 | 200 | 200 | 400 |
| Location | 100 | 200 | 100 | 200 | 200 | 400 |
| Description | 100 | 200 | 100 | 200 | 200 | 400 |
| User name | 100 | 200 | 100 | 200 | 200 | 400 |
| Timezone | 100 | 200 | 100 | 200 | 200 | 400 |

Table 1: Parameter settings of the proposed models.

## 5.2 Evaluation Metrics

We evaluate our approach in the following three commonly used metrics for user geolocation:

- *Acc@161*: The percentage of predicted locations which are within a 161km (100 mile) radius of the actual location (Cheng et al., 2010). This metric is a proxy for accuracy within a metro area.

- *Mean error*: The mean value of error distances in predicted locations (Eisenstein et al., 2010).

- *Median error*: The median value of error distances in predictions (Eisenstein et al., 2010).

Note that higher numbers are better for Acc@161 but lower numbers are better for mean and median errors.

## 5.3 Results

Table 2 presents the performance of user geolocation methods over TWITTERUS and WNUT[2] datasets.

The results show that our proposed method achieves the best performance in terms of all evaluation metrics. The main reason is the effective representation of text, metadata, and network information, and unifying them through a fusion of neural networks.

## 5.4 Ablation Study

To evaluate the contribution of each component in indicating the user's location, we train an individual neural network model for each field. To this end, we feed the final representation of each sub-network to a fully-connected dense layer, activated

by softmax function. We use stochastic gradient descent over shuffled mini-batches with Adam (Kingma and Ba, 2014) and cross-entropy loss as objective function for classification. The parameters of all models are set as follows: batch size = 256, epochs = 5, dropout=0.2, and learning rate = 0.001. Table 3 shows the performance breakdown for each model over the WNUT dataset.

The results conclude that user-declared location in tweet metadata is the most informative field for geolocating users, and model trained on this source achieves the best single source performance. This model can correctly geolocate 44.9% of users with a median error of 41.0km.

Using only tweet text, our model can predict the correct location for 34.9% of all users with a median error of 169.3km. It is noteworthy that this model outperforms the text-based approach IBM.1 (Chi et al., 2016) in terms of all metrics by a large margin.

User network model can correctly geolocate only 18.4% of users. However, our experiments show that excluding user network information declines the performance of the final model in terms of accuracy by 5.1%. Models using other metadata fields provide an accuracy between 2.7% to 10.6%, with description field being the most informative one. Tweet publication time, on the other hand, has the minimum accuracy in predicting user's location. However, mining the temporal patterns of users' posting habits can potentially provide useful information for geolocation inference.

We have also reported the results of our model when it takes only the metadata fields as inputs. The metadata-based model can correctly geolocate 46.1% of the users with a mean error of 1318.3km, and a median error of 37.9km. It shows the ef-

| | | TwitterUS | | | WNUT | | | |
| Category | Acc@161 | Mean | Median | Accuracy | Acc@161 | Mean | Median |
|---|---|---|---|---|---|---|---|---|
| WB-Uniform (Wing and Baldridge, 2014) | TB | 49 | 703 | 170 | - | - | - | - |
| WB-KDTree (Wing and Baldridge, 2014) | TB | 48 | 686 | 191 | - | - | - | - |
| MDN-Shared (Rahimi et al., 2017a) | TB | 42 | 655 | 216 | - | - | - | - |
| MLP+KDTree (Rahimi et al., 2017b) | TB | 54 | 562 | 121 | - | - | - | - |
| MLP+K-Mean (Rahimi et al., 2017b) | TB | 55 | 581 | 91 | - | - | - | - |
| LP-Rahimi (Rahimi et al., 2015b) | NB | 37 | 747 | 431 | - | - | - | - |
| LP-NA (Rahimi et al., 2016) | NB | 50 | 610 | 144 | - | - | - | |
| MadCelB (Rahimi et al., 2015a) | NB | 54 | 709 | 117 | - | - | - | - |
| MadCelW (Rahimi et al., 2015a) | NB | 54 | 705 | 116 | - | - | - | - |
| LP-LR (Rahimi et al., 2015b) | Hyb | 50 | 620 | 157 | - | - | - | - |
| MadCelW-LR (Rahimi et al., 2015a) | Hyb | 60 | 529 | 78 | - | - | - | - |
| MadCelW-MLP (Rahimi et al., 2017b) | Hyb | 61 | 515 | 77 | - | - | - | - |
| GeoCel-BK (Ebrahimi et al., 2017) | Hyb | 66 | 438 | 56 | - | - | - | - |
| Drexel.2 | ? | - | - | - | 7.9 | - | 6161.4 | 4000.2 |
| Drexel.1 | ? | - | - | - | 8.0 | - | 6053.3 | 5714.9 |
| Aist | ? | - | - | - | 9.8 | - | 4002.4 | 1711.1 |
| IBM.1 (Chi et al., 2016) | TB | - | - | - | 22.5 | - | 2860.2 | 630.2 |
| Drexel.3 | ? | - | - | - | 35.2 | - | 3124.4 | 262.7 |
| FujiXerox.3 (Miura et al., 2016) | Hyb | - | - | - | 45.1 | - | 1084.3 | 28.2 |
| FujiXerox.1 (Miura et al., 2016) | Hyb | - | - | - | 46.4 | - | 963.8 | 21.0 |
| FujiXerox.2 (Miura et al., 2016) | Hyb | - | - | - | 47.6 | - | 1122 | 16.1 |
| CSIRO.3 (Jayasinghe et al., 2016) | Hyb | - | - | - | 50.1 | - | 2242.4 | 30.6 |
| CSIRO.2 (Jayasinghe et al., 2016) | Hyb | - | - | - | 52.0 | - | 2071.5 | 23.1 |
| CSIRO.1 (Jayasinghe et al., 2016) | Hyb | - | - | - | 52.6 | - | 1929 | 21.7 |
| Thomas (Thomas and Hennig, 2017) | Hyb | - | - | - | 53.0 | - | 839 | 14.9 |
| Proposed Approach | Hyb | **70.8** | **313.2** | **40.1** | **56.7** | **72.3** | **731.5** | **0** |

Table 2: Performance of Text-based (TB), Network-based (NB), and Hybrid (Hyb) geolocation methods over TwitterUS and WNUT datasets ("-" signifies that no results were published for the given dataset, and "?" signifies that the participant team has not provided descriptions of the proposed system). We have also reported the *Accuracy* of our proposed approach on WNUT dataset to make our results comparable with the existing methods.

fectiveness of utilized metadata fields for user geolocation. Meanwhile, a deeper analysis of metadata fields can further improve the performance of user location prediction. As an example, customized scrapers for social media websites like FourSquare, Swarm, Path, Facebook, and Instagram can be employed as described by (Jayasinghe et al., 2016) to increase the geolocation accuracy.

## 5.5 Error Analysis

As reported in Table 2, our proposed approach achieves quite low median errors over the TwitterUS and WNUT datasets (i.e., 40.1km and 0km, respectively). However, there are some cases with large error distances, which make the mean errors much larger than median errors. Our analysis shows that some notable error distances are related to the following cases: (1) Users from remote areas for which less supervisions are available; (2) Users from small cities/states are misclassified to be in the neighboring larger cities/states; (3) Users from some neighboring cities/states are also misclassified between the two cities/states, which might be the result of business and entertainment connections between them.

Our ablation study demonstrates that the location field highly contributes to the geolocation per-

| Model | Accuracy | Mean | Median |
|---|---|---|---|
| Tweet text | 34.9 | 1674.1 | 169.3 |
| User network | 18.4 | 2551.7 | 789.8 |
| Metadata fields | 46.1 | 1318.3 | 37.9 |
| - Location | 44.9 | 3515.0 | 41.0 |
| - Description | 10.6 | 5540.7 | 3005.4 |
| - Timezone | 6.4 | 5203.6 | 5181.0 |
| - User name | 6.1 | 5836.1 | 3966.7 |
| - UTC offset | 5.3 | 6305.9 | 3727.8 |
| - User language | 4.6 | 9077.1 | 8585.4 |
| - Links | 4.5 | 6543.2 | 6691.6 |
| - Source application | 4.4 | 6950.3 | 6937.7 |
| - Tweet time | 2.7 | 11138.3 | 10165.5 |

Table 3: Performance breakdown for each component over WNUT dataset

formance. However, some prediction errors arise when location fields are incorrect. We found two main cases that result in incorrect location fields: (1) Users who move to a new place (i.e., house) but do not update their locations; (2) Users who visit a new place (e.g., as tourists) and temporarily update their locations. Our proposed model cannot handle these types of errors, since it only supports single location field. A future direction is to extend the current architecture to track location changes and deal with temporal states such as traveling.

Previous network-based methods (Jurgens, 2013; Compton et al., 2014) have demonstrated the effectiveness of users' social relationships for geolocation inference. However, our ablation study shows relatively low accuracy for the user network component. One main reason is that our model is less sophisticated (but more scalable) comparing to the mentioned network-based methods, since it only considers the immediate connected nodes as the network for each user. As a future work of this study, node/graph embeddings such as DeepWalk (Perozzi et al., 2014) can be employed to provide better representation of users' social relationships, and consequently improve the accuracy of network component.

## 6 Conclusion and Future Work

In this paper, we have proposed a unified user geolocation method which relies on a fusion of neural networks. Our joint model effectively utilizes different sources of information including tweet message, users' social relationships, and metadata fields embedded in tweets and profiles. In particular, we employed a neural network model to generate a dense vector representation for each information field and then used the concatenation of these representations as the feature for classification. For modeling tweet message and textual metadata fields, we utilized a bidirectional LSTM network augmented with an attention mechanism to identify the most location indicative words.

We have conducted comprehensive experiments on two standard Twitter geolocation datasets, and demonstrated that our method achieves the best performance in terms of all three evaluation metrics. In an ablation study, we have also trained individual models to investigate the usefulness of each information field in predicting the locations of Twitter users.

As a future work, it would be intriguing to utilize customized scrapers for social media websites (Jayasinghe et al., 2016) to further improve the performance of our geolocation model. It is noteworthy that the proposed model could be modified to infer other user demographic attributes such as gender and age.

Tweet publication time include both date and time, however, only time information is exploited in this work to infer users' geolocations. A future direction is to leverage tweeting behavior over dates for user geolocation. The intuition is that local residents would occasionally post tweets about their home city in a long-term manner, while tourists tend to tweet a lot while visiting the city. Hence, their different tweeting patterns can be easily revealed using date information from their tweet timestamps.

## Acknowledgments

## References

Amr Ahmed, Liangjie Hong, and Alexander J Smola. 2013. Hierarchical geographical modeling of user locations from social media posts. In *Proceedings of the 22nd international conference on World Wide Web (WWW 2013)*, pages 25–36. ACM.

Oluwaseun Ajao, Jun Hong, and Weiru Liu. 2015. A survey of location inference techniques on twitter. *Journal of Information Science*, 41(6):855–864.

Ji Ao, Peng Zhang, and Yanan Cao. 2014. Estimating the locations of emergency events from twitter streams. *Procedia Computer Science*, 31:731–739.

Miriam Cha, Youngjune Gwon, and HT Kung. 2015. Twitter geolocation and regional classification via sparse coding. In *Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM 2015)*, pages 582–585.

Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2010. You are where you tweet: a content-based approach to geo-locating twitter users. In *Proceedings of the 19th ACM International Conference Information and Knowledge Management (CIKM 2010)*, pages 759–768. ACM.

Zhiyuan Cheng, James Caverlee, and Kyumin Lee. 2013. A content-driven framework for geolocating microblog users. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(1):2.

Lianhua Chi, Kwan Hui Lim, Nebula Alam, and Christopher J Butler. 2016. Geolocation prediction in twitter using location indicative words and textual features. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT 2016)*, pages 227–234. ACL.

Ryan Compton, David Jurgens, and David Allen. 2014. Geotagging one hundred million twitter accounts with total variation minimization. In *Proceedings of the 2014 IEEE International Conference on Big-Data (BigData 2014)*, pages 393–401. IEEE.

Clodoveu A Davis Jr, Gisele L Pappa, Diogo Rennó Rocha de Oliveira, and Filipe de L Arcanjo. 2011. Inferring the location of twitter messages based on user relationships. *Transactions in GIS*, 15(6):735–751.

Mohammad Ebrahimi, Elaheh ShafieiBavani, Raymond Wong, and Fang Chen. 2017. Exploring celebrities on inferring user geolocation in twitter. In *Proceedings of the 21st Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2017)*, pages 395–406. Springer.

Mohammad Ebrahimi, Elaheh ShafieiBavani, Raymond Wong, and Fang Chen. 2018a. Leveraging local interactions for geolocating social media users. In *Proceedings of the 22nd Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2018)*, pages 803–815. Springer.

Mohammad Ebrahimi, Elaheh ShafieiBavani, Raymond Wong, and Fang Chen. 2018b. Twitter user geolocation by filtering of highly mentioned users. *Journal of the Association for Information Science and Technology (JASIST)*, 69(7):879–889.

Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. 2010. A latent variable model for geographic lexical variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing (EMNLP 2010)*, pages 1277–1287. ACL.

Cheng Guo and Felix Berkhahn. 2016. Entity embeddings of categorical variables. *CoRR*, abs/1604.06737.

Bo Han, Paul Cook, and Timothy Baldwin. 2012. Geolocation prediction in social media data by finding location indicative words. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 1045–1062. ACL.

Bo Han, Paul Cook, and Timothy Baldwin. 2014. Text-based twitter user geolocation prediction. *Journal of Artificial Intelligence Research (JAIR)*, 49:451–500.

Bo Han, AI Hugo, 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 2016)*, pages 213–217. ACL.

Brent Hecht, Lichan Hong, Bongwon Suh, and Ed H Chi. 2011. Tweets from justin bieber's heart: the dynamics of the location field in user profiles. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 237–246. ACM.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Liangjie Hong, Amr Ahmed, Siva Gurumurthy, Alexander J Smola, and Kostas Tsioutsiouliklis. 2012. Discovering geographical topics in the twitter stream. In *Proceedings of the 21st International Conference on World Wide Web (WWW 2012)*, pages 769–778. ACM.

Mans Hulden, Miikka Silfverberg, and Jerid Francom. 2015. Kernel density estimation for text-based geolocation. In *Proceedings of the 29th Conference on Artificial Intelligence (AAAI 2015)*, pages 145–150.

Sergey Ioffe and Christian Szegedy. 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 37th International Conference on Machine Learning (ICML 2015)*, pages 448–456.

Gaya Jayasinghe, Brian Jin, James Mchugh, Bella Robinson, and Stephen Wan. 2016. Csiro data61 at the wnut geo shared task. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT 2016)*, pages 218–226. ACL.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759*.

David Jurgens. 2013. That's what friends are for: Inferring location in online social media platforms based on social relationships. In *Proceedings of the 7th International Conference on Weblogs and Social Media (ICWSM 2013)*, volume 13, pages 273–282.

David Jurgens, Tyler Finethy, James McCorriston, Yi Tian Xu, and Derek Ruths. 2015. Geolocation prediction in twitter using social networks: A critical analysis and review of current practice. In *Proceedings of the 9th International Conference on Web and Social Media (ICWSM 2015)*, volume 15, pages 188–197.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Jey Han Lau, Lianhua Chi, Khoi-Nguyen Tran, and Trevor Cohn. 2017. End-to-end network for twitter geolocation prediction and hashing. In *Proceedings of the 8th International Joint Conference on Natural Language Processing (IJCNLP 2017)*, pages 744–753.

Rui Li, Shengjie Wang, Hongbo Deng, Rui Wang, and Kevin Chen-Chuan Chang. 2012. Towards social user profiling: unified and discriminative influence model for inferring home locations. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2012)*, pages 1023–1031. ACM.

John Lingad, Sarvnaz Karimi, and Jie Yin. 2013. Location extraction from disaster-related microblogs. In *Proceedings of the 22nd International Conference on World Wide Web (WWW 2013)*, pages 1017–1020. ACM.

Ji Liu and Diana Inkpen. 2015. Estimating user location in social media with stacked denoising auto-encoders. In *Proceedings of the 14th Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2015)*, pages 201–210. ACL.

Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. 2012. Where is this tweet from? inferring home locations of twitter users. In *Proceedings of the 6th International Conference on Web and Social Media (ICWSM 2012)*, volume 12, pages 511–514.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi, and Tomoko Ohkuma. 2016. A simple scalable neural networks based model for geolocation prediction in twitter. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT 2016)*, volume 9026924, pages 235–239. ACL.

Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi, and Tomoko Ohkuma. 2017. Unifying text, metadata, and user network representations with a neural network for geolocation prediction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, volume 1, pages 1260–1272. ACL.

Anastasios Noulas, Salvatore Scellato, Neal Lathia, and Cecilia Mascolo. 2012. Mining user mobility features for next place prediction in location-based services. In *Proceedings of the 12th IEEE International Conference on Data mining (ICDM 2012)*, pages 1038–1043. IEEE.

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2014)*, pages 701–710. ACM.

Afshin Rahimi, Timothy Baldwin, and Trevor Cohn. 2017a. Continuous representation of location for geolocation and lexical dialectology using mixture density networks. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*, pages 167–176. ACL.

Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2015a. Twitter user geolocation using a unified text and network prediction model. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics - 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2015)*, pages 630–636. ACL.

Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2016. pigeo: A python geotagging tool. In *Proceedings of the 54st Annual Meeting of the Association for Computational Linguistics (ACL 2014): System Demonstrations*, pages 127–132. ACL.

Afshin Rahimi, Trevor Cohn, and Timothy Baldwin. 2017b. A neural model for user geolocation and lexical dialectology. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (ACL 2017)*, volume 2, pages 209–216. ACL.

Afshin Rahimi, Duy Vu, Trevor Cohn, and Timothy Baldwin. 2015b. Exploiting text and network context for geolocation of social media users. In *Proceedings of the 14th Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT 2015)*, pages 1362–1367. ACL.

Vineeth Rakesh, Chandan K Reddy, and Dilpreet Singh. 2013. Location-specific tweet detection and topic summarization in twitter. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2013)*, pages 1441–1444. ACM.

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 (EMNLP-CONLL 2012)*, pages 1500–1510. ACL.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research (JMLR)*, 15(1):1929–1958.

Philippe Thomas and Leonhard Hennig. 2017. Twitter geolocation prediction using neural networks. In *International Conference of the German Society for Computational Linguistics and Language Technology*, pages 248–255. Springer.

Fengjiao Wang, Chun-Ta Lu, Yongzhi Qu, and S Yu Philip. 2017. Collective geographical embedding for geolocating social network users. In *Proceedings of the 21th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2017)*, pages 599–611. Springer.

Benjamin Wing and Jason Baldridge. 2011. Simple supervised document geolocation with geodesic grids. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (ACL-HLT 2011)*, pages 955–964. ACL.

Benjamin Wing and Jason Baldridge. 2014. Hierarchical discriminative classification for text-based geolocation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 336–348. ACL.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 15th Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2016)*, pages 1480–1489. ACL.

Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. 2013. Who, where, when and what: discover spatio-temporal topics for twitter users. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (ACM-SIGKDD 2013)*, pages 605–613. ACM.