Changes in Tweet Geolocation over Time: A Study with Carmen 2.0

Jingyu Zhang, Alexandra DeLucia, Mark Dredze
Johns Hopkins University
Introduction

Twitter Geolocation tools are useful for demographic studies in various topics

- Civil unrest
- Natural disasters
- Disease spread

Existing tools identify the location of tweets base on tweet metadata, tweet content, and social networks
Problem Statement

While widely used, geolocation tools tend to be English-centric and are often not evaluated for global coverage or performance across time and language.

We assess the following factors’ impact on geolocation tool Carmen:

- Language
- Country
- Time
Carmen: A Review

Introduced in Dredze et al. (2013), Carmen is a metadata-based geolocation tool that resolve locations from:

- Embedded coordinates in the Geo object
- Matching the Place object to internal location database
- Mapping user profile location string to internal location database
Carmen 2.0

- Compatible with Twitter API v2
- Performance optimization (25x faster geocode resolver)
- Expanded database with GeoNames
  - GeoNames Only
  - GeoNames + Carmen Original
Database Comparison

Carmen Original Database

- 7K location entries
- Inferred from tweets between May 2009 and Aug 2012 (primarily English tweets from US)
- Does not align with external knowledge base

Carmen 2.0

- 73K entries extracted from the GeoNames database
- Alternative names in many languages
- Hierarchical (CITY, ADMIN, COUNTRY) structure compatible with GeoNames
Evaluating Geotagging Performance

Geotagging tools should be able to accurately cover a wide range of locations:

- **Coverage**: for what portion of data can the geotagger propose a location
- **Accuracy**: how well the proposed locations compare to ground truth

We develop *multiple* metrics tailored to geotagging performance
Metrics for Geotagger

Coverage: percentage of data successfully mapped to a location

Accuracy:

- **Match Ratio** of level $L$: percentage of resolved tweets that is correct on level $L$. $L$ is one of \{country, admin, city\}
- **Distance**: geodesic distance between resolved and ground truth location
- **Acc@$K$**: percentage of resolved tweets such that the distance error does not exceed $K$ miles.
Experiments
Ground Truth Data

We introduce **Twitter-Global**, a new geolocation evaluation dataset collected from multiple Twitter API streams

- 15.3M geotagged tweets
- Collected from 2013 to 2021
- Covers a wide range of languages and countries
## Performance across Language

We create two subsets of English and Non-English data from **Twitter-Global**

<table>
<thead>
<tr>
<th>Language</th>
<th>Database</th>
<th>Coverage</th>
<th>(mr\text{_country} )</th>
<th>(mr\text{_admin} )</th>
<th>(mr\text{_city} )</th>
<th>(d)</th>
<th>Acc@10</th>
<th>Acc@100</th>
<th>Acc@1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>GeoNames-Only</td>
<td>49.58%</td>
<td>99.42%</td>
<td>95.63%</td>
<td>47.49%</td>
<td>853.9</td>
<td>0.81</td>
<td>0.85</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>GeoNames-combined</td>
<td>49.63%</td>
<td>99.43%</td>
<td>94.36%</td>
<td>47.69%</td>
<td>58.7</td>
<td>0.81</td>
<td>0.91</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Original</td>
<td>48.14%</td>
<td>99.35%</td>
<td>94.94%</td>
<td>48.90%</td>
<td>46.4</td>
<td>0.78</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>Non-English</td>
<td>GeoNames-Only</td>
<td>41.77%</td>
<td>99.36%</td>
<td>66.50%</td>
<td>20.13%</td>
<td>482.3</td>
<td>0.84</td>
<td>0.88</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>GeoNames-combined</td>
<td>41.78%</td>
<td>99.35%</td>
<td>66.83%</td>
<td>20.27%</td>
<td>105.3</td>
<td>0.84</td>
<td>0.90</td>
<td>0.99</td>
</tr>
<tr>
<td></td>
<td>Original</td>
<td>32.27%</td>
<td>98.95%</td>
<td>75.61%</td>
<td>14.22%</td>
<td>106.2</td>
<td>0.67</td>
<td>0.87</td>
<td>0.99</td>
</tr>
</tbody>
</table>

On Non-English data, GeoNames

- Substantially increased coverage
- Moderate increased accuracy-based metrics
Performance across Countries

We create two subsets of US and Non-US data from Twitter-Global

<table>
<thead>
<tr>
<th>Origin</th>
<th>Database</th>
<th>Coverage</th>
<th>( mr_{country} )</th>
<th>( mr_{admin} )</th>
<th>( mr_{city} )</th>
<th>( d )</th>
<th>Acc@10</th>
<th>Acc@100</th>
<th>Acc@1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>GeoNames-only</td>
<td>50.56%</td>
<td>99.37%</td>
<td>99.87%</td>
<td>53.66%</td>
<td>994.2</td>
<td>0.79</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>GeoNames-combined</td>
<td>50.60%</td>
<td>99.37%</td>
<td>99.87%</td>
<td>53.81%</td>
<td>23.6</td>
<td>0.79</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Original</td>
<td>51.03%</td>
<td>99.93%</td>
<td>99.96%</td>
<td>55.33%</td>
<td>23.7</td>
<td>0.79</td>
<td>0.91</td>
<td>1.00</td>
</tr>
<tr>
<td>non-US</td>
<td>GeoNames-only</td>
<td>42.63%</td>
<td>99.37%</td>
<td>61.51%</td>
<td>18.73%</td>
<td>439.3</td>
<td>0.84</td>
<td>0.89</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>GeoNames-combined</td>
<td>42.65%</td>
<td>99.37%</td>
<td>60.81%</td>
<td>18.88%</td>
<td>121.2</td>
<td>0.84</td>
<td>0.90</td>
<td>0.98</td>
</tr>
<tr>
<td></td>
<td>Original</td>
<td>32.89%</td>
<td>98.45%</td>
<td>66.11%</td>
<td>11.10%</td>
<td>118.0</td>
<td>0.67</td>
<td>0.87</td>
<td>0.99</td>
</tr>
</tbody>
</table>

On Non-US data, GeoNames

- Substantially increased coverage
- Achieved comparable accuracy with original database
Performance over Time

We create subsets of Twitter-Global for each year between 2013-2021

Main findings

- Due to change in metadata availability, Coverage dropped significantly after 2014
- GeoNames provide slightly better coverage regardless of metadata availability
Summary

- Introduced **Carmen 2.0**, an updated version of geolocation tool Carmen backed by an open-source gazetteer, GeoNames
- **Twitter-Global** is a Twitter geolocation evaluation dataset for language, country, and time ablation studies
- Significant **difference in performance in the ablation**, with higher performance for English and US-based tweets
- Geolocation tools should be **robust to language, country of origin**, and available metadata
- More **work is needed for a fine-grained study** on individual languages and countries
Thank you!

Analysis Code: https://github.com/AADeLucia/carmen-wnut22-submission

Carmen: https://github.com/mdredze/carmen-python
Supplementary
Table 1: The statistics of city, county, state, and country-level locations in the original Carmen location database and the new GeoNames database versions developed for Carmen 2.0. The GeoNames-augmented databases have more than 10 times the number of location entries than Original. Percentage refers to portion of the database dedicated to each granularity.