Micromobility in Smart Cities: A Closer Look at Shared Dockless E-Scooters via Big Social Data

Yunhe Feng\textsuperscript{1}, Dong Zhong\textsuperscript{2}, Peng Sun\textsuperscript{3}, Weijian Zheng\textsuperscript{4}, Qinglei Cao\textsuperscript{2}, Xi Luo\textsuperscript{5}, and Zheng Lu\textsuperscript{2}

\textsuperscript{1}University of Washington
\textsuperscript{2}University of Tennessee
\textsuperscript{3}Chinese University of Hong Kong, Shenzhen
\textsuperscript{4}Purdue University
\textsuperscript{5}SLAC National Accelerator Lab
Background

- **Pros**: affordability, easy accessibility via an app, and zero emissions.

- **Cons**: traffic rules, public safety, and parking regulations.

- **11%** of Paris residents reported using e-scooters either frequently or from time to time [1].

- According to NACTO[2], people took **38.5 million trips** on shared e-scooters in 2018.
Motivation

- A missing research perspective from the riders’ comments shared via smartphones, especially via social media apps.

- Social media is a good data source to investigate the e-scooter usages.
  - **Diversity**: multimodal social data (text, image, timestamps, GPS, emoji, etc.) enables detailed profiles of e-scooter sharing services in various aspects.
  - **Scalability**: performing a large-scale study of micromobility with social media data is flexible and effortless.
  - **Transparency**: public social media platforms (e.g., Twitter) addressing potential concerns of research non-reproducibility.
Methodology

- **Use Twitter to collect data**
  - Free APIs to collect tweets of interest
  - The fourth most active social networking platform
  - Users expect the data they post on Twitter to be publicly available

- **Big data analytics**
  - Spatio-temporal Visualization
  - Topic discovery - latent Dirichlet allocation (LDA)
  - Optical character recognition (OCR)
  - Logo detection and recognition
  - Sentiment analysis
Data Collection & Bot Detection

- Twitter's Streaming APIs with keywords of *scooter* and emoji
  - 5.8 million tweets generated by 2.7 million users
  - 178,048 different image URLs (33,851 expired)
  - October 6, 2018 to March 14, 2020

- Remove bots [3]
  - Those who posted more than 525 scooter-tagged tweets, i.e., more than one such tweets per day, during the data collection period
  - Those who posted over 100 scooter-tagged tweets in total and the top three frequent posting intervals covered at least their 90% tweets.
Data Cleanup

• Word co-occurrence analysis
  ❑ Taylor, Swift, Justin, Bieber, Scott, Samuel, Braun, Ariana, Grande, and Borchetta
  ❑ 1,541,815 related tweets were deleted

• Further reduce false positives
  ❑ Tweets containing the word of **Share**
  ❑ Tweets containing shared e-scooter brands including **Bird, Lime, Spin, Bolt, gruv, Lyft, Sherpa, VeoRide, Taxify, Jump, RazorUSA, Scoot Networks**, and **Skip**.
Temporal Distribution

- The United States accounts for more than 82% of all collected tweets.
- Usage peaks in summer and drops in winter.
- New Zealand & Australia ↓, Germany & India ↑, and United Kingdom & France →.
- The most active time during weekdays (weekends) is between 10:00 am to 5:00 pm (12:00 pm and 7:00 pm).
California (28.8%) and Texas (11.7%) account for more than 40% of all tweets.

Six (CA, TX, GA, FL, NC, OH) out of the top ten states with the highest percentages were among the ten most populous states.

After normalizing by state population, we obtained a relatively smooth distribution.
Topic Discovery

• Data preprocessing
  ❑ Each tweet content is a single document.
  ❑ Corrected misspelled words, removed stop words, tokenized words, and lemmatized words.
  ❑ Performed the term frequency inverse document frequency (TF-IDF).

• Latent Dirichlet Allocation (LDA)
  ❑ # of topics = 12 achieving the highest coherence score
  ❑ We clustered the 12 topics into four categories
    ➢ Deployment
    ➢ Stakeholder
    ➢ Operation
    ➢ Emotion
# Topic Discovery

## TABLE I: The extracted topics using the LDA topic model

<table>
<thead>
<tr>
<th>Rank</th>
<th>Stakeholder</th>
<th>Emotion</th>
<th>Deployment</th>
<th>Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rider</td>
<td>Gig</td>
<td>Worker</td>
<td>Company</td>
</tr>
<tr>
<td>1</td>
<td>kid</td>
<td>gig</td>
<td>electr</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>kick</td>
<td>worker</td>
<td>startup</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>adjust</td>
<td>built</td>
<td>compani</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>wheel contractor</td>
<td>san</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>height</td>
<td>gen</td>
<td>market</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>amp</td>
<td>economi</td>
<td>via</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>child</td>
<td>lemon</td>
<td>new</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>light</td>
<td>kmh</td>
<td>launch</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>adult</td>
<td>stabl</td>
<td>ford</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>boy</td>
<td>libbi</td>
<td>tech</td>
<td></td>
</tr>
</tbody>
</table>
Deployment in Cities

- 3359 exact GPS coordinates located in 579 cities.
- Most extensive scooter deployments occurred in large cities at East Coast and West Coast, and other metropolises.
- Three cities (Los Angeles, San Francisco, San Diego) in California contributed more than 15.6% of all GPS-tagged tweet.
Policies and Regulations

• We surveyed policies on shared e-scooters in ten cities generating the most GPS data.
• Common rules
  □ Self-protection requirements, e.g., wearing helmets
  □ Riding behaviors, e.g., no phone usage while riding
  □ Traffic restrictions, e.g., reasonable max speeds from 15 mph to 30 mph
  □ Parking rules, e.g., not blocking sidewalks
Rider Profiles

• A great gender gap in shared e-scooters with 34.86% identified as female and 65.14% as male, being consistent with a study [4] by Portland State University, which reported 34% riders identified as a woman, 64% as a man, and 2% as transgender or non-binary.

• Only 4.17% riders were recognized as kids.

• 83.51% users did not wear a helmet when riding, which might be one of the most common risky behaviors.
• All the top 10 most frequent mentions are e-scooter brands.
• The top 3 of market shares based on tweet text and recognized logos, i.e., Lime, Bird, and Lyft, agree with the top @mentioned accounts.
Transaction Analysis

- Optical Character Recognition (OCR) to extract the trip information.
- Regular expressions to extract the payment and trip duration.
- Median payment and median duration were $3.8 and 15.0 minutes.
- Average payment and duration were $8.9 and 44.3 minutes with standard deviations of 14.6 and 95.7.
Self-reported Injury

**TABLE II: Self-reported injury categories**

<table>
<thead>
<tr>
<th>Head (22.88%)</th>
<th>Trunk &amp; Hands (27.45%)</th>
<th>Leg &amp; Feet (49.67%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chin</td>
<td>Arm</td>
<td>Ankle</td>
</tr>
<tr>
<td>5.23%</td>
<td>8.50%</td>
<td>11.11%</td>
</tr>
<tr>
<td>Eye</td>
<td>Elbow</td>
<td>Heel</td>
</tr>
<tr>
<td>7.19%</td>
<td>5.88%</td>
<td>0.65%</td>
</tr>
<tr>
<td>Mouth</td>
<td>Finger</td>
<td>Knee</td>
</tr>
<tr>
<td>5.88%</td>
<td>2.61%</td>
<td>24.84%</td>
</tr>
<tr>
<td>Nose</td>
<td>Hand</td>
<td>Thigh</td>
</tr>
<tr>
<td>5.88%</td>
<td>11.76%</td>
<td>3.92%</td>
</tr>
<tr>
<td>Others</td>
<td>Others</td>
<td>Others</td>
</tr>
<tr>
<td>11.76%</td>
<td>0.65%</td>
<td>10.46%</td>
</tr>
</tbody>
</table>

- **Knee protective gears** are required because of their highest injury frequency in all body parts.
- **Fingerless gloves** can be a good choice for riders to avoid hand bruises, the second most common injuries, and enable touching smartphone screens at the same time.
- **A helmet with chin protection** is a must because over half of the reported chin wounds were very serious.
Parking Behaviors

- 37.39% e-scooters were docked at right places properly, and the rest **62.61% were in wrong places**.
- Among those e-scooters parked improperly, 34.78% were parked in the middle of sidewalks; 4.78% were placed indoors; 5.65% were vandalized; and 17.39% were parked in other wrong areas.
- **Blocking sidewalks** was the most common improper parking behavior.
- **Vandalism**.

(a) Under water  (b) On fire  (c) Up in trees
Emotion Analysis

- **Emoji**: 62.6% positive, 16.2% neutral, and 21.2% negative
  - Positive
  - Negative

- **Emoticon**: the positive emotions override the negative ones

- **World cloud**

![Word clouds](image-url)
Limitation and Future Work

• **Data Bias**
  - Twitter user demographics are skewed
  - Not all shared e-scooter riders are Twitter users
  - Not all Twitter users tweet their opinions on shared e-scooters

• **Beyond an Overview of Shared E-scooter Usage**
  - Investigate shared e-scooter usage at the state/city level
  - Further explore specified aspects of share e-scooter usage, e.g., injuries, parking behaviors
  - Profile the dynamic changing and evolvement of shared e-scooter usages
  - Data and model fusion (e.g., third-part data, and expert-approved simulators)
Conclusion

• Leveraged massive volumes of heterogeneous Twitter data, including text, @mentions, GPS data, general photos, screenshots of e-scooter apps, emojis, and emoticons, to study e-scooter ridesharing services on a large scale.

• After performing a comprehensive data preprocessing to remove noise and reduce false positives, we summarized 12 popular topics using the LDA topic model.

• For each of the extracted topics, we reported the profound insights and patterns, such as the popularity in different cities, the gender gap of riders, e-scooter market shares, transaction information, injury types, parking behaviors, and emotions from the public.
1. ODOXA, New modes of travel are disrupting transport in Ile-de-France, 2019. [Online]. Available: https://shorturl.at/nBLU0
Thank you!

Q&A