TRAFFIC DATA ANALYTICS

WAZE VS. BLUETOOTH

**Students:** Brooklynn Hauck (SRU); Tracy Liu (CSUST)
**Mentors:** Dr. Lee Han; Nima Hoseinzade; Yuandon Liu; Dr. Kwai Wong
Introduction

Research Goal

Waze and Bluetooth

Steps towards the goal

Data Processing
Calculate Parameters
OpenDIEL
Data Visualization and Analyzation
Picking Conditions

Conclusion
Research Goal

Use **Bluetooth Data** as ground truth

Started out as a very open goal:

- Is the Data from Waze **reliable** to be applied in traffic research?

  - **1** What *day of the week* is the Waze Data most reliable?
  - **2** What *time of day* is Waze Data most reliable?
  - **3** Which *factors* affect the reliability of Waze?
    
    (Traffic Volume, Speed or Length of Segment)
What is Waze?

- A downloadable GPS app for both IOS and Android, provided by Google, that provides turn-by-turn navigation information and user submitted Travel Time and Route Details
- Data is collected by Online data source provided by Waze
- Data has not been tested and is an unreliable source, though Universities have been using it for Research purposes.
What is Bluetooth?

• A wireless technology standard for exchanging data between fixed and mobile devices over short distances
• Data is collected by two Bluetooth detectors with a set distance between the two.
• Bluetooth has been proven to be a reliable source of data.
About Data

Online Waze data source
Capturing real-time json files
Convert to csv file
Waze Travel Time and Speed data

Online Sevierville Bluetooth data
Capturing real-time XML files
Convert to csv file
Bluetooth Speed and Travel Time

Data

Speed & Travel Time

One month

Each second
Data Processing

- Sorts through both Bluetooth and Waze data
- Removes Duplicate Data
- Removes Waze Historical Data
- Merge Bluetooth and Waze Data based on the time of day and Segment

Code was written in R
Removing Historical Data
Calculate Parameters

• We choose these three error formulas as parameters:
  • *MAE* (Mean Absolute Error)
  • *RMSE* (Root Mean Standard Error)
  • *MAPE* (Mean Absolute Percentage Error)

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y - \bar{y}|
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y - \bar{y})^2}
\]

\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y - \bar{y}}{y} \right|
\]

<table>
<thead>
<tr>
<th>Time</th>
<th>bIS</th>
<th>waS</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:01:11</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>00:01:12</td>
<td>NA</td>
<td>34</td>
</tr>
<tr>
<td>00:01:13</td>
<td>45</td>
<td>NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Time</th>
<th>bIS</th>
<th>waS</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:01</td>
<td>40</td>
<td>38</td>
</tr>
<tr>
<td>00:02</td>
<td>44</td>
<td>42</td>
</tr>
<tr>
<td>00:03</td>
<td>34</td>
<td>36</td>
</tr>
</tbody>
</table>

\(y\) = Bluetooth speed sample(after aggregate) in each hour

\(\bar{y}\) = Waze speed sample(after aggregate) in each hour

\(N\) = number of Bluetooth/Waze speed samples(after aggregate) in each hour
The MAE and RMSE of most segments do not exceed 10, the MAPE does not exceed 0.1.

There is big difference in different segments, which may be caused by other factors.
OpenDIEL

• R Code was ran on Comet in OpenDIEL to produce the distribution graphs
  • *Over 1800 graphs total*
• 3 different codes to produce all 1800 graphs
  • *Create Folders*
  • *Produce Graphs*
  • *Separate Graphs (Segments; Days of the week)*
# OpenDIEL Codes

**Create Folders**
- Folder for one Segment and One Day (930 Folders)
- Copies data Specific for that day
- Creates a text file with the Day and Segment

**Produce Graphs**
- Sorts through Bluetooth and Waze Data
- Pulls out all the data for a Specific Day and Segment
- Prints out distribution graphs
  - *Speed*

**Separate Graphs**
- Puts graphs into folder based on the Segment and Day of the week
Data Visualization

• Printed graphs based on:
  - Speed
  - Error Percentage
  - Traffic Volume

• Types of Graphs
  - Distribution Graph
  - Line Graph
  - Box Plot
  - Scatter Plot
Data Analyzation

- What we did?
  - Compare based on different segments
  - Compare based on the day of the week
  - Compare based on the Traffic Volume
  - Picked two segments with good MAPE and two segments with bad MAPE
  - Compare based on the Length of the Segment
  - Compare based on the Speed
Mean MAPE in 31 Segments
Comparison based on the days of the week
Comparison based on the Traffic Volume
Segments Used for Comparison

2

18

8

26
Segment 2

- TN-66 & Boyds Creek Rd to TN-66 & Huffaker Rd (Sevierville, TN)
- Speed Limit: 45 mph
- Length of Segment: 1 mile
Segment 8

- Dolly Parton Pkwy & Veterans Blvd to Dolly Parton Pkwy & Food City (Sevierville, TN)
- Speed Limit: 35 mph
- Length of Segment: 0.4 mile
- Red Lights Passed: 1
Segment 18

- Tn-66 & Huffaker Rd to TN-66 & TN-139 (Sevierville, TN)
- Speed Limit: 45 mph
- Length of Segment: 1.2 mile
Segment 26

- South Blvd to Collier Dr. (Pigeon Forge, TN)
- Speed Limit: 35 mph
- Length of Segment: 0.8 mile
- Red Lights Passed: 1
Comparison based on the Length of a Segment

Segment 8
- Length of Segment: 0.4 mile
- Lights Passed: 1

Segment 18
- Length of Segment: 1.2 mile
- Lights Passed: 1
Comparison based on the Length of a Segment (Number of Stop lights)

Segment 2
- Length of Segment: 1 mile
- Lights Passed: 0

Segment 26
- Length of Segment: 0.8 mile
- Lights Passed: 1
Comparison based on the Speed

1) Speed points are close to the perfect line (Bluetooth speed = Waze speed)
2) Range 40~55 mph

1) Range 25~45 mph
2) Waze Speed < Bluetooth Speed
Comparison based on the Speed

Grouping in 5 speed levels:
- speed $\leq 20$
- 20 < speed $\leq 30$
- 30 < speed $\leq 40$
- 40 < speed $\leq 50$
- 50 < speed

when speed less than or equal to 20, or more than 50 mph, MAPE is higher than other speed levels.
When speed is from 30 to 40 mph, MAPE is very low
Is Waze Reliable?

Yes, but only under these conditions:

- **Day of the week?**
  - *Any day of the week works!*

- **Time of day?**
  - *Day time; Roughly between 7am and 8pm*

- **Traffic Volume?**
  - *Traffic Volume need to be High*
  - *The higher the Traffic Volume the lower the MAPE*

- **Speed?**
  - *Speed needs to be between 30 mph and 45 mph*

- **Length of the segment?**
  - *The Length of the segment should long*
  - *Longer Segments have lower MAPE*
  - *Preferably with no stop lights to pass through*
Acknowledgement

This project has been done during the tenure of the RECSEM program of 2018. The program took place at The University of Tennessee. The program is funded by JICS, NSF, and UTK. Nothing could have been accomplished without the guidance of our mentors, Dr. Lee Han, Dr. Kwai Wong, as well as the graduate students, Nima Hoseinzadeh, Yuandong Liu.
References


Any Questions?