Report on EXIMIUS: A Measurement Framework for Explicit and Implicit Urban Traffic Sensing

Daniel Pattathil, dp865

1 SCIENTIFIC QUESTION

This research paper aims to create a stronger measurement framework in order to estimate traffic sensing following the two generally accepted methods, explicit and implicit. The two types of data utilized for this particular project, EXIMUS, were from vehicular GPS data from 3,000 cars (explicit) and cellular phone data from 3,000,000 users (implicit) all in Hefei, China. This way, there can be an easier determination of a crowdedness level in order to help both on the transportation and telecommunications sides. Both forms of large scale data help inform the researchers as they try to find trends in urban traffic.

2 EXISTING WORK

There already has been a good amount of work done in order to use large scale data to estimate the levels of congestion and flow in urban areas, such as in the other papers referenced from the paper. There has been work done using sensors/GPS, image processing, mobile networks, social media, emissions, and more all in order to try to gauge the actual level of traffic, but generally most previous research takes one of the approaches, either implicit or explicit sensing methods. There have been tangential processes, such as looking at factors such as building occupancy in nearby traffic areas to better estimate crowdedness levels, however, not much work has been done to compare the relative effectiveness of the previous solutions. By using terabytes of data with both categories of methods, they are able to utilize the differences in each to highlight different key points in their paper.

3 AUTHOR'S CONTRIBUTION

The major contribution of the work is in proposing a framework to assess the formally defined explicit and implicit traffic sensing in future works. Going over many of the factors in why both methods can be used, I was interested in seeing the difference both in data and accuracy. I was not surprised to see that having a much better temporal coverage when using the implicit data was helpful for the various street location assessments.

It was beneficial to see that there were many factors that contributed, such as spatial when looking at the roads themselves and temporal when trying to consider rush hours/time of day. When using Mean Absolute Percentage Error (MAPE) as an indicator, it was noted that longer roads such as highways provided more accurate readings and that time intervals of only five minutes for the geo-location data was less accurate than using the longer twenty minute intervals. The new contribution here was the combination of both collection methods in order to attempt to mitigate the weakness of each individual approach. Using this "fusion" type allowed for good estimations, but lacked strength in the prediction for future analysis.

4 STRENGTHS AND WEAKNESSES

I find that the biggest strength of this paper is the combination of the explicit and implicit methods. It is clear by the MAPE benchmarks from the fusion of the two that estimations of the crowdedness level were much more accurate in most road conditions. In addition, I believe that many contextual factors were considered when performing this research and finding the provided outcomes.

There are also a few possible weaknesses in the approach to solving this research problem. I think the first issue was behind the data collections and assumptions made. I noticed that the basis on which both methods were building off were both fairly low accuracy already. In addition to this, there was the factor that all this information had come from one month's worth of data in one particular city. I also question the veracity of the mobile network data, considering that determining speed was harder than total travel time.

5 POTENTIAL SOLUTIONS

I think that a potential future work would be to use different types of explicit and implicit data together in order to gain more meaningful information. There were clearly many types of urban sensing methods used in existing work, and I think the addition of computer vision, social media, and other forms of data analysis could be major benefits. I would also hope to apply this framework on other datasets to prevent any unassessed variables from affecting the validity of EXIMUS. I think these changes could allow for more discernible differences in the previous methodologies compared to the new theory.