2019 UDOT RESEARCH PROBLEM STATEMENT

*** ***Problem statement deadline is Feb. 6, 2019. Submit statements to UTRAC@utah.gov.*** ***

Title: Utilizing machine learning to cross-check traffic data and understand urban mobility **No. (Office Use):** 19.03.01

Written By: Xianfeng Terry Yang, Shandian Zhe **Organization:** University of Utah **Email:** x.yang@utah.edu, zhe@cs.utah.edu **Phone:** 801-585-1290

Submitted By UDOT Employee: Kelly Burns **Email:** kburns@utah.gov **Phone:** 801-386-6118

UDOT Champion (if different): **Email:** **Phone:**

Select ONE Subject Area ☐ Materials/Pavements ☐ Maintenance ☒ Traffic Mgmt/Safety ☐ Structures/Geotech ☐ Planning ☐ Perf Mgmt/Data Analytics ☐ Public Transportation ☐ Other

1. Describe the problem to be addressed:

Traffic data play as a key role in supporting various traffic operation and transportation planning activities. In UDOT Traffic Operation Center (TOC), several on-line platforms such as Freeway Performance Metrics and Performance Measurement System (PeMS), and iPeMS are currently used for data visualization and sharing. PeMS stores point data collected by roadside radar sensors, loop detectors, and microloops while iPeMS provides statewide HERE probe data. Notably, point data in PeMS is good for studying volumes, speeds, and occupancy but it has coverage limitations due to the limited number of sensors are used. In contrast, iPeMS can provide statewide traffic state estimation using probe vehicle information. However, the estimates have a high potential to be biased considering the low prove penetration rate (around 2% of the entire traffic).

By comparing the data from PeMS and iPeMS at the same roadway sites, we have observed significant differences of 5-min speed and flow between the two databases. Therefore, a critical question raises here is how to obtain the "ground truth" based on the information variations. Considering the fact that PeMS data quality is mainly affected by detection accuracy and iPeMS suffers from bias in data, this project will develop a machine learning model to mitigate data variations from different sources and produce estimations of the "ground truth". In 2018, UTRAC supports two projects "Multi-Stage Algorithm for Detection-Error Identification and Data Screening" and "Addressing Big Data Biases", where the first one aims to evaluate the data accuracy in PeMS and the second focuses on assessing HERE biases. This project will ground on the outcomes of the two on-going projects and investigate how to conduct data fusion so as to provide more reliable statewide traffic state estimates.

2. Write the project objective (25 words or less):

Use machine learning technique to correct data in iPeMS and estimate statewide traffic mobility pattern by fusing data across PeMS and iPeMS.

3. Explain why this research is important:
(In response, consider addressing specific UDOT goals, applicability in Utah or other states, etc.)

PeMS data, collected from road sensors, are usually considered to be more "reliable" than estimates based on probe data. Hence, the observed differences of 5-min speed and flow between PeMS and iPeMS indicate high potential of bias and inaccuracy in iPeMS. However, as iPeMS can provide statewide traffic information, it has been used to support many traffic operation tasks when PeMS data are not available. Hence, the lack of correcting iPeMS data can result in unreliable inputs and failure of operation activities.

The proposed work in this project would provide a solution on how to integrate two data resources, grounded on the multi-view machine learning concept. The outcome of this project will yield statewide traffic estimation which absorb the "reliability" of PeMS data and "high network coverage" of iPeMS data. The proposed algorithm can be integrated into current UDOT data visualization platforms in future endeavors.

4. List the major tasks:

1. Implement the data screening algorithm developed from the on-going UTRAC project to a broader area for understanding the performance of PeMS data
2. Investigate the iPeMS data biases and perform cross-check between PeMS and iPeMS.
3. Develop machine learning algorithms (multi-view learning) to estimate statewide traffic mobility pattern (e.g., 5-min speed and flow rate) considering the potential errors in both PeMS and iPeMS data.
4. Develop a work plan for implementing the machine learning algorithm on current UDOT data visualization platforms.
5. Prepare the final project report.

5. List the expected deliverables (reports, manual, specification, design method, training, etc.):

1. Report of PeMS data quality assessment in broader areas of Utah.
2. Report of evaluating data quality of iPeMS by comparing with PeMS.
3. Machine learning algorithm for estimating statewide traffic patterns.
4. Work plan for algorithm implementation in the future.

6. Describe how the research results will be implemented:
(In response, consider addressing UDOT leader support, process or standard improvement, etc.)

The research results of this project will help UDOT TOC coordinate the efforts of managing different databases for better data usage. As both PeMs and iPeMS have similar data visualization functions but with different data processing algorithms. There are two possible ways of implement the proposed algorithm. The first way is to replace the current algorithm in iPeMS by the developed machine learning models. The second possibility is to add one more layer of data visualization on either PeMS or iPeMS for sharing the more reliable statewide traffic state estimates.

Notably, the task 4 of the proposed work is to work with UDOT TOC staffs to develop a plan for results implementations. The co-PI of the project, who is a professor in Computer Science and has a lot experience in developing on-line data processing and visualization platforms, will contribute in providing a feasible data structure for the implementations.

7. Requested from UDOT: $50,000 **Other/Matching Funds:** $40,000 **Total Cost:** $90,000
(or UTA for Public Transportation)

8. Outline the proposed schedule, including start and major event dates:

The schedule assumes an Oct 2019 start day and an 18-month project period:

Oct 2019: project kick off meeting.
Nov 2019 – Mar 2019: Task 1.
Apr 2019 – Jun 2019: Task 2.
July 2019 – Oct 2019: Task 3.
Nov 2019 – Dec 2019: Task 4.
Jan 2021 – Mar 2021: Final report preparation