Adventures in Transportation from a (Big) Data Perspective

Transportation is a part of our daily lives. We ride bikes (sometimes rented), drive cars, ride buses and walk. Items we buy in stores are brought to the stores via trucks and trains. It is well known that transportation is an important part of our lives and economy and has a large impact on the environment. What may be less well known is how diverse and interesting the data generated by transportation infrastructure is. High-definition radar sensors on freeways and arterials (side-roads) count vehicles, but also measure the speed and length of those vehicles; traffic lights report every time a traffic signal changes and, also, when a pedestrian pushes the "walk button"; busses report if they are on time or late, also how many people got on and off the bus, how long the bus waited at a bus stop, and more; infrared sensors count bicycles and pedestrians. The data collected from these sensors is used for management of the transportation system including display on Travel Time signs and Bicycle Count signs (see Figure 1 and Figure 2).

This talk gives a tour through transportation data sources from a data management perspective. Data sources are described, but more importantly issues in obtaining and managing that data are discussed. In particular, transportation data is an excellent example of Big Data Variety. We see variations in data coming from different cities and different transportation agencies, variations in data coming from different types of sensors and also a need to combine data across various data sources in order to more completely understand activity across the transportation system. In addition to the tour of data sources, we discuss how the data is used in practice currently and provide ideas of how it might be used in the future.

We conclude the talk by discussing areas where computer science research is applicable in the transportation data domain. While we do have some preliminary solutions for archiving the varied transportation data, and those solutions will be discussed, this talk is really about describing an interesting, complex application domain (transportation) and highlighting potential research problems in that area.

Speaker Bio: The speaker has 10 years experience working with transportation agencies and managing their data. In this role, she manages a 3TB public archive of transportation data. She is also big data researcher with a PhD in databases and 25 years of experience designing and implementing research data management systems. The combination of experience provides a unique perspective that underlies this talk.

Figure 1 Travel Time Sign in Portland, OR

Figure 2 Bicycle Count Sign on the Hawthorne Bridge in Portland, OR