

DATA QUALITY ASSESSMENT AND REAL-TIME EVALUATION OF GPS PROBE DATA

CROSS-REFERENCE TO RELATED PATENT APPLICATIONS

This patent application claims priority to U.S. provisional application 61/841,452, filed on Jul. 1, 2013, the contents of which are incorporated in their entirety herein.

FIELD OF THE INVENTION

The present invention relates to analyzing GPS data. Specifically, the present invention relates to a system and method of assessing the relevancy of bulk GPS probe data, determining the contribution of additional probe data from different vendors, and performing real-time evaluations of the probe data for vertical commercial applications.

BACKGROUND OF THE INVENTION

Data generated by geographical position systems (GPS) is currently sold in bulk, by the number of data points per day or per month. Generally, this data may be packaged in different ways—for example, in the form of “raw” or unprocessed probe data points, or in the form of processed probe data that reflects traffic speed on a roadway network. Ingests of raw probe data include data points of which many will not be relevant to the purchaser, and there is no current methodology for evaluating how much data in a bulk dataset of raw probe data is pertinent from the collection of information provided by each vendor. Similarly, there is no existing framework in the existing art for determining the value of a data point in a dataset that can be used to comparatively evaluate different vendors.

Raw probe data is useful for extracting information about traffic conditions on roadways, such as for example vehicular speed. Once a subscription to bulk raw probe data from a set (N) of vendors is undertaken, however, there is no current methodology for determining how much further value each additional vendor (N+1) provides for improving the analysis of roadway conditions like traffic flow from speed. In other words, there is no known framework in existence that permits traffic engineers to judge whether the accuracy of data extracted from a GPS dataset can be improved by additional subscriptions to vended probe data.

Additionally, there is no current methodology for performing a real-time evaluation of raw probe data to enable a prediction of data quality and realize a distribution of value extracted from an analysis of the quality of data points in a dataset. Because of the large number of GPS devices in use today, a real-time tool for foreseeing future roadway conditions such as traffic flow from known data would have significant utility in the marketplace, and would enable monetization of the value embedded within datasets comprised of raw probe data.

BRIEF SUMMARY OF THE INVENTION

The present invention provides a system and method of determining quality of raw GPS probe data in a vended dataset. Data is usually provided by GPS firms on a subscription basis, and as noted herein, may be provided in either a raw or unprocessed form, or pre-processed so that traffic speed is already known. Regardless, the present invention provides a framework for assessment of the quality of data in

a dataset to enable evaluation of the data points contained therein, resulting in a number of benefits and objectives as noted throughout this disclosure.

In one embodiment of the present invention, a system and method of assessing a value of traffic information in a set of GPS probe data is disclosed in which incoming raw probe data is initially analyzed to “clean-up” the dataset for removal of unnecessary information. The data is then mapped to roadway links, and smeared to fill in missing values that are an inherent characteristic of GPS datasets. The resulting output is then analyzed to determine the coverage value of data provided by a given vendor, and enable a comparison of a different vendors.

Another embodiment of the present invention involves evaluating a contribution of further vendors of probe data to an existing dataset. This embodiment seeks to determine how much additional value is added by subscribing to a dataset provided by a new vendor. Coverage surfaces are constructed for a full dataset that includes the new vendor, and for a dataset that excludes data provided by the new vendor. The coverage surface excluding the new vendor is subtracted from the first coverage surface to determine the added coverage surface. This added coverage surface is then used to calculate the value of data provided by the new vendor by spatially comparing the number of data points with those provided by other vendors across a common length with a geographical area.

Still another embodiment of the present invention provides a system and method for a real-time performance evaluation of continually-ingested probe data. Historical coverage profiles and data count profiles are built for each vendor, for each day of the week, for raw probe data ingested from a plurality of vendors on a periodic basis. These historical coverage profiles and the data count profiles are updated at specified time intervals, and an evaluation of probe data is performed for all of the vendors on a periodic basis to project a value of probe data for a next incremental time period, so that where the full dataset that includes data from all participating vendors for the time period is valued at a value X, values of contributing datasets are fractions of the value X, proportional to their area of coverage. This embodiment permits a real-time evaluation of probe data to project data quality on a forward-looking basis, and may be used to establish a database of vendors and a framework for monetizing data embedded in raw probe data, such as an auction-based trading platform. Yet another embodiment of the present invention therefore involves commercializing GPS probe data subjected to the above analyses to determine the quality and value of data in a dataset.

It is therefore one objective of the present invention to provide a framework for evaluating how much data in a bulk dataset of raw probe data provided by each vendor is pertinent for determining traffic information, such as speed. It is also an objective of the present invention to provide a framework for determining the value of a data point in a dataset that can be used to comparatively evaluate different vendors. Another objective is a framework for determining how much incremental value is provided by additional vendors for improving the assessment of roadway conditions like traffic flow. Still another objective is to provide a framework for real-time evaluation that can be used to predict traffic conditions and generate further revenue streams from processing of raw probe data.

Other objectives, embodiments, features and advantages of the present invention will become apparent from the following description of the embodiments, taken together with the