Future Connected Vehicles: Challenges and Opportunities for Spatio-temporal Computing

Reem Y. Ali[⋆], Venkata M.V. Gunturi[#], Shashi Shekhar[⋆], Ahmed Eldawy[⋆], Mohamed F. Mokbel[⋆], Andrew J. Kotz[§], William F. Northrop[§]

[⋆]Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN
[#]IIIT-Delhi, New Delhi, India
[§]Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN
{alireem,shekhar,eldawy,mokbel}@cs.umn.edu, gunturi@iiitd.ac.in,
{kotzx029,wnorthro}@umn.edu

ABSTRACT

Modern vehicles are increasingly being equipped with rich instrumentation that enables them to collect location aware data on a wide variety of travel related phenomena such as the real-world performance of engines and powertrain, driver preferences, context of the vehicle with respect to others nearby, and–indirectly–traffic on the transportation network itself. Combined with their increased access to the Internet, these *connected vehicles* are opening up vast opportunities to improve the safety, environmental friendliness, and the overall experience of urban travel. However, significant spatial computing challenges need to be addressed before we can realize the full potential of connected vehicles. This paper presents some of the open research questions under this theme from the perspectives of query processing, data science and data engineering.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications–Spatial databases and GIS, Data mining

Keywords

Connected Vehicles, Spatial and Spatio-temporal Graphs, Spatial and Spatio-temporal Data Mining, Spatial Big Data, Transportation, Spatial Statistics

1. VISION

Modern vehicles are increasingly becoming "smarter." They are typically equipped with rich instrumentation such as GPS receivers, Internet access, wireless local area network (e.g, a dedicated short range communications technology), increasingly powerful electronic control units (ECUs), and engine sensors to periodically measure sub-system properties of the vehicle [11, 7, 9]. Collectively, these technologies enable a vehicle to get *"connected"* to its surroundings (i.e., the transportation network and other vehicles on it) and collect location-aware data on a wide range of travel related aspects including real-world performance of engines, driver preferences, road and traffic conditions (in an indirect fashion) along the journey, etc. Combined with a potential to communicate with cloud computing infrastructure for analytics, connected vehicles open an opportunity to make urban travel safer, more efficient, environmentally friendly and enjoyable. From a safety perspective, a connected vehicle could talk to neighboring vehicles for mutual avoidance of traffic accidents. For instance, DaimlerChrysler, Honda Motors, and General Motors have recently conducted pilot studies on using vehicle-to-vehicle communications for crash avoidance [7, 19]. Similarly, a connected vehicle could also perform prognostics on its engine performance (potentially with the aid of a cloud computing infrastructure) to predict engine failure events.

Connected vehicle technologies could also help in reducing fuel consumption and greenhouse gas (GHG) emissions, thus promoting global economic and environmental stewardship and helping bring the nation closer to the goal of energy independence [3]. A recent research strategic plan [21] from the US Department of Transportation also points to this vision. Data coming from connected vehicles opens up the potential to answer novel routing queries beyond the traditional shortest-distance or earliest arrival path queries. Some sample queries now made possible include: "What route is the most fuel-efficient or has the least emissions?"; "What route is the most eco-friendly for my style of driving?" etc. This data can also be used to improve engines by providing the ability to adapt their real-world behavior to improve fuel economy, emissions or drivability. For instance, engine ECUs–a microprocessor which reads from sensors and controls a series of actuators (e.g., throttle position and spark plug timing) on the internal combustion engine–could use rules learned from data to adapt its behavior according to the situation (driving in a busy downtown versus suburban regions, or driving with frequent stops, etc.). These rules could be personalized for drivers as well (e.g., ECU rules for gentle versus aggressive driving). Such adaptability of rules is not available in current designs [20].

Apart from improved safety and environmental friendliness, connected vehicles can also be used to improve travel-

SIGSPATIAL'15, November 03-06 2015, Bellevue, WA, USA
© 2015 ACM. ISBN 978-1-4503-3967-4/15/11$15.00
DOI: http://dx.doi.org/10.1145/2820783.2820885.

ers' overall mobility experience by helping to reduce traffic congestion through better signal and traffic control [18]. For instance, data from connected cars could be used to provide advance alerts on traffic conditions and hazardous road conditions as sensed by the cars up ahead in the transportation network. Similarly, this data can also be used to study travel patterns and traffic demand in urban environments, which in turn could be used by policy makers to design more efficient transportation networks.

Preliminary success stories of connected vehicles include innovations like Uber, a taxi-hailing app which is estimated to have already captured about 46% of all paid car rides in the first quarter of 2015 [2]. It has an additional value exceeding hundreds of billions of dollars annually according to reports [16, 14] from McKinsey, Cisco and others by developing services, vehicles and transportation networks which can help consumers either avoid traffic congestion or make their travel more economical and environmentally friendly (e.g., ride sharing, adapting the engine, autonomous cars [28]).

However, before we can fully realize the potential of *connected vehicles*, significant spatial computing challenges need to be met. For instance, a sample connected vehicle dataset containing engine measurements of about 100 engine variables, once a minute, over the 100 million US vehicles in existence [26, 27], may have 10^{15} data-items per year [24]. These datasets, which we refer to as *vehicle measurement big data (VMBD)*, contain a collection of trips on a transportation graph such as a road map. Here, each trip is a time-series of attributes such as vehicle location, fuel consumption, vehicle speed, odometer values, engine speed in revolutions per minute (RPM), engine load, and emissions of criteria pollutants like smog-causing nitrogen oxides (NO_X) and GHGs like CO_2.

Computationally, VMBD and other connected vehicle data have spatio-temporal graph (STG) semantics [10, 8], where road intersections can be modeled as vertices and the road segments connecting adjacent intersections are represented as edges. Properties of edges may record engine sub-system measurements such as fuel-consumption, NO_X emissions and other vehicle measurements made while moving on that edge.

The rest of this paper is organized as follows. Section 2 details some spatial computing research problems associated with the vision of connected vehicles. Specifically, Subsection 2.1 presents novel queries on connected vehicle data (computationally modeled as a spatio-temporal graph) and their challenges. Subsection 2.2 describes the challenges of data science using these data. Data engineering challenges are presented in Subsection 2.3. Finally, we conclude our discussion in Section 3.

2. CHALLENGES AND OPPORTUNITIES

Realizing the full potential of connected vehicles raises numerous computational challenges for the state-of-the-art in spatio-temporal graph (hereafter referred to as "STG") query processing, data science and data engineering. Figure 1 summarizes the needs for further research in each of these areas and this section describes them in detail.

2.1 STG Query Processing

2.1.1 Novel Preference Functions for Routing Services

Traditional routing services identify a small set of routes based on limited route properties (e.g., travel-distance or travel-time) available in traditional digital road map datasets. By contrast, VMBD increases tremendously the number of possible preference functions beyond travel-distance and travel-time by allowing drivers to compare routes by their fuel consumption or GHG emissions. However, identifying eco-friendly routes requires addressing several research questions: "What is the computational structure of a fuel-efficient route or a minimum emission route?"; "Does it follow the assumptions of Dynamic Programming?"; "Can the routes be explored in a Greedy fashion?". These questions are challenging since measurements of fuel consumption or emissions, when measured over a long journey, cannot be broken down into corresponding values for single road segments, thus violating an assumption behind traditional shortest-path algorithms [8]. For instance, the fuel consumption or emissions of a vehicle driven by a commuter on a road segment, as derived from his/her annotated GPS traces, would depend on factors such as the initial velocity and acceleration before entering the road segment [8]. Another example illustrating this non-decomposability challenge can be observed in electric hybrid engine powertrain vehicles. Here, the vehicle stores energy via regenerative braking when coming to a stop sign or driving downhill. This energy can later be used for accelerating or climbing uphill leading to an uneven distribution of fuel consumption along the journey.

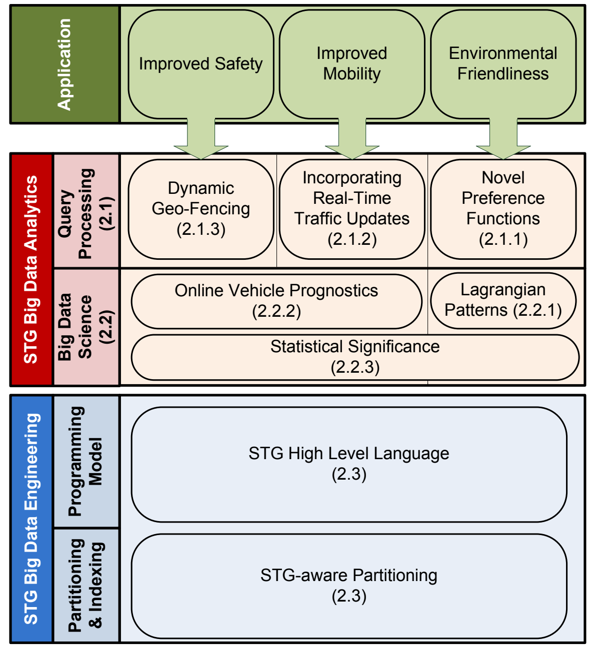

Figure 1: Research needs for realizing future connected vehicles (Best viewed in color)

2.1.2 Incorporating Real-Time Updates

Internet access and vehicle-to-vehicle communication technologies on modern connected vehicles can enable real-time traffic information to be incorporated into routing and navigation services. This can help vehicles avoid segments of idling and congestion. However, it also raises several questions about the ability of current routing algorithms to handle real-time updates. For instance, how much should the algorithm trust traffic updates from each data source (e.g.

speeding or slowing down vehicles, loop detectors, etc.)? Moreover, as new updates arrive, is the algorithm required to re-compute the entire shortest path solution or can it adjust the current available solution? One possible approach to avoid redundant computations would be to provide an upper bound on the maximum time for which each returned shortest path solution is still valid.

2.1.3 Dynamic Geo-Fencing

By allowing vehicles to communicate over dedicated channels, a vehicle can detect when a nearby vehicle has come too close and thus start coordinating its motion with the nearby vehicle to avoid a crash accident. This continuous monitoring of a vehicle's surrounding space is referred to as "dynamic geo-fencing" and is important to ensure driver and vehicle safety. To allow vehicles to collaboratively coordinate their operation to avoid geo-fence crossings, vehicles must be able to accurately identify their location. Location determination systems that rely on GPS signals might not be adequate in tunnels or other regions with weak satellite reception. In those areas, many systems would rely on "dead reckoning" methods to extrapolate the vehicle's current location from the last available reading for the vehicle's location and speed. However, that might not be sufficient in severe weather conditions (e.g. snow, hail, etc.) where vehicles could suddenly change their speeds and become more vulnerable to crashing. Another challenge for allowing motion coordination is the need for richer maps that reveal more detailed information about the different lanes (e.g. lane center-lines) as compared to traditional road maps.

2.2 STG Big Data Science

2.2.1 Lagrangian Pattern Mining on Road Networks

Traditional electronic control units (ECUs) calibrate the engine performance using a pre-selected rule-set that is chosen and fixed by the manufacturer (or after-market modifiers). These rule-sets are designed in laboratories using the engine and vehicle dynamometer experimentation and physics-based forward models [23, 13] to meet fuel efficiency and GHG regulations using a reference "drive-cycle" mimicking on-road driving conditions. By contrast, future connected vehicles will rely on smarter ECUs that allow the selection of appropriate rules based on historic performance and real-time readings from engine on-board sensors and other environmental variables (e.g. weather and real-time traffic conditions) to adapt and optimize the engine behavior for the current real-world conditions. Adaptive engine calibration rules can be developed by identifying relevant patterns in VMBD such as spatio-temporal graph hotspots of high engine emissions or low fuel efficiency. However, this is a challenging task since events of low fuel efficiency and/or high GHG emissions are also dependent on vehicle and driver specific factors such as the driving style or the current engine load. For example, while a road segment might represent a hotspot of high emissions at a given time for a certain driving style (e.g. aggressive acceleration/braking), it might not be the same for a driver moving along the same road segment with a different (e.g. gentler) acceleration/braking pattern. Current related work has only focused on linear hotspots of aggregated counts such as hotspots of road fatalities or crime incidents [22]. Hence, it cannot identify driver or vehicle-dependent patterns. Developing engine calibration rules will require analyzing the data from a traveler's perspective. That is to say, instead of a snapshot view, the data will need to be analyzed from a Lagrangian frame of reference [8] which captures the spatio-temporal perspective of a single vehicle as it travels over the road network. For instance, adaptive calibration rules would require identifying sub-trips on the road network in terms of space, time and engine behavior (i.e, variations in engine-power, gear transmission, brake-torque, etc.) that typically lead to events of high emissions or fuel consumption. In addition, several research questions need to be addressed for identifying such patterns. For instance: "What interest measures can be used to balance the statistical interpretation of the pattern and the scalability of the mining algorithm?"; "How can these rules be learned in a near real-time manner for adapting the engine to the vehicle's real-world conditions?"; "Can the search space for engine calibration rules be reduced by leveraging the physical constraints that govern engine behavior?"; "How would the sampling rate of the data and the duration of the physical phenomenon under investigation influence the choice of the appropriate learning method?".

2.2.2 Online Vehicle Prognostics

VMBD provides unique opportunities for predicting engine malfunctions and forecasting any required maintenance. This can reduce the down-time and effort on behalf of the vehicle's owner. Examples of vehicle prognostics include the ability to predict the occurrence of gasoline engine pre-ignition (i.e., knocking), an engine event that causes reduced engine fuel efficiency due to the sub-optimal placement of combustion [25]. Vehicle prognostics can also be used for the automatic detection of vehicle battery performance degradation and the need for a battery replacement. However, these malfunctions might not be well represented in the training data available (i.e. rare-class problem). Hence, predicting these rare events represents a major challenge for traditional spatio-temporal classification methods.

2.2.3 Ensuring Statistical Significance of Patterns

Due to the large volume of connected vehicle data, current spatio-temporal data science may lead to many spurious patterns that are difficult to interpret using the laws of physics [17]. Such false positives have a high societal cost since they can be used for deciding the engine operation modes at real-time scenarios and can also lead to a waste of time and resources allocated to engine maintenance and design optimization. To avoid this cost, STG big data science needs to ensure the statistical significance of the output patterns. This is a challenging problem since in many scenarios it requires rigorous statistical simulations (e.g. Monte Carlo simulation). Furthermore, one has to decide on a suitable null hypothesis for performing these simulations when the data is generated through a physical process that is governed by known laws. For example, this may require the use of sophisticated physical models for generating multi-dimensional physics-compliant datasets under the null hypothesis. Moreover, connected vehicle data exhibits a big-zero inflation problem in which routinely gathered data show a lot of non-events, that is, where nothing of importance or significance has happened. In popular statistical measures (e.g. correlation), large volumes of non-event data may lead to a near complete wash-out of significant but rare events (e.g. engine malfunctions or non-compliance with predictions from physical models) causing the low-count events to be missed.

2.3 STG Big Data Engineering

Cloud computing platforms are essential to scaling up STG data analytics for handling the huge volume of connected vehicle data. However, STGs used to model connected vehicle data violate the assumption underlying current spatial cloud computing environments (e.g., ESRI GIS Tools for Hadoop [6], SpatialHadoop [5], HadoopGIS [1] that data is embedded in a geometric space (e.g., Euclidean or Spherical). This leads to inconvenience in representing STG operations using the current spatial cloud computing programming models (e.g., SpatialHadoop Pigeon [4], Pregel [15], GraphLab [12]). Furthermore, it reduces the effectiveness of data-partitioning methods (e.g, R-tree) due to the reduced load-balancing. Hence, an STG-aware computational infrastructure is needed to simplify the representation and improve the computational scalability of STG data analytics. This infrastructure should support STG data analytics on both data partitioning and programming model levels. Data partitioning methods need to account for the spatial, temporal and connectivity aspects of STG data to improve load balancing and minimize I/O operations. For instance, an STG-aware data partitioning method may group together nodes that are expected to be visited consecutively on a shortest path query. Novel programming models are also needed to facilitate the representation of STG abstract data types, predicates and operations.

3. CONCLUSION

In this paper, we introduced a vision of future connected vehicles equipped with rich instrumentation and Internet access enabling them to communicate with nearby vehicles and the transportation network infrastructure. Connected vehicles have the potential to improve almost every aspect of roadway travel. They can minimize vehicle crashes, help drivers avoid congestion and traffic jams based on real-time traffic conditions, and provide greener navigation choices. This vision cannot be fully realized without significant computational advances in spatio-temporal graph query processing, data science and data engineering. We encourage the spatial computing community to take on these challenges and research needs to help ensure the vision of connected vehicles becomes a reality.

4. ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1029711, IIS-1320580, 0940818, and IIS-1218168, the USDOD under Grant No. HM0210-13-1-0005, and the University of Minnesota under the OVPR U-Spatial. We are particularly grateful to Kim Koffolt for her help in editing this paper.

5. REFERENCES

[1] A. Aji, F. Wang, H. Vo, R. Lee, Q. Liu, X. Zhang, and J. Saltz. Hadoop gis: a high performance spatial data warehousing system over mapreduce. *Proceedings of the VLDB Endowment*, 6(11):1009–1020, 2013.

[2] A. Bender. Uber's astounding rise: Overtaking taxis in key markets. Forbes `http://goo.gl/PFnVyB`, April 2015.

[3] U. Congress. Energy independence and security act of 2007. `http://en.wikipedia.org/wiki/Energy_Independence_and_Security_Act_of_2007`, Public Law, (110-140), 2007.

[4] A. Eldawy and M. F. Mokbel. Pigeon: A spatial mapreduce language. In *Data Engineering (ICDE), 2014 IEEE 30th International Conference on*, pages 1242–1245. IEEE, 2014.

[5] A. Eldawy and M. F. Mokbel. Spatialhadoop: A mapreduce framework for spatial data. In *Proceedings of the IEEE Intl Conference on Data Eng.*, 2015.

[6] ESRI GIS Tools for Hadoop: Big Data Spatial Analytics for the Hadoop Framework. `http://esri.github.io/gis-tools-for-hadoop/`.

[7] T. Geller. Car talk. *Communications of the ACM*, 58(3):16–18, Mar 2015.

[8] V. Gunturi and S. Shekhar. Lagrangian xgraphs: A logical data-model for spatio-temporal network data: A summary. In *Advances in Conceptual Modeling*, volume 8823 of *LNCS*, pages 201–211. Springer, 2014.

[9] J. Harding and e. al. Vehicle-to-vehicle communications: Readiness of v2v technology for application. Technical Report Report No. DOT HS 812 014, National Highway Traffic Safety Administration, Washington, DC, August 2014.

[10] E. G. Hoel, W.-L. Heng, and D. Honeycutt. High performance multimodal networks. In *Advances in Spatial and Temporal Databases*, pages 308–327, 2005. Springer. LNCS 3633.

[11] H. Kargupta, V. Puttagunta, M. Klein, and K. Sarkar. On-board vehicle data stream monitoring using minefleet and fast resource constrained monitoring of correlation matrices. *New Generation Computing*, 25(1):5–32, 2006. Springer.

[12] Y. Low, J. E. Gonzalez, A. Kyrola, D. Bickson, C. E. Guestrin, and J. Hellerstein. Graphlab: A new framework for parallel machine learning. *arXiv preprint arXiv:1408.2041*, 2014.

[13] B. Lumpp, M. Tanimou, M. McMackin, E. Bouillon, E. Trapel, M. Muenzenmay, and K. Zimmermann. Desktop simulation and calibration of diesel engine ecu software using software-in-the-loop methodology. Technical report, SAE Technical Paper, 2014.

[14] A. Mai and D. Schlesinger. A business case for connecting vehicles. Cisco `http://goo.gl/u4ngwF`, April 2011.

[15] G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. Pregel: a system for large-scale graph processing. In *Proceedings of the ACM Intl Conf. on Management of data*, pages 135–146, 2010.

[16] J. Manyika et al. Unlocking the potential of the internet of things. McKinsey Global Institute `http://goo.gl/qzq5mV`, June 2015.

[17] G. Marcus and E. Davis. Eight (no, nine!) problems with big data. *The New York Times*, 2014.

[18] P. Marshall. Can transportation agencies call on smartphones for traffic data? Government Computer News `http://goo.gl/fd8zO7`, Feb 2014.

[19] P. Marshall. Connected cars: Apps, networks and storage on wheels. Government Computer News `http://goo.gl/MNc5QE`, Feb 2014.

[20] H. Nanjundaswamy and e. al. Development and calibration of on-board-diagnostic strategies using a micro-hil approach. Technical Report 2011-01-0703, SAE Technical Paper, 2011.

[21] U. D. of Transportation. Us dept of transportation research, development, and technology strategic plan fiscal year 2013–2018. `http://goo.gl/wzVu8P`, 2013.

[22] D. Oliver, S. Shekhar, X. Zhou, E. Eftelioglu, M. R. Evans, Q. Zhuang, J. M. Kang, R. Laubscher, and C. Farah. Significant route discovery: A summary of results. In *Geographic Information Science*, pages 284–300. Springer, 2014.

[23] A. Palladino, G. Fiengo, and D. Lanzo. A portable hardware-in-the-loop (hil) device for automotive diagnostic control systems. *ISA Transactions*, 51(1):229–236, 2012.

[24] J. I. Speed. Iot for v2v and the connected car. `http://goo.gl/3b5NSy`, 2014.

[25] J. M. Spelina, J. C. P. Jones, and J. Frey. Recent advances in knock analysis, simulation, and control. *SAE International Journal of Engines*, 7(2014-01-1349):947–955, 2014.

[26] D. Sperling and D. Gordon. *Two billion cars*. Oxford University Press, 2009.

[27] H. Statistics. Federal highway administration. *HM-63, HM-64*, 2008.

[28] B. Zhang. Autonomous cars could save the us $1.3 trillion dollars a year. Business Insider `http://goo.gl/8i7KQD`, 2014.