The Next Generation of Transportation Systems, Greenhouse Emissions, and Data Mining

Hillol Kargupta
University of Maryland
Baltimore County
and Agnik, LLC
Baltimore, MD 21250, USA
hillol@cs.umbc.edu

João Gama
University of Porto
R. de Ceuta 118-6, 4050
Porto, Portugal
jgama@fep.up.pt

Wei Fan
IBM T.J.Watson Research
19 Skyline Drive
Hawthorne, NY 10532
wei.fan@us.ibm.com

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database applications—*data mining*; I.2.6 [**Artificial Intelligence**]: Learning—*classifiers design and evaluation*

General Terms

Algorithms, Design, Economics

Keywords

Transportation Systems, Data Mining

1. INTRODUCTION

Controling Greenhouse gas (GHG) emissions for minimizing the impact on the environment is one of the major challenges in front of the human civilization. Although future concentrations, damages and costs are unknown, it is widely recognized that major emissions reduction efforts are needed. In 1997, the Kyoto Protocol promoted by the United Nations Framework Convention on Climate Change, aimed at fighting global warming. The main goal is *"stabilization of greenhouse gas concentrations in the atmosphere at a level that would prevent dangerous anthropogenic interference with the climate system"* [9]. According to the International Energy Agency [1], energy efficient in buildings, industrial processes and transportation could reduce the world's energy needs in 2050 by one third, and help controlling global emissions of greenhouse gases. The report [1] describes a series of scenarios showing how key energy technologies can reduce emissions of carbon dioxide, the greenhouse gas which is most responsible for climate change.

Of the four primary GHG under scrutiny, carbon dioxide (CO2), and the need to lower carbon emissions in general, is of paramount concern. It is estimated that transportation activities are responsible for approximately 25% to 30% of total U.S. GHG emissions, with the on-highway commercial truck market accounting for over 45% of transportation GHG. However, the transportation sector emissions remain almost entirely unaddressed with respect to GHG and CO2 reduction.

The Intergovernmental Panel on Climate Change (IPCC) provided guidelines for calculating carbon emission offer estimations only for certain common types of fuels; even the estimates are not available novel fuel blends and gaseous fuels such as CNG and LNG. Indeed, these and other references have documented the uncertainty in model-based theoretical carbon emissions calculations and the need for a standardized, consistent method of accurately characterizing CO2 emissions. Moreover, correlating various vehicle performance and traffic parameters may open up new insights resulting in better techniques for controlling emissions. For example, it is widely known that vehicle speed, engine load and state of repair/maintenance play important roles in governing emissions. Mining the emissions data along with the traffic patterns in a metropolitan area, vehicle performance (load, rpm, and vehicle oxygen sensor characteristics) and the driving behavior may provide useful information to design speed limits, traffic signals and fleet maintenance policies. Such advanced analysis of emission data will be possible only when we can directly and accurately measure emissions in the vehicle.

The 2010 ACM SIGKDD panel on "The Next Generation of Transportation Systems, Greenhouse Emissions, and Data Mining" explores this area. It identifies the main areas where data mining researchers and practitioners can contribute in order to reduce the greenhouse emissions from transportation systems.

2. THE ROLE OF DATA MINING IN TRANSPORTATION SYSTEMS

Data mining is likely to play a key role in reducing the amount of GHG emissions and understanding its impact on the environment. Innovative solutions that could help in meeting the Kyoto goal of replacing 30% of fossil fuel usage by 2020 could also bring economic benefits. The broader agenda for the data miners include:

- Reducing greenhouse gases through technological and socio-economic means
- Development of clean and efficient engines and powertrains including hybrid technologies
- Use of alternative fuels for transport applications, in particular hydrogen and fuel cells
- Taking account of cost-efficiency and energy-efficiency considerations
- Developing end-of-life strategies for vehicles and vessels.

KDD'10, July 25–28, 2010, Washington, DC, USA.
ACM 978-1-4503-0055-1/10/07.

2.1 Growing Activity

A growing amount of literature and commercial applications are emerging to address the challenges of the transportation systems and emissions reductions. Several papers [8, 3, 7] dealing with the vehicle health data mining problems are emerging.

The US *National Science Foundation* and *European Research Council* are actively promoting research on *"the sustainable management of both the man-made and the natural environment and its resources."*. To this end, increased knowledge on the interaction between the climate, the biosphere, ecosystems and human activities is sought and new environmentally-friendly technologies, tools and services need to be developed. NSF recently supported the Next Generation Data Mining Summit[1] that explored many of the relevant issues.

For example, the **Green Cars** is an initiative involving research on a broad range of technologies and smart energy infrastructures essential to achieve a breakthrough in the use of renewable and non-polluting energy sources, safety and traffic fluidity. The project MERGE (Mobile Energy Resources for Grids of Electricity) will prepare the European electrical system for the massification of electric vehicles. The aim is to find solutions that will minimize the need to strengthen to electric grid infrastructures and power systems, and thus avoid extra costs that would have to be borne by the users of the electric vehicles. This is the largest research project of the 7^{th} EU Framework Programme.

Advanced data stream mining systems such as InfoSphere Streams from IBM [5] are finding applications in this domain [4]. Wireless emissions monitoring standards for cars are emerging. NASA is developing advanced Integratd Vehicle Health Monitoring (IVHM) program for supporting research and other resources [2].

Commercial systems are also being developed. Examples include the MineFleet system [6] from Agnik, that recently received the 2010 Frost & Sullivan North American Enabling Technology of the Year Award in the area of commercial telematics. MineFleet is a decision support system that offers a powerful onboard data stream mining software for modeling, benchmarking, and monitoring of vehicle health, emissions, driver behavior, fuel-consumption, and fleet characteristics. It is an autonomous onboard distributed data stream mining platform. It continuously monitor data generated from vehicle sensors, analyzes the data streams (manufacturer enhanced parameters, generic parameters, and fault-codes), and generates the predictive models, reports, and alerts.

2.2 The Panel

Understanding the effects of greenhouse emissions requires advanced data analysis techniques for fully exploitation of remotely sensed data. Reducing the carbon footprints of buildings, vehicles, and airplanes would require continuous monitoring of sensors and detecting deviation from desired behavior.

Designing the next generation of transportation network becomes particularly challenging in the context of increasing demand for energy supplies and reducing greenhouse emissions.

[1] www.kd2u.org/NGDM09/

[2] https://c3.ndc.nasa.gov/dl/

A list of research questions to discuss in the experts panel include:

1. What are the key information processing challenges in the next generation of intelligent transportation systems?
2. What are the challenges in making transportation "green"-er and how data mining can help?
3. Why predictive vehicle health monitoring is important and why should data miners care?
4. What are the emerging business models for Green IT that data miners can benefit from?
5. How can data miners help vehicle manufacturers in building better and cleaner vehicles?
6. How can data miners help maintaining and monitoring vehicles after market?
7. What is the current status of the technology and what are the achievable return of investments for this market?
8. What are the projections for the next five years and what can data miners do to help?
9. What are the challenges against large-scale adoption of data mining-based decision support tools for clean vehicles and transportation systems?
10. How can policy makers and funding organizations help?

Sensor networks for highways and vehicles equipped with diagnostic data buses along with the availability of machine-to-machine wireless communication networks are going to make the role of advanced data mining techniques very important in the transportation industry. Computing in itself is under scrutiny from the perspective of its effect on greenhouse emissions and pollution. Data mining technical questions that people would care to think about:

- How do we address multiple and heterogeneous data sources?
- How do we handle data that are not structured or semi-structured?
- How do we construct feature from raw data given the fact that much data is not in feature vector format that we are familiar of.
- How do we handle large amount of streaming data?
- Is parametric model or non-parametric model more suitable for these problems?
- How to handle large amount of missing data?
- What are the limits of data mining approaches to solve these problems?
- What problems that data mining cannot solve?

3. SYSTEM COMPONENT AND DIFFERENT ROLE FROM PHYSICAL SCIENCE

In order to address various challenges and make real-world impact in deployment, data mining models need to work seamlessly with systems, such as data stream systems (Infosphere Streams [5], etc), in order to receive real-time data for training, model tuning as well asl real-time scoring and prediction. In the same time, it is important to understand the different roles between data mining and traditional physical science that are more hypothesis-testing and simulation-based.

4. CONCLUSIONS

The panel will discuss the technologies of future, that might be available in the next 5 years. The main issue of next generation transportation systems are the environmental impacts. Data mining might play a relevant role in efficient adaptive cars and infrastructures by optimizing resources in real-time and at any-time. Overall, we need to develop and explore technology for a sustainable world.

Acknowledgments

João Gama acknowledges the financial support given by the project *Knowledge Discovery from Ubiquitous Data Streams* (PTDC/EIA/098355/2008). Hillol Kargupta would like to thank Agnik and UMBC for supporting this work. He would also like to acknowledge NASA NRA NNX07AV70G for supporting his work.

5. REFERENCES

[1] Energy Business Council. Invest in clean technology. Technical report, International Energy Agency, 2005.

[2] Industrial technologies Directorate-General for Research. New public-private partnerships for research in the manufacturing, construction and automotive sectors. Technical report, European Commision, 2010.

[3] Hillol Kargupta, Vasundhara Puttagunta, Martin Klein, and Kakali Sarkar. On-board vehicle data stream monitoring using minefleet and fast resource constrained monitoring of correlation matrices. *New Generation Computing*, 25(1):5–32, 2007.

[4] Report on the Next Generation of Data Mining for Dealing with Energy, Greenhouse Emissions, and Transportation Challenges. Chandra Bhat, Auroop Ganguly, Johannes Gehrke, Chris Giannella, Mark McGranaghan, Paul Melby, Tom Dietterich, Carla Gomes, Hillol Kargupta, Vipin Kumar, Ashok Srivastava, Philip Yu. October 1-3, 2009.

[5] InfoSphere Streams: *http://www-01.ibm.com/software/data/infosphere/streams.*

[6] Hillol Kargupta, Kakali Sarkar, and Michael Gilligan. MineFleet: An Overview of a Widely Adopted Distributed Vehicle Performance Data Mining System. *Proceedings of the ACM SIGKDD Conference*, 2010.

[7] S. Krishnaswamy, S. W. Loke, A. Rakotonirainy, O. Horovitz, M. M. and Gaber. Towards Situation-awareness and Ubiquitous Data Mining for Road Safety: Rationale and Architecture for a Compelling Application. *Proceedings of Conference on Intelligent Vehicles and Road Infrastructure (IVRIŠ05)*, University of Melbourne, 16-17 February 2005.

[8] A. N. Srivastava, W. Buntine. Predicting Engine Parameters using the Optical Spectrum. Proceedings of the AIAA Electrochemical Conference.

[9] United Nations Framework Convention on Climate Change. Kyoto protocol: Status of ratification. Technical report, United Nations, 2009.