DSCS 6030 - Intro to Data Mining and Machine Learning

Predicting Air Pollutant Emission

Your Name
Date

a. A clear description of the topic.

Emission of air pollutants has been widely discussed regarding environmental concerns. Politically, it has also sparked many new standards and regulations that are currently or will soon be put into place. More recently, the deception of Volkswagen diesel emissions has been big news. Electric cars are also becoming more popular to help combat this issue.

Emissions rates for consumers can be based on their transportation method and the number of vehicles on the road. Consumer expenditure surveys provide data on whether people drive or use public transportation and transportation costs. The emissions measurements are provided by the US EPA. The basis of my project is to use mapping techniques to detect areas with high vehicle ownership (from consumer expenditure data) and predict emissions rates.

b. Background research of related work.

Global emissions, especially carbon dioxide, are predicted to increase 43% by 2035. (Kirkland, 2010) For the US, emissions increased by 2.0 percent from 2012 to 2013. (US EPA, 2014) The trends result from national energy data, agricultural activities and miles traveled by vehicles. The measurements are provided for each state. However, predictions are based on year to year tracking rather than predictions from consumer expenditures.

There is a couple of government agencies that track emissions rate and try to predict emissions based on various factors.

IPCC: http://www.ipcc-data.org/observ/ddc_co2.html

US Trends: http://www3.epa.gov/climatechange/ghgemissions/usinventoryreport.html

How carbon footprint is calculated: https://www.carbonfund.org/how-we-calculate

c. Data sources?

I will use consumer expenditure reports that report household expenditures, including transportation data. Emissions data can be viewed from the US EPA.

Consumer Expenditure Survey: http://www.bls.gov/cex/pumdhome.htm

US EPA Emission Inventories: http://www3.epa.gov/ttn/chief/eiinformation.html

Voluntary Reporting of Greenhouse Gases Program: http://www.eia.gov/oiaf/1605/coefficients.html

Boston data of Municipal energy: https://data.cityofboston.gov/dataset/Municipal-Energy-Data/bcnb-bux2 (narrower scope?)

Data from 2011 (last available data from both):

Vehicle Ownership: http://www.bls.gov/cex/pumd/data/comma/intrvw11.zip

Emissions: http://www.epa.gov/ttn/chief/net/2011inventory.html

d. What algorithms are being used and code sources.

For mapping emission rates, neural networks and the genetic algorithm are being used in multiple studies. I will try to mimic using neural networks to predict results using my gathered data sources.

Neural Networks:

Wiki: https://en.wikipedia.org/wiki/Artificial_neural_network

NeuralNetTools R Package: https://cran.r-project.org/web/packages/NeuralNetTools/index.html

Example in R: https://beckmw.wordpress.com/tag/neural-network/

Neuralnet R Package: https://cran.r-project.org/web/packages/neuralnet/index.html

Example in R: http://www.r-bloggers.com/using-neural-networks-for-credit-scoring-a-simple-example/

Example in R: http://bicorner.com/2015/05/13/neural-networks-using-r/

Research papers on Neural Networks:

Neural Networks and genetic algorithm: http://www.ncbi.nlm.nih.gov/pubmed/23220141

"This paper describes the development of an artificial neural network (ANN) model for the forecasting of annual PM(10) emissions at the national level, using widely available sustainability and economical/industrial parameters as inputs. The inputs for the model were selected and optimized using a genetic algorithm and the ANN was trained using the following variables: gross domestic product, gross inland energy consumption, incineration of wood, motorization rate, production of paper and paperboard, sawn wood production, production of refined copper, production of aluminum, production of pig iron and production of crude steel."

Artificial Neural Network: http://www.scirp.org/journal/PaperDownload.aspx?paperID=33792

"In this paper we are providing a solution to forecast the poison CO2 gas emerged from energy consumption. Four inputs data were considered the global oil, natural gas, coal, and primary energy consumption to build our system. In this paper, we used the Artificial Neural Network (ANN) as successful and powerful tool in handling a time series modeling problem."

e. How are you going to evaluate the success of your project?

To evaluate the successful usage of neural networks and R packages, I will first apply the algorithms to predict the emissions rates from 2011 to see if the predicted emission rates from consumer surveys follow the trends as mentioned in the background research.

Since the latest data from both emissions reports and consumer survey is from 2011, I can apply the algorithm to the consumer survey from 2014 to predict the emission rates for 2014.

f. References.

Antanasijević DZ1, Pocajt VV, Povrenović DS, Ristić MĐ, Perić-Grujić AA. "PM(10) emission forecasting using artificial neural networks and genetic algorithm input variable optimization." 2013 Jan 15;443:511-9. doi: 10.1016/j.scitotenv.2012.10.110. Epub 2012 Dec 4.

Baareh, Abdel Karim. "Solving the Carbon Dioxide Emission Estimation Problem: An Artificial Neural Network Model." Journal of Software Engineering and Applications, 2013, 6, 338-342 http://dx.doi.org/10.4236/jsea.2013.67042 Published Online July 2013 (http://www.scirp.org/journal/jsea)

F. Günther and S. Fritsch. "Neuralnet: Training of Neural Networks". 2010. The R Journal Vol.2/1, June 2010 (http://journal.r-project.org/archive/2010-1/RJournal_2010-1_Guenther+Fritsch.pdf)

Kirkland, Joel. "Global Emissions Predicted to Grow through 2035" 2010. Website: http://www.scientificamerican.com/article/global-emissions-predicted-to-grow/ Accessed: Oct 23, 2015

US EPA "Trends in Greenhouse Gas Emissions" Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2012. US GHG Inventory 2014 Chapter 2 Trends. (http://www3.epa.gov/climatechange/Downloads/ghgemissions/US-GHG-Inventory-2014-Chapter-2-Trends.pdf)