Big Data Drive: Supporting Product Analytics at Ford Motor through the employment of Big Data technologies

David A. Ostrowski
Global Data Insights and Analytics
Agenda

• Introduction
• Projects
  • Fuel Economy Analysis
  • HMI
  • Battery Charge Limits
  • Geographic Analysis
• Conclusion
Introduction : What is BDD?

• The Big Data Drive (BDD) is a special research project in which Ford Motor employees volunteer to have their vehicle data collected via the OpenXC interface and stored on the Ford HPC Hadoop cluster for analysis.

• Initiated in 2014

• ~1200 Vehicles data available and growing
What is BDD? : Background

• Most participants are Ford Motor Employees from metro Detroit with some drivers from Mexico City, and Ford of Europe

• Recent expansions include humanitarian Aid Programs in India, Gambia, Nigeria and South Africa as well as a vendor supported expansion through our dealer networks running in Tampa, Phoenix and Dallas

• The data flow is summarized by the cartoon below
What is OpenXC?

• An open source project allowing consumer devices access to data from any vehicle.

• More information at http://openxcplatform.com/

• This exists as the foundation for Big Data Drive where the data is shipped to the cloud, finally residing in our HPC (Big Data) Cluster.
What is BDD? : Background

BDD provides data from the OpenXC library including:

- Accelerator Pedal Position
- Torque at Transmission
- Transmission Gear Position
- Engine RPM
- Vehicle Speed
- Brake Pedal Status
- Fuel Consumed Since Start
- Steering Wheel Angle
- Odometer
- Ignition Status
- Fuel Level
- Door Status
- Headlamp Status
- Latitude
- Longitude

- BDD Data includes all available CAN data which has been integrated with the standard OpenXC.
- BDD is also merged with Geographical information allowing for characterizations of slope and elevation while driving.
Data Collection (details)

- A Controller Area Network (CAN) interface is installed in each participating vehicle
- Vehicle data is streamed to the phone via Bluetooth
- Data is collected on the phone using the Ford BDD app.
- The BDD app transmits data to Ford servers every day through the participant’s home broadband Internet
- Data is then available for use by Ford Engineers
BDD : Advantages

• Motivated by it’s roots as a mobility project, BDD has become valuable as a testing tool due to the following reasons:
  • Unbiased Data Collection
  • On-Road Conditions
  • Peer groups (Big Data)
  • Economical
Advanced Powertrain – Data Mining analysis

• Experimentation of Data Mining techniques to assist in the engineering process (Spark ML package)

• Initial experiments of applying Decision Tree learning to identify secondary and tertiary effects of CAN variables on dependent variables (per mile mpg)

• Future consideration to other unsupervised methods including clustering as well as dimensionality reduction leveraging principle components analysis
Analysis Design:
Considering current parameters of interest for powertrain study of per mile fuel economy for 1.6L Escape (CAN) Parameters were evaluated over an entire data set of per-mile fuel economy generated over 1839 miles

Highlighted parameters were identified through decision tree analysis to support the highest information gain in a binary classification between low and high fuel economy (threshold of 20 mpg)

\[
\begin{align*}
\text{AirAmb}\_P\_\text{Actl}, \\
\text{BSBattCurrent}, \\
\text{ACCompressorDisp}\_\text{B1}, \\
\text{ACCompressorDisp}\_\text{B2}, \\
\text{ACCompressorStatusEVDC}\_\text{B1}, \\
\text{ACCompressorStatusEVDC}\_\text{B2}, \\
\text{EvapTemp}, \\
\text{EvapTemp}\_\text{UB}, \\
\text{EvapTempSetPt}, \\
\text{EngClnt}\_\text{Te}\_\text{Actl}\_\text{B1}, \\
\text{EngClnt}\_\text{Te}\_\text{Actl}\_\text{B2}, \\
\text{GboxOil}\_\text{Te}\_\text{Actl}, \\
\text{AirAmb}\_\text{Te}\_\text{Actl}\_\text{B1}, \\
\text{AirAmb}\_\text{Te}\_\text{Actl}\_\text{B2}, \\
\text{GPS}\_\text{Speed}, \\
\text{GPS}\_\text{Vdop},
\end{align*}
\]

\[
\begin{align*}
\text{EngAout}\_\text{N}\_\text{Actl}, \\
\text{GPS}\_\text{MSL}\_\text{altitude}, \\
\text{WhlFl}\_\text{W}\_\text{Meas}, \\
\text{WhlFl}\_\text{W}\_\text{Meas}\_\text{UB}, \\
\text{WhlFr}\_\text{W}\_\text{Meas}, \\
\text{WhlFr}\_\text{W}\_\text{Meas}\_\text{UB}, \\
\text{WhlRl}\_\text{W}\_\text{Meas}, \\
\text{WhlRl}\_\text{W}\_\text{Meas}\_\text{UB}, \\
\text{WhlRr}\_\text{W}\_\text{Meas}, \\
\text{WhlRr}\_\text{W}\_\text{Meas}\_\text{UB}, \\
\text{BpedDrvAppl}\_\text{D}\_\text{Actl}, \\
\text{GearPos}\_\text{D}\_\text{Actl}, \\
\text{AirAmb}\_\text{P}\_\text{Actl}, \\
\text{EngAirIn}\_\text{Te}\_\text{Actl}, \\
\text{AirAmb}\_\text{Te}\_\text{ActlFilt}\_\text{B2}
\end{align*}
\]
Considering per-mile mpg in a binary classification problem, out of 30 CAN messages, AirAmbient temp and EngAout were identified as the most significant variables.
Powertrain Calibration – Next Steps

• Determine a means of normalizing the vehicle parameters in order to study secondary and tertiary effects of the relevant CAN messages

• Consider classifiers for predictive behavior

• Tie in both efforts to apply Machine Learning at larger scales
• Application of analytics to BDD 1.0 for the purpose of studying HMI

• Initial questions are:
  How are heated and cooled seats used?
  How are the seat memory controls used?
  Do we need to simplify/ minimize / adjust controls?

• Due to consistently higher content and accessibility to relevant CAN channels the Lincoln MKS was investigated

• Following are samples of the output at vehicle level - looking at broadcast messages acting as a proxy for time.
HMI Analytics (climate seat controls per day, over time)

- Usage in a day
  - Seat Off
  - Seat Heat
  - Seat Cool

Broadcast message Count (proxy for time)

Ambient temp

Days

Wed, 19 Feb 2014

Sat, 13 Dec 2014
HMI Analytics (ambient temp, cooled/heated seat usage)

Cooled seat usage matching temperature peaks

Usage on warm days suggests “therapeutic usage”

Highest heated seat usage matching the temperature valleys
HMI Analytics – Lincoln MKS Seat Control Usage

Used rarely over course Of 229 days

More significant usage: Indicating multiple drivers
HMI Controls: Intermediate Results

• Cooled or heated seats are not used throughout the entire day (ever)
• Cooled seats are used sparingly
  • Too cold?
  • Why aren’t they used on every hot day?
• Some results indicate heated seat usage that is not perfectly correlating to climate
  • Therapeutic usage?
• Seat memory adjustments are used sparingly
• Secondary, tertiary controls (for memory seats) are used rarely
HMI Analytics: Investigation in predictive behaviors

Investigated application of Machine learning methods to support the classification of behaviors of HMI usage

We relied on the Spark ML package, using the Decision Tree library

Feature 0 day:
Feature 1 month
Feature 2 AmbTemp
Feature 3 Day_Night_Status
Feature 4 InCarTemp
Feature 5 EngAout_N_Actl (engine speed)

Example characterization leveraging Heated seat usage as the dependent variable:

If ambTemp < 10 deg:
  if engineSpeed > 10 rpm:
    if winterMonth (sept – april)
      if inCarTemp < 10:
        if vehicleSpeed > 10:
          if (between midnight to noon):
            set the heatedSeats
HMI Analytics: next steps

- Collecting more data for purpose of supporting a more significant sample set

- Investigating utilization of the Lincoln Continental

- Continuing to develop means of being able to significantly characterize the data collection to enable optimal HMI design
Electrical Powertrain System Controls

Gen III Traction Battery Charge Power Limit Increase Benefit Analysis:

• Evaluate the possibility of improving EV range

• Investigate possible benefits from a charge power limit increase.

• Increasing the charge power limit would allow for more energy recovery during braking which would contribute to improving EV range during charge depletion
Electrical Powertrain System Controls

- Calculate braking energy for various power ranges above the current 35kW limit (ie. 35-40, 40-45 kW, ...etc)

- Estimate the potential improvement in EV range given the above braking energy for each power range while considering each driver’s consumption (watt-hr/mile)

- Evaluate statistical relevance for the population of PHEV drivers
• Leveraging the BDD data, the analysis demonstrated improvements in range through increasing specific power charge limits.

• The evaluation also showed that improvements may be even higher for those with lower brake scores (conversely lower for those with high brake scores)
Geographic Analyses - Parking insights

Short summary:
In this experiment we used GPS signal and timestamp to identify parking events in the RIC parking lot. We used this analysis to understand driver parking behaviors.
Geographic Analyses - Parking insights

Parking Density by hour: 3 pm
Geographic Analyses – Parking maneuvers

Parameters:
- Vehicle speed
- Gear position
- GPS signal

EXAMPLE:
- Identify perpendicular parking events
**Geographic Analyses – Parking maneuvers**

**Parameters:**
- Vehicle speed
- Gear position
- GPS signal

**EXAMPLE:**
Identify parallel parking events
Geographic Analyses – Speed insights

Parameters:
Vehicle speed
GPS signal

EXAMPLE 1:
Identify speed designations (Highways vs Local Roads)

EXAMPLE 2:
Identify traffic stops events
Geographic Analyses – Speed insights

Parameters:
- Vehicle speed
- GPS signal

EXAMPLE:
Charting differences in speed across different modes highlights stark contrast in 4whl vs 2whl effectiveness.

Motorcycle agility in dense urban area (Banjul)

Ranger agility in rural, highway-oriented journeys
Future Direction (Environmental Parameters)

Parameters:
- Altitude (m)
- Grade (%)
- GPS signal

EXAMPLE:
Calculate grade values per GPS point and join to the BDD data to analyze fuel consumption trends over a trip.
Future Direction (Environmental Parameters)

Parameters:
- Altitude (m)
- Grade (%)
- GPS signal

EXAMPLE:
Vehicle travelling on I94 W along the bridge over the Rouge River (Allen Park)
Overall Conclusion

• Work to institutionalize analysis, streamline pipeline and assist product development in the internalization and analysis of data.

• Further leverage the parallel environment in order to prepare for increasingly larger data sets.

• This work supports a step into further integration of on-road vehicle data to the product development process.
THANK YOU