MEASURING FREEWAY SAFETY PERFORMANCE:

THE FREEWAY ACCIDENT RISK ANALYSIS (ARA) TOOL

James Marca
Institute of Transportation Studies
University of California Irvine

Research Objective

Map traffic flow conditions to levels of accident risk

For urban freeway mainlines with 3+ directional lanes

Traffic Conditions

- Use VDS loop detector data
  - Loops are ubiquitous
  - 30-sec raw data captures dynamic changes
  - Recent data for each district readily available to TMC staff
  - Historical raw data for most Caltrans Districts available from PeMS data archive

Accident Risk

- Measured by Police-investigated mainline accidents
- Accidents classified by type, e.g.:
  - Severity: injury or property damage only
  - Number of involved vehicles: 1, 2, or 3+
  - Collision location: out of lanes, left, interior, or right lanes

Continuation of PATH 5307

- PATH 5307: Implementation of a Tool for Measuring ITS Impacts of Freeway Safety Performance
  - Developed the overall methodology
- This project:
  - Update models with new data
  - Deploy the new models for Caltrans

Application

- Freeway Accident Risk Analysis (ARA) Tool
- Operations:
  - Suitable for long-term monitoring
  - Enable before and after studies to quantify safety improvements
  - Measure the accident risk impacts of traffic flow changes
- Planning:
  - Identify areas that would benefit from improvements
  - Open access to intermediate data and model predictions

Methodology

- Combine accident and traffic flow data to capture the traffic conditions at time and place of an accident
- Draw a comparable sample of non-accident traffic conditions
- Use probabilistic models to determine what traffic conditions are most likely to be associated with different types of accidents

Input Data

- Traffic flow
  - 30-sec VDS loop data downloaded from PeMS
  - All of 2007 for modelling, 2008 through 2010 for analysis
- Accidents
  - 2007 (modelling) and 2008 (validation) accident data from Federal Highway Safety Information System (HSIS) multistate data repository
  - Original source was the Caltrans accident database

Traffic Flow Variables

- Need 20 minutes of detector data for 3 lanes
  - Number 1 (left) lane
  - One interior lane
  - Right (curb) lane
- Using only Volume and Occupancy
  - Density, consequently speed, is not determinable (Average vehicle lengths unknown in any 30 sec interval)
- Fault tolerance
  - Need 30 out of 40 observations

27 Traffic Flow Variables

- Flow level: 3 mean volumes, by lane
- Turbulence:
  - 3 standard deviations of volume, by lane
  - 3 coefficients of variation of occupancy, by lane
  - 3 coefficients of variation of volume/occupancy, by lane
- Synchronization:
  - 3 correlations of volume, by lane pair
  - 3 correlations of occupancy, by lane pair
  - 3 correlations of volume/occupancy, by lane pair
- Queuing and other systematic changes:
  - 3 autocorrelations of volume, by lane
  - 3 autocorrelations of occupancy, by lane

Data Storage

- All raw loop data from PeMS for 2007 through 2010 evaluated
- 30s raw data ⇔ 27 variables
- Factor of 5 increase in size
- Relational database (PostgreSQL) had poor query performance

Data Storage

- 27 variables (plus raw) stored in CouchDB
- One database per detector
- One document per day
- Unique document key sorts by timestamp
- Stored *views* in each database produce summaries, model output, etc

Map-Reduce Views

- "Map" step:
  - Per-detector views process 30-second data
- "Reduce" step:
  - View output is aggregated to hourly totals
- Aggregation:
  - Detector database output piped into a single, District-wide CouchDB database for cross-detector analyses

Data Storage

- Initial view computation takes some time
  - Approx. 24 hrs per model per district per year
- New data is processed very quickly
  - Hourly chunks of data are processed within seconds
- Very fast per-detector and district-wide query performance: seek time ≈10ms
  - Delay comes from transferring data, not from computing the result.
- Storage space required: ≈800GB

Model Estimation

- Multinomial logit models
  - Regression of logarithm of odds of an event occurring
  - Traffic flow variables are the independent variables
  - Accident types are the events
    - Base category = no accident
    - Events, e.g.:
      - Property damage accident
      - Injury accident
  - Event-based sampling corrections are applied to adjust constants

Forecasting

- Models generate unique prediction functions for each type of accident
- Forecasts driven by streams of loop detector data
  - From stored historical data via CouchDB views
  - From real-time data streams (also possible with CouchDB views)

Model Validation

- 2007 was used for model estimation
- 2008 was used for model validation

Validation Conclusions

- In the aggregate (all of District 12, all of 2008) the models are generally within a few percent of observed accident characteristics
- For short periods of time and for short sections of freeway, the models are wildly wrong
  - Example prediction: an accident may occur with probability 0.0001
  - Example observation: an accident occurs or not
- **Do not use for short term predictions of possible accidents!**

Implement as a website

- Goals:
  - For all District 12 mainline loops (with valid data)
  - Predict relative odds of different accident characteristics at each detector
  - Provide basic analysis tools
- Two components:
  - Process the Data:
  - Respond to queries:

DEMO

Broader Implementation

- Deployment will scale roughly in proportion to the number of loops
- A broader implementation requires only:
  - Process raw data from PeMS (easy)
  - Estimate the model parameters (difficult)
  - Modest storage requirements (500 GB / year / District)

Questions

- Contact:

James Marca
(949) 824-6287 jmarca@uci.edu
Institute of Transportation Studies
AIRB Ste. 4000
University of California
Irvine, CA 92697-3600