Graphical Abstract

Ambulance Dispatch Recommendation System based on Automated Crash Reports from Cell Phones

J. Bradford Burkman,Chee-Hung Henry Chu,Miao Jin,Malek Abuhijleh,Xiaoduan Sun

Ambulance Dispatch Recommendation System based on Automated Crash Reports from Cell Phones

Police receive an automated crash report from a cell phone. An AI system recommends whether to immediately dispatch an ambulance.

Recommendation		Wait for Call from Eyewitness	Immediately Dispatch
Needs	No	Correct	Increased Cost
Ambulance?	Yes	Normal Delay	Prompt Medical Help

1. Political entities provide criteria for tradeoff of lives and money
2. Find metrics for the criteria
3. Choose input data features; some are easily available ($), some are not ($$$)
4. Build a recommendation system based on historical data that satisfies the criteria and saves the most lives

Highlights

Ambulance Dispatch Recommendation System based on Automated Crash Reports from Cell Phones

J. Bradford Burkman,Chee-Hung Henry Chu,Miao Jin,Malek Abuhijleh,Xiaoduan Sun

- Supports transferability and benchmarking of different approaches on a public large-scale dataset. We have attached the code we used to perform the analysis on data from the Crash Report Sampling System (CRSS).
- Novel Application motivated by Emerging Technology: Machine Learning Classification Models for Dispatching Ambulances based on Automated Crash Reports
- New Use of Dataset: Used Crash Report Sampling System (CRSS), which has imputed missing values for some features, but not all of the ones we wanted to use. For the first time we have seen, we used the software the CRSS authors use for multiple imputation (IVEware) to impute missing values in more features, then compared the results with other imputation methods.
- Explicit Incorporation of Imbalanced Costs
- Explicit Incorporation of Political Dimensions
- Consideration of Marginal Effects of Threshold Shifting
- Perennial Machine Learning Challenge: Imbalanced Datasets

Ambulance Dispatch Recommendation System based on Automated Crash Reports from Cell Phones

J. Bradford Burkman[a,b,*,1], Chee-Hung Henry Chu[a], Miao Jin[a], Malek Abuhijleh[a,c] and Xiaoduan Sun[c]

[a] *School of Computing and Informatics, University of Louisiana at Lafayette, 301 E. Lewis St, Lafayette LA, 70503, USA*
[b] *Louisiana School for Math, Science, and the Arts, 715 University Pkwy, Natchitoches LA, 71457, USA*
[c] *Department of Civil Engineering, University of Louisiana at Lafayette, 131 Rex St, Lafayette LA, 70504, USA*

ARTICLE INFO

Keywords:
Automated crash notification
Ambulance dispatch
Emergency medical services
Machine learning
Imbalanced Cost
Imbalanced Data
Imputation

ABSTRACT

Some new cell phones can automatically notify an emergency dispatcher if the phone detects the deceleration profile of a vehicular crash. Most crash notifications come from an eyewitness who can say whether an ambulance is needed, but the automated notification from the cell phone cannot provide that information directly. Should the dispatcher immediately send an ambulance before receiving an eyewitness report? There are three options: Always, Wait, and Sometimes. The "Always" option refers to sending an ambulance to every automatically reported crash, even though most of them will not be needed. In the "Wait" option, the dispatcher sends police, but always waits for a call from an eyewitness (perhaps the police) before sending an ambulance. In the "Sometimes" option, the dispatcher relies on a machine learning recommendation system to decide whether to immediately dispatch an ambulance, reserving the option to send one later based on an eyewitness report.

This paper explores one option for building a machine learning (ML) model for making a recommendation in the "Sometimes" option. Our goal is to build a model that returns, for each feature vector (crash report, sample), a value $p \in [0, 1]$ that increases with the probability that the person needs an ambulance. Then we choose a threshold θ such that we immediately send ambulances to those automated crash reports with $p > \theta$, and wait for eyewitness confirmation for those reports with $p < \theta$. In an actual implementation, the choice of θ is political, not technical, so we consider and interpret several options.

Once a threshold has been chosen, the costs of the false positives (FP) and false negatives (FN) in dispatching ambulances are very different. The cost of sending an ambulance when one is not needed (FP) is measured in dollars, but the cost of not promptly sending an ambulance when one is needed (FN) is measured in lives. Choosing the decision threshold θ is ethically problematic, but governments implicitly choose such a tradeoff when they set budgets for emergency services.

We consider and interpret several options for the decision threshold θ based on the political consideration, "How much will it cost?" How many automated ambulance dispatches are we willing to fund (FP + TP) for each one of them that is actually needed (TP)? We will explore two versions of that question, the total and the marginal.

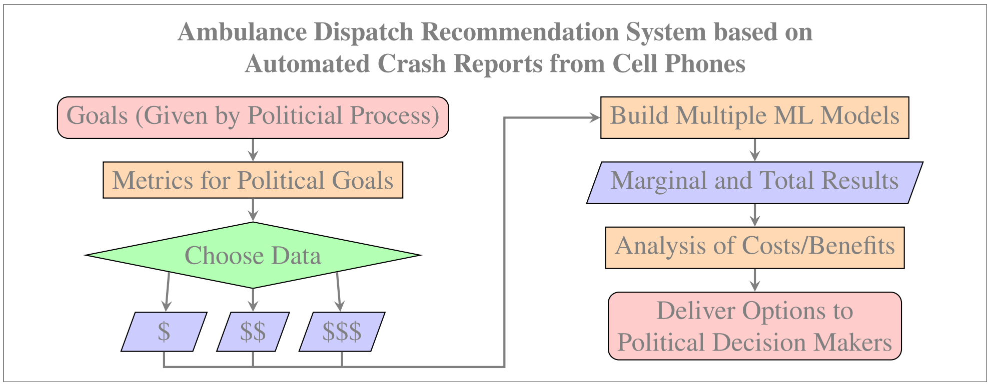
We show that the quality of the model depends highly on the input data available, and we considered three levels of data availability. The "Easy" level includes data the emergency dispatcher has before the notification, like time of day and weather. The "Medium" level adds information about the location and information from the cell service provider about the user, like the age and sex. The "Hard" level adds information that requires having access to records about the vehicle likely to be driven by the cell phone user and detailed and temporal information about the location, like lighting conditions and whether it is currently a work zone.

We used the data of the Crash Report Sampling System (CRSS) to validate our approach. We have applied new methods (for this dataset in the literature) to handle missing data, and we have investigated several methods for handling the data imbalance. To promote discussion and future research, we have included all of the code we used in our analysis.

1. Introduction

bradburkman@gmail.com (J.B. Burkman)
http://www.github.com/bburkman (J.B. Burkman)
ORCID(s):

2. Methods



2.1. Outline

1. Political Goals
2. Metrics for Political Goals
3. The Dataset
4. Choosing Features
5. Building Models
6. Analysis of Cost/Benefit
7. Choosing the Best Model

3. Results

4. Conclusions

5. Discussion

6. Future Work

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of Interest

Declarations of interest: none

Acknowledgements

George Broussard contributed to this work in the NSF Research Experiences for Undergraduates program.

Data Availability

The CRSS data is publicly available at
`https://www.nhtsa.gov/crash-data-systems/crash-report-sampling-system`
All of the code and generated data, tables, and graphs are available at `http://www.github.com/bburkman`

CRediT authorship contribution statement

J. Bradford Burkman: Conceptualization, Investigation, Writing - original draft, Visualization. **Chee-Hung Henry Chu:** Supervision, Methodology, Writing - review and editing. **Miao Jin:** Supervision, Methodology. **Malek Abuhijleh:** Data curation, Investigation, Methodology. **Xiaoduan Sun:** Data curation, Writing - review and editing.

References