Machine Learning Classifiers for Predicting Transit Fraud

Emergent Research Forum (ERF)

Jay Claiborne
Auburn University
jzc0019@tigermail.auburn.edu

Ashish Gupta
Auburn University
ashish.gupta@auburn.edu

Abstract

Fraud related studies cover a wide range of approaches for detection and identifying or mitigating risks associated with it. This study extends the existing body of research and our understanding on fraud with a specific focus on transit media. A large public transit transaction dataset was extracted from production systems of a large U.S Public Transit Authority system. We developed various machine learning models using techniques such as random forests, support vector machines, and artificial neural networks to effectively classify transit fare media fraud into binary groups and compare their relative model performance. We found that random forests and neural networks outperformed support vector machines for transit fraud detection. More data and parameter optimization are needed for improved results.

Keywords

Machine learning, public transit, fraud, digital disruption.

Introduction

The American Public Transportation Association (APTA) reports that in 2016, U.S. public transportation operations provided approximately 10 billion unlinked passenger trips (Neff & Dickens, 2015). This figure marks a steady rise in transit utilization over the last four decades. As transit authorities adopt new operating practices to meet growing demand, information systems are becoming increasingly vital components in a variety of transit operations. Transit authorities operate in a unique business environment. They enjoy a geographically captive market to whom they can largely dictate the choice of payment methods and therefore drive adoption rates (Quibria, 2008). One of the most significant ways technology has impacted transit is the evolution of fare media. Transit authorities use a variety of fare media including paper tickets, tokens, magnetic stripe cards, and smart cards. APTA reports that more than 50% of transit agencies use magnetic cards and that more than 25% use digital smart cards (Neff & Dickens, 2015).

Smart cards utilize radio frequency identification (RFID) and offer several advantages over traditional fare media, including the potential to collect and analyze transaction data to better understand transit system complexities and user travel behaviors (Gokasar et al. 2015). The advent of fare media that can digitally store information, create usage and transaction logs, and be managed remotely has also introduced new challenges to transit agencies. One significant challenge is electronic fraud. Counterfeit media, credit card fraud, and RFID risks associated with smart cards represent significant operational hurdles for the transit industry. With major transit authorities located in dozens of countries, a method to mitigate financial losses due to fraud could represent vast savings to public transit authorities. A 1997 study revealed that self-reported annual losses to theft, fraud, and counterfeiting ranged from roughly $1 million to as much as $1.8 million per year depending on the size of the respondent (TCRP, 1997).

Smart cards offer a means to mitigate the impact of fraud. Modern transit authorities utilizing smart fare media, can remotely deactivate fare products. This process is called “hotlisting” and is used for lost, stolen, and fraudulent cards. Transit firms are particularly vulnerable as they tend to serve a wide range of anonymous customers over large urban areas through numerous small transactions. Uncovering patterns
in those transactions that indicate fraud is typically the goal of marrying analytics to the vast transactional
data. Due to the fiscal impact of credit card fraud, interest in new techniques of detection and prevention
are unlikely to diminish.

This study addresses the problem of fraud detection by focusing specifically on transit related fare media
fraud utilizing machine learning classification techniques. The research questions that we seek to explore
in this study are

Research question 1. Can machine learning models effectively classify transit fare media fraud?

Ensemble methods, such as stacking, are often used to produce superior models in machine learning
research. This study seeks to extend the application of machine learning ensemble methods into transit
fraud prediction.

Research question 2. Which models, versions of models, or combination of models/versions yields the best
results for classifying transit fare media fraud?

Literature Review

As a research concentration, fraud detection offers a number of well documented techniques and associated
challenges. Five distinct types of fraud categories were documented in a review of the São Paulo transit
system in Brazil (Farzin, 2007). Bolton & Hand noted several difficulties in fraud detection research,
including unbalanced class sizes, differing costs of classification errors, and the difficulty of obtaining
sensitive financial data for research (Bolton & Hand, 2002). In previous research they underscored the
post-transactional and constantly evolving nature of fraud as criminals adopt new methods and techniques
(Bolton & Hand, 2001). Phua et. al, reviewed a 10-year span of fraud detection literature including analysis
of sample sizes, performance measures, and methods/techniques by fraud category. They highlighted a
deficiency in the instances of implemented research, studies utilizing temporal or spatial information, and
research utilizing less complex but faster algorithms (Phua et. al, 2010).

This study recognizes credit card fraud, counterfeit media, and RFID risks as three major categories of fraud
relevant to transit. The majority of research in fraud detection has centered around credit card fraud using
supervised techniques. Credit card fraud is the illegal use of credit card information either physically or
virtually (Zareapoor et al. 2012). Credit card fraud is committed in transit by using stolen credit cards to
purchase fare media online or at point-of-sale devices. The transit authority studied, will have lost an
estimated $1 million from 2014-2018 due to credit card chargebacks, which are primarily due to instances
of stolen credit cards used to buy fare media for unauthorized resale.

Counterfeit tickets range from low-tech paper copies of authentic tickets (in systems that use transit
employees to visually inspect patron tickets), to high-tech reproductions using stolen data. A variety of
counterfeit instances have been well documented in the literature and media. In 2012, Italian authorities
seized 2 million counterfeit train tickets worth $35M (Natanson, 2012). In the U.S., a 2011 counterfeit ticket
scheme conducted by a transit industry contractor was uncovered with an estimated $5M loss to the MBTA
(Massachusetts Bay Transportation Authority) (Moskowitz, 2011). A number of researchers have been able
to defraud transit fare collection systems using a variety of electronic counterfeiting techniques.

RFID devices and systems are subject a variety of attacks with varying degrees of impact. One method to
categorize the attacks is to separate them into groups according to the threat perspective. The
vulnerabilities of RFID physical, network transport, application, and strategic layers as well as a class of
multilayer attacks were described by a team of researchers in the Netherlands (Mitrokotsa, Rieback, &
Tanenbaum, 2009). Before RFID smart cards were deployed to the Dutch public transit system, successful
hacks were demonstrated for a variety of vulnerabilities. In separate instances students and hackers were
able to indefinitely extend a card’s life, reverse engineer a card’s internal circuitry to determine the security
algorithm used by the chip, and to execute RFID relay attacks using a homemade tag emulator whose
construction details were freely available on the internet (Felten, 2008).

The variety and complexity of fraud further support the approach of this research. While a great deal of
research deals with fraud prevention, this study elects to identify transit fraud based on transaction
behavior. Correctly classifying transit behavior as indicative of fraud will serve to mitigate the financial and
brand impacts while also possibly identifying new methods/trends in fraud. The classification method of choice for this research is machine learning.

Data

The dataset used for this project was derived from the transaction detail records of a major U.S. transit authority where fraud is discovered at a rate of roughly 1 per 500 tickets/cards that translates to approximately .02% of sales. For pilot testing, a daily aggregation of all fare media product transactions by unique serial number was produced for a continuous fourteen-day period using twelve variables. Approximately 1 million records were aggregated by unique serial numbers and provided summary transaction records for approximately 375,000 distinct fare media products. This data was randomized and sampled for a working dataset of 10,000 records.

To generate enough samples of fraudulent fare media, the continuous fourteen-day sample period was repeated for the first two weeks of each month for the first five months of 2016. This generated a concentrated group of 5,000 fare media records for deactivated cards/tickets. These 5,000 records, along with the 10,000 records from the non-suspect group, were combined to produce a pilot research dataset of 15,000 records. Aggregated fields are the sum of the qualifying transactions for the 14-day sample period. Table 1 lists the fields and a brief description of each column attribute. Prior IRB approval and required authorizations from the transit authority were availed for the purpose of this study. Table 1 provides the listing of variables used in the study developing classification models.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SERIAL_NBR</td>
<td>Each card or ticket is encoded with a unique serial number</td>
</tr>
<tr>
<td>HOTLISTED_FLAG</td>
<td>Has this card/ticket been deactivated 0=No and 1=Yes</td>
</tr>
<tr>
<td>MEDIA_TYPE_DESC</td>
<td>Type of media (paper or plastic contactless smart card)</td>
</tr>
<tr>
<td>RC_DESC</td>
<td>Rider class – type of rider and originating organization</td>
</tr>
<tr>
<td>MODES</td>
<td>Sum of modes used across all modes, facilities, &amp; devices</td>
</tr>
<tr>
<td>USE_TYPES</td>
<td>Sum of use types across all modes, facilities, &amp; devices</td>
</tr>
<tr>
<td>DEVICES</td>
<td>Sum of devices used across all modes, facilities, &amp; device types</td>
</tr>
<tr>
<td>FACILITIES</td>
<td>Sum of facilities visited</td>
</tr>
<tr>
<td>FARE_INTRUMENTS</td>
<td>Sum of fare types utilized (daily, weekly, monthly, etc.)</td>
</tr>
<tr>
<td>FARE_CATEGORIES</td>
<td>Sum of fare categories utilized</td>
</tr>
<tr>
<td>ENTRIES</td>
<td>Sum of entry transactions across all modes, facilities, &amp; devices</td>
</tr>
<tr>
<td>EXITS</td>
<td>Sum of exit transactions across all modes, facilities, &amp; devices</td>
</tr>
<tr>
<td>ENT_vs_EXT_</td>
<td>Difference between summed entries and exits</td>
</tr>
</tbody>
</table>

Table 1. Variable Descriptions

Methods

Through supervised learning, where the dataset trains the computer what output is desired, the machine learns and sharpens its accuracy with every transaction while analyzing new transactions for potential fraud. This ability to use the large datasets to construct and refine algorithms will likely lead to the digital disruption of the current practices in transit fraud detection that are mostly detected using manual screening and offers no automation.
A 2014 evaluation of 179 machine learning classifiers revealed that versions of the random forest, support vector machine, and neural networks modeling techniques consistently outperformed other methods (Fernández-Delgado et al. 2014). Based on these findings, this study focuses on these techniques as a starting point into transit related fraud detection. For the pilot study, the research strategy was to develop a binary classifier (“fraud”, “not fraud”) using a supervised learning approach. Accuracy, precision, recall, and F1-score were calculated for each modeling technique and tabled for comparison.

## Results

Pilot testing revealed that at least some machine learning classifiers effectively identified fraudulent fare media. Random forests and neural networks outperformed support vector machines, which underperformed in the testing areas of precision, recall, and F1 scores shown in table 2.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>AUC Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>0.9720*</td>
<td>0.97*</td>
<td>0.97*</td>
<td>0.97*</td>
<td>0.9660</td>
</tr>
<tr>
<td>SVM</td>
<td>0.9565</td>
<td>0.79</td>
<td>0.70</td>
<td>0.60</td>
<td>0.9602</td>
</tr>
<tr>
<td>ANN</td>
<td>0.9680</td>
<td>0.97*</td>
<td>0.97*</td>
<td>0.97*</td>
<td>0.9745*</td>
</tr>
</tbody>
</table>

*Highest value in each column

In the expanded results, random forests performed consistently well. Support vector machines showed a wide range of results, indicating that additional tuning may be necessary to produce results consistent enough to include in ensembles in the next phase of research.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision Not Fraud / Fraud</th>
<th>Recall Not Fraud / Fraud</th>
<th>F1 Not Fraud / Fraud</th>
<th>AUC Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>.9720*</td>
<td>.98</td>
<td>.97</td>
<td>.98*</td>
<td>.96*</td>
</tr>
<tr>
<td>SVM</td>
<td>.9565</td>
<td>.69</td>
<td>1.0*</td>
<td>.09</td>
<td>.81</td>
</tr>
<tr>
<td>ANN</td>
<td>.9680</td>
<td>1.0*</td>
<td>.92</td>
<td>.95</td>
<td>.99*</td>
</tr>
</tbody>
</table>

*Highest value in each column

There was some indication that the variables USE_TYPE, FACILITIES, and ENT vs EXT were the most impactful.

## Discussion & Conclusion

Transit agencies typically see significant overhead due to fraud and other inefficiencies. With increasing computing power, it is becoming possible to add the artificial intelligence techniques of machine learning to control such costs. This technological advancement creates opportunities for the status quo of transit firms using combined descriptive and predictive analytics to inform business decisions to disrupt that practice and automate the pattern recognition; i.e., machine learning, for predicting fraud (Mohri, Rostamizadeh, & Talwalkar, 2012).

It is through these identified fraudulent patterns that future fraud is predicted and policies can be enacted in efforts to prevent losses. This is the work of predictive analytics and is both quantitative and qualitative (Waller & Fawcett, 2013). It is through the identified fraudulent patterns that future fraud is predicted and policies enacted in efforts to prevent losses. This is the work of predictive analytics and could combine both quantitative and qualitative (Waller & Fawcett, 2013).
We found that machine learning classifiers may be a viable method of screening public transit records for indications of fraud. Of particular interest is the potential importance of the ENT vs EXT variable. This variable was included as a passive identifier of patrons who may be committing fare evasion. While this research is directed toward electronic fraud, this finding may indicate that physical fraud (gate surfing, forcing gates, jumping turnstiles, etc.) may be a useful signal for identifying electronic fraud. Other variables indicative of physical fraud should be considered in the final research.

Additional work needs to be done to tune the classifiers and identify additional relevant variables. While the pilot study was limited to 12 variables, there are dozens of potentially value-added variables available in the dataset. Ideally the work would be expanded across longer periods of time, test multiple variations/tunings of the three primary classifiers, and be tested on data from additional public transit authorities.

References