

Chapter VIII

Using Decision Trees to Predict Crime Reporting

Juliette Gutierrez

Claremont Graduate University , USA

Gondy Leroy

Claremont Graduate University , USA

ABSTRACT

Crime reports are used to find criminals, prevent further violations, identify problems causing crimes and allocate government resources. Unfortunately, many crimes go unreported. The National Crime Victimization Survey (NCVS) comprises data about incidents, victims, suspects and if the incident was reported or not. Current research using the NCVS is limited to statistical techniques resulting in a limited 'view' of the data. Our goal is to use decision trees to predict when crime is reported or not. We compare decision trees that are built based on domain knowledge with those created with three variable selection methods. We conclude that using decision trees leads to the discovery of several new variables to research further.

INTRODUCTION

The financial loss due to violent and personal crimes in 2004 was \$15.85 billion (Sedgwick, 2006) and 57.5% of these crimes were not reported to the police (BJS, 2005). Other costs of unre-

ported crimes include counseling costs, alarms, electronic surveillance equipment and indirect costs such as insurance and taxes (Sedgwick, 2006). An ongoing nationwide survey has been in use since 1973 in order to better understand both reported and unreported crimes. The Na-

tional Crime Victimization Survey (NCVS) is used to gather data on injury, theft, damage, the amount of lost work and other characteristics of the incident, victim and suspect. One of the goals of the NCVS is to understand the quantity of crimes and crime types that are not reported to the police (BJS, 2005). Each year, 45,000 households are interviewed about past incidents where they were the victim and the NCVS is the main source of data on the characteristics of criminal victimizations (NACJD, 2006). In addition, it also describes crime types not reported to law enforcement and the characteristics of violent offenders (NACJD, 2006).

The NCVS classifies each incident as a personal or property crime. Personal crimes include rape, sexual attack, robbery, assault and purse snatching. Property crimes include burglary, theft and vandalism. For example in 2005, 51% of personal crimes and 59% of property crimes were not reported (BJS, 2006a). Table 1 shows the large number of personal crimes, by crime type,

in 2005 and whether or not they were reported. There were a significant percentage of crimes that are not reported.

According to statistics from the Bureau of Justice Statistics (BJS), the criminal justice system does not act in response to many crime incidents because so many crimes are not discovered or reported to the police (BJS, 1967). Our goal is to define new techniques that can help law enforcement evaluate unreported versus reported crime data. Previous research done using the NCVS and descriptive statistics is limited to few variables which show only a limited view of the problem. In contrast, data mining allows for the use of more variables. Moreover, existing work uses descriptive statistics, such as logistic regression or binomial regression, which require a good understanding of these underlying techniques to interpret the outcome. Decision trees, in contrast, reveal which variables are most important and provide an easy to understand overview for users without a data mining background.

Table 1. Number of victimizations, by crime type and whether or not reported (BJS, 2005)

Crime Type	Number of Victimizations	Percentage Reported		
		Yes	No	Unknown
Completed Violence	1,658,660	62	37	1
Attempted/Threatened Violence	3,515,060	41	57	2
Rape/Sexual Assault	191,670	38	62	0
Crimes of Violence	5,365,390	47	51	2
Completed robbery	415,320	61	39	1
Attempted robbery	209,530	36	64	0
Robbery	624,850	52	47	1
Aggravated	1,052,260	62	37	1
Simple	3,304,930	42	55	2
Assault	4,357,190	47	51	2
Completed purse snatching	43,550	51	49	0
Attempted purse snatching	3,260	0	100	0
Pocket picking	180,260	32	67	2
Purse snatching/Pocket picking	227,070	35	64	1

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/using-decision-trees-predict-crime/4296

Related Content

A Generalized Comparison of Open Source and Commercial Database Management Systems

Theodoros Evdoridis and Theodoros Tzouramanis (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 13-27).

www.irma-international.org/chapter/generalized-comparison-open-source-commercial/7899

Effects of Domain Familiarity on Conceptual Modeling Performance

Jihae Suhand Jinsoo Park (2017). *Journal of Database Management* (pp. 27-55).

www.irma-international.org/article/effects-of-domain-familiarity-on-conceptual-modeling-performance/182868

INTECoM: An Integrated Approach to the Specification and Design of Information Requirements

Clare Atkins (2001). *Developing Quality Complex Database Systems: Practices, Techniques and Technologies* (pp. 240-260).

www.irma-international.org/chapter/intecom-integrated-approach-specification-design/8278

Hierarchical Architecture of Expert Systems for Database Management

R. Manjunath (2005). *Encyclopedia of Database Technologies and Applications* (pp. 271-275).

www.irma-international.org/chapter/hierarchical-architecture-expert-systems-database/11158

Theoretical vs. Practical Complexity: The Case of UML

Keng Siau, John Erickson and LihYunn Lee (2005). *Journal of Database Management* (pp. 40-57).

www.irma-international.org/article/theoretical-practical-complexity/3336