Spring 5-14-2015

Sharing and Analyzing Data to Reduce Insurance Fraud

Daniel J. Power
University of Northern Iowa, daniel.power@uni.edu

Mark L. Power
Iowa State University, mpower@iastate.edu

Sharing and Analyzing Data to Reduce Insurance Fraud

Daniel J. Power
University of Northern Iowa
daniel.power@uni.edu

Mark L. Power
Iowa State University
mpower@iastate.edu

ABSTRACT

Insurance fraud is a multi-billion dollar problem. Fraudulent practices occur frequently and often repeatedly. Fraud can be detected and prevented if appropriate data is collected, analyzed and shared among insurance companies. Appropriate decision support and analytics can be developed to routinize fraud detection. Creating these decision support capabilities involves addressing managerial, technological, and data ownership issues. This article examines these issues in the context of using new data sources and predictive analytics to both reduce insurance fraud and improve customer service. Evidence suggests that appropriate sharing of proprietary company data among industry participants and combining that data with external data, including social media and credit history data, can provide advanced data-driven decision support. Cooperative development and deployment of predictive analytics and decision support should reduce insurance costs while improving claims service. A process model is developed to encourage discussion and innovation in fraud detection and reduction.

Keywords

Insurance fraud, decision support, analytics, big data, data sharing.

INTRODUCTION

According to the Coalition Against Insurance Fraud (2015), insurance fraud is a multi-billion dollar problem costing Americans at least $80 billion a year. Preventing and detecting insurance fraud is a critical task at insurance companies. There is some evidence that insurance fraud is actually increasing (cf., Ramos, Kinzie and Epps, 2012), but information technologies to detect fraud also seem to be improving. Managerial processes are also improving data quality and data availability. Overall, the increasing availability of new data sources and more rapid access to historical data has created new possibilities for predicting and detecting insurance fraud. What is called "big data" by technology marketers has created renewed excitement about analytics and decision support. The promises and potential are significant. For example, Accenture Managing Director Eva Dewor (2013) asserts that "advanced analytics can help insurers reduce loss costs and improve their performance." The key question for managers and information technologists is how to make this possibility a reality and improve detection, deter misstatements, while increasing the speed of processing legitimate claims and improving customer satisfaction.

According to McKinsey Global Institute (MGI) and McKinsey's Business Technology Group, cf., Manyika, Chui, Brown, Bughin, Dobbs, Roxburgh and Hung Byers (2011), "big data" can become a basis of competition and a source of innovation. New data sources and real-time data can provide inputs for recommendation decision tools, sentiment analysis, fraud detection, and risk modeling. A key question is what data to use and how to analyze the data to get these results (Power 2014b). Sharing more insurance claims “big” data among companies can potentially reduce insurance fraud, but it is important to assess how this data will be collected, analyzed and shared to protect individual, organizational and societal interests.

This article examines the managerial, technological, and data ownership issues associated with new “big data” sources related to individual insurance claims, pooled frequency of insurance use, types of insurance activities, individual behavioral data, and individual social media activity. This article examines these issues and explores possible solutions for reducing fraud in insurance service systems.

INSURANCE FRAUD OVERVIEW

Broadly, the term fraud means lying and misrepresenting facts. Derrig and Krauss (1994) proposed that the term fraud be used only for criminal acts that are "provable beyond a reasonable doubt, that violate statutes making the willful act of obtaining money or value from an insurer under false pretenses or material misrepresentations a crime." Both fraud broadly and narrowly defined creates problems. Predicting criminal fraud means finding indicators that are highly correlated with fraud activity. Fraud, both criminal and petty, occurs at many points in the insurance cycle -- at the application stage, at the claims stages, and at the final settlement stage. Different types of analytics and decision support seem necessary at each stage in the insurance cycle. For example, at the application stage, decision support may focus on screening to predict fraud or providing an assessment of the risk of fraud and the potential magnitude of subsequent fraudulent claims.

The Insurance Information Institute (http://www.iii.org) identifies five private insurance types as most likely to have questionable claims: 1) Casualty, 2) Vehicle, 3) Property, 4) Workers Compensation and 5) Commercial. According to the National Insurance Crime Bureau (NICB), questionable insurance claims rose by 16 percent from 100,201 in 2011 to 116,171 in 2012. According to the Coalition Against Insurance Fraud (Insurancefraud.org), “Auto insurers lost $15.9 billion due to premium rating errors in private-passenger premiums in 2009.” Also, bogus and abusive auto insurance claims ranged between $4.3 billion and 5.8 billion in 2002, or between 11 percent and 15 percent of total payments (Insurance Research Council, 2008).

Fraud in government health insurance plans is common and usually involves billing for services or supplies that are not delivered. Medicare has no official estimate of the amount of fraud incurred each year, but the Federal Bureau of Investigation cites estimates of Medicare fraud at 3 to 10 percent of all health care billings. In 2011, Medicare expenditures totaled approximately $565 billion. The FBI is the primary agency for exposing and investigating health care fraud, with jurisdiction over both federal and private insurance programs.

Questionable claims and other fraudulent behaviors are a significant problem. A SAS solution brief on Insurance fraud notes “As a relatively low-risk, high-return activity, claims fraud holds particular appeal for organized crime syndicates, which account for a growing proportion of insurance fraud.“ Different data are needed to detect insurance fraud and support decisions to investigate instances of fraud for each insurance type.

INSURANCE INDUSTRY NEW DATA SOURCES

There are many potential internal company and external sources of data for detecting various types of insurance fraud related to different insurance types. The volume of data is large, sometimes unstructured and growing as health insurance coverage expands and as passive sensor and driving behavior data is collected from the operation of motor vehicles. For example, the U.S. Census Bureau reported that in 2012 almost 199 million people in the U.S. had Private Health Insurance and about 101 million had Government Health Insurance. In the U.S., approximately 86% of the driving population is insured (http://www.ircweb.org/).

Fraudulent claims are increasing. An insurance claim is a formal request for payment based on the terms of an insurance policy. Depending upon the type of insurance, many claims can be filed by an insured person or by a requesting third party like a hospital or auto repair shop during a policy period. Claims filed by third party providers of health services generate “big” data. Similar volumes of claims exist for other types of insurance. For example, based on data at FloodSmart.gov, approximately 100,000 flood insurance claims were paid each year from 2008-2012. The average claim was $42,000. Some estimates indicate 5%-10% of the flood insurance claims paid may have involved fraudulent claims. Individual claims can involve substantial payments by an insurer.

In the U.S. the average homeowner files a claim once every 10 years. The frequency of claims activity can indicate fraudulent activity so many years of data must be kept and analyzed when each claim is processed. U.S. insurers track claims in national databases called CLUE and A-PLUS. A number of additional industry sponsored and independent organizations collect and share data. For example, Fast Track Plus™ is a compilation of quarterly insurance industry data by state for private passenger auto and homeowners carriers. LexisNexis markets Accurint® for Insurance to help locate, verify and investigate customers and their claims. The National Insurance Crime Bureau (NICB) maintains vehicle theft databases for use by the insurance industry. Individual insurance companies with multiple products may however have customer data silos with limited metadata that is hard to integrate with data from other insurance companies, with external data or with other internal databases. If fraud mitigation efforts are to be successful, data sharing problems must be overcome.

With the broad use of web-based applications new data can also be captured and shared. Both web-based insurance applications and claims filing can generate machine and interaction data that may help detect fraudulent activity. Text analytics can assist in processing statements about claims to detect discrepancies and patterns. Once there is the suspicion of a fraudulent claim, then analysis of social network data may help identify others involved in similar fraudulent activities. In general, data analytics can support the decision to flag or target an application or claim for further investigation.

PREDICTIVE ANALYTICS AND DECISION SPPORT

Predictive analytics involves systematic analysis of data using quantitative and statistical models to predict and forecast results, trends and behavior patterns. Data-driven decision support emphasizes access to and manipulation of a time-series of internal company data and sometimes external data, cf. Power (2013a) and Power (2013b). Both predictive analytics and data-driven decision support can use new data sources and better access to data in silos. Sharing data among insurers can also provide better data sets.

A number of software vendors provide tools and capabilities to improve fraud management processes, some of the capabilities include 1) analytic techniques like business rules and anomaly detection, 2) predictive analytics to automatically route suspicious cases for review, 3) tools to automatically assemble alerts from multiple monitoring systems and prioritize and route to appropriate team members for investigation, 4) social network analysis to identify targets for investigation, 5) streamlined case management, and 5) advanced text analytics and data mining. These new capabilities which analyze both structured and unstructured data can help identify and predict fraudulent activities that might not be detected using traditional methods.

For many years, screening and scoring models that have been proposed and tested to identify or deter insurance fraud including logistic regression, decision trees, naive Bayes models, neural networks, and weighted criteria models (cf., Derig, 2002; Derrig and Kraus, 1994). Analytics and decision support using one or more of these models can assist in deterring fraudulent claims by flagging opportunities for abuse and by detecting claims with a high likelihood of involving significant fraud.

One simple method of detecting fraudulent claims and insurance scams is through analysis of a specific claim compared to other similar claims. Because claims for certain groups of insured risks tend to fall within a range, a business rule might specify that any claims more than 3 standard deviations above the mean of all claims are flagged for insurance investigators to conduct further review and analysis. If a claim seems unreasonably high, the insurance company may then require additional proof of loss or may conduct a field inspection of the situation. Also, examining and scoring a client's or third parties history of making claims against an insurance policy or across policies can help identify potentially fraudulent claims activity.

A number of additional analytical techniques have been identified as effective in detecting fraud: 1) Calculation of statistical parameters for a data set to identify outliers that might indicate fraud, 2) Classification, using decision trees, to find patterns among data elements, 3) Stratification of data to identify unusual data values, 4) Joining diverse data sources to identify matching values that should not exist, 5) Analysis to identify duplicated transactions, and 6) Checking input parameters on applications and claims for reasonableness and validating entry dates and amounts.

Decision support and predictive screening and scoring models should focus on identifying suspicious cases so investigators and managers can assess the cost of further information gathering and investigation. So what steps should be taken by insurance companies and industry groups?

REDUCING INSURANCE FRAUD

Ultimately managers are responsible for creating and maintaining a fraud detection and management system in a service business like an insurance company. According to Bermingham (2014), "Insurance companies have taken steps to improve the ability to identify and address fraudulent claims, but these efforts are typically fragmented. Because of the lack of a collective industry approach – most carriers work independently. In relation to technology, insurers sometimes lack the proper data mining system to help identify potential fraud and the business processes to follow up on flagged claims activity." He notes “Some carriers are already implementing advanced analytics and claims predictive modeling to help in areas such as workers’ compensation claims and auto bodily injury.”

Reducing fraud is a major task that will use significant resources. There is evidence of some successes. Ramos et. al (2012) report that a "major insurance company decided to implement advanced analytics and claims predictive modeling to support fraud referrals in its workers' compensation and auto bodily injury lines of business, which it had identified as the areas offering the greatest opportunities. The models use both company data and also external data sources to identify potential fraudulent claims. Even before a claim is submitted, the models can identify policyholders who are more likely to submit fraudulent claims, which is useful in underwriting."

Manyika et al. (2011) assert that "big data" and "sophisticated analytics can substantially improve decision-making." The ongoing task is to figure out how to make that happen. Talented professionals will need to figure out answers to the "how" questions for specific companies and industries in the next few years. Managers and legislators will need to clarify the constraints on data analysis in the areas of data privacy, data security, intellectual property, and liability for erroneous results. In some cases, laws related to insurance fraud will also need to be updated.

As Derig (2002) noted, the goal of increasing decision support and analytics in a claims processing system should not be to replace adjusters, but rather to assist and provide more timely information and flag cases for expert evaluation. The philosophy of decision support has emphasized the need to keep human decision makers in the decision loop (cf., Power, 2012; Power, 2013a). The goal of increasing decision support in the insurance cycle is not to automate claims processing and fraud detection.

Verma and Mani (2015) explain the importance of integrating data and the possibilities for including sentiment analysis, text mining and social network analysis to supplement traditional statistical analysis of claims data. Figure 1 proposes an iterative set of tasks that may improve the insurance fraud detection process. Managers at insurance companies should conduct analyses of current processes and consider following the five steps iteratively with the twin goals of reducing fraud while improving customer service.

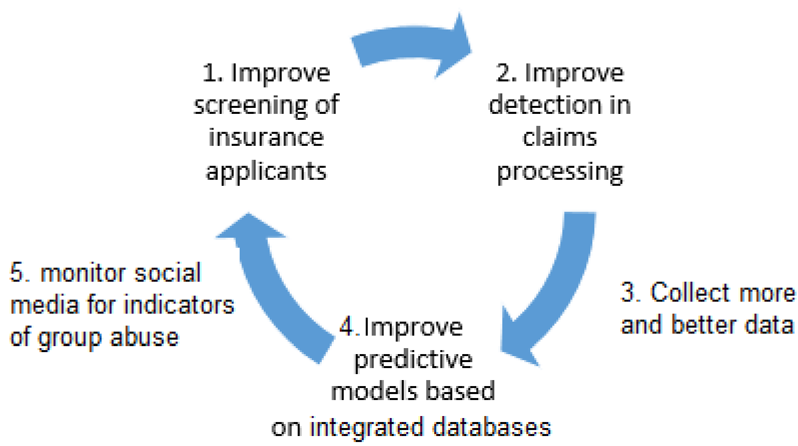

Figure 1. Improving the Fraud Detection Process

First, improve screening of applicants. A weighted screening model based upon historical fraud data can either reject applicants or assign a likelihood score for fraud to each applicant. Managers can determine how much risk they want to assume. Customers that have a high likelihood for fraud can be monitored more closely for fraudulent behaviors.

Second, improve detection and decision support for processing of claims. Identify claims that based on historical data are potentially fraudulent. An analytical model based upon the cost of various actions from rejecting the claim to investigating the claim can recommend follow-on actions. Use text analytics to compare statements made about a claim by the claimant at different times and by others to detect discrepancies Verma and Mani (2015).

Third, collect more and better data, especially track claims and suspicious behavior to improve predictive detection models. As fraudulent claims are identified using other means, the screening and detection models can be improved. Meta-analysis of claims can be linked to data derived from other anti-fraud and fraud detection measures to improve the models.

Fourth, improve predictive models based on integrated databases, create better industry-wide anti-fraud databases that merge data from participating insurers, data from credit bureaus and other, novel data sources. Applicants should be checked for prior/suspected abuses to identify patterns and then evaluated for possible links to other abusers.

Fifth, monitor social media and other sources to identify indications of abuse and consider incorporating social media data in policy application and claims processes.

Fraud detection models and processes can be updated to use new data sources and analytical and decision support capabilities.

CONCLUSIONS

Managers at insurance companies can deploy systems to predict fraudulent behavior that will reduce insurance costs, improve profitability and create opportunities to offer better insurance products and service, cf., Verma and Mani (2015).

Insurance company fraud mitigation is a step-wise process. First, accurate and complete data must be maintained about policy applications and claims. Second, data about claims and coverage must be shared among all actors so that analytical predictive models can be developed. A good example of doing this is the recent release of a Medicare public data set. Greater transparency of data on claims and billing enables creation of expected cost models and expose exaggerated costs and claims to increased scrutiny. Finally, insurance companies need to record data from investigations of possible fraud to provide data that can be analyzed to identify correlations with fraud activities.

The Coalition Against Insurance Fraud (www.insurancefraud.org) regulatory guidelines taskforce encouraged standardization among all jurisdictions in the insurance anti-fraud effort, including mandatory reporting of suspected fraud. Regulations can make it easier for companies to share data, standardize reporting and create a nationwide database. Insurers can benefit from frequent reporting of suspicious activities, claims denied for fraud, claims referred to a state's fraud bureau, and data about the claimant and claim. Assuming safeguards are in place to protect privacy and secure the database, then analysis of the database can identify trends and potentially help reduce fraud. The National Association of Insurance Commissioners NAIC) also has a model fraud act.

Analytics and targeted decision support systems should screen and sort claims into what Derig (2002) calls "bins with associated actions". Ideally this categorization and sorting will also attach likelihood estimates to the category and indicate the magnitude of the suspected claim misrepresentation. Some of the categories will certainly have proof criteria far short of establishing provable criminal activity beyond a reasonable doubt. The cost of gathering more information should also be considered in recommending actions to fraud investigators. Reducing loss costs and improving company profits with analytics and decision support is dependent upon critical success factors like good data and good models (cf., Power, 2013b).

Bermingham (2014) argues insurance carriers "need to create a more streamlined way of sharing information universally. If information was shared in bulk across the entire industry, the data set would be massive and would decrease the possibility of fraudulent claims falling through the cracks." Verma and Mani (2015) argue "Insurance firms always hesitate in implementing analytics because of the initial time investment needed for analytics solutions (p. 9)". The increasing interest in predictive analytics and new data sources like social media data will hopefully encourage increased experimentation with building decision support, analytics and integrated databases related to detection of fraud.

New data sources, metaphorically called "big data" that is shared and used for predictive analytics and decision support can potentially reduce many types of insurance fraud. The large number of claims each year and the sophistication of some fraud activity means that managers will need to continually improve and upgrade systems and procedures for fraud detection and monitoring. No insurance company is immune from fraudulent claims and insurers should share data rather than transferring fraudulent policy applications or claims to other insurance companies using less sophisticated predictive analytics and decision support. Fraud is a global insurance industry problem that must be addressed collectively by industry participants. A discussion of innovative uses of data, improved decision support and analytics is a useful starting point in reducing fraud in many domains, including insurance.

REFERENCES

1. Bermingham, P. (2014) Commentary: The Rise in Insurance Fraud and How to Combat it, *Claims Journal*, February 19 at URL www.claimsjournal.com.

2. Coalition Against Insurance Fraud (2015) Insurance fraud: the crime you pay for, at http://www.insurancefraud.org/fraud-backgrounder.htm#.VS_3VyFVhBc .

3. Derrig, R. A. (2002) Insurance Fraud, *The Journal of Risk and Insurance*, 69, 3, 271-287.

4. Derrig, R. A. and Kraus, L. (1994) First Steps to Fight Workers Compensation Fraud, *Journal of Insurance Regulation*, 12, 3, 390-415.

5. Dewor, E. (2013) Analytics: A critical tool in detecting—and preventing—insurance fraud (Part 3 of 3), September 12 Accenture Insurance Blog at URL http://insuranceblog.accenture.com/analytics-a-critical-tool-in-detecting-and-preventing-insurance-fraud-part-3-of-3

6. Insurance Research Council (2008) Fraud and Buildup Add 13 to 18 Percent in Excess Payments to Auto Injury Claims, Press Release, November 24 at http://www.insurance-research.org/sites/default/files/downloads/IRC_Fraud_NR.pdf.

7. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A. (2011) Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, May, at URL http://www.mckinsey.com/ insights/business_technology/big_data_the_next_frontier_for_innovation .

8. Mohr, A.(2012) How Insurance Companies Detect Insurance Scams, *Investopedia.com*, February 17, 2012 http://www.investopedia.com/financial-edge/0212/how-insurance-companies-detect-insurance-scams.aspx .

9. Power, D. J. (2012) What should managers know about predictive analytics and model-driven decision support? *Decision Support News*, 13, 16, August 19 at URL http://dssresources.com/faq/index.php?action=artikel&id=246.

10. Power, D. J. (2013a) Decision Support, Analytics, and Business Intelligence, Business Expert Press, New York.

11. Power, D. J. (2013b) What are critical success factors for using advanced analytics? *Decision Support News*, 14, 25, December 8 at URL http://dssresources.com/faq/index.php?action=artikel&id=273.

12. Power, D. J. (2014a) How can analytics and decision support reduce insurance fraud? *Decision Support News*, Vol. 15, No. 5, March 2, 2014 at URL http://dssresources.com/faq/index.php?action=artikel&id=292.

13. Power, D. J. (2014b) Using 'Big Data' for analytics and decision support. Journal of Decision Systems, Vol. 23, No. 2, pp. 222-228.

14. Ramos, C., Kinzie, J. and Epps, D. (2012) Reducing claims fraud: A cross-industry issue, *Deloitte Forensic Center* at URL https://www.deloitte.com.

15. Verma, R. and Mani, A.R. (2015) Using Analytics for Insurance Fraud Detection, *FINsights Journal*, Infosys, Issue 10 at URL http://www.infosys.com/FINsights/Documents/pdf/issue10/insurance-fraud-detection.pdf .