

Tapping the Benefits of Text Mining for Fraud

By: James Ruotolo, SAS

As more and more insurers leverage analytical techniques for detecting suspicious claims, the benefits of text mining are becoming more apparent.

Analytical models rely heavily on structured data fields. But within many claim organizations, some of the most useful information is buried within unstructured text fields like claim notes. Fortunately, text analytics can bridge the gap.

Where to Look

Ask any insurance investigator where they look for detailed information about a claim, and the answer is almost always the same: the claim notes. The claims process collects and generates large volumes of text-based information, such as adjuster notes, emails, customer service calls and claimant interviews. In fact, unstructured data can represent up to 80 percent of claims data. This information can be used to help reduce investigation costs and optimize recovery operations.

But most predictive models are built using only structured data elements – meaning your fraud detection decisions could be based on a mere 20 percent of the available information!

Recognizing the Challenge

A recent survey by Accenture identified several key areas of focus for claims executives, including concern over an explosion of unstructured data. Most insurance companies are aware that they need better plans to deal with big data but many organizations still do not have a good handle on how to better leverage their growing repository of text-based information. Experts agree that unstructured data is growing at an exponential rate and insurers are increasingly looking to use both internal sources, like claim notes, as well as external sources including social media.

However, according to a recent survey by the **Coalition Against Insurance Fraud** on the State of Insurance Fraud Technology, only 40% of insurers use text mining as an anti-fraud tool, leaving lots of room for improvement.

Making it Happen

Especially for long-tail claims like workers' compensation, some of the best data is in the claim notes. In many insurance fraud detection projects, anywhere from one-third to one-half of the variables used in the fraud detection model come from unstructured text information.

One common approach to text mining for fraud detection is variable extraction. If a variable is

A recent survey by the Coalition Against Insurance Fraud on the State of Insurance Fraud Technology, only 40% of insurers use text mining as an anti-fraud tool, leaving lots of room for improvement.



**Coalition Against
Insurance Fraud**

useful but does not exist as a discrete field in the data, the variable can be created through the use of text mining technology to extract the relevant information. For example, a flag indicating whether or not an ambulance was called to the scene of an auto accident could be a useful variable as an indicator of accident severity. But if this data is not captured in the claim system, it is often possible to mine the text of the claim notes to discover this information.

Text mining is more than just keyword searching. Good text analytics tools are able to interpret the context in order to determine the difference between “an ambulance was called to the scene” and “no ambulance was called to the scene.” Once the correct information is extracted, a new column for the flag is created and can be leveraged in a traditional predictive model. There are many examples of how text mining can improve fraud detection results.

The good news, according to the Coalition survey, is that many insurers say they plan to invest further in technology in the future. Text mining is the second most sought-after technology, right behind predictive modeling.