Predictive Analytics: A Breakthrough in Subrogation Identification
by Anthony Mattioli, Full Capture Solutions, Inc., East Hartford, Connecticut

Even the best subrogation professionals can’t do anything about claims that they never get to see. Overworked claims handlers can miss opportunities that should have been referred. What’s the solution? The introduction of predictive analytics technology into the claims world might be the answer.

Industry research estimates that 10% of legitimate recovery opportunities may never be identified, pursued or recovered. That could be $1 billion or more - financial opportunities which are embedded in existing claims but, for various reasons, are never captured.

The NASP benchmarking studies have shown a wide range of recoveries as a percentage of paid losses, even within the same line of business. These studies usually prompt management questions along the line of “how did we do versus the benchmarks?” A better question might be “Are we recovering everything that we could be?” The answer to this question depends on the completeness of the referral process.

Companies usually rely on their claims adjusters’ judgments and instincts to evaluate subrogation potential. Already typically overwhelmed with investigating and settling formidable caseloads - due to sheer claim volume, coworker turnover or both - it’s understandable that claims adjusters simply can’t catch everything.

Current Practices

Practices that some insurers have implemented to recapture missed opportunities and increase the flow of files to the subrogation unit have met with mixed results.

Some companies ask subrogation professionals to review all claims to ensure that no opportunities are missed, which is very resource-intensive. Others have tried periodic closed claim blitzes, with subrogation staff stopping what they’re doing and reviewing tons of closed claims - file by file, screen by screen, hour by hour. Anyone who has done this knows how much fun that can be! Vendors are sometimes brought in to manually scour the files, but this can be disruptive to the office, and vendors have a tough time doing a thorough review in a cost-effective manner. Other companies rely on training of the adjusters as their practice, but this can be inconsistently applied and can result in over-referral of “false positives.” Canned reports that list claims with certain loss description codes or dollar values also result in significant “false positives.”

All these efforts are a result of a basic recognition that subrogation opportunities exist in closed files and new opportunities are missed on a daily basis. While each method can generate some financial lift, they are not sustainable control processes and are reactionary in nature, not addressing root causes.

Predictive Analytics Technology

Predictive analytics is an advanced type of enterprise intelligence technology. It includes the application of sophisticated analysis techniques to data. Predictive analytics is an advanced form of “data mining,” but it differs in that it includes much more sophistication in its statistical methods and underlying data science. Predictive analytics reads and analyzes massive amounts of data, discovers meaningful patterns and relationships in the data, and provides actionable decision-making information about the future.

Predictive analytics (and corresponding real-time recommendations) is cutting-edge in the evolution of enterprise intelligence. Over the past twenty years or so, techniques used to analyze data for decision-making have progressed through several phases, as companies have grappled with maximizing the value of data buried within their systems.
Evolution of Enterprise Intelligence Techniques

<table>
<thead>
<tr>
<th>Phase</th>
<th>Technique</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Human Instinct</td>
<td>Experience/Gut feel</td>
</tr>
<tr>
<td>2</td>
<td>Canned Reporting</td>
<td>Transaction registers</td>
</tr>
<tr>
<td>3</td>
<td>Parameterized Reporting</td>
<td>If “x” and/or “y”</td>
</tr>
<tr>
<td>4</td>
<td>Ad hoc Query</td>
<td>Let me specify “x” and “y”</td>
</tr>
<tr>
<td>5</td>
<td>Predictive Analytics</td>
<td>Recognize the patterns</td>
</tr>
<tr>
<td>6</td>
<td>Real-Time Recommendations</td>
<td>Act on it as it happens</td>
</tr>
</tbody>
</table>

In a nutshell, predictive analytics encompasses: 1) data pattern recognition, 2) translation of the patterns into a set of business rules, and 3) actions and decisions.

Applications

Predictive analytics is rapidly being deployed in many industries. The banking and direct marketing industries have been early adopters. Direct marketers evaluate consumer data to identify very narrow buying patterns, so they know who they should market to, and what campaigns are most effective to certain subsets of their markets. Banks can assess future performance of loan portfolios and applicants’ credit worthiness.

The insurance industry has embraced credit scores as a reliable indicator of claim likelihood and has integrated credit scoring into many risk evaluation/underwriting processes. Credit scores and their relationship to losses were discovered using predictive analytics techniques.

Claim Data and Subrogation Identification

One of the biggest obstacles to thorough subrogation identification - whether performed by people or technology - is that claims data comes in different formats: structured, semi-structured and unstructured. Patterns are based on thousands of unique features.

Structured data is standardized, easily entered and handled information, such as numbers or insurer-designated codes for lines of business, line abbreviations, causes of loss, and the like. Often, structured data fields have a defined set of values which are accepted.

Semi-structured data has some structure, but may be irregular and/or incomplete. It doesn’t necessarily conform to a fixed schema. Web-related information, such as HTML formatting and markup, and fields where a brief description of loss may be entered, are examples.

Unstructured data is non-standardized, freeform, explanatory information. Examples of unstructured data include adjuster notes, imaged documents and email messages. This type of data is a goldmine of rich, but hidden, information. Unlocking it and extracting its actionable business value has been a daunting necessity. Typographical and grammatical errors, non-standard abbreviations, synonymy and polysemy issues, page layout inconsistencies, and limitations in automated systems are among the myriad of hindrances to streamlining subrogation recognition.

Historically, one of the difficulties insurers have had employing technology to identify subrogation opportunities is that the tools employed are insufficient to uncover the underlying “story” of each claim. While structured data is relatively accessible (as long as you can gain access to your company’s IT resources), the type of data captured in a structured format in claim databases is meant primarily for financial transactions and reporting. It rarely includes information about the details of the claims uncovered during the investigation that are important for liability assessment and collection, thus has limited value on its own in the predictive analytics process.
How Predictive Analytics Is Applied

Data science experts have turned their expertise to the insurance industry, and have introduced the application of predictive analytics to enhance several areas of claim processing. For example, identification of fraud indicators, litigation potential and other troublesome characteristics of claims can be spotted and acted on much quicker through the use of predictive analytics. But one of the most valuable applications to insurers is the identification of subrogation recovery opportunities. Current predictive analytics capabilities now include the expertise to evaluate all digitized claim data - structured, semi-structured and unstructured data - in order to analyze and evaluate the entire claim “story” in a manner that is truly breakthrough. The wide scope of data enables effective and comprehensive pattern recognition - for both closed and open claims. Applying the results of the pattern recognition, and integrating statutes and liability rules, claim files can be scored for their likelihood of subrogation potential, prioritized and referred quickly to the subrogation professionals. The analytics can be applied on a current basis (for complete and timely referrals) or retrospectively (to identify claims that were missed).

Predictive analytics is a non-intrusive technology. What this means is that it “spiders” over existing data without intruding upon it. It reads the data without penetrating it. Google, Yahoo and other search engines employ a similar underlying technology. They scan over an enormous quantity of website content - all in different formats and structures - and recognize patterns in the content that match whatever you may be searching for. Search engines don’t penetrate the website content, and it doesn’t matter how you structure your website. It reads it quickly and sorts it all based on your preference. The way predictive analytics works over claim data is very similar. It scans over your claim data without invading it, identifying patterns and reports back results and actions that need to be taken.

Conclusion

The insurance industry is overdue for a standardized, aggressive, expeditious and consistently thorough approach to the subrogation identification process. Along with a growing awareness and appreciation among insurers of the tremendous financial value of an effective subrogation process, the introduction of proven predictive analytics technology to the subrogation identification process is an excellent fit with companies’ efforts to maximize recoveries. Not only would it extract more recovery revenue from an insurer’s available claims data - coming closer to 100% subrogation recovery - it would also serve to keep costs down, improving their pricing and ranking among their competition. Use of this new technology for mining claims data by insurance carriers is inevitable. The question is not if, but when.

This article appeared in the Spring/Summer 2005 Issue of the NASP Subrogator.
© NASP