



Claims Predictive Modeling Using Industry Standard Claims Submission Data

CASE 2008 Spring Meeting

Molly Bhattacharyya, Ph.D
March 2008

Discussion topics

- Industry Standard Claims Submission Data for Property/Casualty
- Applications other than for Pricing, Underwriting
- An example

Industry Standard Claims Data

- 98% of Auto insurers, 95% of Home insurers participate
- Data is produced and reported at regular intervals
- One of the most valuable information resources the insurance industry has collectively and collaboratively produced

Current application of the Data

- Current major use is for a comprehensive search of claims histories of applicants for pricing/ underwriting purposes.
- Rating factor relativities

What Data is Reported

- Policy Information
 - Policy Type
 - Inception date, expiration date
 - Renewal indicator
 - Others
- Claims Information
 - Claim type
 - Number of parties involved
 - Date, time of Loss
 - Loss report date
 - Others
- Claimant Information
 - Name , Address, Date of Birth/ Age
 - Relationship to Insured
 - OTHERS

Data (cont'd)

- Loss Information
 - Loss description
 - Loss location
 - Nature of injury
 - Body part
 - CPT, ICD-9 codes
 - Medical amount paid
 - Indemnity amount paid
 - Witness
 - Police Report
 - Hit and run
 - Single vehicle
 - Phantom vehicle
 - OTHERS

Data (cont'd)

- Property Information
 - VIN, Year, Make
 - Anti-theft device
 - Air bag status
 - Odometer reading
 - Residential/ Commercial/ off-premise
 - Fire indicator

- Service Provider Information
 - Names, Addresses, Date of Birth
 - Professional License information
 - Individual/ Business Indicator
 - OTHERS

Applications other than current uses

- Can this data, collected for submission, be used internally ?
- Why ask the question?
 - Level of information rich enough for other uses
 - Data is produced regularly
 - Data is available and accessible
 - Data is accurate, reliable and tested
 - Data is well understood

Predictive Modeling

- Use of statistical and computational techniques to:
 - Extract knowledge from large amounts of data
 - Identify patterns, relationships and correlations across various data elements
 - Make better business decisions

- The process of predictive modeling
 - Begins with collection and cleaning of data
 - Defining and creating the outcome variable, input / predictor variables
 - Creating the model to identify and quantify the relationship between the outcome variable and the predictor variables
 - Verifying the validity of the model
 - Applying results to predict outcome
 - Monitoring model performance

Predictive Modeling (cont'd)

- Models produce output which guides decisions for
 - Rates, rating tiers, surcharges and discounts
 - Underwriting guidelines and rules
 - Marketing targets
 - Claims Management
 - Improved claims handling
 - Better resource allocation

The Claims Process

- Claims notification and assignment
- Claims Investigation and evaluation
- Negotiation and settlement

Decisions at Different Stages in the Claims Process

- Assignment of claims to reps , units
 - Fraud referrals, Subrogation referrals
 - Complex claims identification
 - Claims likely to go to litigation
 - Determine claims settlement amount
-
- Models can be designed and structured to address each of these issues.

Claims Fraud

- Claims fraud is a significant and costly concern
 - According to the Insurance Information Institute and the NICB, it adds about 10% to Loss and loss adjustment expenses, resulting in \$24 billion of losses each year
 - Of this, \$5 billion are attributable to WC claims
 - Costs policy holders an estimated amount of \$200-\$300 a year in increased premium

- P/C insurers detect less than 20% of fraud
- Life/disability insurers detect less than 10% of their fraud
- Health care insurers detect only about 1% of their fraud

Claims Fraud (cont'd)

- Current shape of Fraud Detection
 - Most use manual reviews and industry standard 'red flags'
 - Moving towards predictive systems and technology used by credit card companies and banks to detect fraud

A Claims Fraud Model

- A definition of fraud
 - Independent medical examination
 - Referrals to SIU (Special Investigative Unit)
 - Verified/ Mitigated fraud
- Information on
 - Claim
 - Claimant
 - Policy
 - Bills
 - Payments
 - Others

Claims Fraud Model (cont'd)

- Probability of Fraud (BI) = f (Policy Inception date, Policy Expiration date, Age of Claimant, Injury date time, Injury report date, Hire Date, Injury Type, Nature of Injury, Body Part, ICD-9 codes, CPT codes, Witness, Police Report, ER, AWW, Job class, First payment date, number of distinct medical providers, distance provider-claimant, attorney, date of attorney introduction, ...)
- The algorithm processes and analyzes the data. Weights are calculated for each variable. Weighted values are combined to produce a model score. This model score expresses the likelihood of a claim being fraudulent
- Models are calibrated prior to implementation. Depending on business needs, threshold values of scores are selected. Claims within the threshold scores are routed for review and validation.
- The model score can change over the life of the claim

Claims Fraud Model (cont'd)

Variables known to be predictive

- Report lag, Treatment lag
- Provider type
- Number of providers
- Number of doctor visits
- CPT codes, ICD-9 codes
- Limits, deductibles
- Distance claimant-provider
- First Claim payment date
- Years of employment
- AWW
- Policy Start date – Claim date
- Policy End Date – Claim date
- OTHERS

PROS and CONS

- Most, if not all of these are available from standard industry claims submission data efforts.
 - Additionally, the Loss Description field provides significant scope for knowledge derivation using Text mining
 - External data can be combined with available fields (zip code)
- Not all desired input/ predictor variables are available or can be derived off the Industry Standard claims data
 - a number of reporting fields are optional, and therefore not reported
 - Information is at an aggregated level — granular data provides deeper insights
 - Bills and payments details are not available
 - Lag between event and data collection for reporting

RECAP

- Data prepared for industry standard claims submission contains Claimant, Policy, Loss and other data that have useful predictive value to support decision making
- Data is tested, reliable, well understood, available and accessible
- Fraud example shown. Many other applications
 - Identification of severe medical only claims
 - Identification of claims likely to go to litigation
 - Identification of claims with recovery/ subrogation potential
 - Claim duration analysis