Proceedings of the 3rd INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2008)
J. Li, D. Aleman, R. Sikora, eds.

A DATA MINING FRAMEWORK FOR IDENTIFYING CLAIM OVERPAYMENTS FOR THE HEALTH INSURANCE INDUSTRY

Abhinav Anand
Mining and Modeling Group
West Asset Management
Omaha, NE
aanand@west.com

Dmitriy Khots, Ph.D.
Mining and Modeling Group
West Asset Management
Omaha, NE

Abstract

The private health insurance expenditures in U.S. are expected to hit one trillion dollars by 2010. With over two hundred million insured lives, complexity in claim processing and government regulated timeliness for payment to medical providers; health insurance companies are susceptible for making payment errors. An insurance company covering 10 million lives approximately pays medical providers in excess of $400 million a year in "*identified overpayments*". Current industry practice to identify and recover such over-payments involves ad-hoc querying of medical claims databases and manual auditing of potentially overpaid claims. In this paper, we describe the successful deployment of a data mining framework, predominantly using predictive modeling and association rule discovery techniques to correctly identifying overpaid claims. The prediction problem is challenging due to the large volume of data, high dimension of the feature space and complex relationships among the features. We also discuss methodology to detect abnormal changes in the claims data that help us monitor and control the predictive models. This framework of cost containment taken as a whole, affords health insurance companies the ability to reconsider year-over-year annual premiums and rising healthcare costs.

Keywords: Overpayment identification, logistic regression, association rule discovery, healthcare data, claims processing.

Introduction

The healthcare costs in the U.S. have been rising at over double the inflation rate over the past years [1]. This is, in part, due to inefficiencies existing within medical insurance carrier processes. Payment errors in a trillion dollar industry yield millions of dollars in losses, therefore, preventive measures must be taken to either stop or identify and recover overpaid medical claims (overpayments).

An overpayment is defined as an erroneous payment made by the health insurance carrier to a healthcare provider or an insured individual, such that the amount paid exceeds appropriate, agreed to, service charges (reasonable and customary charges). Overpayments occur for various reasons including duplicate payments, incorrect coordination of benefits, (in)eligibility of the insured, or a mis-interpretation of a contract between the healthcare provider and the insurance carrier.

A standard industry approach to overpayment identification (OPID) consists of a combination of subject matter expert knowledge, ad-hoc querying of claims databases based on various industry leads, and manual auditing of claims tagged as potential overpayments. Overpaid claim recovery process is primarily contingent on state legislature and usually follows a predefined set of steps which results in a practically random work effort.

In this paper we discuss a data mining approach to OPID and recovery that has been successfully deployed at "Accent" a division of West Asset Management - who is a major OPID and Recovery provider. We present the data mining framework around the OPID process, showcase efficiency wins over the existing methods, and discuss future applicability of this framework to other extensions of health insurance cost containment, such as Third Party Liability identification.

Challenges in OPID

The primary challenges faced by the insurance company and its outsourced vendors while identifying overpayments is the non-availability of precise automated business rules for payment of claims, incomplete information about the insured and the volume of claims filed daily. Based on the internal data at Accent, a prominent nationwide Insurance carrier can have more than fifty thousand medical providers with individual contract rates and process anywhere between 1 million and 1.3 million claims weekly. In some cases, sufficient information is not available to make a correct payment decision and the insurance company has to make a full payment as charged by the medical provider. Once these claims have been paid, most insurance companies attempt to identify and recover overpayments either internally or by outsourcing. Third party vendors that provide OPID services have to deal with variability due to client procedural and pricing changes which are not always conveyed to them in a timely fashion.

Almost all groups providing this service rely heavily on subject matter experts who study current billing practices, Medicare guidelines and client payment procedures to devise business rules, commonly referred to as "Edits" to identify potential overpayments. For example, some edits are mere logic statements, "if the patient "X" is a male, he should not have a claim for a hysterectomy procedure." If patient X has this claim, a flag is placed on the account and the claim would be reviewed. These potentials are manually analyzed to ascertain the overpaid amounts. Edits play an important role in restricting the size of claim pool to be manually audited. However, in an effort to minimize false negatives especially ones with high dollar amount, these business rules tend to be "loose." Given the trial and error that goes into development of new edits, there arises a need to use data mining techniques to discover knowledge from the claims data to help create smarter business rules. Association Rule discovery techniques have successfully helped un-earth patterns of overpayments that can then be converted into business rules.

Once these edits have filtered claims into a smaller subset, predictive modeling techniques can help reduce false positives and direct auditors towards claims with highest propensity of overpayment.

One of the most common overpayment types is Coordination of Benefits popularly known as COB in the insurance industry. As defined by the Illinois Department of Financial & Professional Regulation [2], COB is a "method by which two or more carriers or plans can coordinate their respective benefits so the total benefit paid does not exceed 100% of the total allowable expenses incurred." Typical scenarios are when a person has both Medicare and private insurance. Unfortunately while paying a claim, the insurance carrier is not always aware of a secondary insurance for the patient. Using predictive modeling on historical payment profile (features like past payments, co-payment, coinsurance payments, deductibles, etc.) for a patient we are able to predict the probability that a claim can be coordinated with a secondary insurance.

The data mining framework developed and deployed at Accent encapsulates modules to overcome the three major issues currently faced by OPID vendors: (1) in-sufficient and primitive business rules or Edits, (2) variability in client data and (3) high volumes of false positives while manual auditing.

Data Mining Framework

The claims data consists of characteristics of the following major healthcare elements: claim, line item, patient, insured, provider and group. In particular, claim information identifies the incident of services provided to the patient by the provider such as the billed amount, medical diagnosis, claim processing center, etc. Line-items contain information about particular services that are billed with the claim, including specific procedures performed or drugs administered. The other characteristics give the demographic and insurance-related information for the patient, the insured individual (who is not necessarily the patient), the healthcare provider (hospital or physician), and the group (the client of the insurance company, either an individual or a company).

Since the incidence of overpayment is based on a medical claim, the claim becomes the key, and all further analysis is either rolled up to the claim level (in case of line items) or is joined to the claim level (for the rest of the elements).

The entire data mining framework for OPID can be decomposed into two stages: development and production.

Development Stage

The development stage (Figure 1) begins with the in-depth analysis (both statistical and exploratory) of client feed database. Based on the existing Subject Matter Expert (SME) knowledge, a set of edits is developed and enhanced by the association rule discovery algorithms. An edit is a Structured Query Language (SQL) filter that uses business and association rule logic to parse through medical information and create potential overpayments. These potentials are placed into work pools, where they are available for auditing. Once auditing is completed, potentials are classified into non-overpayments and overpayments and then stored in historical files. The history is used as the training set by the predictive models, which employs logistic regression, decision trees and neural networks to predict the probability of a claim being overpaid. The predictive model also uses external data such as medical provider information.

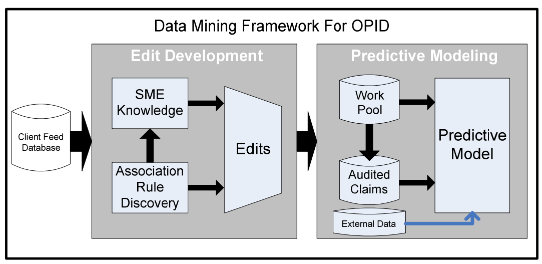

Figure 1: Data Mining Framework - Development

A validated and tested predictive model is then ready for deployment in the production system. The list of predictors explaining the binary target (overpaid vs. non-overpaid) is used by the Inventory Variance Detection System (IVDS) which monitors the variables directly from the client feed database. During the development stage, the benchmark tables containing variable statistics are created so that statistical process control methods could be used in the production stage to monitor the predictor variables. This is used to anticipate model drift as well as to alert operations on changes in client processes.

Production Stage

Figure 2 represents the architecture of production stage of the data mining framework. Accent receives client data either weekly or monthly. Typically each client data feed has claims, items, member, provider and group policy data. The Inventory Variance Detection System (described in next section) monitors every new batch, statistically verifies changes in model variables and if a severe change is detected, generates an operations alert report so that appropriate action can be taken at the auditing floor. Also this system sends feedback to the model manager for model monitoring, model tweaking and even model retiring based on the consistency and severity of the changes detected.

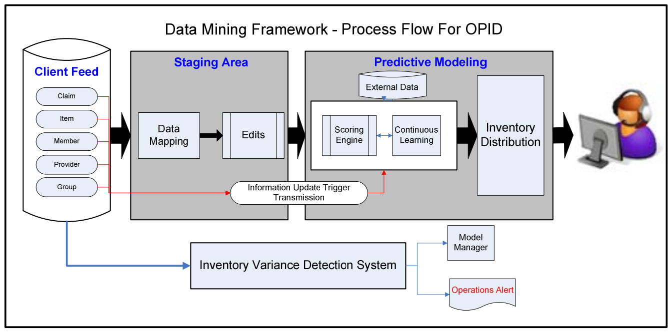

Figure 2: Data Mining Framework – Process Flow

The pre-define SQL filters or edits reduce the size of the client feed and create potentials which are placed into various work pools. The scoring engine scores the potentials in real time and a probability of overpayment is assigned to each potential. Depending on the pool definition, potentials could be assigned into treatment groups based on an inventory distribution algorithm which combines the probability of overpayment with its dollar value. The potentials are then rank-ordered by either the probability of overpayment or the treatment score and distributed to the auditing staff.

Predictive models built for certain pools use information about the insured individual that updates monthly. A re-scoring mechanism has been put in place that activates whenever an update flag is triggered. In this case, the potentials' scores would change depending on more current information.

A small subset of pools experience business driven changes which are continuously detected and relayed to the model manager by the IVDS. Subsequently any static model for such pools would show a steady drift since it was built on samples that did not completely represent the population. The flexibility of our data mining framework has enabled us to experiment with a continuous learning loop for such pools. As new data accumulates, it is used to re-estimate the model coefficients on a floating window of a parametric width. We discuss this experimental methodology in a later section.

Methodology

We will focus on major techniques implemented in the four data mining framework modules: association rule discovery, Inventory Variance Detection System, predictive modeling and scoring, and active learning.

Association Rule Discovery for Edits

Association rule discovery or association mining is the process of identifying medical claim features that appear frequently together [5]. Cluster analysis may be employed to split the historical client data feed into meaningful subgroups. Association mining is then applied to each subgroup. Each acquired relationship (or rule) between features is then examined by the business practitioners to determine its validity.

The antecedence – consequence factor of association is not important in this case, just the fact that certain elements appear together is valuable. The association rules selected for consideration are based on a significant observation count and, therefore, a significant support. The confidence is not used as a measure because the antecedence is not relevant.

As part of the edit enhancement, we have analyzed 20,000 claims that were identified as overpaid by a health insurance carrier itself, prior to the claims being fed into the framework. The claims were split into four groups based on the overpayment type and association mining was performed to identify claim features primarily responsible for the overpayments. Due to a moderate sample size, the rule selection criterion was based on support of 5% or more. Between twenty to thirty rules were identified in each group with the maximum rule support of 20%. These rules were then filtered according to business logic they proposed and the most significant were put for edit development.

As an example, a rule combining a procedure involving treatment of speech disorders performed on a child with a particular type of insurance product was among the top rules selected for edit development. With a strong support of

6.27%, a business case was made that identified a clarification in the contractual relationship for these services when a particular insurance product is used. The following is a sample of rules derived via association mining:

Table 1: Association Rule Discovery Sample

Rule	Support
single & Product A ==> spouse & Other Insurance	19.088
multiple ==> child & Provider Type 1	10.541
Procedure Code 1 ==> multiple & N & Product A	6.5527
Procedure Code 2 & Provider Type 1 & Product A ==> multiple & child & N	6.2678

Note, the confidence values are omitted as irrelevant. The variables present describe various details of the claim, such as the relationship of the patient to the insured, the insurance plan, inpatient / outpatient information, particular procedure and group codes.

Inventory Variance Detection System

The Inventory Variance Detection System has been developed in order to monitor swift changes in the process and also to detect outliers. Various change detection models based on deviation from the mean of a distribution have been proposed for large transactional data [3]. However, all such models assume that the continuous variables are normally distributed. We utilize process control methodology where we use the residual from the mean as the statistic for change detection since the continuous variables in our scenario may not necessarily follow a normal distribution. The “noise” around the mean of the variable can be safely assumed to be normally distributed and so the values of the changes can be compared against historical standard deviations.

A base table is created that contains statistics (such as the mean and the median) for the continuous predictors grouped by a certain time-frame. For categorical predictors, a base table containing statistics for historical proportions of each category is created. This approach allows us to treat all input variables as continuous.

The IVDS keeps track of all changes so that it is possible to distinguish between an outlier and a definite change/shift in the claim population which in turn is a result of process change on the client’s end. We define a process change if a severe change is detected for a large number of predictor variables, followed by a moderate and consequently mild change. An outlier is detected if a severe change is experienced by a single variable, followed by a change of the same type and no changes thereafter, see Figure 3. If any sudden change is detected, it is classified into “mild”, “medium”, and “severe” depending on the number of standard deviations away from the historical mean, and the model manager is notified so that appropriate action can be taken.

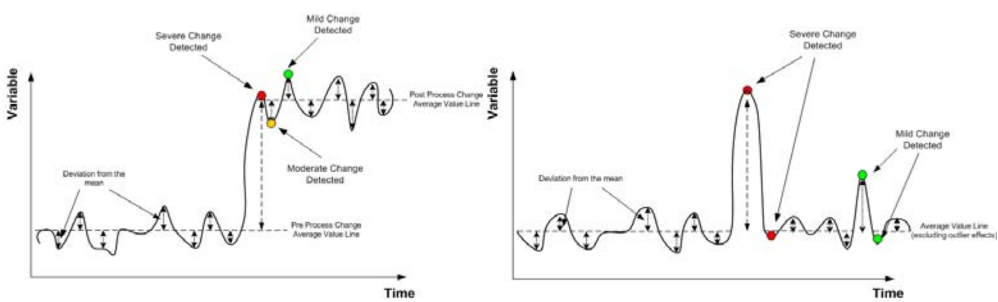

Figure 3: Process Change & Outlier Detection

Predictive Modeling and Scoring

The predominant form of predictive modeling used for the binary target (overpayment vs. non-overpayment) is the logistic regression. Since some of the fields are manually keyed in at the provider billing office, such as diagnosis (ICD-9) codes or the procedure codes, each code needs to be checked against a database for correctness. Also, missing data is handled by various imputation techniques which depend on MNAR or MCAR principles. In insurance, information is usually missing due to a systematic process error, hence it is mostly MNAR. Therefore, useful information is stored in the missing indicator variables.

Due to inconsistency in the data, the joins with demographic information for the patient, insured, provider, and group tend to get cumbersome. For instance, each healthcare client uses their own provider naming/coding system. We use various name standardization algorithms, such as the ones described in [7], to perform accurate joins, in order to minimize missing records. Initial model runs, as well as the Kramer's V tests are used to identify significant variables. We also use standard data mining techniques such as feature selection using clustering and categorical compression using clustering based on Chi-Square tests as well as initial model runs as the final steps of the data preparation.

An interesting piece while developing a predictive model for OPID is the interaction between features. An example scenario is when hospitals over-charge insurance carriers for services of medical professionals, when only one would have been sufficient to perform a procedure. The edits to catch such claims may be very rudimentary, essentially flagging all claims with procedure codes of a certain type. This leads to a lot of false positives and waste of valuable resources auditing correctly priced claims. The predictive model used to score claims in this case pays special attention to the interaction between diagnosis codes (ICD-9) and the procedure codes (CPT) to statistically estimate the need of a second surgeon. The task can be pretty daunting due to increase in feature size. For example, 15,000 possible ICD-9 codes and over 5,000 CPT codes for surgery. A number of techniques on handling and rolling up ICD-9 and CPT codes given in [4] help in reducing the dimensionality of the feature space.

In addition to feature interaction, the modeling process involves investigation of nonlinear dependencies within the data. In addition to allowing polynomial terms in the logistic regression and using large p-values (greater then 0.05) to reject such terms, we use the empirical logits plots. The empirical logits are plotted versus the binned variable in question of nonlinear influence on the target. The graph will show whether nonlinearities exist.

Once the model is complete, it is measured for goodness-of-fit on a recent sample. Since the insurance data is volatile, it is critical to perform this test. If the Kolmogorov-Smirnov statistic remains in acceptable range and the decile distribution remains even from training to the test samples then the model passes the goodness-of-fit test. For instance, consider Table 2. The best model is the Logistic Regression with interactions since its test KS statistic remains close to its train value and the decile distribution remains even. Once the model selection is complete, each historical account is assigned a probability of overpayment. The accounts are then rank ordered by the probability and split into deciles (ten equal buckets). The score code developed as a result of the predictive modeling effort is used by the scoring engine to score new claims in real time.

Table 2: Model Selection

	KS Stat		**Decile Distribution**	
Model	**Train**	**Test**	**Train**	**Test**
Logistic Regression	35.1	37.2	even	top heavy
Logistic Regression with Interactions	37.25	37.42	even	even
Logistic Regression with polynomial terms	40.12	32.5	even	bottom heavy

Continuous Learning

Frequent changes in auditing rules and the volume of claims processed in a given time period result in highly volatile pools of inventory. Hence, any historical information available is rarely representative of the current situation. To deal with this problem, we have utilized an active learning algorithm.

A "master model" is built on all available historical information which uses all data points since last known change. As sufficient new data is transferred to history, a new model on a smaller data set is built – aptly named the slave model. This process repeats for a parametrically defined number of time periods. As a result, a master model receives votes from the slave models and a decision on probability is made by averaging out the predictions. The master model is eventually replaced by a new master as the existing master model goodness-of-fit measures decline by a predefined epsilon. As a result, the process incorporates inventory changes into predictions, making the overall system of models more robust.

Business Impact

Improvements to the overpayment identification process by the newly deployed data mining framework have led to increase in dollars identified as overpayment. Consequently more dollars have been recovered from health providers. In this section we provide a few examples of the "before and after" situations. A small-size client (less than 3 million insured lives) with average overpayment dollars identified as the coordination of benefits of about $50K per month has seen an increase in identified overpayments to $600K per month. With a 50% recovery rate, the annual cost savings for this client went from $300K to $3.6M.

Another mid-size client (over 10 million insured lives) in a particular potential overpayment pool dealing with incorrectly paid claims due to services performed by an assistant surgeon.

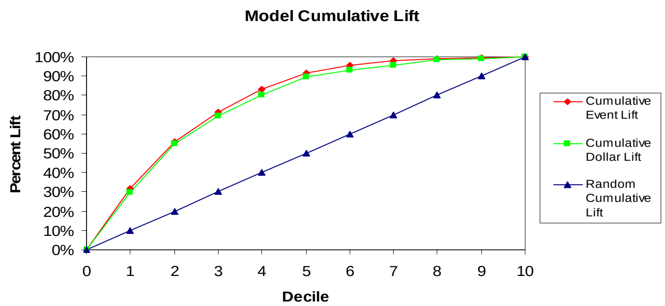

Figure 4: Model Performance

Figure 4 provides the standard cumulative lift chart that gauges logistic regression performance. Over 90% of overpayment events and overpaid dollars are acquired through working half of the inventory. The model deployed in this pool successfully decreased false positives by a factor of 3. This allowed a faster inventory penetration and more efficient allocation of audit resources.

Future Work

This paper establishes a data mining framework for overpayment identification. The real cost savings for health insurance companies occurs when these overpayments are recovered from the healthcare provider (or the insured). An extension of the OPID data mining framework can be applied to increase recoveries. For this, we need to consider two major targets of analysis: a healthcare service instance (which could consist of multiple claims) and the entity that received benefits for this service.

A two-stage modeling effort can be utilized to address this problem. Today, model development is focused on predicting the probability of recovery for a particular service instance. The data used to build the model consists of claim information for overpayments identified via the data mining framework, client-identified overpayments, and external data with geo-demographic attributes related to place of service.

Acknowledgements

We thank Robert Steidle and Curt Becker for their vision and support in building this framework around existing OPID processes. Gediminas Murauskas has given us many helpful suggestions and ideas. Many others have contributed to the success of this project, particularly Jeremy Wortz, Yi Peng, Dave Anderson and James Patterson.

References

1. National Coalition for Healthcare. http://www.nchc.org/facts/cost.shtml
2. Illinois Department of Financial & Professional Regulation, http://www.idfpr.com/DOI/HealthInsurance/coord_benefits.asp
3. Curry, C., Grossman, R., Locke, D., Vejcik., S., and Bugaijski, J. 2007 Detecting Changes in Large Data Sets of Payment Card Data: A Case Study. In Proceedings of ACM SIGKDD 2007, 1018-1022.
4. Scerbo, M., Dickstein, C., and Wilson, A., 2005 Health Care Data and SAS, SAS Publishing.
5. Delgado, M., Sanchez, D., Martin-Bautista, M., Vila, M., 2001, Mining Association Rules With Improved Semantics In Medical Databases, Artificial Intelligence in Medicine, Vol. 21, Issue 1-3, 241-245
6. Hart, M. 2002 Statistical Process Control for Healthcare, Brooks/Cole Pub Co.
7. Loshin, D. 2008 The Simplicity and Complexity of Person Name Standardization, http://www.b-eye-network.com/view/7450