Fraud detection models

This is a document describing the ten fraud detection models incorporated in the Indalyz demo version of the algorithm. Every model is explained in words, and will result in a model score and potentially a corresponding monetary loss that are defined here as well. The last section describes how these are used to construct the total scores that define the ranking in the results screen of the online GUI.

Introduction

The power of the algorithm lies in three main ingredients:

1. State-of-the-art statistical and machine learning techniques are used in the underlying models.
2. Many different kinds of models are used simultaneously, such the fraudsters who manage to stay under the radar in each aspect will still surface when models are combined.
3. All ingredients are fully customizable by the user: weights of one model with respect to one another, the relative importance of severity and monetary losses as well as many model-specific parameters can be provided by (and experimented with) by the user. It is also easy to include more, even client-specific, models to the algorithm.

The model described here is merely a demonstration of what kind of things are possible, and what a fully deployed client implementation might look like. The first ten sections describe the models one by one, after which an explanation of the aggregation of scores is given.

Cost per member

The model “Cost per member” looks for outliers in cost per member among all providers. Typically, this quantity follows a log-normal distribution among providers whose mix of procedures is similar. Different medical specialisms may correspond to very different distributions, but within one specialism, or reference group, such a distribution is typical. This model gives the user the freedom to group providers by a known specialism, by a peer group that is determined by the algorithm (see model: billing pattern), based on the mix of procedures that is billed or over the whole group of providers, without taking specialisms into account. This choice if left to the user in the settings. What defines an outlier is the other user-defined parameter. The choices here are to use everyone above a fixed percentile (90^{th} or 95^{th}) or to find statistical outliers based on the interquartile distance of the actual distribution. These are all non-parametric methods, which is safe because it requires no prior knowledge about the data. Other non-parametric methods for outlier detection can also be used, as well as parametric definitions (that depend on the assumed functional form of the distribution function), but this will require additional coding in the model.
The score of this model is defined as follows: Using the method picked by the user, an upper limit is determined for the cost per member (either per specialism, per determined reference group, or overall, as set by the user). Everybody above this value will be flagged. The score is such that the median cost per member of everybody who is flagged will result in a score of 1. Others are proportionally scored (somebody just above the upper limit will get a score close to zero, providers with a cost per member that is two times further from the upper limit than the median will get a score of two, etc.).

The monetary value is determined by (cost_per_member – upper_limit) * number_of_members and can therefore be regarded as an estimate of the minimum amount of money that is charged too much.

Billing pattern

In this model, which is also the basis for the "determined peer groups" used in the previous models, outliers are identified in their billing pattern, with which we mean the mix of procedure codes they bill. In the data set, 20 different procedures codes are present, so the fraction of the revenue of providers for the twenty different codes spans a 20-dimensional space. A principal component analysis is used to reduces this space to 3 dimensions. In this three dimensional space (which is normalized, such that distances along the three axes mean the same thing), a density based clustering algorithm is used. Providers at too low a local density (i.e. far away from agglomerations of providers) are flagged as outliers.
The score is defined as the distance to the nearest cluster center, relative to the cluster size. Monetary losses are set to zero as it is unclear how to define these.

Treatments on weekends and holidays

Here, the user can decide whether treatments on such days are suspect every time they happen, or whether they are only suspect for providers who do them more than their peers. In case of the former, all such treatments are counted and the model score is the number of treatments on such days. It is again normalized such that the median number of treatments on such days gets score one. In the latter case the distribution of the fraction of all treatments performed on such days is determined and those providers that have more than twice the standard deviation of that distribution more than the average fraction are flagged.
In the former case, all money billed for treatments on holidays and in weekends is added up to form the total monetary loss. In the latter it is the fraction of the money billed for treatments on such days corresponding to the fraction of treatments on such days that was above the limit.

Increasing revenue

When providers just do their job and nothing drastically changes in either their practice or the market share of the insurance company, a steady revenue is expected. This model flags providers for which the revenue per unit time goes up during the year. This can be either with a step function, or as a general increase over time. The method used is to compare the increase of cumulative revenue to the average revenue over the year. If revenue increases over a year, it is higher per unit time at the end than at the start of the year. The ratio of average cumulative revenue to average revenue aggregated over the whole year is roughly normally distributed, with outliers at the high end for providers who underwent an increase in revenue. Everyone above 2 standard deviations above the average is flagged. The score is such that the median is 1 and is linearly proprtional with the difference between the actual ratio and the average ratio plus two standard deviations. Monetary losses are set to zero, as it is undefined what the yearly revenue should be without a steady level.

Seasonality

Patients go on holidays. Therefore, a dip in the summer months for all providers is expected, which is compensated for by a slight increase in revenue in September and October. The pattern that is expected is learned from the whole population, where the fraction of yearly revenue in blocks of two months is determined. The deviation from this seasonal pattern is corrected for statistical fluctuations, based on the volume per provider. If a provider deviates by more than the expected fluctuation, the provider is flagged. The normalization is based on the accepted amount of fluctuation, not on the median of people who are flagged. Monetary losses are set to zero, as it is not clear how to define them.

Combination of procedures

Some combinations of procedures are forbidden. This is an example of a very simple, black and white business rule model. Here the choice is made to define one forbidden combination. Everybody who did that combination is flagged. The score is such that the median is 1 and scale linearly with the number of forbidden combinations. The monetary loss is equal to the number of forbidden combinations times the rate of the cheapest procedure, to get an estimate for the minimum corresponding loss.

The ratio of expensive to cheap versions of the same treatment

For some procedures, there is a simple, cheap version of the treatment, and a more expensive one. The ratio of the two is something that should depend on patient population, and maybe the specialism of the provider. The user gets to determine whether this is determined over all providers, or per specialism. Another choice for th user is to make the score dependent on the number ratio of the cheap and expensive version, or of the revenue for both (so the second is a rate weighted version of the first) and whether everybody above the 90th percentile or only statistical outliers are flagged.
The scores are then correspondingly based on the number ratio or revenue ratio, such that the median outlier gets score 1 and the rest linearly dependent. Monetary values are calculated from the deviating ratios and the rate difference for the cheap and expensive treatments.

Freely billed rates

Some procedures do not have set rates, but are billed freely. The amount billed per freely rated procedure is investigated and an outlier detection is performed. Again, the user can decide whether scoring is done per reference group or over all providers combined. It can also be done per procedure code separately (there may be differences in procedures with different codes and their billed rates) or overall. As usual, one can chose between flagging only statistical outliers, or everybody above the 90th percentile. The scores as well s the monetary amounts are as always: the median score is one, the rest linearly scaled and the monetary loss is defined as the amount of money per procedure higher than the limit, multiplied by the number of such procedures.

Periodic treatments billed too often

There is a procedure code set that is defined as periodic treatment. These should not be given to the same patient more than twice a year. The extra treatments are simply counted and the monetary amounts just the sum of all "extra" treatments. The score is the number of extra treatments divided by the median of that. There is no correction made for market share or practice size.

Billing cycle

For bills submitted by providers (all member bills are exclude), one expects a fairly regular schedule of billing. Some providers will send a bill every week, others every month, quarter, etc. These patterns are detected for all providers and providers who seem to have a regular scheme, but one or more deviations from this pattern are flagged. The user decides whether:

- it is a flag for every provider who sends in more than 1 bill in any given week (if they have at least 40 different weeks in which they bill),
- or for providers with sufficient volume (> median volume) and sufficiently many bills (bills in more than 10 different weeks), flag those with a standard deviation in periods between the bills that are higher than 1 day,
- or to detect providers for whom, in series of billing dates that are regular, the standard deviation in time between bills reduces significantly when "odd dates" are left out.

Scores and monetary losses are based on, in the same order:

- The number of extra bills, normalized to the median and the number of extra bills times the mean amount per bill.
- The standard deviation of days in between bills, normalized to the median. Monetary losses are set to 0, as it is unclear which bills are suspect.
- The number of days that has the desired effect on the statistics of the billing cycle, normalized to the median. The monetary loss is total billed sum of those same days.

Scoring

All scores and monetary values are combined into a total score, as illustrated in the graphic on the next page.
Every model that is turned on will run and results in a model score and a monetary loss (if that is meaningful for the model), see below for more details. Every model also has an assigned weight, set by the user.

Out of the list of model scores, three numbers are calculated per provider:

1. A severity score, which is the weighted average of the model scores (with the weight being the model weight set by the user).
2. The sum of all monetary amounts of the models the provider scored on.
3. The number of different models on which a provider scored.

The total score is a number that can be at most 100. 40 points are given for severity, 40 for monetary loss and 20 for the number of flags. The provider with the highest severity score gets the 40 points somebody with an f times lower severity gets 40/f points for that. The same works for the monetary loss and the number of flags. If one provider would have the highest severity, and the highest monetary loss and the highest number of flags, that provider would get 100. All other situations result in lower scores.

Within the model the score is created in a way that is different for each model, and it may depend on the settings of the model. I can explain all these in detail, but I do not know how sensible that would be for a demo model, made up by me, rather than clients. In essence: of everybody who scores, the one who's severity (which is "how much do you deviate on this model") is equal to the median of all who score on that model will have model score 1. Others are scaled linearly (so providers who are one quarter or twice as bad, get model score 0.25 and 2 respectively). The monetary loss is also something that depends on the model in question. It should give a fairly precise indication of how much money is involved with the deviation of the provider in question.