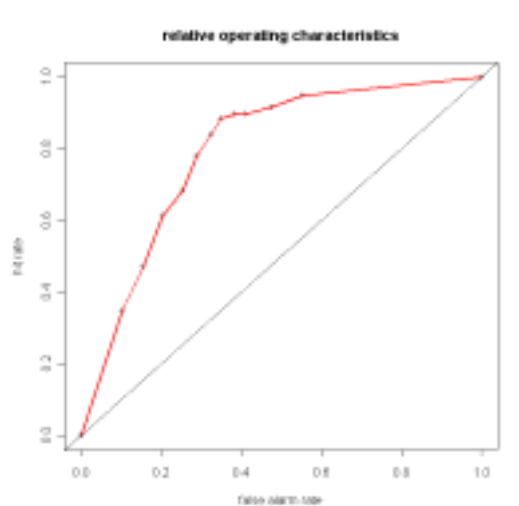


Automatic Fraud Detection

MIKAN

Dr. Martin Fischer
27 rue Sébastopol
12270 B.P. 98802 Nouméa Cedex
Nouvelle Calédonie



Summary

Two self-learning methods for automatic fraud detection are presented. In both cases a classification into two categories (correct YES/NO) is carried out. Both methods show good forecast results. In a test data set hit rates are above 80% with false alarm rates ranging from 15% to 23%.

Introduction

Actual situation

So called “expert systems” are frequently used in the field of automatic fraud detection. A software package emulates the knowledge of experienced people who control such data sets manually. The rules to be applied are developed empirically and have to be coded manually. Thus a software package simulates human experience.

Problem

The need to code the rules manually is a major drawback of such systems. Three main problems are encountered.

1. Since the system is based on human experience it is most likely biased. The subjective criteria of experienced people are emulated which may lead to sub-optimal filter systems.
2. If legislation changes or if only the format of the declaration is modified the expert system has to be adapted. Since the rules are coded manually this requires a substantial amount of update work with the related delays and costs.
3. A severe security problem occurs if unauthorised persons get access to the expert system. In that case it can be used to “tune” false declarations.

Solution

Adaptive self-learning systems may overcome the drawbacks described above. Such systems extract optimal rules automatically from so called training data sets. A number of example data sets including the (known) result is fed into the system and the software “learns” the hidden rules to separate the data into a certain number of classes. The needed number of examples depends very much on the complexity of the problem and of the number of input and output parameters. Typically it is of the order 50 to 500 in difficult cases a few thousand.

Main advantages of self learning adaptive systems are speed, flexibility, and objectivity. The learning cycle requires typically a few minutes of CPU time on a powerful PC.

It is almost impossible for unauthorised people to misuse such a system for targeted fraud. Since such methods develop their capabilities only in conjunction with examples, two components – the software + examples – are needed to use it. This makes it much easier to protect such systems against misuse and hence improves security considerably.

Adaptive systems may permanently adjusted to the actual needs. New findings or constraints can be immediately integrated into the assessment process without the need of additional coding. It is sufficient to present another set of typical examples to adjust the system.

Study

In this study we present results, obtained with two different methods. The first system is based on a non-linear regression algorithm, whereas the second uses a partitioning method.

As test data 994 anonymous declarations were used, of which 255 were suspicious. Each data record contains 71 parameters. The meaning of the parameters was for security reasons hidden.

To train the adaptive systems we used 100 suspicious and 100 non-suspicious data records. The remaining 794 records – containing 155 suspicious and 639 non-suspicious records – were used for verification. It is important to note that we did not mix training- and verification data sets. This means that the verification data sets were completely independent from the records that were used for training. In this way we avoid the problem of so called “artificial skill”. The results, presented in the following sections therefore represent realistically a true forecast situation.

Results

As already mentioned we tested two different systems. In the following we show the results for each method separately.

Non-linear Regression Method

We use a non-linear adaptive regression method to model the relation between the 71 input parameters and the two possible classification results (suspicious YES/NO). To find this relationship the above mentioned 200 training data sets were used. Once a model is found this model is used to classify the remaining 794 records of the verification data set. The predicted classes are compared with the actual results.

The output of regression method is a continuous number between 0 and 1. Results close to 0 indicate correct declarations, whereas predictions close to 1 point to a high probability for fraud. The user has to define a threshold to separate between correct and incorrect declarations. Records for which the predictions yields a value higher than this threshold are classified as suspicious, those that remain below the threshold are assumed to be correct.

Depending on the chosen threshold one obtains a certain ratio between hit-rate (suspicious declarations are correctly classified as being not correct) and false alarm rate (a correct declaration is classified as suspicious).

Plotting those rates against each other for a range of different thresholds yields the so called “relative operating characteristics” (ROC) which is shown below.

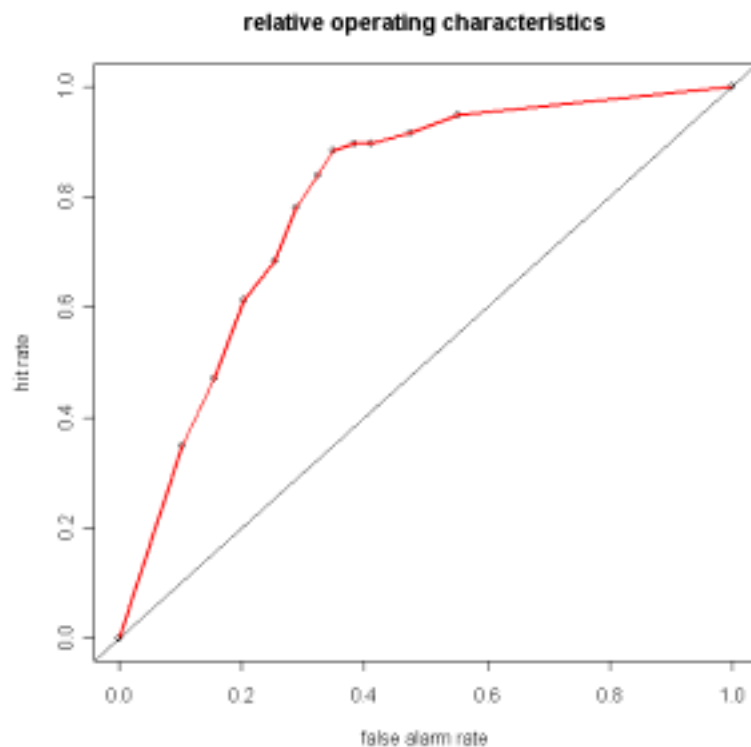


Figure 1: ROC curve for the regression method.

The ROC-curve contains all information about the quality of the classification system. The diagonal represents a classification system based on coincidence. The further the actual ROC curve is above the diagonal, the higher is the quality. The ROC curve is the starting point for risk management methods that may be used to optimise decision making processes.

In the presented example we obtain – for instance – for a threshold of 0.5 a hit rate of 84% and a false alarm rate of 32%. This means that a false declaration is recognised as suspicious with a probability of 84%. The practical meaning of hit rate and false alarm rate are explained in the following example.

Example I

We assume a mean rate of false declarations of 10%, that means out of 100 declarations 10 are on average false. In order to find for sure all 10 falsifications one has to check all declarations, i.e. 100 detailed checks have to be carried out. If instead of that only those are checked, which were classified as suspicious we obtain the following. With the above hit rate and false alarm rate, on average the system will attribute 37 times the label “suspicious”. Among those we may expect to find 8.4 actual fakes. This means we may reduce the checking expenses to by 63%, for the price of not finding 1.6 falsifications out of 10. A higher hit rate can be achieved by choosing a higher threshold, however, this will also yield a higher false alarm rate. With a threshold of 0.2 we obtain for the present system a hit rate of 90% and a false alarm rate of 41%, which will give 46 times the attribute

“suspicious”. In this case the checking expenses can be reduced by 54% and 9 out of 10 falsifications are found on average.

Partitioning Algorithm

The partitioning algorithms determines iteratively optimal criteria to distinguish between different classes. However, those algorithms have difficulties dealing with a large number of input parameters. Therefore “intelligent” pre-processing is necessary to identify the most relevant input parameters and to limit the analysis to those. In the present case we used the regression method, presented in the previous section, to pre-select the relevant input parameters. This means that the results presented in this sections are obtained with a two-tier approach.

Out of the 71 input parameters 6 were identified to be relevant. The partitioning method was applied to those 6 parameters only.

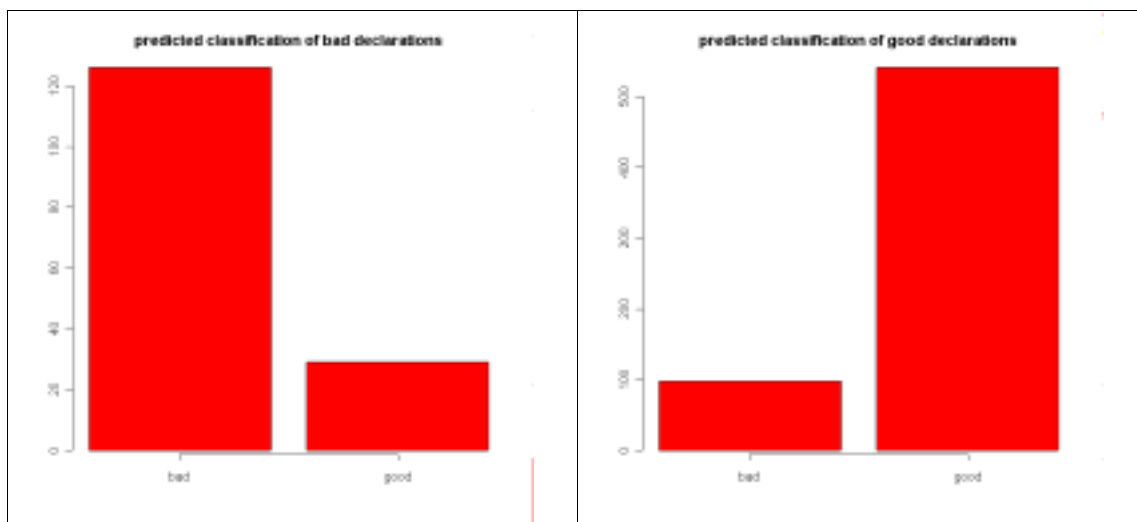


Figure 2: Results obtained with the partitioning algorithm.

In the left part of Fig.2 we show the classification results for those records that were known to be false. Out of 155 false data records 126 were correctly classified as suspicious - 29 falsifications were not found. This yields a hit rate of 81%. In the right part of Fig.2 we show the classification results for the correct data sets. In that case the system gave in 98 out of 639 cases erroneously the attribute “suspicious”. 541 declarations were correctly classified. This yields a false alarm rate of 15%.

Example II

With the same assumptions as in example I the partitioning method gives the following results. In a 100 cases the method will attribute 21 declarations as being suspicious. If only those are checked carefully on will find about 8 falsifications. This means one may reduce the checking expenses by about 80% but on misses 2 out of 10 falsifications.

Conclusion

Both methods have proved that for the present case they can be used for automatic fraud detection. The ration of hit rate to false alarm rate ranges from 85% to 32% and 81% to 15% for the regression and the partitioning algorithm, respectively.

With both methods the expenses for manual checking can be reduced substantially without a dramatic drop in detection-rate. Further optimisation can be achieved with risk management methods (not shown in this article).

Both methods are very flexible and can be applied to numerous classification problems. In both cases unbiased classifications were carried out.