**A prescription fraud detection model – Summary**

Authors : Karca Duru Aral , Halil Altay Güvenir, I˙hsan Sabuncuog˘lu, Ahmet Ruchan Akar

Article history: Received 23 November 2010

Received in revised form 12 September 2011

Accepted 13 September 2011

Abstract :

Aimed to develop a model for detecting cases of prescription fraud and test it on real world data from a large multi-center medical prescription database (Turkey). A novel distance based on data-mining approach for assessing the fraudulent risk of prescriptions.

The results obtained from experiments reveal that the proposed model works considerably well with a true positive rate of 77.4% and a false positive rate of 6% for the fraudulent medical prescriptions.

Concludes that incorporating such a system in health authorities, social security agencies and insurance companies would improve efficiency of internal review to ensure compliance with the law, and radically decrease human-expert auditing costs.

1. Introduction:

Skipping introduction summary as it’s all about how Turkey’s SSA in turkey provides healthcare to citizens and how the manual fraud detection out of a random sample is done. Which is somewhat similar & different to how DSGI does it. However, the problem the paper talks about is exactly the problem we are trying to solve for DSGI.

2. Related Work : (literature and results of Supervised, Unsupervised & Hybrid models on other industry fraud)

2.1 Supervised approaches

Explained neural networks (back propagation to train) ---> output(risk score)

Other researchers tried regression, decision tree, K-Nearest Neighbors, Neural Networks, statistical modelling, stepwise selection of predictors, probit regression

When incorporated decision tree(rule induction, case-based), dynamically updates in response to changes in the manufacturer’s conditions. Enabling a flexible and higher quality decisions, the system is tested on simulation runs which reveals that the proposed model outperforms the existing algorithms in the literature.

2.2 Un-supervised approaches

Very little mention on un-supervised approach in this section, maybe because they implemented this approach and talked in depth in later sections. For now paper just mentions two models, which are SmartSifter (medical insurance domain) & PCA of RIDIT scores (automobile insurance)

2.3 Hybrid approaches

Supervised hybrids:

Supervised neural networks, Bayesian networks, and decision trees are the methodologies mostly used to create hybrids. Combined naive Bayes, C4.5, CART, and RIPPER classifiers.

The results give better efficiency on credit card transactions.

Un-supervised hybrids:

Detection of fraudulent international calls. Tried *k*-means is employed for cluster detection, C4.5 is used for decision tree rule induction, and domain knowledge, then statistical summaries and visualization tools are utilized for rule evaluation.

Describes another medical claim fraud/abuse detection system based on data mining used by a Chilean private health insurance company. The proposed detection system employs multilayer perceptron neural networks (MLP)

3. Proposed approach:

3.1. Data structure :

Graphical user interface

Description automatically generated

3.2. Methodology :

Diagram

Description automatically generated

4. Computational results

The approach proposed here is able to handle both categorical and ordered features. The output of the system is easy to understand and interpret by human users. Besides, the system can learn and process accordingly as the input data shifts. Finally, its core methodology is adoptable to many other areas in health care and possibly in other industries.

Given the performance measurements with a true positive rate of 77.4% and a false positive rate of 6%, we can conclude that the proposed system works reasonably well for the prescription fraud detection problem. Nonetheless, further refinement of the tool would require scaling the risk outputs across all domains. This would mean that incorporating different parameters for different domains would lead to the same risk measurements across all domains.

5. Concluding remarks and further research direction

Our methodology proposes dividing up the 6 dimensional features’ domain into several 2 dimensional sub-domains considering the interaction levels between the features. The methodology consists of populating incidence matrices for each of the above domains and then incorporating a distance based data-mining approach. The risk metrics employed in this data-mining approach return risk measures for each of the domains mentioned above. This risk measure is scaled to be between 0 and 1, in order to give a straightforward definition of the risk level. For each of the domains, the user can specify thresholds. That way, the program alarms for only those prescriptions with risk levels higher than the thresholds. The automated fraud detection methodology gives considerably compatible results with the human expert auditing. The system is flexible enough for an integrated on-line/on-time user interface, and its on-line incorporation is computationally inexpensive, it presents a novel and easy way to keep track of health care transactions in incidence matrices for auditing.