Process Mining in Health Care for the Detection of Fraudulent Data

C.Rohith Bhat, *Research Scholar, Dr.MGR Educational and Research Institute, Maduravoyal, Chennai, India*

rohithbhat2000@gmail.com

Dr.S.Ramamoorthy, *Professor, Dr.MGR Educational and Research Institute, Maduravoyal, Chennai, India*

srm24071959@yahoo.com

Abstract: Process mining has a purpose of extracting process-oriented knowledge from event logs extracted from information systems. It is a new research discipline that has evolved significantly since the early work on idealistic process logs. Over the last years, process mining prototypes have incorporated elements from semantics and data mining and targeted visualization techniques that are more user-friendly to business experts and process owners. However, there were a few studies showing the applications of process mining in manufacturing industry. With such an intensive need for health insurances, however, health care service provider's fraudulent behavior has become a serious problem. In this research, we propose a process data mining framework that utilizes the concept of clinical pathways to facilitate automatic and systematic construction of an adaptable and extensible detection model. We investigated the mining of frequent patterns from clinical instances and the selection of features that have higher discrimination power and also proposed approaches have been evaluated objectively by a real-world data set. The empirical experiments show that our detection model is efficient and capable of identifying some fraudulent and abusive cases that are not detected by a manually constructed detection model.

KEYWORDS: PROCESS MINING, HEALTHCARE, FRAUDULENT DATA

1. INTRODUCTION

Healthcare has become a major focus of concern and even a political, social and economic issue in modern society. The medical expenditure required to meet public demand for high quality and high-technology services is substantial. This phenomenon is likely to become more widespread and more intense due to the increasing average lifespan and decreasing birth rates of humans in many societies. People rely on health insurance systems, which are either sponsored by governments or managed by the private sector, to share the high healthcare costs.

Fraud is the abuse of a profit organization's system without necessarily leading to direct legal consequences [1]. Detecting healthcare fraud and abuse, however, needs intensive medical knowledge. Many health insurance systems rely on human experts to manually review insurance claims and identify suspicious ones. Most of the computer systems that are intended to help detect undesirable behavior require human experts to identify a set of features so as to develop the core of detection models. This results in both system development and claim reviewing being time-consuming, especially for the large government-sponsored national insurance programs in countries such as France, Australia, and Taiwan. The problems have been reported for the health insurance programs of other developed countries [2].

In order to assure the batter operation of a health care insurance system, fraud detection mechanisms are imperative, but highly specialized domain knowledge is required. Furthermore, well-designed detection policies, able to adapt to new trends acting simultaneously as prevention measures, have to be considered. Data mining which is part of an iterative process called knowledge discovery in databases (KDD) [3] [4] can assist to extract this knowledge automatically. It has allowed better direction and use of health care fraud detection and investigative resources by recognizing and quantifying the underlying attributes of fraudulent claims, fraudulent providers, and fraudulent beneficiaries [5]. Automatic fraud detection helps to reduce the manual parts of a fraud screening/checking process becoming one of the most established industry/government data mining applications [6].

In this research, we propose a process-mining framework that utilizes the concept of clinical pathways to facilitate the automatic and systematic construction of an adaptable and extensible detection model. We take a data-centric point of view and consider healthcare fraud and abuse detection as a data analysis process. The theme of our approach is to apply Process-mining techniques to gathered clinical-instance data to construct a model that distinguishes fraudulent behaviors from normal activities. This automatic approach eliminates the need to manually analyze and encode behavior patterns, as well as the guesswork in selecting statistics measures. The proposed framework is evaluated via real-world data to demonstrate its efficiency and accuracy.

II. PROBLEM STATEMENT

The processing of health insurance claims involves three parties: service providers, insurance subscribers, and insurance carriers. The National Health Care Anti-Fraud Association defined healthcare fraud as 'an intentional deception or misrepresentation made by a person, or an entity, with the knowledge that the deception could result in some unauthorized benefit to him or some other entities' and healthcare abuse as 'the provider practices that are inconsistent with sound fiscal, business, or medical practices, and result in an unnecessary cost, or in reimbursement of services that are not medically necessary or that fail to meet professionally recognized standards for health care [7].

III. CHALLENGES:

This concept of clinical pathways shows great promise in detecting fraud and abuse by service providers. A care activity is very likely to be fraudulent if it orders suspiciously. For example, since physicians prefer performing simple, non-invasive tests before performing more complex, invasive tests, there is a high probability that the same set of care activities performed in a different order is fraudulent or abusive. Extensively, to accurately determine the appropriateness of a care activity performed on a particular patient, we must take into account the other activities performed on the patient. For example, while a single ambulant visit is normal, repetitive visits are problematic, especially where the average length of pathway instances is small. This observation initiates our idea that the clinical structures, including care activities and their order of execution, can be used to discriminate between normal and fraudulent cases.

A schematic of our process-mining framework is provided in Fig. 1. Generally, two sets of clinical instances, which are labelled as normal and fraudulent, serve as the input of the module for discovering structure patterns. This module produces a set of structure patterns that have occurred frequently, which then serve as features of clinical instances. Each clinical instance is considered an example that comprises a set of features and a class label (normal or fraudulent). A feature-selection module to eliminate redundant and irrelevant features further filters the resultant data set. The selected features and the data set are finally used to construct the detection model, which is performed by the induction module. The detection model is then used to detect the incoming instances that are fraudulent.

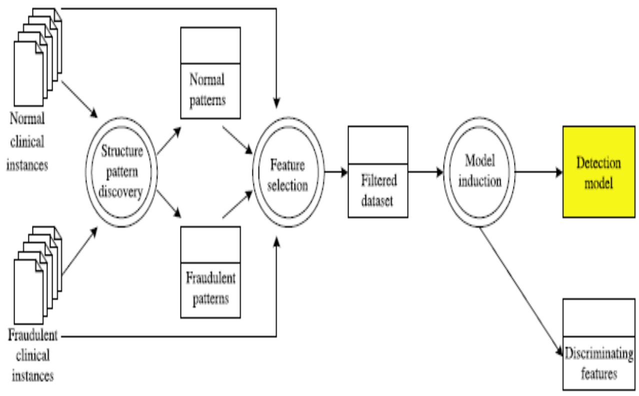

Fig. 1: The process-mining framework.

IV. STRUCTURE PATTERN DISCOVERY

The first step of the proposed framework involves extracting patterns in a way amenable to represent structures of clinical instances. In this section, we explore the entrance problem: the discovery of structure patterns. Typically, a clinical instance is a process instance comprising a set of activities, each of which is a logical unit of work performed by medical staffs. For example, a patient treatment flow may involve measuring blood pressure, examining respiration, and medicine treatment. These activities, each appearing over a temporally extended interval, may execute sequentially, concurrently, or repeatedly. For example, before giving any therapeutic intervention, diagnosis activities are usually executed to verify the condition of a patient. Also, more than one therapeutic intervention may be executed concurrently in order to increase the curative effect in some cases. As a result, if we want to extract structure patterns from clinical instances, we need to take structural characteristics of process—temporally extended intervals and various transitions—into consideration.

In this paper, we apply structure pattern mining techniques proposed in [8] [9] to identify a set of structure patterns from clinical instances.

The structure pattern discovery algorithm is sketched as follows:

MiningStructurePatterns(S: a set of clinical instances): a set of structure graphs

```
{
        Scan S to find the set TGraphSet1 of all activities with minimum support;
        nZ1;
        Repeat {
        nZnC1;
        CandidateSetnZ
        GenerateCandidateGraph(TGraphSetnK1);
```

Scan S to find a subset TGraphSetn of CandidateSetn with minimum support;
} Until TGraphSetnZ:;
Return TGraphSet1gTGraphSet2g/gTGraphSetnK1;
}

V. PATTERN FEATURE SELECTION

In our framework, frequent structure patterns discovered by the algorithm described in Section 4 are regarded as features. In practice, the number of features is often huge (usually more than 10,000). It is widely recognized that the number of features has a strong impact on the efficiency of an induction algorithm, and the inclusion of irrelevant or redundant features may degrade its accuracy. Therefore, it is imperative to reduce the feature set prior to constructing a detection model to decrease the running time of the induction algorithm and to increase the accuracy of the resultant model. This feature selection issue is addressed in this section.

Several studies have addressed the problem of feature selection. The proposed approaches fall into the following two categories: the wrapper model and the filter model. The wrapper model [10] scans through the space of feature subsets in search of the one that has the highest estimated accuracy from an induction algorithm. Specifically, the feature selection algorithm continuously interacts with the underlying induction algorithm, with the aim of choosing a subset of features that achieves the best classification result for the induction algorithm. While these methods have been shown to achieve some success on induction, they suffer from high computation cost and are not applicable to tasks with even a few hundred features.

The filter model introduces a pre-processing step prior to induction. As such, the adoption of the induction algorithm does not interfere with the selection of the feature selection algorithm. A major benefit with the filter model is that it does not need to search through the space of feature subsets as required in the wrapper models, and is therefore efficient for domains containing a large number of features. Three of the most well-known filter methods are RELIEF and the Markov blanket filter. In RELIEF, each feature is individually assigned a weight indicating its relevance to the class label, and a subset of features with the highest weights is selected. It is possible that RELIEF fails to remove redundant features, since two predictive (but highly correlated) features will both be selected. The FOCUS algorithm exhaustively searches all feature subsets in order to identify a minimal set of features that consistently label instances in the training data. This consistency criterion makes FOCUS vulnerable to noise in the training data. Moreover, searching the power set of features also makes this algorithm impractical for domains with a large number of features. A probability framework (the Markov blanket filter) for selecting an optimal subset of features has referred. Theoretically, this method eliminates a feature if it gives no additional information beyond that subsumed by a subset of remaining features (called the Markov blanket). Since finding the Markov blanket of a feature might be computational infeasible, this research resulted in an algorithm that computes an approximation to the optimal feature set.

Since the structure pattern discovery algorithm may generate a large number of structure patterns (or features), we focus our attention on the filter model due to its key advantage on computation cost. In our framework, the discovered structure patterns are regarded as features, each of which denotes whether a specific pattern is supported by an instance. Thus, each instance can be translated as a set of feature values with a class label, and our view on a translated example can be formally described as below.

FeatureSelection(T: a training set; F: a set of features; N: an integer): G: a set of features
// Suppose features in F are listed in ascending order of their sizes
{
//First Stage
GZAncestorPruning(T, F);
//Second Stage
If (jGj*O* N) {
GZ MarkovBlanketFilter(G, N);
}
Return G;
}

It is clear from the above analysis that the running time of the Markov blanket filter algorithm will increase dramatically with larger K. However, a larger K is more likely to subsume the information in the feature, thereby forming a Markov blanket. However, a larger conditioning set, as formed by larger K, may in turn fragment the training set into many small chunks, thereby reducing the accuracy of the probability and hence the cross-entropy estimates. Therefore, there is a tradeoff for setting K in terms of classification accuracy.

If a large extent of redundant information can be eliminated at the first stage, there is a high probability that a smaller conditional set will result in a satisfactory approximation. A smaller conditional set reduces the number of chunks and hence increases the accuracy of cross-entropy estimates, and the running time also decreases dramatically since the computation complexity of the second stage is exponential with K. Therefore, the combined approach is particularly suitable for our problem -a domain with a huge number of structure patterns.

Steps:

(1) Filtering out noisy data: The treatment data of each patient was regarded as an instance, and we removed instances that had missing or noisy attribute values. In this step we removed 77 instances.

(2) Identifying activities: Based on the domain knowledge provided by experts, we identified medical activities in the remaining instances. Some activities, such as examination of blood pressure, were performed routinely and thus discarded. We finally identified 127 medical activities in this step.

(3) Identifying fraudulent instances: Two gynecologists were involved in the identification of fraudulent instances. They examined all instances, among which 906 instances were judged by both gynecologists as fraudulent.

(4) Selecting normal instances: We then randomly selected 906 cases from the remaining instances that both gynecologists considered normal cases. As a result, a total 1812 instances were used in our experiments.

IV. EXPERMENTAL RESULTS

We adopted the Classification Based on Associations algorithm (CBA) as our induction method. Also, in order to evaluate the detection model, we consider two measures, sensitivity and specificity, which are often used in medical diagnosis and in the detection of fraudulent behaviour. The Sensitivity is the proportion of fraudulent cases that are identified as fraudulent by a system, and Specificity is the proportion of the normal cases that are classified as normal by the system. Clearly, a detection system is considered to have good performance if it has both high sensitivity and high specificity.

6.1. Number of features deducted

In order to construct our detection model, patterns are first discovered using the structure pattern discovery algorithm, then translated as features, and finally filtered by the feature subset selection algorithm. Fig. 2 shows the number of features selected in our model. These patterns (features) are discovered at different support thresholds, ranging from 10 to 2% at 2% decrements. Fig. 2a shows the number of initial features (discovered by the structure pattern discovery algorithm) and the number of features that pass the first stage of feature subset selection. Fig. 2b shows the number of features that are eliminated by the first stage of feature subset selection divided by the number of initial features.

As expected, the number of initial features increased as the minimum support decreased. While the number of remaining features still increased moderately as a function of support threshold, a large proportion of features is eliminated by the first stage of feature subset selection. For example, at a support threshold of 2%, an average of 30,701 features is initially discovered while only 3120 features pass the test. Further, as shown in Fig. 2b, the number of eliminated features divided by the number of initial features grows substantially as the minimum support decreases.

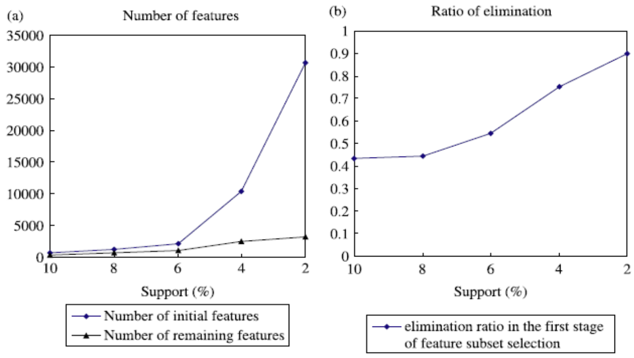

Fig 2: Effects of feature subset selection.

6.2. Prediction power with the first stage of feature subset selection.

We next investigated the sensitivity and specificity of our detection model, which are constructed by features selected by the first stage of feature selection. At support thresholds of 2–6%, because many (more than 1000) features pass the first stage of feature subset selection, we further filter features by applying the Markov blanket filter (the second stage of feature subset selection) with various blanket sizes (KZ0, 1, 2). One thousand features (NZ1000) are finally selected in these cases. Also, since the CBA is most accurate when the minimum support is 1–2% , we set the support and confidence of the CBA to 1 and 50%, respectively. The resultant sensitivity and specificity of our detection model are depicted in Fig. 3.

Fig. 3 shows that the sensitivity and specificity of the detection model increased as the support threshold decreased.

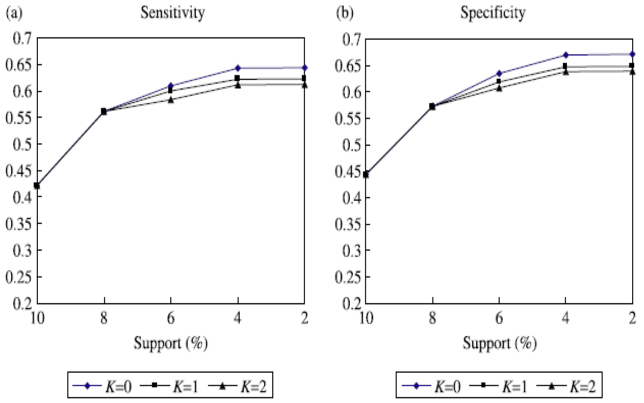

Fig 3: Sensitivity and specificity of the detection model with the first stage of feature subset selection.

This is as expected, since a lower support threshold indicates the discovery of more features and thus the provision of more information for the classification task. The best sensitivity and specificity (64 and 67%, respectively) are obtained at a support threshold of 2%. It is also worth noting that the best sensitivity and specificity are both obtained at a conditioning level of KZ 0. This demonstrates that a great extent of redundant information has been eliminated in the first stage of feature subset selection, and thus a low conditioning level (KZ0) is sufficient to further filter out correlated information.

6.3. Prediction power without the first stage of feature subset selection

We also investigated the sensitivity and specificity of our detection model in which all features were selected by a
Markov blanket filter with various conditioning settings. The settings of this experiment were the same as the previous one except for the omission of the first stage of feature selection. The sensitivity and specificity of the resultant detection model are depicted in Fig. 4.

It can be seen that the best sensitivity and specificity (60 and 64%, respectively) were both obtained at a conditioning level of KZ2. Comparison with the results shown in Fig. 4indicates that the performance of this detection model is slightly worse, which is as expected because the Markov blanket filter uses only approximations to eliminate features. Moreover, the conditioning setting (KZ2) shows that it is necessary to have a higher conditioning level to filter redundant information, resulting in a longer computation time.

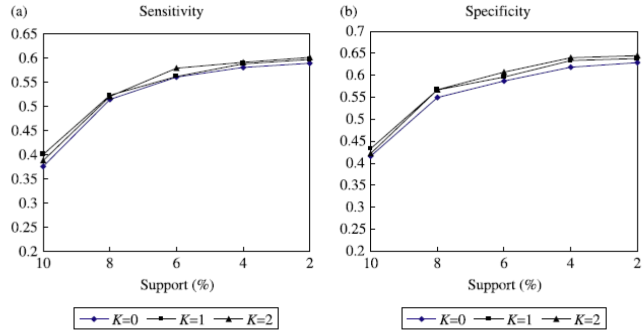

Fig 4: The sensitivity and specificity of the detection model without the first stage of feature subset selection.

6.4. Comparison of detection models

We finally compare our detection model with that proposed by Chan and Lan [11], which was designed to detect suspicious claims in the Taiwan NHI program. The resultant sensitivities and specificities of the two detection models are shown in Fig. 5. Fig. 5 clearly shows that the non-structure detection model, which mainly involves expense features, has high specificity but relatively low sensitivity. This is because normal examples tend to have low expenses, and thus result in a high specificity; whereas fraudulent examples have variable expenses, and thus result in a low sensitivity. Similar conclusions were reported in [11]. Compared with their detection model, our detection model has more balanced values of sensitivity and specificity. Also, the specificity of their detection model is higher than ours, while the sensitivity of our detection model is slightly higher at low support thresholds.

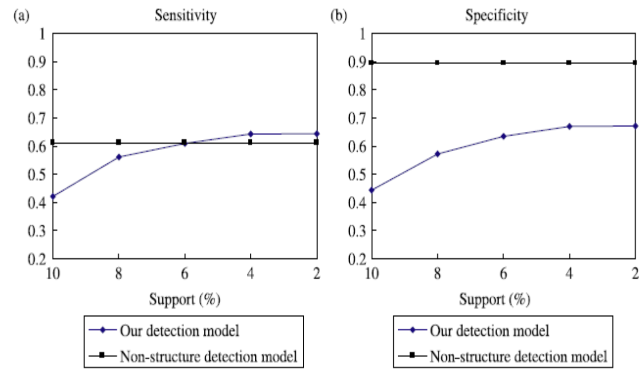

Fig 5: Comparison of our model and the non-structure detection model.

The comparison of sensitivity in Fig. 5 is not intended to demonstrate that one model is better than the other, but rather to illustrate where the differences lie. Of fraudulent examples returned by the non-structure detection model, our detection model captures 69% of the examples on average. Some examples, such as overdose, are not returned by our detection model. In contrast, of the fraudulent examples returned by our detection model, their detection model captures 63% of the examples on average. Some examples, such as those that have repeated ambulant visits while still have low expense, are not returned by their detection model. This illustrates the differences between our structure driven approach and the non-structure driven approach.

VI. CONCLUSION

In this paper, we have facilitated the automatic and systematic construction of systems that detect healthcare fraud. We investigated the mining of frequent patterns from clinical instances and the selection of features that have higher discrimination power. The proposed approaches have been evaluated objectively using a real-world data set. The empirical experiments show that our detection model is efficient and capable of identifying some fraudulent and abusive cases that are not detected by a manually constructed detection model.

VII. FUTURE ENHANCEMENT

This work could be extended in several directions there are many cost factors in healthcare fraud detection, and so building detection models that can be easily adjustable according to site-specific cost policies is important in practice.

REFERENCES

1. Phua C., C., Lee V., Smith K. "Comprehensive Survey of Data Mining-based Fraud Detection Research", Artificial Intelligence Review, (2005).
2. Lassey, M., Lassey, W., & Jinks, M. (1997). Health care systems around the world: Characteristics, issues, reforms. Englewood Cliffs, NJ: Prentice-Hall.
3. Phua, Lee, Smith, and Gayler, "A comprehensive survey of data mining-based fraud detection research", Artificial Intelligenc Review, submitted, 2005.
4. Sokol, Garcia, Rodriguez, West and Johnson, "Using data mining to find fraud in HCFA health care claims,"Top Health Information Management, vol. 22, no. 1, pp. 1-13, 2001.
5. Pflaum and J. S. Rivers, "Employer strategies to combat health care plan fraud," Benefits quarterly, vol. 7, no. 1, pp. 6-14, 1991.
6. AnanshaAsthana, Komal R. Madan, "Educational Data Mining" in National Student Conference on Advances in Electrical and Informatio Communication Technology"AEICT-2014.
7. Hwang, S. Y., Wei, C. P., & Yang, W. S. (2004). Process mining: Discovery of temporal patterns from process instances. Computers in Industry, 53(3), 345–364.
8. Wei, C. P., Hwang, S. Y., & Yang, W. S. (2000). Mining frequent temporal patterns in process databases. Proceedings of international workshop on information technologies and systems, Australia: Brisbane (pp. 175–180).
9. Guidelines to health care fraud (1991). Guidelines to health care fraud. National Health Care Anti-Fraud Association (NHCAA), NHCAA Board of Governors. http//www.nhcaa.org.
10. Blum, A., & Langley, P. (1997). Selection of relevant features and examples in machine learning. Artificial Intelligence, 97(1/2), 245–271.
11. Chan, C. L., &Lan, C. H. (2001). A data mining technique combining fuzzysets theory and Bayesian classifier—An application of auditing the healthinsurance fee. Proceedings of the international conference on artificialintelligence (pp. 402–408).