Your Knowledge Partner™

Medical Fraud Detection Through Data Mining

Megaputer Case Study

Megaputer

Megaputer Intelligence, Inc.
120 West Seventh Street, Suite 310
Bloomington, IN 47404, USA
+1 812-330-0110

www.megaputer.com

Contents

Industry Background

Annual health care expenditures in the US have exceeded $1.3 trillion according to the Health Care Financing Administration. At the same time, the loss by insurance companies and government agencies due to fraudulent healthcare transactions was about $100 billion, amounting to 10% of the nation's annual health care expenditure[1] . Timely detection and prevention of fraud and abuse could help recover enormous amounts of money and return it back to medical institutions and patients, thus improving the quality and decreasing the cost of healthcare for millions of taxpayers.

Typical medical fraud schemes include[2]

- Billing for services not actually performed
- Falsifying a patient's diagnosis to justify procedures that aren't medically necessary
- Misrepresenting procedures performed to obtain payment for non-covered services, such as cosmetic surgery
- "Upcoding" — billing for a more costly service than the one actually performed
- "Unbundling" — billing each stage of a procedure as if it were a separate procedure
- Accepting kickbacks for patient referrals
- Waiving patient co-pays or deductibles and over-billing the insurance carrier or benefit plan (involves both the provider and the patient)

Insurance companies and Medicaid agencies are seeking automated tools to help them reliably identify and flag suspicious activities, separating them from valid transactions. Huge volume and complexity of healthcare transactions significantly complicate the task of their timely and accurate validation of transactions and prevention of fraud. The task of fighting fraud involves two major analytical steps:

1. Discover unknown patterns and relations signifying new fraud schemes. Solving this task helps find past offenders and their fraudulent transactions.
2. Incorporate discovered knowledge as business rules for screening new transactions and flagging and blocking fraudulent transactions in real time.

While there are many automated systems that address the second step of transaction screening in accordance with built-in business rules, the real success of a fraud detection project hinges upon the identification of unknown fraud patterns and relations from the analysis of transactional data. In order to accomplish this task one requires advanced tools for intelligent data analysis, known under the name of data mining solutions. These solutions help investigators quickly uncover new fraud schemes and transform this new knowledge to quantifiable business rules for real-time transaction screening.

[1] US General Accounting Office (GAO) report to Congress.
[2] CIGNA Special Investigations Department.

Case Description

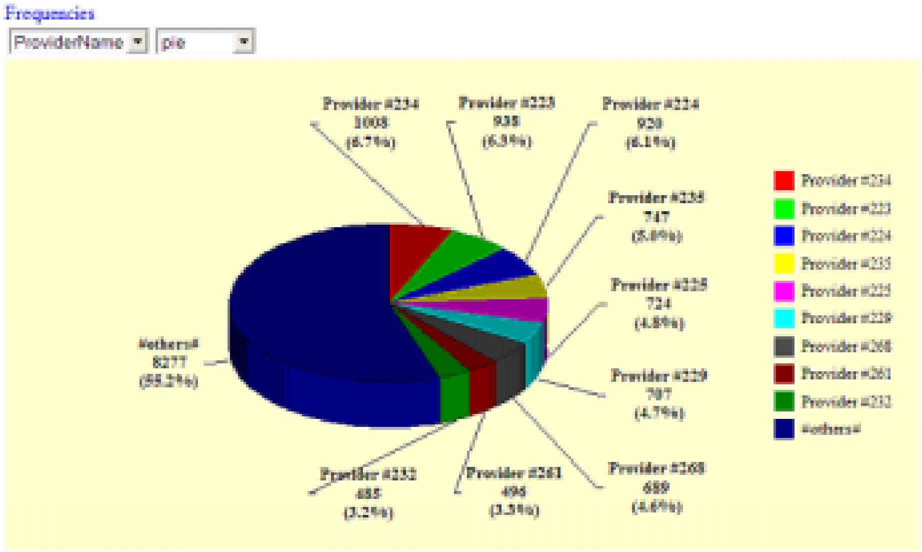
A state Medicaid agency covering about one million people throughout the state wanted to identify and reduce the number of fraudulent medical transactions. To accomplish the task of fraud detection, the agency hired an IT company specializing in medical billing solutions, which developed business rules for capturing known fraud mechanisms based on their background knowledge. Random manual screening of data demonstrated that this system was able to identify about a half of fraudulent transactions, with the second half still being disguised by different fraud mechanisms, which could not be caught by a set of predefined business rules. Megaputer Intelligence was tasked with providing a system capable of identifying unknown fraud schemes directly from the analysis of transactional data and helping the agency significantly increase the share of caught fraudulent transactions.

The data for the analysis had a standard medical data format listing patient name, provider name, date of service, diagnosis, type of procedure, and billed and paid amounts for each procedure. The Medicaid agency wanted to identify and flag suspicious providers and collections of transactions, which could indicate fraud and thus required further investigation.

Implemented Solution

In order to reduce the volume of data to be analyzed and concentrate on more probable candidates for fraud, Megaputer analysts decided to first isolate and study records of patients who received more than 150 procedures during one calendar year. The resulting data contained about 400,000 transactions, which were further explored with the help of advanced analytical algorithms of the PolyAnalyst data mining system.

The Summary Statistics algorithm demonstrated that the remaining records corresponded to 998 patients, 623 types of performed procedures, and 903 providers, 24 of which were larger hospitals performing more than 5,000 individual transactions annually.



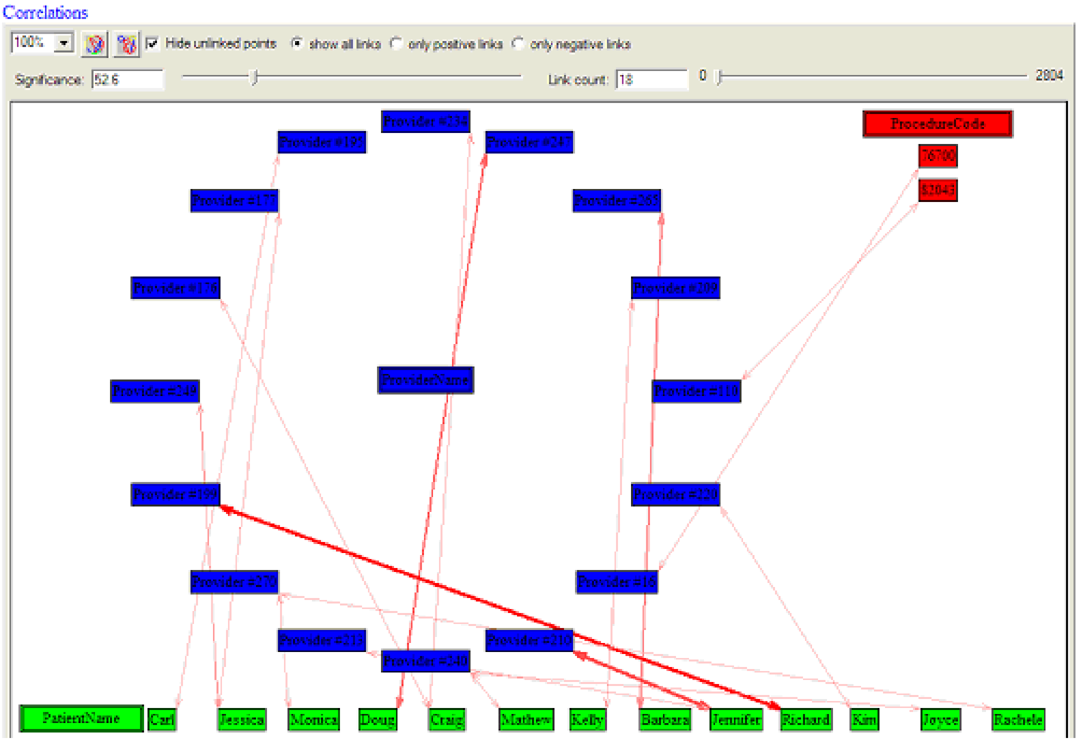
Frequencies

ProviderName ▾ pie ▾

***Figure 1:** Summary Statistics shows the number of healthcare providers and their size*

Provider-Patient Fraud

In order to identify possible provider-patient scams, such as waiving patient co-pays or deductibles and over-billing the insurance carrier or benefit plan, the data was first investigated with the help of the Link Analysis algorithm of PolyAnalyst. Link Analysis reveals and visually displays correlations between values of different attributes: the heavier is the line representing a link, the more correlated are the objects connected by this link.



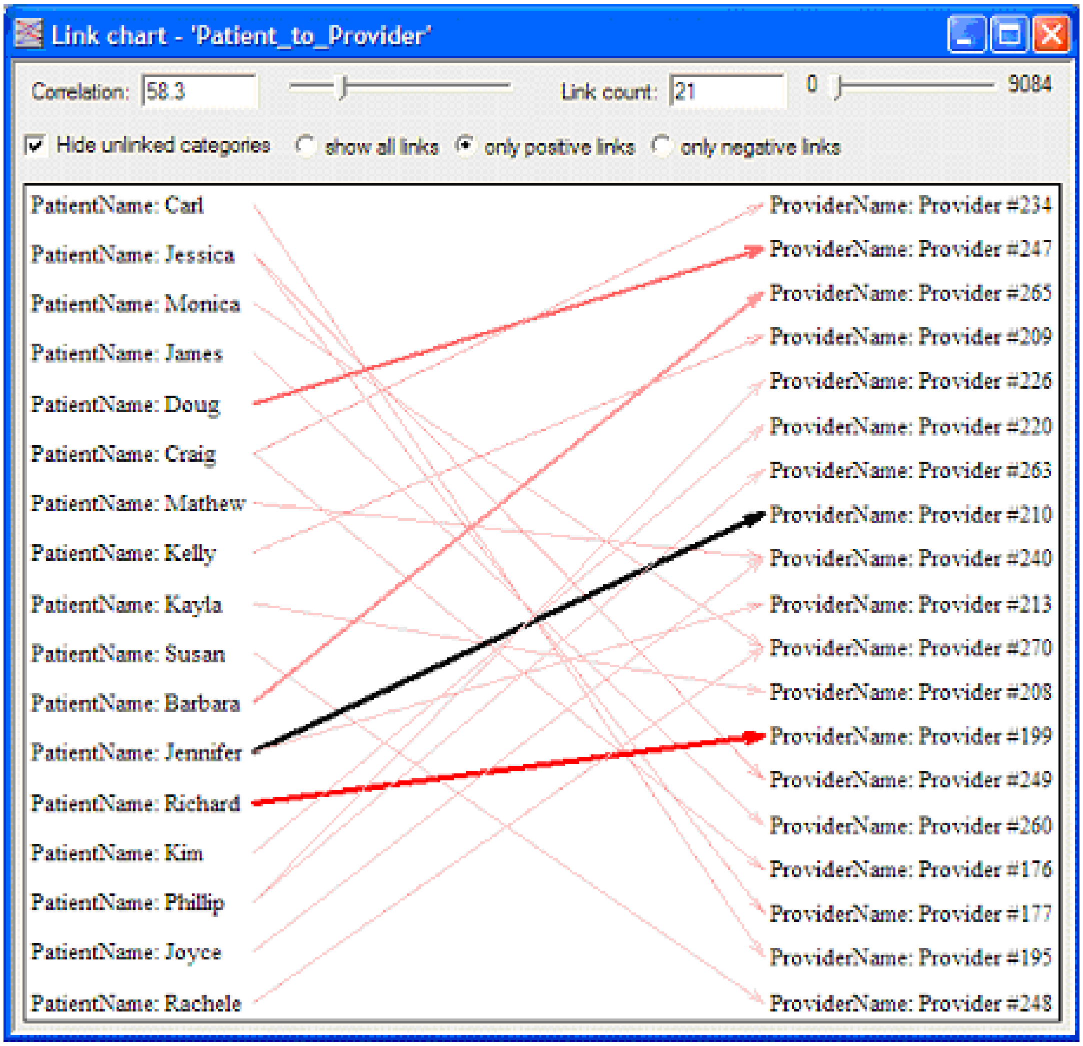
☑ Hide unlinked points ● show all links ○ only positive links ○ only negative links

Significance: 52.6

Link count: 18

Figure 2: *Link Analysis visualizes strong correlations between values of selected attributes*

Figure 2 illustrates the most prominent correlations between individual patients, providers and procedure codes. It is easy to observe that while there are no obvious correlations of individual procedures with either providers or patients, there exist several strongly correlated patient-provider pairs, which deserve further investigation. A more detailed analysis with the help of pairwise Link Chart algorithm of PolyAnalyst displays additional patient-provider pairs in a more organized fashion and allows the user to easily drill through to the underlying transactions and verify the validity of individual transactions.



Link chart - 'Patient_to_Provider'

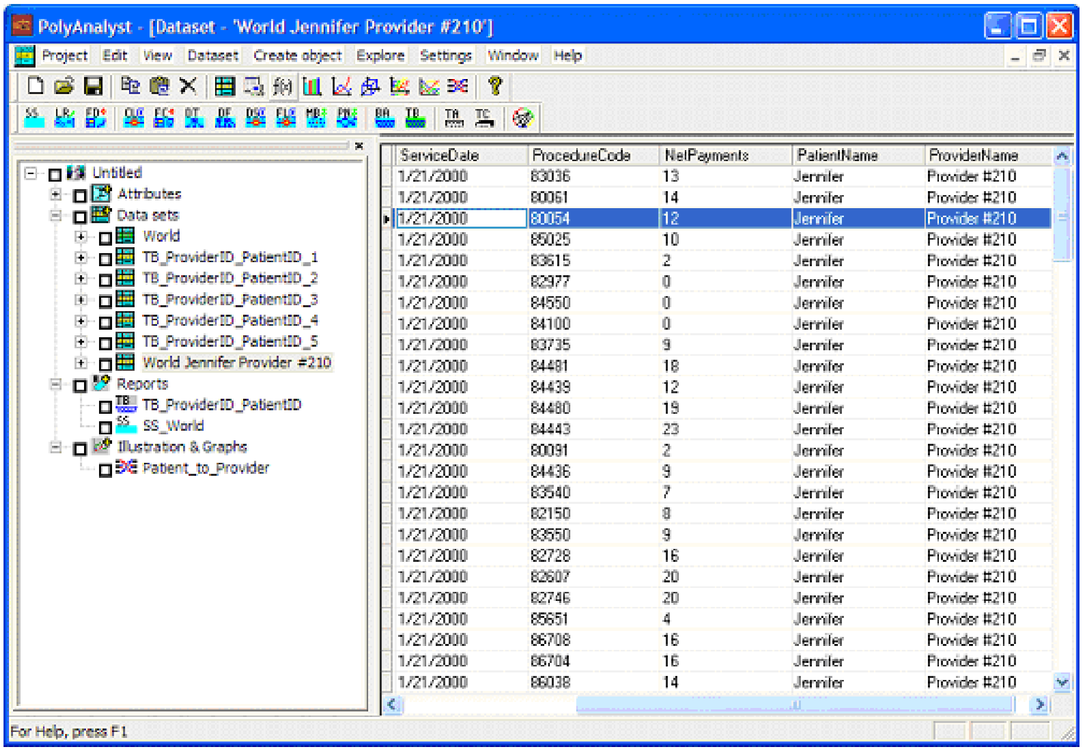
Correlation: 58.3 Link count: 21 0 9084

☑ Hide unlinked categories ○ show all links ● only postive links ○ only negative links

PatientName: Carl
PatientName: Jessica
PatientName: Monica
PatientName: James
PatientName: Doug
PatientName: Craig
PatientName: Mathew
PatientName: Kelly
PatientName: Kayla
PatientName: Susan
PatientName: Barbara
PatientName: Jennifer
PatientName: Richard
PatientName: Kim
PatientName: Phillip
PatientName: Joyce
PatientName: Rachele
ProviderName: Provider #234
ProviderName: Provider #247
ProviderName: Provider #265
ProviderName: Provider #209
ProviderName: Provider #226
ProviderName: Provider #220
ProviderName: Provider #263
ProviderName: Provider #210
ProviderName: Provider #240
ProviderName: Provider #213
ProviderName: Provider #270
ProviderName: Provider #208
ProviderName: Provider #199
ProviderName: Provider #249
ProviderName: Provider #260
ProviderName: Provider #176
ProviderName: Provider #177
ProviderName: Provider #195
ProviderName: Provider #248

Figure 3: Link Chart allows a quick drill-down to underlying transaction records

Figure 4 below illustrates a sample data set obtained with the help of the Link Chart drill-through operation. The performed analysis allowed the agency to flag dozens of suspicious situations and further scrutinize the practices of the involved providers.


PolyAnalyst - [Dataset - 'World Jennifer Provider #210']
Project
Edit
View
Dataset
Create object
Explore
Settings
Window
Help
Untitled
Attributes
Data sets
World
TB_ProviderID_PatientID_1
TB_ProviderID_PatientID_2
TB_ProviderID_PatientID_3
TB_ProviderID_PatientID_4
TB_ProviderID_PatientID_5
World Jennifer Provider #210
Reports
TB_ProviderID_PatientID
SS_World
Illustration & Graphs
Patient_to_Provider
For Help, press F1

ServiceDate	ProcedureCode	NetPayments	PatientName	ProviderName
1/21/2000	83036	13	Jennifer	Provider #210
1/21/2000	80061	14	Jennifer	Provider #210
1/21/2000	80054	12	Jennifer	Provider #210
1/21/2000	85025	10	Jennifer	Provider #210
1/21/2000	83615	2	Jennifer	Provider #210
1/21/2000	82977	0	Jennifer	Provider #210
1/21/2000	84550	0	Jennifer	Provider #210
1/21/2000	84100	0	Jennifer	Provider #210
1/21/2000	83735	9	Jennifer	Provider #210
1/21/2000	84481	18	Jennifer	Provider #210
1/21/2000	84439	12	Jennifer	Provider #210
1/21/2000	84480	19	Jennifer	Provider #210
1/21/2000	84443	23	Jennifer	Provider #210
1/21/2000	80091	2	Jennifer	Provider #210
1/21/2000	84436	9	Jennifer	Provider #210
1/21/2000	83540	7	Jennifer	Provider #210
1/21/2000	82150	8	Jennifer	Provider #210
1/21/2000	83550	9	Jennifer	Provider #210
1/21/2000	82728	16	Jennifer	Provider #210
1/21/2000	82607	20	Jennifer	Provider #210
1/21/2000	82746	20	Jennifer	Provider #210
1/21/2000	85651	4	Jennifer	Provider #210
1/21/2000	86708	16	Jennifer	Provider #210
1/21/2000	86704	16	Jennifer	Provider #210
1/21/2000	86038	14	Jennifer	Provider #210

Figure 4: Results obtained by Transactional Basket Analysis algorithm

Ghost Patient Billing

While Link Analysis algorithms are very efficient tools for identifying some types of fraud, such as *patient-provider fraud*, additional analytical algorithms are better suited for catching other types of fraudulent transactions, for example *ghost patient billing*, which involves offending providers sharing a list of valid patient IDs and billing unsuspecting patients for procedures that were never rendered).

Indeed, the Basket Analysis algorithm of PolyAnalyst provides adequate means to discover groups of providers sharing a large number of patients (or just valid patient IDs as in the *ghost patient billing* fraud scheme). This algorithm simultaneously uncovered groups of patients (or patient IDs) appearing in transactions performed by several providers, and also identified groups of providers rendering services to the same patients. For example, Figure 4 lists five groups of providers sharing some patients, as discovered by the Basket Analysis algorithm (these data sets are named TB_ProviderID_PatientID_xx).

The results obtained by the Basket Analysis algorithm revealed several groups of providers sharing a large number of patients. One particular group of three providers was sharing over 10% of all patients considered in our analysis! Of course, there is always a chance that these providers offer complementary services and the found transactions are perfectly legitimate. Medical transactions can be so involved that a medical fraud investigator has to walk a fine line to separate grain from churn. Nevertheless, the presence of a large number of overlapping patients should definitely raise a red flag for an analyst and require further investigation.

Creating Automated Solution

The final goal of a fraud detection project is not only to discover past fraudulent transactions and create a list of the corresponding offenders, but to screen every new transaction with the discovered business rules and prevent processing suspicious transactions in real time. In some situations this is the only method that can guarantee that an insurance agency does not get charged for fraud. The most suspicious transactions should be scheduled for a more detailed manual scrutiny by an analyst.

To automate the process of knowledge discovery and applying business rules for verification of transactions, Megaputer offers *KnowledgeFactory*™, a unique platform for rapid visual development of reusable push-button analytical solutions. This system allows an analyst to create and distribute advanced analytical scenarios throughout the organization. The system automatically executes these scenarios when certain predefined conditions become true. Consequently, the entire organization gains the capability to quickly and consistently identify new fraud schemes and monitor the validity of submitted transactions in real time.

Conclusion

Fraud Detection is an area of vital importance to numerous organizations: government agencies, banks, credit, insurance, telecom and real estate companies alike have to be able to discern fraudulent activities from the main stream of legitimate business transactions. Inability to catch fraud can become an extremely costly, painful and damaging problem to a business, and thus the need for efficient Fraud Detection solutions should reside high on the "to do" list of every organization.

As fraud schemes get more sophisticated and the volume of transactions grows fast, it becomes increasingly more difficult to discern fraud from the bulk of legitimate transactions. Investigators have to utilize advanced data analysis tools capable of processing large volumes of data, determining rules for separating fraud from legitimate transactions, and detecting unusual events deviating from normal operation patterns. Fraud schemes are rapidly changing and analysts need to be able to discern new fraud patterns without an explicit prior knowledge of these patterns.

The above case study illustrates a methodology for utilizing data mining techniques to discover possible fraudulent cases in healthcare transactions. It is demonstrated how Link Analysis and Basket Analysis algorithms can help the user uncover particular examples of two widespread fraud schemes, provider-patient fraud and ghost patient billing, which cannot be discovered through the application of simple predefined business rules. Similarly, other types of fraud, such as charging for unnecessary services, submitting duplicate or erroneous billing, etc can be discovered through utilizing other data mining algorithms. The case outlined the necessity and a possible implementation of a system for real time monitoring of newly submitted transactions.

The implementation of the discussed techniques can significantly increase the quality and timeliness of detecting medical fraud, as well as real-time flagging of suspicious transactions. Timely discovery and elimination of fraudulent transactions in healthcare and in other domains would save the affected government agencies and commercial companies (and thus, ultimately, taxpayers) hundreds of billions of dollars per year.

Corporate and Americas Headquarters
Megaputer Intelligence, Inc.
120 West Seventh Street, Suite 310
Bloomington, IN 47404, USA
TEL: **+1.812.330.0110**; FAX: **+1.812.330.0150**
EMAIL: info@megaputer.com

Europe Headquarters
Megaputer Intelligence, Ltd.
B. Tatarskaja 38
Moscow 113184, Russia
TEL: **+7.095.951.8079**; FAX: **+7.095.953.5731**
EMAIL: info@megaputer.com

Megaputer
Your Knowledge Partner™