CHEMICAL ENGINEERING TRANSACTIONS
VOL. 33, 2013
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at: www.aidic.it/cet
Guest Editors: Enrico Zio, Piero Baraldi

ISBN 978-88-95608-24-2; **ISSN** 1974-9791
DOI: 10.3303/CET1333026

Application of Bayesian Methods in Detection of Healthcare Fraud

Tahir Ekin[a], Francesca Leva*[,b], Fabrizio Ruggeri [c], Refik Soyer [d]

[a] Department of Computer Information Systems and Quantitative Methods, Texas State University 601 University Dr, McCoy 404, San Marcos, TX, 78666 U.S.A.
[b] Department of Mathematics, Politecnico di Milano, via Bonardi 9, 20133, Milano, Italy.
[c] Consiglio Nazionale delle Ricerche, Istituto di Matematica Applicata e Tecnologie Informatiche, via Bassini 15, 20133, Milano, Italy.
[d] Department of Decision Sciences, School of Business, The George Washington University, 2201 G Street, NW, Duquès Hall Washington, D.C. 20052, U.S.A.
francesca.ieva@polimi.it

The term fraud refers to an intentional deception or misrepresentation made by a person or an entity, with the knowledge that the deception could result in some kinds of unauthorized benefits to that person or entity. Fraud detection, being part of the overall fraud control, should be automated as much as possible to reduce the manual steps of a screening/checking process. In the health care systems, fraud has led to significant additional expenses. Development of a cost-effective health care system requires effective ways to detect fraud. It is impossible to be certain about the legitimacy of and intention behind an application or transaction. Given the reality, the best cost effective option is to infer potential fraud from the available data using mathematical models and suitable algorithms. Among these, in recent years co-clustering has emerged as a powerful data mining tool for analysis of dyadic data connecting two entities. In this paper application of Bayesian ideas in healthcare fraud detection will be presented. The emphasis will be on the use of Bayesian co-clustering to identify potentially fraudulent providers and beneficiaries who have unusual group memberships. Detection of such unusual memberships will be helpful to decision makers in audits.

1. Introduction

The National Health Care Anti-fraud Association (NHCAA) defines health care fraud as “an intentional deception or misrepresentation made by a person or an entity, with the knowledge that the deception could result in some kinds of unauthorized benefits to that person or entity” (NHCAA, 2012). The NHCAA estimated conservatively that at least 3 %, or more than 60 billion dollars, of the US’s annual health care expenditure was lost due to fraud in 2010. Other estimates by government and law enforcement agencies placed this loss as high as 10 % (Aldrich, 2010). In addition to the financial loss, fraud also severely hinders the US health care system from providing quality care to legitimate beneficiaries. Therefore, effective fraud detection is important for improving the quality and reducing the cost of health care services. Abuse and waste only differ from fraud by the degree of the legal intent. Activities that are inconsistent with established practices and result in unnecessary costs to the health care programs can be classified as medical abuse. Failure to document medical records adequately, providing unnecessary services and charging the insurers higher rates are among these activities. It is tough to know the intent for an activity, therefore distinguishing fraud from waste and abuse is challenging, as mentioned in Musal, (2010).
When speaking about fraud, a distinction has to be made between fraud prevention and fraud detection. Fraud prevention describes measures to stop fraud from occurring in the first place. In contrast, fraud detection involves identifying fraud as quickly as possible once it has been perpetrated. Many fraud detection problems involve huge data sets that are constantly evolving. In general, fraud detection comes

Please cite this article as: Ekin T., Ieva F., Ruggeri F., Soyer R., 2013, Applications of bayesian methods in detection of healthcare frauds, Chemical Engineering Transactions, 33, 151-156 DOI: 10.3303/CET1333026

into play once fraud prevention has failed. In what follows, we will focus on statistical methods for identifying fraud. Our focus will be on health care fraud.

Fraud in health care is generally classified into three categories based on the source of the fraudulent activity as provider (hospitals, physicians) fraud, consumer (patients) fraud and insurer fraud. U.S. law identifies the submission of false claims, the payment or receipt of kickbacks and self-referrals as provider fraud (Kalb, 1999). In addition, up-coding (charging for a more expensive service) and unbundling (charging separately for procedures which are initially part of one procedure) are also examples of provider fraudulent activities, as discussed in Li et al. (2008). Consumer fraud are the cases that patients are involved in fraudulent activities such as falsifying documents to obtain extra prescription or misusing their insurance cards. Insurer fraud happens when insurers falsify statements or they simply do not provide the insurance they have collected premiums for.

2. Medical Fraud Data

Recent development of new technologies eased production, collection and storage of high dimensional and complex data. Healthcare has been no exception. Modern medicine generates a great deal of information stored in medical databases.

Medical databases are increasing in size in three ways: (1) the number of records in the database, (2) the number of fields or attributes associated with a record, (3) the complexity of the data itself. Extracting pertinent information from such complex databases for inferring potential fraudulent activities has become increasingly important for fraud detection. Popkoski (2012) gives an account of the amount of information involved in the reimbursement process for Medicare D, which supports the cost of prescription medications to seniors and the disabled in the US. In such a complex process, involving many actors, the possibility of fraud cannot be overlooked. At the same time, quality of medical records should be ensured to avoid, for example, false claims of fraud: a detailed discussion can be found in Gregori and Berchialla (2012).

Data mining, a step in the process of Knowledge Discovery in Databases (KDD), is a method of extracting information from large data sets. Built upon statistical analysis, it can analyze massive amounts of data and provide useful information about patterns and relationships that exist within the data that might otherwise be missed. Data mining techniques have gained attention in the fraud detection literature; see for example, the review by Bolton and Hand (2002). Most of these have been considered for credit card fraud and general insurance fraud. Use of sophisticated data mining tools for health care fraud has been relatively new; see the recent review by in Li et al. (2008). As noted by the authors, these tools extensively include supervised algorithms such as neural networks, decision trees, association rules and genetic algorithms. These methods are successful in modelling particular data sets and stable fraud patterns for which classification of data is easier. However, the rare nature of fraud makes classification a difficult task and hinders the success of supervised algorithms in modelling health care data with dynamic fraud patterns. Therefore, unsupervised methods are proposed to detect abnormal dynamic patterns. Yamanishi et al. (2004) used outlier analysis to investigate the existence of potential fraudulent activities. More recently, Musal (2010) proposed use of cluster analysis for geographical analysis of potential fraud.

The emphasis of previous work in health care has been on types of fraud committed by a single party. Li et al. (2008) point out that there is a newly emerging type of fraud called “conspiracy fraud” which involves more than one party. An important characteristic of conspiracy fraud is the need to deal with dyadic data connecting the involved parties. The important feature of dyadic data is that it can be organized into a matrix where rows and columns represent a symmetric relationship. In health care fraud detection the typical relationship of interest is the one between a provider and a beneficiary. As noted by Li et al. (2008), detection of conspiracy fraud has not gained much attention in the health care fraud literature. In what follows, we consider use of co-clustering methods for detection of conspiracy fraud. In so doing, we propose Bayesian models for describing and capturing the dyadic dynamic that connects providers and beneficiaries. Co-clustering enables us to group providers and beneficiaries simultaneously, that is, the clustering is interdependent. The objective of the proposed approach is to identify potentially fraudulent associations among the two parties for further investigation.

Analysis of health care dyadic data presents many challenges. Due to the high number of beneficiaries involved and many types of services being provided, data size is huge, usually in terabytes. Beneficiaries and providers are not homogeneous since there is a great variety in the services being provided and the monetary charges involved. Furthermore, legal systems and health care procedures may change frequently which lead to changes in definition of fraudulent and legitimate practices. Bayesian approaches are suitable to capture these dynamic patterns; (so called adaptive fraud detection). The health care dyadic data may consist of visitation links associated with pairs of health service providers (doctors) and

beneficiaries (patients), number of visits or insurance claims involving provider-beneficiary pairs or monetary charges associated with provider-beneficiary pairs.
Our development in the next section is based on visitation links but can be easily extended so that other attributes of providers and/or beneficiaries are considered within the model. The proposed Bayesian co-clustering algorithm which is based on Markov chain Monte Carlo methods is general and can be easily adapted to other types of dyadic data. The attractive feature of the Bayesian approach is its incorporation of subjective input such as the medical knowledge into the analysis and the quantification of uncertainty about associations and therefore fraudulent relationships probabilistically. Furthermore, the Bayesian approach can handle missing data in a very straightforward manner.

3. Bayesian Co-clustering

Recently, co-clustering has emerged as a powerful data mining tool that can analyze dyadic data connecting two entities. Such dyadic data are represented as a matrix with rows and columns representing each entity respectively. An important data mining task pertinent to dyadic data is to get a clustering of each entity. Traditional clustering algorithms do not perform well on such problems because they are unable to utilize the relationship between the two entities. In comparison, co-clustering can achieve a much better performance in terms of discovering the structure of data and predicting the missing values by taking advantage of relationships between two entities (Agarwal and Merugu, 2007). Simultaneous clustering of rows and columns of a data matrix was proposed firstly by Hartigan (1972). Earlier work on Bayesian cluster analysis is due to Binder (1978). Bayesian co-clustering approaches have been considered mostly in data mining and machine learning literature; see for example Shan and Banerjee (2008).

In the sequel, we propose a general co-clustering model for healthcare fraud detection. We assume each row and column to have a mixed membership respectively, from which row and column clusters are generated. Each entry of the data matrix is then generated given that row-column cluster, i.e., the co-cluster. Moreover, assume that we have *I* health-care providers and *J* health-care service users or beneficiaries. Let X_{ij} be a binary random variable representing if the provider *i* serves user *j*. In other words, X_{ij} is a Bernoulli random variable

$$X_{ij} = \begin{cases} 1 & \text{if provider i serves beneficiary j} \\ 0 & \text{otherwise} \end{cases}$$

We have $\mathbf{X} = \{X_{ij};\ i = 1,...,I,\ j = 1,...,J\}$, a data matrix of size IxJ. Assume that there are K clusters of providers and L clusters of users. Marginal membership probabilities are denoted by π_{1k}, i = 1,...,K for row clusters and by π_{2l}, l = 1,...,L for column clusters such that

$$\sum_{k=1}^{K} \pi_{1k} = \sum_{l=1}^{L} \pi_{2l} = 1 \tag{1}$$

The latent variables Z_{1i} and Z_{2j} , i = 1,...,I, j = 1,...,J, denote membership to the row (provider) and column (beneficiary) clusters such that $Z_{1i} \in \{1,...,K\}$ and $Z_{2j} \in \{1,...,L\}$. Given $\boldsymbol{\pi_1} = (\pi_{1k};\ k = 1,...,K)$ and $\boldsymbol{\pi_2} = (\pi_{2l};\ l = 1,...,L)$, Z_{1i} and Z_{2j} are independent discrete random variables. Furthermore, given the latent variables Z_{1i} and Z_{2j}, X_{ij}'s are Bernoulli random variables with parameter $\theta_{Z_{1i}Z_{2j}}$, that is,

$$X_{ij} \mid Z_{1i} = k, Z_{2j} = l, \theta_{kl} \sim \text{Be}(\theta_{kl}) \tag{2}$$

and X_{ij}'s are conditionally independent. The co-clustering problem involves assignment of each X_{ij} to a co-cluster defined by the latent pair (Z_{1i} and Z_{2j}).

The Bayesian model involves specification of priors for the unknown parameters $\boldsymbol{\pi_1}$, $\boldsymbol{\pi_2}$ and $\Theta = (\theta_{kl};\ k = 1,...,K,\ l = 1,...,L)$. We can assume independent Dirichlet priors for $\boldsymbol{\pi_1}$ and $\boldsymbol{\pi_2}$ and independent beta priors for elements of Θ. More specifically, we have

$$\begin{aligned} &\pi_1 \sim Dir(\alpha_{1k}; k = 1,...,K), \quad \pi_{2l} \sim Dir(\alpha_{2l}; l = 1,...,L), \\ &\theta_{kl} \sim Beta(a_{kl}, b_{kl}),\ k = 1,...,K,\ \ l = 1,...,L \end{aligned} \tag{3}$$

Straightforward extensions of the model may include assuming $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$ and Θ to be modelled through beneficiary and/or provider specific covariates, as well as different assumption on data distribution, within the exponential family.

Given data matrix $\mathbf{X} = \{X_{ij};\ i = 1,...,I,\ j = 1,...,J\}$, the joint posterior distribution of $\boldsymbol{\pi}_1$, $\boldsymbol{\pi}_2$, Θ and the latent vectors $\mathbf{Z}_1 = \{Z_{1i};\ i = 1,...,I\}$, $\mathbf{Z}_2 = \{ Z_{2j};\ j = 1,...,J\}$ can not be obtained analytically. However, the posterior analysis can be developed by using a standard Gibbs sampler; see for example Casella and George (1992). Implementation of the Gibbs sampler requires the full posterior conditional distributions of $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2$, Θ, $\mathbf{Z}_1$ and $\mathbf{Z}_2$. By successively drawing samples from the full conditionals we can obtain samples from the joint posterior distributions of $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2$, Θ, $\mathbf{Z}_1$ and $\mathbf{Z}_2$.

The full conditionals for θ_{kl}'s can be obtained as (conditionally) independent beta densities given by

$$\theta_{kl} \mid \mathbf{Z}_1, \mathbf{Z}_2, \mathbf{X} \sim Beta\left(a_{kl} + \sum_{i,j} X_{ij} \mathrm{I}\left(Z_{1i} = k, Z_{2j} = l\right), b_{kl} + \sum_{\mathrm{i,j}} \left(1 - X_{ij}\right) \mathrm{I}\left(Z_{1i} = k, Z_{2j} = l\right) \right) \tag{4}$$

The full conditionals of $\boldsymbol{\pi}_1$ and $\boldsymbol{\pi}_2$ are (conditionally) independent Dirichlet distributions are given by

$$\begin{aligned} \pi_1 &\sim D\mathrm{ir}\left(\alpha_{1k} + \sum_{i,j} \mathrm{I}\left(Z_{1i} = k\right); k = 1,..., K \right) \\ \pi_2 &\sim D\mathrm{ir}\left(\alpha_{2l} + \sum_{i,j} \mathrm{I}\left(Z_{2l} = l\right); l = 1,..., L \right) \end{aligned} \tag{5}$$

Finally, the full conditionals of the couple (Z_{1i}, Z_{2j}) can be obtained as

$$p\left(Z_{1i} = k, Z_{2j} = l \mid \boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \boldsymbol{\theta}, X_{ij}\right) = \frac{\theta_{kl}^{X_{ij}} \left(1 - \theta_{kl}\right)^{\left(1 - X_{ij}\right)} \pi_{1k} \pi_{2l}}{\sum_{r=1}^{K} \sum_{c=1}^{L} \theta_{rc}^{X_{ij}} \left(1 - \theta_{rc}\right)^{\left(1 - X_{ij}\right)} \pi_{1r} \pi_{2c}} \tag{6}$$

Once the samples are drawn from the posterior distributions, we can infer co-clusters of providers and beneficiaries by looking at the probabilities of all latent pairs (Z_{1i}, Z_{2j}). Also, by looking at the posterior distributions of θ_{kl}'s we can infer which co-clusters have higher interactions. These posterior distributions help us to identify unusual provider-beneficiary pairings. Furthermore, analysis of the marginal posterior distributions of Z_{1i}'s and Z_{2j}'s enable us to identify unusual memberships in provider and beneficiary clusters. As previously noted, the Bayesian co-clustering model is helpful to flag potential fraudulent activities by detecting unusual co-cluster and/or cluster memberships.

4. Illustration using Simulated Data

In this section we present an implementation of the proposed model using simulated data. In so doing, we test the performance of the proposed approach with a toy example where we simulated the data matrix $\mathbf{X} = \{X_{ij};\ i = 1,...,20,\ j = 1,...,300\}$, assuming the presence of K = 2 clusters of providers and L = 3 clusters of beneficiaries. Simulated actual membership arise from a generating model where $\boldsymbol{\pi}_1$ =(0.9,0.1) and $\boldsymbol{\pi}_2$ =(0.3,0.3,0.4). This means that we expect to find the most part of providers within the first cluster, that is, $P(Z_{1i}=1) >> P(Z_{1i}=2)$, i = 1,...,20. On the other hand, beneficiaries are almost equally distributed over the three clusters. Moreover, the Θ matrix used for simulating data is the following:

$$\Theta = \begin{bmatrix} 0.05 & 0.1 & 0.9 \\ 0.2 & 0.8 & 0.5 \end{bmatrix} \tag{7}$$

Concerning Θ entries, the higher the θ_{kl}, the more likely is the probability that a member of provider cluster *k* serves the members of beneficiary cluster *l*.

Then we analyzed data in order to see whether the procedure can estimate the actual values used to simulate data themselves. Following the development in Section 2, we ran a Gibbs sampler of 10000

iterations, discarding the first 5000 (burn-in) and using the sample of last 5000 iterations for posterior analysis. Moreover, we set diffuse but proper priors with hyperparameters $\alpha_1 = (1,1)$, $\alpha_2 = (1,1,1)$ and $(a_{kl}, b_{kl}) = (1,1)$. The posterior means of the components of the Θ matrix is given below

$$\hat{\Theta} = \begin{bmatrix} 0.06 & 0.10 & 0.90 \\ 0.21 & 0.86 & 0.53 \end{bmatrix} \tag{8}$$

illustrating that the posterior distributions of θ_{kl}'s are estimated accurately by the Bayesian approach.

In Figure 1, posterior membership probabilities of provider 18 and beneficiary 5, whose actual memberships are groups k=2 and l=3 respectively, are shown. The posterior membership probability distributions are illustrated in Figure 2. Posterior medians of π_1 and π_2 distribution are, respectively, (0.84, 0.16) and (0.28, 0.33, 0.39). Thus, we can conclude that π_1 and π_2 are estimated reasonably well.

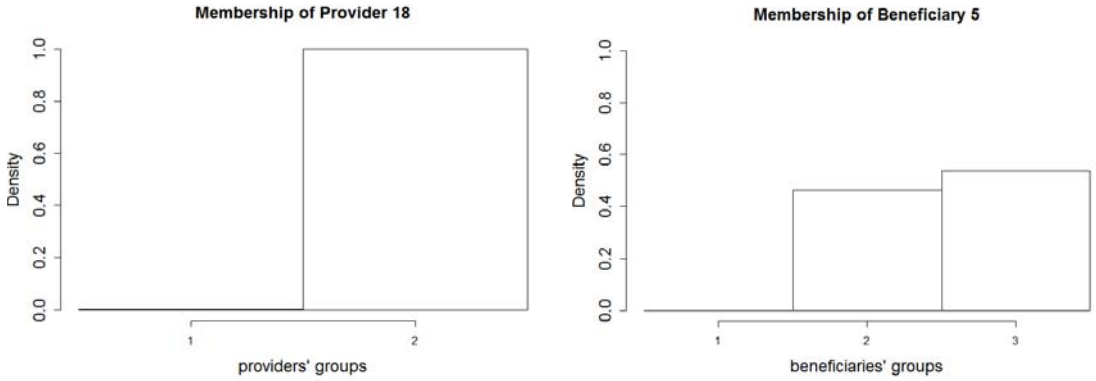

Figure 1: Marginal posterior distributions of memberships of provider 18 (i.e., Z_{1i}, i=18) and beneficiary 5 (i.e., Z_{2j}, j=5).

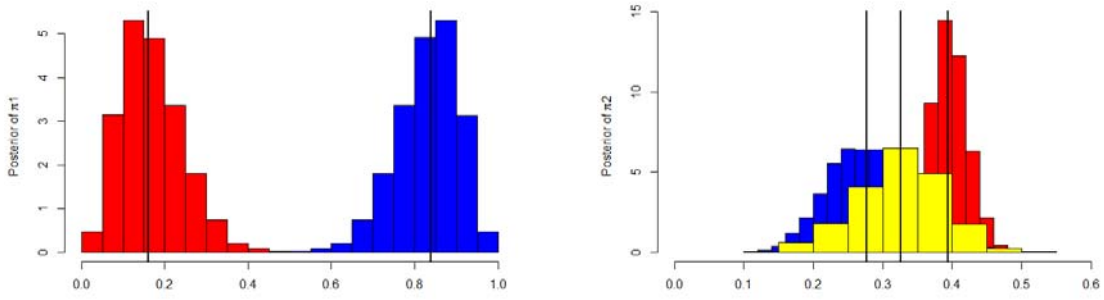

Figure 2: Posterior distribution of membership probabilities π_1 and π_2

5. Conclusions and Further Developments

Use of sophisticated statistical methods in health care fraud detection has been relatively new, mostly because of the difficulty in accessing medical data due to confidentiality and privacy issues. As we have discussed in previous sections, statistical approaches have lot to offer in medical fraud assessment. The statistical methods can be helpful in identifying potential fraudulent behavior as well as in minimizing costs of the subsequent investigation process. The Bayesian approach provides formalism for both quantifying uncertainty about fraudulent behavior as well as for making decisions for investigation of fraud. Potential incorporation of subjective expert knowledge in the Bayesian framework via the prior distributions makes it more attractive in the type of problems we have considered here. For example, in the co-clustering

problem, one can declare his/her prior opinion by assigning high probability to geriatricians being in the same cluster of providers, elderly people being in a cluster of beneficiaries as well as both groups being co-clustered. Moreover, the use of a Bayesian approach would be helpful in determining future evolution of clusters and forecasting possible behavior of new providers/beneficiaries given their characteristics. Finally, with the advances in medical fraud assessment more statistical approaches which combine medical prevention, detection and response efforts would be needed. Integration of information systems that combine different sources could be useful (See Iancu et al. 2012 for a relevant work), and a real time analysis and dynamic monitoring can be a viable option in the near future by use of Bayesian methods.

References

Agarwal D., Merugu S., 2007, Predictive discrete latent factor models for large scale dyadic data. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, 26-35

Aldrich N., 2010, Medicare Fraud Estimates: A Moving Target?, The Sentinel, May 2010, 21-24, Center of Service and Information for Senior Medicare Patrol (SP) projects

Binder D.A., 1978, Bayesian Cluster Analysis, Biometrika, 65 (1), 31-38

Bolton R.J., Hand D.J., 2002, Statistical fraud detection: A review, Statistical Science, 17 (3), 235-255

Casella, G., George, E. I., 1992. Explaining the Gibbs Sampler, The American Statistician, 46, 167-174

Gregori, D., Berchialla, P., 2012, Quality of electronic medical records, in Statistical Methods in Healthcare, Eds. Faltin F., Kenett, R., Ruggeri F., Wiley, Chichester, UK.

Hartigan J., 1972, Direct clustering of a data matrix, Journal of the American Statistical Association, 67(337), 123-129

Iancu P., Adamescu D., Plesu V., Dinu G., Arsene C., Nicola S., Gorunescu L. E. and Gheorghe C. N., 2012, EMSYS - human resources and payroll management integrated information system, Chemical Engineering Transactions, 29, 1633-1638

Kalb P., 1999, Health care fraud and abuse, Journal of the American Medical Association, 282(12), 1163-1168

Li J., Huang K-Y., Jin J., Shi, J., 2008, A survey on statistical methods for health care fraud detection, Health Care Management Science, 11, 275-287

Musal R., 2010, Two models to investigate Medicare fraud within unsupervised databases, Expert Systems with Applications, 37(12), 8628-8633

NHCAA (National Health Care Anti Fraud Association), 2012, What is Health Care Fraud? < www.nhcaa.org/resources/health-care-anti-fraud-resources/consumer-info-action.aspx > accessed: 10.1.2013

Popkoski, M, 2012, Statistical issues in insurance/payor processes, in Statistical Methods in Healthcare, Eds. Faltin F., Kenett R., Ruggeri F., Wiley, Chichester, UK.

Shan H., Banerjee A., 2008, Bayesian Co-Clustering. 8th IEEE International Conference on Data Mining, 530-539

Yamanishi, K., Takeuchi, J., Williams, G., and Milne, P., 2004, On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. Data Mining and Knowledge Discovery, 8(3), 275-300