Detecting anomalous doctors by measuring behavioral volatility using temporal clustering

Authors: Daniel Lasaga, Dan Olson

TABLE OF CONTENTS

Abstract

Health care provider fraud and waste are persistent problems in the public and private health insurance industry with some estimates showing more than 20% of expenditures going to fraud and waste. [1] The dynamism and ephemeral nature of fraud creates difficulties for pinning down models to capture fraud and waste problems among providers.

We propose to use an entity-action temporal clustering model where we quantify discrete actions taken by doctors, cluster this into peer groups and monitor their volatility over time. The behavioral signatures for each doctor in the population established and tracked over time by clustering them into peer groups using Gaussian Mixture Models (GMMs). The relationships of each individual doctor are tracked in relation to peer groups. Normal behavior would be expected to exhibit consistent action profiles over time, while those with inconsistent behavior represent potential higher risk. Suspicious behavior is not simply those that have outlier behavior relative to other doctors, but instead those that show volatility from shifts in behavior. Based on experimental data simulation, our proposed entity-action temporal clustering model suggest superior results to a traditional clustering approach in isolating bad behavior among doctors.

Copyright © 2019 Deloitte Development LLC. All rights reserved.

Intro

Prescription fraud and procedural overutilization are often cited as ubiquitous in the health care. Studies have shown up to 20% of procedures and prescriptions in the US are unnecessary [1]. Combating fraud has been exceedingly difficult for a number of reasons. Health care is a complex and multifaceted industry requiring deep expertise and training. Furthermore, with the shifting nature of fraud and abuse, there is often limited knowledge of how fraudulent behavior exhibits itself. When fraudulent behavior is successfully uncovered, it often shifts quickly away from model estimations. The ephemeral quality of fraud means that supervised models will begin to age as soon as they are completed. While there are a number of methods to tackle fraud detection using supervised methods, increasingly fraud detection pursuits are not complete without unsupervised methodology.

We propose to use a discrete entity-action temporal clustering model in which an entity is composed of the sum of its actions. In this approach, we create of series of temporally shifting clusters designed to estimate peer groups across a range of time. We then track instability in individual behaviors relative to peers. In this case, the entity is an individual doctor and its actions are represented by medical procedures and drug prescriptions claimed for payment to health insurance providers. The paper will first give brief review of literature on unsupervised methods as they pertain to fraud identification in health care. We will then describe in detail the proposed entity-action model and the experimental design to test the approach. Finally, we will look at experimental results ending with a discussion and conclusions.

Existing Work

Literature on fraud detection in health care is extensive, therefore we limit our review here to the past few years. Supervised learning can play an important role in health care fraud detection as is demonstrated by the brief discussion of supervised methodologies here [2]. However, we will otherwise limit ourselves to literature on unsupervised learning because it takes a prominent role in the detection of fraud, waste, and abuse in health care and that is ultimately what our proposed model falls under.

Capelleveen and Liu both propose clustering oriented around claim costs to detect of fraud. Capelleveen's paper explores multivariate clustering of claim costs to determine risk areas [3]. Liu et. al research methods to predict cost overages by finding deviations from expected values in provider peer group clusters [4].

Bauder's paper and Gao both use outlier analysis to explore provider behavioral anomalies through procedure codes and provider characteristics. Bauder, Rosa and Khoshgoftaar explore different unsupervised techniques of identifying provider fraud over three years of aggregated provider data. Their paper determines that Local Outlier Factor was the most effective way to identify fraud [5]. Gao et. al investigate anomalies in behavioral sequences of medical activities. They combine processing of sequences with domain expertise and Local Outlier Factor based approach to arrive at a probability of fraud [6].

Another area of the literature leverages graph analysis to find high-risk relationships among health care providers and their patients. Gangopadhyay and Chen develop graph algorithms to detect communities with suspicious relationships [7]. Juan Liu et. al wrote a well-cited paper that explores options in graph analysis to find latent networks of providers and pharmacies sharing anomalous practices [8].

Copyright © 2019 Deloitte Development LLC. All rights reserved.

While clustering is prevalent in the literature for health care fraud detection, we have not found research which combines clustering medical entities with volatility analysis. More generally, there is research on time series clustering analysis, but it mainly focuses on grouping different sequences together [9], [10]. We did find some research in Aoying Zhou et. al that focuses on tracking the evolution of clusters across time [11]. Our entity-action volatility approach does not appear to have had much exploration in health care or in the wider literature.

Model methodology

We expect most doctors fall into natural peer groups per their specialties. Orthopedic surgeons and podiatrists should naturally have very different peers with distinct procedure/prescription profiles. A simplistic approach would use clustering to estimate peer groups and assess risk using basic anomaly detection: if an individual falls far from all of the cluster centers, then that doctor is considered to be anomalous and high risk. However, there may be legitimate specialists that do not have close peers which would be flagged in the model. For example, specialists in facial reconstructive surgery might share aspects of plastic surgery, dermatology and orthopedic surgeons. Furthermore, there may be doctors perpetrating fraud that position themselves closer to a peer group they should not be near. An example of this may be a pediatrician that appears more closely to a pain management specialist because they are over-prescribing pain medicines. We ideally want to differentiate legitimate specialized doctors from high-risk doctors.

We aim to differentiate legitimate outliers from high-risk outliers by using temporal analysis to further qualify the expected behavior of a doctor. A specialized doctor might not have a close peer group, but they would be expected to show a consistent relationship to the more distance peers around them. Conversely, doctors that shift peer groups should heighten our perception of potential risk. The volatility of behavior over time is a better indicator of risk than whether a doctor is an inlier or outlier relative to his or her peer group at a given time. Our model focuses on behavioral volatility over time of individual doctors per the quantifiable actions they have taken.

Our proposed model consists of three distinct components in order to quantify behavioral volatility. First, we construct behavioral signatures for individual doctors. Secondly, the individual signatures are clustered into peer groups that shift across the date range of the data. Finally, individual doctors are measured against peer groups per time period to assess volatility. In order to arrive at a deeper understanding, we will step through each of these components in detail and how they fit together.

Construction of behavioral signatures

The basis of our entity-action temporal clustering model is the construction of a behavioral signatures from the monthly counts of prescription and procedures prescribed by each doctor. Quantified actions allow us to formulate behavioral signatures from the normalized proportion of different types of procedures and prescriptions claimed. We normalize monthly volumes using simple proportion. If a doctor in a single month prescribes 20 prescriptions for analgesics, 60 for anticoagulants, 120 for antibiotics, and no anti-rheumatics, then we would have a signature of 0.1, 0.3 and 0.6 respectively. The monthly proportional prescriptions per doctor

Table 2: Example action matrix. Note the risky shift in behavior from doctor 003

DOCTOR ID	MONTH	ANALGESIC	ANTICOAGUL	ANTIBIOTICS	ANTI-RHEUM
001	4	0.1	0.3	0.6	0.0
002	4	0.2	0.2	0.6	0.0
003	4	**0.1**	0.4	0.3	0.2
001	5	0.1	0.3	0.6	0.0
002	5	0.2	0.2	0.6	0.0
003	5	**0.5**	0.3	0.1	0.1

Copyright © 2019 Deloitte Development LLC. All rights reserved.

combine to form an action matrix [see Table 2]. Once we have created a normalized action matrix, we run Principal Component Analysis (PCA) to effectively merge together closely related actions and reduce the dimensionality of the entity-action matrix.

Peer group clustering

Given a fully completed action matrix with several years of history, we proceed to formulate a relative monthly measure of peer groups using GMMs. GMMs are an agile generalization of the more commonly known *k*-means clustering. The main advantages of GMMs over *k*-means are that they allow relative measures of association to centroids and they use relative Mahalanobis distance instead of isometric Euclidean distance. Because of this, GMMs can form more flexible cluster definitions with probabilistic associations.

Like a *k*-means model, we will need to specify a number of cluster centers. While there are other ways (such as a Dirichlet process), we derive the initial optimal number of clusters through a grid search. The main focus is to establish stable representations of peer group centers from the initial period. Once the number of clusters is established this will remain constant throughout the analysis. The cluster centers will be seeded by the first time period of the dataset. The model tracks peer group structures progressively over time. Each subsequent time period will use the location of the cluster centers from the prior time period as the start point to establish new centers. Given the way that GMMs are optimized through Estimation Maximization, it guarantees minimum necessary shifting of each cluster from one period to the next. The temporal shifting of cluster centers will account for endogenous and exogenous changes in the behavior and environment of the system of doctors.

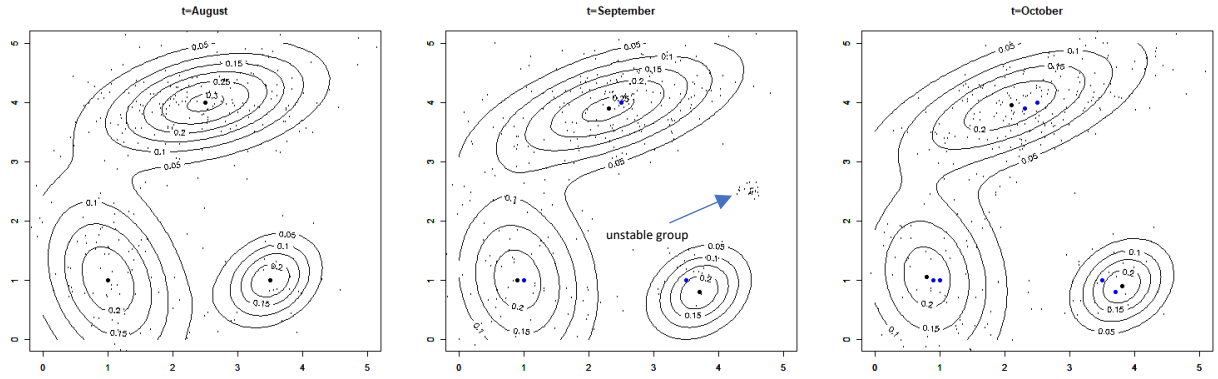

Figure 1: Month over month shifts in peer group behavior. The three larger solid black dots represent peer group centers. September and October show historical peer group centers as solid blue dots. An unstable group is pointed out in September.

Figure 1 shows a hypothetical progression of doctor peer groups, month over month. While individual doctors show more radical movements, we would expect the peer group centers to make more subtle shifts. The group that forms in the center right of the September time period and then disappears from the October data exemplifies ephemeral structures we would not want the GMM clusters to capture because they may actually be suspicious. Instead we want the contours of the peer groups to remain relatively unaffected by the smaller unstable group.

Copyright © 2019 Deloitte Development LLC. All rights reserved.

Temporal doctor monitoring

Establishing cluster centers and enabling them to shift over time allows us to account for national trends or policy changes, which may alter the behavior of subpopulations in the data. The ultimate goal is to monitor individual behavior of doctors relative to their peers and to their own past history. The premise of the model is that doctors that maintain consistent behavior are typically low risk and that doctors with major observed shifts or volatility in behavior are higher risk. The GMM clusters offer a relative measure in which to monitor the behavioral consistency of doctors by tracking behavior from one time period to the next. Consistent outliers or inliers generally should be considered lower risk, and doctors with unstable changes in behavior should be looked at with more scrutiny.

Figure 2: Using GMMs to monitor doctor behavior

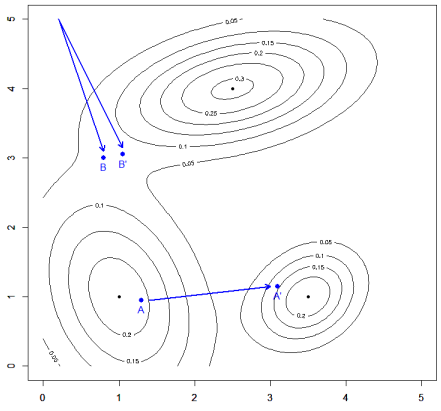

Figure 2 depicts the shift in behavior between time t and t' of two doctors A and B. We see that doctor A is a relative inlier in both periods but switches the cluster it resides in, while doctor B is a clear outlier in its positioning between two separate peer groups but remains so in both periods. While doctor B may exhibit atypical behavior, the small movement between periods indicates consistent behavior over time suggesting there should be less concern that it is an outlier. In contrast, doctor A is acting in more typical behavioral zones, but its major shift between peer groups is unusual and suggests it should be viewed with more scrutiny.

In order to capture and quantify relative behavioral shifts and isolate potential risky patterns, we create a time series for each doctor. Each time period will measure magnitude of change in association to peer group centers. We take the sum of squared changes in probability of doctor_p existing in peer $\text{group}_{k\ldots K}$ at time t. Last we take the mean* volatility for each doctors across the observation range.

$$X = matrix\ of\ behavioral\ signatures\ in\ time\ period\ t\ in\ T\ for\ each\ doctor\ p\ in\ P$$

$$\mu_k = center\ for\ peer\ group\ k\ of\ K\ peer\ groups$$

$$\Sigma_k = covariance\ matrix\ for\ peer\ group\ k\ of\ K\ peer\ groups$$

$$\phi_k = normalizing\ weight\ so\ that\ gaussians\ PDFs\ sum\ to\ 1$$

$$M_{k,t,p} = MultivariateNormalPDF\left(x_{t,p}, \mu_{t,k}, \Sigma_{t,k}, \phi_{t,k}\right)$$

$$Behavioral\ volatility\ measure\ V_p = \frac{\sum_{t=2}^{T}\sum_{k}^{K}\left(M_{k,t,p} - M_{k,t-1,p}\right)^2}{T-1}$$

* Depending on the use case it may be more appropriate to use a recency-weighted mean with a weight of $\frac{1}{T-t+1}$.

Copyright © 2019 Deloitte Development LLC. All rights reserved.

The multivariate normal Probability Density Function (PDF) measures how far away a point is from the center of a cluster taking into account the covariance structure of points that are "members" of the cluster. $M_{k,t,p}$ is the probability that observation $X_{t,p}$ is part of peer group k. The aggregated behavioral volatility measure is sum øf relative changes in association between a doctor and the peer groups. If a doctor's behavior stayed exactly the same between time *t* and *t-1*, then all distances $M_{k,t,p}$ and $M_{k,t-1,p}$ will be the same. Therefore $\sum_k^K \left(M_{k,t,p} - M_{k,t-1,p}\right)^2 \approx 0$. If there are differences between *t* and *t-1*, we end up with a non-zero amount. Because the GMM consists of a series of PDFs and each of those PDFs is modified by vector ϕ such that the GMM integrates to 1, the squared sum of differences between any two consecutive time periods will be bound between zero and one.

Table 3: Doctor to peer group Mahalanobis distances

DOCTOR ID	MONTH	PEER GROUP 1	PEER GROUP 2	PEER GROUP 3	PEER GROUP 4
001	4	0.0	0.0	0.3	0.0
002	4	0.1	0.1	0.0	0.0
003	4	0.0	0.0	**0.3**	**0.0**
001	5	0.0	0.0	0.4	0.0
002	5	0.1	0.1	0.0	0.0
003	5	0.0	0.0	0.0	**0.3**

Using Table 3, we can see how the Mahalanobis distance might look for the example doctors' signatures from Table 2. Doctor 003, due to the change in overall signature, is depicted as shifting relative peer group associations.

The final output of the model is a mean volatility metric that aggregates how unstable a doctor's behavior is across the observation range. The result is a quantification of behavioral volatility that can then be incorporated an indicator to prioritize risk.

Experimental Design

For the purposes of demonstrating how the proposed behavioral anomaly model might perform on real data, we set up an experimental simulation with 2,000 doctors. We then inject particular types of behavior anomalies among a 10% subpopulation and test how well the proposed model detects them.

The simulation starts by defining 10 peer groups, each with a distinct profile of procedures prescribed by those in the peer group. Each peer group is assigned a prominence such that some peer groups are larger and more prevalent than others. Two of the peer groups are assigned points at which they shift behavior over a period of time to mimic real word shifts in practices, policies, or the introduction of new techniques. Peer group 0 was selected to change behaviors between period 10 and 15 and peer group 2 was selected to change between periods 30 and 40. See Figure 3 for the initial peer group profiles and Figure 4 shows the transitions for peer group 0 and 2.

The simulation builds 2,000 doctors, first assigning a mean monthly claim volume per doctor selected from

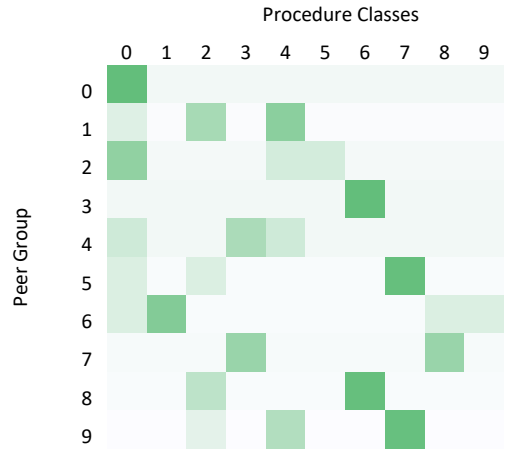

Figure 3: Peer group profiles by probability of procedural class darker greens are more prevalent procedures. Each procedure class contains 10 procedures that can be prescribed.

Figure 4: Peer group profiles set to change over time during the simulation

Copyright © 2019 Deloitte Development LLC. All rights reserved.

$Lognormal(\mu = \ln(2000), \sigma = \ln(1.5))$. Each doctor is then assigned a peer group from which a distinct signature is generated such that any two peer doctors have similar but not identical profiles.

We selected a random 10% subpopulation to assign fraudulent behavior. Fraudulent behavior is given three parameters: the volatility of switching into and out of a fraud state, the speed by which they transition into and out of a fraud state, and the type of fraud state. The volatility is determined by a Markov transition matrix where state change to and from fraud states is determined as between .05 and .5 probabilities. Transition speed is determined to be abrupt (no transition) or gradual (4 period linear transitions). Lastly, there are three fraud states that are randomly assigned to any fraudster doctor:

- Single procedure overutilization: Here a single procedure is promoted far above what it would otherwise be in its peer group while all other activities remain proportionally the same. *Example: a physician starts pushing a particular opioid drug unnecessarily on patience because he receives a kickback from a pharmaceutical company.*
- Profile shifting: A doctor switches from the expected peer group to a profile similar to another peer group. *Example: a pediatrician starts to perform live births but is not a licensed Ob/Gyn.*
- Outlier shifting: A doctor switches from their expected peer group profile to a non-peer group.

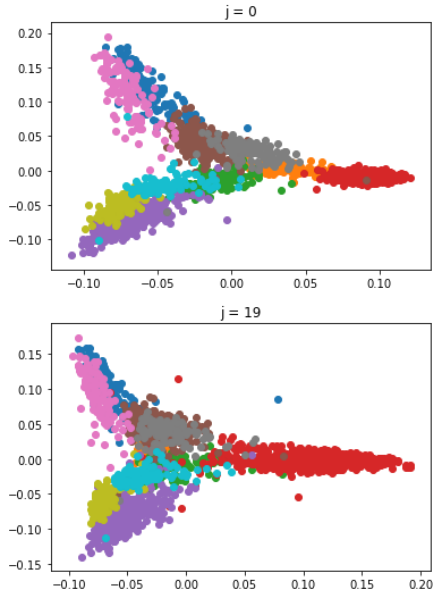

Figure 5: Two-dimensional projections of simulated data with color coded peer groups

Once the 2,000 doctors have been defined, data is simulated over 60 periods to emulate five years of monthly data. See Figure 5 for a 2-dimensional projection of period 0 and period 19 with color coding of the 10 peer groups. Shifts in the peer groups are attributable to peer group 0's shift between period 10 and period 15.

Once the data is simulated, we can then use the resulting dataset to test the ability of different models to isolate fraudulent doctors we injected into the population. We standardize performance using Receiver Operating Characteristic (ROC) curves and corresponding Area Under the Curve (AUC) across the different models and different hyper parameters settings.

Results

Using the simulated data described in the prior section, we ran a traditional Gaussian mixture clustering model to compare against our proposed entity-action temporal clustering model. The traditional GMM approach takes the results of the normalized action matrix and builds a unitary cluster

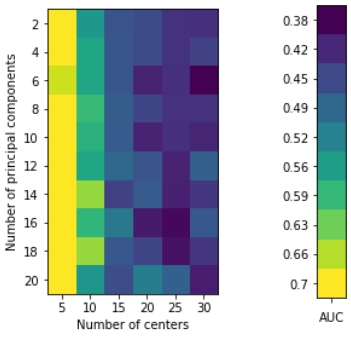

Figure 6: Gaussian mixture model behavioral approach

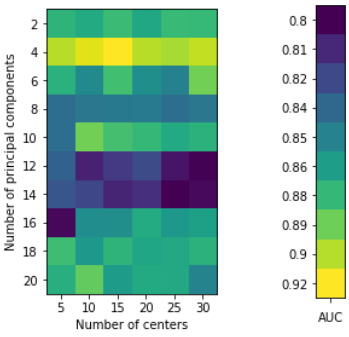

Figure 7: Entity action temporal clustering model

Copyright © 2019 Deloitte Development LLC. All rights reserved.

model over top all periods. It then looks to see how often an entity is found outside expected cluster centers.

The entity-action temporal clustering model easily beats the non-temporal GMM outlier model. We found top performance from the hyperparameter grid search of the basic behavior GMM to be an AUC of 0.70 when we use 5 Gaussian centers and 20 principal components. Using our entity-action temporal clustering approach, performance tops out with an AUC of 0.94 using 15 centers and 4 principal components. See Figures 6 and 7 for full details of the grid search. It is interesting to note that the traditional Gaussian clustering approach seems to focus on having a much smaller number of clusters and it is relatively unaffected by how much we reduce dimensionality. Conversely, the temporal behavioral volatility approach recognizes more centers of activity and needs lower rank data matrix to accomplish a higher performance.

To explore the hyperparameter space for both the traditional model and our proposed model, we took a standard grid search approach between three hyper parameters: regularization, number of clusters and number of principal components after PCA. We then produced ROC curves and precision-recall curves to compare performance at correctly identifying the providers with fraudulent tendencies.

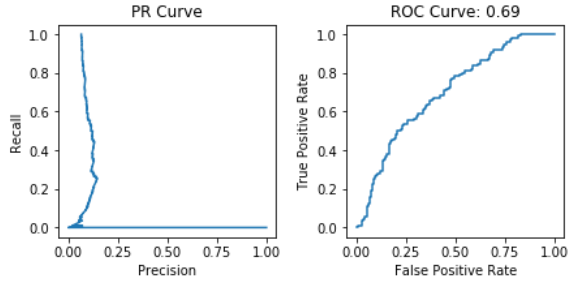

Figure 8: Performance curves from the best case of the traditional Gaussian mixture model approach

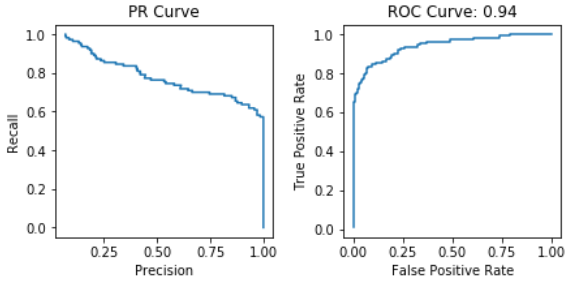

Figure 9: Performance curves from the best case of the proposed entity-action temporal clustering model

We can see the ROC curve and precision-recall curves for the temporal volatility model in Figure 8. In particular the PR curve is strong considering we have a low frequency event – we injected fraudulent activity in less than 10% of the total population. If we break down performance into the different fraud volatility schemes, we see variation in performance. All of the different schemes show good performance. Unsurprisingly, the single drug augmentation and the gradual shifts have the worst performance as they make the anomalous behavior harder to distinguish. Profile outlier shift where a provider shifts to a completely unknown profile is very easy to identify.

Fraud Volatility Type	*AUC GMM*	*AUC Volatility*
Single procedure overutilization	0.63	0.90
Profile peer group shift	0.62	0.95
Profile outlier shift	0.88	0.99
Abrupt shift	0.64	0.98
Gradual shift	0.74	0.91

Copyright © 2019 Deloitte Development LLC. All rights reserved.

Conclusions

It is worth noting that there are several potential weaknesses to the approach. One weakness is that it focuses minimally on detecting shifts in behavior. If an entity is consistently in a fraud state or if fraud is overwhelmingly common, then this approach can break down. However, based on the AUC results from the simulation described herein, our proposed model performed better than a traditional static cluster approach. The temporal clustering concept helps to capture erratic behavior increasing positive identification while allowing the model to adjust for legitimate changes in peer groups that we might expect in many systems reducing false positives.

The example we center on in this paper is a system of doctors. The approach can be expanded beyond systems of doctors to find potentially anomalous behavior in clinics, hospitals, and other health care entities. However, we can also think of this approach beyond health care to cover other systems of entities such as expenses among employees, systems of vendors for large corporations, or other situations where there is a system of complex entities that can be tracked over time. In future research, it might make sense to explore other clustering algorithms in an effort to differentiate peer groups and stress test different simulated scenarios. Any well-rounded fraud identification system should consider assessing potential fraud from multiple angles. Adding more unsupervised approaches, such as the temporal entity action model discussed in this paper, should nicely complement a solution's fraud/anomaly tool belt.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee ("DTTL"), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as "Deloitte Global") does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the "Deloitte" name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms.

References

[1] H. Lyu, "Overtreatment in the United States," *PLoS One,* September 2017.

[2] R. A. Bauder and T. M. Khoshgoftaar, "The Detection of Medicare Fraud Using Machine," in *The Thirty-First International Florida Artificial Intelligence Research Society Conference*, 2018.

[3] G. v. Capelleveen, Poel, Mueller and Hillegersberg, "Outlier detection in healthcare fraud: A case study in the Medicaid dental," *International Journal of Accounting Information Systems,* p. 2016, June.

[4] E. Liu, M. A. Ahmad, C. Eckert, A. Nascimento, M. De Cock, K. Padthe, A. Teredesai and G. McKelvey, "Automatic Detection of Excess Healthcare Spending and Cost Variation in ACOs," *SAIM,* 2018.

Copyright © 2019 Deloitte Development LLC. All rights reserved.

[5] R. A. Bauder and Et.al., "Identifying Medicare Provider Fraud with Unsupervised Machine Learning," in *IEEE International Conference on Information Reuse and Integration for Data Science*, 2018.

[6] Y. Gao, C. Sun, R. Li and Q. Li, "An Efficient Fraud Identification Method Combining Manifold Learning and Outliers Detection in Mobile Healthcare Services," *IEEE Access,* vol. 6, 2018.

[7] A. Gangopadhyay and S. Chen, "Health Care Fraud Detection with Community Detection Algorithms," in *IEEE International Conference on Smart Computing (SMARTCOMP)*, St. Louis, 2016.

[8] J. Liu, E. Bier, A. Wilson, J. A. Guerra-Gomez, K. Sricharan, L. Gilpan, G. Leilani and D. Davies, "Graph Analysis for Detecting Fraud, Waste, and Abuse in Health-Care Data," *AI Magazine,* no. Summer, p. 33, 2016.

[9] S. Rani and G. Sikka, "Recent Techniques of Clustering of Time Series Data: A Survey," *International Journal of Computer Applications,* vol. 52, no. 15, 2012.

[10] S. Soheily-Khah, A. Douzal-Chouakria and E. Gaussier, "Generalized k-means-based clustering for temporal data under weighted and kernel time warp," *Pattern Recognition Letters,* vol. 75, pp. 63 - 69, 2016.

[11] A. Zhou, F. Cao, W. Qian and C. Jin, "Tracking clusters in evolving data streams over sliding windows," *Knowledge and Information Systems,* vol. 15, no. 2, pp. 181-124, 2008.

Copyright © 2019 Deloitte Development LLC. All rights reserved.