Meta-Learning Approach for Credit Card Fraud Detection

Moorissa Tjokro
mmt2167, moorissa.tjokro@columbia.edu

Abstract

Credit card fraud continues to be a significant cost for financial institutions and the advancement of fraud detection can provide significant savings for the financial industry.

The first research paper explored the application of Bayesian network in detecting credit card fraud. This final paper digs further into the problem and limitations of the original method. The paper discusses a proposed meta-learning solution that improves Bayesian Network at both high-level and system-level views. This solution incorporates k-nearest neighbor and decision tree techniques in the implementation of Bayesian network.

In addition to discussing the methodology of the proposed solution, this essay studies other attributes of meta-learning implementations, from the advantages and challenges to key design considerations and performance comparison to state-of-the-art solutions in the market. Finally, we will see what future work can be done for optimizing credit card fraud detection.

1. Overview

Rapid growth of technology over the past decade has created proliferation of credit card use worldwide. The U.S. Department of Justice noted approximately $7.60 billion credit card fraud occurred in 2010. This number increased by over three times as large this year, and is projected four times as large, at $32 billion loss, in 2020 (Figure 1) [1].

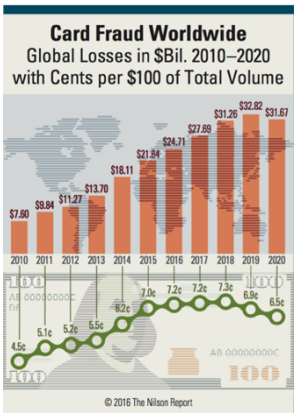

Figure 1

As it brought huge devastations to businesses in finance industry, major companies took actions to prevent further losses. Among which are Bank of America agreed to pay $16.5 billion for resolving financial fraud case [2], and IRS began to observe professionals and academicians who committed fraud, e.g. founder of Bixby Energy Systems deceived more than 1,800 investors and committed multi-million dollar fraud [3].

With billions of dollars incurred by financial institutions every year due to such crime, detecting fraudulent behaviors in credit card transaction systems has been of significant importance.

Detailed implementation of the proposed meta-learning approach for fraud detection is studied along with description of how each of the components works. This includes solution's methodology, advantages, and challenges in improving the existing Bayesian network.

1.1. Problem Description

Credit card fraud detection is the process of identifying credit card transactions that are fraudulent given a set of all purchases and transfers made by one individual using another individual's credit card. The statistics in Figure 1 indicates how credit cards are such a popular target for fraudulent transactions, mainly for earning a lot of money in a very short time and such crime is commonly discovered a few weeks after.

When banks lose money due to credit card fraud, card holders are partially or entirely responsible for the loss. Individuals can either be charged through higher interest rates, higher membership fees, or reduced benefits. It is interest of both financial institutions and card holders to minimize illegitimate use of credit cards. This is why banks and financial corporates started implementing fraud detection in the first place.

1.2. Types of Fraud Techniques

Credit card frauds in the financial sector can be broadly classified into three categories: traditional, merchant, and internet related frauds [4].

A. Traditional credit card frauds

Among the most prevalent techniques in traditional credit card fraud is copying a credit card while holding the secret pin code of the user and charging more money

that the amount that the user agreed, without them being aware of it.

B. Merchant related credit card frauds

This type of fraud usually stems from merchant establishment's owners and their employees. They would pass on the cardholder accounts to fraudsters illegally or sometimes through a 'false' website.

C. Internet credit card frauds

With Internet becoming more popular for credit card fraud, people use more of site cloning models to commit fraud (Figure 2). Site cloning occurs when fraudsters close an entire site from which the customer made a purchase. They will then send a customer receipt like a real company would do. The customer suspects nothing while fraudsters have all the details they need to commit credit card fraud.

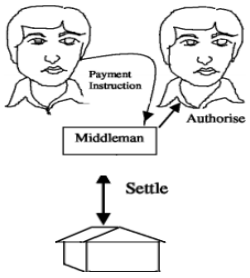

Figure 2

Not only does internet provide free access, but also operations on an international level. This allows them to expand their trans-border beyond just countries, but also economic and political spaces.

2. Limitations of Bayesian Network

Bayesian network technique was discussed in the first paper as the primary means for credit card fraud detection [5]. In order to provide an improved version of the existing approach, this section highlights key limitations of Bayesian network and splits it into two categories: (1) based on the natural characteristics of Bayesian network (internal), and (2) based on its implementation, adoption, or regulation challenges (external).

2.1. Internal

A disadvantage of an approach involving Bayesian network is the fact that there is no universally accepted method for constructing a network from raw data [6]. Two specific weaknesses come out of this lack, the first being that the design of a Bayesian network requires a comparatively large amount of effort. Secondly, the resulting problem causes Bayesian network to be able to only exploit causal influences that are recognized by the person programming it [7].

2.2. External

Key challenges in the *implementation* include:

1. Financial companies don't share their data for a number of reasons. Although Bayesian network deals well with missing information, it requires at least enough amount of reliable data to implement and measure effectiveness.
2. Databases that companies maintain on transaction behavior are huge and growing rapidly. This requires Bayesian network approach to adjust flexibly with demand scalable machine learning systems [8].
3. Easy distribution of models in a networked environment is important to keep the models updated.

In regards to the *adoption*, extensive effort is required for development of this approach. Developing Bayesian network models not only requires collaboration with domain experts but also an extensive iterative development process. Even though the method presents and covers a range of techniques for reducing the burden of expert elicited models, this up-front development effort remains the primary barrier to more widespread adoption of Bayesian networks [5].

Looking at the *regulatory* challenges, with an upsurge in financial accounting fraud in the current economic scenario experienced, financial accounting fraud detection (FAFD) have received considerable attention from the investors, academic researchers, media, the financial community and regulators [5].

3. Meta-Learning Approach

The proposed model for improving the accuracy of Bayesian network approach and addressing some of its limitations is a meta-learning model, consisting of three base classifiers: Bayesian network itself, k-nearest neighbor, and decision tree.

3.1. High-Level Overview

There are four main stages in the meta-learning process. The first stage establishes base classifiers using a training dataset, consisting of 50% fraudulent transactions and 50% legitimate transactions (Figure 3). This can be done on a monthly basis for the first 8 months, for instance between December 2015 to July 2016. This period is where all of the fraudulent transactions for the given month were matched with an equal number of randomly chosen legitimate transactions (Figure 4).

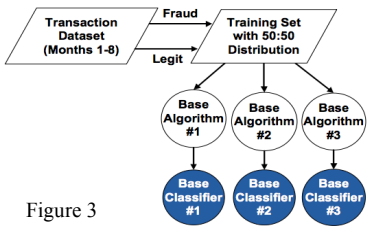

Figure 3

In the second stage, the base classifiers are applied to a validation dataset to generate base predictions. The validation dataset consists of all of the transactions

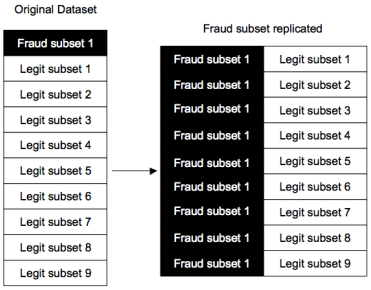

Figure 4

between month 9 and 10. In this example, it would cover the period between August and September 2016.

The predictions from the second stage are then combined with the validation dataset in Stage 3, and a meta-algorithm is applied to this combined dataset to construct a meta-classifier (Figure 5). Since algorithms are run independently, order does not matter in this case. For example, we can use Bayesian network as base classifier

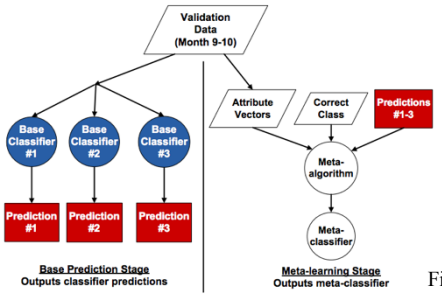

Figure 5

#1, k-nearest neighbor as base classifier #2, and decision tree as base classifier #3. Studies show that the Bayesian network algorithm highlights the improvement for this approach [9].

Finally, in stage 4, the forward predicting test stage, the meta-classifier is applied to the testing dataset (October 2016) to produce the predictions (Figure 6). These predictions are then compared to the existing system predictions to see if this meta-learning approach can improve on fraud detection.

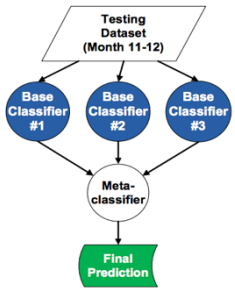

Figure 6

3.2. System-Level View: Basic Building Blocks

As seen in the high-level overview, each prediction (#1, #2, and #3) are produced as a result of running the base algorithms. The following subsections provide a system-level view of the building blocks of the proposed solution, which consist of Bayesian network, k-nearest neighbor, and decision tree techniques.

3.2.1. Bayesian Network

Bayesian network is constructed to model behavior that has been assumed the user is fraudulent and second model under the assumption that the user is a legitimate [10]. The fraud net is set up by using expert knowledge. The user net is set up by using data from non-fraudulent users. By inserting evidence to these networks, the result of any transaction has been classified as fraudulent or non-fraudulent behavior.

A Bayesian Network for Detecting Credit-Card Fraud

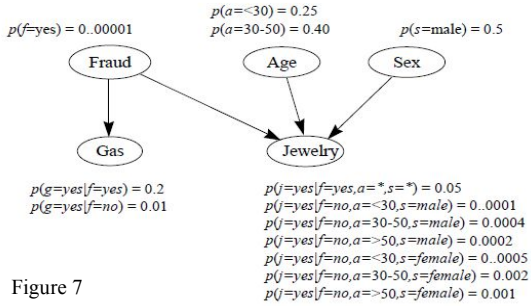

Figure 7

Figure 7 shows a Bayesian network representing joint probability distribution for a fraud detection scene.

The graphical diagram represents relationships and influences among nodes, with some noted instances:

- Direction of arcs: parent node → descendant node
- Parents of node Xi: Pai
- Pa(Jewelry) = {Fraud, Age, Sex}

In the probability of fraud=P (F) then P (NF) =1-P (F) in general and by applying Bayes rule, it gives the probability of fraud for any incoming transaction [11]. The fraud probability that has obtained of training can be used as an alarm level. By classifying patterns expressed in probabilistic terms between predictors and outcome variables, 'Prediction #1' can easily be achieved.

3.2.2. K-Nearest Neighbor

K-Nearest Neighbor (KNN) algorithm is a supervised learning algorithm where the result of new instance query is classified based on majority of k-nearest neighbor category [12]. The technique helps improve the Bayesian network's performance through (1) the distance metric used to locate the nearest neighbors, (2) the distance rule used to derive a classification from k-nearest neighbor, and (3) the number of neighbors used to classify the new sample [13].

KNN-based credit card fraud detection techniques require a distance or similar the measure defined between two data instances [14]. An incoming transaction is classified by calculating of nearest point to new incoming transaction. Then if the nearest neighbor be fraudulent, then the transaction indicates as a fraud. The value of K is used as, a small and odd to break the ties. Larger K values would reduce the effect of noisy data set. In this algorithm, distance between two data instances can be calculated in different ways, e.g. Euclidean distance [15].

Using legitimate as well as fraudulent samples of data for training, the 'Prediction #2' output can be achieved.

3.2.3. Decision Tree

Since Bayesian network heavily exploits causal influences that are recognized during credit card transactions, it has less flexibility in handling the imbalanced data distribution problem. Decision tree provides an effective way of overcoming this limitation by using its cost-sensitive attribute [16].

A cost-sensitive decision tree induction algorithm is useful to identify fraudulent behaviors in any given credit card transactions [16]. In decision tree, the sum of the costs of the child nodes are divided by the number of child nodes after the split so that there will not be a bias to select the variables resulting in more split nodes than the ones resulting in fewer split nodes (Figure 8). Every possible split for each input variable is used in the search for the candidate splits for the best cost reduction, and the split which gives the best cost reduction in the child level is chosen as the split for the node [17].

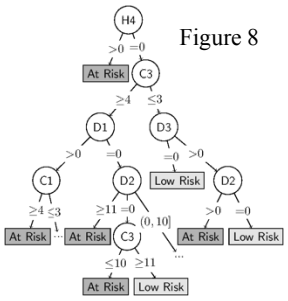

Figure 8

Using misclassification cost calculation of the nodes, both the class of the node and the probabilities of the transactions in the node—whether it is predicted to be fraudulent or normal—are accurately found. The 'Prediction #3' is then also achieved.

3.3. Key Design Considerations

The methodology applied in this paper is developed from a meta-classifier framework introduced by Chan and Stolfo in 1997 [18]. The approach combines results from multiple learners like k-nearest neighbor and decision tree to improve prediction accuracy and utilize their strengths for complementing limitations in the Bayesian network approach (see 2. Limitations of Bayesian Network).

Another key consideration is that choices of the classifiers have been proven accurate through previous studies. For instance, among the various credit card fraud detection methods of supervised statistical pattern recognition, the KNN achieves consistently high performance, without prior assumptions about the distributions. High performance individual algorithm would bring out strength that improves the existing solution [13].Base classifiers were selected based on a diversity metric (Chan, 1997). This entropy-based metric measures both the randomness of the predictions and how different the base classifiers are. The higher the diversity of the base classifiers, the more evenly distributed the predictions are, and thus gives higher accuracy for detecting fraud [19]. The calculated diversity values were then compared between different combinations of algorithms. The best algorithm with highest diversity value was found to be a combination between Bayesian network, k-nearest neighbor, and decision tree [20].

3.4. Novel Aspects

The proposed meta-learning solution is different than the existing Bayesian network in terms of the number of parameters used in the approach. By incorporating two additional models in the meta-learning solution, accuracy have certainly improved.

Research shows that through involving k-nearest neighbor and decision tree in the Bayesian network, the results become much reliable, as the strengths in other algorithms complement for weaknesses in the other. Compared to Bayesian network approach, the proposed solution is able to:

- Achieve a reduction of 20% to 40% in total credit card fraud losses [21].
- Detect credit card fraud in real time.
- Implement easily with commercial databases.
- Quickly and accurately classify transactions.

There are more functionalities and strengths through combining different classifiers, which adds new advantages to the proposed solution.

4. Advantages

This section discusses the advantages of the proposed meta-learning approach by comparing it with state-of-the-art solutions in the market and articulating its novel aspects.

4.1. Comparison with Market Solutions

Many recent research studies focused on applying multiple algorithm techniques in credit card fraud detection. The proposed meta-learning approach was tested through an experiment to see which performs better in terms of fraud catching and false alarm rates. According to Chan and Stolfo, the proposed solution has a fraud catching rate (true positive) of 80% and alarm rates (false positive) of 17% [19].

Brause et al (2011) combined a rule-based technique with a neural network to identify fraudulent credit card transactions. Studies found that the fraud catching rate was 74% with a false alarm rate of 17% [22]. Phua et al (2008) proposed the use of back-propagation neural networks, naïve Bayesian, and C4.5 algorithms as base classifiers, and to combine the base classifiers' predictions using a meta-classifier technique to detect fraudulent automobile insurance claims. Its fraud

catching rate was 80% with a false alarm rate of 16% [23]. Duman and Ozcelik (2010) used a novel combination of the genetic algorithm and the scatter search algorithm to detect credit card fraud in a large Turkish bank. Its fraud catching rate is found to be 76% with a false alarm rate of 16% [24].

The abovementioned methods have all relatively low false alarm rates, but the proposed meta-learning has the highest fraud catching rate, indicating that our proposed method is one of the best for any fraud detection cases.

Phua's method may have a slightly lower false positive rate because it incorporated back-propagation neural network, which studies show as one of the most studied and used method for fraud detection [22].

4.2. Novel Aspects

In addition to statistical measures, the proposed meta-learning solution has an advantage over recent methods. Dataset as an input for base classifiers can be summarized into a data structure that embeds the complexity and performance of the induced training dataset. The resulting representation can serve as a basis to explain the reasons behind the performance of the learning algorithm. As an example, one can build a decision tree from a dataset and collect properties of the tree (e.g. nodes per feature, maximum tree depth, shape, tree imbalance, etc.), as a means to characterize the dataset.

5. Challenges

This section discusses the challenges faced in the implementation, adoption, and regulation for credit card fraud detection.

5.1. Implementation

In order to produce the proposed meta-learning solution, it is necessary to perform an empirical evaluation (e.g. cross-validation) of the candidate algorithms on a problem. Hence, the cost of generating a whole set of meta-examples may be high, depending on the number and complexity of the candidate algorithms, the methodology of empirical evaluation, and the amount of available problems [25].

Predictable factors such as the available amount of training data (relative to the dimensionality of the feature space), the spatial variability of the effective average distance between data samples, and the type and amount of noise in the data can set influence such classifiers to a significant degree [26]. Hence, the implementation has to be conducted carefully.

5.2. Adoption

A key limitation of the current meta-learning strategy is the lack of effective metrics to guide the adoption of base classifiers that will produce the best meta-classifier [27].

The approach for adopting the proposed meta-learning algorithms also generally involve costly trial-and-error procedures, or require expert knowledge, which is not always easy to acquire [28].

Meta-learning approaches for automatic algorithm adoption sometimes assume that the features used to represent meta-instances are sufficiently relevant. However, some features may not be directly relevant, and some features may be redundant or irrelevant. An attribute to successful adoptions would be to create framework in which accuracy was measured and achieved on a limited number of datasets, limited number of classifiers, and their parameter settings [29].

5.3. Regulation

Major operational limitations involve regulatory breach, such as regulatory reporting or account segregation. This includes whether the unit operates in a tight regulatory environment with multiple legal entities and global reach, or the unit has a loose regulatory environment with few legal entities and a local or national focus [30].

Financial institutions can also potentially face a regulatory issue as it often develops custom fraud detection systems targeted to their own asset bases. Most of these systems employ advanced machine learning and statistical analyses to produce pattern-directed inference systems [31]. Using models of anomalous behaviors, the proposed solution would require analysis of large and inherently distributed databases of information about transaction behaviors to produce models of fraudulent credit card transactions.

6. Conclusion

With meta-learning algorithms emerged as a powerful data mining technique for detecting credit card fraud, this research study confirms that the proposed solution would create optimal for handling uncertainty in complex domains, classifying, and recognizing patterns of fraudulent transactions.

This paper has shown that single-learning approaches, such as Bayesian networks, can always be improved using meta-learning techniques. Meta-classifier always has better performance than any of its constituent.

The model can be further improved in the future through studies, experiments, and reuse of previous experience or from analysis of other problems, Improved methods would eventually relieve humans from most of the work and realize the goal of computer programs that perform an accurate credit card fraud detection.

7. References

[1] Mizes, Howard. "Credit Card Fraud Solutions." SciVee (n.d.): n. pag. Card Fraud. Rocherster University, Computer Science Department, 3 Dec. 2013. Web. 15 Nov. 2016.

[2] Overfelt, Maggie. "Mad Hacker Rush to Create $14B in Fraud before New Cards Take over." CNBC. CNBC, 06 May 2016. Web. 15 Nov. 2016.

[3] Rajesh Parekh, Jihoon Yang, and VasantHonavar, "Constructive Neural-Network Learning Algorithms for Pattern Classification" IEEE 2000.

[4] Franfurt, Gorby, Jaakkola, and Young, R.A. "Combining location and expression data for principled discovery of genetic regulatory network models." Pac Symp Biocomput, 2002.

[5] Tjokro, Moorissa. "Bayesian Networks for Credit Card Fraud." *Midterm for COMS W4995: Topics in Computing, Fall 2016*. Ed. Moorissa M. Tjokro. New York, 2016. 15 Oct. 2016.

[6] Patidar, Raghavendra, and Lokesh Sharma. Credit Card Fraud Detection Using Neural Network. 2231-2307 ed. Vol. 1. Jaipur, India: International Journal of Soft Computing and Engineering (IJSCE), June 2011. Print. Issue-NCAI2011.

[7] Jarryd-Lee, Mandy. "A Geek's Guide to Machine Learning and Risk Analytics and Decisioning | Provenir." Provenir A Geeks Guide to Machine Learning and Risk Analytics and Decisioning Comments. Provenir, 06 Oct. 2016. Web. 13 Nov. 2016.

[8] Sohn, S.Y. Meta Analysis of Classification Algorithms for Pattern Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(11): 1137-1144, 1999.

[9] Luttrell, S. P. "Partitioned mixture distribution: An adaptive Bayesian network for low-level image processing." IEE Proc Vision, Image Signal Process, 141 (4) (1994), pp. 251–260.

[10] Eisenstein, E.L., Alemi. A comparison of three techniques for rapid model development: an application in patient risk-stratification. Proc/AMIA Annu Fall Symp (1996), pp. 443–447.

[11] Leamont, Maggie. "Mad Hacker Rush to Create $14B in Fraud before New Cards Take over." CNBC. CNBC, 06 May 2016. Web. 15 Nov. 2016.

[12] Lagon, Gifford, Jaakkola, and Young, R.A. "Combining location and expression data for principled discovery of genetic regulatory network models." Pac Symp Biocomput, 2002.

[13] Bearn, T. S. Jaakkola, and R. A. Young. "Difference between Neural Network and Bayesial Learning". Pac Symp Biocomput, 2001.

[14] A. J. Hartemink, D. K. Gifford, T. S. Jaakkola, and R. A. Young. "Using graphical models and genomic expression data to statistically validate models of genetic regulatory network." Pac Symp Biocomput, 2001.

[15] Heckerman DE. "A tutorial on learning with Bayesian network." MSR-TR-95-06. 1996. Redmond, WA, Microsoft Research.

[16] Todorovski, L., Dzeroski, S. Combining Classifiers with Meta Decision Trees. Machine Learning 50 (3), 223-250, 2003.

[17] P. K. Chan and S. J. Stolfo, "Experiments in Multistrategy Learning by Meta-Learning." Proceedings of the second international conference on Information and knowledge management, pp. 314-323, 1993.

[18] McCloskey, J. "Credit Card Fraud Detection Using Bayesian and Neural Network." First International NAISO Congress on Neuro Fuzzy Technologies, Havana, Cuba. 2002.

[19] Chan, Liu. "Credit Card Fraud Solutions." SciVee (n.d.): n. pag. Card Fraud. Rocherster University, Computer Science Department, 3 Dec. 2013. Web. 15 Nov. 2016.

[20] R. Brause, T. Langsdorf, M. Hepp "Neural Data Mining for Credit Card Fraud Detection, "International Conference on Tools with Artificial Intelligence; (1999). (103-106).

[21] Widmer, G. Recognition and Exploitation of Contextual Clues via Incremental Meta-Learning. In Proceedings of the Thirteenth International Conference on Machine Learning (ICML-96), 1996B.

[22] Widmer, G. Tracking Context Changes through Meta-Learning. Machine Learning, 27(3): 259-286, 1997.

[23] Jitendra Dara,Laxman Gundemoni, "Credit Card Security And E-Payment." 2006. The New England Debit Card Task Force "Best Practice Guide for Managing Debit Card Fraud." IEEE July 2005.

[24] David J. Montana, "Neural Network Weight Selection Using Genetic Algorithms." Bolt Beranek and Newman Inc. July 2003.

[25] Philip K. Chan, Wei Fan, Andreas L.Prodromidis, and Salvatore J, "Distributed Data Mining in Credit Card Fraud Detection" IEEE. December 1999.

[26] Sushmito Ghosh and Douglas L. Reilly, "Credit Card Fraud Detection with a Neural-Network." Nestor, Inc. IEEE (1994).

[27] Rajesh Parekh, Jihoon Yang, and VasantHonavar, "Constructive Neural-Network Learning Algorithmsfor Pattern Classification" IEEE 2000.

[28] Mubeena Syeda,Yan-Qing and Yi-Pan,"Parellel Granular Network For Credit Card Fraud Detection". IEEE 2002.

[29] Erik Bothelius, "Fraud detection in the Internal Account System for Payment Service Providers." May 8, 2005.

[30] D.WHITLEY,"Genetic Algorithm And Neural Network."2003.

[31] Ting, K. M., Witten I. H. Stacked generalization: When does it work?. In Proceedings of the 15th International Joint Conference on Artificial Intelligence, pp 866-873, Nagoya, Japan, Morgan Kaufmann, 1997.