

MỘT SỐ THUẬT TOÁN ỨNG DỤNG TRONG PHÁT HIỆN GIAN LẬN THANH TOÁN THẺ TÍN DỤNG

Nguyễn Thị Hồng Loan

Trường Đại học Tài nguyên và Môi trường Hà Nội

Tóm tắt

Hiện nay, hình thức thanh toán bằng thẻ tín dụng (dưới dạng thẻ vật lý hoặc online) đang được ưa chuộng. Đi cùng với sự nở rộ của hình thức thanh toán này là các loại tội phạm về gian lận và lừa đảo trong thanh toán thẻ. Các ngân hàng cần phải xây dựng các hệ thống phòng ngừa và kiểm tra gian lận trong mỗi giao dịch. Khai phá dữ liệu là một trong những kỹ thuật được ứng dụng trong các hệ thống này. Bài báo trình bày một số thuật toán trong khai phá được đánh giá là hiệu quả nhất trong các hệ thống phát hiện gian lận giao dịch thẻ tín dụng.

Từ khoá: Phát hiện gian lận; Khai phá dữ liệu

Abstract

Some algorithms applied in credit card fraud detection

In recent years, credit card transactions including physical credit card payment and online payment are increasingly used in many countries in the world. This convenience payment method attracts more and more crimes, especially credit card frauds. Consequently, many banks in the world have developed systems that prevent and detect fraud for each credit card transaction. Data mining is one of techniques applied in these systems. This paper analyses some popular algorithms used in detecting credit card fraud systems.

Keywords: Fraud detection; Data mining

1. Giới thiệu

Các hình thức thanh toán của ngân hàng ngày càng đa dạng và thay đổi theo xu hướng phát triển của công nghệ. Cụ thể, hầu hết các ngân hàng ở Việt Nam đã cho phép khách hàng thanh toán trực tuyến thông qua dịch vụ ngân hàng điện tử hoặc qua thẻ tín dụng. Trong đó, thẻ tín dụng được coi là phương thức thanh toán ngày càng phổ biến và tiện lợi. Theo Nilson Report, số lượng giao dịch bằng thẻ tín dụng trên toàn thế giới ngày càng tăng, đến năm 2016 có đến 257,17 tỷ giao dịch liên quan đến thẻ tín dụng [1]. Chính vì khách hàng ngày càng ưa chuộng hình thức thanh toán qua thẻ tín dụng nên số vụ lừa đảo, gian lận liên quan đến loại hình thanh toán này có xu hướng tăng. Theo số liệu của Nilson, năm 2015 cả thẻ giới bị mất 28,8 tỉ đô la do các lừa đảo hoặc gian lận về thẻ tín dụng. Trong đó số vụ gian lận ở Châu Á Thái Bình dương là lớn nhất [2]. Tuy nhiên, một

vài năm trở lại đây, các vụ gian lận về thanh toán thẻ tín dụng đã hạn chế một cách đáng kể. Để làm được điều đó, các ngân hàng đã phải xây dựng các hệ thống ngăn ngừa và phát hiện gian lận trong các giao dịch, trong đó có giao dịch thẻ tín dụng. Hệ thống này hoạt động theo quy tắc chung: 1) lưu trữ thông tin giao dịch qua thẻ của khách hàng; 2) phân tích dữ liệu giao dịch và khoanh vùng phạm vi sử dụng thẻ; 3) khi một giao dịch được tiến hành, hệ thống dựa vào dữ liệu trong quá khứ để phỏng đoán xem giao dịch có gì bất thường không, từ đó đưa ra các cảnh báo cần thiết.

Các nhà nghiên cứu đã đưa ra nhiều kỹ thuật để phát hiện gian lận trong thanh toán thẻ tín dụng như: phát hiện gian lận dựa vào cây quyết định; sử dụng mạng Nơ ron để phán đoán các giao dịch bất thường; dựa vào cây quyết định để phán đoán email giả mạo;... Trong bài báo này, tác giả tập trung phân tích và giới

thiệu các thuật toán khai phá được đánh giá là hiệu quả nhất.

2. Khai phá dữ liệu và phát hiện gian lận

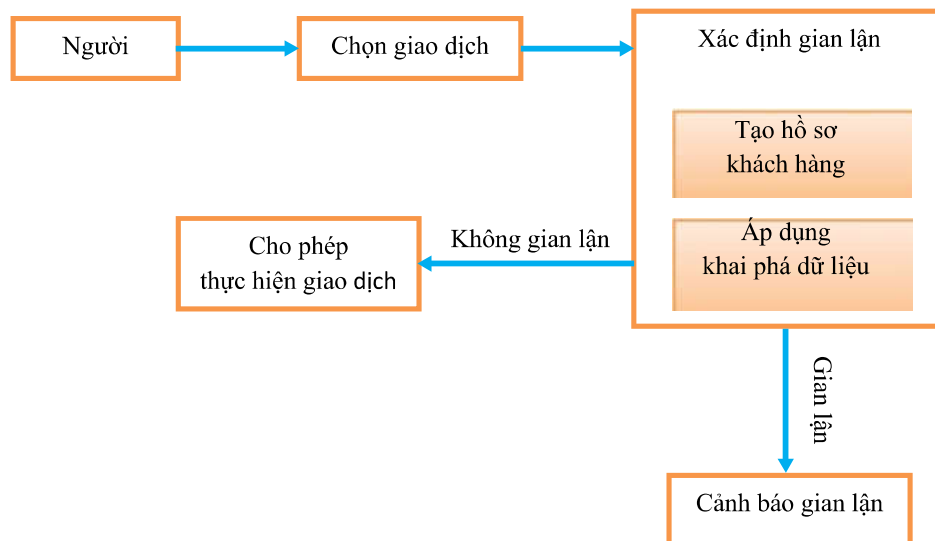
Khai phá dữ liệu là quá trình tính toán để tìm ra các mẫu trong các bộ dữ liệu lớn liên quan đến các phương pháp tại giao điểm của máy học, thống kê và các hệ thống cơ sở dữ liệu [3]. Ngân hàng là một trong những ngành có thể áp dụng rất tốt các kỹ thuật khai phá dữ liệu, bởi ngân hàng sở hữu lượng dữ liệu khổng lồ về các giao dịch cũng như thông tin liên quan đến hoạt động tiền tệ khác. Các công cụ khai phá dữ liệu hoạt động bằng cách thu thập dữ liệu từ

những nguồn, sau đó phân tích và đưa ra các mẫu để từ đó hiểu ý nghĩa của dữ liệu, đồng thời giúp nhân viên ngân hàng đưa ra quyết định.

Để xác định gian lận cần phải tiến hành qua hai bước:

- *Bước 1.* Ngân hàng thu thập các thông tin về chủ thẻ cũng như các giao dịch mà họ đã thực hiện để lập hồ sơ khách hàng.

- *Bước 2.* Thực hiện các kỹ thuật khai phá dữ liệu để nhận diện gian lận. Đối với mỗi giao dịch, nếu phát hiện dấu hiệu gian lận thì đưa ra những cảnh báo, ngược lại để các giao dịch tiến hành bình thường.



Hình 1: Quy trình phát hiện gian lận [3]

3. Một số thuật toán phát hiện gian lận

3.1. Cây quyết định

Phương pháp cây quyết định dùng để giải quyết các bài toán đệ quy hoặc phân cụm bằng cách thể hiện dữ liệu dưới dạng cây. Theo cách biểu diễn này, mỗi cây có một nút gốc, các nút trong và các nút lá. Mỗi nút được gán nhãn bằng tên các thuộc tính, mỗi cạnh được gán nhãn là giá trị của các thuộc tính. Để dự đoán giá trị của một bản ghi, cần phải duyệt cây từ nút gốc. So sánh giá trị của các thuộc tính ở nút gốc và các thuộc tính của bản ghi. Dựa vào kết quả của

phép so sánh để đi duyệt các nút kế tiếp. Quá trình so sánh và chuyển nút được thực hiện liên tục cho đến khi gặp các nút lá [4]. Cây quyết định là phương pháp dễ hiểu, dễ thực thi hơn so với các thuật toán phân cụm khác.

Đối với bài toán xác định gian lận, cây quyết định được ứng dụng trong dò tìm địa chỉ email và IP của người sử dụng thẻ [5]. Hệ thống so sánh những địa chỉ mua sắm trong quá khứ của chủ thẻ với địa điểm mua sắm hiện tại, nếu có dấu hiệu bất thường sẽ tiến hành dò tìm địa chỉ email cũng như địa chỉ IP của người đang thực hiện giao dịch.

Cây quyết định là phương pháp đầu tiên và cũng là hiệu quả nhất được áp dụng trong các hệ thống xác định gian lận trong giao dịch thẻ tín dụng.

3.2. Mô hình Markov ẩn (HMM)

Mô hình Markov ẩn (Hidden Markov Model - HMM) là mô hình thống kê trong đó hệ thống được mô hình hóa là một quá trình Markov với các tham số không biết trước và nhiệm vụ là xác định các tham số ẩn từ các tham số quan sát được, dựa trên sự thừa nhận này. Các tham số của mô hình được rút ra sau đó có thể sử dụng để thực hiện các phân tích kế tiếp, ví dụ cho các ứng dụng nhận dạng mẫu [3].

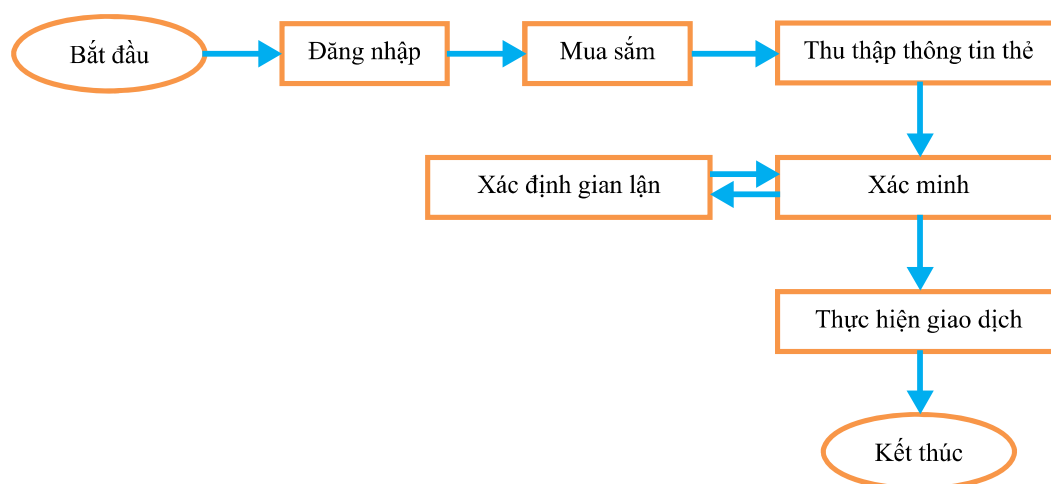
Trong phát hiện gian lận, HMM được sử dụng để phát hiện giao dịch khả nghi bằng cách tính toán hành vi tiêu tiền của khách hàng dựa trên dữ liệu giao

dịch trong quá khứ như: số tiền, địa chỉ IP, địa chỉ giao hàng, địa chỉ giao dịch gần đây nhất. Có thể phân xu hướng tiêu tiền của chủ thẻ thành ba loại:

- Tiêu tiền ít
- Tiêu tiền trung bình
- Tiêu nhiều tiền

Bước đầu tiên trong phán đoán gian lận là phân tích hồ sơ và giao dịch trong quá khứ để biết chủ thẻ thuộc nhóm nào trong các xu hướng mua sắm [6]. Quy trình xác định gian lận được tiến hành qua hai bước như sau:

- *Bước 1.* Huấn luyện HMM bằng các dữ liệu giao dịch trong quá khứ.
- *Bước 2.* Đưa dữ liệu giao dịch hiện tại vào chạy HMM để kiểm tra giao dịch hiện tại có tuân theo quy luật thông thường không.

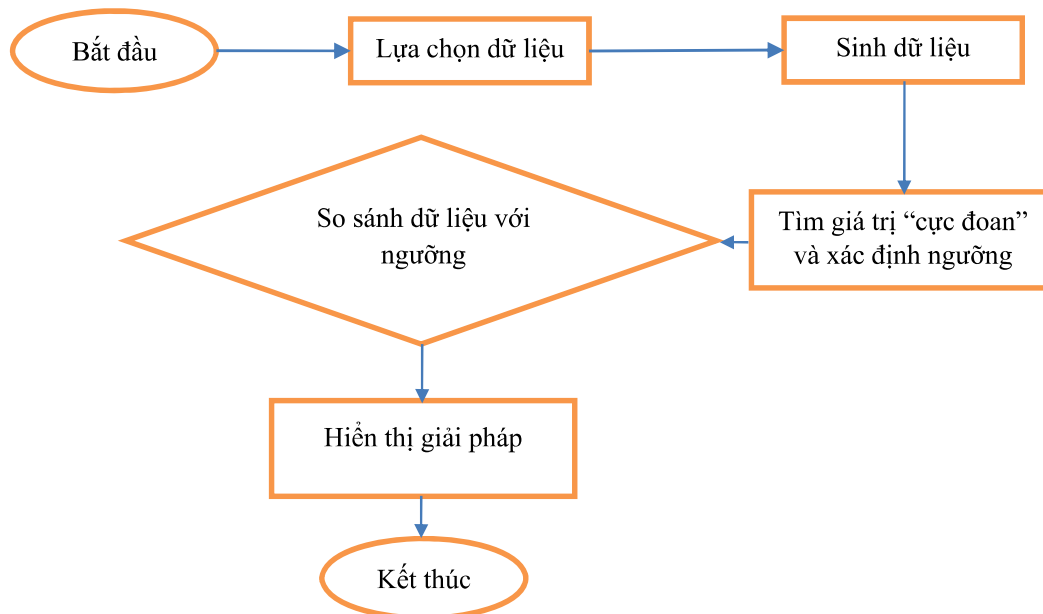


Hình 2: Xác định gian lận sử dụng HMM [6]

3.3. Giải thuật di truyền

Giải thuật di truyền là sự mô phỏng lại quá trình tiến hóa di truyền trong tự nhiên. Một cách chính xác, đó là giải thuật chỉ ra các tập cá thể được hình thành, ước lượng và biến đổi ra sao. Cụ thể là làm thế nào để lựa chọn ra cá thể tái tạo và cá thể bị loại bỏ. Mục đích chính của giải thuật là tìm ra giải pháp tối ưu và đưa ra kết luận về giao dịch nghi ngờ gian lận dựa trên kho dữ liệu của ngân hàng [7].

Bước đầu tiên trong quá trình xác định gian lận là lựa chọn bộ dữ liệu, sau đó chuẩn hóa dữ liệu và thêm các dữ liệu về thông tin chủ thẻ. Bước tiếp theo, tính giá trị “cực đoan” dựa vào mức độ sử dụng thẻ thường xuyên, số dư, mức độ thấu chi, nơi sử dụng thẻ của những giao dịch trong quá khứ. Khi một giao dịch được tiến hành, dữ liệu giao dịch đó được mang ra để so sánh với giá trị “cực đoan”, từ đó đưa ra phán đoán giao dịch có gian lận hay không.



Hình 3: Xác định gian lận sử dụng giải thuật di truyền [8]

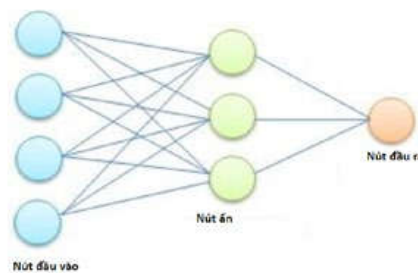
3.4. Mạng Nơ ron

Mạng Nơ ron cũng là một trong những thuật toán thường được áp dụng cho các hệ thống phát hiện gian lận trong thanh toán thẻ tín dụng. Sử dụng mạng Nơ ron trong xác định gian lận thanh toán được ví như áp dụng nguyên lý tư duy của con người để phán đoán sự việc. Bộ não người lưu trữ kinh nghiệm có được từ cuộc sống, khi một hiện tượng xảy ra, dựa vào kinh nghiệm có sẵn, con người sẽ suy luận để đưa ra phán đoán. Cách thức hoạt động của mạng Nơ ron cũng như vậy.

Trong bài toán nhận diện gian lận, thuật toán này chia thông tin thành nhiều nhóm. Nhóm đầu tiên, lưu thông tin thu thập, nghề nghiệp. Nhóm thứ hai, lưu thông tin thanh toán như số lần thanh toán nhiều tiền, mức độ thường xuyên của những lần mua sắm nhiều tiền, nơi mua sắm. Những dữ liệu này được dùng để phán đoán xem giao dịch trong tương lai có đúng do chủ thẻ thực hiện hay

không [9].

Mô hình xác định gian lận sử dụng mạng Nơ ron chia thành ba tầng:



Hình 4: Xác định gian lận sử dụng mạng Nơ ron [9]

- **Tầng đầu.** Tầng này bao gồm các nút đầu vào để kiểm tra thông tin chủ thẻ
- **Tầng ẩn.** Chức các nút thực hiện giải thuật mạng Nơ ron để phán đoán giao dịch hiện tại có gian lận hay không.
- **Tầng đầu ra.** Đưa ra kết quả của việc phân tích, thường đưa ra giá trị nằm trong khoảng từ 0 tới 1.

Bảng 1. Đánh giá các thuật toán phát hiện gian lận

Thuật toán	Ưu điểm	Nhược điểm
Cây quyết định	Phù hợp với để xác định gian lận thẻ tín dụng qua các giao dịch trực tuyến	Dễ thực thi

Thuật toán	Ưu điểm	Nhược điểm
Mô hình Markov ẩn	Phát đoán nhanh	Chi phí cao Độ chính xác thấp Khả năng làm việc với lượng dữ liệu lớn kém
Giải thuật di truyền	Xử lý nhiễu tốt Dễ dàng tích hợp vào các hệ thống Dễ cài đặt	Cần dùng thêm các công cụ tri thức mở rộng để cài đặt Khó hiểu
Mạng Nơ ron	Có khả năng rút ra các luật và dự đoán các hoạt động trong tương lai dựa vào tình huống hiện tại Khả năng phán đoán nhanh Xử lý nhiễu hiệu quả	Đối với mạng Nơ ron lớn, thời gian xử lý lớn Khó cài đặt Dữ liệu phi số cần được chuyển đổi và chuẩn hóa

4. Kết luận

Thẻ tín dụng ngày càng được sử dụng rộng rãi trong các hoạt động mua sắm và thanh toán. Đi cùng với nó, các tội phạm liên quan tới loại hình thanh toán này cũng tăng và đa dạng về hình thức. Mục đích chính của bài báo là đưa ra những thuật toán đã và đang được áp dụng hiệu quả trong các hệ thống phát hiện gian lận, lừa đảo liên quan tới thẻ tín dụng. Hình thức hiệu quả nhất là dựa vào dữ liệu trong quá khứ để phán đoán cho giao dịch hiện tại. Trong các kỹ thuật đã trình bày, cây quyết định và mô hình Markov ẩn được đánh giá là phương pháp hiệu quả nhất.

TÀI LIỆU THAM KHẢO

- [1]. T. N. report (2017). *Card & Mobile payment Industry Statistics*. [Online]. Available: https://nilsonreport.com/publication_chart_and_graphs_archive.php?1=1&year=2017.
- [2]. *The Nilson Report* (2017). [Online]. Available: https://nilsonreport.com/upload/content_promo/The_Nilson_Report_Issue_1118.pdf.
- [3]. Mr.P.Matheswaran, Mrs.E.SivaSankari ME, Mr.R.Rajesh (2015). *Fraud Detection in Credit Card Using Data Mining Techniques*. IJRSET, vol. II, no. I.
- [4]. S. al, (2015). *Credit Card Fraud Detection Using Decision Tree Induction*

Algorithm. International Journal of Computer Science and Mobile Computing, vol. 4, no. 4.

- [5]. Dinesh L. Talekar, K. P. Adhiya (2014). *Credit Card Fraud Detection Using Decision Tree For Tracing Email And Ip*. IJCSI International Journal of Research(IJMER), vol. 4, no. 9.

- [6]. Dinesh L. Talekar, K. P. Adhiya (2014). *Credit Card Fraud Detection System: A Survey*. International Journal Of Modern Engineering Research(IJMER), vol. 4, no. 9.

- [7]. Pooja Chougule#1, A.D. Thakare, Prajakta Kale, MadhuraGole, PriyankaNanekar (2015). *Genetic K-means Algorithm for Credit Card Fraud Detection*. International Journal of Computer Science and Information Technologies (IJCSIT), vol. 6, no. 2.

- [8]. Dr. R.Dhanapal1, Gayathiri.P2 (2012). *Fraud Detection of Credit Card Payment System by Genetic Algorithm*. International Journal of Scientific & Computer Science Issues, vol. 9, no. 5.

- [9]. Raghavendra Patidar, Lokesh Sharma (2011). *Credit Card Fraud Detection Using Neural Network*. International Journal of Soft Computing and Engineering (IJSCE), vol. 1.

BBT nhận bài: 10/8/2018; Phản biện xong: 05/9/2018