시간가중치와 거리기반 도표를 이용한 신용카드 도난 분실 탐지 기법

나용찬[0] , 나연묵
단국대학교 컴퓨터공학과
{wowchany, ymnah}@dankook.ac.kr

A Detection Technique for Credit-card Robbery using Time Weight and Distanced-based Graph

Yongchan Nah[0] Yunmook Nah
Dept. of Computer Engineering, Dankook University

요 약

최근들어 경제활동의 증가로 대부분의 성인들은 몇 장의 신용카드를 소지하고 있을 것이다. 이에 따른 신용카드의 도난 분실 사고는 카드사의 문제가 되고 있다. 기존의 탐지 시스템은 도난신고 등의 일반적인 탐지와 갑작스런 사용 액수의 증가를 탐지하여 도난 분실 카드를 판별하였다. 이것은 소액의 부정거래 탐지가 어렵다는 단점이 있다. 본 논문에서 제시하는 탐지 시스템은 outlier 기법을 사용하여 training set을 만들고 시간가중치와 거리기반 도표를 이용하여 도난 분실 카드를 탐지한다. 금액, 시간 도표에서 거래요구시간의 차를 계산하여 가중치를 주고 장소, 소비종류 도표에서는 training set에서 얻은 자료인 저녁 8시를 기준으로 소비종류의 배열을 바꾼다. 제안된 시스템은 소액의 부정거래 탐지에도 우수하고 이전의 시스템보다 정확함을 장점으로 한다.

1. 서 론

최근 경제활동의 급성장으로 신용카드는 성인이라면 몇 장 정도는 소지를 하고 있을 것이다. 그러나 이에 따른 많은 문제점도 대두되고 있는 것이 사실이다. 본 논문에서는 도난 분실카드에 대한 부정거래를 데이타마이닝의 한 분야인 outlier와 거리기반[5]의 가중치를 이용하여 이상거래의 탐지를 좀 더 효과적으로 할 수 있는 방법을 제시한다.

기존의 outlier기법들은 모두 outlier자체를 탐지하는 것이 아닌 연관규칙[4], 순차패턴[7], 분류, 군집화[2] 등의 방법에서 추가적으로 나온 것이었다. 따라서 위의 방법들에서는 제거의 대상이었고 의미있는 대상이 아니었다. 그러나 현재 금융, 조세, 의학정보 등의 분야에서 outlier는 관심이 되는 분야로 떠올랐고 현재 많은 연구가 진행되고 있다.

outlier를 찾는 방법은 크게 중첩 루프(nested-loop) 방법과 분할 기반(partition-based) 방법[3]으로 나누어 진다. 최근에는 계산의 횟수가 줄어든 분할 기반 방법이 많이 사용되고 있다.

두 방법 모두 계산에 의존하고 있고 데이터의 양이 많을 경우 성능의 감소가 발생한다. 더구나 결과 자체가 어떤 의미를 가지지 못하기 때문에 추가적인 해석 작업이 필요하다.

본 논문에서는 outlier의 기법을 이용하여, 실 데이터를 얻고 그것을 시간 가중치와 간단한 도표를 통해 시각적이고 정확한 방법을 제시한다.

일정기간의 실 데이터인 training set을 이용해 기존 도난 분실카드의 행태를 살피고 들어온 사용자의 정보와 기존의 자료를 더해 도난 분실 카드를 탐지하는 시스템을 구현한다. 특별히 시간 가중치와 금액, 장소, 소비종류 등을 통해 기존의 담당자의 경험에 의존하는 시스템보다 더 효과적인 탐지를 할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 소개하고, 3장에서는 도난 분실 카드에 대한 탐지 방법을 제안한다. 마지막으로 4장에서는 결론과 향후 연구방향에 대해 기술한다.

2. 관련연구

2.1 Outlier 탐지

이전의 outlier[1] 탐지는 데이타마이닝 기법에서 많이 연구되어 오지 않은 기법이다. 기존 데이터마이닝의 몇몇 기법에서는 각 데이터의 유사성이 강조되어 숨겨져 있는 의미를 찾는 것에 대한 연구가 지배적이었다. 그러나 유사하지 않은 데이터의 내용 중에서 희소한 것이 중요한 의미를 가질 수 있다는 것이 알려지면서 많은 연구

가 진행되고 있다. 실제로 카드사기, 탈세, 부정진료, 금융사기 등의 분야에서 많은 연구가 행해지고 있다. 일반적으로 자주 사용되는 방법 중 중첩 루프 방법은 구현이 간단하나 시간 복잡도가 $O(N^2)$로 복잡하고 I/O 장치의 부담이 많은 것이 단점이다. 이에 반해 분할 기반 방법은 정확한 계산이 아닌 대략적인 계산으로 탐지하므로 시간 복잡도와 I/O 장치의 부담이라는 측면에서 중첩 루프 방법 보다 우수하다.

2.2 거리기반 측정

두 위치간의 거리[5]를 측정하는 방법은 유클리디안(euclidean) 거리, 맨하탄(manhattan) 거리 등이 있는데 어떤 방법을 쓰느냐에 따라 특성이 달라진다. 본 논문에서는 유클리디안 거리를 쓰기로 한다. 유클리디안 거리 D 의 정의는 다음과 같다.

$$D = | \chi 2 - \chi 1 |$$

χ : 시간

위의 식에서 $\chi 1$ 은 최종거래시간이며 $\chi 2$ 는 현재 들어온 거래요청시간이다. 이때 D 의 값이 이전에 계산한 값보다 작아지면 가중치를 증가한다.

2.3 클러스터링

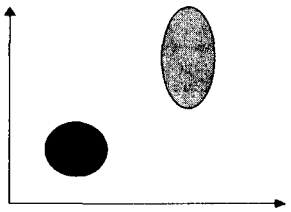

그림 1. 클러스터링

그림 1 에서 보듯이 진한 부분이 점들이 밀집해 있는 부분이고 옅은 부분이 점들이 흩어져 있는 부분이다. 시각적인 도표로 보여지는 여러 도표 위의 점들은 각각의 의미를 갖는다. 그것을 보기 쉬운 하나의 중심점들로 표현하는 것이 클러스터링 기법이다[6].

클러스터링에도 여러 가지 방법이 있으나 outlier 분야에서는 상한(upper bound)과 하한(lower bound)을 가지고 클러스터링하는 방법과 서로 이웃하는 점들과의 거리를 가지고 클러스터링 하는 두 가지 방법이 많이 사용되고 있다.

본 논문에서는 시간, 금액 도표와 장소, 품명 도표에서 얻어진 내용들이 원점에서 가까울수록 도난 분실 카드의 가능성이 높다고 판단하고 있다. 이런 자료들을 비교의 자료로 가지고 있어 새로운 자료가 발생하였을 때 판단의 근거로 삼는다.

3. 도난, 분실카드 탐지방법

3.1 제안 시스템 구조

기존 카드사에서 쓰는 시스템은 고객의 카드에 대한 미리 정해져 있는 기본적인 검사인 카드사의 규약, 신용불량 카드, 한도액수 초과 등의 검사를 하고 그 뒤는 단순히 액수만을 비교하여 탐지하는 방법을 사용했다. 그러나 이것은 담당자의 경험이나 직관에 의존하므로 정확한 탐지가 될 수 없었고 도난, 분실 카드라 해도 큰 액수가 아닌 작은 액수의 불법적인 사용은 탐지가 어려워 대부분의 경우 사후 처리에 머무르는 수준이었다.

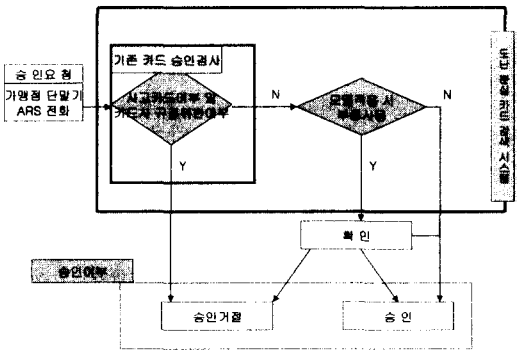

그림 2. 도난분실 탐지 시스템 구조도

그림 2는 본 논문에서 제시하는 시스템 구조로 기존의 시스템보다 정확성을 높일 수 있도록 설계되었다. 큰사각형의 안쪽 사각형이 기존의 검색시스템이고 바깥의 굵은 사각형이 본 논문에서 제시하는 검색시스템이다. 일단 기존의 검색을 수행한 후 제안된 시스템에 적용한다.

3.2 탐지 기법

도난 분실 카드의 탐지 단계는 다음과 같다.

단계 1. 실 데이터를 기반으로 training set 구축
단계 2. 기존의 카드 승인검사
단계 3. 시간, 금액 도표를 이용 도난 분실 여부탐지
단계 4. 장소, 품명 도표를 이용 도난 분실 여부탐지
단계 5. 거래 요청점에 사용자 확인 메시지 발송

3.2.1 training set 구축

정확한 탐지를 위해 실제 카드사의 한달 간 의 도난 분실카드의 사용시간, 금액, 장소, 품목 등을 조사하여 탐지의 기초자료로 사용한다.

본 논문에서는 이미 training set이 구축되어 있다는 가정 하에 시작을 한다.

3.2.2 시간, 금액 도표

그림 3에서 제시된 시간 금액 도표에서는 시작점을 기준으로 training set에서 얻은 한계시간 3시간 범위 안에서 최종 거래 시간과의 거리를 기준으로 거래요청의 입력거리가 짧아지면 도난 분실 카드 가능성의 가중치를 높인다. 최종 거래 시간 후 3시간이 지난 거래 발생 시 탐지 대상에서 제외한다. 또 한계시간 사이의 총 거래횟수를 세어 5회 이상일 때 가중치를 증가시킨다. 금액은 고객의 기존 한도액을 저장하여 기존 한도액 초과 시 기준을 증가시키고 그때에 가중치를 증가한다. 입력되는 점들이 원점에 가까울수록 도난 분실 카드의 가능성이 높다.

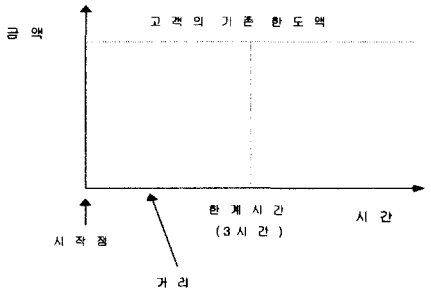

그림 3. 금액 시간 도표

3.2.3 장소, 소비종류 도표

그림 4의 장소 소비종류 도표에서도 training set의 자료에 의거해 낮 시간엔 귀금속, 전자제품, 명품점 등의 순서로 배열하고 밤 시간엔 유흥업소, 접대비의 순서로 배열한다. 이때 밤과 낮으로 배열의 내용이 바뀌는 시간은 training set의 값인 오후8시를 기준으로 하고 입력된 장소와 소비종류가 원점에 밀집되어 질 수록 도난 분실 카드의 가능성이 높다.

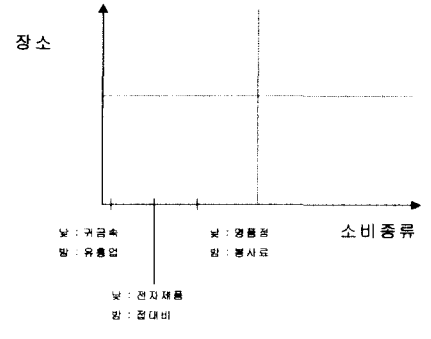

그림 4. 장소 소비종류 도표

위의 그림 3, 4 에서 이상거래로 탐지되는 카드는 사용자 확인 메시지를 요청점에 발송하여 본인 확인 작업을 거친다.

4. 결론

이전의 도난 분실 카드의 탐지는 단순한 요청발생시의 금액만을 보고 탐지하는 수준이었다. 본 논문은 이를 좀 더 발전시켜 시간가중치와 거리기반 도표를 이용하여 기존의 방법보다 정확히 탐지해 내는 방법을 제안하였다. 이전의 방법에서는 액수만을 탐지하기 때문에 소액의 계속적인 도난 분실 카드는 탐지하기 어려웠다. 그러나 본 논문에서 제안한 방법은 소액의 부정거래도 시간가중치와 거래횟수의 누적으로 탐지가 가능하다.

실제적인 효율의 비교를 위해 실제 데이터가 입수되는 대로 각각의 탐지효율을 비교할 예정이다. 실제 탐지되는 정보들의 정확도를 이전의 시스템과 비교하여 얼마나 좋아지는 지에 대한 비교가 필요하다.

향후 연구과제는 보다 정확한 결과를 유지시키는 것에 대한 연구인데 예를 들어 장소 품명 도표의 경우 품명의 순서에 따라 정확도의 차이를 보이게 될 것이다. 이를 위해서 보다 정확한 training set의 구현이 필요하고 실제의 데이터를 주기적으로 갱신시키는 작업이 필요하다.

참고문헌

[1] 박성진, "고객성향을 통한 고객관리", 마이크로 소프트웨어, 2001년 5월, pp.218-222.

[2] 조순이, 이도헌 "스키마간 연관성을 이용한 테이블 군집화 기법", 정보과학회 2001 춘계학술대회논문집 , 제28권1호, 한국정보과학회, 2001년 4월, pp.85-87

[3] Kyuseok Shim, Sridhar Ramaswamy, Rajeev RastogiEfficient, "Algorithms for mining outliers from large datasets," in Proc. of SIGMOD, Dallas, Texas, 2000.

[4] R. agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Database," in Proc. of SIGMOD, 1993, pp.207-216.

[5] Edwin Knorr and Raymound Ng, "Algorithm for Mining Distance-based Outliers in Larg Datasets", in Proc. VLDB, Sept. 1998 pp.392-403.

[6] Markus M.Breuning, Hans-Peter Kriegel, Raymond T.Ng, Lörg Sander, "LOF:Identifying Density-Based Local Outliers," in Proc. SIGMOD Dallas, TX, 2000.

[7] Minos N.Garofalakis, Rajeev Rastogi and Kyuseok Shim, "SPIRIT: Sequencial Pattern Mining with Regular Expression Consyraints," in Proc. VLDB, Edinbrugh, Scotland, UK, 1999.