퍼지 및 지능시스템학회 논문지 2005, Vol. 15, No. 5, pp. 545-551

신용카드 사기 검출을 위한 비용 기반 학습에 관한 연구

Cost-sensitive Learning for Credit Card Fraud Detection

박래정

Lae-Jeong Park

강릉대학교 정보전자공학부 전자공학전공

Department of Electronics Engineering

Kangnung National University

요 약

사기 검출의 주목적은 사기 거래로 인해 발생하는 손실을 최소화하는 것이다. 하지만, 사기 검출 문제의 특이한 속성, 즉 불균형하고 중첩이 심한 클래스 분포와 비균일한 오분류 비용으로 인해, 실제로 희망하는 거절율 동작 영역에서의 분류 비용 측면의 최적 분류기를 생성하는 것이 용이하지 않다. 본 논문에서는, 특정 동작 영역에서의 분류기의 분류 비용을 정의하고, 진화 탐색을 이용하여 이를 직접적으로 최적화함으로써, 실제 신용카드 사기 검출에 적합한 분류기를 학습할 수 있는 비용 기반 학습 방법을 제시한다. 신용카드 거래 데이터를 사용한 실험을 통해, 제시한 방법이 타 학습 방법에 비해 비용에 민감한 분류기를 학습할 수 있는 효과적인 방법임을 보인다.

Abstracts

The main objective of fraud detection is to minimize costs or losses that are incurred due to fraudulent transactions. Because of the problem's nature such as highly skewed, overlapping class distribution and non-uniform misclassification costs, it is, however, practically difficult to generate a classifier that is near-optimal in terms of classification costs at a desired operating range of rejection rates. This paper defines a performance measure that reflects classifier's costs at a specific operating range and offers a cost-sensitive learning approach that enables us to train classifiers suitable for real-world credit card fraud detection by directly optimizing the performance measure with evolutionary programming. The experimental results demonstrate that the proposed approach provides an effective way of training cost-sensitive classifiers for successful fraud detection, compared to other training methods.

Key Words : fraud detection, cost-sensitive learning, classifier evaluation

1. 서 론

최근 신용카드의 사용 증가와 함께, 사기(fraudulent) 거래로 인한 신용카드 회사와 은행들의 경제적 손실이 증가하고 있으며, 이러한 손실을 줄이기 위한 조기 검출 시스템을 운영하고 있다. 효과적인 사기 검출기법에 관한 다양한 연구가 데이터 마이닝과 기계 학습 분야에서 진행되어 왔다[1,2].

대부분의 사기 검출 문제는 클래스간의 분포가 매우 유사(overlapping)하고, 클래스 비율이 몹시 불균형(unbalanced)하다. 즉, 사기 거래와 정상 거래의 구분이 용이하지 않고, 사기 거래 수가 정상(legitimate) 거래 수에 비해 월등히 적다. 사기 검출을 위한 분류기 학습 과정에서, 단순한 오분류 샘플 개수의 최소화 성능 척도, 예를 들면, 정확도(accuracy)를 사용하는 경우, 분류기가 희망하는 사기 검출 성능을 갖도록 기대하거나 조정하기가 어렵다[3]. 따라서, 최근 검출율(true positive rate)와 오검출율(false positive rate)간의 적절한 타협(trade-off)을 평가하고 조절할 수 있는 ROC(Receiver Operating Characteristics) 곡선 등을 이용하여 분류기의 성능을 평가하는 척도가 많이 사용하고 있다[3,4].

한편, 사기 검출 문제의 또 다른 특징은 각 거래의 오분류 비용(misclassification cost)이 균일(uniform)하지 않다는 점이다. 특히 신용카드 사기 검출 문제는 정상(사기) 거래를 사기(정상) 거래로 분류할 때의 비용이 일정한 고정값이 아니라, 오분류되는 거래에 따라 달라진다. 그러므로, 신용 카드 사기 검출용 분류기의 학습 방법은 비균일한(non-uniform) 오분류 비용을 고려함으로써, 비용 측면에서 최적의 분류기를 학습할 수 있어야 한다.

위의 측면에서 기존의 연구를 살펴보면, [5],[6]은 일반적인 분류기 학습시의 장애 요인인, 불균형하고 겹쳐있는 정상과 사기 클래스 분포를 극복하기 위한 학습 방법을 제시하였으며, [7],[8],[9]는 검출 성능 향상(높은 검출율과 낮은 오검출율)을 위하여 하이브리드 학습 방법과 복수개의 분류기를 함께 사용하는 하이브리드 분류 방법을 사용하였다. 실제 필드에서의 신용 카드 사기 검출 시스템의 필요성과 요구는,

접수일자 : 2005년 5월 6일
완료일자 : 2005년 8월 23일
감사의 글 : 이 논문은 2001년도 강릉대학교 학술연구조성비 지원에 의하여 수행되었습니다.

사기 거래로 인한 손실을 줄이는 것이므로, 사기 검출 분류기의 학습 방법은 비용 최소화에 맞추어져야 한다. 하지만, [10]을 제외하고는 비용을 최소화하기 위한 사기 검출용 분류기 학습 방법에 관한 연구는 의외로 적은 편이다. 본 논문에서는 신용 카드 사기 검출 분류기를 비용측면에서 최적화하기 위하여, 사기 검출기의 동작 구간에서의 평균 분류 비용에 관한 목적함수를 정의하고, 이를 이용한 특정 동작 구간에서의 사기 검출기의 비용 최소화 학습 방법을 제시한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존의 신용 카드 사기 검출 연구들을 살펴보고, 3장에서 사기 검출기의 특정 동작 구간에서의 평균 분류 비용 평가의 필요성과 평가 지수에 대하여 설명한다. 4장에서는 본 논문에서 제시하는 사기 검출 분류기의 평균 분류 비용 최소화 학습 방법을 살펴보고, 5장에서는 실제 신용카드 거래 데이터를 사용하여 제시한 학습 방식에 의한 사기 검출 분류기의 성능을 실험 분석하고, 효과 및 장단점에 대하여 논의하며, 6장에서 결론을 맺는다.

2. 신용 카드 사기 검출 관련 기존 연구들

신용카드 사기 검출 관련한 기존의 연구를 살펴보면, Dorronsoro 등은 신용카드 사기 검출에서 불균형적이고 겹쳐있는 클래스 분포에 의한 신경망 학습 문제를 해결하기 위하여 비선형 discriminant analysis 방법에 기반한 학습방법을 제안하였다[5]. Park 등은 불균형적인 클래스 분포로 인한 오검출을 줄이기 위하여 신경망과 결정 트리(decision tree)의 하이브리드 기법을 제안하였다[6]. Brause 등은 입력 특징 중 심볼 특징(symbol feature)과 연속 특징들을 효과적으로 다루기 위하여 연관 규칙과 신경망 분류기를 함께 사용하였다[7]. Maes 등은 신경망과 Bayesian 네트워크 기법의 성능 및 장단점을 ROC 그래프를 이용하여 비교 분석하였으며[8], Park은 사기 검출을 위한 신경망 분류기 학습에서의 MSE 최소화의 문제점을 지적하고, 진화기법을 이용하여 일정한 오검출율 조건 하에서의 검출율 최대화 방법을 제시하였다[9].

대부분의 신용카드 사기 검출 연구가 검출율/오검출율 측면의 성능 향상에 초점을 맞추고 있으며, 분류 비용 측면의 성능 향상에 관한 연구는 많지 않다. 비용 기반(cost-sensitive) 사기 검출 분류기 학습에 관한 연구로는, Chan 등이 불균형한 학습 데이터를 다소 균형된 클래스 분포를 가진 복수개의 부분 학습 데이터로 나누고, 이를 사용하여 복수개의 선형 분류기를 생성하고, 생성된 선형 분류기의 메타학습을 통해 최종 분류기를 생성하는 방법을 제시하였다[10]. 또한, 분류기 학습과정에서 오분류 비용이 큰 오분류 학습 데이터의 비중을 조정하여 분류기들을 비용에 민감하게(cost-sensitive) 만드는 방법인 AdaCost 기법[11]을 채택하여 분류 비용을 최소화하였다. Phua 등도 Bagging과 Stacking 기법을 사용하여 신경망, Naive Bayesian, 결정트리 분류기의 성능을 비용 모델과 오분류 비용 측면에서 분석 고찰하였다[12]. 이들 방법들은 복수개의 분류기와 모든 데이터의 가중치를 저장하고 조절(update)하여야 하는 복잡한 학습 구조를 가지며, 최종 분류기의 특정 동작 영역, 예를 들면, 오검출율 범위 1%-2% 에서의 분류 비용 성능을 선택적으로 최적화하기 어려운 단점을 갖고 있다.

한편, 실제 신용카드 사기 검출기는 특정 동작점 혹은 구간에서 운영되므로, 분류기의 공정하고 의미있는 성능 비교 및 평가를 위해서는 출력 문턱값을 조정하면서 분류기의 성능을 분석 평가하여야 하는데, 대부분 이러한 성능 분석의 결과가 나타나 있지 않다. 또한, 위 연구의 대부분이 상대적으로 적은 거래 데이터 수를 사용하거나, 클래스 불균형이 상당히 완화되어 있는 데이터 (사기 비율이 5% 이상)를 사용하였으므로, 제시한 방법들이 실제 신용카드 사기 검출 환경에서의 성능과 차이를 보일 가능성이 높다.

3. 비용 곡선 기반 분류기 성능 평가

3.1 특정 동작 구간에서의 성능 평가

정상과 사기 클래스 비율의 변동, 사기 거래 패턴의 변동, 정상/사기 클래스 내의 데이터 분포 변동 등에 의해서 입력 데이터 특성은 시간에 따라 약간씩 변동하게 된다. 이러한 데이터 변동에 따라서 사기 검출기의 동작점 (operating point, 즉, 검출율 및 오검출율)이 변동되고, 이에 따라 분류 비용 성능도 함께 달라진다. 학습 과정에서 희망하였던 동작점과 분류 비용 성능이 운영 중에 달라질 수밖에 없다. 그러므로, 사기 검출 분류기의 학습 과정에서 이러한 동작점의 변동을 감안하여, 특정 동작점인 검출율/오검출률에서의 성능이 아니라, 입력 거래 특성 변동에 의해 예상되는 동작 영역(operating range)에서의 "평균적"인 성능을 기준으로 사기 검출 분류기를 학습하는 것이 바람직하다. 일반적인 학습 방법과 다르게, 특정 동작 영역에서의 평균 성능을 정의하고, 이를 기준으로 분류기를 학습하는 방법이 필요하다.

3.2 비용 곡선(Cost curve) 기반한 분류기 성능 평가

입력 변동과 이에 따른 동작점 변동에 의한 분류기의 분류 비용 성능을 효과적으로 분석 비교하기 위해서 Provost와 Fawcett가 제안한 ROC 공간상의 등성능 라인(iso-performance line)[3]이나, Drummond와 Holte가 제안한 비용 곡선(cost curve)[13]의 이용이 가능하다. 비용 곡선은 분류기의 평균 분류 비용(expected cost)를 명확하게 표현하기 위해 제안되었는데, ROC 표현의 이중(daul) 표현이긴 하나, 평균 비용기준으로 분류기의 성능을 표현하는 경우에 ROC 곡선보다 상대적 잇점을 가지고 있다[13].

비용 곡선을 살펴보면, 고정 판단 문턱값 θ을 갖는 분류기의 동작점을 검출율과 오검출율, {TP, FP}로 표현하면, 그 분류기의 정규화된 평균 오분류 비용 NE[C]는 아래와 같이 표현된다.

$$NE[C] = \frac{p(+)(1-TP)C(-|+) + p(-)FPC(+|-)}{p(+)C(-|+) + p(-)C(+|-)} = (1-TP+FP)PCF(+) + FP \quad (1)$$

여기서 $p(+)$와 $p(-)$는 정상과 사기의 priori 확률, 즉 클래스 비율을 나타내며, $C(+|-)$와 $C(-|+)$는 미검출 분류 비용과 오검출 분류 비용을 나타내며, $PCF(+) = \frac{p(+)C(-|+)}{p(+)C(-|+) + p(-)C(+|-)}$는 확률-비용함수로 불리우며, 클래스 비율과 오분류 비용을 함께 포함한 '동작 환경'을 나타낸다. 식 (1)에서 보듯이, {TP, FP}의 동작점을 갖는 분류기는 확률-비용 함수 PCF(+)와 NE[C]를 x, y축으로 하여 하나의 라인으로 표시 가능하며, 그림 1에서 보듯이, 판단 문턱값 θ을 바꾸면서 얻어지는 여러 개의 비용 라인들로 구성되는 최저 포락선(envelope)이 바로 해당 분류기의 비용 곡선이 된다. 비용 곡선을 이용하면, NE[C]

측면에서 "동작 환경"에 따른 분류기 최적 성능을 쉽게 표현할 수 있으므로, 복수개의 분류기의 성능 비교에 효과적으로 이용될 수 있다. 즉, 특정 클래스 비율과 오분류 비용 비율에서 분류기간의 상대적 성능 비교가 명확하게 표현 가능하며, 또한 관심있는 특정 PCF(+) 범위에서의 분류기의 평균 성능 비교 평가가 가능하다. 예를 들면, 그림 1에서의 빗금친 부분의 면적이 최소가 되도록 분류기를 학습한다면, 해당 PCF(+) 범위 내에서의 평균 NE[C]가 최적인 분류기를 학습할 수 있다.

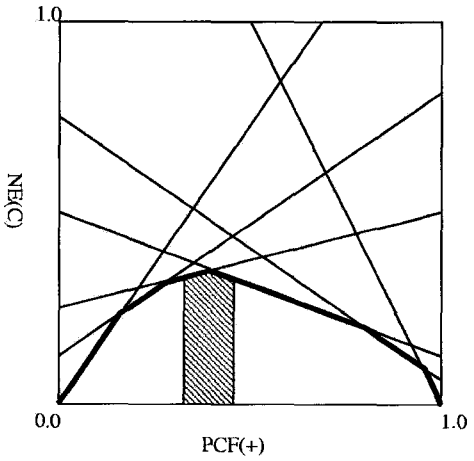


그림 1. 여러 판단 문턱값 θ에 의한 비용 직선과 비용 곡선(굵은 선으로 표시)

Fig. 1 A cost curve (thick lines) formed by six cost lines, each corresponding to a decision threshold θ.

하지만 비용 곡선을 신용카드 사기 검출에 적용하기에는 몇 가지 제약이 있다. 첫째, 비용곡선은 고정된 클래스 오분류 비용을 가정하고 있지만, 실제는 승인 요청 금액이 거래마다 다르다. 즉, 오분류 비용 $C(+|-)$와 $C(-|+)$가 고정된 값이 아니고 거래마다 다른 값을 가진다. 둘째로, PCF(+)는 클래스 비율변동과 오분류 비용의 변화를 표현할 수 있지만, 실제로 클래스 비율의 변동은 미비하며, 분류기의 성능 변동을 가져오는 입력 변동의 주요인은 사기와 정상 데이터의 클래스 분포(within-class distribution)의 변동이므로, 동작점의 변동을 PCF(+) 변화로 표현하여 사용하는 것은 적절하지 않다.

4. 사기 검출 분류기의 평균 분류 비용 최소화 학습

4.1 평균 분류 비용 평가 지수 (PWSC)

매개 변수화하기 어려운 입력 분포 변동에 대한 사기 검출기의 평균 성능을 평가하는 것보다 입력 분포 변동을 반영하면서 측정이 가능한 변수를 기준으로 사기 검출기의 동작점 범위를 정의하고, 해당 동작 범위에서 분류기의 평균 오분류 비용을 평가하는 것이 사용자 입장에서 보다 현실적이고 실질적인 접근이다. 본 논문에서는 이러한 조건을 만족시키는 변수로 거절율(alarm rate)[1]을 사용하는데, 이는 입력 거래 특성 변동에 따른 사기 검출기의 동작점 변동을 반영하며, 분류 결과로부터 쉽게 계산할 수 있다. 또한, 필드에서는 의심스러운 거래에 대한 조사 한계를 결정하고, 분류기의 판단 문턱값 조정 여부를 판단하는 요소이기도 하다[9]. 따라서, 분류기 M의 허용 동작 구간은 거절율로 표현하고, 학습 과정에서 사용하는, 해당 거절율 구간에서의 평균 분류 비용을 최소화하기 위한 함수 PWSC(partial weighted sum of cost)는 아래와 같이 주어진다.

$$PWSC(M) = \int_{r_0}^{r_n} \overline{C(r)} \cdot w(r)dr \qquad (2)$$

여기서 $r(0<r<1)$은 분류기 M의 거절율로서, 분류기 출력의 판단 문턱값 θ에 의해 결정되며, $\Delta r(=r_n - r_0)$는 입력 변동을 감안한 허용 거절율 구간이며, $w(r)$는 Δr 구간에서의 분류기의 동작점(거절율)에 대한 가중치(weighting factor)를 반영하기 위한 것으로서, 일정 기간동안의 운영 결과를 이용하여 결정할 수 있으나, 여기서는 균일한 상황을 가정한다. $\overline{C(r)}$는 거절율 r에서의 분류기 M의 검출로 인한 이득과 오검출로 인한 손실을 반영하기 위하여 식 (3)과 같이 주어진다.

$$\overline{C(r)} = \alpha(N_F(r) - \gamma N_T(r)) + (A_F(r) - \gamma A_T(r)) \qquad (3)$$

여기서, $N_F()$와 $A_F()$는 오검출한 정상 거래 수와 금액 합을 나타내며, $N_T()$와 $A_T()$는 검출한 정상거래 거래 수와 금액 합을 나타낸다. 식 (3)의 첫 항은 거래 건수 측면에서 오검출을 감소하고 검출을 증가시키기 위한 항이며, 두 번째 항은 오검출 금액을 감소하고 검출 금액을 증가시키기 위한 항으로서, 거래 건수와 금액 측면을 함께 고려하여 분류기의 분류 비용을 최적화한다. α는 두 항의 비중을 결정하는 상수이며, γ는 검출과 오검출의 비중을 결정하는 상수로서, 예비 실험 과정에서 trial-and-error 방법을 사용하여 결정한다.

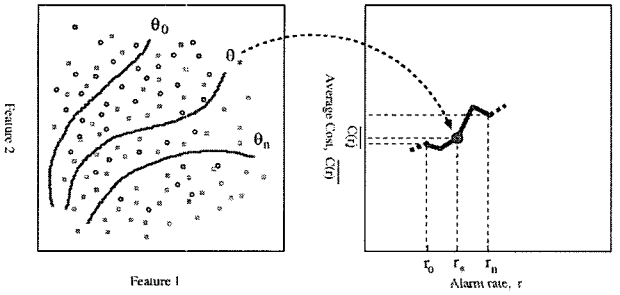
그림 2는 허용 동작 범위 $[r_0, r_n]$[2]가 주어질 때, 분류기 M의 PWSC를 계산하기 위한 과정을 도식적으로 나타낸다. 그림 2(a)에서처럼 분류기 M의 판단 문턱값 θ를 $[r_0, r_n]$에 대응되는 구간 $[\theta_0, \theta_n]$에서 변경하면서[3] 그림 2(b)에의 $\overline{C(r_i)}, i=0,\ldots,n$을 계산하고, 이를 이용하여 분류기 M의 PWSC를 계산한다. 데이터를 이용하여 계산하므로, $[r_0, r_n]$ 구간에 유한한 개수의 $(r_i, \overline{C(r_i)})$가 주어지므로, 식 (4)의 근사식을 이용하여 계산한다.

1) 총 거래 중 사기 거래로 판단하여 거절하는 비율(%)

2) 일반적으로 사기 검출기의 현재 동작점 r_*과 허용 동작점 범위 $[r_0, r_n]$은 총 거래수 대비 거절 거래수가 일정 범위 이내가 되도록, 예를 들면, 일 단위 혹은 주간단위의 거절 거래의 조사 처리 가능 건수를 넘지 않도록 사기 검출 분류기의 사용자에 의해 결정된다.

3) 분류기의 판단 문턱값(decision threshold) θ를 변화시킴에 따라 사기 검출기의 동작점이 달라진다. 분류기의 출력값이 θ보다 작으면 정상거래, θ보다 크면 사기거래로 분류한다. 일반적으로 θ를 높이면 검출율과 오검출율이 줄어들고, 낮추게 되면 검출율과 오검출율이 함께 증가하게 된다.

$$PWSC(M) = \int_{r_0}^{r_n} \overline{C(r)} \cdot w(r)dr \simeq \sum_{i=0}^{n-1} w(r_i) \cdot \frac{\overline{C(r_i)} + \overline{C(r_{i+1})}}{2} \cdot (r_{i+1} - r_i) \quad (4)$$



(a) (b)

그림 2. (a) 특징 공간상에서의 사기 검출 분류기 M의 θ에 따른 분류 경계선, (b) θ에 대응되는 거절율 구간 $[r_0, r_n]$과 $\overline{C(r)}$ 곡선.

Fig. 2 (a) Decision boundaries on the feature space associated with decision thresholds θ of classifier M, (b) $\overline{C(r)}$ curve on a range of alarm rates $[r_0, r_n]$ corresponding to $[\theta_0, \theta_n]$.

4.2 진화 연산을 이용한 분류기 학습

4.1절에서 설명한 PWSC 지수가 최소가 되도록 분류기를 학습함으로써, 특정 동작 구간 $[r_0, r_n]$에서의 최적의 평균 분류 비용 성능을 갖는 분류기를 생성할 수 있다. 하지만, 결정 트리(decision tree)나 신경망 분류기[4]를 사용하는 경우, 분류기의 학습 파라미터 (결정트리의 각 노드의 분류 경계, 신경망 분류기의 가중치(weights))에 대한 PWSC의 변화율 계산이 불가능하므로, 오류역전파 학습 방법(backpropagation)과 같은 기울기 감소(gradient-descent) 학습 방법을 사용할 수 없다. 따라서, PWSC를 최소화하기 위해서 기울기 정보를 사용하지 않으며, 근사 최적해 탐색 능력을 갖고 있는 탐색 기법인 진화 프로그래밍(evolutionary programming, EP [14])을 사용한다. EP는 국소점 (local optima) 문제를 극복할 수 있는 최적화 기법으로 알려져 있으며 신경망 학습에서도 주어진 비용 함수를 최적화하는 가중치를 찾는데 성공적으로 적용되어 왔다[15].

본 논문에서 사용하는 EP의 기본 알고리즘은 그림 3과 같다. 간단히 소개하면, EP는 μ개의 개체로 구성된 집단(population) P를 사용하는 탐색 알고리즘으로서, 각 개체 $\vec{a}_i$는 n차원의 $\vec{x}_i$ 벡터와 $\vec{v}_i$로 구성되어 있다. $\vec{x}_i$ 벡터는 최적화하고자 하는 파라미터로서, 신경망을 학습하고자 하는 경우, 신경망의 가중치 벡터에 대응되고, 따라서 개체 $\vec{a}_i$는 하나의 신경망 분류기를 나타낸다. 동작과정을 살펴보면, μ개의 개체를 초기화하고 비용함수 $f()$ 즉 PWSC 지수를 사용하여 적합도(fitness)를 평가한 후, 각 개체를 변이 연산자(mutation), $m()$을 사용하여 새로운 μ개의 개체 $\vec{a}_i'$로 구성된 집단 P'를 구성한다. 새로 구성된 집단내의 각 개체를 평가한 다음에, P와 P'의 합집합 (2μ의 개체) 에서 확률적인 경쟁 과정[5]을 통하여 다음 세대의 집단을 구성함으로써, 점점 적합도가 높은 개체들을 탐색하게 된다. 변이 연산자 $m()$는 현재 개체로부터 새로운 개체를 생성하기 위한 가우시안 형태의 변이 연산자로서, 일반적으로 식 (5)와 같이 주어진다.

$$\begin{aligned} v_i' &= v_i + v_i \cdot N_i(0,1), \\ x_i' &= x_i + v_i' \cdot N_i(0,1), \end{aligned} \quad i \in \{1, \ldots, n\} \quad (5)$$

여기서, $N_i(0,1)$는 평균이 0이고 분산이 1인 가우시안 분포를 갖는 난수 생성기이다. 살펴본 바와 같이, EP 알고리즘은 한 세대의 변이와 선택과정을 자연계의 진화과정처럼 오랜 기간 동안 반복적으로 수행함으로써, 원하는 수준의 적합도를 갖는 개체를 찾고자 하는 탐색 기법이다. EP 알고리즘은 수렴 속도가 다소 늦은 단점이 있으나, PWSC 지수처럼 기울기 정보를 이용할 수 없는 문제에 효과적으로 적용 가능한 장점을 갖고 있다.

```
t = 0
초기화 P(0) := {a_1(0), ..., a_μ(0)}, where
    a_k = (x_i, v_i, i ∈ {i, ..., n})
평가 P(0) = {Φ(a_1(0)), ..., Φ(a_μ(0))}, where
    Φ(a_k(0)) = f(x_k(0))
종료 조건이 만족될 때까지 아래 동작을 반복
    P'(t) := 변이 P(t), where
        a_k'(t) := m(a_k(t)), k ∈ {1, ..., μ}
    평가 P'(t) := {a_1'(t), ..., a_μ'(t)}
        {Φ(a_1'(t)), ..., Φ(a_μ'(t))}, where
            Φ(a_k'(t)) = f(x_k'(t))
    P(t+1) := 선택{P(t) ∪ P'(t)}
    t = t+1
종료
```

그림 3. EP의 기본 알고리즘

Fig. 3 Pseudo code of EP algorithm.

5. 실험 결과 및 분석

5.1 데이터 이해 및 처리

실험에 사용한 신용 카드 거래 데이터는 다음과 같다. 1년간의 거래 중에서 선별적으로 수집된 약 5만 건의 거래를 학습 데이터로 사용하였다. 이후 4개월 동안의 약 900만 건의 거래 중에서 1개월간의 데이터는 $\overline{C(r)}$의 파라미터 α와 γ를 결정하기 위하여 사용하고, 나머지 3개월간의 데이터를 사기 검출 분류기의 일반화 성능 평가에 사용하였다. 학습 데이터의 사기 대 정상 비율은 1:4정도이고, 테스트 데이터의 경우는 약 1:1700정도로서, 신용카드 사기 검출 문제가 상당히 불균형한 문제임을 알 수 있다. 거래 횟수와 빈도, 요청 금액과 요청 시간 등을 이용한 6개의 특징을 추출하여 분류기의 입력으로 사용하였다.

4) 거의 대부분의 상용 사기 검출 시스템은 비선형 분류 특성과 연속 출력(continuous output)이 가능한 신경망 분류기를 사용한다.

5) 적합도가 높은 개체가 다음 세대의 집단에서 살아남을 확률이 높게끔 경쟁 과정을 구성한다. 즉, 확률적 적자 생존의 (survival of the fittest) 선택 과정을 구현한다.

5.2 신경망 기반 사기 검출기의 학습

사기 검출 방식은 개인의 정상적인 사용 행태 특징을 추출하고, 거래 발생시 이와의 차이를 근거로 사기를 검출하는 형태와, 개인별 구분없이 사기 거래의 공통 패턴을 추출하고 이와 유사한 거래를 사기로 판별하는 형태로 나눌 수 있는데, 실험에 사용한 데이터에는 사용자마다의 일정 기간이상의 연속 거래 데이터가 제공되지 않은 관계로, 후자 형태를 채택하였다. 사기 검출 분류기의 신경망 모델은 1개의 은닉층을 갖고 시그모이드(sigmoid) 함수를 사용하는 다층 퍼셉트론(multi-layer perceptron, MLP)을 사용한다. 보다 분류 및 근사화 능력이 우수한 신경망 모델을 사용하여 검출 성능을 다소 향상할 수 있으나, 이에 대한 실험은 생략하였다. 수행한 모든 실험의 EP의 실험 파라미터인 개체 수와 총 반복 횟수는 적절한 수렴 속도를 위하여 각각 30과 1000으로 설정하였다. 가우시안 변이의 최적화 과정에서 해당 거절율 구간에 포함되는 $\overline{C(r_i)}$의 개수가 일정 개수 이하인 경우의 분류기에 벌점(penalty)를 부가함으로써, 실제 운영중 거절율과 $\overline{C(r)}$의 변동이 심할 가능성이 높은 분류기는 최종 탐색 대상에서 제외되도록 하였다.

5.3 비용 모델

학습된 분류기의 테스트 데이터에 대한 최종 성능 평가를 위한 분류 비용 계산 모델은 표 1과 같다. [10]에서처럼, 사기로 판단하는 경우에는 조사비용 포함 비용 overhead를 고려하고, 오검출시의 경우에는 발생하게 되는 손실 및 그로 인해 향후 예상되는 비용을 반영하는 false_alarm_loss를 포함하였다. 실험에서는 오검출시 거래의 수수료 감소만을 포함하였다. 표 1에서 볼 수 있듯이, 사기로 판단하는 경우에도, 해당 거래의 amount(i)가 overhead보다 작을 경우에는, 즉, 사기로 판단함으로써 발생하는 비용보다 작은 경우는 비용 감소를 위해서 모두 승인하는 후처리 규칙을 적용한다.

표 1. 신용카드 거래 i의 비용 계산 모델
Table. 1 Cost calculation model for credit card transaction i.

정상/사기	분류 결과	비용	
정상	정상(승인)	0	
	사기(거절)	0	if amount(i) < overhead
		overhead + false_alarm_loss(i)	if amount(i) > overhead
사기	정상(승인)	amount(i)	
	사기(거절)	amount(i)	if amount(i) < overhead
		overhead	if amount(i) > overhead

5.4 실험 결과 및 논의

제시하는 학습 방법을 평가하기 위하여 다른 두 기준과 비교 실험을 수행하였으며, 실험 조건은 다음과 같다. 신경망의 은닉층 노드 개수를 6과 12로 증가시키면서, 각각 30개의 신경망 분류기를 생성하였다. 이하, 모든 결과 값은 30개 신경망 분류기에 의한 평균값이다. PWSC의 파라미터 α와 γ는 예비 실험과정을 통하여 각각 2×10^5과 0.3으로 결정하였다. 실제 상황과 동일한 환경을 만들기 위하여 테스트 데이터에 대한 일반화 성능 계산시의 동작 거절율 r_*과 거절율 구간 $[r_0, r_n]$은 사기 검출 시스템의 운영자의 정보로 부터 0.17%와 [0.1%, 0.25%]로 설정하였다. 각 분류기의 판단 문턱값 θ_*은 1개월간의 데이터를 이용한 예비 실험에서 거절율 r_*을 갖게끔 고정한 상태에서, 3개월간의 테스트 데이터 (A달, B달, C달로 구분) 각각에 대하여 분류기의 거절율과 분류 비용을 측정하였다.

제시한 방법의 성능을 검증하기 위하여, 신경망 분류기 학습에 널리 사용되는 MSE 기준으로 분류기를 학습한 경우와 동일한 거절율 구간 $[r_0, r_n]$에서 사기 거래의 "검출율"이 최대가 되도록 학습한 분류기의 경우도 함께, 동일한 조건으로 실험하여 결과를 비교하였으며, 그 결과는 표 2와 같다. 표 2는 무검출 경우 (모든 거래를 승인하는 경우)에 계산한 거래당 분류 비용 대비, MSE 최소화 기준, 검출율 최대화 기준, PWSC 최소화 기준 각각으로 학습한 분류기의 거래당 분류 비용의 차이를 나타낸다. 표1의 비용 계산 모델을 사용하고, overhead와 수수료는 1000원과 2.5%로 설정하여 분류 비용을 계산하였다. 결과에서 (-)값은 무검출 경우보다 비용이 적음을 나타내는 것으로, 상대적 이익을 의미한다.

표에서 알 수 있듯이, PWSC 기준으로 학습한 분류기의 평균 성능이 가장 우수한 반면, MSE 기준으로 학습한 경우가 가장 낮은 성능을 나타낸다. 이러한 결과는, 모든 입력 데이터의 MSE를 최소화하는 학습 방법이 특정 동작 구간에서의 성능의 최적화에 적절하지 않음을 나타내며, 이는 탐색 공간에서 MSE 최소화 기준과 PWSC 최소화 기준이 일대일 대응하지 않기 때문이다. 제시한 방법의 성능과 MSE 기준 경우와의 차이는, 3개월 모두에서 통계적으로 유의하다(one-sided tail p값 < 0.0005). 검출율 최대화의 경우도 MSE 기준보다는 높은 성능을 보이나, 여전히 PWSC 기준보다 낮은 성능을 나타내며, 성능 차이도 2달 (A달, C달)에 대해서 통계적으로 유의한 결과를 보인다(one-sided tail p값 < 0.0005, B달의 경우 p값은 0.0136). 매달 거래 건수를 감안할 때, 작은 평균 분류 비용의 차이도 막대한 비용 차이를 나타내므로, 제시한 비용 기준 학습 방식의 분류 비용 감소는 중요한 의미를 가진다.

12개의 은닉층 노드 개수의 실험 결과를 살펴보면, 각 학습 기준간의 성능 차이는 은닉층 노드 개수가 6개인 경우와 동일한 경향을 보인다. 특이한 점은, PWSC 학습 기준을 제외하고는 6개인 경우와 비교하여 3개월 모두에 대해서 다소 낮은 성능과 큰 표준편차를 보이는 것이다. 특히, MSE 기준으로 학습한 경우가 특정 동작 구간에서의 성능을 최적화하는 학습하는 방식에 비해, 신경망 분류기가 overfitting할 가능성이 높기 때문인 것으로 추측된다. 또한, 오검출로 인한 손실을 결정짓는 수수료율 변경 (2.0%, 2.5%, 3.0%)에 따른 실험에서도 세 학습 기준의 성능 차이는 동일한 경향을 보였다. 그림 4에 A, B, C 연속 3개월 거래 데이터에 대한 세 학습 기준으로 학습한 분류기의 동작점 (거절율과 평균 분류 비용) 변화를 도시하였다. 세 분류기 모두가 동작점 (r_* = 0.17%)에서 크게 벗어나지 않으며, 허용 구간 $[r_0, r_n]$내의 거절율을 보인다. 제시한 학습 방법에 의한 분류기가 3개월 모두에 대하여 매달 거절율은 낮지만, 분류 비용 성능이 높음을 알 수 있다. 그림 5는 검출율과 오검출율 측면에서 세 학습 기준에 의한 분류기의 평균 성능을 나타낸다. x축이 오검출율이고 y축이 검출율로서, 분류기의 성능을 나타내는 점이 그래프의 좌측 상단에 위치할수록 좋은 분류 성능을 의미

한다. 그림에서 볼 수 있듯이, PWSC 학습 기준을 사용한 분류기가 비슷한 검출율인 경우에도 낮은 오검출율을 나타내며, 이러한 낮은 오검출율 성능이 타 방법보다 상대적으로 큰 평균 분류 비용의 감소를 가져온 것으로 추정된다.

제시한 분류 비용 감소를 위한 분류기 학습 방법의 단점은, 첫째 $\overline{C(\gamma)}$의 α와 γ를 결정하기 위한 예비 실험 과정이 필요하다는 것이다. 여러 α와 γ 값으로 실험한 결과를 분석해 보면, γ 값에 분류기 성능이 다소 민감한 변화를 보이므로 trial-and-error 방식이외에 체계적인 결정 방법이 요구된다. 또한, 진화 탐색 기법을 사용하기 때문에 분류기 생성을 위한 학습 시간이 많이 걸리는 단점이 있다. 한 개의 분류기 학습에 소요되는 시간이 짧을수록 주어진 기간 동안 최적의 분류기를 찾을 가능성이 높아지므로, 분류기 학습의 수렴 속도 개선을 위한 추후 연구가 필요하다.

표 2. 세 학습 기준에 의한 사기 검출 분류기 (PWSC 기준, MSE 기준, 검출율 기준으로 학습한 분류기)의 분류 비용 성능 비교 (단위는 원/거래).

Table 2. Comparison results of performances of three classifiers that are trained with PWSC criterion, MSE criterion, and detection rate criterion, respectively, in terms of costs per transaction (Won/transaction).

은닉층 노드 개수	학습 기준	A 달	B 달	C 달	3개월 평균
6	MSE 최소화	-5.71∓5.88	-3.42∓5.37	-1.91∓5.48	-3.68
	검출율 최대화	-7.83∓3.60	-6.55∓2.76	-3.08∓3.14	-5.82
	PWSC 최소화	-11.26∓3.55	-7.87∓1.78	-5.61∓2.76	-8.25
12	MSE 최소화	-4.01∓8.08	-2.13∓7.79	-0.70∓7.21	-2.28
	검출율 최대화	-5.78∓4.22	-6.03∓3.57	-2.17∓3.18	-4.66
	PWSC 최소화	-11.50∓2.55	-8.12∓1.92	-4.06∓2.53	-7.89

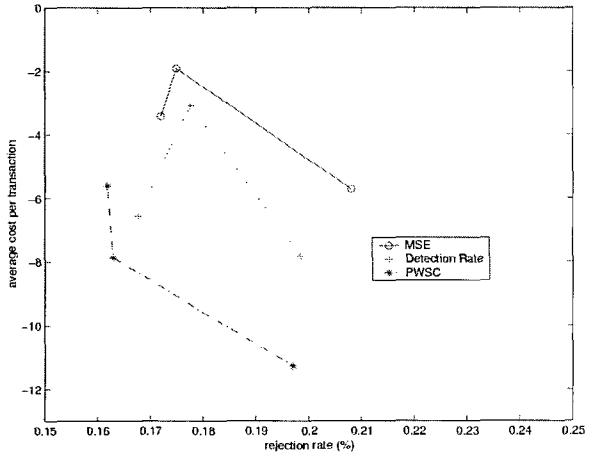


그림 4. PWSC, MSE, 검출율 기준으로 학습한 분류기의 3개월 데이터에 대한 성능(은닉층 노드 개수는 6)

Fig. 4 Performance of three classifiers that are trained with PWSC criterion, MSE criterion, and detection rate criterion, respectively, for three consecutive months (6 hidden nodes).

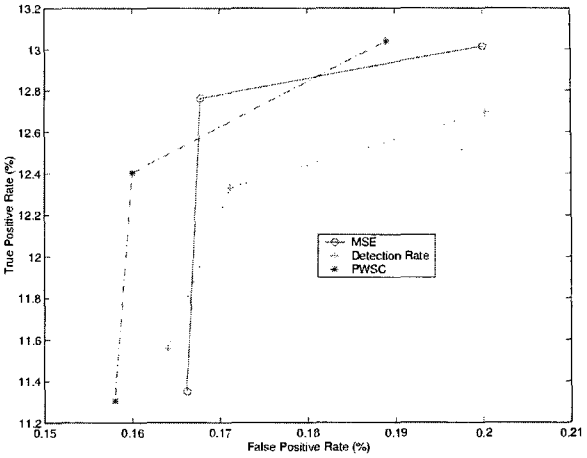


그림 5. ROC 그래프상에서의 PWSC, MSE, 검출율 기준으로 학습한 분류기의 3개월 데이터에 대한 성능 (은닉층 노드 개수는 6)

Fig. 5 Performance of three classifiers that are trained with PWSC criterion, MSE criterion, and detection rate criterion, respectively, in an ROC graph for three consecutive months (6 hidden nodes).

6. 결 론

본 논문에서는 사기 검출기의 분류 비용을 최소화하기 위한 학습 방법을 제시하였다. 일반적인 분류기의 학습 방법과 다르게, 사기 검출기의 특정 동작 구간에서의 분류 비용을 최적화하기 위한 평가 함수를 정의하고, 진화 기법을 이용하여 희망 동작 구간에서 최적의 분류 비용을 갖는 분류기를 생성하는 방법을 제시하였다. 아주 높은 정상과 사기 비율을 갖는 실제 신용카드 거래 데이터를 이용한 실험을 통해서, 제시한 방법이 선택적으로 해당 동작 구간에서의 분류 비용을 효과적으로 감소시키고, 데이터 분포의 변동이 있는 일정 기간 동안의 데이터 환경에서 강건하고(robust), 우수한 성능을 갖는 분류기를 생성할 수 있음을 보였다. 또한, 제시한 방법은 사기 검출 문제이외에도 클래스 분포의 겹쳐진 정도가 심하고 비용에 기반한 분류가 요구되는 응용 분야에도 효과적으로 적용할 수 있다.

앞으로 사기 검출 분류기의 수렴 속도를 개선하기 위하여, 복수개의 작은 크기의 분류기로 구성된 모듈 구조의 분류기에 적용하여 연구할 계획이다.

참 고 문 헌

[1] M. Weatherford, "Mining for fraud," *IEEE Intelligence Systems*, July/August Issue, pp. 4-6, 2002.

[2] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review," *Statistical Science*, vol. 17, no. 3, pp. 235-255, 2002.

[3] F. Provost and T. Fawcett, "Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions," *Proc.*

of Int'l Conf. Knowledge Discovery and Data Mining, pp. 43-48, 1997.
[4] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," vol. 30, no. 7. pp. 1145-1159, *Pattern Recognition*, 1997.
[5] J. R. Dorronsoro, F. Ginel, C. Sanchez, and C. S. Cruz, "Neural fraud detection in credit card operations," *IEEE Trans. on Neural Networks*, vol. 8, no. 4, pp. 827-834, 1997.
[6] L.-J. Park, S.-A. Kim, H.-J. Cho, T.-S. Kim, and B.-H. Wang, "A credit card fraud detection system based on hybrid of neural networks and decision tree," *Proc. of the Korea-U.S. Science and Technology Symposium*, 1998.
[7] R. Brause, T. Langsdorf, and M. Hepp, "Neural data mining for credit card fraud detection," *Proc. of 11th IEEE Int'l Conf. on Tools with Artificial Intelligence*, pp. 103-106, 1999.
[8] S. Maes, K. Tuyls, B. Vanschoenwinkel, and B. Manderick, "Credit card fraud detection using Bayesian and neural networks," *Proc. of the 1st Int'l NAISO Congress on Neuro Fuzzy Technologies*, 2002.
[9] 박래정, "신용카드 사기 검출을 위한 신경망 분류기의 진화 학습," *퍼지및지능시스템학회 논문지*, vol. 11, no. 5, pp. 400-405, 2001.
[10] P. K. Chan, W. Fan, A. L. Prodromidis, and S. J. Stolfo, "Distributed data mining in credit card fraud detection," *IEEE Intelligent Systems*, vol. 14, no. 6, pp. 67-74, 1999.
[11] W. Fan, S. J. Stolfo, K. Zhang, and P. K. Chan, "AdaCost, Misclassification cost-sensitive boosting," *Proc. of 16th Int'l Conf. Machine Learning*, pp. 97-105, 1999.
[12] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection: classification of skewed data," *Sigkdd explorations*, vol. 6, no. 1, pp. 50-59, 2004.
[13] C. Drummond and R. C. Holte, "What ROC curves can't do (and cost curves can)," Proc. of ECAI's 2004 Workshop on ROC Analysis in AI, 2004.
[14] D. B. Fogel, Evolutionary computation: *Toward a new philosophy of machine intelligence*, Wiley-IEEE press, 1999.
[15] X. Yao, "Evolving artificial neural networks," *Proceedings of the IEEE*, vol. 87, no. 9, pp. 1423-1447, 1999.

저 자 소 개

박래정(Lae-Jeong Park)
1991년: 서울대학교 전기공학과 공학사
1993년: 한국과학기술원 전기및전자공학과 공학석사
1997년: 한국과학기술원 전기및전자공학과 공학박사
1997년~1999년: LG종합기술원 정보기술연구소 선임연구원
2000년~현재: 강릉대학교 정보전자공학부 전임강사, 조교수

관심분야 : 기계 학습, 진화 연산, 데이터 마이닝, 패턴 인식
Phone : +82-33-640-2389
Fax : +82-33-646-0740
E-mail : ljpark@kangnung.ac.kr