

신용카드 불법현금용통 적발을 위한 축소된 앙상블 모형

이화경

다음 커뮤니케이션
(hmire@ycos.co.kr)

한상범

홍익대학교 산업공학과
(sbhan@hongik.ac.kr)

지원철

홍익대학교 산업공학과
(jhee@hongik.ac.kr)

불법현금용통 적발모형 개발에 앙상블 접근방법을 사용하였다. 불법현금용통은 국내 신용카드사의 손익에 영향을 미치며 최근 국제화되고 있음에도 불구하고 학문적인 접근이 이루어지지 않았다. 부정행위 적발모형(Fraud Detection Model, FDM)은 데이터 불균형 문제로 인하여 좋은 성능을 얻기 어려운데, 다수의 모형을 결합하는 앙상블이 대안으로 제시되어 왔다. 앙상블에 포함된 모형들의 다양성이 보장된다면 단일모형에 비해 더 좋은 성능을 보인다는 점은 이미 인정되고 있으며, 최근 연구 결과는 학습된 모든 기본모형들을 사용하는 것보다 적절한 기본모형들만 선택하여 앙상블에 포함시키는 것이 바람직하다는 것이다.

본 논문에서는 효과적인 불법현금용통 적발을 위하여 축소된 앙상블 기법을 사용하는데, 정확성과 다양성 척도를 사용하여 앙상블에 참여할 기본모형을 선택하는 것이다. 다양성은 앙상블을 구성하는 기본모형들 사이의 불일치(Disagreement or Ambiguity)를 의미하는데, FDM에 내재된 데이터 불균형 문제를 고려하여 두 가지 측면에 중점을 두었다. 첫째, 학습 자료의 추출 과정에서 다양성을 확보하기 위한 소수 범주의 과잉추출 방법과 적절한 훈련 방법에 대해 설명하였다. 둘째, 소수범주에 초점을 맞추어 기존의 다양성 척도를 효과적인 척도로 변형시키고, 전진추가법과 후진소거법의 동적 다양성 계산법을 도입하여 앙상블에 참여할 기본모형을 평가하였다. 실험에 사용된 학습 알고리즘은 신경망, 의사결정수와 로짓 회귀분석이었으며, 동질적 앙상블과 이질적 앙상블을 구성하여 성능평가를 하였다. 실험 결과 불법현금용통 적발모형에 있어 축소된 앙상블은 모든 기본모형이 포함된 앙상블과 성능 차이가 없었다. 축소된 앙상블은 앙상블 구성의 복잡성을 감소시키고 구현을 용이하게 한다는 점에서 FDM에서도 유력한 모형 수립 접근방법이 될 수 있음을 보였다.

논문접수일 : 2010년 02월 10일 논문수정일 : 2010년 02월 14일 게재확정일 : 2010년 03월 07일 교신저자 : 지원철

1. 서론

신용카드사의 부정행위방지시스템(Fraud Detection System, FDS)은 필수적으로 구비해야 할 시스템의 하나가 되었다. 이는 2003년을 전후 한 신용카드 대란 사태 이후 건전한 신용카드 사용문화를 구축하기 위해 정부 당국이 신용카드사들과 함

께 많은 노력을 기울인 결과이다. 현재 국내 신용카드사들은 부정행위의 대표적 형태인 분실·도난, 위변조 등에 대비할 수 있도록 사고적발 모형(Fraud Detection Model, FDM) 룰베이스 및 사고의심건 조사시스템 등으로 구성되는 FDS를 대부분 갖추고 있으며, 신용카드 부정행위 사고율도 전 세계에서 가장 낮은 수준을 유지하고 있다(금융감

* 이 논문은 2006년 홍익대학교 학술연구진흥비에 의하여 지원되었음.

독원, 2002~2006).

하지만, 최근 신용카드 사고유형 및 경로에 두 가지 큰 변화가 일어나고 있다. 첫 번째는 전자상거래 활성화로 인하여 신용카드를 이용한 전자결제 증가함에 따라 해킹에 의한 정보유출 사고가 빈번해지고 유출된 신용카드 정보를 이용한 위변조 사고가 증가한 것이다. 지금까지 위변조 사고는 해외거래에서 많이 발생하였으나 국내에서도 안심할 수 없는 단계에 이른 것으로 판단된다. 두 번째는 신용카드 가맹점들의 비정상 매출 행위이며, 대표적인 부정행위로 불법현금유통(속칭 카드깡)과 유사수신(불법자금모집)이 있다. 이와 같은 불법행위는 신용카드 선진국들에서는 볼 수 없는 형태로 국내 금융환경의 후진성과 단기간에 고수익을 실현하려는 잘못된 투자성향이 결합되어 나타난 특이한 현상이며, 최근에는 일본 중국 등에서도 발견되고 있어 점차 국제화되고 있는 부정행위로 볼 수 있다. 불법현금유통(Illegal Cash Accommodation, ICA)은 ‘여신전문금융업법’에 의해 금융질서문란으로 처벌이 가능하지만, 신용카드 가맹점을 매개체로 하여 조직적 음성적으로 이루어지기 때문에 적발이 쉽지 않다는 어려움이 있다. 금융감독 당국이 2005년부터 지속적인 적발을 독려하여 2006년부터는 가맹점 부정행위로 인한 제재조치가 감소하였으나, 최근 글로벌 금융위기로 인한 경기 침체로 다시 급증하는 추세에 있다(금융감독원 2002~2006).

카드회원이 이용대금을 정상적으로 결제한다면 신용카드사에 불법현금유통으로 인한 피해는 발생하지 않는다. 하지만, 국내 한 카드사의 분석 결과 불법현금유통 적발회원은 일반회원에 비해 카드 이용대금을 연체하는 비율이 5배 높았고, 최종적으로 대손 처리되는 비율은 3배가 높았다. 현재 국내에서 이루어지는 불법현금유통의 규모가 수

조원으로 추산되고 있음을 감안할 때 신용카드사에 상당한 손실을 주는 요인으로 볼 수 있는데, 아직까지 체계적인 대응방안에 대한 연구가 이루어지지 않았다. 따라서 본 논문에서는 불법현금유통을 신용카드 FDS의 주요 방지 대상의 하나로 인식하고, 기존의 FDS에서 주로 취급하였던 분실·도난, 위변조와는 부정행위의 성격이 완전히 다르다는 점에 유의하여 신용카드 불법현금유통 적발 모형의 개발을 시도한다.

신용카드 부정행위적발 모형에 관한 연구로 Fawcett and Provost(1996)는 사용자 프로파일을 활용하는 방안을 제시하였고, Brause et al.(1999)은 신경망을 활용하였다. 부정행위적발 모델링에 대한 체계적이고도 집중적인 연구를 수행한 Columbia 대학의 JAM 프로젝트 (Stolfo et al., 1997; Chan et al, 1999)에서는 Meta-Learning 접근방법을 사용하였다. 최근의 Chiu and Tsai (2004), Chen et al.(2005), Panigrahi et al.(2009) 등의 연구에서는 보다 복잡하고 다양한 모델링 기법들이 활용되고 있다. 이들 연구의 부정행위 대상은 분실·도난 및 위변조이었으며, 최신의 다양한 모델링 기법들이 사용된 것은 문제의 복잡성을 보여 주는 것이다.

전형적인 이진 분류문제임에도 불구하고 부정행위적발 모델링은 데이터 불균형(Imbalanced Data) 문제로 인한 어려움이 있으며, 특히 신용카드 부정행위 적발 및 네트워크 침입방지 문제는 방대한 거래량을 고려할 때 가장 극단적인 데이터 불균형 문제라고 할 수 있다. 비록 인공적으로 생성한 데이터에 대한 실험이지만 Japkowicz et al.(2002)의 연구에 의하면 데이터 불균형 문제는 학습 자료의 크기에 의해 완화될 수 있으며, 다수의 모형을 결합해 사용하는 앙상블(Ensemble 또는 Mixture-of-Experts) 모형이 일반화 능력을 높이는데 효과적이었다. 본 논문에서도 국내 신용카드사에서 실제

로 발생한 불법현금유통 사고데이터를 활용하여 적발 모델을 개발함에 있어 앙상블 접근방법을 사용한다. 앙상블 모형은 참여하는 단일 모형 수가 많을수록 일반화 능력 제고에 도움이 되는 것으로 알려져 있다. 하지만, 신용카드의 경우 부정행위의 실시간 적발이 이루어져야만 효과적인 사고방지가 가능한데, 사고적발 모형에 참여하는 기본모형의 수가 너무 많을 경우 처리 부담이 너무 커 FDS의 실제적 구현에 어려움이 따른다. 따라서 본 논문에서는 불법현금유통 적발모형을 개발함에 있어 실시간 처리 상황에서의 적용가능성을 높이기 위해 앙상블에 참여하는 기본모형의 수를 제한하면서도 일반화 능력을 손상시키지 않는 축소된 앙상블(Ensemble Size Reduction) 접근방법의 유효성을 검증하고자 한다.

본 논문의 구성은 제 2장에서 연구 배경으로 불법현금유통에 대한 소개와 부정행위 적발에 관련된 과거 연구들을 소개한다. 제 3장에서는 정확도와 다양성 척도를 사용한 축소된 앙상블 모형의 구성에 대해 설명한다. 제 4장에서는 본 논문에서 사용된 실험 데이터인 국내 신용카드사의 불법현금유통 자료와 실험 결과에 대해 설명하고 마지막으로 제 5장에서 결론과 향후 연구방향에 대해 설명한다.

2. 연구 배경

2.1 불법현금유통

신용카드 관련 부정행위는 가맹점 관련 여부에 따라 크게 세 가지로 분류할 수 있다. 첫 번째 부정 유형은 신용카드 발급 과정에서 발생하는 명의 도용(Fraudulent Application)과 미수취(Never Received Issues)이며, 두 번째는 타인 명의로 발급된

신용카드를 부정 사용하는 것으로 분실·도난(Lost and Stolen Cards) 및 위변조(Counterfeit)로 대표된다. 이러한 유형의 부정행위는 피해가 해당 신용카드 보유 회원에 국한된다는 점에서 회원 중심의 부정 유형으로 분류할 수 있다. 과거 많이 발생했던 우편·통신판매(MO/TO), 다중압인(Multiple Imprints), 계좌대여(Account Takeover) 등의 부정 유형은 신용카드 가맹점의 전자식 단말기 보급과 전자상거래 확대에 인하여 최근에는 거의 발생하지 않는다. 하지만, 전자상거래의 확대는 신용카드 부정행위가 일어나는 과정에 큰 변화를 가져왔다. 즉, 전자상거래 사이트로부터 대량의 개인정보를 유출시켜 일시적으로 사용하거나 대량의 위변조 카드를 제조하는 것이다. 이는 과거 위변조를 위하여 신용카드 단말기를 스킴밍(Skimming)하는 것에서 진화된 형태로 볼 수 있으며 대량의 신용카드 정보유출(Personal Information Leakage, PII)은 대형 사고로 이어질 가능성이 크며, 피해 범위도 여러 신용카드사에 동시에 발생한다는 점에서 문제의 심각성이 있다.

마지막 유형은 가맹점 중심의 부정행위로 불법현금유통(Illegal Cash Accommodation, ICA, 속칭 카드깡), 매출표유통(Sales Slip Circulation, SSC), 위장가맹점(Fake Merchant) 및 유사수신 행위 등이 있다. 이와 같은 신용카드 가맹점의 비정상 매출은 신용카드사에 직접적인 피해가 없다는 특징이 있지만, 탈세 수단으로 악용될 수 있으며, 위장가맹점 또는 유사수신으로 피해를 입은 다수의 회원들이 궁극적으로 신용카드사에 민원을 제기한다는 점에서 신용카드사들도 방치할 수 없는 부정행위들이다. 특히, 불법현금유통을 저지른 회원의 신용카드 이용대금은 악성연체로 이어지면서 대손 처리될 확률이 매우 높아서 신용카드사에 상당한 손실을 입히므로 최근 카드사들이 불법현금유통

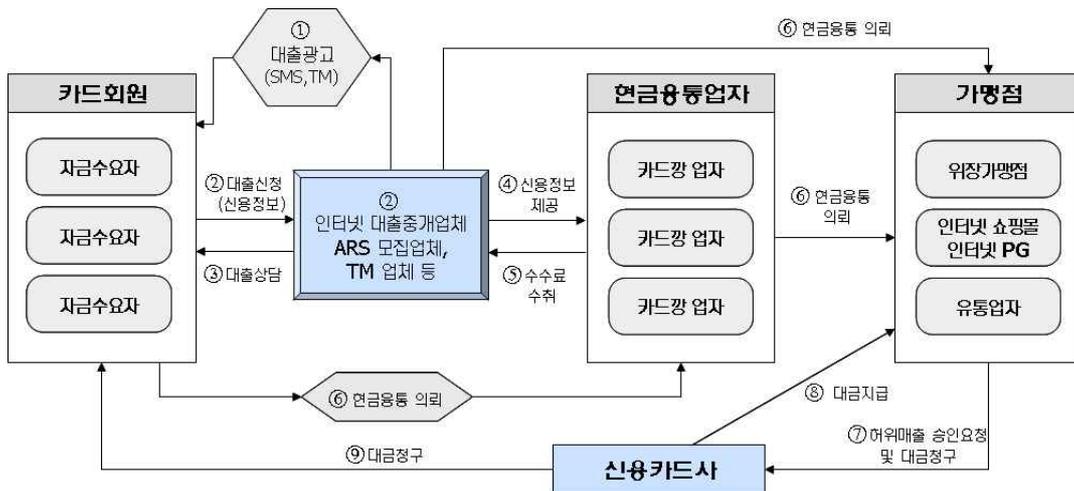
통의 방지에 많은 관심을 가지고 있다.

불법현금유통이란 가맹점이 실제 물품의 판매나 용역의 제공 없이 신용카드에 의한 거래를 가장하여 매출을 발생시키고, 허위매출금을 신용카드 회사에 청구하는 방법으로 현금을 유통하는 행위로, 여신전문금융업법 제 70조에 의거하여 가맹점주는 3년 이하의 징역 또는 2천만 원 이하의 벌금을, 이용고객은 금융질서 문란자로 7년 간 금융거래상의 제한을 받게 된다(금융감독원, 2002~2006). 2002년부터 금융기관의 부실채권 관리가 강화되면서 신용카드 회원에 대한 현금서비스 한도가 축소됨에 따라 급전수요를 만족시키지 못하는 회원들이 신용판매 승인한도를 현금화하는 속칭 카드깡을 통해 15~20%의 수수료를 부담하면서 자금유통을 하게 되었다. 또한, 대부업자들의 입장에서 보면 고율의 수수료와 함께 채무불이행 리스크를 카드사로 이전시킬 수 있다는 이점이 있으므로 급속하게 확산되었다(조성목, 2004).

고전적인 불법현금유통은 명품, 귀금속, 전자제품, 농산품, 상품권 등과 같은 고가의 환금성 물품

을 매개로 이루어져 왔으나, 최근 종교단체에의 기부금, 여행상품 및 항공권, 해외 도박 사이트 등을 이용하는 수법으로 발전하고 있으며, 특히 최근 불법현금유통 업자들이 가맹점 개설 및 운영비용이 들지 않는 인터넷 상의 결제대행서비스(Payment Gateway, PG) 및 경매 사이트를 이용하고 있어 적발에 어려움이 있다. <그림 1>은 현재 다양한 경로를 통해 이루어지고 있는 불법현금유통의 구조를 도식화한 것으로 실물거래의 유무여부와 일반 가맹점을 이용하는지 아니면 인터넷 상의 쇼핑물, PG 또는 경매를 이용하는지에 따라 구분할 수 있다. 불법현금유통 거래가 많은 가맹점들은 다음과 같은 특징이 있다.

- ① 평균 거래단가가 동일 업종에 비해 현저히 높고 일정 금액(예를 들어 100만원) 이상인 경우
- ② 가맹점 개설은 2~3개월에 불과하지만 일일 거래금액이 가맹점 한도에 육박하는 경우
- ③ 거래시간이 주간에 집중되면서 특히 은행 마



<그림 1> 불법현금유통의 발생 경로

- 감시간 이전에 모든 거래가 종료되는 경우
- ④ 동일 신용카드가 짧은 시간 안에 수회에 걸쳐 고액을 분할 승인 요청하는 거래가 많은 경우
 - ⑤ 판매물품에 비해 고액거래가 많고 동일 상품이 다른 금액으로 판매된 사례가 많은 경우
 - ⑥ 거래 금액이 50, 100만 원등으로 단수 금액이 거의 없는 경우
 - ⑦ 동일한 인터넷 IP로 수십 개의 신용카드로 다수의 거래 승인 요청을 한 경우
 - ⑧ 매출 건수가 많음에도 검색사이트에 등록이 되지 않았거나 인터넷 홈페이지가 부실한 경우

2.2 관련 연구

부정행위 적발 모델링은 전형적인 이진 분류문제이면서 데이터 불균형(Imbalanced Data) 문제이다. 즉, 목표변수의 두 범주에 속하는 모집단의 개체 수에 현격한 차이가 있는 경우 데이터 불균형 문제라고 하는데, 신용카드 부정행위(조성목, 2004; Chan et al., 1999; Chen et al., 2005; Panigrahi et al., 2009), 보험사기 적발(안철경 외 2인, 2002; 유상진 박문로, 2005; 김정동 박종수, 2006; 허준 김종우, 2007; Weisberg and Derrig, 1991; Stijin et al., 2004), 이동통신 단말기 부정사용(Fawcett and Provost, 1997), 자금세탁 방지, 네트워크 침입방지(Lee et al., 1999), Remote Alarm Sensing(Bruzzzone and Serpico, 1997), Oil Spill Detection(Kubat et al., 1998), Scene Classification(Yan et al., 2003), Protein Database(Radivojac et al., 2004), 자금세탁 방지 및 마케팅 분야의 Response Modeling 등의 분야에서 다양하게 나타난다. 특히, 신용카드 부정행위 문제의 경우 신용카드 거래량의 방대함을 고려할 때 가장 극단적인

데이터 불균형 문제라고 할 수 있다.

데이터 불균형은 학습 자료의 구성과 모형 성능 지표의 선택에 큰 영향을 미친다. 일반적인 이진분류 문제에서와 같이 학습 자료를 50 : 50으로 구성하여 정확도(Accuracy)에 의해 모형 성능을 평가하는 경우에는 좋은 예측모형을 얻을 수 없다. 이는 데이터 불균형 문제에서는 소수 범주의 정확한 분류가 더 중요한 경우가 많은데 다수 범주의 비율이 높으면 높을수록 정확도 척도가 다수 범주에 의해 왜곡되어지는 현상이 심하게 나타나고, 경우에 따라서는 학습이 전혀 이루어지지 않는 경우도 발생한다. 따라서 소수 범주의 오분류율, 소수 범주와 다수 범주의 정확도를 동시에 고려하는 기하학적 평균, 또는 LIFT 척도 등이 성능지표로 사용된다.

데이터 불균형 문제에서 좋은 성능의 모형을 얻기 위한 학습 자료의 구성에 대한 기존 연구들은 세 가지 접근방법으로 정리할 수 있다(Chawla et al., 2004; Batista, et al., 2004; 강필성 외 2인, 2005; 허준, 김종우, 2007). 첫째, 소수 범주의 수에 맞추어 다수 범주의 데이터를 추출하는 과소추출(Under-Sampling) 방법인데, 다수 범주에서 추출되는 데이터에 의해 모형 성능이 크게 영향을 받는 단점이 있으므로 소수 범주의 수가 매우 적은 경우에는 좋은 결과를 얻을 수 없다. 둘째는 다수 범주로부터의 추출할 표본크기를 정하고 이에 맞추어 소수 범주로부터 추출하는 과잉추출(Over-Sampling) 방법이다. 과잉 추출을 위해서 복원추출 또는 부족한 소수 범주의 데이터를 인공적으로 생성하는 방법이 사용되는데, 데이터 수가 증가되므로 학습시간이 증가한다는 단점이 있다(Chawla et al., 2002; Guo and Viktor, 2004). 마지막으로 오분류 비용(Misclassification Cost)을 사용하는 것이다. 즉 소수 범주에 대한 오분류에 다수 범주에 대한 오

분류보다 더 큰 오분류 비용을 적용하는 것이다. 이는 원래 데이터 분포를 왜곡시키지 않는 장점이 있는 반면에 데이터 불균형이 심할 경우에는 샘플링 방법에 비해 효과가 적다는 단점이 있다(강필성 외 2인, 2005; 허준 김중우, 2007).

주어진 불균형 데이터 문제에 대한 최적의 학습 자료 추출방법은 사전에 알 수 없다. 하지만, Japkowicz et al.(2002; Estabrooks et al., 2004)은 텍스트 분류 문제에 다양한 비율의 과잉, 과소추출에 의한 학습 자료를 생성하여 다중 모형(Mixture-of-Experts)을 구성한 후, Adaboost 기법을 사용하여 20개의 C4.5로 구성된 모형과 비교하였다. 결과는 다중 모형 즉 앙상블 모형이 학습 및 검증 단계에서 모두 좋은 성능을 보였으며, 특히 검증 단계의 성능이 더 좋았다. 강필성 외 2인(2005)의 연구에서도 SVM 앙상블 모형을 사용하여 유사한 결과를 얻었으므로 불법현금유통 모델링에서도 앙상블 모형이 유효할 것으로 판단된다.

앙상블 또는 위원회(Committee) 학습 방법은 지난 십여 년 동안 데이터마이닝 분야에서 많은 관심을 받아 왔으며 다수의 기본모형(Base Model)들을 결합하여 예측모형의 정확성 및 안정성을 높여 일반화 능력을 제고시키는 것으로, 첫째 기본모형들을 어떻게 얼마나 생성할 것인가와 둘째 기본모형들의 예측치를 어떻게 결합하여 최종 결과를 얻을 것인가의 두 문제를 해결하여야 한다(Hansen and Salamon, 1990; Ditterich, 2000; Rooney and Patterson, 2004; Zhao et al., 2005). 앙상블 모형이 효과를 얻기 위해서는 정확성을 유지하면서 가능한 다양한 기본모형들을 결합하는 것이 좋다고 알려져 있으며, 다양성을 얻기 위해 사용되는 방법은 기본모형별로 적용되는 학습계수, 입력변수 및 학습자료 들을 다르게 사용하는 것이다(Opitz, 1999). 학습 자료의 다양성을 추구하는 방

법으로 Boot-strapping 샘플을 사용하는 Bagging (Breiman, 1996, 1998)과 적응적 샘플링을 사용하는 Boosting (Freund and Schapiro, 1998)이 있으며, 기본모형의 적절한 결합에 의해 모형 성능을 개선하려는 노력은 Meta-Learning 또는 Stacked Ensemble이라고 부른다(Wolfert, 1996).

앙상블에 대한 과거 연구들은 앙상블을 구성하는 기본 모형들의 다양성을 확보하기 위한 여러 시도가 있었지만 기본 모형들 사이의 다양성과 정확성은 서로 상충관계에 놓일 경우가 많으므로 앙상블 구성할 때 양자에 대한 적절한 고려가 필요하다(Rooney et al., 2004). 다양성을 고려한 앙상블에 관한 연구의 대부분은 UCI 레포지터리의 벤치마킹 자료를 이용한 것이다. 물론, 연구의 객관성을 확보하기 위하여 벤치마킹 자료를 사용하는 것은 바람직하지만, 기존 연구에 사용된 데이터의 규모가 매우 작다는 사실을 고려할 때, 해당 연구 결과가 현실세계에서도 그대로 적용될 수 있다는 보장은 없다. 특히, 대용량 데이터로부터 유용한 지식을 발견한다는 데이터마이닝 관점에서 볼 때 현실의 대용량 데이터를 이용한 연구가 필요하다.

3. 축소된 앙상블을 위한 평가척도

불법현금유통 적발 문제에 예측모형의 일반화 능력을 제고시키기 위한 앙상블 학습 방법을 적용하기 위해서는 기본 모형들의 다양성 확보 문제와 생성된 기본 모형들 중에서 앙상블에 참여할 기본 모형의 효과적인 선정 문제를 해결하여야 한다. 앙상블은 복잡한 문제에 대해 기본모형이 갖는 편이(Bias)와 분산(Variance)의 딜레마를 완화시킬 수 있어, 이론적으로 앙상블 내의 기본모형들이 서로 독립이라면, 기본모형의 수를 무한대로 증가시키면 앙상블 모형의 오분류율도 0에 수렴하게 된다.

하지만 많은 수의 기본모형으로 구성되는 복잡도 높은 앙상블 모형은 현실적으로 구현에 어려움이 있어 활용도가 떨어지는 단점이 있으며, 특히 신용카드 부정행위방지 시스템은 실시간으로 운영되므로 처리 부담을 줄이기 위해서는 앙상블에 참여하는 기본모형의 수를 최소화하면서 부정행위 예측의 정확성을 보장할 수 있어야 한다.

앙상블을 구성할 기본모형들이 준비되면 기본 모형들의 출력값들을 결합하는 방법으로 Majority Voting과 단순 평균 등이 많이 사용되었으며, 최근에는 기본모형들의 출력값으로 상위 예측 모형을 구성하는 것을 메타 학습 또는 Stacking이라고 한다. 기본모형들을 결합할 때, 학습된 기본모형들을 모두 사용하여 왔으나, 최근 다양성 척도를 고려하여 앙상블의 성능향상에 도움이 되는 기본모형들만 포함시키는 연구들이 있었다. 이를 축소된 앙상블(Ensemble Size Reduction, ESR) 기법이라고 하며 첫째 정확도(Accuracy)가 높은 기본모형들이 결합되어야 하고 둘째, 같은 입력 데이터에 대해 가능한 서로 다른 결과를 내는(Diversity) 다양한 기본모형들이 결합되어야 좋은 결과를 기대할 수 있다. 하지만, 정확도가 높은 모형들만의 결합은 상대적으로 다양성이 떨어져 기대한 앙상블 효과를 얻기 어렵게 하는 상충작용이 존재한다. 따라서 식 (1)과 같이 기본모형의 정확성과 다양성을 동시에 고려하는 평가척도를 사용하여 앙상블 모형에의 포함 여부를 결정하는 것이 필요하다 (Krogh and Vedelsby, 1995; Zhou et al., 2003; Hernandez et al., 2005; Rooney et al., 2006).

$$\text{공헌도} = \alpha \times \text{정확성} + (1 - \alpha) \times \text{다양성}, 0 \leq \alpha \leq 1 \quad (1)$$

본 논문에서도 다양한 기본모형의 생성 후 축소

된 앙상블을 구성함에 있어 동일한 공헌도를 사용하지만 불법현금용통 적발 모델링이 갖는 극단적인 데이터 불균형이라는 특수성을 감안하여 소수 범주에 보다 중점을 두는 정확성 및 다양성 평가 척도를 사용하고자 한다.

이진분류 문제의 정확성은 <표 1>과 같은 오분류표에서 $(N00 + N11) / NT$ 로 정의된다. 하지만 심각한 데이터 불균형이 존재하는 신용카드 불법현금용통 적발과 같은 FDM 문제의 성능 측정은 일반적인 정확성 척도만으로 성능을 평가하기 어렵다. 이는 정상건의 수(Na0)가 사고건(Na1)에 비해 압도적으로 많기 때문에 특이도(Specificity)가 대부분의 모델에서 매우 높은 수치를 기록해 전체 정분류율을 왜곡시키기 때문이며, 오분류비 문제 때문에 민감도(Sensitivity) 역시 모형 성능 척도로 활용하기 어렵다. 불법현금용통의 효과적인 적발을 위해서는 N11을 최대화하면서 N01을 최소화하는 모형이 가장 좋은 모형이다. 따라서 본 논문에서는 정상건을 사고건이라고 잘못 예측하는 제1종 오류(False Positive Rate)와 사고건을 정상건이라고 잘못 예측하는 제2종 오류(False Negative Rate)를 모두 반영할 수 있는 평가척도인 가중효율성(Weighted Efficiency, WE)을 사용한다(Optiz, 1996). 모형이 사고라고 예측한 것 중에서 실제 사고였던 비율을 ‘모델적중률’이라고 부르고, 실제 사고건들 중에서 모형이 사고라고 예측한 비율을 ‘사고적중률’이라고 부른다면 WE는 다음과 같이 정의된다.

<표 1> FDM의 오분류표

		예측		합계
		정상(0)	사고(1)	
실제	정상(0)	$N00$	$N01$	$Na0$
	사고(1)	$N10$	$N11$	$Na1$
합계		$Np0$	$Np1$	NT

$$WE = \text{정분류율} \times \text{모형적중률} \times \text{사고적중률} \quad (2)$$

$$\begin{aligned} \text{where 정분류율} &= (N00 + N01) / NT \\ \text{모형적중률} &= N11 / Np1 \\ \text{사고적중률} &= N11 / Na1 \end{aligned}$$

양상블에 참여할 기본모형 선정을 위한 평가척도인 공헌도를 구성하는 또 다른 척도인 다양성에 관한 연구로는 Kuncheva et al.(2002)이 대표적이다. 이 연구에서는 10개의 다양성 척도에 대한 관계 분석 및 양상블 모형의 정확도에 미치는 영향을 분석하였지만, 다양성의 고려가 정확도 향상에 기여하는 것은 사실이지만 어떤 다양성 척도가 가장 바람직한가에 대한 결론은 내리지 못했다. 본 연구에서는 분산 개념에 기초하여 Krogh and Vedelsby (1995)가 처음 제안한 다양성 척도인 애매성 척도(Ambiguity Measure)를 사용한다. 하지만, 첫째 데이터 불균형을 고려하여 소수 범주에 중점을 두는 방안과 둘째 애매성 척도를 동적으로 계산하는 방안을 도입하여 다양성 척도를 확장한다.

입력자료 $X = \{x_1, x_2, \dots, x_N\}$, $x_k \in R^n$, 정상(Non-Fraud, NF)과 사고(Fraud, F)를 구분하는 범주 $C = \{C_{NF}, C_F\}$ 및 이에 대한 분류함수인 기본모형 $f_i(X)$, $i = 1, 2, \dots, L$ 가 주어졌을 때 양상블 모형 $\bar{f}(\cdot)$ 의 출력값을 기본모형의 출력값들의 단순 평균으로 구한다면, i 번째 기본 모형의 다양성(DV, Diversity) 척도는 양상블 모형과의 출력값 차이의 제곱 합으로 구한다. 본 연구에서는 불법현금유통 문제의 특성을 반영하여 다양성 척도인 DV를 범주별로-정상(DvN)과 사고(DvF) - 다음과 같이 분할한다.

$$DV_i = \sum_{k=1}^N \{f_i(x_k) - \bar{f}(x_k)\}^2 \quad (3)$$

$$DvN_i = \sum_{k \in C_{NF}} \{f_i(x_k) - \bar{f}(x_k)\}^2 \quad (4)$$

$$DvF_i = \sum_{k \in C_F} \{f_i(x_k) - \bar{f}(x_k)\}^2 \quad (5)$$

$$\text{where } \bar{f}(\cdot) = \frac{1}{L} \sum_{j=1}^L f_j(\cdot)$$

이는 데이터 불균형 문제로 인해 소수 범주에 대한 예측의 다양성이 반영되는 못하는 현상을 방지하기 위한 것으로, 다양성 척도로 소수 범주만을 고려한 다양성 척도인 DvF와 다수 범주의 다양성과 소수범주의 다양성을 동일하게 반영하는 척도인 DvH를 추가한 것이다. 따라서 양상블에 참여할 기본모형 선정을 위한 공헌도를 구성하는 정확성과 다양성 척도를 정리하면 다음과 같다.

$$\widetilde{WE}_i = WE_i / \sum_{k=1}^L WE_k \quad (6)$$

$$\widetilde{DV}_i = DV_i / \sum_{k=1}^L DV_k \quad (7)$$

$$\widetilde{DvF}_i = DvF_i / \sum_{k=1}^L DvF_k \quad (8)$$

$$\widetilde{DvH}_i = (\widetilde{DvN}_i + \widetilde{DvF}_i) / 2, \quad (9)$$

$$\text{where } i = 1, 2, \dots, L$$

식 (3)~식 (5)에서 다양성 척도를 구할 때 사전에 생성된 기본모형 전체를 사용하여 양상블을 구성한다고 가정하였다. 이와 같이 다양성 척도를 정적으로 구할 경우 최종 양상블 모형에 참여하는 기본모형들 사이의 다양성이 보장되지 않을 수 있다. 따라서 본 논문에서는 실제로 양상블에 참여한 기본모형들과 후보 기본모형들 사이의 다양성을 동적으로 측정하는 방안을 도입한다. 동적 다양성을 위해서 최고 성능을 보이는 기본모형으로부터 시작하여 양상블 참여 모형의 수를 하나씩 증가시켜가는 전진추가법(Forward Addition)과 모든 기본모형이 포함된 양상블로부터 출발하여 공헌도가 가장 낮은 기본모형을 하나씩 제거해 나가는

후진소거법(Backward Elimination)의 두 가지 방법을 적용한다. ESL을 앙상블에 참여하는 기본모형들의 집합으로 $\phi(f)$ 를 기본모형 f 의 공헌도로 각각 정의하면 전진추가법과 후진소거법은 다음과 같이 정리할 수 있다.

전진추가법
1. $ESL \leftarrow \{\emptyset\}; cmax = 0; \phi(f_{cmax}) = 0;$ 2. for each $f_i, i \in ESL, i = 1, 2, \dots, N$ do; 2.1. if $\phi(f_i) \geq \phi(f_{cmax}), cmax \leftarrow i;$ 2.2. end; 3. $ESL \leftarrow ESL + \{f_{cmax}\}$ 4. Reconstruct ensemble $\bar{f}(\cdot)$ using ESL 5. Evaluate accuracy of $\bar{f}(\cdot)$ 6. if satisfied then stop; otherwise goto 2;

후진소거법
1. $ESL \leftarrow \{f_i, i = 1, 2, \dots, N\}; cmax = 0; \phi(f_{cmax}) = \infty;$ 2. for each $f_i, i \in ESL, i = 1, 2, \dots, N$ do; 2.1. if $\phi(f_i) < \phi(f_{cmax}), cmax \leftarrow i;$ 2.2. end; 3. $ESL \leftarrow ESL - \{f_{cmax}\}$ 4. Reconstruct ensemble $\bar{f}(\cdot)$ using ESL 5. Evaluate accuracy of $\bar{f}(\cdot)$ 6. if satisfied then stop; otherwise goto 2;

4. 실험 설계

4.1 실험 자료

<표 1>은 본 연구에서 실험에 사용한 자료를

정리한 것이다. 국내 신용카드사의 실제 거래 자료를 6개월 간 수집하여 학습 자료로 사용하였고, 7개월째 승인자료를 1차 테스트 자료로 사용하였다. 또, 학습에 사용한 기간으로부터 6개월 후의 1개월 자료를 2차 테스트 자료로 사용하였다. <표 1>의 승인건당 승인금액의 평균 및 중앙값에서 볼 수 있듯이 불법현금응통의 건당 금액이 정상건에 비해 상당히 높음을 알 수 있다. 승인금액이 높다고 불법현금응통이라고 볼 수 없지만 불법현금응통의 평균 승인금액이 높은 점과 분할 승인에 의해 적발을 회피하려는 시도를 방지하여야 한다는 점을 고려하여 승인금액 30만 원 이상의 승인건만을 실험에 사용하였다. 수집된 실험 자료의 정상 승인건은 797만 건이었으며, 정상 대비 사고 비율은 약 1,196 : 1이었다. 1, 2차 테스트 자료는 각기 122만 및 120만 건이었으며, 특히 2차 테스트 자료의 정상 대비 사고 비가 876 : 1로 낮아졌는데 이는 불법현금응통 적발 업무가 본격화되어 사고적발이 증가하였기 때문이다.

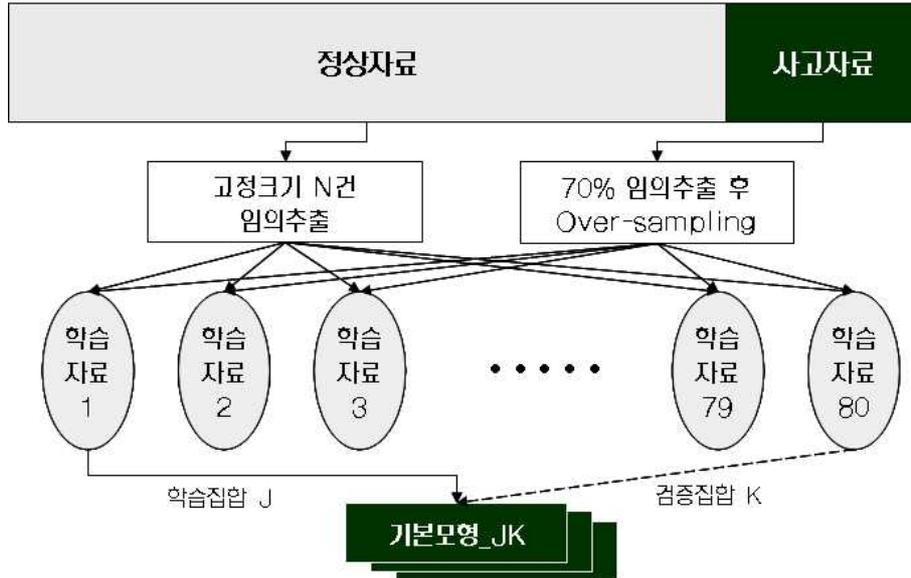
4.2 기본모형의 다양성 확보

앙상블 모형의 일반화 능력을 높이기 위한 기본모형들의 다양성을 확보하기 위하여 본 논문에서는 다음과 같이 접근하였다. 첫 번째는 수집된 대용량 승인자료를 활용한 다양한 학습 자료의 생성이다. <그림 2>에서와 같이 데이터 불균형을 해결

<표 1> 실험 자료

(단위 : 건, 천원)

	불법현금응통 건수	정상 : 사고	승인금액 평균		승인금액 중앙값	
			정상	사고	정상	사고
학습자료	6,666	1,196 : 1	601	1,659	330	1,128
1차 테스트 자료	1,056	1,156 : 1	575	1,970	327	1,475
2차 테스트 자료	1,376	876 : 1	647	2,089	338	1,500



<그림 2> 학습 자료의 생성방법

하기 위하여 소수 범주의 과잉 추출 방법을 사용하였는데 정상 대 사고비가 1000 : 1을 상회하므로 다수 범주의 과소추출의 경우에는 추출된 학습 자료들의 모집단 대표성에 심각한 문제가 있다고 판단했기 때문이다. 각 학습 자료는 정상건에서 40만건을 임의추출하고 불법현금용통 사고건 즉 소수 범주에서 70%를 임의 추출한 후 정상 대 사고비가 10 : 1이 되도록 복원 추출하였다. <그림 2>에 서와 같이 학습 자료를 생성한 것은 모집단에 대한 대표성을 유지하면서 학습 자료의 다양성을 보장하기 위한 것이다.

학습 자료는 총 80개를 생성하였는데 40개는 기본모형의 학습(Train)에 사용하고 나머지 40개는 학습 과정에서 검증용(Validation)으로 사용하였다. 의사결정수와 신경망의 경우 모형의 과잉적합을 방지하기 위해서는 검증용 학습 자료의 사용이 매우 중요하며 또 신용카드 부정행위 적발 모형에

있어 어떤 카드의 사고건은 얼마든지 다른 카드의 정상적인 거래일 수 있다는 사실에 착안하여 가능하면 학습과정에서 정상건에 대한 오분류율을 최대한 억제하기 위하여 별도로 학습 자료와 같은 크기의 검증용 자료를 사용하는 것이다. 모형의 성능 테스트는 학습 자료와는 다른 기간의 데이터를 사용하여 2차에 걸쳐서 수행하였다.

기본모형의 다양성을 확보하기 위한 또 다른 방법은 다양한 학습 알고리즘의 사용이다. 기본모형의 학습은 40개의 학습 자료를 임의 선정한 후, 로짓 회귀분석(Logit Regression, LR), 의사결정수(Decision Tree, DT) 및 신경망(Neural Networks, NN)의 세 가지 알고리즘을 SAS사의 데이터 마이닝 툴인 Enterprise Miner V4.3을 사용하여 각각 학습시켰다. 의사결정수 모형을 위해서는 C4.5에 해당하는 엔트로피 기준을 사용하였으며(Quilan, 1996), 신경망 모형을 위해서는 다계층 퍼셉트론

(Multi-Layered Perceptron)을 사용하였다. 의사결정수는 자체적인 변수선택 기능을 가지고 있고, 로짓분석의 경우 Stepwise 절차(유의수준 5%)를 이용하였기 때문에 두 모형을 위한 입력변수로는 생성한 모든 입력변수를 사용하였다. 따라서 학습자료에 따라 선택된 입력변수의 차이가 발생되었다. 신경망의 경우 변수선택을 위한 절차가 없으므로 의사결정수와 로짓분석에서 선택된 변수들을 정리하여 인위적으로 조정하였다. 앙상블을 구성하기 위해 기본 모형들의 출력값을 결합하는 방법으로는 단순평균(Simple Average)을 사용하였다. 모형의 성능평가는 출력값 0.5 이상을 사고건으로 판단하여 계산된 정확성 척도인 가중효율성(WE)을 사용하였다.

4.3 축소된 앙상블

앙상블 모형의 일반화 능력을 손상시키지 않으면서 앙상블에 참여하는 기본모형의 수를 최소화하려면 기본모형들의 정확성과 다양성이 적절히 조화되어야 한다. 이를 위해 본 논문에서는 두 척도의 가중평균인 공현도를 이용하여 앙상블에 참여하는 기본모형들을 결정하는데, 정확성 척도로는 가중효율성(WE)을 사용하고 다양성 척도로는 애매성(Ambiguity)을 세 가지 관점에서 변형한 척도를 사용한다. 앙상블에 참여하는 기본모형들의 최적 선택 또는 정확성과 다양성의 최적조합에 의한 기본모형의 선택 문제에 GASEN(Zhou et al., 2002)과 같이 유전자 알고리즘을 사용할 수도 있지만 본 연구의 학습자료 크기를 감안하면 계산 부담을 감당하기 어렵다. 따라서 본 연구에서는 공현도를 구성하는 파라미터 α 의 값을 $\alpha = 0.0, 0.25, 0.5, 0.75, 1.0$ 의 다섯 가지를 사용하여 앙상블 모형에 참여할 기본모형들을 선택하는 실험을 하였다.

따라서 본 연구에서 앙상블 구성원을 선택하기 위해 사용한 공현도의 구성은 <표 2>와 같이 모두 13가지 경우이며, 괄호 안 숫자는 제 2장에서 설명한 수식 번호이다.

<표 2> 공현도의 구성

구 분	정확성 척도	다양성 척도	α
조합 1	$\overline{WE}(6)$	-	1
조합 2	-	$\overline{DV}(7)$	0
조합 3	-	$\overline{DvF}(8)$	0
조합 4	-	$\overline{DvH}(9)$	0
조합 5	$\overline{WE}(6)$	$\overline{DV}(7)$	0.25
조합 6	$\overline{WE}(6)$	$\overline{DV}(7)$	0.5
조합 7	$\overline{WE}(6)$	$\overline{DV}(7)$	0.75
조합 8	$\overline{WE}(6)$	$\overline{DvF}(8)$	0.25
조합 9	$\overline{WE}(6)$	$\overline{DvF}(8)$	0.5
조합 10	$\overline{WE}(6)$	$\overline{DvF}(8)$	0.75
조합 11	$\overline{WE}(6)$	$\overline{DvH}(9)$	0.25
조합 12	$\overline{WE}(6)$	$\overline{DvH}(9)$	0.5
조합 13	$\overline{WE}(6)$	$\overline{DvH}(9)$	0.75

실험은 13가지 선택 기준에 따라 공현도가 높은 순서대로 상위 m 개의 기본모형으로 축소된 앙상블을 구성하는 방식으로 진행하였는데, $m = 5, 10, 15, 20$ 의 네 가지 조합을 사용함으로써 실험 조건 별로 각각 52가지의 축소된 앙상블 모형을 구성한 후, 최대 성능을 보인 단일모형과 생성한 모든 기본모형을 참여시킨 앙상블 모형과 함께 비교를 하였다. 본 논문의 실험 조건은 학습 알고리즘과 다양성 척도 계산 방법을 의미한다. 학습 알고리즘은 신경망, 의사결정수와 로짓 회귀분석의 세 가지를 사용하였는데 동일 알고리즘만으로 앙상블을 구성하는 동질적 학습(Homogeneous Learning)과 학습된 알고리즘들을 혼합하여 앙상블을 구성하

는 이질적 학습(Heterogeneous Learning)으로 구분할 수 있다. 다양성 척도 계산 방법은 정적 다양성과 동적 다양성의 두 가지이며, 동적 다양성은 다시 전진추가법과 후진소거법으로 구분하였다.

5. 실험 결과

축소된 앙상블의 유효성을 검증하기 위해 본 논문에서 사용한 기본모형은 신경망 의사결정수 및 로짓분석의 알고리즘별로 40개씩으로 합계 총 120개이다. <표 3>은 정확성 척도인 가중효율성을 사용하여 학습 알고리즘별로 단일모형의 성능에 대한 기본 통계를 나타낸 것이다. 신경망 모형이 학습에서는 가장 좋은 성능을 보였으며, 테스트에서는 의사결정수가 가장 좋은 성능을 보였다. 반면에 로짓분석의 경우 학습 및 테스트에서 가장 적은 분산을 보임으로써 모형의 성능은 떨어지지만 안정성은 가장 좋은 것으로 나타났다. 일반적으로 정상건 대비 사고건의 비 즉 사고율이 동일하게 유지된다면 시간이 지남에 따라 모형의 성능이 하락

하는 것이 일반적일 것이다. 1차 테스트 보다 2차 테스트의 성능이 더 좋았던 것은 2차 테스트 기간 동안 불법현금유통 적발건이 많아서 사고율이 1차 테스트 기간에 비해 높았던 것이 원인으로 판단된다.

<표 4>는 단일모형과 앙상블 모형의 성능 비교를 나타낸 것이다. 신경망, 트리(의사결정수), 로짓 알고리즘 별로 최고 성능을 보인 단일모형과 학습된 40개의 기본모형 전부를 이용하여 동질적 앙상블을 구성한 경우를 비교한 것이다. 추가로 생성한 기본모형 120개 전부로 앙상블을 구성한 경우를 함께 비교하였다. <표 4>에서 모형 명칭으로 신경망은 NN으로, 의사결정수는 TREE로, 로짓 회귀분석은 REG로 각각 표시했으며, _Best는 해당 알고리즘 모형 중 최고성능을 의미하며 _40은 동질적 앙상블 모형을 HETERO_120은 모든 기본모형을 사용한 앙상블을 의미한다. 또, 모델적중률, 사고적중률 및 오분류율은 가중효율성(WE)의 계산근거를 보인 것이다. 학습 과정에서는 NN_40이 가장 좋은 성능을 보였으며, 1, 2차 테스트 모두 HETERO_120이 가장 좋은 성능을 보였다. 특히, 신경망은 앙상블을 구성했을 경우 1, 2차 테스트에서 상당한 성능 개선을 보인 반면, 로짓분석의 경우는 앙상블이 단일모형보다 성능이 좋지 않았다. 이는 로짓 회귀분석이 안정성은 좋으나 심한 데이터 불균형이 존재하는 경우에는 바람직한 분류 알고리즘이라고 할 수 없음을 의미하며, 부정행위 적발 모형 개발에서 신경망 앙상블의 유효성을 입증하는 것으로 볼 수 있다. 하지만 <표 3>에서와 같이 신경망 단일모형들의 분산이 가장 크다는 점을 고려하면 단일모형을 사용할 경우 주의가 요구된다.

축소된 앙상블을 구성하기 위하여 사용한 13가지 공현도 척도에 따라 실험한 결과를 부록에 그래프로 나타내었다. 각 공현도 척도에 따라 각 알

<표 3> 학습된 기본모형의 통계

구 분		신경망	의사결정수	로짓분석
학습	최대값	3.156%	2.739%	1.819%
	최소값	2.360%	2.075%	1.774%
	평균	2.691%	2.365%	1.799%
	분산	0.235%	0.275%	0.016%
1차 테스트	최대값	1.180%	1.106%	0.967%
	최소값	0.722%	0.874%	0.880%
	평균	0.906%	0.991%	0.919%
	분산	0.141%	0.146%	0.095%
2차 테스트	최대값	1.403%	1.572%	1.384%
	최소값	0.991%	1.165%	1.206%
	평균	1.152%	1.348%	1.299%
	분산	0.162%	0.159%	0.107%

<표 4> 앙상블과 단일모형 성능 비교

구 분	순서	모형	WE	모델적중률	사고적중률	오분류율
학습	1	NN_40	4.33%	4.82%	91.5%	1.93%
	2	HETRO_120	3.88%	4.89%	80.7%	1.69%
	3	NN_Best	3.16%	3.62%	89.5%	2.55%
	4	TREE_40	2.78%	3.65%	78.0%	2.21%
	5	TREE_BEST	2.74%	3.68%	76.0%	2.14%
	6	REG_BEST	1.82%	2.68%	69.9%	2.73%
	7	REG_40	1.80%	2.68%	69.0%	2.69%
1차 테스트	1	HETRO_120	1.54%	2.54%	61.9%	1.94%
	2	NN_40	1.36%	2.46%	56.2%	1.83%
	3	NN_Best	1.18%	1.79%	67.9%	3.01%
	4	TREE_40	1.12%	1.93%	59.3%	2.45%
	5	TREE_BEST	1.11%	1.88%	60.4%	2.56%
	6	REG_BEST	0.97%	1.65%	60.4%	2.92%
	7	REG_40	0.90%	1.50%	61.9%	3.29%
2차 테스트	1	HETRO_120	2.07%	3.33%	63.9%	2.79%
	2	NN_40	1.93%	3.36%	58.9%	2.56%
	3	TREE_BEST	1.57%	2.50%	65.3%	3.81%
	4	TREE_40	1.55%	2.41%	67.2%	4.07%
	5	NN_Best	1.40%	2.40%	60.8%	3.71%
	6	REG_BEST	1.38%	2.23%	64.8%	4.25%
	7	REG_40	1.30%	2.09%	65.1%	4.56%

고리즘 별로 단일모형의 최대성능과 비교하여 축소된 앙상블의 성능개선 비율을 정적 다양성(STATIC), 전진추가법(FORWARD) 및 후진소거법(BACKWARD)으로 구별하여 정리한 것이다. 부록 <그림 1>~<그림 3>에서 볼 수 있듯이 다양성 계산 방법에 상관없이 신경망(NN)과 이질적 학습(Hetero)의 경우 큰 폭의 성능 개선이 이루어진 반면 의사결정수(TREE)와 로짓분석(REG)의 성능 개선은 10% 정도에 머물렀다. 특히 로짓분석의 경우에는 앙상블 구성 기본모형의 수가 5개일 경우 2차 테스트에서 단일모형보다 성능이 못한 경우도 발생하였다. 13가지 공현도 척도 관점에서 살펴보면 정확도나 다양성만 고려한 경우보다는

양자를 동시에 고려하는 것이 더 좋았으며, 특히, 정적 다양성 보다는 동적 다양성 즉 전진추가법이나 후진소거법을 적용했을 때 효과가 더 좋았다. 다양성 척도도 정상건과 사고건을 모두 감안한 다양성(DV)보다 사고건 즉 소수범주의 다양성에 중점을 두는 척도가 더 좋았으며 특히 사고건만 고려한 DvF가 가장 효과가 좋았다. 특히, 정적 다양성 계산의 경우 DvF를 사용한 경우, 즉 공현도 조합 3, 8, 9, 10일 때 HETERO 앙상블이 신경망 앙상블 보다 성능개선 효과가 좋았으며 동적 다양성의 경우에는 공현도 조합에 상관없이 HETRO 앙상블의 성능개선 효과가 가장 좋았다.

가장 효과적인 공현도 조합을 파악하기 위하여

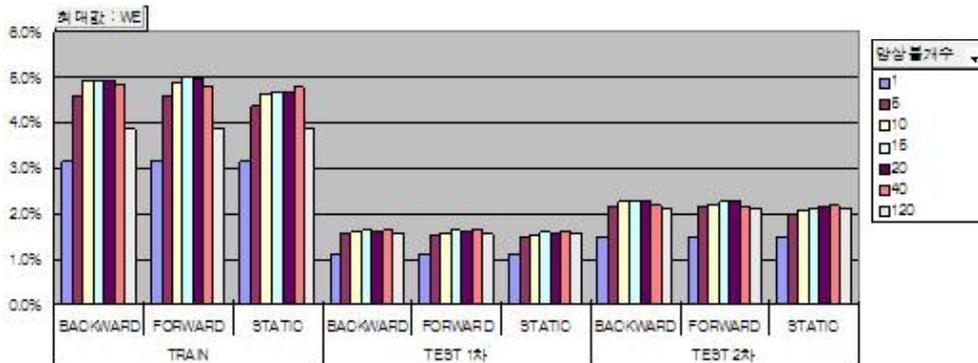
<표 5> 상위 5위의 선택된 공현도 조합

구분	학습		1차 테스트		2차 테스트	
HETERO, NN, TREE & REG						
순위	조합	빈도	조합	빈도	조합	빈도
1	6	11	9	10	10	12
2	1	10	8	9	9	10
3	5	9	10	7	3	7
4	7	7	7	6	8	5
5	8, 10	4	3	5	11	4
HETRO & NN						
순위	조합	빈도	조합	빈도	조합	빈도
1	6	10	8	8	10	9
2	5	8	9	4	9	6
3	7	3	10	4	3	4
4	8	3	3	4	8	4
5	9	1	11,12	3	11,12	2

<표 5>에 각 실험 조건, 즉 주어진 다양성 계산방법과 앙상블 참여 기본모형의 수에서 가장 좋은 성능을 보인 공현도 조합의 선택 회수를 상위 5위 까지 집계하여 나타냈다. 학습과정에서는 소수 범주와 다수 범주의 다양성을 모두 고려하는 공현도 조합 5, 6, 7이 상위에 위치하고 소수 범주의 다양성만 고려하는 8, 9, 10은 5위 정도였다. 하지만, 1, 2차 테스트에서는 공현도 조합 8, 9, 10이 상위로

선택되었으며, 소수범주의 다양성만 고려하는 조합 3도 상위 5위안에 포함되었다. 특히, 성능개선이 컸던 이질적 학습과 신경망으로 한정된 경우에도 조합 3, 8, 9, 10이 효과적인 것으로 나타났다.

부록 <그림 4>~<그림 6>은 소수 범주에 중점을 두는 공현도의 조합 3, 8, 9, 10에 대하여 앙상블 참여 기본모형 수에 따른 성능 변화를 보인 것이다. 이질적 앙상블의 경우가 성능개선에 가장 유리하며, 특히 앙상블 참여 기본모형의 수가 5개로 적을 경우에는 동적 다양성을 사용하는 것이 유리하다는 것과 생성한 모든 기본모형을 사용하지 않아도, 즉 축소된 앙상블의 성능이 더 좋은 경우도 있음을 알 수 있다. <그림 3>은 동적 다양성의 유효성 여부를 파악하기 위하여 각 학습 알고리즘별로 다양성 계산 방법에 따라 앙상블을 구성하여 최고 성능을 보인 경우를 그림으로 나타낸 것이다. 동적 다양성 즉 다양성 척도 계산에 전진추가법이나 후진소거법을 적용하는 것이 정적 다양성 보다 학습과 1, 2차 테스트에서 모두 약간씩 유리한 것으로 나타났지만 통계적으로 유의하지는 않았다. 또, 로짓 분석의 경우 앙상블에 의한 효과가 다양성 계산 방법을 변화시켜도 거의 없는 것으로 나타났는데 원인으로 다양한 로짓 기본모형의 생성



<그림 3> 동적 다양성의 성능 비교

<표 6> 앙상블 참여 기본모형에 따른 성능변화

최대값 (WE)		앙상블내 기본모형 개수							
구 분	알고리즘	1	5	10	15	20	40	120	최대값
TRAIN	HETERO	3.16%	4.59%	4.95%	4.99%	4.98%	4.84%	3.88%	4.99%
	NN	3.16%	4.28%	4.45%	4.49%	4.45%	4.33%	0.00%	4.49%
	TREE	2.74%	3.14%	3.13%	2.97%	3.01%	2.78%	0.00%	3.14%
	REG	1.82%	1.81%	1.81%	1.80%	1.81%	1.80%	0.00%	1.82%
	최대값	3.16%	4.59%	4.95%	4.99%	4.98%	4.84%	3.88%	4.99%
1차 TEST	HETERO	0.93%	1.57%	1.59%	1.64%	1.61%	1.65%	1.54%	1.65%
	NN	0.93%	1.39%	1.35%	1.38%	1.43%	1.33%	0.00%	1.43%
	TREE	1.11%	1.29%	1.24%	1.22%	1.23%	1.12%	0.00%	1.29%
	REG	0.89%	0.93%	0.93%	0.92%	0.92%	0.90%	0.00%	0.93%
	최대값	1.11%	1.57%	1.59%	1.64%	1.61%	1.65%	1.54%	1.65%
2차 TEST	HETERO	1.40%	2.17%	2.26%	2.27%	2.28%	2.19%	2.13%	2.28%
	NN	1.40%	1.87%	1.90%	1.93%	1.94%	1.82%	0.00%	1.94%
	TREE	1.51%	1.72%	1.69%	1.62%	1.62%	1.55%	0.00%	1.72%
	REG	1.26%	1.35%	1.34%	1.31%	1.32%	1.30%	0.00%	1.35%
	최대값	1.51%	2.17%	2.26%	2.27%	2.28%	2.19%	2.13%	2.28%

에 실패한 것인지 아니면 로짓분석이 데이터 불균형이 심한 부정행위 적발 모형에 부적합한지의 여부는 추가 연구가 필요하다고 판단된다.

마지막으로 <표 6>는 다양성 계산방법에 상관없이 앙상블 참여 기본모형의 수에 따른 최고 성능을 학습 알고리즘 별로 요약한 것이다. HETRO가 학습 및 테스트 과정 모두 가장 좋은 성능을 보였으며 다음으로 신경망이 좋았다. 40개의 기본모형들로 구성된 HETERO의 경우가 1차 테스트에서 가장 좋은 성능을 보인 경우를 제외하면 HETERO와 신경망 모두 앙상블 참여 기본모형의 수가 15, 20일 때 120개로 구성된 앙상블보다 더 좋은 성능을 보였다. 이는 축소된 앙상블이 적어도 성능 훼손 없이 구성될 수 있음을 보여주는 것이다. 이론적으로 앙상블에 참여하는 모형의 수가 증가할수록 앙상블의 성능은 개선되는 것으로 알려져 있지만 현실에서 앙상블 참여 모형의 수를 무한정

늘릴 수 없다면 다양한 기본모형의 생성이 앙상블의 효과를 보장하는 중요한 요인이다. 따라서 축소된 앙상블도 앙상블에 참여하는 기본모형의 다양성을 확보하는 것이 중요하다.

6. 결론

신용카드 불법현금유동 적발 모형의 개발에 있어 축소된 앙상블 기법이 유효한가를 실험하기 위하여 국내 신용카드사로부터 수집한 대용량 실거래 자료를 사용하였다. 앙상블을 구성하는 기본모형들의 다양성을 확보하기 위하여 데이터 불균형을 감안하여 소수범주 과잉추출에 의한 학습자료 40개 검증자료 40개를 생성하였고, 신경망 의사결정수 로짓분석의 세 가지 학습 알고리즘의 사용과 학습 시 입력변수의 다양성이 보장되도록 하였다. 또, 축소된 앙상블을 구성하는 기본모형의 선택기

준으로 정확성과 다양성을 동시에 고려하는 공현도를 사용하였다. 정확성 척도로는 가중효율성(Weighted Efficiency)을 사용하였고 다양성 척도로는 애매성(Ambiguity)을 데이터 불균형을 감안하여 변형시켜 사용하였으며 다양성의 계산에 동적 계산 방법을 고려하였다.

본 연구의 실험 결과는 다음과 같이 요약할 수 있다. 첫째, 축소된 앙상블에 참여하는 기본모형의 수를 효과적으로 줄이기 위해서는 정확성과 다양성을 동시에 고려하여 앙상블 모형의 구성원을 선택하는 것이 효과적이다. 둘째, 학습 알고리즘의 관점에서 볼 때 신경망, 의사결정수 및 로짓 분석을 혼합하는 이질적 학습(Heterogeneous Learning)이 다양성 확보 측면 및 성능 측면에서 모두 동질적 학습 보다 효과적이었다. 셋째, 목표변수의 범주별 빈도 차이가 매우 큰 부정행위 적발모형의 경우, 목표변수 예측의 다양성 즉 사고건을 다양하게 예측할 수 있도록 앙상블 구성하는 것이 학습 자료 전체 즉 정상건과 사고건 모두에 대하여 다양성을 측정하는 것보다 효과적이다. 넷째, 다양성을 계산방법으로 동적 다양성을 사용하는 것이 적은 수의 기본모형으로 앙상블을 구성할 때 유리하다. 하지만 동적 다양성을 계산하는 진진추가법과 후진소거법의 차이는 유의하지 않았다. 마지막으로 본 논문에서는 반년이 경과한 뒤의 모형 성능 변화 추이를 살펴보았는데 축소된 앙상블은 모든 기본모형을 사용한 앙상블과 비교한 결과 성능 측면에서 불리하지 않았다.

본 논문에서는 실세계의 대용량 데이터를 사용하여 데이터 불균형이 심한 분류 문제에 있어서도 성능 손실 없이 축소된 앙상블 기법을 적용할 수 있음을 보였다. 하지만 본 논문의 실험 결과는 사용된 데이터에 대해서 제한적일 수밖에 없으므로 다른 분야에 적용할 경우에는 유의하여야 한다. 따

라서 앞으로 신용카드 불법현금유통 적발 문제 이외에 보험, 자금세탁 등의 다른 부정행위 적발 응용분야에서도 검증되어야 하며 해당 분야에서도 좋은 결과를 기대할 수 있을 것이다. 또, 현재 대용량 데이터에의 적용에 한계가 있는 SVM을 이질적 앙상블 구성에 포함시키는 연구는 부정행위 적발 모형의 성능 개선에 큰 도움이 될 것으로 기대된다. 본 논문에서는 다양성 척도로 애매성을 사용하였는데 Kuncheva and Whitaker(2002)의 연구에서 언급된 다양성 척도들을 검증하여 부정행위 적발에 축소된 앙상블 기법을 적용할 때 가장 적절한 척도를 찾는 것도 앞으로 필요한 연구이다.

참고문헌

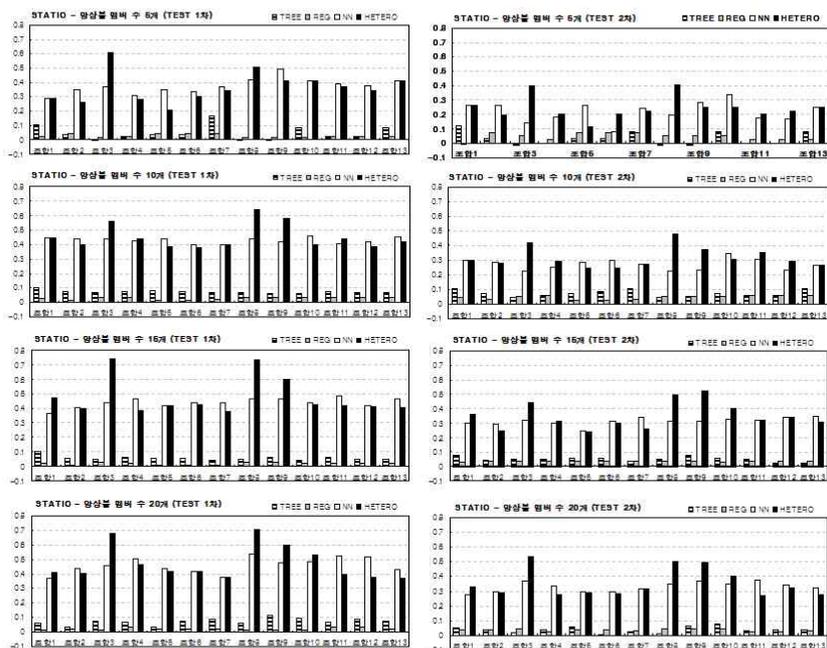
- 금융감독원 보도자료, 신용카드사 경영실적, (2002~2006).
- 안철경, 조혜원, 김경환, 국내외 보험사기관리 실태 분석 : 선진사례 및 설문분석을 중심으로, 보험개발원, (2002).
- 강필성, 이형주, 조성준, “데이터 불균형 문제에서의 SVM 앙상블 기법의 적용”, *한국정보과학회 추계학술대회논문집*, 31권 2호(2005), 706~708.
- 김정동, 박종수, “자동차보험 사기 적발 모형에 관한 연구”, (2006).
- 유상진, 박문로, “데이터마이닝 기법을 활용한 의료보험 진료비 청구 삭감분석 시스템 개발 및 구현에 관한 연구”, *Information System Review*, Vol.7(2005), 275~295.
- 조성목, “신용카드불법거래 유형 및 대응방안”, 신용카드 30호(2004).
- 허준, 김종우, “불균형 데이터 집합에서의 의사결정나무 추론”, *Information System Review*, Vol.9(2007), 45~65.
- Breiman, L., “Bagging Predictors”, *Machine Lear-*

- ning, Vol.24(1996), 123~140.
- Breiman, L., "Arcing Classifiers", *Annals of Statistics*, Vol.26(1998), 801~849.
- Batista, G., Pati, R. C. and Monard, M. C., "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data", *SIGKDD Exploration*, Vol.6(2004), 20~29.
- Brause, R., T. Langsdorf, and M. Hepp, "Neural Data Mining for Credit Card Fraud Detection", *Proceeding of the 11th IEEE International Conference on Tools with Artificial Intelligence*, November Vol.8, No.10(1999) 103~105.
- Bruzzone, L. and S. B. Serpico, "Classification of Imbalanced Remote-sensing Data by Neural Networks", *Pattern Recognition Letters*, Vol.18(1997), 1323~1328.
- Chan, P. K., W. Fan, A. L. Prodromidis and S. J. Stolfo, "Distributed Data Mining in Credit Card Fraud Detection", *IEEE Intelligent Systems*, Vol.14, No.6(1999), 67~74.
- Chawla, N. V., K. W. Boywer, L. O. Hall and W. P. Kegelmeyer, "SMOTE : Synthetic Minority Over-sampling Technique", *Journal of Artificial Intelligence Research*, Vol.16(2002), 321~357.
- Chawla, N. V., N. Japkowicz and A. Kolcz, "Editorial : Special Issue on Learning from Imbalanced Data Sets", *SIGKDD Exploration*, Vol.6(2004), 1~6.
- Chen, R. C., S. T. Luo, X. Liang and V. C. S. Lee, "Personalized approach based on SVM and ANN for detecting credit card fraud", *Proceedings of the IEEE International Conference on Neural Networks and Brain*, October(2005), 810~815.
- Chiu, C. and Chieh-Yuan Tsai, "A Web Services-Based Collaborative Scheme for Credit Card Fraud Detection", *Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service*, March Vol.28, No.31(2004), 177~181.
- Dietterich, T., "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees : Bagging, Boosting and Randomization", *Machine Learning*, Vol.40, No.2(2000), 139~157.
- Estabrooks, A., T. Jo and N. Japkowicz, "A Multiple Resampling Method for Learning from Imbalances Data Sets", *Computational Intelligence*, Vol.20, No.1(2004), 18~36.
- Fawcett, T. and F. Provost, "Combining Data Mining and Machine Learning for Effective User Profile", *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, OR, AAAI(1996), 8~13.
- Fawcett, T. and F. Provost, "Adaptive Fraud Detection", *Data Mining and Knowledge Discovery*, Vol.1(1997), 291~316.
- Freund, Y and R. Shapiro, "A Decision-theoretic Generalization of On-line Learning and an Application to Boosting", *Journal of Computer and System Sciences*, Vol.55(1997), 119~139.
- Guo, H. and H. L. Viktor, "Learning from Imbalanced data Sets with Boosting and Data Generation : The DataBoos-IM Approach", *SIGKDD Exploring*, Vol.6(2004), 30~39.
- Hansen, L. and P. Salomon, "Neural Network Ensembles", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.12, (1990), 993~1001.
- Hernandez, C., M. Fernandez, and M. Oritz, "New Experimental Ensembles of Multilayer Feedforward for Classification Problem", *Int'l Joint Conf. on Neural Networks*, (2005).

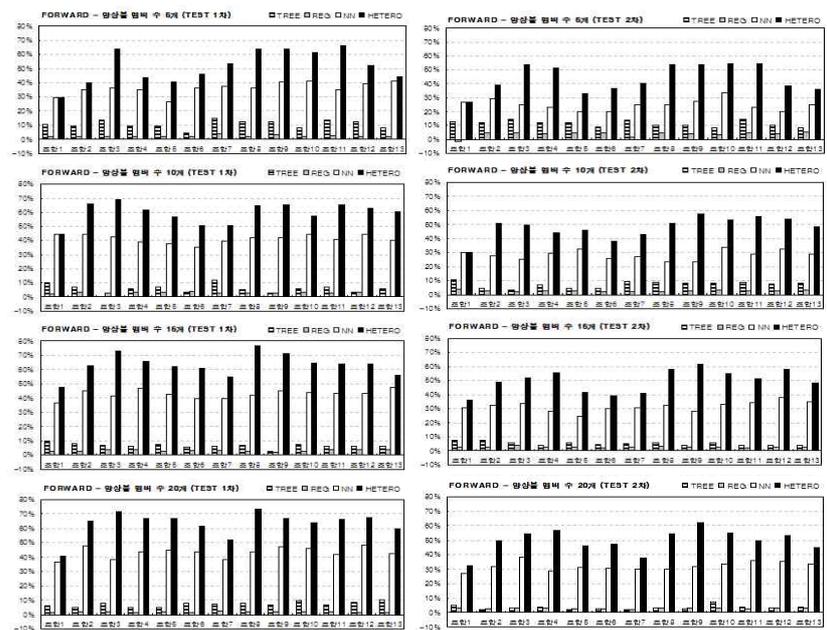
- Japkowicz N. and S. Stephen, "The Class Imbalance Problem : A Systematic Study", *Intelligent Data Analysis*, Vol.6, No.5(2002) 429~450.
- Krogh, A. and J. Vedelsby, "Neural Networks Ensembles, Cross Validation, and Active Learning", *Advances in Neural Information Processing Systems*, (1995), 231~238.
- Kuncheva, L. and C. J. Whitaker, "Measures of Diversity in Classifier Ensembles and Their Relationship with the Ensemble Accuracy", *IEEE Transactions on System, Man and Cybernetics*, Vol.32, No.2(2002), 146~156.
- Kubat, M., R. Holte and S. Matwin, "Machine Learning for the Detection of Oil Spills in Satellite Radar Images", *Machine Learning*, Vol.30(1998), 195~215
- Lee, W., S. Stolfo, and K. Mok. "A Data Mining Framework for Building Intrusion Detection Models," *Proceedings of the IEEE Symposium on Security and Privacy*, Oakland, CA, May(1999).
- Opitz, D., "Feature Selection for Ensembles", *Proc. of the 16th National Conf on Artificial Intelligence*, AAAI, (1999), 379~384.
- Opitz, D. and J. Shavlik, "Actively Searching for an Effective Neural Network Ensembles", *Connection Science*, Vol.8, No3(1996), 337~353.
- Quilan, J. R., "Bagging, Boosting, and C4.5", *Proc. of the 13th National Conf on Artificial Intelligence*, (1996), 725~730.
- Panigrahi, S., A. Kundu, S. Sural and A. K. Majumdar, "Credit card fraud detection : A fusion approach using Dempster-Shafer theory and Bayesian learning", *Information Fusion*, Vol.10(2009), 354~363.
- Radivojac, P., V. N. V. Chawla, K. A. Dunker and Z. Obradovic, "Classification and Knowledge Discovery in Protein Databases", *Journal of Biomedical Informatics*, Vol.37(2004), 224~239.
- Rooney, N., D. Patterson and C. Nugent, "Reduced Ensemble Size Stacking", *Proc. of the 16th IEEE International Conference on Tools with Artificial Intelligence(ICTAI)*, (2004), 266~271.
- Rooney, N., D. Patterson and C. Nugent, "Pruning Extension to Stacking", *Intelligent Data Analysis*, Vol.10(2006), 47~66.
- Stijn, V., R. A. Derrig and G. Dedene, "A Case Study of Applying Boosting Naive Bayes to Claim Fraud Diagnosis", *IEEE Transactions on Knowledge and Data Engineering*, (2004), 612~620.
- Stolfo, S. J., W. Fan, W. Lee, A. Prodromidis and P. K. Chan, "JAM : Java Agents for Meta-Learning over Distributed Databases", *Proc. of 3rd Int'l Conference on Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, CA., (1997), 74~81.
- Weisberg, H. I. and R. A. Derrig, "Fraud and Automobile Insurance : A Report on the Baseline Study of Bodily Injury Claims in Massachusetts," *Journal of Insurance Regulation*, Vol.9(1991), 427~541.
- Wolpert, D., "Stacked Generalization", *Neural Networks*, Vol.5(1992), 241~259.
- Yan, R., Y. Liu, R. Jin and A. Hauptman, "On Predicting Rare Classes with SVM Ensembles in Scene Classification", *IEEE International Conference on Acoustics, Speech and Signal Processing*, (2003).
- Zhao, Y., J. Gao, and X. Yang, "A Survey of Neural Network Ensembles", *International Conference on Neural Networks and Brain*, (2005), 438~442.

Zhou, Z. H., J. Wu, and W. Tang, "Ensembling
Neural Networks : Many could be better

than all", *Artificial Intelligence*, Vol.137,
No.1(2002), 239~263.

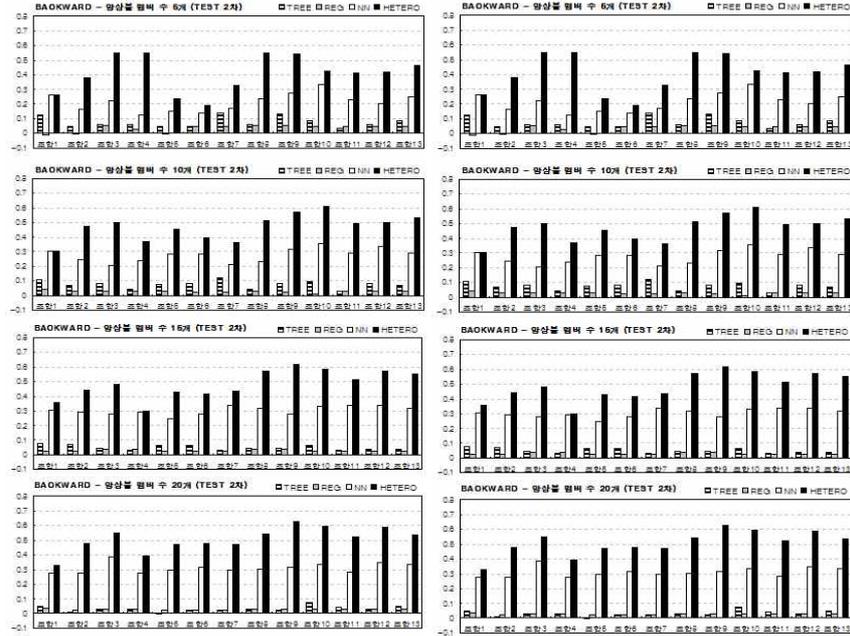


(부록 <그림 1>) 정적 다양성에 의한 축소된 앙상블의 성능개선 효과

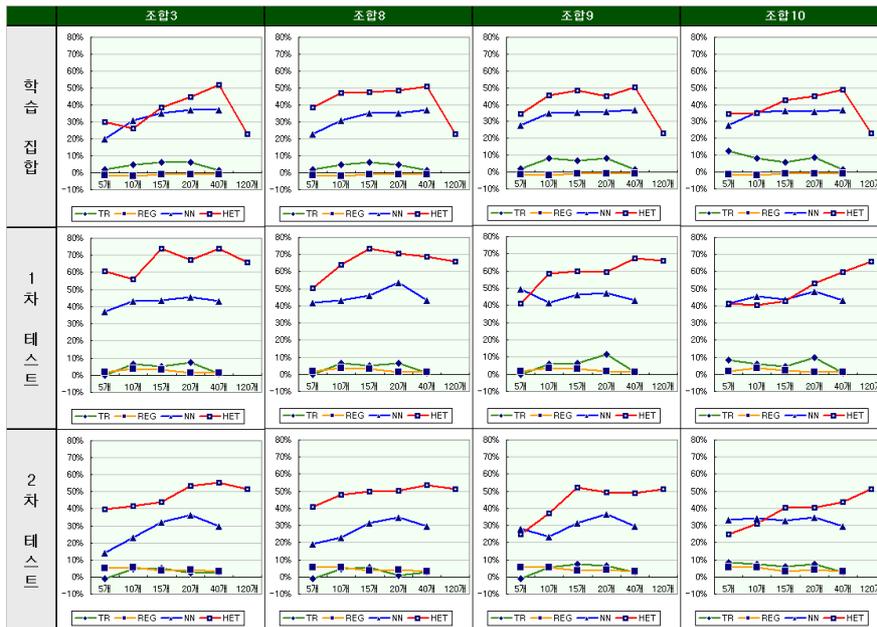


(부록 <그림 2>) 전진추가법에 의한 축소된 앙상블의 성능개선 효과

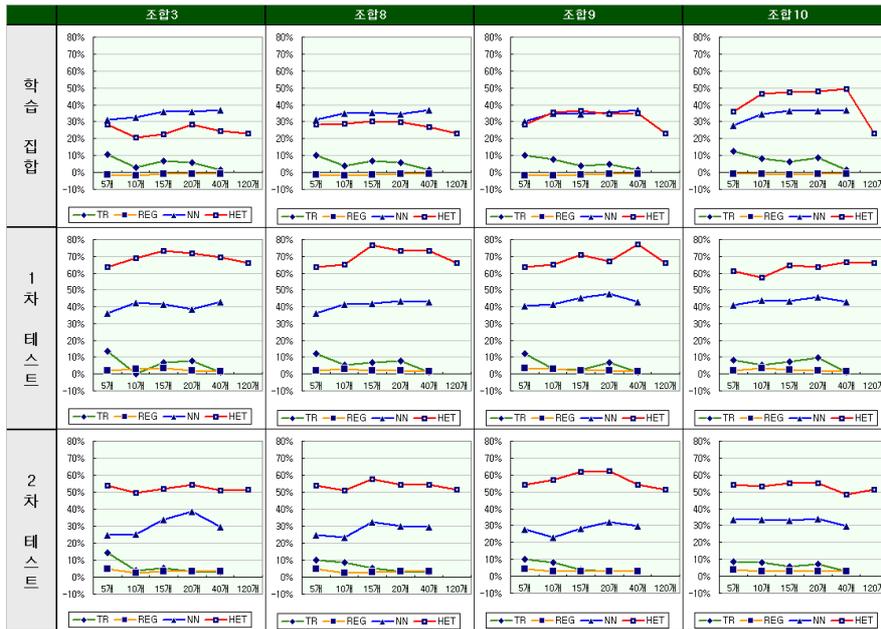
신용카드 불법현금응통 적발을 위한 축소된 앙상블 모형



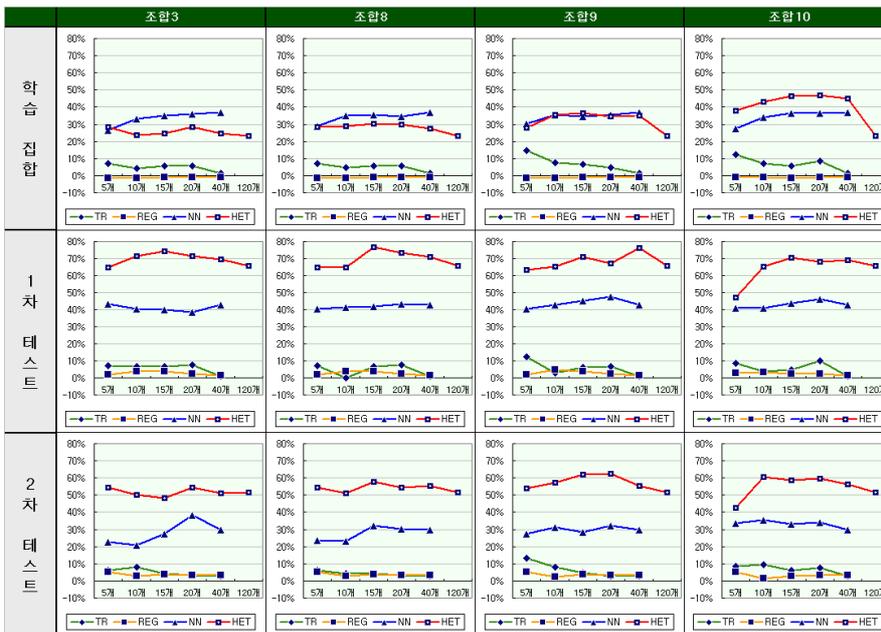
(부록 <그림 3>) 후진소거법에 의한 축소된 앙상블의 성능개선 효과



(부록 <그림 4>) 기본모형 수에 의한 앙상블의 성능 변화(정적다양성)



(부록 <그림 5>) 기본모형 수에 의한 앙상블의 성능 변화(전진추가법)



(부록 <그림 6>) 기본모형 수에 의한 앙상블의 성능 변화(후진소거법)

Abstract

Illegal Cash Accommodation Detection Modeling Using Ensemble Size Reduction

Hwa Kyung Lee* · Sang Bum Han** · Won Chul Jhee**

Ensemble approach is applied to the detection modeling of illegal cash accommodation (ICA) that is the well-known type of fraudulent usages of credit cards in far east nations and has not been addressed in the academic literatures. The performance of fraud detection model (FDM) suffers from the imbalanced data problem, which can be remedied to some extent using an ensemble of many classifiers. It is generally accepted that ensembles of classifiers produce better accuracy than a single classifier provided there is diversity in the ensemble. Furthermore, recent researches reveal that it may be better to ensemble some selected classifiers instead of all of the classifiers at hand.

For the effective detection of ICA, we adopt ensemble size reduction technique that prunes the ensemble of all classifiers using accuracy and diversity measures. The diversity in ensemble manifests itself as disagreement or ambiguity among members. Data imbalance intrinsic to FDM affects our approach for ICA detection in two ways. First, we suggest the training procedure with over-sampling methods to obtain diverse training data sets. Second, we use some variants of accuracy and diversity measures that focus on fraud class. We also dynamically calculate the diversity measure-Forward Addition and Backward Elimination. In our experiments, Neural Networks, Decision Trees and Logit Regressions are the base models as the ensemble members and the performance of homogeneous ensembles are compared with that of heterogeneous ensembles. The experimental results show that the reduced size ensemble is as accurate on average over the data-sets tested as the non-pruned version, which provides benefits in terms of its application efficiency and reduced complexity of the ensemble.

Key Words : Illegal Cash Accommodation, Fraud Detection System, Diversity Measure, Ensemble Size Reduction, Data Mining

* Daum Communication

** Dept. of IE, Hong Ik University

저 자 소개



이화경

홍익대학교 산업공학과를 졸업하고, 동 대학원에서 데이터마이닝 전공으로 석사 학위를 취득하였으며, 현재 다음커뮤니케이션 데이터마이닝팀에 재직 중이다. 주요 관심 분야는 데이터마이닝, 웹로그 마이닝, 고객행태분석 등이다.



한상범

대우증권, 한국공항공에서 근무하였으며 서울산업대학에서 산업공학 석사학위를 취득하였으며, 현재 홍익대학교 박사과정에 있다. 관심분야는 BSC, e-Business 및 데이터마이닝 등이다.



지원철

현재 홍익대학교 산업공학과 교수로 재직하고 있으며, 서울대 경영학 학사, 한국과학기술원(KAIST)에서 산업공학 석사, 경영정보시스템 박사학위를 취득하였다. 한국데이터마이닝학회 회장을 역임하였으며, 관심분야는 Fraud Detection System, 신용평가, CRM, 의료정보시스템 등이다.