



A Hybrid Data Mining Classifier for Prediction Churn Customers

Boo-Sik Kang*

Division of Service Management, Mokwon University

A B S T R A C T

The activities for customer relationship retention are most essential process of customer relationship management. If we improve the rate of churn customers about 5%, it is known that the profit rate of the company raises about 100%. Prediction activities must be preceded for management of the churn customers. If we can predict churn customers in advance, we can minimize the loss because of removing the churn factor. Data sets about churn customers have an imbalanced characteristic that the number of the churn customers is less remarkably than the number of the retention customers. One of the main issues on the models to predict churn customers is to increase prediction performance. This research deals with decision tree, neural networks, and SVM models those are mainly used to predict the churn customers and known as good prediction performance, and proposes a hybrid weighted data mining classifier. To test the prediction performance of proposed method, this study used 'churn data sets' in UCI Machine Learning Repository. Experimental results showed that the method was effective for prediction of churn customers.

© 2014 KKITS All rights reserved

KEYWORDS : Hybrid data mining, Churn predictions, Imbalanced datum, Decision trees, Neural networks, Support vector machines

ARTICLE INFO: Received 29 May 2014, Accepted 13 June 2014.

1. 서론

*Corresponding author is with the Division of Service Management, Mokwon University, 88 Doanbuk-ro Seo-gu Daejeon, 302-729, KOREA.

E-mail address: bookang@mokwon.ac.kr

마케팅 패러다임이 고객관계관리를 중시하는 방향으로 변화하고 있다. 고객관계관리의 중추적인 프로세스 중 하나가 고객관계유지이다. 고객이탈을

을 5%만 개선시켜도 기업의 수익률은 거의 100% 가까이 증가하는 것으로 알려져 있다[1]. 고객 이탈율을 개선하기 위해 선행해야 할 활동은 이탈고객의 예측이다. 사전에 이탈고객을 예측할 수 있다면 이에 대한 대비를 하거나 더 나아가 이탈 요인을 제거함으로써 이탈고객에 대한 기업의 손실을 최소화 할 수 있을 것이다.

이탈고객 데이터는 이탈고객의 사례 수가 유지고객의 사례 수에 비해 현저하게 작은 불균형 특성을 가진다. 이러한 상황하의 성능분석에서는 이탈 및 유지고객에 대한 예측의 정확성과 더불어 적은 클래스에 해당하는 이탈고객의 정확한 분류가 다수 클래스에 해당하는 유지고객의 정확한 분류에 비하여 큰 가치를 갖는다[2]. 클래스 불균형을 다루기 위해 널리 사용하는 방법 중의 하나는 오버샘플링 기법이다[2]. 오버샘플링은 훈련 데이터 집합이 똑같은 수의 긍정사례와 부정사례를 가질 때까지 긍정적 사례들을 복제한다. 클래스 불균형을 해결하기 위해 하이브리드 데이터마이닝 기법 [3]이나 여러 분류기법들의 예측을 종합함으로써 분류 정확성을 향상시키는 하이브리드 앙상블 기법도 사용된다[2].

이탈고객 예측 모형의 기본 분류기로는 데이터 마이닝의 여러 기법 중에서도 비교적 예측력이 좋은 것으로 나타난 의사결정나무기법, 인공신경망기법, SVM(Support Vector Machine) 기법 등이 많이 활용되고 있다[3].

이 연구에서도 불균형 특성을 갖는 이탈고객의 분류 예측을 위해 SVM 모형, 의사결정나무 모형, 인공신경망 모형을 사용한다. 클래스 불균형 문제를 다루기 위해 오버샘플링 기법도 적용하고, 각 기법의 예측결과에 가중치를 곱하여 종합하는 하이브리드 기법도 제시한다. 단일 분류기의 가중치를 구하기 위해 사례의 실제값과 예측값의 오차를 최소화하기 위한 비선형모형을 제시하고 이의 해

를 구함으로써 가중치를 찾는 방안도 제시한다.

2. 관련 연구

2.1 클래스 불균형 문제

균형이 맞지 않는 클래스 분포를 갖는 데이터 집합은 실제 응용에서 많이 나타나며, 기업 이탈고객의 경우에도 유지고객의 수에 비해 그 수가 현저하게 적은 불균형 특성을 갖는다. 예를 들면, 이탈고객의 수가 15%에서 20%라면 유지고객의 수는 80%에서 85% 정도를 차지한다. 이러한 응용에서는 소수 클래스의 정확한 분류가 다수 클래스의 정확한 분류에 비해 큰 가치를 가지며[2], 분류기의 예측 성능을 개선하는 데 있어 정확성의 향상 만큼이나 소수 클래스의 예측 성능 향상이 중요하다고 할 수 있다.

소수 클래스의 정확한 분류가 더 중요한 클래스 불균형 문제를 해결하기 위한 다양한 방법이 연구되어 왔다[4]. 샘플링은 클래스 불균형 문제를 다루기 위해 널리 사용되는 또 다른 방법이다 [2]. 샘플링의 아이디어는 사례들의 분포를 변경하여 소수 클래스가 훈련 집합에서 더 잘 표현되도록 하는 것이다. 샘플링 기법에는 다수 클래스를 적게 추출하는 언더샘플링[5], 소수 클래스를 중복해서 추출하는 오버샘플링[6] 기법 등이 있다. SVM으로 클래스 불균형을 해결하고자 하는 연구도 있다. [3]은 SVM, 인공신경망, 의사결정나무의 하이브리드 기법으로 예측성능을 높이는 방안을 제시하였다.

또한, 여러 분류기의 예측 결과를 종합하여 투표하는 하이브리드 앙상블 기법도 클래스 불균형 문제를 해결하는 좋은 대안이 됨을 보였다[2].

2.2 분류 모형

[3]은 클래스 불균형 특성을 갖는 이탈고객의 예측을 위해 데이터마이닝의 여러 분류모형 중에서도 의사결정나무, 인공신경망, SVM의 하이브리드 기법을 제안하였다.

의사결정나무 모형의 장점은 모형을 통해 나온 결과를 비교적 쉽게 이해할 수 있으며, 변수간의 교호관계를 잘 나타내며, 변수의 종류에 관계없이 사용할 수 있고, 계산속도가 빠르며, 대형자료처리에 용이한 점을 들 수 있다[7]. 대표적인 의사결정나무 기법중 하나는 기계학습 분야에서 활발하게 사용되고 있는 C4.5[8]이다. C4.5는 하향식 반복적 분할정복 원리를 이용하여 의사결정나무를 생성한다. 훈련데이터는 의사결정나무 뿌리에서부터 시작하여 분할된다. 분할을 위한 속성선택 기준으로는 엔트로피 척도가 사용되며, 정보이득 값이 크도록 학습 자료가 분할을 하게 된다. 의사결정나무를 생성한 후에는 과도적합 문제를 피하기 위해 나무 가지치기 작업을 하게 되며, 이후 최종 의사결정나무를 생성한다[8][9].

인공신경망은 복잡한 구조를 가진 자료의 분류 및 예측에 사용되는 비선형모형 중 하나이다[7]. 인공신경망은 입력층, 은닉층, 출력층 3계층으로 구성된다. 신경망의 역전파 알고리즘은 훈련자료의 속성 간 관계나 패턴을 학습을 통해 찾아낸다. 신경망 학습을 위해서는 은닉층의 은닉노드 수, 학습률, 모멘텀 등의 파라미터를 조정해 가면서 예측 성능이 가장 우수한 파라미터를 찾아 나간다.

SVM 모형은 지지도 벡터(Support Vector)라 불리는 학습사례의 부분집합을 이용하여 종속변수를 가장 잘 분류하는 의사결정 경계인 최대마진 초평면(Hyperplane)을 표현하는 기법이다[2]. 본래의 좌표공간 x 에 있는 데이터를 선형 의사결정 경계를 사용할 수 있는 새로운 좌표공간 $\phi(x)$ 로 변환시킴으로서 초평면을 찾게 된다. 비선형 SVM에서 변

환함수로 다항식 커널함수, RBF 커널함수, 탄젠트 커널함수 등이 사용되며, 고차원의 데이터에도 잘 작동되는 특징이 있다 [2][9].

3. 하이브리드 데이터마이닝 가중치모형

다음 <그림 1>은 불균형 데이터 특성을 가진 이탈고객 예측을 위한 하이브리드 데이터마이닝 가중치 모형이다.

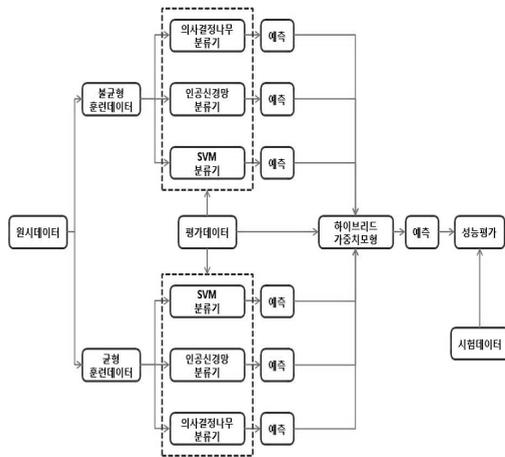


그림 1. 하이브리드 데이터마이닝 가중치 모형
Figure 1. Framework for hybrid data mining weighting model

3.1 데이터 구성

본 연구에서 사용하는 데이터는 UCI Machine Learning Repository[10]에서 제공하는 churn 데이터 집합이다. 이 데이터는 이동통신 고객의 유지(retention)와 이탈(churn) 여부를 나타내는 1개의 클래스 속성과 고객과 관련한 기타 속성 20개로 구성되어 있으며, 5,000개의 사례를 가지고 있다. 전체 사례 중 유지고객은 4,293명(약 86%)이고 이탈고객은 707명(약 14%)으로 이탈고객의 수가 현저하게 적은 클래스 불균형 데이터 집합이다.

전체 데이터는 학습데이터(4,000개)와 시험데이터(1,000개)로 분리하였다. 시험데이터는 전체 데이터를 무작위 샘플링을 하여 유지 사례 수를 850개(85%), 이탈 사례 수를 150개(15%)로 구성하였다. 시험데이터는 제안된 기법의 성능을 평가하기 위해 사용된다.

학습데이터는 다시 훈련데이터 (3,000개)와 검증데이터(1,000개)로 분리하였다. 검증데이터는 학습데이터에서 무작위 샘플링을 통해 유지 사례 수가 850개(85%), 이탈 사례 수는 150개(15%)로 구성하였다. 검증데이터는 훈련데이터로 만들어진 분류기들의 최적 파라미터와 이 연구에서 제안하고 있는 하이브리드 가중치 기법의 최적 가중치를 학습하기 위해 사용된다.

훈련데이터는 2,593개의 유지 사례와 407개의 이탈 사례를 가진 불균형 데이터이다. 불균형 훈련데이터는 추가적인 변환없이 3,000개의 훈련데이터가 사용된다. 균형 데이터는 유지와 이탈의 사례수가 동일하게 구성된 데이터로 이 연구에서는 407개의 이탈 사례를 반복 재생하는 오버샘플링을 실시하여 2,593개의 유지 사례와 2,593개의 이탈 사례를 갖는 5,186개의 균형 훈련데이터를 생성하였다.

3.2 분류기 학습

불균형 훈련데이터 3,000개를 이용하여 의사결정나무 분류기, 인공신경망 분류기, SVM 분류기를 학습한다. 10겹 상호검증시험에 의하여 각 분류모델을 생성한 후 검증데이터에 적용하여 예측성능을 검증받고, 우수한 성능이 나타나는 파라미터를 설정한다.

균형 훈련데이터 5,186개에 대해서도 분류기 학습 및 최적 파라미터 찾는 과정을 실행한다.

3.3 하이브리드 데이터마이닝 가중치기법

단일 분류기의 예측 결과는 각 분류기의 가중치와 곱해져서 계산된다. i번째 사례의 j번째 분류기의 예측 결과를 C_{ij} 라 하고, j번째 분류기의 가중치를 w_j 라 두면, m개 분류기의 예측 결과를 종합한 결과는 식(1)과 같다.

$$Y_i = \sum_{j=1}^m w_j C_{ij} \tag{1}$$

가중치는 사례의 실제값과 계산값의 예측 에러(오차)의 총합이 최소화되도록 구해진다. 예측 에러를 측정하는 기준으로는 식(2)과 같은 MSE(Mean Square forecast Error)가 많이 사용된다[11]. 식(2)에서 X_i 는 사례의 실제값을 Y_i 는 단일 분류기의 예측값의 가중평균을 나타내고, n은 사례 수를 의미한다.

$$MSE = \sum_{i=1}^n (X_i - Y_i)^2 / n \tag{2}$$

하이브리드 데이터마이닝 가중치 기법에서는 MSE를 최소화하는 분류기의 가중치를 찾는다. 이는 <그림 2>의 비선형모형으로 표현할 수 있다.

$$Min z = \sum_{i=1}^n \left(X_i - \sum_{j=1}^m w_j C_{ij} \right)^2 / n$$

s. t.

$$w_j \geq 0, \sum_{j=1}^m w_j = 1, j = 1, \dots, m$$

그림 2. 가중치를 구하는 비선형모형
Figure 2. Nonlinear model for computing weights

<그림 2>의 비선형문제의 해 w_j 가 분류기 j 의 가중치가 된다. 하이브리드 가중치 모형에서 이탈고객의 예측은 클래스의 유지고객(예: 0), 이탈고객(예: 1)의 중간값(0.5)을 경계로 유지와 이탈을 예측한다.

3.4 예측 성능 평가

예측 분류기의 성능 평가를 위해 주로 사용되는 척도는 정확도이다. 정확도는 전체 사례에서 클래스를 적합하게 분류한 비율을 나타낸다. 불균형 데이터의 경우에는 분류기의 예측에 있어 소수 클래스를 무시하는 경향이 있다[2]. 이 경우 소수 클래스의 정확한 분류가 다수 클래스의 정확한 분류에 비해 더 큰 가치를 갖는다[2]. 다수 클래스(유지)의 정확한 분류를 민감도(sensitivity)라 하고, 소수 클래스(이탈)의 정확한 분류를 특이도(specificity)라 할 때 불균형 데이터의 분류성능 평가를 위해서 클래스(유지 및 이탈)의 정확한 분류를 나타내는 정확도와 더불어 특이도를 성능 척도로 사용하기도 한다[2][3].

리콜(recall)과 정밀도(precision)는 어떤 특정 클래스(클래스 불균형 문제의 소수 클래스에 해당)의 성공적인 검출이 다른 클래스들의 분류에 비하여 훨씬 중요한 응용에서 널리 사용되는 두 가지 측정기준이다. 정밀도와 리콜 수치를 모두 최대화시키는 모델을 구축하는 것은 분류 알고리즘의 중요도전과제로, 정밀도와 리콜은 식(3)의 F1 기준으로 요약할 수 있다[2].

$$F1 = \frac{2rp}{r+p} = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (3)$$

TP는 분류모델에 의해 올바르게 예측된 소수 클래스 사례 수에 해당하고, FP는 분류모델에 의해

소수클래스로 잘못 예측된 다수 클래스 사례 수이고, FN은 분류모델에 의하여 다수클래스로 잘못 예측된 소수클래스 사례 수를 나타낸다. F1 측정치가 높다는 것은 정밀도와 리콜 모두가 상당히 크다는 것을 보장한다[2].

본 연구에서는 가중치 투표 모형의 예측 성능 평가를 위해 F1 척도를 이용한다.

4. 실험 및 결과 분석

4.1 데이터 학습

본 연구에서는 공개 데이터마이닝 도구인 WEKA[12]를 사용하여, 이동통신 회사의 이탈고객 데이터인 UCI churn 데이터 집합을 이용하여 실험하였다. 신경망모형은 WEKA의 MultilayerPerceptron을, 의사결정나무모형으로 WEKA의 J48을, SVM모형은 LibSVM[13]을 WEKA에 연결하여 사용하였다.

WEKA에서 균형 및 불균형 훈련데이터에 대해 의사결정나무(DT), 인공신경망(ANN), SVM 분류기를 학습하였다.

훈련데이터를 이용한 단일분류기의 학습 시에 10겹 상호검증시험을 하였고 검증데이터에 우수한 예측성능을 가지는 파라미터를 찾는 실험을 반복하였다. 실험결과 최종으로 의사결정나무의 파라미터는 WEKA의 기본 값을 사용하고, 인공신경망은 은닉노드 50, 학습률 0.3, 모멘텀 0.2로, SVM의 경우 균형 데이터에는 poly커널, gamma 1.0, degree 3.0, coef 1.0을, 불균형 데이터에는 poly 커널, gamma 1.0, degree 2.0, coef 3.0을 사용하였다.

하이브리드 가중치 기법에서 사용할 가중치를 구하기 위해 가중치 비선형모형을 구성하고 모형의 해는 엑셀의 해 찾기 기능[11]을 사용하여 최적해를 구하였다.

4.2 결과 분석

균형 훈련데이터(5,186) 및 불균형 훈련데이터(3,000)로 학습한 각 분류기의 검증데이터에 대한 결과는 <표 1>과 같다.

표 1. 검증데이터에 대한 분류기 성능
Table 1. Prediction results of classifiers to validation data

훈련데이터	분류기	민감도	특이도	정확도	F1
균형	DT	0.931	0.793	0.91	0.726
	ANN	0.948	0.613	0.898	0.643
	SVM	0.9	0.327	0.814	0.345
불균형	DT	0.991	0.6	0.932	0.726
	ANN	0.972	0.427	0.89	0.538
	SVM	0.968	0.427	0.887	0.531

단일 분류기를 검증데이터에 적용하여 성능평가를 한 결과 의사결정나무, 신경망, SVM의 순으로 성능이 좋게 나타났다. 그러나 전반적으로 F1 측정치가 적게 나타나고 있어 클래스 불균형 문제에 대한 취약점을 가지고 있음을 알 수 있다.

6개 분류기의 검증데이터 사례의 실제값과 가중치 기법에 의해 계산된 값의 MSE 최소값을 갖는 가중치는 <그림 2>의 비선형모형을 이용하여 구하였다. 엑셀의 해찾기 기능을 사용하여 가중치를 구한 결과는 다음 <표 2>와 같았다.

표 2. 단일분류기에 대한 가중치
Table 2. Weights of single classifiers

분류기	균형			불균형		
	DT	ANN	SVM	DT	ANN	SVM
가중치	0.242	0.189	0	0.401	0.089	0.079

하이브리드 가중치 기법을 시험데이터에 적용하여 예측 성능평가를 한 결과는 <표 3>에 제시되어 있다.

표 3. 시험데이터에 대한 분류기 성능
Table 3. Prediction results of classifiers to test data

분류기	민감도	특이도	정확도	F1	
균형	DT	0.932	0.787	0.91	0.724
	ANN	0.931	0.653	0.889	0.638
	SVM	0.896	0.42	0.825	0.419
불균형	DT	0.994	0.653	0.943	0.775
	ANN	0.973	0.54	0.908	0.638
	SVM	0.958	0.493	0.888	0.569
하이브리드 가중치 기법	0.988	0.753	0.953	0.828	

실험결과 하이브리드 가중치 기법의 F1 측정치가 0.828로 단일 분류기의 가장 높은 F1 측정치 0.775에 비해 상당히 높음을 알 수 있다. F1 측정치가 높다는 것은 리콜과 정밀도 모두가 크다는 것을 보장한다[2]. 리콜과 정밀도는 특정 클래스의 검출이 다른 클래스의 분류에 비해 훨씬 중요한 응용에서 널리 사용되는 측정기준으로 본 연구에서 제안한 하이브리드 데이터마이닝 가중치 기법이 효과적임을 알 수 있다.

5. 결 론

기업의 고객유지율을 5%만 개선시켜도 기업 성과가 거의 100% 정도 높아지는 것으로 알려져 있다. 따라서 이탈고객의 방지는 고객관리에 있어 중추적인 프로세스라고 할 수 있다. 고객이탈관리를 방지하기 위해서는 고객이탈의 정확한 예측이 선행되어야 한다.

일반적으로 기업의 이탈고객 데이터는 이탈 사례 수가 유지 사례 수보다 현저하게 적은 클래스 불균형 특성을 가진다. 클래스 불균형 데이터의 예측 성능을 향상시키기 위해 데이터마이닝 기법을 활용한 다양한 방안들이 연구되었다. 단일 분류기를 결합한 하이브리드 기법이나 불균형 데이터의 소수 클래스를 오버샘플링하여 균형데이터로 학습하는 방법 등이 제안되었다.

본 연구에서는 불균형 데이터의 예측에 많이 사용되는 의사결정나무, 신경망, SVM 기법의 하이브리드 가중치 기법을 제안하였다. 제안된 기법은 균형 훈련데이터와 불균형 훈련데이터를 학습하여 예측한 결과와 단일 분류기의 가중치를 곱하여 가중평균을 구하고 클래스를 예측하였다. 단일 분류기의 가중치는 사례의 실제값과 하이브리드 기법의 계산값 사이의 오차를 최소화할 수 있는 비선형모형을 구성한 후 해를 찾아 구하였다.

제안된 기법을 이동통신 회사의 이탈고객 데이터인 UCI churn 데이터 집합에 적용하여 실험하였다. 실험결과 소수 클래스의 검출 성능을 측정하는데 많이 사용하는 F1 측정치가 상당히 높아짐을 알 수 있었다. F1측정치가 높다는 것은 정밀도와 리콜의 수치가 높은 것을 의미하며, 제안된 기법이 클래스 불균형 특성을 갖는 이탈고객 예측에 효과적임을 알 수 있었다.

본 연구에서 사용한 데이터마이닝 기법은 의사결정나무, 인공신경망, SVM 기법으로, 향후 다른 데이터마이닝 기법까지 포함하여 예측 성능을 높이기 위한 연구로 확장할 필요가 있으며, 더욱 많은 사례와 응용 분야에 대한 추가적인 실험을 통해 일반화 시킬 필요가 있다.

References

- [1] H.-S. Kim, Y.-G. Kim, and C.-Y. Park, *Customer relationship management: Principle and practical application*, SciTech, 2011.
- [2] H.-S. Yong, Y.-M. Na, J.-S. Park, H.-U. Seung, M.-S. Lee, S.-J. Lee, and R. Choi, *Introduction to data mining*, Infinitybooks, 2007.
- [3] J.-S. Lee, and J.-G. Kwon, *A hybrid SVM classifier for imbalanced data sets*, Journal of Intelligence and Information Systems, Vol. 19, No. 2, pp.125-140, 2013.
- [4] N. V. Chawha, N. Japkowicz, and A. Kolcz, *Editorial: Special issue on learning from imbalanced data sets*, SIGKDD Explorations, Vol. 6, No. 1, pp. 1-6, 2004.
- [5] M. Kubat, and S. Matwin, *Addressing the curse of imbalanced traing sets: One sided selection*, Proceedings of the 14th International Conference on Machine Learning, pp. 179-186, 1997.
- [6] X. Chen, B. Gerlach, and D. Casasent, *Prunung support vectors for imbalanced data classification*, Proceedings International Joint Conference on Neural Networks, pp. 1883-1888, 2005.
- [7] T.-R. Lee, J.-Y. Koo, H.-J. Park, K.-H. Lee, and D.-W. Choi, *Data mining*, KNOU Press, 2004.
- [8] J. R. Quinlan, *C4.5: Programs for machine learning*, Morgan Kaufmann, 1993.
- [9] B.-S. Kang, and J.-H. Cho, *Prediction performance improvement of KOSDAQ business bankruptcy using SVM preprocessor*, Journal of The Korea Knowledge Information Technology Society, Vol. 7, No. 4, pp. 21-27,2012.
- [10] K. Bache, and M. Lichman, *UCI machine learning repository* [<http://archive.ics.uci.edu/ml>], University of California, School of Information and Computer Science, 2013.
- [11] K.-H. Jeong, and C.-H. Back, *Management science with excel 5th Ed.*, BnmBooks, 2013.
- [12] I. H. Witten, E. Frank, and M. A. Hall, *Data mining: Practical machine learning tools and techniques 3rd Ed.*, Morgan Kaufmann, 2011.
- [13] C. C. Chang, and C. J. Lin, LIBSVM - A library for support vector machines [<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>], 2001.

하이브리드 데이터마이닝 기법을 이용한 이탈고객 예측

강부식

목원대학교 서비스경영학부

요 약

고객관계유지 활동은 고객관계관리의 가장 중추적인 프로세스로, 고객이탈율을 5%만 개선시켜도 기업의 수익률은 거의 100%가까이 증가하는 것으로 알려

져 있다. 고객이탈관리를 위해서는 이탈고객 예측활동이 선행되어야 한다. 이탈고객을 사전에 예측할 수 있다면 이에 대한 대비를 하거나 고객의 이탈요인을 사전 제거함으로써 고객이탈에 따른 손실을 최소화하는 것이 가능하다. 기업의 이탈고객 데이터는 이탈고객 수가 유지고객 수에 비해 사례가 현저하게 적은 불균형 데이터 특성을 가지고 있다. 이탈고객 예측 모형의 주요 이슈 중 하나는 예측력을 높이는 것이다. 본 연구에서는 불균형 특성을 가진 이탈고객에 대한 예측력을 높이는 하이브리드 데이터마이닝 기법에 대해 살펴본다. 이탈고객 예측에 많이 사용되며, 예측력이 비교적 높은 것으로 알려진 의사결정나무모형, 신경망 모형, SVM모형을 이용한 하이브리드 데이터마이닝 가중치 기법을 제안한다. 제안된 기법의 성능 평가를 위해 UCI churn 데이터 집합에 적용하였다. 실험결과 제안된 하이브리드 데이터마이닝 가중치기법이 불균형 데이터 특성을 가진 이탈고객을 예측하는 데 있어 효과적임을 알 수 있었다.

information systems, customer relationship management, data mining, and service quality management. He is a life member of the KKITS.

E-mail address: bookang@mokwon.ac.kr

감사의 글

이 논문은 2013년도 목원대학교 연구년 지원에 의하여 연구되었음.



Boo-Sik Kang received the bachelor's degree in the Department of Industrial Engineering from the Kyung Hee University in 1985. He received the MS degree and the Ph.D. degree in the Department of Industrial Engineering from KAIST in 1989 and 2000, respectively. From 1989 to 2001, he was a researcher at Korea Telecom. He has been a professor in the Division of Service Management at Mokwon University since 2001. His current research interests include intelligent