ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org

J Intell Inform Syst 2015 March: 21(1): 29~45 http://dx.doi.org/10.13088/jiis.2015.21.1.29

불균형 데이터 환경에서 변수가중치를 적용한 사례기반추론 기반의 고객반응 예측*

김은미
부산대학교 경영대학 시간강사
(keunmi100@pusan.ac.kr)

홍태호
부산대학교 경영대학 교수
(hongth@pusan.ac.kr)

고객반응 예측모형은 마케팅 프로모션을 제공할 목표고객을 효과적으로 선정할 수 있도록 하여 프로모션의 효과를 극대화 할 수 있도록 해준다. 오늘날과 같은 빅데이터 환경에서는 데이터 마이닝 기법을 적용하여 고객반응 예측모형을 구축하고 있으며 본 연구에서는 사례기반추론 기반의 고객반응 예측모형을 제시하였다. 일반적으로 사례기반추론 기반의 예측모형은 타 인공지능기법에 비해 성과가 낮다고 알려져 있으나 입력변수의 중요도에 따라 가중치를 상이하게 적용함으로써 예측성과를 향상시킬 수 있다. 본 연구에서는 프로모션에 대한 고객의 반응여부에 영향을 미치는 중요도에 따라 입력변수의 가중치를 산출하여 적용하였으며 동일한 가중치를 적용한 예측모형과의 성과를 비교하였다. 목욕세제 판매데이터를 사용하여 고객반응 예측모형을 개발하고 로짓모형의 계수를 적용하여 입력변수의 중요도에 따라 가중치를 산출하였다. 실증분석 결과 각 변수의 중요도에 기반하여 가중치를 적용한 예측모형이 동일한 가중치를 적용한 예측모형보다 높은 예측성과를 보여주었다. 또한 고객 반응예측 모형과 같이 실생활의 분류문제에서는 두 범주에 속하는 데이터의 수가 현격한 차이를 보이는 불균형 데이터가 대부분이다. 이러한 데이터의 불균형 문제는 기계학습 알고리즘의 성능을 저하시키는 요인으로 작용하며 본 연구에서 제안한 Weighted CBR이 불균형 환경에서도 안정적으로 적용할 수 있는지 검증하였다. 전체데이터에서 100개의 데이터를 무작위로 추출한 불균형 환경에서 100번 반복하여 예측성과를 비교해 본 결과 본 연구에서 제안한 Weighted CBR은 불균형 환경에서도 일관된 우수한 성과를 보여주었다.

주제어 : 고객반응, 예측모형, 사례기반추론, 변수 가중치, 불균형 데이터

논문접수일 : 2014년 11월 7일 논문수정일 : 2014년 12월 23일 게재확정일 : 2014년 12월 28일
투고유형 : 국문급행 교신저자 : 홍태호

1. 서론

기업은 신규고객의 획득 및 기존고객을 유지하기 위해 다양한 프로모션을 제공하고 있다. 모든 고객에게 무차별적으로 제공하는 프로모션은 불필요한 마케팅 비용지출은 물론이고 고객과의 관계도 악화시킬 수 있기 때문에 프로모션을 제공할 목표고객의 선정은 매우 중요한 업무이다. 고객반응 예측모형은 프로모션에 반응할 것 같은 고객을 예측하고, 반응할 가능성이 높은 고객을 선별하여 프로모션을 제공할 목표고객을 효과적으로 선정할 수 있도록 하여 프로모션의 효과를 극대화 할 수 있도록 해준다. 반응고객을 선별하기 위해 기업은 고객의 방대한 데이터를 다양한 데이터 마이닝 기법에 적용하여 예측모형을 구축해 왔으나 프로모션에 대한 고객의 반

* 이 논문은 2012학년도 부산대학교 박사후연수과정 지원사업에 의하여 연구되었음.

응률은 10%미만으로 대부분 비반응 고객으로 나타나는 불균형 데이터이다. 불균형 데이터는 두 범주의 비율이 현저히 차이가 나는 경우로 기계학습의 성능을 저하시키는 요인으로 작용하기 때문에 두 범주의 비율을 비슷하게 하기 위해 샘플링 방법이나 오분류 비용을 통해 전체 예측모형의 성과를 향상시키고자 하고 있다(Barandela et al., 2003).

데이터 마이닝 기법 중 사례기반추론은 적용이 쉽고 간단하며 모형의 갱신이 실시간으로 이루어질 수 있다는 장점으로 인해 학계와 실무에서 주목받고 있다(Ahn et al., 2005). 사례기반추론(Case-Based Reasoning; CBR)은 새로운 사례와 가장 비슷한 사례를 통해 문제를 해결하기 때문에 고객반응 예측모형에 사례기반추론을 적용함으로써 기존고객의 주요특성을 신규고객에게도 효과적으로 반영할 수 있다는 장점을 가지고 있다. 사례기반추론은 일반적으로 타 인공지능기법에 비해 예측성과가 상대적으로 낮게 나타난다는 문제점이 제기되고 있으나 입력변수의 중요도에 따라 가중치에 차이를 두어 예측성과를 향상시킬 수 있는 것으로 알려져 있다(Roh et al., 2005). 모든 변수에 동일한 가중치를 적용하게 되면 중요한 변수를 예측모형에 보다 많이 반영하지 못하며 중요하지 않은 변수를 중요한 변수와 같이 생각할 수 있기 때문에 모형 구축 시 사용되는 변수의 중요도에 따라 가중치를 달리 적용하여 중요한 변수를 모형에 보다 많이 반영하여 예측성과를 향상시킬 수 있다.

사례기반추론에서 입력변수의 가중치를 선정하기 위해 전문가 집단의 의견을 반영하거나(Park and Han, 2002; Hong and Cho, 2009), 유전자알고리즘을 통해 가중치를 적용한(Shin and Han, 1999; Chiu, 2002) 연구가 이루어져 왔다. Park and Han(2002)은 기업의 부도예측을 위한 모형을 구축하기 위해 사례기반추론을 적용하였으며 AHP를 적용하여 전문가 집단의 의견을 입력변수의 가중치로 적용하였다. Hong and Cho (2009)은 기업의 부도예측을 위해 사례기반추론 모형을 제시하였으며 입력변수의 가중치를 결정하기 위해 AHP를 통한 전문가 집단의 의견을 적용하였다. Shin and Han(1999)는 회사채 평가를 위한 사례기반추론의 입력변수 가중치를 선정하기 위해 유전자알고리즘을 적용하였으며, Chiu et al.(2003)은 보험회사의 고객관계관리를 위해 사례기반추론의 입력변수에 가중치를 선정하기 위해 유전자알고리즘을 활용하였다. 그러나 전문가에 의해 부여되는 입력변수의 가중치는 아무리 전문가라 하더라도 가중치를 효과적으로 부여하는 것은 쉽지 않으며 또한 최적화를 위해 많이 적용되는 유전자알고리즘은 복잡한 계산과 시간이 필요하다.

따라서 본 연구에서는 사례기반추론을 적용하여 고객반응 예측모형을 구축하며, 로짓모형의 계수를 적용하여 변수에 대한 중요도를 계산하고 입력변수의 가중치로 적용하고자 한다. 로짓모형에서 나타나는 변수의 계수는 종속변수에 얼마나 영향을 미치는지를 나타내기 때문에 변수의 중요도를 판단할 수 있으므로 입력변수의 가중치로 활용할 수 있다. 사례기반추론을 적용한 예측모형의 성과를 향상시키기 위해 본 연구에서는 입력변수에 동일한 가중치를 적용한 사례기반추론(Pure CBR) 모형과 입력변수의 중요도에 따라 상이한 가중치를 적용한 사례기반추론(Weighted CBR) 모형을 제시하고 고객반응 예측모형의 예측성과를 비교하였다.

또한 실생활에서는 기업의 프로모션 또는 다양한 마케팅 프로그램에 대한 반응고객의 비율

이 비반응 고객에 비해 매우 적으므로 이와 같은 불균형 데이터에서 반응예측성과가 일정한 모형을 필요로 한다. 따라서 본 연구에서 제안한 Weighted CBR에 기반한 고객반응예측 모형을 불균형 데이터 환경에서도 성과가 기대치보다 높은 상태를 유지하며 안정적인지를 실증적으로 검증하여 실무에서도 활용가능성이 높은 고객반응 예측모형을 제안하였다.

2. 이론적 배경

2.1. 고객반응 예측

기업은 다양한 마케팅 프로모션을 통해 고객의 구매를 촉진하고자 하며 최근에는 새로운 제품은 물론이고 기존제품에도 마케팅 프로모션은 필수사항으로 인식되고 있다(Chan et al., 2010). 기업은 프로모션에 대한 고객의 반응을 미리 예측하여 프로모션을 제공할 목표고객을 효과적으로 선정하고자 한다. 잘 선정한 목표고객은 기업의 수익을 증가시킬 수 있지만, 목표고객을 잘못 선정하는 것은 마케팅 비용의 증가는 물론이고 고객과의 관계도 악화시킬 수 있기 때문이다(Cönül et al., 2000). 기업은 고객 데이터베이스에 저장되어 있는 고객정보, 구매정보, 상품정보 등을 기반으로 프로모션에 대한 고객의 반응확률을 미리 예측하고 반응확률이 높은 고객을 대상으로 프로모션을 제공한다(Kim et al., 2008). 프로모션에 대한 고객의 반응을 예측하기 위해 전통적으로 로짓이나 판별분석과 같은 통계적인 기법이 적용되었으며, 방대하고 비선형적인 복잡한 데이터의 분석을 위해 의사결정나무, 인공신경망, 로짓, SVM과 같은 데이터 마이닝 기법들도 많이 적용되고 있다(Cheung et al., 2003; Shin and Cho, 2006).

Ahn and Kim(2008)은 온라인 쇼핑몰 고객의 구매예측과 주가예측을 위해 로지스틱 회귀분석, 인공신경망, 사례기반추론 등의 단일모형을 개발하였으며 사례기반추론에 사용된 k-최근접 이웃법의 최적의 이웃 수를 찾기 위해 유전자알고리즘을 적용하여 예측모형의 성과를 향상시키고자 하였다. Hong and Park(2009)는 미국의 통신판매 데이터를 대상으로 로짓, 인공신경망, SVM 등의 단일모형을 제시하였으며 이들 단일모형을 사례기반추론으로 통합하여 예측모형의 성과를 향상시키고자 하였다. Cui et al.(2006)은 다이렉트 마케팅 데이터를 사용하여 인공신경망, CART, 로짓모형, 베이지안 네트워크 등의 다양한 모형을 적용하여 고객반응 예측모형을 구축하였다. Kim and Street(2004)은 CoIL Challenge 2000 데이터를 통해 예측모형을 구축하였으며 유전자알고리즘과 인공신경망을 적용하였다. Chiu(2002)는 보험회사의 고객데이터를 사용하여 고객의 반응예측 모형을 구축하였으며 사례기반추론에 유전자알고리즘을 통합하였다. Coenen et al.(2000)은 통신판매 고객데이터를 사용하여 반응모형을 구축하였으며 예측성과를 향상시키기 위해 C5.0 알고리즘과 사례기반추론을 통합하였다. 반응고객과 비반응 고객을 분류하기 위해 C5.0 알고리즘을 적용시켰으며 분류된 고객의 특성을 기반으로 재정렬하여 메일을 보낼 목표고객을 선정하기 위해 사례기반추론을 적용하였다.

2.2. 사례기반추론

사례기반추론은 과거에 적용되었던 사례와 그 결과들을 참조하여 새로운 사례에 대한 결과값

을 예측하는 것으로, 새로운 사례와 가장 비슷한 과거의 사례를 일부 추출하여 추출한 사례의 특정지식을 통해 문제를 해결하는 방식이다. 사례기반추론은 상대적으로 적은 데이터를 가진 분야에서도 적용이 가능하며 복잡하거나 덜 구조화된 분야에서도 쉽게 적용할 수 있다. 또한 사례기반추론은 새로운 사례가 데이터베이스에 추가되더라도 특별한 학습과정을 거치지 않고 모형의 갱신이 즉각적으로 이루어질 수 있다는 장점이 있다(Shin and Han, 1999). 일반적으로 사례기반추론은 5개의 단계로 이루어진다(Bradley, 1994). 새로운 사례가 입력사례로 주어지면 사례베이스를 검색하여 새로운 사례와 가장 유사한 과거의 사례를 추출(retrieve)하게 된다. 추출한 사례의 결과들을 참조하여 새로운 사례에 대한 추천결과를 적용(adaptation)하고, 추천결과를 검증(validation)하게 된다. 검증단계를 통해 제시된 새로운 결과는 새로운 사례로 저장되어 미래의 문제해결을 위해 재사용 될 수 있도록 사례베이스의 갱신(update)이 이루어진다.

새로운 입력사례와 비슷한 사례를 사례베이스에서 추출하기 위해 유사도를 측정하게 되며 유클리드(Euclidean) 거리가 많이 사용되고 있다. 유클리드 거리는 식(1)과 같이 나타낼 수 있다. 유사도 기준에 의해 입력사례와 가장 가까운 기존 사례를 찾아내는 방법으로는 k-최근접이웃법이 가장 널리 활용된다(Jarmulak et al., 2000; Chiu, 2002).

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} \quad (1)$$

사례기반추론 기법에서는 유사도를 측정하기 위한 속성선택과 함께 각 속성들의 가중치를 어떻게 부여할 것인지가 중요하다. 목표변수와 관련성이 적은 속성들이 관련성이 높은 속성들과 같은 중요도로 사용된다면 예측성과에 부정적인 영향을 줄 수 있기 때문에 관련성이 적은 속성에는 낮은 가중치를 적용하고 관련성이 많은 속성에는 높은 가중치를 적용하여 사례기반추론 모형의 예측성과를 향상시킬 수 있다(Kolodner, 1993). 각 변수의 중요도에 따라 가중치를 적용한 유클리드 거리는 식(2)와 같이 계산된다.

$$D_{weight} = \sqrt{w_1(x_1 - y_1)^2 + w_2(x_2 - y_2)^2 + \cdots + w_n(x_n - y_n)^2} \quad (2)$$

Park and Han(2002)는 AHP를 적용하여 전문가 집단의 의견을 통해 입력변수의 가중치를 결정하였으며, Hong and Cho(2009)은 기업부도예측을 위한 사례기반추론 모형을 제시하였으며 입력변수의 가중치를 위해 AHP를 통한 전문가 집단의 의견을 통해 가중치를 결정하였다. 하지만 전문가라 하더라도 정확한 가중치를 결정하는 것은 쉬운 일이 아니다. 이에 대한 대안으로 유전자 알고리즘을 통해 입력변수의 가중치를 적용한 연구가 다음과 같이 많이 이루어졌다. Shin and Han(1999)은 회사채평가에 사례기반추론을 적용하였으며 입력변수의 가중치를 위해 유전자알고리즘을 적용하였으며 Chiu et al. (2003)은 공장의 만기일 할당문제에서 입력변수의 가중치를 결정하기 위해 유전자알고리즘을 적용하였다. Chiu(2002)은 보험회사의 고객관계관리에서 입력변수의 가중치를 적용하기 위해 유전자알고리즘을 적용하였으며 Kim(2004)은 유전자알고리즘을 적용하여 모형구축을 위한 입력변수와 입력변수의 가중치를 선정하였다. 또한 Chung and Suh(2006)은 암환자의 진료비 예

측모형을 개발하기 위해 사례기반추론을 적용하였으며 Relief-F 기법과 인공신경망 기법에서 제공해 준 상대적 중요도의 평균값을 해당 속성의 가중치로 사용하였다.

2.3. 불균형 데이터 처리

불균형 데이터는 하나의 범주에 속하는 데이터의 수가 다른 범주에 속하는 데이터의 수와 현저한 차이를 보일 때를 의미하며 데이터 마이닝의 분류 및 예측문제에서 흔히 발생되는 문제 중 하나이다. 부도예측, 침입탐지, 사기적발, 이탈고객, 고객반응 등의 데이터에서 나타나는 불균형 문제는 기계학습 알고리즘의 성능을 저하시키는 요인으로 작용한다(Kang et al., 2004). 모형의 전체적인 오분류를 줄이기 위해 다수의 범주로 패턴분류를 많이 하게 되어 소수범주는 다수범주로 취급되기 쉽기 때문에 불균형 데이터에서는 지도학습을 통한 패턴 인식이 어렵다(Weiss and Provost, 2001). 또한 불균형 데이터에서 다수범주는 정확하게 분류하면서 소수범주는 무시하는 경향이 있다(Jo and Japkowicz, 2004). 따라서 다수범주와 소수범주가 최소한 패턴을 인식할 수 있는 수준의 비율을 유지하는 것이 좋으나 실제 데이터의 분포는 그렇지 않은 경우가 대부분이다.

이러한 불균형 문제를 해결하기 위한 대표적인 방법으로는 샘플링을 이용한 방법과 오분류를 조정하는 방법이 있다. 샘플링 방법에는 다수범주 집단에서 임의로 샘플링하여 소수범주와 균형을 이루게 하는 under sampling과 소수범주 집단을 반복적으로 복사하여 다수범주 집단과 균형을 이루게 하는 over sampling이 있다. Under sampling은 데이터의 잡음(noise)을 제거하여 예측성과를 향상시키는 장점이 있으나 데이터에 대한 정보손실에 대한 우려가 있으며 over sampling은 이상치가 선택되었을 경우 계속적인 확산의 우려가 존재한다(Liu et al., 2006). 또다른 불균형 데이터의 해결방법인 오분류 비용의 조정은 원래의 데이터를 그대로 유지하면서 오분류에 가중치를 주어 데이터의 불균형을 해결하는 방법이다. 오분류 비용을 이미 알고 있다는 가정하에 오분류에 패널티를 부과한다. 즉 소수범주의 오분류 비용을 높여서 소수범주의 특성을 더 잘 학습할 수 있도록 하여 불균형을 해결할 수 있다.

Ha et al.(2005)는 다이렉트 마케팅의 고객데이터를 적용하여 고객반응 예측모형을 구축하였으며 불균형 데이터를 해결하기 위해 구매액을 적용하였다. 구매액에 가중치를 두고 구매액을 재계산하여 구매액에 따라 고객을 재정렬하고 상위 20%의 고객을 선정하여 데이터의 불균형을 완화시켰다. 여전히 존재하는 불균형 문제는 랜덤으로 추출하여 균형데이터로 조정하고 모형을 구축하였다. Yu and Cho(2006)도 Ha et al.(2005)이 제안한 가중 구매액을 적용하여 균형데이터로 조정하였다. Lee and Cho(2007)는 반응 모델링에서 불균형 데이터를 완화시키기 위해 새로운 학습법으로 1-SVM과 LVQ-ND를 제안하고 반응률을 변화시켜가며 반응률에 따라 어느 모형의 성과가 좋은지를 비교하였다. Hwang and Cho(2007)는 클러스터링 기법을 적용한 데이터의 전처리를 통해 참조 셋의 수를 적게 하여 반응모델 구축 시 방대한 데이터 및 패턴에 대한 문제를 해결하고자 하였다. Jang et al.(2008)은 소수 범주에 대한 과대 표본 추출, 다수 범주에 대한 과소 표본 추출, 이상치 제거 후 과소 표본 추출, 오분류 비용 조정 등의 불균형 해소 기법

들을 유전자 알고리즘을 활용하여 결합적으로 활용하여 불균형 문제를 해결하고자 하였다. Lee and Kwon(2013)는 SVM, 인공신경망, 의사결정나무 기법 등으로 하이브리드 모델을 구축하여 다수범주의 예측성과는 유지하면서 불균형 데이터의 분류문제에서 흔히 나타나는 소수범주에 대한 예측성과를 향상시켰다.

3. 연구모형

본 연구에서는 사례기반추론을 적용하여 프로모션에 대한 고객반응 예측모형을 구축한다. 사례기반추론은 타 인공신경망 기법보다 상대적으로 예측성과가 낮다고 알려져 있으나 사례기반추론은 입력변수와 입력변수의 가중치에 따라 예측성과를 향상시킬 수 있다. 사례기반추론에 기반한 반응고객 예측모형의 성과를 높이기 위하여 사용된 변수의 가중치를 결정하는 방법은 앞의 이론연구에서처럼 전문가를 이용한 AHP 결과와 유전자알고리즘의 이용이 주로 사용되고 있으며, 본 연구에서는 사용 편의성이 높은 로짓모형을 이용하여 사례기반추론 모형의 가중치를 결정하도록 한다. 따라서 관련성이 적은 입력변수에는 낮은 가중치를 적용하고 관련성이 높은 입력변수에는 높은 가중치를 적용하여 예측성과를 향상시키고자 사례기반추론의 두 가지 모형을 제시한다. 하나는 입력변수에 동일한 가중치를 적용하여 예측모형을 구축한 사례기반추론(Pure CBR) 모형을 제시하였으며 두 번째로 입력변수의 중요도에 따라 상이한 가중치를 적용하여 예측모형을 구축한 사례기반추론(Weighted CBR) 모형을 제시하고 모형의 성과를 비교하였다. 또한 Weighted CBR의 예측성과와 인공신경망 모형, SVM 모형과도 예측성과를 비교하였다.

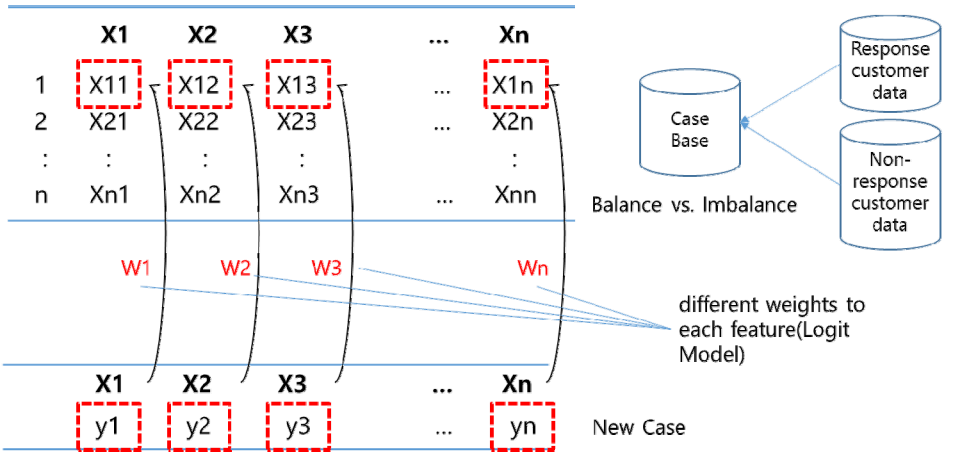
예측모형의 입력변수선정을 위해 로짓의 stepwise방법을 적용하였으며 프로모션에 대한 고객의 반응여부를 종속변수로 하여 입력변수를 선정하고, 선정된 입력변수에 가중치를 적용하기 위해 로짓의 계수를 적용하여 각 변수에 대한 가중치를 계산하였다. 로짓의 계수는 각 변수가 종속변수에 영향을 미치는 정도를 나타내기 때문에 입력변수와 종속변수와의 관련성을 파악할 수 있으며 이들을 사례기반추론의 입력변수에 대한 가중치로 적용하여 Weighted CBR 모형을 구축하였다. 가중치는 식(3)과 같이 전체 변수에 대한 로짓계수 및 변수의 중요도 합을 각 변수의 로짓계수 및 중요도로 나누어 각 변수에 대한 상대적인 가중치를 계산하였다.

$$Weight = \frac{|x_i|}{\sum_{i=1}^{n} |x_i|} \times 100 \qquad (3)$$

x_i : i th Logit coefficients and variable importance of input variables

로짓에서 선정된 변수는 종속변수에 음의 영향을 미칠 수도 있으므로 로짓계수의 절대값을 사용하여 가중치를 계산하였다. 가중치를 각 입력변수에 적용하여 변수들간의 유클리드 거리를 계산하여 가중치를 적용한 유클리드 거리를 통해 유사한 사례를 추출하고, k-최근접이웃법을 활용하여 사례기반추론을 적용하였다.

고객데이터에서 <Figure 1>과 같이 반응고객과 비반응 고객을 균형과 불균형으로 추출한 사례베이스를 새로운 사례에 적용할 수 있도록 하였으며 본 연구에서 제안한 Weighted CBR은 로짓모형을 적용하여 입력변수의 중요도에 따른



〈Figure 1〉 The proposed CBR model with the different weights

가중치를 산출하여 적용하였다.

사례기반추론을 불균형 데이터 환경에 적용하기 위해 반응고객과 비반응 고객으로 이루어져 있는 전체데이터에서 무작위로 100개의 데이터를 추출하여 Pure CBR과 Weighted CBR을 적용하였다. 사례기반추론 모형은 알고리즘의 계산속도 등을 비교하면 사용의 편의성이 매우 높은 기계학습임에도 불구하고 예측성과 측면에서는 비교적 인공신경망에는 못 미치는 것으로 알려져 왔다. 그러나 반응고객 예측모형을 적용할 때에는 비반응 고객의 데이터가 훨씬 많은 환경에서 모형을 구축하고 적용할 수 있다면 실무적으로 매우 유용할 것으로 기대된다. 따라서 본 연구에서 제안된 Weighted CBR에 기반한 고객반응 예측모형을 100개의 데이터를 100번 반복하여 추출하고 예측성과의 평균을 비교하였으며 균형 데이터 환경에서의 성과와 불균형 환경에서의 성과를 비교하였다. 중요 변수에 대한 가중치를 차별화하게 되면 불균형 데이터 환경에서도 유의한 유사 사례를 찾게 됨으로써 제안된 고객반응 예측모형의 성과를 높임과 동시에 불균형비에 따른 강건한(robust) 모형을 개발할 수 있게 된다.

4. 실험 및 결과분석

4.1. 데이터

본 연구에서는 소비재에 대한 고객의 구매행태를 추적하는 아시아의 시장조사기관에서 제공하고 있는 판매데이터를 사용하였다(Shmueli et al., 2007). 실험에 사용된 데이터는 600명의 고객에 대한 거래데이터와 고객데이터로 이루어져 있으며 총 45개의 변수로 구성되어 있다. 또한 이 데이터는 매년 갱신되는 인구통계학적 데이터, 보유중인 내구재를 통한 부유지수, 그리고 매달 갱신되는 제품목록 및 브랜드에 대한 구매데이터의 정보와 기업에서 제공하는 다양한 프로모션 활동에 대한 고객의 구매비율을 제시하고 있다. 기업은 묶음판매, 가격할인, 쿠폰제공, 사은품 제공 등과 같은 프로모션을 제공하였으

며 묶음판매에 의한 고객의 구매비율을 제시하였다. 본 연구에서는 프로모션에 대한 고객의 반응예측을 위하여 묶음판매 프로모션에 대한 고객의 반응여부를 종속변수로 하였다. 묶음판매 프로모션에 의해 구매가 발생한 고객을 반응고객으로 하였으며, 묶음판매 프로모션에 의한 구매가 발생하지 않은 고객은 비반응 고객으로 하여 총 600명의 고객 중 반응고객은 299명, 비반응 고객은 301명으로 분류되었다.

4.2. 고객반응 예측모형

본 연구에서는 프로모션에 대한 고객의 반응예측을 위해 먼저 로짓의 stepwise forward를 통해 변수를 선정하였다. 총 45개로 구성되어 있는 고객의 인구통계학적 정보와 구매정보 등의 데이터를 통해 <Table 1>과 같이 8개의 변수가 선정되었다.

로짓에 의해 선정된 변수는 TV의 시청여부(X1), 연속해서 구매한 브랜드의 건수(X2), 묶음판매가 아닌 다른 프로모션에 활동에 대한 구매비율(X3), 가격대가 저렴한 제품에 대한 구매비율(X4), 천연재료의 제품에 대한 구매비율(X5), 미백제품에 대한 구매비율(X6), 글리세린 제품에 대한 구매비율(X7), 석탄산 제품에 대한 구매비율(X8)이 프로모션에 대한 고객의 반응여부에 유의한 영향을 미치는 변수로 선정되었다.

로짓에 의해 선정된 변수를 사용하여 인공신경망 모형, SVM 모형, 그리고 사례기반추론을 통해 예측모형을 구축하였다. 인공신경망 모형에서는 데이터를 훈련용(360개), 평가용(120개), 검증용(120개)으로 분할하여 학습률과 모맨텀을 각각 0.1로 하였으며 은닉노드를 1, 4, 8, 12, 17으로 변화시켜가며 시행착오(trial-and-error)방법을 통해 성과가 우수한 모형을 찾았다. 각 층에 대한 출력함수는 시그모이드 함수를 적용시켰으며 neuroshell 2 4.0을 사용하였다. 인공신경망에 대한 자세한 내용은 Rumelhart and McClelland (1986)을 참고한다.

SVM 모형을 위해서는 비선형 함수로 가장 많

〈Table 1〉 The input variables for Logit model

Variables	Descriptions	B[1]	S.E.[2]	Wald[3]	Sig.
X1	Cable & Satalite/ Non Cable & Satalite	0.474	0.242	3.827	0.050
X2	Brand Runs	0.121	0.016	56.693	0.000
X3	Purchase volume other promotion %	5.584	2.156	6.707	0.010
X4	Price Category-wise Volume % (ANY SUB-POPULAR)	1.615	0.560	8.317	0.004
X5	Proposition-wise Volume % (ANY FRESHNESS)	3.561	0.847	17.691	0.000
X6	Proposition-wise Volume % (ANY FAIRNESS)	2.634	1.044	6.373	0.012
X7	Proposition-wise Volume % (ANY GLYCERINE)	-6.071	2.245	7.310	0.007
X8	Proposition-wise Volume % (ANY CARBOLIC)	-1.401	0.552	6.445	0.011
Constant		-2.669	0.370	52.148	0.000

[1] coefficients [2] standard errors [3] Wald statistics (The higher Wald statisstic value, the more significant the coefficient)

이 사용되는 가우시안 RBF 커널함수를 사용하였으며 학습용(360개), 평가용(120개), 검증용(120개)으로 분할하여 모형을 구축하였다. 상한값을 나타내는 C와 가우시안 RBF 커널함수의 중요한 모수인 σ의 설정값을 변화시켜가며 SVM 모형에 대한 전체성과가 가장 우수한 모형을 찾았다. 모수 C={1, 20, 40, 60, 80, 100}, σ^2 ={0.5, 0.7, 1, 2.5, 5, 10, 50}로 설정하여 SVM 공개 소프트웨어인 LIBSVM(Chang and Lin, 2001)을 사용하였다. SVM에 대한 자세한 내용은 Vapnik (1995)을 참고한다.

Pure CBR에서는 로짓에 의해 선정된 8개의 변수를 사용하여 새로운 데이터와 유클리드 거리를 계산하여 변수들 간의 유사성을 측정하였다. k-최근접이웃법을 적용하여 k를 1부터 11까지 변화시켜 가며 최적의 이웃 수를 정하였으며 각 변수들의 중요도에 대한 가중치는 동일하게 적용하였다. Weighted CBR에서는 로짓에 의해 선정된 변수의 로짓계수를 적용하여 각 변수의 중요도에 대한 가중치를 상이하게 적용하여 모형을 구축하였으며 Pure CBR에서와 같이 k-최근접이웃법을 적용하여 k를 1부터 11까지 변화시켜가며 최적의 이웃 수를 정하였다. k-최근접이웃법은 Microsoft Excel의 VBA(Visual Basic for Application)을 이용해 사례기반추론 알고리즘을 구현하여 실험을 수행하였다.

반응고객과 비반응 고객의 비대칭적 분포에 따른 불균형 데이터를 대상으로 한 고객반응 예측모형의 성능의 차이를 검증하기 위해서 Pure CBR과 Weighted CBR을 전체 600개의 데이터에서 무작위로 100개의 데이터를 추출하였으며 이를 100번 반복하여 Pure CBR과 Weighted CBR를 적용하였다. 일반적으로 비반응 고객이 반응고객보다 절대적으로 많은 점을 반영하여 100번의 반복 추출에서는 비반응 고객의 수가 항상 많도록 하여 실험을 진행하였다.

4.3. 실험결과

본 연구에서는 프로모션에 대한 고객의 반응을 예측하기 위해 인공신경망 모형, SVM 모형, 그리고 사례기반모형을 구축하였으며 사례기반모형을 입력변수의 가중치를 동일하게 적용한 Pure CBR 모형과 입력변수의 가중치를 상이하게 적용한 Weighted CBR 모형을 구축하였으며 예측모형의 성과는 <Table 2>와 같이 나타났다.

로짓을 통해 입력변수를 선정한 후 모형을 구축한 <Table 2>에서는 입력변수의 가중치를 동일하게 적용한 인공신경망 모형, SVM 모형, 그리고 Pure CBR에서의 예측성과는 75.00% 로 인공신경망 모형에서 가장 높게 나타났으며 Pure CBR의 예측성과가 71.67% 로 가장 낮게 나타났다. 사례기반모형의 Pure CBR과 Weighted CBR의 예측성과를 비교해 본 결과, 로짓모형의 계수를 입력변수의 가중치로 적용한 Weighted CBR의 예측성과가 높게 나타났으며 입력변수에 가중치를 적용함으로써 예측성과가 4%이상 향상된 결과를 얻을 수 있었다. 또한 본 연구에서 제

〈Table 2〉 The results of the prediction model of customer response with balanced data

ANN		SVM		Pure CBR	Weighted CBR
Training	Validation	Training	Validation		
75.56%	75.00%	74.17%	74.17%	71.67%	75.83%

〈Table 3〉 The results of CBR models with imbalanced data

	Pure CBR	Weighted CBR
Average	65.73%	72.64%
Standard Deviation	0.1108	0.1048
Smallest Value	40.91%	50.00%
Median	65.30%	73.33%
Largest Value	91.30%	95.24%
1stQuartiles	57.14%	66.35%
3rdQuartiles	73.74%	80.00%

〈Table 4〉 Comparative analysis of the performance of CBR models according to the ratio of imbalanced data

Ratio of imbalanced data	within 5%		within 10%		within 15%		within 20%	
	Pure CBR	Weighted CBR	Pure CBR	Weighted CBR	Pure CBR	Weighted CBR	Pure CBR	Weighted CBR
Average	59.05%	68.46%	61.52%	68.93%	66.77%	75.33%	63.86%	67.53%
Max	80.00%	95.24%	76.19%	78.95%	84.21%	85.71%	84.00%	88.00%
Min	50.00%	50.00%	45.00%	60.00%	50.00%	69.57%	40.91%	50.00%
Number of data sets	12		5		7		5	

안한 Weighted CBR 모형과 인공신경망 모형, SVM 모형의 예측성과와 비교한 결과 Weighted CBR의 성과가 다른 모형에서보다 우수한 예측성과를 보여주었으며 입력변수에 가중치를 적용하지 않았을 때보다 입력변수에 가중치를 적용했을 때 예측성과가 향상되는 것을 볼 수 있다.

전체 600개의 데이터에서 반응고객과 비반응고객의 비율을 고려하지 않고 무작위로 100개의 데이터를 추출한 불균형 환경에서 Pure CBR과 Weighted CBR를 적용하였으며 이를 100번 반복한 예측성과는 <Table 3>과 같다. 불균형 환경에서의 Weighted CBR의 예측성과는 72.64%로 균형 환경에서의 Pure CBR의 예측성과 71.67%보다 높은 예측성과를 보여주었으며 불균형 데이터 환경에서도 입력변수에 가중치를 적용한 Weighted CBR의 성과가 우수하다는 것을 볼 수 있다.

<Table 4>는 불균형비에 따른 Pure CBR과 Weighted CBR의 예측성과이다. 실제 프로모션에 대한 고객 데이터는 반응고객이 비반응 고객에 비해 매우 적기 때문에 반응고객이 비반응 고객보다 적은 경우만 예측성과를 비교하였다. 반응고객과 비반응 고객에 속한 데이터의 비율 차이가 5%, 10%, 15%, 20%이내인 경우의 예측성과 모두 입력변수에 상이한 가중치를 적용한 Weighted CBR에서 동일한 가중치를 적용한 Pure CBR보다 높은 예측성과를 보여주었다. 또한 불균형의 비율이 높아진다고 Weighted CBR의 예측성과가 낮아지지는 않으므로 불균형 데이터에서 반응예측의 문제점을 해결할 수 있다.

5. 결론 및 향후 연구과제

본 연구에서는 프로모션에 대한 고객의 반응을 예측하기 위해 사례기반추론 모형을 적용하였다. 사례기반추론은 적용이 쉽고 간단하며 모형의 갱신이 실시간으로 이루어질 수 있다는 장점이 있으나 모형의 예측성과가 타 인공지능기법에 비해 상대적으로 낮게 나타나는 문제점이 있었다. 사례기반추론 모형을 위한 입력변수의 선정과 입력변수에 대한 가중치는 모형의 예측성과를 향상시켜 줄 수 있을 것이며 본 연구에서는 목욕세제 판매데이터를 적용하여 예측성과를 비교해 보았다.

모형의 입력변수는 로짓의 stepwise를 통해 선정하였으며 입력변수의 가중치를 동일하게 적용했을 때와 변수의 중요도에 따라 입력변수의 가중치를 상이하게 적용했을 때의 사례기반추론 모형을 균형 데이터와 불균형 데이터 환경에서 비교해 보았다. 이를 위해 로짓모형을 통해 선정된 8개의 변수에 로짓의 계수를 적용하여 입력변수의 가중치를 적용하였다. 입력변수에 동일한 가중치를 적용하게 되면 중요하지 않은 변수도 중요한 변수와 동일하게 반영되어 모형을 구축하게 되며 중요한 변수는 보다 많이 반영되지 못하여 예측성과를 떨어뜨릴 수 있다. 예측모형의 성과를 향상시키기 위해 균형 데이터 환경에서 입력변수의 가중치를 동일하게 적용한 Pure CBR과 중요도에 따라 입력변수의 가중치를 다르게 적용한 Weighted CBR의 예측성과를 비교한 결과 가중치를 적용한 Weighted CBR에서 높은 예측성과를 나타냈다.

또한 프로모션에 대한 고객의 반응률은 10% 미만으로 매우 낮게 나타나며 비반응 고객이 대부분을 차지하는 불균형 데이터를 이루고 있다. 불균형 데이터는 기계학습 알고리즘의 성능을 저하시키는 요인이며 소수범주에 대한 예측성과가 낮아지는 문제점이 존재한다. 불균형 데이터에서도 사례기반추론 기법이 유용하게 적용될 수 있는지 확인하기 위하여 반응고객과 비반응고객을 구분하지 않고 100명의 고객을 무작위로 추출하였으며 이를 100번 반복하여 Pure CBR과 Weighted CBR의 예측성과를 비교하였다. 불균형 데이터에서 입력변수의 가중치를 상이하게 적용한 Weighted CBR이 균형데이터에서 입력변수에 동일한 가중치를 적용한 Pure CBR에서보다 높은 예측성과를 보여주었으며 이는 불균형 데이터 환경에서도 Weighted CBR이 우수함을 보여주었다.

사례기반추론에서는 새로운 사례와 유사한 사례를 추출하기 위해 유사도를 측정하게 되며, 이는 예측성과에 중요한 영향을 미친다. 예측모형의 성과를 향상시키기 위해 유사도를 보다 정확하게 측정해야 하며 본 연구에서는 가장 많이 활용되고 있는 유클리드 거리를 통해 유사도를 측정하였다. 유클리드는 변수들 간의 단위 등이 다를 경우 큰 값을 갖는 변수가 거리의 값을 주도할 수 있다는 문제점이 존재하며 이를 해결하기 위해 향후 연구에서는 유클리드 거리뿐 아니라 마할라노비스(Mahalanobis) 거리, 민코우스키(Minkowski) 거리와 같은 방법을 적용해 볼 수 있을 것이다. 또한 유사도 기준에 의해 입력사례와 가장 가까운 기존사례를 찾아내는 방법으로 k-최근접이웃법을 적용하였으며 최적의 k를 찾기 위해 1부터 11까지 변화시켜가며 휴리스틱하게 최적의 이웃 수를 찾았으나 예측성과를 향상시킬 수 있는 최적의 이웃 수를 모형에서 자동으로 찾을 수 있도록 모형을 구축해야 할 것이다. 모형구축을 위한 입력변수를 로짓의 stepwise를

통해 선정하였으나 보다 많은 반응고객을 구별하여 예측성과를 향상시키기 위해 의미있는 변수를 선정하기 위한 방법을 향후 연구에서 제시해 보아야 할 것이며 입력변수에 대한 가중치를 부여하는 방법도 함께 생각해 보아야 할 것이다.

참고문헌(References)

Ahn, H. and K. -j. Kim, "Using genetic algorithms to optimize nearest neighbors for data mining," *Annals of Operations Research*, Vol.163, No. 1(2008), 5~18.

Ahn, H., K. -j. Kim, and I. Han, "Purchase Prediction Model using the Support Vector Machine," *Journal of Intelligence and Information Systems*, Vol.11, No.3(2005), 69~81.

Allen, B. P., "Case-based reasoning: Business applications," *Communications of the ACM*, Vol. 37, No. 3(1994), 40~42.

Barandela, J., S. Sanchez, V. Garcaa, and E. Rangel, "Strategies for Learning in Class Imbalance Problems," *Pattern Recognition*, Vol. 36, No.3(2003), 849~851.

Chan, S. L., W. H. Ip, and V. Cho, "A Model for Predicting Customer Value from Perspectives of Product Attractiveness and Marketing Strategy," *Expert Systems with Applications*, Vol. 37, No. 2(2010), 1207~1215.

Chang, C. -C. and C. -J. Lin, *LIBSVM -- A Library for Support Vector Machines*, 2001. Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm (Accessed6November,2014).

Cheung, K. -W, J. T. Kwok, M. H. Law, and K. -C. Tsui, "Mining Customer Product Ratings for Personalized Marketing," *Decision Support Systems*, Vol. 35, No. 2(2003), 231~243.

Chiu, C., "A Case-based Customer Classification Approach for Direct Marketing," *Expert Systems with Applications*, Vol. 22, No. 2(2002), 163~168.

Chiu, C., P. -C. Chang, and N. -H. Chiu, "A case-based expert support system for due-date assignment in a wafer fabrication factory," *Journal of Intelligent Manufacturing*, Vol. 14, No. 3~4(2003), 287~296.

Chung, S. and Y. Suh, "Development of a Medial Care Cost Prediction Model for Cancer Patients Using Case-Based Reasoning," *Asia Pacific Journal of Information Systems*, Vol.16, No.2(2006), 69~84.

Coenen, F., G. Swinnen, K. Vanhoof, and G. Wets, "The Improvement of Response Modeling: Combining Rule-induction and Case-based Reasoning," *Expert Systems with Applications*, Vol. 18, No. 4(2000), 307~313.

Cönül, F. F., B. D. Kim, and M. Shi, "Mailing Smarter to Catalog Customer," *Journal of Interactive Marketing*, Vol. 14, No .2(2000), 2~16.

Cui, G., M. L. Wong, and H. -K. Lui, "Machine Learning for Direct Marketing Response Models: Bayesian Networks with Evolutionary Programming," *Management Science*, Vol. 52, No. 4(2006), 597~612.

Ha, K., S. Cho, and D. MacLachlan, "Response Models based on Bagging Neural Networks," *Journal of Interactive Marketing*, Vol. 19, No. 1(2005), 17~30.

Hong, H. and S. Cho, "Case-Based Reasoning Approaches by Considering Variable Covariance Structure and Variable Weight: Corporate

Bankruptcy Prediction," *Korean Management Review*, Vol.38, No.5(2009), 1165~1184.

Hong, T. and J. Park, "Integrating the Customer Response Model in Direct Marketing Using Case-Based Reasoning," *The Journal of Information Systems*, Vol.18, No.3(2009), 375~399.

Hwang, S. and S. Cho, "Clustering-based Reference Set Reduction for k-nearest Neighbor," *Lecture Notes in Computer Science*, Vol. 4492(2007), 880~888.

Jang, Y. S., J. W. Kim, and J. Hur, "Combined Application of Data Imbalance Reduction Techniques Using Genetic Algorithm," *Journal of Intelligence and Information Systems*, Vol.14, No.3(2008), 133~154.

Jarmulak, J., S. Craw, and R. Rowe, "Self-optimizing CBR Retrieval," Proceedings of the Twelfth IEEE International Conference on Tools with Artificial Intelligence, Vancouver, Canada, (2000), 376-383.

Jo, T. and N. Japkowicz, "Class Imbalances versus Small Disjuncts," *SIGKDD Explorations Newsletter*, Vol. 6, No. 1(2004), 40~49.

Kang, P., H. ‐j. Lee, and S. Cho, "SVM Ensemble Techniques for Class Imbalance Problem," *Proceedings of Korea Information Science Society Conference*, Vol.31, No.2(2004), 706~708.

Kim, D., H. -j. Lee, and S. Cho, "Response Modeling with Support Vector Regression," *Expert Systems with Applications*, Vol. 34, No. 2(2008), 1102~1108.

Kim, K. -J., "Toward global optimization of case-based reasoning systems for financial forecasting," *Applied intelligence,* Vol. 21, No. 3(2004), 239~249.

Kim, Y. and W. N. Street, "An Intelligent System for Customer Targeting a Data Mining Approach," *Decision Support Systems*, Vol. 37, No. 2(2004), 215~228.

Kolodner, J., *Case-Based Reasoning*, Morgan Kaufman Publishers, 1993.

Lee, H. -j. and S. Cho, "Focusing on Non-respondents: Response Modeling with Novelty Detectors," *Expert Systems with Applications*, Vol. 33, No. 2(2007), 522~530.

Lee, J. S. and J. G. Kwon, "A Hybrid SVM Classifier for Imbalanced Data Sets," *Journal of Intelligence and Information Systems*, Vol.19, No.2(2013). 125~140.

Liu, Y., A. An, and X. Huang, "Boosting Prediction Accuracy on Imbalanced Datasets with SVM Ensembles," *Lecture Notes in Computer Science*, Vol. 3918(2006), 107~118.

Park, C. -S. and I. Han, "A Case-Based Reasoning with the Feature Weights Derived by Analytic Hierarchy Process for Bankruptcy Prediction," *Expert Systems with Applications*, Vol. 23, No. 3(2002), 255~264.

Roh, T. -H., M. -H. Yoo, and I. -G. Han, "Integrating rough set theory and case-based reasoning for the corporate credit evaluation," *The Journal of Information Systems*, Vol.14, No.1(2005), 41~65.

Rumelhart, D. E. and J. L. McClelland, *Parallel Distributing Processing: Exploration in the Microstructure of Cognition*, Cambridge, MA: MIT Press. Vol. 1(1986).

Shin, H. and S. Cho, "Response Modeling with Support Vector Machine," *Expert Systems with Applications*, Vol. 30, No. 4(2006), 746~760.

Shin, K. -s., and I. Han, “Case-based reasoning supported by genetic algorithms for corporate bond rating,” *Expert Systems with Applications*, Vol. 16, No. 2(1999), 85~95.

Shmueli, G., N. R. Patel, and P. C. Bruce, *Data Mining for Business Intelligence*, Wiley, 2007.

Vapnik, V., The Nature of Statistical Learning Theory, Springer, 1995

Weiss, G. M. and F. Provost, “The effect of class distribution on classifier learning,” *Technical Report*, Department of Computer Science, Rutgers University, 2001.

Yu, E. and S. Cho, "Constructing Response Model using Ensemble based on Feature Subset Selection,” *Expert Systems with Applications*, Vol. 30, No. 2(2006), 352~360.

Abstract

Response Modeling for the Marketing Promotion with Weighted Case Based Reasoning Under Imbalanced Data Distribution

Eunmi Kim* · Taeho Hong**

Response modeling is a well-known research issue for those who have tried to get more superior performance in the capability of predicting the customers' response for the marketing promotion. The response model for customers would reduce the marketing cost by identifying prospective customers from very large customer database and predicting the purchasing intention of the selected customers while the promotion which is derived from an undifferentiated marketing strategy results in unnecessary cost. In addition, the big data environment has accelerated developing the response model with data mining techniques such as CBR, neural networks and support vector machines. And CBR is one of the most major tools in business because it is known as simple and robust to apply to the response model. However, CBR is an attractive data mining technique for data mining applications in business even though it hasn't shown high performance compared to other machine learning techniques. Thus many studies have tried to improve CBR and utilized in business data mining with the enhanced algorithms or the support of other techniques such as genetic algorithm, decision tree and AHP (Analytic Process Hierarchy). Ahn and Kim(2008) utilized logit, neural networks, CBR to predict that which customers would purchase the items promoted by marketing department and tried to optimized the number of k for k-nearest neighbor with genetic algorithm for the purpose of improving the performance of the integrated model. Hong and Park(2009) noted that the integrated approach with CBR for logit, neural networks, and Support Vector Machine (SVM) showed more improved prediction ability for response of customers to marketing promotion than each data mining models such as logit, neural networks, and SVM.

This paper presented an approach to predict customers' response of marketing promotion with Case Based Reasoning. The proposed model was developed by applying different weights to each feature. We

* Part-time Lecturer, College of Business, Pusan National University
E-mail : keunmi100@pusan.ac.kr
** Corresponding author: Taeho Hong
2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan 609-735, Korea
Tel: +82-51-510-2531, E-mail: hongth@pusan.ac.kr

deployed logit model with a database including the promotion and the purchasing data of bath soap. After that, the coefficients were used to give different weights of CBR. We analyzed the performance of proposed weighted CBR based model compared to neural networks and pure CBR based model empirically and found that the proposed weighted CBR based model showed more superior performance than pure CBR model.

Imbalanced data is a common problem to build data mining model to classify a class with real data such as bankruptcy prediction, intrusion detection, fraud detection, churn management, and response modeling. Imbalanced data means that the number of instance in one class is remarkably small or large compared to the number of instance in other classes. The classification model such as response modeling has a lot of trouble to recognize the pattern from data through learning because the model tends to ignore a small number of classes while classifying a large number of classes correctly. To resolve the problem caused from imbalanced data distribution, sampling method is one of the most representative approach. The sampling method could be categorized to under sampling and over sampling. However, CBR is not sensitive to data distribution because it doesn't learn from data unlike machine learning algorithm. In this study, we investigated the robustness of our proposed model while changing the ratio of response customers and nonresponse customers to the promotion program because the response customers for the suggested promotion is always a small part of nonresponse customers in the real world. We simulated the proposed model 100 times to validate the robustness with different ratio of response customers to response customers under the imbalanced data distribution. Finally, we found that our proposed CBR based model showed superior performance than compared models under the imbalanced data sets. Our study is expected to improve the performance of response model for the promotion program with CBR under imbalanced data distribution in the real world.

Key Words : Customer response, prediction model, Case-based reasoning, feature weights, imbalanced data

Received : November 7, 2014 Revised : December 23, 2014 Accepted : December 28, 2014
Type of Submission : Fast Track Corresponding Author : Taeho Hong

저 자 소 개

김 은 미

현재 부산대학교 시간강사로 재직하고 있다. 부산대학교 경영학과에서 경영학석사와 박사학위를 취득하였으며, 주요 관심분야는 데이터마이닝, 고객관계관리, 지식경영, Social Networks 등이다. 인터넷전자상거래연구, 정보시스템연구, Information Systems Review, Expert Systems with Applications, 지능정보연구, 지식경영연구 등에 논문을 게재하였다.

홍 태 호

현재 부산대학교 경영대학 부교수로 재직하고 있다. KAIST에서 산업공학사를 취득하였고 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 딜로이트 컨설팅에서 컨설턴트로 재직했으며, 주요 관심분야는 데이터마이닝, CRM, Business Intelligence 그리고 Social Networks 등이다. Expert Systems, Expert Systems with Applications, Asia Pacific Journal of Information Systems, 그리고 정보시스템연구, 지능정보연구, Information Systems Review 등을 비롯한 국내외 학술지에 논문을 발표하였다.