

SOM을 이용한 온라인 게임 제공업체의 고객 이탈방지 방법론

김재경* · 채경희** · 송희석***

A Defection Prevention Procedure using SOM for On-line Game Providers

Jae-kyeong Kim* · Kyung-hee Chae** · Hee-seok Song***

■ Abstract ■

The retention of customer is an increasingly pressing issue in today's competitive environment. The proposes of this paper is a personalized defection detection and the procedure of prevention based on economic analysis of customer defection possibility, and behaviour state transition cost. This procedure is based on the observation that potential defectors have a tendency to take a couple of months or weeks to gradually change their behaviour before their eventual withdrawal. In this procedure, the SOM(Self-Organizing Map) is used to determine the possible states of customer behaviour from past behaviour data, and to prevent the defection of potential defectors, the proposed procedure recommends the desirable behaviour state for the next period based on the analysis of transition cost, and likelihood of defection. The case study has been conducted for a Korean on-line game provider to evaluate of this procedure.

Keyword : Customer Relationship Management, Customer Retention, Data Mining, Self-Organizing Map, Economic Analysis

논문접수일 : 2004년 7월 31일 논문게재확정일 : 2004년 10월 15일

* 경희대학교 경영대학

** 경희대학교 경영대학 일반대학원 e비즈니스전공

*** 한남대학교 경상대학 경영정보학과

1. 서론

최근 치열한 경쟁 환경으로 인해 고객 유지에 대한 관심이 증대되고 있으며, 신규 고객 유치보다는 기존 고객을 유지하는 것이 더 중요하게 되었다 [22]. 신규 고객은 초기 유치비용이 높으나, 기존 고객은 유치비용이 들지 않으며, 재구매를 통해 기업과 제품에 대해 이해하고 있음으로 해서 긍정적 구전을 통해 자발적으로 기업 마케팅을 수행할 수 있기 때문이다 [25]. 고객 유지 방법에 주로 활용되고 있는 기술 중 하나가 데이터마이닝이다. 최근 많은 기업들이 데이터베이스와 데이터웨어하우스를 이용하여 숨겨진 사실을 발견해 내기 위해 노력하고 있으며, 데이터마이닝은 데이터에서 숨겨진 사실을 발견하고, 모델을 찾아내는 것에 그 목적이 있다 [3]. 데이터마이닝은 기업이 방대한 자료 분석을 통해 고객의 패턴과 성향을 알 수 있는 매우 유용한 방법이며, 많은 연구를 통해 데이터마이닝의 활용과 효과가 입증되고 있다 [28]. 특히 통신산업과 서비스산업의 경우 고객의 인구통계학적 자료뿐 아니라, 총 사용시간, 사용 횟수 등 요금 청구를 위한 많은 양의 행위데이터를 보유하고 있다. 이러한 행위데이터는 미래의 고객행위를 예측하기 위한 정보를 포함하고 있다.

본 연구는 데이터마이닝 기술을 통해 이탈가능성이 높은 고객을 사전에 발견하여, 이탈을 방지하기 위한 방법론을 수립하는 것에 그 목적이 있다. 이 때, 이탈가능성 및 고객행위유도 비용을 고려함으로써, 보다 효율적으로 고객 개개인의 이탈을 방지할 수 있는 방법을 수립할 수 있도록 한다. 본 연구에서 제시한 방법론은 고객이 이탈하기 몇 주 혹은 몇 달 전에 특이한 공통의 행위패턴을 보인다는 것에 착안하였다. 이와 같은 고객의 이탈 신호를 사전에 파악함으로써, 기업은 고객이 이탈하기 전에 고객의 이탈을 예측할 수 있을 뿐 아니라, 고객으로부터 이탈 신호가 보이면, 기업은 즉시 신호를 보인 고객의 행위를 기존의 비이탈 고객의 행위 상태로 유도함으로써 고객의 이탈을 사전에 막을

수 있다. 또한, 이탈 신호를 보이지 않은 고객이라도 이탈 가능성을 감소시킴으로써 이탈을 미리 예방할 수 있게 된다.

하지만, 고객의 행위를 파악하고 예측한다는 것은 다음과 같은 이유로 어렵게 된다. 고객의 행위 데이터는 몇 주, 혹은 몇 달과 같이 다기간 (multi-period) 데이터이며, 한 가지 요인만으로 예측할 수 있는 것이 아니다. 또한 행위 예측에 활용되는 데이터는 비선형적이기 때문에 본 논문에서는 이와 같은 문제점들을 극복하고, 고객의 행위 상태를 정의하기 위한 방법으로 데이터마이닝 기법의 하나인 SOM(Self-Organizing Map)을 사용하였다. SOM은 복잡하고 다양한 차원의 데이터를 이해하기 쉽고 가시적인 형태로 보여준다는 장점이 있다. 또한 실제 데이터의 위치를 그대로 보존하는 특성이 있다 [6]. SOM을 통해 고객의 상태가 정의되면, 다음으로 각 고객 집단별 이탈 가능성을 파악한다. 이탈 가능성에 기반하여, 이탈 가능성이 높은 고객은 이탈 가능성이 낮은 방향으로 유도되도록 한다. 본 논문에서는 고객의 행위 유도를 위하여 수익률을 사용한다. 이상적인 방법은 모든 고객을 이탈 가능성이 가장 낮은 집단으로 유도하는 것이나, 이는 지나치게 많은 마케팅 비용이 소요될 뿐 아니라 비효율적이다. 따라서 수익률을 사용하여, 수익률 대비 이익이 높은 방향으로 행위를 유도한다.

온라인 환경에서 고객의 이탈 전 특이 행동을 발견하기 쉬운 분야 중의 하나가 온라인 게임 산업이다. 특이 행동이라는 것은 기존의 행동과 다른 행동을 의미하므로, 정기적인 사이트 방문은 특이 행동 파악에 있어 매우 중요한 요소인데, 이는 일정한 패턴이 없고, 간헐적으로 방문하는 고객에서는 특이 행동을 찾아내기 어렵기 때문이다. 온라인 게임 산업의 고객은 다른 온라인 산업의 고객에 비해 정기적으로 방문하는 경우가 많으므로, 사용자 행위 데이터 수집이 용이할 뿐 아니라, 사용자 행위 데이터가 자세히 기록됨에 따라 특이 행동을 파악하기도 쉽다. 통계에 의하면 인터넷 사용자들

은 다른 사이트보다 게임 사이트에 자주 오랫동안 방문하고 있으며, 인터넷 사용자들의 19%가 정기적으로 인터넷 게임을 사용하는 것으로 나타났다 [3]. 따라서 본 연구에서 제시한 방법론을 적용하기에 가장 적절한 분야로 판단되어, 온라인 게임 제공업체의 고객 데이터를 통해 제시한 방법론의 활용성과 타당성을 검증해 보았다.

본 연구의 구성은 다음과 같다. 2장에서는 고객 이탈 방지와 관련된 선행 연구와 사용될 알고리즘들에 대해 간단히 살펴보았다. 3장에서는 이탈방지 방법론에 대해 설명하고 있으며, 4장에서는 실제 사례에 적용한 결과를 분석하였다. 마지막으로 5장에서는 연구의 정리와 추후 연구 방향을 제시하였다.

2. 연구배경

2.1 고객 이탈 방지에 관한 선행연구

데이터마이닝 기술을 활용하여 잠재적인 이탈 고객들을 발견하기 위한 많은 연구들이 선행되어 왔다. 이 연구들을 적용 분야, 연구 목적, 연구에 활용된 기법, 분석에 사용된 데이터 등의 기준으로 정리하면 <표 1>과 같다.

Berson et al.[2], Reghavan et al.[23], Datta et al.[11] 등의 연구에서는 주로 인구통계학적 변수를 활용한 예측 모델을 구축하였으며, 이탈 방지 방법론에 초점을 맞춘 것이 아닌, 정확성이 높은 모형의 구축과 모형을 활용한 이탈 고객 예측이 주된 목적이었다. 반면, Ng and Liu의 연구[22]에서는 데이터베이스에 저장된 로그 데이터를 변수로 사용하여 잠재적 이탈 고객을 예측하기도 하였는데, 이 모든 연구들은 단지 이탈 고객을 발견하는 것에 초점을 맞추고 있다. 비록 고객이 이탈하기 전에 이탈 가능성이 높은 고객을 발견할 수 있다 하더라도 고객의 이탈 요인이 통제되거나 해결되지 않으므로 결국 이탈은 발생하게 된다. 따라서 이와 같이 고객의 이탈 요인을 통제하기 위한 연구가 Kim et al.[15], Song et al.[24]을 통해 이루어지기도 하였다. 그러나 Kim et al.[15], Song et al.[24]의 연구에서는 경영학에 있어 중요한 측면인 효율성 측면과 비용 측면이 고려되고 있지 않으므로 제시한 이탈 방지 전략이 여러 가능한 이탈 방지 전략 중에서 적절하다는 것을 입증하기 어렵다. 본 연구는 기존의 Song et al.[24] 연구를 더욱 확대시켜서 고객 통제 및 유도를 위해, 비용에 기반 한 경제적 분석을 함으로써 효율적인 이탈 방지 방법의 제시를 가능하게 한다.

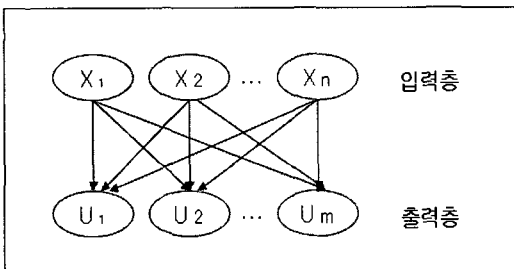
<표 1> 이탈 방지 절차에 관한 기존 연구

기존 문헌	적용 분야	연구 목적	활용 기법	데이터
Berson et al. (2000)	무선통신산업	캠페인을 실시한 고객 그룹과 실시하지 않은 고객 그룹을 대상으로 이탈 확률 분석	의사결정나무	인구통계학적 정보 및 고객 불만 접수 사항
Datta et al. (2000)	무선통신산업	대규모 무선통신 산업을 위한 자동화 된 이탈 분석 및 예측 모델 개발	인공신경망	사용 및 계좌 정보와 같은 행위 데이터와 인구통계학적 정보
Ng and Liu (2000)	위성방송산업	여러 가지 기법을 활용하여 통합적으로 활용함으로써 이탈 고객 예측	의사결정나무, 분산분석	위성 방송 신청 수 및 기간별 데이터 총 40개
Reghavan et al. (2000)	인터넷서비스 제공산업(ISP)	인터넷 서비스 제공자를 위한 이탈 고객 예측 모델 구축	로지스틱 회귀분석	총 10개의 사용기록 데이터 및 인구통계학적 데이터
Kim et al. (2003)	온라인 게임 산업	SOM과 마코프체인을 이용한 동적 이탈방지 방법론	SOM, 의사결정나무, 마코프체인	사용시간, 접속횟수, 오류횟수 등의 사용자 행위데이터
Song et al. (2004)	온라인 게임 산업	온라인 게임 사용자에 대한 이탈가능성 분석 및 이탈방지 방법론 개발	SOM, 의사결정나무	사용시간, 접속횟수, 오류횟수 등의 사용자 행위데이터

2.2 Self-Organizing Map(SOM)

SOM은 본 연구에서 제시하고 있는 방법의 가장 핵심이 되는 분석 방법으로써, Kohonen[18-20]에 의해 제시, 개발되어 Kohonen Maps이라고도 알려져 있다. SOM은 고차원의 데이터를 이해하기 쉬운 저차원의 뉴런(neuron)으로 정렬하여 지도(map)의 형태로 형상화한다. 이러한 형상화는 입력 변수의 위치 관계 그대로 보존한다는 특징이 있다. 다시 말해 실제 공간의 입력 변수가 가까이 있으면, 지도상에서도 가까운 위치에 있게 된다. 이러한 SOM의 특징으로 인해, 입력 변수의 정보와 그들의 관계가 지도상에 그대로 나타나게 되는 것이다.

SOM 모델은 <그림 1>과 같이 두 개의 인공신경망 층으로 구성되어 있다. 하나는 입력층이고, 다른 하나는 출력층이다. 입력층은 입력 변수의 개수와 동일하게 뉴런 수가 존재하고 출력층은 사용자가 미리 정해놓은 군집의 수만큼 뉴런 수가 존재한다. 입력층의 자료는 학습을 통하여 출력층에 정렬되는데, 이를 지도라 부른다. 입력층에 있는 각각의 뉴런은 출력층에 있는 각각의 뉴런들과 연결되어 있다. 따라서 출력층의 각 뉴런은 입력층의 뉴런 수와 같은 수의 연결고리가 존재하며 이 연결고리는 연결 가중치(synaptic weight)를 가진다. SOM에서 각각의 뉴런은 n -차원의 가중치 벡터(weight vector), $m = [m_1, m_2, \dots, m_n]$ 에 의해 표현되어진다. 이 때 n 은 입력층의 뉴런의 수와 같다.



<그림 1> Kohonen Maps의 기본 개념

SOM은 반복해서 학습되어진다. 각 학습 단계마다 입력층의 데이터 집합으로부터 하나의 표본 벡터(sample vector) x 가 임의로 선택되어 그것과 SOM의 모든 가중치 벡터의 거리가 유클리드 방법 등에 의해 계산되고 비교된다. 입력층에 가장 가까운 가중치 벡터를 선택하여 BMU(Best-Matching Unit)라 명명한다. BMU는 $\|x - m_c\| = \min_i \{\|x - m_i\|\}$ 와 같은 방법으로 산출되며, $\|\cdot\|$ 은 유클리드 거리를 측정한 것이고, m_c 는 BMU의 가중치 벡터이다. BMU가 선택된 후 SOM의 가중치 벡터가 개선됨으로써 BMU는 입력층에 있는 입력 벡터에 더 가깝게 이동한다. BMU의 위상학적 이웃(topological neighbors) 또한 이와 같은 방법으로 선택되고 이동한다. 자세한 내용은 Kohonen의 연구[18-20]를 참조하기 바란다.

SOM은 전문검색 마이닝(full-text mining)과 금융 데이터 분석 등을 통해 가치 있는 마이닝 도구가 입증되었다. 또한 여러 기술 분야에서의 패턴 발견, 이미지 분석, 프로세스 모니터링, 오류 진단 방법 등으로 활용한 결과가 성공적임에 따라 그 적용 분야도 다양해졌다[27]. 특히 동적 프로세스 모니터링 분야에서 SOM은 프로세스의 운영 상태를 나타내기 위한 방법으로 사용되었다. Tryba and Goser[26], 그리고 Kasslin의 연구[14]에서는 SOM을 활용하여 제지 산업의 복잡한 프로세스를 모니터링하고 통제하였다. 이 사례에서는 종이의 질을 좌우하는 원료의 배합, 온도, 습도, 시간 등 복잡하고 다양한 프로세스를 SOM을 통하여 이해하기 쉬운 2차원의 지도로 나타냄으로써 보다 쉽게 현재 상태를 이해하고, 통제한다. 또한 운영 시점, 예를 들어 현재 운영 상태에서와 과거의 자료가 지도에서 궤도로써 시각적으로 표현된다. 이를 통해 동적 프로세스를 추적하는 것이 가능한 것이다. SOM은 프로세스를 쉽게 이해할 수 있도록 함으로써, 각 변수와 그들 간의 상호작용을 동시에 분석할 수 있도록 도와준다[6]. 본 연구는 SOM의 이러한 개념을 도입하여 변수와 변수들 간의 차이와 상호작용을 분석해 보고자 한다.

2.3 의사결정나무 분석

의사결정나무분석은 SOM에 의해 나타난 군집들을 레벨링(labeling)할 때 사용되며, 예측과 분류를 위한 보편적이고 강력한 데이터마이닝 도구이다[2]. 인공신경망 분석(Feed forward neural networks)과 함께 다양한 분야에서 많이 활용되고 있는데[9], 인공신경망 분석과는 달리 나무구조로 규칙을 표현하기 때문에 규칙이 이해하기가 쉽게 나타난다.

경우에 따라서는 얼마나 잘 분류하거나 예측했는지에 대한 결과만이 중요하기도 하다. 즉, DM 발송회사가 발송하게 될 고객을 선정할 때, 모델이 어떻게 구성되었는지 그 과정보다는, 얼마나 자신의 메일에 잘 대답을 해줄 수 있는 집단을 선정했는지에 대한 결과가 중요할 것이다. 이런 경우 분석 과정의 해석은 어려우나 예측력이 뛰어난 인공신경망 분석이 더 유용할 수도 있다. 하지만, 어떤 경우에는 왜 이런 결정을 하게 되었는지 설명하는 것도 중요할 수 있다. 예를 들어, 카드신청자의 카드 발급을 거절해야 하는 경우 그것의 결과를 설명할 수 없는 인공신경망 분석보다 이유를 설명해 줄 수 있는 의사결정나무분석이 더 유용할 것이며, 의사결정나무 분석은 과정이 단순 명료하게 나타나므로 이러한 경우 용이하게 설명해 줄 수 있다.

본 연구에서 제시한 방법론은 의사결정나무 분

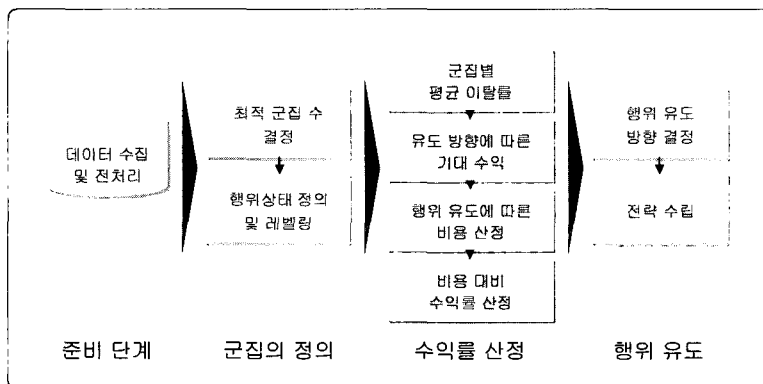
석의 이와 같은 설명력을 이용하여, SOM에 의해 분류된 고객 집단의 상태를 알 수 있게 한다. 이와 같은 고객의 상태와 과거 데이터 분석을 바탕으로 고객의 이탈 가능성을 파악하게 되는 것이다. 의사결정나무 분석 알고리즘으로는 CHAID[13], CART[1], C4.5[5], QUEST[21] 등이 있다.

3. 고객 이탈 방지 방법

3.1 분석 절차

본 연구에서 제시하고 있는 방법론은 Tryba and Goser[26]와 Kasslin의 연구[14]에서 기본 개념을 획득하였다. SOM은 각각의 변수가 독립적이지 않아도, 또한 비선형적이라도 모든 데이터를 동시에 관찰할 수 있다[6]. 본 연구에서 대상으로 하고 있는 인터넷 온라인 게임 사이트의 고객 데이터는 서로 완전히 독립적이지 않으며, 시계열 데이터이고, 여러 요인들로 인해 고객 상태를 한눈에 파악하기 어렵다. SOM은 이러한 특징을 갖는 데이터를 잘 처리해 줄 뿐만 아니라, SOM의 지도는 위치보존 특성을 가지고 있으므로 군집간 거리를 쉽게 가늠해 볼 수 있다는 장점이 있다.

제시한 방법론의 전체적인 절차는 <그림 2>와 같다. 이와 같은 절차는 고객의 상태를 파악하고, 고객의 행위를 유도하기 위한 절차이다. 다시 말해,



<그림 2> 전체적인 분석 과정

이 절차의 핵심은 고객의 행위유도 방향을 결정하는데 있으며, 이 때 고객 간 거리를 통해 비용을 유추해 봄으로써 최종 방향을 결정하게 된다.

전체적인 절차는 크게 4단계로 이루어지며, 첫 번째 단계에서는 고객 행위와 관련된 웹로그(web log) 데이터를 수집하여 분석에 적절한 단위로 정리한다. 두 번째 단계에서는 SOM을 이용하여 고객을 군집 별로 나누고 의사결정나무 분석을 이용하여 정의하게 된다. SOM을 이용하여 군집을 나눌 때에는 먼저, 나누게 될 군집의 수를 결정해야 하는데, 군집의 수가 지나치게 많으면 과잉적합(overfitting)되어 일반화 기능을 상실하게 되고, 수가 지나치게 적으면 이탈, 비이탈 고객에 대한 식별 능력을 상실하게 된다. 따라서 여러 가지 경우로 군집의 수를 나누어 본 후, 분석하여 오류를 살펴봄으로써 적절한 군집 수를 결정한다. 이 때, 오류는 2종 오류, 즉 이탈한 고객을 이탈하지 않았다고 생각하는 것에 비해 1종 오류, 즉 이탈하지 않은 고객을 이탈했다고 보는 것이 더 위험하므로 1종 오류를 5% 이내로 제한하는 범위 내에서 2종 오류가 가장 적은 군집의 수를 채택하였다. 최적 군집 수가 결정되면 고객의 행위 상태에 대해 정의한다. SOM은 군집을 나누게 된 과정에 대해 설명하지 못하므로 의사결정나무 분석을 실시하여 군집을 나누게 된 과정, 즉 의사결정나무의 규칙을 통해 군집별 특성을 살펴보고, 각 군집을 이해한다.

세 번째 단계에서는 각 군집의 평균 이탈률을 통해 비이탈 고객의 이탈 가능성을 예상해 본다. 즉 군집내 대부분의 사람들이 이탈했다면, 비록 이탈하지 않은 고객이라도 이탈 위험이 높은 고객으로 분류한다. 군집별 이탈 가능성이 파악되면, 다음으로 수익률을 분석한다. 고객을 이탈 가능성이 높은 집단으로부터 이탈 가능성이 낮은 집단으로 고객 행위를 유도하는 것이 목적이나, 집단 간의 거리가 먼 경우 지나치게 많은 마케팅 비용이 소요되므로 거리를 비용으로 계산하여, 비용대비 기대 수익률이 가장 높은 방향으로 유도한다.

마지막 단계에서는 고객 집단별 행위 유도 방법

을 제시하게 된다. 세 번째 단계에서 산출된 비용 대비 기대 수익률을 기반으로 행위 유도 방향을 결정한 후, 고객이 현재 속해있는 집단과 목표 집단 간의 세부적인 차이를 살펴본다. 그리고 고객을 목표 집단으로 유도하기 위해서 어떠한 조치가 필요한지 결정하게 되며, 그것을 바탕으로 마케팅 캠페인을 수립하게 된다. 자세한 내용은 다음 절에서 세부적으로 살펴보도록 한다.

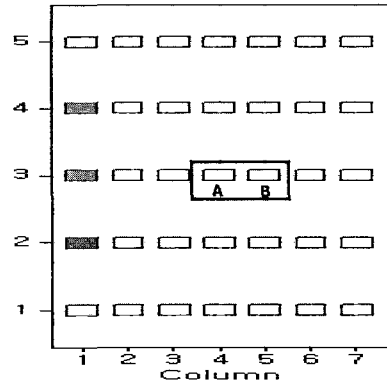
3.2 데이터 수집 및 전처리

분석에 사용될 변수는 영역 전문가에 의해 선택된다. 또한 데이터의 값들은 오랜 시간에 걸쳐 데이터베이스에 저장된 웹로그 파일로부터 수집된다. 웹로그 데이터의 전처리 과정에 대한 자세한 내용은 Cooley의 논문[10]을 참조하기 바란다.

본 연구에서 제시하고 있는 방법을 설명하기 위하여, 전처리 과정을 통과한 온라인 게임 업체의 데이터는 한 달 동안의 총 사용 횟수, 한 달 동안의 총 접속 횟수, 한 달 동안의 총 에러 횟수, 한 달 평균 사용시간 등이다. 에러는 로그아웃하고 스스로 종료한 것이 아닌 본인의 컴퓨터 혹은 프로그램의 에러, 잘못된 조작 등으로 강제 종료된 것을 말한다. 이러한 변수들의 선택은 영역 전문가들에 의해 이루어지며, 마케팅 담당자 혹은 의사결정자들에 의해 결정되어 진다. SOM의 학습을 위해 선택된 변수와 데이터들은 정규화 과정을 거쳐야 하는데, 상세한 정규화 과정은 Chakraborty의 논문[8]에 설명되어 있으며, 변수 선택 및 데이터 전처리 과정은 Hall and Smith[12], Kohavi and John[17], Almuallim and Dietterich[7], Kira and Rendell[16] 등을 참조하기 바란다.

온라인 게임 업체에서 사용하고 있는 웹로그 데이터 추출 과정은 다음과 같다. 우선 데이터베이스로부터 분석을 위한 표본 ID를 추출한다. 그리고 나서, 표본으로 선정된 고객 ID의 한 달 동안의 로그 데이터를 모두 수집한다. 수집된 로그 데이터 중에서 로그인 데이터와 로그아웃 데이터 및 정상

종료 여부 데이터를 추출한다. 다음으로, 각 ID별 총 로그인 횟수를 세어 접속 횟수에 저장한다. 또한, 각 ID별 로그아웃 시간에서 로그인 시간을 감한 후 그것의 합을 총 사용 시간에 저장하고, 각 ID별 정상 종료 여부 중, 비정상 종료의 경우를 세어 에러 횟수에 저장한다. 이 때, 로그인 시간은 존재하지만 로그아웃 시간이 없거나 또는, 로그인 시간과 로그아웃 시간이 같은 경우는 비정상 종료로 간주하여 에러 횟수에 포함한다. 마지막으로 총 사용 시간을 접속 횟수로 나누어 평균 사용 시간에 저장한다. 이와 같은 단계를 수행하고 나면 결과적으로 고객의 ID, 한 달 동안의 총 접속횟수 및 총 사용시간, 에러, 평균 사용시간 등이 나타나게 된다.



<그림 3> SOM의 학습을 통해 결정된 고객 군집

<표 2> 오분류표

실제값 \ 예측값	비이탈	이탈
비이탈	정확한 예측	오류 영역 2
이탈	오류 영역 1	정확한 예측

3.3 군집의 정의

3.3.1 최적 군집수 결정

분석을 위한 데이터가 준비되었으면, 다음으로 준비된 데이터를 사용하여 고객을 군집별로 나눈다. 각 고객의 한 달 동안의 접속횟수, 총 사용시간, 에러, 평균 사용시간 등을 기반으로 SOM을 사용하여 군집을 형성한다. SOM은 다차원의 시계열 데이터 분석에 유용하며, 분석결과가 <그림 3>과 같이 지도의 형태로 나타나는데, 지도상에 가깝게 위치한 집단은 실제적으로도 가깝게 존재하는 위치보존 특성을 가지고 있다. 예를 들어 <그림 3>의 A군집과 그 이웃인 B군집에 속해 있는 사람은 실제로도 유사한 특성을 지니고 있게 된다. 이러한 SOM의 특성은 행위유도 시 발생하는 비용을 계산하는데 사용된다.

SOM을 사용하여 군집을 나누기에 앞서, 먼저 적절한 군집의 수를 정해야 하는데, <표 2>와 같은 오분류표를 이용하여 결정한다. 오분류표가 작성되면, '오류 영역 1'이 95% 신뢰수준을 만족시키는 범위, 즉 오류가 5%를 넘지 않는 범위 내에서 '오류 영역 2'가 가장 적게 나타나는 군집의 수를 최적 군집수로 선택하게 된다.

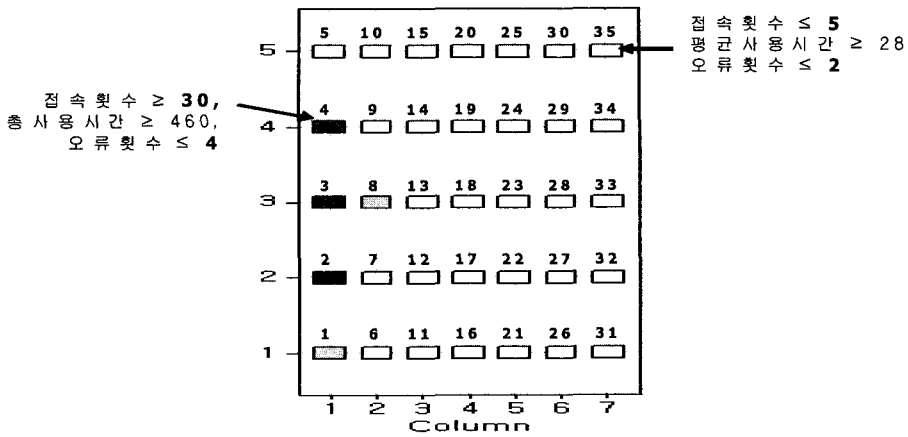
예를 들어, 군집을 4×4와 4×6, 7×5로 나눈 결과 <표 3>와 같은 오분류표가 나왔다고 가정하자. 이 때 최적 군집수에 대한 결정을 해야 하는데, 앞서 언급했듯이 이탈하지 않은 고객을 이탈했다고 보는 것의 더 위험한 의사결정이므로, '오류 영역 1'의 값이 5% 이하인 범위 내에서 '오류 영역 2'의 값이 가장 낮은 7×5가 최적의 군집 형태가 된다.

<표 3> 각 군집수에 따른 오분류표의 예

실제값 \ 예측값	4×4		4×6		7×5	
	비이탈	이탈	비이탈	이탈	비이탈	이탈
비이탈	-	0.5	-	0.9	-	0.6
이탈	0.8	-	0.3	-	0.4	-

3.3.2 행위상태 정의 및 레벨링

최적 군집수를 결정하고 SOM을 학습시켰으면, 다음으로 각 군집에 어떤 특성을 가진 고객들이 분포되어 있는지, 어떤 기준에 의해 군집이 나뉘어졌는지 살펴봐야 한다. SOM을 통해 학습된 데이터들은 지도상에 서로 간의 유사성에 따라 군집의 형태로 분포되고, 이러한 군집들은 의사결정나무



<그림 4> SOM의 고객 군집 및 정의

분석에 의해 그 특성이 설명된다. 의사결정나무 분석 결과, 발견되는 규칙들은 리프노드(leaf load)인 SOM의 각 군집이 분석에 사용된 네 가지 변수의 속성 중 어느 부분에 속하는지 이해하기 쉽고 간단하게 표현해 준다. <그림 4>는 각 군집별 ID(Cluster ID)와 각 ID에 대한 설명을 보여준다. <그림 4>의 예에서 ID '4'는 총 접속횟수가 30회 이상이고, 총 사용시간이 460분 이상이며, 오류는 4회 이하인 고객들이 속해 있는 군집이다. 이와 같은 각 군집의 특성은 고객의 행위 유도 방안 수립 시 필요하게 된다.

3.4 수익률 산정

SOM을 통해 고객을 군집화 하고 나면, 각 군집에 속해있는 고객이 몇 명이나 이탈했는지를 계산하여 군집별 이탈률을 산정할 수 있다. 이것을 바탕으로 그 군집에 속해 있는 고객의 이탈 가능성을 추정해 볼 수 있는 것이다. 또한 이탈률이 낮은 집단을 발견함으로써, 이탈률이 높은 군집에 속하는 고객의 행위 유도 방향을 결정할 수 있게 된다. 이 때, 가장 이탈률이 낮은 집단으로 고객 행위를 유도하는 것이 좋으나, 그에 따르는 마케팅 비용을 간과 할 수 없다. 따라서 고객의 행위를 어느 방향으로 유도할지 결정하기 위해 고객의 행위를 유도

할 경우 발생하는 수익과 비용을 고려하여 수익률을 산정하고, 가장 수익률이 높은 방향으로 고객의 행위를 유도할 필요가 있다. 본 연구에서는 이와 같이 수익과 비용에 기반 한 행위 유도 방법을 제시하며, 세부과정은 다음과 같다.

- 단계 1 : 각 군집별 평균 이탈률 계산
- 단계 2 : 다른 군집으로 고객 행위를 유도할 경우 발생하는 수익(gain) 계산
- 단계 3 : 행위를 유도에 따라 발생하는 비용 계산
- 단계 4 : 비용 대비 기대 수익률 계산

3.4.1 군집별 평균 이탈률 계산

SOM을 통해 나타난 각 군집의 평균 이탈률을 계산해 봄으로써, 이탈 가능성을 추정할 수 있다. 즉, 어느 한 고객이 속해 있는 군집의 평균 이탈률이 90%라 하면, 아직 이탈하지 않은 고객이라 할 지라도 이탈가능성이 매우 높은 고객으로 분류된다. 이와 같이 계산된 이탈률은 다음 단계에서 계산되는 비용과 함께, 최종적으로 고객의 행위 유도 방향을 설정하기 위한 기준이 된다.

군집 X의 이탈률 P_x 를 수식으로 정리하면 다음과 같다 : $P_x = D_x / N_x \times 100$.

여기에서 N_x 는 군집 X에 속한 고객의 수, D_x 는 군집 X내의 이탈 고객의 수이고, 이 때 각각의

군집은 X, Y, Z ... 등으로 구분되며, 각 군집에 속한 개인은 i 로 정의한다. 즉, X그룹에 속한 고객은 x_i 로 표시된다.

군집의 평균 이탈률(P_x)계산에 대한 예를 들어 보면 <표 4>과 같다. 군집 '1'에 '가, 나, 다, 라, 마' 고객이 속해 있다고 가정하자. '가' 고객은 이탈, '나' 고객은 비이탈, '다', '라', '마' 고객은 이탈한 경우, 이 군집의 평균 이탈률은 80%가 된다. 이와 같은 방법으로 군집 '2'의 평균 이탈률은 50%가 된다.

<표 4> 군집별 평균 이탈률 예

군집 ID	고객 ID	이탈 여부	평균 이탈률
1	가	이탈	80%
	나	비이탈	
	다	이탈	
	라	이탈	
	마	이탈	
2	바	이탈	50%
사	비이탈		
3	아	이탈	25%
	자	비이탈	
	차	비이탈	
	카	비이탈	

3.4.2 고객 행위 유도 시 발생하는 수익 산정
 각 군집의 평균 이탈률을 토대로 목표 방향으로 그 방향으로 유도했을 경우의 수익($G_{x,y}$)을 계산하면 다음과 같다: $G_{x,y} = P_x - P_y$. 즉, 수익은 이탈률이 높은 곳에서 낮은 곳으로 고객을 유도하였을 경우 발생하는 이탈률의 감소량으로 정의된다.

<표 4>의 예를 이용하여 행위 유도에 따라 발생 가능한 기대 수익을 계산하면 <표 5>와 같다. 기본 개념은 이탈 가능성이 높은 집단에 속해 있는 고객 또한 이탈 위험이 높으며, 따라서 이탈 가능성이 낮은 집단으로 유도한다는 것이다. 즉, 이탈하지 않은 고객이 이탈 가능성이 높은 집단에 속해 있다면, 이탈이 발생하기 전에 이탈 가능성이 낮은 집단으로 유도하여 이탈할 확률을 줄여보자는 것이 본 연구에서 제시한 방법의 목적이다. <표 4>의 예를 살펴보면 군집 '1'의 '나' 고객은 현재 비이

탈 고객이지는 않지만, '나' 고객이 속해 있는 군집은 평균 이탈률이 80%로 매우 높다. 따라서 '나' 고객 또한 이탈할 가능성이 매우 높으므로 이탈률이 낮은 '2' 또는 '3'으로 유도해야 한다. 이 때, '2'의 방향으로 유도하는 것이 바람직한지, 아니면, '3'으로 유도하는 것이 바람직한지는 다음 단계에서 계산되는 비용을 고려하여 결정하게 된다.

<표 5> 고객 행위 유도 시 발생하는 수익

고객 행위 유도 방향	획득 가능한 수익
1 → 2	80 - 50 = 30(%)
1 → 3	80 - 25 = 55(%)
2 → 3	50 - 25 = 25(%)

3.4.3 고객의 행위 유도 시 발생하는 비용 산정

고객의 행위 유도는 수익과 비용을 모두 고려하여 이탈률은 높고 비용은 적은 방향, 다시 말해 비용 대비 수익률이 가장 높은 방향으로 이루어지게 된다. 고객의 행위를 원하는 방향으로 유도하려면, 기업에서는 그에 상응하는 조치를 취하여야 할 것이고, 이와 같은 기업 행동에는 비용이 발생하게 된다.

SOM을 통해 생성된 지도는 실제 가까이 존재하는 변수를 지도상에도 가깝게 정렬시키므로 지도 상에서 거리가 멀게 나타난 군집은 실제로도 관계가 멀다고 볼 수 있다. 다시 말해 거리가 먼 집단으로 고객을 유도할 경우 실제로도 큰 비용이 발생한다는 것이다. 그러므로 고객과 유도하고자 하는 집단과의 거리를 계산하여 비용을 예상해 볼 수 있다. 군집 X의 고객 x_i 를 군집 Y로 유도하고자 할 때, 예상되는 비용을 $C_{x,y}$ 라 하면, $C_{x,y}$ 는 다음과 같이 정의 된다: $C_{x,y} = \sum_{i=1}^{N_x} D(x_i, y_j) / N_x$. 이 때, $D(x_i, y_j)$ 는 x_i 와 y_j 와의 유클리드 거리를 의미한다. $C_{x,y}$ 가 계산되면, 이 값을 이용하여 군집간 거리($C_{x,y}$), 즉 군집 간 유도비용을 계산할 수 있는데, 이는 다음과 같이 정의 된다:

$$C_{x,y} = \sum_{i=1}^{N_x} C_{x_i,y} / N_x$$

<표 4>에 나타나는 각 고객 간의 거리를 <표 6>와 같다고 가정해보자. <표 6>에서의 각 고객간 거리는 이해를 돕기 위해 임의의 수치를 활용하지만, 실제 사례에서는 프로그램을 통하여 비용 ($C_{x,y}$)이 자동적으로 계산된다. <표 6>를 이용하여 $C_{x,y}$ 를 계산해 보면 <표 7>과 같으며, <표 7>을 기반으로 군집간 거리, 즉 비용($C_{x,y}$)을 계산하면 <표 8>과 같다.

<표 6> 각 고객 간의 거리 예

		군 집 2		군 집 3			
		바	사	아	자	차	카
군 집 1	가	1	2	5	5	8	7
	나	2	3	5	5	7	6
	다	3	1	7	5	7	6
	라	2	1	7	7	8	8
	마	1	1	6	5	5	7

<표 7> 행위 유도할 고객과 목표 군집간의 거리 예

		군 집 2	군 집 3
군 집 1	가	1.50	6.25
	나	2.50	5.75
	다	2.00	6.25
	라	1.50	7.50
	마	1	5.75

<표 8> 군집간 거리 예

	군 집 2	군 집 3
군집1	1.7	6.3

위의 예를 살펴보면 <표 7>의 '가'와 군집 '2'와의 거리는 1.50인 것으로 나타나 있다. 이는 <표 6>의 '가'와 '바'의 거리, 그리고 '가'와 '사'의 거리를 합한 후, 군집 '2'에 포함된 고객의 수, 2로 나눈 값이다. 이와 같은 방법으로 군집 '1'에 속해있는 각각의 고객과 군집 '2' 및 '3'과의 거리를 계산한 결과가 <표 7>에 나타나 있다. <표 7>의 값을 이용하여 군집 '1'에 속해 있는 각각의 고객과 군집 '2'와의 거리를 모두 합한 후, 군집 '1'의 수, 5로 나

눈 값이 군집 '1'과 군집 '2'와의 거리 1.7이다.

3.4.4 비용 대비 수익률 산정

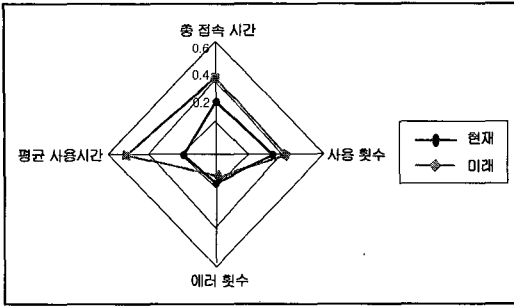
고객의 행위 유도에 따라 발생될 수 있는 수익을 계산하고, 그에 따른 비용을 계산하였으면, 마지막 단계로 비용 대비 수익률을 산정하여 고객의 행위 유도 방향을 결정하게 된다. 군집 X에서 군집 Y로 이동시 발생 가능한 비용 대비 수익률 $E_{x,y}$ 의 정의는 다음과 같다: $E_{x,y} = G_{x,y} / C_{x,y}$. 앞서 제시된 <표 5>와 <표 8>의 예를 활용하여 비용 대비 수익률을 계산한 결과가 <표 9>에 나타나 있다. <표 9>을 살펴보면, 수익률은 군집 '1'에서 '3'으로 유도할 때 가장 높으나, 비용을 고려할 경우 군집 '2'로 유도할 때가 수익이 높은 것을 알 수 있다.

<표 9> 비용 대비 수익률 예

현재의 군집	목표 군집	수익률	비용	비용 대비 수익률
군집 1	군집 2	30	1.7	17.65
	군집 3	55	6.3	8.73

3.5 고객의 행위 유도

비용 대비 수익률을 산정하고, 그 결과를 바탕으로 어느 군집을 목표로 행위를 유도할 것인지 결정해야 한다. <표 9>의 경우에서는 군집 '3'보다 군집 '2'로 행위를 유도할 경우 수익이 더 크므로 군집 '2'를 목표 군집으로 정한다. 목표 군집이 결정되면, 앞서 정의해 놓은 고객 특성을 이용하여 행위 유도 전략을 수립할 수 있다. <그림 5>는 마케팅 캠페인 설계를 위한 사례로써, 현재 고객의 행위 상태(현재)와 유도하고자 하는 행위 상태(미래) 즉, 목표 군집을 비교하여 보여준다. <그림 5>는 현재 고객의 사용 시간이 크게 부족함을 알 수 있다. 이와 같은 고객의 이탈을 방지하기 위한 방법으로는 사용 시간을 늘릴 수 있는 무료 사용 쿠폰 발송이나, 포인트 적립, 새로운 게임의 일정기간 무료 사용과 같은 캠페인 예를 들 수 있을 것이다.



〈그림 5〉 행위 유도를 위한 마케팅 캠페인 설계의 예

4. 사례 연구

3장에서 제시한 이탈방지 모형의 과정들을 실제 온라인 게임 제공업체 사례에 적용하여 보았다. 데이터 분석을 위하여 Visual Basic 6.0으로 프로그래밍 하였으며, SOM과 의사결정나무 분석은 SAS사의 e-Miner를 사용하였다.

4.1 데이터

K 온라인 게임 제공업체의 데이터베이스에 저장되어 있는 400명 고객에 대한 114,736건의 웹로그 거래 데이터를 수집하였으며, 이탈 고객과 비이탈 고객의 비율은 40%와 60%로 구성하였다. 수집된 데이터는 전처리 과정을 통해 한달 동안의 총 사용시간, 총 사용 횟수, 에러 횟수 및 한 달 평균 사용시간을 <표 10>와 같이 정리하였다. 기간은 한 달 혹은 한 주가 될 수도 있으나, 고객의 행위 정보를 충분히 파악하고, 변화를 관찰하기 위해 한 달이 적절하다고 판단되었다. 또한 인공지능경망 분석과 본 논문에서 제시되고 있는 모형의 성능을 비교하기 위해 데이터를 학습용 데이터 60%와 검증용 데이터 40%로 나누었다.

〈표 10〉 전처리 과정을 거친 고객의 행위데이터

고객 ID	접속횟수	사용시간(분)	오류	평균 사용시간
행복나무	17	490	3	28.82
HIYSY	4	119	3	29.75
MISUN0707	5	150	1	30.00
[순순이]선영	24	736	4	30.67
天上天下	17	528	10	31.06
사신망키	10	313	3	31.30
		∴		

4.2 최적 군집 수

K 온라인 게임사의 고객을 가장 적절하게 분류하는 최적 군집 수를 결정하기 위하여 6×6, 5×7, 7×5, 6×8, 8×6 등과 같은 군집수를 임의로 정한 후, 각 경우마다 오분류표를 작성해 보았으며, 결과는 <표 11>과 같다.

앞의 분석과정 설명에서 언급했듯이, 최적 군집 수는 오류영역 1이 5%를 넘지 않는 범위 내에서 오류영역 2가 가장 적은 것을 선택하기로 하였다. 이와 같은 전제를 토대로 K 온라인 게임 제공업체의 경우에 있어서는 오류영역 1이 4.56%인, 5×7이 가장 적절한 군집 분류 방법으로 채택되었다. 5×7으로 군집을 분류하면 <표 12>와 같은 결과가 도출된다. 5×7, 총 35개의 군집 중에서, 포함 된 고객 수가 하나도 없는 군집을 제외하고 22개의 군집이 고려되었다. 또한 포함하고 있는 고객 수가 지나치게 적으면 그 군집에 대한 평균 이탈률을 신뢰할 수 없으므로, 적정수준 이하의 고객 수를 포함하고 있는 군집은 분석에서 제외시켜야 할 것이다. 군집수를 제외시켜야 할지, 포함할지는 분석 목표나 의사결정자의 결정에 의해 달라질 수 있을 것이다.

〈표 11〉 각 군집수에 따른 오분류표

실제값 \ 예측값	6×6		5×7		7×5		6×8		8×6	
	비이탈	이탈	비이탈	이탈	비이탈	이탈	비이탈	이탈	비이탈	이탈
비이탈	-	11.67	-	17.62	-	15.51	-	14.52	-	13.33
이탈	11.90	-	4.56	-	6.90	-	6.67	-	8.57	-

<표 12> 유효 군집과 그에 속한 고객 수

유효 군집명	이탈률(%)	군집에 포함된 고객 수(명)
0	74.19	31
1	57.14	8
2	45.45	11
3	100	2
4	100	1
8	78.72	47
9	37.50	15
10	25.00	8
11	16.67	8
12	100	6
15	48.84	43
16	23.33	29
17	18.18	12
18	25.00	8
20	100	13
21	100	8
22	23.81	43
23	16.00	25
24	6.67	15
26	100	11
29	36.36	22
32	100	9

<표 13> 유도 방향 및 현재 군집과 목표 군집의 비교

현재 군집	목표 군집	현 군집과 목표 군집과의 차이점
0	1	접속횟수의 증가
1	2	
2	9	평균 사용시간의 증가
3	9	
8	9	총 사용시간의 증가
9	17	
10	11	
11	24	접속 횟수 및 총 사용시간의 증가
12	11	총 사용시간의 증가
15	23	평균 사용시간의 증가
16	17	
17	24	
18	17	접속 횟수 및 총 사용시간의 증가
20	24	
21	24	총 사용시간의 증가
22	23	
23	24	총 사용시간 증가 및 사용 오류의 감소
26	24	사용자 오류의 감소
29	23	
32	24	

<표 14> 행위 유도 방안

군집간 차이점	행위 유도를 위한 캠페인
접속횟수의 증가	<ul style="list-style-type: none"> • 사이트의 수시 업데이트 및 그에 따른 소식 발송 • 새로운 게임 정보 제공 및 무료 사용 쿠폰 제공
평균 사용시간의 증가	<ul style="list-style-type: none"> • 사용시간에 따른 누적 점수 도입 • 누적 점수에 따라 상품 및 무료 쿠폰 발송
총 사용시간의 증가	<ul style="list-style-type: none"> • 저렴한 월 정액권 판매 및 할인권 증정 • 사용 누적 점수에 따라 상품 및 무료 쿠폰 발송
접속횟수 및 총 사용시간의 증가	<ul style="list-style-type: none"> • 새로운 게임 정보 제공 및 무료 사용 쿠폰 제공
사용 오류의 감소	<ul style="list-style-type: none"> • 서버 오류인지, 사용자 오류인지 분석 • 사용자가 많은 시간대의 접속에 따른 오류 발생의 경우에는 사용자가 적은 시간대에 사용할 수 있는 쿠폰 발송

4.3 고객 행위 유도 전략 수립

각 유효 군집의 이탈률과 전처리 과정을 통해 생성된 K 온라인 게임사의 고객의 네 가지 변수를 사용해 생성된 비용을 기반으로 Visual Basic 6.0 프로그램을 활용하여 현재 군집과 목표 군집을 생성하였다. 또한 의사결정나무 분석을 통해 생성된 규칙을 통해 현재 군집과 목표 군집간의 차이를 살펴보았는데, 그 결과는 <표 13>과 같이 나타났다.

각 군집의 목표 군집 및 목표 군집과의 차이를 기반으로 <표 14>에서 고객 행위의 변화를 유도하기 위한 몇 가지 방법을 제시해 보았다. <표 13>에서 군집 '0'은 수익률 분석 결과 군집 '1'로 유도하는 것이 가장 적절하다는 결론이 나왔다. 이러한 사실을 기반으로 군집 '0'과 군집 '1'의 속성을 비교해본 결과 두 군집간의 차이는 접속횟수이며,

'0'에서 '1'로 균집을 유도하기 위해서는 접속횟수를 증가시켜야 한다는 사실을 알 수 있었다. 이 때, 접속횟수를 증가시킬 수 있는 몇 가지 방법을 살펴보면, 그 중 하나가 사이트를 수시로 업데이트함으로써 고객에게 새로운 정보를 지속적으로 제공하는 방법이 있을 수 있다. 그 외에도 소식을 발송하여 새로운 게임 정보를 제공함으로써 관심을 유발시키거나, 무료 쿠폰 제공하여 게임을 경험하게 하여 지속적인 사용을 유도하는 방법 등을 고려해 볼 수 있을 것이다. <표 14>는 각 경우에 따라 수행 가능한 행위 유도 방안을 보여주고 있다.

4.4 성과 측정

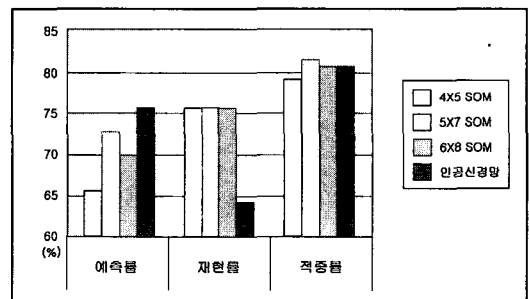
본 연구에서 제시하고 있는 이탈 방지 모형의 성과를 전체적으로 검증하기에는 어려움이 있으므로 이탈 가능성만을 검증하고자 한다. 이를 위하여, 이 절에서는 제시한 모형과 예측력이 뛰어난 것으로 알려져 있는 인공신경망과 비교하여 그 정확도에 대해 평가해 보도록 하겠다. 앞서 말했듯이 SOM 분석 방법은 분석에 앞서 분석하고자 하는 균집수를 먼저 결정하여야 한다. 만일 균집의 수를 너무 적게 결정한다면, 분석을 통해 얻고자 하는 정보를 잃을 수 있을 것이다. 다시 말해, 분석에 앞서 너무 적은 균집의 수를 설정하게 되면, 비록 집단의 이탈 가능성은 낮더라도 그 속에 포함되어 있는 고객은 높은 이탈 가능성을 내포하고 있을 수 있다는 것이다.

반면에 지나치게 적은 균집의 수를 설정하여 분석을 실시하게 되면 집단의 이탈 가능성은 그 의미를 잃게 된다. 그러므로 SOM의 분석에 앞서 적절한 균집의 수를 결정하는 것은 매우 중요하다. 앞서 최적 균집수를 결정하기 위해 바이탈 고객이 이탈하였다고 예측하는 에러를 최소화 하는 방법을 사용하였으나, 과연 균집의 수가 적절하였는지 평가하기 위해 일반적으로 많이 사용되고 있는 예측률(precision), 재현률(recall) 및 적중률(hit ratio)을 사용하였다.

정확성 평가를 위해 인공신경망 분석에 의한 결과를 비교하였는데, 그 결과는 <표 15> 및 <그림 6>에 나타나 있다. 비교 결과 몇몇의 균집 수 중에 5X7이 가장 예측 정확도가 우수하였으므로 적절하게 균집수가 결정되었음을 알 수 있다. 또한 인공신경망과 비교하였을 때, 예측의 정확성이 뒤지지 않음을 확인할 수 있었다.

<표 15> 예측 정확성 측정 비교 결과

분석기법	예측률	재현률	적중률
4×5 SOM	65.5	76	78.7
5×7 SOM	73.1	76	82.7
6×8 SOM	70.3	76	81.3
MN(MLP)	76.2	64	81.3



<그림 6> 예측 정확성 측정 비교 결과

5. 결 론

신규 고객을 획득하는 것보다 기존 고객을 유지하는데 소모되는 마케팅 비용이 더 저렴하다는 것 외에, 고객 충성도 및 캠페인의 효과성 등 부가적인 효과로 인해 기존 고객의 유지에 대한 중요성이 증대되고 있다. 그러나 인터넷환경에서는 기업 경쟁이 더욱 치열해지고 고객이 선택할 수 있는 폭이 다양해짐에 따라 고객 이탈은 더욱 빈번하게 이루어지고 있다. 본 연구는 이러한 고객의 이탈을 실시간으로 감지하여, 이탈의 원인을 발견함으로써 이탈이 발생하기 전에 고객의 행위를 유도하여 이탈을 방지할 수 있는 방법을 제시하고 있다.

제시한 방법론을 우리나라 K 온라인 게임 제공

업체의 사례를 예로 들어 설명하였으며, 회사 데이터를 이용하여 고객 이탈 가능성을 분석하였을 뿐 아니라, 이탈을 감소시키기 위하여 경제성 분석에 기반 한 캠페인 전략까지 제시하였다. 본 연구의 의의는 기존에 시행되고 있는 캠페인이 과거데이터 분석에 기반 한 일회성 캠페인이거나 변화하는 고객의 시계열적인 특성을 반영하기 어려운 것인데 반해, 정확한 원인 파악을 통해 보다 체계적이고 효율적인 캠페인을 위한 기준을 제시했다는 데 있다.

고객 행위변화 타지를 위해서는 웹로그, 구매 데이터베이스 등의 대용량의 고객 행위 데이터에 대한 분석이 요구되기 때문에 데이터 마이닝 기법의 활용이 필수적이다. 그러나, 대부분의 데이터 마이닝 연구는 예측 및 분류의 정확성이 높은 모델을 개발하는데 초점이 맞추어져 있으며, 고객의 이탈 원인을 밝힌다거나, 행위를 유도하고자 하는 연구는 지극히 부족한 상황이다. 따라서 본 연구는 이탈 가능한 고객을 사전에 예상하여 각각의 잠재적인 이탈 고객에게 개인화된 캠페인을 실시하므로 개인화된 마케팅을 수행할 수 있도록 하였다. 본 연구에서는 한 달 동안의 실제 데이터를 활용하여 새로운 고객 이탈 방지 절차를 제시함으로써, 효율적인 마케팅 캠페인을 수행하기 위한 근간을 제시하였다. 그러나 마케팅 수행 결과를 분석하고 그 효과를 실험에 반영하지 못한 한계점이 있다. 따라서 앞으로의 연구에서는 본 논문에서 제시된 이탈 방지 절차를 자동화하고, 실제 캠페인을 계획하고 실시하여, 그 효과성을 검증해 보고자 한다. 또한 효과성 검증을 위한 기준을 연구하여, 고객 행위 유도에 대해 효과가 높은 캠페인에 대해서는 정규화된 메뉴얼로 가시화 하고, 효과가 없는 캠페인에 대해서는 인터뷰나 설문조사를 통하여 고객의 요구사항을 반영함으로써 보다 실제적인 캠페인 전략을 제시하고자 한다. 본 논문에서 제시한 방법론에서는 SOM의 군집간의 거리를 사용하여 비용으로 계산하였다. 또한 군집별 고객의 이탈률을 이용하여 수익을 정의하였다. 제시한 경제성 분석의 의미가 있으려면, 이론적인 검증이 쉽지 않으므로,

다양한 방법에 의한 비용 및 수익을 정의하고 본 연구에서 제시하고 있는 수익 및 비용이 적절하다는 것을 실험적으로 검증할 필요가 있다. 본 연구에서는 온라인 게임 사이트의 사례를 통해 적용해 보았으나, 이외에도 통신 산업, 인터넷 서비스 및 콘텐츠 산업에도 적용해 볼 수 있을 것으로 기대된다.

참 고 문 헌

- [1] Beiman, L., J. Friedman, R. Olshen and C. Stone, *Classification and Regression Trees*, Wadsworth International Group, C.A, 1984.
- [2] Berson, A., S. Smith and K. Thearling, *Building data mining applications for CRM*. McGraw-Hill, 2000.
- [3] Draenos, S., *Gamers Log On*, Upside Magazine, January, 2001.
- [4] Fayyad, U., G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [5] Quinlan, J., *C4.5: Programs for Machine Learning*, Morgan Kaufmann, C.A., 1993.
- [6] Alhoniemi, E., J. Hollmen, O. Simula and J. Vesanto, "Process monitoring and modeling using the Self-Organizing Map," *Integrated CAE*, Vol.6, No.1(1999), pp.3-14.
- [7] Almuallim, H. and T. Dietterich, "Learning Boolean concepts in the presence of many irrelevant features," *Artificial Intelligence*, Vol.69(1994), pp.279-306.
- [8] Chakraborty, G., and B. Chakraborty, "A novel normalization technique for unsupervised learning in ANN," *IEEE transactions on neural networks*, Vol.11, No.1(2000), pp.253-257.
- [9] Chung, Y., W. Pottenger and B. Schatz, "Automatic Subject Indexing Using an Associative Neural Network," In *Proceedings of the ACM Conference on Digital Libraries*,

- DL-99(1998), pp.59-68.
- [10] Cooley, R., B. Mobasher and J. Srivastava, "Data preparation for mining world wide web browsing patterns," *Knowledge and information systems*, Vol.1, No.1(1999), pp.5-32.
- [11] Datta, P., B. Masand, D. Mani and B. Li, "Automated cellular modeling and prediction on a large scale," *Artificial Intelligence Review*, Vol.14(2000), pp.485-502.
- [12] Hall, M., and L. Smith, "Practical feature subset selection for machine learning," *Proceedings of the 21st Australian Computer Science Conference*, 1998, pp.181-191.
- [13] Kass, G. "An Exploratory Technique for Investigating Large Quantities of Categorical Data," *Applied Statistics*, Vol.29(1980), pp.119-127.
- [14] Kasslin, M., J. Kangas, and O. Simula, "Process state monitoring using Self-organizing maps," *Artificial Neural Networks*, Vol.2(1992), pp.1531-1534.
- [15] Kim, Y., H. Song and S. Kim, "A dynamic procedure for defection detection and prevention based on SOM and a Markov chain," *International Conference of Korea Intelligent Information Systems Societ*, 2003, pp.141-148.
- [16] Kira, K., and L.A. Rendell, "The feature selection problem : Traditional methods and a new algorithm" *Proceedings AAAI-92(1992)*, pp.129-134.
- [17] Kohavi, R. and G. John, "Wrappers for feature subset selection," *Artificial Intelligence*, Vol.97 (1997), pp.273-324.
- [18] Kohonen, T., "The Self-Organizing Map," *Proceedings of the IEEE*, Vol.78, No.9(1990), pp.1464-1480.
- [19] Kohonen, T. "Self-Organizing and Associative Memory," Berlin, Springer-Verlag, 1995.
- [20] Kohonen, T., E. Oja, O. Simula, A. Visa and J. Kangas, "Engineering applications of the Self-Organizing Map," *Proceedings of the IEEE*, Vol.84, No.10(1996), pp.1358-1384.
- [21] Loh, W. and Y. Shih, "Split selection methods for classification trees," *Statistica Sinica*, Vol.7, No.8(1997), pp.815-840.
- [22] Ng, K. and H. Liu, "Customer retention via data mining," *Artificial Intelligence Review*, Vol.14, No.6(2000), pp.569-590.
- [23] Raghavan, N., R.M. Bell, and M. Schonlau, "Defection Detection : Using Online Activity Profiles to Predict ISP Customer Vulnerability," In *Proceedings of the sixth International Conference of Knowledge Discovery and Data Mining*, 2000, pp.506-515.
- [24] Song, H., J. Kim, Y. Cho and S. Kim, "A personalized defection detection and prevention procedure based on the self-organizing map and association rule mining : Applied to online game sit," *Artificial Intelligence Review*, Vol.21, No.2(2004), pp.161-18.
- [25] Trubik, E., and M. Smith, "Developing a model of customer defection in the Australian banking industry," *Managerial Auditing Journal*, Vol.15, No.5(2000), pp.199-208.
- [26] Tryba, V. and K. Goser, "Self-organizing feature maps for process control in chemistry," *Proceedings of international conference on artificial neural networks (ICANN 91)*, 1991, pp.847-852.
- [27] Vesanto, J., "SOM-based data visualization methods," *Intelligent data analysis*, Vol.3(1999), pp.111-126.
- [28] Yeo, A., K. Smith, R. Willis and M. Brooks, "Modelling the effect of premium changes on motor insurance customer retention rates using neural networks," *Lecture notes in computer science*, Springer-Verlag, 2001, pp.390-399.