실시간 CRM을 위한 분류 기법과 연관성 규칙의 통합적 활용: 신용카드 고객 이탈 예측에 활용

이지영[a], 김종우[b]

[a] 한양대학교 경영대학원, 133-791 서울시 성동구 행당동 17
Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail: cruise2jio0@hanyang.ac.kr

[b] 교신저자, 한양대학교 경영대학, 133-791 서울시 성동구 행당동 17
Tel: +82-2-2220-1067, Fax: +82-2-2220-1169, E-mail:kjw@hanyang.ac.kr

Abstract

이탈 고객 예측은 데이터 마이닝에서 다루는 주요한 분제 중에 하나이다. 이탈 고객 예측은 일종의 분류(classification) 문제로 의사결정나무추론, 로지스틱 회귀분석, 인공신경망 등의 기법이 많이 활용되어왔다. 일반적으로 이탈 고객 예측을 위한 모델은 고객의 인구통계학적인 정보와 계약이나 거래 정보를 입력변수로 하여 이탈 여부를 목표 변수로 보는 형태로 분류 모델을 생성하게 된다. 본 연구에서는 고객과의 지속적인 접촉으로 발생되는 추가적인 사건 정보를 활용하여 연관성 규칙을 생성하고 이 결과를 기존의 방식으로 생성된 분류 모델과 결합하는 이탈 고객 예측 방법을 제시한다. 제시한 방법의 유용성을 확인하기 위해서 특정 국내 신용카드사의 실제 데이터를 활용하여 실험을 수행하였다. 실험 결과 제시된 방법이 기존의 전통적인 분류 모델에 비해서 향상된 성능을 보이는 것을 확인할 수 있었다. 제시된 예측 방법의 장점은 기존의 이탈 예측을 위한 입력 변수들 이외에 고객과 회사간의 접촉을 통해서 생성된 동적 정보들을 통합적으로 활용하여 예측 정확도를 높이고 실시간으로 이탈 확률을 갱신할 수 있다는 점이다.

Keywords:

이탈 고객 예측; 의사결정나무추론; 분류; 연관성 규칙; 실시간 CRM

1. 서론

국내 신용카드 산업은 시장의 급격한 팽창과 신용카드사간의 적극적인 경쟁의 결과로 상당히 빠른 속도로 포화상태에 이르렀다. 신용카드사간의 치열한 경쟁과 경제 생활 모든 곳곳에 신용카드 사용이 확대되는 상황 속에서 경쟁력 있는 고객 관리는 기업의 수익가치와 직결되는 문제로서 그 중요성이 더욱더 부각되고 있다. 또한 고객별로 다사(多社)카드를 소지하는 경향이 일반적인 상황에서 고객관리는 그 어느 때보다 중요하다고 할 수 있다. 고객관계관리(CRM, Customer Relationship Management)를 적극적으로 활용하기 위해서는 무엇보다도 획득한 고객 데이터를 효과적이고 즉각적으로 분석하여, 고객 관리 활동에 반영하는 것이 중요하다. 특히 경쟁 환경이 급변하고 치열해 짐에 따라 고객관계관리 중 이탈 고객을 방지함으로써 수익 가치가 높은 로열티 고객을 늘려나가는 과정이 더욱더 중요해지고 있다.

본 연구에서는 신용카드사의 고객 이탈 확률 예측력을 높이기 위한 방안을 제시하고자 한다. 기존의 고객 이탈 예측에 많이 활용되었던 정적인 정보(인구 통계학적인 속성 및 신용카드 이용 정보, 연체 정보) 기반의 분류(classification) 모델에 추가하여, 고객의 상태 변화에 대한 사건 정보인 동적 정보(연체 상태 변화, 한도 변화 등)를 기초로 연관성 규칙을 생성하고 이를 기존의 분류 모델과 통합적으로 활용하여 이탈 확률의 예측력을 높이고자 한다. 본 논문의 구성은 다음과 같다. 2장에서는 이탈 고객 예측과 관련한 기존의 연구들을 살펴본다. 3장에서는 본 연구에서 제시하는 이탈 고객 예측 방안을 제시하고, 실제 국내 신용카드사의 데이터를 활용하여 실험한 결과를 제시하도록 한다. 4장에서는 결론을 제시한다.

2. 이탈 고객 예측

2.1 신용카드사의 이탈 고객 관리의 필요성

고객 이탈 방지의 궁극적인 목적은 고객 이탈 원인을 파악하고 이를 통하여 이탈 가망 고객에 대한 관리 활동을 강화하여 이탈 고객을 최소화시킴으로써 기업의 가치를 극대화시키는데 있다고 할 수 있다. 이탈 고객의 정보를 잘 활용하면 기업 활동에 있어 잘못된 점을 파악할 수 있으며, 이러한 문제점을 적절히 개선한다면 이탈은 기업에게 있어 위협이 아닌 더 견고한 토대 위에서 고객과의 관계를 쌓아 갈 수 있는 좋은 기회가 될 수 있다. 또한 이탈 고객 분석은 신규 고객을 확보하는데 더 많은 비용이 소요되고 기존 고객의

휴면화 등으로 인해 지속적인 수익 창출이 어려운 상황에서 고객과의 관계를 돈독히 함으로써 가치가 높은 로열티 고객을 확보할 수 있으므로 수익 증대라는 기업의 목표 달성 차원에서 매우 중요하다[1]

현재 신용카드 시장에서 한 가지 카드만을 사용하는 고객은 거의 찾아보기 힘들며, 다사(多社)카드를 소지하고 있는 고객들이 대부분이다. 여기서 고객 이탈은 단순히 신용카드 시장에서의 이탈을 의미하는 것이 아니라 다른 경쟁사로의 전환을 의미한다. 고객 보유율 5% 신장이 수익률 120%의 증대를 가져오는 것으로 보고되고 있는 신용카드 산업에서는 특히 기존 고객을 유지, 관리하는 것이 매우 중요하다고 볼 수 있다[4].

2.2 데이터마이닝을 활용한 고객 이탈 예측

고객 이탈 예측에 관한 연구들은 주로 의사결정나무추론, 인공신경망, 로지스틱 회귀분석 등의 데이터마이닝 분류 기법을 활용한 고객 이탈 예측에 관하여 연구가 많이 진행이 되어왔다. 이건창 등(2001)의 연구에서는 은행고객 이탈을, 이건창 등(2002)의 연구에서는 신용카드사 고객 이탈을 C5.0, 로지스틱 회귀분석, 인공신경망을 사용하여 예측하고자 하였다[3,4]. 이훈영 등(2006)의 연구에서는 로지스틱 회귀분석을 사용하여 생명보험사의 고객 이탈을 예측하고자 하였다[6]. Chu et al.(2006)의 연구에서는 통신회사 가입자 이탈 예측에 C5.0을 사용하였다[8].

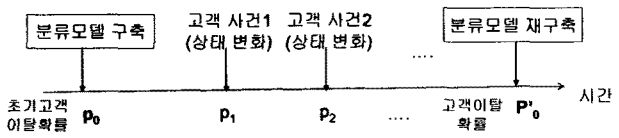
앞에 제시한 연구들이 주로 다른 분류 기법들의 성능을 비교하였다면, 다른 일부 연구에서는 다수의 분류 모델을 결합적으로 활용하는 방안들을 제시하였다. 이재식 등(2006)의 연구에서는 자동차 보험 고객 이탈 예측을 위한 다중모형 접근법을 제시하였고[5], Wei and Chiu(2002)의 연구에서는 multi-classifier class-combiner 접근법을 활용하여 무선통신 서비스 고객 이탈 예측을 꾀하였다[11]. 이밖에도 먼저 군집분석을 적용하고, 각 군집에 대하여 분류 기법을 적용하는 연구들도 행해졌는데, 이에는 Hung et al.(2006)의 연구, 김재경 외(2004)의 연구 등이 있다[2,9]. Hung et al.의 연구에서는 무선통신 서비스 고객 이탈을, 김재경 외의 연구에서는 온라인게임 고객 이탈을 다루었다.

이외에도 고객이탈 예측과 고객 이탈 방지를 위해서 다양한 기법들을 복합적으로 활용한 연구들이 존재한다. 예를 들어, NG & Liu의 연구에서는 의사결정나무 추론, deviation analysis, 연관성 규칙 등을 이용하여 통신 중계 서비스 고객 이탈 방지에 대하여 연구하였고[10], Chiang et al.(2003)의 연구에서는 이탈 원인을 찾기 위해 연관성 규칙을 사용하였다[7]. 두 연구 모두 연관성 규칙을 사용하였으나, 앞의 연구에서는 통신 서비스 이탈 기업과 함께 이탈할 기업을 찾기 위해서 연관성 규칙을 활용하였고, Chiang et al.의 연구에서는 이탈의 원인을 찾기 위해서 연관성 규칙을 사용하였으나, 본 연구에서와 같이 연관성 규칙을 활용하여 이탈 확률 정확도를 높이기 위한 연구는 미미한 형편이다.

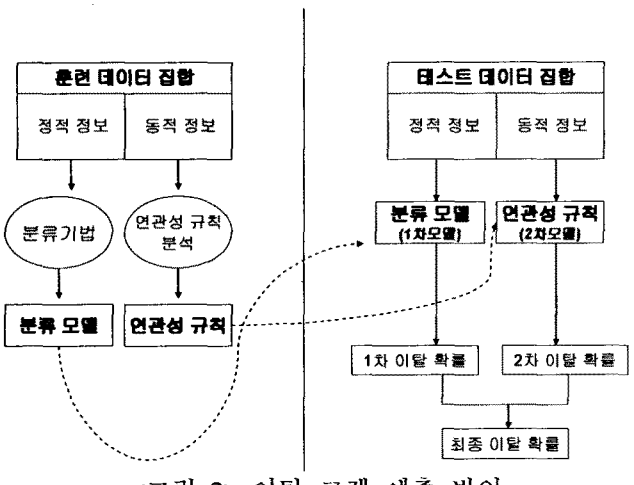
3. 통합적 이탈 고객 예측 방안

3.1 이탈 고객 예측 방안

본 연구에서 제시하는 이탈 고객 예측 방안은 일종의 다중모델이다. 하지만 기존의 대부분의 다중모델 접근법들이 동일한 데이터 집합을 기초로 하는 분류 모델들의 결합적 활용 방안(예를 들어, C5.0과 인공신경망의 결합)을 제시했던데 반해서, 본 연구에서 제시한 기법은 기존에 많이 연구되었던 데이터마이닝의 분류 모델의 예측 정확도를 향상시키기 위해서 분류 모델 생성 이후에 발생되는 고객의 동적 정보를 이용하도록 기법이 설계되었다. 즉, 기존에는 과거 데이터 집합을 활용한 분류 모델이 생성된 이후에는 특정 고객의 이탈 확률에 대한 예측값이 모델의 재생성하기 전까지는 동일하나, 본 연구에서는 초기 분류 모델 생성 이후에 발생되는 고객의 상태 변화(사건 발생)에 따라서 이 확률값을 지속적으로 보정하도록 한다(<그림 1> 참조).



<그림 1> 시간에 따른 이탈 확률의 보정



<그림 2> 이탈 고객 예측 방안

본 연구에서 제시하는 이탈 확률 예측 절차는 <그림 2>와 같다. 기존의 분류 모델을 생성하기 위해서 활용되었던 고객의 정적 정보를 사용하여

의사결정나무 추론, 로지스틱 회귀분석, 인공신경망으로 분류 모델(1차 모델)을 생성하고, 고객 이탈에 직접적으로 영향을 끼칠 수 있는 고객의 상태 변화(동적 정보)들을 변수로 선정하여, 이러한 상태 변화와 이탈 여부간의 연관성 규칙을 생성한다(2차 모델). 1차 모델의 예측결과와 2차 모델의 예측결과를 종합하여 최종으로 고객 이탈을 예측한다.

3.2 실험

3.2.1 이탈 기준의 정의

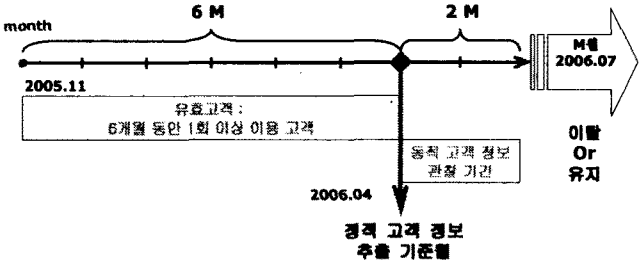
이탈 고객을 예측하기 위해서 가장 먼저 해야 할 일은 이탈 고객의 기준을 설정하는 것이다. 즉, 고객의 어떤 상태를 이탈로 볼 것인가에 대한 정의를 내려야 한다. 본 연구는 단순히 신용카드 고객의 이탈 분석이 아니라, 로열티 고객의 이탈 확률을 예측하고자 하는 것이 목적이다. 따라서, 유효/유지/이탈 고객의 정의에서 제시하듯이 모집단을 최근 6개월(2005년 11월 ~ 2006년 4월) 동안 신용카드 이용 실적이 있으며, 신용카드 이용 한도가 0보다 큰 개인 신용카드 고객을 유효 고객이라 정의하였다. 이러한 유효 고객에서 예측하고자 하는 기준월(2006년 7월)에 고객이 이탈을 했느냐 안 했느냐에 따라 이탈 고객과 유지 고객으로 구분하였다.

<표 1> 분석 표본 집단 선정

구 분	내 용		
분석 모집단 (유효 고객)	개인 고객 & 이용 한도 > 0 & 가입일 <= 2005년 11월 & 2005.11 ~ 2006.04 까지 1회 이상 신용카드 이용 고객		
표본추출법	임의 추출 방식		
분석 표본집단	이탈 고객	2006.07 이탈한 고객	9,102명
	유지 고객	2006.07 유지한 고객	14,000명

3.2.2 데이터 준비

이탈 기준에 맞춰 모집단 유효 고객 중에서 임의 추출 방식을 이용하여 이탈 고객과 유지 고객간의 비율을 2:3으로 하여 유지 고객 14,000명, 이탈 고객 9,102명을 추출하였다. 유지 고객과 이탈 고객의 비율은 유효 고객에서 이탈 고객의 비율이 매우 낮기 때문에 동일 비율이 아닌 2:3의 비율로 추출하였다. 훈련 데이터 집합과 테스트 데이터 집합은 임의로 50%씩 분할하였다. 실험 데이터 집합에 대한 상세한 내용은 <표 1>과 같다.
표본 집단을 추출한 후 고객의 정적 정보와 동적 정보를 추출하기 위하여 데이터 추출시점을 구분하였다. <그림 3>에서 보는 바와 같이 고객의 정적 정보 중에서 고객의 인구통계학적인 정보 및 신용카드 이용한도, 고객 등급에 관한 데이터는 2006년 4월 기준 데이터이며, 고객의 신용카드 이용정보 및 연체정보는 2005년 11월부터 2006년 4월까지 총 6개월간의 관찰기간을 갖는다. 고객 동적 정보는 이탈 또는 유지 전 2개월간의 고객 행동에 대한 사건 정보로서 연체 변화, 이용형태 변화, 고객등급 변화 등의 변수를 선정하였다.



<그림 3> 데이터 추출 시점

<표 2>는 본 연구에서 선정한 정적 정보들의 목록으로, 인구통계학적 정보 12항목, 신용카드 이용 정보 9항목, 신용카드 이용한도, 연체정보 6항목, 고객등급 관련 항목 4개, 총 32개의 변수를 설정하였다. <표 3>은 이탈과 관련 있을 것으로 예상되는 고객의 상태 변화, 즉 사건 데이터로, 2차모델 구축에 활용되었다.

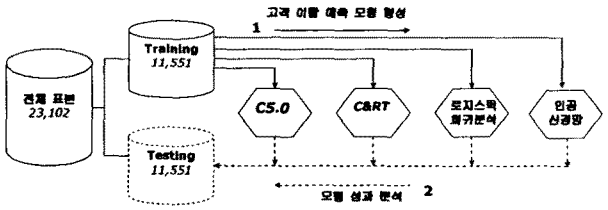
<표 2> 1차 모델 변수

NO	변 수 명	변 수 설 명
1	성별	1: 남 2: 여
2	나이	만 20세 ~69세로 한정, 그룹화하지 않음
3	결혼여부	1: 기혼 2: 미혼 3: 무응답
4	가입후경과	가입일로부터 현재까지 경과일수
5	가족카드	1: 가족카드발급 2: 가족카드 미발급
6	직업등급	Prime, Super Prime 등5개의 등급으로 분류
7	직군코드	전문직, 공무원, 학생 등 23개로 분류
8	자택주소	서울특별시, 도, 광역시 14개로 분류
9	직장주소	서울특별시, 도, 광역시 14개로 분류
10	근속년수	직장 근속년수 (카드 가입신청 시 고객이 직접 기입)
11	주택형태	1: 아파트 2: 단독 3: 고급빌라 4: 연립(다세대) 5: 오피스텔 6: 기타 7: 무응답
12	대금청구지	1: 직장 2: 자택 3: 기타
13	총 한도	2006년 4월 기준 신용카드 총 이용 한도
14	일시불(6M)	2005.11~2006.04, 6개월 일시불 금액
15	할부(6M)	2005.11~2006.04, 6개월 할부 금액
16	현금(6M)	2005.11~2006.04, 6개월 현금 금액

17	총 이용액(6M)	2005.11~2006.04, 6개월 총 이용금액
18	월평균일시불 이용액	6개월 일시불 이용금액의 월평균 이용금액
19	월평균할부이 용액	6개월 할부 이용금액의 월평균 이용금액
20	월평균현금이 용액	6개월 현금 이용금액의 월평균 이용금액
21	월평균총이용 액	6개월 총 이용금액의 월평균 이용금액
22	이용형태	1: 일시불+할부+현금서비스 모두 이용 2: 현금서비스만 이용 3: 할부만 이용 4: 일시불만 이용 5: 일시불+할부만 이용 6: 기타 (다른 조합)
23	연체횟수	2005.11~2006.04 연체한 횟수, 0~6까지
24	최근3M연체횟 수	2006.02~2006.04 연체한 횟수, 0~3까지
25	최근1M연체여 부	1: 2006년 4월 연체, 0: 2006년 4월 미연체
26	연체금액	2005.11~2006.04까지 연체 금액 합계
27	최근3M연체금 액	2006.02~2006.04까지 연체 금액 합계
28	최근1M연체금 액	2006년 4월 연체금액
29	리스크등급	1~15등급까지
30	할부수수료등 급	1~7등급까지
31	CA수수료등급	1~12등급까지
32	론등급	1~16등급까지

<표 3> 2차 모델 변수

NO	변수명	변수설명
1	한도상승경험 여부	정적 정보 추출월 = M월(2005.04) 이후 이탈 예측월(2006.07) 사이 즉, 동적정보 관찰기간 2M동안 1회 이상 이용한도 상승 여부
2	한도하락경험 여부	2M동안 1회 이상 이용한도 하락 여부
3	한도계속상승 여부	2M동안 이용한도 계속 상승
4	한도계속하락 여부	2M동안 이용한도 계속 하락
5	월평균일시불 이용금액_상승 여부	M월 이후 월평균 일시불 이용금액 1회 이상 상승여부
6	월평균할부이 용금액_상승여 부	M월 이후 월평균 할부 이용금액 1회 이상 상승여부
7	월평균현금이 용금액_상승여 부	M월 이후 월평균 현금 이용금액 1회 이상 상승여부
8	월평균총이용 금액_상승여부	M월 이후 월평균 총 이용금액 1회 이상 상승여부
9	월평균일시불 이용금액_하락 여부	M월 이후 월평균 일시불 이용금액 1회 이상 하락여부
10	월평균할부이 용금액_하락여 부	M월 이후 월평균 할부 이용금액 1회 이상 하락여부
11	월평균현금이 용금액_하락여 부	M월 이후 월평균 현금 이용금액 1회 이상 하락여부
12	월평균총이용 금액_하락여부	M월 이후 월평균 총 이용금액 1회 이상 하락여부
13	일시불이용금 액_0	M월 이후 일시불 이용금액 0
14	할부이용금액_ 0	M월 이후 할부 이용금액 0
15	현금이용금액_ 0	M월 이후 현금 이용금액 0
16	총이용금액_0	M월 이후 총 이용금액 0
17	연체연속여부	M월 연체고객 -> M+1 연체 유 **and** M+2 연체 유
18	재연체여부	M월 연체고객 -> M+1 연체 유 **or** M+2 연체 유
19	연체단절여부	M월 연체고객 -> M+1 연체 무 **and** M+2 연체 무
20	리스크등급 하락여부	M월 이후 리스크등급 1회 이상 하락여부
21	할부수수료등 급 하락여부	M월 이후 할부수수료등급 1회 이상 하락여부
22	CA수수료등급 하락여부	M월 이후 CA수수료등급 1회 이상 하락여부
23	론등급 하락여부	M월 이후 론등급 1회 이상 하락여부
24	리스크등급 상승여부	M월 이후 리스크등급 1회이상 상승여부
25	할부수수료등 급 상승여부	M월 이후 할부수수료등급 1회이상 상승여부
26	CA수수료등급 상승여부	M월 이후 CA수수료등급 1회이상 상승여부
27	론등급 상승여부	M월 이후 론등급 1회 이상 상승여부



<그림 4> 1차 이탈 고객 예측 모델 설계

3.2.3 정적 정보를 이용한 1차 이탈 고객 예측 모델

1차 모델 생성에는 데이터마이닝 기법 중에서 의사결정나무(C5.0과 CART) 기법과 로지스틱 회귀분석, 인공신경망을 사용하였다(<그림 4> 참조). 데이터마이닝 도구는 SPSS의 클레멘타인을 활용하였다. 4가지 1차 이탈 고객 예측 모델 예측 정확도는 <표 4>와 같다. <표 4>에서 알 수 있듯이 인공신경망이 전체 예측오류율과 소수집단(이탈집단) 예측오류율이 가장 낮은 것으로 나타났다. 또한 나머지 3가지 모델, C5.0, CART와

로지스틱 회귀분석은 예측오류율, 소수집단 예측오류율이 유사하게 나타났으며, 근사하게나마 C5.0의 오류율이 낮은 것으로 나타났다.

<표 4> 1차 단일 모형간의 성과 분석 비교(단위: %)

데이터마이닝 기법	C5.0	CART	로지스틱 회귀분석	인공 신경망
예측오류율	28.85	28.97	28.91	24.09
소수집단예측오류율	52.89	53.09	52.98	41.40

3.2.3 동적 정보를 이용한 2차 이탈 고객 예측 모델

2차 모델 생성에는 연관성 규칙 분석을 위한 Apriori 알고리즘을 이용하여 연관성 규칙을 생성하였다. 즉, 연관성 규칙 분석 알고리즘에서는 'IF 조건, THEN 결론' 형태의 연관성 규칙이 생성되는데, 이탈여부를 '결론'으로 하는 연관성 규칙을 생성하고, 이들 연관성들을 일반적인 연관성 규칙 평가 기준인 지지도(Support), 신뢰도(Confidence), 향상도(Lift)를 기준으로 평가하였다. 평가결과, 신뢰도 기준으로 90% 이상인 상위 네 개의 연관성 규칙을 추출하였다.
추출된 연관성 규칙을 바탕으로 이탈 예측 정확도를 테스트 데이터 집합을 통해서 확인한 결과는 <표 5>와 같다. <표 5>의 빈도를 기초로 2차 예측 모델의 예측오류율은 26.12 % (= (208+2,812)/11,551), 소수집단예측오류율은 61.78% (=2,812/4,551)로 전체적인 예측오류율은 인공신경망을 제외한 나머지 1차 모델보다 좋으나, 소수집단예측오류율에서는 가장 떨어지는 것으로 나타났다.

<표 5> 2차 모델의 교차표

		예측값		
	빈도	유지	이탈	합계
실제값	유지	6,792	208	7,000
	이탈	2,812	1,739	4,551
	합계	9,604	1,947	11,551

3.2.4 최종 이탈 고객 예측 모델

본 연구에서는 최종 모델을 2개 생성하였는데, 1차 모델 중 가장 성과가 좋은 인공신경망과 2차 모델을 결합하여 하나의 최종 모델을 생성하였고, 1차 모델 중 두 번째로 성과가 좋은 C5.0과 2차 모델을 결합하여 또 하나의 최종 모델을 생성하였다(<그림 5> 참조). 1차 모델에서 만들어진 이탈 확률을 2차 모델의 이탈 확률로 보정하기 위해서 다음과 같은 식을 사용하였다. 즉,

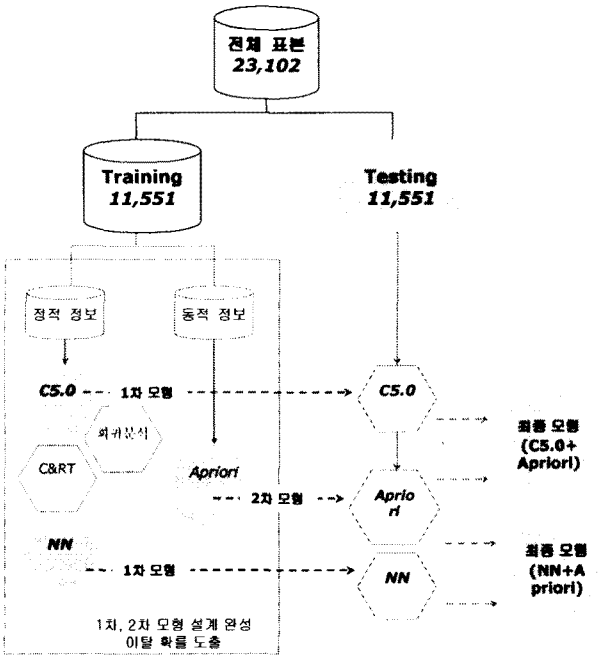
$P_1(c)$ = c고객에 대한, 1차 이탈 고객 예측 모델 결과로 도출된 이탈 확률

$P_2(c)$ = c고객에 대한, 2차 이탈 고객 예측 모형 결과로 도출된 이탈 확률
이라면, 최종 이탈 확률 $P(c)$는 다음과 같이 예측된다.

① $P_2(c) > 0.5$ 인 경우(이탈로 예측한 경우),
$P(c) = P_1(c) + (1- P_1(c)) \times P_2(c)$ (식 1)

② $P_2(c) < 0.5$ 인 경우(유지로 예측한 경우),
$P(c) = P_1(c) - P_1(c) \times P_2(c)$ (식 2)

③ $P_2(c) = 0.5$ 인 경우.
$P(c) = P_1(c)$ (식 3)



<그림 5> 최종 이탈 고객 예측 모델

(식 1)은 $P_2(c)$ 가 0.5 초과인 경우로, 2차 모델에서 이탈로 예측한 경우이다. 이 경우에는 1차 모델의 이탈 확률에 $(1 - P_1(c)) \times P_2(c)$를 더해서 확률을 증가시킨다. (식 2)의 경우는 2차 모델에서 유지로 예측한 경우이므로, 1차 모형의 이탈 확률에서 $P_1(c) \times P_2(c)$를 빼줌으로써 이탈 확률을 보정하고 있다. (식 1), (식 2), (식 3)을 통해서 계산된 최종 모델의 이탈 확률 $P(c)$가 0.5 이상인 경우, 이탈로 예측하고 그렇지 않은 경우 유지로 예측한다.

3.4 실험 결과 분석

본 연구의 최종 모델의 예측오류율은 <표 6>과 같다. (C5.0+연관성) 최종 모델의 경우, 예측오류율 25.36%, 소수집단예측오류율 41.33%로, 기존의 1차 모델, 2차 모델보다 모두 향상되었다. (인공신경망+연관성) 최종 모델의 경우는 전체적인 예측오류율은 미세하게나마 감소했지만,

소수집단예측오류율은 41.40%에서 26.81%로 대폭 상승했음을 알 수 있다. 전체 예측오류율이 일부 감소한 것은 2차 모델을 통해서 이탈 확률을 상향시키는 경우가 하향시키는 경우보다 많기 때문으로 생각된다.

<표 6> 1차 모델, 2차 모델, 최종 모델의 성과 비교

모델	예측 오류율	소수집단 예측오류율
1차 모형 - C5.0	28.84%	52.88%
1차 모형 - CART	28.96%	53.08%
1차 모형 - 로지스틱 회귀분석	28.91%	52.97%
1차 모형 - 인공신경망	24.09%	41.40%
2차 모형 - 연관성	26.14%	61.78%
최종모형 - C5.0+연관성	25.35%	41.31%
최종모형 - 인공신경망+연관성	26.56%	26.81%

본 연구가 가지는 의의는 다음과 같다. 기존의 고객 이탈 예측에 활용되던 데이터마이닝 분류 모델들이 특정 시점에서 과거 데이터를 활용해서 일괄처리(batch) 형태로 생성되는데, 이 경우는 미래 시점에서 새로운 데이터 집합으로 다시 학습되기 전까지는 특정 고객에 대하여 동일한 예측 확률을 제공한다. 본 연구에서는 분류 모델 생성 이후에 발생하는 고객의 상태 변화로부터 이탈 예측 확률을 보정할 수 있는 방안을 제시하였고, 이러한 방안의 유용성을 신용카드사의 고객 이탈 데이터를 통해서 확인하였다. 고객의 상태 변화에 따라서 민첩하게 대응하는 할 수 있는 실시간 CRM의 중요성이 강조되고 있지만, 실시간 데이터마이닝 모델 갱신이나 수정에 대한 연구는 아직 미진한 형편이다. 본 연구에서는 데이터마이닝 분류 모델의 예측 확률을 고객의 동적 정보를 기초로 한 연관성 규칙 분석 기법을 활용해서 보정할 수 있음을 실제 데이터를 통해서 확인했다는데 그 의의가 있다.

4. 결론

고객 이탈을 예측하기 위해서 의사결정나무추론, 로지스틱 회귀분석, 인공신경망 등의 데이터마이닝 분류 기법들이 많이 사용되었다. 이러한 분류 기법들은 고객의 정적 정보, 즉, 인구통계학적인 정보, 신용카드 이용 정보, 고객 신용 정보 등을 바탕으로 기계학습을 통해서 분류 모델을 생성한다. 본 연구에서는 분류 기반 예측 모델 생성 이후에 발생된 고객의 동적 정보, 즉 상태 변화 정보에 근거한 연관성 규칙을 통해서, 이탈 확률을 보다 정확하게 보정하는 방안을 제시하였다. 또한 실제 국내 한 신용카드사의 데이터를 통해서 제시한 방법의 유용성을 확인하였다. 본 연구에서 제시된 방법은 고객 이탈 예측 모델의 실시간 갱신을 가능하게 함으로써, 고객의 상태 변화에 민감하게 반응할 수 있는 기반을 제공한다.

추후 연구 과제는 다음과 같다. 먼저 본 연구에서 제시한 통합적 활용 방안에 대한 보다 정교화가 필요하다. 예를 들어, 1차 모델의 예측 확률과 2차 모델의 예측 확률을 결합하는 방법을 본 연구에서 제시했는데, 다양한 다른 방법들을 적용하여 비교하는 것이 필요하다. 두 번째로 연관성 분석을 위해서 대표적인 Apriori 알고리즘을 사용하였는데, 사건 연관성 패턴을 분석하는 다른 방법들을 적용해서 성능을 비교하는 것이 필요하다. 또한 고객 이탈 문제가 발생하는 다른 영역에 추가적인 적용이 필요하다. 예를 들어, 인터넷 서비스 고객 이탈 등에 적용해보는 것도 흥미로운 연구 주제가 될 것으로 생각된다.

참고문헌

[1] 김상용, 송지연, 이기순 (2005). "CRM 고객데이터 분석을 통한 이탈고객 연구," 한국마케팅학회 한국마케팅저널.

[2] 김재경, 채경희, 송희석 (2004), "SOM을 이용한 온라인 게임 제공업체의 고객이탈방지 방법론," 경영과학, 제21권, 제3호, pp. 85-99.

[3] 이건창, 권순재, 신경식 (2001), "은행고객 세분화를 통한 이탈고객 관리분석-가계성 예금을 중심으로" 한국지능정보시스템학회논문지, 제7권, 제1호, pp. 177-197.

[4] 이건창, 정남호, 신경식 (2002), "신용카드 시장에서 데이터 마이닝을 이용한 이탈고객분석," 한국지능정보시스템학회논문지, 제8권, 제2호, pp. 15-35.

[5] 이재식, 이진천 (2006), "다중모델을 이용한 자동차 보험 고객의 이탈예측," 한국지능정보시스템학회논문지, 제12권, 제2호, pp. 167-183.

[6] 이훈영, 양주환, 류치환 (2006), "고객의 이탈 가능성과 LTV를 이용한 고객등급화 모형개발에 관한 연구," 한국지능정보시스템학회논문지, 제12권, 제4호, pp. 109-126.

[7] Chiang, D., Y. Wang, S. Lee, C. Lin (2003), "Goal-oriented Sequential Pattern for Network Banking Churn Analysis," Expert Systems with Applications, Vol. 25, pp.293-302.

[8] Chu, B., M. Tsai, C. Ho (2006), "Toward a Hybrid Data Mining Model for Customer Retention," Knowledge-Based Systems (Article in Press).

[9] Hung, S., D.C. Yen, H. Wang (2006), "Applying Data Mining to Telecom Churn Management," Expert Systems with Applications, Vo. 31, pp. 515-524.

[10] NG, K and H. Liu (2000), "Customer Retention via Data Mining," Artificial Intelligence Review, Vol. 14, pp.569-590.

[11] Wei, C.P. and I.T. Chiu (2002). "Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach," Expert Systems with Applications, Vol. 23, pp. 103-112.