

고객의 이탈 가능성과 LTV를 이용한 고객등급화 모형개발에 관한 연구*

이훈영
경희대학교 경영학과
(hylee@khu.ac.kr)

양주환
경희사이버대학교 벤처농업경영학과
(seriyjh@empal.com)

류치훈
경희대학교 경영학과
(chihuni04@hanmail.net)

성공적인 고객관계관리(CRM : customer relationship management)를 수행하기 위해서는 효과적인 고객 등급화가 필요하다. 일반적으로 고객등급화는 고객별로 LTV를 산정한 다음 일정한 비율로 고객을 분류하여 등급을 정하는 방법이 사용되어 왔다. 그러나 이러한 방법은 등급간의 이질성을 명확하게 반영하지 못하기 때문에 적지 않은 문제점을 내포하고 있다. 본 논문에서는 Holistic Profit을 이용해서 고객을 등급화 하는 방법을 제시하고, A 생명보험회사의 고객자료를 이용해서 이를 검증하였다. Holistic Profit은 신용대출 승인정책에서 승인입계점수(Cutoff Point) 책정에 활용되고 있는 방법들 중의 하나이다. 요약하면, 본 논문의 목적은 Holistic Profit을 활용하여 보다 효과적이고 과학적인 방법으로 고객 등급화 하는 방법의 개발과 검증에 있다. 본 논문에서 제시된 방법을 사용해서 고객을 등급화 함으로써 기업은 보다 효과적인 고객관계관리(CRM)와 마케팅 활동을 수행할 수 있을 것으로 기대된다.

논문접수일 : 2006년 09월 게재확정일 : 2006년 12월 교신저자 : 이훈영

1. 서론

기업들이 성공적인 고객관계관리(CRM : customer relationship management)를 수행하기 위해서는 고객의 등급화가 반드시 선행되어야 한다. 등급에 의해 분류된 고객들에 대해서는 등급별로 상이한 마케팅전략이 적용되어야 하며, 기업은 이를 통해 경쟁우위를 확보할 수 있다. 일반적으로 고객의 등급화는 고객평생가치(LTV : Lifetime value)를 기준으로 이루어지고 있다. 그러나 고객의 평생가치를 척도로 하여 등급을 분류할 때 사용하는 분류 기준값을 찾는 것은 쉽지 않다. 일반적인

방법으로는 고객들의 LTV를 산정한 후 그에 따른 고객의 분포도를 산출하고 도출된 분포정보를 이용하여 일정한 비율로 고객을 분류하고 등급을 정하는 단순한 방법이 있다.

그러나 이렇게 단순하게 적용되는 일정한 비율에 따른 고객등급화는 등급별로 고객가치의 차이가 명확하지 않다는 문제점을 내포하고 있다. 즉, 등급 내 고객들의 동질성 혹은 등급간 고객들 간의 이질성에 대한 불분명한 분류가 문제점으로 제시되어 왔다.

본 연구의 목적은 기업이 고객을 등급화 하는 방법에 있어 기존의 문제점을 해결하고 보다 효율

* 본 논문은 2003년도 경희대학교 교내 연구비지원에 의하여 연구되었음(khu-20030326).

적이고 과학적인 방법으로 고객의 등급을 구분하는 분류 기준값을 찾는 데에 있다. 본 논문에서는 기존의 신용대출승인정책에서 적용되고 있는 방법인 Holistic Profit을 등급화에 적용하여 고객의 등급을 구분하는 분류 기준값을 산출한 다음 그 효과성을 검증하였다.

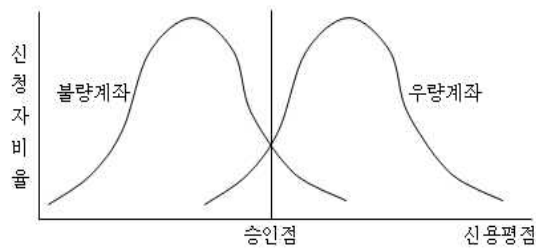
본 논문의 구성은 다음과 같다. 제 1장의 서론에 이어 2장에서는 신용평점을 이용한 대출승인 분야에서 활용되고 있는 Holistic Profit의 개념을 설명하고 이에 대한 기존의 연구에 대하여 논한다. 제 3장에서는 Holistic Profit 개념을 적용하여 어떻게 고객 등급화를 할 것인가에 대한 접근방법을 살펴본 후, 제 4장에서는 실증분석을 보충고객 데이터를 이용하여 LTV를 산출하고 Holistic Profit을 이용하여 등급 분류 값을 구하고 고객을 등급화 하는 과정을 실증분석 하였다. 제 5장에서는 단순히 일정한 비율로 고객을 분류하고 등급을 구분하는 방법과 Holistic Profit을 이용하여 등급을 구분하는 방법을 비교하여 어느 방법이 고객등급화에 유용한가를 분석하였다. 마지막으로 결론에서는 본 연구의 결과를 요약하면서 연구의 한계와 향후 연구 과제를 제시하였다.

2. Holistic Profit의 개념

일반적으로 금융기관들은 대출을 희망하는 고객의 위험을 측정하기 위해 개인의 속성을 기초로 미리 작성된 신용 평점표를 이용하여 신용평점을 산출한다. 즉, 평가대상자의 속성이나 속성수준별로 할당되는 신용점수와 이에 대한 가중치를 설정한 신용 평점표를 이용하여, 개인별로 속성에 따른 점수와 가중치를 적용해서 구한 합계 점수가

특정한 개인의 신용평점이다.

이렇게 해서 대출신청자에 대한 신용평점이 산출되면 사전에 정해진 대출 승인점과 비교한 후, 대출여부를 결정하게 된다. 즉, 미리 지정된 대출 승인점과 같거나 크면 대출이 승인되고 작으면 대출은 기각된다는 것이다. 가상적인 우량과 불량 고객의 분포와 그에 따른 승인점을 그래프로 나타내면 [그림 1]과 같다. 효과적인 신용평점시스템은 고객의 리스크를 정확하게 반영하는 평점산출 모형과 승인점 판단 방법을 갖추어야 한다.



[그림 1] 우량/불량계좌와 대출 승인점

일반적으로 신용평점시스템의 효과성을 측정하는 방법으로 흔히 사용되는 전통적인 척도들로는 변별도(Divergence), KS(Kolmogorov Smirnov) 통계량 그리고 트레이드오프 곡선(Trade-off Curve) 등이 있다(Bruce Hoadley, 1996). 정보이론에서 파생된 변별도 또는 정보가치(Information Value) 척도는 신용평점이 어떻게 우량과 불량집단을 잘 분리하는가에 대한 통계적인 척도이다. KS 통계량은 우량과 불량 분류에 대한 비모수통계 추정방법이며, 또한 트레이드오프 곡선은 그래프적 접근방법으로 신용평점 값에 따라 누적우량분포와 누적불량분포를 그래프 상에 표시하는 방법이다. 이러한 트레이드오프 곡선은 ROC (Receiver Operating Characteristic) 곡선으로도

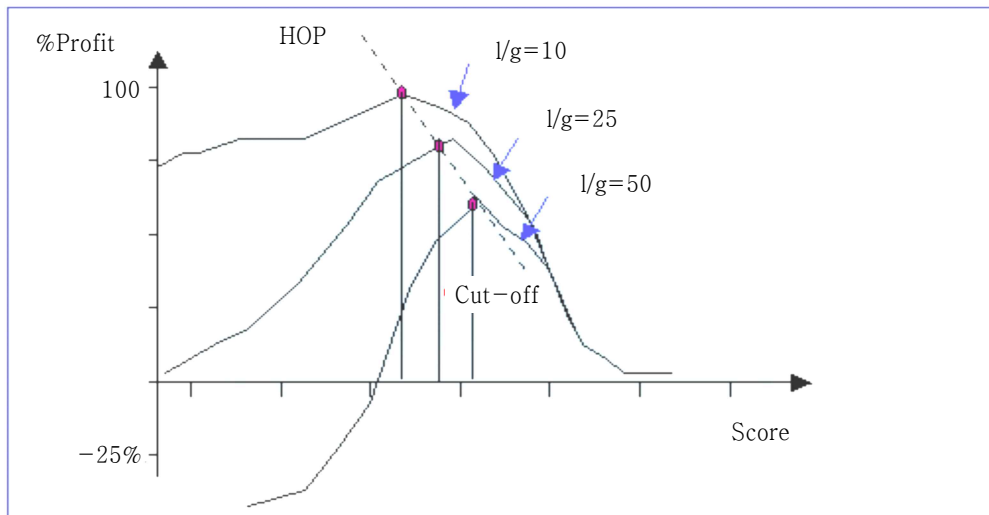
불려진다.

이러한 전통적인 접근방법은 유용한 척도이지만, 신용평점의 성능에 대한 효과적인 척도로서는 부족하다. 따라서 이를 보완하는 새로운 형태의 측정도구가 개발되었는데 이것이 Holistic Profit 곡선과 Scan(Hoadley B. and Oliver RM, 1998)이다. Holistic Profit 곡선과 각 곡선상의 극대점을 연결한 선, 즉 Scan은 경제적 모수(Economic Parameter)에 대한 정보 없이 신용평점 의사결정 모델에 적용해서 계산되고 해석될 수 있다는 장점을 가지고 있다. 즉, Holistic Profit을 이용한 방법은 경제적 모수인 이득과 손실을 별도로 추정할 필요가 없다. 왜냐하면 경제적 모수를 고려하지 않아도 되는 상대적인 이익인 Holistic Profit의 개념을 도입하였기 때문이다. Holistic Profit은 완전정보 하에서의 이익과 불완전한 정보 하에서

의 기대이익의 비율이다. 따라서 모든 손실과 이득의 비율(Loss/Gain)값들에 대해 신용평점을 독립변수로 하는 함수인 Holistic Profit 곡선이 존재한다.

[그림 2]에서 각각의 Holistic Profit 곡선은 손실과 이득의 비율이 추정되거나 가정된 상태에서 X축으로 나타나는 각각의 승인점수(score)에 따른 예상 기대수익, 즉 불완전정보 하에서의 기대수익과 완전정보 하에서의 수익의 비율을 나타내고 있다. 또한 Holistic Profit Scan은 Holistic Profit 곡선의 극대점을 연결한 선이 된다. 따라서 대출 의사결정의 기준이 되는 점수, 즉 기준점(cutting score)은 손실과 이득의 비율에 준하여 이루어져야만 기대수익을 극대화 할 수 있다.

또한 일반적으로 대출승인점이 낮아지면 대출 계좌 수가 증가하여 이익이 증가하나 신용평점이



자료 : INTERACT96, Fair and Isaac Forum(MAY 12~15, 1996)

[그림 2] Holistic Profit 곡선의 형태

낮은 고객에 대한 대출을 승인함으로써 손실이 발생할 가능성도 높아진다. 따라서 대출관리자의 입장에서 서로 상충되는 목적(기대이익과 기대손실)과 신용위험도 등을 고려하여 승인정책을 수립하여야 한다(Oliver and Wells, 2001).

기대이익을 산출하는 문제는 정확하게 알려지지 않은 경제적 모수인 이득(g)과 손실(l)의 경제적 비율($\omega^*=l/g$)에 의존한다. 이득과 손실에 대한 값을 명확히 알고 있지 않을 경우, Holistic Profit(HP)이라는 상대적 이익을 측정할 적도를 사용할 수 있다. 완전한 정보 하에서의 기대이익은 $g(\text{이득}) * P_G(\text{대출계좌 중 우량계좌 비율})$ 이다. 왜냐하면, 완전한 정보 하에서 모든 우량계좌들만을 승인하고 그 결과로 각각의 계좌로부터 이득(g)을 얻을 수 있기 때문이다. Holistic Profit을 산정하는 방법을 식으로 나타내면 [그림 3]과 같다.

[그림 3]의 식에서 볼 수 있듯이, Holistic profit은 완전정보 하에서의 기대이익과 불완전정보

하에서의 기대이익간의 비율을 말한다. 식에서 S는 신용 평점표를 이용하여 산출된 신용평점이며, 분모의 $g * P_G$ 는 완전정보 하에서의 기대이익으로 기업이 고객들 중에 우량(good)고객과 불량(bad)고객을 확실하게 구분할 수 있는 경우의 수익이다. 이는 전체 Gain중에 우량(good)고객이 가져다주는 gain의 portion이 전체의 gain 중에 얼마나 차지하는지를 나타내는 지수이며 $g * P_G = \text{gain} * (\text{우량고객 수} / \text{전체고객 수})$ 로 나타낼 수 있다. 반면에 분자는 불완전한 정보, 즉 불량고객이 발생할 수 있다는 가정하의 기대수익을 말한다. 식에서 F_G 는 우량(good)고객집단의 누적분포 함수를 나타내고 F_B 는 불량(bad)고객집단의 누적분포 함수를 나타낸다. ω^* 는 손실과 이익의 비율로 한번의 손실(loss)이 포함하는 수익(gain)의 횟수로 설명된다. ω_0 는 모집단의 오즈비로 P_G / P_B 를 나타낸다. P_G 는 전체고객 중에서 우량(good)고객이 차지하는 비율이고 P_B 는 전체고객 중에서 불량

HolisticProfit=기대이익/완전정보하에서의이익

$$H(s|\omega^*) = \frac{\left[g \times P_G \times \{1 - F_G(s)\} \times \left(\frac{\omega^*}{\omega_0} \right) \times \{1 - F_B(s)\} \right]}{g \times P_G}$$

s : 신용평점(cutting score)
 g : 계좌가 승인되고 우량일 경우의 이득
 l : 계좌가 승인되고 불량일 경우의 손실
 P_G : 대출포트폴리오내의 우량비율
 P_B : 대출포트폴리오내의 불량비율
 ω_0 : $P_G / P_B =$ 모집단비율(Odds)
 ω^* : loss/gain
 F_G : 우량누적확률분포함수
 F_B : 불량누적확률분포함수

[그림 3] Holistic Profit의 산출식

(bad)고객이 차지하는 비율이다.

3. Holistic Profit을 이용한 고객등급화

개별고객의 LTV(Lifetime Value)를 기반으로 고객을 등급화 하기 위해 본 연구에서는 Holistic

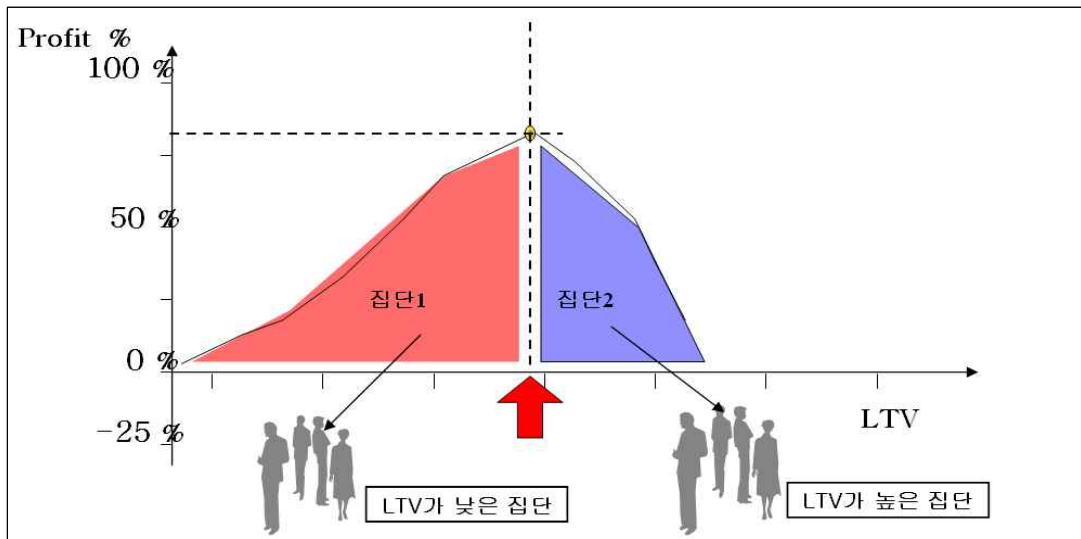
Profit 곡선그래프에서 사용되는 독립변수인 신용 평점(score) 대신 고객별 LTV(Lifetime Value)를 적용하였으며, 고객의 우량 및 불량 여부는 고객 이탈모형에 의해 도출된 고객이 이탈할 가능성을 나타내는 이탈율을 이용하였다. 고객의 LTV를 이용하여 Holistic Profit을 정의하면 다음과 같이 정의할 수 있다.

Holistic Profit = 기대이익 / 완전 정보 하에서의 이익

$$H(v|w^*) = \frac{g \times P_G \times \{1 - F_G(v)\} \times \left(\frac{w^*}{w_0}\right) \times \{1 - F_B(v)\}}{g \times P_G}$$

v : 고객별 LTV
 $F_G(v)$: 유지고객집단의 LTV 누적분포함수
 $F_B(v)$: 이탈고객집단의 LTV 누적분포함수
 w^* : $1/g$ (l : 고객이탈 손실, g : 고객유지 이득)
 w_0 : P_G/P_B =모집단비율(Odds)
 P_G : 고객집단내의 유지비율
 P_B : 고객집단내의 이탈비율

[그림 4] 고객의 LTV와 적용한 Holistic Profit의 정의



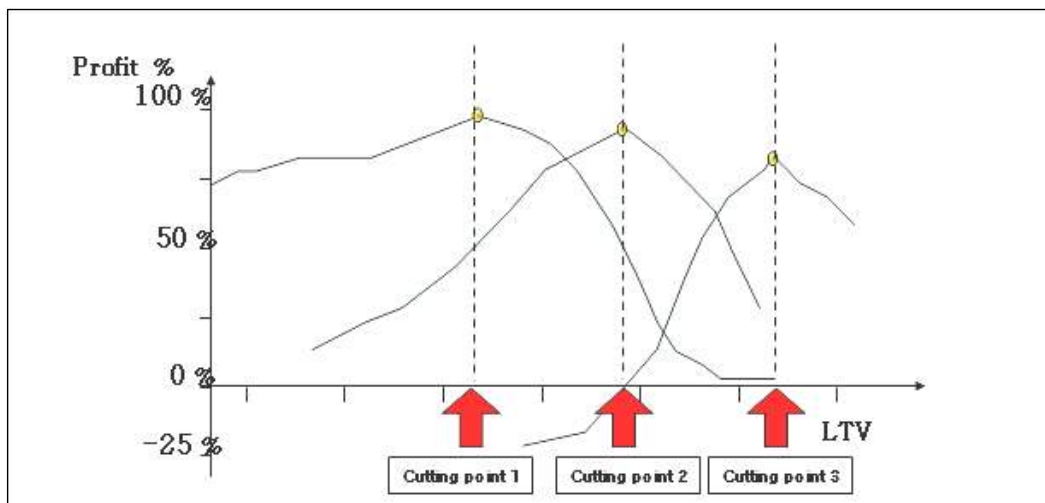
[그림 5] Holistic Profit 곡선

또한 주어진 손실대비 이득에 대한 비율($\omega^* : l/g$)에 대해 고객의 LTV 값에 따라 Holistic Profit 을 나타내면 일반적으로 [그림 5]와 같은 Holistic Profit 곡선을 구할 수 있다.

[그림 5]의 Holistic Profit 곡선을 살펴보면 LTV가 증가함에 따라 Holistic Profit이 함께 증가하다가 Holistic Profit이 최대화되는 V값 이상으로 LTV가 증가하면 Holistic Profit은 감소하는 모양을 하는 것을 볼 수 있는데 Holistic Profit이 최대화되는 LTV값이 고객의 등급을 구분하는 분류기준점(Cutting Point)으로 본다. 이는 완전 정보 하에서 기대이익을 가장 잘 설명해주는 LTV 값이다. 이렇게 해서 분류기준점 왼쪽에 속한 그룹을 하나의 등급으로 정한 후 나머지 오른쪽에 속한 고객들을 다시 둘로 구분하는 과정을 반복 실행하여 등급을 구분한다. 따라서 본 연구에서는 고객등급화를 위해 Holistic Profit을 최대화하는 LTV를 연속적으로 구하여 전체 고객집단을 적절한 등급으로 분류하는 방법을 제안하고 있다.

[그림 6]은 전체 고객집단을 4개의 집단으로 등급을 분류하기 위해 3회에 걸쳐 연속적으로 Holistic Profit 곡선을 도출하여 3개의 분류기준점(cutting point)을 산출하고 이를 기준으로 고객등급화를 하는 과정을 나타낸 그림이다.

따라서 고객의 등급을 결정하기 위한 분류기준을 산출하는 방법을 요약하면 먼저 전체 고객집단을 대상으로 Holistic Profit 곡선을 형성하여 하나의 분류점(cutting point)을 도출하고, 분류점이상의 고객군을 대상으로 다시 새로운 Holistic Profit 곡선을 구하여 두 번째 분류점을 산출한다. 이와 같은 방식으로 세 번째와 네 번째 분류점을 구하면 전체 고객집단을 5개의 고객집단으로 등급화 할 수 있다. 이때 주의할 점은 이미 분류가 된 고객데이터는 추가적인 Holistic Profit 곡선을 도출할 때에는 제외시킨다. 본 연구에서는 이러한 고객등급화 접근방법으로 전체고객집단을 5개의 등급으로 구분하기 위하여 4번의 Holistic Profit곡선을 산출하여 4개의 분류기준점을 산출하였다.



[그림 6] 고객등급화를 위한 분류기준점 산출방법

4. 실증분석

국내 A 생명보험회사의 고객 데이터베이스에서 추출한 고객데이터 19,162명의 자료를 이용해서 실증분석을 실시하였다. 고객평생가치(LTV)를 산출하는 모형 중에서 금융회사 특히 보험사 고객의 LTV 산정에 가장 적합한 Berger(1998)의 모형을 적용해서 고객별 LTV를 산출하였다. 또한 산출된 고객별 LTV를 토대로 Holistic Profit을 도출한 후 그 결과를 그래프로 나타내어 고객등급화를 위한 4개의 분류기준점을 도출하였다. 이를 토대로 전체 고객집단을 5개의 등급으로 분류한 결과와 일정한 비율에 의한 고객등급화 방법을 비교·분석하였다.

4.1 고객별 LTV 산출

본 논문에서 고객별 LTV를 산출하기 위해 적용된 모형은 Berger and Nasr의 Case 2a(Berger, P.D. and Nasr, N.I, 1998) 모형을 응용한 것으로 분석하고자 하는 데이터가 보험회사, 특히 생명보험회사의 고객 데이터에 적합한 모형이다.

Berger와 Nasr(1998)의 LTV 측정모형은 일년에 여러 번 판매를 가정하는 모형으로서 3개월 혹은 6개월 단위로 회원권을 갱신하는 헬스클럽과 같은 기업에 적용할 수 있는 모형이다. 이 모형을 기본모형으로 선택한 이유는 국내 생명보험의 특성과 관계가 있다. 미국이나 유럽의 경우에는 월납보다는 3개월납이나 6개월납, 그리고 연납을 선호하는 경향이 있지만 국내의 경우는 월납이 거의 대부분을 차지하고 있다. 따라서 이러한 모형을 이용하여 일년에 12번 판매가 일어나는 상품에 대한 고객의 평생가치를 측정하는 것과 같은 방법으로 생명보험회사 고객의 고객평생가치를 측정

하고자 한다.

이 모형을 생명보험회사에 적용하는 경우, GC' 는 매월 고객이 기업에 제공하는 공헌이익으로 합계보험료가 되며, M' 는 매월 기업이 개별 고객을 창출하고 유지하는데 드는 판매촉진비용으로 보험회사의 신계약상각비 및 보험유지비가 된다.

그러나 이러한 방식으로 생명보험회사 고객들의 LTV를 측정하면 합계보험료(GC')와 거래기간(pm)에 따라서 고객평생가치(LTV)가 결정된다. 즉, 합계보험료가 크면 클수록 거래기간이 길면 길수록 LTV가 크다. 따라서 본 연구에서는 두 가지 사항을 추가해서 LTV를 산출하였다.

첫째는 보험상품의 종류를 반영하였다. 생명보험상품은 상품의 종류에 따라 보장의 정도, 자금유입기간 등에 차이가 있기 때문에 기업수익에 미치는 영향도 상품별로 차이가 발생하게 된다. 따라서 본 연구에서는 상품의 종류에 따라 가중치를 부여하는 방식으로 상품별 차이를 합계보험료에 반영하였다. 가중치를 부여하는 방법은 타 상품에 비하여 보험회사 이익과 고객 이익이 평균정도인 상품에 기준 가중치(weight=1)를 설정한다. 그리고 보험사의 입장에서 보다 수익성이 높은 상품은 가중치를 높이고, 수익성이 낮은 상품의 가중치는 낮추는 방법으로 가중치를 부과했다. 이렇게 함으로써 고객이 가입한 보험 상품별로 해당 고객의 LTV에 미치는 영향을 반영하였다.

둘째는 일반적으로 지금까지 제안된 LTV 측정 모형에서 고객유지율은 전체 고객의 평균 유지율이다. 즉, 현재까지 전체 고객들의 연간 혹은 월 평균 유지율을 모든 고객들에게 동일하게 적용하였다. 그러나 고객들의 유지율은 고객별로 서로 다르며, 이를 동일하게 간주하여 LTV를 측정한다면 정확한 LTV를 측정했다고 할 수 없다. 따라

$$CLTV = \left[GC' \times \sum_{i=0}^{pn} \left\{ \frac{(r')^i}{(1+d)^{i/p}} \right\} - \left\{ M' \times \sum_{i=1}^{pn} \left[\frac{(r')^{i-1}}{(1+d)^{(i-0.5)/p}} \right] \right\} \right]$$

[그림 7] 고객평생가치 측정모형

서 본 연구에서는 분석대상 자료를 이용해서 고객 이탈 예측모형을 개발한 다음 이 모형으로부터 얻은 고객 이탈율을 반영하여 개별 고객들의 LTV를 산출하도록 했다. 본 연구에서 활용한 고객이탈 예측모형은 로지스틱 회귀분석이다.

보험 상품의 종류와 개별 고객의 유지율을 고려하여 다음과 같은 고객평생가치 측정모형을 사용하였다.

여기서 r' 은 고객이탈예측모형을 통해서 도출된 개별 고객의 유지율(1-이탈율)을 나타낸다. GC' 는 합계보험료를 나타내며, p 는 연간 판매주기로 매월 보험료를 납입하므로 12가 된다. M' 는 보험회사의 신계약상각비 및 보험유지비가 되며, d' 는 미래의 가치를 현재가치로 환산해주는 할인율이다. 그리고 n 은 보험료 납입기간으로서 단위는 년이 된다. 이러한 방식으로 유지고객들을 대상으로 생명보험회사 고객들의 LTV를 측정할 결과를 정리하면 <표 1>과 같다.

<표 1> 유지 고객들의 LTV에 관한 기술통계량 (단위 : 원)

구 분	LTV
평 균	1,831,008
표준편차	3,554,582
최 소 값	10,239
중 간 값	793,170
최 대 값	144,493,068

4.2 로지스틱(logistic) 회귀분석을 이용한 고객 이탈모형

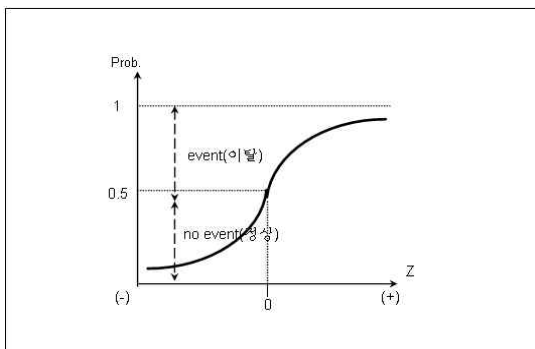
Holistic Profit 곡선을 이용하여 고객을 등급화하기 위해서는 고객의 우량 및 불량 여부를 알아야 한다. 고객의 우량 및 불량 여부는 고객 이탈모형을 적용하여 도출된 고객 이탈율을 이용하였다. 본 연구에서는 고객이탈 예측모형을 구하기 위하여 로지스틱 회귀모형을 이용하였다. 로지스틱(Logistic) 회귀모형은 어떤 하나의 사건이 발생할 확률을 측정하기 위한 다변량 기법(Multivariate Technique)이다. 판별분석이 정규분포라는 기본 가정을 충족시켜야 하는 통계적인 한계점이 있으나 로지스틱 회귀모형은 다변량 판별분석에 비해 자료의 분포나 변수의 척도에 대한 가정이 거의 없으며, 판별분석의 가정이 충족되었을 경우에는 보다 더 좋은 모형을 구축할 수 있다.

로지스틱 회귀모형에서 어떤 하나의 사건이 발생할 확률을 직접 측정할 수 있는데 한개 이상의 독립변수가 있을 경우, 로지스틱 회귀모형은 다음과 같다.

$$prove(event) = \frac{1}{1 + e^{-x}}$$

여기서 e 는 자연로그(≈2.718)이며, $Z = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_P * X_P$ 로서 선형결합이다. Z 의 값은 실수를 취하며, 이때 로지스틱 회귀곡선(Logistic Regression Curve)은 S자형이

며, 0과 1 사이의 확률값을 가진다. S자형 모양은 정규분포의 누적확률곡선과 유사하다. 따라서 독립변수와 확률간의 관계가 비선형이다. 로지스틱 회귀모형의 모수(Parameter)는 최대우도법(Maximum-likelihood Method)을 사용해서 산출하였다.



[그림 8] 로지스틱 회귀분석 확률곡선

생명보험회사의 고객 데이터베이스에서 추출한 자료의 27개 변수 중에서 날짜를 나타내는 변수인 계약일자, 만기일자, 실효일자, 업적월도, 최종연월, 최종납입일자 등의 변수들은 납입기간이나 보험기간, 유지상태 등을 통해 충분히 파악할 수 있으므로 분석에서 제외하였다. 또한 유지상태 변수에서 기실효, 당월실효, 시효소멸을 이탈(1)로 하고, 정상유지를 정상(0)으로 하는 새로운 목적변수를 생성하였다.

27개의 변수 중에서 계약일자 등 날짜를 나타내는 변수 6개와 유지상태 변수를 제외한 20개의 변수들을 모두 분석에 사용하기가 어려워 이 중에 분석에 유용한 변수들을 1차적으로 선별하기 위하여 교차분석과 상관분석, T-test, 그리고 스텝와이즈 로지스틱 회귀분석을 실시하였다. 먼저 9개의 범주형 변수들을 대상으로 목적변수(이탈, 정상)와 교차분석을 실시하여, 유의수준 0.05에서 유의하지 않은 변수들을 제외시켰다. 그 결과 <표 2>

와 같이 미아고객 여부를 나타내는 변수가 χ^2 값이 0.7395이고, 유의수준 0.05에서 유의확률 p값이 0.3898로 통계적으로 유의하지 않게 나타나 분석에 적용되는 변수목록에서 제외시켰다.

<표 2> 범주형 변수의 교차분석 결과

변수명	χ^2 test	
	χ^2 값	p-value
계약자 성별	244.1614	<.0001
계약자 거주지	102.7958	<.0001
계약자 직업	422.3366	<.0001
납입방법	76.54	<.0001
수금방법	8693.2617	<.0001
사원계약여부	910.5838	<.0001
미아고객여부	0.7395	0.3898
설계사 성별	32.8388	<.0001
상품종류	61.4644	<.0001

다음으로 12개의 연속형 변수들을 대상으로 목적변수(이탈, 정상)와 상관분석을 실시하고 T-test를 실시하였다. 그 결과 <표 3>과 같이 유의수준 0.05에서 모든 변수들이 통계적으로 유의하게 나타났다.

<표 3> 연속형 변수의 상관분석 및 T-test 결과

변수명	Correlation		T-test	
	상관계수	p-value	t값	p-value
계약자 연령	-0.03299	<.0001	4.67	<.0001
납입기간	0.0959	<.0001	-13.62	<.0001
보험기간	-0.08847	<.0001	12.56	<.0001
보험금	-0.0172	<.015	2.43	<.015
합계보험료	-0.02655	<.0002	3.76	<.0002
예정이율	0.14843	<.0001	-21.23	<.0001
총납입회수	-0.35053	<.0001	52.93	<.0001
상품가입회수	0.42227	<.0001	-65.88	<.0001
설계사연령	0.03733	<.0001	-5.19	<.0001
과거실효회수	0.20826	<.0001	-30.11	<.0001
대출회수	0.19824	<.0001	-28.6	<.0001

마지막으로 20개의 변수들 모두를 사용해서 로지스틱 회귀분석을 스텝와이즈(stepwise) 방식으로 실행하여 유의수준 0.05에서 최종적으로 선정된 변수들을 추출한 결과 <표 4>와 같이 계약자 직업과 대출회수를 제외한 나머지 변수들이 채택되었다.

<표 4> 로지스틱 회귀분석 Stepwise 결과

변수명	Logistic Reg.	
	Estimate	pr.>x ²
수금방법	8274.7972	<.0001
상품가입 개수	1658.624	<.0001
총납입회수	932.9383	<.0001
예정이율	1441.8445	<.0001
납입방법	474.432	<.0001
과거 실횠회수	329.7945	<.0001
미아고객여부	131.4855	<.0001
납입기간	92.4631	<.0001
계약자 연령	63.1486	<.0001
보험기간	63.6723	<.0001
상품종류	58.7734	<.0001
설계사 성별	40.718	<.0001
사원계약여부	30.6493	<.0001
계약자 거주지	54.3534	<.0001
설계사 연령	14.6799	0.0001
계약자 성별	9.3405	0.0022
합계보험료	5.6961	0.017
보험금	13.0965	0.0003

이상의 분석결과를 바탕으로 네 가지 분석 방법 모두에서 선택된 변수들을 사용하여 고객이탈 예측모형을 구축하였다. 그 결과를 정리하면 <표 5>와 같다.

<표 5> 모형구축에 선택된 변수

변수명	교차 분석	상관 분석	T-test	로지스틱	최종 여부
계약자 성별	○			○	○
계약자 직업	○			×	×
계약자 거주지	○			○	○
상품종류	○			○	○
납입방법	○			○	○
수금방법	○			○	○
사원계약여부	○			○	○
미아고객여부	×			○	×
설계사성별	○			○	○
계약자 연령		○	○	○	○
납입기간		○	○	○	○
보험기간		○	○	○	○
보험금		○	○	○	○
합계보험료		○	○	○	○
예정이율		○	○	○	○
총 납입회수		○	○	○	○
설계사 연령		○	○	○	○
상품가입 개수		○	○	○	○
과거 실횠회수		○	○	○	○
대출회수		○	○	×	×

4.3 Holistic Profit를 이용한 고객 등급화

Holistic Profit 곡선을 이용하여 고객을 등급화 하기 위해서는 고객을 우량고객과 불량고객으로 구분하여야 하는 데 본 연구에서는 이탈여부에 따라 우량고객과 불량고객으로 구분하였다. 따라서 이탈유무를 나타내는 변수인 목적변수를 기준으로 이탈(1)이면 불량(Bad)고객으로 유지(0)이면 우량(Good)고객으로 분류를 하되, 데이터의 특성 상 유지율을 중심으로 데이터를 분류하여 상위 50%는 우량고객으로 하위 50%는 불량고객으로 가정하여 우량/불량고객의 비율을 동일하게 적용

<표 6> 고객집단의 우량고객, 불량고객, 전체의 확률 및 누적확률 분포

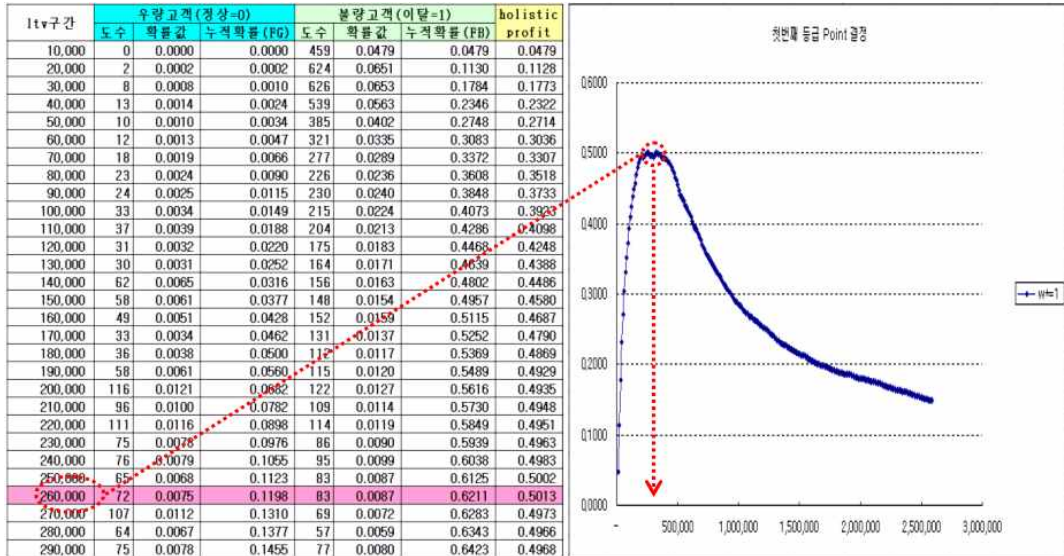
ltv 구간	우량고객(정상 = 0)			불량고객(이탈 = 1)		
	도수	확률값	누적확률(FG)	도수	확률값	누적확률(FB)
10,000	0	0.0000	0.0000	459	0.0479	0.0479
20,000	2	0.0002	0.0002	624	0.0651	0.1130
30,000	8	0.0008	0.0010	626	0.0563	0.1784
40,000	13	0.0014	0.0024	539	0.0563	0.2346
50,000	10	0.0010	0.0034	385	0.0402	0.2748
60,000	12	0.0013	0.0047	321	0.0335	0.3083
70,000	18	0.0019	0.0066	277	0.0289	0.3372
80,000	23	0.0024	0.0090	226	0.0236	0.3608
90,000	24	0.0025	0.0115	230	0.0240	0.3848
100,000	33	0.0034	0.0149	215	0.0224	0.4073
110,000	37	0.0039	0.0188	204	0.0213	0.4286
120,000	31	0.0032	0.0220	175	0.0183	0.4468
130,000	30	0.0031	0.0252	164	0.0171	0.4639
140,000	62	0.0065	0.0316	156	0.0163	0.4802
150,000	58	0.0061	0.0377	148	0.0154	0.4957
:	:	:	:	:	:	:
47,620,000	0	0.0000	0.9997	0	0.0000	1.0000
47,630,000	0	0.0000	0.9997	0	0.0000	1.0000
47,640,000	0	0.0000	0.9997	0	0.0000	1.0000
47,650,000	0	0.0000	0.9997	0	0.0000	1.0000
47,660,000	0	0.0000	0.9997	0	0.0000	1.0000
144,493,069	3	0.0003	1.0000	0	0.0000	1.0000
	9581	1.0000	2.0000	9581	1.0000	2.0000

하였다. 또한 이득과 손실비율(1/g)도 1로 가정하여 Holistic Profit을 산출하였다.

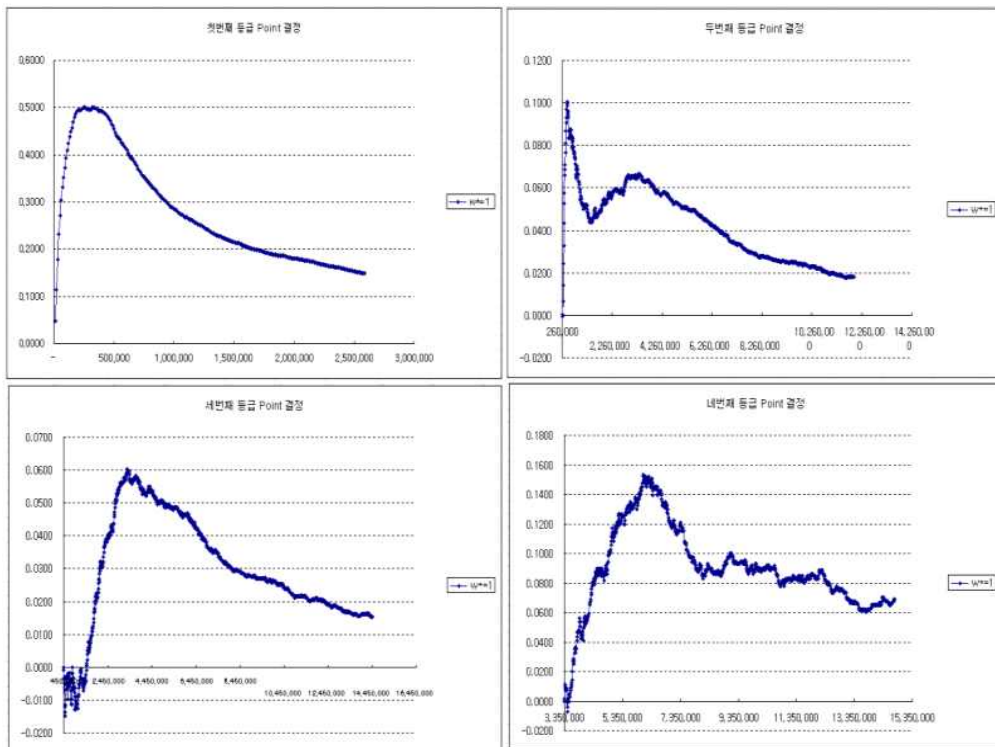
Holistic Profit 곡선을 이용하여 고객을 등급화하기 위한 첫번째 분류기준점(Cutting Point)을

도출하기 위해서 필요한 전체 고객집단, 우량고객, 불량고객에 대한 확률분포와 누적확률분포에 대한 산출결과는 <표 6>과 같다.

고객 LTV의 전체 구간에 대해 Holistic Profit



[그림 9] 고객등급화를 위한 분류기준점 산출



[그림 10] Holistic Profit을 이용한 고객등급화 분류기준점 산정

분포를 그려보면 [그림 9]와 같다. 그래프가 처음에는 지속적으로 상승하다가 특정한 LTV값을 지나면서 점차 Holistic Profit이 줄어드는 형태를 보인다. 여기서 첫 번째 분류기준점은 Holistic Profit이 최대화되는 LTV값이 된다. 왜냐하면 바로 그 분류기준점이 기업의 이익을 극대화하면서 우량고객과 불량고객의 수를 적정하게 구분해 주는 분류기준점이 되기 때문이다.

분석결과 가장 하위등급인 5등급과 차 하위 등급인 4등급을 구분하는 고객들의 등급분류 기준점은 LTV값이 260,000원인 것으로 나타났다. 즉, LTV가 260,000원 이하인 고객들은 5등급으로 분류하고 5등급에 속한 고객데이터를 제외한 나머지 고객자료를 대상으로 연속적으로 Holistic Profit 곡선을 적용하여 4등급, 3등급, 2등급, 1등급으로 고객집단을 분류하였다. 고객등급화의 분류기준점을 산출한 결과는 [그림 10]과 같다.

5. 고객등급화 방법의 비교분석

본 논문에서 제시한 Holistic Profit에 의한 고객 등급화 방법의 효과성을 검증하기 위해서 가장 일반적으로 적용되고 있는 일정 비율에 의한 등급

화 방법과 비교 분석되었다. 즉, 분석 대상이 되는 고객의 LTV를 일정한 비율로 나누어 고객등급화를 하는 방법과 Holistic Profit 곡선을 이용하여 고객 등급화 하는 방법을 비교분석하여 과연 Holistic Profit 곡선을 이용한 고객등급화가 더 효과적인지를 검증하고자 하였다. 구체적인 검증방법으로는 고객등급별로 우량고객과 불량고객의 비율을 파악해보고, 고객 등급간 LTV 차이검증을 통하여 등급 내 고객의 동질성과 등급간 고객의 이질성을 F-검정을 통해 분석하였다. 또한 동일 집단군 검증을 통해 몇 개의 그룹으로 분류가 되는지를 함께 검증해보았다.

5.1 일정한 비율로 고객등급화

일정한 비율로 고객을 등급화 방법은 LTV에 따른 고객의 분포도를 산출하고 그러한 분포정보를 이용하여 일정한 비율로 고객을 구분하였다. 먼저 고객을 등급화 하는 방법으로 전체 고객들의 LTV를 산출한 후 LTV값의 크기에 따라 순서대로 나열하고 20%씩 일정한 비율로 고객을 5등급으로 분류하였다. 분류된 고객등급별 프로파일은 <표 7>과 같다.

우량고객이 14,686명(77%)이고 불량고객이 4,476

<표 7> 일정한 비율에 의해 분류된 고객등급 프로파일

등급	구간				합	최소값	최대값	평균	우량	불량
	1347106	이상	144493068	이하						
1	1347106	이상	144493068	이하	3832	1347106	144493068	4106070	3608	224
2	623956	이상	1946732	이하	3833	623956	1346732	913041	3619	214
3	297475	이상	623952	이하	3832	297475	623952	455998	3560	272
4	912285	이상	297450	이하	3833	91285	297405	186673	2980	853
5	186	이상	91254	이하	3832	186	91254	37941	919	2913
합계					19162	186	144493068	1139883	14686	4476

명(23%)인 전체 19,162명의 LTV를 산출하였고 이를 일정비율로 나누어 등급화 하였다. 1등급의 경우 전체 LTV값이 1,347,106원 이상 144,493,068원 이하인 고객들의 집단으로 LTV 평균값은 4,106,070원 이었다. 우량고객은 3,608명이고 불량고객은

224명으로 불량고객이 1등급에서 5.85%를 차지 하였다. 전반적으로 상위등급일수록 우량고객의 비율이 높고 불량고객의 비율이 낮았다. 또한 F-검정을 통해 고객 등급간 고객 LTV의 차이검증을 실시하여 <표 8>과 같은 결과를 도출하였다.

<표 8> 일정한 비율에 의해 분류된 고객등급간 고객LTV 평균 차이검증

	제곱합	자유도	평균제곱	F	유의확률
집단-간	43840213371832000	4	10960053342958000	2249.640853	0.00
집단-내	93331227363483700	19157	4871912479171		
합계	137171440735316000	19161			

<표 9> 일정한 비율에 의해 분류된 고객 등급간 동질성 검증

등급	N	유의수준 = 0.05에 대한 부집단			
		1	2	3	4
Scheffe(a, b)	5	3832	37941.03		
	4	3833	186672.91		
	3	3832		455998.14	
	2	3833			913040.55
	1	3832			

일정한 비율에 의해 분류된 고객 등급간 LTV 평균 차이를 검증한 결과 유의수준 0.05에서 통계적으로 유의하였으며, F값은 2249.64로 도출되었다. 또한 분산분석에서 특히 어느 집단 간의 차이가 명확한 지를 분석하는 동일집단군 검증을 실시하여 <표 9>와 같은 결과를 얻었다. 즉, 동질성 검증을 실시한 결과, 5개의 고객등급으로 고객을 분류했으나 5등급과 4등급간에는 서로 큰 차이가 없이 1개의 집단에 속하는 것으로 나타나 고객등급의 분류는 5개의 등급보다는 4개 등급이 더 바람직함을 알 수 있었다.

결론적으로 일정한 비율에 의해 고객을 등급화 하는 경우를 분석한 결과 동질성 검증에서 5개 등급으로 정확하게 구분되지 않아 현재의 고객등급 구분에 한계가 있는 것으로 분석되었다.

5.2 Holistic Profit을 이용한 고객등급화

본 논문에서 제시한 Holistic Profit을 이용하여 전체 고객을 등급화 하는 방법을 이용해서 전체 고객을 5개의 등급으로 분류하였고 이에 따른 고객등급별 프로파일은 <표 10>과 같다.

<표 10> Holistic Profit을 이용하여 분류된 고객등급 프로파일

등급	구간				합	최소값	최대값	평균	우량	불량
	1	2	3	4						
1	6074675	이상	144493068	이하	595	6074675	144493068	11841244	567	28
2	3351132	이상	6069121	이하	825	3351132	6069121	4434795	775	50
3	450183	이상	3342777	이하	8258	450183	3342777	1162833	7769	489
4	260029	이상	449944	이하	2385	260029	449944	348986	2186	199
5	186	이상	259879	이하	7099	186	259879	99054	3389	3710
합계					19162	186	144493068	1139883	14686	4476

등급1에 속한 고객들은 LTV 값이 6,074,675원 이상 144,493,068원 이하인 고객들로서 이들의 LTV 평균값은 11,841,244원 이었다. 여기에 속한 고객중에 우량고객은 567명이고 불량고객은 28명 으로서, 불량고객이 전체의 4.70%인 것으로 나타났다. 이는 일정한 비율에 의해 분류된 1등급의

불량고객비율 5.85%보다 약 1%정도 작게 나타났다. 전반적으로 상위등급일수록 우량고객의 비율이 높고 불량고객의 비율이 낮았다. 또한 F-검정을 통해 고객 등급간 LTV에 대한 차이 검증을 실시한 결과는 <표 11>과 같다.

Holistic Profit을 이용하여 고객 등급간 LTV

<표 11> Holistic Profit을 이용하여 분류된 고객 등급간 LTV 평균 차이검증

	제곱합	자유도	평균제곱	F	유의확률
집단-간	86282185769952900	4	21570546442488200	8120.12199	0.00
집단-내	50889254965363100	19157	2656431328776		
합계	137171440735316000	19161			

<표 12> Holistic Profit을 이용하여 분류된 고객 등급간 동질성 검증

등급	N	유의수준 = 0.05에 대한 부집단				
		1	2	3	4	5
Scheffe(a, b)	5	7099	99053.84			
	4	2385		348986.31		
	3	8258			1162833.00	
	2	825				4434794.69
	1	595				

평균 차이를 검증한 결과 유의수준 0.05에서 통계적으로 유의하였으며, F값은 8,120.12로 매우 높게 나타났다. 일정한 비율에 의해 분류된 고객 등급간 LTV 평균 차이검증에서 도출된 F값인 2,249.64과 비교하여 F값이 매우 높게 나타나 등급간에 차이가 더욱 명확하다고 판단할 수 있다.

동질성 검증을 실시한 결과는 <표 12>와 같으며 고객등급의 분류정도를 분석한 결과, 5개의 고객 등급이 모두 명확하게 분류됨을 확인할 수 있었다.

결론적으로 Holistic Profit 곡선을 이용하여 고객을 등급화 하는 방법이 일정한 비율에 의해 등급화 하는 방법보다 고객을 좀더 효과적으로 분류함과 동시에 등급간에 차별성도 보다 명확한 것으로 분석되었다.

6. 결론

산업이 발달하고 경쟁이 치열해지며, 고객들의 요구가 점차 다양해짐에 따라 기업들의 고객에 대한 관심은 더욱 더 증가하고 있다. 이에 기업들은 고객의 중요성을 새롭게 인식하고 CRM(Customer Relationship Management)시스템 구축에 박차를 가하고 있다. 성공적인 CRM활동을 위해서는 고객집단을 효과적으로 분류하여 이를 바탕으로 차별화된 마케팅활동을 수행할 필요가 있다. 일반적으로 고객을 분류하는 기준으로 고객평생가치(LTV)가 활용되고 있다. 그러나 고객평생가치를 적용한다고 해도 등급을 나누기 위해 고객을 분류하는 기준점을 찾는 것은 용이하지 않다. 그동안 고객별 LTV를 산정한 다음 그에 따른 고객의 분포도를 산출하고, 산출된 분포정보를 이용하여 일정한 비율로 고객을 분류하고 등급을 정하는 것이 하나의 방법으로 활용되어 왔다.

본 논문에서는 금융기관의 신용대출 승인정책에서 광범위하게 이용되고 있는 Holistic Profit을 활용하여 보다 효율적이고 과학적인 방법으로 고객을 등급화 하는 방법을 제안하고 그 효과성을 검증하였다. 비율을 이용하여 고객을 등급화 하는 일반적인 방법과 비교하여 고객등급별 우량(Good)고객 및 불량(Bad)고객 비율, 고객 등급간 LTV 차이검증(F값)을 통해 분석한 결과, 본 연구에서 제시된 Holistic Profit을 활용하여 고객을 등급화 하는 방법이 일정한 비율로 고객을 등급화 하는 방법 보다 더 효과적이라는 실증적인 결과를 얻을 수 있었다. 따라서 기업은 기존의 일정한 비율에 의해 고객을 등급화 하는 방법을 개선하여 Holistic Profit을 적용한 고객등급화를 활용한다면 보다 효율적이고 과학적인 방법으로 고객집단을 등급화할 수 있을 것으로 기대된다.

그러나 본 연구에도 적지 않은 한계점이 있다. 본 연구의 대표적인 한계점으로는 첫째, Holistic Profit을 활용하여 고객집단을 분류하기 위해 고객별로 산출한 고객평생가치(LTV)의 산출방법에 있다. 효과적인 고객의 등급화를 위해서는 정확한 LTV의 산출이 전제가 되어야 하는데 본 연구에서는 고객이 하나의 생명보험상품에 가입하고 있다는 전제하에서 LTV를 측정하였다. 2005년 생명보험협회의 통계자료를 살펴보면 국내 가구당 평균 생명보험 가입건수가 3.8건에 이르고 있다. 따라서 고객별로 복수의 생명보험 상품을 가입하고 있는 경우가 대부분으로, 고객별 LTV를 산출하는 데에 이를 반영하지 못한 점이다.

둘째, LTV가 지나치게 낮은 고객들이 많아 Holistic Profit을 활용한 고객등급화 방법에 대한 평가가 제대로 이루어지지 못하였다는 점이다. 분석 자료의 특성상 LTV가 100,000원 미만으로 상당히 낮은 LTV 고객들의 빈도가 굉장히 높게 나

타났다. 이러한 고객들의 특징을 살펴보면 보험료 납입이 1년도 채 남지 않은 고객 즉, 향후 1년 이내에 보험료 납입기간이 끝나서 완납이 되는 고객들이다. 이러한 LTV의 편중으로 인하여 Holistic Profit을 활용하여 등급을 구분하는 기준점을 산출하기 위해서 여러 가지 가정을 할 수 밖에 없었다.

셋째, 고객의 LTV를 이용하여 고객을 등급화하고자 하는 경우 고객의 신용위험을 추가하여 복합적으로 고려하는 연구가 필요하다. 즉, 신용위험과 동시에 고객가치를 반영하여 고객을 등급화하고 또한 마케팅 차원에서 어떻게 활용할 것인가에 대한 추가적인 분석이 이루어져야 할 것이다. 향후 연구는 이러한 제한점들을 보완해서 수행하면 보다 일반화할 수 있는 결과를 도출할 것으로 기대된다.

참 고 문 헌

[1] 양주환, 이훈영, 안성식 “신용대출 시장에서

의 최적대출 전략에 관한 실증적 연구”, *경영 관련 통합학술대회, 경영과학회 발표*, 2002.

[2] 양주환, “신용대출 시장에서의 최적대출 승인 정책에 관한 실증 연구”, *경희대학교 박사학위 논문*, 2002.

[3] Berger, P. D. and Nasr, N. I., “Customer Lifetime Value: Marketing Models and Applications”, *Journal of Interactive Marketing*, Vol.12 No.1, 1998.

[4] Bruce Hoadley(1996), “The Holistic Profit Scan-A Better Way to Measure Score Performance”, INTERACT96, Fair & Isaac Forum.

[5] Hoadley B. and Oliver RM(1998), “Business Measures of Scorecard Benefit”, *IMA Journal of Mathematics Applied in Business and Industry*.

[6] RM Oliver and E Wells(2001), “Efficient Frontier Cutoff Policies in Credit Portfolios”, *Journal of the Operational Research Society*, Vol.52.

Abstract

A Model for Effective Customer Classification Using LTV and Churn Probability : Application of Holistic Profit Method

Lee, HoonYoung* · Yang, JooHwan** · Ryu, Chi Hun*

An effective customer classification has been essential for the successful customer relationship management. The typical customer rating is carried out by the proportionally allocating the customers into classes in terms of their life time values. However, since this method does not accurately reflect the homogeneity within a class along with the heterogeneity between classes, there would be many problems incurred due to the misclassification. This paper suggests a new method of rating customer using Holistic profit technique, and validates the new method using the customer data provided by an insurance company. Holistic profit is one of the methods used for deciding the cutoff score in screening the loan application. By rating customers using the proposed techniques, insurance companies could effectively perform customer relationship management and diverse marketing activities.

Key words : Customer Relationship Management (CRM), Customer Life Time Value (LTV), Customer Rating, Holistic Profit, Churn Probability

* School of Business Administration, KyungHee University

** KyungHee Cyber University