

RFM에서 등급부여 방법에 관한 연구

류귀열¹ · 문영수²

¹서경대학교 컴퓨터과학과 · ²한국과학기술정보연구원

접수 2013년 1월 21일, 수정 2013년 2월 6일, 게재확정 2013년 2월 13일

요약

본 논문은 RFM (recency frequency monetary) 모델에서 등급을 매기는 방법을 정규분포를 이용하여 6등급모델을 제안하고 NDSL (national discovery for science leaders) 자료를 이용하여 현재 많이 사용되고 있는 5등급모델과 10등급모델을 비교하였다. 제안 모델이 5등급모델과 10등급모델에 비해 고객그룹들을 쉽게 세분화할 수 있다는 사실을 알 수 있었다. 제안된 모델은 대칭적으로 등급을 부여하고, 고객분포를 이용하기 때문에 고객특성을 잘 반영함으로써 경계값들이 명확하게 구분되는 특징을 가지고 있다. 또한 등급 값을 보면 쉽게 어느 위치에 속하고 있는 지 알 수 있으며, 고객세분화 후 고객의 RFM값들의 확인으로 고객의 특성을 쉽게 알 수 있는 장점이 있다. 향후 군집분석 등의 통계적 등급부여 방법들과 비교연구와 가중치 부여 방법에 관한 연구가 필요하다.

주요용어: 고객관계관리, 고객세분화, 충구매액, 최근성, 최빈성.

1. 서론

기업들은 고객들의 다양한 욕구를 충족시키기 위해 많은 노력을 경주해 왔다. 그러나 기업의 제한된 자원으로 현대 사회에서 다양하고 끊임없이 변화하는 개인의 욕구를 만족시키면서 전체시장에서 경쟁우위를 보이는 일은 거의 불가능에 가깝다. 따라서 어느 기업이든 전체시장에서 모든 기업들과 경쟁하기 보다는 자사가 경쟁우위를 가지고 있는 제품이나 서비스를 가지고 자사가 가장 유망한 세분시장에서 경쟁하는 것이 유리하다. 또한 정보사회의 진전에 따라 고객 욕구는 매우 다양하게 변하고 있으며 이는 기업에게는 매우 큰 도전이 되고 있다. 왜냐하면 고객의 욕구에 잘 대응하지 못하는 기업들은 미래를 보장받지 못하고, 규모가 작을지라도 고객의 욕구에 잘 대응하는 기업들은 빠른 시간에 효과적으로 시장을 확보할 수 있는 역동적인 시대가 펼쳐져 있기 때문이다.

다양하고 복잡한 소비자 욕구를 파악하기 위해 기업들은 CRM (customer relationship management; 고객관계관리) 시스템을 도입하고, 이를 기반으로 하는 마케팅을 실시하는 기업이 늘고 있다. CRM은 고객의 욕구에 대한 광범위하고 심층적인 지식을 바탕으로 고객에게 적합한 차별적인 제품이나 서비스를 제공함으로써 고객과의 관계를 지속적으로 강화하는 활동을 말한다. CRM은 고객에 대한 정보를 지속적으로 수집하여 심층적인 분석을 통해, 우수고객을 유지하고 잠재고객을 발굴하고 새로운 시장을 개발하는 등 효과적인 마케팅 도구로 활용되고 있다. 많은 기업들은 CRM을 통해 고객만족을 실현하여 기업가치 극대화를 도모하고 있다. 실제로 CRM을 도입하여 고객만족도 향상을 통한 매출증대, 고객이탈로 인한 비용 최소화, 고객유지비용 감소 등을 실현하고 있다.

¹ (136-704) 서울시 성북구 정릉동 산16-1, 서경대학교 컴퓨터과학과, 부교수.

² 교신저자: (130-741) 서울시 동대문구 회기로 66, 한국과학기술정보연구원, 선임연구원.

E-mail: youngsum@kisti.re.kr

일반적으로 시장은 이질적 욕구를 가진 다양한 고객들의 집합으로 구성되어 있다. 기업들은 개별적인 욕구를 만족시키기에는 인적 물적 자원의 한계를 가지고 있다. 또한 고객 개인별 욕구는 상이하지만 제품에 대한 태도, 의견, 구매행동 등의 면에서 유사한 소비자 집단들도 존재하고 있다. 따라서 비슷한 성향을 가진 고객들의 집단과 다른 성향을 가진 고객들의 집단으로 분리하여 하나의 집단으로 묶는 것이 필요한 데 이를 고객세분화라고 한다. 고객세분화는 CRM에서 매우 중요한 작업이고, Song과 Kim (2001)에서 보듯이 마케팅에서 많이 사용하는 전략인 STP (segmentation targeting position) 전략에서 가장 기초적인 작업이다.

본 논문은 고객세분화에서 많이 사용되고 있는 RFM (recency frequency monetary) 모형 중에서 등급을 구분하는 방법을 제안하고, 기존 사용하였던 방법들과 NDSL (national discovery for science leaders) 자료를 이용하여 비교한다. 본 논문의 구성은 다음과 같다. 2절에서는 고객세분화와 RFM 관계를 설명하고 RFM 해석을 통한 3절에서는 새로운 등급방법을 제안하고 4절에서는 NDSL 자료를 이용하여 제안된 방법과 5등급모델과 10등급모델과 비교할 것이다. 5절에서 결론을 내리고자 한다.

2. 고객세분화와 RFM 해석

고객세분화는 시장을 일정한 기준으로 동질적인 세분시장으로 구분하는 것을 말한다. 목적은 과학적인 방법으로 고객을 정확히 이해하여 기업의 경쟁력을 높이는 것이다. 고객세분화에서 고려되어야 할 사항으로는 판별하기 쉬워야 하며, 분류에 따라 적절하고 차별적인 대응이 가능하여야 한다. 일반적인 세분화 기준으로는 연령, 성별, 소득 등의 인구통계적 기준, 사회계층, 라이프스타일 등 심리변수적 기준, 구매패턴, 구매동기 등의 소비자행동적 기준, 지역에 의한 기준 등이 있다. 이들은 한 가지 변수로만 고객을 분류하기 때문에 한계를 가지고 있다. 이를 개선하고 고객을 과학적이고 다면적으로 평가하기 위해, 고객이 얼마나 최근에 구매했는가를 평가하는 최근성과, 고객이 얼마나 자주 구매했는가를 평가하는 최빈성, 고객이 얼마나 많이 구매했는가를 평가하는 총구매액 등 3가지 요인으로 고객을 세분화하는 방법인 RFM 모형이 개발되어 많이 사용되고 있다. RFM 모형은 $RFM = a \cdot R + b \cdot F + c \cdot M$ 으로서 R, F, M의 등급을 매기는 문제와 가중치 a, b, c 를 결정하는 문제가 있다. 등급을 매기는 이유는 R, F, M의 단위로부터 발생하는 문제를 없애고 동일한 기준으로 평가하기 위해서 이다. 많이 사용되는 방법은 일률적으로 5등급 또는 10등급 등으로 나누는 방법이다. 가중치 a, b, c 를 결정하는 방법에는 Lee (2002)와 같이 파레토 법칙을 이용하는 방법과 회귀분석을 이용하는 방법, Hodgson (1998)과 같이 일정한 값을 적용하는 방법 등이 있다. 그러나 Figure 2.1에서 보듯이 R, F, M의 가중치인 계수를 결정하는 문제는 고객들이 분포되어 있는 대로 세분화된 그룹들이 잘 구분되는 사영을 하도록 결정되어야 함을 보여 준다.

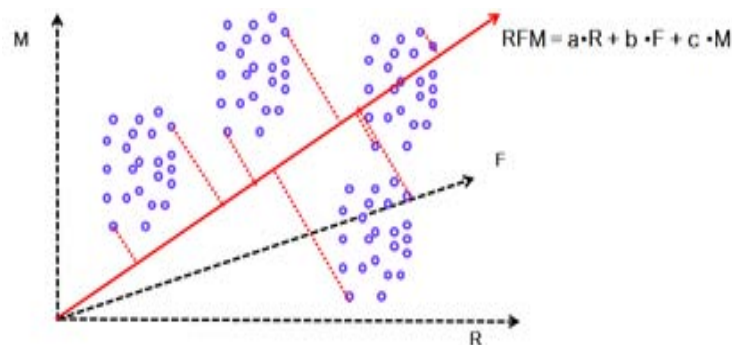


Figure 2.1 Projection of RFM

이 외에도 군집분석을 이용한 세분화, 의사결정나무를 이용한 세분화, 다차원척도법을 이용한 세분화, 데이터마이닝을 이용한 세분화, 인공지능망 분석을 이용한 세분화, LTV (life time value)를 이용한 세분화 등이 있다. 고객세분화관련 연구는 MacQueen (1967), Lee 등 (2004), Lee 등 (2010), Park (2010), Cho와 Park (2011a, 2011b, 2011c, 2012) 등이 있다.

3. RFM에서 등급값 부여방법 제안

RFM 모형에서 각 요인별 등급을 매기는 방법은 일반적으로 일률적으로 20%씩 나누는 5등급 또는 10%씩 나누는 10등급 등을 많이 사용하였다. 이러한 방법들은 확일적으로 나누는 방법으로 R, F, M의 분포를 무시하는 경향이 있다. 분포를 고려하여 등급을 매기는 군집분석을 통하여 등급값을 부여 하는 방법이 있으나, 군집의 특징을 쉽게 이해하지 못하는 문제점이 있다. 우리는 분포를 고려하고 쉽게 이해할 수 있는 등급 값 부여방법을 제안하고 이에 따르는 가중치 선정과 세분화 그룹의 결과를 NDSL 자료를 이용하여 비교하려고 한다. Rogers (1962)는 신제품을 수용하는 세분그룹을 통계분포를 이용하여 제시하였다. Rogers가 제시한 모형은 신제품을 수용하는 시간이 $\sim \mu - 2\sigma$ 까지의 상위 2.5%를 혁신층, 그 다음 $\mu - 2\sigma \sim \mu - \sigma$ 의 상위 13.5%를 조기수용층, 그 다음 $\mu - \sigma \sim \mu$ 의 34%를 조기 다수층, 그 다음 $\mu \sim \mu + \sigma$ 의 34%를 후기다수층, 마지막 $\mu + \sigma \sim$ 의 16%를 후발수용층으로 구분하여 그룹들의 특성을 설명하였다. Rodgers 모형은 Seo와 Kim (2011) 등과 같이 신제품 마케팅과 혁신수용모델 등에 널리 활용되고 있는데, 왜냐하면 통계적 분포를 이용하여 구분함으로써 이론적으로 그룹의 특성이 잘 구분되기 때문이다. Table 3.1은 Rodgers의 요약 표이다.

Table 3.1 Experimental design of a, b, c

Name	innovators	early adopters	early majority	late majority	laggards
Adoption time	$\sim \mu - 2\sigma$	$\mu - 2\sigma \sim \mu - \sigma$	$\mu - \sigma \sim \mu$	$\mu \sim \mu + \sigma$	$\mu + \sigma \sim$
Proportion	2.5%	13.5%	34%	34%	16%

우리가 제안하는 등급값 부여 방법은 Rogers 모형을 활용하여 $\mu + 2\sigma$ 의 상위 2.5%에 6점 최우수 그룹, 그 다음 $\mu + \sigma \sim \mu + 2\sigma$ 의 상위 13.5%에 5점 우수 그룹, 그 다음 $\mu \sim \mu + \sigma$ 의 34%에 4점 상위 그룹, 그 다음 $\mu - \sigma \sim \mu$ 의 34%에 3점 하위그룹, 그 다음 $\mu - 2\sigma \sim \mu - \sigma$ 의 하위 13.5%에 2점 차하위 그룹, $\sim \mu - 2\sigma$ 의 2.5%에 1점 최하위그룹으로 부여하는 6등급 부여 방법이다. 이러한 방법은 균등하게 구분하는 것 보다 일반적으로 중심에 많이 모이는 경향을 띄고 있는 고객 분포를 이용하기 때문에 고객특성을 잘 반영하고 경계값들이 명확하게 구분되고, 등급 값을 보면 쉽게 어느 위치에 속하고 있는지 알 수 있다. 또한 고객세분화 후 고객의 RFM값을 역 추적함으로써 고객 및 그룹의 특성을 쉽게 알 수 있는 장점이 있다. 5등급으로 구분하는 경우 상위 20%를 동일한 등급으로 평가함으로써 최우수 고객이라 불릴 수 있는 최상위 그룹을 상위 20%에 속하게 하여 구분하기 어렵고, 고객들이 많이 분포되어 있는 중앙에서는 동일한 비율로 나누기 때문에 지나치게 세분화되어 경계값의 폭이 작아지는 경향이 있다. 10등급으로 구분하는 경우 지나치게 세분화되어 경계값이 뚜렷하지 않는 경향이 있다. 군집분석으로 구분하는 방법은 차이가 있는 그룹을 분류할 수 있으나, 속하는 그룹의 특징과 위치를 쉽게 알 수 없는 단점이 있다. 본 논문은 우리가 제안한 6등급 방법과 5등급과 10등급으로 구분하는 방법과 비교할 것이다. Figure 3.1은 우리의 제안 모델과 5등급모델, 10등급모델의 분포를 비교하는 그림이다. 제안 모델은 상위 그룹과 하위 그룹을 자세히 세분화하고 있음을 알 수 있다.

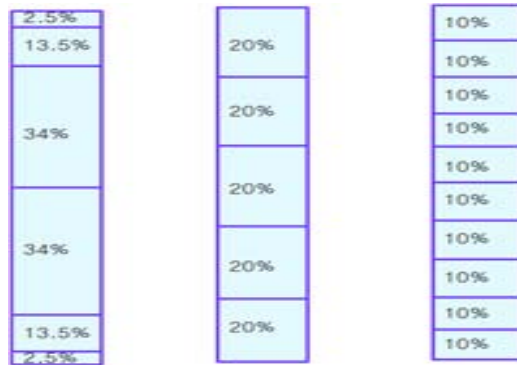


Figure 3.1 Distributions of proposed model, 5 level model, and 10 level model

4. 자료분석

제안된 등급부여 모형을 적용하기 위해 NDSL 자료를 사용하였다. NDSL은 과학기술분야 논문, 특허, 연구보고서, 동향분석, 산업표준 정보를 제공하는 사이트이다. NDSL의 1년간 로그인 이용자는 2012년 5월 31일 기준 57,476명이나 조사대상 기간인 2011년 6월 1일에서 2012년 5월 31일 사이에 NDSL이 제공하는 콘텐츠 사용경험이 있는 가입자는 23,595명이다. 분석 대상은 조사 대상 기간 중 콘텐츠 사용경험이 있는 고객으로 하였다. 조사 변수 R은 조사시점 2012년 5월 31일 기준과 가장 최근에 접속한 날짜와 차이로 정의하였고, F는 조사기간 중 접속 횟수로 정의하였고, M은 NDSL에서 제공하는 다섯 가지 콘텐츠의 상세보기와 원문보기 합으로 정의하였다. 제안 모델과 5등급, 10등급의 경계 값을 추정한 결과 Table 4.1과 같다. Table 4.1의 결과를 살펴보면 다음과 같다. 제안 모델과 5등급모델은 겹치는 등급은 없으나 10등급모델은 너무 상세하게 나뉘므로써 겹치는 그룹이 F와 M요인에서 발생했다. 제안 모델에서는 F와 M요인에서 각각 0인 고객은 최하위 2.5%에 속하는 고객임을 알 수 있으나, 5등급모델에서는 하위 20%에 속하는 그룹으로 다른 값을 가진 고객과 구분할 수 없는 한계를 가짐을 알 수 있다.

Table 4.1 Values of R, F, and M

value	R			F			M		
	Proposed 1 Model	5level	10level	Proposed Model	5level	10level	Proposed Model	5level	10level
10			38			12			48
9			39 49			7 11			23 47
8			50 83			6			13 22
7			84 144			4			8 12
6	15		145 207	35		3	155		6 7
5	14 48	49	208 257	8 34	6	3	29 154	23	4 5
4	49 207	50 144	258 321	4 7	4 5	2	6 28	8 22	3
3	208 362	145 257	322 351	2 3	3	2	2 5	4 7	2
2	363 503	258 351	352 374	1	2	1	1	2 3	1
1	504	352	375	0	1	0	0	1	1

RFM함수 $RFM = a * R + b * F + c * M$ 에서 한국과학기술정보연구원은 중요도를 M 다음으로 F, 마지막으로 R로 선정하였기 때문에, 가중치를 추정하기 위해 제약조건으로 $b \geq c \geq a$ 를 사용하였다. 또한 가중치의 값을 중요도를 비교평가하기 위해 소수점 첫째자리까지 추정하기로 하였다. 따라서 가중치 추정을 위한 실험은 Table 4.2와 같이 8가지 경우가 발생한다.

Table 4.2 Experimental design of a, b, c

	a	b	c
Case 1	0.1	0.1	0.8
Case 2	0.1	0.2	0.7
Case 3	0.1	0.3	0.6
Case 4	0.2	0.2	0.6
Case 5	0.1	0.4	0.5
Case 6	0.2	0.3	0.5
Case 7	0.2	0.4	0.4
Case 8	0.3	0.3	0.4

Case 1에서 제안 모델과 5등급모델과 10등급모델에 대한 RFM의 분포가 Figure 4.1과 4.2, 4.3에 나와 있다. 제안 모델은 RFM의 분포가 5개 그룹으로 뚜렷하게 구분되지만 5등급모델의 분포는 구분 정도가 낮고 10등급모델의 분포는 그룹으로 구분되지 않는다는 사실을 보여 준다. 제안 모델은 다른 가중치에서도 잘 구분되고 있으나, 5등급모델은 다른 가중치에서는 그룹을 찾기 어려운 분포형태를 보였다. 10등급모델은 8가지 경우 모두 그룹을 찾기 어려운 분포형태를 보였다. 제안 모델은 분포를 이용하여 특징을 살려서 6등급으로 구분함으로써 RFM의 분포에서 쉽게 그룹을 나눌 수 있었고 5등급모델이나 10등급모델은 확일적으로 구분함으로써 RFM의 분포에서 쉽게 그룹을 나눌 수 없었다. Figure 2.1에서 설명되어 있는 것처럼 제안 모델이 5등급모델과 10등급모델보다 사영을 잘하고 있어서 우수하다는 사실을 알 수 있다.

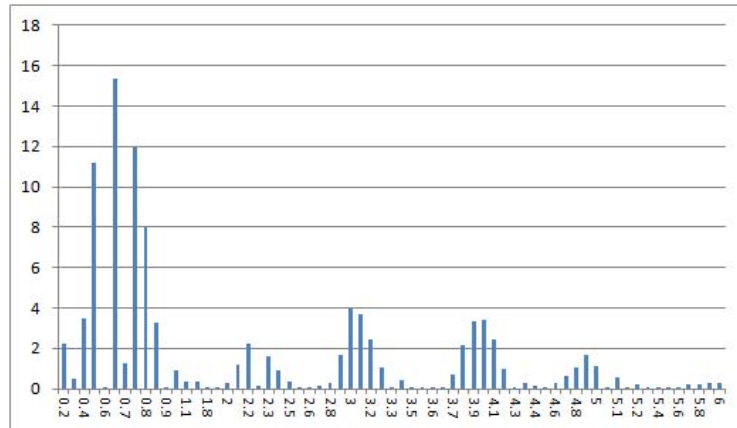


Figure 4.1 RFM with proposed model when $a=0.1, b=0.1,$ and $c=0.8$

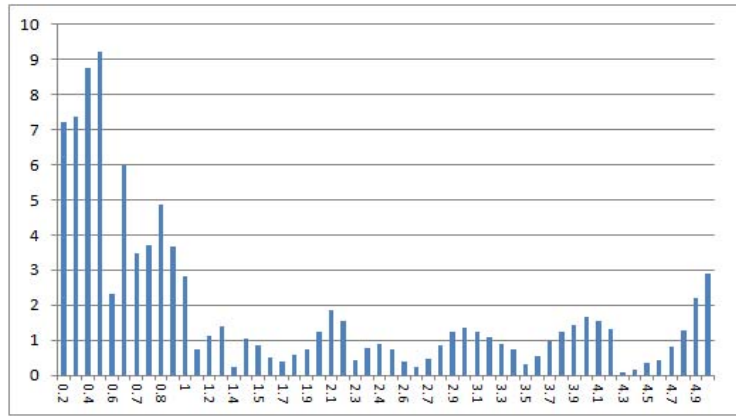


Figure 4.2 RFM with 5 level when $a=0.1$, $b=0.1$, and $c=0.8$

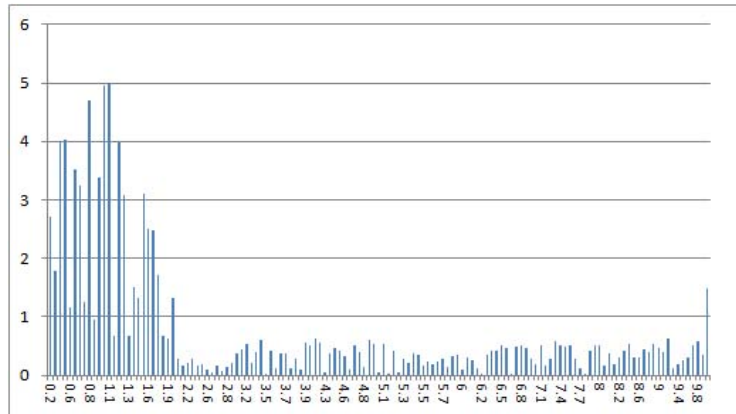


Figure 4.3 RFM with 10 level when $a=0.1$, $b=0.1$, and $c=0.8$

제안 모델을 이용하여 나머지 일곱 가지 가중치에 대해 추정된 RFM의 분포는 Figure 4.4에서 Figure 4.10에 나와 있다.

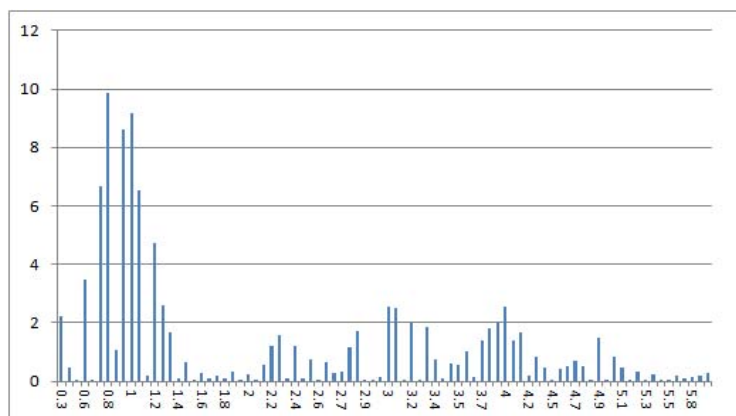


Figure 4.4 RFM with proposed model when $a=0.1$, $b=0.2$, and $c=0.7$

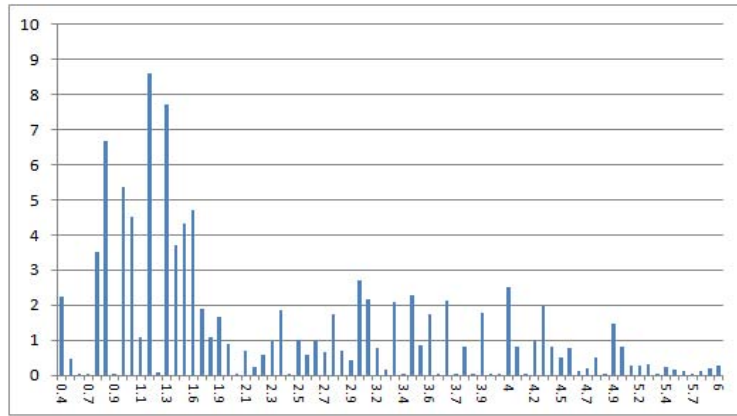


Figure 4.5 RFM with proposed model when $a=0.1$, $b=0.3$, and $c=0.6$

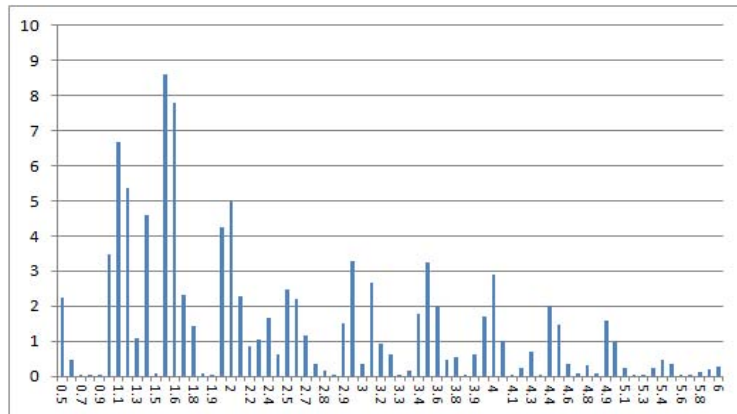


Figure 4.6 RFM with proposed model when $a=0.1$, $b=0.4$, and $c=0.5$

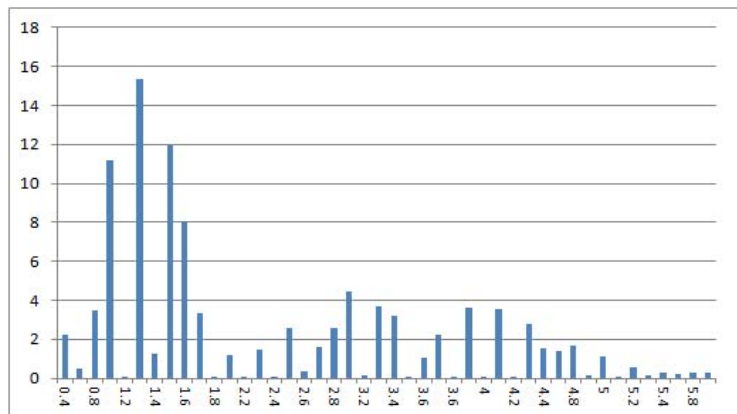


Figure 4.7 RFM with proposed model when $a=0.2$, $b=0.2$, and $c=0.6$

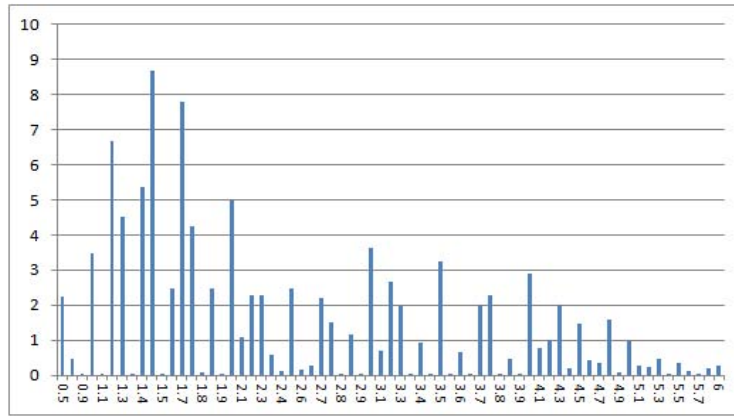


Figure 4.8 RFM with proposed model when $a=0.2$, $b=0.3$, and $c=0.5$

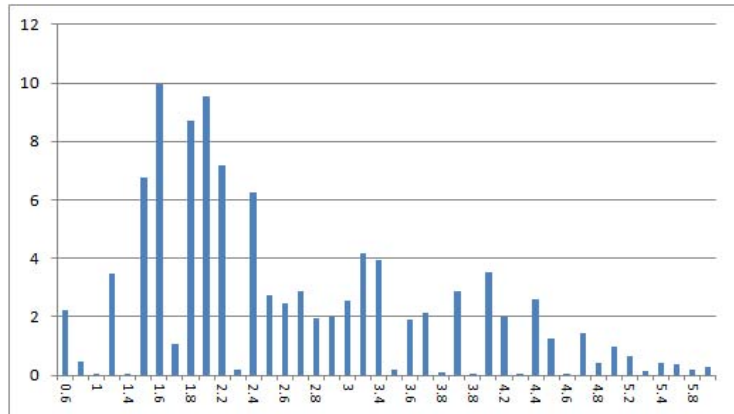


Figure 4.9 RFM with proposed model when $a=0.2$, $b=0.4$, and $c=0.4$

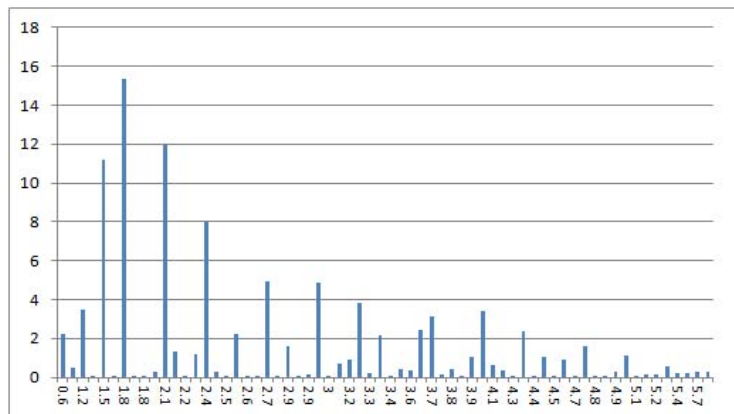


Figure 4.10 RFM with proposed model when $a=0.3$, $b=0.3$, and $c=0.4$

위의 분포를 보면 가중치를 변화함에 따라 RFM의 분포 형태도 달라짐을 알 수 있다. 목적에 따라 그룹의 수를 5개로 제한한다면 $a=0.1$ 과 $b=0.1$, $c=0.8$ 또는 $a=0.1$ 과 $b=0.2$, $c=0.7$ 이 타당한 모형으로 추정할 수 있으며, 그룹의 수를 7개로 제한한다면 $a=0.1$ 과 $b=0.4$, $c=0.5$ 가 타당한 모형으로 추정할 수 있으며, 그룹의 수를 3개로 제한한다면 $a=0.2$ 과 $b=0.2$, $c=0.6$ 이 타당한 모형으로 추정할 수 있다. 따라서 회사의 상황이나 목적에 따라 적절한 고객그룹의 개수를 결정하고 RFM의 분포를 보면서 가중치를 결정할 수 있다.

5. 결론

고객세분화는 시장을 일정한 기준으로 동질적인 세분시장으로 구분하는 것을 말한다. 목적은 과학적인 방법으로 고객을 정확히 이해하여 기업의 경쟁력을 높이는 것이다. 고객을 과학적이고 다면적으로 평가하기 위해, 고객이 얼마나 최근에 구매했는가를 평가하는 최근성과, 고객이 얼마나 자주 구매했는가를 평가하는 최빈성, 고객이 얼마나 많이 구매했는가를 평가하는 총 구매액 등 3가지 요인으로 고객을 세분화하는 방법인 RFM 모형이 개발되어 많이 사용되고 있다. RFM 모형은 R, F, M의 등급을 매기는 문제와 가중치 a , b , c 를 결정하는 문제가 있다. 본 논문은 등급을 매기는 방법을 정규분포를 이용하여 제안하고 NDSL 자료를 이용하여 5등급모델과 10등급모델을 비교하였다. 자료 분석 결과 제안 모델이 5등급모델과 10등급모델에 비해 고객그룹들을 쉽게 세분화할 수 있다는 사실을 알 수 있었다. 제안된 모델은 6등급모델로 대칭적으로 등급을 부여하고, 고객분포를 이용하기 때문에 고객특성을 잘 반영함으로써 경계 값들이 명확하게 구분되는 특징을 가지고 있다. 또한 등급 값을 보면 쉽게 어느 위치에 속하고 있는 지 알 수 있으며, 고객세분화 후 고객의 RFM값들의 확인으로 고객의 특성을 쉽게 알 수 있는 장점이 있다. 또한 우리가 제안한 모델을 이용하면 Figure 4.1에서 Figure 4.10까지 보듯이 RFM의 분포가 명확하게 구분되는 경향이 있기 때문에 목표에 따라 세분화 그룹을 적절히 구축할 수 있다. 향후 연구로는 군집분석 등 통계분석에 의한 등급구분 방법과의 비교 연구가 필요하며, 제안된 등급구분 방법을 지속적으로 적용하여 다양한 경우에서의 결과를 분석하는 것이 필요하다. 또한 가중치를 추정하는 방법에 대한 추가연구가 필요하다.

참고문헌

- Cho, K. H. and Park, H. C. (2011a). A study on insignificant rules discovery in association rule mining. *Journal of the Korean Data & Information Science Society*, **22**, 81-88.
- Cho, K. H. and Park, H. C. (2011b). A study on decision tree creation using intervening variable. *Journal of the Korean Data & Information Science Society*, **22**, 671-678.
- Cho, K. H. and Park, H. C. (2011c). A study on removal of unnecessary input variables using multiple external association rule. *Journal of the Korean Data & Information Science Society*, **22**, 877-884.
- Cho, K. H. and Park, H. C. (2012). A study on the segmentation of real estate customer using RFMP. *Journal of the Korean Data & Information Science Society*, **22**, 877-884.
- Hodgson, R. S. (1988). *Direct mail and mail order handbook*, 3rd Edition, Probus Publishing, Chicago.
- Lee, K. T. (2002). *Study on the customer segmentation techniques for LTV maximization in an eCRM environment*, Master Thesis, Jeonju university, Jeonju.
- Lee, S. Y., Choi, S. B., Kim, K. K. and Kang, C. W. (2004). Study on development the optimal RFM model for customer segmentation. *Journal of the Korean Data Analysis Society*, **6**, 1829-1840.
- Lee, Y. J., Kang, C. Y., Kim, K. K. and Choi, S. B. (2010). Customer segmentation and model evaluation using CFRM. *Journal of the Korean Data Analysis Society*, **12**, 3283-3293.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 281-297.
- Park, H. C. (2010). Weighted association rules considering item RFM scores. *Journal of the Korean Data & Information Science Society*, **21**, 1147-1154.

- Rogers, E. M. (1962). *The diffusion of innovations*, The Tree Press, New York.
- Seo, Y. M. and Kim, H. J. (2011). Empirical study for the adoption attitudes of new product between generations and countries. *Journal of the Korea Contents Association*, **11**, 405-415.
- Song, Y. S. and Kim, H. S. (2001). *New marketing*, Moonyoungsa, Seoul.

A study on proposing a method for grouping R, F, and M in RFM model

Gui-Yeol Ryu¹ · Young-Soo Moon²

¹Department of Computer Science, SeoKyeong University

²Department of ASTI Business Team, KISTI

Received 21 January 2013, revised 6 February 2013, accepted 13 February 2013

Abstract

The object of study is to propose a method for grouping R, F, and M in RFM model. Our model uses 6 levels using standard normal distribution. First level is upper 2.5% and second level next 13.5%, third level next 34%, fourth level next 34%, fifth level next 13.5%, sixth level next 2.5%. Values are symmetric and limits are clear. We compare proposed model with traditional 5 level model and 10 level model using NDSL data of KISTI. Proposed model divides most clearly the distribution of the RFM function for all cases of weights, because it uses the distribution of customers. Comparison studies of our model with grouping using cluster analysis and studies on weights of RFM model are needed.

Keywords: CRM, frequency, monetary, recency, segmentation.

¹ Associate professor, Department of Computer Science, SeoKyeong University, Seoul 136-704, Korea.

² Corresponding author: Senior researcher, Department of ASTI Team, KISTI, Seoul 130-741, Korea.
E-mail: youngsum@kisti.re.kr