

“RFT 모델 기반의 개선된 고객 세분화 방안”

Y-Set(와이셋)

우현우(정보산업공학과, 2014147049)

임선우(응용통계학과, 2015122029)

배수한(응용통계학과, 2015122020)

키워드: RFT, RFM, 로지스틱, CRM, 고객세분화, 패커스

1. 서론 Introduction

기업 경영에 있어 고객관계관리(CRM)는 **고객의 가치**를 이해하여 기업에 이윤이 되는 고객을 보유하게 해준다. 그러나 현실적으로 모든 고객이 높은 수익을 가져다주지 않는다(장민석 & 김형중, 2018). 따라서 **고객을 세분화**하여 특정 마케팅 활동의 대상이 되는 핵심 고객군을 선정하고, 그들 각각에 대한 효과적인 마케팅을 시행할 필요가 있다. 특히 기업의 입장에서 새로운 고객을 유인하는 것보다 **기존 고객을 유지하는 것이 더욱 중요하다**(Kotler & Armstrong, 2006). 기업이 새로운 신규고객을 유치하는데 드는 비용이 기존 고객을 관리하고 유지하는 비용의 5배 이상이 소요된다고 알려져 있다(전양주 외, 2011). 또한 모든 고객에게서 동일한 수준의 가치가 발생하는 것이 아니며 특정 고객들에게서 대부분의 가치가 발생한다. 따라서 **그 고객에게 집중**하여 세일즈와 마케팅을 전개할 필요가 있다(김상국, 2016).

본 연구는 온라인 배달용기 전문 쇼핑몰 '패커스'의 고객 데이터를 기반으로 고객 세분화 모델을 제시한다. **고전적 방법론인 'RFM 모델'**로 기존 고객을 분류하고, 그 한계점을 극복하고자 **이를 개선한 'RFT 모델'**을 새롭게 만들어 마케팅에 적극 활용가능한 새로운 고객 세분화 방안을 제시한다.

2. 연구방법 Methodology

2.1 전처리

패커스에서 제공한 고객 정보를 기반으로 RFM과 RFT 모델에 필요한 형태로 전처리를 진행한다. 본 연구에서는 2017년 1월 2일부터 2019년 6월 19일까지를 기준으로 서로 다른 고객 ID의 4955명을 대상으로 하고 있다(기준일자 이후부터 data누락 발생). 이때 R, F, M, 그리고 T에 대한 의미와 정의는 다음과 같다.

R(Recency, 거래 최근성): 해당 고객이 얼마나 최근에 상품을 구매했는가?

F(Frequency, 거래 빈도): 해당 고객이 얼마나 자주 상품을 구입했는가?

M(Monetary, 거래 규모): 해당 고객이 어느 정도로 금액을 총 사용했는가?

T(Time interval, 거래 간격): 해당 고객이 상품을 구매하는 시간 간격이 어느 정도인가?

[table1. R, F, M, T 정의]

변수	정의
R	2019년 6월 19일과 가장 최근 거래일의 차이 [day]
F	정해진 기간 내 총 구매횟수 [회]
M	정해진 기간 내 총 구매누적액 [원]
T	바로 다음 재구매까지 걸린 기간 중 최댓값 [month]

일반적인 RFM 모델을 통해 크게 신규 고객, 기존 고객, 이탈 고객으로 구분할 수 있다. 본 연구에는 기존 고객에 집중하는 데에 그 목적이 있어 기존 고객만을 연구 대상으로 하고 있다. 따라서 ($R \geq 4$ and $F=1$) 또

는 (R=1 and F>=4 and M>=4)에 해당하는 고객을 데이터셋에서 제거한다. RFM 모델을 개선한 RFT 모델에서는 T 값이 존재하기 위해 (F=1)에 해당하는 고객(일시적 고객)을 데이터셋에서 제거한다.

2.2 RFM 모델

[table2. RFM Score Board]

Score	R(Recency)	F(Frequency)	M(Monetary)
5	마지막 결제일이 20일 미만	총 주문횟수 20회 이상	총 소비액 243만원 이상
	마지막 결제일이 20일 이상 48일 미만	총 주문횟수 10회 이상 20회 미만	총 소비액 120만원 이상 243만원 미만
4	마지막 결제일이 48일 이상 122일 미만	총 주문횟수 5회 이상 10회 미만	총 소비액 47만원 이상 120만원 미만
	마지막 결제일이 122일 이상 365일 미만	총 주문횟수 2회 이상 5회 미만	총 소비액 18만원 이상 47만원 미만
3	마지막 결제일이 365일 이상	총 주문횟수 1회	총 소비액 18만원 미만

고객별로 각 요소별 Score를 주기 위해 R, F, M은 4분위수를 기준으로 구간을 나눈다. 그러나 M의 1과 2 Score를 구분하는 기준은 1회만 구매한 고객들의 총 소비액 평균으로 한다. 이를 토대로 고객 가치(Customer Value of RFM)는 각 Score의 가중치합으로 구성한 뒤 Logistic 모델에 적합시킨다.

$$W_R = \frac{\sum R}{\sum R + \sum F + \sum M} \quad W_F = \frac{\sum F}{\sum R + \sum F + \sum M} \quad W_M = \frac{\sum M}{\sum R + \sum F + \sum M}$$

$$Customer Value of RFM = \frac{1}{W_R} R + \frac{1}{W_F} F + \frac{1}{W_M} M = 0.28R + 0.36F + 0.36M$$

2.3 RFT 모델

RFM 모델링이 완료된 이후, 해당 모델을 통해 분류된 고객층마다의 1회당 평균 소비금액을 확인한다. **배달 용기와 같은 소매품의 특성상, 고객층마다 1회 구매할 때의 금액에 차이가 적다면 총 구매액(M)은 구매빈도(F)로 충분히 설명 가능하기 때문에 그 의미가 떨어진다.** 따라서 고객층마다 해당 금액에 차이가 없을 경우에 '총구매액(M)'을 '바로 다음 재구매까지 걸린 기간 중 최댓값(T)'으로 대체한다.

[table3. RFT Score Board]

Score	R(Recency)	F(Frequency)	T(Time interval)
5	마지막 결제일이 20일 미만	총 주문횟수 21회 이상	재구매까지 걸리는 시간의 최대 1개월
	마지막 결제일이 20일 이상 34일 미만	총 주문횟수 9회 이상 21회 미만	재구매까지 걸리는 시간의 최대 2개월
4	마지막 결제일이 34일 이상 63일 미만	총 주문횟수 5회 이상 9회 미만	재구매까지 걸리는 시간의 최대 3개월 이상 6개월 미만
	마지막 결제일이 63일 이상 365일 미만	총 주문횟수 3회 이상 5회 미만	재구매까지 걸리는 시간의 최대 6개월 이상 12개월 미만
3	마지막 결제일이 365일 이상	총 주문횟수 2회	재구매까지 걸리는 시간의 최대 12개월 이상

요소별 Scoring에 있어서 RFM모델과 동일한 방식으로 R과 F를 4분위수를 기준으로 구간을 나눈다. T는 1, 2, 3, 6, 12개월을 기준으로 구간을 나눈다. 이를 토대로 고객 가치(Customer Value of RFT)는 각 Score의 가중치합으로 구성한 뒤 Logistic 모델에 적합시킨다.

$$W_R = \frac{\sum R}{\sum R + \sum F + \sum T} \quad W_F = \frac{\sum F}{\sum R + \sum F + \sum T} \quad W_T = \frac{\sum T}{\sum R + \sum F + \sum T}$$

$$Customer Value of RFT = \frac{1}{W_R} R + \frac{1}{W_F} F + \frac{1}{W_T} T = 0.36R + 0.36R + 0.28T$$

3. 연구결과 Result

3.1 RFM 모델 해석 및 평가

[table4. RFM 척도의 Logistic Summary]

변수	Estimate	Standard error	p-value
Intercept	-1.2	0.1	<0.001
R value	0.17	0.04	<0.001
F value	0.46	0.07	<0.001
M value	0.48	0.06	<0.001
AIC	3812.8		

$$\text{Logistic RFM} = \frac{P(Y = i|X)}{1 - P(Y = i|X)} = -1.2 + 0.17 \times R \text{ value} + 0.46 \times F \text{ value} + 0.48 \times M \text{ value}, \quad i = 1, 2, 3, 4$$

가중치합으로 구성된 Customer Value of RFM을 로지스틱모델로 적합시킨 Logistic RFM은 위와 같다. 해당 모델은 고객의 R, F, M value에 따라 고객층을 분류한다. 5개의 고객층 구분 구간은 앞서 계산한 Customer Value of RFM을 기준으로 하였다. 구간별 누적분포는 아래와 같다.

[table5. Logistic Regression 적합에 있어 RFM 모델 분류 기준]

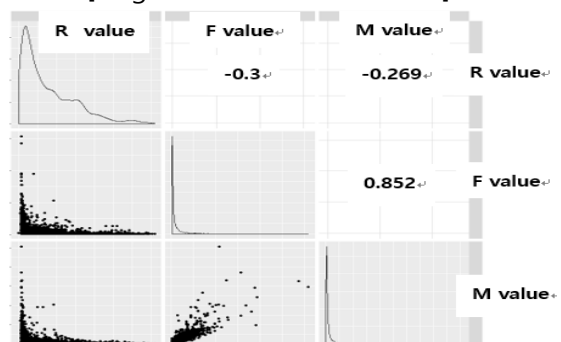
RFM score interval	기준	고객 수	비율	누적비율
5	RFM score 5점	9	0.002	1
4	RFM score 4점 이상 5점 미만	346	0.073	0.998
3	RFM score 3점 이상 4점 미만	554	0.116	0.925
2	RFM score 2점 이상 3점 미만	1304	0.275	0.809
1	RFM score 1점 이상 2점 미만	2535	0.534	0.534

그러나 각 고객의 1회 평균 소비금액들의 고객군별 평균이 유사함을 알 수 있다. 또한 F와 M간의 0.852의 높은 correlation의 선형성을 확인하였다. 따라서 금액과 관련한 M을 고객 분류 기준으로 삼는 데에는 문제가 있다고 할 수 있다. 따라서 M을 T 변수로 대체하였다.

[table6. RFM 고객군별 평균소비금액]

RFM score interval	기준	1회 평균 소비금액
5	RFM score 5점	146,979
4	RFM score 4점 이상 5점 미만	143,116
3	RFM score 3점 이상 4점 미만	132,332
2	RFM score 2점 이상 3점 미만	141,422
1	RFM score 1점 이상 2점 미만	99,566

[diagram1. Correlation Matrix]



3.2 RFT 모델 해석 및 평가

[table7. RFT척도의 Logistic Summary]

변수	Estimate	Standard error	p-value
Intercept	-4.2	0.27	<0.001
R value	0.7	0.07	<0.001
F value	0.7	0.06	<0.001
T value	0.35	0.06	<0.001
AIC	1449.8		

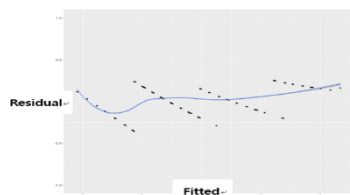
$$\text{Logistic RFT} = \frac{P(Y = i|X)}{1 - P(Y = i|X)} = -4.2 + 0.7 \times \text{R value} + 0.7 \times \text{F value} + 0.35 \times \text{T value}, \quad i = 1, 2, 3, 4$$

가중치합으로 구성된 Customer Value of RFT를 로지스틱모델로 적합시킨 Logistic RFT은 위와 같다. 해당 모델은 고객의 R, F, T value에 따라 고객층을 분류한다. 5개의 고객층 구분 구간은 앞서 계산한 Customer Value of RFT를 기준으로 하였다. 구간별 누적분포는 아래와 같다.

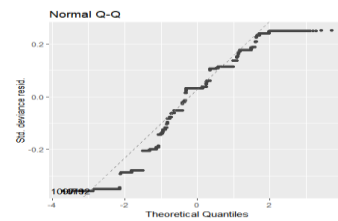
[table8. Logistic Regression 적합에 있어 RFT 모델 분류 기준]

RFT score interval	기준	고객 수	비율	누적비율
5	RFT score 5점	7	0.004	1
4	RFT score 4점 이상 5점 미만	275	0.144	0.996
3	RFT score 3점 이상 4점 미만	610	0.319	0.852
2	RFT score 2점 이상 3점 미만	791	0.413	0.533
1	RFT score 1점 이상 2점 미만	228	0.12	0.12

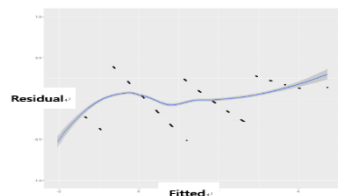
3.3 모델 검증



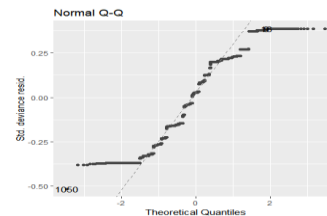
[diagram2. RFT Logistic Residual Plot]



[diagram3. RFT Logistic Normal QQplot]



[Diagram4. RFT Logistic Residual Plot]



[diagram5. RFT Logistic Normal QQplot]

위 그래프를 통해 Logistic Model의 결과물이 잔차의 정규성, 선형성, 등분산성 가정을 크게 위반하지 않음을 알 수 있다.

4. 결론 Conclusion

본 연구는 기존 고객을 고전적인 RFM 모델로 고객 세분화를 시행하였다. 이후 '패커스' 기업의 주 제품인 소매품이 가진 특성을 기반으로 해당 모델을 **RFT 모델로 개선하였다**. 이는 어떤 고객군이더라도 한 번 구매할 때에 소비하는 금액에 차이가 적기 때문에, F(Frequency, 구매빈도)가 M(Monetary, 총 소비금액)을 충분히 설명할 수 있기 때문이다. 따라서 **M을 대체할 T(Time interval, 다음 재구매까지 시간 간격의 최댓값)**를 고객 판단 변수로서 활용하였다.

본 연구에서 개발한 RFT 모델은 다음의 의미를 가진다.

1. 소매품 판매 기업의 고객을 분류함에 있어서 해당 모델은 '유의미한 고객 세분화 성능'을 보여준다. 기존의 RFM 모델은 가장 최하위 등급의 고객이 53%, 그 다음 하위 등급의 고객까지 81%의 누적 비율을 보이고 있다. 즉, 대부분의 고객을 하위 등급으로 판정해버리는 극단적인 결과를 나타내고 있다. 하지만 RFT 모델은 기존 고객들을 중간 수준의 등급에 고객의 수가 많도록 현실성 있게 분류해주고 있다. (중형 분포)
2. 기존의 RFM 모델에서 제기되는 문제점인 F와 M간의 '다중공선성을 해결'해주고 있다. 대부분 방문 빈도가 많을수록 자연스레 누적액이 높아지기 때문에 발생하는 이 문제는 실제 모델을 구현하는 데에 있어서 유의하지 않다는 결과가 자주 나온다. 따라서 RFT 모델을 사용한다면 이러한 유의성 문제를 해결할 수 있다.
3. 기존의 '패커스'에서 회원들에게 제공하는 고객 등급이 낮더라도 '우수한 고객'을 찾아낼 수 있다. 기존의 회원등급제는 분기동안의 총 결제 누적금액으로 결정이 되고 있다. 하지만 해당 모델을 통해 총 결제 누적금액이 적어서 낮은 회원 등급을 가진 고객이더라도, 이들의 '꾸준함'이라는 가치를 찾아낼 수 있다. 이는 단순히 금액만 보지 않고 시간적인 요소를 반영하여 기업의 마케팅에도 충분히 활용할 수가 있다.

따라서 본 연구는 기존 RFM 모델의 **한계점을 개선한 새로운 고객 세분화 모델(RFT모델)**을 제시하고, 기업이 금액 이외의 요소를 활용한 **새로운 고객 분류 체계를 통해 마케팅과 전략 측면에서 도움**을 줄 것으로 기대한다.

5. 참고문헌 References

- 장민석, 김형중. (2018). 빅데이터를 활용한 은행권 고객 세분화 기법 연구. 한국디지털콘텐츠학회 논문지, 19(1), 85-91.
- Kotler P. & Armstrong G. (2006). Principles of marketing, (11th Ed.) Upper Saddle River: New Jersey: Prentice-Hall.
- 전양주, 김철중, 임수진. (2011.) NIPPON 코퍼레이션(일본기업의 성숙시장에서의 실패와 교훈), MJ 미디어.
- 김상국. (2016). 고객 세분화 프레임워크에 관한 연구. 한국컴퓨터정보학회 학술발표논문집, 24(2), 63-64.