

K-means를 적용한 고객 가치 사례 분석

Case Study of Customer Value Analysis using K-means

이동준^a, 장시환^b, 류종석^c, 최황용^c, 김성수^{d*}

Dong-Jun Lee^a, Si-Hwan Jang^b, Jong-Seok Ryu^c, Hwang-Yong Choi^c, Sung-Soo Kim^{d*}

^a Heungkuk Life Insurance, Assistant Manager, 68, Saemunan-ro, Jongno-gu, Seoul 03184, Republic of Korea

^b ETRI, Senior Researcher, 218, Gajeong-ro, Yuseong-gu, Daejeon, 34129, Republic of Korea

^c Division of Energy Resource and Industrial Engineering, Kangwon National University, Student, Chuncheon, 24341, Republic of Korea

^d Division of Energy Resource and Industrial Engineering, Kangwon National University, Professor, Chuncheon, 24341, Republic of Korea

Received 20 September 2024; Revised 10 October 2024; Accepted 10 October 2024

Abstract

Customer identification for company is very valuable for direct marketing and increase of profit to target the population who are to become most profitable customer to the company based on target customer analysis and customer segmentation. Customer value analysis involves seeking the profitable groups of customers through analysis of customer's attributes. Data mining techniques can help to accomplish to extract or detect hidden customer values and behaviors from big data.

The objective of this paper is to propose customer value analysis based on RFM (R: Recency, F: Frequency, M: Monetary) model to identify the profitable segments (top target customer) of customer based on customer' underlying characteristics. We use the case study of S-company (122 customers with 6639 transactions from 2017/09/01 to 2018/08/31) to show the procedure of customer value analysis based on RFM model. We show how we can make the scores of RFM attributes and segment customers. K-means is one of the most important technique in data mining. K-means is used for five group market segmentations based on valid index intra-cluster distance which is a popular and efficient data clustering method. Our experiments and simulation results show the 26 top target customers out of 122 customers. We also propose the product recommend system based on RFM model for efficient marketing strategy with high priority.

Keywords: Custer Value Analysis, Data Clustering, RFM model, K-means

1. 서론

1.1 연구의 배경 및 목적

고객관계관리(Customer relationship management, CRM)는 기업이 고객을 확보하고, 유지하며 이윤을 추구하기 위해 고객과 의미 있는 소통을 하고 이해하기 위한 경쟁력 확보 수단이다.

Ngai & et al. (2009)^[2]는 CRM에서 목표 고객 분석 및 고객 분류를 위한 식별(identification), 직접 마케팅을 통한 유도(attraction), 충

성고객 집중 마케팅, 불만 관리를 통한 유지(retention), 고객 생애 가치 및 판매 시장 분석을 통한 개발 및 발전(development) 시키는 선순환과정을 Fig. 1과 같이 표현하였다. 이를 달성하기 위해 분류(classification), 군집(clustering), 예측(forecasting), 회귀(regression), 순서 발견(sequence discovery), 시각화(visualization), 연관(association) 데이터 마이닝 기술을 적용할 수 있고 고객 관계 관리와의 상관 관계를 다음 Fig. 1과 같이 표시하였다.

* Corresponding author. Tel.: +82-33-250-6283

fax: +82-504-135-6768

E-mail address: kimss@kangwon.ac.kr (Sung-Soo Kim).

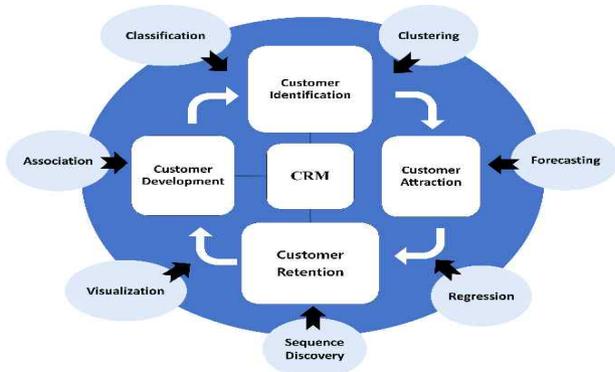


Fig. 1 CRM and Data Mining Technology (Ngai & et al. 2009)

Fig. 1의 고객 식별 단계에서 가장 중요한 것은 고객 가치 분석이다. 즉, 분류 또는 군집을 통한 세분화는 개별 고객 군에 대한 이해와 니즈 파악을 통한 상품 개발에 활용될 수 있다. 이를 통하여 전체 고객을 차별화와 핵심고객 마케팅 전략을 통하여 효율적 관리가 가능하다(Ngai & et al. 2009)^[12].

고객 식별을 위해서는 고객가치 분석 또는 시장 세분화를 위해 RFM (R: Recency, F: Frequency, M: Monetary) 모델과 데이터 클러스터링(군집) 방법을 적용할 수 있다. 이 RFM 모델은 고객이 최근에 거래할 수록 높게 평가하는 특징 R, 고객의 거래 횟수가 클수록 높게 평가하는 특징 F, 고객의 거래 대금이 클수록 높게 평가하는 특징 M을 적용한다. 이와 관련된 기존 연구는 다음과 같다.

Cheng & Chen (2009)^[2]과 Hosseini & et al. (2010)^[4]는 시장 세분화를 위해 K-means를 적용하였다. 특히, Cheng & Chen (2009)^[2]는 특징 R, F, M 데이터를 구간별로 약 20%씩 구분하여 가장 높은 점수 5점에서 가장 낮은 점수 1점까지 점수화 하여 데이터 클러스터링 방법을 적용하였다. Tsai & Chiu (2014)^[14], Kuo & et al. (2006)^[10], Chan (2008)^[1], Jonker & et al. (2004)^[5]는 유전자 알고리즘을 적용하였다. Chiu & et al. (2009)^[3]는 K-means와 파티클군집최적화(Particle swarm optimization, PSO)를 적용하였다. Niknam & Amiri(2010)^[13]는 파티클 군집최적화(PSO), 개미군집최적화(Ant colony optimization)와 K-means를 활용하였다.

이와 같이 다양한 알고리즘을 사용하여 고객 가치 분석과 시장 세분화를 활용한 다양한 연구가 진행되고 필요성이 제기되고 있다. 따라서 본 논문의 목적은 효율적이고 적용하기 쉬운 K-means를 사용하여 어떻게 고객 가치 분석과 시장 세분화를 할 수 있는지 사례를 적용하여 분석하고 설명하였다.

2절에서는 고객 세분화 데이터 클러스터링 문제의 수학적 모델을 설명하였다. 3절에서는 고객 가치분석을 위한 시장 세분화 방법과 이를 기반으로 하는 제품 추천시스템을 설명하였다. 4절에서는

S-company 사례를 활용한 실험결과 및 분석 결과를 서술하였다.

2. 고객 가치 시장 세분화 분석 문제

고객 가치 분석 또는 시장 세분화의 목적은 효과적인 마케팅 전략 수립의 핵심 정보가 되는 시장 세분화를 위해 RFM 모델을 적용한 데이터 클러스터링 방법으로 고객가치의 차이에 따라 클러스터링 하는 것이다. 이를 통하여 최근성 (Recency, R), 거래 횟수 (Frequency, F), 거래 대금 크기 (Monetary, M)가 우수한 고객 그룹을 식별한다. 이 우수고객 그룹을 중심으로 효과적인 마케팅 전략을 세울 수 있다.

본 연구에서 사례로 사용한 S-company 기업에 가장 이익을 많이 창출할 수 있는 고객(거래처)은 누구 인가? 우수 고객 그룹이 추가적인 주문을 할 수 있도록 상품 주문 시스템 개발을 통한 직접적인 마케팅으로 고객을 유도할 수 있다. 우수고객이 원하는 것은 무엇이고 어떻게 추가적인 상품 구매를 유도할 것인가? 고객 1대1 마케팅, 충성도 또는 불만족 관리 프로그램을 통한 고객을 유지할 수 있다. 거래 횟수와 규모가 확장되어 수익이 증가할 수 있도록 고객을 개발하고 발전시킬 수 있다.

고객 가치 분석 및 시장 세분화를 위한 데이터 클러스터링 문제 (즉, n 명의 고객을 K개의 그룹으로 클러스터링 하는 문제), Krishna & Murty (1999)^[9], Kim & et al, (2017)를 수리적으로 정립화 할 수 있다. 거래처 고객 데이터 집합 $X = \{x_1, x_2, \dots, x_n\}$ 는 데이터 i (x_i)로 구성된다($i=1,2,\dots,n$). 각각의 x_i 는 d차원(특징, attribute)으로 구성되는데 $x_{ij} = [x_{i1}, \dots, x_{id}]$ 는 데이터 i의 특징 데이터(고객이 최근에 거래할 수록 높게 평가하는 특징 R, 고객의 거래 횟수가 클수록 높게 평가하는 특징 F, 고객의 거래 대금이 클수록 높게 평가하는 특징 M 3가지) j의 값을 표현한 것이다. 또한, k ($k=1,2,\dots,K$)개의 클러스터 서브 집합 $C = \{C_1, C_2, \dots, C_k\}$ 로 서로 겹치지 않는 클러스터로 구성된다. 각 클러스터 집합은 적어도 한 개의 데이터가 존재한다. 데이터 클러스터링 해를 평가할 때 각 클러스터 내에서 평균과 소속된 데이터 사이의 유클리드 거리 (Intra-cluster distance)의 총합을 나타내는 식(1)을 적용하고 최소화 하는 것이고 제한식 (2)-(6)으로 나타낼 수 있다.

$$\text{Minimize } \sum_{k=1}^K S(k) \quad (1)$$

$$s.t \ w_{ik} = \begin{cases} 1 & \text{고객 } i \text{가 그룹 } k \text{에 포함된 경우} \\ 0 & \text{그렇지 않을 경우} \end{cases} \quad (2)$$

$$\sum_{k=1}^K w_{ik} = 1, i = 1, 2, \dots, n \quad w_{ik} \in \{0, 1\} \quad (3)$$

$$\sum_{k=1}^K w_{ik} \geq 1, k = 1, 2, \dots, K \quad (4)$$

$$C_{kj} = \frac{\sum_{k=1}^K w_{ik} x_{ij}}{\sum_{k=1}^K w_{ik}} \quad (5)$$

$$S(k) = \sum_{i=1}^n w_{ik} \sqrt{\sum_{j=1}^d (x_{ij} - c_{kj})^2} \quad (6)$$

만약 고객 i (x_i)가 그룹 k 에 포함되었을 경우, 의사결정변수 w_{ik} 를 1로 표시하고 그렇지 않을 경우 0으로 표시하여 식(2)와 같이 정의할 수 있다. 데이터 클러스터링 해 표현 매트릭스를 $W=\{w_{ik}\}$ 로 나타낼 수 있다. x_i 가 하나의 고객 클러스터 k 에 포함되는 여부에 따라 식(3)과 같이 표현할 수 있다.

식(4)는 고객 클러스터 k 에 적어도 하나 이상의 고객 데이터 i (x_i)가 포함되어 있는 것을 표현하였다. 식(5)의 C_{kj} 는 $C_{kj}=(C_{k1}, C_{k2}, \dots, C_{kd})$ 의 클러스터 k 에서 특징 데이터 j 의 평균값을 나타낸다. 식(6)은 클러스터 k 의 거리(각 고객 클러스터 내에서 평균과 소속된 고객 데이터 사이의 거리)의 합을 나타낸 것이다.

고객가치분석 문제는 RFM 모델을 적용하여 고객을 세분화하고 중요 타겟 고객(target customer, most profitable customer)을 구별화하여 효과적인 마케팅 전략을 수립하는데 적용할 수 있다. Recency (R)는 최신 거래가 많을 수록 중요 고객 (모든 고객의 최신 거래를 20%씩 등급화 하여 1~5로 점수화)이다. Frequency (F)는 일정 기간 동안 거래 횟수가 많을 수록 중요 고객 (모든 고객의 거래 횟수를 20%씩 등급화 하여 1~5로 점수화)이다. Monetary (M)는 거래 대금이 많을 수록 중요 고객 (모든 고객의 거래 대금을 20%씩 등급화 하여 1~5로 점수화)이다. 예를 들어, Table 1은 본 논문에서 사용할 S-company의 122개 거래처 중 10명의 고객 $G_1 \sim G_{10}$ 의 3가지 특성을 점수화 한 것이다. G_6 거래처 고객은 RFM 점수가 (5, 5, 5)점으로 가장 우수한 거래처로 예상할 수 있다. G_9 고객은 RFM 점수가 (2, 1, 1)점으로써 기업의 가장 우수하지 못한 거래처인 것으로 예상할 수 있다.

Table 1 RFM Attribute Score of Guest $G_1 \sim G_{10}$

Guest ID	R	F	M
G_1	3	2	3
G_2	5	4	4
G_3	4	4	5
G_4	3	1	2
G_5	2	4	5
G_6	5	5	5
G_7	2	3	4
G_8	4	4	4
G_9	2	1	1
G_{10}	3	1	2

3. 시장 세분화 방법과 제품 추천시스템

3.1절에서는 고객 세분화를 위해 데이터 클러스터링을 어떻게 적용할 수 있는지를 서술하고 3.2절에서는 3.1절을 통하여 얻어진 군집 정보에 따른 적합한 상품 추천 과정을 설명한다.

3.1 시장 세분화 방법

시장 세분화 수, 즉 고객 그룹 수(K) 결정 후 고객 그룹 해를 평가할 때는 같은 그룹 내의 평균과 데이터 간의 거리(Intra-cluster distance)를 평가 기준으로 식(1)~(6)을 적용하여 Fig. 2의 K-means 데이터 클러스터링을 적용한다. K-means는 초기해를 임의적으로 선택한다. Fig. 3>과 같이 $n=10$ 개의 거래처 고객이 $K=3$ 개의 그룹으로 랜덤하게 할당하여 생성하고 해 표현을 할 수 있다.

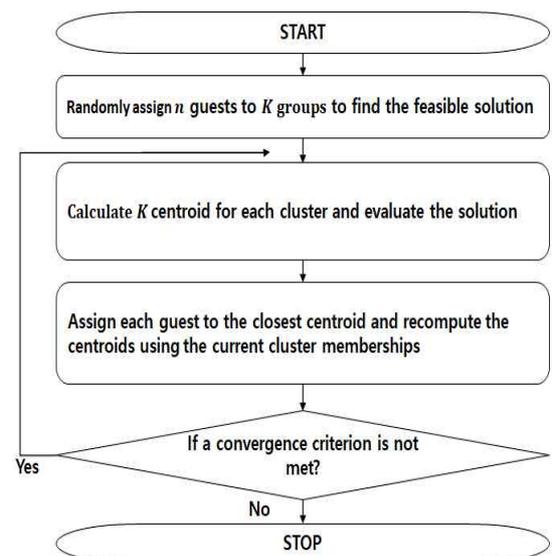


Fig. 2 K-means for Market Segmentation

Cluster	G_1	G_2	G_3	G_4	G_5	G_6	G_7	G_8	G_9	G_{10}	Evaluation value
1	0	0	0	1	0	0	0	0	1	1	intra-cluster distance 9.2716
2	1	0	0	0	0	0	1	0	0	0	
3	0	1	1	0	1	1	0	1	0	0	

Fig. 3 Solution Representation of 10 Guests for Three Clusters

각 그룹에 소속된 거래처 고객 데이터들의 특징 R, F, M의 평균과 소속 R, F, M 데이터 간의 거리의 합을 평가기준으로 현재의 해를 평가하여 식(1)을 최소화할 수 있는 해를 탐색한다. 거래처 고객 i 의 특징 데이터 x_{ij} 를 각 그룹 k 의 특징 j 의 평균 c_{kj} 을 기준으로 재할당하여 새로운 그룹 해를 탐색하여 다시 평가한다. 이러한 과정들을 통하여 더 좋은 새로운 해가 탐색 되지 않을 때 (각 그룹의 평균점 c_{kj} 가 변화하지 않을 때)까지 반복하여 해를 탐색한다.

Table 1 RFM 특성 점수에 적용하여 식(1)~(6)을 설명하면 다음과 같다. $n=10$ 명의 고객이 $K=3$ 개의 그룹으로 클러스터링 한다고 할 때, 만약, Fig. 3과 같은 해 표현과 같이 G_4, G_9, G_{10} 이 그룹 1, G_1, G_7 이 그룹2, 그리고 G_2, G_3, G_5, G_6, G_8 이 그룹 3에 포함될 때 $w_{12}, w_{23}, w_{33}, w_{41}, w_{53}, w_{63}, w_{72}, w_{83}, w_{91}, w_{101}$ 은 1로 식(2)를 만족한다. 10명의 고객이 그룹 1, 2, 3 중에 한 그룹에만 소속되므로 식(3)을 만족한다. $\sum_{i=1}^{10} w_{i1} = 3, \sum_{i=1}^{10} w_{i2} = 2, \sum_{i=1}^{10} w_{i3} = 5$ 이므로 한 그룹에 적어도 1명의 고객 이상이 소속되므로 식(4)를 만족하게 된다. 각 그룹 k 의 3가지 R, F, M 특징 j 의 평균값 C_{kj} 를 식 (5)로 계산한다. 3개의 고객 그룹과 3개의 특징이 존재하므로 $C_{11} \sim C_{13}, C_{21} \sim C_{23}, C_{31} \sim C_{33}$ 의 9개의 평균이 계산된다. $C_{11}=2.67, C_{12}=1, C_{13}=1.67, C_{21}=2.5, C_{22}=2.5, C_{23}=3.5, C_{31}=4, C_{32}=4.2, C_{33}=4.6$.

각 그룹의 각 특징의 평균과 소속 데이터와의 클러스터 내의 거리 $S(k)$ 를 식(6)에 따라 각 고객 그룹(1 ~ 3)의 $S(1), S(2), S(3)$ 값이 계산된다. 즉, 그룹 1의 특징의 평균 (C_{11}, C_{12}, C_{13}), (2.67, 1, 1.67)과 소속 고객 G_4, G_9, G_{10} 데이터와의 클러스터 내의 거리 $S(1) = 1.8856$ 을 식(6)으로 계산한다. 그룹 2의 특징의 평균(C_{21}, C_{22}, C_{23}), (2.5, 2.5, 3.5)과 소속 고객 G_1, G_7 데이터와의 클러스터 내의 거리 $S(2)=1.7321$ 를 식(6)으로 계산한다. 그룹 3의 특징의 평균 (C_{31}, C_{32}, C_{33}), (4, 4.2, 4.6)와 소속 고객 G_2, G_3, G_5, G_6, G_8 데이터와의 클러스터 내의 거리 $S(3) = 5.6539$ 를 식(6)으로 계산한다. 데이터 클러스터링 해의 평가 값은 식(1)을 사용하여 $S(1)+S(2)+S(3)=9.2716$ 로 계산된다.

Table 1의 거래처 $G_1 \sim G_{10}$ 의 10개의 데이터의 RFM 3개 특징 데이터에서 3개의 그룹 각 특징의 평균 C_{11}, C_{12}, C_{13} , (2.67, 1, 1.67), C_{21}, C_{22}, C_{23} , (2.5, 2.5, 3.5), C_{31}, C_{32}, C_{33} , (4, 4.2, 4.6)까지의 거리를 각각 계산한다. 각 데이터에서 가장 가까운 평균이 소속된 클러스터에 각 데이터를 포함시켜 해를 재구성하고 그 해를 평가하여 기존 해와 비교한다. 더 이상 좋은 해가 탐색되지 않을 때 (즉, 각 클러스터 중심 평균이 변하지 않을 때)까지 계속 진행한다.

3.2 제품 추천시스템

Liu & Shih (2005)^[11]는 RFM 모델 기반으로 상품추천시스템을 제안하였다. 본 논문에서도 Fig. 4와 같이 3.1절의 고객 가치분석을 위한 RFM 모델 시장 세분화를 기반으로 하는 제품 추천시스템을 제안한다. 거래처 고객의 데이터를 수집하고 각 고객의 RFM 특징을 점수화 한다. RFM 각 특징의 가중치를 기업 전략 또는 여러 관계 부서의 담당자들이 협의하여 결정한다. 결정된 가중치 적용 후 RFM 모델 기반 고객 가치 분석을 통하여 군집화 한다. 군집 순위 결정 후, 각 군집에 적합한 상품을 각 군집에 소속된 거래처 고객에게 추천할 수 있다. 즉, 새로운 거래처 또는 잠재 고객에게 고객 가치 순위가 결정된 군집의 RFM 점수 기준을 알려 준 뒤, 이를 바탕으로 어느 군집에 포함될 수 있는지 결정하고 해당 군집에 포함된 고객들이 가장 많이 구입했던 상품을 탐색하여 추천하거나 가장 유사한 기존 고객들이 가장 선호했던 상품을 추천할 수 있다. 또한, 새로운 고객의 거래 데이터도 추가적으로 수집되어 추후 분석에 적용된다.

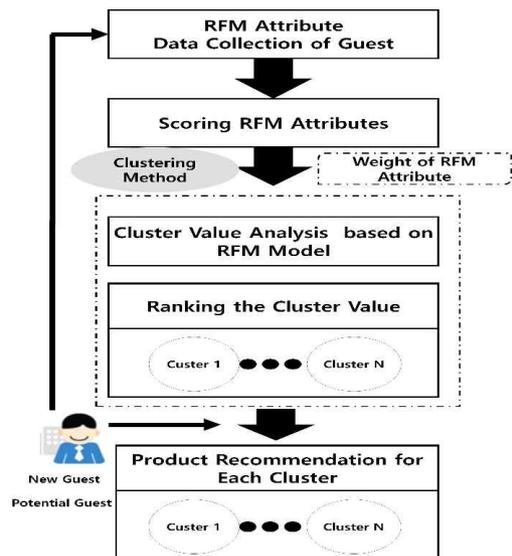


Fig. 4 Product Recommendation based on Customer Value Analysis

4. S-company 사례 적용 및 분석

본 절에서의 S-company 실제 사례 (Kim, S et al. 2019^[8], Kim, J. et al. 2018^[6]) 실험은 윈도우10 Pro 프로세서: Intel(R) Core™ i7-8700K CPU @ 3.70GHz 3.70 GHz 메모리(RAM): 16GB, 64비트 운영 체제, x64 기반 프로세서 운영 체제, Visual C++ 환경에서 실험 하였다.

S-Company의 2017/09/01~2018/08/31 기간 동안 122개 고객 거래처의 6639회 거래를 분석하였다. 122개 거래처 중 하나인 거래처 고객 #1번의 경우 1년 동안 Table 2와 같이 총 37회 거래가 있었고 같은 출고 일자가 9개 중복(2017년 11월7일 2개, 12월26일 2개, 2018년 1월22일 3개, 1월30일 2개, 3월26일 3개, 5월25일 2개, 7월9일 2개)되어 같은 거래처에서 동일한 날짜 하루에 일어난 중복 거래는 1회로 간주하고 거래 대금도 1회로 합산하였다. 즉, 총 28회 거래 횟수 Frequency가 있는 것으로 계산하였다. 거래처 고객 1번의 1년간 총 거래 대금 Monetary는 55,821,000원이다.

122개 거래처 고객의 Recency 계산 방법은 #1~#122번 고객의 거래가 최근까지도 날짜별로 빈번히 발생하여 하루 차이로 Recency의 단계를 차별화하는 것은 문제가 있다고 판단하였다. 판단 근거는 가장 마지막 최근 거래일인 2018년 8월31일 하루에 총 122개 거래처 중 17개의 거래처가 거래해서 Recency가 높은 것으로 평가하는 것은 문제가 있다. 즉, 오랫동안 거래를 하지 않다가 2018년8월31일 최근 하루에 갑자기 거래했다고 Recency 평가를 좋게 평가한다는 것은 타당하지 않다. 따라서, Recency에서 122개의 거래처를 5등급으로 차별화하기 위해 각 거래처에 대하여 최근 3회까지의 거래가 일어난 시점부터 현재까지의 평균 일자로 차등화하였다. 예를 들어 고객 거래처 #1번은 최근 3회 거래가 2018년 8월 6, 23, 29일 발생하였다. 예를 들어 고객 거래처 #1번은 최근 3회 거래가 2018년 8월 6, 23, 29일 발생하였다. 최근 3회까지의 거래가 일어난 시점부터 현재까지의 평균 Recency $6+23+29=58$ 이고 $58/3=19.33$ 로 계산하여 Recency는 8월 19일을 기준으로 결정하였다.

결과적으로 거래처 고객 #1의 Table 2의 37회 거래를 데이터로 활용하여 3가지 특성을 정리해 보면 최근성 R은 2018년8월19일, 거래 횟수 F는 28회, 거래 대금 M은 55,821,000원으로 정리할 수 있다. 나머지 121개 거래처 고객도 #1 고객과 같은 방식으로 각각의 특성을 정리할 수 있다. Table 3은 S-company 122개 거래처 고객 중 G_1~G_10의 거래처 고객의 RFM 특성을 나타낸 것이다.

Table 2 Transactions of Guest G_1

Date of Transactions	Monetary
2017-09-14	1570000
2017-09-18	1780000
2017-09-26	1335000
2017-10-31	890000
2017-11-07	378000
2017-11-07	2100000
2017-11-13	1335000
2017-11-27	1780000
2017-12-26	1335000
2017-12-26	2100000
2018-01-09	890000
2018-01-22	1570000
2018-01-22	2100000
2018-01-22	2780000
2018-01-30	1335000
2018-01-30	2100000
2018-02-06	890000
2018-02-20	420000
2018-03-07	1170000
2018-03-20	1570000
2018-03-26	1780000
2018-03-26	2100000
2018-03-26	1570000
2018-04-09	1335000
2018-04-18	1780000
2018-04-23	189000
2018-04-30	445000
2018-05-08	1170000
2018-05-25	1335000
2018-05-25	1570000
2018-06-11	420000
2018-07-09	1170000
2018-07-09	1780000
2018-07-23	2670000
2018-08-06	5600000
2018-08-23	1390000
2018-08-29	89000

Table 3 RFM Attributes for Guest $G_1 \sim G_{10}$

Guest ID	R	F	M
G_1	2018-08-19	28	55,821,000
G_2	2018-08-24	41	88,195,000
G_3	2018-08-23	44	151,536,000
G_4	2018-08-17	19	40,222,000
G_5	2018-08-16	41	116,831,000
G_6	2018-08-24	90	290,598,000
G_7	2018-08-10	36	79,233,000
G_8	2018-08-22	41	88,976,000
G_9	2018-08-13	15	35,835,000
G_{10}	2018-08-19	19	45,220,000

이와 같이 S-Company의 122개 거래처 고객의 Recency (구간 2018. 3.29 ~ 2018. 8. 26), Frequency (구간 3회 ~ 99회), Monetary (구간 1,335,000원 ~ 372,289,000원)은 거래처 RFM 점수 기준에 따라 5등급으로 등급화한다. 즉, R, F, M 특징값은 122개 거래처 고객이 약 20%씩 나누어질 수 있도록 1~5 점수 값으로 결정한다. 예를 들어, 최근 거래 기준일이 2018년 8월 24일 이후이거나, 거래 횟수가 50회 이상일 때, 또한 거래 대금이 112,494,000원 이상인 상위 20% 거래처 고객에게 각각의 특징 R, F, M의 가장 높은 점수인 5점을 부여하였다. 이와 같이 특징 R, F, M의 4점(상위 40%~20%), 3점(상위 60%~40%), 2점(상위 80%~60%), 1점(상위 100%~80%) 순으로 점수를 부여하였다. Table 3의 데이터를 이와 같은 기준에 따라 122개 거래처 고객의 데이터를 1~5점으로 정리할 수 있는데, 본 논문 2절에서 예제로 사용한 Table 1은 S-company의 122개 고객 중 10개 $G_1 \sim G_{10}$ 의 거래처 고객의 RFM 특성을 1~5점으로 나타낸 것이다. R, F, M 각 특징의 가중치는 각 기업의 마케팅 전략과 상품에 따라 또는 관련 임직원들이 협의 통하여 결정할 수 있다 (Hosseini & et al. 2010^[4], Tsai & Chiu 2004^[14], Liu & Shih 2005^[11]). 본 논문에서는 Cheng & Chen (2009)^[2]와 마찬가지로 각 특징이 특별한 가중치로 정할 수 없는 경우 같은 가중치를 갖는 것으로 가정하였다.

이와 같이 S-company 총거래 6639를 122개 고객 거래처별로 특성 R, F, M을 정리하여 Fig. 5와 같이 3차원 그래프로 표시하였다. 특성 R(y축), F(x축), M(z축)간의 상대적 비교를 위해 Fig. 6 Monetary와 Frequency 2차원, Fig. 7 Monetary와 Recency 2차원, Fig. 8 Frequency와 Recency 2차원 그래프로 표시하였다. Fig. 6 ~ Fig. 8을 통하여 122개 거래처 고객 중 S-company에 기여도가 높을 것으로 예상되는 고객 #13, 기여도가 중간으로 예상되는 고객 #14, 기여도가 낮을 것으로 예상되는 #52을 1차적으로 분석할 수 있다. 122개 거래처 중 13번 거래처는 M이 매우 높은 372,289,000원이고 F가 매우 높은 99

회, R은 2018년 8월 24일로 상대적으로 최근 거래가 이루어져 S-company에 가장 수익성이 높을 것으로 예상되는 고객이다. 52번 거래처는 M이 1,335,000원이고 F가 3회, R은 2018년 3월 29일로 상대적으로 최근 거래가 이루어지지 않아 가장 수익성이 낮을 것으로 예상되는 고객이다. 거래처 14는 M이 118,629,000원으로 비교적 높은 범위로 상대적으로 높고, F가 60회로 비교적 높으나, R은 2018년 7월 25일로 상대적으로 낮아 최근 거래가 이루어지지 않았으므로 새로운 거래가 이루어질 수 있도록 유도하는 마케팅 전략 적용으로 R을 올릴 수 있는 조치를 취해야 한다. 이와 같은 분석은 데이터의 수와 특징 수가 적을 경우, 부분적으로 가능하나 데이터 수와 특징 수가 커질 경우, 상세한 분석에는 한계가 있다. 따라서, 데이터 수와 특징 수가 커져 빅데이터가 되면 거래처의 그룹별 데이터 분석이 필요하다.

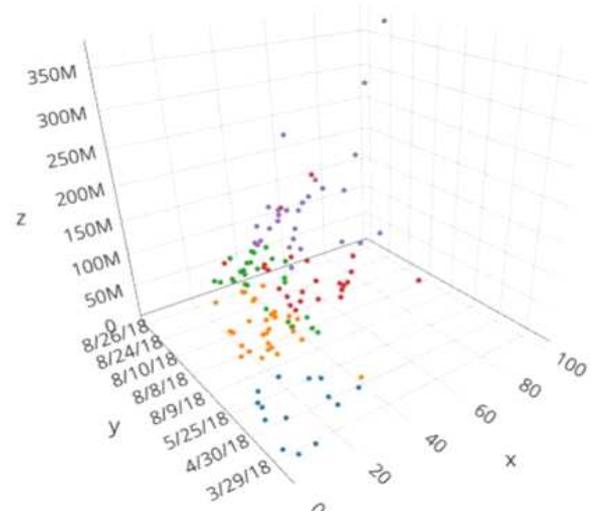


Fig. 5 Recency (Y axis), Frequency (X axis), Monetary (Z axis) Graph

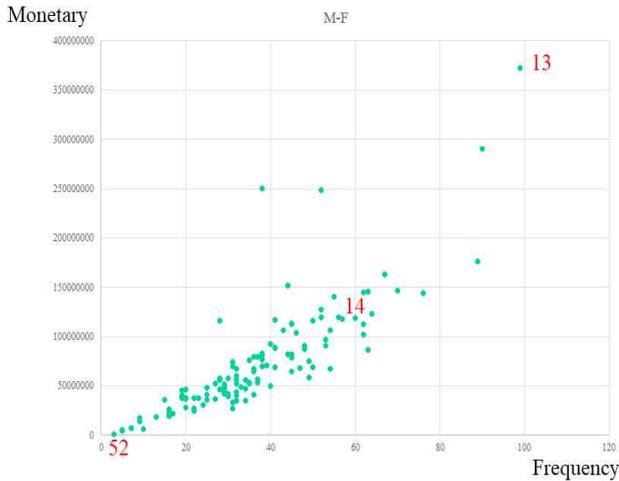


Fig. 6 Monetary & Frequency Graph

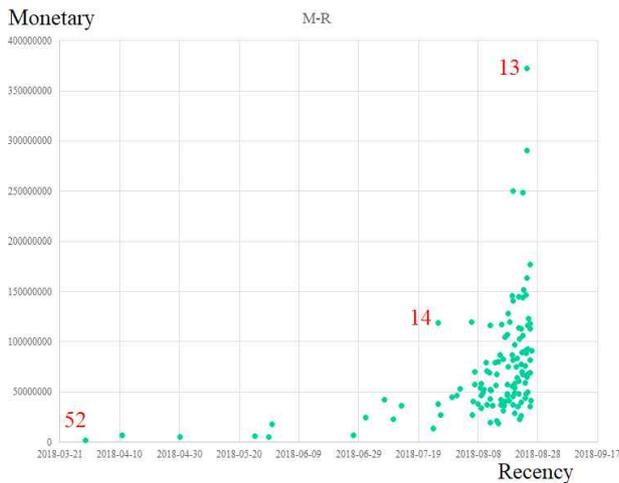


Fig. 7 Monetary & Recency Graph

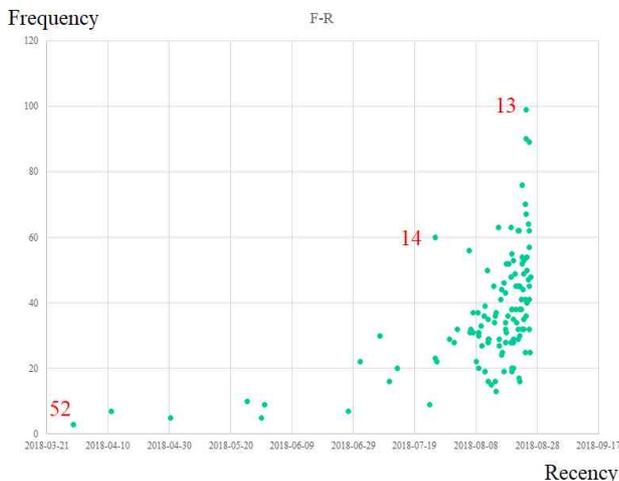


Fig. 8 Frequency & Recency Graph

본 논문에서 가장 보편적이고 효율적인 데이터 클러스터링 방법인 K-means를 사용하여 S-Company의 122개 거래처 고객을 RFM 모델을 적용하여 실험 분석하였다. Table 4는 S-company의 122개 고객 거래처를 RFM 모델을 기반으로 5개 그룹으로 고객 세분화를 한 결과이다. 분석에 따르면 C_4 그룹(Very high) 거래처는 26개 거래처가 포함되어 전체 고객 122 거래처 중 약 21.3%를 차지하였다. 26개 거래처의 최근성 R은 5점 만점 기준 평균 4.615385로 최근성이 높고, 거래 횟수 F는 평균 4.615385로 거래 횟수가 높으며, 평균 4.461538로 거래 대금 크기 M 거래 대금 총액이 큰 핵심 고객으로 중점적으로 관리해야 할 중요 대상 고객이다. 즉, RFM 평균 C_{41} , C_{42} , C_{43} 은 (4.615385, 4.615385, 4.461538)이고 원점으로부터의 거리는 7.906256으로 가장 멀다. C_2 그룹(Very low)은 전체 122 거래처 중 31개로 약 25.4%를 차지하였는데, R은 5점 만점 기준 평균 1.806452로 최근성이 낮고, F는 평균 1.16129로 거래 횟수가 낮으며 평균 1.225806로 M 거래 대금 총액이 낮다. 즉, RFM 평균 C_{21} , C_{22} , C_{23} 은 (1.806452, 1.16129, 1.225806)이고 원점으로부터의 거리는 2.472744으로 원점에서 가까워 S-company의 수익에 상대적으로 도움이 되지 못하는 거래처 그룹이다.

122개 각 거래처 고객의 그룹 소속은 Fig. 9로 표시할 수 있다. 5개의 거래처 그룹 $C_1 \sim C_5$ 은 1~5로 표시하였다. 즉, 가장 우수 그룹(Very high), 클러스터 C_4 그룹의 거래처 26개는 ‘4’으로 표시되고, 거래처 고객 번호는 # 2, 3, 6, 8, 11, 13,16, 22, 29, 32, 40, 45, 48, 51, 66, 68, 75, 79, 84, 91, 94, 96, 99, 109, 114, 115이다. Fig. 6 ~ Fig. 8에서 예상했던 거래처 #13이 포함되어 있다.

본 논문 1절에서 설명한 바와 같이 데이터 마이닝 기술 분류와 군집을 통하여 식별된 26개의 중요 거래처 고객은 중점 관리가 필요하고 추가적인 상품 구매 마케팅을 통한 고객 유도와 유지 및 발전시킬 수 있도록 한다. 가장 S-company에 도움이 되지 못하는 C_2 그룹 (Very low) 거래처는 ‘2’로 표시하였는데, Fig. 6 ~ Fig. 8에서 예상했던 거래처 #52이 포함되어 있다. 이 31개의 거래처 고객의 관리 및 유지 비용을 최소화할 계획이다. 거래처 #14는 우수한 C_3 그룹 (High)에 포함되어 있는데, 이 그룹에 포함된 거래처는 최우수 C_4 그룹(Very high)의 수준까지 올라올 수 있도록 적극적인 마케팅으로 RFM 특성을 향상시킬 필요가 있다.

Table 5과 Table 6은 5개 그룹의 특징 RFM의 1~5점수대와 실제 범위를 보여준다. 즉, 거래처 #13이 포함된 최우수 고객 그룹 C_4 의 26개 거래처 고객의 최근성 R은 2018. 8. 21. ~ 2018. 8. 26. (5점 스케일로 4-5점)이고, 거래 횟수 F는 40-99회 (5점 스케일로 4-5점)이고,

거래 대금 M은 67,775,000~372,289,000원 (5점 스케일로 3~5점)으로 최근성, 거래 횟수와 거래 대금이 모두 높다 (R ↑, F ↑, M ↑).

거래처 #14이 포함된 우수 고객 그룹 C₃의 18개 거래처 고객의 최근성 R은 2018. 7. 25. ~ 2018. 8. 20. (5점 스케일로 1~3점)이고, 거래 횟수 F는 38~63회 (5점 스케일로 3~5점)이고, 거래 대금 M은 69,077,000 ~250,414,000원 (5점 스케일로 3~5점)으로 거래 횟수와 거래 대금은 높은 편이나 최근 거래가 상대적으로 낮으므로(R ↓, F ↑, M ↑) 최근 거래를 높여 최근성을 올릴 수 있도록 그룹 C₃을 유도하는 것이 필요하다.

또한, S-Company에 가장 기여도가 낮다고 판단되고 거래처 #52이 포함된 그룹 C₂의 31개 거래처 고객의 최근성 R은 2018. 3. 29. ~ 2018. 8. 22. (5점 스케일로 1~4점)이고, 거래 횟수 F는 3~31회 (5점 스케일로 1~2점)이고, 거래 대금 M은 1,335,000~46,132,000원 (5점 스케일로 1~2점)으로 최근성, 거래 횟수와 거래 대금이 모두 낮다 (R ↓, F ↓, M ↓).

Table 4 Five Groups for 122 Guests

Centroid of Cluster	C ₁	C ₂	C ₃	C ₄	C ₅
R	1.65	1.80	2.44	4.61	3.81
F	2.50	1.16	4.44	4.61	2.77
M	2.85	1.22	4.38	4.46	2.66
Distance between centroid and origin	4.13	2.47	6.70	7.90	5.42
Rank of each cluster	Low	Very Low	High	Very High	Medium
No of guests for each cluster	20	31	18	26	27

Table 5 Five Groups of RFM Score for 122 Guests

			Minimum	Median	Maximum		Count
Five Groups	very low	Recency	1	2	4	R ↓	31
		Frequency	1	1	2	F ↓	
		Monetary	1	1	2	M ↓	
	low	Recency	1	2	2		20
		Frequency	2	3	3		
		Monetary	2	3	5		
	mini um	Recency	3	4	5		27
		Frequency	2	3	4		
		Monetary	1	3	4		
	high	Recency	1	3	3		18
		Frequency	3	5	5		
		Monetary	3	4	5		
	very high	Recency	4	5	5	R ↑	26
		Frequency	4	5	5	F ↑	
		Monetary	3	5	5	M ↑	

[1]	5	4	4	2	3	4	1	4	2	2	4	3	4	3	3	4	5	5	2	2	5	4	1	5	2	2	1	2	4	3
[31]	5	4	1	2	5	1	3	2	2	4	1	3	1	3	4	2	1	4	1	5	4	2	2	3	2	2	5	1	5	2
[61]	2	2	5	2	2	4	5	4	1	5	2	2	5	4	1	1	5	4	2	2	5	3	4	1	5	1	2	5	5	
[91]	4	5	3	4	5	4	1	3	4	1	2	1	5	3	3	2	5	3	5	3	5	2	4	4	1	3	4	2	5	
[121]	5	3																												

Fig. 9 Solution Representation for 122 Guests

Table 6 Five Groups of RFM Attributes for 122 Guests

			Mini mum	Media n	Max imum	Count
5 Groups	very low	Recency	2018 -03-29	2018-0 8-09	2018 -08-22	31
		Frequency	3	19	31	
		Monetary	1335 000	269240 00	4613 2000	
	low	Recency	2018 -07-07	2018-0 8-09	2018 -08-15	20
		Frequency	27	32	37	
		Monetary	4027 1000	526815 00	1161 01000	
	medi um	Recency	2018 -08-17	2018-0 8-21	2018 -08-25	27
		Frequency	25	32	49	
		Monetary	3501 9000	558210 00	8327 8000	
	high	Recency	2018 -07-25	2018-0 8-17	2018 -08-20	18
		Frequency	38	50	63	
		Monetary	6907 7000	105590 000	2504 14000	
	very high	Recency	2018 -08-21	2018-0 8-24	2018 -08-26	26
		Frequency	40	54	99	
		Monetary	6777 5000	113010 500	3722 89000	

5. 결론

본 논문에서는 RFM 모델 기반 고객가치분석을 위한 시장 세분화를 S-company 사례 (2017/09/01 ~ 2018/08/31 기간 동안 122개 거래처 고객의 6639회 거래 데이터를)를 활용하여 5개 그룹으로 클러스터링 하여 분석하였다. 가장 우수 그룹으로 식별된 26개의 중요 거래처 고객은 중점 관리가 필요하고 추가적인 상품 구매 마케팅을 통한 고객 유도와 유지 및 발전시킬 수 있다. 거래처 고객의 데이터를 수집하고 고객 세분화 RFM 모델 데이터를 수집하고 고객 세분화 RFM 모델 데이터 분석을 위한 기준치는 같은 것으로 가정하였고, 각 특징의 등급화 과정과 점수화를 설명하였다. 각 RFM 특성의 가중치는 각 회사의 전략과 시장 상황에 따라 임원과 각 부서의 협의에 따라 결정하는 것이 적절하다.

데이터 클러스터링 방법은 가장 보편적으로 사용되고 적용이 수월한 K-means를 적용하여 군집의 고객 가치 순위를 결정하였다. 이를 통하여 S-company에 추가적인 수익을 창출할 수 있는 우수 고객을 식별하여 효과적인 CRM 마케팅 전략 수립에 활용할 수 있다. 또한, 제안한 고객가치분석 방법과 시스템은 잠재 거래처 고객

들에게 분석된 군집의 정보를 활용하여 적절한 상품을 추천할 수 있다.

6. 감사의 글

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2024년도 문화체육관광 연구개발사업으로 수행되었음(과제명 : 중소 게임 기업의 게임 제작 검증 효율화를 위한 AI 기반의 대규모 게임 자동검증 기술 개발, 과제번호 : RS-2024-00393500, 기여율: 100%)

References

- [1] Chan, C. C. H., 2008, Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer, *Expert systems with applications*, 34(4), 2754-2762.
- [2] Cheng, Chen., 2009, Classifying the segmentation of customer value via RFM model and RS theory, *Expert systems with Application*, 36 (2009) 4176-4184.
- [3] Chiu, C. Y., Chen, Y. F., Kuo, I. T., Ku, H. C., 2009, An intelligent market segmentation system using k-means and particle swarm optimization, *Expert Systems with Applications*, 36(3), 4558-4565.
- [4] Hosseini, S. M. S., Maleki, A., Gholamian, M. R., 2010, Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty, *Expert Systems with Applications*, 37(7), 5259-5264.
- [5] Jonker, J. J., Piersma, N., Van den Poel, D., 2004, Joint optimization of customer segmentation and marketing policy to maximize long-term profitability, *Expert Systems with Applications*, 27(2), 159-168
- [6] Kim, J., Lee, D., Kang, B., Kim, S., 2018, Big Data Optimization Analysis for Market Segmentation, Poster in Department of System and Management Engineering, Kangwon National University, 2018. 11.
- [7] Kim, S., Baek, J., and Kang, B., 2017, Hybrid Simulated Annealing for Data Clustering, *Journal of Society of Korea Industrial and Systems Engineering*, 2017, Vol. 40, No. 2, pp. 92-98.
- [8] Kim, S., Kang, B., and Lee, D., 2019, Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer, *Expert systems with applications*, 34(4),

2754-2762.

- [9] Krishna, K. and Murty, M.N., 1999, Genetic K-means algorithm, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 29, No. 3, pp.433-439.
- [10] Kuo, R. J., An, Y. L., Wang, H. S., Chung, W. J., 2006, Integration of self-organizing feature maps neural network and genetic K-means algorithm for market segmentation, Expert systems with applications, 30(2), 313-324.
- [11] Liu, Shih., 2005, Integrating AHP and data mining for product recommendation based on customer lifetime value, Information & Management 42 (2005) 387-400.
- [12] Ngai, et al., 2009, Application of data mining techniques in customer relationship management: A literature review and classification, Expert systems with Application, 36 (2009) 2592-2602.
- [13] Niknam, T., Amiri, 2010, An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis, Applied soft computing, 10(1), 183-197.
- [14] Tsai, C. Y., Chiu, C. C., 2004, A purchase-based market segmentation methodology, Expert Systems with Applications, 27(2), 265-276.