

RFID 데이터를 이용한 고객 쇼핑 동선 패턴 분석

양승준* · 정인철** · 권영식***

Shoppers' Shopping Path Pattern Analysis using RFID Data

Seungjoon Yang* · In-Chul Jung** · Young S. Kwon***

■ Abstract ■

As the retail industry has been challenged by stiff competition, the retailer becomes more interested in better understanding consumers' in-store behavior to gain and sustain competitive advantage. Consumers' shopping paths provide valuable clues to understanding customers' in-store behavior, which has been a long standing research issue in business.

This study is to explore the shopping path patterns in a grocery using RFID technology and clustering method. To this end, we designed the RFID systems, affixing active RFID tags to the bottom of grocery carts. The tag emit signal that is received by receptors installed at various location throughout the store. The RFID systems provide the time and location of the cart while consumers shop around the store. The point of sale data are matched with the cart movement records to provide a complete picture of each shopping path. To find the distinctive patterns of consumers' shopping paths, we proposed the distance-index matrix using dijkstra method and normalization method to conduct the clustering in order to handle the problem in measuring the similarity among shopping paths, which is raised by the spatial nature of consumer movement in a grocery. After analyzing the RFID data obtained in one of the groceries in a major Korean retailer, we could successfully identify several distinctive patterns of shopping paths, which prove to provide the valuable implications for store management.

Keyword : Consumer Behavior, RFID, Shopping Path, Pattern Analysis, Clustering

1. 서론

이마트, 롯데마트와 같은 대형 할인유통매장은 점포수와 매출액이 크게 늘어나고 있으며 이에 따라 업체들 간의 경쟁은 더욱 치열해 지고 있다[2, 3]. 이러한 경쟁에서 선두를 지키기 위하여 유통업체들은 고객의 구매상품, 구매량, 상표번호도와 같은 고객의 구매 행동변수와 연령, 소득, 주거지역, 성별, 라이프스타일 등의 고객특성변수를 파악하고 이를 바탕으로 시장을 세분화하여 고객이 누구인지 어떤 구매 행동을 보이는지를 파악하는데 많은 노력을 기울이고 있다.

고객의 소비 의사결정에 관한 연구에 따르면 고객 소비 행동이 매대와 상품의 배치 등 매장환경에도 크게 영향을 받는 것으로 파악됨에 따라 고객 행동 분석을 통한 매장 환경 개선에도 관심이 높아지고 있는 상황이다(Newman et al.[9]). 이는 넓은 면적과 다양한 상품 구색을 갖추고 있는 대형 유통 매장의 특성상, 고객들이 어떻게 이동하며 쇼핑을 하는지에 대한 이해가 매장의 관리와 마케팅 전략 수립에 있어 중요한 정보로 사용될 수 있기 때문이다.

이에 따라 쇼핑동선을 파악하는 것은 고객의 소비행동을 파악하는 첫 단계로서 매우 중요한 의미를 가지고 있다. 이때 쇼핑 동선은 쇼핑물에서 고객이 입장하여 퇴장할 때까지 매장 안에서 상품의 구매를 위해 쇼핑하는 이동 경로를 말한다. 이동 경로를 추적하면 고객이 쇼핑을 하면서 구매한 상품과 매대 위치, 시간, 지나간 통로들을 파악하여 시간대별로 매장내의 어느 지역이 판매가 활발히 이루어지고, 고객들이 붐비는 매대 또는 오랫동안 체류하는 매대는 어느 곳인가를 알 수 있다. 최종적으로는 이런 정보를 근거로 매대 배치와 상품 배치를 변경하여 쇼핑 환경을 개선시켜 매출을 증대시킬 수 있다. 그러나 현재 유통매장에서는 이러한 정보를 얻기 위해, 일부 고객들을 대상으로 관찰자가 따라다니며 동선과 구매 행동을 관찰하는 스파게티 다이어그램 방식을 활용하고 있다. 그러

나 이는 시간과 비용 상의 제약 하에 제한된 샘플 고객을 대상으로 하기 때문에 개략적인 정보를 파악하는데 그칠 수밖에 없다.

최근 RFID 기술의 진보로 태그와 리더기의 성능이 안정화되고 가격도 저렴해 짐에 따라 유통매장에서의 RFID의 현실적인 적용이 용이해지고 있다. 따라서 본 연구에서는 국내의 한 대형 슈퍼마켓을 대상으로 매장에서 고객위치 데이터를 획득하기 위한 RFID 시스템을 설계하고 고객의 이동 패턴을 파악하기 위한 방법을 제안하고 결과를 분석하여 시사점을 찾는 것을 목적으로 한다.

2. 관련연구

그동안 고객 행동 정보를 얻기 위한 연구는 관찰이나 설문조사 방식을 통하여 상품 배치에 따른 매출액이나 브랜드 전환의 상관관계 분석, 비계획적인 구매에 대한 요인의 탐색, 구매 빈도와 고객의 인구 통계학적 특성의 분석, 매장 내 정보 프로세싱과 의사결정과정, 쇼핑 고객을 추적하여 구매 행동의 다양성에 대한 인류학적(anthropological) 연구(Underhill[13]) 등이 있다. 이러한 연구들은 고객의 집단적인 구매 행동에 대한 행동과학적 또는 인류학적(anthropological)연구들로 고객 집단의 구매 행동에 미치는 변수간의 관계규명에 초점을 맞추고 있다. 또한 컴퓨터와 경영 과학적 분석기술이 발전함에 따라 Farley and Ring[4]는 확률 모형(stochastic model)을 이용하여 고객이 한 매대에서 다른 매대로 이동하는 확률을 추정함으로써 고객동선의 패턴을 파악하고자 하는 연구를 수행하였다. 이는 고객동선분석에 대한 최초의 계량적인 접근방식을 제안했다는 점에서 의의가 있으나 오늘날과 같이 많은 고객이 내왕하며 많은 매대가 있는 대형 유통매장에 실제로 적용하는 데에는 한계가 있다.

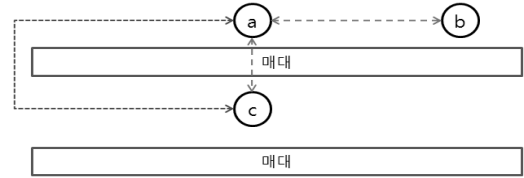
근래에는 RFID와 영상 기술의 발전으로 고객 개인의 매장내 움직임을 파악 할 수 있게 되었다. Larson et al.[6]은 쇼핑카트에 RFID 태그(이하 태

그)를 부착하여 슈퍼마켓 매장에서 얻은 RFID 데이터를 이용하여 고객 쇼핑시간의 장단에 따라 3개의 고객 그룹으로 나누어 각 그룹의 특징적인 동선을 파악하고자 하였다. 이는 RFID 기술의 적용 가능성을 보인 최초의 탐색적인 연구로써 의미가 크다. Liao and Lin[7]도 RFID 기술을 이용하여 고객의 동선을 파악하고자 하였다. 카트에 RFID를 부착, 매장을 그리드(Grid)화 하여 고객의 동선을 추적하고 연관규칙 기반의 Transaction Mining 방법론을 적용하여 분석하였다. 이와 비슷한 연구로 Gil et al.[5]은 480명의 슈퍼마켓 고객의 동의 하에 CCTV 영상을 통한 고객 움직임을 분석하여 고객을 5종류의 고객 군으로 분류하고 매장 배치가 구매 행동에 미치는 영향, 특징적인 매장 이동 또는 구매 행동의 패턴 등을 분석하고자 하였다. 또한 서성보 외[1]은 RFID를 이용하여 매장의 고객들을 추적하는 시스템을 가정하여 얻은 동선의 빈발 발생 패턴을 발견하는 순차패턴 알고리즘을 적용, 의미 있는 패턴을 발견하고자 하였다.

이들 연구에서와 같이 데이터로부터 특징 패턴을 찾아내기 위해서는 일반적으로 군집화 기법을 이용하는데, Larson et al.[6]과 Gil et al.[5] 연구에서도 군집화를 위해 유클리드 거리기반의 통계적 군집화 기법을 이용하였다. 그러나 매장에서의 고객 위치 데이터와 같이 공간적 특성을 갖는 데이터의 경우 유클리드 거리기반의 군집화 과정 중 유사도 측정에 문제점을 내포하고 있다. 예를 들어, [그림 1]에서와 같이 a-c의 유클리드 거리는 a-b의 유클리드 거리보다 훨씬 가까우나 실제 매장에서 고객 이동 동선 관점에서 볼 때 a-c는 a-b보다 더 많은 이동 거리를 갖는다. 이와 같이 유클리드 거리를 이용할 경우 고객들 간에는 동선 유사도가 달라지는 문제점이 있다.

또한 유클리드 거리 기반 군집화 기법을 이용할 경우 입력 변수의 개수가 동일해야 하는데 고객마다 동선길이가 다르기 때문에 유클리드 거리기반 군집화 기법을 사용할 수 없거나 동일한 길이의 RFID 데이터를 획득하도록 하는 RFID 시스템 설

계상의 문제점이 있다. 본 연구는 이러한 문제점을 해결하기 위한 제안이다.



[그림 1] 매장 환경에서의 유클리드 거리 사용 문제점

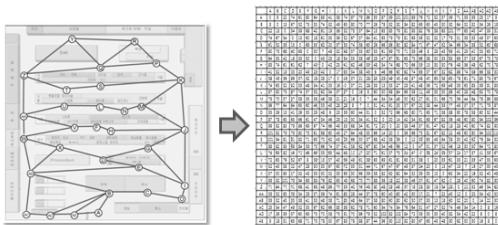
3. 고객 쇼핑 동선 패턴 분석

3.1 공간적 제약을 고려한 고객동선 패턴 군집화

제 2장 관련 연구에서 언급된 바와 같이 대부분의 연구에서는 유클리드 거리에 기반을 두어 유사도를 계산하여 군집화를 하는데 유클리드 거리는 매장내의 매대나 가판대와 같은 장애물 제약 여부에 따라 고객 동선들 간의 유사도가 달라진다는 문제점이 있다. Tung et al.[12]은 이와 같은 문제를 풀기위하여 COD(clustering with obstructed distance)-CLARANS 군집화 알고리즘을 제안하였다. COD-CLARANS는 군집화 알고리즘인 CLARANS를 기반으로 하여 두 객체 또는 객체와 군집의 중심 사이에 장애물이 존재할 경우 장애물을 우회하는 최단 경로를 거리로 사용하는 새로운 거리 함수를 적용함으로써 장애물로 인한 거리 문제를 처리한다. 데이터 전처리 단계에서 가시성 그래프(Visibility Graph)를 생성하여 장애물이 있는 두 점사이의 최단 경로를 거리 계산하였다. 이를 위해 BSP(Binary-Space-Partition) 트리를 사용하여 두 점사이의 가시성 여부를 결정하는데 두 점이 일직선으로 서로 연결이 가능하면 서로 가시성이 있다고 판단한다. 가시성을 기준으로 모든 데이터와 장애물의 꼭짓점을 그래프 노드로 표현을 하여 그래프로 만들어 각 노드의 거리를 계산하였다. Lim et al.의 연구[8]에서는 장애물 문제를 해결하기 위하

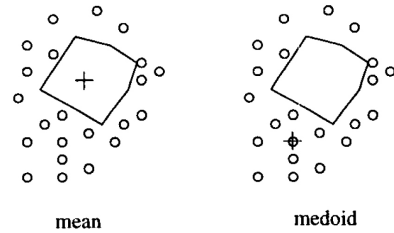
여 밀도 기반의 군집화 기법인 DBSCAN-O를 제안하였다. DBSCAN-O는 COD-CLARANS의 가시 그래프(Visibility Graph)를 통하여 인덱스를 구성하고, 알고리즘 상에서 이를 참조하는 방식을 사용한다.

장애물 제약 조건을 고려하는 군집화 방법에서는 장애물을 고려하지 않는 경우보다 알고리즘의 계산량이 크게 증가하게 된다. 반복적인 유사도 계산이 수행되는 군집화 방법에서, 이러한 현상은 분석 방법의 실제 적용에 문제를 일으킬 수 있다. 따라서 본 연구에서는 고객동선간의 유사도를 계산하기 위하여 거리-인덱스행렬(distance-index matrix)을 제안한다. 매장안의 각 RFID 리피터(이하 리피터)는 하나의 노드(node)로 표현되며, 매대에 걸리지 않아 고객이 이동에 지장이 없는 노드끼리는 직접 연결이 되며, 매대에 걸려 이동할 수 없는 노드끼리는 중간에 다른 노드를 거치는 최단 경로로 연결된다. 최단 경로는 다익스트라(dijkstra) 알고리즘으로 구하였다. [그림 2]는 매장 안의 모든 리피터에 대하여 이러한 연결 관계와 리피터가 31개일 경우, 31×31 행렬의 형태를 갖는 거리-인덱스 행렬을 보이고 있다.



[그림 2] 거리 인덱스 행렬

본 연구에서는 군집화를 위해 k-medoids 알고리즘을 사용하였다. 그러나 전통적인 k-means 군집화의 경우 [그림 3]과 같이 매장 내에서 매대나 가판대에 가로막혀 고객이 이동하기 어려운 지역에 군집중심이 발생할 수 있기 때문에 제안한 거리-인덱스 행렬을 이용한 클러스터링 알고리즘을 개발 하였다.



[그림 3] Mean vs Medoid[12]

이때 사용된 알고리즘은 대표적인 k-medoids 알고리즘의 하나인 PAM(Partitioning Around Medoids) 알고리즘이다. 생성된 거리-인덱스 행렬은 [그림 4]의 알고리즘 절차에서 각 객체간 거리를 구하는 계산 과정에 활용된다.

입력 : k : 군집화 수
D : 동선 데이터

출력 : k-군집화 결과

반복절차 :

- (1) (초기 값 객체 선정) 초기 값으로 k개의 동선을 데이터 D에서 임의로 선정
- (2) (반복) 최종 군집의 변화가 없을 때까지
- (3) (객체 군집 배정) 초기 값을 중심으로 각 객체에 대하여 k개의 군집 중심좌표와의 거리를 산출 후 가장 가까운 군집으로 배정
- (4) (새로운 중심 좌표 선정) 새로운 군집의 중심좌표 재계산
- (5) (수렴조건 점검) 새로 산출된 중심좌표의 값과 이전 좌표 값을 조건 비교하여 변경 여부 결정

[그림 4] PAM 알고리즘

본 연구에서 고객의 동선 추적을 위하여 사용하는 RFID 시스템에서는, 군집화 과정 중 각 객체의 위치가 리피터의 위치로 표현 된다. 즉, 고객의 매장 내 이동 경로는 리피터 좌표의 조합으로 이루어지며, 따라서 미리 구축된 거리-인덱스 행렬을 이용하는 방식은 기존의 방식과 비교하여 알고리즘의 수행 상 계산량을 크게 낮출 수 있는 장점을 가진다.

3.2 고객 동선 정규화

고객의 구매 행동에 따라 매장에서 여러 지역을 이동하는 고객이 있는 반면, 상대적으로 적은 지역을 움직이는 고객이 있으므로 고객 이동 동선의 형태와 길이는 고객마다 다르다. 그러나 특징적 동선의 패턴을 찾기 위한 동선 군집화를 위해서는 동선의 길이가 같아야 한다. 본 연구에서는 이러한 문제를 해결하기 위하여, 시간적 정규화(temporal normalization) 기법을 이용하여 각 고객의 전체 쇼핑 시간을 n 등분 하여 기준 시간 축을 만들고, 시간 축에 매칭 되는 동선 정보를 추출하는 방식을 사용한다. <표 1>은 이러한 과정을 통해 전 처리된 동선 데이터를 보여준다. <표 1>은 약 964초의 동선을 가지는 고객에 대하여 $n = 400$ 일 경우의 예시인데, “duration”은 총 쇼핑 시간을 $n = 400$ 으로 나누었을 때의 값, 즉 시간 간격이며, “scan_time”은 이동 중인 카드에 부착된 태그로부터의 신호가 리피터에 검출된 시각을 나타내고 있다. “calc_time”은 정규화 과정을 통해 새롭게 계산된 검출 시간을

보인다. repeater_id는 매장에 설치된 리피터 ID를 나타낸다. 매대에 여러 개의 리피터가 설치되어 있어 고객들이 카트를 끌고 지나가게 되면 리피터 ID를 통해 이동 위치를 순차적으로 파악할 수 있다. 각기 다른 길이를 가진 고객의 동선은 설정한 n 으로 표준화 된 동일한 길이의 동선 형태를 가지게 하기 위하여, 예를 들어 전체 쇼핑 시간을 $n = 400$ 으로 나누어 duration은 2.41이 되고 이 시간을 단위 시간으로 나누어 [그림 5]과 같이 쇼핑 동선을 정규화 한다. 이러한 과정을 통해서 최초의 scan_time의 데이터에서 정규화 과정을 통해 calc_time으로 새롭게 이동 동선이 탐지된 것으로 가정을 하는 것이다.

4. 분석프로세스와 실험

4.1 분석 프로세스 설계

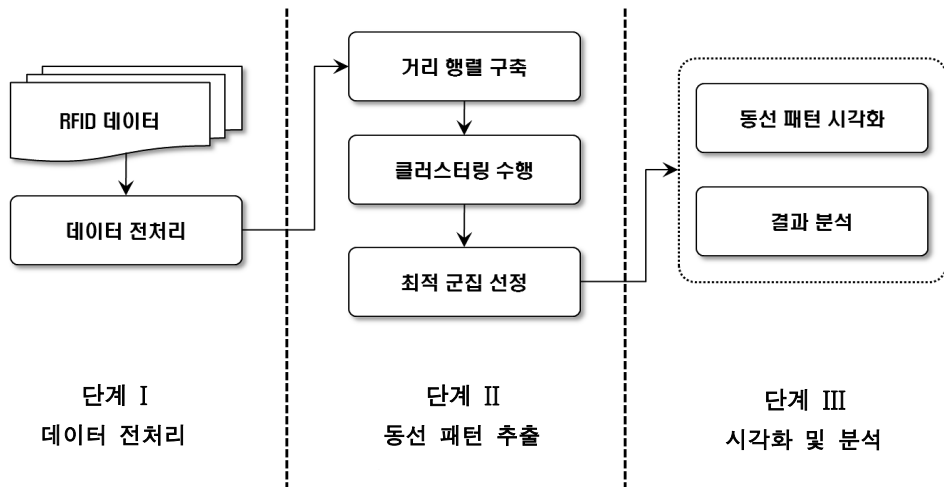
[그림 6]은 분석 프로세스를 보인다. 단계 I에서는 RFID 시스템을 통해 얻어진 동선 데이터를 전

<표 1> 전처리 데이터의 예

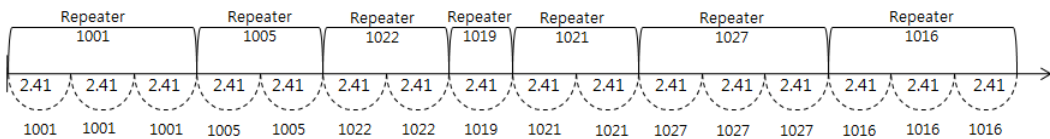
sequence	calc_time	duration	scan_time	repeater_id
1	09:54:38	7.23	09:54:38	1001
2	09:54:45	4.82	09:54:48	1005
3	09:54:50	4.82	09:54:51	1022
4	09:54:55	2.41	09:54:56	1019
5	09:54:57	4.82	09:54:58	1021
6	09:55:02	7.23	09:55:03	1027
7	09:55:09	7.23	09:55:11	1016

전처리
====>

sequence	calc_time	duration	scan_time	repeater_id
1	09:54:38	2.41	09:54:38	1001
2	09:54:40	2.41	09:54:38	1001
3	09:54:43	2.41	09:54:38	1001
4	09:54:45	2.41	09:54:48	1005
5	09:54:48	2.41	09:54:48	1005
6	09:54:50	2.41	09:54:51	1022
7	09:54:52	2.41	09:54:51	1022
8	09:54:55	2.41	09:54:56	1019
9	09:54:57	2.41	09:54:58	1021
10	09:55:00	2.41	09:54:58	1021
11	09:55:02	2.41	09:55:03	1027
12	09:55:04	2.41	09:55:03	1027
13	09:55:07	2.41	09:55:03	1027
14	09:55:09	2.41	09:55:11	1016
15	09:55:12	2.41	09:55:11	1016
16	09:55:14	2.41	09:55:11	1016



[그림 5] 분석 프로세스



[그림 6] 동선 전처리 과정

처리한다. 고객이 태그를 부착한 카트나 장바구니를 이용하여 매장을 이동하게 되면, 매장의 매대에 부착된 리피터를 통하여 이동 경로가 추적된다. 예를 들어 12:30:10에 A고객의 카트가 3번 리피터에서 인식되고, 12:30:15에 3번 리피터에서, 12:30:20에 5번 리피터에서, 12:30:25에 10번 리피터에서 인식된다면 A고객의 15초간 이동 경로는 3 → 5 → 10으로 표현된다. 단계 I에서는 이렇게 얻어진 데이터를 분석에 적합한 형태로 변환하게 되며 이때 시간적 정규화 과정을 거쳐 고객들의 동선은 동일한 길이로 변환된다.

단계 II는 군집화 단계로, 단계 I를 통하여 전처리된 데이터에 대한 동선 패턴 추출을 실시하는 단계이다. 동선간의 유사도를 측정하기 위하여 거리 인덱스 행렬을 만들고 이를 이용하여 PAM 알고리즘으로 동선군집화를 한다. 최적의 군집 수 k 를 결정하기 위하여 일반적으로 많이 쓰이는 SSE(Sum of Square Error)를 이용하였다(Tan et al.[11],

Sugar[10], Larson[6]).

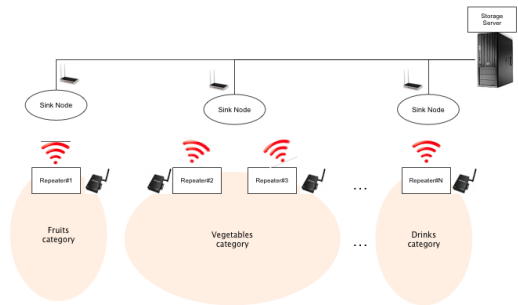
단계 III은 동선 패턴 군집화 결과를 시각화하고 결과를 분석하는 과정으로 매장 운영 의사결정자에게 매대의 재배치, 운영 인력 관리, 상품 카테고리 관리 등에 대한 통찰력을 제공한다.

4.2 고객 위치 데이터의 획득을 위한 RFID 시스템

고객의 구매 행동분석을 위해서는 고객의 이동 동선 정보와 구매이력 정보가 필요하기 때문에, 구매 현황 및 물품에 대한 정보는 기존의 POS 시스템과 주문 DB를 통해서 수집하고, 고객 개인마다 매장에서 이루어지는 이동 동선 정보를 수집하기 위해서 [그림 7]과 같이 매장내의 매대에 리피터와 매장 천정에 리더기 역할을 하는 싱크 노드(sink node)를 설치하고 고객의 쇼핑 카트에 태그를 설치하였다.



[그림 7] 매장과 쇼핑가트에 설치된 RFID 장비



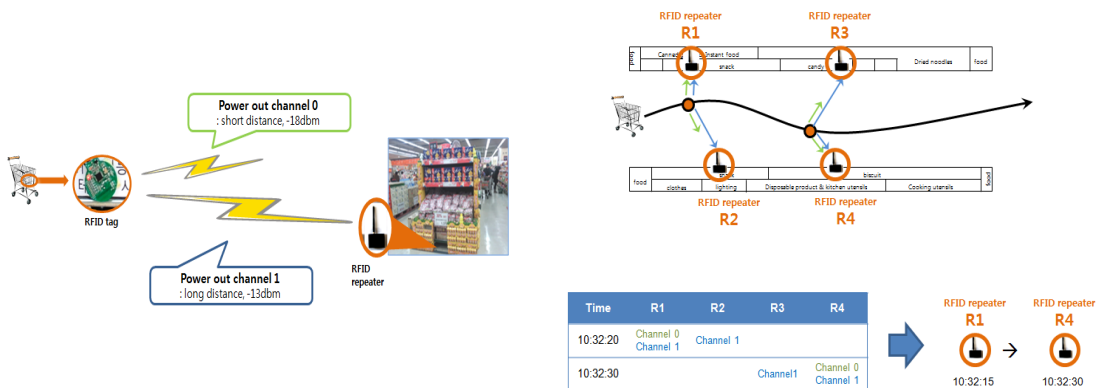
[그림 8] 리피터와 리더기

동선 정보 획득을 위해 2.4GHz 주파수 대역을 사용하는 Ultra-low Power Wireless System One-Chip을 능동형 태그 형태로 개발 하였다. 능동형 태그는 쇼핑 카트에 부착하여 일정한 주기로 데이터를 송신하며, 매장 내부의 주요 위치마다 리피터를 설치하여 쇼핑 카트가 지나갈 때 태그로부터 데이터를 수신하여 싱크 노드로 전송을 하고 다시 정보를 동선 저장 서버에서 처리할 수 있도록 데이터를 가공하여 서버로 전송하도록 RFID 시스템 설계를 하였다([그림 8] 참조).

리피터 가격이 리더기 가격보다 훨씬 저렴하기 때문에 10여 개의 리피터로부터 수신된 데이터를 1개의 RFID 리더기가 수신하도록 시스템을 구성함으로써 시스템 설치비용을 절감 하였다. 리피터의 설치 위치는 진열 상품의 분류기준(대분류, 중분류, 소분류)을 근거로 하여 적절히 설치 지점을 정

하였고, 이때 RFID의 전파는 설치 지역의 위치와 설치 구조물 모양, 그리고 구조물의 종류에 따라 수신율이 달라지기 때문에, 위치별로 리피터의 수신 감도 조절을 하였으며, 능동형 태그는 쇼핑카트의 전면부 하단에 부착하였다([그림 7] 참조). 매장통로에 있는 쇼핑카트의 위치를 정확히 탐지하기 위하여 태그를 2가지 채널로 설계하였다. 채널 1은 짧은 유효 거리로 신호를 보내며 채널 2는 긴 거리까지 신호를 보냄으로써 쇼핑카트가 어느 매대에 가까이 있는지를 파악 할 수 있다. [그림 9]는 2가지 채널 신호를 이용하여 정확한 이동 위치를 파악하는 과정을 보여주고 있다.

고객의 쇼핑 카트에 부착된 고유한 ID를 갖는 능동형 태그를 통해 고객의 이동 위치 및 시간 데이터가 수집되며 고객이 쇼핑을 끝내고 계산대에서 계산을 마치면, 고객 정보(예를 들어, Family



[그림 9] 태그 2채널 신호를 이용한 쇼핑 카트위치 파악과정

Card ID)와 태그 ID를 매핑 하여 개별 고객의 이동정보와 시간 및 구매 상품 정보 등을 얻을 수 있다.

4.3 RFID 동선 데이터의 획득과 전처리

국내의 S사의 슈퍼마켓에 설치한 RFID 시스템을 가동하여 고객동선 데이터를 획득하였다. 데이터는 201*년 1월 22일부터 2월 20일 사이에 수집되었으며, 리피터에서 인식된 태그의 인식 건수는 8,428,542건이었다. 하루 약 500여명, 총 16,640명의 고객 데이터가 획득되었으며 고객 당 평균 쇼핑 시간은 약 31분, 평균 구매 금액은 59,842원이었다. 매장의 면적은 총 1,300㎡이며, 168개의 리피터가 설치되었으며, 300대의 쇼핑 카트에 태그가 부착되었다. [그림 10]은 매대의 배치와 설치된 RFID를 보이고 있다.



[그림 10] 매대 배치와 설치된 리피터

고객동선의 길이가 고객마다 다르기 때문에 동선 길이 정규화를 실시하였다. 이때 적절한 n 의 값을 선택해야 하는데 n 값을 너무 적게 설정할 경우 동선 데이터의 손실이 있을 수 있기 때문이다. 전체 쇼핑 시간을 n 으로 나눈 “duration” 값이 실제 RFID 검출 간격보다 크면 기준 시간에 매칭되는 데이터의 간격이 넓어지기 때문이다. 본 연구에서

n 을 500으로 설정하였는데, 이는 RFID 검출 횟수를 동선의 길이로 보았을 때, 하위 약 90%를 기준으로 설정한 값이다. 가장 길이가 긴 동선을 기준으로 정규화를 진행한다면, 데이터의 손실은 방지할 수 있으나 군집화 과정에서 계산량이 급격히 늘어나는 단점을 가지게 된다.

정규화 과정에 따른 데이터 손실을 관측하기 위하여 인식된 리피터 수에서 결측치를 나눈 값으로 손실률을 정의하였다. 여기서 인식된 리피터의 수는, 한 고객의 동선에서 연속으로 중복된 값을 제외한 리피터 검출 개수를 의미한다. 즉, 어떠한 고객의 동선이 $A \rightarrow B \rightarrow B \rightarrow C$ 로 표현될 때, 연속된 중복 값인 B 를 제외한 A, B, C 세 개가 인식된 리피터 수가 된다. 결측치는 정규화 과정을 거친 후, 원래 데이터와의 인식된 리피터 개수의 차이를 뜻한다. 즉, 정규화 후에 위 고객의 동선이 $A \rightarrow B \rightarrow B \rightarrow B$ 로 표현된다면, 중복을 제외한 A, B 두 개가 인식된 리피터의 수가 되며, 이 경우 결측치는 3에서 2를 뺀 값인 1이 된다. 이 경우, 손실률은 $1/3$ 으로 약 33%가 된다. <표 2>는 본 연구에서 n 을 500으로 정규화를 진행하였을 때의 손실률을 보여준다. 정규화된 데이터가 원 데이터와 약 2% 미만의 차이를 보이고 있음을 알 수 있다.

<표 2> 정규화 에 따른 데이터 손실률
($n = 500$ 의 경우)

인식된 리피터 수	결측치	퍼센트(%)
555,789	11,212	1.9774

4.4 PAM 알고리즘을 이용한 동선군집화

군집화를 위해서 거리 인덱스 행렬을 구하였다. 이를 위해 본 연구가 수행된 S사의 슈퍼마켓의 매대를 기준으로 724×630 픽셀의 평면에 리피터의 위치를 좌표로 설정하였다. 이를 바탕으로 고객의 매대 사이의 이동 거리를 구하였는데, 중간에 매대에 걸리지 않아 고객이 이동에 지장이 없는 매대와 매대 사이의 거리는 두 매대 사이의 직선거리로 두

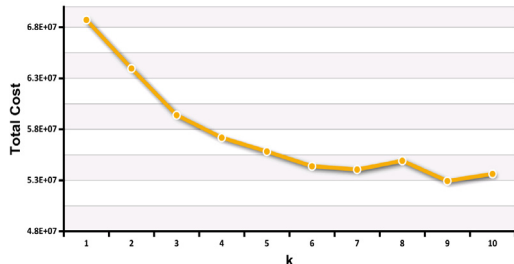
있고, 매대에 걸려 고객이 바로 이동할 수 없는 매대와 매대 사이의 거리는 중간에 다른 노드를 거치는 다익스트라(dijkstra) 최단경로로 하였다.

[그림 11]은 이러한 과정을 거쳐 구축된 거리 인덱스 행렬의 일부를 보이고 있다. 거리 인덱스 행렬의 행과 열은 각각의 리피터 ID를 나타낸다. 본 연구에서는 총 168개의 리피터에서 얻어진 고객 쇼핑 동선 데이터를 이용하여 매대와 매대 사이의 최단거리를 나타내는 168×168의 거리 인덱스 행렬을 구하였다.

1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1	0	25	111	125	121	0	160	158	165	171	171	164	255	500	133	352	406	223	227	294	189	440	571	489	423	282	442	373	
2	25	0	136	100	132	58	165	165	200	322	296	402	374	280	500	358	373	231	192	312	275	153	524	482	474	437	351	417	303
3	111	136	0	134	102	178	143	15	144	165	160	286	230	144	500	222	181	96	130	170	211	101	388	416	412	375	235	305	289
4	125	100	134	0	32	42	179	149	180	222	196	362	374	180	500	250	227	133	86	213	170	41	428	392	374	297	251	317	250
5	121	132	102	32	0	78	147	117	148	193	194	270	262	148	200	298	192	96	4	129	146	15	386	346	292	239	293	293	213
6	160	165	178	42	74	0	201	146	222	264	238	344	316	222	500	300	280	173	137	254	215	108	468	434	416	379	236	304	292
7	165	165	143	179	147	221	0	128	95	137	111	217	389	95	500	172	206	146	170	221	256	146	339	360	267	420	260	230	310
8	158	161	175	149	117	149	128	0	129	171	146	251	223	129	500	227	206	112	146	199	226	133	373	404	404	360	284	363	303
9	295	280	144	190	148	222	95	129	0	42	50	122	164	96	500	144	173	141	148	192	221	117	335	340	264	261	231	331	324
10	207	302	196	222	199	204	137	171	42	0	66	80	112	50	500	129	161	159	191	170	211	159	284	324	342	375	235	295	288
11	271	296	190	196	146	230	111	145	50	66	0	106	168	16	500	94	123	121	153	142	177	189	200	230	300	340	181	251	254
12	277	307	206	231	144	237	251	122	80	104	0	112	122	500	129	164	173	227	170	213	230	234	244	265	296	183	229	243	
13	240	274	228	274	242	316	189	225	124	112	106	112	0	94	500	16	143	143	170	143	181	211	182	232	236	303	172	176	
14	255	286	144	136	146	222	95	129	86	50	16	122	94	0	500	78	111	102	141	126	161	147	244	234	232	325	189	238	238
15	500	500	500	500	500	500	500	500	500	500	500	500	500	500	0	500	500	500	500	500	500	500	500	500	500	500	500	500	500
16	203	292	229	229	229	229	173	207	144	139	94	159	16	78	500	0	133	127	160	129	165	186	146	144	247	81	81	80	
17	302	327	191	227	192	206	206	206	177	161	127	159	146	111	500	153	0	96	132	15	132	164	197	227	286	214	24	188	127
18	208	231	95	131	98	175	140	130	141	155	131	171	143	195	500	127	96	0	96	81	116	86	290	323	317	260	120	260	169
19	223	268	173	98	81	137	178	148	148	159	157	207	176	141	500	163	132	96	0	117	78	32	202	229	226	260	222	163	
20	227	212	192	210	254	221	169	182	176	142	174	142	128	500	128	15	0	117	0	117	149	212	242	236	199	39	179	112	
21	298	373	211	173	144	215	256	227	211	177	213	181	361	500	163	118	78	117	0	119	251	219	211	184	78	144	77	22	
22	188	193	301	83	31	195	146	116	117	159	163	229	211	147	500	166	164	96	32	140	110	0	360	320	311	274	180	254	107
23	489	524	708	414	382	448	339	371	333	284	260	214	182	244	500	166	129	320	221	252	340	0	182	144	187	173	107	114	
24	417	482	418	382	360	414	369	403	440	324	290	244	212	274	500	196	227	323	297	242	218	329	182	0	11	185	203	25	142
25	469	474	412	374	342	416	387	425	358	342	309	282	230	282	500	224	246	317	279	239	201	311	164	19	0	147	197	27	114
26	452	413	425	353	358	378	420	380	381	375	341	256	325	500	247	274	280	242	190	146	270	197	185	147	0	180	169	81	11
27	528	511	413	391	374	380	380	380	381	415	411	405	403	405	500	471	54	120	196	261	276	189	173	233	197	160	140	163	71
28	442	417	395	317	295	359	331	344	331	285	251	205	173	235	500	157	169	200	221	176	144	254	107	75	57	80	140	0	47
29	375	393	398	259	239	333	330	384	284	254	209	176	230	500	160	127	193	193	112	77	107	174	142	124	27	15	47	0	1

[그림 11] 구축된 거리 인덱스 행렬

다음으로 군집수에 따라 SSE(Sum of Square Error)를 구하였는데 [그림 12] 에서와 같이 군집 수는 11까지 측정하였다.



[그림 12] 군집 수에 따른 SSE 변화

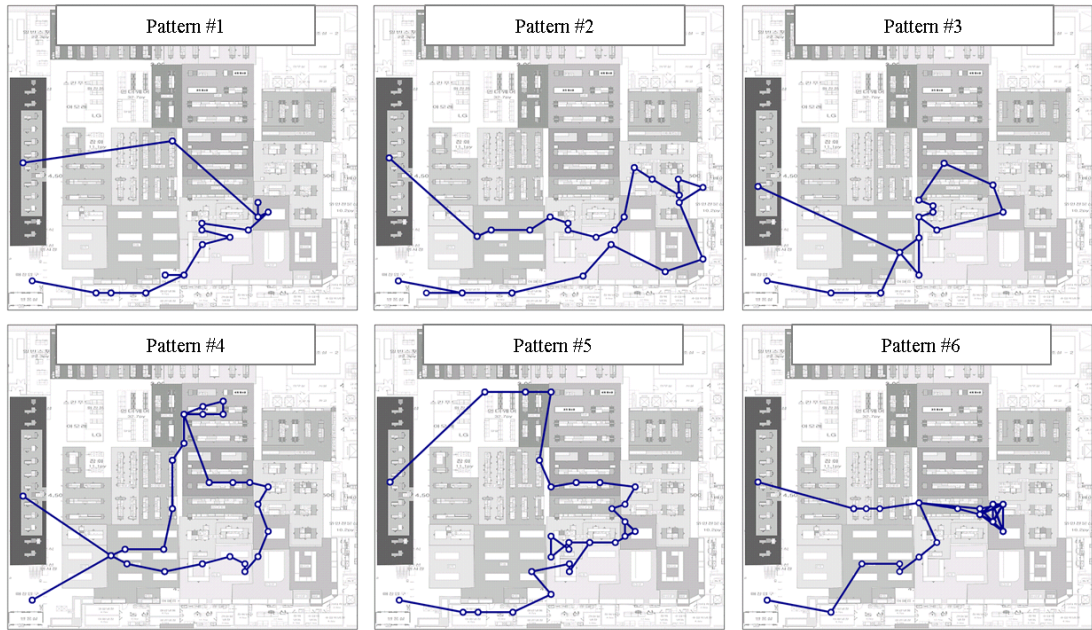
4.4 결과 시각화 및 분석

RFID 시스템의 시각화모듈을 통하여 [그림 13]

과 같이 6개의 동선 패턴을 얻었다. 패턴 #1은 2718명, 패턴 #2는 3481명, 패턴 #3은 2262명, 패턴 #4는 2426, 패턴 #5는 2926명, 패턴 #6은 2827명의 고객이 속한다.

<표 3>은 6개 그룹의 고객의 특징을 보이고 있다. 각 패턴별 평균 구매 금액과 평균 쇼핑시간을 나타내었다.

패턴별 고객 이동 동선의 특징을 살펴보면 <표 4>와 같은 특징을 보이고 있다. 패턴 #1의 경우에는 주로 채소 및 닭고기 코너를 지나 생선 코너와 견해산물 코너를 둘러보고 냉장냉동, 조미식품, 주방용품 순으로 쇼핑 동선이 이루어지는 것으로 나타났다. 전반적인 모습은 매장을 반시계 방향으로 순회하는 형태를 나타내고 있다. 패턴 #2의 경우에는 채소와 과일 코너를 지나 닭고기, 생선, 소고기 코너를 통과하여 조리식품, 냉장냉동, 데일리, 조미식품 코너에서 쇼핑을 한 후 다시 채소 코너를 지나 매장을 빠져 나가는 쇼핑 동선을 보여주었다. 전반적으로 매장 하단을 반시계 방향으로 이동하는 모습을 보여주고 있다. 패턴 #3의 경우에는 채소와 닭고기 코너를 지나 견해산물, 조미식품, 냉장냉동, 인스턴트 코너를 둘러보고 견해산물, 생선, 채소 코너를 지나가는 쇼핑 동선을 보여주었다. 다른 패턴들과 비교하여 이동 영역이 좁은 특징을 가지고 있다. 패턴 #4는 채소와 주방용품, 건강·차 코너를 지나 과자 코너에서 쇼핑을 한 후, 인스턴트, 데일리, 냉장냉동과 생선 코너를 지나 견해산물과 채소 코너를 지나가는 동선을 보여주었다. 다른 패턴들과 비교하여 매장의 상하 이동이 긴 형태를 보여주었으며, 반시계 방향이 아닌 시계 방향으로 쇼핑 동선이 형성되었다. 패턴 #5는 채소와 닭고기 코너를 지나 생선, 견해산물, 조미식품, 냉장냉동 코너를 통과하여 인스턴트와 데일리, 주방용품, 세제 코너에서 쇼핑을 한 후, 수예, 애완원예용품 코너를 지나가는 동선을 보여주었다. 매장을 크게 반시계 방향으로 순회하는 모습을 보여주었다. 패턴 #6은 채소와 인스턴트 코너를 지나 냉장냉동, 데일리 코너에서 쇼핑을 한 후, 주방용품, 청소욕



[그림 13] 쇼핑 동선 패턴 시각화 결과

<표 3> 패턴별 특징

	패턴 1	패턴 2	패턴 3	패턴 4	패턴 5	패턴 6
고객수(명)	2,718	3,481	2,262	2,426	2,926	2,827
평균 구매금액(원)	60,578	61,173	61,670	60,778	56,728	58,345
평균 쇼핑시간(초)	1,940	1,524	2,190	1,975	1,724	2,081

<표 4> 패턴별 동선의 특징

패턴	이동 특징	이동 방향 및 특징
패턴 1	채소 → 닭고기 → 생선 → 견해산물 → 냉장냉동 → 조미식품 → 주방용품	반시계 방향
패턴 2	채소 → 과일 → 닭고기 → 생선 → 소고기 → 조리식품 → 냉장냉동 → 유가공품 → 조미식품 → 채소	반시계 방향
패턴 3	채소 → 닭고기 → 견해산물 → 조미식품 → 냉장냉동 → 인스턴트 → 견해산물 → 생선 → 채소	반시계 방향
패턴 4	채소 → 주방용품 → 건강 · 차 → 과자 → 인스턴트 → 유가공품 → 냉장냉동 → 생선 → 견해산물 → 채소	시계 방향
패턴 5	채소 → 닭고기 → 생선 → 견해산물 → 조미식품 → 냉장냉동 → 인스턴트 → 유가공품 → 주방용품 → 세제 → 수예 → 애완원예용품	반시계 방향
패턴 6	채소 → 견해산물 → 인스턴트 → 냉장냉동 → 조미식품 → 유가공품 → 주방용품 → 청소욕실	반시계 방향

실 코너를 통해 매장을 빠져나가는 쇼핑 동선을 보였는데, 다른 쇼핑 동선 패턴들과 비교하여 좁은 지역에서 동선이 형성되는 특징을 보여주었다.

<표 5>는 고객들의 각각의 고객 쇼핑 동선에 따른 상품 분류별 구매 건 수를 나타낸 것이다. 매우 적은 구매 건 수를 보여준 수예, 완구, 주류 상

〈표 5〉 고객 쇼핑 동선에 따른 상품 별 판매 건 수

	Pattern #1	Pattern #2	Pattern #3	Pattern #4	Pattern #5	Pattern #6
건강, 차	72	104	74	201	173	110
건해산물	313	349	261	215	208	391
과일	1050	1243	815	623	726	650
과자	679	812	584	1548	1544	1269
냉장냉동	601	914	521	459	419	868
닭고기	564	720	440	294	444	419
데일리	973	1464	802	884	1053	1473
돼지고기	190	278	139	101	101	205
생선	827	929	631	392	500	571
세계	128	135	121	324	384	168
소고기	420	579	301	192	232	300
애완원예용품	12	23	16	52	107	18
위생용품	95	92	61	220	237	147
음료	184	242	169	294	277	373
인스턴트	426	569	367	483	402	747
인테리어	36	45	36	120	151	47
자동차	15	14	10	37	64	12
조리식품	542	996	450	400	467	768
조미식품	565	666	533	483	439	767
주방용품	339	365	326	475	441	430
채소	2869	3496	2254	1726	2255	2110
청소욕실	171	211	191	260	215	238

품을 제외한 나머지 상품들에 대한 각각의 구매 건 수를 나타내었다. <표 5>를 살펴보면, 상품 구매 성향을 알 수 있다. 이러한 결과는 매장 내 다양한 마케팅 전략 수립에 활용할 수 있을 것으로 기대된다. 예를 들어, 패턴 #4를 살펴보면, 전체 그룹과 비교하여 채소 품목의 구매가 상당히 저조한 것을 알 수 있다. 반면에, 과자의 경우는 전체 그룹이나 다른 패턴들과 비교하여 상당히 높은 구매 건수를 보여준다. 따라서 패턴 #4에 해당하는 고객들에게는 소고기 상품에 대한 할인 쿠폰을 증정하는 등의 프로모션을 통하여 상품군간 판매량의 조절이 가능할 것이다. 또한 [그림 12]의 동선 시각화 결과와 비교하여 동선과는 가까우나, 구매가 많이 일어나지 않는 상품에 대하여 집중적인 프로

모션을 진행한다면, 동선을 고려하지 않았을 경우보다 상품에 노출될 확률이 높기 때문에, 프로모션에 대한 높은 반응률을 기대할 수 있을 것이다.

<표 6>은 패턴 #4를 분석한 예이다. 동선은 패턴 #4의 동선을 상품 군에 매칭한 결과이고 구매 상품은 패턴 #4에 속하는 고객들이 구매상 상위 15개의 상품을 나타낸 것이다. 표를 살펴보면, 패턴 #4의 경우 고객들이 건강, 차 매장과 건해산물 매장을 지나가지만 실제 구매는 많이 이루어지지 않은 것을 살펴볼 수 있다. 이러한 경우, 건강, 차 상품과 건해산물 상품에 대한 프로모션 활동이 필요함을 알 수 있다.

또한 행사상품 선정 이외에도 쿠폰 발행과 전단지 작성에도 이러한 결과들을 활용할 수 있다. 기

존의 고객 구매 분석은 주로 장바구니 분석과 같이 고객의 구매 이력을 분석하여 이루어지고 있는데, 이러한 분석 방법은 현재 판매 기회를 잃고 있는 상품에 대한 분석이 어려운 한계점을 가진다.

〈표 6〉 패턴 #4의 동선과 구매 상품

동선	구매 상품 및 구매 건수	
출입구	채소	1726
채소	과자	1548
주방용품	데일리	884
건강, 차	과일	623
과자	인스턴트	483
건강, 차	조미식품	483
인스턴트	주방용품	475
데일리	냉장냉동	459
인스턴트	조리식품	400
냉장냉동	생선	392
생선	세제	324
냉장냉동	닭고기	294
건해산물	음료	294
채소	청소욕실	260
POS	위생용품	220

본 연구에서 제안하는 분석 시스템을 통해, 매장 내 고객들의 주요 쇼핑 동선을 찾아낼 수 있다면, 고객들의 동선을 고려하여 쿠폰 발행이나 전단지 작성이 가능하기 때문에, 기존에 간과하고 있었던 판매가 저조한 상품에 대한 프로모션 방안을 마련할 수 있다. 전단지 작성 역시, 동선을 고려하여 고객의 쇼핑 동선 패턴에 따라 전단지 레이아웃을 작성하는 등의 활용이 가능하다.

매장 내 고객 쇼핑 동선 패턴을 이용하여 어떤 지역이 고객들이 이동하는 주요 경로인지를 파악하는 분석도 가능하다. 이러한 주요 영역은 행사 상품 선정과 연계하여 이벤트 진행을 위한 장소 선정에 활용할 수 있는데, 이러한 방법은 기존에 직관력에 의존하는 장소 선정에 비교하여 정량적인 분석 결과를 토대로 얻어진 결과라는 장점을 가진다. 또한 각 쇼핑 동선 패턴의 구매 데이터와 연

계, 요일별이나 시간대별로 주요 영역을 파악하여 이벤트를 진행하는 고도화된 마케팅 전략 수립이 가능할 것이다.

5. 결 론

본 연구에서는 RFID 동선 추적 시스템을 설계, 구현하고 매장에서의 고객이동 데이터를 획득하고 동선 패턴을 추출하기 위하여 분석 프로세스를 설계하였다. 공간적 제약을 고려한 고객 동선들 간의 유사도를 구하기 위하여 거리 인덱스 행렬과, 길이가 다양한 고객동선을 군집화하여 패턴을 찾아내는 방법을 제안하고 국내 대형마트의 한 슈퍼마켓매장에 적용하여 그 결과를 분석하고 시사점을 제시하였다.

그간 매장관리자의 직관 또는 경험에 의존하던 고객동선 패턴을 과학적이고 정량적인 방법으로 구함으로써 매장내의 매대, 상품배치 등과 같이 매장운영 전반에 대한 보다 정량적인 정보와 시사점을 얻을 수 있음을 보였다. 차후에는 분석 결과의 활용 효과에 대한 검증과 더욱 다양한 활용 방안에 대한 연구가 필요할 것이다. 태그와 리더퍼 부착 등 RFID 시스템 설치에 소요되는 시간과 PAM 알고리즘을 이용한 군집화 수행 및 거리-인덱스행렬 작성에 걸리는 시간을 단축시킬 수 있다면 매장현업에서의 시스템 활용도는 더 커질 것으로 기대된다.

참 고 문 헌

- [1] 서성보, 이용미, 이준욱, 남광우, 류근호, 박진수, "RFID 데이터 스트림에서 이동궤적 패턴의 탐사", 『한국공간정보시스템학회논문지』, 제11권, 제1호(2009), pp.127-136.
- [2] 심재현, 이성호, "대형할인점의 입지선정을 위한 의사결정에 관한 연구", 『대한토목학회논문집』, 제28권, 제5D호(2008), pp.705-712.
- [3] 유 군, "대형 할인점의 점포특성이 고객 구매

- 행동에 미치는 영향에 관한 연구”, 『한양대학교 경영학과 학위논문』, 2008.
- [4] Farley, J. U. and L. A. Ring, “Stochastic Model of Supermarket Traffic Flow”, *Operations Research*, Vol.14, No.4(1966), pp.555-567.
- [5] Gil, J., E. Tobar, M. Lemlij, A. Rose, and A. Penn, “The Differentiating Behavior of Shoppers : Clustering of Individual Movement Trace in a Supermarket”, *Proceeding of the 7th International Space Syntax Symposium*, 2009.
- [6] Larson, J. S., E. T. Bradlow, and P. S. Fader, “An exploratory look at supermarket shopping paths”, *Intern. J. of Research in Marketing*, Vol.22(2005), pp.359-414.
- [7] Liao, I. and W. C. Lin, “Shopping path analysis and transaction mining based on RFID technology”, *RFID Eurasia 1st Annual*, (2007), pp.1-5.
- [8] Lim, H. S., H. S. Kim, H. S. Yong, S. H. Lee, and S. S. Park, “Density Based Spatial Clustering Method Considering Obstruction”, *Korea Multimedia Society*, Vol.6, No.3(2003), pp.375-383.
- [9] Newman, A. J., D. K. C. Yu, and D. P. Oulton, “New insights into retail space and format planning from customer-tracking data”, *Journal of Retailing and Consumer Services*, Vol.9, No.5(2002), pp.253-258.
- [10] Sugar, C., “Techniques for clustering and classification with applications to medical problems”, *Unpublished Phd dissertation, Stanford University*, 1998.
- [11] Tan, P. N., M. Steinbach, and A. K. Jain, *Introduction to Data Mining*, Addison Wesley, Pearson International Edition, 2006.
- [12] Tung, A. K. H., J. Hou, and J. Han, *Spatial Clustering in the Presence of Obstacles*, Data Engineering, Proceedings 17th International Conference, 2001.
- [13] Underhill, P., *Why We Buy*, N. Y., 1999.

◆ 저 자 소 개 ◆



양 승 준 (seven@lgcns.com)

동국대학교에서 산업시스템공학석사와 석사학위를 취득하였다. 현재는 LG CNS에 근무하고 있으며, 주요 관심분야는 데이터마이닝, 고급 분석, Big Data 분석, Data Warehouse, OLAP, 분산 저장 시스템 등이다.



정 인 철 (uhhaha@gmail.com)

동국대학교에서 산업시스템공학석사와 박사학위를 취득 하였고 주요 관심 분야로는 기계학습, 데이터마이닝, Big Data 분석, 지능형 정보시스템, 에이전트 등이다.



권 영 식 (yskwon@dgu.edu)

동국대학교 산업시스템공학과 교수로 재직 중이며 주요 관심 연구분야는 비즈니스인텔리전스, IT전략 등이다.