韓國IT서비스學會誌
第9卷 第1號
2010年 3月

고객군의 지리적 패턴 발견을 위한 데이터마트 구현과 시각적 분석에 관한 연구*

조재희** · 하병국***

Buying Pattern Discovery Using Spatio-Temporal Data Mart and Visual Analysis*

JaeHee Cho** · ByungKook Ha***

■ Abstract ■

Due to the development of information technology and business related to geographical location of customer, the need for the storage and analysis of geographical location data is increasing rapidly. Geographical location data have a spatio-temporal nature which is different from typical business data. Therefore, different methods of data storage and analysis are required. This paper proposes a multi-dimensional data model and data visualization to analyze geographical location data efficiently and effectively. Purchase order data of an online farm products brokerage business was used to build prototype datamart. RFM scores are calculated to classify customers and geocoding technology is applied to display information on maps, thereby to enhance data visualization.

Keyword : Geographical Pattern of Customer, Data Modeling, Data Visualization, Data Mart

논문투고일 : 2010년 01월 17일 논문수정완료일 : 2010년 03월 05일 논문게재확정일 : 2010년 03월 08일
* 이 논문은 2009년도 광운대학교 교내 학술연구비 지원에 의해 연구되었음.
** 광운대학교 경영학부 교수
*** 광운대학교 경영정보학과 박사과정

1. 서 론

최근 들어 지리정보는 새로운 서비스로 주목 받고 있다. Google, MS 등은 'Google Earth', 'Live Search Map' 등의 지도검색 서비스를 통해 각종 지리정보와 길 찾기, 실시간 교통정보, 공사 정보 등을 제공하고 있으며, 점차 그 영역을 확대해가고 있다.

지리정보가 새로운 서비스 요소로 부각되고 있는 가운데, 지리정보를 활용한 고객관계관리(Customer Relationship Management)기술인 gCRM(geographic CRM)에 대한 관심이 증대 되고 있다[5]. 몇몇 사례를 살펴보면, 비씨카드사는 주거형태, 주변상권 등 카드사용에 영향을 미치는 공간적 특성에 따라 고객을 세분화한 후 마케팅에 활용하고 있다[4]. 국민은행은 GIS 점포전략시스템을 구축하여 영업과 관련된 내·외부 정보를 전자지도상에 기록하고, 이 정보를 토대로 각종 영업전략에 활용하고 있다[4].

이처럼 지리정보, 기업내부 데이터, 주변 지역의 외부데이터 등이 결합된 gCRM이 주목 받고 있지만, 그 한계 또한 보이고 있다. 지리정보시스템(Geographic Information Systems)을 기반으로 단일 기준의 고객정보를 지도에 점, 분포, 위치, 밀도의 형태로 표현하는 수준이다. 그리고 GIS 기반으로는 대용량의 데이터를 탐색하거나 패턴을 발견하는 데는 적절하지 않다[7]. 이는 사용자와 시스템의 상호작용이 부족하기 때문이다[8]. 이에 다각적이고, 심층적인 데이터 분석 기능을 제공하는 OLAP(On-Line Analytical Processing)과 GIS를 통합 하려는 연구노력이 필요하다[16].

본 연구에서는 지리정보를 활용한 고객관계관리에 OLAP 기능의 결합이 필요하다는 것을 사례를 통한 데이터마트 구현과 시각적 분석을 통하여 그 필요성을 이야기하고자 한다. 이를 위해 온라인 농산품 쇼핑몰인 'H'사의 1년 간의 주문데이터 22,214건을 사용하였다. 먼저 분석을 위해 분석용 다차원 모델을 제안 하고, 심도 있는 분석을 위해 CRM의 고객세분화 기법 중 RFM분석을 실시하였다. 고객의 지리정보분석을 진행하기 위하여 시각화 분석을 바탕으로 하였다. 지리적 시각화분석을 위해 고객의 주소정보를 지오코딩(Geocording)하여 지도 좌표데이터로 생성하였다. 제안된 모델에 따라 데이터마트를 구현하였다. 구현된 마트를 기반으로 시각적 분석 툴인 Tableau 5.0[26] 을 이용하여 분석을 수행하였다.

2. 관련 연구

2.1 gCRM에 관한 연구

gCRM은 일반적 CRM과 GIS의 결합으로, 일반적 CRM에 공간적 혹은 지리적 차원을 부가한 것이다[9]. gCRM은 고객, 상점들의 위치, 체인점의 접근성, 각 영업 지역의 잠재력과 같이 지역들 간의 관계 및 그런 관계들의 특징에 관심을 둔다는 점에서 일반적 CRM과 차이가 있다[13].

오늘날 일반적으로 사용되고 있는 gCRM 시스템들은 주로 상업지역 분석 및 타겟지역의 추출, 고객정보 분석 및 고객세분화, 상점 및 서비스 지점의 관리, 지도와 레포트를 이용한 분석 등의 기능을 제공한다.

이런 기능을 발휘하기 위해서는 고객의 정확한 위치 데이터의 습득, 고객 주소 데이터의 정제, 외부 데이터와의 통합이라는 기술적 전제 조건을 충족시켜야만 한다. 그러나 고객 위치 정보가 대부분 질문지 형식으로 수집되며, 외부 데이터는 개인 사생활 침해와 같은 법적 문제와 연루되기 때문에 분석 중심의 gCRM을 위한 전제 조건을 만족시키기란 쉽지 않다[2].

하지만 개인 모바일 GPS, RFID, 신용카드, IC교통카드, 개인 이동통신 단말기 등 기술의 발전으로 고객의 위치정보를 수집하고 활용할 수 있는 기회가 점차 늘어나고 있다.

2.2 고객세분화

세분화는 고객들의 기본적인 구매행태, 즉 고객이 무엇을, 언제, 어디서, 어떻게 구매하는지에 대

한 정보를 파악하는 것이다[3]. 세분화의 목적은 고객 집단별로 차별화 전략과 집중화 전략을 실시함으로써 핵심고객에게 서비스를 극대화하고자 하는 것이다.

세분화 기법으로는 <표 1>과 같이 전통적인 단일 기준에 의한 직관적 세분화 방법과 다기준 스코어에 의한 세분화 방법으로 구분할 수 있으며[11], 이러한 세분화의 기준변수로는 인구통계학적 변수(연령, 성별, 가족 수, 교육수준, 소득, 종교, 사회계층 등)와 고객 행동적 변수(추구혜택, 상표충성도, 가격민감도 등)가 있다[3].

〈표 1〉 고객세분화 기법

분류	상세기법
전통적인 기준	- 상품에 의한 세분화 - 연령에 의한 세분화 - 구매액에 의한 세분화 - 기타 고객 프로파일에 의한 세분화 등
다기준 스코어	- RFM에 의한 세분화 - LTV에 의한 세분화 - 통계모형에 기반한 세분화

어떤 세분화 방법을 선택할 것인가의 문제는 고객세분화의 목적, 보유하고 있는 고객 관련 데이터의 종류, 고객세분화를 위한 잠재적 기준 보유 여부 등에 따라 달라질 수 있다.

이이 중 RFM 분석은 비교적 원리가 간단하며 높은 반응률을 보이기 때문에 실무적으로 많이 활용되고 있다[12]. RFM 분석은 고객이 얼마나 최근에 구매를 했는지를 나타내는 최근성(Recency), 얼마나 자주 구매했는가를 나타내는 빈도(Frequency), 얼마나 구매했는가를 나타내는 구매액(Monetary Amount) 3개 요인을 분석하여 정량적으로 고객을 등급 분류하는 방법이다[15]. Hughes(2005)는 두가지 등급부여 방법을 제안했는데, 그것은 정 5등급 방법(Exact Quintiles)과 직분류 방법(Hard Coding)이다. 전자는 정확히 같은 크기의 5개 그룹을 생성하는 방법(각 그룹은 전체 고객 중 정확히 20%를 차지)이고, 후자는 각 셀마다 임의의 데이터 범위를 할당하여 분류하는 방법이다. 어떤 방법을 사용하더라도 많은 등급을 사용할수록 더 세부적이고 과학적인 분석이 가능할 것이다. 그러나 많은 그룹을 분류하는데 많은 비용이 소요될 수 있어 10등급 이상으로 그룹을 세분화 할 필요는 없으며, 대상의 특성을 고려한 모델 설계자의 직관에 의해 등급 수가 결정된다[15].

RFM 점수는 고객의 수익기여도를 나타내는 세 가지 지표들의 선형결합을 구한 점수(Score)이다.

RFM 점수는 최근 구매시기(R), 총 구매수량(F), 총 구매금액(M) 고객의 수익기여도를 나타내는 세 가지 지표들의 선형결합을 구한 점수(Score)이다. 즉 가장 최근에, 얼마나 자주, 상품 구매에 많은 비용을 들였는가에 대한 개별 정보를 수치화 시켜 각각의 요인(R, F, M)에 가중치를 부여한 후 합산하여 구하는 방법으로 가장 일반적인 RFM모형은 다음과 같다.

$$\text{RFM가중합} = W_1 \times R + W_2 \times F + W_3 \times M$$
$$\text{RFM점수} = (\text{RFM가중합}/\ 5) \times 100$$

여기서 W_1, W_2, W_3은 각 요인에 대한 가중치가 된다. 각 변수들의 가중치를 산출하는 방식에는 크게 통계적 추정방식, 거래 비율 분할 방식 그리고 보정 지수 산출방식이 있다. 먼저 통계적 추정방식이란 회귀분석과 같은 통계적 분석 방법을 통해 각 변수의 가중치를 추정하는 것이다. 즉, 수익성을 종속 변수로 R, F, M을 각각 독립변수로 설정하여 다중 회귀분석을 수행함으로써 각 변수들에 대한 계수(Coefficient) 값을 추정해 낼 수 있다. 둘째, 거래 비율 분할 방식이란 R, F, M의 각 점수 구간(보통 1~5점)에 포함되는 고객들의 비율을 바탕으로 가중치를 산정하는 방식이다. 일반적으로 4등급과 5등급의 고객 비율을 평균하여 각 R, F, M의 가중치로 활용한다. 끝으로 보정 지수 산출은 특정 변수의 점수가 지나치게 과장되거나, 특정 변수의 의미가 잘못 해석될 수 있는 위험을 해소하기 위해 변수들의 점수를 보정할 수 있는 지수를 제공하는 방법이다[1].

2.3 다차원 모델 개발을 위한 방법론

분석용 데이터베이스인 데이터웨어하우스를 구축하기 위해서는 테이블 형태(2차원)의 운영계 데이터베이스를 다차원적으로 분석할 수 있는 형태로 변환시킨 것이 다차원 모델이다. 다차원 모델은 비즈니스 정보에 대한 논리적 모델로서 전적으로 사용자의 분석목적에 맞도록 사용자의 관점에서 설계하며 비즈니스를 표현하는 가장 자연스러운 형태이다[6].

분석계 데이터베이스인 데이터웨어하우스 또는 데이터마트를 구현하는 데 중요한 이슈 중의 하나는 질의, 탐색, 보고, 분석을 지원하는 다차원 데이터 모델을 개발하는 방법에 관한 것이다[14]. 지금까지의 다차원 데이터 모델 개발 방법은 기본적으로 세가지(데이터 중심적, 목적 중심적, 사용자 중심적) 방법이 존재한다[7, 18].

데이터 중심 방법은 소스 데이터를 분석하여 논리적인 데이터 스키마(스타, 스노우플레이크, 스타ER, ME/R)를 설계하는 것이다. 데이터 중심적 접근 방법은 소스 데이터만을 분석하여 사용자의 정보 요구에 적절히 부응할 수 없다는 문제점을 갖는다.

목적 중심적 방법은 데이터웨어하우스 또는 데이터마트를 기업의 전략과 비즈니스 목적에 맞추는 것을 강조한다. 기업의 전략적 목적과 고객 서비스를 확인하고, 이와 관련된 주요 비즈니스 프로세스를 SOM(Semantic Object Model) 스키마를 적용하여 분석한 후, 각 프로세스에서 요구되는 차원과 측정지표를 확인하여 다차원 데이터 모델을 구성한다.

마지막으로 사용자 중심적 방법은 사용자 요구사항 분석 프로세스에 초점을 맞추어 최종 사용자의 참여를 촉진시킨다. Westerman(2001)은 월마트에서 개발한 방법을 기술하면서[24], 일반적인 비즈니스 요구에 기반하여 우선 프로토타입을 개발하고 이에 대한 사용자의 요구사항을 수집 및 분류하여 프로토타입에 적용하는 사용자 중심적 방법을 제안하였다.

앞에서 살펴본 다차원 데이터 모델 개발을 위한 세 가지 방법론은 상호보완적이며, 이들을 통합하는 연구들이 진행되고 있다. 따라서 다차원 데이터 모델 개발 시 데이터 관점, 목적 관점, 사용자 관점을 균형적으로 반영하려는 노력이 필요하다.

본 연구는 데이터 중심적 방법과 사용자 중심적 방법을 혼용하여 구축하였다. 즉, 일차적으로 가용한 데이터를 중심으로 프로토타입(데이터마트와 분석화면)을 완성하였고, 향후 인터넷몰 관계자에게 프로토타입과 그 결과물을 제공하여 피드백을 얻고, 그 피드백을 기초로 수정 · 보완 작업을 진행하였다. 그리고 다차원 정보분석을 위해 OLAP의 기본적인 데이터 접근 방식인 드릴다운(Drill-Down), 드릴업(Drill-Up), 드릴어크로스(Drill-Across), 드릴쓰루(Drill-Through)를 활용한 분석을 실시하였다.

2.4 데이터 시각화 연구

시각화는 시각적 이미지를 사용하여 데이터를 표현하는 프로세스로 정의할 수 있으며[17], 표현하려는 데이터의 특성에 따라 크게 과학적 시각화와 데이터 시각화로 구분된다[19]. Tegaden(1999)은 비즈니스 데이터 시각화를 '비즈니스 데이터를 시각화하기 위해 시각화 기술을 사용하는 것'으로 정의하였다. 비즈니스 데이터는 다른 유형의 데이터와 비교하여 추상적, 이산적, 다차원적이기 때문에 상대적으로 시각화가 어렵다[22].

Tegaden(1999)은 데이터 시각화 설계에서 가장 중요한 것은 적절한 시각적 표현의 선택으로 데이터 시각화 설계자는 관련 데이터와 업무의 이해를 수반해야 한다고 하였다. 이를 간단히 정리하면 다음과 같다.

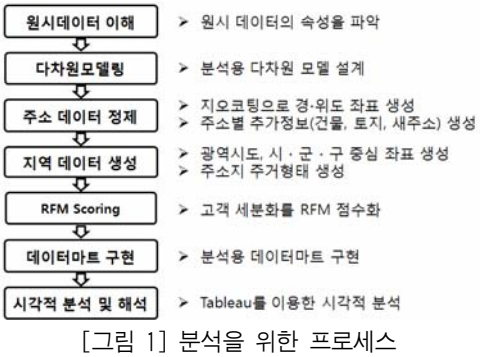
첫째, 설계자는 시각화의 기초가 되는 원시 데이터의 속성을 이해해야 한다. 데이터 속성은 데이터 소스, 데이터 유형(이산적 또는 연속적), 데이터 차원성, 데이터 수준(명목, 순차, 등간, 비율), 데이터 신뢰성 등을 포함한다. 데이터의 속성을 이해하는 한 방법으로 데이터 모델링이 유용할 수 있다[21].

둘째, 시각적 표현 형식은 관련 업무에 따라 결정되므로 설계자는 의사결정자에 의해 수행되는 업무에 집중해야 한다. 인지적합이론에서는 데이터의 표현 형식과 의사결정 업무 사이의 적합이 문제해결 성과에 긍정적인 영향을 미친다는 것이다[23]. 데이터 표현 형식과 업무가 의사결정자에게 동일한 정보와 프로세스를 요구할 때 인지적합이 발생하고, 일괄된 정식표상을 만들어 결과적으로 의사결정자의 신속하고 정확한 의사결정을 돕는다.

마지막으로, 설계자는 적절한 시각화 구성요소를 사용해야 한다. 시각화 구성요소는 시각적 기호 (Visual Glyphs)와 시각적 속성(Visual Property)으로 구성된다[10]. 시각적 기호에는 점, 선, 면, 아이콘 등이 있으며, 시각적 기호에 데이터 개체 값을 부여하면 시각적 속성이 된다. 기본적으로 시각적 속성에는 위치(X 좌표, Y 좌표, Z 좌표 등), 크기(높이, 길이, 폭 등), 색, 방위, 모양, 애니메이션 등이 있다[20].

3. 분석을 위한 다차원 모델

고객군의 지리적 패턴 발견과 시각적 분석을 위하여 [그림 1]과 같은 프로세스를 수행하였다. 1) 원시데이터의 각 필드의 속성을 파악하였다. 2) 분석을 위한 다차원 모델을 설계하였다. 3) 공간데이터인 주소를 지오코딩하여 경·위도 좌표를 생성하고 주소정제 및 추가 데이터를 생성하였다. 4) 생성된 추가 데이터를 바탕으로 지리적 데이터를 생성하였다. 5) 고객 세분화를 위한 RFM Scoring을 실시하였다. 6) 생성된 데이터를 바탕으로 데이터마트를 구현하였다. 7) 시각적 분석 툴인 Tableau 5.0을 이용하여 분석을 실시하였다.



[그림 1] 분석을 위한 프로세스

3.1 원시데이터 이해

본 연구에 사용된 데이터는 인터넷 농산품 쇼핑몰에서 2008년 9월 25일부터 2009년 9월 24일 동안 발생한 주문정보 22,214건이다. 원시데이터는 <표 2>와 같이 10개의 필드로 구성되어 있다.

〈표 2〉 원시데이터 필드

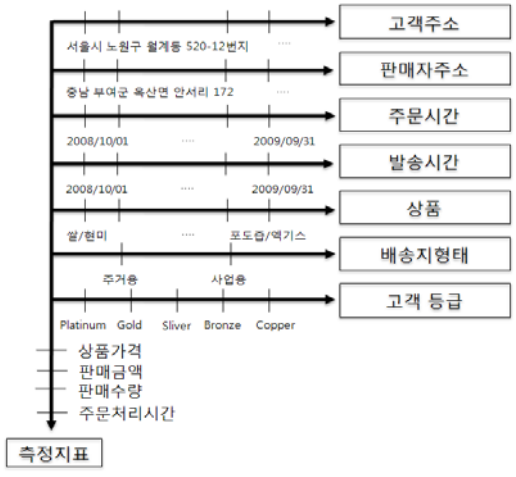
필드명	설 명
일련번호	주문 일련번호
주문일	년/월/일 시:분:초
발송일	년/월/일 시:분:초
판매자주소	농산물 판매자 주소
고객주소	주문 배송지 주소
카테고리1	상품의 소분류
상품명	엑기스, 쌀 10kg, 백미 20kg
매출액	주문 수량 × 상품 단위가격
상품가격	쇼핑몰에서 판매 되는 가격
수량	한 상품에 대한 주문 수량

3.2 다차원 모델 설계

고객별 주문정보가 내포하고 있는 지리적 정보를 기반으로 고객 분석을 하기 위해서는 [그림 2]와 같이 8개 차원을 갖는 다차원 모델이 필요하다. <표 3>은 다차원 모델의 세부구조를 설명하고 있다. 각 차원과 차원항목, 측정지표, 비즈니스 룰(측정지표의 관계식, 차원항목의 계측구조식)을 정리한 것이다.

공간차원인 고객주소와 판매자 주소는 '서울 광진구 모진동 214번지' 등과 같은 서술적인 요소와 이를 지도위에 표현하기 위해 '(127.073056, 37.547144)'와 같은 지리좌표를 포함한다.

이는 Rivest 외 2인(2001)이 말한 비지리적 공간



[그림 2] 다차원 모델 구조

차원, 지리적공간차원, 혼합형 공차원항목의 특성만을 가지고 공간차원의 유형을 설명하고 있으나, 이는 차원항목이 항목의 특성을 기술하는 애트리뷰트(Atribute)를 가질 수 있다는 사실을 간과한 것이다[8]. 비지리적 공간차원항목인 고객주소와 판매자주소는 애트리뷰터로 지리좌표를 가짐으로써 지리적인 표현과 분석이 가능해진다. 이러한 구조를 <표 5>에서 '서울 광진구 모진동 214번지 2층 (127.073056, 37.547144)'으로 표기하였다.

고객·판매자 주소 차원은 '광역시·도'→'시·군·구'→'읍·면·동' 순으로 각 레벨 사이에 계층구조를 가지며 상위레벨인 행정구역을 지리적 분석단위로 설정 하였다. 주문시간과 발송시간은 '년, 월, 일, 시, 분, 초' 순으로 각 레벨 사이에 계층구조를 가지며 분석은 최소 '시' 단위로 실시하였다. 상품 또한 '카테고리'→'카테고리 2'→'상품명'의 순으로 각 레벨사이에 계층구조를 가진다. 배송지형태는 배송지의 건물과 주소지의 상호를 기반으로 주거용과 사무용으로 설정하였다. 고객등급은 RFM 점수를 바탕으로 5그룹으로 세분화 하였다.

3.3 주소 데이터 정제 및 지역데이터 생성

지리적 정보를 시각적으로 표현하기 위해서는 명목적 공간(주소)에 대한 지리좌표 데이터를 추가해야 한다. 그 대표적인 방법으로 주소정보를 지리좌표인 경도와 위도로 변환하는 지오코딩(Geocoding)을 들 수 있다. 예를 들어 고객이 '서울시 강남구 대치동 **2차 아파트'에 대한 경도와 위도(127.049539, 37.498201) 좌표 값을 자동으로 찾아주는 것이다. 이

〈표 3〉 다차원 모델 세부구조

차원	차원항목	비즈니스 룰			
측정지표	상품가격 , 판매수량				
	주문처리시간	발송시간-주문시간			
고객주소	서울 광진구 모진동 214번지 (127.073056, 37.547144), …	광역시·도	시·군·구	읍·면·동	합계
		서울, …	광진구, …	모진동, …	
판매자주소	전북 완주군 화산면 종리 용수마을 산하농원 (127.23183, 36.005722), …	광역시·도	시·군·구	읍·면·동	합계
		전북, …	완주군, …	화산면, …	
주문시간	2008/9/25~2009/9/24	년/월/일/시			
발송시간	2008/9/25~2009/9/24	년/월/일/시			
상품	해바라기부사/10kg/32과, ….	카테고리 2		카테고리 1	
		사과, …		과일/견과류, …	
배송지형태	아파트, 빌라, 주택, 맨션, …	주거용			
	학교, 공장, 건물, 회사, …	사업용			
고객등급	Platinum, Gold, Silver, Bronze, Copper	RFM점수			

를 이용하여 고객의 위치를 지도상에서 확인할 수 있게 되며, 보다 지리적 속성을 잘 이해할 수 있다.

지오코딩을 실시함에 있어 중요한 부분이 정확한 주소이다. 입력한 주소가 오타나 단어누락으로 잘못 입력되었을 때 좌표 값을 찾을 수 없다. 이를 보완하는 것이 주소정제 작업이다. 주소정제에는 오타나 단어누락 뿐만 아니라 행정구역 변경이나 새로운 주소 체계로 변경되었을 때 기존의 내용을 변경해주는 것을 말한다. 이런 지오코딩 및 정제 서비스 또한 새로운 비즈니스로 선보이고 있다. 국내에서는 약 5개 정도의 전문솔루션 업체가 있는 것으로 파악된다. 본 연구에서는 그 중 오픈메이트사의 지오코딩 전문솔루션 서비스[25]를 사용하였다.

오픈메이트의 지오코딩 솔루션을 선택한 이유로는 지오코딩과 주소정제를 한번에 해결 할 수 있는 서비스이며, 추가적으로 활용할 수 있는 데이터가 제공 되었다. 추가적으로 제공된 데이터로는 건물종류, 지번면적, 소유구분, 지대가격 등 40개 항목이다. 그 중 건물종류를 활용하여 고객의 주거형태를 구분하는데 참고하였다. 현재 오픈메이트의 지오코딩 솔루션은 제한적으로 무료 서비스를 실시하고 있다.

3.4 고객 세분화를 위한 RFM점수

본 연구에서는 고객의 수익성 분석에 활용되는 RFM 모형을 사용하여 고객 세분화를 실시하였다. 연구에서 R, F, M의 점수는 정 5등급(Exact Quintiles)을 적용 및 산출하여 <표 4>와 같이 R, F, M 점수 기준표를 완성하였다. 본 연구에서는 가중치 선정을 위해 파레토 법칙을 적용한 거래비율분할방식을 사용하였다. 거래비율분할방식은 먼저 상위 20% 고객의 최근성, 빈도, 구매금액을 가중치의 기준으로 정하고 각각의 전체 값에서 상위 20% 고객이 차지하는 비율을 구한다. 이때 최근성은 기간의 숫자값과 역의 의미를 가지므로 다음과 같이 {(전체평균 최근성-상위 20%의 평균 최근성)/전체평균 최근성}으로 비율을 계산한다. <표 5>는 본 연구에서 사용한 온라인 농산품 쇼핑몰 고객의 RFM 가중치를 산정한 표이다.

〈표 4〉 R, F, M 점수 기준표

점수	R	F	M
1점	257일~	1회	~25,000원
2점	166~256일	2회	25,100~36,000원
3점	103~165일	3회	36,100~55,000원
4점	58~102일	4회	55,100~100,000원
5점	1~57일	5회 이상	100,100원~

〈표 5〉 R, F, M 가중치 산정표

RFM	계산방법	전체	상위20%	비율	가중
R	가중평균	152.33일	120.42일	0.209	16%
F	빈도비율	22,214건	10,828	0.487	38%
M	금액비율	8**,***,935원	5**,***,285원	0.601	46%
합계				1.297	100%

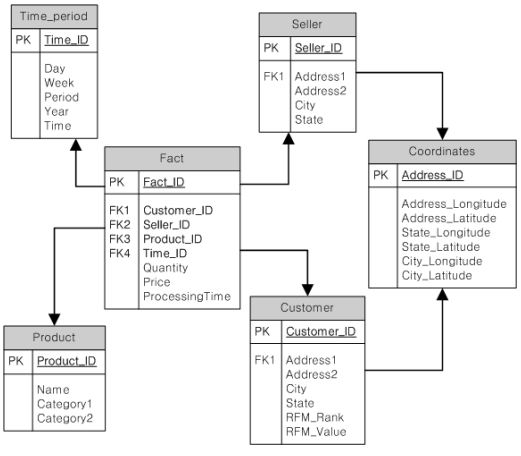
〈표 6〉 RFM 값에 따른 고객 등급

RFM값	등급	고객수	비율
20~36	Copper	3,499	33.09%
37~52	Bronze	2,970	28.09%
53~68	Silver	2,258	21.35%
69~84	Gold	1,030	9.74%
85~100	Platinum	818	7.74%

앞 제 2.2절에서 살펴본 RFM 모형에 따르면 RFM은 최하 20에서 최고 100까지의 값을 가진다. Hughes (2005)의 직분류(Hard Coding)방법을 사용하여 RFM 값을 16점 단위로 구분하여 고객을 5등급으로 나눌 수 있다. 이와 같은 방식을 사용하여 온라인 농산물 쇼핑몰 고객 등급별 고객수를 <표 6>과 같이 정리하였다.

3.5 데이터마트 구축

지도 위에 고객데이터를 시각적으로 표현하기 위하여 제안된 다차원 모델을 반영하여 [그림 3]과 같이 1개의 사실 테이블과 5개의 차원 테이블을



[그림 3] 스노우플레이크 스키마

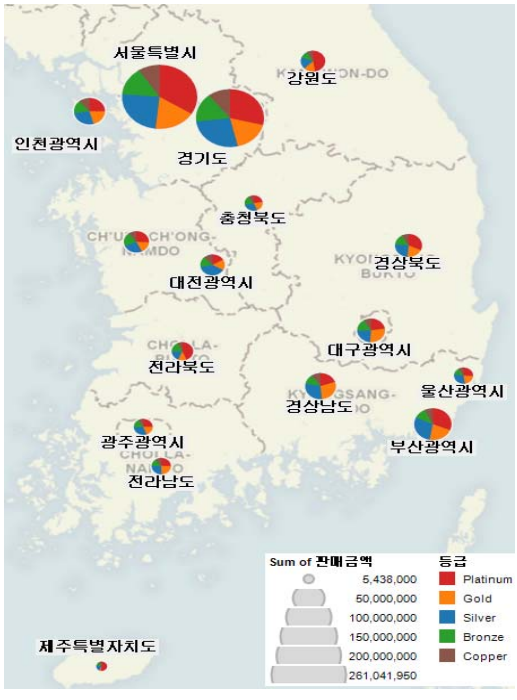
갖는 스노우플레이크 스키마를 설계하였다.

사실(fact) 테이블은 시간(Time ID), 상품(Product ID), 고객(Customer ID), 판매자(Seller ID)를 속성으로 가지며 이들은 각각의 차원 테이블과 조인되어 있다. 그리고 측정지표인 상품가격(Price), 판매수량(Quantity), 주문처리시간(Processing Time)을 속성으로 갖는다. 지리좌표(Coordinates) 테이블은 주소의 경도(Address_Longitude), 주소의 위도(Address_Latitude), 광역시·도의 경도(State_Longitude), 광역시·도의 위도(State_Latitude), 시·군·구의 경도(City_Longitude), 시·군·구의 위도(City_Latitude)의 속성을 갖는다. 지리좌표(Coordinates) 테이블은 향후 다른 공간 데이터와의 연계를 위하여 고객(Customer) 테이블과 판매자(Seller) 테이블로부터 분리되었다. 이를 기반으로 관계형 데이터베이스를 구축하였다.

4. 분석 및 해석

OLAP 툴 중 시각적 분석 기능이 우수한 Tableau Software사의 Tableau 5.0을 사용하여 데이터마트를 분석한 결과이다. OLAP의 기본적인 드릴다운, 드릴쓰루를 활용한 분석으로 [그림 4]~[그림 7]은 사용자가 직접 툴을 조작하여 분석을 실시한 화면이다.

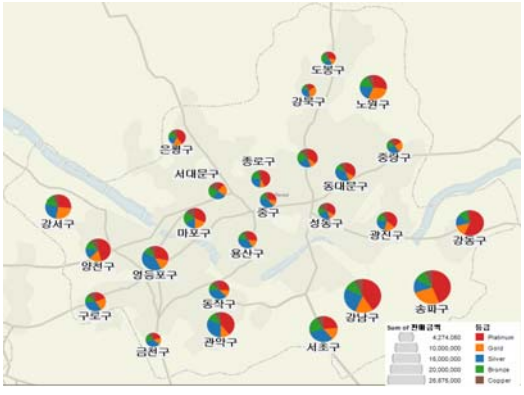
[그림 4]는 탐색적으로 고객을 분석하는 것으로



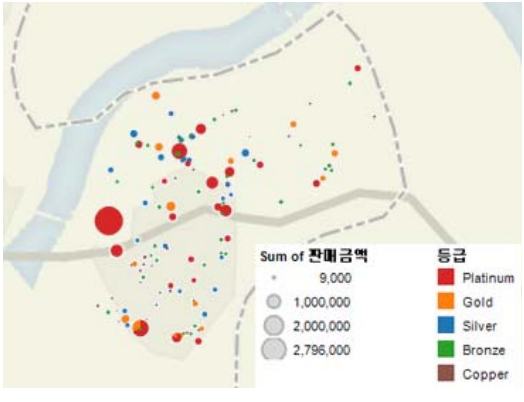
[그림 4] '광역시·도' 별 고객군 현황

지리정보의 상위레벨인 각 '광역시·도'별로 고객군의 비율이 어떠한지를 한눈에 확인할 수 있다. 이를 위해 X축에는 광역시·도의 경도좌표를 Y축에는 광역시·도의 위도좌표를 위치시켜야 한다. Tableau 5.0에서는 Web Map Service(WMS)를 지원하므로 경·위도 좌표가 입력되면 자동으로 지도를 생성하여 사용할 수 있다. 그리고 단순히 점으로 표현될 때 보다 집계정보를 활용하여 사용자의 분석력을 향상 시킬 수 있다. 원의 크기는 각 지역별 매출의 합을 나타내고 있고, 각 조각의 색은 고객군을 나타내고 있다.

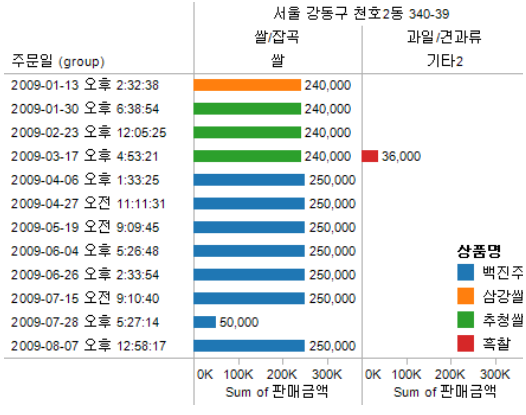
[그림 4]에서 보는 봐와 같이 각 지역별 매출 규모가 서울, 경기, 부산 순으로 많은 것을 알 수 있으며 특히 서울, 경기 지역이 월등히 많다는 것을 확인할 수 있다. 그렇다면 많은 매출을 보이는 서울지역 내에서의 현황이 궁금하다. 이는 분석의 단위를 서울지역으로 드릴다운 하여 분석을 실시할 수 있다. 이를 위해 X, Y축에 시·군·구의 경·위도 좌표를 위치시키고 서울지역만을 선택하



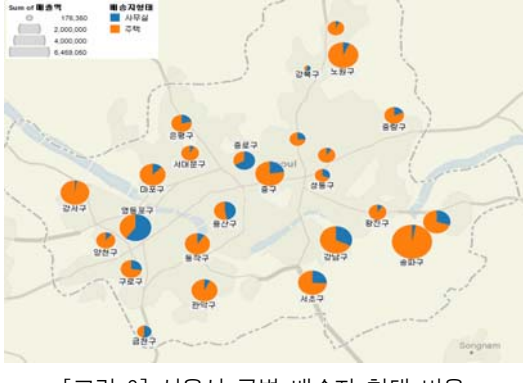
[그림 5] 서울시 구별 고객군 매출 현황



[그림 6] 서울 강동구의 고객군 분포



[그림 7] 고객 구매 패턴 분석



[그림 8] 서울시 구별 배송지 형태 비율

면 [그림 5]와 같은 화면을 생성할 수 있다.

[그림 5]는 서울지역의 각 구별, 고객군별 매출 현황이다. 같은 서울지역 내에서도 강남, 송파, 강동 지역이 많은 매출을 보이는 것으로 확인된다. 그리고 다른 지역에 비해 강동구의 고객을 보면 Platinum 고객군이 다른 지역에 비해 많은 것을 알 수 있다.

[그림 6]은 강동구 지역의 고객군 위치를 나타내고 있다. 색깔은 각 고객군을 의미하며, 원의 크기는 구매금액을 크기로 나타내고 있다. 눈에 띄는 Platinum 고객을 발견할 수 있다. 그리고 발견된 고객을 선택하여 구매이력을 드릴쓰루 하면 [그림 7]과 같은 구매패턴을 발견할 수 있다. 고객은 쌀 제품을 정기적으로 구매하고 있으며, 쌀 제품 중에서 '백진주' 제품을 선호하는 것을 알 수 있다.

[그림 4]~[그림 7]의 일련의 활동들은 기존의 GIS와 비교한다면 정보를 탐색하는 과정에서 사용자가 방대한 분량의 데이터에 접근하여 드릴다운(Drill-Down), 드릴쓰루(Drill-Through)의 방식을 사용한 분석이 가능하다는 것을 보여 주었다. 또한 하나의 분석도구에서 여러 형태의 시각적 분석이 가능하다는 잘 보여준다.

[그림 8]은 상품 중 매출 규모 상위 30개 제품을 대상으로 서울지역의 각 구의 배송지 형태별 판매금액 비율을 나타내고 있다.

배송지 형태를 주거용, 상업용으로 구분하여 색으로 표현 하였다. 원의 크기는 지역의 판매금액을 나타내고 있다. [그림 8]에서 확인할 수 있는 것은 대부분 주거용 배송지의 비율이 높은 것을 알 수 있다. 특이점으로 종로구와 영등포구는 상

업용이 절반 이상을 차지하는 것을 확인할 수 있다. 두 지역의 배송지 형태별 특성을 확인하기 위해 두 지역의 상품 구매이력을 확인하였다.

[그림 9]는 두 지역의 상업용 배송지의 상품 목록을 나타내고 있다. 상품명을 자세히 살펴보면, 가공류의 제품들이라는 것을 확인할 수 있다. 상품의 특성과 지역의 특성을 고려하였을 때, 간편식 위주의 구매패턴을 생각할 수 있다.

[그림 10]은 2009년 4월부터 9월까지 서울지역의 고객군별로 카테고리1 수준에서의 상품 구매를 지역적으로 표현한 것이다. 좌측(X축)으로 갈수록

상품카테고..	상품명	영등포구	종로구
기타4	야채스프(유기농무농약)		
딸기	딸기원액(산하유기농)		
복분자	천환경무농약인증 복분자 생과5kg,10..		
기능식품	합초환800g (200g 추가증정)		
	홍화씨분말가루(500g)		
	리필용홍화씨환500g		
	홍화씨환(200g)		
냉동	유기아이스 딸기(2009년설향)		
	아이스딸기(설향2009년)		
엑기스	가시오갈피엑기스(100gx40봉)		
	가시오갈피엑기스(100gx60봉)		
	천환경 단비네 꿀복분자 원액		
	가시오가피엑기스		
음료	천환경 단비네복분자 생생즙(100%)		

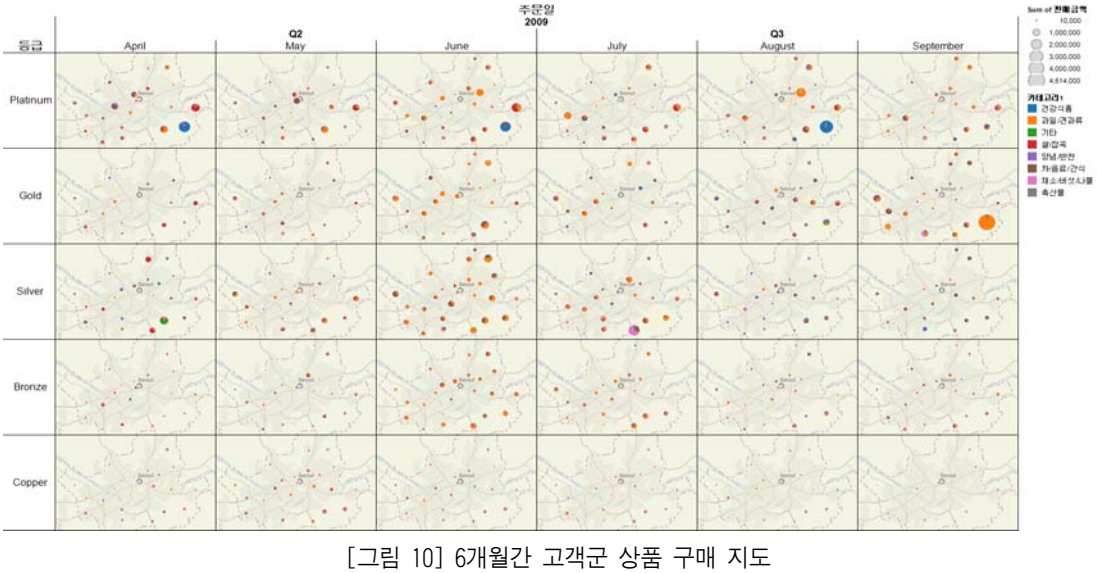
[그림 9] 상업용 배송지 상품목록(영등포구, 종로구)

시간의 변화를 나타내고 있으며, 위에서 아래(Y축)쪽으로 각 고객의 등급을 나타내고 있다. 원의 크기는 주문횟수를 나타내고, 상품의 분류는 파이 차트를 이용하여 각 조각으로 표현되었다. 상단의 Platinum 고객군은 시간의 변화에 큰 변화 없이 주기적으로 일정 수준의 주문을 보이고 있다. 지리적으로 보면 은평구, 노원구, 강남구, 강동구에서 일정한 구매의 형태를 확인할 수 있다. 그 반면에 Silver, Bronze 고객군에서는 8월, 9월 주문량이 주는 것으로 확인되었다. Cooper 고객군은 주문이 거의 발생하지 않는 것을 확인할 수 있다. 이는 Cooper 고객군의 특성상 신규 고객이거나 이탈 고객이 다수 포함되어 있어 고객분석의 요구가 필요하다고 판단된다.

[그림 10]은 기존의 GIS가 단일차원의 시각적 분석을 제공하였다면, OLAP 기능을 활용하여 시간, 공간, 상품, 고객의 다차원적 시각분석이 가능하다는 것을 보여주고 있다.

5. 결론 및 향후 과제

지리정보를 활용한 고객관계관리를 위한 단일 GIS 기반의 gCRM의 한계를 극복하고자 OLAP의 활용 가능성을 사례를 통해 제시하였다. 이를 위해 다차원모델을 설계하고, 데이터마트를 구축하



[그림 10] 6개월간 고객군 상품 구매 지도

였으며, 분석 도구를 통하여 분석과 해석을 실시하였다. 분석의 결과를 실무 담당자에게 제시하여 다음과 같은 의견을 받았다.

"고객 위치정보를 고객관계관리 시스템에 포함시키면 시각화 효과가 커서 지역별 프로모션을 위한 대상 고객발굴에 좋을 것 같다. 현재 본사는 오프라인 매장이 없고, 전량 택배를 통해 고객에게 우송되므로 지리적 영향을 크게 받지 않는다. 그러나 농가체험 상품이 증가추세 인데, 이러한 상품 분석 및 개발에 있어서 지역과 거리 요인을 고려할 필요가 있다."

본 연구의 목적은 지리정보를 활용한 고객관계관리에 OLAP 기능의 결합의 필요성을 제시 하고자 하였다. 그리고 사례를 통한 테이터마트 구현과 시각적 분석을 제시함으로서 논지를 뒷받침하였다. 기존의 GIS 기반에서는 단일 기준의 고객정보를 지도에 표현하였다면 이번 연구에서는 다차원적으로 지도상에 표현하였다. 그리고 대용량의 데이터를 탐색하는 데는 적절하였다. 물론 지리적 패턴을 발견하지는 못 하였으나 연구의 범위가 지리적 패턴 발견을 위한 데이터마트 구현과 시각적 분석이 가능하다는 것을 제안하는 수준에서는 충분하였다.

지리정보의 활용분야는 다양하고, 기회가 많이 있다고는 하지만 아직 적용분야의 한계가 존재하는 것은 사실이다. 서론에서 말한 적용 사례들 또한 오프라인 비즈니스들이 대부분이다. 앞으로 지리정보를 활용한 고객관계관리의 비즈니스 활용에 대한 연구가 필요하다.

본 연구를 실시함에 있어 한계로는 OLAP 기능과 GIS 기능을 동시에 만족하는 분석도구가 없어 시각적 분석이 띄어난 Tableau 5.0을 사용하였으나, 지리적 속성을 파악하기에는 지도 서비스(WMS)의 수준이 낮았다. 그리고 아파트와 같은 다층구조의 지리정보를 활용하기에 한계가 있었다. 이런 면에서 본다면 다양한 지도의 활용으로 지리적 속성(위치, 건물, 지형, 인구밀도, 상업지 등)을 활용할 수 있는 GIS의 장점을 활용할 수 있는 연구가 요망된다.

참 고 문 헌

[1] 김형수, 김영걸, 박찬욱, 「CRM 고객관계관리 전략 원리와 응용」, 사이텍미디어, 2009.

[2] 박기호, 정재곤, 황명화, "P2P LBS 기술을 이용한 gCRM의 영역 확대에 관한 연구", 「한국GIS학회 춘·추계학술대회」, 2003.

[3] 박진수, 장남식, 황유섭, "S카드사의 가맹점 분류체계 정비를 통한 고객세분화 전략", 「Information Systems Review」, Vol.10, No.3(2008), pp.89-109.

[4] 윤용근, "gCRM 기법을 이용한 상권분석 연구", 「정보기술연구」, 2005, pp.29-43.

[5] 정보통신산업진흥원, 「S/W시장 분석 : g-CRM 시장 동향 및 전망」, 정보통신산업진흥원, 2008.

[6] 조재희, 박성진, 「OLAP 테크놀로지」, 시그마컨설팅그룹, 1999.

[7] 조재희, 서일정, "이동객체의 데이터 시각화를 통한 이동패턴 분석에 관한 연구", 「한국IT서비스학회지」, Vol.6, No.1(2007), pp.127-140.

[8] 조재희, 서일정, "시공간데이터 분석을 위한 다차원 모델과 시각적 표현에 관한 연구", 「Journal of Information Technology Applications and Management」, Vol.13, No.1(2006), pp.137-147.

[9] 황정래, 강혜영, 이기준, "시산과 공간 데이터 마이닝을 적용한 gCRM 구조", 「한국정보과학회」, 학술발표논문집, Vol.30, No.1(2003).

[10] Card, S. K., J. D. Mackinlay, and B. Sheiderman, *Readings in Information Visualization : Using Vision to Think*, San Francisco, CA : Morgan Kaufmann, 1999.

[11] Don Peppers and Martha Rogers, *Enterprise One to One 1rd Edition*, New York : Currency Doubleday, 1999.

[12] Gattuso, G., "Clothing Retailer Finds Database a Good Fit", Direct Marketing, Vol.5

(1994), pp.26-27.

[13] GISMO, *Geo-Marketing at your fingertips with GISMO*, 6th EC-GE and GIS workshop, Presentation Document, 2000.

[14] Guo, Y., S. Tang, Y. Tong, and D. Yang, *Triple-Driven Data Modeling Methodology in Data Warehousing : A Case Study*, In Proc. 9th ACM international workshop on Data warehousing and OLAP, 2006, pp.59-66.

[15] Hughes, M., *Strategic Database Marketing, 3rd Ed*, New York ; McGraw-Hill, 2005.

[16] Kuijpers, B. and A. A. Vaisman, *A Data Model for Moving Objects Supporting Aggregation*, in Data Engineering Workshop, IEEE 23rd International Conference, 2007, pp.546-554.

[17] Latham, R, *The Dictionary of Computer Graphics and Virtual Reality 2nd ed.* Springer-Verlag, NY, 1995.

[18] List, B., R. Bruckner, K. Machaczek, and J. Schiefer, *A comparison of data warehouse development methodologies : Case study of the process warehouse*, In Proc. DEXA 2002, pp.203-215.

[19] Schroeder, W., K. Martin, and B. Lorensen, *The Visualization Toolkit : An Object-Oriented Approach to 3D Graphics 2nd edition.* Upper Saddle River, NJ; Prentice Hall, 1998.

[20] Tegarden, D. P., "Business information visualization", *Communication of the AIS*, Vol.1, No.4, 1999.

[21] Teorey, T. J., *Database Modeling and Design : The Entity-Relationship Approach*, Morgan Kaufmann, San Mateo, CA, 1990.

[22] VDI, *Visualization Design. Visible Decisions*, Inc. Toronto, CA, 1997.

[23] Vessey, I., "Cognitive Fit : A Theory Based Analysis of the Graphs versus Tables Literature", *Decision Sciences*, Vol.22, No.2(1991), pp.219-240.

[24] Westerman, P., *Data Warehousing using the Wal-Mart Model*, Morgan Kaufmann, 2001.

[25] http://www.geocoding.co.kr.

[26] http://www.tableausoftware.com.

◈ 저 자 소 개 ◈

조 재 희 (mis1@kw.ac.kr)

연세대학교에서 경영학 학사, Miami University (Ohio)에서 경영학 석사, University of Nebraska at Lincoln에서 경영정보학 박사를 취득하였다. 펜타시스템테크놀로지와 STM(LG CNS의 전신)에서 컨설턴트로 근무하였고, 현재 광운대학교 경영학부 교수로 재직 중이다. 주요 관심분야는 데이터자산의 전략적 활용, 다차원모델링, OLAP, 데이터웨어하우스, 비즈니스 인텔리전스 등이다.

하 병 국 (djkoogy@kw.ac.kr)

동국대학교 경제학 학사, 광운대학교 경영대학원 경영학 석사학위를 취득하였다. 현재 광운대학교 대학원에서 경영정보학 전공으로 박사과정에 재학 중이다. 주요 관심분야는 비즈니스 인텔리전스, CRM(Customer Relationship Management) 등이다.