韓國 컴퓨터情報學會 論文誌
第10卷 第3號, 2005. 7.

2005-10-3-1-10

실 위치지정자 자격으로서의 멤버 특성을 활용한 개인화 작업

이 덕 근*, 유 한 주**, 최 인 수***

Personalization Using Member Properties in the Physical Locator

Deok-Keun Lee*, Han-Ju Yu**, In-Soo Ch***

요 약

가상 위치지정자는 물리적 위치지정자의 콘텐츠를 기반으로 하는 논리적인 위치지정자이다. 이 콘텐츠들은 물리적인 위치지정자에 존재하는 멤버 특성인 것이 보통이다. 가상 위치지정자를 사용하여 CRM에서 가장 중요한 기술인 개인화를 이룩할 수 있으나, 본 연구에서는 집계수준을 늘려서 분석을 용이하게 하기 위하여 실 위치지정자를 사용하는 방법을 제안하고 있다. 이러한 실 위치지정자를 활용함으로써 성공적인 e-비즈니스를 수행할 수 있는 것이다.Abstract

Abstract

A virtual locator is a logical locator based on the contents of a physical locator. These contents can be existing member properties in the physical locator. Using virtual locator, we can accomplish personalization which is the technology area associated most often with CRM. In this study, however, what are called virtual locators in many OLAP models would be treated as physical locators for many unique aggregation levels. By using physical locators, we can bring a successful e-business.

▸ Keyword : Personalization, Virtual/Physical locator, Aggregation level, CRM, OLAP

• 제1저자 : 이덕근
• 접수일 : 2005.05.27, 심사완료일 : 2005.07.11
* 숭실대학교 산업·정보시스템공학과 박사수료
** 숭실대학교 산업·정보시스템공학과 박사과정
*** 숭실대학교 산업·정보시스템공학과 교수
※ 본 연구는 숭실대학교 교내 연구비 지원으로 이루어졌음.

I. 서 론

오늘날 기업과 고객과의 관계에서 발생하는 정보의 수요는 기업간의 시장 경쟁이 치열해 짐에 따라 증폭되고 있다. 기업간의 경쟁에서 살아남기 위해서는 얼마나 고객의 요구에 발 빠르게 반응하는가가 가장 중요한 관건이 된다[1,2,9]. 이러한 환경변화에 맞춰 나타난 것이 e-비즈니스이다. e-비즈니스란 어떤 것인가? 여러 정의가 있을 수 있겠으나 비즈니스와 고객 간의 새로운 관계를 정립하기 위해 인터넷 정보를 활용하도록 설계된 비즈니스를 e-비즈니스로 정의하고자 한다. 즉, 인터넷 스피드로 운영되는 비즈니스라 하겠다.

고객관계관리(Customer Relationship Management, 이하 CRM으로 기록)는 e-비즈니스의 핵심활동이다. 현재 CRM이 강조되는 이유는 비즈니스 철학이 오늘날 고객 위주의 철학으로 바뀌고 있기 때문이라고 생각한다. 인터넷 정보를 손쉽게 접할 수 있는 고객은 조금만 좋은 조건을 제시하는 경쟁기업이 있다면 그 기업으로 마음을 돌리기 때문에 과거 어느 때보다도 한층 상세하게 고객을 알아야만 한다. 고객과 지속적인 비즈니스 관계를 유지하는 것만이 기업이 경쟁에서 이기는 유일한 길이라고 생각한다.

기술면에서 볼 때 CRM에 가장 관련 있는 분야는 개인화(personalization) 분야이다. 개인화를 하자면 고객 각 개인에게 기업의 웹사이트를 접촉시켜야 한다. 이렇게 하면 일 대 일로 영업과 판매를 할 수 있고 고객이 다시 기업의 웹사이트로 찾아오게끔 할 수 있기 때문이다. 이러한 개인화를 성공 시키자면 고객에 대한 정보를 기업이 잘 수집·정리 하고 통합할 수 있게 하여야 하며 또한 고객에 대한 모든 가능한 정보를 유효하게 활용하게끔 하여야 하는데, 이를 가능하게 하는 것이 데이터 웨어하우스 (Data Warehouse)이다[2].

데이터 웨어하우스는 "다년간 기업의 운영시스템에 집적된 데이터를 주제영역별로 통합하여 별도의 프로그램 없이 다양한 각도에서의 분석을 가능케 하는 시스템"이다. 보다 실무적인 측면에서 보면 데이터 웨어하우스는 "의사결정을 위해 필요한 분석 정보를 자동으로 모아두고, 결합시켜 사용자가 원하는 시점에 원하는 형태로 제공하는 시스템"이라고 정의할 수 있겠다[3,9]. 데이터 웨어하우스에서의 데이터베이스는 거의 대부분이 관계형 데이터베이스로 되어 있는데, 이 점이 CRM의 관점에서 문제가 된다. 즉, 운용 데이터의 처리를 최적화 시키는 것을 목적으로 할 때에 관계형 데이터베이스가 적합할지는 몰라도 데이터의 분석업무에는 적합하지 않다는 문제가 생긴다는 것이다. 관계형 데이터베이스로 인해 데이터 웨어하우스에 야기되는 이러한 문제점을 극복하도록 설계된 것이 다차원 데이터베이스이다[1,3,4,5,10].

다차원 데이터베이스와 맞물려 긴밀하게 활용되는 툴(tool)로서 OLAP(OnLine Analytical Pro cessing)이 있다. OLAP은 사용자가 의사 결정에 필요한 지식을 찾아내기 위해 대량의 데이터를 쉽게 분석할 수 있도록 도와주는 데이터베이스 툴이고, 이 OLAP은 다차원 데이터베이스의 클라이언트(client)로 활약 하고 있다[1,5,8,9]. 본 연구에서는 OLAP을 활용하여 새로운 개념의 개인화 분석모델을 구축하는 것을 목적으로 하고 있다.

II. CRM과 개인화(Personalization)

오늘날 기업은 과거의 대량생산체제의 기업위주의 경영에서 고객위주의 경영으로 그 성격이 바뀌었다. 그 이유는 수많은 경쟁업체의 발생뿐만이 아니라 고객의 생활수준도 높아졌기 때문이다. 따라서 기업은 고객의 유치와 관리를 위해 항상 고객의 니즈를 파악하고 그것에 합당한 응답을 해야 한다. 그런데 문제는 이 과정에 필요한 시간의 폭이 점차 줄어들고 있다는 것이다. 그만큼 고객의 성향과 환경적인 요인의 변화가 빠르다는 것이다. 따라서 기업이 성공하기 위해서는 고객에 대한 사전 행동 및 바라는 것을 빠르게 예상하여야 한다. 또 수많은 고객들의 니즈를 파악하고 유지하기 위해서는 각각의 고객에 대한 관리가 필요로 하게 되는데 이러한 것들을 가능하게 하는 방법론이 CRM이다[2,6].

CRM은 고객과 관련된 기업의 내외부 데이터를 분석, 통합하여 고객 특성에 기초한 마케팅 활동을 계획하고, 지원하며, 평가하는 프로세스를 일컫는 말이다. CRM에서는 먼저 고객 데이터를 세분화시키며, 이에 기초하여 신규고객을 획득하고, 우수고객을 유지하며, 고객가치를 증진시키고, 잠재고객을 활성화시키며, 평생고객화를 이룩함으로써 적극적

으로 고객을 관리하고 고객의 가치를 극대화시킬 수 있게 된다[1,2,6].

이상에서 보다시피 CRM에서는 고객과 회사를 1 대 1로 연결시키는 개인화 부문이 핵심부문으로 자리 잡고 있음을 인식하여야 한다. 개인화를 달성시킬 기존의 OLAP에는 개개인의 패턴을 찾는 기법이 많이 결여되어 있다. 이 이유는 OLAP은 차원의 멤버와 멤버속성들의 수준에 대한 유연성을 확보하기 힘들기 때문이다. 다시 말하면 OLAP에서는 분석자의 기준에 맞는 세분화가 이뤄지지 않는다는 뜻이다. 유연성을 확보하기 위해 가상차원이라는 개념을 도입할 수 있지만 이 가상차원도 약점을 노출하고 있다.

본 연구에서는 고객 세분화의 한계점을 극복하고 새로운 개념의 개인화 분석모델을 만들기 위해 LC모델(Located Contents Model)의 개념을 활용하고 있다. LC모델에는 위치지정자(locator)와 콘텐츠(contents)가 있는데, 여기서 콘텐츠로 표현되는 부분이 바로 CRM의 핵심인 개인화를 가장 잘 표현할 수 있는 부분이 된다[1]. 본 연구에서는 CRM적인 분석 관점에서 중요하게 인식하여 활용할 수 있는 콘텐츠들을 다루는 효율적인 분석 방법을 제시하고자 한다.

III. LC Model과 가상차원

3.1 LC Model의 소개

LC(Located Contents)모델은 논리와 수학에 바탕을 둔 소프트웨어 데이터 모델이다. LC 모델에서, 위치지정자(locator)는 차원과 유사하며, 로케이션(location)은 차원 각각의 멤버의 교점과 유사하고, 콘텐츠(content)는 변수(variable), 사실(fact), 수식(formula), 측정값(measure) 또는 속성(attribute)과 유사한 개념이다[1]. 여기서 위치지정자와 콘텐츠는 고정된 개념이 아니라 분석자의 입장에 따라, 분석자의 주관에 따라서 달라질 수 있다. 다음과 같은 고객(Customer) 테이블을 예로 들어보자.

customer_id	name	address	gender	age_lies	married	children_id	job_id	education	yearly_income
1	Sheri Nowmer	2439 Bailey Road	F	20_lies	Y	4	7	High School Degree	40k~50k
2	Derrick Whelply	2219 Dawing Avenue	M	20_lies	Y	1	6	High School Degree	70k~80k
3	Jeanne Derry	7540 First Ave.	F	20_lies	Y	0	7	Bachelors Degree	50k~60k
4	Michael Spence	337 Tosca Way	M	60_lies	Y	2	5	High School Degree	20k~30k
5	Maya Gutierrez	8669 Via Neruda	F	40_lies	Y	3	7	Bachelors Degree	40k~50k
6	Robert Damstra	1619 Stillman Court	F	50_lies	N	0	10	Bachelors Degree	60k~90k
7	Rebecca Kanagaki	2660 D Mt. Hood Circle	F	30_lies	Y	0	5	High School Degree	40k~50k
8	Kim Brunner	6064 Brodia Court	M	50_lies	Y	2	1	Bachelors Degree	50k~60k
9	Brenda Blumberg	7560 Trees Drive	M	20_lies	N	0	2	High School Degree	20k~30k
10	Darren Stanz	1019 Kanwal Rd.	M	20_lies	Y	4	1	Bachelors Degree	40k~50k
11	Jonathan Murrain	5423 Camby Rd.	M	30_lies	N	0	9	High School Degree	60k~70k
12	Jewel Creek	1792 Belmont Rd.	F	30_lies	Y	1	1	High School Degree	40k~50k
13	Peggy Medina	3795 Keller Ridge	M	30_lies	Y	0	7	High School Degree	40k~50k

그림 1. 고객 테이블
Fig 1. Customer Table

고객에 대한 평가를 하는 데에는 일반적으로 고객 테이블 외에 시간(Time), 제품(Product), 사실(Fact) 테이블을 필요로 한다.

product_id	product_name	brand_name	product_price	product_cost	product_revenue
1	Washington Berry Juice	Washington	70	40	30
2	Washington Mango Drink	Washington	60	32	28
3	Washington Apple Drink	Washington	50	32	18
4	Jeffers Oatmeal	Jeffers	50	35	15
5	Jeffers Corn Puffs	Jeffers	200	120	80
6	Blue Label Canned Beets	Blue Label	300	200	100
7	Blue Label Creamed Corn	Blue Label	70	50	20
8	Green Ribbon Canned Mixed Fruit	Green Ribbon	30	22	8
9	Green Ribbon Canned Peaches	Green Ribbon	20	10	10
10	King Rosy Sunglasses	King	80	34	46
11	Queen Eyeglass Screwdriver	Queen	80	50	30
12	Queen City Map	Queen	60	32	28
13	Club Low Fat Cottage Cheese	Club	20	14	6

그림 2. 제품 테이블
Fig 2. Product Table

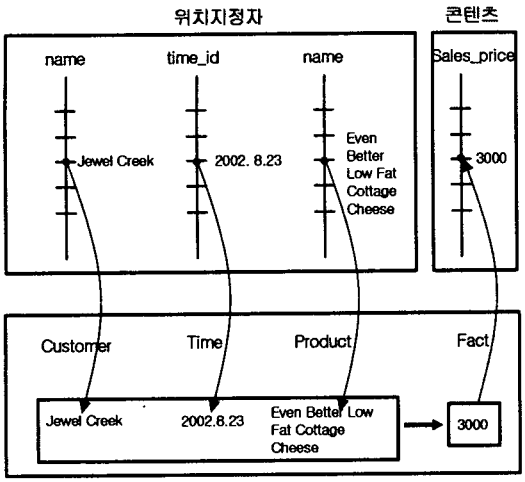
time_id	the_date	the_day	the_month	the_year	day_of_month	week_of_year	month_of_year	quarter
367	2002-01-01	Tuesday	January	2002	1	1	1	Q1
368	2002-01-02	Wednesday	January	2002	2	1	1	Q1
369	2002-01-03	Thursday	January	2002	3	1	1	Q1
370	2002-01-04	Friday	January	2002	4	1	1	Q1
371	2002-01-05	Saturday	January	2002	5	1	1	Q1
372	2002-01-06	Sunday	January	2002	6	2	1	Q1
373	2002-01-07	Monday	January	2002	7	2	1	Q1
374	2002-01-08	Tuesday	January	2002	8	2	1	Q1
375	2002-01-09	Wednesday	January	2002	9	2	1	Q1
376	2002-01-10	Thursday	January	2002	10	2	1	Q1
377	2002-01-11	Friday	January	2002	11	2	1	Q1
378	2002-01-12	Saturday	January	2002	12	2	1	Q1
379	2002-01-13	Sunday	January	2002	13	3	1	Q1

그림 3. 시간 테이블
Fig 3. Time Table

sales_date	Product_ID	Order_Qty	sales_price	sales_revenue	customer_id
2002-07-25	12	150	9000	4200	247
2002-08-02	12	150	9000	4200	7
2002-08-06	76	150	4500	1800	247
2002-08-06	80	200	16000	6000	108
2002-08-07	33	50	3000	1200	130
2002-08-09	76	150	4500	1800	132
2002-08-10	76	30	900	360	152
2002-08-13	32	250	20000	6250	99
2002-08-14	23	200	16000	10000	244
2002-08-15	50	150	12000	7500	281
2002-08-15	96	250	7500	2500	94
2002-08-16	68	70	700	70	215
2002-08-17	46	200	8000	1000	123
2002-08-17	51	100	15000	4000	186
2002-08-17	92	120	2400	720	31
2002-08-20	36	90	45000	16200	27
2002-08-21	33	50	3000	1200	281
2002-08-22	22	80	3200	800	280
2002-08-23	1	50	3500	1500	83

그림 4. 사실 테이블
Fig 4. Fact Table

이 테이블들을 이용하여 "어떠한 고객이 언제, 어떤 제품을 얼마나 사 갔는가?"에 대한 것을 LC 모델로 나타내어 보면 (그림 5)와 같다.



Customer.name, Time.time_id, Product.name

그림 5. 전형적인 고객의 판매 정보
Fig 5. Typical customer sales data

(그림 5)는 위치지정자로 고객 테이블의 name과 시간 테이블의 time_id, 그리고 제품 테이블의 name을 두고 콘텐츠로 사실 테이블의 sales_price를 둔 것을 나타내고 있다. (그림 5)에서는 "Jewel Creek라는 고객이 2002년 8월 23일 Even Better Low Fat Cottage Cheese를 3000달러어치 사들였다."라는 내용을 나타내고 있다. 그러나 CRM적인 분석기준에서는 고객, 시간, 제품의 세 가지 위치지정자와 하나의 사실 테이블로 분류하는 기준으로서는 분석자가 원하는 여러 평가 사항들을 표현하기가 힘들다. 예를 들면, Customer라는 위치지정자에서의 gender, married, job, yearly_income 등의 콘텐츠로서 표현되는 것들이 중요한 평가 기준이 된다고 분석자가 판단하였다면 분석자는 이러한 기준을 가지고 평가하기를 바랄 것이다. LC 모델을 이러한 관점에서 다음과 같이 다시 표현할 수 있다.

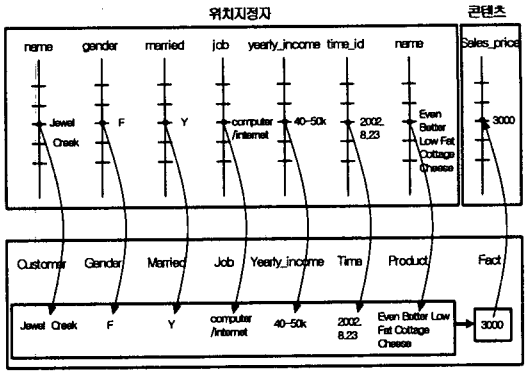


그림 6. 본 연구에서 제안하고 있는 고객의 판매 정보
Fig 6. Proposed customer sales data

(그림 6)은 위치지정자로 고객 테이블의 name, gender, married, job, yearly_income과 시간 테이블의 time_id와 제품 테이블의 name을 두고 콘텐츠로서 사실 테이블의 salse_price를 두고 있다.

(그림 6)은 "Jewel Creek라는 결혼하여 자녀가 하나있고 컴퓨터/인터넷 분야의 직업을 가진 여성 고객이 2002년 8월 23일 Even Better Low Fat Cottage Cheese를 3000달러어치 사들였다."라는 내용을 보여준다.

이상의 내용처럼 CRM적인 분석에서는 고객에 대한 성별이나 결혼여부, 직업, 연간 수입, 연령, 학력 등의 측정값으로써 단순하게 표현하였던 항목들이 활용가치가 뛰어난 분석 기준이 될 수 있다. 고객에 대한 가치 평가를 할 때 위와 같은 분석 기준들이 CRM적인 사고에서는 분석자에게 보다 정확한 고객에 대한 평가와 미래 지향적인 분석 기준들을 보여줄 수 있다. 따라서 이러한 LC모델의 특징과 같이 더 많은 분석에 사용 할 수 있는 측정값들을 하나의 위치지정자로서 독립시켜 CRM적인 측면에서의 의사결정시 분석을 효과적으로 하는 방법을 다루고자 한다.

3.2 위치지정자와 콘텐츠의 구별

3.1에서 위치지정자와 콘텐츠가 고정된 것이 아니라 분석자의 관점과 기준에 따라서 달라질 수 있다는 것을 보았다. 그러나 우리는 보고자 하는 컬럼에 대해서 이를 확고하게 하나의 위치지정자로 잡느냐에 대해서는 머뭇거리게 될 것이다. 즉, 위치지정자로 잡을 수 있는 어떤 기준을 생각해 볼 필요가 있다는 뜻으로 달리 말할 수 있다. 따라서 이 부분에서는 분석하고자 하는 컬럼에 대하여 독립된 위치지정자로 잡는 기준을 제시하고자 한다.

3.2.1 1:N 관계

(그림 2)의 제품 테이블을 보자. brand_name 컬럼과 product_name 컬럼의 멤버들 간의 카디널리티 관계(cardinality relationship)가 기준이 되는데, 이 테이블에서 하나의 brand_name에 여러 개의 product_name이 속해 있는 것을 볼 수가 있다. 예를 들면 Washington이라는 brand_name에 Washington Berry Juice, Washington Mango Drink, Washington Apple Drink의 세 개의 product_name이 속해 있다. 아래 (그림 7)은 예제를 위한 두 가지 product_name과 brand_name 컬럼의 일부를 추출한 것이다.

product_name	brand_name
Washington Berry Juice	Washington
Washington Mango Drink	Washington
Washington Apple Drink	Washington
Jeffers Oatmeal	Jeffers
Jeffers Corn Puffs	Jeffers
Blue Label Canned Beets	Blue Label
Blue Label Creamed Corn	Blue Label
Green Ribbon Canned Mixed Fruit	Green Ribbon
Green Ribbon Canned Peaches	Green Ribbon
King Rosy Sunglasses	King
Queen Eyeglass Screwdriver	Queen
Queen City Map	Queen
Club Low Fat Cottage Cheese	Club
Club Low Fat Sour Cream	Club
Club Blueberry Yogurt	Club
Red Spade Chicken Hot Dogs	Red Spade
Red Spade Pimento Loaf	Red Spade
Blue Medal Large Eggs	Blue Medal
Blue Medal Egg Substitute	Blue Medal
Golden Low Fat Waffles	Golden
Golden Orange Popsicles	Golden
Golden Frozen Corn	Golden
Faux Products Silky Smooth Hair Conditioner	Faux Products
Faux Products Childrens Cold Remedy	Faux Products
Denny Screw Driver	Denny
Ship Shape Extra Lean Hamburger	Ship Shape
Ship Shape Seasoned Hamburger	Ship Shape

그림 7. 제품 테이블의 name과 brand
Fig 7. Product.name and Product.brand

이러한 1:N의 관계가 깊으면 깊을수록 brand_name 컬럼과 product_name컬럼을 분리·독립시키지 않고, brand_name 컬럼을 product_name 컬럼의 하나의 위치지정자에 있어서의 상위수준(upper level)으로 정의하는 것이 분석적으로나 설계상으로 그리고 관계를 보기에 타당하다[1]. (그림 2)의 제품 테이블에서 예를 들것 같으면 product_name은 잎 멤버(leaf member), brand_name은 product_name의 상위수준 멤버, Product는 product_name와 brand_name를 멤버로 갖는 하나의 위치지정자이다.

3.2.2 M:N 관계

두 번째로 M:N의 관계를 생각해 보자. (그림 1)의 고객 테이블을 보자. 실질적으로 고객 평가를 할 경우 콘텐츠로서 구분되는 고객에 대한 성별이나 결혼여부, 직업, 연간 수입, 연령, 학력 등의 여러 가지 측정값들이 아주 중요한 분석 기준이 된다. 그래서 이러한 것들을 하나의 독립된 축인 위치지정자로 활용하여 사용하는 경우가 많다. 예를 들어 분석자가 고객 연령별 남녀 성비를 보고자 한다. 다수의 연령대별로 다수의 남녀가 있으므로 이 관계는 M:N의 관계라고 할 수 있다. 즉, 20대의 여성이 여러 명 있을 것이고 반대로 여러 명의 여성이 20대의 연령대에 존재할 것이다. 이러한 M:N의 관계가 깊으면 깊을수록 두 개의 위치지정자 Gender와 Age_Ties로 분리·독립시키는 것이 타당하다. 이것은 〈표 1〉과 같이 표현 할 수 있다.

표 1. 분리된 위치지정자 Gender와 Age_Ties
Table 1. Gender and Age_Ties as separate locator

		Gender	
		M	F
Age-Ties	20		
	30		
	40		
	50		
	60		

각각의 연령별 남녀의 수치와 성비를 손쉽게 알아볼 수 있을 것이고 타겟이 되는 고객에 대해서 보다 신경을 쓸 수 있을 것이다. 그러나 만약 이러한 분석 기준이 위치지정자 Customer에서 1:N의 관계로 표현 된다면 (그림 8)과 같이 표현 할 수 있을 것이다.

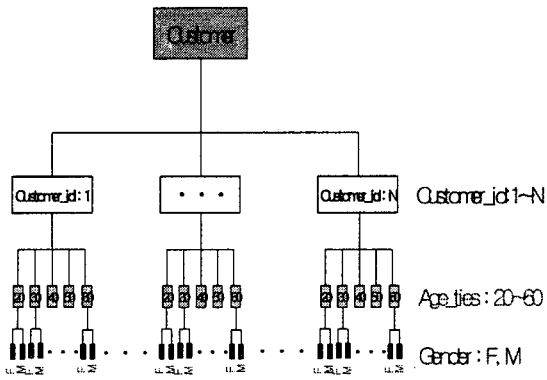


그림 8. 위치지정자 Customer에서의 연령별 남녀 성비의 1:N 관계
Fig 8. A 1-N relationship between Age_ties and Gender

(그림 8)과 같은 1:N의 관계였을 때 고객 연령대에 10대와 70대가 추가 된다면 이를 표현해 주기 위해 고객 수의 네 배만큼의 구별을 정의 해주어야 한다[1]. 그렇지만 〈표 1〉과 같은 M:N의 관계로 정의 한다면 우리가 다루고자하는 것들의 증가에 따른 영향을 받지 않게 될 것이고 이러한 추가 작업은 쉽게 할 수 있을 것이다.

이상을 종합해 볼 때 CRM적인 분석 관점에서는 콘텐츠로서 표현하였던 중요하게 활용할 수 있는 것들에 대하여 위치지정자로서 바꾸어 생각하는 것이 가장 효율적인 방법이라고 생각한다. 덧붙여 말하자면 CRM적인 분석뿐만 아니라 다른 분석 방법에서도 가장 중요한 결정 요소는 미래 시장에 있을 법한 변화 원인들을 알아내고, 변화 원인에 대한 의사 결정을 하는 것이다. 이러한 것들에 현재의 모델을 만드는 최적화 방법을 어떻게 구성하느냐에 관하여 말해질 수 있는데 이 부분에서, 우리들은 같은 논리적인 스키마를 표현하는 두 가지의 물리적인 방법에 대하여 논하였다. 다음 부분에서는 이러한 바탕을 기반으로 기존에 콘텐츠들로서 구분 되었던 측정값들을 하나의 독립된 위치지정자로 만드는 방법에 대하여 알아보겠다.

3.3 가상 위치지정자의 활용

가상 위치지정자란 물리적 위치지정자의 내용을 기반으로 하고 있다. 가상 위치지정자의 내용은 물리적인 위치지정자에 있는 기존 멤버 속성이나 물리적 위치지정자에 있는 테이블의 열이 될 수 있다. 일예로 멤버 속성은 가상 위치지정자의 생성 전에 정의하여야 한다.[5] 위치지정자를 구성하는 요소들을 멤버라고 하는데 이 멤버들의 특성을 설명해 주는 항목들은 멤버 속성이라고 한다.[1,5] 가상 위치지정자를 사용하는 중요한 이유는 LC 모델에서 이야기 하는 콘텐츠들에 가까운 성질을 가지는 멤버들을 일반적인 위치지정자처럼 지정하여 분석 하고자 하기 때문이다. 즉, 물리적인 위치지정자에서 수준으로서 정의해 주지 못하는 멤버들을 가상의 독립된 위치지정자로서 다루기 위하여 가상 위치지정자를 사용한다고 볼 수 있다. 예를 들면 (그림 9)에서의 위치지정자 Product의 멤버인 brand와 name은 서로의 관계를 계층으로 나타낼 수 있지만, 위치지정자 Customer의 멤버인 gender나 married, yearly_income들과 같은 콘텐츠에 가까운 성질을 가지는 멤버들은 수준을 지정해 줄 수가 없다. 가상 위치지정자를 사용하면 큐브에 있는 위치지정자의 멤버의 속성을 기준으로 큐브 데이터를 분석할 수 있으므로 사용자에게 있어서는 또 다른 분석 축이 생긴다는 장점을 갖게 된다.

3.1에서 기술한바 있는 고객, 제품, 시간, 사실 테이블을 이용하여 SALES 큐브를 만들어 보면 (그림 9)와 같다.

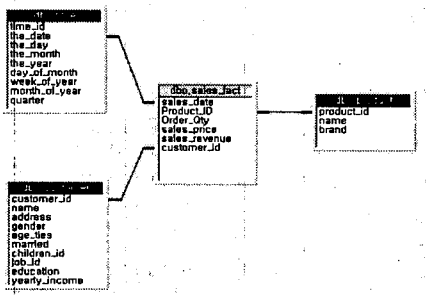


그림 9. SALES 큐브
Fig 9. SALES Cube

이 큐브는 Time, Product, Customer의 3개의 위치지정자로 구성되어 있으며, 또한 이 3개의 위치지정자의 키(customer_id, sales_date, product_id)와 콘텐츠(order_qty, sales_price, sales_revenue)로 구성되는 Sales_Fact라는 사실 테이블로 구성되어 있다. 이 SALES 큐브의 데이터를 살펴보면 (그림 10)과 같다.

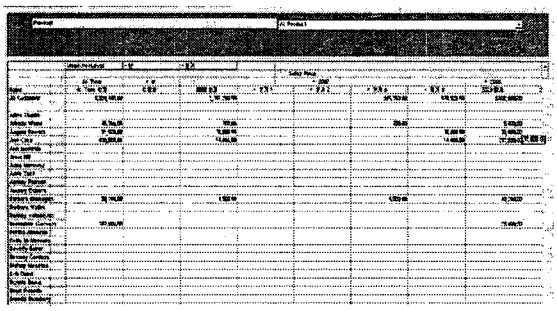


그림 10. 전형적인 큐브에서의 데이터의 표현
Fig 10. Typical cube data expression

(그림 10)에서는 개개인의 고객에 따른 시간별 매출 현황만을 볼 수 있다. 따라서 분석자가 고객의 성별과 연령대, 결혼여부, 교육 수준별, 직업별, 연간 수입에 따른 매출현황과 같은 사항을 알고자 할 경우, 이러한 사항들은 Customer라는 하나의 위치지정자에서 수준으로 잡아줄 수 없는 멤버들에 관련되기 때문에 수준별 분석을 위주로 하는 OLAP 도구는 역량을 발휘할 수 없게 된다. 따라서 전술한바와 같이 이러한 멤버들을 하나의 위치지정자로서 표현하기 위하여 가상 위치지정자를 이용하게 되었다. 가상 위치지정자를 이용한 분석 데이터를 살펴보면 (그림 11)과 같다.

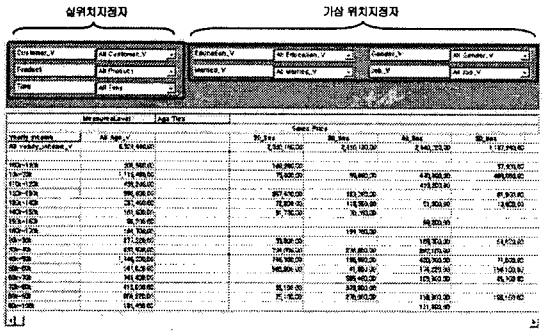


그림 11. 가상 위치지정자를 이용한 분석
Fig 11. Analysis using virtual locators

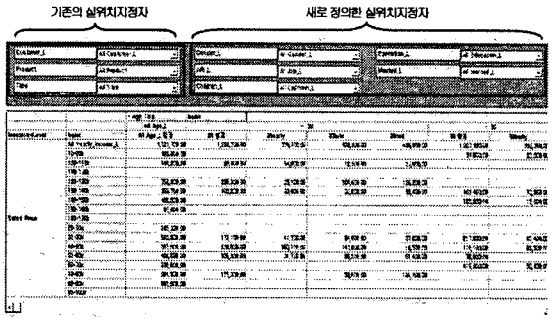


그림 12. 재구성한 실위치지정자를 이용한 분석
Fig 12. Analysis using physical locators

(그림 11)은 (그림 10)에 비해서 확연히 분석하고자 하는 기준을 분석자나 사용자들에게 보여줄 수 있도록 만들어 준다. 여기서는 연령대에 따른 연간소득별 매출현황에 관하여 나타내었으며, 이 큐브는 각각의 실제의 위치지정자와 각각의 가상의 위치지정자들을 이용하여 모든 조합에 따른 분석을 행할 수 있다. 그러므로 가상 위치지정자를 이용한 방법은 분석이 기준이 될 수 없었던 멤버들을 CRM적인 관점에서 잘 나타내고 있다고 할 수 있다. 가상 위치지정자는 확인한 바와 같이 모든 기준에 대한 분석 데이터들을 분석자에게 보여줄 수가 있다. 그러나 만약 분석자가 각 연령대를 초반·중반·후반의 계층으로 분할하여 분석하고자 할 경우 가상 위치지정자는 이러한 상황을 정의 해줄 수가 없다. 가상 위치지정자는 기존 위치지정자의 멤버속성들을 멤버로 멤버를 위치지정자로서 승격시켜 정의 한 것이기 때문에 계층으로서 표현하는 것은 불가능하다. 즉, 가상 위치지정자를 여러 수준으로 지정해 주는 것은 불가능하다. 결과적으로 가상 위치지정자를 사용하는 방법은 분석 기준의 한계가 명확하기 때문에 분석자는 한계 이상의 세밀한 분석을 정의할 수 없다.

따라서 이러한 가상 위치지정자의 한계점을 해결하기 위해 분석의 기준이 되는 측정값들을 모두 독립된 가상이 아닌 실위치지정자로 재구성하였다.(그림 12참조)

실제 독립된 위치지정자로서 큐브를 만들면 (그림 3-12)에서 보는 바와 같이 수준을 재정의 해줄 수 있다는 커다란 이점이 생긴다. 이 점으로 미루어 보았을 때 독립된 위치지정자로서 큐브를 만들게 되면 가상 위치지정자로서 큐브를 재정의 했을 때 보다 세밀하고 정교한 분석을 행할 수 있다는 것을 알 수 있었다.

IV. 사례 연구

본 연구에서 제시하는 실위치지정자를 사용하여 LC 모델에서 이야기 하는 콘텐츠들에 가까운 성질을 가지는 멤버들을 정의하는 방법을 채택하면 가상 위치지정자를 이용하여 콘텐츠와 같은 멤버들을 정의할 때 생기는 한계를 해결할 수 있다. Foodcakes International(FCI)에서 실제 사용하는 데이터를 바탕으로 본 연구에서 제시하고 있는 방법의 우월성을 증명하고자 한다.

CRM적인 기준에서 비즈니스 분석가는 기존 고객과의 원활한 관계와 잠재적인 고객의 창출에 관한 포인트를 잡기 위해 고객에 대한 세밀한 분석을 해야 하고, 그에 따라 일목요연하게 정리된 리포트를 확인해야 한다. FCI의 데이터를 이용하여 이미 3.3에서 SALES 큐브를 생성한바 있는데, 이 큐브는 3개의 위치지정자 Time, Product, Customer로 구성되어 있으며, 또한 이 3개의 위치지정자의 키(customer_id, sales_date, product_id)와 콘텐츠(order_qty, sales_price, sales_revenue)로 구성되는 Sales_Fact라는 사실 테이블로 구성되어 있다.

비즈니스 분석가가 2002년 3분기에서의 고등학교 졸업의 학력을 가진 결혼한 30대 여성의 직업과 연간수입에 따른 매출현황을 보고자 할 때에는 전형적인 큐브를 갖고서는 이를 나타내기가 불가능 하다는 것을 이미 3장에서 설명한 바 있다. 따라서 이 해결책으로 먼저 가상 위치지정자를 사용하는 방법을 들어보기로 한다. 가상의 위치지정자로 규정짓게 되는 것들은 위치지정자 Customer의 멤버들인 married, age_ties, gender, education, job_id, yearly_income이다. 가상 위치지정자로서 표현된 데이터를 살펴보면 (그림 13)과 같다.

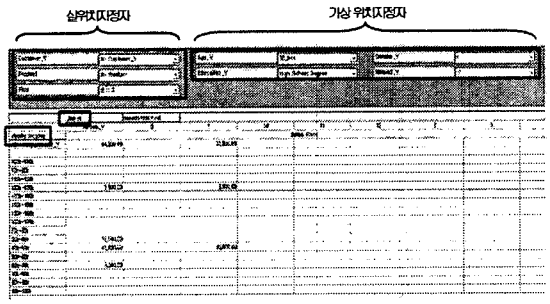


그림 13. 가상 위치치정자를 이용하여 분석한 데이터
Fig 13. Case analysis using virtual locators

(그림 13)에서 보는 바와 같이 분석 조건에 해당하는 여성들의 총 매출액은 $64,200.00이고, $40-50k의 연간 수입을 가진 고객들에게 총 매출액의 2/3가 넘는 $45,000.00의 매출을 올렸으며, 직업 아이디가 1인 즉, 컴퓨터/인터넷 종사자가 전체 분석 조건의 매출 중 절반 이상인 $33,500.00의 매출을 차지한다는 것을 알 수 있다. 즉, 2002년 3분기에서는 연간 수입이 $40-50k이며 컴퓨터/인터넷 종사자인 여성 고객이 가장 큰 매출을 차지하고 있다는 것을 종합적으로 알 수 있다. 이로써 분석자는 분석 기준에 따른 만족할 만한 결과 값을 얻을 수 있게 된다. 이러한 분석 결과를 얻은 후에 분석가가 좀더 세밀한 분석을 하기를 원한다면, 즉 일예로 분석자는 2002년 3분기에서의 인문계 고등학교 졸업의 학력을 가진 결혼한 30대 초반의 여성의 직업별 연간 수입에 따른 매출현황을 보고자 한다면 분석가의 분석 기준에 합당하는 결과 값을 얻기 위해서는 연령대와 학력을 세분화 시켜야 한다. 즉, 수준을 재정의 해야만 한다. 그러나 가상 위치지정자를 사용한 방법으로는 3.3에서 전술한 바와 같이 수준을 재정의 하는 것이 불가능하기 때문에 위치지정자 Customer의 멤버로서 표현 되었던 married, age_ties, gender, education, job_id, yearly_income를 모두 독립된 실위치지정자로 재정의 하여야 한다.

독립된 위치지정자로서 멤버들을 재정의 하기로 결정한 후 분석기준에 합당한 결과 값을 얻기 위해서 연령대를 초반·중반·후반으로 세밀화 하였고, 학력도 고등학교는 인문계, 공업 고등학교, 상업 고등학교로 대학은 2년제와 4년제로 대학원은 석사와 박사로서 세분화 하였다. 덧붙여서 결혼한 고객의 자녀수에 따른 분석기준이 생길 수 있다는 미래 지향적인 생각에 의해 children_id도 독립된 위치지정자로서 정의하여 추가하였다. 이렇게 구성한 새로운 SALES 큐브의 스키마를 살펴보면 (그림 14)와 같다.

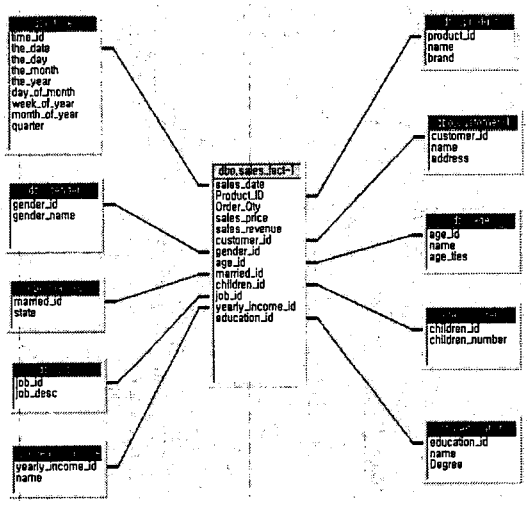


그림 14. 각각의 콘텐츠들을 독립된 실위치지정자로 정의한 SALES 큐브
Fig 14. Case SALES cube using physical locators

최초 Time, Product, Customer의 3개의 위치지정자와 Sales_Fact라는 사실 테이블로 구성되어 있었던 것에 비하여 (그림 14)에서는 Sales_Fact 테이블을 중심으로 Time, Product, Customer, Married, Age, Gender, Education, Jobs, Yearly_income, Children의 10개의 위치지정자가 키로서 연결되어 있는 것을 확인 할 수 있다. 실위치지정자로서 표현된 데이터를 살펴보면 (그림 15)과 같다.

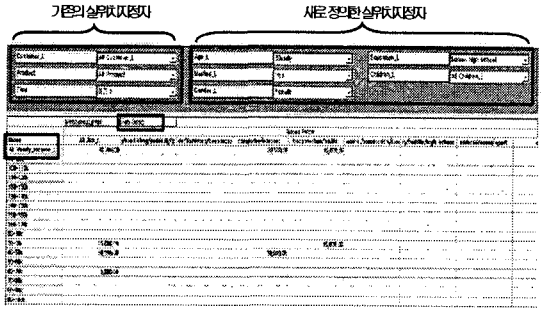


그림 15. 독립된 실위치지정자로서 분석한 데이터
Fig 15. Analysis with 5 newly defined physical locators

(그림 15)에서 보는 바와 같이 2002년 3분기에서의 인문계 고등학교 졸업의 학력을 가진 결혼한 30대 초반의 여성들의 총 매출액은 $48,700.00이였고 이중 연간 수입이 $40-50k이고 컴퓨터/인터넷 직업 종사자의 고객이 $30,500.00원으로 가장 높은 비율의 매출을 올리게 한 것을 확인 할 수 있다.

이처럼 분석 환경이 분석자의 기준에 따라 추가적으로 세분화되었을 때에 가상 위치지정자를 사용하여 분석하면 합당한 값을 얻을 수 없다. 이에 비해 실위치지정자를 사용하였을 때에는 분석 기준에 합당한 값을 얻을 수 있을 뿐만 아니라 추가적인 분석 기준을 확장시킬 수도 있다. 결과적으로 수준을 정의해줄 수 없는 멤버들을 분석하기 위해서는 가상 위치지정자로 정의해 주는 것보다는 실제의 독립된 실위치지정자로서 정의해 주는 것이 CRM적인 관점에서는 보다 효율적이라는 것을 확인할 수 있었다.

V. 결론

데이터웨어하우스의 방대한 데이터를 사용자의 분석 관점에 따라서 다양하게 비교·분석함으로써 경영을 지원하는 OLAP 기술은 CRM에서의 핵심부분인 개인화를 달성하기에는 부족하다고 하겠다. 그 이유는 분석항목의 수준을 분석자의 관점에 맞게 세분화할 수 없다는데 있다. CRM에서는 OLAP에서 콘텐츠로서 표현되는 고객 각각의 성별, 결혼유무, 직업, 연간수입, 연령, 학력, 취미 등의 항목에 대해 중요하게 다룬다. 그러나 OLAP은 분석수준의 유연성의 결여로 인해 CRM에서 중요시 여기는 이러한 부분들에 대한 세밀한 분석을 할 수가 없다. 따라서 OLAP에서 CRM의 핵심인 개인화를 달성하기 위해서는 콘텐츠로서 표현되는 항목에 대하여 유연하게 다룰 수 있어야 한다. 본 연구에서는 CRM적인 분석 관점에서 중요하게 활용 가능한 항목들에 관하여 OLAP에서 지원하는 가상 위치지정자가 아닌 실제의 위치지정자로서 다루는 방법을 제시하고 실제 예제를 통하여 실위치지정자의 우월성을 확인하였다.

본 연구에서 제시하는 방법을 사용하면 기존 OLAP의 한계였던 유연성을 확보 할 수 있으므로 고객 각각의 패턴을 찾고 고객의 행동에 관하여 예상할 수 있는 데이터마이닝 분야까지도 다룰 수 있게 된다. 다시 말하자면 OLAP이 데이터마이닝의 부분적인 역할까지도 수행 할 수 있게 됨으로써 OLAP상에서 시스템적으로 효율적인 CRM을 다룰 수 있게 된다는 것이다. 이러한 점에서 본 연구에서 제시하는 방법은 기업의 CRM 구축에 필요한 시간과 비용을 줄일 수 있을 것으로 생각하며, CRM에 있어서 OLAP의 적용분야를 더욱 넓힐 수 있을 것으로 예상한다.

실 위치지정자를 사용함으로써 분석기준을 높이게 되면 큐브(cube)상의 희박성(sparsity) 문제가 생기게 되는 점에 대하여서는 추후 연구가 필요하리라 생각된다.

참고문헌

[1] Erik Thomsen, "OLAP Solutions", 2nd Edition, John Willy and Sons, 2002.
[2] Berson, Smith, Thearling, Building Data Mining Applications for CRM, McGraw Hill, 2000.
[3] Inman, W.H., "Building the Data Warhouse," 2nd Edition, John Willy and Sons, 1996.
[4] Oracle, "Data Warehousing Concept Paper," Oracle Korea, 1997.
[5] 조재희, 박성진, "OLAP 테크놀로지", Sigma Insight, 2003.
[6] David M. Kroenke, "Database processing", Ninth Edition, Pearson Education International, 2003.
[7] Turban, Raiiner, Potter, "Introduction To Information Technology", John Willy and Sons, 2001.

[8] N.-S. Koutsoukis et al., "IAdopting on-line analytical processing for decision modeling:the interaction of information and decision technologies", Elsevier science B.V., Decision Support Systems, 26권, 1호, 1쪽 1999.
[9] 이덕근, 오미화, 조재훈, 최인수, "DW에서의 질의어 처리 성능향상을 위한 데이터 구조화 방법", 한국컴퓨터정보학회 제10권, 33호, pp. 7-13, 2005.
[10] 최정우, 최인수, "웹 상에서의 비정형 시스템 의 자율적 설계", 한국컴퓨터정보학회 제8권, 2호, pp. 64-74, 2003.

저 자 소 개

이 덕 근

1985년 경기대학교 경영학과 졸업(학사)
1988년 숭실대학교 중소기업대학원 졸업(공학석사)
2004년 숭실대학교 산업·정보시스템공학과 수료(공학박사)
2001년~현재 부품소재통합연구단 소장
〈관심분야〉 MIS, 정보시스템 분석·설계, DW, OLAP, CRM

유 한 주

2001년 남서울대학교 산업공학과 졸업(학사)
2004년 숭실대학교 MIS 졸업(공학석사)
2005년~현재 숭실대학교 산업·정보시스템공학과 재학(공학박사)
〈관심분야〉 MIS, 정보시스템 분석·설계, DW, OLAP, CRM

최 인 수

1985년 서울대학교 산업공학 졸업(공학박사)
1975년 영국 College of Librarianship, Wales (Diploma) 객원교수
1990년 日本 東洋大學 經營學部 객원교수
1980년~현재 숭실대학교 산업·정보시스템공학과 교수
〈관심분야〉 객체지향데이터베이스, 정보시스템 분석·설계, ERP, MIS, DW, OLAP, CRM