
CRM 데이터 웨어 하우스 구축 모형에 관한 연구

정진택*

목 차

- I. 서론
- II. CRM분석의 활용
- III. Implications of CRM requirements on DataWarehousing
- IV. Challenges For CRM Technologies
- V. Conclusions and Future Areas For Research

참고문헌

Key Words: CRM, 데이터 마이닝

Abstract

It is far more expensive for companies to acquire new customers than it is to retain customers. As a result, companies are turning to Customer Relationship Management (CRM) in order to make decisions about managing the relationship and the profitability of those customer relationships. CRM is a strategy that integrates the concepts of Knowledge Management, Data Mining and Data Warehousing in order to support the organization's decision-making process to retain long-term and profitable relationships with its customers. This paper examines the design implications that CRM poses to data warehousing. We then present a robust data warehouse schema to support CRM analyses and decisions. For example, the proposed schema could be used to calculate customer profitability and to identify social networks of influence between customers. The paper also discusses future areas for research pertaining to CRM data warehousing and data mining.

* 한성대 정경학부 교수 -jungit@hansun.ac.kr

I. 서론

한 명의 신규고객을 확보하는데 드는 비용을 기존 고객에게 투입하면 5명을 유지할 수 있다. 기존의 고객들에게 새로운 제품을 판매하는 것이 새로운 고객보다 2배 이상 수익성이 높다 (Winer et al 2001). 따라서 최적 마케팅의 목적은 모든 고객이 동일하지 않다는 전제에서 출발하여 고객에는 특정 고객은 타 고객보다 수익성이 높으며 일부 고객은 전혀 수익을 내지 못하며 어떤 고객은 현재 뿐 아니라 미래에도 결코 기업의 수익에 도움을 주지 않는다. 따라서 현재 또는 장차 수익성이 있거나 장차 수익성 있는 고객에게 투자를 확대하고 비수익성 고객에게는 투자를 제한하는 것이 보다 효과적일 것이다. 고객이 원하는 것과 기업과 어떻게 상호작용하기 원하는지를 진정으로 이해할 때 고객 충성도 및 이와 연관된 수익성 개선을 이룰 가능성이 높아지고 기업은 고객을 지속적으로 유지할 수 있는 가장 효과적인 방법을 고안해야 할 필요성이 대두되었다. 개별화된 고객과의 관계수립 및 관리를 위해 기업들은 고객관계관리 (Customer Relation Management, CRM) 기법과 관련 기술에 주목하기 시작했다.

고객관계관리는 관계형 마케팅, 또는 고객관리라 불리어지며 영업 자동화 (Sales Force Automation)에서 출발하였다. (Zohoff 2001, Spangler 1999) 90년대 각광을 받았던 영업직 자동화는 (SFA) 영업직 사원을 위해서는 어느 정도 도움을 주었으나 접촉관리 수준에 머물러 기업의 관계형 마케팅에 기여하기 보다는 낡은 방식의 업무처리 속도를 증가시켰을 뿐이다. 이와는 달리 CRM은 영업, 마케팅과 정보시스템 등 고객관련

부서를 융합시켜 원활한 고객서비스 지원이 가능하도록 정보를 교류하는 역할은 하고 있다. 고객관계관리는 신규고객 획득, 기존 고객 유지 및 고객수익성을 증대하기 위하여 지속적인 커뮤니케이션을 통해 고객 욕구를 이해하고 지속적인 고객과의 관계를 통하여 장기적으로 이윤을 극대화 전략으로 정의된다. 또한 기업 내 조직 구조와 조직 문화 속에서 자의적 일반 원칙이 아닌 실질적인 고객선호도에 기반 한 적극적이고 수익 창출적인 중장기적 고객관계를 지원하기 위해 기업 내 지식과 기술을 획득 및 활용하는 전략으로 정의되기도 한다. CRM은 지식관리 또는 기업지식 및 기술을 이용하여 기업이 무엇보다도 제품선정, 마케팅전략, 고객과의 관리에 대한 의사 결정을 가능하게 해주는 수단이다. 그러한 의사결정 지원시스템의 핵심에는 데이터웨어하우스, 데이터마트, 데이터마이닝기법 등이 있다. 데이터웨어하우스와 데이터마트는 고객행위에 대한 데이터 수집에 이용되고, 데이터마이닝은 고객의 행위 뿐만 아니라 전문 지식을 기반한 모델을 이용하여 고객에 관련된 중요한 특성을 파악하기 위해 이용된다. 프로세스를 성공적으로 수행하기 위해서는 먼저 높은 이익 잠재력을 가진 고객들을 포함하고 있는 시장군을 정의해야 한다. 다음으로는 이러한 고객들에 긍정적으로 영향을 미치는 캠페인을 수립하고 집행한다. 우선 시장군을 알기 위해서는 잠재고객 및 그들의 구매행태에 대한 많은 데이터가 필요하다.

본 논문은 고객관계관리에 관련된 기술적 디자인 이슈를 다룬다. 본 논문의 논의 전개는 CRM 분석의 활용분야에 대해 살펴보고 CRM 기술의 요구

사항과 테크닉분야 및 CRM 기저에 깔려있는 기술적 요구사항에 대해 살펴본다. CRM 기술도전에서는 CRM에 관련된 데이터웨어하우스, 데이터마트,

데이터마이닝 기법이 직면한 도전에 대해 논의한 후 향후 연구방향에 대해 언급한다.

II. CRM 분석의 활용

데이터웨어하우스, 데이터마트를 활용하여 기업들은 고객특화 전략수립을 위한 의사결정을 하기 위해 고객프로파일, 고객세분화, 교차분석을 수행할 수 있다. 예컨대, 기업은 통합된 데이터마트를 활용하여 고객의 과거 및 미래가치를 결정하여 고객을 다음과 같은 사분면 중에 한 분면으로 나눌 수 있다.

- 고객관계를 단절해야 할 경우 : 고객으로 인해 창출되는 수입보다 소요되는 투입 비용이 훨씬 상회하는 경우
- 고객관계를 재정립 되어야 할 경우 : 잠재적인 고객가치는 있으나 이를 위해서 기업의 관리가 필요한 경우
- 고객관계에 착수해야 할 경우
- 고객관계에 적극적으로 투자해야 할 경우 (Verhoff 2002)

기업은 사분면에 상응하는 고객전략을 사용하여 고객관계를 관리할 수 있다. 표1, 2에서는 데이터마트로 수행할 수 있는 고객세분화 타입을 예시하고 있다. 특히 1/4분면에 속하는 고객 세분화시 주의가 요망된다. 1/4분면에 속하는 고객은 추가적으로 1) 좀 더 가치 있는 고객을 위한 벤치마킹 사례로 사용되는 경우 2) 제품 품질개선과 작업과정 효율성 증진을 위한 아이디어 제공으로 사용되는 경우 3) 기업에 전혀 도움이 되지 않는 경우

〈표1〉 고객 세분화

		과거가치	
		낮다	높다
미래가치	높다	재정립(리엔지니어링)	적극적 투자
	낮다	관계단절	관계착수

〈표2〉 고객 세분화 대응 전략

		과거가치	
		낮다	높다
미래가치	높다	상승판매활동 및 가치추가	우선고객, 특혜고객조치
	낮다	비용절감, 가격상승	고객 충성도 유지를 위한 신규 기회 포착을 위한 관계착수

통합데이터를 사용하는 것이 반드시 고객행태의 결정요인을 확정하는 것은 아니다. 그러나 구체적 데이터를 분석해야 할 수 있다.(Verhoff 2002) 게다가 고객세분화는 현재 및 확장된 가구(HOUSEHOLD)개념에 의해 보다 확장될 수 있다. 고객의 확장된 가구를 확인할 수 있는 하우스홀드분석을 통해 가구 또는 관계양식(고객간 잠재적 영향도 관계)에 따라 개인을 세분화할 수 있다. 가구분석을 통해 얻을 수 있는 혜택은 아래와 같다.

- 교차판매 및 상승판매 기회의 확인
- 중요한 라이프이벤트의 확인
- 잠재적 영향도 관계의 확인

전통적으로 대다수의 기업은 전통적인 가구로 분류되는 고객데이터만을 분석해왔다. 그러나 실질적으로 세가지 형태의 가구가 존재한다.

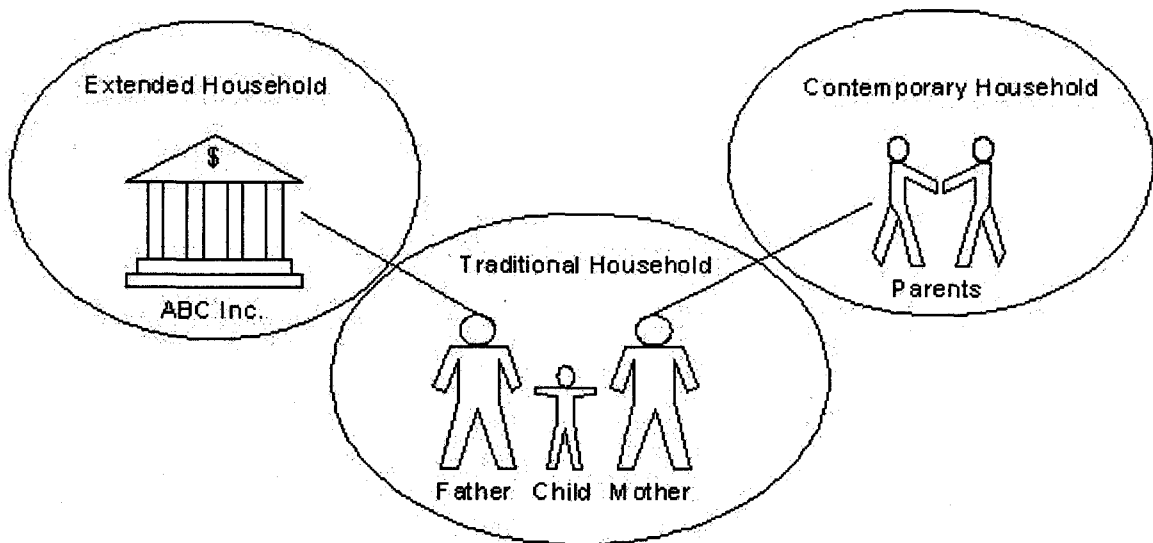
- 전통적인 가구
- 현재 가구
- 확장된 가구

전통적 가구란 가족구성원이 동일한 주소에 거주하는 경우를 말한다. 현재 가구란 상이한 주소에 거주하고 있는 가족구성원 고객간의 관계를 의미한다. 반면에 확장된 가구란 전통적 가구의 구성원과 기업간의 연계를 나타내기 위해 사용된다. 표1은 고객중심기업구축에 사용된 세가지 형태의 가구를 보여주고 있다.

그림1에 묘사된 모델의 의미는 기업들은 한 분야에 속한 멤버에 관련된 의사결정이 다른 분야에 속하는 관련 고객에게 부정적인 영향을 끼치지 않도록 잠재 영향도 관계를 관리하고자 한다. 예컨대 한 고객이 비수익성 분야에 속해있다면 기업은 그 고객에 대한 가격을 상승하려고 할 것이다. 그러나 만일 기업이 동일한 비수익성 고객이 좀 더 수익성 있는 분야에 속해있는 고객(부모, 중소기업)영향을 미친다는 사실을 인지하고 있다면 기업은 양쪽 고객 모두를 놓칠 수도 있는 위험을 감수하면서까지 고객에 대한 가격을 상승시키지는 않을 것이다.

가구분석은 분명 가치있는 분석이기는 하나 데이터웨어하우징에서 도전적인 문제로 간주되고 있

〈그림 1〉 Non-Traditional Household



다. 분명 실질적인 CRM전략을 성공적으로 수행하기 위해서는 기업들은 일관적이고 통일된 고객에 대한 기업차원의 시각을 가지고 있어야 한다. 이는 서로 다양한 시스템으로부터 제공되는 데이터가 통합되어야 한다는 것을 의미한다. 이러한 통합

된 품질높은 데이터소스에 대한 요구는 CRM에 핵심기술로써 데이터웨어하우징과 관련된 데이터마트에 대한 기술을 요구한다. 다음 절에서는 CRM을 지원하기 위해 필요한 데이터웨어하우징과 관련된 데이터마트기술에 대해 서술하겠다.

III. Implications of CRM Requirements on Data Warehousing

CRM은 고객인 기업과 거래하는 때 순간마다 생성되는 데이터 수집에 의존하기 때문에 CRM을 지원하는 데이터웨어하우스 디자인 요구사항에 대해 신중한 고려가 요구된다. 데이터웨어하우스 스키마에서 핵심적인 고객 DIMENSION은 확고한 고객분석과 탐색을 수행하기 위해 가장 극소화된 형태로 저장되어야 한다. 예컨대, 전화번호 전체를 동일 필드에 저장하기보다 전화번호를 구분하여 지역코드, 전화번호, 내선번호 등의 국지화된 요소를 갖는 각기 다른 필드에 저장해야 한다. 의사결정권자는 지역코드에 따라 통신판매에서 어떤 고객을 텔레마케팅에 포함시켜야 할지 결정할 수 있다. 동일한 주장을 우편번호 필드를 주요 지역코드와 부수적인 지역코드로 구분할 때 적용된다. 추가적으로 데이터웨어하우스는 고객들을 확장된 가구(익스텐디드 하우스홀드)로 세분화시킬 수 있어야 한다. 가구분석에 가치를 상이한 고객 세그먼트에 속한 고객간에 존재하는 영향도 관계를 결정할 수 있다는데 있다. 한 세그먼트가 다른 세그먼트보다 좀 더 수익을 창출할 수 있다 라는 의사결정을 할 수 있다. 이러한 사회적네트워크는 기업들이 고객을 유지하는 능력에 미치는 영향이 지대하기 때문에 고객획득과 관리에 중요하다.

위에서 언급하였듯이 신규고객확보에 소요되는 비용이 현재 고객을 유지하는데 소요되는 비용에 약 5배 이상 소요된다. 단골고객은 신규고객에 비해 두 배 이상 매출을 발생시킬 수 있다. 그러므로 고객 유지도를 측정할 수 있는 능력과 고객 이탈의 근본원인을 파악할 수 있다면 우리는 고객 유지를 개선할 수 있는 핵심적인 요인이 된다. 이는 데이터웨어하우스 디자인이 기업들에게 이러한 핵심요인을 모니터링하고 분석할 수 있는 능력을 제공해야 한다. 추가적으로 고객프로파일에 기반하여 고객을 한 개 이상의 고객세그먼트에 속한 구성원으로 분류될 수 있다. 그리고 시간의 흐름에 따라 고객멤버십이 바뀔 수도 있다.(멤버십 이동성) 게다가 CRM분석은 고객구매행위를 예측할 수 있으나 예측은 일반적으로 고객이 특정구매행위를 보여줄 수 있는 가능성을 적시하는 계량적인 가치를 지닌 숫자로 표시된다. 예컨대 높은 고객 이탈점수는 고객이 기업에서 이탈될 가능성이 높다는 뜻이고 낮은 이탈점수는 고객이 충성도 있는 고객으로 계속 남아있을 것이라는 것을 의미한다. CRM분석에서는 이러한 데이터특성을 데이터웨어하우스 스키마가 고객분류를 효율적으로 갱신하고 모니터링할

수 있는 스키마를 가지고 있어야 한다는 것을 의미한다. 달리 말해 고객이 속한 고객 세그먼트는 시간이 흐름에 따라 변화하고 이런 고객의 동태적 이동성은 KIMBELL의 최소 DIMENSION기법이나 아웃트리거 기법을 이용하여 모델링할 수 있다. 대다수의 기업들은 고객분류의 변화과정의 흐름을 측정하기를 원하기 때문에 최고로 현실적인 접근 방식은 DIMENSION모델을 유지하는데 3, 2, 1타입의 조합을 이용하는 것이다. 이러한 접근방식을 채택함으로써 기업들은 고객분류에 완벽한 흐름과 현재 고객분류에 대하여 알 수 있게 된다.

아웃트리거(브릿지데이터)접근방식의 단점은 정확한 결과치를 얻기 위해 사용해야 하는 검색문의 복잡성을 증가시키는데 있다. 이는 잘못된 검색문이 시행되어 부정확한 결과치를 고객에게 제공할 가능성이 증가한다는 것을 의미한다. 그러므로 신속하고 간편한 분석용 레코드선별을 위해서는 다

양한 고객점수가 데이터테이블에 아웃트리거방식이 아닌 최소 DIMENSION방식으로 직접적으로 연계되어 있어야 한다.

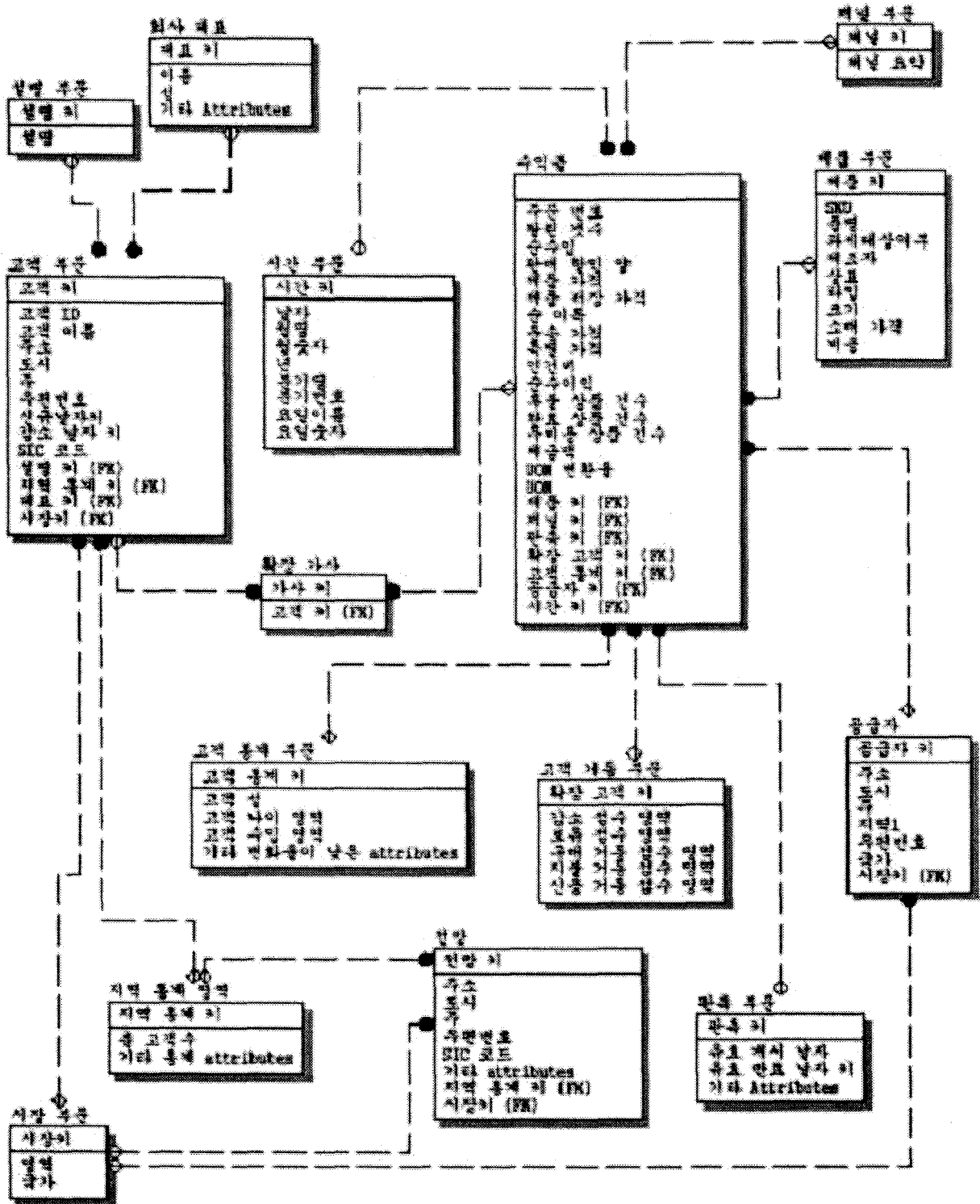
추가적으로 데이터는 무선, 유선, 전화, 이미지, 개인적 방문 등의 복수의 채널에서 수집되고 있다. 이러한 상이한 채널은 나름대로의 채널특성을 가지고 있어 고객소비행태행위를 분석하는데 적절한 의사결정을 위해 고려할 사항이다. 그러므로 데이터웨어하우스는 기업과 거래시의 선호도와 데이터테이블속에 저장되어 있는 구매거래를 위한 채널 선호도 등을 파악할 수 있어야 한다.

CRM분석에 사용되는 대다수의 데이터는 기업 내부에서 생성되지만 어떤 데이터는 기업 외부에서 내부에서 생성되는 속도와는 다른 속도로 생성되어진다. 따라서 데이터웨어하우스에서 추가적인 요구사항으로 데이터웨어하우스 디자인은 기업의

〈표3〉 CRM 데이터웨어하우스를 위한 디자인 최소사항

1.	고객유지율모니터링기능
2.	고객 이탈의 근본원인 파악 기능
3.	고객 점수화 기능
4.	고객을 복수의 확장구개정과의 연계 기능
5.	고객을 복수의 고객세그먼트에 세분화시키는 기능
6.	고객 세그먼트의 변화과정과 고객 점수 유지 기능
7.	고객 캠페인의 시간대에 따른 고객반응 평가 기능
8.	과거 구매행위에 기반한 미래 구매행위 예측 기능
9.	상이한 고객 관계와 집단간 고객 충성도 패턴 이해 기능
10.	인구 통계학적 분석 기능
11.	추세 분석 기능
12.	고객 수익성 분석 기능
13.	제품 수익성 분석 기능
14.	데이터의 외부소스를 포함한 복수의 소스데이터를 통합하는 기능
15.	효율적인 데이터 갱신과 유지 보수 기능

〈그림 2〉 A Starter Model for CRM



부소스데이터를 상이한 갱신속도로 데이터웨어하우스와 데이터마트에 통합할 수 있어야 한다. 이상의 선행연구결과에 따라 표 3은 CRM분석을 지원하기 위해 데이터웨어하우스에 필요한 최소 요구사항에 대하여 요약하고 있다. 그림 3은 표 3에 제시된 CRM요구사항을 지원하기 위한 데이터웨어하우스 초기 디자인 모델을 제시하고 있다.

표 2에서 언급된 모델은 고객수익성분석, 가구수익성분석, 인구통계적수익성분석, 제품수익성분석, 채널수익성분석 및 PROMOTION수익성분석 등 다양한 종류의 수익성 분석에 이용될 수 있다. 예를 들어 데이터테이블에서 어떤 거래에 대한 수익성은 아래와 같이 계산될 수 있다.

- (1) 총이윤=총수입-제조비용-마케팅비용-제품저장비용
- (2) 순이윤=총이윤-화물비용-특수비용-OVERHEAD비용
- (3) 총마진=총이윤/총수입

총마진에 대한 계산을 제외한 데이터테이블에 저장되어 있는 모든 계산은 정확한 계산에 대한 접근이 가능하다. 총이윤과 총수입은 데이터테이블에 저장되어 있지만 총마진에 대한 결과는 계산이 명시적으로 이루어지지 않기 때문에 명시적으로 저장되어 있지 않다. 총마진은 합계에 대한 비율로 계산할 수 있다. 영업할인액수와 조세관련계산은 고객들이 이러한 계산을 포함하거나 제외할 수 있는 융통성을 제공할 수 있도록 명시적으로 이루어진다. 수익성분석은 고객, 제품, 채널, 프로모션 또는 다른 디멘전에 의하여 집산화할 수 있다. 수익성분석결과 고객은 네 개의 4/4분면 가운데 한 개의 4/4분면에 세분화되어 세분화시키는데 사용될 수 있다. 이에 따라 고객관계관리를 위한 대책을 마련할 수 있다. 추가적으로 표 3에서 언급된 모델은 정시도착 아이템 숫자 불량률 제로 제품의 숫자 등 경영성과지표를 계산하기 위해 사용될 수 있다. 또한 총 물량에서 명시적으로 저장된 경영성과지표 숫자를 차감함으로써 보완적인 계산을 할 수 있다. 이러한 경영성과지표는 고객관계를 모니터링

〈표 4〉 데이터웨어하우스 모델 DIMENSION 정의

채널 DIMENSION	고객이 고객과 거래할 수 있는 다양한 수단을 저장한다.(인터넷, 상점 등)
고객 DIMENSION	고객관련 정태적 정보를 저장
고객행위 DIMENSION	고객의 동태적 점수화필드를 저장
고객인구통계학적 DIMENSION	고객의 동태적 인구통계학적 특성을 저장
COMMENT DIMENSION	고객이탈이유와 다른 DIMENSION에 필요한 COMMENT를 저장
기업 대표 DIMENSION	고객관리계정에 책임있는 기업실무자(영업실무자) 정보
카운티 인구통계학적 DIMENSION	고객이 유치하고 있는 카운티의 외부 인구통계학적 특성
확장된 가구 DIMENSION	고객이 한 개 이상의 확장된 가구에 속할 데이터를 제시
시장 DIMENSION	고객과 공급자가 속한 지역과 국가를 나타냄
제품 DIMENSION	기업이 판매하는 제품을 나타냄
PROMOTION DIMENSION	기업이 제공하는 PROMOTION을 나타냄
프로스펙트 DIMENSION	외부의 선행리스트에서 제공되는 예측에 관련된 정보를 저장
공급자 DIMENSION	제품을 공급하는 업체에 관련된 정보를 제공
TIME DIMENSION	스키마전반에 걸쳐 사용되는 일반적인 날짜정보를 나타냄

하고 관리하는데 중요하고 또한 고객만족도와 장차 고객유지에 영향을 미칠 수 있다.

시간 디멘전에 있는 날짜 관련 외부키 (FOREIGN KEY)를 사용하여 다른 디멘전에 연결될 수 있다는 점을 주목한다. 동태적 고객디멘전 (고객인구통계학적 디멘전과 고객행위디멘전)은 정보가 시간에 따라 변화하기 때문에 두 개의 DIMENSION으로 구분되어 졌다. 추가적으로 이러한 DIMENSION들은 고객들이 손쉽게 데이터 테이블을 탐색하고 데이터테이블에 있는 확장된 고객키에 기반하여 고객세그먼트를 구축할 수 있도록 아웃트리거기법이 아닌 최소디멘전으로 구축되어 있다. 이러한 고객세그먼트는 수익성있는 고객의 구매성향과 인구통계학적 고객을 프로파일하기 위해 사용되어 질 수 있다. 이러한 정보는 유사한 프로파일과 함께 잠재고객에 대한 캠페인을 사용되어 질 수 있다. 이러한 접근 방식의 또 다른 장점은 고객의 구매행위 점수와 인구통계학에 있어서의 변화과정은 데이터테이블의 일부로 저장되고 고객인구통계학 또는 고객행위DIMENSION에 대한 1, 2, 3타입기법을 사용하지 않고 확고한 분석을 제공할 수 있다. 이러한 접근방식의 결과 비교적 느린 변화를 보여주는 DIMENSION은 카운터 인구통계학적DIMENSION, 제품DIMENSION과 공급자DIMENSION이다. 이러한 세 개의 비교적 느린 DIMENSION을 효율적으로 갱신하는 방법은 기록물을 단순히 레코드를 재작성하는 일본타입 방식이다. CRM분석을 위하여 가장 필요한 제품DIMENSION필드는 데이터테이블에서 저장되어 있는 비용정보이다. 표 2에서 수익성 데이터테

이블은 축적된 단지 단면을 보여줄 뿐이다. 왜냐하면 수익성데이터는 분석과정에 서로 다른 분석과정에서 획득되어지기 때문에 수익성데이터테이블은 축적된 단상에 불과하다. 표 2에 제시된 데이터 웨어하우스모델은 기업들로 하여금 아래와 같은 기능을 수행할 수 있도록 도와준다.

- (1) 동일산업계에서 선호하는 고객세그먼트에 속한 고객들의 소비성향과 동일한 소비성향을 가질 가능성
- (2) 다른 사업계에서 기업이 많은 제품을 판매 또는 판매할 제품과 동일한 타입의 제품을 사용하고 있는 가능성
- (3) 동일 산업계에서 현재 좋은 고객이 긴 하지만 동일한 타입의 제품을 실제로 구매하지 않고있는 고객, 이는 고객이 장차 어떤 다른 제품을 구매할지도 모르는 제품을 결정하는데 사용한다.
- (4) 제품을 구매 또는 구매할 가능성이 있는 지정학적 위치에 따른 가능성.
- (5) 어떤 제품을 함께 판매할것인가를 결정, 이는 산업과 지정학적 위치에 따라 프로모션을 정의하기 위해 사용되어 질 수 있다.
- (6) 고객수익성
- (7) 고객점수화
- (8) 고객이 이탈하는 이유를 결정하기 위한 근본 원인분석

이상에서 분석한 결과는 기업들로 하여금 제품 출시, 마케팅전략 및 고객과의 거래에 대한 의사결정을 내리는데 도움을 줄 수 있다.

IV. Challenges For CRM Technologies

CRM기술이 직면하고 있는 몇가지 문제와 데이터웨어하우스, 데이터마트 및 데이터마이닝이 직면하고 있는 몇 가지 문제에 대하여 논의하고자 한다.

1. 데이터마이닝과 고객분류

CRM기술이 직면하고 있는 한가지 흥미로운 과제는 발견지향 데이터마이닝기법이다. 발견지향 데이터마이닝기법은 데이터간의 관계특성에 관한 어떤 선입관없이 데이터속에서 새로운 관계나 중요한 규칙을 발견하는 것이다. SPINGER 등은 데이터마이닝 알고리즘에 대한 세 가지 분류로 나누고 있다.

- (1) 수학기반기법
- (2) 거리기반기법
- (3) 논리기반기법

수학기반기법에서 가장 보편적으로 사용되고 있는 데이터마이닝기법은 인공신경망과 선형판별분석이다. 또한 논리기반 데이터마이닝기법에서 가장 많이 쓰이는 기법은 의사결정나무와 규칙추출 기법이다. 의사결정나무와 규칙추출 데이터마이닝 기법의 장점은 이들은 전문가 지식을 명시적으로 모델링하고 추론과정이 분류과정에 포함되어 있고 따라서 고객들에 의하여 손쉽게 이해되고 해석되어 진다. 의사결정나무와 규칙추출 데이터마이닝 테크닉과는 달리 인공신경망 데이터마이닝기법에서 모델의 추론과정을 해석하기는 쉽지 않다. 고객을 0.5이상의 기준값에 기반하여 고객 세분화집단

의 구성원의 멤버로 구성할 수 있지만 이러한 분류를 위한 기준이 사용자에게 명백하게 알려져있지 않다. 인공신경망 네트워크기법의 또다른 단점은 높은 비율의 의연성 오류(FALSE NEGATIVE)가 날지도 모른다는 것이다. 의연성 오류란 예기치 못한(기대치 못한) 특성의 관찰로 정의되어 진다. 인공신경망 데이터마이닝테크닉의 또다른 장점은 선형 판별 분석보다 낮은 오류율을 보여주며 서로 다른 영역에 대하여 상대적으로 일관성있는 예측을 제공해 준다. 그러나 이 모든 데이터마이닝테크닉들도 한 개 이상의 고객 세분화집단에 속해 있는 데이터(고객정보)를 모델링하는데 불충분하다.

KIM 등은 관측데이터를 한 개 이상의 분류에 세분화시키는 가능성있는 분석틀을 제공해 주고 있다. 그들에 따르면 현재 유전자 알고리즘을 이용하여 복수의 분류자를 통합하는 두 가지 방식이 있다. 1) 연속통합방식 2) 동시통합방식, 연속통합분류자에서 한 분류자에서 나온 결과물은 다음 분류자의 INPUT으로 순차적으로 진행된다. 연속통합분류자와는 달리 분류될 입력데이터가 복수의 분류자의 동시에 입력된다. 이러한 복수의 분류자 각각으로부터 나온 결과는 알고리즘을 통하여 통합되어 진다. 분류자들로부터 나온 결과물은 다음 세 가지 방식가운데 하나로 분류되어진다. 1) 추상적 방식 2) 순위적방식 3) 측정방식

추상적방식은 클래스를 의미하고 순위적방식은 각각의 클래스에 대한 순위를 나타내고 측정방식은 INPUT패턴과 클래스간의 관계를 나타낼 가능성을 의미한다. 비록 KIM 등에 의해 수행되어진 초기연구가 가능성이 있긴 하지만 순위적 추상적

방법을 고려하지 않고 있다. 그러므로 복수의 분류 자테크닉을 위한 과제는 복수의 분류과정을 수행하며 바람직한 종류의 OUTPUT을 고려하는 방식(추상, 순위, 측정 또는 이들의 조합)

끝으로 발견지향 데이터마이닝과정은 데이터마이닝 알고리즘에 선택, 가설설정, 모델평가 및 수정과정으로 이루어진다. 이러한 과정은 시간이 많이 소요될 수 있다. 그러므로 발견과정에 대한 한 가지 과제는 과정을 좀 더 체계화하고 좀 더 생산적으로 하는 방식을 발견하는 것이다.

2. 집체화

CRM기술에 있어서 또다른 과제는 데이터 STRUCTURE에 저장된 데이터의 크기가 매우 크지만 접속시간성능은 효율적이어야 하는데 있다. 고객들로 하여금 성과수준을 유지하면서 저장된 정보를 효율적으로 활용할 수 있는 한 가지 기법은 집체정보와 CUBE를 미리 구축하는 것이다. 집체정보는 데이터를 간결하고 통합된 요약형태로 저장할 수 있도록 한다. 일반적으로 CUBE는 모든 가능한 부분집합에 대한 집체정보를 계산한다. 그 결과 희귀 CUBE와 데이터베이스 폭발현상을 가져올 수 있다. 데이터 폭발은 다섯 개 이상의 DIMENSION을 가진 희귀데이터에 대한 집체정보를 미리 계산할 때 발생하는 흔히 볼 수 없는 문제이다. 만일 모든 집체정보가 미리 계산되어진다면 데이터베이스가 훨씬 커질 것이고 메모리제약으로 인하여 추가적인 디스크접속을 초래할 것이기 때문에 성능이 저해될 수 있다. 데이터베이스크기가 커질수록 각각의 측정가능한 데이터들이 검색에서 결코 사용되어지지 않는 복수의 추가정보를 초래할 수 있다. 그 결과 데이터스트럭처는 기하급수적으로 증가할 수 있다. 그러므로 CRM을

지원하기위한 CUBE를 구축할 때 아래와 같은 점의 주의가 요망된다.

- 1) 모든 집체정보의 미리 구축하는 것을 피할 것.
- 2) 하이퍼CUBE대신 멀티CUBE방식을 사용할 것,

하이퍼 CUBE방식은 데이터값이 모든 DIMENSION의 조합에 입력된다. 또한 하이퍼 CUBE는 단일 데이터테이블만을 취급한다. 반면에 멀티CUBE방식은 서로 다른 DIMENSION을 가진 복수의 데이터테이블을 취급할 수 있다. 또한 멀티CUBE방식은 매우 희귀한 데이터를 저장하는데 좀 더 효율적이고 데이터 폭발현상을 감소시킬 수 있다.

추가적으로 KIMBELL의 최소DIMENSION기법을 사용하여 데이터레코드 생성시에 알려지지 않았던 특성들을 데이터테이블에 링크되어 있는 별도의 DIMENSION으로 이동함으로써 데이터테이블의 희귀데이터를 최소화하는데 이용되어질 수 있다. 이러한 접근방식이 데이터베이스 폭발효과를 최소화하기 위하여 사용되어질 수 있지만 어떤 집체화정보가 미리 구축되어져야 하는 방법을 결정하는 유리스트릭을 개발할 필요성이 있다. 관계형 OLAP 데이터웨어하우스를 위한 집체화정보 및 CUBE를 구축하는 방법에 관한 많은 연구가 이루어졌지만 대다수의 연구결과는 상업용 MOLAP 소프트웨어에 반영되어 있지 않다. LI 등에 의한 연구가 MOLAP를 위한 집체화정보를 구축하는 일반 알고리즘을 제공하고 있지만 MOLAP CUBE나 다른 OLAP작업을 구축하는 어떤 알고리즘도 제공하지 못하고 있다.

V. Conclusions and Future Areas For Research

본 연구과제에서 우리는 고객수익성, 고객유지, 채널수익성 및 고객간 영향도를 위한 사회적 네트워크의 발견 등의 CRM관련 분석을 지원하는 확고한 데이터웨어하우스스키마를 제시하였다. 또한 CRM 데이터웨어하우스를 위한 몇 가지 연구과제에 대하여 논의하였다. 이러한 논의 과제는 향후 중요한 연구과제를 제시하고 있다. 예컨대 MASSEY 등에 따르면 기업들이 지식자원을 활용하고 거래중심으로부터 시장중심 또는 관계형 고객중심으로 이전하는 방법에 관한 보편적 이론이나 경험적 연구가 거의 없었다고 지적하고 있다. 추가적으로 고객세분화(고객분류)가 웹상에서의 개인화, 일대일마케팅 및 제품제공 등을 위한 강력한 분석수단을 제공하고 있지만 고객이 복수의 고객세그먼트의 속할 경우에 단일 분류기법을 사용하는 데에는 제한이 있다. 그러므로 향후 연구를 위한 중요한 과제는 복수의 고객세분화집단에 속하는 고객들을 효율적이고도 효과적으로 분석할 수 있는 좀 더 개선된 알고리즘을 개발하는 것이다. 이러한 종류의 연구는 단일분류기법에 기반된 연구보다 복수분류기법에 기반된 연구가 좀 더 정확하기 때문에 가치있는 연구가 될 것이다.

복수고객세분화멤버십에 대한 또다른 접근방식은 세분화분석을 위한 서로 다른 스키마에 적합성을 조사하는 것이다. 달리 말해 4번에 의하여 추천된 스타스키마가 어떤 타입의 분석에 부적절할 것인가 달리 말해 데이터간에 미리 지정된 관계를 채택하고 있는 스타스키마가 선입관없이 데이터속의 새로운 관계를 발견하고자 하는 지식발견과정의 부정적인 영향을 미칠 것인가. 만일 그렇다면 어떤 스키마가 어떤 분석 타입에 가장 적합할 것인가. 어떤 스키마(플랫스키마)보다 나은 분석시간, 보다 나은 분석결과 그리고 보다 나은 분석 해석을 제공할 것인가. CRM데이터웨어하우스스키마와 데이터마트타입이 특정한 종류의 분석에 적합한지 알려주는 분류표를 개발하는 것도 의미있는 연구가 될 것이다. 그러한 연구결과의 의의는 데이터마이닝을 지원하기 위한 데이터스트럭처 구축하는 규칙을 발견하는 것이다. 서로 다른 CRM관련 분석들이 선택되고 서로 다른 데이터마트 타입(멀티CUBE, 하이퍼CUBE 등)을 사용하여 선택된 데이터CUBE타입에 적합한 CUBE를 구축하는 유리스틱을 개발할 수 있다.

참고문헌

- [1] Anindya Datta and Helen Thomas, The cube data model: a conceptual model and algebra for on-line analytical processing in data warehouses. *Decision Support Systems* 27 (1999), pp. 289-301.
- [2] Lloyd Alan Fletcher, Going beyond the buzzword: what exactly is CRM? *Learned Publishing* 14 (2001), pp. 213-222.
- [3] Barton J. Goldenberg, *CRM Automation*. Prentice Hall, New Jersey (2002).
- [4] Claudia Imhoff, Lisa Loftis and Jonathan G. Geiger, *Building the Customer-Centric Enterprise*, Wiley Computer Publishing, New York, 2001.
- [5] Eunju Kim, Wooju Kim and Yillbung Lee, Combination of multiple classifiers for the customer's purchase behavior prediction. *Decision Support Systems Article In Press* 975 (2002).
- [6] Ralph Kimball and Margy Ross, *The Data Warehouse Toolkit* (second edition), Wiley Computer Publishing, New York, 2002.
- [7] Jianzhong Li and Jaideep Srivastava, Efficient Aggregatin Algorithms for Compressed Data Warehouses. *IEEE Transactions on Knowledge And Data Engineering* 14 3 (2002), pp. 515-529.
- [8] M. Sweiger, M. R. Madsen, J. Langston, and H. Lombard, *Clickstream Data Warehousing*, Wiley, 2002.
- [9] Anne P. Massey, Mitzi M. Montoya-Weiss and Kent Holcom, Re-engineering the customer relationship: leveraging knowledge assets at IBM. *Decision Support Systems* 32 2 (2001), pp. 155-170.
- [10] Hamid R. Nemati, David M. Steiger, Lakshmi S. Iyer and Richard T. Herschel, Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems* 33 (2002), pp. 143-161.
- [11] T. S. Raghu, P. K. Kannan, H. R. Rao and A. B. Whinston, Dynamic profiling of consumers for customized offerings over the Internet: a model and analysis. *Decision Support Systems* 32 2 (2001), pp. 117-134.
- [12] Michael J. Shaw, Chandrasekar Subramaniam, Gek Woo Tan and Michael E.

- Welge, Knowledge management and data mining for marketing. *Decision Support Systems* 31 1 (2001), pp. 127 – 137.
- [13] J.P. Shim, Merrill Warkentin, James F. Courtney, Daniel J. Power, Ramesh Sharda and Christer Carlsson, Past, present, and future of decision support technology. *Decision Support Systems* 33 (2002), pp. 111 126.
- [14] William E. Spangler, Jerrold H. May and Luis G. Vargas, Choosing data-mining methods for multiple classification: Representational and performance measurement implications for decision support. *Journal of Management Information Systems* 16 1 (1999), pp. 37 62.
- [15] Panos Vassiliadis and Timos Sellis, A Survey on Logical Models for OLAP Databases. *SIGMOD Record* 28, 4 (December 1999), pp. 64 69.
- [16] Peter C. Verhoef and Bas Donkers, Predicting customer potential value an application in the insurance industry. *Decision Support Systems* 32, 2 (2001), pp. 189 – 199.
- [17] Russell S. Winer, A framework for customer relationship management. *California Management Review* 43 4 (2001).
- [18] Robert Winter, The Current and Future Role of Data Warehousing in Corporate Application Architecture. www.crm-forum.com, (2000).