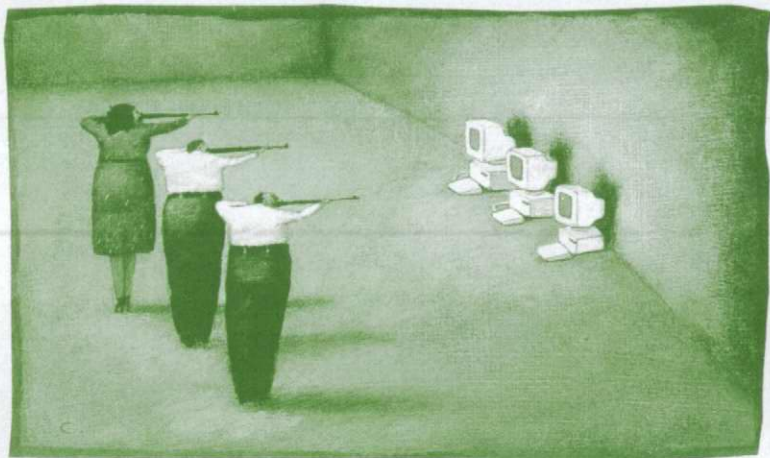


풍부한 경험, 지식이 성공 열쇠



데이터 마이닝이 기업의 수익을 창출하는 가장 주목받는 프로세스로 대두된 것은 이미 오래 전이다. 그러나 데이터 마이닝을 가장 효율적이고 기업에 맞게 사용하기 위해 필요한 요소가 무엇인지에 대한 고민들이 부족한 것은 사실이다. 따라서 데이터 마이닝과 관련하여 업계에서 논란이 되고 있는 주요 이슈를 한국 SAS의 자료를 중심으로 살펴본다. <편집자>

데

이터 마이닝(Data Mining)은 흔히 얘기하듯 데이터에서 금맥을 찾아내는 일이라 할 수 있다. 즉 기존의 축적된 데이터에서 기업의 경쟁력을 높여줄 수 있는 다양하고 효과적인 '정보'를 찾아내는 작업이라고 할 것이다.

이러한 데이터 마이닝이라는 개념과 실제 도구가 국내에 소개된 것은 그리 오래전 얘기가 아니다. 국내에서의 데이터 마이닝은 통신/금융산업의 이탈방지, 카드 혹은 휴대폰 도난도용방지, 유통의 상품진열분석 등을 위해 사용되다가 현재는 크게 고객관계관리(CRM)의 비즈니스 업무 분석, 전략 도출-고객데이터웨어하우스 구축-고객데이터 가치 분석-고객별 전략 실행(call center, SFA, Campaign 등)-평가 부분 중 고객 가치분석 및 고객별 전략 실행 결과 분석에 데이터 마이닝 기법이 사용되고 있다.

DW, OLAP, Mining의 통합 연계 활용

OLAP과 데이터 마이닝을 별개의 정보기술로 생각하는 경향이 아직까지는 많지만 별개의 기능과 특성을 지닌 것이 아니라 하나의 주제를 구현하기 위해 서로 상호보완적으로 활용되어야 하는 관계라는 사실을 점차 인식하고 있다. OLAP의 시각화 기

법과 데이터 마이닝의 분석기법을 적절히 활용함으로써 효과적으로 원하는 분석을 수행할 수 있다. 즉, OLAP은 데이터를 레포팅하고 탐색하기 위한 강력하고 빠른 기능을 제공하는가 하면 데이터 마이닝은 데이터에 대한 숨겨진 패턴을 찾는 기능을 담당하는 것이다.

가령, 고객 세분화를 위해 기업에 따라 OLAP을 사용할 수도 있고, 데이터 마이닝을 사용할 수 있다. 여기에서 고객 세분화를 위해 OLAP이라는 기법을 이용했을 때, 데이터 마이닝이라는 기법을 이용했을 때, 그리고 OLAP의 장점과 데이터 마이닝 장점을 결합하여 이용했을 때 어떤 차이가 있는지를 살펴봄으로써 고객 세분화라는 목적을 달성하는데 가장 적합한 방법이 무엇인지에 대한 답을 구할 수 있다.

가. OLAP 활용예: 단일 기준에 의한 고객 세분화

OLAP을 이용한 고객세분화는 다차원 분석 결과 정보를 바탕으로 한 세분화로 인구통계학적, 심리분석적 등의 정보나 RFM(Recent, Frequency, Monetary)에 의해 고객 세분화를 위한 기술로 적용된다. 여기에서 OLAP의 장점과 단점을 살펴보면 <표 1>과 같다.

〈표 1〉 OLAP의 장점 및 한계

OLAP의 장점	OLAP의 한계
<p>동시에 여러 차원으로 데이터를 탐색할 수 있는 시각적 기법 탁월</p> <ul style="list-style-type: none"> ● Identify Clusters! 유사한 특성을 지닌 고객군을 시각적으로 쉽게 파악 가능 ● Identify Trend! 고객 데이터의 시간 흐름에 따른 변동 사항을 시각적으로 쉽게 파악 가능 ● Identify Outliers! 고객 데이터의 이상치를 시각적으로 쉽게 파악 가능 	<ul style="list-style-type: none"> ● 마케터나 분석가의 감에 의존하는 경향 ● 연속 변수 handling의 한계 (그룹 변수를 생성하여 처리) ● 향후 고객 행동에 대한 예측 요인에 의한 세분화가 불가능

나. 데이터 마이닝의 활용 예: 모델에 의한 고객 세분화

OLAP이 데이터 탐색과정을 거쳐 사람의 직감이나, 단일한 기준에 의해 고객을 세분화한다면, 데이터 마이닝은 고객의 행동 예측 모델에 의해 고객을 세분화한다. Clustering, Classification, Regression 등의 다양한 분석 기법을 이용하여 고객 세분화를 실시하며, 다른 마이닝 결과물(고객 이탈 점수, 부실 예측 점수 등)을 활용한 고객 세분화가 가능하다.

그럼 여기에서 다시 데이터 마이닝을 이용한 고객세분화의 장점과 단점을 살펴보면 〈표 2〉와 같다.

〈표 2〉 데이터 마이닝의 장점 및 한계점

데이터 마이닝의 장점	데이터 마이닝의 한계점
<p>데이터 마이닝에서 제공하는 다양한 기법들을 이용하여 패턴 인식 기능 탁월</p> <ul style="list-style-type: none"> ● Clustering ● Association ● Decision Tree ● Neural Network 등 <p>다양한 요인을 복합적으로 고려한 고객 세분화 가능</p>	<p>데이터 마이닝에 의하여 얻어진 정보를 다차원적으로 검색하거나 리포팅하는 기능이 있어서 OLAP보다 한계가 있음</p>

여기에서 주목해야 할 것은 OLAP이나 데이터 마이닝 각각을 이용한 고객 세분화는 완벽하지 못하다는 점이다. OLAP의 단점을 데이터 마이닝이 보완해야 하고, 데이터 마이닝의 단점을 OLAP이 보완했을 때 보다 만족스러운 고객세분화가 이루어질 수 있다는 점을 주시할 필요가 있다. 그럼, 데이터 마이닝 기법과

OLAP이 어떠한 상호 보완적인 관계를 맺고 있는가를 다음의 그림에서 살펴보자

〈표 3〉 OLAP과 데이터 마이닝 통합 활용 예

패턴1	Mining then Cubing 고객세분화(Mining) 후 결과분석(OLAP)
패턴2	Cubing then Mining + Mining then Cubing 탐색 및 추출(OLAP) - 고객세분화(Mining) - 결과분석
패턴3	Cubing While Mining 탐색 및 추출(OLAP) - 고객점수별 예측(Mining) - 결과분석(OLAP) - 고객 세분화(Mining) - 결과분석(OLAP)

DW와 Mining의 메타 데이터 공유 문제

데이터 마이닝 작업은 반복적인 프로세스이며 전체 작업의 70% 이상을 차지하는 데이터 처리과정의 효율성을 위하여 분석을 위한 원천 데이터에 대한 효과적인 액세스 과정이 분석가의 생산성을 높일 수 있다.

DW와의 메타 데이터 공유를 통하여 원천 데이터와 파생 데이터에 대한 관리 및 신속한 분석 데이터 생성을 가능케 한다.

Macro Mining과 Micro Mining

흔히들 데이터 마이닝 제품을 처리할 수 있는 데이터의 크기, C/S 지원여부, 다양한 모델링 기법들에 의해 Macro Mining과 Micro Mining으로 분류한다.

데이터 마이닝을 도입하는 업체에 따라 Macro 혹은 Micro Mining을 지원하는 제품을 선택할 수 있다.

각각의 특성을 살펴보면 다음과 같다.

가. Micro Mining

- PC Standalone Type Package
- 소량의 데이터 처리/데이터 소스 액세스 제한적임
- 단일 기법만을 제공하는 경우가 많음
- SPSS, Scenario, DataMind, Business Objects 등

나. Macro Mining

- PC, Unix, Mainframe까지의 다양한 이기종 지원
- 대용량 데이터 처리, 다양한 모델링 기법 지원
- 고객관계관리, 리스크 관리 등의 전략적 프로젝트에 사용됨
- SAS, IBM 등

다양한 모델링 테크닉의 활용

데이터 사이의 가장 적합한 연관관계, 예측을 위해서는 다양한 모델링 기법들이 제공되어야 한다. 주로 사용되는 데이터 마이닝 테크닉으로는 다음과 같다

〈표 4〉 데이터 마이닝 테크닉 사용 기법 (출처: Meta Group)

데이터 마이닝 테크닉	사용현황
Decision Tree	47%
Regression Analysis	38%
Data Visualization	38%
Neural networks	32%
Induction	30%
Clustering	14%

〈표 5〉 데이터 마이닝 기법간의 상대적인 장점

기 법	장 점
Decision Tree	<ul style="list-style-type: none"> ● 명확한 규칙의 형태로 제공되어 모델의 설명력 우수 ● 결과에 대한 규칙제공으로 중요차원 발견 가능 ● 변수간 상관관계에 대한 이해 용이
Neural Network	<ul style="list-style-type: none"> ● 다양한 비즈니스 문제를 다룰 수 있음(예측, 세분화) ● 다양한 데이터 형태 처리(연속 변수, 분류 변수) ● 설명변수와 종속 변수간의 비선형 관계에 대한 설명력
Regression	<ul style="list-style-type: none"> ● 분석결과에 대한 설명 가능
Clustering	<ul style="list-style-type: none"> ● 데이터에 대한 사전 지식이 없이도 학습 가능 ● 데이터가 갖는 특성을 찾아 분석을 위한 시작점을 제공하는 역할 수행
Memory Based Reasoning	<ul style="list-style-type: none"> ● 다양한 데이터 형태 지원 ● 쉽게 설명 가능한 형태의 결과를 제공 ● 최소의 노력으로 Training Set 유지

성공적인 데이터 마이닝 프로젝트를 위한 고려 사항

● 데이터 마이닝 툴의 선택

데이터 마이닝을 위한 통합된 환경을 제공하는가?
다양한 업무에 지속적으로 적용 가능한가?

● 비즈니스 파트너

전사적인 IT 환경에서의 프로젝트 경험과 역량은 충분한가?
데이터 마이닝에 대한 이론적 배경과 컨설팅 인력의 수준은?

● 프로세스 정립

마이닝 전문가 양성을 위한 마인드는 가지고 있는가?
마이닝 결과를 활용할 수 있는 조직 프로세스는 갖췄나?

그 중 첫번째 고려 사항에 대한 이유는 기업들이 구축했거나 구축 예정인 데이터 웨어하우스를 기반으로 데이터 웨어하우스

가 제공하는 정제된 데이터를 이용했을 때 만이 데이터 마이닝은 제 효과를 발휘할 수 있기 때문이다.

즉, 대용량의 데이터 웨어하우스를 관리하는 데 중요한 데이터 인 메타 데이터를 원활히 이용할 수 있는가의 여부가 성공적인 프로젝트 수행에 결정적인 요인이 되고 있다는 것이다. 메타 데이터를 효과적으로 사용할 수 있는 데이터 마이닝 툴은 대용량 데이터 관리와 정보 창출에 있어 능률적인 작업과 신속한 결과를 가져다 준다.

두번째 고려 사항이 데이터 마이닝 프로젝트에서 중요한 이유는 데이터라는 방대한 창고에서 그 데이터의 패턴을 찾아내는 작업은 하루아침에 이루어지는 경험과 지식이 아니다. 같은 데이터를 가지고 작업을 한다 해도 현업부서에서 원하는 결과를 얻기는

그리 쉬운 일이 아니기 때문이다.

바꿔 말한다면 데이터 마이닝 프로젝트야 말로 풍부한 경험과 지식이 뒷받침 되어야만 성공할 수 있는 프로젝트라고 할 수 있다.

단순히 데이터를 모으고 그 데이터를 분석하는 종래의 방식과는 차별화된, 데이터의 보이지 않는 패턴을 찾아내고자 하는 데이터 마이닝은 앞서 예로든 분야 뿐 아니라 다양한 분야에서 적용이 될 것이며, 그러한 데이터 마이닝 프로젝트를 추진하거나 진행중인 기업에서는 무엇보다도 앞서 서술한 고려 사항을 어떻게 소화해 낼 것인가 라는 물음과 그 해답을 주는 파트너 선정에 신중을 기해야 할 것이다. 