

Data Mining의 예측기능을 이용한 효과적인 eCRM

강래구^{*} · 김승언^{**} · 정채영^{***}

^{*} 조선대학교 전산통계학과 · ^{**}서강정보대학 보건의료정보과

Effective eCRM using prediction function of Data Mining

Rae-Goo Kang^{*} · Seung-Eon Kim^{**} · Chai-Yeoung Jung^{***}

^{*} Dept of Computer Science & Statistics, Chosun University

^{**} Dept of Health Medical Information, Seokang College

E-mail : kangrg@hanmail.net

요 약

최근 들어 고객정보의 체계적인 분석과 고객의 다양한 패턴을 발견하고 분석 및 예측을 하기 위한 목적으로 많은 기업들이 eCRM을 빠르게 도입하고 있고 과거에 주로 사용되던 통계적 과정을 자동화하여 일반인들도 쉽게 양질의 결과를 추출하고 예측 할 수 있는 데이터마이닝으로 점점 대체되고 있는 추세이다. 이러한 데이터마이닝이 대표적으로 이용되고 있는 분야가 eCRM이다. 본 논문에서는 A할인점의 고객 데이터와 1년간의 매출 데이터를 기반으로 데이터마이닝을 통해 이듬해 고객 기여도를 예측 하는 실험을 하여 실제 데이터와 예측된 데이터와의 비교를 통해 데이터마이닝이 eCRM에 얼마나 효과적인지를 입증하였다.

ABSTRACT

Because many corporations computerize process figure enemy who is introducing eCRM fast and are used mainly at past by purpose to detect and analyze and forecast systematic analysis of customer information and various pattern of customer recently, ordinary peoples are trend that is alternated gradually by data mining that can draw and forecast result of good quality easily. Field that this data mining is used representatively is eCRM. In this treatise customer data of A discount store and sale data of 1 years experimenting that forecast customer contribution to base next year through data mining actuality data and data mining through comparison with predicted data are how effective to eCRM prove.

키워드

Data Mining, eCRM, CRM, 유전자알고리즘

1. 서 론

2000년대에 들어서면서 기업 마케팅은 매스 마케팅에서 퍼스널마케팅으로 대체 되고 있다.[1] 이러한 퍼스널마케팅의 대표적인 방법이 고객관계관리(CRM : Customer Relationship Management)이다. 기업의 관심사가 고객 행위에 대한 다양한 이해로 바뀌어감에 따라 기업의 가장 중요한 필

수 요소로 고객관계관리가 중요시 되고 있다.[2][3]

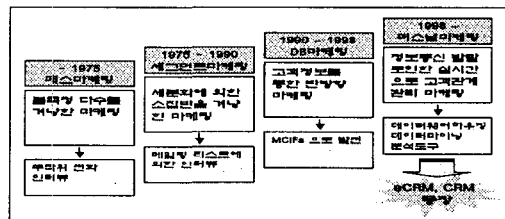


그림 1. 마케팅 발달과정

***교신저자 · cyjung@chosun.ac.kr

정보통신산업의 비약적인 발달로 인해 CRM을 기반으로 한 eCRM(Electronic Customer Relationship Management)이 새롭게 등장하였다.

eCRM의 기본적인 방법론이나 사상은 CRM과 크게 다르진 않지만 고객정보수집 및 커뮤니케이션이 기존 CRM과 달리 첨단 정보통신 기술과 인터넷 중심으로 이루어져 있다는 것이 특징이다.

최근에 eCRM이 붐을 일으키는 이유 중의 하나는 기업이 과거에 비해 엄청난 양의 고객 데이터를 축적하게 되었고 그것을 가능하게 하는 데이터웨어하우스, 데이터마이닝 같은 기술이 많이 보급되고 연구되었기 때문이라 할 수 있다.[1][4][6]

이러한 기술의 발달로 인해 과거에 주로 사용되어 오던 회귀분석, 판별분석과 같은 전통적 통계기법에서 벗어나 대량의 데이터 터미에서 보다 빠르고 편리하게 원하는 데이터를 추출하고 예측할 수 있게 되었다.

본 논문에서는 데이터마이닝 기법 중 유전자알고리즘을 이용하여 우수고객을 예측 해내는 실험을 하였다.

II. 데이터마이닝을 통한 eCRM

eCRM은 고객과의 관계를 효과적으로 관리하는 기법을 의미하며 기업과 고객간의 상호교류를 관리하는 일종의 프로세스이다. 이러한 프로세스를 보다 효과적으로 관리하여 기업의 생존과 지속적 성장을 위해 통합된 고객중심 마케팅 전략으로 발전시켜 나가기 위해서는 무엇보다도 프로세스의 자동화가 필요하며 그 중심에 데이터마이닝이 있다.

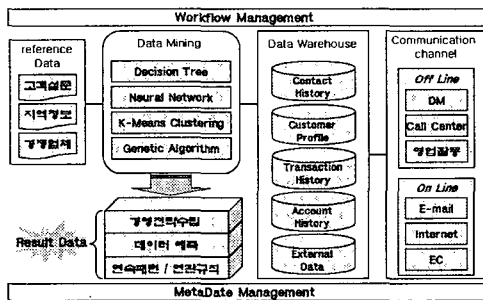


그림 2. eCRM Architecture

그림 2는 eCRM의 기본 구조를 나타내고 있다. 기업의 On/Off Line 채널을 통해 얻어진 고객개인의 각종 데이터들을 데이터마이닝 기법을 통해 보다 빠르고 편리하게 숨겨진 지식이나 예상치 못했던 새로운 정보를 발견하고 이를 분석하여 신규고객확보, 기존고객의 유지 및 이탈방지, 고객이 기업에 주는 가치의 최대화 등 다양한 영업활동이 가능하도록 해준다. 이러한 데이터마이닝 기술을 기반으로 한 eCRM을 영업 및 마케팅

활동에 최대한 활용함으로써 효과적으로 고객관리가 가능하게 되었고 보다 공격적인 마케팅 수단으로 활용 할 수 있게 되었다.[6]

2.1 데이터마이닝이란

데이터마이닝의 중요한 특징은 발견과 예측이다. 기업이 보유하고 있는 일일 거래자료, 고객자료, 상품자료, 피드백 자료와 기타 외부 자료를 포함하여 사용가능한 모든 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 예측함으로써 과거 회고적이 아닌 예측모형을 만들 수 있는 능력이 데이터마이닝의 가장 큰 매력이자 장점이다.

그렇다면 회귀분석이나 판별분석과 같은 전통적 통계기법과 데이터마이닝의 차이는 무엇일까?

데이터마이닝과 통계기법의 가장 큰 차이점은 데이터마이닝은 통계적 과정을 자동화 하여 통계전문가가 아니더라도 대량의 데이터 속에서 필요한 자료를 일반인들도 쉽게 양질의 결과를 추출하고 예측 할 수 있는 기능을 제공하는 특징이 있다.[1]

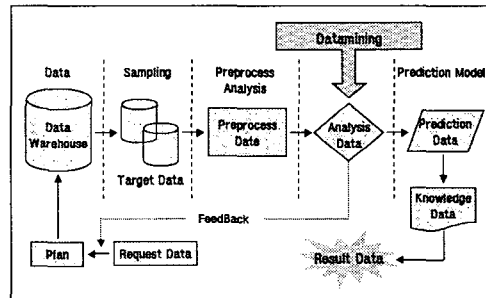


그림 3. Datamining Process

그림 3은 데이터마이닝의 처리 과정으로써 요청된 데이터를 데이터마이닝 기법을 통해 원하는 결과를 자동으로 도출하게 된다.

효과적인 마케팅은 기업의 성패를 좌우하는데 큰 영향을 미치게 된다. 이러한 마케팅을 위해 통계전문가가 아닌 일반 실무자들도 데이터마이닝 시스템을 통해 쉽고 빠르게 원하는 데이터를 추출 해냄으로써 개별 고객을 상대로 보다 질 좋은 마케팅이 가능하게 될 것이다.

이러한 데이터마이닝 기법이 대표적으로 이용되고 있는 분야가 바로 eCRM이다.

2.2 유전자알고리즘

최근 들어 데이터마이닝 기법은 통계와 같은 기존의 고전적 기법에서 신경망, 의사결정나무, 유전자알고리즘과 같은 차세대 기법으로 점점 대체되고 있다. 이러한 차세대 기법 중 본 논문에서 사용한 유전자알고리즘은 1975년 John Holland에

의해서 처음 소개 되었다.

자연계에서 적자생존의 원리에 따라 세대가 지나면서 우량의 형질을 지닌 개체가 생성되는 과정을 모방한 알고리즘으로 탐색, 예측과 같은 예상모델을 구현하는데 적절한 알고리즘이다.

집단이라는 탐색공간이 초기화되고 집단의 개체는 적합도 함수에 의해 평가된다. 이 적합도에 의해 다음 세대에 선택될 개체의 선택확률이 정해진다. 이 과정에서 선택, 교배, 돌연변이 연산을 통해 적합도가 우수한 해를 탐색하게 되고 이렇게 세대가 되풀이되면서 가장 우수한 해를 찾아내게 된다.[7]

<표 1> 실험에 사용한 연산자

선택(Selection)	Roulette wheel
교배(Crossover)	Edge Recombination
돌연변이(Mutation)	Inversion

본 논문에서 사용한 GA연산자는 표 1과 같다.

```

Genetic Algorithm
{
  generation = 1;
  While (종료조건에 만족할 때까지 반복)
  {
    For (i = 1 부터 임의의 수  $\ell$  까지 반복)
    {
      염색체  $P_1, P_2$  선택; //선택
      pop(i) = crossover( $P_1, P_2$ ); //교배
      pop(i) = mutation(pop(i)); //돌연변이
    }
    pop(i), ..., pop( $\ell$ )를 population
    내의  $\ell$  개의 염색체와 대치;
  }
  가장 우수한 염색체를 return;
}
    
```

그림 4. 유전자알고리즘의 수행 과정

그림 4 은 유전자알고리즘 수행과정을 간략화하여 나타낸 그림이다.

III. 시스템 설계

본 논문에서는 A할인점의 On/Off Line상의 고객데이터 약 5만개와 2004년 01월부터 2004년 12월까지의 매출 데이터 중 공산품 데이터 약 100만개를 사용하여 실험하였다. 고객 데이터와 매출 데이터를 유전자알고리즘을 이용한 데이터마이닝 기법을 적용하여 2005년도 우수고객 상위 50명을 예측함과 동시에 2005년도 상반기 동안 실제 고객 기여도 순위와의 비교를 통해 eCRM을 위한 데이터마이닝 기법이 얼마나 정확하게 결과를 예측해 내는지를 증명하고자한다.

이 중 고객 데이터 추출은 총 고객 데이터 총 매출 기간동안 적어도 한달에 두 번 이상 매장을 직접 방문 하였거나 소매품에 Log-In 한 내역이 있는 고객들을 자동으로 추출하여 실험 하였다.

데이터마이닝을 위해 실험에 사용할 유전자알고리즘의 구현은 Oracle 기반의 SQL문으로 Procedure와 Trigger를 직접 작성하여 데이터베이스 내에서 실행하여 결과 값은 View로 Return 하였다.

```

-----
Procedure name : GA_mining_P01
Date : 2005/09/09
Rae-Goo Kang
-----

CREATE OR REPLACE
PROCEDURE GA_mining_P01 (v_date IN CHAR) AS

/*****
= Declare Local Variables =
*****/

BEGIN

OPEN CRT1:          /* CURSOR 로 선언된 유전자알고리즘 CALL */
LOOP
FETCH CRT1 INTO
S_STR_CODE, S_MEM_CODE, S_ITM_CODE, S_SALE_SUM, S_SALE_ON;
EXIT WHEN CRT1%NOTFOUND;
IF S_STR_CODE = 'a' THEN          /* On Line 매출내역 */
UPDATE TEMP1
SET SALE_RANK = SALE_RANK + 1,
SALE_SUM = S_SALE_SUM
WHERE STR_CODE = S_STR_CODE
AND MEM_CODE = S_MEM_CODE
AND SALE_DATE = S_SALE_DATE;
IF SQL%NOTFOUND THEN
INSERT INTO T_LOD
(STR_CODE, MEM_CODE, SALE_DATE, ITM_CODE)
VALUES
(S_STR_CODE, S_MEM_CODE, S_SALE_DATE, S_ITM_CODE);
END IF;          /* IF SQL =/
ELSE          /* On Line 매출내역 */
UPDATE TEMP1
SET BEST_RANK = SALE_RANK + 1,
SALE_SUM = S_SALE_SUM
WHERE STR_CODE = S_STR_CODE
AND MEM_CODE = S_MEM_CODE;
END LOOP;
COMMIT;

EXCEPTION WHEN NO_DATA_FOUND THEN NULL;
END GA_mining_P01;
    
```

그림 5. SQL문으로 작성한 유전자알고리즘 Procedure의 일부

그림 5는 데이터마이닝 작업을 위해 직접 작성하여 사용한 유전자알고리즘 Procedure의 일부를 보여주고 있다.

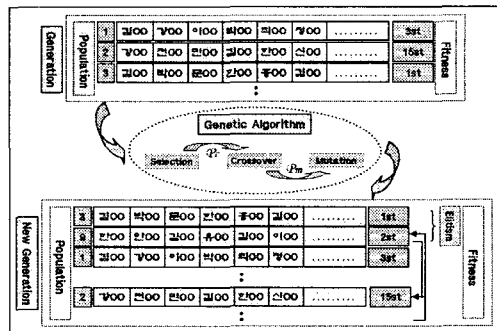


그림 6. 유전자알고리즘 진행 과정

그림 6은 유전자알고리즘이 본 실험에서 진행되는 과정을 나타낸 그림이다. 본 실험에서는 무작위 초기화법을 이용하여 최초 집단을 생성하였

고 검색체의 크기는 고객 데이터수와 일치시켰으며 개체의 크기는 1000으로 한정하였고 세대 수는 제한을 두지 않고 100세대 동안 값의 변화가 없을 경우 자동으로 종료되도록 설계하였다. 그리고 교배 확률(P_c)과 돌연변이 확률(P_m)은 0.7 과 0.1을 사용하였다.

3.1 실험 결과

실험은 P-4 2.4GHz에서 Oracle 9를 기반으로 데이터마이닝 작업을 진행하였다.

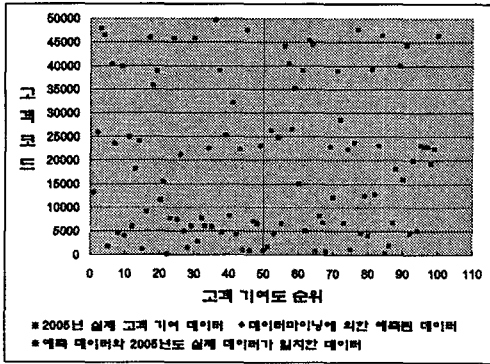


그림 7. 실제 데이터와 예측된 데이터 비교

그림 7은 2005년도 상반기 실제 고객 기여도 순위와 2004년 1월부터 12월까지 고객과 매출 데이터를 유전자알고리즘을 이용한 데이터마이닝 기법으로 예측한 순위 데이터와의 비교를 그래프로 나타내고 있다. 2005년도 실제 고객 기여도 데이터는 파랑색 사각형으로 표시하였고 예측된 데이터는 빨강색 마름모로 표시하였다.

그림 7에서 알 수 있듯이 데이터마이닝 기법으로 예측한 고객 50명 중 43명이 2005년도 상반기 실제 고객 데이터에 일치하였고 또한 50명에 포함되지 않은 나머지 7명도 실제 고객 데이터 80명 안에 모두 포함되는 결과로 예측하였다.

즉, 데이터마이닝을 통해 미래의 결과를 미리 예측해본 데이터를 추 후 실제 데이터에 비교해 보았을 때 거의 비슷한 결과를 예측하고 있음을 알 수 있다. 이와 같이 유전자알고리즘을 이용한 데이터마이닝 기법은 데이터마이닝의 가장 큰 특징 중 하나인 예측이라는 측면에 부합한 차세대 마이닝 기법 중 하나임을 실험을 통해 입증하였다.

IV. 결론

많은 기업들이 고객정보 분석과 이를 근거로 쉽고 편리하게 고객의 다양한 패턴을 발견하고 분석 및 예측하기 위해 eCRM을 빠르게 도입하고

있다.

과거엔 고객관리가 통계학자들이나 전문적인 통계패키지에 의해 관리됨에 따라 일반 실무자들이 접근하기엔 많은 어려움이 있었다.

그러나 2000년 이후 정보통신 분야의 급격한 발전을 기반으로 통계적 과정을 자동화하여 통계 전문가가 아닌 일반인들도 쉽게 양질의 결과를 추출하고 예측 할 수 있는 데이터마이닝 기법으로 점점 대체되고 있는 추세이다.

본 논문에서는 효과적인 eCRM을 위해 다양한 차세대 데이터마이닝 기법 중 하나인 유전자알고리즘을 이용하여 A사의 고객 데이터와 2004년도 매출 데이터를 기반으로 2005년도 상반기 우수 고객을 예측 해내는 실험을 하였다.

실험 결과 예측한 고객 50명 중 43명이 2005년 실제 데이터 상위 50명에 포함되었고 나머지 7명은 상위 80명 안에 포함되는 결과를 보임으로써 데이터마이닝 기법이 eCRM에 얼마나 효과적인지를 입증하였다.

또한, 본 논문에서는 데이터마이닝을 위한 유전자알고리즘을 SQL문을 이용하여 직접 구현함으로써 추후 본 논문에서 구현한 알고리즘을 이용하여 직접 eCRM System 개발도 가능할 것이다.

참고문헌

- [1] Berson.Alex, Building Data Mining Applications for Crm, McGraw-Hill. 1999
- [2] Yim CK, Kannan PK., "Consumer behavioral loyalty: a segmentation model and analysis" Journal of Business Research. Vol.44(2). pp.75-92, 1999
- [3] Kohli R, Piontek F, Ellington T, VanOsdol T, Shepard M, Brazel G, "managing customer relationships through e-business decision support applications: a case of hospital -physician collaboration" Decision Support System, Vol.32(2), pp. 171-187, 2001
- [4] 사와노보리 히데아키, "e-CRM 마케팅" 국립증권경제연구소, 2000
- [5] Fayyad, U. M, "Advances in Knowledge Discovery and Data Mining", MIT Press, 1996
- [6] 박주석, "성공적인 CRM 구축에 영향을 미치는 요인에 관한 연구", 경영과 컴퓨터, 2000
- [7] Hon, K. K. B., and H. Chi, "A New Approach of Group Technology Part Families Optimization", Annals of the CIRP, 1994