
성공적인 eCRM, CRM을 위한 데이터마이닝 기법

강래구* · 임희경* · 정채영**

Datamining technique for successful eCRM, CRM

Rae-Goo Kang* · Hee-Kyoung Lim* · Chai-Yeoung Jung**

이 논문은 2006년도 조선대학교 학술연구비의 지원을 받아 연구되었음

요 약

고객관리가 기업의 성패를 좌우하는 중요한 화두로 떠오르면서 보다 쉽고 편리하게 고객의 다양한 패턴을 발견하고 예측하기 위해 많은 기업들이 CRM과 eCRM을 빠르게 도입하고 있다. 과거엔 고객관리가 통계학자들이나 전문적인 통계패키지에 의해 관리되어 왔으나 정보통신 분야의 급격한 발달을 기반으로 통계적 과정을 자동화시킨 데이터마이닝 기법으로 점점 대체되고 있는 추세이다. 이러한 데이터마이닝이 대표적으로 이용되고 있는 분야가 CRM, eCRM이다. 본 논문에서는 A할인점의 고객 데이터와 2004년도 매출 데이터를 기반으로 유전자알고리즘을 이용한 데이터마이닝을 통해 2005년도 우수 고객을 예측하였고 실제 고객 데이터와의 비교를 통해 데이터마이닝이 eCRM에 얼마나 효과적인지를 입증하였다.

ABSTRACT

To customer management finds and forecast customer's various pattern more easily and conveniently rising by important topic that control corporation's success and failure, many corporations are introducing CRM and eCRM fast. At past, customer management had been managed by statisticians or special statistics package but it is trend been alternating gradually by datamining technique to do to computerize statistics process based on sudden development of IT. Field that this datamining is used representatively is CRM, eCRM. This paper applied datamining using GA referencing customer data of discount store and sale data of 2004 years. Forecasted 2005 years melancholy customer by datamining and proved datamining through comparison with actuality customer data is how effective to eCRM.

키워드

데이터마이닝, 고객관계관리, eCRM, CRM, 유전자알고리즘

I. 서 론

70년대 이전까지 기업의 마케팅은 시장 전체 불특정 다수를 겨냥하여 대량으로 상품과 서비스를 제공하는 매

스마케팅이 중심이었다. 그 후 보다 효과적인 마케팅을 위해 세분화에 의한 소집단을 겨냥한 세그먼트 마케팅 기법이 사용되어 왔다.

2000년대에 들어서면서 기업 마케팅 기법의 화두로 퍼

* 조선대학교 전산통계학과

** 교신저자

스널 마케팅이 대두되고 있다. 이러한 퍼스널 마케팅의 대표적인 방법이 고객관계관리(CRM : Customer Relationship Management)이다.[1][2]

정보통신산업의 비약적인 발달로 인해 최근엔 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 있다는게 특징이다.[3][4]

본 논문에서는 eCRM과 CRM을 eCRM으로 통일하여 표현하였다.

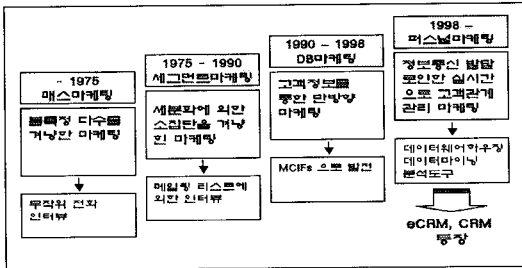


그림 1. 마케팅 발달과정
Fig. 1. Marketing development process

eCRM이 붐을 일으키는 가장 큰 이유는 데이터웨어하우스, 데이터마이닝과 같은 기술이 보급되고 연구되었기 때문이라 할 수 있다.

데이터웨어하우스란 복잡한 질의 및 분석을 위해 대량의 정보를 모아놓은 거대한 데이터 창고를 의미한다. 그리고 그 곳에 저장되어 있는 대량의 데이터를 다양한 알고리즘을 이용하여 새로운 연관관계나 패턴을 발견하고 유용한 정보들을 추출하는 과정을 데이터마이닝이라 한다.[1][5]

이러한 데이터마이닝의 발달로 인해 과거에 주로 사용되어 오던 회귀분석, 판별분석과 같은 전통적 통계기법에서 벗어나 대량의 데이터 터미에서 보다 빠르고 편리하게 원하는 데이터를 추출하고 예측 할 수 있게 되었다.[6]

본 논문에서는 데이터마이닝 기법 중 유전자알고리즘을 이용하여 A할인점의 우수 고객을 예측하였고 예측한 데이터와 실제 데이터와의 비교를 통해 데이터마이닝이 eCRM에 얼마나 효과적인지 실험을 통해 입증하고자 한다.

II. eCRM을 위한 데이터마이닝

eCRM은 고객과의 관계를 효과적으로 관리하는 마케팅 기법을 의미하며 기업과 고객간의 상호교류를 관리하는 일종의 프로세스이다.

기업의 생존과 지속적인 성장을 위해 프로세스를 보다 효과적으로 관리하여 통합된 고객중심 마케팅 전략으로 발전시켜 나가기 위해서는 무엇보다도 프로세스의 자동화가 필요하며 그 중심에 데이터마이닝이 있다.

갈수록 다양해지는 시장여건에 효과적으로 대응하기 위해 기업의 On/Off Line 채널에서 획득한 대량의 자료로부터 데이터마이닝을 이용하여 새로운 정보를 발견하고 예측하여 신규고객확보, 기존고객의 유지 및 이탈방지 등에 적극적으로 이용할 수 있게 되었고 보다 공격적인 마케팅 수단으로 활용할 수 있게 되었다.[7]

이러한 데이터마이닝이 대표적으로 이용되고 있는 분야가 바로 eCRM이다.

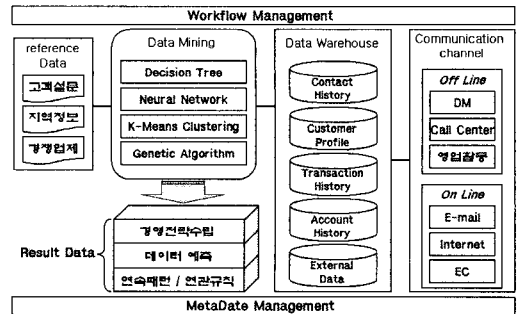


그림 2. eCRM Architecture
Fig. 2. eCRM Architecture

그림 2는 eCRM의 전체적인 구성도를 나타내고 있다.

2.1 데이터마이닝이란

데이터마이닝의 중요한 특징은 발견과 예측이다. 기업이 보유하고 있는 대량의 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 예측함으로써 과거 회고적이 아닌 예측모델을 만들 수 있는 능력이 데이터마이닝의 가장 큰 장점이다.

그렇다면 회귀분석이나 판별분석과 같은 전통적 통계기법과 데이터마이닝의 차이는 무엇일까?

과거에는 통계학자들에 의해 통계기법을 기반으로 의미있는 데이터를 찾기 위해 데이터베이스를 분석 해왔으

나 데이터마이닝은 이러한 과정을 자동화하여 통계전문가가 아니더라도 대량의 데이터 속에서 필요한 자료를 손쉽게 추출하고 예측 할 수 있는 기능을 제공하는 특징이 있다. 이러한 특징으로 인해 오늘날 데이터마이닝이 주목을 받고 있는 것이다.[1]

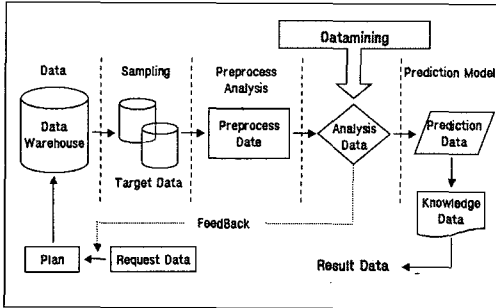


그림 3. Datamining Process
Fig. 3. Datamining Process

그림 3은 데이터마이닝의 처리 과정을 보여주고 있다.

2.2 유전자알고리즘

최근에 데이터마이닝은 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 소개되었다. 자연계에서 적자생존의 원리에 따라 세대가 지나면서 우량의 형질을 지닌 개체가 생성되는 과정을 모방한 알고리즘으로 정보의 탐색이나 예측과 같은 모델을 구현하는데 적절한 알고리즘이다.

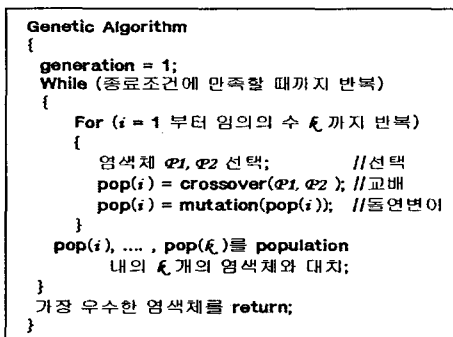


그림 4. 유전자알고리즘의 실행 과정
Fig. 4. Execution process of Genetic Algorithm

그림 4는 유전자알고리즘 수행과정을 간략하게 나타낸 그림이다.

집단이라는 탐색공간이 초기화되고 집단의 각 개체는 적합도 함수에 의해 평가된다. 이 과정에서 선택, 교배, 돌연변이 연산을 통해 적합도가 우수한 해를 탐색하게 되고 세대가 되풀이되면서 가장 우수한 해를 찾아내게 된다.[8]

표 1. 실험에 사용한 연산자
Table 1. Operator used in experiment

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

본 논문에서 사용한 GA연산자는 표 1과 같다.

III. 시스템 설계

본 논문에서는 A할인점의 2004년 01월부터 2004년 12월까지의 On/Off Line 상의 매출 데이터 중 공산품 데이터 약 100만개와 고객 데이터 약 5만개 중 매출 기간동안 적어도 한달에 두 번 이상 매장을 방문 하였거나 쇼핑몰에 Log-In 한 내역이 있는 고객만을 대상으로 데이터마이닝을 적용하여 2005년도 우수고객을 순위와 관계없이 50명을 예측함과 동시에 2005년도 상반기 실제 고객 데이터 중 고객 기여도가 높은 상위 50명과의 비교를 통해 eCRM을 위한 데이터마이닝이 얼마나 정확한 결과를 예측해 내는지 증명하고자한다.

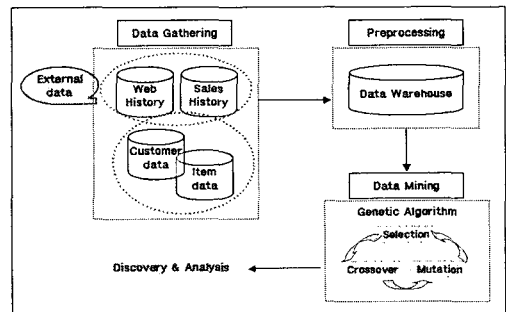


그림 5. 시스템 구성
Fig. 5. System composition


```

Oracle SQL Plus
SQL> select min as '사',
2 min_code as 'Pre(code)',
3 min_name as 'Pre(name)',
4 min_code as '2005(code)',
5 min_name as '2005(name)'
6 from temp1;

```

sq	Pre(code)	Pre(name)	2005(code)	2005(name)
1	13281	박병*	13281	박병*
2	23522	김정*	23922	김정*
3	87827	김정*	87827	김정*
4	86875	정희*	86875	정희*
5	1881	김옥*	1881	김옥*
6	8882	김희*	8882	김희*
7	2351	박정*	2351	박정*
8	8654	이진*	8654	이진*
9	39981	김정*	39981	김정*
10	882	이승*	882	이승*
11	25842	이정*	25842	이정*
12	5978	지정*	5978	지정*
13	18187	이정*	18187	이정*
14	24275	신정*	24275	신정*
15	1189	김민*	1189	김민*
16	9228	정민*	9228	정민*
17	86811	김정*	86811	김정*
18	35882	김정*	35882	김정*
19	38978	김정*	38978	김정*
20	11528	이정*	11528	이정*
21	15431	이정*	15431	이정*
22	178	이정*	178	이정*
23	7714	신정*	7714	신정*
24	85388	이정*	85388	이정*
25	7429	이정*	7429	이정*
26	21172	이정*	21172	이정*
27	4982	이정*	4982	이정*
28	837	이정*	837	이정*
29	23817	이정*	23817	이정*
30	12174	이정*	12174	이정*
31	39988	이정*	39988	이정*
32	28462	이정*	28462	이정*
33	4492	이정*	4492	이정*
34	22388	이정*	22388	이정*
35	182	이정*	182	이정*
36	23758	이정*	23758	이정*
37	81891	이정*	81891	이정*
38	831	김정*	831	김정*
39	15428	이정*	15428	이정*
40	8118	이정*	8118	이정*
41	39266	이정*	39266	이정*
42	4089	이정*	4089	이정*

그림 9. SQL Plus로 조회한 결과
Fig. 9. Result that inquire by SQL

그림 9는 데이터마이닝 작업을 수행한 결과 값을 저장해 놓은 View를 Oracle SQL Plus에서 Select 한 결과를 보여주고 있다.

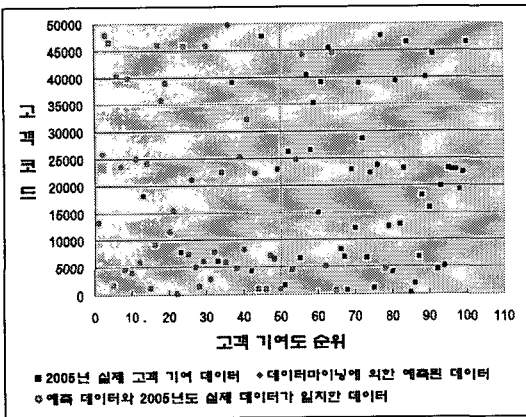


그림 10. 실제 데이터와 예측된 데이터 비교
Fig. 10. Actuality data and predicted data comparison

그림 10은 그림 9에서 Select 한 결과 값을 근거로 2005년도 상반기 실제 고객 기여도 순위와 2004년 1월부터 12월까지 고객과 매출 데이터를 유전자알고리즘을 이용한 데이터마이닝 기법으로 예측한 데이터와의 비교를 그래프로 나타내고 있다. 2005년도 실제 고객 기여도 데이터

는 파랑색 사각형으로 표시하였고 예측한 데이터는 빨간색 마름모로 표시하였다.

그림 9와 10에서 알 수 있듯이 데이터마이닝을 통해 예측한 고객 50명 중 43명이 2005년도 상반기 실제 고객 기여도 상위 50명 안에 정확히 일치함으로써 86%의 예측률을 보였고 또한 50명 안에 포함되지 않은 나머지 7명도 상위 78명 안에 모두 포함되었다.

2005년도 실제 고객 기여도 상위 50명 중 데이터마이닝을 통해 예측하지 못한 7명은 본 실험 조건에 만족하지 않은 고객이 다수 포함되어 있는 것으로 확인되었다. 실험을 위한 고객 데이터 추출 조건은 2004년 1월부터 12월까지 정상적인 매출이 일어난 고객 데이터에 한하여 실험하였으나 예측하지 못한 7명 중에는 2005년도에 새로 등록된 고객과 2004년도 매출이 없었던 고객이 다수 포함되어 있었다.

이러한 조건을 감안 한다면 유전자알고리즘을 이용한 데이터마이닝 기법에서 보여준 86%의 예측률은 결코 낮은 수치라 할 수 없을 것이다.

이와 같이 유전자알고리즘을 이용한 데이터마이닝 기법은 데이터마이닝의 가장 큰 특징 중 하나인 예측이라는 측면에 부합한 차세대 마이닝 기법 중 하나임을 실험 결과를 통해 입증하였다.

V. 결 론

많은 기업들이 성공적인 마케팅을 위해 고객정보의 체계적인 분석과 다양한 패턴을 발견하고 분석 및 예측을 하기 위해 고객관계관리로 불리는 CRM과 eCRM을 빠르게 도입하고 있다.

과거엔 고객관리가 통계학자들이나 전문적인 통계패키지에 의해 관리되어 왔으나 2000년 이후 정보통신분야의 급격한 발달을 기반으로 통계적 과정을 자동화하여 통계전문가가 아닌 일반인들도 쉽게 양질의 데이터를 추출하고 예측 할 수 있는 데이터마이닝으로 점점 대체되고 있는 추세이다.

본 논문에서는 효과적인 eCRM을 위해 다양한 차세대 데이터마이닝 기법 중 유전자알고리즘을 이용하여 A사의 고객 데이터와 2004년도 매출 데이터를 기반으로 2005년도 우수 고객을 예측 해내는 실험을 하였다. 실험 결과 예측한 고객 50명 중 43명이 2005년 실제 고객 기여도 데

이더 상위 50명 안에 포함되었고 나머지 7명은 상위 78명 안에 포함되는 결과를 보임으로써 데이터마이닝이 eCRM에 얼마나 효과적인지를 입증하였다.

이러한 마이닝 기법의 발달로 인해 SAS나 SPSS, MINITAB과 같은 전문적인 통계 패키지를 습득한다거나 통계전문가가 아니더라도 데이터마이닝을 통해 비전문가 누구나 원하는 정보에 대한 분석과 예측을 보다 쉽고 편리하게 할 수 있을 것이다.

또한, 본 논문에서는 데이터마이닝을 위한 유전자알고리즘을 SQL문으로 직접 작성하여 구현함으로써 추후 본 논문에서 구현한 알고리즘을 이용하여 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), 1999
- [3] Kohli R, Piontek F, "managing customer relationships through e-business decision support applications: a case of hospital-physicion collaboration" Decision Support System, Vol.32(2), pp. 171-187, 2001
- [4] 사와노보리 히데아키, "e-CRM 마케팅" 국일증권경제연구소, 2000
- [5] J. Widom, "Research Problems in Data Warehousing", Proceedings of the Fourth International Conference on Information and Knowledge Management, pp.25-30, 1995
- [6] Fayyad, U. M, "Advances in Knowledge Discovery and Data Mining", MIT Press, 1996
- [7] 박주석, "성공적인 CRM 구축에 영향을 미치는 요인에 관한 연구", 경영과 컴퓨터, pp.262-265, 2000
- [8] Hon, K. K. B., and H. Chi, "A New Approach of Group Technology Part Families Optimization", Annals of the CIRP, 1994

저자소개

강 레 구(Rae-Goo Kang)



2005년 조선대학교 전자공학과 공학사
2005년 - 조선대학교 전산통계학과 석사과정

※관심분야: 신경망,인공지능,정보보호,Bioinformatics
데이터베이스

임 희 경(Hee-Kyoung Lim)



2003년 조선대학교 전산통계학과 박사과정수료
2006년 - 서강정보대학 컴퓨터정보과 겸임교수

※관심분야: 신경망,인공지능,정보보호,Bioinformatics

정 채 영(Chai-Yeoung Jung)



1989년 조선대학교 대학원 전산전공 학박사
1986년 - 조선대학교 컴퓨터통계학과 교수

※관심분야: 신경망,인공지능,정보보호,Bioinformatics
※ e-mail : cyjung@mail.chosun.ac.kr ※062)230-6625