

eCRM에서 최적화된 고객 집단 편성 알고리즘에 관한 연구

A Study on Optimized Customer-Group-Organization Algorithm in eCRM

이재훈, 이성주

조선대학교 컴퓨터공학부

Jae-Hoon Lee, Sung-Joo Lee

Dept. of Computer Science Chosun University

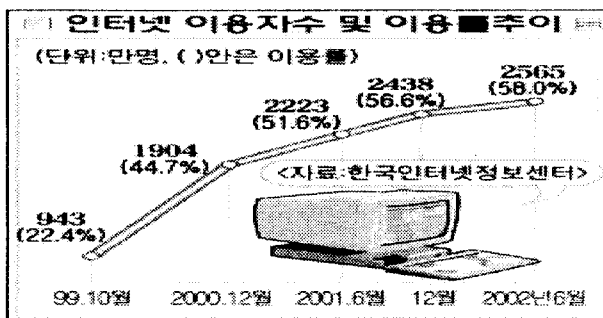
E-mail : nuridepo@stmail.chosun.ac.kr, sjlee@chosun.ac.kr

요 약

인터넷을 비롯한 온라인의 발전은 기업들의 고객중심 마케팅 기법중 하나인 고객관계관리(CRM: Customer Relationship Management)를 다양하게 발전시켰고 최근 가장 큰 이슈가 되고 있다. eCRM이란 CRM에서 인터넷을 이용해 기존의 시스템을 재구성하는 것을 말하는데 고객만족을 극대화하면서 동시에 관련 비용을 절감할 수 있는 새로운 고객관리라고 할 수 있다. 본 논문은 고객 성향 분석과 개인화에서 얻어진 일련의 정보를 다시 한번 더 가공함으로써 고객 집단 편성을 최적화 할 수 있음을 보였다.

1. 서론

인터넷은 이제 우리 생활에 있어 매우 밀접한 환경이 되고 있다. 세계적으로 인터넷을 사용하는 인구가 급속하게 증가하고 있고 국내도 2500만 이상의 인구가 인터넷을 사용하고 있다는 통계가 보도된 바 있다[그림 1][1].



[그림 1] 국내 인터넷이용자 및 이용률추이(2002. 6월)

이에 따라 인터넷만을 채널로 하는 닷컴사들이 엄청난 속도로 증가하고 있으며 기존의 기업들도 하나둘씩 인터넷 기반 영역으로의 사업 확장을 서두르고 있다. 이러한 e-business 기업들에게 있어서 가장 중요한 이슈로 떠오르는 것은 다름 아닌 고객관계관리(CRM: Customer Relationship

Management)이다.

인터넷은 기존의 오프라인 비즈니스와는 다른 여러 변화를 가져오고 있다. 기존에 고객들은 자신이 원하는 상품과 서비스를 제공하는 기업들을 선택하는데 있어 선택의 폭이 넓지 않았다. 여러 기업들이 제공하는 상품과 서비스를 비교하여 선택할 수 있는 제반 정보도 부족했으며 제공업체를 전환하는 데 소요되는 비용과 노력이 만만치 않아 쉽게 전환할 수 없는 어려움을 가지고 있었다. 그러나 인터넷 세상에서의 모습은 이와는 다르다. 그야말로 고객이 왕이 되는 시장이 바로 e-Business 시장인 것이다. 고객이 인터넷 사이트에 방문하여 얻는 정보와 경험의 만족도에 따라 고객이 그 기업에게 갖는 충성도가 결정될 수 있으며 이는 곧 기업의 매출로 연계될 수 있다.

이러한 치열한 경쟁 속에서 기업이 경쟁우위를 갖기 위해서는 자사 웹사이트를 방문한 고객들을 얼마나 잘 유지하고 충성도를 높이느냐가 관건이 되는 것이다. e-Business 프로세스 기반으로 수익성있는 고객을 선별하고 고객의 만족을 통하여 가치를 극대화하는 고객과의 관계 정립을 eCRM 이라고 정의할 수 있다.

본 논문은 고객 분류 최적화 알고리즘으로 분

류된 고객의 정보를 개인화와 마이닝을 이용한 또 한번의 가공으로 집단 편성의 최적화를 도모할 수 있으며 이로 인해 기업의 고객 관리가 더욱 강화될 수 있음을 제시하였다.

본 논문은 다음과 같은 구성으로 되어있다. 2장에서 관련 연구로 데이터 마이닝의 작업 유형과 고객 분류 최적화 알고리즘을 살펴본다. 3장에서 고객 집단 편성 최적화 알고리즘을 제안하고 4장에서 결론 및 향후 연구과제를 제시한다.

2. 관련 연구

2.1 데이터 마이닝의 작업 유형

2.1.1 연관규칙

“어떤 상품들이 함께 잘 팔리냐?” “A제품을 구입한 고객에게 어떤 제품을 함께 팔 수 있을까?” “카드사가 유형간에는 어떤 관계가 있나?” “와 같이 한 항목을 알 경우 다른 항목을 예측할 수 있다면, 이들의 관계는 종속관계가 존재한다. 연관규칙을 발견하는 작업이란 데이터 안에 존재하는 항목간의 종속관계를 찾아내는 작업이며, 마케팅에서는 손님의 장바구니에 들어있는 품목간의 관계를 알아본다는 의미에서 장바구니 분석(market basket analysis)이라고 한다.

연관규칙은 ‘항목A’ & ‘항목B’ => “항목C”의 형태로 표현되는데, 이는 ‘항목A와 항목B를 포함한 거래는 항목C도 포함한다.’로 해석한다. 그러나 이러한 연관 규칙들은 실제 데이터를 분석할 때에는 100% 신뢰도를 가지는 경우가 거의 드물기 때문에 통상적으로 확률이나 도표등을 이용하여 정량화한다.

일반적으로 연관규칙은 제품이나 서비스의 교차판매(cross selling), 매장진열(display), 첨부우편(attached mailings), 사기적발(fraud detection) 등의 다양한 분야에 활용된다.

2.1.2 연속규칙

연속규칙이란 연관규칙에 시간관련 정보가 포함된 형태이다. 예를 들어 “새 냉장고를 구입한 고객 중 한달 이내에 새 오븐을 구입하는 경향이 많다” 와 같이 시간성에 순차적으로 나타나는 사건이나 거래의 종속관계를 의미한다.

즉 연속규칙에는 시간의 흐름이 있기 때문에 연관규칙에 비해 더 구체적이며, 목표(target)마케팅이나 일대일(one-to-one)마케팅에 바로 활용할 수 있다. 그러나 연속규칙을 찾기 위해서는 데이터에 거래일이나 거래품목 이외에도 고객의 구매 이력(history) 속성이 반드시 필요하다는 조건이 만족되어야 한다.

2.1.3 분류규칙

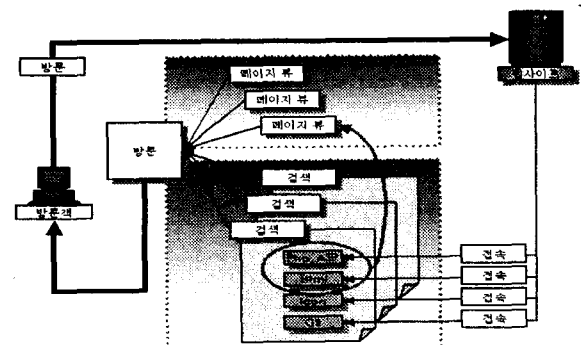
분류는 데이터마이닝에서 가장 많이 사용되는 작업으로 부류값이 포함된 과거의 데이터로부터 부류별 특성을 찾아내어 분류모형을 만들고, 이를 토대로 새로운 레코드의 부류값을 예측하는 것을 의미한다. 예를들어 신용카드 회사의 고객 신용평가 모형이 있다 하자. 이 회사에서는 지금까지의 거래를 토대로 고객들의 신용을 <우수, 보통,불량>으로 분류했다. ‘불량’ 평가를 받은 고객층의 특성중 하나가 “25~30세 가량의 미혼남으로 월 평균 수입이 200만원 이하인 고객” 이라면, 신규 카드 가입자들의 신용평가지 이러한 규칙을 활용함으로써 보다 객관적인 의사결정을 유도할 수 있다.

2.1.4 데이터 군집화

군집화란 레코드들을 유사한 특성을 지닌 몇 개의 소그룹으로 분할하는 작업을 뜻한다. 작업의 특성이 분류작업과 흡사하다고 생각할 수 있으나, 분석하고자 하는 데이터에 부류가 포함되어 있지 않다는 점에서 차이가 있으며, 다른 데이터마이닝 작업을 위한 선행작업으로서의 역할을 수행하는 경우가 많다.

2.2 고객 분류 최적화 알고리즘

일반적으로 특정 웹 페이지를 보기 위한 사용자의 요구로, 웹 서버는 해당 웹 페이지와 관련된 여러 파일 등에 접근하게 된다. 따라서 사용자가 요청하는 특정 웹 페이지뿐만 아니라 해당 웹 페이지와 관련된 이미지 파일, 이미지 데이터, 모든 연관 파일 등에 대한 정보가 로그 파일에 저장되는 것이다. 이러한 웹 로그 트래킹을 측정하는 단위는 히트(Hits), 페이지뷰(Page View), 체류 시간(Duration Time), 세션(Session), 방문자(Visitor) 등이 있는데 현재 페이지뷰를 측정단위로 가장 많이 이용하고 있다. 로그 분석 방법을 [그림 2]와 같은 방법으로 설계하고 사용자의 페이지뷰를 측정 단위로 사용한다.



[그림 2] 페이지뷰/히트/접속/방문

고객 분류		의 의	매출 기여도	
기존 고객	현재 고객	귀빈 고객	현재 거래가 활발한 고객	A
		철새 고객	잊어버릴 만하면 다시 오는 고객	A
		꼬리 고객	점점 발길이 줄어드는 고객	B
	과거 고객	휴면 고객	잠시, 겨울잠을 자는 고객	B
		이별 고객	뭔가 불만이 있어 가 버린 고객	C
		변심 고객	다른 상표가 좋다고 가 버린 고객	C
잠재 고객	신규 고객	희망 고객	앞으로 계속 올 것 같은 고객	A
		유보 고객	언제가 꼭 올 것 같은 고객	B
		해성 고객	한두 번 왔지만, 다시는 안 올 고객	C
	미래 고객	미지 고객	몰라서 안 오는 고객	B
		오해 고객	잘못 알아서 안 오는 고객	B
		주저 고객	알지만, 망설이며 안 오는 고객	B

[표 1] 고객의 분류

고객의 로그를 분석하여 고객을 분류를 하면 다.

[표 1]과 같이 12부류의 고객이 편성될 수 있다.

A등급 고객은 고객에게는 보다 적극적인 관심을 기울여야 한다. 개별적인 접촉과 알아보기 등이 이들을 고정 고객화하는 데 도움이 된다.

B등급 고객들은 잠재력이 충분한 고객들로서, 인내심을 가지고 꾸준하게 접근해야 한다. 고객의 반응을 성급하게 지대하지 말고, 직접 우편이나 광고·홍보를 통해 정보를 전달하는 등 동기 유발을 위한 끊임없는 자극이 필요하다.

C급 고객들을 붙잡기는 쉽지 않지만 그들이 등을 돌린 이유를 탐색할 필요가 있다.

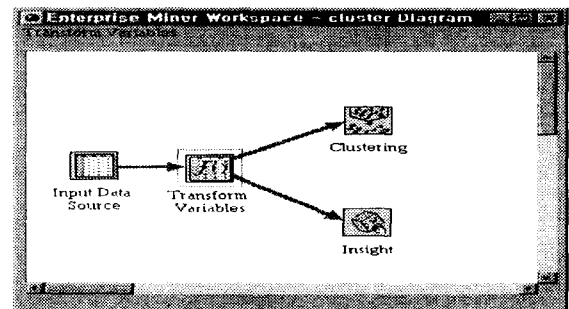
3. 고객 집단 편성 최적화 알고리즘

본 논문은 고객의 로그를 분석하여 분류된 고객의 유형을 데이터 마이닝 기법을 적용하여 가공함으로써 집단 편성을 최적화 할 수 있음을 보였다. [표 1]에서 A, B, C 등급으로 분류된 고객의 특성을 파악하여 최적화된 집단을 다시 편성한다.

STEP 1. Transformation Variable node를 사용해 새로운 분석변수를 생성을 먼저 수행한다. [그림 3]처럼 Input Data Source Node는 Transform Variable node와 연결되어 있고 Transform Variable node는 Insight node와 Clustering node와 연결되어 있다.

STEP 2. 자료의 선택 및 설정한다.

- ① Input Data Source node를 연다.
- ② DUNGAREE 데이터 셋을 선택한다.
- ③ STORID를 ID변수로 설정한다.
- ④ Input Data Source node를 저장하고 닫는다.



[그림 3] Clustering Diagram

이러한 형태의 판매 데이터를 군집화하는 하나의 접근방법은 제품 분류와 제품 판매량을 독립으로 간주하는 방법이다.

Enterprise Miner의 Clustering node는 k-means 군집분석 방법을 사용하고 있다. 많은 군집화 방법(k-means 군집분석 포함)이 입력변수(판매량)와 case(지점)과의 거리를 측정하여 분석의 도구로 사용하고 있다. 거리에 기초하는 방법은 입력변수의 scale에 매우 민감하다. 그러므로 자료에 대한 적절한 변환이 더 좋은 결과를 보이는 경우가 흔히 나타난다.

STEP 3. 군집화를 위한 새로운 변수의 생성

- ① Transform Variable node를 연다.
- ② Action → Create Variable을 선택한다.
- ③ Name Field에 FA_RATIO라고 입력한다.
- ④ Define을 선택한다.
- ⑤ 수식을 정의하는 부분에 수식을 입력한다. 새로 생성된 FA_RATIO 변수는 각 지점의

original jean에 대한 fashion jean의 상대적 scale을 나타내며 이 예에서 log 변환을 한 이유는 분석에 보다 적절한 scale이라고 간주되었기 때문이다.

위의 절차를 반복하여 다음과 같은 변수를 생성한다.

- LE_RATIO=LOG(LEISURE/ORIGINAL)
- ST_RATIO=LOG(STRECH/ORIGINAL)
- SALESTOT=FASHION + LEISURE + STRETCH + ORIGINAL

[그림 4]는 새로운 변수들로 얻어낸 결과이다.

Name	Factor	Mean	Std. Dev.	Skew.	Kurtosis	CV
SALESTOT	Yes	4302.107	366.2132	-3.77	15.62	0
ST_RATIO	Yes	-1.63959	0.995919	-2.74	8.66	-0
LE_RATIO	Yes	0.030369	0.299889	0.08	-0.19	0
PA_RATIO	Yes	-3.07575	0.806011	-3.10	18.80	-0
FASHION	Yes	92.4627	32.61626	0.44	0.72	0
LEISURE	Yes	1916.419	350.84	-0.17	0.61	0
STRETCH	Yes	444.0406	111.6901	0.35	0.60	0
ORIGINAL	Yes	1049.385	291.608	-0.20	0.19	0

[그림 4] 새로운 변수들이 생성된 결과

4. 결론

본 연구에서는 고객 패턴을 웹 마이닝을 통하여 고객 정보 추출을 최적화하는 알고리즘을 사용하여 얻어낸 최적 고객 분류를 다시 한번 가공하여 최적화된 집단 편성을 할 수 있음을 보였다. 이를 통해 온라인과 오프라인을 이용하여 고객분류를 자동화 할 수 있고 기업의 의사결정에 많은 영향을 미칠 것으로 사료된다. 제시한 알고리즘을 통해 기업은 고객형태를 쉽게 분류할 수 있을 것이며 고객을 관리하는 지침이 되고 기업 간 고객 관계에 많은 영향을 줄 것으로 사료된다. 그러나 집단 편성을 최적화할 수 있다 하더라도 [표 1]의 C등급에 해당하는 고객들은 기업의 노력에도 불구하고 A등급 또는 B등급으로 전향이 어려울 것으로 예상된다. 하지만 C등급 고객들의 형태와 패턴을 파악함으로써 A, B등급 고객들을 C등급 고객으로 전향하지 않을 수 있을 것이다.

기업은 이러한 고객 집단에 대한 적절한 운용과 각 개인의 성향에 따른 방침을 더욱 강화하여 고객 중심의 경제 활동을 도모할 수 있을 것이

참고문헌

- [1] 코리아 인터넷마케팅센터, <http://www.webpro.co.kr>
- [2] Accrue software Inc, "Web Mining Whitepaer", Accrue White Paper, 2000
- [3] 삼성SDS IT Review, "효율적인 로그분석을 활용한 eCRM 마케팅 전략", 2002
- [4] 손소영, 신형원, "범주형 자료에 대한 데이터 마이닝 분류기법 성능 비교", IE Interfaces Vol. 12, No. 4, pp. 551-556, December 1999
- [5] ㈜웹로그, "웹로그 사용자 설명서", 2001
- [6] ㈜웹로그, "Log Cluster 사용자 설명서", 2001