

데이터 마이닝을 이용한 eCRM 시스템의 피드백 시기 결정

홍정연⁰ 박승수
이화여자대학교 컴퓨터학과
(jungyun, sspark)@ewha.ac.kr

Determining Proper Feedback Time in eCRM System using Data Mining

Jung Yon Hong⁰ Seung Soo Park
Dept. of Computer Science & Engineering, Ewha Womans University

요 약

고객 관리를 위한 CRM 시스템이 인터넷 환경에서 eCRM 시스템으로 발전하게 됨에 따라 보다 효과적이고 인터넷 환경에 맞는 eCRM 시스템을 개발하는 것에 대한 관심이 높아지고 있다. 기존의 CRM 시스템은 고객을 분석하고 분석한 결과를 일괄적으로 캠페인에 적용하고 또한 한꺼번에 캠페인 결과를 피드백하여 분석함으로써 다음 캠페인에 이용할 수 있도록 하고 있다. 그러나 eCRM 시스템은 24시간 캠페인을 가능하게 하였으며 그때 그때 수집되는 데이터를 모아 일정한 시간을 주기로 하여 피드백하여 결과에 반영하고 있다. 보다 개인화(personalization)된 eCRM을 설계하고 이를 효과적으로 이용하기 위해서는 좀더 지능적인 피드백과 분석이 필요하다.

이에 본 연구에서는 데이터 마이닝을 이용하여 단순히 일정 시간주기가 아닌 의미있는 데이터량을 정함으로써 그 데이터량이 되었을 때 피드백을 수행하도록 하고 있다. 그럼으로써 개인화를 수행하는데 있어서 좀더 적시에 정확한 정보를 추천할 수 있도록 제안하고자 한다.

1. 서 론

CRM이 널리 소개되고 그 결과가 큰 이익을 창출하는 것이 기업들에게 알려지면서 고객관계에 대한 인식이 달라지고 CRM에 대한 관심이 높아지고 있다. 그리고 인터넷의 일반화는 고객과 더욱 가까워 질 수 있는 기회가 되었으며 고객과의 접촉, 고객의 요구 및 만족 등 고객과 관련된 모든 서비스를 효과적으로 제공하기 위한 필요성이 절실했다.[1]

인터넷의 발달은 CRM을 eCRM으로 발전하게 하였으며 고객과의 주요 접점이 이메일과 전자 게시판이 되었다. 이것은 캠페인과 캠페인 후 관리를 보다 쉽게 할 수 있도록 하였으며 한꺼번에 피드백을 받는 일반적인 CRM의 캠페인 관리와는 달리 캠페인이 24시간 이루어지고 그 결과를 빠르게 분석해야 되었다. 따라서 캠페인 결과를 적절하게 피드백하여 반영하기 위한 방법들이 제시되어야 한다. 예를 들어 너무 자주 피드백을 해주면 시스템에 오버헤드가 발생하고 너무 오랫동안 하지 않았을 경우에는 효과적인 캠페인이 이루어질 수 없게 된다.

또한 eCRM을 통해 더욱 그 중요성이 대두된 개인화를 효과적으로 이용해 direct-selling과 교차판매(cross-selling)를 높이기 위해서는 기존의 일정한 시간주기를 두는

피드백 시기 보다 좀더 지능적인 피드백 시기를 고려해 보는 것이 필요하다.

본 논문에서는 웹 로그와 고객의 인구 통계학적 데이터를 가지고 24시간 운영되는 웹 페이지에서 실시간으로 얻어지는 고객의 구매 데이터를 캠페인 결과에 반영하기 위하여 기존에 사용하고 있는 한 달 또는 한 주를 주기로 하는 피드백 시기 대신에 데이터 마이닝을 이용하여 고객변화의 내용을 효과적으로 반영할 수 있는 피드백 데이터 양을 찾아냄으로써 시스템 오버헤드를 줄이고 고객에게 적시에 정확한 정보를 전달하고 추천하여 더 많은 이익을 창출할 수 있도록 제한하고자 한다.

본 논문의 구성은 2장에서는 eCRM과 개인화에 대해 알아보고, 3장에서는 시스템 구성과 의미 있는 피드백 데이터량을 찾기 위한 데이터 마이닝 과정에 대해 설명하였다. 4장에서는 실험을 통한 분석과정과 실험 결과를 분석하였고, 5장에서는 결론 및 향후 연구에 대해 논의한다.

2. 관련 연구

2.1 eCRM

인터넷 기반의 온라인 CRM을 eCRM이라고 한다. CRM은 경쟁사와 차별화되는 고객에 대한 광범위하고

* 본 연구는 BK21(train korea 21)의 지원에 의해서 수행되었음

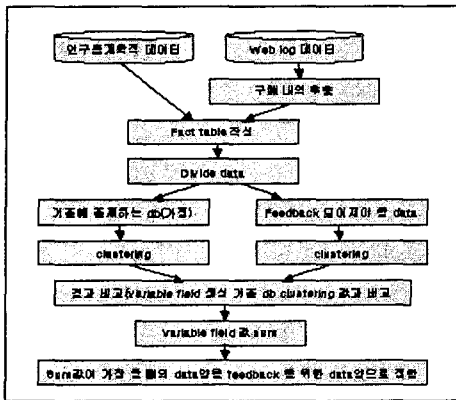
심층적인 이해를 바탕으로 고객의 개별적 요구를 충족할 수 있는 차별적 제품과 서비스를 제공함으로써 신규고객을 확보하고 기존고객과의 관계를 지속적으로 강화해 나가는 통합적이고 전사적인 마케팅 시스템이다[2].

고객에 대한 이해와 접근방식은 오프라인 CRM과 동일하나, 고객정보획득 및 커뮤니케이션 방법에는 차이가 있다[3]. 즉, eCRM이란 e-Business환경에서 적용되는 CRM을 의미한다. 따라서 eCRM은 기존의 오프라인 CRM과는 달리 인터넷을 통해 고객 데이터를 수집하고, 고객과 커뮤니케이션 할 수 있다는 특징이 있으며[5], 실시간 반응(real-time reaction), 실시간 가격 책정(real-time pricing)등을 CRM에 도입할 수 있다는 장점이 있다[6].

2.2 개인화(personalization)

개인화란 일반적으로 인터넷 비즈니스 상의 모든 상품과 광고, 메뉴 등의 콘텐츠를 고객 개개인의 성향과 취향에 맞도록 웹 사이트를 개발하여 고객에게 필요한 정보를 제공하며 획일적인 서비스가 아닌 고객과 친근한 일대일 관계를 맺기 위한 웹 전략이며 웹 기술이다[4].

3. 시스템 구성



[그림 1] 시스템 구성도

3.1 데이터 집합

데이터는 SPSS사에서 제공하는 샘플 데이터로 인구통계학적 데이터와 웹 로그 데이터이다. 인구통계학적 데이터는 이 쇼핑몰에 가입된 회원으로서 회원의 신상에 관한 정보가 저장되어 있다.

웹 로그 데이터는 어떤 회원이 로그인 한 상태에서 웹 서핑을 하고 상품을 구매했는지에 대한 정보가 저장되어 있다.

3.2 전처리 과정

웹 로그 데이터로부터 상품을 구매한 경우에 한하여 회원 아이디와 상품구매 정보를 추출하고 이 정보와 인구통계학적 데이터, 상품 카테고리 정보 등을 이용하여 fact table을 만들고 구매한 시간 순서로 정렬한다.

3.3 데이터 양에 따른 군집 변화 정도 측정

fact table을 기존에 존재하는 데이터 베이스(실험에 이용하기 위해 가정)와 피드백 되어져야 할 데이터로 나눈다. (여기서 피드백 데이터는 이미 저장된 데이터를 사용하지만 웹 상에서 시간 순으로 꾸준히 저장되고 있다고 가정한다.)

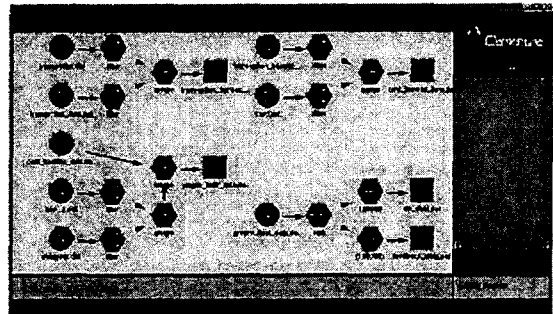
- (1) 기존에 존재하는 데이터 베이스를 클러스터링 하여 구매경향이 서로 비슷한 고객을 군집화 한다.
- (2) 일정시간 순으로 저장되고 있다고 가정한 피드백 할 데이터의 양을 피드백 할 데이터 전체의 $1/n$ 으로 정하고 기존의 데이터베이스와 합친 후 클러스터링 한다.($n=1,2,3,\dots$), (일정 데이터 양이 모였을 때 피드백을 위한 데이터 마이닝 수행 의미).
- (3) (1)과 (2)를 비교하여 군집에 변화가 있는 데이터 레코드에 대해 variable값을 1로 setting한다. 그리고 데이터 양이 $i/n(i=1,2,3,\dots)$ 의 경우에 대한 variable sum 값을 구한다.
- (4) n 값을 달리하여 데이터 양에 변화를 주면서 이와 같은 과정을 반복한다.

3.4 피드백을 위한 데이터 양 결정

위와 같은 과정을 반복하여 실험하여 n 값에 따른 variable sum값을 구하고 이것을 그래프로 그린다. 그래프의 증가량이 두드러지거나 값이 증가하다가 감소하는 부분의 데이터량을 피드백을 위한 데이터 양으로 결정한다.

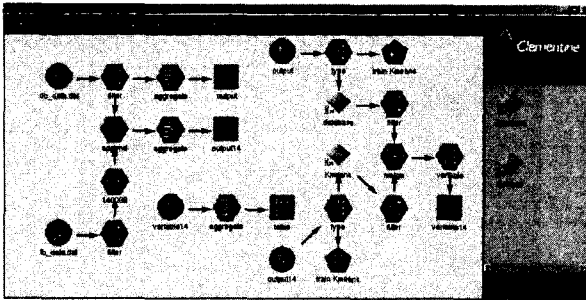
4. 실험

본 논문에서는 실험을 위해 SPSS사에서 제공하는 샘플 웹 로그 데이터와 인구 통계학적 데이터, 상품 카테고리 정보를 사용하였으며, 로그 파일의 전처리 과정과 마이닝 과정을 위하여 SPSS사의 데이터 마이닝 툴인 클레멘타인[7][8]을 사용하였다.



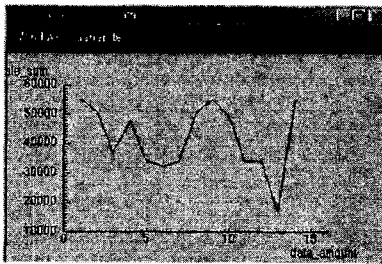
[그림 2] 전처리 과정을 위한 작업 화면

[그림 2]는 전처리 과정을 위한 작업으로 웹 로그 데이터로부터 추출한 상품 구매 정보와 인구통계학적 데이터, 그리고 상품 카테고리 정보를 이용하여 fact table을 만들었다.



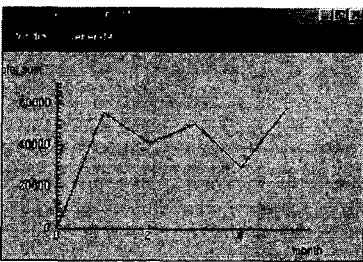
[그림 3] 효과적인 feedback 데이터 양을 찾기 위한 작업 화면

[그림 3]은 효과적인 피드백 데이터 양을 찾기 위해 fact table로부터 클러스터링을 이용하여 variable이라는 field를 생성하고 그 sum값을 구하여 그 변화 추이를 살피는 실험을 나타낸 것이다. 데이터 양을 달리하여 실험 할 때마다 sum값을 구하는데 이것을 variable sum 값으로 나타낸다. 여기서 variable sum값은 클러스터에 변화가 일어난 레코드 수의 합을 나타낸다.



[그림 4] variable sum 값의 변화 그래프

[그림 4]는 실험결과로 얻어진 variable sum값의 변화를 그래프로 표시한 것이다. 이 경우 변화 곡선이 상승하다 감소하는 지점을 효과적인 feedback 데이터 양으로 정할 수 있다.



[그림 5] 피드백 주기를 한달 주기로 했을때의 variable sum 그래프

[그림 5]는 한 달을 주기로 했을 때의 variable sum 그래프로서 같은 데이터를 가지고 의미있는 데이터 양을 주기로 했을 때와 비교분석하기 위해 실험한 것이다. 이 데이터는 2001년 8월부터 12월 까지 5달의 데이터로 한 달을 주기로 했을 때는 5번의 마이닝이 수행되고 의미있는 데이터 양을 주기로 했을 때는 [그림 4]에서 보는 바와 같

이 상승 곡선이 최고점인 3곳에서 마이닝이 수행된다. 두 결과를 비교하면 마이닝을 수행할 때마다 업 데이터 되는 레코드 양이 데이터 양을 주기로 했을 때보다 더 많고, 또한 한 달을 주기로 했을 때는 그 달의 사용자의 구매 성향에 따라 업 데이터 되는 데이터 양의 변화가 심한 것을 볼 수 있는 데 이것은 효과적이지 않다.

따라서 효과적인 피드백 데이터 양은 한 번의 피드백에 의한 결과 반영이 시간을 주기로 하는 피드백에 의한 결과 반영에 비해 더 많은 변화를 반영할 수 있다. 이것은 결과 반영을 위해 보다 적은 데이터 마이닝 분석 과정을 수행한다는 것을 의미한다. 따라서 시스템 오버헤드를 줄일 수 있다는 장점이 있다. 또한 보다 정확한 고객 분석을 통하여 적절하게 웹 페이지 구성을 한다거나, 고객에게 맞는 상품 추천을 할 수 있다. 이것은 24시간 운영되는 쇼핑몰의 효과적인 운영을 위해 적은 시스템 오버헤드로 적시에 고객에게 알맞은 상품 추천이 가능하다는 것이다.

5. 결론 및 향후 연구

본 논문에서는 쇼핑몰의 샘플 데이터를 이용하여 eCRM 시스템을 위한 피드백 시기 결정 연구를 하였다. 이것은 기존에 시간을 주기로 하는 피드백 시기 결정과는 다르게 데이터 양을 이용하여 피드백 시기 결정을 위한 또 다른 방법을 제안 하고 있다.

데이터 양을 이용한 피드백시기 결정은 데이터 마이닝을 이용하여 효과적인 데이터 양을 찾아냄으로서 보다 적은 데이터 마이닝 분석 수행으로 보다 효과적인 시스템 운영과 고객에게 적합한 상품을 추천함으로써 이익을 높일 수 있다는 것에 착안한 것이다.

향후 과제로는 전체 데이터 베이스와 피드백 데이터 간의 양적인 비율에 관한 연관성을 분석해 보고 논문을 쓸 계획이다.

6. 참고문헌

- [1] Chambers, B., and J. Fenner, "Complete CRM Integrating and Automating Process", *Imaging and Document Solutions*, Feb. 2000.
- [2] Berry, M. J. and G. Linoff, *Master Data Mining: The Art and Science of Customer Relationship Management*, John Wiley & Sons, 2000.
- [3] Brown, S. A., *Customer Relationship Management*, John Wiley & Sons, 2000.
- [4] Mobasher, B., R. Cooley, and H. Srivastava. Automatic Personalization Based on Web Usage Mining. *ACM*, Vol. 43(8). PP142-151, August 2000
- [5] Elsenpeter, R. C. and Velte, T.J., *e-Business A Beginner's Guide*, McGraw-Hill, Berkeley 2001
- [6] Harmon, P., M. Rosen, and M. Guttman, *Developing e-Business Systems and Architectures: A Manager's Guide*, Morgan Kaufmann Publishers, San Francisco, 2001.
- [7] Khabaza, T., and D. Sigerson. WebCAT: the Clementine Application Template for WebMining and Analytical eCRM, web-mining workshop paper, 1st SIAM International Conference on Data Mining, Chicago, April 7, 2001
- [8] Clementine User Guide Version 5