

웹 개인화를 통한 자동화된 뉴스레터 시스템

*김계숙, *박우수, **권오현, *박규석
*경남대학교 컴퓨터공학과, **동명정보대학교 컴퓨터공학과

Automatic Newsletter System with Web Personalization

*Gye-Sook Kim, *Woo-Su Park, **Oh-Hyun Kwon, *Kyoo-Seok Park
*Dept. of Computer Engineering, Kyungnam University
**Dept. of Computer Engineering, Dongmyong University of Infor. Tech.
E-mail : {edy100,bgbest}@unicath.net, trats@sarang.net, kspark@kyungnam.ac.kr

요약

본 논문에서는 웹 데이터 마이닝을 통하여 웹 사이트를 방문한 사용자의 콘텐츠 유형에 따른 정보를 조사하고, 필터링 과정을 통해 분류화하고, 이러한 과정을 통해 얻은 정보를 이용하여 뉴스레터를 발송하며, 발송된 뉴스레터로부터의 콘텐츠 유형에 따른 CTR(Click Through Rate)과 사용자 반응을 추적하여 이러한 정보를 분석하고 사용자 프로파일 및 웹 사이트로부터 분류화된 정보, 그리고 추적된 정보와 함께 뉴스레터 콘텐츠를 재구성하는 개인화된 자동화 뉴스레터 시스템을 설계하고 구현한다.

1. 서론

인터넷의 발달과 함께 이메일의 사용이 증가하게 되고, 이를 바탕으로 한 전자상거래 사업이 발전 일로에 있다[4]. 따라서, 기업이 고객을 만나는 중요한 채널은 웹사이트이며, 인터넷을 통해 적은 비용으로 이메일 마케팅을 할 수 있다.

이메일 마케팅은 eCRM의 중요한 부분을 차지하고 있으며, 이메일 마케팅 분야에 데이터마이닝 기술을 적용함으로써 효과적인 고객관리를 할 수 있다.

기존의 뉴스레터는 문서들이 지닌 의미와 문서간의 상호작용을 고려하여 뉴스레터를 구현하였지만, 동적으로 변하는 사용자 개개인의 취향을 반영하기에는 어려운 점이 많다. 따라서, 데이터 마이닝 기술을 통하여 개개인에게 맞는 뉴스레터 시스템을 구현함으로써 사용자들에게 보다 친숙하고 유용한 정보를 제공할 수 있다.

본 논문에서 제안하는 시스템은 각 사용자의 콘텐츠 유형에 따라서 사용자의 URL을 필터링 한 후 분류화(classification) 과정을 거쳐서, 뉴스레터를 통하여 획득할 수 있는 정보와 통합하여 다양한 정보를 추출 할 수 있다. 이렇게 추출된 정보를 이용하여 웹

사이트를 통한 사용자의 개인화된 정보와 뉴스레터를 통한 사용자의 CTR을 비교하여 뉴스레터를 통한 효과를 분석할 수 있으며, 이를 이용하여 뉴스레터 내용을 재구성하고 개인화된 뉴스레터 시스템을 통하여 사용자에게 보다 긍정적이고 효과적인 시스템을 설계하는 방법을 제시한다.

2. 관련연구

웹을 통하여 사용자의 탐색행동을 추적 할 수 있기 때문에 사용자의 취향을 알 수가 있어, 사용자가 원하는 정보를 제공할 수 있다. 일반적으로 웹 데이터마이닝은 웹 콘텐츠 마이닝(Web Contents Mining), 웹 구조적 마이닝(Web Structural Mining), 웹 유시지 마이닝(Web Usage Mining)으로 구분할 수 있다. 여기서는 사용자들의 접속 유형을 발견하는 것을 목적으로 하는 Web Usage Mining을 이용한다. 웹 마이닝 단계는 전처리 과정, 패턴 탐색, 패턴분석으로 구분된다. 전처리 과정 단계에서는 획득된 데이터를 분석에 적합한 형태로 정제하며, 이로 인해 원천 데이터에 비해 용량이 감소하게 된다. 패턴탐색 과정으로는 기술적 통계분석 방법, 연관 분석, 클러스터링, 분류화, 순

차적 분석 등의 기법이 있다. 마지막, 패턴분석 과정에서는 웹 데이터 마이닝을 이용하여 분석된 정보를 사용자들이 보다 더 쉽게 이해할 수 있도록 통계, 시각화, 사용성 분석, 데이터베이스 쿼링 등을 이용한다 [2].

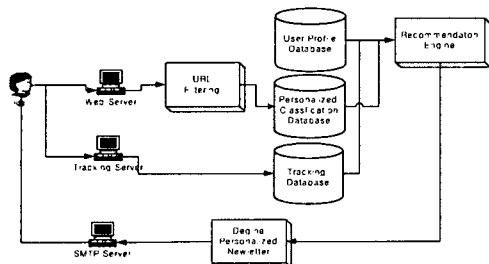
웹 유지지 마이닝에 적용 가능한 기술로는 탐사하고자하는 지식의 형태에 따라 분류화, 클러스터링, 연관규칙(association rule), 순차패턴(sequential pattern) 탐사 등이 있다.

위와 같은 분류에서처럼 각 사이트의 특성에 따라서 사용하는 기법이 다양하다. Yahoo에서는 personal My Yahoo 서비스, 신문사에서는 filtering & rating 기법을 이용하여 개인화된 콘텐츠를 제공하고 있으며, Alexa에서는 고객의 방문 빈도에 따른 사이트를 추천하는 서비스를 시행하고 있다. 그리고 순차 패턴 탐사를 통하여 웹을 재구성하는 방법[6]도 있다.

3 개인화된 자동화 뉴스레터 시스템

3.1 개인화를 위한 스코어링

본 논문에서 제시하는 시스템은 웹 사이트에 접근한 사용자들이 어떤 특정 콘텐츠에 대해 관심을 가지고 있는지를 분류 하고 이것을 토대로 하여, 자동화 뉴스레터 시스템을 구현하는데 목적이 있다. <그림 1>은 본 논문에서 제시한 전체적인 개인화된 자동화 뉴스레터 시스템의 구조도이다.



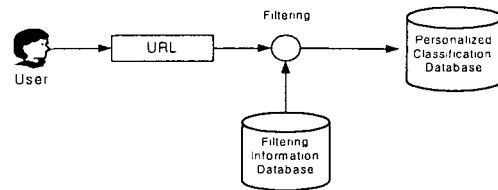
<그림 1> 개인화된 자동화 뉴스레터 시스템

본 논문에서는 웹 사이트로의 접근에 대한 깊은 통찰력은 제공하지 못하지만, 가장 자주 접근되는 데이터 즉, 사용자의 콘텐츠 유형에 따라 URL을 필터링하고 분류화 한다.

3.1.1 웹 사이트에서의 스코어링

사용자의 콘텐츠 유형에 따라 정보를 획득하는 가장 간단한 방법은 Session이나 Cookie를 이용해서 서

버 파일 내에 정보를 얻기에 필요한 구문을 삽입하는 방법이 있으나, 여기서 제시하는 방법은 서버파일 내에 구문을 삽입하지 않고 사용자 URL을 이용해서 획득하는 방법을 이용한다. 이 방법의 이점은 URL을 처리하기 위해서 하나의 서버파일만 존재하면 되므로, 콘텐츠의 삽입, 수정 및 삭제 시 좋은 이점을 가져다 준다.



<그림 2> URL의 추출

웹 사이트를 통해 사용자로부터 접근되는 데이터가 모두 필요한 것은 아니다. 뉴스레터 시스템을 구현하기 위해서는 마케팅에 필요한 데이터로 정제할 필요가 있다. 사용자로부터 생성된 URL 데이터와 시스템을 구현하기 위해 <그림 2>와 같이 필터링 정보 데이터베이스(Filtering Information Database)로부터 필요한 데이터에 대해 정의된 데이터 목록과 비교하여 필요한 정보를 필터링 한 후 각 사용자에 대한 데이터 분류에 맞게끔 콘텐츠 필드 값을 증가 시킨다.

(표 1) 사용자 URL

사용자 ID	URL
1	/board.asp forumname=data_Caption
2	/movie/crying53.asx
3	/communication/view_board.asp no=15
1	/music_video/ly12.asx

(표 2) 필터링 목록

filtering list	data content
/board.asp?formunname=data_Caption	Caption
/link/	Link
/movie/	movie
/music_video/	music_video

(표 3) Personalized Classification Database

사용자 ID	Caption	Link	Movie	Music_video
1	21	0	12	20
2	32	1	9	11
3	43	2	4	4
4	22	1	5	45

(표 1)은 사용자로부터 요청되어진 URL 문자열이며, 사용자로부터 전달된 Request Method는 GET 방식이다. 그리고, 요청한 페이지 부분인 url부분과 query부분으로 나눌 수 있다.

(표 2)는 (표 1)에서 요청되어진 문자열을 필터링하기 위하여 정의된 콘텐츠 부분이며, 이러한 정의된 콘텐츠와 일치된 문자열이 있을 경우 (표 3)에서와 같이 스코어링을 한다.

3.1.2 뉴스레터로부터의 스코어링

웹 사이트를 통하여 사용자 취향에 대한 정보를 얻고, 개인화된 뉴스레터 시스템을 구현 할 수 있지만, 뉴스레터를 통하여 접근하는 사용자의 접근 패턴이 다를 수 있다. 그러므로, 정확한 개인화된 뉴스레터 시스템을 구성하기 위하여, 뉴스레터를 통한 사용자 접근 패턴을 알아 볼 필요가 있다. 본 논문에서는 사용자의 콘텐츠 유형에 따른 CTR에 관한 방법과 뉴스레터를 통하여 또 다른 정보를 얻는다.

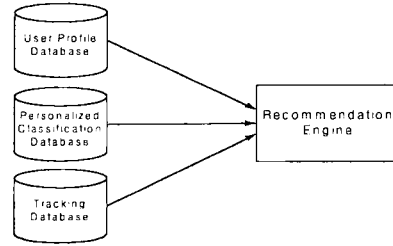
<그림 1>과 같은 방법으로 뉴스레터에 대한 개인화된 분류화 데이터베이스(Personalized Classification Database)를 구축한다. 이러한 방법을 통하여 사용자가 어떤 콘텐츠에 관심을 가지고 웹 사이트를 방문했는지 알 수 있고, 이를 통하여 개인화된 뉴스레터 시스템을 구성하는데 좋은 정보를 제공한다. 하지만 뉴스레터를 통한 또 다른 이점은 Html 구문 내에 다양한 태그를 삽입함으로써 캠페인에 참가한 고객의 반응을 세밀하게 추적하고, 수행된 마케팅의 효과를 분석할 수 있으며, 그 결과를 데이터베이스에 축적함으로써 보다 정교한 마케팅이 가능해 진다.

3.2 개인화된 뉴스레터 시스템

일반적으로 이메일 캠페인 설계시 캠페인 마케터가 어플리케이션 서버의 기능을 이용하여 설계 한다.

본 논문에서 제안하는 시스템은 일반적인 타겟 데이터 베이스[3]의 설계과정 없이, 웹 사이트 가입시 생성된 사용자 프로파일과 웹 사이트로부터 획득된

개인화된 데이터, 그리고 뉴스레터를 통하여 얻어진 트래킹 데이터를 이용하여 Recommendation 엔진을 이용한 사용자 취향에 맞는 콘텐츠를 구성할 수 있다.



<그림 3> Recommendation Engine

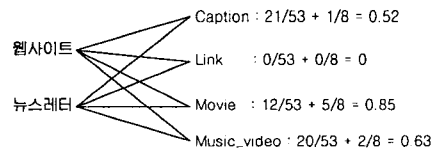
<그림 3>은 각 데이터베이스로부터 사용자에 대한 정보를 가져와서 개인에게 적합한 뉴스레터 콘텐츠 구조를 설계하며, (표 4)와 (표 5)로부터 각각의 콘텐츠에 대한 비율을 추정할 수 있다.

(표 4) 웹 사이트로부터의 PCD

사용자 ID	Caption	Link	Movie	Music_video
1	21	0	12	20
2	32	1	9	11
3	43	2	4	4
4	22	1	5	45

(표 5) 뉴스레터로부터의 PCD

사용자 ID	Caption	Link	Movie	Music_video
1	1	0	5	2
2	5	0	3	3
3	1	0	4	4
4	3	0	5	3



<그림 4> 사용자 ID가 1인 경우

<그림 4>에서와 같이 사용자 ID가 1인 경우는 각 콘텐츠의 값에 따라서, Movie -> Music_video -> Caption -> Link 순으로 나타나며, 이 값을 이용해서

뉴스레터의 구조를 사용자 관심도에 따라 자동적으로 구성할 수 있다. 그리고, 사용자 프로파일을 이용하여 뉴스레터 대상자를 선별할 수가 있다. 예로서, 나이가 20대인 사람이나, 여성과 같이 특정대상자에게 보낼 경우, 사용자 프로파일을 이용함으로써 대상자를 선별할 수 있다.

SMTP 서버를 통하여 개인화된 뉴스레터를 발송하며, 발송된 뉴스레터는 추적과정을 통하여 추가적인 정보를 얻을 수 있다.

메일 발송후 생기는 에러를 처리함으로써 개인에게 전달하는 뉴스레터의 수를 줄일 수 있기 때문에 SMTP 서버의 효율을 향상을 시킬 수가 있다.

한편, 뉴스레터 송신시 생기는 대부분의 에러는 다음과 같이 분류할 수 있다.

- ① domain이 존재 하지않는 경우,
- ② 사용자가 존재하지 않는 경우,
- ③ 연결이 되지 않는 경우,

①과 ②는 사용자가 메일주소를 잘못 적은 경우이므로 뉴스레터 송신후 사용자가 웹사이트를 방문할 때 이 부분을 다시 작성하는 폼(form)을 제시함으로써 수정할 수가 있다. ③의 경우는 송신을 여러번 시도후에도 계속 발생 할 경우로 ①과 ②의 경우와 같은 방법으로 해결을 한다.

뉴스레터를 발송함으로써 얻을 수 있는 정보는 다음과 같다.

- ① 발송한 총 메일
- ② 발송에 실패한 메일
- ③ 발송에 성공한 메일
- ④ 수신자가 열어본 메일
- ⑤ 뉴스레터를 통해 웹사이트를 방문한 수
- ⑥ 메일 수신을 거부한 수

위와 같은 정보를 이용하여 뉴스레터의 성공여부를 판별할 수 있는 자료를 얻을 수 있으며, 각 뉴스레터를 송신하기 전에 수신자를 판별하기 위한 ID와 token정보를 삽입함으로써 로그인 과정을 거치지 않고 개개인의 사용자 및 뉴스레터를 식별할 수 있다.

4. 결론 및 향후과제

본 논문에서는 웹 상의 사용자 행동을 분석하여 개인화된 DB를 구축하고, 개인화 추천 엔진을 이용하여

개인화된 뉴스레터를 생성, 발송 및 추적 후 개인화를 병행할 수 있는 자동화된 뉴스레터 시스템을 제시하였다. 본 논문에서 제시한 추천엔진 외에 피드백(feedback)된 정보를 이용하여 콘텐츠 분류별 가중치를 분석하고 피드백시키는 콘텐츠별 가중치 부여에 대한 연구, 웹 사이트 구조와 디자인이 개인화에 미치는 영향 등에 대한 연구가 필요하다.

[참고문헌]

- [1] 프레드릭 뉴웰, 삼성전자 글로벌마케팅연구소, 인터넷 시대의 고객관계경영 CRM.com, 21세기북스.
- [2] http://www.ecminer.com/m4_webDataMining.html
- [3] 김성민의 4명, 데이터마닝 기법을 활용한 이메일 마케팅 시스템의 구축, 데이터베이스 연구회지, 17권 1호, 2001.3
- [4] 김길평, 인터넷상에서 디렉트메일광고의 효과에 관한 연구, 한성대학교 논문집 1998. 17p~37p
- [5] 무빅스, <http://movics.net/>
- [6] 고경자, 김인철, 적응형 웹 사이트를 위한 웹 로그 마이닝, 한국정보처리학회 춘계 학술 논문집 제8권 제1호, 2001
- [7] Bamshard Mobasher, Robert Cooley, Jaideep Srivastava, Automatic Personalization Based on Web Usage Mining, <http://maya.cs.depaul.edu/~mobasher/personalization/index.html>
- [8] I-Yuan Lin, Xin-Mao Huang and Ming-Syan Chen, Capturing User Access Patterns in the Web for Data Mining, Proceedings of the 11th IEEE International conference, 1999
- [9] Bing Liu, Wynne Hus, Lai-Fun Mun and Hing-Yan Lee, Finding Interesting Patterns Using User Expectations, IEEE Transactions on Knowledge & Data Engineering, V.11 N.6, 1999