

사용자 정보를 이용한 이메일 자동 생성 시스템의 설계

박일호*

*(주)리테일테크

e-mail:ilho@nate.com

A Design of Automatic Generation System of E-mail using User Information

Il-Ho Park*

*Technical Research Institute, Retail Tech Co., LTD.

요 약

인터넷의 급속한 발전으로 매우 많은 정보와 다양한 콘텐츠가 개인 사용자들에게 제공되고 있다. 또한, 개인을 고객으로 하는 각종 포털 사이트들이 많이 생성됨에 따라 고객 개인을 위한 고객화된 정보가 매우 중요한 하나의 콘텐츠로 작용하고 있다. 하지만, 현재 국내외 전자상거래를 주목적으로 하는 포털 사이트의 시스템들은 고객에 관한 정보를 단순한 개인 프로파일로 활용하고 있을 뿐 고객 정보를 이용하여 더 많은 개인화된 새로운 정보를 창출하지 못하고 있다. 본 논문은 고객에 대한 정보를 세분화하고 분석하여 제3의 개인화 정보를 생성하여 자동으로 각 개별 고객에게 개인화된 정보를 제공해 줄 수 있는 시스템을 설계 및 구현하고자 한다. 고객의 이메일 오픈율과 마우스 이벤트 정보를 분석 및 계산하여 개별 고객에게 고객의 관심정보 및 관심 컴포넌트를 생성한 후 관심정보와 관심 컴포넌트를 이용하여 개별 고객의 관심 정보를 고객이 선호하는 이메일 규격 및 양식에 맞게 에이전트를 통해 자동으로 재구성하여 푸쉬해 주는 개인화 메일 자동 생성 에이전트 시스템을 개발하고자 함이 본 논문의 목적이다.

1. 서론

최근 전자상거래의 발전은 인터넷의 대중화를 통하여 가속화되고 있으며 이로 인하여 전자상거래에서의 고객 유지 및 관리에 대해 많은 연구가 활발히 진행되고 있다.

고객과 기업이 지속적인 커뮤니케이션을 통해 고객 행동을 이해하여 전자적으로 더욱 효율적인 고객관리를 통하여 기업의 브랜드와 상품 및 서비스에 만족할 수 있도록 경영적 측면에서의 접근 중 하나가 전자적 고객 관계 관리이다[1].

이러한 e-CRM(electronic-Customer Relationship Management)에서의 중요한 고객관리 도구 중의 하나가 전자우편(e-mail)이다. 최근 각 기업의 웹사이트에서는 e-CRM 시스템을 도입하여 잠재 고객 파악, 신규 고객 획득, 기존 고객 유지 및 고객 수익성 증대를 위해 여러 가지 서비스들을 e-mail로 제공하고 있다. 따라서 e-CRM에서의 가장 효율적이고 직접적인 고

객관리 도구가 e-mail이라 할 수 있다.

개별 고객마다의 잠재적 관심정보를 정확히 파악하여 관심정보를 제공하게 되면 고객의 서비스에 대한 만족도를 더욱 높일 수 있을 뿐만 아니라 고객에 대한 효율적인 서비스 관리가 지속적으로 유지될 수 있다. 따라서 본 논문은 고객의 이메일 오픈율과 마우스 이벤트 정보를 분석 및 계산하여 개별 고객에게 고객의 관심정보 및 관심 컴포넌트를 생성한 후 관심정보와 관심 컴포넌트를 이용하여 개별 고객의 관심 정보를 고객이 선호하는 이메일 규격 및 양식에 맞게 에이전트를 통해 자동으로 재구성하여 푸쉬해 주는 개인화 메일 자동 생성 에이전트 시스템을 설계하고 개발하고자 함이 본 논문의 목적이다.

2. 관련연구

2.1 Personal WebWatcher

카네기 멜론 대학에서 구축한 Personal WebWatcher

는 웹 브라우저 상위에서 실행하면서 사용자의 행동을 모니터링하여, 개인 사용자에게 적응성을 더욱 부여한 에이전트이다[2].

Personal WebWatcher는 사용자의 컴퓨터 내부에서 백그라운드로 실행되면서, 사용자가 보는 웹 브라우저와 사용자의 행동을 관측하고, 이를 기반으로 사용자의 관심 문서를 모은다. 이렇게 모아진 관심 문서들을 분석함으로써 사용자의 관심파일을 만들고, 사용자가 정보 검색 시에 이 관심파일을 이용하여 사용자의 관심 문서를 예측하여 제안한다.

2.2 WiseWire

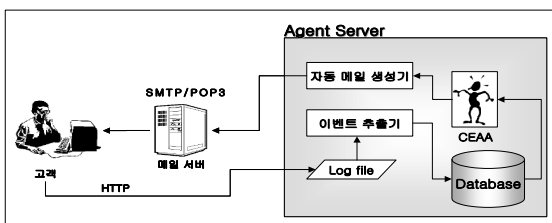
WiseWire는 WiseWire사에서 만든 지능형 에이전트로, 인터넷에서 정보를 검색할 때, 사용자의 기존 검색 패턴을 분석하여 얻은 프로파일을 이용하여, 사용자들에게 각 개인이 원하는 정보를 보다 쉽게 검색할 수 있도록 하는 기능을 제공한다[4].

WiseWire는 관심 정보와 비 관심 정보를 기반으로 사용자의 관심(Profile)을 기계학습을 이용하여 학습하고, 이렇게 학습된 사용자의 관심을 이용하여 사용자 적응력 있는 검색을 수행한다.

3. 고객화 이메일 자동 생성 에이전트

3.1 시스템의 구조

본 논문에서 제안하는 고객화 이메일 자동 생성 에이전트 시스템인 CEAA(Customizing E-mail Auto-generation Agent)는 고객의 잠재적 관심정보를 추출하여 고객이 선호하는 이메일 양식을 자동으로 생성하여 관심정보를 제공해주는 시스템이다. 그림 1은 CEAA의 전체적인 시스템 구조를 보이고 있다.



[그림 1] CEAA의 시스템 구조도

고객은 이메일을 통한 에이전트와의 상호작용을 통해 자신의 잠재적 관심정보를 에이전트에게 전달하게 된다. 따라서, 고객은 자신도 미처 인식하지 못하는 관심정보를 이메일을 받고 확인하는 동안 에이전트

가 추출하게 된다.

고객이 클라이언트로 CEAA로부터 이메일을 받고 확인하는 과정에서 고객의 잠재적 관심정보가 로그 파일로 CEAA의 데이터베이스에 저장되며 저장된 로그 파일의 정보가 CEAA의 이벤트 추출기에 의해 관심정보 컴포넌트를 추출한다. 자동 메일 생성기는 추출된 관심 정보 컴포넌트를 이용하여 고객화된 관심정보를 이메일로 생성하게 된다.

에이전트 서버는 클라이언트인 고객이 이메일을 통해 자신의 관심 정보에 대한 이벤트를 발생시킬 때마다 로그 파일 형태로 전달된 이벤트 정보에서 관심정보를 추출하여 데이터베이스에 저장하고 저장된 고객의 관심정보를 기반으로 고객화된 이메일을 생성하게 된다. 에이전트 서버는 CEAA와 이벤트 추출기, 자동 메일생성기 및 데이터베이스로 구성되어 있다.

3.2 개인화 메일 생성 알고리즘

제안하는 개인화 메일 자동 생성 에이전트 시스템은 관심 컴포넌트가 크게 텍스트와 이미지로 분류된다. 따라서 잠재적 관심 컴포넌트의 속성을 각각 추출하여 이를 관심 정보로 이용해야 하므로 모든 관심 컴포넌트에 대한 속성을 변수 값으로 정의한다.

가) 주 관심 정보 추출 알고리즘

I_d : 세부 정보 오픈에 따른 관심도

W_i : 세부관심정보 우선순위 가중치

$$I_d = \frac{1}{W_i} \quad (\text{단, } W_i \text{는 처음 클릭한 타이틀(1)부터 마지막 클릭한 타이틀(n)까지의 가중치})$$

나) 관심 속성 측정 알고리즘

고객이 개인화 이메일 자동생성 에이전트로부터 메일을 받았을 때 자신의 관심정보를 마우스로 클릭하게 되면 텍스트와 이미지의 속성값을 개인화 이메일 자동생성 에이전트 서버에서 받아 관심도 측정을 하게된다. 마우스 클릭 이벤트에서 추출할 수 있는 속성에 대한 변수는 표 1과 같이 정의한다.

[표 1] 마우스 클릭 이벤트 관심 컴포넌트 속성 변수

속성 값	색상	크기	폰트	모양
컴포넌트 종류				
텍스트 카운트 변수	Cx_{tc}	Cx_{ts}	Cx_{tf}	
이미지 카운트 변수		Co_{ts}		Co_{tf}

즉, Cx_{tc} 는 사용자가 특정한 색상을 갖는 텍스트를 클릭하였을 때마다 1씩 증가시키는 카운트 변수이며 x 값은 색상의 RGB 코드가 된다. Cx_{ts} 는 특정한 크기를 갖는 텍스트를 클릭하였을 때의 카운트 변수이며 x 값은 텍스트의 크기값이 된다. Cx_{tf} 는 폰트에 대한 카운트 변수이며 x 값은 해당 폰트명이 된다.

텍스트 카운트 변수들을 이용하여 사용자가 클릭한 텍스트에 대한 관심도를 다음과 같은 식으로 계산할 수 있다.

Ix_{tc} : 클릭한 텍스트의 특정 색상에 대한 관심도

Ix_{ts} : 클릭 텍스트의 특정 크기에 대한 관심도

Ix_{tf} : 클릭 텍스트의 특정 폰트에 대한 관심도

$$Ix_{tc} : \frac{Cx_{tc}}{\sum_{i=1}^n Ci_{tc}} \text{ (단, } i \text{는 메일양식에 사용된 텍스트 색상의 갯수) (식1)}$$

$$Ix_{ts} : \frac{Cx_{ts}}{\sum_{i=1}^n Ci_{ts}} \text{ (단, } i \text{는 메일양식에 사용된 텍스트 크기의 갯수) (식2)}$$

$$Ix_{tf} : \frac{Cx_{tf}}{\sum_{i=1}^n Ci_{tf}} \text{ (단, } i \text{는 메일양식에 사용된 텍스트 폰트의 갯수) (식3)}$$

이미지 카운트 변수들을 이용하여 사용자가 클릭한 이미지에 대한 관심도를 다음과 같은 식으로 계산할 수 있다.

Iy_{is} : 클릭 이미지의 특정 크기에 대한 관심도

Iy_{if} : 클릭 이미지의 특정 모양에 대한 관심도

$$Iy_{is} : \frac{Co_{Is}}{\sum_{i=1}^n Co_{Is}} \text{ (단, } i \text{는 메일양식에 사용된 이미지 색상의 갯수) (식4)}$$

$$Iy_{if} : \frac{Co_{If}}{\sum_{i=1}^n Co_{If}} \text{ (단, } i \text{는 메일양식에 사용된 이미지 모양의 갯수) (식5)}$$

다) 관심 메일 추정 알고리즘

T_{Eo} : 세부정보 오픈 시간

T_{Ec} : 세부정보 닫은 시간

T_{Ei} : $T_{Ec} - T_{Eo}$: 메일 조회 시간

$Q_1 \leq T_{Ei} \leq Q_3$ (LogFile의 Q_1 :1사분위수 Q_3 :3사분위수)

$Max(T_{Ei}) \leq Cur(T_{Ei})$ 이면 $Cur(T_{Ei}) = 1$
 $Ave(T_{Ei}) \leq Cur(T_{Ei}) \leq Max(T_{Ei})$ 이면 $Cur(T_{Ei}) = 0.75$

$Min(T_{Ei}) \leq Cur(T_{Ei}) \leq Ave(T_{Ei})$ 이면 $Cur(T_{Ei}) = 0.5$

$Cur(T_{Ei}) \leq Min(T_{Ei})$ 이면 $Cur(T_{Ei}) = 0.25$

사용자의 세부정보조회 시간이 데이터베이스에 저장되어 있는 가장 큰 값보다 클 경우 현재의 세부정보조회에 가장 큰 가중치를 갖고, 평균과 비교하여 가중치의 비율을 점차적으로 낮게 하여 사용자의 관심도를 측정한다.

사용자의 의도하지 않은 행위로 인한 관심정보의 오류를 최소화하기 위하여 로그파일에 기록된 조회시간의 1사분위수와 3사분위수 사이의 시간을 관심정보로 활용하여 오류정보를 배제한다.

4. CEAA 시스템 구현

관심정보 추출을 위해 파일로 저장된 로그정보를 전체로그데이터에서 세부정보조회 시간의 일사분위수와 삼사분위수 사이의 유효조회 시간을 적용하여 유효한 관심데이터를 추출한다. 사분위수 버튼을 선택하면 총 로그파일에 저장된 데이터의 개수에서 일사분위수와 삼사분위수를 결정한다. 유효 데이터의 범위를 계산하고 추출버튼을 선택하여 개인의 관심정보를 관심 데이터베이스에 저장한다.

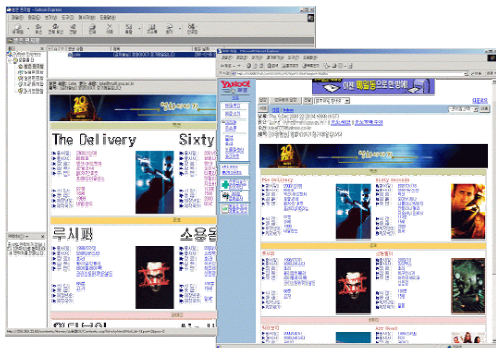
관심데이터베이스는 개인화 이메일 발송에 필요한 개인이 선호하는 장르 및 글꼴, 색상, 이미지와 텍스트의 위치결정에 필요한 각 개인의 정보를 갖고 있을 뿐만 아니라, 선호하는 제작국가, 감독 및 배우에 대한 정보도 함께 가지고 있어 다단계의 개인화 메일을 생성할 수 있게 로그정보를 가공하여 각 항목에 대한 개인화된 관심률을 가지고 있다.

그림 2는 개인화된 이메일의 생성화면으로 데이터베이스화된 개인의 로그정보를 기준으로 다중고객에게 발송될 메일내용과 메일양식을 생성하는 화면이다. 먼저 발송될 메일을 선택하면 표준메일형태의 HTML 문서가 텍스트로 화면의 중앙에 생성된다. 개인정보를 이용하여 메일 발송대상을 지정하면 우측하단에 발송대상 명단이 보이며 각 대상을 클릭하면 개인의 관심정보를 볼 수 있다.



[그림 2] 개인화 이메일 생성

에이전트에 의해 생성되고 발송되어진 메일은 그림 3과 같이 아웃룩익스프레스나 웹메일을 통하여 받아 볼 수 있으며, 각각의 텍스트와 이미지, 제목에는 사용자의 관심정보를 알 수 있도록 로그정보를 가지고 있다. 사용자는 메일에서 세부정보를 보기 위하여 클릭하는 행위만으로 자신도 모르는 사이에 관심정보를 로그파일에 쓰게 되는 것이다.



[그림 3] 개인화된 이메일

5. 결론

본 논문은 사용자가 정보를 주기적으로 받기 원하는 해당 기업의 사이트로부터 자신만의 특정한 관심정보를 자신이 선호하는 이메일 양식과 더불어 자동 생성하여 푸쉬해 주는 에이전트 시스템을 제안하였다.

고객에 대한 관심도 측정은 관심 컴포넌트를 추출하여 그 관심 컴포넌트에 가중치를 부여하여 관심도 측정 알고리즘에 의해 계산되어진다.

관심 컴포넌트는 개별 고객의 마우스 이벤트와 메일 오픈 시간, 주 관심정보 추출, 세부 관심정보 추출로 나누어지며 마우스 클릭 이벤트는 텍스트의 색상과 길이, 이미지의 색상, 크기 및 모양으로 다시 분류된다.

또한, 세부관심 정보 추출을 위해 마우스 클릭 이벤트가 발생한 객체에 대한 특징을 찾아내어 클러스터링 함으로써 고객의 관심정보를 확장하고 추론할 수 있도록 하였으며 이렇게 각 관심 컴포넌트들의 조합 및 분류로 인해 생성된 각 고객의 개인 관심정보를 기반으로 개인화된 관심정보를 고객의 선호 이메일 양식을 자동으로 생성하여 제공해 준다.

따라서 각 고객은 본인의 개인 관심정보를 이메일로 받아서 확인 할 때마다 관심 컴포넌트들이 재생성되므로 지속적인 고객 관심도를 시스템은 재 계산하여 매 번 업데이트된 관심 정보 이메일을 제공받을 수 있다.

향후 연구 과제로는 관심 컴포넌트들을 확장시켜 마우스 클릭 이벤트뿐만 아니라 마우스 드래그 이벤트 및 키보드 이벤트 등을 이용하여 사용자의 관심정보 추출을 하여 더욱 관심도의 정확성을 높일 수 있는 알고리즘에 대한 연구가 요구된다.

[참고문헌]

- [1] R. Forsyth, "Components of a CRM System and a Value-Driven Approach to Implementation", CRM Forum Report, 1999.
- [2] Dunja Mladenic, Personal WebWatcher : Implementation and Design, *Technical report IJS-DP-7472*, October, 1996.
- [3] Thorsten Joachims, "A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization", March 1996.
- [4] Bruce Krulwich, "Learning document category description through the extraction of semantically significant phrases", *Center for Strategic Technology Research Andersen Consulting LLP 100 South Wacker Drive, Chicago, IL 60606*, 1995.
- [5] Trinity College Dublin., *Broadcorn Eireann research Ltd.*, "Software Agent : A Review", 27, May, 1997.
- [6] Yezdi Lashkari, Max Metral, Pattie Maes, "Collaborative Interface Agents", *Conference of the American Association for Artificial Intelligence*, Seattle, August 1994.
- [7] Dunja, Mladenic "Feature subset selection in text-learning", *Department for Intelligent Systems*, J.Stefan institute, 1997.