

무선 단말기용 XML기반 맞춤 서비스 시스템 설계

송민영* 이기호
이화여자대학교 컴퓨터학과
{navyblue, khlee}@mm.ewha.ac.kr

A Design of Personalization Service System for Wireless Device based on XML

Minyoung Song* Kiho Lee
Dept. of Computer Science & Engineering, Ewha Womans University

요 약

최근 E-Business가 활성화됨에 따라 고객의 특성을 파악해서 고객 개인의 관심에 부합되는 개인화 된 정보나 서비스를 제공할 것이 요구되고 있다. 무선 인터넷을 이용한 서비스가 증가하고 있지만 대부분의 서비스 시스템들은 사용자 개인의 성향은 고려하지 않고 모든 사용자에게 획일적인 서비스를 제공한다. 무선 환경일수록 이러한 무분별한 광고는 오히려 고객의 만족도를 감소시킬 수 있다. 따라서 각각의 고객에게 취향과 관심 분야에 따른 차별화 된 서비스가 필요하다. 기존의 e-mail 시스템들은 모든 사용자들에게 단지 질의한 응답만을 제공하거나 똑같은 광고성 메일을 전달한다. 즉, 개인의 성향은 고려하지 않은 응답 결과를 보여주었다.

이에 본 논문에서는 휴대하기 편리한 이동 단말기의 특성을 이용하여 시,공간적 제약을 극복하고 작은 단말기 액정화면을 통해 정보를 일일이 검색해야 하는 번거로움을 덜어줄 수 있는 XML 기반의 무선 단말기용 맞춤 서비스 시스템을 설계하였다. 이를 위해 e-mail 헤더 정보를 이용하여 사용자별로 분류하였고 텍스트마이닝 기법을 적용해 추출된 토픽과 사용자 프로필 정보를 통해 예측된 사용자의 관심분야에 따른 카테고리를 계산하여 템플릿에 매핑함으로써 맞춤 서비스를 제공하는 시스템을 설계한다. 이로 인해 무선에서 제공하는 서비스의 질을 향상시키고 사용자에게 편리함과 흥미를 유발할 수 있다.

1. 서 론

무선환경과 인터넷을 이용한 서비스가 증가하고 있고 모든 고객을 개별적인 시장으로 보는 1 대 1 서비스의 개념이 등장했으며 이를 실제로 실현하고 있다.

인터넷을 통한 상거래의 변혁에 따른 기업의 마케팅 환경 변화를 지원할 수 있는 혁신적인 솔루션을 제공하기 위해서는 e-mail이나 웹을 통하여 온라인으로 접근해오는 단순고객을 충실고객으로 바꿀 수 있도록 지원하는 시스템의 개발이 필수적이다. 그러나 이와 유사한 관련 연구들이 다각적으로 활발히 진행되고 있지만 아직까지 무선 단말기에서의 개인화 서비스에 대한 개발은 미흡한 실정이다.

이에 본 논문에서는 이러한 변화에 부응하기 위해 획일적인 서비스가 아닌 사용자 자신의 관심분야에 맞는 개인화된 무선 단말기용 맞춤 서비스 시스템을 구축함으로써 작은 액정 화면에서 정보검색을 위한 불필요한 시간과 노력의 낭비를 줄일 수 있는 무선 단말기용 맞춤 서비스 시스템을 제안한다.

또한 휴대하기 편리한 이동 단말기의 특성을 이용해서 시, 공간상의 제약을 극복하고 무선 환경에서 얻을 수 있는 헤더정보와 내용에 대한 텍스트마이닝, 그리고 등록된 사용자 프로필 정보를 통해, 유용한 정보를 추출하여 무선에서 제공하는 서비스 질을 향상시키고 사용자에게 특성화된 서비스를 제공한다.

이를 위해 e-mail 헤더 정보를 통해 얻을 수 있는 사용자 ID별로 개인별 디렉토리를 만들고 사용자의 요구에 해당하

는 내용에 대한 텍스트마이닝 기법을 적용하여 카테고리를 계산한다. 여기에 사용자 프로필에 저장되어 있는 정보를 토대로 사용자의 성향을 예측하여 선별된 정보를 추천하는 것이 가능하도록 한다.

이와 같이 본 논문에서는 무선에서 각 고객에게 1대 1로 맞춤화 된 서비스를 제공함으로써 편리성을 제공하고 시간과 노력의 낭비를 줄이며 서비스에 대한 친밀감과 흥미를 유발할 수 있도록 하였다.

2. 관련연구

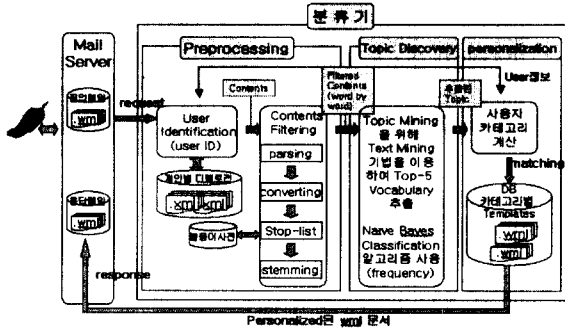
웹상에서 각 고객에게 1대 1로 차별화 된 서비스 제공을 위해 등록된 사용자 프로필 정보와 쇼핑패턴을 토대로 사용자의 관심도와 필요한 시기를 예측해서 e-mail로 개인화된 광고 및 추천 서비스를 제공[1]하는 개인화 서비스에 대한 기존의 여러 연구들이 있다. 본 논문에서는 무선에서의 맞춤 서비스를 제공하는 방안을 제시하였다.

이를 위해 텍스트마이닝 기법을 적용하여 분류를 하였다. 텍스트마이닝은 자연어로 구성된 비구조적인 텍스트 안에서 패턴 또는 관계를 추출하여 지식을 발견하는 것으로, 기존의 텍스트 데이터로부터 새로운 정보를 이끌어내기 위한 과정이라고 정의될 수 있다.[2] 마이닝 기법 중 Naive-Bayes 분류 알고리즘을 사용하여 토픽을 추출하였다. Naive-Bayes 분류기는 다른 기법들에 비해 더욱 나은 성능을 보인다고 제안되었다[3][4][5][6]. 따라서 일반적으로 많이 쓰는 기법을 사용하였다.

3. 맞춤 서비스 시스템 설계

3.1 시스템 구성도

기존 e-mail 시스템들은 모든 사용자에게 단지 질의한 응답만을 제공하거나 모든 사용자에게 똑같은 광고성 메일을 전달한다. 즉, 사용자의 개인 성향이 전혀 고려되지 않은 응답 결과를 보여주었다. 본 논문에서는 결과에 대한 만족을 높이고자 다음과 같은 시스템을 제안한다. 이 시스템은 크게 전처리, 주제어 추출, 맞춤 서비스 세 부분으로 구성되며, 자세한 내용은 다음 <그림 1>과 같다.



<그림 1> 전체 시스템 구성도

무선 단말기를 통해 WML 형식으로 된 문서를 입력받은 후 사용자 ID로 각 사용자를 구별하여 개인별 디렉토리에 데이터를 저장하고 텍스트형태의 문서를 XML로 변환한다. XML문서에서 주제어와 내용부분을 추출하여 필터링 과정을 거쳐 단어 집합을 얻는다. 이렇게 처리된 단어들 중 토픽을 선정하고, 선정된 단어들의 분류(Classification) 과정을 통해 얻어진 사용자 카테고리 정보를 프로파일에 저장한다. 마지막으로 사용자마다 Time history별로 기록된 사용자 프로파일을 바탕으로 카테고리를 계산하여 각각의 WML 형식으로 된 템플릿에 매핑시킨 후 맞춤형 서비스를 제공한다.

3.2 전처리

3.2.1 User Identification

각 개인별로 서비스를 제공하기 위해서는 각각의 고객들을 다른 고객들로부터 구별해야 한다.

사용자가 이동 단말기를 통하여 보낸 메일은 WML 형식으로 입력받아 Web Server 에 텍스트형태로 저장되고 이 문서를 다시 XML 형식으로 변환한다.

헤더는 메시지를 전달하기 위한 일종의 규칙으로 구성되는 다음과 같다.

Return-Path	< adhong@alldons.com >
Received	From mail.alldons.com (mail (aaa.bbb.ccc.ddd) by mail.alldons.com(8.9.3/8.9.3) with ESMTP id FPK5C00 TDH for <adong@maxaworld.com>; Mon, 14 Feb 2001 20:38:36 (KST)
Message-ID	<00601b4d2b230de4f38082217de4@sociality.mm.saltu.sc.kr>
Reply-To	"Hong alldons" <alldons@alldons.com>
From	"Hong alldons" <alldons@alldons.com>
To	alldons@alldons.com
Subject	금주개별영화
DATE	Mon, 14 Feb 2001 20:38:36 +0900 (KST)
MIME-Version	1.0
Content-Type	Text/plain; charset="euc-kr"
X-Priority	3
X-MSMail-Priority	Normal
X-Mailer	Microsoft Outlook Express 4.72.2106.4
X-MimeOLE	Produced By Microsoft MimeOLE V4.72.2106.4
Content-Transfer-Encoding	8bit
X-MIME-Autoconverted	From base64 to 8bit by mail.alldons.com id TAA10649

<표 1> 헤더의 구성

이렇게 전달되어진 파일은 "From: "헤더의 내용을 보고 사용자를 구별하여 각각의 개별 디렉토리에 정보를 저장하고,

텍스트 형태로 된 e-mail 파일을 XML문서로 변환한다. 다음 그림은 입력으로 들어온 WML문서에서 XML문서로 변환된 예를 보여주고 있다.

```
<wml>
<card id="card1">
  <do type="script">
    <do href="#card2"/>
  </do>
  <input type="text" name="subject" format="M"/>
</card>
<card id="card2">
  <do type="script">
    <do href="http://www.maxaworld.com/forward.cgi" method="post">
      <do type="form" name="contents" value="#subject%#(contents)"/>
    </do>
    <do type="button" label="user">
      <do href="#card1"/>
    </do>
  </do>
  <input type="text" name="contents" format="M"/>
</card>
</wml>
```

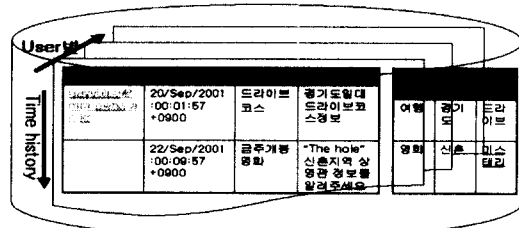
<그림 2> WML문서의 예

```
<?xml version="1.0" encoding="KSC5601" ?>
<DOCTYPE MAIL [
  <ELEMENT MAIL (제목, 보낸시간, 내용)>
  <ELEMENT 제목 (#PCDATA)>
  <ELEMENT 보낸시간 (#PCDATA)>
  <ELEMENT 내용 (P)>
]>
<MAIL>
<제목> 금주개별영화 </제목>
<보낸시간>
<ADDRESS>adong@maxaworld.com</ADDRESS>
</보낸시간>
<내용>
<P> "The Hole" 신촌지역 상영관 정보를 알려주세요</P>
</내용>
</MAIL>
```

<그림 3> XML문서의 예

XML(eXtensible Markup Language) 문서는 사용자가 태그를 작성해 줄 수 있어서 필드를 확장할 수 있는 유연성(flexibility)과 확장성을 지닌다는 이점을 가지고 있다.

개인별 디렉토리에 사용자 ID와 메일 보낸 시간(Date), 주제(Subject), 내용(Contents) 그리고 사용자가 보낸 메일의 내용을 바탕으로 계산된 카테고리 정보들이 들어가게 된다. 이 개인별 디렉토리는 사용자 프로파일로 사용된다. 이렇게 분류된 고객은 Time history별로 DB에 저장된다.



<그림 4> 사용자 프로파일의 예

3.2.2 Contents Filtering

이 과정은 주제어를 추출하기 위한 전처리 작업으로 입력 문서인 XML문서로부터 다음의 과정을 수행한다.

- 파싱(Parsing) - 입력된 문서를 xml 파서를 통해 파싱하고 토큰을 재생성한다.
- 대, 소문자 변환(Converting) - 1단계에서 얻은 의미 있는 토큰 중 영문 소문자는 대문자로 변환한다.
- 불용어 제거(Stop-list) - 불용어를 제거한다.
- 스템밍(Stemming) - Porter Suffix Stemmer를 이용하여 접미사를 제거하고 어근을 추출하여 다시 저장한다.

User ID	Subject	Contente
monica@hanmail.net	음식	신촌에 수제비 전문점
Apple@hanbox.com	여행정보	경기도 지역의 드라이브 코스를 알려주세요
havyblue@m.m.ewha.ac.kr	금주개봉영화	"The Hole" 신촌지역 상영관 정보를 알려주세요

THEHOLE(신촌지역의 금주개봉영화)

<그림 5> Contents Filtering 예

3.3 주제어 추출

전 단계에서 word by word로 저장된 filtered contents에서 토픽이 어느 문서 카테고리에 속하는지 분류하기 위해 주제어를 추출하는 과정으로 자연언어로 구성된 비구조적 텍스트 안에서 패턴이나 관계를 추출하여 지식을 발견하는 텍스트마이닝 기법 중 Naive Bayes Classification 알고리즘을 이용하여 토픽을 추출한다.

Naive-Bayes classification rule은 다음과 같다.

$$Pr(C_i|d) = \frac{Pr(C_i)d \times Pr(d)}{Pr(d)}$$

여기서, Pr(d)는 모든 클래스에 대해 같기 때문에 무시하고, Pr(C_i)와 Pr(d|C_i)는 다음과 같다.

$$Pr(C_i) = \frac{\text{클래스 } C_i \text{ 에 속한 문서의 총수}}{\text{데이터 집합에서 문서의 총수}}$$

$$Pr(d|C_i) = \prod_{w \in d} Pr(w|C_i)$$

이 과정을 거쳐 상위 5개의 vocabulary를 추출한다.

3.4 개인화 서비스

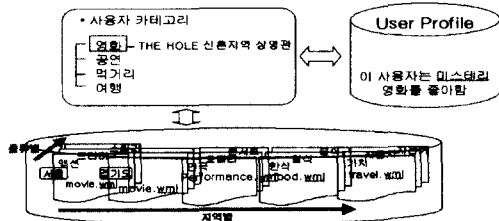
개인화 서비스 단계는 수집된 고객의 프로파일 정보를 이용하여 고객지향 맞춤 서비스를 제공하는 단계이다.

3.4.1 사용자 카테고리 계산

토픽을 기반으로 카테고리별로 분류된 토픽셋에 매핑시킨다. 고객의 질문을 파악하고 이를 효율적으로 분석하고 개인화 된 서비스를 제공하기 위해서 각각의 응답 페이지에 대해서 분류를 해 둔다. 예를 들어 영화 카테고리의 경우 장르별(액션, 드라마, 스릴러, 미스터리, 코믹, 애니메이션,...) 지역별(서울, 경기, 강원,...) 상영관 정보를 분류한다. 고객이 질의 메일을 보낼 경우 토픽을 계산하여 관련된 카테고리별로 분류하는 작업이 필요하다.

3.4.2 사용자 프로파일 관리

본 논문에서 제안하는 시스템은 사용자마다 각기 다른 토픽셋을 매핑하게 되므로, 사용자 프로파일을 이용하여 카테고리별로 분류한다. 프로파일에는 앞서 설명한 개인별 디렉토리에 user ID정보를 이용하여 접근하고 사용자가 관심있는 분야(예를 들어, 영화의 장르, 음식의 종류 등)를 분석할 수 있는 정보가 저장된다.

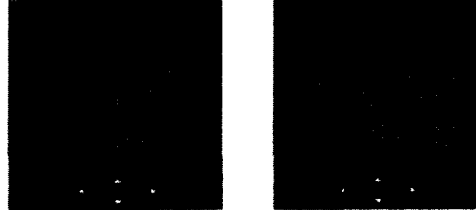


<그림6> 사용자 카테고리 계산의 예

3.4.3 WML 파일로 변환

이동 단말기로 응답 메일을 전송하기 위해 WML 형식의 파일로 변환 과정을 거친 후 변환된 파일을 각각의 고객에게 응답 메일로 전달한다.

4. 사례



<그림 7> UP.Simulator를 통해 본 입력 메일과 맞춤형 응답메일의 예

질의로 들어온 메일에 대한 응답 뿐 아니라 사용자 프로파일에 저장된 정보를 이용하여 사용자가 관심을 갖는 장르의 영화를 함께 추천하는 맞춤형 응답메일을 전송해준다.

5. 결론 및 향후 연구

무선 인터넷을 이용한 서비스가 증가하고 있지만 대부분의 서비스 시스템들은 사용자 개인의 성향은 고려하지 않고 획일적인 서비스만을 제공한다.

이런 문제를 해결하기 위해 본 논문에서 제안한 무선 단말기용 맞춤 서비스 시스템에서는 개인별 취향에 맞는 정보를 제공해 줌으로써 사용자가 직접 작은 단말기 화면을 통해서 웹사이트를 검색하지 않더라도 미리 사용자의 성향을 예측하여 생성한 페이지를 e-mail로 제공해 주는 1 대 1 서비스를 실현해준다.

향후 연구로 무선에서 이 방법을 적용하여 뉴스 서비스나 전자 상거래에 접목시킨다면 맞춤화 된 정보의 제공이 가능해 질 것이다. 그러나 이 시스템이 적용되기 위해서는 각 상품별 카테고리를 자동으로 생성해 줄 수 있는 방법이 우선적으로 필요하며 사용자별 성향을 보다 자세하고 정확하게 분류하는 방법이 필요하다.

참고문헌

- [1] 김동휘, 김순자 “에이전트 기반 비교쇼핑 시스템에서의 개인화 된 광고와 추천 서비스 방안” 한국정보과학회 춘계 학술발표논문집, 27권 2000.
- [2] Lee Hing Yan, “Text Mining-Knowledge Discovery from Text”, Trend in Knowledge Discovery from Databases, 29th June 1999.
- [3] S.Chakrabarti, B. Dom, R.Agrawal, and P.Raghavan. Using Taxonomy, Discriminants, and Signatures for Navigating in Text Databases. In Proc. Of the 23rd Int'l Conf. On Very Large Databases, pages 446-455, 1997
- [4] D.D. Lewis and M. Ringuette. A comparison of two learning algorithms for text categorization. In In Third Annual Symposium on Document Analysis and Information Retrieval, pages 81-92, 1994.
- [5] K. Lang. News Weeder: Learning to Filter Net-News. In Proc. Of the 12th Int'l Conf. On Machine Learning, pages 331-339, 1995
- [6] Andrew McCallum and Kamal Nigam. A Comparison of Event Models for Naive Bayes Text Classification. In AAAI-98 Workshop on “Learning for Text Categorization”, 1998.