

정보기술응용연구  
제 3 권 제 1 호  
2001년 3월

## 전자신문 제공업자를 위한 인터넷 상에서의 개인화된 광고 기법

하성호\*

### 요 약

인터넷과 월드와이드웹 (WWW)의 폭발적인 성장은 온라인 기반의 전 세계적인 전자상거래의 발전을 촉발시켰다. 상거래를 위한 매체로서의 인터넷의 급속한 도입은 기업과 소비자간에 새로운 대화 채널로서의 웹 기반 광고의 중요성을 부각시키고 있다. 제대로 된 웹 광고가 올바르게 소비자에게 전달된다면 웹 광고의 효과는 극대화되고 기업의 수익은 지대하게 증대될 수 있다. 따라서 본 논문은 인터넷 상에서 전자신문 서비스를 제공하는 기업을 위한 지능화된 고객 서비스의 일환으로 개인화 (personalization) 된 광고 기법을 제공하고자 한다. 전자신문은 통상적으로 정치, 경제, 스포츠, 문화 등의 여러 섹션으로 구분되어 제공되어지는데, 기계 학습 (machine learning) 기법을 도입하여 고객이 전자신문 페이지에 접속, 검색하고 읽어보는 기록을 분석하고 적절한 웹 광고 (특히, 배너 광고)를 선택한 뒤, 그것을 고객이 해당 웹 사이트에 재접속할 때 제시함으로써 웹 광고에 대한 고객의 반응을 극적으로 향상시키고자 한다. 이를 위하여 현재 대한민국의 인터넷 전자신문 제공업자 중 유명한 한 곳을 선정하여 본 논문에서 제시한 방법론을 적용하고자 한다.

\*) 한국과학기술원 산업경영연구소 연구원

## 1. 서론

인터넷과 웹 기술의 상거래에의 광범위한 적용은 기업들이 가상적인 온라인 시장에서 더 빠르고 편리하게 비즈니스 정보를 교환하고 금융 거래를 할 수 있도록 만들고 있다. 이것은 일반적인 상행위에 따르는 거래 비용을 절감하는 효과를 가져왔는데 이를 통한 비용 절감 혜택은 거래 당사자들에게 돌아감으로써 인터넷 비즈니스에 대한 관심을 촉발시키는 계기가 되고 있다. 따라서 인터넷 비즈니스의 경쟁이 심화되는 상황에서 기업이 경쟁력을 제고하고 경쟁우위를 확보하기 위해서는 고객의 욕구를 만족시킬 수 있는 개별화된 서비스를 제공할 수 있어야만 한다.

이러한 새로운 대화 채널로서의 인터넷의 뚜렷한 이점 (targetability, tracking, deliverability, flexibility, interactivity [21])에도 불구하고 현재의 인터넷 광고는 모든 고객에게 동일한 광고 메시지를 뿌리고 있다. 이것은 결국 고객의 광고에 대한 반응을 약화시키는 결과를 가져올 뿐이다.

인터넷 기술이 발전함에 따라 광고 기법도 아울러 발전하고 있는데 원시적인 형태의 웹 광고는 웹사이트 그 자체였다. 하지만 웹이 점점 상업적인 사이트로서의 가능성을 주목받기 시작하면서 단순히 웹사이트를 만드는 행위 자체로서 인터넷 고객을 유인하기에는 불충분하다는 것이 인식되기 시작했다. 따라서 오늘날 사용되는 인터넷 광고 모델은 배너 광고, 이메일, 채팅/커뮤니티, 후원 (sponsorship), 푸시 기법 (push technology) 등에 이르고 있다 [2]. 특히, 웹 배너 광고가 현재의 온라인 광고 시장에서 선도적인 위치를 점하고 있다. 웹 배너 광고의 효과를 향상시키기 위해 웹 광고를 운영하는 사이트는 적시에, 적절한 고객에게 적절한 광고 메시지를 제공할 수 있는 능력이 필요하다. 다시 말해 어떤 고객이 전자신문의 스포츠 섹션을 주로 읽는다면, 그 고객에게는 스포츠 관련 광고를 보여주는 것이 웹 광고의 효과를 극대화하는 방법이 될 것이다.

성공적인 온라인 마케팅 기법은 고객을 식별하고 고객의 취미나 취향을 이해하고 예측하며 적절한 광고를 선정하여 개인화된 방법으로 직접 고객이 웹사이트에 접근할 때 제시할 수 있어야 한다 [11]. 이것이 최근의 마케팅 전략에서 특히 주목받고 있는 일대일 마케팅 (one-to-one marketing) 개념이다 [17]. 게다가 일대일 마케팅에서는 개개 고객의 욕구를 이해하고 고객 세분화를 통해 고객을 구별하며 개인화된 제품이나 서비스를 제공하여 고객과의 관계를 더욱더 공고히 하는 것을 중요시하는데 이를 고객관계관리 (CRM, customer relationship management)로 부르기도 한다 [16].

지금까지의 경향을 보면 웹사이트 방문자와 고객의 정보를 수집하는 가장 효

과적인 방법은 “고객 등록 화면”의 초기 프로파일을 이용하는 것이다 [15]. 고객이 특정 사이트에 회원으로 등록하면서 제공하는 인구 통계적인 정보나 개인의 취향 등과 같은 정보는 해당 고객에게 어떤 광고를 제시하는 것이 좋을 것인지를 판단하는 기본적인 자료로 활용되어졌다. 그러나 온라인을 통한 사생활 노출에 대한 거부감은 정확한 고객 정보의 수집 자체를 어렵게 하고 있다. 그것은 온라인 광고의 효과를 감소시키고 있다. 게다가 온라인 고객으로 하여금 그들의 기호가 바뀔 때마다 웹사이트를 방문하여 고객 정보를 변경하도록 요구하는 것은 고객들을 귀찮게 하는 일이고 심지어는 해당 사이트를 떠나는 이유가 되기도 한다. 따라서 고객의 정보를 획득하는 또 다른 방법은 웹 서버에 저장되어 있는 로그파일을 이용하는 방법이다. 웹 서버 로그를 마이닝함으로써 사생활의 직접적인 노출이라는 고객들의 불안을 감소시키고 그들의 웹사이트 이용 패턴을 추적하여 관심 사항을 파악하고 적절한 광고를 제시할 수 있다.

본 논문에서는 인터넷 상에서 전자신문을 제공하는 기업을 대상으로 일대일 마케팅이 가능하도록 웹 광고를 선택할 수 있는 기법을 제시하고자 한다. 이 기법에 따르면 웹사이트 고객의 취향을 반영하는 적절한 광고를 실시간으로 선정하여 제시함으로써 고객의 관심을 촉발하고 따라서 광고에 대한 반응을 향상시킬 수 있다. 전자신문 사이트에 접속한 고객의 섹션 이용 패턴을 분석하면 목표 고객의 관심사를 알 수 있게 되고 그 관심사에 가장 밀접한 광고를 선정하여 집중 제시하게 된다. 즉, 웹사이트 광고를 개개 고객에 대해 맞춤 (customization) 함으로써 웹사이트 콘텐츠의 개인화를 구현하게 되는 것이다. 목표 고객의 섹션 이용 패턴이 변하면 고객의 관심도가 변화했다는 것을 의미하게 되고 이런 변화에 맞게 웹 광고는 적절하게 다시 선정되어진다. 이것을 위해 고객의 웹사이트 이용에 대한 정보를 담고 있는 서버 로그 파일을 마이닝하고 퍼지 논리에 입각하여 적절한 웹 배너를 선정하는 방법을 이용한다. 퍼지 논리는 본 논문에서 제시하는 고객 세분화와 광고 카테고리 분류, 적절한 광고의 선택 등에서 조정될 수 있는 애매함 (fuzziness) 을 처리할 수 있도록 도입되어진다.

본 논문의 구성은 다음과 같다. 2장에는 전자상거래 분야에서 연구되고 있는 일대일 웹 마케팅의 현재 모습을 보여주고, 3장에서는 광고 선택기의 구성과 광고 선택 절차를 제시하며, 4장에서는 광고 선택기의 실제 적용을 예시하고 마지막 5장에서는 본 논문의 결론과 향후 연구 과제를 간략하게 설명한다.

## 2. 기존 문헌 연구

인터넷을 활용하는 온라인 서비스 제공업자들은 일반적으로 마케팅의 관점에서 자신들의 웹사이트 효율성을 평가한다. 이런 관점에서 웹 마케팅은 온라인 상점들이 고객을 획득하고 유지시키는 일련의 행위라고 정의되어진다 [19]. 온라인 마케팅 기법에는 데이터베이스 마케팅 (database marketing), 일대일 마케팅 (one-to-one marketing), 목표 광고 (ad targeting) 이 있다.

데이터베이스 마케팅은 기업이 고객들을 대상으로 맞춤 서비스를 제공하려는 시도이다. 우편번호, 소득, 직업과 같은 인구통계자료에 근거 고객을 세분화하고, 세분화된 고객 그룹에 마케팅 활동을 집중한다.

일대일 마케팅은 개개의 고객을 특별히 독립된 개체로 간주하여 전통적인 마케팅의 비개인성 (impersonal)을 극복하고자 한다. 현재, 온라인 상점들의 일대일 마케팅 전략을 구현하는데 있어 지금까지 많은 웹 기반의 개인화된 추천 시스템 (recommendation system)이 제안되어지고 있다 [1, 4, 12, 18]. 개인화는 개인 프로파일을 통해 제품이나 웹 페이지를 필터링함으로써 이루어진다. 필터링에는 두 가지 접근 방법이 있는데, 콘텐츠 기반 (content-based)과 협동 필터링 (collaborative filtering)이 그것이다. 콘텐츠 기반의 추천 시스템은 특정 고객이 과거에 좋아했던 것과 유사한 제품이나 서비스를 제공한다. 일반적으로 제품이나 서비스, 프로파일은 속성 공간 (feature space)에서 벡터로 표현되어지고 유클리디안 거리 (Euclidean distance) 와 같은 표준 척도를 통해 계산되어지는 유사성 정도에 근거하여 필터링을 한다. 협동 필터링은 목표 고객 (target customer) 과 유사한 여타 고객들이 선호하는 제품이나 서비스를 추천한다. 이 방법은 고객 클러스터링을 통해 기호가 유사한 고객 세그먼트를 파악하고 그 세그먼트의 정보를 활용한다.

목표 광고는 고객의 과거 행태에 기반하여 어떤 고객에게 어떤 광고를 집중하면 될 것인지를 결정한다. 인터넷 광고가 광고주에게 매력적으로 보이게 된 이면에는 다음과 같은 요인이 작용한다.

- 온라인 서비스 제공업자가 고객의 인터넷 이용 행태를 수집하고 분석하여 적절한 광고를 선택할 수 있게 한다. 인터넷은 목표 고객에 집중할 수 있도록 하며 따라서 광고비 대비 광고의 효과를 증진시킨다.
- 맞춤 광고가 가능하다. 서비스 제공업자가 고객 정보를 축적함에 따라 온라인 상점, 제품, 가격, 판매 과정이 개개 고객을 위한 맞춤식이 되어진다.
- 인터넷 광고비가 비용면에서 기존의 텔레비전, 라디오, 인쇄 매체 광고 보다 더 저렴하다. 이것이 인터넷 광고의 매력적인 부분이다.
- 인터넷 광고의 큰 장점은 기업들이 광고의 성공 여부를 이전보다 쉽게 모니터링할 수 있고 따라서 광고비 지출을 효과적으로 통제할 수 있다는데 있다.

다.

- 인터넷 광고는 지역성을 극복하고 전세계적인 홍보가 가능해진다.

최근의 포레스트 조사에 의하면 배너 광고가 여타 온라인 광고보다 인기 면에서 우위를 보이고 있다. 이 같은 배너 광고의 성장은 상기되어진 장점들을 이들 광고가 가지고 있다는 것에서 기인한다. 배너 광고는 CPM (cost per thousand) 기반으로 가격이 책정이 되어지는데 고객의 광고에 대한 관심과 집중은 클릭의 형태로 사이트 서버에 로그 파일로 흔히 저장이 되어진다.

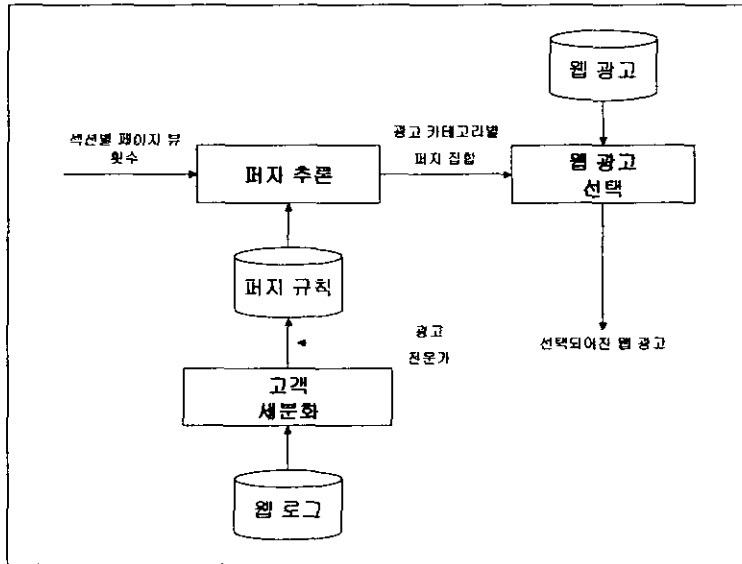
본 연구에서 분석의 대상이 되고 있는 전자신문 서비스의 특징을 약술해보면, 전자신문 서비스는 기존의 종이 신문을 전자화하여 인터넷 매체를 통해 서비스하는 형태를 취하는데 종이 신문이 시시각각 변하는 최신 뉴스를 즉각 갱신하지 못하는 단점을 가지고 있는데 반해 전자신문은 인터넷 환경의 상호 작용성을 충분히 활용하여 즉각 새로운 뉴스로의 갱신이 가능하다. 또 문서, 음성, 영상 등의 멀티미디어 도구를 활용하여 콘텐츠를 제공할 수 있어 훨씬 더 풍부하고 생생한 정보를 고객에게 제공할 수 있다. 전자신문이 웹의 하이퍼링크 기능을 통해 지면을 구현하다보니 클릭 스트림의 깊이가 커질 수 있는 단점을 가지고 있지만, 고객관계관리의 관점에서 고객의 취향을 분석하고 적절한 광고를 적시에 제공할 수 있는 것은 분명 그만의 장점이다.

전자비즈니스 환경에서 고객의 클릭 스트림 자료에 내포되어진 정보를 분석하는 것은 웹 마케팅의 효율성을 증진시키기 위해 매우 중요하게 인식되어지고 있다. 그와 같은 노력의 일환으로 최근에는 웹 서버 로그 분석 연구가 활발히 이루어지고 있다. 그 중에는 데이터마이닝 기법을 이용하여 고객들의 웹사이트 접근 행태를 모델링하는 연구도 많이 이루어지고 있다 [5, 8].

### 3. 웹 광고 선택 기법

본 연구에서 제안하는 웹 광고 선택을 위한 시스템 (일명, 웹 광고 선택기)는 [그림-1]과 같이 크게 세 개의 부분으로 나누어진다: 고객 세분화, 퍼지 추론, 웹 광고 선택. 먼저 고객 세분화 부분은 인공신경망의 일종인 자기조직화특징 맵 (SOM, self-organizing map)을 이용하여 웹 서버 로그 파일을 마이닝하고 그 결과, 웹 사이트를 이용하는 고객을 유사한 취향을 보이는 유한 개수의 세그먼트로 분할한다. 퍼지 추론 부분은 목표 고객의 전자신문 섹션별 페이지 뷰 횟수를 입력 벡터의 형태로 받아들여서 추론한 뒤 각각의 광고 카테고리별 퍼지 집합과

확정값을 얻는다. 이 퍼지 집합은 웹 광고 선택 부분으로 다시 넘어가는데 여기서는 해밍 거리 (Hamming distance)를 계산하고 적절한 광고를 선택한 후 목표 고객에게 제시하게 된다.



[그림-1] 웹 광고 선택 절차

### 3.1 웹사이트를 이용하는 고객 세분화

어떤 광고가 목표 고객에게 가장 적절한가? 이 질문은 웹 광고의 효과를 극대화하기를 기대하는 광고주를 대신하여 웹 광고를 집행하는 서비스 제공자들에게는 가장 중요한 질문이다. 서비스 제공자는 웹사이트 이용자에게 가장 적절한 광고를 제시하기 위해서 고객의 웹사이트 이용 패턴을 파악할 수 있는데 이것은 고객의 로그 파일을 마이닝 (Web mining) 함으로써 가능하다.

웹 마이닝은 지금까지 두 가지 방향으로 연구 접근되어지고 있다 [7]. 하나는 웹 콘텐츠 마이닝 (Web content mining) 이라고 불리는 것으로 웹사이트에 산재되어있는 자료로부터 정보를 추출하는 과정이다. 최근 몇 년 사이에 웹에서 얻을 수 있는 반구조화된 자료를 조직화하는 정보 검색 (information retrieval) 도구들이 속속 개발되어지고 있다.

두 번째는 웹 사용 마이닝 (Web usage mining) 이라고 하는 것인데 이것은 웹 서버로부터 고객의 접근 패턴을 자동 추출하는 것이다. 인터넷에서 가상 상점을 운영하고 있는 소매상들은 웹 서버 로그에 모여진 고객의 자료를 분석하여

고객의 평생 가치를 (life time value) 결정하고 여러 제품들을 망라하는 마케팅 전략을 구현하며 자신들의 판매 촉진 행위의 효과를 평가하기도 한다. 웹 사용 마이닝을 활용하면 웹사이트를 어떻게 디자인하는 것이 고객들의 시선을 유도할 수 있는 것인지에 대한 정보를 획득할 수도 있다. 아울러 효과적인 고객과의 대화관리나 웹 서버 구조에 대한 통찰을 얻을 수 있다.

고객의 접근 이력을 추출하기 위해서는 먼저 서버 로그 파일에 대한 정제 작업과 고객 트랜잭션 파악 작업이 선행되어야 한다 [8]. 데이터 정제 작업은 로그 파일에 축적되어있는 자료 중에서 필요없는 레코드를 제거하는 작업이다. 이와 같이 불필요한 레코드의 제거는 URL (uniform resource locator) 에 나타나는 파일 이름의 확장자를 확인함으로써 용이하게 진행되어질 수 있는데, gif, jpeg, GIF, JPEG의 파일 확장자를 가진 로그 레코드를 모두 제거한다. 개개 고객별 트랜잭션의 확인은 별도의 로그인 이력을 관리하는 시스템의 존재로 어렵지 않게 수행되어질 수 있는데, 트랜잭션 확인 문제에 대한 연구는 [6, 7] 을 참조하면 된다.

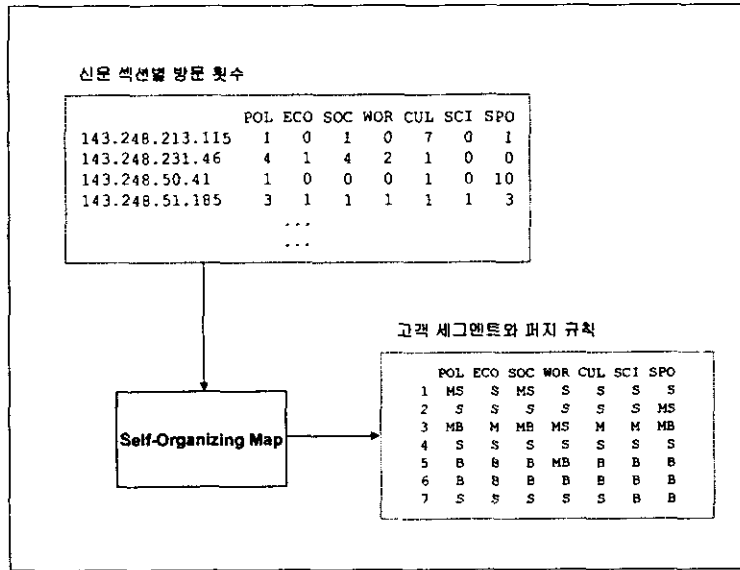
고객별 웹사이트 접근 이력이 파악되어지면 경로 분석, 연관 규칙 추출, 순차 패턴, 클러스터링, 분류 등 여러 가지 형태의 후속 마이닝 작업이 가능한데 본 연구에서는 고객 클러스터링을 하기 위해 정제된 서버 로그 파일에서 고객별로 다음의 속성을 파악해낸다.

- 고객 번호 혹은 고객 아이디
- 개개 고객이 섹션별 페이지를 최근에 언제 방문하였는가?
- 개개 고객이 섹션별 페이지를 몇 차례에 걸쳐 방문하였는가?

[그림-2]는 고객 세분화 과정을 보여준다. 본 연구에서는 전자신문의 섹션을 적절하게 7개 분야로 나누었는데 정치 (POL), 경제 (ECO), 사회 (SOC), 국제 (WOR), 문화 (CUL), 과학과 기술 (SCI), 스포츠 (SPO) 섹션이 그것이다. 이 7개의 섹션 구분은 나중에 퍼지 규칙의 전건부 변수로 (antecedent) 사용되는데 고객 세분화를 위해 각 고객의 섹션별 방문 횟수를 세어서 SOM의 입력 변수로 사용한다.

SOM은 인공지능망을 학습하는데 비교사 학습 알고리즘 (unsupervised learning)을 사용한다 [3, 10]. 비교사 학습에서는 신경망에 입력 데이터를 주는 것만으로 그들에 대하여 비슷한 것끼리 모아서 자동적으로 클러스터링할 수가 있다. 일반적으로 웹 마이닝은 학습 단계 (training phase) 와 이용 단계 (use phase) 의 두 단계를 거친다. 학습 단계에서는 웹 접근 데이터를 SOM을 이용하여 분석하고 고객 행태 모델을 수립하는 단계이다. 이 단계는 일반적으로 많은

시간이 소요되고 데이터 분석가의 판단이 요구되어지는 단계이다. 모델이 수립되고 나면 웹 마이닝은 이용 단계에 들어가는데 이 단계에서는 학습 단계에서 수립된 모델이 개개의 고객 상황에 적용되어진다.



[그림-2] SOM을 이용한 고객 세분화 과정

SOM을 이용한 고객 세분화의 결과, 7개의 특징적인 고객 그룹이 도출되어졌다 (상기 [그림-2] 참조). SOM의 학습에서 입력 뉴런의 개수는 7, 출력 뉴런의 개수를 10보다 작게 설정을 했는데 출력 뉴런의 개수를 10보다 작게 한 것은 이론적인 근거에 입각해서라기보다 고객 관리의 편리상 10개 이하의 고객 그룹이 적절하다고 판단되어졌기 때문에 취한 조치이다.

도출되어진 고객 그룹에 대한 분석은 다음의 [표-1]로서 요약되어진다.



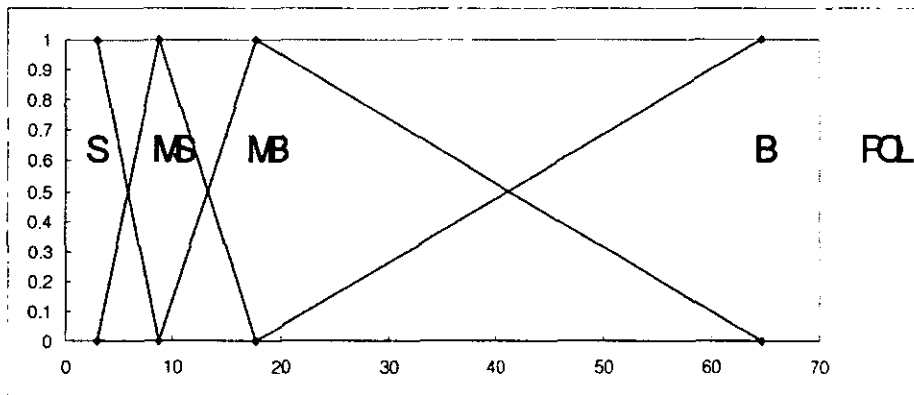
[표-1] 고객군과 전자신문 섹션별 평균 접속 횟수

고객군	섹션						
	정치	경제	사회	국제	문화	과학기술	스포츠
1	8.73	4.68	5.34	1.77	1.91	1.93	2.87
2	60.68	70.84	53.32	21.05	46.37	47.11	127
3	3.37	2.74	2.76	1.04	1.76	2.01	10.22
4	17.66	18.45	14.34	4.99	9.45	11.52	29.84
5	2.75	2.64	2.72	1.12	1.31	1.3	1.54
6	51.6	60.52	45.56	17.48	36.92	36.64	103.16
7	89.69	111.81	82.94	32.75	69.88	76.31	182.81

상기 표는 각 고객군에 속하는 고객들이 평균적으로 섹션별로 얼마나 접속을 하였는지를 보여준다. 이를 바탕으로 섹션별 삼각 퍼지 수를 도출할 수 있는데, 고객군에 의한 섹션별 방문 횟수는 S (small), MS (medium small), M (medium), MB (medium big), B (big) 와 같은 삼각 퍼지 숫자로 표현되어질 수 있다. 예를 들어 [그림-3]은 정치 (POL) 섹션의 퍼지 분할을 보여준다. 정치 섹션의 삼각 퍼지 숫자는 다음의 값을 가진다.

$$MS = (2.9, 8.7, 17.6)$$

$$MB = (8.7, 17.6, 64.6)$$



[그림-3] 정치 섹션 (POL)의 퍼지 분할

### 3.2 웹 광고의 카테고리화

목표 고객에게 제시되어질 웹 광고를 저장하는 데이터베이스를 구축하기 위해

서 실제 약 60개 정도의 웹 광고를 모으고 분류하였다. 이들 광고는 다음의 6개 카테고리로 분류가 되어졌는데, 간략하면 다음과 같다.

- MAG : 잡지, 신문 등 뉴스 관련 광고
- ECO : 비즈니스, 금융, 주식 관련 광고
- COM : 컴퓨터, 정보 통신 관련 광고
- SPO : 스포츠 관련 광고
- ENT : 취미, 오락 등의 일상 생활 관련 광고
- RAN : 임의성 광고

광고 분류는 광고 전문가에 따라서 다를 수 있음에 주의할 필요가 있다. 따라서 광고주가 웹 광고를 의뢰할 때 상기 6개 범주에 따라 자신들의 광고를 분류하고 광고 전문가가 이 분류에 따라 상응한 퍼지 집합을 할당하는 것이 바람직하다. 이 6개 카테고리는 퍼지 규칙의 후건부 (right-hand side) 변수로 사용되어 지는데 규칙의 전건부 (left-hand side) 가 어느 정도 참일 경우에 실행이 촉발 (fire) 되어진다. 후건부 변수인 광고 카테고리는 0.0 ~ 1.0 사이의 연속적인 소속 값 (membership degree) 을 가지며 S (small), M (medium), B (big)의 삼각 퍼지 숫자를 가진다. "RAN" 이라는 범주가 포함된 이유는 목표 고객에게 제시할 적절한 광고를 선정하지 못할 경우에 대응하기 위함이다. "RAN" 범주가 다른 범주보다 큰 퍼지 값을 가지면 웹 광고 선택기가 임의로 선택한 광고를 고객에게 제시한다.

### 3.3 퍼지 규칙과 퍼지 추론 엔진

본 논문에서 제시하는 퍼지 추론 엔진은 목표 고객의 전자신문 섹션별 방문 횟수를 입력 벡터 값으로 받아들인다. 이 입력 벡터를 미리 구축되어져 있는 퍼지 규칙 베이스에 적용하여 퍼지 추론을 한 뒤, 개개 광고 카테고리별 퍼지 집합과 비퍼지화의 결과 값인 확정치를 얻어낸다.

#### • 퍼지 규칙 디자인

본 연구에서는 32개의 퍼지 규칙이 광고 전문가에 의해 도출되어졌으며 다음의 [그림-4]에 퍼지 규칙의 샘플이 예시되어져있다.

	POL	ECO	SOC	WOR	CUL	SCI	SPO	THEN	MAG	ECO	COM	SPO	ENT	RAN
R1	S	S	S	S	S	S	MS	THEN	S	S	S	B	S	S
R2	B	B	B	B	B	B	B	THEN	B	B	B	B	B	S
R3	B	B	B	MB	B	B	B	THEN	M	M	B	B	B	S
R4	B	B	B	B	B	B	B	THEN	B	B	B	B	B	S
R5		B						THEN	M	M	S	S	S	S
R6				MS				THEN	S	S	S	S	S	M
				...								...		
				...								...		

[그림-4] 샘플 퍼지 규칙

퍼지 규칙은 IF-THEN 형식으로 기술되어지는데 전건부 변수는 섹션별 방문 횟수를 삼각 퍼지 숫자로 표시한 것이고 후건부 변수는 웹 광고 범주별 삼각 퍼지 숫자로 표시한 것이다. 7개의 전건부 변수와 6개의 후건부 변수를 가지고 있기 때문에 7 입력 6 출력의 IF-THEN 규칙이다.

먼저, 광고 전문가는 추출된 7개의 고객 세그먼트에 적절한 광고가 각각 무엇 일 것인지를 결정한다. 예를 들어 [그림-4]의 첫 번째 규칙인 R1을 보면 고객 세그먼트 2에 속하는 고객들이 스포츠 섹션에 관심을 가지고 있으므로 그들에게는 스포츠 관련 광고가 적절하다는 것을 보여준다. 이런 규칙에 더하여 전건부에 한 개의 삼각 퍼지 숫자를 가지는 특별한 규칙이 생성되어 규칙 베이스에 추가되어 지는데 이것은 퍼지 규칙의 적용 범위를 일반화시키기 위함이다. 예를 들어 [그림-4]의 규칙 R6을 보면 국제 (WOR) 섹션에 관심을 가지고 있는 고객에게는 임의로 선택한 광고를 제시함이 적절하다고 예시하고 있다.

• 퍼지 추론 엔진

섹션별 방문 횟수에 입각하여 적절한 웹 광고를 추론하기 위하여 Mamdani의 MIN-MAX 연산이 퍼지 추론법으로 사용되어진다. 방문 횟수 벡터가 (POL, ECO, SOC, WOR, CUL, SCI, SPO) = (x<sub>1</sub>, x<sub>2</sub>, x<sub>3</sub>, x<sub>4</sub>, x<sub>5</sub>, x<sub>6</sub>, x<sub>7</sub>)의 형태로 주어질 때, 다음의 식 (1)은 Mamdani의 추론법을 묘사하기에 충분하다 [9, 14].

$$W_k = \mu_{A_1}(x_1) \wedge \mu_{A_2}(x_2) \wedge \mu_{A_3}(x_3) \wedge \mu_{A_4}(x_4) \wedge \mu_{A_5}(x_5) \wedge \mu_{A_6}(x_6) \wedge \mu_{A_7}(x_7)$$

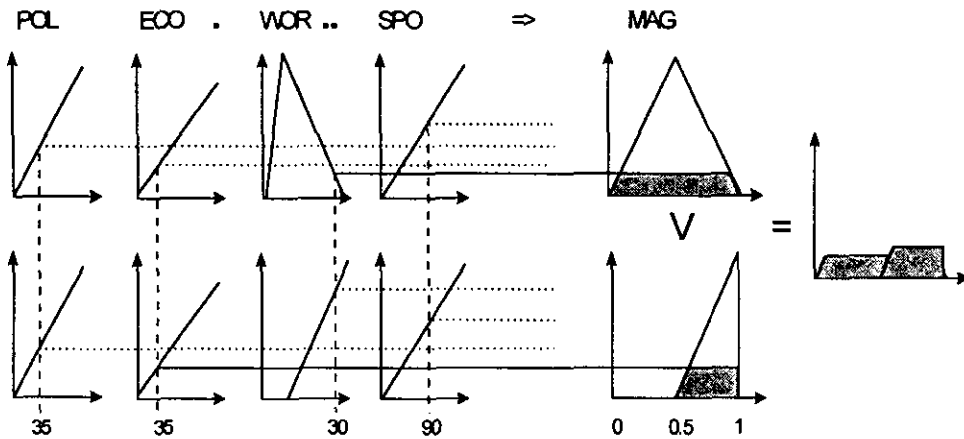
$$\mu_C(z) = \bigvee_{j=1}^{\text{rule \#}} [W_j \wedge \mu_{C_j}(z)] \quad (1)$$

여기서  $x_i$ 는 방문 횟수 벡터에서  $i$ 번째 섹션에 대한 방문 횟수를 나타내고,  $\mu_{A_i}(x_i)$ 는  $i$ 번째 섹션 퍼지 집합의  $x_i$ 에 있어서의 소속값을 나타내고,  $W_k$ 는  $k$ 번째 규칙의 전건부의 적합도를 나타내며,  $\mu_C(z)$ 는 웹 광고의  $z$ 번째 범주의 퍼지 집합을 나타낸다.

퍼지 추론의 과정은 다음과 같이 요약되어진다.

- i) 방문 횟수 벡터에 대한 각 규칙의 적합도를 구한다. 각 퍼지 규칙에는 7개의 전건부 변수가 있기 때문에 7개의 소속 함수 값이 계산되어지고 이들 소속 함수 값의 MIN ( $\wedge$ ) 연산에 의해 각 규칙의 적합도를 구할 수 있다.
- ii) 단계 I에서 구한 적합도를 후건부의 퍼지 집합에 적용하여 각각의 규칙의 추론 결과를 구한다.
- iii) 단계 2에서 구한 각각의 규칙의 추론 결과를 MAX ( $\vee$ ) 연산을 통해 모은다.

예를 들어, 방문 횟수 벡터가  $X = (\text{POL}, \text{ECO}, \text{SOC}, \text{WOR}, \text{CUL}, \text{SCI}, \text{SPO}) = (35, 35, 35, 30, 35, 35, 90)$ 로 주어졌다면 "MAG" 범주에 대한 퍼지 집합은 어떻게 구할 수 있는가? 자세한 과정이 [그림-5]에 도식화되어 있다. Mamdani의 추론법은 R3 규칙에 의해 "국제 (WOR)", R4 규칙에 의해 "경제 (ECO)"가 최소 적합도를 가짐을 보여준다. 나머지 광고 범주에 대한 퍼지 집합 도출은 동일한 방식으로 진행되어진다.



[그림-5] "MAG" 범주에 대한 퍼지 추론

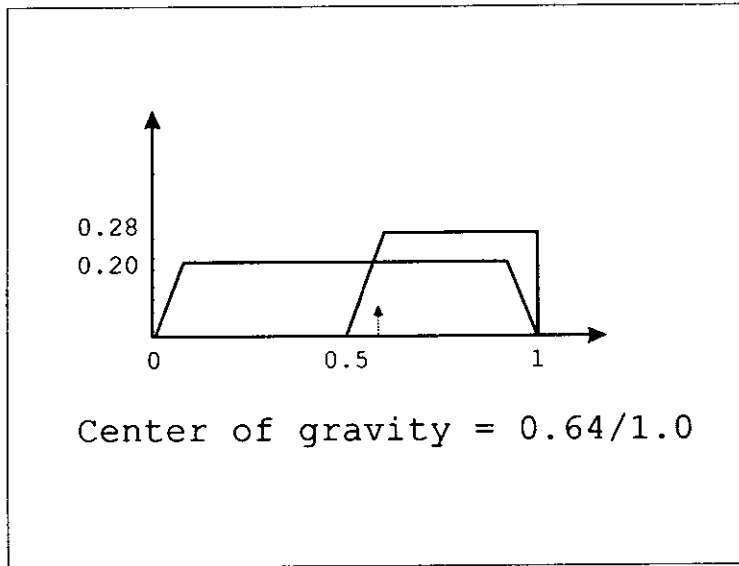
• 비퍼지화 (defuzzification)

비퍼지화는 퍼지 제어 (fuzzy control) 와 퍼지 전문가 시스템 (fuzzy expert system)에 많이 사용되어지는데 중심 좌표 (central gravity), 중앙값 (median) 과 같은 대표값을 구해 주어진 퍼지 집합을 확정치로 변환하는 조작이다 [13]. 다시 말해 비퍼지화는 모든 퍼지 집합을 실수의 집합으로 매핑시키는 방법을 제공한다.

비퍼지화는 퍼지 추론의 결과 도출된 광고 카테고리별 퍼지 집합을 실수 (real number)로 전환시킨다. 본 논문에서는 무게중심법 (central gravity method, centroid method)을 사용하는데 다음의 식 (2)에 의해 무게중심 ( $Z_{defuzzified}$ )이 계산되어진다.

$$Z_{defuzzified} = \frac{\int \mu_c(z) \cdot z \, dz}{\int \mu_c(z) \, dz} \quad (2)$$

무게중심법에 의해 "MAG" 범주의 비퍼지화는 [그림-6]과 같이 묘사되어진다.



[그림-6] 광고 범주 “MAG”에 대한 무게중심

• 웹 광고 선택

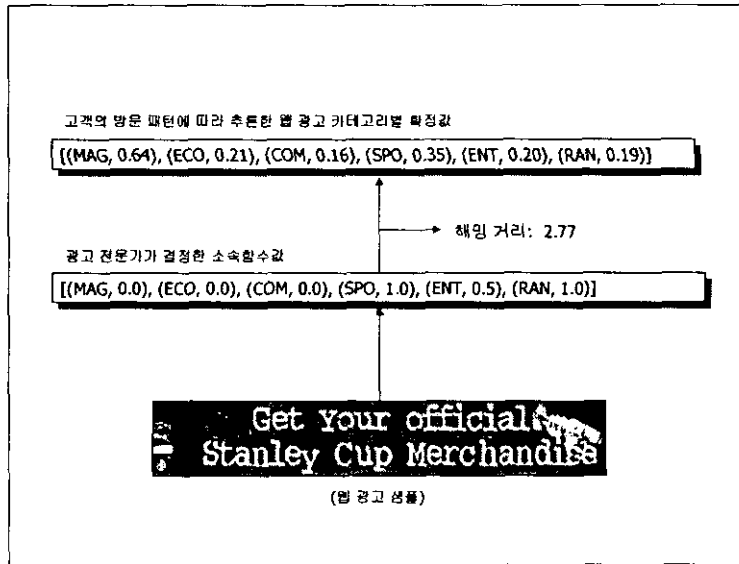
퍼지 추론을 거친 후, 목표 고객의 웹 페이지 방문 패턴에 따르는 광고 카테고리별 퍼지 집합과 비퍼지수가 계산되어진다. 그리고 웹 광고 데이터베이스에 저장되어져 있는 광고는 광고 전문가에 의해 이미 적절한 퍼지 집합이 주어져 있으므로 목표 고객에 제시되어질 적절한 광고의 선택은 광고 카테고리별 퍼지 집합과 광고의 퍼지 집합사이의 해밍 거리 (Hamming distance)를 계산함으로써 결정되어진다. 해밍 거리가 가까울수록 두 퍼지 집합은 유사해지는데 목표 고객의 웹 페이지 방문 패턴에 걸맞는 광고일수록 해밍 거리는 가까워진다 [20]. 따라서 해밍 거리가 가까운 광고를 우선적으로 해당 고객에게 제시하도록 한다. 해밍 거리는 다음의 식 (3)에 의해 계산되어진다.

$$d(A, B) = \sum_{\substack{i=1 \\ x_i \in X}}^n |\mu_A(x_i) - \mu_B(x_i)| \tag{3}$$

여기서  $d(A, B)$ 는 퍼지 집합  $A, B$ 의 해밍 거리를 나타내고,  $\mu_A(x_i)$ 는 퍼지 집합  $A$ 의  $x_i$ 에 있어서의 소속값을,  $\mu_B(x_i)$ 는 퍼지 집합  $B$ 의  $x_i$ 에 있어서의 소속값을 각각 나타낸다.

[그림-7]은 샘플 광고의 퍼지 집합과 목표 고객의 방문 패턴으로부터 추론한

개인화된 퍼지 집합사이의 해밍 거리를 보여준다. 실제로 이 샘플 광고는 전미 아이스하키 리그에서 기념품을 팔기 위한 광고이다. 이 광고는 광고 전문가에 의해 "SPO" 카테고리에 소속하는 정도가 1.0, "ENT" 카테고리에 소속하는 정도가 0.5, 무작위적인 성격 (RAN) 이 1.0의 값을 가지는 것으로 결정되어졌다.



[그림-7] 두 퍼지 집합의 해밍 거리

#### 4. 시스템의 실제 실행 예

프로토타입 시스템의 구축이 끝난 후, 목표 고객을 상대로 시스템의 작동 예를 들어보면 다음의 [표-2]와 같다. 이 표에 의하면 고객 번호가 "128.134.80.61" 인 고객은 전자신문의 각 섹션별로 다음과 같은 접속 횟수를 가지는 것으로 집계되어졌다.

$$(POL, ECO, SOC, WOR, CUL, SCI, SPO) = (9, 6, 4, 4, 1, 4, 2)$$

상기 방문 횟수를 퍼지 추론 엔진에 입력시켜 본 결과, "MAG" 카테고리에 대한 퍼지 집합을 구하기 위해서 1번 퍼지 규칙이 최소 적합도를 가지는 것으로 파악이 되어졌고, 실제 0.266846 이라는 소속값을 가지는데 이것은 "MAG" 카테고리에 대한 퍼지 집합에서 입력 접속 횟수가 규칙 1에 소속하는 정도를 일컫는

다.

[표-2] 목표 고객의 웹 사이트 방문 패턴과 퍼지 추론

128.134.80.61	9	6	4	4	1	4	2
0.000000	0.969765	0.030235	0.000000				
0.888651	0.000000	0.111349	0.000000				
0.513410	0.486590	0.000000	0.000000				
0.266846	0.733154	0.000000	0.000000				
0.645161	0.000000	0.000000	0.000000				
0.757301	0.000000	0.242699	0.000000				
0.996364	0.003636	0.000000	0.000000				
rule# = 1, min_value = 0.266846							
:							

본 논문에서 제안한 방법론대로 목표 고객의 섹션별 접속 횟수에 근거하여 추론한 결과 해당 고객에게는 다음과 같은 종류의 광고가 가장 적합함을 알 수 있었다.

{(MAG, 0.769), (ECO, 0.769), (COM, 0.590), (SPO, 0.500), (ENT, 0.446), (RAN, 0.410)}

“MAG” 계열의 광고와 “ECO” 계열의 광고가 가장 적합하다는 퍼지 추론의 결과를 가지고 실제 어떤 광고가 선정될 것인가는 광고 퍼지 집합과의 해밍 거리를 구해보면 알 수 있는데 결과는 다음과 같다.

[표-3] 목표 고객에게 제시되어질 광고와 적합도

BG: MAG (ABC NEWS, 10117_abc0603_introl0.gif) matched with 3.125735
BG: ECO (Stock, 10852_musicblvd0610.gif) matched with 3.125735
BG: MAG (NewsDay, NWDY1042.gif) matched with 3.125735
BG: ECO (Investment, PCFN1059.gif) matched with 3.125735
BG: MAG (PUSH TECH, individual36.gif) matched with 3.125735

[표-3]은 목표 고객의 퍼지 집합과 해밍 거리가 가장 가까운 5개의 광고를 보여주고 있다. 계산되어진 해밍 거리는 3.125735 인데, 5개 광고의 해밍 거리가 같기 때문에 임의 순서로 목표 고객이 웹사이트에 접속할 때 제시하거나, 목표 고



객이 현재 서핑하고 있는 섹션의 종류에 따라 광고 제사 순서를 결정하는 것이 가능하다면 섹션 특성을 반영하는 것도 의미가 있다. 즉, 광고의 적합도가 같다면, 목표 고객이 "ECO" 섹션을 서핑할 때는 "ECO" 계열의 광고를 "MAG"에 선행하여 뿌리는 것이다. [표-3]에서 첫째 행의 "ABC NEWS" 는 광고의 내용이고, "10117\_abc0603\_intro10.gif"는 웹 광고 데이터베이스에 저장되어있는 해당 광고 파일의 이름이다.

## 5. 결론 및 향후 연구

웹 광고의 효과를 증진시키기 위해 본 논문에서는 기계학습을 활용하는 웹 광고 선택기를 제안했다. 웹 광고 선택기는 3개의 주요 기능을 가지고 있는데, 첫째는 고객 세분화 기능으로 웹 서버 로그 파일을 SOM을 이용, 마이닝함으로써 웹사이트 고객을 적절한 개수의 고객 세그먼트로 나누었다. 둘째는 퍼지 추론 기능인데 목표 고객의 전자신문 섹션별 방문 횟수를 입력 값으로 받아서 추론한 결과 퍼지 집합과 비퍼지화된 확정값을 얻어낸다. 세 번째 기능인 웹 광고 선택 기능은 목표 고객에 적절한 광고를 선택하기 위해 해밍 거리 척도를 계산한다.

본 논문에서 제안한 기법은 향후 연구를 위해 세 가지 방향에서 확장되어질 수 있다.

먼저, 고객의 방문 패턴을 클러스터링 하는데 이용된 기본 변수가 섹션별 사이트 방문 횟수에 대한 RFM (recency, frequency, monetary) 값이다. RFM 변수가 비교적 좋은 분석 결과를 도출해주지만, RFM 이외에 보다 나은 예측력을 가진 변수가 있다면 모델의 예측력을 향상시킬 수 있다는 측면에서 고객의 행태를 분석할 수 있는 다른 변수의 탐색이 필요하다. 또 본 논문에서는 고객의 섹션 방문 횟수만을 사용했는데 이것이 judgemental bias (cognitive importance가 반영되지 않았을 수 있음) 에 의한 활용 상의 한계를 가져올 수 있으므로 향후 연구에서는 극복해야할 과제가 된다.

둘째는 퍼지 규칙의 도출과 유지에 대한 이슈이다. 퍼지 추론은 충분한 수의 퍼지 집합과 규칙이 정의되어질수록 정확한 결과를 가져온다. 그러나 퍼지 집합의 수가 커질수록 퍼지 시스템의 복잡도가 증가되고 견고성은 감소한다. 뿐만 아니라 시스템의 수정 변경 또한 어려워진다. 따라서 적당한 수의 퍼지 집합과 규칙의 유지는 새로운 이슈가 될 수 있다.

셋째, 보다 장기간에 걸쳐 여러 형태의 온라인 서비스 제공자에 대한 경험적인 비교 분석이 수행되어진다면 유사한 웹 마케팅 전략의 평가를 도울 수 있을 것

이다.

## 참 고 문 헌

- [1] Aggarwal, C. C., Wolf, J. L., Wu, K-L., and Yu, P. S., "Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering", KDD-99 Proceedings, 1999, pp. 201-212.
- [2] Allen, C., Kania, D., and Yaeckel, B., Internet World Guide to One-To-One Web Marketing, John Wiley & Sons, Inc., Canada, 1998.
- [3] Berry, M. J. and Linoff, G., Data Mining Techniques for Marketing, Sales and Customer Support, John Wiley & Sons Inc., 1997.
- [4] Borchers, A., Herlocker, J., Konstan, J., and Riedl, J., "Ganging up on Information Overload", Computer, Vol. 31, No. 4, 1998, pp. 106-108.
- [5] Buchner, A. G. and Mulvenna, M., "Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining", SIGMOD Record, Vol. 27, No. 4, 1998, pp. 54-61.
- [6] Chen, M. S., Park, J. S., and Yu, P. S., "Efficient Data Mining for Path Traversal Patterns in Distributed Systems", Proceedings of the 16th International Conference on Distributed Computing Systems, May 27-30, Hong Kong, 1996, pp. 385-393.
- [7] Cooley, R., Mobasher, B., and Srivastava, J., "Web Mining: Information and Pattern Discovery on the World Wide Web", Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence, Newport Beach, California, Nov. 3-8, URL: <<http://www-users.cs.umn.edu/~mobasher/webminer/survey/survey.html>>, 1997.
- [8] Cooley, R., Mobasher, B., and Srivastava, J., "Data preparation for Mining World Wide Web Browsing Patterns", International Journal of Knowledge and Information Systems, Vol. 1, No. 1, 1999, pp. 5-32.
- [9] Cox, E., The Fuzzy Systems Handbook: A Practitioners Guide to Building, Using, and Maintaining Fuzzy Systems, AP Professional, MA, 1994.
- [10] Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., Advances in Knowledge Discovery and Data Mining, AAAI press, 1996.
- [11] Kalakota, R. and Whinston, A. B., Electronic Commerce: A Managers Guide, Addison-Wesley, MA, 1996.
- [12] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L. R., and Riedl, J., "GroupLens: Applying Collaborative Filtering to Usenet News", Communications of the ACM, Vol. 40, No. 3, 1997, pp. 77-87.

- [13] Ma, M., Kandel, A., and Friedman, M., "A new approach for defuzzification", *Fuzzy Sets and Systems*, Vol. 111, No. 3, 2000, pp. 351-356.
- [14] McNeill, F. M. and Thro, E., *Fuzzy Logic: A Practical Approach*, Academic Press, Inc., MA, 1994.
- [15] Mena, J., *Data Mining Your Website*, Butterworth-Heinemann, MA, 1999.
- [16] Peppard, J., "Customer Relationship Management (CRM) in Financial Services", *European Management Journal*, Vol. 18, No. 3, 2000, pp. 312-327.
- [17] Peppers, D. and Rogers, M., *The One to One Future: Building Relationships One Customer at a Time*, Bantam Doubleday Dell Publishing, 1997.
- [18] Resnick, P. and Varian, H. R., "Recommender Systems", *Communications of the ACM*, Vol. 40, No. 3, 1997, pp. 56-58.
- [19] Schafer, J. B., Konstan, J. A., and Riedl, J., "E-commerce Recommendation Applications", *Data Mining and Knowledge Discovery*, Vol. 5, No. 1-2, 2001, pp. 115-153.
- [20] Zadeh, L. A., Fu, K-S., Tanaka, K., and Shimura, M., (Ed.), *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, Academic Press, Inc., MA, 1975.
- [21] Zeff, R. and Aronson, B., *Advertising on the Internet*, John Wiley & Sons, Inc., Canada, 1997.

## **Personalized Advertising Techniques on the Internet for Electronic Newspaper Provider**

Sung Ho Ha

### **Abstract**

The explosive growth of the Internet and the increasing popularity of the World Wide Web have generated significant interest in the development of electronic commerce in a global online marketplace. The rapid adoption of the Internet as a commercial medium is rapidly expanding the necessity of Web advertisement as a new communication channel. If proper Web advertisement could be suggested to the right user, then effectiveness of Web advertisement will be raised and it will help company to earn more profit. So, this article describes a personalized advertisement technique as a part of intelligent customer services for an electronic newspaper provider. Based on customers history of navigation on the electronic newspapers pages, which are divided into several sections such as politics, economics, sports, culture, and so on, appropriate advertisements (especially, banner ads) are chosen and displayed with the aid of machine learning techniques, when customers visit to the site. To verify feasibility of the technique, an application will be made to one of the most popular e-newspaper publishing company in Korea.

◆ 저자소개 ◆



하성호 (Sung Ho Ha)

연세대학교 경영학과를 졸업하고 (1990), (주)LG-EDS에서 시스템 분석가로 활동하다가 한국과학기술원 산업공학과에서 경영정보시스템 전공으로 석사 (1998), 박사 (2001) 학위를 취득하였다. 현재는 한국과학기술원 산업경영연구소 연구원으로 재직하고 있다. 주요 관심 분야는 전자상거래, 인터넷 마케팅, 고객관계관리, 데이터마이닝, 의사결정지원시스템 등이다.

E-mail : hash@major.kaist.ac.kr