

# Decision Tree 를 이용한 고객 취향 관리 시스템

최중훈<sup>0</sup>, 이은, 공은배  
충남대학교 컴퓨터공학과  
{jhoon, elee, keb}@ce.cnu.ac.kr

## Customer Relationship Management System using Decision Tree

Jong-Hoon Choi<sup>0</sup>, Eun Lee, Eun-Bae Kong  
Dept. of Computer Engineering, Chungnam National University

### 요 약

인터넷의 활성화로 많은 사람들이 인터넷을 이용하고 이에 따라 인터넷을 이용한 서비스도 홍수를 이루고 있다. 이에 따라 인터넷을 상업적 목적으로 사용하는 서비스도 증가하고 있다. 그러나 많은 인터넷 서비스들이 고객들에게 획일적이고 일률적인 서비스만을 제공한다. 각각의 고객에게 취향과 관심 분야에 따른 차별화 된 서비스가 필요로 한다. 각 고객에게 1 대 1 로 차별화 된 service 를 제공하기 위해서 먼저 각 고객을 구별하고 그 고객의 취향과 관심분야의 파악을 위해서 인터넷에서의 행동을 관찰한다. 또한 고객의 관리를 위해 고객을 필요에 따라 그룹화 하고, 고객과 직접 접촉을 통해 고객 정보를 파악할 수도 있다. 파악된 고객 정보의 효율적 저장과 분석을 위해서 decision tree 를 이용해 학습을 한다. 고객의 행동의 특성상 incremental 한 학습 알고리즘을 사용하며 고객의 선호도를 이용한 decision tree 를 이용한다. 학습된 결과를 이용해서 1 대 1 서비스를 제공함으로써 고객에서 편리성을 제공하고 서비스에 대한 친밀감과 고객의 흥미를 유발할 수 있다.

### 1. 서론

인터넷의 사용은 이제 더 이상 컴퓨터를 전공하는 사람의 전유물이 아니다. 이제 남녀노소를 불문하고 모두에게 인터넷은 친숙한 존재가 되어 가고 있다. 인터넷의 대중화에 따라서는 인터넷을 통한 각종 서비스들 또한 홍수를 이루고 있다. 이러한 인터넷의 확산에 따라 인터넷을 통해 제품을 판매하는 쇼핑몰이나 대금을 받고 서비스나 정보를 판매하는 등의 인터넷을 상업적으로 이용하는 서비스들이 많이 생겨나고 있다. 그러나 이러한 서비스들은 모두 고객에게 똑같은 서비스를 제공한다. 각각의 고객에 대한 차별화 된 서비스를 제공하는 곳은 거의 없다. 고객에게 차별화 된 서비스를 제공하기 위해서는 고객의 취향이나 관심분야 등을 파악하는 것이 필요하게 된다. 고객의 취향을 파악하는 방법으로는 다음 2 가지 방법이 있다.

- Customization : 고객 스스로가 자신의 정보와 관심분야를 입력하도록 하고 이를 바탕으로 서비스를 한다.
- Personalization : 능동적으로 고객의 행동을 관찰하여 각 고객의 관심 분야와 취향을 파악하여 서비스를 한다.

이러한 2 가지 방법은 모두 장단점이 존재 한다. Customization 은 고객 정보를 정확히 얻을 수 있고 고객 정보 파악을 위한 별도의 오버헤드가 필요로 하지 않는다. 그러나 고객 스스로 자신을 정보를 입력해야 하므로 고객이 자신의 정보를 밝히기 전에는 서비스를 할 수 없으며 고객이 자신의 정보를 밝히기 꺼리는 경우 서비스가 어렵게 되며 고객의 관

심분야와 취향의 변화에 민감하게 반응 할 수 없다.

Personalization 은 능동적으로 고객의 취향을 파악하므로 고객이 인지하지 못하는 사이에도 차별화 된 서비스가 가능해지고 고객의 변화하는 취향과 관심분야에도 쉽게 반응할 수 있다. 하지만 personalization 은 고객 행동을 관찰하고 분석하는 오버헤드가 더 필요로 하게 된다.

본 논문에서는 Personalization 에 초점을 두고 1 대 1 서비스를 하는 방법에 대해서 살펴본다. 고객이 인터넷에서 행하는 행동을 관찰하고 이를 분석하여 고객의 취향과 관심 분야를 파악하고 이를 바탕으로 각 개인에게 personalize 된 서비스를 제공한다. 또한 관찰한 고객의 행동을 효율적으로 저장, 분석하기 위해서 decision tree 를 이용해 학습을 하는 기술에 대해서 살펴 보도록 한다.

### 2. 1 대 1 마케팅

1 대 1 마케팅[1] 이란 각 고객들에게 1:1 로 차별화 된 서비스를 제공하는 것을 말한다. 예를 들어 인터넷 쇼핑 물에서는 고객이 자주 쇼핑하는 상품이나 관심을 갖는 제품을 각 개인별로 관찰을 하고 있다가 그 고객이 접속을 했을 경우 관심 있는 제품을 먼저 보여줄 수 있다. 인터넷 신문에서는 고객이 관심을 가지는고 자주 읽는 기사를 파악하고 있다가 그 고객이 접속하면 그 기사를 먼저 서비스 하는 것을 말한다. 이러한 1 대 1 서비스를 제공하기 위해서 4 단계가 필요로 한다. 고객의 구별, 고객의 그룹화, 고객과의 대화, 고객에게 1 대 1 서비스 제공의 4 단계에 대해서 살펴보면 다음과 같다.

\* 본 연구는 과학기술부/한국과학재단 지정 충남대학교 소프트웨어연구센터의 지원에 의한 것입니다.

**2.1 고객의 구별(identify)**

각 개인별로 서비스를 제공하기 위해서는 각각의 고객들을 다른 고객들로부터 구별해야 한다. 고객의 취향을 파악하기 위해서 고객의 인터넷에서의 행동을 관찰한다. 고객을 구별해 내는 것은 관찰된 행동이 누구의 행동인가를 파악하기 위해서이다. 이러한 행동을 보인 고객이 누구인지 알아야 그 고객이 다시 접속했을 때 그 고객에게 1 대 1 서비스를 제공할 수 있다. 즉 관찰된 행동과 고객을 연결하기 위해서 고객의 구별이 필요로 하게 된다.

고객을 구별하기 위해서는 고객이 제공하는 정보를 이용할 수도 있고 특정 하드웨어를 이용할 수도 있다. 특정 하드웨어를 이용하는 방법은 IP address, hard disk serial number, smart card 등을 이용하는 것이다. 이것은 익명으로 고객을 구별할 수 있으나 하드웨어에 의존하므로 이전에 방문한 고객과 동일인이라고 확신할 수 없다. 고객이 자신의 정보를 알리거나 혹은 고객과의 대화를 통해 알아 낸 경우에는 고객의 이름, 아이디, 패스워드, 주소, 전화번호, 이메일 주소 등으로 다른 고객으로부터 그 고객을 구별해 낼 수 있다. 또는 쿠키를 이용하여 고객을 구별 할 수도 있다. 흔히 사용되는 방법이 아이디와 암호를 이용하는 방법이다. 고객 정보를 이용해 구별하므로 고객을 확실하게 구별할 수 있으나 고객이 접속 할 때 마다 자신의 정보를 입력하는 불편이 따른다.

고객의 익명을 보장하면서 정확히 구별을 하는 것이 가장 좋은 방법이나 익명으로 고객의 행동을 관찰하고 1 대 1 서비스를 제공하는 데는 한계가 있다. 그러나 인터넷 사용자들은 자신이 누구이며 자신의 정보가 남에게 알려지는 것을 꺼리는 사람도 있다.

고객의 구별은 최대한 고객에게 편리한 방법으로 구현되어야 하며 적은 고객 정보로 정확히 그 고객을 구별해 낼 수 있는 것이 좋다. 이렇게 고객을 구별 하여 그 고객이 인터넷에서 무엇을 하는지 어떤 곳에 관심이 있는지 고객의 행동을 관찰하여 고객의 구별 정보와 함께 저장, 분석 해야 한다.

**2.2 고객의 분류(differentiate)**

고객의 정보와 그 고객의 인터넷에서의 행동의 수집하였으면 그러한 고객들을 관리하기 위한 방법으로 고객을 분류할 수 있다. 수 많은 고객들을 1 명씩 관리하기란 쉽지 않으므로 많은 고객들을 그룹으로 분류해서 관리하는 것이 훨씬 효율적이다. 고객을 분류하는 방법에는 여러 가지가 있을 수 있다. 고객의 profile 에 의해서 분류를 할 수 있고 서비스를 제공하는 제공자의 필요에 따라 분류될 수도 있다. 고객의 학력이나 직업, 신분 등으로 분류하거나 비슷한 취향이나 관심분야 등으로, 또는 고객의 인터넷 사용시간이나 서비스 이용의 충성도 등으로 고객을 그룹화 할 수 있다.

이러한 고객의 분류는 단순히 고객 관리의 효율성 뿐만 아니라 서비스의 질적인 향상에도 영향을 미칠 수 있다. 예를 들어 어떤 고객의 행동이 관찰되었을 경우 비슷한 취향이나 같은 그룹의 다른 고객도 곧 같은 행동을 보일 것이라는 예측이 가능하게 해준다.

**2.3 고객과 대화(interaction)**

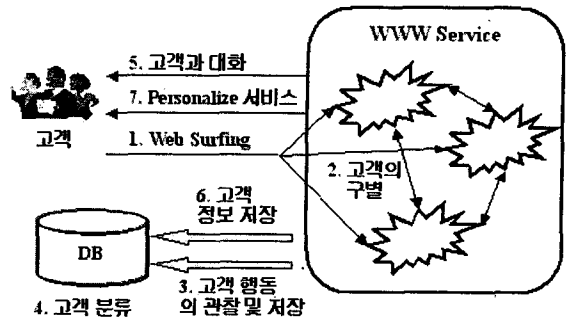
고객과의 대화는 서비스 제공자가 고객과 직접 접촉하는 과정이다. 고객을 구별하고 고객의 행동을 관찰하는 것으로 얻을 수 있는 고객의 정보에는 한계가 있다. 고객과의 접촉은 고객과 직접 대화하는 것이므로 좀 더 정확하고 많은 정보를 수집할 수 있다. 따라서 고객과 대화를 하기 전에 어떤 정보를 얻을 것인가를 정하고 그 행식을 미리 itemize 해야 한다. 또한 고객의 정보뿐만 아니라 서비스에 대한 고객의 만족도나 불만 등을 파악할 수도 있다. 고객과의 대화는 곧 정보가 되고 정확한 정보는 곧 이익이 될 수 있다. 즉 "Dialogue = Profit" 이라는 등식이 성립하게 된다.

고객과의 대화에서 중요한 것은 고객이 자발적으로 대화에 참여하도록 해야 한다는 것이다. 서비스 제공자가 억지로 접촉을 시도하거나 강압에 의해 대화에 참여 시킨다면 그 고객은 오히려 서비스에 대한 거부감을 가질 수도 있다. 따라서 고객 스스로가 서비스 제공자와 자연스럽게 대화할 수 있도록 해야 한다. 또한 고객과 대화를 할 경우 고객으로부터 얻는 정보가 어디에 쓰일 것인지 프라이버시 침해 문제가 없는지 고객에게 분명히 밝혀야 하며 접촉 후 고객이 자신에게 이익이 있다고 느끼도록 해 주어야 한다.

**2.4 1 대 1 서비스**

1 대 1 서비스 단계는 수집한 고객의 profile 과 인터넷에서의 행동을 이용해서 고객에게 서비스를 제공하는 단계이다. 고객의 취향이나 관심분야에 따라서 각각의 고객에게 차별화 된 서비스를 제공한다. Personalize 된 서비스는 고객에게 편리성을 제공하고 고객이 서비스에 대해 친밀감을 가질 수 있게 한다.

앞의 4 단계를 이용한 personalize 서비스를 그림으로 표시하면 [그림 1]과 같다.



[그림 1] Personalize 서비스

고객은 먼저 서비스를 제공하는 곳에 접속한다. 이때 서버쪽에서는 개개인의 고객들을 구별 한다. 고객이 로그인이나 특별한 과정 없이 익명으로 구별 할 수도 있고 처음 접속할 때 고객 아이디, 패스워드나, 이메일, 주민등록번호 등의 고객 정보로 구별 할 수도 있다. 고객이 접속 후 인터넷에서 surfing 을 하는 동안 그 과정들을 관찰을 하면서 저장한다. 저장된 정보를 바탕으로 고객들을 분류 하고 필요에 따라서 고객과 직접 접촉하여 대화를 통해 고객 정보를 수집하거나 고객의 만족도나 불만 등을 파악한다. 수집한 정보들을 바탕으로 각각의 고객에게 personalize 된 서비스를 제공한다.

이러한 1 대 1 마케팅 은 1 회성의 단순한 관계가 아니라 고객들로 하여금 자신이 지속적으로 특별하게 대우 받는다는 생각을 갖게 하고 서비스 제공자와 고객 사이에 1 대 1 의 관계가 성립된다고 볼 수 있다.

**3. 고객 행동의 관찰과 분석**

단순히 1 명의 고객의 행동을 관찰하고 저장하는 일은 간단한 일이다. 그러나 고객이 무수히 많아 질 경우 각 고객이 방문한 사이트나 인터넷에서 하는 행동 등을 모두 저장하는 것은 저장 용량이나 그 내용을 분석하는데 있어서 매우 비효율적이다. 고객의 방문 정보나 행동을 효율적으로 분석해 내는 것에 대해서 살펴본다.

**3.1 고객 행동의 관찰**

고객이 서버에 접속하여 각각의 페이지를 방문 할 때 서버에서는 각 고객 마다 각각 구별화를 하고 그 고객이 어떤 페이지를

이지를 방문하고 얼마나 오래 그 페이지에 머무르는지 관찰한다. 고객 행동의 관찰에는 쿠키를 이용하거나 웹 서버에서 제공하는 log 를 이용하거나 특정 프로그램을 작성하는 등의 여러 가지 방법이 있을 수 있다.

고객의 행동을 관찰하고 이를 효율적으로 분석하기 위해서는 먼저 고객이 방문한 페이지에 대한 정보가 필요하다. 이를 위해서 각각의 웹 페이지에 대해서 classify 를 해 둔다. 예를 들어 신문사의 사이트 일 경우 모든 기사가 그대로 학습을 위한 training example 이 될 수 없으므로 기사를 classify 를 한다. "정치, 경제, 사회, 문화, 스포츠..." 등과 같이 기사에 해한 class 들을 정의해 놓는다. 그리고 각 기사에 대해서 그 기사가 어떤 class 에 속하는지 태그를 붙여 놓는다. 고객이 신문사에서 기사를 읽을 경우 각 페이지의 기사를 그대로 저장하지 않고 그 기사가 속하는 class 를 저장한다. 고객의 행동을 관찰하기 전에 웹 페이지에 대한 classify 작업이 선행되어야 한다. 이러한 작업이 선행되지 않는다면 고객이 방문한 페이지에 대해서 그 페이지가 무엇과 관계된 페이지 인가를 판단하는 작업이 더 필요로 하게 된다. 본 논문에서는 모든 페이지에 대한 classify 가 되어 있고 각 페이지에는 class 에 대한 태그가 되어 있다고 가정한다.

### 3.2 Decision tree 를 이용한 학습

고객의 행동을 관찰한 정보를 파일이나 또는 DB 등으로 저장해 해야 한다. 단순히 저장만 한다면 저장 용량도 커질 뿐만 아니라 그 정보에 대한 분석이나 정보를 이용해서 서비스를 하는데 어려움이 따를 수 있다. 그러나 각각의 고객 마다 방문한 페이지와 그 페이지에 접속한 횟수 체류 시간 등을 모두 저장하면 그 용량이 커질 뿐 아니라 이를 이용해 고객의 취향을 파악하기 어렵다. 따라서 효율적인 분석과 적용을 위해서 관찰한 고객의 행동을 decision tree 를 이용해 학습을 한다. 신문사 사이트를 예로 들면 앞에서 말한 것처럼 모든 기사들에 대한 classify 가 되어있으므로 고객이 정치에서 대통령과 관계된 기사를 10 분, 스포츠에서 프로야구와 관계된 기사를 12 분 동안 보았다면 다음과 같이 정보를 수집할 수 있다.

- [정치->대통령, 10]
- [스포츠->프로야구, 12]

이렇게 수집된 정보를 training example 로 해서 decision tree 에 학습을 한다.

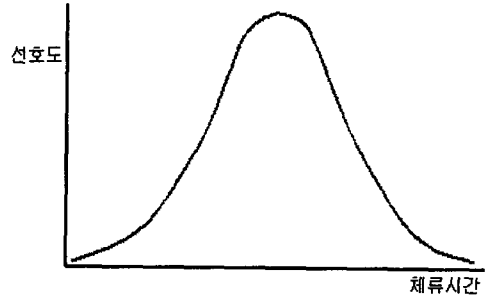
Decision tree 로 학습을 하기 위해서는 많은 양의 training example 이 있어야 한다. 그러나 고객의 행동의 관찰로 얻어지는 데이터는 한번에 많은 양을 수집할 수 없다. 적은 양의 데이터를 지속적으로 꾸준히 관찰을 해야 한다. 따라서 많은 training example 로 한번 학습을 하는 알고리즘이 아닌 지속적으로 입력되는 training example 을 계속 학습할 수 있는 incremental 한 학습 알고리즘[2] 이 필요하다.

### 3.3 선호도 값을 이용한 decision tree

앞에서 예로는 신문사 사이트에서 고객이 기사를 보는 시간은 그 기사에 대한 선호도로 볼 수 있다. 그러나 고객이 기사를 보는 시간이 너무 짧거나 너무 길 경우에는 오히려 선호도가 없다고 볼 수 있다. 기사를 보는 시간이 너무 짧은 것은 기사에 흥미가 없거나 다른 기사를 보기 위해 지나치는 것으로 볼 수 있으며 기사를 보는 시간이 너무 긴 경우는 페이지를 열어둔 채로 다른 작업을 하는 것이라 볼 수 있다. 따라서 체류시간과 선호도관계를 그래프로 표현하면 [그림 2]와 같다.

### 3.4 학습 결과의 분석과 적용

고객의 행동을 관찰하고 decision tree 에 학습을 했으면 이를 이용해서 고객에게 personalize 된 서비스를 제공해야 한다. 고객의 선호도를 decision tree 서 파악할 수 있으므로 제공하고자 하는 서비스를 그 선호도에 따라서 서비스를 할 수 있다. 신문사의 경우 그날의 기사들이 어느 class 에 속하는가 정해



[그림 2] 고객의 체류시간과 선호도

져 있고 각 고객의 class 에 대한 선호도가 있으므로 각각의 기사에 그 고객에 대한 순위를 정할 수 있다. 이렇게 정해진 순위에 따라 고객에게 기사를 서비스 할 수 있다. 하지만 각 기사에 대해서 신문 편집자가 중요도를 부여 할 수 있다. 중요한 기사는 1 면 톱기사가 되고 별로 중요하지 않은 기사는 다른 면에 나타날 수 있다. 편집자 즉, 서비스 제공자가 각각의 기사에 대해서 중요도를 부여 한 것과 각 고객의 선호도를 어떻게 절충하는가 하는 문제가 발생한다. 고객의 선호도에 따른 기사와 편집자의 중요도에 따른 기사를 따로 서비스 할 수도 있고 고객의 선호도와 편집자의 중요도를 이용해 다시 선호도를 계산해서 서비스를 할 수도 있다. 고객의 선호도에 따라 모든 기사에 순위를 정할 수 있으나 고객이 접속할 때마다 그 순위를 계산 해야 한다. 기사가 많을 경우 이를 계산 하는 것도 시간이 걸리는 문제이다. 따라서 모든 기사에 순위를 정하지 않고 선호도가 몇 개의 class 에 대한 기사만 순위를 정해 서비스 하고 나머지 기사는 편집자의 중요도에 의해 서비스를 할 수도 있다.

## 4. 결론 및 향후과제

1 대 1 서비스를 위해서 고객을 구별하고, 고객의 행동을 관찰하고 비슷한 성향의 고객을 그룹화하고, 고객과 접촉을 통한 정보 수집 등의 단계를 거쳐 수집한 정보를 이용해 personalize 된 서비스를 제공한다. 수집한 정보의 효율적인 분석과 서비스를 위해서 decision tree 를 이용한다. decision tree 알고리즘은 incremental 한 알고리즘을 사용하며 decision tree 에 선호도를 추가하여 서비스를 제공한다.

Personalize 된 서비스를 제공함으로써 고객에게 편리성을 제공하고 맞춤 서비스를 할 수 있다. 고객이 자신이 원하는 것을 찾아가는 번거로움을 줄이고 취향에 따라 원하는 것을 미리 파악하여 서비스 한다. 이러한 서비스는 고객이 더욱 친밀감을 가질 수 있고 흥미를 유발할 수 있다.

고객의 정보를 수집하는 과정에서 고객이 거부감을 갖거나 자신의 정보가 유출되는 것에 대한 철저한 대비가 필요로 된다. 고객 정보의 수집은 고객이 자신의 정보를 자발적으로 제공할 수 있도록 해야 하며 개인의 프라이버시가 침해 당하지 않도록 고객 정보에 대한 보안과 서비스를 제공할 때도 역시 정보 보안이 이루어져야 한다.

## 5. 참고 문헌

- [1] Don Peppers, Martha Rogers, and Bob Dorf, "One to one Fieldbook" The Complete Toolkit for Implementing a 1To1 Marketing Program, Bantam Doubleday Dell Pub(P), 1998
- [2] PAUL E. UTGOFF, "Incremental Induction of Decision Trees", 1989
- [3] Tom M. Mitchell, "Machine Learning", The McGraw-Hill Companies, 1997