

사용자 성향 분석을 통한 광고시스템¹

송강수⁰ 정해권⁰⁰ 이원돈⁰

⁰충남대학교 컴퓨터과학과

{ksong⁰, wdlee}@cnu.ac.kr, haggai⁰⁰@sorient.com

Advertisement System Based On User Preference

Kang Soo Song⁰ Hae Gwon Jeong⁰⁰ Won Don Lee⁰

Dept. of Computer Science Chungnam National University

요약

사용자의 성향을 분석하여 그 결과를 이용하는 기업 활동은 매우 유용하고 효과적이다. 이미 분석된 사용자의 성향에 대한 이용 분야는 매우 다양하다. 그 중에서 사용자와 기업 간의 상호 작용이 많은 부분이 광고 분야이다. 사용자의 성향을 알고 있고 그 사용자에게 광고를 하였을 때 광고 효과를 최상으로 이끌어 낼 수 있다. 기업은 광고비를 지출하면서 그것이 최상의 효과를 가지기를 원한다. 사용자는 자신이 원하지 않는 광고는 바로 폐기 한다. 이러한 두 가지의 딜레마에서 양쪽의 집단을 동시에 만족시킬 수 있는 기법이 본 논문에서 제안하고 실험한 사용자 분석을 통한 광고 시스템이다. 이러한 시스템을 위해서 사용자의 성향을 찾아내기 위한 데이터의 집합에서 법칙을 적용하여 사용자의 성향에 맞게 분류한다. 그 지정한 분류 안에서 광고를 광고비를 기반으로 한 결정 방법을 적용하여 가장 공평한 광고가 나가게 시스템을 구성하고 실험하였다.

1. 서론

본 논문에서 제안하고자 하는 것은 인터넷을 매개체로 하는 모든 프로그램에서 이용이 가능한 광고 시스템을 제안한다. 이 광고 시스템은 기본적으로 사용자의 성향을 분석하여 그 사용자에게 맞는 분류를 확인하고 그 분류에 맞는 광고를 선택적으로 보내는 것이다. 이것을 통하여 사용자가 얻게 되는 것은 기본적으로 자신이 현재 관심 있어 하는 분야의 광고를 보게 됨으로써 필요한 정보를 검색하는데 시간을 들일 필요가 없어지게 되며, 광고주 입장에서는 관심을 갖고 있는 사용자를 대상으로 광고를 실시하므로 비용을 줄이고 광고의 효과를 높일 수 있다. 이러한 것을 위하여 기존에 이용하는 방식은 CRM(Customer Relationship Management)이다. 이것은 경영학적인 측면에서 고객의 자원(관계)을 경영하는 기법으로써 기업이 경영하는데 필요한 고객 측의 정보를 최대한으로 이용하여 기업이 경영을 위하여 취하는 행위에 대한 정보를 제공하는 것이다. 이러한 작업은 사용자의 요구하는 바를 어떠한 방법을 통해서든지 받아들여서 그 받아 들인 정보를 분석하여 적절한 행위를 요청한다. 이러한 절차는 시간이 너무 오래 걸릴 수 있다. 그래서 새로운 기법

인 E-CRM(Electronic – Customer Relationship Management)가 앞으로 이용될 가능성이 높아지고 있다. 이것은 CRM의 확장된 것으로써 CRM의 정보 수집하는 부분을 Electronic 기법을 이용하여 수집하여 기업의 활동까지 모두 자동화 하는 시스템을 의미한다 [1].

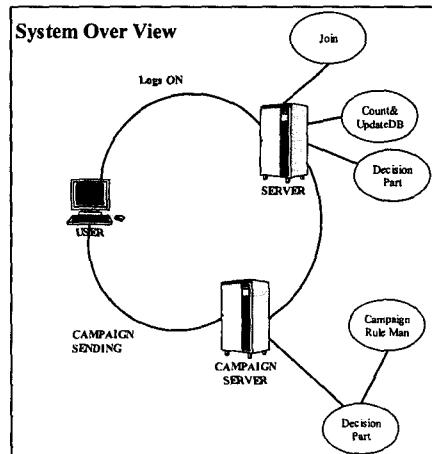


그림 1: 제안된 시스템의 모델

¹ 본 논문은 2003년도 충남대학교 발전기금 재단의 지원에 의하여 연구되었음

2. 가중치와 난수를 이용한 선택 방법

이러한 것을 확인하고 실험하기 위하여 기본적으로 몇 가지를 가정하였다. 우선 각각의 모든 사용자는 그들의 기본적인 정보인 성별, 나이, 거주지역 등을 제공하고 있다. 그리고 그들은 자신의 컴퓨터를 이용하여 정보검색과 방문기록을 가지고 있으며 본 시스템을 위하여 제공하기로 했다고 가정했다. 또한 광고는 모두 광고비를 기준으로 하여 각각 일정한 비율로 광고가 나가게 된다. 우선 사용자는 프로그램을 실행하면 인터넷 방문 기록인 URL(Uniform Resource Locator)를 검색하여 각각의 URL을 분류한다. 이때 적용하는 기준은 본 시스템에서 이용할 수 있는 분석엔진에서 수행한다. 각각의 URL은 세 개의 단계의 분류로 분류되어 이것은 이미 분석되어 있는 것을 이용한다. 최상의 분류에는 열두 가지 분류가 있으며 그것의 아래에는 최대한 세 단계의 깊이로 되어 있다. 그리고 광고주는 자신의 광고를 광고 트리 중 한곳에 자신의 광고를 위치하게 된다. 사용자의 URL 분석 결과를 가지고 광고를 결정하여 광고를 보내 주는 형태이다. 다음의 그림 1은 이러한 시스템의 구성과 동작 모습을 보여주고 있다.

사용자 정보 분석을 통한 광고 분류의 결정시 여러 개의 분류 중 한 개를 선택해야 한다. 이때 사용자측의 가중치는 전체의 분류 정보 중 어떠한 분류정보가 어느 정도를 가지고 차지하고 있는지를 기반으로 하여 가중치를 만들고 거기에서 난수를 이용하여 하나를 선택한다. 다음의 알고리즘과 표는 이러한 과정을 보여주고 있다.

```

Weight = 0.0
Total # of counts = A
Do until the last category
Count of a category = a[i]
Ratio = a[i]/A
Weight[i] = Weight[i] + ratio
Loop

```

Weight determination algorithm

여기에서 weight는 가중치이며 난수를 이용하여 결정할 때 용이하게 하기 위하여 weight 범위를 나타내었다. 이러한 알고리즘을 이용하여 실제의 실험에서는 다음의 표를 이용하였다.

표 1: 분류 별 가중치와 가중치 범위

Category Name	Category Count (E)	Category Ratio (Φ)	Category Weight Range
Category A	α	$\frac{\alpha}{\sum E}$	$[0, \frac{\alpha}{\sum E})$
Category B	β	$\frac{\beta}{\sum E}$	$[\frac{\alpha}{\sum E}, \frac{\alpha+\beta}{\sum E})$
...
Category X (last)	χ	$\frac{\chi}{\sum E}$	$[1 - \frac{\chi}{\sum E}, 1]$

이 정보를 이용하여 난수를 이용하여 한 개의 분류를 결정하는 모습은 다음과 같다.

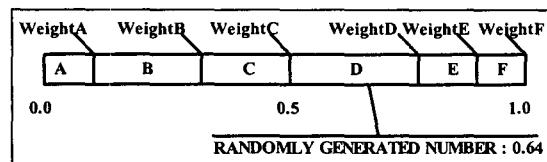


그림 2: 난수를 이용하여 한 개의 분류를 결정하는 모습

3. 광고가 나가는 과정

2장에서와 같이 분류를 선택하고 난 후 그 분류아래에 있는 광고를 선택하는 것은 전체의 광고비와 광고주의 요청에 의한 몇 가지 요구사항에 맞추어 결정 해야 한다. 만약에 결정된 분류에 광고가 없을 경우 그것보다 상위 분류로 가서 광고를 결정했다. 그리고 광고를 결정한 이후에 광고가 나가기 바로 직전에 광고주의 요구사항에 맞는지 확인한 후 광고가 나갔다. 광고가 나간 후에는 광고의 비용이 줄어들며 그것을 기반으로 하여 광고의 가중치는 모두 다시 계산하였다. 이때 적용되는 알고리즘은 3.1에서 제안한 것을 이용한다. 다음의 표 2는 이러한 것을 적용한 예를 보여주는 표이다.

표 2

adverID	Total investment	Expense per hit	RATIO	Weight Range
Adver01	100	10	0.16	[0,0.16)
Adver02	200	10	0.33	[0.16,0.49)
Adver03	300	10	0.50	[0.49,1]

4. 실험 및 결과

실험을 위하여 사용자의 기본 정보는 임의로 지정하였으며 사용자의 컴퓨터는 본 학과의 실습실에 있는 컴퓨터가 각각 하나의 사용자라고 가정하고 실험을 하였다. 다음의 표는 사용자의 컴퓨터에서 추출한 자료의 분류 전체 중 일부를 선택하여 나타낸 것이다.

표 3: 추출된 분류

The Gathered Category Data Ratio(%)						
Category \ UserID	pc008	pc009	pc013	pc017	pc018	pc021
A(1)	8.78	7.25	16.52	5.38	7.07	7.65
B(2)	9.46	7.73	0.00	30.93	12.45	15.51
C(3)	0.00	0.97	0.00	0.41	0.48	0.87
D(4)	43.24	10.14	25.22	13.20	23.02	27.61
E(5)	2.03	10.83	8.57	10.72	8.45	9.83
F(6)	0.00	0.48	0.00	0.00	0.21	0.34
G(7)	8.11	28.50	21.74	8.25	9.52	4.97
H(8)	0.68	4.63	9.57	6.60	5.74	3.29
I(9)	3.38	10.14	7.83	2.68	9.23	6.97
J(10)	1.35	0.00	0.00	1.86	2.31	0.39
K(11)	12.16	0.00	5.22	0.82	4.20	2.13
L(12)	8.11	3.86	0.87	5.77	10.66	13.63
M(13)	1.35	4.63	0.00	12.78	5.43	4.86
N(14)	1.35	10.83	3.48	0.82	1.25	1.95

이 표를 기반으로 하여 광고를 선택하고 광고를 보냈다. 그리고 이러한 양상을 시간의 흐름에 따라 나타내기 위하여 각각의 열두 개의 분류에 있는 총 광고 금액의 변화를 나타낸 그래프이다.

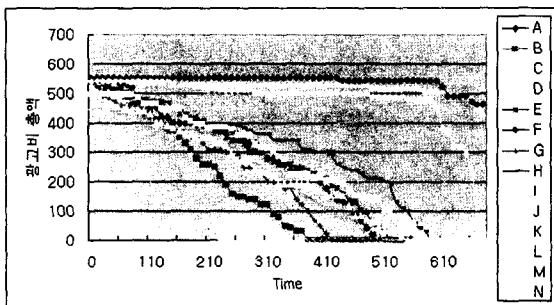


그림 3: 시간의 흐름에 따른 각 분류 별 광고비의 변화

5. 결론

본 논문에서는 광고를 위하여 사용자의 성향을 분석하고 그것을 기반으로 광고하는 시스템을 실험하였다. 그림 3에서 보듯이 이러한 기법을 도입하면 비용을 들인 광고가 원하는 사용자에게만 보내어 지고 그렇게 함으로써 광고의 효과가 기존의 방법보다 더 좋다는 것을 알 수 있다. 이러한 사용자의 성향을 파악하기 위하여 여기에 좀더 이용할 수 있는 부분으로는 자주 쓰는 프로그

램의 분석이 여기에 추가가 되면 좀더 확실한 시스템으로 갈 수 있을 것으로 사료된다. 이러한 기법이 광고 시스템에서만 이용하는 것이 아니라 어떠한 사이트에서 각 사용자의 성향에 따라 개인화 된 서비스를 제공할 수 있는 기반을 이루게 된다. 또한 유비쿼터스 환경에서 이러한 기법이 적용되면 더욱더 효율적으로 사용자의 성향을 분석하고 움직이면서 정보를 수집하며 그것을 분석하는 기법의 부분으로 이용할 수도 있다.

6. 참고문헌

- [1] Shan L. Pan, Jae-Nam Lee, "Using e-CRM for a unified view of the customer" Communications of the ACM, Volume 46 Issue 4, April 2003
- [2] 이용배, "장르기반의 문서 자동분류 방법", 충남대학교 대학원, 2003
- [3] Eric J. Glover, Kostas Tsoutsouliklis, Steve Lawrence, David M. Pennock, Gary W. Flake, "Using web structure for classifying and describing web pages" Proceedings of the eleventh international conference on World Wide Web, May 2002
- [4] Magdalini Eirinaki, Michalis Vazirgiannis, "Web mining for web personalization" ACM Transactions on Internet Technology (TOIT), Volume 3 Issue 1, February 2003
- [5] 이운상, "고객관계관리(Customer relationship management CRM) 시스템이 경영성과에 미치는 영향" 중앙대학교 대학원, 2001
- [6] Wendy Gersten, Rudiger Wirth, Dirk Arndt, "Predictive modeling in automotive direct marketing: tools, experiences and open issues" Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, August 2000
- [7] 조성훈, "Data mining for CRM", 한국데이터베이스 학회, 2001년 춘계
- [8] 김광용, "Web Data Mining과 E-CRM", 한국지능정보시스템 학회, Vol.1 No.1, 2001
- [9] 이훈영, "CRM 가치 창조를 위한 Marketing Paradigm의 변화", 한국데이터베이스학회, Vol.4 No.1, 2001
- [10] 이진실, "E-mail 채널관리와 CRM", 한국데이터베이스학회, Vol.4 No.1, 2001