

중복 추천 문제를 반영한 다중 캠페인의 최적화

(Optimization of Multiple Campaigns Reflecting Multiple Recommendation Issue)

김 용 혁[†] 문 병 로^{**}
(Yong-Hyuk Kim) (Byung-Ro Moon)

요 약 개인화된 마케팅에서 고객 만족과 마케팅 효율을 최대화하는 것은 중요하다. 개인화된 캠페인 수행됨에 따라 여러 캠페인이 동시에 수행되곤 한다. 이 논문에서 우리는 동시에 여러 개인화된 캠페인을 수행할 때 발생하는 중복 추천 문제를 제기한다. 이는 특정 고객에게 상당히 많은 양의 캠페인이 쏟아지게 되는 문제를 말한다. 이 이슈를 해결하기 위한 다중캠페인 할당 문제를 모델링 한다. 그리고 이 문제의 해결 방법으로 동적계획법을 비롯한 여러 휴리스틱 알고리즘들을 제안한다. 필드 데이터의 실험을 통해 제기된 문제 모델의 중요성과 제안된 알고리즘의 효율성을 입증한다.

키워드 : 개인화된 마케팅, 다중 캠페인 할당, 반응 억제 함수, 동적계획법, 휴리스틱 알고리즘

Abstract In personalized marketing, it is important to maximize customer satisfaction and marketing efficiency. As personalized campaigns are frequently performed, several campaigns are frequently run simultaneously. The multiple recommendation problem occurs when we perform several personalized campaigns simultaneously. This implies that some customers may be bombarded with a considerable number of campaigns. We raise this issue and formulate the multi-campaign assignment problem to solve the issue. We propose dynamic programming method and various heuristic algorithms for solving the problem. With field data, we also present experimental results to verify the importance of the problem formulation and the effectiveness of the proposed algorithms.

Key words : personalized marketing, multi-campaign assignment, response suppression function, dynamic programming, heuristic algorithms

1. 서 론

고객관계관리(Customer Relationship Management; CRM)는 반복적으로 제품이나 서비스를 구매하는 충실한 고객(loyal customer)의 확보와 유지에 중요하다. 회사들은 자사의 수익과 고객의 만족을 최대화하기 위해 고객에게 개인화된 서비스를 제공하고자 노력해왔는데 그 대표적인 예가 일대일 마케팅이다. 인터넷과 이동통신이 급속하게 발전함에 따라 이 일대일 마케팅 시장이 확장되었다. 개인화된 캠페인에서는 캠페인 주제에 가장 매력적인 고객을 목표로 정한다. 그리고 각 고객에게 가

장 매력적인 콘텐츠를 제공하고 주어진 광고에 가장 적당한 고객들을 대응시킴으로써 고객의 충성도(loyalty)를 높이게 된다[1]. 따라서 이 때 캠페인에 대한 고객 선호도를 정확히 예측하는 것은 중요한 문제가 되고 이것은 또 다른 독립된 연구 주제가 된다. 협동적 필터링¹⁾(Collaborative Filtering; CF)[2-3], 클러스터링[4]과 근접이웃 알고리즘[5]을 포함한 다양한 데이터 마이닝 기법[6-7]들이 캠페인에 대한 고객 선호도를 예측하기 위해 사용되어 왔다[8]. 특히, CF는 빠르고 단순하므로 전자상거래에서 개인화를 위해 널리 사용되고 있고 [9-10] 이 CF를 기반으로 하는 많은 고객 선호도 예측 방법이 제안되어 왔다[11-13]. 전자상거래에서의 추천 시스템에 대한 보다 많은 내용은 논문 [14]에 잘 정리되어 있다.

개인화된 캠페인을 종종 수행함에 따라 여러 캠페인

· 이 논문은 2005년도 두뇌한국21사업에 의하여 지원되었음. 본 연구를 위한 데이터를 마련해주신다 고생한 권영근, 최윤석 학생께 감사의 말을 전합니다.

[†] 비 회 원 : 서울대학교 기계항공공학부
yhdffy@soar.snu.ac.kr

^{**} 정 회 원 : 서울대학교 컴퓨터공학부 교수
moon@soar.snu.ac.kr

논문접수 : 2004년 8월 19일

심사완료 : 2005년 3월 2일

1) 협동적 필터링은 개인화된 추천을 위한 표준 방법으로 입증되어 있다 [9-10]. 협동적 필터링을 사용하는 대표적인 회사는 NetPerceptions, Inc.이다.

을 동시에 수행하곤 하는데 이 때 어떤 한 캠페인에 대해 매력적인 고객이 다른 캠페인에 대해서도 매력적인 경우가 빈번히 발생한다. 만약 우리가 이러한 문제를 고려하지 않고 개별적으로 캠페인을 수행해 버린다면, 어떤 고객에게는 상당히 많은 양의 캠페인이 쏟아지게 될 것이다. 이러한 문제를 중복 추천 문제라 부르기로 한다. 고객에 대한 추천 수가 많아질수록 캠페인에 대한 흥미가 평균적으로 떨어지게 될 것이고 결국은 분별없는 캠페인들이 고객 만족뿐만 아니라 마케팅 효율성을 낮추게 된다. 이는 고객의 충성도를 감소시키게 되고 기존의 충실한 고객들마저 떠나버리는 크나큰 문제를 종종 야기할 수 있다. 불행히도 지금까지의 전형적인 방법들은 단일 캠페인의 효율성에만 초점을 맞추고 있었고 다중 캠페인 환경에서의 중복 추천에 대한 위험을 전혀 고려하지 않았다. 여러 캠페인이 동시에 수행되는 상황에서는 주의 깊게 고객들을 캠페인들에 분배하는 작업이 필요하게 된다.

본 논문에서는 중복 추천 문제를 고려한 다중 캠페인 할당 문제(Multi-Campaign Assignment Problem; MCAP)를 모델링하고 그 주제를 해결하기 위한 많은 알고리즘들을 제안한다. 또한, 필드 데이터를 가지고 행한 실험을 통해 제안된 알고리즘들의 효율성과 제한점을 검증한다.

이 논문의 나머지는 다음과 같이 구성된다. 2절에서는 다중 캠페인 할당 문제를 자세히 묘사한다. 3절에서는 그 문제 해결을 위한 다양한 알고리즘들을 제안한다. 문제의 핵심 요소들 중의 하나인 반응 억제 함수에 대한 설명이 4절에서 제시된다. 실험 결과 및 분석을 5절에서 제시하고 마지막으로 6절에서 결론을 내린다.

2. 다중 캠페인 할당 문제

이 절에서는 다중 캠페인 할당 문제를 정의하고자 한다. 문제를 엄밀하게 정의하기에 앞서 서론에서 정의 없이 사용되었던 캠페인을 먼저 정의한다. 캠페인이란 일반적으로 '특정 목적을 이루고자 하는 일련의 계획된 커뮤니케이션 활동'으로 정의를 내릴 수 있다. 좀더 구체적으로 언급하면 캠페인은 '비교적 많은 수의 소비자를 대상으로 일정 기간 동안 구체적 목적을 수행하기 위해 조직적으로 전개되는 활동'이라고 할 수 있으며, 여기에서 구체적인 목적이 상품이나 서비스의 프로모션에 있다면 이를 광고 캠페인이라고 한다. 마케팅 분야에서는 광고라는 수식어 없는 그냥 캠페인이란 용어가 곧 광고 캠페인을 의미한다. 이 논문에서도 광고 캠페인이란 뜻으로서 캠페인이란 용어를 사용한다.

2.1 문제의 작은 예

다중 캠페인 할당 문제는 캠페인의 효과를 최대화하

는 고객-캠페인 할당을 찾는 문제이다. 독립적인 캠페인과의 주된 차이는 캠페인에 대한 고객의 반응이 중복 추천에 영향을 받는다는 사실이다. 다음에서 이 문제의 중요성을 보여주는 간단한 예를 제시한다. 표 1-4는 중복 추천 문제를 감안한 다중 캠페인 마케팅의 중요성을 보여주는 간단한 예제 데이터를 보여준다.

표 1 고객의 캠페인 선호도

	캠페인1	캠페인2	캠페인3
고객1	100	98	80
고객2	70	48	13
고객3	50	77	62
고객4	34	4	61
고객5	9	90	70

표 2 고객의 반응도 함수값

	캠페인1	캠페인2	캠페인3
반응도	1.0	0.7	0.5

표 3 개별 캠페인 할당 예 (선호도·반응도 값 = 415.5)

	캠페인1	캠페인2	캠페인3
고객1	1	1	1
고객2	1	0	0
고객3	1	1	1
고객4	0	0	0
고객5	0	1	1

표 4 다중 캠페인 할당 예 (선호도·반응도 값 = 483)

	캠페인1	캠페인2	캠페인3
고객1	1	1	0
고객2	1	0	0
고객3	0	1	1
고객4	1	0	1
고객5	0	1	1

표 1은 고객의 캠페인 선호도를 예시적으로 표시한 것으로, 캠페인은 3 개, 고객은 5 명으로 구성된다. 각 캠페인 별로 정확히 3 명을 추천하는 해야 하는 상황이라 가정한다. 또한, 표 2에 의하면 중복 추천이 없는 경우 1.0, 두 개의 중복 추천이 발생할 경우 0.7, 세 개의 중복 추천이 발생할 경우 0.5의 비율로 각각의 캠페인에 대한 반응도가 감소한다고 가정된다.

지금까지의 전형적인 마케팅 방법으로 캠페인별로 별도로 추천을 행하면, 즉 다른 캠페인과의 연관성을 전혀 고려하지 않고 주어진 캠페인에서 선호도가 높은 순으로 추천을 행하면, 표 3에 표시된 바와 같이 고객 1과 고객 3은 세 개의 캠페인 중복 추천이 발생하고 고객 5는 두 개의 캠페인 중복 추천이 발생한다. 따라서 이 개

별 캠페인을 최적으로 수행한 방식에서는 추천 고객마다 선호도에 반응도를 곱한 값들의 합인 목적 값은 415.5가 된다. 하지만 표 4에 표시된 바와 같이 중복추천 문제를 감안하여 추천 고객을 적당히 분산시킨 방식에서는 목적 값은 483이 되어 효율이 훨씬 더 높아진다. 이 예는 중복추천 문제를 인지하고 이를 감안한 다중 캠페인 할당 방식이 마케팅의 효율을 높일 수 있음을 보여준다.

2.2 문제의 형식화

이제 다중 캠페인 할당 문제를 보다 엄밀하고 구체적으로 기술하고자 한다. N 을 고객의 수, $C=\{1,2,\dots,M\}$ 를 고객의 집합, K 를 캠페인의 수라고 하자. 문제를 위한 입력과 출력, 제한조건과 평가 함수를 다음에서 자세히 설명한다.

<입력>

각 고객에 대해, 각 캠페인에 대한 선호도가 주어지고 각 캠페인은 가중치 값을 가지게 된다. 중복 추천에 관련된 반응 억제 함수를 R 로 나타낸다.

- $f_1, f_2, \dots, f_K : C \rightarrow [0, \infty)$: 각 캠페인의 선호도 함수 (실제로는 벡터). 여기에서는 양의 실수에 의해 표현되는 양의 선호도만을 고려한다. 캠페인에 대한 선호도는 CF와 같은 이미 존재하는 선호도 예측 방법에 의해 얻을 수 있다. $f_j(i)$ 는 캠페인 j 에 대한 고객 i 의 예측된 선호도 값을 의미한다.
- $R : N \rightarrow [0, 1]$: 추천 수에 관련된 반응 억제 함수. 함수 값 $R(k)$ 를 한 고객이 k 개의 추천을 가질 때 선호도 감소 인자로 정의한다. 만약 H_i 가 고객 i 에 대한 추천 수라면 캠페인 j 에 대한 고객 i 의 실제 예측 선호도는 $R(H_i) \cdot f_j(i)$ 가 된다.
- w_1, w_2, \dots, w_K : 각 캠페인의 가중치. (캠페인 j 의 가중치는 w_j 이고 양수 값이 된다.) 각 캠페인 j 는 자신의 가중치 w_j 를 가진다. 캠페인의 중요도가 이 가중치에 의해 결정된다. 형식화된 다중캠페인 할당 문제에서 캠페인 j 에 대한 고객 i 의 선호도는 결국 가중치를 곱한 값 $w_j \cdot R(H_i) \cdot f_j(i)$ 가 된다.

<제한조건>

문제에 대한 제한조건으로 각 캠페인에 대한 추천 수의 최소 값과 최대 값이 있다. 캠페인 j 에서 추천 수는 P_j 와 P^j 사이에 있어야 한다. 여기에서 P_j 와 P^j 는 각각 캠페인 j 의 최소 및 최대 추천 수이다. 이들 제한조건에 있어서 각 캠페인에 대한 최대 추천 수와 최소 추천 수 사이의 차이, 즉, $P^j - P_j$ 는 사업 상황에 따라 다소 작을 수도 있고 반대로 클 수도 있다.

<출력>

출력은 이진 캠페인 할당 행렬 $M=(m_{ij})$ 이다. 여기에서 m_{ij} 는 캠페인 j 의 고객 i 로의 할당을 나타낸다; 캠페인 j 가 고객 i 로 할당되면 $m_{ij}=1$ 이 된다. 그림 1은 캠페인 할당 행렬의 한 예를 보여준다.

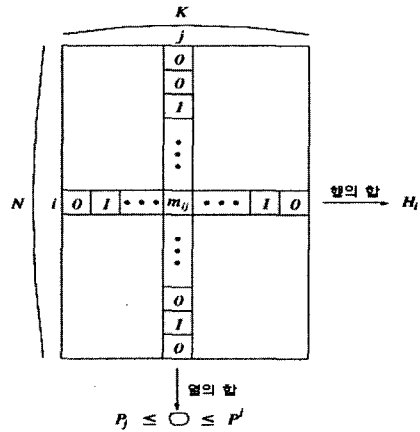


그림 1 캠페인 할당 행렬 $M=(m_{ij})$

<평가>

캠페인 j 에 대한 선호도 합을 캠페인 j 에 대한 추천 고객들의 실제 예측 선호도들의 합으로 정의한다. 캠페인 할당 행렬 M 의 적합도 F 는 다음과 같은 모든 캠페인에서의 선호도 합들의 가중치 합이 된다.

$$F(M) = \sum_{j=1}^K (w_j \cdot \sum_{i \in C} R(H_i) \cdot f_j(i)).$$

문제의 목적은 F 를 최대화하는 행렬 M 을 찾는 것이다.

<용어>

이 논문의 나머지에서 사용되는 용어를 아래에서 자세히 제시한다.

- $w = (w_1, w_2, \dots, w_K) \in \mathbb{R}^K$: 캠페인 가중치 벡터
- $f_j : C \rightarrow [0, \infty)$: 캠페인 j 에 대한 선호도 함수
- $R : N \rightarrow [0, 1]$: 반응 억제 함수(편의상, $R(0)=0$ 이라 가정한다.)
- $p^* = (P^1, P^2, \dots, P^K) \in \mathbb{IN}^K$: 상한 제한조건 벡터
- $p_* = (P_1, P_2, \dots, P_K) \in \mathbb{IN}^K$: 하한 제한조건 벡터
- $M = (m_{ij}) : N \times K$ 이진 캠페인 할당 행렬
- $M' = (m'_{ij}) : N \times K$ 실수 행렬, 여기에서 $m'_{ij} = f_j(i) \cdot m_{ij}$
- $m_i = (m_{i1}, m_{i2}, \dots, m_{iK})$: 행렬 M 의 i 번째 열벡터
- $m'_i = (m'_{i1}, m'_{i2}, \dots, m'_{iK})$: 행렬 M' 의 i 번째 열벡터
- $1_n : n$ 차원 벡터 $(1, 1, \dots, 1)$
- $H_i = m_i \cdot 1_K^T$: 고객 i 가 받는 추천 수

$\sigma_i = m'_i \cdot w^T$: 고객 i 의 추천 받은 캠페인에 대한 선호도의 가중치 합

<형식적인 정의>

보다 형식적으로 다중 캠페인 할당 문제(MCAP)를 다음과 같이 정의한다.

정의 1. 다중 캠페인 할당 문제는 다음 식을 만족하는 캠페인 할당 행렬 $M=(m_{ij})$ 을 찾는 문제이다.

$$\max(R(1_K M^T) \cdot M \cdot w^T) \text{ subject to } p_* \leq 1_N M \leq p^*$$

여기에서 $R(x_1, x_2, \dots, x_n) = (R(x_1), R(x_2), \dots, R(x_n))$ 이고 각 n 차원 벡터는 $1 \times n$ 행렬로 가정한다.

3. 제안하는 방법들

3.1 동적계획법

동적계획법(Dynamic Programming; DP)은 다양한 문제들에 유용하게 사용되어 왔다[15-16]. 특히, 할당 문제에 있어서도 배낭 문제(Knapsack Problem)와 같은 유명한 NP-완전 문제에 DP를 적용할 수 있다. 하지만 모든 할당 문제에 DP를 적용할 수 있는 것은 아니다.

예를 들어, 이차 할당 문제(Quadratic Assignment Problem)에는 DP를 적용하기가 어렵다. 제안된 다중 캠페인 할당 문제가 배낭 문제와 같은 선형 목적 함수를 가지고 있지는 않지만 제한 조건을 가지는 조합 최적화 문제로서 DP를 사용해서 MCAP의 최적 캠페인 할당 행렬을 찾을 수 있다. 여기에서는 최적치 함수 $S_i(v)$ 를 고정된 제한조건 벡터 v 와 고객 집합 $\{1, 2, \dots, i\}$ 를 가지는 다중 캠페인 할당 문제의 최적 적합도로 정의한다. 여기에서 v 의 각 원소 v_j 는 캠페인 j 에서의 추천 수를 의미한다. 최적성 원리에 의해 정의에 적합한 점화 관계는 다음과 같다.

점화식(recurrence relation): 각 $i = 1, 2, \dots, N$ 와 $v \leq p^*$ 에 대해

$$S_i(v) = \max_{m_i, \forall j, m_i \leq v_j} (S_{i-1}(v - m_i) + R(H_i) \cdot \sigma_i).$$

경계 조건(boundary condition): $S_0(0_K) = 0$ 이고 $0_K \neq v \leq p^*$ 일 때 $S_0(v) = -\infty$.

그림 2는 얻어진 점화 관계에 기반을 둔 DP 알고리즘의 의사코드를 보여준다. 1부터 N 까지 임의의 수로 고객들에 번호를 붙이면 추천에 관련된 제한조건 벡터

```

DP(R, p*, p*)
{
    /* 배열을 초기화한다. (경계 조건 처리) */
    S_0(0_K) = 0; // 0_K은 K차원 벡터 (0, 0, ..., 0)
    0_K ≠ v ≤ p*를 만족하는 각 v에 대해, // v = (v_1, v_2, ..., v_K) ∈ N^K
        S_0(v) = -∞;
    /* 점화식에 따라 배열을 채운다. */
    for i = 1 to N // 순차적으로 채운다.
        v ≤ p*를 만족하는 각 v에 대해,
            S_i(v) = max_{m_i, \forall j, m_i \leq v_j} (S_{i-1}(v - m_i) + R(H_i) \cdot \sigma_i);
            L_i(v) = argmax_{m_i, \forall j, m_i \leq v_j} (S_{i-1}(v - m_i) + R(H_i) \cdot \sigma_i); // 역방향 링크
    /* 최적의 값을 얻어낸다. */
    optimum_value = max_{v, p_* \leq v \leq p^*} S_N(v);
    /* 역방향 링크를 통해 최적 할당 행렬 M=(m_{ij})을 만든다. */
    rv = argmax_{v, p_* \leq v \leq p^*} S_N(v);
    for i = N to 1 // 역방향으로 최적의 할당을 찾아간다.
        m_i = L_i(rv);
        rv = rv - m_i;
    /* 결과를 리턴한다. */
    return optimum_value and M=(m_{ij});
}
    
```

그림 2 다중 캠페인 할당 문제를 위한 동적계획법

```

M = 0;
AVL 트리를 생성한다;
각 캠페인 j에 대해,
    높은 이득 값을 가지는 상위 α 명의 고객을 찾고,
    그 중에서 각 고객 i에 대해 노드 (g(i,j), (i,j))를 AVL 트리에 넣는다;
AVL 트리가 비거나 모든 캠페인이 다 채워질 때까지 다음을 반복한다
{
    최대 이득 노드 (i,j)를 선택하고 AVL 트리에서 삭제한다;
    만약 g(i,j)가 양이 아니고 모든 캠페인이
        최소 추천량에 대한 제한 조건을 만족한다면,
        반복문을 빠져 나간다;
    만약 캠페인 j가 다 채워지지 않았다면,
        고객 i에게 캠페인 j를 추천한다; // mij ← 1
        각 채워지지 않은 캠페인 k에 대해,
            AVL 트리에서 이득 값 g(i,k)를 갱신한다;
}
// 실험에서 α는 N/2으로 정했다.
    
```

그림 3 건설적인 할당 알고리즘 CAA (AVL 트리를 사용한 예)

v 를 취하고 먼저 고객 i 에게 m_i 벡터를 할당한다. 그러면 다시 고객 $i-1$ 에게 남은 추천량 $v - m_i$ 에서 m_{i-1} 벡터를 할당한다. 이러한 작업을 고객 1까지 반복한다. 즉, 이미 행해진 일부 할당을 가진 상태에서 남은 제한조건 벡터와 다음에 고려될 고객을 관계시킨다. 이러한 순차적인 방법으로 각 가능한 제한조건 벡터 v 에 대한 최적 값을 유지한다. 이미 할당된 벡터와 남은 추천량은 남은 과정의 최적화에 필수적이다. 이 알고리즘은 $O(NK \cdot \prod_{j=1}^K P^j)$ 공간이 필요하고 K 차원 벡터 m_i 가 가질 수 있는 상태의 최대 수는 2^K 이므로 이 알고리즘은 $O(NK2^K \cdot \prod_{j=1}^K P^j)$ 시간이 걸리게 된다. 만약 K 가 고정된 수라면 이 방법은 다항식 시간이 걸리는 알고리즘이 된다. 그러나 K 가 작지 않고 대부분의 모든 P^j 들이 $\Omega(N)$ 일 때, 이 알고리즘을 현실적인 시간 안에 수행하기 어렵게 된다. 최적 할당 행렬은 동적계획법 과정 동안에 저장된 역방향 링크 $L_i(v)$ 에 의해 얻어진다.

제한된 DP 알고리즘을 통해 캠페인 수인 K 의 값이 작은 수치로 고정되어 있는 경우 MCAP을 N 에 대한 $K+1$ 차 다항식 시간 안에 풀 수 있게 된다. 하지만 이 알고리즘은 K, N 값이 커짐에 따라 그 시간 복잡도가 급격하게 커지므로 오직 작은 K, N 값을 가지는 문제에 만 그 응용이 가능하다. 그러므로 큰 크기의 문제를 합리적인 시간 안에 효율적으로 다루기 위해서는 다른 휴리스틱 알고리즘이 필요하게 된다. 하지만 DP 알고리즘은 최적해를 보장하므로 다음 절에서 제안되는 다른 휴리스틱의 최적성 평가에 또한 유용하게 쓰일 수 있다. 5 절에서는 이러한 목적으로 DP가 사용될 것이다.

3.2 휴리스틱 알고리즘

3.2.1 건설적인 할당 알고리즘

이 절에서는 건설적인 휴리스틱 알고리즘을 제안한다. 이 알고리즘은 어떠한 고객도 캠페인 추천을 받지 않은 상황에서 시작하여 탐욕적(greedy)인 방식으로 고객에게 캠페인을 반복적으로 할당한다. 이 알고리즘을 건설적인 할당 알고리즘(Constructive Assignment Algorithm; CAA)이라 부르기로 한다. (고객 i , 캠페인 j) 쌍의 이득 $g(i,j)$ 을 고객 i 에게 캠페인 j 를 할당함에 의한 적합도 이득의 양으로 정의한다. 초기에 이득 $g(i,j)$ 는 $wf_j(i)$, 즉, 캠페인 가중치와 캠페인 j 에 대한 고객 i 의 선호도 곱과 같다. 일반적인 경우에 이득 값은 다음과 같은 식으로 계산된다.

$$g(i,j) = R(H_i+1) \cdot (\sigma_i + wf_j(i)) - R(H_i) \cdot \sigma_i$$

이 이득 값은 일반적으로 실수 값을 가지게 되는데 본 논문에서는 실수 값을 가지는 이득의 효율적인 관리를 위해 AVL (Adel'son-Vel'skii Landis) 트리[17]를 사용한다.²⁾ 그림 3은 AVL 트리를 사용한 CAA의 템플릿을 보여준다. 먼저, 각 캠페인에 대한 가장 매력적인 고객 α 명을 선택하고 이들을 AVL 트리에 삽입한다. 그런 후에 반복적으로 다음을 수행한다. 최대 이득을 가지는 (고객 i , 캠페인 j) 쌍을 선택한다. 만약 이득이 양수이고 캠페인 j 가 최대 추천수를 넘기지 않았다면 고객 i 에 캠페인 j 를 할당한다. 할당이 이루어지면 다른 캠페인에 대한 고객 i 의 이득 값들을 갱신한다. 자연스럽게

2) 비록 본 논문에서는 이득 값을 가지는 노드들을 효율적으로 다루기 위한 자료구조의 한 예로서 AVL 트리를 사용하였지만 힙(heap)이나 다른 개선된 균형 트리(balanced tree)를 사용하여 같은 시간 복잡도를 가지면서도 실제적으로는 더 빠른 구현이 가능하다.

반응 억제 함수는 단조 감소라 가정한다. 그러면 최대 이득이 0 아래로 떨어진 이후에 어떤 이득 값도 양수가 될 수 없게 되므로 모든 캠페인이 최소 추천 수 제한 조건을 만족하는 한 알고리즘을 종료한다. 알고리즘 수행동안 AVL 트리에는 $O(NK)$ 개의 노드가 있게 된다. 그러므로 AVL 트리에서 한 노드의 추가, 삭제에 $O(\log(NK))$, 즉, $O(\log M)$ 시간이 걸린다. 그러면 이 알고리즘의 시간 복잡도는 $O(NK^2 \log M)$ 이 된다. 만약 이 알고리즘이 AVL 트리와 같은 트리 표현의 자료구조를 사용하지 않고 단순하게 구현된다면 $O(N^2 K^3)$ 의 시간이 걸리게 된다.

3.2.2 반복 향상 알고리즘

본 논문에서는 또한 반복 향상 휴리스틱을 제안한다. 모든 고객이 적당한 수의 캠페인을 할당 받은 후에 이 휴리스틱을 수행한다. 이 휴리스틱은 일련의 패스(pass)들을 진행한다. 각 패스 동안에 휴리스틱은 초기 해를 개선시켜 새로운 해를 만든다. 이 과정을 어떠한 개선도 얻어지지 않을 때까지 반복한다.

A_j 와 B_j 를 각각 캠페인 j 에 대한 추천받은 고객의 집합, 추천받지 않은 고객의 집합이라고 하자. A_j 에 있는 각 고객 a 에 대해 캠페인 j 의 이득 g'_a 를 고객 a 에 대한 할당을 취소함으로써 얻는 적합도 이득으로 정의한다. B_j 에 있는 각 고객 b 에 대해 캠페인 j 의 이득 g'_b 를 고객 b 에 캠페인 j 를 할당함으로써 얻는 적합도 이득으로 정의한다. 형식적으로,

$$g'_a = R(H_a - 1) \cdot (\sigma_a - w f_j(a)) - R(H_a) \cdot \sigma_a, \quad a \in A_j, \text{ 이고}$$

$$g'_b = R(H_b + 1) \cdot (\sigma_b + w f_j(b)) - R(H_b) \cdot \sigma_b, \quad b \in B_j.$$

그림 4는 반복 향상 휴리스틱의 템플릿을 보여준다. 반복 향상의 한 패스를 각 캠페인에 대해 수행한다. 하나의 캠페인이 주어지면 휴리스틱은 교환할 때 최대 이득 합을 가지는 같은 크기의 추천 고객과 비추천 고객 집합을 선택하고 그 두 집합을 교환한다. 이 부분 집합 쌍을 교환한 후에 새로운 해에서 시작하여 다른 패스를 수행한다. 한 패스의 시간 복잡도는 $O(NK \log N)$ 이 된다.

// A_j, B_j : 각각은 각 캠페인 j 에 대한 추천 고객의 집합, 추천받지 않은 고객의 집합. 개선이 없을 때까지 다음을 반복한다.

{

 각 캠페인 j 에 대해,

 각 $a \in A_j, b \in B_j$ 에 대해 이득 값 g'_a, g'_b 를 계산한다;

g'_a 들과 g'_b 들을 내림차순 정렬한다.

$\sum_{i=1}^k (g'_a + g'_b)$ 값을 최대화하는 k 를 선택한다;

 부분집합 $\{a_1, a_2, \dots, a_k\}$ 와 $\{b_1, b_2, \dots, b_k\}$ 를 바꾼다.

}

그림 4 반복 향상 휴리스틱

4. 반응 억제 함수

다중 캠페인 추천의 경우에 고객 반응률은 추천 수가 증가함에 따라 감소한다. 이 절에서는 중복 추천에 따라 반응률 감소를 반영하는 반응 억제 함수를 도입한다. 함수는 단조 감소 함수이어야 한다. 최적의 반응 억제 함수는 상황에 따라 다르고 이는 또 다른 연구 주제가 된다. 대신에 본 논문에서는 많은 반응 억제 함수들을 고안했다. 이러한 함수들은 비록 최적화 작업을 통하여 구한 최적의 함수는 아니지만 현실적으로 사용가능한 함수 모델이다. 반응 억제 함수는 실제로 캠페인을 수행하는 상황에 많이 의존하는 함수이므로 그때그때마다 캠페인 상황에 따라 그 함수가 다르게 된다. 또한, 이러한 점을 고려하여 현재의 상황에 맞는 최적의 반응 억제 함수를 찾기 위한 하나의 방법으로 독립 캠페인의 실험을 통한 반응 억제 함수의 최적 예측 모델을 새로운 최적화 문제로서 제시한다.

4.1 기본적인 반응 억제 함수들

그림 5는 5 개의 기본적인 반응 억제 함수를 보여준다. 이들 함수는 최대 1의 값을 가지는 음이 아닌 단조 감소

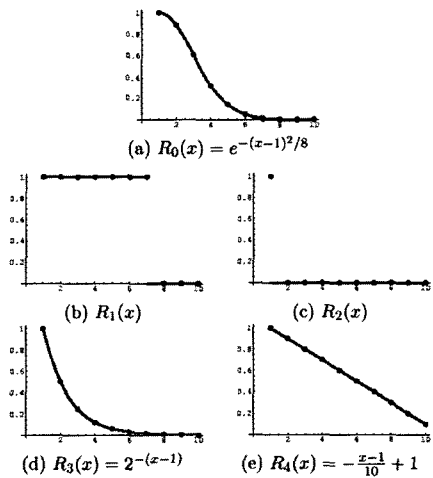


그림 5 기본적인 반응 억제 함수들 ($R_i(x) = 0, x \geq 11$)

함수들이다. 이 논문에서는 주요한 반응 억제 함수로 함수 R_0 를 사용하고(그림 5(a)), 부가적인 연구를 위해 다른 함수들을 사용한다. R_0 는 가우시안 함수로부터 고안하였다. 이 함수에서 캠페인에 대한 선호도는, 예를 들면, 4 개의 캠페인을 한 고객에게 동시에 추천하였을 때 1/3로 감소된다. R_1 과 R_2 는 단순한 스텝 함수들이고 R_3 과 R_4 는 각각 지수 함수적으로, 선형적으로 감소하는 함수들이다.

4.2 좋은 반응 억제 함수 찾기

이 절에서는 독립적인 캠페인의 결과로부터 최적의 반응 억제 함수를 추정하는 방법을 하나의 최적화 문제로서 소개한다. 만약 선호도가 정확히 예측되었다고 가정하면, 우리는 독립적인 캠페인의 결과가 중복 추천에 따른 반응을 감소를 반영한다고 기대할 수 있다. 그러면 좋은 반응 억제 함수를 찾는 문제는 실제 반응에 가장 유사한 실제 예측 선호도를 찾는 문제가 된다. 우선, 다음의 두 조건을 가정한다. (i) 반응 억제 함수 R 은 음이 아닌 단조 감소 함수이다. (ii) 각 고객은 주어진 K 캠페인 이외의 다른 추천을 받지 않는다. 반응 억제 함수를 찾기 위해 독립적인 캠페인의 수행은 필수 선행사항이다. 독립적인 캠페인의 결과 행렬을 $\Gamma=(\gamma_{ij})$ 라 하자. 여기에서 γ_{ij} 는 캠페인 j 에 대한 고객 i 의 실제 반응이다. 일반적으로 우리는 $R(1)=1.0$ 이라고 가정하므로 반응 억제 함수 R 은 $K-1$ 차원 음이 아닌 실수 벡터에 대응한다. 반응 억제 함수 R 이 주어지면 결과 행렬에 대한 유사성을 다음과 같이 정의한다.

$$\frac{\sum_{k=1}^K \sum_{(i,j) \in C^2} \Delta r_{i,j,k} \Delta r_{i,j,k}}{\sqrt{\sum_{k=1}^K \sum_{(i,j) \in C^2} \Delta^2 r_{i,j,k}}}$$

여기에서 $\Delta r_{i,j,k} = \gamma_{ik} - \gamma_{jk}$ 이고 $\Delta r_{i,j,k} = R(H_i)f_k(i) - R(H_j)f_k(j)$ 이다. 그러면 가장 그럴듯한 반응 억제 함수를 추정하는 문제는 이 유사성을 최대화하는 함수 R 을 찾는 최적화 문제가 된다. 최적의 반응 억제 함수들은 고객의 성향에 의존하며 캠페인 상황에 따라 다를 수 있다.

5. 실험 결과

5.1 입력과 파라미터들

이메일 마케팅은 인터넷 마케팅을 위한 가장 유망한 도구 중의 하나로 인식되고 있다. 그 반응률은 직접적인 우편물이나 배너 광고보다도 훨씬 좋은 것으로 알려져 있다[18]. 요즘 이메일 마케팅 회사는 고객의 허락을 얻고 있고 이메일 마케팅의 이러한 종류는 합법적이고 유망한 사업 모델로 잘 설립되어 있다. 본 논문에서는 이

메일 마케팅으로부터 얻어진 필드 데이터를 사용했다. 48,559 명의 고객과 33 개 캠페인으로 구성된 필드 데이터로부터 변형된 CF로 예측된 선호도 값을 사용했다. 이 연구에 사용된 전체 데이터 집합은 Optus Inc.에 의해 제공된 것을 사용하였다. 개인 정보는 많은 독립 변수와 고객이 이메일에 반응했는지를 나타내는 하나의 종속 변수로 구성된다. 각 선호도 값은 다음과 같이 예측된다. 선호도 예측을 위해 본 논문에서는 각 캠페인에 대한 실제 반응 값을 가지고 있는 8,650 명의 훈련 고객(C_i)을 사용했다. 각 고객은 86 개의 이진변수로 구성된 개인 정보를 가진다. 이와 동일한 정보가 논문 [19]에서 사용되었다. 캠페인이 주어지면 각 고객 i 에 대해 예측된 선호도 값은 $\sum_{k \in C} \rho_{ik} \cdot f(k)$ 이다. 여기에서 ρ_{ik} 는 고객 i 와 훈련 고객 k 사이의 상관도 값이고 $f(k)$ 는 고객 k 의 실제 반응도이다.

다른 언급이 없다면 반응 억제 함수로서 가우시안 함수 R_0 를 사용한다. 모든 캠페인에 대한 가중치는 동일하게 정하였다. 그리고 각 캠페인에 대한 최대 추천 수를 동일하게 2,428 명, 전체 고객 수의 5%로 정하였고 최소 추천 수는 0으로 정하였다.

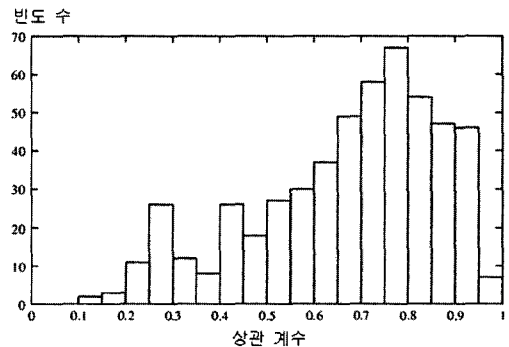


그림 6 각 캠페인 쌍의 피어슨 상관 계수에 대한 히스토그램

캠페인에 대한 고객의 평균 선호도는 4.67이었고 선호도의 표준 편차는 5.05이었다. 이는 분포가 정규적이지 않다는 것을 나타낸다. 본 논문에서는 또한 각 캠페인 쌍에 대해 선호도 값들의 피어슨 상관 계수를 조사했다. 여기에서 상관계수는 캠페인을 고객들에 대한 선호도 값의 벡터로 표현했을 때 캠페인 쌍 간의 상관 계수를 의미한다. 그림 6은 이 값들의 히스토그램을 보여준다. 전체 528 쌍 중에서 422 쌍(약 80%)이 0.5보다 큰 상관 계수를 보여주었다. 필드 데이터의 이러한 속성은 중복 추천 문제의 위험성을 드러낸다. 그러므로 이 결과는 MCAP 모델링의 필요성에 대한 충분한 이유를 제공해 준다.

5.2 결과 분석

먼저 제안한 다중 캠페인 형식화가 기존의 독립적인 캠페인 방식에 비해 어느 정도의 성능 차이를 보이는지를 앞에서 제안한 다양한 휴리스틱을 사용하여 실험하였고 사용된 휴리스틱 간의 성능에 대한 비교 분석을 행한다. 표 5는 독립적인 캠페인과 다중 캠페인 형식화 하에서 다양한 다중 캠페인 휴리스틱들의 성능을 보여준다. 독립적인(Independent) 캠페인의 결과는 다른 캠페인들과의 관계를 고려하지 않은 33 개의 독립적인 캠페인들로부터 나온다. 다중 캠페인 휴리스틱에는 두 가지의 초기화 방법이 있다: 랜덤(Random) 초기화 혹은 3.2.1절의 CAA가 바로 그것이다. 초기화 후에 반복 항상 휴리스틱을 적용함으로써 초기화로 얻어진 할당의 품질을 향상시킨다. 랜덤 초기화 후에 반복 항상 휴리스틱을 적용한 방법과 CAA 후에 반복 항상 휴리스틱을 적용한 방법을 각각 Random-I와 CAA-I로 적기로 한다. 제안한 반복 항상 휴리스틱은 적용할 캠페인의 순서를 임의로 선택한다. 따라서 같은 입력에 대해서도 다른 결과를 생성하게 된다. 한편, CAA는 항상 같은 결과를 출력하는 결정적 알고리즘이다. CAA를 제외한 모든 방법은 1,000 번 수행되었다.

표 5 여러 알고리즘의 비교

방법	최고*	평균*	표준편차	수행시간†
Independent	-	10068.36	-	-
Random	8076.55†	7902.69†	61.38†	-
Random-I	31713.46†	31280.43†	155.01†	29.77
CAA	33048.49	33048.49	-	14.72
CAA-I	33245.83†	33212.29†	12.45†	14.72+21.17

* 각 값은 2절에서 설명된 적합도 값 $F(M)$ 을 의미한다.
 † 1,000 번 수행한 결과. CAA는 결정적 알고리즘이다.
 ‡ 랜티엄 III 1 GHz.

독립적인 캠페인이 다중 캠페인 형식화에서 랜덤 할당보다는 좋았다 하지만 다른 다중 캠페인 휴리스틱들에는 비교될만하지 않았다. 반복 항상과 결합된 랜덤 초기화 방법(Random-I)은 CAA보다 많이 느릴 뿐만 아니라 1,000 번 수행한 결과에서 제일 좋은 해조차도 CAA의 결과보다 나빴다. 이는 CAA 성능이 반복 항상 휴리스틱 보다 월등히 좋음을 보여준다. 그림 7은 CAA의 노드 추가 과정 중에서의 적합도 값의 변화를 보여준다. 2,400 노드 삽입 부분에서 CAA는 독립적인 캠페인을 능가했다. 33,300 노드 삽입 이후에 CAA는 반복 항상 휴리스틱의 성능을 앞질렀다. 하지만 반복 항상 휴리스틱은 CAA의 결과를 향상시키는데 효율적이었다. 즉, CAA가 반복 항상 휴리스틱 보다 월등히 좋기는 하지만 CAA는 결정적 알고리즘이고 항상 같은 품질의

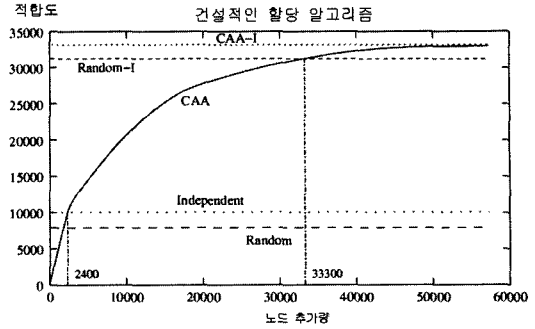
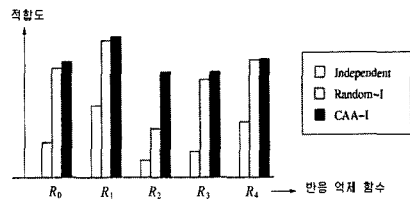


그림 7 노드 추가에 따른 CAA의 성능

결과를 만들게 되는데 이 결과를 반복 항상 휴리스틱을 통해 작지만 결과를 개선시킬 수 있었다. CAA-I 휴리스틱의 적합도 값은 독립적인 캠페인의 것보다 3 배 이상이고 이 결과는 다중 캠페인 접근 방식이 얼마나 중요한지를 잘 보여준다.

또한, 본 논문에서는 반응 억제 함수를 하나로 고정하지 않고 다양한 반응 억제 함수에 대해 실험을 수행했다. 반응 억제 함수의 형태는 상황에 따라 달라질 수 있으므로 하나의 함수로 고정하지 않고 다양한 형태의 실험을 통해 분석하는 것이 의미가 있기 때문이다. 그림 8은 다른 반응 억제 함수에 따른 결과를 보여준다. 구체적인 함수의 모습은 그림 5에서 제시하였다. 사용된 모든 반응 억제 함수에 대해 다중 캠페인 모델의 결과는 독립적인 캠페인 결과에 비해 의미 있는 개선을 보여주었다. 중복 추천에 따른 고객의 반응이 느리게 감소함에 따라 두 캠페인 모델 간의 성능 차이는 더 작게 된다. 반응도 억제 함수 R_1 은 7 개의 중복 추천에 대해서도 동일하게 높은 반응을 보이도록 정해졌다. 놀랍게도 이 R_1 에 대해 다중 캠페인의 결과는 여전히 독립적인 캠페인의 결과에 비해 약 2 배정도 좋았다. 이는 중복 추천이 얼마나 빈번히 일어날 수 있는지를 보여주는 좋은



방법	R_0	R_1	R_2	R_3	R_4
Independent	10068.36	20582.88	4883.20	7413.27	15923.29
Random-I	31280.43	39041.82	14031.61	27977.04	33691.92
CAA-I	33212.29	40218.96	30209.76	30229.19	33800.37

그림 8 다양한 반응 억제 함수에 따른 결과
 1,000 번 수행의 평균 결과

결과이고 제한한 다중 캠페인 모델의 중요성을 다시 한번 뒷받침해 준다.

추가적으로 최적해를 보장하는 DP 알고리즘과 제한된 휴리스틱 알고리즘들을 비교한다. 비록 동적계획법이 느리다 할지라도 이는 항상 최적해를 보장하므로 다른 휴리스틱의 최적성을 평가하는데 의미가 있게 된다. 동적계획법의 실행시간이 굉장히 크기 때문에 우리는 1,000 명의 고객과 3 개의 캠페인을 가지는 인스턴스로 실험을 제한했다.³⁾ 표 6은 이들의 성능을 보여준다. 33 개 캠페인 가운데 3 개의 캠페인을 임의로 선택했고 고객의 집합은 100 명부터 1,000 명까지의 크기를 가지는 집합을 만들기 위해 원래 고객 집합에서 임의로 추출했다. 추천 고객의 최대 수는 전체 고객 수의 절반으로 정했고 최소 추천 수는 0으로 정했다. 작은 크기의 문제들에 대해서도 CAA-I는 Random-I의 성능을 능가했다. CAA-I는 동적계획법에 비해 비교할 수 없을 만큼 훨씬 빠를 뿐만 아니라 그 결과가 최적해에 아주 근접해 있음을 알 수 있다. 이러한 특성은 문제의 크기, 즉, 고객 수에 따라 그 성향이 크게 달라지지 않음을 볼 수 있다. 그림 9에서 이 결과를 그림으로 도식화하였다. 비록 이

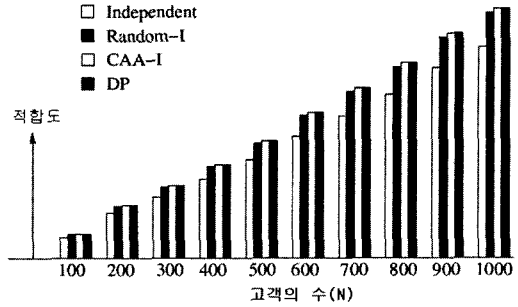


그림 9 표 6의 도식화

결과로부터 CAA-I가 큰 테이타에 대해서도 근사 최적해를 찾을 수 있음을 보장하지는 못한다 할지라도 CAA-I가 매력적이고 실용적인 휴리스틱임을 보여준다. 48,559 명의 고객과 33 개 캠페인을 가지는 5.1절에 제시된 필드 데이터 문제에 대해 불행히도 동적계획법의 막대한 시간 요구량 때문에 그 최적해를 알지 못한다.

마지막으로 추천 캠페인 수에 따른 고객의 분포를 조사했다. 그림 10(제한된 알고리즘을 구현한 상업적 프로그램으로부터 얻어진 그림)은 독립적인 캠페인과 CAA-I 각각에 대한 분포를 보여준다. 결과는 33 개 캠페인으로부터 얻어졌다. 두 방법이 극단의 대조를 보여주고 있다. 독립적인 캠페인에서 3 개의 캠페인보다 많이 추천 받은 고객이 상당수가 있는 반면 CAA-I에는 거의 없음을 관찰할 수 있다. 극단적인 경우로, 우리는 독립적인 캠페인에서 33 개 가운데 28 개의 캠페인을 할당받은 고객을 발견할 수 있었다. 하지만 CAA-I에서 캠페인은 고객들에게 훨씬 골고루 할당되었다.

3) 앞 절에서 언급했지만 캠페인 수 K 가 주어지면 시간 복잡도가 N 에 대해 $K+1$ 차 다항식이 된다. 그래서 표 6에서 보여준 결과는 $K=3$ 인 경우에 $N=1000$ 인 작은 경우에도 불구하고 펜티엄 III 1 GHz에서 약 70,000 초라는 많은 시간이 걸리게 된다. 필드 데이터인 약 50,000 명 정도의 고객을 갖는 경우라면 그 수행 추정 시간이 약 14,000 년 정도가 걸리게 된다. 이 시간은 캠페인 수가 3인 작은 경우에 해당하고 실제 필드 데이터인 33 개의 캠페인에 적용할 경우 그 시간은 막대하게 커지게 된다. 이는 동적계획법이 실제로 사용하기에 얼마나 비효율적인지를 보여준다고 할 수 있다.

표 6 작은 데이터 집합에 대한 결과 비교

방법	$N=100$	$N=200$	$N=300$	$N=400$	$N=500$
Independent	286.18	628.33	848.42	1093.58	1371.02
Random-I*	331.99	721.71	992.20	1279.07	1598.69
CAA-I* (시간*)	337.51 (0.0011)	734.27 (0.0024)	1009.87 (0.0037)	1304.16 (0.0050)	1632.04 (0.0063)
DP* (시간*)	337.51 (9.26)	734.64 (145.82)	1010.09 (755.04)	1304.43 (2336.41)	1632.20 (5654.93)

방법	$N=600$	$N=700$	$N=800$	$N=900$	$N=1000$
Independent	1697.14	1979.68	2280.35	2651.34	2945.03
Random-I*	1989.91	2324.90	2663.31	3077.42	3421.86
CAA-I* (시간*)	2030.68 (0.0079)	2368.01 (0.0094)	2726.54 (0.0110)	3132.84 (0.0126)	3480.66 (0.0142)
DP* (시간*)	2031.22 (11553.92)	2368.44 (19469.34)	2726.99 (28838.84)	3135.04 (46011.66)	3483.02 (71479.06)

캠페인 수 $K=3$. 각 캠페인에 대해 최대 50% 추천, 최소 0% 추천. 동일한 캠페인 가중치, 즉, 각 캠페인 j 에 대해 $w_j = 0.33$.

반응 억제 함수 $R(x) = R_0(x)$.

* 1,000 번 시도의 평균 결과, † 최적해, * 펜티엄 III 1 GHz

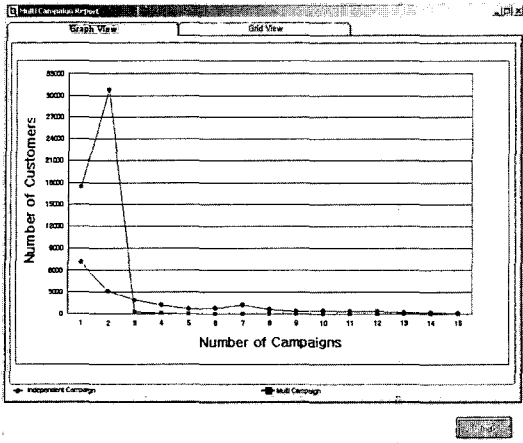


그림 10 추천 캠페인 수에 따른 고객의 분포

6. 결론

이 논문의 대표적인 성과는 다음과 같다. 첫째, 본 논문에서는 다중 캠페인 할당 문제(MACP)를 처음으로 제안하고 형식화했다. 저자가 아는 한 이것은 최적화 관점에서 다중 캠페인 상황을 공략하는 첫 번째 모델이다. 둘째, 본 논문에서는 MCAP을 해결하기 위한 동적계획법을 고안했다. 이 알고리즘은 얻어진 해의 최적성을 보장한다. 마지막으로, 본 논문에서는 MCAP을 해결하기 위한 여러 휴리스틱들을 제안했다. 또한, 다양한 실험을 통해 이들의 성능을 비교 검증하였다. CAA는 독립적인 휴리스틱 알고리즘으로서 좋은 성능을 보여주었다. 뿐만 아니라 제안된 DP 알고리즘에 의해 최적해를 아는 문제들에 대해서 CAA는 최적해에 아주 가깝게 도달했다. 또한 CAA를 반복 향상 알고리즘으로 그 성능을 개선할 수 있었다. 고객 1,000 명과 3 개의 캠페인을 가지는 예제에 대해 CAA-I는 DP 알고리즘이 걸린 시간의 5백만 분의 1시간 안에 근사 최적해를 찾아내었다. 우리가 작은 수의 고객에 대해 작은 집합의 캠페인을 수행한다면 DP가 선택될지 모른다. 하지만 대부분의 실제적인 상황에서 CAA-I가 실용적인 선택이 된다. 실용적인 시간 안에 CAA-I를 수행할 수 있기 때문이다; 대략 50,000 명의 고객과 33 개의 캠페인을 가지는 필드 데이터에 대해서 CAA-I는 펜티엄 III 1GHz에서 불과 약 36 초 정도 밖에 소요되지 않았다.

5절에서는 MCAP 모델을 통해 많은 수의 고객에게 캠페인이 몰리는 현상을 피할 수 있음을 알 수 있었다. 5 개 이상의 캠페인 추천을 받은 고객들은 대부분 회사의 입장에서는 충실한 고객이다. 하지만 이런 분별없는 캠페인은 높은 가치를 지니는 고객의 충성도를 감소시키고 최악의 경우에 기존의 충실한 고객들이 회사를 떠

나버리는 크나큰 문제 일으킬지 모른다. MCAP 모델은 전반적인 고객 반응을 향상시킬 뿐만 아니라 적어도 고객의 충성도에 피해를 일으키지 않게 한다.

비록 DP가 현재 형태로는 비실용적이라 할지라도 차원 감소 기법에 의해 보다 큰 문제를 다룰 수 있을 것이다. 라그랑지 승수(Lagrange multiplier)의 사용은 차원 감소의 좋은 후보가 되는데 이를 앞으로의 연구과제로 남긴다. 이 연구가 다중 캠페인 할당 최적화를 위한 깊은 연구의 동기를 제공하기를 바란다.

참고 문헌

- [1] R. Dewan, B. Jing, and A. Seidmann. One-to-one marketing on the internet. In *Proceedings of the 20th International Conference on Information Systems*, pages 93-102, 1999.
- [2] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61-70, 1992.
- [3] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the Computer Supported Collaborative Work Conference*, pages 175-186, 1994.
- [4] A. K. Jain and R. C. Dubes. *Algorithms for Clustering Data*. Prentice Hall, 1988.
- [5] C. Feustel and L. Shapiro. The nearest neighbor problem in an abstract metric space. *Pattern Recognition Letters*, 1:125-128, 1982.
- [6] M. S. Chen and P. S. Han, J. Yu. Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering*, 8(6):866-883, 1996.
- [7] M. Goebel and L. Gruenwald. A survey of data mining and knowledge discovery software tools. *SIGKDD Explorations*, 1:20-33, 1999.
- [8] M. J. A. Berry and G. Linoff. *Data Mining Techniques for Marketing, Sales, and Customer Support*. John Wiley & Sons, Inc, 1997.
- [9] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 230-237, 1999.
- [10] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordan, and J. Riedl. GroupLens: applying collaborative filtering to usenet news. *Communications of the ACM*, 40:77-87, 1997.
- [11] C. C. Aggarwal, J. L. Wolf, K. L. Wu, and P. S. Yu. Horting hatches an egg: A new graph-theoretic approach to collaborative filtering. In *Knowledge Discovery and Data Mining*, pages

201-212, 1999.

[12] D. Greening. Building consumer trust with accurate product recommendations. Technical Report LMWSWP-210-6966, LikeMinds White Paper, 1997.

[13] U. Shardanand and P. Maes. Social information filtering: Algorithms for automating "word of mouth," In *Proceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems*, volume 1, pages 210-217, 1995.

[14] J. B. Schafer, J. A. Konstan, and J. Riedi. Recommender systems in e-commerce. In *Proceedings of the ACM Conference on Electronic Commerce*, pages 158-166, 1999.

[15] R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.

[16] S. E. Dreyfus and A. M. Law. *The Art and Theory of Dynamic Programming*. Academic Press, 1977.

[17] G. M. Adel'son-Vel'skii and E. M. Landis. An algorithm for the organization of information. *Soviet Mathematics Doklady*, 3:1259-1262, 1962.

[18] *Email Marketing Maximized, Insight Report 2000*. Peppers and Rogers Group, 2000.

[19] Y. K. Kwon and B. R. Moon. Personalized email marketing with a genetic programming circuit model. In *Proceedings of the Genetic and Evolutionary Computation Conference*, pages 1352-1358, 2001.



김 용 혁

1999년 서울대학교 전산과학전공 학사
 2001년 서울대학교 전기컴퓨터공학부 석사.
 2005년 서울대학교 전기컴퓨터공학부 박사.
 현재 서울대학교 기계항공공학부 두뇌한국21 박사후과정 및 반도체 공동 연구소 연구원. 관심분야는 그래프 분할,

최적화 이론, 알고리즘 설계



문 병 로

1985년 서울대학교 계산통계학과 학사
 1987년 KAIST 전산학과 석사. 1994년 펜실바니아 주립대 박사. 1987년~1991년 (주)LG전자 중앙연구소 선임연구원
 1994년~1995년 UCLA VLSI CAD Lab 박사후연구원. 1996년~1997년 8월

(주)LG반도체 DT연구소 책임연구원. 1997년 9월~현재 서울대학교 컴퓨터공학부 부교수. 관심분야는 최적화, 유전 알고리즘, 알고리즘 디자인 및 분석, 스케줄링 등