

온라인 배너 광고 강화학습의 최적 탐색-활용 전략: 구전효과의 영향

김 범 수*, 유 건 재**, 이 준 겐***

목 차

요약	3. 모형
1. 서론	4. 분석 결과
2. 이론적 배경	5. 강건성 분석(robustness test)
2.1 온라인 배너 광고	6. 결론 및 시사점
2.2 구전효과	References
2.3 강화학습과 활용-탐색 전략	Abstract

요약

온라인 배너 광고 산업에서는 일반적으로 복수의 배너 대안이 제작된다. 이때 중요한 의사결정은 어떤 광고 배너 대안을 선택해서 고객에게 노출하느냐 하는 것이다. 각 배너 대안을 고객이 클릭할 확률을 미리 알 수 없기 때문에 경영자는 실험적으로 여러 대안을 노출한 후, 고객의 클릭 여부에 따라 각 대안의 클릭 확률을 추정하며 최적의 대안을 찾아야 하고 이것은 온라인 광고와 관련된 강화학습 프로세스이다. 이 과정에서의 주요 의사결정 문제는 축적된 추정 클릭 확률 지식을 이용해서 최적의 대안을 노출하는 활용 전략과, 잠재적으로 더 우수한 대안을 찾기 위해 새로운 대안을 시도해보는 탐색 전략의 최적 균형점을 찾는 것이다. 본 연구는 구전효과와 대안의 수가 이러한 최적 탐색-활용 전략에 미치는 영향을 분석하였다. 이는 고객이 노출된 배너를 클릭하는 경우 관련 제품을 주위에 홍보하는 과정을 통해 광고 배너의 클릭률이 높아지는 구전효과를 온라인 광고 관련 강화학습에 추가하여 구현한 것이다. 분석을 위해 Multi-Armed Bandit 모형을 이용한 시뮬레이션 기법을 사용하였다. 분석 결과, 구전효과의 크기가 커지고 배너 대안의 수가 적을수록 광고 강화학습의 최적 탐색 수준이 높아지는 것이 관측되었다. 이는 구전효과에 의해 고객이 광고 배너를 클릭할 확률이 증가함에 따라 기존에 축적했던 추정 클릭률 지식의 가치가 낮아지고, 따라서 새로운 대안을 탐색하는 것의 가치가 증가하기 때문으로 분석되었다. 또한 광고 대안의 수가 작을 경우에는 구전효과 크기가 커질 때 최적 탐색 수준이 더 큰 폭으로 증가하는 경향을 발견하였다. 최근 온라인 구전으로 인해 구전효과의 영향이 커지는 시점에서 본 연구는 의미 있는 시사점을 제공한다.

표제어: 온라인 배너 광고, 강화학습, 활용-탐색, 구전효과, 시뮬레이션

접수일(2024년 05월 15일), 수정일(2024년 06월 07일), 게재확정일(2024년 06월 15일)

* 제1저자, 서강대학교 경영대학 교수, bk4498@sogang.ac.kr

** 공동저자, 홍익대학교 경영대학, 교수, gy52@hongik.ac.kr

*** 교신저자, 서강대학교 경영대학 교수, jklee@sogang.ac.kr

1. 서론

최근 온라인 상거래 규모가 급속하게 증가하면서, 온라인 쇼핑과 관련된 온라인 구전효과가 학계와 업계에서 주목을 받고 있다. 국내 2023년 온라인쇼핑물 거래액은 약 229조 원으로 지난 2017년 약 94조 원이었던 것에 비해 두 배 이상 증가했다 (통계청, 2024). 이러한 거래액 증가에 힘입어 온라인 시장에서 소비자의 구매의도에 영향을 미치는 다양한 디지털 마케팅 활동 또한 활발히 진행되고 있다. 그중에서도 온라인 구전은 소비자 행동과 비즈니스 전략을 형성하는 주요한 요인으로 주목받고 있다. 온라인 구전은 소셜 미디어 등을 중심으로 사용자들이 제품에 대해 다양한 상호작용을 하며 전파 속도가 빠르다는 특징을 보인다 (Radighieri and Mulder, 2014).

온라인 배너 광고의 주요한 과제는 최적의 광고 배너를 찾아내는 것이다. 일반적으로 온라인 배너의 디자인 시안은 전달 메시지, 디자인, 크기 등이 다른 복수의 대안이 제작되지만 (Namin et al., 2020), 기업은 고객이 각 대안을 클릭할 확률 (클릭률)을 사전에 파악하기 어렵기 때문에 실험과 시행착오를 클릭률을 추정하고 최적을 찾아내려는 노력을 한다. 기업이 특정 시점에 노출한 대안이 클릭되면 그 대안의 추정 클릭률은 높아지고, 클릭되지 않으면 추정 클릭률은 낮아진다. 따라서 기업은 시점별로 여러 대안을 실험적으로 노출해본 후 각 대안의 클릭수를 바탕으로 클릭률을 추정해야 한다 (Urban et al., 2014). 이렇게 실험을 통해 각 대안의 확률을 추정하고 갱신하면서 최적의 대안을 찾아가는 과정은 강화학습 (reinforcement learning)의 주요 과제이다.

온라인 광고에서 기업은 최적의 배너를 파악해서 정해진 기간 동안 고객이 광고 배너를 클릭하는 총 횟수를 최대화해야 한다. 이를 위해 각 시점에서 노출할 대안을 선택하는 것이 주요 의사결정 문제이고, 이는 강화학습의 활용-탐색 전략과 관련된다. 추정

클릭률이 높은 대안을 다시 노출하는 것은 고객의 클릭 확률을 높이기 위해 축적된 지식을 활용 (exploitation)하는 전략이다. 그러나 추정 클릭률은 실제 클릭률과 다를 수 있기 때문에, 실제 클릭률이 더 높은 잠재적 대안을 발견하기 위해서 추정 클릭률이 낮은 대안을 노출하는 것을 시도해 볼 수 있고, 이것이 탐색 (exploration) 전략이다. 이는 지금까지 방문해본 식당 중에 최고의 식당을 다시 가느냐, 아니면 아직 시도해보지 않았지만 잠재적으로 나올 수 있는 새로운 식당을 시도해보느냐의 문제와 비슷하다.

디지털 시대에 구전효과는 온라인 광고의 효과에 영향을 미치는 주요 요인으로 부상하고 있으며 온라인 공간을 이용한 온라인 구전의 영향력은 더욱 커지고 있다 (Hanssens et al., 2015). 소비자 결정에 직접적인 영향을 미치고, 브랜드 가시성을 확대하며, 경제적으로 효율적인 마케팅 해법을 제공하는 온라인 구전은 디지털 마케팅에서 필수적인 현상이다. 최근 온라인 구전과 관련된 연구에 관한 관심이 증가하고 있지만, 온라인 광고의 강화학습과 관련된 온라인 구전의 영향에 관한 연구는 찾아보기 어렵다. 온라인 광고의 강화학습에 관련해서는 최적의 광고나 웹사이트 대안을 찾기 위한 기술적 방법에 관한 연구가 주로 진행되었다 (Hauser et al., 2014; Liberali and Ferecatu, 2022). 하지만 최근 온라인 광고의 온라인 구전 영향을 고려하면 온라인 광고의 활용-탐색 전략과 관련한 온라인 구전 영향 연구가 필요한 시점이다. 따라서 본 연구에서는 온라인 구전이 온라인 배너 광고의 강화학습에 있어 최적의 활용-탐색 전략에 미치는 영향을 분석하고자 한다.

본 연구의 목적은 다음과 같다. 첫째, 온라인 배너 광고의 강화학습에 있어 구전효과가 최적 활용-탐색 수준에 미치는 영향을 분석한다. 구체적으로, 온라인 구전효과의 크기에 따라 최적 활용-탐색 수준이 어떻게 달라지는지를 살펴본다. 둘째, 구전효과의 영향이 광고 배너 대안의 개수에 의해 어떻게 달라지는

가를 분석한다. 이를 위해 배너 대안이 적어지거나 많아지는 경우에 구전이 최적 활용-탐색 수준에 미치는 영향이 어떻게 변하는지 살펴본다. 이를 통해 주로 최적 대안을 찾는 효율적 방법을 기술적으로 개발하는 최근 연구 주제를 넘어서서, 구전효과와 대안의 수가 최적 전략 탐색에 경영학적으로 미치는 영향을 분석하고자 한다.

2. 이론적 배경

2.1 온라인 배너 광고

온라인 광고에는 크게 검색 광고, 이메일 광고 그리고 배너 광고가 있다. 다양한 온라인 광고 방법 중 사용자들의 직접적인 검색을 통해 노출되는 검색 광고나 수신 여부를 확인하기 어려울 수 있는 이메일 광고에 비해 배너광고는 이용자들이 웹사이트에 접속할 시 강제적으로 노출이 되기 때문에 적은 노력으로 광고 메시지를 전달할 수 있다는 장점이 존재한다.

온라인 배너 광고는 인터넷 사이트의 일부 부분에 광고를 게재하는 형태가 일반적이는데, 초창기에는 단순 문구를 게재하는 정적 배너 위주로 작성되었으나 최근에는 움직이는 이미지나 애니메이션 효과를 포함한 동적 배너와 인터랙티브 (interactive) 배너 등 광고 사이트와 소비자를 연결할 수 있도록 클릭을 유도하는 다양한 배너들이 활용되고 있다.

온라인 배너 광고의 효과성과 관련해서는 다양한 연구가 진행되었다. 온라인 배너 광고 효과성의 측정 지표로 클릭률 (Click through rate)을 활용하여 Li and Bukovac (1999)은 정적인 이미지의 광고보다 동적인 애니메이션으로 구성된 배너 광고가 사용자들의 클릭을 더욱 유도한다고 주장했으며 Lothia et al. (2003)은 적절한 색으로 작성된 배너 광고가 극단적으로 색을 많이 사용하거나 색을 아예 사용하지

않은 배너보다 클릭률이 높은 것을 밝혀 냈다. 이 외에도 배너의 크기 (Chandon et al., 2003), 배너의 형태 (North and Ficorilli, 2017), 메시지 길이 (Robinson et al., 2007) 등 온라인 배너의 특성이 광고 효과에 미치는 영향에 관한 다양한 연구가 진행되었다.

최근에는 이용자 맞춤 온라인 배너 광고의 적절성에 관한 연구 또한 활발하게 진행되고 있다. Stanaland and Tan (2010)은 목표 중심적인 소비자의 경우 단순한 형태의 배너를 선호하는 반면 경험 중심적인 소비자의 경우 복잡하고 상호작용이 가능한 배너를 선호한다고 주장했다. Iyer et al. (2005)은 이용자들이 유용하다고 생각하는 웹사이트의 배너 광고가 클릭률이 높다는 점을 밝혀냈으며 Chu et al. (2009)은 컨텍스트 매칭 (context-matching) 기법으로 배너 광고를 맥락에 적합한 웹사이트에 게재함으로써 클릭률을 향상시킬 수 있다는 결과를 확인했다.

이와 같이 온라인 배너 광고의 효과성에 관한 연구를 살펴보면 배너의 형태뿐만 아니라 다양한 성향의 소비자들에게 어떠한 배너를 노출시킬 것인지에 대한 의사결정 또한 매우 중요하다는 것을 확인할 수 있다. 본 연구에서는 다양한 형태로 디자인된 배너 시안들 가운데 무엇을 어느 시점에 노출할 것인지를 모형화하여 여러 상황에 최적화된 온라인 배너 노출 전략을 분석한다.

2.2 구전효과

구전은 개인 소비자들이 서로 제품 및 서비스에 대한 의견, 정보 및 소식을 공유하고 소통하는 것을 의미한다 (Dichter, 1996). 기존 문헌들을 살펴보면 소비자 행동 및 의사결정에 구전이 미치는 영향에 관한 다양한 연구가 오래 진행되었으며, 대부분의 연구는 구전이 소비자 행동에 유의미한 영향을 미친다는 주장에 동의한다 (Bone, 1995; Duan et al., 2008; Liu, 2006)

여기에 더해 최근에는, 웹 2.0의 지속적인 발전과 모바일기기의 확산으로 소비자들이 온라인을 통해 제품 및 서비스의 정보를 검색하고 전달하며 구매와 관련한 의사결정을 내리는 등 인터넷이 소비자 구매 활동에 커다란 영향력을 미치고 있다. 이와 관련한 대표적 현상이 온라인 구전이다.

전통적인 구전이 대면 상호작용 중에 일어나는 것과 달리, 온라인 구전은 소셜 미디어, 포럼, 리뷰 웹사이트, 블로그 등 다양한 매체를 통해 발생한다. 또한 사용자들 사이에서 자유롭게 의견을 올리고 그 의견에 대한 추천 및 비판이 온라인에서 발생하며 전통적 구전과 비교해 구전의 전파 속도가 매우 빠르다는 특징을 보인다 (Radighieri and Mulder, 2014).

온라인 구전은 구전 전파 속도 외에도 전통적 구전과 다른 특성을 보인다. 첫째, 인터넷을 통해 전파되는 온라인 구전은 지역적 한계를 넘어선다. 즉 온라인 구전은 접근이 용이하고 다양한 소비자들 사이의 효율적인 소통을 가능하게 한다 (Bakos and Dellarocas, 2011; Floyd et al., 2014). 둘째, 온라인 구전은 많은 양의 다양한 의견을 동시에 제공해 소비자들에게 덜 편향적이고 균형 잡힌 정보를 제공한다 (Senecal and Nantel, 2004). 셋째, 전통적 구전이 대체로 잡지 혹은 신문에 기고하는 "전문가"들로부터 시작되는 반면 온라인 구전은 전문가 외에 다수의 일반 소비자들도 참여함으로써 구전 정보원의 유형도 다양하고 각 정보원에 따라 구전이 구매의도에 미치는 영향도 다르다. 구체적으로 살펴보면 전통적 구전에서는 전문가의 의견이 소비자들의 구매의도에 커다란 영향을 미치는 반면 (Sorensen and Rasmussen, 2004) 온라인 구전의 경우 소비자들은 다른 일반 소비자들을 더욱 유사하고 친숙하다고 생각하여 전문가의 의견보다 오히려 일반 소비자들의 의견을 신뢰한다 (Hanssens et al., 2015; Zhang et al., 2010). 마지막으로 온라인 구전의 경우 구전 공유 매체의 디자인, 리뷰 형태, 이용자 등 판매자 혹은

구전 매체 운영자가 구전을 관리하기가 용이하기 때문에 마케터가 온라인 구전 관련 전략을 구축하고 실행하는데 있어 더욱 적극적인 형태를 보인다 (Floyd et al., 2014).

구전효과가 광고 산업에 미치는 영향은 온라인 구전으로 인해 커지고 있다. 관련하여 다수의 연구에서 온라인 구전이 매출에 긍정적 영향을 미친다는 것을 보였다 (Duan et al., 2008; Gu et al., 2012; Liu, 2006). 그러나 현재까지는 강화학습과 관련해서 구전효과와 영향 분석은 여전히 부족하다. 따라서 구전효과가 광고 배너 선택의 최적 활용-탐색 전략에 미치는 영향에 관한 연구는 학문적으로나 실무적으로 많은 시사점을 제공할 수 있다.

2.3 강화학습과 활용-탐색 전략

최적의 대안을 찾기 위한 일련의 과정은 강화학습의 주요한 연구 주제이다 (Sutton and Barto, 2018). 강화학습은 보상을 얻을 확률 분포를 알 수 없는 여러 대안이 주어진 경우 특정 대안의 선택을 통해 보상을 얻을 확률을 추정하는 시행착오를 통해 스스로 보상 확률을 학습하는 알고리즘이다 (Sutton and Barto, 2018). 최적의 강화학습 전략은 보상을 최대화하는 탐색과 활용의 수준을 결정하는 것이다. 강화학습은 에이전트가 환경과 상호작용하여 최적의 행동 정책을 학습하여, 주어진 문제에서 최대의 보상을 얻기 위해 사용된다. 성공적인 적용을 위해서는 문제 설정, 알고리즘 선택, 탐험과 활용의 균형, 자료수집 등이 필수적이다 (Sutton and Barto, 2018; Mnih et al., 2015). 특히 탐험과 활용의 균형을 최적화하기 위한 지속적인 연구는 강화학습의 성능 향상에 필수적이다.

온라인 광고에 있어 강화학습의 어려움은 추정 클릭률이 실제 클릭률과 다르고, 특정 대안의 시도 횟수를 높일 경우에만 그 대안의 추정 클릭률의 정확도가 올라간다는 것이다. 따라서 축적한 지식을

활용하기 위해 추정 클릭률이 높은 대안의 노출 빈도를 높이면 다른 대안의 추정 정확도를 높일 기회를 잃어버리게 된다. 이로 인해 탐색과 활용의 교환 관계가 발생하게 된다. 이러한 교환관계에서 성과를 최대화하는 탐색-활용의 최적 수준을 찾기 위해 온라인 광고와 관련된 강화학습 모형에서는 Multi-Armed Bandit (MAB) 모형 등을 이용해서 최적해 탐색을 위한 기술적 방법론 등이 주로 연구되었다 (Urban et al., 2014).

MAB 모형을 이용한 온라인 광고의 강화학습과 관련된 주요 최신 연구 흐름 중 하나는 광고의 여러 속성이 고객의 클릭률에 미치는 영향에 관한 것이다. 예를 들어, Schwartz et al. (2017)는 MAB 모형을 이용하여 고객의 클릭률에 영향을 미치는 온라인 배너 구조와 속성을 밝히고 현장 실험을 통한 결과를 보였다. 또한 Namin et al. (2020)은 배너의 크기와 종류 (정적, 동적)가 고객의 클릭률에 유의미한 영향을 준다는 것을 실증 데이터를 사용해서 보였다. 이와 더불어 Ferecatu and De Bruyn (2022)과 같이 MAB 기반의 실험을 통해 의사결정자의 탐색-활용 전략 선택에 영향을 미치는 심리적 특성에 관한 연구도 최근 활발히 진행되고 있다. 그러나 이들 연구는 구전효과를 고려한 최적 탐색-활용 수준을 직접적으로 보여주지는 않았다.

온라인 광고에서 탐색은 추정 클릭률이 낮은 대안의 노출 빈도를 높이는 것을 의미한다. 이를 통해 탐색은 여러 대안들을 시도해보고 추정 클릭률의 정확도를 높이고 더 좋은 대안을 찾을 수 있다. 반면에 활용은 추정된 클릭률이 높은 대안의 노출 빈도를 높이는 것을 의미한다. 이를 통해 활용은 현재까지 축적된 지식을 이용해서 배너 클릭률을 높일 수 있다. 활용의 수준이 너무 높으면 잠재적으로 더 나은 대안을 찾는 것을 포기하는 것을 의미하고, 탐색의 수준이 너무 높으면 대안 찾기와 관련된 축적된 지식을 활용하지 않는 것을 의미한다. 따라서 탐색과 활용의 균형점을 찾는 것이 강화학습 전략의 주

요 목적 중의 하나이고 이는 조직의 장기 성과에 영향을 미친다 (Smith and Tushman, 2005).

강화학습 연구는 기본적으로 탐색과 활용의 균형을 가져가는 것이 최적이라는 것을 밝혀냈다. 축적된 지식을 활용하는 것과 새로운 지식을 탐색하는 것 사이에 적절한 균형점을 찾아내는 것이 성과를 최대화할 수 있는 방법이다 (March, 1991). 균형점은 고정된 것이 아니라 연구 모형에서 설정한 환경에 의해서 달라진다. 온라인 배너 광고의 탐색-활용 연구에서는 대부분 각 대안의 초기 클릭률이 처음 설정한 값에서 변하지 않는 조건을 설정하였다. 강화학습의 다른 연구 분야 소수의 연구에서 각 대안이 성공할 확률이 변하는 환경에서 최적 전략을 살펴봤다. 예를 들어 Posen and Levinthal (2012)은 외부 충격에 의해서 대안의 성공률이 무작위로 변화하는 조건에서 최적 탐색-활용 전략에 관해 연구했고, 환경 변화가 있는 경우 상황에 따라 탐색보다는 활용 전략이 더 좋은 성과가 낼 수 있음을 보였다.

하지만 온라인 광고에서는 구전효과로 인해 고객이 배너를 클릭하는 원인이 배너 대안의 클릭률을 변화시키는 인과관계가 존재한다. 인과관계가 존재하는 경우의 클릭률 변화는 무작위적으로 성공률이 변화하는 것과 다른 양상을 보일 것이고, 최적 탐색-활용 전략이 달라지는 결과로 이어질 수 있다. 그럼에도 불구하고 이러한 인과관계가 존재하는 경우의 최적 탐색-활용 전략에 관한 연구는 찾아보기 어렵다. 따라서 본 연구에서는 구전효과에 의한 인과적 클릭률 변화가 있을 경우의 최적 전략에 대해 분석하고자 한다.

3. 모형

강화학습과 관련된 연구는 실증적으로 검증하기 어려운 특징을 가지고 있기 때문에 많은 연구에서 시뮬레이션 방법론을 사용한다 (Burton and Obel,

2011). 본 연구에서는 Multi-Armed Bandit (MAB) 모형 (Robbins, 1952)을 활용한 시뮬레이션 모형으로 온라인 광고에서의 구전효과를 분석한다. MAB 모형은 강화학습의 탐색-활용 문제를 연구하기 위해 널리 사용되고 있다 (Posen and Levinthal, 2012; Shahrokhi and Ching 2019). MAB 모형은 매우 다양한 변주가 있지만, 본 연구에서는 전통적인 (classical) MAB 모형을 기본적으로 사용한다. 따라서 총 배너의 수는 고정되어 있고, 한 번에 하나의 배너가 노출되며, 노출되지 않는 배너는 클릭되지 않는다는 가정을 한다 (Mahajan and Teneketzis, 2008). 그러나 전통 MAB 모형과 다르게 고객이 노출된 배너를 클릭하는 경우, 구전효과에 의해 다음 기간의 배너들의 클릭률이 확률적으로 증가할 수 있으며 이는 Luger et al. (2018)과 Posen and Levinthal (2012)의 모형의 가정과 유사하다.

본 연구에서의 시뮬레이션 모형은 다음과 같은 순서로 진행된다. (0) 초기 설정. 이후 전체 기간 1~T 동안의 각 t에서 다음을 실행함. (1) 노출할 대안 i 선택 및 노출 (탐색 수준에 따라 확률적으로 선택됨), (2) 고객의 클릭 여부 결정 (대안 i의 실제 클릭률에 따라 확률적으로 결정됨), (3) 대안 i에 대한 추정 클릭률 갱신 (고객이 클릭했으면 추정 클릭률 높임, 클릭하지 않으면 낮춤), (4) 구전효과에 따라 대안들의 실제 클릭률 변화 (고객이 노출된 배너 클릭한 경우 구전효과로 인해 대안들의 실제 클릭률 확률적으로 증가).

MAB 모형에서 기업은 N개의 광고 배너 대안을 가지고 있다. 그러나 각 대안을 고객이 클릭할 실제 확률을 사전에 알지 못하고 고객의 클릭 여부에 의해 추정하면서 최적의 대안을 찾아야 한다. 초기 추정 클릭률은 실제 대안들의 클릭률의 평균이고, 추정 클릭률 갱신은 March (1991)의 방식을 따라 계산한다. t시점에서 대안 i를 노출하는 경우 대안 i의 추정 클릭률은 $q_{i,t+1} = q_{i,t} + (\sigma - q_{i,t}) / (k_{i,t} + 1)$ 로 갱신된다. $q_{i,t}$ 는 t시점의 대안 i에 대한 추정 클릭률이고, σ 는

고객의 클릭 여부이고, $k_{i,t}$ 는 t 시점까지 대안 i를 노출한 횟수이다. 만약 고객이 노출된 대안을 클릭하면 $\sigma=1$ 이고 그 대안의 추정 클릭률은 높아지고, 클릭하지 않으면 $\sigma=0$ 이고 그 대안의 추정 클릭률은 낮아진다. 기업은 1에서 T사이의 각 시점에서 고객에게 노출할 대안을 선택하고 그 결과에 따라 대안의 클릭률을 추정하고 갱신한다.

활용-탐색 수준에 의한 대안 선택 확률은 Softmax 선택 기준을 따른다 (Luce, 1959). Softmax 기준은 다양한 연구에서 광범위하게 사용되고 있는 검증된 방법이고 (Posen and Levinthal 2012; Park et al., 2018), 실제 사람의 의사결정과 유사하다 (Daw et al., 2006). 기본적으로 추정 클릭률이 높은 대안이 낮은 대안보다 노출될 확률이 높다. 그러나 확률의 상대적 크기는 기업의 탐색-활용 전략의 수준에 따라 달라진다. 기업이 대안 i를 노출할 확률은 $m_i = e^{(q_i / (\tau^{10}))} / \sum_{j=1, \dots, N} [e^{(q_j / (\tau^{10}))}]$ 로 계산한다. 기업의 탐색 전략 수준은 τ 에 의해서 결정된다. 탐색 수준이 높으면 추정 클릭률이 낮은 대안을 노출할 확률이 상대적으로 높고, 탐색 수준이 낮으면 추정 클릭률이 높은 대안을 노출할 확률이 상대적으로 높다. 탐색 수준을 0으로 낮추면 의사결정 시점에서의 추정 클릭률이 가장 높은 대안을 선택하게 되며, 이는 greedy (탐욕) 전략이 된다 (Auer et al., 2002). 본 연구에서는 기존 연구에 근거하여 매우 낮은 탐색에서 매우 높은 탐색까지 7가지 수준의 탐색-활용 전략을 사용하였고, 이에 대응하는 τ 는 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35이다 (Posen and Levinthal 2012).

온라인 광고 환경에서의 구전효과는 고객이 노출된 배너를 클릭하게 되면 다음 시기의 배너 클릭률이 높아지는 것으로 구현된다. 이는 배너를 클릭한 고객이 직접 구전 혹은 온라인 구전을 통해 제품을 주변에 홍보한다는 가정에 의한 것이다. 고객이 특정 배너를 홍보하는 것이 아니라 광고 대상 제품을 홍보한다는 점을 고려해서, 클릭률이 높아지는 배너는 무작위로 선정되고 클릭률 증가는 특정 분포를

통해 정해지게 된다. 따라서 노출된 광고 배너가 선택되게 되면 다음 시기에 어떤 배너 대안의 클릭률이 얼마나 높아질지는 확률적으로 결정된다.

본 연구에서는 구전효과의 수준을 '없음', '중간', '높음'의 세 단계로 설정하였다. 모형에서 초기 클릭률은 Beta (1,50)을 따르고, 최대 클릭률은 0.5로 설정되었다. 이는 클릭률이 높은 대안을 Beta 분포를 이용해서 제한적으로 발생시키는 이전 연구들의 설정과 같다 (Posen and Levinthal 2012, Uotila, 2017). 구전효과가 작동할 경우 선택된 대안의 클릭률 증가는 최대 클릭률과 현재 클릭률의 차이에 Beta 분포에서 무작위 추출한 값을 곱해서 결정하였다 (구전효과 '중간'은 Beta (1,30), '높음'은 Beta (1,15)에서 추출함). 평균 초기 클릭률은 0.02이고 평균 최종 클릭률은 구전효과 '없음'은 0.02, '중간'은 0.16, '높음'은 0.29의 수치를 보인다.

구전효과의 영향은 의사결정자가 몇 개의 광고 배너 대안을 가지고 있느냐에도 영향을 받는다. 대안의 수가 적다면 클릭률이 가장 높은 최적의 대안을 찾는 것이 더 용이하게 된다. 그러나 최적의 대안을 찾기 위해 탐색의 수준을 높이면 현재까지 추정된 최적 대안을 포기하는 것이기 때문에 탐색-활용의 교환관계는 여전히 존재한다. 본 연구에서는 구전효과가 존재하는 상황에서 대안의 개수가 최적 강화학습 전략에 미치는 영향에 대해서 분석하고자 한다. 이를 위해 광고 배너 대안의 수를 적음 (5개), 중간 (10개), 많음 (30개)로 설정하여 연구를 진행한다. 이는 현재 업계에서 가장 큰 온라인 광고 배너 광고 회사인 Google AdSense와 Facebook audience network이 제공하는 온라인 광고 가이드라인에서 적정 광고 시안을 5개 정도로 제안하고, Schwartz et al. (2017)의 MAB를 이용한 온라인 광고 배너 클릭 실험에서 12개의 대안을 이용한 것을 참고한 것이다.

따라서 구전효과의 크기가 세 가지 수준, 배너 대안의 개수가 세 가지 수준으로 이들의 조합으로 총

9개의 시나리오를 분석한다. 이를 통해 구전효과와 배너 대안 개수의 영향을 종합적으로 고려한 최적 탐색-활용 전략의 경향을 파악할 것이다.

4. 분석 결과

본 연구는 온라인 광고의 강화학습에서의 구전효과와 영향을 분석하기 위해 MAB 시뮬레이션 모형을 사용하여 다양한 수준의 구전효과와 광고 배너 대안 개수에서 최적 탐색-활용 전략을 분석하였다. 전체 기간 T는 1,000으로 설정했다. 시뮬레이션 횟수는 10,000번이고 모든 결과값은 10,000번의 평균이다.

탐색-활용 전략의 성과는 '총클릭횟수'에 의해 결정된다. 총클릭횟수는 전체 1,000 기간 중 고객이 배너를 클릭하는 횟수이다. 9개 시나리오 각각의 환경에서 7단계의 탐색-활용 전략 중 총클릭횟수가 가장 많은 전략이 최적 전략이 된다.

강화학습의 핵심은 각 광고 배너의 실제 클릭률 추정 정확도이다. 이를 평가하기 위해 '추정정확도'를 사용하였다. 추정정확도는 '1 - 각 대안의 추정 클릭률 오차의 제곱의 합'으로 정의되고, 이는 기존 연구의 평가척도와 동일하다 (Posen and Levinthal, 2012). 추정정확도가 높으면 실제 클릭률과 추정 클릭률의 오차가 적다는 것을 의미한다.

각 시나리오 환경에서 탐색-활용 수준에 따른 선택 행태는 'Greedy' 수준으로 추정했다. Greedy는 추정 클릭률이 가장 높은 대안을 선택할 확률이다. 따라서 Greedy가 높으면 추정되는 최고의 대안을 노출할 확률이 높다. 탐색의 수준이 낮아지면 Greedy 수준이 높아지게 된다.

결과 분석과 해석을 위해서 'NonGreedy성공'의 지표를 활용했다. NonGreedy성공은 추정 클릭률이 최고가 아닌 대안을 선택했을 때, 고객이 노출된 대안을 클릭하는 비율을 측정한 것이다. 따라서 이 지

표는 탐색 전략을 사용할 때의 성공률을 근사적으로 보여준다.

Tab. 4-1은 분석 결과를 요약한 정보를 보여주고, Fig. 4-1은 구전효과 수준과 대안의 수에 따른 최적 탐색 수준의 변화를 보여준다. 먼저 전체 성과의 변화를 살펴보면, 구전효과가 커질수록 그리고 대안의 수가 적을수록 총클릭횟수가 높아짐을 알 수 있다. 이는 구전효과가 커지면 시간이 지남에 따라 평균 클릭률이 증가하고, 대안의 수가 적으면 최적 대안을 찾는 것이 상대적으로 더 수월해지기 때문으로 해석할 수 있다.

강화학습의 최적 탐색-활용 수준을 살펴보면, 극단적 탐색 혹은 활용이 아니라 탐색과 활용이 균형을 이루는 지점이 최적일 됨을 알 수 있다. 이는

March (1991) 연구 결과와 일치한다. 탐색의 수준을 높이면 추정 클릭률이 낮은 대안도 선택 횟수가 높아져 추정의 정확도가 높아지지만, 그 과정에서 추정 클릭률이 높은 대안을 포기해야 하는 비용이 발생한다. 따라서 최적 전략은 활용-탐색 수준이 두 극단의 사이에서 결정된다. 예를 들어 구전효과가 없고 대안의 수가 30인 시나리오에서는, 추정 클릭률이 가장 높은 대안을 선택하는 Greedy 수준은 탐색의 수준이 높아질수록 0.879에서 0.242으로 낮아지고, 추정정확도는 0.986에서 0.988로 높아진다. 이 경우에는 추정 클릭률이 가장 높은 대안을 포기하는 것에 비해서 추정정확도가 높아지는 정도가 유의미하게 높아지지 않는다. 따라서 탐색 수준이 2가 최적 전략이 된다. 그럼에도 불구하고 탐색 수준은 극

Tab. 4-1. Summary of Analysis Results

대안 탐색 수	구전효과: 없음				구전효과: 중간				구전효과: 높음			
	총클릭 횟수	추정 정확도	Greedy	NonGreedy 성공	총클릭 횟수	추정 정확도	Greedy	NonGreedy 성공	총클릭 횟수	추정 정확도	Greedy	NonGreedy 성공
1	46.4	0.986	0.879	0.020	60.1	0.915	0.901	0.023	79.5	0.674	0.915	0.025
2	47.6	0.987	0.711	0.023	63.8	0.909	0.763	0.029	92.0	0.615	0.820	0.032
3	44.1	0.987	0.552	0.024	62.7	0.914	0.622	0.034	96.3	0.602	0.717	0.039
30 4	40.2	0.988	0.432	0.024	57.8	0.926	0.493	0.036	94.9	0.623	0.614	0.045
5	36.2	0.988	0.344	0.024	52.1	0.938	0.393	0.035	90.3	0.662	0.517	0.048
6	33.6	0.988	0.287	0.023	48.0	0.947	0.324	0.035	84.2	0.711	0.434	0.050
7	31.0	0.988	0.242	0.023	44.4	0.954	0.272	0.034	78.2	0.752	0.366	0.051
	평균	0.987	0.492	0.023		0.929	0.538	0.032		0.663	0.626	0.041
1	41.2	0.996	0.887	0.017	83.5	0.820	0.922	0.022	142.3	0.202	0.933	0.022
2	41.3	0.997	0.731	0.019	92.7	0.802	0.836	0.032	166.2	0.071	0.887	0.032
3	39.3	0.997	0.591	0.020	95.0	0.808	0.734	0.041	176.4	0.038	0.835	0.043
10 4	36.4	0.998	0.484	0.020	92.2	0.831	0.628	0.048	179.9	0.058	0.772	0.054
5	34.3	0.998	0.411	0.020	89.4	0.852	0.541	0.053	179.4	0.092	0.713	0.066
6	32.0	0.998	0.356	0.020	83.8	0.879	0.463	0.054	177.5	0.158	0.648	0.077
7	30.1	0.998	0.316	0.020	79.9	0.897	0.405	0.056	173.6	0.227	0.587	0.086
	평균	0.998	0.539	0.020		0.841	0.647	0.044		0.121	0.768	0.054
1	35.5	0.999	0.885	0.014	133.5	0.650	0.930	0.020	237.5	0.156	0.942	0.021
2	35.9	0.999	0.746	0.016	145.4	0.628	0.876	0.030	256.8	0.133	0.910	0.028
3	34.2	0.999	0.629	0.017	147.5	0.642	0.809	0.042	267.3	0.141	0.881	0.038
5 4	31.9	0.999	0.543	0.017	149.8	0.660	0.747	0.056	271.4	0.167	0.850	0.051
5	30.7	0.999	0.484	0.017	147.7	0.690	0.682	0.065	272.1	0.201	0.817	0.065
6	29.4	1.000	0.441	0.017	143.5	0.723	0.619	0.074	273.0	0.239	0.783	0.081
7	27.9	1.000	0.408	0.016	141.7	0.748	0.569	0.082	272.0	0.279	0.749	0.097
	평균	0.999	0.591	0.016		0.677	0.747	0.053		0.188	0.847	0.054

단인 1이 아닌 2가 최적이 됨을 확인할 수 있다. 구전효과가 중간이고 대안의 수가 5인 시나리오를 살펴보면, Greedy 수준은 탐색 수준이 높아지면 0.930에서 0.569로 줄어들고, 추정정확도는 0.650에서 0.748으로 늘어나며, 최적 활용 수준은 중간인 4에서 결정된다. 이와 같이 활용의 수준을 높이면 Greedy 수준은 낮아지지만 추정정확도 수준이 높아지며, 둘의 교환관계로 인해 최적 활용 수준은 중간 수준에서 결정됨을 알 수 있다. 실험 시나리오의 설정 환경에 따라 최적 수준은 달라진다.

강화학습에서의 구전효과의 영향을 살펴보면, 구전효과가 높아질수록 최적 탐색 수준이 높아짐을 알 수 있다. 구전효과 크기가 없음, 중간, 높음 수준으로 커짐에 따라 대안의 수가 30인 경우에 최적 탐색 수준은 2, 2, 3으로 증가하고, 대안의 수가 10인 경우는 최적 탐색 수준이 2, 3, 4로 증가하고, 대안의

수가 5인 경우에는 최적 탐색 수준이 2, 4, 6으로 증가한다.

구전효과가 높아짐에 따라 추정정확도는 낮아지는 경향을 보인다. 예를 들어, 대안의 수가 10인 경우 평균 추정정확도는 0.998에서 0.841, 그리고 0.121로 낮아진다. 이는 고객이 배너를 클릭하고 이를 타인에게 전파하는 구전효과가 있을 경우 각 배너의 실제 클릭률이 초기 클릭률에서 벗어나서 증가하기 때문이며, 따라서 과거에 축적한 지식의 정확도가 떨어지는 결과가 나타난다. 클릭률이 증가하는 대안과 증가 폭은 확률적으로 결정되기 때문에 구전효과가 커지면 추정정확도는 낮아지게 된다. 결과적으로 구전효과가 높아지면, 기존에 축적했던 지식의 효용 가치가 상대적으로 떨어지고 새로운 대안을 탐색해서 잠재적으로 더 좋은 대안을 찾는 것의 가치가 상대적으로 높아지게 된다.

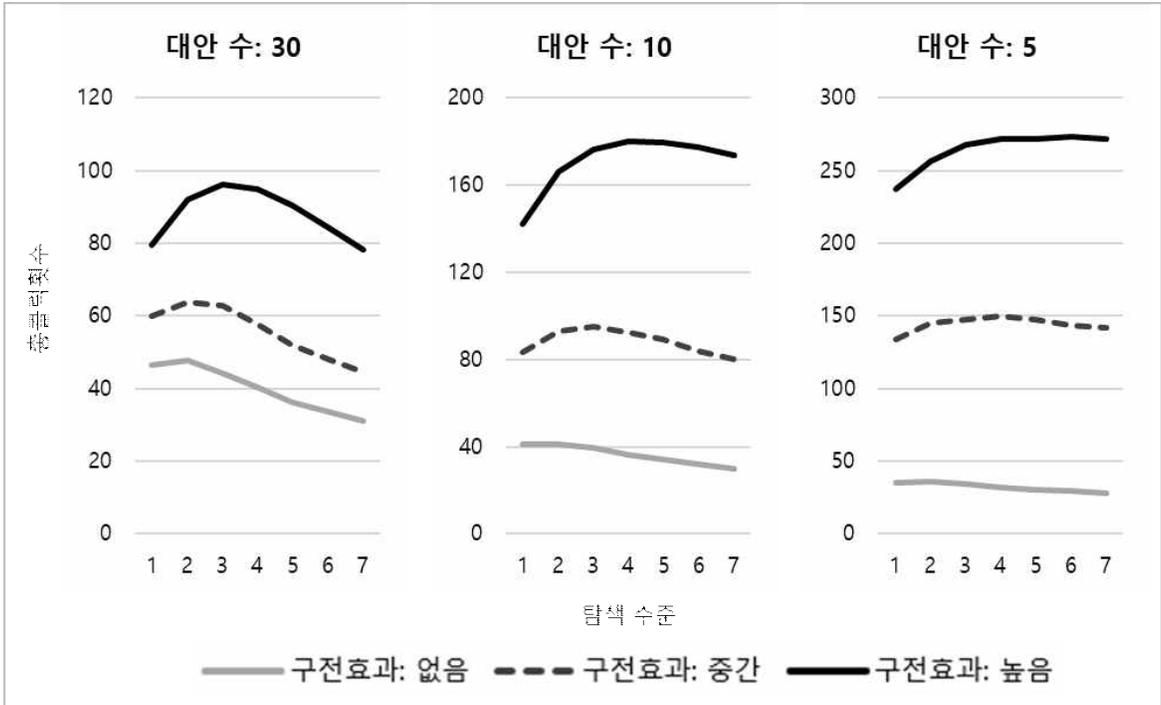


Fig. 4-1. Changes in Optimal Exploration Level Based on Word-of-Mouth Effects and the Number of Alternatives

NonGreedy성공은 추정 클릭률이 최고가 아닌 대안을 탐색해서 노출할 경우에 그 선택이 성공하는 (고객이 클릭할) 비율을 보여주는 지표로써 탐색 전략의 가치를 간접적으로 나타낸다. 분석 결과에서 구전효과가 높아지면 NonGreedy성공 지표가 높아지는 것이 나타났다. 예를 들어 대안의 수가 10인 경우 평균 NonGreedy성공은 구전효과 증가에 따라 0.020에서 0.044, 그리고 0.054로 증가한다. 즉 구전효과가 커짐에 따라, 특정 시점까지 추정된 최고의 대안이 아닌 다른 대안을 탐색적으로 선택하는 모험이 성공할 가능성이 커지는 것이다. 종합해보면, 구전효과가 커질수록 그동안 축적했던 지식의 가치는 낮아지고 새로운 것을 탐색해보는 것의 가치는 높아지기 때문에 최적 탐색의 수준이 높아지는 것을 알 수 있다.

다음으로 광고 배너 대안의 수의 영향을 살펴보면, 대안의 수가 줄어들수록 최적 탐색 수준이 높아짐을 알 수 있다. 예외적으로 구전효과가 없는 경우에는 대안의 수 감소에 따라 추정정확도와 NonGreedy성공 지표가 의미 있게 변화하지 않았고 최적 탐색 수준은 2로 유지되었다. 따라서 구전효과가 없는 경우에는 광고 배너 대안의 수가 최적 탐색-활용 전략에 큰 영향을 미치지 않는 것을 알 수 있다. 그러나 구전효과가 중간이거나 높은 경우에는 대안의 수가 줄어들 경우 최적 탐색 수준이 높아지는 경향을 보였다. 구전효과가 중간인 경우 대안의 수 감소에 따라 최적 탐색 수준은 2, 3, 4로 증가하였고, 구전효과가 큰 경우에는 최적 탐색 수준은 3, 4, 6으로 증가하였다.

이와 관련하여 분석 지표를 살펴보면, 구전효과가 없는 경우는 추정정확도가 유의미하게 변화하지 않았다. 구전효과가 중간인 경우는 대안의 수 감소에 따라 평균 추정정확도는 0.929에서 0.841, 그리고 0.677로 감소하는 경향을 보였다. 구전효과가 있는 경우에는 고객이 노출된 배너를 클릭해서 구전효과가 작동하는 경우 실제 클릭률이 증가해서 기존 추

정 클릭률에서 더 벗어나게 된다. 대안의 수가 적으면 하나의 대안의 오차가 전체 평균에 미치는 영향이 높기 때문에 추정정확도가 낮아지게 된다. 특히 하계도, 구전효과가 높은 경우에는 대안의 수가 중간인 10인 경우에 추정정확도가 가장 낮았다. 구전효과에 의해 낮아진 추정의 정확도는 다음 시기에 대안을 선택하고 고객의 클릭 여부 확인에 의해 다시 회복되게 된다. 대안의 수가 적은 경우에는 오차가 높아진 대안을 선택할 확률이 상대적으로 높아지게 되므로 추정정확도를 높일 가능성이 있다. 결과적으로 대안의 수는 추정정확도에 상반되는 두 가지 영향을 미친다. 대안의 수가 적으면 한 대안의 추정 오차가 전체에 미치는 영향이 커지므로 추정정확도가 낮아지고, 반대로 오차가 큰 대안을 선택해서 강화학습을 통해 추정정확도를 높일 가능성이 커진다. 이 두 가지 힘의 영향은 상황에 따라 달라지고, 구전효과가 높은 경우에는 이들의 영향으로 대안의 수가 중간인 10인 경우에 추정정확도가 가장 낮게 나타났다.

NonGreedy성공 지표는 대안의 수 감소에 따라 높아지는 경향성을 가진다. 이는 대안이 적기 때문에 최고 추정 클릭률을 가지는 대안이 아닌 다른 것을 시도해 볼 경우 우연히 실제 클릭률이 높은 대안을 선택할 확률이 커지기 때문으로 해석할 수 있다. 결과적으로 대안의 수가 적어지면 최적 탐색 수준은 높아지는 경향을 보이게 된다. 또한 이 효과는 구전효과가 높은 경우에 더 커진다. 구전효과가 중간인 경우 NonGreedy성공은 0.032에서 0.044, 그리고 0.053으로 증가했지만 구전효과가 높은 경우에는 0.041에서, 0.054, 0.054로 증가했다.

5. 강건성 분석 (robustness test)

시뮬레이션 결과의 강건성을 총 기간, 초기 클릭률의 분포, 대안의 수, 구전효과에 영향을 등에 대해

분석하였고 최적 탐색-활용 수준의 경향은 변하지 않는 것을 검증하였다. 먼저 총 기간을 1,000이 아닌 500과 3,000으로 분석하였다. 다음으로 초기 클릭률을 Beta (1,50)이 아닌 Beta (1,30)과 Beta (1,100)을 사용하여 분석하였다. 그리고 대안의 수를 3, 20, 50을 사용하여 분석하였다. 마지막으로 구전효과의 영향을 더 다양한 수준으로 세분화하고 넓은 범위를 사용하여 분석하였다. 결과적으로 모든 분석에서 구전효과의 영향이 커질수록, 대안의 수가 적어질수록 최적 탐색 수준이 높아지는 것을 확인하였다.

6. 결론 및 시사점

본 연구는 온라인 배너 광고 강화학습의 최적 대안 선택에서 구전효과의 크기와 대안의 수가 최적 활용-탐색 전략 수준에 미치는 영향을 분석하였다. 분석 결과 구전효과의 크기가 커질수록 그리고 광고 배너 대안의 수가 적을수록 최적 탐색 수준이 높아지는 것을 보였다. 이는 구전효과에 의해 고객이 광고 배너를 클릭할 확률이 증가함에 따라 기존에 축적했던 추정 클릭률 지식의 가치가 낮아지고 따라서 새로운 대안을 탐색하는 것의 가치가 증가하기 때문이다. 광고 배너 대안의 수가 적어지면 추정 클릭률 지식의 가치는 낮아지고 새로운 대안의 탐색 성공률은 높아지는 경향을 보였다. 따라서 대안의 수가 많은 경우에는 상대적으로 구전효과의 영향이 적었고, 대안의 수가 적은 경우에는 구전효과 크기가 커질 때 최적 탐색 수준이 더 큰 폭으로 증가하는 경향을 보였다.

본 연구는 강화학습과 관련하여 구전효과의 영향을 분석했다는 점에서 학문적 시사점을 지닌다. 배너 광고와 관련된 기존 강화학습 연구에서는 각 대안의 클릭 확률이 초기 설정 값에서 변하지 않는다는 가정을 하는 경우가 대부분이었다. 광고 분야 이외의 강화학습 관련 연구에서는 고객 클릭률이 변하

는 경우를 다룬 연구가 존재한다. Rojas-Cordova et al. (2023)은 문헌연구를 통해 환경 조건과 조직 조건에 따라 최적 탐색-활용 수준이 달라질 수 있음을 보이면서, 성공률 (본 연구에서의 고객 클릭률에 대응되는 개념)이 동적으로 변하는 경우 탐색이 조직 성과를 높이는 경우와 반대로 탐색이 조직 성과를 낮추는 경우에 대해 분석하고, 이는 조직의 복잡도에 따라 다를 수 있음을 보였다. 그러나 이와 관련된 일련의 연구에서는 성공률이 무작위로 변하는 경우만을 분석했고, 고객이 대안을 클릭하는 행위에 의해서 대안들의 클릭률이 변하는 인과관계가 있을 경우에 대한 분석은 다루지 않았다. 결론적으로 온라인 광고와 관련된 연구 혹은 다른 강화학습 관련 연구에서도 이러한 인과관계가 존재하는 경우의 최적 탐색-활용에 관련된 연구는 찾아보기 어렵다.

온라인 구전으로 인해 구전효과의 영향이 지속적으로 중요해지는 것을 고려하면, 고객의 클릭 행위와 대안의 클릭률 변화라는 인과관계를 고려한 구전효과가 최적 탐색-활용 전략에 미치는 영향과 그 방향성을 밝혀내는 것은 학문적 의의가 크다고 할 수 있다. 또한 본 연구는 광고 배너 대안의 수가 구전효과와 연계되어 강화학습에 복합적으로 미치는 영향에 대해서도 분석하였다. 이 과정에서 대안의 수가 지식의 가치를 증가시키기도 하고 감소시키기도 하는 두 가지 상반된 효과에 대해서도 분석함으로써 관련 분야의 연구에 학문적으로 공헌하였다.

본 연구는 경영자에게 최적 광고 배너 선택 지침을 제공한다는 점에서 실무적 시사점을 지닌다. 경영자는 복수의 배너 대안의 클릭률을 사전에 알아낼 수 없기 때문에 시행착오를 통해 실험적으로 최적의 대안을 추정할 수밖에 없다. 이러한 환경에서 최적의 탐색-활용 수준과 구전효과와 대안의 수의 영향의 방향성에 관한 본 연구는 경영자가 성과를 높이기 위해 최적 대안을 선택하는 데 있어서 중요한 수단으로 작용할 것이다. 구전효과가 크고, 제작된 대안이 적은 경우 최적의 추정 대안이 아닌 다른 대안

을 시도함으로써 온라인 광고의 성과와 효과를 높일 수 있다는 것은 실무적으로 유용한 지침으로 활용될 것이다.

본 연구는 온라인 배너 광고의 맥락에서 구전효과가 있을 경우 최적 탐색-활용 전략에 대해서 분석하였지만, 연구 결과는 다른 광고 방식에도 적용할 수 있다. 기업이 여러 광고 대안 중 하나를 노출하고 고객의 반응을 즉각적으로 확인할 수 있는 환경에서는 본 연구의 결과를 확장해서 적용할 수 있다. 따라서 앱 푸시 광고, 유튜브의 중간 광고, 온라인 검색 광고, 웹사이트 팝업 광고 등에서 최적 광고 대안을 찾아야 하는 문제에서는 본 연구의 결과를 활용할 수 있을 것이다.

본 연구는 구전효과와 대안의 수가 최적 탐색-활용에 미치는 영향을 분석했지만, 대안의 초기 클릭률의 분포는 대안의 수에 의해 영향을 받지 않는다는 가정을 하였다. 이 경우에도 대안의 수가 많아지면 높은 초기 클릭률을 가지는 대안이 확률적으로 발생할 가능성은 높아진다. 그러나 대안의 수가 많아지는 경우에 초기 클릭률의 분포 자체가 변하는 가능성에 대해서는 고려하지 않았다. 실무에서는 고객의 선호를 미리 알 수 없기 때문에, 많은 광고 배너 대안을 제작해서 그중에 고객이 특별히 선호하는 높은 클릭률의 대안이 잠재적으로 있기를 기대할 수 있다. 대안의 수가 많아지면 실제 클릭률을 추정해야 하는 선택지의 수가 많아지게 때문에 최적 대안을 찾는 강화학습 과정에 어려움이 있을 수 있다. 그럼에도 불구하고, 각 대안의 초기 클릭률 분포가 대안의 수에 의해 변하는 경우에 관한 연구는 의미 있는 결과를 보여주었다고 판단된다.

이 외에도 경쟁사의 배너의 영향, 제한된 예산이 미치는 영향 등에 관해서는 본 연구에서 다루지 않았다. 시뮬레이션 연구의 특성상 모든 변수를 다 고려하지 못하기 때문에 제한적으로 중요한 변수를 다룬 점이 본 연구의 한계점이라 할 수 있다. 향후 연구에서는 온라인 광고와 관련해서 구전효과와 대안

의 수 이외에 다른 의미 있는 변수를 추가해서 본 연구를 확장할 수 있을 것이다.

또한 본 연구는 구전효과에 의한 클릭률 증가 크기를 임의로 설정하였다는 한계가 있다. 이는 구전효과 크기에 따라 탐색-활용 전략의 최적 수준이 변화하는 방향성과 경향을 보여주기 위함이었지만, 실제 산업계에서 체감하는 구전효과에 의한 클릭률 변화를 직접적으로 모형의 모수 (parameter)로 이용하는 못하였다. 향후 연구에서 실제 온라인 광고 업계에서의 구전효과 크기를 추정하여 모형에 적용한다면, 더욱 실용적인 결과를 보여줄 수 있을 것이다.

[References]

- [1] 2011. 통계청 (2024), <https://kosis.kr>.
- [2] Auer, P., Cesa-Bianchi, N. and Fischer, P. (2002), Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47 (2), 235-256.
- [3] Bakos, Y., and Dellarocas, C. (2011), Cooperation without Enforcement? A Comparative Analysis of Litigation and Online Reputation as Quality Assurance Mechanisms. *Management Science*, 57(11), 1944-1962.
- [4] Bone, P.F. (1995), Word-of-Mouth Effects on Short-Term and Long-Term Product Judgements. *Journal of Business Research*, 32(3), 213-223.
- [5] Burton, R.M. and Obel, B. (2011), Computational modeling for what-is, what-might-be, and what-should-be studies-and triangulation.

- Organization Science*, 22(5), 1195-1202.
- [6] Chandon, J.L., Chtourou, M.S., and Fortin, D.R. (2003), Effects of configuration and exposure levels on responses to web advertisements. *Journal of Advertising*, 43(3), 217-229.
- [7] Chu, W., Park, S-T., Beaupre, T., Motgi, N., Phadke, A., Chakraborty, S., and Zachariah, J. (2009), A case study of behavior-driven conjoint analysis on Yahoo! Front Page Today module. Proc. 15th ACM SIGKDD *International Conference Knowledge Discovery Data Mining*, 1097-1104.
- [8] Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., and Dolan, R. J. (2006), Cortical substrates for exploratory decisions in humans. *Nature*, 441(7095), 876-879.
- [9] Dichter, E. (1966), How Word-of Mouth Advertising Works. *Harvard Business Review*, 44(6), 147-166.
- [10] Duan, W., Gu, B., and Whinston A.B. (2008), The dynamics of online word-of-mouth and product sales - An empirical investigation of the movie industry. *Journal of Retailing*, 84(2), 233-242.
- [11] Ferecatu, A., and De Bruyn, A. (2022), Understanding managers' trade-offs between exploration and exploitation. *Marketing Science*, 41(1), 139-165.
- [12] Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y., and Freling, T. (2014), How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90(2), 217-232.
- [13] Gu, b., Park, J., and Konana, P. (2012), Research Note - The Impact of External Word-of-Mouth Sources on Retailer Sales of High-Involvement Products, *Information Systems Research*, 23(1), 182-196.
- [14] Hanssens, D. M., Villanueva, J., and Yoo, S. (2015), Word-of-mouth and marketing effects on customer equity. *Handbook of Research on Customer Equity in Marketing*, 178-198.
- [15] Hauser, J.R., Liberali, G., and Urban, G.L. (2014), Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph. *Management Science*, 60(6), 1594-1616.
- [16] Iyer, G., Soberman, D., and Villas-Boas, J.M. (2005), The targeting of advertising. *Marketing Science*, 24(3), 461-476.
- [17] Li, H., and Bukova, J.L. (1999), Cognitive impact of banner ad characteristics: An experimental study. *Journalism and Mass Communication Quarterly*, 76(2), 341-353.
- [18] Liberali, G., and Ferecatu, A. (2022), Morphing for consumer dynamics: Bandits meet hidden markov models. *Marketing Science*, 41(4), 769-794.
- [19] Liu, Y. (2006), Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3), 74-89.

- [20] Lothia, R., Donthu, N., and Hershberger, E.K. (2003), The impact of content and design elements on banner advertising click-through rates. *Journal of Advertising Research*, 43(4), 410-418.
- [21] Luce, R. (1959), *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.
- [22] Luger, J., Raisch, S., and Schimmer, M. (2018), Dynamic balancing of exploration and exploitation: The contingent benefits of ambidexterity. *Organization Science*, 29(3), 449-470.
- [23] March, J.G. (1991), Exploration and exploitation in organizational learning. *Organization Science*, 2(1), pp.71-87.
- [24] Mahajan, A., and Teneketzis, D. (2008), Multi-armed bandit problems. In *Foundations and applications of sensor management* (pp. 121-151). Boston, MA: Springer US.
- [25] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., and Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529-533.
- [26] Namin, A., Hamilton, M.L., and Rohm, A.J. (2020), Impact of message design on banner advertising involvement and effectiveness: An empirical investigation. *Journal of Marketing Communications*, 26(2), 115-129.
- [27] North, M., and Ficorilli, M. (2017), Click Me: An Examination of the Impact Size, Color, and Design has on Banner Advertisements Generating Clicks. *Journal of Financial Services Marketing*, 22, 99-108.
- [28] Park, J.W., Kim, J.B., and Choi, Y.L. (2018), A Study on Performance Improvement with Combination of Softmax and UCB Algorithm, *Journal of The Korea Society of Information Technology Policy & Management*, 10 (1), 649-654 (박주원, 김재범, 최용락. (2018), Softmax 와 UCB 알고리즘의 결합을 통한 성능 개선 연구-베이지안 확률론을 이용하여. *한국 IT 정책경영학회 논문지*, 10(1), 649-654).
- [29] Posen, H.E. and Levinthal, D.A. (2012), Chasing a moving target: Exploitation and exploration in dynamic environments. *Management Science*, 58(3), 587-601.
- [30] Radighieri, J.P., and Muler, M. (2014), The Impact of source effects and message valence on word of mouth retransmission. *International Journal of Market Research*, 56(2), 249-263.
- [31] Robbins, H. (1952), Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5), 527-535.
- [32] Robinson, H., Wysocka, A., and Hand, C. (2007), Internet Advertising Effectiveness: the Effect of Design on Click-through Rates for Banner Ads. *International Journal of Advertising*, 26(4), 527-541.

- [33] Rojas-Cordova, C., Williamson, A.J., Pertuze, J. A., and Calvo, G. (2023), Why one strategy does not fit all: a systematic review on exploration-exploitation in different organizational archetypes. *Review of Managerial Science*, 17(7), 2251-2295.
- [34] Schwartz, E.M., Bradlow, E.T., and Fader, P.S. (2017), Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, 36(4), 500-522.
- [35] Senecal, S., Nantel, J. (2004), The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2), 159-169.
- [36] Shahrokhi T.S., and Ching, A.T. (2019), A Heuristic Approach to Explore: The Value of Perfect Information. *Johns Hopkins Carey Business School Research Paper*, (19-05).
- [37] Smith, W.K. and Tushman, M.L.. (2005), Managing strategic contradictions: A top management model for managing innovation streams. *Organization Science*, 16(5), 522-536.
- [38] Sorensen, A.T., and Rasmussen, S.J. (2004), Is any publicity good publicity? A Note on the Impact of Boor Reviews. *NBER Working paper*. Stanford University, (2004).
- [39] Stanaland, A.J.S., and Tan, J. (2010), The impact of surfer/seeker mode on the effectiveness of website characteristics, *International Journal of Advertising*, 29(4), 569-595.
- [40] Sutton, R.S., and Barto, A.G. (2018), *Reinforcement learning: An introduction*. MIT press.
- [41] Uotila, J. (2017), Exploration, exploitation, and variability: Competition for primacy revisited. *Strategic Organization*, 15(4), pp.461-480.
- [42] Urban, G.L., Liberali, G., MacDonald, E., Bordley, R., and Hauser, J.R. (2014), Morphing banner advertising. *Marketing Science*, 33(1), 27-46.
- [43] Zhang, Z., Ye, Q., Law, R., and Li, Y. (2010), The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4), 694-700.



Bumsoo Kim (bk4498@sogang.ac.kr)

Bumsoo Kim is an associate professor in the Sogang Business School at Sogang University, Seoul Korea. His research interests include Bayesian stochastic models, decision analysis, forecasting models and applications on topics related to online markets and digital environments.



Gun Jea Yu (gy52@hongik.ac.kr)

Dr. Gun Jea Yu has a Ph.D. in Industrial Labor and Relations from Cornell University and is now an Associate Professor at the College of Business Administration in Hongik University. His current research is focused on innovation and R&D management.



Joonkyum Lee (jkle@sogang.ac.kr)

Joonkyum Lee is a professor at Sogang Business School at Sogang University. His research focus is on operations management and innovation management. He received his Ph.D. from the Johnson School at Cornell University.

Optimal Exploration-Exploitation Strategies in Reinforcement Learning for Online Banner Advertising: The Impact of Word-of-Mouth Effects

Bumsso Kim*, Gun Jea Yu**, Joonkyum Lee***

ABSTRACT

One of the most important decisions for managers in the online banner advertising industry, is to choose the best banner alternative for exposure to customers. Since it is difficult to know the click probability of each banner alternative in advance, managers must experiment with multiple alternatives, estimate the click probability of each alternative based on customer clicks, and find the optimal alternative. In this reinforcement learning process, the main decision problem is to find the optimal balance between the level of exploitation strategy that utilizes the accumulated estimated click probability information and exploration strategy that tries new alternatives to find potentially better options. In this study we analyze the impact of word-of-mouth effects and the number of alternatives on the optimal exploration-exploitation strategies. More specifically, we focus on the word-of-mouth effect, where the click-through rate of the banner increases as customers promote the related product to those around them after clicking the exposed banner, and add it to the overall reinforcement learning process. We analyze our problem by employing the Multi-Armed Bandit model, and the analysis results show that the larger the word-of-mouth effect and the fewer the number of banner alternatives, the higher the optimal exploration level of advertising reinforcement learning. We find that as the probability of customers clicking on the banner increases due to the word-of-mouth effect, the value of the previously accumulated estimated click-through rate knowledge decreases, and therefore the value of exploring new alternatives increases. Additionally, when the number of advertising alternatives is small, a larger increase in the optimal exploration level was observed as the magnitude of the word-of-mouth effect increased. This study provides meaningful academic and managerial implications at a time when online word-of-mouth and its impact on society and business is becoming more important.

Keywords: On-line banner advertising, reinforcement learning, exploration-exploitation, word-of-mouth effect, Simulation

* First Author, Professor, Sogang Business School, Sogang University

** Co-Author, Professor, College of Business Administration, Hongik University

*** Corresponding author, Professor, Sogang Business School, Sogang University