

키워드검색광고 포트폴리오 구성을 위한 통계적 최적화 모델에 대한 실증분석

양홍규

연세대학교 기술정책협동과정
(yang.hongkyu@yonsei.ac.kr)

홍준석

경기대학교 경영정보학과
(juehong@kyunggi.ac.kr)

김우주

연세대학교 정보산업공학과
(wkim@yonsei.ac.kr)

본 논문은 키워드검색광고와 관련하여 의사결정자인 광고주의 입장에서 분석한 통계모델 기반 검색엔진최적화(Search Engine Optimization)논문이다. 일반적으로 키워드입찰은 노출순위를 대상으로 하는 입찰가액에 의해 이루어지고 있다. 그런데, 대부분 광고주는 수천 개 이상의 많은 키워드를 관리함에 있어, 매시간적으로 바뀌는 키워드별 입찰가액을 통해 입찰광고시스템을 관리하고 있는데, 사실상 시간과 인력자원측면에서 비효율적이다. 따라서, 본 논문에서는 기존의 입찰가액을 중심으로 하는 입찰시스템에 대해 의문점을 제기하고, 새로운 관점에서 노출순위를 의사결정변수로 하는 새로운 검색광고모델을 제정하여 제시하였다. 새로운 검색광고모델에 대한 최적화실증분석을 위해 예측모델과 최적화모델을 제시하였다. 연구과정은 우선 키워드의 특성에 따라 키워드그룹을 원천·제조브랜드·유통브랜드의 범주화기준을 제시한 후, PC와 모바일 매체별로 대표 키워드 선정한 후 노출순위와 클릭률이 비선형분포임을 보였고, 통계적 관계를 검토하였다. 클릭률예측 및 입찰가액예측을 위한 통계적 시나리오를 제시하였고, 적합성 분석을 통해 최적의 예측모델을 선정한 후, 선정된 예측모델을 기반으로 하여 클릭률과 기대이익(전환율)에 관한 최적화목적함수를 정의하고 실증분석을 진행하였다. 분석결과, 본 논문에서 제시한 검색광고모델은 클릭률 기반의 클릭수와 전환율 기반의 기대이익으로 표현되는 최적화모델 모두에서 개선효과가 있음을 확인하였다. 다만, 기대이익 최적화모델의 경우에는 핵심키워드임에도 불구하고 기대이익이 낮아 광고에서 배제되는 문제를 있음을 확인하고 대안을 제시했다. 마코브체인분석을 통해 핵심경유키워드 개념을 도입하였고, 최적화목적함수에 대해 핵심경유키워드의 기회이익을 반영한 최적화수정모델을 제시하여 적용가능성을 확인하였다. 본 논문은 키워드입찰시스템의 의사결정변수를 노출순위의 관점으로 전환하는 새로운 모델을 제안하였고, 키워드 범주별 및 노출순위 기반의 통계적 예측을 제시하고, 포트폴리오 구성에서의 최적화실증분석을 통해 노출순위 기반 예측모델의 유효성을 확인함과 동시에, 키워드간의 확산효과를 포함하는 수정모델제시 등 전략적인 입찰을 제안한 점에 시사점이 있다.

주제어 : 키워드검색광고, 키워드입찰, 클릭률, 확산효과, 최적화분석, 포트폴리오, SEO

논문접수일 : 2019년 5월 27일 논문수정일 : 2019년 6월 21일 게재확정일 : 2019년 6월 25일

원고유형 : 일반논문(급행) 교신저자 : 김우주

1. 서론

인터넷 사용환경의 개선 및 사용자수의 확대로 인해 온라인거래가 급격하게 확대되었고, 인터넷과 온라인거래의 증가는 새로운 광고의 수

단으로서 온라인광고시장의 확대를 가져왔다. 우리나라의 인터넷 이용자수(유무선포함)는 2017년 말 기준 4,528만명, 인터넷 이용률은 90.3% 수준을 기록하는 등 지속적인 성장세를 보이고 있고¹⁾, 온라인쇼핑몰 거래금액은 2017년 91.3조

원에 이어 2018년도는 110조원 초과 달성할 것으로 예상된다²⁾. 이렇듯 온라인거래가 확대되면서 자연스럽게 온라인검색광고시장은 2018년에는 2.7조원으로 전년대비 약14.2%성장하였고, 모바일검색광고시장은 1.7조원으로 전년대비 약 28% 급성장하였다³⁾.

키워드검색광고, 유료검색, 혹은 검색엔진광고(SSA, Sponsor's Search Advertising)는 고객이 쉽게 연상하거나 자주 사용하는 키워드를 활용하여 광고노출을 확대시키는 마케팅기법이다. 키워드검색과정을 사용자의 입장에서 살펴보면, 사용자는 네이버나 네이트 등의 검색엔진에서 관심 있는 키워드를 입력하고 검색하면, 검색엔진은 해당 키워드와 연관되어 있는 사이트를 보여준다. 이때 사용자는 출력되는 사이트에 노출되며, 클릭을 통해 관심 있는 사이트로 이동한다.

이번에는 검색과정을 광고주의 입장에서 살펴보자. 광고주는 사용자가 검색할 때의 가장 높은 위치에 광고할 사이트를 노출시키기 위하여 다른 광고주와 노출위치(Rank)를 놓고 서로 입찰을 하게 되는데, 가장 높은 금액을 제시한 광고주가 가장 높은 위치에 사이트가 노출되는 등 입찰가액에 따라 광고주의 사이트가 결정된다. 광고주의 노출은 검색엔진결과 페이지(SERP: Searching Engine Result Page)에 노출되는데, 사용자가 검색 후 SERP에 도달할 때 노출(Impression)이 집계된다. 사용자가 해당 SERP에서 광고주의 사이트를 클릭하게 되면 광고주는 클릭당 입찰가액(CPC)을 검색엔진에 광고비로 지급한다. 문제는 이러한 키워드입찰은 언제든 지 참여할 수 있는 시스템이며, 입찰가액을 수정

하는 즉시 입찰순위가 변동되는데, 사전에 입찰가액에 따른 노출순위를 알 수 없는 확률적인 관계가 존재하는 실시간개념의 광고시스템이다. 클릭수를 노출수로 나눈 값을 클릭률(CTR)이라고 하는데, 노출이 많은 키워드라도 광고주의 제품과 관련성이 없거나 클릭률 또는 기대이익이 낮으면 의미가 없다. 따라서 광고주는 사용자의 선호도가 높은 키워드를 추출하여, 새로운 키워드로 개발하며, 수 많은 키워드의 클릭률과 구매 전환율을 높이는 포트폴리오를 구축하고, 방문한 사용자를 관리하고, 자기 고유의 브랜드 인지도를 올리고자 노력하게 된다.

그런데, 검색엔진광고(SSA)은 매우 독특한 입찰시스템을 갖고 있다. 일반적인 입찰시스템은 최고가입찰(GFP, Generalized First Price)와 차하위입찰(GSP, Generalized Second Price)방식으로 구분되는데, 일반 입찰이 공개적인 입찰을 통해 입찰가액을 낙찰가로 지급하는 것인 반면, SSA 입찰시스템은 동일 대상에 대하여 차하위입찰가액을 낙찰가로 하는 방식이다. 결국, SSA 입찰시스템은 GSP방식의 비밀입찰시스템인데(Kamijo, 2013), 이는 노출순위별 입찰가액 뿐만 아니라 상대방의 제시하는 입찰가를 확인할 수 없는 입찰방식이다. 이렇듯, GSP과금방식 및 비밀입찰로 인한 노출순위 불확실성, 그리고 예산의 제약과 대량의 키워드관리 그리고 의사결정환경의 불확실성과 복잡성으로 인하여 유인부합성(incentive compatibility)이 낮아 전략적 접근이 필요하다(Abhishek and Kartik, 2013).

그런데, 광고주의 입장에서 볼 때, 일반적으로 수천개 이상의 수많은 키워드를 관리하게 되는

1) Korea Internet & Security Agency(KISA), <http://isis.kisa.or.kr/>

2) Statistics Korea, <http://kostat.go.kr/>

3) Report on Ads Expenditure by Korea Broadcast Advertising Corp (KOBACO)

데, 개별 입찰가를 일일이 판단하여 입찰에 참여하는 것은 시간과 자원소모가 많게 되는 문제가 있고, 또한 입찰가 역시 시간대에 따라 또는 고객의 수요변동에 등락이 계속되기 때문에 동일한 노출순위의 경우라도 입찰가가 계속 바뀌므로 입찰가와 노출순위간의 일관된 관계성을 가정하기가 어렵다. 따라서, 본 논문에서는 이러한 문제점이 예측모델과 최적화분석에서의 불확실성을 가져온다는 점을 지적하고, 효율적인 SSA 입찰시스템이 구현될 수 있도록 새로운 형태의 의사결정SSA모델을 제안하였다.

본 논문의 공헌내용은 다음과 같다. 첫째로 입찰가액이 아닌 노출순위를 독립변수(의사결정변수)로 채택한 SSA모델을 제시하였다. 기존 논문의 의사결정에서 접근하는 SSA모델에서는 입찰가액을 독립변수로 하는 모델을 가정하였지만, 실무적으로 입찰가격과 노출순위의 관계성이 명확하지 않은 반면, 노출순위와 클릭률의 관계성은 일관성이 있고, 시스템적으로도 간접적이긴 하지만 쉽게 구현가능한 수단이 있기 때문이다. 둘째로 개별키워드의 특성에 따라 키워드의 분류기준을 제시하였고 범주별 대표키워드를 선정하여 분석함으로써 키워드입찰을 범주 또는 그룹단위로 진행할 수 있는 이론적 배경을 제공하였다. 셋째로, 키워드 포트폴리오의 실증분석에 의한 유효한 통계적모델을 제시하였다. 통계적 모델시나리오를 이론적으로 구성하고, 실제 자료에 기반하여 가장 적합한 통계적모델을 제시하였으며, 선정된 예측모델에 기반하여 클릭률과 기대이익의 최적화 실증분석을 통해 개선효과가 있음을 보였다. 마지막으로, 클릭률 뿐만 아니라 광고주의 목적에 맞도록 기회이익을 반영한 전환율모델을 제시하면서, 키워드 확산효과와 클릭률의 유효성을 확인하였고, 이러한 키워드확산

효과를 반영하는 수정된 최적화모델을 제시함과 동시에, 키워드포트폴리오를 위한 광고전략에 포함할 것을 제안하였다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구문헌을 소개하고, 3장에서는 이론적배경이 되는 새로운 SSA모델을 재정의하고, 4장에서 연구대상자료요약과 연구분석방법을 설명하고, 5장에서 최적화분석에 활용할 수 있는 노출순위기반의 예측모델에 대한 이론적 배경과 가정 및 시나리오를 제시하며, 모델적합도검증을 통하여 최적예측모델을 선정하였다. 제6장에서는 선정된 예측모델에 기반하여 최적화분석을 실시한 결과를 보고하고, 기존 최적화모델의 문제점을 지적함과 동시에 수정된 최적화모델을 제시하고, 제7장에서는 결과의 의미와 연구의 한계 및 추가 논의사항을 기술하였다.

2. 연구문헌

2.1 키워드 범주화

키워드를 분류하는 방식에 대한 기존의 논문은 의미중심과 형식중심 접근으로 구분할 수 있다. 의미중심의 접근은 소비자가 정보를 검색하는 인지적인 부분이 포함되며, 형태중심의 접근은 키워드의 매칭정도와 키워드의 길이 등이 포함한다.

의미중심은 사용자의 키워드 검색목적에 따라 정보목적(informational), 탐색목적(navigational), 거래목적(transactional)으로 구분할 수 있는데, 약 80%의 검색이 정보목적이며, 각각 약10%정도가 나머지목적이다 (Jansen and Spink, 2007). 키워드는 상품이나 광고주에 대한 인지도에 따라 원천

제품(Generic), 소매브랜드(Retailer), 제조브랜드(Brand)의 3개 범주를 갖는데, 소매브랜드정보를 포함하는 키워드는 클릭률과 전환율의 증가와 관련이 있지만, 브랜드중심의 키워드에서는 클릭률과 전환율이 오히려 감소하였다 (Ghose and Yang, 2009). 광고주의 상품과 관련하여 키워드를 범주화할 수 있는데, 이 경우 광고주의 상품의 원천(generic-relevant)키워드, 광고주브랜드(focal-brand)키워드, 경쟁사브랜드(competing brand) 키워드의 3개 범주로 구분하여 분석하기도 하였다 (Du, Zhang and Zheng, 2017).

형태중심접근의 경우에는 키워드 매칭정도와 관련한 것으로, 키워드별로 일치(exact), 유사(broad), 문구포함(phase)등 3가지 형태로 키워드로 구분하였는데, 특히 의미중심의 범주별로 많은 키워드를 형태적으로 분류하여 입찰진행하는 것이 더 효율적이라고 분석하였고 (Du et al., 2017). 키워드의 길이와 관련하여 사용자가 키워드를 자세하게 입력하면 클릭률 또는 전환율이 영향을 받을 수 있다고 가정할 수 있는데, 연구결과 큰 영향은 없는 것으로 확인되었다 (Ghose and Yang, 2009).

키워드범주화는 의미중심과 형식중심의 결합하는 방식으로 진행되고 있다. 특히 인공지능의 결합으로 세만틱기반 검색이 확대되고 있어, 다양한 의미 또는 검색의 이해관계를 포함하는 방식의 검색이 확대되고 있는데, Gong, Abhisek and Li (2018)은 이를 키워드모호성(Keyword ambiguity)으로 설명한다. 그는 키워드모호성을 검색어관심의 관련성정도로 정의하고, 검색결과에 미치는 영향을 분석하였다. 분석결과에 의하면 상위에 노출된 키워드의 경우 키워드모호성이 높은 키워드에서 높은 클릭률을 보였는데, 소비자가 모호한 검색결과이후 유료검색으로 대체

하는 확산효과를 확인하였다.

키워드범주화에 있어 개별키워드가 과연 노출과 클릭률을 적절히 반영하고 있는 키워드인가에 대한 검토가 필요하다는 측면에서 의미와 형태를 포괄하는 키워드의 개발이 매우 중요하다. 이를 위해 키워드를 쇼핑의도를 포함하는 키워드를 자동적으로 추출하여 순위를 집계하는 알고리즘이 중요한데, 투자할 가치가 있는 키워드를 식별할 수 있는 이러한 알고리즘의 성능이 우수한 것으로 보고되고 있다 (Kim, Kim and Jung, 2014). 일반적으로 광고주가 관리하는 키워드는 소규모 쇼핑몰이라도 수천개에 이르는데, 특정 키워드를 포함하는 키워드내 사용자의 검색의도를 분석한 맞춤형 검색서비스로서 다단계 유사도 검색순위화 알고리즘등의 검색솔루션이 필요하다 (Kim and Kim, 2012)

광고주의 키워드선정과정은 사용자가 선호하는 키워드를 추출하는 과정과 키워드로 개발하는 과정, 키워드를 적용하는 과정으로 나눌 수 있다. 키워드추출은 SNS 또는 다양한 검색어를 인공지능기반의 알고리즘으로 접근하는 방식이므로 본 논문의 의도인 검색엔진최적화(SEO)접근과 다르고, 검색엔진자체에서 키워드별 검색량과 트렌드정보를 제공하므로 본 논문에서는 제외한다. 기존 논문에서는 키워드범주에 따라 클릭률등의 성과를 분석했지만, 키워드범주간의 효과와 범주별 확산효과를 최적화함수로 반영하지 못하였다. 따라서, 본 논문은 키워드의 개발과 적용측면에서 키워드의 범주화개념을 중심으로 포트폴리오구성함에 있어서, 형태보다는 의미중심 범주화를 중심으로 범주를 정의하고, 각 범주별 특성과 확산효과를 검토한다.

2.2 성과측정

키워드검색광고에 있어서 대표적인 성과측정 기준은 클릭률과 전환율이다. 클릭률(CTR)은 일정기간의 클릭수를 노출수로 나눈 값이고, 전환율(CVR)은 클릭수중에서 실제 구매활동이 이루어진 클릭수의 비율이다. 클릭률이 소비자와의 상호작용 가능성을 확보하기 위한 목적인 반면, 전환율은 판매자의 특정목적에 대한 달성가능성을 측정하는 목적이다 (Schlangenotto, Poniatowski and Kundisch, 2018), 전환율이 중요한 이유는 키워드의 손익과 직접적으로 연과 되어 있기 때문이다. 즉, 클릭률은 고객의 관심을 유도하기 위한 목적인데 반해, 전환율은 직접 구매가 이루어지는 비율을 의미하므로, 실제 손익에 영향을 미친다.

기존 논문에 의하면 많은 논문에서 클릭률과 노출순위의 관계성은 인정하지만, 전환율과 노출순위와의 관계성에 대해서는 다른 결과를 보이고 있다. Ghose and Yang (2009)은 노출순위가 상승함에 따라 클릭률과 전환율이 상승하지만, 기존의 클릭당 가치와 노출순위가 독립적이지 않다는 점을 확인하였다. 즉, 노출순위가 높다고 해서 키워드별 이익이 높은 것은 아니라는 것을 밝혔다. 키워드별로 키워드의 의미범주에 따라 다른 결과를 가져올 수 있다고 하면서, 과거의 클릭률보다는 현재의 입찰가액(CPC)이 더 중요하다고 강조하였다. Schlangenotto et al., (2018)는 노출순위와 클릭률간의 관계성은 확인하였지만, 노출순위와 전환율에 대해서는 일관성이 없음을 보였다. 이에 반해 광고주의 브랜드이미지에 따라 클릭률과 전환율이 영향을 받을 수 있다. 즉, 우량기업이 열등기업에 비해 상대적으로 낮은 순위에 입찰해도 높은 클릭률을 보이는 반면, 열

등기업은 소비자인지도가 낮으므로 높은 순위에 대한 인센티브가 강하며 이익을 확보할 수 있을 수 있다 (Jerath, Ma, Park, Srinivasan, 2011; Jeziorski, 2018). 클릭률간의 인과관계와 제품관여도 수준에 따른 클릭률의 관계에 따라 성과가 다르게 나타난다. 관련연구에 따르면, 조회수는 클릭률에 큰 영향을 주지 못했는데, 조회수와 클릭수 및 클릭률의 세가지 척도를 동시에 고려하는 것이 효과적이며, 광고주는 키워드의 광고목적을 명확히하여 광고전략을 세우는 것이 중요하다 (Jeon, 2017).

일반키워드와 유료키워드간의 관계 역시 중요한 부분이다. 사용자는 최초 일반검색키워드를 통해 검색하지만, 점차 유료브랜드로 이전하는 효과가 있음을 확인하였다 (Rutz and Bucklin, 2011; Abhishek, 2012; Gong et al., 2018). 다만, 이들 논문에서는 일반검색키워드와 유료키워드를 비교하였지만, 유료키워드간 비교는 이루어지지 않았다.

본 논문에서는 목표성가로 클릭률접근과 전환율접근에 기초하여 모델링하되, 기존 논문에서 언급한 기대이익 최적화방안의 문제점을 제시하고, 새로운 형태의 성과측정모델을 제시한 후 전략적 대안을 제시하였다.

2.3 통계적 예측모델

대부분 클릭률과 전환률에 대한 예측모델은 노출, 입찰가액, 순위, 클릭률, 전환률, 키워드길이 등을 각각 조합하여 독립변수로 구성하고, 이렇게 구성된 독립변수들을 선형 또는 비선형 혹은 이를 혼합한 다양한 통계적 예측모델을 가정한다.

우선 선형예측모델을 적용한 대표적인 문헌은

Rutz and Bucklin(2011)인데, 클릭률과 전환율에 대한 예측모델에서 노출, 전기클릭률, 순위, 인지도 등의 다양한 독립변수를 중심으로 하는 베이저안 접근선형모델을 구성하였다. Abhishek and Kartik (2013)의 경우는 키워드간 확산효과를 분석하기 위하여 클릭수예측모델의 경우에 선형모델을 가정하였다. Bae and Park (2018)은 매체특성과 광고특성별 광고효과를 확인하기 위하여 광고규격, 매체유형, 텍스트여부, 광고기간 등을 독립변수로 하여 클릭률을 예측하는 선형모델을 가정하여 분석하였다.

최근에는 머신러닝기법이 도입과 함께, SVM 기법(Support Vector Machines) 또는 LFM기법(Latent Factor Model)등이 클릭률예측에 도입되었다. 예를 들어, 광고주는 광고대상선정을 위해 사용자의 이벤트클릭예측모델이 필요한데, 특히, 클릭률회귀모델을 머신러닝의 분류모델 방식으로 적용하는 방식인 다중기준 선형계획회귀분석(Multiple Criteria Linear Programming Regression)은 광고대상간 중첩되는 부분과 차이가 있는 부분을 선형적인 거리측정방식으로 클릭률을 추정하는 모델이 제안되었다(Wang, Suphamitmongkol and Wang, 2013).

비선형모델은 클릭률 또는 순위를 예측함에 있어서 광범위하게 제시되었다. 이는 실제자료가 확률분포상 선형적이지 않은 부분도 있었지만, 클릭률이나 순위가 [0,1]의 단위구간값을 갖고 있어 선형보다는 비선형모델이 적합한 이유도 있다. Ghose and Yang (2009)는 클릭이 일어날 확률과 구매가 일어날 확률을 가정하고, 이항분포와 베이지안분포에 기초하여 클릭률과 전환율에 대한 확률함수를 정의한 예측모델을 구성하였다. 클릭률의 확률분포는 노출순위, 유통브랜드와 제조브랜드 포함유무, 키워드의 길

이, 시간 등을 포괄하는 로지스틱함수로 정의하였고, 전환율의 확률분포는 클릭률확률분포의 변수외에 품질지수를 포괄하는 로지스틱함수로 정의하였다. 입찰가격은 전환율확률분포에 사용된 전체변수의 지수함수로 예측모델을 정의하였다. Gopal, Li and Sankaranarayanan (2011)은 키워드광고를 유료매체와 기타매체로 구분하고, 각각의 상호관계를 분석하면서 매체간 카니발리즘의 존재를 밝혔는데, 특히, 클릭률, 순위, 노출, 입찰가액변수 각각이 독립변수와 종속변수가 되어 상호 영향을 주고받는 복합적인 형태의 비선형예측모델을 구성하였다. 클릭률예측과정에서는 순위와 노출, 직전 클릭률 및 상대매체의 노출 등을 독립변수로 하는 로지스틱함수를 제안하였고, 기타 변수는 단순회귀와 로그기반회귀분석을 제시하였다. Abhishek and Kartik (2013)는 노출순위예측모델에서 입찰가액을 독립변수로 하는 베르누이법칙에 따른 통계모델을 구성하였고, 이항분포의 확률로 Weibull분포를 가정하였다. 다만, 모델의 복잡성으로 인하여 회귀분석이나 최대우도법을 적용하기가 어렵다고 가정하고, 파라미터 추정에 있어 일반화적률추정법(GMM, Generalized Method of Moments)을 적용하였다. Du et al. (2017)은 클릭률과 입찰가액에 대한 예측과정에서 직전노출순위, 키워드길이, 범주별 키워드유무, 매칭정도, 이들 변수간 교차효과를 반영하였는데, 클릭률은 로지스틱함수로 정의하였고, 입찰가격은 지수함수로 정의하여 예측모델을 적용하였다. 앞선 연구에서 대부분의 비선형통계예측은 일반화선형회귀분석(Generalized Linear Regression)이나 베타회귀분석이 제안되고 있다.

지금까지 검토한 기존 논문의 통계예측모델은 클릭률 또는 입찰가액을 예측하기 위해 다양한

독립변수를 조합하는 방식으로 도입하여 예측의 타당성을 확인하는 논문이 대부분이다. 이에 비해, 광고주의 의사결정과정을 반영한 대표적인 논문은 다음과 같다. Cholette, Özlük and Parlar (2012)는 노출순위를 입찰가격요인과 비가격경쟁요인의 두가지 요인에 의한 베타분포모델로 정의한 후, 베타분포의 적률생성함수(MGF, Moment Generating Function)에 기반한 노출순위 예측모델을 제안하였다. Selçuk and Özlük (2013)는 Cholette et al. (2012)의 노출순위예측모델에서 출발하여, 제1차 적률생성함수값을 산출한 후, 여기에 자연클릭률을 곱하여 클릭률을 예측하는 모델을 제안하였다. 자연클릭률이란 노출 순위1위일때의 클릭률을 말한다.

본 논문에서는 의사결정변수에 대한 논의와 더불어 자연클릭률을 반영하는 통계예측모델의 실증분석 및 타당성을 병행하여 검토하였다.

2.4 최적화모델

키워드검색의 최적화모델은 목적과 제약조건에 따라 다양한 형태의 모델이 제시되어왔는데, 대부분 예산범위내에서 최적화를 추구하는 예산 최적화(Budget Optimization)문제이다. 최적화모델의 목적함수는 기본적으로 키워드전체의 클릭률을 극대화하는 최적화와 기대이익을 극대화하는 전환율최적화로 구분할 수 있다. 목적함수는 목적에 맞게 형태로 수정되고 있는데, 대표적으로 다음과 같다.

Abhishek and Kartik(2013)은 예산제약조건하에서의 수익을 최대화하는 모델을, Cholette et al. (2012)는 예산제약조건하에서 기대이익을 최대화하는 모델을 제안하였다. Selçuk and Özlük (2013)는 목표클릭률달성의 제약조건하에서 비

용을 최소화하는 모델은, DasGupta and Muthukrishnan (2010)는 최대한의 키워드를 활용하면서 예산의 효율성을 최대화하는 키워드를 찾는 확률모델을 제시하였고, Gopal et al. (2011)는 클릭수를 극대화하는 모델을 제시하면서, 매체별로 클릭수를 최대화하는 예산배분모델을 제시하였다. Feldman, Muthukrishnan, Pal and Stein (2006)은 경쟁적인 가격의 다중키워드(multi keyword) 최적화문제를 정보가 불확실한 상황에서의 단일입찰가전략(Uniform strategy)을 제시하면서 최적예산배분모델을 제시하였다.

일반적인 최적화목적함수 이외에, 실제 사용자가 클릭후 구매전환하는 과정의 클릭스트림데이터를 통해 광고주가 목적하는 목적함수를 구현할 수 있다. 클릭스트림데이터는 사용자가 클릭이후 사이트를 검색하면서 남긴 로그데이터로 유입시간, 체류시간, 페이지 기록, 구매전 검색자료 등의 전자적 활동기록을 의미한다. 이러한 클릭스트림 데이터를 기반으로 사용자의 탐색행동을 유형화하고 탐색패턴등을 분석하기 위한 순차패턴 마이닝기법등을 통해 구매전환개선등의 전략적 효과를 평가할 수 있다 (Choi and Nam, 2019).

본 논문에서는 클릭률과 기대이익(전환율)의 최적화모델을 기반으로 하되, 새롭게 클릭스트림데이터에 대한 마코브체인분석을 적용하여 최적화모델을 수정제안한다.

2.5 통계적추론

통계적추론방법은 크게 빈도주의적 접근법과 베이지안 접근법으로 나누어진다.

빈도주의적접근법은 대표적으로 최대우도법을 적용하는 추론방법인데, 종속변수가 일정범위내

에서 제한된 비율값으로 표시되는 경우에는 베타 회귀방법을 적절하며, 관측값으로부터 최적의 파라미터를 추출하는 최대우도법이 적용된 베타 회귀방법은 선형회귀보다 우월하다고 보고하였다 (Ferrari and Cribari-Neto, 2004; Smithson and Verkuilen, 2006; Paolino, 2001)

베이지안접근법으로 최대사후확률법이 대표적인데, 최대우도법이 절차적으로 단순하지만, 수렴되지 않는 가능성이 있고, 바이어스가 크며, 초기값의 문제로 최적해가 발견되지 않을 가능성이 있고, 경계값의 문제와 사전확률을 고려하지 않는 우도값의 일반화문제가 있으므로, 일정 범위구간의 제한된 비율값이 종속변수인 경우에는 베이지안 접근에 의한 베타분포가 적절하다고 강조하였다 (Liu and Eugenio, 2018).

본 논문에서는 모델시나리오를 테스트함에 있어서 양대 대표적인 통계적 추론방법을 사용하여 모델선정을 진행하였다.

3. 키워드검색광고(SSA)모델구축

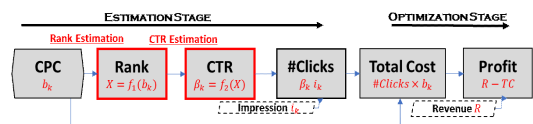
본 장에서는 최적화분석을 위해 가장 기본이 되는 프레임인 기존SSA모델의 문제점을 지적하고, 새로운 시각에서 키워드검색광고(SSA)모델을 재정의하였다.

3.1 기존 SSA모델 (SSA0모델)

기존 논문에서 제시하는 SSA모델(SSA0모델)은 입찰가액에서 시작된다 (<Figure 1> 참조). 광고주가 입찰에 참여하여 높은 노출순위를 확보하기 위해 입찰가액을 제시하면 검색엔진결과페이지(SERP)에서의 노출순위(Rank)가 결정되고,

노출순위에 따라 클릭률(CTR)이 종속변수로 예측된다. 클릭률은 노출순위의 종속변수로 예측된다고 가정한다. 클릭이 이루어지면 낙찰가액(CPC)이 광고비로 지급된다. 이렇게 볼 때 SSA0 모델은 입찰가액이 의사결정변수인 모델이고, 입찰대상은 노출순위이다. 참고로 입찰가액과 낙찰가액은 의미가 다르다. 입찰가액은 입찰시 제출한 가액이고, 낙찰가액(CPC)은 다른 광고주의의 키워드 입찰과정에서 차하위입찰가(GSP) 시스템에 의해 최종 클릭할 때 마다 광고비로 지급하는 금액이다. 내용상 구별의 의미가 없으면, 입찰가액으로 통일하여 사용한다.

SSA0모델은 크게 예측단계(Estimation Stage)와 최적화단계(Optimization Stage)로 구분된다. 예측단계는 입찰가액에 기반하여 노출순위를 추정하는 예측과정(Rank Estimation)과, 추정된 노출순위에 기반하여 클릭률을 예측하는 과정(CTR Estimation)을 포함하며, 최적화단계는 예상수익과 예산제약을 반영한 광고주의 특정 목적함수를 최대화하는 단계인데, 이 과정에서 최적 입찰가액(CPC)를 찾는 과정으로 구성된다.



<Figure 1> Classical SSA Model (SSA0 Model)

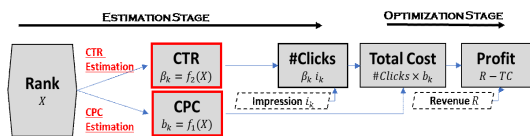
그런데, SSA0모델은 입찰가액(CPC)을 의사결정변수로 함에 있어 몇 가지 단점이 있다. 우선, 입찰가액(CPC)과 입찰순위(Rank)간의 관계성에 대한 의문이다. 시간대별로 또는 시기적으로 동일한 노출순위에 대해 입찰가액의 변동폭이 크다는 점은 과연 노출순위가 입찰가액의 함수라고 할 수 있는가의 의문이 있다. 둘째로, 광고주

들이 수많은 키워드를 관리하고 있는데, 다중 키워드(Multi Keywords)에 대해 개별 키워드별로 입찰가액을 일일이 감안하여 입찰에 참여하기는 실무적으로 사실상 불가능하다는 점이다. 셋째로, 입찰가액과 노출순위가 불확실성으로 인하여, 최종적인 클릭률의 예측이 과연 적절할 것인가에 대한 문제가 발생한다.

이와 별도로, 입찰가액을 독립변수로 정의하여 클릭률을 직접 추정하는 모델을 가정할 수 있다. 그러나, 이 경우 클릭률은 입찰대상이 될 수 없으므로 모델화하기 어렵다. 설사 적용한다 하더라도, 입찰가액과 노출순위를 독립변수로 하고 클릭률을 종속변수로 할 경우, 노출순위만 유의하게 인정되는 문제가 발생된다. 따라서 본 논문에서는 입찰가액으로 클릭률을 직접예측하는 시나리오는 제외한다.

3.2 새로운 SSA모델 구성(SSA1모델)

기존 SSA0모델의 문제점을 해결하기 위해 본 논문에서는 <Figure 2>에서와 같이 노출순위(Rank)를 의사결정변수로 하는 새로운 시각에서 SSA1모델을 제안한다.



<Figure 2> Novel SSA Model (SSA1 Model)

SSA1모델은 크게 예측단계와 최적화단계를 갖는다. 예측단계는 노출순위(Rank)에 기반하여 클릭률(CTR)과 입찰가액(CPC)를 추정하는 단계이고, 최적화단계는 예상수익과 예산제약을 반영하는 목적함수를 최대화하는 과정이다. 제안

한 SSA1모델을 살펴보면, 노출순위와 클릭률 관계는 대부분의 기존논문에서 관계성이 높다고 인정하고 있으며, 입찰가액과의 관계에 비해 일관성이 높아 예측력이 높은 모델의 구성이 가능하다. 특히, 클릭률예측과정에서 입찰가액을 배제하므로 클릭률추정이 독립적으로 이루어지는 장점이 있다. 반면에 노출순위와 입찰가액간의 관계는 노출순위 이외의 시기적 및 환경적 요인으로 인하여 일정한 관계성을 찾기가 쉽지 않다. 이를 위해 본 논문에서는 각각의 관계에 대한 가정을 전제로 예측시나리오를 구축하고 최적의 예측모델을 구성하였다.

3.3 예측모델선정절차

SSA0모델과 달리 SSA1모델은 노출순위(Rank)를 의사결정변수로 하는 새로운 접근방식이다. 따라서 본 논문에서는 최적화분석에 필요한 클릭률과 입찰가액의 예측모델을 선정함에 있어서 각각의 가정과 이론적 배경에 의한 시나리오를 제시하고, 모델선정의 기준이 되는 정보기준, 파라미터 예측기간 및 통계적 추론별로 모델적합성(Goodness of Fit, GOF)테스트를 통해 가장 적합한 예측모델을 선정한다. 이렇게 선정된 예측모델에 기반하여 궁극적으로 포트폴리오기반의 최적화목적함수를 구성하고, 제약조건에 따른 실증분석을 진행한다.

4. 연구대상자료

4.1 연구자료 및 범주화

실증분석을 진행할 연구대상자료는 최근 범정부차원에서 추진하고 있는 자동차중고부품 할

성화정책과 관련하여 국토부와 한국자동차해체 재활용협동조합이 공동으로 구축한 중고부품 쇼핑몰의 키워드 검색자료를 활용하였다. 분석대상 광고입찰시스템은 국내최대의 검색포탈서비스를 제공하는 네이버의 광고시스템의 파워링크를 중심으로 하였다.

〈Table 1〉 Keyword Category

Key-word Category	Key-word Korean	Features
Generic	<k1> 자동차 중고 부품	Keyword which best describe business objective. <k1> clarifies that the transaction items are “Used Parts for Car”.
Abstract	<k2> 자동차 부품	<k2>meaning is one step broader than <k1> <k2> clarifies that the transaction targets are “Parts for car”. <k2> is not clear whether the parts are “Used” or “New” one.
	<k3> 중고 부품	<k2>meaning is one step broader than <k1> <k3> clarifies that the transaction targets are “Used Parts”. <k3> is not clear whether the parts are for “Car” or “Other machine” or “Electric Appliances”, etc.
Retailer Brand	<k4> 지파크	Sponsor’s unique retailer brand <k4> is the unique name of Internet shopping mall as transaction platform

연구대상 키워드 선정은 사업목적을 정확하게 표현하는 정도에 따라 <Table 1>에서와 같이 (사업)원천키워드, 추상적 키워드, 유통키워드의 세 가지의 키워드범주를 정의하고, 각각 대표키워드를 선정하였다. 의미기반의 범주만을 고려했고 정확하게 매칭되는 키워드만 고려하였다. 예를 들어, ‘자동차중고부품<k1>’키워드에 대해 ‘자동차중고부품판매점’ 또는 ‘자동차중고부품

사이트’ 등의 확장문구를 포함하는 키워드는 전혀 고려하지 않고, 정확하게 동일한 키워드만을 대상으로 한다. 확장문구가 포함되는 키워드는 확장된 의미기반을 제공하지만, 수요가 낮아 노출이 낮다는 단점이 있다.

분석대상기간은 검색키워드별로 2016.7.1.~2018.12.31.까지 30개월간의 주간 평균자료를 PC 키워드와 모바일키워드로 구분하여 네이버광고시스템의 입찰통계시스템과 로그통계에서 각각 추출하였다. 추출된 자료는 주간 노출, 평균 입찰가액(CPC), 클릭수, 평균클릭률(CTR), 평균노출순위(Rank), 광고비용, 평균전환율(CVR)에 대한 통계자료이다(<Table 2> 요약자료 참고).

〈Table 2〉 Dataset Summary

PC Keyword

Category	Key word	Cases	Weekly Mean				Weekly SD	
			Impres-sion	#Clicks	Rank	CTR	Rank	CTR
Generic	<k1>	128	1,159	296	3.93	25.51%	1.29	2.79%
Abstract	<k2>	129	792	25	4.41	3.16%	2.20	1.60%
	<k3>	129	257	59	1.63	23.00%	0.59	4.72%
Retailer	<k4>	129	1,192	494	1.29	41.43%	0.46	11.81%

Mobile Keyword

Category	Key word	Cases	Weekly Mean				Weekly SD	
			Impres-sion	#Clicks	Rank	CTR	Rank	CTR
Generic	<k1>	129	2,036	412	3.23	20.23%	0.54	2.48%
Abstract	<k2>	126	703	40	3.07	5.72%	1.38	3.41%
	<k3>	130	344	77	1.66	22.50%	0.56	3.82%
Retailer	<k4>	130	1,683	479	1.25	28.46%	0.44	20.69%

유효키워드의 노출은 네이버의 경우 ‘파워링크’라는 검색엔진결과페이지(SERP)에서 이루어

진다. 네이버에서 노출되는 PC키워드의 경우 SERP상에서 가장 위쪽이 1위이고 가장 아래가 10위이고, 모바일키워드의 경우는 가장 아래쪽의 키워드가 3위 내지 5위이다. 분석목적상 비교 가능성과 분석을 위해 상대비율로 표시하는데, 가장 위쪽이 가장 높은 점수를 갖도록 하기 위해, 가장 위쪽을 1.0, 가장 아래쪽을 0.0점으로 척도를 변화하고, 이를 ‘Rank점수(Rank Score)’ 변수로 정의하였다.⁴⁾ 또한 베타회귀분석은 종속변수가 (0,1) 구간범위의 비율값이어야 하는데, Rank점수와 클릭률값이 [0,1]구간값이므로 경계값문제가 발생한다. 따라서 본 논문에서는 $((y(n-1)+0.5))/n$ 식을 적용하여 (0,1)범위의 단위함수값으로 변환하였다. (Smithson and Verkuilen, 2006).

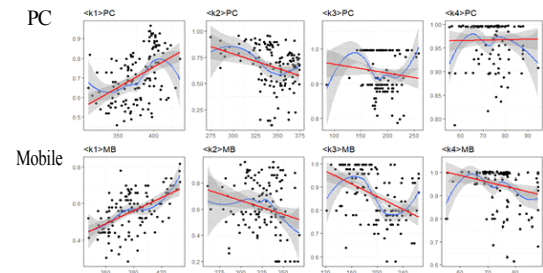
4.2 변수관 관계분석

4.2.1 입찰가액(CPC)와 Rank점수(Rank Score)

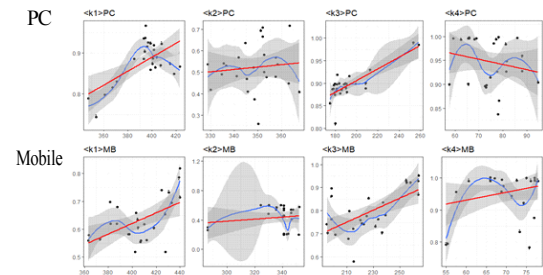
입찰가액과 Rank점수의 관계는 키워드별 특성과 고객 선호 등 입찰가격 이외의 요소에 영향을 많이 받는 것으로 보인다. 입찰가액과 Rank점수의 관계를 최근 30개월<Figure 3> 및 6개월<Figure 4>의 대상기간별로 보면 시간에 따른 변화를 볼 수 있다.

자동차중고부품<k1>의 경우에는 소비자 선호도가 높은 원천키워드로 많은 검색과 노출이 발생하는 키워드이다. 따라서 여전히 입찰경쟁이 높은 편으로 상관계수가 정(+)의 관계를 갖는다. 지파츠<k4>는 브랜드키워드인데, 광고주의 독특한 브랜드 이미지와 관련되어 있다. 일반적으로 초기 기업의 경우 브랜드이미지가 낮으므로 소

비자의 인지도가 낮으므로 키워드경쟁이 낮다. 광고주의 입장에서는 이미지확대 차원에서 입찰 참여를 하지만, 가격이 높아질수록 다른 광고주들이 자신의 브랜드를 홍보하게 되는 것이므로, 유통브랜드 키워드의 경우에는 굳이 높은 순위의 노출위치를 확보하기 위해 높은 입찰가를 제시할 필요성이 낮다. 노출순위는 높고 입찰가액은 낮게 되므로, 무리해서 입찰할 필요가 적어 상관관계가 낮다. 최근 들면서 모바일키워드에 대한 수요가 높아진 것으로 보인다. 추상적인 키워드의 경우에는 장기적으로 Rank점수가 낮아지고 있어 키워드 선호도가 점차 감소하고 있음을 볼 수 있다.



<Figure 3> Plot: Rank Score(dv) on CPC(iv): 30 months



<Figure 4> Plot: Rank Score(dv) on CPC(iv): 6 months

4) 네이버검색광고 파워링크의 PC키워드 노출수는 10개이고, 모바일키워드는 3~5개이므로, 각각의 Rank점수는 $[1-(\text{Rank점수}/10)]$ 와 $[1-(\text{Rank}/5)]$ 로 변환하여 ‘Rank점수’ 변수를 도출하였다.

<Table 3> Linear Regression:

Rank Score(dv) on CPC(iv) of 30 month period

Predictors	<k1>PC		<k2>PC		<k3>PC		<k4>PC	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	-0.0724	0.608	1.6151	<0.001	0.9814	<0.001	0.9601	<0.001
CPC	0.0021	<0.001	-0.0028	0.001	-0.0003	0.196	0.0001	0.961
Observations	128		129		129		129	
R ² / R ² adjusted	0.196 / 0.190		0.077 / 0.070		0.013 / 0.005		0.000 / -0.008	

Predictors	<k1>MB		<k2>MB		<k3>MB		<k4>MB	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	-0.372	0.006	1.292	<0.001	1.1326	<0.001	1.161	<0.001
CPC	0.0024	<0.001	-0.0021	0.016	-0.0014	<0.001	-0.0029	0.009
Observations	129		126		130		130	
R ² / R ² adjusted	0.277 / 0.271		0.046 / 0.039		0.183 / 0.176		0.052 / 0.044	

<Table 4> Linear Regression:

Rank Score(dv) on CPC(iv) of 6 month period

Predictors	<k1>PC		<k2>PC		<k3>PC		<k4>PC	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	0.2078	0.235	0.1356	0.858	0.6457	<0.001	1.0284	<0.001
CPC	0.0017	0.001	0.0011	0.612	0.0013	<0.001	-0.0011	0.321
Observations	26		26		26		26	
R ² / R ² adjusted	0.393 / 0.368		0.011 / -0.030		0.642 / 0.627		0.041 / 0.001	

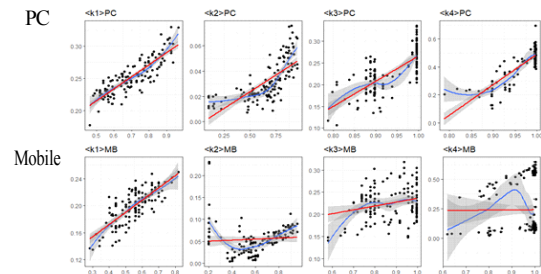
Predictors	<k1>MB		<k2>MB		<k3>MB		<k4>MB	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	-0.1387	0.578	-0.0387	0.955	0.2064	0.195	0.7849	<0.001
CPC	0.0019	0.004	0.0014	0.494	0.0026	0.001	0.0024	0.237
Observations	26		24		26		27	
R ² / R ² adjusted	0.291 / 0.262		0.022 / -0.023		0.377 / 0.351		0.055 / 0.018	

입찰가액과 Rank점수간 단순선형회귀분석결과를 보면 전체적으로 입찰가액의 영향력이 매우 낮거나 상관계수가 매우 낮은 것을 볼 수 있다 (<Table 3>, <Table 4> 참조). 6개월 기간의 회귀분석자료에서도 볼 때, 전체적으로 상관계수가 증가되었지만 상관계수가 여전히 낮다. 결국, 입찰가액이 노출순위를 예측하는 독립변수로서의 역할에 대한 의문이 제기되는 부분이다.

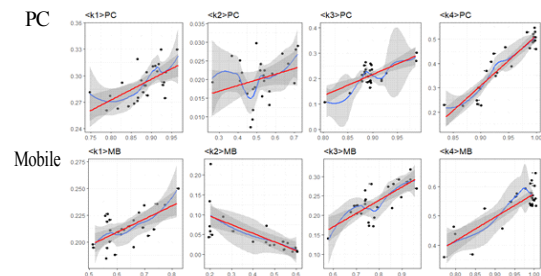
4.2.2 클릭률(CTR)과 Rank점수(Rank Score)

Rank점수와 클릭률간의 그래프를 PC키워드와 모바일키워드로 구분하여 최근 30개월<Figure 5> 및 6개월<Figure 6>의 기간별로 표시하였는다.

클릭률과 Rank점수간의 그래프를 보면, 전체적으로 정(+)의 관계를 보이고 있다. 다만, 자동차부품<k2>모바일키워드는 6개월 그래프를 보면 부(-)의 관계로 바뀌었는데, 이는 노출이 급감하고 클릭률이 매우 낮아 불확실성이 높은 상황에서 광고주가 노출순위를 점차 낮추었기에 나타나는 현상으로 분석되었다. 30개월 회귀분석에서는 상관계수가 높았지만, 6개월 회귀분석에서는 더욱 이런 현상이 나타난 것을 알 수 있다. 지프츠<k4>모바일키워드는 노출과 클릭률이 급격하게 증가하면서 관계패턴이 바뀌고 있다. 30개월의 그래프에서는 회귀선 위쪽과 아래쪽으로 2개의 다른 관계의 모집단이 분리되어 보이는데, 최근 6개월의 그래프에서 보면 점차 안정화 단계에 들어선 것으로 보인다. 이러한 영향으로 회귀분석의 상관계수역시 점차 증가된 것을 확인할 수 있다 (<Table 5>, <Table 6> 참조).



<Figure 5> CTR(dv) on Rank Score(iv): 30 months



<Figure 6> CTR(dv) on Rank Score(iv): 6 months

(Table 5) Linear Regression:

CTR(dv) on Rank Score(dv) of 30 month period

Predictors	<k1>PC		<k2>PC		<k3>PC		<k4>PC	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	0.1227	<0.001	-0.0026	0.408	-0.3132	<0.001	-1.6899	<0.001
rRanking1	0.1861	<0.001	0.0527	<0.001	0.5791	<0.001	2.1817	<0.001
Observations	128		129		129		129	
R ² / R ² adjusted	0.730 / 0.727		0.513 / 0.509		0.524 / 0.521		0.715 / 0.713	

Predictors	<k1>MB		<k2>MB		<k3>MB		<k4>MB	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	0.0997	<0.001	0.0475	<0.001	0.1491	<0.001	0.2236	0.26
rRanking1	0.1813	<0.001	0.0119	0.409	0.0868	0.004	0.0166	0.937
Observations	129		126		130		130	
R ² / R ² adjusted	0.604 / 0.601		0.006 / -0.003		0.063 / 0.056		0.000 / -0.008	

(Table 6) Linear Regression:

CTR(dv) on Rank Score(dv) of 6 month period

Predictors	<k1>PC		<k2>PC		<k3>PC		<k4>PC	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	0.0827	0.146	0.0124	0.039	-0.5434	0.003	-1.5489	<0.001
rRanking1	0.2371	0.001	0.0151	0.17	0.8435	<0.001	2.0642	<0.001
Observations	26		26		26		26	
R ² / R ² adjusted	0.375 / 0.349		0.077 / 0.039		0.479 / 0.457		0.864 / 0.858	

Predictors	<k1>MB		<k2>MB		<k3>MB		<k4>MB	
	Estimat	p	Estimat	p	Estimat	p	Estimat	p
(Intercept)	0.1348	<0.001	0.1397	<0.001	-0.0333	0.525	-0.2273	0.035
rRanking1	0.124	0.001	-0.2164	<0.001	0.3409	<0.001	0.8038	<0.001
Observations	26		24		26		27	
R ² / R ² adjusted	0.386 / 0.360		0.493 / 0.470		0.537 / 0.518		0.696 / 0.684	

4.2.3 평가기준 및 분석도구

클릭률과 입찰가액의 예측시나리오에 대한 모델적합성평가(GOF, Goodness of Fit)기준은 일반적으로 모델평가에 적용되는 평균제곱근오차(RMSE)과 AIC(Akaike Information Criterion)로 정하였다. 클릭률예측시나리오의 경우에는 독립변수가 Rank점수로 동일하므로 RMSE에 의해 평가하며, 입찰가액예측 시나리오의 경우는 모델별로 변수의 개수가 다르므로 RMSE와 AIC를 기준을 동시에 고려한다.

통계패키지는 R을 사용하였고, 최적화모델은 엑셀의 Evolutionary Solver를 활용하였다.

5. 실증분석을 통한 예측모델선정

본장에서는 SSA₁모델에서의 최적화분석에 활용 통계적 예측모델시나리오를 제시하고, 모형적합성검증을 거쳐 최적예측모델을 선정한다. 클릭률예측모델과 CPC예측모델 각각에 대한 구축과정을 설명한다.

5.1 클릭률예측모델 선정

5.1.1 이론적 배경

클릭률은 클릭수를 노출수로 나눈 값이므로 [0,1]의 단위구간값을 갖는 확률변수이다. 클릭률이 높다는 것은 사용자의 클릭이 높은 것을 의미하므로, 키워별로 선호도와 관계가 높다고 볼 수 있다. 따라서 광고주는 선호도가 높은 키워드에 대해 노출순위를 높이려는 경쟁이 높다고 볼 수 있다. 결국, Rank점수가 높을 수록 클릭률이 증가한다고 가정할 수 있는 근거가 된다. 이론적으로 보면, 클릭률의 최대값인 1.0근처에서는 Rank 점수가 아무리 높아도 클릭률의 증가폭은 감소되어 최대값은 1.0을 초과할 수 없다. 이러한 관계는 Rank점수와 클릭률관계가 S자곡선(Sigmoid Curve)의 특성을 갖는다는 것을 의미하는데, 이를 함수로 표현하면 <Equation 1>의 로지스틱함수(Logistic Function)이다. 참고로, $Z\beta$ 는 Rank 점수를 로짓함수로 표현할 때의 벡터수식이며, S자곡선(Sigmoid Curve) 계수의 벡터표시이다.

$$E[\beta_k] = logistic(Z) = \frac{1}{1 + \exp(-Z\beta)} \quad \text{(Equation 1)}$$

제한된 범위내에서의 비선형 비율분포는 베타 분포로 설명하는 것이 가장 적절하다 (Ferrari and

Cribari-Neto, 2004). 따라서 클릭률의 경우 [0,1]의 제한적 범위내의 비율값이므로 경계값을 변환하면 베타분포를 가정하여 설명하기가 용이해진다. 앞서 클릭률의 경우 Rank점수와 S자특성의 관계성을 갖고 있으므로, 확률분포의 평균이 S자곡선인 일정한 함수형태의 분포를 갖는다고 가정할 수 있는데 이를 수식으로 표현하면 <Figure 2>와 같다.

$$\beta_i \sim B(\mu_i, \phi)$$

$$\text{logit}(\mu_i) = Z\beta \Leftrightarrow \mu_i = \frac{1}{1 + \exp(-Z\beta)}$$

(Equation 2)

따라서, 본 논문에서는 클릭률이 베타분포를 따른다고 가정하고, Rank점수와 클릭률간의 S자곡선특성을 반영한 logit함수를 링크함수로 사용하는 베타회귀분석을 통해 클릭률예측 시나리오를 구성하였다.

5.1.2 클릭률예측모델 시나리오

본 논문에서는 노출순위(Rank점수)에 기반한 클릭률예측모델시나리오를 제시하였는데, 동 예측모델의 적합성을 기존문헌에 제시한 ‘CPC를 독립변수로 하는 시나리오’와 비교 검토하기 위해 <Table 7>의 시나리오를 비교테스트 한다.

<Table 7> CTR Estimation Scenario

SN	Model Scenario	Decision Variable	Description
S1	Natural CTR model	CPC	$CPC \xrightarrow{reg} \widehat{Rank}$ $\widehat{Rank} \xrightarrow{\text{natural CTR}} \widehat{CTR}$
S2	2-Step CPC Model	CPC	$CPC \xrightarrow{reg} \widehat{Rank} \xrightarrow{reg} \widehat{CTR}$
S3	Rank Model	Rank점수	$\widehat{Rank} \xrightarrow{reg} \widehat{CTR}$

<시나리오1>은 입찰가액을 기반으로 클릭률을 예측하는 시나리오이다. 이 모델은 Selçuk and Özlük (2013)이 제안한 클릭률예측모델을 검증하기 위해 시나리오에 포함하였다. 예측방식은 베타회귀분석에 의한 Rank점수를 예측한 후, 자연클릭률을 적용하여 클릭률을 예측하는 시나리오이다. 참고로 자연클릭률은 1순위일때의 클릭률을 의미한다. <시나리오2>는 입찰가액을 독립변수로 하여 Rank점수를 추정한 후, 추정된 Rank점수를 기반으로 클릭률을 추정하는 시나리오이다. <시나리오3>은 Rank점수를 독립변수로 하여 직접 클릭률을 예측하는 모델이다. 참고로 <시나리오3 (Rank모델)>은 최적화분석을 위해 클릭률예측과 동시에 CPC예측이 필요하다.

5.1.3 최적클릭률예측모델 선정

클릭률예측시나리오의 모델적합도검증은 파라미터를 추정하기 위한 관측치의 대상기간과 통계적추론 방법에 따라 적합도가 각기 달라질 수 있다. 따라서 각각의 시나리오에 대하여 파라미터의 추정기간을 3개의 기간으로 나누어서 각각 30개월(2016.7~2018.12), 12개월(2018.1~2018.12) 및 6개월(2018.7~2018.12)의 기간별로 진행하고, 통계적추론방법은 최대우도법(MLE)과 최대사후확률법(MAP)에 따라 모델적합도(GOF, Goodness of Fit)검증하였다. 정보기준은 RMSE로 하였고, 최적모델은 가장 낮은 RMSE를 갖는 모델로 선정한다.

적합성검증 결과, 노출순위(Rank점수)를 독립변인으로 하는 <시나리오3>이 입찰가액을 독립변인으로 하는 <시나리오1,2>보다 RMSE가 낮았다. 또한 분석대상기간이 짧을 수록 우수한 적합도를 보였는데, 이는 키워드에 대한 소비자의

(Table 8) GOF Test Results:

-RMSE Scores of CTR Estimation Scenarios-

(a) Parameter Estimation period 2016.7-2018.12 (30mo)

Key word	PM	Model1		Model2		Model3	
		MLE	MAP	MLE	MAP	MLE	MAP
<k1>	PC	0.078%	0.077%	0.066%	0.067%	0.020%*	0.022%
	MB	0.053%	0.053%	0.047%	0.049%	0.025%*	0.027%
<k2>	PC	0.022%	0.022%	0.021%	0.026%	0.010%*	0.015%
	MB	0.117%	0.117%	0.117%	0.121%	0.117%*	0.122%
<k3>	PC	0.221%	0.221%	0.221%	0.224%	0.104%*	0.106%
	MB	0.164%	0.164%	0.145%	0.146%	0.136%*	0.138%
<k4>	PC	1.386%	1.387%	1.382%	1.382%	0.347%*	0.348%
	MB	4.230%	4.229%	4.120%*	4.120%*	4.257%	4.262%

(b) Parameter Estimation period 2018.1-2018.12 (12mo)

Key word	PM	Model1		Model2		Model3	
		MLE	MAP	MLE	MAP	MLE	MAP
<k1>	PC	0.042%	0.044%	0.042%	0.046%	0.026%*	0.032%
	MB	0.029%	0.029%	0.020%	0.029%	0.015%*	0.026%
<k2>	PC	0.004%	0.004%	0.004%	0.034%	0.003%*	0.032%
	MB	0.233%	0.233%	0.227%	0.263%	0.135%*	0.173%
<k3>	PC	0.138%	0.139%	0.136%	0.147%	0.089%*	0.101%
	MB	0.155%	0.155%	0.153%	0.162%	0.076%*	0.086%
<k4>	PC	1.306%	1.310%	1.303%	1.306%	0.146%*	0.151%
	MB	2.829%	2.832%	2.581%*	2.585%	2.899%	2.895%

(c) Parameter Estimation period 2018.7-2018.12 (6mo)

Key word	PM	Model1		Model2		Model3	
		MLE	MAP	MLE	MAP	MLE	MAP
<k1>	PC	0.035%	0.042%	0.034%	0.062%	0.024%*	0.055%
	MB	0.046%	0.044%	0.025%	0.065%	0.015%*	0.053%
<k2>	PC	0.003%	0.003%	0.003%	0.111%	0.003%*	0.106%
	MB	0.235%	0.235%	0.231%	0.335%	0.120%*	0.236%
<k3>	PC	0.147%	0.150%	0.135%	0.184%	0.091%*	0.121%
	MB	0.165%	0.167%	0.163%	0.195%	0.091%*	0.111%
<k4>	PC	1.168%	1.181%	1.167%	1.178%	0.160%*	0.166%
	MB	0.464%	0.489%	0.449%	0.450%	0.138%*	0.142%

* Minimum RMSE

선호도 및 입찰경쟁이 지속적으로 변화하고 있고 가장 최근의 자료가 소비자의 선호도를 가장 잘 설명함을 의미한다. 또한 최대우도법(MLE)접

근이 기존모델을 더 잘 설명하였다. (<Table 8> 참조).

따라서, 본 논문에서는 관측자료 대상기간 6개월의 최대우도법(MLE)법에 따른 Rank기반 <시나리오3> 을 채택하여 최적화 실증분석을 진행한다.

5.2 입찰가액(CPC)예측모델 선정

5.2.1 이론적 배경

Rank점수와 입찰가액의 관계는 한마디로 정의하기 어렵다. 이는 키워드입찰의 경쟁시간대와 키워드의 특성, 고객의 선호 등 입찰가액 외적인 요소가 많은 영향을 주어 변동성이 크기 때문이다. 따라서 입찰가액 예측시나리오는 일정 기간 평균개념에서의 분석하되, 키워드가 갖을 수 있는 기본적인 가정을 먼저 검토하였다.

일반적으로 선호도가 높은 키워드의 경우에는 노출순위에 대한 경쟁이 높으므로 입찰가액이 높아진다. 즉, Rank점수가 높을 수록 입찰가액이 증가한다고 가정할 수 있다. 이론적으로 볼 때 Rank점수의 최대값인 1.0근처에서는 입찰가액은 체증적으로 증가하게 된다. 따라서 선호도가 높은 키워드에서는 이론적으로 독립변수 Rank점수에 대하여 종속변수 입찰가액이 ‘역S자곡선 (Inverse Sigmoid Curve)’의 관계성을 갖는다고 가정할 수 있다. 이를 함수식으로 표현하면 <Equation 3>과 같이 logit함수가 된다.

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \text{logistic}^{-1}(p)$$

<Equation 3>

이와 반대되는 가정이 존재한다. 입찰가액이 상승함에 따른 심리적인 이유로 상한가격을 형

성할 수 있다. 이러한 경우에는 S자곡선(Sigmoid Curve)의 성격을 갖는다고 가정할 수 있다. 이 경우 함수식은 <Equation 2>에서와 같다. 이 외에도 노출순위에 대해 대체키워드를 개발하거나, 입찰가격 이외의 경쟁요소인 카페나 블로그 활성화로 우회하거나, 모바일등의 키워드로 선회하거나 사용자의 선호도가 낮아지거나 또는 키워드에 대한 의미변화 등으로 키워드별로 다른 패턴을 가정할 수 있다.

5.2.2 입찰가액(CPC)예측모델 시나리오

본 논문에서는 다음과 같이 Rank점수를 기반으로 하여, 선형회귀분석외에 비선형회귀분석의 가정을 감안하여 <Table 9>의 시나리오를 구성하고, 적합성검증을 통해 최적의 모델을 선정하기로 한다.

<Table 9> CPC Estimation Scenarios

SN	Regression Scenario	Scenario Formula
S1	Linear (LR)	$\widehat{CPC} = \beta_0 + \beta_1 Z$
S2	Polynomial (PR2)	$\widehat{CPC} = \beta_0 + \beta_1 Z + \beta_2 Z^2$
S3	Polynomial (PR3)	$\widehat{CPC} = \beta_0 + \beta_1 Z + \beta_2 Z^2 + \beta_3 Z^3$
S4	Inverse Sigmoid (ISR)	$\widehat{CPC} = \beta_0 + \beta_1 \ln(Z/(1-Z))$
S5	Beta (BR)	$\text{logit}(\widehat{CPC}) = \beta_0 + \beta_1 Z$

<시나리오1>은 단순하면서도 기준을 제시하는 전형적인 선형성을 가정하였고, <시나리오2>와 <시나리오3>은 확장된 선형개념으로 각각 2차 및 3차 비선형의 다항회귀분석(Polynomial Regression)을 가정한다. 2차함수는 단조함수이지만 3차함수는 증감을 반영한다고 가정한다. 외

부 환경적인 원인을 반영한다고 볼 때 적절할 수 있는 시나리오이다. <시나리오4>는 선호도가 높은 키워드의 특성인 ‘역S자곡선(Inverse-Sigmoid)’ 특성을 갖는다고 가정한다. 선호도가 높은 키워드는 Rank점수가 높을수록 입찰가액이 체증적으로 커지게 된다고 가정할 수 있기 때문이다. <시나리오5>는 심리적 한계가 적용될 수 있다는 가정하에 S자곡선(Sigmoid)을 가정하였다. 키워드 특성에 따라 노출순위가 높아지더라도 심리적인 상한선이 있으므로 로지스틱함수 특성에 따라 입찰가액의 상한을 유지한다고 가정한다.

5.2.3 최적의 입찰가액(CPC)예측모델 선정

입찰가액(CPC)예측시나리오의 모델적합도검증은 파라미터를 추정하기 위한 관측치의 대상기간에 따라 적합도가 각기 달라질 수 있다. 따라서 각각의 시나리오에 대하여 파라미터의 추정기간을 3개의 기간구분으로 나누어 30개월(2016.7~2018.12), 12개월(2018.1~2018.12) 및 6개월(2018.7~2018.12)의 기간별로 진행하였고, 선택을 위한 정보기준은 RMSE와 AIC를 병행하였는데, 최적모델은 가장 낮은 RMSE 또는 AIC를 갖는 모델을 선정한다.

입찰가액(CPC)예측모델의 적합성검증 결과는 <Table 10>과 같이 RMSE기준으로 평가할 경우 3차 다항회귀분석(PR3)에서 RMSE가 가장 낮았다. 그러나 파라미터 변수 개수의 증가에 따른 과적합문제가 발생하는데, 이를 반영한 AIC점수에 의한 평가결과는 혼조세를 보였다. 전체적으로 분석대상기간이 짧을 수록 RMSE와 AIC점수가 낮아지면서 더 우수한 적합도를 보이는 경향이 있다.

적합성 결과에 따라, 본 논문에서는 관측자

료 추정기간을 6개월로 정하고, RMSE가 가장 낮은 다항회귀분석(PR3)시나리오를 기본모델로 정하되, 비교 검토를 위하여 AIC점수에서 상대적으로 우위를 보인 역S자곡선회귀분석(ISR)시나리오를 채택하여 각각 최적화 실증분석을 진행한다.

〈Table 10〉 GOF Test Result:

-RMSE & AIC Scores of CPC Estimation Scenarios-

Parameter Estimation period 2016.9-2018.12 (24mo)

SN	Key word	RMSE					AIC				
		LM	PR2	PR3	IS	BR	LM	PR2	PR3	IS	BR
1	<k1>PC	5.55%	5.47%	5.47%*	5.52%	5.58%	-300	-301	-299	-301	-305*
2	<k1>MB	4.36%	4.29%	4.27%*	4.33%	4.46%	-351	-352*	-351	-352	-341
3	<k2>PC	4.39%	4.13%	4.10%*	4.37%	4.39%	-349	-360*	-359	-350	-357
4	<k2>MB	4.48%	4.31%	4.29%*	4.49%	4.48%	-331	-337*	-336	-331	-336
5	<k3>PC	5.80%	5.79%	5.57%*	5.75%	5.80%	-291	-289	-295*	-293	-289
6	<k3>MB	7.27%	7.05%	6.79%*	6.90%	7.28%	-244	-249	-254	-255*	-245
7	<k4>PC	1.65%	1.61%	1.61%*	1.66%	1.65%	-553	-555*	-554	-552	-551
8	<k4>MB	1.57%	1.55%	1.54%*	1.56%	1.57%	-562	-564	-563	-565*	-562

Parameter Estimation period 2018.1-2018.12 (12mo)

SN	Key word	RMSE					AIC				
		LM	PR2	PR3	IS	BR	LM	PR2	PR3	IS	BR
1	<k1>PC	3.52%	3.49%	3.25%*	3.65%	3.51%	-195	-193	-199*	-191	-198
2	<k1>MB	4.20%	3.84%	3.83%*	4.16%	4.29%	-176	-184*	-182	-177	-178
3	<k2>PC	2.54%	2.53%	2.48%*	2.54%	2.54%	-228	-227	-227	-229	-229*
4	<k2>MB	3.75%	3.64%	3.49%*	3.74%	3.75%	-173	-174	-176	-173	-176*
5	<k3>PC	4.24%	3.27%	3.24%*	3.85%	4.24%	-175	-200*	-199	-185	-176
6	<k3>MB	4.72%	4.33%	4.32%*	4.55%	4.72%	-164	-171*	-169	-168	-164
7	<k4>PC	1.77%	1.59%	1.58%*	1.73%	1.77%	-266	-275*	-274	-269	-266
8	<k4>MB	1.74%	1.72%	1.72%*	1.74%	1.74%	-268	-267	-265	-268*	-267

Parameter Estimation period 2018.7-2018.12 (6mo)

SN	Key word	RMSE					AIC				
		LM	PR2	PR3	IS	BR	LM	PR2	PR3	IS	BR
1	<k1>PC	3.21%	2.52%	2.44%*	3.58%	3.07%	-99	-110*	-109	-93	-100
2	<k1>MB	4.18%	3.94%	3.76%*	4.13%	4.32%	-85	-86	-87	-86	-89*
3	<k2>PC	2.19%	2.12%	2.06%*	2.19%	2.19%	-119	-119	-118	-119	-119*
4	<k2>MB	3.62%	3.53%	2.16%*	3.63%	3.62%	-85	-84	-106*	-85	-88
5	<k3>PC	2.74%	1.93%	1.91%*	2.00%	2.68%	-107	-123	-122	-124*	-108
6	<k3>MB	3.94%	3.70%	3.66%*	3.74%	3.94%	-88	-90	-88	-91*	-89
7	<k4>PC	1.98%	1.91%	1.65%*	1.91%	1.98%	-124	-124	-130*	-126	-124
8	<k4>MB	1.46%	1.45%	1.45%*	1.47%	1.46%	-146*	-144	-142	-145*	-145

* Minimum of RMSE and AIC

6. 최적화분석

최적화분석은 전체클릭률을 최대화하는 클릭률최적화모델과 기대이익을 최대화하는 전환율 최적화분석으로 구분된다.

6.1 클릭률최적화

클릭률최적화모델의 목적함수는 분석대상 키워드 포트폴리오 클릭률의 최대화이며, 제약조건은 예산범위이다(<Equation 4> 참조). 최적화실증분석의 대상기간은 2019년1분기이며, 비교가능성을 높이기 위해 분석대상기간 키워드별 노출수와 클릭광고비 합계로 노출수와 비용예산을 목적함수와 제약조건에 도입하였고, 최소 CPC는 현재 네이버의 최저입찰가에 광고주의 품질지수를 감안하여 60원으로 정하였다.

$$z^* = \underset{z}{\operatorname{argmax}} \sum_{k \in K} \left\{ i_k E[\beta_k(z)] / \sum_{j \in K} i_j \right\}$$

$$s.t. \sum_k i_k E[b_k(z)] E[\beta_k(z)] \leq B, b_k \geq 60$$

(Equation 4)

최적화분석결과를 보면 상기 최적화모델은 <Table 11>에서와 같이 클릭률기준으로는 +2.8~3.5%P, 클릭수기준으로 약 +9.0~11.3%의 개선효과를 보였다. 이를 통해 Rank기반 입찰모델의 유효성을 확인할 수 있었다.

구체적으로 보면, ‘자동차부품<kw>’와 ‘중고부품<k3>’ 또는 ‘자동차중고부품<kq>’의 입찰에서 노출순위를 줄여 비용을 줄이고, 여기서 보충된 예산을 모두 클릭효율이 높은 ‘지파츠<k4>’ 키워드 중심으로 포트폴리오를 구성하는 입찰전략을 진행할 것을 강조하고 있다. 참고로, ISR분석은 단조증가함수이므로 CPC를 과대평가하는 경향이 있다. 따라서 ‘자동차중고부품<k1>’ 모바일의 경우 CPC증가대비 클릭수의 증가효율이 낮기 때문에 포트폴리오를 최소한으로 할 것을 추천한다. 이에 비해 PR3의 경우는 다항회귀의 특성상 증가와 감소를 동시에 고려하고 있어 보다 유연하다.

(Table 11) CTR Optimization Test Results

Key word	PM	Actual Data			(A)Rank based SSA with PR3 (CPC)			(B)Rank based SSA with ISR(CPC)		
		Rank Score	CTR	#Clicks	Rank Score	CTR	#Clicks	Rank Score	CTR	#Clicks
<k1>	PC	72.0%	24.7%	3,839	65.9%	24.1%	3,756	63.4%	23.6%	3,673
	MB	53.1%	22.1%	7,053	99.0%	26.0%	8,315	1.0%	14.6%	4,673
<k2>	PC	54.9%	2.3%	200	1.0%	1.3%	112	91.8%	2.8%	245
	MB	41.1%	5.3%	287	76.3%	1.1%	59	1.0%	18.0%	980
<k3>	PC	96.8%	27.2%	784	59.8%	5.7%	165	98.2%	29.1%	839
	MB	93.4%	29.7%	1,329	28.0%	10.4%	466	95.7%	29.6%	1,325
<k4>	PC	91.8%	27.5%	5,878	99.0%	49.6%	10,598	99.0%	49.6%	10,598
	MB	92.8%	55.3%	19,688	98.2%	56.2%	19,991	99.0%	56.8%	20,232
Total			31.0%	39,058		34.5%	43,462		33.8%	42,565

6.2 전환율최적화

기대이익(전환율)최적화모델의 목적함수는 분석대상 키워드 포트폴리오 전체기대이익의 최대화이며, 제약조건은 예산범위이다(<Equation 5> 참조). 목적함수에서의 수익금액과 전환율은 비교가능성을 위해 각각 2019년 1분기의 클릭당 평균수익(r_k)과 평균구매전환율(v_k)과 동일하게 적용하였고, 기타 제약조건은 클릭률제약조건과 동일하게 하였다.

$$z^* = \operatorname{argmax}_z \sum_{k \in K} i_k E[\beta_k(z)] (r_k v_k - E[b_k(z)])$$

$$\text{s.t. } \sum_k i_k E[b_k(z)] E[\beta_k(z)] \leq B, b_k \geq 0$$

(Equation 5)

분석결과를 보면 최적화모델에 의한 기대이익은 72~73%수준의 광고비로 약 47.7-48.5%의 기대이익 개선효과를 보였다.

이러한 개선효과를 구체적으로 검토해보면 자동차중고부품<k1> 모바일키워드가 손실이 발생

하고 있으므로, 기대이익을 목적함수로 하는 최적화모델에서는 해당 키워드를 배제하고 그 대신 수익성이 높은 ‘지프츠<k4>’로 제안하기 때문에 기대이익이 크게 증가하였다.

■ 전환율최적화에서의 문제점

전환율기반의 기대이익모델에서는 몇 가지 문제점이 발견된다. 우선, 자동차중고부품<k1> 모바일 키워드는 최근 들어 모바일사용의 급증으로 노출과 클릭률이 급격히 증가하고 있는 주요 원천키워드임에도 포트폴리오에서 배제되고 있고, 둘째로 회원가입을 통해 광고주의 사이트를 이용해 본 경험이 있는 고객은 재방문의 가능성이 높으므로 고객입장에서 선호도가 높은 키워드광고를 통해 우선 방문할 수 있도록 유도할 필요가 있는데도 불구하고 단기이익을 추구하는 과정에서 이를 반영하지 못하는 점, 마지막으로 키워드간의 확산효과를 간과하였다.

따라서 광고주의 입장에서는 단순히 구매전환율에만 의존하는 기대이익에서 탈피하여, 클릭률최적화와 병행하거나, 목적함수에 키워드간의

(Table 12) CVR Optimization Test Results

Key word	PM	Actual Data			(A)Rank based SSA with PR3(CPC)			(B)Rank based SSA with ISR(CPC)		
		Rank Score	#Clicks	Exp Profit	Rank Score	#Clicks	Exp Profit	Rank Score	#Clicks	Exp Profit
<k1>	PC	72.0%	3,839	1,063	99.0%	4,953	1,230	95.4%	4,814	1,110
	MB	53.1%	7,053	-823	0.0%	0	0	25.4%	0	0
<k2>	PC	54.9%	200	-45	0.0%	0	0	44.0%	0	0
	MB	41.1%	287	-67	76.0%	59	2	68.4%	0	0
<k3>	PC	96.8%	784	318	98.1%	841	295	99.0%	864	360
	MB	93.4%	1,329	-2	28.0%	465	72	1.0%	291	18
<k4>	PC	91.8%	5,878	5,936	99.0%	10,598	10,747	99.0%	10,598	10,715
	MB	92.8%	19,688	5,846	99.0%	20,232	5,806	99.0%	20,232	5,861
Total			39,058	12,226		37,148	18,153		36,798	18,064

확산효과를 반영하거나, 회원가입 또는 이벤트 사이트 응모 등에서 기회이익을 반영하는 가입 전환율 또는 응모전환율을 반영하는 목적함수로 수정하는 등 전략적인 접근이 필요하다.

이중에서 키워드간의 확산효과를 가져오는 키워드를 찾는 방법과 이를 최적화함수에 반영하는 대안을 검토한다.

6.3 확산효과와 핵심경유키워드

키워드 확산효과를 가져오는 키워드를 핵심경유키워드(Core Transit Keyword, CTK)라고 정의하자. 핵심경유키워드는 비록 현재 기대손실이 예상되는 키워드이지만, 향후 기대이익이 높은 키워드로의 확산을 유도하는 키워드가 된다. 예를 들면, 자동차중고부품<k1>모바일키워드는 기대손실이 발생한다고 해서 혹은 기대이익이 낮다는 이유로 해서 일방적으로 포트폴리오 입찰에서 배제되지만, 이러한 원천키워드는 광고주가 원하는 유통키워드로 전이될 가능성이 매우 높다. 따라서 핵심키워드를 파악하는 방법과 최적화함수의 모델화가 필요하다.

우선 핵심키워드는 마코브체인을 통해 확산효과를 확인할 수 있다. <Table 13>은 2019년1분기에 2회이상 방문한 고객중에서, 처음 클릭하여 유입된 키워드와 이후에 유입된 키워드를 분석하여 전이확률(Transition Probability)을 구성하여 확산효과를 보여주는 마코브체인자료이다⁵⁾.

분석결과 대부분의 키워드에서 최초 유입키워드와 이후 유입키워드가 동일한 비율이 높았는데, ‘키워드충성도’가 높다는 측면에서 고객의

<Table 13> Markov Chain Results

PC Keyword

tot	FROM					Steady State
	<kk1>	<kk2>	<kk3>	<kk4>	<kk5>	
<kk1>	66.8%	9.2%	30.8%	6.4%	13.3%	27.1%
<kk2>	2.6%	40.0%	0.9%	0.3%	0.8%	2.0%
<kk3>	2.7%	2.5%	37.7%	0.7%	1.7%	2.9%
<kk4>	4.4%	1.7%	6.3%	76.5%	4.7%	16.4%
others	23.5%	46.7%	24.3%	16.1%	79.4%	51.6%

Mobile Keyword

tot	FROM					Steady State
	<kk1>	<kk2>	<kk3>	<kk4>	<kk5>	
<k1>	55.8%	11.1%	19.0%	6.1%	10.3%	17.5%
<k2>	0.7%	31.0%	0.5%	0.2%	0.3%	0.5%
<k3>	3.2%	7.0%	27.7%	1.2%	1.4%	2.3%
<k4>	8.8%	8.8%	14.4%	74.8%	7.1%	23.4%
others	31.4%	42.1%	38.4%	17.7%	80.9%	56.3%

인식이 쉽게 바뀌지 않는 것을 알 수 있다. 또한, 전이확률을 검토해보면 자동차중고부품<kk1>키워드가 PC와 모바일키워드에서 ‘핵심경유키워드’ 역할을 담당하고 있고, 궁극적으로 지파츠<kk4>키워드로 확산효과를 일으키고 있음을 확인할 수 있다. 이는 마코브체인의 안정상태에서 확인할 수 있는데, PC키워드의 안정상태(Steady State)는 {kk1,kk2,kk3,kk4,기타}={27.1%,2.0%,2.9%,16.4%,51.6%}이고 모바일의 경우는 {kk1,kk2,kk3,kk4,기타}={17.5%,0.5%,23.4%,23.4%,56.3%}인 점에서, 자동차중고부품<k1>, 지파츠<k4>는 확산효과와 핵심인 핵심경유키워드(CTK)이다.

5) 마코브체인분석에서 <kk1>키워드는 ‘자동차중고부품’, <kk2>키워드는 ‘자동차부품’, <kk3>키워드는 ‘중고부품’의 단어로 시작하는 모든 키워드로 구성하였고, <k4>는 ‘지파츠’ 또는 ‘GPARTS’를 포함한 키워드이다. 따라서 본 연구의 연구대상 키워드의 범위보다 크다. 연구대상 키워드와 구분하기 위해 <kk>기호를 사용한다.

6.3 최적화함수의 수정제안

이번 챕터에서는 기대이익기반의 전환율최적화함수의 문제점을 해결하기 위해 목적함수를 수정하여 확산효과를 반영하는 대안을 제시한다.

키워드 확산효과를 전환율최적화의 목적함수에 반영하는 방법은 광고주가 임의로 중요하다고 판단되는 키워드나 가입전환율이나 기타 목적을 반영한 키워드별로 가중평균을 임의의 책정하는 방법을 적용할 수 있다. 그러나 이러한 방법은 임의적이므로, 이론적인 접근을 통해 포트폴리오에 자동으로 편입될 수 있는 알고리즘을 찾을 필요가 있다. 본 논문에서는 마코브체인에 의한 전이확률 혹은 안정상태를 활용하여 기회이익(Expected Opportunity Profit, EOP)을 추정한 후, 이를 목적함수에 반영하는 대안을 제안한다.

우선 키워드별로 최적화알고리즘과정에서 예상클릭수가 정해졌다고 가정하자. 키워드별 예상클릭은 마코브체인의 전이확률에 의해 다른 키워드에 영향을 주는 확산효과가 나타난다고 가정할 수 있다. 만일 이때 증가하는 키워드의 클릭수에 해당 키워드의 클릭당 공헌이익을 곱한 예상기회이익(EOP)을 목적함수를 수정하여 반영하면, 전환율최적화 문제는 자동적으로 해결될 수 있다. 물론 최적화알고리즘은 다른 제약 조건에 따라 최적값을 찾을 때까지 반복되는데 이때마다 예상기회이익을 반영하여 노출순위를 결정함으로써 포트폴리오에 자동적으로 편입되게 된다.

예상기회이익(EOP)를 계산하기 위해 예측된 키워드별 클릭수벡터 $N = [n_k]$, 공헌이익벡터 $M = [m_k]$, 마코브체인 전이확률벡터 $\delta = [p_k]$ 라고 가정하면, 첫번째 키워드 클릭수가 다른 키

워드 전이되는 클릭수는 전이확률에 따라 $n_1 \delta = [n_1 p_k]$ 가 되고, 예측된 키워드별로 해당 공헌이익인 $(n_1 \delta) M' = [n_1 p_k m_k]$ 의 기회이익을 내포한다. 이 기회이익은 구매전환이 완료된 예측기회이익이므로, 매년 재구매율(δ_k)만큼 구매율이 감소된다. 이를 무한급수형태로 표시하고, 시간가치를 무시하면 키워드별 기대이익은 $E[EOP_k(z)]$ 로 정의된다 (<Equation 6> 참조).

$$E[EOP_k(z)] = \sum_{k \in K} i_k E[\beta_k(z)] \frac{1}{1 - \delta_k} \sum_{j \in K} p_j m_j \quad \langle \text{Equation 6} \rangle$$

전환율최적화의 목적함수는 기존의 키워드별 기대이익에 <Equation 6>의 키워드별 기대기회이익(EOP)합계를 최대화하는 목적함수를 구성할 수 있다. 결국 수정된 최적화모델은 <Equation 7>과 같다.

$$\begin{aligned} z^* &= \operatorname{argmax}_z \\ &\sum_{k \in K} \{ i_k E[\beta_k(z)] (r_k v_k - E[b_k(z)]) + E[EOP_k(z)] \} \\ &\quad - E[EOP(z^*)] \\ \text{s.t. } &\sum_k i_k E[b_k(z)] E[\beta_k(z)] \leq B, \quad b_k \geq 0 \end{aligned} \quad \langle \text{Equation 7} \rangle$$

참고로 최적화반복 알고리즘에 반영된 기대기회이익(EOP)은 포트폴리오에 구성되는 노출순위를 유도하기 위해 목적함수에 추가된 금액이므로, 최종 기대이익을 추정하는 때에는 확정된 노출순위에 의한 기대기회이익 $E[EOP_k(z^*)]$ 를 차감한다.

〈Table 14〉 Revised CVR Optimization Test Results

Key word	PM	Actual Data			(A)Rank based SSA with PR3(CPC)			(B)Rank based SSA with ISR(CPC)		
		Rank Score	#Clicks	Exp Profit	Rank Score	#Clicks	Exp Profit	Rank Score	#Clicks	Exp Profit
<k1>	PC	72.0%	3,839	1,063	68.6%	3,844	3,222	95.4%	4,811	3,768
	MB	53.1%	7,053	-823	100.0%	8,361	1,614	0.0%	4,644	962
<k2>	PC	54.9%	200	-45	0.0%	111	54	0.0%	111	46
	MB	41.1%	287	-67	76.3%	59	13	0.0%	0	0
<k3>	PC	96.8%	784	318	94.4%	733	619	99.1%	867	756
	MB	93.4%	1,329	-2	28.0%	465	317	0.0%	286	180
<k4>	PC	91.8%	5,878	5,936	100.0%	11,077	19,229	100.0%	11,077	19,165
	MB	92.8%	19,688	5,846	100.0%	20,515	10,060	99.9%	20,488	10,091
(-)Contribution							16,916			16,489
Total			39,058	12,226		45,165	18,213		42,285	18,479

수정최적화모델을 의한 최적화결과를 보면 일반 전환율최적화에서는 기대손실로 인해 포트폴리오에서 배제되지만, 확산효과를 반영하는 수정전환율모델에서는 미래의 기회이익을 반영하므로 포트폴리오에 포함되며, 결국 기대손이익시 개선효과가 있음을 확인할 수 있다. 분석결과를 보면 확산효과에 의한 최적화모델은 핵심경유키워드(CTK)에의 투자유도와 함께, 기대이익 약47.7-48.5%의 개선효과가 나타난다(<Table 14> 참조).

7. 결론 및 함의

본 논문은 의사결정주체자인 광고주의 입장에서 통계적 모델에 기반한 포트폴리오구성의 예측모델을 제시하고, 실증자료에 기반하여 최적

화방안과 전략에 대해 연구분석한 논문이다. 키워드입찰시스템은 차하위입찰가(GSP)방식의 비밀입찰시스템으로, 대량의 키워드관리 및 입찰 불확실성과 제약사항이 많아 유인부적합성(Incentive Incomplete)이 높기 때문에 전략적인 접근이 필수적인 광고매체이다. 그런데, 기존의 입찰가액이 독립변수인 모델은 입찰가액과 노출 순위간의 설명력이 불확실하기 때문에 최적화모델을 정립하기가 쉽지 않다는 문제가 있다. 따라서 본 논문에서는 입찰주체인 광고주의 입장에서 노출순위를 의사결정변수로 하는 새로운 검색광고모델을 재정의하고, 예측모델과 최적화모델을 제시하여 실증분석을 진행하였다. 본 논문에서 제시한 클릭률/입찰가액(CPC)예측모델을 기반으로 한 클릭률최적화분석결과 클릭수는 약 +9.0~11.3%의 개선효과를 보였고, 전환율최적화 분석결과 약49%의 기대이익 개선효과를 확인하

였는 바, 본 논문에서 제시한 예측모델과 최적화 모델의 유효성을 확인할 수 있었다. 전환율최적화에서의 기대이익 최적화는 주의해서 활용할 필요가 있는데, 중요한 확산효과를 가져오는 키워드임에도 당장 기대손실이 발생하는 이유로 키워드광고에서 배제할 가능성이 있기 때문이다. 따라서 광고주는 대안으로써 클릭률최적화와 병행하거나, 구매전환율 이외의 회원가입전환율 또는 회원가입 또는 이벤트사이트 응모전환율 등의 개념을 도입하여 이를 기회이익으로 목적함수에 반영하는 등 전략적인 접근이 필요하다. 키워드간의 확산효과를 반영하는 방안중 하나로 전환율최적화모델의 수정모델을 제안하였다. 이러한 수정모델은 핵심경유키워드(CTK)에의 투자를 전략적으로 유지하여 기대이익이 높은 키워드로의 확산효과를 주도한다. 이렇듯 최적화 목적함수는 광고주의 목적에 따라 수정될 수 있다고 판단된다.

본 논문의 공헌내용은 우선, 광고주의 입장에서 노출순위를 의사결정변수로 하는 검색광고모델을 최초로 재정의하여 제시하였다는 점이며, 둘째로는 키워드를 분류기준을 제시하고 범주별 대표키워드를 선정하여 키워드입찰을 그룹단위로 진행할 수 있는 이론적 배경을 제공하였고, 셋째로, 통계적 모델에 기반한 클릭률예측과 입찰가액(CPC)예측모델을 제시하여 개선효과를 보임으로써 노출순위기반 예측모델의 유효성을 확인하였으며, 마지막으로, 광고주의 입장에서 클릭률을 포함하여 광고주의 목적에 맞도록 기회비용을 반영한 전환율모델을 제시하면서, 키워드 확산효과를 반영하는 키워드광고전략 및 기존 최적화모델의 수정모델을 제시하고 적용가능성을 확인하였다.

본 연구의 가장 큰 한계는 키워드별 경쟁자를

포함한 전체 입찰가액과 노출순위 및 클릭률 등의 입찰자료 확보가 어렵다는 점이다. 개별 광고주 본인의 키워드 입찰관련 자료와는 달리, 키워드광고시스템에서 이루어지고 있는 전체 광고주의 입찰현황에 대한 통계자료는 네이버와 같은 광고운영주체가 보유하고 있는 비공개자료이다. 따라서, 전체적인 CPC와 Rank 및 클릭률은 물론 노출이나 품질지수에 대한 정보가 부족하여 정확한 모델설정에는 제약이 있다.

향후의 연구방향으로는 우선 정부가 정책적인 차원에서 친환경 재활용부품사용에 대해 적극적으로 확대하는 정책을 추진 중에 있는데, 향후 이러한 정책이 키워드검색광고에 미치는 영향을 분석하는 것이 필요하다고 판단된다. 또한 예측 모델에 있어 노출과 클릭률, Rank등에서의 시간 흐름에 따른 관계분석, 클릭을 통해 유입된 사용자의 체류시간, 체류 페이지와 구매와 가입에 이르는 페이지 추적, 클릭한 상품에 대한 추적, 유입키워드별 재구매율 또는 재방문 페이지와 체류시간등 머신러닝을 활용한 다양한 예측모델분석이 추가될 수 있다. 키워드는 생명력이 있는 생물체와 같다. 키워드간 확산효과에 대해 추가적으로 시간적인 선호도의 변화와 로그스트림데이터를 분석함으로써 향후 유효한 키워드의 발굴과 예측모델에 반영하는 분석이 유의미하다고 판단한다.

참고문헌(References)

- Abhishek, V. "Essays on Online Advertising." University of Pennsylvania, 2012.
- Abhishek, V., and H. Kartik. "Optimal Bidding in Multi-Item Multislot Sponsored Search

- Auctions.” *Operations research* 61, no. 4 (07/01 2013).
- Bae S. D., and D-H. Park. “The Effect of Mobile Advertising Platform through Big Data Analytics - Focusing on Advertising, and Media Characteristics.” *Journal of Intelligence and Information Systems* 24(2) (2018.06): 37-57.
- Choi K. B., and K. W. Nam. “Analysis of shopping website visit types and shopping pattern.” *Journal of Intelligence and Information Systems* 25(1) (2019.03): 85-107.
- Cholette, S., Ö. Özlük, and M. Parlar. “Optimal Keyword Bids in Search-Based Advertising with Stochastic Advertisement Positions.” [In English]. *Journal of Optimization Theory and Applications* 152, no. 1 (Jan 2012): 225-44.
- DasGupta, B., and S. Muthukrishnan. “Stochastic Budget Optimization in Internet Advertising.” *arXiv e-prints*. (2010). Accessed January 01, 2010.
- Du, X., M. Su, X. M. Zhang, and X. Zheng. “Bidding for Multiple Keywords in Sponsored Search Advertising: Keyword Categories and Match Types.” *Information Systems Research* 28, no. 4 (2017): 711-22.
- Feldman, J., S. Muthukrishnan, M. Pal, and C. Stein. “Budget Optimization in Search-Based Advertising Auctions.” *arXiv e-prints*. (2006). Accessed December 01, 2006.
- Ferrari, S., and F. Cribari-Neto. “Beta Regression for Modelling Rates and Proportions.” *Journal of Applied Statistics* 31, no. 7 (2004/08/01): 799-815.
- Ghose, A., and S. Yang. “An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets.” [In English]. *Management Science* 55, no. 10 (Oct 2009): 1605-22.
- Gong, J., V. Abhisek, and B. Li. “Examining the Impact of Keyword Ambiguity on Search Advertising Performance: A Topic Model Approach.” *MIS Quarterly* 42, no. 3 (2018): 805-29.
- Gopal, R., X. Li, and R. Sankaranarayanan. “Online Keyword Based Advertising: Impact of Ad Impressions on Own-Channel and Cross-Channel Click-through Rates.” *Decision Support Systems* 52, no. 1 (2011/12/01): 1-8.
- Jansen, B. J., and A. Spink. “The Effect on Click-through of Combining Sponsored and Non-Sponsored Search Engine Results in a Single Listing.” Paper presented at the Proceedings of the 2007 Workshop on Sponsored Search Auctions, WWW Conference, 2007.
- Jeon, M. J, “A Study on CTR in Relation to Product Involvement for Keyword Search Advertising Strategy in E-commerce”, *Journal of International Trade & Commerce* 13(3) (2017.06): 543-562.
- Jerath, K., L. Ma, Y-H. Park, and K. Srinivasan. “A “Position Paradox” in Sponsored Search Auctions.” *Marketing science (Providence, R.I.)* 30, no. 4 (07/01 2011): 612-27.
- Jeziorski, P. “Advertiser Prominence Effects in Search Advertising.” *Management science* 64, no. 3 (03/01 2018): 1365-83.
- Kamijo, Y. “Bidding Behaviors for a Keyword Auction in a Sealed-Bid Environment.” *Decision Support Systems* 56 (2013): 371-78.
- Kim, M. G., N. G. Kim and I. H. Jung. “A Methodology for Extracting Shopping-Related Keywords by Analyzing Internet

- Navigation Patterns”, *Journal of Intelligence and Information Systems* 20(2) (2014.06): 123-136.
- Kim S. Y., and G. W. Kim. “Ontology-based User Customized Search Service Considering User Intention.” *Journal of Intelligence and Information Systems* 18(4) (2012.12): 129-143.
- Liu, F., and E. C. Eugenio. “A Review and Comparison of Bayesian and Likelihood-Based Inferences in Beta Regression and Zero-or-One-Inflated Beta Regression.” *Statistical methods in medical research* 27, no. 4 (04/01 2018): 1024-44.
- Paolino, P. “Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables.” *Political Analysis* 9, no. 4 (2001): 325-46.
- Rutz, O. J., and R. E. Bucklin. “From Generic to Branded: A Model of Spillover in Paid Search Advertising.” *Journal of marketing research* 48, no. 1 (02/01 2011): 87-102.
- Schlangenotto, D., M. Poniowski, and D. Kundisch. “What Drives Paid Search Success? A Systematic Literature Review.” Paper presented at the Americas Conference on Information Systems 2018: Digital Disruption, AMCIS 2018, 2018.
- Selçuk, B., and Ö. Özlük. “Optimal Keyword Bidding in Search-Based Advertising with Target Exposure Levels.” *European Journal of Operational Research* 226, no. 1 (2013): 163-72
- Smithson, M., and J. Verkuilen. “A Better Lemon Squeezer? Maximum-Likelihood Regression with Beta-Distributed Dependent Variables.” [In English]. *Psychological Methods* 11, no. 1 (Mar 2006): 54-71.
- Wang, F., W. Suphamitmongkol, and B. Wang. “Advertisement Click-through Rate Prediction Using Multiple Criteria Linear Programming Regression Model.” *Procedia Computer Science* 17 (2013/01/01/ 2013): 803-11.

Abstract

An Empirical Study on Statistical Optimization Model for the Portfolio Construction of Sponsored Search Advertising(SSA)

Hognkyu Yang* · Juneseok Hong** · Wooju Kim***

This research starts from the four basic concepts of incentive incompatibility, limited information, myopia and decision variable which are confronted when making decisions in keyword bidding. In order to make these concept concrete, four framework approaches are designed as follows; Strategic approach for the incentive incompatibility, Statistical approach for the limited information, Alternative optimization for myopia, and New model approach for decision variable.

The purpose of this research is to propose the statistical optimization model in constructing the portfolio of Sponsored Search Advertising (SSA) in the Sponsor's perspective through empirical tests which can be used in portfolio decision making. Previous research up to date formulates the CTR estimation model using CPC, Rank, Impression, CVR, etc., individually or collectively as the independent variables. However, many of the variables are not controllable in keyword bidding. Only CPC and Rank can be used as decision variables in the bidding system. Classical SSA model is designed on the basic assumption that the CPC is the decision variable and CTR is the response variable. However, this classical model has so many huddles in the estimation of CTR. The main problem is the uncertainty between CPC and Rank. In keyword bid, CPC is continuously fluctuating even at the same Rank. This uncertainty usually raises questions about the credibility of CTR, along with the practical management problems. Sponsors make decisions in keyword bids under the limited information, and the strategic portfolio approach based on statistical models is necessary.

In order to solve the problem in Classical SSA model, the New SSA model frame is designed on the basic assumption that Rank is the decision variable. Rank is proposed as the best decision variable in

* Graduate Program in Technology Policy, Yonsei University

** Department of Management Information Systems, Kyonggi University

*** Corresponding Author: Wooju Kim

Department of Information and Industrial Engineering, Yonsei University

50 Yonsei-ro Seodaemun-Gu, Seoul 03722, Korea

Tel: +82-2-2123-4010, Fax: +82-2-364-7807, E-mail: wkim@yonsei.ac.kr

predicting the CTR in many papers. Further, most of the search engine platforms provide the options and algorithms to make it possible to bid with Rank. Sponsors can participate in the keyword bidding with Rank. Therefore, this paper tries to test the validity of this new SSA model and the applicability to construct the optimal portfolio in keyword bidding.

Research process is as follows; In order to perform the optimization analysis in constructing the keyword portfolio under the New SSA model, this study proposes the criteria for categorizing the keywords, selects the representing keywords for each category, shows the non-linearity relationship, screens the scenarios for CTR and CPC estimation, selects the best fit model through Goodness-of-Fit (GOF) test, formulates the optimization models, confirms the Spillover effects, and suggests the modified optimization model reflecting Spillover and some strategic recommendations.

Tests of Optimization models using these CTR/CPC estimation models are empirically performed with the objective functions of (1) maximizing CTR (CTR optimization model) and of (2) maximizing expected profit reflecting CVR (namely, CVR optimization model). Both of the CTR and CVR optimization test result show that the suggested SSA model confirms the significant improvements and this model is valid in constructing the keyword portfolio using the CTR/CPC estimation models suggested in this study. However, one critical problem is found in the CVR optimization model. Important keywords are excluded from the keyword portfolio due to the myopia of the immediate low profit at present. In order to solve this problem, Markov Chain analysis is carried out and the concept of Core Transit Keyword (CTK) and Expected Opportunity Profit (EOP) are introduced. The Revised CVR Optimization model is proposed and is tested and shows validity in constructing the portfolio. Strategic guidelines and insights are as follows; Brand keywords are usually dominant in almost every aspects of CTR, CVR, the expected profit, etc. Now, it is found that the Generic keywords are the CTK and have the spillover potentials which might increase consumers awareness and lead them to Brand keyword. That's why the Generic keyword should be focused in the keyword bidding.

The contribution of the thesis is to propose the novel SSA model based on Rank as decision variable, to propose to manage the keyword portfolio by categories according to the characteristics of keywords, to propose the statistical modelling and managing based on the Rank in constructing the keyword portfolio, and to perform empirical tests and propose a new strategic guidelines to focus on the CTK and to propose the modified CVR optimization objective function reflecting the spillover effect in stead of the previous expected profit models.

Key Words : Sponsored Kearch Advertising, CTR, Spillover, Optimization, Keyword Bidding

Received : May 27, 2019 Revised : June 21, 2019 Accepted : June 25, 2019

Publication Type : Regular Paper(Fast-track) Corresponding Author : Wooju Kim

저 자 소개



양홍규

연세대학교에서 BBA과정 학사학위와 MBA과정 석사학위를 취득하였고, 동 대학교 기술정책협동 박사과정에 있다. 현재는 주식회사 슈펙스비엔피 자동화사업부 사장으로 재직중이다. 주요 연구 관심분야는 금융공학, 지능형 전자상거래, 머신러닝, 빅데이터 등이다.



홍준석

서울대학교 경영학과를 졸업한 후, KAIST 경영과학과에서 석사학위를, 테크노경영대학원에서 박사학위를 취득하였다. 현재 경기대학교 경영정보학과 교수로 재직 중이다. 주요 관심분야는 딥러닝을 활용한 지능형 의사결정 및 에이전트, 빅데이터 플랫폼 및 분석 시스템 등이다.



김우주

연세대학교 BBA과정 학사 학위를 취득하였고, KAIST에서 경영과학 석사, 박사 학위를 취득하였다. 현재 연세대학교 정보산업공학과 교수로 재직 중이며, 주요 연구 관심분야는 지능형 시맨틱웹, 지식관리 및 인공지능, 빅데이터 등이다.