

## 검색결과와 최적 조정 비율 분석

우수한<sup>1</sup> · 이은희<sup>2</sup> · 김기훈<sup>2\*</sup>

<sup>1</sup>CJ푸드빌(주), <sup>2</sup>고려대학교 경영대학

### Analysis of the Optimal Degree of Search Result Modification

Soohan Woo<sup>1</sup> · Eun hee Lee<sup>2</sup> · Kihoon Kim<sup>2</sup>

<sup>1</sup>CJ Foodville, <sup>2</sup>Korea University Business School

#### ■ Abstract ■

Naver, a leading search engine in South Korea, may show modified and reorganized search results for some trendy and popular keywords; when popular words such as the titles of soap operas and films are searched for, all the detailed and well-organized information regarding them can be presented. By recognizing that search engines may modify and reorganize search results for some popular keywords, we mathematically model the impact of the degree of modification of search results on the search engine's profit to derive its optimal modification degree. We show how the optimal degree of search result modification may change according to the different shapes of the search engine's advertising revenue function.

Keywords : Search Engine, Search Quality, Search Result Modification, Two-Sided Market

### 1. 연구 배경 및 관련된 문헌 연구

검색엔진은 인터넷상에 있는 정보 중 원하는 자료를 쉽게 찾아볼 수 있도록 해 주는 소프트웨어이다. 대한민국의대표적인 검색엔진인 네이버는 통합검색

이라는 방법으로 검색결과 화면을 보여주며 다년간 국내 시장점유율 1위를 기록하고 있고, 세계적으로 널리 사용되는 구글은 페이지랭크라는 기술력을 통한 검색결과 화면을 보여준다. 네이버에서 사용하는 통합검색이라는 방법은 검색어의 특성에 따라

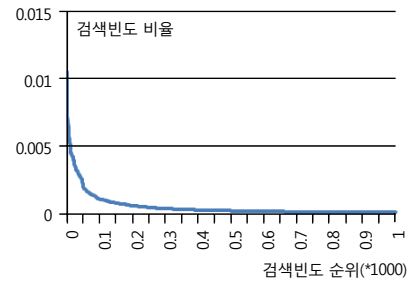
논문접수일 : 2014년 03월 23일 논문게재확정일 : 2014년 09월 02일

논문수정일(1차 : 2014년 06월 20일, 2차 : 2014년 08월 27일)

\* 교신저자 kihoon@korea.ac.kr

해당하는 정보를 상위에 노출시켜주는 방법이다. 그런데 이러한 방법은 검색어 별로 상이한 결과를 보여준다. 예를 들어 드라마 제목이나 영화 제목과 같이 사람들이 많이 찾는 검색어는 이와 같은 단어를 검색했을 때 예고편 영상이나 작품 줄거리 혹은 출연배우 정보 등과 같은 추가적인 데이터베이스가 포함되어 있다. 다시 말해서 통합검색은 사전에 정해진 특정 단어가 검색되었을 때 결과 화면 상단에 해당되는 연관 정보를 노출시켜 주는 것이다. 이와 같은 추가적인 데이터베이스에 관련된 정보를 결과 화면에 보여주기 위해서는 사람의 임의적인 조정이 필요하다. 비록 연관되는 데이터베이스를 단순하게 연결하여 화면에 노출하는 방식이 사용된다고 할지라도 연관된 데이터베이스의 인터넷 주소를 찾아서 입력하는 등의 프로그램적인 조정이 필요한 것이다. 따라서 네이버 통합검색에서 검색결과에 수동적인 조정을 했다고 가정하며 수학적 모델링을 통해 단순화 시킨 상황에서 검색엔진의 성과에 어떤 영향을 미치는지 살펴볼 것이다.

한편, 사용자들이 입력하는 검색어의 검색빈도는 소수의 특정 단어에 높게 나타나고, 다수의 단어들에 대해 낮게 나타나는 특징을 갖고 있음을 실제 검색포탈의 자료조사를 통해 알 수 있다. [그림 1]에서는 네이버 키워드 스테이션 [http://searchad.naver.com/CMKS02/CMKS0200\\_A01/](http://searchad.naver.com/CMKS02/CMKS0200_A01/); 네이버 키워드 스테이션에서 제공하는 검색어 통계를 활용하여 수집한 2011년 6월 검색된 약 6만 8천여 개의 검색어에 대한 검색횟수 중 1위부터 1000위까지 상위 검색어를 표본으로 추출하였고, y축은 개별 검색어에 대한 검색 횟수의 전체 검색 횟수에 대한 비율을, x축은 1000개의 표본 중 상위 몇 퍼센트에 해당 단어가 위치하는지를 보여준다. 본 연구는 [그림 1]의 빈도분포와 같은 일반적인 확률분포를 가정하고, 검색엔진이  $\alpha$ 만큼의 확률로 검색될 상위 검색어에 대해서 조정을 할 경우에 이익에 어떤 영향을 줄 지를 간단하게 모델링하여, 최적의 조정 확률을 구하게 된다. 그러면, 그 확률로부터 최적의 조정 비율을 구할 수 있게 된다.



[그림 1] 네이버 검색 빈도 통계

검색엔진의 품질을 측정하는 기존 문헌들은 주로 문헌정보학에서 많이 다루어 졌으며 크게 설문조사를 이용한 주관적인 품질 혹은 사용되는 기능적인 품질과 관련되어 있다. 지금까지 행해졌던 연구에서 검색엔진의 품질은 인터뷰를 통한 사용자의 만족도[14]를 활용해 측정하거나, 인터넷 자료의 색인 작업이나 순위 결정 요소들과 같은 검색엔진의 기술적인 요소들을 통해 측정했다[9]. 또한 Lewandowski[10]는 검색엔진의 품질을 측정하는 방법으로 보유하고 있는 자료들을 업데이트 하는 색인의 최신성을 고려했고 Lewandowski and Hochstotter[11]는 색인 품질과 검색 특성의 품질, 검색엔진의 사용성에 대한 측정으로 품질을 측정할 수 있다고 했다. Edosomwan[6]은 7개의 검색엔진에 10개의 쿼리를 입력하여 정확성과 반응시간을 통해 품질을 비교하였다. Hawking and Craswell et al.[7]은 TREC이라고 불리는 방법론을 사용해 20개의 검색엔진의 효과성을 평가했다. 역시 TREC의 데이터를 이용하여, Clarke et al.[5]는 정보 검색에 있어서 정보의 유사성과 다양성 등을 평가할 수 있는 방법을 제안하였다. Vaughan[15]은 웹 검색엔진 성능을 평가하는 측정요소들을 제안했는데, 결과 순위의 질과 최고 순위의 페이지를 찾을 수 있는 능력, 검색된 결과의 안정성을 요소들로 제안했다. Xie and Wang et al.[16]은 일반적인 서비스의 질을 측정하는 SERVQUAL 프레임워크를 응용한 인터넷 검색엔진의 품질 차원을 제시했고 설문조사를 통해 품질을 조사했다. 본 연구는 검색결과와 품질을 평가하는 데 초점을 두기 보다는, 검색엔진이 결과를 조정할 때, 수익 함수의 형태에 따라 조정 비율이 어떻게

달라질 수 있는 지를 분석하는 데 의의를 두고 있다.

검색엔진은 사용자와 광고주 또는 사용자와 정보 제공자를 양면으로 두는 양면 시장의 플랫폼으로 생각해 볼 수 있다. 양면시장은 비교적 최근 들어 주목 받고 있는 이론으로 하나 혹은 여러 개의 플랫폼이 두 개의 상호 네트워크 효과를 갖고 있는 서로 다른 사용자 그룹을 갖고 있는 형태이다[2]. 자세히 설명하면 양면시장 이론은 네트워크 효과와 네트워크 외부효과에서부터 시작된 이론으로써, 하나의 플랫폼을 기준으로 양쪽으로 연결되어 있는 참여자가 존재하고, 상대방의 참여자 수가 늘어날수록 이득을 얻게 되는 네트워크 효과로 인해서 각 참여자들이 시장에 참여하고자 하는 동기를 얻게 되는 특징이 있다[1]. 검색엔진과 네트워크 효과에 대한 Argenton and Prüfer[3]의 연구에서는 검색엔진끼리 검색에 대한 정보를 공유함으로써 검색의 질과 소비자 잉여 나아가 사회전체의 후생(welfare)에도 긍정적인 효과를 미칠 수 있음을 지적하였다. Kim[8]은 하위 검색엔진의 경우, 사용자의 질문이 복잡하고 검색엔진 사이트의 기존 온라인 콘텐츠가 빈약할수록, 지식검색서비스로 인한 시장점유율 증가가 커지는 현상에 대해 연구하였다. 특히, Taylor[12]와 White[16]는 유료 검색광고 서비스와 검색결과와의 관계를 다루었다. 두 논문에서는 유료광고로 인해 검색결과에 질이 저하될 수 있음을 보였다. 하지만, 유료 광고는 검색엔진 사이트의 주요한 수입원이다. 따라서 유료 광고 때문에 검색결과에 질이 다소 저하되더라도 검색엔진은 유료광고를 하지 않을 수 없는 tradeoff가 존재함을 알 수 있었다. 또한, Telang et al.[13]의 논문은 검색엔진의 수익이 검색엔진의 효용 또는 질에 비례한다는 가정하에 경쟁을 묘사하였다. 본 연구는 사용자들의 검색어에 수동적으로 조정을 가했을 경우를 가정하였다. 수동적 조정을 좋아하는 사용자와 좋아하지 않는 사용자의 비율에 따라 수동적인 조정이 검색엔진의 수익에 미치는 영향도 달라질 것이다. 모델링으로 검색엔진이 최대 수익을 낼 때 수동조정 비율의 최적값을 찾아볼 수 있었다.

## 2. 모 형

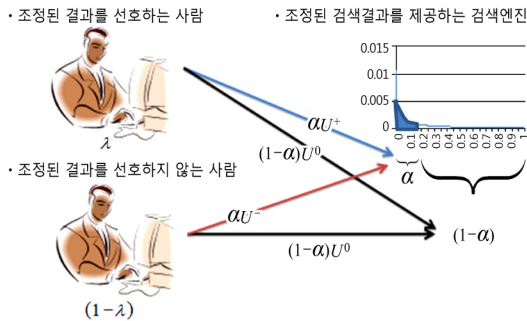
본 연구는 한국의 네이버와 미국의 구글처럼 검색엔진 시장을 선도하는 경우를 고려하여 독점적인 검색엔진을 고려할 것이다. 우선, 검색엔진에 접속하는 사용자의 유형은 두 가지라고 가정한다. 하나는 검색엔진이 수동적인 조정을 통해 보여주는 결과 화면을 선호하는 유형( $\lambda$ )이다.  $\lambda$ 를 조정된 검색어를 통하여 높은 효용을 얻을 수 있는 확률로 해석해도 같은 효용함수를 가지게 한다. 이들은 새롭게 변화하는 세상에 관심이 많으며 그렇기 때문에 최신 정보를 찾기를 원한다. 검색결과에 수동적인 조정의 목적은 사용자들의 검색 서비스 사용 만족감을 높이기 위함이다. 그렇게 하기 위해서 검색엔진은 사회에서 이슈가 되고 있는 검색어들에 대해서 사용자가 원하는 최신의 다양한 자료를 보기 좋게 제공한다. 따라서 조정을 선호하는 사람들은 새로운 트렌드가 반영된 검색결과 페이지에 만족하며 원하는 정보를 얻을 수 있다. 반면 다른 사용자 유형의 사람들( $1-\lambda$ )은 조정된 결과를 원하지 않는다. 그들은 기존의 널리 알려진 지식을 얻기 원하기 때문에 새로운 트렌드가 반영되어 최신의 정보를 포함하고 있는 조정된 검색결과를 선호하지 않으며 원하는 정보를 얻지 못한다. 먼저 모형에 사용되는 변수를 기호로 정의하면 다음과 같다.

〈표 1〉 모델링에 사용된 결정 변수 및 매개 변수 정의

변수	설명
$\alpha \in [0, 1]$	검색엔진에서 조정된 검색어가 검색될 확률
$\lambda \in [0, 1]$	조정된 검색결과를 선호하는 사람들의 비율
$1-\lambda$	조정된 검색결과를 선호하지 않는 사람들의 비율
$U^0$	검색결과가 조정되지 않았을 때 느끼는 효용
$U^+$	검색결과에 조정의 의해 긍정적인 영향을 받는 사람들이 느끼는 효용
$U^-$	검색결과에 조정의 의해 부정적인 영향을 받는 사람들이 느끼는 효용

본 연구에서 결정할 변수인  $\alpha$ 는 사용자가 검색하는 검색어 빈도를 확률분포로 가정 했을 때 누적확률 분포값을 나타낸다. 그러므로, 사용자가 검색한 검색

어는  $\alpha$ 의 확률로 조정된 결과 화면을 보여줄 것이다. 검색어의 빈도에 따른 확률분포를 구하여  $\alpha$ 를 대입하면 전체 검색어 중에 어느 정도의 검색어까지 조정해야 하는지를 구할 수 있기 때문에, 최적 조정 비율도 구할 수 있게 된다. 이 논문에서는 편의상  $\alpha$ 를 최적 조정 비율로 부른다.



[그림 2] 조정된 검색결과에 따른 사용자의 효용 변화

사용자들이 검색엔진을 사용하면서 얻는 효용은  $U$ 변수로 정의한다.  $U$ 변수는 3가지 경우를 가정할 수 있다. 첫 번째는 사용자가 검색한 검색어가 검색엔진의 조정 범위에 들어있지 않아서 영향을 받지 않는 경우이며  $U^0$ 변수로 정의한다. 이것은 일반적인 사용자들이 일반적인 검색엔진에서 얻는 효용이라고도 볼 수 있다. 두 번째는 사용자가 검색한 검색어가 조정 범위에 포함 되어 있을 경우이다. 만약 사용자가 조정된 검색결과를 선호하는 유형이라면 효용은 일반적인 상황보다 증가할 것이고 우리는 증가한 효용을  $U^+$ 로 정의한다. 한편 사용자가 조정된 검색결과를 선호하지 않는 유형이라면 조정된 검색결과에 불편함을 느낄 것이고 원하는 정보를 찾기 위해 많은 시간과 노력이 투자될 것이다. 그렇기 때문에 우리는 이와 같을 때 사용자의 효용이 감소한 것으로 볼 수 있으며  $U^-$ 로 정의한다. 세 개의 효용 관련 매개 변수의 상대적인 크기는  $U^+ > U^0 > U^-$ 라고 가정한다. 두 가지 유형으로 나누어진 전체 사용자들이 검색결과가 조정된 검색엔진을 사용함에 따라 얻는 효용을 수학적으로 표현 할 수 있다. 먼저, 조정된 검색결과를 선호하는 사용자가 얻는 기대 효용은  $E[U_1] = \alpha U^+ + (1-\alpha)U^0$ 이다. 반면에, 조정된 검색

결과를 선호하지 않는 사용자의 기대 효용  $E[U_2] = \alpha U^- + (1-\alpha)U^0$ 은 조정비율이 커지는 만큼 효용이 감소할 것이다. 두 그룹의 사용자 모두 조정되지 않은 검색결과  $(1-\alpha)$ 는 사용자의 효용에 영향을 끼치지 못한다. 두 가지 유형의 사용자로부터 얻는 효용을 평균화하면,  $E[U] = \lambda E[U_1] + (1-\lambda)E[U_2]$ 가 된다.

검색엔진의 주된 수입원인 광고료는 사용자의 수에 상응하여 증가할 것이기 때문에, 검색엔진의 수익은 사용자가 얻는 평균 효용인  $E[U]$ 에 비례한다고 가정한다. 139페이지의 [그림 3]을 보면 네이버의 순방문자 수와 광고매출액의 관계를 보면 볼록함수 형태로 증가함을 관찰할 수 있다. 순방문자 수가 평균 효용과 비례한다고 생각하면, 검색엔진의 광고수익이 검색엔진의 평균 효용과 비례한다고 설정할 수 있다. 우리는 수익 함수가 선형인 경우와 제곱근의 형태인 두 가지 경우를 고려하며, 검색결과가 조정된 검색어가 검색될 확률인  $\alpha$ 가 증가할수록 더 많은 비용이 든다고 가정하여 볼록 함수(convex function)인 이차 함수의 형태를 비용 함수에 사용한다. Telang et al.[13] 역시 이차 함수 형태의 비용 함수를 사용하였다. 그리하여, 검색엔진은 다음의 최적화 문제를 풀어 이익을 극대화 한다 :

$$\underset{\alpha \in [0, 1]}{\text{Maximize}} f(\alpha) = g(\lambda E[U_1] + (1-\lambda)E[U_2]) - k\alpha^2.$$

### 3. 모형 분석

이 절에서는 우선 일반적인 수익함수에 대한 최적 조정 비율을 알아본 후, 수익함수가평균효용의 증가에 따라 선형, 오목, 그리고 볼록 형태로 증가할 때의 구체적인 최적 조정 비율을 알아본다. 마지막으로, 세 가지 다른 수익함수에서의 최적 조정 비율 값을 서로 비교해 본다.

#### 3.1 일반적인 수익함수

일반적인 수익함수일 때의 최적 조정 비율을 구하기 전에, 식을 간단하게 전개하기 위해서,  $d = \lambda U^+$

$+ (1-\lambda)U^- - U^0 = E[U^m] - U^0$ 라고 정의하면,

$$f(\alpha) = g(d\alpha + U^0) - k\alpha^2$$

$E[U^m]$ 은 검색자가 조정된 검색어를 찾았을 때 얻을 수 있는 기대 효용으로 해석할 수 있으므로,  $d$ 는 조정된 검색어를 통한 기대 효용과 원래 효용과의 차이라고 볼 수 있다. 이익 함수가  $\alpha$ 에 대한 오목함수라면,  $f'(\alpha) = 0$ 이 되는 해를 찾아서, 그 해가 1을 넘지 않으면, 그 해가 최적 조정 비율이 된다 만약 1을 넘으면, 1이 최적 조정 비율이 된다. 이익 함수가  $\alpha$ 에 대한 볼록함수라면,  $f(0)$ 과  $f(1)$ 의 값을 비교하면 최적 조정 비율을 찾을 수 있다. 위의 내용을 정리하면 아래 Theorem 1에서와 같이 일반적인 수익함수에 대한 최적 조정 비율을 구할 수 있다. Theorem 1의 증명은 부록에서 제공된다.

**Theorem 1.**

만약  $d^2g''(d\alpha + U^0) < 2k$ 이 항상 성립하면, 최적의 조정 비율  $\alpha^* = \min(\alpha^m, 1)$ . 여기에서,  $\alpha^m$ 은 일차 조건(first-order condition)을 만족하는 해이다. 즉,  $f'(\alpha^m) = 0$ 을 만족한다. 만약  $d^2g''(d\alpha + U^0) \geq 2k$ 이 항상 성립하면, 최적의 조정 비율은 0과 1중에서 더 큰 이익을 가져다 주는 값을 선택하면 된다. 즉,  $\alpha^* = \operatorname{argmax}_{x \in \{0, 1\}} f(x)$ .

Theorem 1에서 보듯이 수익함수가 선형이거나 오목함수이면  $g''(\cdot) \leq 0$ 이 성립하게 되어, 이익함수가 오목함수가 된다. 만약 수익함수가 볼록함수라면  $g''(\cdot) > 0$ 이 되어서, 조건에 따라 이익함수가 오목 또는 볼록 함수일수도 있다.

**3.2 선형의 수익함수**

$g(x) = x$ 일 때는, 이익함수는 아래와 같은 형태를 가지게 된다:

$$f(\alpha) = d\alpha + U^0 - k\alpha^2.$$

$f(\alpha)$ 는  $\alpha$ 에 대한 오목 함수(concave function)이므로  $\alpha$ 에 대하여 미분한 식을 0으로 두면 이익 함수를 최대화하는 조정 확률을 찾을 수 있다.

$$f'(\alpha) = d - 2k\alpha = 0 \Rightarrow \alpha^{\max} = d/2k$$

한편 최적 조정 확률  $\alpha^*$ 는 0과 1사이이므로 선형의 수익함수를 가정했을 때의 최적의 조정 확률은 아래의 Proposition 1로 정리된다.

**Proposition 1.**

수익함수가 선형일 때, 검색결과와 수동적인 조정의 최적 확률인  $\alpha^*$ 는 다음과 같이 결정 된다.

$$\alpha^* = \begin{cases} 0 & \text{if } d \leq 0 \\ 1 & \text{if } d > 0 \text{ and } k \leq d/2 \\ d/2k & \text{if } d > 0 \text{ and } k > d/2 \end{cases}$$

Proposition 1은 조정을 했을 때의 기대 효용이 조정하지 않았을 때의 효용보다 클 때에만 검색엔진이 조정 결과를 조정할 필요가 있음을 보여준다. 그리고, 두 효용간의 차이가 커질수록 최적 조정 확률(따라서 비율)이 커진다는 것도 보여준다. 즉,  $U^+$ 나  $\lambda$ 가 커질수록  $\alpha^*$ 는 증가한다.

**3.3 오목(제곱근) 형태의 수익함수**

선형 수익함수 이외에 우리는 제곱근의 형태를 갖고 있는 수익함수에 대해서 생각해 볼 것이다(즉,  $g(x) = \sqrt{x}$ ). 이 함수는 사용자들의 기대 효용이 늘어날수록 (그리하여 사용자의 수가 증가할수록) 검색엔진의 수익 증가율은 감소함을 의미한다. 다시 말해서 더 많은 사용자들이 검색엔진을 사용함에 따라 검색엔진의 수익은 증가하지만 그 증가하는 경향은 초반에 많이 증가하고 후반으로 가면 갈수록 적게 증가하는 것이다. 즉, 이것은 사용자 수가 어느 정도 일정 규모 이상이 되면 사용자 수 증가를 통한 광고 수익의 증가가 제한적임을 가정하는 것이다. 즉, 검색엔진의 이익은

$$f(\alpha) = \sqrt{d\alpha + U^0} - k\alpha^2.$$

목적함수를  $\alpha$ 에 대해서 미분하면(목적함수는  $\alpha$ 에 대하여 concave하므로),

$$f'(\alpha) = \frac{d}{2\sqrt{d\alpha + U^0}} - 2k\alpha.$$

만약  $d$ 가 양수가 아니라면, 목적함수는 모든  $\alpha \in (0, 1)$ 에 대해서 감소하게 되므로, 목적함수를 최대화 시키는  $\alpha$ 는 0이다. 즉, 조정된 결과가 주는 평균 효용이 원래보다 감소하므로 조정을 하지 않는 것이 최적의 결정이 된다.  $d$ 가 양수인 경우에는  $f'(0)$ 이 양수이므로,  $f'(1)$ 의 부호에 따라  $\alpha^*$  값이 변화하게 된다. 즉,

$f'(1) = \frac{d}{2\sqrt{d+U^0}} - 2k \geq 0$ 이면  $\alpha^* = 1$ ;  $f'(1) < 0$ 이면  $\alpha^* \in (0, 1)$ 이 된다.

그리고,  $f'(1) < 0$ 일 때의  $\alpha^*$ 는  $f'(\alpha) = 0$ 을 만족한다. 위 내용을 토대로 수익함수가 오목함수인 제공근 형태일 때, 최적 조정 확률에 대하여 정리하면 아래의 Proposition 2를 얻을 수 있다.

**Proposition 2**

검색결과의 최적 조정 확률,  $\alpha^*$ 는 다음과 같이 결정 된다.

$$\alpha^* = \begin{cases} 0 & \text{if } d \leq 0; \\ 1 & \text{if } d > 0 \text{ and } k \leq \frac{d}{4\sqrt{d+U^0}}; \\ \in (0, 1) & \text{if } d > 0 \text{ and } k > \frac{d}{4\sqrt{d+U^0}}. \end{cases}$$

마지막 조건에 해당하는 최적 조정 확률  $\alpha^*$ 는

$$\frac{d}{2\sqrt{d\alpha + U^0}} - 2k\alpha = 0 \tag{1}$$

을 만족하는 실수해이다.

수익함수가 선형일 때와 비슷하게 최적 조정 확률 (따라서 비율)은 조정에 따른 평균 효용의 향상 정도( $d$ )와 조정 비용 계수인  $k$ 의 크기에 따라 달라진다. 즉, 조정에 따른 평균 효용이 조정 전 보다 크고 조정 비용 계수가 일정 수준 이상인 경우에 최적 조정 비율은 0과 1사이에 존재하게 된다.

$\lambda$ 가 증가할 때 최적 조정 비율이 증가하는 지를 알아보기 위하여, 식 (1)을  $\alpha$ 에 대하여 정리하면

$$g(\alpha, d) := 16dk^2\alpha^3 + 16k^2\alpha^2 U^0 - d^2 = 0 \tag{2}$$

식 (2)에 음함수 이론(implicit function theorem)을 이용하면,

$$\frac{d\alpha}{d\lambda} = \frac{d\alpha}{dd} \frac{dd}{d\lambda} = -\frac{g_d}{g_\alpha} \frac{dd}{d\lambda} = \frac{2d - 16k^2\alpha^3}{48dk^2\alpha^2 + 32k^2\alpha U^0} \frac{dd}{d\lambda} > 0$$

식 (2)로부터  $16k^2\alpha^3 < d$ 를 얻게 되어 위 미분 값이 양임을 보일 수 있다. 즉, 조정된 검색결과를 좋아하는 검색자의 비율이 증가할수록(또는 검색자들이 조정된 결과를 좋아하는 확률이 증가할수록) 최적 조정비율이 증가함을 알 수 있다.

Proposition 1과 2를 통하여 수익함수가 선형일 때와 concave (제공근 형태)일 때의 최적 조정 비율을 비교해 볼 수 있다. 수익함수가 선형일 때는 조정된 결과를 보는 검색자들이 얻는 효용( $E[U^m]$ )이 커져도 비례하여 검색엔진이 수익을 얻을 수 있지만, 제공근 형태일 때에는 조정으로 인한 효용이 높아지게 되면 효용의 증가 대비 수익의 상승폭이 줄어들게 된다. 이런 특성에 의하여 Theorem 2를 얻을 수 있다. Theorem 2의 증명은 부록에서 제공된다.

**Theorem 2.**

$d = E[U^m] - U^0 > 0$ 라고 가정하자. 수익함수가 선형일 때의 최적 조정 비율을  $\alpha_1^*$ , 제공근 형태일 때의 최적 조정 비율을  $\alpha_c^*$ 라고 하자. 조정 비용 계수인  $k$ 가 매우 작다면(즉,  $k \leq \min\left(\frac{d}{2}, \frac{d}{2\sqrt{d+U^0}}\right)$ ),

$\alpha_c^* = \alpha_i^* = 1$ . 만약,  $k > \min\left(\frac{d}{2}, \frac{d}{2\sqrt{d+U^0}}\right)$  라면, 두 조정 비율의 차이는 조정이 없는 경우에 검색자가 느끼는 효용( $U^0$ )과 조정이 있을 경우에 검색자가 얻는 평균 효용( $E[U^m]$ ), 그리고 조정 비용 계수  $k$ 의 정도에 의하여 다음과 같이 결정된다.

If  $E[U^m] < \frac{1}{4}$ ,  $\alpha_i^* < \alpha_c^*$ ;  
 If  $U^0 \geq \frac{1}{4} (\therefore E[U^m] \geq \frac{1}{4})$ ,  $\alpha_c^* < \alpha_i^*$ ;  
 If  $E[U^m] < \frac{1}{4}$ ,  $U^0 < \frac{1}{4}$ , and  
 $k < \frac{d^2}{2(0.25 - U^0)}$ ,  $\alpha_c^* < \alpha_i^*$ ;  
 If  $E[U^m] \geq \frac{1}{4}$ ,  $U^0 < \frac{1}{4}$ , and  
 $k \geq \frac{d^2}{2(0.25 - U^0)}$ ,  $\alpha_i^* < \alpha_c^*$

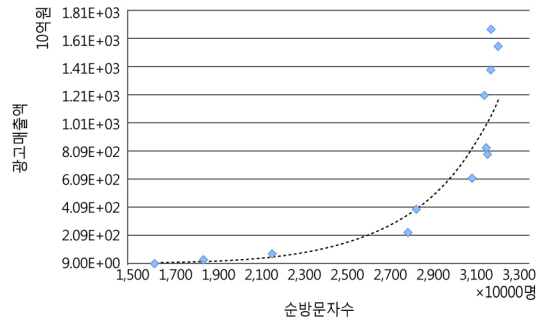
Theorem 2에서 보듯이, 검색엔진의 기본 효용( $U^0$ )가 일정 수준을 넘을 경우에는, 선형 함수일 때에 더 많은 검색어에 대한 결과를 조정하게 된다. 그리고, 기본 효용이 낮고 조정으로 인한 효과가 일정 수준 이상일 경우에는 비용 계수의 크기에 따라서 최적 조정 비율의 상대적 크기가 결정된다.

### 3.4 불록(제곱) 형태의 수익 함수

검색엔진의 광고 수익은 사용자가 증가하면서 광고 단가가 높아질 경우에는 불록 함수 형태를 가질 수 있다. 즉, 수익함수가 불록 함수일 때는 사용자들의 기대효용이 증가할수록 검색엔진의 수익 증가는 증가함을 의미한다. 실제로 네이버의 순방문자수에 따른 매출액의 증가 추이를 보면 [그림 3]처럼 불록 함수의 형태를 띠는 것을 알 수 있다.

$g(x) = x^2$  형태의 간단한 불록 함수를 수익함수로 고려하면, 검색엔진의 이익은

$$f(\alpha) = (\alpha d + U^0)^2 - k\alpha^2 = (d^2 - k)\alpha^2 + 2dU^0\alpha + (U^0)^2.$$



[그림 3] 네이버의 순방문자 수와 광고매출액의 관계 (2001년~2013년)

만약  $d^2 \geq k$ 이면,  $f(\alpha)$ 는 볼록함수 또는 선형함수가 되어서  $\alpha^* = 1$ 이 된다.  $d^2 < k$ 이면,  $f(\alpha)$ 는 오목함수가 되어서  $f'(\alpha) = 0$ 을 만족하는 해를 찾아서 그 해가 1을 넘지 않는 지를 검토하면 된다. 그리하여, 아래와 같은 최적 조정 비율을 구할 수 있다.

### Proposition 3

검색결과와 최적 조정 확률,  $\alpha^*$ 는 다음과 같이 결정 된다.

$$\alpha^* = \begin{cases} 1 & \text{if } k \leq d(d + U^0); \\ \frac{du^0}{k - d^2} & \text{otherwise.} \end{cases}$$

Proposition 3의 결과를 Proposition 1에서 제시된 수익함수가 선형일 때의 최적 조정 비율과 비교하여, 다음의 Theorem 3를 얻을 수 있다. Theorem 3의 증명은 부록에서 제공된다.

### Theorem 3.

$d = E[U^m] - U^0 > 0$ 라고 가정하자. 수익함수가 선형일 때의 최적 조정 비율을  $\alpha_i^*$ , 제곱 형태일 때의 최적 조정 비율을  $\alpha_c^*$ 라고 하자. 조정 비용 계수인  $k$ 가 매우

1) 네이버의 순방문자 수는 wikipedia의 코리안 클릭 자료를 인용하였으며([http://ko.wikipedia.org/wiki/대한민국의\\_포털\\_사이트](http://ko.wikipedia.org/wiki/대한민국의_포털_사이트)), 광고매출액은 네이버의 IR 사이트(<http://www.navercorp.com/ko/ir/businessReport.nhn>)에서 각 연도별 사업보고서를 참조하였다.

작다면 (즉,  $k \leq \min\left(\frac{d}{2}, d(d+U^0)\right)$ ),  $\alpha_v^* = \alpha_i^* = 1$ . 만약,  $k > \min\left(\frac{d}{2}, d(d+U^0)\right)$  라면, 두 조정 비율의 차이는 조정이 없는 경우에 검색자가 느끼는 효용( $U^0$ )과 조정이 있을 경우에 검색자가 얻는 평균 효용( $E[U^m]$ ), 그리고 조정 비용 계수  $k$ 의 정도에 의하여 다음과 같이 결정된다.

$$\text{If } E[U^m] < \frac{1}{2}, \alpha_v^* < \alpha_i^*;$$

$$\text{If } U^0 \geq \frac{1}{2} (E[U^m] \geq \frac{1}{2}), \alpha_i^* < \alpha_v^*;$$

$$\text{If } E[U^m] < \frac{1}{2}, U^0 < \frac{1}{2}, \text{ and}$$

$$k < \frac{d^2}{1-2U^0}, \alpha_i^* < \alpha_v^*;$$

$$\text{If } E[U^m] \geq \frac{1}{2}, U^0 < \frac{1}{2}, \text{ and}$$

$$k \geq \frac{d^2}{1-2U^0}, \alpha_v^* < \alpha_i^*$$

Theorem 2에서와 유사하게, 검색엔진의 기본 효용인  $U^0$ 가 0.5보다 크면 수익함수가 볼록할 때, 더 큰 조정 비율을 갖게 된다. 그리고, 조정으로 인한 효용의 증가가 일정 수준 이상일 경우에는 조정 비용 계수의 크기에 따라서, 최적 조정 비율의 비교 결과가 달라지게 된다.

이번에는 Proposition 2에 나와 있는 제공근 수익함수형태에 따른 최적 조정 비율과 비교하여, 다음의 Theorem 4을 얻을 수 있다.

#### Theorem 4.

$d = E[U^m] - U^0 > 0$ 라고 가정하자. 수익함수가 제공근일 때의 최적 조정 비율을  $\alpha_c^*$ , 제공 형태일 때의 최적 조정 비율을  $\alpha_v^*$ 라고 하자. 조정 비용 계수인  $k$ 가 매우 작다면 (즉,  $k \leq \min\left(\frac{d}{4\sqrt{d+U^0}}, d(d+U^0)\right)$ ),  $\alpha_v^* = \alpha_c^* = 1$ . 만약,  $k > \min\left(\frac{d}{4\sqrt{d+U^0}}, d(d+U^0)\right)$  라면, 두 조정 비율의 차이는 조정이 없는 경우에 검색자가

느끼는 효용( $U^0$ )과 조정이 있을 경우에 검색자가 얻는 평균 효용( $E[U^m]$ ), 그리고 조정 비용 계수  $k$ 의 정도에 의하여 다음과 같이 결정된다.

$$\text{If } E[U^m] < \left(\frac{1}{4}\right)^{\frac{2}{3}}, \alpha_v^* < \alpha_c^*;$$

$$\text{If } U^0 \geq \left(\frac{1}{4}\right)^{\frac{2}{3}} (\therefore E[U^m] \geq \left(\frac{1}{4}\right)^{\frac{2}{3}}), \alpha_c^* < \alpha_v^*;$$

$$\text{If } E[U^m] \geq \left(\frac{1}{4}\right)^{\frac{2}{3}}, U^0 < \left(\frac{1}{4}\right)^{\frac{2}{3}}, \text{ and}$$

$$k < \frac{d^2}{1-4^{\frac{2}{3}}U^0}, \alpha_c^* < \alpha_v^*;$$

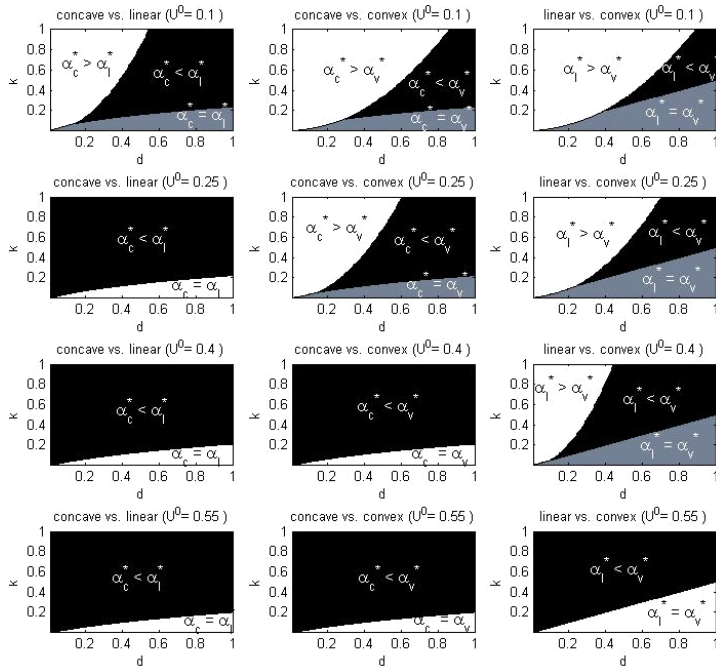
$$\text{If } E[U^m] \geq \left(\frac{1}{4}\right)^{\frac{2}{3}}, U^0 < \left(\frac{1}{4}\right)^{\frac{2}{3}}, \text{ and}$$

$$k \geq \frac{d^2}{1-4^{\frac{2}{3}}U^0}, \alpha_v^* < \alpha_c^*;$$

앞서 보아왔던 Theorem 2와 3의 결과 형태와 유사한 결과를 Theorem 4에서 발견할 수 있다. Theorem 4의 증명은 부록에서 제공된다.

지금까지의 결과를 그래프로 표현하면 아래 [그림 4]와 같은 결과를 얻을 수 있다. [그림 4]는 검색엔진의 기본 효용  $U^0$ 의 값이 증가할 때, 세 가지 다른 수익함수하의 최적 조정 비율을 비교한 것이다.  $U^0$  값이 작을 때에는 세 영역이 존재한다. 조정으로 인한 효용의 증가가 조정 시 필요한 비용에 비해 크지 않을 때에는 (왼쪽 가운데 및 윗부분), 수익함수의 한계 증가도가 감소할수록 더 큰 조정 비율을 갖게 된다(즉, 오목 > 선형 > 볼록). 조정비용이 조정으로 인한 효용의 증가에 비해 적을 때에는(오른쪽 아랫부분), 수익함수의 형태에 관계없이 같은 조정비율을 가지게 된다. 조정 비용과 조정으로 인한 효과가 비례하는 영역에서는, 수익함수의 한계 증가도가 증가할수록 더 큰 조정 비율을 갖게 된다(즉, 볼록 > 선형 > 오목). 그리고,  $U^0$ 의 값이 증가하면서 언급한 첫 번째 영역이 점점 줄어들면서, 세 번째 영역이 커지게 된다. 즉, 검색엔진의 기본 효용이 높다면, 수익함수의 한계 증가도가 증가하는 볼록 함수의 경우에 가





[그림 4] 세 가지 수익함수 형태에 따른 최적 조정 비율 비교( $k$ : 조정 비용 계수,  $d = E[U^m] - U^0$ )

장 큰 조정 비율을 갖게 된다. 이 결과는 검색엔진의 수익이 사용자의 증가에 따라 어떤 함수 형태를 띠는 지에 따라서, 조정 비율을 증가 또는 감소시킬 수 있음을 보여 주는 것이다.

#### 4. 결론 및 향후 연구 방향

본 연구는 검색엔진의 수익함수가 변화할 때 검색 결과의 수동적인 조정 비율을 증가 또는 감소시켜야 하는 지에 대한 가이드라인을 제시한다. 먼저, 최적의 조정 확률(또는 비율)을 수익함수가 일반적인 경우에 해를 구할 수 있는 조건을 제시한다. 그리고, 수익 함수가 선형, 오목, 그리고 볼록 함수 형태일 때 최적 조정 비율이 어떻게 결정되고 변화하는 지를 분석한다. 검색엔진이 제공하는 기본적인 효용(조정이 없는 경우)이 클 경우에는, 수익함수가 더 큰 한계 증가율을 가진 경우에 더 많은 조정을 하게 된다. 검색엔진의 기본적인 효용이 작고 조정된 검색어를 통한 효용의 증가가 일정 수준 이상일 때에는 조정 비용

계수의 수준에 따라서 최적 조정 비율의 크기가 달라지게 된다. 만약, 조정 비용 계수가 크다면 한계 증가율이 작을수록 더 많은 조정을 하게 된다. 조정 비용 계수가 작다면 한계 증가율이 클수록 더 많은 조정을 하게 된다. 이는 검색엔진의 수익이 사용자의 수가 증가함에 따라 어떤 형태를 보여주는 지에 따라 다른 최적의 조정 비율이 존재함을 의미한다. 가령, 성장기에는 수익 함수가 선형형태를 가지다가 성숙기에는 오목함수가 되고, 기본적인 검색엔진의 효용이 크다면, 최적의 조정 비율을 줄이는 방향으로 가는 것이 검색엔진의 이익을 최대화 시킬 수 있게 된다.

향후 연구방향으로는 첫째, 두 개의 검색엔진이 서로 경쟁할 경우의 최적 조정 비율을 구해볼 수 있다. 즉, 두 경쟁중인 검색엔진에 대해 최대의 이익을 얻을 수 있는 검색엔진의 최적 조정비율의 균형에 대해 연구해볼 수 있다. 둘째, 본 논문을 확장하여 특화된 서비스제공과 관련된 연구를 해 볼 수 있다. 전체 고객 안에는 특화된 서비스를 좋아하는 사람과 좋아하지 않는 사람이 일정 비율로 있을 수

있으며, 서비스 제공자는 고객의 만족도를 높이기 위해 특화된 서비스를 제공할 수 있다. 이와 같은 경우에 고객의 유형에 따라 서비스 제공자의 최적 조정 비율을 분석해 볼 수 있다.

## 참 고 문 헌

- [1] 장대철, 정영조, 안병훈, “양면시장에서의 진입가능성 연구”, 『한국경영과학회지』, 제31권, 제4호(2006), pp.105-123.
- [2] 김도훈, “양면시장형 컨버전스 산업생태계에서 플랫폼 경쟁에 관한 진화게임 모형”, 『한국경영과학회지』, 제35권, 제4호(2010), pp.55-79.
- [3] Argenton, C. and J. Prüfer, “Search Engine Competition with network externalities,” *Journal of competition Law and Economics*, Vol.8, No.1(2012), pp.73-105.
- [4] Armstrong, M., “Competition in two sided markets,” *The RAND Journal of Economics*, Vol.37, No.3(2006), pp.668-691.
- [5] Clarke, L.A., C.M. Kolla, G.V. Cormack, O. Vechtomova, A. Ashkan, S. Büttcher, and I. MacKinnon, “Novelty and Diversity in Information Retrieval Evaluation,” *SIGIR '08 Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, (2008), pp.659-666.
- [6] Edosomwan, J. and T.O. Edosomwan, “Comparative analysis of some search engines,” *South African Journal of Science*, Vol.106, No.11-12(2010), pp.1-4.
- [7] Hawking, D., N. Craswell, et al., “Measuring search engine quality,” *Information Retrieval*, Vol.4, No.1(2001), pp.33-59.
- [8] Kim, K. and E. Tse, “Dynamic search engine competition with a knowledge-sharing service,” *Decision Support systems*, Vol.52, No.2(2012), pp.427-437.
- [9] Lewandowski, D., “Web searching, search engines and Information Retrieval,” *Information Services and Use*, Vol.25, No.3(2005), pp.137-147.
- [10] Lewandowski, D., “A three-year study on the freshness of web search engine databases,” *Journal of Information Science*, Vol.34, No.6(2008), p.817.
- [11] Lewandowski, D. and N. Hochstotter, “Web searching : A quality measurement perspective,” *Web Search*, (2008), pp.309-340.
- [12] Taylor, G., “Search Quality and Revenue Cannibalisation by Competing Search Engines,” *Journal of Economics and Management Strategy*, Vol.22, No.3(2013), pp.445-467.
- [13] Telang, R., U. Rajan, and T. Mukhopadhyay, “The Market Structure for Internet Search Engines,” *Journal of Management Information Systems*, Vol.21, No.2(2004), pp.137-160.
- [14] Van Couvering, E., “Is relevance relevant? Market, science, and war : Discourses of search engine quality,” *Journal of Computer Mediated Communication*, Vol.12, No.3(2007), pp.866-887.
- [15] Vaughan, L., “New measurements for search engine evaluation proposed and tested,” *Information Processing and Management*, Vol.40, No.4(2004), pp.677-691.
- [16] White, A., “Search Engines : Left Side Quality versus Right Side Profits,” *International Journal of Industrial Organization*, Vol.31(2013), pp.690-701.
- [17] Xie, M., H. Wang, et al., “Quality dimensions of Internet search engines,” *Journal of Information Science*, Vol.24, No.5(1998), p.365.

## 〈부 록〉

### A. Theorem 1의 증명

$d^2g''(d\alpha + U^0) < 2k$ 이 항상 성립하면, 이익함수가 오목함수의 형태이므로 일차 조건을 만족하는 조정 비율을 찾은 후, 그 값이 1보다 작거나 같으면 최적의 조정 비율이 된다. 그렇지 않으면, 1이 최적 조정 비율이 된다.  $d^2g''(d\alpha + U^0) \geq 2k$ 이 항상 성립하면, 이익함수가 볼록함수(weakly convex)이므로 0과 1 중에서 더 큰 이익을 가져다주는 값이 최적의 조정 비율이 된다.

### B. Theorem 2의 증명

우선,  $E[U^m] < \frac{1}{4}$  일 때를 고려하면,  $\frac{d}{2} < \frac{d}{4\sqrt{d+U^0}}$  이므로  $k \leq \frac{d}{2}$  일 때는  $\alpha_i^* = \alpha_c^* = 1$ ;  $\frac{d}{2} < k \leq \frac{d}{4\sqrt{d+U^0}}$  이면  $\alpha_i^* < \alpha_c^* = 1$ ;  $k > \frac{d}{4\sqrt{d+U^0}}$  이면 식 (1)에 나와 있는  $f'(\alpha) = \frac{d}{2\sqrt{d\alpha+U^0}} - 2k\alpha$ 는  $\alpha_i^*$ 에서 양수가 되어  $\alpha_i^* < \alpha_c^*$ 이 성립하게 된다.  $E[U^m] \geq \frac{1}{4}$  이면,  $k \leq \frac{d}{4\sqrt{d+U^0}}$  이면  $\alpha_i^* = \alpha_c^* = 1$ ;  $\frac{d}{4\sqrt{d+U^0}} < k \leq \frac{d}{2}$  이면  $\alpha_c^* < \alpha_i^* = 1$ ;  $k > \frac{d}{2}$  이고  $U^0 \geq \frac{1}{4}$  라면,  $f'(\alpha_i^*) < 0$ 이 성립하여  $\alpha_c^* > \alpha_i^*$ ;  $k > \frac{d}{2}$  이고  $U^0 < \frac{1}{4}$  라면  $\alpha_i^* = \frac{d}{2k}$ 을  $f'(\alpha)$ 에 대입하면,  $f'(\alpha_i^*) = \frac{d}{2\sqrt{\frac{d^2}{2k} + U^0}} - d$ 이므로  $\frac{d^2}{2k} + U^0 < \frac{1}{4}$  일 때는 양의 값을 가지게 되고,  $\alpha_c^* < \alpha_i^*$ 이 성립하게 된다. 그 반대의 경우에는  $\alpha_c^* \leq \alpha_i^*$ 이 성립한다.

### C. Theorem 3의 증명

먼저  $E[U^m] < \frac{1}{2}$  이라고 가정하자. 그러면,  $d(d+U^0) < \frac{d}{2}$  이므로  $k \leq d(d+U^0)$  일 때는  $\alpha_i^* = \alpha_v^* = 1$ ;  $d(d+U^0) < k \leq \frac{d}{2}$  이면  $\alpha_v^* < \alpha_i^* = 1$ ;  $k > \frac{d}{2}$  이면  $\alpha_v^* < \alpha_i^* \Leftrightarrow d^2 < k(1-2U^0)$ 와  $\frac{d}{1-2U^0} < \frac{1}{2}$ 에서  $\alpha_v^* < \alpha_i^*$ 이 성립함을 알 수 있다. 이번에는  $E[U^m] \geq \frac{1}{2}$  라고 가정하자.  $k \leq \frac{d}{2}$  이면  $\alpha_i^* = \alpha_v^* = 1$ ,  $\frac{d}{2} < k \leq d(d+U^0)$  이면  $\alpha_i^* < \alpha_v^* = 1$ .  $k > d(d+U^0)$  이고  $U^0 \geq \frac{1}{2}$  라면,  $\alpha_v^* < \alpha_i^* \Leftrightarrow d^2 > k(1-2U^0)$ 에서  $\alpha_i^* < \alpha_v^*$ 임을 알 수 있다.  $k > d(d+U^0)$  이고  $U^0 < \frac{1}{2}$  라면,  $k \leq \frac{d^2}{1-2U^0}$  이면  $\alpha_i^* \leq \alpha_v^*$ ;  $k > \frac{d^2}{1-2U^0}$  이면  $\alpha_v^* < \alpha_i^*$ 임을 쉽게 알 수 있다.

### D. Theorem 4의 증명

먼저  $E[U^m] < \left(\frac{1}{4}\right)^{\frac{2}{3}}$  이라고 가정하자. 그러면,  $d(d+U^0) < \frac{d}{4\sqrt{d+U^0}}$  이므로  $k \leq d(d+U^0)$  일 때는  $\alpha_c^* = \alpha_v^* = 1$ ;  $d(d+U^0) < k \leq \frac{d}{4\sqrt{d+U^0}}$  이면  $\alpha_v^* < \alpha_c^* = 1$ ;  $k > \frac{d}{4\sqrt{d+U^0}}$  이면  $f'(\alpha)$ 가  $\alpha_v^*$ 에서 양수이므로  $\alpha_v^* < \alpha_c^*$ 이 성립함을 알 수 있다. 이번에는  $E[U^m] \geq \left(\frac{1}{4}\right)^{\frac{2}{3}}$  라고 가정하자.  $k \leq \frac{d}{4\sqrt{d+U^0}}$  이면  $\alpha_i^* = \alpha_v^* = 1$ .  $\frac{d}{4\sqrt{d+U^0}} < k \leq d(d+U^0)$ 이

면  $\alpha_c^* < \alpha_v^* = 1$ .  $k > d(d + U^0)$ 이고  $U^0 \geq \left(\frac{1}{4}\right)^{\frac{2}{3}}$  라면,  $f'(\alpha_v^*) < 0$ 이므로  $\alpha_c^* < \alpha_v^*$ 임을 알 수 있다.  $k > d(d + U^0)$ 이고  $U^0 < \left(\frac{1}{4}\right)^{\frac{2}{3}}$  라면,  $k < \frac{d^2}{1 - 4^{\frac{2}{3}} U^0}$  이면  $f'(\alpha_v^*) < 0$ 이므로  $\alpha_c^* < \alpha_v^*$ ;  $k \geq \frac{d^2}{1 - 4^{\frac{2}{3}} U^0}$  이면  $f'(\alpha_v^*) \geq 0$ 이므로,  $\alpha_v^* \leq \alpha_c^*$ 가 성립한다.