

온라인 상품 카테고리 내 주요 가격대 식별

Identifying the Main Price Ranges of Online Product Category

김준우*, 임광혁**

동아대학교 산업경영공학과*, 배재대학교 전자상거래학과**

Jun Woo Kim(kjunwoo@dau.ac.kr)*, Kwang Hyuk Im(khim@pcu.ac.kr)**

요약

최근 많은 소비자들이 관심 있는 물품 카테고리에 대한 정보를 얻기 위한 목적으로 종합 쇼핑몰이나 가격 비교 사이트를 방문하고 있다. 하지만, 이러한 웹 사이트들은 종종 이들에게 많은 상품들과 판매자가 포함된 지나치게 방대한 정보를 제공하여 소비자들의 구매 결정을 효과적으로 지원하지 못한다. 따라서 현대 온라인 쇼핑 에이전트들은 검색된 정보를 사용자에게 제공하기 전에 보다 지능적인 방법으로 이를 가공할 필요가 있다. 본 논문은 특정 물품 카테고리 내에서 많은 상품들이 분포하고 있는 주요 가격대를 식별하는 방법을 제안하고자 한다. 이를 위해 한 개 카테고리 내 상품의 가격들을 벡터로 표현하고, 여기에 k-means 군집 분석을 적용하여 서로 비슷한 가격 벡터들을 포함하는 군집을 형성한 다음, 각 군집에서 주요 가격대를 추출하는 방법을 적용하였다. 일반적으로 가격은 소비자들의 구매 결정에서 가장 중요한 요인 중 하나이기 때문에, 추출된 주요 가격대들은 온라인 쇼핑 이용자들이 효과적으로 상품을 검색하는데 도움이 될 것으로 기대된다.

■ 중심어 : | 온라인 쇼핑 | 가격대 | 쇼핑 에이전트 | 상품 카테고리 | 군집 분석 |

Abstract

In recent, many consumers visit the online shopping malls or price comparison sites to collect the information on the product category that they are interested in. However, the volumes of the data provided by such web sites are often too enormous, and significant number of consumers have trouble in making purchase decision based on the plethora of products and sellers. In this context, modern online shopping agents need to process the retrieved information in more intelligent way before providing them to the users. This paper proposes a novel approach for identifying the main price ranges hidden in a single product category. To this end, the price of an item in the category is represented as a row vector and k-means clustering analysis is applied to the price vectors to produce the clusters that consists of the product items with similar price vectors. Then, the main price ranges of the product category can be identified from the result of clustering analysis. In general, the price is one of the most important factors in the consumers' purchase decision, and the identified main price ranges will be helpful for the online shoppers to find appropriate items effectively.

■ keyword : | Online Shopping | Price Range | Shopping Agent | Product Category | Clustering Analysis |

* 이 논문은 동아대학교 교내연구비 지원에 의하여 연구되었음

접수번호 : #120913-008

접수일자 : 2012년 09월 13일

심사완료일 : 2012년 11월 12일

교신저자 : 임광혁, e-mail : khim@pcu.ac.kr

I. 서론

이용의 편의성, 저렴한 가격 및 다양한 제품 정보에 대한 높은 접근성 등의 장점으로 인하여 온라인 쇼핑이 활성화된 지 오래이다[11][15][16]. 이에 따라 인터넷을 통해 일반 소비자들에게 물품을 판매하는 다양한 온라인 소매상들이 출현하였고, 최근에는 이들이 취급하는 물품들에 대한 정보를 편리하게 취합하여 조회할 수 있는 통합 온라인 쇼핑몰이나, shopbot 또는 비교 쇼핑 에이전트(comparative shopping agent)를 활용한 가격 비교 사이트를 통해 구매 전 관심 물품에 대한 정보를 수집하는 경우도 많아졌다[10].

하지만 온라인에서 거래되는 물품의 종류와 온라인 판매자 수의 폭발적인 증가로 인하여 가격 비교 사이트에서 검색되는 정보의 양도 크게 증가하였다. 더구나, 동일 상품에 대해서도 다양한 온라인 판매자들이 서로 다른 가격을 제시하는 가격 분산성(dispersity)은 소비자들의 구매 의사 결정을 보다 복잡하게 만드는 경우가 많다[2][3]. 결과적으로 구매할 물품에 대한 사전 지식이 충분하지 않은 소비자들의 경우, 종종 인터넷을 통해 수집한 정보를 이용하여 구매할 상품을 구체적으로 정하는데 어려움을 느끼는 경우가 발생하고 있다[14]. 따라서 인터넷에서 수집된 방대한 물품 정보들을 효과적으로 가공하여 일반 소비자들이 편리하게 활용할 수 있도록 하는 방법들이 필요하다[8].

본 논문에서는 특정 물품 카테고리 내 상품들의 주요 가격대를 탐색하는 방법을 제안하고자 한다. 주요 가격대란 특정 제품 카테고리 내 많은 상품들의 가격이 분포하는 가격 범위를 의미하는 것으로, 이러한 정보는 특정 물품 카테고리에 관심을 가진 소비자가 카테고리 내 상품 검색 범위를 좁히는데 도움이 될 것으로 생각된다.

일례로 노트북 컴퓨터를 구매하려고 하나, 이 물품 카테고리에 대한 사전 지식이 없는 소비자를 가정해보자. 인터넷을 통하여 관련 정보를 수집하는 경우, 이 소비자는 옥션[1]이나 지마켓[9] 등과 같은 통합 온라인 쇼핑몰이나 다나와[4] 같은 가격 비교 사이트 또는 일반 포털 사이트의 쇼핑 정보를 참조하게 될 것이다.

하지만 보통 하나의 물품 카테고리 정보를 요청할 경우, 수십~수백 가지 종류의 상품이 검색되며, 한 개 상품에 대해서 다시 많은 판매자들이 서로 다른 가격을 제시하고 있어, 이러한 방대한 정보를 개인 소비자가 일일이 분석하기 어렵다. 특히, 일반적으로 구매 결정에 있어 가격이 중요한 요인이 되는데, 사전 지식이 부족한 개인 소비자의 경우, 위 정보만 가지고 어느 정도 예산 범위의 상품에 초점을 맞추어야 할지 결정하는 것이 곤란하다.

이러한 점을 감안하여, 현행 온라인 쇼핑몰들은 일반적으로 각 상품의 최저가를 방문자들에게 제시하고, 최저가 기준 정렬 기능을 제공한다. 나아가, 물품 카테고리 내 상품의 가격대를 등간격으로 이산화하여 특정 구간 선택 시, 최저가가 해당 구간에 속하는 상품들을 제시하기도 한다. 하지만 이러한 방법의 경우, 동일 상품의 가격도 넓은 구간에 분포할 수 있어 최저가 기준으로 하위 구간에서 검색된 상품이라도 실제로는 그보다 상위 구간 가격으로 구매할 수 있는 경우가 빈번하다. 특히 상품의 최저가는 종종 특정 프로모션을 통해서만 얻어지거나, 제한된 사양의 제품인 경우가 많다. 나아가, 기존의 최저가 기준 가격 이산화를 사용하는 경우, 소비자들은 현재 물품 카테고리 내 상품들이 주로 어떤 가격대에 분포하는지 알기 어렵고, 해당 카테고리 내에서 비교적 저렴한 상품이나 고가의 상품들을 구매하기 위한 예산 규모를 파악하기 곤란하다.

반면, 예를 들어 노트북 컴퓨터가 주로 30~50만원, 60~70만원, 90~110만원의 주요 가격대에 많이 분포한다고 가정하고 이러한 점을 알려주면 소비자는 주요 가격대 중 하나를 선택하여, 가격 분포가 비슷한 상품의 그룹을 함께 조회할 수 있고, 해당 카테고리 내 상품들이 대강 어떤 가격대에 분포하는지 파악이 용이할 것이다.

본 논문은 2장에서 현대 온라인 쇼핑 환경 및 쇼핑 에이전트에 대한 기존 문헌들에 대하여 간단히 소개하고, 3장에서 동일 카테고리 내 상품 정보를 다루는 방법과 함께, 주요 가격대를 추출하는 방법에 대해 설명한다. 이 과정에서 개별 상품의 분산된 가격을 벡터로 표시하고, 이 벡터들에 k-means 군집을 적용하며, 주요 가격대 추출 결과를 평가하기 위한 지표를 제안한다. 이어

4장에서는 실제 국내 가격 비교 사이트에서 수집한 울트라씬(ultra-thin) 노트북 카테고리에 이러한 방법을 적용한 결과에 대해 살펴보고, 끝으로 5장에서 결론 및 향후 연구 과제를 제시한다.

II. 연구 배경

온라인 쇼핑이 활성화되고 이를 통해 거래되는 상품이 늘어나면서 개인 소비자의 온라인 쇼핑을 효과적으로 지원할 수 있는 가격 비교 사이트와 같은 쇼핑 에이전트에 대한 연구가 활발히 이루어지고 있다[12]. 초창기에는 온라인 쇼핑을 통해 정보 공유가 활성화되어 한 제품의 가격이 통일되는 일물일가 현상이 나타날 것이라는 예측이 있었다[7]. 그러나 현실적으로는 여러 가지 요인들로 인하여 오히려 예전보다 동일 제품에 대하여 판매자마다 서로 다른 가격을 제시하는 가격 분산성이 널리 나타났고, 이는 개인 소비자들의 정보 검색을 혼란스럽게 만드는 측면이 있다[2][3][7].

더구나 동일 물품 카테고리 안에서 비슷한 기능과 사양의 상품들이 거래되면서 방대한 상품 및 판매자들에 대한 정보를 개인 소비자가 자세히 확인하여 활용하는 것이 대단히 어려워져, 가격 비교 사이트를 비롯한 온라인 쇼핑 에이전트들로서는 고객의 구매 의사 결정 과정을 보다 효과적으로 지원할 수 있는 부가적인 기능이 요구된다[8].

이에 따라 Shopbot 또는 비교 쇼핑 에이전트로도 불리는 이러한 현대 가격 비교 사이트들의 부가 기능에 대한 연구들이 일부 이루어져왔고, 이러한 예로 상품/판매자 정보와 개별 고객의 선호도/행태 정보 사이의 관계에 강화 학습(reinforcement learning)을 적용하여 개인화된 추천 상품 랭킹을 생성하는 방법[6], 특정 물품 구매 선택 시, 최대의 가격 절감 효과를 거둘 수 있는 부가 구매 물품 추천 시스템[5][6] 등에 대한 연구들이 수행되어 왔다.

하지만 이러한 기존 연구들은 주로 소비자가 해당 사이트에 가입하여 선호도나 행태에 관한 정보가 알려져 있거나, 구체적으로 구매할 상품을 결정한 상태임을 가

정한다. 반면, 최근 가격 비교 사이트에서는 가입이나 로그인 없이 특정 물품 카테고리 전체에 대한 일반적인 정보를 수집하고자 하는 사용자들의 접속 빈도도 높으며, 오픈 마켓 등의 등장으로 인하여 가격 비교 서비스들은 구매자와 판매자 사이의 단순 중개 역할만 하는 경우가 많다. 따라서 대상 물품 카테고리에 대한 사전 지식이 충분하지 않아 아직 구체적인 구매 결정을 하지 못하는 익명의 소비자들에게 보다 효과적인 정보를 취합해 줄 수 있는 방법이 필요하다.

본 논문에서는 하나의 물품 카테고리 내에서 많은 물품들이 관찰되는 주요 가격대를 추출하여 소비자들에게 제시함으로써 소비자들이 예산을 조정하고 구체적으로 검색 대상 상품을 좁혀나가는 데 도움을 주기 위한 방법을 제안하고자 한다. 이러한 목적을 위하여 카테고리 내 각 물품의 가격 분포 정보를 적절한 형태로 나타내고, 대표적인 데이터마이닝 기법 중의 하나인 k-means 군집 방법[13]이 적용된다.

III. 동일 카테고리 내 주요 가격대 추출

1. 동일 물품 카테고리 내 상품 정보

어떤 물품을 구매하고자 하나, 그 물품에 대한 충분한 사전 지식이 없는 소비자들은 전통적으로 오프라인 매장 방문이나 지인들의 입소문을 통해 관련 정보를 수집하는 경우가 많았다. 반면, 최근에는 인터넷을 통해 여러 가지 상품들의 사용 후기나 가격대, 사양 등을 참고하는 경우가 늘었다. 하지만 동일 물품 카테고리 내에도 상품의 종류와 판매자가 많아지면서 어느 정도 사전 지식 없이는 상품 관련 정보 검색 결과를 이용하는 데 곤란을 겪는 경우도 많다. 예를 들어, 국내 유명 가격 비교 사이트인 다나와.com에서 울트라씬 노트북 컴퓨터 카테고리 상품들을 검색할 경우, 2011년 7월 기준으로 총 95개의 상품이 조회된다. 이들은 기본적으로 인기도 순으로 정렬되어 조회되고, 사용자의 설정에 따라 최저가 낮은 순 또는 높은 순 등의 정렬도 가능하다. 목록에서 상품 하나를 클릭할 경우에는 그 상품의 판매자 및 제시 가격 목록이 조회되고, 이러한 상품별 정보

들은 [표 1]에 요약되어 있다.

[표 1]에서 상품에 따라 수십 명의 판매자에 의해 취급되고, 동일 상품에 대하여 서로 다른 판매자가 제시하는 가격의 차이가 많게는 수십만 원에 이른다는 것을 볼 수 있다. 따라서 이러한 방대한 상품 관련 정보들은 때로 일반 소비자에게 혼란을 줄 수 있다.

조회되는 상품 개수를 줄이기 위하여 일반적으로 온라인 쇼핑 관련 웹사이트들은 사양이나 제조사, 가격 등의 기준으로 조회 상품을 필터링하는 기능을 제공한다. 하지만 이러한 기능을 효과적으로 이용하기 위해서는 소비자가 해당 카테고리 내 상품에 대하여 어느 정도의 사전 지식을 가지고 있어야 한다.

표 1. 울트라씬 노트북 카테고리 내 상품

ID	판매자 수	최저가(원)	최고가(원)	가격분포(원)
1	50	1,026,000	1,349,000	323,000
2	44	498,900	633,000	134,100
3	48	833,460	940,000	106,540
4	59	820,030	1,225,350	405,320
5	67	908,180	1,499,640	591,460
...
95	5	1,034,110	1,255,190	221,080
최소	1	378,900	445,550	0
최대	70	1,418,770	2,174,030	1,160,600
평균	27.5	810,495	1,079,922	269,428

특히, 가격 기준 검색의 경우 소비자가 직접 검색할 가격대를 입력해야 하는데 카테고리 내 상품들에 대한 사전 지식이 충분하지 않은 경우, 적절한 가격 범위 설정이 어려울 수 있다. 또한 가격 범위를 설정했다 하더라도 최저가만을 기준으로 하는 가격대 기준 필터링의 특성 상, 필터링 결과의 유용성이 떨어지기 쉽다.

예를 들어 소비자가 80~100만원 범위 상품을 조회해보고 싶은 경우, 최저가를 기준으로 [표 1]의 상품 1, 95는 최저 100만원 초반에 구매가 가능하여 이 소비자의 관심을 끌 가능성이 있음에도 필터링에서 제외된다. 반면, 상품 5의 경우에는 가격대가 150만원 가량의 고가 에까지 분포함에도 최저가를 기준으로 검색 결과에 포함된다. 이러한 점을 보완하기 위하여 최근 다나와.com 과 같은 가격 비교 사이트에서는 카테고리 내 상품 조회 시, [그림 1]의 하단과 같이 몇 개의 가격대를 제시하

여 소비자가 간편히 선택할 수 있도록 하고 있다.



그림 1. 가격 비교 사이트 제시 가격대

하지만 이러한 사이트 제시 가격대에는 여전히 최저가만을 이용한다는 단점이 존재하고, 가격대들을 어떻게 추출할 것인지에 대한 고려가 부족하며, 최저가가 하위 가격대에 속하는 상품이라도 실제로는 그보다 상위 가격대에서 취급되거나 하는 경우가 빈번하다. 이러한 맥락에서 본 논문에서는 한 개 물품 카테고리에 대하여 어떤 가격대를 추출하여 소비자에게 제시하는 것이 바람직한가에 대하여 논의해보고, 바람직한 주요 가격대 추출 방법을 제안하고자 한다.

2. 주요 가격대 추출 전략

현대 온라인 쇼핑 환경을 분석하는 데 있어 중요하게 고려해야 할 사항 중의 하나는 동일 상품도 여러 판매자에 의해 취급되고, 이들이 상이한 가격을 제시한다는 가격 분산성이다. 이러한 상품의 분산된 가격을 표현하기 위해 본 논문에서는 상품 1개의 가격 정보를 (1)과 같은 행벡터로 나타낸다. 이 행 벡터는 가격 구간을 n개로 분할한 후, 한 개의 상품 k의 가격 정보를 각 구간 i(i=1, 2, ..., n)에 속하는 판매자들의 상대도수 f_{ki} 로 나타낸다.

$$p_k = [f_{k1} f_{k2} f_{k3} \dots f_{kn}] \tag{1}$$

이러한 가격 벡터 형태의 분산 가격 정보를 갖는 상품 k를 어떤 가격대 $R = [r_{\min}, r_{\max}]$ 에 소속시켰다고 가정해보자. 가격 벡터 p_k 에서 0이 아닌 값을 갖는 원소 중 가장 낮은 구간의 시작값을 s_k , 가장 높은 구간의 종료값을 f_k 라고 할 때, 상품 k에 대한 가격대 R의 적

합성은 $s_k = r_{\min}$, $f_k = r_{\max}$ 일 때 가장 높다. 반대로 s_k, f_k 가 각각 r_{\min}, r_{\max} 에서 점점 멀어질수록 상품 k 는 가격대 R 로 특징짓기 부적합하다. 이 점을 반영하여 본 논문에서는 (2)와 같은 부적합도 지표를 통해 임의의 상품 k 에 대한 가격대 R 의 적합성을 측정하고자 한다.

$$U_R(k) = |s_k - r_{\min}| + |f_k - r_{\max}| \quad (2)$$

아울러, 총 N 가지 상품으로 구성된 물품 카테고리의 상품들의 가격대를 R_1, R_2, \dots, R_c 로 분류할 경우, 개별 상품 k 는 가장 많은 판매가격들이 속해 있는 가격대에 소속되고, 이러한 가격대들을 해당 물품 카테고리에 적용할 경우의 전체적인 부적합도를 (3)과 같이 모든 상품들의 부적합도 평균으로 계산한다. 단, $U_{rk}(k)$ 는 상품 k 가 소속된 가격대 R_k 에 대한 상품 k 의 부적합도이다.

$$U = \frac{\sum_{i=1}^k U_{rk}(i)}{N} \quad (3)$$

특정 물품 카테고리 상품들에 대하여 상품 가격대가 주어질 때, 이 가격대들이 얼마나 적합한지는 (3)의 전체 부적합도 U 로 평가할 수 있고, 이 값이 작을수록 바람직한 가격대로 상품들이 분류되고 있음을 의미한다.

3. 군집 분석을 통한 주요 가격대 추출

앞에서는 상품 가격대들이 주어지는 경우, 그에 대한 평가지표로 전체 부적합도를 제안하였다. 이제, 물품 카테고리 내에서 상품 가격대의 개수 및 각각의 범위를 정하는 방법이 필요하다. 전체 부적합도가 낮으려면, 가격 분포가 서로 비슷한 상품들이 함께 같은 가격대에 소속되어야 하기 때문에 본 논문에서는 (2)와 같이 추출한 가격 벡터에 군집 분석을 적용한다. 일반적으로 단일 상품에 대해 서로 다른 판매자가 상이한 가격을 제시하더라도 그 차이에는 어느 정도 한계가 있고, 판매자 별 제시 가격에 큰 차이가 있는 상품들이라면, 사

양에 중대한 차이가 있어 서로 다른 상품으로 보는 것이 타당할 것이다. 이에 따라 일반적으로는 단일 상품에 대한 가격은 가격 분할 구간 중, 인접한 몇 개 구간에 몰려있을 개연성이 크며, 군집 분석을 적용하여 서로 유사한 가격 벡터를 군집화할 경우, 가격의 분포 구간이 비슷한 상품들이 같은 군집을 형성하게 되고, 이러한 군집의 중심점(centroid)을 계산할 경우, 해당 군집 내 상품들이 일반적으로 분포하는 인접한 가격 구간들을 식별할 수 있다. 본 논문에서는 이렇게 식별된 가격 구간들을 해당 군집 소속 상품들의 주요 가격대로 제시한다.

실제 분석에는 대표적인 군집 분석 알고리즘 중 k-means 알고리즘을 사용하였고, 가격 벡터들은 군집 알고리즘을 적용하기 전에 한 차례 정규화 과정을 거치도록 하였다. 정규화는 가격 벡터에 포함된 원소들의 값을 카테고리 내 상품들의 해당 원소 최대값으로 나누어 실행하고, 이 작업은 특정 몇몇 구간에 대한 벡터의 원소값들이 군집 과정에서 지나치게 큰 영향력을 발휘하는 것을 제어하기 위한 절차이다.

무감독 군집 방법인 k-means 알고리즘은 일반적으로 군집의 개수 k 를 사전에 결정하는 것이 필요하다. 그러나 대상 데이터의 구조에 대한 상세한 지식이 없는 상태에서 군집의 개수를 정하는 것은 다소 까다롭다. 이에, 반복적인 과정을 통해 k 를 정하는 것이 일반적이다. 단순한 방법으로는 k 를 2, 3, 4, ... 로 늘려가면서 군집 분석 결과에 대한 평가 지표인 오차제곱합의 감소 추세가 느려지는 시점의 k 를 선택하는 것 등이 있다.

이와 유사하게, 한 물품 카테고리 내 주요 가격대를 추출하기 위해 k-means 알고리즘을 적용할 때도 몇 개의 가격대를 식별할 것인지가 문제가 될 수 있다. 앞에서 언급하였듯이, 물품 카테고리 내 상품들의 정규화 가격 벡터에서 0이 아닌 원소들의 분포가 서로 인접한 위치에서 관찰되는 경우에는 이 문제는 간단히 해결된다. 그 이유는 k 값이 작을 때는 각 정규화 벡터 군집의 중심점 벡터들의 분포 구간이 상당히 넓지만, k 값을 늘려갈수록 결국 중심점 벡터들은 인접한 몇 개의 구간에서 0이 아닌 원소들을 갖게 되기 때문이다. 따라서 이러한 군집들이 관찰되는 시점에서의 k 값을 선택할 수 있다.

반면, 실제 한 개 물품 카테고리 내에는 가격의 분포가 지극히 넓은 상품들이 상당수 존재하는데, 이러한 경우들은 외형상으로는 동일한 상품이나 사양이나 구매 조건 또는 제공 혜택의 차이 등으로 인해 서로 다른 상품으로 취급하는 것이 더 바람직한 경우라고 생각된다. 따라서, 본 논문에서는 가격 분포가 넓은 상품들은 제외하고, 어느 정도 밀집된 가격 분포를 갖는 상품들에 대해서만 분석을 실시하고자 한다.

IV. 울트라씬 노트북 카테고리 적용 결과

[표 1]의 울트라씬 노트북 상품들의 가격 정보를 나타내기 위하여 가격을 5만원 단위로 분할하여 [0, 45만원), [45만원, 50만원), ..., [130만원, 135만원), [135만원 이상)의 20개 구간으로 나누는 경우, 카테고리 내 상품들의 가격 정보는 [표 2]와 같은 행벡터로 표현된다.

표 2. 울트라씬 노트북 카테고리 내 상품들의 가격 벡터

ID	분산 가격 정보
1	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.32 0.36 0.18 0.10 0.00 0.00 0.04 0.00]
2	[0.00 0.07 0.52 0.30 0.11 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00]
3	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.33 0.48 0.19 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00]
4	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.22 0.10 0.07 0.14 0.17 0.14 0.14 0.02 0.02 0.00 0.00]
5	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.10 0.01 0.34 0.19 0.06 0.04 0.12 0.03 0.03]
...	...
95	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.60 0.00 0.00 0.00 0.20 0.20 0.00 0.00]

이러한 가격 벡터들을 관찰한 결과, 상품들의 가격 분산 형태는 대단히 다양하였다. 표현의 편의성을 위해 단일 상품의 가격 벡터에서 0이 아닌 원소를 갖는 구간 중, 가장 낮은 구간을 $\min(i)$, 가장 높은 구간을 $\max(i)$ 라고 하자. 이 때, 가격 분포 범위의 길이 $\max(i) - \min(i) + 1$ 을 95개 상품 각각에 대하여 계산한 결과는 표 3과 같고, 동일 카테고리 물품들 중에서도 1개 구간에만 가격이 분포하는 상품부터 20개 구간 중, 10개 이상

의 구간에 걸쳐 가격이 분포하는 상품들까지 가격 분산의 정도가 대단히 다양함을 알 수 있다.

앞 장에서 설명하였듯이, 분포 길이가 너무 길거나 반대로 짧은 경우에는 주요 가격대를 추출하는데 적합하지 못하다. 그 이유는 분포 길이가 짧은 경우, 특정 구간만을 너무 강조할 수 있으며, 분포 길이가 긴 상품이 분석에 포함될 경우에는 만들어진 군집의 중심점 벡터 역시 그 분포 범위가 지나치게 넓을 수 있기 때문이다. 따라서 본 논문에서는 분포 범위 길이가 평균값 5.56과 인접한 3~6에 걸쳐 있는 42개 상품들의 가격 벡터를 이용하기로 한다. 참고로 가격 분포가 좁은 상품들의 경우에는 추후 추출된 주요 가격대에 적당히 분배하고, 가격 분포가 넓은 상품들은 두 개 이상의 상품으로 나누는 것을 고려해볼 수 있다.

표 3. 상품들의 가격 분산

분포 범위 길이	상품 개수
1	10
2	8
3	12
4	9
5	9
6	12
7	10
8	4
9	9
10이상	12
평균	5.56

선택된 상품들의 가격 벡터 원소들의 값을 각각 해당 구간에서의 최대값으로 나누어 정규화한 후, 정규화 가격 벡터들에 k값을 2에서부터 늘려가며 k-means 알고리즘을 적용한 결과, k=6일 때 각 군집의 평균적인 분포가 어느 정도 인접한 구간들로 모이는 것을 볼 수 있었다. 따라서 이 물품 카테고리의 상품들의 경우, 6개 정도의 가격대를 식별하여 주는 것이 바람직하다는 점을 알 수 있고, 이 때, 얻어진 6개의 군집 중, 첫 번째 군집에 소속된 상품들의 정규화 가격 벡터가 [표 4]에 예시로 나와있다. 여기서 이 군집 내 상품들의 경우 대부분 10번~16번 구간에서 가격이 형성되고 있음을 볼 수 있다. 이 군집의 중심점 벡터는 [0 0 0 0 0 0 0 0.06 0.45 0.42 0.41 0.26 0.19 0.19 0.12 0.01 0 0 0]으로 얻어

지고, 중심점 백터에서 값이 0이 아닌 원소들 중, 양쪽 끝의 작은 빈도의 구간을 제외할 때, 이 군집의 가격대는 90만원~120만원으로 표시할 수 있다. 이와 같은 요령으로 6개 군집 전체의 가격대를 표시한 것이 [표 5]에 요약되어 있다. 또한, [표 5]와 같이 추출된 주요 가격대에 개별 상품들을 적절히 소속시킨 경우, (2)~(3)을 이용하여 계산한 전체 부적합도는 약 8.75만원으로 계산되었다.

반면, 해당 가격 비교 사이트에서 해당 물품 카테고리의 상품 목록을 조회할 때 기본적으로 제시하는 가격대 분류는 '50만원 미만', '50만원~80만원', '80만원~110만원', '110만원~140만원', '140만원 이상'의 5개로, 30만원 단위로 가격을 균일하게 분할하고 있는데, 이 5개의 가격대에 물품들을 소속시켰을 때 얻어지는 전체 부적합도는 약 17.75만원으로, 구간을 균일하게 분할하여 가격대를 나누는 것이 단순하고 직관적이기는 하나, 이 경우 분산 가격을 갖는 온라인 상품들은 복수의 가격대에 분포하거나 한 가격대의 극히 일부 구간에만 분포하는 등의 문제가 많이 발생할 수 있음을 볼 수 있다. 이러한 측면에서 실제 상품들의 가격 분포를 고려하는 지능적인 주요 가격대 산출이 유용할 것으로 생각된다.

표 4. k=6일 때 얻은 클러스터 1

ID	분산 가격 정보
18	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.60 0.22 0.22 0.24 0.00 0.00 0.00 0.00]
51	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.35 1.00 0.51 0.00 0.31 0.00 0.00 0.00 0.00 0.00]
17	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.38 0.38 0.00 0.00 0.15 0.00 0.00 0.00 0.00]
21	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.49 0.07 0.36 0.47 0.43 0.35 0.00 0.00 0.00 0.00]
15	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.34 0.64 0.42 0.37 0.46 0.25 0.05 0.00 0.00]
20	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.34 0.36 0.49 0.05 0.24 0.08 0.00 0.00 0.00 0.00]
26	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.91 0.47 0.28 0.00 0.12 0.00 0.13 0.00 0.00]
44	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.23 0.41 0.18 0.30 0.26 0.22 0.59 0.00 0.00 0.00]
66	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.36 0.87 0.17 0.38 0.00 0.00 0.23 0.00 0.00]
75	[0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.59 0.56 0.00 0.42 0.00 0.00 0.38 0.00 0.00]

표 5. 군집 결과에서 얻은 주요 가격대

주요 가격대	군집	소속 상품 개수
0~ 700,000	2	7
550,000~ 750,000	3	2
700,000~1,000,000	5	9
800,000~1,100,000	4	6
850,000~1,200,000	1	10
900,000~1,350,000	6	6

이렇게 얻어진 주요 가격대 및 관련 정보는 다양한 방법으로 실제 온라인 쇼핑몰 방문자의 구매 의사결정을 지원할 수 있다. 첫째, 온라인 쇼핑몰 방문자가 특정 물품 카테고리를 선택할 경우, [표 5]와 같은 정보를 우선적으로 제시함으로써 고객들로 하여금 해당 카테고리 내 상품들이 대강 어떤 가격대에 주로 분포하는지를 알려주어 구매 예산 수립에 도움을 줄 수 있다. 두 번째로 방문자가 [표 5]를 보고 특정 주요 가격대를 선택할 경우, 소속 상품들의 목록을 보여주는 것을 생각해볼 수 있다. 이 때 조회되는 상품들은 대부분 주요 가격대에서 일반적인 구매가 가능한 상품들이며, 기존의 등간격 가격대에서처럼 실제로는 다른 가격대에서 구매해야 하는 상품들이 없다. 끝으로, [표 2]의 정보를 사용하지 않고, 기존의 등간격 가격대나 최저가 기준 정렬 등을 통하여 특정 상품 정보를 조회하는 고객의 경우, 해당 상품과 같은 주요 가격대에 속하는 상품들을 함께 추천하여 고객이 관심있어 하는 상품과 가격 분포가 비슷한 것들을 제시하는 것이 가능하다.

V. 결론 및 향후 연구 방향

기존 온라인 쇼핑 환경에서는 동일 카테고리 내의 다양한 상품들을 취급하기 위하여 각 상품의 최저가를 중요하게 취급해왔다. 그러나 동일 상품에 대하여 다양한 판매자가 다양한 가격을 제시하는 가격 분산성은 이러한 최저가를 기준으로 하는 방법의 한계점을 야기시키는 요인이 되어왔다.

이에 따라, 본 논문에서는 온라인 쇼핑에서 동일 카테고리로 분류된 상품들의 가격 분포를 고려하여 비슷한 상품들의 가격 분포를 의미하는 주요 가격대를 추출

하는 것이 필요하다는 것을 지적하고, 가격대 분류 평가 방법과 함께 주요 가격대 추출 방법을 제안하였다. 실험 결과, 현재 사용되는 등간격의 가격대와 상품 최저가를 이용한 필터링에서 동일 그룹 내에 가격 분포가 상이한 상품들이 함께 소속되는 경우가 많은 반면, 이렇게 추출된 주요 가격대 기준으로 동일 그룹에 소속된 상품들끼리는 전체적인 가격 분포가 유사하다는 점을 볼 수 있었고, 이러한 측면에서 주요 가격대는 온라인 쇼핑물 방문자의 구매 의사결정 과정을 보다 효과적으로 지원할 수 있을 것으로 기대된다.

나아가, 주요 가격대를 특정 물품 카테고리를 선택한 소비자에게 제시하여, 소비자가 해당 카테고리 내의 상품들이 주로 어떤 가격대들에 분포하고 있는지를 파악할 수 있으며, 자신의 예산이나 구매 용도에 맞는 주요 가격대를 선택하여 보다 효과적으로 상품 정보를 검색 및 수집하는데 사용할 수 있을 것으로 기대된다.

반면, 카테고리 내 상품들 중 가격의 분포가 극히 넓거나 서로 분리된 분포를 갖는 경우에는 전체적으로 주요 가격대 추출 알고리즘의 성능을 떨어뜨릴 수 있음이 관찰되어, 앞으로 이러한 부분에 대한 처리 방법이 필요하다. 아울러, k-means 군집 알고리즘을 사용하여 산출된 주요 가격대들은 서로 중첩되는 경우가 많이 발생하였는데, 이는 소비자들이 한 눈에 주요 가격대 중 자신에 가장 맞는 것을 선택하는데 있어 혼란을 줄 수 있다. 따라서 향후에는 기존의 균일한 가격대처럼 가격을 이산화할 수 있는 주요 가격대 산출 방법에 대한 연구가 필요할 것으로 생각된다.

참 고 문 헌

- [1] <http://www.auction.co.kr>
- [2] M. T. Baye, J. Morgan, and P. Scholten, "Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site," *The Journal of Industrial Economics*, Vol.52, No.4, pp.463-496, 2004.
- [3] M. R. Baye, J. Morgan, and P. Scholten, "Temporal Price Dispersion: Evidence from an Online Consumer Electronics Market," *Journal of Interactive Marketing*, Vol.18, No.4, pp.101-115, 2004.
- [4] <http://www.danawa.com>
- [5] R. Garfinkel, R. Gopal, B. Pathak, and F. Yin, "Shopbot2.0: Integrating Recommendation and Promotions with Comparison Shopping," *Decision Support Systems*, Vol.46, No.1, pp.61-69, 2008.
- [6] R. Garfinkel, R. Gopal, A. Tripathi, and F. Yin, "Design of a Shopbot and Recommender System for Bundle Purchases," *Decision Support Systems*, Vol.42, No.3, pp.1974-1986, 2006.
- [7] M. Haynes and S. Thompson, "Price, Price Dispersion and Number of Sellers at a Low Entry Cost Shopbot," *International Journal of Industrial Organization*, Vol.26, No.2, pp.459-472, 2008.
- [8] S. Y. Hong, J. W. Kim, and Y. H. Hwang, "Fuzzy-Semantic Information Management System for Dispersed Web Information," *Journal of Computer Information Systems*, Vol.52, No.1, pp.96-105, 2011.
- [9] <http://www.gmarket.co.kr>
- [10] G. G. Lim, J. Y. Kang, J. K. Lee, and D. C. Lee, "Rule-based Personalized Comparison Shopping Including Delivery Cost," *Electronic Commerce Research and Applications*, Vol.10, No.6, pp.637-649, 2011.
- [11] M. Limayem, M. Khalifa, and A. Frini, "What Makes Consumers Buy from Internet? A Longitudinal Study of Online Shopping," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, Vol.30, No.4, pp.421-432, 2000.
- [12] B. Pathak, "A Survey of the Comparison Shopping Agent-based Decision Support

Systems," Journal of Electronic Commerce Research, Vol.11, No.3, pp.178-192, 2010.

[13] P. N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison Wesley, 2005.

[14] F. Vachon, "Can Online Aids Support Non-cognitive Web Shopping Approaches?," *International Journal of Business and Management*, Vol.6, No.10, pp.16-27, 2011.

[15] 강대기, 이제선, 함호상, "웹 상의 온라인 비교 쇼핑을 위한 상품 정보 자동 색인 및 검색 시스템의 설계 및 구현에 대한 연구," 한국CALS/EC학회지, 제3권, 제2호, pp.57-71, 1998.

[16] 김경희, "인터넷 쇼핑가치, 점포이미지와 고객만족 및 재방문의도 - 구매제품유형별 비교," 한국콘텐츠학회논문지, 제8권, 제3호, p.173-181, 2008.

임 광 혁(Kwang-Hyuk Im)

정회원



- 1995년 2월 : 한국과학기술원 전산학과(공학사)
- 2000년 8월 : 한국과학기술원 산업공학(공학석사)
- 2006년 2월 : 한국과학기술원 산업공학(공학박사)

▪ 2006년 3월 ~ 2008년 2월 : 삼성전자(주) 반도체연구소 책임연구원

▪ 2008년 3월 ~ 현재 : 배재대학교 전자상거래학과 조교수

<관심분야> : 지식서비스, 경영정보시스템, 데이터마이닝, 지능정보시스템, 전자상거래, 고객관계관리

저 자 소 개

김 준 우(Jun-Woo Kim)

정회원



- 2001년 2월 : 한국과학기술원 산업공학과(공학사)
- 2003년 8월 : 한국과학기술원 산업공학(공학석사)
- 2009년 8월 : 한국과학기술원 산업공학(공학박사)

▪ 2009년 9월 ~ 2011년 2월 : 한국기술교육대학교 산업경영학부 대우교수

▪ 2011년 3월 ~ 현재 : 동아대학교 산업경영공학과 조교수

<관심분야> : 지능형 시스템, 데이터마이닝, 메타 휴리스틱, 조합 최적화, 융합 콘텐츠