

협업적 전자상거래 비즈니스 모델 구현을 위한 유사상품 추천 시스템 개발

A Similar Product Recommendation System Development for Implementing a Collaborative Commerce Model

최상현(Sang Hyun Choi)*, 전영준(Young Jun Jeon)**

*경상대학교 산업시스템공학부 조교수

**경상대학교 산업시스템공학과 석사과정

Abstract

We developed a similar product recommendation system for implementing a collaborative commerce model between the cooperating companies. The system is based on a similar product finding algorithm. The main idea of the proposed algorithm is using a multi-attribute decision making(MADM) to find the utility values of products in same product class of the companies. Based on the values we determine what products are similar. The system helps the companies to recommend products in accordance with the customer's preferences regarding product specifications.

1. 서론

최근 디지털 시대를 맞아 마케팅 환경의 가장 큰 특징으로 '기업과 고객간의 더 한층 밀접한 상호 작용'을 들 수 있으며, 기업은 고객으로부터 얻을 수 있는 정보를 바탕으로 각각의 고객에게 각각 다른 제품과 서비스를 제공해야 하는 과제를 안게 되었다[5]. 이러한 과제는 곧 개인화된 상품 추천으로 이어졌고 현재까지 많은 종류의 추천 알고리즘과 시스템 개발에 대한 연구가 진행되고 있다. 그러나 개인별 선호를 반영하는 연구는 아직도 부족하다.

특히, 구매 빈도가 낮고, 가격이 높으며, 구매의 사결정시 많은 평가기준 및 장시간이 소요되는 고관여 상품을 대상으로 한 추천 시스템의 연구는 거의 찾아볼 수가 없다. 고관여 상품에는 컴퓨터, 자동차, 가구 등이 있는데 이러한 고관여 상품의 구매 시 소비자는 상품에 대한 다양한 정보를 능동적으로 탐색하고, 대안들에 대해 자세히 비교·평가하며, 여러 가지 평가기준 중 몇 가지 속성에 대해 특별히 고려를 하는 특징이 있다[6].

본 논문에서는 기업간 협업 전자상거래 비즈니스

모델 구현을 위한 비즈니스절차를 제안하고, 개인화된 추천을 위한 효용범위 기반 유사상품 추천 알고리즘을 이용한 유사상품 추천 시스템을 개발하였다. 그리고 알고리즘과 시스템의 검증을 위해 실증적 실험을 수행하였다.

2. 추천을 위한 유사성 척도에 대한 기존연구

고객에 대한 성향을 파악한 후 고객에게 적당한 상품 또는 서비스를 추천해주는 시스템을 추천 시스템이라 한다. 오늘날 추천 시스템에 관한 많은 연구가 활발히 진행되고 있으며, 추천 시스템에서 가장 중요한 것은 고객의 선호도를 정확하게 분석하고 정제하여 정확한 예측 능력으로 고객이 원하는 가장 적절한 상품을 추천해 줄 수 있는 능력이다[2]. 이는 곧 개인화된 추천 방법론의 연구로 이어지는데 그 중에서 가장 대표적인 두 방법론으로 내용기반 필터링[13]과 협업 필터링[10]을 들 수 있다.

Montaner 등은 인터넷상의 지능형 추천시스템에 대한 최신의 분류를 제시하였는데, 여기에서도 추천의 방법론을 크게 내용기반 추천과 협업 필터링 그리고 이들을 혼합한 방법론으로 나누고 있다[12]. 그리고 추천을 위한 주요 유사성 척도에는 코사인 유사도, 나이브 베이지안 분류, 피어슨 상관계수 등이 사용되고 있다. [표1]은 추천 시스템을 추천 방법론, 추천 도메인 및 유사성 척도에 대해 재구성한 표이다.

Method	System	Recommending Domain	Similarity Measure
Content Based Filtering	Amalthaea	Web recommender	Cosine Similarity
	Let's Browse	Web recommender	Cosine Similarity
	Letizia	Web recommender	Cosine Similarity
	Syskill & Webert	Web recommender	naive Bayesian classifier
			Nearest Neighbor
			Cosine Similarity
	Webmate	Web recommender	Cosine Similarity
	LaboUr	Document recommender	naive Bayesian classifier
			Nearest Neighbor
	MovieLens	Movie recommender	Cosine Similarity
News Dude	Netnews recommender	Nearest Neighbor	
		Cosine Similarity	
		naive Bayesian classifier	
NewT	Netnews filtering	Cosine Similarity	
Re:Agent	E-mail filtering	Nearest Neighbor	
Collaborative Filtering	Bellcore Video Recon	Movie recommender	Pearson r correlation
	GroupLens	news recommender system	Pearson r correlation
	LaboUr	Document recommender	Pearson r correlation
	MovieLens	Movie recommender	Cosine Similarity
	Ringo/FireFly	Music recommender	Pearson r correlation
	Smart Radio	Music lists recommender	Pearson r correlation
Hybrid	Fab	Web recommender	Cosine Similarity
	Personal WebWatcher	Web recommender	naive Bayesian classifier
	WebWatcher	Web recommender	Cosine Similarity
	Anatagonomy	news recommender system	Cosine Similarity
	Krakatoa Chronicle	news recommender system	Cosine Similarity
	News Weeder	news recommender system	Cosine Similarity

[표1] 추천 시스템 분류

그러나 추천 시스템에 사용되는 주요 유사성 척도인 코사인 유사도, 피어슨 상관계수, 나이브 베이지안 분류, 유클리디안 거리 등은 다음과 같은 단점들을 가진다. 먼저, 코사인 유사도의 경우 군집 영역이 원점에 대하여 떨어져 있고, 군집 영역들 간에도 충분히 떨어져 있을 때에만 가용하다[3]. 피어슨 상관계수의 경우 항목의 종류가 많은 데이터일 경우 희소성(sparsity)문제가 발생, 항목의 종류뿐만 아니라 사용자 수가 많을 경우 알고리즘 수행 속도가 느려지는 규모성(scalability)문제가 발생한다[7, 14, 15]. 나이브 베이지안 분류의 경우 속성 값들이 범주형 데이터일 때 적용가능하고, 속성들에 대해 가중치를 고려하지 않는 문제와 하나의 레코드에 각 속성들은 서로 독립적이다 라는 가정의 제한이 존재한다[7]. 유클리디안 거리의 경우 각 데이터를 점으로 대응시켜야 하는데 카테고리 데이터의 경우 이 같은 점 대응이 어렵게 된다. 그리고 수치 값이 아닌 데이터에 적용이 어려우며 데이터의 수가 많아질수록 상당히 많은 양의 계산과정이 필요하여 처리 속도의 저하를 가져오기도 한다[4, 16].

특히 개인별로 속성들의 가중치를 다르게 반영

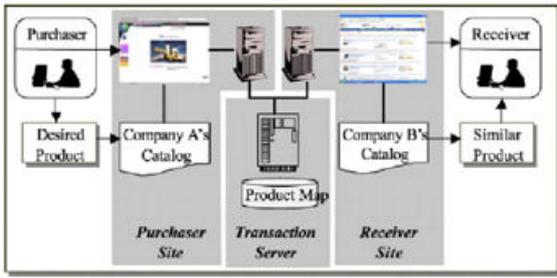
하는 개인화 된 유사성 척도는 거의 없다. 본 논문에서는 상품에 대한 여러 사양 값에 대해 고객이 선호하는 정도를 수치로 표현할 수 있는 가산효용함수(Additive Utility)의 개념을 활용한 효용범위 기반 유사상품 추천 알고리즘(An utility range-based product recommendation algorithm)을 활용하고자 한다[9]. 또한, 이 알고리즘은 적은 양의 데이터에도 적용 가능한 방법으로써 사양별로 다른 가중치를 갖는 문제에 잘 적용될 수 있으며 이와 같은 사양들에 대한 가중치는 사용자들의 선호를 반영할 수 있는 것이다. 예를 들어, 동일한 컴퓨터를 구매할 때, 고객들이 CPU 유형, 브랜드, 가격 등을 고려한다면, 어떤 고객은 CPU 유형을 가장 중요하게 생각할 수 있지만, 또 다른 고객은 브랜드를 가장 중요하게 생각할 수 있다. 이와 같이 일반적으로 상품의 사양에 대해 고객들은 동일하지 않은 선호 정보를 가지고 있으며, 상품 사양에 대한 가중치를 특정한 값으로 측정하기는 상당히 어려운 일이다. 이때, 고객이 상품 사양의 가중치에 대한 정보를 가중치간의 상관관계로써 표현되는 불완전 정보를 제공하는 것이 더욱 수월하다는 것이 알고리즘의 기본 개념이며[8, 9, 11], 불완전 정보에 대한 자세한 내용은 Choi and Kim[8] 논문을 참고하기 바란다.

3. 유사상품 추천 문제 및 절차

3.1 적용 가능 비즈니스 모델

효용범위 기반 유사상품 추천 알고리즘은 고객별로 사양에 대한 가중치를 다르게 반영하여 보다 개인화된 추천을 용이하게 하므로 그 적용 가능 분야가 다양하다. 그 중 대표적인 몇 가지를 제시하자면 다음과 같다.

첫째, 특정 기업에서 특정 사양 값을 갖는 상품을 추천하고자 할 때 고객이 원하는 사양 값을 가지는 상품들을 추천해 줄 수 있다. 둘째, 협업 기업간 상품 맵을 공유함으로써 동일하게 정의된 상품 클래스 내에 있는 상품들에 대해 협업 기업간 상호 보완적 상품 추천이 가능하다([그림 1]). 셋째, 협상을 위한 중개 에이전트 시스템에 응용되어 상품에 대한 다중 속성을 고려하여 협상에 참여한 두 에이전트의 상호 이익을 최대화하기 위한 에이전트 시스템에 적용이 가능하다[1]. 넷째, 특정 기업에서 고객이 구매하고자 하는 상품의 재고가 없을 때, 가장 유사한 대체 상품의 추천이 가능하다.



[그림 1] 협업 비즈니스 모델의 3가지 구성요소(Choi & Cho, 2004)

3.2. 유사상품 추천을 위한 문제 정의

효용범위 기반 유사상품 추천 알고리즘은 동일한 사양의 집합에 의해 정의될 수 있는 단품 단위(SKUs, Stock Keeping Units)의 클래스에 속해 있는 상품들을 사용자의 선호에 따라 분류할 수 있는 방법을 제안한다. 이 알고리즘을 활용하게 되면, 동일한 사양의 집합을 갖는 클래스로 구분한 후, 각 단품의 사양 값에 따라 세부분류를 수행할 수 있으며, 이때 사양 값은 사용자의 효용 값으로 변환되며 사양별 가중치를 반영하여 가중 합을 구하면 상품별 효용 값을 구할 수 있다. 이때 사용자마다 다른 사양별 가중치를 반영함으로써 사용자들의 각자 다른 선호를 반영하는 것이 가능해진다.

현재 하나의 기업이 보유하고 있는 상품에 대해 단품 단위 수준에서 K개의 클래스 중에서 동일한 하나의 클래스에 속해있는 상품들은 동일한 사양들의 집합에 의해 평가될 수 있으며, 각 상품들은 다양한 사양 값들을 갖게 된다. 동일한 클래스에 속해있는 상품들의 효용 값을 구하기 위해서 다음과 같이 용어를 정의한다.

$I = \{i\}_{i=1,N}$: 하나의 클래스로 분류되는 N개의 상품들의 집합

$J = \{j\}_{j=1,M}$: 하나의 클래스에 속해있는 상품들의 특성을 나타내는 M개의 사양들의 집합

w_j : j 번째 사양의 중요도

x_{ij} : 하나의 클래스내의 i번째 상품에 대한 j번째 사양의 값

다기준 의사결정 분석(MCDM, Multi-Criteria Decision Making)은 유한한 개수의 속성에 대해서 의사결정 대안을 평가하기 위한 방법이기 때문에, 유한한 사양의 값들로 이루어진 상품의 평가문제에 잘 적용될 수 있다. 가장 잘 알려진 방법은 M개의 속성 값을 갖는 상품 $P_i = (x_{i1}, \dots, x_{iM})$ 을 평가하기 위해 가치 함수 v 에 대한 가중가산형태로 이용하는 것이다.

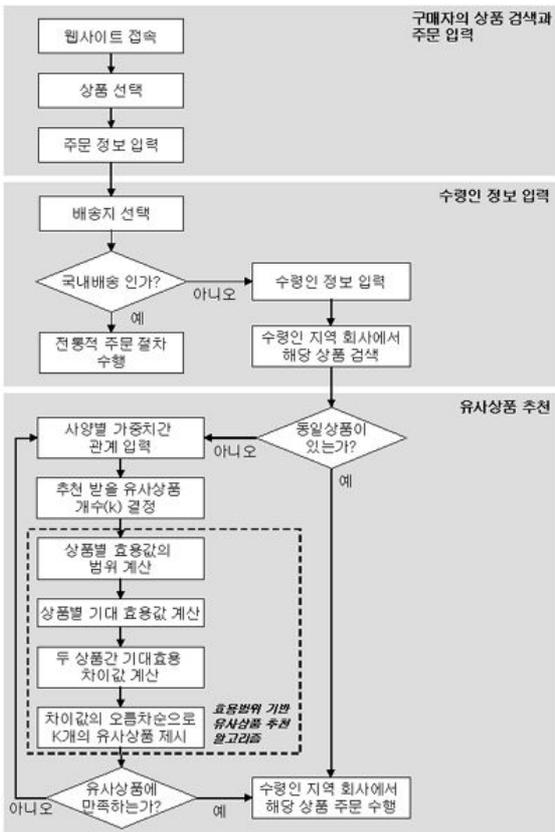
$$v_i(P_i) = \sum_{j=1}^M w_j v_{ij}(x_{ij}) \quad (\text{식 1})$$

여기서 v_{ij} 는 $v_{ij} : x_{ij} \rightarrow [0, 1]$ 의 조건을 만족시키는 한계 효용가치 함수이다. 결국, 효용범위 기반 유사상품 추천 알고리즘에서는 고객이 선택한 상품과 동일한 클래스에 속하며 대상 사이트에서 보유하고 있는 상품들의 효용의 범위를 구하고 고객이 선택한 상품 P_s^A 와 가장 유사한 효용 값을 갖는 유사상품 P_s 를 찾아내어 고객에게 제안하는 것이 목적이다.

상품들의 효용 값을 구하기 위해서 의사결정 모수인 w_j 와 $v_j(\cdot)$ 가 의사결정자에 의해 모두 정확한 숫자 값으로 주어질 수 있다면, (식 1)을 사용해서 P_s^A 와 P_s 의 가치 함수 값을 쉽게 계산할 수 있으나, 실제 문제에서 가중치의 정확한 값을 구하는 것은 상당히 어려운 일이다. 대신에 의사결정자는 가중치에 대한 정보를 가중치 간의 관계식이나 가중치 값의 범위 등의 불완전 정보의 형태로 제공하는 것이 용이하다.

3.3. 유사상품 추천 절차

본 절차는 국가가 다르고, 협업 관계에 있는 두 국가 내의 쇼핑몰에서 상품의 클래스 및 사양정보를 공유하는 비즈니스 모델 하에, 고객이 구매지에서 선택한 상품에 대해 배송지의 상품들 중 이 상품과 동일한 분류에 속하는 유사상품을 찾고자 할 때 수행된다. 유사상품 발견의 절차를 그림으로 표현하면 [그림 2]와 같다.



[그림 2] 유사상품 발견 절차

먼저 구매고객은 자국 쇼핑몰에서 원하는 상품을 선택하고 주문정보를 입력하게 된다. 그리고 배송지를 선택하게 되는데, 이때 수령인이 구매자 자신이거나 또는 국내에 있는 다른 사람일 경우 전통적 주문 절차에 의해 주문처리가 이루어진다. 그러나 수령인이 국외 거주자일 경우 상품 검색이 이루어진 사이트와 협업관계에 있는 해당 배송지의 사이트의 상품 리스트에서 동일 상품이 있는지 검색하게 된다. 이때, 동일 상품이 있을 경우 그 쇼핑몰에서 상품을 구매하고 해당 수령지로 배송주문을 하게 된다. 그러나 동일 상품이 없을 경우 유사상품 추천의 절차를 거치게 되는 것이다.

유사상품 추천의 처리 과정은 먼저 고객으로부터 사양별 가중치간의 관계를 불완전 정보의 5가지 표현 형태를 이용하여 입력받게 되고, 그다음 추천받고자 하는 유사상품의 개수를 지정받게 된다. 여기까지가 고객으로부터의 입력 정보를 받는 단계이고 이 정보들을 통해 효용범위 기반 유사상품 추천 알고리즘을 거치게 된다.

효용범위 기반 유사상품 추천 알고리즘의 세부 처리 순서는 다음의 5단계로 구성된다.

제 1단계 : 상품별 효용 값의 범위 계산 - 고객이 선택한 상품과 수령지에서 보유하고 있는 상품에 대한 사양 값을 취합하고 이 값에 근거하여 사양별

효용 값을 계산한다. 이때, i 번째 상품에 대한 j 번째 사양에 대한 효용가치 v_{ij} 는 (식 2)와 같이 구할 수 있으며, 이는 $[0, 1]$ 사이의 값으로 표현된다.

$$v_{ij} = \frac{x_{ij} - \min_i x_{ij}}{\max_i x_{ij} - \min_i x_{ij}}$$

j : 값이 클수록 효용가치가 높은 사양

$$\text{또는, } v_{ij} = \frac{x_{ij} - \max_i x_{ij}}{\min_i x_{ij} - \max_i x_{ij}}$$

j : 값이 작을수록 효용가치가 높은 사양

(식 2)

제 2단계 : 상품별 기대 효용값 계산 - 상품 분류표상 동일 클래스에 있는 상품들의 효용 값의 범위를 계산한다. 이때, 사양 값에 대한 가중치 w_j 에 대한 정확한 값을 의사결정자에 의해 입력 받는 것이 매우 어려우므로 사양에 대한 가중치간 불완전 정보를 의사결정자로부터 입력 받아 제약식으로 사용하여 선형계획문제를 푸는 방법으로 효용 값의 증가산 범위를 구하게 된다. 그리고 3.2절에서 언급한 5가지 형태의 불완전 정보를 이용하여 사양에 대한 상대적인 가중치 정보로서 입력된 불완전 정보를 모아서 가중치간 관계집합 Φ_w 로 정의한다. 이와 같이 얻어진 정보를 활용하여 i 번째 상품에 대한 효용 값의 범위, $[v_i(\min), v_i(\max)]$,를 다음과 같은 수식에 의해 계산한다.

$$v_i(\min) = \min \sum_{j=1}^M w_j x_{ij}$$

$$\text{and, } v_i(\max) = \max \sum_{j=1}^M w_j x_{ij}$$

subject to Φ_w (식 3)

결국, i 번째 상품의 기대 효용 값, $E[v_i]$, 은 $[v_i(\min) + v_i(\max)]/2$ 으로 구할 수 있다.

제 3단계 : 두 상품 간 기대효용 차이 값 계산 - 고객이 선택한 상품의 기대효용 $E[v_s]$ 와 수령지의 각 상품들의 기대효용 $E[v_i]$ 의 차이의 절대 값 $DE[v_{si}]$ 를 구한다. $DE[v_{si}] = |E[v_s] - E[v_i]|$ (식 4)

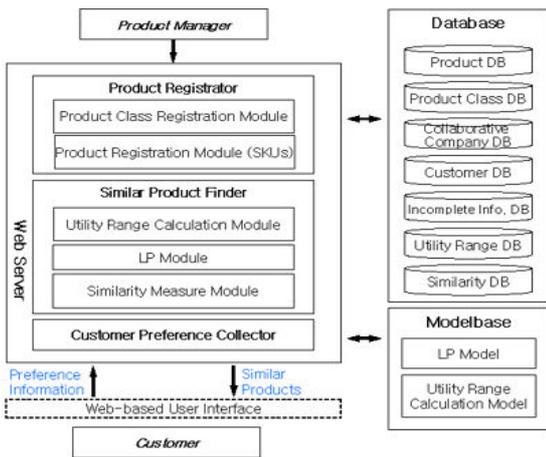
제 4단계 : 차이 값의 오름차순으로 k 개의 유사상품 제시 - 수령지의 각 상품들에 대한 $DE[v_{si}]$ 값이 작을수록 유사도가 높은 상품으로 정의하고, 고객이 선택한 k 개의 상품을 $DE[v_{si}]$ 값의 오름차순으로 유사상품을 제시한다.

이상으로 불확실상황에서의 가산효용함수의 개념을 적용한 효용범위 기반 유사상품 추천 알고리즘

과 이를 이용한 유사상품 발견 절차에 대하여 살펴 보았으며, 다음 장에서는 이를 바탕으로 구현한 시스템을 소개하고자 한다.

4. 시스템 구현

본 시스템은 크게 상품 등록기(Product Registrator), 고객 선호 수집기(Customer Preference Collector), 유사상품 탐색기(Similar Product Finder)로 구성이 되며 이들은 또 다시 몇 개의 하위모듈들로 이루어진다. [그림 3]은 본 시스템의 아키텍처를 나타낸다.



[그림 3] 시스템 아키텍처

본 시스템은 웹 기반의 사용자 인터페이스를 바탕으로 상품 관리자 측면의 프로세스와 고객 측면의 프로세스가 각각의 모듈들과 연계된다.

4.1. 상품 등록기

상품 등록기는 Product Class Registration Module과 단품 단위의 Product Registration Module로 이루어진다. 상품의 등록에 관련된 작업을 담당하는 사람을 상품 관리자(Product Manager)라고 하고, 이 상품 관리자에 의해 새로운 상품 클래스가 정의되고, 상품 클래스별 사양들이 결정된다.

상품 관리자는 새로운 단위 상품을 등록하기에 앞서 그 상품의 해당 클래스를 먼저 정의해 주어야 한다. 이미 정의된 상품 클래스가 있으면 새로운 단위 상품을 해당 클래스 아래에 상품 등록을 하고, 정의된 클래스가 없다면 다음과 같은 절차에 의해 새로운 상품 클래스를 생성한다.

먼저, 상품 클래스 아이디와 이름을 정의하고, 상품 클래스에 대한 사양을 정의한다. 이때, 가격 등과 같이 수치화 할 수 있는 사양을 정량적 사양이라고 하고, 제조사 등과 같이 비수치화 된 사양을 정성적 사양이라 정의한다. 다음으로 각 사양들에 대해

B.F.L.(Better For Larger, 크면 클수록 좋은 값. 예, 컴퓨터 메모리)과 B.F.S.(Better For Smaller, 작으면 작을수록 좋은 값. 예, 가격)을 지정한다. 마지막으로 정성적 사양에 대해서 효용 값 계산을 위한 점수화 과정인 카테고리 값 지정(예를 들어, 컴퓨터 메모리의 경우 SDR256, SDR512, DDR128, DDR256, DDR512에 대해 각각 1, 2, 3, 4, 5의 값을 지정)을 거치게 된다.

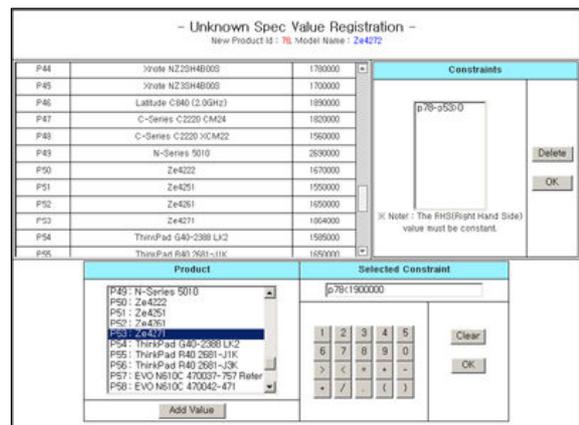
No.	Spec Name	Category Value	Determine B.F.L. or B.F.S.	Delete
1	Price	Define	B.F.S.	Delete
2	Brand	Defined	B.F.L.	Delete
3	RAM	Defined	B.F.L.	Delete
4	HDD	Defined	B.F.L.	Delete
5	ODD	Defined	B.F.L.	Delete
6	Display	Defined	B.F.L.	Delete

※ B.F.L : Better For Larger
B.F.S : Better For Smaller

[그림 4] 정의된 상품 클래스 사양

상품 클래스와 세부 사양이 정의되고 나면 본격적으로 새로운 단위 상품들을 등록하게 된다. 이때, 상품의 등록은 사양 값을 알고 있는 상품을 등록하는 경우와 일부 또는 전체의 사양 값을 모르는 상품의 등록으로 나누어진다.

먼저, 사양 값을 이미 알고 있는 상품의 등록은 해당 사양 값이 정량적 사양 값일 경우 수치 값을 바로 입력하고, 정성적 사양 값일 경우 해당 카테고리 값을 선택하게 된다. 이에 반해, 정확한 사양 값을 모르는 상품의 경우 [그림 5]와 같이 다른 상품들과의 관계 정보를 5가지 유형의 불완전 정보로 설정해줌으로써 사양 값을 표현한다. 이 관계식은 LP Module을 통해 사양 값의 최대값과 최소값을 구해낼 때, 제약식으로 이용된다.

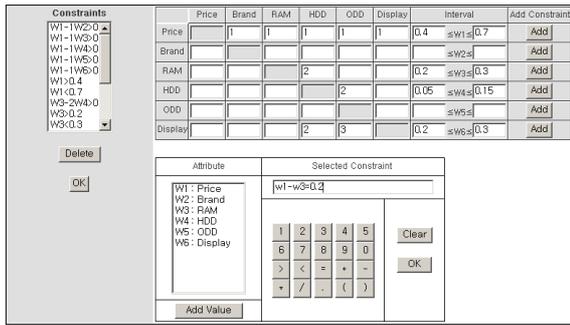


[그림 5] 부정확한 사양 값을 갖는 상품의 등록

모든 단위 상품의 등록이 완료되면 정확한 사양 값을 갖는 상품에 대해서는 (식 2)를 사용하여 상품별 상대적 효용 값을 계산한다.

4.2. 고객 선호 수집기

유사 상품 추천을 위해서는 고객으로부터 먼저 상품별 사양의 가중치에 대한 선호 정보를 입력받아야 한다. 그러나 3.2절에서 언급하였듯이 선호정보는 주로 불완전하고, 부정확한 형태로 기술되는 것이 보통이다. 그렇기 때문에 본 시스템에서는 불완전 가중치 정보를 효율적으로 표현할 수 있는 도구를 제작, 적용하였다.



[그림 6] 불완전 가중치 정보 입력 양식

4.3. 유사상품 탐색기

상품의 유사도 측정을 위해 사용되는 효용 값의 범위를 구하기 위해 LP Module을 통한 최대화 및 최소화 과정을 거치게 된다. 본 시스템에서는 자동적인 선형계획문제의 정식화를 지원하고 있으며 이는 LP Module을 통해 수행된다. 정식화된 선형계획문제는 내부 어플리케이션인 Lindo API를 통해 솔루션이 탐색된다. 그리고 계산된 최대값과 최소값은 상품별로 데이터베이스에 저장이 된다.

[그림 2]의 유사상품 발견 절차에 따라 고객이 원하는 상품에 대해 유사 상품의 개수를 지정하고 각 모듈의 내부 처리를 통해 유사상품이 제시되게 된다. [그림 7]은 이러한 유사상품 발견의 결과를 나타낸다.

Product ID	Model Name	Price	Brand	RAM	HDD	ODD	Display
54	ThinPad 640-C388 Lx2	1595000	LG-BM(2)	DDR5(2G)	60GB 이(1)5	COMB(0)5	14inch 이(1)2 15inch 미(1)3

We have found 5 similar products at the B mall respectively by Utility Range-based Approach.

Product ID	Model Name	Price	Brand	RAM	HDD	ODD	Display
13	PC-G-GR070	199000	Sony(5)	DDR5(2G)	30GB 이(1)2 ~ 60GB 미(1)3	COMB(0)5	14inch 이(1)2
4	EVO N100C 470051-341	1429000	Compaq(3)	DDR(256G)	30GB 이(1)2 ~ 60GB 미(1)3	COMB(0)5	14inch 이(1)2 ~ 15inch 미(1)3
6	EVO N1025V 470059-399	163000	Compaq(3)	DDR5(2G)	30GB 이(1)2 ~ 60GB 미(1)3	COMB(0)5	14inch 이(1)2 ~ 15inch 미(1)3
23	드림북 N100L4	1429000	TG(2)	DDR(256G)	30GB 이(1)2 ~ 60GB 미(1)3	COMB(0)5	14inch 이(1)2 ~ 15inch 미(1)3
24	드림북 N102L4	1429000	TG(2)	DDR(256G)	30GB 이(1)2 ~ 60GB 미(1)3	COMB(0)5	14inch 이(1)2 ~ 15inch 미(1)3

[그림 7] 유사상품 발견 결과

5. 실험적 연구

본 논문에서 다루고 있는 효용범위 기반 유사상품 추천 알고리즘의 정확도와 사용자들의 시스템 사

용 만족도를 측정하기 위해 실증 실험을 수행하였다.

5.1. 실험 환경

실험은 본 시스템에 사용된 알고리즘에 대한 사전 지식이 전혀 없으며, 각기 서로 다른 선호도를 가진 8명의 학생을 대상으로 수행하였다. 그리고 비교 대상이 되는 방법론으로는 유클리디안 거리 척도를 이용한 추천방법론에 의한 유사상품 제시를 사용하였다.

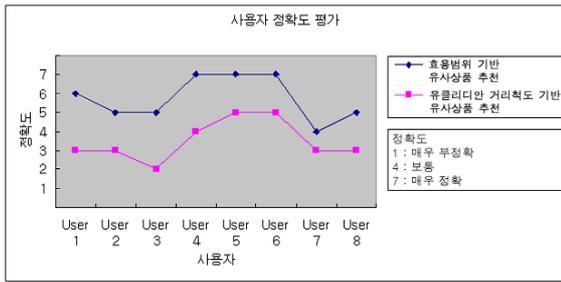
실험에 사용된 평가척도로는 사용자가 선택한 상품과 추천된 유사상품들의 일치정도를 사용자가 직접 7점 척도(1점 : 매우 부정확, 4점 : 보통, 7점 : 매우 정확)로써 정확도를 표시하도록 하였고, 유사상품 추천 과정을 수행 중 시스템 사용에 대한 입출력 과정에서 사용자들이 느낀 만족도를 마찬가지로 7점 척도(1점 : 매우 불만족, 4점 : 보통, 7점 : 매우 만족)로써 표시하도록 하였다.

실험에 사용된 상품 클래스는 노트북 컴퓨터이고, 상품의 사양으로는 가격, 브랜드, 메모리, 하드디스크 용량, 광학드라이브 종류, 디스플레이 크기로 한정하였다. 실험의 편의상 상품의 등록과 사양에 대한 카테고리 값들의 정의는 사전에 시스템 개발자에 의해 수행되었다.

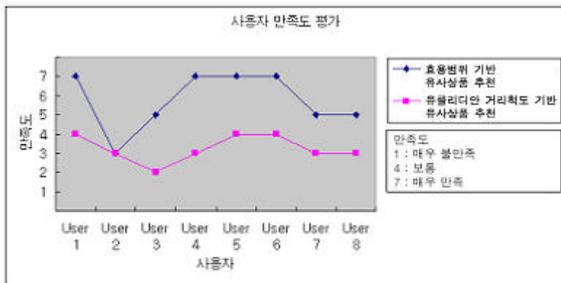
앞부분은 실험 참가자들에게 문제 상황 설명을 통해 대신하고 실제 실험은 [그림 2]의 유사상품 추천에 대해서만 다음과 같은 순서로 진행하였다. 이때 추천받을 유사상품 개수(k)는 5개로 통일하였다. 먼저 실험 참가자가 [그림 6]의 입력 양식을 통해 자신의 사양별 선호정보를 입력한다. 그리고 시스템이 이를 이용하여 각각의 내부 모듈들과 효용범위 기반 유사상품 추천 알고리즘에 의한 5개의 유사상품을 제시하면, 실험 참가자는 제시된 유사상품들에 대해 정확도와 만족도를 표시한다. 다음으로 비교대상이 되는 유클리디안 거리척도를 이용한 유사상품 5개에 대해서도 마찬가지로 정확도와 만족도를 표시하는 방식으로 진행하였다.

5.2. 실험결과

[그림 8]의 정확도 평가 결과 그래프와 [그림 9]의 만족도 평가 결과 그래프에서 알 수 있듯이 본 시스템에서 사용된 효용범위 기반 유사상품 추천 알고리즘에 의한 추천 결과가 유클리디안 거리척도 방법론에 의한 추천 결과보다 정확도와 만족도 모두 높게 나타난 것을 확인 할 수 있다.



[그림 8] 정확도 평가 결과



[그림 9] 만족도 평가 결과

두 추천 기법 간 유용성에 관한 통계적 유의성 검정을 위해 대응표본T검정을 실시하였다. [표2], [표3]과 같이 정확도와 만족도 모두 유의수준 5% 내에서 두 기법 간의 평균에 관한 차이가 통계적으로 유의한 것으로 나타났다. 이러한 결과는 실험 참가자들과의 인터뷰에서 본 방법론이 개인별 선호를 반영할 수 있도록 설계되었기 때문으로 조사되었다.

정확도	실험인원(명)	평균	표준편차	평균의 표준오차	T-Value	P-Value
효용범위 기반 유사상품 추천	8	5.750	1.165	0.412	9.00	0.000
유클리디안 거리척도 기반 유사상품 추천	8	3.500	1.069	0.378		
대응차	8	2.250	0.707	0.250		

[표2] 정확도에 대한 통계적 검증

만족도	실험인원(명)	평균	표준편차	평균의 표준오차	T-Value	P-Value
효용범위 기반 유사상품 추천	8	5.750	1.488	0.526	5.92	0.00
유클리디안 거리척도 기반 유사상품 추천	8	3.250	0.707	0.250		
대응차	8	2.500	1.195	0.423		

[표3] 만족도에 대한 통계적 검증

6. 결론 및 추후 연구

본 논문에서는 효용범위 기반 유사상품 추천 알고리즘을 바탕으로 시스템을 개발하였으며 개발된 시스템의 성능 검증을 위해 실험적 연구를 수행하였다. 본 시스템의 특징은 고객별로 사양에 대한 가중치를 다르게 입력받아 보다 개인화된 추천을 수행할 수 있으며, 웹 브라우저를 통한 그래픽 사용자 인터페이스와 자동화된 모듈 활용을 통해 사용자 편의성

을 증대시켰다. 시스템을 통한 사용자 실험을 거친 결과 기존의 추천 방법론에 비해 향상된 정확도와 만족도를 얻을 수 있었다.

본 논문의 추후 연구 과제로는 하나의 상품을 선택하고 그와 유사한 상품을 검색하는 현재의 기능 이외에 특정 사양 값을 만족하는 상품 추천을 위해서 사양 값에 대해서 검색을 제한하는 방향으로의 시스템 개발이 필요하며, 실제 기업들에 적용해보는 실험이 수행된 후 시스템 개발이 진행되어야 할 것이다.

참고문헌

- [1] 박상현, 양성봉, "다중 속성 협상과 상호 이익을 위한 중개 에이전트 시스템", 「정보과학회지」, 제31권, 제3호(2004), pp.308-316.
- [2] 박지선, 김택현, 류영석, 양성봉, "추천 시스템을 위한 2-way협동적 필터링 방법을 이용한 예측 알고리즘", 「정보과학회지」, 제29권, 제10호(2002), pp.669-675.
- [3] 이성환, 「패턴인식의 원리 1권」, 홍릉과학출판사, 1994
- [4] 이정수, 윤소정, 오경환, "행동 분석을 이용한 적응형 정보여과 에이전트", 「인지과학」, 제9권, 제2호(1998)
- [5] 이혁수, "디지털 시대의 새로운 마케팅 패러다임", 「주간경제」, 제578호(2000)
- [6] 이훈영, 「e-마케팅 플러스」, 무역경영사, 2002
- [7] Billsus, D. and M.J. Pazzani, "Learning collaborative information filters", *International Conference on Machine Learning*, Vol.15(1998), pp.46-54.
- [8] Choi S.H. and S.H. Kim, "An interactive procedure for multiple attribute group decision making with incomplete information", *European Journal of Operational Researches*, Vol.118, No.1(1999), pp.139-152.
- [9] Choi, S.H. and Y.H. Cho, "An utility range-based similar product recommendation algorithm for collaborative companies", *Expert Systems with Applications*, Vol.27(2004), pp.549-557.
- [10] Goldberg, D., D. Nichols, B.M. Oki and D. Terry, "Using Collaborative Filtering to Weave an Information Tapestry", *Communications of the ACM*, Vol.35, No.12(1992), pp.61-70.
- [11] Kim J.K. and S.H. Choi, "An utility range

2005 한국경영과학회/대한산업공학회 춘계공동학술대회
2005년 5월 13일~14일, 충북대학교

based interactive group support system for multiattribute decision making", *Computers and Operations Research*, Vol.28, No.5(2001), pp.485-503.

[12] Montaner, M., B. Lopez, and J.L.D. Rosa, "A Taxonomy of Recommender Agents on the Internet", *Artificial Intelligence Review*, Vol.19(2003), pp.285-330.

[13] Mooney, R. and L. Roy, (2000). "Content-Based Book Recommending Using Learning for Text Categorization", *ACM Conference on Digital Libraries*, Vol.5(2000), pp.195-204

[14] Sarwar, B., G. Karypis., J. Konstan., and J. Reidl, "Analysis of recommendation algorithms for e-commerce", *ACM Conference on Electronic Commerce (2000)*, pp.158-167.

[15] Sarwar, B., G. Karypis., J. Konstan., and J. Reidl, "Item-based Collaborative Filtering Recommendation Algorithms", *International WWW Conference, (2001)*, Hong-Kong.

[16] Schafer, J.B., J.A. Konstan and J. Riedl, "E-Commerce Recommendation Applications", *Data Mining and Knowledge Discovery*, Vol.5(2001), pp.114-153.