

# 온라인 분석 처리(OLAP) 기반 홈쇼핑 상품 트렌드 분석 방법

박한샘\*, 권경락\*, 강대현\*, 이정민\*, 정인정\*

\*고려대학교 컴퓨터정보학과

e-mail : {park11232000, helpnara, internetkbs, wjdals543, chung}@korea.ac.kr

## Home Shopping Product Trend Analysis Method Based on On-Line Analytical Processing (OLAP) cube

Hansaem Park\*, Kyunglag Kwon\*, Daehyun Kang\*, Jeungmin Lee\*, In-Jeong Chung\*

\*Dept. of Computer and Information Science, Korea University

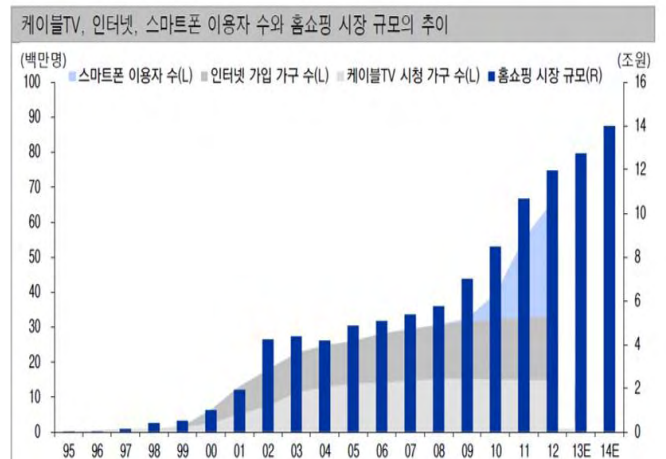
### 요 약

최근 웹 2.0의 폭발적인 성장과 스마트기기의 대중화 및 모바일 서비스의 활성화로 인하여 다양하고 방대한 양의 정보들이 생성되었다. 또한, 현재 산업분야에서는 이와같은 방대한 양의 데이터들을 처리하기 위하여 데이터웨어하우스와 OnLine Analytical Processing(OLAP)을 통한 정보 분석 사례가 많아지고 있다. 특히, 의사결정자들은 이러한 수많은 정보들 중에서 의사결정에 도움이 되는 정보들을 찾는 것을 목표로 하지만 아직까지도 의사결정자들은 자신들이 원하는 정보들을 찾는 데 많은 어려움을 겪고 있다. 따라서, 최근에 수많은 정보들을 효과적으로 활용하기 위한 다양한 연구가 수행어지고 있고 의사결정자들의 올바른 의사결정을 도와주는 시스템에 대한 중요도가 나날이 급증하고 있다. 본 논문에서는 이러한 의사결정자들의 올바른 의사결정을 위해 OLAP을 활용하여 TV 홈쇼핑에서 발생하는 수많은 정보들을 분류 목적에 따라 다차원적으로 분석하고 분석된 정보들을 바탕으로 하여 TV 홈쇼핑에서 판매하고 있는 상품의 트렌드를 분석한다.

### 1. 서론

최근 스마트기기의 대중화와 웹 2.0의 급속한 발전으로 인하여 시공의 제약 없이 온라인 서비스를 실시간으로 이용할 수 있게 되었다. 이로 인하여, 기존의 판매자와 직접 대면하던 소매 유통방식에서 인터넷, 모바일을 통한 유통채널로 확장되어졌고 현재 국내 소매유통시장을 이끌어가고 있다. TV 홈쇼핑은 소비자가 쇼핑을 위해 직접적으로 점포를 방문하지 않고도 집안에서 정보전달매체를 이용하여 상품을 구매하는 판매방식으로, 바쁜 일상생활 속에서 편리함을 추구하는 소비자들로부터 각광받을 뿐만 아니라 언제 어디서든 홈쇼핑 방송을 보거나 상품을 구매할 수 있도록 모바일 어플리케이션을 개발함으로써 급속하게 성장하고 있다[1]. 그림 1<sup>1</sup>과 같이 TV 홈쇼핑의 급속한 성장으로 인해 TV 홈쇼핑 기업의 매출 손실을 감소시키기 위해서 TV 홈쇼핑 의사결정자들의 역할이 매우 중요해졌다. 이러한 의사 결정자들의 올바른 의사결정을 돕기 위해 소비자들의 개인 성향에 맞는 홈쇼핑 상품을 다차원적으로 분석할 필요가 있다. 따라서 정

보를 다각적으로 분석할 수 있는 데이터웨어하우스



및 OLAP 시스템의 구축이 필연시 되어졌다[2].

<그림 1> 케이블 TV, 인터넷, 스마트폰 이용자 수와 홈쇼핑 시장 규모의 추이

OLAP은 데이터웨어하우스와 함께 사용되는 가장 효율적인 도구이며 최근 다양한 분야에서 의사결정을 효과적으로 지원하기 위해 많이 사용되어지고 있다[3].

<sup>1</sup> <http://www.etrade.co.kr>

OLAP은 기존 OnLine Transaction Processing(OLTP)에서 제공하는 데이터 삽입, 삭제, 업데이트 등에 관한 트랜잭션 뿐만 아니라 데이터를 다차원적으로 분석하기 위해 데이터 큐브 모델을 구축하고 OLAP 연산을 사용함으로써 기존에 발견되지 않았거나 알지 못했던 새로운 사실들을 도출해 낼 수 있다[4].

본 논문에서는 국내의 특정 하나의 TV 홈쇼핑에서 발생하는 수많은 정보들을 바탕으로 데이터웨어하우스를 구축하고 다차원적 분석을 위하여 데이터 큐브 모델을 생성한 후, OLAP 연산을 통해 의사결정에 도움이 되는 상품의 트렌드를 분석하는 방법을 제안한다.

마지막으로 제안한 방법을 활용함으로써 상품의 트렌드를 분석할 수 있음을 파일럿 실험을 통하여 보인다.

## 2. Data Warehouse and Online Analytical Processing (OLAP)

데이터 웨어하우스는 방대한 데이터의 저장소를 의미하는 것으로 의사 결정자들에게 요구되는 정보를 효율적으로 제공하고 사용자의 관점에서 주제별로 (Subject-Oriented) 통합 및 가공하기 위하여 사용된다. OLAP은 OLTP와 다른 개념으로써 최종사용자(End-User)가 데이터웨어하우스의 다차원정보에 직접 접근하여 대화식으로 원하는 정보를 분석하고 이를 의사 결정에 활용하는 과정으로 정의할 수 있다[2, 4]. 초기 데이터웨어하우스 구축의 초점은 사용자의 의사결정을 지원할 수 있는 정보기반의 구축에 있었던 반면, OLAP은 정보를 다각적인 관점에서 효과적인 분석 및 활용하는 측면에 초점을 맞추었다. 또한 데이터웨어하우스 구축은 데이터의 일관성, 정확성, 가용성 등의 측면을 중시하는 반면, OLAP은 최종사용자의 분석적 요구사항이나 모델링 측면을 중시하였다. 그러나 IT 기술의 발달과 스마트기기의 보편화로 인하여 방대한 양의 정보가 축적되었고 이와 같은 정보들의 가치가 점점 더 중요시 되면서 데이터웨어하우스는 구축의 관점보다는 분석 및 활용의 관점이 강해짐에 따라 급속도로 OLAP과의 연계성이 높아지게 되었다[5]. 또한, 데이터 웨어하우스의 구축에 있어서 OLTP 시스템은 데이터 분석보다는 처리에 적합하게 설계되어 만족할만한 데이터 분석 속도 및 질의 처리 결과를 얻기 어렵다. 따라서 데이터 웨어하우스에서는 다차원 데이터 분석을 위하여 OLAP 시스템을 활용한다. OLAP은 OLTP와 달리 데이터 항목들을 다양한 시점에서 조회하고 분석하는 다차원 분석을 지원하며 이로 인해 중요한 의사결정에서 요구되는 복잡한 질의들을 효과적으로 처리할 수 있다[6]. OLAP은 크게 관계형 데이터베이스를 기반으로 하는 Relational OLAP(ROLAP)과 다차원 데이터베이스 기반의 Multi-dimensional OLAP(MOLAP)과 마지막으로 위의 두가지 방법의 단점을 보완하기 위하여 두가지 방법을 결합시킨 Hybrid OLAP(HOLAP)이 있다[7]. OLAP의 형태 및 방식에 따라서 데이터웨어하우스의 구조와 데이터

의 추출 및 가공 방식이 달라질 수 있다. MOLAP은 다차원적인 분석을 빠르게 할 수 있는 반면, 원시데이터를 볼 수 없고, 대용량의 데이터를 취급하기에는 역부족이다. 또한 데이터를 로딩 하는데 많은 시간을 소비한다. ROLAP은 원시데이터의 취급과 대용량의 데이터를 취급할 수 있는 확장성을 제공한다.

## 3. 제안하는 방법

본 논문에서 제안하는 방법은 다음과 같다. 국내 대표적인 TV 홈쇼핑 사이트 중 하나를 통하여 상품명, 상품에 대한 고객 평가, 상품의 카테고리에 대한 정보 등을 수집한 후, 이와 같은 정보들을 다차원 분석 기법인 OLAP을 활용하여 TV 홈쇼핑 상품의 트렌드를 분석 및 파악하고 이를 활용하여 의사결정자의 의사결정을 보다 효율적으로 할 수 있도록 도와준다.

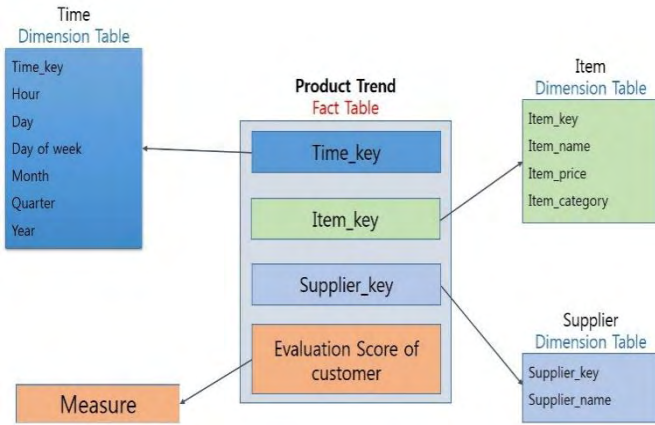
제안한 방법은 크게 1)데이터 수집, 2)TV 홈쇼핑 데이터를 활용한 다차원 데이터 모델 구성, 3)OLAP 연산을 이용한 TV 홈쇼핑 상품 트렌드 분석으로 크게 3 단계로 나눈다.

### 3.1. 데이터 수집

현재 국내 홈쇼핑 사이트에서는 다양한 정보들을 제공한다. 예를 들어, 상품명, 상품이 속한 카테고리, 고객평가 점수(품질, 디자인, 가격, 배송)과 상품 가격, 상품을 평가한 참여고객의 수 등이 있다. 본 논문에서는 고객들의 성향에 맞는 상품의 트렌드를 분석하기 위하여 사실테이블에 측도값(Measure)으로써 상품에 대한 고객평가 점수를 수집하고, 각 차원 테이블에 들어갈 측도값으로는 상품명, 날짜 및 시간, 홈쇼핑 이름, 상품가격과 상품의 카테고리 와 같은 데이터를 수집한다. 또한, 수집된 다양한 데이터를 바탕으로 하여 다차원 데이터 큐브 모델을 다음 단계에서 구성한다. 구성하기 위하여 각 차원 테이블에 측도값(Measure)으로 맵핑되어 질 수 있는 정보들을 수집한다.

### 3.2. TV 홈쇼핑 데이터를 활용한 다차원 데이터 모델 구성

데이터웨어하우스와 OLAP 도구는 다차원 데이터 모델을 기초로 하며, 이 모델은 데이터 큐브의 형태로써 제공된다[8-10]. 데이터 큐브는 데이터가 여러 차원으로 모델링되어지고 차원(Dimension)테이블과 사실(Fact)테이블로 정의된다. 본 논문에서는 상품의 트렌드를 분석하기 위하여 시간(Time), 상품(Item), 홈쇼핑(Supplier)이라는 3개의 차원테이블을 정의하고, 각 차원테이블에 따라 다양한 OLAP 연산을 통하여 다차원적인 분석을 수행한다. 또한 사실 테이블은 3개의 차원 테이블에 대한 키값과 상품의 트렌드를 분석하기 위하여 상품에 대한 고객평가를 측도(Measure)값으로 갖는 테이블을 말한다. 또한, 다차원 모델을 구성하기 위하여 본 논문에서는 스타 스키마(Star schema)를 다차원 분석 모델로 사용하였다[8-10].

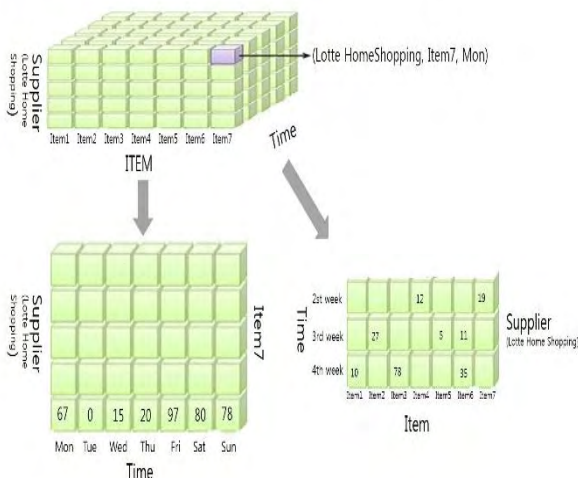


<그림 2> TV 홈쇼핑 Cube 를 위한 스타 스키마

그림 2 은 본 논문에서 TV 홈쇼핑 Cube 를 위해 정의된 스타 스키마를 표현한다. 위의 스키마에서 보는 바와 같이 사실 테이블은 3 개의 차원 테이블의 각각에 대한 키로써 연결되고 기업의 의사결정자들은 분석 목적과 원하는 관점에 따라 차원을 선택하고 각 차원 테이블에서의 개념에 대한 계층 정도를 조절함으로써 사실 테이블의 측도인 고객평가로 표현되는 분석 결과를 제공받을 수 있다.

### 3.3. TV 홈쇼핑 OLAP Cube 연산 및 상품 트렌트 분석

OLAP 은 방대한 양의 데이터들을 관리하며, 요약과 집계를 위한 도구를 제공한다. 또한 OLAP 에서는 데이터를 다각적으로 분석하기 위하여 롤-업(Roll-up), 드릴-다운(Drill-down), 슬라이스(Slice)와 다이스(Dice) 등과 같이 다양한 연산들을 제공한다. 본 논문에서는 각 상품이 어떠한 시간에 고객평가 점수가 높은지에 대해 분석하기 위하여 OLAP 연산 중 슬라이스와 롤-업 연산을 사용한다(그림 3 참조).



<그림 3> TV 홈쇼핑 Cube 에서의 OLAP 연산

(Slice, Roll-up)

### 4. 파일럿 실험 및 결과

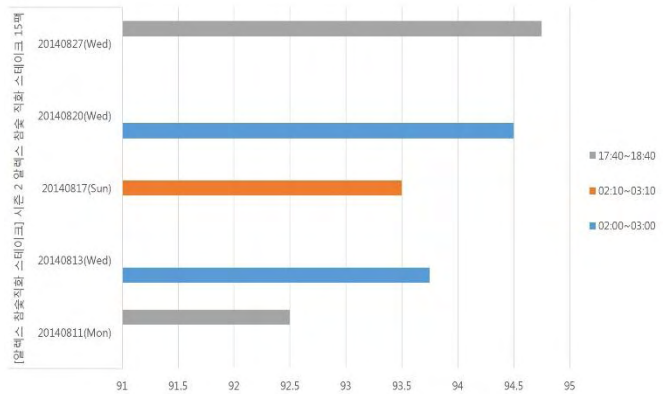
본 논문에서는 TV 홈쇼핑 데이터 기반으로 상품의 트렌드를 분석하기 위한 파일럿 실험을 수행하였다. 수집된 TV 홈쇼핑 데이터는 국내 대표적인 홈쇼핑 사이트 중 한 곳으로부터 2014 년 8 월 5 일부터 9 월 2 일까지의 TV 홈쇼핑 상품 데이터(상품명, 상품의 판매 날짜 및 시간, 고객평가 점수 등)를 수집하였다. 또한 OLAP 연산을 수행하기 위하여 Excel 에서 제공해주는



분석도구인 Pivot 테이블을 이용하였다[7].

<그림 4> 국내 TV 홈쇼핑 웹사이트

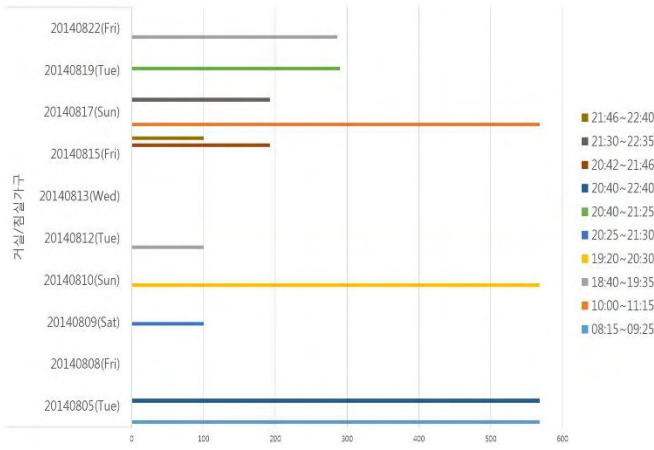
그림 4 와 같이 국내 TV 홈쇼핑에서는 판매 상품



에 대한 여러가지 정보들을 제공한다.

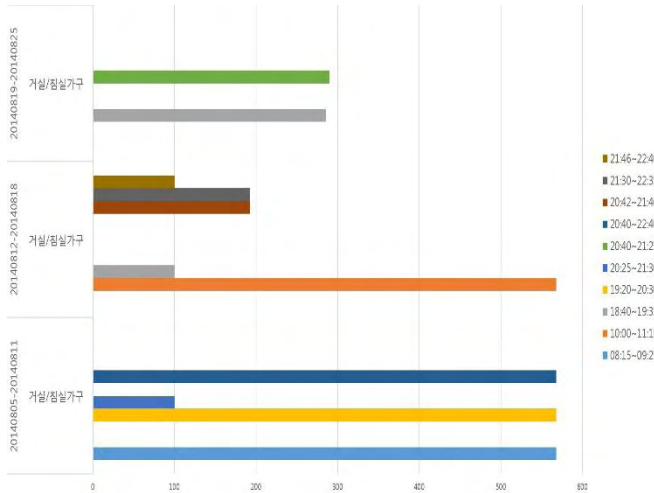
<그림 5> 특정 상품에 대한 OLAP 연산

그림 5 는 특정 상품(알렉스 참숯직화 스테이크)에 대한 8 월 5 일부터 9 월 2 일까지의 고객평가가 누적된 그래프이다. 위 그림과 같이 특정 상품(알렉스 참숯직화 스테이크)과 같은 경우에는 대체적으로 수요일에 판매되었을때 소비자들에게 좋은 고객평가를 받은 것을 알 수 있다. 또한, 아이템 차원을 높여서 봤을 경우는 그림 6 과 같다.



<그림 6>상품 카테고리별 OLAP 연산

위 그림과 같이 아이템 차원을 높여 상품의 카테고리(거실/침실가구)로 보았을 때에는 일요일 및 화요일에 상품이 판매되었을 경우 소비자들에게 좋은 고객평가를 받은 것을 알 수 있다. 또한 시간의 차원을 일주일 단위로 높여서 봤을 경우에는 다음과 같다.



<그림 7>주 단위 상품 OLAP 연산

그림 7은 시간 차원을 높인 후 특정 상품 카테고리의 상품 트렌드 분석한 결과이다. 그림에서 보는 바와 같이 거실/침실가구 상품은 8월말보다는 8월초에 더 좋은 고객평가를 받은 것을 알 수 있다.

본 논문에서는 제안한 방법을 사용하여 TV 홈쇼핑 데이터에 적용시킨 후 상품의 트렌드를 분석하였다. 분석한 결과는 비록 적은 양의 데이터이지만 상품의 트렌드를 분석하기에 적합한 것을 파일럿 실험을 통하여 보였다. 향후에는 TV 홈쇼핑 데이터 양을 충분히 수집한 후 제안한 방법을 적용할 것이다.

## 5. 결론 및 향후 과제

본 논문에서는 TV 홈쇼핑에서 판매되고 있는 상품을 대상으로 하여 OLAP을 활용함으로써 상품의 트렌드를 분석하는 방법을 제안하였다. 파일럿 실험에서는 제안한 방법을 활용하여 상품의 시간별, 상품의

카테고리별, 주단위 상품 카테고리별 고객평가 점수로 TV 홈쇼핑 상품의 트렌드를 분석하였다.

향후 연구로는 다양하고 더 많은 양의 데이터를 수집하여 제안한 방법의 성능을 평가하고 타당성을 검증할 것이다. 또한, 단순히 상품의 트렌드 분석 뿐만 아니라 특정 날씨에 따른 상품의 트렌드 변화 분석을 할 수 있는 방법을 제안하고자 한다.

## 사사

본 논문은 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 수행된 BK21 플러스 사업의 연구 결과임 (No. T1300572)

## 참고문헌

- [1] 최현정, 박준홍, and 나건, "TV 홈쇼핑 산업에서의 새로운 가치창출에 관한 연구," *디지털 디자인학연구*, vol. 13, pp. 353-362, 2013.
- [2] 양정연, 이익훈, 이수경, and 이상구, "입시 관리의 의사결정 지원을 위한 데이터 웨어하우스와 OLAP 도구의 설계 및 구현," *한국정보과학회 학술발표논문집*, pp. 173-180, 2004.
- [3] M. C. Tremblay, R. Fuller, D. Berndt, and J. Studnicki, "Doing more with more information: Changing healthcare planning with OLAP tools," *Decision Support Systems*, vol. 43, pp. 1305-1320, 2007.
- [4] S. Chaudhuri and U. Dayal, "An overview of data warehousing and OLAP technology," *ACM Sigmod record*, vol. 26, pp. 65-74, 1997.
- [5] E. F. Codd, S. B. Codd, and C. T. Salley, "Providing OLAP (on-line analytical processing) to user-analysts: An IT mandate," *Codd and Date*, vol. 32, 1993.
- [6] W. H. Inmon, "Building the data warehouse," *John wiley & so Golfarelli, M., Maio, D., & Rizzi, S.(1998). The dimensional fact model: Aconceptual model for data warehouses. International Journal of CooperativeInformation Systems*, vol. 7, pp. 215-247, 2005.
- [7] A. Hamoud and T. A. Obaid, "Using OLAP with Diseases Registry Warehouse for Clinical Decision Support," 2014.
- [8] J. Han and M. Kamber, *Data Mining, Southeast Asia Edition: Concepts and Techniques*: Morgan kaufmann, 2006.
- [9] D. Zhang, C. Zhai, and J. Han, "Topic Cube: Topic Modeling for OLAP on Multidimensional Text Databases," in *SDM*, 2009, pp. 1124-1135.
- [10] P. Zhao, X. Li, D. Xin, and J. Han, "Graph cube: on warehousing and OLAP multidimensional networks," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, 2011, pp. 853-864.