

인터넷 쇼핑몰에서 구매시점의 추천

장은실, 이용규
동국대학교 컴퓨터공학과

Recommendation of Buying Points for Internet Shopping Malls

Eun Sill Jang, Yong Kyu Lee
Dept. of Computer Engineering, Dongguk University

요 약

최근 인터넷 쇼핑몰에서 상품을 구매하는 고객들에게 편의성과 효율성을 제공하기 위하여 구매자들의 선호도나 가격에 맞는 상품을 추천해 주는 연구들이 활발하게 진행되고 있다. 그러나 이러한 상품을 추천하는 연구들은 다양하게 발전하고 있지만 추천된 상품들의 구매시점에 관한 연구는 찾아보기 어렵다. 이에 본 논문에서는 인터넷 쇼핑몰의 적극적인 마케팅 일환으로 상품을 구매할 시점을 추천해 주는 방안을 제안한다. 이를 위하여 과거의 판매 기록 데이터베이스에 있는 판매가격의 기준 시계열 패턴과 유사한 시계열 패턴을 정규화 변환된 유사도로써 검색한다. 검색된 과거 가격 패턴을 기준으로 미래 가격 패턴을 분석하여, 미래 가격 패턴의 변화에 따라 상품 구매시점을 추천한다. 또한 본 논문에서는 이러한 구매시점을 추천하는 상품 추천 시스템을 설계한다.

1. 서론

전자상거래의 보편화에 따라 인터넷 쇼핑몰들의 적극적인 마케팅 정책으로 구매자가 최선의 구매를 할 수 있도록 다양한 서비스를 제공하고 있다[10]. 이러한 서비스들은 구매자들의 선호도를 고려하거나 원하는 가격에 맞는 상품을 추천해 주는 것이 대부분이었고, 구매자들이 원하는 상품을 언제 구매하는 것이 가장 좋은 상품을 저렴하게 구매할 수 있는지에 대한 서비스는 거의 없었다. 그러나 인터넷 쇼핑몰에서 구매자들에게 구매할 상품에 대한 정보와 함께 언제 구매하는 것이 가장 저렴하게 상품을 구매할 수 있는지, 구매시점을 추천해 준다면 구매자 입장에서 유용한 정보를 제공받게 되는 것이다.

이에 본 논문에서는 구매자가 구매할 계획이 있는 상품에 대하여 저렴하게 구매할 수 있도록 판매가격의 흐름을 패턴으로 분석하여 판매가격 패턴의 변화에 따라 상품을 구매하기 가장 적절한 시점을 알려주는 구매시점의 추천 방안을 제안한다. 판매가격 패턴의 변화를 알아내기 위하여 과거의 판매 기록 데이터베이스에 있는 판매가격들을 이용하여 판매가격의 패턴과 유사한 변화를 보이는 패턴들을 검색한다. 검색한 과거 가격 패턴을 기준으로 미래의 판매가격 패턴을 분석하여 미래 판매가격 패턴의 변화에 따라 구매

시점을 추천한다. 또한 이러한 구매시점을 추천하는 상품 추천 시스템을 설계한다.

2. 관련연구

본 절에서는 인터넷 쇼핑몰의 상품 추천에 대한 소개와 유사한 가격 패턴 검색에 이용되는 시계열 데이터의 유사도 측정에 대한 관련 연구들을 소개한다.

2.1 상품 추천

최근 인터넷의 발전에 따라 인터넷 쇼핑몰들이 급증하게 되었고, 구매자들의 인터넷 쇼핑몰에 대한 이용률도 높아지게 되어 인터넷 쇼핑몰에서는 구매자들에게 적극적으로 상품을 추천하기 시작하였다. 구매자들이 원하는 상품에 대한 정보를 보다 정확하게 제공하기 위하여 구매자들의 선호, 관심, 구매기록과 같은 정보를 기초로 하여 구매자에게 가장 알맞은 정보를 제공하려는 기술들이 연구되기 시작했다.

최근의 소비시장은 적극적인 마케팅의 일환으로 상품들이 빠르게 출시되고, 그에 따라 상품들의 가격 변동 주기도 짧아졌다[10]. 따라서 인터넷 쇼핑몰에서는 이와 같은 추세를 반영하여 고객들이 상품을 보다 저렴하게 구매할 수 있도록 과거 판매가격의 변동 주기를 분석하여 구매시점 정보를 제공해야 한다. 그러나

아직까지는 구매자들이 원하는 상품을 정확하게 추천 해주려는 연구들이 대부분이고, 언제 구매하는 것이 좋은 상품을 가장 저렴하게 구매할 수 있는지에 대한 연구는 찾아보기 어렵다.

2.2 시계열 데이터의 유사도 측정

시계열 데이터(Time-Series Data)는 일정한 시간 주기마다 얻어진 연속된 실수 값들로 이루어진 데이터이며, 이러한 예로 주가 데이터, 환율 데이터, 기온 데이터, 제품 판매량 데이터, 의료 측정 데이터, 기업 성장률 데이터 등이 있다[1]. 시계열 데이터의 유사성 문제에 대한 많은 연구들이 있으며[4], 이들은 유사도 기반의 시계열 데이터 질의 방법에 대한 연구들이다 [1, 2, 5, 6, 8]. 유사한 시계열 데이터를 검색하는 기법에는 유클리디안 거리(Euclidean distance)[1, 6, 8, 9], 스케일링(Scaling)[2], 쉬프팅(Shifting)[2], 정규화(Normalization)[2, 5, 6, 8], 이동 평균(Moving Average)[2, 3, 6, 7, 8], 타임 워핑(Time Warping)[5, 8], 경계표 유사성 측정법(Landmark Similarity Measurement)[4], 모양 기반의 검색(Shape Based Search)[8] 등이 있다.

유클리디안 거리만을 이용하게 되면 원하는 시계열 데이터를 검색하지 못하는 경우가 빈번하게 발생한다. 따라서 유클리디안 거리를 이용하여 유사한 시계열 데이터를 검색하는 연구들은 유사한 시계열 데이터의 효과적인 검색을 위하여 연속되는 k개 데이터들의 평균값을 내어 순차적으로 나열하는 변환으로 전체 시계열 데이터의 잡음을 제거하는 k-이동 평균 변환과 상대적인 값을 계산하여 유사한 패턴 검색을 쉽게 하는 정규화 변환 기법을 이용하여 유사도가 높은 패턴을 검색하였다[2, 6, 7, 8].

따라서 본 논문에서는 유사도를 측정하는데 일반적으로 널리 사용되는 k-이동 평균 변환과 정규화 변환 기법을 이용한 유클리디안 거리를 계산하여 유사한 판매가격 패턴을 검색한다.

3. 유사한 가격 패턴 검색

본 절에서는 시계열 데이터인 판매가격에서 유사한 가격 패턴을 검색하기 위하여 시계열 데이터의 정규화 변환과 유사도에 대하여 살펴본다.

3.1 시계열 데이터의 정의

판매가격 전체 시계열 데이터를 D라 할 때, 기준 시계열 데이터를 Q라 하고, 부분 시계열 데이터는 S라 한다. 전체 시계열 데이터 D에서 길이가 n인 기준

시계열 데이터 Q를 선택한다. 전체 시계열 데이터의 나머지 부분 중에서 길이가 n인 모든 부분 시계열 데이터 S를 기준 시계열 데이터 Q와 비교하여 유사도 기준을 만족하는 유사한 판매가격 패턴의 부분 시계열 데이터 S를 검색한다.

3.2 가격 패턴 검색 방법

전체 시계열 데이터 D에서 기준 시계열 데이터 Q와 유사한 패턴을 갖는 부분 시계열 데이터 S를 검색하는 과정은 다음의 정규화 변환과 유사도 계산을 반복하여 수행한다.

3.2.1 시계열 데이터의 정규화

[1 단계] 전체 시계열 데이터 D에서 잡음(noise)의 영향을 제거하기 위하여 전체 시계열 데이터에 대해 k-이동 평균 변환 과정을 수행한다.

이러한 k-이동 평균 변환을 수행하는 식은 다음과 같다.

$$S_k[j] = \frac{1}{k} \times \sum_{i=j}^{j+k-1} S[i] \dots\dots\dots \text{식(1)}$$

S = 시계열 데이터 (0 ≤ i < Len(S))

k = 이동 평균 계수 (1 ≤ k < Len(S))

S_k[j] = k-이동 평균 변환 시계열 (0 ≤ j < Len(S)-k+1)

[2 단계] 전체 시계열 데이터에서 유사한 패턴을 쉽게 파악하기 위하여 정규화 변환 과정을 수행한다.

이러한 정규화 변환을 수행하는 식은 다음과 같다.

$$S'[i] = \frac{S_k[i] - \frac{Max(S) + Min(S)}{2}}{\frac{Max(S) - Min(S)}{2}} \dots\dots\dots \text{식(2)}$$

S[i] = 시계열 데이터 (0 ≤ i < Len(S))

S'[i] = 정규화 변환 시계열 (0 ≤ i < Len(S))

3.2.2 시계열 데이터의 유사도

전체 시계열 데이터의 나머지 부분 중에서 길이가 n인 모든 부분 시계열 데이터와 기준 시계열 데이터와의 유클리디안 거리를 계산하여 유사도 기준을 만족하는 부분 시계열 데이터를 검색한다.

부분 시계열 데이터 S와 기준 시계열 데이터 Q와의 유사도 측정에 사용되는 유클리디안 거리를 구하는 식은 다음과 같다.

$$L(S,Q) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \dots\dots\dots \text{식(3)}$$

S={a₁, a₂, a₃,....., a_n}, Q={b₁, b₂, b₃,....., b_n}




4. 구매시점 추천 방법

본 절에서는 앞 절에서 소개한 가격 패턴 검색 방법을 이용하여 과거의 판매가격에서 유사한 패턴으로 검색된 과거 가격 패턴을 기준으로 미래 가격 패턴을 분석한다. 분석한 미래의 가격 패턴 변화에 따른 구매시점 추천 방법에 대하여 살펴본다.

4.1 기본 패턴의 종류

기본 패턴은 표 1과 같이 3가지 종류의 패턴으로 구분한다. 구분된 패턴으로 미래 판매가격의 패턴 변화에 따라 구매시점을 추천한다.

표 1. 기본 패턴의 종류

패턴의 종류	설명
	판매가격이 상승하는 패턴
	판매가격이 유지되는 패턴
	판매가격이 하락하는 패턴

4.2. 과거 가격 패턴의 윈도우 사이즈

시계열 패턴의 길이는 주간 단위로 한다. 즉, 시계열 패턴의 길이가 5이면 5주가 된다. 이 때 시계열 패턴의 길이를 윈도우 사이즈라 한다.

본 논문에서는 과거 가격 패턴의 윈도우 사이즈 길이를 3에서 6으로 하여 기존 시계열 패턴과 유클리디안 거리를 계산한 뒤, 유사한 가격 패턴을 검색한다.




4.3 미래 가격 패턴의 윈도우 사이즈

과거의 판매가격에서 검색한 유사한 가격 패턴을 기준으로 미래의 가격 패턴을 분석하는데, 이 때 미래 가격 패턴의 윈도우 사이즈는 1주에서 5주 기간의 길이로 한다.

4.4 구매시점 추천의 방법

표 2는 과거 가격 패턴을 기준으로 분석된 미래 가격 패턴에 따라 추천한 구매시점을 나타낸 표이다.

표 2. 미래 가격 패턴에 따라 추천한 구매시점

패턴	추천된 구매시점
	현재 시점에 구매
	언제든지 구매 가능
	당분간 구매 대기

즉, 미래 가격 패턴이 상승하는 패턴이면 앞으로 판매가격이 상승할 것이므로 현재 시점에 구매하도록 추천하고, 유지되는 패턴이면 앞으로 판매 가격이 유지될 것이므로 언제든지 구매 가능으로 추천하며, 하락하는 패턴이면 앞으로 판매가격이 하락할 것이므로 당분간 구매를 대기하도록 추천한다.

5. 구매시점 추천의 성능 실험

본 절에서는 상품 구매시점을 추천하기 위하여 미래 가격 패턴의 유사도와 예측율에 대한 성능 실험을 한다.

5.1 실험 환경

실험을 하기 위하여 실제 거래된 중고 자동차 판매 정보를 수집하여 데이터베이스로 구축하였고, 여기에는 동일한 자동차 모델에 대한 554건의 판매 정보가 있으며, 1년간의 데이터가 저장되어 있다.

본 논문에서는 위의 판매 정보 데이터베이스에서 판매가격을 주간 단위로 평균을 구하여 얻어진 50개의 시계열 데이터를 대상으로 하고, 실험에 사용되는 기존 시계열 패턴은 10개를 선정하여 실험을 한다.

5.2 미래 가격 패턴의 유사도

다음 그림 1은 과거의 가격 패턴을 기준으로 분석된 미래 가격 패턴의 유클리디안 거리의 유사도를 윈도우 사이즈별로 나타낸 그림이다.

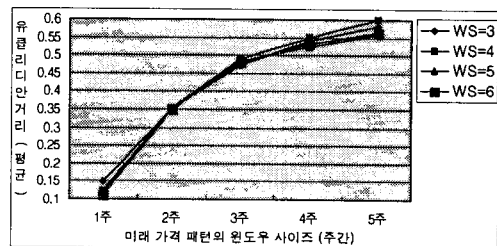


그림 1. 미래 가격 패턴의 유사도

즉, 미래의 가격 패턴의 윈도우 사이즈가 짧을수록 유클리디안 거리의 유사도가 높음을 확인할 수 있다. 또한 과거 가격 패턴의 윈도우 사이즈에는 비슷한 유사도를 보이고 있다.

5.3 미래 가격 패턴의 예측율

다음 그림 2는 과거의 가격 패턴을 기준으로 분석된 미래 가격 패턴의 상승, 유지, 하락의 변화를 예측한 비율을 윈도우 사이즈 별로 나타낸 그림이다.

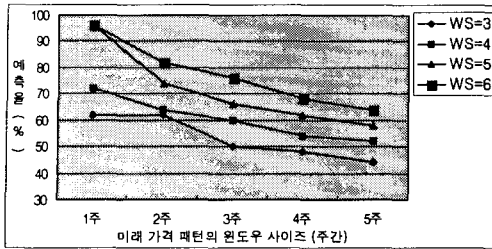


그림 2. 미래 가격 패턴의 예측율

즉, 미래 가격 패턴의 윈도우 사이즈는 짧을수록 예측율이 높았고, 1주에서 3주 사이에 60% 이상의 예측율을 보였다. 또한 과거 가격 패턴의 윈도우 사이즈는 6, 5, 4, 3 순으로 높은 예측율을 보이고 있다.

6. 상품 추천 시스템의 설계

앞 절에서 설명한 구매시점을 추천하기 위한 에이전트를 포함한 상품 추천 시스템의 전체 구조는 그림 3과 같으며, 상품 추천 시스템에는 상품 등록 서브시스템, 상품 추천 서브시스템, 구매시점 추천 에이전트, 판매 관리 서브시스템으로 구성되어 있다.

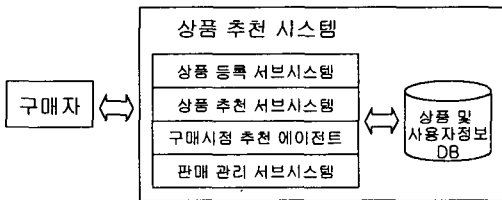


그림 3. 상품 추천 시스템의 전체 구조

구매자가 상품을 구매하려고 하면 상품 추천 시스템에서는 기존에 등록되어 있는 상품들 중에서 구매자가 원하는 상품을 검색한다. 검색한 상품은 상품 추천 서브시스템과 구매시점 추천 에이전트를 거쳐 가장 구매하기 적절한 시점 정보와 함께 상품 정보를 구매자에게 제공한다. 이렇게 제공된 상품 판매에 대한 모든 정보는 판매관리 서브시스템에서 관리한다.

7. 결론 및 향후 연구

본 논문에서는 판매가격의 패턴 변화에 따라 상품을 가장 저렴하게 구매하기 좋은 시점을 추천하는 방안을 제안하였다. 즉, 과거의 판매가격에서 가장 유사한 판매가격 패턴을 검색하여 검색한 판매가격 패턴을 기준으로 미래의 판매가격 패턴을 분석한 뒤, 판매가격의 패턴 변화에 따라 구매시점을 추천하였다. 또한 이에 대한 성능을 실험하였고, 구매시점을 추천하

는 상품 추천 시스템을 설계하였다.

이에 본 논문에서 제안한 상품 추천 시스템은 구매자의 선호도만을 고려하여 상품을 추천하던 시스템에서 구매하기 좋은 시점도 함께 추천하여 인터넷 쇼핑몰에서 구매자가 상품을 구매하는데 있어 편의성과 효율성을 제공해 줄 것으로 본다.

향후에는 3가지로 구분된 미래 가격 패턴의 종류를 확장하여 실험하고, 구매시점 추천의 효율성을 입증하는 연구가 필요하며, 본 논문에서 제안한 상품 추천 시스템과 기존에 연구되었던 선호도를 함께 제공하는 시스템을 연구할 것이다.

[참고문헌]

- [1] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient Similarity Search in Sequence Databases," Proc. of the 4th International Conference of Foundations of Data Organization and Algorithms(FODO), Chicago, Illinois, USA, pp. 69-84, 1993
- [2] R. Agrawal, K. Lin, H. S. Sawhney, and K. Shim, "Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases," Proc. of the 21st International Conference on Very Large Databases, Zurich, Switzerland, pp. 490-501, 1995
- [3] C. Chatfield, The Analysis of Time Series: An Introduction, 6th edition, CRC Press, 2003
- [4] C. S. Perng, H. Wang, S. R. Zhang, and D. S. Parker, "Landmarks: A New Model for Similarity-Based Pattern Querying in Time Series Databases," Proc. of the IEEE 16th International Conference on Data Engineering(ICDE'2000), San Diego, California, USA, pp. 33-42, 2000
- [5] D. Rafiei and A. Mendelzon, "Similarity-Based Queries for Time-Series Data," Proc. of the 1997 International Conference on Management of Data, ACM SIGMOD, Tucson, Arizona, USA, Vol. 26, No. 2, pp. 13-25, 1997
- [6] D. Rafiei and A. O. Nendelzon, "Querying Time Series Data Based On Similarity," IEEE Transactions On Knowledge And Data Engineering, Vol. 12, No. 5, pp. 675-693, 2000
- [7] 노웅기, 김상욱, 황규영, 심규석, "시계열 데이터베이스에서 임의 계수의 이동평균 변환을 지원하는 서브시퀀스 매칭 알고리즘," 정보과학회논문지, Vol. 27, No. 09, pp. 469-485, 2000
- [8] 원정임, 윤지희, 김상욱, 박상현, "시계열 데이터베이스에서 복수의 모델을 지원하는 모양 기반 서브시퀀스 검색," 정보처리학회논문지, Vol. 10-D, No. 04, pp. 577-590, 2003
- [9] 이상준, 이석호, "정규 거리에 기반한 시계열 데이터베이스의 유사 검색 기법," 정보과학회논문지, Vol. 31, No. 1, pp. 23-29, 2004
- [10] 인터넷 신문역, <http://www.jeonju.ac.kr/~sybaek/dp308.htm>, 1998.