

가격패턴에 기반한 구매시점의 추천 방법

장은실*, 이용규**

A Method of Recommending Buy Points Based on Price Patterns

Eun Sill Jang*, Yong Kyu Lee**

요약

인터넷 쇼핑몰에서 구매자들이 원하는 상품을 추천해 주는 연구들은 활발하게 진행되고 있지만, 상품의 구매시점 추천에 대한 연구는 찾아보기 어렵다. 이에 본 논문에서는 구매자에게 상품의 구매시점을 추천해 줄 수 있는 방법을 제안한다. 이를 위해서 과거의 판매 기록으로부터 정규화 된 가격패턴 데이터베이스를 구축하고, 이로부터 가격패턴을 분석한 뒤, 앞으로의 가격 변화를 예측하여 구매자에게 제시한다. 또한 가격 변화에 영향을 미치는 요소들의 분석과 이 결과를 예측에 반영하기 위하여 회귀분석을 활용한다. 그리고 성능평가를 통하여 제안하는 방법의 활용 가능성이 높음을 보인다.

Abstract

Even though much research has been performed to recommend favorite items to the buyers in the internet shopping mall, to the best of our knowledge, it is hard to find previous research on the recommendation of buy points. In this paper, we propose a method which can be used to recommend buy points of an item to the buyers. To do this, a database containing normalized price patterns is constructed from the archive of past prices. Then, the future price pattern is retrieved from the database based on the similarity. Here, regression analysis is used to find and analyze the elements that affect the price. We also present performance results showing that the proposed method can be useful for shopping malls.

▶ Keyword : 구매시점(Buy Point), 추천(Recommendation), 가격패턴(Price Pattern), 회귀분석(Regression Analysis)

• 제1저자 : 장은실 • 교신저자 : 이용규
• 접수일 : 2007.10.15, 심사일 : 2007.10.30, 심사완료일 : 2007.11.10.
* 동국대학교 산업기술연구원 전임연구원, ** 동국대학교 컴퓨터공학과 교수

I. 서론

전자상거래의 발전은 전자상거래를 이용하는 사용자에게 보다 유용한 정보를 제공하기 위한 연구로 활발히 나타나고 있다[1]. 유용한 정보를 제공한다는 것은 고객의 수요를 예측하여 정보를 제공하는 것으로써 경영 및 경제학적 관점에서 생산성 향상을 기대할 수 있다[2]. 전자상거래에서의 수요 예측은 사용자에게 제공하는 여러 정보들 중에서도 보다 저렴하게 상품을 구매할 수 있도록 하는 구매시점을 예측하는 것이다. 따라서 구매시점에 대한 연구는 더욱 실질하게 되었지만, 현재의 가격 예측을 위한 연구들은 대부분이 주가 예측에 대한 연구들이며, 전자상거래에서 구매자를 위한 상품 가격 예측에 대한 연구는 부족하다. 만약, 인터넷 쇼핑물에서 구매자들에게 구매할 상품에 대한 정보와 함께 언제 구매하는 것이 가장 저렴하게 상품을 구매할 수 있는지를 제공하여 구매 결정을 도와준다면 구매자 입장에서 유용한 정보를 제공받게 되는 것이다. 이에 본 논문에서는 구매자 입장에서 저렴하게 상품을 구매할 수 있도록 하는 구매시점 추천 방법을 제안한다.

본 논문에서는 구매 계획이 있는 상품에 대하여 구매자에게 정보를 제공할 때, 저렴하게 구매할 수 있도록 구매시점 정보를 함께 제공한다. 구매시점 정보를 추출하기 위하여 상품들의 판매가격 패턴을 분석하고, 패턴의 변화에 따라 상품을 구매하기에 가장 적절한 시점을 추출하여 제공한다. 처리 절차와 방법은 다음과 같다.

첫째, 판매가격 패턴의 변화를 알아내기 위하여 과거의 판매 기록 데이터베이스에 있는 판매가격들을 이용하여 판매가격의 패턴과 유사한 변화를 보이는 패턴들을 검색한다. 둘째, 검색한 과거 가격패턴을 분석하여 미래의 판매가격 패턴을 예측한 뒤, 미래 판매가격 패턴의 기술기에 따라 구매 결정을 도와주는 구매시점에 대한 정보를 제공한다.

기존의 구매시점 연구[3][4]들은 판매가격 패턴으로 구매시점 정보를 제공한다는 것에 의의를 두었으나, 본 논문에서는 첫째, 회귀분석을 이용하여 판매가격에 영향을 미치는 요인을 찾아내고, 둘째, 찾아낸 요인이 판매가격에 영향을 미치는 정도에 따라 가중치를 설정하여 셋째, 설정된 가중치에 따라 미래 판매가격 패턴의 기술기를 조정하여 다양한 형태의 구매시점 정보를 제공한다.

본 논문의 구성은 다음과 같다. 2장에서는 인터넷 쇼핑물에서의 상품 추천, 회귀분석, 시계열 데이터의 유사도 측정에 대한 관련 연구를 소개하고, 3장에서는 과거 판매가격의 기준이 되는 패턴과 유사한 변화를 보이는 패턴을 정규

화된 유사도로써 검색하는 과정을 설명한다. 4장에서는 검색된 가격패턴을 기준으로 미래의 가격패턴의 변화를 예측하여, 구매 결정을 도와주도록 추천 정보를 제공하는 구매시점 추천 방법을 제안한다. 5장에서는 윈도우 사이즈 분석과 적응률 분석에 대한 성능 실험을 하며, 6장에서는 결론을 맺고 향후 연구 과제를 제시한다.

II. 관련 연구

2.1 상품 추천

구매자들의 인터넷 쇼핑물에 대한 이용률이 높아지게 되자 인터넷 쇼핑물에서는 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 연구되었다. 이러한 기술들은 지능형 에이전트의 개인화 기능을 통하여 구매자의 선호도나 관심의 이동을 획득하여 구매자가 좋아할 만한 상품들의 추천 기능으로 이어진다. 즉, 구매하기 원하는 상품에 대한 정보를 제공하여 어떤 상품을 구입할 것인지를 결정할 수 있도록 도와준다.

고객에 대한 성향을 파악한 후 고객에게 적당한 상품 또는 서비스를 추천해 주는 시스템을 추천 시스템이라 한다. 오늘날 추천 시스템에 관하여 많은 연구가 활발히 진행되고 있으며, 추천 시스템에서 가장 중요한 것은 고객의 선호도를 정확하게 분석하여 고객이 원하는 가장 적절한 상품을 추천해 주는 것이다. 이는 곧 개인화된 추천 방법론의 연구로 이어지는데, 그 중에서 가장 대표적인 방법론으로 인구통계학적 기반 기법(Demographic Information-Based Technique), 지식 기반 추천 기법(Knowledge-Based Technique), 내용기반 필터링 기법(Content-Based Filtering Technique), 협업 필터링 기법(Collaborative Filtering Technique), 효용기반 추천 기법(Utility-Based Technique) 등이 있다 [5].

인터넷 쇼핑물에서는 가격 변동 주기가 짧아지는 추세를 반영하여 고객들이 상품을 보다 저렴하게 구매할 수 있도록 과거 판매가격의 변동 주기를 분석하여 구매시점을 추천하는 정보를 제공해야 한다. 그러나 아직까지는 구매자들이 원하는 상품을 정확하게 추천해주려는 연구들이 대부분이고, 언제 구매하는 것이 좋은 상품을 가장 저렴하게 구매할 수 있는지에 대한 연구는 매우 부족하다. 따라서 구매자 입장

에서 저렴하게 상품을 구매할 수 있도록 하는 구매시점 정보 제공에 대한 연구가 필요한 것이다.

2.2 회귀분석

회귀분석(Regression Analysis)은 자료에 포함된 한 변수와 또 다른 하나 이상의 변수들 간의 연관 관계를 적절한 함수식을 표현하여 자료 분석을 간편하게 하는 통계 기법이다[6].

회귀분석 방법의 기능과 용도는 기술도구로써 회귀식 추정 및 추정된 회귀식의 정확도 평가, 특정 독립변수의 종속 변수에 대한 독립적 영향력 측정, 변수들 간의 복잡한 인과 관계 메커니즘 파악 등에 사용된다. 추리도구로써는 모수 추정 및 신뢰구간 추정, 모집단에 관한 가설검증 등에 사용된다[7].

회귀분석은 사람의 라이프스타일과 수명과의 함수관계를 밝혀 이들 사이의 관계를 연구하거나, 제조환경과 생산성과의 함수관계를 밝혀 이들 사이의 관계를 연구하거나, 신제품 개발시의 시장규모 예측, 관광지 개발시 수요예측, 고객 만족도에 영향을 미치는 각종 변수들의 영향력을 검증하는데 사용된다[7].

회귀분석의 변수는 다른 변수에 영향을 주는 변수인 독립변수(Independent Variable)와 다른 변수의 영향을 받는 변수인 종속변수(Dependent Variable)로 이루어진다. 독립변수는 원인변수, 설명변수, 예측변수라고도 하고, 종속 변수는 반응변수, 결과변수라고도 한다[8].

회귀분석의 종류에는 독립변수가 하나일 때, 독립변수와 종속변수간의 선형관계에 관한 분석인 단순회귀분석(Simple Regression)과 종속변수에 영향을 미치는 변수가 여러 개일 때, 이들 독립변수들과 종속변수간의 선형관계에 관한 분석인 다중회귀분석(Multiple Regression)이 있다[9].

회귀방정식이 독립변수와 종속변수 간의 관계를 잘 말해 주는지 즉 종속 변수를 얼마나 잘 설명해 주는지를 적합도 검증 방법을 통하여 회귀 모형의 정확도를 평가하고, t-검증과 F-검증을 통하여 유의성 검증을 평가한다. 적합도란 도출된 표본회귀선이 자료를 얼마나 잘 나타내고 있는가의 정도를 말한다. 이러한 적합도를 측정하는 방법은 기본적으로 두 가지가 있는데 그 중 하나가 추정의 표준오차, 즉 잔차를 가지고 평가하는 방법이고, 나머지 하나가 결정계수를 가지고 평가하는 방법이다[8][9].

2.3 시계열 데이터의 유사도 측정

시계열 데이터(Time Series Data)는 일정한 시간 주기마다 얻어진 연속된 실수 값들로 이루어진 데이터로 증가, 하락, 기온, 제품 판매량, 의료 측정, 기업 성장을 등과 관련된 데이터가 그 예이다[10]. 현재 이러한 시계열 데이터의 유사성 문제[11]와 유사도 기반의 시계열 데이터 질의 방법에 대한 관련 연구들이 진행 중에 있다[10][12][13][14][15]. 유사한 시계열 데이터를 검색하는 기법에는 유클리디안 거리(Euclidean Distance)[10][14][15][16][17], 스케일링(Scaling)[12], 쉬프팅(Shifting)[12], 정규화(Normalization) [12][13][14][15][16], 이동 평균(Moving Average)[12] [14][15][18][19], 타임 워핑(Time Warping)[13][15], 경계표 유사성 측정법(Landmark Similarity Measurement) [11], 모양 기반 검색(Shape Based Search)[15] 등이 있다.

유클리디안 거리만을 이용하게 되면 원하는 시계열 데이터를 검색하지 못하는 경우가 빈번하게 발생한다. 따라서 유클리디안 거리를 이용하여 유사한 시계열 데이터를 검색하는 관련 연구들에서는 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 검색하는 방법을 사용하였다[12][14][15][19].

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

III. 가격패턴 검색

본 절에서는 시계열 형태인 판매가격에서 유사한 가격패턴을 검색하기 위한 시계열 데이터의 정규화 변환 과정과 유사도에 대하여 살펴본다. 본 과정은 구매를 의뢰한 현재 시점을 중심으로 볼 때 과거 가격패턴에서 이루어진다.

3.1 시계열 데이터

과거의 판매 기록 데이터베이스에 있는 판매가격을 전체 시계열 데이터로 이용한다. 이 중에서 기준 시계열 데이터를 정하고, 이와 유사한 부분 시계열 데이터를 검색하게 되는데, 이때 표현되는 각각의 시계열 데이터를 그림으로 나타내면 다음 (그림 1)과 같다.

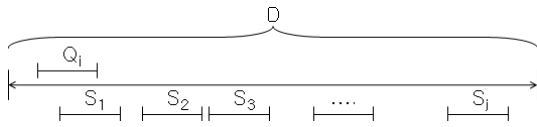


그림 1. 시계열 데이터 구성
Fig. 1 Time Series Data

D = 판매가격 전체 시계열 데이터,
Q_i = 기준 시계열 데이터, S_j = 부분 시계열 데이터

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. 시계열 데이터의 검색 과정을 정리하면 다음과 같다.

- ① D에서 Q를 선택 (Q의 길이는 n)
- ② D에서 S와 Q를 비교 (S의 길이는 n)
- ③ 유사도 기준을 만족하는 유사한 판매가격 패턴의 S를 검색

전체 시계열 데이터 D에서 기준 시계열 데이터 Q와 유사한 패턴을 갖는 부분 시계열 데이터 S를 검색하기 위해서 시계열 데이터의 정규화 변환과 유사도 계산 과정을 반복하여 수행한다.

3.2 시계열 데이터의 정규화

[1단계] 전체 시계열 데이터 D에서 잡음(noise)의 영향을 제거하기 위하여, 전체 시계열 데이터에 대해 k-이동 평균 변환 과정을 식(3.1)을 이용하여 수행한다. (Len(S)은 시계열 데이터의 길이)

$$S_k(j) = \frac{1}{k} \times \sum_{i=j}^{j+k-1} S[i] \dots\dots\dots (3.1)$$

S = 원본의 시계열 데이터 (0 ≤ i < Len(S))
k = 이동 평균 계수 (1 ≤ k < Len(S))
S_k(j) = k-이동 평균 변환된 시계열 데이터 (0 ≤ j < Len(S)-k+1)

[2단계] 전체 시계열 데이터에서 유사한 패턴을 검색하기 위하여, 식(3.2)를 이용하여 정규화 변환 과정을 수행한다.

$$S'(i) = \frac{S[i] - \frac{Max(S) + Min(S)}{2}}{\frac{Max(S) - Min(S)}{2}} \dots\dots\dots (3.2)$$

Max(S) 및 Min(S)의 S = k-이동 평균 변환된 시계열 데이터 (0 ≤ i < Len(S))
S(i) = 시계열 데이터 (0 ≤ i < Len(S))
S'(i) = 정규화 변환 시계열 (0 ≤ i < Len(S))

정규화 변환을 수행하는 알고리즘은 다음과 같다.

```

알고리즘 : normalization_calculation
입력 : 전체 시계열 데이터, 이동 평균 계수
출력 : 정규화 변환 시계열
{
  /* 전체 시계열 데이터 D에서 잡음(noise)의 영향을 제거
  하기 위하여 전체 시계열 데이터에 대해 k-이동 평균
  변환 과정을 수행 */
  while(전체 시계열 데이터) {
    k-이동 평균 변환의 식(3.1) 계산;
    순차적으로 전체 시계열 데이터를 갱신;
  }
  /* k-이동 평균 변환 과정을 수행한 전체 시계열 데이터
  에서 유사한 패턴을 쉽게 파악할 수 있도록 시계열
  데이터의 상한과 하한을 각각 +1과 -1로 고정하는 정
  규화 변환 과정을 수행 */
  while(k-이동 평균 변환 과정을 수행한 전체 시계열 데이터) {
    정규화 변환의 식(3.2) 계산;
    순차적으로 전체 시계열 데이터를 갱신;
  }
  return 정규화 변환 시계열;
}
    
```

알고리즘 1. 시계열 데이터의 정규화 알고리즘
Algorithm 1. Normalization Algorithm of Time Series Data

3.3 시계열 데이터의 유사도

전체 시계열 데이터에서 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 거리는 식(3.3)을 이용하여 측정한다.

$$L(S, Q) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2} \dots\dots\dots (3.3)$$

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

유사도를 수행하는 알고리즘은 다음과 같다.

```

알고리즘 : similarity_search
입력 : 정규화 변환 시계열 데이터,
        부분 시계열 데이터, 기준 시계열 데이터
출력 : 유사 검색 시계열 데이터
{
  /* 전체 시계열 데이터 D에서 부분 시계열 데이터 S와
  기준 시계열 데이터 Q와의 유사도 측정에 사용되는
  유클리디안 거리를 구하여 유사도 계산 과정을 수행
  */
  while(전체 시계열 데이터) {
    S와 Q와의 유클리디안 거리로 유사도 측정 식(3.3) 계산;
    if(유사도 < 임계치)
      유사 패턴 추출;
  }
  return 유사 패턴;
}
    
```

알고리즘 2. 시계열 데이터의 유사도 알고리즘
Algorithm 2. Similarity Algorithm of Time Series Data

IV. 구매시점 추천

중고 상품을 경매 사이트에서 구매하는 경우 동일 상품이더라도 판매되는 가격은 항상 다르게 나타난다. 과거에 거래된 판매 데이터베이스를 이용하여 가격의 흐름을 미리 예측한 후 구매자에게 상품을 저렴하게 구매할 수 있는 시점 정보를 제공한다면 구매자에게 유용한 정보를 제공하게 되는 것이다. 따라서 본 절에서는 구매자가 원하는 상품을 가장 저렴하게 구매할 수 있는 시점을 추천해 주기 위한 방법을 제안한다. 제안 방법은 과거 가격패턴을 기준으로 미래의 가격패턴 변화를 예측하여 미래 판매가격 패턴의 변화에 따라 구매시점을 추천한다. 구매시점의 추천 과정은 구매를 의뢰한 현재 시점을 중심으로 볼 때 미래 가격패턴에서 이루어진다.

4.1 가격패턴

4.1.1 과거 가격패턴과 미래 가격패턴

〈그림 2〉는 구매를 의뢰한 현재 시점을 중심으로 과거 가격패턴과 미래 가격패턴을 표현한 것이다. 과거 가격패턴을 유사도에 의해 분석한 뒤, 미래 가격패턴의 변화를 예측하여 구매자에게 구매 결정을 위한 구매시점 정보를 제공할 때, 정보를 요청한 현재 시점을 중심으로 구매시점의 패턴을 분석한다.

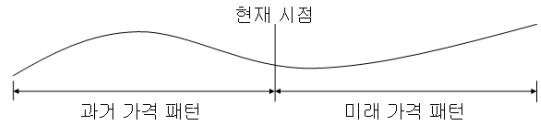


그림 2. 과거 가격패턴과 미래 가격패턴
Fig. 2 Past Price Pattern and Future Price Pattern

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

본 논문에서는 과거 가격패턴의 윈도우 사이즈 길이를 3에서 6으로 하여 기준 시계열 패턴과 유클리디안 거리를 계산한 뒤, 유사한 가격패턴을 검색한다. 이 때 윈도우 사이즈 길이를 3에서 6의 길이로 정한 것은 패턴의 길이가 짧을수록 패턴이 단순해져서 대부분의 패턴과 유사한 패턴으로 검색이 되고, 패턴의 길이가 길어질수록 패턴이 복잡해져서 유사한 패턴이 검색되지 않기 때문이다.

과거의 판매가격에서 검색한 유사한 가격패턴을 기준으로 미래의 가격패턴을 예측하는데, 이 때 예측하는 미래 가격패턴의 기간은 1주에서 5주 기간을 본다.

4.1.2 패턴의 종류

3가지 종류의 기본 패턴을 기준으로 미래 판매가격의 패턴 변화에 따른 구매시점을 추천하게 되는데, 1단계 패턴 결합에 따라 다음 〈표 1〉과 같이 9가지로 표현할 수 있다.

표 1. 패턴의 종류
Table. 1 Types of Patterns

기본 패턴	확장된 패턴		
 판매가격이 상승하는 패턴	 상승후 상승(II)	 상승후 유지(IIM)	 상승후 하락(ID)
 판매가격이 유지되는 패턴	 유지후 상승(MI)	 유지후 유지(MM)	 유지후 하락(MD)
 판매가격이 하락하는 패턴	 하락후 상승(DI)	 하락후 유지(DM)	 하락후 하락(DD)

4.1.3 패턴의 일반화

시계열 데이터의 복잡하고 다양한 패턴을 단순 회귀식으로 표현하여 패턴을 일반화한다. 〈그림 3〉은 상승 패턴을 일반화 시킨 예로써, ① 패턴의 시작 부분인 m과 끝 부분인 n을 직선으로 이어 ②의 일반화된 패턴으로 유도하여 식 (4.1)로 표현한다. 또한 이러한 패턴의 일반화는 상승, 유지, 하락 패턴에도 동일한 방법으로 적용한다.

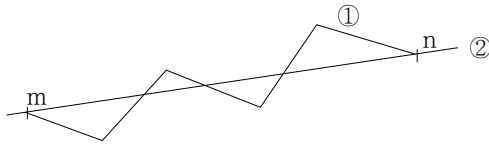


그림 3. 상승패턴의 일반화 예
Fig. 3 Generalization of Increment Pattern

$$Y_i = \alpha + \beta X_i$$

X_i = 시간(주간단위), Y_i = 판매가격,
 α = 상수, β = 기울기

4.2 구매시점 추천

〈표 2〉는 과거 가격패턴을 기준으로 예측된 미래 가격패턴에 따라 구매시점의 추천 정보를 나타낸 표이다.

표 2. 미래 가격패턴에 따라 추천한 구매시점 정보
Table. 2 Recommendation of Buying Point for Future Price Pattern

미래 가격패턴	추천된 구매시점 정보
상승하는 패턴	현재 시점에 구매
유지하는 패턴	언제든지 구매 가능
하락하는 패턴	당분간 구매 대기

즉, 미래 가격패턴이 상승하는 패턴이면 앞으로 판매가격이 상승할 것이므로 '현재 시점에 구매'하도록 추천하고, 유지되는 패턴이면 앞으로 판매 가격이 유지될 것이므로 '언제든지 구매 가능'으로 추천하며, 하락하는 패턴이면 앞으로 판매가격이 하락할 것이므로 '당분간 구매 대기'하도록 추천한다.

4.2.1 패턴의 기울기에 의한 추천

〈표 3〉은 미래 가격패턴의 상승, 유지, 하락에 따른 기본적인 구매시점의 정보에서 미래 가격패턴의 기울기 정도에 따라 구매시점 정보를 추천 정도로 나타낸 것이다. 이는 단순한 구매시점 정보 제공에서 기울기에 따라 구매시점에 대한 정보를 추천 정도로 제공해 주는 것이다. 패턴의 기울기 범위는 평균 기울기값(0.15)에서 2배 이상의 변화 폭에 따라 세분화한 것이다. 추천 정도는 패턴의 기울기가 유지될 정도로 0과 1사이의 값으로 표현되는데, 1에 가까울수록 추천의 강도는 높은 것이다.

표 3. 패턴의 기울기에 따른 추천 내용
Table. 3 Recommendation for Slope of Pattern

구분	패턴	기울기(β)	구매시점의 추천	추천 정도(r)
상승	급격 상승	$\beta \geq 0.6$	현재시점에 구매	$r \geq 0.69$
	보통 상승	$0.3 \leq \beta < 0.6$		$0.62 \leq r < 0.69$
	완만 상승	$0.15 \leq \beta < 0.3$		$0.56 \leq r < 0.62$
유지		$-0.15 \leq \beta < 0.15$	언제든지 구매 가능	$r \geq 0.81$
하락	완만 하락	$-0.3 \leq \beta < -0.15$	당분간 구매대기	$0.58 \leq r < 0.64$
	보통 하락	$-0.6 \leq \beta < -0.3$		$0.64 \leq r < 0.70$
	급격 하락	$\beta \leq -0.6$		$r \geq 0.70$

4.2.2 가중치에 의한 기울기 조정

실제 거래된 중고 자동차 판매 정보의 데이터베이스를 이용하여 SPSS 통계 프로그램[20]으로 회귀분석을 한 결과 판매월이 가장 영향력이 큰 요인으로 나타났고, 그 다음은 주행거리로 나타났다.

〈표 4〉의 분석된 계수를 보면 3월의 베타계수의 값이 가장 크므로 통계 모형에 대한 신뢰도 또한 가장 높은 변수임을 알 수 있다. 각 독립변수의 통계적 유의성을 판단하기 위하여 독립변수마다 t값이 산출되어 있다. t값이 클수록 유의성이 높다는 것을 의미하는데, 본 예의 경우에는 3월, 4월, 5월, 6월, 7월, 9월, 10월이 통계적으로 유의하게 나타났다.(유의확률(0.05) 이는 주행거리와 판매월 중에서는 3월, 4월, 5월, 6월, 7월, 9월, 10월이 가장 의미가 있다는 것이다. 〈표 4〉에 2월과 12월의 자료가 없는 것은 통계 분석을 통하여 의미가 많이 떨어지는 변수는 제거되었기 때문이다.

표 4. 분석된 계수
Table. 4 Analyzed Coefficient

판매월	표준화 계수	t	유의확률
	베타		
1월	0.028	0.997	0.319
3월	0.256	7.605	0.000
4월	0.176	5.394	0.000
5월	0.206	5.753	0.000
6월	0.154	4.515	0.000
7월	0.109	3.090	0.002
8월	0.064	1.863	0.063
9월	0.104	3.035	0.002
10월	0.144	4.173	0.000
11월	0.058	1.660	0.097

판매가격에 영향을 미치는 요인들 중에서 판매월이 가장 의미가 있다는 SPSS 통계 프로그램의 분석 결과에 따라 판

매월에 가중치를 부여한다. 가중치를 부여한 조정 기울기 표현식은 식(4.2)와 같다.

조정 기울기 $Y_i = \alpha + \beta(w)X_i$ 24
 X_i = 시간(월단위), Y_i = 판매가격,
 α = 상수, β = 기울기, w = 가중치

4.2.3 구매시점 추천 방법

판매월이 9월과 10월에 소폭 상승하였으므로 상승 가중치(w) = 2로 하여 기울기를 조정한다. 또한 3월부터 8월까지는 하락 가중치(w) = -2로 하여 기울기를 조정한다. 여기서 기울기는 전체 가격패턴의 평균 기울기 값인 0.15에서 2배씩 조정하도록 한 것이다. 기울기의 가중치 변수를 2로 한 것은 패턴의 기울기가 유지될 확률이 가장 높았기 때문이다.

구매자가 9월에 구매를 의뢰하게 된다면 9월은 가격패턴이 소폭 상승함으로 가중치 2를 반영하여 기울기를 조정하게 된다. 조정된 기울기는 가격패턴이 완만 상승이었다면 보통 상승으로, 보통 상승이었다면 급격 상승으로 패턴을 수정한다. 이렇게 수정된 패턴으로 다음 <표 5>와 같이 구매시점을 추천한다.

표 5. 가중치를 반영한 구매시점의 추천 내용
 Table. 5 Recommendation of Buying Point Reflected Weight

구분	조정전 패턴	조정후 패턴	기울기(β)	구매시점의 추천	추천 정도(r)
상승	급격 상승	초급격 상승	$\beta \geq 1.2$	현재시점에 구매	$r \geq 0.71$
	보통 상승	급격 상승	$0.6 \leq \beta < 1.2$		$0.69 \leq r < 0.71$
	완만 상승	보통 상승	$0.3 \leq \beta < 0.6$		$0.62 \leq r < 0.69$
	유지	완만 상승	$0.15 \leq \beta < 0.3$		$0.56 \leq r < 0.62$
유지			$-0.15 \leq \beta < 0.15$	언제든지 구매 가능	$r \geq 0.81$
하락	유지	완만 하락	$-0.3 \leq \beta < -0.15$	당분간 구매대기	$0.58 \leq r < 0.64$
	완만 하락	보통 하락	$-0.6 \leq \beta < -0.3$		$0.64 \leq r < 0.70$
	보통 하락	급격 하락	$-1.2 \leq \beta < -0.6$		$0.70 \leq r < 0.77$
	급격 하락	초급격 하락	$\beta \leq -1.2$		$r \geq 0.77$

V. 성능 실험

본 절에서는 예측한 미래 가격패턴의 유사도, 예측율을 통하여 구매시점 추천에 대한 신뢰성을 확인하고, 추천 적중률과 절감된 구매비 평균값에 대한 성능 실험을 통하여 제안하는 방법의 활용 가능성이 높음을 살펴본다.

5.1 실험 환경

성능 실험을 위하여 실제 거래된 중고 자동차 판매 정보를 수집하여 데이터베이스로 구축하였다(21). 데이터베이스에는 동일한 자동차 모델에 대한 1271건의 판매 정보가 있으며, 3년간의 데이터가 저장되어 있다. 본 논문에서는 위의 판매 정보 데이터베이스를 기반으로 판매가격을 주간 단위로 평균을 구한 150개의 시계열 데이터와 기준 시계열 패턴 9개를 선정하여 실험하였다.

<표 6>은 실험에 사용되는 기준 시계열 패턴 9가지와 각 패턴의 기울기이다. 패턴의 기울기는 상승의 3가지 패턴, 유지의 3가지 패턴, 하락의 3가지 패턴에 대응되는 값이다.

표 6. 실험에 사용되는 패턴과 각 패턴의 기울기
 Table. 6 Slope of Patterns and Patterns

패턴			기울기
상승후 상승 (II)	상승후 유지 (IM)	상승후 하락 (ID)	$\beta \geq 0.6$ $0.3 \leq \beta < 0.6$ $0.15 \leq \beta < 0.3$
유지후 상승 (MI)	유지후 유지 (MM)	유지후 하락 (MD)	$-0.15 \leq \beta < 0.15$
하락후 상승 (DI)	하락후 유지 (DM)	하락후 하락 (DD)	$-0.15 \geq \beta > -0.3$ $-0.3 \geq \beta > -0.6$ $\beta \leq -0.6$

5.2 실험 방법

윈도우 사이즈 분석 실험을 통하여 구매시점 추천에 대한 신뢰성을 확인하고, 적중률 분석 실험을 통하여 구매자가 얻을 수 있는 이윤 등에 대한 성능 실험을 한다.

성능 평가 척도로 일반적으로 많이 사용하는 MAE (Mean Absolute Error)(22)[23]를 이용하여 오차 정도를 측정한다. MAE를 통하여 산출된 E 값은 그 값이 작을수록 오차의 크기가 작으며, MAE 공식은 식(5.1)과 같다. MAE는 150건에 대하여 구매시점 추천없이 구매가 이루어졌을 때와 구매시점 추천 정보에 의해 구매가 이루어졌을 때와의 오차를 측정하는데 사용한다.

$$E = \frac{1}{n} \sum_{i=1}^n (|b_i - r_i|) / b_i \dots\dots\dots (5.1)$$

n = 추천의 수
 b_i = 실제 패턴의 기울기
 r_i = 추천 패턴의 기울기

5.3 실험 결과

5.3.1 윈도우 사이즈 분석

과거의 가격패턴을 기준으로 분석된 미래 가격패턴의 유클리디안 거리의 유사도를 윈도우 사이즈별로 나타낸 그래프는 <그림 4>와 같다. 즉, 미래의 가격패턴의 윈도우 사이즈가 짧을수록 유클리디안 거리의 유사도가 높음을 확인할 수 있다. 또한 과거 가격패턴의 윈도우 사이즈에는 비슷한 유사도를 보이고 있다.

과거의 가격패턴을 기준으로 분석된 미래 가격패턴의 상승, 유지, 하락의 변화를 예측한 비율을 윈도우 사이즈 별로 나타낸 그래프는 <그림 5>와 같다. 즉, 미래 가격패턴의 윈도우 사이즈는 짧을수록 예측율이 높았고, 1주에서 3주 사이에 60% 이상의 예측율을 보였다. 또한 과거 가격패턴의 윈도우 사이즈는 6, 5, 4, 3 순으로 높은 예측율을 보이고 있다.

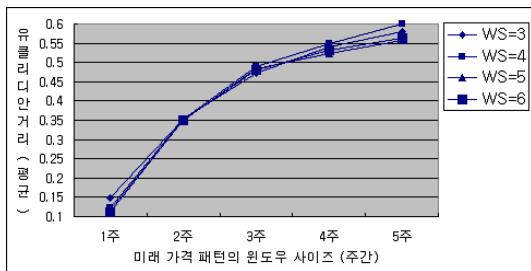


그림 4. 미래 가격패턴의 유사도
 Fig. 4 Similarity of Future Price Pattern

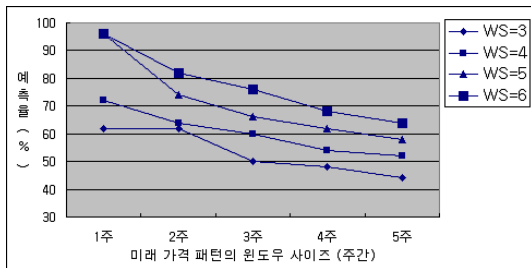


그림 5. 미래 가격패턴의 예측율
 Fig. 5 Prediction Rate of Future Price Pattern

5.3.2 패턴별 오차 및 적중률 분석

<그림 6>은 E값을 구하는 공식에 의해서 산출된 그래프이다. 즉, 150건에 대해 구매시점 추천없이 구매가 이루어졌을 때와 구매시점 추천 정보에 의해 구매가 이루어졌을 때와의 오차값을 나타낸 것이다. 그래프를 보면 가격이 상승할 때보다는 가격이 유지되거나 하락할 때 비교적 E값이 작음을 알 수 있다. <그림 7>은 <그림 6>의 그림에서 나타난 값들의 평균을 구한 것으로 MAE 공식에 의해서 산출된 그래프이다.

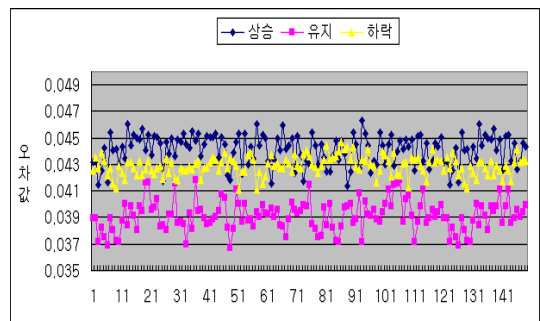


그림 6. 패턴별 오차값
 Fig. 6 Error of Patterns

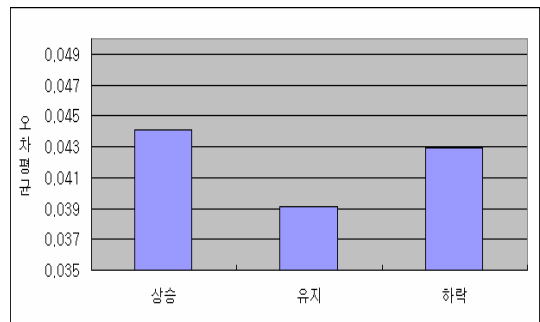


그림 7. 패턴별 오차 평균
 Fig. 7 Error Average of Patterns

구매시점 추천의 적중률을 비교한 그래프는 <그림 8>과 같다. 이 그래프는 기본 패턴 정보만으로 추천했을 때보다 기울기나 판매가격에 영향을 미치는 요인인 판매월에 가중치 조절을 하여 추천했을 경우 더 높은 적중률을 보이고 있음을 확인할 수 있다.

구매시점을 추천할 때 1주부터 5주 사이의 가격 변화에 따라 구매시점을 추천할 경우의 적중률을 보여주는 그래프는 <그림 9>와 같다. 이 그래프에서 보듯이 구매를 의뢰한 시점에 가까운 미래에 대한 추천에 대한 적중률이 높음을

확인할 수 있다. 또한 가격 상승할 때보다는 하락이나 유지에 대한 구매시점 추천에 대하여 높은 적중률을 보이고 있다.

구매자 입장에서 구매시점 추천에 대한 정보를 제공 받았을 때 절감된 구매비 평균값을 비교한 그래프는 <그림 10>과 같다. 이 그래프는 기본 패턴 정보만으로 추천받았을 때보다는 기울기나 판매가격에 영향을 미치는 요인인 판매월에 가중치 조정을 하여 추천받았을 경우 구매비가 더 많이 절감됐음을 확인할 수 있다.

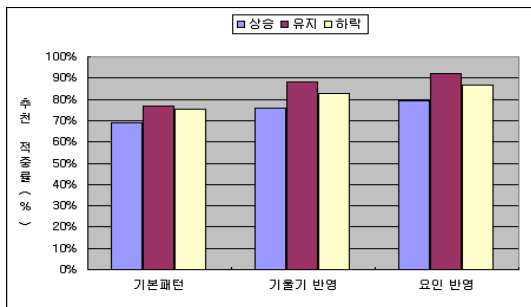


그림 8. 추천 적중률 비교
Fig. 8 Comparison for Recommendation Hit Rate

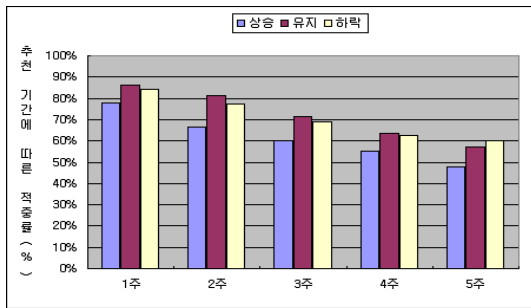


그림 9. 추천 기간에 따른 적중률
Fig. 9 Hit Rate for a Period of Recommendation

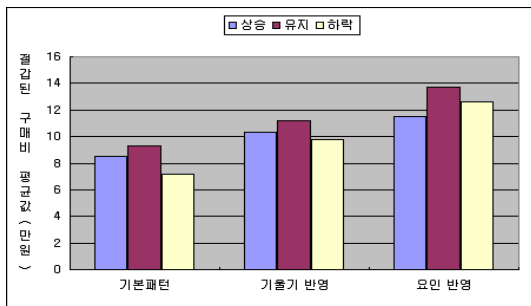


그림 10. 절감된 구매비 평균값
Fig. 10 Saved Buying Cost Average

VI. 결론 및 향후 연구

본 논문에서는 상품을 저렴하게 구매할 수 있도록 구매시점 정보를 제공하는 방안을 제안하였다. 구매시점 정보를 알아내기 위하여 과거의 판매 기록 데이터베이스에 있는 판매가격들을 이용하여 판매가격의 패턴과 유사한 변화를 보이는 패턴들을 검색하고, 검색한 과거 가격패턴을 기준으로 미래의 판매가격 패턴을 예측하여 미래 판매가격 패턴의 변화에 따라 구매 결정을 도와주는 구매시점에 대한 정보를 제공한다. 이때 미래 판매가격 패턴 변화의 예측은 회귀분석을 이용하여 얻어진 판매가격에 영향을 미치는 판매월에 가중치를 반영하여 처리한 구매시점 정보를 추천하였다.

본 논문에서 제안한 구매시점의 추천은 성능 실험을 통하여 미래 가격패턴의 윈도우 사이즈가 짧을수록 추천한 구매시점의 적중률이 높은 것으로 나타났다. 즉, 구매시점 추천에 대한 적중률은 가까운 미래일수록 높다는 것이다. 패턴에 대한 적중률은 유지, 하락, 상승 순으로 높게 나타났다. 이러한 적중률은 기본 패턴 정보만으로 추천받았을 때보다는 기울기나 판매가격에 영향을 미치는 요인인 판매월에 가중치 조정을 하여 추천받았을 경우 구매자에게 구매비를 5% 정도 절감되게 하는 것으로 나타났다.

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

향후에는 패턴의 기울기에 따른 구매시점 추천시 기울기에 대한 검증과, 추천 정도에 대한 실험을 통하여 구매 추천의 효율성 입증에 대한 실험을 할 것이며, 표준편차와 삼각함수를 이용하여 구매시점의 추천에 대한 신뢰성을 실험할 것이다. 또한 본 논문에서 제안한 구매시점 추천에 대한 기능을 포함한 구매시점 추천 에이전트 및 상품 추천 시스템과 기존에 연구되었던 선호도를 함께 제공하는 시스템을 연구할 것이다.

참고문헌

- [1] 인터넷 신문역, <http://search.mk.co.kr>, 2005.
- [2] 이순룡, 제품·서비스 생산관리론, pp. 156-188, 법문사, 2005.
- [3] E. S. Jang, Y. K. Lee, "A Method of Recommending Buying Points for Internet Shopping Malls," KES2006, Lecture Notes in Artificial Intelligence (LNAI), vol.

- 4251, pp. 992-1001, 2006.
- [4] 장은실, 이용규, "인터넷 쇼핑물에서 구매시점의 추천," 한국멀티미디어학회 '04 춘계 학술발표논문집, pp. 491-494, 순천향대학교, 2004.
- [5] 안재명, RFID기반 쇼핑동선 분석을 이용한 실시간 선호 상품 추천에 관한 연구, 숭실대학교 대학원 박사학위논문, 2005.
- [6] D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, 3rd edition, pp. 13-66, Wiley-Interscience, 2001.
- [7] 신건권, "회귀분석기법을 적용한 호텔기업의 원가 추경과 예측," 회계연구, Vol. 6, No. 1, pp. 245-273, 2001.
- [8] 전용진, SPSS 통계, pp. 315-352, 학문사, 2002.
- [9] 고상복, 소프트웨어 프로젝트 개발 일정 예측을 위한 통계적 다중회귀 분석 모형 연구, 중앙대학교 대학원 박사학위논문, 2005.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] D. Raffei 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.
- [14] D. Raffei 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.
- [15] 원정임, 윤지희, 김상욱, 박상현, "시계열 데이터베이스에서 복수의 모델을 지원하는 모양 기반 서브시퀀스 검색," 정보처리학회논문지, Vol. 10-D, No. 4, pp. 577-590, 2003.
- [16] 유승근, 이상호, "한국 주식 데이터를 이용한 서브시퀀스 매칭 방법의 효과성 평가," 한국정보처리학회논문지, Vol. 12-D, No. 3, pp. 355-364, 2005.
- [17] 이상준, 이석호, "징구 거리에 기반한 시계열 데이터베이스의 유사 검색 기법," 정보과학회논문지, Vol. 31, No. 1, pp. 23-29, 2004.
- [18] C. Chatfield, The Analysis of Time Series: An Introduction, 6th edition, pp. 169-201, CRC Press, 2003.
- [19] 노웅기, 김상욱, 황규영, 심규석, "시계열 데이터베이스에서 임의 계수의 이동평균 변환을 지원하는 서브시퀀스 매칭 알고리즘," 정보과학회논문지, Vol. 27, No. 9, pp. 469-485, 2000.
- [20] 채서일, 김선철, 최수호, SPSS/WIN을 이용한 통계분석(3판), pp. 76-251, B&M books, 2005.
- [21] 오토와이즈, <http://www.autowise.co.kr/>, 2007.
- [22] 고민정, 경매 시스템을 위한 시계열 분석 기반의 낙찰 예정가 추천 방법, 동국대학교 대학원 박사학위논문, 2005.
- [23] 박성은, 입찰시스템을 위한 원가산정법 기반의 입찰가 추천 에이전트, 동국대학교 대학원 박사학위논문, 2005.

저 자 소 개

장은실
 2001년 동국대학교 컴퓨터교육전공(석사)
 2007년 동국대학교 컴퓨터공학과(공학박사)
 2007년~현재 동국대학교 산업기술연구원
 전임연구원
 관심분야: 유비쿼터스 물류관리 시스템, 데이터베이스 시스템, e-비즈니스 시스템, 텔레매틱스 등



이용규
 1986년 동국대학교 전자계산학과(학사)
 1988년 한국과학기술원 전산학과(석사)
 1996년 Syracuse University(전산학박사)
 1978년~1983년 행정직 국가공무원
 1988년~1993년 국방정보체계연구소 선임연구원
 1996년~1997년 한국통신 선임연구원
 2002년~2003년 콜로라도대학교 컴퓨터학과 방문교수
 1997년~현재 동국대학교 컴퓨터공학과 교수
 관심분야: 데이터베이스 시스템, XML, 정보검색, e-비즈니스 시스템

