

# 데이터마이닝 알고리즘을 이용한 제품수명주기 예측 : 의류산업 적용사례

이슬기<sup>1</sup> · 강지훈<sup>1</sup> · 이한규<sup>1</sup> · 주태우<sup>1</sup> · 오시연<sup>2</sup> · 박성욱<sup>2</sup> · 김성범<sup>1\*</sup>

<sup>1</sup>고려대학교 산업경영공학과 / <sup>2</sup>삼성에버랜드 패션사업부

## Prediction of Product Life Cycle Using Data Mining Algorithms : A Case Study of Clothing Industry

Seulki Lee<sup>1</sup> · Ji Hoon Kang<sup>1</sup> · Hankyu Lee<sup>1</sup> · Tae Woo Joo<sup>1</sup> · Shawn Oh<sup>2</sup> · Sungwook Park<sup>2</sup> · Seoung Bum Kim<sup>1</sup>

<sup>1</sup>School of Industrial Management Engineering, Korea University

<sup>2</sup>Fashion Business of Samsung Everland Inc.

Demand forecasting plays a key role in overall business activities such as production planning, distribution management, and inventory management. Especially, for a fast-changing environment of the clothing industry, logical forecasting techniques are required. In this study, we propose a procedure to predict product life cycle using data mining algorithms. The proposed procedure involves three steps : extracting key variables from profiles, clustering, and classification. The effectiveness and applicability of the proposed procedure were demonstrated through a real data from a leading clothing company in Korea.

**Keywords:** Product Life Cycle, Demand Forecasting, Clothing Industry, Data Mining

### 1. 서론

비즈니스에서의 수요예측은 제품이나 서비스의 수요 양과 시기를 추정하는 것으로 생산, 재고관리 및 마케팅 과정에서 효과적인 전략을 구사할 수 있는 수단으로 이용되며 전반적인 기업 활동에 기반이 된다(Kim, 2008; Min and Jeong, 2009). 최근 불확실한 경영환경으로 인해 수요예측의 중요성이 더욱 강조되고 있으며(Park *et al.*, 2012) 이에 대한 관심은 농업, 건설업, 제조업뿐 아니라 음식점, 병원 등 서비스업까지 확장되고 있다(Kim, 2008). 본 연구에서 중점적으로 다룬 의류산업은 시장 진입 장벽이 낮아 다른 산업에 비해 더욱 경쟁적이며, 따라서 의류산업에서의 수요 예측은 경쟁우위 확보를 위해 필수적이라 하겠다. 실제로 많은 기업들이 활용하고 있는 수요 예측방

법은 대부분 과거의 경험이나 직관에 의존하거나, 단순히 기초통계 분석에 그치는 수준이다. 더욱이 의류 제품은 종류가 매우 다양하고 제품수명이 짧기 때문에 기존 수요예측방법을 그대로 사용하기가 어렵다(Thomassey and Happiette, 2007).

예측 기법은 크게 정성적 기법과 계량적 기법으로 나눌 수 있다. 정성적 기법은 분석 시 사용할 수 있는 자료의 수가 매우 적거나 없는 경우에 사람들의 경험이나 사전지식 같은 주관적인 요소를 활용하여 예측하는 방법이고, 계량적 기법은 서비스나 제품 판매 등의 활동으로부터 얻어지는 데이터를 기반으로 예측하는 방법이다. 예를 들면 수요량이나 매출액 같은 정량적인 자료를 이용하여 미래의 수요를 예측하는 방법이다. 예측기법은 <Figure 1>과 같이 분류할 수 있다(Kim *et al.*, 2011; Lim and Oh, 1992).

본 연구는 미래창조과학부의 재원으로 한국연구재단의 기초연구사업(2013007724)과 지식경제부 정보 통신 기반 구축 사업의(NIPA-2011-(B1110-1101-0002)) 지원을 받아 수행됨.

\* 연락저자 : 김성범 교수, 136-701 서울시 성북구 안암동 5가 1번지 고려대학교 산업경영공학과, Tel : 02-929-5888, Fax : 02-3290-3397,

E-mail : sbkim1@korea.ac.kr

2014년 1월 6일 접수; 2014년 5월 9일 수정본 접수; 2014년 5월 17일 게재 확정.

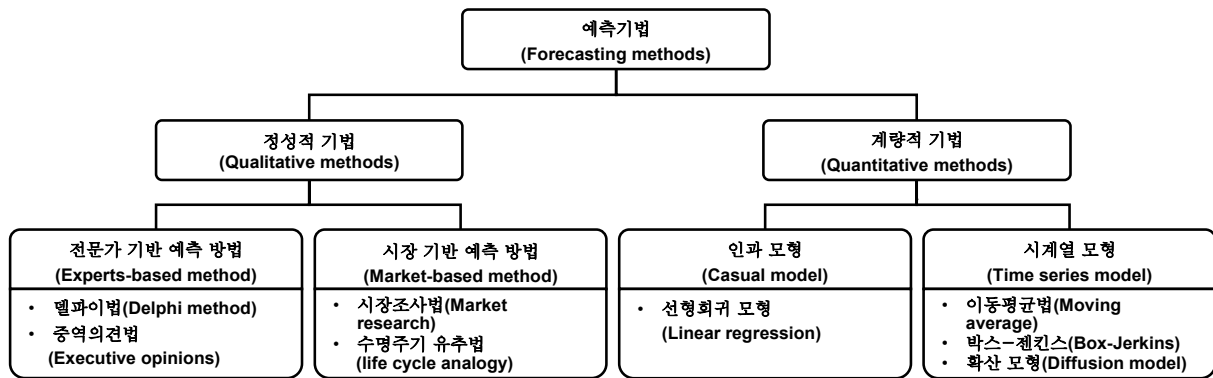


Figure 1. A conceptual diagram of a variety of forecasting methods

정성적 기법 중 전문가 기반 예측 방법은 전문가들의 합치된 의견에 기반을 두어 예측하는 델파이 기법처럼 대부분 전문가들의 의견이나 직관에 의존한다(Koo and Min, 2013). 시장 기반 예측 방법인 수명주기 유추법은 예측하고자 하는 제품과 유사한 특징을 갖고 있는 과거 제품의 수명주기 정보를 이용하여 수요를 추정하는 방법으로 주로 신제품 개발에 사용되고 있다(Kim et al., 2011).

계량적 기법 중 대표적인 인과 모형인 선형회귀 모형은 통계 이론을 바탕으로 독립변수와 종속변수간의 관계를 선형 또는 비선형으로 표현하는 방법으로 모델 구축 절차가 비교적 간단하여 널리 쓰이고 있다. 또한 시계열 모형은 물가지수, 상품 판매고와 같은 경제관련 자료뿐만 아니라 국가나 인구 또는 특정 지방의 기상을 예측하는 경우 등 다양한 분야에서 사용되고 있다(Kim, 2006). 특히 박스-젠킨스(Box-Jenkins) 시계열 모형은 시계열 자료의 추세나 계절성을 잘 반영하는 효과적인 모델로 꾸준히 발전되어 왔다(Bowerman et al., 2005). 확산 모형(Diffusion model)은 일반적으로 새로운 제품의 판매량을 예측하기 위한 제품수명주기(Product life cycle) 생성에 초점을 맞추고 있다(Mahajan, 1986). 확산 모형의 주목적은 서비스나 제품이 출시되었을 때 잠재적 수요자들에게 혁신이 퍼지는 정도를 시간에 따른 간단한 함수식으로 나타내는 것이다(Mahajan and Muller, 1979). 즉, 신제품이 출시되어 포화상태까지의 과정을 추정하는 모델로 사용되며, 해당 품목에 대한 누적시장규모를 시간에 따른 S자 형태의 곡선으로 가정한다(Kim et al., 2000). 확산 모형은 주로 정보통신 시장 예측과 같은 중장기 수요예측에 사용되고 있다(Choi and Kang, 2001).

위와 같이 여러 예측 방법들이 존재하지만, 본 연구에서 중점적으로 다룰 의류는 제품의 종류가 많고 교체시기가 매우 빠르기 때문에 기존 예측 방법들을 그대로 적용하는 것은 여러 가지 문제를 야기할 수 있다(Thomassey and Fiordaliso, 2006).

본 연구에서는 의류산업에 적용할 수 있는 수요예측의 일환으로 제품수명주기를 예측하는 방법을 제안하였다. 제품수명주기란 하나의 제품이 시장에 도입된 후 사라질 때까지의 판매량 곡선을 의미하는데(Buzzell, 1966), 기업은 제품수명주기를 통해 제품의 수명과 잠재성을 파악할 수 있고, 특히 제품수

명주기상의 각 단계에 적합한 전략을 효과적으로 수립할 수 있다(Ye and Kim, 2005). 즉, 제품수명주기의 예측을 통해 시간에 따른 제품 판매량 변화패턴을 추정함으로써 기업은 생산, 재고, 유통, 마케팅 관점에서 제품수명주기 예측결과를 전략적인 의사결정 지표로 사용할 수 있다. 이와 같이 정확한 제품수명주기 예측은 기업의 이익을 극대화 시키는 수단이 될 수 있다.

기존 정성적 기법인 수명주기 유추법은 현재 몇몇 의류기업에서 적용하고 있는 방법이지만, 디자이너나 전문가의 주관에 의존하기 때문에 예측 결과에 대한 신뢰도가 떨어지며 객관적이지 못하다는 한계가 존재한다. 또한 정량적 기법인 확산 모형은 앞서 기술했듯이 중장기 예측에 적합하다고 알려졌기 때문에, 제품수명이 짧고 종류가 다양한 의류산업에는 적합하지 않다.

따라서 본 연구에서는 제품수명주기가 짧으면서도 종류가 많은 경우에 효과적으로 사용할 수 있는 제품수명주기 예측 방법론을 제안하고자 한다. 이는 수요예측의 일환으로 제품출시 초반에 향후 제품수명주기를 예측하는 방법이다. 특히, 경영적인 측면에서는 제품출시 이전이나 제품의 수요가 본격적으로 확대되는 출시 초반시기에 수요예측이 중요한데(Hong and Koo, 2013), 제안방법론은 이런 점을 잘 반영할 수 있으므로 기업에 실질적인 이익을 가져다 줄 수 있다. 제안방법은 데이터를 기반으로 한 알고리즘을 사용하여 제품의 판매패턴을 정의한 후 예측모델을 만들어 미래 제품수명주기를 예측하는 방법으로 크게 파생변수생성, 군집화, 예측모델의 3단계로 이뤄진다.

의류산업의 경우 생산액 대비 재고비율이 타 제조업의 평균보다 약 2배를 웃돌 만큼 재고문제가 심각하다(KIET, 2012). 의류산업의 유통시장은 백화점, 대리점, 아울렛과 같이 단계별로 존재하는데, 본 연구에서 제안하는 예측기법의 결과에 따라 유통채널 변경과 할인을 통한 효과적인 재고관리가 가능할 것으로 예상 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 본 연구에서 제안하는 제품수명주기 예측 방법에 대해 서술하였고, 제 3장에서는 실제 국내 의류브랜드의 판매 데이터에 제안기법을 적용

하여 효과를 입증하였다. 제 4장에서는 본 연구의 결론 및 기대효과와 함께 본 연구의 한계점에 대해 논의하고 향후 연구 방향을 모색하였다.

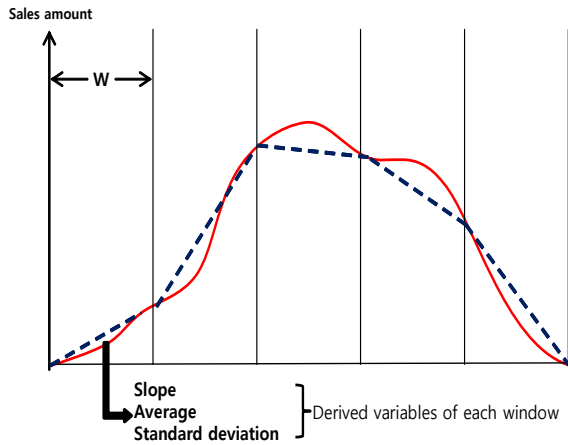
## 2. 제품수명주기 예측 방법론

앞 장에서도 언급했듯이 기존 수요예측 방법은 상황에 따라, 목적에 따라 다양하게 연구되어왔다. 본 연구에서 제안하는 수요예측 방법은 데이터 기반의 제품수명주기 예측으로, 과거에 판매된 제품 데이터 기반의 알고리즘을 이용하여 예측모형을 구축한 후, 미래 제품의 속성정보와 출시 초반의 판매 패턴 정보를 이용하여 향후 제품수명주기를 예측하는 방법이다. 제안하는 제품수명주기예측 방법은 <Figure 2>와 같이 데이터마이닝 알고리즘인 군집화와 예측모형을 이용한 다음의 4단계

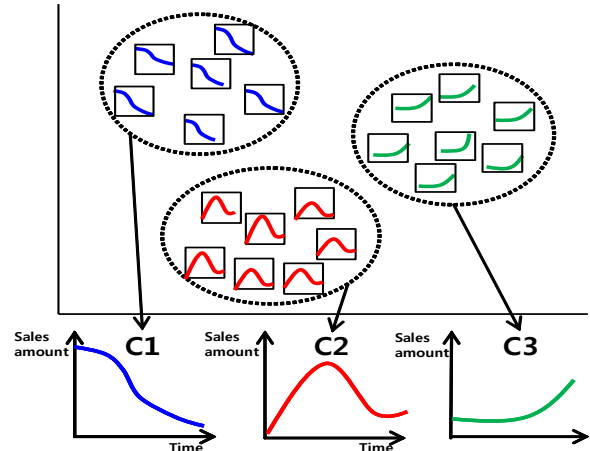
로 이루어진다.

### 2.1 Step 1 : 판매정보를 요약하는 파생변수 생성

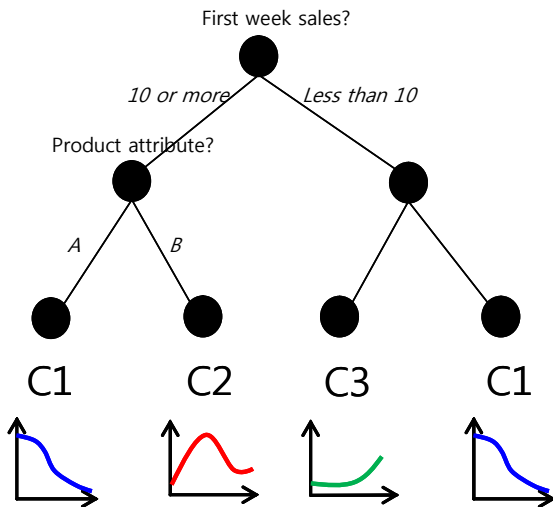
<Figure 2>(a)에서 보여주듯이 첫 번째 단계는 판매량 시계열 데이터의 특징을 잘 반영할 수 있는 변수를 생성하는 것이다. 먼저 시계열 데이터를 적절한 수의 구간으로 나누고 각 구간별로 기울기, 평균, 표준편차를 구한다. 이 파생변수들은 각 구간에서 판매량의 확산속도, 크기, 일관성을 반영한다. 구간 크기  $W$ 는 사용자가 정하는 값으로, 2부터 제품의 수명이 종료되는 시점까지로 정할 수 있다.  $W$ 가 1인 경우는 각 시점마다의 판매량을 그대로 반영하는 것으로, 구간이 생성되지 않으므로 이에 따라 기울기, 평균, 표준편차를 생성할 수 없기 때문에  $W$ 는 2부터 사용하도록 한다. 또한  $W$ 가 제품수명주기 종료시점까지로 할 경우는 구간 분할 없이 전체 패턴을 요약하는 것이



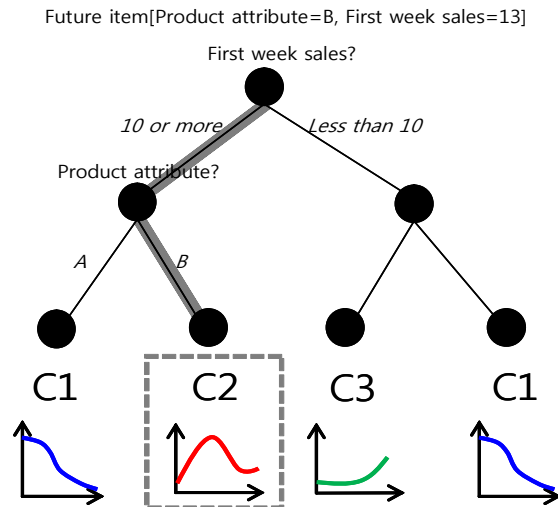
(a) Step 1 : Generating derived variables from each profile



(b) Step 2 : Clustering and labeling of profiles



(c) Step 3 : Building a classification model and validating the accuracy of the model



(d) Step 4 : Predicting class of the future item

Figure 2. Overview of the proposed method

되어 구간의 의미가 없어진다.  $W$ 를 정하는 방법은 딱히 정해지지 않았으며 예측하고자 하는 제품이나 데이터의 특성을 파악하여 수요반응이 빠르며 수명주기가 짧은 경우는  $W$ 를 작게 잡아 작은 변동을 반영할 수 있도록 하고, 수요반응이 비교적 느리고 수명이 긴 경우는  $W$ 를 크게 잡을 것을 제안한다.

## 2.2 Step 2 : 판매패턴 유사 제품 군집화와 레이블링

군집분석은 대표적인 비지도학습 방법으로 군집 내 분산은 최소화하고 군집 간 분산은 최대화하여 가장 유사한 관측치끼리 그룹화 하는 방법이다(Gorden, 1999).  $k$ -평균군집화는 가장 보편적으로 쓰이는 군집기법으로, 군집들 내부의 분산을 최소화 하는 방향으로 각각의 관측치를  $k$ 개의 군집들 중 하나에 배정 하는 기법이다.  $k$ -평균군집화 기법을 적용하기 위해서는 사전에 사용자가  $k$ 를 결정해야하는데, 사전지식으로 결정하거나 서로 다른  $k$ 값으로 구한 군집들을 비교하여 정할 수 있다(Shmueli *et al.*, 2008). 본 연구에서는 Step 1에서 생성한 파생변수들을 이용하여  $k$ -평균군집화를 실행하였고 군집결과를 바탕으로 각 군집에 속하는 관측치에 범주를 부여하였다. Step 2는 각 관측치의 제품수명주기가 유사한 패턴을 가지고 있는 것들끼리 묶어 소수의 군집으로 만드는 과정이며, 군집화와 레이블링을 통해 생성된 범주는 각 관측치가 어떤 수명주기 패턴에 속하는지를 알려주는 지표가 된다.

## 2.3 Step 3 : 제품 속성정보와 초반판매량을 이용한 예측모델 생성 및 검증

이번 단계에서는 전 단계에서 생성한 범주 정보를 종속변수로써 활용하여 제품수명주기를 분류하는 모델을 구축한다. 본 연구에서는 분류 문제에 널리 사용되고 있는 의사결정나무 알고리즘을 적용하였다. 의사결정나무 모델은 결과에 대한 해석이 용이하며, 분류에 영향을 주는 변수에 대한 이해가 가능하다는 장점이 있다(Curram and Mingers, 1994). 무엇보다도 독립변수의 형태가 연속형과 범주형이 섞여있는 데이터에도 특별한 변환 없이 적용할 수 있다는 장점이 있다(Hastie *et al.*, 2001).

본 연구에서는 의사결정나무 모델로 C4.5 알고리즘을 사용하였는데 이는 Ross Quinlan에 의해 개발된 ID3 알고리즘에서 발전된 알고리즘으로 기계학습 분야에서 널리 사용되고 있다(Shin and Song, 2003). 분류모델 구축 시 독립변수로는 제품의 여러 속성변수와 출시 초반 판매량을 정보를 사용하였다. 속성변수의 대부분은 범주형이고 초반 판매량 변수는 연속형으로 독립변수가 연속형과 범주형 변수의 혼합으로 구성됨을 알 수 있다. 모델 구축 시 사용하는 종속변수로는 Step 2에서 군집화를 통해 생성된 범주 정보를 사용한다.

예측모델의 성능을 검증하기 위해 예측모델 생성과정에서 주어진 데이터를 모델 구축에 쓰이는 학습 데이터와 검증에 사용되는 검증 데이터로 나누었다.

## 2.4 Step 4 : 새로운 제품의 제품수명주기 범주 예측

Step 1부터 Step 3까지의 단계적인 과정을 거쳐 구축된 제품 수명주기 분류모델은 새로운 제품의 수명주기를 예측하는데 사용된다. 즉, 새로운 제품의 속성정보와 초반 판매량 정보를 구축된 모델에 넣으면 해당 제품의 제품수명주기가 예측되는 것이다. 이를 통해 새로운 제품의 제품수명주기 패턴을 예측할 수 있다.

## 3. 의류제품의 데이터를 이용한 실제 사례

본 연구에서 제안하는 제품수명주기 예측방법의 효용성을 입증하기 위해 국내 의류기업의 실제 제품별 판매 데이터를 적용해 보았다. 실험에 사용한 데이터는 남성 브랜드 데이터로, 2011년부터 2012년까지 총 2년간 가을시즌에 판매된 제품의 주별 판매 데이터로 매 시즌마다 약 500여 개의 제품이 포함되었다. 각각의 제품은 총 7개의 속성정보를 갖고 있는데 이는 품목, 색상, 디자인, 소재, 패턴, 핏(fit), 가격이다.

Step 1부터 Step 3까지는 예측모델의 구축 및 성능검정 과정으로 2011년 가을시즌 데이터를 사용하였고 Step 4에서는 구축된 예측모델을 통해 2012년 동일시즌 데이터의 제품속성정보와 출시 후 초반의 판매량을 이용하여 새로운 제품의 범주를 예측하고 그 결과를 그림으로 표현하여 모델의 실질적인 활용가능성을 확인하였다.

### 3.1 Step 1 : 파생변수 생성

시계열 판매 데이터의 특징을 잘 반영할 수 있도록 구간을 나누어 주는데 구간의 수  $W$ 를 결정하는 것은 전문가의 의견을 반영하거나 데이터의 특성을 파악한 후 정한다. 실험에 사용한 남성복의 판매 데이터는 출시 후 17주에 제품의 누적판매 비율이 90%에 도달하며 수요반응이 정적인 경향을 보이고 데이터의 판매패턴이 급격하게 변하지 않아 비교적 넓은 구간으로  $W$ 를 정하면서도 제품판매의 90% 이상을 반영하기 위해  $W$ 를 6주 정하고 18주까지의 판매 데이터를 사용하였다. 즉, 1~6주, 7~12주, 13~18주로 총 3개의 구간에서 각각 판매량에 대한 기울기, 평균, 표준편차를 구하여 총 9개의 파생변수를 생성하였다.

### 3.2 Step 2 : $k$ -평균군집화와 레이블링

Step 1에서 생성한 9개의 판매량 요약변수를 이용하여  $k$ -평균군집화를 진행하였다. 적절한 군집 수에 대한 사전 정보가 없기 때문에 본 연구에서는 최적의 군집 개수를 정하는 방법 중 하나인 실루엣 통계량 기법을 이용하였다(Rousseeuw, 1987). 관측치  $i$ 에 대한 실루엣 통계량은 다음의 아래 식으로 계산된다.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

$$\bar{s} = \frac{1}{n} \sum_{i=1}^n s(i) \quad (2)$$

식 (1)에서  $a(i)$ 는  $i$ 관측치가 속한 군집에서  $i$ 를 제외한 다른 모든 관측치들과의 평균 거리이고  $b(i)$ 는  $i$ 관측치가 속하지 않는 군집 중  $i$ 관측치와 가장 가까운 다른 군집의 관측치들과의 평균거리이다. 즉,  $a(i)$ 는 군집의 응집정도를 나타내는 것으로 작을수록 군집이 잘 형성된 것이고,  $b(i)$ 는  $i$ 관측치가 속한 군집과 가장 근접한 군집과의 분리정도를 의미하므로 큰 값을 가질수록 군집이 잘 형성된 것이다.  $s(i)$ 는  $b(i)$ 와  $a(i)$ 의 차이를 표준화시킨 값으로 [-1, 1]의 값을 갖으며, 1에 가까울수록 군집이 잘 형성된 상태를 의미한다.

최적의 군집 수  $k$ 를 찾기 위해 식 (2)에서 나타내는  $s(i)$  값의 평균값  $\bar{s}$ 를 이용한다. 이때 실험적으로  $k$ 를 2부터 차례대로 변경하여  $k$ -평균군집화를 실행하고, 각 경우에 대해서  $\bar{s}$  값을 각각 구했을 때 최대의  $\bar{s}$  값을 갖는  $k$ 가 최적의 군집수를 나타낸다. 이 값이 일반적으로 0.5 이상이면 군집이 잘 되었다고 판단한다(Kaufman and Rousseeuw, 2009).

<Figure 3>은 Step 1에서 생성한 9개의 판매량 요약변수로  $k$ 를 2부터 20까지 바꿔가며 얻은 실루엣 통계량 값을 보여주고 있는데, 군집수가 2일 때 실루엣 통계량이 0.57로 가장 높으며 군집수가 4일 때 0.52로 두 번째로 높음을 알 수 있다. 비록 실루엣 통계량 값이  $k = 2$ 일 때 가장 높게 나왔으나 제품수명주

기 패턴을 2개로 정했을 경우 너무 단순화 될 수 있다고 판단하여 군집수를 4개로 결정하였다.

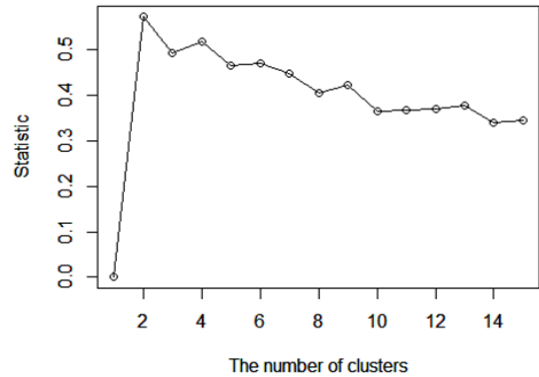


Figure 3. Silhouette plot to determine the number of clusters

<Figure 4>는 4개의 군집에 속한 제품들의 시간에 따른 실제 판매량을 나타낸다. 굵게 표현된 선은 군집에 속한 제품의 평균 판매량으로 각 군집을 대표하는 제품수명주기를 나타낸다. 본 실험에서는 각 군집의 제품수명주기 특징을 반영하여 각 군집을 Slow-up, Fast-up, High Peak, Never-up으로 명명하였다. Slow-up에 속한 제품들은 출시 후 어느 정도 시점이 지난 후 판매량 최고치를 기록하는 패턴을 보이며 Fast-up은 출시 후 대략 6주 이내에 판매량 최고치를 보이고 이후 하락하는 패턴을 보이고 있다. High Peak에 속한 제품들은 출시 직후 빠른 속도로 판매량이 증가하다가 8주 정도에 정점을 찍고 다시 빠르게

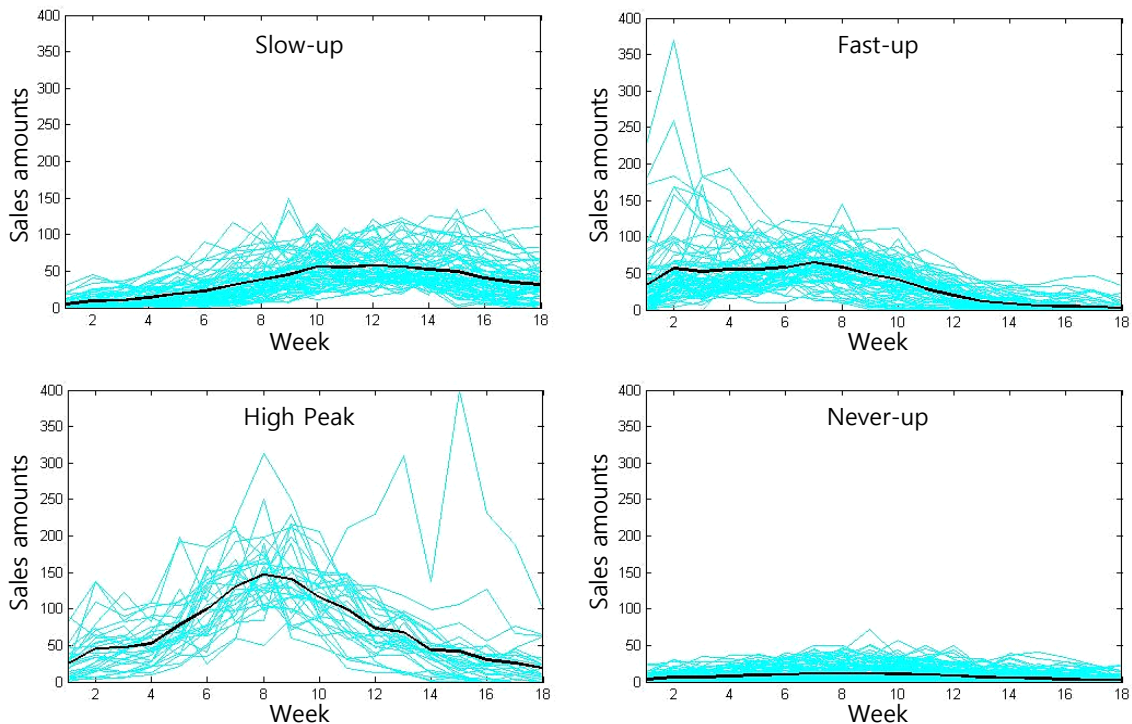


Figure 4. Results of clustering on product life cycle of 2011's items

하락하는 경우이며 Never-up은 출시 이후 내내 판매량이 미미한 제품들로 구성되어 있는 군집이다.

### 3.3 Step 3 : 의사결정나무 분류모델 구축

본 단계에서는 새로운 제품이 출시된 초반시점에 향후 어떤 판매패턴을 보일지를 예측하기 위해 전 단계에서 얻은 군집결과를(Slow-up, Fast-up, High Peak, Never-up) 종속변수로 두고 분류모델을 구축하였다. 제품의 7가지 속성정보와 출시된 후 초반 5주간의 주별 판매량 정보(총 12개의 독립변수)를 이용하여 의사결정나무 모델을 구축하였다.

본 연구에서 의사결정나무 구축 시에 고려해야 할 점은, 종속변수인 범주형 자료의 불균형이 존재한다는 문제와 각 범주마다 기업에 주는 가치가 다르다는 것이다. 본 실험에서는 Never-up의 제품 개수가 나머지 3개 범주들에 속하는 제품의 개수에 비해 약 2배를 웃도는 불균형문제가 존재하는데, 아무런 조치를 취하지 않고 모델을 구축하면 대다수의 제품이 Never-up으로 분류되는 문제가 발생한다. 이러한 결과는 기업의 이익관점에서 문제를 초래하는데, 예를 들면 Slow-up을 Never-up으로 분류한 결과를 토대로 판매활동에 조치를 취한다면 더 팔수 있음에도 못 팔게 되는 판매실기가 발생하여 기업에 손해를 끼치게 된다. 이러한 문제를 해결하기 위해 의사결정나무 구축 시 비용민감학습(Cost-sensitive learning)을 적용하였다.

비용민감학습은 모델에 주어진 관측치가 어떤 특정 범주에 속할 확률이 낮더라도 비용이 높은 경우는 그 범주로 예측할 수 있는 분류모델을 생성하는 방법이다(Kim, 2004). 즉, 분류 과정에서 특정 범주를 더 잘 분류할 수 있도록 가중치를 주어 모델을 구축하는 것이다. 비용민감학습을 적용하기 위해서는 먼저 비용행렬(Cost matrix)을 정의해야 한다. 비용행렬은 범주가 C개 존재 할 경우에 각 원소가  $C_{i,j}$  값을 갖는  $C \times C$  크기의 행렬로 정의되고  $C_{i,j}$  값은 실제 j범주에 속하는 관측치를 i범주로 예측할 경우에 발생하는 비용을 의미한다. 이 값은 특별한 제약 없이 보통 실수 값으로 설정하며 일반적으로 대각행렬의 값은 0으로 둔다(Turney, 1995). 본 연구에서는 Slow-up, Fast-up, High Peak 범주를 Never-up보다 잘 예측하는 것에 가중치를 주고 있기 때문에 Never-up을 제외한 나머지 범주의 비용을 주며 반복 실험을 통하여 <Table 1>와 같은 비용행렬을 얻을 수 있었다.

분류모델의 정확도를 측정하기 위하여 데이터를 7:3으로 나누어 각각 학습 데이터와 검증 데이터로 사용하였다. 모든 범주에 대하여 일관적인 예측모델을 구성하기 위해, 4개의 군집 Slow-up, Fast-up, High Peak, Never-up에서 각각 70%의 학습 데이터를 무작위로 추출하여 모델을 구축하고 나머지 30%의 데이터로 모델분류성능을 검증하였다. 검증 결과는 <Table 2>과 같이 나타나며 Slow-up의 경우 19개 중 12개, Fast-up은 22개 중 17개, High Peak은 8개중 5개, Never-up은 89개중 60개를 잘 분류하고 있다.

Table 1. Cost matrix

		Predicted class			
		Slow-up	Fast-up	High Peak	Never-up
Actual class	Slow-up	0	1	1	5
	Fast-up	1	0	1	1
	High Peak	2	3	0	1
	Never-up	1	1	1	0

Table 2. Confusion matrix for prediction accuracy

		Predicted class			
		Slow-up	Fast-up	High Peak	Never-up
Actual class	Slow-up	12	2	0	5
	Fast-up	2	17	3	0
	High Peak	1	2	5	0
	Never-up	22	1	6	60

<Figure 4>에서 보듯 Slow-up과 Never-up의 초반 판매패턴은 유사하기 때문에 두 범주 간의 분류가 혼동 될 확률이 높다. 특히, Slow-up 제품을 Never-up으로 분류할 경우 판매실기가 발생하는데, 이에 따른 손해를 줄이기 위해 <Table 1>과 같이 실제 Slow-up에 속한 제품들이 Never-up으로 예측될 때의 비용을 가장 높게 주었다. 이 영향 때문에 Never-up제품 중 22개 제품이 Slow-up으로 잘못 예측되었었지만<Table 2>, Never-up을 Slow-up으로 예측하는 경우 보다 Slow-up을 Slow-up을 제대로 예측할 때의 기회비용이 더 높기 때문에 비용행렬적용과 그에 따른 결과는 유용하다고 판단된다.

### 3.4 Step 4 : 2012년 제품의 제품수명주기 예측

본 단계에서는 앞서 구축한 예측모델에 2012년 제품의 5주까지의 초반판매량과 7가지 속성정보를 변수로 사용하여 제품수명주기 범주를 예측하였다. 실제로는 2012년 실제 판매량 데이터가 있기 때문에 각 범주로 예측된 제품의 2012년 실제 판매량 데이터의 전체적인 패턴을 <Figure 5>와 같이 볼 수 있었다. 결과를 보면 2011년 데이터를 이용한 4개의 군집이 2012년에도 잘 나타남을 볼 수 있다. 다만 몇몇 제품들이 해당 범주 내의 다른 제품 판매패턴과 상이한 패턴을 보이고 있지만 대부분은 유사함을 볼 수 있었다. 초기 판매량과 속성정보만으로 예측한 2012년 제품의 수명주기와 2011년 군집분석으로부터 얻은 4개의 범주와의 패턴 유사성을 보기 위해 2011년(실선), 2012년(점선) 각각의 판매량 데이터의 평균을 <Figure 6>과 같이 비교해 보았다. 그림에서 확인 할 수 있듯이 모든 4개의 범주에서 평균 판매량 패턴이 비슷함을 보여주고 있다. 다만 High Peak 범주의 경우 2011년과 2012년 모두 8주차에 판매량이 최고점을 달성하는데 2012년의 경우에는 그 패턴을 13주 정도까지 유지하다가 하락함을 볼 수 있다. 이는 <Figure 5>에서 확인할 수 있듯이 소수의 제품이 13주에 급격한 판매를 보이기 때문이라고 판단된다.

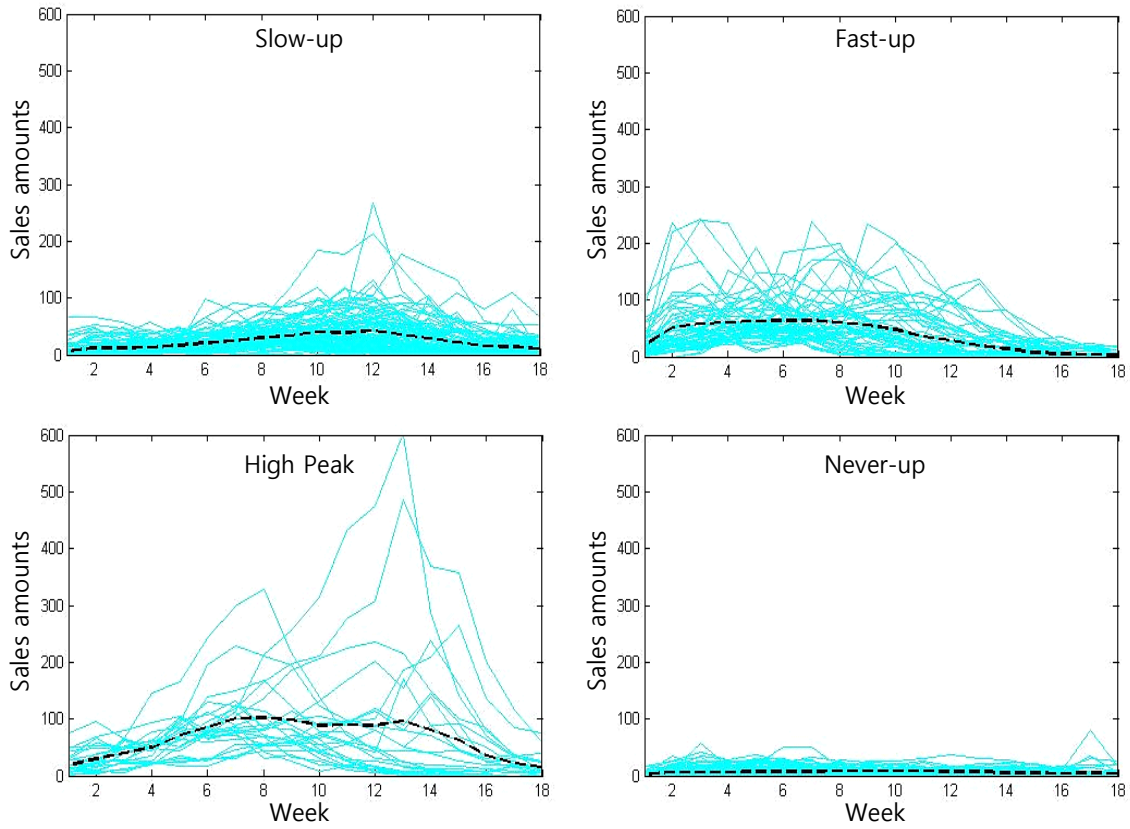


Figure 5. Classification results represented in actual product life cycle of 2012's items

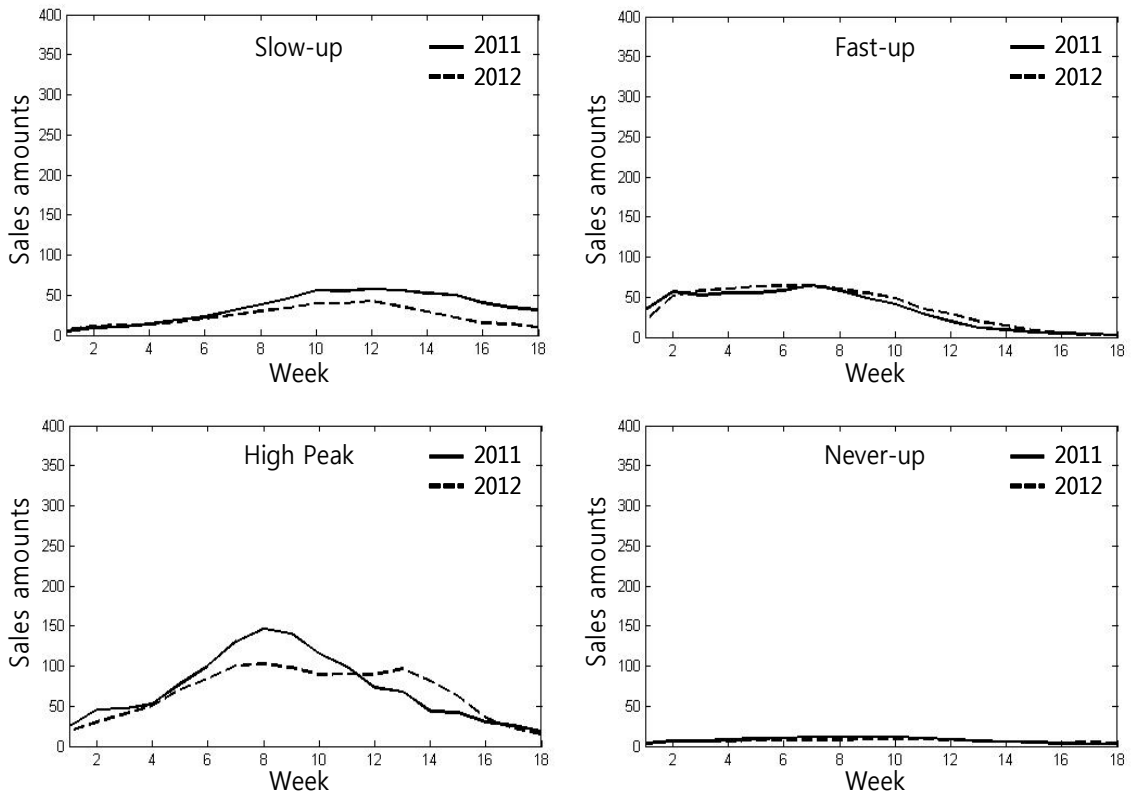


Figure 6. Comparisons of product life cycle of 2011's item with classification results on 2012

## 4. 결 론

통상적으로 의류산업에서의 수요예측은 디자이너와 전문가의 직관에 의해 진행되어 왔다. 본 연구에서는 데이터를 기반으로 객관적이며 의류산업에 적용 가능한 수요예측 기법을 제안하였다. 과거 제품의 판매패턴을 군집화하고 그 결과를 종속변수로 사용하여 분류모델을 구축하였으며 미래 데이터가 왔을 때 향후에 어떤 판매패턴으로 판매될지를 예측하는 기법을 제안하였다. 본 연구에서 제안한 판매주기패턴 예측 기법은 제품 출시 초반에 제품수명주기를 가능하여 적절한 공급과 재고관리를 가능하게 할 수 있다. 본 연구에서는 의류산업의 예를 들어 기술했지만 대부분의 유통업 분야에 확장 적용할 수 있을 것으로 보인다.

제안기법의 한계점으로는, 처음에 설정한 기간까지의 제품수명주기만 예측할 수 있다는 점이다. 실제 예에서는 18주까지의 판매패턴을 보았는데, 특히 Slow-up에 속하는 제품들의 경우는 18주가 넘어가도 제품수명주기가 끝나지 않는 것을 알 수 있다. 그러나 Step 1에서  $W$  설정 할 때 18주에 전체 판매량의 90% 이상이 소진되는 것을 확인하였으므로 예측결과를 마케팅 전략이나 유통관리의 의사결정의 지표로 사용하는 것에는 큰 무리가 없을 것이라 판단된다. 또한 제안기법은 판매패턴 군집화를 기반으로 진행되기 때문에, 군집화 과정이 중요하다. 본 연구에서는  $k$ -평균군집화를 이용하여 모든 제품을  $k$ 개로 분할하였다. 향후 연구방향으로는 군집화 과정에서 이상치 탐지를 고려하여 더욱 완고한 군집을 생성하여 더욱 정확한 모델을 구축하는 것이다. 즉, 제품수명주기가 어떤 군집에도 속하지 않는 제품은 제외하고 판매패턴이 유사한 제품을 잘 군집하여 이 결과를 바탕으로 예측모델을 생성하여 예측성능이 더 높은 모델을 구축할 수 있을 것이다. 끝으로 본 연구에서는  $W$ 를 설정할 때 모든 구간을 일정하게 하였는데 데이터 특성에 맞는 유동적인  $W$ 크기 적용 방법에 대한 고안이 향후 연구되어야 할 것으로 보인다.

## 참고문헌

- Bowerman, B. L., O'Connell, R. T. and Koehler, A. B. (2005), *Forecasting, time series, and regression : an applied approach*, South-Western Pub.
- Buzzell, R. D. (1966), Competitive behavior and product life cycles, *New ideas for successful marketing*, 46-68.
- Choi, J. H. and Kang, H. C. (2001), A Study on the Demand Forecasting using Diffusion Models and Growth Curve Models, *The Korean journal of applied statistics*, **14**(2), 233-243.
- Curram, S. P. and Mingers, J. (1994), Neural networks, decision tree induction and discriminant analysis : an empirical comparison, *Journal of the Operational Research Society*, 440-450.
- Gorden, A. D. (1999), *Classification*, Chapman and Hall/CRC, New York, USA.
- Hastie et al. (2001), *The Element of Statistical Learning*, Springer.
- Hong, J. S. and Koo, H. Y. (2013), Constrained NLS Method for Long-term Forecasting with Short-term Demand Data of a New Product, *Journal of the Korean Operations Research and Management Science Society*, **33**(1), 45-59.
- Kaufman, L. and Rousseeuw, P. J. (2009), *Finding groups in data : an introduction to cluster analysis*, Wiley.com.
- KIET (2012), Apparel stock market status and Implications, Policy report of Ministry of Knowledge Economy, <http://www.korea.kr/archive/expDocView.do?docId=32370>.
- Kim, G. S., Kim, J. h., Kim, H. T., Suh, C. J. and Erh, Y. Y. (2011), *Production and operations management system of the green era*, 200-212, Bobmunsa, Seoul, Korea.
- Kim, J. H. (2004), ROC and Cost Graphs for General Cost Matrix Where Correct Classifications Incur Non-zero Costs1), *Communications for statistical applications and methods*, **11**(1), 21-30.
- Kim, K. H., Kim, J. S., Kang, H. I. and Jun, C. H. (2000), Market Forecasting Modeling Using the Diffusion Model for the Strategic Items in Information/Telecommunication Area, *Electronics and Telecommunications Trends*, **15**(6), 178-189.
- Kim, O. N. (2008), Demand forecasting system, how to build one, *LG Economic Research Institute*, Seoul, Korea.
- Kim, U. C. (2006), *Modern Statistics*, 4th, 427-428, Youngji publishers, Seoul, Korea.
- Koo, H. Y. and Min, D. K. (2013), Forecasting renewable energy using delphi survey and the economic evaluation of long-term generation mix, *Journal of the Korean Institute of Industrial Engineers*, **39**(3), 183-191.
- Lim, J. I. and Oh, H. S. (1992), A Study on the New Product Forecasting Methodology, *Journal of the Korean Institute of Industrial Engineers*, **18**(2), 51-63.
- Mahajan, V. and Muller, E. (1979), Innovation diffusion and new product growth models in marketing, *The Journal of Marketing*, 55-68.
- Mahajan, V. (Ed.) (1986), *Innovation diffusion models of new product acceptance*, Ballinger.
- Min, J. H. and Jeong, C. W. (2009), A GA-based Classification Model for Predicting Consumer Choice, *Journal of the Korean Operations Research and Management Science Society*, **34**(3), 29-41.
- Park, S. B., Kim, J. W., Jeon, S. I., Kim, C. W., Choi, E. J., Lee, C. H. and Huh, Y. S. (2012), Effective demand forecasting methods and practices, *Samsung Economic Research Institute*, Seoul, Korea.
- Rousseeuw, P. J. (1987), Silhouettes : a graphical aid to the interpretation and validation of cluster analysis, *Journal of computational and applied mathematics*, **20**, 53-65.
- Shin, S. C. and Song, M. S. (2003), Bias Reduction in Split Variable Selection in C4. 5, *Communications of the Korean statistical society*, **10**(3), 627-635.
- Shmueli, G., Patel, N. R. and Bruce, P. C. (2008), Data mining for business intelligence : Concepts, techniques, and applications in Microsoft Office Excel with XLMiner, Wiley.com.
- Thomassey, S. and Fiordaliso, A. (2006), A hybrid sales forecasting system based on clustering and decision trees, *Decision Support Systems*, **42**(1), 408-421.
- Thomassey, S. and Happiette, M. (2007), A neural clustering and classification system for sales forecasting of new apparel items, *Applied Soft Computing*, **7**(4), 1177-1187.
- Ye, J. S. and Kim, M. S. (2005), *New Marketing*, 295-307, Pakyoungsa, Seoul, Korea.
- Turney, P. (1995), Cost-sensitive classification : Empirical evaluation of a hybrid genetic decision tree induction algorithm, *Journal of Artificial Intelligence Research (JAIR)*, **2**.