

부품서비스 관점에서 분배 알고리즘을 적용한 수요예측 엔진의 설계 및 개발에 관한 연구

이 영

계명대학교 경영공학과

A Design and Development of Demand Forecasting Engine by applying Distribution Algorithms based on Parts Services

Young Rhee

Keimyung University, Industrial and Management Engineering

In this study, a forecasting engine from the user perspective is studied and developed. Characteristics of forecasting engine can be divided into a few categories, an algorithms for predicting variety of situations and the depth of algorithms based on the number and the types of data. Then applying a variety of algorithms that most closely match the predicted values for the actual value that deduce criteria for selecting an appropriate forecasting algorithm is to organize. Through the forecast quality assessment, the suggested distribution algorithm compared to the existing demand forecast algorithms is good indicators for its accuracy.

Keywords : Forecast Engine, User Perspective, Quality Assessment

1. 서 론

경영전략이나 경영혁신에 대한 근본 개념은 가치창조를 통한 제품의 차별화로 시작되었으나, 제품차별화를 넘어서 서비스 분야, 즉 고객에 대한 서비스의 확대로 이어지고 있다. 이러한 맥락에서 제조업에서는 서비스부품의 신속한 공급이 기업 차별화 방법으로 사용되고 있으며, 이를 뒷받침하는 경영전략이 회사에서 회자되는 말로 즉납률(즉시납부율)을 향상시키기 위해 과학적으로 수요를 예측하여 신속한 공급정책을 세우는 것이다. 이는 부품에 대한 수요예측의 결과가 효율적인 재고 관리 뿐만 아니라 미래의 계획생산(planned production)을 할 수 있어, 비용절감 등으로 기업경영의 핵심요소로 자리 잡고 있다는 것이다.

수요예측은 기업의 부가가치 창출과정에서 있어서

필요한 다양한 의사결정의 기초를 제공해 준다는 점에서 중요한 의의를 갖는다. 제품이나 서비스에 대한 미래의 고객 수요를 추정하는 것은 거의 대부분 과거 또는 현재의 자료나 정보에 근거하여 이루어지게 되며, 이 경우에 특히 환경변수의 정확한 분석이 요구된다. 수요예측의 내용은 고객과 시장의 특성, 제품이나 서비스 종류, 수량, 품질 등이 포함된다[2]. 수요예측을 통한 효과는 다음과 같다. 첫째, 기업의 경영정책과 계획생산이 환경변화에 따라 바르게 이루어져 자재 수급계획이 일정해지며, 고객수요 변화에 대응할 수 있어 기업의 신용도와 서비스 수준이 높아진다. 둘째, 재고에 대한 투자비용을 효율적으로 관리할 수 있다. 셋째, 계획생산이 가능하여 예산편성에 관한 이점과 생산비절감, 고용의 안정을 이룰 수가 있다. 일반적으로 정확한 수요예측을 위한 수요예측기법의 선택은 자료의 형태, 예측의

목적 혹은 예측의 성격에 따라 결정된다. 그에 따라 바람직한 예측 모형이 갖추어야 할 기준들을 다음과 같이 제시하고 있다[5, 7~9]. 먼저 어느 정도의 정확성이 요구되는지를 알고 수요예측 값을 얻기 위해 소요되는 시간 및 노력과 그 예측의 정확성과의 절충(trade off)을 시키는 과정이 가장 먼저 요구된다. 가장 적은 시간과 비용을 들인다면 이동평균법을 이용하여 예측을 할 수 있지만, 이 경우 예측의 정확도가 낮아 잘못된 예측으로 인한 비용이 발생할 수 있다[3]. 반대로 ARMA(Auto-regressive moving average) 등과 같은 정확한 예측기법을 사용하여 예측을 하더라도 수익이 낮으면 오히려 예측 자체의 비용이 더 클 수도 있다. 그리고 예측하고자 하는 기간의 길이(forecast horizon)에 따라 예측모형을 선택해야 되는데, 수요예측기간을 기준으로 분류 했을 때, 크게 단기, 중기, 장기 수요예측으로 분류되어 선택하는 것이다. 단기 수요예측에는 이동평균법, 지수평활법 등이 해당되며, 중기 수요예측에는 추세분석, 시계열 예측법 등이 해당되고, 장기 수요예측에는 ARIMA(Auto-regressive integrated moving average), ARMA 등이 해당된다. 다음으로 이용 가능한 자료의 종류에 따라 적용 가능한 모형을 선별한다. 이용 가능한 자료의 종류란 자료의 개수를 의미하는 것인데, 일반적으로 ARIMA의 경우 50개 이상의 자료가 있어야만 그 자료를 신뢰할 수 있다고 하며, 시계열분해법의 경우 40여개 이상의 자료가 있어야 한다고 알려져 있다[8]. 즉, 과거 데이터 개수가 충분하지 불충분한지에 따라 선택할 수 있는 예측기법이 바뀔 수 있는 것이다. 마지막으로 앞서 언급한 기준에 따라 선별된 예측기법 알고리즘이 어느 정도로 복잡한 것인가를 알아야 한다. 같은 비용과 시간, 그리고 정확성을 가진다면 가급적 쉬운 모델을 선택하는 것이 용이하다.

수요예측 엔진의 구성은 크게 세 가지로 구분할 수 있는데, 첫째는 엔진 내부의 예측알고리즘에 대한 다양한 구성과 알고리즘의 깊이(depth)라고 할 수 있다. 이를 좀 더 부연하자면 알고리즘에 관하여는 이미 다양한 연구가 진행되어 데이터의 수와 형태에 따라 적용할 수 있는 기준이 있으며, 이를 체계화하는 것이 문제로 남아있다. 알고리즘의 깊이는 프로그래머로서 예측오차 등을 감소 또는 축소시키기 위한 개발자적 마인드라고 할 수 있다. 둘째는 다양한 알고리즘을 적용하여 실제 값과 가장 근접하는 예측 값을 유추해 내는 적절한 수요예측 알고리즘을 선택하는 기준을 마련하는 것이다. 그러나 실제 값과 예측 값은 사건의 발생순서가 뒤바뀐 상황이기 때문에 이는 과거의 데이터를 통한 검증은 적용할 수 밖에 없다. 마지막으로 이상과 같이 얻어진 예측 값으로서 미래의 상황을 투영(projection)할 기반이 담겨져 있어야

한다는 것이다.

본 연구는 제 2장의 선행연구를 통하여 수요예측 알고리즘 소개, 예측정확도를 판정하는 방법론이 소개되며, 제 3장에서는 수요예측 엔진의 주요 구성모듈에 대한 설명이 있다. 제 4장에서는 수요예측 엔진의 화면 구성과 수요예측 결과의 품질지수가 소개되며, 제 5장에서 결론을 맺는다.

2. 선행연구

서비스 부품을 생산하는 제조업에서는 일정한 규칙의 유무를 떠나, 시간의 흐름에 따라 주문량이 변하는 현상의 자료이므로 시간의존적인 시계열 데이터라고 할 수 있다. 이런 시계열 데이터를 가지고 모형을 설정하여 미래를 예측하는 분석을 시계열 분석이라 한다. 시계열 분석은 과거 데이터를 사용하기 때문에 서비스 부품의 과거 주문 데이터로부터 데이터의 형태를 설명하고, 과거 움직임에 기초하여 규칙을 찾아내는 방법이라 말할 수 있다. 시계열 데이터에는 연속된 또는 특정주기를 가지는 값들 사이에 얼마간의 자기상관성이 발견되는데, 추세요인, 순환요인, 계절요인 그리고 불규칙요인 등의 기준에 근거하여 시계열의 패턴을 몇 가지 측면에서 분석할 수 있다. 시계열분석을 크게 나누면 평활법, 분해법, 조정법의 3가지로 구분할 수 있고, 평활법의 대표적인 것이 이동평균법(Moving Average), 지수평활법(Exponential Smoothing)이다. 그 외에 Box-Jenkins 방법은 조정법(Control Method)의 대표적인 것이고, 분해법(Decomposition Method)은 정부의 경제지표에 많이 이용된다[5]. 대표적으로 실무에서 사용되는 예측기법은 주로 이동평균법, 지수평활법, 추세분석, 시계열분해법이며, 일반적으로 잘 알려져 있는 이동평균법과 지수평활법의 모형은 생략하고 추세분석 및 시계열분해법 모형을 언급하고자 한다.

$$F_t = a + bx \tag{1}$$

$$b = \frac{n \sum X_i Y_i - \sum X_i \sum Y_i}{n \sum X_i^2 - \sum X_i \sum X_i} \quad a = \frac{\sum Y_i - b \sum X_i}{n}$$

추세분석, 시계열분해법은 식 (1)과 같이 최소자승법에 의해 예측 값이 결정되며 시계열분해법의 경우 별도로 계절지수를 산출하여 이를 가법 또는 승법처리 하여 예측치를 산출한다. 식 (1)에서 X 는 시점(차수), Y 는 X 시점의 수요량, 그리고 n 은 총 데이터 개수를 의미한다. 시계열분해법의 경우 식 (1)의 결과 예측지수 F_t 에 계절지수 ω 를 가법 또는 승법처리 하여 예측치를 산출한다.

가법모형은 F_t 에 계절지수 ω 를 더하는 방법으로 일정한 변동폭을 지닌 예측대상일 때 사용하며, 승법모형은 F_t 에 계절지수 ω 를 곱하는 형태로 시간의 흐름에 따라 점차로 일정한 추세를 가지고 증가하며 또한 변동폭도 증가하는 경향을 보일 때 사용한다. 이 때 계절지수 ω 는 산출하는 방법이 몇 가지가 있는데 그 중 이동평균 비율법이 가장 많이 사용된다. 이 방법은 일평 중심이동평균법이라고도 하는데 각 기의 실측치를 그 기의 중심 이동평균치로 나누어 계절비를 얻고, 동일 기에 속하는 모든 계절비를 평균하여 계절지수를 얻는다[8~9].

예측값의 정확도는 예측오차(forecast error)에 의하여 결정되며, 오차가 작을수록 정확도는 높아진다. 정확도를 측정하는 방법으로는 오차제곱합(SSE : sum of squared error), 오차제곱평균(MSE : mean square error), 평균오차제곱근(RMSE : root mean square error), 평균오차제곱근비율(RMSPE : root mean squared percentage error), 절대평균오차비율(MAPE : mean absolute percentage error), 테일계수(Theil-U coefficient) 등이 있다. 이 중에서 MAPE는 계량모델 간에 예측오차의 비교가 용이하고, 그 신뢰도가 높기 때문에 자주 이용되고 있다[10~12].

근래에는 신경망 모형의 신뢰성이 부각됨에 따라 국내에서도 1990년 1월부터 1996년 5월까지의 일별 수요 실측치를 이용하여 단기 천연가스 수요예측을 위해 시계열분석과 인공신경망을 이용한 사례도 있다[5]. 기존의 경영활동에서 수요예측의 범위는 전사적 차원의 기업 전략을 수립이었다면, 최근에는 좀 더 범위를 확장하여 SCM(Supply Chain Management) 구축을 위한 협업적 수요예측 모형을 개발하려는 연구도 진행된 바 있다[8, 11].

3. 수요예측 엔진의 설계

수요예측을 하기 위한 준비단계는 데이터의 유실 여부를 판단하여 결측치에 대한 보정 단계부터 시작된다. 결측치를 유추 및 보정하는 다양한 방법론이 제시되고 있다[4]. 이후는 데이터 패턴분석 단계 등이 있다. 본 연구에서의 수요예측엔진의 기본 모듈은 다양한 수요예측 알고리즘을 포함하는 알고리즘 풀(pool)과 최적의 알고리즘을 선정하기 위한 모듈, 그리고 이 후는 얻어진 예측치를 통한 미래에 투영하는 분배알고리즘 모듈로 구성된다.

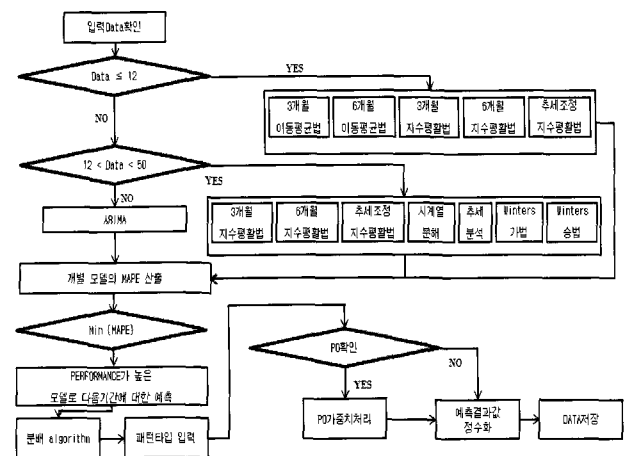
3.1 알고리즘 풀

서비스부품은 시간의 흐름에 따라 주문량이 변하는 현상이 관측됨으로서, 이런 시계열 자료를 가지고 모형을

설정하여 미래를 예측하는 분석을 시계열 분석이라 한다. 시계열 분석은 분석의 전초 단계로 데이터의 계절성과 추세 존재 유무를 먼저 파악하도록 한다. 계절성이 존재하면 시계열 분해, Winters 모형, ARIMA(자기회귀이동평균 : autoregressive integrated moving average)로 예측을 할 수 있으며 계절성과 추세가 존재하지 않을 경우에는 단일 지수평 활법이나 이동평균법으로 예측이 가능하다. 계절성은 없으나 추세가 존재할 경우에는 추세분석이나 이중 지수 평활법, ARIMA를 이용한다.

계절성과 추세 존재를 알아보기 위해서는 주어진 관측값에 대해 모든 가능한 한 쌍의 평균편차 사이의 상관관계수인 자기상관계수(auto correlation coefficient)를 계산하는 것과 시계열도표를 이용하는 방법이 있다. 자기상관계수의 자기라는 말은 동일한 시계열 내에 존재하는 값들 사이의 상관관계 정도를 계산하기 때문에 붙여진 것이다. 추정된 자기상관계수(estimated auto correlation coefficient)는 두 확률변수에서 한 쌍의 관측값 사이에 통계적 관계를 나타내는 방향과 세기를 의미한다. 분해법의 적용은 시계열의 변동요인이 각각 분해될 수 있다는 가정 하에서 시계열을 분석하는 방법으로써 이론적으로 많은 약점을 가지고 있고, 변동요인들이 쉽게 분해될 수 없는 것이 사실이다. 이에 시계열 분석에 대한 새로운 접근 방법을 정립했고, 또 이 방법을 더욱 일반화시킴으로써 ARIMA 모형화를 토대로 하는 통계적인 이론 체계를 구축하는 것으로 이 분석방법은 관찰된 시계열 자료를 하나의 시계열 모형집단으로부터 구축된 표본으로 간주하여 이들이 어떤 확률적 성질을 만족하는가를 조사하고 통계적 추정 및 검정을 통하여 적절한 시계열모형을 수립하는 것이다.

본 연구에서의 수요예측의 편의상 기능을 수동예측 기법과 자동예측기법으로 구분하여 구성하였는데 구현 상에는 자동예측기법이 수동예측기법을 포함하고 있다



〈그림 1〉 수요예측 알고리즘 풀

고 보면 된다. 자동예측기법은 <그림 1>에서 보듯이 알고리즘 풀을 적용 가능한 데이터의 수와 연관되어 알고리즘을 분류한 형태로서 가장 소수의 데이터부터 시작하여 50개 이상의 시계열 데이터가 주어졌을 때 적용되는 알고리즘을 분류했다는 것이다. 자동예측기법은 수요예측에 관하여 부족한 지식으로도 효율적인 적용이 가능하다.

수동예측기법은 하나의 알고리즘 풀에 있는 모든 알고리즘을 인위적으로 지정하게 구성하였다.

3.2 알고리즘 선정방법

시계열 데이터가 어떤 확률적 성질을 가지고 있으며 어떤 알고리즘이 적합한가를 찾기 위한 알고리즘 식별(algorithm identification), 모수 추정(Parameter Estimation), 적합성 진단(Diagnostic Checking)의 세 단계를 거친다. 알고리즘 식별은 제시된 시계열 데이터에 대하여 적절한 것이라고 생각되는 알고리즘 풀 안에서 관련 알고리즘에 대한 관련 통계량 등을 비교하여 가장 적절한 하나의 알고리즘을 선정하는 것이다. 알고리즘을 선정하는데 주의해야 할 점은 모수 절약의 원칙(principle of parsimony)에 충실해야 한다는 것이다. 이것은 관찰된 데이터를 적절히 표현하면서 모수의 수가 가장 작은 알고리즘을 선정하는 것으로써 가능한 한 단순하게 표현하려는 의도이다. 선정된 알고리즘은 하나 이상의 모수를 포함하고 있는데 이 모수는 제시된 데이터로부터 추정할 수 있다. 모수의 추정방법으로는 적률법, 최우추정법, 최소제곱법 등이 주로 사용하였다. 알고리즘이 선정되면 본래의 시계열 데이터를 설명하는데 문제가 없는지 검증해야 하는데 이를 알고리즘의 적합성 진단이라고 한다. 만일 모든 가정과 조건이 만족한다면 알고리즘은 적절한 것이며 그 알고리즘을 기초로 미래 값을 예측할 수 있다. 그러나 가정 및 조건이 만족되지 않는다면 다시 알고리즘 식별 단계로 되돌아가서 위의 세 단계를 반복하여야 한다.

주어진 서비스 부품 주문 데이터에 여러 가지 후보 알고리즘이 존재할 경우 일반적으로 사용되는 알고리즘의 선택기준은 잔차, 자기상관함수 검정통계량, Box-Ljung 검정 통계량 등을 가지고 모형을 선택할 수 있다. 본 연구에서 수요예측 알고리즘의 선택 기준은 통계분야에서 예측력 비교 기준으로 예측오차에 근거한 적합척도인 MAPE를 적용하였다.

$$MAPE = \frac{1}{n} \sum \frac{|a_t - f_t|}{a_t} \times 100(\%) \quad (2)$$

각 알고리즘에 대하여 MAPE를 식 (2)를 이용하여 계산하는데 a_t 와 f_t 는 t 시점에서의 실제치과 예측치를 의미한다. 예를 들어 3개월 이동평균법의 적용은 t 시점 이전의 3개월의 데이터는 실측치가 이미 존재하기 때문에 $t-3$ 시점부터 예측치를 구한 후, 이에 상응하는 MAPE를 계산하여 3개월 이동평균법의 MAPE로 값을 가진다. 이러한 과정을 알고리즘 풀 안의 모든 수요예측 알고리즘에 적용하여 각각의 MAPE를 계산한 후에 가장 작은 MAPE를 갖는 알고리즘이 선정되며 이를 이용하여 t 시점의 수요예측량을 취하는 방법이다.

3.3 분배알고리즘

미래의 수요 예측치를 순차적으로 취하는 방법은 t 시점의 예측치, f_t 를 구하고, $t+1$ 시점의 예측치는 t 시점에서 구한 예측치를 적용하여 $t+1$ 시점의 예측치, f_{t+1} 을 계산하는 일련의 순환과정이 통상적으로 적용되고 있으나 예측치를 가지고 예측치를 구한다는 점에서 다른 방법을 적용하기로 하였다.

본 연구에서 제시하는 분배 알고리즘은 “모든 시계열 자료는 과거의 패턴이 미래에도 동일하다”는 가정에서 시작한다. 예측을 수행할 시점의 수요는 과거 동일 시점의 실측치와 강한 상관관계가 있다고 보는 것에서 분배 알고리즘은 시작하며, 과거 동일 시점의 실측치를 각 연도별로 다른 비율로 가져와 이를 예측치로 결정한다. 분배 알고리즘의 모형은 다음과 같은 단계로 진행된다.

Step 1 : 현재 시점을 기준으로 차기 예측치 산출

이동평균법, 시계열분해법 등 일반적인 예측기법을 통해 현재 시점을 기준으로 차기 예측치, f_{t+1} 을 산출한다. 차기 예측치의 값은 분배 알고리즘에 직접적인 영향을 끼치기 때문에 가급적 정확도가 높은 예측기법을 사용하는 것이 좋으나 단순한 이동평균법 등을 이용하여도 무방하다.

Step 2 : 증감비율(RFA : rate of forecasting amount) 산출

산출된 차기 예측치가 과거 같은 시점의 수요에 비해 얼마만큼 증가 또는 감소하였는지 알기 위해 식 (3)과 같이 예측치를 과거 같은 시점의 수요에 w_i 만큼 가중치를 적용하여 합산한 값으로 나눠 RFA를 구한다. w_i 는 연도별 가중비율이라 정의하며, w_1, w_2, w_3 는 0이 될 수 없다는 것은 최소 24개의 데이터가 요구됨을 의미한다. 연도별 가중비율은 불규칙적인 수요의 변화정

도를 평활화하기 위한 한 가지 방법으로서 최근 데이터에 좀 더 많은 가중치를 줄 수 있는 방법이 될 수 있다. 아래 식 (3)에서 t 는 현재 시점을 의미하며 $f_{(t+1)}$ 은 현재를 기준으로 차기 예측치, $a_{(t+1)-12i}$ 는 예측 시점으로부터 1~3년 전 과거 동일 시점의 실측치를 의미한다.

$$RFA = \frac{f_{t+1}}{\sum_{i=1}^3 \omega_i a_{(t+1)-12i}} \quad (3)$$

where $\sum_{i=1}^3 \omega_i = 1 \quad \omega_1, \omega_2, \omega_3 \neq 0$

Step 3 : RFA를 적용한 예측치 산출

최종적으로 예측치를 산출하는 모형을 식 (4)와 같이 나타냈다. 과거 동일 시점의 실측치를 각 연도별로 ω_i 만큼의 가중치를 적용하여 합산한 값에 식 (3)에서 구한 증감비율을 곱하여 이를 예측치로 산출하게 된다. 증감비율을 곱하는 목적은 과거 실측치를 합산한 결과에 추세 요인을 보정하기 위함이다.

$$F_{t+2} = RFA \cdot \sum_{i=1}^3 \omega_i A_{(t+2)-12i} \quad (4)$$

$$F_{t+n} = RFA \cdot \sum_{i=1}^3 \omega_i A_{(t+n)-12i}$$

분배 알고리즘의 정확성을 검증하기 위해 본 연구는 대표적인 농기계 제조업체인 D사의 부품데이터를 사용하였으며, 자료는 2000년 1월부터 2008년 12월까지의 수집하였다. 하지만 이 중 2004년 이전과 2008년의 데이터는 결측치가 많아 시계열 자료라고 볼 수 없어, 표본 기간은 2004년 1월부터 2007년 12월까지로 한정하였으며, 2007년의 데이터는 실측치가 없는 것으로 가정하고 해당 기간을 예측기간으로 설정하였다. D사의 총 부품종류는 약 65,000여종 중 4년간의 표본기간 내에서 결측치와 이상치(outlier)가 포함되어 있지 않고 평균 수요가 25 이상인 표본으로 범위를 한정하여 최종적으로 57종의 표본을 수집하였다.

Step 1에서 현재 시점을 기준으로 차기 예측치를 산출할 때는 일반적 예측기법을 활용한다고 언급한 바 있다. 본 연구에서는 그 중 예측오차에 대하여 신뢰할 만한 시계열분해법을 이용하여 차기 예측치를 산출하였으며, 이를 이용하여 분배 알고리즘을 적용하였다. 비교대상이 되는 수요예측기법은 3기간 단순이동평균법, 3기간 단순 지수평활법, 추세분석법, 시계열분해법이다. 그리고 서론에서도 언급한 바와 같이 수요예측 기법선

택을 위한 기준을 만족하는지 검증하기 위해 57개 표본에 대해 아래의 <표 1>과 같은 기준으로 분류하여 수요예측을 수행하였다. 여기서 전체 수요 평균이란 2004년 1월부터 2007년 12월까지의 모든 수요에 대한 평균을 의미한다.

<표 1> 표본 분류 기준

	데이터 크기
실험기준 1	24개 데이터를 이용(2005년 1월~2007년 12월)
실험기준 2	36개 데이터를 이용(2004년 1월~2007년 12월)

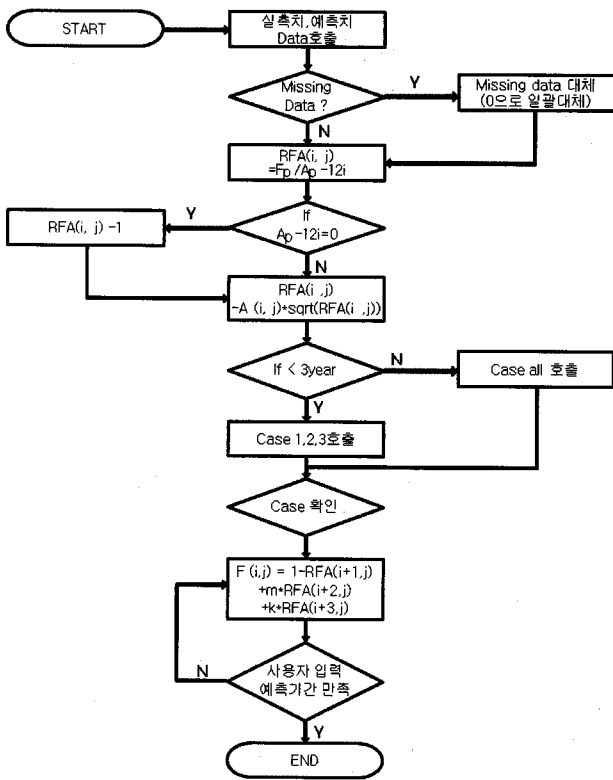
앞서 언급한 연도별 가중비율 ω_i 는 수요예측 대상에 따라 적합한 비율이 다르다. 이는 시계열 자료를 구성하고 있는 추세, 순환, 계절, 우연 요인 등의 내부적 요인과 대상의 외부적 요인에 해당하는 시장의 상황, 판매촉진(promotion) 등에 의해 결정된다고 짐작된다. 따라서 본 연구에서는 실험계획을 통해 가장 낮은 오차를 보였던 <표 2>의 비율에 대해서만 범위를 한정하여 표본적으로 분배알고리즘의 타당성을 실험하였다.

<표 2> 연도별 가중비율

가중비율	ω_1	ω_2	ω_3
적용비			
9 : 1 비율	0.9	0.1	0
8 : 2 비율	0.8	0.2	0
7 : 3 비율	0.7	0.3	0
5 : 5 비율	0.5	0.5	0
5 : 3 : 2 비율	0.5	0.3	0.2

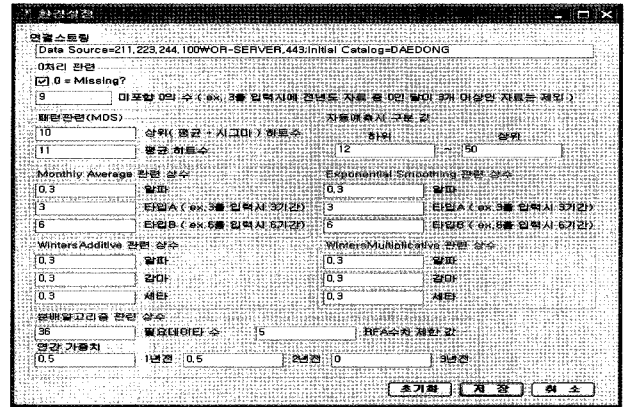
이와는 다르게 실제 수요예측엔진을 구성하는 과정에서는 w_1, w_2, w_3 의 값을 환경설정에서 수동으로 지정하는 경우와 자동으로 모든 경우의 수를 조합하는 방법을 적용하여 예측오차를 구한 후에 이를 토대로 가장 적절한 일종의 가중치 w_1, w_2, w_3 를 찾도록 구현하였다. 분배 알고리즘의 정확성을 검증하기 위한 방법으로는 MAPE 값을 사용되었으며, MAPE는 모든 실측치와 예측치의 차이를 이용하여 산출하지만 분배알고리즘의 경우 f_{t+1} 시점까지는 일반 예측기법을 이용하므로 MAPE 값의 차이가 없다. 따라서 본 연구에서는 분배알고리즘이 적용되는 최종 예측기간 1년간의 MAPE 값을 산출하여 정확성을 검증하였다.

<그림 2>는 수요예측엔진의 한 모듈로 분배알고리즘을 구현한 흐름도를 보여주고 있다.



<그림 2> 분배알고리즘 흐름도

측 시 인정하는 데이터의 범위, 알고리즘에 사용되는 상수들 기본적으로 설정 할 수 있도록 기능을 부여하였고, 초기는 기본값으로서 시작하여 자동으로 처리되도록 하였다.



<그림 4> 환경설정 내용

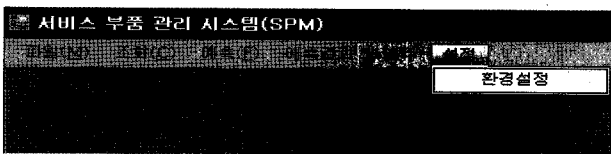
환경설정 후에는 메뉴바에서 <그림 5>와 같이 예측 기능의 자동예측 버튼을 클릭하면 월간 수요예측을 위한 단계로 진입한다.

4. 수요예측 엔진 개발

수요예측 엔진을 개발하기 위한 환경은 닷넷환경의 C# 프로그램 언어와 데이터베이스 관리를 위하여 MS SQL 를 사용하였다.

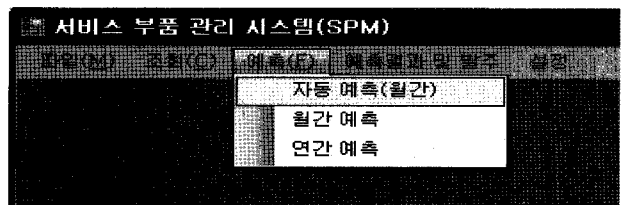
4.1 화면설정

본 개발에서 수요예측시스템의 시작은 환경을 설정해야 하는데, <그림 3>과 같이 프로그램 상단 메뉴의 설정에서 환경설정 버튼을 클릭함으로써 시작된다.



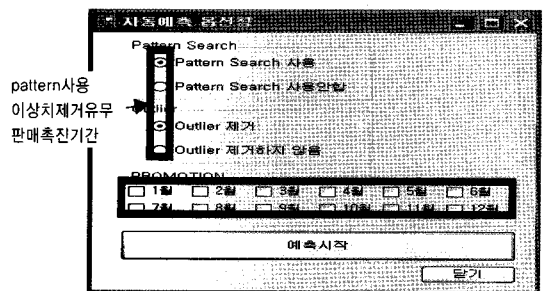
<그림 3> 수요예측 환경설정

환경설정은 <그림 4>와 같이 기본적으로 DB에 데이터를 저장하는 기능과 초기화하는 기능을 갖추었고, 이는 서버의 인터넷 주소 변경 과 포트번호를 변경할 수 있게 구성되었다. 결측치 처리 관련, 패턴처리 관련, 예



<그림 5> 자동예측(월간)

자동예측 단계에서는 자동예측에 대한 옵션 창이 준비 되는데, <그림 6>과 같이 자동예측(월간)의 기본 화면으로 Pattern Search사용의 유무, 이상치의 유무 및 판매측진 기간 등을 선택하도록 구성하였다. 설정 완료 후 예측 시작 버튼을 누르면 실행된 결과 데이터는 DB에 저장되며, 예측결과 및 발주의 기능에서 확인할 수 있다.



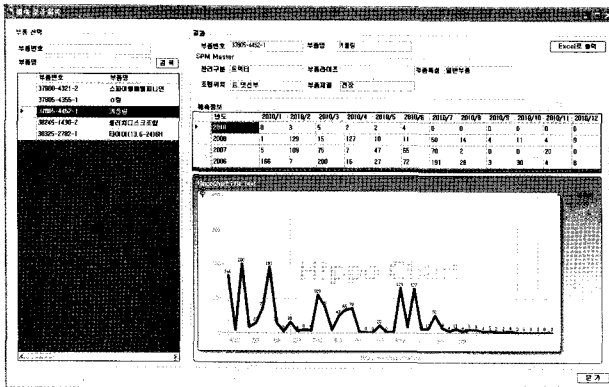
<그림 6> 자동예측(월간) 옵션창

<그림 7>에서 보듯이 월간예측의 기본 화면으로 사용자가 직접 예측을 할 수 있도록 구성하였고, 인위적으로 모든 알고리즘의 선택이 가능하도록 되어있다. 운영자가 직접 예측해보고자 하는 알고리즘을 부품마다 입력할 수 있게 구현하였고, 자동예측(월간)과 비교할 수 있게 하였다.



<그림 7> 월간 수등예측 화면

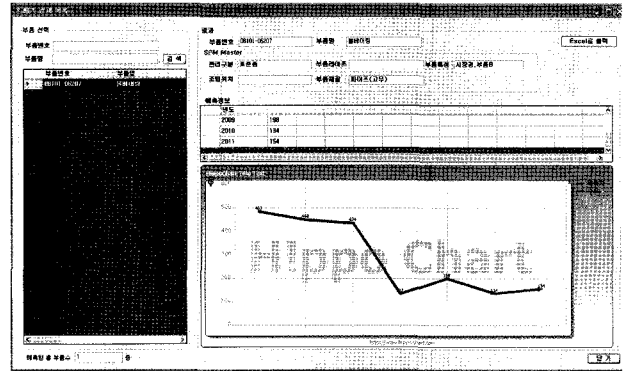
이후의 예측된 결과는 예측결과 목록에서 확인할 수 있도록 구성되었는데, 자동예측(월간)과 다르게 사용자가 예측을 하고 난 이후의 예측 결과를 DB에 저장하지 않는다. 분배 알고리즘을 적용한 12개월의 결과는 <그림 8>과 같이 표현되며 엑셀로도 출력이 가능하도록 구현하였다.



<그림 8> 12개월 예측결과

최종적으로 연간예측의 기능을 보완하였는데, <그림 9>와 같이 기본화면으로 사용자가 연간예측을 할 수 있도록 하였다. 연간예측은 판매촉진기간을 설정하는 기능

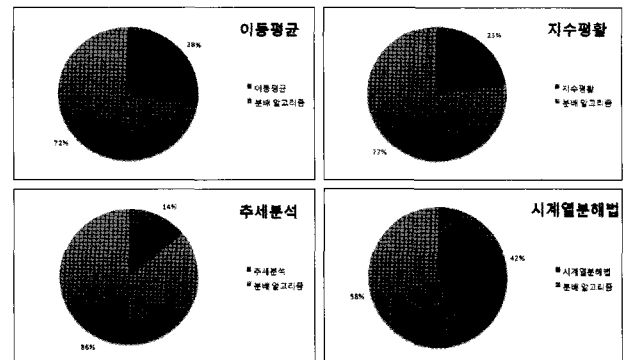
을 제외 하고, 월간예측의 모든 기능과 동일하며, 주문자료가 최대 7년 밖에 없어 자료의 불충분으로 정확한 예측은 기대하기 어려움이 있었다.



<그림 9> 연간 예측결과 목록

4.2 수요예측 품질평가

본 연구의 개발에서 최적의 수요예측 엔진을 구현하기 위한 조건을 두 가지로 밝히고자 한다. 첫째는 알고리즘 풀에서 최적 알고리즘을 선택하는 과정인데, 과거의 자료를 통한 통계적 실험을 바탕으로 단순히 예측오차를 최소화시키는 알고리즘을 선택하였다. 둘째는 선택된 알고리즘을 이용하여 미래의 수요를 투영하는 알고리즘인 분배알고리즘에 수요예측에 대한 정확도가 판별된다고 할 수 있다.



<그림 10> 알고리즘 간 MAPE 비교

미래 수요를 투영하는 분배알고리즘의 정확성과 데이터 크기별 차이를 알아보기 위해 D사의 농기계부품 57개의 표본을 추출하여 그 중 36개월 이상 실험을 한 결과, 분배 알고리즘과 비교 대상 예측알고리즘에 대해서 1:1로 MAPE 값을 비교한 후, MAPE 값이 상대적으로 적게 계산된 알고리즘의 수를 카운트 한 결과를 <그림 10>과 같은 얻었다. MAPE 값이 적다는 의미는

수요예측의 정확도가 있음을 의미한다. <그림 10>에서 알 수 있듯이 대체적으로 분배알고리즘은 다른 예측기법들에 비해 MAPE 값이 적은 것을 알 수 있다.

단, 시계열분해법의 경우, 단순히 MAPE 값이 적은 부품수를 고려할 때는 분배알고리즘이 시계열분해법에 비해 우세하게 나타났지만, 미니탭을 이용한 <표 3>과 같이 가설검정으로 비교해보면 $P(T \geq t)$ 의 확률이 0.39로 유의수준 $\alpha=0.1$ 보다 크기 때문에 통계적으로 유의한 차이는 없다고 할 수 있다.

<표 3> 가설검정 결과

통계량	비교대상	시계열분해법	분배알고리즘
평균		97.19	95.53
분산		4922.07	4701.12
관측수		57	57
자유도		53	
t 통계량		0.28	
$P(T \geq t)$ 단측 검정		0.39	
t 기각치 단측 검정		1.67	

데이터 크기별로 분류하여 기본적인 통계량을 비교할 때, <표 4>에서 보듯이 이동평균법과 지수평활법은 데이터의 크기에 영향을 받지 않으므로 통계량이 차이가 없으며, 추세분석과 시계열분해법의 MAPE 값이 데이터 크기가 작을 때 좀 더 높을 것으로 기대했으나 대부분의 표본이 2005년부터 좀 더 안정적인 움직임을 보여 예측 오차가 더 낮게 나왔다. 그리고 데이터 크기와 분배 알고리즘의 정확성은 서로 상관관계가 없는 것으로 분석되었다.

<표 4> 분배알고리즘과 타 알고리즘과 통계량 비교

	비교대상	평균	편차	최대값	최소값
36개월 (3년)	이동평균	131.83	109.83	463.65	41.7
	지수평활	154.97	161.73	808.93	40.36
	추세분석	170.98	150.27	582.74	40.98
	시계열분해	97.19	70.15	395.4	26.11
	분배 알고리즘	95.53	68.56	365.79	21.91
24개월 (2년)	이동평균	131.83	109.83	463.65	41.7
	지수평활	154.97	161.73	808.93	40.36
	추세분석	115.24	83.03	428.79	42.06
	시계열분해	86.53	37.4	206.14	39.24
	분배 알고리즘	81.22	42.51	197.34	33.57

부품 주문에 관한 데이터의 평균 수요가 분배 알고리즘의 정확성에 영향을 끼치는지 알아본 결과 <표 5>와 같은 결과를 얻었다. 평균 수요의 크기는 분배 알고리즘의 정확성에 상관성을 보이지 않으며, 평균 수요가 1000 이상일 때 대체로 MAPE 값이 적은 수치를 보여 정확성이 있음을 알 수 있다.

<표 5> 평균 수요별 분배 알고리즘의 MAPE 평균

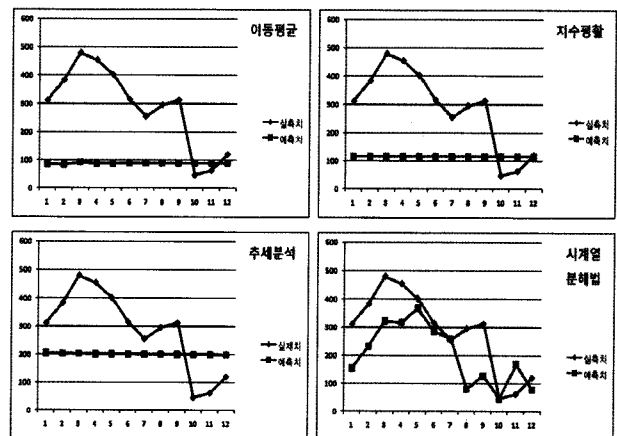
평균 수요 / 데이터 크기	$\bar{x} \geq 1000$	$300 \geq \bar{x}$	$100 \geq \bar{x}$	$\bar{x} < 100$
36개월	63.62	93.23	106.5	94.56
24개월	58.41	84.85	87.61	80.15

<표 6>은 연도별 가중비율 ω_t 를 달리 적용했을 때 MAPE 값이 가장 적은 적용비율이 어떤 것인가를 알아본 것이다. 기본적으로 몇 가지 경우를 적용하여, 데이터 크기와는 상관없이 5:5비율을 적용했을 때 MAPE 값이 가장 적었으며, 다음으로 9:1비율을 적용했을 때 적었다. 24개월 데이터 크기는 3년간의 자료로 한정해서 조사할 수 밖에 없었다.

<표 6> MAPE가 낮은 적용 비율이 채택된 개수

적용 비율 / 데이터 크기	9:1	8:2	7:3	5:5	5:3:2	합계
36개월	11	2	3	32	9	57
24개월	18	3	6	30	-	57

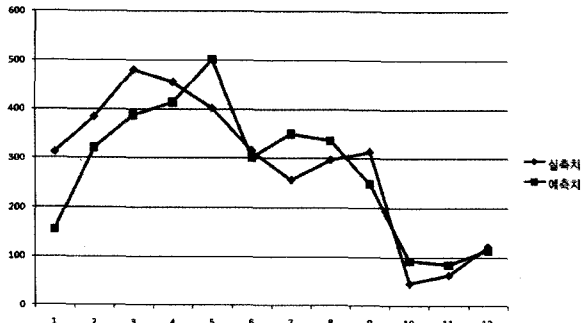
분배알고리즘과 기타 알고리즘에 대한 12개월간의 미래 수요를 투영하였다. <그림 11>은 랜덤 추출한 한 표본에 대해 기존 예측 알고리즘을 적용하여 산출된 예측치와 실측치에 대해 꺾은선 그래프로 나타낸 것이다.



<그림 11> 기존 예측기법 예측결과

<그림 11>에서 확인 되듯이 이동평균법, 수평활법, 추세분석법을 적용하면 12개월의 예측치가 일직선으로 표현되며, 시계열분해법은 실측치에 근사됨을 알 수 있다.

<그림 12>는 같은 표본을 이용하여 분배 알고리즘을 적용한 예측치와 실측치를 그래프로 나타낸 것이다. 이상의 데이터는 가중 비율을 5:3:2로 적용한 것이다.



<그림 12> 분배알고리즘 예측결과

실제 이론 및 실무적 과정에서 가장 많이 적용되는 수요예측기법과 본 연구에서 제시하는 분배 알고리즘을 비교한 결과 시계열 예측법과는 유사한 결과를 보여 주나, 분배 알고리즘의 MAPE 값이 <표 4>에서 보듯이 시계열분해법에 비해 평균적으로 약 6.76%가 낮은 것으로 분석되었고, 그 외 이동평균법, 지수평활법 및 추세분석에 비해서는 상당히 우수한 것으로 분석되었다.

분석 단계에서의 실험 계획에 준하여 연도별 가중비율이 주로 5:5비율을 사용했을 때 정확성이 높은 것으로 조사되어, 최근 2년간의 과거 수요가 예측을 하는데 유효하게 작용하는 것으로 보인다. 연구 결과 분배 알고리즘은 최소 24개의 데이터를 활용하여 시계열분해법 이상의 정확도를 보이고 있으며, 과거의 데이터를 일정 비율로 분배하는 방식을 갖기 때문에 단기예측 또는 장기예측 모두 MAPE가 동일하다는 점에서 분배 알고리즘은 수요예측 선택기준을 모두 만족한다고 할 수 있다. 그리고 시계열분해법과 비교하여 오차의 크기가 큰 차이를 보이지 않기 때문에 같은 오차라면 가급적 좀 더 쉬운 모형을 선택한다는 기준을 들어 분배 알고리즘의 적용 폭을 넓힐 수 있다고 사료된다.

5. 결 론

고객의 서비스 만족률, 재고유지, 비용 감소, 기업의 신용도를 높이기 위해서는 효과적인 수요예측이 이루어져야 한다. 본 연구는 수요예측 엔진을 구성하는 중요 요소로서 적절한 수요예측 알고리즘을 선택하는 방

법론과 미래의 수요에 투영시키기 위한 분배알고리즘을 소개 하였다.

수요예측 엔진의 구성은 알고리즘 풀을 구성하기 위한 예측 알고리즘에 대한 다양한 구성과 알고리즘에 관한 분석이라 할 수 있다. 이후의 프로그램 선택 기준을 실제 값과 가장 근접하는 예측 값을 유추해 내는 적절한 수요예측 알고리즘을 선택하는 것이다. 최적 알고리즘의 선택은 과거의 데이터에 대한 검증과 보완을 통해 예측오차를 기준으로 적용함을 구현하였다. 수요예측 알고리즘에 의한 예측 값과 분배알고리즘에 의한 예측 오차의 상호 보완적인 과정을 통하여 수요예측 엔진이 구현 가능하였다. 본 연구에서 수요예측 알고리즘의 선정과정과 분배알고리즘의 정확성을 검증하는 과정에서 몇 가지 문제점이 발견되었는데, 심각한 결측치가 존재하는 상황을 극복하기 위하여 결측치 보정을 통하여 보다 정확성을 기할 수 있다는 것이다. 실제 본 연구에서는 언급하지 않았으나 결측치 보정에 관한 모듈이 개발되어 수요 예측 실시 이전에 구현되었음을 밝혀드립니다.

서비스 수준과 재고비용의 대립된 관점에서 예측의 오차를 줄이기 위한 연구는 계속 진행되어야 하는데, 본 연구는 이러한 일련의 과정을 실제 기업의 주문자료 등을 적용하였고, 화면의 구성 또한 사용자의 편리성에 근거하여 구성하였다.

참고문헌

- [1] 권재현; "SCM 구축을 위한 협업적 수요예측 모형 개발 : 통신장비 제조 산업의 협업 수요예측 실제 사례 모형 연구", 인천대학원 석사학위논문, 2003.
- [2] 오홍용; "시계열 분석과 인공지능망을 이용한 단기 천연가스 수요예측", 한국과학기술연구원 석사학위논문, 1997.
- [3] 윤영진, 백종현; "소매업에서의 수요예측기법의 활용 사례 연구", 한국경영컨설팅학회, 8(4) : 2008.
- [4] 이 영; "서비스 부품 매출량을 이용한 데이터 분석과 그룹화에 관한 연구", 계명대 산업기술연구소, 32(1) : 37-42, 2009.
- [5] 이덕기; 예측방법의 이해, 초판, 고려정보산업, 1999.
- [6] 이재식; "Box-Jenkins 시계열 분석법을 이용한 국내 정유제품의 단기수요 예측", 한국과학기술연구원 석사학위논문, 1979.
- [7] 이충기, 송학준; "최적 시계열 수요예측 모델선정에 관한 연구", 한국관광학회, 31(6) : 2007.
- [8] 임석현, 최종열, 김창대; 생산운영관리, 삼영사, 2 : 2003.

- [9] 조신섭, 손영숙, 시계열분석, 초판, 을곡출판사, 1999.
- [10] Burger, G. et al.; "A practitioners guide to time-series methods for tourism demand forecasting-a case study of Durban, South Africa," *Tourism Management*, 22(4) : 2001.
- [11] Johnston F. R.; "Forecasting for items with Intermittent Demand," *Journal of Operation Research Society*, 47(1) : 113-121, 1996.
- [12] Kulendran, N. and Witt, S. F.; "Forecasting the Demand for International Business Tourism," *Journal of Travel Research*, 41(3) : 2003.
- [13] Yue, X. and Liu, J.; "Demand forecast sharing in a dual-channel supply chain," *European Journal of Operation Research*, 174(1) : 646-667, 2006.