

분배모형에 의한 철도 수요예측에서 영향인자에 대한 연구

오석문*, 홍순흠
한국철도기술연구원

A study on the effect factors of the railway passenger demand forecasting by the disaggregate model

Seog-Moon Oh*, Soon-Heum Hong
Korea Railroad Research Institute

Abstract - 본 논문에서는 철도 수요예측 문제의 유형을 목적에 따라 3가지로 분류하였고, 최근 철도자원을 재고관리 차원에서 접근하고자 하는 시각에 따라 분배모형으로써 적용필터를 사용하는 방법의 타당성에 대해 설명하였다. 또 철도 승객수요의 주요 특징을 분석하였으며, 철도 승객수요 예측의 요구사항 및 방법론을 대규모 재고관리 시스템의 일반적 요구사항에 따라 정리하였다. 영향인자에 대한 분석으로 요일별 계절변동 지수를 정량적으로 산정하였다. 적용필터를 이용한 철도 승객수요 예측의 예제를 제시하였으며, 예측에의 정확성에 대한 비교를 제시하였다.

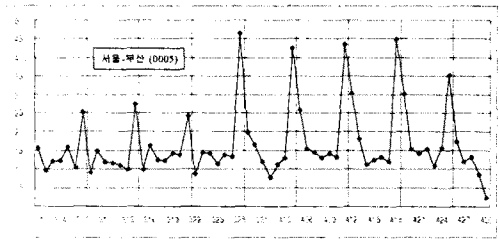


그림1 철도 승객수요의 시계열 패턴

1. 서론

대부분의 산업에서 수요예측의 문제는 계획업무의 기초가 되는 중요한 단계이고 이런 점은 철도산업에서 역시 마찬가지이다. 철도산업에서 수요예측의 문제는 목적에 따라 다음과 같이 분류될 수 있다. (1) 신선(新線)을 건설할 때 국토개발 차원에서 최선의 노선 및 선로용량을 결정하기 위하여 주요 기점간의 철도수요 예측, (2) 결정된 선로의 노선 및 용량에 대해 서비스 계획을 수립하기 위한 승객수요의 예측 (3) 결정된 서비스 계획에 대해 고객의 요구에 대한 자원할당량을 최적으로 조정하기 위한 승객수요의 예측.

철도수요의 예측에 대해 지금까지 국내에서 발표된 대부분의 논문 및 연구보고서들은 (1)의 문제에 집중되어 있고, 일부가 (2)의 문제에 대해 언급하고 있다. 또 방법론에 있어서도 대부분 Aggregate 모형을 따르고 있다.

본 논문에서는 (2)와 (3)의 문제에 접근하기 위해 분배모형(disaggregate)에 대한 필요성에 대해 언급하고, 분배모형으로써 적용필터를 사용하는 방법의 타당성을 실험 데이터를 통해 분석하였다. 또 철도수요 문제에서 발생하는 영향인자에 대해 정량적인 통계분석을 실시하고, 각 인자와 수요예측 모형의 관계에 대해 설명하였다.

2. 본론

2.1 철도 승객수요의 특징

시계열(time series)이란 시간의 경과에 따라 변동하는 변수를 관측한 결과들의 집합을 의미한다. 시계열에서 나타나는 변동은 성질을 달리하는 몇 가지 변동이 합성되어있는 것으로 본다[1].

철도 승객수요의 가장 큰 특징은 전형적으로 시계열 모형의 특성을 보인다는 점으로 이것은 (1)의 문제를 위해 정의되는 철도수요가 인과적 요인(causal factor)에 의해 설명이 잘되는 것과는 대조가되는 큰 차이이다. 그림 1은 철도 승객수요의 시계열 패턴으로 새마을호 0005 열차의 1998년 3월1일~4월30까지 서울~부산 승객의 실적 데이터이다.

철도 승객수요의 특징은 또 분명한 계절변동(seasonal variation, S(t))을 보인다는 점이다. 계절변동이 순환변동(cyclical factor, C(t))과 구분되는 점은 12개월 또는 1주일 단위로 비교적 단기의 반복적인 패턴을 보인다는 것이다[2]. 그림 1에서 시계열의 눈금은 매 7일 단위로 되어있다. 이 그림에서 보는 바와 같이 철도 승객수요는 주 단위(weekly-base) 계절변동을 분명히 드러내고 있다. 이러한 특성은 특정한 열차(0005)에서만 발견되는 것은 아니다. 그림 2는 같은 기간동안 서울역에서 09:00~12:00에 출발한 모든 승객(목적지가 부산인 열차에 탑승한 승객수의 총합)에 대한 시계열 데이터에서 자기상관계수(autocorrelation coefficient)를 구한 것으로 매 7-lag에서 거의 1에 가까운 자기상관(autocorrelation)을 보이고 있다.

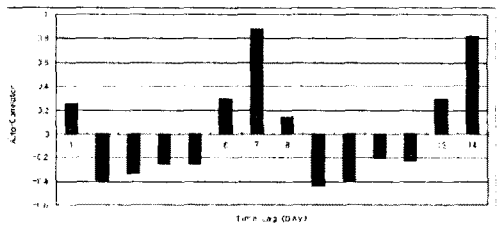


그림 2 철도 승객수요 패턴의 자기상관계수

철도 승객수요에는 주 단위의 계절변동 외에도 다양한 영향인자들이 존재한다. 이러한 인자들은 주로 시간적인 요소들로서 간략히 정리해보면 방학/휴가, 명절, 연휴 등으로 분류할 수 있고, 이러한 세부적인 사항들에 대한 각 OD간의 영향력은 철도 영업계획 작성단계에서 세부적으로 분석되어야 한다.

철도 승객수요의 또 다른 특징은 입석승객에 의해 발생한다. 현재 승객들은 열차가 출발하기 30일 전부터 출발직전까지 예약을 할 수 있다. 그림 3은 0111열차의 3월1일~4월30까지 DCP(Data Collection Points)별 탑승실적을 나타낸다. 그림에서 위쪽 그림은

좌석승객의 실적을 나타내고, 아래쪽 그림은 입석승객을 나타낸다. 그림에서 나타난 바와 같이 우리나라 승객의 예약 행태(behavior)는 출발전 수일 이내에 집중적으로 나타나고, 특히 입석승객의 경우에는 대부분의 승객이 출발직전에 나타나는 패턴을 보인다.

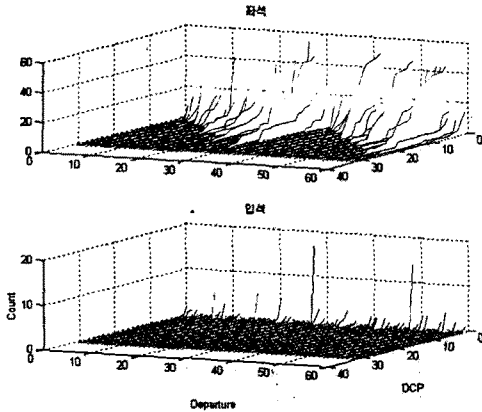


그림 3 DCP별 탑승실적

이러한 특성은 (2)와 (3)의 목적에 따른 수요예측의 경우 DCP별 탑승실적을 적절하게 사용하는데 어려움이 있음을 의미한다.

마지막으로 철도 승객수요는 시장에 따라 다른 패턴을 보인다. 시장(market)은 분석가의 고려대상이 되는 범위로서 특정 OD에 대한 시간대별 또는 열차별의 승객 수요의 집합이라고 할 수 있다. 즉 그림 1에서와는 다른 패턴들이 시장에 따라 다양하게 존재한다.

2.2 승객수요예측의 목적 및 방법론

(2)와 (3)의 목적에 따른 승객수요의 예측 값은 분석가가 정의하는 각 시장별로 제시되어야 한다. 이러한 방법은 지금까지 많이 수행되어왔던 철도 전체승객의 탑승 실적(passenger and/or passenger-km)에 대한 총합적 예측모형(aggregate forecast model)과는 다른 점이다. 승객수요의 예측 값을 각 시장별로 제시하기 위해서는 분배모형(disaggregate forecast model)에 따른 예측방법이 고려되어야 한다. 또 분배모형에 따른 예측 절차는 (2)와 (3)의 업무절차를 시스템화 할 경우 필수적인 과정이다.

정보시스템이 급속하게 발전한 현재에는 총합적 예측이 필요한 사례에서도 광범위하고 세부적인 자료를 활용할 수 있는 분배모형의 방법론이 주목을 받고있다[3].

현재 우리 철도에서 (2)의 문제에 대한 영업주기는 1년 이상으로 되어있고, 외국의 경우 6개월 또는 1년으로 되어있다. 이 경우 예측 모델은 각 시장별로 6개월 또는 1년을 대상으로 한 예측 값을 제시해야 한다. 또 (3)의 문제의 경우에는 대략 1일~1개월까지의 예측 값을 시장별로 제시해야 한다.

(3)의 문제에 대해서는 특별히 그림 3에서와 같은 DCP별 탑승실적의 시계열 데이터를 활용할 수 있다. 외국의 경우 시장별 자원관리 종료시점은 출발 1일 혹은 2일전까지이므로 예측모형은 최소 출발 1~2일전에는 가능한 최고의 정확성을 가진 값을 제시해야 한다. 반면 DCP별 탑승실적의 변화는 그림에서 보는 바와 같이 그 기간에 급격히 변화하므로 이 정보를 효율적으로 활용할 수 있는 방안은 연구과제라고 할 수 있다.

(2)와 (3)의 문제를 목적으로 한 승객수요 예측모형이 시스템화 될 경우에는 시장의 특성을 면밀히 분석하고, 정확한 모형을 적용할 뿐만 아니라, 예측후의 다양

한 분석절차를 제공해야 한다. 외국의 경우에는 예측 값에 대한 영업전략을 what-if 시뮬레이션으로 구성하여 수익성이 높은 영업계획을 작성하고 있다.

(2)와 (3)의 문제에 대한 이상과 같은 승객수요의 특성은 온라인 운용을 위한 필수적인 요구사항으로 받아들여지고 있다. 대규모의 재고관리 시스템(large-scale multi-item inventory control systems)을 위한 수요예측 방법론의 주요 요구사항을 정리하면 적응성(adaptiveness), 반복적 추정(recursive estimation), Fading memory, 경제성(computational economy), 강인성(robustness), 자기 초기화(self-initialization) 등을 들 수 있다[4].

다음 절에서는 지금까지 분석한 승객수요 예측의 특징, 목적 및 방법론을 고려하여 영향인자를 고려하고, 적응필터(adaptive filter)를 이용한 승객수요 예측의 예를 제시하였다.

2.3 영향인자를 고려한 수요예측

승객수요의 대표적인 영향인자인 요일을 계절변동 지수로 나타내기 위해 승법모형에서 이동평균을 대비하는 classical decomposition 방법을 사용하였다[1],[2]. 분석대상은 그림 2에서 사용한 데이터이고, 결과는 표 1과 같다. 표에서 보는 바와 같이 요일별 계절변동 지수는 {월}, {화}, {수}, {목}, {금}, {토}, {일}에 따라 다르게 나타나고 있다.

표 1 classical decomposition에 의한 계절변동 지수

Weedly-based Seasonal Index	Day							Total
	월	화	수	목	금	토	일	
1주	137%	120%	118%	112%	103%	64%	4%	
2주	137%	115%	119%	118%	100%	50%	5%	
3주	140%	123%	120%	120%	111%	40%	5%	
4주	140%	131%	121%	121%	127%	14%	2%	
5주	132%	135%	131%	123%	116%	10%	4%	
6주	166%	118%	155%	118%	93%	19%	5%	
7주	151%	150%	112%	124%	106%	3%	5%	
8주			129%	74%	107%	18%	7%	
MA	141%	125%	124%	118%	107%	23%	5%	98%
SI	142%	127%	129%	120%	106%	22%	5%	70%

MA: Moving Average
SI: Seasonal Index
Adjustment Factor: 750/692 = 1.01156

이러한 요일별 계절변동 지수에 따라 별도의 데이터 그룹으로 시계열을 구성(data granulation)하면 예측의 정확성 향상을 기대할 수 있다[5].

앞 절에서는 철도 승객수요 예측과 같은 대규모 재고관리 시스템에서의 수요예측 방법론으로 적응필터를 사용하는 것이 적합한 방법임을 밝혔다.

적응필터에 의한 수요예측 방법은 [6]에 의해 체계화된 이후 수많은 연구들이 뒤를 이었다[7]~[10]. 이 방법은 이동 평균법이나 지수 평활법 등과 같이 관측 값과 가중치의 곱에 대한 합으로 표현되는 예측방법이다.

$$S_{t+1} = \sum_{i=1}^N w_i x_{t-i+1} \quad (1)$$

$$w_i = w_i + 2ke_{t+1}x_{t-i+1} \quad (2)$$

(k: training coefficient, e: error)

적응필터를 사용하는 방법에서는 (2)식에 따라 가중치를 최적화 하는 과정을 추가적으로 수행한다. 현재까지의 관측 값에 의해 사후 예측 값(posteriori prediction)을 구하고, 사후 예측 값과 관측 값 사이의 오차를 최소화시키는 가중치를 steepest descent method를 이용하여 찾아내는 절차를 따른다.

이와 같은 과정은 시변 모수추정(time varying parameter estimation)의 전형적인 문제이고, 이러한 시변 모수추정의 문제가 칼만필터에 의해 효과적으로 다루어진다는 사실에 의해 이 분야의 연구가 진행되었다[11]~[13].

그림 5에서 적응필터를 이용한 승객수요의 예측에 대

한 예제를 제시하였다.

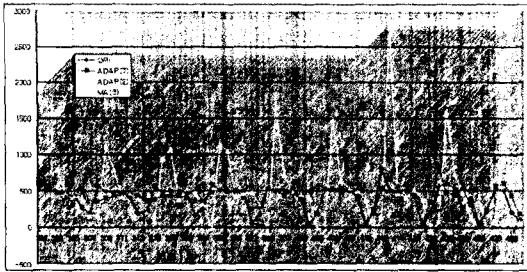


그림 5 철도 승객수요 예측의 예

예측 대상은 서울역에서 오전에 출발하여 부산으로 향하는 승객의 3월에서 4월까지의 수요이다. 그림에서 ORI는 관측 값, ADAP(7)와 ADAP(2)는 각각 자기회기분석 모형인 AR(7)와 AR(2)를 모형으로 한 경우이다. 이 그림의 결과를 제곱평균오차(MSE, Mean Square Error)로 정리하면 다음과 같다.

기법	ADAP(7)	ADAP(2)	MA(3)	MA(2)
MSE	13,071	253,990	53,258	54,565

$$MSE = \frac{\sum_{i=1}^n (\hat{x}_i - x_i)^2}{n} \quad (3)$$

적용필터에서 가중치의 수를 결정할 때는 관측 값의 잠재패턴(underlying pattern)을 고려하여 결정해야 한다. 위의 표에서 가중치의 수에 따라 MSE의 값이 현저하게 다르게 나타나고 있는데 이러한 이유는 그림 5에서 관측 값의 잠재패턴이 7일을 주기로 한 요일별 계절변동 지수가 분명한 것이 주요한 원인이 된다.

3. 결 론

이 논문에서는 철도 승객수요 예측의 목적을 경우에 따라 분류하여 각각의 특징에 대하여 논하였다. 특히 (2)와 (3)의 문제에 대해서는 대규모 재고관리 시스템을 위한 수요예측의 일반적인 요구사항과 비교하여 철도 승객수요 예측의 목적 및 방법론을 정리하였다.

우리 철도에서 승객수요의 영향인자 중 가장 큰 요인인 요일별 계절변동 지수를 정량적으로 분석하였다. 또 철도 승객수요를 적용필터를 이용하여 예측하는 예제를 제시했다.

차후 요일별 계절변동 지수와 적용필터를 동시에 적용하여 예측의 정확성을 향상시킬 수 있을 것으로 사료된다.

{참 고 문 헌}

[1] 김연형 편저, "시계열 분석과 예측", 자유아카데미, 1990, 6
 [2] S. C. Wheelwright, S. Makridakis, "Forecasting Method for Management", John Wiley & Sons, 1978
 [3] M. Ben-Akiva, S. R. Lerman, "Discrete Choice Analysis", MIT Press, 1997
 [4] Derek W. Bunn, "A Comparison of Several Adaptive Forecasting Procedures", OMEGA, Vol. 8, No. 4, PP. 485~491, 1980
 [5] 한국철도기술연구원, "철도종합정보 시스템 엔지니어링 기술개발 보고서", 1998
 [6] Steven C. Wheelwright, Spyros Makridakis, "An Examination of the use of Adaptive Filtering in Forecasting", Operational Research Quarterly, Vol.

24, No. 1, PP. 55~64, 1973

[7] Stuart Bretschneider, Robert Carbone, Richard L. Longini, "An Adaptive Approach to Time Series Forecasting", Decision Science, Vol. 10, PP. 232~244, 1979.

[8] E.R. Golder, J.G. Settle, "On Adaptive Filtering", Operational Research Quarterly, Vol. 27, No. 4, PP. 857~867, 1976

[9] Spyros Makridakis, Steven C. Wheelwright, "Adaptive Filtering : An Integrated Autoregressive /Moving average Filter for Time Series Forecasting", Operational Research Quarterly, Vol. 28, No. 2, PP. 425~437, 1977

[10] Steinar Ekern, "Forecasting with Adaptive Filtering : a Critical Re-examination", Operational Research Quarterly, Vol. 27, No. 3, PP. 705 - 715, 1976

[11] Douglas R. Kahl and Johannes Ledolter, "A Recursive Kalman Filter Forecasting Approach", Management Science, Vol. 20, No. 11, PP. 1325~1333, 1983, Nov.

[12] Greg Welch and Gary Bishop, "An Introduction to the Kalman Filter", <http://www.cs.unc.edu/~welch>, Sep. 17 1997

[13] Peter S. Maybeck, "Stochastic models, estimation, and control", Academic Press, Vol. 1, PP. 1~16, 1979