

# 신경망을 이용한 철도 수요 예측

## Forecasting the Demand of Railroad Traffic using Neural Network

신영근\*                      정원교\*\*                      박상성\*\*\*                      장동식\*\*\*\*  
Shin, Young-Geun      Jung, Won-Gyo      Park, Sang-Sung      Jang, Dong-Sik

---

### ABSTRACT

Demand forecasting for railroad traffic is fairly important to establish future policy and plan. The future demand of railroad traffic can be predicted by analyzing the demand of air, marine and bus traffic which influence the demand of railroad traffic. In this study, forecasting the demand of railroad traffic is implemented through neural network using the demand of air, marine and bus traffic. Estimate accuracy of the demand of railroad traffic was shown about 84% through neural net model proposed.

---

### 1. 서 론

중전 철도청의 만성 적자 해소를 위해 지난 2005년 철도공사로 전환하여 효율적 운영으로 이윤을 극대화하려고 하고 있으나 설립 초기 경영 자립에 어려움을 겪고 있다. 이를 위해 향후 2015년 내에 경영수지 흑자 전환을 목표로 최근 효율적인 인력 운용, 적자역 운영개선, 보유자산 매각, 전철 운임 현실화 등의 방안을 담은 경영개선 대책을 내놓았다. 이렇게 제시한 대책은 기업의 체질을 개선하고 불필요한 비용, 인력의 감축을 토대로 경영상의 효율성을 이끌어 내자는 것으로 단기적으로 보면 기업의 내부 효율성을 높여 수익의 급격한 증가를 이끌어 낼 수 있으나 이는 기업의 경영 효율성으로부터 비롯된 수익의 증가라기보다는 비용을 줄임으로써 발생하게 되는 수익의 증가이다. 이러한 대책은 기업 내외부의 희생으로부터 비롯되는 것으로 이에 대한 반발이 필연적으로 뒤따르게 되는 결과를 낳게 되며 장기적으로 기업의 경영 개선이 제대로 이루어졌다고 보기는 힘들다. 그러므로 이러한 대책을 급격하게 진행하기 보다는 장기적으로 수익을 높일 수 있는 방안을 함께 진행하여 지속적인 경영 개선을 이루어내는 것이 시급한 문제라 하겠다.

일반적으로 기업의 수익 증가는 보유자원의 효율적인 활용으로 최적의 서비스를 제공함으로써 수요를 충족시키는데서 비롯된다고 할 수 있다. 이를 위해서 철도 수송 서비스에 있어서 철도 시설의 투자 계획 수립이나 사업 평가에 큰 영향을 주는 철도 수송에 대한 정확한 수요 예측이 필요하다. 최근 철도 분야에 대한 민간 부분의 투자가 활성화되면서 수요 예측의 정확성이 더욱 요구되고 있는 실정이다.

수요예측을 하는데 있어 다양한 통계기법들이 광범위하게 사용되고 있다. 이러한 기법에는 MDA (Multivariate Discriminant Analysis), Logit Analysis, Regression Analysis 등이 있으며 시간에 따른 변동이 있는 철도 수송 수요와 같은 데이터는 Stochastic Process로 모델링되어 예측할 수 있다.[1]

---

※ 본 연구는 2007년 두뇌한국 21 사업에 의하여 지원되었음

\* 고려대학교, 석박사통합과정, 비회원

E-mail : toctop@korea.ac.kr

TEL : (02)3290-3900 FAX : (02)929-5888

\*\* 고려대학교, 석사과정, 비회원

\*\*\* 고려대학교, 연구전임강사, 비회원

\*\*\*\* 고려대학교, 교수, 비회원

시간 변동 데이터는 특징에 따라 정상, 비정상 시계열로 나눌 수 있으며 각각의 경우 서로 다른 예측방법이 사용된다. 다음과 같은 특징을 만족하면 정상 시계열 데이터로 간주하며 그렇지 못할 경우 비정상 시계열이라고 한다.[2]

- 1) 변수가 시간에 따라 균형으로 회귀하려는 성질을 가질 때
- 2) 아래의 세 가지 조건 만족하는 경우
  - 평균이 일정할 때
  - 분산이 존재하고 일정할 때
  - 두 시점사이에 존재하는 변수 간의 공분산이 시점보다는 시차에만 의존할 때

정상 시계열 데이터를 고려한 예측에는 자기회귀과정(AutoRegressive), 이동평균과정(Moving Average), 혼합된 자기회귀이동평균과정(AutoRegressive Moving Average) 등의 모형이 사용된다. 그러나 실제 현상에서 나타나는 시계열은 정상이 아닌 비정상인 경우가 대부분이다. 비정상 시계열은 시간에 따라 변하는 확률구조, 특히 평균과 공분산의 행태가 시간에 따라 달라지는 특징을 갖는다. 시계열의 비정상성은 현실에서 관측된 자료로부터 미래의 값을 예측하는데 정상을 전제로 개발된 예측모형을 적용하는데 장애가 된다. 따라서 비정상 시계열의 분석을 위해서는 평균이나 분산 등에서 나타나는 비정상성을 설명할 수 있는 더 일반화된 모형이 요구된다. 시계열은 대부분 어떤 회귀추세(Regression Trend)를 포함하거나 분산이 일정치 않고 위치에 따라 다르거나 아니면 어떤 계절적 요인에 의한 계절성 변동(Seasonal Variation)을 포함하게 된다. 이러한 비정상 시계열의 정상화를 위해 차분이나 계절 차분을 통하여 회귀 추세나 계절적 변동을 제거하기도 하였지만 이러한 차분 절차와 효과를 모형에 포함시키는 것이 더 효과적일 수 있다. 이를 위해 사용되는 모형이 자기회귀누적이동평균과정(AutoRegressive Integrated Moving Average) 등이 있다. 앞서 언급한 시계열 예측 방법은 철도 수요 예측에 적용하기 위해 필요한 가정이나 제약에 대한 분석 결여와 정확도 문제로 인해 세밀한 보완과정을 거쳐야 한다.

그러나 앞서 언급한 다양한 통계 기법들은 독립변수들의 선형적 결합을 통해 예측 모델을 표현하고 있어 현실에서 발생하는 데이터를 표현하고 예측하기에는 어려움이 있다. 이러한 단점을 보완하기 위해 최근 인공지능망 기법을 이용한 다양한 예측 사례가 많이 연구되고 있다. 신경망은 인간의 두뇌를 구성하고 있는 기본적인 구성요소인 뉴런과 기능적으로 유사한 디지털 뉴런을 서로 연결한 추론모델이다.[3] 신경망은 자체적으로 학습기능을 이용해 자료 및 상황의 변화에 따라 스스로의 기능을 개선시킬 수가 있다. 그리고 변수들간의 선형관계나 어떠한 확률 분포도 가정하고 있지 않기 때문에 다른 예측 방법들에 비해 현실 데이터에 대한 적응력이 크다고 할 수 있다. 이러한 특징으로 인해 본 연구에서는 신경망을 이용한 철도 수송 수요 예측에 대해 알아보하고자 한다.

본 연구에서는 국내 수송 수단 중 철도의 대체재로 이용될 수 있는 항공, 해운, 공로의 수요를 이용해 철도의 수요를 예측해 보았다.

## 2. 신경망 모델과 그 응용

신경망 모델은 노드, 링크, 활동 함수, 학습 알고리즘 등의 요소로 구성된 정보처리 시스템이다. 일반적으로 신경망 모델에는 입력 노드, 은닉 노드, 출력 노드 등 3가지 처리 마디가 존재한다. 입력노드는 시스템 외부로부터 데이터를 받아 시스템 내로 이들을 전송한다. 시스템 내에 자리 잡고 있는 은닉 노드는 전송된 입력값을 넘겨받아 그것들을 처리한 뒤 즉각적인 결과를 산출해낸다. 출력노드는 입력값과 현재 시스템 상태에 기준하여 시스템 출력값을 산출해낸다.

### 2.1 신경망 구조

신경망은 세 가지 노드 즉, 입력, 은닉, 출력 노드의 구조로 이루어진다. 입력 노드를  $x=(x_1, x_2, \dots, x_j)$ , 은닉 노드를  $z=(z_1, z_2, \dots, z_j)$ , 출력 노드를  $y=(y_1, y_2, \dots, y_k)$ 라고 하면 신경망 모형은 다음과 같이 수식화 가능하다.

$$z_j = f_j(a_j + \sum_{n=1}^i w_{nj} x_n) \quad (1)$$

$$y_k = f_k(a_k + \sum_{n=1}^j w_{nk} z_n) \quad (2)$$

식1의  $a_j, w_{nj}$ , 식2의  $a_k, w_{nk}$ 는 각각 입력노드와 은닉노드, 은닉노드와 출력노드 간의 연결강도이다. 은닉노드에서 입력노드에서 주어진 값들에 연결강도를 가중한 합을 계산한 후 이 값을 함수  $f_j$ 로 비선형 변환하여 출력노드로 전달하게 된다. 이때 함수  $f_j$ 는 Sigmoid 활성화 함수로 다음과 같이 나타낼 수 있다.

$$f_j = \frac{1}{1 + \exp(-x)}, 0 < f_j < 1 \quad (3)$$

은닉노드와 출력노드 간의 연결도 입력노드와 은닉노드 사이의 과정과 같이 이루어진다.

## 2.2 신경망 학습

신경망은 학습을 통해 초기 연결강도 값을 데이터에 적합한 값으로 변환하는 과정을 거치게 된다. 활성화 함수를 통과하여 나온 출력값과 목표값 간의 오차를 다음과 같이 구할 수 있다.

$$e = \frac{\sum_{n=1}^k (y_k - t_k)^2}{2} \quad (4)$$

학습절차로는 BP(Back-Propagation)알고리즘을 이용한다. 학습 과정은 다음과 같다. 출력 노드에서 입력노드 순으로 노드와 노드 사이의 연결강도를 오차 e를 최소로 하는 값으로 구해서 다시 설정하게 된다. 연결강도를 조정하면 후 다시 입력을 넣어 계산하게 되면 출력값은 앞서 나온 값보다 오차가 적은 값이 나오게 된다.

## 3. 연구모형 설계

### 3.1 신경망 모형 설계

본 연구에서는 국내 교통 수단인 항공, 해운, 공로의 수송 실적이 철도에 미치는 영향을 분석하여 이를 통해 철도 수요를 예측하는 시스템을 제안하고자 한다. 연구에서 사용된 신경망의 구조도는 그림 1과 같다.

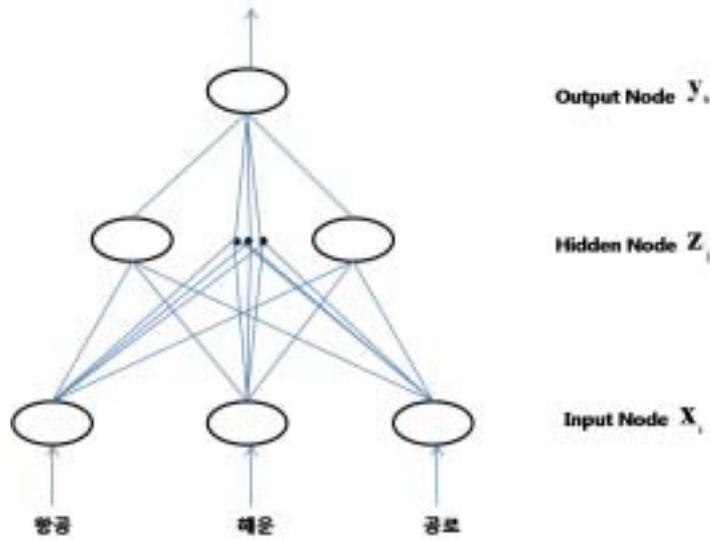


그림 1. 신경망 구조도

그림 1에서 보면 항공, 해운, 공로의 수요 데이터가 입력노드를 통해 들어가면 연결강도에 따라 값이 조절되어 출력노드로부터 출력값이 나오게 된다. 이때 해당 출력값과 목표값과의 오차를 계산하여 기존 연결강도를 출력노드로부터 입력노드까지 조절하게 되는 것이다. 이와 같은 과정을 통해 철도 수요 예측이 이루어진다.

### 3.2 데이터 및 실험 설계

본 연구에서는 2000년부터 2004년 월간 국내 여객 수송 실적을 사용하였다. 데이터는 국가교통데이터베이스에 공개된 자료를 사용하였다. 신경망 모형은 입력값과 출력값이 0에서 1사이 값일 때 최적의 성능을 보여주기 때문에 공개된 자료를 모두 0에서 1사이 값으로 변환하여 실험을 진행하였다.[4] 2000년부터 2003년까지의 데이터를 Training Data로 선택하여 신경망 모형을 설정하였으며 2004년 데이터를 Testing Data로 하여 모형의 예측 정확도를 검증하였다. 일반적으로 신경망 모형의 성과 측정에는 오분류율, 평균절대오차(MAE), 평균제곱오차의 제곱근(RMS)로 측정이 가능하다. 오분류율은 샘플에서 오분류된 수치를 총 샘플수로 나눈 값이다. 그리고 MAE는 실제 목표값과 출력값의 차이를 평균한 것이다. RMS는 이러한 값들을 토대로 모형의 정확도를 나타내는 값으로 오차값이 0일 때 모형이 100%의 예측정확도를 나타내는 것을 의미한다. RMS는 다음과 같이 정의할 수 있다.

$$e = \frac{\sqrt{\sum_{k=1}^N (y_k - t_k)^2}}{N} \quad (5)$$

$y_k, t_k$ 는 각각 출력값과 목표값을 나타내고 N은 샘플의 총 수를 나타낸다.

## 4. 실험 결과 및 분석

본 장에서는 신경망을 이용해 항공, 해운, 공로 수송 데이터로부터 철도 수요를 예측 정확도를 검증해보았다. 실험 결과는 표 1과 같다.

표 1. 철도 수요 예측 결과

	Number of Hidden Node	MAE	RMS
Training	2	0.19924	0.22564
	4	0.17334	0.22510
	6	0.16773	0.21477
Testing	2	0.19062	0.22004
	4	0.16650	0.16357
	6	0.16170	0.15911

신경망 모형에서 은닉노드의 처리 단위가 많은 경우 자료 내의 노이즈마저 모형 내에서 설명하려고 하는 Overfitting 문제를 보인다. 즉, 신경망 모형은 훈련에 이용한 자료를 과도할 정도로 정확하게 설명하려고해 자료의 모분포 또는 시스템에서 창출되는 표본의 자료에 대한 일반화가 어렵게 되며 예측성능 또한 좋지 못한 정확도를 보인다. 이러한 이유에서 본 연구에서는 최적의 은닉노드 처리단위를 구하기 위해 은닉노드가 각각 2, 4, 6개인 경우로 나누어 실험을 진행하였으며 각각 반복학습 500회를 수행하였다. 표 2에서 볼 수 있듯이 은닉노드의 처리단위가 6개인 경우 MAE, RMS 모두 가장 적은 값을 가지는 것으로 나왔다. 본 연구에서는 RMS는 처리단위가 6개인 경우 0.15911로 구축된 모형이 84.089%의 정확성을 가지고 있음을 알 수 있다.

## 5. 결론

철도청의 공사 전환 이후 다양한 수익 증대 방안이 제시되고 시도되고 있는 실정이다. 단기적으로 큰 수익을 가져다 줄 수 있는 방안이 있지만 이러한 방안은 대부분 기업 내외부의 회생을 요구하는 부분이 큰 것이 사실이다. 이에 장기적으로 수익을 지속적으로 증가시킬 수 있는 방안이 요구되고 있는 실정이다. 본 연구에서는 이러한 장기적인 수익 창출의 필요성을 인식하고 수익 창출을 위해 필요한 시설 투자 계획이나 사업평가에 큰 영향을 주는 철도수송의 수요예측에 대한 연구를 진행하였다. 최근 민간 분야의 철도 분야 투자가 활성화되면서 수요 예측의 정확성이 더욱 요구되고 있어 이러한 연구의 필요성이 상당히 크다고 하겠다.

철도 수송 실적에 큰 영향을 미치는 항공, 해운, 공로 수송 실적의 데이터를 통해 향후 철도 수요를 예측할 수 있는 방안을 신경망 모델을 통해 알아보았다. 2000년부터 2004년의 데이터로 신경망 모형을 통해 실험을 진행해본 결과 철도 수요 예측의 정확도는 84.089%인 것으로 나타났다. 이는 본 연구에서 제안한 모형이 철도 수송 수요 예측에 있어 매우 유용하다는 것을 보여준다. 그러나 본 연구는 다음과 같은 문제점을 가진다. 첫째, 지난 2005년 고속열차의 개통으로 인해 2004년 이전까지 항공, 해운, 공로가 미치는 영향이 2005년 이후부터 달라졌지만 이에 대한 공개된 자료가 없어 현실에 필요한 수요예측 결과를 제공하지 못하였다. 이 부분은 향후 최근 실적치가 공개되게 되면 본 모형의 예측력을 평가할 수 있을 것으로 생각한다. 둘째, 신경망 모형의 모수값을 결정하는 데 있어 기존의 연구 문헌에서 좋은 결과를 보이는 수치를 사용하였다는 것이다. 모수값 결정의 경우 실험 전 미리 지정을 해주어야 하는 부분으로 실험에서 사용한 값이 모든 경우에 최적의 값이라고 말할 수 없다.

## 참고문헌

1. 이덕규, 홍대화, 김학배, 우광방(1998), "Data Granulization 을 이용한 수송수요예측에 관한 연구", 한국철도학회 춘계학술대회 논문집, pp.211-218
2. 이상열(2001), "시계열분석의 원리", 자유아카데미
3. 이건창, 김명중, 김혁(1994), "기업도산예측을 위한 귀납적 학습지원 인공신경망 접근:MDA, 귀납적 학습방법, 인공신경망 모형과 성과비교" 경영학연구 23(4).

4. Michael Berry, Gordon Linoff(1997), "Data Mining Techniques", Wiley Computer Publish