

회귀모형과 신경망모형을 이용한 아파트 가격 모형에 관한 연구

A Study on Apartment Price Models Using Regression Model and Neural Network Model

김태훈 (Kim, Taehun) 경성대학교 상경대학 디지털비즈니스학부 e-비즈니스 전공
홍한국 (Hong, Hankuk) 동의대학교 상경대학 경영정보·인터넷비즈니스학부 경영정보전공

※ 주요단어: 신경망모형, 아파트 가격, 특정가격함수, 회귀모형

목차

I. 서론

II. 선행 연구

III. 연구 모형

1. 분석 데이터와 변수
2. 연구 모형과 절차

IV. 회귀모형을 이용한 아파트 가격 추정

1. 기초 통계량 분석
2. 회귀모형을 이용한 분석

V. 신경망모형을 이용한 아파트 가격 추정

1. 신경망모형의 고찰
2. 신경망모형을 이용한 분석

VI. 회귀모형과 신경망모형의 비교 분석

VII. 결론

I. 서론

우리나라 주택에서 상당한 부분을 차지하고 있고 주택 건설의 대부분¹⁾을 차지하고 있는 아파트에 대한 연구는 과거에서 현재까지 계속 이어오고 있다. 특히, 다양한 아파트 특성들을 이용하여 아파트 가격을 추정하고 예측하는 연구 또한 많은 실정이다. 그렇지만 이러한 연구들 대부분이 회귀모형에 지나치게 의존하고 있다. 회귀모형은 등분산성, 독립성, 정규성 등의 가정과 다중공선성의 문제, 특이값에 민감하게 반응하는 문제점 등이 있으며, 또한 선행 연구들 대부분이 모수의 비선형을 고려하여 수행한 연구는 많지 않다. 그리고 회귀모형을 적용함에 있어서 최적의 회귀모형을 선택하였는가에 의문점이 있을 수도 있고, 다양한 아파트 특성들에 대한 정성적 평가의 문제 제기도 있을 수 있으리라 판단된다.

그러나 회귀모형은 단점보다 장점이 많은 모형이다. 본 연구는 회귀모형을 부정하기보다는 새로운 모형을 도입하여, 회귀모형의 문제점을 극복하고 회귀모형과 상호보완적인 모형을 소개하고자 본 연구를 수행하였다. 현재까지 인공지능 분야에서 널리 이용되어 왔던 신경망모형 (Neural Network Model)은 입력변수가 불완전하고 변동폭이 넓은 경우에도 해석이 가능하며, 데이터 수가 적거나 불규칙한 경우라도 사례의 반복학습을 통해 오차를 줄여나가기 때문에, 데이터 수에 민감한 영향을 받는 회귀모형보다 정밀한 산정이 가능하다 (박우열, 차정환, 강경인, 2002). 이러한 신경망모형에 아파트 특성들을 도입하여 아파트 가격을 정밀하고 유효하게 예측하고, 기존의 회귀모형들과 비교하는 것은 아파트 가격에 대한 연구 분야에 큰 의미가 있다 하겠다. 그리고 주택에 관한 기존 연구와 신규 연구에 신경망모형이 활용될 수 있으리라 판단된다.

다양한 아파트 특성들에 대하여, 신경망모형을 이용하여 아파트 가격을 예측하고 기존의 회귀모형과 비교하는 것이 본 연구의 주목적이다. 또한 회귀와 신경망모형의 상호보완성을 찾는 것은 본 연구의 부차적인 목적이 된다. 아파트 특성들은 주변에서 쉽게 이용가능한 데이터를 위주로 하였다. 2004년 6월 기준으로 서울시 송파구와 도봉구의 아파트 매매가격들과 12개의 아파트 특성들을 수집하였다. 아파트 매매가격들 (즉, 매매 하한가, 일반 거래가, 매매 상한가) 대신에, 새로운 측정방법을 이용하여 하나의 매매가격으로 추정하였으며, 대표성을 가지도록 하였다.

연구 내용은 다음과 같이 구성된다. 2절은 아파트 가격 모형에 대한 선행 연구를 본 연구관점에서 비교한 것이다. 즉, 기존의 회귀모형과 신경망모형을 아파트 가격관점

1) 2003년 건설 주택수는 585,382호인데, 아파트는 468,763호 (80%)를 차지하고 있다 (통계청의 통계정보시스템 <http://kosis.nso.go.kr/>)

에서 비교하였다. 3절은 본 연구의 연구 모형과 수집된 데이터에 대하여 설명하였다. 4절은 회귀모형을 이용하여 아파트 가격을 추정하고, 5절은 신경망모형을 이용하였다. 6절은 회귀모형과 신경망모형을 비교 분석하였고 그에 대한 설명을 부가하였다. 7절은 본 연구의 기여, 한계점, 그리고 향후 연구 과제를 제시하였다.

II. 선행 연구

기존의 아파트 가격 책정 방법은 원가연동제, 아파트 채점표, 시장자료접근법, 특정 가격함수법 (Hedonic Price Function Method) 등이 있다 (안병욱, 2001)²⁾. 사업지 인근의 기존 아파트 가격에 대해 우열 가중률을 적용하여 산출하는 것이 아파트 채점표 방법이고, 시장조사에 의해 시행하는 것이 시장자료접근법이다. 이러한 방법들은 과학적이거나 수리적인 도구를 활용하지 않고 필요에 따라서 자의적으로 이용되기도 한다 (안병욱, 2001). 특정가격함수법은 과학적이고 수리적인 기법으로 평가되고 있다. 이것은 아파트 가격과 아파트 특성간의 관계를 회귀모형으로 표현한 것으로써 아파트 가격을 종속변수로, 아파트 특성들을 독립변수로 설정하는 것이다. 기존의 주택가격 관련 연구들은 대부분 이와 같은 회귀모형을 이용하여 주택 매매가격의 예측 및 특성 가격 추정을 가능하게 하는 주택가격모형을 설정하였다 (구본창, 손현영, 2001).

종속변수와 독립변수들과의 변화관계를 함수로 표현하는 데 있어서, 지금까지는 주로 회귀모형이 사용되어 왔으나, 최근에는 신경망이 많이 사용되고 있다 (이문규, 허해숙, 1995). 특히, 회귀모형과 신경망모형의 비교분석 연구들이 다양한 분야에서 수행되었으며, 그러한 연구의 결과 대부분이 회귀모형보다 예측 측면에서 신경망 모형이 더 우수한 것으로 평가되고 있다. 즉, 기업도산예측 분야, 원가예측 분야, 주택 프로젝트 사업비 예측 (박우열, 차정환, 강경인, 2002), 홍수위 예측등이 있다.

<표 1> 은 기존의 아파트 가격 예측을 위한 회귀모형과 신경망 모형을 비교하였고, 각 모형의 장단점이 정리되어 있다 (박우열, 차정환, 강경인, 2002; 이문규, 허해숙, 1995; 최병남, 1994). <표 1> 에서 회귀모형은 아파트 가격 예측에 대한 최근 3년간 (2002년 ~ 2004년) 의 연구로써, 다양한 아파트 특성들 (예를 들어, 물리적 요소, 입지적 요소, 환경적 요소 등) 을 이용하여 아파트 가격을 예측한 것이다. 이러한 회귀모형은 과거 10년 이상 사용되었고, 계속해서 이러한 모형에 관한 연구가 지속될 것이다. 이에 비해 아파트 가격 예측을 위한 신경망모형에 대한 연구는 거의 전무한 실정이다. 국외에서는, Do & Grudniski (1993), Worzala, Lenk, & Silva (1995), Nguyen &

2) 국토연구원 (2004)에서는 주택수요분석으로 주택수요함수, 특성감안 가격함수, 주택가격 예측으로 ARIMA 모형, VAR 모형, 연립방정식 모형으로 분류하고 있다 (pp276-321).

Cripps (2001) 등이 신경망 모형을 통해 특정 지역에 대한 판매가격을, 다양한 특성을 이용하여 예측하였고, Buja and Lee (2001) 는 Harrison and Rubinfeld 가 연구한 보스턴 지역의 주택 가격을 데이터 마이닝에 적용하여 예측하였다. 국내에서는 아직까지 이러한 연구가 나타나지 않고 있다. 즉, 이러한 모형에 대한 이해가 충분하지 않은 실정이다. 주택 관련 연구 분야에서는 초기 사업비 예측과 토지평가 모형에 신경망모형이 이용되었다 (박우열, 차정환, 강경인, 2002; 최병남, 1994). 특히 박우열, 차정환, 강경인 (2002) 연구의 문제점은 학습 데이터의 개수가 너무 적다는 것이다. 방법론적인 문제는 전혀 없으나, 너무 적은 학습 데이터 개수로 인해서 예측 신뢰가 떨어진다.

<표 1> 회귀모형과 신경망모형의 주택분야 선행 연구

분석 방법	분석방법의 장점	분석방법의 단점	연구자	연구내용	특징
회귀 모형	<ul style="list-style-type: none"> 신경망모형에 비해 해석이 쉽고 이해하기 쉬움 예측에 대한 설명력이 어느 정도 가능함 	<ul style="list-style-type: none"> 가정과 제약 조건이 많음 특이값과 데이터 수에 민감 비선형인 경우에 문제 발생 	구본창, 손현영(2002), 이변송, 정의철, 김용현(2002), 정성훈, 강준모(2002), 최열, 공윤경(2003), Bin(2004)	<ul style="list-style-type: none"> 다양한 아파트 특성들을 이용하여 아파트 가격에 대한 예측모형을 개발 	<ul style="list-style-type: none"> 아파트 가격 예측모형에 사용됨
신경망 모형	<ul style="list-style-type: none"> 계약조건 (예를 들어, 다중공선성, 분포 등)에 무관함 데이터 수에 둔감 주어진 문제가 제대로 정의되지 않았거나 문제에 대한 이해가 부족할 때 유용 비선형 함수관계를 포착할 수 있음 다각도의 문제점 처리에 용이 복잡한 분야의 자료에서도 좋은 결과 도출 가능 연속형과 이산형 자료 모두 처리가 가능 	<ul style="list-style-type: none"> 학습인자를 잘못 선택할 경우 오차율이 더 큼 적절한 은닉층과 은닉마디의 수를 결정하기 위해서 시행착오적인 방법 구사 해석이 어렵고 복잡함 과잉추정 (Overfitting) 의 문제 바람직하지 못한 결과치로 수렴 가능성 	Do & Grudniski (1993)	<ul style="list-style-type: none"> 판매가를 다양한 특성을 통하여 예측 출력변수: 판매가 (Selling Price) 입력변수: 주택연식 (Age), 욕실수, 방수, 총면적, 층개수, 차고개수, 벽로개수, 토지 면적 	<ul style="list-style-type: none"> 지역: The southwestern part of San Diego, California country 가구: 242 single-family homes 기간: 1991년 1월 ~ 1991년 9월
			Worzala, Lenk, & Silva (1995)	<ul style="list-style-type: none"> 판매가를 다양한 특성을 통하여 예측 출력변수: 판매가 입력변수: 욕실수, 방수, 주택연식, 토지 면적, 총면적, 벽로 (Fireplace) 개수, 차고개수, 스타일 (bi-level, tri-level, 4-level, 2-story, ranch) 	<ul style="list-style-type: none"> 지역: Fort Collins metro area 가구: 288 single-family homes 기간: 1993년 11월 ~ 1994년 1월
			Nguyen & Cripps (2001)	<ul style="list-style-type: none"> 판매가를 다양한 특성을 통하여 예측 출력변수: 판매가 입력변수: 주거 면적, 방수, 욕실수, 주택연식, 판매분기, 차고유무 	<ul style="list-style-type: none"> 지역: Rutherford Country, Tennessee 가구: 3,906 single-family homes 기간: 1993년 1월 ~ 1994년 6월 신경망 모형과 회귀 모형 비교
			Buja and Lee (2001)	<ul style="list-style-type: none"> 평균집값을 다양한 특성을 통해 예측 출력변수: 평균집값 입력변수: 범죄율, 주택용 부지비율, 중대형 버스비율, 하천인접여부, 산화질소농도, 방의 평균개수, 1940년 이전의 주택비율, 근무지까지의 가중거리, 주요도로까지의 인접성, 세율, 초·중고등학교 교사비율, 흑인비율, 중·하류층 비율 	<ul style="list-style-type: none"> 보스턴 지역의 주택 가격 데이터 예측 506개의 관측치 13개의 독립변수 (12개는 연속형, 1개는 이산형)
			박우열, 차정환, 강경인(2002)	아파트의 초기사업비 예측을 위해서 회귀모형과 신경망 모형을 비교	아파트 가격예측모형에 사용된 것은 아님
			최병남 (1994)	인공신경망을 이용하여 토지평가모형	

				개발 • 지역별 토지특성 변수 17~19개	
--	--	--	--	----------------------------	--

III. 연구 모형

1. 분석 데이터와 변수

본 연구의 데이터 수집 시점은 2004년 6월 15일에서 6월 30일까지이며, 분석 대상 아파트는 서울시 송파구와 도봉구 전체로 한정하였다. 수집 원천 (Source) 은 국민은행 아파트 시세 홈페이지와 닥터아파트 아파트 시세 홈페이지이며, 두 홈페이지에서 수집가능한 모든 변수들을 정리하였다. 즉, 구, 동, 아파트명, 브랜드명, 평형, 매매하한가, 일반거래가, 매매상한가, 방수, 욕실수, 세대수, 현관구조, 입주시기, 난방, 지하철 거리 등이 수집되었다³⁾. 이러한 변수들 중에서 구, 동, 현관구조, 난방을 범주화시켜 코딩하였다. 그리고 브랜드명은 시공회사를 의미하는 것으로써, 대한건설협회에서 제공한 2003년도 시급(도급) 순위 자료에 의거하여 순위 번호로 범주화시켰다⁴⁾. 입주시기는 현재 시점까지의 사용기간 (월 단위) 으로 환산하였다.

<표 2> 분석 대상 데이터

ID	구 코드	동 코드	브랜드 코드	평형	매매 하한가	일반 거래가	매매 상한가	추정 매매가격	방수	욕실 수	세대 수	현관 구조 코드	입주시기 계산: 사용기간 (월)	난방 코드	지하철 거리 (분)
1	1	1	12	34	37,000	39,000	41,500	39,295	3	2	62	2	6	1	3
2	1	1	12	44	45,500	50,000	52,500	49,820	4	2	32	2	6	1	3
3	1	1	999	13	30,750	30,750	32,000	31,087.5	2	1	1650	2	275	2	2
...
753	2	17	1	20	11,500	12,000	12,500	12,015	2	1	21	2	65	2	3
754	2	17	1	24	14,250	14,600	15,000	14,621	2	1	158	2	65	2	3
755	2	17	1	28	16,250	17,000	17,250	16,917.5	3	1	65	2	65	2	3

3) 국민은행 홈페이지에서는 매매상한가, 일반거래가, 매매하한가를 중심으로 수집하였고, 닥터아파트 홈페이지에서는 아파트 특성들을 중심으로 수집하였다. 그러나 하나의 홈페이지에만 존재하는 일부 아파트들에 대해서는 닥터아파트 홈페이지를 기준으로 자료를 수집하였다. 또한 국민은행 홈페이지에 존재하지 않는 아파트 가격에 대해서는 닥터아파트 홈페이지의 매매상한가와 매매하한가를 구한 다음, 일반거래가를 매매상한가와 매매하한가의 평균값으로 추정하였다. 그리고 데이터가 없는 경우 (예를 들어, 지하철 거리에 대한 데이터 등), 결측치 (Missing Value) 로 처리하였다.

4) 2003년도 시급(도급) 순위 자료를 100위까지만 활용하고 100위 미만 기업들에 대해서는 기타 (즉, 999) 로 분류하였다. 그리고 도수가 많은 한양과 동아는 997, 998 로 코딩하였다. 그리고 대기업의 경우, 다른 계열사로 시공된 아파트일 경우 같은 브랜드로 나타내었다. 예를 들어, 삼성물산, 삼성중공업 등은 같은 브랜드로 표시하였다.

<표 2> 는 수집된 자료를 분석에 맞게 변환하여 최종적으로 구한 데이터이며, 총 755개의 관측치 (Observation) 를 가진다. <표 2> 에서 음영으로 나타난 것이 분석 대상 변수들이며 <표 3> 에서 자세히 설명하였다. 즉, <표 3> 은 아파트 가격을 예측하기 위하여 사용된 아파트 특성들을 정리한 것이다.

<표 3> 변수 구성

구분	변수명	단위	변수설명
종속(출력) 변수	추정 매매가격	만원	세대수를 이용하여 매매가격을 새롭게 추정
	구코드	더미	송파구=1, 도봉구=2
독립(입력) 변수	동코드	더미	가락동=1, 거여동=2, ..., 쌍문동=16, 창동=17
	브랜드코드	더미	현대=1, 대우=2, 삼성=3, 기타=999
	평형	평	아파트 평형
	방수	개	방의 개수
	욕실수	개	욕실의 개수
	세대수	세대	동일 평형에 대한 유형 (예를 들어, 33평 A형)의 개수
	현관구조코드	더미	계단=1, 복도=2, 타워=3
	사용기간	월	입주시기 (준공년수)를 월단위로 환산
	난방코드	더미	지역=1, 개별=2, 중앙=3
	지하철거리	분	지하철까지 도보로 걸리는 시간

<표 2> 의 추정 매매가격은 기존의 매매상한가와 매매하한가의 평균으로 계산하는 것의 단점을 보완하여, 새로운 측정방법으로 추정된 것이다. 일반적으로 아파트 평균 매매가격을 추정할 경우 매매상한가와 매매하한가의 평균가격을 적용하는 데, 이 때 전체 세대 중 상한가와 하한가 세대가 반반이라는 것을 전제로 한다. 그러나 아파트 가격의 범위 (Range) 를 아는 것이 목적이 아니라, 아파트 가치를 추정하고자 할 때에는 문제가 발생한다 (박진우, 이기재, 김재관, 김진억, 2004).

<표 4> 서울시 구별 아파트 가격대별 세대수의 비율

지역	매매하한가 세대비율	일반거래가 세대비율	매매상한가 세대비율
서울시 전체	0.2	0.54	0.26
송파구	0.19	0.54	0.27
도봉구	0.18	0.61	0.21

<표 4> 는 서울시 전체와 서울시 각 구별로 동일 단지 내 동일 평형 아파트 세대들을 매매상한가, 매매하한가, 일반거래가로 나누다고 했을 때 각 범주에 해당되는 구성비를 나타낸 것이다⁵⁾. 이러한 구성비를 이용하여 송파구와 도봉구 아파트의 매매가격

5) 이러한 자료는 국민은행 청약사업팀이 서울시내 전체 아파트 단지에서 표본으로 추출하여 조사한 결과이다

을 추정하면 다음의 식 (1.1) 과 같다:

$$\begin{aligned} \text{송파구 A 아파트 추정 매매가격} &= \frac{A_1 * M * 0.19 + A_2 * M * 0.54 + A_3 * M * 0.27}{M} \\ \text{도봉구 B 아파트 추정 매매가격} &= \frac{B_1 * N * 0.18 + B_2 * N * 0.61 + B_3 * N * 0.21}{N} \end{aligned} \quad (1.1)$$

A_1, B_1 : 하한가 A_2, B_2 : 일반거래가 A_3, B_3 : 상한가 M, N : 각 아파트 평형별 세대수

<표 2> 에서 ID 가 1인 관측치의 추정 매매가격을 계산하면 다음과 같다:

$$\frac{37,000 * 62 * 0.19 + 39,000 * 62 * 0.54 + 41,500 * 62 * 0.27}{62} = 39,295 \text{ (만원)}$$

기존의 매매상한가와 매매하한가의 평균가격은 식 (1.1) 의 새로운 추정 매매가격보다 과소 추정되는 (Underestimated) 것으로 조사되었으며 (박진우, 이기재, 김재광, 김진익, 2004), 본 연구에서는 식 (1.1) 을 이용하여 <표 2> 의 추정 매매가격을 도출하였고, 이것을 분석에 이용하였다.

2. 연구 모형과 절차

본 연구는 아파트 가격을 예측하기 위하여 회귀모형과 신경망모형을 이용하였다. 이러한 연구 모형과 그 절차는 다음과 같이 요약할 수 있다: 연구목적에 기반하여 회귀모형과 신경망모형의 관점에서 선행 연구들을 고찰한다. 그리고 분석 대상이 되는 데이터를 수집한다. 수집된 데이터 중에서 매매가격들의 대표값을 추정한다 (즉, 추정 매매가격). 그리고 신경망모형에 대한 이론적 고찰을 수행한 후, 데이터를 두 가지 모형에 적합시킨다. 최종적으로 두 모형에 대한 비교·분석을 수행한다. 이를 통하여 결론을 유도하고, 본 연구의 기여와 한계점을 기술한다.

IV. 회귀모형을 이용한 아파트 가격 추정

1. 기초 통계량 분석

(박진우, 이기재, 김재광, 김진익, 2004, p83).

분석 대상 데이터에 대한 기초 통계량을 살펴보면 <표 5> 와 같다. 평형의 평균값은 34평이며, 최소 8평부터 102평까지 분포하고 있다. 추정 매매가격의 평균은 3억 7천만원이며, 최소 6천 백만원부터 16억 4천 2백만원까지 분포되어 있다. 분석 지역으로 보면, 송파구 453개, 도봉구 302개 아파트가 분석대상이다. 동코드 6 (삼전동), 7 (석촌동) 은 도수가 적기 때문에 분석에서 제외될 수 있을 것이다. 또한 현관구조코드에서 타워형 현관구조의 도수가 다른 구조에 비해서 상당히 적음을 알 수 있다. 마지막으로 브랜드코드에서 기타의 도수가 전체 도수의 44%를 차지하고 있다. 기타라고 분류된 아파트들은 주택공사, 시영, 영세업체 등이 시공한 것들이다. 특히, 재개발 아파트 등이 포함되어 있기 때문에, 아파트 가격이 상당히 왜곡되어 있을 것이다.

<표 5> 분석 대상 데이터의 기초 통계량

변수명	평균	표준편차	최소값	최대값
평형	34.29934	15.45415	8	102
추정 매매가격	37085.47	26781.9	6172.5	164250
방수	3.059603	0.914311	1	8
욕실수	1.580133	0.507114	1	3
세대수	191.396	350.4435	1	4320
사용기간	128.1762	83.30553	2	353
지하철거리	6.67947	4.59997	1	20
변수명	코드값	도수	백분율	
구코드	1	453	60%	
	2	302	40%	
동코드	1	71	9.40%	
	
	6	1	0.13%	
	7	2	0.26%	
	
	17	131	17.35%	
브랜드코드	1	102	13.51%	
	
	999	331	43.84%	
현관구조코드	1	485	64.24%	
	2	261	34.57%	
	3	9	1.19%	
난방코드	1	218	28.87%	
	2	376	49.80%	
	3	161	21.32%	

2. 회귀모형을 이용한 분석

본 연구에 사용된 회귀모형의 초기 완전모형 (Initial Full Model) 은 다음 식 (1.2)

와 같다. 식 (1.2) 처럼 이산형 변수는 더미변수로 변환하였다⁶⁾.

$$\begin{aligned} \text{추정 매매가격}_i = & \beta_0 + \beta_1 \text{평형}_i + \beta_2 \text{방수}_i + \beta_3 \text{욕실수}_i + \beta_4 \text{세대수}_i + \beta_5 \text{사용기간}_i \\ & + \beta_6 \text{지하철거리}_i + \beta_7 \text{구코드} + \beta_8 D_{\text{동1}} + \dots + \beta_{23} D_{\text{동16}} + \beta_{24} D_{\text{브랜드1}} + \dots + \beta_{62} D_{\text{브랜드39}} \\ & + \beta_{63} D_{\text{현관구조1}} + \beta_{64} D_{\text{현관구조2}} + \beta_{65} D_{\text{난방1}} + \beta_{66} D_{\text{난방2}} + \epsilon_i \quad (1.2) \end{aligned}$$

이러한 식 (1.2) 을 SAS REG 로 적합시킨 후, 모형에 대한 회귀 진단 (Regression Diagnostics) (허명회, 서혜선, 2001) 을 수행한 결과 다중공선성 (Multicollinearity), 특이값 (Outliers) 그리고 잔차의 이분산성⁷⁾ 등의 문제가 나타났다. 이러한 문제들을 해결하기 위하여 다음의 방법을 사용하였다: 첫째, <표 5> 에서 관측도수가 적은 경우와 브랜드가 기타인 경우를 삭제하였고⁸⁾, 다중공선성의 문제를 해결하기 위하여 변수 선택 (Model Selection) 방법에 의하여 축소모형을 선택하였다⁹⁾. 둘째, 잔차가 이분산성이지만 비선형성을 보이지 않았기 때문에, 일반적으로 사용되는 종속변수에 대한 로그 변환 (즉, Semi-로그변환) 을 수행하였다. 셋째, 표준화 잔차 (Student Residual) 와 Cook's D 통계량에 의해 영향력이 큰 특이값을 제거하였다.

이러한 과정들을 거쳐서 최종 모형을 구했으며, 최종 모형의 모수 추정치는 SAS GLM 을 통하여 수행하였다. 그 결과는 <표 6> 과 <표 7> 에 제시되어 있다. <표 6> 은 모형의 변수들이 유의하다는 결과를 보여주고 있다. 그리고 최종 모형의 설명력 (R^2) 은 95% 이고 오차항의 표준편차 ($\hat{\sigma}$) 는 0.132990 (즉, Root Mean Square Error; RMSE) 으로 나타난다.

<표 6> 분산분석표

Source	자유도	제곱합 (Sum of Squares)	평균제곱 (Mean Square)	F 값	Pr > F
모형	54	124.0355214	2.2969541	129.87	<.0001
오차	345	6.1017714	0.0176863		
합	399	130.1372928			

모형의 변수	자유도	제곱합	F 값	Pr > F
동	13	67.00754068	87.63	<.0001
브랜드	34	17.97414	9.01	<.0001
난방	2	1.610634	25.95	<.0001

6) 각 더미변수들과 다른 변수들 사이에 상호작용은 없는 것으로 가정하였다.

7) 다중공선성에서는 분산확대인자 (Variance Inflation Factor) 값이 10을 넘는 것이 있었으며, 잔차와 \hat{y} (즉, 추정 매매가격의 예측치) 의 플롯 (Plot) 이 이분산성을 나타내었다.

8) 기타를 제외한 경우와 포함한 경우를 비교한 결과, 제외한 경우의 설명력이 더 높았고 맬로우즈 C_p 통계량이 작은 것으로 나타났다.

9) 현관구조, 세대수를 제외한 경우가 가장 좋은 통계량을 가져서 모형에서 현관구조, 세대수를 삭제하였다. 그리고 구와 동은 선형관계로 나타나기 때문에 구와 동 중에서 구를 제외시켰다. 또한 최열, 공윤경 (2003) 의 연구에서 제시된 사용기간의 제곱을 모형에 적합시켜 보았으나, 다중공선성의 문제를 유발하여 모형에는 포함시키지 않았다.

평형	1	33.15479	135.95	<.0001
방수	1	1.23040854	66.71	<.0001
욕실수	1	1.902222	103.33	<.0001
사용기간	1	0.205728	7.63	0.0061
지하철거리	1	0.950052	53.72	<.0001

<표 7> 에서 송파구 지역의 추정치들은 양(+) 의 값을 가진 반면에, 도봉구 지역의 추정치들은 음(-) 의 값을 나타내고 있다. <표 2> 에서 ID 가 1인 관측치를, 추정치를 이용하여 계산하면 40,070 만원으로 나타난다¹⁰⁾. 이것의 잔차는 775만원 (즉, 39,295 만원 - 40,070 만원) 으로 계산된다.

<표 7> 최종 아파트 가격 모형의 모수 추정표

변수	코드값	추정치	표준오차	t 값	Pr > t
절편		8.926768399	0.05639761	158.28	<.0001
동	가락동	0.548696386	0.03740812	14.67	<.0001
	거여동	0.270449971	0.04476348	6.04	<.0001
	마천동	0.214082994	0.04629697	4.62	<.0001
	문정동	0.403227949	0.04027121	10.01	<.0001
	방이동	0.459283174	0.04848518	9.47	<.0001
	송파동	0.593400867	0.04261336	13.93	<.0001
	신천동	0.551602784	0.04063475	13.57	<.0001
	오금동	0.496819039	0.04245754	11.7	<.0001
	잠실동	0.630705881	0.09981138	6.32	<.0001
	풍납동	0.288580838	0.03477648	8.3	<.0001
	도봉동	-0.017398814	0.04979628	-0.35	0.727
	방학동	-0.061841962	0.04972796	-1.24	0.2145
	쌍문동	-0.29007178	0.03437364	-8.44	<.0001
	창동	0	.	.	.
브랜드	현대(1)	-0.021783881	0.03581562	-0.61	0.5434
	대우(2)	0.094367588	0.05213717	1.81	0.0712
	삼성(3)	0.229500326	0.04547231	5.05	<.0001
	5	-0.003225193	0.04945898	-0.07	0.948
	8	-0.380620466	0.09385563	-4.06	<.0001
	12	-0.312129793	0.08939758	-3.49	0.0005
	14	0.110197024	0.05872571	1.88	0.0614
	15	-0.046403857	0.0823797	-0.56	0.5736
	19	0.08112839	0.04692929	1.73	0.0848
	20	-0.116522676	0.07265361	-1.6	0.1097
	22	-0.354092539	0.0908822	-3.9	0.0001
	23	-0.231803571	0.11048738	-2.1	0.0366
	26	0.067267486	0.04897649	1.37	0.1705
	27	-0.097717332	0.07153927	-1.37	0.1729
	28	-0.036723849	0.05135064	-0.72	0.475
	29	0.316926311	0.07926277	4	<.0001

10) = 8.926768399 (절편) + 0.548696386 (가락동) -0.312129793 (브랜드 12) + 0.198540709 (지역 난방) + 34 평* 0.0142463 + 3개 * 0.140444488 (방수) + 2개 * 0.192683042 (욕실수) + 6개월 * -0.000447508 (사용기간) + 3 분 * -0.01728737 = 10.5984. 이것의 e 누승 (즉, $e^{10.5984}$) 을 구하면 40,070 만원이 계산된다.

	30	-0.291453507	0.09041569	-3.22	0.0014
	32	-0.013422411	0.09445618	-0.14	0.8871
	33	-0.109987989	0.08129766	-1.35	0.177
	35	-0.064208199	0.05856793	-1.1	0.2737
	40	-0.226930715	0.08482464	-2.68	0.0078
	41	0.018104159	0.11094522	0.16	0.8705
	46	-0.078475424	0.08692665	-0.9	0.3673
	47	-0.074002385	0.09512338	-0.78	0.4371
	48	-0.014248851	0.06261258	-0.23	0.8201
	49	0.135297746	0.07224872	1.87	0.062
	50	-0.0029255	0.10804777	-0.03	0.9784
	55	0.030545763	0.11069837	0.28	0.7828
	63	-0.114805785	0.0595244	-1.93	0.0546
	68	0.040606045	0.06398874	0.63	0.5261
	72	-0.000094117	0.09670156	0	0.9992
	82	-0.422361826	0.08186459	-5.16	<.0001
	85	0.099251297	0.08605901	1.15	0.2496
	997	0.198687735	0.04973202	4	<.0001
	998	0	.	.	.
난방	지역	0.198540709	0.02857099	6.95	<.0001
	개별	0.050856737	0.02501643	2.03	0.0428
	중앙	0	.	.	.
평형		0.0142463	0.00122182	11.66	<.0001
방수		0.140444488	0.0171947	8.17	<.0001
욕실수		0.192683042	0.01895532	10.17	<.0001
사용기간		-0.000447508	0.00016202	-2.76	0.0061
지하철거리		-0.01728737	0.0023587	-7.33	<.0001

V. 신경망모형을 이용한 아파트 가격 추정

1. 신경망모형의 고찰

1) 신경망모형의 개요

신경망모형은 인간 두뇌의 신경세포를 모형화하여 지능을 구현하고자 하는 기법이다. 신경망은 병행적으로 상호작용하는 여러 개의 계산 요소들 (입력층, 은닉층, 출력층의 노드 (Node)) 로 이루어져 있으며, 각 계산요소는 가중치 합 (Weighted Sum) 과 같은 단순한 계산만을 수행한다 (Harvey, 1994). 그리고 대부분의 신경망 모형은 대응되는 통계모형을 가지고 있다 (강현철 외 5명, 1999).

2) 신경망모형의 구조

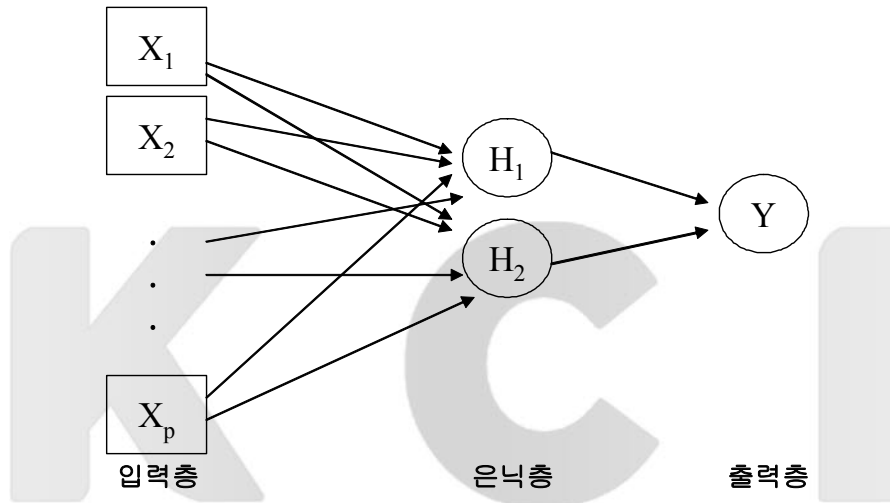
본 연구에서 사용된 신경망의 구조, 알고리즘, 학습인자 등에 대해서 간단히 살펴본다.

(1) 신경망의 구조: 다층 퍼셉트론 (Multi-layer Perceptron) (강현철 외 5명, 1999

참조)

다층 퍼셉트론은 입력층과 출력층 사이에 하나 이상의 중간층 (은닉층) 이 존재하는 신경망으로, <그림 1> 에 나타난 것과 같은 계층 구조를 갖는다. <그림 1> 은 입력층과 출력층 그리고 하나의 은닉층에 두 개의 은닉마디를 가지는 다층 퍼셉트론 구조를 보여주고 있는데, 이를 식 (1.3) 으로 표현하면 다음과 같다

$$\begin{aligned} H_1 &= f_1 (b_1 + w_{11}X_1 + w_{21}X_2 + \cdots + w_{p1}X_p) \\ H_2 &= f_2 (b_2 + w_{12}X_1 + w_{22}X_2 + \cdots + w_{p2}X_p) \\ Y &= g (b_0 + w_{10}H_1 + w_{20}H_2) \end{aligned} \quad (1.3)$$



<그림 1> 다층 퍼셉트론 구조

위의 수식에서 f_1 , f_2 와 g 를 각각 활성화함수와 출력활성함수라고 한다. 활성화함수는 로지스틱 회귀모형에서 연결함수 (확률과 선형함수를 연결하는 함수) 의 역함수와 유사한 의미를 가지며 가장 보편적으로 로지스틱 함수가 사용된다.

다층 퍼셉트론은 종종 범용근사자 (Universal Approximator) 라고 불리기도 하는데, 이는 이론적으로는 적절한 활성화함수와 적당한 수의 은닉마디를 가지는 다층 퍼셉트론은 모든 비선형 곡선을 매우 정확하게 근사시킬 수 있다는 것을 의미한다. 예를 들어, 두 변수 X 와 Y 의 관계가 비선형성을 가지는 경우, 신경망모형은 활성화함수와 은닉마디의 수만을 결정하면 모형을 쉽게 설정할 수 있다. 예를 들어, 로지스틱 함수와 은닉마디의 수를 3개로 지정하면 다음의 식 (1.4) 를 유도할 수 있다.

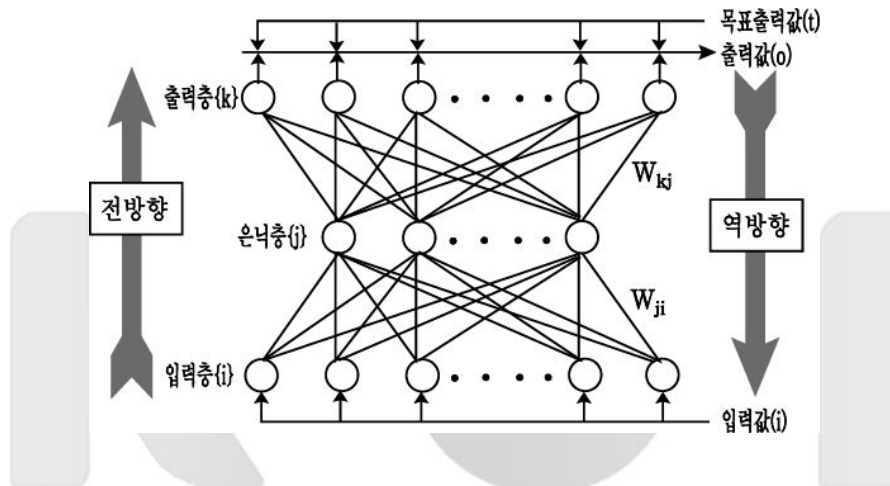
$$H_i = \frac{1}{1 + e^{-(b_i + w_i X)}}, i = 1, 2, 3, Y = b_0 + w_{10}H_1 + w_{20}H_2 + w_{30}H_3 \quad (1.4)$$

비선형 회귀모형으로 적합하는 경우에는 다음과 같은 식 (1.5) 를 미리 파악하고 있어야 한다.

$$Y = a + b_1X + b_2X^2 + \dots + b_pX^p \quad \text{또는} \quad Y = ae^{bx} \quad (1.5)$$

즉, 여러 개의 입력변수를 가지고 있는 경우에는 이러한 함수 형태를 미리 파악한다는 것 자체가 대부분 거의 불가능하다. 따라서 신경망모형은 많은 입력변수를 가지고 있으며 입력변수와 출력변수간의 관계가 복잡한 비선형 형태를 가질 때 특히 유용하다.

(2) 역전파 (Back Propagation) 알고리즘



<그림 2> 역전파 알고리즘 11)

역전파 알고리즘은 신경망의 계수 추정을 위해 개발된 기울기 하강 (Gradient Descent) 알고리즘 중의 하나이다 (강현철 외 5명, 1999). <그림 2> 에서와 같이, 1 단계는 입력층의 각 노드에 입력패턴을 주면, 이 신호는 각 노드에서 변화되어 은닉층에 전달되고 마지막으로 출력층에서 신호를 출력하게 된다. 2단계는 출력값(o)과 목표출력값(t)을 비교하여 식 (1.3) 에 의해 에러값 ($e = o - t$) 을 구한다. 3단계는 이 오차를 줄여나가는 방향으로 연결강도를 조절하고, 상위층에서 역전파하여 하위층에서는 이를 근거로 다시 자기층의 연결강도를 조정해 나간다. 연결강도를 조절하는 데에는 활성화함수가 사용된다 (김대수, 1993).

11) 김대수, 1993, p92.

$$error_p = \sqrt{\sum_j (O_{pj} - T_{pj})^2}, \quad ERROR = \sum_p error_p \quad (1.6)$$

$error_p$: 입력선택 p 에 대한 신경망의 에러
 O_{pj} : 입력패턴 p 에 대한 출력층 j 번째 신경세포의 실제 출력
 T_{pj} : 입력패턴 p 의 j 번째 성분
 $ERROR$: 모든 패턴에 대한 신경망의 오차

(3) 학습 인자

역전파 알고리즘의 학습은 일반적으로 상당히 긴 학습시간이 요구되며, 빠르고 올바르게 학습시키기 위해서는 초기 연결강도, 학습률 (Learning Rate), 은닉층의 노드수, 모멘텀 (Momentum) 등 여러 요인을 고려해야 한다: 일반적으로 초기 연결강도는 $-0.5 \sim 0.5$ 범위의 값으로 설정되며, 학습률은 일반적으로 $0.001 \sim 10$ 사이의 값을 사용한다. 모멘텀은 신경망의 연결 가중치 조절식에 관성을 줌으로써, 학습시간을 단축시키고 학습 성능을 향상시키기 위해 고안된 것이다 (박우열, 차정환, 강경인, 2002).

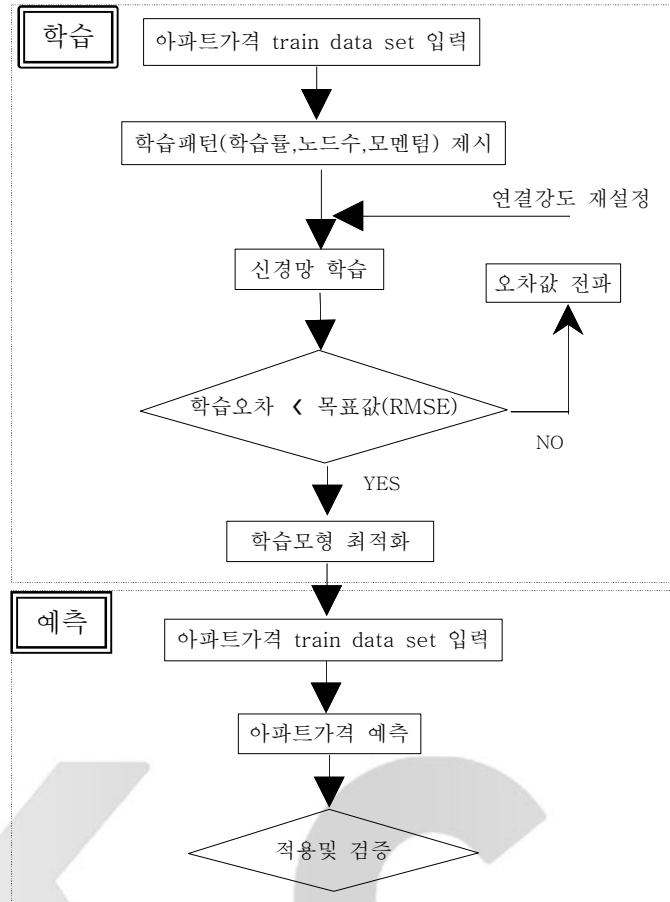
2. 신경망모형을 이용한 분석

1) 신경망모형의 구성

본 연구에서는 QwikNet¹²⁾ 신경망 프로그램을 사용하여 구축하였다. 신경망모형을 이용한 아파트 가격 예측 과정은 회귀분석에서와 동일한 입출력 데이터를 사용하였다¹³⁾. 동, 브랜드, 평형, 방수, 욕실수, 세대수, 입주시기, 지하철거리 등 8개의 입력변수와 출력변수인 추정 매매가격간의 관계를 신경망의 구조로 연결하였다. 신경망 모형의 대략적인 흐름은 <그림 3> 과 같다.

12) <http://qwiknet.home.comcast.net/>

13) 최종적으로 이용된 변수는 축소모형을 통해 결정되었다. 회귀모형과 동일하게 기타 브랜드인 경우를 제외시키고 적합시켰다. 그리고 입력변수도 회귀모형과 같이 신경망모형을 가장 잘 나타내는 변수선택 과정을 거쳐서 최종적으로 선택하였다. 일반적으로 회귀모형으로 변수선택 등을 수행한 후, 그 모형을 신경망모형에 적합시키는 것이 에러값과 정확도 측면에서 좋다.



<그림 3> 신경망의 학습과 예측 흐름도

신경망모형을 구축하기 위해서는 먼저 학습을 통한 모형의 최적화 작업이 필요하다. 최적화 작업은 미리 설정된 오차한계 이내로 학습오차를 최소화하는 것을 의미하며, 가능하면 적은 학습반복횟수에 목표 모형의 특성에 적응하도록 하는 것이다. 이러한 신경망 모형의 최적화 작업을 위한 요소로는 입력변수, 출력변수, 학습반복 횟수, 학습률, 모멘텀, 은닉층의 노드수를 설정해야만 한다 (박우열, 차정환, 강경인, 2002).

본 연구에서 적용한 실험 방법은 Ten-fold Cross Validation 방법으로써, 전체 데이터 Set 을 Set1 ~ Set10 의 10개 Set 로 나누어서, 총 10번의 학습이 이루어지도록 하였다. 각 Set 는 총 424 개의 데이터 중 90% (384개) 의 Train 데이터와 10% (40개) 의 Test 데이터로 구성되었다.

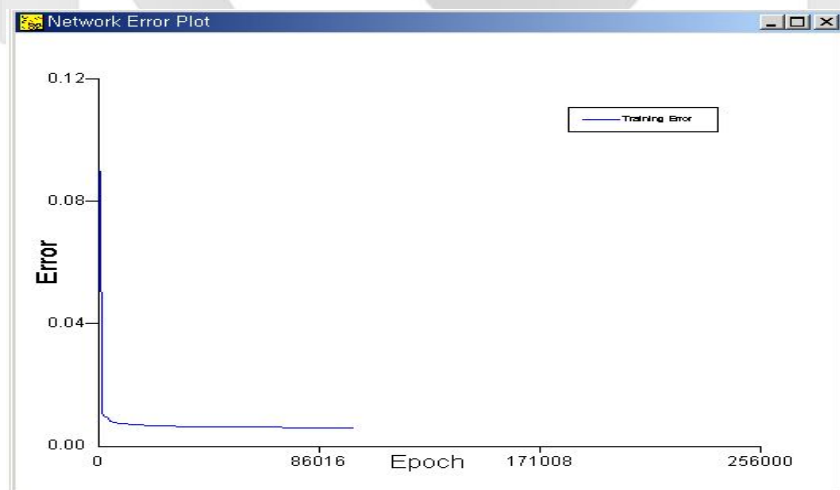
1. Train (S1, S2, S3, S4, S5, S6, S7, S8, S9), Test (S10)
2. Train (S1, S2, S3, S4, S5, S6, S7, S8, S10), Test (S9)
3. Train (S1, S2, S3, S4, S5, S6, S7, S9, S10), Test (S8)
4. Train (S1, S2, S3, S4, S5, S6, S8, S9, S10), Test (S7)

5. Train (S1, S2, S3, S4, S5, S7, S8, S9, S10), Test (S6)
6. Train (S1, S2, S3, S4, S6, S7, S8, S9, S10), Test (S5)
7. Train (S1, S2, S3, S5, S6, S7, S8, S9, S10), Test (S4)
8. Train (S1, S2, S4, S5, S6, S7, S8, S9, S10), Test (S3)
9. Train (S1, S3, S4, S5, S6, S7, S8, S9, S10), Test (S2)
10. Train (S2, S3, S4, S5, S6, S7, S8, S9, S10), Test (S1)

학습률, 모멘텀, 은닉층의 노드수를 결정하기 위해서 다음과 같은 실험을 수행하였다. 학습률을 0.1, 0.3, 0.5 으로, 모멘텀을 0.7, 0.8, 0.9 로, 은닉층의 노드수를 5, 10, 15 가지로 나누어 총 27가지 경우를 학습시킨 후, 모델을 평가하여 최적 모델을 선택하였다. 학습반복횟수는 100,000 번으로 설정하였으며, 에러값 (RMSE; 즉, 식 (1.3) 의 ERROR) 이 0.001 이하가 될 경우는 학습을 중단시키도록 설정하였다.

2) 신경망모형의 적용 및 검증

입력한 자료를 학습률, 모멘텀, 은닉층의 노드수에 의해 학습시킨 결과는 다음과 같다: 학습률 0.5, 모멘텀 0.8, 은닉층의 노드수 10 일 때 에러값이 0.0059 로 가장 낮아, 최적의 학습 모형으로 선정되었다. 그리고 예측의 정확도는 92.7% 정도로 높은 편이다. 신경망 학습에서 범주형 변수의 경우, 변수값의 개수만큼 변수의 개수로 확장하여 학습한다.



<그림 4> 학습률 0.5, 모멘텀 0.8, 은닉층의 노드수 10 일 때 학습에러값의 변화

<그림 4> 는 학습률 0.5, 모멘텀 0.8, 은닉층의 노드수 10 일 때, 학습에러값의 변화를 나타내는 것이다. 신경망모형이 아파트 가격 예측에 효과적으로 적합되고 있음을 보여주고 있다. 즉, 에러값이 신경망 학습초기에 급격한 변동을 보이는 것은

입력층, 은닉층, 출력층의 연결가중치를 처음에 설정할 때 난수를 발생하여 할당하기 때문이다. 그리고 어느 정도 안정된 근사 해에 도달하면 에러값이 크게 변동되지 않는다.

<표 8> 은 10개의 Data Set 에 대하여 선정된 학습모형의 에러값과 정확도이다. 평균 에러값은 0.0063 이고, 예측의 정확도는 92.7%이다. <표 2> 에서 ID 가 1인 관측치를, 신경망모형으로 추정하면 39,937 만원으로 나타난다¹⁴⁾. 이것의 잔차는 632만원 (즉, 39,295 만원 - 39,937 만원) 으로 계산된다.

<표 8> 10개의 Data Set 에 대한 에러값과 정확도

구분	set 1	set 2	set 3	set 4	set 5	set 6	set 7	set 8	set 9	set 10	평균
에러값 (RMSE)	0.0059	0.0061	0.0068	0.0057	0.0067	0.0059	0.0071	0.0058	0.0064	0.0065	0.0063
정확도 (# Correct)	92.7	93.1	91.9	91.7	93.8	92.1	93.1	94.2	91.9	92.1	92.7

VI. 회귀모형과 신경망모형의 비교 분석

두 모형의 정확도를 비교하기 위하여 RMSE 를 사용하였다 (김용재, 이세현, 2002; 손민우, 이길성, 2003; Worzala et al,1995). 본 연구에서 회귀모형의 RMSE 는 0.132990 이며, 신경망모형은 0.0063 으로, 전체적인 면에서 회귀모형보다 더 모형을 잘 적합시키고 있다고 볼 수 있다. 그리고 회귀모형의 설명력은 95%이고 모형에 사용된 변수들 (즉, 동, 브랜드, 난방, 평형, 방수, 욕실수, 사용기간, 지하철거리) 이 통계적으로 유의하였다. 신경망 모형의 예측 정확도는 92.7% 이다.

그리고 본 연구의 이차적인 목적으로써, 특정 모형의 우수성을 소개하는 것도 중요하지만, 상호보완적인 측면을 찾아서 이용하는 것 또한 중요하다.

다음의 신규 아파트를 송파구에 신규로 분양한다고 했을 경우, 추정 매매가격, 매매상한가, 일반거래가, 매매하한가를 예측하여 보자¹⁵⁾:

▶ 풍납동, 브랜드코드 (=3), 33평형, 방수 3개, 욕실수 2개, 200세대, 복도식 현관, 개별 난방, 지하철거리 5분

14) ID=1 인 것의 가격은 39,295 이다. 이것을 신경망모형에서는 0과 1사이로 척도화시킨다 (즉, (매매가격-최소 매매가격)/(매매가격의 범위) 이다). 이것을 변환시키면 0.2095 이다 (<표 5> 참조). 신경망모형에 적합시켜서 나온 값은 0.2136이다. 이것을 계산하면 39,937 (즉, $0.2136 = (x-6172.5)/(164250-6172.5)$, $x=39,937$) 이 나온다.

15) 신규 아파트이기 때문에 사용기간은 없다. 그리고 풍납동에는 브랜드코드 (=3) 인 아파트가 존재하지 않는다.

<표 9> 회귀모형과 신경망모형의 예측 비교 (단위: 만원)

구분	회귀모형				신경망모형			
	매매상한가	일반거래가	매매하한가	추정 매매가격	매매상한가	일반거래가	매매하한가	추정 매매가격
풍납동	45,763 (54세대)	43,726 (108세대)	40,824 (38세대)	43,743	45,719 (54세대)	43,682 (108세대)	40,780 (38세대)	43,499

회귀모형으로 예측할 경우에는 43,743 만원으로 추정되었고, 신경망모형에서는 43,499 만원으로 추정되었다. 그리고 매매상한가, 일반거래가, 매매하한가의 추정가격¹⁶⁾은 <표 9> 에 제시되어있다. 두 모형의 차이는 44만원으로 거의 비슷하게 추정되었다.

예로 제시된 두 가지 모형의 상호보완적인 측면을 고려하면 다음과 같다: 첫 번째, 회귀모형과 신경망모형을 통하여 추정된 시장가격을 참조하여 적정 가격대를 구성하는데 지원 (Decision Support) 할 수 있을 것이다. 예컨대 두 모형의 추정 매매가격을 결합하여 사용할 수 있는 데, 두 모형으로부터 산출된 가격의 평균을 사용한다든지 하는 것이다. 두 번째, 신경망모형에 근거하여 매매가격을 추정하였을 경우, 그 추정 과정을 회귀 모형을 이용하여 설명할 수 있을 것이다. 즉, 신경망모형은 해석이 어렵기 때문에 본 예에서와 같이 유사한 결과를 가지는 회귀모형을 이용하여 설명할 수 있을 것이다. 세 번째, 많은 변수를 신경망모형에 도입하는 경우 회귀 모형의 변수 증감법 (Stepwise Method) 과 같은 변수 선택 (Selection of Variables) 방법을 통해서 축약된 신경망모형을 도출할 수 있다. 네 번째, 본 연구에서는 고려하지 않았지만, 비선형 문제가 발생할 경우이다. 회귀모형을 비선형성 발견에 이용하고, 신경망모형은 모형을 수립하는 데 이용하게 된다

신경망모형이 새로운 모형이라고 해서 항상 좋기만 하지는 않으므로 과용은 금물이며 상호보완적인 측면을 고려해야 할 것이다. 때때로 신경망모형이 전통적인 통계적 방법에 비해 다소 못한 것으로 평가되는 경우가 있는데, 어떤 모형이 나은지는 실제 상황에 적용해 봐야 알 수 있을 뿐이다 (허명희, 이용구, 2003).

VII. 결론

본 연구는 아파트 가격을 예측하기 위하여 회귀모형과 신경망모형을 사용하였고,

16) 각 동별로 매매상한가, 매매하한가, 일반거래가와 추정 매매가격의 차이 평균을 계산하여 이를 모형 추정값에 더해서 구하였다. 예를 들어, 회귀모형의 경우 도봉동의 일반거래가와 추정매매가격의 차이 평균은 -3이다. 그래서 도봉동 회귀모형 추정값 32,212 에 -3을 하여 일반거래가 32,209를 계산하였다.

두 모형을 비교·분석하였다. 이러한 두 모형이 상호 배타적인 관계가 아니라 상호보완적인 측면에서 의사결정지원에 활용된다면, 효과적으로 아파트 가격을 정밀하게 예측할 수 있을 것이다. 그리고 두 모형에서 나오는 결과를 비교하여 적정 가격대를 산출할 수 있도록 도와줄 것이다.

본 연구의 주요한 기여 (Contribution) 는 기존의 분양가 산정 모형 (안병욱, 2001) 에 신경망 모형을 도입한 것이다. 또한 주택에 관한 기존 연구와 신규 연구에 신경망모형이 활용될 수 있으리라 판단된다. 그리고 또 다른 기여점은 두 모형의 상호보완성을 분석하였다는 것이다.

본 연구는 다음과 같은 한계점을 가진다: 첫째, 주택 및 부동산학에 대한 근본적 이해가 불충분하다. 둘째, 수집 대상 데이터가 두 개의 구에만 한정되어 있다. 셋째, 이론적으로 의미 있는 변수들 (예를 들어, 교육적 요인, 환경적 요인 등) 을 추가하지 못하였다. 넷째, 회귀모형의 경우 상호작용을 고려하지 못하였다.

향후 연구 과제는 다음과 같다: 첫째, 모형의 수정과 확대가 필요하다. 추가적인 연구를 통해서 모형을 수정하고 새로운 모형을 개발해야 할 것이다. 둘째, 주택에 관련된 시계열 자료를 이용하여, 가격 예측 모형으로 신경망모형을 개발하는 것이다. 현재 시계열 자료에 대한 주택 가격 예측 모형으로 ARIMA 모형, VAR 모형 등이 사용되고 있다 (국토연구원, 2004). 이 모형들 뿐만 아니라 다중회귀신경망을 이용하여 시계열 패턴을 예측할 수 있을 것이다 (임슬기, 김광희, 이태호, 1994).

참고문헌

1. 강현철·한상태·최종후·김차용·김은석·김미경. 1999. SAS Enterprise Miner를 이용한 데이터마이닝: 방법론 및 활용 제2판. 자유아카데미.
2. 구본창·송현영. 2001. “아파트 특성에 따른 가격결정모형 연구: 분당신도시를 대상으로”. 주택산업연구원 주산연 2001 보고서.
3. 국토연구원. 2004. 공간분석기법. 한울아카데미.
4. 김대수. 1993. 신경망 이론과 응용 (I) 제2판. 하이테크정보.
5. 김용재·이세현. 2002. “탄산 가스 아크 용접에서 회귀 분석과 인공 신경망을 이용한 아크 센서 모델 개발”, 대한용접학회지 20(6): pp52-58.
6. 박우열·차정환·강경인. 2002. “신경망이론을 이용한 공동주택 초기사업비 예측에 관한 연구”. 대한건축학회논문집 구조계 18(7): pp155-162.
7. 박진우·이기재·김재광·김진익. 2004. “아파트 가격조사를 위한 측정방법”. 조사연구 5(1): pp79-91.

8. 손민우·이길성. 2003. “신경망 이론과 회귀분석을 통한 홍수위 예측”. 대한토목학회 논문집 B 23(3): pp147-155.
9. 안병욱. 2001. “특성가격함수법을 이용한 아파트 분양가 책정방법 연구”. 대한건축학회 학술발표논문집 21(1): pp103-106.
10. 이문규·허해숙. 1995. “선형함수 fitting 을 위한 선형회귀분석, 역전파신경망 및 선형 Hebbian 신경망의 성능비교”. 한국경영과학회지 20(3): pp17-29.
11. 이변송·정의철·김용현. 2002. “아파트 단지특성이 아파트 가격에 미치는 영향 분석”. 국제경제연구 8(2): pp21-25.
12. 임슬기·김광희·이태호. 1994. “다중회귀신경망을 이용한 시계열패턴의 예측에 관한 연구”. 대한전자공학회 학술발표회 논문집 3(2): pp54-59.
13. 정성훈·강준모. 2002. “아파트 전세 가격 결정요인 연구: 수원시를 사례로”. 한국지역개발학회지 14(2): pp57-76.
14. 최병남. 1994. “인공신경망에 의한 토지평가시스템”. 국토연구 22: pp37-50.
15. 최열·공윤경. 2003. “재건축 가능성에 따른 공동주택의 가격변화에 관한 연구”. 대한건축학회논문집 계획계 19(8): pp123-130.
16. 허명희·이용구. 2003. 데이터마이닝 모델링과 사례. SPSS 아카데미.
17. 허명희·서혜선. 2001. SAS 회귀분석 제3판. 자유아카데미.
18. Bin, O. 2004. "A Prediction Comparison of Housing Sales Prices by Parametric versus Semi-parametric Regressions". *Journal of Housing Economics* 13: pp 68-84.
19. Buja, A. and Lee, Y. S. 2001. "Data Mining Criteria for Tree-based Regression and Classification". *Proceedings of the 7th ACM SIG KDD International Conference on Knowledge Discovery and Data Mining*: pp27-36.
20. Do, A. Q., and Grudniski, G. 1993. "A Neural Network Analysis of the Effect of Age on Housing Values". *Journal of Real Estate Research* 8(2): pp253-264.
21. Harvey, R. L. 1994. Neural network principles. Prentice-Hall.
22. Nguyen, N. and Cripps, A. 2001. "Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks". *Journal of Real Estate Research* 22(3): pp313-336.
23. O'Leary, DE. 1998. "Using Neural Networks to Predict Corporate Failure". *International Journal of Intelligent Systems in Accounting, Finance and Management* 7: pp187-197.
24. Worzala, E. M., Lenk, M., and Silva, A. 1995. "Exploration of Neural

Networks and Its Application to Real Estate Valuation". *Journal of Real Estate Research* 10(2): pp185-202.

ABSTRACT

A Study on Apartment Price Models Using Regression Model and Neural Network Model

Taehun Kim and Hankuk Hong

※ Keywords: Apartment Price, Hedonic Price Model, Neural Network Model, Regression Model

It is in the present situation that there are many studies that estimate and presume apartment price by using various characteristics of apartment. Most of these studies depend on Regression model excessively, but Regression model has more merit than demerit. Thus this study does not deny Regression model but attempts to introduce new model. That is to say, this study is performed by means of the necessity to overcome the problems of Regression model and to introduce complementary model.

The main purpose of this study is to presume apartment price by using neural network model of various characteristics of apartment and to compare it with regression model. And examining the complementary aspect of regression model and neural network model is the secondary purpose of this study.

In the characteristics of apartment, data easily available in the surroundings are put first in importance. We collected apartment sale price of Songpa-gu and Dobong-gu in Seoul and the characteristics of 12 apartments on the basis of 6/2004. We unified apartment sale prices (that is, sale lowest limit price, general trade price, sale upper limit price) into one sale price by using new measurement method.

It is significant in the study field of apartment price to estimate apartment price precisely and effectively by introducing neural network model and to compare it with the existing regression model. It is judged that neural network model can be applied to the past studies and new studies concerning housing.