

# 회귀모형과 신경망모형을 이용한 아파트 가격 모형에 관한 연구

홍한국\* · 서보라\*\* · 김태훈\*\*\*

## 요약문

다양한 아파트 특성들을 이용하여 아파트 가격을 추정하고 예측하는 연구 또한 많이 존재하고 있는 실정이다. 그렇지만 이러한 연구들 대부분이 회귀모형에 지나치게 의존하고 있는 실정이다. 그러나 회귀모형은 단점보다 장점이 많은 모형이다. 본 연구는 회귀모형을 부정하기보다는 새로운 모형을 도입하여, 회귀모형의 문제점들을 극복하고 회귀모형과 상호보완적인 모형을 도입할 필요성에 의해서 본 연구를 수행한 것이다.

다양한 아파트 특성들에 대하여 신경망모형을 이용하여 아파트 가격을 예측하고, 기존의 회귀모형과 비교하는 것이 본 연구의 주목적이다. 또한 회귀모형과 신경망모형의 상호 보완적인 측면을 규명하는 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)을 새로운 측정방법을 이용하여 하나의 매매가격으로 추정하였으며, 대표성을 가지도록 하였다.

신경망모형을 도입하여 아파트 특성들을 이용하여 아파트 가격을 정밀하고 유효하게 예측하고, 기존의 회귀모형들과 비교하는 것은 아파트 가격에 대한 연구 분야에 큰 의미가 있다 하겠다. 그리고 주택에 관한 기존의 연구와 신규 연구에 신경망모형이 활용될 수 있으리라 판단된다.

Keywords: 신경망 모형, 회귀모형

## I. 서론

우리나라 주택에서 상당한 부분을 차지하고 있고 주택 건설의 대부분을 차지하고 있는 아파트에 대한 연구들은 과거에서 현재까지 계속 이어오고 있다. 특히, 다양한 아파트 특성들을 이용하여 아파트 가격을 추정하고 예측하는 연구 또한 많이 존재하고 있는 실정이다. 그렇지만 이러한 연구들 대부분이 회귀모형에 지나치게 의존하고 있는 실정이다. 회귀모형은 등분산성, 독립성, 정규성 등의 가정과 다중공선성의 문제, 특이값에 민감하게 반응하는 문제 등이 제기되고 있으며, 또한 선행 연구들 대부분이 모수의 비선형에 대한 연구는 전무한 실정이다. 그리고 회귀모형을 적용함에 있어서 최적의 회귀모형을 선택하였는가에 의문점이 있을 수도 있고, 다양한 아파트 특성들에 대한 정성적 평가에 대한 문제 제기도 있을 수 있으리라 판단된다. 그러나 회귀모형은 단점보다 장점이 많은 모형이다. 본 연구는 회귀모형을 부정하기보다는 새로운 모형을 도입하여, 회귀모형의 문제점들을 극복하고 회귀모형과 상호보완적인 모형을 도입할 필요성에 의해서 본 연구를 수행한 것이다. 현재까지 인공지능 분야에서 널리 이용되어 왔던 신경망모형(Neural Network Model)은 입력변수가 불완전하고 변동폭이 넓은 경우에도 해석이 가능하며, 데이터 수가 적거나 불규칙할 경우라도 사례를 통한 반복학습을 통해 그 오차를 줄여가기 때문에 데이터 수에 민감한 영향을 받는 회귀모형보다는 정밀한 산정이 가능하다(박우열, 차정환, 강경인, 2002). 이러한 신경망모형을 도입하여 아파트 특성들을 이용하여 아파트 가격을 정밀하고 유효하게 예측하고, 기존의 회귀모형들과 비교하는 것은 아파트 가격에 대한 연구 분야에 큰 의미

가 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 상호 보완적인 측면을 규명하는 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 매매가격들을 새로운 측정방법을 이용하여 하나의 매매가격으로 추정하였으며, 대표성을 가지도록 하였다.

## II. 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 출처는 국민은행 아파트 시세 홈페이지와 닥터아파트 아파트 시세 홈페이지이며, 두 홈페이지에서 수집가능한 모든 변수들을 정리하였다. 즉, 구, 동, 아파트명, 브랜드명, 평형, 매매하한, 일반거래가, 매매상한, 방수, 욕실수, 세대수, 현관구조, 입주시기, 난방, 지하철 거리(분) 등이 수집되었다. 이러한 변수들 중에서 구, 동, 현관구조, 난방을 범주화시켜 코딩하였다. 그리고 브랜드명은 시공회사를 의미하는 것으로써, 대한건설협회에서 제공한 2003년도 시급(도급) 순위 자료에 의거하여 순위 번호로 범주화시켰다. 그리고 대기업의 경우, 다른 계열사로 시공된 아파트일 경우 같은 브랜드로 나타내었다. 예를 들어, 삼성물산, 삼성중공업 등은 같은 브랜드로 표시하였다. 입주 시기는 현재 시점까지의 사용기간(월 단위)으로 환산하였다.

\* 동의대학교 경영정보학과 부교수, \*\* 동의대학교 경영정보학과 석사, \*\*\* 경성대학교 e비즈니스학과 조교수

[표 1] 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

[표 1] 은 수집된 자료를 분석에 맞게 변환하여 최종적으로 구한 데이터이다. 총 755개의 관측치 (Observation) 를 가진다. [표 1] 에서 음영으로 나타난 것이 분석 대상 변수들이며 [표 2] 에서 자

세히 설명하였다. 즉, [표 2] 는 아파트 가격을 예측하기 위하여 사용된 아파트 특성들을 정리한 것이다.

[표 2] 변수 구성

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

[표 1] 의 추정 매매가격은 기존의 매매상한과 매매하한의 평균으로 계산하는 것의 단점을 보완하여, 새로운 측정방법으로 추정된 것이다. 일반적으로 아파트 평균매매가격을 추정할 경우 상한가와 하한가의 평균가격을 적용하는 데, 이 때 전

체 세대 중 상한가와 하한가 세대가 반반이라는 것을 전제로 한다. 그러나 아파트 가격의 범위 (Range) 를 아는 것이 목적이 아니라, 아파트가치 추정 시 문제가 발생한다(박진우 외 3명, 2004).

[표 3] 서울시 구별 아파트 가격대별 세대수의 비율

지역	하한가 세대비율	일반거래가 세대비율	상한가 세대비율
서울시 전체	0.2	0.54	0.26
송파구	0.19	0.54	0.27
도봉구	0.18	0.61	0.21

[표 3] 은 서울시 전체와 서울시 각 구별로 동일 단지 내 동일 평형 아파트 세대들을 상한가, 하한가, 일반거래가로 나누다고 했을 때 각 범주에 해당되는 구성비를 나타낸 것이다 (박진우 외 3명, 2004). 이러한 구성비를 이용하여 송파구와 도봉구 아파트들의 매매가격을 추정하면 다음의 식 (1.1) 과 같다:

$$\begin{aligned} \text{송파구 A 아파트 추정 매매가격} & \quad (1.1) \\ &= \frac{A_1 * M * 0.19 + A_2 * M * 0.54 + A_3 * M * 0.27}{M} \end{aligned}$$

$$\begin{aligned} \text{도봉구 B 아파트 추정 매매가격} \\ &= \frac{B_1 * N * 0.18 + B_2 * N * 0.61 + B_3 * N * 0.21}{N} \end{aligned}$$

A<sub>1</sub>, B<sub>1</sub>: 하한가 A<sub>2</sub>, B<sub>2</sub>: 일반거래가  
A<sub>3</sub>, B<sub>3</sub>: 상한가 M, N: 각 아파트 평형별 세대수

[표 1] 에서 ID 가 1인 관측치의 추정 매매가격을 계산하면 다음과 같다:

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

기존의 상한가와 하한가의 평균가격 추정은 새로운 추정 매매가격보다 과소추정되는 (Underestimated) 것으로 조사되어 있으며 (박진우 외 3명, 2004), 본 연구에서는 식 (1.1) 에서 이용하고 계산된 [표 1] 의 추정 매매가격을 사용하였다.

## 2. 연구 모형과 절차

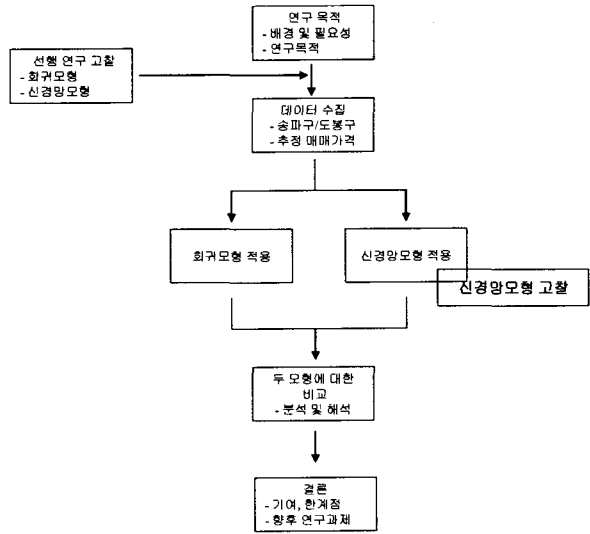
본 연구는 아파트 가격을 예측하기 위하여 신경망모형과 회귀모형을 이용하였다. 이러한 연구 모형과 그 절차를 <그림 1> 과 같이 요약할 수 있다: 연구목적에 기반하여 회귀모형과 신경망모형관점에서 선행 연구들을 고찰한다. 그리고 분석 대상이 되는 데이터를 수집한다. 수집된 데이터 중에서 매매가격들의 대표값을 추정한다 (즉, 추정 매매가격). 그리고 신경망모형에 대한 이론적 고찰을 수행한 후, 데이터를 두 가지 모형에 적합시킨다. 최종적으로 두 모형에 대한 비교를 수행하고 분석한다. 이를 통하여 결론을 유도하고, 본 연구의 기여와 한계점을 기술한다.

### III. 회귀모형을 이용한 아파트 가격 예측

#### 1. 기초 통계량 분석

분석 대상 데이터에 대한 기초 통계량을 살펴보면 평형의 평균값은 이며, 최소 8평부터 102평까지 분포하고 있다. 추정 매매가격은 평균 3억 7천 만원으로 아파트 가격이 조사되었으며, 최소 6천 백만원부터 16억 4천 2백만원까지 분포되어 있다. 분석 지역으로 보면, 송파구의 453개 아파트를, 도봉구의 302개 아파트를 대상으로 한다. 동코드 6 (즉, 삼전동), 7 (석촌동) 은 도수가 적기 때문에 분석에서 제외될 것임을 예상할 수 있다. 또한 현관구조코드에서 타워형 현관구조의 도수가 다른 구조에 비해서 상당히 적음을 알 수 있다. 마지막으로 브랜드코드에서 기타의 도수가 전체 도수의 44%를 차지함을 알 수 있다. 기타라고 분류된 아파트들은 주택공사, 시영, 영세업체 등이 시공한 것들이다. 특히, 재개발 아파트 등이 포함되어 있기

때문에, 아파트 가격이 상당히 왜곡되어 있을 것이다.



[그림 1] 본 연구 모형과 절차

#### 2. 회귀모형을 이용한 분석

본 연구에 사용된 회귀모형의 초기 완전모형 (Initial Full Model) 은 다음 수식과 같다. 이산형 변수는 더미변수로 변환하였다.

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

앞의 수식을 SAS REG 로 적합시키 후, 모형에 대한 회귀 진단 (Regression Diagnostics) (허명희, 서혜선, 2001) 을 수행한 결과 다중공선성 (Multicollinearity), 특이값 그리고 잔차의 이분산성 등의 문제가 나타났다. 이러한 문제들을 해결하기 위하여 다음의 방법을 사용하였다: 첫째, 관측도수가 적은 경우와 브랜드가 기타인 경우를 삭제하였고, 다중공선성의 문제를 해결하기 위하여 변수 선택 (Model Selection) 방법에 의하여 축소모형을 선택하였다. 둘째, 잔차가 이분산성이지만 비선형성을 보이지는 않기 때문에 일반적으로 사용되는 로그 변환을 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 추정은 SAS GLM 을 통하여 수행하였다. 그 결과는 [표 4] 과 [표 5] 에 제시되어 있다. [표 4] 는 모든 변수들의 유의하다는 결과를 보여주고 있다. 그리고 최종 모형의 설명력 (R<sup>2</sup>) 은 95% 이고 오차항의 표준편차 ( $\hat{\sigma}$ ) 는 0.132990 으로 추정된다. 한편, 각 변수들의 기여율은 등 (51.3%), 평형 (25.5%), 브랜드 (13.3%) 순으로 나타나고 있다.

[표 4] 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	기여율	F 값	Pr > F
동	13	67.00754068	51.3%	87.63	<.0001
브랜드	34	17.97414	13.3%	9.01	<.0001
난방	2	1.610634	1.2%	25.95	<.0001
평형	1	33.15479	25.5%	135.95	<.0001
방수	1	1.23040854	1%	66.71	<.0001
욕실수	1	1.902222	1.4%	103.33	<.0001
사용기간	1	0.205728	0.1%	7.63	0.0061
지하철거리	1	0.950052	0.7%	53.72	<.0001

[표 6] 에서 송파구 지역의 추정치들은 양(+)의 값을 나타낸 반면에, 도봉구 지역의 추정치들은 음(-)의 값을 나타내고 있음을 알 수 있다. 이러한 추정치를 이용하여 [표 2] 의 ID = 1 을 계산

하여 지수변환 (즉,  $e^{\hat{y}}$ ) 을 수행하면, 40,070만원으로 추정된다. 그것의 잔차는 775만원(즉, 39,295 만원 - 40,070 만원) 으로 나타난다.

[표 5] 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
	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	

#### IV. 신경망모형을 이용한 아파트 가격 모형

##### 1. 신경망을 이용한 분석

###### 1.1 신경망 구성

본 연구에서는 아파트 가격 예측을 위한 신경망 모델(역전파 알고리즘)의 학습에 필요한 프로그램은 QwikNet32 신경망 프로그램을 사용하여 구축하였다. 신경망 모델을 이용한 아파트 가격 예측 과정은 회귀분석에서와 동일한 입출력 데이터를 사용한다. 동, 브랜드명, 평형, 방수, 욕실 수, 세대 수, 입주 시기, 지하철 거리(분) 등 8개의 입력값과 출력값인 일반거래가 간의 관계를 신경망의 구조로 연결한다. 학습 후에 기억된 연결강도를 조절하여 학습에 사용되지 않았던 평가데이터를 가지고 아파트 가격을 예측하게 된다.

우선 신경망 모델을 구축하기 위해서는 신경망 모델의 최적화 작업이 필요하다. 신경망 모델의 최적화란 미리 설정한 오차한계 이내로 학습오차를 최소화하는 것을 의미하며, 가능하면 적은 학습반복횟수에 목표 모델의 특성에 적응하도록 하는 것이다. 이러한 신경망 모델의 최적화 작업을 위한 요소로는 입력변수, 출력변수, 학습반복횟수, 학습률, 모멘텀, 은닉층의 노드수를 설정해야만 한다.

본 연구에서는 적용한 실험 방법은 Ten-Fold Cross Validation 으로 전체 데이터 set을 Set1,

Set2, Set3, Set4, Set5, Set6, Set7, Set8, Set9, Set10 등 10개의 set으로 나누어 총 10번의 학습이 이루어지게 된다. 각 set은 총 424개의 데이터 중 90%(384개)의 train 데이터와 10%(40개)의 test 데이터로 구성하였다.

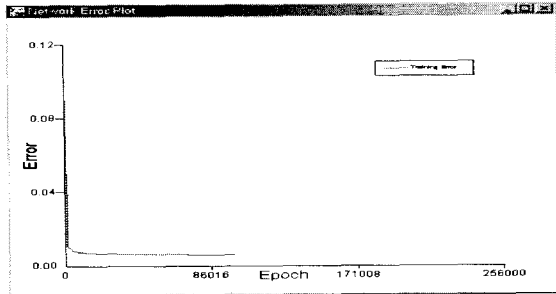
1. Train (S1, S2, S3, S4, S5, S6, S7, S8, S9), Test (S10)
2. Train (S1, S2, S3, S4, S5, S6, S7, S8, S10), Test (S9)
3. Train (S1, S2, S3, S4, S5, S6, S7, S9, S10), Test (S8)
4. Train (S1, S2, S3, S4, S5, S6, S8, S9, S10), Test (S7)
5. Train (S1, S2, S3, S4, S5, S7, S8, S9, S10), Test (S6)
6. Train (S1, 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 노드수의 결정을 위해서 다음과 같이 실험을 하였다. 학습률을 0.1, 0.3, 0.5로, 모멘텀을 0.7, 0.8, 0.9로, 은닉층의 노드수를 5, 10, 15가지로 나누어 총 27가지 경우를 학습시킨 후 모델을 평가하여 최적의 모델을 선택하였다. 학습반복횟수는 기본적으로 100,000번을 설정하였으며, 에러값(RMS Error)이 0.001값 이하가 될 경우는 학습을 중단하는 것으로 설정하였다.

## 2.2 모델의 적용 및 검증

입력한 실측자료를 학습률, 모멘텀 및 은닉층 노드수를 각각 3가지로 나누어 모두 27가지로 모델을 학습시킨 결과, 학습률, 모멘텀, 은닉층의 노드수를 변화시켜 실험한 결과, <그림 3> 의 오른쪽 하단에 있는 Traing info를 보면, 학습률 0.5, 모멘텀 0.8, 은닉노드수 10일 때 에러값(RMS error)이 0.0059로 가장 낮아, 최적의 학습 모델로 선정되었다. 예측의 정확도(% Correct)는 92.7% 정도로 높은 편이다. 신경망 학습에서는 동, 브랜드명 등과 같은 범주형 데이터들은 수치 데이터로 변환이 필요하다. 한 범주형 속성의 경우, 속성값의 수 만큼 속성의 수를 확장하여 학습하였는데, 전체 속성의 수는 60개로 확장되었다. <그림 7>의 상단에 있는 Inputs의 개수 60은 이를 나타낸다.

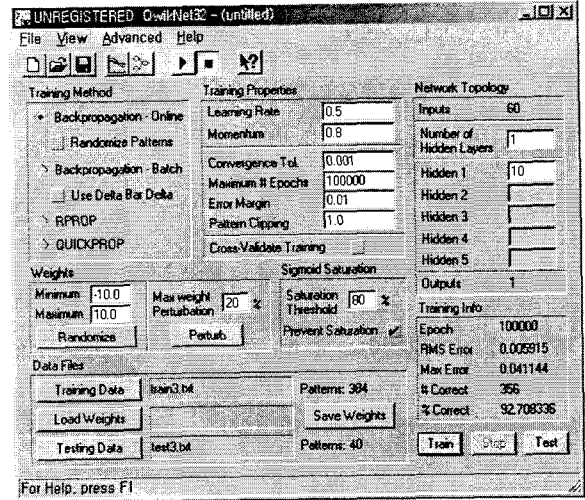
<그림 2>은 학습률 0.5, 모멘텀 0.8, 노드수 10일 때의 학습에러의 변화를 나타내고 있는데, 신경망 모델이 아파트 가격의 예측을 효과적으로 하고 있음을 보여주고 있다.



[그림 2]. 학습률 0.5, 모멘텀 0.8, 노드수 10일 때 에러값의 변화

에러값이 신경망 학습초기에 급격한 변동을 보이는 것은 입력층, 은닉층, 출력층의 연결가중치를 처음에 설정할 때 난수를 발생하여 할당하였기 때문에 어느 정도 안정된 근사 해에 도달하기까지 에러값이 크게 변동되기 때문인 것으로 사료된다.

<표 6>은 10개의 data set에 대한 선정된 학습 모델의 에러값과 정확도이다. 평균 에러값(RMS Error)이 약 0.0063수준이고 예측의 정확도는 92.7%이다.



[그림 3]. 학습모델; 학습률 0.5, 모멘텀 0.8, 노드수 10일 때 에러값(RMS Error;0.0059)

<표 8> 10개의 data set에 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 (RMS Error)	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

## V. 회귀모형과 신경망모형의 비교 분석

<표 7> 에서 ID 가 1인 관측치를 회귀모형은 40.700만원으로, 신경망모형은 39,937만원으로 추정하였다. 회귀모형의 잔차는 775만원, 신경망모형은 632만원으로 143만원의 차이가 나타났다. 본 연구의 신경망모형 예측 정확도가 92.7%이고, 평균 에러값이 0.63%로, 전체적인 면에서 회귀모형보다 더 모형

을 잘 적합시키고 있다고 볼 수 있지만, 더 좋다고 판단할 수는 없다. 그러나 본 연구의 이차적인 목적으로써, 특정 모형의 우수성을 소개하는 것도 중요하지만, 상호보완적인 측면을 찾아서 이용하는 것도 또한 중요하다.

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

<표 7> 회귀모형과 신경망모형의 예측 비교 (단위: 만원)

구분	회귀모형				신경망모형				가격 민감도측정		이익 지수법 가격
	매매상한가	일반거래가	매매하한가	추정 매매가격	매매상한가	일반거래가	매매하한가	추정 매매가격	소비자 인지적정 가격	소비자 인지수용 가격	
풍납동	45,763 (54세대)	43,726 (108세대)	40,824 (38세대)	43,743	45,719 (54세대)	43,682 (108세대)	40,780 (38세대)	43,499	x1	x2	x3

- ▶ 풍납동, 브랜드코드 (=3), 33평형, 방수 3개, 욕실 수 2개, 200세대, 복도식 현관, 개별 난방, 지하 철거리 5분

회귀모형으로 예측할 경우에는 43,743 만원으로 추정되었고, 신경망모형에서는 43,499 만원으로 추정되었다. 그리고 매매상한가, 일반거래가, 매매하한가의 추정가격은 <표 7>에 제시되었다. 두 모형의 차이는 44만원으로 거의 비슷하게 추정되었다. 가격 민감도 측정은 실제 소비자의 행태분석을 통해서 수집된 결과치를 나타내는 것이다. 이것은 소비자 인지적정가격과 소비자 인지수용가격으로 구성되며, 이익지수법 가격은 기업 입장의 공급이익가격을 나타낸다. 이익지수법을 사용하여 소비자 구매 의향율과 공급자의 단위 이익의 점점 분석을 통해 소비자 요구와 공급자 이익이 일치하는 이상점점 가격을 분석할 수 있고, 회귀모형과 신경망모형을 통해 추정된 시장가격을 참조하여 적정 가격대를 구성하는데 지원할 수 있다.

## VI. 결론

본 연구는 아파트 가격을 예측하기 위하여 회귀모형과 신경망모형을 사용하였고, 두 모형을 비교·분석하였다. 이러한 두 모형이 상호 배타적인 관계가 아니라 상호보완적인 측면에서 의사결정지원에 활용된다면, 효과적으로 아파트 가격을 정밀하게 예측할 수 있을 것이다. 그리고 기업의 마케팅의 전략으로 소비자 행태 분석을 수행한다면, 소비자 인지적정가격과 소비자 인지수용가격들을 조사하여, 두 모형에서 나오는 결과 비교하여 적정 가격대를 산출할 수 있도록 도와줄 것이다. 본 연구의 주요한 기여는 기존의 분양가 산정 모형(안병욱, 2001)에 신경망 모형을 도입한 것이다. 또한 주택에 관한 기존 연구와 신규 연구에 신경망모형이 활용될 수 있다. 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. 이 모형들 뿐만 아니라 다중회귀신경망을 이용하여 시계열 패턴을 예측할 수 있을 것이다.

## 참고문헌

1. 강현철·한상태·최종후·김차용·김은석·김미경. 1999. SAS Enterprise Miner를 이용한 데이터마케팅: 방법론 및 활용 제 2판. 자유아카데미.
2. 김대수. 1993. 신경망 이론과 응용 (I) 제2판. 하이테크정보.

3. 김태훈·이희석·김재운·전진오·이은식. 1999. "주택 사업 분석 시스템 구축: 서울지역 아파트 가격 데이터를 중심으로". 정보기술과 데이터베이스 저널 6(2): pp.115-130.
4. 박우열·차정환·강경인. 2002. "신경망이론을 이용한 공동주택 초기사업비 예측에 관한 연구". 대한건축학회논문집 구조계 18(7): pp155-162.
5. 이문규·허해숙. 1995. "선형합수 fitting 을 위한 선형회귀분석, 역전파신경망 및 선형 Hebbian 신경망의 성능비교". 한국경영과학회지 20(3): p p17-29.
6. 허명희·서해선. 2001. SAS 회귀분석 제3판. 자유아카데미.
7. Bin, O. 2004. "A Prediction Comparison of Housing Sales Prices by Parametric versus Semi-parametric Regressions". *Journal of Housing Economics* 13: pp 68-84.
8. Buja, A. and Lee, Y. S. 2001. "Data Mining Criteria for Tree-based Regression and Classification". *Proceedings of the 7<sup>th</sup> ACM SIG KDD International Conference on Knowledge Discovery and Data Mining*: pp27-36.