

## 국내 남자 프로농구 승패 예측 모형 비교 연구

구승환 · 김현수\* · 장성용(서울산업대학교)

이 연구의 목적은 농구경기의 승패에 영향을 미치는 요인을 팀별, 경기별로 분석하는 것이다. 이를 검증하기 위해 로짓회귀분석 및 인공신경망 모형으로 승패예측을 하였다. 승패에 영향을 미치는 요인을 분석하기 위한 경기 자료로 2007-2008년도 시즌 동안 한국농구연맹이 주최한 270경기를 사용하였다. 게임에 영향을 미치는 요인을 분석하기 위해 회귀분석을 실시하였으며, 검증을 위해 게임의 승패 예측에 관한 두 가지 실험을 실시하였다. 첫 번째 실험은 팀별로 승패에 영향을 미치는 요인을 분석하기 위해 기존의 연구처럼 270경기를 승리집단과 패배집단으로 나누어 총 540경기에 대한 자료를 사용하여 집단의 승리/패배를 예측하는 것이었다. 두 번째 실험은 경기별로 승패에 영향을 미치는 요인을 분석하기 위해 승리/패배 집단으로 나누지 않고 경기 내용을 그대로 사용하였다. 실험 결과 인공신경망 모델 및 로짓회귀모델 모두 85% 이상의 예측 정확도를 보였으며, 승패의 두 개 그룹으로 구분하지 않은 경우가 구분한 경우보다 우수한 예측결과를 보였으며, 유의미하게 영향을 미치는 변수만을 사용했을 때가 모든 변수를 사용했을 때보다 승패예측 정확도가 우수하였다. 따라서 이 연구에서 선정된 변수가 농구경기 결과에 유의한 영향을 미치고 있음을 알 수 있었다.

주요어: 로짓회귀분석, 인공신경망, 농구, 승패예측, 승패요인

### 서 론

2007-2008년도 한국 프로 농구 경기는 역대 최다관중 기록을 세우며 성공적인 시즌을 마쳤다. 이와 같이 많은 시민들의 관심이 늘면서 경기의 결과에 대한 예측 및 요인에 관한 연구도 증가하고 있다. 스포츠 경기에서 예측결과가 주는 의미는 여러 가지가 있다. 첫째, 각 팀의 감독 및 코칭스태프에게 팀 경기력에 대한 전반적인 평가를 바탕으로 새로운 작전을 구상할 수 있도록 도움을 준다. 둘째, 선수에게는 경기 기록요인별 평가 자료를 제시함으로써 미래의 경기력을 향상시키는데 필요한 정보를 제공하는 역할을 한다. 셋째, 스포츠 토트 등과 같은 투표권의 이용자에게 실제 자료를 통한 예측 모형을 제공함으로써 보다 신뢰할 수 있고 예측율도 높은 투자정보를 주는 것 등이다.

농구경기의 승패예측에 관한 국내 연구를 살펴보면 박제영(1994)은 농구경기의 승패요인을 알아보기 위하여 제 40회 세계군인 농구선수권대회를 대상으로 경기집계표상

의 14개 변수를 사용해 승리한 팀과 패배한 팀의 총 68게임에 대한 평균값과 표준편차를 구하고, 승패에 미친 비중을 알아보기 위하여 정준판별분석을 실시하였다. 최현희와 임백빈(2005)도 14회 아시안게임의 남자농구경기의 경기분석을 통하여 두 팀 간의 전력 차가 많고 적응의 기준을 최종 점수 10점차로 정하여 최종 점수 차가 10점차 이하로 종료된 경기의 승리 팀/패배 팀, 10점차 초과로 종료된 경기의 승리 팀/패배 팀의 4개 집단으로 나누어 득점차 분류의 승/패에 따른 정준판별분석을 실시하였다. 한필수(2006) 역시 여자 프로농구를 대상으로 국제선수평가그룹에서 새롭게 제시한 변수를 사용하여 정준판별분석을 실시하였다. 한재순 등(2006)은 WKBL (Women's Korean Basketball League) 공식경기기록표에서 사용하는 공격과 수비에 관련된 9 가지 기록요인이 경기승패에 미치는 영향을 WKBL의 2004-2005 겨울리그, 2005 여름리그, 2005-2006 겨울리그의 경기를 대상으로 상관분석 및 회귀분석을 통해 규명하였다. 김세형 등(2008)은 KBL홈페이지에 공개되는 공격변인과 수비변인 각각 7개를 사용하여 '06-'07 리그를 대상으로 로짓 회귀분석과 의사결정나무 분석을 통해 승패결과를 예측했다. 박제영(2008)은 2006-2007시즌의 국내 남자농구 270경기를 대상으로 승

논문 투고일: 2008. 5. 8.

논문 수정일: 2009. 9. 17.

게재 확정일: 2009. 10. 23.

\* 저자 연락처: 김현수(hskim@snu.ac.kr)

리와 패배를 결정하는 요인을 정준판별분석을 통해 분석하였다. 그리고 Stefani(1977, 1980)는 대학농구팀을 대상으로 Least Squares 및 Improved Least Squares를 이용하여 각각 69%, 70%의 예측률을 보였다.

한편 인공신경망을 사용한 스포츠경기 승패예측 국내연구를 살펴보면 최형준과 김주학(2006)은 2005년도 영국 워블던 테니스 대회의 경기결과를 토대로 총 7개의 경기요인을 사용하여 다층 신경망모형을 만들었고, 이를 역전파 알고리즘을 사용하여 분석한 결과 기존의 예측기법들인 로짓 회귀분석 및 정준판별분석에 비해 높은 적중률을 보인다고 밝혔다. 김주학 등(2007)은 2006년 독일 월드컵의 64 경기를 대상으로 승패와 관련한 기록요인을 집수화하여 변수로 사용하고, 이를 인공신경망 모형을 이용하여 예측하였다. 이밖에도 Abbott & Douglas M(2005)은 몬타나 대학교의 풋볼경기를 대상으로 153명의 학생에게 50개의 설문조사를 한 뒤, 설문결과 값을 토대로 6개의 독립변수를 사용하여 인공신경망을 활용해 경기결과를 예측한 결과 75%이상의 정확도를 보였다.

이상과 같이 스포츠 경기의 승패를 예측하는 국내 외 연구들은 대부분 정준판별분석, 로짓 회귀모형 또는 Least Square를 이용하여 경기의 승패에 영향을 미치는 요인분석 및 경기예측을 시도하였다. 몇몇 스포츠경기에서 인공신경망을 이용하여 예측을 한 연구는 보고되었으나 인공신경망으로 농구경기의 승패에 영향을 주는 요인을 검증한 연구는 없다. 따라서 이 연구의 목적은 첫째, 집계 표상의 다양한 계량적인 경기 변수들 중 농구경기의 승패에 가장 크게 영향을 미치는 요인을 찾아내고 이를 인공신경망모형에 입력하여 승패 결과를 예측하여 실제 결과와 비교분석함으로써 이 연구에서 제시한 농구경기의 승패에 영향을 미치는 요인이 옳은지를 검증하는 것이었다.

둘째, 경기 결과예측을 위한 자료 투입시 기존의 연구와 같이 팀별로 승리 및 패배 집단으로 나누어 분석하는 것<sup>1)</sup>과 경기별로 홈팀 및 원정팀으로 구분하여 분석했을 때<sup>2)</sup> 어느 요인이 승패결과에 영향을 미치는지 검토하는 것이었다.

1) 270경기가 이루어졌을 때, 이를 승리집단, 패배집단으로 나누어 540개의 경기결과 자료로 사용해 분석.

2) 270경기가 이루어졌을 때, 이를 승리집단, 패배집단으로 나누지 않고 270개의 경기결과를 그대로 입력 자료로 사용해 분석.

## 연구방법

### 분석방법

이 연구는 회귀분석을 통해 농구경기의 승패를 결정짓는 요인을 찾아내었다. 그리고 이를 검증하기 위하여 로짓회귀분석, 정준판별분석, 그리고 인공신경망을 사용하여 승패예측을 시도, 비교하였다. 검증을 위해 인공신경망을 추가로 사용한 이유는 일반적으로 인공신경망이 로지스틱회귀모형이나 정준판별분석보다 예측력이 높다고 하기 때문이다.

기존의 예측 모형 중 로짓회귀 분석은 종속변수가 0 또는 1일 때 사용하는 분석으로 흔히 승패여부와 같은 이항반응이 있을 경우에 사용된다. 즉, 독립변수가 주어졌을 때  $y$ 값이 1일 확률을 구하는 것이다. 로짓회귀 분석은 odds<sup>3)</sup>를  $\ln(odds)$ 로 선형모형화한 것이며, 아래와 같은 식으로 표현할 수 있다.

$$\ln \frac{p(y=1)}{1-p(y=1)} = \alpha + \beta x$$

판별분석은 소속집단(종속변수)이 이미 알려진 케이스에 대하여 그 소속집단을 판별하는데 효과적이라고 생각되는 여러 개의 설명변수(독립변수)들의 측정값을 가지고 독립변수들의 선형판별함수를 만들어 소속집단을 모르는 케이스에 대해 독립변수 값을 가지고 어느 집단에 속하는지 판별하는데 사용된다.

이 연구에서 사용한 인공신경망<sup>4)</sup>은 Artificial Neural Network의 MLP(Multi-Layer Perceptron)<sup>5)</sup>를 이용하였으며, 오차를 줄이기 위하여 오류역전파(Back propagation) 알고리즘<sup>6)</sup>을 사용하였다. 학습시 활성화함수로서 일반적으로

3) odds =  $p / (1-p)$ 로 사건이 발생할 확률이 발생하지 않을 확률의 몇 배인지를 나타냄.

4) 인공신경망이란 뇌기능의 특성 몇 가지를 컴퓨터 시뮬레이션으로 표현하는 것을 목표로 하는 수학 모델이다. 신경망이라고도 하며, 생물학이나 뇌과학과 엄밀하게 구별하려는 경우에는 앞에 '인공'을 붙인다.

5) MLP : 다층퍼셉트론이라고 하며, 인공신경망에서 은닉층이 여러개 있는 모형이다.

6) 역전파알고리즘 : 다층퍼셉트론 입력층에 들어간 입력 signal은 각 연결강도(weight)와 곱해지고 각 노드(node)에서 더해진다. 출력층에서는 결과값과 실제값을 비교하여 오차가 작은 방향으로 노드간의 연결강도를 조절하는(delta rule)을 사용하여 학습이 이루어진다.

가장 많이 사용되고 있는 시그모이드 함수<sup>7)</sup>를 사용하였으며, 은닉층과 출력층의 개수를 달리하여 실험하였다. 사용된 인공신경망 승패 예측 모형은 <그림 1>과 같다.

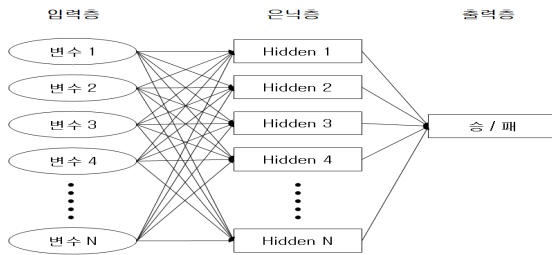


그림 1. 입력층이 N개, 은닉층이 N개, 출력층이 1개인 인공신경망 예 (N-N-1)

승패예측 모형의 비교를 위해 두 가지 실험을 실시하였다. 첫 번째 실험은 팀별 승패예측의 모형검증이다. 한 시즌의 270경기를 승리집단과 패배집단으로 나누었고 총 540경기에 대한 자료를 입력하여 집단의 승리/패배를 예측하였다. 두 번째 실험은 경기 내에서 승패요인을 검증하기 위해 승리/패배 집단으로 나누지 않고 경기 내용을 그대로 입력하여 사용하였다. 그 이유는, 기존연구 관점에서 집단을 나누어 분석했을 시 발생 가능한 양 팀의 승리 또는 패배로 예측될 수 있는 문제를 배제하기 위함이다. 실제로 10%의 자료를 로짓 회귀분석한 결과 27개 경기 중 10경기의 예측이 양 팀 모두 승리 또는 양 팀 모두 패배로 예측되는 것을 알 수 있었다. 이점을 보완할 수 있는 두 번째 실험은 좀 더 우수한 모형이 될 수 있을 것이다.

경기 자료는 2007-2008시즌 프로농구 정규리그 270

경기를 대상으로 KBL홈페이지에 공개된 선수별 상세 집계표를 <그림 2>를 참조하여 입력하였다. 자료 중 60%의 경기는 학습용으로, 20%의 경기는 검증용으로, 20%의 경기는 모형테스트용으로 나누었다.

또한, 기존의 연구에서는 정준판별분석 또는 로짓 회귀 분석으로 예측하였을 경우 전체 자료를 토대로 분류적중률을 알아보았으나 이 연구는 실제 자료를 예측용으로 사용했다는 점에서 기존의 연구와는 차별을 둘 수 있을 것이다.

인공신경망을 구현하기 위하여 Neural Connection 2.0 툴을 사용하였으며, 로짓회귀 분석 및 판별분석을 위해 SPSS 12.0 툴을 사용하였다.

## 실험 설계

이 연구는 농구경기 승/패에 영향을 주는 요인을 찾기 위해 회귀분석을 실시하였으며, 이를 검증하기 위해 인공신경망 모형을 활용하였다. 실험에 사용된 변수와 자료의 수는 <표 1>과 같다.

표 1. 실험에서 사용한 변수

실험구분	입력변수	출력변수	훈련	검증	Test
실험 1-1	14	1	324	108	108
실험 1-2	14	2	324	108	108
실험 1-3	9	1	324	108	108
실험 1-4	9	2	324	108	108
실험 2-1	28	1	162	54	54
실험 2-2	28	2	162	54	54
실험 2-3	14	1	162	54	54
실험 2-4	14	2	162	54	54

No	선수명	SCORE						Min	2P	2PA	%	3P	3PA	%	FG%	FT	FTA	%	DA	REBOUNDS			Ast	Stl	GD	BS	PF			TF	TO
		1Q	2Q	3Q	4Q	EQ	Tot													Off	Def	Tot					W/FT	w/oFT	Tot		
15	루로	5	15	14	4	2	40	00:44:59	16	23	70	1	4	25	63	5	8	63	4	10	7	17	3	0	0	2	4	1	5	0	3
9	임재현	0	5	2	3	2	12	00:36:42	1	3	33	1	2	50	40	7	7	100	0	1	2	3	2	1	0	0	3	2	5	0	3
10	문경은	0	9	6	8	0	23	00:32:22	4	6	67	3	10	30	44	6	6	100	0	0	1	1	3	1	0	0	1	2	3	0	1
7	방성운	2	0	3	7	0	12	00:31:51	2	5	40	2	6	33	35	2	2	100	0	6	6	9	3	0	0	0	1	2	3	0	0
1	노경석	0	0	0	0	3	3	00:25:00	0	1	0	1	6	17	14	0	0	0	0	2	1	3	3	0	0	0	1	0	1	0	0
4	키부 스콧	5	0	0	2	0	7	00:13:08	2	4	50	0	0	0	50	3	3	100	0	0	2	2	0	0	0	1	4	1	5	0	1
13	전희철	0	0	0	2	0	2	00:17:39	1	3	33	0	4	0	14	0	0	0	0	0	2	2	1	0	0	1	5	0	5	0	0
5	장택영	0	0	0	0	0	0	00:10:56	0	1	0	0	2	0	0	0	0	0	0	1	0	1	1	0	0	0	0	2	2	0	0
21	김종학	0	0	0	0	0	0	00:06:23	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	0	0
9명	TOTAL	12	29	25	26	7	99	03:45:00	26	47	55	8	34	24	42	23	26	88	4	17	21	38	16	2	0	4	20	11	31	0	8

그림 2. KBL 집계표

7) 시그모이드 함수 : 0과 1에 무한히 가까워 지는 형태를 가진 함수다. 따라서 어떤 값이 시그모이드 함수를 거치게 되면 미분 가능한 형태로 변환된다.

표 2. 선행연구에서 사용된 변수

연도	1994	2000	2005	2005	2006	2007	2008	2008
저자	박제영	박제영 외 2	최현희 임백빈	이근모 이재형	한재순 외 3	박제영	김세형 외 3	김세형
분석방법	판별분석	판별분석 D/T <sup>8)</sup>	판별분석	판별분석	판별분석	판별분석	로짓회귀 D/T	로짓회귀 D/T
종속변수	승/패	승/패	승/패	승/패	승/패	승/패	승/패	승/패
2점슛 성공률	0	0	0	0	0	0	0	0
3점슛 성공률	0	0	0	0	0	0	0	0
공격 리바운드	0	0	0	0	0	0	0	0
덩크슛 성공률		0	0	0			0	0
로스트 볼	0							
루즈볼 캐치	0							
속공 성공률			0	0				
수비 리바운드	0	0	0	0	0	0	0	0
블록 슛	0	0	0	0		0	0	0
실책			0	0				
어시스트	0	0	0	0	0	0	0	0
인터셉트	0	0	0	0	0	0	0	0
자유투 有과울	0	0					0	0
자유투 성공률	0		0		0	0	0	0
자유투 無과울	0	0					0	0
전체 리바운드			0	0				
좋은 수비	0		0	0	0		0	0
턴 오버		0			0	0	0	0
테크니컬 파울	0	0					0	0
파울 총 횟수			0	0		0		
필드골 성공률			0					

인공신경망의 특성상 은닉층의 개수에 따라 예측값이 다르게 나타나기 때문에 본 실험에서는 은닉층의 수를 변화시켜가며 (1, 2, 3, 4, 5, 10, 15, 20, 50, 100, 200) 모형을 구축해보았다. 모형에 사용된 입력자료의 경우 14개 혹은 28개 모두 사용되거나 승패에 요인을 미치는 변수만을 선정하여 사용하는 경우를 고려하였다. 출력층은 1개 아니면 2개이다.

검증 실험은 실험 1과 실험 2로 분류하여 각각의 실험에는 입력변수 및 출력변수의 수에 따라 4개의 세부실험이 존재한다. 예를 들어 실험 1-1에서는 14개의 입력변수와 1개의 출력변수, 그리고 은닉층 11개가 사용된다. 실험 1-2에서는 14개의 입력변수와 2개의 출력변수, 그리고 은닉층 11개가 사용된다. 따라서 모든 실험 경우의 수는  $2 \times 4 \times 11 = 88$ 이 된다.

## 변수 선정

〈표 2〉는 기존연구에서 사용한 변수를 정리한 것이다. 이 연구는 기존연구에서 많이 사용한 변수 중 유의미한

것으로 검증된 변수를 살펴보았다. 기록 자료는 KBL 홈페이지에서 추출하였다. 이 실험에서 사용한 입력 변수는 프로농구 집계표의 자료중 김세형(2008)의 논문에서 사용한 공격변인 7개와 수비변인 7개이다(표 3).

표 3. 실험에 투입한 변수

변수명		비고
공격 변수 7 개	2점슛 성공률	성공수 / 시도수
	3점슛 성공률	성공수 / 시도수
	자유투 성공률	성공수 / 시도수
	덩크슛 성공률	성공수 / 시도수
	공격 리바운드	횟수
	어시스트	횟수
	턴오버	횟수
수비 변수 7 개	수비 리바운드	횟수
	가로채기	횟수
	굿 디펜스	횟수
	블록 슛	횟수
	자유투有 반칙	횟수
	자유투無 반칙	횟수
	테크니컬 파울	횟수
종속변수	승, 패	1, 0

8) D/T : Decision Tree

실험 1-3, 1-4에서 사용한 변수는 <표 3>의 변수 중 다중회귀 분석에서 유의미하게 나타난 2점 슛 성공률, 3점 슛 성공률, 자유투 성공률, 공격 리바운드, 수비 리바운드, 가로채기, 블록 슛, 자유투有 반칙, 턴 오버 9가지 변수이다. 이 연구에서는 이를 '07-'08 국내 남자프로농구경기의 승패에 영향을 준 팀별 요인으로 제시하며, 이를 실험 1의 독립변수로 사용하여 요인 검증 및 승패를 예측하고 모형을 비교 검증하였다(<표 4>).

표 4. 14개의 변수 중 유의미하게 나타난 변수 (팀별)

입력변수	비표준화계수		t	유의 확률
	B	표준오차		
상수	-2.806	0.221	-12.719	0.000
2점슛 성공률	2.024	0.232	8.711	0.000
3점슛 성공률	1.874	0.170	11.020	0.000
자유투 성공률	0.614	0.142	4.336	0.000
덩크슛 성공률	-0.003	0.037	-0.069	0.945
공격 리바운드	0.026	0.005	5.180	0.000
어시스트	-0.007	0.004	-1.626	0.105
턴오버	-0.038	0.005	-7.736	0.000
수비 리바운드	0.049	0.004	12.854	0.000
가로채기	0.049	0.006	7.931	0.000
굿 디펜스	0.026	0.015	1.760	0.079
블록 슛	0.027	0.008	3.286	0.001
자유투有 반칙	-0.014	0.005	-2.951	0.003
자유투無 반칙	0.006	0.007	0.771	0.441
테크니컬 파울	-0.013	0.030	-0.428	0.669

실험 2-3, 2-4에서 사용한 변수는 다중회귀 분석을 통해 유의미하게 나타난 변수인 홈팀 2점 슛 성공률, 3점 슛 성공률, 자유투 성공률, 공격리바운드, 수비리바운드<sup>9)</sup>, 자유투有 반칙, 턴 오버, 원정팀 2점 슛 성공률, 3점 슛 성공률, 자유투성공률, 공격리바운드, 수비리바운드, 자유투有 반칙, 턴 오버의 14가지 변수이다. 이 연구에서는 이를 07-08 국내 남자프로농구경기의 승패에 영향을 끼친 경기별 요인으로 제시하며, 이를 실험 2의 독립변수로 사용하여 요인 검증 및 승패를 예측하고 모형을 비교 검증하였다(<표 5>).

<표 6>은 위의 <표 4>와 <표 5>에 사용된 회귀분석의 모형을 요약한 것이다.

9) 실험1에서 유의확률의 인정범위는 0.05%수준으로 정했으며, 실험2에서 유의확률의 인정범위는 0.1%로 정했다. 이유는 실험2에서 변수가 2배로 늘어났기 때문에 유의확률의 인정범위도 2배로 정했다. 일반적으로 유의확률 0.1%까지는 유의미한 것으로 본다.

표 5. 28개의 변수 중 유의미하게 나타난 변수 (경기별)

입력변수	비표준화계수		t	유의 확률
	B	표준오차		
상수	0.520	0.577	0.900	0.369
H_2점슛 성공률	1.896	0.414	4.585	0.000
H_3점슛 성공률	1.563	0.258	6.064	0.000
H_자유투 성공률	0.645	0.181	3.555	0.000
H_덩크슛 성공률	0.001	0.047	0.015	0.988
H_공격 리바운드	0.017	0.007	2.334	0.020
H_어시스트	-0.005	0.005	-0.906	0.366
H_턴오버	-0.030	0.010	-3.071	0.002
H_수비 리바운드	0.013	0.007	1.854	0.065
H_가로채기	0.002	0.012	0.193	0.847
H_굿 디펜스	-0.019	0.022	-0.832	0.406
H_블록 슛	0.001	0.012	0.092	0.927
H_자유투有 반칙	-0.024	0.006	-3.807	0.000
H_자유투無 반칙	0.000	0.010	-0.033	0.974
H_테크니컬 파울	-0.041	0.037	-1.085	0.279
A_2점슛 성공률	-1.636	0.376	-4.354	0.000
A_3점슛 성공률	-1.831	0.258	-7.107	0.000
A_자유투 성공률	-0.513	0.185	-2.768	0.006
A_덩크슛 성공률	0.016	0.045	0.366	0.715
A_공격 리바운드	-0.023	0.007	-3.277	0.001
A_어시스트	-0.001	0.006	-0.127	0.899
A_턴오버	0.048	0.010	4.773	0.000
A_수비 리바운드	-0.026	0.008	-3.413	0.001
A_가로채기	-0.012	0.011	-1.042	0.298
A_굿 디펜스	-0.017	0.021	-0.819	0.414
A_블록 슛	0.005	0.010	0.458	0.647
A_자유투有 반칙	0.030	0.006	5.054	0.000
A_자유투無 반칙	0.004	0.010	0.407	0.685
A_테크니컬 파울	0.039	0.041	0.952	0.342

표 6. 모형 요약 (표4, 표5)

모형	R	수정된 R <sup>2</sup>	표준오차	Durbin-Watson
14개 변수	0.705	0.484	0.360	1.941
28개 변수	0.826	0.646	0.298	1.988

## 연구결과

이 연구에서는 회귀분석을 통해 국내 남자 프로농구 경기의 승패에 영향을 미치는 팀별/경기별 요인을 찾아보고, 이를 검증하기 위해 승패예측을 하였다. 또한 승패예측을 다양한 모형으로 비교 검증함으로써 추후 승패예측에 관한 연구에 참조될 수 있는 모형을 제시하였다.

## 모형별 결과 분석 비교

### 실험 1 (팀별)

입력변수로 모든 변수를 사용한(공격변인 7개, 수비변인 7개) 값은 <표 7>과 같다. <표 7>의 모형을 보면 여러 가지 인공지능망 모형이 있는데 예를 들면 14-3-1은 입력변수가 14개, 은닉층이 3개, 출력층이 1개임을 나타낸다. 모형은 출력층이 1개인 경우와 2개인 경우로 나뉜다. 출력층이 1개인 경우는 <그림 1>과 같이 종속변수가 한 개로 0 또는 1의 값이 입력되며, 출력층이 2개인 경우는 종속변수가 두 개로 한 쪽의 출력이 0, 다른 한쪽의 출력이 1이다. 결과를 보면 출력층이 1개인 모형이 대체로 예측력이 우수하다는 것을 알 수 있었다. <표 8>은 다중회귀분석을 통해 유의미하게 나타난 변수를 입력변수로 사용하여 나온 결과 값이다.

실험 1-1, 1-2는 농구경기의 모든 변수를 사용한 결과 값이다. <표 7> 로짓 회귀모형의 예측력이 87.96%, 정준판별분석의 예측력이 77.78%, 인공지능망모형의 평균

표 7. 실험 1-1, 1-2 결과

모형	예측값	모형	예측값
14-1-1	80.56%	14-1-2	80.56%
14-2-1	82.41%	14-2-2	81.48%
14-3-1	82.41%	14-3-2	82.41%
14-4-1	81.48%	14-4-2	82.41%
14-5-1	82.41%	14-5-2	79.63%
14-10-1 <sup>10)</sup>	86.11%	14-10-2	79.63%
14-15-1	84.26%	14-15-2	80.56%
14-20-1	82.41%	14-20-2	81.48%
14-50-1	83.33%	14-50-2	81.48%
14-100-1	80.56%	14-100-2	79.63%
14-200-1	81.48%	14-200-2	76.85%
로짓회귀	87.96%	정준판별	77.78%

표 8. 실험 1-3, 1-4 결과

모형	예측값	모형	예측값
9-1-1	83.33%	9-1-2	81.48%
9-2-1	84.26%	9-2-2	85.19%
9-3-1	82.41%	9-3-2	83.33%
9-4-1	82.41%	9-4-2	84.26%
9-5-1	81.48%	9-5-2	79.63%
9-10-1	81.48%	9-10-2	81.48%
9-15-1	85.19%	9-15-2	85.19%
9-20-1	84.26%	9-20-2	84.26%
9-50-1	84.26%	9-50-2	85.19%
9-100-1	83.33%	9-100-2	81.48%
9-200-1	80.56%	9-200-2	85.19%
로짓회귀	88.89%	정준판별	70.37%

10) 인공지능망 14-10-1모형은 <그림 1>처럼 입력층이 14개, 은닉층이 10개, 출력층이 1개인 모형이다.

예측력이 81.52%로 나타나는 것을 확인 할 수 있었다.

이 연구에서 승패에 영향을 미친다고 제시한 팀별 변수를 입력변수로 사용한 실험 1-3, 1-4의 결과를 보면 로짓 회귀모형의 예측력이 88.89%, 정준판별분석의 예측력이 70.37%, 인공지능망모형의 평균 예측력이 83.17%로 나타났다. 이는 실험 1-1과 1-2보다 인공지능망 모형과 로지스틱 회귀 모형의 예측력이 우수하게 나타났기 때문에, 이 연구에서 제시한 팀별 요인이 농구경기의 승패에 영향을 미치고 있다는 것을 증명할 수 있었다.

모형 간 예측력을 비교해보면 로짓회귀분석이 가장 우수하였고, 인공지능망, 정준판별분석 순으로 나타났다.

### 실험 2 (경기별)

각 팀의 공격변인 14개와 수비변인 14개의 자료 28개를 동시에 입력변수로 사용하여 얻은 결과 값은 <표 9>와 같다. 모형은 실험 1과 마찬가지로 출력층이 1개인 경우와 2개인 경우로 분류된다. 결과를 살펴보면 경기의 승패에 유의미하게 영향을 끼칠 것이라고 나타나는 변수 14개를 입력변수로 사용한 결과 예측력이 우수한 것을 알 수 있었다. <표 10>은 다중회귀분석을 통해 유의미하게 나타난

표 9. 실험 2-1, 2-2 결과

모형	예측값	모형	예측값
28-1-1	85.19%	28-1-2	81.48%
28-2-1	87.04%	28-2-2	83.33%
28-3-1	85.19%	28-3-2	88.89%
28-4-1	88.89%	28-4-2	87.04%
28-5-1	90.74%	28-5-2	85.19%
28-10-1	85.45%	28-10-2	88.89%
28-15-1	88.89%	28-15-2	90.74%
28-20-1	90.74%	28-20-2	92.59%
28-50-1	87.04%	28-50-2	92.59%
28-100-1	87.04%	28-100-2	88.89%
28-200-1	85.19%	28-200-2	81.48%
로짓회귀	85.19%	정준판별	-

표 10. 실험 2-3, 2-4 결과

모형	예측값	모형	예측값
14-1-1	88.89%	14-1-2	90.74%
14-2-1	87.04%	14-2-2	83.33%
14-3-1	90.74%	14-3-2	88.89%
14-4-1	85.19%	14-4-2	87.04%
14-5-1	90.74%	14-5-2	85.19%
14-10-1	90.74%	14-10-2	87.04%
14-15-1	92.59%	14-15-2	90.74%
14-20-1	90.74%	14-20-2	90.74%
14-50-1	87.04%	14-50-2	88.89%
14-100-1	92.59%	14-100-2	83.33%
14-200-1	88.89%	14-200-2	88.89%
로짓회귀	92.59%	정준판별	-

변수를 입력변수로 사용하여 나온 결과이다.

실험 2-1, 2-2의 결과는 농구경기의 모든 변수를 사용한 결과 값이다. <표 9>에서 로짓 회귀모형의 예측력이 85.19%, 인공신경망모형의 평균 예측력이 87.39%로 나타났다. 실험2에서 정준판별분석을 실시하지 않은 이유는 판별분석은 독립변수들을 토대로 결과 값이 어떤 집단에 포함 될 것인가를 알아보는 것이기 때문에 집단을 나누지 않은 실험2의 경우에는 의미가 없기 때문이다.

이 연구에서 제시한 경기별 변수를 입력변수로 사용한 실험 2-3, 2-4의 결과를 살펴보면 로짓 회귀모형의 예측력이 92.59%, 인공신경망 모형의 평균 예측력이 88.64%로 나타나 로짓회귀 모형과 인공신경망 모형의 예측력이 증가한 것을 알 수 있었으며, 이 연구에서 제시한 경기별 요인이 농구경기의 승패에 영향을 미치고 있다는 것을 증명할 수 있었다.

모형 간 예측력을 비교해보면 실험1에서와 마찬가지로 로짓회귀분석이 가장 우수하였다.

## 결 론

이 연구에서는 '07-'08시즌 국내 남자프로농구 경기의 승패에 영향을 미치는 요인을 팀별 요인과 경기별 요인으로 나누어 검토해 보았다. 요인 분석을 위해 다중회귀 분석을 실시하였으며, 이를 검증하고 승패예측을 통한 모형간의 비교를 위해 로짓회귀분석, 판별분석, 인공신경망 모형을 사용하여 승패예측을 하였다.

승패 예측 결과 이 연구에서 제시한 변수만을 사용하였을 때 예측 결과가 높은 것을 확인 할 수 있다. 예측결과가 높아진다는 것은 외부의 또 다른 영향 즉, 왜곡을 줄인 것이라고 볼 수 있다. 이는 이 연구에서 제시한 요인들이 경기의 승패에 많은 영향을 미치고 있다는 것을 반증한다.

실험 1은 '07-'08시즌 국내 남자프로농구경기의 270경기를 승리한 팀과 패배한 팀으로 나누어 팀별로 승리 및 패배에 영향을 미치는 요인을 찾아보고 검증하는 내용이었다. 실험결과 모든 변수를 투입변수로 사용한 실험1-1, 1-2보다 실험1-3, 1-4의 인공신경망 모형과 로지스틱 회귀 모형의 예측력이 우수하게 나타났기 때문에, 이 연구에서 제시한 팀별 요인이 농구경기의 승패에 영향을 미치고 있다는 것을 증명할 수 있었다.

실험 2는 '07-'08시즌 국내 남자프로농구경기의 270경

기를 승리한 팀과 패배한 팀으로 나누지 않고 경기 자체 내에서 승리 및 패배에 영향을 미치는 요인을 찾아보고 검증하는 내용이었다. 실험결과 모든 변수를 투입변수로 사용한 실험 2-1, 2-2보다 실험 2-3, 2-4의 인공신경망과 로지스틱 회귀 모형의 예측력이 우수하게 나타났기 때문에, 이 연구에서 제시한 팀별 요인이 농구경기의 승패에 영향을 미치고 있다는 것을 증명할 수 있었다.

이 연구는 기존의 프로농구 승패예측에서 사용되지 않았던 인공신경망을 사용했다는 것과 '07-'08 한국 프로농구 경기 기록집계표상의 변수 중에서 승패결과에 유의한 변수를 찾았다는 것에 의의가 있다. 또한, 기존 연구와는 다르게 팀별, 경기별로 승패요인을 분석해보고 이를 검증하기 위한 방법으로 승패예측이라는 새로운 시도를 해봤다는 데 들 수 있을 것이다.

승패 집단을 나누어 팀별로 분석했을 때(실험1)에는 기존 연구와 비슷하게 2점 슛 성공률, 3점 슛 성공률, 자유투성 공률, 공격리바운드, 수비리바운드, 가로채기, 블록슛, 자유투 반칙, 턴 오버의 9가지 변수가 승패에 영향을 미치는 것으로 나타났다. 승패 집단을 나누지 않고 경기 그 자체를 분석했을 때(실험2)에는 기존 연구에서 사용된 변수에서 가로채기와 블록슛이 제거 된 것을 확인할 수 있었다. 이는 '07-'08시즌 경기 내에서 가로채기와 블록슛이 승패와 직접적인 관련이 없다는 의미로 해석 할 수 있다 즉, 가로채기와 블록 슛이 득점으로 바로 연결되지 않았음을 알 수 있다.

승리패배집단 분류방법과 홈팀/원정팀으로 나누어 분류하는 방법을 비교해보면 홈팀/원정팀으로 나누어 분류하는 것이 더욱 효과적이라는 것을 알 수 있다.

모형간의 비교를 살펴보면 로짓회귀 분석 > 인공신경망 > 정준판별분석의 순으로 우수한 예측력을 가지는 것으로 나타났다. 인공신경망을 개별적으로 볼 경우 로짓 회귀분석보다 예측력이 우수한 것도 있었지만 평균으로 나타내면 그렇지 않았다. 추후 연구에서 인공신경망을 사용한다면 은닉층의 개수는 5-20개가 적당하다고 판단된다.

이 연구의 한계점으로는 승패를 결정짓는 요인을 찾아내 인공신경망과 로짓회귀 분석을 통해 예측을 시도했으나 일반적인 이론과는 반대로 인공신경망의 예측 값이 로짓회귀 분석과 큰 차이가 나지 않았다는 점이다. 이는 인공신경망의 특성상 연구에 사용된 경기 수가 적기 때문과 은닉층의 범위가 너무 넓었기 때문이라고 판단된다. 향후 연구에서는 과거 5년 동안의 자료를 가지고 분석하여 요인을 찾아내고 예측한다면 좀 더 정확한 모형 비교가 될 것으로 사료된다.

## 참고문헌

- 김세형, 강상조, 박재현, 김혜진(2008). 한국프로농구 경기 기록 분석에 의한 승패결정요인. 한국체육측정평가학회지, 10(1), 1-12.
- 김세형(2008). 프로농구 경기기록 분석을 통한 승패 예측모형 개발. 미간행 석사학위논문. 한국체육대학교대학원.
- 김주학, 노갑택, 박종성, 이원희(2007). 신경망분석을 이용한 축구경기 승패예측모형개발 -2006독일월드컵대회를 중심으로. 체육과학연구원, 18(4), 54-63.
- 박제영(1994). 농구경기의 승패요인 분석에 관한 연구 - 제 40회 세계군인 농구선수권대회를 중심으로. 수원대학교 논문집, 12, 205-215.
- 박제영(2008). 2006-2007시즌 한국프로 농구경기의 승패 요인분석. 한국체육과학회지, 17(2), 129-138.
- 박제영(2007). 제15회 아시아경기대회 남자농구경기의 승패요인 분석. 한국사회체육학회지, 30, 941-950.
- 박제영, 김차용, 지용범(2000). 99-2000시즌 한국 프로농구 경기의 승패요인 -의사결정나무 분석을 이용하여. 한국사회체육학회지, 14(1), 327-338.
- 이근모, 이재형(2005). 국제 남자농구 경기의 승패요인 분석. 한국사회체육학회지, 24, 553-565.
- 최현희, 임백빈(2005). 농구경기시 득점차 분류에 의한 승패 요인 분석. 한국스포츠리서치, 16(6), 223-235.
- 최형준, 김주학(2006). 인공신경망을 이용한 2005년도 영국 워블던 테니스 대회의 경기결과 예측에 관한 연구. 한국체육학회지, 45(3), 459-467.
- 한계순, 이근일, 김종필, 천윤석(2006). 한국여자 프로농구 리그(WKBL) 경기의 공격 및 수비요인이 승패에 미치는 영향. 체육과학연구논총, 16(1), 57-71.
- 한필수(2006). 2006 겨울리그 자료를 이용한 경기상황별 승패 요인 분석. 코칭능력개발지, 8(3), 85-94.
- Abbott, Douglas M. (2005). *Analysis of a predictive model for college choice by high school football players*. University of Montana.
- Stefani, R. T. (1977). *Football and Basketball Predictions Using Least Squares*. Systems, Man and Cybernetics, IEEE Transactions on, 7(2), 117-121.
- Stefani, R. T. (1980). *Improved Least Squares Football, Basketball, and Soccer Predictions*. Systems, Man and Cybernetics, IEEE Transactions on, 10(2), 116-123.

## A Comparison Study on the Prediction Models for the Professional Basketball Games

Seung-Hwan Gu, Hyun-Soo Kim, & Seong-Yong Jang  
Seoul National University of Technology

The purpose of this paper was to analyze the winning factors for the professional baseball games using the logit regression model and artificial neural network(ANN). All the data on 270 games of 2007-2008 Korea Basketball League is used for the model development and determination of the winning factors. Regression analysis has been done to evaluate the factors to affect the win or loss and two types of experiments has been done to predict the game results. In the first experiment, 270 games are divided into win and loss groups, and result in 540 games in total. In second experiment, 270 games are not divided into two groups. All cases of experiment are  $2 \times 4 \times 11 = 88$  in all. Sixty percentage of the data are used for model building, 20% of the data for model validation and the rest of the data for the model prediction. The experiment result shows that all ANN models and logit model proposed in existing research gives more than 85% of prediction accuracy, the case of no classification into two groups has better prediction accuracy than the case of classification and the proposed ANN model using the only significant factors gives better results compared to those using all the possible factors. This study proves that the selected factors affect the game results significantly.

**key words:** Logit Regression Model, Artificial Neural Network, Basketball, Prediction Model, Winning Factors

