Journal of the Korean Data & Information Science Society
2009, **20**(6), 1119–1127

한국데이터정보과학회지

신경망을 이용한 우승자 예측모형[†]

민대기[1] · 현무성[2]

[1]덕성여자대학교 통계학과 · [2]강남대학교 사회체육학과

접수 2009년 9월 10일, 수정 2009년 11월 20일, 게재확정 2009년 11월 24일

요 약

골프경기에서 상금이나 평균타수와 같은 척도에는 명확한 기록이 정의되어 있지만 누가 우승을 할 것인가 하는 관점에서는 Tiger Woods나 Phil Mickelson 그리고 Steve Stricker 등 2009년에 3승 이상을 한 선수를 제외하면 과연 누구일까 하는 의문을 갖게 될 것이다. 왜냐하면 워낙 선수층이 두터워 백지한창 차이의 실력을 갖춘 우승후보 선수들이 많고, 다른 종목보다 정신력이 결과에 많은 영향을 미치기 때문이다. 본 연구에서는 복잡한 비선형 형태의 자료를 파악하는데 아주 유용한 도구인 신경망을 이용하여 2009년 PGA자료를 바탕으로 우승자 예측모형에 대하여 연구를 하였다.

주요용어: 결합함수, 로지스틱모형, 신경망, 포아송모형. 활성함수.

1. 서론

복잡한 비선형모형에 예측이 뛰어난 신경망은 그 이론적 구성이 비선형 로지스틱 회귀, 최적화 방법 등 광범위하고 다양하며, 결과를 해석하기가 용이하지 않아 적절하게 활용하기에 어려움이 있다. 강현철 (2006)은 은닉 층과 은닉마디가 증가하면서 유연성이 뛰어나고 다양한 모형적합성이 가능하나 복잡한 모형일수록 추정하는 계수가 증가하여 계수추정과 최적화에 어려움이 있다고 하였다. 그러나 신경망은 복잡한 구조를 가진 자료에서의 예측문제를 해결하기 위해서 사용되는 뛰어난 비선형모형의 하나로 분류된다. 신경망모형을 이용한 예측모형 연구를 살펴보면 Kim과 Lee (2003)의 신용평가모형과 Cho와 Park (2008)의 보험회사 이탈고객 분석에 관한 것 등이 있다. PGA (Professional Golf Association)골프에서 선수들의 경기능력는 상금 액수나 평균타수로 표시되며 그 경기능력는 일반적으로 드라이브거리, 페어웨이 정확도, 그린 적중 율, 평균 퍼팅 수에 의해서 결정된다. 하지만 우승 가능 권에 속한 선수 중 우승을 경험한 선수와 하지 못한 선수의 차이는 어떻게 구별 할 수 있을까 하는 의문이 든다. 왜냐하면 단순하게 우승 경험자와 못한 자를 0과 1로 이분하기에는 정보의 손실이 너무 많고 그 우승 횟수를 0부터 n까지의 카운트한 자료는 포아송 분포형태로 보이기 때문이다. 그러나 n이 충분히 크지 않을 때는 반응변수를 이분하여 모형설정을 하는 것이 바람직 하기 때문에 우승 가능성에 대한 예측모형을 자료의 반응변수 형태에 따라 포아송 회귀모형과 신경망모형을 이용하여 연구하였다. 신경망은 입력변수와 출력변수에 따라 결합함수와 활성함수를 선택하여 모형을 구축한다. 본 연구에서는 반응변수 0,1의 형태를 신경망에 적용하기 때문에 일반적으로 가장 많이 사용되는 MLP (Multilayer Perceptron)와 일반적으로 분류모형에 많이 쓰이는 NRBF (Normalized Radial Basic Function)모형의 5개 신경망을 적용 후 결과를 비교 하였다.

[†] 본 연구는 2008년도 덕성여자대학교와 강남대학교 교내연구 지원을 받았음.
[1] 교신저자: (132-714) 서울특별시 도봉구 근화교길19, 덕성여자대학교 통계학과, 부교수.
E-mail: dkmin@duksung.ac.kr
[2] (446-702) 경기도 용인시 기흥구 구갈동 111번지, 강남대학교 사회체육학과, 부교수.

2. 연구방법

2.1. 연구목적

일반적인 선수들의 경기능력, 즉 상금 액수나 평균타수는 드라이브거리, 정확도, 그린 적중 율, 평균 퍼팅 수에 의해서 결정된다. 본 연구에서는 우승 가능 권에 속한 선수 중 우승을 경험한 선수와 하지 못한 선수의 변별력에 대한 예측모형을 연구 하였다.

2.2. 반응변수와 예측변수의 선택 및 기술적 분석

데이터 수집은 2009년 PGA자료를 인터넷을 통하여 하였다. 선수들의 해당년도에 대한 종합적인 경기능력를 측정할 수 있는 상금, 평균타수, Fedex 포인트 등 몇 가지 가능한 항목이 있지만 우승가능 확률을 예측하기 위하여 10위권 경험자 중 우승 횟수를 반응변수를 설정 하였다. 예측변수로는 평균 스코어나 상금 등에 가장 중요한 영향을 주는 Gir (green in regulation), Avegputt (average putts per green in regulation), Avedist (average driving distance), Hitratio (driving accuracy), Scrambling , Ssave (sand save) 등의 기록을 수집하였다.

2.2.1. 선택된 변수설명 및 기술적 분석

Gir는 선수가 그린 위에 파보다 2타 적은 타수에 볼을 올릴 수 있는지를 측정한다. 즉 선수가 파4에서 경기를 하면 2번안에 볼이 그린 위에 도달할 수 있는지를, 파3에서는 1번에, 파5에서는 3번안에 볼이 그린 위에 도달 할 수 있는지를 측정한다. 그러므로 Gir통계는 전체 홀 중 몇 홀이 그린 위에 규정보다 적은 타수에 도달했는지를 비율로 나타내면 일반적으로 아이언샷의 정확도로 해석된다. Avegputt는 온 그린 시 그린위에서의 평균퍼팅수로 일반적으로 홀 당 퍼팅수로 표시하며 라운드 당 전체 퍼팅수와 산출방식이 다르다. Avegputt는 라운드당 퍼팅 수보다 평균스코어와의 관계도 훨씬 높게 나타난다. Avedist는 경기에서 대부분의 선수들이 드라이버로 티샷을 할 만 한 두 홀을 선택하여 라운드 마다 거리를 측정한다. 이 수치는 선수들의 파워에 대한 측정 자료로 해석되며 최근에는 과학적인 운동방법과 기구의 발달로 과거 5년 전보다 평균비거리가 20야드 정도 증가 되었다. 우승권에 접하기 위해서는 필요적으로 갖춰야할 요소의 하나이다. Hitratio은 파3을 제외한 각 홀에서 티샷을 한 볼이 페어웨이에 떨어졌는지를 %로 나타낸다. 그린에 볼을 올리기 위해서는 러프보다는 페어웨이 공이 놓인 것이 훨씬 유리하며, 그린위에서 핀의 위치에 따라 페어웨이의 어느 지점에 공을 도착시킬까 하는 것은 스코어를 결정하는 매우 중요한 요인이다. Scrambling은 규정된 타수 안에 볼이 그린 위에 올라가지 못한 경우에 그 홀에서 파 이상의 좋은 성적을 낸 경우를 비율로 표시하며 일반적으로 그린주변에서의 숏게임 능력으로 평가된다. Berry (2001)는 이 수치를 퍼팅능력과 같이 해석 하였으나 PGA 상금 랭킹 상위권에 속한 선수들의 1-2미터 내의 짧은 거리에 대한 퍼팅능력을 감안한다면 숏게임의 능력에 대한 해석이 적당 할 것이다.

표 2.1 예측변수와 반응변수의 정의

변수	변수정의	변수역할
First	우승횟수	반응변수
Avedist	평균 드라이브거리	예측변수
Avegputt	그린적중 시 평균 퍼팅수	예측변수
Gir	그린적[a]q 율	예측변수
Hitratio	드라이브 정확도	예측변수
Scrratio	scrambling	예측변수
Ssave	Sand Save	예측변수

표 2.2 변수의 기술 통계량

변수	N	평균	표준편차	최소값	최대값
Avegputt	151	1.77	0.02	1.73	1.85
Avedist	151	289.09	7.97	268.10	312.20
Gir	151	64.82	2.17	58.08	70.54
Hitratio	151	62.88	5.42	48.11	73.54
Scrratio	151	58.57	3.46	49.67	68.18
Ssave	151	50.40	6.23	30.77	64.43

2.2.2. 반응변수의 설정 및 분포

10위권 경험자에 속한 151명 즉 최소한 1번 이상 각 대회의 10위권 이내에 진입한 적이 있는 선수에 대한 우승횟수의 분포는 표 2.3 에서 알 수 있듯이 정규분포를 따르지 않고 심하게 한쪽으로 치우쳐 있다. 2009년도의 우승기록을 살펴보면 Tiger Woods가 6번으로 가장 많이 하였고 Steve Stricker와 Phil Mickelson이 각 각 3번을 했으며 양용은을 포함한 5명의 선수가 2번 우승을 했다. 우승을 한번 한 선수로는 Sean O'Hair를 포함한 19명 이었다. 상금과 우승횟수를 비교하면 대체로 비례하지만 Jim Furyk이나 나상욱과 같이 10위 안에 전체시즌 투어 중 11번, 9번이나 기록되었으나 우승이 없는 선수도 있었다. 이번연구의 목적은 우승 확률 예측모형 개발이지만 반응변수를 단순하게 우승횟수가 1번 이상인 선수와 그렇지 않은 선수로 나누는 경우에, 우승권에 많이 접근한 Jim Furyk, 나상욱 같은 선수를 우승이 없는 선수로 단순하게 분류하는 것은 많은 정보의 손실을 가져 올 수 있는 단점이 있다. 또한 포아송 회귀모형을 적용하기 위하여 우승횟수를 반응변수로 설정 했을 경우 151명의 전체대상 중 124명이 우승이 한번 도 없는 0에 해당 되어 전체 관측치중 82%가 0에 집중됨을 알 수 있었다.

표 2.3 우승횟수 (first)의 기술적 분석

우승 횟수	빈도수	퍼센트
0	124	82.12
1	19	12.58
2	5	3.31
3	2	1.32
6	1	0.66

표 2.4 반응변수 winner의 기술적 분석

우승경험자 (winner)	빈도수	퍼센트
1	27	17.88
0	124	82.12

3. 모형 및 결과분석

3.1. 포아송 회귀 분석 및 변수선택

포아송 회귀모형은 0또는 그 이상 양수의 발생건수를 갖는 반응변수에 적합하다. 종전에는 일반적으로 카운트자료를 일반회귀 모형으로 분석을 시도 하였으며 분포가 한쪽으로 치우친 경우 변수변환을 통하여 정규분포를 만든 다음 일반회귀 모형분석을 시도 하였다. 포아송 회귀모형은 이에 반하여 이산 형 카운트 자료나 심하게 편향되어 있는 자료에 변환하지 않고 분석을 할 수 있는 장점이 있다. 단점으로는

포아송 회귀모형은 평균에 비하여 분산이 아주 큰 경우 적합하지 못한 점을 갖고 있다. 우승횟수를 나타내는 반응 자료는 0에 전체의 82%가 발생되어 있고 3에는 2번 6에는 1번만 발생하여 0에 심하게 편중되어 있는 것을 볼 수 있다. 표2.2에서 언급한 설명변수 중 Gir과 Hitratio는 반응변수에 대하여 유의하지 않아 모형에서 제외되었다. Avegputt는 척도를 구간으로 하는 경우 계수를 이용한 반응변수에 대한 해석이 현실적이지 않아 사분위로 나누어 순서 형 변수로 적용하였다. SAS의 Proc Genmod를 이용한 결과는 표 3.1과 같다. 표 3.1의 결과를 보면 그린 적중 시 퍼팅수가 우승평균에 가장 영향을 많이 미치고 그 다음으로 숏게임 능력이 우승에 영향을 주는 것으로 나타났다. 일반적으로 그린 적중률이 상금과 평균타수에 가장 영향력 있는 변수로 나타난 것과 다른 발견이었다. 우승권에 있는 선수들은 실력이 우수하여 그린 적중률에 의해서는 우승 가능성에 대한 변별력이 없어 보인다.

표 3.1 포아송 회귀 모형결과

변수	자유도	추정치	표준오차	왈드 카이제곱	P-값
편차	1	-38.59	6.597	34.22	<.0001
Avedir	1	0.075	0.019	15.15	<.0001
Avegputt	1	0.422	0.181	5.40	0.0201
Scrratio	1	0.239	0.0449	29.56	<.0001
Scale	0	1.000	0.000		

표 3.2의 피어슨 카이제곱에 근거한 P-값이 0.2 이상이므로 모형의 적합도는 만족하다고 할 수 있으나, 예측 값과 실제 관측 값을 비교해 본 결과 0이 아닌 상위 값에 대한 오차가 많이 존재 하였다. 전반적으로 많은 관측 치에 0값이 존재하고 그 예측 값에 대한 오차는 적으나, 상위 값에 대한 오차가 많이 존재하여 우승 확률예측에 적합한 모형이라 할 수 없었다.

표 3.2 모형의 적합도 검정

항목	자유도	값	값/자유도
이탈도	147	104.451	0.710
척도이탈도	147	104.451	0.710
피어슨카이제곱	147	158.427	1.078
척도피어슨	147	158.427	1.078

3.2. 신경망 모형 분석

박우창 (2000)은 신경망은 단순히 복잡한 비선형 로지스틱회귀함수라 할 수 있으며 네트웍을 각각의 층 안에서 다른 층과 다른 단위로 나누어 많은 비선형 함수가 만들어 질 수 있고 특정 모형의 선택을 통하여 데이터에 적합 된다고 하였다. 은닉 층과 출력 층의 결합함수와 활성함수는 신경망모형의 가장 중요한 요소이다. Martignon (2005)는 신경망은 여러 가지 모형이 있으나 MLP이 가장 많이 사용되고 있으며 입력 층과 출력 층의 형태에 따라 ORBFEQ (Ordinary Radial Basic Function with equal Widths and Heights), ORBFUN (Ordinary Radial Basic Function with unequal Widths), NRBFEQ (Normalized Radial Basic Function with Equal Weights and Heights), NRBFEV (Normalized Radial Basic Function with Volumes), NRBFUN (Normalized Radial Basic Function with Unequal Weights and Heights), NRBFEH (Normalized Radial Basic Function with equal Heights and unequal Weights), NRBFEW (Normalized Radial Basic Function with equal Weights and unequal Heights) 등이 있다고 하였다. 본 연구에서는 2009년 PGA 10위권 경험자 리스트에 있는 151명 전원을 훈련자료로 이용 하였다. 특히 151명중 우승 경험이 있는 선수가 27명으로, 로지스틱 회귀에는 적합하

나 신경망을 이용하기에 충분히 큰 자료가 아니므로 분할하여 검증자료로 이용 할 수가 없었다. 훈련자료를 이용하여 위에서 언급한 NRBF (Normalized Radial Basic Function)의 5개의 신경망결과를 비교하여 적합도가 우수한 2개의 모형을 선택 하였고, 두 개의 모형에 MLP와 로지스틱 회귀모형을 추가하여 다시 적합도를 비교한 후, 마지막으로 새로운 2007년 PGA자료를 검증자료로 적용하여 최종 모형을 결정 하였다. 2008년 자료에는 Tiger Woods가 부상으로 중도에서 대회를 포기하는 이유로 우승횟수에 대한 분포가 2009년과 아주 다르게 보여서, 2007년 자료를 검증자료로 선택 하였다.

3.2.1. NRBF 모형비교

NRBF 신경망에는 은닉층이 한 개만 있으면 MLP과 비슷하나 은닉층에서 결합함수로 다음과 같은 원형기준 함수를 사용한다는 점이 다르다.

$H_j = \exp(\eta_j)$ 여기서
$\eta_j = \log(abs(a_j)) - b_j^2 \sum_i^p (w_{ij} - x_i)^2$
x_i 표준화 입력값, a_j 높이, b_j 편의, w_{ij} 중점, p 입력의 수

Sarma (2005)는 신경망모형에서 가장 적합한 모형 선택과정은 반응변수의 결과를 고려한 여러 NRBF모형을 설정 후 모형평가 항목의 통계량이 가장 작은 것을 택한다 하였다. 특히 NRBF의 5가지 모형 선택에서는 특정하게 어떤 경우에 어떤 모형이 적합하다는 규칙이 없기 때문에 실험 자료를 이용한 특정 항목의 오차통계량이 작은 NRBFEV, NRBFEQ을 택했다.

표 3.3 NRBF의 5가지 모형비교

	NRBFUN	NRBFEH	NRBFEV	NRBFEW	NRBFEQ
오분류율	0.17	0.18	0.16	0.17	0.17
AIC	152.32	149.27	143.98	146.33	143.82
SBC	224.74	212.63	207.35	212.71	201.15

3.2.2. MLP, 로지스틱 함수와 NRBEQ, NRBEV의 비교

표 3.3의 NRBF 아키텍처 5모형 중 에서는 NRBEQ와 NRBFEV의 결과가 다른 모형보다 AIC (Akaike Information Criterion), SBC (Schwarz Bayesian Criterion) 그리고 오분류율에서 우수하여 최종 모형후보로 선택 하였다. 오분류율은 반응변수가 범주형 자료인 경우 실제 값과 관측 값이 얼마나 일치하지 않는가를 비율로 나타낸다. 본 연구에서는 일반적으로 최종모형설정 시 비교하는 여러 통계량 중에서 AIC, SBC 그리고 오분류율을 모형 선택의 기준으로 사용 하였다. 이렇게 선택한 두 개의 모형과 신경망에서 가장 일반적으로 사용하는 MLP와 로지스틱 모형을 비교하였다.

표 3.4 NRBF 모형과 MLP, Logistic 모형비교

방 법	모 형	SBC	실험자료 오분류율	검증자료 오분류율
회귀	Logistic	210.765	0.166	0.262
신경망	NRBFEQ	208.558	0.179	0.277
신경망	NRBFEV	201.978	0.154	0.253
신경망	MLP	214.842	0.152	0.285

위 테이블에서 보듯이 NRBFEV 모형이 SBC통계나 오차율에서 가장 적합한 모형으로 나타났다. 본 연구의 목적이 우승자예측을 위한 모형개발이고 MLP 모형이 오분류비률에서 NRBFEV와 같이 우수

하므로 SAS E-MINER의 Lift Chart 윈도우의 적발반응 (Captured response) 그래프에 정확선 (exact)을 추가하여 두 모형의 반응도를 비누적 (non-comulative) 항목에서 비교하여 보았다. 적발반응은 각 등급에서 반응 개수 즉 1로 예측한수를 전체 반응 개수로 나눈 비율이다. 그림 3.1에 추가한 정확도는 실험 자료를 근거로 반응도를 계산한 것으로 MLP와 NRBFEV에 의한 그래프중 정확선과 가까운 값을 나타낼수록 바람직한 모형이라 할 수 있겠다.

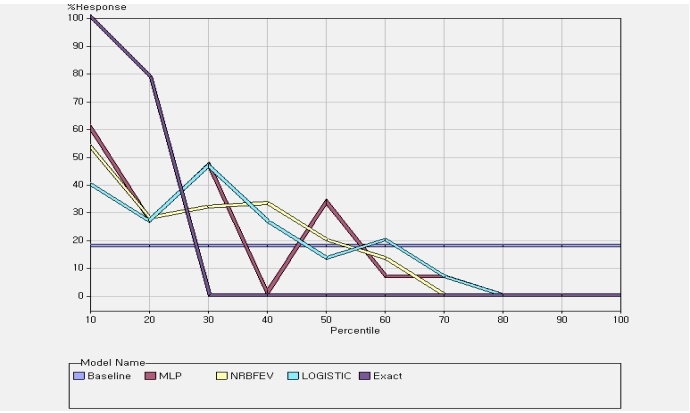

그림 3.1 각 모형의 반응도 비교

위 그림 3.1에서 보듯이 실험 자료에 근거한 정확도는 30%에서 0값에 이르며 상위 20%까지는 MLP구조 모형이 NRBFEV보다 우수 한 것으로 나타났으나 20%에서 30%까지는 NRBFEV가 정확도에 근사하게 나타났다. 상위권자의 우승확률에 관한 정확도에서는 MLP 모형이 더 적합한 것으로 보였다.

표 3.5 NRBEF와 MLP의 정확도비교

등급	합계	NRBFEV에의한누적정확도	NRBFEV에 의한정확도	MLP에 의한누적정확도	MLP에 의한정확도
10	15.1	29.63	29.63	33.33	33.33
20	15.1	45.19	15.56	48.15	14.81
30	15.1	62.96	17.78	74.07	25.93
40	15.1	81.48	18.52	74.07	0.00
50	15.1	92.59	11.11	92.59	18.52
60	15.1	100.00	7.41	96.30	3.70
70	15.1	100.00	0.00	100.00	3.70
80	15.1	100.00	0.00	100.00	0.00
90	15.1	100.00	0.00	100.00	0.00
100	15.1	100.00	0.00	100.00	0.00

그러나 2007년 PGA자료를 검증자료로 하여 모형에 적용한 결과 아래의 그림 3.2에서 보듯 NRBFEV모형에 의한 상위 20%에 대한 적중률이 MLP를 이용한 적중률보다 높게 나왔다. 그래서 최종적으로 적중률, AIC, SBC, 오차분류비율 등을 고려하여 NRBFEV모형을 예측모형으로 결정 하였다.

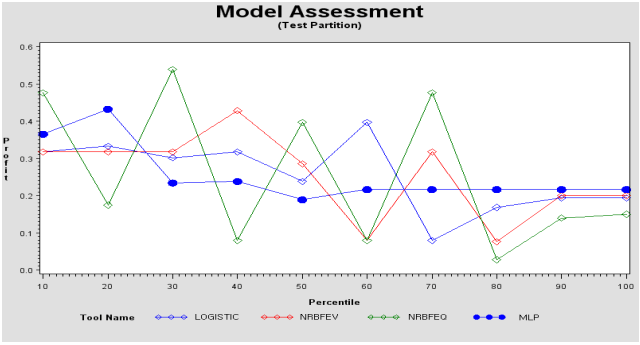

그림 **3.2** 검증자료를 이용한 모형비교

3.2.3. 최종 NRBFEV모형에 의한 분석결과

1개의 은닉층과 3개의 은닉마디를 가진 NRBFEV가 실행 된 결과는 아래와 같이 보여 질 수 있다.

$$H11 = f(1.76 - 0.20adedir - 0.09scrratio + 1.46avegrputt1 + 0.84avegrputt2 + 0.64avegrputt3)$$

$$H12 = f(1.87 - 0.68avedir - 0.54scrratio + 1.12avegrputt2 + 0.21avegrputt3)$$

$$H13 = f(1.52 + 0.10avedir + 0.73scrratio + 1.17avegrputt1 + 2.56avegrputt2)$$

$$P(Y = 1) = g(-0.72H11 - 4.68H12 - 0.53H13)$$

검증자료에 대한 예측은 로지스틱회귀나 모든 신경망 모형에서 훈련자료에 비하여 다소 적중률이 떨어졌다. 이는 신경망에 대한 과적합성 문제보다는 2007년의 우승에 기여하는 주요 골프기술의 기술통계량이 2009년과 다소 다른 특성을 지닌 것으로 생각된다.

4. 결론 및 기대효과

본 연구는 신경망을 이용한 PGA TOURNAMENT 우승확율 예측에 관한 것이다. 신경망은 로지스틱 회귀모형에 비하여 복잡한 비선형모형에 정확도가 뛰어나고, 특히 예측변수간의 상관도가 존재하여 다중 공선성이 있는 경우에 장점이 있어 이를 적용하였다. 신경망 모형 중 예측모형에 많이 적용되는 NRBF모형과 일반적으로 가장 많이 사용되는 MLP 모형을 적용하여 결과를 비교하였다. 본 연구에서는 특히 일반적인 적중률보다는 상위 그룹에 대한 적중률에 근거하여 최종 모형을 결정 하였다. 결론적으로 MLP 모형보다는 NRBFEV가 실험자료나 검증자료에 대하여 상위 그룹의 예측력이 안정성이 있었고 AIC나 SBC에서도 다른 모형보다 좋은 결과를 나타냈다.

참고문헌

강현철, 한상태, 최종후, 이성건, 김은석, 엄익현, 김미경 (2006). <데이타마이닝 방법론>, 자유아카데미, 서울

박우창, 승현우, 용환승, 최기헌 (2000). <데이터 마이닝 개념 및 기법>, 자유아카데미, 서울.

Berry, S. M. (2001). How ferocious is Tiger?. *Chance*, **14**, 51-56.

Cho, M. and Park, E. (2008). Analyzing customer management data by data mining: Case study on Churn prediction models for Insurance company in Korea. *Journal of Korean Data & Information Science Society*, **19**, 1007-1018.

Kim, K. and Lee, C. (2003). A study of data mining optimization model for credit evaluation. *Journal of Korean Data & Information Science Society*, **14**, 825-836.

Martignon, R. (2005). *Neural network modeling using SAS Enterprise Miner*, authorhouse, Cary, North Carolina.

Sarma, K. S. (2007). *Predictive modeling with SAS Enterprise Miner*, SAS Institute Inc., Cary, North Carolina.

Journal of the Korean Data & Information Science Society
2009, **20**(6), 1119–1127

한국데이터정보과학회지

Prediction of a winner in PGA tournament using neural network[†]

Daekee Min[1] · Moo Sung Hyun[2]

[1]Department of Information & Statistics, Duksung Women's University
[2]Department of Physical Education, Gangnam University

Received 10 September 2009, revised 20 November 2009, accepted 24 November 2009

Abstract

In PGA golf, total prize money and average score are good response variable related to golf skills such as driving distance, green in regulation and putts per green in regulation. But it's not easy to predict the winner of coming tournament. Thus I applied Neural Networks which has pretty good advantages for non-linear complex modeling to binary data. In neural network architectures, I applied NRBF and MLP architecture model for binary data which represent who had a win or not.

Keywords: Activation function, combination function, neural networks, Poisson regression.

[†] This research was supported by Duksung Women's University and Gamnam University Research Fund.
[1] Corresponding author: Associate Professor, Department of Information & Statistics, Duksung Women's University, Seoul 132-714, Korea. E-mail: dkmin@duksung.ac.kr
[2] Professor, Department of Physical Education, Gangnam University, Gyeonggi_do 446-702, Korea.