

스포츠 데이터 분석에 적용된 통계 및 데이터마이닝 기법에 대한 연구동향

박성건 · 박주영*
송실대학교

The Research Trend of the Statistical Analysis and Data Mining Techniques Applied to Sports Data Analysis

Park, Sung-Geon · Park, Jooyoung
Soongsil University

Abstract

The purpose of this paper is to analyze and compare statistical analysis and data mining techniques used for sports data analysis by investigating previous studies, and to provide the analyzed results to researchers who study sports data analysis. The analysis was performed based on 84 literatures collected from ACM Portal, IEEE xplore, Pubmed, and RISS for Higher Education. Statistics are used to standardize the records for performance evaluation of players and teams and to identify related factors of win and loss based on the statistical methods such as t-test, ANOVA, correlation, regression and so on, while data mining and machine learning algorithms are used to analyze unstructured data such as image and text, and to predict the outcome of games. In conclusion, sports data analysis has become one of the promising research fields as the technical convergence of sports and other studies progresses quickly. We predict that research and development related to sports data analysis will increase continually in the future.

Key words : Sport Data Analysis, Strategy and Tactics, Data Mining

I. 서론

스포츠 데이터 분석은 최근 스포츠 비즈니스 시장

에서 새롭게 발전하고 있는 분야이며, 기존의 체육학, 통계학, 컴퓨터공학, 미디어공학 등이 복합적으로 상호작용하는 융합적인 학문 분야이다. 스포츠는 모든

* dsqueen@ssu.ac.kr

학문, 기술, 산업과 융합이 가능한 특징을 가진 콘텐츠로서, 정보통신기술이 발전하면서 전문운동선수의 경기력과 관련된 시설 및 장비를 개발하고, 현장에 적용되는 사례가 증가하고 있다(박성진, 2013; 정지명, 이진희, 신선운, 2011).

스포츠 데이터 분석 연구 분야의 가장 큰 이슈는 ‘경기력 분석’이다. 경기력 분석은 해당 종목에 대한 이해와 충분한 경험, 그리고 과학적인 지식이 필요하다. Hughes(1986)의 연구에서 정의된 경기력 분석의 목적은 1) 기술적 평가, 2) 전술적 평가, 3) 움직임 분석, 4) 통계적 분석 정보의 도출이며, 경기력 분석을 통해 획득된 정보는 선수와 팀의 기술 및 전술적 평가, 움직임 분석을 위해 선수와 지도자에게 제공할 수 있다고 보고하였다(Hughes, 1988). 스포츠 현장의 지도자들과 스포츠과학자들은 경기력 분석의 정확성과 객관성을 확보하기 위해 해당 스포츠를 경험적인 요소로 관찰하고 있지만, 이러한 경험적 관찰에는 한계가 있다(Alamar, 2013). 다시 말해서, 경기력 분석의 목적은 선수의 기록 단축 및 승패와 직결된 요인들을 발견하여 경험적 요소의 단점을 보완해줄 수 있는 경기력 향상과 관련된 과학적인 연구를 수행하는 것이라고 정의할 수 있다.

경기력 분석은 크게 선수의 움직임에 대해 운동역학적으로 분석하는 것과 부호화를 통한 기록 분석을 하는 것으로 구분할 수 있다. 운동역학적 분석은 스포츠 동작을 수행할 때 발생하는 움직임의 문제점을 발견하여 부상의 잠재적인 위험 요소들을 기술적으로 정의하는데 활용된다. 부호화 분석은 경기 내용에 대한 효율적인 피드백 정보를 제공하기 위해 해당 스포츠의 경기력과 관련된 변인들을 ‘지표(indicator)’로 설정한 후, 정성적인(qualitative) 데이터를 정량적인(quantitative) 데이터로 변환하는 분석 방법이라고 정의할 수 있다(Hughes & Franks, 2008). 운동역학적 분석과 부호화 분석은 접근하는 방법에는 차이가 있지만, ‘경기력 향상을 위한 과학적인 분석의 추구’라는 궁극적인 목

적은 동일하다(O'Donoghue, 2009). Hughes와 Bartelt(2008)의 연구에서는 운동역학적 분석과 부호화 분석의 목적이 유사하기 때문에, 경기력 분석 시 운동역학과 부호화 분석을 복합적으로 수행한다면 보다 효과적인 분석이 가능하다고 제안하기도 하였다.

경기력 분석이 필요한 이유는 선수 및 지도자들이 경기가 끝난 후 경기에서 승리하거나 패배한 원인을 분석하고, 분석된 결과에 따라 적절한 피드백 제공이 가능해야 한다. Franks와 Miller (1991)의 연구에서는 국가대표급 축구 지도자들을 대상으로 경기(90분)가 끝난 직후 주요 경기내용의 기억 상태에 대해 실험한 결과, 경기 내용의 약 45%를 기억해내었고, 전반전(45분)이 끝난 직후 경기 내용의 약 59%를 기억하였다고 보고되고 있으므로(Laird & Waters, 2008), 경기력 분석은 선수들의 경기력 향상에 긍정적인 영향을 줄 수 있다고 판단된다.

최근, 경기력 분석과 관련된 연구를 수행하는 국제 학술단체 중 국제경기력분석학회(International Society of Performance Analysis in sports)에서 발표된 주요 연구 이슈는 1) 운동신경 조절, 2) 신경 네트워크, 3) 운동학습과 피드백, 4) 운동수행능력 평가(엘리트 선수와 심판원에 대한 분석), 5) 코칭 과정, 6) 코칭 행동, 7) 디지털 분석, 8) 스포츠 경기력 데이터베이스 관리, 9) 움직임 분석, 10) 부호화 분석, 11) 스포츠 소프트웨어 시스템, 12) 스포츠과학과 경기력이었으며(Alamar, 2013), 국제스포츠공학회(International Sports of Engineering Association)에서 제안하고 있는 연구 분야는 1) 스포츠장비 디자인 설계, 2) 실험 장비 개발, 3) 시뮬레이션 모델링 기법, 4) 개발된 장비의 성능 평가, 5) 선수 부상예방, 6) 경기력 향상과 관련 연구 분야이다. 즉, 스포츠 데이터 분석에서 다루고 있는 연구 분야는 전통적인 스포츠과학과 정보통신기술, 공학 분야가 결합된 형태의 다학제(Multidisciplinary) 또는 융합적인(Convergence)형태로 발전하고 있다(박성진, 2013).

표 1. 분석된 문헌

구분	연도	저자	문헌명	연구 방법
국내	1994	서재순	귀납적 추론을 이용한 프로야구 승패 예측시스템 개발에 관한 연구	추론학습 알고리즘을 보완한 이산화 알고리즘 Chimerge를 사용하여 두 팀 간의 승패 예측
	1998	허준희, 정태충	프로야구 경기 예측 시뮬레이터에서의 역전과 알고리즘을 이용한 투수 교체시기 예측 모듈 개발	역전과 알고리즘으로 투수 교체 시기를 예측
	1999	문성훈 등	내용 기반 스포츠 뉴스 비디오 시스템 설계 및 구현	자막 및 특징정보에서 추출한 시맨틱 정보를 반영한 내용 기반 스포츠 뉴스 비디오 검색 시스템을 설계
	2001	이기청	남자배구 선수들의 공격 수행 능력 평가	데이터베이스 구축 후 신경회로망 모델 분석
	2002	강상조	체육, 스포츠과학 분야의 학문적 성장: 통계적 방법 적용의 역사	체육, 스포츠과학에 통계 방법을 미국 전문학술지와 한국체육학회지를 비교
		엄한주 등	통계모형의 전문스포츠 현장 적용 사례	군집, 다중회귀, Markov로 승패 확률 분석
	2003	홍석미 등	혼합형 기계학습 모델을 이용한 프로야구 승패 예측 시스템	휴리스틱 함수의 혼합형 모델 사용
		오광모, 이장택	데이터마이닝을 이용한 한국프로야구 선수들의 연봉에 관한 모형 연구	신경망분석, 의사결정트리, 회귀분석을 사용
	2006	한필수	2006 겨울리그 데이터를 이용한 경기상황별 승·패 요인 분석	t-검정 후 집단별 승패 변인에 대한 선형식을 찾아 판별분석 실시
		이기봉, 이영석	한국과 덴마크 여자 핸드볼 팀의 공격 방법 비교	핸드볼 경기의 16가지 경기지표를 설정하고, 지표에 대한 빈도분포표를 제시
		이기봉, 이영석	한국과 덴마크 여자 핸드볼 팀의 공격 방법 비교	영상 자료를 분석하여 사용하여 경기력 항목의 단순 시도와 성공을 카운트
		이만규	세이버 매트릭스를 적용한 프로야구 타자의 경기력과 연봉과의 관계	OPS, ISO, SECA 등과 연봉의 빈도분석, 단순회귀, 중다회귀분석을 실시
	2007	김혜진	사회연결망(social network analysis: SNA)을 이용한 스포츠 경기분석	사회연결망 이론을 스포츠분석에 적용할 수 있는 방법과 선행연구결과들의 사례를 제시
		이경용	K-리그, 프리미어리그, 프리메라리가, 세리에A의 반칙유형 비교·분석	빈도, 교차, 카이검정을 실시하여 반칙 유형과 시간대를 분석
	2008	김세형 등	한국프로농구 경기기록 분석에 의한 승패결정 요인	승/패와 공격/수비 변인의 차이를 로지스틱 회귀 분석과 의사결정나무 분석 실시
		김채희, 윤형기	경기력향상을 위한 세계컬링대회 경기내용 분석	카이제곱을 사용하여 한국팀과 외국팀의 작전을 엔드별, 포지션별, 전/후반으로 비교
		송종우	한국 프로스포츠 선수들의 연봉에 대한 다변량적 분석	팀 기여도와 연봉과의 관계를 분석하기 위해 다중선형회귀모형, 트리모형을 이용
	2008	이순호	탁구 경기내용분석 프로그램 개발	탁구경기분석 소프트웨어를 개발 후 프로그램을 이용하여 경기분석
		이현재 등	실시간 축구 경기 데이터 분석 및 처리	연구자의 조작적 정의에 의해 축구 경기 데이터를 저장할 수 있는 소프트웨어 프로그램을 활용하여 경기요인의 빈도분석

	최재일, 정구영	여자프로골퍼의 스포츠 상해 분석	KLPGA에 등록된 선수 중 서울·경기 지역에 소속된 프로선수 220명을 대상으로 수집된 상해시기, 동작, 응급처치방법에 대해 교차분석	
2009	김민철	시계열분석을 통한 프로야구 관중현황 예측모델연구	입장관객 수를 예측하기 위해 시계열분석 실시	
	김현경	아시아 주니어 컬링 여자선수의 경기기술 분석	엔드별, 포지션별, 선/후공별, 전/후반으로 구분하여 분석하기 위해 빈도분석 실시	
	천영진	남자 프로배구 경기에서 득점 과정을 이용한 경기 분석	빈도 차이를 알아보기 위해 카이제곱 실시	
	김민철	시계열분석을 통한 프로야구 관중현황 예측모델연구: 부산롯데자이언츠를 대상으로	1928-2008년까지의 롯데 정규리그 홈경기 입장객을 대상으로 시계열 분석을 통해 향후 5년간의 홈관중 수 예측	
국내	2010	김세중 등(a)	데이터 마이닝 기법을 이용한 한국 프로농구 경기 분석	CART 알고리즘을 사용하여 경기승패에 미치는 요인을 분석
		김세중 등(b)	한국프로농구경기 2000~2009시즌 PLAY-OFF 진출 팀의 경기 자료 데이터를 이용한 경기 분석	플레이오프에 진출한 팀 승패 패턴, 요소 분석
		강우용 등	탁구경기 내용분석을 위한 득실점 유형분석 및 체계화	공격, 실책, 경기외적 요인을 구분하여 빈도분석
		안재석	K리그와 EPL의 지역별점유율과 공격방향 분석을 통한 경기력 향상에 관한 연구	기술통계 및 비모수통계(Mann-Whitney U)
		이중철	웹 기반 스포츠 기술지도 시스템 모형 개발	웹 기반 경기영상을 활용한 상담-지도 모형 제시를 통해 효율적인 피드백 제공 시스템 개발에 관한 연구사례 제시
	2011	천영진, 김규태	캘리포포인트제에서 여자 프로배구 득점 내용 및 추이 분석	득점 내용과 득점 추이 카이제곱 검정 실시
정호석 등		웹 캐스트와 텍스트 마이닝을 이용한 축구 경기의 심층분석	텍스트 마이닝의 연관규칙, 성장도 분석, 패스파인더 네트워크 분석 기법을 사용	
김기진 등		국가대표 마라톤선수의 도로러닝시 운동생리학적 지표 분석	마라톤 선수의 훈련 전·후 혈액 변인 변화 Independent t-test와 paired t-test를 실시	
이강현 등		탁구경기 내용분석기법의 적용 및 활용방안 탐색	경기의 승/패배, 득/실점으로 구분하여 시도횟수, 성공횟수를 백분율(%)로 제시	
2012	강선영 등	제10회 밴쿠버 장애인동계올림픽 참가 대한민국 선수의 스포츠상해 실태 분석	밴쿠버올림픽 참가선수의 대회기간 중 1,000회 경기 및 훈련시간 당 상해 건수 조사	
	이규민 등	2010년 남아공월드컵 조별 예선경기에서 나타난 슈팅요인 결과 분석	PASW를 사용하여 교차분석을 실시	
국외	1971	Gale, D.	Optimal strategy for serving in tennins	테니스 서브 득점을 높이기 위한 전략을 확률을 기반 분석 방법 제시
	1991	Franks, I.M. & Miller, G.	Training coaches to observe and remember	영상 정보를 활용하여 경기 종료 후 경기내용의 기억 여부 집단 비교 실험 연구
	1993	Ghosh, A & Steckel, J.	Roles in the NBA: there's always room for a big man, but his role has changed	농구의 각종 지표에 대한 통계적 분석을 통해 센터의 역할 변화에 대해 연구
		Ghosh, A & Steckel, J.	Roles in the NBA: there's always room for a big man	NBA 센터 포지션의 선수의 기록(득점, 리바운드, 파울 등)을 시계열로 분석한 연구

	McGarry & Frank	A Stochastic approaches to predicting competition squash match play	Markov모델 기반 스쿼시 선수의 행동을 분석 및 예측하기 위한 연구
1994	P. K. Korkia et al.	An epidemiological investigation of training and injury patterns in British triathletes	1990년 시즌, 영국 트라이애슬론 선수들의 훈련, 시합의 총 거리와 부상경험과의 관계를 역학적으로 조사한 연구
	D'Ottavio & Gastagna	Physiological load imposed on elite soccer referees during actual match paly	이탈리아 세리에-A의 프로축구선수 대상으로 축구 경기시 심박수, 운동량을 Oshashi (1988)등의 측정방법을 사용하여 분석
2001	Murray & Hughes	Tactical performance profiling in elite level senior squash	스쿼시 코트 분석 소프트웨어 프로그램을 설계하여 경기 시 공의 접촉면을 분석
	O'Donghue & Ingram	A notational analysis of elite tennis strategy	테니스 코트의 면적, 성별에 따라 경기 중 발생하는 랠리 횟수의 차이를 통계 검증
	O'Donghue, P.G.	Performance models of ladies and men's singles tennis at the Australian Open	테니스 경기에서 승리하기 위한 전략의 남녀 차이를 통계적으로 비교 분석
2002	Edelmann-Nusser et al.	Modelling and prediction of competitive swimming performance in swimming upon neural networks	2000년 시드니 올림픽 200m 배영 여자 결승전 경기를 인공신경망 분석으로 경기결과를 예측
	Sampaio & Janeira	Importance of Free-Throw Performance in Game Outcome During the Final Series of Basketball Play-offs	경기 수준을 Superior, Intermediate, Inferior 3개로 구분하여 승패 팀의 자유투 성공률 two-way with ANOVA 분석
2003	Magni. M et al.	Match performance of high-standard soccer players with special reference to development of fatigue	Yo-Yo 간헐적 회복 테스트 실험에 대해 t-test를 사용하여 선수회복능력 집단 평가
2004	Andersen et al.	Mechanisms of head injuries in elite football	노르웨이와 아이슬란드 프로리그에서 발생한 머리 부상에 대해 1000시간 동안 경기당 부상 발생 횟수, 부위 등을 빈도분포로 제시
	Maffulli et al.	Long term sport involvement and sport injury rate in elite young athletes	훈련이나 시합시 두통을 호소한 경험을 로지스틱 회귀분석을 사용하여 분석
2005	McCrory et al.	Prevalence of headache in Australian footballers	호주 축구선수(남,여)들의 두통 경험 여부를 설문지로 자료를 수집하여 비교 분석
	Ortega, E. Cardenas et al.	Differences Between Winning and Losing Teams in Youth Basketball Games (14-16 Years Old)	청소년 농구 경기에서 승/패 선수들의 기술과 팀 전술 차이를 비교하기 위해 사전에 정해진 지표를 기반으로 횟수, 성공률 분석
	Martin, L.	Modeling the interaction in game sports -relative phase and moving correlations.	핸드볼 경기에서 볼의 획득 시간과 득점과의 상관관계, 득점을 높이기 위한 확률을 분석
2006	McGarry	Identifying patterns in squash contests using dynamical analysis and human perception, International Journal of Performance Analysis of Sport	스쿼시 선수들의 움직임의 공간 활용도를 기준으로 비교 분석
	Tscholl et al.	Causation of injuries in female football players in top-level tournaments	2003 피파 여자월드컵 경기 영상과 기록을 통해 부상발생과 테클과의 상관관계 분석
	Patrick E. McSharry	Altitude and athletic performance (statistical analysis using football results)	고도에 따른 경기력과 경기결과를 통계적(선형회귀, 포아송 분포)으로 분석
2007	Lloyd Smith et al.	Data Mining in Sports: Predicting Cy Young Award	1967-2006년 MLB 투수 데이터 기반 베이지안 분류 모델을 이용 사이영상 수상자 예측

2008	Helena Andersson et al.	Elite football on artificial turf versus natural grass: Movement patterns, technical standards, and player impressions	인공 및 천연잔디에서 축구선수의 움직임, 볼 트래킹 등 특징을 비교 분석
	Shafizadeh, M.	Qualitative analysis of aggressive behaviors in the adolescent, youth and adult soccer World Cups	체질량, 연령, 경기출전수, 포지션, 경기라운드 등의 변수 경로분석, MANOVA 통계 분석
2009	Johnson & Doherty.	Investigation of growth, development, and factors associated with injury in elite schoolboy footballers	맨체스터-U 유소년 축구선수들을 대상으로 선수들의 골연령과 부상과의 관계를 연구
2010	Nora Wiiuma et al.	Does Relative Age Effect Exist among Norwegian Professional Soccer Players?	노르웨이 프로축구선수들의 연령과 부상, 부상시기의 관계 분석
	Sun Jing et al.	Study of Association Rule Mining on Technical Action of Ball Games	Apriori 알고리즘 미들웨어를 사용하여 테니스 선수의 기술 및 전술적 움직임을 분석
	Z. Ivanković et al.	Analysis of basketball games using neural networks	세르비아 남자 농구리그 2005-2010시즌 890경기를 대상으로 신경망 분석 실시
2011	Slavko Roganae et al.	Position-specific and Team-ranking-related Morphological Characteristics in German Amateur Soccer Players - a Descriptive Study	독일 아마추어 축구선수들의 신체적 특징, 포지션 위치, 리그 랭킹과의 상관관계 분석
	Dvorak J et al.	Injuries and illnesses of football players during the 2010 FIFA World Cup	2010 월드컵 32개 팀 선수를 대상으로 질병, 부상 발생 횟수와 부위, 내용 빈도분석
	Bayraktar B et al.	Injury evaluation of the Turkish national football team over six consecutive seasons	2000~2005년 터키 축구 국가대표팀의 경기, 훈련 시간에 따른 부상발생 횟수와 종류 분류, 원인을 훈련시간과 경기당 발생 빈도로 제시
	Gustavo Gonçalves Arian et al.	The Brazilian Football Association (CBF) model for epidemiological studies on professional soccer player injuries	훈련시간에 따른 부상발생 여부와 회복에 소요된 시간들을 의료기관의 데이터베이스를 기반으로 기술통계 분석
2012	Pedro, O.S. Vaz De Melo et al.	Forecasting in the NBA and Other Team Sports: Network Effects in Action	MLB, NBA를 대상 네트워크 기반 예측모델을 이용해 기존 예측정확도를 14% 개선
2013	Ganeshpillai & Guttag	A Data-driven Method for In-game Decision Making in MLB	MLB 2006-2010 정규 및 포스트 시즌 경기 데이터 (투구 속도, 스트라이크, 볼, 홈런, 타격, 볼넷 등) 기반 선행회귀모형 모델 제시
	Sung Geon Park & Soowon Lee.	Curling analysis based on the possession of last stone per end	이분형 로지스틱 회귀분석(종속: 경기결과, 독립: 엔드별 선 · 후공 결과)을 실시하여 후공이 경기결과에 미치는 영향 분석

효율적인 스포츠 데이터 분석을 연구하기 위해서는 관련 선행연구들을 대상으로 연구방법에 대한 분석이 필요하다. 선행연구 문헌에 대한 리뷰(review)는 기존 연구에 대한 비판적인 능력 향상과 새로운 연구자(대학원생 등)들의 창의적인 연구 설계에 긍정적인 영향을 주어 그 중요성이 강조되고 있지만, 스포츠 데이터 분석과 관련된 선행연구에 대한 국내 논문은 매우 부족하다.

따라서, 본 연구의 목적은 스포츠 데이터 분석과 관련된 선행연구 리뷰(review)를 통해 스포츠 데이터 분석을 위해 사용되는 통계 및 데이터마ining 기법을 비교·분석하여 연구자들에게 스포츠 데이터 분석과 관련된 기초정보를 제공하는데 있다.

II. 연구방법

1. 자료수집

본 연구에서는 스포츠 연구 분석을 위해 통계 분석 및 데이터마이닝 기법을 사용한 스포츠 관련 국내·외 논문 70편(국내: 35편, 해외: 35편), 문헌 14편(국내: 4편, 해외: 10편) 총 84편(국내: 39편, 해외: 45편)을 분석하였다. 자료수집은 ACM Portal, IEEE xplore, Pubmed, RISS for Higher Education을 이용하여 수행하였으며, 수집된 문헌은 국내/외, 저자명, 논문명, 데이터분석 방법, 연구결과로 구분하여 <표 1>과 같이 정리하였다.

스포츠 데이터 분석에서 다루는 연구 분야는 매우 광범위하기 때문에, 본 연구에서는 1) 스포츠 데이터 분석의 개념 및 연구 분야, 2) 통계 및 데이터마이닝 연구 적용사례 비교, 3) 스포츠 데이터 분석의 연구동향 등 세 가지의 카테고리로 구분하여 선행 연구사례 및 연구동향을 제시하였다.

III. 연구결과

스포츠 데이터 분석은 선수의 행동을 설명하거나 승패와 관련된 요소들을 발견하기 위한 수단으로 활용되고 있다(Lames et al., 2009; Martin, 2006; McGarry, 2006). 또한, 영상과 경기 데이터를 기반으로 분석된 정보는 선수 개인의 운동능력 향상을 도와줄 뿐만 아니라 선수단 전체의 경기력 향상에 기여한다고 알려져 있다(Franks, 1997; Greene, 2008).

일반적으로, 스포츠 데이터 분석 방법은 기록 방식(notational or computing), 경기력 지표(performance indicators), 해당 종목의 특성(individual or team)에 따라 달라진다. 데이터 수집 및 분석 시기는 경기 전



그림 1. 스포츠 데이터 분석 연구 분야의 분류

(before) · 중(playing) · 후(after)에 진행되며, 수집 및 분석된 정보는 사용자의 특성(선수 및 지도자, 방송관계자, 시청자 및 스포츠 팬 등)에 따라 해당 콘텐츠로 제작되어 제공된다(Alamar, 2013). 스포츠 데이터 분석은 야구와 축구를 중심으로 발전하였으며, 우수한 선수를 발굴 및 평가하거나, 스타 선수 및 팀을 비교하기 위한 기준으로 사용되고 있다(O'Donogues, 2009). 야구 데이터 분석의 경우, 1850년대~1900년대 미국 최초의 야구 기자(Henry Chadwick)가 야구경기를 기록하면서 시작되었다고 알려져 있지만, 1920년대까지 데이터 분석은 이루어지지 못하였다. 그 후, 여러 학자 및 분석가에 의해 야구 기록을 통계적으로 분석할 수 있는 모델들이 제시되기 시작하였다. 야구 저술가이자, 통계학자인 빌 제임스(George Williams Bill James)는 좌/우 투수의 성적 비교, 홈 경기와 원정 경기의 승률, 주자 상황에 따른 타자 성적 등과 같은 요인들을 수학적 모델로 표현할 수 있는 야구 경기력 지표인 '세이버메트릭스(Sabermetrics)'를 개발하여 현장에 적용하였다. 이러한 수학적 모델은 IT기술과 결합하면서 스포츠 현장에 과학적인 분석이 가능하도록 이끌었다. 과학적인 방법에 의해 분석된 정보는 유망주 발굴, 데이터 기반의 선수 연봉 산정, 선수 및 팀의 작

전분석을 통한 감독의 의사결정 등 다양한 분야에 활용되고 있다(Cost et al., 2009).

축구 데이터 분석은 1960년대부터 부호화 분석방법을 사용하기 시작하였는데, Reep과 Benjamin(1968)은 1953년부터 1968년까지 수집된 3,213개의 축구 경기 데이터를 실증적인 방법(empirical methods)을 이용하여 패스(passing)와 쏘트(shooting)의 횟수와 골(goal)과 관계에 대해 분석하였다. Reilly & Thomas (1976)는 경기 중 선수들의 활동지역과 움직임 강도를 분석하였으며, Reilly (1977)는 축구 경기에서 요구되는 생리적인 변인들을 정의하여 수집된 데이터를 데이터베이스(DB)로 구축하였다. Hughes와 Colleagues(1988)는 Church와 Hughes(1986)에 의해 개발된 시스템을 사용하여 1986년 월드컵 결승전에서 승리한 팀과 패배한 팀의 경기 패턴에 대해 연구한 결과, 승리한 팀은 미드필더 지역에서의 활동이 많았고, 패배한 팀은 수비 지역에서의 볼 소유 시간이 많고 선수들이 패스보다는 드리블을 하는 횟수가 승리한 팀보다 높게 나타났다고 보고하였다. 축구 데이터 분석에 사용되는 지표는 볼 점유율, 슈팅, 패스, 파울, 득점, 경고/퇴장 개수 등이며, 최근에는 영상 트래킹 시스템을 활용하여 선수의 움직임 패턴과 활동 정보(거리, 운동량)를 분석한 정보를 추가적으로 제공하고 있다.

스포츠 데이터 분석의 연구 분야는 <그림 1>과 같이, 학문적 관점에 따라 세 가지 카테고리(공학적 관점) 또는 네 가지 카테고리(스포츠과학적 관점)로 구분될 수 있다. 공학적 관점의 스포츠 데이터 분석은 1) 구조화된 데이터의 관리, 2) 데이터를 활용하기 위한 예측 분석 모델, 3) 의사결정을 위한 정보 시스템 개발로 세분화될 수 있다(Hughes & Barlett, 2008; Schumaker et al., 2010; Solieman, 2006).

첫째, 구조화된 데이터 관리는 팀의 중간 관리자를 위해 선수의 의료 및 경기 데이터를 포함한 모든 데이터를 효율적으로 관리하기 위한 방법에 대해 연구하는 분야이다. 데이터 관리자에게 제공되는 주요 정

보는 데이터 관리를 위한 시설 구축 및 유지보수에 필요한 비용은 얼마나 소요되는지, 수집된 정보를 이용하여 어떤 정보를 제공해야 하는지, 그리고 정보의 보안을 위한 시스템 구축을 어떻게 해야 하는가에 대한 기술적인 부분이다(O'Donoghue, 2009).

둘째, 데이터를 활용하기 위한 예측 분석 모델은 프로팀들이 유망주를 발굴하거나 우수한 선수를 선발하기 위해 연구하는 분야이다. 예측 분석 모델은 수집되는 데이터의 양(volume)과 질(quality)에 따라 모델의 정확성과 효율성에 차이가 발생할 수 있기 때문에 수집되는 데이터가 매우 중요하여 데이터베이스 구축 기술과 연관되어 있다(Hughes & Franks, 2008).

셋째, 의사결정을 위한 정보 시스템은 효율적인 구단 관리를 위해 필요한 각종 통계정보(관중 수입, 기념품 판매, TV 중계권료, 광고 등에 의한 구단 총 수입과 선수단 연봉 등의 구단 운영을 위한 총 지출 금액), 구단 경영과 관련된 모든 정보를 관리할 수 있는 시스템을 개발하거나 관련 기술을 연구하는 분야이다(Alamar, 2013).

스포츠과학적 관점의 스포츠 데이터 분석은 1) 코칭 행동 분석, 2) 기술적 효과 분석, 3) 전술 분석, 4) 움직임 분석으로 세분화된다(Gilbert & Trudel, 2004; Hughes & Probert, 2006).

첫째, 코칭 행동 분석은 지도자의 지식 함양, 경기 현장에서의 의사결정 지원, 훈련 프로그램 개발 등에 도움을 주기 위한 방안을 연구한다(Gilbert & Trudel, 2004; Kahan, 1999).

둘째, 기술적 효과 분석은 경기에서 승리하기 위해 어떤 기술이 적용되어야 하는지, 또는 선수의 기술적인 숙련도를 객관적으로 평가하기 위한 방법에는 어떠한 것이 있는지 등을 연구한다. 예를 들면, 축구 경기에서 기술적인 효율성을 평가하기 위해 정확도를 - 3점에서 3점의 양적 지표로 정의하여 정량적으로 평가하는 것이다(Hughes & Probert, 2006).

셋째, 전술 분석은 전략과 작전 분석을 포함하는데

전략(strategy)은 선수 또는 팀의 강점을 최대한 활용하고, 단점을 보완할 수 있는 계획을 수립하는 것이며, 작전(tactics)은 경기 중 발생하는 각 상황에 대처하기 위해 수행하는 선수 또는 지도자가 의사결정을 지원하는 것을 의미한다. 전략은 경기 전과 후에 감독 또는 전략분석관에 의해 진행되는 반면에, 작전은 경기가 진행되는 동안 지도자 또는 선수에 이루어진다(Fuller & Anderson, 1990). 예를 들면, 축구의 공격성과 경기력과의 연관성을 설명하기 위해 심판의 카드 횟수와 파울 수를 분석한 것은 전술 분석과 관련된 연구인 것이다(Shafizadeh, 2008).

넷째, 움직임 분석은 해당 스포츠를 수행하는데 필요한 선수의 신체적 능력과 부상의 위험을 인지하거나 이와 관련된 부상 발생 원인을 연구하는 분야이다(O'Donoghue, 2008). Lee(2008)의 연구에서는 효율적인 움직임 분석을 위해 단순 반복적인 동작인 마이너 기술(Minor Skill: 달리기, 점프), 경기결과와 직접적인 연관이 있는 기술을 반복적으로 수행하는 동작인 메이저 기술(Major Skill: 골프 스윙), 한 종목을 수행하기 위한 전체의 기술인 이벤트 기술(Event Skill: 멀리뛰기)로 구분하여 부상의 발생 원인을 세분화된 선수의 움직임 분석을 통해 가능하다고 제안하기도 하였다.

결과적으로, 스포츠 데이터 분석은 '기록 단축 및 기록의 향상, 그리고 경기에서 승리하기 위한 코칭 행동, 선수의 움직임 분석을 통해 자신의 강점은 극대화하고, 상대의 강점을 무력화시킬 수 있는 방법을 과학적으로 접근하고 연구하는 분야'라고 정의할 수 있다. 또한, 스포츠 데이터 분석과 관련된 연구 분야는 '경기력 향상'과 관련된 장비의 설계 및 개발, 과학을 통한 훈련 방법의 개선을 위해 전통적인 스포츠과학에서 정보통신, 공학 분야가 결합된 형태인 다학제 또는 융합적인 연구 분야라고 정의할 수 있다.

2. 통계 및 데이터마이닝 연구적용 사례 비교

스포츠 데이터 분석을 위해 통계적 분석과 데이터마이닝 기법이 활용되고 있는데, 스포츠 통계는 선수와 팀을 평가하고, '경기력 향상'과 관련된 변수를 찾거나, 어떤 특징을 가진 집단을 비교하는 것을 목적으로 한다(엄한주, 조정환, 신승윤, 2002). 스포츠 데이터 분석을 위한 통계 기법은 연구 목적과 가설에 따라 다르게 적용된다. 특성이 다른 두 모집단의 평균을 비교할 때는 독립표본(independent) t-test를 사용하고, 동일한 집단에 속한 피검자들을 반복적으로 측정한 자료의 평균을 비교할 때는 종속표본 t-test를 사용한다. 예를 들면, 어떤 지도자가 새로운 훈련방법을 제안하였다고 가정하고, 그 훈련의 효과를 검증하고자 할 때는 paired t-test가 독립표본 t-test를 적용하는 것보다 적합하다고 볼 수 있는데, paired t-test는 사전·사후 측정을 통해 처치효과(treatment effect)를 검증할 수 있기 때문이다.

분산(variance)은 집단의 평균으로부터 거리가 얼마나 떨어져 있는지를 나타내는데, 종속변수와 독립변수가 각각 한 개일 때 일원분산분석(One-way ANOVA), 종속변수가 한 개이고 독립변수가 두 개이면 이원분산분석(Two-way ANOVA)을 사용한다. 어떤 연구자가 '알코올 섭취가 타자의 반응시간에 미치는 영향'에 대해 검증하고자 할 때 독립변수는 알코올 섭취(1개), 종속변수는 반응시간(1개)이기 때문에 일원분산분석을 사용하는 것이 적합한 분석방법이 된다(Williams & Wagg, 2004).

회귀분석(Regression)은 한 변수를 종속변수로 다른 변수를 독립변수로 설정하여 이들 간의 관계를 분석($Y = a + \beta X_i + \epsilon_i$)하는데, 스포츠 데이터는 대부분 2개 이상의 독립변수가 사용되기 때문에 다중회귀분석(multiple regression)을 사용하게 된다. 예를 들어, '프로야구선수의 경기력(8개의 항목)이 뛰어들수록 선

수연봉은 높아질 것이다.’ 라는 연구가설을 수립하고, 그 연구가설을 검증한다고 가정할 때 연봉은 종속변수가 되고, 선수경력, 지난 시즌 타점, 타율, 출루율, 도루, 출전 경기 수, 홈런, 타율은 독립변수가 된다. 이를 회귀모형으로 표현하면 다음과 같다(이만규, 2006).

$$Y = a + \beta_1 X_1 (\text{경력}) + \beta_2 X_2 (\text{타점}) + \dots + \beta_8 X_8 (\text{타율}) + \epsilon_i$$

또한, 회귀분석에서 종속변수가 두 범주로 구성된 명목척도(nominal scale)인 경우에는 로지스틱 회귀분석(logistic regression analysis)을 사용하게 되는데, 로지스틱 회귀분석의 목적은 어떤 사건이 발생할 확률을 예측하고자 할 때 사용되기 때문에, 결과를 해석할 때 주의해야 한다. Sung Geon Park과 Soowon Lee(2013)의 연구에서는 컬링 경기에서 엔드별(1~10 엔드) 선공(first stone) 및 후공(last stone)과 경기 결과(승·패)와의 관계를 알아보기 위해 종속변수는 경기결과, 독립변수는 엔드별 선·후공의 결과로 설정하여 이분형 로지스틱 회귀분석(binary logistic regression)을 실시한 결과, 컬링 경기의 1엔드와 10엔드 후공을 획득한 팀이 선공인 팀보다 경기에서 최종적으로 승리할 확률이 높게 나타났다.

이러한 통계적 분석 방법들은 연구자가 사전에 조작적 정의를 통해 경기결과와 연관된 경기력 지표를 설정하고, 경기력 지표의 발생 횟수와 성공 횟수를 기록하여 성공률 등을 승리한 팀과 패배한 팀을 비교·분석하거나(강우용, 2010; 김현경, 2009; 이기봉, 이영석, 2008; 한필수, 2006), 유망주 발굴과 우수한 선수를 판단하기 위해 기록을 표준화하여(Thorn & Palmer, 1985), 선수가 공격적 또는 방어적인 플레이 유형인지를 파악 및 평가함으로써(Ghosh & Steckel, 1993; O'Donoghue & Ingram, 2001), 팀 승리에 긍정적인 영향을 주는 요인을 분석하여 선수 및 지도자들에게 피드백을 제공할 수 있는 방법(Csataljay et al., 2008; Hughes

& Clarke, 1995; Nora Wiiuma et al., 2010; Pedro et al., 2012) 등을 발견할 수 있는 수단으로 활용 가능하다. 또한, 프로팀들과 연구자들이 선수의 부상예방 및 발생 원인에 대한 특징을 파악하기 위한 공동연구사례는 주로 해외 연구결과를 통해 보고되고 있다(Andersen et al., 2004; Johnson & Doherty., 2009; McCrory et al., 2005; Nora Wiiuma et al., 2010).

데이터마이닝(data-mining)은 데이터로부터 유용한 정보를 추출하는 것을 연구하는 통계학(statistics)과 인공지능(artificial intelligence)의 합류점에 위치하고 있는 학문 분야이다(Pregibon, 1999). 스포츠 데이터 분석에서 데이터마이닝이 적용되는 주요 목적은 1) 향후 경기력 예측, 2) 훈련 방법의 장·단점 파악, 3) 경기력 개선 효과의 평가, 4) 코칭과정의 지원, 5) 동기부여 제공 등이다(Leonardo De Marchi, 2011).

인공지능(Artificial Intelligence)은 판단, 추론, 학습 등 인간의 뇌 기능을 갖춘 컴퓨터 시스템을 연구하는 분야인데, 인공지능은 기계학습, 음성인식, 자연어 처리, 컴퓨터 비전(영상분석 시스템)등을 포함하고 있다(Bartlett, 2004). 대표적인 인공지능 기법으로는 전문가 시스템(Expert systems)과 인공신경망(Artificial neural networks: ANNs)이 있다(Negnevsky, 2002). 전문가 시스템은 어떤 분야의 전문가가 처리하던 일을 컴퓨터가 대신 처리하는 기술이다. 예를 들면, ‘조금 더 가까이 패스해’ 라는 말은 수치화된 데이터가 아니기 때문에 컴퓨터가 처리할 수 없지만, 규칙 기반 전문가 시스템은 이러한 문제들을 해결해줄 수 있는 기술이라고 이해하면 된다. 인공신경망(ANNs)은 인간의 뇌 기능의 특성을 컴퓨터로 표현할 수 있는 수학적 모델인데, 축구 경기와 같이 복잡한 정보들을 분석하거나 경기결과를 예측할 때는 인공신경망 분석이 사용된다(O'Donoghue & Williams, 2004). 또한, 많은 선행 연구에서는 신경망 분석 기법인 Kohonen Feature Map(KFM), 패턴인식, 상황추출 등을 사용하여 축구에서 득점력이 높은 패턴을 찾거나 럭비 경기의 승패 요인을 추출하는

연구를 수행하기도 하였다(Bauer & Schöllhorn, 1997; Lees et al., 2003; Lees & Barton, 2004; Perl, 2001; Yan & Wu, 2000).

기계학습(Machine Learning)은 인공지능의 한 분야로 알고리즘(algorithm)을 사용하여 주어진 데이터를 컴퓨터에 입력하여 학습시킨 후 새로운 데이터가 주어졌을 때 데이터를 판별(분류)하거나 예측할 때 사용되는 데(Schumaker et al., 2010), 지도학습(supervised learning)과 비지도학습(unsupervised learning)방법으로 구분된다. 지도학습은 주어진 데이터를 바탕으로 회귀모형을 만들고, 만들어진 회귀모형에서 새로운 데이터를 처리하는 과정으로 이루지는 반면에<그림 2, 표 2>, 비지도학습은 데이터를 주고 컴퓨터가 그 규칙성을 직접 찾아가도록 두는 방법론의 차이가 있다<그림 3>.

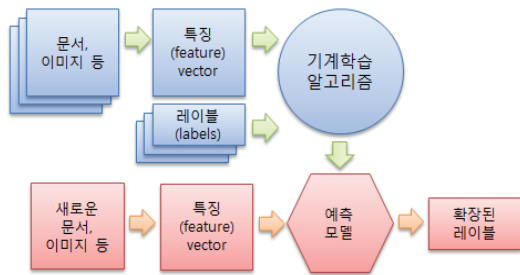


그림 2. 지도학습 모델(supervised learning model)

표 2. 지도학습 모델의 평가 방법

실제 결과분류			
		참	거짓
추론결과 /분류	참	tp(true positive)	fp(false positive)
	거짓	fn(false negative)	tn(true negative)

$$\text{precision} = \frac{tp}{tp + fp} \quad \text{recall} = \frac{tp}{tp + fn}$$

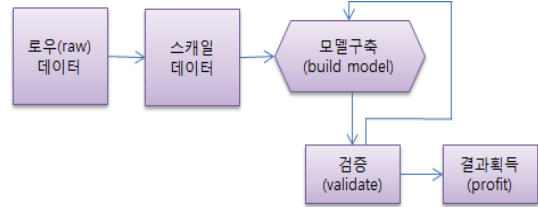


그림 3. 비지도학습 모델(unsupervised learning model)

스포츠 데이터 분석에서 데이터마이닝 기법이 적용된 국내 연구사례를 살펴보면, 휴리스틱(heuristic)과 역전파 알고리즘(back-propagation algorithm), 의사결정 나무 모형 등을 이용하여 경기결과(승·패)를 분석 및 예측하거나(김세중, 허종관, 도재현, 2010a; 김세중, 허종관, 이강웅, 2010b; 김세형, 강상조, 박재현, 김혜진, 2008; 서재순, 1994; 홍석미, 정겨숙, 정태충, 2003), 특정 포지션(야구의 투수)의 경기 중 적절한 투입 시기를 예측한 결과를 제시하였고(허준희, 정태충, 1998), 신경망을 이용한 분석 모델을 제시하기도 하였다(오광모, 이장택, 2003; 이기칭, 2001). 또한, 정호석, 이종욱, 유재학, 이한성, 박대희(2011)는 축구의 인터넷 문자 중계정보와 도메인 지식을 기반으로 연관규칙 마이닝, 성장도 분석, 패스파인더 네트워크 분석 기법을 이용하여 2010년 남아공 월드컵대회 경기 중 스페인 국가대표 경기를 분석하였다.

스포츠 데이터 분석에서 데이터마이닝 기법이 적용된 해외 연구사례를 살펴보면, 데이터베이스를 구축하여 해당 스포츠의 경기지표를 사전에 설정한 후, 경기 지표에 대한 통계 정보를 기반으로 선수의 행동반경 및 동작, 전술적 움직임을 분석하거나 예측하기도 하였다(Choi et al., 2006; McGarry & Frank, 1993; Murray & Hughes, 2001). 또한, 훈련시간에 따른 부상발생 여부와 회복에 소요된 시간들을 연구하기 위해 관련 의료기관의 데이터베이스를 활용한 연구사례가 있다.

결과적으로, 통계학은 데이터로부터 두 변수간의 관계를 평가하는 것에 대해 중점을 두고 있는 반면, 데이터마이닝은 통계학에서 처리하지 못하는 텍스트,

영상 등 비정형화된 데이터를 분석하고, 기계학습을 통해 예측 모델을 제시한다는 점에 차이가 있다. 다시 말해서, 통계 분석은 두 변수 간의 차이와 관계성을 규명하거나 평가할 때 사용하고, 데이터마ining은 데이터베이스에 저장된 대용량의 데이터를 기반으로 패턴, 상관관계, 추세 등을 알고자 할 때 주로 사용된다.

3. 스포츠 데이터 분석의 연구동향

최근, 스포츠 데이터 분석은 프로선수의 가치를 객관적이고 과학적으로 평가하기 위한 방법으로 그 중요성이 강조되고 있다. 또한, 미국 메이저리그와 유럽 축구리그에서 활약하는 국내 스포츠 스타들을 분석한 다양한 콘텐츠가 방송을 통해 대중들에게 알려지면서 스포츠 데이터 분석은 엘리트 선수만을 위한 것이 아니라, 스포츠 문화를 즐기기를 위한 모든 사람들을 위한 콘텐츠로서 주목받고 있다. IBM의 스포츠 분석가 Matthew가 ‘스포츠 분석은 IT기술의 발전으로 인해 빠른 데이터 수집과 대용량 정보 처리가 가능해짐에 따라, 향후 스포츠 환경은 크게 달라질 것’이라고 말한 것처럼, 최근 IT분야에서 주목받고 있는 빅데이터(big data) 분석 기술을 스포츠 데이터 분석에 적용하기 위한 관련 기술들이 개발되고 있다. 특히, IBM은 ‘슬램트래커(SlamTracker)’라는 빅데이터 분석 기술을 적용한 테니스 경기 분석 시스템을 개발하여 지난 8년간의 그랜드 슬램 테니스 대회에서 발생한 경기 데이터를 수집하여 분석하였다. 이 시스템은 테니스 한 경기당 발생하는 4,100만 건의 데이터를 실시간으로 분석하여 해당 경기의 주요 선수 정보, 경기 결과의 사전 예측 정보, 사전에 예측된 정보와 실제 정보의 차이를 비교한 결과 등을 제공하고 있다(IBM, 2013).

LA다저스의 류현진 투수를 분석한 사례가 있다. 류현진 선수는 2013년 메이저리그 데뷔 첫 해, 정규리그 동안 192이닝, 평균자책점 3.00, 30경기 등판 중 22번의 퀄리티 스타트(6이닝 당 3실점 이하 경기 수)등

데뷔 첫해 매우 우수한 성적을 기록하였다. 하지만, 여기서 주목할 기록은 9이닝 당 볼넷 개수이다. 9이닝 당 볼넷 개수의 의미는 투수가 타자와의 대결에서 정면 승부를 하는지, 또는 장타를 의식하여 소극적으로 투구했는지를 평가할 수 있는 지표이다. 2013년 9월 기준, 류현진 선수의 9이닝 당 볼넷(Based on ball)개수는 평균 2.3개(2013년 9월 기준)로 일본 출신의 텍사스 레인저스 선발투수인 다르빗슈(4.2개)보다 낮은 수치를 기록하였다. 또한, 이 데이터를 두 선수의 메이저리그에 진출하기 전 자국리그의 3년간(2010-2012) 평균 기록과 비교해보면, 그 차이가 더욱 명확해진다. 류현진 선수는 최근 3년간의 9이닝 당 볼넷수는 2.3개로 메이저리그 진출 후 변함이 없었지만, 다르빗슈는 1.9개에서 4.2개로 크게 증가하였다. 즉, 류현진은 타자와의 대결에서 적극적이고 공격적인 투구를 한 반면, 다르빗슈는 소극적인 투구를 하였다는 것을 객관적으로 평가할 수 있다(MBC, 2013).

이처럼, 야구, 농구, 축구, 테니스와 같은 프로스포츠는 점점 높은 수준의 수학적 알고리즘을 기반으로 한 분석 방법들이 연구되고 있고, 스포츠 현장에 제시되고 있다. 실제로 2013년 10월 메이저리그 사무국은 ‘골드 글러브’ 후보자 선정을 위해 미국야구연구협회(Society for American Baseball Research)에서 제공하는 수비지표(SABR Defensive Index)를 25% 반영하겠다고 발표하였다. 야구에서는 미국 메이저리그에서만 사용되어오던 ‘세이버메트릭스’가 국내 프로야구에서도 적극적으로 사용되고 있고, 이와 비슷한 지표인 ‘카스포인트’를 개발하여 사용하고 있다. ‘카스포인트’는 ‘MBC SPORTS Plus’와 ‘Cass’가 공동 개발한 것으로 경기 중 발생하는 주요 결과를 점수로 환산하여 투수와 타자를 객관적으로 평가할 수 있는 한국형 야구 선수평가 지표 ‘라고 볼 수 있다. 이 포인트 제도는 사전에 정해진 경기 지표의 배점에 따라 매경기 점수를 누적하고, 시즌이 종료된 후에 랭킹 순위를 정하는 방식을 채택하고 있다.

추가적으로, 심판원과 경기 임원들에 대한 분석이 새로운 이슈로 떠오르고 있다. 심판원의 수행능력은 경기결과를 결정짓는데 매우 큰 요소로 작용될 수 있다. 따라서, 심판원과 임원의 수행능력은 연구에 있어서 매우 중요한 부분이다(Harshorn, 2009). Mellick(2005)의 연구에서는 심판원의 위치가 시각적인 정보의 획득에 영향을 주기 때문에 경기 판정에 많은 영향을 줄 수 있다고 보고되고 있다. 따라서, 심판원들에 대한 체력상태는 경기에 많은 부분에 영향을 줄 수 있기 때문에 심판원들을 위한 전문적인 체력 훈련 프로그램과 과학적인 심판원 평가 방식에 대한 연구가 필요하다(D'Ottavio & Gastagna, 2001), 국내 연구에서는 관련 연구가 매우 부족한 상황이다. 이처럼, 스포츠 데이터 분석은 전통적인 스포츠과학을 벗어나 수학과 통계학, 컴퓨터공학, 정보통신기술(IT) 등이 융·복합적으로 적용되고 있는 새로운 분야로서 인정받고 있다. 또한, 스포츠 데이터 분석은 선수, 지도자, 팀 등을 과학적으로 평가할 수 있는 새로운 지표의 개발과 같은 창의적인 연구가 요구되고 있기 때문에 향후 관련 연구가 증가할 것으로 예측된다.

IV. 결론

본 연구는 스포츠 데이터 분석과 관련된 선행연구들을 리뷰(review)하여 스포츠 데이터 분석에 사용된 통계 및 데이터마이닝 기법을 분석하였다. 분석을 통해 도출된 결론은 다음과 같다.

첫째, 스포츠 데이터 분석은 선수의 기록 향상 또는 경기에서의 승패 요인을 발견하여 경기력 향상을 위한 선수, 지도자 등의 행동을 분석하는 연구 분야라고 정의할 수 있다. 또한, 이 분야는 스포츠과학을 기반으로 컴퓨터공학, 통계학, 미디어공학 등 다학제, 융합적인 연구 분야이다.

둘째, 통계학은 데이터로부터 두 변수간의 관계를 평가하고, 변인간의 관계성을 찾는 것에 대해 중점을 두고 있는 반면, 데이터마이닝은 통계학에서 처리하지 못하는 텍스트 및 영상과 같은 비정형화된 데이터와 대용량의 데이터 처리 및 분석, 기계학습을 통한 예측 모델을 제시할 때 적합한 기술이다. 따라서, 연구자의 연구가설과 연구목적에 따라 통계 방법 및 데이터마이닝 기법을 적절히 사용해야 할 필요가 있다.

셋째, 스포츠 데이터 분석은 타 학문 및 기술과 학술적·기술적인 융·복합화의 가속화로 인해 해당 스포츠의 새로운 지표 개발에 대한 요구와 정보의 대중화로 인해 빠르게 발전할 수 있는 분야이다.

하지만, 국내 스포츠 데이터 분석에 대한 학문적·기술적인 수준은 시작 단계이기 때문에 발전 가능성이 매우 높은 분야이다. 또한, 국내 프로스포츠 선수들의 해외 진출 증가는 더욱 많은 스포츠 콘텐츠 개발을 요구할 것이며, 이는 관련 기술의 개발과 학술적 연구를 촉진하는 계기가 될 것으로 사료된다.

참고문헌

- 강상조(2002). 체육, 스포츠과학 분야의 학문적 성장: 통계적 방법 적용의 역사, 한국통계학회 2002년 추계 학술발표회 논문집. 서울.
- 강선영, 조창욱, 장지훈, 이상식(2012). 제10회 벤쿠버 장애인동계올림픽 참가 대한민국선수의 스포츠상해 실태 분석. 한국체육과학회지, 21(6), 1315-1325.
- 강우용, 오윤경, 이강현(2010). 탁구경기 내용분석을 위한 득실점 유형분석 및 체계화. 코칭능력개발지, 12(3), 25-38.
- 김기진, 안나영, 홍창배, 김상현, 김재훈, 변자영, 주영식, 조장래, 조창모, 고진호, 박주식, 김홍수(2011). 국가대표 마라톤선수의 도로러닝시 운

- 동생리학적 지표 분석. 코칭능력개발지, 13(3), 131-138.
- 김민철(2009). 시계열분석을 통한 프로야구 관중현황 예측모델연구: 부산롯데아이언즈를 대상으로. 한국스포츠산업 · 경영학회지, 14(1), 17-25.
- 김세중, 허종관, 도재현(2010a). 데이터 마이닝 기법을 이용한 한국 프로농구 경기 분석. 한국체육과학회지, 19(2), 1413-1420.
- 김세중, 허종관, 이강웅(2010b). 한국프로농구경기 2000-2009시즌 PLAY-OFF 진출 팀의 경기 자료 데이터를 이용한 경기분석. 한국체육과학회지, 19(4), 1405-1411.
- 김세형, 강상조, 박재현, 김혜진(2008). 한국프로농구 경기기록 분석에 의한 승패결정요인. 한국체육측정평가학회지, 10(1), 1-12.
- 김채희, 윤형기(2008). 경기력향상을 위한 세계컬링대회 경기내용 분석. 한국체육과학회지, 17(2), 47-57.
- 김현경(2009). 아시아 주니어 컬링 여자선수의 경기기술 분석. 한국체육과학회지, 18(1), 1109-1117.
- 김혜진(2007). 사회연결망(social network analysis: SNA)을 이용한 스포츠 경기분석. 한국체육측정평가학회지, 9(1), 99-112.
- 문성훈, 정선호, 양재동(1999). 내용 기반 스포츠 뉴스 비디오 시스템 설계 및 구현. 한국정보과학회 1999년도 봄 학술발표논문집 제26권 제1호(B), 서울.
- 박성진(2013). 경기력 향상을 위해 스포츠 현장에 적용된 스포츠IT융합기술의 연구동향. 한국스포츠과학기술학회지, 1(1), 65-81.
- 서재순(1994). 귀납적 추론을 이용한 프로야구 승패 예측시스템 개발에 관한 연구. 경희대학교 대학원 석사학위논문.
- 송종우(2008). 한국 프로스포츠 선수들의 연봉에 대한 다변량적 분석. 응용통계연구, 21(3), 441-453.
- 안재석(2010). K리그와 EPL의 지역별점유율과 공격방향 분석을 통한 경기력 향상에 관한 연구. 코칭능력개발지, 12(3), 17-23.
- 엄한주, 조정환, 신승윤(2002). 통계모형의 전문스포츠 현장 적용 사례. 한국통계학회 2002년 추계 학술발표회 논문집. 서울.
- 오광모, 이장택(2003). 데이터마이닝을 이용한 한국 프로야구 선수들이 연봉에 관한 모형연구. 한국스포츠사회학회지, 16(2), 295-309.
- 이강현, 오윤경, 강우용(2012). 탁구경기 내용분석기법의 적용 및 활용방안 탐색. 코칭능력개발지, 14(3), 115-124.
- 이경용(2007). K-리그, 프리미어리그, 프리메라리가, 세리에A의 반칙유형 비교 · 분석. 명지대학교 기록과학대학원 석사학위논문.
- 이규민, 김주학, 김진환, 홍성진(2012). 2010년 남아공월드컵 조별 예선경기에서 나타난 슈팅요인 결과 분석. 체육과학연구, 23(2), 265-273.
- 이기봉, 이영석(2006). 한국과 덴마크 여자 핸드볼팀의 공격 방법 비교. 코칭능력개발지, 8(1), 11-22.
- 이기청(2001). 남자배구 선수들의 공격 수행 능력 평가. 한국운동역학회, 10(2), 35-46.
- 이만규(2006). 세이버 매트릭스를 적용한 프로여자타자의 경기력과 연봉과의 관계. 국민대학교 스포츠산업대학원 석사학위논문.
- 이순호(2008). 탁구 경기내용분석 프로그램 개발. 체육과학연구, 19(4), 82-93.
- 이중철(2010). 웹 기반 스포츠 기술지도 시스템 모형 개발. 체육과학연구, 21(2), 1123-1132.
- 이현재, 강경호, 김혜진, 박재현, 박두순, 홍민(2008). 실시간 축구 경기 데이터 분석 및 처리. 한국인터넷정보학회 2008 정기총회 및 추계학술발표대회, 제9권 제2호, 서울.
- 정지명, 이건희, 신선운(2011). 스포츠IT융합 인력 수요전망. 서울: 체육과학연구원 연구용역보고서.
- 정호석, 이종욱, 유재학, 이한성, 박대희(2011). 웹 캐스트와 텍스트 마이닝을 이용한 축구 경기의 심층분석. 한국콘텐츠학회논문지, 11(10), 59-68.
- 천영진(2009). 남자 프로배구 경기에서 득점 과정을 이용한 경기 분석. 한국사회체육학회지, 38,

- 1305-1312.
- 천영진, 김규태(2011). 랠리포인트제에서 여자 프로 배구 득점 내용 및 추이 분석. *한국체육과학회지*, 20(2), 729-737.
- 최재일, 정구영(2008). 여자프로골퍼의 스포츠 상해 분석. *한국체육과학회지*, 17(4), 1433-1441.
- 한필수(2006). 2006 겨울리그 데이터를 이용한 경기 상황별 승·패 요인 분석. *코칭능력개발지*, 8(3), 85-94.
- 허준희, 정태충(1998). 프로야구 경기 예측 시뮬레이션에서의 역전과 알고리즘을 이용한 투수 교체시기 예측 모듈 개발. *한국정보과학회 봄 학술발표논문집*, 제25권 제1호, 서울.
- 홍석미, 정경숙, 정태충(2003). 혼합형 기계학습 모델을 이용한 프로야구 승패 예측 시스템. *정보과학회 컴퓨팅의 실제 논문지*, 9(6), 693-698.
- Alamar, Benjamin. G.(2013). *Sports Analytics: A guide for coaches, managers, and other decision makers*. New york: Columbia University Press.
- Andersen, T., Arnason, A., Engebretsen, L., and Bahr, R.(2004). Mechanisms of head injuries in elite football. *British Journal of Sports Medicine*, 38(6), 690-696.
- Barbacci, S.(2002). Labanotation: a universal movement notation language. *Journal of Science Communication*, 1, 1-11.
- Bartlett, R. M.(2004). Artificial intelligence in performance analysis. In P. O'Donoghue & M. Hughes (eds.) *Notational Analysis of Sport VI*, (In press).
- Bauer, H.U. & Schöllhorn, W.I.(1997). Self-organizing maps for the analysis of complex movement patterns. *Neural Processing Letters*, 5, 193-199.
- Bayraktar, B., Dinç, C., Yücesir, I., and Evin, A.(2011). Injury evaluation of the Turkish national football team over six consecutive seasons. *Ulus Travma Acil Cerrahi Derg.* 17(4). 313-7.
- Belloti, B.(1992). *The Points Created Basketball book*. New Jersey, Night Work Publishing.
- Carling, C., Bloomfield, J., Nelson, L., and Reilly, T.(2008). The role of motion analysis in elite soccer: contemporary performance measurement techniques and work rate data. *Sports Med*, 38(10). 839-862.
- Choi, H., O'Donoghue, P. G., and Hughes, M. D.(2008). A comparison of whole match and individual set data in order to identify valid performance indicators for real-time feedback in men's single tennis matches. In A. Lees, D. Cabello, G. Torres (Eds.), *Science and Racket Sport IV*. (pp. 227-31), London: Routledge.
- Croucher, J. S.(1982). The effect of the tennis tie-breaker. *Research Quarterly for Exercise and Sport*, 53(4). 331-336.
- D'Ottavio, S. & Gastagna, C.(2001). Physiological load imposed on elite soccer referees during actual match play. *Journal of Sports Medicine and Physical Fitness*, 41(1), 27-32.
- Dvorak, J., Junge, A., Derman, W., and Schwellnus, M.(2011). Injuries and illnesses of football players during the 2010 FIFA World Cup. *British Journal of Sports Medicine*, 45(8), 626-630.
- Edelmann-Nusser, J., Hohmann, A., and Henneberg, B.(2002). Modelling and prediction of competitive swimming performance in swimming upon neural networks. *European Journal of Sport Science*, 2(2), 1-10.
- Franks, I. M. & Miller, G.(1991). Training coaches to observe and remember. *Journal of Sports Sciences*, 9(3), 285-297.
- Franks, I. M.(1997). Use of feedback by coaches and players. In T. Reilly, J. Bangsbo, M. Hughes (eds), *Science and Football III*. (pp. 267-78), London: E and FN Spon.
- Gabrel. B. Cost.(2009). *Practicing Sabetmetrics*. NC:

- McFarland.
- Gale, D.(1971). Optimal strategy for serving in tennis. Mathematical Association of America, 44(4), 197-199.
- Ganeshapillai Gartheeban & John Guttag(2013). A data-driven method for in-game decision making in MLB: when to pull a starting pitcher. ACM, 973-979.
- Ghosh, Avijit. & Steckel, Joel H. (1993). Roles in the NBA: There's always room for a big man, but his role has changed. Interfaces, 23(4), 43-55.
- Gilbert, W.D. & Trudel, P.(2004). Analysis of coaching science research published from 1970-2001, Research Quarterly for Exercise and Sport, 75(4), 388-399.
- Grant, A., Reilly, T., Williams, M., and Borrie, A.(1998). Analysis of the goals scored in the 1998 World Cup. The FA Coaches Association Journal, 2(1), 18-22.
- Gustavo Gonçalves Arliani, Paulo Santoro Belangero, Jose Luiz Runco, and Moisés CohenI.(2011). The Brazilian Football Association (CBF) model for epidemiological studies on professional soccer player injuries. Clinics (Sao Paulo), 66(10), 1707-1712.
- Helena, Andersson, Björn Ekblom, and Peter Krstrup.(2008). Elite football on artificial turf versus natural grass: Movement patterns, technical standards, and player impressions. Journal of Sports Sciences, 26(2), 113-122.
- Hughes, M. & Franks, I. (2007). The Essential of Performance Analysis. New York, Madson. Routledge.
- Hughes, M.(1993). Notational analysis of football, In T. Reilly, J. Clarys and A. Stibbe (Eds), Science and Football 2 (pp. 151-159). OX. London: Tatlor & Francis.
- IBM. IBM SlamTracker, October 28, 2013, from http://www.wimbledon.com/en_GB/slamtracker
- Ivanković, Z., Racković, M., Markoski, B., Radosav, D., and Ivković, M.(2010). Analysis of basketball games using neural networks. 11th IEEE International Symposium on Computational Intelligence and Informatics, (pp. 251-256), Budapest, Hungary: IEEE Computer society.
- Korkia, P.K., Tunstall-Pedoe, D.S., and Maffulli, N.(1994). An epidemiological investigation of training and injury patterns in British triathletes. British Journal of Sports Medicine, 28(3), 191-196.
- Laird, P. & Waters, L.(2008). Eye-witness recollection of sports coaches. International Journal of Performance Analysis of Sport, 8(1), 76-84.
- Lees, A. & Barton, G.(2004). A characterisation of technique in the soccer kick using a Kohonen neural network analysis. In Proceedings of the 5th World Congress of Science and Football. (In press).
- Lees, A., Barton, G., and Kershaw, L.(2003). The use of Kohonen neural network analysis to qualitatively characterize technique in soccer kicking. Journal of Sports Sciences, 21, 243-244.
- Lloyd Smith, Bret Lipscomb, and Adam Simkins(2007). Data Mining in Sports: Predicting CY Young Award Winners. Journal of Circuits, System and Computers, 22(4), 115-121.
- Maffulli, N., Baxter-Jones, A., and Grieve, A. (2005). Long term sport involvement and sport injury rate in elite young athletes. Archives of Disease in Childhood, 90(5), 525-527.
- Martin, L. (2006). Modelling the interaction in game sports-relative phase and moving correlations. Journal of Sports Science and Medicine, 5(4), 556-560.
- MBC. (2013). 류현진 특집-데이터 분석-[스포츠 매거진], 서울: MBC.

- McCrory. P., Heywood. J., and Coffey. C.(2005). Prevalence of headache in Australian footballers. *British Journal of Sports Medicine*, 39(2), e10.
- McGarry, T. & Frank, I. M.(1994). Stochastic approaches to predicting competition squash match play. *Journal of Sports Sciences*, 12(6), 573-884.
- McGarry, T. and Perl, J.(2004). Models of sports contests: Markov processes, dynamical systems and neural networks. In *Notational Analysis of Sport: Systems for Better Coaching and Performance in Sport* (edited by M.D. Hughes and I.M. Franks), pp. 227-242. London: Routledge.
- McGarry, T.(2006). Identifying patterns in squash contests using dynamical analysis and human perception. *International Journal of Performance Analysis of Sport*, 6(2), 134-147.
- McGarry, T., Khan, M.A., and Frank, IM.(1999). On the presence and absence of behavioural traits in sport: an example from championship squash match-play. *Journal of Sports Science*, 17(4), 297-311.
- McSharry & Patrick. E.(2007). Altitude and athletic performance: statistical analysis using football results. *British Medical Journal*, 335(7633), 1278-1281.
- Mohr, M., Krstrup P., and Bangsbo. J. (2003). Match performance of high-standard soccer players with special reference to development of fatigue. *Journal of Sports Science*, 21(7), 519-528.
- O'Donoghue, P.(2002). Performance models of ladies and men's singles tennis at the Australian Open. *International Journal of Performance Analysis of Sport*, 2(1), 73-84.
- O'Donoghue, P.G. & Ingram, B.(2001). A notational analysis of elite tennis strategy. *Journal of Sports Sciences*. 19(2), 107-15.
- O'Donoghue. P.(2009). *Research Methods for Sports Performance Analysis*. New York: Routledge.
- Palut, Y. & Zonone, P.G.(2005). A dynamical analysis of tennis: Concepts and data. *Journal of Sports Sciences*, 23(10), 1021-1032.
- Pedro, O. S. Vaz De Melo., Virgilio, A. F. Almeida., and Antonio, A. F. Louriero.(2012). Forecasting in the NBA and Other Team Sports: Network Effects in Action. *ACM Transactions on Knowledge Discovery from Data*, 6(3), 13-27.
- Perl, J.(2001). Artificial neural networks in sports: New concepts and approaches. *International Journal of Performance Analysis in Sport*, 1, 106-121.
- Reep, C. & Bejamin, B.(1968). Skill and chance in association football. *Journal of the Royal Statistical Society, Series A (General)* 131(4), 581-585.
- Robert, P. Schumaker, Osama, K. Solieman, and Hsinchun. Chen.(2010). *Sports Data Mining*. New York, Springer.
- Shafizadeh, M.(2008). Qualitative analysis of aggressive behaviors in the adolescent, youth and adult soccer World Cups. *International Journal of Performance Analysis of Sport*, 8(3), 40-48.
- Sun Jing, Yu Wenshuang, and Zhao Huiqun.(2010). Study of Association Rule Mining on Technical Action of Ball Games. *International Conference on Measuring Technology and Mechatronics Automation Vol.3*, (pp. 539-542), China: IEEE Computer society.
- Sung Geon Park & Soowon Lee.(2013). Curling analysis based on the possession of last stone per end. 6th Asia-Pacific Congress on Sports Technology(APCST), Vol. 60. (pp. 391-396). Hong Kong: International Sports Engineering Association.
- Thomas Reilly & A. Mark Williams(1993). Computer and video aided analysis of football games. In T. R. J. Clarys & A. Stibbe (Eds), *Science and*

- Soccer(second ed.) (pp. 167-174). NY: Routledge.
- Tscholl, P., O’Riordan, D., Fuller, C.W., Dvorak, J., Gutzwiller, F., and Junge, A.(2007). A Causation of injuries in female football players in top-level tournaments. *British Journal of Sports Medicine*, 1, 8-14.
- Yan, B. & Wu, Y.(2000). The ANN-based analysis model of the sports techniques. In *Proceedings of the XVIII International Symposium on Biomechanics in Sports*, Vol. 2. (edited by Y. Hong & D.P. Johns), pp. 585-589. Hong Kong: The Chinese Univ.

논문투고일 : 2016. 10. 31.

심 사 일 : 2016. 12. 26.

심사완료일 : 2016. 12. 30.