

게임 데이터를 이용한 지표 개발과 승패예측모형 설계[†]

구지민¹ · 김재희²

¹²덕성여자대학교 정보통계학과

접수 2017년 1월 11일, 수정 2017년 3월 10일, 게재확정 2017년 3월 14일

요 약

스포츠의 새로운 분야로 자리 잡고 있는 e-스포츠는 국내 뿐 아니라 해외에서도 많은 인기를 얻고 있다. 그 중 AOS (aeon of strife) 장르의 게임들은 대표적인 e-스포츠 대회 중 하나로 주목받으며, 방송 및 미디어 매체는 다양한 통계 지표를 활용한 게임 중계를 실시하고 있다. 본 논문에서는 AOS 장르의 게임인 리그오브레전드의 게임 데이터를 이용한 통계적 분석으로 게임 내 지표를 개선하고 승패예측을 위한 승패예측모형을 설계한다. 인자 분석을 통해 구한 인자로 기존의 지표를 개선하는 새로운 지표를 창출하고, 판별 분석, 인공신경망, SVM을 이용한 승패예측모형을 추정해 모형 간 비교를 실시하였다. 그 결과, 게임 내 포지션의 특성을 반영한 인자 점수로 새로운 지표를 제안하였으며, 세 가지 승패예측모형은 모두 평균 95%의 높은 정분류율을 보였다.

주요용어: 게임 데이터, 승패예측모형, 인공신경망, 인자 분석, 판별 분석, SVM.

1. 서론

온라인 게임은 가상세계라는 방대한 플랫폼을 바탕으로 스포츠, RPG (role playing game), FPS (first person shooting) 등 다양한 장르의 개발을 이루었다. 그 중 RTS (real-time strategy) 장르의 특별한 형태인 AOS (aeon of strife) 장르 게임은 최근 e-스포츠 산업의 성장과 더불어 많은 유저 (user)들의 관심을 받게 되었으며, e-스포츠 산업의 대표 주자로 떠오르며 국내 및 해외 대회를 개최하였다. MOBA (multiplayers online battle arena) 또는 ARTS (action real time strategy) 로도 불리는 AOS 게임은 플레이어의 전략과 운영이 게임의 승패를 좌우하는 중요한 요소로 작용한다. e-스포츠 대회에 참가하는 프로게임단은 코치진을 구성해 게임의 내적 분석과 전략을 세우며 팀의 성적을 향상시키기 위한 노력을 거듭하고 있다. 선수들은 게임 영상을 리플레이하며 피드백과 토의를 진행해 향후 게임 운영의 개선점을 찾고 있다. 최근 스포츠 게임에서의 데이터 분석도 활발하며 팀의 능력 파악과 전략 개발에 대한 연구 결과가 있다. Kim and Lee (2016), Cho (2016)는 야구 게임에서 승리모형을 다루고 Kang 등 (2015)은 배구팀의 전략을 세우기위한 정보를 얻기위한 모형을 제안하였다.

e-스포츠 시장이 활성화됨에 따라 e-스포츠 콘텐츠를 제공하는 방송·웹진 등의 미디어는 게임의 승패 예측과 더불어 선수들의 경기 지표에 관심을 두고 있다. 그 중 AOS 장르 게임인 리그오브레전드 (League of Legends)는 10명의 선수가 한 게임에 참여하기 때문에 게임 내에서 영향력 있는 선수를 선

[†] 이 논문은 2016년도 정부 (미래창조과학부)의 재원으로 한국연구재단의 연구 (NRF:2014R1A1A3050229) 지원과 산업통상자원부 (MOTIE)와 한국에너지기술평가원 (KETEP)의 지원을 받아 수행한 연구 과제입니다 (No. 20161210200610).

¹ (01369) 서울시 도봉구 삼양로 144길 33, 덕성여자대학교 정보통계학과, 학사과정.

² 교신저자: (01369) 서울시 도봉구 삼양로 144길 33, 덕성여자대학교 정보통계학과, 교수.

E-mail: jaehee@duksung.ac.kr

정하기 위한 과정에서 게임 데이터를 이용한 지표를 활용해왔다. 하지만 객관성의 측면에서 기존 지표가 가지는 문제점이 제기되었고, 이를 보완하기 위한 노력들이 지속되고 있다.

본 연구에서는 AOS 장르의 게임인 리그오브레전드의 게임 데이터를 이용한 통계적 분석을 실시해 게임에서 사용될 수 있는 지표를 탐색하고, 승패를 예측하는 모형을 설계하고자 한다. 게임사에서 제공하는 기존 지표와 인자 분석 결과를 바탕으로 만들어진 새로운 지표를 비교하며, 통계적 방법과 인공지능경망, SVM을 이용한 승패예측모형을 설계한다. 각 모형 간의 비교를 통해 모형 별 특징을 정리하고 승패 예측모형의 활용 방안을 제언한다. 본 연구는 게임 데이터에 대해 통계패키지 R을 활용하여 통계적 분석을 실시하며 e-스포츠 분야의 지표 연구를 활성화하고, 게임 데이터의 활용성을 높이는데 기여할 수 있을 것으로 기대한다. 본 논문의 구성은 다음과 같다. 2절에서는 게임 분야의 지표와 승패 관련 연구를 살펴보고 본 연구의 필요성을 설명한다. 3절에서는 본 연구에서 사용된 게임 데이터를 설명하고 4절에서는 분석 모형에 대한 설계와 식을 설명한다. 5절에서는 분석 결과와 모형 간 비교를 보여주며 6절에서는 향후 연구에 대한 보완을 제안한다.

2. 게임 데이터 관련 기존 연구

온라인 게임의 성장과 함께 e-스포츠 문화가 생성되면서 e-스포츠와 관련된 많은 연구들이 이루어졌다. Park (2009)는 e-스포츠의 가치와 스포츠로서의 e-스포츠를 바라보고 e-스포츠의 문제점을 해결하기 위한 방안을 제시하였으며, Oh 등 (2012)은 e-스포츠에 대한 연구 동향을 알아보고 연구 주제와 분석 방법을 정리하였다. e-스포츠 분야의 승패예측모형을 설계하기 위해 기존 스포츠 산업에서의 승패 예측과 관련한 연구를 살펴보았다. Gu 등 (2009)은 다중회귀분석을 이용해 경기에 영향을 미치는 요인을 분석하고 회귀분석, 인공지능경망을 이용한 승패 모형을 설계하였다. 또한 Kim 등 (2012)은 로지스틱 회귀분석, 의사결정나무분석을 통해 2010-2011 시즌의 한국남자프로농구 경기기록을 이용한 승패결정 요인을 분석하였다.

게임 산업에서도 방대한 양의 게임 데이터를 이용한 연구들이 진행되었다. Ryu 등 (2009)은 온라인 게임의 주요 관리 지표인 신규 이용자, 동시접속자 등에 대한 분석 및 예측을 위해 생존모형과 Bass의 확산 모형을 사용하였다. Lee (2008)는 게임 내의 중요 요소들을 도출하고 AHP (analytic hierarchy process)를 이용한 평가 요인 간의 중요도를 분석하였다. 이후 게임에서의 로그 데이터를 처리하는 기술이 발달하면서, 로그 데이터를 이용한 분석 연구도 생겨나기 시작했다. Kim 등 (2015)은 FPS 게임 서버의 로그를 분석해 클라이언트 단의 보안 솔루션 로그와 서버 로그를 융합한 치팅 탐지 모형을 설계하였으며, Kim 등 (2013, 2014)은 온라인 게임의 로그 데이터를 분석해 게임봇을 검출하는 방법과 그 과정을 소개하였다.

이와 같이 e-스포츠와 온라인 게임 산업에서는 이용자의 행동 분석과 함께 게임 데이터를 이용한 다양한 연구가 이루어지고 있음을 확인할 수 있었다. 따라서 본 연구에서는 e-스포츠 현장에서 활용될 수 있는 승패 예측 모형의 설계와 함께 기존의 게임 지표를 개선하여 게임 운영의 전략 수립과 발전에 도움을 주고자 한다.

3. 리그오브레전드 게임 데이터 설명

3.1. 분석 데이터의 출처 및 설명

본 연구에서는 e-스포츠를 대표하는 게임 중 하나인 라이엇게임즈의 리그오브레전드 게임 데이터를 이용해 분석을 실시하였다. AOS장르의 게임인 리그오브레전드는 개발자 사이트를 통해 게임 API (application programming interface)를 제공하고 있으며, 이를 이용한 게임 데이터의 분석이 가능하다.

분석 데이터는 리그오브레전드 한국서버의 2016 시즌에서 6.8 패치 버전 이전까지 플레이 된 5×5 랭크 게임으로 소환사의 협곡 맵에서 진행된 게임을 선택하였으며, 챌린저와 다이아몬드 랭크에 속한 선수들의 게임 데이터를 사용하였다.

이러한 기준을 만족하는 데이터로 $n = 376$ 개의 게임 데이터가 선택되었으며, 통계적 분석을 실시하기 위해 플레이어 별 데이터와 팀 별 데이터를 분리하는 과정을 거쳤다. 경기 데이터는 오픈 소스 R을 이용한 통계 분석을 실시하였다.

3.2. 변수의 선정

리그오브레전드 게임 내 소환사의 협곡 맵은 5×5 게임으로 5명의 플레이어가 한팀이 되어 상대방의 적 기지를 부수면 승리하는 게임이다. API에서는 한 게임이 끝나게 되면 시간 대 별로 기록된 게임 데이터를 가져올 수 있으며, 한 번의 게임 데이터에는 총 184개의 변수가 포함되어 있다. 플레이어 개인이 게임 내 미치는 영향력에 대해 인자 분석을 실시하기 위해 데이터를 가공하는 과정을 거쳐 17 개의 연속형 변수와 승패를 나타내는 이산형 변수 1 개, 포지션 분류를 위한 문자형 변수 1개를 선택하였으며, 데이터 가공 과정에서는 e-스포츠 대회 중계 방송 혹은 웹진에서 언급되는 지표들을 참고하였다. 그 결과 선택된 변수의 유형과 설명은 Table 3.1과 같다. 각 변수는 수행 과정에 앞서 전처리 과정을 거치며 표준정규화를 이용한다.

승패예측모형에 투입 될 변수를 선택하기 위해 플레이어 개인 데이터를 팀 단위 데이터로 통합하는 과정을 거쳤다. 플레이어 개인 데이터에 포함된 17개 변수가 모형을 설명하는 변수로 전부 투입 될 경우 모형의 신뢰도가 떨어질 수 있다는 점을 고려해 개인 데이터를 팀 단위의 데이터로 변환하였다. 그 결과 10개 연속형 변수와 승패를 나타내는 1 개의 이산형 변수를 정의할 수 있었다. 팀 단위 데이터의 타입과 설명은 Table 3.2와 같다.

Table 3.1 Variables of League of Legends game's player data

Name (Continuous)	Description	Variable
champLevel	Champion level achieved	X_1
kills	Number of kills	X_2
deaths	Number of deaths	X_3
assists	Number of assists	X_4
totalDamageDealt	Total damage dealt	X_5
largestCriticalStrike	Largest Critical Strike	X_6
totalDamageDealtToChampions	Total damage dealt to champions	X_7
totalDamageTaken	Total damage taken	X_8
totalHeal	Total heal amount	X_9
minionsKilled	Minions killed	X_{10}
goldEarned	Gold earned	X_{11}
wardPlaced	Number of wards placed	X_{12}
wardKilled	Number of wards killed	X_{13}
totalUnitsHealed	Total units healed	X_{14}
totalTimeCrowdControlDealt	Total dealt crowd control time	X_{15}
time	Milliseconds into the game the frame occurred	X_{16}
neutralMinionsKilled	Neutral minions killed	X_{17}
Name (Discriminant)	Description	Variable
winner	Flag indicating whether or not the participant won	Y
lane	Participant's lane (Legal values : MID, TOP, JUNGLE, BOT)	K

Table 3.2 Variables of League of Legends game's team data

Name (Continuous)	Description	Variable
kills	Number of kills the team	X_1
deaths	Number of deaths the team	X_2
assists	Number of assists the team	X_3
totalDamageDealt	Total damage dealt the team	X_4
minionsKilled	Number of Minions the team killed	X_5
goldEarned	the team Gold earned	X_6
wardPlaced	Number of wards the team placed	X_7
wardKilled	Number of wards the team killed	X_8
inhibitorKills	Number of inhibitors the team destroyed	X_9
towerKills	Number of towers the team destroyed	X_{10}
Name (Discriminant)	Description	Variable
winner	Flag indicating whether or not the participant won	Y

3.3. 기존 캐리레이팅 계산

캐리 레이팅 (carry rating)이란 게임 내에서 각 포지션이 수행하는 주요 역할을 반영해 선수의 실력을 나타내는 지표이다. 리그오브레전드 게임 대회가 신설된 이후로 MVP 선수를 선정하는 데에 주로 사용되었으며, 각 팀의 키 플레이어와 선수 랭킹을 매길때에도 캐리 레이팅을 이용하고 있다.

Table 3.3 Carry rating

Position	Name	Description	Formula
TOP	D(T)PM	Total damage per minute	(Total damage dealt to champions) + (Total damage taken to champions)/time
JUNGLE	KA%	Contribution to kill	(Number of kills)+(Number of assists) + (Total number of kills)
MIDDLE	DPM	Damage per minute	(Total damage dealt to champions)/time
BOT-AD	DPM	Damage per minute	(Total damage dealt to champions)/time
BOT-SUPPORT	APG	Number of Assist per game	Number of assists

게임 초기에는 선수들을 평가하는데 있어 식 3.1의 KDA (kill-death-assist) 수치와 MVP (most valuable player) 선수를 투표하는 판정단의 채점을 MVP 선정의 주요 지표로 사용하였다. KDA수치는 다음과 같이 계산되어 진다. 하지만 KDA 수치는 게임 내 플레이 요소들이 반영되기보다는 게임의 일부 요소만이 반영되었으며, MVP 선정단의 채점 역시 주관적인 판단이라는 유저들의 지적이 있었다. 실제로 경기 내에서 적에게 가한 데미지가 낮았음에도 불구하고 MVP로 선정된 선수도 있었으며, MVP 판정단의 판정으로 유저들의 예측과는 다른 선수들이 선정되기도 하였다.

$$KDA = \begin{cases} \frac{\text{Number of kills} + \text{Number of assists}}{\text{Number of deaths}}, & \text{if Number of deaths} > 0, \\ (\text{Number of kills} + \text{Number of assists}) \times 1.2, & \text{if Number of deaths} = 0. \end{cases} \quad (3.1)$$

리그오브레전드 게임은 이와 같은 문제점을 개선하기 위해 리그오브레전드 챔피언스 코리아 (이하 LCK)에 포지션 별 세분화 된 지표로 캐리 레이팅을 도입하였으며, LCK 2015 Summer 시즌부터 시범적으로 사용하였다.

4. 모형의 설계

4.1. 캐리 레이팅 모형

현재 사용되고 있는 캐리 레이팅 지표는 게임 데이터 중 일부 요소만을 반영하기 때문에 누락된 요소들은 캐리 레이팅 계산에 이용되지 않고 있다. 본 연구는 결손 데이터의 수를 줄이고 데이터가 가지는 변수를 최대한으로 설명하기 위해 인자 분석을 이용한 캐리 레이팅 지표의 개선안을 제시한다.

데이터는 표면적으로 드러나는 요소 외에 잠재적으로 존재하는 설명 요인들이 존재한다. 여러 개의 변수가 관측되는 데이터의 경우 변수 간의 관계에 대한 분석이 필요한데, 이 과정에서 사용할 수 있는 분석이 바로 다변량 통계 분석이라 할 수 있다. 그 중 인자 분석은 서로 관련이 있는 변수들 속에서 변수들을 잘 설명하는 인자를 찾아내는 분석 방법으로 Kim (2015)를 참조한다. 다중인자모형은

$$X - \mu = LF + \epsilon \quad (4.1)$$

으로 변수들을 공통 인자 F 와 특수 인자 ϵ 의 선형 결합으로 표현한다. 여기서 μ 는 X 의 평균벡터로 $X = (X_1, X_2, \dots, X_p)'$, $p = 17$ 이고, F 는 공통인자벡터, ϵ 은 특수인자, L 은 인자적재행렬을 나타낸다.

분석 과정에서는 376개의 게임 중 승리한 경우에 해당하는 유저들의 게임 내 데이터를 가져와 이를 각 포지션 별로 나누고, 포지션 별로 나뉜 데이터에 따라 각 포지션의 특성을 설명하는 공통 인자들을 찾는다. 또한 인자 모형에 대한 적합으로 인자 점수를 추정하여 이를 캐리 레이팅의 개선된 지표로 설정한다. 추정된 인자 점수는 기존의 캐리 레이팅과의 비교를 통해 차이점을 보여준다.

4.2. 승패예측모형

현재까지 예측 모형은 금융, 주식 등의 분야에서 주로 연구되었으며 회귀분석, 마코프 연쇄를 이용한 다양한 승패예측모형의 연구가 이루어져왔다. 본 연구에서는 기존의 통계적 예측 기법인 판별 분석, 인공신경망, SVM을 이용한 예측 모형을 설계해 각 모형 간 비교를 통해 예측 모형의 차이점을 알아보고 최선의 모형을 선택한다. 승패 예측의 기준으로 사용되는 변수는 1 (승리)과 0 (패배)으로 표현된 winner 변수이다. 본 연구에서는 정분류율을 모형의 예측 성능을 판단하는 평가 지표로 사용한다.

판별 분석은 이미 알려진 집단으로 구성된 자료들로부터 정보를 얻어 집단을 구별할 수 있는 함수를 결정하는 것이다. 판별 분석은 분석하고자 하는 자료의 변수들이 서로 상관성이 높지 않고 이상점이 없으며 종속변수가 질적 변수일 때 사용할 수 있다. 다변량 정규분포를 따르며 공분산 행렬이 동일한 경우는 선형 판별함수를 사용하지만 공분산 행렬이 다른 경우에는 이차판별함수를 사용한다. 이는 분산 간 차이 정도에 의해 결정되며, 통계적 방법으로 분산 동일성 검정을 실시해 모형을 결정한다. 이차판별함수는

$$Q(X) = \frac{1}{2} \ln \frac{|\Sigma_1|}{|\Sigma_2|} - \frac{1}{2} (\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2) X - \frac{1}{2} X' (\Sigma_1^{-1} - \Sigma_2^{-1}) X \quad (4.2)$$

으로 표현되며, 표본을 이용하는 경우 μ_1, μ_2 대신 표본평균 \bar{X}_1, \bar{X}_2 을 이용한다. Σ_1, Σ_2 는 각각 승리그룹과 패배그룹의 공분산행렬로, Σ_1, Σ_2 에 대해서는 공분산행렬추정량 S_1, S_2 을 사용하게 된다. 본 분석에서 $Q(X)$ 의 표본함수인 $Q_s(X)$ 는

$$Q_s(X) > \ln \frac{p_1}{p_2} \quad (4.3)$$

이면 X (플레이어)를 승리그룹에 분류하고, 그렇지 않으면 X 를 패배그룹에 분류한다. 여기서 p_1, p_2 는 각각 플레이어가 승리, 패배그룹에서 발생하는 사전확률이다. 본 분석에서는 승패 예측 모형을 결정하기 위해서 게임이 종료된 시점에서의 팀별 데이터를 이용하며 분석 과정을 통해 모형의 모수를 추정한다.

인공신경망은 생물의 신경망을 모방하여 컴퓨터에 구현한 학습 알고리즘이다. 신경망 (neural network)에서 각 노드들은 네트워크를 형성하며 생물의 뉴런 역할을 한다. 인공 노드들은 입력층, 은닉층, 출력층으로 구분되어지며 데이터의 입력 값에 따라 원하는 출력이 나올 수 있도록 링크의 가중치를 조정하는 방법으로 학습을 수행한다. 본 연구에서는 게임 내 팀 데이터 변수들을 입력층으로, 출력층의 처리단위의 수는 0 (패배)과 1 (승리)의 값으로 부여한 단일 은닉층 신경망 (single-hidden-layer neural network)를 이용한다. 은닉층 처리단위의 수는 입력층과 출력층에 사용된 단위 수의 평균을 사용하였으며 소수점은 반올림하였다.

신경망의 장점은 학습이 가능하다는 점인데 이 학습 과정을 통해 비용 함수를 최소화 할 수 있다. 비용 함수는 최적해를 찾기 위해 확률 과정을 수반하며 비용 함수를 찾는 과정에서는 시그모이드 함수가 사용된다. 시그모이드 함수는 로지스틱 함수의 특별한 형태로 미분이 되지 않는 계단식의 함수를 미분 가능하도록 곡선형태로 바꾸어 주는 활성화 함수이다. 주로 은닉층의 입출력 특성을 표현하는데 사용되며, 은닉층에서는 입력된 데이터의 합성값들이 활성화 함수를 거치며 입력된 값에 내재하고 있는 패턴과 특성을 알아낸다. 시그모이드 활성화 함수의 식은

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4.4)$$

이고, 입력층의 벡터 (플레이어 데이터)로 표현되는 $X = (x_1, x_2, \dots, x_p)'$ 를 활성화 함수 (4.4)에 대입하면 은닉층의 j 번째 뉴런은

$$H_j = \frac{1}{1 + \exp[-\sum_{u=1}^n w_j^u x_j^u]} \quad (4.5)$$

와 같이 표현된다. 수식에서 w 는 입력층에서 은닉층으로 향하는 가중치로, 가중치가 포함된 값을 계산해 은닉층의 출력값을 $(H_1, H_2, \dots, H_k)'$ 과 같이 나타낼 수 있다. 은닉층의 출력값과 출력층으로 향하는 가중치 벡터인 $V = (v_1, v_2, \dots, v_k)'$ 를 적용하면 출력층의 i 번째 패턴은

$$Y_i = \sum_{j=1}^k V_j H_j \quad (4.6)$$

와 같이 표현할 수 있다.

SVM (support vector machine)은 다른 범주에 속한 데이터 간의 간격이 최대가 되도록 하는 선 또는 평면을 찾는 분류 모형으로 Hastie 등 (2001)을 참조한다. 입력된 데이터가 선형으로 구분되지 않는 경우 데이터를 고차원으로 대응시켜 카테고리를 분류하는 선 또는 평면을 찾을 수 있다. 시험 데이터 (training data)는 $\{(x_1, y_1), \dots, (x_m, y_m)\}$ 으로 표기한다. 여기서 $x_i = (x_{i1}, \dots, x_{ip})' \in \mathbb{R}^p$, $y_i \in \{-1, 1\}$ 이다. 고차원 평면 기하연구에 의해 모든 i 에 대해 $y_i f(x_i) \geq 0$ 을 만족하는 $f(x) = x'\beta + \beta_0$ 를 찾을 수 있다.

$$\begin{aligned} &\text{minimize} && \frac{1}{2} \|\beta\|^2 + \gamma \sum_{i=1}^m \xi_i, \quad m = \left\lceil \frac{n}{3} \right\rceil \\ &\text{subject to} && \xi_i \geq 0, \quad y_i(x_i'\beta + \beta_0) \geq 1 - \xi_i \quad \forall i, i = 1, \dots, m \end{aligned} \quad (4.7)$$

여기서 $\{\xi_1, \xi_2, \dots, \xi_m\}$ 는 여유 변수 (slack variable)로 $\xi_i \geq 0$, $\sum_{i=1}^m \xi_i \leq \text{constant}$ 를 만족한다. γ 는 예측과정에서의 오류를 줄이기 위한 패널티 항이다. γ 에 대한 최적값은 교차타당성 (cross-validation) 계산을 통해 구하며, γ 는 특성공간 (feature space)를 확장시키는 역할을 한다. 이 과정에서 데이터를 고차원으로 변환하는 대신 고차원에서 벡터의 내적 연산으로 계산한 값과 같은 값을 반환하는 커널

함수 (kernel function) 을 사용한다. 본 분석에서는 가우시안 RBF 커널함수 (Gaussian radial basis function)

$$k(x, x') = \exp(-\sigma \|x - x'\|^2) \quad (4.8)$$

을 이용한다. 가우시안 RBF 커널은 일반적으로 데이터에 대한 사전 정보가 없을 때 주로 이용되며 다양한 데이터의 형태에 잘 적용된다는 장점이 있다. 본 분석에 사용되는 게임데이터는 다변량 자료로 팀의 승패를 0과 1로 구분하기 위해 가우시안 RBF 커널 함수를 이용한다.

5. 데이터 분석 및 결과

5.1. 캐리 레이팅 모형

5.1.1. 인자 분석 결과

인자 점수를 생성하기 위해 각 포지션 별로 인자 분석을 실시하였다. 대표적으로 승리한 팀에서 탑 (TOP) 포지션에 해당하는 플레이어의 데이터를 추출해 프로맥스 (promax) 회전된 인자분석을 실시한 결과는 Table 5.1과 같다.

Table 5.1 Factor loadings of the TOP position

Variable	Variable Name	Factor1	Factor2	Factor3	Factor4	Factor5
X1	champLevel	0.48	0.541			
X2	kills			0.949		
X3	deaths	0.635	-0.137			0.146
X4	assists	0.809	-0.171			
X5	totalDamageDealt		0.842			0.143
X6	largestCriticalStrike	-0.256	0.263	0.198	0.209	0.128
X7	totalDamageDealtToChampions	0.408	0.309	0.383		
X8	totalDamageTaken	0.805			0.498	
X9	totalHeal	0.457			0.739	
X10	minionsKilled		1.176	-0.108		-0.235
X11	goldEarned	0.38	0.515	0.243		
X12	wardsPlaced	0.511	0.4	-0.134		
X13	wardsKilled	0.291	0.504	-0.196		0.137
X14	totalUnitsHealed				0.447	
X15	totalTimeCrowdControlDealt	0.292				-0.163
X16	time	0.71	0.371	-0.116		
X17	neutralMinionsKilled		-0.144			0.901

인자 1 (Factor 1)은 방어력 기반의 챔피언 성장 인자를 의미하며 챔피언의 레벨 X_1 과 골드 획득량 X_{11} 을 포함한다. 특징적으로 적에게 받은 총 데미지 X_8 가 포함되는데, 이는 방어력 기반으로 얼마나 적의 데미지를 잘 받고 아군을 보호하는 자세를 취했는가를 의미한다. 따라서 인자 1은 탑 포지션이 얼마나 방어력에 충실했는가를 보여주는 인자로 판단 할 수 있다. 인자 2는 공격력 기반의 챔피언 성장 인자를 의미한다. 인자 1과의 차이점은 적에게 받은 총 데미지 변수 X_8 이 생략되고 적에게 가한 피해량 X_5 가 포함되었다는 점이다. 적에게 가한 피해가 높을수록 방어력 기반의 챔피언 성장보다는 공격력 위주의 챔피언 성장이 이루어졌음을 뜻하며, 미니언 처치 수 X_{10} 를 포함해 챔피언 성장이 얼마나 잘 되었는가를 보여준다. 인자 3은 챔피언 처치 인자로 탑 포지션이 적을 얼마나 처치했는지를 의미한다. 특히 킬 횟수 X_2 변수를 포함하는 것으로 보아 챔피언 처치 횟수로 얼마나 성장하였는가를 보여준다.

Figure 5.1은 프로맥스 회전된 탑 포지션의 인자 패턴이다. 인자 1의 축과 가장 가까운 변수는 적에게 받은 총 데미지 변수 X_8 이며, 인자 2의 축과 가장 가까운 변수는 미니언 처치 수 X_{10} 와 골드 획득량

X_{11} 으로 나타났다. 오른쪽 그림은 인자 점수 결과로 탑 포지션에서 인자 1과 인자 2에 해당하는 점수를 산점도로 인자 1과 인자 2의 값이 높을수록 경기 내에서 좋은 플레이를 보였다고 할 수 있다.

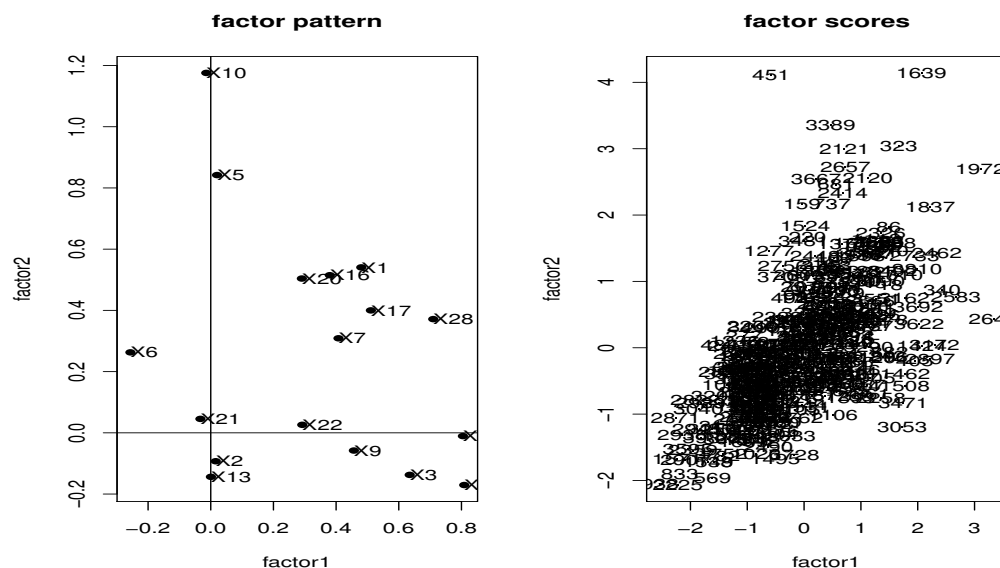


Figure 5.1 Factor pattern and factor scores of the TOP position

탑 포지션에서 실시한 인자 분석과 동일한 방법으로 정글 (JUNGLE), 미드 (MIDDLE), 바텀 (BOTTOM) 포지션을 대상으로 인자 분석을 실시하였고, Table 5.2와 같은 인자 분석 결과를 얻을 수 있었다. 게임 내에서 바텀 포지션은 원거리 딜러와 서포터라는 두 가지 포지션으로 나눌 수 있다. 바텀 포지션의 인자 분석을 실시한 결과 인자 1은 원거리 딜러 (AD) 인자를 나타내었고, 인자 2는 서포터 (SUP) 인자로 판단할 수 있었다. 따라서 바텀 포지션은 두 가지 포지션으로 세분화하여 표현하였다.

Table 5.2 Factor loadings of all positions

Variable	TOP		JUNGLE		MIDDLE		BOT-AD	
	Factor1	Factor2	Factor1	Factor2	Factor1	Factor2	Factor1	Factor2
X_1	0.48	0.541	0.6	0.266	0.864	0.114	0.677	0.556
X_2			-0.306	0.978			0.338	-0.164
X_3	0.635	-0.137	0.648		-0.172	0.869	0.193	0.548
X_4	0.809	-0.171	0.819		0.286	0.492	-0.244	0.765
X_5		0.842	0.232	0.386	0.956	-0.148	1.049	
X_6	-0.256	0.263	-0.159	0.172	-0.149		0.712	-0.148
X_7	0.408	0.309	0.257	0.704	0.525	0.167	0.795	
X_8	0.805		0.796			0.669	0.218	0.786
X_9	0.457		0.462	-0.101		0.246	-0.145	0.598
X_{10}		1.176	0.322	0.612	1.146	-0.265	0.98	-0.199
X_{11}	0.38	0.515			0.469		0.951	
X_{12}	0.511	0.4	0.48	0.522	0.84		0.763	0.381
X_{13}	0.291	0.504	0.637	-0.166	0.647	0.242	-0.354	0.872
X_{14}			0.618	-0.121	0.588	0.208		0.771
X_{15}	0.292		0.256				-0.271	0.297
X_{16}	0.71	0.371	0.251	0.106	0.16		0.398	0.167
X_{17}		-0.144	0.893		0.761	0.322	0.467	0.842

5.1.2. 캐리 레이팅의 비교

기존의 캐리 레이팅과 인자분석을 이용해 새롭게 도출한 캐리 레이팅의 비교를 실시하였다. 기존 캐리 레이팅 계산법으로 각 포지션 별 가장 높은 점수를 받은 10명의 유저번호인 ID를 가져왔으며, 인자 분석에 의한 각 포지션 별 캐리 레이팅 계산으로 높은 점수를 받은 10명의 ID를 가져왔다. 캐리 레이팅 계산 과정 중 데이터 내에서 원거리 딜러와 서포터의 구분이 되어 있지 않은 점을 고려해 군집 분석을 통해 바텀 포지션을 2개의 그룹으로 분리하였고, 분리된 데이터로 각 포지션 별 캐리 레이팅을 계산하였다.

Table 5.3 Carry Rating using the existing indicators

	TOP		JUNGLE		MIDDLE		BOT(AD)		BOT(SUP)	
	ID	C.R	ID	C.R	ID	C.R	ID	C.R	ID	C.R
1	451	12149.8	779	0.947	828	9268.14	1306	9644.40	656	34
2	681	9694.19	1868	0.938	28	8452.86	1117	8635.50	403	32
3	1639	8134.94	346	0.909	124	8141.433	3116	8622.26	1282	29
4	737	8013.08	276	0.870	1839	7626.18	3089	7809.81	2461	29
5	159	7987.54	1544	0.857	1637	7453.35	1208	7809.21	2629	28
6	1277	7689.52	591	0.857	1320	7356.50	2585	7786.38	983	27
7	1524	7222.94	150	0.828	881	7325.68	3694	7689.86	1038	27
8	220	7099.78	721	0.826	1676	7283.65	3606	7680.42	3338	27
9	1736	7034.02	961	0.815	1898	7188.92	3809	7388.79	3104	26
10	323	6907.55	1128	0.813	324	7006.60	90	7382.74	3286	26

Table 5.4 Carry Rating using the factor scores

	TOP		JUNGLE		MIDDLE		BOT(AD)		BOT(SUP)	
	ID	C.R	ID	C.R	ID	C.R	ID	C.R	ID	C.R
1	1639	4.142	1040	3.857	3806	2.994	1306	3.991	1974	3.750
2	451	4.117	3570	3.279	1637	2.989	1117	3.889	1640	3.076
3	3389	3.353	3463	3.003	1973	2.930	1970	3.743	2645	3.041
4	323	3.039	3171	2.909	1975	2.904	2585	3.476	1838	2.645
5	2121	2.998	1638	2.782	124	2.656	1636	3.395	2461	2.570
6	2657	2.730	259	2.741	324	2.620	2642	3.284	321	2.511
7	1972	2.691	1669	2.717	1839	2.533	3809	2.917	2581	2.446
8	2120	2.564	2388	2.662	28	2.231	1825	2.851	1824	2.195
9	3667	2.540	2260	2.573	828	2.155	90	2.832	338	2.178
10	681	2.460	826	2.541	2644	2.058	3116	2.810	19	2.177

Table 5.3과 Table 5.4를 비교한 결과 각 포지션 별로 동일하게 추출된 ID가 존재하였다. 미드 포지션의 인자 1은 미드 포지션 성장 인자를 의미하는데, 기존의 캐리 레이팅과 비교한 결과 포지션의 성장이 높을수록 분당 데미지가 높게 측정되었다. 기존의 캐리 레이팅이 분당 데미지만을 반영한다면 개선된 캐리 레이팅은 미드 포지션의 성장과 관련된 요소들이 반영되므로 게임의 상세한 부분을 표현한다고 볼 수 있다. 탑 포지션의 경우에는 공격력 위주의 챔피언 성장 인자 점수를 캐리 레이팅으로 판단하였다. 스포링 시즌 6.8 패치 버전에서는 탱커형 챔피언이 유행하였기 때문에 방어력 기반의 챔피언 성장 인자가 더 적합할 것으로 예상되었으나 이와 달리 공격력 위주의 챔피언 성장에서의 캐리 레이팅 계산이 기존의 캐리 레이팅 계산과 유사한 결과를 보였다. 이는 프로 경기가 아닌 일반 유저들의 게임 데이터라는 점에서 밴픽 (ban-pick) 과정과 게임 운영 방식의 차이로 인한 결과가 발생한 것으로 사료된다.

캐리 레이팅 비교 과정에서 가장 큰 차이를 보이는 부분은 정글 포지션과 서포터 포지션이다. 정글 포지션은 킬 관여율로 불리는 캐리 레이팅을 채택하고 있는데, 킬과 어시스트 요인 외에 게임 내적 요소들이 반영되지 않는다는 점이 문제로 지적되었다. 이를 개선하기 위해 캐리형 정글 포지션 인자를 선택해 인자 점수에 의한 캐리 레이팅 계산을 실시하였다. 서포터 포지션 역시 게임 당 어시스트 수로 계산된 기존의 캐리 레이팅 대신 서포터 포지션 인자 점수를 선택했다. 어시스트 횟수 뿐 아니라 서포터의 시야

장악 능력을 반영하기 때문에 인자 분석으로 계산된 캐리 레이팅이 기존보다 개선된 지표로 사용될 수 있을 것으로 보인다.

5.2. 승패예측모형

5.2.1. 판별 분석 결과

판별 분석을 실시하기에 앞서 모형을 결정하기 위해 공분산행렬에 대한 동일성 검정을 실시하였다. 검정 결과 유의수준 5%에서 그룹 간 공분산행렬이 모두 같다는 귀무가설을 기각하므로 그룹 간 공분산행렬이 다르다는 결과에 근거하여 두 그룹을 분리하는 이차판별함수를 선택하였다.

Table 5.5 Means of variables

Group	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
0 (Defeat)	15.95	26.07	25.96	438252.55	571.70	44979.26	67.86	23.02	0.09	1.38
1 (Win)	26.03	16.05	45.00	542299.39	621.33	55251.62	70.28	24.90	1.07	5.79

Table 5.5는 그룹 별 변수들의 평균을 나타낸 것이다. 변수 별 평균을 살펴보면 죽은 횟수 변수 X_2 를 제외한 나머지 변수들의 평균이 승리 (1) 그룹에서 높게 나타나는 것을 알 수 있다.

Table 5.6 Classification using quadratic discriminant function

	Classification		
		0 (Defeat)	1 (Win)
	Original	92.46%	7.54%
		2.27%	97.73%

판별 분석에 의한 분류 결과는 Table 5.6과 같다. 판별 모형의 정분류율은 94%로 높은 분류 성능을 보였다. 특히 기존의 승리 그룹을 정확하게 분류할 확률은 98%로 매우 높게 나타났다. 반면 패배를 승리로 분류할 경우는 잘못된 판단을 내릴 수 있게 하는 위험한 오류인데, 이 경우에서의 확률이 7.54%로 높게 나타나는 특징을 보였다.

5.2.2. 인공신경망 분석 결과

신경망 분석은 모든 변수를 투입하는 신경망 (입력층-은닉층-출력층) 모형으로 신경망 (10-6-2) 모형을 설정하였다. 출력층 처리단위는 0 (패배)와 1 (승리)의 값으로 부여하였으며, 은닉층 개수는 입력층과 출력층 처리단위의 평균값인 6을 사용하였다.

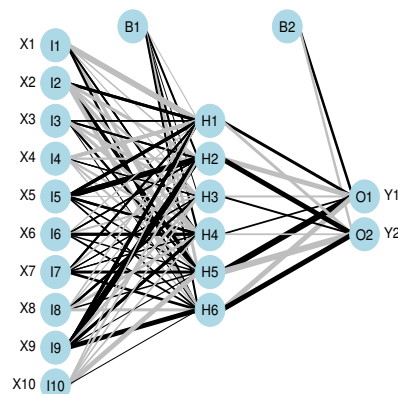


Figure 5.2 Single-hidden-layer neural network for League of Legends game data

Figure 5.2는 신경망 (10-6-2)모형을 적용한 결과를 그림으로 그린 결과이다. 죽은 횟수를 나타내는 변수 X_2 가 패배 (0)를 결정하는데 주로 작용하는 것을 알 수 있으며, 적에게 가한 데미지 변수 X_4 는 승리 (1)를 결정하는데 크게 작용하는 요소임을 확인할 수 있다.

Table 5.7 Classification using neural network

		Classification	
		0 (Defeat)	1 (Win)
Original	0 (Defeat)	95.13%	4.87%
	1 (Win)	4.2%	95.8%

Table 5.7을 보면 인공신경망 모형에 의한 분류 결과 승리 그룹과 패배 그룹을 정확히 분류할 확률은 약 95%로 높게 나타났다. 패배 그룹과 승리 그룹을 정확히 구별하는 분류율의 차이는 각각 95.13%, 95.8%로 큰 차이를 보이지 않는다. 오분류 경우를 살펴보면 패배를 승리로 분류할 확률과 승리를 패배로 분류할 확률이 각각 4.87%, 4.2%로 오류율의 차이 또한 적게 나타났다.

5.2.3. SVM 분석 결과

SVM을 이용해 승패 예측 모형을 적합한 결과는 Table 5.8에서 보여준다. SVM 모형에 의한 분류 결과 승리 그룹과 패배 그룹을 분류할 확률은 약 94%로 높게 나타났다. 패배 그룹과 승리 그룹을 구별하는 분류율은 각각 94.7%, 94.27%로 큰 차이를 보이지 않으며, 패배를 승리로 분류할 확률이 5.3%로 승리를 패배로 분류할 확률인 5.73%와 큰 차이를 보이지 않았다.

Table 5.8 Classification using SVM

		Classification	
		0 (Defeat)	1 (Win)
Original	0 (Defeat)	94.7%	5.3%
	1 (Win)	5.73%	94.27%

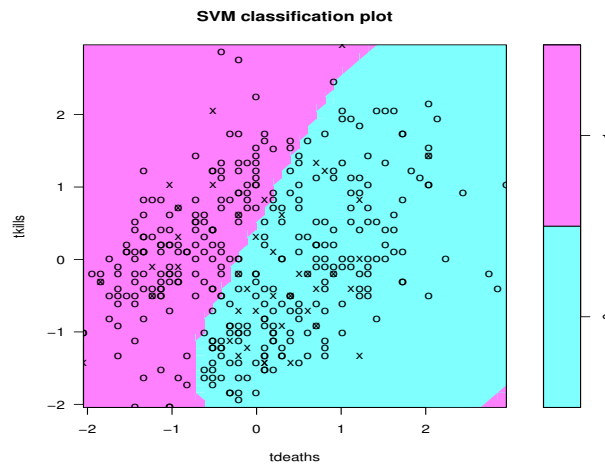


Figure 5.3 Classification plot by SVM

Figure 5.3을 보면 죽은 횟수 (tdeaths)가 증가하면 0(패배)그룹으로 분류될 확률이 높으며, 킬 횟수 (tkills)가 증가하면 1(승리)그룹으로 분류될 확률이 높은 것을 보아 분류가 잘 되어보인다.

5.2.4. 승패예측모형 간 비교

판별 분석과 인공신경망, SVM을 이용한 승패예측모형을 설계하고 각 모형의 예측율을 알아보았다. 모든 모형에서 평균 95%의 예측율을 보였으며 모형 간 정분류율은 크게 차이가 나지 않는 것을 알 수 있었다. 하지만 모형의 예측율에서 패배를 승리로 분류하거나, 승리를 패배로 분류하는 오류율의 차이가 있음을 알 수 있었다. Table 5.9는 승패예측모형 별 오류율을 정리한 표이다. 패배를 승리로 분류하는 오류는 게임 운영에 필요한 전략을 세우는데 있어 위험요소가 될 수 있다. 판별 분석을 이용한 승패예측모형에서 패배를 승리로 분류하는 오류가 7.54%로 가장 높게 나타났으며, 인공신경망 모형이 4.87%로 가장 낮은 수치를 보였다. 인공신경망 모형의 경우 판별 분석과 달리 공분산행렬을 사용하지 않고 학습과정을 통해 최종 모형이 선택되었기 때문에 이 과정에서 오류율이 달라진 것으로 생각된다.

Table 5.9 Comparison of misclassification of models

Original-Classification	Discriminant	Neural Network	SVM
0-1 (Defeat-Win)	7.54%	4.87%	5.3%
1-0 (Win-Defeat)	2.27%	4.2%	5.73%

6. 결론

본 연구는 e-스포츠 분야의 게임인 리그오브레전드의 게임 데이터를 이용해 캐리 레이팅 지표를 개선하고 승패예측모형을 적합해 보았다. 인자 분석을 통해 만든 새로운 지표와 기존 지표의 차이를 비교해 본 결과 기존 지표에 포함되지 않았던 요소들을 보다 다양하게 반영할 수 있었으며, 게임 내 포지션을 대표하는 인자를 선택할 수 있었다.

승패예측모형은 통계적 방법인 판별 분석, 학습 과정을 포함하는 인공신경망 모형과 SVM을 이용한 모형으로 모형 간 차이를 비교해보고, 게임 승패 분석에 가장 적합한 모형을 선택할 수 있었다. 이러한 모형은 게임 승패 예측에 활용될 수 있다. 판별 분석의 경우 평균 95%의 확률로 승패를 예측하였으나, 승리를 패배로 분류하거나 패배를 승리로 분류하는 오류의 정도가 다른 두 가지 모형에 비해 높게 나타났다. 또한 패배와 승리를 분류하는데 확률의 차이가 크게 나타난다는 특징을 파악할 수 있었다. 인공신경망 모형과 SVM을 이용한 모형은 평균 94%의 확률로 승패를 예측했다. 특히 판별 분석과 다르게 패배를 승리로 분류하거나 승리를 패배로 분류하는 오류율의 차이가 적고, 패배와 승리를 분류하는 확률의 차이가 크지 않았다. 이는 판별 분석에서는 공분산행렬을 활용해 분류 모형을 적합하지만, 인공신경망과 SVM은 학습 과정을 통한 추정과 예측 과정이 포함된다는 점에서 분류 확률의 차이가 나타난 것으로 보인다.

향후 연구에서는 학습을 통한 추정과 공분산행렬을 포함한 식의 차이가 어떠한 원인에 의해 발생하는지를 파악해야 할 것이다. 공분산행렬에 따른 분류 정확도의 변화를 알아보기 위해 모의실험이 필요할 것이며, 승패예측모형의 오류율을 줄이기 위한 분석이 이어져야 할 것이다. 또한 시간대별로 게임의 승패 확률을 알아보기 위해서는 다변량시계열분석이 필요해 보인다. 시간을 요인으로 반영하는 승패예측모형을 설계해 현재 모형들과의 차이를 알아보고, 정확도를 향상시킨 모형의 개발을 기대한다.

References

- Cho, D. (2016) The winning probability in Korean series of Korean professional baseball. *Journal of the Korean Data & Information Science Society*, **27**, 663-676.
- Gu, S. H., Kim, H. S. and Jang, S. Y. (2009). A comparison study on the prediction models for the professional basketball games. *Korean Journal of Sport Science*, **20**, 704-711.
- Hastie, T., Tibshirani, R. and Friedman, J. (2001). *The elements of statistical learning*, Springer Verlag, Germany.
- Kang, B., Huh, M. and Choi, S. (2015) Performance analysis of volleyball games using the social network and text mining techniques. *Journal of the Korean Data & Information Science Society*, **26**, 619-630.
- Kim, J. H. (2015). *R multivariate statistical analysis*, Kyowooosa, Seoul.
- Kim, J. Y. and Lee, H. J. (2013). A study of gamebot detection using online game log data analysis. *Proceedings of the Korea Information Science Society 2013 Fall Conference*, 680-682.
- Kim, J. Y. and Lee, H. J. (2014). Gamebot detecting rule verification and gamebot detection using online game log data. *Proceedings of the Korea Information Science Society 2014 Winter Conference*, 835-837.
- Kim, S.-K. and Lee, Y.-H. (2016) The estimation of winning rate in Korean professional baseball league. *Journal of the Korean Data & Information Science Society*, **27**, 653-661.
- Kim, S. H. and Lee, J. W. (2012). Estimating the determinants of victory and defeat through analyzing records of Korean pro-basketball. *Journal of the Korean Data & Information Science Society*, **23**, 993-1003.
- Kim, S. M. and Kim, H. K. (2015). A research on improving client based detection feature by using server log analysis in FPS games. *Journal of the Korea Institute of Information Security and Cryptology*, **25**, 1465-1475.
- Korea Council of Sport for All, Sports Encyclopedia, Available: <http://portal.sportall.or.kr> (downloaded 2016, Aug. 24).
- League of Legend Developers Web Site, Available: <https://developer.riotgames.com/>.
- Lee, C. S. (2008). Evaluation model of on-line game using analytic hierarchy process. *Journal of Global E-Business Association*, **9**, 109-127.
- Oh, S. S. and Kim, D. H. (2012). Analysis of the academic research trend of e-sports. *Journal of Korean Society for Wellness*, **7**, 113-121.
- Park, B. I. (2009). e-Sports value and the controversial issues and solutions for a problem of e-Sports from a sportive point of view. *Journal of Sport and Leisure Studies*, **36**, 101-120.
- Ryu, S. I. and Park, S. J. (2009). Indicator analysis and prediction methods of online games using parametric method and Fourier analysis. *Proceedings of Asia Pacific Journal of Information Systems*, **2**, 466-481.

Development of game indicators and winning forecasting models with game data[†]

Jimin Ku¹ · Jaehee Kim²

¹²Department of Information and Statistics, Duksung Women's University

Received 11 January 2017, revised 10 March 2017, accepted 14 March 2017

Abstract

A new field of e-sports gains the great popularity in Korea as well as abroad. AOS (aeon of strife) genre games are quickly gaining popularity with gamers from all over the world and the game companies hold game competitions. The e-sports broadcasting teams and webzines use a variety of statistical indicators. In this paper, as an AOS genre game, League of Legends game data is used for statistical analysis using the indicators to predict the outcome. We develop new indicators with the factor analysis to improve existing indicators. Also we consider discriminant function, neural network model, and SVM (support vector machine) for make winning forecasting models. As a result, the new position indicators reflect the nature of the role in the game and winning forecasting models show more than 95 percent accuracy.

Keywords: Discriminant analysis, factor analysis, game data, neural network, support vector machine, winning forecasting model.

[†] This research was supported by Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (NRF: 2014R1A1A3050229). This work was also supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20161210200610).

¹ Undergraduate student, Department of Information and Statistics, Duksung Women's University, Seoul 132-714, Korea.

² Corresponding author: Professor, Department of Information and Statistics, Duksung Women's University, Seoul 01369, Korea. E-mail: jaehee@duksung.ac.kr