

데이터 마이닝을 이용한 AOS 장르 게임의 승패 예측<sup>1)</sup>구지민<sup>0</sup> 강현다 유건아

덕성여자대학교

kjm0426@duksung.ac.kr gusek25@duksung.ac.kr kyeonah@duksung.ac.kr

## Prediction of winning of the AOS genre game by using Data Mining

Ji-Min Ku<sup>0</sup> Hyun-Da Kang Kyeonah Yu

Duksung Women's University

## 요 약

AOS(Aeon of Strife)장르는 국내 및 해외에서 가장 인기 있는 온라인 게임 장르 중 하나이며 다수의 플레이어와 함께 게임을 진행하는 게임의 특성 상 서버에 기록되는 대용량의 로그 데이터는 플레이어의 성향, 게임 패턴 등 게임 운영에 필요한 다양한 정보를 담고 있다. 본 논문에서는 사용자가 기존의 게임 데이터를 이용해 게임의 전략적 분석을 실시하고, 분석 결과를 통해 게임 플레이를 발전시킬 수 있게 하기 위해 데이터 마이닝 기법을 이용한 AOS 게임 데이터 분석을 진행하였다. 실험은 기존의 통계적 승률 예측 기법인 로지스틱 회귀분석과 판별분석 및 인공신경망을 이용하여 이루어졌으며 실험 결과, 모든 경우에 평균 87% 이상의 정확도로 승패 예측이 가능함을 보였으며, 각 모형 간 예측률의 차이를 비교해 볼 수 있었다.

## 1. 서 론

온라인 게임이 발전됨에 따라 새로운 장르의 게임들이 생겨나기 시작했다. 대표적으로 MOBA(Multiplayer Online Battle Arena), ARTS(Action Real-time Strategy)등으로 불리는 AOS(Aeon of Strife)장르는 2000년대 초반에 확립된 장르로, 10년 안팎에 불과한 역사에도 불구하고 국내 및 해외에서 가장 인기 있는 온라인 게임 장르 중 하나이다. 다수의 플레이어가 함께 게임을 진행하는 특성상 대용량의 게임 데이터가 서버에 저장되고, 게임사는 데이터를 이용해 게임의 질적 수준 향상과 플레이어의 요구 사항을 업데이트에 반영한다. 게임 유저들은 게임 개발사가 제공하는 API를 이용해 캐릭터의 특성을 분류하고 게임 뱀픽물을 분석 하는 등 활발한 통계적 연구가 이어지고 있다. 본 연구에서는 데이터 마이닝 기법을 통해 AOS 게임 데이터를 이용한 게임의 승률 예측 모형을 추정하고자한다. 본 연구에서 이용하는 데이터 마이닝 기법은 금융, 스포츠 분야에서 승률 예측에 널리 사용되고 있는 로지스틱 회귀분석, 판별분석 등의 통계학적 방법과 인공지능 분야의 학습 방법인 인공신경망이다. AOS 게임 데이터는 게임의 진행 시간과 함께 기록된 경기의 내용이다. 게임 데이터를 일정 시간으로 등분하고 세 가지 방법의 분석을 이용하여 각 타임라인에 대한 게임의 전반적인 운영 방향과 패턴을 알아본다. 또한 시간대별 최적의 승률을 예측하는 모형을 제안하고자 한다. 분석은 오픈 소스 통계 프로그램인 R을 이용해 실시하였으며 세 가지 방법에 대한 예측 정확도를 비교하였다. 분석된 결과는 게임의 리플레이 기능과 함께 이용되어 게임 내용을 확인하고, 차후 게임 전략을 세우는데 도움을 줄 것으로 기대한다.

## 2. 문제 정의

## 2.1 분석 과정의 설계

1) 본 논문은 한국연구재단 연구비지원에 의한 논문임 (과제번호:NRF-2011-0024567)

본 연구에서 사용되는 데이터는 게임의 진행 시간과 함께 기록된 경기의 세부 내용이다. 데이터의 구조 상 다변량시계열분석을 실시 할 수 있지만 본 연구는 시간이 지날수록 감소하는 변량의 효과를 모형 설계에 고려하고자 하였다. AOS 게임의 특성 상 게임 진행 시간이 길어질수록 아이템 빌드의 한계가 나타나거나 캐릭터의 재생 시간이 증가하는 등 변수들이 가지는 영향력이 변하게 된다. 따라서 2분마다 기록되는 게임 데이터를 이용해 각 타임라인에 해당하는 최적의 예측 모형을 결정하고, 아울러 통계 분석을 이용해 각 모형의 정분류율(Correct Rate)을 제공한다.

각 모형의 정분류율은 표본분할을 통한 오분류율(Misclassification rate) 계산을 이용하였다. 표본분할법은 기존 데이터를 분석용과 검증용으로 분류한 후, 분석용 데이터에서 유도된 판별함수를 이용해 검증용 데이터로 정분류율을 계산하는 것이다. 본 연구에서는 분석용과 검증용 데이터를 70:30의 비율로 나누었으며, 정분류율은  $(1 - \text{오분류 되는 경우의 수} / \text{전체 표본의 수})$ 로 계산한 후 백분율 단위로 나타내었다. 본 연구에서는 정분류율을 모형의 예측 성능을 판단하는 평가 지표로 사용한다.

## 2.2 연구 데이터

본 연구에서는 통계 분석에 이용되는 표본으로 리그오브레전드 경기의 게임 플레이 데이터를 선정하였다. 선정 기준은 다음과 같다.

- 리그오브레전드 챔피언스 코리아 스프링 시즌 경기에 속한 게임 - 데이터 분석 중 게임 플레이의 운영적 측면을 반영하기 위하여 현재 리그오브레전드 프로리그 이므로 활동하고 있는 선수들의 경기를 선택하였다.
- 총 플레이 시간이 40분 이내인 게임 - 게임 시간이 30-45분으로 진행될수록 캐릭터가 죽은 후 재생성 되는 시간이 길어지기 때문에 한 번의 싸움으로 게임이 끝나는 경우가 많다. 따라서 분석 결과의 정확성을 높이기 위해 40분 이내에 종료된 게임을 선택하여 모든 정글 몬스터가 등장하는 시간인 20분부터 40분

까지의 게임 데이터를 이용하였다. 이 구간을 2분 간격으로 나누어 총 10개의 구간을 단계별로 분석하였다.

위의 기준을 만족하는 데이터로 각 그룹 당 50개 씩 총 100개의 표본을 추출하였다. 승률에 영향을 주는 변수들은 대전 기록을 통해 확인할 수 있는 변수 중 연속형 변수 6가지( : 팀 골드 획득량,  $X_2$ : 킬 횟수,  $X_3$ : 파괴한 타워의 개수,  $X_4$ : 바론 처치 횟수,  $X_5$ : 드래곤 처치 횟수,  $X_6$ : 역제기 파괴 횟수)를 선택하였다.

### 2.3 변수의 선정

앞에서 제시한 6개의 독립 변수 중에 모형 분석에 사용될 수 있는 변수는 단계적 변수 선택법을 이용해 다음과 같이 선정하였다.

- 로지스틱 회귀모형의 단계적 선택법에서는 AIC(Akaike Information Criterion)값이 가장 작게 나타나는 모델을 사용하였다.
  - 판별분석은 klaR패키지의 stepclass 함수를 이용하였으며, 정확도 값이 가장 높게 나타나는 모형을 최종 모형으로 선택하였다.
  - 인공신경망에서는 1) 모든 변수를 이용, 2) 로지스틱 회귀모형에서 선정된 변수를 이용, 3) 판별분석에서 선정된 변수를 이용하여 3가지 모형으로 선택하였다.
- 각 시간대 별 데이터는 표준 정규화를 통해 변수값의 분포를 표준화하는 작업을 거쳤다.

### 2.4 데이터 마이닝 기법

데이터 마이닝이란 데이터로부터 암묵적이고 잠재적으로 유용한 정보를 추출하는 것으로 통계학적 이론에 기반한 기법들과 인공지능의 기계학습에 기반한 기법들이 있다. 본 연구에서 사용하는 데이터는 결과가 승, 패로 분류되어 있기 때문에 대표적인 분류 기법인 로지스틱 회귀분석, 판별분석, 인공신경망을 이용한다.

로지스틱 회귀분석은 범주형 데이터인 종속변수를 대상으로 한다. 종속변수에 영향을 주는 변수가 무엇이고 그 영향력은 어떤지를 나타내는 선형식을 구하여 독립변수 값이 주어졌을 때 종속변수가 어느 집단으로 분류될 수 있는가를 예측하는 통계적 방법이다.

판별분석은 로지스틱 회귀분석과 같은 목적을 지니며 종속변수가 2개 이상인 경우에 사용한다. 분석하고자 하는 데이터의 특성이 각 독립변수 간에 상관성이 높지 않고 이상점이 없으며 종속변수가 질적 변수일 때 사용할 수 있다.

인공신경망은 생물의 신경망을 모방하여 컴퓨터에 구현한 것이다. 신경망에서 뉴런을 흉내 낸 노드들은 입력층, 은닉층, 출력층으로 구분되어 나열되는데 데이터의 입력에 해당하는 값에 대해 원하는 출력이 나올 수 있도록 링크의 가중치를 조정하는 방법으로 학습을 수행한다.

## 3. 실험 결과

### 3.1 로지스틱 회귀분석 결과

2.3절의 방법으로 선정된 변수들을 이용해 구축된 로지스틱 회귀 모형과 예측률(단위: 백분율(%))은 표 1과 같다.

표 1. 로지스틱 회귀모형과 예측률

구간	time(분)	Logistic function	C.R
1	20-22	$0.18+2.95X_1+0.68X_5$	83.3
2	22-24	$0.14+3.71X_1+1.2X_5$	83.3
3	24-26	$1.04+3.78X_1+2.55X_3+1.10X_5$	83.3
4	26-28	$1.18+6.51X_1+3.2X_3$	90.0
5	28-30	$0.70+4.48X_1+2.39X_3$	90.0
6	30-32	$-0.06+4.13X_1+2.07X_3$	83.3
7	32-34	$1.9+2.4X_1+6.8X_3+1.16X_5$	93.3
8	34-36	$-0.55+2.41X_1+2.85X_2+4.59X_3$	93.3
9	36-38	$-0.13+5.47X_3+1.52X_4$	93.3
10	38-40	$0.6+5.17X_3$	96.7
평균			88.98

게임 전반적으로 예측률은 평균 80-90%으로 나타났으며 시간이 흐를수록 예측률이 상승하는 경향을 보였다. 하지만 게임 중반 시간에 해당하는 구간6에서 예측률이 다소 감소하게 되며, 구간7부터는 다시 증가하게 된다. 이는 게임 중반 부분 캐릭터 재생성 시간이나 한 번의 싸움으로 인한 게임의 승패 변화가 원인으로 판단된다. 구간1,2의 모형을 제외한 나머지 시간대 모형에서  $X_3$ , 즉 파괴한 타워의 개수에 해당하는 변수가 게임 승리에 가장 큰 영향을 미치는 것으로 나타났다.

### 3.2 판별분석 결과

선정된 변수들을 이용해 구축된 판별 모형과 예측률은 표 2와 같다.

표 2. 판별모형과 예측률

구간	time(분)	Discriminant function	C.R
1	20-22	$1.38X_1$	83.3
2	22-24	$1.17X_1+0.06X_5$	83.3
3	24-26	$1.58X_1$	66.7
4	26-28	$1.68X_1$	83.3
5	28-30	$0.86X_1+1.06X_3$	90.0
6	30-32	$1.00X_1+0.87X_3$	83.3
7	32-34	$1.24X_3+0.85X_5$	86.7
8	34-36	$1.65X_3+1.05X_4$	93.3
9	36-38	$1.29X_3+0.77X_5$	93.3
10	38-40	$1.01X_3+0.64X_4$	93.3
평균			87.98

판별 모형의 경우 선형판별분석을 수행하여 추정값을 얻을 수 있었다. 선택된 변수들은 1-2개로 적은 변수들이 선택되었다. 게임 초반에는  $X_1$ , 즉 팀별 골드획득량이 주요 변수에 해당하나, 게임 중반 이후에는  $X_3$ , 즉 파괴한 타워의 개수가 승패 예측에 가장 큰 영향을 미친다는 것을 확인할 수 있었다.

### 3.3 인공신경망 분석 결과

인공신경망 모형은 총 세 가지의 경우를 가지고 비교해 보았다. 신경망 모형1은 모든 변수를 투입하는 주효과 모형을 이용하였으며 모형2와 모형3은 각각 로지스틱 회귀분석과 판별분석에서 사용된 변수들을 입력층으로 이용한 모형을 이용하였다. 출력층의 처리단위의 수는 1

(승리)과 0(패배)의 값으로 부여하였다. 은닉층 처리단위의 수는 입력층과 출력층 처리단위의 평균을 사용하였으며 소수점은 반올림 하였다. R의 nnet 패키지를 이용하여 분석한 결과는 표 3과 같다.

표 3. 인공신경망 모형1, 모형2, 모형3의 구성도와 예측률

구간	입력층-은닉층-출력층			Correct Rate		
	모형1	모형2	모형3	모형1	모형2	모형3
1	5-4-2	2-2-2	1-2-2	86.7	76.7	83.3
2	6-4-2	2-2-2	2-2-2	83.3	80.0	80.0
3	6-4-2	2-3-2	1-2-2	86.7	86.7	90.0
4	6-4-2	4-3-2	1-2-2	86.7	90.0	90.0
5	6-4-2	2-2-2	2-2-2	90.0	91.6	91.6
6	6-4-2	2-2-2	2-2-2	86.7	90.2	90.2
7	6-4-2	3-3-2	2-2-2	91.7	93.3	96.7
8	6-4-2	3-3-2	2-2-2	96.7	96.7	98.4
9	6-4-2	2-2-2	2-2-2	96.7	93.3	86.7
10	6-4-2	1-2-2	2-2-2	95.0	75.0	83.3
평균				90.02	87.35	89.02

인공신경망 모형중에는 전체 변수를 입력층으로 사용한 모형 1의 경우가 가장 예측률이 우수했으며 이는 통계적 방법에 의한 승리 예측보다 높은 예측률을 보였다. 특히 구간8 이후에서는 평균 96%에 해당하는 높은 예측 성능을 보이는 것을 확인할 수 있었다. 게임 중반에 해당하는 구간6에서는 캐릭터 재생성 시간이나 한 번의 싸움으로 인한 게임의 승패 변화로 인해 예측률이 다소 감소하는 경향이 있었다. 이는 로지스틱 회귀분석과 판별분석과도 일치하는 결과이며 5가지 방법의 결과를 시간대별로 나타낸 그래프는 그림 1과 같다.

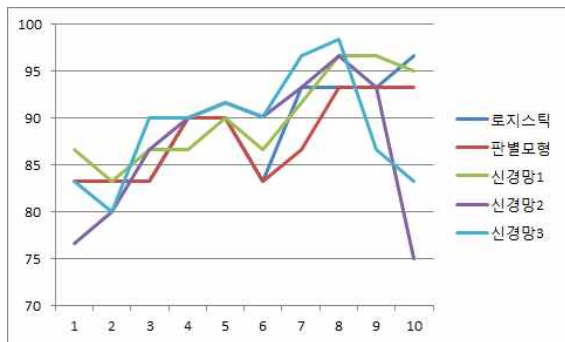


그림 1. 시간대별 예측률 비교

게임 초반의 경우 모든 모형이 90%이하의 예측률을 보이는 것을 알 수 있으며, 시간이 지날수록 예측률이 상승하는 것을 확인할 수 있었다. 특이점은 모든 모형들의 예측률이 구간6에서 감소한다는 점이다. 모형과 시간에 따른 예측률을 이원배치 분산분석으로 비교한 결과, 모형과 시간 별 예측률이 유의미한 차이를 보이는 것을 확인할 수 있었다.

### 3.4 승패 예측 모형의 활용

본 연구는 승패 예측 모형의 결과 게임의 특징적인 패턴을 찾아낼 수 있었다. 대부분의 모형이 구간3~8에서 승패 예측률의 상승 곡선을 보여주고 있는 것으로 보아

플레이어는 구간3~8에 해당하는 시간 사이에 승률의 상승을 유도하기 위한 게임 운영을 실시하는 것이 좋으며, 구간9 이후에는 한 번의 싸움을 유도하여 게임을 역전하거나 승기를 굳히는 것이 중요한 전략으로 사용될 것으로 사료된다.

또한 본 연구는 유도된 예측 모형을 바탕으로 다음과 같이 승패 예측 모형의 활용 방안을 제시하고자 한다.

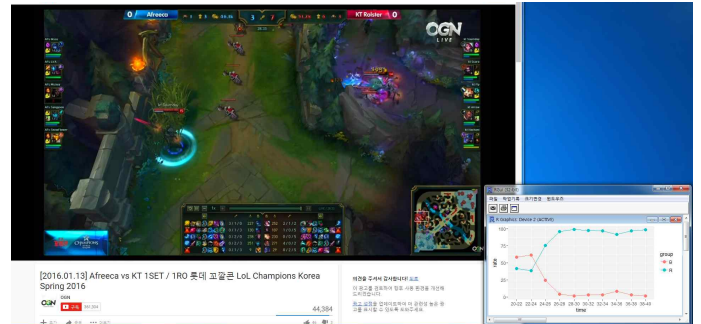


그림 2. 경기 영상과 승패 예측률의 실시간 비교

실제 게임 경기를 진행하는 과정에서는 해설자가 주관적 판단으로 승패를 예측한다. 그러나 그림 2와 같이 실제 경기 화면과 본 연구의 시간대 별 승패 예측 모형을 활용한다면 게임 플레이어는 본인의 경기 화면과 승패 예측률을 실시간으로 비교하여 게임을 모니터링 할 수 있다. 플레이어는 게임의 확률이 변하는 시간대를 확인해 게임의 흐름이 기울어진 원인을 찾을 수 있으며, 향후 게임 운영 방식을 개선할 수 있을 것으로 기대한다.

## 4. 결 론

본 연구에서는 데이터 마이닝 기법들을 이용해 AOS 게임인 리그오브레전드의 데이터를 분석하고 승리 예측 모형을 제시하였다. 각 시간대별 게임 데이터를 활용해 시간대별로 적합한 분류 모형을 설계하였으며, 단계적 선택법을 이용해 최종 변수를 선택하였다. 이 결과 모든 모형이 평균 87%로 높은 예측률을 보였으며, 모형과 시간 별 예측률이 유의미한 차이를 보이는 것을 확인할 수 있었다. 시간에 따라 예측률의 차이가 보여지므로 게임 초반 단계에서는 판별분석 중 단계적 선택법을 이용해 변수를 선택한 신경망 모형을 사용하는 것이 가장 높은 예측 확률을 보이는 것을 알 수 있었다.

향후 연구에서는 인공신경망 모형의 예측력을 높이고, 승패 요인을 수치적으로 나타낼 수 있는 실험을 실시할 것이다. 또한 모형의 변수가 될 수 있는 요인을 데이터를 통해 더 찾아내어 모형의 적합성을 향상시킬 것이다.

## 참고문헌

- [1] 김재희, R다변량통계분석, 서울:교우사, 2015
- [2] 서민구, R을 이용한 데이터 처리&분석 실무, 서울:(주)도서출판 길벗, 2014
- [3] 최규석, 박종진, 인공지능시스템, 경기:21세기사, 2008
- [4] B. D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, 2004
- [5] Christian, R. Nils, R. Karsten, L. Uwe, L. Gero, S. Marc, Z., Classification and visualization, The klaR Package, <http://www.statistik.uni-dortmund.de>