Winner Prediction for Real-time Strategy Games through Feature Selection Based on a Genetic Wrapper

Seung-Soo Shin
School of Software, Kwangwoon University
Seoul, Repulic of Korea
tmd5565@gmail.com

Yong-Hyuk Kim
School of Software, Kwangwoon University
Seoul, Repulic of Korea
yhdfly@kw.ac.kr

ABSTRACT

We propose a method that can predict the game winner using a long short-term memory (LSTM) model that is trained using the match data of a real-time strategy game (Clash Royale). Feature selection based on a genetic wrapper is utilized to identify an excellent feature subset. Subsequently, the performance of a model trained using the data of the aforementioned subset is compared with that of an existing model. The model, which is trained using data that includes the entire features before feature selection, exhibited a winner prediction accuracy of about 60%. Conversely, the model, which is trained using the optimal subset data identified through the application of the genetic wrapper feature selection, exhibited a winner prediction accuracy of about 75%. Based on this comparison result, it was found that the proposed method increased the winner prediction accuracy by about 15% compared to the existing method. In excellent chromosomes explored through genetic wrapper feature selection, the spatial feature was always selected. Simultaneously, the temporal feature was always excluded.

CCS CONCEPTS

• **Computing methodologies** → **Genetic algorithms**; feature selection.

KEYWORDS

genetic algorithm, feature selection

ACM Reference Format:
Seung-Soo Shin and Yong-Hyuk Kim. 2021. Winner Prediction for Real-time Strategy Games through Feature Selection Based on a Genetic Wrapper. In *2021 Genetic and Evolutionary Computation Conference Companion (GECCO '21 Companion), July 10–14, 2021, Lille, France.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3449726.3462735

1 INTRODUCTION

The modern online game industry has been growing constantly; match data can be easily stored and retrieved owing to an increasing number of users and the development of communication technologies and cloud systems. Based on the vast amount of data, chain effects, such as winner prediction, meta-analysis, and intelligent agent establishment, have occurred. In particular, research on winner prediction can provide tactical information to e-sports viewers and be the basis of the development of intelligent agents. Accordingly, research on winner prediction methods based on a combination of match data and machine learning techniques has been conducted. Hodge et al. [1] proposed a winner prediction model for Dota2, a multiplayer online battle arena game. This model combines match data with logistic regression and random forest and can be used during e-sports broadcasting. In this process, the problem of insufficient samples of professional match data was solved through combined utilization of the match data of general users. However, data that can hamper machine learning can be generated when raw match data are used. Such data result in a decrease in the performance of a model and an increase in the amount of training time. To solve the problem of the model using raw data, genetic wrapper feature selection was applied in previous study [2].

2 DATA

Clash Royale[1] is a real-time strategy tower-rush game in which two users place cards that perform fixed actions in real time to destroy towers of the opponent. This game was selected as a demonstration e-sports event for the 18th Asian Games Jakarta-Palaembang. As of 2020, it was regarded as a global mega-hit game because the number of its accumulated registered users exceeded 300,000,000. As for the data used in this study, 130,596 pvp logs collected from the top 100 players between 20/03/2020-13/04/2020, were obtained. Data were collected from the official open API[2] of Clash Royale and RoyaleAPI[3], which is a website providing pvp logs. Match data comprise card's ID played by users and x,y coordinates information according to time. Table 1 shows an example of match data in a sequence type. As for the collected data, 64 features (including 47 properties, 12 stats, 3 spatial, and 2 temporal features for cards according to the rows) were preprocessed to have a sequence type. Standards for properties and stats of cards were prepared based on DeckShopPro[4], the Clash Royale internet forum. As the number of games in the collected data exceeded 100,000 cases, only a range of 85% to 100% was extracted from and used in each game. Moreover, the difference in row length or the number of card placements varies according to the length of the game. However, the length of the input sequence in the machine learning model used in this

Table 1: Examples of match data

Index	Card ID	Bloc	Time	Coordinates
0	28000011	Ally	3.18	(13 , 8)
1	26000023	Ally	3.56	(16 , 8)
2	26000046	Opponent	3.81	(6 , 23)
3	26000000	Ally	4.94	(8 , 11)
4	28000008	Opponent	6.88	(8 , 12)

[1] https://supercell.com/en/games/clashroyale
[2] https://developer.clashroyale.com
[3] https://royaleapi.com
[4] https://www.deckshop.pro/card

GECCO '21 Companion, July 10–14, 2021, Lille, France

ACM ISBN 978-1-4503-8351-6/21/07.
https://doi.org/10.1145/3449726.3462735

study should be fixed. Therefore, padding, which adds values of 0 to the number of the maximum placements when the length of a row is less than the number of the maximum placements, was applied to increase the rows.

3 METHOD

As the match data used in this study were in the form of a sequence, a training process based on a recurrent neural network model (RNN) was found to be appropriate. A pilot experiment using partial data as well as RNN, LSTM, and Bi-LSTM models was conducted; the LSTM, which showed the best performance was selected as the target model in this study. This model was designed to have a many-to-one structure (the input consists of sequence type's stacked rows of each match and the output is the winner of a target game). Feature selection is a technique used to increase performance and reduce processing time in machine learning. Therefore, this study used the wrapper method, which can generate and compare feature subsets to identify the optimal subset. A generational genetic algorithm, which is a global optimization algorithm, was used as the logic for generating and selecting subsets for the wrapper method. In terms of chromosome encoding for the genetic wrapper method, the status of features selected by a target chromosome among the entire 64 features of match data was expressed as a 64-bit string. The fitness of each chromosome is the accuracy of the model trained with the feature subset represented by the chromosome. Each chromosome from the population is matched with a random one and generates two offspring by using one-point crossover and bit-wise mutation of rate 0.1. Among the four chromosomes, the best two ones are selected for the next generation. An experiment was conducted for 100 generations. The number of initial training epochs for the model was set as 25, and it was increased twice every time the generation increased by 25 to reduce time. Through this process of controlling the training epoch number according to the generation, the training time was reduced by about 53% compared to that with the number of epochs fixed at 100.

4 RESULTS AND DISCUSSION

Table 2 compares the model trained using all the features and the model applying feature selection based on our genetic wrapper. It was evaluated that the model trained using all the features showed accuracy and F_1 score of about 60%. Conversely, the model trained features of the best chromosome derived through the application of feature selection based on the genetic wrapper showed a higher accuracy and a higher F_1 score by about 15%. Moreover, excellent chromosomes, which had a fitness of 0.7 or more, representing better performance by 10% than that of the existing model, were stored. A total of 84 excellent chromosomes were stored. Fig. 1 shows a graph of the selection frequencies of features of excellent chromosomes among the entire 64 features. The average of 31 features was selected from excellent chromosomes, and they are represented as a dotted line in the graph. The yellow dot on the bar graph indicates features selected by the best chromosomes. This graph shows that several features with high selection frequencies were selected based on the dotted line. A notable point is that the y coordinate feature was always selected and that temporal feature was always excluded. Fig. 2 shows a graph of the best fitness in the population and the average fitness of an offspring according to generation. The entire fitness increased by about 7%. The fitness of offspring according to generation exhibited a fluctuation with a gradually increasing amplitude. It is placed upward in the right way when it is expressed through linear regression. Therefore, the maximum and minimum values are located in the latter part of a generation. The minimum and maximum values were calculated to be about 0.49 and 0.65, respectively.

Table 2: Comparison between the method using all the features and that using the features selected by our genetic wrapper (LSTM is used as a classifier)

Case		Loss	Accuracy	Precision	Recall	F_1 score
Using all the features	Train	0.6449	0.6074	0.6449	0.6074	0.6074
	Evaluate	0.6507	0.6031	0.6169	0.6031	0.6059
Genetic wrapper	Train	0.5009	0.7551	0.7558	0.7551	0.7550
	Evaluate	0.5252	0.7488	0.7509	0.7488	0.7496

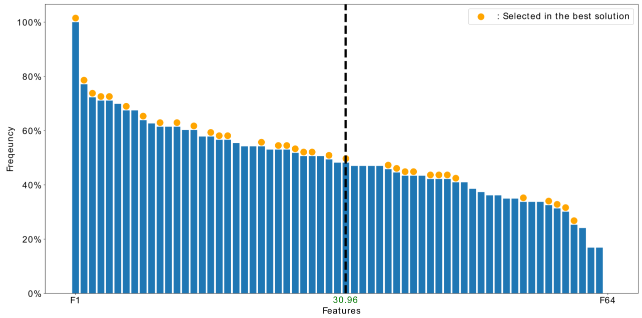

Figure 1: Selection frequencies of features (dotted line indicates the average number of selected features)

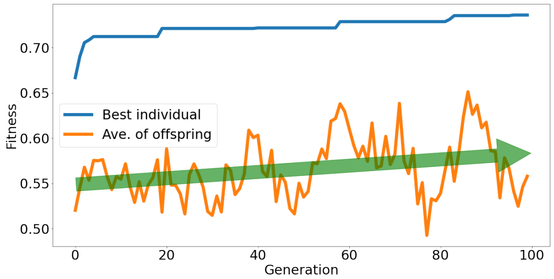

Figure 2: Best fitness in the population and average fitness of offspring according to generation

This study applied a genetic wrapper to a real-time strategy winner prediction model based on the LSTM and obtained a classifier that exhibited better performance than the existing one. Further studies can be conducted to compare their performances of the method of feature selection by wrapper and that by filter. In addition, if images of game replay data are utilized as input, we can expect to get a better model than the presented model.

ACKNOWLEDGMENTS

This research was supported by the MIST (Ministry of Science and ICT), under the National Program for Excellence in SW (2017-0-00096), supervised by the IITP (Institute for Information & communications Technology Promotion)

REFERENCES

[1] V. Hodge, S. Devlin, N. Sephton, F. Block, A. Drachen, and P. Cowling, "Win prediction in esports: Mixed-rank match prediction in multi-player online battle arena games," *arXiv preprint arXiv:1711.06498*, 2017.

[2] D.-H. Cho, S.-H. Moon, and Y.-H. Kim, "A daily stock index predictor using feature selection based on a genetic wrapper," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pp. 31–32, 2020.