

SIAR: An Effective Model for Predicting Game Propagation

Tianyi Wang¹, Guodong Ye¹, Xin Liu^{1*}, Rui Zhou^{1*}, Jinke Li¹, and Tianzhi Wang¹

School of Information Science & Engineering, Lanzhou University, China
zr@lzu.edu.cn, xliu2019@lzu.edu.cn,
WWW home page: <http://dslab.lzu.edu.cn/>
(*Co-corresponding authors)

Abstract. The COVID-19 pandemic has revitalized focus on predictive models, but scant research has been devoted to modeling game transmission, and current models are inadequate in this regard. To predict the spread of games within the population, this paper proposes the "addicted individuals", a new group based on the three groups of the SIR model. We applied the SIAR model, designed based on differential equations, to predict game transmission within this population. The SIAR model was validated on an existing dataset and compared with the traditional SIR model, demonstrating its greater accuracy.

Keywords: game propagation, addicted individuals, differential equations, BFGS algorithms

1 Introduction

Recently, there have been significant improvements in the quality of life of individuals through the use of predictive models[7][2]. Notably, models developed using machine learning and artificial intelligence techniques have become increasingly popular. Such models can learn and reveal patterns hidden in large datasets to anticipate future trends and behaviors. These include intelligent voice assistants, smart home controls, and personalized recommendation systems, all of which are built on predictive modeling technology. They aim to enhance people's lives by making them more convenient, efficient, and comfortable. Predictive models are increasingly expanding into diverse fields, such as healthcare, finance, and education, enhancing people's services and providing more accurate decision support. Therefore, predictive models are playing an increasingly important role in enhancing the quality of life for individuals. In particular, the application of predictive models in controlling infectious diseases has been under the spotlight during the COVID-19 pandemic. For instance, Cooper et al. made significant contributions to the COVID-19 management using their predictive models for control and prevention[6].

Although some models may perform well under specific circumstances, they tend to be ineffective in predicting the spread of games due to the characteristics of game propagation they overlook. Avid gamers commonly experience game

addiction, leading to their affinity towards specific games even after prolonged periods. As a result, the number of players likely remains stable once the game has been in the market for a while, contrasting the spread of infectious diseases. Game development is resource-intensive; both human and material resources are required to create successful games. Additionally, the number of active players is the key revenue driver for game companies. Thus, it is vital for game companies to predict the spread of games among the population. Regrettably, most existing models do not account for or clarify the game addiction phenomenon. The current challenge is to design models that incorporate this phenomenon of game addiction.

The problem solved by predictive models actually involves time series [1]. Various techniques have been proposed, one of which is based on deep learning techniques such as LSTM (Long Short Time Memory)[3] and RNN (Recurrent Neural Network)[20], which has shown good promise in fitting time series data. However, for smaller datasets (only a few hundred points), performance may degrade significantly due to over-fitting[21]. In addition, these models have poor explanatory power, especially for fluctuations in predicted outcomes. Another technique is based on mathematical models, such as the SIR (Susceptible, Infectious, or Recovered) model, which is considered a superior method for predicting the propagation of contagious phenomena like COVID-19[6] and games. In real game scenarios, there are usually players who are very enthusiastic with a particular game and keen to promote their favourite game to those around them. However, these specific features are usually not captured by SIR models, which are often crucial when predicting the spread of games.

To tackle these challenges mentioned above, this paper proposes a new population classification based on the SIR model to elucidate the addiction phenomenon and introduces its own SIAR model, a prediction system that employs a system of differential equations. This model has demonstrated high accuracy in forecasting game spread among the population. Briefly, the contributions of this paper can be summarised as follows:

1. Through an analysis of the communication characteristics of games, we have identified a new group of individuals, referred to as the "addicted ones." This group is essential in explaining the observed phenomenon of a game's player count stabilizing after the fervor for the game has subsided.
2. We propose a novel extension of the SIR model that incorporates a new group of 'Addicted individuals' into our system of differential equations, leading to an improved model's ability to predict the spread of games in the population and provide additional explanatory power. The proposed model has coined the SIAR model.
3. Our proposed SIAR model has been tested on existing datasets and compared against current models, which has demonstrated its superior performance in predicting the spread of games among the population.

2 Related Works

Kermack and McKendrick were the first to propose a mathematical model for describing the spread and control of infectious diseases in a population[9][10][11]. The model categorizes the population into three groups: susceptible, infected, and recovered, and describes the transmission of infectious diseases in the population. Mathematical tools like calculus and difference equations[8] are used to derive the fundamental equations and basic laws of infectious disease spread in the population.

The COVID-19 pandemic prompted the widespread use of disease transmission models, including SIR models, to predict the spread and control of the virus. Cooper et al. applied SIR models to predict the spread of COVID-19[6], while Mwalili et al. used SEIR models to predict the spread of COVID-19 propagation[14]. B Shayak et al. considered the lag between asymptomatic infected individuals and COVID-19 symptoms, added them to the SIR model, and derived predictions and analytical results for the spread and prevalence of the virus by numerical simulation and fitting to actual data. The results highlight the importance of planning and allocating resources for epidemic management [17]. Benjamin F. Maier et al. used the SIR-X model based on SIR models to explain the phenomenon of sub-exponential growth in mainland China during the early stages of the COVID-19 epidemic and to show that this growth was a direct consequence of epidemic control policies[13]. AK Singh et al. introduce a algorithm that uses the differential evolution algorithm in combination with Adam–Bashforth–Moulton method to learn the parameters in a system of variable-order fractional SIR model, which can predict the confirm COVID-19 cases in India considering the effects of nationwide lockdown and the possible estimate of the number of infliction inactive cases after the removal of lockdown on June 1, 2020[18]. And Chen et al. utilized the α -path-based approach to determine the uncertainty distributions and expected values of the solutions. They also applied the method of moments to estimate the parameters and developed a numerical algorithm to solve the model. The proposed model was then used to describe the development trend of COVID-19 in Hubei province by analyzing infected and recovered data[4]. Ram et al. developed a customized age-structured SIR model by considering the social contact and distancing measures in Washington, USA[15].

Previous studies indicate that a variety of SIR-based models have significantly aided in forecasting the transmission of communicable illnesses. Nonetheless, there has been little research done so far on anticipating the propagation of games in societies, and current models are inadequate in projecting the spread of games.

3 The SIR Model

The SIR model is frequently utilized to forecast the spread of infectious diseases in epidemics and has shown success in modeling the spread of COVID-19 [6]. The model categorizes individuals into three groups:

a) Susceptible individuals (S): Defined as individuals lacking immunity who do not currently have the disease but are at risk of contracting it when exposed to infected individuals.

b) Infected individuals (I): Individuals who have contracted the disease and can pass it on to susceptible individuals.

c) Removed individuals (R): Individuals who have recovered from the disease and, consequently, have developed immunity.

Then SIR model is controlled by the following ODE systems:

$$\left. \begin{aligned} \frac{dS(t)}{dt} &= -\frac{\beta IS}{N}, \\ \frac{dI(t)}{dt} &= \frac{\beta IS}{N} - \gamma I, \\ \frac{dR(t)}{dt} &= \gamma I, \end{aligned} \right\} \text{SIR Model's ODE System} \quad (1)$$

where $S(t)$, $I(t)$, and $R(t)$ represent the number of susceptible, infected, and removed individuals at time t , respectively. Here, β represents the rate of infection of susceptible persons by infected individuals per unit of time, while γ represents the rate of recovery of each infected individual per unit of time.

What's more, Fig. 1 provides a graph depicting the conversion of the three populations in the SIR model.

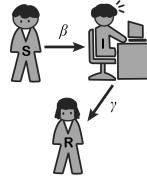


Fig. 1. Conversion of the Three groups in SIR model

By utilizing the SIR model for game prediction, we will redefine the meanings of the three populations:

a) Susceptible individuals (S): People who have yet to be introduced to the game and who are likely to be recommended and subsequently become players.

b) Infected individuals (I): Current players of the game who have a chance of losing interest, in addition to recommending the game to non-players.

c) Removed individuals (R): Past players of the game who have lost interest and will no longer play.

4 Methods

4.1 SIAR Model

Although the SIR model is effective in modeling the spread of infectious diseases, it does not perform well when predicting the diffusion of non-medical phenomena, such as the popularization of a new game.

Therefore, in our modified model, we introduce a new category of individuals:

- Addicted individuals (**A**): People excessively hooked on the game. They have a low probability of losing interest in the game and a high probability of recommending it to others.

The modified model is called the SIAR model, and its differential equations are as follows:

$$\left. \begin{aligned} \frac{dS(t)}{dt} &= -\frac{\beta IS + \epsilon AS}{N}, \\ \frac{dI(t)}{dt} &= \frac{\beta IS + \epsilon AS}{N} - \gamma I - \alpha I, \\ \frac{dR(t)}{dt} &= \gamma I + \theta A, \\ \frac{dA(t)}{dt} &= \alpha I - \theta A \end{aligned} \right\} \text{SIAR Model's ODE System} \quad (2)$$

Here, β and γ have the same meanings as in the previous SIR model. The parameter ϵ represents the probability that each addicted individual will promote the game to others. The parameter α represents the probability of a regular player becoming addicted, while θ represents the probability of an addicted person losing interest in the game.

In order for the model to be better understood, the relationship between the four groups in the model is shown in Fig. 2.

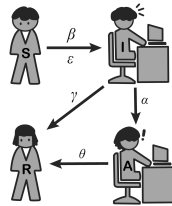


Fig. 2. Conversion of the four groups in SIAR model

4.2 The Solution of SIAR Model

To achieve our aim of predicting the spread of the game across the population, we first defined a loss function to quantify the accuracy of our predictions:

$$Loss(X_{pred}, X_{real}) = \sum_{i=1}^n (x_{pred,i} - x_{real,i})^2 \quad (3)$$

where x_{pred} is a vector indicating the predicted number of individuals who will play the game per day, and x_{real} is a vector indicating the real number of individuals who will play the game per day.

To determine the optimal parameters of our model, we utilize the BFGS algorithm[12], a numerical method to minimize multivariate functions. The BFGS algorithm is a Newton-like method that estimates the Hessian matrix of the objective function so that the search direction of the method can be progressively updated. This method has higher accuracy and a faster convergence rate when compared to the gradient descent algorithm[16].

The central concept of the BFGS algorithm involves an iterative method that approximates the Hessian matrix in the following algorithmic form:

$$B_{k+1} = B_k + \Delta B_k, \quad k = 0, 1, 2, \dots \quad (4)$$

To initiate the iterative process, we choose B_0 to be the identity matrix, I . The rate of BFGS convergence can be increased by appropriately selecting ΔB_k as:

$$\Delta B_k = \alpha \mathbf{u}\mathbf{u}^T + \beta \mathbf{v}\mathbf{v}^T \quad (5)$$

Incorporating Newton's condition results in the following expression:

$$\mathbf{y}_k = B_k \mathbf{s}_k + (\alpha \mathbf{u}^T \mathbf{s}_k) \mathbf{u} + (\beta \mathbf{v}^T \mathbf{s}_k) \mathbf{v} \quad (6)$$

Setting $\alpha \mathbf{u}^T \mathbf{s}_k = 1$, $\beta \mathbf{v}^T \mathbf{s}_k = -1$, $\mathbf{u} = \mathbf{y}_k$, $\mathbf{v} = B_k \mathbf{s}_k$, yields:

$$\alpha = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}, \quad \beta = -\frac{1}{\mathbf{s}_k^T B_k \mathbf{s}_k} \quad (7)$$

Incorporating all the above-step results in the computation formula for ΔB_k :

$$\Delta B_k = \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{B_k \mathbf{s}_k \mathbf{s}_k^T B_k}{\mathbf{s}_k^T B_k \mathbf{s}_k} \quad (8)$$

After obtaining ΔB_k , as an additional step, we can obtain B_{k+1} by applying the recursive formula:

$$B_{k+1} = B_k + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} - \frac{B_k \mathbf{s}_k \mathbf{s}_k^T B_k}{\mathbf{s}_k^T B_k \mathbf{s}_k} \quad (9)$$

Here, $\mathbf{y}_k = B_k^{-1} \Delta x_k$ is the change in gradient, \mathbf{s}_k represents the change in the optimization variables and B_k is the approximation of the Hessian of the gradient.

5 Experiment Results

To further evaluate the validity of the model, we performed a series of tests using the dataset to assess its accuracy and reliability. We compared the model’s performance to other established models, as well as analyzed its ability to generalize to new data and make accurate predictions. Our testing methodology involved a rigorous and comprehensive approach, to ensure the model’s soundness was thoroughly assessed. The results of these tests confirmed the robustness of the model, indicating its suitability for use in real-world applications.

5.1 Dataset

The dataset we used is provided by MCM[5], which contains the number of people playing the game ”Woddle” from Jan 07 2022 to Dec 31, 2022, which is shown in Fig. 3.

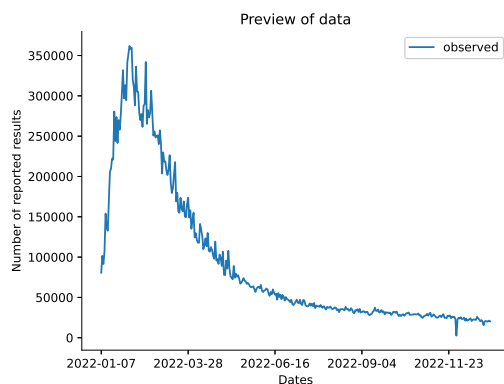


Fig. 3. The preview of dataset

In order to uphold the scientific validity of the experiment, two partitioning methods were implemented on the dataset.

- The first method involved dividing the 359-day dataset into two parts: the initial 299 days were assigned for training the model’s parameters, while the last 60 days were reserved for testing the model’s effectiveness. This method was utilized to assess the model’s predictive performance.
- The second method employed all 359 days of the dataset for both training and testing purposes. This segmentation approach evaluated the model’s ability to accurately fit the dataset.

5.2 Experiment Implementation

For the implementation of the model, we use the `optimize.minimize` function and `integrate.odeint` function of the `scipy` library[19]. The `optimize.minimize` function is used to optimize the parameters of the model and `integrate.odeint` is used to solve the differential equations.

Upon completion of the model training phase, the following formula was utilized to determine the model's level of error:

$$Error(Y_{pred}, Y_{real}) = 2 \sum_{i=1}^n \frac{|y_{pred,i} - y_{real,i}|}{y_{pred,i} + y_{real,i}} \quad (10)$$

5.3 Analysis

Results in the case of splitting dataset After analyzing the outcomes of the SIR and SIAR models with the split dataset, as illustrated in Fig. 5 and 4 and Table 1, it is evident that the SIAR model offers significantly better predictions of the game's transmission phenomenon compared to the SIR model. Through the examination of both SIR and SIAR models, as depicted in Fig. 5 and 4 and Table 1, it is evident that the SIAR model provides a better representation and prediction of the game's transmission patterns compared to the SIR model.

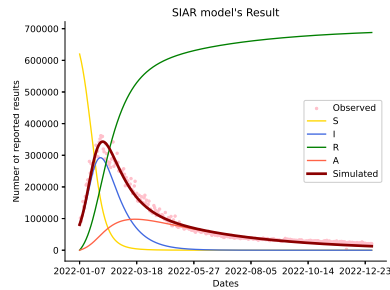


Fig. 4. Results of the SIAR model after splitting the training and test sets

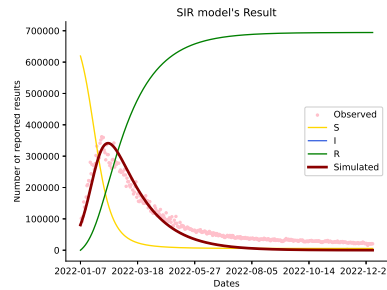


Fig. 5. Results of the SIR model after splitting the training and test sets

Table 1. Error table for SIR and SIAR model

	SIR	SIAR
Train Set Error	0.045	0.004
Test Set Error	0.48	0.099

Results in the case of all dataset When we trained and tested the performance using the full dataset, the results are shown in Fig. 7 and 6 and Table 2. Analysis of the results shows that our SIAR model has a better ability to fit the propagation data of the game compared to the SIR model.

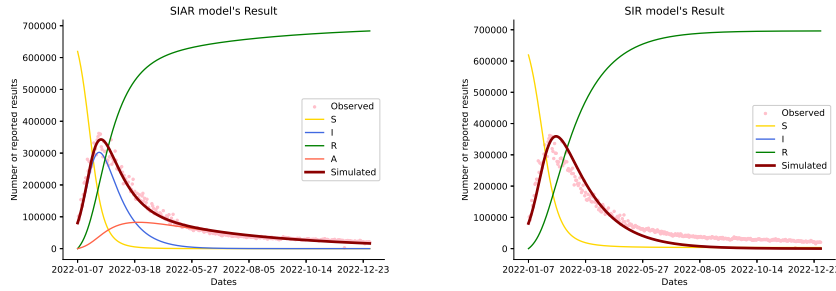


Fig. 6. Results of the SIAR model using all the dataset **Fig. 7.** Results of the SIR model using all the dataset

Table 2. Error table for SIR and SIAR model

	SIR	SIAR
All dataset Error	0.214	0.004

Under both scenarios, it became apparent that the SIR model exhibits a poor fit with the data during the latter half of the time period. We attribute this result to the fact that during this time frame, the majority of players demonstrated strong loyalty towards the game, making the likelihood of becoming disenchanted with the game extremely low. The SIR model is unable to accommodate this framework, as it does not account for loyal players, thereby impeding its ability to predict and fit the game’s transmission in this condition. On the other hand, our SIAR model demonstrated outstanding success in accommodating prevalent loyal players, fitting the dataset consistently well.

6 Conclusion

This paper introduces a SIAR model that predicts the propagation of games among the population. Our model is constructed based on the SIR model while devoted to the gaming-specific attributes in the population. The implementation of our model can lead to more accurate and comprehensible predictions of gaming patterns among crowds. Additionally, we conducted multiple comprehensive tests to establish the dependability and validity of our model.

References

1. Adhikari, R., Agrawal, R.K.: An introductory study on time series modeling and forecasting. arXiv preprint arXiv:1302.6613 (2013)
2. Brynjolfsson, E., McAfee, A.: The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company (2014)
3. Cao, J., Li, Z., Li, J.: Financial time series forecasting model based on ceemdan and lstm. *Physica A: Statistical mechanics and its applications* **519**, 127–139 (2019)
4. Chen, X., Li, J., Xiao, C., Yang, P.: Numerical solution and parameter estimation for uncertain sir model with application to covid-19. *Fuzzy optimization and decision making* **20**(2), 189–208 (2021)
5. COMAP: Predicting wordle results. <https://www.mathmodels.org/Problems/2023/MCM-C/index.html>
6. Cooper, I., Mondal, A., Antonopoulos, C.G.: A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals* **139**, 110,057 (2020)
7. Domingos, P.: The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books (2015)
8. Hethcote, H.W.: The mathematics of infectious diseases. *SIAM review* **42**(4), 599–653 (2000)
9. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character* **138**(834), 55–83 (1932)
10. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character* **141**(843), 94–122 (1933)
11. Kermack, W.O., McKendrick, A.G.: Contributions to the mathematical theory of epidemics-i. 1927. *Bulletin of mathematical biology* **53**(1-2), 33–55 (1991)
12. Liu, D.C., Nocedal, J.: On the limited memory bfgs method for large scale optimization. *Mathematical programming* **45**(1-3), 503–528 (1989)
13. Maier, B.F., Brockmann, D.: Effective containment explains subexponential growth in recent confirmed covid-19 cases in china. *Science* **368**(6492), 742–746 (2020)
14. Mwalili, S., Kimathi, M., Ojiambo, V., Gathungu, D., Mbogo, R.: Seir model for covid-19 dynamics incorporating the environment and social distancing. *BMC Research Notes* **13**(1), 1–5 (2020)
15. Ram, V., Schaposnik, L.P.: A modified age-structured sir model for covid-19 type viruses. *Scientific reports* **11**(1), 15,194 (2021)
16. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016)
17. Shayak, B., Sharma, M., Rand, R.H., Singh, A.K., Misra, A.: Transmission dynamics of covid-19 and impact on public health policy. *MedRxiv* pp. 2020–03 (2020)
18. Singh, A.K., Mehra, M., Gulyani, S.: A modified variable-order fractional sir model to predict the spread of covid-19 in india. *Mathematical Methods in the Applied Sciences* **46**(7), 8208–8222 (2023)
19. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern,

- R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17**, 261–272 (2020). DOI 10.1038/s41592-019-0686-2
20. Zaremba, W., Sutskever, I., Vinyals, O.: Recurrent neural network regularization. arXiv preprint arXiv:1409.2329 (2014)
 21. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM* **64**(3), 107–115 (2021)