Advances and Applications in Mathematical Sciences
Volume 21, Issue 7, May 2022, Pages 4197-4209

HYBRIDIZATION OF DRAGONFLY ALGORITHM FOR VIDEO POPULARITY FORECAST

NEETI SANGWAN and VISHAL BHATNAGAR

University School of Information
Communication and Technology
Guru Gobind Singh Indraprastha
University Dwarka
New Delhi-110078, India

and

Maharaja Surajmal Institute of
Technology Janakpuri
New Delhi-110058, India
NSUT East Campus
(Formerly Ambedkar Institute of Advanced
Communication Technologies
and Research) New Delhi-110031, India

Abstract

There is intense development of competition among the publishers of the online content to pull the attention of the viewers. It has been perceived that some amount of online content earns the popularity while the rest content stays obscure by the large folk. Endeavors are applied by the researchers to know the future popularity pattern that a content would achieve in future. Distinct researchers provided the methods to forecast the popularity of the content utilizing various intrinsic and extrinsic parameters. In this paper, authors proposed the approach to optimize the process of anticipating the popularity of the video content available online. Authors provide the hybrid approach based on dragonfly metaheuristic algorithm to optimize the popularity of the videos.

2020 Mathematics Subject Classification: 68T09.

Keywords: Video, Popularity, Regression, Meta-heuristic, Prediction.

Corresponding author; E-mail: neetisangwan@gmail.com, vishalbhatnagar@yahoo.com

Received September 25, 2021; Accepted November 21, 2021

1. Introduction

With technological enhancement, online content is growing at an exponential rate. Predicting the future popularity of the online data is required to prioritize the data on the basis of its future usefulness. Future popularity helps deciding that data need to be kept in the database and to discard it. Prominent amount of online content consists of videos [1] [2]. Prior anticipation of the future fame helps the online video service providers, advertisers, online social networks to make appropriate decision and gain maximum. Several methods exist that can be used in order to forecast the popularity to be gained by the videos in future. Feature selection procedure can also be applied prior to the prediction for further optimized results. It selects the features that have prominent influence on the prediction performance and saves the resources.

The process of popularity prediction can be optimized using nature-inspired algorithms (NIA). NIA helps selecting the influential features from large number of features. Optimized feature selection leads to better results for prediction in less computational requirements. The task of optimized forecasting of the video popularity can be seen as two-step process: feature selection and regression based prediction. Feature selection process helps in extracting the features that have major impact on the results. Regression method is applied to find the relation among distinct features of the videos. The initial step included for prediction is to preprocess the dataset and to select the features that affect the results significantly. Afterwards, regression has been performed to predict for popularity. In this paper, distinct NIAs are exploited to pursue this process in the right direction. In this paper, we use GA, PSO and BA hybridized with DA to provide two-fold feature selection following the regression based prediction.

In section 2 literature related to the popularity prediction and nature-inspired algorithms has been provided. Section 3 deals with the NIAs utilized in the proposed work, distinct proposed approaches and results obtained from these proposed approaches on a dataset are discussed. Section 4 concludes the work and provide some future directions and scope of the work.

2. Literature Review

Expansion in the number of online users prompt the researchers to concentrate towards predicting the popularity of the online content. The essence of this section is to explore the contribution of the distinct researchers in video popularity prediction. This section also illuminates the distinct aspects related to video popularity prediction.

Various researchers revealed about the dependency of the future popularity of the video on its early views and other properties of social networks.

Logarithmic correlation existed between early and future popularity [1]. Multivariate linear regression [2] and Evolution Pattern and Burst Prediction based Multivariate Linear (EPBP_ML) regression model [3] has been given by the researchers for popularity prediction. Figeureigo et al. [4] generalized the evolution pattern of video fame. Distinct approaches: K-spectral clustering [5], log normal distribution [6], power law [7] has been utilized to find popularity patterns. Trzcinski et al. [8] proposed Support Vector Regress or (SVR) to predict popularity using visual and social features of videos. Relation between video propagation and popularity in the micro-blogging system has been represented in [9].

Gamma distributions [10], Hawkes Intensity Process Insights Explorer (HIPie)-an interactive visualization system [11], Multifactor differential influence (MFDI) prediction model [12] are given by the researchers for making video related predictions. An approach has been presented for video transcoding time prediction and scheduling for HTTP adaptive streaming videos [13]. Researchers provided Complexity class based transcoding time prediction for video sequences using artificial neural network [14]. Suspicious human activity recognition system has been presented in [15].

In this paper, Nature Inspired Algorithms (NIA) are also investigated to find its advantages for video popularity forecast. NIAs have proved their benefit in different fields including optimization problems [16]. Genetic algorithm (GA) [17] provided a compact algorithm based on the biological genetic process. It mimics the selection, crossover and mutation behavior of the simple Genetic process. Harmony search (HS) [18] is an algorithm based on improvement in music performance process. It works on the principle of searching for better harmony in music. Grey Wolf Optimization (GWO) [19] idealized the leadership levels and hunting procedure of the grey wolves. Particle Swarm Optimization (PSO) is an algorithm motivated from the movement of the birds' flock. Using this algorithm, result can be optimized by updating the position of different birds [20]. Bat Algorithm (BA) is inspired by the pattern that bats exploited while looking for the food [21]. Dragonfly Algorithm (DA) follows the pattern of communication done by the dragonflies to guide the swarm towards food and keep away from the enemies without collision with each other [22]. Artificial Bee Colony (ABC) [23] is a method that mimics the intelligent behavior of bees' swarm to locate the best food source. Cuckoo Search (CS) [24] utilized the analogy of breeding behavior of the biological cuckoo species. It idealized the way in which cuckoo lays the egg and put the egg in arbitrarily chosen host nest. Host of nest may identify the egg of cuckoo and discard the egg or leave the nest. Nest with high quality eggs is selected to pass it to the next generation. Whale Optimization Algorithm (WOA) [25] is inspired from the hunting method adopted by the whales. Whales find its prey location and target them. Other search agents update their position according to the position of best search agent. Moth Flame Optimization (MFO) [26] exploited the concept of the movement of moths in night by keeping up a fixed angle to the moon. Elephant Herding Optimization (EHO) [27] is the idealization of herding behavior of elephants. Different groups of elephants live together under the initiative of an authority. Male elephants are supposed to leave their family. Two operators are maintained to manage the group by the leader and to separate the male elephant from the family on the growth. Ant Lion Optimizer (ALO) [28] utilized the hunting method of the antlions in which antlions digs a hole in the sand and hides them under the hole. Crow Search Algorithm [29] idealized the intelligent behavior of crow to find the food and hiding that food for future use. In this search agent update its position of hiding the food. Hybrid ABC/DA (HAD) [30] utilized the ABC and DA to overcome the issue of slow convergence. Linearly Decreasing Weight Particle Swarm Optimization (LDWPSO) [31] is a nature inspired optimization of Convolutional Neural Network (CNN). Hyper parameters of CNN are optimized using LDWPSO.

On exploring this literature survey, it has been noticed that NIAs can be utilized for optimization in distinct fields. In this paper, the authors have utilized the characteristics of distinct NIAs to improve video popularity forecast by optimizing the features of the video to participate in the prediction.

3. Proposed Work

As described in previous section, NIAs finds its utility in distinct fields for several engineering problems: feature selection, image processing, optimization. In this paper, distinct NIAs are utilized to propose the hybrid approaches that optimize the process of popularity prediction of the online video content. Proposed approaches exploited the GA, PSO, BA to be hybridized with DA to provide optimized selection of the features. These selected features are then participating in the prediction process to give improved results.

3.1 Empirical Evaluation

For evaluation of the proposed approaches i.e. GA-DA-SVR, PSO-DA-SVR and BA-DA-SVR, we worked upon the video dataset namely: Facebook 2015. Data set consists of 1470 videos with 702 distinct features.

3.1.1 Particle Swarm Optimization PSO is a metaheuristic method introduced by Eberhat and Kennedy [20] that makes use of flying pattern of the group of birds in a search space. In this, the movement of each particle is regulated by its current position, its best position, its velocity and the best position of the flock. Each particle in a search space is represented by D-dimensional vector. Position vector of the i^{th} particle can be represented by $P_i = (P_{i1}, P_{i2}, \ldots, P_{iD})$. Particle best position can be calculated by comparing the fitness value of current position with the fitness value of previous best position of that particle and represented by $B_i = (B_{i1}, B_{i2}, \ldots, B_{iD})$. The updation of the best position achieved by the particle is given by equation (1).

$$B_i = P_i \text{ if } f(P_i) > f(B_i) \tag{1}$$

Best position value achieved among all the particles is known as Swarm best position, represented as $S_i = (S_{i1}, S_{i2}, \ldots, S_{iD})$. The value of S_i is updated comparing the best value achieved by all the particles in the swarm at each instance. Best position vector of all the particles are compared on the basis of fitness value as in equation (2).

$$S_{id} = B_{id} \text{ if } f(B_{id}) > f(S_{id}) \quad (2)$$

The best position of a particle and swarm are responsible for finding the rate at which position of the particle will be updated. The rate of change of position for particle is represented by $R_i = (R_{i1}, R_{i2}, \ldots, R_{iD})$. The rate of change of position at timestamp $n+1$ can be calculated as in equation (3)

$$R_{id}(n+1) = w * R_{id}(n) + C_1 * rand(\ )B_{id}(n) - P_{id}(n) + C_2 * rand(\ )$$
$$S_{id}(n) - P_{id}(n) \quad (3)$$

where d ranges from 1 to D, n is the previous timestamp, w is the inertial constant, C_1, C_2 are the constants used to increase or decrease the ability of the particles to search and rand( ) generates a random number ranging between 0 and 1. The next position of i^{th} particle will be updated as in equation (4):

$$P_{id}(n+1) = P_{id}(n) + R_{id}(n+1) \quad (4)$$

3.1.2 Bat Algorithm

BA is a fascinating metaheuristic algorithm that mimics the procedure pursued by the group of bats in order to look for food [21]. It is based on two special characteristics of bats: echolocation and orientation mechanism. Capability of echolocation in bats pulled the attention of the researchers to take their time. Echolocation is the phenomenon in which bats radiate an uproarious and short beat of sound whose echo comes back on hitting an object. In this manner bats are able to estimate the distance from the object. Using orientation mechanism, bats are able to distinct between the food and obstacles. Orientation mechanism makes bats to search food even in dark. A bat flies in D-dimensional search space with some velocity $V_i = (V_{i1}, V_{i2}, \ldots, V_{iD})$ at position $P_i = (P_{i1}, P_{i2}, \ldots, P_{iD})$. Updation of velocity of a bat depends on the frequency and pulse rate of emitted pulse. Velocity of each bat is refurbished utilizing the velocity of the bat at timestamp n, bat's best position $B_i = (B_{i1}, B_{i2}, \ldots, B_{iD})$, swarm's best position $S_i = (S_{i1}, S_{i2}, \ldots, S_{iD})$ and frequency F_i. Velocity at timestamp

$n+1$ can be calculated using the equation (5) and equation (6), where d ranges from 1 to D.

$F_i = F\min + (F\min - F\max)\beta$ where β is random vector ranging $[0, 1]$ (5) $V_{id}(n+1) = (S_{id}(n) - B_{id}(n)) + F_{id}(n)$. (6) As bat moves towards the target, loudness decreases and rate of pulse emission increases. Loudness L_i and pulse emission rate E_i has been updated in each iteration using equation (7) and equation (8) respectively. $L_i(n+1) = C_3 L_i(n)$ where C_3 is constant (7) $E_i(n+1) = E_i(0)[11 - \exp(-C_4(n))$ where C_4 is constant (8)

Position of bat is updated with increase in timestamp from n to $n+1$ by using equation (9)

$$P_{id}(n+1) = P_{id}(n) + V_{id}(n+1) \tag{9}$$

3.1.3 Dragonfly Algorithm

DA is a swarm based method that finds out the global optimum solution to the problem [22]. It makes the use of distinct standards followed by the dragonflies to move in a group without colliding with each other. Dragonflies maintain communication among each other to fascinate the swarm towards the food and deflected all the dragonflies in a group from their enemies. The DA consists of D-dimensional search space in which dragonflies move. A repetitive procedure is followed to determine the movement of candidate solutions (dragonflies) by calculating separation between each dragonfly and its neighbors, alignment of each dragonfly with respect to its neighbor, attraction, cohesion towards the food and repulsion from the enemies. Separation $S_i = (S_{i1}, S_{i2}, \ldots, S_{iD})$ maintained among the dragonflies to avoid collision is dependent on the current location of each dragonfly and its neighbor. It can be mathematically represented by equation (10) where $D_i = (D_{i1}, D_{i2}, \ldots, D_{iD}), D_j = (D_{j1}, D_{j2}, \ldots, D_{jD})$ are the position of i^{th} and j^{th} dragonfly (neighbor of i^{th} dragonfly varies from $j = 1$ to N) respectively and N determines the neighborhood size.

$$S_i = \sum_{j=1}^{N} (D_i - D_j) \tag{10}$$

Alignment A_i, is the aligning motion of i^{th} dragonfly to match its velocity with the velocity of other search agents present in the neighborhood, can be calculated as in equation (11) where V_j denotes the velocity of j^{th} neighboring dragonfly and N is the neighborhood size.

$$A_i = \frac{\sum_{j=1}^{N} V_j}{N} \tag{11}$$

Cohesion is another property of the dragonflies in which each search agents fly towards the neighborhood's center of mass. Cohesion can be mathematically represented by equation (12).

$$C_i = \sum_{j=1}^{N} \frac{(D_i - D_j)}{N} \tag{12}$$

Where D_i and D_j denotes the current position of i^{th} dragonfly and j^{th} neighboring dragonfly respectively. Attraction, Atti is the ability of i^{th} search agent to get attracted towards the food. It can be calculated using position of food, D_{food} and current position of i^{th} individual D_i and presented in equation (13).

$$A_{tti} = D_{food} - D_i \tag{13}$$

Similarly, each dragonfly maintains the distance from the enemy. Distraction is the ability of each search agent to fly away from an enemy. It can be calculated using the position of enemy and the current position of the search agent. In equation (14), Distraction of i^{th} dragonfly, D_{isatti} is presented where D_{enemy} denotes the position of enemy and D_i signifies the current location of the i^{th} dragonfly.

$$D_{isatti} = D_{enemy} - D_i \tag{14}$$

DA uses Step vector and position vector for feature selection. Equation (15) defines the Step vector where s, a, c, f, e, n represents separation weight, alignment weight, cohesion weight, food factor, enemy factor, inertial weight and iteration number respectively.

$$\Delta D_i(n+1) = sS_i + aA_i + cC_i + fA_{tti} + eD_{isatti} + w\Delta D_i(n) \tag{15}$$

Position vector is updated in an iteration by summation of its step vector and previous position vector and represented using equation (16).

$$D_i(n+1) = D_i(n) + \Delta D_i(n+1) \tag{16}$$

3.2 Approaches

In this paper, authors proposed three approaches formulated on the hybridization of the algorithms discussed in previous section: PSO, GA, BA, DA. Proposed hybrid approaches has been discussed in this section.

Approach 1. GA-DA-SVR: This approach optimizes the prediction process by selecting the prominent features using nature-inspired algorithms: Genetic Algorithm and DA. After this two-fold selection process, selected features undergo regression for the prediction results.

Approach 2. PSO-DA-SVR: Using this approach, relevant features are shortlisted using nature-inspired algorithms: PSO and DA. After this two-fold selection process, selected features undergo regression for predicting the popularity of the videos.

Approach 3. BA-DA-SVR: BA-DA-SVR is an approach that utilizes two nature-inspired algorithms: BAT and DA to select the most influential set of features among all the features of the videos. After this two-fold selection process, selected features undergo regression to generate the popularity score of all the considered videos.

4. Results

For simulation of the proposed approaches, Facebook 2015 data set has been utilized. This dataset consists of 1470 videos with 702 distinct temporal and visual features. Simulation of the approaches has been conducted on the Google Colaboratory that provide Graphical Processing Unit and Tensor Processing Unit support. Packages of python that are used in the experimentation includes Numpy, Pandas, Scikit learn, Json. and Spacy. Performance of the Base algorithm (SVR) and proposed approaches is evaluated on the basis of Coefficient of determination (R^2) at different regularization factor, $C = 1000, 100, 10, 0.1$ R^2 score is a parameter that evaluate the performance of distinct approaches for popularity prediction by measuring the extend of a model to predict the dependent variables using independent variables. Table 1 manifested the value of R^2 score using distinct SVR kernels, namely: linear, Radial Basis Function (RBF) and number of prominent features selected to participate in each approach. Figure 1 provides the graphical representation of R^2 score using SVR kernel: RBF.

Table 1. Performance evaluation using distinct kernels.

Approach	R^2					Features selected
	SVR Kernel					
	Linear	RBF	RBF	RBF	RBF	
		C=1000	C=100	C=10	C= 0.1	
SVR	99.71	78.25	78.25	76.76	41.46	702
GA-DA-SVR	99.06	82.34	82.34	80.72	44.09	220
PSO-DA-SVR	99.69	82.54	82.54	80.94	44.06	115
BA-DA-SVR	99.59	82.72	82.72	81.26	44.58	103

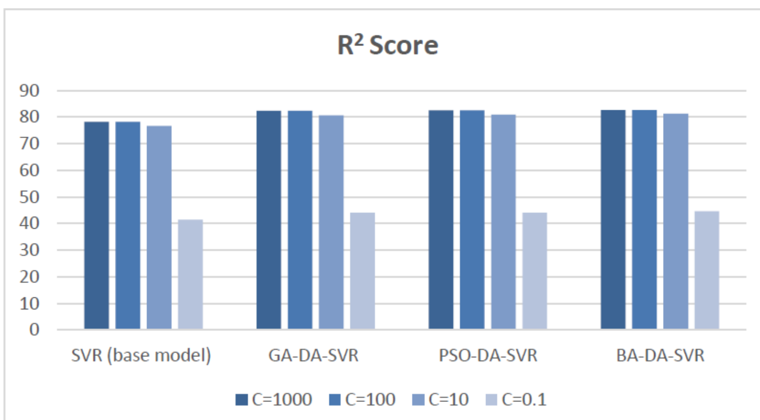

Figure 1. R^2 score for 'RBF' kernel.

5. Conclusion and Future Work

In this paper, DA is hybridized with GA, PSO and BA to propose three novel hybrid approaches for optimized popularity prediction. Authors implemented GA-DA-SVR, PSO-DA-SVR and BA-DA-SVR approaches using google colaboratory, a hosted jupyter notebook environment. Results obtained for the proposed approaches are contrasted with each other and base SVR model for popularity prediction. Performance is evaluated on the basis of two parameters: R2 score and the number of features needed in each approach for making predictions. From the results it has been observed that approach BA-DA-SVR performs better than SVR, GA-DA-SVR, PSO-DA-SVR using 'RBF' kernels in terms of R2 score. Also, BA-DA-SVR requires less number of features than other discussed approaches to give optimized results.

In this paper, authors intend to find the different hybrid nature-inspired algorithms to forecast the fame to be gained by the online videos. But, still predictions can be further improved by expanding the generations and data points. Other NIAs and prediction methods can be investigated.

References

[1] G. Szabo and B. A. Huberman, Predicting the popularity of online content, Commun. ACM 53(8) (2010), 80-88.

[2] H. Pinto, J. M. Almeida and M. A. Gonçalves, Using early view patterns to predict the popularity of YouTube videos, in Proceedings of the sixth ACM international conference on Web search and data mining-WSDM' 13, 2013.

[3] C. Li, J. Liu and S. Ouyang, Characterizing and predicting the popularity of online videos, IEEE Access 4 (2016), 1630-1641.

[4] F. Figueiredo, F. Benevenuto and J. M. Almeida, The tube over time: Characterizing popularity growth of YouTube videos, in Proceedings of the fourth ACM international conference on Web search and data mining - WSDM' 11, 2011.

[5] F. Figueiredo, J. M. Almeida, M. A. Gonçalves and F. Benevenuto, Trend Learner: Early prediction of popularity trends of user generated content, Inf. Sci. (Ny) 349(350) (2016), 172-187.

[6] Y. Borghol, S. Mitra, S. Ardon, N. Carlsson, D. Eager and A. Mahanti, Characterizing and modelling popularity of user-generated videos, Perform. eval. 68(11) (2011), 1037-1055.

[7] M. Cha, H. Kwak, P. Rodriguez, Y.-Y. Ahn and S. Moon, Analyzing the video popularity characteristics of large-scale user generated content systems, IEEE ACM Trans. Netw. 17(5) (2009), 1357-1370.

[8] T. Trzcinski and P. Rokita, Predicting popularity of online videos using support vector regression, IEEE Trans. Multimedia 19(11) (2017), 2561-2570.

[9] Z. Wang, L. Sun, C. Wu and S. Yang, Guiding internet-scale video service deployment using microblog-based prediction, in 2012 Proceedings IEEE INFOCOM, 2012.

[10] A. Tatar, M. D. de Amorim, S. Fdida and P. Antoniadis, A survey on predicting the popularity of web content, J. Internet Serv. Appl. 5(1) (2014).

[11] Q. Kong, M.-A. Rizoiu, S. Wu and L. Xie, Will this video go viral: Explaining and predicting the popularity of YouTube videos, in Companion of the Web Conference 2018 on The Web Conference 2018 - WWW' 18, 2018.

[12] Z. Tan and Y. Zhang, Predicting the top-N popular videos via a cross-domain hybrid model, IEEE Trans. Multimedia 21(1) (2019), 147-156.

[13] P. Agrawal, A. Zabrovskiy, A. Ilangovan, C. Timmerer and R. Prodan, FastTTPS: fast approach for video transcoding time prediction and scheduling for HTTP adaptive streaming videos, Cluster Comput., 2020.

[14] A. Zabrovskiy, P. Agrawal, R. Matha, C. Timmerer and R. Prodan, ComplexCTTP: Complexity class based transcoding time prediction for video sequences using artificial neural network, in 2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM), 2020.

[15] A. S. Ben-Musa, S. K. Singh and P. Agrawal, Suspicious Human Activity Recognition for Video Surveillance System, in Proc. of the International Conference on Control, Instrumentation, Communication and Computing Technologies (ICCICCT-2014), 2014.

[16] I. Fister, I. Fister Jr, X.-S. Yang and J. Brest, A comprehensive review of firefly algorithms, Swarm Evol. Comput. 13 (2013), 34-46.

[17] D. E. Goldberg and J. H. Holland, Genetic algorithms and machine learning, Mach. Learn. 3(2-3) (1988), 95-99.

[18] Z. W. Geem, J. H. Kim and G. V. Loganathan, A new heuristic optimization algorithm: Harmony Search, Simulation 76(2) (2001), 60-68.

[19] S. Mirjalili, S. M. Mirjalili and A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014), 46-61.

[20] R. Eberhart and J. Kennedy, A new optimizer using particle swarm theory, in MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 2002.

[21] X.-S. Yang, A new metaheuristic bat-inspired algorithm, in nature inspired cooperative strategies for optimization (NICSO 2010), Berlin, Heidelberg: Springer Berlin Heidelberg (2010), 65-74.

[22] S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems, Neural Comput. Appl. 27(4) (2016), 1053-1073.

[23] D. Karaboga and B. Basturk, On the performance of artificial bee colony (ABC) algorithm, Appl. Soft Comput. 8(1) (2008), 687-697.

[24] X.-S. Yang and S. Deb, Cuckoo search via Lévy flights, in 2009 World Congress on Nature and Biologically Inspired Computing (NaBIC), 2009.

[25] S. Mirjalili and A. Lewis, The whale optimization algorithm, Adv. Eng. Softw. 95 (2016), 51-67.

[26] S. Mirjalili, Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm, Knowledge Based System 89 (2015), 228-249.

[27] G.-G. Wang, S. Deb and L. dos S. Coelho, Elephant herding optimization, in 2015 3rd International Symposium on Computational and Business Intelligence (ISCBI), 2015.

[28] S. Mirjalili, The ant lion optimizer, Adv. Eng. Softw. 83 (2015), 80-98.

[29] A. Askarzadeh, A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm, Comput. Struct. 169 (2016), 1-12.

[30] W. A. H. M. Ghanem and A. Jantan, A cognitively inspired hybridization of artificial bee colony and dragonfly algorithms for training multi-layer perceptrons, Cognit. Comput. 10(6) (2018), 1096-1134.

[31] T. Serizawa and H. Fujita, Optimization of convolutional neural network using the linearly decreasing weight particle swarm optimization, arXiv [cs.NE], 2020.