The Use of Genetic Algorithms for Automated Machine Learning in Trend Prediction in Time Series Data: A Review

James Taljard
University of Cape Town
Cape Town, South Africa
tljjam001@myuct.ac.za

ABSTRACT

In recent times, machine learning algorithms have been applied to an array complex real-world problems in various knowledge domains. In order the effectively use these algorithms, optimal algorithm hyperparameters need to be selected as the selection of these hyperparameters has significant impact on the performance of these machine learning algorithms. AutoML aims to automate the selection of these algorithms and their hyperparameters in order to democratise the use of machine learning methods by allowing users to need little-to-no knowledge of machine learning theory in order to effectively use machine learning models. Genetic algorithms have been used successfully used in AutoML. This paper focuses on the analysis of literature regarding the use genetic algorithms for AutoML when applied to problems of forecasting time series data, and particularly, trend prediction. We find that genetic algorithms have been successfully applied in AutoML for time series prediction tasks. However, its applicability to trend prediction is still largely unstudied.

KEYWORDS

AutoML, Genetic Algorithms, Hyperparameter Optimization, Trend Prediction, Time Series.

1 INTRODUCTION

In recent years, the use of machine learning (ML) algorithms has become prevalent in multiple application domains including natural language processing, computer vision, object detection, and advertising. This is because ML, and more specifically deep-learning (DL), algorithms can be applied to various problem and data types while often achieving better results than human experts [7,17]. However, designing and implementing an effective ML model usually requires human ML experts to manually select the model algorithm and tune its hyperparameters, usually through a trail-and-error approach, in order to achieve an optimal model architecture. Common ML algorithm hyperparameters include learning rate, batch size, number of hidden neural network (NN) layers, and number of neurons per hidden layer in the NN [9]. These hyperparameters are used to alter the behaviour of the ML algorithm in some specific way and must be set by the model designer before training can begin [15].

Recently, there has been an increasing interest in automating the various stages of the development of an ML model. These stages include data preparation, algorithm selection, and hyperparameter selection [7]. This automated machine learning (AutoML) has democratised the use of ML in various application domains by allowing non-experts to utilize various ML algorithms effectively within their knowledge domain [10]. Hyperparameter optimisation (HPO) is a subclass of AutoML which aims to the automate the selection of ML algorithm hyperparameters that will give an optimal ML model for a given data set and ML algorithm. Various methods are used for HPO which include grid search, Bayesian optimisation, and meta-heuristic methods like particle swarm optimisation and genetic algorithm [6].

A genetic algorithm (GA) is a subclass of evolutionary algorithms that draw inspiration from the concept of natural selection. A typical GA begins with a group of chromosomes, known as the population; a chromosome is an encoding of a solution to a given problem. The algorithm generates a new population from the old one by generating new chromosomes using the genetic operations of selection, crossover, and mutation. New chromosomes are evaluated by some fitness score in order to determine whether or not it should be added to the population. Once the algorithm has terminated, the fittest chromosome is the final solution to the problem[13,16,18] The use of GAs has become a popular method for HPO since they have the capacity to solve non-convex, non-continuous, non-smooth optimisation problems [17]

A time series is a series of values of a quantity taken at successive time intervals, the time intervals are often equal in length. A trend in time series data is a long-term increase or decrease in the data and understanding these trends can improve our forecasting ability. Due to the complex nature of real-world systems that produce time series data and the problems associated with them, the use of various ML algorithms to perform the task of time series analysis and trend prediction have been demonstrated with varying levels of success [4]. However, the hyperparameters of these algorithms must usually still be manually tuned by ML experts.

The focus of this paper is the review of literature regarding the use of GA for AutoML for trend prediction. In section 2.1 we analyse the studies regarding the use of GAs for AutoML in time series forecasting tasks. We will analyse the current state of research surrounding the use of GAs for AutoML in trend prediction in section 2.2. In section 2.3, we summarise the literature regarding the use of a GA for AutoML in for time series prediction. Section 2.4 details the GA-based AutoML and HPO software tools and platforms. Section 2.5 will analyse the important evaluation metrics that are used. Section 3 presents the main discussions and recommendations drawn. Finally, in section 4, we provide conclusions of the result of this paper.

2 USE OF GENETIC ALGORITHMS FOR AUTOMATED MACHINE LEARNING

The literature regarding the use of GAs for AutoML has been organized based on the application of the AutoML technique in the literature under review. Broadly speaking, these applications are the use of GAs for AutoML in time series prediction tasks and the use of GAs for AutoML for trend prediction tasks. The majority of the literature reviewed focuses on the use of GAs for HPO in the ML model development pipeline.

2.1 Genetic Algorithms for AutoML for Time Series Prediction Tasks

The majority of literature reviewed deals with the use of GAs for HPO of various ML algorithms for time series prediction tasks. Generally, each study focuses on the use of a GA for HPO of a single ML algorithm and compares this model to other standard ML algorithms. These standard ML algorithms usually have manually selected hyperparameters. These hyperparameters are either selected by using past studies as a guide, using common heuristic to select hyperparameters, or leaving them as their default values [2,3]. We will briefly analyse the literature surrounding the use of a GA for HPO of specific ML algorithms. These ML algorithms are support vector machine (SVM), artificial neural network (ANN), recurrent neural network (RNN), long short-term memory (LSTM), and gated recurrent unit (GRU).

2.1.1 Support Vector Machine

The use of GAs for automated hyperparameter optimisation in SVMs for prediction tasks in time series data, as well as support vector regression (SVR), has shown promising results when compared to traditional forecasting models like least-mean squared algorithm, ANNs, SVR models, bagging, random forest, SVMs, and AdaBoost [5,19].

2.1.2 Artificial Neural Networks

The use of ANNs for time series forecasting have been shown to be better than many traditional forecasting models like ARIMA as ANNs can be used to solve complex, non-linear problems [8]. The use of GA for HPO for ANNs was successful when compared to ANNs with hand selected hyperparameters. [2] Due to the relative simplicity of ANNs, a hybrid GA-ANN model that uses a GA to optimise the hyperparameters selection for the ANN can be a useful comparison metric for newer GA HPO hybrid models like the study by Almalaq et al. [1]. The complexity of real-world forecasting problems regarding time series data means that ANNs are not powerful enough to efficiently handle these problems. Kieu et al. [8] shows that use of GAs for HPO of ANNs is less effective for time series prediction that the use of GAs for HPO for either RNN or LSTM.

2.1.3 Recurrent Neural Networks

Kieu et al. [8] tested the use of GAs for HPO for ANN, RNN, and LSTM models. The use of GAs to optimise hyperparameters of RNNs showed better than the hybrid GA-ANN method, but was not better that the combination of GA for HPO and LSTM.

2.1.4 Long Short-Term Memory

Majority of the literature reviewed focuses on the use of GAs for HPO of LSTM models. An LSTM is a state-of-the-art RNN which makes it extremely useful for solving time series problems and is regarded as one of the most critical deep-learning algorithms due to its long-term memory characteristic [8]. GAs can be used for HPO of LSTM models. Common hyperparameters include window size (or number of time lags), number of hidden neurons per layer, number of hidden layers, and number of epochs [1,3,8]. Some studies fixed the number of hidden layers and applied HPO only to window size and number of hidden neurons per layer [1,3]. In most studies, a hybrid ML model that uses GAs for HPO for LSTM is proposed, GA-LSTM. In these studies, this hybrid approach outperformed various other ML models - including a standard LSTM and a GA-ANN hybrid model - in time series prediction tasks [1,3]. However, Kieu et al. [8] reported very similar results between the use of GA for HPO LSTM and RNN.

2.1.5 Gated Recurrent Unit

Gated recurrent unit (GRU) is designed to handle sequential data such as time series. The structure of GRU is slightly simpler that LSTM, hence, training speed of GRU is slightly faster than LSTM. Noh et al. [12] successfully combined GRU and GA for HPO as a hybrid model GA-GRU in order to forecast product demand. The GA-GRU model outperformed all other forecasting methods that it was compared against. These other forecasting methods included ARIMA, RNN, GRU, and a GA-LSTM hybrid model where the LSTM hyperparameters are optimized using a GA [12].

2.2 Genetic Algorithms for AutoML for Trend Prediction

Despite a tremendous amount of research into the use of GAs for AutoML, in general, and the large number of studies that focus on its use in time series forecasting application, there has been very little research into the use of GAs for AutoML, especially HPO, in trend prediction of time series data. Studies of trend prediction often make use of statistical or ML models where the hyperparameters are based on experimentation or selected using expert knowledge [9,11].

Table 1: A comparison of various studies that use GAs for AutoML/HPO for time series forecasting

Reference	**Application of GA to AutoML**	**ML Algo. Optimised**	**Hyperparameters Optimised**	**Application of ML Algorithm**	**Data Set**	**Results**	**Strengths**	**Limitations**
[8]	GA for HPO of NNs for time series forecasting	ANN RNN LSTM	Number of epochs. Number of hidden layers. Number of nodes per layer.	Time series forecasting	Time series of daily maximum temperatures at Cheongja Station, South Korea. (1976-2015).	Study focuses on the use of ML algorithms with the use of GA for HPO. Results show the effectiveness of NNs rather than GA for AutoML. GA and LSTM shown to be effective for prediction.	Application to time series prediction. Both one- and multi-step-ahead prediction studied. Analysis of GA for HPO in multiple ML algorithms for same problem.	Main focus of study was not on the use of GA for AutoML. Thus, results do not pertain to the use of GAs for AutoML. Does not focus on trend prediction.
[3]	Use of GA for HPO for LSTM models for time series forecasting.	LSTM	Number of time lags/window size. Number of hidden layers. Number of nodes per layer.	Time series forecasting	Time series of RTE Power Consumption in France. (Jan 2008-Dec 2016).	Study showed that LSTM with use of GA had lower forecast errors compared to best ML technique.	Focus was on ML application to forecasting, not GAs for AutoML.	Focus on short- and medium-term monthly forecasting horizons. Focus on only a single ML algorithm (LSTM). Does not focus on trend prediction.
[2]	Use of GA for HPO for NNs for time series forecasting.	ANN	Not clear, just refers to optimisation of NN structure/architecture.	Time series forecasting	Time series of economic and social indicators in Singapore. (1975-1994).	Study showed use of GAs for hyperparameter selection of NNs performed better than conventional NN.	Analysis of model performance with and without GA hyperparameter selection.	Paper was first published in 2000 and the idea of HPO is quite rudimentary. Comparison NN was designed using outdated heuristics that may be far from optimal. Does not focus on trend prediction.
[19]	Use of GA to HPO of SVR algorithms for time series forecasting.	SVR	C – Penalty factor γ ε	Time series forecasting	Monthly sales volume of trucks and small cars of a Taiwanese car-manufacturer, including other input variables such as GDP per capita and CPI. (2003-2009).	The study showed that the hybrid GA-SVR approach outperformed ANNs and traditional ML techniques by a more than 50% reduction in MAPE on average.	Illustrates the potential benefit of applying a GA for HPO. Comparison of various ML forecasting techniques.	Construction of comparison ML algorithms not detailed, and these may sub-optimal. These ML algorithms are relatively simple and newer, better techniques have been proposed. Does not focus on trend prediction.
[12]	Use of GA for HPO of GRUs for time series forecasting.	GRU	Window size/Number of time lags. Number of neurons in hidden state. Batch size. Number of epochs. Initial learning rate	Time series forecasting	Customer daily product demand of a top Brazilian retailer. (Jan 2014-Jul 2016).	This study proposes a hybrid GA-GRU model for forecasting product demand. This hybrid model outperformed all other ML models including a hybrid GA-LSTM model.	The study provides many comparison algorithms including a GA-LSTM hybrid.	Only studied one-step-ahead forecasting Use of k-fold cross-validation for time series data is not advised. Only deals with univariate data. Does not focus on trend prediction.
[5]	Use of GA for HPO and feature extraction of SVMs for time series forecasting.	SVM	C – Penalty factor. σ - Bandwidth of Gaussian RBF kernel.	Time series forecasting	World stock indices from Brazil, US, UK, Germany, Japan, China. Data on the currencies USD, EUR, CNY.	The study proposes a hybrid GA-SVM method, GAENSEMBLE, that will use GAs for both feature extraction from the data, and HPO. The proposed method outperformed various other methods including bagging, AdaBoost, random forest, SVM.	The study shows the applicability of GA for HPO to SVM algorithms.	Improvement in proposed method cannot be attributed to HPO solely. Use of 10-fold cross-validation. Extremely long run times compared to other methods. Does not focus on trend prediction.
[1]	Use of GA for HPO for LSTM models for time series forecasting.	LSTM	Number of hidden neurons in each layer, Window size/Number of time lags.	Time series forecasting	Individual household electricity consumption in KW with one minute resolution. (Dec 2006-Nov 2010) Data consumption for single commercial building in KW with five-minute resolution. (Jan 2012-Dec 2012)	This study proposes a hybrid GA-LSTM method and compares it to various other ML algorithms. The proposed method outperformed all other ML models studied including a hybrid GA-ANN	Study provides comparison of two GA hybrid ML methods, GA-LSTM and GA-ANN for time series prediction on the same data set. Reduces complexity of the HPO search space by not optimising the number of hidden layers by means of the GA.	Use of 10-fold cross-validation instead of walk-forward validation. Does not focus on trend prediction.

2.3 Analysis of Studies Regarding the Use of Genetic Algorithms for AutoML for Time Series

An analysis of the studies that were reviewed in the paper is provided above in Table 1. These studies' main focus was on the use of GAs for HPO of various ML algorithms for time prediction tasks. Each study was analysed based on a number of criteria. These criteria form the columns of the table above.

2.4 Genetic Algorithm based AutoML Software Tools and Platforms

Many software tools and platforms exist that either (attempt to) completely automate the entire ML pipeline or are used to perform a specific aspect of the AutoML pipeline such as HPO or feature extraction.

2.4.1 DEAP

DEAP – Distributed Evolutionary Algorithms in Python – is the most common software tool used to implement a GA for the task of HPO [1,8,12]. DEAP implements various evolutionary algorithms. It can be integrated with sklearn and parallelization mechanism like multiprocessing [17].

2.4.2 TPOT

TPOT is an AutoML tool than can be used to optimize ML pipelines using GAs. TPOT is a python tool that is built on top sklearn and is therefore easy to implement TPOT across various ML models [17].

2.4.3 Nevergrad

Nevergrad is an open-source Python library that can be used to tune all types of hyperparameters by choosing different optimisers. Nevergrad includes a wide array of optimisers included fast-GA and particle swarm optimisation [17].

2.5 Evaluation Metrics for AutoML

Throughout the literature, several key evaluation metrics were identified. These metrics are used to measure and compare the performance of different AutoML methods.

2.5.1 Root Mean Square Error

Root mean square error (RMSE) is the most common evaluation metric for comparing different kinds of AutoML methods [3,8,12]. Since errors are squared before taking the average, RMSE gives a higher weighting to large errors. Hence, RMSE is a useful metric when large errors are undesired [14].

2.5.2 Mean Square Error

Mean square error (MSE) is very similar to RMSE except that the RMSE is simply the square root of MSE [12].

2.5.3 Mean Absolute Error

The mean absolute error (MAE) is another common evaluation metric for AutoML [12]. Unlike RMSE, MSE does not give higher weighting to larger errors. If the size of the error is irrelevant, then using MAE is appropriate [14].

2.5.4 Mean Absolute Percentage Error

The mean absolute percentage error (MAPE) is the proposed evaluation metric regarding the use of GAs for AutoML in some studies [2,19]. MAPE is similar to MAE but the mean error is represented as a percentage as opposed to a positive real number.

In terms of GA for AutoML, the evaluation metric is implemented as the GA's fitness functions. This is used in determining the probability of selection for crossover as well as probability of being culled. These evaluation metrics are not common to AutoML problems and the application of these evaluation metrics to a problem must be carefully considered.

3 DISCUSSIONS AND RECOMMENDATIONS

Based on the literature that has been reviewed, a number of key areas of discussion and their following recommendations can be drawn.

The majority of papers that analyse the use GAs for AutoML in time series forecasting either focus on one of two aims:

1. Improving a single ML algorithm through the use of a GA-based HPO method, or
2. Analyse the performance of multiple ML algorithms that use GA-based HPO, which all use the same data, the same GA-based method for HPO, and evaluation metrics to solve the same problem.

Each of these two cases present a problem when analysing the results of the various studies. The first problem is that for papers with a focus on the first aim, comparing the results of the study with another was difficult as there were large differences in many aspects of the various studies such as data and evaluation metrics used. The second problem stems from studies that focus on the second aim listed above. These studies usually do not compare the results against the same ML algorithms where the hyperparameters are manually selected. In future studies, a combination of these two aims should be adopted in order to not only compare the effectiveness of GAs for HPO of ML algorithms, but also the effectiveness of GAs for HPO amongst various ML algorithms.

In many papers, the use of k-fold cross-validation was used which has been shown to not be appropriate for testing of time series data. Thus, in future studies, walk-forward validation should be used when dealing with time series data.

In some papers that compared HPO for an ML algorithm against the same ML algorithm with hand-tuned hyperparameters, the selection of those hyperparameters was done poorly by either using default values or selecting them from some previous literature. This could have resulted in the performance of HPO for ML algorithms being overstated.

4 CONCLUSIONS

In this paper, we reviewed literature regarding the use of GAs in AutoML for trend prediction in time series data. Due to the lack of literature around the use of GAs in AutoML for trend prediction, studies regarding applications to time series forecasting formed the bulk of the analysis. Specifically, the use of GAs for HPO in time series forecasting was analysed. The studies that were analysed show that GAs are an effective tool for HPO, a sub-class of AutoML, of various forecasting models. The literature suggests that a combination of GA for HPO with LSTM and GRU models can provide forecasting models that outperform various other forecasting methods that use hand-tuned hyperparameters. The effectiveness of a GA-LSTM or GA-GRU for trend prediction is yet to be established, however, it not be a leap-of-faith to postulate that these same methods will have a similar performance when applied to trend prediction.

REFERENCES

[1] Abdulaziz Almalaq and Jun Jason Zhang. 2019. Evolutionary Deep Learning-Based Energy Consumption Prediction for Buildings. *IEEE Access* 7, (2019), 1520–1531. DOI:https://doi.org/10.1109/ACCESS.2018.2887023

[2] Goh Bee-Hua. 2000. Evaluating the performance of combining neural networks and genetic algorithms to forecast construction demand: The case of the Singapore residential sector. *Construction Management and Economics* 18, 2 (2000), 209–217. DOI:https://doi.org/10.1080/014461900370834

[3] Salah Bouktif, Ali Fiaz, Ali Ouni, and Mohamed Adel Serhani. 2018. Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies* 11, 7 (2018). DOI:https://doi.org/10.3390/en11071636

[4] Dattatray P. Gandhmal and K. Kumar. 2019. Systematic analysis and review of stock market prediction techniques. *Computer Science Review 34*, 100190. DOI:https://doi.org/10.1016/j.cosrev.2019.08.001

[5] Rafael Thomazi Gonzalez, Carlos Alberto Padilha, and Dante Augusto Couto Barone. 2015. Ensemble System based on Genetic Algorithm for stock market forecasting. In *2015 IEEE Congress on Evolutionary Computation, CEC 2015 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 3102–3108. DOI:https://doi.org/10.1109/CEC.2015.7257276

[6] Baosu Guo, Jingwen Hu, Wenwen Wu, Qingjin Peng, and Fenghe Wu. 2019. The Tabu_genetic algorithm: A novel method for hyper-parameter optimization of learning algorithms. *Electronics (Switzerland)* 8, 5 (May 2019), 579. DOI:https://doi.org/10.3390/electronics8050579

[7] Xin He, Kaiyong Zhao, and Xiaowen Chu. 2021. AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems* 212, (2021), 106622. DOI:https://doi.org/10.1016/j.knosys.2020.106622

[8] Trang Thi Kieu Tran, Taesam Lee, Ju Young Shin, Jong Suk Kim, and Mohamad Kamruzzaman. 2020. Deep learning-based maximum temperature forecasting assisted with meta-learning for hyperparameter optimization. *Atmosphere* 11, 5 (2020), 1–21. DOI:https://doi.org/10.3390/ATMOS11050487

[9] Kouame Hermann Kouassi and Deshendran Moodley. 2020. *Automatic deep learning for trend prediction in time series data*. Retrieved from http://arxiv.org/abs/2009.08510

[10] Jason Liang, Elliot Meyerson, Babak Hodjat, Dan Fink, Karl Mutch, and Risto Miikkulainen. 2019. Evolutionary Neural AutoML for Deep Learning. 9, (2019). DOI:https://doi.org/10.1145/3321707.3321721

[11] Tao Lin, Tian Guo, and Karl Aberer. 2017. Hybrid neural networks for learning the trend in time series. In *IJCAI International Joint Conference on Artificial Intelligence*, International Joint Conferences on Artificial Intelligence, 2273–2279. DOI:https://doi.org/10.24963/ijcai.2017/316

[12] Jiseong Noh, Hyun Ji Park, Jong Soo Kim, and Seung June Hwang. 2020. Gated recurrent unit with genetic algorithm for product demand forecasting in supply chain management. *Mathematics* 8, 4 (2020). DOI:https://doi.org/10.3390/math8040565

[13] Laurits Tani, Diana Rand, Christian Veelken, and Mario Kadastik. 2021. Evolutionary algorithms for hyperparameter optimization in machine learning for application in high energy physics. *European Physical Journal C* 81, 2 (February 2021), 170. DOI:https://doi.org/10.1140/epjc/s10052-021-08950-y

[14] Rafael Thomazi Gonzalez. *Using Deep Learning and Evolutionary Algorithms for Time Series Forecasting*.

[15] Jonathan Waring, Charlotta Lindvall, and Renato Umeton. 2020. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artificial Intelligence in Medicine 104*, 101822. DOI:https://doi.org/10.1016/j.artmed.2020.101822

[16] Darrell Whitley. 1993. *A Genetic Algorithm Tutorial*.

[17] Li Yang and Abdallah Shami. 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing* 415, (2020), 295–316. DOI:https://doi.org/10.1016/j.neucom.2020.07.061

[18] Steven R. Young, Derek C. Rose, Thomas P. Karnowski, Seung Hwan Lim, and Robert M. Patton. 2015. Optimizing deep learning hyper-parameters through an evolutionary algorithm. *Proceedings of MLHPC 2015: Machine Learning in High-Performance Computing Environments - Held in conjunction with SC 2015: The International Conference for High Performance Computing, Networking, Storage and Analysis* (November 2015). DOI:https://doi.org/10.1145/2834892.2834896

[19] Fong-Ching Yuan. 2012. Parameters Optimization Using Genetic Algorithms in Support Vector Regression for Sales Volume Forecasting. *Applied Mathematics* 03, 10 (2012), 1480–1486. DOI:https://doi.org/10.4236/am.2012.330207