Automatic Machine Learning for Trend Prediction: A review of the literature

Adam Lewison
lwsada002@myuct.ac.za
University Of Cape Town
Cape Town, South Africa

ABSTRACT

Machine learning techniques and algorithms are used in almost every application domain such as image recognition, object detection, and financial applications. Building a high-quality machine learning model relies on human expertise. Automatic Machine Learning (AutoML) is the process of automating different stages of the Machine Learning Pipeline. This automation has been shown to produce superior Machine Learning Models. In this paper, we present a survey for the various methodology and past work making use of AutoML for time series data to solve the problems of forecasting and trend prediction. Special attention is given to a particular subset of the Machine Learning pipeline which involves algorithm selection and Hyper-Parameter Optimization. We look at a state-of-the-art methodology to optimise this step known as *Combined Algorithm And Hyper-parameter Selection* (CASH).

KEYWORDS

Automatic Machine Learning, Particle Swarm Optimisation, Time Series, Trend Prediction

1 INTRODUCTION

Due to the importance of time series forecasting in many areas, a lot of research has gone into the development of forecasting models and in improving prediction accuracies [1]. The use of machine learning systems have shown, in many cases, to outperform classical statistical methods. [24]. Machine learning has made a lot of progress in the last decade , but the performance depends on the selection of the right model architecture and training algorithm [31]. Hence, automating and optimising these choices has shown to provide superior results in time series problems [23]. Arguably in certain instances it is more valuable to forecast the trend of a time series than to forecast the absolute value [24]. However it is shown that research is limited for the use of AutoML for time series trend prediction [23]. The structure of this paper is set out to look at AutoML in general (Section 2), what problem it aims to solve and a development of the research of AutoML. According to [8] the primary function of automated machine learning is hyper-parameter optimization (HPO), so a more in-depth survey of the step of the Machine Learning Pipeline known as *Model Generation* (Figure 1), in particular the use of Combined Algorithm And Hyper-parameter Selection (Section 2.2) for *Model Generation.* We survey various Black Box Optimization techniques (Section 3) in order to solve the CASH problem with a more detailed look at the optimisation technique: Particle Swarm Optimisation (Section 4). We then explore the use of Machine Learning techniques (with an emphasis on AutoML) as applied to solving time series problems (Section 5). Results of previous studies of ML and AutoML when applied to time series problems are tabulated in Table 1. And finally we explore various open source software frameworks which enable the practitioner to apply AutoML methods to time series data (Section 6).

2 AUTOMATIC MACHINE LEARNING

Building a high-quality machine learning model relies on human expertise. The human expert (usually a data scientist) will approach a machine learning problem as outlined in Figure 1. The decisions of the data scientist in each of these steps affect the performance and quality of the developed model [9]. Furthermore, users of machine learning tools are often non-experts who require off-the-shelf solutions [9]. The automation of different steps of the machine learning pipeline reduces the human effort necessary for applying machine learning, improves the performance of machine learning algorithms as well as improves the reproducibility and fairness of scientific studies [8]. AutoML is beneficial for the ML expert due to the automation of tasks like hyperparameter optimization (HPO) and also for the Domain expert who is now able to build their own ML pipeline without a data scientist.

2.1 The Development of AutoML

As described by [32], Automatic Machine Learning (AutoML) has been evolving for a number of decades. As early as the 1990s, solutions were available for Hyper-Parameter Optimisation for selected classification algorithms using the brute-force [26] grid search technique. This was quickly improved using a greedy best-first approach adaptation of grid search [21]. By the early 2000s, the first approaches for automatic feature selection were introduced into AutoML literature [27]. In 2009 *Full model selection* [12] attempted to completely automate the full ML pipeline - that is: data pre-processing, feature selection and classification while optimizing the Hyper-Parameters of each step. A more detailed description of the ML pipeline is covered by [14]. Starting from 2011 Bayesian Optimisation for Hyper-Parameter tuning [5] as well as model selection [28]. The work of [28] developed *Auto-Weka* framework for Combined Algorithm And Hyper-parameter Selection (CASH). Automatic Feature engineering without domain knowledge was introduced into AutoML in 2015 [19]. In recent years there has been a primary research focus on Neural Architecture Search (NAS) which automates the stages of model generation and model evaluation, many papers have done an in-depth survey of NAS [10, 30]. There is also much research into automation of the rest of the ML pipeline [9, 14]. The rest of this paper will focus on the review of

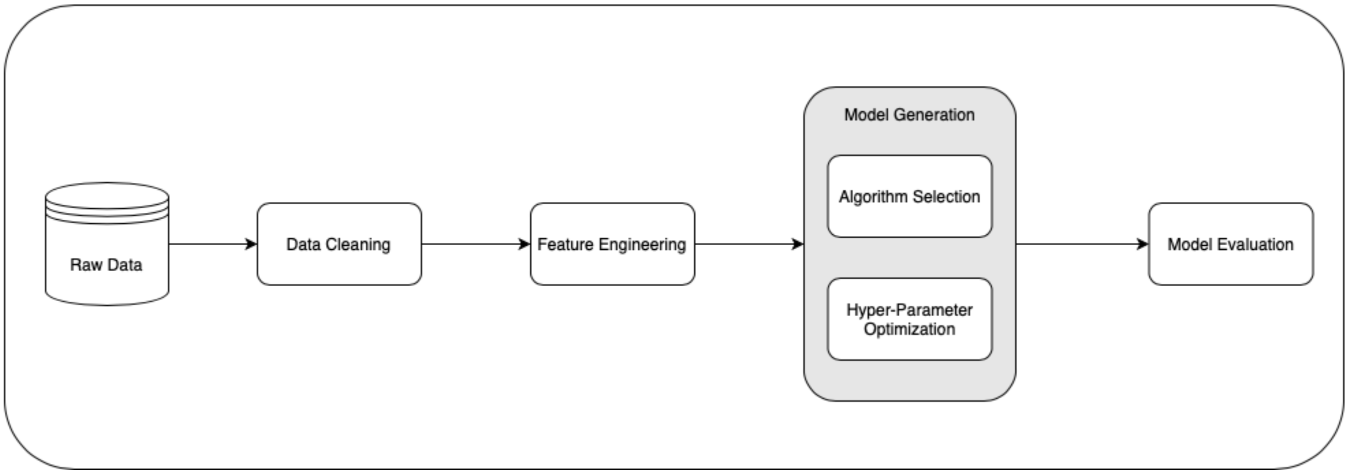

Figure 1: Machine Learning pipeline as used by most AutoML frameworks as described by [9, 14, 26, 32]

the research and development of the Combined Algorithm And Hyper-parameter Selection problem.

2.2 Combined Algorithm And Hyper-parameter Selection

Combined Algorithm And Hyper-parameter Selection was first introduced in [28]. In this method, the selection of an algorithm along with the optimisation of its hyper-parameters are executed simultaneously, instead of first selecting an algorithm and then optimising its hyper-parameters. Essentialy, CASH automates *Model Generation* as depicted in Figure 1. This problem can be formulated as a black box optimization problem leading to a minimization problem [32]. This converts the problem of CASH into the problem of solving the Black Box Optimization Problem.

3 SOLVING THE BLACK BOX OPTIMIZATION PROBLEM

The following various optimization problems strategies can be used can be used to solve the Black Box Optimization Problem:

- **Grid Search**
- **Random Search** [3]
- **Bayesian Optimization**
- **Sequential Model-based Optimization (SMBO)** [5, 15]
- **Evolutionary Algorithms**

Grid search is considered a brute force solution in which all combinations of of the input space are evaluated. While Bayesian Optimization is one of the state-of-the-art black- box optimization techniques which is tailored for expensive objective functions [9]. In recent studies [23] a hybrid solution Bayesian Optimisation and Hyper-Band (BOHB) was selected as the optimisation technique of choice due to its stable and competitive results when limited compute resources are available. Of all the evolutionary algorithms for solving the black box optimization problem, this review only takes a closer look at Particle Swarm Optimisation (PSO) in Section 4.

4 PARTICLE SWARM OPTIMISATION FOR AUTOML

Originally developed to simulate simple social behaviour of individuals in a swarm, particle swarms can also be used as an optimizer [20].

Particle swarm optimization (PSO) is a population based search algorithm inspired by the social behaviour of birds in a flock. PSO contains a swarm of particles, where each particle is a potential solution to an optimization problem. The swarm traverses a search space searching for an optimum solution. As the algorithm executes, each particle is attracted to both the best position it has found so far, as well as the best position found around it [1].

Particle Swarm Optimisation was used among the first AutoML methods dealing with the *full model selection problem* [11] i.e. finding the best combination of data preprocessing, feature engineering and model selection, together with the optimization of all of the associated hyper-parameters (See Figure 1). In [32] PSO was implemented in the Optunity framework, used as the optimisation technique to solve the CASH problem. As depicted in Table 2, this combination of Optunity using PSO scored the highest average accuracy among any other AutoML framework and optimisation technique.

5 AUTOML FOR TIME SERIES

An essential area of the study of time series is forecasting which involves estimating future values of a time based sequence of historical data (time series) [1].

Before machine learning techniques were adopted, trend prediction methods included Hidden Markov Models and multi-step ahead prediction. However, underlying the raw data of a time series is a complex data generating system. This system is the result of of uncertain behaviours and dynamic causalities of the underlying multivariate real-world processes [17]. This results in time series exhibiting characteristics of uncertainties, multivariate correlation, non-linearity and chaotic fluctuation [6]. It was shown that machine learning methods are more suitable to adequately model such

Author	Study		Model		Performrmance	
Kouassi and Moodley [23]	**Objective**	TS TP	**Type**	CASH	**Measure**	RMSE
			Model	*Automatic*		
	Data	Voltage TS	**OT**	BOHB	**Value**	35,72
	Objective	TS TP	**Type**	CASH	**Measure**	RMSE
			Model	*Automatic*		
	Data	Methane TS	**OT**	BOHB	**Value**	27,05
	Objective	TS TP	**Type**	CASH	**Measure**	RMSE
			Model	*Automatic*		
	Data	NYSE TS	**OT**	BOHB	**Value**	43,58
	Objective	TS TP	**Type**	CASH	**Measure**	RMSE
			Model	Automatic		
	Data	JSE TS	**OT**	BOHB	**Value**	16,21
Wen et al. [29]	**Objective**	TS TP	**Type**	CNN	**Measure**	f1 score
			Model	LSTM		
	Data	S&P 500 TS	**OT**	*Manual*	**Value**	63,05%
	Objective	TS TP	**Type**	CNN	**Measure**	f1 score
			Model	LSTM		
	Data	IBM Stock TS	**OT**	*Manual*	**Value**	53,60%
Almalaq and Zhang [2]	**Objective**	TS Forecasting	**Type**	HPO	**Measure**	RMSE (kW)
			Model	LSTM		
	Data	Residential Energy TS	**OT**	GA	**Value**	0.1943
	Objective	TS Forecasting	**Type**	HPO	**Measure**	RMSE (kW)
			Model	LSTM		
	Data	Commercial Energy TS	**OT**	GA	**Value**	0.427
Lipton et al. [25]	**Objective**	Medical Diagnosis	**Type**	RNN	**Measure**	Micro f1 score
			Model	LSTM		
	Data	Medical TS	**OT**	*Manual*	**Value**	29,38%
Lin et al. [24]	**Objective**	TS TP	**Type**	Neural Network	**Measure**	RMSE
			Model	LSTM		
	Data	Stock Price	**OT**	*Manual*	**Value**	9,735
	Objective	TS TP	**Type**	Neural Network	**Measure**	RMSE
			Model	TreNet		
	Data	Stock Price TS	**OT**	*Manual*	**Value**	7,545

Table 1: A summary of various papers using Automated Machine Learning for Trend Prediction or Forecasting using Time Series Data. For neatness, the following new abbreviations were introduced: TS (Time Series), TP (Trend Prediction)

sytems [6]. Artifical Neural Networks (ANNs) and State Vector Machines (SVMs) have been widely applied for time series prediction in finance, for example in stock price prediction [16]. In more recent work [24] demonstrated superior results in time series trend prediction using a novel hybrid neural network for learning the trend (See Table 1). As is the case with all machine learning models, the performance depends on the selection of the right model architecture and training algorithm [31]. Hence, automating and optimising these choices has shown to provide superior results in time series problems [23].

5.1 Trend Prediction

In time series trend prediction we are interested in understanding and forecasting the evolving trend of a time series i.e. the upward or downward pattern. According to [24] in certain cases it is more valuable to forecast the trend of a time series than to forecast the absolute value. For example predicting stock price trends is preferred over the prediction of the actual future price of the stock [4].

5.2 Previous Work

As explained in [23], there is a wide variety of the use of Machine Learning methods for time series prediction [1, 24, 25, 29]. Research is also available on the use of AutoML for other time series problems [2]. However, there is limited research on the use of AutoML for trend prediction.

Table 1 provides a summary of various studies which implemented AutoML methodologies in order to solve time series problems.

G	Framework	CASH OT	Average accuracy	Group Average
1	Scikit-learn	Grid Search	0,6964	0,829575
	Scikit-learn	Random Search	0,8746	
	HpBandSter	BOHB	0,8725	
	Optunity	PSO	0,8748	
2	Auto-Sklearn	Random Search	0,80853	0,818935
		SMAC	0,82606	
	TPOT	Genetic Prog.	**0,8304**	
	HPSKLEARN	Hyperopt	0,81075	

Table 2: Overview of results from [32]. G is abbreviated for Group. Group 1 is a select set of frameworks to run CASH optimisation using the corresponding Optimisation Technique (OT). Group 2 is a set of AutoML frameworks capable of building complete ML pipelines.

6 SOFTWARE TOOLS

Several tools have been implemented on top of widely used centralized machine learning packages. A popular selection has been tabulated in Table 3

7 CONCLUSIONS

In this paper we developed the idea of AutoML for solving time series problems. An survey of previous research shows that there appears to be a gap in the literature for Automated and Optimised Machine Learning algorithms for Time Series Trend Prediction. This is a promising area of research as we have shown Machine Learning methodology is superior to classical statistical forecasting methodology in the realm of time series, both with forecasting and Trend Prediction (Table 1). While at the same time it was shown that that solving the CASH problem with Particle Swarm Optimisation lead to machine learning models with competitive accuracy (Table 2). There is a wide variety of open-source software frameworks that are well suited to perform experimentation for the use of AutoML on time series data (Table 3). The author recommends using Evolutionary Algorithms (e.g: Particle Swarm Optimisation) as the solver of the CASH problem in order to generate machine learning models that perform trend prediction on time series data.

ACKNOWLEDGMENTS

The author would like to thank Dr Deshendran Moodley for his insight and guidance.

REFERENCES

[1] Salihu Aish Abdulkarim. 2018. *Time Series Forecasting using Dynamic Particle Swarm Optimizer Trained Neural Networks.* Ph.D. Dissertation. http://hdl.handle.net/2263/70388

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

[3] R. L. Anderson. 1953. Recent Advances in Finding Best Operating Conditions. *J. Amer. Statist. Assoc.* 48 (12 1953). Issue 264. https://doi.org/10.2307/2281072

[4] George Atsalakis and Kimon Valavanis. 2009. Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Systems with Applications* 36 (09 2009), 10696–10707. https://doi.org/10.1016/j.eswa.2009.02.043

[5] James Bergstra, R Bardenet, Yoshua Bengio, Balázs Kégl, and Rémi Bardenet. 2011. Algorithms for Hyper-Parameter Optimization. https://hal.inria.fr/hal-00642998

[6] Changqing Cheng, Akkarapol Sa-Ngasoongsong, Omer Beyca, Trung Le, Hui Yang, Zhenyu Kong, and Satish Bukkapatnam. 2015. Time Series Forecasting for Nonlinear and Nonstationary Processes: A Review and Comparative Study. *IIE Transactions* (01 2015). https://doi.org/10.1080/0740817X.2014.999180

[7] Marc Claesen, Jaak Simm, Dusan Popovic, Yves Moreau, and Bart De Moor. 2014. Easy Hyperparameter Search Using Optunity. arXiv:1412.1114 [cs.LG]

[8] Frank. editor. Hutter, Lars. editor. Kotthoff, and Joaquin. editor. Vanschoren. 2019. *Automated Machine Learning.* Springer International Publishing. 101822 pages. https://doi.org/10.1007/978-3-030-05318-5

[9] Radwa Elshawi, Mohamed Maher, and Sherif Sakr. 2019. Automated Machine Learning: State-of-The-Art and Open Challenges. (6 2019). http://arxiv.org/abs/1906.02287

[10] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. 2019. Neural Architecture Search: A Survey. arXiv:1808.05377 [stat.ML]

[11] Hugo Jair Escalante. 2020. Automated Machine Learning – a brief review at the end of the early years. arXiv:2008.08516 [cs.LG]

[12] Hugo Jair Escalante, Manuel Montes, and Luis Villaseñor. 2009. Particle Swarm Model Selection for Authorship Verification. https://doi.org/10.1007/978-3-642-10268-4_66

[13] Matthias Feurer, Aaron Klein, Katharina Eggensperger, Jost Springenberg, Manuel Blum, and Frank Hutter. 2015. Efficient and Robust Automated Machine Learning. In *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (Eds.). Curran Associates, Inc., 2962–2970. http://papers.nips.cc/paper/5872-efficient-and-robust-automated-machine-learning.pdf

[14] Xin He, Kaiyong Zhao, and Xiaowen Chu. 2021. AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems* 212 (1 2021). https://doi.org/10.1016/j.knosys.2020.106622

[15] Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. 2011. Sequential Model-Based Optimization for General Algorithm Configuration. https://doi.org/10.1007/978-3-642-25566-3_40

[16] Huseyin Ince. 2007. Kernel principal component analysis and support vector machines for stock price prediction. *Iie Transactions* 39 (03 2007). https://doi.org/10.1080/07408170600897486

[17] Peng Jiang, Cheng Chen, and Xiao Liu. 2016. Time series prediction for evolutions of complex systems: A deep learning approach. *2016 IEEE International Conference on Control and Robotics Engineering (ICCRE).* https://doi.org/10.1109/ICCRE.2016.7476150

[18] Haifeng Jin, Qingquan Song, and Xia Hu. 2019. Auto-keras: An efficient neural architecture search system. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* 1946–1956.

[19] James Max Kanter and Kalyan Veeramachaneni. 2015. Deep feature synthesis: Towards automating data science endeavors. *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA).* https://doi.org/10.1109/DSAA.2015.7344858

[20] J. Kennedy and R. Eberhart. [n.d.]. Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks.* https://doi.org/10.1109/ICNN.1995.488968

[21] Ron Kohavi and George H. John. 1995. Automatic Parameter Selection by Minimizing Estimated Error. https://doi.org/10.1016/B978-1-55860-377-6.50045-1

[22] Brent Komer, J. Bergstra, and C. Eliasmith. 2014. Hyperopt-Sklearn: Automatic Hyperparameter Configuration for Scikit-Learn.

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

[24] Tao Lin, Tian Guo, and Karl Aberer. 2017. Hybrid Neural Networks for Learning the Trend in Time Series. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17.* 2273–2279. https://doi.org/10.24963/ijcai.2017/316

[25] Zachary C. Lipton, David C. Kale, Charles Elkan, and Randall Wetzel. 2017. Learning to Diagnose with LSTM Recurrent Neural Networks. arXiv:1511.03677 [cs.LG]

[26] Randal S. Olson, Nathan Bartley, Ryan J. Urbanowicz, and Jason H. Moore. 2016. Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science. arXiv:1603.06212 [cs.NE]

[27] B. Samanta. 2004. Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. *Mechanical Systems and Signal Processing* 18 (5 2004). Issue 3. https://doi.org/10.1016/S0888-3270(03)00020-7

[28] Chris Thornton, Frank Hutter, Holger H. Hoos, and Kevin Leyton-Brown. 2013. Auto-WEKA: Combined Selection and Hyperparameter Optimization of Classification Algorithms. *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.* https://doi.org/10.1145/2487575.2487629

[29] Min Wen, Ping Li, Lingfei Zhang, and Yan Chen. 2019. Stock Market Trend Prediction Using High-Order Information of Time Series. *IEEE Access* 7 (2019). https://doi.org/10.1109/ACCESS.2019.2901842

[30] Martin Wistuba, Ambrish Rawat, and Tejaswini Pedapati. 2019. A Survey on Neural Architecture Search. arXiv:1905.01392 [cs.LG]

Author	Software	OT	Language	ML Library
[13]	Auto-sklearn	Bayesian Optimisation	Python	scikit-learn
[28]	Auto-Weka	SMAC	Java	Weka
[26]	TPOT	Genetic Algorithm	Python	scikit-learn
[18]	Autokeras	Bayesian Optimisation	Python	Keras, TensorFlow
[7]	Optunity	Variety	Python, R, MATLAB	Generic
[22]	Hyperopt-sklearn	Bayesian Optimisation	Python	scikit-learn

Table 3: Summary of popular open-source software libraries which are used to solve the Combined Algorithm And Hyperparameter Selection (CASH) problem. Where OT stands for Optimisation Technique.

[31] Peter Zhang, Eddy Patuwo, and Michael Hu. 1998. Forecasting With Artificial Neural Networks: The State of the Art. *International Journal of Forecasting* 14 (03 1998), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

[32] Marc-André Zöller and Marco F. Huber. 2021. Benchmark and Survey of Automated Machine Learning Frameworks. arXiv:1904.12054 [cs.LG]