Analysis on Trends of Automated Machine Learning

Yo Seob Lee¹

¹School of ICT Convergence, Pyeongtaek University, Pyeongtaek-si, Gyeonggi-do, Republic of Korea

Abstract- Recently, machine learning has been applied to many fields. Many steps are required to apply machine learning to real problems. There are many automation solutions to simplify these parts. With the Automated Machine Learning solution, you can create simpler solutions, create solutions quickly, and provide a high-performance model. In this paper, we will analyze the recent trend of Automated Machine Learning. Keywords –Automated Machine Learning, Auto ML

I. INTRODUCTION

Demand is increasing rapidly in recent years than supply to machine learning professionals. To alleviate this gap, user-friendly machine learning software is available for non-experts. The first step in simplifying machine learning is to develop a simple and unified interface to various machine learning algorithms. Although non-experts made machine learning easy to experiment, many expertise was needed to create high-performance machine learning models, and deep learning models are difficult for non-experts to adjust. Solutions to automate machine learning have emerged to address these problems. In this paper, we will analyze the recent trend of Automated Machine Learning.

The rest of the paper is organized as follows. Proposed embedding and extraction algorithms are explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

II. AUTOMATED MACHINE LEARNING

Automated Machine Learning(Auto ML) is a process that automates end-to-end processes that apply machine learning to real problems. Typical machine learning applications apply data preprocessing, feature engineering, feature extraction, and feature selection methods for data sets to be used in machine learning. After this preprocessing step, algorithm selection and hyper parameter optimization are performed to maximize the predictive performance of the final machine learning model.

Figure 1 shows the general machine learning process for map learning. Through the models learned from the data, the results required from the new data can be predicted. By automating end-to-end processes that apply machine learning, you can create simpler solutions, quickly create solutions, and deliver models that are better than manually designed models.

Table -1	Auto	ML Solutions	

Solution	Open Source
Autosklearn	Yes
AutoWeka	Yes
Google Cloud Auto ML	No
H2O	Yes
ТРОТ	Yes

Table 1 shows Auto ML solutions. Most solutions offered as commercial services do not support open source.



A Machine Learning Process

Table 2 shows the automation objectives of Auto ML. Automation goals include automated data preparation and processing (from raw data and other formats), automated feature engineering, automated model selection, hyper parameter optimization (from learning algorithms and characterizations), automated pipeline selection and complexity constraints), automated selection of evaluation criteria / verification procedures, automated problem checking, automated analysis of the results obtained, and user interface and visualization for automated machine learning.

Table -2 Targets	of Automation	of Auto ML
------------------	---------------	------------

Target	Details		
Automated data preparation and	- Automated column type detection		
ingestion	- Automated column intent detection		
	- Automated task detection		
Automated feature engineering	- Feature selection		
	- Feature extraction		
	- Meta learning and transfer learning		
	- Detection and handling of skewed data and/or missing		
	values		
Automated model selection			
Hyper parameter optimization of the learning algorithm and featurization			
Automated pipeline selection under time, memory, and complexity constraints			
Automated selection of evaluation metrics / validation procedures			
Automated problem checking	- Leakage detection		
	- Misconfiguration detection		
Automated analysis of results obtained			
User interfaces and visualizations for automated machine learning			

Automated data preparation and processing includes automated column type detection (from Boolean, discrete numbers, continuous numbers or text), automated column tent detection (target / label, layered fields, numeric functions, categorized text functions), Automated job detection (in binary classification, regression, clustering or ranking).

Automated feature engineering consists of feature selection, feature extraction, meta learning and transfer learning, detection and processing of distorted data and / or missing values.

Automated problem checking consists of leakage detection and incorrect configuration detection.

III. AUTOMATED MACHINE LEARNING SOLUTIONS

3.1. AutoWEKA

AutoWEKA is a method to select machine learning algorithm and hyper parameter at the same time. It is combined with the WEKA package to automatically generate models suitable for various data sets.

3.2. Auto-sklearn

Auto-sklearn is an extension of AutoWEKA that uses the Python library scikit-learn. This tool replaces the regular scikit-learn classifier and regression analyzer.

3.3. TPOT

TPOT optimizes the machine learning pipeline using genetic programming. TPOT automates the most tedious part of machine learning by intelligently exploring thousands of possible pipelines to find the best fit for your data. Once the TPOT has completed the search, you can create a pipeline by providing the Python code for the best pipeline found.



An example TPOT pipeline

3.4. H2O AutoML

H2O AutoML provides automated model selection and ensembles for H2O machine learning and data analysis platforms. H2O AutoML has designed an easy-to-use interface that automates the process of educating various candidate models for non-technical users to easily use machine learning software. It can also be a useful tool for advanced users by providing a simple wrapper that performs many modeling tasks that typically require a large number of lines of code. H2O's AutoML can be used to automate machine learning workflows, including automatic learning and tuning of multiple models within a user-specified time limit.

3.5. Google Cloud AutoML

Google's Cloud AutoML is a solution that allows developers with limited machine learning expertise to leverage Google's cutting-edge transfer learning and neural-structure search technology to deliver high-quality models that meet business needs. As shown in Table 3, there are products that provide vision, natural language processing, and translation.

Product	Feature	
AutoML Vision	- Easily detect broad sets of objects in your	
	images	
	- Extract text	
	- Content moderation	
AutoML Natural	- Extract actionable insights	
Language	- Multimedia multilingual support	
	- Content classification relationship graphs	
AutoML Translation	- Translate many languages	
	- Language detection	
	- Simple integration	

Table -3 Google Cloud AutoML Products

Table 3 shows Auto ML solutions. Most solutions offered as commercial services do not support open source.

IV.CONCLUSION

In this paper, we analyze the current technology trend of AutoML. AutoML is a technology that allows companies without artificial intelligence development capabilities to create and apply artificial intelligence to services. As the market for AI services grows rapidly and the need for AutoML solutions grows, I think the analysis can help us identify future AutoML trends.

V. REFERENCES

- [1] B. Corona, M. Nakano, H. Pérez, "Adaptive Watermarking Algorithm for Binary Image Watermarks", Lecture Notes in Computer Science, Springer, pp. 207-215, 2004.
- [2] A A. Reddy and B. N. Chatterji, "A new wavelet based logo-watermarking scheme," Pattern Recognition Letters, vol. 26, pp. 1019-1027, 2005.
- [3] P. S. Huang, C. S. Chiang, C. P. Chang, and T. M. Tu, "Robust spatial watermarking technique for colour images via direct saturation adjustment," Vision, Image and Signal Processing, IEE Proceedings -, vol. 152, pp. 561-574, 2005.
- [4] F. Gonzalez and J. Hernandez, " A tutorial on Digital Watermarking ", In IEEE annual Carnahan conference on security technology, Spain, 1999.
- [5] D. Kunder, "Multi-resolution Digital Watermarking Algorithms and Implications for Multimedia Signals", Ph.D. thesis, university of Toronto, Canada, 2001.
- [6] J. Eggers, J. Su and B. Girod," Robustness of a Blind Image Watermarking Scheme", Proc. IEEE Int. Conf. on Image Proc., Vancouver, 2000.
- [7] Barni M., Bartolini F., Piva A., Multichannel watermarking of color images, IEEE Transaction on Circuits and Systems of Video Technology 12(3) (2002) 142-156.
- [8] Kundur D., Hatzinakos D., Towards robust logo watermarking using multiresolution image fusion, IEEE Transcations on Multimedia 6 (2004) 185-197.
- [9] C.S. Lu, H.Y.M Liao, "Multipurpose watermarking for image authentication and protection," IEEE Transaction on Image Processing, vol. 10, pp. 1579-1592, Oct. 2001.
- [10] L. Ghouti, A. Bouridane, M.K. Ibrahim, and S. Boussakta, "Digital image watermarking using balanced multiwavelets", IEEE Trans. Signal Process., 2006, Vol. 54, No. 4, pp. 1519-1536.
- [11] P. Tay and J. Havlicek, "Image Watermarking Using Wavelets", in Proceedings of the 2002 IEEE, pp. II.258 II.261, 2002.
- [12] P. Kumswat, Ki. Attakitmongcol and A. Striaew, "A New Approach for Optimization in Image Watermarking by Using Genetic Algorithms", IEEE Transactions on Signal Processing, Vol. 53, No. 12, pp. 4707-4719, December, 2005.
- [13] H. Daren, L. Jifuen, H. Jiwu, and L. Hongmei, "A DWT-Based Image Watermarking Algorithm", in Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 429-432, 2001.
- [14] C. Hsu and J. Wu, "Multi-resolution Watermarking for Digital Images", IEEE Transactions on Circuits and Systems- II, Vol. 45, No. 8, pp. 1097-1101, August 1998.
- [15] R. Mehul, "Discrete Wavelet Transform Based Multiple Watermarking Scheme", in Proceedings of the 2003 IEEE TENCON, pp. 935-938, 2003.
- [16] Automated machine learning, https://en.wikipedia.org/wiki/Automated_machine_learning.
- [17] Julien Jouganous, et al, A brief overview of Automatic Machine Learning solutions (AutoML), https://hackernoon.com/a-brief-overview-of-automatic-machine-learning-solutions-automl-2826c7807a2a.
- [18] AutoML, https://www.ml4aad.org/automl/.
- [19] AutoWEKA, http://www.cs.ubc.ca/labs/beta/Projects/autoweka/.
- [20] Auto-Sklearn, http://automl.github.io/auto-sklearn/stable/.
- [21] TPOT, http://epistasislab.github.io/tpot/
- [22] H2O AutoML, http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html.
- [23] Cloud AutoML, https://cloud.google.com/automl/.