Meta-information for Data Interference Reduction in Classification and Forecasting

David Afolabi, Sheng-Uei Guan, and Ka Lok Man,

***Abstract*—To achieve interference-less machine learning, therefore avoiding the negative impact of excessive noise/outlier on training and testing accuracy, we establish four fundamental hypotheses for application in classification and forecasting tasks. A spatial transformation, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN), and a temporal transformation, such as Empirical Mode Decomposition (EMD), have shown the potential to form meaningful representations of the data in classification and forecasting tasks respectively. Using these hypotheses, the dataset is preprocessed to generate meta-information. This meta-information is utilized to guide the model building stage and noise reduction is evident. Several learning algorithms—for instance, the Constructive Backpropagation (CBP) for classification and the long short-term memory (LSTM) neural network for forecasting—have been augmented and tested on real-world benchmark datasets and our results reported in several research proceedings reveal significant performance enhancement when conditions for these hypotheses are satisfied. This paper presents an overview of the techniques and potential areas of application.**

***Index Terms*—meta-learning, interference, machine learning, DBSCAN, EMD, classification, forecasting.**

I. Introduction

THE popularity of machine learning has soared in the last decade due to recent advancements in learning algorithms and enabling hardware infrastructure. Even though the performance of these algorithms are currently suitable for real-world application in some areas such as computer vision, natural language processing, and forecasting, these systems' reliability depends on the quality of the data that is been modelled. Therefore, the impact of noise and outliers within the data can be detrimental to machine learning algorithm's accuracy and training time [1]. To mitigate the effect of such interference present in the dataset, we propose a meta-learning technique carried out at the preprocessing stage which is independent on the machine learning algorithm selected to model the data. Meta-learning is a branch of machine learning that focuses on the automatic and flexible learning of informative concepts or knowledge mined from given data (or related data) in an efficient manner to improve performance, whereby such system includes a process to monitor the learning progress. The process involves extracting inherent patterns and trends within the data to create meta-information on how the data is structured and this information is used to guide the learning algorithm as it builds an appropriate model [2]. This meta-learning technique for interference reduction is extended for application in classification [3], [4] and forecasting tasks [5], [6]. The following section details the fundamental hypotheses that inspire this potential solution to noise/outlier interference, and then the specific meta-learning processes with regards to classification and forecasting are expounded. The discussion section highlights the observations in this research and the concluding section presents potential real-world applications.

Manuscript received January 30, 2018; revised February 11, 2018. This research is supported by China Jiangsu Provincial Science and Technology programme under Grant No. BK20131182.

D. Afolabi, S-U. Guan, and K. L. Man are with the Department of Computer Science and Software Engineering, Xian Jiaotong-Liverpool University, Suzhou 215123, China. e-mail: {david.afolabi09, steven.guan, ka.man}@xjtlu.edu.cn.

II. Fundamental Hypotheses

Our proposed method is established on four main hypotheses:

- Cluster assumption for meta-learning.
- Recurrent trend patterns for meta-learning.
- Hierarchical learning of local and global clusters/trends.
- Meta-learning for noise identification and elimination.

A. *Cluster assumption for meta-learning*

The cluster assumption [7] is based on the hypothesis that samples of the dataset with higher similarity should be clustered with other neighbouring samples having similar labels. If a group of unidentified patterns with high similarity can be clustered in a region, the whole cluster can be assigned with the identity of the closest labelled data. This form of semi-supervised learning techniques has a number of advantages such as:

1) Automatic labelling of a huge set of unlabelled data which is usually the easiest type of data to collect. This technique is also efficacious to unsupervised learning problems.
2) Identification of new concepts from unanticipated clusters (i.e. knowledge discovery which can contribute to accurate learning) and anomalies such as data outliers (which can cause learning interference and increase error rate).

Most interference would occur within a single cluster containing multiple class patterns thus leading to a longer time needed for training and also some amount of unlearning of useful discriminatory factors that could have been learnt at an earlier epoch [8]. Therefore, by decomposing the entire dataset into arbitrary-shaped clusters, based on the structure of the classification dataset, we tag each training sample via its meta-attribute in order to guide the machine learning algorithm model building phase. The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [9] is utilized for clustering because of its ability to form irregularly shaped clusters which is representative of how real-world dataset clusters are shaped.

ISBN: 978-988-14048-8-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

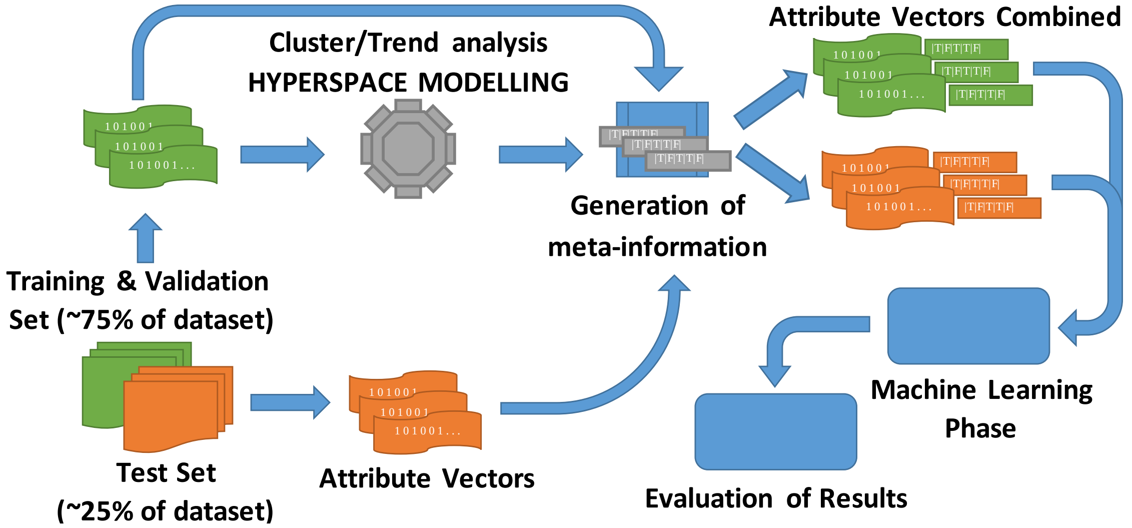

Fig. 1. Meta-information generation work-flow.

B. Recurrent trend patterns for meta-learning

Identification of characteristic trends within complex time series data is possible by means of common transformation algorithms such as Empirical Mode Decomposition (EMD) [10] and Fourier transform [11]. EMD results in a decomposition of an original signal into several building-block trends from high to low frequencies, called the intrinsic mode functions, which can be summed up to reconstruct the original signal. These derived structural primitive trends contain repeating patterns that can be information bearing, or on the other hand noise-inducing, but enables easier detection of components necessary for improved prediction accuracy. This approach offers the following advantages:

1) It enables the generation of meta-information so that noise/interference inducing components can be eliminated.
2) Meta-information is introduced into the learning process as an exogenous input to guide the learning process.

C. Hierarchical learning of local and global clusters/trends

As real-world information tends to be structured in a hierarchy of concepts, it is intuitive to learn on small/simpler clusters before tackling a complex encompassing cluster; or as in time series, learning short-term patterns before long-term trends. A related research has proposed an arbiter meta-learning technique in which random subsets of the training data are used to train several learning algorithm models before integrating the solution to obtain a better or at least equal predictive performance [12]. The technique in [12] can be highly susceptible to randomization; therefore, the previous two hypotheses can be applied to mitigate this effect. This machine learning follows a hierarchically structured manner and presents the following advantages:

1) Prevent interference between various class clusters that lie relatively close to each other.
2) Learning a hierarchy of concepts especially in data with taxonomy tree-like structure.
3) Provides a local and global view on the data being modelled.
4) Allows learning of short-term and long-term trends separately in order to avoid erosion of previously learned patterns.

D. Meta-learning for noise identification and elimination

Meta-information is derived and introduced into the learning process based on the inherent structure/distribution of pattern clusters or component signal trends within the data to tackle the problem of interference and noise within input attributes. Rather than simply recombining existing input attributes, these meta-attributes are derived through supervised or unsupervised techniques. In this research, a density-based clustering and signal decomposition technique (for classification and time series tasks respectively) are selected to derive the meta-attributes which inform the machine learner about inherent characteristics of each training sample and how it relates to other samples.

III. Meta-information

As mentioned, the process of meta-information generation involves decomposing the original data into structural primitives to provide additional insight into the relationship between classes, attributes and trends of the training set. This principle is applied using different techniques for the classification and forecasting tasks.

A. In classification

Real-world data categories or clusters are not necessarily formed as definable geometric shapes in a spatial dimension. From [13], it is stated that complexity reduction can be achieved by recursive segmentation and learning of formed clusters. Thus using hyperplanes or hypersphere [14] may not be sufficient to define the irregularly shaped cluster boundary. Conversely, a viable solution derived from hyperplanes or hypersphere can get very complex in order to accurately characterize an arbitrarily shaped group of related

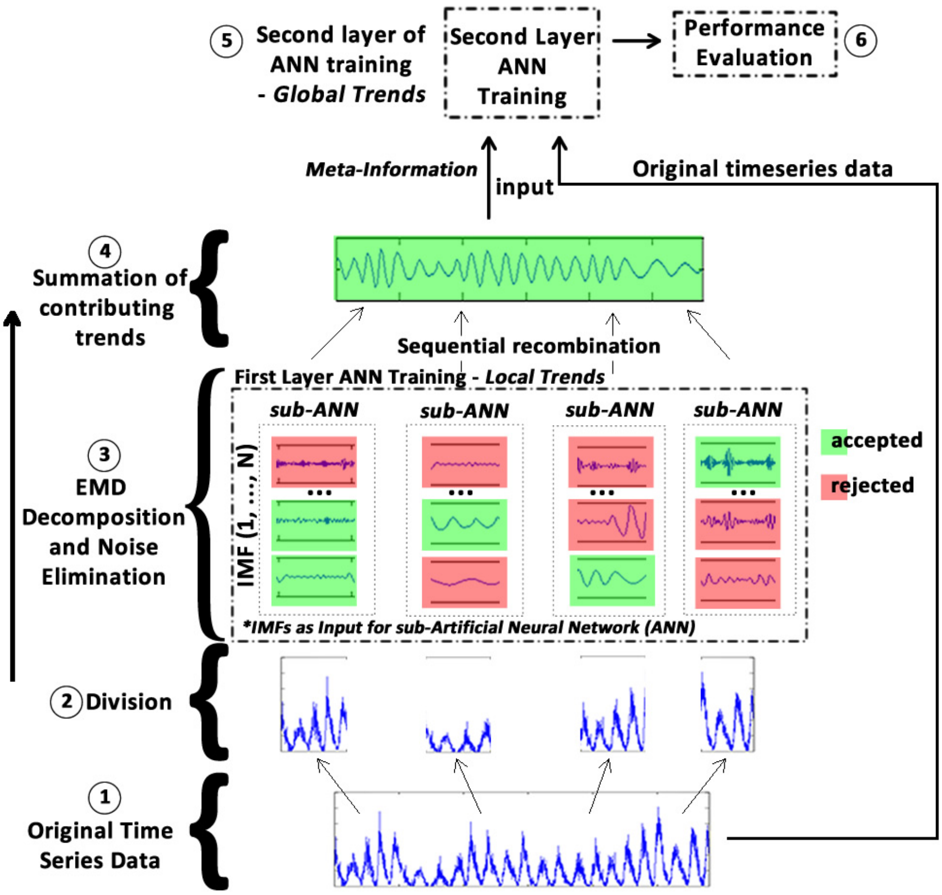

Fig. 2. Hierarchical meta-learning using the Empirical Mode Decomposition [6]

data points. For classification tasks, the DBSCAN density-based clustering technique is utilized for meta-information generation. DBSCAN algorithm requires a radius and a minimum number of points parameters to form clusters within the data. The radius specifies a region which can be identified as dense based on the number of points located within this area, and the minimum number of points to identify it as a dense region. DBSCAN designates either of this labels to each instance/point in the data: core, border, or noise point. A 'core' label is assigned when there is at least the minimum number of points within that radius. A 'border' label is assigned if it does not have the minimum number of data points around it but is within the radius of at least one core point. Finally, a point can be labelled as 'noise' when there are no other points within its radius. Also, a unique cluster-identifier is given to all points within every formed cluster. This information about the intrinsic structure of the data is appended with or substituted into the input attributes and the model is trained (see Figure 1). A meta-learning modified Constructive Backpropagation (CBP) neural network is used in [3], [4] and performs better than the traditional approach using the original CBP for classification on several real-world datasets.

B. In forecasting

While other studies in forecasting have applied forward search [15] and autoregressive integrated moving average-noise [16], we utilize EMD to create building-block trends by iteratively decomposing the time series data into several components known as the intrinsic mode functions (IMF). In the hierarchical structure for trend discovery, as shown in Figure 2, the time series data is divided into sequential partitions which are further decomposed by EMD. After which, a time series processing artificial neural network (ANN) is used for each partition to eliminate local interfering IMF components in the lower layer of the hierarchy. This sub-level training and generation of meta-information is completed by only summing beneficial IMF components of each partition and sequentially recombining the output of all the sub-ANN. The derived meta-information is used as an exogenous input to the higher level ANN to guide the learner by identifying useful information-bearing trends as the model's performance is finally optimized on the global trends of the time series data. From our research in [6], we applied this principle using the non-linear autoregressive exogenous (NARX) artificial neural network and the long short-term memory (LSTM) recurrent neural network; this technique showed performance accuracy enhancement compared to the unmodified NARX and LSTM networks.

IV. Discussion

By decomposing classification datasets into arbitrarily shaped clusters or time series data sequence into constituent trends, meta-information can be generated and used to guide the learning process. Therefore, better generalization can be observed through this structural-based learning using local and global learners. Decomposition with relatively high seasonality or meaningful clusters improves performance and lowers dependence on additional hidden-neurons. Notable drawbacks include an adverse effect on the performance due to excessive noise reduction and dependence on the successful decomposition of intrinsic pattern/clusters or trends/seasonality in classification and forecasting respectively [3], [5], [4], [6].

V. Conclusion

These techniques are proposed in order to address the challenges of obtaining classification data that is labelled/well-structured or forecasting data with a high signal to noise ratio. Applying these meta-learning and semi-supervised learning techniques to organise and make sense of huge quantities of data has become more important especially in the era of big data. Potential areas of application include data mining, function approximation, feature selection, knowledge handling, and training optimization.

References

[1] S. Weaver, L. Baird, and M. Polycarpou, "Using localizing learning to improve supervised learning algorithms," *IEEE Transactions on Neural Networks*, vol. 12, no. 5, pp. 1037–1046, 2001.

[2] C. Lemke, M. Budka, and B. Gabrys, "Metalearning: a survey of trends and technologies," *Artificial Intelligence Review*, vol. 44, no. 1, pp. 117–130, Jun. 2015.

[3] D. O. Afolabi, S.-U. Guan, F. Liu, K. L. Man, and P. W. Wong, "Class Interference Reduction through Meta-attribute Reinforced Learning," in *International Conference on Computing and Technology Innovation*, Luton, UK, May 2015.

[4] D. O. Afolabi, S.-U. Guan, B. Wu, and K. L. Man, "Evaluating the contribution of meta-attribute to noise reduction in machine learning," Dubai, UAE, Jan. 2016.

[5] D. O. Afolabi, S.-U. Guan, K. L. Man, and P. W. H. Wong, "Meta-learning with Empirical Mode Decomposition for Noise Elimination in Time Series Forecasting," in *Advanced Multimedia and Ubiquitous Engineering: FutureTech & MUE*, J. J. H. Park, H. Jin, Y.-S. Jeong, and M. K. Khan, Eds. Singapore: Springer Singapore, 2016, pp. 405–413, dOI: 10.1007/978-981-10-1536-6_53.

[6] D. Afolabi, S.-U. Guan, K. L. Man, P. W. H. Wong, and X. Zhao, "Hierarchical Meta-Learning in Time Series Forecasting for Improved Interference-Less Machine Learning," *Symmetry*, vol. 9, no. 11, p. 20, Nov. 2017. [Online]. Available: http://www.mdpi.com/2073-8994/9/11/283

[7] P. Mallapragada, R. Jin, A. Jain, and Y. Liu, "SemiBoost: Boosting for Semi-Supervised Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2000–2014, Nov. 2009.

[8] J. Hua Ang, S.-U. Guan, K. C. Tan, and A. A. Mamun, "Interference-less neural network training," *Neurocomputing*, vol. 71, no. 16-18, pp. 3509–3524, Oct. 2008.

[9] X. Xu, M. Ester, H.-P. Kriegel, and J. Sander, "A distribution-based clustering algorithm for mining in large spatial databases," in *Data Engineering, 1998. Proceedings., 14th International Conference on*. IEEE, 1998, pp. 324–331.

[10] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454. The Royal Society, 1998, pp. 903–995.

[11] W. W.-S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. USA: Addison-Wesley Pearson Higher Ed, 2006.

[12] B. Merk, C. V. Bratu, and R. Potolea, "Meta-learning enhancements by data partitioning," in *Intelligent Computer Communication and Processing, 2009. ICCP 2009. IEEE 5th International Conference on*. IEEE, 2009, pp. 59–62.

[13] K. Ramanathan and S. U. Guan, "Clustering and combinatorial optimization in recursive supervised learning," *Journal of Combinatorial Optimization*, vol. 13, no. 2, pp. 137–152, Dec. 2006.

[14] J. Song, S.-U. Guan, and B. Zheng, "Incremental hyper-sphere partitioning for classification," vol. 5, no. 2, pp. 72–88.

[15] M. Pesentia and M. Pirasa, "A Modified Forward Search Approach Applied to Time Series Analysis," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII. Beijing, China: ISPRS, 2008, pp. 787–792, beijing.

[16] W.-k. Wong and R. B. Miller, "Repeated Time Series Analysis of ARIMANoise Models," *Journal of Business & Economic Statistics*, vol. 8, no. 2, pp. 243–250, Apr. 1990.

ISBN: 978-988-14048-8-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)