



# Editorial note of the special issue on emerging topics in artificial intelligence selected from IEA/AIE2021

Jerry Chun-Wei Lin<sup>1</sup> · Ali Selamat<sup>2</sup>

Published online: 14 February 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

The 34th International Conference on Industrial, Engineering, and Other Applications of Applied Intelligent Systems (IEA/AIE) was held in Kuala Lumpur, Malaysia (virtually), July 26–29, 2021. After reviewing all accepted papers, the authors of the top-9 accepted papers were given the opportunity to submit the extension of their accepted papers for this special issue. They were also given the opportunity to significantly expand their contributions to include manuscripts suitable for publication in *Applied Intelligence*. All of the submitted and approved papers had a common focus, namely, dealing with applications of various artificial intelligence approaches and techniques. For a problem of practical and industrial importance, some of the articles focused more on the fundamentals, while others presented a particularly practical solution for a particular application area.

Kawalerowicz et al. presented a paper which introduced the CBOP concept, which uses historical build results and metrics derived from the software repository to create and use a model that classifies and source code changes made by the developer. To investigate the CBOP concept, the authors conduct a small experiment with repeated measurements and two conditions, and reproduce the experiment in a real-world, business-oriented software development project. In this early evaluation of CBOP, the Failed Build Ratio (FBR) - the ratio of failed build results to all other build results - is evaluated. Surprisingly, the study shows a tiny increase in FBR when CBOP practices are applied, even if the impact is small. The authors propose the function of the authority principle in this phenomenon. The authority concept is rarely

mentioned in the context of software development practices, but this explanation is compelling. To explore other possible explanations for the low decline in FBR, the authors analyze the predictive model to determine its quality. In addition, the authors conduct a survey inspired by the Technology Acceptability Model (TAM) among experiment participants and industry experts to determine the acceptability of CBOP and the tool, and describe their results. Compared to developers who participated in the experiment, industry experts who did not use CBOP show higher acceptance.

In the first technique presented in the paper by Huynh et al., RNN structures are used in addition to Conv1D. In the second method, the authors use DisllBert for two instances of case-sensitive words. It can help vectorize features (such as title, abstract, and keywords) and then use Conv1d to extract the features. In addition, the authors provide a novel computational technique for similarity evaluation of target and scope with other features, which can be called as DislBertAims. It can help to continuously update the weights of the similarity calculation and then find more data. In terms of top-k accuracy ( $k=1, 3, 5, 10$ ), the experimental results show that the second strategy could achieve better performance (62.46%, 90.32%, 94.89%, 97.96%) than the best performance in previous studies (50.02%, 78.89%, 86.27%, 93.23%). Interestingly, the top strategy in this work has better top-1 accuracy than the best from the previous study by 12.44%.

Sarkar et al. presented several solutions to eliminate the false alarm in the industry. The proposed unsupervised solution, the Mul-Generaons Tree (MGTree) method, minimized false-positive alarms and was equally effective on small and large data sets. Several industrial datasets, including those from Yahoo, AWS, GE, and machine sensors, were evaluated using MGTree. From the empirical evaluation, it shows that MGTree outperforms Isolaon Forest (iForest), One Class Support Vector Machine (OCSVM), and Ellipse Envelope in terms of anomaly detection accuracy (True-Posive and False-Posive). Weighted Time-Window Moving

---

✉ Jerry Chun-Wei Lin  
jerrylin@ieee.org

Ali Selamat  
aselamat@utm.my

<sup>1</sup> Western Norway University of Applied Sciences, Bergen, Norway

<sup>2</sup> University of Technology Malaysia, Johor Bahru, Malaysia

Estimation (WTM) is a time series prediction technique that does not depend on the static features of the dataset and has been tested on many time series datasets. The hybrid combination of WTM and MGTTree, Uni-variate Mul-Generations Tree (UVMGTTree), outperforms OCSVM, iForest, SARIMA, and Elliptic Envelope in identifying anomalies in time series datasets. The developed technology can have a tremendous impact on predictive maintenance and monitoring the health of industrial systems across all sectors, saving the operations team significant time and effort in addressing false alarms.

The paper presented by Dang et al. focuses on the identification of DNA motifs that satisfy the 2-optimality postulate. The authors present a novel hybrid genetic (HG1) algorithm based on elitism and local search strategies. To maintain the balance between exploration and exploitation, a unique elitism approach and a longest distance strategy are implemented. Based on the proposed balance between exploration and exploitation, a novel hybrid genetic algorithm (HG2) is constructed. Simulation results show that these algorithms generate high quality DNA molecules. The HG2 algorithm generates the highest quality DNA molecules.

In their paper, Phan et al. introduce an approach to local search called Potential Solution Improving (PSI), which aims to improve the performance of MOEAs by improving specific potential solutions on proximity fronts. The fundamental bottleneck in NAS is the expensive computation required to train a large number of candidate designs to evaluate their correctness. Recently, Synaptic Flow was introduced as a measure that accurately describes the performance of deep neural networks without performing a training period. Therefore, the authors propose that our PSI algorithm can use this training-free parameter as a proxy for network precision during local search phases. In experiments, the well-known MOEA Non-dominated Sorting Genetic algorithm II (NSGA-II) is used in conjunction with the training-free PSI local search to solve NAS problems generated by the standard NAS-Bench-101 and NAS-Bench-201 benchmarks. Experimental results confirm the efficiency gains achieved by our proposed strategy, which reduces the computational cost by a factor of four compared to the standard methodology.

Rahimi and Granmo present the first reinforcement learning framework based on the Tsetlin machine. The designed approach combines the value iteration technique with the regression Tsetlin machine to approximate the value function. The authors present a modified feedback mechanism of the Tsetlin machine that responds to the dynamic nature of the value iteration to provide an accurate estimate of the state value outside of the policy. The Tsetlin machine is able to unlearn and recover from the initial incorrect results. Transferring the naturally continuous nature of learning state

values to the proposed architecture of the Tsetlin machine via probabilistic updates is a significant difficulty but the authors solve this issue. This technique is accurate in the off-policy, but much slower than neural networks in the on-policy. By incorporating multilevel time difference learning in conjunction with high-frequency propositional logic patterns, the designed model can reduce the performance difference. Although the developed system is built on single-stage AND rules in propositional logic, it outperforms equivalent neural network models, as shown by many Gridworld examples. Finally, the authors propose how the class of models learned by our Tsetlin machine for the gridworld problem can be transformed into a graph structure that is more comprehensible. The graph structure captures the approximation to the state value function and the associated policy discovered by the Tsetlin machine.

Ahmed et al. presented a study which presents an embedded architecture based on distributed learning for transaction classification. The model converts transaction data into frequent sets and supports encode-decode model, thus, the model can capture related vectors in low dimensions while maintaining context and location. The authors explored a high-dimensional collection of transactional data to evaluate attention-based techniques and federated learning. The proposed approach reduces the global loss function to increase decision bounds while maintaining privacy and security. In the experiment, four datasets are used for comparison. The data are randomly selected and delivered to different clients for each dataset. To test active learning, the authors run each experiment with five random subsets of the dataset. The size of the training dataset remains constant between rounds, but the test dataset is not scored between rounds. Using the F1-score and the percentage of the dataset used, the developed techniques are compared to the best performing baseline method. The proposed model performed better than the baseline model in terms of percentage gains and output classes.

Kiran et al. extended the state-of-the-art approach by introducing the high-dimensional SHUI miner algorithm (HDSHUI-miner). The developed approach explores a set of unique pruning techniques to reduce the search space and computational cost required to find the target object sets. HDSHUI-Miner outperforms SHUI-Miner in terms of memory consumption, performance, and scalability, as shown by experimental results on seven real-world datasets. Finally, the authors present two real-world case studies to demonstrate the application of the proposed approach.

In the paper authored by Jodlowiec et al., first a language for universal and adaptable data modeling EGG is presented, along with methods for adaptation (extensions and generalizations). According to the first use case scenario, Extended Generalized Graph (EGG) can replace existing

data modeling languages for domain-specific data modeling activities. Second, the study presents and implements (for some data modeling languages, including RDF, XML, RDBM, UML, and AOM) a unique approach to measure and compare data modeling languages by mapping their metamodels to the EGG metamodel. Consequently, based on the second application scenario, the EGG metamodel can be used as a reference metamodel for research to compare the expressiveness of data modeling languages. It can also facilitate the selection of a data modeling language for a particular domain-specific data modeling activity. Third, the EGG described in this research prevents the transformation of reality to meet the data modeling language requirements because it is sufficiently generic for domain data modeling. The complete abstract syntax of the extended generalized

graph is presented and formulated in terms of its implementations in the Association-Oriented Metamodel and the Unified Modeling Language. Each semantic category of the abstract syntax is characterized. The study also presents two complete concrete syntaxes for the Extended Generalized Graph. The case studies of social network modeling and knowledge modeling demonstrate the applicability and utility of EGG. The abstract syntax is compared to a number of other metamodels. A quantitative measure is given as a summary of the comparative study of the case study models first developed in different metamodels and then articulated in the EGG metamodel.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.