eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Cost-Sensitive Weighting and Imbalance-Reversed Bagging for Streaming Imbalanced and Concept Drifting in Electricity Pricing Classification

Wing W. Y. Ng, *Senior Member, IEEE,* Jianjun Zhang, Chun Sing Lai*, *Student Member, IEEE,* Witold Pedrycz, *Fellow, IEEE,* Loi Lei Lai*, *Fellow, IEEE,* and Xizhao Wang, *Fellow, IEEE,*

Abstract—In data streaming environments such as smart grid, it is impossible to restrict each data chunk to have the same number of samples in each class. Hence, in addition to concept drift, classification problems in streaming data environments are inherently imbalanced. However, streaming imbalanced and concept drifting problems in power system and smart grid have rarely been studied. Incremental learning aims to learn correct classification for future unseen samples from the given streaming data. In this work, we propose a new incremental ensemble learning method to handle both concept drift and class imbalance issues. The class imbalance issue is tackled by an imbalance-reversed bagging method which improves the true positive rate while maintains a low false positive rate. The adaptation to concept drift is achieved by a dynamic cost-sensitive weighting scheme for component classifiers according to their classification performances and stochastic sensitivities. The proposed method is applied to a case study for the electricity pricing in Australia to predict whether the price of New South Wales will be higher or lower than that of Victorias in a 24-hour period. Experimental results show the effectiveness of the proposed algorithm with statistical significance in comparison to the state-of-the-art incremental learning methods.

Index Terms—Electricity Pricing, Imbalanced Classification, Incremental Learning.

Wing W. Y. Ng and Jianjun Zhang are with Guangdong Provincial Key Lab of Computational Intelligence and Cyberspace Information, School of Computer Science and Engineering, South China University of Technology, Guangzhou, China (e-mail: wingng@ieee.org, jjzhangscut@gmail.com).

Chun Sing Lai is with School of Civil Engineering, Faculty of Engineering, University of Leeds, LS2 9JT, U.K. and also with Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou, 510006, China (e-mail: c.s.lai@leeds.ac.uk).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6R 2V4 AB Canada, also with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia, and also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland (e-mail: wpedrycz@ualberta.ca).

Loi Lei Lai is with Department of Electrical Engineering, School of Automation, Guangdong University of Technology, Guangzhou, 510006, China (e-mail: l.l.lai@ieee.org).

Xizhao Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: x-izhaowang@ieee.org).

*Chun Sing Lai and Loi Lei Lai are corresponding authors.

I. INTRODUCTION

ELECTRICITY pricing plays a key role in determining short-term operating schedules and bidding strategies in competitive electricity markets [1]. Hence, many data-driven machine learning methods have been developed to predict short-term electricity market prices [2][3][4][5]. However, current methods tend to predict the exact value of prices while in some situations of the electricity market, it is not necessarily for every participant to know the exact value. For examples, demand-side market participants may not react until prices exceed certain thresholds considering the on/off nature of most electric loads [2]; some facilities only purchase electricity from the grid if the electricity price is below the marginal cost of operating the on-site electricity generation equipment [6]. In these situations, participants are not interested in knowing the exact value of prices and the price forecasting problem is turned into a classic price classification problem. The task of the electricity pricing classification problem is to classify future prices into several classes of interest, for instances, whether the future price is higher than a predefined threshold so that one should turn off most electric loads, or whether the prices in a city will be higher than the other city so that a better schedule of electricity transmission between these two cities should be planned ahead of time.

Electricity pricing classification problem is not an easy task due to its streaming nature [7]. Data generated from the grid form a data stream, which introduces new challenges to traditional machine learning approaches, such as limited training and testing time, constraint of memory usage, and a single scanning of incoming samples [8]. More importantly, concept drift and class imbalance problems are prevalent in streaming environment which may lead to accuracy drop of classifiers. Concept drift refers to the change of the joint probability distribution between inputs and true classifications in different time moments in a data streaming setting [9]. When concept drift occurs, it changes statistical characteristics of target concept over time which results in classifiers trained using outdated samples would yield very poor generalization capability on samples in the future. Ensemble methods are often applied to relieve this problem because of their high performance and usefulness for streaming learning owing to the ease of being integrated with drift detection methods and dynamic updates [10].

Class imbalance problems occur when the number of sam-

ples in at least one class is either much more or less than other classes. When class imbalance happens, traditional classifiers yield poor generalization performance on the minority class (In a two-class problem, a class containing more training samples than the other class is referred as the majority class while the other one is referred as the minority class.) [11]. Therefore, proper techniques like data processing should be employed to deal with the class imbalance problem. Data processing is one of the key elements for the successful operation of complex systems such as smart grids [12][13]. In the context of smart grid, classification with machine learning has been applied to fault cause identification [14][15], future electricity market prices [2][3][4][5], electrical machines [16][17], power quality disturbances classification [18], and cyber-attacks detection [19][20]. However, to date, the concepts of imbalanced streaming data and concept drifting in smart grid have rarely been studied. Seldom work has been done on imbalanced classification for power system problems. Authors in [21] claimed to be the first researches on investigating the outliers in electricity demand time series with imbalanced classification techniques. To assess power system short-term voltage stability, an oversampling technique and a cost-sensitive learning method are applied to deal with the predictions of the rarely-occur instability events [22].

Very few efforts have been made to deal with both concept drift and class imbalance problems. Existing methods can be distinguished into two types. One is to retrain a new model using the most recent samples so that the trained classifier can react to the concept change fast, for example the SERA (SElectively Recursive Approach) [23]. The SERA reserves all the minority samples seen so far, from which the most relevant ones are selected to combine with the most recent majority samples so that a pre-selected post-balance ratio is met. A classifier or ensemble is trained from this rebalanced dataset. The other type is to dynamically update the model, for example the Learn++.CDS (Concept Drift with SMOTE), Learn++.NIE (Non-stationary and Imbalanced Environments) [24] and the DWMIL (Dynamic Weighted Majority for Imbalance Learning) [25]. The CDS rebalances the most recent data chunk (A data chunk refers to a block of consecutive samples in between some time interval for the learning model to train or to predict.) using the SMOTE (Synthetic Minority Oversampling TEchnique) [26] to tackle the class imbalance problem by generating new samples along a line connecting a minority sample and its nearest minority sample, while the NIE uses a bagging-variation method to create several relatively balanced dataset to train a classifier ensemble. Regarding the adaptation to the concept drift, both the CDS and the NIE apply a dynamic weight assignment scheme so that classifiers yielding high performance on the current data environment receive high weights. The major drawback of the Learn++ family is that all classifiers are maintained which increases the computational costs and lowers the prediction speed. To avoid this kind of problem, the DWMIL applies a time-decay function to its weight assignment scheme so that the weight of each classifier decreases automatically. When weights are lower than a threshold, corresponding classifiers are removed so that the number of classifiers maintained is much lower than the

number of time moments.

The major concern of classifier training is their generalization abilities for future unseen samples in incoming data stream. However, current learning methods do not take generalization error of the classifiers into account when training classifiers. Therefore, we propose an incremental ensemble of ensembles learning with a Cost-sensitive Weighting and an Imbalance-reversed Bagging, i.e. CWIB, to deal with both concept drift and class imbalance issue, which significantly enhances the performances than the state-of-the-art methods in terms of accuracy, F1-measure, and G-mean and ranks the first in terms of all performance metrics applied in this work. The CWIB relieves the class imbalance problem by applying an imbalance-reversed bagging method which builds a set of diversified base classifiers to form a component ensemble classifier. In comparison with methods building a single classifier with each data chunk and update weights of classifiers (e.g. CDS in the experiment), the proposed CWIB yields significantly better results in accuracy, F1-measure, and G-mean value. This shows the effectiveness and satisfactory results of the CWIB using ensemble of classifiers. Ensemble of classifiers usually yields lower error rate in comparison with a single classifier [27]. Moreover, training an ensemble of classifiers using the current chunk looks to be very time-consuming, but these component classifiers are independent from each other and can be trained in parallel as suggested in [28]. In this way, time consumption will be roughly similar to that of training a single classier. Therefore, training an ensemble of classifier instead of a single classifier with each data chunk is a better choice. Then, component classifiers are fused together to form the final ensemble for the CWIB using the weighted sum method. To adapt to the concept changes across time, the weight of each component classifier is computed according to their cost-sensitive classification performances and stochastic sensitivities with respect to the current data chunk. Major contributions of this work are as follows:

1) An imbalance-reversed bagging (IRB) method is proposed to relieve the class imbalance issue in a data chunk. The IRB boosts the true positive rate while maintains a relatively low false positive rate.

2) A new cost-sensitive stochastic sensitivity measure (ST-SM) is proposed to weight samples in different classes differently based on their ST-SM and a cost computed by the imbalance ratio.

3) A dynamic cost-sensitive weighting scheme based on the cost-sensitive ST-SM is applied to compute fusing weights of component classifiers. A larger weight is assigned to a component classifier yielding a good cost-sensitive performance on the current data chunk.

4) A fixed size of classifier ensemble is maintained, which is much smaller than the time moments and requires both less computational resources and less storage.

The paper is structured as follows. The CWIB is proposed in Section II. Section III shows experimental results and discussion. We conclude the paper in Section IV.

II. Cost-Sensitive Weighting And Imbalance-Reversed Bagging

The Algorithm 1 shows procedures for training the CWIB at time t. The CWIB training method consists of two components: one is to handle the class imbalance issue and the other one is to dynamically assign different weights to each component classifier for adaptation to changes in data. The overall procedure of the CWIB is as follows:

At time moment t, the current H^{t-1} is an ensemble of ensembles which consists of a set of component classifiers fused by a weighted sum while each component classifier consists of a set of base classifiers fused by a simple majority voting. When a new data chunk arrives, a new component classifier is trained using the IRB. The new component classifier h^t is expected to be more relevant to the current data environment, thus its weight is set to 1 (the largest weight). Existing component classifiers in H^{t-1} are then weighted according to their classification performance based on a cost-sensitive loss and their stochastic sensitivities on the current data chunk. Then, the H^{t-1} is combined with the h^t along with newly computed weights to from the H^t. If the number of component classifiers in H^t is larger than the pre-selected ensemble size, the worst performing component classifier yielding the smallest weight is removed. The IRB and the dynamic weighing scheme are proposed in Sections II-A and II-B, respectively.

Algorithm 1 Training CWIB at time moment t

Require:

   t, current time moment; u, number of base classifiers in a component classifier; D^t, current data chunk; H^{t-1}, current classifier ensemble; M, maximum ensemble size.

Ensure:

   H^t, classifier ensemble at time t.

1: Apply the IRB on D^t to build the component classifier h^t and set its weight to 1.

2: Compute the weight of each component classifier in H^{t-1} based on its cost-sensitive loss and stochastic sensitivity

3: Combine h^t and H^{t-1} with their newly computed weights to form H^t.

4: If the number of component classifiers in H^t is larger than M then removes the component classifier yielding the smallest weight.

A. Imbalance-Reversed Bagging

When a new data chunk arrives, u datasets are sampled from the original data chunk with replacement based on a probability distribution. The probability of a minority (majority) sample being sampled is equal to the number of samples in the majority (minority) class divided by the total number of samples in the current data chunk. Such that, the minority samples becomes the majority in the sampled dataset. In this way, the class imbalance is reversed which forces the base classifier being built using this dataset to bias to the minority class for improving its true positive rate. Then, a component

classifier of the CWIB is built by fusing all u base classifiers using a simple majority voting.

However, in electricity pricing problems, the number of samples in the minority class may be larger than that of the majority class in some data chunks. In these anomaly cases, the probability of sampling will not be reversed as aforementioned to let the system to keep focus on the original minority class. The random sampling with replacement from both classes creates diversified training datasets for base classifiers. By favoring the minority class in the IRB, each base classifier may yield a high false positive rate. The bagging of diversified base classifiers maintains a low false positive rate [29] to relieve this problem. By applying the IRB, the representation of the minority class is enhanced while a relatively low false positive rate is maintained. The algorithm of the IRB is given in Algorithm 2.

Algorithm 2 Imbalance-Reversed Bagging

Require:

   u, number of base classifiers in a component classifier; D, current data chunk.

Ensure:

   A classifier ensemble h.

1: Reverse the imbalance ratio of D if the number of minority samples is not larger than that of the majority class (anomaly case).

2: Randomly sample u datasets with replacement from D according to the imbalance ratio in Step 1.

3: Build u base classifiers using the u training datasets and form the classifier ensemble h with a simple majority vote fusion

---

B. Dynamic Weight Assignment

The IRB proposed in the previous section builds a new component classifier whenever a new data chunk arrives. The adaptation to concept drifts in the non-stationary streaming data environment is achieved by a dynamic weighting of component classifiers according to their classification performances and stochastic sensitivities with respect to the current data chunk. The weight is ranged between $[0, 1]$. A larger weight is assigned to a component classifier if it yields a higher classification performance for the current data environment with smaller stochastic sensitivity with respect to small input perturbations. The final classifier ensemble of the CWIB is fused by the weighted sum method as follows:

$$H^t(x_b^t) = sign(\sum\nolimits_{j=1}^{l^t} w_j^t h_j^t(x_b^t)) \qquad (1)$$

where x_b^t, $h_j^t(x_b^t)$, w_j^t and l^t denote the bth training sample, the predicted output of the jth component classifier given x_b^t, the weight of the jth component classifier and the number of component classifiers, at time t, respectively. For simplicity, the time t will be ignored in the following part of this section because all computations are finished within the same time moment. Therefore, Equation (1) is rewritten as follows:

$$H(x_b) = sign(\sum_{j=1}^{l} w_j h_j(x_b)) \qquad (2)$$

The classification performance of h_j (the jth component classifier) is evaluated by a cost-sensitive loss function. In class imbalance problems, misclassifying a minority sample is usually more costly than misclassifying a majority one. Therefore, a misclassification of a minority sample yields a larger penalty in the loss function. The logistic loss function is used in this work:

$$\varphi(h_j(x_b), y_b) = log(1 + exp(-y_b h_j x_b)) \qquad (3)$$

where $y_b \in \{-1, +1\}$ and $h_j(x_b) \in \{-1, +1\}$ denote the true label and the predicted label of x_b, respectively. Then, the cost-sensitive loss function is defined as follows:

$$L_j = \sum_{b}^{m} C_b \varphi(h_j(x_b), y_b) \qquad (4)$$

The misclassification cost (C_b) is equal to $\frac{N^-}{N^+}$ if x_b belongs to the minority class and 1 otherwise where $N^-(N^+)$ denotes the number of majority samples (minority samples). Then, the classification weight ($w_{c|j}$) is inversely proportional to the cost-sensitive loss and written as follows:

$$w_{c|j} = \frac{exp(-L_j)}{\sum_j exp(-L_j)}, j = 1, 2, \cdots, l^{t-1} \qquad (5)$$

where l^{t-1} denotes the ensemble size of H at time moment $t-1$. The weight of the newly trained component classifier at t is equal to 1. Therefore, only weights for the l^{t-1} component classifiers in H^{t-1} need to be computed.

On the other hand, the sensitivity of a component classifier is evaluated by the cost-sensitive stochastic sensitivity measure (ST-SM). The ST-SM [30] has been widely applied in different applications, for instances neural network architecture selection [30], sample selection [31], MLPNN training [32], feature selection [33], steganalysis [34], and business intelligence [35]. The ST-SM of the jth component classifier is defined as the expectation of squared differences between outputs of training samples and samples located within a distance of Q in each dimension:

$$E_{S_Q}((\Delta y)^2) = \frac{1}{m} \sum_{b=1}^{m} E[(f(x_b + \Delta x) - f(x_b))^2] \quad (6)$$

where $\Delta x \in [-Q, +Q]^n$ denotes the perturbation of the training sample and f is the real-valued outputs before thresholding to $\{-1, +1\}$ of a component classifier. Intuitively, the ST-SM measures the fluctuation of classifier outputs with respect to input perturbations, that is, it measures the stability of the classifier. Therefore, a classifier yielding a large ST-SM value is easily affected by small perturbations of inputs and more unstable. As a result, a smaller weight should be assigned to a classifier yielding a higher ST-SM value.

In this work, we propose the cost-sensitive ST-SM which is defined as follows:

$$S = \frac{1}{m} \sum_{b=1}^{m} C_b E[(f(x_b + \Delta x) - f(x_b))^2] \qquad (7)$$

A quasi-Monte Carlo-based method is adopted to calculate the cost-sensitive ST-SM of a classifier as in [32]. Specifically, Δx is generated via an n-dimensional Halton sequence [36] with each coordinate ranging from $[-Q, Q]$ using MATLAB and 50 Halton points are used in the calculation of the expectation term in (7). According to experiments in [32], 50 Halton points yield only around 4% estimation error and the computational time is fast. Higher number of Halton points can be used for more accurate estimation but with higher computational costs.

It is difficult to automatically select the Q value theoretically. In implementations, $Q = 0.1$ is usually used which indicates a maximum of 10% of deviation from the training samples for dataset with input features being normalized to $[0, 1]$.

In the theory of the Localized Generalization Error Model [30], a good classifier should minimize both the classification error and the ST-SM. Therefore, the proposed weighting scheme assigns larger weights to classifiers yielding smaller cost-sensitive ST-SM values. The sensitivity weight is inversely proportional to the cost-sensitive ST-SM and written as follows:

$$w_{s|j} = \frac{exp(-S_j)}{\sum_j exp(-S_j)}, j = 1, 2, \cdots, l^{t-1} \qquad (8)$$

Then, the fusion weight of the final ensemble of the CWIB is defined as the combination of the classification weight and the sensitivity weight as follows:

$$W_j = \eta W_{c|j} + (1 - \eta) W_{s|j}, j = 1, 2, \cdots, l^{t-1} \qquad (9)$$

where η is a trade-off coefficient between the classification performance and the stability of component classifiers. In our experiment, $\eta = 0.5$ is used to represent an equal importance of these two factors. The final decision of the ensemble is the weighted sum of outputs of all component classifiers:

$$H^t(x_b) = sign(\sum_j w_j h_j(x_b)) \qquad (10)$$

III. EXPERIMENTAL STUDIES WITH ELECTRICITY PRICE CLASSIFICATION

As mentioned in [4], electricity price is a complex signal due to its characteristics of nonlinearity, time variant, and non-stationary behavior. More robust and accurate price classification and forecasting methods are still needed. As an example, for electricity price forecasting, authors in [4] proposed a complex electricity price forecasting technique based on feature selection and cascaded neuro-evolutionary algorithm (CNEA). The CNEA consists of cascaded forecasters, with each forecaster made up of an evolutionary algorithm and neural network. The adjustable parameters in the feature selection algorithm and the CNEA are fine-tuned with an iterative search procedure. However, the data segmentation for model training, i.e. optimal data size for training, was not well studied. To predict the day-ahead price, authors used a rule-of-thumb and the model was trained according to previous 50 days of data. The nature of the data, i.e. data imbalance or concept drift has not been considered prior training the model.

The work in [3] investigated several data mining approaches for electricity price classification. This includes correlation-based feature selection, multilayer perceptron, K-nearest neighbors etc. Similar to forecasting problems, the data segmentation for classification problems has been arbitrary. For example, the authors used 20 historical days for the model training, with an argument as a fair comparison to the previous work in [2]. Evidently, it has been observed that previous research efforts have not considered electricity market price classification in data streams.

In this section, the CWIB is compared with existing methods designed for imbalanced data streaming classification problems with concept drift. The electricity dataset used in the experiment is introduced in Section III-A. Section III-B studies effects of parameters to the CWIB. Section III-C presents and discusses experimental results of the CWIB and other methods. The CWIB is compared with the following state-of-the-art methods where the default values for the parameters are used as suggested in the literatures:

- CDS [24]: When a new data chunk arrives, a new classifier is trained using data rebalanced by the SMOTE. Each classifier is weighted based on a time-decay function and its performance on current data chunk.
- NIE [24]: The differences between CDS and NIE are that NIE trains a sub-ensemble when a new data chunk arrives and NIE uses different error metrics to evaluate its sub-ensembles. By using different error metrics, the NIE can be distinguished into three variations, which are WRM (weighted recall measure), FM (F1-score measure), and GM (geometric-mean measure).
- SERA [27]: When a new data chunk arrives, the SERA trains a new ensemble using the current data chunk and the most relevant historical minority samples. All minority samples seen so far are preserved and those with the smallest Mahalanobis distance from current minority samples are selected as part of the training samples so that a pre-selected post-balance ratio is met.
- DWMIL [25]: The DWMIL trains a new sub-ensemble for each data chunk using UnderBagging and weights each sub-ensemble based on their performance to the current data chunk. The weights are reduced based on both a poor performance and /or the age of the sub-ensemble over time.

The numbers of component classifiers and base classifiers used in the CWIB are set to be 10 and 5, respectively. Larger numbers of component and/or base classifiers could be used to better adapt to the gradual drifts if higher computational costs are allowed. Radial Basis Function Neural Networks (RBFNN) with 10 hidden neurons are used as the base classifiers in all algorithms for fair comparisons. Neural networks have been successfully applied in future electricity forecasting [5][37]. RBFNN is used here for its universal approximation capability [38] and its fast training speed compared with other types of neural networks, e.g. multilayer perceptrons. Ten independent runs are performed for all methods to reduce random effects. The AUC (Area Under Curve), the F1-measure, the G-mean (geometric mean), and the Accuracy are used to compare performances of each method.

A. Electricity Dataset

The Electricity Pricing dataset Elec2 [39] is used in our experiment to simulate the concept drifting and class imbalance environment, which originally contains $45,312$ samples drawn from 7 May 1996 to 5 December 1998 with one sample for every half an hour from the electricity market in New South Wales, Australia. Samples with missing features have been removed so the remaining dataset contains only $26,975$ samples. This dataset provides time and demand fluctuations in the price of electricity in New South Wales, Australia. The day, period, New South Wales electricity demand, Victoria electricity demand, and the scheduled electricity transfer between the two states are used as the input features to predict whether the price of New South Wales will be higher or lower than that of Victorias in a 24-hour period. Usually, a data chunk consists of 336 samples. However, samples with missing values are removed. Therefore, some data chunks may consist of fewer samples.

The concept drifts in this dataset is natural and unavoidable because the electricity prices change with demand over different time periods. Moreover, the imbalance ratio between two classes changes over time and the majority and the minority classes may swap over time. Table I shows the characteristics of the dataset. The imbalance ratio is defined as the ratio of the number of minority samples over that of majority samples. A special case of concept drift for this dataset occurs when the imbalance ratio exceeds 1 because of the minority positive class becomes the majority class in some time moments. Fig. 1 shows the imbalance ratios of the Elec2 dataset over time, where the y-axis represents the imbalance ratio which is unitless because it is a ratio of two numbers and their units cancel each other. The x-axis represents the time moment/ time step of each data chunk arriving which contains roughly one week of data (since some data with missing values are removed). The effects of imbalance drift on the performance of learning models have been systematically analysed in [40][41], showing that without properly handling the drift of imbalance, changes in imbalance status would negatively affect the performance. In this work, we propose to apply the IRB to handle the class imbalance problem to avoid severe performance deterioration caused by the drift of class imbalance.

TABLE I: Characteristics of the Elec2 dataset

Size of data chunk	# features	# time moments	Imbalance ratio
328~329	5	82	0.27~1.63

B. Effects of Different Parameters

The CWIB uses two parameters: the number of component classifiers and the number of base classifiers in a component classifier. Experiments are carried out to show different behaviors of the CWIB by using different sets of parameters.

Fig. 2a and 2b show the true positive rate (TPR) and false positive rate (FPR) of both the minority and majority class by
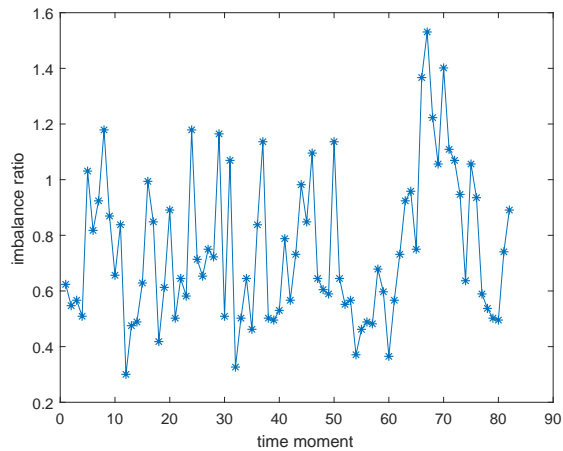
Fig. 1: Imbalance ratios of Elec2 dataset over time

varying numbers of component classifiers and base classifiers, respectively, where TPR and FPR are unitless because they are defined as the ratio of two numbers and their units cancel each other. From Fig. 2a, with the increment of the number of component classifiers, the true positive rate of the minority class increases while the false positive rate of the minority class decreases. Both curves of the true positive rate and the false positive rate of the minority class tend to converge when the number of component classifiers is around 10. In contrast, the true positive rate of the majority class decreases and the false positive rate of the minority class increases when the number of component classifiers increases. This is because the IRB reverses the class imbalance ratio and the representation of minority class is enhanced while the representation of majority class is diminished. The performance gained on the minority class is higher in comparison with the minor classification performance loss on the majority class. So, the overall performance of the CWIB is enhanced. Therefore, the number of component classifiers is set to 10 in our experiments to yield relatively high true positive rates and relatively low false positive rates for both classes.

From Fig. 2b, with the increment of the number of base classifiers, all four curves are quite stable and start to converge when the number of base classifiers is around 5. The number of base classifiers seems to have very minor effects on the performance of the CWIB. Hence it is set to 5 to maintain a low computational cost and achieve a high classification performance.

C. Experimental Results

Fig. 3a to 3d show average values of the four performance metrics of different methods over 10 independent runs over time, respectively, where all four metrics are unitless because they are all calculated by certain operations on unitless quantities. Table II shows the mean and the standard deviation values of different metrics for different methods over all data chunks. The bolded value of each column indicates the best result yielded for this metric and the symbol "*" indicates a statistically significant difference between the CWIB and

the corresponding method by Student's t-test with 95% confidence. The number in the parenthesis is the rank of the method in terms of corresponding performance metric. The last column gives the average rank of each method over four metrics.

From Table II, the CWIB yields the best average rank in terms of all metrics. The CWIB outperforms the FM, GM, WRM, and SERA significantly in terms of all metrics. In comparison with the CWIB, both the DWMIL and the CDS yields only a bit worse performances in AUC , but much worse performances in all F1-measure, G-mean, and Accuracy (at least 2.74% differences). The high values of G-mean and F1-measure yielded by the CWIB indicate that the combination of the IRB and the cost-sensitive weighting scheme in the CWIB enhances accuracies of both classes (i.e. true positive rate and true negative rate). The SERA yields a relatively high Accuracy (ranks the third) but very poor ranks in terms of the AUC, the F1-measure, and the G-mean (all ranks the seventh). It may be due to the fact that decision boundaries created by the SERA are too biased to the majority class. This makes the SERA classify most samples as the majority class to achieve a high average accuracy but ignore the performance on the minority class. In contrast, the CWIB enhances the representation of the minority samples by applying the cost-sensitive weighting scheme, reversing the imbalance ratio, and at the same time a bagging method is employed to maintain a low false positive rate. Numerical results confirm the effectiveness of the proposed CWIB.

From Fig. 3a to 3d, performances of all methods fluctuate severely because of the type of concept drift is unknown and can be highly complicated. Moreover, the swapping between the minority and the majority classes further increases the difficulty of this learning task. The SERA yields the worst performance in all metrics and sometimes yields 0 value in terms of G-mean because the SERA maintains too many outdated minority samples which consistently deteriorate its performance. The rest of methods perform similarly and fluctuates severely as time varies.

In summary, experimental results show that the CWIB is effective and yields statistically significantly better results in comparison with state-of-the-art methods. Moreover, the CWIB uses fewer storage and computational costs by using a small fixed size ensemble in comparison to the very large ensemble size (equal to t) used by the CDS, the WRM, the FM, and the GM for a large t and the variable ensemble size used by the DWMIL.

IV. CONCLUSIONS AND FUTURE WORK

Concept drifts are prevalent in data streaming classification problems, such as electricity price classification. The consideration of concept drift and imbalanced data for electricity price is a novelty of this work. In general, the concept drifting problem is more complicated when numbers of samples in different classes are imbalanced. Therefore, the CWIB is proposed to deal with these two problems simultaneously. The CWIB dynamically weights component classifiers according to their cost-sensitive classification performances and stochastic sensitivities with respect to input perturbations. New
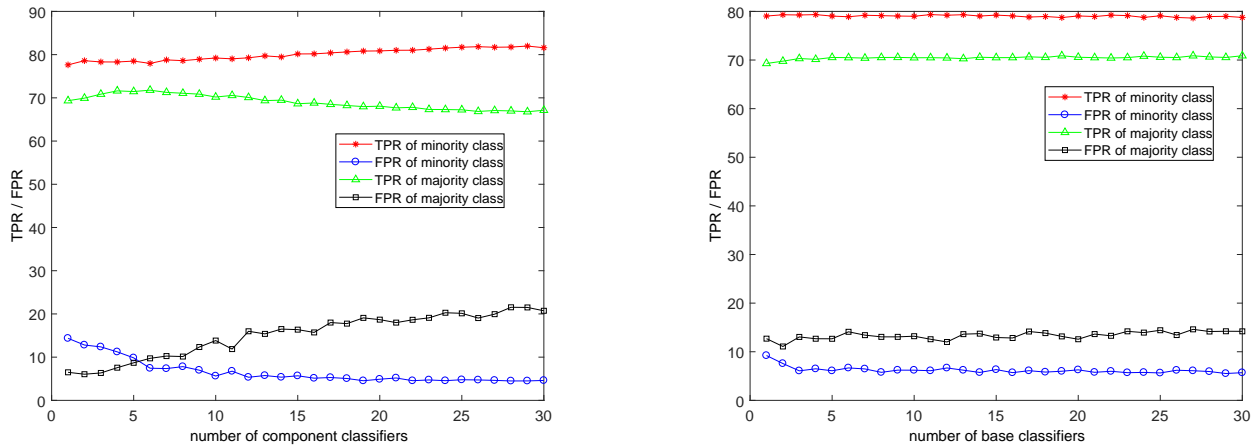
(a) TPR and FPR of both classes by varying the number of component classifiers

(b) TPR and FPR of both classes by varying the number of base classifiers

Fig. 2: Different behaviors of the CWIB by varying the parameters

TABLE II: Performance of different methods on the Elec2 dataset

	Accuracy	AUC	F1-measure	G-mean	Mean Rank
CWIB	**75.80±0.52(1)**	**85.47±0.36(1)**	**73.47±0.44(1)**	**72.20±0.37(1)**	**1**
DWMIL	73.06±1.12*(2)	85.14±0.37(3)	68.68±1.30*(6)	64.74±1.48*(6)	4.25
CDS	72.37±0.54*(6)	85.40±0.29(2)	70.00±0.60*(3)	68.90±0.68*(3)	3.5
FM	72.15±0.51*(7)	83.30±0.42*(5)	69.66±0.49*(4)	68.41±0.53*(4)	5
GM	72.84±0.58*(4)	83.37±0.45*(4)	70.37±0.60*(2)	69.13±0.59*(2)	3
WRM	72.41±0.41*(5)	83.29±0.41*(6)	69.47±0.36*(5)	67.63±0.44*(5)	5.25
SERA	72.98±0.44*(3)	71.83±0.75*(7)	65.53±0.62*(7)	57.70±1.73*(7)	6

component classifier is trained using the Imbalance-Reversed Bagging (IRB) method to cope with the imbalance issue in a data chunk. The ensemble of the CWIB maintains a constant size by removing the component classifier yielding the smallest weight.

Experimental results show that the proposed CWIB yields better Accuracy, AUC, G-mean, and F1-measure than state-of-the-art methods with statistical significance on an electricity pricing dataset. This shows that the proposed method is useful to energy and power researches when the classification problems has a dataset in a streaming form with class imbalance occurring in data chunks, e.g. prediction of outliers of electricity demand, fault diagnostic in power distribution system, and stochastic renewable energy generation, e.g. wind and solar. They are the future areas to be studied.

On the other hand, removing the component classifier yielding the smallest weight may not be optimal because it may reduce the diversity of the classifier ensemble. In our future works, we will research on the possibility of adding time as a component of the weight computation. The diversity between base classifiers and between component classifiers may also be added to the weight computation to enhance the overall diversity of the ensemble of ensembles of the CWIB. In addition to that, the RBFNN has been used as the base classifier. Multiple types of classifiers may be used to create the classifier ensemble for the CWIB. The optimal classifier combination and selection method will be one of our important future works.

REFERENCES

[1] A. J. Conejo, F. J. Nogales, and J. M. Arroyo, "Price-taker bidding strategy under price uncertainty," *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 1081-1088, 2002.

[2] H. Zareipour, A. Janjani, H. Leung, A. Motamedi, and A. Schellenberg, "Classification of future electricity market prices," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 165-173, 2011.

[3] D. Huang, H. Zareipour, W. D. Rosehart, and N. Amjady, "Data mining for electricity price classification and the application to demand-side management," *IEEE Transactions on Smart Grid*, vol. 3, no. 2, pp. 808-817, 2012.

[4] N. Amjady and F. Keynia, "Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm," *IEEE Transactions on Power Systems*, vol. 24, no. 1, pp. 306-318, 2009.

[5] M. Rafiei, T. Niknam, and M.-H. Khooban, " Probabilistic forecasting of hourly electricity price by generalization of ELM for usage in improved wavelet neural network," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 1, pp. 71-79, 2017.

[6] L. M. Saini, S. K. Aggarwal, and A. Kumar, "Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in national electricity market," *IET Generation, Transmission and Distribution*, vol. 4, no. 1, pp. 36-49, 2010.

[7] J. Gama, I. Zliobaite, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A survey on concept drift adaptation," *ACM Computing Surveys*, vol. 46, no. 4, pp. 1-37, 2014.

[8] D. Brzezinski, and J. Stefanowski, "Reacting to different types of concept drift: the accuracy updated ensemble algorithm," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 1, pp. 81-94, 2014.

[9] B. Krawczyk, L. L. Minku, J. Gama, J. Stefanowski, and M. Wozniak, "Ensemble learning for data stream analysis: a survey," *Information Fusion*, vol. 37, no. C, pp. 132-156, 2017.

[10] H. M. Gomes, J. P. Barddal, F. Enembreck, and A. Bifet, "A survey on ensemble learning for data stream classification," *ACM Computing Surveys*, vol. 50, no. 2, 23, 2017.

(a) Accuracy

(b) AUC
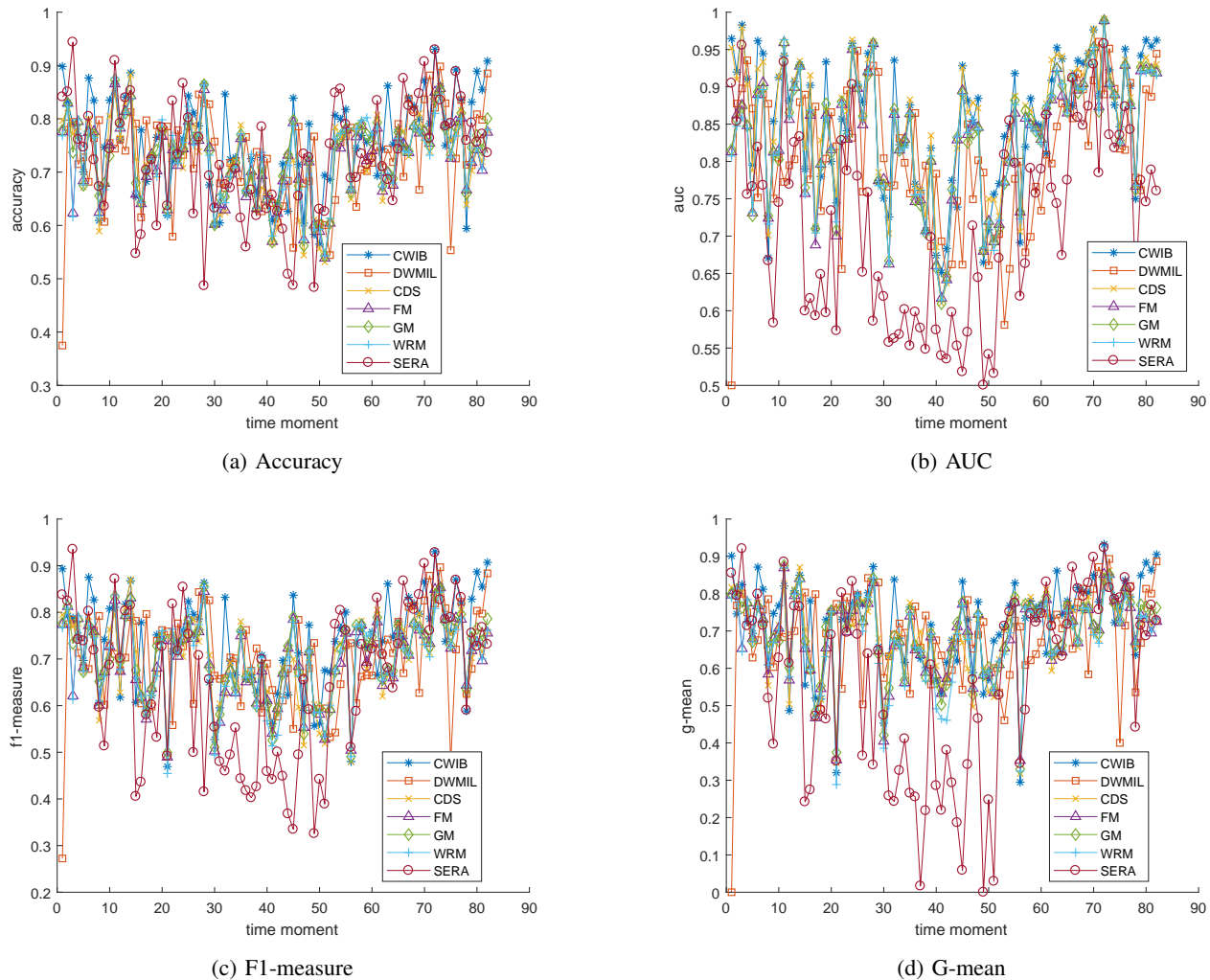

(c) F1-measure

(d) G-mean

Fig. 3: Experimental results of different methods for the Elec2 dataset

[11] P. Branco, L. Torgo, and R. P. Ribeiro, "A survey of predictive modeling on imbalanced domains," in *ACM Computing Surveys*, vol. 49, no. 2, article 31, 2016.

[12] X. Dai and Z. Gao, "From model, signal to knowledge: a data-driven perspective of fault detection and diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 9, pp. 2226-2238, 2013.

[13] D. Alahakoon, and X. Yu, "Smart electricity meter data intelligence for future energy systems: a survey," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 1, pp. 425-436, 2016.

[14] L. Xu, M.-Y. Chow, and L. S. Taylor, "Power distribution fault cause identification with imbalanced data using the data mining-based fuzzy classification *E*-algorithm," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 164-171, 2007.

[15] L. Xu and M.-Y. Chow, "A classification approach for power distribution systems fault cause identification," *IEEE Transactions on Power Systems*, vol. 21, no. 1, pp. 53-60, 2006.

[16] M. Cococcioni, B. Lazzerini, and S. L. Volpi, "Robust diagnosis of rolling element bearings based on classification techniques," *IEEE Transactions on Industrial Informatics*, vol. 9, no. 4, pp. 2256-2263, 2013.

[17] R. Razavi-Far, M. Farajzadeh-Zanjani, and M. Saif, "An integrated class-imbalance learning scheme for diagnosing bearing defects in induction motors," *IEEE Transactions on Industrial Informatics*, 2017.

[18] F. A. S. Borges, R. A. S. Fernandes, I. N. Silva, and C. B. S. Silva, "Feature extraction and power quality disturbances classification using smart meters signals," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 824-833, 2016.

[19] S. Pan, T. Morris, and U. Adhikari, "Classification of disturbances and cyber-attacks in power systems using heterogeneous time-synchronized

data," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 650-662, 2015.

[20] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630-1638, 2017.

[21] F. J. Duque-Pintor, M. J. Fernandez-Gomez, A. Troncoso, and F. Martinez-Alvarez, "A new methodology based on imbalanced classification for predicting outliers in electricity demand time series," *Energies*, vol. 9, no. 9, p. 752, 2016.

[22] C. Zhu, C. Lu, Z. Y. Dong, and C. Hong, "Imbalance learning machine-based power system short-term voltage stability assessment," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2533-2543, 2017.

[23] S. Chen, and H. He, "SERA: selectively recursive approach towards nonstationary imbalanced stream data mining," in *Proceedings of International Joint Conference on Neural Networks*, 2009.

[24] G. Ditzler, and R. Polikar, "Incremental learning of concept drift from streaming imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 10, pp. 2283-2301, 2013.

[25] Y. Lu, Y. M. Cheung, and Y. Y. Tang, "Dynamic weighted majority for incremental learning of imbalanced data streams with concept drift," in *Proceedings of Twenty-Sixth International Joint Conference on Artificial Intelligence*, pp. 2393-2399, 2017.

[26] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "S-MOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321-357, 2002.

[27] J. Gao, W. Fan, J. Han, and P. S. Yu, "A general framework for mining concept-drifting data streams with skewed distributions," in *SIAM International Conference on Data Mining*, 2007.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2850930, IEEE Transactions on Industrial Informatics

9

[28] K. Wu, A. Edwards, W. Fan, and K. Zhang, "Classifying imbalanced data streams via dynamic feature group weighting with importance sampling," in *SIAM International Conference on Data Mining*, 2014.

[29] M. A. Tahir, J. Kittler, and F. Yan, "Inverse random under sampling for class imbalance problem and its application to multi-label classification," *Pattern Recognition*, vol. 45, no. 10, pp. 3738-3750, 2012.

[30] D. S. Yeung, W. W. Y. Ng, D. Wang, E. C. C. Tsang, and X.-Z. Wang, "Localized generalization error model and its application to architecture selection for radial basis function neural network," *IEEE Transactions on Neural Networks*, vol. 18, no. 5, pp. 1294-1305, 2007.

[31] W. W. Y. Ng, J. Hu, D. S. Yeung, S. Yin, and F. Roli, "Diversified sensitivity-based undersampling for imbalance classification problems," *IEEE Transactions on Cybernetics*, vol. 45, no. 11, pp. 2402-2412, 2014.

[32] D. S. Yeung, J.-C. Li, W. W. Y. Ng, and P. P. K. Chan, "MLPNN training via a multiobjective optimization of training error and stochastic sensitivity," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 5, pp. 978-992, 2016.

[33] W. W. Y. Ng, D. S. Yeung, M. Firth, E. C. C. Tsang, and X.-Z. Wang, "Feature selection using localized generalization error for supervised classification problems using RBFNN," *Pattern Recognition*, vol. 41, no. 12, pp. 3706 - 3719, 2008.

[34] W. W. Y. Ng, Z.-M. He, D. S. Yeung, and P. P. K. Chan, "Steganalysis classifier training via minimizing sensitivity for different imaging sources," *Information Sciences*, vol. 281, pp. 211 - 224, 2014.

[35] W. W. Y. Ng, X.-L. Liang, J. Li, D. S. Yeung, and P. P. K. Chan, "LG-Trader: stock trading decision support based on feature selection by weighted localized generalization error model," *Neurocomputing*, vol. 146, no. 1, pp. 104 - 112, 2014.

[36] L. Kocis, and W. J. Whiten, " Computational investigations of low discrepancy sequences," *ACM Transactions on Mathematical Software*, vol. 23, no. 2, pp. 266-294, 1997.

[37] C. Wan, Z. Xu, Y. Wang, Z. Y. Dong, and K. P. Wong, "A hybrid approach for probabilistic forecasting of electricity price," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 463-470, 2014.

[38] J. Park, and I. W. Sandberg, "Universal approximation using radial-basis-function networks," *Neural Computation*, vol. 3, no. 2, pp. 246-257, 2014.

[39] M. Harries, "SPLICE-2 comparative evaluation: Electricity pricing," *School of Computer Science and Engineering, Univ. New South Wales, New South Wales, Australia*, Tech. Rep. 9905, 1999.

[40] S. Wang, L. L. Minku, and X. Yao, "A learning framework for online class imbalance learning," in *Computational Intelligence and Ensemble Learning*, pp. 36-45, 2013.

[41] S. Wang, L. L. Minku, and X. Yao, "A systematic study of online class imbalance learning with concept drift," in *IEEE Transactions on Neural Networks and Learning Systems*, doi: 10.1109/TNNLS.2017.2771290.

Mr. Jianjun Zhang received his bachelor degree from South China University of Technology in 2015. He is currently a Ph.D. candidate of the Department of Computer Science and Engineering, South China University of Technology, China. His current research interests are neural networks and machine learning for imbalanced and non-stationary environments.

Dr. Chun Sing Lai (S'11) received the B.Eng. degree (Hons.) in electrical and electronic engineering from Brunel University London, U.K., in 2013. He is currently an EPSRC Research Fellow with the School of Civil Engineering, University of Leeds and also a Visiting Research Fellow with the School of Automation, Guangdong University of Technology, China. He is currently completing the D.Phil. degree in engineering science with the University of Oxford, U.K. He is Secretary of the IEEE Smart Cities Publications Committee. He organized the IEEE SMC Workshop on Smart Grid and Smart City, SMC 2017 in Canada. His current interests are in data analytics and energy economics for renewable energy and storage systems.

Professor Wing W. Y. Ng (S'02, M'05, SM'15) received his B.Sc. and Ph.D. degrees from Hong Kong Polytechnic University in 2001 and 2006, respectively. He is now a Professor in the School of Computer Science and Engineering, South China University of Technology, China. His major research directions include neural networks, deep learning, smart grid, smart health care, smart manufacturing, and non-stationary information retrieval. Dr. Ng is currently the deputy director of the Guangdong Provincial Key Laboratory of Computational Intelligence and Cyberspace Information and an associate editor of the International Journal of Machine Learning and Cybernetics. He is the principle investigator of three China National Nature Science Foundation projects and a Program for New Century Excellent Talents in University from China Ministry of Education. Dr. Ng served as the Board of Governor of IEEE Systems, Man and Cybernetics Society in 2011 - 2013.

Professor Witold Pedrycz (M'88, SM'90, F'99) is Professor and Canada Research Chair (CRC‒Computational Intelligence) in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. In 2009 Dr. Pedrycz was elected a foreign member of the Polish Academy of Sciences. In 2012 he was elected a Fellow of the Royal Society of Canada. Witold Pedrycz has been a member of numerous program committees of IEEE conferences in the area of Fuzzy Sets and Neurocomputing. In 2007 he received a prestigious Norbert Wiener award from the IEEE Systems, Man, and Cybernetics Society and in 2013 a Killam Prize. He is a recipient of the IEEE Canada Computer Engineering Medal 2008. In 2009 he has received a Cajastur Prize for Soft Computing from the European Centre for Soft Computing for "pioneering and multifaceted contributions to Granular Computing".

His main research directions involve computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data mining, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and Software Engineering. He has published intensively in this area. He is also an author of 16 research monographs covering various aspects of Computational Intelligence and Software Engineering.

Dr. Pedrycz is intensively involved in editorial activities. He is an Editor-in-Chief of Information Sciences (Elsevier), WIREs Knowledge Discovery and Data Mining (Wiley), and Granular Computing (Springer). He currently serves on the Advisory Board of IEEE Transactions on Fuzzy Systems and is a member of a number of editorial boards.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2018.2850930, IEEE Transactions on Industrial Informatics

10

Professor Loi Lei Lai (M'87, SM'92, F'07) received B.Sc., Ph.D., and D.Sc. from University of Aston and City, University of London respectively. Presently he is University Distinguished Professor at the Guangdong University of Technology, China. He was Director of Research and Development Centre, Pao Yue Kong Chair Professor, Vice President, Professor & Chair in Electrical Engineering and Fellow Committee Evaluator for State Grid Energy Research Institute, China; Zhejiang University, China; IEEE Systems, Man and Cybernetics Society (IEEE/SMCS); City, University of London; and IEEE Industrial Electronics Society respectively. He was awarded an IEEE Third Millennium Medal, IEEE Power and Energy Society (IEEE/PES) UKRI Power Chapter Outstanding Engineer Award in 2000, IEEE/PES Energy Development and Power Generation Committee Prize Paper in 2006 & 2009, IEEE/SMCS Outstanding Contribution Award in 2013 & 2014 and the Most Active Technical Committee Award in 2016. He is a Fellow of IET and IEEE & National Distinguished Expert in China (In the Honorary list of the National Distinguished Expert in 2015, less than 1% of experts were selected.) and Distinguished Expert in State Grid Corporation of China.

Professor Xizhao Wang (M'01, SM'06, F'12) received the Ph.D. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1998. He is currently a Professor with the Big Data Institute, Shenzhen University, Shenzhen, China. His current research interests include uncertainty modeling and machine learning for big data. He has edited more than ten special issues and published three monographs, two textbooks, and more than 200 peer-reviewed research papers. By the Google scholar, the total number of citations is over 5000. He is on the list of Elsevier 2015/2016 most cited Chinese authors. Dr. Wang is the Chair of the IEEE SMC Technical Committee on Computational Intelligence, the Editor-in-Chief of Machine Learning and Cybernetics Journal, and Associate Editor for a couple of journals in the related areas. He was a recipient of the IEEE SMCS Outstanding Contribution Award in 2004 and a recipient of the IEEE SMCS Best Associate Editor Award in 2006.