Chapter 18
From Ensemble Learning to Meta-Analytics: A Review on Trends in Business Applications

Mohammad Nazmul Haque and Pablo Moscato

Abstract Ensemble learning has been applied in different areas to improve the predictive performances using multiple learners. The two core building blocks, diversity and combination rule, which play a significant role in ensemble learners. The ensemble approach can be divided into two broad groups based on the variation of base classifiers: homogeneous and heterogeneous ensemble. We conducted a comprehensive review of the ensemble learning used for data analytics. The study has proceeded from the feature selection to classification. We found that the ensemble learning helps to overcome the problem associated with the dimensionality and class imbalance of data. For this reason, the ensemble approach found to be more suitable for the classification of high-dimensional data. Then we move towards the meta-analytics using ensemble learning. Our comprehensive review of the metaheuristics-based ensemble learning for homogeneous and heterogeneous ensemble found a substantial number of applications of ensemble learning from these categories. The in detail study of the ensemble learning in business applications able to identify four successful application areas: purchasing and marketing, predictive analytics, business process management, and customer churn prediction. From these application areas, we observe that majority of the approaches built homogeneous ensembles with dynamic selection for single objective optimization. Despite these success in various application domains, ensemble learning could face challenges in analytics in the future. We concluded the chapter with identifying those difficulties and some trends to overcome them for ensemble learning with meta-analytics.

Keywords Classifier ensembles · Ensemble learning · Meta-analytics · Failure prediction · Multi-objective optimization · Group-ensemble learner · Machine learning · Customer churn prediction

M. N. Haque (✉) · P. Moscato
School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW, Australia
e-mail: Mohammad.Haque@newcastle.edu.au; Pablo.Moscato@newcastle.edu.au

P. Moscato, N. J. de Vries (eds.), *Business and Consumer Analytics: New Ideas*,
https://doi.org/10.1007/978-3-030-06222-4_18

18.1 Introduction

Recent advances in machine learning theory have extended the research areas to increase the capabilities of basic learning methods. This has also been the consequence of a philosophical change. Quoting Leo Breitman [12]:

> The best solution could be an algorithmic model, or maybe a data model, or maybe a combination. But the trick to being a scientist is to be open to using a wide variety of tools.

who proposed the successful *random forest* approach for classification [10], a "cultural change" has been slowly happening over the last two decades. This change is not only supported by performance, and there is an important scientific quest in understanding why the collective several learning agents can perform substantially better than individual ones. It is perhaps a good analogue with memetic algorithms, a metaheuristic method that aims at combining the virtues of different algorithmic approaches to an optimization problem, so that some sort of "emergent phenomena" comes out of the synergy of the techniques employed by the individual agents.

These new learning methods, the ones that are going to occupy our interest in these pages, have been called "meta-learning schemes" or "meta-classifiers" or "ensembles". In machine-learning paradigm, ensemble data mining methods strategically advance the power of committee methods, or combine models to achieve better prediction accuracy than any of the individual models could obtain [67]. The fundamental goal when designing an ensemble is to develop it in such a way that it will provide independent models. The final aim is that the combined model will produce better performance than the individual models.

Here the inducer is the classifier, applicable for a specific domain. This *"No Free Lunch"* phenomenon presents a dilemma for researchers and practitioners, namely how to select the best classifier for a specific domain. The ensemble learning

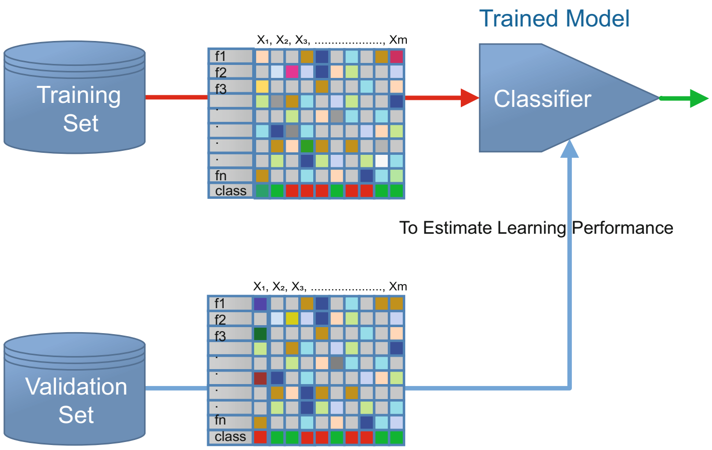


Fig. 18.1 The learning process of a classifier. The classifier learns from the training data. The validation data is used to estimate the learning performance of the trained model. This trained classifier model is then used for classifying samples from the unknown test set

methods (defined in Definition 18.2) can overcome this dilemma by combining the output of many independent classifiers that perform well in certain domains, but are sub-optimal in others. We can then generally define this type of learning approach as follows:

Definition 18.1 (Learning Task) Given a set of pairs $\{(\mathbf{x_i}, \omega_i)\}$, where $\mathbf{x_i} \in S$ is a sample in a given set S (which could be finite or infinite) and ω_i is a target/class label associated with sample $\mathbf{x_i}$, the task is to find a mathematical model that can "predict" the class label for any other sample in S. The learning process is illustrated in Fig. 18.1.

Definition 18.2 (Ensemble Learning) An ensemble of learning machines is a set of adaptive entities that deliver partial solutions to a given problem, and then integrate these solutions in some manner to construct a final or complete solution to the original problem.

Ensemble methods have been used by researchers from various disciplines such as pattern recognition, statistics, and machine learning. Over the last decade, ensemble-based systems have drawn rising attention, gaining popularity, and spreading the application areas into a broad spectrum. Some promising areas to apply ensemble methods are: business, financial, biomedical, remote sensing, genomic, incremental learning, and other types of rapidly growing data analytics problems. A typical ensemble method for classification tasks contains the following building blocks [72]:

- **Training set:** It is a labelled dataset used for training the ensemble.
- **Validation set:** Another independent set of samples and their associated labels which used to estimate how well the classifiers perform after being trained. This set helps you to estimate properties of the models.
- **Test set:** The final dataset of samples and their labels which is used to estimate the performance of the ensemble.
- **Base Classifier:** The base classifier C is an induction algorithm to form a generalized relationship among sample $\mathbf{x}$ and its corresponding label ω. In many cases we have a discrete set of possible class labels (e.g. $\Omega = \{\omega_1, \cdots, \omega_c\}$).
- **Diversity Generator:** We aim at obtaining several independent models of the mapping between the samples and the class labels, so an algorithm is generally needed for generating the diverse models from individual learners after being trained with labelled training data even when the same technique is used for all of them.
- **Combiner:** Another algorithm (also called the *fusion method*) is responsible for integrating the outputs of a set of base classifiers such that the whole set would deliver stronger generalization ability than any individual classifier.

There are very little rules to "guide" how this process should be conducted. In practice, perhaps the simplest approach for ensemble learning is to "weight" the outputs of several individual models, and then to combine them in such a way that the obtained accuracy is maximized. This is a case where "meta-analytic" techniques can be used, guiding the process of combining the information provided by the independent classifiers. The next subsection clarifies the issues related to the design of such an ensemble.

18.1.1 Basic Design Techniques for Ensemble of Classifiers

Though, every ensemble method combines multiple classifiers outcome into a single decision, the building paradigms of the ensemble of classifiers usually differ from each other. They generally have different diversity generation mechanisms and in the strategy for combining them. The process involved in the ensemble learning is illustrated in Fig. 18.2. We will briefly discuss the basic algorithmic blocks.

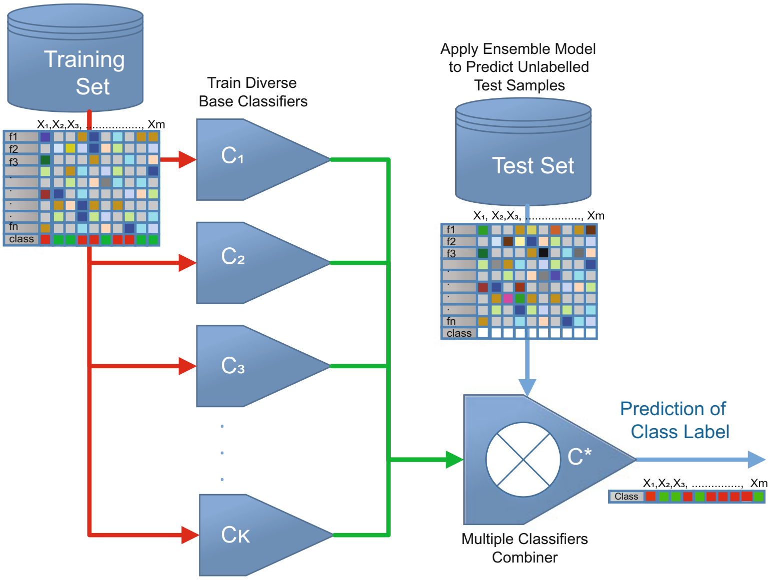


Fig. 18.2 A generic ensemble learning process. Diverse base classifiers are trained on training data. The combination rule (also known as combiner method, fusion approach) combines the learning of multiple base classifiers to create the ensemble. The ensemble is then used for predicting the class label of the samples in the test data

18.1.1.1 Diversity in Ensemble of Classifiers

In a way somewhat analogous to the need of low correlation of features for improving classification accuracy, it is also the case that this is not to say that we do not want them to be highly correlated with the correct output, but it is clearly the case that some variation of outputs is expected and that individual classifiers can predict the wrong output for some samples. The diversity can be achieved in many ways. Based on how the ensemble employs the training set, we can observe three sources of diversity.

- **Different Classifier:** To attain the diversity among base classifiers, this method is used to train all single classifiers using same training set. They use different learning algorithms as base classifiers, or their variations of parameters of the base classifiers. This is known as the *heterogeneous ensemble method*.
- **Different Training Set:** Another diversity generation mechanism can result from using different training sets (obtained from the original training set by re-sampling techniques, such as bagging and AdaBoost) to train base classifiers. These approaches use multiple instances of a base classifier trained on different training samples. This is called *homogeneous ensemble method*.
- **Ensemble Feature Selection:** Diversity can also be generated by training the individual classifiers with datasets that consist of different feature subsets, like the Random Subspace Method (RSM). In the ensemble feature selection, an ensemble method is used to draw feature subspace. Base classifiers are then trained using those sub-samples of features.

Ensemble methods can use any one of the alternatives mentioned above to generate diverse individual base classifiers. To determine the final decision, the process of combining the outcomes comes next in the process.

18.1.1.2 Designing the Combination Rule

The combination rule is the second key component of the ensemble method. It helps to reach into a final decision from all participating single classifier's outcome. The combination rule can be mainly designed in two ways:

- **Classifier Fusion:** In this case, each classifier is trained over the entire feature space. Then, results of individual classifiers are combined in an appropriate manner to reach a final decision. The fusion method can use majority voting, weighted majority voting, summation, product, maximum and minimum, fuzzy integral, the Dempster–Shafer based combiner, or the decision templates to decide the final outcome.
- **Classifier Selection:** This design approach for a combination method uses domain expert classifiers in the training phase. Here, each classifier is trained to become an expert in a specific part of the total feature space. In the decision prediction stage, only one base classifier is used, instead of combining the decision of all available classifiers. The selection of a final predictor classifier, among available expert single classifiers, can be done in two manners. They are the *Static Classifier Selection (SCS)* method and the *Dynamic Classifier Selection (DCS)* method.

– *Static Classifier Selection (SCS):* In the SCS method, the region of competence for a single classifier is defined during the training period. Therefore, one classifier is chosen for prediction task.
– *Dynamic Classifier Selection (DCS):* On the other hand, the regions of competence of participating single classifiers are defined during the prediction phase, based on the characteristics of the testing dataset.

18.2 Review on the Ensemble Learning for Data Analytics

In this section, we focus on the selection of important features (i.e. feature selection) and in the classification methods aimed at achieving better accuracy using ensemble methods. The canonical concept of ensemble approaches is to improve the prediction accuracy by using a set of individual classifiers. The underlying assumption is that distinct base classifiers in an ensemble "bring together" different patterns observed in the data. It is then the combination of these patterns that helps to better predict the class label. We review the literature that deals with the enhancement of classification accuracy by the adoption of an ensemble-based approach. We will also review the relevant literature where we have observed that metaheuristics have been employed to create or guide the ensemble. This review helps to explain the effectiveness of the ensemble method data analytics and in particular for feature selection and effective classification.

Researchers have published several reviews and empirical analysis on the ensemble methods since date. Most of them discussed and tried to explain why the ensemble of classifiers (EoC) outperforms the results of the best of its single classifier component. For instance, Jordan and Jacobs [37] proposed a statistical hierarchical tree-structured mixture model of classifiers utilizing ideas from mixture model estimation and generalized linear model theory. They presented an Expectation Maximization (EM) algorithm for adjusting the parameters of the architecture using an iterative approach to maximize the supervised learning performance. They have not emphasized important issues (such as convergence and consistency of the final result) in the paper. Dietterich [18] evaluated ensemble methods named Bayesian averaging, bagging, and boosting. The article had addressed three principal reasons that explain why the ensemble methods outperform any single classifier. They identified that when the amount of available training data is too small compared to testing data, then classifiers can lead to an "under-fitting" problem. On the other hand, classifiers that only use a local search based heuristic may, quite obviously of course, get stuck in local optima. They provide experimental evidence for choosing AdaBoost algorithm for ensemble creation. However, the study did not examine the interaction between AdaBoost and the properties of the underlying learning algorithm. Valentini and Masulli [82] presented a brief overview of ensemble methods. Moreover, they proposed a taxonomy based on the combination rule for the base classifiers in ensembles. After that, Polikar [69] examined the context in which the accuracy of an ensemble method is more useful than their individual classifier algorithms. They also analysed the various components of an ensemble system as well as various procedures by which individual predictions could be combined. Recently Krawczyk et al. [45] proposed an ensemble of fuzzy classifiers and used a GA for selecting samples for the under-sampling approach. They applied a metaheuristic-based approach to select samples for balancing medical image dataset, and they have improved the classification performances.

Roli et al. [73] proposed two design methods based on the so-called *"overproduce" and "choose"* paradigm with metaheuristics. In this scheme, *overproduce* is the generation of a large set of candidate ensemble classifiers. Then *choose* is the extraction of the best performing sub-ensemble. They have used three heuristic search algorithm named forward search, backward search, and tabu search as their chosen methods. Even though these design techniques demonstrated some compelling features, they do not claim any best choice method among three and do not show how to design the optimal ensemble. Similarly, Kotsiantis et al. [43] described various classification algorithms and the recent effort for improving classification accuracy by using ensembles of classifiers. They have discussed some algorithms based on Artificial Intelligence (Logic-based techniques, Perceptron-based techniques) and Statistics (Bayesian Networks, Instance-based techniques). The book by Kuncheva [48] was entirely devoted to the field of the model combination. The book covers the topics of multiple binary classifier systems and their base classifier combination methods. Meanwhile Tulyakov et al. [81] proposed an ensemble of classifiers method to increase the performance of pattern recognition applications. They introduced a retraining operation which trained the optimal ensemble classifier with another set of training data and adjusts associated weight. They have also used local neighbourhood search like k-NN to identify similar sub-samples of the test set to be used for the retraining of the classifier combinations. Such effects showed a significant effect on the performance of combinations. Oza and Tumer [68] reviewed ensemble methods for diverse classes of statistical classification algorithms and their various real-world application domains. They surveyed applications of ensemble methods to problems that have historically been most representative of the difficulties in classification. In particular, the survey covers the applications of ensemble methods to remote sensing, person recognition, one-vs-all recognition, and medicine. Recently, Galar et al. [26] developed an empirical analysis of different aggregations to combine outputs of binarization strategies for the dataset. They formed a dual study: firstly, they employed different base classifiers to monitor the suitability and capability of each combination within each classifier. Then, compared the performance of these ensemble techniques with the classifiers' themselves. The proposed research tested several well-known algorithms such as Support Vector Machines, Decision Trees, Instance Based Learning, or Rule-based Systems. Experimental evidence supported that the goodness of the binarization techniques about the base classifiers were most robust methods within this framework.

Collectively, these research papers showed that ensemble methods are a promising approach for use on incremental learning, data fusion, feature selection, learning with missing features, and multi-class classification domains. They also provided answers to the question: *"Why should we choose ensemble methods instead of* *improving individual classification performance?"* They have justified that ensemble of classifiers is an efficient method for improving classification accuracy.

18.2.1 Ensemble Learning in Feature Selection

Feature selection method is an essential step in classification. It plays a key role in classification performance enhancement. Literature suggests two main categories of feature selection procedure named *ranking* and *subset selection* method. Ranking feature selection methods applied each feature's worth to order them and select top k number of features. On the other hand, feature subset selection process uses the searching method to find a better subset of features that explains original feature most.

18.2.1.1 Population-Based Approaches

Many approaches are based on Genetic Algorithms (GAs), a population-based heuristics, to manage search space created from the ensemble of feature selection method. Kuncheva and Jain [46] proposed two basic ways to utilize the power of GAs to create a multiple-classifier system. Their method started with a GA version that selects disjoint feature subsets and the second variant selects overlapping feature subsets. They created the ensemble with three-classifier systems and basic types of individual classifiers. Oliveira et al. [65] applied GA to combine the predictions of different feature selection methods. They proposed an ensemble feature selection method based on a hierarchical multiobjective GA. In their first level, a set of robust classifiers applied for feature selection. Next level of GA combines the features to build a powerful ensemble. The proposed method is applied for handwritten digit recognition, using three different feature sets and neural networks (MLP) as classifiers. Experimental result showed the power of the proposed strategy. Later on, Minaei-Bidgoli et al. [59] applied GAs on feature selection to improve the prediction performance. First, they used feature selection method for decreasing computational cost and increasing classifier efficiency. The feature selection optimized the prediction accuracy of ensemble using GA. Their approach was more efficient than feature subset selection method, and also adaptable to analyse different attributes. Subsequently, Oliveira et al. [64] also dealt with feature selection by using GAs for a classification ensemble building in the similar manner of Oliveira et al. [65]. They experimented on handwritten digit recognition problem and compared with traditional Bagging and Boosting. The proposed method produced higher prediction accuracy. Recently, Ebrahimpour and Eftekhari [21] proposed a fuzzy-based CFS score with an ensemble of feature ranking algorithms. The experimental result revealed that their approach is more suitable than metaheuristics for extremely large search space.

18.2.1.2 Reduction of Dimensionality

On the other hand, many researchers used different types of feature selection methods to create an ensemble to reduce the dimensionality problem in the dataset. Cunningham and Carney [15] argue that feature subset-selection has emerged as a useful technique for creating diversity in classification ensembles. They proposed an entropy measure of the outputs of the ensemble members as a useful measure of the ensemble diversity, and they evaluated it on a medical prediction problem. Experimental results showed enhanced prediction performance on the ensemble and the entropy measure of diversity. Additionally, it expressed a relationship with the change in diversity and breadth of the ensemble. Similarly, Blanco et al. [6] proposed an ensemble of wrapper-based method for gene selection and classification in gene expression datasets. They employed the Naïve-Bayes classification algorithm in a wrapper form. Their design reduces the number of genes that have been selected with a similar accuracy than other conventional approaches. Nikulin et al. [62] created an ensemble that is capable of greater prediction accuracy than any of their individual members by means of utilizing selected features. They found a large number of relatively small and balanced subsets where representatives from the larger pattern are to be selected randomly. They have tested the ensemble strategy against datasets of the PAKDD-2007 data-mining competition. Their ensemble method produced higher accuracy than single classifiers. Recently Seijo-Pardo et al. [75] experimented on different ensemble configuration for feature selection. The ensemble combined the features selected from seven feature ranking algorithms to increase the stability of the selection process. Their experimental results suggested the usage of aggregated feature subset of the ensemble with Random Forest (RF) classifier for data classification.

18.2.1.3 Class Imbalance

Several other works in the literature have reported the issue of class imbalances in the dataset and have proposed alternative methods to deal with the problem. Duangsoithong and Windeatt [20] presented a bootstrap feature selection for ensemble classifiers. They selected optimal features from the full dataset before bootstrapping on selected data. Then they have applied ensemble classifier to evaluate the performance on UCI machine learning repository and a causal discovery datasets. The results showed that a bootstrap feature selection algorithm provides slightly better accuracy than traditional feature selection for ensemble classifiers. Akbaş et al. [2] has implemented different ensemble methods for feature selection for classification of a colon cancer dataset. They have evaluated the performance improvement of individuals with ensemble classification methods. Their results showed that the classification accuracy for the colon cancer can be increased by reducing the features using the ensemble methods. Recently, Yang et al. [93] proposed an ensemble-based wrapper method (feature selection) which is suitable for highly imbalanced class distribution. First, they have eliminated the imbalanced nature of the dataset by sampling and creating multiple stable subsets of actual data. Then, they have evaluated feature subsets using an ensemble of base classifiers each trained on a balanced dataset. Their test results indicate that ensemble-based feature selection by wrapper method outperformed original wrapper algorithms for imbalanced class data.

According to result from these papers, we can conclude that feature selection method can be a potential approach for resolving problems in the dataset. Imbalanced class data and dimensionality are two verdicts for classification algorithms, which can be eliminated using feature selection methods. Ensemble methods can lead to better selection of features from a large number of features and can circumvent problems related to imbalanced class distributions.

18.2.2 Ensemble Learning in Classification

A popular method for creating a perfect classifier from a set of training data is to build several classifiers, and then to combine their predictions for achieving better accuracy. Researchers have employed different types of ensembles of classifier methods. We begin this section of literature survey with embedded feature selection in classifiers then popular ensemble classifiers including tree-based structures.

Tsymbal et al. [80] created an ensemble consisting of multiple classifiers constructed by randomly selecting feature subsets. They conducted experiments on a set of 21 real-world and simulated datasets. In many cases, the ensembles have higher accuracy than the single classifier model. Namsrai et al. [61] introduce an ensemble classification method for classifying cardiac arrhythmia disease. They have applied feature subset selection process to obtain a number of feature subsets from the original dataset. Then they have built classification models using each feature subset and combined using a voting approach. They have considered both classification error rate and feature selection rate (which means the frequency of a specific feature in feature subsets) to calculate the score of the each classifier in the ensemble. The ensemble method improves the classification accuracy significantly. In contrast, Srimani and Koti [76] conducted a classification analysis on five medical datasets, and their results show that individual classifiers perform better than some cases than an ensemble. They have concluded with comments that only selected classifiers needed to be used for each dataset and to achieve optimal accuracy with the medical dataset; researchers need to choose the classifier accurately. They had selected default options for classifiers and did not use any optimization techniques to find the optimal solution for ensemble classifiers. This was missing in their conducted research.

There are some researches on creating the ensemble of classifiers only using tree based structures. In this regard, Zhang et al. [99] proposed an ensemble classification system based on diversified multiple trees which addressed the uncertainty of microarray data quality problem. The proposed diversified multiple decision tree algorithms (DMDT) can determine most informative features from abundant features. Then, they use this unique diversity value as an ensemble combination function. The test results showed that the DMDT is more accurate on average than other well-known ensemble of classifiers on microarray datasets. Likewise, Hu et al. [35] proposed maximally diversified multiple decision tree algorithm (MDMT) based on an ensemble method for robust microarray classification. They are concerned with the noise susceptibility of different decision tree algorithms and MDMT for microarray dataset classification. The experimental results showed that ensemble decision tree methods tolerate the noise values better than a single tree does. They recommended decision tree based ensemble classifiers to deal with noisy microarray datasets. Similarly, Osareh and Shadgar [66] proposed an ensemble method combining Rotation Forest and AdaBoost techniques which in turn preserve the accuracy and diversity in microarray datasets. They have applied five different feature selection algorithms to determine a concise subset of informative genes. Then they have applied decision tree based ensemble classifier using selected features as training set. The experiment showed that the proposed ensemble classifiers outperform not only the conventional machine learning classifiers but also the classifiers generated by two widely used ensemble learning methods, that is, bagging and boosting. Previously, Galar et al. [27] tackled the classifier learning problem with datasets that suffer from imbalanced class distributions and provide empirical comparisons based on ensemble taxonomy. They reviewed the state of the art on ensemble techniques in the framework of imbalanced binary datasets. Their results show empirically that ensemble-based algorithms are advantageous since they outperform the mere use of pre-processing techniques before learning the classifiers. Koutanaei et al. [44] proposed hybrid ensembles for feature selection and classification for credit scoring datasets. They evaluated three feature selection method using the performances of SVM classifiers. Afterwards they used ensemble of classifiers using neural networks for final classifier creation and revealed better performances with the feature subset selected by PCA approach.

These works of literature advocate that ensemble of classifiers are more effective methods than other combinations of classifiers like boosting and bagging. It is also possible to bring inherent power from embedded feature selection methods for imbalanced data classification, noise susceptibility, and multi-class classifier. In a word, the ensemble of classifiers methods is suitable enough to deal with high-dimensional biological datasets.

18.3 Ensemble Learning with Metaheuristics: Moving Towards Meta-Analytics

A current subject of intense research in data classification is the combination of several classifier systems. They can be built following either the same or different models and/or datasets building approaches. A large number of available single classifiers can create an enormous search space for ensemble combinations. To get the better combination for creating an ensemble from this NP-hard problem [34], researchers applied different types of metaheuristics strategies. In this extent, we found that the population-based metaheuristics have been applied in mainly three levels of the ensemble of classifiers (EOC) method. The application level of metaheuristics in the EOC can be categorized as: *decision combination level, feature selection level*, and *classifier formation* or *creation level*.

Firstly, we will discuss the literature where metaheuristic has been employed in the base classifier creation, next to the feature selection and finally as the combination methods.

Zhang and Bhattacharyya [97] demonstrated the potential of a genetic programming (GP) metaheuristic as a base classifier algorithm in building ensembles in the context of large-scale data classification. An ensemble built upon base classifiers trained with GP significantly outperformed its counterparts built upon individual base classifiers. Wang and Wang [84] proposed an ensemble of classifiers where each individual classifier is trained on a particular weighting over the training examples. They incorporated a genetic algorithm (GA) to search a substantial weighting space. They tested the algorithm on the UCI benchmark datasets and discovered the ensemble method as robust and consistent for face detection application. After a while, Ranawana and Palade [71] presented a current overview of Multi-Classifier Systems (MCSs) and provided an outline roadmap for MCS. They also presented a case-study of the MCS theoretical issues, and simple guidelines for the selection of different paradigms. Moreover, they introduced a novel optimization of the traditional majority voting combination method which uses a genetic algorithm. A couple of years later, Kim and Oh [40] proposed a hybrid genetic algorithm (a type of memetic algorithm) for classifier ensemble selection. They used two local search operations to improve offspring prior to replacement. They parameterized the local search operations to control the computation time and found it as an effective approach. Later on, Espejo et al. [31] survey existing literature about the importance of genetic programming for classification. They have shown the different ways in which evolutionary algorithm can help in the creation of accurate and reliable classifiers. Haque et al. [32] proposed to use GAs to find the best combination of ensembles from a large pool of base classifiers. The ensemble learning with the metaheuristic (population-based search) approach was able to find better ensembles than the state-of-the-art ensemble learner for image and biological data classification.

As explained in Sect. 18.1.1.1, we can classify the ensemble technique in three categories based on the diversity generation mechanism. Moreover, based on the variations of base classifiers used, the ensemble of classifiers can be categorized into *homogeneous* (as shown in Fig. 18.3) and *heterogeneous* (shown in Fig. 18.4) ensemble method. The feature selection ensemble can be embedded into both of them as the diversity generation methods. Now we will review works that fall into those two categories and that use population-based metaheuristics search algorithms.

18.3.1 Homogeneous Ensemble Learning

Kim et al. [39] proposed a meta-evolutionary (ME) based homogeneous ensemble of classifier method. It uses two-level evolutionary search through the feature selection and ensemble creation. The author applied an ME for feature selection using an ensemble method which works like Boosting algorithm. The fitness of the ensemble is updated with the cost based on its predictive accuracy, as determined by majority voting with equal weight among base classifiers. They used a variable number of ANNs (Artificial Neural Network) to form the ensemble and used majority voting as their combination function. Experimental results on 15 datasets suggested that the weighted ensemble is more effective than single classifiers and classic ensemble methods. They have tested their method on smaller datasets (maximum dimensions in the dataset were 3772 features, 27 samples), it needs to verify the suitability of their approach on large datasets.

Fig. 18.3 The structure of homogeneous ensemble

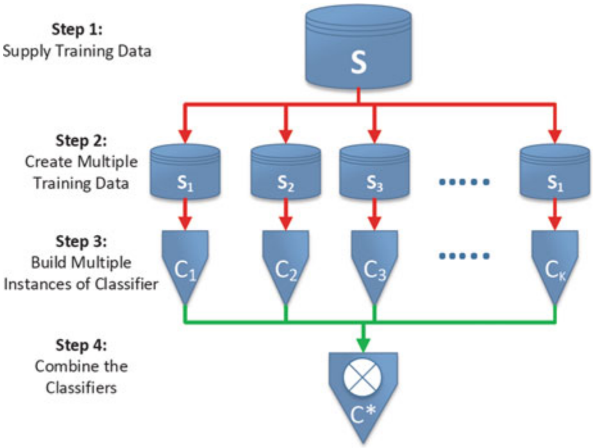
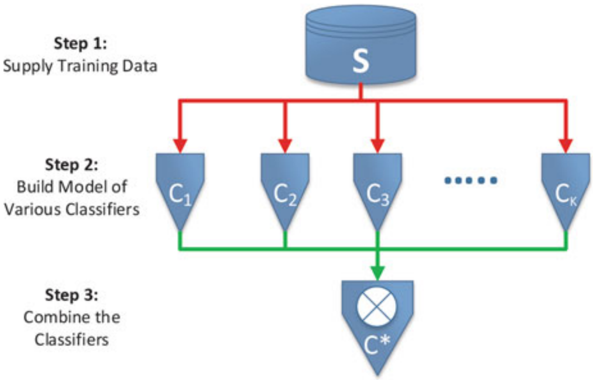


Fig. 18.4 The structure of heterogeneous ensemble



Cleofas et al. [14] proposed a GA-based ensemble to deal with the imbalance class problem in their datasets. They have applied the GA on feature selection level for balancing the class distribution by under-sampling the majority class samples. They used a m-dimensional chromosome, which represents all features in the training set of the m samples. Then they applied features selected by GA to the ensemble of classifiers. This method is impractical to apply on high-dimensional datasets due to encoding all features into the chromosome.

Gaber and Bader-El-Den [24] proposed a genetic algorithm based random forest (GARF) ensemble method. They used heuristic chromosome (HC) where an individual represents an Ensemble (Random Forest), and each gene represents a Random Tree. They stored pre-computed classification results for all random trees. Then, they evolve new ensemble using pre-generated trees (EV-Ensemble) to evaluate the fitness of new random forest using the HC. They have also applied an Indirect-GA to create random forest classifier. Their experiment showed that the performance of random forest could be boosted when using the genetic algorithm.

More recently, Oh and Gray [63] proposed a GA Robust Ensemble (GA-RE), which used the GA in the classifier formation level. It used variable sized individuals in the range of 5–20 decision stumps trees. But the stump tree they used, it contained only two terminal nodes which possibly loss informative features. Instead of using decision stumps, they could try using decision trees. Zhang et al. [96] proposed a homogeneous ensemble of classifier selected by a GA approach for image data classification. They select the homogeneous base classifiers created by varying the parameters using GA to experiment with SVM and NN based homogeneous ensemble of classifiers approach.

18.3.2 Heterogeneous Ensemble Learning

Heterogeneous ensemble method uses different types of classifiers to create an ensemble combination. Some researcher uses single instance from each of the different classifiers and other uses multiple instances of some classifier algorithms. Genetic algorithms have been applied on feature selection, classifier selection, and combination level of the heterogeneous ensemble.

In [25], the authors have proposed an interesting individual structure of their GA. They have represented each individual using a 3-dimensional incidence matrix. In the incidence matrix, one dimension represents features, another dimension represents classifiers, and the other dimension represents combination methods. They also have proposed associative genetic operators to work with the individual structure. Unfortunately, the whole potential of this idea is still to be tested as the authors have applied their method only on tiny dataset with six (6) features. The approach may not be entirely suitable for the high-dimensional dataset, because it encodes all feature in each individual.

Xu and He [92] proposed a genetic algorithm based ensemble classifier. The GA has been applied on both feature selection and decision combiner level simultaneously. They have applied it to a real-world multi-sensor dataset. They have randomly selected features for each base classifier and used very few numbers of features to build the classifier. There is scope for enhancing this system by adopting an appropriate feature selection method. However, the method has outperformed feature level and decision level fusion methods, it needs to be tested on larger datasets.

After that, Thammasiri and Meesad [79] proposed a GA-based classifier ensemble method to select the appropriate classifiers. They use majority vote in order to increase ensemble classification accuracy. Their evaluation showed that the proposed ensemble classification models selected by the GA yield the highest performance and are efficient in building the ensemble. The maximum feature count for the applied datasets was 30.

Later on, Lertampaiporn et al. [50] applied heterogeneous ensemble methods using GAs for feature selection. They have used Support Vector Machines (SVMs), k-Nearest Neighbour (k-NN), and Random Forests (RFs) to create the ensemble and they applied on large biological datasets. They have applied SMOTE-bagging methods to balance the dataset and used correlation-based feature (CFS) selection method with GA search method. Their experiment shows better performance than single classifiers. More recently, Haque et al. [33] proposed another population-based search metaheuristic, named Differential Evolution (DE) algorithm, for optimizing the weights of heterogeneous base classifiers for the weighted voting in ensemble learning. This approach has performed better in predicting heart disease in compared with other ensemble learning methods.

18.3.3 A Partial Summary Based on Three Streams

We can provide a partial summary of the studies in this subsection based on the application of the metaheuristics in different algorithmic components of the ensemble learning framework. We have observed that some methods applied the metaheuristics to find the combination rule that could enhance the generalization performance (e.g. [79]). There are some studies, which used it for feature selection level. They try to optimize the feature subset to gain better classification performances for base classifiers using various feature subsets (e.g. [14, 39, 50, 79]). The third category of ensemble learning uses metaheuristics for creating the classifier (e.g. [24, 39, 63]). The authors used several tree-based classifiers to create the ensemble. The GA has been applied here to form an optimal classifier using the trained tree-based models. The application of GA in the classifier formation level is common for homogeneous ensemble creation. There exists some study which applied the genetic algorithm to find best base classifiers for the classifier selection models (e.g. [42, 74]). They used only one base classifier selected by the associated metaheuristics. Searching the ensemble combination space for a large pool of base classifiers using metaheuristic was proposed in [32, 33], where the combinatorial space of the base classifiers is explored.

18.4 Ensemble Learning in Business Analytics

The successful invasion of ensemble learning in life science, engineering, management has inspired the business domain as well. We could broadly divide the applied areas of ensemble learning for business and consumer analytics into *marketing, finance, business process management, credit scoring, consumer behaviour, customer relation management*, and *churn prediction*. The key characteristics of ensemble learning used in business analytics for the last decade are presented in Table 18.1. We represent the information in chronological order so that it is easy to see the evolution of the techniques in a number of fronts. The following subsections organize the discussion by application area and summarize the types of business applications in a word cloud[1] of Fig. 18.5.

18.4.1 Purchasing and Marketing

The process of market analytics uses the purchase history information or other behavioural characteristics to find the most valuable customers from their responses to marketing campaign questions or data from market research. Meta-analytics along with metaheuristics has proven its success in this arena for quite a long time. Blaszczynski et al. [7] applied ensemble of weak learners to predict customer behaviour for internet users. Govindarajan [30] compared heterogeneous ensemble learning (created with via the Bagging method [11]) with Radial Basis Functions

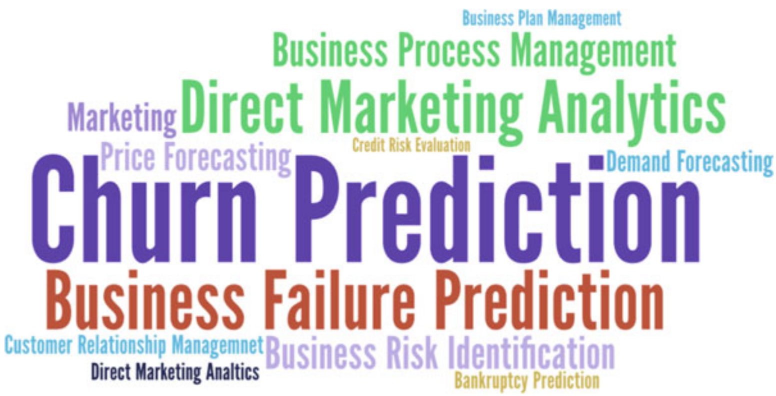


Fig. 18.5 Word cloud view of the types of business analytics used ensemble learning in the last 10 years (from 2007 to 2017). The most frequent application has the largest size, and the least frequent application has the smallest size in the word cloud

[1] We created the Word Clouds using the Pro Word Cloud add-ins for Microsoft Word 2016 found at: https://store.office.com/en-us/app.aspx?assetid=WA104038830.

Table 18.1 A survey of ensemble learning methodologies and applications in business for the last 10 years (from 2007 to 2017 sorted by year)

Application area and paper	Ensemble technique	Key characteristics
Customer Relationship Management [49]	Homogeneous ensemble learning	Used SVM ensembles to effectively manage customer relationship from a risk avoidance perspective by identifying high risk customers
Business risk identification [58]	Neural network and rule-based ensembles	Applied the ensemble classifiers to predict the potential defaults for a set of personal loan accounts
Business risk identification [94]	Metaheuristics-based ensemble learner	Applied classifiers to generate knowledge and aggregated into an ensemble output using an evolutionary programming (EP) technique
Price forecasting [95]	Homogeneous ensemble	Proposed an empirical mode decomposition (EMD)-based neural network ensemble learning model for modelling and forecasting the spot price of world crude oil
Direct marketing analytics [38]	Homogeneous ensemble learning	Compared two fundamental ways of developing ensemble models—sampling and feature selection using homogeneous types of ensemble
Direct marketing analytics [98]	Group-ensemble learner	Proposed a 3-level ranking model (Group-Ensemble) to solve crucial data mining problems like data imbalance, missing value, and time-variant distribution for marketing data analytics
Churn prediction [8]	Ensemble learning	Proposed an ensemble classification models using probability estimation trees (PETs) with weighted voting fusion based on lift measure
Churn prediction [9]	Rotation-based ensemble classifiers	Rotation Forests, feature extraction is applied to feature subsets in order to rotate the input data for training base classifiers, while RotBoost combines Rotation Forest with AdaBoost
Demand forecasting [56]	Homogeneous ensemble of regression	Proposed a balanced-sampling-based ensemble of support vector regression (SVR) forecasting method to improve the predictive accuracy in demand forecasting for supply chain management
Business failure prediction [53]	Multiple models combination ensemble	Formulated an ensemble learner by incorporating the feature selection methods in the representation level, a hybrid of principal component analysis (PCA) at the modelling level, and weighted majority voting at the fusion level
Business failure prediction [54]	Multiple models combination ensemble	Formulated an ensemble learner using case-based reasoning (CBR) predictor models with feature-bagging as diversity generation approach. The CBR ensemble for multiple case representations used majority voting for producing final prediction
Churn prediction [87]	Cost-sensitive ensemble learning	Proposed an ensemble learning with dynamic classifier selection of cost-sensitive learning from imbalanced data

(continued)

Table 18.1 (continued)

Application area and paper	Ensemble technique	Key characteristics
Marketing [87]	Homogeneous ensemble learning	Comparison of performances in customer choice modelling is done with three ensemble learning approach: "Bagging", "Boosting", and "MultiBoosting" evaluated with bias–variance decomposition
Business process management [23]	Ensemble-based clustering method	Proposed a clustering method for discovering performance-oriented process models where separate prediction models were built on different subsamples of the data
Business plan management [19]	Metaheuristics-based ensemble learner	The genetic programming-based ensemble of classifiers is compared against an intelligent model of business plans' appraisal system. Subject matter experts were satisfied with the Reliability and the accuracy of the selected plan
Business failure prediction [55]	Two-stage ensemble	Four feature spaces are created to build different models using multivariate discriminant analysis (MDA) and logit. Afterwards, two levels of ensembles are implemented: one with each of the same type of models and another with two ensembles. Final decision made using majority voting fusion approach
Churn prediction [88]	Ensemble learning with dynamic classifier selection	Train classifiers on different sub-sample of the data using resampling technique to balance the class distribution. Then, dynamically select a proper classifier ensemble for each test sample
Churn prediction [1]	Homogeneous ensemble learning	Compared the performance of four popular ensemble methods: "Bagging", "Boosting", "Stacking", and "Voting". Four known base classifiers are used: C4.5 Decision Tree (DT), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Reduced Incremental Pruning to Produce Error Reduction (RIPPER)
Direct marketing analytics [3]	Ensemble of classifiers	Used an ensemble classification method which removes imbalance in the data, using a combination of clustering and under-sampling
Direct marketing analytics [30]	Bagging classifier	Designed both homogeneous and heterogeneous ensembles using RBF and SVM as base classifiers. Bagging of RBF exhibited the best performance on direct marketing data
Churn prediction [89]	Dynamic classifier selection ensembles	Proposed a feature-selection-based dynamic transfer ensemble (FSDTE) model to introduce transfer learning theory for utilizing the customer data. Iterative feature generation was used for training the base classifier. Finally, base classifier was selected dynamically

(continued)

Table 18.1 (continued)

Application area and paper	Ensemble technique	Key characteristics
Churn prediction [5]	Ensemble selection	Proposed a decision-centric framework to create churn models using heterogeneous base classifiers. Then, pairwise combinations of base classifiers performances were assessed to select the best classifier for decision-making
Business process management [100]	Clustering ensemble	Proposed an entropy-based clustering ensemble method for mining different types of preference patterns with a priority-based schedule algorithm for dynamic resource allocation in multi-instance process contexts
Churn prediction [90]	Multiple classifiers ensemble	First base classifier models were generated by using different subset of the data. Then, final decisions were combined based on weighted voting and supplied threshold
Churn prediction [4]	Heterogeneous ensemble selection	Proposed a heterogeneous ensemble method where the selection of best base classifier subset was done using soft set based method to combine the decisions for final result
Business process management [16]	Multi-tier ensemble learning	Proposed a multi-view ensemble learning approach with a clustering-based trace abstraction method and a context- and probability-aware stacking method as decision fusion
Business failure prediction [83]	Heuristic-based ensemble selection	Proposed a two-stage selective ensemble model with manifold learning algorithm, feature selection and ensemble selection were used to predict business failure
Bankruptcy prediction [22]	Metaheuristic-based ensemble learning	Proposed a two-stage ensemble learning (with SVMs) for feature selection and classification using several metaheuristics (including artificial bee colony (ABC), genetic algorithm (GA), and sequential forward selection (SFS)) to predict bankruptcy
Credit risk evaluation [17]	Hybrid Homogeneous ensemble learning	Proposed a feature selection-based hybrid-bagging algorithm (FS-HB) for improved credit risk evaluation

(RBF) and Support Vector Machine (SVM) classifiers. They found that the RBF bagging approach performed better on analysis of direct marketing data. Kim [38] compared two fundamental ways of developing ensemble models (using sampling and feature selection) for homogeneous types of ensemble. Different types of homogeneous ensemble learners are created with Artificial Neural Network (ANN), Naive Bayes (NB), and SVM as base classifiers and applied on vehicle insurance policy direct marketing data. Their experimental outcome suggested the ensembles with Naive Bayes (NB) and SVM classifiers marginally improved performance. Similar approach has been used in [85] for profiling household and customer targeting from existing customer database. Zhang et al. [98] proposed a 3-level ranking model for ensemble learner with Bagging, RankBoost, and Expending Regression Tree to solve crucial data mining problems (data imbalance, missing value, and time-variant distribution) for design marketing strategies. Amini et al. [3] used ensemble learner to increase the prediction accuracy and improve the response rate for the bank's direct marketing.

18.4.2 Predictive Analytics

Forecasting plays a major role in making the business decision. The ability to predict the demand or forecast the price, identify the risks, and predict the failure at an early stage helps businesses to be less prone to failure. A homogeneous ensemble with empirical mode decomposition (EMD)-based neural network was proposed by Yu et al. [95] to forecast the spot price of world crude oil. Liu et al. [56] proposed a homogeneous ensemble learning method using SVMs as base regression method to forecast the customer demand in supply chain management. These forecasting will help the business to act in advance to gain more. Risk identification in business contributes towards a successful business plan and to reduce future loss. Matthews and Scheurmann [58] proposed an artificial neural network method and rule-based ensembles to predict the potential defaults for a set of personal loan accounts. Yu et al. [94] applied ensemble classifiers to generate knowledge and aggregated into an ensemble output using an evolutionary programming (EP) technique to identify corporate financial risk. The identification of risk is a crucial step which influences the process of business plan management. In this arena, Li et al. [55] proposed two-stage ensemble using four feature selection and two classifiers (multivariate discriminant analysis (MDA) and logit). The final decision was made using majority voting fusion approach for the business failure prediction. Wang and Wu [83] also used similar two-stage ensembles for classifier model and feature selection, but with heuristics, to predict the business failure.

18.4.3 Business Process Management

Business process management (BPM) is a systematic approach to redesign the workflow of an organization to make it more efficient, effective, and adaptive. The BPM requires analysis of the existing business to identify candidate workflow for improvement and is a good area for analytics. Ensemble learning techniques are showing a very strong footprint in this regard. Folino et al. [23] proposed an ensemble learning-based clustering method for discovering performance-oriented process models to help the BPM. Zhao et al. [100] also proposed a clustering ensemble, but using entropy measure for mining different types of preference patterns and dynamic resource allocation in multi-instance process contexts. Subsequently, Cuzzocrea et al. [16] proposed multi-tier ensemble learning approach with a clustering-based trace abstraction method and a context- and probability-aware stacking method as decision fusion for identifying the deviations between business processes.

18.4.4 Customer Churn Prediction

Customer churn prediction plays a significant role in deciding the company's Customer Relationship Management (CRM) strategy. Among all the business applications using ensemble learning, churn prediction standouts for many applications in compared with others. In [8], they proposed an ensemble classification models using probability estimation trees (PETs) with a weighted voting fusion based on lift measure for customer churn prediction. Bock and den Poel [9] proposed rotation-based ensemble classifiers named RotBoost which combines Rotation Forest with AdaBoost in customer churn prediction. Xiao et al. [87] introduced a cost-sensitive ensemble learning with dynamic classifier selection from imbalanced data. They applied the method in popular Germany credit scoring and a branch of Sichuan Telecommunication data. The experimental result analysis suggested the usage of both the accuracy and diversity of the ensemble simultaneously in the process of selection to improve the classification performances. Xiao et al. [88] proposed ensemble learning with dynamic classifier selection and resampling technique to balance the class distribution in churn data. Abbasimehr et al. [1] used four different homogeneous ensemble learning formulated with C4.5 Decision Tree (DT), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Reduced Incremental Pruning to Produce Error Reduction (RIPPER) as a base classifier in each setting. In their experiments, boosting ensemble learning approach formulated with any of the four base classifiers performed better than other approaches in churn prediction. Xiao et al. [89] proposed dynamic classifier selection Ensembles and iterative feature generation. Baumann et al. [5] also proposed ensemble selection, with a different approach, by considering pairwise combinations of base classifiers performances for churn prediction. Xiao et al. [90] proposed Multiple Classifiers Ensemble based on weighted voting and supplied threshold. Awang et al. [4] proposed heterogeneous ensemble selection where the selection of best base classifier subset was performed using soft set based method.

From those applications of ensemble learning in business, we identify some important characteristics of ensembles. Some used multi-tier ensemble for simultaneous selection of feature selection method and the base classifiers. A substantial number of ensembles were built on dynamic selection approach. The majority of the ensembles proposed in these areas were in the category of homogeneous ensemble learning. While this volume presents a multi-objective approach, most ensemble learning techniques so far are based on a single objective. It is clear that we are moving towards a meta-analytic frontier with the combination of more powerful heuristics, metaheuristics, and the increasing availability of software codes that provide access to a large number of base classifiers. The phenomenon is also seen

Fig. 18.6 Word cloud view of the types of approaches used in business analytics for the last decade. The size of the word depends on the frequency of the approach in the literature, i.e. most frequent to least frequent method size varies from the largest to smallest. Note the conspicuous appearance of the word "Metaheuristics" in this set

from the word cloud of the types of methods used in business analytics in Fig. 18.6. Here, the metaheuristic appeared nearly equal to the Ensemble learner.

18.5 Future Challenges and Conclusion

Despite the success of some existing ensemble learning, take these as the pioneers of the more advanced meta-analytics methods to come in the future, they show not only the opportunities were given but some of the challenges ahead. Some of them stem from the inherent characteristics of the datasets. The recent advent of big data and stream datasets brings particular challenges to ensemble learning techniques. For instance, they add an extra layer of problem-domain difficulties, e.g. redundancy, noise, heterogeneity of data sources, lack of annotation, and imbalanced data [101]. For eliminating the redundancy in data, traditional pairwise approaches such as the use of Euclidean distance-based algorithms are not fast enough for big data and for data streams. Data gathered from various sources for Big Data analytics (especially in business analytics tasks) sometimes poses the problem of syntactic (distribution of values) and semantic (understanding) heterogeneity between data gathered from different sources [51].

Knowledge discovery from imbalanced class dataset is a challenging topic in data mining. This problem takes place when the number of cases that represent one class is remarkably lower than the number on other classes. This is not uncommon when samples of one class occur with remarkably lower probability than in the other one. The ensemble of classifiers process needs to find a way to handle these situations. Li [52] proposed a bagging of classification to creating the ensemble learner. They used the minority class data maximally without creating synthetic data or making changes to the existing classification systems. The experimental results using real world imbalanced data are advocated for the efficacy of the proposal. In [77], a review of existing methods for classification of data with imbalanced class distribution was provided. This work gives an overview of the classification of imbalanced data regarding the application domains and the nature of the problem. They also addressed how to obtain objective and measure accuracy, the learning difficulties with conventional classifier learning algorithms, and finally the class imbalance problem in the presence of multiple classes. For this reason, diversity is particularly crucial for ensembles. Mirza et al. [60] proposed an ensemble of online sequential extreme learning machine (ESOS-ELM) approach that employs neural networks and two different learning schemes to handle the imbalance ratio in data streams. However, the algorithm presented some limitations due to the assumption that no drift takes place on the minority class.

This diversity in real-world applications has served along with a growth of research from researchers. Hence, Kuncheva [47] have studied ten statistics to determine diversity among binary classifier outputs. They have examined the relationship between the ensemble accuracy and different measures of diversity, and among the measures themselves. Their results disagreed with some proven relationship between accuracy and diversity measures in building classifier ensembles in real-life pattern recognition problems. Brown et al. [13] first reviewed the various attempts to provide a formal interpretation of error diversity, including several heuristics and qualitative explanations in the literature. They surveyed the various techniques used for creating various ensembles, and categorized them, forming an exploratory taxonomy of diversity creation methods. They introduced the concept of implicit and explicit diversity generation methods and proposed some new directions that may prove useful in understanding classification error diversity.

Another challenge for creating an ensemble of classifiers is the combination method. It plays a significant role in the process of ensemble classifiers final decision and classification accuracy. In this regard, Jain et al. [36] explored the principle of several classifiers combination method. They also mentioned different types of combination process like feature sets, training sets, classification methods, or different training sessions joined together. The result is fused by a number of classifiers with the hope of improving overall classification accuracy. The nearest mean method for classifier combination gave the best overall result. This was also the best result of the entire experiment. Kleinberg [41] bridged the gap between the theoretical prediction expressed by stochastic discernment and practical explanation of algorithmic implementation. He also tested and found out that stochastic discrimination outperformed both boosting and bagging in the majority of benchmark problems.

In machine learning, many researchers used the ensemble of classifiers to improve the accuracy of individual classifiers by mixing many of them. Neither of these learning methods alone solves the class imbalance problem. The ensemble algorithms need to be specially designed to deal with these problems. Other relevant problems with the data characteristics need to be handled in the pre-processing step of the ensemble learning. The recent advances of redundancy elimination techniques from big data (faster version of minimum-Redundancy- Maximum-Relevance (Fast-mRMR) [70]) have opened new opportunities for ensemble learning by reducing the redundancy in big data. To reduce the noise, García-Gil et al. [29] proposed a homogeneous ensemble and a heterogeneous ensemble filter approach for big data. Research on ensemble learning for Big Data is advancing through solving the different problems with peculiar characteristics [28] and proposing new methods and approaches to tackle the big data using ensemble learning [57, 78, 86, 91].

We expect the future research in the big data analytics will continue their investigation to address these issues. From this comprehensive review, it is evident that the ensemble learning has the power to deal with the analytics of big data. Additionally, Meta-heuristics will increase the capability of ensemble learning into a new dimension. Hence, the ensemble learning is posed as an effective way of utilizing the power of meta-heuristics to solve real-world big data analytics problems.

Acknowledgements Pablo Moscato acknowledges previous support from the Australian Research Council Future Fellowship FT120100060 and Australian Research Council Discovery Projects DP120102576 and DP140104183.

References

1. Abbasimehr H, Setak M, Tarokh MJ (2014) A comparative assessment of the performance of ensemble learning in customer churn prediction. Int Arab J Inf Technol 11(6):599–606
2. Akbaş A, Turhal U, Babur S, Avci C (2013) Performance improvement with combining multiple approaches to diagnosis of thyroid cancer. Engineering 5(10):264
3. Amini M, Rezaeenour J, Hadavandi E (2015) A cluster-based data balancing ensemble classifier for response modeling in bank direct marketing. International Journal of Computational Intelligence and Applications 14(4)
4. Awang MK, Makhtar M, Rahman MNA, Deris MM (2016) A new customer churn prediction approach based on soft set ensemble pruning. In: SCDM, Springer, Advances in Intelligent Systems and Computing, vol 549, pp 427–436
5. Baumann A, Lessmann S, Coussement K, Bock KWD (2015) Maximize what matters: Predicting customer churn with decision-centric ensemble selection. In: ECIS
6. Blanco R, Larrañaga P, Inza I, Sierra B (2004) Gene selection for cancer classification using wrapper approaches. International Journal of Pattern Recognition and Artificial Intelligence 18(08):1373–1390
7. Blaszczynski J, Dembczynski K, Kotlowski W, Pawlowski M (2006) Mining direct marketing data by ensembles of weak learners and rough set methods. In: DaWaK, Springer, Lecture Notes in Computer Science, vol 4081, pp 218–227
8. Bock KWD, den Poel DV (2010) Ensembles of probability estimation trees for customer churn prediction. In: IEA/AIE (2), Springer, Lecture Notes in Computer Science, vol 6097, pp 57–66
9. Bock KWD, den Poel DV (2011) An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. Expert Syst Appl 38(10):12,293–12,301
10. Breiman L (2001) Random forests. Machine Learning 45(1):5–32
11. Breiman L, Breiman L (1996) Bagging predictors. In: Machine Learning, pp 123–140
12. Breiman L, et al (2001) Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science 16(3):199–231

13. Brown G, Wyatt J, Harris R, Yao X (2005) Diversity creation methods: a survey and categorisation. Information Fusion 6(1):5–20
14. Cleofas L, Valdovinos RM, García V, Alejo R, Universitario C, Valle U (2009) Use of Ensemble Based on GA for Imbalance Problem. In: 6th International Symposium on Neural Networks, ISNN 2009 Wuhan, China, May 26-29, 2009 Proceedings, Part II, Springer Berlin Heidelberg, pp 547–554
15. Cunningham P, Carney J (2000) Diversity versus quality in classification ensembles based on feature selection. In: European Conference on Machine Learning, Springer, pp 109–116
16. Cuzzocrea A, Folino F, Guarascio M, Pontieri L (2016) A multi-view multi-dimensional ensemble learning approach to mining business process deviances. In: IJCNN, IEEE, pp 3809–3816
17. Dahiya S, Handa S, Singh N (2017) A feature selection enabled hybrid-bagging algorithm for credit risk evaluation. Expert Systems
18. Dietterich T (2000) Ensemble methods in machine learning. In: Multiple Classifier Systems, Lecture Notes in Computer Science, vol 1857, Springer Berlin Heidelberg, pp 1–15
19. Dounias G, Tsakonas A, Charalampakis D, Vasilakis E (2013) Effective business plan evaluation using an evolutionary ensemble. In: DATA, SciTePress, pp 97–103
20. Duangsoithong R, Windeatt T (2010) Bootstrap feature selection for ensemble classifiers. In: Proceedings of the 10th industrial conference on Advances in data mining: applications and theoretical aspects, Springer-Verlag, Berlin, Heidelberg, ICDM'10, pp 28–41
21. Ebrahimpour MK, Eftekhari M (2017) Ensemble of feature selection methods: A hesitant fuzzy sets approach. Applied Soft Computing 50:300–312
22. Fallahpour S, Lakvan EN, Zadeh MH (2017) Using an ensemble classifier based on sequential floating forward selection for financial distress prediction problem. Journal of Retailing and Consumer Services 34:159–167
23. Folino F, Guarascio M, Pontieri L (2012) Context-aware predictions on business processes: An ensemble-based solution. In: NFMCP, Springer, Lecture Notes in Computer Science, vol 7765, pp 215–229
24. Gaber MM, Bader-El-Den M (2012) Optimisation of Ensemble Classifiers using Genetic Algorithm. In: Graña M, Toro C, Posada J, Howlett RJ, Jain LC (eds) Advances in Knowledge-Based and Intelligent Information and Engineering Systems, IOS Press
25. Gabrys B, Ruta D (2006) Genetic algorithms in classifier fusion. Applied Soft Computing 6(4):337–347
26. Galar M, Fernández A, Barrenechea E, Bustince H, Herrera F (2011) An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. Pattern Recognition 44(8):1761–1776
27. Galar M, Fernández A, Barrenechea E, Bustince H, Herrera F (2012) A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 42(4):463–484
28. García S, Ramírez-Gallego S, Luengo J, Benítez JM, Herrera F (2016) Big data preprocessing: methods and prospects. Big Data Analytics 1(1):9
29. García-Gil D, Luengo J, García S, Herrera F (2017) Enabling Smart Data: Noise filtering in Big Data classification. ArXiv e-prints 1704.01770
30. Govindarajan M (2015) Comparative study of ensemble classifiers for direct marketing. Intelligent Decision Technologies 9(2):141–152
31. Espejo P, Ventura S, Herrera F (2010) A survey on the application of genetic programming to classification. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 40(2):121–144
32. Haque MN, Noman N, Berretta R, Moscato P (2016) Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification. PLoS ONE 11(1):e0146,116.

33. Haque MN, Noman N, Berretta R, Moscato P (2016b) Optimising weights for heterogeneous ensemble of classifiers with differential evolution. In: 2016 IEEE Congress on Evolutionary Computation (CEC), pp 233–240
34. Hernández-Lobato D, Martínez-Muñoz G, Suárez A (2006) Pruning in ordered regression bagging ensembles. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp 1266–1273
35. Hu H, Li J, Wang H, Daggard G (2008) Robustness analysis of diversified ensemble decision tree algorithms for Microarray data classification. 2008 International Conference on Machine Learning and Cybernetics pp 115–120
36. Jain AK, Duin RPW, Mao J, Member S (2000) Statistical Pattern Recognition : A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(1):4–37
37. Jordan MI, Jacobs RA (1994) Hierarchical Mixtures of Experts and the EM Algorithm. Neural Computation 6(2):181–214
38. Kim Y (2009) Boosting and measuring the performance of ensembles for a successful database marketing. Expert Syst Appl 36(2):2161–2176
39. Kim Y, Street WN, Menczer F (2006) Optimal ensemble construction via meta-evolutionary ensembles. Expert Systems with Applications 30(4):705–714
40. Kim YW, Oh IS (2008) Classifier ensemble selection using hybrid genetic algorithms. Pattern Recognition Letters 29(6):796–802
41. Kleinberg E (2000) On the algorithmic implementation of stochastic discrimination. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(5):473–490
42. Ko AH, Sabourin R, Britto AS, Jr (2008) From dynamic classifier selection to dynamic ensemble selection. Pattern Recognition 41(5):1718–1731
43. Kotsiantis S, Zaharakis I, Pintelas P (2006) Machine learning: a review of classification and combining techniques. Artificial Intelligence Review 26(3):159–190
44. Koutanaei FN, Sajedi H, Khanbabaei M (2015) A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services 27:11–23
45. Krawczyk B, Galar M, Jeleń Ł, Herrera F (2016) Evolutionary undersampling boosting for imbalanced classification of breast cancer malignancy. Applied Soft Computing 38:714–726
46. Kuncheva L, Jain L (2000) Designing classifier fusion systems by genetic algorithms. IEEE Transactions on Evolutionary Computation 4(4):327–336
47. Kuncheva L, Whitaker C (2003) Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Machine Learning 51(2):181–207
48. Kuncheva LI (2014) Combining Pattern Classifiers: Methods and Algorithms, 2nd edn. John Wiley & Sons, Inc.
49. Lai KK, Yu L, Wang S, Huang W (2007) An intelligent CRM system for identifying high-risk customers: An ensemble data mining approach. In: International Conference on Computational Science (2), Springer, Lecture Notes in Computer Science, vol 4488, pp 486–489
50. Lertampaiporn S, Thammarongtham C, Nukoolkit C, Kaewkamnerdpong B, Ruengjitchatchawalya M (2013) Heterogeneous ensemble approach with discriminative features and modified-SMOTEBagging for pre-miRNA classification. Nucleic acids research 41(1):e21
51. L'Heureux A, Grolinger K, ElYamany HF, Capretz M (2017) Machine learning with big data: Challenges and approaches. IEEE Access PP(99)
52. Li C (2007) Classifying imbalanced data using a bagging ensemble variation (BEV). In: Proceedings of the 45th annual southeast regional conference, ACM, New York, NY, USA, ACM-SE 45, pp 203–208
53. Li H, Sun J (2011) Principal component case-based reasoning ensemble for business failure prediction. Information & Management 48(6):220–227
54. Li H, Sun J (2012) Case-based reasoning ensemble and business application: A computational approach from multiple case representations driven by randomness. Expert Syst Appl 39(3):3298–3310

55. Li H, Sun J, Li J, Yan X (2013) Forecasting business failure using two-stage ensemble of multivariate discriminant analysis and logistic regression. Expert Systems 30(5):385–397
56. Liu Y, Wei W, Wang K, Liao Z, Gao J (2011) Balanced-sampling-based heterogeneous SVR ensemble for business demand forecasting. In: ICIC (1), Springer, Lecture Notes in Computer Science, vol 6838, pp 91–99
57. Ma C, Zhang HH, Wang X (2014) Machine learning for big data analytics in plants. Trends in plant science 19(12):798–808
58. Matthews C, Scheurmann E (2008) Ensembles of classifiers in arrears management. In: Soft Computing Applications in Business, Studies in Fuzziness and Soft Computing, vol 230, Springer, pp 1–18
59. Minaei-Bidgoli B, Kortemeyer G, Punch W (2004) Optimizing classification ensembles via a genetic algorithm for a web-based educational system. In: Fred A, Caelli T, Duin R, Campilho A, de Ridder D (eds) Structural, Syntactic, and Statistical Pattern Recognition, Lecture Notes in Computer Science, vol 3138, Springer Berlin Heidelberg, pp 397–406
60. Mirza B, Lin Z, Liu N (2015) Ensemble of subset online sequential extreme learning machine for class imbalance and concept drift. Neurocomputing 149:316–329
61. Namsrai E, Munkhdalai T, Li M, Shin JH, Namsrai OE, Ryu KH (2013) A Feature Selection-based Ensemble Method for Arrhythmia Classification. Journal of Information Processing Systems 9(1):31–40
62. Nikulin V, Mclachlan GJ, Ng SK (2009) Ensemble Approach for the Classification of Imbalanced Data. In: AI 2009: Advances in Artificial Intelligence, Springer, pp 291–300
63. Oh DY, Gray JB (2013) GA-Ensemble: a genetic algorithm for robust ensembles. Computational Statistics 28(5):2333–2347
64. Oliveira L, Morita M, Sabourin R, Bortolozzi F (2005) Multi-objective genetic algorithms to create ensemble of classifiers. In: Coello Coello CA, Hernández Aguirre A, Zitzler E (eds) Evolutionary Multi-Criterion Optimization, Lecture Notes in Computer Science, vol 3410, Springer Berlin Heidelberg, pp 592–606
65. Oliveira LS, Sabourin R, Bortolozzi F, Suen CY (2003) Feature selection for ensembles: A hierarchical multi-objective genetic algorithm approach. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 2, IEEE Computer Society, Washington, DC, USA, ICDAR ’03, pp 676–680
66. Osareh A, Shadgar B (2013) An Efficient Ensemble Learning Method for Gene Microarray Classification. BioMed Research International 2013:1–10
67. Oza NC (2006) Ensemble data mining methods. In: Wang J (ed) Encyclopedia of Data Warehousing and Mining, Idea Group Reference, vol 1, pp 448–453
68. Oza NC, Tumer K (2008) Classifier ensembles: Select real-world applications. Information Fusion 9(1):4–20
69. Polikar R (2006) Ensemble based systems in decision making. Circuits and Systems Magazine, IEEE 6(3):21–45
70. Ramírez-Gallego S, Lastra I, Martínez-Rego D, Bolón-Canedo V, Benítez JM, Herrera F, Alonso-Betanzos A (2017) Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data. International Journal of Intelligent Systems 32(2):134–152
71. Ranawana R, Palade V (2006) Multi-Classifier Systems: Review and a roadmap for developers. International Journal of Hybrid Intelligent Systems 3(1):35–61
72. Rokach L (2009) Ensemble-based classifiers. Artificial Intelligence Review 33(1-2):1–39
73. Roli F, Giacinto G, Vernazza G (2001) Methods for designing multiple classifier systems. In: Kittler J, Roli F (eds) Multiple Classifier Systems, Lecture Notes in Computer Science, vol 2096, Springer Berlin Heidelberg, pp 78–87
74. Santana A, Soares R, Canuto A, Souto MCPd (2006) A dynamic classifier selection method to build ensembles using accuracy and diversity. In: Neural Networks, 2006. SBRN ’06. Ninth Brazilian Symposium on, pp 36–41
75. Seijo-Pardo B, Bolón-Canedo V, Alonso-Betanzos A (2017) Testing different ensemble configurations for feature selection. Neural Processing Letters pp 1–24

76. Srimani PK, Koti MS (2013) Medical Diagnosis Using Ensemble Classifiers - A Novel Machine-Learning Approach. Journal of Advanced Computing pp 9–27
77. Sun Y, Wong AKC, Kamel MS (2009) Classification of Imbalanced Data: A Review. International Journal of Pattern Recognition and Artificial Intelligence 23(04):687–719
78. Tang Y, Wang Y, Cooper KM, Li L (2014) Towards big data Bayesian network learning-an ensemble learning based approach. In: Big Data (BigData Congress), 2014 IEEE International Congress on, IEEE, pp 355–357
79. Thammasiri D, Meesad P (2012) Ensemble Data Classification based on Diversity of Classifiers Optimized by Genetic Algorithm. Advanced Materials Research 433-440:6572–6578
80. Tsymbal A, Puuronen S, Patterson DW (2003) Ensemble feature selection with the simple Bayesian classification. Information Fusion 4(2):87–100
81. Tulyakov S, Jaeger S, Govindaraju V, Doermann D (2008) Review of classifier combination methods. In: Marinai S, Fujisawa H (eds) Machine Learning in Document Analysis and Recognition, Studies in Computational Intelligence, vol 90, Springer Berlin Heidelberg, pp 361–386
82. Valentini G, Masulli F (2002) Ensembles of learning machines. In: Marinaro M, Tagliaferri R (eds) Neural Nets, Lecture Notes in Computer Science, vol 2486, Springer Berlin Heidelberg, pp 3–20
83. Wang L, Wu C (2017) Business failure prediction based on two-stage selective ensemble with manifold learning algorithm and kernel-based fuzzy self-organizing map. Knowl-Based Syst 121:99–110
84. Wang X, Wang H (2006) Classification by evolutionary ensembles. Pattern Recognition 39(4):595–607
85. Wang Y, Xiao H (2011) Ensemble learning for customers targeting. In: KSEM, Springer, Lecture Notes in Computer Science, vol 7091, pp 24–31
86. Wang Y, Yu C (2016) Research on the database marketing in the big data environment based on ensemble learning. Economics 12(6):21–32
87. Xiao J, Xie L, He C, Jiang X (2012) Dynamic classifier ensemble model for customer classification with imbalanced class distribution. Expert Syst Appl 39(3):3668–3675
88. Xiao J, Wang Y, Wang S (2013) A dynamic transfer ensemble model for customer churn prediction. In: BIFE, IEEE Computer Society, pp 115–119
89. Xiao J, Xiao Y, Huang A, Liu D, Wang S (2015) Feature-selection-based dynamic transfer ensemble model for customer churn prediction. Knowl Inf Syst 43(1):29–51
90. Xiao J, Jiang X, He C, Teng G (2016) Churn prediction in customer relationship management via GMDH-based multiple classifiers ensemble. IEEE Intelligent Systems 31(2):37–44
91. Xie L, Draizen EJ, Bourne PE (2017) Harnessing big data for systems pharmacology. Annual Review of Pharmacology and Toxicology 57:245–262
92. Xu R, He L (2008) GACEM: Genetic Algorithm Based Classifier Ensemble in a Multi-sensor System. Sensors 8(10):6203–6224
93. Yang P, Liu W, Zhou BB, Chawla S, Zomaya AY (2013) Ensemble-based wrapper methods for feature selection and class imbalance learning. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, pp 544–555
94. Yu L, Lai KK, Wang S (2008a) An evolutionary programming based knowledge ensemble model for business risk identification. In: Soft Computing Applications in Business, Studies in Fuzziness and Soft Computing, vol 230, Springer, pp 57–72
95. Yu L, Wang S, Lai KK (2008b) An EMD-based neural network ensemble learning model for world crude oil spot price forecasting. In: Soft Computing Applications in Business, Studies in Fuzziness and Soft Computing, vol 230, Springer, pp 261–271
96. Zhang L, Wang X, Moon WM (2015) PolSAR images classification through GA-based selective ensemble learning. In: Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, pp 3770–3773
97. Zhang Y, Bhattacharyya S (2004) Genetic programming in classifying large-scale data: an ensemble method. Information Sciences 163(1–3):85–101

98. Zhang Z, Chen Q, Ke S, Wu Y, Qi F (2010a) Ranking potential customers based on group-ensemble. In: Strategic Advancements in Utilizing Data Mining and Warehousing Technologies, IGI Global, pp 355–365
99. Zhang Z, Li J, Hu H, Zhou H (2010b) A robust ensemble classification method analysis. In: Arabnia HR (ed) Advances in Computational Biology, Advances in Experimental Medicine and Biology, vol 680, Springer New York, pp 149–155
100. Zhao W, Liu H, Dai W, Ma J (2016) An entropy-based clustering ensemble method to support resource allocation in business process management. Knowl Inf Syst 48(2):305–330
101. Zhou L, Pan S, Wang J, Vasilakos AV (2017) Machine learning on big data: Opportunities and challenges. Neurocomputing 237:350–361