Front. Comput. Sci.
https://doi.org/10.1007/s11704-019-8208-z

REVIEW ARTICLE

A survey on ensemble learning

Xibin DONG[1], **Zhiwen YU** (✉)[1], **Wenming CAO**[2], **Yifan SHI**[1], **Qianli MA**[1]

1 School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China
2 Department of Computer Science, City University of Hong Kong, Hong Kong SAR 999077, China

Abstract Despite significant successes achieved in knowledge discovery, traditional machine learning methods may fail to obtain satisfactory performances when dealing with complex data, such as imbalanced, high-dimensional, noisy data, etc. The reason behind is that it is difficult for these methods to capture multiple characteristics and underlying structure of data. In this context, it becomes an important topic in the data mining field that how to effectively construct an efficient knowledge discovery and mining model. Ensemble learning, as one research hot spot, aims to integrate data fusion, data modeling, and data mining into a unified framework. Specifically, ensemble learning firstly extracts a set of features with a variety of transformations. Based on these learned features, multiple learning algorithms are utilized to produce weak predictive results. Finally, ensemble learning fuses the informative knowledge from the above results obtained to achieve knowledge discovery and better predictive performance via voting schemes in an adaptive way. In this paper, we review the research progress of the mainstream approaches of ensemble learning and classify them based on different characteristics. In addition, we present challenges and possible research directions for each mainstream approach of ensemble learning, and we also give an extra introduction for the combination of ensemble learning with other machine learning hot spots such as deep learning, reinforcement learning, etc.

Keywords ensemble learning, supervised ensemble classification, semi-supervised ensemble classification, clustering ensemble, semi-supervised clustering ensemble

Received March 12, 2016; accepted July 21, 2016

E-mail: zhwyu@scut.edu.cn; wenmincao2-c@my.cityu.edu.hk

1 Introduction

Ensemble learning methods exploit multiple machine learning algorithms to produce weak predictive results based on features extracted through a diversity of projections on data, and fuse results with various voting mechanisms to achieve better performances than that obtained from any constituent algorithm alone [1]. From Fig. 1 (modified with reference to Fig. 6 of the essay "Understanding the Bias-Variance Tradeoff" by Fortmann-Roe S. See the author's personal website), we observe that the total error of the learning model declines continuously until reaching the bottom, followed by a rapid upward trend, when the model complexity increases. The trends of the bias and variance are opposite: the former drop dramatically before remaining steady, while the latter holds steady before rising heavily. We deduce that when increasing the model complexity to improve the performance of this model, we are dedicated to reaching a delicate balance between bias and variance. The earliest work of ensemble learning can date back to the last century [2–4]. Dasarathy and Sheela [2] proposed to utilize component classifiers trained from different categories to constitute a composite classification system, thereby enhancing the performance of identification systems. Kearns [3] investigated the equivalent problem about the relationships between the weak learning algorithms and the strong learning algorithms in PCA learning model. Afterward, Schapire and Robert [4] explored the feasibility of incorporating multiple weak learning models into a high-precision model. The past decades have witnessed ensemble learning drawing increasing attention, and researchers have carried out a great amount of exploration and innovation, and in some international machine learning competitions like Kaggle, KDD-Cups, etc., ensemble learning has achieved exceptionally satisfactory performance. To put it more specifically, ensemble learning aims to integrate various machine learning algorithms into a unified framework seamlessly. Thus, the complementary information of each algorithm is effectively utilized to allow better performance of the overall model. From this perspective, ensemble learning is extremely extensible to combine with diverse machine learning models for different types of tasks, such as common classification tasks, clustering tasks, etc. Generally speaking, existing ensemble learning methods can be grouped into four categories: supervised ensemble classification, semi-supervised ensemble classification, clustering ensemble, and semi-supervised clustering ensemble. We would discuss these methods in the following sections from their research progress and algorithm applications to challenges encountered and give some potential future directions. We also give a brief introduction to the main research issues of ensemble classification and clustering ensemble in Fig. 2 and Fig. 3 to make readers know the main content of this survey more clearly.

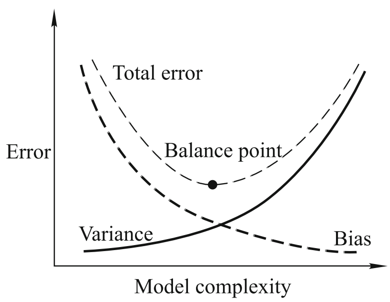


Fig. 1 The relationship between learning curve and model complexity

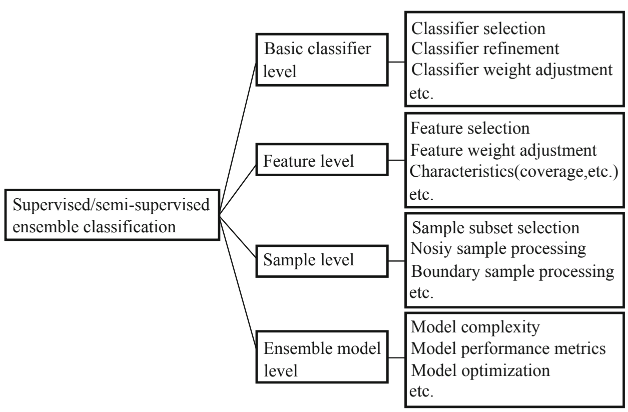


Fig. 2 Main research issues of ensemble classification

The rest of this survey is organized as follows. In Section 2, we review the supervised ensemble classification (in short, we call it ensemble classification). Next, we give an overview of semi-supervised ensemble classification in Section 3. Section 4 introduces the researches of clustering ensemble. In Section 5, we provide a conspectus of semi-supervised clustering ensemble. In Section 6 we give a more specific introduction for the combination of ensemble learning with deep learning and other hot spots in machine learning, and in Section 7, we give a summary and discussion about the survey.

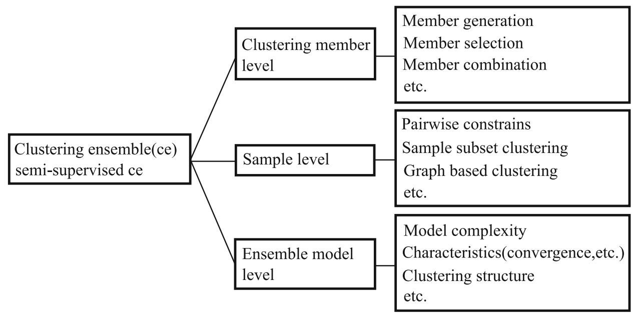


Fig. 3 Main research issues of clustering ensemble

2 Supervised ensemble classification

Figure 4 illustrates the main idea of a typical ensemble classification model, which consists of two steps: (1) generating classification results using multiple weak classifiers, and (2) integrating multiple results into a consistency function to get the final result with voting schemes. The widely-used ensemble classification methods include bagging [5], AdaBoost [6], random forest [7], random subspace [8], gradient boosting [9]. The Bagging method generates sample subsets by randomly sampling from the training data set, and then uses these obtained subsets to train the basic models for integration. The training of basic models in the Bagging model is performed in a parallel manner. From Fig. 5, we can see that AdaBoost focuses on samples that are misclassified through adjusting weights of samples iteratively, thereby improving classification performances of basic models for the final integration. It is worth noticing that the training of basic models in AdaBoost is conducted in a tandem manner instead of in a parallel manner. Random Forest trains multiple decision tree models from two perspectives: the sample dimension and the feature dimension. As a result, it alleviates the problem that decision trees are prone to over-fitting by integrating voting results of multiple decision trees. The training of basic models in the Random Forest is parallel, which is similar to that in Bagging. Random Subspace constructs a set of feature subspaces via randomly sampling features, and then trains basic classifiers in these subspaces to generate multiple results before being fused into the final result. The basic models in

Random Subspace are trained parallelly. As Fig. 6 shows, Gradient Boosting randomly samples to get sample subsets and then each learner is constructed and trained to reduce the residuals generated by the previous learner. Consequently, Gradient Boosting can make the sum of the final residuals from the integrated models small enough, thereby forcing the prediction close to the actual value. Similar to that in AdaBoost, basic models in the gradient boosting are trained in a tandem way. In addition to the above-mentioned methods, there are some other methods in the ensemble classification category.

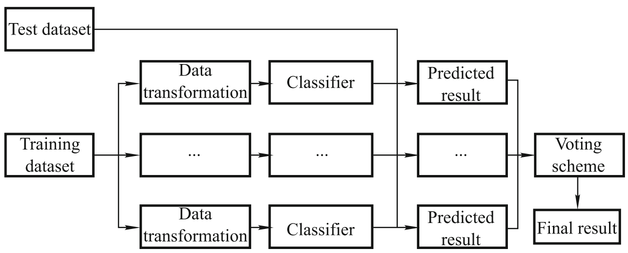


Fig. 4 The framework of ensemble classification

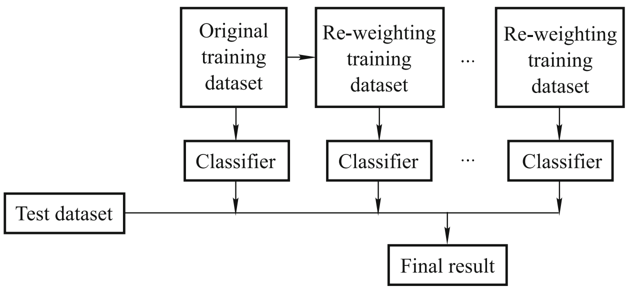


Fig. 5 The framework of AdaBoost

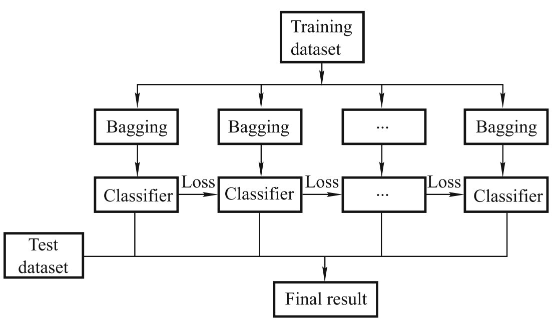


Fig. 6 The framework of Gradient Boosting

At the sample level, ensemble classification methods can be grouped into two categories concerning the manner that how the sample subsets are selected. Specifically, Garcia-Pedrajas [10] combined an instance selection algorithm with Boosting method to effectively reduce the space complexity when constructing basic classifiers, and focus on the instances which were difficult to learn using a biased distribution. The instance selection algorithm in this method is also suitable for other basic classifiers such as KNN and SVM. Afterward, Garcia-Pedrajas et al. [11] fused Boosting method and Random Subspace method into a unified framework, in which a series of subspaces were obtained. In each subspace, misclassified instances were utilized to generate supervised projections that span a space where the next classifier was trained using all instances. The final results were obtained by using such serial operations.

At the feature level, some works carried out studies on characteristics of features embedded in the original data from different aspects [12, 13]. Kuncheva et al. [12] investigated three feature criteria including usability, coverage, and diversity for random subspace ensemble, these criteria were used for searching appropriate ensemble size and feature size to improve the accuracy and diversity of classifiers. Ye et al. [13] proposed a stratified sampling method to divide features into two groups: one with strong information and the other with weak information. With these two groups, one can construct multiple feature subspaces by proportionally sampling from each group. This special sampling makes it possible that every subspace contains enough informative features for training better classifiers and increasing diversities of classifiers. These above studies demonstrate that the characteristics of features can be utilized to effectively improve the performance of an ensemble classification model in terms of prediction accuracy.

Different from works that explored characteristics of features, some works focused on feature subset selection, feature extraction, redundancy feature removal, etc. Empirical evidence has demonstrated that such explorations can effectively reduce dimensions of features, and decrease negative effects brought about by noisy data. For example, Bryll et al. [14] used Bagging technique to do sampling on the attributes of instances and used feature subsets to train basic classifiers, where the performance on training data set was used for adjusting the size of attribute subsets and the number of voters. Based on co-training algorithm [15] which tries to generate basic classifiers from different perspectives, Wang et al. [16] extended co-training to the multi-view situation, in which random subspaces of feature space were chosen to train multiple classifiers, since classifiers showed various sensitivities to subspaces and provided complementary information for each other. Besides, Yaslan and Cataltepe [17] proposed a relevant random subspace co-training approach, in which features were drawn with probabilities proportional to corresponding relevance that were quantified by the mutual information between features and class labels. Zhang and Zhang [18] introduced random discriminant information into canonical correlation analysis for feature extraction in multi-view ensemble learning. Guo et al. [19] presented a dynamic rough subspace based selective ensemble method, in which the rough set theory was adopted to reduce the searching space and the dimensionality of features. Windeatt et al. [20] used a feature ranking scheme to remove irrelevant features in multilayer perceptron ensemble. Additionally, researches at the feature level include how to design and construct new features for classifier construction. Rodriguez et al. [21] designed Rotation Forest method to extract features from the original features using coordinate rotation transformation. Takemura et al. [22] constructed pattern-spectrum based features using mathematical morphology computation, in which the spectrum parameters could help quantify analysis of tumor shapes. Amasyali and Ersoy [23] presented extended space forest method by performing different transformations on pairs of original features to generate random feature combinations. Concerning the missing feature processing, Polikar et al. [24] proposed Learn++.MF method to solve the problem of missing features with random subspace method, in which labels of data with missing features were predicted by the classifier trained on data that contained the corresponding missing feature. The performance of this method experienced a gradual decline trend when the number of missing data increased. Concerning the feature weight adjustment, Nanni and Lumini [25] used particle swarm optimization algorithm [26] to assign optimal weights for features in each random subspace, thereby minimizing the error rate in the training process.

At the basic model level, many basic models are trained for prediction, then the ensemble model fuses the predictive results from these basic models via a consensus function, however, not all basic models are beneficial to the final result of the integration. As shown in Fig. 7, it is expected that basic models beneficial to the integration performances are selected, and models that are redundant or have negative effects are removed. Specifically, Zhou and Tang [27] proposed GASEN-b algorithm using bit strings to represent the appearance of tree classifiers in the ensemble and adopted genetic algorithm on bit strings to conduct classifier selection. Diao et al. [28] developed a thought of feature selection to facilitate the selection of basic models by considering classifiers as features after converting ensemble predictions into training samples. Yu et al. [29] proposed a progressive selection process by considering the sample space and the feature space simultaneously, in which a cost function was designed to fuse the current and perennial information to serve the classifier selection sequentially. Recently, they propose the hybrid incremental ensemble learning (HIEL) approach [30] which takes into consideration the feature space and the sample space simultaneously for classifier selection to handle noisy dataset. Dos Santos et al. [31] combined optimization process with the dynamic selection strategy to select the most confident subsets of classifiers with high accuracy. Hernández-Lobato et al. [32] used statistical techniques to evaluate confident levels of basic classifier subsets for voting results, in which the polling of classifiers would be halted once the remaining failed to change the final result with probabilities above the specified confidence level. Moreover, it is an important issue that how to design and utilize effective methods to integrate results from basic models. Martínez-Muñoz et al. [33] studied several pruning strategies and found that an appropriately ordered classifier aggregation could better the model in terms of accuracy and robustness. De Stefano et al. [34] exploited Bayesian network to merge the response of decision tree ensembles, in which performance was improved and the number of classifiers was considerably reduced. Rahman and Verma [35] proposed a cluster-oriented hierarchical ensemble classification algorithm which integrated classifiers generated by applying clustering algorithms to the data in multiple layers. In [36], a hierarchical ensemble classification algorithm was further proposed which was based on clustering confidence vectors. In general, these studies focus on filtering basic models using statistical knowledge, optimization process, etc., and adopting various schemes to integrate predictions from basic models to obtain the final prediction. There are also some other studies that introduce different types of basic models into ensemble classification, and these basic models include neural networks, support vector machines, etc. These works [37–39] show that the ensemble method can be easily compatible with other machine learning methods. Specifically, Zhang and Suganthan [37] introduced support vector machine into oblique decision tree ensemble [38] to help get the testing hyperplane for internal nodes to do classification. Zhou et al. [39] proposed a neural network based ensemble method, where the weights of networks were updated by genetic algorithm. Apart from the abovementioned researches, some works concentrate on the refinement of basic models. Since the K-nearest neighbor algorithm cannot sufficiently utilize the information embedded in the feature space, Yu et al. [40] proposed a hybrid KNN classification approach to relieve this limitation, in which random subspace method was adopted for the ensemble. Moreover, Li et al. [41] proposed a random subspace evidence classifier which used the information of both the whole feature space and the random subspace to calculate basic belief for classification.

There have been lots of studies at the level of the entire ensemble model, which has some characteristics including complexity, sparsity, stability, etc. For example, Hernández-Lobato et al. [42] explored the effect of the size of a parallel ensemble on the aggregated predictive result, in which the minimum number of classifiers was estimated. Wang et al. [43] found that with the fuzziness of basic classifiers becoming higher, the ensemble classification model could achieve better generalization ability when handling data sets with complex boundaries. Kuncheva [44] used the kappa-error diagrams to analyze the property of classifiers ensemble which showed that accuracies of basic models played an important role in improving performance of ensemble classification model. Gao and Zhou [45] introduced the concept which called approximation stability to help evaluate the ensemble model stability. Yin et al. [46] studied the sparsity and diversity of ensemble model, and introduced the concept called diversity contribution ability for classifier selection and weight adjustment. Zhang and Suganthan [47] increased diversities of the tree classifiers in random forest by concatenating different rotation spaces into a higher space in the root node, and searched the best rotation method for child nodes. Li et al. [48] studied the connection between the diversity and the generalization performance of ensemble models, and found that the diversity could be used for the ensemble regularization. Zhang et al. [49] constructed basic classifiers by resampling pairwise constraints to improve diversities of models. Zhou and Li [50] studied the ensemble diversity from the perspective of information theory and found that ensemble diversity could be decomposed over the basic classifiers. Recently, a tree matching diversity measurement was proposed by Sun and Zhou [51] to consider the structural diversity and the behavioral diversity concurrently.

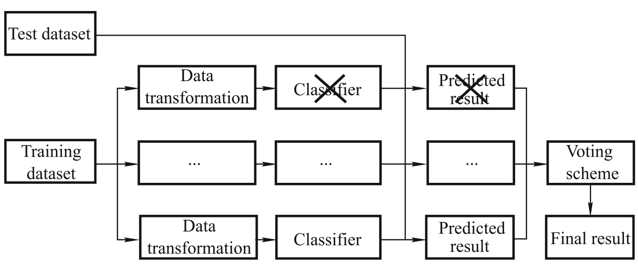


Fig. 7 The framework of model selection in ensemble method

Overall, these above researches mainly focus on characteristics of ensemble models and use the complementary information of basic models to improve the overall performance. Apart from these works, some works are dedicated to designing new measurements for performances of ensemble methods. Specifically, Mao et al. [52] designed an objective function to approximate the ensemble error to reduce the difficulty of balancing the diversity and the accuracy of ensemble model. Yu et al. [53] proposed a non-parametric test method to evaluate the suitability of ensemble methods on different data sets. Furthermore, some works studied optimization processing from different aspects such as features, basic models, etc. For instance, Kim and Cho [54] used an evolutionary algorithm to find optimal discriminant ensembles of features and classifiers. Qian et al. [55] used a bi-objective formulation to better the generalization performance and prune basic models simultaneously. Some researches focus on relaxing limitations of existing methods, such as Zhou and Feng [56] presented the multi-grained cascade forest method to rival deep neural networks (DNN) by achieving comparable performances, in which a deep forest ensemble with a cascade structure was generated for representation learning. They also proposed encoder forest [57] which adopted trees ensemble for auto-encoder and obtained lower reconstruction errors with faster training speed when compared with DNN based auto-encoders. More researches about deep forest can be seen in [58–60]. To find the optimal subspace combinations for traditional Random Subspace method, Yu et al. [61] proposed a hybrid adaptive ensemble classification algorithm to explore the optimal random subspace set.

Because ensemble classification methods have advantages in terms of accuracy, stability, and generalization, they are widely adopted to solve all kinds of problems, such as multi-instance learning, multiple-label learning, imbalance learning, etc. Specifically, Zhou and Zhang [62] transformed multi-instance representations into single-instance learning algorithms, in which the basic classifiers were trained using bags that represented features factor generated by clustering on all samples. Zhu et al. [63] proposed an active learning ensemble framework which selectively labeled instances from data streams with the guide of minimum variance principle. Brzezinski and Stefanowski [64] presented an accuracy updated ensemble method which combined accuracy-based weighting mechanisms with the incremental nature of Hoeffding Trees to solve different types of concept drifts in data stream mining. Muhlbaier et al. [65] designed a dynamically weighted consult and a voting mechanism to integrate classifiers for the incremental learning of new new class. Xiao et al. [66] proposed a dynamic classifier ensemble selection strategy to reduce the bias in classification error and obtained good performance when handling noisy data. Galar et al. [67] conducted a thoroughly empirical comparison of representative ensemble classification approach dealing with imbalanced problems. Liu et al. [68] presented EasyEnsemble method and BalanceCascade method, which generated subsets of major class firstly, and trained classifiers using the subsets, thus information of major data could be fully utilized. Sun et al. [69] proposed an evolutionary undersampling based bagging method which generated a set of accurate and diverse basic classifiers for ensemble to deal with imbalance data. Li et al. [70] adopted the wiener process which characterized the particles macroscopic movement into over-sampling to synthesize new samples for certain minority class. The regularity exhibited in the wiener process forced the generated samples to follow the distribution of the original minority class, which expanded the range of attribute values in the training data set. As a result, a stable and robust decision region can be constructed for the classifier to achieve better performance on an imbalanced data set. Abawajy et al. [71] introduced a large iterative multitier ensemble scheme tailored for handling big data. In this method, except the second-level ensemble modules using basic classifiers as members, ensemble modules in the higher level were composed of low-level ensembles instead of basic classifiers. Li et al. [72, 73] presented a selective ensemble of classifier chains method to reduce the computational cost and the storage cost arose in multi-label learning by decreasing the ensemble size. In this method, with F1-score as the performance criterion, an upper bound for the empirical risk was converted into a convex optimization problem using math-norm regularization. In addition, ensemble classification methods are widely used in the biomedical field, such as protein kinase-specific phosphorylation sites prediction [74], protein-ligand binding site localization [75], protein function prediction [76], breast cancer cell identification [77], etc. Apart from applications in the biomedical field, ensemble classification methods have been applied to the intelligent transportation area, such as pedestrian detection [78,79], vehicle type recognition [80], traffic flow prediction [81]. Furthermore, ensemble classification methods are adopted in pattern recognition applications, such as face recognition [82], hand-printed character recognition [83], multi-label image/video annotation [84], speaker verification [85], gait recognition [86], image retrieval [87] and network intrusion detection [88]. Besides, ensemble classification methods can also be adopted for social applications such as noise differentiation [89], customer relationship management [90], sentiment analysis [91], etc.

Based on the above-mentioned works, we find that there are still many challenges in the ensemble classification field that are worth further exploring in the future. Specifically, although most methods mainly consider improving the accuracy of the model at the architecture level of ensemble models, there are quite a few researches on determining the appropriate model size and reducing the complexity of the model to increase the training speed. Moreover, ensemble classification models contain many characteristics like the diversity, accuracy, generalization and so on, and these characteristics are conflicting in improving performances of the models in certain cases. Hence, there are many existing methods exploring to combine these characteristics using mixture models or multi-objective functions to optimize them simultaneously, but there is a lack of researches on theoretically analyzing relationships among these characteristics. Moreover, some researches have proved that performances of ensemble classification models can be further improved by taking the interconnection and feedback between different levels such as sample level, feature level, etc. into account and optimizing these levels simultaneously, which needs more researches. Besides, it is also a feasible research issue of performing optimization at a higher level such as the classifier collection level of ensemble model. Finally, it is necessary to expand the practical applications of ensemble classification to handle multiple-type data that may be semi-structured and unstructured, or continuous and discrete.

3 Semi-supervised ensemble classification

Semi-supervised ensemble classification methods have drawn extensive attention in the past years. Unlike ensemble classification methods, semi-supervised ensemble classification methods focus on expanding the training set and utilizing these expanded training set, Fig. 8 shows that the semi-supervised mechanism can help capture more accurate underlying data distribution by introducing more informative data [1]. Specifically, in semi-supervised ensemble classification, a classifier is first trained using limited labeled data. Secondly, this classifier is used to assign pseudo-tags to unlabeled data. These pseudo-labeled data are used to update the classifiers together with the original labeled data. Lastly, the results from the classifiers are fused to get the final prediction using a certain voting scheme. Extensive experimental results demonstrate that semi-supervised ensemble classification methods outperform other traditional ensemble classification methods in the case where the labeled data is insufficient. We would give a brief overview of semi-supervised ensemble classification models in the following.

Among existing semi-supervised ensemble classification models, some works focus on mining hidden structures, distributional information, dependencies and other characteristics of the data. As shown in Fig. 9 (RS denotes random subspace, NG denotes neighborhood graph, and SC denotes semi-supervised classifier), Yu et al. [92] proposed a graph-based semi-supervised ensemble classification method that effectively tackled the underlying structures of the high-dimensional data which was hard to characterized by graph method directly. This method constructed a neighborhood graph in each feature subspace and trained a semi-supervised linear classifier on the learnt graph for integration. They also proposed a multi-objective subspace selection process to generate the optimal combination of feature subspaces, and used unlabeled data sets to generate an auxiliary training set based on the sample confidence to improve the performance of the classifier ensemble [93]. Gharroudi et al. [94] presented a semi-supervised multi-label classification method by combining both bagging technique and random subspace strategies to construct multi-label classification models for ensemble. Lu et al. [95] presented a semi-supervised rotation forest algorithm that utilized both the discriminative and local structural information embedded in the labeled and unlabeled data to provide better class separability for classification.

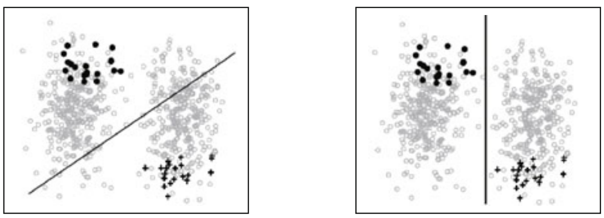


Fig. 8 Decision boundary adjusted by semi-supervised method

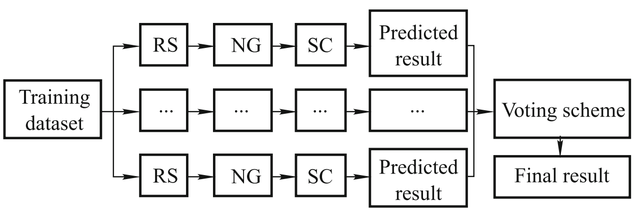


Fig. 9 The framework of semi-supervised ensemble classification in subspace

Furthermore, some works explore different manners to effectively select high-quality unlabeled data. Specifically, Wang and Chen [96] used an active data selection strategy to select unlabeled data with more probability of improving the model performance. Soares et al. [97] proposed a cluster-based boosting method which utilized a cluster-based semi-supervised optimization method to overcome potential incorrect label estimation for unlabeled data. Some researches focus on the effective use of unlabeled data to ameliorate the ensemble performance. Particularly, Woo and Park [98] used a label propagation method to predict labels and then constructed ensemble classifiers using these expanded labeled data, thereby classifiers with good diversity and accuracy were obtained. Zhang and Zhou [99] proposed UNDEED method, a semi-supervised classification method to increase the classifier accuracy on labeled data and diversity on unlabeled data simultaneously, we show some of the experiment results in Fig. 10, it shows that when semi-supervised mechanism is applied, the model performance including accuracy and stability on most of the UCI data sets can be improved (names of datasets according to the abbreviation are diabetes, heart, wdbc, austral, vote, vehicle, hepatitis, ionosphere, colic, credit_g, g241n). In addition, some other works focus on loosening limitations of existing methods. For example, Alves et al. [100] presented the social-training method using social choice functions that work with rank aggregation for heterogeneous classifier ensemble, and it was demonstrated that this method could sufficiently exploit the information about the label data when compared with traditional semi-supervised methods. Since existing methods seldom consider the optimization on the unlabeled data, Yu et al. [101] proposed a progressive semi-supervised ensemble approach which fused a progressive generation process and a self-evolutionary sample selection process into a unified framework to enrich the training set with unlabeled data.

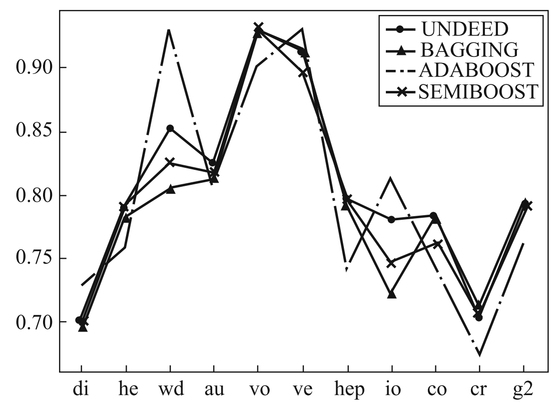


Fig. 10 Comparison of supervised and semi-supervised ensemble classification

With respect to applications of semi-supervised ensemble classification methods in the data mining area, especially when dealing with issues such as data stream processing, multi-label learning, high-dimensional data processing, semi-supervised ensemble classification can effectively better model performance by introducing semi-supervised mechanism into ensemble model. For example, Hosseini et al. [102] proposed a semi-supervised ensemble classification algorithm to classify instances of a large-scale non-stationary data stream with concept drift. Wang and Li [103] applied an incremental learning technique to avoid unnecessary repetition training and improve the accuracy of basic models for time-varying data streams. Yu et al. [104] designed an adaptive semi-supervised classifier ensemble framework which incorporated an adaptive feature selection process, adaptive weighting process and auxiliary training set generation process for high-dimensional data classification. In the biomedical field, Li and Zhou [105] proposed a computer-aided diagnosis and treatment system based on the co-forest method which used the random forest method to extend the co-training paradigm. In pattern recognition area, Guz et al. [106] investigated the performance of semi-supervised ensemble classification models with respect to self-training and co-training method for sentence segmentation of speech. Shi et al. [107] presented a semi-supervised ensemble classification method using tolerance rough set to approximate concepts existed in documents and extract an initial set of negative examples when there was no labeled negative example for text classification. Semi-supervised ensemble classification methods are also employed to other related fields, such as fault classification in micro-grids system [108], automatic mine detection [109], urban pollutant monitoring [110], spam short message service detection [111] and so on.

It is a key research point that how to effectively make use of unlabeled data, which is accompanied by a series of challenges. Apart from that, since the incorrectly-labeled samples have negative effects on the performance of models in the label propagation process, it is imperative to develop more effective schemes to reduce negative effects of these samples, which needs us to make more efforts. Furthermore, most existing semi-supervised learning methods focus on handling unlabeled data from the same field with labeled data. We can explore to use different fields of data to expand the training sets and model the relationships between unlabeled data and labeled data by combining different methods, such as graph theory, probability theory, etc. Additionally, since the traditional models need to be retrained when introducing new data for the semi-supervision method, it would be a feasible direction to design a model with efficient update ability.

4 Clustering ensemble

Clustering ensemble algorithm works by generating a series of clustering partition using clustering algorithms and combining the partitions together to get the consensus solution. Clustering ensemble methods have better performances in terms of accuracy, robustness, and stability when compared with single clustering algorithms because they can make full use of the information provided by its clustering members. The related work can date back to Strehl and Ghosh's research [112]. As shown in Fig. 11, a typical framework of the clustering ensemble framework includes two stages: the clustering ensemble member generation and the consistency function partition. In the former stage, one can produce multiple clustering solutions with various clustering algorithms after performing different transformations on the original data set. In the latter stage, some special consensus functions are exploited to fuse clustering solutions from ensemble members and get the final clustering result.

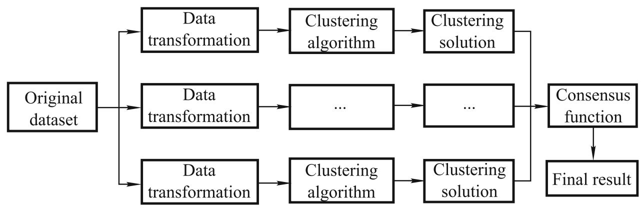


Fig. 11 The framework of clustering ensemble

There are extensive researches on clustering ensemble methods, and we can classify these existing works into the following types according to their emphasis during the clustering process. The first type of research works focuses on designing new algorithms for the clustering member generation process and the combination process. For example, Yang et al. [113] adopted the nearest neighbor method to fill the category information for missing samples and generate basic partitions with a good balance between the quality and the diversity for clustering ensemble. Wu et al. [114] proposed a clustering ensemble algorithm based on data subset, when compared with point-based clustering ensemble methods, experiment showed the method could significantly reduce the computational complexity when the data set is large. Franek and Jiang [115] formulated the complex clustering ensemble as the Euclidean median problem by mapping and clustering embeddings of ensembles in vector space, and the consensus clustering was obtained by utilizing an inverse transformation from the vector space to clustering space. Yu et al. [116] proposed a graph-based consensus clustering scheme to effectively discover the class information about gene expression data. In this method, random subspace and correlation clustering algorithms were combined to enhance diversities of clustering ensemble. In [117], Yu et al. presented a hybrid clustering ensemble framework, which was based on the random transformations in the sample dimensions and the feature dimensions, in this method, the clustering technique was adopted to assign different weights for clustering solution according to the confidence level, followed by the normalized cut algorithm as the consensus function to generate the final partition. Afterward, Yu et al. [118] proposed a triple spectral-clustering based consensus clustering framework which adopted spectral clustering along with normalized cut serving as the consensus function. To further enhance the robustness, stability and accuracy of clustering ensemble methods, Yu et al. [119] introduced fuzzy theory to hybrid fuzzy clustering ensemble algorithms. In this paper, four types of schemes were contained to differentiate samples from different cancers, the first two schemes generated basic ensemble partitions in the sample and attribute dimension, respectively, and the last two schemes combined the first two in a serial and a parallel way, respectively. Moreover, in [120], Yu et al. designed a noise immune clustering ensemble framework to address issues brought about by noisy data sets. This framework utilized affinity propagation and normalized cut techniques by adopting multiple distance functions to evade effect of noise. Ayad and Kamel [121] converted the voting problem into regression problem involving multi-response and multi-input variables, and randomized generation techniques were adopted for generating basic partitions, the proposed method utilized an information theoretic algorithm to obtain the consensus clustering and the number of clusters from learned ensemble representations with the help of bipartite matching and cumulative voting. Zhang et al. [122] investigated the adjusted rank index (ARI) and proposed two measurements based on ARI to calculate the consistency of clustering members for clustering ensemble. Fred and Jain [123] introduced the evidence accumulation clustering (EAC) concept to combine multiple clustering results, in which each partition was treated as independent evidence of data organization, and these learned partitions were combined via certain voting mechanism before adopting the hierarchical agglomerative clustering technique to obtain the final partition. To explore the scalability of EAC from the perspective of theoretical analysis to reduce the space complexity, Lourenco et al. [124] learned the compact representation of the co-association matrix which exploits the inherent sparseness and constructed clustering partitions via the split-and-merge strategy. In summary, this category of researches focuses on investigating delicate clustering ensemble algorithms from different perspectives rather than simply aligning clustering results obtained from traditional algorithms.

The second type of research works theoretically analyzes clustering ensemble model characteristics such as stability, diversity, convergence, etc. For example, Amasyali and Ersoy [125] investigated different factors that produced a great influence on performances of clustering ensemble algorithms. Those factors included clustering algorithms, the number of features, the size of ensemble models, and the fusion policy of clustering results and so on. Fern and Brodley [126] used the random projection for ensemble clustering and analyzed the impacts of qualities and diversities of individual clustering result on the final prediction. Kuncheva and Whitaker [127] explored the relationship between diversity and accuracy of ensemble model, followed by investigating the stability of clustering ensemble algorithms with respect to various initialization parameters and conditions [128]. Shi et al. [129] proposed a transfer CES (TCES) algorithm to make use of the relationship between quality and diversity in source dataset, and they proposed a transfer CE framework (TCE-TCES) based on TCES to obtain better clustering results. Topchy et al. [130] studied the convergence of clustering ensemble algorithms, which indicated that the consensus solutions could converge to a potential clustering structure as the number of integration partitions increased. Wang [131] proposed co-association (CA) tree method which used the hierarchical data structure to reduce the ensemble model complexity. Hore et al. [132] explored the feasibility of merging multiple clustering centroids that were obtained from data in a scalable framework. The purpose of this type of researches was to improve performances of clustering ensemble algorithms and provide theoretical supports.

The third type of researches focuses on the selection of clustering results from ensemble models. Fern and Lin [133] studied how to effectively select clustering results for ensemble based on diversities and qualities of clustering. In this work, they proposed three methods to combine diversity and quality: the first method combined them by a joint objective function, the second method separated clustering member into different groups and selected high-quality solutions in each group, the third method utilized points to represent the average quality and diversity of a pair of clustering solutions, followed by selecting solutions with a convex hall. Besides, Azimi and Fern [134] proposed an adaptive selection algorithm by using characteristics of data sets for clustering results selection. The method generated a subset of ensemble members based on the diversities between ensemble members and consensus partitions and then combined them to obtain the final output. Wang et al. [135] designed a clustering result selection strategy based on the rough set theory, in which significant attributes of data were used to find optimal subsets of clustering members. In [136], Yu et al. considered clustering solutions as sample features, and used weighting function to combine several features selection algorithms to select informative clustering solutions.

Additionally, Yu et al. [137] conducted researches on clustering structures for the clustering ensemble method. Specifically, they introduced the cluster structure concept and extracted the cluster structures from different data sets, in which the re-sampling technique and graph theory were utilized to construct a unified cluster structure. To solve the structural ensemble problem for heterogeneous data, Yu et al. [138] developed a cluster structure ensemble framework which was based on Gaussian Mixture models. In this method, representative clustering structures were selected and generalized to form a unified cluster structure. Yu et al. [139] also designed a distribution based distance function to quantify similarities between cluster structures and chose representative structures to generate the final result with distributional normalized hyperplane algorithm.

Clustering ensemble methods have been widely used to solve a diversity of real-world problems. In the data mining area, Yang and Jiang [140] presented a hidden Markov model (HMM) based clustering ensemble method to reduce impacts of initialization and model selections on performances for temporal data clustering, in which a hierarchical clustering refinement was utilized. Yang and Chen [141] adopted a hierarchical clustering ensemble algorithm which used HMM-based partitioning to help mine the intrinsic structure of data sets such as cluster numbers from temporal data. In the bioinformatics field, Yu and Wong [142] designed a special tumor discovery framework which adopted perturbation technique and cluster validity index to help explore the number of classes of tumor gene expression data. Besides, Yu et al. [143] presented a random double clustering-based fuzzy clustering ensemble framework for cancer discovery, in which the fuzzy extension model was contained to generate fuzzy metrics used for consensus partition. The clustering ensemble methods could be applied to other related research field, such as DNA microarray analysis [144], gene expression data analysis [145], image segmentation [146–149], significant region detection [150], biomedical text clustering [151], speaker recognition [152], internet security [153], filtering recommendation [154], etc.

In the above-mentioned work, we summarize challenges and potential research directions for clustering ensemble. Concerning traditional clustering ensemble algorithms, there is a lack of theoretical design principles for sample allocations for each clustering member. In addition, in the case where the prior information is not available, it is still a problem that how to determine the numbers of clustering members and final clustering members, which needs to be further explored. Moreover, because the clustering results of ensemble members need to be fused, the time complexity is extremely high, especially when dealing with high-dimensional data. In this setting, it is imperative to develop efficient algorithms to reduce the time complexity of clustering ensemble models. Besides, it is necessary to explore the technical combination between ensemble clustering methods and other methods such as semi-supervised mechanism, etc., which allow making full use of the prior knowledge in data sets.

5 Semi-supervised clustering ensemble

Semi-supervised clustering ensemble algorithms focus on utilizing prior knowledge such as cannot-link and must-link to instruct the clustering ensemble process. In other words, semi-supervised clustering ensemble can be treated as a technical combination of semi-supervised clustering and ensemble learning. Thus, it allows fusing advantages of both techniques to improve the accuracy and robustness of the model, compared with traditional clustering ensemble methods.

In the field of semi-supervised clustering ensemble, some works focus on optimizing the generation process and the selection process of clustering members. As shown in Fig. 12, Yu et al. [155] developed an incremental semi-supervised clustering ensemble framework (RS denotes random subspace, CS denotes constraint subset, and EECP is a constraint propagation method), in which the random subspace and the constraint propagation technique respectively helped to deal with high-dimensional data and incorporate the prior knowledge. Besides, this method designed an incremental clustering member selection process to effectively eliminate redundant members. Yu et al. [156] designed a random subspace based semi-supervised clustering ensemble scheme to fuse clustering solutions into a unified solution. In this method, transitive closures were introduced to expand constraints to obtain clustering solutions on different data sets, and the label propagation process is adopted to disseminate pairwise constraints. Wei et al. [157] proposed a hybrid semi-supervised clustering ensemble algorithm, in which the prior knowledge including class labels and pairwise constraints were utilized to generate basic clustering partitions. Based on these partitions, a metric function was designed to consider the spatial distribution for feature extractions, followed by integrating into a consensus function. Inspired by Chameleon [158] which performed hierarchical clustering via dynamic modeling, Xiao et al. [159] proposed a semi-supervised clustering ensemble model. We show part of the experiment results in Table 1, it shows that when compared with traditional clustering ensemble algorithm such as voting proposed by Zhou and Tang [160], CECH achieve comparable performance, and we can see that semi-supervision mechanism can further increase the accuracy of the model on UCI data sets, moreover, it is also worth noticing that the performance of semi-supervised ensemble method improves with the increase of the percentage of pairwise constraints, and outperforms voting method obviously. More details can be seen in [159].

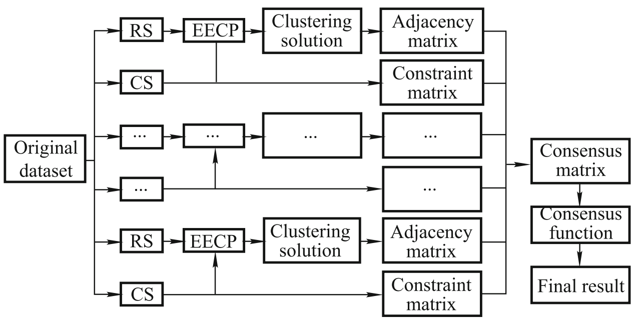


Fig. 12 The framework of incremental semi-supervised clustering ensemble

Table 1 Semi-supervised clustering ensemble compared with clustering ensemble

	Voting	CECH	SCECH0%	SCECH5%	SCECH10%
Iris	0.8853	0.8853	0.8847	0.9172	0.9412
Wine	0.9564	0.9548	0.9552	0.9583	9613
Glass	0.4714	0.4849	0.4862	0.5256	0.5418
Ionosphere	0.7177	0.7177	0.7175	0.7253	0.7295
Vehicle	0.4224	0.4456	0.4455	0.4455	0.4518
Diabetes	0.6656	0.6656	0.6681	0.6752	0.6911
Cmc	0.4319	0.4761	0.4760	0.4802	0.4898
Segment	0.5101	0.6903	0.6909	0.6915	0.6918
Sonar	0.5579	0.5556	0.5556	0.5623	0.5686

Some researchers studied the way to integrate clustering members into final prediction via some special voting mechanisms. Zhang et al. [161] presented collaborative training by using tri-training [162] as consensus function for semi-supervised clustering ensemble. Wang et al. [163] introduced the normal mutual information to semi-supervised clustering ensemble. Yu et al. [164] proposed an adaptive ensemble member weighting process to associate different weight values with different ensemble members. Yang et al. [165] implemented a parallel multi-ant colonies algorithm using MapReduce technique to improve the performance of semi-supervised clustering ensemble. In addition, related works such as Iqbal et al. [166] engaged supervision in clustering ensemble using two parameters for clustering consensus partition, where the first parameter described the compatibility between the dataset and the clustering algorithm and the second parameter provided the user feedback on the partitions. Chen et al. [167] conducted convergence analysis of semi-supervised clustering ensemble methods, which showed that the accuracy of semi-supervised clustering ensemble could converge with the increase of the number of basic clustering members. To reduce the difficulty of high dimensionality metric learning in clustering, Yan and Domeniconi [168] projected the data and the constraints in multiple subspaces and tried to learn the distance between data points in subspaces for ensemble. Apart from abovementioned researches, semi-supervised clustering ensemble has also been applied to other fields. For example, Mahmood et al. [169] adopted semi-supervised clustering ensemble methods to categorize the web videos and used the genetic algorithm(GA) to help iterate the clustering ensemble process for social media mining [170]. Additionally, Junaidi and Fink [171] utilized semi-supervised clustering ensemble for character labeling. In the medical field, Yu et al. [172] treated prior knowledge as constraints in clustering ensemble for cancer classification. Afterward, they designed a double-layer selection [173] which was applied to tumor discovery.

From the above-mentioned works, we observe that there are relatively few studies on semi-supervised clustering ensemble. Specifically, there are quite a few researches on semi-supervised hypotheses, such as smoothness, clustering density and manifold in semi-supervised clustering ensemble. Considering the aspect of algorithms, since existing semi-supervised clustering ensemble methods have the following disadvantages: failing to make full use of the constraint information and optimize the constraint selection, we can further explore how to effectively overcome the above shortcomings. Moreover, the semi-supervised mechanism can be refined by introducing unlabeled data from other sources. Thus, it sounds a promising research direction to capture the underlying relationships between multi-source data. Finally, it is worth to explore the refinement of semi-supervised mechanism by other machine learning methods (such as transfer learning, active learning, etc.) for clustering ensemble in certain rational formations.

6 New direction

As we can see, ensemble learning is a relatively mature machine learning issue, compared with recent machine learning hot spots like deep learning, reinforcement learning, transfer learning, etc. Previous sections show that ensemble learning is more of a frame ideology, so it becomes possible for ensemble learning to combine with other machine learning methods seamlessly. We investigate relevant works in recent years, and we find that there are a number of investigations which have proved that ensemble learning can successfully fuse with deep learning, reinforcement learning, etc., where the performance can be effectively improved with the introduction of ensemble mechanism. Relevant work such as Krogh and Vedelsby [174] defined the variation of output of ensemble networks averaged over unlabeled data as ambiguity, and then used the ambiguity in combination with cross-validation to estimate the ensemble generalization error, thus the reduction of the error helps to improve the performance. To optimize the determination process of deep classification model structure and the combination of multi-modal feature abstractions, Yin et al. [175] proposed multiple-fusion-layer based ensemble classifier of stacked auto-encoder (MESAE) for recognizing emotions, in which deep learning is used for guiding autoencoder ensemble. Moreover, based on the assumption that different convolutional neural network (CNN) architectures learn different levels of semantic representations, Kumar et al. [176] developed a new feature extractor by ensembling CNNs that were initialized on a large data set of natural images. Experiment showed that the ensemble of CNNs can extract features with a higher quality, compared with traditional CNNs. Liu et al. [177] applied ensemble of convolutional neural network models with different architectures for visual traffic surveillance systems.

As for the transfer learning issue, related works like [178] used ensemble method to combine outputs of various selective layer based transference conditions of deep learning model. Experiments show it can reduce the effect of negative feature transference on image recognition tasks. Nozza et al. [179] used ensemble methods to reduce the cross-domain generalization error of domain adaptation problem in sentiment classification tasks. Liu et al. [180] design an ensemble transfer learning framework which used AdaBoost to adjust the weights of the source data and target data, this method achieved good performance on UCI data sets when the training data are insufficient.

Relevant works about reinforcement learning like [181] combined multi-objective optimization and ensemble techniques to boost solving performance in reinforcement learning. Specifically, reward signals created by reward sharping were combined using ensemble method, thus the sample complexity in reinforcement learning could be reduced. To solve the problem that existing ensemble algorithms in reinforcement learning are not compatible with nonlinearly parameterized value functions, Chen et al. [182] proposed an ensemble network architecture for deep reinforcement learning, in which the temporal ensemble stabilized the training process by reducing the variance of target approximation error and the ensemble of target values reduced the overestimate. The method got good performance in OpenAI Gym environment.

With the above-mentioned researches, we find that ensemble learning is more than a specific algorithm, which makes it easy to combine ensemble method with other machine learning algorithms, and we also give a brief introduction for the combination of ensemble learning with other machine learning techniques in Fig. 13 according to Sections 1–6, for providing readers some references on research issues.

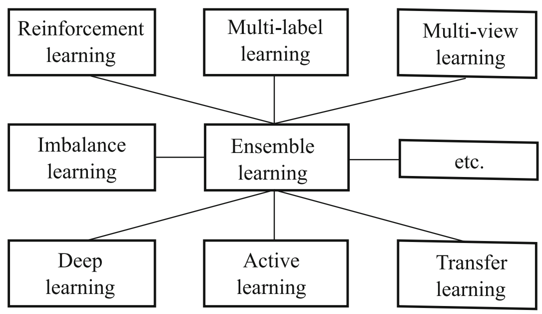


Fig. 13 The combination of ensemble learning with other machine learning issues

7 Summary

In this paper, we investigated the research progress in various branches of ensemble learning, and categorized ensemble learning methods from different perspectives. Besides, we introduced challenges and feasible research directions for ensemble learning. However, there are still more efforts to make to further improve performances of ensemble models, especially in the case where data contains complex patterns. We consider that readers would have a preliminary understanding of these existing ensemble learning approaches, and conduct ensemble learning from different aspects through our paper. We expect to throw some light on this field by providing some suggestions for future ensemble learning directions.

Acknowledgements The authors are grateful for the constructive advice received from the anonymous reviewers of this paper. The work described in this paper was partially funded by grants from the National Natural Science Foundation of China (Grant Nos. 61722205, 61751205, 61572199, 61502174, 61872148, and U1611461), the grant from the key research and development program of Guangdong province of China (2018B010107002), the grants from Science and Technology Planning Project of Guangdong Province, China (2016A050503015, 2017A030313355), and the grant from the Guangzhou science and technology planning project (201704030051).

References

1. Zhou Z H. Ensemble Methods: Foundations and Algorithms. Chapman and Hall/CRC, 2012
2. Dasarathy B V, Sheela B V. A composite classifier system design: concepts and methodology. Proceedings of the IEEE, 1979, 67(5): 708–713
3. Kearns M. Learning boolean formulae or finite automata is as hard as factoring. Technical Report TR-14-88 Harvard University Aikem Computation Laboratory, 1988
4. Schapire, Robert E. The strength of weak learnability. Machine Learning, 1990, 5(2): 197–227
5. Breiman L. Bagging predictors. Machine Learning, 1996, 24(2): 123–140
6. Hastie T, Rosset S, Zhu J, Zou H. Multi-class adaboost. Statistics and its Interface, 2009, 2(3): 349–360
7. Breiman L. Random forests. Machine Learning, 2001, 45(1): 5–32
8. Ho T K. Random decision forests. In: Proceedings of the 3rd International Conference on Document Analysis and Recognition. 1995, 278–282
9. Friedman J H. Stochastic gradient boosting. Computational Statistics and Data Analysis, 2002, 38(4): 367–378
10. Garcia-Pedrajas N. Constructing ensembles of classifiers by means of weighted instance selection. IEEE Transactions on Neural Networks, 2009, 20(2): 258–277
11. Garcia-Pedrajas N, Maudes-Raedo J, Garcia-Osorio C, Rodriguez-Díez J J, Linden D E, Johnston S J. Supervised subspace projections for constructing ensembles of classifiers. Information Sciences, 2012, 193(11): 1–21
12. Kuncheva L I, Rodriguez J J, Plumpton C O, Linden D E, Johnston S J. Random subspace ensembles for FMRI classification. IEEE Transactions on Medical Imaging, 2010, 29(2): 531–542
13. Ye Y, Wu Q, Huang J Z, Ng M K, Li X. Stratified sampling for feature subspace selection in random forests for high dimensional data. Pattern Recognition, 2013, 46(3): 769–787
14. Bryll R, Gutierrez-Osuna R, Quek F. Attribute bagging: improving accuracy of classifier ensembles by using random feature subsets. Pattern Recognition, 2003, 36(6): 1291–1302
15. Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: Proceedings of the 11th Annual Conference on Computational Learning Theory. 1998, 92–100
16. Wang J, Luo S W, Zeng X H. A random subspace method for co-training. In: Proceedings of 2008 IEEE International Joint Conference on Neural Networks. 2008, 195–200
17. Yaslan Y, Cataltepe Z. Co-training with relevant random subspaces. Neurocomputing, 2010, 73(10–12): 1652–1661
18. Zhang J, Zhang D. A novel ensemble construction method for multi-view data using random cross-view correlation between within-class examples. Pattern Recognition, 2011, 44(6): 1162–1171
19. Guo Y, Jiao L, Wang S, Liu F, Rong K, Xiong T. A novel dynamic rough subspace based selective ensemble. Pattern Recognition, 2015, 48(5): 1638–1652
20. Windeatt T, Duangsoithong R, Smith R. Embedded feature ranking for ensemble MLP classifiers. IEEE Transactions on Neural Networks, 2011, 22(6): 988–994
21. Rodriguez J J, Kuncheva L I, Alonso C J. Rotation forest: a new classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(10): 1619–1630
22. Takemura A, Shimizu A, Hamamoto K. Discrimination of breast tumors in ultrasonic images using an ensemble classifier based on the AdaBoost algorithm with feature selection. IEEE Transactions on Medical Imaging, 2010, 29(3): 598–609
23. Amasyali M F, Ersoy O K. Classifier ensembles with the extended space forest. IEEE Transactions on Knowledge and Data Engineering, 2013, 26(3): 549–562
24. Polikar R, Depasquale J, Mohammed H S, Brown G, Kuncheva L I. Learn++.MF: a random subspace approach for the missing feature problem. Pattern Recognition, 2010, 43(11): 3817–3832
25. Nanni L, Lumini A. Evolved feature weighting for random subspace classifier. IEEE Transactions on Neural Networks, 2008, 19(2): 363–366
26. Kennedy J, Eberhart R C. A discrete binary version of the particle swarm optimization algorithm. Computational Cybernatics and Simulation, 1997, 5(1): 4104–4108
27. Zhou Z H, Tang W. Selective ensemble of decision trees. In: Proceedings of International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing. 2003, 476–483
28. Diao R, Chao F, Peng T, Snooke N, Shen Q. Feature selection inspired classifier ensemble reduction. IEEE Transactions on Cybernetics, 2014, 44(8): 1259–1268
29. Yu Z, Wang D, You J, Wong H S, Wu S, Zhang J, Han G. Progressive subspace ensemble learning. Pattern Recognition, 2016, 60: 692–705
30. Yu Z, Wang D, Zhao Z, Chen C P, You J, Wong H S, Zhang J. Hybrid incremental ensemble learning for noisy real-world data classification. IEEE Transactions on Cybernetics, 2017, 99: 1–14
31. Dos Santos E M, Sabourin R, Maupin P. A dynamic overproduce-and-choose strategy for the selection of classifier ensembles. Pattern Recognition, 2008, 41(10): 2993–3009
32. Hernández-Lobato D, Martínez-Muñoz G, Suárez A. Statistical instance-based pruning in ensembles of independent classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 364–369
33. Martínez-Muñoz G, Hernández-Lobato D, Suárez A. An analysis of ensemble pruning techniques based on ordered aggregation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 245–259
34. De Stefano C, Folino G, Fontanella F, Di Freca A S. Using bayesian networks for selecting classifiers in GP ensembles. Information Sciences, 2014, 258: 200–216
35. Rahman A, Verma B. Novel layered clustering-based approach for generating ensemble of classifiers. IEEE Transactions on Neural Networks, 2011, 22(5): 781–792
36. Verma B, Rahman A. Cluster-oriented ensemble classifier: impact

of multicluster characterization on ensemble classifier learning. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(4): 605–618

37. Zhang L, Suganthan P N. Oblique decision tree ensemble via multisurface proximal support vector machine. IEEE Transactions on Cybernetics, 2015, 45(10): 2165–2176
38. Tan P J, Dowe D L. Decision forests with oblique decision trees. In: Proceedings of Mexican International Conference on Artificial Intelligence. 2006, 593–603
39. Zhou Z H, Wu J, Tang W. Ensembling neural networks: many could be better than all. Artificial Intelligence, 2002, 137(1–2): 239–263
40. Yu Z, Chen H, Liu J, You J, Leung H, Han G. Hybrid k-nearest neighbor classifier. IEEE Transactions on Cybernetics, 2016, 46(6): 1263–1275
41. Li H, Wen G, Yu Z, Zhou T. Random subspace evidence classifier. Neurocomputing, 2013, 110(13): 62–69
42. Hernández-Lobato D, Martínez-Muñoz G, Suárez A. How large should ensembles of classifiers be? Pattern Recognition, 2013, 46(5): 1323–1336
43. Wang X Z, Xing H J, Li Y, Hua Q, Dong C R, Pedrycz W. A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning. IEEE Transactions on Fuzzy Systems, 2015, 23(5): 1638–1654
44. Kuncheva L I. A bound on kappa-error diagrams for analysis of classifier ensembles. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(3): 494–501
45. Gao W, Zhou Z H. Approximation stability and boosting. In: Proceedings of International Conference on Algorithmic Learning Theory. 2010, 59–73
46. Yin X C, Huang K, Hao H W, Iqbal K, Wang Z B. A novel classifier ensemble method with sparsity and diversity. Neurocomputing, 2014, 134: 214–221
47. Zhang L, Suganthan P N. Random forests with ensemble of feature spaces. Pattern Recognition, 2014, 47(10): 3429–3437
48. Li N, Yu Y, Zhou Z H. Diversity regularized ensemble pruning. In: Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2012, 330–345
49. Zhang D, Chen S, Zhou Z H, Yang Q. Constraint projections for ensemble learning. In: Proceedings of the 23rd National Conference on Artifical Intelligence-Volume 2. 2008, 758–763
50. Zhou Z H, Li N. Multi-information ensemble diversity. In: Proceedings of International Workshop on Multiple Classifier Systems. 2010, 134–144
51. Sun T, Zhou Z H. Structural diversity for decision tree ensemble learning. Frontiers of Computer Science, 2018, 12(3): 560–570
52. Mao S, Jiao L, Xiong L, Gou S, Chen B, Yeung S K. Weighted classifier ensemble based on quadratic form. Pattern Recognition, 2015, 48(5): 1688–1706
53. Yu Z, Wang Z, You J, Zhang J, Liu J, Wong H S, Han G. A new kind of nonparametric test for statistical comparison of multiple classifiers over multiple datasets. IEEE Transactions on Cybernetics, 2017, 47(12): 4418–4431
54. Kim K J, Cho S B. An evolutionary algorithm approach to optimal ensemble classifiers for DNA microarray data analysis. IEEE Transactions on Evolutionary Computation, 2008, 12(3): 377–388
55. Qian C, Yu Y, Zhou Z H. Pareto ensemble pruning. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence. 2015
56. Zhou Z H, Feng J. Deep forest: towards an alternative to deep neural networks. 2017, arXiv preprint arXiv:1702.08835
57. Feng J, Zhou Z H. AutoEncoder by forest. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. 2018
58. Zhang Y L, Zhou J, Zheng W, Feng J, Li L, Liu Z, Zhou Z H. Distributed deep forest and its application to automatic detection of cash-out fraud. 2018, arXiv preprint arXiv:1805.04234
59. Feng J, Yu Y, Zhou Z H. Multi-layered gradient boosting decision trees. In: Proceedings of Advances in Neural Information Processing Systems. 2018, 3555–3565
60. Pang M, Ting K M, Zhao P, Zhou Z H. Improving deep forest by confidence screening. In: Proceedings of the 18th IEEE International Conference on Data Mining. 2018, 1194–1199
61. Yu Z, Li L, Liu J, Han G. Hybrid adaptive classifier ensemble. IEEE Transactions on Cybernetics, 2015, 45(2): 177–190
62. Zhou Z H, Zhang M L. Solving multi-instance problems with classifier ensemble based on constructive clustering. Knowledge and Information Systems, 2007, 11(2): 155–170
63. Zhu X, Zhang P, Lin X, Shi Y. Active learning from stream data using optimal weight classifier ensemble. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2010, 40(6): 1607–1621
64. Brzezinski D, Stefanowski J. Reacting to different types of concept drift: the accuracy updated ensemble algorithm. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(1): 81–94
65. Muhlbaier M D, Topalis A, Polikar R. Learn^{++}.NC: combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes. IEEE Transactions on Neural Networks, 2009, 20(1): 152–168
66. Xiao J, He C, Jiang X, Liu D. A dynamic classifier ensemble selection approach for noise data. Information Sciences, 2010, 180(18): 3402–3421
67. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging, boosting, and hybrid-based approaches. IEEE Transactions on Systems Man and Cybernetics Part C, 2012, 42(4): 463–484
68. Liu X Y, Wu J, Zhou Z H. Exploratory under-sampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2009, 39(2): 539–550
69. Sun B, Chen H, Wang J, Xie H. Evolutionary under-sampling based bagging ensemble method for imbalanced data classification. Frontiers of Computer Science, 2018, 12(2): 331–350
70. Li Q, Li G, Niu W, Cao Y, Chang L, Tan J, Guo L. Boosting imbalanced data learning with wiener process oversampling. Frontiers of Computer Science, 2017, 11(5): 836–851
71. Abawajy J H, Kelarev A, Chowdhury M. Large iterative multitier ensemble classifiers for security of big data. IEEE Transactions on Emerging Topics in Computing, 2014, 2(3): 352–363
72. Li N, Zhou Z H. Selective ensemble of classifier chains. In: Proceedings of International Workshop on Multiple Classifier Systems. 2013, 146–156

73. Li N, Jiang Y, Zhou Z H. Multi-label selective ensemble. In: Proceedings of International Workshop on Multiple Classifier Systems. 2015, 76–88
74. Yu Z, Deng Z, Wong H S, Tan L. Identifying protein-kinase-specific phosphorylation sites based on the Bagging-AdaBoost ensemble approach. IEEE Transactions on Nanobioscience, 2010, 9(2): 132–143
75. Yu D J, Hu J, Yang J, Shen H B, Tang J, Yang J Y. Designing template-free predictor for targeting protein-ligand binding sites with classifier ensemble and spatial clustering. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013, 10(4): 994–1008
76. Yu G, Rangwala H, Domeniconi C, Zhang G, Yu Z. Protein function prediction using multilabel ensemble classification. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013, 10(4): 1
77. Daliri M R. Combining extreme learning machines using support vector machines for breast tissue classification. Computer Methods in Biomechanics and Biomedical Engineering, 2015, 18(2): 185–191
78. Oliveira L, Nunes U, Peixoto P. On exploration of classifier ensemble synergism in pedestrian detection. IEEE Transactions on Intelligent Transportation Systems, 2010, 11(1): 16–27
79. Xu Y, Cao X, Qiao H. An efficient tree classifier ensemble-based approach for pedestrian detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2011, 41(1): 107–117
80. Zhang B. Reliable classification of vehicle types based on cascade classifier ensembles. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(1): 322–332
81. Sun S, Zhang C. The selective random subspace predictor for traffic flow forecasting. IEEE Transactions on Intelligent Transportation Systems, 2007, 8(2): 367–373
82. Su Y, Shan S, Chen X, Gao W. Hierarchical ensemble of global and local classifiers for face recognition. IEEE Transactions on Image Processing, 2009, 18(8): 1885–1896
83. Zhang P, Bui T D, Suen C Y. A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. Pattern Recognition, 2007, 40(12): 3415–3429
84. Xu X S, Xue X, Zhou Z H. Ensemble multi-instance multi-label learning approach for video annotation task. In: Proceedings of the 19th ACM International Conference on Multimedia. 2011, 1153–1156
85. Hautamaki V, Kinnunen T, Sedlák F, Lee K A, Ma B, Li H. Sparse classifier fusion for speaker verification. IEEE Transactions on Audio Speech and Language Processing, 2013, 21(8): 1622–1631
86. Guan Y, Li C T, Roli F. On reducing the effect of covariate factors in gait recognition: a classifier ensemble method. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(7): 1521–1528
87. Tao D, Tang X, Li X, Wu X. Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(7): 1088–1099
88. Hu W, Hu W, Maybank S. AdaBoost-based algorithm for network intrusion detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2008, 38(2): 577–583
89. Zhang P, Zhu X, Shi Y, Guo L, Wu X. Robust ensemble learning for mining noisy data streams. Decision Support Systems, 2011, 50(2): 469–479
90. Yu L, Wang S, Lai K K. Developing an SVM-based ensemble learning system for customer risk identification collaborating with customer relationship management. Frontiers of Computer Science, 2010, 4(2): 196–203
91. Fersini E, Messina E, Pozzi F A. Sentiment analysis: Bayesian ensemble learning. Decision Support Systems, 2014, 68: 26–38
92. Yu G, Zhang G, Yu Z, Domeniconi C, You J, Han G. Semi-supervised ensemble classification in subspaces. Applied Soft Computing, 2012, 12(5): 1511–1522
93. Yu Z, Zhang Y, Chen C L P, You J, Wong H S, Dai D, Wu S, Zhang J. Multiobjective semisupervised classifier ensemble. IEEE Transactions on Cybernetics, 2019, 49(6): 2280–2293
94. Gharroudi O, Elghazel H, Aussem A. A semi-supervised ensemble approach for multi-label learning. In: Proceedings of the 16th IEEE International Conference on Data Mining Workshops (ICDMW). 2016, 1197–1204
95. Lu X, Zhang J, Li T, Zhang Y. Hyperspectral image classification based on semi-supervised rotation forest. Remote Sensing, 2017, 9(9): 924
96. Wang S, Chen K. Ensemble learning with active data selection for semi-supervised pattern classification. In: Proceedings of 2007 International Joint Conference on Neural Networks. 2007, 355–360
97. Soares R G F, Chen H, Yao X. A cluster-based semi-supervised ensemble for multiclass classification. IEEE Transactions on Emerging Topics in Computational Intelligence, 2017, 1(6): 408–420
98. Woo H, Park C H. Semi-supervised ensemble learning using label propagation. In: Proceedings of the 12th IEEE International Conference on Computer and Information Technology. 2012, 421–426
99. Zhang M L, Zhou Z H. Exploiting unlabeled data to enhance ensemble diversity. Data Mining and Knowledge Discovery, 2013, 26(1): 98–129
100. Alves M, Bazzan A L C, Recamonde-Mendoza M. Social-training: ensemble learning with voting aggregation for semi-supervised classification tasks. In: Proceedings of 2017 Brazilian Conference on Intelligent Systems (BRACIS). 2017, 7–12
101. Yu Z, Lu Y, Zhang J, You J, Wong H S, Wang Y, Han G. Progressive semi-supervised learning of multiple classifiers. IEEE Transactions on Cybernetics, 2018, 48(2): 689–702
102. Hosseini M J, Gholipour A, Beigy H. An ensemble of cluster-based classifiers for semi-supervised classification of non-stationary data streams. Knowledge and Information Systems, 2016, 46(3): 567–597
103. Wang Y, Li T. Improving semi-supervised co-forest algorithm in evolving data streams. Applied Intelligence, 2018, 48(10): 3248–3262
104. Yu Z, Zhang Y, You J, Chen C P, Wong H S, Han G, Zhang J. Adaptive semi-supervised classifier ensemble for high dimensional data classification. IEEE Transactions on Cybernetics, 2019, 49(2): 366–379
105. Li M, Zhou Z H. Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 2007, 37(6): 1088–1098
106. Guz U, Cuendet S, Hakkani-Tur D, Tur G. Multi-view semi-supervised learning for dialog act segmentation of speech. IEEE Transactions on Audio Speech and Language Processing, 2010,

18(2): 320–329

107. Shi L, Ma X, Xi L, Duan Q, Zhao J. Rough set and ensemble learning based semi-supervised algorithm for text classification. Expert Systems with Applications, 2011, 38(5): 6300–6306
108. Abdelgayed T S, Morsi W G, Sidhu T S. Fault detection and classification based on co-training of semi-supervised machine learning. IEEE Transactions on Industrial Electronics, 2018, 65(2): 1595–1605
109. Saydali S, Parvin H, Safaei A A. Classifier ensemble by semi-supervised learning: local aggregation methodology. In: Proceedings of International Doctoral Workshop on Mathematical and Engineering Methods in Computer Science. 2015, 119–132
110. Shao W, Tian X. Semi-supervised selective ensemble learning based on distance to model for nonlinear soft sensor development. Neurocomputing, 2017, 222: 91–104
111. Ahmed I, Ali R, Guan D, Lee Y K, Lee S, Chung T. Semi-supervised learning using frequent itemset and ensemble learning for SMS classification. Expert Systems with Applications, 2015, 42(3): 1065–1073
112. Strehl A, Ghosh J. Cluster ensembles: a knowledge reuse framework for combining partitionings. Journal of Machine Learning Research, 2002, 3(3): 583–617
113. Yang F, Li X, Li Q, Li T. Exploring the diversity in cluster ensemble generation: random sampling and random projection. Expert Systems with Applications, 2014, 41(10): 4844–4866
114. Wu O, Hu W, Maybank S J, Zhu M, Li B. Efficient clustering aggregation based on data fragments. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012, 42(3): 913–926
115. Franek L, Jiang X. Ensemble clustering by means of clustering embedding in vector spaces. Pattern Recognition, 2014, 47(2): 833–842
116. Yu Z, Wong H S, Wang H. Graph-based consensus clustering for class discovery from gene expression data. Bioinformatics, 2007, 23(21): 2888–2896
117. Yu Z, Wong H S, You J, Yu G, Han G. Hybrid cluster ensemble framework based on the random combination of data transformation operators. Pattern Recognition, 2012, 45(5): 1826–1837
118. Yu Z, Li L, You J, Wong H S, Han G. SC^3: triple spectral clustering-based consensus clustering framework for class discovery from cancer gene expression profiles. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2012, 9(6): 1751–1765
119. Yu Z, Chen H, You J, Han G, Li L. Hybrid fuzzy cluster ensemble framework for tumor clustering from biomolecular data. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2013, 10(3): 657–670
120. Yu Z, Li L, Liu J, Zhang J, Han G. Adaptive noise immune cluster ensemble using affinity propagation. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(12): 3176–3189
121. Ayad H G, Kamel M S. On voting-based consensus of cluster ensembles. Pattern Recognition, 2010, 43(5): 1943–1953
122. Zhang S, Wong H S, Shen Y. Generalized adjusted rand indices for cluster ensembles. Pattern Recognition, 2012, 45(6): 2214–2226
123. Fred A L N, Jain A K. Combining multiple clusterings using evidence accumulation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(6): 835–850
124. Lourenco A, Fred A L N, Jain A K. On the scalability of evidence accumulation clustering. In: Proceedings of the 20th International Conference on Pattern Recognition. 2010, 782–785
125. Amasyali M F, Ersoy O. The performance factors of clustering ensembles. In: Proceedings of the 16th IEEE Signal Processing, Communication and Applications Conference. 2008, 1–4
126. Fern X Z, Brodley C E. Random projection for high dimensional data clustering: a cluster ensemble approach. In: Proceedings of the 20th International Conference on Machine Learning (ICML-03). 2003, 186–193
127. Kuncheva L I, Whitaker C J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Machine Learning, 2003, 51(2): 181–207
128. Kuncheva L I, Vetrov D P. Evaluation of stability of k-means cluster ensembles with respect to random initialization. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(11): 1798–1808
129. Shi Y , Yu Z , Chen C L P, You J, Wong H S, Wang Y D, Zhang J. Transfer clustering ensemble selection. IEEE Transactions on Cybernetics, 2018, PP(99): 1–14
130. Topchy A P, Law M H C, Jain A K, Fred A L. Analysis of consensus partition in cluster ensemble. In: Proceedings of the 4th IEEE International Conference on Data Mining (ICDM'04). 2004, 225–232
131. Wang T. CA-tree: a hierarchical structure for efficient and scalable coassociation-based cluster ensembles. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2011, 41(3): 686–698
132. Hore P, Hall L O, Goldgof D B. A scalable framework for cluster ensembles. Pattern Recognition, 2009, 42(5): 676–688
133. Fern X Z, Lin W. Cluster ensemble selection. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2008, 1(3): 128–141
134. Azimi J, Fern X. Adaptive cluster ensemble selection. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence. 2009, 992–997
135. Wang X, Han D, Han C. Rough set based cluster ensemble selection. In: Proceedings of the 16th International Conference on Information Fusion. 2013, 438–444
136. Yu Z, Li L, Gao Y, You J, Liu J, Wong H S, Han G. Hybrid clustering solution selection strategy. Pattern Recognition, 2014, 47(10): 3362–3375
137. Yu Z, You J, Wong H S, Han G. From cluster ensemble to structure ensemble. Information Sciences, 2012, 198: 81–99
138. Yu Z, Li L, Wong H S, You J, Han G, Gao Y, Yu G. Probabilistic cluster structure ensemble. Information Sciences, 2014, 267(5): 16–34
139. Yu Z, Zhu X, Wong H S, You J, Zhang J, Han G. Distribution-based cluster structure selection. IEEE Transactions on Cybernetics, 2017, 47(11): 3554–3567
140. Yang Y, Jiang J. HMM-based hybrid meta-clustering ensemble for temporal data. Knowledge-Based Systems, 2014, 56: 299–310
141. Yang Y, Chen K. Temporal data clustering via weighted clustering ensemble with different representations. IEEE Transactions on Knowledge and Data Engineering, 2010, 23(2): 307–320
142. Yu Z, Wong H S. Class discovery from gene expression data based on perturbation and cluster ensemble. IEEE Transactions on Nanobioscience, 2009, 8(2): 147–160

143. Yu Z, Chen H, You J, Liu J, Wong H S, Han G, Li L. Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2015, 12(4): 887–901
144. Avogadri R, Valentini G. Fuzzy ensemble clustering based on random projections for DNA microarray data analysis. Artificial Intelligence in Medicine, 2009, 45(2): 173–183
145. Mimaroglu S, Aksehirli E. DICLENS: divisive clustering ensemble with automatic cluster number. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2012, 9(2): 408–420
146. Alush A, Goldberger J. Ensemble segmentation using efficient integer linear programming. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 1966–1977
147. Li H, Meng F, Wu Q, Luo B. Unsupervised multiclass region cosegmentation via ensemble clustering and energy minimization. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(5): 789–801
148. Zhang X, Jiao L, Liu F, Bo L, Gong M. Spectral clustering ensemble applied to SAR image segmentation. IEEE Transactions on Geoscience and Remote Sensing, 2008, 46(7): 2126–2136
149. Jia J, Liu B, Jiao L. Soft spectral clustering ensemble applied to image segmentation. Frontiers of Computer Science, 2011, 5(1): 66–78
150. Rafiee G, Dlay S S, Woo W L. Region-of-interest extraction in low depth of field images using ensemble clustering and difference of Gaussian approaches. Pattern Recognition, 2013, 46(10): 2685–2699
151. Huang X, Zheng X, Yuan W, Wang F, Zhu S. Enhanced clustering of biomedical documents using ensemble non-negative matrix factorization. Information Sciences, 2011, 181(11): 2293–2302
152. Bassiou N, Moschou V, Kotropoulos C. Speaker diarization exploiting the eigengap criterion and cluster ensembles. IEEE Transactions on Audio Speech and Language Processing, 2010, 18(8): 2134–2144
153. Zhuang W, Ye Y, Chen Y, Li T. Ensemble clustering for internet security applications. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 2012, 42(6): 1784–1796
154. Tsai C F, Hung C. Cluster ensembles in collaborative filtering recommendation. Applied Soft Computing, 2012, 12(4): 1417–1425
155. Yu Z, Luo P, You J, Wong H S, Leung H, Wu S, Zhang J, Han G. Incremental semi-supervised clustering ensemble for high dimensional data clustering. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(3): 701–714
156. Yu Z, Kuang Z, Liu J, Chen H, Zhang J, You J, Wong H S, Han G. Adaptive ensembling of semi-supervised clustering solutions. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(8): 1577–1590
157. Wei S, Li Z, Zhang C. Combined constraint-based with metric-based in semi-supervised clustering ensemble. International Journal of Machine Learning and Cybernetics, 2018, 9(7): 1085–1100
158. Karypis G, Han E H S, Kumar V. Chameleon: hierarchical clustering using dynamic modeling. Computer, 1999, 32(8): 68–75
159. Xiao W, Yang Y, Wang H, Li T, Xing H. Semi-supervised hierarchical clustering ensemble and its application. Neurocomputing, 2016, 173: 1362–1376
160. Zhou Z H, Tang W. Clusterer ensemble. Knowledge-Based Systems, 2006, 19(1): 77–83
161. Zhang J, Yang Y, Wang H, Mahmood A, Huang F. Semi-supervised clustering ensemble based on collaborative training. In: Proceedings of International Conference on Rough Sets and Knowledge Technology. 2012, 450–455
162. Zhou Z H, Li M. Tri-training: exploiting unlabeled data using three classifiers. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(11): 1529–1541
163. Wang H, Yang D, Qi J. Semi-supervised cluster ensemble based on normal mutual information. Energy Procedia, 2011, 13: 1673–1677
164. Yu Z, Luo P, Liu J, Wong H S, You J, Han G, Zhang J. Semi-supervised ensemble clustering based on selected constraint projection. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(12): 2394–2407
165. Yang Y, Teng F, Li T, Wang H, Zhang Q. Parallel semi-supervised multi-ant colonies clustering ensemble based on mapreduce methodology. IEEE Transactions on Cloud Computing, 2018, 6(3): 857–867
166. Iqbal A M, Moh'D A, Khan Z. Semi-supervised clustering ensemble by voting. Computer Science, 2012, 2(9): 33–40
167. Chen D, Yang Y, Wang H, Mahmood A. Convergence analysis of semi-supervised clustering ensemble. In: Proceedings of International Conference on Information Science and Technology. 2014, 783–788
168. Yan B, Domeniconi C. Subspace metric ensembles for semi-supervised clustering of high dimensional data. In: Proceedings of European Conference on Machine Learning. 2006, 509–520
169. Mahmood A, Li T, Yang Y, Wang H, Afzal M. Semi-supervised clustering ensemble for Web video categorization. In: Proceedings of International Workshop on Multiple Classifier Systems. 2013, 190–200
170. Mahmood A, Li T, Yang Y, Wang H, Afzal M. Semi-supervised evolutionary ensembles for web video categorization. Knowledge-Based Systems, 2015, 76: 53–66
171. Junaidi A, Fink G A. A semi-supervised ensemble learning approach for character labeling with minimal human effort. In: Proceedings of 2011 International Conference on Document Analysis and Recognition. 2011, 259–263
172. Yu Z, Wongb H S, You J, Yang Q, Liao H. Knowledge based cluster ensemble for cancer discovery from biomolecular data. IEEE Transactions on Nanobioscience, 2011, 10(2): 76–85
173. Yu Z, Chen H, You J, Wong H S, Liu J, Li L, Han G. Double selection based semi-supervised clustering ensemble for tumor clustering from gene expression profiles. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2014, 11(4): 727–740
174. Krogh A, Vedelsby J. Neural network ensembles, cross validation and active learning. In: Proceedings of the 7th International Conference on Neural Information Processing Systems. 1994, 231–238
175. Yin Z, Zhao M, Wang Y, Yang J, Zhang J. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. Computer Methods and Programs in Biomedicine, 2017, 140: 93–110
176. Kumar A, Kim J, Lyndon D, Fulham M, Feng D. An ensemble of fine-tuned convolutional neural networks for medical image classification. IEEE Journal of Biomedical and Health Informatics, 2017, 21(1): 31–40
177. Liu W, Zhang M, Luo Z, Cai Y. An ensemble deep learning method for vehicle type classification on visual traffic surveillance sensors.

IEEE Access, 2017, 5: 24417–24425

178. Kandaswamy C, Silva L M, Alexandre L A, Santos J M. Deep transfer learning ensemble for classification. In: Proceedings of International Work-Conference on Artificial Neural Networks. 2015, 335–348

179. Nozza D, Fersini E, Messina E. Deep learning and ensemble methods for domain adaptation. In: Proceedings of the 28th IEEE International Conference on Tools with Artificial Intelligence (ICTAI). 2016, 184–189

180. Liu X, Liu Z, Wang G, Cai Z, Zhang H. Ensemble transfer learning algorithm. IEEE Access, 2018, 6: 2389–2396

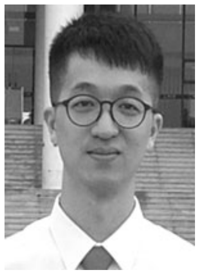
181. Brys T, Harutyunyan A, Vrancx P, Nowé A, Taylor M E. Multi-objectivization and ensembles of shapings in reinforcement learning. Neurocomputing, 2017, 263: 48–59

182. Chen X L, Cao L, Li C X, Xu Z X, Lai J. Ensemble network architecture for deep reinforcement learning. Mathematical Problems in Engineering, 2018, 2018: 1–6

Xibin Dong is a Master candidate in the School of Computer Science and Engineering in South China University of Technology, China. His research interests include machine learning, data mining. He is mainly working on the imbalance learning.

Zhiwen Yu is a professor in School of Computer Science and Engineering, South China University of Technology, China. He is a distinguishable member of CCF (China Computer Federation), a senior member of IEEE and ACM, and the vice chair of ACM Guangzhou chapter. He is an associate editor of IEEE Transactions on Systems, Man, and Cybernetics: Systems. Dr. Yu obtained the PhD degree from City University of Hong Kong, China in 2008. The research areas of Dr. Yu focus on artificial intelligence, data mining, machine learning and pattern recognition. Until now, Dr. Yu has been published more than 130 referred journal papers and international conference papers, including more than 30 IEEE Transactions papers.

Wenming Cao received MS degree from the School of Automation, Huazhong University of Science and Technology (HUST), China in 2015. He received PhD degree at the Department of Computer Science, City University of Hong Kong, China. His research interests include data mining and machine learning.



Yifan Shi is a Master candidate in the School of Computer Science and Engineering in South China University of Technology, China. His research interests include machine learning and data mining. He is mainly working on ensemble clustering.

Qianli Ma received the PhD degree in computer science from the South China University of Technology, China in 2008. He is an associate professor with the School of Computer Science and Engineering, South China University of Technology, China. From 2016 to 2017, he was a Visiting Scholar with the University of California at San Diego, USA. His current research interests include machine-learning algorithms, data-mining methodologies, and time-series modeling and their applications.