



Nature-inspired framework of ensemble learning for collaborative classification in granular computing context

Han Liu¹ · Mihaela Cocea²

Received: 13 June 2018 / Accepted: 25 July 2018 / Published online: 30 July 2018
© The Author(s) 2018

Abstract

Due to the vast and rapid increase in the size of data, machine learning has become an increasingly popular approach of data classification, which can be done by training a single classifier or a group of classifiers. A single classifier is typically learned by using a standard algorithm, such as C4.5. Due to the fact that each of the standard learning algorithms has its own advantages and disadvantages, ensemble learning, such as Bagging, has been increasingly used to learn a group of classifiers for collaborative classification, thus compensating for the disadvantages of individual classifiers. In particular, a group of base classifiers need to be learned in the training stage, and then some or all of the base classifiers are employed for classifying unseen instances in the testing stage. In this paper, we address two critical points that can impact the classification accuracy, in order to overcome the limitations of the Bagging approach. Firstly, it is important to judge effectively which base classifiers qualify to get employed for classifying test instances. Secondly, the final classification needs to be done by combining the outputs of the base classifiers, i.e. voting, which indicates that the strategy of voting can impact greatly on whether a test instance is classified correctly. In order to address the above points, we propose a nature-inspired approach of ensemble learning to improve the overall accuracy in the setting of granular computing. The proposed approach is validated through experimental studies by using real-life data sets. The results show that the proposed approach overcomes effectively the limitations of the Bagging approach.

Keywords Machine learning · Ensemble learning · Classification · Bagging · Random forests · Granular computing

1 Introduction

Machine learning is a branch of artificial intelligence, which has become increasingly popular in the big data era. In particular, machine learning is typically categorized into supervised learning and unsupervised learning. Supervised learning is generally aimed at learning from labelled data, which means that the value of the decision attribute (dependent variable) is provided by domain experts based on the

values of the condition attributes (independent variables), with respect to each training instance (data point). In contrast, unsupervised learning is generally aimed at learning from unlabelled data, which means that none of the training instances is provided with a decision attribute and the learning is simply on the basis of the condition attributes. In practice, supervised learning is involved in classification and regression tasks, and unsupervised learning is involved in association and clustering tasks.

The rest of this paper will focus on classification tasks. In the context of machine learning, classification can be achieved by using a single classifier or an ensemble of classifiers. A single classifier is typically learned by using a single learning algorithm, such as ID3 (Quinlan 1986) and C4.5 (Quinlan 1993). However, it has been proven in machine learning literature (Breiman 1996; Freund and Schapire 1996) that each single learning method has its own advantages and disadvantages. In particular, each single algorithm may involve bias if the learning strategy is fully or partially based on heuristics. Also, data is timely

✉ Han Liu
LiuH48@cardiff.ac.uk
Mihaela Cocea
mihaela.cocea@port.ac.uk

¹ School of Computer Science and Informatics,
Cardiff University, Queen's Buildings, 5 The Parade,
Cardiff CF24 3AA, UK

² School of Computing, University of Portsmouth,
Buckingham Building, Lion Terrace, Portsmouth PO1 3HE,
UK

changed in reality, and such changes in training data are very likely to result in variance regarding the performance of a learning algorithm. In fact, a small change in training data may lead to a huge difference in terms of learning performance, especially when the learning algorithm is sensitive to the change in training data (Breiman 1996).

Due to the issues mentioned above, ensemble learning has been increasingly used towards improving the overall accuracy of classification. Bagging is one of the most popular approach of ensemble learning (Kononenko and Kukar 2007), which involves sampling with replacement towards using different versions of a training set and combining the classifications made by different classifiers learned from different versions of the training set towards classifying an instance, i.e. majority voting. In this way, the performance of learning has shown to be more stable in terms of classification accuracy (Kononenko and Kukar 2007). Another popular approach of ensemble learning is Boosting (Freund and Schapire 1996), which involves sequential training of base classifiers and combining the classifications made by the base classifiers through weighted voting, i.e. a base classifier is trained at each iteration; then the training data is modified to give more weights on incorrectly classified instances for training another base classifier at the next iteration; finally all the base classifiers are combined to jointly classify each new instance. In this way, the performance of learning can be boosted but the ensemble classifier may also be trained to overfit validation data (Kononenko and Kukar 2007). In this paper, we focus the study on the Bagging approach, since our proposal aims to enable the competition among base classifiers, which are trained on the same sample by using different learning algorithms in parallel.

Although the Bagging approach leads to the improvement of the overall accuracy of classification compared with the use of a single learning algorithm, there are still some issues that impact the performance of learning. In particular, we argue that the Bagging approach can be advanced in two ways: (a) it is necessary to judge effectively the degree to which a base classifier qualifies to get employed for classifying instances in the testing stage—in other words, a metric for judging the suitability of the individual classifiers for a particular instance is needed; (b) the voting criteria in the last stage need to be revisited towards more accurately judging the final classification. In this paper, we incorporate the above two ways towards reducing the bias in base classifiers selection and voting and advancing the Bagging approach by taking advantage of nature-inspired strategies (Liu et al. 2016; Liu and Cocea 2017b), which assume that the highest weight for a classifier or class means the highest chance for the classifier or class to be selected, i.e. the highest weight does not guarantee that the classifier or class is the winner in classifier selection or voting.

The rest of this paper is organized as follows: Sect. 2 provides a review of the background and recent developments of the Bagging approach. We also identify in this section the limitations of the approach that highlight the need for further development. In Sect. 3, we propose a nature-inspired framework of ensemble learning towards advancing the Bagging approach by addressing the previously identified limitations. We also justify how granular computing concepts are used for designing the ensemble learning framework. In Sect. 4, we conduct an experimental study by using 15 UCI data sets, and discuss the results in terms of the effect of the nature-inspired approach on the classification accuracy. In Sect. 5, we highlight the contribution of this paper and suggest research directions for further improvements in this area.

2 Related work

In this section, we describe the procedure of the Bagging approach and provide a review of the background and recent developments of this approach. Also, we identify the current limitations of the Bagging approach that indicate the need for further development.

2.1 Background of Bagging

Bagging (Breiman 1996) is a popular method of ensemble learning, which stands for bootstrap aggregating. As mentioned in Sect. 1, the Bagging approach involves sampling with replacement towards getting different versions of training data. In particular, this approach typically takes n samples, with each sample of size m , where m is the size of the original training set, in which the instances from the training set are randomly selected into each sample set. This indicates that some instances in the training set may appear more than once in one sample set and some other instances may never appear in that sample set. On average, each sample set is expected to contain 63.2% of the training instances (Kononenko and Kukar 2007; Liu and Gegov 2015; Liu and Cocea 2018).

In the training stage, a chosen learning algorithm learns a base classifier from each of the sample sets. In the testing stage, each of the base classifiers makes an independent classification, and the final classification is then made by combining the outputs of the base classifiers through majority voting (equal voting), i.e. the most frequently occurring class is assigned to the instance being classified. The detailed procedure of Bagging is illustrated in Fig. 1.

As mentioned in Sect. 1, Bagging addresses the issue that a small change in the training data leads to a great impact on the performance of learning, especially when the chosen learning algorithm is very sensitive to changes in training data (Breiman 1996). In particular, the random

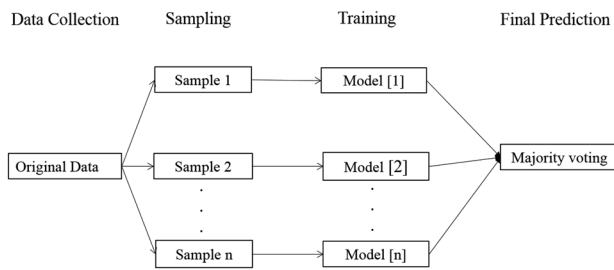


Fig. 1 The procedure of Bagging (Liu et al. 2017)

forests algorithm (Breiman 2001) is a special example of the Bagging approach, where decision trees must be the base classifiers learned in the training stage. Also, the random forests algorithm is an extension of the random decision forests algorithm, which was developed in Ho (1995) and is based on the random subspace method (Ho 1998). In other words, the random decision forests algorithm involves random selection of a subset of attributes, such that attribute selection at each iteration of decision tree learning (Liu et al. 2017) is on the basis of the selected subset of attributes at that iteration. Consequently, the random forests method involves the combination of Bagging and random feature subset selection—it is this combination that has led to great advances in decision tree learning in the setting of ensemble learning (Breiman 2001; Kononenko and Kukar 2007; Tan et al. 2005).

Further to the introduction of the Bagging idea, this approach of ensemble learning has been used for advancing machine learning techniques in various ways. In particular, a combination of Bagging and Boosting was proposed in Zheng and Webb (1998) towards achieving more advanced Boosting referred to as ‘multiple Boosting’. Bagging was also used in Skurichina and Duin (1998) towards advancing the performance of linear classifiers. Furthermore, Bagging was used in Skurichina and Duin (2002) by combining it with Boosting and the Random Subspace method towards advancing further the performance of linear classifiers. Bagging was also used jointly with Boosting in Borra and Ciacio (2002) towards improving non-parametric learning methods. In addition, Bagging was combined with Geographical Information Systems in Rizzoli et al. (2002) for lyme disease risk prediction in Trentino, Italian Alps, and was also used to achieve advances in distributed learning (Chawla et al. 2002) and computer vision (Draper and Baek 1998).

2.2 Recent developments

In recent years, the Bagging approach has been advanced through the incorporation of competitive learning towards effective employment of base classifiers. In particular, an extended framework of Bagging was developed in Liu and

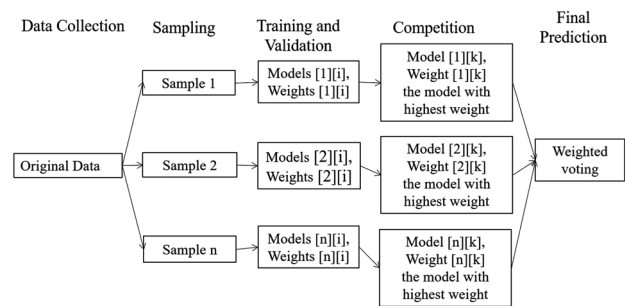


Fig. 2 The procedure of advanced Bagging (Liu et al. 2017)

Gegov (2015). The details of the extended framework are illustrated in Fig. 2.

In this extended framework, there are multiple learning algorithms employed, which means that there are multiple base classifiers learned from each sample of the training data. This is in order to involve competition among the base classifiers learned from the same sample of the training data. In the competition stage, from each sample, the learned base classifiers are put in a group, and within each group, the base classifiers are evaluated in terms of their quality (weight) by using the validation data, and the base classifier that has the highest weight is employed for getting involved in the testing stage towards classifying unseen instances. The experimental results reported in Liu and Gegov (2015) show that the employment of multiple learning algorithms for involving competitive learning leads to improvement of the overall accuracy of classification, in comparison with the traditional Bagging approach that employs only a single learning algorithm for each training sample.

On the other hand, the extended framework of Bagging, which was introduced in Liu and Gegov (2015), also involves modification of the voting strategy. In particular, as mentioned in Sect. 2.1, the traditional Bagging approach typically employs majority voting for final classification of an instance. Some other ensemble learning approaches, such as Boosting, employ weighted voting for classifying an instance, and the overall accuracy of a base classifier (estimated by using validation data) is typically used to contribute towards increasing the weight of a class.

However, as argued in Liu and Gegov (2015), neither majority voting nor weighted voting would be effective enough in measuring the confidence of a classification decision, as outlined in the following example. In particular, majority voting is to select the most frequently occurring class for classifying an instance, whereas weighted voting is to select the most highly weighted class for the same purpose. For example, let's consider three classifiers A, B and C, which are used for classifying an instance to either ‘Positive’ or ‘Negative’; A gives ‘Positive’ as the classification output with the weight of 0.8, while both B and C give ‘Negative’

as the classification with weights of 0.55 and 0.2, respectively. In this example, the final classification when using majority voting would be ‘Negative’ since the frequency for the ‘Negative’ class is higher than the frequency for the ‘Positive’ class, i.e. frequency of 2 for ‘Negative’ from classifiers B and C, and frequency of 1 for ‘Positive’ from classifier A. In contrast, the final classification when using weighted voting would be ‘Positive’ since the weight of the ‘Positive’ class is higher than the weight of the ‘negative’ class, i.e. the weight for ‘Positive’ is 0.8, while the weight for ‘Negative’ is $(0.55 + 0.2 = 0.75)$.

Although weighted voting has been considered more effective than majority voting, there are still research questions to investigate about how the weight of a class is established. In traditional ensemble learning, the typical way is by measuring the overall accuracy of a base classifier based on validation data and then adding the overall accuracy towards increasing the total weight of a class. However, as described in Liu and Gegov (2015); Kononenko and Kukar (2007), the overall accuracy of a classifier can not reflect the confidence of a classifier in classifying instances of a single class. In other words, a classifier may be confident in classifying instances of one class but not confident in classifying instances of other classes. Therefore, the use of precision/recall instead of overall accuracy has been investigated theoretically and empirically in Liu and Gegov (2015).

From theoretical perspectives, as argued in Liu and Gegov (2015); Liu et al. (2015), precision is considered more effective than recall in measuring the confidence of a classifier in classifying instances of a particular class. In particular, precision with respect to a class reflects the percentage of instances that are correctly classified to a particular class that the classifier assigns to the instances, whereas recall with respect to a class reflects the percentage of instances of a particular class that are correctly classified. In accordance with the above definitions, high recall could result from the case that a class has a low frequency. For example, as illustrated in Liu and Gegov (2015), while there are 5 out of 20 instances that belong to the ‘Positive’ class, a classifier correctly classifies the 5 instances to the ‘Positive’ class, but also incorrectly classifies another 5 instances to the ‘Positive’ class. In this case, precision with respect to the ‘Positive’ class is 50% and recall with respect to this class is 100%. In other words, precision is the proportion of instances correctly classified as a particular class from all the instances classified as that class (5 out of 10 in the above example), while recall is the proportion of instances correctly classified as a particular class from all the instances belonging to that class (5 out of 5 in the above example).

In fact, in real applications, it is impossible to know the actual class to which an unseen instance belongs. From this point of view, while the confidence of an individual classification is measured, it is more appropriate to look at the

precision with respect to the class given by a classifier as an individual classification. In other words, it is known to which class (the target class) a classifier assigns an unseen instance, and is also achievable to count the frequency that an individual classification is correct while the target class is assigned to the unseen instance by the classifier. The experimental results reported in Liu and Gegov (2015) show that precision is more effective than recall and overall accuracy in terms of measuring the confidence of an individual classification from a classifier, towards improving the performance of ensemble classification through more intelligent voting.

3 Nature-inspired ensemble framework

In this section, we propose to adopt nature-inspired techniques towards advancing further the Bagging approach. In particular, we adopt natural selection for more effective employment of base classifiers. Also, we employ precision to measure the confidence of an individual classification from each base classifier, towards increasing the weight of a class used for voting. However, the voting is inspired naturally by taking the weight of a class as the chance of selecting this class towards classifying an instance.

3.1 Key features

The nature-inspired framework of Bagging is illustrated in Fig. 3. Comparing with the framework illustrated in Fig. 2, the main modifications are in terms of employment of base classifiers and voting, which are shown in the last two layers of the framework, namely ‘Selection’ and ‘Final Prediction’, as illustrated in Fig. 3.

In terms of employment of base classifiers, natural selection is adopted to employ a base classifier within each group of base classifiers learned from the same sample of training data, which means that the weight of a base classifier is taken as the chance of employing this classifier to get involved in the testing stage. In contrast, in the framework illustrated in

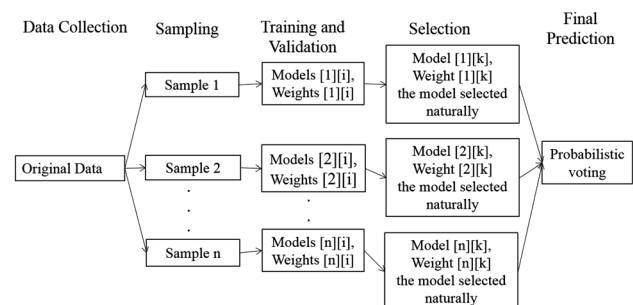


Fig. 3 Nature-inspired Bagging

Fig. 2, heuristic selection is adopted for employing a base classifier within each group of base classifiers, which means that the base classifier of the highest weight within its group is certainly employed to get involved in the testing stage.

On the other hand, in terms of voting towards final classification, probabilistic voting is adopted in the nature-inspired framework illustrated in Fig. 3. In this context, the weight of a class is measured by adding the precision values of the base classifiers that give this class as their individual classifications, and the weight is used as the chance of selecting this class towards classifying an instance. In contrast, in the framework illustrated in Fig. 2, weighted voting is adopted towards classifying an instance, which means that the class of the highest weight is certainly selected and assigned to the instance being classified.

3.2 Justification

The ensemble learning framework is partially designed in the setting of granular computing, which is a paradigm of information processing (Yao 2005b). From a philosophical perspective, granular computing is a way of structured thinking (Yao 2005b). From a practical perspective, granular computing is considered as a way of structured problem solving (Yao 2005b).

In general, granular computing involves two main operations, namely, granulation and organization (Yao 2005a). The former operation is aimed at decomposition of a whole into different parts, whereas the latter operation is aimed at integrating several parts into a whole. In computer science, the concepts of granulation and organization have been popularly used to achieve the top-down and bottom-up approaches, respectively (Liu and Cocea 2017a; Liu et al. 2018). In the context of ensemble learning, the Bagging approach involves random sampling of training data with replacement, which essentially follows the principle of information granulation. Also, the Bagging approach involves combining the independent outputs of base classifiers for classifying each new instance, which essentially follows the principle of organization.

In granulation and organization, the main aim is to deal with granules and granularity (Pedrycz 2011; Pedrycz and Chen 2011, 2015a, b, 2016), which are two main concepts of granular computing. A granule generally represents a large

particle, which consists of smaller particles that can form a larger unit. In the setting of the nature-inspired ensemble learning, a group of base classifiers needs to be trained on each sample (as illustrated in Fig. 3), and the best base classifier within each group needs to be selected and added into the ensemble for classifying new instances in the testing stage. In this context, the ensemble, which consists of finally selected base classifiers, is viewed as a granule in the top level of granularity. Since each of these base classifiers is selected from a group of classifiers trained on a specific one of the samples drawn from the original training data, each of these groups has a hierarchical relationship to the ensemble. From this point of view, each of the above groups is viewed as a granule in the second level of granularity. In addition, each of base classifiers in the ensemble needs to make an independent classification in the testing stage, so the set of independent classifications from these base classifiers can be viewed as a granule, which is horizontally correlated to the ensemble (another granule) in the top level of granularity.

On the other hand, the strategies of base classifiers selection and voting are designed through nature inspiration. For example, probabilistic voting (Liu et al. 2016; Liu and Cocea 2017b) is viewed to be inspired naturally in the setting of computational intelligence, since this kind of voting is made on the basis of the hypothesis that the class of the highest weight only has the best chance of being selected towards classifying an instance. In other words, it is not guaranteed that the class of the highest weight is certainly selected and assigned to the instance being classified. The procedure of probabilistic voting is illustrated below:

- Step 1: calculating the weight W_i for each single class i .
- Step 2: calculating the total weight W over all classes.
- Step 3: calculating the percentage P_i of weight W_i for each single class i , i.e. $P_i = W_i \div W$.
- Step 4: Randomly selecting a single class i with the probability P_i towards classifying an unseen instance.

The following example relating to Bayes Theorem is used for the illustration of the above procedure:

Inputs(binary): x_1, x_2, x_3
Output(binary): y

Probabilistic correlation (induced from training data):

$$P(y = 0|x_1 = 0) = 0.6, P(y = 1|x_1 = 0) = 0.4, P(y = 0|x_1 = 1) = 0.5, P(y = 1|x_1 = 1) = 0.5;$$

$$P(y = 0|x_2 = 0) = 0.4, P(y = 1|x_2 = 0) = 0.6, P(y = 0|x_2 = 1) = 0.8, P(y = 1|x_2 = 1) = 0.2;$$

$$P(y = 0|x_3 = 0) = 0.5, P(y = 1|x_3 = 0) = 0.5, P(y = 0|x_3 = 1) = 0.6, P(y = 1|x_3 = 1) = 0.4;$$

While $x_1 = 0, x_2 = 0, x_3 = 1, y = ?$

Following Step 1, the weight W_i for each single value of y is:

$$W_0 = P(y = 0 | x_1 = 0, x_2 = 0, x_3 = 1) = P(y = 0 | x_1 = 0) \times P(y = 0 | x_2 = 0) \times P(y = 0 | x_3 = 1) = 0.6 \times 0.4 \times 0.6 = 0.144$$

$$W_1 = P(y = 1 | x_1 = 0, x_2 = 0, x_3 = 1) = P(y = 1 | x_1 = 0) \times P(y = 1 | x_2 = 0) \times P(y = 1 | x_3 = 1) = 0.4 \times 0.6 \times 0.4 = 0.096$$

Following Step 2, the total weight $W = W_0 + W_1 = 0.144 + 0.096 = 0.24$.

Following Step 3, the percentage P_i of weight for each single value of y is:

$$\text{Percentage for } y = 0: P_0 = 0.144 \div 0.24 = 60\%$$

$$\text{Percentage for } y = 1: P_1 = 0.096 \div 0.24 = 40\%$$

Following Step 4, $y = 0$ (60% chance) or $y = 1$ (40% chance).

In the above illustration, weighted voting would result in 0 being assigned to y due to its higher weight shown in Step 4. In particular, in the context of weighted voting, Step 4 would indicate that over the total weight the percentage of the weight for y to equal 0 is 60% and the percentage of the weight for y to equal 1 is 40%. Therefore, weighted voting would choose to assign y the value of 0. However, in the context of probabilistic voting, Step 4 would indicate that y could be assigned either 0 (with 60% chance) or 1 (with 40% chance). In this way, the bias in voting can be reduced effectively towards improvement of the overall accuracy of classification in ensemble learning.

The probabilistic voting approach illustrated above is very similar to natural selection which is one step of the procedure of genetic algorithms (Man et al. 1996; Chen and Chung 2006; Maity et al., in press), i.e. each class is viewed as an individual and the probability of a class being selected is viewed as the fitness of an individual involved in natural selection. In particular, the way of selecting a class involved in Step 4 of the above procedure is inspired by the Roulette Wheel Selection (Lipowski and Lipowska 2012). In this paper, the nature selection strategy is also used as a technique for employing base classifiers as mentioned in Sect. 3.1. In this context, each base classifier is viewed as an individual and the chance of a base classifier being employed is viewed as the fitness of an individual.

The motivation for incorporating nature-inspired characteristics into the Bagging framework is mainly to deal effectively with the uncertainty due to the incompleteness of training and validation data. In particular, it is fairly difficult to guarantee in practice that the collected data covers a complete pattern. In other words, it is highly possible that a base classifier covers an incomplete pattern, which means that a pattern may exist but has not been learned yet, due to the incompleteness of training data. On the other hand, as mentioned in Sect. 3.1, the weight of a base classifier is

measured by using validation data. If the validation data is of low completeness, it is very possible that some pattern has been poorly learned but has never been tested, i.e. it is not

possible to reflect accurately the confidence of a classifier in classifying instances covered by this part of the learned pattern.

On the basis of the above argumentation, while the test set has some instances covered by the pattern that has been poorly learned or even not learned at all, it is very likely to result in incorrect classifications if the above pattern has not been covered in the validation data either. From this point of view, the weight of a base classifier measured by using validation data can not be completely trusted, and the incorporation of nature inspired characteristics is thus necessary towards uncertainty handling. The same also applies to the measure of the weight of a class towards voting. The experimental results reported in Liu et al. (2016); Liu and Cocea (2017b) have shown that the use of probabilistic voting leads to an improvement of classification accuracy, in comparison with the use of weighted voting.

4 Experimental results

In this section, we report an experimental study, which is conducted by using 15 data sets retrieved from the UCI repository (Lichman 2013).

The experimental study involves the incorporation of multiple algorithms into the nature-inspired Bagging framework, which is compared with the random forests method and the framework (illustrated in Fig. 2) that also incorporates multiple algorithms but has no nature inspired characteristics. The purpose is to show that the incorporation of nature inspired characteristics for both selecting base classifiers and voting would lead to improvement of classification accuracy, comparing with the case that the employment of base classifiers is through the selection of the one of the highest weight within each group of base classifiers and that weighted voting is used for final classification. In addition, this study also aims to show that the nature inspired framework of Bagging is capable of outperforming the random forest method.

In this study, the nature-inspired framework of ensemble learning, which involves C4.5, Naive Bayes and K nearest neighbour for learning base classifiers, is compared with the framework illustrated in Fig. 2, which also involves only C4.5, Naive Bayes and K nearest neighbour for learning base classifiers, but uses weighted voting. Thus, both approaches

Table 1 Data sets

Dataset	Attribute types	Number of attributes	Number of instances	Number of classes
Credit-g	Discrete, continuous	20	1000	2
Hepatitis	Discrete, continuous	20	155	2
Lung-cancer	Discrete	56	32	3
Breast-cancer	Discrete	9	286	2
Labor	Discrete, continuous	17	57	2
Spect	Discrete	22	267	2
Sonar	Continuous	60	208	2
Sick	Discrete, continuous	30	3772	2
Postoperative	Discrete, continuous	8	90	3
Sponge	Discrete, continuous	45	76	12
Cylinder-bands	Discrete, continuous	39	512	2
Haberman	Continuous	3	306	2
Solar-flare-2	Discrete	13	1066	2
Supermarket	Discrete	217	4627	2
Contact-lenses	Discrete	4	24	3

use the competitive selection of base classifiers, thus allowing us to assess the influence of natural selection of base classifiers and probabilistic voting on the classification performance. In addition, the nature-inspired framework is also compared with the random forests method due to its popularity in real applications. The characteristics of the 15 data sets used in this study are described in Table 1, which show to be more diverse—some data sets include only discrete or continuous attributes and the others include both types of attributes. The values of discrete attributes are nominal and thus more straightforward to deal with, whereas the ones of continuous attributes are numerical leading to more complex computation during classifiers training.

The experiments are conducted by partitioning a data set into a training set and a test set in the ratio of 70:30. For each data set, the experiment is repeated 10 times in terms of data partitioning and the average accuracy is taken for comparative validation. In terms of parameters setting, the ensemble size is set to 10, i.e. 10 samples are drawn from the training data, so 10 classifiers (each one trained on a sample) are obtained to make up the ensemble. The value of K for the nearest neighbour algorithm is set to 3.

The results of the experimental study are shown in Table 2. In particular, the second column indicated the used of the Random Forest algorithm; the third column, i.e. Heuristic Bagging, indicates that the advanced Bagging

Table 2 A comparison of classification accuracy rates for different data sets based on different methods

Dataset	Random forest (%)	Heuristic Bagging (%)	Nature-inspired Bagging (%)
Credit-g	72	68	71
Hepatitis	85	86	91
Lung-cancer	70	83	89
Breast-cancer	65	75	79
Labor	88	75	91
Spect	71	92	92
Sonar	77	77	75
Sick	98	93	94
Postoperative	59	75	92
Sponge	91	92	95
Cylinder-bands	65	62	67
Haberman	66	75	81
Solar-flare-2	100	100	100
Supermarket	63	62	65
Contact-lenses	43	75	86

approach (illustrated in Fig. 2) is adopted, where three learning algorithms (C4.5, Naive Bayes and K nearest neighbour) are employed for learning base classifiers and weighted voting is used for final classification. In contrast, the last column indicates that the nature inspired framework of Bagging is adopted, where the same algorithms (C4.5, Naive Bayes and K nearest neighbour) are employed for learning base classifiers and probabilistic voting is used for final classification.

When comparing the results of the two Bagging approaches (heuristic and nature inspired), we notice that nature-inspired approach outperforms the heuristic one in 12 out of 15 cases; the performance of the two approaches is the same for 2 cases, i.e. ‘spect’ and ‘solar-flare-2’, and the heuristic approach outperforms the nature inspired one in one case, i.e. ‘sonar’.

The results shown in Table 2 indicate that the nature inspired framework of Bagging outperforms the random forest method in 11 out of 15 cases, i.e. ‘hepatitis’, ‘lung-cancer’, ‘breast-cancer’, ‘labor’, ‘spect’, ‘postoperative’, ‘sponge’, ‘cylinder-bands’, ‘haberman’, ‘supermarket’ and ‘contact-lenses’. Also, there is one case that the nature inspired framework of Bagging performs the same as the random forest method, i.e. solar-flare-2. In the rest of the three cases, the nature inspired framework of Bagging performs slightly worse than the random forest method, i.e. ‘credit-g’, ‘sonar’ and ‘sick’.

The results in the experimental study show that incorporation of nature-inspired characteristics leads to advances in classification performance, comparing with the case that the same algorithms are employed for learning base classifiers

but both the employment of base classifiers and voting are based on heuristics (weight of classifiers/classes). The above descriptions indicate that the incorporation of nature-inspired characteristics can result in effective reduction of bias on the employment of base classifiers and voting and thus lead to advances in ensemble classification.

On the other hand, the results show that in several cases the employment of multiple algorithms for learning base classifiers and the incorporation of nature-inspired characteristics for employing base classifiers and voting fail to achieve that the performance is better than the traditional Bagging approach (i.e. the random forests method in this experimental study). This would indicate two points:

1. The employed learning algorithms need to be complementary to each other in terms of learning base classifiers from the same sample of training data, i.e. if the base classifier learned by an algorithm is not good enough, then the other base classifier learned from the same sample by the other algorithm needs to be good enough for the overall accuracy of classification to be good enough; if all algorithms lead to poor base classifiers, the overall accuracy of classification will be poor; thus, consideration needs to be given to which algorithms are selected to be part of the ensemble learning—a possible way for these decisions is the use of a measure to assess the ability of an algorithm to learn from a given data set; this idea has been explored in Liu et al. (2017), however, further research is needed to incorporate it within the nature inspired framework;
2. The incorporation of nature-inspired characteristics needs to lead to the effective trade-off between bias and variance, i.e. a large increase of variance needs to be avoided despite the reduction of bias. This point can also be supported by the experimental results shown in Table 2, which indicate that in several cases the incorporation of nature-inspired characteristics for employing base classifiers and voting fails to outperform the advanced Bagging approach (illustrated in Fig. 2).

5 Conclusion

In this paper, we proposed a nature-inspired framework of ensemble learning for advancing the Bagging approach in the setting of granular computing. In particular, we identified two critical points that have an impact on classification accuracy: (a) the selection of base classifiers for classifying unseen instances, and (b) the strategy of voting for final classification. In order to address the above two points, we incorporated nature inspired characteristics in both the training and testing stages. In the training stage, natural selection has been used for employing more effectively base classifiers,

based on the hypothesis that the base classifier of the highest weight has the best chance to be of the highest quality leading to the best performance of classification. In the testing stage, probabilistic voting is used for final classification, based on the hypothesis that the class of the highest weight has the best chance of being selected towards classifying correctly an unseen instance. Our experimental results show that the nature-inspired framework of ensemble learning mostly outperforms random forests (based on the Bagging approach illustrated in Fig. 1) and the advanced Bagging approach (illustrated in Fig. 2), in terms of classification accuracy.

The performance of ensemble learning can potentially be improved even further. In particular, precision has been used in this paper to measure the confidence of an individual classification from a base classifier, but the measure of confidence can actually be done in more depth. In other words, precision is just a measure that reflects the capability of a classifier in classifying instances of a particular class. In the big data era, a class can be very general and covers a very broad range of patterns, which means that a class may need to be specialized into sub-classes towards in-depth analysis of the confidence of an individual classification from a classifier. From this point of view, it is worth exploring the idea of instance-based evaluation of a classifier towards advancing further the performance of ensemble classification, comparing with class-based evaluation through the use of precision.

Instance-based evaluation is generally achieved through looking at the performance of a classifier only on the instances that are highly similar to the current unseen instance. For example, in the context of rule learning, instances that are covered by the same rule are considered to be highly similar to each other. Also, in the context of instance-based learning, instances that are highly similar to each other are likely to be grouped together. In future research, we will investigate the adoption of instance-based evaluation through the above two learning approaches, i.e. rule learning and instance-based learning, towards in-depth evaluation of classifiers in terms of their confidence of an individual classification. In addition, it is also worth to investigate the effectiveness of adopting the proposed framework of ensemble learning in the context of multi-attribute decision making (Xu and Wang 2016; Liu and You 2017; Chatterjee and Kar 2017; Lee and Chen 2008; Zulueta-Veliz and Garca-Cabrera 2018), and incorporate fuzzy set theory related techniques (Zadeh 1965; Wang and Chen 2008; Chen et al. 2012, 2009; Chen and Chen 2011; Chen and Tanuwijaya 2011; Chen and Chen 2001; Chen and Chang 2011; Chen et al. 2013) into the proposed framework to achieve fuzzy ensemble learning (Nakai et al. 2003).

Acknowledgements The authors acknowledge support for the research reported in this paper through the research development fund at the University of Portsmouth.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Borra S, Ciaccio AD (2002) Improving nonparametric regression methods by bagging and boosting. *Comput Stat Data Anal* 38(4):407–420
- Breiman L (1996) Bagging predictors. *Mach Learn* 24(2):123–140
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Chatterjee K, Kar S (2017) Unified granular-number-based ahp-vikor multi-criteria decision framework. *Granul Comput* 2(3):199–221
- Chawla NV, Moore TE, Hall LO, Bowyer KW, Kegelmeyer WP, Springer C (2002) Distributed learning with bagging-like performance. *Pattern Recogn Lett* 24(1–3):455–471
- Chen SJ, Chen SM (2001) A new method to measure the similarity between fuzzy numbers. In: *IEEE international conference on fuzzy systems*. Melbourne, Australia, pp 1123–1126
- Chen SM, Chang YC (2011) Weighted fuzzy rule interpolation based on ga-based weight-learning techniques. *IEEE Trans Fuzzy Syst* 19(4):729–744
- Chen SM, Chen CD (2011) Handling forecasting problems based on high-order fuzzy logical relationships. *Expert Syst Appl* 38(4):3857–3864
- Chen SM, Chung NY (2006) Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *Int J Inf Manage Sci* 17(3):1–17
- Chen SM, Tanuwijaya K (2011) Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques. *Expert Syst Appl* 38(12):15,425–15,437
- Chen SM, Wang NY, Pan JS (2009) Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships. *Expert Syst Appl* 36(8):11,070–11,076
- Chen SM, Munif A, Chen GS, Liu HC, Kuo BC (2012) Fuzzy risk analysis based on ranking generalized fuzzy numbers with different left heights and right heights. *Expert Syst Appl* 39(7):6320–6334
- Chen SM, Chang YC, Pan JS (2013) Fuzzy rules interpolation for sparse fuzzy rule-based systems based on interval type-2 gaussian fuzzy sets and genetic algorithms. *IEEE Trans Fuzzy Syst* 21(3):412–425
- Draper BA, Baek K (1998) Bagging in computer vision. In: *IEEE international conference on computer vision and pattern recognition*. Santa Barbara, CA, USA, pp 144–149
- Freund Y, Schapire RE (1996) Experiments with a new boosting algorithm. In: *Machine learning: proceedings of the thirteenth international conference*, Bari, Italy, pp 148–156
- Ho TK (1995) Random decision forests. In: *Proceedings of the 3rd international conference on document analysis and recognition*, Montreal, QC, pp 278–282
- Ho TK (1998) The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell* 20(8):832–844
- Kononenko I, Kukar M (2007) *Machine learning and data mining: introduction to principles and algorithms*. Horwood Publishing Limited, Chichester
- Lee LW, Chen SM (2008) Fuzzy multiple attributes group decision-making based on the extension of topsis method and interval type-2 fuzzy sets. In: *Proceedings of the 2008 international conference on machine learning and cybernetics*, Kunming, China, vol 6, pp 3260–3265
- Lichman M (2013) UCI machine learning repository. <http://archive.ics.uci.edu/ml>
- Lipowski A, Lipowska D (2012) Roulette-wheel selection via stochastic acceptance. *Phys A* 391(6):2193–2196
- Liu H, Cocea M (2017a) Fuzzy information granulation towards interpretable sentiment analysis. *Granul Comput* 2(4):289–302
- Liu H, Cocea M (2017b) Granular computing based approach for classification towards reduction of bias in ensemble learning. *Granul Comput* 2(3):131–139
- Liu H, Cocea M (2018) *Granular computing based machine learning: a big data processing approach*. Springer, Berlin
- Liu H, Gegov A (2015) Collaborative decision making by ensemble rule based classification systems. Springer, Switzerland, pp 245–264
- Liu P, You X (2017) Probabilistic linguistic todim approach for multiple attribute decision-making. *Granul Comput* 2(4):332–342
- Liu H, Gegov A, Cocea M (2017) Unified framework for control of machine learning tasks towards effective and efficient processing of big data, chap 6. Springer, Berlin, pp 123–140
- Liu H, Cocea M, Ding W (2018) Multi-task learning for intelligent data processing in granular computing context. *Granul Comput* 3(3):257–273
- Liu H, Gegov A, Cocea M (2015) Hybrid ensemble learning approach for generation of classification rules. In: *International conference on machine learning and cybernetics*. Guangzhou, China, pp 377–382
- Liu H, Gegov A, Cocea M (2016) Nature and biology inspired approach of classification towards reduction of bias in machine learning. In: *International conference on machine learning and cybernetics*. Jeju Island, South Korea, pp 588–593
- Maity S, Roy A, Maiti M (In press) A rough multi-objective genetic algorithm for uncertain constrained multi-objective solid travelling salesman problem. *Granul Comput*
- Man KF, Tang KS, Kwong S (1996) Genetic algorithms: concepts and applications. *IEEE Trans Ind Electron* 43(5):519–534
- Nakai G, Nakashima T, Ishibuchi H (2003) A fuzzy ensemble learning method for pattern classification. *J Japan Soc Fuzzy Theory Intell Inf* 15(6):671–681
- Pedrycz W (2011) Information granules and their use in schemes of knowledge management. *Sci Iran* 18(3):602–610
- Pedrycz W, Chen SM (2011) *Granular computing and intelligent systems: design with information granules of higher order and higher type*. Springer, Heidelberg
- Pedrycz W, Chen SM (2015a) *Granular computing and decision-making: interactive and iterative approaches*. Springer, Heidelberg
- Pedrycz W, Chen SM (2015b) *Information granularity, big data, and computational intelligence*. Springer, Heidelberg
- Pedrycz W, Chen SM (2016) *Sentiment analysis and ontology engineering: an environment of computational intelligence*. Springer, Heidelberg
- Quinlan JR (1986) Induction of decision trees. *Mach Learn* 1:81–106
- Quinlan JR (1993) *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc, San Francisco
- Rizzoli A, Merler S, furlanello C, Genchi C (2002) Geographical information systems and bootstrap aggregation (bagging) of tree-based classifiers for lyme disease risk prediction in trentino, italian alps. *J Med Entomol* 39(3):485–492
- Skurichina M, Duin RP (1998) Bagging for linear classifiers. *Pattern Recogn* 31(7):909–930
- Skurichina M, Duin RP (2002) Bagging, boosting and the random subspace method for linear classifiers. *Pattern Anal Appl* 5(2):121–135
- Tan PN, Steinbach M, Kumar V (2005) *Introduction to data mining*. Addison-Wesley Longman Publishing Co., Inc, Boston
- Wang HY, Chen SM (2008) Evaluating students' answerscripts using fuzzy numbers associated with degrees of confidence. *IEEE Trans Fuzzy Syst* 16(2):403–415

- Xu Z, Wang H (2016) Managing multi-granularity linguistic information in qualitative group decision making: an overview. *Granul Comput* 1(1):21–35
- Yao J (2005a) Information granulation and granular relationships. In: IEEE international conference on granular computing. Beijing, China, pp 326–329
- Yao Y (2005b) Perspectives of granular computing. In: Proceedings of 2005 IEEE international conference on granular computing, Beijing, China, pp 85–90
- Zadeh L (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Zheng Z, Webb GI (1998) Multiple boosting: A combination of boosting and bagging. In: Proceedings of the 1998 international conference on parallel and distributed processing techniques and applications, Las Vegas, Nevada, pp 1133–1140
- Zulueta-Veliz Y, Garca-Cabrera L (2018) A choquet integral-based approach to multiattribute decision-making with correlated periods. *Granul Comput* 3(3):245–256

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