

ALGEBRAIC FUSION OF MULTIPLE CLASSIFIERS FOR HANDWRITTEN DIGITS RECOGNITION

HUIHUANG ZHAO^{1,3}, HAN LIU^{2*}

¹College of Computer Science and technology, Hengyang Normal University
Hengyang 421008, China

²School of Computer Science and Informatics, Cardiff University
Queen's Buildings, 5 The Parade, Cardiff CF24 3AA, United Kingdom

³Hunan Provincial Key Laboratory of Intelligent Information Processing and Application
Hengyang, 421002, China

E-MAIL: happyday.huihuang@gmail.com, liuh48@cardiff.ac.uk(*corresponding author)

Abstract:

Recognition of handwritten digits is a very popular application of machine learning. In this context, each of the ten digits (0-9) is defined as a class in the setting of machine learning based classification tasks. In general, popular learning methods, such as support vector machine, neural networks and K nearest neighbours, have been used for classifying instances of handwritten digits to one of the ten classes. However, due to the diversity of handwriting styles from different people, it can happen that some handwritten digits (e.g. 4 and 9) are very similar and are thus difficult to distinguish. Also, each single learning algorithm may have its own advantages and disadvantages, which means that a single algorithm would be capable of learning some but not all specific characteristics of handwritten digits. From this point of view, a method for handwritten digits recognition is proposed in the setting of ensemble learning, towards encouraging the diversity among different classifiers trained by different learning algorithms. In particular, the image features of handwritten digits are extracted by using the Convolutional Neural Network architecture. Furthermore, single classifiers trained respectively by K nearest neighbours and random forests are fused as an ensemble one. The experimental results show that the ensemble classifier was able to achieve a recognition accuracy of $\geq 98\%$ using the MNISSET data set.

Keywords:

Image processing; Handwritten digits recognition; Machine learning; Ensemble learning; Multi-classifier fusion

1 Introduction

A number of studies involving deep learning approaches have claimed state-of-the-art performances in a considerable number of tasks. These include, but are not limited to, image classification, natural language processing, speech recognition, and text classification. The models used in the said tasks employ the softmax function at the classification layer. However, there have been studies[1] conducted that takes a look at an alternative to softmax function for classification the support vector machine (SVM).

As we know, a single learning algorithm may have its own advantages and disadvantages in classification. For example, a single algorithm may be capable of learning some but not all specific characteristics of handwritten digits. Also, some instances of handwritten digits such as '4' and '9' could be very similar and are thus difficult to distinguish. From this point of view, we propose to adopt instance based recognition of handwritten digits in the setting of ensemble learning, towards encouraging the diversity among different classifiers trained by different learning algorithms.

In the whole procedure of recognition, it does not just involve combining classifiers trained by different machines learning algorithms. Prior to the classifiers training stage, the image features of handwritten digits are extracted by using the Convolutional Neural Network architecture. By analyzing those features, we found K nearest neighbours and random forests are complementary to each other. So, single classifiers trained respectively by K nearest neighbours and random forests are fused as an ensemble one.

The rest of this paper is organized as follows. In section 1, our motivation and aims are introduced. In section 2, we introduce some related work in this context. We describe the proposed algorithm in section 3. Experimental results are presented in section 4. and conclusion are shown in section 5.

2 Relate work

This section provides a review of the applications of convolutional neural networks for image classification, an overview of handwritten digits recognition and a review of traditional machine learning methods alongside potential improvements through the use of granular computing concepts.

2.1 Convolutional Neural Network

In machine learning, a convolutional neural network (CNN, or ConvNet) is a class of deep feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery [2].

CNN work is good at solving computer vision problems such as recognition of 3D objects, image denoising, image segmentation and character classification [3]. CNNs use a variation of multi-layer perceptrons designed to require minimal preprocessing. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics.

CNNs use relatively little pre-processing compared to other image classification algorithms. This independence from prior knowledge and human effort in feature design is a major advantage.

Convolutional architectures also seem to benefit extracting features from image data. In our method, the image features of handwritten digits are extracted by using the Convolutional Neural Network architecture.

2.2 Handwritten digits recognition

It has been intensely studied for many years in the field of handwriting recognition. Numerous results have been achieved by researchers who have used different algorithms, such as K-Nearest-Neighbors (KNNs), Support Vector Machines (SVMs) [4], Genetic algorithms [5], Neural Networks (NNs), Convolutional Neural Networks (CNNs), etc [2].

The reason why handwritten digit recognition is still an important area is due to its vast practical applications and financial implications. The industry demands a decent recognition rate with the highest reliability. the higher recognition rate on

handwritten digits means the higher accuracy for recognition of handwritten data, which usually exist in numeral strings. When we consider recognizing a numeral string of a ZIP code, the recognition probability of this string is the multiplication of the recognition probability of each isolated digit (assuming that each digit is correctly separated from the numeral string by the segmentation process). However, it is unrealistic to have a handwritten digit recognition system with 100% recognition accuracy. Hence, the reliability is much more important than the recognition accuracy in real-life systems [6]. [7] proposed one variant of RF called Forest-RI to the recognition of handwritten digits from the well known MNIST database. A strictly pragmatic approach based on different parametrization values in Random Forest is studied.

2.3 Review of machine learning methods

At present, the most popular algorithms in machine learning mainly include multi-layer perceptron (MLP), Naive Bayes (NB), K nearest neighbour (KNN) [8, 9], C4.5, and Random forests. Also, those machine learning algorithms have been used popularly in Handwritten digits recognition tasks.

MLP belongs to approaches of neural network learning to build a multi-layer architecture, which consists of input, hidden and output layers. In the input layer, each node represents an input feature. In the output layer, the single node represents the class output. Also, there are also several hidden layers which are also referred to as perceptron layers. Due to the constraints on hardware performance, there are typically only one or two hidden layers involved in a MLP model. In the MLP classification stage, the mapping from inputs to outputs is poorly transparent [8, 10].

NB belongs to approaches of Bayesian learning, which aims to learn the conditional relationships between feature values and class labels. the NB algorithm can show how each input feature relates to a class in the context of probability theory, but the NB algorithm does not involve identification of the correlation between different input features [11, 12].

KNN belongs to approaches of lazy learning. In the prediction stage, an instance is classified by finding k training instances that are the most similar ones to the unseen instance. The similarity between two instances can be calculated by some distance functions, such as including Euclidean, cosine, Chi square, and Minkowsky [8].

C4.5 belongs to approaches of decision tree learning. It involves self evaluation of features on an iterative basis. The C4.5 algorithm has a root node and a number of internal nodes. Each of these nodes is labeled a feature and also known as a non-leaf node, which indicates that at each iteration of decision tree

learning, a feature needs to be selected towards labelling a non-leaf node of a tree [8, 13].

Random forest (RF) is an ensemble learning approach that is aimed at training different decision trees with high diversity, i.e. it is a combination of decision trees such that they can classify each instance in a collaborative way [14].

2.4 Granular Computing

Granular computing is paradigm of information processing. From philosophical perspectives, it is viewed as a way of structured thinking, whereas it is considered as a way of structured problem solving [15]. In general, granular computing involves two main operations: granulation and organization. The former operation is aimed at decomposing a whole into different parts, whereas the latter is aimed to integrate several parts into a whole. In practice, the concepts of granulation and organization are used respectively for designing the top-down and bottom-up approaches from a computer science perspective [16]. In the setting of ensemble learning, the decomposition of training data into a number of samples through ransom sampling with replacement (bootstrap sampling) [17] is viewed as a practical application of the granulation operation, and the combination of multiple classifiers for outputting a final class label is considered as a practical application of the organization operation.

On the other hand, two main concepts, which are usually involved in granular computing theory and applications, include granule and granularity [18]. In general, a granule represents a collection of smaller units that can form a larger unit. Different granules may be located in the same or different levels of granularity, so these granules would involve horizontal relationships, such as mutual exclusion, correlation and mutual independence, if they are located in the same level of granularity. Otherwise, these granules would involve hierarchical relationships. In the context of ensemble learning, an ensemble of classifiers is viewed as a granule. Also, if the combination of classifiers is involved in different levels, then each of the levels is viewed as a level of granularity.

In addition, the performance of learning algorithms can be measured in different levels, i.e. classifier, class and instance levels. From this point of view, evaluation of a machine learning method could be viewed essentially as a multi-granularity operation task. In Section 3, we will introduce how to measure the confidence of classifiers in the instance level.

3 Algebraic fusion based ensemble learning

In this section, we illustrate the key features of our proposed approach of ensemble learning through fusion of multiple clas-

sifiers. Also, we justify the theoretical and practical significance of the ensemble learning approach in terms of increasing the diversity among different classifiers.

3.1. Key features

The proposed framework of fusion based ensemble learning is illustrated in Fig 1.

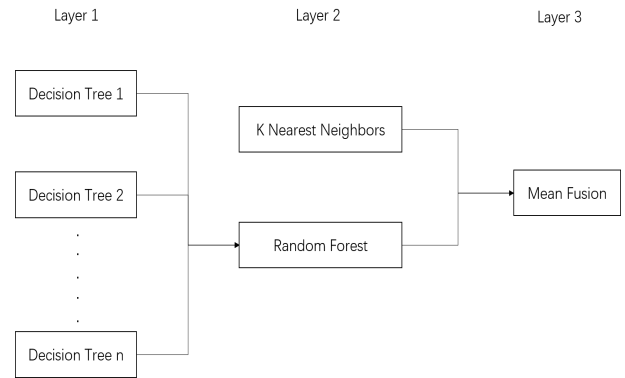


FIGURE 1. Fusion of K nearest neighbor and random forest

In Layer 1, a number n of decision trees are trained as base classifiers to make up a random forest. In particular, the Bagging (bootstrap sampling) approach is adopted to draw training samples with replacement, which means that some training instances may be selected more than once into the sample but some others may never be selected. Also, attribute selection for labelling each node of a decision tree is done on the basis of a randomly selected attribute subset drawn through random subspace sampling. More details on random forests can be found in [14].

In Layer 2, the parameters of the K nearest neighbors and random forest approach are set towards training an optimal KNN classifier and an optimal ensemble of decision trees, in order to achieve more effective fusion of the two classifiers in Layer 3.

In Layer 3, the trained KNN classifier is combined with the random forest one through algebraic fusion, which is essentially to combine the hidden outputs (the posterior probability for each class) of the classifiers in a fixed rule, e.g. mean, max, min and median [19].

3.2. Justification

In traditional machine learning, it is a usual practice to measure the confidence of classification in the classifier level, i.e.

accuracy. In some application areas such as sentiment analysis, the confidence is measured in the class label, i.e. precision, recall and F-measure. However, in handwritten digits recognition, the writing styles of different people may have high diversity, which is likely to result in ambiguous instances, e.g. 4 and 9 can be very similar and it is thus difficult to distinguish some handwriting of instances that belong to these two classes. Therefore, it is necessary to undertake instance based measure of confidence.

In the setting of ensemble learning, the most popular way of combining classifiers is voting. In particular, voting based ensemble classification could be achieved through choosing the class that the majority of the classifiers provide as their outputs, i.e. majority voting. Another way of voting, which is referred to as weighted voting, is achieved by taking the class output alongside the weight of each single classifier and deriving the class that obtains the highest weight for finally classifying an instance. The weight of a classifier is typically estimated by using the overall accuracy (confidence) of this classifier evaluated on a validation set, e.g. boosting [20]. There are also ways of measuring the confidence in the class level by using the precision or recall for a specific class [21]. Again, the confidence measured for a classifier or each single class can not simply represent the confidence in classifying an instance, due to the high diversity among different instances.

In the proposed framework of ensemble learning, we employ the KNN approach, since it is essentially an approach of instance based reasoning. The employment of the RF approach is helpful to increase the diversity of different decision tree classifiers in the first iteration of ensemble learning, since these trees are learned from different training samples and feature subsets. In the second iteration, the hidden outputs (probability for each class) of the KNN and RF classifiers are combined, which is essentially aimed at measuring the confidence of each classifier in the instance level, i.e. the probability for the class that a classifier outputs to classify an instance represents the confidence of the classifier on the instance. Since both KNN and RF show to be capable of training good classifiers and their different strategies of learning are also likely to result in high diversity between classifiers, we propose to fuse the two classifiers in an algebraic way, which will be specified in Section 4.

4 Experimental results and discussion

In this section, we report an experimental study using the MNIST data set, which is essentially a 10-class (0-9) classification task in the machine learning context.

The MNIST database (Modified National Institute of Standards and Technology database) is a large database of handwritten digits that is commonly used for training various image processing systems [22]. In this experimental study, we first evaluate the performance for three popular methods of standard learning, namely C4.5, NB, KNN, and MLP, in order to select the best performing one of the trained classifiers for fusion at a later stage. Secondly, we design to boost the performance of decision tree learning through the RF ensemble method. Finally, we fuse the KNN and RF classifier to boost the performance further. In terms of parameters setting, the K value for KNN is set to 3 and the trained random forest consists of 100 decision trees. Also, the MLP classifier is trained through 100 iterations with 2 hidden layers and 10 units in each layer. The KNN and RF classifiers are fused through averaging their hidden outputs (probability for each class), i.e. the mean rule of algebraic fusion. All the experiments are conducted by using 10-fold cross validation.

The results on the miniset data set is shown in Table 1 in terms of classification accuracy. The results indicate that the nature of the KNN method through instance based learning generally leads to much better performance than MLP, C4.5 and NB. Also, the high performance of the RF method again shows the effectiveness of increasing the diversity among different decision tree classifiers trained on different training samples and feature subsets, which leads to a significant improvement over the method of learning a single decision tree as normally expected. On this basis, the fusion of KNN and RF classifiers leads to a further improvement, which indicates that the different learning strategies between the KNN and RF methods can really result in diversity between their trained classifiers.

TABLE 1. Results

Method	Accuracy
MLP	0.61
C4.5	0.88
NB	0.81
KNN	0.97
RF	0.97
Fusion	0.98

5. Conclusions

In this context, a novel method for handwritten digits recognition method which combines KNN and Random Forest is proposed in the setting of ensemble learning, towards encour-

aging the diversity among different classifiers trained by different learning algorithms. During our approach, the Convolutional Neural Network is used to extract the the image features. Random Forest used to increase the diversity of different decision tree classifiers. And KNN classifier is combined with the Random Forest one through algebraic fusion. By comparing with four popular methods of standard learning, namely MLP, C4.5, NB and KNN, our proposed ensemble approach can achieve a better recognition accuracy of $\geq 98\%$ using the MNISSET data set. It is also proved ensemble learning which aims to encourage the diversity among different classifiers is very useful. In future, we will investigate to select an optimal set of features to boost the performance further.

Acknowledgements

This work was supported by National Natural Science Foundation of China (61503128), Science and Technology Plan Project of Hunan Province (2016TP102), Scientific Research Fund of Hunan Provincial Education Department (14B025, 16C0311), and Hunan Provincial Natural Science Foundation(2017JJ4001).

References

- [1] M. Elleuch, R. Maalej, and M. Kherallah, "A new design based-svm of the cnn classifier architecture with dropout for offline arabic handwritten recognition," *Procedia Computer Science*, vol. 80, pp. 1712–1723, 2016.
- [2] R. Plamondon and S. N. Srihari, "Online and off-line handwriting recognition: a comprehensive survey," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 1, pp. 63–84, 2000.
- [3] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification," in *Document Analysis and Recognition (ICDAR), 2011 International Conference on*. IEEE, 2011, pp. 1135–1139.
- [4] G. Guo, S. Z. Li, and K. L. Chan, "Support vector machines for face recognition," *Image and Vision computing*, vol. 19, no. 9-10, pp. 631–638, 2001.
- [5] C. De Stefano, F. Fontanella, C. Marrocco, and A. S. Di Freca, "A ga-based feature selection approach with an application to handwritten character recognition," *Pattern Recognition Letters*, vol. 35, pp. 130–141, 2014.
- [6] L. Guo, P.-S. Ge, M.-H. Zhang, L.-H. Li, and Y.-B. Zhao, "Pedestrian detection for intelligent transportation systems combining adaboost algorithm and support vector machine," *Expert Systems with Applications*, vol. 39, no. 4, pp. 4274–4286, 2012.
- [7] S. Bernard, S. Adam, and L. Heutte, "Using random forests for handwritten digit recognition," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, vol. 2. IEEE, 2007, pp. 1043–1047.
- [8] T. M. Mitchell *et al.*, "Machine learning. wcb," 1997.
- [9] H. Liu, A. Gegov, and M. Cocea, "Nature and biology inspired approach of classification towards reduction of bias in machine learning," in *Machine Learning and Cybernetics (ICMLC), 2016 International Conference on*, vol. 2. IEEE, 2016, pp. 588–593.
- [10] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.
- [11] M. Hazewinkel, "Bayes formula," *Encyclopedia of mathematics*, 2001.
- [12] D. Barber, *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.
- [13] H. Liu, A. Gegov, and M. Cocea, *Rule based systems for big data: a machine learning approach*. Springer, 2015, vol. 13.
- [14] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [15] Y. Yao, "Perspectives of granular computing," in *Proceedings of 2005 IEEE International Conference on Granular Computing*, Beijing, China, 25-27 July 2005, pp. 85–90.
- [16] J. Yao, "Information granulation and granular relationships," in *IEEE International Conference on Granular Computing*, Beijing, China, 25-27 July 2005, pp. 326–329.
- [17] L. Breiman, "Bagging predictors," *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [18] H. Liu and M. Cocea, "Fuzzy information granulation towards interpretable sentiment analysis," *Granular Computing*, vol. 2, no. 4, pp. 289–302, 2017.
- [19] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall/CRC, 2012.

- [20] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Machine Learning: Proceedings of the Thirteenth International Conference*, Bari, Italy, 3-6 July 1996, pp. 148–156.
- [21] H. Liu and A. Gegov, *Collaborative Decision Making by Ensemble Rule Based Classification Systems*. Switzerland: Springer, 2015, pp. 245–264.
- [22] X.-X. Niu and C. Y. Suen, "A novel hybrid cnn–svm classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.