EXTREME LEARNING MACHINE'S THEORY & APPLICATION

Predicting consumer sentiments using online sequential extreme learning machine and intuitionistic fuzzy sets

Hai Wang · Gang Qian · Xiang-Qian Feng

Abstract Predicting consumer sentiments revealed in online reviews is crucial to suppliers and potential consumers. We combine online sequential extreme learning machines (OS-ELMs) and intuitionistic fuzzy sets to predict consumer sentiments and propose a generalized ensemble learning scheme. The outputs of OS-ELMs are equivalently transformed into an intuitionistic fuzzy matrix. Then, predictions are made by fusing the degree of membership and non-membership concurrently. Moreover, we implement ELM, OS-ELM, and the proposed fusion scheme for Chinese reviews sentiment prediction. The experimental results have clearly shown the effectiveness of the proposed scheme and the strategy of weighting and order inducing.

Keywords Sentiment prediction · Extreme learning machine · OS-ELM · ensemble learning · Intuitionistic fuzzy set · Induced aggregation operator

1 Introduction

With the development of Web technology, more and more people tend to express their opinions and emotions on the Web, as well as absorb others' feeling online. People prefer to share their sentiments about a product or a service via electronic word-of-mouth reviews [5, 37]. The solution of automatically predicting the sentiment orientation of reviews has been widely applied to many fields. For instance, it could help potential customers to go through reviews about the concerned product to obtain overall orientations [30]. Also, the orientations of reviewers have important implications for managers to be aware of their brand building, product development, etc. [7]. In addition, it can help governments to analyze and monitor public opinions.

Sentiment prediction, or sentiment classification, as a special case of text classification for subjective texts, is becoming a hotspot for handling promptly increasing reviews. A scheme for (consumers') review sentiment classification aims at automatically judging what sentiment orientation, positive, negative, or neutral, a review is. There are mainly two strategies for prediction: machine learning techniques and semantic orientation aggregation [28]. The former follows traditional text classification techniques, such as Naïve Bayes (NB) [18] and support vector machine (SVM) [13]. The second strategy is to classify features into two classes, positive and negative, and then aggregate the overall orientation of a review. There are only a few studies focused on sentiment classification for Chinese reviews. Ye et al. [35] first worked on this area at 2005. Tan and Zhang [28] presented an experimental study with four feature selection methods and five machine learning methods. Wang et al. [30] presented a novel feature selection method based on utilizing mean and variance to improve Fisher's discriminant ratio. Zhang et al. [37] studied sentiment classification of restaurant reviews written in Cantonese.

Generally, there are some debates on feature selection and classification techniques in review sentiment classification. Pang et al. [13] pointed out that binary-based features have shown better performance that frequency-based features, while the opposite conclusion could be found in

H. Wang (✉) · G. Qian · X.-Q. Feng
School of Computer Science and Technology,
Nanjing Normal University,
Nanjing, China
e-mail: wanghai17@sina.com

X.-Q. Feng
Jiangsu Research Center of Information Security and Privacy
Technology, Nanjing, China

[30]. SVM performs the best in [13] on unigram features, whereas NB performs better than SVM in [6] using unigram features as well. Moreover, sentiment classification is severely dependent on domains or topics [28]. Xia et al. [31] argued that this is mainly because different types of features have distinct distributions and would probably vary in performance among different machine learning algorithms. Thus, they first combined distinct feature sets and separate machine learning classifiers to improve accuracy of sentiment classification. Multiple classifiers system (MCS) that fuses a pool of component classifiers is a pop issue which has more robustness, accuracy and generality than single classifier [1].

Besides above-mentioned debates, however, there are some disadvantages among existing techniques. First, due to uncertainties of outputs of component classifiers, it is reasonable to consider fuzziness of outputs, while most of the classification schemes harden the outputs of a component classifier by a maximum operator. Second, although Xia et al. [31] employed MCS to improve accuracies, only the degree of support of a test pattern belongs to one class with respect to a component classifier is considered, while the degree of a test pattern does not belong to the class is not taken into accont. If we want to measure both aspects and their fuzziness, the theory of intuitionistic fuzzy sets (IFSs) [2] is the very selection. In addition, we investigate weighting strategy in MCS in Chinese review sentiment prediction by proper operators of IFSs.

Huang et al. [9, 10, 11] presented a fast learning algorithm referred to as extreme learning machine (ELM) for single-hidden layer feedforward networks (SLFNs). Some developments of ELM are presented in [8, 17, 21, 23, 24, 39]. The hidden layer of ELM needs not to be iteratively tuned and therefore result in extremely fast generalization. Therefore, we employ ELM to study Chinese review sentiment prediction. Meanwhile, it is hard to prepare a high-quality corpus with enough reviews suit for training classifiers. Thus, a rational compromise can be done by organizing a small amount of reviews to train original classifiers, when new labeled reviews arrive, new and more accurate classifiers could be retrained by online sequential learning techniques. In addition, since word-of-mouth expressions may change in a long time sequence, it is necessary to add new reviews to capture the changes. Thus, a further development of ELM, online sequential extreme learning machine (OS-ELM) [19], is used as the component classifier.

In this paper, we aim at making an intensive study of the effectiveness of ensemble learning schemes, that is, ensemble of OS-ELMs (EOS-ELMs), for predicting Chinese review sentiments based on intuitionistic fuzzy aggregation. A novel multi classifiers fusion algorithm is proposed in which outputs of component classifiers are equivalently represented by a set of intuitionistic fuzzy values at first and the fusion process is interpreted as aggregation of IFSs. Further, we set up experiments to seek empirical answers to the following questions:

(1) Are ELM and OS-ELM effective to Chinese review sentiment classification?
(2) Can the performances of Chinese review sentiment classification benefit from the proposed ensemble learning techniques?
(3) Among all the versions of proposed fusion algorithm and other comparable algorithms, which generally performs the best?
(4) When weighting component classifiers, which is the better way in area of Chinese review sentiment classification?

The remainder of this paper is organized as follows. Section 2 reviews and presents some preliminaries on ELM, OS-ELM, IFS and induced operators. In Sect. 3, we present the Parallel EOS-ELM algorithm under intuitionistic fuzzy framework and the specific process of intuitionistic fuzzy fusion method. Section 4 evaluates the performance of the proposed scheme with experiments, to show the excellent results. Some further in-depth discussions on both the theoretical and application issues are given in Sect. 5. The conclusions are given in Sect. 6.

2 Preliminaries

2.1 ELM and OS-ELM

ELM [10], as a learning algorithm for SLFNs, randomly selects weights and biases for hidden nodes, and analytically determines the output weights by finding least square solution. OS-ELM [19, 25] proposed by Liang et al. was developed for SLFNs with additive and radial basis function hidden nodes on the basis of ELM. Consider N arbitrary distinct samples $(\mathbf{x}_i, \mathbf{t}_i) \in \mathbf{R}^n \times \mathbf{R}^m$. If a SLFN with L hidden nodes can approximate these N samples with zero error, it then implies that there exist β_i, $\mathbf{a}_i$, and b_i such that

$$f_L(\mathbf{x}_j) = \sum_{i=1}^{L} \beta_i G(\mathbf{a}_i, b_i, \mathbf{x}_j) = \mathbf{t}_j, \quad j = 1, 2, \ldots, N \qquad (1)$$

where $\mathbf{a}_i$ and b_i are the learning parameters of the hidden nodes, β_i is the output weight, and $G(\mathbf{a}_i, b_i, \mathbf{x}_j)$ denotes the output of the ith hidden node with respect to the input $\mathbf{x}_j$.

Assume the data $\aleph = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}^m\}_{i=1}^{N}$ presents to the network (L hidden nodes) sequentially (one-by-one or chunk-by-chunk with fixed or varying chunk size). There are two phases in the OS-ELM algorithm, initialization phase and sequential phase.

Initialization phase: A small chunk of training data is used to initialize the learning, $\aleph_0 = \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^{N_0}$ from the given training set $\aleph$, and $N_0 \geq L$ where $L = rank\ (\mathbf{H}_0)$.

(a) Randomly assign the input parameters.

(b) Calculating the initial hidden layer output matrix $\mathbf{H}_0$:

$$\mathbf{H}_0 = \begin{bmatrix} G(\mathbf{a}_1, b_1, \mathbf{x}_1) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ G(\mathbf{a}_1, b_1, \mathbf{x}_{N_0}) & \dots & G(\mathbf{a}_L, b_L, \mathbf{x}_{N_0}) \end{bmatrix}_{N_0 \times L} \quad (2)$$

(c) Estimating the initial out put weight $\beta_{(0)}$:$\beta_{(0)} = \mathbf{P}_0\,\mathbf{H}_0^T\,\mathbf{T}_0$, where $\mathbf{P}_0 = (\mathbf{H}_0^T\,\mathbf{H}_0)^{-1}$, and $\mathbf{T}_0 = [\mathbf{t}_1, \dots, \mathbf{t}_{N_0}]^T$.

(d) Set $k = 0$. (k: a parameter indicates the number of chunks of data that is prepared to the network.)

Sequential learning phase: Present the $(k + 1)$th chunk of new samples, $\aleph_{k+1} = \{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=\sum_{j=0}^{k} N_j + 1}^{\sum_{j=0}^{k+1} N_j}$, and N_{k+1} denotes the number of samples in the $(k + 1)$th chunk.

(e) Compute the partial hidden layer output matrix $\mathbf{H}_{k+1}$

$$\mathbf{H}_{k+1} = \begin{bmatrix} G\left(\mathbf{a}_1, b_1, \mathbf{x}_{\sum_{j=0}^{k} N_j + 1}\right) & \dots & G\left(\mathbf{a}_L, b_L, \mathbf{x}_{\sum_{j=0}^{k} N_j + 1}\right) \\ \vdots & \ddots & \vdots \\ G\left(\mathbf{a}_1, b_1, \mathbf{x}_{\sum_{j=0}^{k+1} N_j}\right) & \dots & G\left(\mathbf{a}_L, b_L, \mathbf{x}_{\sum_{j=0}^{k+1} N_j}\right) \end{bmatrix} \quad (3)$$

(f) Calculate the output weight $\beta_{(k+1)}$:

$$\mathbf{T}_{k+1} = \left[\mathbf{t}_{\sum_{j=0}^{k} N_j + 1}, \dots, \mathbf{t}_{\sum_{j=0}^{k+1} N_j}\right]^T \quad (4)$$

$$\mathbf{P}_{k+1} = \mathbf{P}_k - \mathbf{P}_k \mathbf{H}_{k+1}^T \left(\mathbf{I} + \mathbf{H}_{k+1} \mathbf{P}_k \mathbf{H}_{k+1}^T\right)^{-1} \mathbf{H}_{k+1} \mathbf{P}_k \quad (5)$$

$$\beta_{(k+1)} = \beta_{(k)} + \mathbf{P}_{k+1} \mathbf{H}_{k+1}^T \left(\mathbf{T}_{k+1} - \mathbf{H}_{k+1} \beta_{(k)}\right) \quad (6)$$

(g) Set $k = k + 1$. Go to (e).

Furthermore, there are some studies on the ensemble schemes of OS-ELM [8, 17, 21]. Corresponding fusion strategies will be specified and implemented by a unique form hereinafter.

2.2 IFS and induced operators

Since the introduction of fuzzy sets, several extensions of this concept have been defined. The most accepted one might be Atanassov's intuitionistic fuzzy sets (IFSs) [2]. The concept of IFS is as follows.

Definition 1 [2] Let X be an ordinary finite nonempty set. An IFS in X is an expression A given by

$$A = \{\langle x, \mu_A(x), v_A(x)\rangle | x \in X\} \quad (7)$$

where $\mu_A : X \to [0, 1]$, $v_A : X \to [0, 1]$ with the condition: $0 \leq \mu_A(x) + v_A(x) \leq 1$, for all x in X.

The functions $\mu_A(x)$, $v_A(x)$, and $\pi_A(x) = 1 - \mu_A(x) - v_A(x)$ denote, respectively, the degree of the membership, non-membership, and hesitation of the element x in the set A. Let Ω be the set of all IFVs. The ordered pair $\alpha(x) = (\mu_\alpha(x), v_\alpha(x))$ is called an intuitionistic fuzzy value (IFV) [32], where $\mu_\alpha(x), v_\alpha(x) \in [0, 1]$ and $\mu_\alpha(x) + v_\alpha(x) \leq 1$. In this paper, IFV is abbreviated as $\alpha = (\mu, v)$. Two operations on IFVs are as follows:

Definition 2 [32] Let $\alpha = (\mu_\alpha, v_\alpha)$ and $\beta = (\mu_\beta, v_\beta)$ be two IFVs, $\lambda > 0$ then

(1) $\alpha \oplus \beta = (\mu_\alpha + \mu_\beta - \mu_\alpha \mu_\beta, v_\alpha v_\beta)$

(2) $\lambda \alpha = (1 - (1 - \mu_\alpha)^\lambda, v_\alpha^\lambda)$.

Xu [32] proposed the following process to compare two IFVs:

Definition 3 Let $\alpha = (\mu_\alpha, v_\alpha)$, $\beta = (\mu_\beta, v_\beta)$ be two IFVs, $S(\alpha) = \mu_\alpha - v_\alpha$ and $S(\beta) = \mu_\beta - v_\beta$ be the scores of α and β, respectively, and $H(\alpha) = \mu_\alpha + v_\alpha$ and $H(\beta) = \mu_\beta + v_\beta$ be the accuracy degrees of α and β, respectively, then

(1) If $S(\alpha) > S(\beta)$, then $\alpha > \beta$;

(2) If $S(\alpha) = S(\beta)$ and $H(\alpha) > H(\beta)$, then $\alpha > \beta$;

(3) If $S(\alpha) = S(\beta)$ and $H(\alpha) = H(\beta)$, then $\alpha = \beta$.

To aggregate IFVs, several aggregation operators [32] were proposed, such as intuitionistic fuzzy arithmetic averaging (IFAA) operator, intuitionistic fuzzy weighted averaging (IFWA) operator, and intuitionistic fuzzy ordered weighted averaging (IFOWA) operator. Operators for classifiers fusion are cited here.

Definition 4 Let $\alpha_j = (\mu_j, v_j)(j = 1, 2, \dots, n)$ be a collection of IFVs with the weighting vector $w = (w_1, w_2, \dots, w_n)^T$ such that $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1$. An IFWA operator of dimension n is a mapping IFWA: $\Omega^n \to \Omega$, and

$$\text{IFWA}_w(\alpha_1, \alpha_2, \dots, \alpha_n) = \overset{n}{\underset{j=1}{\oplus}} \left(w_j \alpha_j\right) \quad (8)$$

Especially, if $w = (1/n, 1/n, \dots, 1/n)^T$, then the IFWA operator is reduced to the IFAA operator:

$$\text{IFAA}(\alpha_1, \alpha_2, \dots, \alpha_n) = \frac{1}{n} \overset{n}{\underset{j=1}{\oplus}} \left(\alpha_j\right) \quad (9)$$

The induced ordered weighted averaging operator proposed by Xu and Da [33] reorders the elements by an order inducing vector, and the associated weighting vector weights the element in the reordered position. We define an induced intuitionistic fuzzy ordered weighted averaging (I-IFOWA) operator for classifier fusion.

Definition 5 Let $\alpha_j = (\mu_j, v_j)(j = 1, 2, \ldots, n)$ be a collection of IFVs. An I-IFOWA operator of dimension n is a mapping I-IFOWA: $\Omega^n \to \Omega$, furthermore,

$$
\begin{aligned}
&\text{I} - \text{IFOWA}_w(\langle u_1, \alpha_1 \rangle, \langle u_2, \alpha_2 \rangle, \ldots, \langle u_n, \alpha_n \rangle) \\
&= \mathop{\oplus}\limits_{j=1}^{n} \left(w_j \alpha_{\sigma(j)} \right) = \left(1 - \prod_{j=1}^{n} \left(1 - \mu_{\sigma(j)} \right)^{w_j}, \prod_{j=1}^{n} \left(v_{\sigma(j)} \right)^{w_j} \right)
\end{aligned}
\tag{10}
$$

where $w = (w_1, w_2, \ldots, w_n)^T$ is the weighting vector, $w_j \in [0, 1]$ and $\sum_{j=1}^{n} w_j = 1, \alpha_{\sigma(j)}$ is the α_j value of the pair $\langle u_j, \alpha_j \rangle$ having the jth largest $u_j (u_j \in [0, 1])$, u_j is referred to as the order inducing variable.

We can obtain the IFWA, IFAA operators by choosing different manifestations of the weighting vector and the order inducing vector $u = (u_1, u_2, \ldots, u_n)$.

3 Integrating the I-IFOWA operator and EOS-ELM for sentiment prediction

3.1 Problem description

The problem of predicting consumer sentiments can be formally defined as an online sequential sentiment prediction problem similar to [3].

The labeled Chinese review corpus consists of N samples $\aleph = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}^m\}_{i=1}^{N}$, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T$ is the random vector corresponding to the features $\{X_1, X_2, \ldots, X_n\}$. In binary-based feature representation, $\mathbf{x}_i$ is derived by encoding the presence or absence of features in the ith review, that is, $x_{ij} \in \{0, 1\}$, $j = 1, 2, \ldots, n$. While in frequency-based feature representation, $\mathbf{x}_i$ is derived by encoding the frequencies of features in the ith review, that is, $x_{ij} \in \{0\} \bigcup Z^+, j = 1, 2, \ldots, n$, Z^+ is the set of positive integer. Similar to [23], we represent the overall sentiment by the class label $\mathbf{t}_i$ coded by m bits: $\mathbf{t}_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T$, where $m = 2$ if "neutral" is exclusive, or

$m = 3$ if "neutral" is inclusive. For a pattern of class k, only t_{ik} is "1" and the rest is "-1," $k = 1, 2, \ldots, m$. The online sequential sentiment prediction problem can be defined as follows.

Online Sequential Sentiment Prediction: A corpus of N reviews arrives sequentially. In initial phase, N_0 reviews $\aleph_0 = \{(\mathbf{x}_i, t_i)\}_{i=1}^{N_0}$ are given. Others would arrive one-by-one or chunk-by-chunk (with fixed or varying chunk size). If $N_0 = N$, sequential prediction reduces to batch prediction. We want to predict the overall sentiment of a new review by a learned mapping $f: \mathbf{R}^n \to \mathbf{R}$.

3.2 Parallel EOS-ELM

Parallel EOS-ELM means OS-ELMs in the ensemble are parallel, thus can be implemented by concurrent technique, each thread of which generates one OS-ELM. A typical parallel EOS-ELM can be found in [17]. In the paper, as we try to use the induced operator to fuse the outputs of OS-ELMs, the order inducing variables needs to be evaluated. Similar to existing fusion methods with weighting procedure, we argue a intuitive hypothesis that "better" classifier should be assigned bigger weight. The concept "better" is measured by some quantitative criteria such as norm of output weights $\|\beta\|$ (just for ELM-correlative method), accuracy, time, etc. It is agreed commonly that smaller $\|\beta\|$ may lead to better generalization performance [21, 39, 10]. Therefore, the smaller $\|\beta\|$ is, the bigger weight should be assigned. Accuracy should be another rational measurement. In some special area, time is a crucial criterion, then we can use time as the order inducing variable. Note that if norm of weights or time is used, then the weighting vector is ascending, whereas if accuracy is used, the weighting vector is descending. Parallel EOS-ELM consists of Q OS-ELM networks with the same number of hidden nodes and the same activation function for each node. The proposed Parallel EOS-ELM is illustrated in Table 1. In order to increase the diversity of the ensemble, some other techniques can be used. For example, each OS-ELM can be trained with distinct samples.

Table 1 Algorithm of parallel EOS-ELM

Input: The data $\aleph = \{(\mathbf{x}_i, \mathbf{t}_i) | \mathbf{x}_i \in \mathbf{R}^n, \mathbf{t}_i \in \mathbf{R}^m\}_{i=1}^{N}$ presents to the network sequentially (one-by-one or chunk-by-chunk); Size of ensemble Q; Number of hidden Nodes L; $N_0 \geq L$.

Output: Q OS-ELMs; order inducing variables u.

Begin

    For $q = 1, 2, \ldots, Q$

        (a) Sequential learning OS$-$ELM$_q$.

        (b) The order inducing variable u_q is set by accuracy of OS$-$ELM$_q$ or $u_q = \left\| \beta^{(q)} \right\|$.

    End for

End

3.3 Intuitionistic fuzzy fusion

When an ensemble is prepared, we could use it to predict class label of a test pattern, that is, an unlabeled Chinese review. The components of outputs of a classifier can be viewed as degree of support, belief, certainty, possibility, etc., not necessarily coming from a statistical classifier [16]. In this paper, we use the I-IFOWA operator for two aspects of purpose. First, we present a generalized form of weighting fusion schemes which include all existing weighting fusion function as special cases of it. Thus, we can compare among different weighting strategies by an uniform formula. Additional, we use IFVs to represent outputs of component classifiers because we try to consider both the degree of a test pattern belongs to a class and the degree of the test pattern does not belong to the class. We develop the following fusion schemes just based on the hypothesis that "better" classifier should be assigned bigger weight.

Let us begin with analyzing the outputs of OS-ELMs. Assume that $\{OS\text{-}ELM_q\}_{q=1}^Q$ is the set of Q OS-ELMs deriving by Parallel EOS-ELM. We simply denote the outputs of those Q OS-ELMs by, given test sample $\mathbf{x} \in \mathbf{R}^n$,

$$DP(\mathbf{x}) = \begin{bmatrix} y_{1,1}(\mathbf{x}) & \cdots & y_{1,m}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ y_{Q,1}(\mathbf{x}) & \cdots & y_{Q,m}(\mathbf{x}) \end{bmatrix}_{Q \times m} \qquad (11)$$

where $DP_q(\mathbf{x}) = [y_{q,1}(\mathbf{x}) \ldots y_{q,m}(\mathbf{x})]$ is the output of OS-ELM_q, such that $y_{q,j}(\mathbf{x}) \in [-1, 1]$ and m is the number of pattern classes. In practice, this interval may not always hold. If $y_{q,j}(\mathbf{x}) \notin [-1, 1]$, we normalize $DP_q(\mathbf{x})$ by $y_{q,j}(\mathbf{x}) = y_{q,j}(\mathbf{x})/\max_j\{|y_{q,j}(\mathbf{x})|\}$. Motivated by [15], we call the matrix decision profile (DP). As described in Sect. 3.1, the ideal DP of an EOS-ELM for a test pattern $\mathbf{x}$ of class j should be:

$$DP(\mathbf{x}) = \begin{bmatrix} -1 & \cdots & 1 & \cdots & -1 \\ \vdots & & \vdots & & \vdots \\ -1 & \cdots & 1 & \cdots & -1 \end{bmatrix}_{Q \times m} \qquad (12)$$

However, the ideal case in (12) would not usually happen because of the fuzzinesses and uncertainties of both training data and test data. Thus, in the decision making phase, our task is to develop a process to handle them. The maximum operator and weighted vote function are rational but elementary resolutions. We utilize theories of IFSs to represent the fuzzinesses and uncertainties by the degree of membership and degree of non-membership, simultaneously.

As we use "1" to represent the correct label and "−1" to represent the incorrect label, the output of OS-ELM_q includes rich information. Let X be the set of samples of class j. The degree of $y_{q,j}(\mathbf{x})$ approximates to "1" figures the degree of $\mathbf{x}$ belongs to class j, denoted by

$$\mu_{q,j}(\mathbf{x}) = P(\mathbf{x} \in X) = \frac{y_{q,j}(\mathbf{x}) - (-1)}{1 - (-1)} = \frac{y_{q,j}(\mathbf{x}) + 1}{2}; \qquad (13)$$

while the degree of $\mathbf{y}_{q,j}(\mathbf{x})$ approximates to "−1" implies the degree of $\mathbf{x}$ does not belong to class j, denoted by

$$v_{q,j}(\mathbf{x}) = P(\mathbf{x} \notin X) = \frac{1 - y_{q,j}(\mathbf{x})}{1 - (-1)} = \frac{1 - y_{q,j}(\mathbf{x})}{2}. \qquad (14)$$

Actually, we calculate the possibilities of both $\mathbf{x}$ belongs to class j and $\mathbf{x}$ does not belong to class j. Because they both indicate some evidences of whether $\mathbf{x}$ belongs to class j, motivated by processes of forming an IFV suggested by [14, 20], we transform $y_{q,j}(\mathbf{x})$ into an IFV $\alpha_{q,j}(\mathbf{x}) = (\mu_{q,j}(\mathbf{x}), v_{q,j}(\mathbf{x}))$ based on the degree of approximation, where $q = 1, 2, \ldots, Q, j = 1, 2, \ldots, m$. The corresponding IFS is

$$A_{q,j} = \left\{ \langle \mathbf{x}, \mu_{q,j}(\mathbf{x}), v_{q,j}(\mathbf{x}) \rangle | \mathbf{x} \in X \right\} \qquad (15)$$

where $\mu_{q,j}(\mathbf{x})$ represents the degree of $\mathbf{x} \in X$, $v_{q,j}(\mathbf{x})$ represents the degree of $\mathbf{x} \notin X$. Note that the transformation enables us to use intuitionistic fuzzy theories to fuse the EOS-ELM. We clarify the equivalence of the transformation in the following theorem.

Theorem 1 *Let $y_{q,j}$ be an entry of outputs of an OS-ELM. IFV $\alpha_{q,j}(\mathbf{x}) = (\mu_{q,j}(\mathbf{x}), v_{q,j}(\mathbf{x}))$ is the transformation result using (13, 14, 15). Then, $y_{q,j}$ is equal to the expected value of the pair $(\mu_{q,j}(\mathbf{x}), v_{q,j}(\mathbf{x}))$. If a piece linear utility function u is assumed, then the utility of $y_{q,j}$ is equal to the expected utility of $\alpha_{q,j}(\mathbf{x})$.*

Proof According to the procedure of transformation, the expected value of $\alpha_{q,j}(\mathbf{x})$:

$$E\left((\mu_{q,j}(\mathbf{x}), v_{q,j}(\mathbf{x}))\right) = -1 \times \frac{1 - y_{q,j}}{2} + 1 \times \frac{y_{q,j} + 1}{2} = y_{q,j}$$

Additional, if a piecewise linear utility function u is assumed for the class of a test pattern. If a test pattern belongs to class j, the utility is denoted by $u(1)$; if a test patten does not belong to class j, the utility is denoted by $u(-1)$. Then, the expected utility of $(\mu_{q,j}(\mathbf{x}), v_{q,j}(\mathbf{x}))$ is:

$$u\left((\mu_{q,j}(\mathbf{x}), v_{q,j}(\mathbf{x}))\right) = \mu_{q,j}(\mathbf{x})u(1) + v_{q,j}(\mathbf{x})u(-1)$$

the utility of $y_{q,j}$ can be calculated by [34]:

$$\begin{aligned} u(y_{q,j}) &= u(1) - \frac{1 - y_{q,j}}{1 - (-1)}(u(1) - u(-1)) \\ &= \frac{y_{q,j} + 1}{2}u(1) + \frac{1 - y_{q,j}}{2}u(-1) \\ &= \mu_{q,j}(\mathbf{x})u(1) + v_{q,j}(\mathbf{x})u(-1) \end{aligned}$$

which complete the proof. □

It is commonly known that any piecewise continuous function can be arbitrarily approximated by piecewise

linear functions. Thus, our transformation results are equivalent to the original data if any piecewise continuous utility function is used.

Therefore, the DP of EOS-ELM can be equivalently transformed into the following form

$$\text{IFDP}(\mathbf{x}) = \begin{bmatrix} \alpha_{1,1}(\mathbf{x}) & \dots & \alpha_{1,m}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \alpha_{Q,1}(\mathbf{x}) & \dots & \alpha_{Q,m}(\mathbf{x}) \end{bmatrix}_{Q \times m} \quad (16)$$

and the qth row, $\alpha_q(\mathbf{x}) = [\alpha_{q,1}(\mathbf{x}), \dots, \alpha_{q,m}(\mathbf{x})]$, represents the output of OS-ELM$_q$, the jth column represents the degree of $\mathbf{x}$ belongs to the jth class and does not belong to the jth class.

The task of classifier combination is to derive the fused output of Q classifiers for final prediction, and the fusion process can be denoted by

$$D(\mathbf{x}) = \Theta(IFDP(\mathbf{x})) = \Theta(\alpha_1(\mathbf{x}), \dots, \alpha_Q(\mathbf{x})) \quad (17)$$

where Θ is regarded as aggregation operator. Recur to the operators mentioned in Sect. 2.2, the order inducing variables $u_q(q = 1, 2, \dots, Q)$ derived in Sect. 3.2 and the weighting vector w of the ensemble, spontaneously, we present the classifier fusion algorithm in Table 2.

There are some differences between the proposed fusion scheme and other fusion schemes. We take into account the degree of support and nonsupport at the same time, whereas existing fusion schemes consider only the degree of support. As seen in Definition 5, we use nonlinear combination rather than linear combination (e.g., sum).

3.4 Decision making using the I-IFOWA operator

All the existing weighting strategies can be seen as special cases of the proposed fusion scheme if Θ in 17 is the I-IFOWA operator implemented by different manifestations of order-inducing variables or weighting vector.

(1) If $w_q = 1/Q$, $q = 1, 2, \dots, Q$, the I-IFOWA operator reduces to the IFAA operator. Thus, component classifiers are indistinguishable, as seen in [17, 27].

(2) If $w_q = u_q / \sum_{q=1}^{Q} u_q$, $q = 1, 2, \dots, Q$, obviously, $\sum_{q=1}^{Q} w_q = 1$ and the order in $\{w_q\}_{q=1}^{Q}$ is equal to the order in $\{u_q\}_{q=1}^{Q}$, thus, the I-IFOWA operator reduces to the IFWA operator, which implement the weighted average strategy as seen in [8, 26].

(3) If $w_q = 2/Q$, $q = 1, 2, \dots, Q/2$ and $w_q = 0$, $q = Q/2 + 1, Q/2 + 2, \dots, Q$, we actually select the first half "better" classifiers from the ensemble for decision making, which is the weighting strategy used in [21]. We can generate some more versions of the I-IFOWA operator based on other fusion strategies.

(4) Ng and Abramson [22] suggested to use the best individual classifier for the final decision. This can be implemented by the maximum operator. The intuitionistic fuzzy maximum operator can be obtained by setting $w_1 = 1$ and $w_q = 0$, $q = 2, \dots, Q$.

(5) Recently, another hot research issue is the selective fusion scheme [38, 12], which could perform better than using all component classifiers. In the intuitionistic fuzzy fusion framework, the I-IFOWA operator may help to select the best M from Q component classifiers by proper weighting vectors. If the M classifiers are treated identically, then, we set the weighting vector as: $w_j = 1/M$, $j = 1, 2, \dots, M$ and $w_j = 0$, $j = M + 1, \dots, Q$; If we want to weight the M classifiers, then weighting strategies can be used.

Other versions of the I-IFOWA operator can also be derived by similar discussion, and other fusion algorithms can be constructed.

3.5 The proposed procedure

Based on preparation above, we present the paradigm of the proposed procedure step by step as a brief illustration. A online predicting system can be implemented for product supplier or consumers based on the paradigm.

Step 1 Initialization. This step includes preparing training corpus and test reviews, selecting classification features,

Table 2 Algorithm of the I-IFOWA operator based fusion

Input: Outputs of L classifiers (DP($\mathbf{x}$)); Order inducing variables $u_q(q = 1, 2, \dots, Q)$; weighting vector w.

Output: class label of $\mathbf{x}$

Begin

    1. Form IFDP($\mathbf{x}$) using (13)–(15).

    2. For $j = 1, 2, \dots, m$

    Aggregate information that support $\mathbf{x}$ belongsto class j or not using the I-IFOWA operator:

$\alpha_j(\mathbf{x}) = \text{I-IFOWA}_w(\langle u_1, \alpha_{1j}(\mathbf{x}) \rangle, \dots, \langle u_Q, \alpha_{Qj}(\mathbf{x}) \rangle)$

    End for

    3. The predict label: $y(\mathbf{x}) = \arg\max_{j}\{\alpha_j(\mathbf{x})\}$

End

forming input vector for each sample and normalization if needed. The training reviews and test reviews should be relevant to the same topic written in Chinese. For example, if we want to predict reviews sentiments about a hotel, then the training reviews should also be reviews about hotels, even hotels in the same city or exactly the same hotel.

Step 2 Training. Train EOS-ELM by Parallel EOS-ELM using training corpus. the outputs of training phase are an ensemble of ELM network models with distinct input and output parameters and order inducing variables.

Step 3 Prediction. Given a test pattern $\mathbf{x}$, utilize the trained EOS-ELM to generate Q outputs and construct the $IFDP(\mathbf{x})$ in (16). Then, the final decision could be made using the I-IFOWA operator based fusion schemes.

Step 4 Conclusion. After all reviews are labeled, the overall consumer sentiments, as the most important index for product supplier and consumers, can be calculated by evaluating the ratio of positive reviews. Then, an alarm mechanism can be set up for product manager, while the threshold of positive ratio should be confirmed according to management science.

4 Performance evaluation

4.1 Experimental setup

Experiments are conducted on three open datasets in distinct domains. Datasets [29] used in this section are summarized in Table 3. Each dataset has 2,000 positive samples and 2,000 negative samples. For each simulation, we randomly select 60% samples of each class for training, and the rest are used for testing.

As feature selection and reduction are not what we concerned in this study, we simply use algorithms proposed in literatures to obtain the input vector of a sample. We combine sets of n adjacent Chinese characters into n-gram features in a review as in [36, 35]. As bigrams usually outperform unigrams and trigrams [37], we use only bigrams in this study. Then, each Chinese review is represented by a vector with each value of which is fixed by binary-based weight, due to the better results reported in [37, 31]. Obviously, there are a big amount of features need to be reduced so that the complexity of computation can be alleviated and the accuracy may be improved as well.

Therefore, we adopt a feature selection method based on improved fisher's discriminant ratio [30] for finally selecting a certain size of features. This method takes use of the conditional mean and conditional variances in statistics to improve the famous fisher linear discriminant. We construct feature set for each dataset, independently, and select 1,000, 1,500 and 1,000 features for HOTEL, BOOK, and NBC, respectively.

In the following experiments, the additive hidden node is used, the activation function of ELM is the sigmoidal function: $y = 1 (1 + e^{-x})$. The input weights and biases are randomly generated from the range $[-1, 1]$. For initialization phase, the number of training data N_0 is $L + 100$, where L is the network size.

Some criteria, such as Training times, Testing times and Standard Deviation (SD), are used to evaluate performances of the proposed methodology and compare with other existing method. Each unit of experiments runs 50 trials for the mean and SD are derived.

4.2 Experiments

We conduct all experiments in a Matlab 7.7.0 environment running in a desktop with CPU 2.93 GHz, and 3GB RAM. The code of ELM [40] is downloaded from Internet, we modified it to output soft predicted labels.

We first focus on parameters of ELM and OS-ELM. There is only one parameter for ELM, that is,number of hidden nodes, need to be optimized. We conduct 50 trials of simulations to get the mean accuracy for each dataset with a series of numbers to find out the optimal choice. Table 4 shows the simulations. The first row of Table 4 means numbers of hidden nodes. The optimal number of hidden nodes are selected as 200 in all three datasets. We did not select bigger number as it may lead to computational complexity though it presents more accurate results. The number of hidden nodes in each OS-ELM in EOS-ELM is set to be equal to that in corresponding dataset as well.

SVM and NB are the most frequently used machine learning methods in recent researches of sentiment classification. SVM, as a family of supervized learning algorithms which select models that maximize the error margin of a training set, have been successively used in many pattern classification problems. We impose the Matlab

Table 3 Summary of three datasets

Dataset name	Domain	# Samples	# Features
HOTEL	Hotel	2,000 + 2,000	1,000
BOOK	Book	2,000 + 2,000	1,500
NBC	Notebook	2,000 + 2,000	1,500

Table 4 Accuracies with distinct numbers of hidden nodes using ELM

Dataset	40	80	120	160	200	240
HOTEL	72.03	75.05	76.86	78.75	80.13	81.09
BOOK	65.57	69.31	72.61	73.48	75.52	76.43
NBC	71.50	76.27	78.40	79.33	80.82	80.98

code for the SVM classifier from [4]. NB assumes that features are mutually independent. Then, conditional probabilities can be simply calculated by $P(\mathbf{x}|\mathbf{y}_j) = P([\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]|\mathbf{y}_j) \approx \prod_{k=1}^{m} P(\mathbf{x}_k|\mathbf{y}_j)$. Then, final decision of NB can be obtained by the maximum index of $\prod_{k=1}^{m} P(\mathbf{x}_k|\mathbf{y}_j)P(\mathbf{y}_j)$, where $j = 1, 2, \ldots, m$. Laplace smoothing is used to prevent infrequently occurring words form being zero probabilities.

We conducted experiments with ELM, OS-ELM, SVM, and NB in Chinese review sentiments prediction in three datasets. Table 5 shows numerical results of several criteria.

In Table 5, it is clear that ELM and OS-ELM are effective algorithms for sentiment prediction of Chinese reviews. In order to improve performances of ELM and OS-ELM, we exam the proposed fusion scheme, comparing with some existing fusion schemes. As we have on idea about that which of norm of output weight and accuracy is better to measure component classifier, we try both measurements in Parallel EOS-ELM. Specific fusion methods are categorized into two groups.

Group 1 includes some existing methods which use straightway outputs of component classifiers:

(1) Arithmetic average (AA) As can be seen in [17], the final prediction can be made by $\arg\max \frac{1}{Q}\sum_{q=1}^{Q} f^{(q)}(\mathbf{x})$, where $f^{(q)}(\mathbf{x})$ is the output of OS-ELM$_q$ with the input pattern $\mathbf{x}$.

(2) Weighted average (WA) Different from AA, weights of component classifiers are taken into account. Frequently, weights are represented by accuracies of component classifiers.

(3) First half arithmetic average (FHAA) Liu and Wang [21] used the first half "better" component classifiers

for final decision. The first half is induced by norm of output weights. But the first half is not distinguished with each other.

Group 2 is formed by proposed intuitionistic fuzzy fusion scheme with some strategies of weighting and two strtegies of order inducing. These kinds of methods are implemented by the I-IFOWA operator. Outputs of component classifiers are transformed into IFVs at first. Corresponding to Group 1, it includes:

(1) Intuitionistic fuzzy arithmetic average (IFAA) The ensemble is formed by Parallel EOS-ELM, and outputs of OS-ELMs are transformed into IFPD before fusion. Similarly, we have

(2) NIFWA represents the intuitionistic fuzzy weigh-ted average (IFWA) method of Parallel EOS-ELM which order inducing variables and weighting vector are obtained by norms of output weighting, actually, we weight the ensemble of OS-ELMs by norms of output weighting; if order inducing variables and weighting vector are obtained by accuracies of OS-ELMs, IFWA is denoted by AIFWA, that is, we weight the ensemble by accuracies in this case.

(3) NIFFHAA represents the method of using I-IFOWA operator to fuse the first half "better" component classifiers when order inducing variables are obtained by norms of output weighting; AIFFHAA represents the method of using I-IFOWA operator to fuse the first half "better" component classifiers if order inducing variables are obtained by accuracies of OS-ELMs.

Note that order inducing variables is useful in (3) of Group 2 only. How to use the I-IFOWA operator to implement Group 2 has been mentioned in Sect. 3.4.

Table 5 Performances on three datasets of ELM, OS-ELM, SVM and NB

Dataset	Algorithm	Time (s)				Accuracy (%)			
		Training	SD	Testing	SD	Training	SD	Testing	SD
HOTEL	ELM	**0.5677**	0.0270	0.1036	0.0145	**82.43**	0.0070	80.13	0.0093
	OS-ELM	0.8870	0.0303	**0.1021**	0.0134	82.38	0.0085	**80.46**	0.0093
	SVM	9.4807	0.1390	3.7734	0.0630	80.74	0.71	79.36	0.41
	NB	21.45	0.3127	38.71	0.1089	75.09	0.99	70.09	0.99
BOOK	ELM	**0.6448**	0.0197	0.1359	0.0149	78.29	0.0114	75.52	0.0137
	OS-ELM	1.0109	0.0266	**0.1344**	0.0151	78.42	0.0103	76.09	0.0131
	SVM	11.35	0.1575	6.062	0.0822	**81.27**	0.7695	**80.77**	0.4027
	NB	32.05	0.5768	58.22	0.1155	79.83	0.0107	74.83	0.0107
NBC	ELM	**0.5672**	0.0225	**0.1000**	0.0146	**83.57**	0.0060	80.82	0.0088
	OS-ELM	0.8802	0.0202	0.1057	0.0163	83.47	0.0076	80.54	0.0105
	SVM	8.6000	0.1286	3.5922	0.0595	82.72	0.8347	**81.29**	0.4153
	NB	21.56	0.3556	38.75	0.1232	80.43	0.0081	75.57	0.0081

Bold values indicate the best performances, the highest accuracies or the smallest times

We test these 8 fusion methods on each dataset, the performances are shown in Tables 6, 7, 8, respectively. In these tables, accuracies of best individual classifier (BIC) and Oracle are also presented. The "Oracle" work as follows: assign the correct class label to **x** if and only if at least one individual classifier produces the correct class label of **x** when its decision is hardened. Numbers in the first row of Tables 6, 7, 8 mean sizes of ensemble.

5 Discussions

Based on numerical results, we make in-depth discussion about the questions raised in Sect. 1. Accuracies obtained in Sect. 4 are higher than using the same dataset (HOTEL) in [41], which proofs that the feature selection process is effective.

5.1 Effectiveness of single classifiers

As can be seen in Table 5, ELM produces the highest training accuracies on HOTEL and NBC, and OS-ELM wins the highest testing accuracy on NBC. SVM seems to be the best classifier on BOOK as both training accuracies and testing accuracies are much higher than others. While the SDs of ELM, OS-ELM and NB are much lower than that obtained by SVM, which means ELM, OS-ELM, and NB are more stable than SVM. Moreover, training time and testing time are significant criteria in practical applications, especially in ensemble learning. Table 5 shows that ELM and OS-ELM are at least 10 times faster than SVM, 30 times faster than NB with respect to training time. Moreover, the advantage of testing time of ELM and OS-ELM is more prominent. Regrading SDs of training time and testing time, ELM and OS-ELM are more outstanding. Generally, we can argue that ELM and OS-ELM are effective in the application of Chinese review sentiment prediction. Similar to classification result of [19], OS-ELM cost slightly more training time than ELM. The testing time of them are almost equivalent. In most case, accuracies of OS-ELM outperforms that of ELM. This is another reason why we choose OS-ELM.

5.2 Effectiveness of ensemble learning

Comparing values of BIC in Tables 6, 7, 8, we can see that each of the fusion methods improve performances on the

Table 6 Accuracies of fusion methods with respect to size of ensemble on HOTEL (%)

Group	Algorithm	4	10	20	30	40	50	60
Group 1	BIC	79.44	79.99	80.02	80.66	80.35	80.63	80.47
	AA	**81.52**	82.29	82.10	82.66	82.40	82.76	82.52
	WA	81.50	82.30	82.11	82.63	82.39	82.76	82.53
	FHAA	80.24	82.09	82.05	82.47	82.19	82.94	82.51
Group 2	IFAA	81.34	**82.47**	**82.80**	**83.73**	**83.44**	**83.29**	82.98
	NIFWA	81.32	81.95	82.26	81.60	81.60	81.77	81.32
	AIFWA	81.33	82.04	82.33	82.96	82.72	82.28	81.71
	NIFFHAA	80.65	82.11	82.43	82.89	82.97	83.23	**83.40**
	AIFFHAA	80.61	81.97	82.30	82.63	82.21	81.64	81.02
	Oracle	89.84	82.51	93.31	93.95	94.04	94.14	94.30

Bold values indicate the best performances, the highest accuracies or the smallest times

Table 7 Accuracies of fusion methods with respect to size of ensemble on BOOK (%)

Group	Algorithm	4	10	20	30	40	50	60
Group 1	BIC	76.38	76.64	76.94	76.27	76.87	76.66	76.92
	AA	**78.78**	80.68	81.25	81.65	81.65	81.76	81.87
	WA	**78.78**	**80.69**	81.25	81.66	81.66	81.75	81.86
	FHAA	78.03	79.46	80.78	81.21	81.72	81.94	82.04
Group 2	IFAA	78.61	80.58	**81.70**	**82.28**	82.53	82.85	**83.08**
	NIFWA	77.94	78.68	79.74	79.43	78.67	78.13	77.45
	AIFWA	78.12	80.42	81.51	81.96	**82.54**	82.37	82.63
	NIFFHAA	78.33	80.04	80.73	81.41	82.03	**82.91**	82.66
	AIFFHAA	78.28	79.69	79.84	80.45	80.42	80.57	80.38
	Oracle	92.70	96.17	97.24	97.42	97.65	97.62	97.70

Bold values indicate the best performances, the highest accuracies or the smallest times

Table 8 Accuracies of fusion methods with respect to size of ensemble on NBC (%)

Group	Algorithm	4	10	20	30	40	50	60
Group 1	BIC	80.96	81.24	81.38	81.54	81.76	81.56	81.72
	AA	**82.56**	83.28	83.79	83.57	83.55	83.63	83.90
	WA	**82.56**	83.27	83.79	83.57	83.55	83.63	83.91
	FHAA	80.08	81.20	81.59	82.45	82.47	82.53	82.51
Group 2	IFAA	82.53	83.20	**83.93**	**83.85**	83.58	83.64	83.93
	NIFWA	**82.56**	83.03	83.46	83.05	82.96	82.86	82.57
	AIFWA	82.51	**83.29**	83.72	83.82	83.46	83.25	82.93
	NIFFHAA	81.51	82.92	83.59	83.50	**83.61**	**83.72**	**83.94**
	AIFFHAA	81.25	82.06	81.87	82.34	82.81	82.28	82.13
	Oracle	89.94	92.21	92.88	93.33	93.44	93.52	93.55

Bold values indicate the best performances, the highest accuracies or the smallest times

basis of best individual classifiers. The highest improvements of three datasets are 3.09%, 6.25%, and 2.55% respectively, all of which are obtained by fusion methods proposed in this paper. While the lowest improvements of three datasets are 2.08%, 2.40%, and 1.60% respectively. Further, the highest improvements occur when sizes of ensemble are 40, 30, and 20, respectively. The gaps between BIC and Oracle mentioned the potential improvements of fusion. However,, gaps between BIC and Oracle are much bigger than gaps between BIC and the fusion methods, which means fusion methods used in experiments did not improve so much. This may caused by many reasons such as dependence of component classifiers, noises of data, etc.

5.3 Regarding fusion methods

Fusion methods with the same weighting strategies are further compared. With regard to AA and IFAA, it is clear that accuracies obtained by IFAA are always higher than that obtained by AA when size of ensemble is bigger than 20. Thus, AA can only suit small size of ensemble, while IFAA performs much better if the size is bigger. That is, if the size of ensemble increases, IFAA suffers from the followed noise and conflict much milder than AA. AIFWA outperforms WA in most of cases in HOTEL and BOOK whereas NIFWA fails to outperform WA in many cases. NIFFHAA performs better than FHAA 20 times and worse than FHAA once. However, AIFFHAA performs better than FHAA 10 times and worse than FHAA 11 times. Thus, we conclude that the intuitionistic fuzzy fusion methods proposed in this paper is generally better than three existing fusion methods.

5.4 Regarding weighting and order inducing strategies

We now focus on how weighting strategies or order inducing strategies influence the performance of fusion. First, the existing weighting methods (weighting by accuracy and

weighting by norms of output weight) are invalid in Chinese reviews sentiment prediction. In other words, arithmetic average is the best weighting strategy. Thus, highest accuracies occur 3 times when AA is used, 10 times when IFAA is used, 5 times when NIFFHAA is used, only 3 times when methods with weighting strategies are used. AA performs nearly equivalent to WA because differences of accuracies of component classifiers are very low, thus the accuracy benefit nearly nothing from the weighting process. Second, accuracy performs better if it is used as weight of component classifier (approximate the performance of AA), performs worse if it is used as order inducing variable. This is concluded by the fact that NIFAA performs even worse when the size of ensemble is bigger in all three datasets. Moreover, norm of output weight performs better if it is used as order inducing variable (inducing the first half could obtain considerable performances), performs worse if it is used as weight of component classifier. Finally, the selective methods usually outperforms methods fusing the whole ensemble though it actually use partial (a half) component classifiers for fusion. FHAA outperforms AA for 4 times and NIFFHAA outperforms IFAA for 5 times.

6 Conclusions

This paper has combined OS-ELM and intuitionistic fuzzy theories to develop a novel ensemble learning methodology to conduct this problem. Under the framework of intuitionistic fuzzy fusion, an ensemble of OS-ELMs as well as their order inducing variables are parallelly trained at first. Then, outputs of OS-ELMs are equivalently transformed into IFPD of the test pattern. Associated with proper order inducing variables and weighting vector, the I-IFOWA operator is used to fusion degree of membership and non-membership simultaneously. It has been shown that the existing fusion strategies are just the special cases of the generalized framework of ensemble learning. The proposed

scheme has been successfully used in Chinese review sentiment prediction with the excellent performance.

References

1. Alexandre LA, Campilho AC, Kamel M (2001) On combining classifiers using sum and product rules. Pattern Recogn Lett 22(12):1283–1289
2. Atanassov K (1986) Intuitionistic fuzzy sets. Fuzzy Set Syst 20(1):87–96
3. Bai X (2011) Predicting consumer sentiments from online text. Decis Support Syst 50(4):732–742
4. Chang C, Lin C (2001) Libsvm: a library for support vector machines. In: http://www.csie.ntu.edu.tw/∼cjlin/libsvm
5. Cheung C, Shek S, Sia C (2004) Virtual community of consumers:why people are willing to contribute? In: Proceeding of the 8th Pacific-Asia conference on information systems, pp 2100–2107
6. Cui H, Mittal V, Datar M (2006) Comparative experiments on sentiment classification for online product reviews. In: Proceedings of the twenty-first national conference on artificial intelligence(AAAI). Boston
7. Dellarocas C (2003) The digitization of word of mouth: Promise and challenges of online feedback mechanisms. Manage Sci 49(10):1407–1424
8. Heeswijk M, Miche Y, Lindh-Knuutila T, Hilbers P, Oja THE, Lendasse A (2009) Adaptive ensemble models of extreme learning machine for time series prediction. In: International conference on artificial neural networks (ICANN2009), vol. 5769, pp. 305–314
9. Huang GB, Zhu QY, Mao KZ, Siew CK, Saratchandran P, Sundararajan N (2006) Can threshold networks be trained directly?. IEEE Trans Circuits Syst II 53(3):187–191
10. Huang GB, Zhu QY, Siew CK. Extreme learning machine: Theory and applications. Neurocomputing 70
11. Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: A new learning scheme of feedforward neural networks. In: Proceedings of international joint conference on neural networks (IJCNN2004), vol 2, pp 985–990. Budapest, Hungary (25–29 July, 2004)
12. Jackowski K, Wozniak M (2010) Method of classifier selection using the genetic approach. Expert Syst 27(2)
13. Joachims T, Nedellec C, Rouveirol C (1998) Text categorization with support vector machines: learning with many relevant. In: Proceedings of the European conference on machine learning, vol 1398. Dortmund, Germany, pp 137–142
14. Khatibi V, Montazer GA (2009) Intuitionistic fuzzy set vs fuzzy set application in medical pattern recognition. Artif Intell Med 47(1):43–52
15. Kuncheva L (2001) Using measures of similarity and inclusion for multiple classifier fusion by decision templates. Fuzzy Set Syst 122(3):401–407
16. Kuncheva L, Bezdek J, Duin R (2001) Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recogn 34(2):299–314
17. Lan Y, Soh Y, Huang G (2009) Ensemble of online sequential extreme learning machine. Neurocomputing 72:3391–3395
18. Lewis D (1998) Naive (bayes) at forty: the independence assumption in information retrieval. Lect Note Comput Sci 1398:4–18
19. Liang N, Huang G, Saratchandran P, Sundararajan N (2006) A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Trans Neural Netw 17(6):1411–1423
20. Liu HW, Wang GJ (2007) Multi-criteria decision-making methods based on intuitionistic fuzzy sets. Eur J Oper Res 179(1):220–233
21. Liu N, Wang H (2010) Ensemble based extreme learning machine. IEEE Signal Process Lett 17(8):754–757
22. Ng K, Abramson B (1992) Consensus diagnosis: An simulation study. IEEE Trans Syst Man Cybern B Cybern 22(5):916–928
23. Rong H, Huang GB, Ong Y (2008) Extreme learning machine for multi-categories classification applications. In: 2008 international joint conference on neural networks (IJCNN2008), Hong Kong, pp 1709–1713
24. Rong H, Ong Y, Tan A, Zhu Z (2008) A fast pruned-extreme learning machine for classification problem. Neurocomputing 72:359–366
25. Rong HJ, Huang GB, Sundararajan N, Saratchandran P (2009) Online sequential fuzzy extreme learning machine for function approximation and classification problems. IEEE Trans Syst Man Cybern B Cybern 39(4):1067–1072
26. Sun Z, Au K, Choi T (2007) A neuro-fuzzy inference system through integration of fuzzy logic and extreme learning machines. IEEE Trans Syst Man Cybern B Cybern 37(5):1321–1331
27. Sun Z, Choi T, Au K, Yu Y (2008) Sales forecasting using extreme learning machine with applications in fashion retailing. Decis Support Syst 46(1):411–419
28. Tan S, Zhang J (2008) An empirical study of sentiment analysis for chinese documents. Expert Syst Appl 34(4):2622–2629
29. Tan SB (2010) Chinese sentiment corpuses. In: http://www.searchforum.org.cn/tansongbo/corpus-senti.htm
30. Wang S, Li D, Song X, Wei Y, Li H (2011) A feature selection method based on improved fisher's discriminant ratio for text sentiment classification. Expert Syst Appl 38(7):8696–8702
31. Xia R, Zong C, Li S (2011) Ensemble of feature sets and classification algorithms for sentiment classification. Inf Sci 181(6):1138–1152
32. Xu Z (2007) Intuitionistic fuzzy aggregation operators. IEEE Trans Fuzzy Syst 15(6):1179–1187
33. Xu Z, Da Q (2003) An overview of operators for aggregating information. Int J Gen Syst 18(9):953–969
34. Yang J, Sen P (1996) Preference modelling by estimating local utility functions fro multiobjective optimisation. Eur J Oper Res 95(1):115–138
35. Ye Q, Lin B, Li Y (2005) Sentiment classification for chinese reviews: A comparison between svm and semantic approaches. In: Proceedings of the fourth international conference on machine learning and cybernetics (ICMLC2005), vol. 4
36. Ye Q, Zhang Z, Law R (2009) Sentiment classification of online reviews to travel destinations by supervised machine learning approaches. Expert Syst Appl 36(3):6527–6535
37. Zhang Z, Ye Q, Zhang Z, Li Y (2011) Sentiment classification of internet restaurant reviews written in cantonese. Expert Syst Appl 38(6):7674–7682
38. Zhou Z, Wu J, Tang W (2002) Ensembling neural networks: Many could be better than all. Artif Intell 137:239–263
39. Zhu Q, Qin A, Suganthan P, Huang G (2005) Evolutionary extreme learning machine. Pattern Recogn 38(10):1759–1763
40. Zhu QY, Huang GB (2004) Source codes of ELM algorithm. In: http://www.ntu.edu.sg/home/egbhuang/. School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
41. Zhai ZW, H, Xu PJ (2010) An empirical study of unsupervised sentiment classification of chinese reviews. Tsinghua Sci Technol 15(6): 702–708