Neurocomputing 74 (2011) 2520–2525

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Upper integral network with extreme learning mechanism

Xizhao Wang *, Aixia Chen, Huimin Feng

College of Mathematics and Computer Science, Hebei University, Baoding 071002, China

ARTICLE INFO

Available online 12 May 2011

Keywords:
Fuzzy measure
Fuzzy integral
Upper integral
Multi-attribute classification
Possibility distribution
Extreme learning machine

ABSTRACT

The upper integral is a type of non-linear integral with respect to non-additive measures, which represents the maximum potential of efficiency for a group of features with interaction. The value of upper integrals can be evaluated through solving a linear programming problem. Considering the upper integral as a classifier, this paper first investigates its implementation and performance. Fusing multiple upper integral classifiers together by using a single layer neural network, this paper considers a upper integral network as a classification system. The learning mechanism of ELM is used to train this single layer neural network. A comparison of performance between a single upper integral classifier and the upper integral network is given on a number of benchmark databases.

1. Introduction

The classical integral has two significant features: its measure is additive and its integrand has linear property. During the recent several decades, various extensions ranging from the non-additive measures to non-linear integrand have been addressed by many scholars. The following is three typical cases. In the early seventies, Sugeno [1] introduced the concept of fuzzy measure and fuzzy integral, which generalized the usual definition of a measure by replacing the additivity property with a weaker condition, i.e. the monotonicity. The Choquet integral, which is a direct extension of the Lebesgue integral, was presented by Murofushi and Sugeno [2]. The upper and lower integrals, which aim to describe and formulate the interaction existing in a group of features, were given in [3,4] by Wang et al. This paper makes an attempt to investigate the implementation and performance of the upper integral as a classifier.

Fusion of multiple classifiers can improve the performance of classification [5]. Any integral defined on a finite space can be used as an aggregation tool in information fusion [6]. And it is also used in non-linear classification, which may be understood as a multi-regression with categorical objective attribute [7]. Recently, there have been several attempts to use fuzzy integral as an aggregation operator in classification problems [8–12,24], and the results were quite encouraging. In this kinds of methods, a fusion of decisions of different sensors or classifiers are merged into a final decision by a non-linear integral with respect to a non-additive measure. A fuzzy measure expresses the weights of each sensor or each classifier for a given class, and the weights on each subset of sensors or classifiers as well. In the non-linear classification, Sugeno integral with respect to λ-measures or possibility measures was used by Keller [10,11], while Choquet integral with respect to any fuzzy measures was used by Grabish and Sugeno [8].

In this paper, we will use upper integral with respect to nonnegative set function in non-linear classification. The motivation is stated as follows. It is well-known that any indeterminate integral of a nonnegative function can be identified by a decomposition [3]. The integration value is the inner product of two (2^n-1) dimensional vectors: one is the set function and the other is a decomposition of the integrand. The upper integral is an extreme decomposition by which the largest integration values can be obtained [3]. We expect that the multi-attribute classification using upper integral have better performances.

From references we can see a number of approaches to determine the set function, such as [2,6,8,13,15,16,18]. Each approach has its own advantages and disadvantages and it is hard to generally say which one is better. It implicitly indicates that there is still a need to explore new approaches to determine set functions for non-linear integral. This paper makes a first attempt to select set functions and to establish an upper integral network by using ELM.

From the existing references we know that ELMs have the strong capability of learning and approximation. Form example, Huang et al. in [26] discussed the universal approximation property by using incremental constructive feedforward networks with random hidden nodes. And then, Huang et al. investigated the convex incremental ELM and the enhanced random search based incremental ELM in [27,28] respectively. Furthermore, Feng et al. in [29] addressed the error minimized ELM with growth of hidden nodes. It is expected that the ELM can provide a better way to establish upper integral classifier system.

In the following, we briefly present non-additive measures and fuzzy integrals at first, and then introduce the general background of classification based on fuzzy integrals. After this, our method

* Corresponding author.
E-mail address: xizhaowang@ieee.org (X. Wang).

0925-2312/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2010.12.034

based on upper integral and possibility theory is presented. Lastly, some tests on real data are provided.

2. Fuzzy measures and fuzzy integrals

In this section, we will present some basic definitions which are used in this paper. The definitions of fuzzy measures and integrals will be presented in the restrictive case of finite spaces since here we deal with the feature spaces which are usually finite.

2.1. Fuzzy measures

Fuzzy measures have been introduced by Sugeno [1] in the early seventies.

Definition 2.1. Let X be a non-empty finite set and $\mathcal{F}$ be the power set of X. A fuzzy measure μ defined on the measurable space $(X,\mathcal{F})$ is a set function $\mu:\mathcal{F}\to[0,1]$ verifying the following axioms:

(1) $\mu(\emptyset)=0,\mu(X)=1$
(2) $A\subset B\Rightarrow \mu(A)\le\mu(B)$
$(X,\mathcal{F},\mu)$ is said to be a fuzzy measure space.

Fuzzy measures include particular cases as probability measures, possibility and necessity measures, belief measures and plausibility measures, etc.

2.2. Fuzzy integrals

The notion of fuzzy integral, which is the logical continuation of the notion of fuzzy measure, has also been introduced by Sugeno [1]. Another non-linear integral, the so called Choquet integral, which is the extension of the Lebesgue integral, has been presented by Murofushi and Sugeno [2]. The upper integral and the lower integral, which are two extreme specified indeterminate integrals, has been introduced by Wang et al. [4].

Definition 2.2. Let $(X,\mathcal{F},\mu)$ be a fuzzy measure space, with $X=\{x_1,x_2,\dots,\ x_n\}$. Let f be a measurable function from X to $[0,1]$, and without loss of generality, assume that $0\le f(x_1)\le f(x_2)\le\cdots f(x_n)\le 1$, and $A_i=\{x_i,x_{i+1},\dots,x_n\}$. The Sugeno integral and the Choquet integral of $f$with respect to the measure μ are defined respectively as

$$(S)\int f\,d\mu=\bigvee_{i=1}^{n}(f(x_i)\wedge\mu(A_i))$$

$$(C)\int f\,d\mu=\sum_{i=1}^{n}(f(x_i)-f(x_{i-1}))\mu(A_i)$$

where $f(x_0)=0$. The symbols $\wedge$ and $\vee$ denote the minimum and maximum operators, respectively.

Definition 2.3. $X=\{x_1,x_2,\dots,x_n\}$, and we take its power set as $\mathcal{F}$. In this case, any function defined on X is measurable. The upper integral and the lower integral of f with respect to the set function μ can be defined as follows

$$(U)\int f\,d\mu=\sup\left\{\sum_{j=1}^{2^n-1}\lambda_j\cdot\mu(E_j)|f=\sum_{j=1}^{2^n-1}\lambda_j\cdot\chi_{(E_j)},\ \lambda_j\ge 0\right\}$$

$$(L)\int f\,d\mu=\inf\left\{\sum_{j=1}^{2^n-1}\lambda_j\cdot\mu(E_j)|f=\sum_{j=1}^{2^n-1}\lambda_j\cdot\chi_{(E_j)},\ \lambda_j\ge 0\right\}$$

where some λ_j may be zero, and $\chi_{(E_j)}$ is the characteristic function of set $E_j, j=1,2,\dots,2^n-1$, E_j are subsets of X arranged in such a way: the binary expression of $j,(j)_2=b_n^{(j)},b_{n-1}^{(j)},\dots,b_1^{(j)}$, is determined by

$$b_i^{(j)}=\begin{cases}1, & x_i\in E_j,\\ 0, & x_i\notin E_j.\end{cases}$$

That is $E_1=\{x_1\},E_2=\{x_2\},E_3=\{x_1,x_2\},E_4=\{x_3\},E_5=\{x_1,x_3\},E_6=\{x_2,x_3\}$, $E_7=\ \{x_1,x_2,x_3\},\dots$.

The upper integral is a type of non-linear integral with respect to non-additive measures, which represents the maximum potential of efficiency for a group of features with interaction. When a set function μ and an integrand f are given, the calculation of the upper integral or the lower integral is essentially a linear programming problem. In this paper, we consider the implementation and performance of the upper integral classifier. In the following, we will present some advantages of upper integral by an classical example.

Example 2.1. [6] There are three workers a, b and c working for $f(a)=10, f(b)=15$ and $f(c)=7$ days, respectively, to manufacture a kind of products. Their efficiencies of working alone are 5, 6 and 8 products per day, respectively. Their joint efficiencies are not the simple sum of the corresponding efficiencies given above, but are listed in Table 1.

Here, inequality $\mu(\{a,b\})>\mu(\{a\})+\mu(\{b\})$ means that a and b have good cooperation. Similarly, a and c have bad relationship and are not suitable for working together.

(1) Without a manager, they begin to work from the same day. During the first $f(c)$ days, all workers work together with efficiency $\mu(\{a,b,c\})$, and during next $f(a)-f(c)$ days, workers a and b work together with efficiency $\mu(\{a,b\})$, during the last $f(b)-f(a)$ days, only b works with efficiency $\mu(\{b\})$. Thus, the total number of products manufactured by these workers during these days is the Choquet integral of f with respect to μ.

$$(C)\int f\,d\mu=f(c)\mu(\{a,b,c\})+[f(a)-f(c)]\mu(\{a,b\})+[f(b)-f(a)]\mu(\{b\})$$
$$=198$$

(2) With a manager, the maximizing total products is 236, which is in fact the value of the upper integral $(U)\int f d\mu$. The optimal schedule is: a and b work together for 10 days, b works with c for 5days, and c works alone for 2 days. The upper integral is evaluated through solving the following linear programming problem.

$$\begin{aligned}\max\quad & 5\mu_1+6\mu_2+14\mu_3+8\mu_4+7\mu_5+16\mu_6+18\mu_7\\ s.t.\quad & \mu_1+\mu_3+\mu_5+\mu_7=10\\ & \mu_2+\mu_3+\mu_6+\mu_7=15\\ & \mu_4+\mu_5+\mu_6+\mu_7=7\\ & \mu_j\ge 0, j=1,\dots,7\end{aligned}$$

Here, we can see that the value of the upper integral is greater than the value of Choquet integral. The upper integral represents the maximum potential of efficiency for a group of features

Table 1
The joint efficiencies of a, b and c.

Workers	Products per day
$\{a,b\}$	14
$\{a,c\}$	7
$\{b,c\}$	16
$\{a,b,c\}$	18

with interaction. So, we attempt to characterize the performance of the classifier based on upper integral.

2.3. Properties of fuzzy integrals

Generally speaking, the fuzzy integrals can be considered as N-place operators, and we will denote $\int f\,d\mu$ by $\mathsf{F}_\mu(a_1,\ldots,a_n)$ with $a_i = f(x_i), i=1,\ldots,n$. In this subsection, we review some properties of fuzzy integrals.

Property 2.1. *For every* F_μ, $\mathsf{F}_\mu(a,\ldots,a)=a$.

Property 2.2. *For the particular measure* μ_{min} *defined by* $\forall B \in \mathcal{F}, B \neq X, \mu_{min}(B)=0$, *and* $\mu_{min}(X)=1$ (*resp.* μ_{max} *defined by* $\forall B \in \mathcal{F}, B \neq \emptyset, \mu_{max}(B)=1$, *and* $\mu_{max}(\emptyset)=0$) *reduces to the minimum operator* (*resp. maximum*).

Property 2.3. *Let f,g be two functions on X and* μ *be measure on* $(X,\mathcal{F})$. *Then,* $f(x) \leq g(x)$ *for every* $x \in X \Rightarrow \mathsf{F}_\mu(f) \leq \mathsf{F}_\mu(g)$.

Property 2.4. *Let* μ,ν *be two measures on* $(X,\mathcal{F})$. *Then,* $\mu(B) \leq \nu(B)$ *for every* $B \in \mathcal{F} \Rightarrow \mathsf{F}_\mu(f) \leq \mathsf{F}_\nu(f)$.

Property 2.5. *Using Properties* 2.3 *and* 2.4, *we can deduce*

$$\bigwedge_{i=1}^{n}(a_1,\ldots,a_n) \leq \mathsf{F}_\mu(a_1,\ldots,a_n) \leq \bigvee_{i=1}^{n}(a_1,\ldots,a_n)$$

Property 2.6. *For every additive measure* μ, *the Choqute integral reduces to the usual Lebesgue integral, i.e.* $(C)\int f\,d\mu = \sum_{i=1}^{n} f(x_i) \times \mu(\{x_i\})$.

Property 2.7. *The Sugeno integral is median*

$$(S)\int f\,d\mu = median(f(x_1),\ldots,f(x_n),\mu(A_2),\ldots,\mu(A_n))$$

with $A_i = \{x_i,x_{i+1},\ldots x_n\}$ *as before.*

Property 2.8. *For every fuzzy measure, fuzzy integrals are continuous functionals, i.e. for every sequence of functions* $\{f_n\}_{n\in N}$ *on X we have*

$$\lim_{n\to\infty}\mathsf{F}_\mu(f_n) = \mathsf{F}_\mu(\lim_{n\to\infty} f_n)$$

Referring to [3,4], we find that the upper integral and lower integral not only have the Properties 2.3 and 2.4, but also the following Properties 2.9–2.11.

Property 2.9. *For any* $c \in [0,\infty), (U)\int cf\,d\mu = c(U)\int f\,d\mu, (L)\int cf\,d\mu = c(L)\int f\,d\mu$.

Property 2.10. *Let* μ *be a generalized fuzzy measure. For any given measurable function* $f: X \to (0,\infty)$

$$(L)\int f\,d\mu \leq (C)\int f\,d\mu \leq (U)\int f\,d\mu$$

Property 2.11. $(U)\int f\,d\mu = 0$ *if and only if for every set A with* $\mu(A) > 0$, *there exists* $x \in A$ *such that* $f(x)=0$, *that is* $\mu(\{x|f(x)>0\})=0$.

3. Classification by fuzzy integral

3.1. Possibility theory

In the early fifties, the economist G.L.S. Shackle proposed the minimum/maximum algebra to describe degrees of potential surprise. In 1978, professor Zadeh first introduced possibility theory as an extension of his theory of fuzzy sets and fuzzy logic [19]. And then, Dubois and Prade further contributed to its development [20–22].

Possibility theory is an uncertainty theory devoted to the handling of incomplete information and it is an alternative to probability theory. It is similar to probability theory because it is based on set functions, and differs from probability theory by the use of a pair of dual set functions (possibility and necessity measures) instead of only one. Besides, it is not additive and makes sense on ordinal structures. In Zadeh's view, possibility distributions were meant to provide a graded semantics to natural language statements. However, possibility and necessity measures can also be the basis of a full-fledged representation of partial belief that parallels probability [20].

Possibility theory has been applied in many fields, such as interval analysis, data analysis, database querying, belief revision, argumentation, case-based reasoning, etc.

In the following, we review the basic notions of possibility theory including possibility distribution, possibility measure and necessity measure.

Definition 3.1 (*Zadeh [19]*). Let X be a variable taking values in the universe of discourse U, with a generic value of X denoted by μ. Informally, a possibility distribution Π_X is a fuzzy relation in U, which acts as an elastic constraint on the values that may be assumed by X, thus, if π_X is the membership function of Π_X, we have

$$Poss\{X=u\} = \pi_X(u), \quad u \in U$$

where the left-hand member denotes the possibility that X may take the value u and $\pi_X(u)$ is the grade of membership of u in Π_X. When it is used to characterize Π_X, the function $\pi_X: U \to [0,1]$ is referred to as a possibility distribution function.

- $\pi_X(u)=0$ means that state u is rejected as impossible;
- $\pi_X(u)=1$ means that state u is totally possible (plausible or unsurprising).

If the universe of discourse U is exhaustive, at least one of its elements should be the actual, so that at least one state is totally possible. Distinct values may simultaneously have a degree of possibility equal to 1.

Example 3.1. Let X be the age of a president. If X is a real-valued variable and $50 \leq X \leq 70$. In this case, the possibility distribution of X is the uniform distribution defined by

$$\pi_X(u) = \begin{cases} 1, & u \in [50,70] \\ 0, & \text{elsewhere} \end{cases}$$

Definition 3.2 (*Wang and Klir [23]*). Let X be a non-empty finite set and $\mathcal{F}$ be the the power set of X. A possibility measure Pos defined on the measurable space $(X,\mathcal{F})$ is a set function $Pos: \mathcal{F} \to [0,1]$ verifying the following axioms:

(1) $Pos(\emptyset)=0, Pos(X)=1$
(2) $Pos(\bigcup_i A_i) = \sup_i Pos(A_i)$ for any sequence $\{A_i\}$ of sets in $\mathcal{F}$

$(X,\mathcal{F},Pos)$ is said to be a possibility space.The dual set function Nec, which is defined by

$$Nec(A) = 1 - Pos(A^c) \text{ for any } A \in \mathcal{F}$$

is called a necessity measure on $\mathcal{F}$. $(X,\mathcal{F},Nec)$ is said to be a necessity space.

3.2. General background of fuzzy integral classifier

In multi-attribute classification, the fuzzy integral is used as an aggregation operator, and the fuzzy measure plays the role of weights of importance on attributes. The learning process of the classification using fuzzy integral, which is based on possibility theory, is briefly described as follows. One can refer to [8,13] for more details.

Given an unknown sample x, which is $(x_1, \ldots, x_N)$, and we want to determine the most probable class which x belongs to. The function $\Phi(C_j)$, described by the possibility distribution $\pi(C_j|x)$, is called the discriminant function. So, it is sufficient to search the maximum of the function $\Phi(C_j)$, which can be viewed as the matching degree of sample x with class C_j. And similarly, the function $\phi_i(C_j)$, described by the possibility distribution $\pi(C_j|x_i)$, is the partial matching degree of x with class C_j with respect to attribute x_i. Then we can obtain the function $\Phi(C_j)$ by fuzzy integral:

$$\Phi(C_j) = \mathsf{F}_{\mu_j}(\phi_1(C_j), \cdots, \phi_N(C_j))$$

where $\Phi(C_j)$ represents the certainty degree that x belongs to the class C_j. The Sugeno integral, Choquet integral and upper integral, as well as any other fuzzy integrals, can be used here. The fuzzy measures μ_j, defined on the set of attributes, represent the weights of importance of individual attributes. For each class, we use a different fuzzy measure.

In order to show the rationality of fuzzy integral for multi-attribute classification problems, Grabisch presented the main characteristics of fuzzy integral for aggregation. More details on these aspects can be found in [13,14,16,17].

(1) Fuzzy integrals are always located between minimum and maximum, so they are suitable for multi-attribute classification.
(2) Particularly, the usual weighted sum is covered by Choquet integral. OWA (ordered weighted average) operators are particular cases of Choquet integral. In addition, weighted minimum and maximum coincide with Sugeno integral with respect to a necessity and a possibility measure, respectively.
(3) Fuzzy integrals are the only operators that can clearly handle interaction among elements, because fuzzy measures are defined not only on the different elements, but also on all subsets of them.

3.3. Learning process

Possibility theory is an alternative to probability theory, and differs from probability theory by the use of a pair of dual set functions (possibility and necessity measures) instead of probability measure. The following learning process is based on possibility theory and genetic algorithm. More contents can be found in Ref. [8].

Consider the learning problem now. As above, we want to classify an unknown sample x by computing the conditional possibility $\pi(C_j|x)$ for each class and choose the class with the highest possibility value. Using Cox's axioms for defining conditional measures, we have

$$\pi(C_j|x_i) = \pi(x_i|C_j) \forall i,j$$

where the conditional possibility distribution $\pi(x_i|C_j)$ have described the class C_j with respect to feature x_i. So, the assignment of all $\pi(x_i|C_j)$ should be done in the learning problem, firstly.

Assume that we have l_j samples $x^{j1}, \ldots, x^{jl_j}$ belonging to class C_j, and similarly for all classes $C_1, \ldots, C_m$, we denote the total number of samples by $l = \sum_{j=1}^{m} l_j$ and use indices i,j,k to denote respectively a feature, a class and a sample. The learning process includes two parts: the learning of mN possibility distribution $\pi(x_i|C_j)$ and the learning of fuzzy measures μ_j.

Learning of the possibility distributions: The learning of a given $\pi(x_i|C_j)$ is as follows. All the samples of class j will be used to construct a "possibilistic histogram". The construction is the following: first, we construct a classical histogram with h boxes $p_1, \ldots, p_h$ from the samples, here $p_r = n_r/l_j$, with n_r the number of samples in box r, and search the tightest possibility distribution $\pi_1, \ldots, \pi_h$ having the same shape as the histogram. Without loss of generality, assuming that $p_1 \geq \cdots \geq p_h$, this is given by $\pi_r = \sum_{s=r}^{h} p_s$. Finally the continuous shape of $\pi(x_i|C_j)$ is obtained by a linear interpolation of the values.

Establishing the upper integral network with ELM: As we know, the difficult problem of non-linear integral classifiers is the determination of the set functions defined on the power set of attributes. In this paper, the extreme learning machine [25–29] is used to determine the non-additive set function based on least square error technique. The scheme of upper integral classifier can be briefly listed as follows. D is a given training data, T is the testing data and h is the number of boxes.

(1) For samples from class j in D, determine frequency histogram for each attribute. For continuous attribute i, determine h boxes and the corresponding frequencies p_i in class j samples. For nominal attribute i, regard each value of attribute i as a box and the corresponding frequencies p_i in class j samples.
(2) Rearrange the p_i, determine the possibility distribution $\pi(x_i|C_j)$ of each attribute. If the attribute is continuous, the possibility distribution is obtained by the linear interpolation.
(3) Use ELM to learn the weights of an upper integral network. We will give more details about this learning process.
For l arbitrary distinct sample (x_i,t_i), where $x_i = (x_{i1}, \ldots, x_{in})^T \in R^n$, and $t_i = (t_{i1}, \ldots, t_{im})^T \in R^m$, standard single hidden layer feedforward networks with N hidden nodes and activation function $g(x)$ are mathematically modeled as
$$\sum_{j=1}^{N} \beta_j g\left((U)\int f(x_i) d\mu_j\right) = o_i$$
where μ_j is the set function connecting the ith hidden node and the input nodes, $\beta_j = (\beta_{j1}, \ldots, \beta_{jm})^T$ is the weight vector connecting the ith hidden node and the output nodes. This can be showed in Fig. 1.
It is equivalent to minimizing the cost function
$$E = \sum_{i=1}^{l} \left\| \sum_{j=1}^{N} \beta_j g\left((U)\int f(x_i) d\mu_j\right) - t_i \right\|$$
where $\|\cdot\|$ denotes the norm of the vector.
According to extreme learning machine, set functions $\mu_j (j=1, \ldots, N)$ are randomly chosen from any intervals of R^n, and $\beta_j (j=1, \ldots, N)$ are learned to minimize the cost function E. In this approach, the set functions are randomly generated and the huge task of learning set functions is avoided. However, due to the existence of weights β_j, the upper integral with respect to the set function can also show itself effectively and smoothly.
(4) Test the classifier on given data sets, write down the classification result.

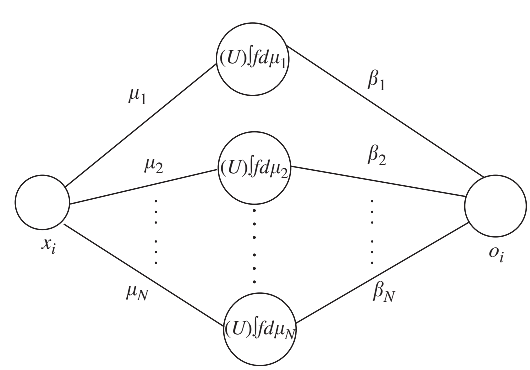

Fig. 1. The network based on extreme learning machine.

4. Test on real data

In order to investigate how well upper integral classifier works, an experimental study is conducted on some UCI machine learning databases which has been extensively used in testing the performance of different kinds of classifiers. Here the data sets are selected according to the criterion that the product of the number of attributes and the number of classes is not too large in comparison with the size of data set for the non-additive set function defined on the power set of attribute set. The information about data sets used in our experiments is summarized in Table 2.

In our experiments, 10-fold cross validation is performed on each data set. The size of training sets is roughly 60 percent of data sets.

First, the possibility histogram is constructed for each attribute according to the samples of class j. A histogram is a graphical data analysis technique for summarizing the distributional information of a variable. If an attribute is continuous, the attribute is divided into h boxes (the value of h between 7 and 15 is appropriate. If the size of samples of class j is not large enough, the value is lower). $p_1,\ldots,p_h$ are the frequencies in each box. If the attribute is nominal, each box is corresponding to an attribute value. Assuming that $p_1 \geq \cdots \geq p_h$ and $\pi_r = \sum_{s=r}^{h} p_s$, then $\pi(x_i|C_j)$ is obtained by a linear interpolation of the values π_r. The linear interpolation is needed only for continuous attributes.

From Table 3, it can be seen that the upper integral classifier can work well for both nominal and continuous attributes. A comparison between the single upper integral classifier and our proposed upper integral network is conducted. From which, we can see that, in comparison with the single upper integral classifier, the upper integral extreme learning machine shows much better performance. The following table indicates the experimental results on a number of benchmark databases.

The bottle-neck problem of non-linear integral classifiers is the determination of set function defined on the power set of attributes. It is obvious that the computational complexity will exponentially increase with the change of attribute number. And therefore, the upper integral classifier is not suitable for classification task with many attributes. How to replace the set function defined on all attributes with a simplified non-additive measure is really meaningful and significant in the non-linear integral application to classification.

5. Conclusions

Motivated by most effectively using the information from each attribute, this paper proposed the upper integral classifier network. From the definition it is known that the upper integral plays such a role that how to reasonably arrange the finite resources can maximize the efficiency. It is to maximize the possibility of sample belonging to each class. The possibility is obtained according to the possibility histogram. The upper integral classifier can work well both on nominal and continuous attributes from the experimental results on real data set. The essential difficulty in upper integral classifier design is the determination of non-additive set functions. Extreme learning machine technique provides an efficient way to overcome this difficulty. The results of comparison between the single upper integral classifier and our proposed upper integral network with ELM show that the later has much better performance.

Table 2
Data sets used for classification.

Data set	Attribute number	Class	The distribution of classes	Size
Pima	8	2	268/500	768
Sonar	60	2	111/97	208
Thyroid	5	3	150/35/30	215
Car	6	4	1210/384/69/65	1728
Breast	10	2	258/241	699

Table 3
Comparison between single upper integral and ELM upper integral.

Data set	Single upper integral	ELM upper integral
Pima	76.56/75.43	78.80/78.00
Sonar	96.28/95.26	98.34/96.99
Thyroid	83.38/80.29	88.98/86.91
Car	76.67/72.11	80.08/76.34
Breast	78.33/76.02	80.26/78.18

Acknowledgments

This research is supported by the national natural science foundation of China (60903088, 60903089), by the natural science foundation of Hebei Province (F2010000323), by the Key Scientific Research Foundation of Education Department of Hebei Province (ZD2010139).

References

[1] M. Sugeno, Theory of fuzzy integrals and its applications, Doct. Thesis, Tokyo Institute of Technology, 1974.
[2] M. Sugeno, T. Murofushi, Choquet integral as an integral form for a general class of fuzzy measures, in: Second IFSA Congress, Tokyo, 1987, pp. 408–411.
[3] Z.-Y. Wang, K.-S. Leung, M. -L Wong, J. Fang, A new type of nonlinear integrals and computational algorithm, Fuzzy Sets and Systems 112 (2000) 223–231.
[4] Z.-Y. Wang, W. Li, K.-H. Lee, K.-S. Leung, Lower integrals and upper integrals with respect to nonadditive set functions, Fuzzy Sets and Systems 159 (2008) 646–660.
[5] B. Biggio, G. Fumera, F. Roli, Multiple classifier systems for robust classifier design in adversarial environments, International Journal of Machine Learning and Cybernetics 1 (1–4) (2010) 27–41.
[6] Z.-Y. Wang, K.-S. Leung, J. Wang, A genetic algorithm for determining nonadditive set function in information fusion, Fuzzy Sets and Systems 102 (1999) 463–469.
[7] K. Xu, Z.-Y. Wang, P.-A. Heng, K.-S. Leung, Classification by nonlinear integral projections, IEEE Transaction on Fuzzy Systems 11 (2) (2003) 187–201.
[8] M. Grabisch, M. Sugeno, Multi-attribute classification using fuzzy integral, in: First IEEE International Conference on Fuzzy Systems, San Diego, March 8–12, 1992, pp. 47–54.
[9] M. Grabisch, F. Dispot, A comparison of some methods of fuzzy classification on real data, in: Second International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, July 17–22, 1992, pp. 659–662.
[10] J.-M. Keller, B. Yan, Possibility expectation and its decision making algorithm, in: First IEEE International Conference on Fuzzy Systems, Saniego, March 8–12, 1992, pp. 661–668.
[11] H. Tahani, J.-M. Keller, Information fusion in computer vision using fuzzy integral, IEEE Transactions on Systems, Man and Cybernetics 20 (3) (1990) 733–741.
[12] L.J. Wang, An improved multiple fuzzy NNC system based on mutual information and fuzzy integral, International Journal of Machine Learning and Cybernetics 1 (2) (2011) 25–36.
[13] M. Grabisch, J.-M. Nicolas, Classification by fuzzy integral: performance and tests, Fuzzy Sets and Systems 65 (1994) 255–271.
[14] M. Grabisch, The application of fuzzy integrals in multicriteria decision making, European Journal of Operational Research 89 (1996) 445–456.
[15] J. Zhu, X. Li, W. Shen, Effective genetic algorithm for resource-constrained project scheduling with limited preemptions, International Journal of Machine Learning and Cybernetics, in press, doi:10.1007/s13042-011-0014-3.
[16] M. Grabisch, Fuzzy integral in multicriteria decision making, Fuzzy Sets and Systems 69 (1995) 279–298 (Special Issue on 5th IFSA Congress).
[17] M. Grabisch, H.T. Ngnyen, E.A. Walker, Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference, Kluwer, Dordrecht, 1995.
[18] Z.-Y. Wang, K.-S. Leung, J. Wang, A genetic algorithm for determining nonadditive set functions in information fusion, Fuzzy Sets and Systems 102 (1999) 463–469.
[19] L.-A. Zadeh, Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3–28.
[20] D. Dubois, H. Prade, Possibility Theory, New York, Plenum, 1988.
[21] D. Dubois, H. Prade, Possibility theory: qualitative and quantitative aspects, in: D.M. Gabbay, P. Smets (Eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems, vol. 1, Kluwer Academic, Dordrecht, 1998, pp. 169–226.
[22] D. Dubois, H. Prade, R. Sabbadin, Decision-theoretic foundations of possibility theory, European Journal of Operational Research 128 (2001) 459–478.
[23] Z.-Y. Wang, G.-J. Klir, Fuzzy Measure Theory, Plenum Press, New York, 1992.

[24] H. Fang, M.-L. Rizzo, H.-G. Wang, K.-A. Espya, Z.-Y. Wang, A new nonlinear classifier with a penalized signed fuzzy measure using effective genetic algorithm, Pattern Recognition 43 (2010) 1393–1401.
[25] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[26] G.-B. Huang, L. Chen, C.-K. Siew, Universal approximation using incremental constructive feedforward networks with random hidden nodes, IEEE Transactions on Neural Networks 17 (4) (2006) 879–892.
[27] G.-B. Huang, L. Chen, Convex incremental extreme learning machine, Neurocomputing 70 (2007) 3056–3062.
[28] G.-B. Huang, L. Chen, Enhanced random search based incremental extreme learning machine, Neurocomputing 71 (2008) 3460–3468.
[29] G. Feng, G.-B. Huang, Q. Lin, R. Gay, Error minimized extreme learning machine with growth of hidden nodes and incremental learning, IEEE Transactions on Neural Networks 20 (8) (2009) 1352–1357.

Xizhao Wang received the B.S and M.S degrees in mathematics from Hebei University, Baoding, China, in 1983 and 1992, respectively, and the Ph.D. degree in computer science from Harbin Institute of Technology, Harbin, China, in 1998.

From 1983 to 1998, he worked as a Lecturer, an Associate Professor, and a Full Professor in the Department of Mathematics, Hebei University. From 1998 to 2001, he worked as a Research Fellow at the Department of Computing, Hong Kong Polytechnic University, Kowloon. Since 2001, he has been the Dean and Professor of the Faculty of Mathematics and Computer Science, Hebei University. His main research interests include inductive learning with fuzzy representation, fuzzy measures and integrals, neuro-fuzzy systems and genetic algorithms, feature extraction, multi-classifier fusion, and applications of machine learning. So far, he has published over 80 international journal papers and completed over 20 funded research projects, and supervised over 50 doctors or masters.

Prof. Xi-Zhao Wang is an IEEE senior member, an associate editor of IEEE Transactions on SMC part B, an associate editor of International Journal of Information Sciences, Chair of IEEE SMC Baoding Chapter; Chair of IEEE SMC Technical Committee on Computational Intelligence; Chair of IEEE SMC TC on Computational Intelligence; Executive member of Chinese Association of Artificial Intelligence; Prof. Xi-Zhao Wang is the General Co-Chair of the 2002, 2003, 2004, 2005, 2006 and 2007 International Conference on Machine Learning and Cybernetics, co-sponsored by IEEE SMC.

Aixia Chen is a lecturer of the Department of Mathematics and Computer Science, Hebei University, China. She received his Bachelor Degree and Master Degree from Hebei University in June 2004 and June 2007, respectively. Her main research interests include fuzzy reasoning, fuzzy measures and fuzzy integrals.

Huimin Feng is a lecturer of the Department of Mathematics and Computer Science, Hebei University, China. She received his Bachelor Degree and Master Degree from Hebei University in June 2002 and June 2005, respectively. Her main research interests include fuzzy reasoning, fuzzy measures and fuzzy integrals.