© 2019. This manuscript version is made available under the CC-BY-NC-ND 4.0 license
http://creativecommons.org/licenses/by-nc-nd/4.0/

Moderate deviation and restricted equivalence functions for measuring similarity between data

A. H. Altalhi[a], J. I. Forcén[b], M. Pagola[c], E. Barrenechea[c], H. Bustince[d], Zdenko Takáč[e]

[a]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia, e-mail:ahaltalhi@kau.edu.sa
[b]Das-nano – Veridas, Poligono Industrial Talluntxe II, Tajonar, Spain, e-mail:jiforcen@das-nano.es
[c]Departament of Statistics, Computer Science and Mathematics and with the Institute of Smart Cities, Universidad Publica de Navarra, Pamplona, Spain, e-mails: {miguel.pagola,edurne.barrenechea}@unavarra.es
[d]Departament of Statistics, Computer Science and Mathematics and with the Institute of Smart Cities, Universidad Publica de Navarra, Pamplona, Spain, King Abdulaziz University, Jeddah, Saudi Arabia, e-mail: bustince@unavarra.es
[e]Slovak University of Technology
Institute of Information Engineering, Automation and Mathematics
Bratislava, Slovakia, e-mail: zdenko.takac@stuba.sk

Abstract

In this work we study the relation between moderate deviation functions, restricted dissimilarity functions and restricted equivalence functions. We use moderate deviation functions in order to measure the similarity or dissimilarity between a given set of data. We show an application of moderate deviate functions used in the same way as penalty functions to make a final decision from a score matrix in a classification problem.

Keywords: Deviation; Aggregation function; Restricted equivalence function; Moderate deviation function; Penalty function

1. Introduction

Aggregation functions are used to fuse several input values into one single output value [1, 8, 18]. This fusion step is crucial in many algorithms and applications such as classification systems [11, 13], pattern recognition [17], image processing [4] or decision making [16].

A very relevant problem is that of choosing the best output value to represent a given set of inputs. This problem has been widely considered in the literature. In particular, the notion of penalty function was introduced to tackle with it [8]. Penalty functions determine the output by means of a minimization procedure and allow for defining the notion of penalty-based aggregation function [7]. However, the need of imposing convexity conditions, so that the minimization procedure can be carried out, leads to several difficulties from an analytical point of view (see for more details [6]).

More recently, it was introduced the notion of deviation-based aggregation functions [9], based on Daróczy's deviation functions [10]. The latter functions measure the difference or deviation between two real values x and y and, in this sense, they are closely related to some well-known notions such as equivalence or restricted equivalence functions (REFs) [3]. In particular, by aggregating deviation functions in an appropriate way, it is possible to measure up to what extent a given value is different from a set of inputs. This is similar to the way in which similarity measures can be built aggregating restricted equivalence functions [3], and it can be used as a basis to determine the best value to represent the given inputs.

Having this idea in mind, the objective of this work is twofold:

1. To study the relation between moderate deviation functions and restricted equivalence functions, and
2. to build deviation-based functions using moderate deviation function based on restricted equivalence functions in order to determine which is the best output to replace a given set of inputs.

In particular, we show how we can use deviation-based functions in the same way as penalty functions, but without getting involved in the convexity problems linked to the latter. In order to prove the usefulness of our approach, we present an illustrative example consisting of an image classification problem. In this case, the final decision on to which class a given image belongs to is made by deviation-based functions. We then compare the results to those obtained by other usual techniques.

The structure of the paper is as follows: first, we present some preliminary concepts that help making the paper self-contained. Section 3 presents a study of the relation between moderate deviation functions and restricted equivalence functions. Section 4 describes the way moderate deviation functions can be used as penalty functions and some examples are presented. Section 5 exhibits a practical example on the use of non-symmetric penalty based aggregations generated from moderate deviation functions and REFs, in the aggregation of the scores in an image classification problem. Finally most important conclusions and future research are described in Section 6.

2. Preliminaries

We start recalling some basic notions which are necessary for our subsequent developments.

Definition 1. An automorphism is a bijective and increasing function $\varphi : [0,1] \to [0,1]$.

Definition 2. A negation N is a decreasing function $N : [0,1] \to [0,1]$ such that $N(0) = 1$ and $N(1) = 0$.

A negation N is non-vanishing if $N(x) = 0$ if and only if $x = 1$. A negation N is strict if it is continuous and strictly decreasing. A negation N is strong if, for every $x \in [0,1]$ it holds that $N(N(x)) = x$.

Definition 3. An implication function is a mapping $I : [0,1]^2 \to [0,1]$ such that

$(I1)$ I is decreasing in the first component;
$(I2)$ I is increasing in the second component;
(BC) $I(0,1) = (1,1) = I(0,0) = 1$ and $I(1,0) = 0$ (Border Conditions).

Among the different properties that we can request to implication functions, the following are of special interest for us.

(OP) $I(x,y) = 1$ if and only if $x \leq y$ (Ordering Property);
(CP) $I(x,y) = I(N(y),N(x))$ for some strong negation N and for every $x,y \in [0,1]$ (Contrapositive Property);
$(P1)$ $I(x,y) = 0$ if and only if $x = 1$ and $y = 0$.

For an in-depth analysis of the notion of implication function as well as these and other properties, see [2].

Regarding the notion of moderate deviation function, we follow the approach given in [9].

Definition 4. A function $D : [0,1]^2 \to \mathbb{R}$ is called a moderate deviation function, if it satisfies:

(MD1) D is non-decreasing in the second component;
(MD2) D is non-increasing in the first component;
(MD3) $D(x,y) = 0$ if and only if $x = y$.

The set of all moderate deviation functions is denoted by $\boldsymbol{MD}$.

The notion of moderate deviation is closely related to those of restricted equivalence function [3] and restricted dissimilarity function [5] that we recall now.

Definition 5. A function $R : [0,1]^2 \to [0,1]$ is called a restricted equivalence function, if it satisfies:

(R1) $R(x,y) = 0$ if and only if $\{x,y\} = \{0,1\}$;
(R2) $R(x,y) = 1$ if and only if $x = y$;
(R3) $R(x,y) = R(y,x)$ for all $x,y \in [0,1]$;
(R4) If $x \le y \le z$, then $R(x,z) \le R(x,y)$ and $R(x,z) \le R(y,z)$ for all $x,y,z \in [0,1]$.

Definition 6. A function $d : [0,1]^2 \to [0,1]$ is called a restricted dissimilarity function, if it satisfies:

(d1) $d(x,y) = d(y,x)$ for all $x,y \in [0,1]$;
(d2) $d(x,y) = 0$ if and only if $x = y$;
(d3) $d(x,y) = 1$ if and only if $\{x,y\} = \{0,1\}$;
(d4) If $x \le y \le z$, then $d(x,y) \le d(x,z)$ and $d(y,z) \le d(x,z)$ for all $x,y,z \in [0,1]$.

Definition 5 and Definition 6 are related via strong negations, as the following result shows.

Proposition 1. [5] Let $R : [0,1]^2 \to [0,1]$ be a restricted equivalence function and N be a strong negation. Then the function $d : [0,1]^2 \to [0,1]$ defined by:

$$d(x,y) = N(R(x,y))$$

is a restricted dissimilarity function.

If we focus on REFs, it is possible to obtain them using automorphisms and implication functions, as two results proven in [4].

Theorem 1. [4] Let $\phi_1, \phi_2 : [0,1] \to [0,1]$ be two automorphisms. Then, the function:

$$R(x,y) = \phi_1^{-1}(1 - |\phi_2(x) - \phi_2(y)|)$$

is a restricted equivalence funcion.

Theorem 2. [4] Let $R : [0,1]^2 \to [0,1]$ be a function. the following statements are equivalent.

1. R is a restricted equivalence function.
2. There exists a function $I : [0,1]^2 \to [0,1]$ which verifies $(I1)$, (OP), (CP) and $(P1)$ and such that

$$R(x,y) = \min(I(x,y), I(y,x)) \, .$$

3. Moderate deviation and restricted equivalence functions

Next, we study the relation between moderate deviation functions, restricted dissimilarity functions and, based on Proposition 1, also restricted equivalence functions.

Let $D : [0,1]^2 \to \mathbb{R}$ be a moderate deviation function. The following properties will be of use later:

(MP) $D(x,y) = M_p$ if and only if $x = 0$ and $y = 1$ for some positive real number M_p;

(MN) $D(x,y) = -M_n$ if and only if $x = 1$ and $y = 0$ for some positive real number M_n.

Note that if D satisfies properties (MP) and (MN), then $D : [0,1]^2 \to [-M_n, M_p]$.

A construction method of restricted equivalence functions from moderate deviation functions is given in the following theorem.

Theorem 3. Let $N : [0,1] \to [0,1]$ be a strong negation and M_n, M_p be positive real numbers. Let $D : [0,1]^2 \to [-M_n, M_p]$ be a moderate deviation function satisfying properties (MP) and (MN) w.r.t. M_p and M_n, respectively. Let $A : [0, \max\{M_n, M_p\}]^2 \to [0, \infty[$ be a non-decreasing function such that $A(0,0) = 0$; $A(x,y) \neq 0$ whenever $x \neq 0$ and $y \neq 0$; $A(x, M_n) < A(M_p, M_n)$ whenever $x < M_p$

and $A(M_p, y) < A(M_p, M_n)$ whenever $y < M_n$. Then the function $d : [0,1]^2 \to [0,1]$ defined by

$$d(x,y) = \frac{A\Big(D\big(\min(x,y), \max(x,y)\big), -D\big(\max(x,y), \min(x,y)\big)\Big)}{A(M_p, M_n)} \qquad (1)$$

for all $x, y \in [0,1]$, is a restricted dissimilarity function.

Moreover the function $R : [0,1]^2 \to [0,1]$ defined by

$$R(x,y) = N\left(\frac{A\Big(D\big(\min(x,y), \max(x,y)\big), -D\big(\max(x,y), \min(x,y)\big)\Big)}{A(M_p, M_n)}\right) \qquad (2)$$

for all $x, y \in [0,1]$, is a restricted equivalence function.

Proof: Since $0 \le D\big(\min(x,y), \max(x,y)\big) \le M_p$ and $0 \ge D\big(\max(x,y), \min(x,y)\big) \ge -M_n$, for all $x, y \in [0,1]$, the functions d and R are well-defined.

(d1) The symmetry of d follows from the symmetry of $\min$ and $\max$.

($d2$) Directly follows from ($MD3$).

($d3$) Let $x \le y$. Then $d(x,y) = 1$ if and only if $D(x,y) = M_p$ and $-D(y,x) = M_n$ which only holds if $x = 0$ and $y = 1$. Similarly, for $x > y$, $d(x,y) = 1$ if and only if $D(y,x) = M_p$ and $-D(x,y) = M_n$ which only holds if $x = 1$ and $y = 0$. Hence, ($d3$) is proved.

($d4$) Observe that if $x \le y \le z$, then

$$0 \le D(x,y) \le D(x,z) \qquad \text{and} \qquad D(z,x) \le D(y,x) \le 0,$$

hence

$$d(x,y) = \frac{A(D(x,y), -D(y,x))}{A(M_p, M_n)} \le \frac{A(D(x,z), -D(z,x))}{A(M_p, M_n)} = d(x,z).$$

Similarly can be proved that $d(y,z) \le d(x,z)$, thus axiom ($d4$) is proved.

The proof regarding the restricted equivalence function straightforwardly follows from Proposition 1. □

Theorem 4. Under the same assumptions as in Theorem 3, the function d :

$[0,1]^2 \to [0,1]$ defined by

$$d(x,y) = \frac{A\Big(-D\big(\max(x,y),\min(x,y)\big), D\big(\min(x,y),\max(x,y)\big)\Big)}{A(M_n,M_p)} \tag{3}$$

for all $x,y \in [0,1]$, is a restricted dissimilarity function. Moreover the function $R: [0,1]^2 \to [0,1]$ defined by

$$R(x,y) = N\left(\frac{A\Big(-D\big(\max(x,y),\min(x,y)\big), D\big(\min(x,y),\max(x,y)\big)\Big)}{A(M_n,M_p)}\right) \tag{4}$$

for all $x,y \in [0,1]$, is a restricted equivalence function.

Proof: The proof is similar to that of Theorem 3. □

Remark 1. (i) Equations (2) and (4) can be reformulated in the following way, respectively:

$$R(x,y) = N\left(\frac{A\Big(\max\big(D(x,y),D(y,x)\big), -\min\big(D(x,y),D(y,x)\big)\Big)}{A(M_p,M_n)}\right) \tag{5}$$

$$R(x,y) = N\left(\frac{A\Big(-\min\big(D(x,y),D(y,x)\big), \max\big(D(x,y),D(y,x)\big)\Big)}{A(M_n,M_p)}\right) \tag{6}$$

and similarly also Equations (1) and (3).

(ii) It is worth pointing out that though d and R are both symmetric functions, the function A need not be symmetric. For symmetric A, Equations (2), (4), (5) and (6) are equivalent and can be simplified as follows:

$$R(x,y) = N\left(\frac{A\big(|D(x,y)|,|D(y,x)|\big)}{A(M_p,M_n)}\right). \tag{7}$$

The same holds for Equations (1) and (3).

Example 1. (i) Consider Theorem 3 and let $N(x) = 1-x$, $A(x,y) = (1-w)x + wy$ for some $w \in]0,1[$ and $D(x,y) = y-x$. Then $M_p = M_n = 1$ and by Eq. (1) we obtain $d(x,y) = |y-x|$ and by Eq. (2) we have $R(x,y) = 1-|y-x|$.

(ii) Now consider the same N and A as in item (i) and let moderate deviation function is given by (see Example 3.3 in [9]):

$$D_{\varepsilon,\delta}(x,y) = \begin{cases} y-x+\varepsilon, & \text{if } y > x, \\ 0, & \text{if } y = x, \\ y-x-\delta, & \text{if } y < x, \end{cases}$$

for some positive constants ε and δ. Then $M_p = 1+\varepsilon$, $M_n = 1+\delta$, i.e., $D_{\varepsilon,\delta} : [0,1]^2 \to [-1-\delta, 1+\varepsilon]$, and by Eq. (1) we obtain:

$$d(x,y) = \begin{cases} \frac{y-x+(1-w)\varepsilon+w\delta}{1+(1-w)\varepsilon+w\delta}, & \text{if } x < y, \\ 0, & \text{if } x = y, \\ \frac{x-y+(1-w)\varepsilon+w\delta}{1+(1-w)\varepsilon+w\delta}, & \text{if } x > y, \end{cases} = \begin{cases} \frac{|y-x|+(1-w)\varepsilon+w\delta}{1+(1-w)\varepsilon+w\delta}, & \text{if } x \neq y, \\ 0, & \text{if } x = y. \end{cases}$$

Consequently, by Eq. (2) we have:

$$R(x,y) = \begin{cases} \frac{1-|y-x|}{1+(1-w)\varepsilon+w\delta}, & \text{if } x \neq y, \\ 1, & \text{if } x = y. \end{cases}$$

Clearly, item (i) is a special case of item (ii).

Now, an inverse construction, i.e., the construction of moderate deviation functions from restricted equivalence functions, is given.

Theorem 5. Let $f_p, f_n : [0,1] \to [0,\infty[$ be non-decreasing functions such that $f_p(x) = 0$ if and only if $x = 0$ and $f_n(x) = 0$ if and only if $x = 0$. Let $N_p, N_n : [0,1] \to [0,1]$ be non-vanishing negations and $R_p, R_n : [0,1]^2 \to [0,1]$ be restricted equivalence functions. Then the function $D : [0,1]^2 \to [-f_n(1), f_p(1)]$ defined by

$$D(x,y) = \begin{cases} f_p\left(N_p\left(R_p(x,y)\right)\right), & \text{if } x \leq y, \\ -f_n\left(N_n\left(R_n(x,y)\right)\right), & \text{if } x > y, \end{cases} \tag{8}$$

for all $x, y \in [0,1]$, is a moderate deviation function.

Proof: Clearly, the function D is well-defined.

$(MD1)$ Let $y_1 < y_2$. We need to consider the following three cases:

1. Let $x \leq y_1 < y_2$. Since $R_p(x,y_1) \geq R_p(x,y_2)$, we have

$$D(x,y_1) = f_p\left(N_p\left(R_p(x,y_1)\right)\right) \leq f_p\left(N_p\left(R_p(x,y_2)\right)\right) = D(x,y_2).$$

2. If $y_1 < y_2 < x$, then $R_n(x,y_1) \leq R_n(x,y_2)$ and consequently:

$$D(x,y_1) = -f_n\left(N_n\left(R_n(x,y_1)\right)\right) \leq -f_n\left(N_n\left(R_n(x,y_2)\right)\right) = D(x,y_2).$$

3. If $y_1 < x \leq y_2$, then $D(x,y_1) < 0 \leq D(x,y_2)$. Hence, D is non-decreasing in the first component.

The proof of $(MD2)$ is similar.

Finally, since N_n, N_p are non-vanishing, $D(x,y) = 0$ if and only if ($R_p(x,y) = 1$ and $x \leq y$) or ($R_n(x,y) = 1$ and $x > y$), which only holds when $x = y$. Thus $(MD3)$ is proved. □

Moderate deviation functions are not symmetric, however, under some conditions, the construction by Eq. (8) secures the symmetry property given in the following proposition.

Proposition 2. Under the assumptions of Theorem 5, if $R_p = R_n = R$, $N_p = N_n = N$ and $f_p = f_n = f$, then Eq. (8) is equivalent to:

$$D(x,y) = \text{sign}(y-x) f\left(N\left(R(x,y)\right)\right) \tag{9}$$

and D satisfies:

$$D(y,x) = -D(x,y)$$

for all $x, y \in [0,1]$.

Proof: Immediately follows from (8). □

Remark 2. If $R_p = R_n$, $N_p = N_n$ and $f_p = f_n$, Eq. (8) is simplified by Eq. (9), otherwise (8) can be reformulated as follows:

$$D(x,y) = \frac{\text{sign}(y-x)+1}{2} f_p\left(N_p\left(R_p(x,y)\right)\right) - \frac{\text{sign}(x-y)+1}{2} f_n\left(N_n\left(R_n(x,y)\right)\right). \tag{10}$$

Example 2. (i) Under the conditions of Theorem 5, let

$$f_p = \begin{cases} 0, & \text{if } x = 0, \\ s_1, & \text{if } x \in]0,1[, \\ s_2, & \text{if } x = 1, \end{cases} \qquad f_n = \begin{cases} 0, & \text{if } x = 0, \\ t_1, & \text{if } x \in]0,1[, \\ t_2, & \text{if } x = 1, \end{cases}$$

where $0 < s_1 < s_2$ and $0 < t_1 < t_2$. Then the moderate deviation function obtained by Eq. (8) is, for any strong negations N_p, N_n and any restricted equivalence functions R_p, R_n, as follows:

$$D(x,y) = \begin{cases} 0, & \text{if } x = y, \\ s_1, & \text{if } x < y \text{ and } \{x,y\} \neq \{0,1\}, \\ s_2, & \text{if } x = 0, y = 1, \\ -t_1, & \text{if } x > y \text{ and } \{x,y\} \neq \{0,1\}, \\ -t_2, & \text{if } x = 1, y = 0. \end{cases}$$

(ii) If $N_p(x) = N_n(x) = 1 - x$, $f_p(x) = sx$, $f_n(x) = tx$ and $R_p(x,y) = R_n(x,y) = 1 - |y - x|$, then

$$D(x,y) = \begin{cases} s(y-x), & \text{if } x \leq y, \\ t(y-x), & \text{if } x > y. \end{cases}$$

(iii) Let $N_p(x) = N_n(x) = 1 - x$, $f_p(x) = sx^{q_1}$, $f_n(x) = tx^{r_1}$, $R_p(x,y) = 1 - |y-x|^{q_2}$ and $R_n(x,y) = 1 - |y-x|^{r_2}$, where $s, t, q_1, q_2, r_1, r_2 > 0$. Then

$$D(x,y) = \begin{cases} s(y-x)^{q_1 q_2}, & \text{if } x \leq y, \\ -t(x-y)^{r_1 r_2}, & \text{if } x > y. \end{cases}$$

Theorem 6. Let M_p, M_n be positive real numbers. The following statements are equivalent:

(i) A function $D : [0,1]^2 \to [-M_n, M_p]$ is a moderate deviation function satisfying properties (MP) and (MN) w.r.t. M_p and M_n, respectively.

(ii) There exist restricted equivalence functions $R_1, R_2 : [0,1]^2 \to [0,1]$ such that

$$D(x,y) = \begin{cases} M_p - M_p R_1(x,y), & \text{if } x \leq y, \\ M_n R_2(x,y) - M_n, & \text{if } x > y, \end{cases} \tag{11}$$

for all $x, y \in [0,1]$.

Proof: $(ii) \Rightarrow (i)$ Immediately follows from Theorem 5 for $f_p(x) = M_p x$, $f_n(x) = M_n x$ and $N_p(x) = N_n(x) = 1 - x$ for all $x \in [0,1]$.

$(i) \Rightarrow (ii)$ Let us define R_1 and R_2, for all $x, y \in [0,1]$, as follows:

$$R_1(x,y) = \frac{M_p - D(\min(x,y), \max(x,y))}{M_p}, \qquad R_2(x,y) = \frac{M_n + D(\max(x,y), \min(x,y))}{M_n}.$$

A trivial verification shows that R_1 and R_2 are restricted equivalence functions. The proof is completed by applying R_1 and R_2 into Eq. (11). ☐

Taking into account Theorems 1 and 2, the following two corollaries are straightforward.

Corollary 1. Let $\phi_1, \phi_2, \psi_1, \psi_2 : [0,1] \to [0,1]$ be automorphisms and let M_p, M_n be positive real numbers. Then the function

$$D(x,y) = \begin{cases} M_p - M_p \phi_2^{-1}(1 - |\phi_1(x) - \phi_1(y)|) & \text{if } x \leq y, \\ M_n \psi_2^{-1}(1 - |\psi_1(x) - \psi_1(y)|) - M_n, & \text{if } x > y, \end{cases}$$

is a moderate deviation function satisfying properties (MP) and (MN) w.r.t. M_p and M_n, respectively.

Corollary 2. Let M_p, M_n be positive real numbers. The following statements are equivalent:

(i) A function $D : [0,1]^2 \to [-M_n, M_p]$ is a moderate deviation function satisfying properties (MP) and (MN) w.r.t. M_p and M_n, respectively.

(ii) There exist functions $I_1, I_2 : [0,1]^2 \to [0,1]$ verifying $(I1)$, (OP), (CP) and $(P1)$ such that

$$D(x,y) = \begin{cases} M_p - M_p \min(I_1(x,y), I(y,x)), & \text{if } x \leq y, \\ M_n \min(I_2(x,y), I_2(y,x) - M_n, & \text{if } x > y, \end{cases}$$

for all $x, y \in [0,1]$.

4. Moderate deviation functions used as penalty functions

In this section we use moderate deviation functions in a similar way as penalty functions are used to measure the similarity or dissimilarity between a given set of data [7, 18]. The main idea is, given a set of numbers, to determine another number which represents all of them and which is the most similar to all of them in the sense determined by the deviation function. That is, we look for the value of y which makes the sum $D(x_1, y) + \ldots + D(x_n, y)$ to be as close to 0 as possible. In this way, for every point in the hypercube $[0,1]^n$, we can find an output y which can be understood to be the most similar to all the inputs.

Theorem 7. Let M_p, M_n be positive real numbers and $D : [0,1]^2 \to [-M_n, M_p]$ be a moderate deviation function defined by Eq. (11). Let $F : [0,1]^{n+1} \to \mathbb{R}$ be the function given by:

$$F(x_1, \ldots, x_n, y) = D(x_1, y) + \ldots + D(x_n, y).$$

Then

(i) If R_1, R_2 are continuous, then, for each n-tuple $(x_1, \ldots, x_n) \in [0,1]^n$, there exists $y \in [0,1]$ such that $F(x_1, \ldots, x_n, y) = 0$.

(ii) If R_1, R_2 are strictly monotone, then, for each n-tuple $(x_1, \ldots, x_n) \in [0,1]^n$, there exists at most one minimum of the function $|F(x_1, \ldots, x_n, \cdot)|$.

Proof: (i) Since the continuity of R_1, R_2 and $(R2)$ implies the continuity of D and consequently also continuity of the function $F(x_1, \ldots, x_n, \cdot)$, the proof follows from the observation that $F(x_1, \ldots, x_n, 0) \leq 0$ and $F(x_1, \ldots, x_n, 1) \geq 0$.

(ii) Observe that the strict monotonicity of R_1, R_2 implies the strict monotonicity of D and consequently also strict monotonicity (increasingness) of the function $F(x_1, \ldots, x_n, \cdot)$. □

Corollary 3. Let M_p, M_n be positive real numbers, $D : [0,1]^2 \to [-M_n, M_p]$ be a moderate deviation function defined by Eq. (11) and let R_1, R_2 be continuous strictly monotone restricted equivalence functions. Let $F : [0,1]^{n+1} \to \mathbb{R}$ be the function given by $F(x_1, \ldots, x_n, y) = D(x_1, y) + \ldots + D(x_n, y)$ and $f : [0,1]^n \to [0,1]$ be the function given by:

$$f(x_1, \ldots, x_n) = \arg_y \min |F(x_1, ..., x_n, y)|.$$

Then, f is idempotent and for each n-tuple $(x_1, \ldots, x_n) \in [0,1]^n$ such that there exist $i, j \in \{1, \ldots, n\}$ with $x_i \neq x_j$, it holds $f(x_1, \ldots, x_n) \in [x_{\sigma(k)}, x_{\sigma(k+1)}[$ and the following statements are equivalent:

(i) $f(x_1, \ldots, x_n) = y$;

(ii) $F(x_1, \ldots, x_n, y) = 0$;

(iii) $\sum_{i=1}^{k} \big(M_p - M_p R_1(x_{\sigma(i)}, y)\big) = \sum_{i=k+1}^{n} \big(M_n - M_n R_2(x_{\sigma(i)}, y)\big)$;

where $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ is a permutation such that $x_{\sigma(1)} \leq \ldots \leq x_{\sigma(n)}$ and $k \in \{1, \ldots, n-1\}$ satisfies

$$\sum_{i=1}^{n} D(x_{\sigma(i)}, x_{\sigma(k)}) \leq 0 \qquad \text{and} \qquad \sum_{i=1}^{n} D(x_{\sigma(i)}, x_{\sigma(k+1)}) > 0. \tag{12}$$

Proof: First observe that $f(x_1,\ldots,x_n)$ is symmetric. Moreover, due to $(MD1)$ and $(MD2)$ we have: if k satisfies Eq. (12), then for all $j \in \{1,\ldots,k\}$ and all $m \in \{k+1,\ldots,n\}$ it holds

$$\sum_{i=1}^{n} D(x_{\sigma(i)}, x_{\sigma(j)}) \leq 0 \qquad \text{and} \qquad \sum_{i=1}^{n} D(x_{\sigma(i)}, x_{\sigma(m)}) > 0,$$

hence $f(x_1,\ldots,x_n) \in [x_{\sigma(k)}, x_{\sigma(k+1)}[$.

The equivalence of (i) and (ii) immediately follows from Theorem 7. The equivalence of (ii) and (iii) follows from the observation:

$$F(x_1,\ldots,x_n,y) = \sum_{i=1}^{k} \left(M_p - M_p R_1(x_{\sigma(i)}, y)\right) + \sum_{i=k+1}^{n} \left(M_n R_2(x_{\sigma(i)}, y) - M_n\right).$$

Finally, the idempotency of f is straightforward. □

Remark 3. As stated above, for continuous strictly monotone R_1 and R_2, the function $f(x_1,\ldots,x_n)$ defined as in Corollary Eq. (3) is idempotent. Moreover, for n-tuples $(x_1,\ldots,x_n) \in [0,1]^n$ such that there exist $i, j \in \{1,\ldots,n\}$ with $x_i \neq x_j$, the computation of $f(x_1,\ldots,x_n)$ consists of the following two steps:

1. To obtain a switch point: first we need to obtain a switch point k, which is the greatest number from $\{1,\ldots,n-1\}$ satisfying:

$$\sum_{i=1}^{n} D(x_{\sigma(i)}, x_{\sigma(k)}) \leq 0.$$

2. To solve an equation: we obtain $f(x_1,\ldots,x_n) = y$ as a solution of equation

$$\sum_{i=1}^{k} \left(M_p - M_p R_1(x_{\sigma(i)}, y)\right) = \sum_{i=k+1}^{n} \left(M_n - M_n R_2(x_{\sigma(i)}, y)\right)$$

where $y \in [x_{\sigma(k)}, x_{\sigma(k+1)}[$.

Example 3. Let us consider the moderate deviation function D defined by Eq. (11).

(i) If $R_1(x,y) = R_2(x,y) = 1 - |y-x|$, then (see Example 2 (ii))

$$D(x,y) = \begin{cases} M_p(y-x), & \text{if } x \leq y, \\ M_n(y-x), & \text{if } x > y \end{cases} \tag{13}$$

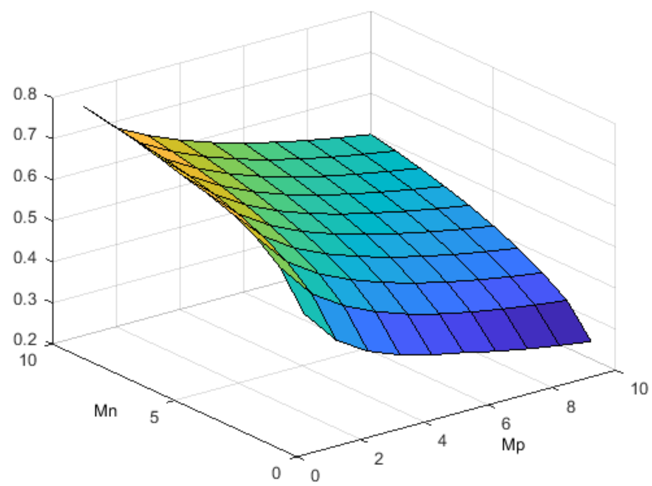

Figure 1: f function in Eq. (14) for different M_p and M_n values.

and by Corollary 3 we have:

$$f(x_1,\dots,x_n)=\frac{M_p\sum_{i=1}^{k}x_{\sigma(i)}+M_n\sum_{i=k+1}^{n}x_{\sigma(i)}}{kM_p+(n-k)M_n}. \tag{14}$$

It is easy to observe that depending on the values of the parameters M_p and M_n the resulting deviation-based function is different. In Figure 1 is depicted function f, Eq. (14), considering as input vector $x=(0.1,0.3,0.5,0.7,0.9)$ when the values of M_p and M_n range from 1 to 10.

For high values of M_p the values of the vector x that are greater than y are highly penalized, so we can see that the higher M_p is, the lower the value of the resulting function. On the contrary for higher values of M_n the final value of the function is bigger.

Moreover, if $M_n=M_p$, we have

$$f(x_1,\dots,x_n)=\frac{M_p\sum_{i=1}^{n}x_{(i)}}{nM_p}=\frac{\sum_{i=1}^{n}x_{(i)}}{n},$$

that is, f is arithmetic mean.

For instance, let $M_p = 4$, $M_n = 10$, $x_1 = 0.6$, $x_2 = 0.6$, $x_3 = 0.1$, $x_4 = 0.8$, $x_5 = 0.3$. Then $k = 2$, since $\sum_{i=1}^{5} D(x_{\sigma(i)}, x_{\sigma(2)}) = 4 \cdot 0.2 + 0 - 10 \cdot 0.3 - 10 \cdot 0.3 - 10 \cdot 0.5 = -10.2 < 0$ and $\sum_{i=1}^{5} D(x_{\sigma(i)}, x_{\sigma(3)}) = 4 \cdot 0.5 + 4 \cdot 0.3 + 0 - 10 \cdot 0 - 10 \cdot 0.2 = 1.2 > 0$. Consequently, $f(0.6, 0.6, 0, 1, 0, 8, 0, 3) \in [0.3, 0.6[$, which is in accordance with the calculation of $y = f(0.6, 0.6, 0.1, 0.8, 0.3)$ by item (iii) of Corollary 3:

$$\begin{aligned} 4(y-0.1)+4(y-0.3) &= 10(0.6-y)+10(0.6-y)+10(0.8-y) \\ 8y-1.6 &= 20-30y \\ y &= 0.568421 \end{aligned}$$

(ii) Let $R_1(x,y) = 1-(y-x)^2$ and $R_2(x,y) = 1-|y^2-x^2|$. Then

$$D(x,y) = \begin{cases} M_p(y-x)^2, & \text{if } x \leq y, \\ M_n(y^2-x^2), & \text{if } x > y \end{cases}$$

and by Corollary 3 we have $f(x_1, \ldots, x_n) = y$, where $y \in [x_{\sigma(k)}, x_{\sigma(k+1)}[$ is a solution of the equation:

$$y^2\left(kM_p+(n-k)M_n\right)+y\left(-2M_p\sum_{i=1}^{k}x_{\sigma(i)}\right)+M_p\sum_{i=1}^{k}x_{\sigma(i)}^2-M_n\sum_{i=k+1}^{n}x_{\sigma(i)}^2=0.$$

For instance, let $M_p, M_n, x_1, \ldots, x_5$ be as in item (i). Then $k = 4$, since $\sum_{i=1}^{5} D(x_{\sigma(i)}, x_{\sigma(4)}) = -1.44 < 0$ and $\sum_{i=1}^{5} D(x_{\sigma(i)}, x_{\sigma(5)}) = 3.28 > 0$. Consequently, $f(0.6, 0.6, 0.1, 0.8, 0.3) \in [0.6, 0.8[$, which is in accordance with the calculation of $y = f(0.6, 0.6, 0.1, 0.8, 0.3)$ by item (iii) of Corollary 3:

$$\begin{aligned} 4(y-0.1)^2+4(y-0.3)^2+8(y-0.6)^2 &= 10(0.64-y^2) \\ 26y^2-12.8y-3.12 &= 0 \\ y &= 0.671115 \end{aligned}$$

Similarly, if $R_1(x,y) = 1-|y^2-x^2|$ and $R_2(x,y) = 1-(y-x)^2$, then we have

$$D(x,y) = \begin{cases} M_p(y^2-x^2), & \text{if } x \leq y, \\ -M_n(y-x)^2, & \text{if } x > y \end{cases}$$

and the equation

$$y^2\left(kM_p-(n-k)M_n\right)+y\left(-2M_n\sum_{i=k+1}^{n}x_{\sigma(i)}\right)-M_p\sum_{i=1}^{k}x_{\sigma(i)}^2-M_n\sum_{i=k+1}^{n}x_{\sigma(i)}^2=0,$$

if $R_1(x,y)=R_2(x,y)=1-(y-x)^2$, we have

$$D(x,y)=\begin{cases} M_p(y-x)^2, & \text{if } x\le y,\\ -M_n(y-x)^2, & \text{if } x>y\end{cases}$$

and the equation

$$y^2\left(kM_p+(n-k)M_n\right)+y\left(2M_n\sum_{i=k+1}^{n}x_{\sigma(i)}-2M_p\sum_{i=1}^{k}x_{\sigma(i)}\right)+M_p\sum_{i=1}^{k}x_{\sigma(i)}^2-M_n\sum_{i=k+1}^{n}x_{\sigma(i)}^2=0,$$

if $R_1(x,y)=R_2(x,y)=1-|y^2-x^2|$, we have

$$D(x,y)=\begin{cases} M_p(y^2-x^2), & \text{if } x\le y,\\ M_n(y^2-x^2), & \text{if } x>y\end{cases}$$

and the equation

$$y^2\left(kM_p+(n-k)M_n\right)-M_p\sum_{i=1}^{k}x_{\sigma(i)}^2-M_n\sum_{i=k+1}^{n}x_{\sigma(i)}^2=0.$$

(iii) Now we show that for R_1,R_2 which are neither continuous nor strict monotone there can exist infinitely many minima of the function $|F(x_1,\dots,x_n,\cdot)|$, and the value of the minima is not equal to 0. Let $M_p=10$, $M_n=2$ and

$$R_1(x,y)=R_2(x,y)=\begin{cases} 0, & \text{if } \{x,y\}=\{0,1\},\\ 1, & \text{if } x=y,\\ 0.5, & \text{otherwise.}\end{cases}$$

Then

$$D(x,y)=\begin{cases} 10, & \text{if } x=0, y=1,\\ 5, & \text{if } x<y \text{ except the case } x=0,y=1,\\ 0, & \text{if } x=y,\\ -2, & \text{if } x=1,y=0,\\ -5, & \text{if } x>y \text{ except the case } x=1,y=0\end{cases}$$

and $k=1$. Consequently, the set of minima of the function $|F(x_1,\ldots,x_n,\cdot)|$ is $]0.1,0.2[$ where the value of the minima is equal to 1.

5. Illustrative example: Moderate deviate functions for aggregation of a score matrix

In this section we describe an application of moderate deviation functions when they are used in order to determine the output which best represents a given set of inputs.

This example is framed in an image classification scenario, but can be extended to any classification problem. In an image classification problem, we are given several sets of images and a new image and we must determine to which of the given sets it belongs to. In this sense, image classification is a multi-class problem that can be tackled with an ensemble of binary classifiers [13]. One of the most widely used methods to deal with this kind of problems is the One-vs-One ensemble method (OVO). Recall that OVO ensemble divides an m class problem into $m(m-1)/2$ binary problems. Each binary problem is faced by a binary classifier, called base classifier, which is responsible for distinguishing from one class C_i from another class C_j. In the training phase each binary classifier is trained with images of its corresponding classes i and j, that is, the images which belong tho the classes i and j are used to fix the parameters the classifiers in order to distinguish when they belong to the class i and when they belong to the class j. When a new instance should be classified, all the base classifiers are considered and their outputs are stored in a score matrix R:

$$R=\begin{pmatrix} - & r_{12} & \cdots & r_{1m} \\ r_{21} & - & \cdots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \cdots & - \end{pmatrix} \tag{15}$$

being $r_{ij}\in[0,1]$ the confidence value in favour of class C_i given by the classifier that distinguishes between $\{C_i,C_j\}$; on the contrary, the confidence for C_j is calculated by $r_{ji}=1-r_{ij}$ (just in case this value is not provided by the classifier).

A common method to decide to which class belongs an instance is the following: aggregate the confidences of each row with the arithmetic mean and select the row with the highest value (this method is usually known as

Method	Performance
Arithmetic mean	65.47
$M_p = 10$, $M_n = 1$	64.67
$M_p = 1$, $M_n = 10$	65.72

Table 1: Percentage of correctly classified test images with different deviation functions

the weighted vote).

$$\text{Class} = \arg\max_{i=1,\ldots,m} \left\{ \frac{\sum_{1 \le j \ne i \le m} r_{ij}}{m-1} \right\} \tag{16}$$

But, there exist some cases in which the confidences of certain classes are underestimated by the binary classifiers; for example if there are few examples of a certain class, all of the binary classifiers related to this class usually returns very low confidence values for said class [14]. Therefore the arithmetic mean may not be the most adequate aggregation operator. Instead, we propose to use operators derived from moderate deviate functions in order to fuse the values in each row of the score matrix $\boldsymbol{R}$.

To test this proposal we carried out an experiment in an image classification problem. We use the cifar-10 data-set [15] (10 categories with 5000 images for training and 1000 for test per class of 32x32). We process the images with Bag-of-features method and extract a vector of 128 features for each image (see more details in [12]). The problem is divided in a One-vs-One strategy as explained before and the base classifiers (linear kernel Support Vector Machine L2-SVM) are trained with the images of the training set. Finally for each image in the test set we calculate the score matrix with the confidences of the classifiers, i.e. for each image we have a 10×10 matrix. In Table 1 we show the percentage of correctly classified images of the test set using three different methods of the score matrix, the first row by means of the arithmetic mean and in the second and third rows considering Eq.(13) with $M_p > M_n$ (i.e. penalizing higher values) and $M_n > M_p$, respectively.

Analyzing the class selected in each case, we noticed that in 9375 cases of 10000 the three different methods made the same prediction; in 615 cases one method predicts a different class from the other two and only in 10 cases the three predictions are different. As we can see in this case the best option is to use the function that penalizes the lower values. This means that some confidences are underestimated and the function build with the moderate deviation function provides a better representative value of each row. But a simple question arises: How can we set the correct values for the moderate deviation function?. Due to M_p and M_n are the parameters that modify the behaviour of the function we can learn their ideal values by means of the gradient descent algorithm. For this example we have fixed the value of $M_p = 1$ and the gradient descent algorithm learns the value of M_n. We have used a set of 5000 images of the training set, and the gradient descent algorithm tries to minimize the following cost function:

$$J = \sum_{i=0}^{m} \begin{cases} 0 & \text{if the image is correctly classified} \\ (\sum_{1 \leq j \neq i \leq m} r_{ij} - \sum_{1 \leq j \neq z \leq m} r_{zj})^2 & \text{if the image is classified as i but it is of class z} \end{cases} \tag{17}$$

Initializing $M_n = 1$ and the step $\alpha = 0.1$ the algorithm converges in very few epochs. The algorithm obtain a value of $M_n = 29$, which achieves a performance of 65.68 in the test set of images. Therefore, the gradient descent algorithm finds a suited value of M_n for this problem. The performance achieved is similar to the best case of Table 1.

In this example we have seen that there are some cases in which the values to be aggregated have some bias, therefore the arithmetic mean is not the most suitable aggregation, because it will also produce biased results. Therefore, operators based on moderate deviation functions can be used to overcome the problem of biased values. We have observed that even, we can calculate through a simple gradient descent algorithm, the most fitted parameter of the moderate deviation function to a problem; so, by simply changing the operator an improvement of the results is obtained.

6. Conclusions

In this work we have studied the relation between deviation functions and restricted equivalence functions. We have provided a construction theorem to generate moderate deviation functions from restricted equivalence functions. Also we have shown how to use moderate deviation functions as penalty functions. We have applied moderate deviation based penalty for the aggregation of a score matrix in an image classification problem. We have shown that this type of function can be fitted to the data therefore obtain better accuracy results than common aggregation operators like the average mean.

Aknowledgements

This work was supported in part by the Spanish Ministry of Science and Technology under project TIN2016-77356-P (AEI/FEDER,UE) and by grants VEGA 1/0420/15 and APVV-14-0013.

[1] G. Beliakov, H. Bustince, T. Calvo Sánchez, A Practical guide to averaging functions, Springer (2016).

[2] H. Bustince, P. Burillo, F. Soria, Automorphisms, negations and implication operators, Fuzzy Sets Syst. 134 (2) (2003) 209-–229.

[3] H. Bustince, E. Barrenechea, M. Pagola, Restricted equivalence functions, Fuzzy Sets Syst. 157 (17) (2006) 2333–2346.

[4] H. Bustince, M. Pagola, E. Barrenechea, Construction of fuzzy indices from fuzzy DI-subsethood measures: Application to the global comparison of images, Information Sciences 177 (3) (2007) 906–929.

[5] H. Bustince, E. Barrenechea, M. Pagola, Relationship between restricted dissimilarity functions, restricted equivalence functions and normal E_N-functions: Image thresholding invariant, Pattern Recognition Letters 29 (4) (2008) 525–536.

[6] H. Bustince, G. Beliakov, G.P. Dimuro, B. Bedregal, R. Mesiar, On the definition of penalty functions in aggregations, Fuzzy Sets Syst. 323 (2016) 1–18.

[7] T. Calvo, G. Beliakov, Aggregation functions based on penalties, Fuzzy Sets Syst., 161 (10) (2010) 1420–1436.

[8] T. Calvo, R. Mesiar, R.R. Yager, Quantitative weights and aggregation, IEEE Trans. Fuzzy Syst., 12 (1) (2004) 62–69.

[9] M. Decký, R. Mesiar and A. Stupňanová, Deviation-based aggregation functions, Fuzzy Sets Syst. 332 (2017) 29–36.

[10] Z. Daróczy, Über eine KLasse von Mittelwerten, Publ. Math. (Debr.), 19 (1972) 211–217.

[11] R.M.O. Cruz , R. Sabourin , G.D.C. Cavalcanti , Dynamic classifier selection: recent advances and perspectives, Information Fusion 41 (2018) 195–216.

[12] J. I. Forcén, M. Pagola, H. Bustince, J. M. Soto-Hidalgo and J. Chamorro-Martínez, Adding fuzzy color information for image classification, in: Proceedings of the 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2017, pp. 1–6.

[13] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes, Pattern Recognition 44 (8) (2011) 1761–1776.

[14] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, F. Herrera, A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches, IEEE Transactions on Systems, Man, and Cybernetics, Part C, 42 (4) (2012) 463–484.

[15] A. Krizhevsky, Learning multiple layers of features from tiny images, Science Department University of Toronto Tech. (2009).

[16] D. Paternain, H. Bustince, M. Pagola, P. Sussner, A. Kolesárová, R. Mesiar, Capacities and overlap indexes with an application in fuzzy rule-based classification systems, Fuzzy Sets Syst. 305 (2016) 70–94,.

[17] G. Selvachandran, H. Garg, M.H.S. Alaroud, A. R. Salleh, Similarity Measure of Complex Vague Soft Sets and Its Application to Pattern Recognition, International Journal of Fuzzy Systems, 20 (6) (2018) 1901-–1914.

[18] R.R. Yager, Toward a general theory of information aggregation, Information Sciences 68 (3) (1993) 191–206.