Atlantis Studies in Uncertainty Modelling, volume 3

Joint Proceedings of the 19th World Congress of the International Fuzzy Systems Association (IFSA), the 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and the 11th International Summer School on Aggregation Operators (AGOP)

Applicability of Ordinal Sums of Conjunctive and Disjunctive Functions in Classification

*Miroslav Hudec[a,c] and **Radko Mesiar**[b] and **Erika Mináriková**[a]

[a]Faculty of Economic Informatics, University of Economics in Bratislava, Bratislava, Slovak Republic
miroslav.hudec@euba.sk, erika.minarikova@euba.sk
[b]Faculty of Civil Engineering, Slovak University of Technology, Bratislava, Slovak Republic
radko.mesiar@stuba.sk
[c]Faculty of Economics, VSB - Technical University of Ostrava, Ostrava, Czech Republic
miroslav.hudec@vsb.cz

Abstract

Classification relies on the rules expressed by domain experts, or on the labeled attribute explaining the output classes. However, such information is not always available. In this work, we explore classification according to aggregation functions of mixed behaviour by the variability in ordinal sums of conjunctive and disjunctive functions. By this approach, entities are classified into three classes: *yes*, *no* and *maybe* including inclination to the classes *yes* and *no*. By the proposed approach, domain experts (or ordinal users) explain classification linguistically, without stating IF–THEN rules and labeled output. The applicability is illustrated by two examples. The discussion of the results and further research activities conclude the paper.

Keywords: Ordinal sums, Conjunctive functions, Disjunctive functions, Averaging functions, Classification.

1 Introduction

In classification by the rule-based systems, domain experts (or ordinal users) should explain classification by the IF-THEN rules matching input with output attributes. A rule-based system should be consistent, which is not always an easy task [2]. In a supervised classification, a labeled output attribute should be available (usually expressed by the categorical values, or discretized numerical values) [1]. Consequently, by e.g., decision trees the solution is obtained and explained. Classification by neural networks has shown efficiency even beyond human-level performance [3, 11]. It holds true when well-designed (and of sufficient size) sets of input-output data are available for learning and validating.

In various classification tasks ranging from everyday activities to medical or business decisions, entities are classified into two classes, which we can mark as *yes* or *no*. However, we might need a third class marked as *maybe*. In addition, an entity might slightly or significantly incline to the one of two extreme poles.

In flexible classification tasks, users are able to express vague, but relevant explanation regarding the aggregation of the input attributes, like: *low values mean no interest at all (resp. reduced interest but not rejection), high values mean full acceptance (resp. increased interest), whereas a mix of high and low values indicates medium interest*. In these tasks, values of output labels might not be available. Hence, aggregation functions of mixed behaviour should be considered. This work evaluates the recently proposed theoretical work on classification by ordinal sums of conjunctive and disjunctive functions [15] on examples and proposes a further theoretical extension.

The remainder of paper is organized as follows. Section 2 briefly elaborates classification with uncertainty. Section 3 introduces ordinal sums and proposes further contribution. Section 4 is devoted to experiments, whereas Section 5 discusses obtained results and the implication for the future research. Finally, Section 6 concludes the paper.

2 Classification into three classes

Generally, the goal is to classify into several classes, whereas binary classification separates entities into two classes. When task deals with the uncertainty, classes should have flexible borders [17]. Three classes *yes*, *no* and *maybe* cover clear acceptance, clear rejection, and entities we cannot classify into these two extreme poles, respectively.

In Figure 1 low values of two attributes indicate belonging to the class *no*, high values to the class *yes* and remaining to the class *maybe*. In this direction, two

possible classification spaces are drawn. Nevertheless, low values can be desired (e.g., cost). It is matter of scaling and inversion. Next, attributes can be atomic or compound. These cases are covered in experiments elaborated in Section 4.
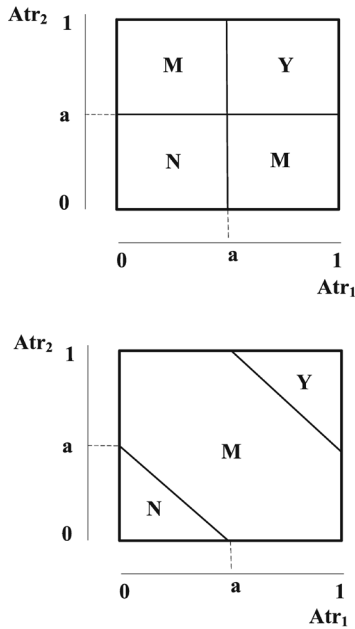

Figure 1: Classification into three classes - two possible realizations.

The motivation in this work is formalizing users explanations for classification expressed linguistically, without stating IF-THEN rules or assigning a labeled output attribute. The proposed classification is beneficial when users search for transparent and explainable classification, but are not able to provide clear requirements. The comparison with other classification approaches is discussed in [15]. The next section explores ordinal sums for solving this demand.

3 Ordinal sums of conjunctive and disjunctive functions

The main classification of aggregation functions is due to [8]: conjunctive, averaging, disjunctive and mixed ones. Generally, we can express conjunctive functions as $0 \leq A(\mathbf{x}) \leq x_i$ for each $i \in \{1,...,n\}$, averaging functions as $x_i \leq A(\mathbf{x}) \leq x_j$ for some $i, j \in \{1,...,n\}$, disjunctive functions as $x_i \leq A(\mathbf{x}) \leq 1$ for each $i \in \{1,...,n\}$ and mixed as remaining aggregation functions. Conjunctive functions attenuate low values, disjunctive emphasize high values, whereas averaging provide solution between the lowest and highest value.

Ordinal sums in their origin were considered as ex-

tension methods for semigroups [6], or for posets [5]. Later, in the framework of fuzzy sets theory, they were considered to build new t–norms/t–conorms from the scaled versions of existing ones [16].

The ordinal sum of conjunctive and disjunctive functions has been proposed by De Baets and Mesiar [7] as follows.

For an n–ary aggregation function $B : [0,1]^n \to [0,1]$ and $[a,b] \subset R$, denote $B_{[a,b]}(\mathbf{x}) = a + (b-a) \cdot B(\frac{\mathbf{x}-a}{b-a})$ Note that then $B_{[a,b]}$ is an n–ary aggregation function on $[a,b]$. For $B_1,...,B_k : [0,1]^n \to [0,1], k \geq 2$, and $0 \leq a_0 < a_1 < ... < a_k = 1$. Let $A_i : [a_{i-1}, a_i]^n \to [a_{i-1}, a_i]$ be given by $A_i = (B_i)_{[a_{i-1},a_i]}$. Then the ordinal sum $A : [0,1]^n \to [0,1], A = (< a_{i-1}, a_i, A_i >)|i = 1,...,k$ is given by

$$A(\mathbf{x}) = \sum_{i=1}^{k} (A_i(a_i \wedge (a_{i-1} \vee \mathbf{x})) - a_{i-1}) \qquad (1)$$

is an aggregation function on $[0,1]$. If all $B_1,...,B_k$ are t–norms (t–conorms, copulas, means) then also A is a t–norm (t–conorm, copula, mean).

Equivalently, $A(\mathbf{x}) = \sum_{i=1}^{k} (a_i - a_{i-1}) \cdot B_i(1 \wedge (0 \vee \frac{\mathbf{x}-a_{i-1}}{a_i-a_{i-1}}))$. For our purposes, $n = k = 2$ is considered. Denoting $a_1 = a(a_0 = 0, a_2 = 1)$, we have two next forms of ordinal sums[15]
(i) $B_1, B_2 : [0,1]^2 \to [0,1]$,

$$A(x,y) =$$
$$a \cdot B_1(1 \wedge \frac{x}{a}, 1 \wedge \frac{y}{a}) + (1-a) \cdot B_2(0 \vee \frac{x-a}{1-a}, 0 \vee \frac{y-a}{1-a})$$
$$(2)$$

(ii) $A_1 : [0,a]^2 \to [0,a], A_2 : [a,1]^2 \to [a,1]$,

$$A(x,y) = A_1(a \wedge x, a \wedge y) + A_2(a \vee x, a \vee y) - a \qquad (3)$$

Then:

- if $(x,y) \in [0,a]^2$, $A(x,y) = a \cdot B_1(\frac{x}{a}, \frac{y}{a}) = A_1(x,y)$,

- if $(x,y) \in [a,1]^2$, $A(x,y) = a + (1-a) \cdot B_2(\frac{x-a}{1-a}, \frac{y-a}{1-a}) = A_2(x,y)$,

- if $(x,y) \in [0,a] \times [a,1]$, $A(x,y) = a \cdot B_1(\frac{x}{a}, 1) + (1-a) \cdot B_2(0, \frac{y-a}{1-a}) = A_1(x,a) + A_2(a,y) - a$,

- if $(x,y) \in [a,1] \times [0,a]$, $A(x,y) = a \cdot B_1(1, \frac{y}{a}) + (1-a) \cdot B_2(\frac{x-a}{1-a}, 0) = A_1(0,y) + A_2(x,a) - a$.

If B_1 is a conjunctive and B_2 is a disjunctive aggregation function, then A is conjunctive on $[0,a]^2$ and disjunctive on $[a,1]^2$. Moreover, if B_1 has a neutral element $e = 1$, i.e., B_1 is a semicopula [10], and B_2 has

a neutral element $e = 0$, i.e., B_2 is a dual semicopula, then, for $(x,y) \in [0,1]^2 \setminus ([0,a]^2 \cup [a,1]^2)$ it holds $A(x,y) = x + y - a \in [\min(x,y), \max(x,y)]$, i.e., A is averaging on this domain. It reflects our demands for classification into three aforementioned classes. The next is variation of conjunctive, disjunctive and averaging functions in ordinal sums.

Product t-norm, its dual probabilistic sum t-conorm and variations of averaging functions

These functions are representative of strict functions. Note that product t–norm is expressed as $C_P(x,y) = x \cdot y$, whereas its dual t–conorm as $D_P(x,y) = x + y - x \cdot y$.

In order to keep the expected value on edges of subintervals $[0,a]^2$ and $[a,1]^2$, when $a = 0.5$, the product t–norm on $[0,0.5]$ is expressed as [15]

$$C_P(x,y) = A_1(x,y) = 2x \cdot y. \tag{4}$$

The dual observation holds for t–conorm on $[0.5,1]$:

$$D_P(x,y) = A_2(x,y) = -1 + 2x + 2y - 2x \cdot y. \tag{5}$$

The averaging function expressed by the geometric mean for $[0,a] \times [a,1] \cup [a,1] \times [0,a]$ is as follows

$$G(x,y) = 2x \cdot y \tag{6}$$

The averaging function expressed by the arithmetic mean for $[0,a] \times [a,1] \cup [a,1] \times [0,a]$ is as follows

$$W(x,y) = x + y - \frac{1}{2} \tag{7}$$

The logical perspective of averaging aggregation functions [9] considers global *andness* and *orness* where the arithmetic mean W is a logically neutral function (*andness* and *orness* values are equal to 0.5, regardless weights assigned to the attributes). The other averaging functions are either conjunctively or disjunctively polarized. Thus, it is possible to formalize diverse behaviours in the class *maybe*.

Geometric mean covers inclination towards the pessimistic or more restrictive classification into the class *maybe* due to conjunctively polarization (*andness* measure is greater than 0.5).

In this work, another representative of averaging function, quadratic mean is introduced for $[0,a] \times [a,1] \cup [a,1] \times [0,a]$ as follows

$$Q(x,y) = \sqrt{x^2 + y^2 - \left(\frac{1}{2}\right)^2} \tag{8}$$

The graphical interpretation can be seen in Figure 2.

This function handles situations, when the inclination is to a less restrictive classification in the averaging part for covering the intensities of belonging to the class *maybe*.
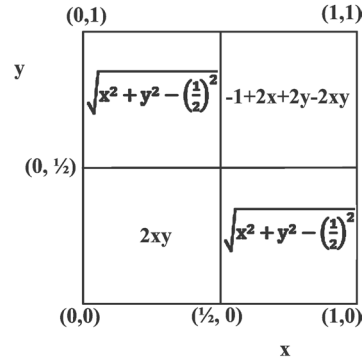

Figure 2: The graphical interpretation of (8) for product t-norm, probabilistic sum t-conorm and quadratic mean.

Łukasiewicz t–norm, its dual Łukasiewicz t–conorm and arithmetic mean

These functions are representative of nilpotent functions. Note that Łukasiewicz t–norm is $C_L(x,y) = \max(0, x + y - 1)$ and its dual t-conorm is $D_L(x,y) = \min(1, x + y)$. Applying $AM(x,y) = A_1(x, \frac{1}{2}) + A_2(\frac{1}{2}, y) - \frac{1}{2} = x + y - \frac{1}{2}$ and adjusting value 0.5 to edges of subintervals, we can observe that for two attributes ($n = 2$) this aggregation behaves as [15]

$$AM_L(x,y) = \mathrm{med}(0, 1, x + y - \frac{1}{2}). \tag{9}$$

This model corresponds with the second classification space depicted in Figure 1.

When the neutral behaviour of class *maybe* is not the case, we can manage averaging behaviour by other averaging functions keeping the nilpotent behaviour for the conjunctive and disjunctive parts of a classification space.

4 Experiments on data

This section provides two experiments demonstrating the applicability of the proposed classification model.

4.1 Atomic predicates in classification

A human resource manager is preparing a rewarding model for motivating sellers in a company. Two key at-

tributes are the sum of arranged orders in money units and the average time spent on persuasion by telephone calls. The requirement for rewarding is as follows:

When amount of arranged sells is high and persuasion time is short the reward should be full (in this task 100 money units), for low amounts of arranged sells and long persuasion time no reward is provided, otherwise a medium reward should be provided ensuring that similar workers are similarly treated.

Data are in a matrix form of $D_{[n,3]}$ where n is the number of evaluated workers. Attributes are worker ID, sum of arranged orders, whose values are denoted by x and length of calls denoted by y. Both attributes are drawn on different scales, thus the scaling into the unified interval is advisable [1]. The transformation from one scale into another is a matter of scaling [4]. The simplest form for transforming into the $[0,1]$ is by

$$x_{iN} = \frac{x_i - min_x}{max_x - min_x} \qquad (10)$$

where $max_x = max_{i=1...n}(x_i)$ and $min_x = min_{i=1...n}(x_i)$.

In practice, outliers (errors, or extremal, but valid values) cause skewed normalization. In this example, domain expert knowledge resolves the problem when stating: values lower than D_L are unacceptable work performance, whereas values higher than D_H are excellent. The opposite holds for the time spent during persuasion (lower values are preferable). This transformation is depicted in Figure 3.

The following function transforms values of the arranged sells into the $[0,1]$ interval.

$$x_{iN} = \begin{cases} 0 & \text{for } x_i \leq D_{Lx} \\ \frac{x_i - D_{Hx}}{D_{Hx} - D_{Lx}} & \text{for } D_{Lx} \leq x_i \leq D_{Hx} \\ 1 & \text{for } x_i \geq D_{Hx} \end{cases} \qquad (11)$$

where D_{Lx} is the limiting value for the unacceptable work performance and D_{Hx} is the limiting value for the excellent performance for this attribute (see Figure 3). Analogously, the following function transforms values of the persuasion time into the $[0,1]$ interval. Note that in this case, lower values are preferable.

$$y_{iN} = \begin{cases} 0 & \text{for } y_i \geq D_{Hy} \\ 1 - \frac{y_i - D_{Ly}}{D_{Hy} - D_{Ly}} & \text{for } D_{Ly} \leq y_i \leq D_{Hy} \\ 1 & \text{for } y_i \leq D_{Ly} \end{cases} \qquad (12)$$

Regarding the arranged sells, a manager says that sells lower or equal to 400 stand for a weak performance, whereas sells higher or equal to 600 are excellent. Thus, we apply $D_{Lx} = 400$ and $D_{Hx} = 600$ into Eq. (11). In a same way, a manager says that the average time longer than 12 is a weak performance, whereas
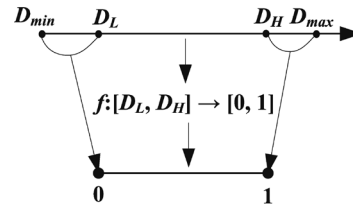

Figure 3: A scaling from the real values into the unit interval.

average time shorter than 8 is excellent. Thus, we apply $D_{Ly} = 8$ and $D_{Hy} = 12$ into Eq. (12).

No further information is provided. Based on the existing information, we apply Łukasiewicz t–norm, its dual t–conorm and arithmetic mean for the averaging behaviour. The input data, belonging to classes *yes* (value 1), *no* (value 0) or *medium* with the intensity in $]0,1[$ and reward are shown in Table 1, where **sol** stands for solution by (9).

ID	SELL	TIME	SOL (9)	REWARD
W1	694	7	1.00	100.00
W2	552	9	1.00	100.00
W3	693	10	1.00	100.00
W4	540	9	0.95	95.00
W5	574	11	0.62	62.00
W6	411	7	0.56	55.50
W7	664	13	0.50	50.00
W8	371	7	0.50	50.00
W9	500	10	0.50	50.00
W10	1000	20	0.50	50.00
W11	537	11	0.44	43.50
W12	585	14	0.43	42.50
W13	369	9	0.25	25.00
W14	498	11	0.24	24.00
W15	472	11	0.11	11.00
W16	418	10	0.09	9.00
W17	455	11	0.03	2.50
W18	324	12	0.00	0.00
W19	474	15	0.00	0.00
W20	445	11	0.00	0.00

Table 1: Classification of workers, when $a = 0.5$, A_1 is Łukasiewicz t-norm, A_2 is Łukasiewicz t-conorm and averaging behaviour is expressed by arithmetic mean.

Worker *W20* gets no reward, because performances of sell and time are unsatisfactory, i.e., values of 0.23 and 0.25, respectively, which is a clear belonging to the class N (Figure 1, the second classification space). Worker *W17* has received a very low reward, because performances are sufficient to weakly belong to the class M (i.e, 0.28 and 0.25). The opposite holds for be-

longing to the class Y. For a clear high performance of one attribute and a clear low performance for another attribute the reward is 0.5 (see *W10*).

4.2 Compound predicates in classification

In this example, a client of a real estate agency wishes to evaluate flats. Two compound attributes are distance and comfort. The former is a result of aggregating atomic distances (e.g., *short distance to school, work, grocery shop, public transport lines, park, etc.*). The latter is a result of aggregating atomic attributes like: *size approx. 200 m^2 and strong preference of balcony and (spacious basement or else spacious larder)*. Clearly, both attributes get values from the unit interval. The requirement is as follows:

High values of both attributes should be emphasized, but the full satisfaction is when at least one of them is ideal. When both are low, the satisfaction decreases, but when one attribute is unacceptable, the flat should be rejected. When one attribute is low and another one is high, the inclination is more to the acceptance than to rejection.

To meet this explanation, we apply product t-norm, its dual probabilistic sum t-conorm and quadratic mean for the averaging behaviour.

The input data and the solution are shown in Table 2.The full satisfaction is when the first attribute is ideally satisfied and another one is satisfied with a degree greater than or equal to 0.5 (compare flats *F2* and *F6*). The same holds for the dual case of rejection. Both high, but not ideal degrees do not ensure belonging to the class Y, i.e., *F5*) The averaging behaviour holds for low and high values, but the inclination is towards class Y (compare *F11* and *F14*).

5 Discussion

The experiments have demonstrated the applicability of ordinal sums in classification. These functions are able to handle imprecise requirements for classification expressed linguistically. The examples have solved by the representative functions of nilpotent (resp. strict) functions, i.e. Łukasiewicz t–norm and its dual t–conorm (resp. product t-norm and its dual probabilistic t-conorm) as well as by the chosen averaging function (arithmetic mean resp. quadratic mean). The other combinations can be constructed to cover further requirements for classification. For instance, nilpotent conjunction for class N and strict disjunction for class Y. Hence, this is a robust classification approach capable to cover diverse requirements posed linguistically.

The question is, whether other functions in these cat-

FLAT	DISTANCE	COMFORT	SOL
F1	0.93	1	1.000
F2	1	0.89	1.000
F3	1	1	1.000
F4	1	0.57	1.000
F5	0.96	0.95	0.996
F6	1	0.31	0.920
F7	0.35	0.94	0.870
F8	1	0	0.866
F9	0	1	0.866
F10	0.75	0.52	0.760
F11	0.8	0.2	0.656
F12	0.51	0.62	0.628
F13	0.76	0.15	0.592
F14	0.5	0.5	0.500
F15	0.39	0.53	0.428
F16	0.23	0.58	0.3732
F17	0.12	0.58	0.317
F18	0.48	0.23	0.221
F19	0.22	0.5	0.220
F20	0.38	0.17	0.129
F21	0.17	0.06	0.020
F22	0.08	0.1	0.016
F23	0	0.23	0.000
F24	0	0	0.000
F25	0	0.43	0.000

Table 2: Classification of flats, when $a = 0.5$, A_1 is product, A_2 is probabilistic sum and averaging behaviour is covered by quadratic mean (8).

egories could be a better fit. The theory offers several parametrized families of t–norms and t–conorms. More about these families can be found in, e.g., [4, 16]. These families usually cover basic t–norms as limiting cases. Several families do not cover both strict and nilpotent behaviour, or one of these behaviours is only for a particular value of parameter. A suitable family for our purpose is the Schweizer and Sklar family [18].

The answer can be reached by machine learning (ML), when values of output attribute are available. In this way, the domain experts provides their contextual implicit knowledge, which is managed by the interactive ML (iML) approach [13] and has been proven as being useful within several scenarios [12, 14]. The future direction should consider human-in-the-loop who provides inputs for selecting the most suitable categories of functions and consequently ML finds the best values for parameters. Thus, the solution becomes transparent and explainable. This work examines classification into three classes. The future work should consider classification into the arbitrary number of classes.

6 Conclusion

This work has examined classification into three classes *yes*, *no*, *maybe* by the ordinal sums of conjunctive and disjunctive functions. This approach is suitable when requirements for classification are expressed linguistically, when the expert knowledge is not available in the form of IF-THEN rules, or an output labeled attribute is not set.

Domain experts assign information how to classify high values and low values of the considered attributes, and how to classify low and high value. When values of attributes are not in the unit interval, domain expert might also explain the transformation into the unit interval linguistically. This work further proposes enveloping quadratic mean into the ordinal sums to complete the behaviour of assigning entities into the class *maybe* (inclinations to a more restrictive or optimistic evaluation). Next, two experiments have demonstrated the applicability on the different scenarios.

The representative nilpotent and strict functions were applied in this work. If for a subset of data the result of classification is available, machine learning approaches might be helpful in recognizing the most suitable parameters for nilpotent or strict functions.

Acknowledgement

References

[1] C. Aggarwal, Data Mining, Springer, Cham, 2015.

[2] J. Alonso, C. Castiello, , C. Mencar, Interpretability of fuzzy systems: Current research trends and prospects, in: J. Kacprzyk, W. Pedrycz (Eds.), Springer handbook of computational intelligence, Springer, Berlin Heidelberg, 2015, pp. 219–237.

[3] T. Araújo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polónia, A. Campilho, Classification of breast cancer histology images using convolutional neural networks, PloS one 12 (6) (2017) e0177544.

[4] G. Beliakov, A. Pradera, T. Calvo, Aggregation functions: a guide for practitioners, Springer, Berlin Heidelberg, 2007.

[5] G. Birkhoff, Lattice Theory, 3rd edition, AMS Colloqium publications, Vol. XXV,American Mathematical Society, Providence, 1967.

[6] A. Clifford, Naturally totally ordered commutative semigroups, American Journal of Mathematics 76 (1954) 631–646.

[7] B. De Baets, R. Mesiar, Ordinal sums of aggregation operators, in: Technologies for Constructing Intelligent Systems 2, 2002, pp. 137–147.

[8] D. Dubois, H. Prade, On the use of aggregation operations in information fusion processes, Fuzzy sets and systems 142 (1) (2004) 143–161.

[9] J. Dujmović, Soft Computing Evaluation Logic: The LSP Decision Method and Its Applications, John Wiley & Sons, 2018.

[10] F. Durante, C. Sempi, Semicopulae, Kybernetika 41 (3) (2005) 315–328.

[11] A. Esteva, B. Kuprel, R. Novoa, J. Ko, S. Swetter, H. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (7639) (2017) 115–118.

[12] D. Girardi, J. Küng, R. Kleiser, M. Sonnberger, D. Csillag, J. Trenkler, A. Holzinger, Interactive knowledge discovery with the doctor-in-the-loop: a practical example of cerebral aneurysms research, Brain Informatics 3 (3) (2016) 133–143.

[13] A. Holzinger, Interactive machine learning for health informatics:when do we need the human-in-the-loop?, Brain Informatics 3 (2016) 119.131.

[14] A. Holzinger, M. Plass, M. Kickmeier-Rust, K. Holzinger, G. Crişan, C. Pintea, V. Palade, Interactive machine learning: experimental evidence for the human in the algorithmic loop, Applied Intelligence 49 (7) (2019) 2401–2414.

[15] M. Hudec, E. Mináriková, R. Mesiar, A. Saranti, A. Holzinger, Classification by ordinal sums of conjunctive and disjunctive functions for explainable ai and interpretable machine learning solutions, Knowledge-Based Systems 220 (2021) id. 106916.

[16] E. Klement, R. Mesiar, E. Pap, Triangular norms, Kluwer, Dordrecht, 2000.

[17] A. Meier, N. Werro, A fuzzy classification model for online customers, Informatica 31 (2007) 175–182.

[18] B. Schweizer, A. Sklar, Associative functions and triangle inequalities, Publicationes Mathematicae Debrecen 8 (1961) 169–186.