Multiset Neurons

Luciano da Fontoura Costa

Multiset Neurons

Luciano da Fontoura Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

1st Nov. 2021

Abstract

The present work reports a comparative performance of artificial neurons obtained in terms of the real-valued Jaccard and coincidence similarity indices and respectively derived functionals. The interiority index and classic cross-correlation are also included for comparison purposes. After presenting the basic concepts related to real-valued multisets and the adopted similarity metrics, including the generalization of the real-valued Jaccard and conicidence indices to higher orders, we proceed to studying the response of a single neuron, not taking into account the output non-linearity (e.g. sigmoid), respectively to the detection of gaussian two-dimensional stimulus in presence of displacement, magnification, intensity variation, noise and interference from additional patterns. It is shown that the real-valued Jaccard and coincidence approaches are substantially more robust and effective than the interiority index and the classic cross-correlation. The coincidence-based neurons are shown to have the best overall performance respectively to the considered type of data and perturbations. The reported concepts, methods, and results, have substantial implications not only for pattern recognition and machine learning, but also regarding neurobiology and neuroscience.

1 Introduction

A great deal of human perception and cognition, as well as of many other living beings, critically rely on neuronal transduction and processing of several types of information. From a simplified mathematical perspective, a neuron has been understood as a cell specialized in processing and transmitting signals. In a very simplified approach to modeling neuronal operation, known as *integrate-and-fire*, neuronal dynamics can be though as involving the two following main stages: (i) *integration:* an inner product between the input stimulus and the respective synaptic weights, yielding an accumulated value; and (ii) *fire:* the subsequent application of a non-linear function, such as a sigmoid, over that value, eventually yielding an action potential (e.g. [1, 2]).

This type of operation can be complemented, regarding the geometrical/shape aspects of neuronal operation, in terms of the concept of *receptive field* (e.g. [3, 4]) defined with respect to some input stage space. For instance, several of the ganglion cells of the retina (e.g. [5]) have been characterized by respective antagonic receptive fields defined on the visual space (scene) or along the retina surface (retinotopic). Cortical neuronal cells often operate on topographical mappings of the visual field (e.g. [3, 4]). The mathematical modeling of these receptive fields therefore provides an effective manner for representing, modeling, and better understanding neuronal operation according to a systemic representation which is directly related to the concepts of correlation, convolution and point-spread functions (e.g. [6, 7, 8, 9]).

In addition to its dynamic properties along time, the shape of receptive fields has been understood to play an important role in detecting and processing patterns. Indeed, a more elaborated dendritic arborization will tend to have enhanced chances of receiving more synaptic connections. The importance of the neuronal geometry seems to be so important that it often adapts to the type of function the neuron performs (e.g. [10, 11, 12]). Among the several possible interrelationships between neuronal shape and function, we have that the alignment and similarity between the visual signal and the neuronal two-dimensional distribution of synaptic weights tend to result in higher neuronal activation, therefore providing some kind of *template matching* or *matched filtering*.

In the present work, we re-evaluate the functioning of single neurons in terms of recently introducted multiset-based similarity indices capable of operating on real-valued data [13, 14]. More specifically, instead of using the traditional inner product, we apply the real-valued Jaccard, interiority, and coincidence similarity metrics [13, 14, 15]. Of particular interest is the fact that, though extremely simple, these metrics implement an action that, though analogous to the classic inner product,

is non-linear as a consequence of the use of the maximum and minimum binary operators. By 'binary operator' it is meant the mathematical understanding of an operation involving two arguments.

We start by presenting the inner product, its properties, basic multiset concepts (e.g. [16, 17, 18, 19, 20, 21]), as well as the recently introduced real-valued Jaccard and coincidence indices [13, 22, 14]. This presentation is performed first respectively to one-dimensional input, and then extended to two- and multidimensional synaptic inputs. In addition to discussing the intrinsic, though limited, ability of the real product between two scalars in providing information about their respective similarity, we also show how the real-valued Jaccard index can be derived in a logical manner starting from the totally strict similarity comparison provided by the Kronecker delta function.

Unlike in a recent study [13], which approached the subject of similarity more generally in terms of correlation-like perspective, the neuronal perspective adopted in this work allowed attention to be focused on similarity comparisons where one of the arguments is kept constant, therefore corresponding to stable synaptic weights. In addition, for generality's sake, additional results are reported regarding the generalization of the multiset similarity indices to higher orders, yielding a generic similarity function that converges to the Kronecker delta product for infinite order.

A systematic approach is then proposed and applied for comparing the performance of neurons in pattern recognition, while adopting the standard cross-correlation as well as the interiority, real-valued Jaccard and coincidence indices [13, 22, 14]. The comparison is performed with respect to varying pattern position, intensity, scale, noise levels, presence of additional interfering patterns, and false positives resulting from completely noisy data.

Several interesting results are report that, all in all, confirm that the coincidence index provides the most strict and detailed recognition, followed by the real-valued Jaccard and interiority indices. The classic cross-correlation resulted almost useless for the considered task and type of data. These results have many implications and applications to several related areas, some of which are also briefly discussed.

2 Product and Similarity

Given any two real values x and y, their product constitutes one of the most frequently performed algebraic operation in science and technology, not to mention daily activities. Yet, there are some quite interesting properties of the product xy that, perhaps as a consequence of being so ubiquitous, are not commonly realized.

Let's start with the product sign rule:

$sign\{x\}$	$sign\{y\}$	$sign\{xy\}$
$-$	$-$	$+$
$-$	$+$	$-$
$+$	$-$	$-$
$+$	$+$	$+$

(1)

Logically, the above rules can be conceptualized as the *identity* operation of Boolean Algebra (e.g. [16]).

It follows that the classic product between two real values is capable of expressing whether the two values x and y head toward the same direction along the real line, in which case $sign\{xy\} = +1$, or if they oppose one another, yielding $sign\{xy\} = -1$. As such, the product operation can be understood to quantify, in its signal, the similarity of the relative orientations of the two operands.

This important property of the classic real product hints at a yet more important respective feature, namely the fact that *the classic real product provides measurement of similarity* between the *signs* (or direction) of two signed values x and y [13]. This particular feature of the product contributes strongly to capacity of the inner product for quantifying the similarity between two vectors. More specifically, the traditional inner product between any two vectors $\vec{v}$ and $\vec{p}$ in an $N-$dimensional space can be written as:

$$\langle \vec{v}, \vec{p} \rangle = \sum_{i=1}^{N} v_i p_i = |\vec{v}|\,|\vec{p}| \cos(\theta) \tag{2}$$

where θ is the smallest angle between the two vectors. Provided the magnitudes of $\vec{v}$ and $\vec{p}$ are kept constant, the inner product will provide an indication of the angular and orientation similarity between these two vectors. Observe that the inner product is a bilinear operation.

When translated to scalar values, the inner product becomes:

$$\langle x, y \rangle = xy = |x|\,|y| \cos(\theta) \tag{3}$$

which makes it clear that the scalar version of the inner product is the product of the two scalar arguments.

In this case, the cosine similarity becomes:

$$\cos(\theta) = \frac{\langle x, y \rangle}{|x||y|} = \frac{xy}{|x||y|} = \pm 1 \tag{4}$$

Observe also that, provided $|x| \leq 1$ and $|y| \leq 1$, it will follow that $-1 \leq xy \leq 1$.

In the case of $\vec{x}$ and $\vec{y}$ being vectors in $\mathcal{R}^N$, the respective cosine similarity can be expressed as:

$$\cos(\theta) = \frac{\langle \vec{x}, \vec{y} \rangle}{|\vec{x}|\,|\vec{y}|} \tag{5}$$

This expression implies that the cosine similarity between two vectors can be understood as corresponding to a normalized version of the inner product between those two vectors. As a consequence, the inner product between two *versors* (vectors with unit magnitude) is identical to the respective cosine similarity.

Despite its intrinsic ability for quantifying similarity between the sign of values, as well as it extensive application in operations as the inner product, the real product has two important shortcomings. First, it is relatively difficult to be implemented in computational hardware or even in analog circuits. Second, it has been shown that the real product tends to be too tolerant regarding the provided indication of similarity [14, 13, 23], as illustrated in Figure 1 with respect to comparison between versors.

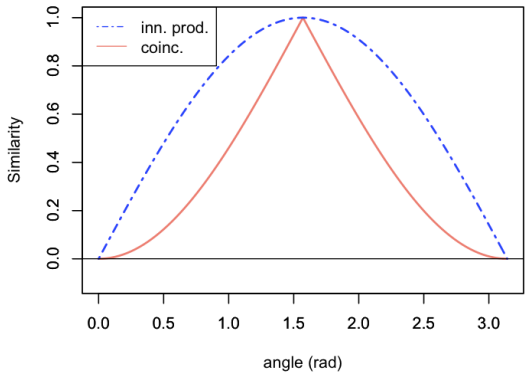

Figure 1: The pairwise similarity between versors in $\mathcal{R}^2$ as quantified by the inner product (dashed blue) and coincidence (solid salmon) similarities. The versor $\vec{x} = [0, 1]$ is compared to versors making angles from 0 to π with the horizontal axis. As expected, both similarity indices reach their respective peak at $\pi/2$, but the inner product similarity, which in the case of versors is identical to the cosine similarity, is markedly less strict in comparing similarity, providing substantially higher values than the coincidence similarity in all cases except for angles 0 and $\pi/2$.

The substantially high similarity values generated by the inner product similarity between two vectors immediately implies that that it tends to provide a relatively coarse, or little strict, quantification of the relationship between the two vectors. Another feature of particular interest in the similarity profiles such as those in Figure 1 concerns the quality factor (in analogy with filter theory, e.g. [24]) of the quantification, which can be understood as being proportional to the peak value divided by the standard deviation or other dispersion measurements, therefore providing an indication of the sharpness of the profile. Another important property of a similarity profile is its magnitude of the derivative at and around its peak. Regarding the former, we have that it is zero for the inner product similarity (as expected with any smooth function) and infinite for the coincidence. At the same time, the magnitude of the derivatives around the peak are very small in the case of the inner product similarity, and particularly high for the coincidence similarity. The fact that the derivative of the coincidence similarity profile diverge at its peak can be easily circumvented, if necessary for analytical and theoretical studies, by representing the profile in terms of a truncated Fourier series, which is necessarily analytical (has any derivative).

One problem of having a smooth (or 'blunt') similarity peak, as is the case with the inner product similarity, consists in the fact that the identification of its position by using derivative is highly susceptible to any level of noise. That is so because the derivatives at and around the peak have very small magnitudes (smooth) and can therefore be severely disturbed by the noise during the derivative, as this operation emphasized the high frequency content of the curve. The sharp and intense derivative peak resulting from the coincidence can hardly have its correct position disturbed by any reasonable level of signal noise.

On the other hand, a smoother similarity comparison profile tends to favor *generalization* of the comparison, a property that is often expected at some level in neuronal networks and pattern recognition. In the case of Figure 1, the relatively higher similarity values provided by the inner product, respectively to the coincidence, similarity means that input patterns that are more different to the one used as a reference (or template) will imply larger similarity values, therefore implying larger generalization. Observe that the generalization property is opposite to accuracy in the similarity comparison, which means that either one of these two properties is prioritized, or a suitable balance between them needs to be achieved. To any extent, as it can be appreciated from Figure 1, the coincidence similarity already presents a substantial ability for generalization, yielding substantially high (though much smaller than the cosine similarity) for input with angles reasonably near $\pi/2$.

Table 1 provides a qualitative relative comparison between the several properties respectively characterizing the inner product (or cosine) and coincidence similarities.

However, being more or less strict does not necessarily imply an advantage or shortcoming of a given similarity index, unless these trends are extreme. Indeed, there are situations in which it may be interesting to implement a more yielding quantification, which tends to favor false negatives. At the same time, a more strict similarity quantification can be particularly interesting in other situations in which false positives have higher costs and need to be minimized and/or enhance accuracy is expected in the quantification. In summary, the localization of the position of the peak tends to be substantially more robust and accurate in the case of the coincidence than the inner product similarity.

property	inner product	coincidence
values	typ. higher	typ. lower
peak shape	smooth	sharp
peak concavity	convex	concave
strictness	lower	higher
peak localiz. accur.	lower	higher
magn. deriv. at peak	lower	higher
deriv. at peak	0	∞
quality factor	lower	higher
false neg. prob.	lower	higher
false pos. prob.	higher	lower
generalization	higher	lower

Table 1: Relative comparison of the main properties of the inner product (cosine) and coincidence similarities.

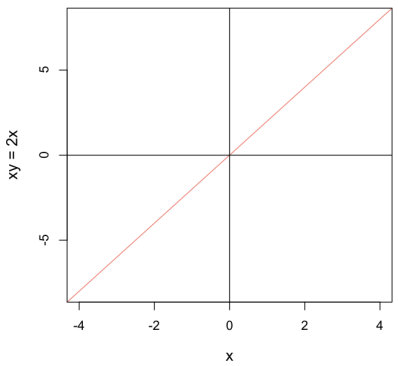

Figure 2: The quantification of similarity between two real values as implemented by the real product xy, with $y = 2$ has severe limitations. The curve should indicate how similar to 2 the values of x are, but monotonically increasing values are obtained instead.

For these reasons, it becomes particularly interesting to consider similarity measurements involving one or more parameters that can be used to control how strict the quantification is, so that this can be adapted to a wide range of situations and applications. Approaches such as *multiresolution* or *multiscale* have been developed with that finality in mind. The present work describes a related approach in which a parameter D is used to control how strict the Jaccard and coincidence similarity indices are.

3 Multiset Similarities in One-Dimensional Spaces

The considered neuronal application of similarity comparison addressed in this work provides an interesting perspective from which to address this issue and its related aspects. More specifically, if we consider that the similarity is to be measured between the synaptic weight y and respective input x, we can simplify the otherwise binary operation as an operation only on x (i.e. a function of x), with y being understood as a parameter. Figure 2 illustrates the real product seen from this perspective, assuming synaptic weight $y = 2$.

This result well-illustrates the limitation of the traditional real product for quantifying similarity. Though the similarity will increase for $|x|$ increasing from 0 to 1, it will continue to increase thereafter. In fact, we have that:

$$\lim_{x \to \infty} xy = \infty \tag{6}$$

with $y = 2$.

As developed in [13], the prototypical function for quantifying similarity in the most strict manner possible consists in the Kronecker delta function, which can be written as:

$$\delta_{x,y} = \begin{cases} 1 & whenever\ x = y \\ 0 & otherwise \end{cases} \tag{7}$$

Though this function cannot provide information about the alignment of the values x and y, it can be readily generalized as:

$$\tilde{\delta}_{x,y} = \begin{cases} 1 & whenever\ x = y \\ -1 & whenever\ x = -y \\ 0 & whenever\ |x| \neq |y| \end{cases} \tag{8}$$

Figure 3 illustrates this function for $y = 2$, i.e. $\tilde{\delta}_{x,2}$.

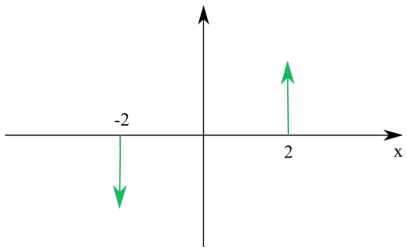

Figure 3: The Kronecker function modified to take into account the relative alignment of the real values x and $y = 2$, which is now reflected in the respective sign.

The binary Kronecker function generalized to quantify signed similarity is shown in Figure 4.

The problem with this generalized Kronecker delta function is that it is simply too strict in its evaluation of the similarity between two real values x and y.

We have from Equation 4 that, when applied on real values x and y in $\mathcal{R}$, the cosine similarity effectively acts as the Kronecker delta function on those two values, therefore presenting rather limited potential for comparing the similarity between x and y.

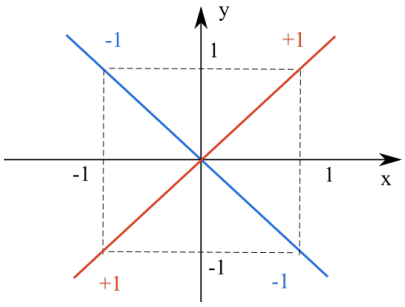

Figure 4: The binary Kronecker function modified to take into account the relative alignment of the real values x and y.

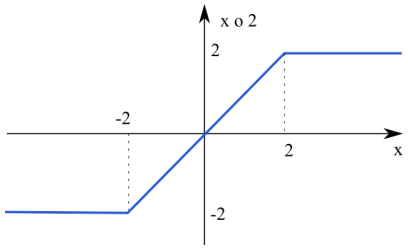

Figure 5: The operation $x \sqcap y$ assuming $y = 2$, i.e. $x \sqcap 2$. Now we have that $-2 \leq x \sqcap 2 \leq 2$ The obtained profile can be understood as being analogous to the 'receptive field' of a respective neuron operating in a one-dimensional input space.

Interestingly, there is another particularly interesting possibility to quantify the similarity between two real values taking possibly negative values [25, 26, 27, 13]. In particular, the basic scalar version of the operator in [26, 27] follows the same sign rules as the above discussed inner product while involving only the minimum binary operation:

$$x \sqcap y = s_{xy} \min \{s_x x, s_y y\} \tag{9}$$

where $s_x = sign(x)$, $s_y = sign(y)$, and s_{xy} is the *conjoint sign function* $s_{xy} = s_x s_y$.

This operator has also been verified [28, 22, 13] to correspond to the signed intersection between multisets taking real, possibly negative values, being dierctoy related to real-valued adaptations of the Jaccard similarity index. In particular, it can be understood as a modification of the intersection between two functions in order to consider the common or shared area of the functions with respect to the horizontal axis.

Interestingly, this product has surprising properties, including: (i) it is extremely simple to be implemented [26, 27], e.g. in analog electronic circuits [29]; (ii) it is conceptually simple; (iii) it obeys the sign rules in Equation 1; (iv) its magnitude is bound by the absolute value of the minimum between x and y; (v) unlike the cosine similarity when applied to 1D spaces ($x, y \in \mathcal{R}$), the real-valued Jaccard index is not limited to yielding ± 1 values, but is instead capable of providing a detailed quantification of the respective similarity.

It is therefore interesting to consider this function from the neuronal perspective, i.e. with one of its values kept constant so as to correspond with the respective weights. Figure 5 illustrates the operation $x \sqcap y$ for $y = 2$, which can be understood as being analogous to the 'receptive field' of a respective neuron in the one-dimensional space $\mathcal{R}$.

It is now clear that this operation, when one of its argument is kept constant, corresponds to a clipped version of the real product $2x$ as in our previous example. The saturation observed for $x > 2$ is a critical feature in which it implies $x \sqcap y$ to become bound by the fixed value.

However, maximum similarity will be observed for any value of x larger than 2. An interesting manner to circumvent this saturation problem consists of adopting the following normalization:

$$J(x, y) = \frac{s_{xy} \min \{s_x x, s_y y\}}{\max \{s_x x, s_y y\}} = \frac{f \sqcap g}{f \tilde{\sqcup} g} \tag{10}$$

so that $-1 \leq x \sqcap y \leq 1$. This normalized version of the operation $x \sqcap y$ has been verified to correspond to the *real-valued Jaccard index* applied to two real scalar values [13, 22, 14].

Observe that, as with the standard Jaccard similarity index, the respective multiset version above capable of operation on real values is not defined for the comparison between two null sets or feature vectors, as it diverges to $0/0$.

Figure 6 illustrates both the function $n(x, y) = \max \{s_x x, s_y y\}$ and the resulting real-valued Jaccard index.

The normalizing function that constitutes the denominator has a direct correspondence with the generalized multiset concept of absolute union [28]. Observe that this function increases linearly with x. As a consequence, the division by the normalizing function will penalize the similarities for $|x| > 2$, yielding to two respective peaks in the real-valued Jaccard index $J_R(x, 2)$, providing enhanced quantification selectivity. Interestingly, this resulting index can therefore be understood as a less strict version of the generalized Kronecker delta (compare Figs. 3 and 6b), while being also more strict than the real product (as a consequence of the saturation).

The developments presented above make it clear that it is possible to define an infinity of other similarity indices. For instance, it is possible to control the sharpness of the similarity peaks by using other products and normalizing functions.

As an example, if even sharper peaks are required, we

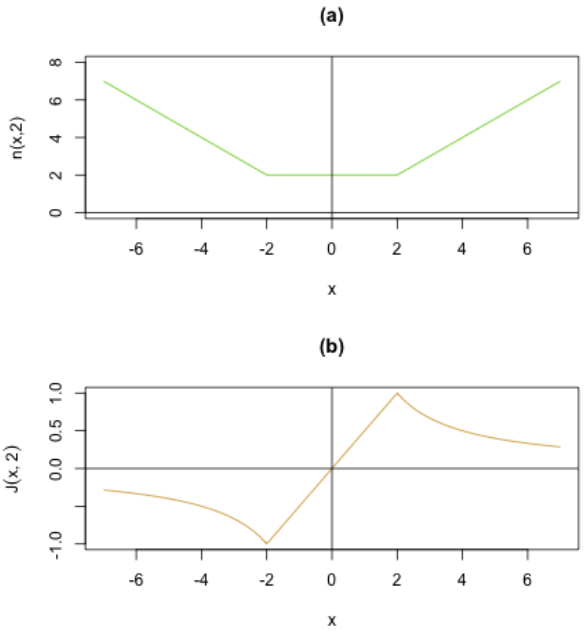

Figure 6: The normalizing function $n(x, 2) = \max\{s_x x, 2\}$ and the normalized operator which corresponds to the real-valued Jaccard index $J(x, 2)$ applied to real scalar values.

can make:

$$J_3(x, y) = J(x, y)^3 \tag{11}$$

Figure 7 illustrates this function for $y = 2$.

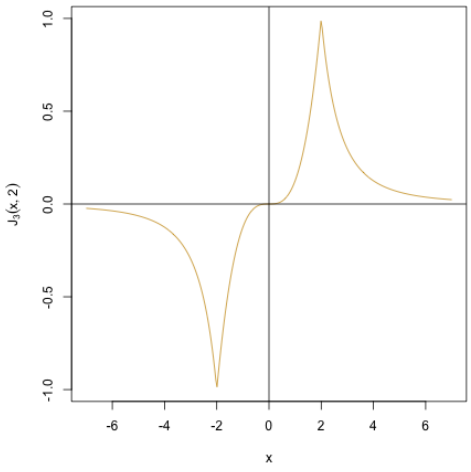

Figure 7: The similarity function $J_3(x, 2)$, when compared to the real-valued Jaccard index, is characterized by a sharper peak, therefore implying even more strict similarity quantification.

The above development can be generalized to any non-negative integer degree D odd as:

$$J_D(x, y) = J(x, y)^D \tag{12}$$

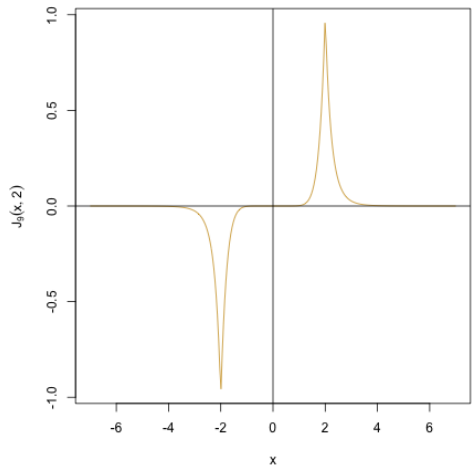

Figure 8: Similarity quantification through the function $S_{D=9}(x, 2)$.

An adaptation can be also implemented in case we need to consider D even. It is also possible to have less strict similarity comparison by adopting $D < 1$, which also requires some adaptation of Equation **??**.

Observe that the similarity function $S_D(x, y)$ tends to the generalized Kronecker delta function when $D \to \infty$, D odd, i.e.:

$$\lim_{D\to\infty} S_D(x, y) = \tilde{\delta}(x, y) \tag{13}$$

However, for simplicity's sake, we will consider only the real-valued Jaccard index in our subsequent performance analysis, which can be understood as the above construction when $D = 0$. A more systematic study of higher values of D will be reported opportunely.

Given that the Jaccard similarity index has been verified not to be able to take into account the relative interiority (or overlap, e.g. [30]) between the two compared vectors, the *coincidence similarity* has been proposed as a respective complementation, consisting of the product between the real-valued Jaccard and interiority indices:

$$\mathcal{C}_R(\vec{x}, \vec{y}) = \mathcal{J}_R(\vec{x}, \vec{y})\ \mathcal{I}_R(\vec{x}, \vec{y}) \tag{14}$$

Since the coincidence index does not distinguish from the respective real-valued Jaccard index for the one-dimensional input (i.e. $x, y, \in \mathcal{R}$), we now assume that the two values to be compared are vectors in an $N-$dimensional space, i.e. $\vec{x}, \vec{y} \in \mathcal{R}^N$.

In this case, the real-valued Jaccard index can be expressed as:

$$\mathcal{J}_R(\vec{x}, \vec{y}) = \frac{\sum_{i=1}^{N} s_{x_i y_i} \min\{s_{x_i} x_i, s_{y_i} y_i\}}{\sum_{i=1}^{N} \max\{s_{x_i} x_i, s_{y_i} y_i\}\, dx} \tag{15}$$

The interiority index for real valued vectors can be expressed [13, 14, 22] as:

$$I(\vec{x},\vec{y}) = \frac{\sum_{i=1}^{N} \min\{s_{x_i}x_i, s_{y_i}y_i\}\, dx}{\max\{S_{\vec{x}}, S_{\vec{y}}\}} \tag{16}$$

where:

$$S_{\vec{x}} = \sum_{i=1}^{N} s_{x_i}x_i \tag{17}$$

$$S_{\vec{y}} = \sum_{i=1}^{N} s_{y_i}y_i \tag{18}$$

4 Multiset Similarities in Two-Dimensional Spaces

Having presented and discussed the properties of the inner product and real-valued Jaccard similarity indices respectively to comparing two real values x and y, we now extend that discussion to two dimensional spaces, so that now we are interested in comparing the similarity between two real-valued vectors $\vec{x} = [x_1, x_2]^T$ and $\vec{y} == [y_1, y_2]^T$, with $\vec{x}, \vec{y} \in \mathcal{R}^2$.

Figure 9 presents the cosine similarity calculated between a reference vector $\vec{y} = [1, 2]^T$ and vectors $\vec{x} = [x_1, x_2]^T$ with $4 \leq x_1, x_2 \leq 4$.

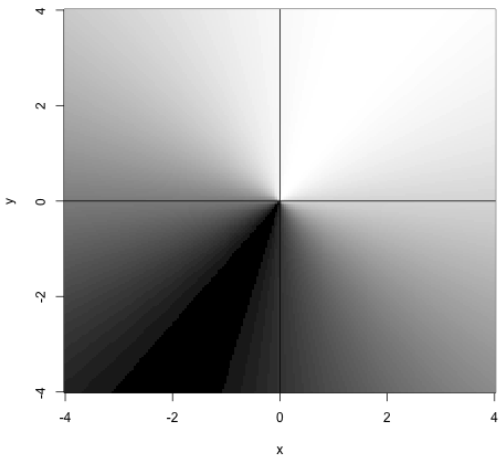

Figure 9: Cosine similarity values (shown in grayscale from dark to bright) obtained for the cosine similarity index between a reference vector $\vec{y} = [1, 2]^T$ and vectors $\vec{x} = [x_1, x_2]^T$ with $4 \leq x_1, x_2 \leq 4$.

The maximum similarity takes place for the angle sector containing the vector $\vec{y}$, but any other vector with the same angle will imply identical cosine similarity, therefore illustrating the fact that the cosine similarity cannot distinguish between any two vectors with the same orientation but distinct magnitudes. In addition, observe that the gray levels undergo rather little variations for vectors with orientations similar to that of $\vec{y} = [1, 2]^T$. These issues can have severe impact on the pattern recognition performance of individual neurons based on the cosine similarity. Analogous implications are expected for $N-$dimensional input.

Figure 10 depicts the similarity values obtained for the real-valued Jaccard index considering the same comparison problem as before.

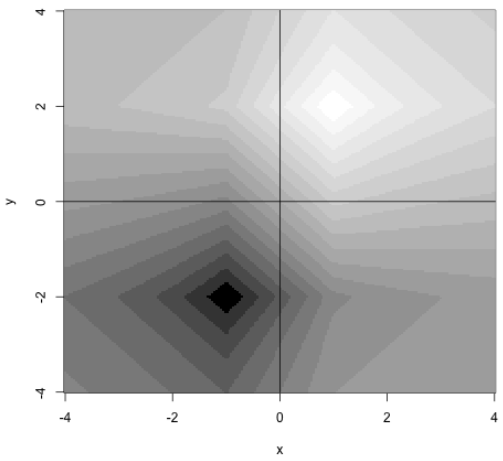

Figure 10: Jaccard similarity values obtained when comparing two real-valued vectors $\vec{x} = [x_1, x_2]^T$ and $\vec{y} = [1, 2]^T$, with $4 \leq x_1, x_2 \leq 4$.

Unlike the results obtained for the cosine similarity, now we have a well-defined and delimited peak (bright gray levels) corresponding to the position $\vec{y} = [1, 2]^T$. As expected, a minimum peak is also observed at $\vec{y} = [-1, -2]^T$. The enhanced specificity and strictness of the comparison implemented by the real-valued Jaccard index, when compared to the cosine similarity results in Figure 9, are striking. As in the one-dimensional case discussed in the previous section, the real-valued Jaccard similarity index has been able to quantify the input similarity with great accuracy while preserving a good level of generalization.

Figure 11 presents the coincidence similarity values obtained for the same comparison problem. An even more strict comparison can be obtained. Observe also, in comparison with the surface in Figure 10, the distinct shape of the level-set contours, which reflect the incorporation of the interiority index into the Jaccard similarity quantification.

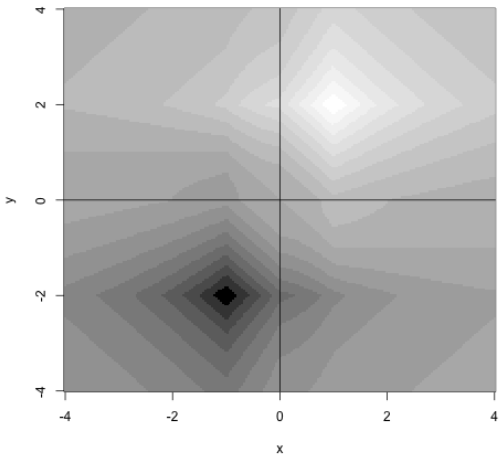

Figure 11: The coincidence similarity obtained while comparing two real-valued vectors $\vec{x} = [x_1, x_2]^T$ and $\vec{y} = [1, 2]^T$, with $4 \leq x_1, x_2 \leq 4$.

Even stricter, sharper comparisons can be obtained by using $D > 1$. Figure 14 presents the similarity surface obtained for the real-valued Jaccard index with $D = 9$. Substantially sharper peaks are obtained at the expense of reduced generalization capabaility.

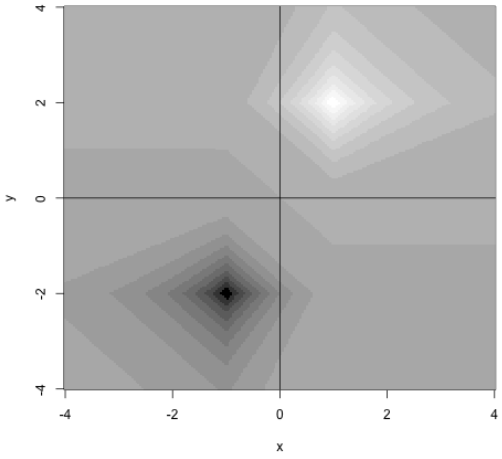

Figure 12: The Jaccard similarity obtained for $D = 9$ while comparing two real-valued vectors $\vec{x} = [x_1, x_2]^T$ and $\vec{y} = [1, 2]^T$, with $4 \leq x_1, x_2 \leq 4$..

As described in [15], both the real-valued Jaccard and real-valued coincidence indices can be generalized to incorporate a parameter α, $0 \leq \alpha \leq 1$, controlling the relative contribution of the pairwise features with the same or opposite signs on the resulting similarity values. In case $\alpha > 0.5$, the contribution of the pairwise features with the same sign will be enhanced, with the opposite taking place for α. When $\alpha = 0.5$, this index becomes identical to its parameterless version. The availability of the parameter α has been verified to enhance the level of details and modularity when of the application of the coincidence for translating datasets described by respective features into complex networks (e.g. [15]).

Figures **??** and Figures **??** presents the coincidence values obtained for the same comparison problem as above, but with $\alpha = 0.7$ and $\alpha = 0.3$, respectively.

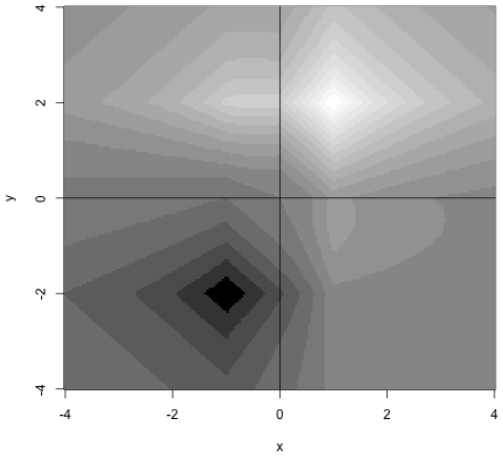

Figure 13: The coincidence similarity obtained for $D = 1$ with $\alpha = 0.7$ while comparing two real-valued vectors $\vec{x} = [x_1, x_2]^T$ and $\vec{y} = [1, 2]^T$, with $4 \leq x_1, x_2 \leq 4$. The maximum and minimum obtained coincidence values are 0.691 and -0.296, respectively, confirming the enhancement of the contribution of pairwise features with the same sign.

The effect of emphasizing the relevance of pairwise features have the same or opposite signs is marked in these figures, confirming the importance of the additional parameter α in controlling how aligned or anti-aligned pairs of features are taking into account, which can lead to enhanced comparison details. Observe also that the adoption of $\alpha \neq 0.5$ implies in the positive and negative peaks to become asymmetric, with the peak with the lower magnitude becoming shallower and less sharp.

5 Single Neuron Comparison

In this section, we perform a comparison of single neurons defined respectively to the real-valued Jaccard, interiority, and coincidence indices, as well as to the classic inner product. The similarity indices are considered for implementing the synaptic efficiency and dendritic integration

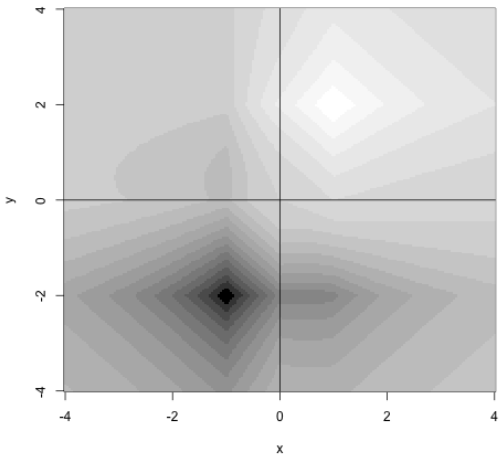

Figure 14: The coincidence similarity obtained for $D = 1$ with $\alpha = 0.3$ while comparing two real-valued vectors $\vec{x} = [x_1, x_2]^T$ and $\vec{y} = [1, 2]^T$, with $4 \leq x_1, x_2 \leq 4$. The maximum and minimum obtained coincidence values are 0.296 and -0.691, respectively, confirming the enhancement of the contribution of pairwise features with the same sign.

of stimuli up to the implantation cone. Therefore, the intrinsic non-linearity of the latter is not considered in this work. The non-linearity here is accounted by the multiset-based operations implemented at each synapsis.

This comparison is developed by taking into account several possible effects commonly found regarding pattern recognition by single neuronal cells, including: (a) relative position displacements; (b) stimulus size variation (scaling); (c) stimulus intensity variation; (d) noise; and (e) presence of more than a single pattern in the stimulus.

The reference input stimulus will be a circularly symmetry two-dimensional gaussian function centered at the stimulus space, given as:

$$g(x, y) = e^{-0.5\left(\frac{d(x,y)}{\sigma}\right)^2} \tag{19}$$

$$\textit{where: } d(x, y) = \sqrt{x^2 + y^2} \tag{20}$$

Unless otherwise stated, we adopt $\sigma = 100$ in an 200×200 image support.

Figure 15 presents the values of the four considered methods respective to relative displacements from 0 to 30 discrete steps (pixels). Full similarity has been duly identified by all methods regarding null displacement, as could be expected. However, as soon as one of the patterns shifts, the values of all indices are decreased. The sharpest decrease is verified for the coincidence approach, which is known [14, 13] to provide a more strict quantification of pairwise similarity.

The classic cross correlation presented the slowest decrease between all methods, except for displacements above 12 pixes, in which case all the indices values are already very small. This is in agreement with the identification of product based similarities [14, 13] to be particularly tolerant to pairwise differences. The real-valued Jaccard approach yielded the second fastest decreasing values.

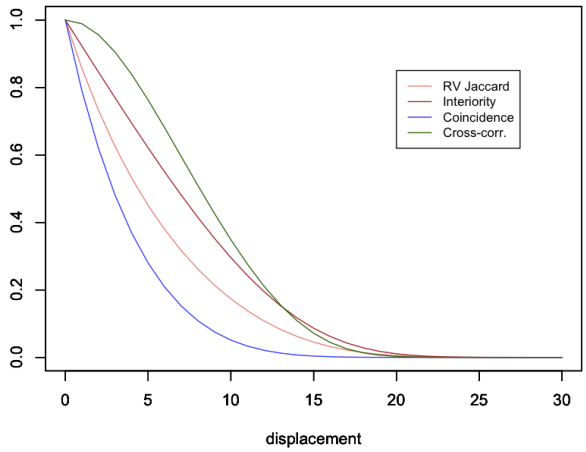

Figure 15: Similarity values obtained by the four considered methods respectively to relative displacements of two identical gaussians. The coincidence method allowed the fastest, and therefore most strict, quantification of the similarity, while the classic cross correlation yielded the most tolerant and least discriminative results.

Next, we analyze the similarity quantification in terms of varying intensities of one of the two identical gaussians, though one of them was displaced by 2 pixels along both axes in order to impose a more challenging similarity quantification. The considered intensity changes varied in a range from 0 to 3. The results are depicted in Figure 16.

Particularly interesting results can be discerned from this figure. Of greatest notice is the complete insensitivity of the classic cross-correlation method to the intensity variations. Though this feature can be helpful in some applications where intensity variance is desired, it will completely fail in cases where more strict quantifications of similarity are required to take into account also the relative intensities. The best results in this sense have been obtained with respect to the coincidence methodology, followed by the real-valued Jaccard approach. The interiority yielded a counter intuitive result, in the sense that it presented the smallest value precisely when the two compared patterns have the same intensity. That is so because of the 2 pixels displacements along the two axes.

It is also worth noticing that the two multiset-based methodologies present two main behaviors. From intensities ranging from 0 to 1, meaning that one of the patterns is less intense than the reference, both these methods

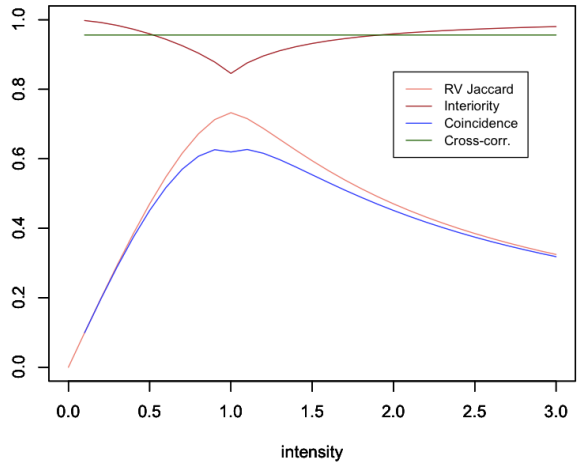

Figure 16: Quantification of the similarity between two circularly symmetric gaussians with the same dispersion, but one of them multiplied by a scaling factor from 0 to 3. In order to impose a more challenging demand, one of the gaussians was always shifted by 2 pixels along each axes. The best results are again observed for the coincidence method, followed by the real-valued Jaccard and interiority approaches. The classic cross correlated revealed to be completely insensitive to the intensity changes.

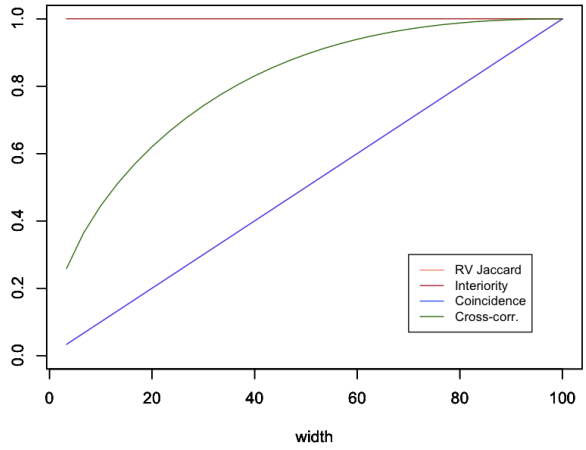

Figure 17: The similarity between the two patterns in terms of the variation of the width, corresponding to the standard deviation, of one of them. The similarity for the real-time Jaccard and coincidence approaches varies linearly from 0 to 1. The cross correlation presents a steep initial variation followed by a saturation. The kept constant at 1, which is expected given that one of the patterns is always interior to the other.

present an almost linear increase up to identical intensity. The maximum similarity value 1 was not obtained in this case because of the small imposed relative displacement of 2 pixels along each axes. From this peak, the similarity values then decrease progressively as the intensities, which now correspond to magnifications, increase.

The results of the study of the effect of the pattern width (or scaling) on the respective matching is shown in Figure 17. The width, which corresponded to the standard deviation of the circularly symmetric gaussian, given in Equation 19, varied from 0 to 100.

Both the real-valued Jaccard and the coincidence match values presented a linear increase from 0 to 1. Recall that, in this experiment, both compared patterns correspond to circularly symmetric gaussians with $\sigma = 100$ pixels.

The classic cross-correlation presented an initially steep increasing profile followed by a saturation. As expected, the interiority index was kept constant with value 1, reflecting the fact that one of the patterns is always interior to the other in this particular experiment. The real-valued Jaccard and coincidence indices represent a suitable choice in case the similarity is to reflect the width discrepancy in a linear manner. The classic cross correlation again resulted more tolerant to the implemented variation, reaching relatively high values sooner than the multiset-based methods.

We now proceed to the consideration of additive symmetric uniform noise to one of the patterns. More specifically, the following noise levels are added:

$$X\,[x,y] = X\,[x,y] + \frac{i}{N_{ns}}\,[u(x,y) - 0.5] \qquad (21)$$

with $i = 0, 1, \ldots, N_{ns}$ and where $u(x,y)$ is a scalar uniform random field taking values in [0.1]. We henceforth adopt $N_{ns} = 20$. A total of 20 experiments were performed for each of these levels, the respective average and standard deviation being then considered as results.

Figure 18 illustrates the similarity values obtained by the four methods with respect to increasing levels of noise. Several aspects of interest can be identified from this figure. First, we have that the interiority similarity accounts for the slowest decreasing similarity values. This can be explained by the fact that the noisy versions of one of the patterns, despite being jagged, will be mostly interior to the other.

The fastest decreasing profiles are those obtained for the real-valued Jaccard and coincidence methods, which also resulted very similar one another. This indicates that these two multiset-based approaches are the most sensitive to the pattern modifications induced by the increasing levels of noise. The classic cross-correlation yielded an intermediate result between the interiority and multiset-based methods, again reflecting its increase tolerance to perturbations.

The last considered type of perturbation concerns the signed addition of whole gaussian patterns into one of the images. From 1 to 5 such patterns have been added into one of the images at uniformly random positions. The

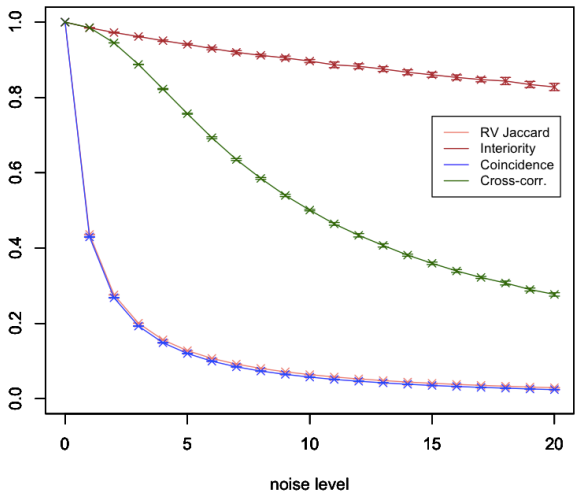

Figure 18: Similarity values obtained by the four considered methods respectively to increasing noise levels. The interiority approach is the most tolerant, followed by the classic cross-correlation, and then the two multiset-based methods. Though these curves correspond to respective averages ± standard deviations, the latter are generally very small to be visualized.

patterns can be added while being multiplied by +1 or −1, chosen in uniformly random manner. The results are shown in Figure 19.

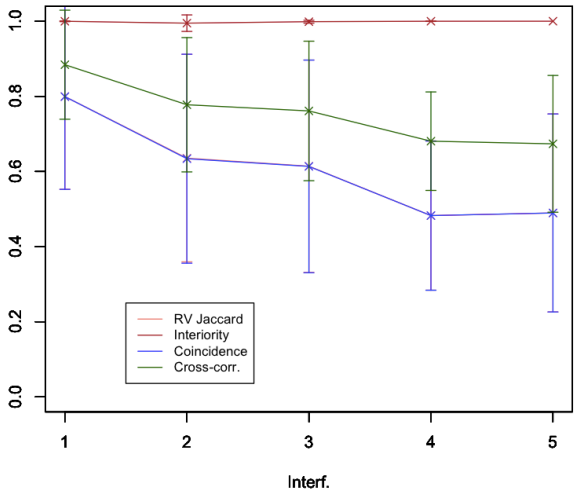

Figure 19: Similarity values obtained in presence of added interference corresponding to signed addition of from 1 to 5 gaussian patterns at uniformly random positions in the image. The curves correspond to the average ± standard deviations. Identical results have been obtained for the real-valued Jaccard and coincidence based similarities.

While the interiority and classic cross-correlation presented total tolerance to the added interference, a moderate discrimination can be observed in the case of the real-valued Jaccard and coincidence results, which present total overlap in the figure.

All in all, the several analysis reported in this section further substantiated the tendency of the multiset-based approaches to provide a more accurate and discriminative quantification of the stimulus recognition than the interiority and cross-correlation based methods. Significant differences in the specificity of the response have been observed in several cases, especially varying intensities, widths, and noise. Given their markedly more strict and discriminative characteristics, the real-valued and Jaccard, and even more so the coincidence approach, therefore correspond to the best choice, among the considered possibilities, for implementing more strict pattern recognition with high levels of accuracy.

There is an important issue to be further discussed here, and it regards the interplay between discriminative and tolerant (or invariant) performance. One first important point concerns the fact that these seem to be opposite properties, in the sense that a neuron that is too tolerant will provide no specific response, and vice versa. Another critical issue concerns the fact that, taken independently, neither of these two properties are necessarily good or bad. As in an engineering problem, the best solution will be that which best suits the specific requirements.

However, in the context of effective recognition of several types of patterns in typical applications, in presence of all the considered perturbations, perhaps the most proper solution is a balanced combination of discriminative and tolerant abilities. Actually, there is a formal solution to this duality between specificity and generality that is not so often realized. It concerns the fact that it is indeed possible to achieve both characteristics in a synergistic manner, not as a kind of trade-off or balance. This solution consists of having sets of neurons, each of which highly discriminative and specific, whose combined operation provides for the requested levels of tolerance and generalizations.

Thus, while each instance of the presented pattern will be accurate and specifically identified by successive individual cells, at the overall group level substantial tolerance will be achieved for several instances and perturbations of the presented stimuli. Nevertheless, this ideal architecture can only be achieved at expense of substantial informational resources, be then biological or artificial. These flexible *and* highly discriminative ensembles, which constitute the ideal solution for many circumstances, are henceforth denominated *synergistic neuronal systems*.

The immediate consequence of the above considerations is that it becomes critical to have the means for implementing strict, discriminative pattern recognition at the smallest informational and energetic expenses. From this perspective, the multised-based similarity identification constitute a particularly interesting resource given their

conceptual and informationally simple operation, allied to their substantially more strict and discriminative operation as verified in this work with respect to several perturbations and in [23] with respect to coexisting patterns.

Interestingly, it has been proposed recently that the multiset operations can be implemented in extremely efficient manner in analog electronics, using only a few operational amplifiers and analog switches [29], which makes the mutiset-based approaches, and in particular the coincidence index, components of choice for the development of real-time pattern recognition systems.

It remains an issue of great interest to contemplate how befitted for implementation in biological hardware the multiset operations ultimately are.

6 Concluding Remarks

The present work has developed a study of the application of the multiset-derived similarity operations, especially the real-valued Jaccard and coincidence indices, to artificial neurons. More specifically, these indices are considered for substituting the inner product performed between the image stimulus and respective matrix of synaptic weights.

After presenting an overview of the related multiset concepts and developments that led to the real-valued Jaccard and coincidence index, including new results regarding higher order respective versions, we proceeded to a systematic comparison of artificial neurons performing pattern recognition in presence of several types of perturbations. More specifically, the pattern to be recognized is stored in the synaptic weights, while the similarity comparison is performed by using the several considered indices.

The results largely confirm the enhanced potential of the coincidence index, followed by the real-valued Jaccard index, for performing strict similarity quantification. This makes these types of artificial neurons primary choices for implementations and applications involving strict patter recognition. The duality between specificity and generality in this type of task has also been discussed, and it has been argued that the ideal solution is to have large ensembles of highly specific and strict neurons, each of which adapted for taking into account specific geometric transformations so as to allow respective invariance.

Now, a particularly interesting issue arises regarding the fact that, given the substantial advantages of neurons based on the coincidence or real-valued Jaccard indices, why would they have not been adopted in biological neuronal networks aimed at effective pattern recognition? Why would the otherwise much less efficient inner product be instead implemented by the dendritic integration of the synaptic input?

There are at least two possible answers to this important question. First, we have that the biological hardware would be intrinsically unsuitable for implementing the multiset-related operations. Interestingly, recent developments have shown that these operations can be very effectively implemented in analog electronics [29], but this does not necessarily extend to biology, though much of the neuronal operation is a correlate of electric and even electronic counterparts. If it happens that biology is intrinsically unsuitable for performing multiset operations, these alternatives remain still valid for implementations in other types of hardware.

The second possible answer is that the biological neuronal cells actually implement multiset-related functions. Indeed, consider the profile of the operation $x \sqcap y$ shown in Figure 5, which is the basis for all effective indices developed and applied in the present work. This function, which resembles a sigmoid, could be applied not at the implantation cone, but at each of the synapses. Indeed, the observed saturation could correspond to the saturation of the synaptic activation and/or of the local polarization of the interior of the cell. The sum corresponding to the numerator of Equation 10 would then correspond to the combination of the diffusive charge effect at the implantation cone.

As for the denominator of that same equation, it is possible that other intracellular mechanisms are activated by the synaptic activity that effectively contribute to the inhibition of the action potential. These inhibitory effects could be similarity integrated at the implantation cone, accounting for the denominator in Equation 10. There are other possible mechanisms that could account for the implementation of multiset-like neuronal operations. For instance, the denominator of Equation 10 could correspond to inhibitory effects received from other cells associated to the same receptive field that would therefore counterbalance the net depolarization of the excitatory cell implementing the numerator integration.

Though these are currently hypothetical, further consideration and experimental developments can help verifying these possibilities.

The concepts, methods, and results reported in the present work have several potential implications in a wide range of areas — including neuroscience, pattern recognition and deep learning — therefore paving the way to a large number of further developments. Some examples include further studies of the possible relationships with biological cells, the consideration of other types of stimuli, as well as the evaluation of the here introduced higher order versions of the real-valued Jaccard and coincidence indices.

Acknowledgments.

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) and FAPESP (grant 15/22308-2).

References

[1] S. Haykin. *Neural Networks And Learning Machines*. McGraw-Hill Education, 9th edition, 2013.

[2] Warren Mcculloch and Walter Pitts. A logical calculus of ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5:127–147, 1943.

[3] D. H. Hubel and T. N. Wiesel. *Brain and Visual Perception: The Story of a 25-Year Collaboration*. Oxford University Press, Oxford, 2004.

[4] D. Hubel and T. Wiesel. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

[5] M. H. Turner, G. W. Schwartx, and F. Rieke. Receptive field center-surround interactions mediate context-dependent spatial contrast encoding in the retina. *eLIfe*, 7:eLife 2018;7:e38841, 2018.

[6] E. O. Brigham. *Fast Fourier Transform and its Applications*. Pearson, 1988.

[7] K. R. Rao, D. N. Kim, and J. J. Hwang. Integer fast fourier transform. In *Fast Fourier Transform - Algorithms and Applications. Signals and Communication Technology*, pages 111–126. Springer, Dorcrecht, 2010.

[8] A. V. Oppenheim and R. Schafer. *Discrete-Time Signal Processing*. Pearson, 2009.

[9] C. Phillips, J. Parr, and E. Riskin. *Signals, Systems and Transforms*. Pearson, 2013.

[10] S. Ramon y Cajal. *Recollections of My Life*. The MIT Press, Cambridge, Mass., 1996.

[11] R. Friedman. Measurements of neuronal morphological variation across the rat neocortex. *Neuroscience Letters*, 734, 2020.

[12] W. B. Grueber, C.-H. Yang, B. Ye, and Y.-N. Jan. The development of neuronal morphology in insects. *Current Biology*, 730–738, 2005.

[13] L. da F. Costa. On similarity. https://www.researchgate.net/publication/355792673_On_Similarity, 2021. [Online; accessed 21-Aug-2021].

[14] L. da F. Costa. Further generalizations of the Jaccard index. https://www.researchgate.net/publication/355381945_Further_Generalizations_of_the_Jaccard_Index, 2021. [Online; accessed 21-Aug-2021].

[15] L. da F. Costa. Coincidence complex networks. https://iopscience.iop.org/article/10.1088/2632-072X/ac54c3, 2022. J. Phys.: Compl, 3 015012.

[16] J. Hein. *Discrete Mathematics*. Jones & Bartlett Pub., 2003.

[17] D. E. Knuth. *The Art of Computing*. Addison Wesley, 1998.

[18] W. D. Blizard. Multiset theory. *Notre Dame Journal of Formal Logic*, 30:36—66, 1989.

[19] W. D. Blizard. The development of multiset theory. *Modern Logic*, 4:319–352, 1991.

[20] P. M. Mahalakshmi and P. Thangavelu. Properties of multisets. *International Journal of Innovative Technology and Exploring Engineering*, 8:1–4, 2019.

[21] D. Singh, M. Ibrahim, T. Yohana, and J. N. Singh. Complementation in multiset theory. *International Mathematical Forum*, 38:1877–1884, 2011.

[22] L. da F. Costa. Multisets. https://www.researchgate.net/publication/355437006_Multisets, 2021. [Online; accessed 21-Aug-2021].

[23] L. da F. Costa. Comparing cross correlation-based similarities. https://www.researchgate.net/publication/355546016_Comparing_Cross_Correlation-Based_Similarities, 2021. [Online; accessed 21-Oct-2021].

[24] K. S. S. Kumar. *Electric Circuits and Networks*. Pearson Education India, 2009.

[25] B. Mirkin. *Mathematical Classification and Clustering*. Kluwer Academic Publisher, Dordrecth, 1996.

[26] C. E. Akbas, A. Bozkurt, M. T. Arslan, H. Aslanoglu, and A. E. Cetin. L1 norm based multiplication-free cosine similiarity measures for big data analysis. In *IEEE Computational Intelligence for Multimedia Understanding (IWCIM)*, France, Nov. 2014.

[27] C. E. Akbas, A. Bozkurt, A. E. Cetin, R. Cetin-Atalay, and A. Uner. Multiplication-free neural networks. In *Signal Processing and Communications Applications Conference (SIU)*, Malatya, Turkey, May. 2015.

[28] L. da F. Costa. Generalized multiset operations. https://www.researchgate.net/profile/Luciano-Da-F-Costa, 2021. [Online; accessed 10-Nov-2021].

[29] L. da F. Costa. Multiset signal processing and electronics. https://www.researchgate.net/publication/355954430_Multiset_Signal_Processing_and_Electronics, 2021. [Online; accessed 21-Nov-2021].

[30] M. K. Vijaymeena and K. Kavitha. A survey on similarity measures in text mining. *Machine Learning and Applications*, 3(1):19–28, 2016.