An Experimental Study on Asymmetric Self-Organizing Map

Dominik Olszewski

Faculty of Electrical Engineering,
Warsaw University of Technology, Poland
olszewsd@ee.pw.edu.pl

Abstract. The paper presents an extension of the justification for use of the asymmetric Self-Organizing Map (SOM). We claim that it can successfully applied in the wider area of research than the textual data analysis. The results of our experimental study in the fields of sound recognition and heart rhythm recognition confirm this claim, and report the superiority of the asymmetric approach over the symmetric one, in both parts of our experiments.

Keywords: Self-Organizing Map, Asymmetric Self-Organizing Map, Sound recognition, Heart rhythm recognition.

1 Introduction

The Self-Organizing Map (SOM) [1] is an example of the artificial neural network architecture. It was introduced by T. Kohonen, and it can be also interpreted as a visualization technique, since the algorithm performs a projection from multidimensional space to 2-dimensional space, this way creating a map structure. The location of points in 2-dimensional grid aims to reflect the similarities between the corresponding objects in multidimensional space. Therefore, the SOM algorithm allows for visualization of relationships between objects in multidimensional space. The asymmetric version of the SOM algorithm was introduced in [2]. However, the justification provided in this paper was related to the hierarchical associations in textual data. The aim of this paper is to show that similar phenomenon can be found in case of the sound signals and human heart rhythm signals. Consequently, the same assertion referring to hierarchical asymmetric relationships in data can be used to justify the use of the asymmetric SOM algorithm version. In other words, we can assert that our paper extends the range of application of the asymmetric SOM method.

The rest of this paper is organized as follows: in Section 2, the traditional symmetric version of the SOM algorithm is described; in Section 3, the asymmetric relationships in data sets are discussed; in Section 4, the asymmetric version of the SOM algorithm is presented; in Section 5, our experimental results are reported; while Section 6 summarizes the whole paper, and gives some concluding remarks.

H. Yin, W. Wang, and V. Rayward-Smith (Eds.): IDEAL 2011, LNCS 6936, pp. 42–49, 2011.

2 Symmetric Self-Organizing Map

The SOM algorithm provides a non-linear mapping between a high-dimensional original data space and a 2-dimensional map of neurons. The neurons are arranged according to a regular grid, in such a way that the similar vectors in input space are represented by the neurons close in the grid. Therefore, the SOM technique visualize the data associations in the input high-dimensional space.

It was shown in [3] that the results obtained by the SOM method are equivalent to the results obtained by optimizing the following error function:

$$e(\mathcal{W}) = \sum_r \sum_{x_\mu \in V_r} \sum_s h_{rs} D(x_\mu,\ w_s) \tag{1}$$

$$\approx \sum_r \sum_{x_\mu \in V_r} D(x_\mu,\ w_r)\ +\ K \sum_r \sum_{s \neq r} h_{rs} D(w_r,\ w_s)\ , \tag{2}$$

where x_μ are the objects in high-dimensional space, w_r and w_s are the prototypes of objects on the grid, h_{rs} is a neighborhood function (for example, the Gaussian kernel) that transforms non-linearly the neuron distances (see [1] for other choices of neighborhood functions), $D(\bullet,\ \bullet)$ is the squared Euclidean distance, and V_r is the Voronoi region corresponding to prototype w_r. The number of prototypes is sufficiently large so that $D(x_\mu,\ w_s) \approx D(x_\mu,\ w_r)\ +\ D(w_r,\ w_s)$.

According to equation (2), the SOM error function can be decomposed as the sum of the quantization error and the topographic error. The first one minimizes the loss of information, when the input patterns are represented by a set of prototypes. By minimizing the second one, we assure the maximal correlation between the prototype dissimilarities and the corresponding neuron distances, this way assuring the visualization of the data relationships in the input space.

The SOM error function can be optimized by an iterative algorithm consisting of two steps (discussed in [3]). First, a quantization algorithm is executed. This algorithm represents each input pattern by the nearest neighbor prototype. This operation minimizes the first component in equation (2). Next, the prototypes are arranged along the grid of neurons by minimizing the second component in the error function. This optimization problem can be solved explicitly using the following adaptation rule for each prototype [1]:

$$w_s\ =\ \frac{\sum_{r=1}^{M} \sum_{x_\mu \in V_r} h_{rs} x_\mu}{\sum_{r=1}^{M} \sum_{x_\mu \in V_r} h_{rs}}\ , \tag{3}$$

where M is the number of neurons, and h_{rs} is a neighborhood function (for example, the Gaussian kernel of width $\sigma(t)$). The width of the kernel is adapted in each iteration of the algorithm using the rule proposed by [4], i.e., $\sigma(t)\ =\ \sigma_i (\sigma_f/\sigma_i)^{t/N_{iter}}$, where $\sigma_i\ \approx\ M/2$ is typically assumed in the literature (for example, in [1]), and σ_f is the parameter that determines the smoothing degree of the principal curve generated by the SOM algorithm [4].

3 Asymmetry in Data

The problem of asymmetry in data analysis was widely studied in the literature. The research of A. Okada and T. Imaizumi [5] is focused on using the dominance point governing asymmetry in the proximity relationships among objects, represented as points in the multidimensional Euclidean space. They claim that ignoring or neglecting the asymmetry in proximity analysis discards potentially valuable information. On the other hand, B. Zielman and W. Heiser in [6] consider the models for asymmetric proximities as a combination of a symmetric similarity component and an asymmetric dominance component. The author of [7], introduces the asymmetric version of the well-known k-means clustering algorithm. Finally, the paper [2] proposes the asymmetric version of the SOM algorithm, which was an inspiration for our research.

When an analyzed data set appears to have asymmetric properties, the symmetric measures of similarity or dissimilarity (for example, the most popular Euclidean distance) does not apply properly to this phenomenon, and for most pairs of data points, they produce small values (similarities) or big values (dissimilarities). Consequently, they do not reflect accurately the relationships between objects. The asymmetry in data set arises, for example, in case, when the data associations have a hierarchical nature. The hierarchical connections in data are closely related to the asymmetry. This relation has been noticed in [8]. In case of the dissimilarity, when it is computed in the direction – from a more general entity to a more specific one – it should be greater than in the opposite direction. As stated in [2], asymmetry can be interpreted as a particular type of hierarchy.

An idea to overcome this problem is to employ the asymmetric similarities and dissimilarities. They should be applied in algorithms in such way, so that they would properly reflect the hierarchical asymmetric relationships between objects in the analyzed data set. Therefore, it should be guaranteed that their application is consistent with the hierarchical associations in data. This can be achieved by use of the asymmetric coefficients, inserted in the formulae of symmetric measures. This way, we can obtain the asymmetric measures on the basis of the symmetric ones. The asymmetric coefficients should assure the consistence with the hierarchy. Hence, in case of the dissimilarities, they should assure greater values in the direction – from more general concept to more specific one.

This paper points out that the phenomenon of the hierarchy-caused asymmetry occurs in a wider range of applications than the text analysis, as it was presented in [2]. Our experimental study concerns the sound signals clustering and human heart rhythm clustering, and confirms the existence of the same phenomenon.

3.1 Asymmetric Coefficients

Asymmetric coefficients convey the information provided by asymmetry. Two coefficients were introduced in [9]. The first one is derived from the fuzzy logic similarity, and the second one formulated on the basis of the Kullback-Leibler divergence. Both of these quantities are widely used in statistics and probability theory. In our experimental study, we have used the first of these coefficients.

Hence, the fuzzy-logic-based asymmetric coefficient is formulated as follows:

$$a_i = \frac{|f_i|}{\max_j \left(|f_j|\right)} , \tag{4}$$

where f_i are the features of objects in the analyzed data set (f_i are the entries of the vectors representing the objects), and $|\bullet|$ is the L_1-norm meaning the number of objects possessing the feature given as the argument.

This coefficient takes values in the $[0, 1]$ interval. Intuitively speaking, it will become large for general (broad) concepts with large L_1-norm.

Note that the asymmetric coefficients must be computed and assigned to each feature of every object in the analyzed data set.

4 Asymmetric Self-Organizing Map

In order to formulate the asymmetric version of the SOM algorithm, we will refer to the error function (2). As it was stated in Section 2, the results produced by the SOM method are identical to the results obtained by optimizing the function (2).

The asymmetric SOM algorithm is derived in three steps:

Step 1. Transform a symmetric dissimilarity (for example, the Euclidean distance) into a similarity:

$$S_{ij}^{\mathrm{SYM}} = C - d^2 \left(x_i, x_j\right) , \tag{5}$$

where $d^2 \left(x_i, x_j\right)$ is the squared Euclidean distance between objects x_i and x_j, and the constant C is the upper boundary of the squared Euclidean distance.

Step 2. Transform the symmetric similarity into the asymmetric similarity:

$$S_{ij}^{\mathrm{ASYM}} = a_i \left(C - d^2 \left(x_i, x_j\right)\right) , \tag{6}$$

where a_i is the asymmetric coefficient defined in Subsection 3.1, in (4), and the rest of notation is described in (5). The asymmetric similarity defined this way, with use of the asymmetric coefficient guarantees the consistency with the asymmetric hierarchical associations among objects in the data set.

Step 3. Insert the asymmetric similarity in the error function (2), in order to obtain the energy function, which needs to maximized:

$$E\left(\mathcal{W}\right) = \sum_r \sum_{x_\mu \in V_r} \sum_s h_{rs} a_i \left(C - d^2 \left(x_i, x_j\right)\right) , \tag{7}$$

where the notation is explained in (2), (5), and (6). The energy function (7) can be optimized in the similar way as the error function (2). Firstly, we run the quantization algorithm, which generates the SOM prototypes w_s. Secondly, the energy function is maximized by solving the set of linear equations $\partial E(\mathcal{W})/\partial w_s = 0$. This system of linear equation can be solved explicitly, by using the following updating formula for the SOM prototypes w_s:

$$w_s = \frac{\sum_{r=1}^{M} \sum_{x_\mu \in V_r} h_{rs} a_\mu x_\mu}{\sum_{r=1}^{M} \sum_{x_\mu \in V_r} h_{rs} a_\mu}, \tag{8}$$

where a_i is the asymmetric coefficient, h_{rs} is a neighborhood function (for example, the Gaussian kernel), and the rest of notation is the same as in (3). This updating formula is similar to the adaptation rule (3), with the difference that the asymmetric coefficient is inserted. In case of use of the Gaussian kernel, its width $\sigma(t)$ can be adapted, like it is done in (3). An important property of the asymmetric SOM algorithm is that it maintains the simplicity of the traditional symmetric approach, and does not increase the computational complexity.

5 Experiments

Our experimental study aims to confirm that the asymmetric version of the SOM algorithm can be applied in the wider area of research than it was proposed in [2], and the justification referring to the textual data analysis can be extended to the other types of data, for example, to sound signals and human heart rhythm signals, which were the subject of our empirical study.

In case of both parts of our experiments, we have compared the results obtained with use of the symmetric and asymmetric SOM algorithms. As the basis of the comparisons, i.e., as the evaluation metrics, we have used the accuracy degree [7], and the entropy measure [2].

- **Accuracy degree.** This evaluation metric determines the number of correctly assigned objects divided by the total number of objects in the data set. Firstly, the centroids of the clustered data are computed, and next, each object is assigned to the cluster represented by the centroid – nearest to this object. Finally, the number of correctly assigned objects is divided by the total number of all objects. The accuracy degree assumes values in the interval $[0, 1]$, and naturally, greater values are preferred.
- **Entropy measure.** This evaluation metric determines the number of overlapping objects divided by the total number of objects in the data set. This means, the number of objects, which are in the overlapping area between clusters, divided by the total number of objects. In other words, it determines the uncertainty for the classification of objects that belong to the same cluster. The entropy measure assumes values in the interval $[0, 1]$, and, smaller values are desirable.

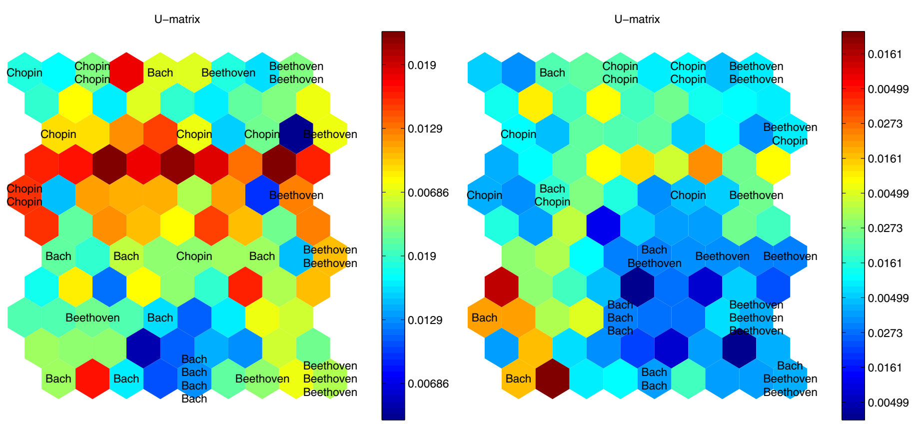
The sound signals, we have analyzed, were the piano music recordings, and the human heart rhythm signals were analyzed on the basis of the ECG recordings derived from the MIT-BIH ECG Databases.

5.1 Piano Music Composer Clustering

In this part of our experiments, we have tested our enhancement to the SOM algorithm and the classical SOM forming three clusters representing three piano music composers: Johann Sebastian Bach, Ludwig van Beethoven, and Fryderyk Chopin. Each music piece was represented with a 20-seconds sound signal sampled with the 44100 Hz frequency. The entire data set was composed of 32 sound signals. The feature extraction process was carried out according to the traditional Discrete-Fourier-Transform-based (DFT-based) method. The DFT was implemented with the fast Fourier transform (FFT) algorithm. Sampling signals with the 44100 Hz frequency resulted in the 44100/2 Hz value of the upper boundary of the FFT result range.

The results of this part of our experiments are demonstrated in Fig. 1 and in Table 1. Figure 1 presents the U-matrices generated by the symmetric (Fig. 1(a)) and asymmetric (Fig. 1(b)) SOM algorithms. Table 1, in turn, presents the accuracy degrees and the entropy measures corresponding to the symmetric and asymmetric SOM approaches.

The results of this part of our experimental study report the superiority of the asymmetric SOM algorithm over the symmetric counterpart. The asymmetric approach leads to the higher clustering accuracy measured on the basis of the accuracy degree (0.9375 vs. 0.8438), and also, it leads to the lower cluster overlapping determined on the basis of the entropy measure (0.1563 vs. 0.2500).



(a) Symmetric SOM (b) Asymmetric SOM

Fig. 1. Piano Music Composer Clustering Maps

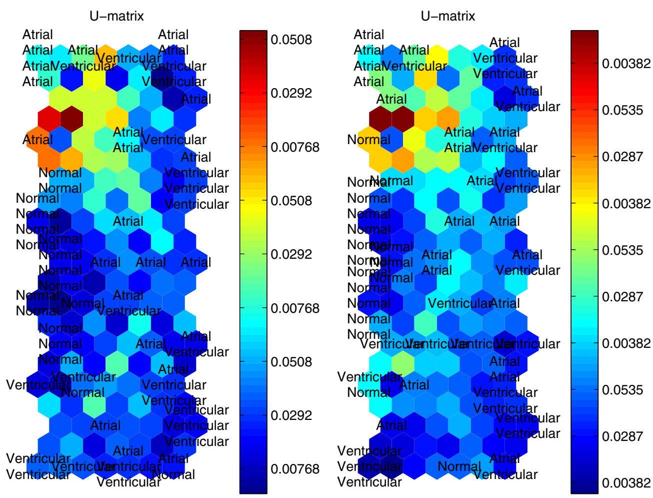
Table 1. Accuracy degrees and entropy measures of the piano music composer clustering

	Symmetric SOM	Asymmetric SOM
Accuracy degree	$27/32 = 0.8438$	$30/32 = 0.9375$
Entropy measure	$8/32 = 0.2500$	$5/32 = 0.1563$

5.2 Human Heart Rhythm Clustering

In this part of our experiments, we have investigated asymmetric SOM and the traditional SOM approach forming three clusters representing three types of human heart rhythms: normal sinus rhythm, atrial arrhythmia, and ventricular arrhythmia. This kind of clustering can be interpreted as the cardiac arrhythmia detection and recognition based on the ECG recordings. In general, the cardiac arrhythmia disease may be classified either by rate (tachycardias – the heart beat is too fast, and bradycardias – the heart beat is too slow) or by site of origin (atrial arrhythmias – they begin in the atria, and ventricular arrhythmias – they begin in the ventricles). Our clustering recognizes the normal rhythm, and, also, recognizes arrhythmias originating in the atria, and in the ventricles. We analyzed 20-minutes ECG holter recordings sampled with the 250 Hz frequency. The entire data set was composed of 63 ECG signals. The feature extraction was carried out in the same way, like it was done with the piano music composer clustering.

The results of this part of our experiments are presented in Fig. 2 and in Table 2, which are constructed in the same way as in Subsection 5.1.



(a) Symmetric SOM (b) Asymmetric SOM

Fig. 2. Human Heart Rhythm Clustering Maps

It is clear that, in the case of the ECG recording clustering, the asymmetric SOM method, again, outperformed the symmetric one, by providing the higher clustering quality (accuracy degree: 0.7778 vs. 0.7143), and the lower clustering uncertainty (entropy measure: 0.2540 vs. 0.2857).

Table 2. Accuracy degrees and entropy measures of the human heart rhythm clustering

	Symmetric SOM	Asymmetric SOM
Accuracy degree	$45/63 = 0.7143$	$45/63 = 0.7143$
Entropy measure	$18/63 = 0.2857$	$16/63 = 0.2540$

6 Summary

The paper presented the results of the experimental study on the asymmetric SOM algorithm, in the fields of piano music composer clustering, and human heart rhythm clustering. According to our experimental results, the asymmetric SOM outperforms the symmetric one, in both studied cases, by providing the higher clustering accuracy, and lower entropy measure. This means that our results confirmed that the hierarchy-caused asymmetric relationships also occur in the analyzed data sets. This conclusion extends the justification of use of the asymmetric SOM algorithm beyond the textual data analysis, which was the aim of this paper.

References

1. Kohonen, T.: Self-Organizing Maps, 3rd edn. Springer, Heidelberg (2001)
2. Martín-Merino, M., Muñoz, A.: Visualizing Asymmetric Proximities with SOM and MDS Models. Neurocomputing 63, 171–192 (2005)
3. Heskes, T.: Self-Organizing Maps, Vector Quantization, and Mixture Modeling. IEEE Transactions on Neural Networks 12(6), 1299–1305 (2001)
4. Mulier, F., Cherkassky, V.: Self-Organization as an Iterative Kernel Smoothing Process. Neural Computation 7(6), 1165–1177 (1995)
5. Okada, A., Imaizumi, T.: Multidimensional Scaling of Asymmetric Proximities with a Dominance Point. In: Advances in Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization, pp. 307–318. Springer, Heidelberg (2007)
6. Zielman, B., Heiser, W.J.: Models for Asymmetric Proximities. British Journal of Mathematical and Statistical Psychology 49, 127–146 (1996)
7. Olszewski, D.: Asymmetric k-Means Algorithm. In: Dobnikar, A., Lotrič, U., Šter, B. (eds.) ICANNGA 2011, Part II. LNCS, vol. 6594, pp. 1–10. Springer, Heidelberg (2011)
8. Muñoz, A., Martin, I., Moguerza, J.M.: Support Vector Machine Classifiers for Asymmetric Proximities. In: Kaynak, O., Alpaydın, E., Oja, E., Xu, L. (eds.) ICANN 2003 and ICONIP 2003. LNCS, vol. 2714, pp. 217–224. Springer, Heidelberg (2003)
9. Muñoz, A., Martín-Merino, M.: New Asymmetric Iterative Scaling Models for the Generation of Textual Word Maps. In: Proceedings of the International Conference on Textual Data Statistical Analysis JADT 2002, pp. 593–603 (2002)