A modification of the SOM algorithm to deal with non-symmetric proximities

Manuel Martín-Merino
Universidad Pontificia de Salamanca, C/Compañía 5, 37002, Salamanca, Spain
mmartinmac@upsa.es
Alberto Muñoz
Universidad Carlos III de Madrid, C/Madrid, 126, 28903 Getafe, Spain
albmun@est-econ.uc3m.es

Abstract - *Self Organizing Maps (SOM) are often applied to the visualization of high dimensional data. The primary source of information to this aim is the (symmetric) distance data matrix. However, there are many interesting problems in which distances or proximities are inherently asymmetric.*
In this paper we extend the SOM algorithm to deal successfully with asymmetric relations. The new models are tested on the challenging problem of word relation visualization using real datasets. The experimental results show that the asymmetric variants proposed outperform the original SOM algorithm.

Key words - Self Organizing Maps, asymmetry, visualization algorithms, textual data analysis

1 Introduction

Self Organizing Maps (SOM) [10] are non-linear visualization techniques helpful to discover meaningful relationships in multivariate data sets. They have been applied to a wide range of applications including text mining problems [11, 15]. Several variants of the original SOM algorithm have been proposed in the literature [10, 18] but usually they rely on the use of symmetric measures such as the Euclidean. Therefore they are not expected to handle asymmetric relations appropriately.

However, there are a number of relevant applications such as text mining in which relations are inherently asymmetric. Consider for instance the problem of word relation visualization. The relation between a broad term such as 'statistics' and a specific term like 'bayesian' is asymmetric in the sense that 'statistics' subsumes the semantic meaning of 'bayesian' but not conversely. In this case, a symmetric similarity would suggest that 'statistics' is hardly related to 'bayesian' which is not true [13, 17]. Therefore new dissimilarities should be proposed that reflect more accurately this kind of proximities.

Some Multidimensional Scaling (MDS) algorithms have been proposed to deal with asymmetric measures (see for instance [6, 21] and references therein). However, these algorithms derive the object proximities considering only the symmetric component of the similarity matrix, losing important information about the data set (see [14] for details).

In this paper we start from an interesting case in text mining to motivate the main ideas of the paper: we first study the impact of asymmetry over the quality of the textual maps and present new dissimilarities that are less sensitive to this problem. Next new versions of the SOM algorithm are proposed that incorporate the new dissimilarities keeping the simplicity of the original SOM. The derivation of the algorithms from an energy function provides a theoretical foundation for the new models. Finally, the algorithms proposed are tested on the interesting problem of word relation visualization.

This paper is organized as follows. In section 2 we study the problem of asymmetry. Section 3 proposes new versions of the SOM algorithm that are able to deal with asymmetric relations. In section 4 the new algorithms are tested using two real textual collections. Finally section 5 gets conclusions and outlines future research trends.

2 Asymmetry

In this section we study the effect that the asymmetry has on visualization algorithms based on symmetric distances. Next the relation between asymmetry and the L_1 norm is established. Finally the meaning and relevance of asymmetry in the field of textual data analysis is discussed.

Consider a set of n objects and let $S = (s_{ij})$ be the similarity matrix made up of object proximities. If a dissimilarity matrix (δ_{ij}) is given instead it can be transformed easily into a similarity using any of the transformations provided in [6] $(s_{ij} = 1 - \delta_{ij})$. Asymmetry arises when $s_{ij} \neq s_{ji}$. In this case the dissimilarity matrix can be decomposed into a symmetric and skew-symmetric component $(S = M + A)$ [21] where $m_{ij} = (s_{ij} + s_{ji})/2$ and $a_{ij} = (s_{ij} - s_{ji})/2$. The first term represents the object proximities and the second one the deviation from symmetry (it equals 0 if S is symmetric).

When asymmetry arises the symmetric similarities usually considered in the literature produce often too small values and fail to reflect the object proximities [17, 14]. To get a deeper insight into this problem we are going to study a text mining example.
Consider a collection of scientific papers where, the broad term 'statistics' appears for instance in 500 documents while the more specific term 'bayesian' appears only in a subset of 10 documents. Obviously, the relation between 'statistics' and 'bayesian' is highly asymmetric in the sense that 'statistics' subsumes the semantic meaning of 'bayesian' but not conversely.

Consider now a symmetric similarity such as the cosine that has been widely used in the information retrieval literature [19, 5]. This measure is equivalent to the Euclidean distance (commonly used by SOM algorithms) if the objects are normalized previously by the L_2 norm. Moreover, the cosine similarity represents somewhat the behavior of a broad range of symmetric similarities over textual data [5, 14].
Now, if the cosine similarity is computed for the example considered above we get a value of 0.14 which is close to 0. This would suggest that 'statistics' is hardly related to 'bayesian', which is not true. Notice that other distances such as the χ^2 [12] are similarly affected by the same problem [5].

The previous example suggests that commonly used symmetric similarities become meaningless (too small) when asymmetry grows large. Moreover, this bias toward small values tends to reduce the variance of the similarity histogram. In particular, the cosine similarity has a standard deviation as low as 0.03 for the datasets considered in this paper. Consequently, the similarities become almost constant over textual data and any algorithm based on distances will be highly distorted [4].

Next we interpret the asymmetry in terms of the object L_1 norm.
Let $\vec{x}_i$ be the vector space representation [2] of term i and $|\vec{x}_i|$ the L_1 norm which is proportional to the term frequency in the database. Consider a similarity such as the fuzzy logic defined as:

$$s_{ij} = \frac{| \vec{x}_i \wedge \vec{x}_j |}{| \vec{x}_i |} \tag{1}$$

where $\wedge$ denotes the standard fuzzy logic intersection and $||$ the L_1 norm. Obviously this similarity is asymmetric and the skew-symmetric component can be written as:

$$a_{ij} \propto | \vec{x}_j | - | \vec{x}_i | \; . \tag{2}$$

This equation suggests that the asymmetry is a property associated to individual objects and may be modeled by the following coefficient of asymmetry: $\omega_i = \frac{|\vec{x}_i|}{\max_k |\vec{x}_k|}$. In the context of text mining this coefficient will become large for broad terms that appear in a wide range of documents.
The previous equation shows that the asymmetry becomes large just for relations between broad terms (large L_1 norm) and specific terms (small L_1 norm). Therefore, the asymmetry will be an important factor in many applications such as text mining in which the L_1 norm obeys a Zipf's law [2]. In this case the L_1 norm histogram is very skew and a_{ij} will become large quite often.
In addition it has been pointed out in [1] that for sparse databases such as textual datasets, the relation among specific (low norm) terms can only be established through relations with related broader terms. Therefore if this kind of asymmetric relations are underestimated, the position of specific terms in the map will become meaningless. To avoid this problem the proximities corresponding to asymmetric relations should be compensated proportionally to the degree of asymmetry (a_{ij}) as we will see in the next section.

3 Asymmetric variants of Self Organizing Maps

In this section we first introduce shortly the SOM algorithm proposed originally by [10]. Next we propose two asymmetric variants that take advantage of the asymmetry to reflect accurately the object proximities. Finally a brief remark about related work is given.

The SOM [10] is a nonlinear visualization technique for high dimensional data. Input vectors are represented by neurons arranged according to a regular grid (usually $1D$-$2D$) in such a way that similar vectors in input space become spatially close in the grid.
From a practical point of view the SOM is equivalent to the algorithm resulting from the optimization of the following quantization error [18]:

$$E(\mathcal{W}) = \sum_r \sum_{x_\mu \in V_r} \sum_s h_{rs} D(\vec{x}_\mu, \vec{w}_s) \, , \tag{3}$$

where D denotes the square Euclidean distance and V_r is the Voronoi region corresponding to prototype $\vec{w}_r$. h_{rs} is a neighborhood function (for instance a Gaussian kernel) that transforms nonlinearly the neuron distances (see [10] for other possible choices). The kernel width is adapted in each iteration and determines the degree of smoothing of the principal curve [16]. The error function (3) is minimized when objects that are close together in input space (according to the Euclidean distance) are mapped to neighboring neurons in the grid.
Notice that the error function proposed by Heskes [18] assumes that the Voronoi regions are computed considering the distance $D(\vec{x}_\mu, \vec{w}_r) = \sum_s h_{rs} D(\vec{x}_\mu, \vec{w}_s)$. If the Euclidean distance is considered instead, the derivation of the SOM algorithm from an error function is only possible for the discrete case.

The SOM energy function may be optimized by an iterative algorithm made up of two steps [18]. First a quantization algorithm is run that represents each pattern by the nearest neighbor prototype. Next, the prototypes are organized along the grid of neurons by minimizing the error function (3). The optimization problem can be solved explicitly resulting in a simple iterative adaptation rule for each prototype [10].

Next, two asymmetric variants of the original SOM are proposed. The goal of the new models is to improve particularly the position of the more specific terms (low L_1 norm). To this aim, a new asymmetric similarity based on the Euclidean distance is defined that reflects accurately the proximities among specific and broader terms (corresponding to asymmetric relations). Next an energy function which incorporates the asymmetric similarity is introduced. Finally the error function is optimized keeping the simplicity of the original algorithm.

Let $d(\vec{x}_i, \vec{x}_j) = \|\vec{x}_i - \vec{x}_j\|^2$ be the square Euclidean distance usually considered in the Self Organizing Maps. This dissimilarity can be easily transformed into a similarity [21, 6] using for instance the following transformation: $s_{ij} = K - \|\vec{x}_i - \vec{x}_j\|^2$, where the constant K is an upper bound for the square Euclidean distances. Now an asymmetric index is defined as follows:

$$s_{ij} = (K - \|\vec{x}_i - \vec{x}_j\|^2)\omega_i \,, \tag{4}$$

where ω_i denotes the asymmetry coefficient defined in section 2. The object proximities induced by the symmetric component of s_{ij} have now the following form:

$$s_{ij}^{(s)} = (K - \|\vec{x}_i - \vec{x}_j\|^2)\frac{\omega_i + \omega_j}{2} \tag{5}$$

When the object relations are highly asymmetric then $\omega_i \gg \omega_j$ or conversely. In this case the similarity (5) compensates the value of the Euclidean proximity proportionally to the degree of asymmetry defined by equation (2). This can be easily seen considering that the degree of asymmetry is proportional to $|\omega_i - \omega_j|$ (see section 2) and under the above conditions it can be approximated by $\max(\omega_i, \omega_j) \approx \omega_i + \omega_j$.
Obviously, the equation (5) compensates only the similarities between related terms such as for instance 'statistics' and 'bayesian' because the Euclidean proximity is not zero. Otherwise, the similarities between unrelated terms such as 'statistics' and 'petal' will remain unaltered because the Euclidean proximity is zero in this case.
Substituting the similarity (5) into equation (3) the error function for the asymmetric SOM can be written as:

$$E(\mathcal{W}) = \sum_r \sum_{x_\mu \in V_r} \sum_s h_{rs}\omega_\mu (K - \|\vec{x}_\mu - \vec{w}_s\|^2) \,. \tag{6}$$

As we have mentioned earlier, this asymmetric error decomposes into a symmetric component that represents the object proximities and a skew-symmetric component that represents the deviation from symmetry. The skew-symmetric component can be neglected in this case because the sum of the non-diagonal elements of a skew-symmetric matrix equals 0 [21]. In the new error function the Euclidean proximities in input space are compensated proportionally to the degree of asymmetry (ω_μ). Therefore the corresponding distances along the grid of neurons will shrink reflecting more accurately the object proximities.

The error function (6) can be optimized in two steps as in the symmetric case. First a quantization algorithm is run that generates the SOM prototypes $\vec{w}_s$. Next the function error is maximized with respect to the weights $\vec{w}_s$. This yields a simple adaptation rule for the network prototypes:

$$\vec{w}_s = \frac{\sum_{r=1}^{M} \sum_{x_\mu \in V_r} \omega_\mu h_{rs} \vec{x}_\mu}{\sum_{r=1}^{M} \sum_{x_\mu \in V_r} \omega_\mu h_{rs}} \tag{7}$$

where h_{rs} is a Gaussian kernel of parameter $\sigma(t)$ which is adapted using the same rules considered for the symmetric version. Notice that the the simplicity of the original SOM algorithm is maintained.

The SOM algorithm proposed earlier improves particularly the associations induced among objects of disparate L_1 norm (asymmetrically related). However, according to equation (5) the objects of large and similar L_1 norm get also closer in the grid of neurons. This behavior may eventually increase the overlapping among the main topics of the database which is an undesirable effect. To avoid this problem an alternative similarity is defined as follows:

$$s_{ij} = (K - \|\vec{x}_i - \vec{x}_j\|^2)[1 + (\omega_i - \omega_j)^2] \,, \tag{8}$$

where ω_i, ω_j denote the asymmetry coefficients defined in section 2. This similarity becomes larger than the Euclidean proximity measure just when ($\omega_i \neq \omega_j$). In this case, the similarity is compensated proportionally to the degree of asymmetry $|\omega_i - \omega_j|$. Substituting this similarity into equation (3) we get the error function to be optimized:

$$E(\mathcal{W}) = \sum_r \sum_{x_\mu \in V_r} \sum_s h_{rs} (K - \|\vec{x}_\mu - \vec{w}_s\|^2)[1 + (\mid \vec{x}_\mu \mid - \mid \vec{w}_s \mid)^2] \,. \tag{9}$$

The optimization of this error is quite complex, because the parameter $\mid \vec{w}_s \mid$ depends on the prototype coordinates. To overcome this problem, the dataset is first transformed to a feature space via the function $\phi : \mathbb{R}^p \to \mathcal{F}$ defined as follows: $\phi(\vec{x}) = (\vec{x}, |\vec{x}|)$. In this feature space, the optimization is easier because the error function (9) can be derived independently with respect to the prototype and the L_1 norm coordinates. Using this trick and solving the set of linear equations $\frac{\partial E(\mathcal{W})}{\partial |\vec{w}_s|} = 0$, we get the following updating rule for the prototype L_1 norm coordinate:

$$| \vec{w}_s | = \frac{\sum_r \sum_{x_\mu \in V_r} h_{rs} \alpha'_{\mu s} | \vec{x}_\mu |}{\sum_r \sum_{x_\mu \in V_r} h_{rs} \alpha'_{\mu s}} , \qquad (10)$$

where $\alpha'_{\mu s} = (K - \|\vec{x}_\mu - \vec{w}_s\|^2)$ and h_{rs} is a Gaussian kernel defined as usual. Similarly, solving the set of linear equations $\frac{\partial E(\mathcal{W})}{\partial \vec{w}_s} = 0$, we get the following updating rule for the network prototypes:

$$\vec{\omega}_s = \frac{\sum_r \sum_{x_\mu \in V_r} h_{rs} \vec{x}_\mu [1 + (| \vec{x}_\mu | - | \vec{w}_s |)^2]}{\sum_r \sum_{x_\mu \in V_r} h_{rs} [1 + (| \vec{x}_\mu | - | \vec{w}_s |)^2]} \qquad (11)$$

In the las years several authors (see for instance [8, 7]) have proposed new variants of the SOM algorithm that are able to work directly from a (symmetric) dissimilarity matrix. In previous papers [14, 13] we have reported that usual mapping algorithms that works from a dissimilarity matrix such as Sammon, can be extended to the asymmetric case using a procedure similar to the one employed here. This suggests that the algorithms mentioned above can be extended to the asymmetric case, just substituting in (4) the Euclidean distance by the corresponding dissimilarity matrix and considering that the coefficient of asymmetry can be defined easily for a given asymmetric measure [14]. However, notice that the SOM algorithms based exclusively on distances are usually more intensive computationally.

We finish this section with a brief comment about the related work.
As far as we know, no asymmetric version of the SOM has been proposed earlier in the literature. However, the multidimensional scaling (MDS) community have proposed several models to deal with asymmetric measures in the context of psychometric or sociometric data (see for instance [6, 21] and references). Those algorithms optimize a quadratic error measure of the form $\sum_{ij}(\delta_{ij} - d_{ij})^2$, where δ_{ij} and d_{ij} denote the asymmetric dissimilarities in input and output spaces. However, it has been pointed out in the literature [21, 14] that the optimization of this error function is equivalent to build two maps that approximate independently the symmetric and skew symmetric components of the dissimilarity matrix (δ_{ij}). Therefore the map that visualizes the object proximities is exclusively derived from the symmetric component of δ_{ij} and is degraded by asymmetry as well. Thus, the contribution of the work presented here is to improve the map that visualizes the object proximities taking advantage of the information conveyed by the asymmetry.

4 Experimental results

In this section we apply the proposed algorithms to the construction of word maps that visualize term semantic relations. First we describe briefly the textual collections used in the experiments.
The first collection, is made up of 2000 *scientific abstracts* retrieved from three commercial

databases 'LISA', 'INSPEC' and 'Sociological Abstracts'. A thesaurus created by human experts is available which allow us to exhaustively check the term associations created by the SOM algorithms. The second collection is made up of 6702 abstracts corresponding to the journals of the ACM digital library. In this case, no thesaurus is available for the collection and therefore the evaluation must rely on unsupervised measures. This a real and interesting problem not previously considered in the literature.

Assessing the performance of algorithms that generate word maps is not an easy task. In this paper the maps are evaluated from different viewpoints through several objective functions. This methodology guaranty the objectivity and validity of the experimental results.
The first measure considered is the Spearman rank correlation coefficient [3] (Sp.). This coefficient checks if the neighbor's ordering in input space is preserved in the map. A complementary measure is the Sp. coefficient taking into account only the 10% of the first nearest neighbors. Notice that the first nearest neighbors of specific terms are frequently broad terms [17, 13]. Therefore, this index provides more specific information about the preservation of dissimilarities corresponding to asymmetric relations.

The second group of measures quantifies the agreement between the semantic word classes induced by the map and the thesaurus. Therefore, once the objects have been mapped, they are grouped into topics with a clustering algorithm (for instance PAM [9]). Next the partition induced by the map is evaluated through the following objective measures:
The F measure [2] has been widely used by the Information Retrieval community and evaluates if words from the same class according to the thesaurus are clustered together. The entropy measure [17] evaluates the uncertainty for the classification of words from the same cluster. Small values suggest that the clusters are tight and so the overlapping among different topics in the map is smaller. Finally the Mutual Information [20] is a nonlinear correlation measure between the word classification induced by the thesaurus and the word classification given by the clustering algorithm. This measure gives more weight to specific words and therefore provides valuable information about changes in the position of specific terms.

	Scientific abstracts					ACM corpus	
	Sp.	Sp. 10%	F	E	I	Sp.	Sp. 10%
[1] *Symmetric SOM*	0.43	0.64	0.70	0.38	0.23	0.43	0.74
[2] Asymmetric SOM	0.57	0.76	0.78	0.35	0.27	0.51	0.76
Improvement in %	33	16	11	8	17	19	3
[3] Asymmetric SOM (L_1 norm difference)	0.37	0.78	0.74	0.31	0.22	0.48	0.79
Improvement in %	-14	22	6	18	-4	12	7

Parameters: Nneur = 88, niter = 30; $\sigma_i^1 = \sigma_i^2 = 30$, $\sigma_i^3 = 33$; $\sigma_f^1 = \sigma_f^2 = 3$, $\sigma_f^3 = 2$.

Table 1: Empirical evaluation of the asymmetric SOM algorithms for a collection of scientific abstracts and the journals of the ACM digital library.

Table 1 shows the experimental results for the two problems considered: The abstracts of scientific journals and the ACM digital library. The symmetric SOM algorithm (row 1) has been taken as reference because it has been widely applied in text mining problems (see for instance the WEBSOM [11]). The SOM topology has been chosen linear because the empirical evidence suggests that the network organization is better. The value of the parameters for each algorithm is specified in table 1. Term vectors have been codified using the vector space model [2] and normalized by the L_2 norm. The primary conclusions are the following:

The first asymmetric version of SOM proposed in section 3 (row 2) outperforms the symmetric counterpart. In particular the Mutual Information (I) is improved a 17% which suggests that the position of specific terms in the map is significantly better in the asymmetric model. This fact helps to avoid that the more specific terms (low norm) concentrate in some specific area of the map regardless of their semantic meaning (see [4] for a detailed analysis of this problem). Consequently the overlapping among terms belonging to different topics is reduced in the map ($\Delta E = 8\%$). Finally the overall word map quality (F) is a 10% better than in the symmetric version.
The unsupervised measures (Sp.) and (Sp. 10) show that the organization of the network is even better than for the classic algorithm. This suggests that it is easier to preserve the asymmetric similarity probably because the histogram is smoother.

The second asymmetric version of SOM (row 3) improves also the map generated by the symmetric counterpart. Notice that as it was suggested in section 3 the overlapping in the map is reduced more than in the previous algorithm ($\Delta E = 18\%$). However, the (Sp.) and I measures suggest that the network organization is more problematic particularly for terms of medium L_1 norm. Finally we point out that the overall word map quality (F) is improved a 6%.
The experimental results for the journals of the ACM digital library collection corroborate the superiority of the asymmetric algorithms proposed in this paper.

As a conclusion, the empirical evidence shows that incorporating asymmetry helps to improve significantly the term associations suggested by the SOM algorithm. However, notice that the asymmetry is not the only factor that distorts the distances in text mining problems (see for instance [1]).

Finally figure 1 shows a visual map generated by our asymmetric SOM for the first textual collection considered in this paper. For the sake of clarity only a small subset of terms that belong to two different topics have been drawn. The SOM prototypes have been projected using the Sammon algorithm [10] and those one corresponding to the neighboring neurons have been joined together by continuous trace. Terms with L_1 norm > 30 and ≤ 30 are visualized in different colors.
The figure 1 shows that the terms are spread along the map regardless of the frequency (L_1 norm). The term associations induced by the map are satisfactory even for words with disparate degree of generality (L_1 norm). Notice also that the network organization is satisfactory.

5 Conclusions and future research trends

In this paper we have proposed two asymmetric variants of the SOM algorithm that improve the visualization of the object proximities when relations are asymmetric. The algorithms have been tested in the challenging problem of word relation visualization using real problems such as the ACM digital library. The word maps have been exhaustively evaluated through several objective functions.

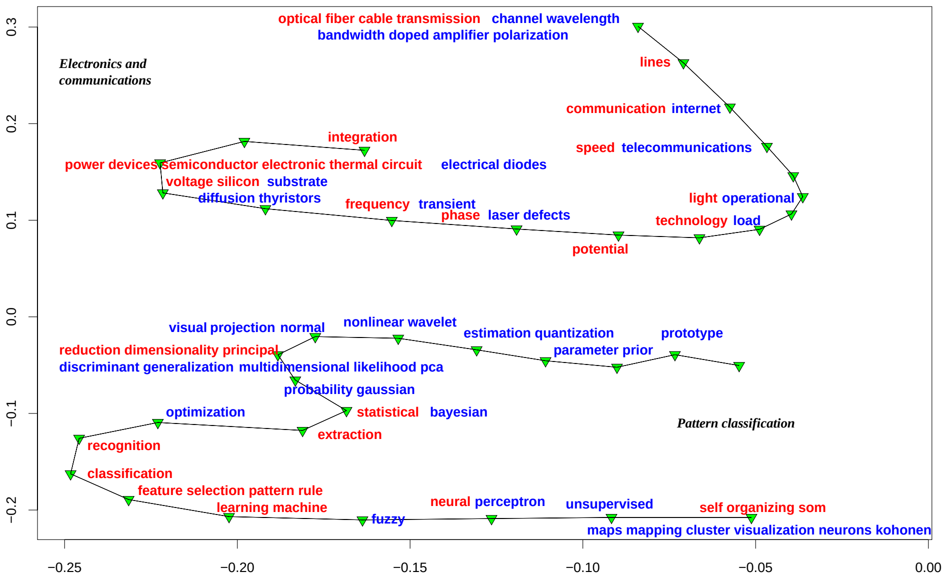

Figure 1: Word map generated by the asymmetric SOM for a collection of scientific abstracts.

The experimental results show that the asymmetric algorithms improve significantly the map generated by a SOM algorithm that relies solely on the use of a symmetric distance. In particular, our asymmetric models achieve a remarkable improvement of the position of specific terms in the map. Besides, the new models keep the simplicity of the original SOM algorithm.
Future research will focus on the development of new asymmetric techniques for classification purposes.

References

[1] C. C. Aggarwal and P. S. Yu. Redefining clustering for high-dimensional applications, IEEE Transactions on Knowledge and Data Engineering, 14(2):210-225, 2002.

[2] R. Baeza-Yates and B. Ribeiro-Neto. Modern information retrieval. Addison Wesley, Wokingham, UK, 1999.

[3] J. C. Bezdek and N. R. Pal. An index of topological preservation for feature extraction, Pattern Recognition, 28(3):381-391, 1995.

[4] A. Buja, B. Logan, F. Reeds and R. Shepp. Inequalities and positive default functions arising from a problem in multidimensional scaling, Annals of Statistics, 22, 406-438, 1994.

[5] Y. M. Chung and J. Y. Lee. A corpus-based approach to comparative evaluation of statistical term association measures, Journal of the American Society for Information Science and Technology, 52, 4, 283-296, 2001.

[6] T. F. Cox and M. A. A. Cox. Multidimensional scaling. Chapman & Hall/CRC, 2nd edition, USA, 2001.

[7] A. E. Golli, B. Conan-Guez and F. Rossi. A Self Organizing Map for Dissimilarity Data. In proceedings of IFCS 2004, pages 61-68, Chicago, U.S.A.. July 2004.

[8] T. Graepel and K. Obermayer. A stochastic self-organizing map for proximity data. *Neural Computation*, 11:139-155, 1999.

[9] L. Kaufman and P. J. Rousseeuw. Finding groups in data. An introduction to cluster analysis. John Wiley & Sons. New York. 1990.

[10] T. Kohonen. Self-organizing maps. Springer Verlag, Berlin, Second edition, 1995.

[11] T. Kohonen, S. Kaski, K. Lagus, J. Salojarvi, J. Honkela, V. Paatero and A. Saarela. Organization of a massive document collection, IEEE Transactions on Neural Networks, 11(3):574-585, 2000.

[12] L. Lebart, A. Morineau and J. F. Warwick. Multivariate descriptive statistical analysis. John Wiley, New York, 1984.

[13] M. Martín-Merino and A. Muñoz. Self organizing map and Sammon mapping for asymmetric proximities, ICANN, LNCS 2130, 429-435, Springer Verlag, 2001.

[14] M. Martín-Merino and A. Muñoz.Visualizing asymmetric proximities with SOM and MDS models, Neurocomputing, 63, 171-192, 2005.

[15] D. Merkl. Text classification with self-organizing maps. Some lessons learned. Neurocomputing, 21:61-77, 1998.

[16] F. Mulier and V. Cherkassky. Self-organization as an iterative kernel smoothing process. Neural Computation, 7, 1165-1177, 1995.

[17] A. Muñoz. Compound key word generation from document databases using a hierarchical clustering ART model. Journal of Intelligent Data Analysis, 1(1), 25-48, 1997.

[18] E. Oja and S. Kaski. Kohonen Maps. Elsevier, Amsterdam, 1999.

[19] M. Rorvig. Images of similarity: A visual exploration of optimal similarity metrics and scaling properties of TREC topic-document sets, Journal of the American Society for Information Science, 50, 8, 639-651, 1999.

[20] Y. Yang and J. O. Pedersen. A comparative study on feature selection in text categorization, Proc. of the 14th International Conference on Machine Learning, Nashville, Tennessee, USA, July, 412-420, 1997.

[21] B. Zielman and W.J. Heiser. Models for asymmetric proximities, British Journal of Mathematical and Statistical Psychology, 49:127-146, 1996.