On Using Asymmetry Information for Classification in Extended Dissimilarity Spaces

Yenisel Plasencia-Calaña[1,2], Edel B. García-Reyes[1], Robert P.W. Duin[2], and Mauricio Orozco-Alzate[3]

[1] Advanced Technologies Application Center,
7ma A ♯ 21406, Playa, Havana - 12200, Cuba
`{yplasencia,egarcia}@cenatav.co.cu`
[2] Faculty of Electrical Engineering, Mathematics and Computer Sciences,
Delft University of Technology, The Netherlands
`r.duin@ieee.org`
[3] Departamento de Informática y Computación, Universidad Nacional de Colombia - Sede Manizales. Kilómetro 7 vía al Aeropuerto, Campus La Nubia – Bloque Q, Piso 2, Manizales, Colombia
`morozcoa@unal.edu.co`

Abstract. When asymmetric dissimilarity measures arise, asymmetry correction methods such as averaging are used in order to make the matrix symmetric. This is usually needed for the application of pattern recognition procedures, but in this way the asymmetry information is lost. In this paper we present a new approach to make use of the asymmetry information in dissimilarity spaces. We show that taking into account the asymmetry information improves classification accuracy when a small number of prototypes is used to create an extended asymmetric dissimilarity space. If the degree of asymmetry is higher, improvements in classification accuracy are also higher. The symmetrization by averaging also works well in general, but decreases performance for highly asymmetric data.

Keywords: asymmetric dissimilarity, dissimilarity spaces, prototype selection, extended dissimilarity spaces.

1 Introduction

Dissimilarity representations [1] arose as an alternative to feature-based representations when the definition and extraction of good features is difficult or intractable while a robust dissimilarity measure can be defined more easily for the problem at hand. Research in this field has focused on several topics: prototype selection [2,3] or generation [4], classification in dissimilarity spaces [5,6], among others. One open issue corresponds to the information usage in dissimilarity matrices: they can be asymmetric but most of the traditional classification and clustering methods are thought for symmetric dissimilarity matrices. In case

of asymmetry, the typical approach is to symmetrize the matrix with any known symmetrization method, and then apply the methods on the symmetric variant. This might carry a loss of useful information.

Asymmetric dissimilarity or similarity measures can arise in several situations; see [7] for a general analysis of the causes of non-Euclidean data. Measures resulting from a matching process may appear to be asymmetric due to a suboptimal procedure. Also, measures designed using expert knowledge for the problem might not be symmetric. One example is fingerprint matching [3], where measures are often asymmetric. When various dissimilarity matrices are combined, the final matrix can also be asymmetric. One of the most widely used methods for symmetrization is the average method. In [1], before embedding asymmetric dissimilarity matrices into Pseudo-Euclidean spaces, the average method is used to make the matrix symmetric. In [2], the dissimilarity matrix is symmetrized using the average method in order to allow the use of some prototype selection algorithms in the dissimilarity space (DS). Other authors, in the context of kernel-based classification, proposed the use of a positive semidefinite matrix $K^t K$, where K denotes a nonsymmetric kernel [8].

Different variants of the Multidimensional Scaling algorithm have incorporated asymmetry in an intuitive way, by defining a skew symmetric term [9]. In [10], the authors proposed modifications to Self Organizing Map and Sammon Mapping in order to deal with asymmetric proximities showing that the proposed algorithms outperformed their symmetric variants. In [11], the authors compared several symmetrization methods of asymmetric kernel matrices for their use in the context of Support Vector Machines. They also proposed a simple supervised symmetrization method that outperformed the other methods compared.

One question that arises is whether the asymmetry information can be useful for classification in dissimilarity spaces, instead of ignoring it or using a symmetrization method. Another question is how we can use the asymmetry information in the context of classification in dissimilarity spaces. In this paper we propose a new approach for using asymmetry information in what we called the extended asymmetric dissimilarity space (EADS). As the dimension of the EADS space is twice the dimension of the original DS, the use of prototype selection is needed in order to reduce the dimensions before the EADS is constructed. Results are provided comparing classification errors in both the DS and EADS for four standard asymmetric dissimilarity data sets.

2 Dissimilarity Space and Extended Dissimilarity Space

Dissimilarity representations arose from the idea that the classes are constituted by similar objects, so the nearness information is more fundamental than features to discriminate between the classes [1]. In this context, the DS was proposed in [1] as follows. Let $R = \{r_1, r_2, ..., r_k\}$ be the representation set: a collection of prototypes that may be a subset of the training set T. Let d be a dissimilarity measure for the problem at hand. The DS is created by a mapping of the objects to the space defined by the dissimilarities to the prototypes, where each

dimension corresponds to the dissimilarities to a given prototype. The representation d_x of an object x is:

$$d_x = [d(x, r_1) \ d(x, r_2) \ ... \ d(x, r_k)]. \tag{1}$$

The DS was postulated as a Euclidean vector space, making suitable the use of traditional classifiers for feature spaces like Bayesian ones. The cardinality of the representation set defines the dimension of the DS. For the reduction of the representation set, prototype selection methods are used. They allow one to determine the desired tradeoff between classification accuracy and representation cardinality.

In this section we present the EADS. The motivation for this proposal is that when projecting asymmetric data in the DS, asymmetry information is lost because we only use dissimilarities from the objects to the prototypes, and not from prototypes to objects. If the matrix is previously symmetrized, we are also neglecting the asymmetry present in the data. In order to take advantage of the asymmetry information in both directions, we explore the use of an extended representation of the initial asymmetric dissimilarity matrix in an EADS. We propose to create the EADS using the prototypes selected from the original dissimilarity matrix as it is given. Then, having those prototypes $R = \{r_1, r_2, ..., r_k\}$, the representation of an object in the EADS is defined by:

$$d_x = [d(x, r_1) \ d(x, r_2) \ ... \ d(x, r_k) \ d(r_1, x) \ d(r_2, x) \ ... \ d(r_k, x)]. \tag{2}$$

In order to represent the training set and the objects submitted for classification in the EADS, we need to measure the dissimilarities from the objects to the prototypes and from the prototypes to the objects. As a result, the dimension of the EADS space is twice the dimension of the DS. Classifiers can be trained in the EADS in the same way they are trained in the DS.

3   Experiments

In this section the discriminative power of the EADS is compared to the discriminative power of the non-symmetrized version and the one symmetrized by averaging. Classification errors are presented using different numbers of prototypes in DS and EADS. Prototypes are the same for both spaces, but in the DS only dissimilarities in one direction are used. In the EADS, dissimilarities from the two directions are used. This leads to a space of dimension twice the size of the DS dimension.

3.1   Data Sets and Experimental Setup

For the experiments we used four data sets: Chickenpieces-20-60, Chickenpieces-35-45, CoilYork, and Zongker. Some important characteristics of the data sets

can be found in Table 1. The Asymmetry column shows an asymmetry coefficient ac for each data set, this was computed using the following equation:

$$ac = \frac{1}{n(n-1)} \frac{\sum |d_{ij} - d_{ji}|}{\min(d_{ij}, d_{ji})}, i = 1...n-1, j = i+1...n, \qquad (3)$$

where n is the number of objects in the data set. We assume that dissimilarities between different objects will not be zero. In case it is known beforehand that in the data it may exist any dissimilarity with value zero between two different objects, a term with a very small value such as 0.0001 must be added in the denominator to avoid the indefinite result of the division by zero.

Table 1. Characteristics of the data sets, the $|X|$ column is the number of training objects, and $|T|$ is the number of test objects

| Data sets	# Classes	# Obj. per class	Asymmetry	$|X|$	$|T|$
ChickenPieces-20-60	5	117,76,96,61,96	0.05	222	224
ChickenPieces-35-45	5	117,76,96,61,96	0.08	222	224
CoilYork	4	4x72	0.009	144	144
Zongker	10	10x200	0.18	400	1600

As classifier it was used the Linear Bayes Normal (BayesL) in both the DS and EADS, it is a simple and fast classifier that is optimal for normally distributed classes with equal covariances. Experiments were repeated twenty times using equal-sized random partitions for training and testing for ChickenPieces and CoilYork data sets, and twenty and eighty percent for training and testing respectively in the Zongker data set. Results were averaged over the twenty experiments. As prototype selectors, two different methods are used: the systematic forward selection optimizing the leave-one-out nearest neighbour error on the training set as in [2] (FS+NN error), and the random selection. The methods selected 5, 10, 15, 20 and 25 prototypes. The BayesL and prototype selectors were trained using the training data, and the classification results were computed in the test set for the DS and EADS generated using the prototypes selected with the different methods. Regularization parameter of BayesL is 0.01.

3.2   Results and Discussion

Figure 1 shows the curves of error rates for increasing number of prototypes in the original asymmetric representation in the DS and the representation in the EADS. Figure 2 shows the curves of error rates for an increasing number of prototypes comparing the symmetrized representation in the DS using the average and the representation in the EADS. Solid lines represent the approaches in EADS; dashed lines represent the approaches in DS. The same symbol is assigned for the results in DS and EADS using the same prototype selector. Standard deviations are between 0.007 and 0.08.
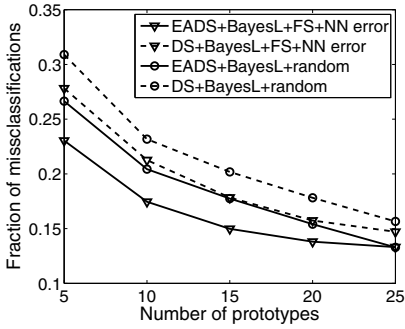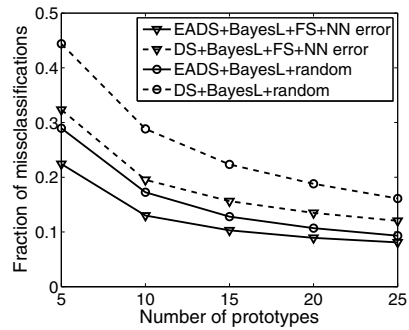
(a) ChickenPieces-20-60 data set
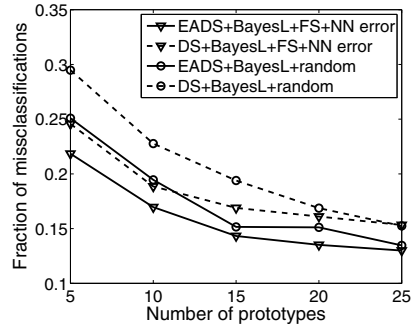
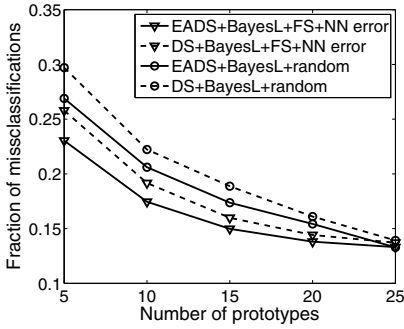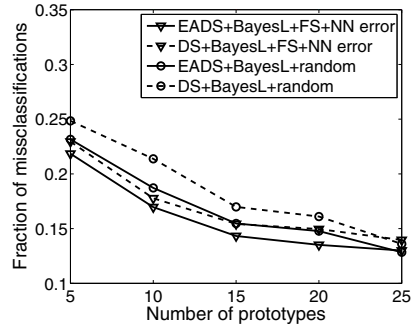(b) ChickenPieces-35-45 data set
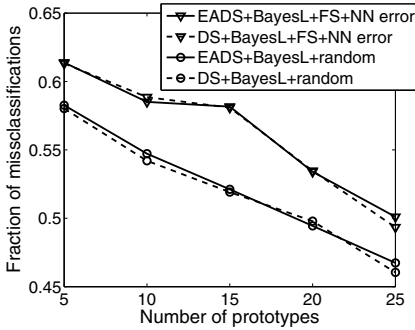
(c) CoilYork data set

(d) Zongker data set

Fig. 1. Classification results for the original asymmetric representation in the DS and the EADS in the data sets, the dimension of the associated DS is equal to the number of prototypes, and the dimension of the associated EADS is twice the number of prototypes

From the results in Fig. 1 we can see that in three of the four data sets —the ChickenPieces-20-60, ChickenPieces-35-45, and Zongker— classification in EADS outperforms classification in DS using both the systematic and the random prototype selectors. These are the data sets with the higher degree of asymmetry as measured by the asymmetry coefficient. In the CoilYork data set, which has the smallest asymmetry degree, the results in the EADS were a little worse than those in the DS. Except for the CoilYork data set, when the number of prototypes increases, the difference between the error rates in EADS and DS decreases. This implies that the asymmetry information is more useful if small sets of prototypes are used, and having more dimensions compensates for not using asymmetry information.
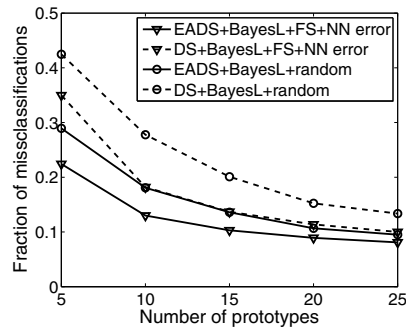
(a) ChickenPieces-20-60 data set

(b) ChickenPieces-35-45 data set
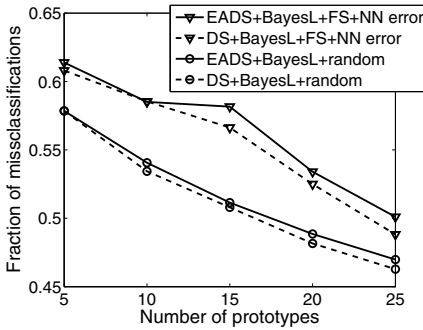
(c) CoilYork data set

(d) Zongker data set

Fig. 2. Classification results for the representation in the DS symmetrized by averaging and the EADS in the data sets, the dimension of the associated DS is equal to the number of prototypes, and the dimension of the associated EADS is twice the number of prototypes

From Fig. 2 we can see that once the dissimilarities are symmetrized by averaging, incorporating the asymmetry information does not improve classification in the same extent as by using the non-symmetrized version. This shows that the symmetrization by averaging is a good alternative for dealing with asymmetric data. In the CoilYork data set, the EADS performed worse than the DS using the symmetrized dissimilarities. In this case, where the asymmetry coefficient has a very small value, the use of asymmetry information leads to a slight decrease in classification performance. The symmetrization by averaging becomes less useful when the asymmetry degree of the data increases as it can be deduced from the similar classification errors in the original DS (see Fig. 1, (d)) and the DS symmetrized by averaging (see Fig. 2, (d)) in the Zongker data.

From the results, we made a characterization of the relationship between the amount of asymmetry present in each data set measured by the asymmetry coefficient and the improvements obtained in classification in the EADS compared to the non-symmetrized DS. First, we sorted the asymmetry coefficients of each data set in increasing order, and plotted the classification improvements in EADS compared to DS measured by the differences between the curves for the same prototype selection method in both spaces. The sum of these differences was plotted for each data set with its related asymmetry coefficient, see Fig. 3.
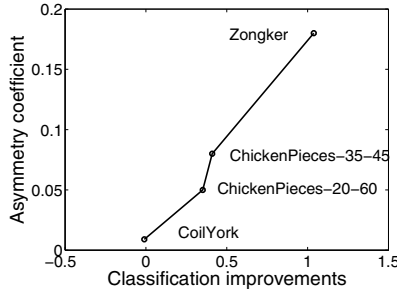

Fig. 3. Classification improvements in EADS compared to DS as a function of the asymmetry coefficient

In the function we can see a positive linear correlation between the two variables, as the value of the asymmetry coefficient increases, the value of the improvements in classification also increases. The value of the correlation coefficient was 0.99. This means that it is important to take the asymmetry information into account in order to improve classification rates when the asymmetry degree is perceivable, and while the data is more asymmetric the classification improvement increases. In the CoilYork dataset we obtained a negative value of improvement equal to -0.01, since the EADS performed slightly worse than the DS.

4   Conclusions

We proposed the EADS that proved to be suitable for exploiting the asymmetry information from the dissimilarities. This is especially useful for small prototype sets. For a data set with a very small degree of asymmetry, it might not be necessary and can even be slightly detrimental to use asymmetry information. Another conclusion is that the symmetrization by averaging is a good alternative for dealing with asymmetric data, although it becomes less useful when the asymmetry degree of the data increases. In our results, the improvements achieved in classification in EADS are positively correlated to the degree of asymmetry in each data set. The use of EADS can be beneficial when one has a very small set of informative prototypes with a highly asymmetric data set. The symmetrization operation may depend on the cause of asymmetry, e.g. averaging can be

good for asymmetry caused by noise, the minimum can be useful for asymmetry caused by a shortest path optimization to compute the dissimilarities. Further work will be devoted to study these operations and the EADS for asymmetry caused by expert knowledge incorporated in the dissimilarity measure, noise or suboptimal procedures.

References

1. Pekalska, E., Duin, R.P.W.: The Dissimilarity Representation for Pattern Recognition: Foundations and Applications (Machine Perception and Artificial Intelligence). World Scientific Publishing Co., Inc., River Edge (2005)
2. Pekalska, E., Duin, R.P.W., Paclík, P.: Prototype selection for dissimilarity-based classifiers. Pattern Recogn. 39(2), 189–208 (2006)
3. Bunke, H., Riesen, K.: Graph Classification Based on Dissimilarity Space Embedding. In: da Vitoria Lobo, N., Kasparis, T., Roli, F., Kwok, J.T., Georgiopoulos, M., Anagnostopoulos, G.C., Loog, M. (eds.) S+SSPR 2008. LNCS, vol. 5342, pp. 996–1007. Springer, Heidelberg (2008)
4. Orozco-Alzate, M., Duin, R.P.W., Castellanos-Domínguez, G.: A generalization of dissimilarity representations using feature lines and feature planes. Pattern Recogn. Lett. 30(3), 242–254 (2009)
5. Pekalska, E., Duin, R.P.W.: Beyond traditional kernels: Classification in two dissimilarity-based representation spaces. IEEE Trans. Syst. Man Cybern. C, Appl. Rev. 38(6), 729–744 (2008)
6. Pekalska, E., Paclik, P., Duin, R.P.W.: A generalized kernel approach to dissimilarity-based classification. J. Mach. Learn. Res. 2, 175–211 (2002)
7. Duin, R.P.W., Pękalska, E.: Non-Euclidean Dissimilarities: Causes and Informativeness. In: Hancock, E.R., Wilson, R.C., Windeatt, T., Ulusoy, I., Escolano, F. (eds.) SSPR & SPR 2010. LNCS, vol. 6218, pp. 324–333. Springer, Heidelberg (2010)
8. Schölkopf, B., Mika, S., Burges, C.J.C., Knirsch, P., Müller, K.R., Rätsch, G., Smola, A.J.: Input space versus feature space in kernel-based methods. IEEE Trans. Neural Netw. 10(5), 1000–1017 (1999)
9. Okada, A., Imaizumi, T.: Nonmetric multidimensional scaling of asymmetric proximities. Behaviormetrika 14(21), 81–96 (1987)
10. Martín-Merino, M., Muñoz, A.: Self Organizing Map and Sammon Mapping for Asymmetric Proximities. In: Dorffner, G., Bischof, H., Hornik, K. (eds.) ICANN 2001. LNCS, vol. 2130, pp. 429–435. Springer, Heidelberg (2001)
11. Muñoz, A., de Diego, I.M., 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)