

Dissimilarity-based Classification of Data with Missing Attributes

M. Millán-Giraldo*, Robert P.W. Duin[†] and J.S. Sánchez*

Emails: mmillan@uji.es, r.p.w.duin@tudelft.nl and sanchez@uji.es

*Institute of New Imaging Technologies

Dept. de Llenguajes y Sistemes Informàtics, Universitat Jaume I

Av. Sos Baynat s/n, 12071 Castellón de la Plana, Spain

[†]Pattern Recognition Lab

Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology

P.O. Box 5031, 2600GA Delft, The Netherlands

Abstract—In many real world data applications, objects may have missing attributes. Conventional techniques used to classify this kind of data are represented in a feature space. However, usually they need imputation methods and/or changing the classifiers. In this paper, we propose two classification alternatives based on dissimilarities. These techniques promise to be appealing for solving the problem of classification of data with missing attributes. Results obtained with the two approaches outperform the results of the techniques based in the feature space. Besides, the proposed approaches have the advantage that they hardly require additional computations like imputations or classifier updating.

I. INTRODUCTION

Data classification is one of the main tasks in pattern recognition. It aims to predict a label for every object according to the class that it belongs to, using a classification model that has been built from a training set. Usually, objects in classification systems are represented by features, also called *attributes*. In many predictive modeling applications for some objects not contain all attribute values may be given. There are several reasons why this can happen, such as delays or failures in communication lines, noise, measurement costs, or simply because they do not exist. In biomedical applications e.g., it is very common that medical records have some lacking values, or in telemedicine data may arrive incomplete. In urgent cases, there might be no time to perform other tests.

Generally, the lack of attributes generates statistical variations in the data and consequently deterioration in the classification model and, as a consequence, classification accuracy is reduced. Furthermore, incomplete objects hamper data handling and analysis. Bias may result from differences between missing and complete data. Most techniques, proposed in the literature, to classify data with missing attributes employ the representation in the feature space. Three methods are commonly used in this space to handle objects with missing attributes:

- 1) *Skipping incomplete objects*: it simply discards the incomplete objects in the dataset in order to create a new complete dataset. It is application dependent whether this is an option.

- 2) *Imputation*: this is probably the most frequently used approach. It estimates a value from the entire dataset to fill the missing attribute. Most common imputation techniques are mean, median, random and Hot deck [11].
- 3) *Projection*: in this case the space is reduced to one dimension less for each missing attribute. This requires a special computation of the classifier in the reduced space.

Recently, Zhang [12] proposed a partial imputation technique. It consists of the imputation of missing data using complete objects in a small neighborhood of the incomplete ones. Delavallade and Ha [13] proposed a new approach, using the entropy to estimate the missing values. Their results showed that this approach can outperform other traditional imputation techniques, as mean and mode. Farhangfar et al. [2] studied the influence of the imputation of missing attributes on the classification error for five imputation methods: mean, the Hot deck method, Naïve Bayes, multiple imputation and framework method using Naïve Bayes and Hot deck. Their analysis shows that, in general, imputation is beneficial for the classification of objects with missing attributes.

Generally, methods for incomplete data classification based on features require additional computations like imputation or classifier updating. Consequently, these methods tend to be computationally expensive.

This paper focuses on investigating the use of the dissimilarity representation [7] in the classification of incomplete data. For this purpose, we propose two simple approaches for which the object representation is based on dissimilarity values. The first one uses a classifier trained with complete objects in the dissimilarity space and applies this classifier to objects with missing data using the possibility to compute object dissimilarities even if objects are incompletely given. The second approach is similar, but uses a correction factor that is applied to the dissimilarities of incomplete objects. The aim of this factor is to compensate the lack of the missing attributes. These two techniques require much less additional computations than those based on a feature space, because they do not need imputation or recomputed classifiers.

The paper is organized as follows. Section II provides an introduction to classification in the dissimilarity space, a de-

scription of the proposed approaches and also the description of an imputation technique in the feature space that is used in our analysis. Section III explains details of the experimental study, a description of the employed datasets, the experimental setup conducted and it also includes the experimental results and their respective analysis. Finally, Section IV presents a summary of the conclusions and discusses future works.

II. CLASSIFICATION IN DISSIMILARITY SPACE

A. Introduction to dissimilarity-based classification

In the dissimilarity space (DS), objects are represented by pairwise dissimilarities with other objects instead of attributes as in the feature space (FS). Hereby, every object is represented by a vector of dissimilarities to other objects [6].

In dissimilarity-based classification, the training set T of n objects, $T = \{\vec{x}_1, \dots, \vec{x}_n\}$, and the representation set R of r objects, $R = \{\vec{p}_1, \dots, \vec{p}_r\}$, are employed to build the classifier W in the dissimilarity space. R is a set of r prototypes that contains all classes, usually $R \subseteq T$. Various methods have been proposed in the literature to select the prototypes, e.g. in [5], random and systematic selection procedures were studied for the normal density-based quadratic classifier.

Given a dissimilarity measure d , the proximity between the object $\vec{x}_i \in T$ and the prototypes of R is $D(\vec{x}_i, R) = \{d(\vec{x}_i, \vec{p}_1), \dots, d(\vec{x}_i, \vec{p}_r)\}$, which is a vector with r distances that associates \vec{x}_i with all objects in the representation set R . Therefore, the proximity $D(T, R)$ is a dissimilarity matrix of size $n \times r$, which refers objects in the training set to all objects in the representation set. This matrix is interpreted as a set of row vectors that is used to build the classifier in the dissimilarity space of dimension r .

Given a test set S with s objects, its representation in the dissimilarity space is obtained by calculating the distances between its objects and prototypes in R , $D(S, R)$, which is a $s \times r$ matrix.

The dissimilarity measure is small when the objects x_i and p_h (where $i < n$, and $h < r$) are similar. It should be larger when the objects are more different. The distance $d(\vec{x}_i, \vec{p}_h) = 0$ when \vec{x}_i and \vec{p}_h are identical. In this paper the dissimilarity representation is derived from the feature space. We will use Euclidean distances. So objects with a zero distance in the feature space will also have a zero distance in the dissimilarity space as all their dissimilarities in the feature space are equal.

Classifiers computed in a dissimilarity space may perform well in comparison with those computed in a feature space [5]. This is partially caused by the non-linear relation between these spaces. When a linear classifier in the dissimilarity space reaches a certain performance, a more complex non-linear classifier in the feature space is required to achieve the same performance. Therefore, a simple linear classification model in the dissimilarity space could more easily separate the classes than the same model in the feature space.

B. Dissimilarity approaches for missing attributes

We propose two simple alternatives to classify incomplete data using the dissimilarity representation.

- *Using the original classifier:* the classifier W is built using the dissimilarity matrix $D(T, R)$, where T and R are the training set and the representation set, respectively, both with complete objects. This technique employs a classifier in the dissimilarity space to classify objects with missing attributes, regardless which attribute is missing. Assume \vec{x}_i is an object from test set S to be classified. Let the attribute j be missing, that is $\vec{x}_i^j = \{x_{i1}, \dots, x_{i(j-1)}, x_{i(j+1)}, \dots, x_{ik}\}_{1 \times (k-1)}$, where k is the number of features. To classify \vec{x}_i^j in the dissimilarity space, distances to the objects in the representation set should be computed. For this purpose, the attribute j of objects in the representation set R is ignored, that is, R is projected to one dimension less, R^j . Afterward, distances between \vec{x}_i^j and R^j are calculated by $D(\vec{x}_i^j, R^j) = \{d(\vec{x}_i^j, \vec{p}_1^j), d(\vec{x}_i^j, \vec{p}_2^j), \dots, d(\vec{x}_i^j, \vec{p}_r^j)\}$ and this vector is classified by the original classifier.

- *Applying correction:* like the previous approach, this technique also employs the trained classifier W with $D(T, R)$ to classify objects independent of which attribute is missing. The difference between them is that this approach uses a correction factor α to compensate the lacking contribution of a missing attribute j . This factor is applied to the distances of \vec{x}_i^j to create the new compensated object in the dissimilarity space \vec{x}_i^j as follows:

$$\vec{x}_i^j = \alpha \cdot D(\vec{x}_i^j, R^j)$$

where the value for α is obtained from the training set. Let j be the missing attribute of \vec{x}_i^j , T^j and R^j the training and representation sets, respectively, where the attribute j for all their objects is ignored. Therefore, the dissimilarity matrix with missing attributes is $D(T^j, R^j)$. Then we estimate α_j by optimizing the following criterion:

$$\alpha_j = \arg \min_{\alpha} \sum_{i=1}^n (D(\vec{x}_i, R) - \alpha D(\vec{x}_i^j, R^j))^2 \quad (1)$$

Equation (1) is optimized by:

$$\alpha_j = \frac{\sum_{i=1}^n D(\vec{x}_i^j, R^j) D(\vec{x}_i, R)}{\sum_{i=1}^n D(\vec{x}_i^j, R^j)^2} \quad (2)$$

Herewith an optimal correction factor α_j is obtained for every possible missing attribute j .

C. Imputation using support vector regression

Support vector machines (SVMs) [10] are learning models frequently used for classification. They usually represent the data in a space with a higher dimension than the original feature space. To this end, objects are mapped by a kernel function to this space. Possible kernel functions are the sigmoid, polynomials and radial basis functions. The purpose of the SVM is to find the optimal hyperplane with the maximum distance from the closest training objects. These closest points are called support vectors and they lie on the margin hyperplanes in feature space. In SVR, the regression

function is obtained using the support vectors and a non-linear error function.

Let $(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n) \in \mathbb{R}^n$ be the vectors of the training set, with two classes ($y_i \in \{-1, 1\}$). Assume the linear function of the hyperplane that separates the two classes is described by:

$$f(\vec{x}, \vec{w}) = \langle \vec{w}, \vec{x} \rangle + b, \text{ with } \vec{w} \in \mathbb{R}^n \text{ and } b \in \mathbb{R}$$

where \vec{w} is the perpendicular normal vector to the hyperplane and $\frac{b}{\|\vec{w}\|}$ is the hyperplane offset to the origin. Consider an ε -insensitive error function E_ε as the non-linear function, where the error increases linearly with respect to distance to the insensitive area, that is given by:

$$E_\varepsilon : \begin{cases} 0 & |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases}$$

where $|\xi|$ is the absolute difference between the predicted output and measured value. The optimal regression function is described by [14]:

$$C \sum_{i=1}^n (\xi_i + \hat{\xi}_i) + \frac{1}{2} \|\vec{w}\|^2,$$

where C is a regularization parameter, $\xi_i \geq 0$ and $\hat{\xi}_i \geq 0$ are the slack variables representing upper and lower limits, respectively, of the insensitive region.

To employ the imputation technique with SVR, assume the object \vec{x}_i does not have the attribute j , that is, \vec{x}_i^j . Complete objects of the training set are used to estimate the missing attribute j . Consider y_i is the class of \vec{x}_i^j . If the object \vec{x}_i^j is unlabeled, y_i may be estimated using any classification method such as *Projection* (Section I).

- For the complete training objects and \vec{x}_i^j , the attribute j is hold as output attribute (label), and original labels are considered as input attributes.
- Complete objects of the training set are used to build the optimal hyperplane.
- The object \vec{x}_i^j is mapped to that new feature space, in order to predict its output attribute (missing attribute).

In [8], Honghai et al. studied the imputation technique using SVR and compared it with the mean, median, the mean of the two closest neighbor values and the value of the nearest neighbor techniques. Their results showed that the SVR technique obtained the highest precision with regards to the others methods.

III. EXPERIMENTS AND RESULTS

The main goal of our experiments is to evaluate the proposed approaches in order to see the behavior of the classification error when a dissimilarity-based representation is applied to classification of missing data. Besides, another aim of the experiments is to compare our techniques with other methods in features space, and to observe whether they improve conventional approaches or not. For this purpose, we described the employed datasets in the experiments and the experimental setup. Afterwards, we present the results and a discussion.

A. Datasets

Experiments are carried out over twelve real datasets, from Ripley [3], Library¹ and UCI [1] repositories, each one with two classes. Table I shows a summary of datasets, which are organized according to the number of objects. They are normalized by the rank $[0, +1]$ with the aim of avoiding the influence of the different scales of feature values over results. All features have numerical values. To analyze the effect of the relevance of missing attributes in the classification error, the least and most relevant attributes for each dataset are considered as missing attributes. Attribute relevance is estimated using the Jeffreys-Matusita distance [4].

TABLE I
CHARACTERISTICS OF THE REAL DATA SETS USED IN THE EXPERIMENTS

Data set	Features	Objects	Source
crabs	6	200	Ripley
sonar	60	208	UCI
laryngeal1	16	213	Library
breast	9	277	UCI
intubation	17	302	Library
liver	6	345	UCI
spect	44	349	Library
wbc	30	569	UCI
australian	42	690	UCI
laryngeal2	16	692	Library
pima	8	768	UCI
german	24	1000	UCI

B. Experimental setup

The experiments are organized as follows:

- New objects are generated from the original datasets, due to the small size of these datasets and with the aim to consider 5000 objects for the test set. The objects are created introducing random noise, in one or more random features of the objects from the original dataset. The probability of introducing noise to each attribute of each new object was 0.1. The amount of noise added to each attribute is randomly chosen uniformly in $[0, 0.5]$ taking into account that data is already normalized and guaranteeing that the values are not out of the range $[0, 1]$ to keep them normalized.
- For each dataset, experiments are repeated five times. Each time, different training and test sets are chosen from the dataset in a random way to estimate a general behaviour of the datasets.
- The training set size is $k \times c$, where k is the number of features and c the number of classes, as all datasets here analyzed have two classes, the training set size is $2k$.
- The training set is built with objects from the original dataset, whereas, test set is built with remaining original objects and with the new generated objects, for a total of 5000 test objects.
- The representation set is chosen randomly from the training set. Its size was half of training set size, that is k , in order to employ a small representation set.

¹Library: http://www.informatics.bangor.ac.uk/~kuncheva/activities/real_data_full_set.htm

- The proximity measure used in the calculation of dissimilarity matrices is the Euclidean distance.
- Three classifiers are employed in the experiments: Fisher linear discriminant, 1-Nearest Neighbor rule (1NN) and Naïve Bayes classifier (NBC), for Naïve Bayes classifier each axis is split in bins and the objects classification is done according to the class with maximum posterior probability.
- The classification errors obtained over the five runs are averaged.

The generation of new synthetic data for the test set could lead to changes in the class distribution and therefore, can confuse the true performance of classifiers. However, although this is not the purpose of this study, this phenomenon could simulate, in certain way, the concept change occurring in many real datasets and which could be extended to future research.

C. Analysis of the Results

The objectives of this paper are to illustrate the potentials of dissimilarity-based representation for classifying objects suffering from missing attributes, and to analyze the possible advantages over procedures based on features. For this purpose, we compared the proposed approaches with two procedures used for feature representations. One of them employs the *Projection* method (Section I). The other one performs the *Imputation* with SVR (Section II-C). We compared with these procedures as they are frequently employed in the literature and besides, the imputation has shown its powerful ability of resolving problems of missing data classification [15] [8] [16]. Moreover, we also took the obtained results by applying dissimilarity and feature representations for complete objects, as the baselines.

We employed a method of ranking [9] in order to evaluate the performance of the proposed approaches and compared them with the already mentioned ones. To apply this method, we consider the classification error averaged over the 5 runs. To this end, we organize the results obtained by each procedure, classifier and dataset, from the smallest to highest error, ranking them by 1, 2, 3, and so on. Then, the sums of ranks (Borda counts) are used instead of the error values, and the ranks over all datasets, procedures and classifiers are averaged.

Tables II and III show the values obtained by applying the ranking method, by selecting as the missing attribute the least (respectively, the most) relevant attribute. In these tables, DS and FS stand for dissimilarity and feature space, respectively. Rows correspond to classifiers and columns to procedures. Lower values correspond to better overall performances for procedure-classifier combinations. Values in bold indicate the best result for each table. Underlined results are the best ones considering only procedures for missing attributes.

Analyzing the classifiers separately:

- **Fisher:** This classifier showed better results in the dissimilarity space. The significant difference between the obtained results in both spaces is possibly related to the fact that a linear classifier in the dissimilarity space corresponds to a non-linear classifier in the feature space. When the missing attribute is the least relevant one, it is

worth to apply the compensation of factor α . However, when this is the most relevant one, the effect of compensation is not significant; this may be due to the fact that the optimization procedure for α is unsupervised.

- **1NN:** When the missing attribute is the least relevant one, the nearest neighbor classifier works better with the *Projection* technique in the feature space than with the other methods. So it seems, for this classifier is probably better to ignore the attribute than take it into account, to impute it in the feature space or classify it in the dissimilarity space. This is possibly related to overtraining. In case the most relevant attribute is missing, as expected, none of the methods used for handling missing attributes improve the result obtained with the complete set in the feature space. In general, the classifier showed better results in the feature space than in the dissimilarity space.
- **NBC:** As with 1NN rule, this classifier also works better in the feature space. When the missing attribute is the least relevant one, it is better to apply the *Projection* procedure in the feature space than other procedures for missing attributes. For the most relevant attribute, it is better to impute it using SVR than to discard it with the *Projection* technique. However, the missing data classification using NBC is not better than for complete objects.

We expected that results obtained by using the correction factor α would outperform the results when it is not used, as occurred with the 1NN classifier, since it is applied in order to compensate the lack of the attribute. However, for Fisher classifier this correction is useful when the missing attribute is the least relevant one, and for NBC when it is the most relevant one. To estimate whether the chosen α value was the most suitable one for our approach, we varied its value in the range $[0.5, 3]$ and we observed the error behaviour. Figure 1 shows the classification error in varying α for the intubation dataset, for missing the least (left) and most (right) relevant attributes. When α is slightly over 1, the error is minimum. The straight line indicates α calculated for the Fisher classifier in the experiments using (2). For this dataset, the Fisher classifier obtained the smallest error when the missing attribute was the least relevant one. However, when it was the most relevant one, the use of α was more favorable for NBC. In most datasets, we obtained suitable α values. In order to see the improvement by applying the correction for the Fisher classifier, we plotted the relation between the error using the correction (y-axes) and the error without using it (x-axes). Figure 2 shows this relation for the twelve datasets, for the least (left) and most (right) relevant attributes. Points under the line indicate that the correction improves the classification error. In most datasets, the correction improves the classification when the missing attribute is the least relevant. Whereas, when it is the most relevant, it is better to classify objects without the correction. This may happen because important information on class differences is not taken into account by optimizing the correction.

TABLE II
TABLE OF MEAN RANKS OVER ALL CLASSIFIERS, WHEN THE MISSING ATTRIBUTE IS THE LEAST RELEVANT

	Complete (DS)	Complete (FS)	Using original classifier (DS)	Applying correction and α^j (DS)	Projection (FS)	Imputation SVR (FS)
Fisher	5.333	12.833	7.125	<u>5.875</u>	10.750	11.333
1NN	9.458	7.417	11.083	9.250	7.292	8.250
NBC	13.917	7.500	13.917	14.000	7.750	7.917

TABLE III
TABLE OF MEAN RANKS OVER ALL CLASSIFIERS, WHEN THE MISSING ATTRIBUTE IS THE MOST RELEVANT

	Complete (DS)	Complete (FS)	Using original classifier (DS)	Applying correction and α^j (DS)	Projection (FS)	Imputation SVR (FS)
Fisher	4.833	10.833	<u>6.792</u>	7.625	12.833	12.750
1NN	8.042	6.208	10.583	9.750	8.000	8.000
NBC	12.917	7.000	14.000	13.750	8.917	8.167

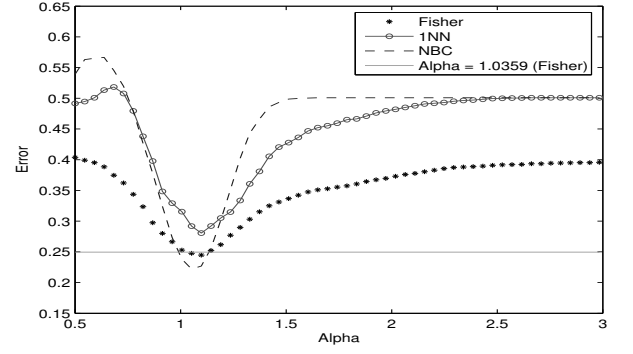
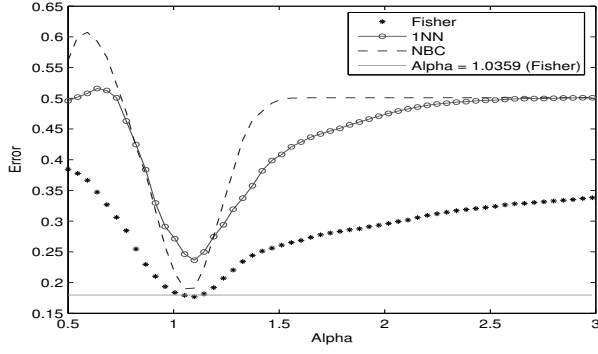


Fig. 1. Classification error obtained with the application of the factor α , for the intubation dataset, when the missing attribute is the least (left) and most (right) relevant.

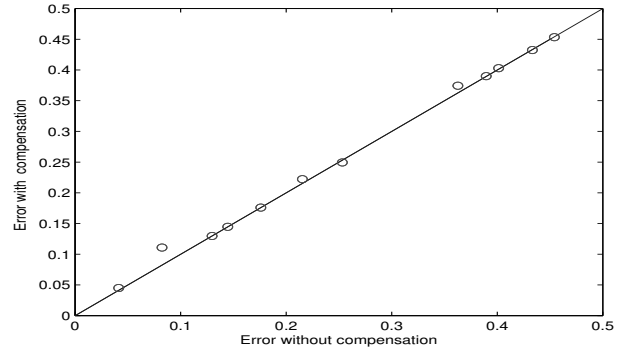
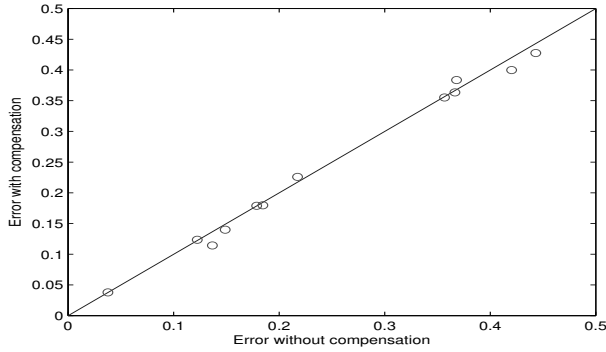


Fig. 2. Effect on the classification error by applying correction to incomplete objects in the dissimilarity space for the twelve datasets, when the missing attribute is the least (left) and most (right) relevant.

IV. CONCLUSIONS AND FURTHER RESEARCH

In this paper, we introduced two simple strategies using the dissimilarity representation for the classification of incomplete objects. Missing attributes considered in the experiments were the least and most relevant ones using the twelve real datasets. Results obtained with both approaches are promising since, unlike the methods of handling and classifying incomplete data in feature spaces, they have the following advantages:

- Classifiers hardly need to be updated, since the original trained classifiers in the dissimilarity space can be used to classify any incomplete object regardless of which

attribute is missing.

- Imputation techniques are not needed.
- As additional computations are not required, processing time is less.

Besides, for the twelve datasets the proposed approaches show, in general, better results than those of the *Projection* technique and the *Imputation* method using SVR. From the three classifiers studied in the experiments, Fisher classifier shows better performance for the proposed techniques, whereas 1NN and NBC classifiers showed better results in the feature space. In spite of this, the proposed approaches using

the Fisher classifier outperform other classifiers in the feature space. Moreover, the results obtained with both approaches for incomplete objects are better than the results for complete objects in the feature space.

Our results are based on just a single missing attribute. In case more attributes are missing, the obtained dissimilarities will be more severely affected. It has to be studied to what extend a correction can still be used and when a recomputation of the classifier has to be preferred.

ACKNOWLEDGEMENT

This work was supported in part by the Spanish Ministry of Education and Science under grants TIN2009-14205, CSD2007-00018 (Consolider-Ingenio 2010), and by the FET programme within the EU FP7 under the project "Similarity-based Pattern Analysis and Recognition - SIMBAD" (contract 213250). The work was undertaken while the first author was a guest researcher in ICT Group at Delft University of Technology, The Netherlands, (grant BES-2007-16184).

REFERENCES

- [1] A. Asuncion and D. J. Newman, UCI Machine Learning Repository, School of Information and Computer Science, University of California, Irvine, CA, 2007. <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- [2] A. Farhangfara, L. Kurganb and J. Dy, Impact of imputation of missing values on classification error for discrete data. *Pattern Recognition*, vol. 41, 2008, 3692 - 3705.
- [3] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge University Press, 1996.
- [4] Bruzzone, L., Roli, R., Serpico, S.B.: An extension of the Jeffreys–Matusita distance to multiclass cases for feature selection. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 33(6), 1995, 1318-1321.
- [5] E. Pekalska and R.P.W. Duin, Dissimilarity-based classification for vectorial representations, in: Y.Y. Tang, S.P. Wang, G. Lorette, D.S. Yeung, H. Yan (eds.), Proc. of the 18th Int. Conf. on Pattern Recognition (ICPR2006, Hong Kong, China, August 2006), vol. 3, IEEE Computer Society Press, Los Alamitos, 2006, 137-140.
- [6] E. Pekalska and R.P.W. Duin, Dissimilarity representations allow for building good classifiers, *Pattern Recognition Letters*, vol. 23, no. 8, 2002, 943-956.
- [7] E. Pekalska and R.P.W. Duin, The Dissimilarity Representation for Pattern Recognition. *Foundations and Applications*, World Scientific, Singapore, 2005.
- [8] F. Honghai, C. Guoshun, Y. Cheng, Y. Bingru and C. Yumei, A SVM Regression Based Approach to Filling in Missing Values, *LNC3 - Knowledge-Based Intelligent Information and Engineering Systems*, Springer Berlin - Heidelberg, vol. 3683, 2005, 581-587.
- [9] M. Friedman, The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance, *Journal of the American Statistical Association*, vol. 32(200), 1937, 675-701.
- [10] N. Cristianini and J. Shawe-Taylor, Support Vector Machines and other kernel-based learning methods. *Cambridge University Press*, UK, 2000.
- [11] R.J.A. Little, D.B. Rubin, Statistical Analysis with Missing Data. Wiley, New York, 1987.
- [12] S. Zhang, Parimputation: From imputation and null-imputation to partially imputation. *IEEE Intelligent Informatics Bulletin*, vol. 9(1), 2008, 32-38.
- [13] T. Delavallade and T.H. Dang, Using Entropy to Impute Missing Data in a Classification Task. In: *IEEE International Conference on Fuzzy Systems*, London, 2007, 1-6.
- [14] V.N. Vapnik, The Nature of Statistical Learning Theory. NY, Springer-Verlag, 1995.
- [15] W. Xian, L. Ao, J. Zhaohui, and F. Huanqing, Missing value estimation for DNA microarray gene expression data by Support Vector Regression imputation and orthogonal coding scheme. *BMC Bioinformatics*, vol. 7(1), 2006.
- [16] Y. Zhang, Y. Liu, Data Imputation Using Least Squares Support Vector Machines in Urban Arterial Streets, *IEEE Signal Processing Letters*, vol. 16 (5), 2009, 414-417.