# Comparing dissimilarity measures for probabilistic symbolic objects

D. Malerba, F. Esposito & M. Monopoli
*Dipartimento di Informatica,*
*Università di Bari, Italy.*

## Abstract

Symbolic data analysis generalizes some standard statistical data mining methods, such as those developed for classification and clustering tasks, to the case of symbolic objects (SOs). These objects, informally defined as "aggregated data" because they synthesize information concerning a group of individuals of a population, ensure confidentiality of original data, nevertheless they pose new problems which finds a solution in symbolic data analysis. A by-product of working with aggregate data is the possibility of dealing with data from complex questionnaires, where multiple answers are possible or constraints among different answers exists. Comparing SOs is an important step of symbolic data analysis. It can be useful either to cluster some SOs or to discriminate between them, or even to order SOs according to their degree of generalization. This paper presents a comparative study aiming at evaluating the degree of *dissimilarity* between the objects of a restricted class of symbolic data, namely Probabilistic Symbolic Objects. To define a ground truth for the empirical evaluation, a data set with understandable and explainable properties has been selected. In the experiment, only two dissimilarity measures, among the seven ones we have studied, seems to have a more stable behaviour.

## 1 Symbolic data analysis

Most of statistical data mining techniques are designed for a relatively simple situation: the unit for statistical analysis is an individual (e.g., a person or an object) described by a well defined set of random variables (either qualitative or quantitative), each of which result in just a *single* value. However, in many situations data analysts cannot access the single individuals (*first-order objects*). A typical situation is that of census data, which raise privacy issues in all governmental agencies that distribute them. To guarantee that data analysts cannot identify an individual or a single business establishment, data are made available in aggregate form, such as “the schools attended by students living in a given enumeration district or census tract and commuting by train for more than one hour per day are in Bredbury and Brinnington”. Aggregated data describe more or less homogeneous classes or groups of individuals (*second-order objects*) by means of *set-valued* or *modal* variables (for a formal definition see [1]). A variable $Y$ defined for all elements $k$ of a set $E$ is termed *set-valued* with the domain $\mathcal{Y}$ if it takes its values in $\boldsymbol{P}(\mathcal{Y})=\{U \mid U \subseteq \mathcal{Y}\}$, that is the power set of $\mathcal{Y}$. When $Y(k)$ is finite for each $k$, then $Y$ is called *multi-valued*. A single-valued variable is a special case of set-valued variable for which $|Y(k)|=1$ for each $k$. When an order relation $\prec$ is defined on $\mathcal{Y}$ then the value returned by a set-valued variable can be expressed by an interval $[\alpha,\beta]$, and Y is termed an *interval* variable. More generally, a *modal* variable is a set-valued variable with a measure or a (frequency, probability or weight) distribution associated to $Y(k)$.

The description of a class or a group of individuals by means of either set-valued variables or modal variables is termed *symbolic data* or *symbolic object*. More specifically a *Boolean symbolic object* (BSO) is described by set-valued variables only, while a *probabilistic symbolic object* (PSO), which is a specific case of modal SO, is described by modal variables with a relative frequency distribution associated to each of them.

A set of symbolic data, which involves the same variables to describe different (possibly overlapping) classes of individuals, can be described by a single table, called symbolic data table, where rows correspond to distinct symbolic data while columns correspond descriptive variables. Symbolic data tables are more complex to be analysed than standard data tables, since each item at the intersection of a row and a column can be either a finite set of values, or an interval or a probability distribution. The main goal of the research area known as symbolic data analysis is that of investigating new theoretically sound techniques to analyse such tables [2].

Most of techniques currently developed in symbolic data analysis are extensions of statistical methods, where the computation of dissimilarity (or conversely, similarity) measures is crucial. Many proposals of dissimilarity measures for BSOs have been reported in literature; an extensive review of their definitions is reported in [3], while a preliminary comparative study on their suitability to real-world problems is reported in [4]. They have been implemented in a software package developed for the three-years ESPRIT project SODAS[1] (Symbolic Official Data Analysis System), concluded in November 1999. The recently started three-years IST project ASSO (Analysis System of Symbolic Official Data) (http://www.assoproject.be/) is intended to improve the SODAS prototype with respect to several aspects, one of which is the extension of dissimilarity measures to PSOs.

[1] The SODAS software can be downloaded from: http://www.ceremade.dauphine.fr/~touati/sodas-pagegarde.htm.

In this paper a set of dissimilarity measures are proposed for the case of PSOs defined by multi-valued variables. Their definitions are based on different measures of divergence between two discrete probability distributions, which are associated to each *Y*(*k*) for some multi-valued variable *Y*. In Section 2 some coefficients measuring the divergence between discrete probability distributions are briefly presented, and some dissimilarity measures are defined by symmetrizing such coefficients. Possible aggregations of dissimilarity measures computed for each multi-valued variable are proposed in Section 3. An empirical evaluation of dissimilarity measures between PSOs is reported in Section 4.

## 2 Comparison functions for discrete probability distributions

Let *S* be a sample space and $Y_p'$, $Y_q'$ two random variables on *S* with the same discrete space $\mathcal{Y}$ (e.g., $\mathcal{Y}$ = {0, 1, 2, ...}). Let *p(y)* and *q(y)* denote the probabilities induced on the point $y \in \mathcal{Y}$ by the probabilities assigned to outcomes of the sample space *S* through the functions $Y_p'$ and $Y_q'$, respectively. The induced probabilities for each $y \in \mathcal{Y}$ define two probability distributions, henceforth denoted as *P* and *Q*, associated to the two random variables. For instance, if *S*={*red*, *white*, *black*} and both $Y_p'$ and $Y_q'$ associate *red* to 0, *white* to 1 and *black* to 2, then $\mathcal{Y}$ = {0, 1, 2} and *P* and *Q* will be both triples of real numbers representing the probabilities (p(0),p(1),p(2)) and (q(0), q(1), q(2)), respectively. For the sake of simplicity, in the following we will denote the two probability distributions *P* and *Q* as follows: *P*=(red:p(0), white:p(1), black:p(2)) and *Q*=(red:q(0), white:q(1), black:q(2)). For instance, we may have: *P*=(red: $0.\overline{3}$ , white: $0.\overline{3}$ , black: $0.\overline{3}$ ) and *Q*=(red:0.1, white:0.2, black:0.7).

Given two discrete probability distributions *P* and *Q*, we are interested in comparing them. Indeed, if *p* and *q* are two rows of a symbolic data table and *Y* is a multi-valued modal variable which contributes to the description of the PSOs in the same data table, then each probability distribution may be associated to *Y(p)* and *Y(q)*, respectively. By comparing *P* and *Q* we assess the similarity between the two SOs *p* and *q* when the variable *Y* alone describes them. The aggregation of partial similarities computed for each variable $Y_j$ describing a set of SOs will be defined in the next section, while in this section we introduce some comparison functions *m(P,Q)* for probability distributions, most of which belongs to the large family of "convex likelihood-ratio expectations" introduced by Csiszàr [5], Ali and Silvey [6].

This family of dissimilarity coefficients is defined as the expected value of a continuous convex function Φ of the likelihood ratio *r=p(x)/q(y)*, $E_P[\Phi(r)]$, where Φ(1) = 0. It obeys the discriminating property, according to which the measure of divergence between the two discrete events should not decrease for any refinement of both the discrete events and their two distributions [7].

- The *KL-divergence* is a measure of the difference between two probability distributions [8]. It is defined as $m_{KL}(P,Q) := \sum_{y \in Y} q(y) \log(q(y)/p(y))$ and measures to which extent the distribution *P* is an approximation of the distribution $Q$ or, more precisely, the loss of information if we take $P$ instead of $Q$. Stated differently, this is a measure of divergence of P (the subject) from Q (the referent). The KL-divergence is generally greater than zero, and it is zero only when the two probability distributions are equal. However, it is impossible to define in absolute terms whether $Q$ is a good approximation of $P$ by looking at $m_{KL}(P,Q)$. It is asymmetric, that is $m_{KL}(P,Q) \neq m_{KL}(Q,P)$ in general, and it is not defined when $p(y)=0$. In the special case of $p(y)/q(y)=0$, it is typically set $q(y)log(q(y)/p(y))=0$.

- The $\chi^2$*-divergence* defined as $m\chi^2(P,Q):=\Sigma_{y\in Y} |p(y)-q(y)|^2/p(y)$, is strictly topologically stronger then KL-divergence since the inequality $m_{KL}(P,Q) \leq m\chi^2(P,Q)$ holds, i.e. the convergence in $\chi^2$-divergence implies convergence in KL-divergence, but the converse is not true [9]. Similarly to the KL-divergence, it is asymmetric and is not defined when $p(y)=0$.
- The Hellinger Coefficient is a similarity-like coefficient. It is given by $m^{(s)}(P,Q):=\Sigma_{y\in Y} q^s(y).p^{1-s}(y)$ where $s$ is a positive exponent with $0 < s < 1$. Hellinger's special case $s=½$ yields the symmetric coefficient $m_B^{(½)}(P,Q):=\Sigma_{y\in Y} (q(y).p(y))½$ which has been termed *Bhattacharyya coefficient* [10,11]. This is a well known measure of the similarity (correlation) between two arbitrary statistical distributions. It is a sort of overlap measure between the two distributions: when their overlap is zero (one), they are completely distinguishable (indistinguishable).
  From this similarity-like measure, a dissimilarity coefficient has been derived, namely the *Chernoff's distance of the order s* defined as $m_C^{(s)}(P,Q) := -\log m^{(s)}(P,Q)$. Moreover, it is related to KL-divergence through its slope at s=0, that is $\partial m_C^{(s)}(P,Q)/ \partial s|_{s=0}=m_{KL}(P,Q)$, it is smaller than $m_{KL}(P,Q)$ and it is less sensitive than the KL-divergence to outlier values in $Y$. In particular a single value $y\in Y$ which has a non zero probability $Q$ but zero probability $P$, causes $m_{KL}(P,Q)$ to diverge, whereas $m_C^{(s)}(P,Q)$ remains finite. Indeed Chernoff's distance diverges only when the two distributions have zero overlap, i.e., the intersection of their support is empty [12].
  The *Rènyi's divergence* (or *information gain*) of order $s$ between two probability distributions $P$ and $Q$ is given by $m_R^{(s)}(P,Q):=-\log m^{(s)}(P,Q)/(s-1)$. This generalized information measure was originally introduced for the analysis of memory-less sources. It is noteworthy that, as $s\rightarrow 1$, the Rènyi's divergence approaches the KL-divergence [13].
- The *variation distance*, given by $m_1(P,Q):= \Sigma_{y\in Y} |p(y)-q(y)|$, is also known as *Manhattan distance* for the probability functions $p(y)$ and $q(y)$ and coincides with the *Hamming distance* when all features are binary. Similarly, it is possible to use *Minkowski's* $L_2$ (or *Euclidean*) *distance* given by $m_2(P,Q):=\Sigma_{y\in Y} |p(y)-q(y)|^2$ and, more in general, the Minkowski's $L_p$ distance with $p \in \{1,2,3, \ldots \}$. All measures $m_p(P,Q)$ satisfy the metric properties and in particular the symmetry property. The main difference

between $m_1$ and $m_p$, $p>1$, is that the former does not amplify the effect of single large differences (outliers). This property can be important when the distributions $P$ and $Q$ are estimated from noisy data.

- The *K–divergence* is given by $m_K(P,Q) := \Sigma_{y\in Y}\, q(y)\log(q(y)/(½p(y)+½q(y)))$ [14], which is an asymmetric measure. It is closely related to $m_{KL}(P,Q)$ through the following relationship $m_K(P,Q):=m_{KL}(P,½P+½Q)$. It has been proved that K-divergence is also related to the variation distance $m_1(P,Q)$ by the inequality: $m_K(P,Q)\leq m_1(P,Q) \leq 2$. Therefore, the K-divergence is upper bounded.

### 2.1 Symmetrization of non-symmetric coefficients

Henceforth, the term *dissimilarity measure d* on a set of objects $E$ refers to a real valued function on $E\times E$ such that $d_a^*=d(a,a)\leq d(a,b)=d(b,a)<\infty$ for all $a,b\in E$. Conversely, a *similarity measure s* on a set of objects $E$ is a real valued function on $E\times E$ such that $s_a^*=s(a,a)\geq s(a,b)=s(b,a)\geq 0$ for all $a,b\in E$. Generally, $d_a^* = d^*$ and $s_a^* = s^*$ for each object $a$ in $E$, and more specifically, $d^* = 1$ while $s^* = 0$. Studies on their properties can be limited to dissimilarity measures alone, since it is always possible to transform a similarity measure into a dissimilarity one with the same properties.

Some of the divergence coefficients defined above do not obey all the fundamental axioms that dissimilarities must satisfy. For instance, the KL-divergence does not satisfy the symmetric property. Nevertheless, a symmetrized version, termed *J-coefficient* (or *J-divergence*), can be defined as follows $J(P,Q) := m_{KL}(P,Q) + m_{KL}(Q,P)$. Alternatively, many authors have defined the *J-divergence* as the average rather than the sum $J(P,Q):=(m_{KL}(P,Q)+m_{KL}(Q,P))/2$. Generally speaking, for any (*possible*) non-symmetric divergence coefficient $m$ there exists a symmetrized version $\underline{m}(P,Q)= m(Q,P) + m(P,Q)$ which fulfils all axioms for a dissimilarity measure, but typically not the triangle inequality. The symmetrized versions of the other coefficients just presented are:

- $\underline{m\chi^2}(P,Q) := \Sigma_{y\in Y}\, |p(y)-q(y)|^2/p(y) + \Sigma_{y\in Y}\, |q(y)-p(y)|^2/q(y)$ obtained by symmetrizing $\chi^2$ – divergence;
- $\underline{m_C}^{(s)}(P,Q) := -\log m^{(s)}(P,Q) - \log m^{(s)}(Q,P)$ obtained by symmetrizing Chernoff's distance;
- $\underline{m_R}^{(s)}(P,Q) := \log m^{(s)}(P,Q)/(s-1) + \log m^{(s)}(Q,P)/(s-1)$ obtained by symmetrizing Rènyi's distance;
- $L(P,Q):=\Sigma_{y\in Y}\, q(y)\log(q(y)/(½p(y)+½q(y)))+\Sigma_{y\in Y}\, p(y)\log(p(y)/(½p(y)+½q(y))$ obtained by symmetrizing the K – divergence.

Obviously, in the case of Minkowski's $L_p$ coefficient, which satisfies the properties of a dissimilarity measure and, more precisely of a metric (triangular inequality), no symmetrization is required.

## 3 Dissimilarity measures for probabilistic symbolic objects

Many methods have been reported in the literature to derive dissimilarity measures from a matrix of observed data, or, more generally, for a set of BSOs [3]. In the following, only some measures proposed for PSOs in the new ASSO project are briefly reported. Let $a$ and $b$ be two PSOs:

$a = [Y_1 \in A_1] \wedge [Y_2 \in A_2] \wedge \ldots \wedge [Y_n \in A_n]$ and $b = [Y_1 \in B_1] \wedge [Y_2 \in B_2] \wedge \ldots \wedge [Y_n \in B_n]$

where each variable $Y_j$ is modal and takes values in the domain $\mathcal{Y}_j$ and $A_j$ , $B_j$ are subsets of $\mathcal{Y}_j$. A dissimilarity function between $a$ and $b$ can be built by aggregating dissimilarity coefficients between probability distributions through the generalized and weighted Minkowski's metric:

$$d_p(a,b) = \sqrt[p]{\sum_{k=1}^{n} \left[c_k \; m\left(A_k, B_k\right)\right]^p}$$

where, $\forall k \in \{1, \ldots, n\}$, $c_k > 0$ are weights with $\Sigma_{k=1..n}\, c_k=1$ and, by setting $P=A_k$ and $Q=B_k$, $m(P,Q)$ is a dissimilarity coefficient between probability distributions.

Alternatively, the dissimilarity coefficients can be aggregated through the product. Therefore, by adopting appropriate precautions and considering only the Minkowski's $L_p$ distance, we obtain the following normalized dissimilarity measure between PSOs:

$$d'_p(a,b) = 1 - \frac{\prod_{i=1}^{n} \left( \sqrt[p]{2} - \sqrt[p]{\sum_{y_i} \left| p(y_i) - q(y_i) \right|^p} \right)}{\left(\sqrt[p]{2}\right)^n} = 1 - \frac{\prod_{i=1}^{n} \left( \sqrt[p]{2} - \sqrt[p]{L_p} \right)}{\left(\sqrt[p]{2}\right)^n}$$

where each $y_i$ corresponds to a value of the $i$-th variable domain.

Note that this dissimilarity measure is symmetric and normalized in [0,1]. Obviously $d'_p(a,b) = 0$ if $a$ and $b$ are identical and $d'_p(a,b) = 1$ if the two objects are completely different.

## 4 Experimental evaluation

In the Section 2, several dissimilarity measures between PSOs have been proposed. They have never been compared in order to understand both their common properties and their differences. In this section, an empirical evaluation is reported with reference to the "Abalone Database", available at the UCI Machine Learning Repository (http://www.ics.uci.edu/~mlearn/MLRepository.html). The database contains 4,177 cases of marine crustaceans, which are described by means of the nine attributes, namely sex (nominal), length (continuous), diameter (continuous), height (continuous), whole weight (continuous), shucked weight (continuous), viscera weight (continuous), shell weight (continuous), and number of rings (integer). The age in years of an abalone can be obtained by adding 1.5 to the number of rings. Generally this data set is used for prediction tasks, where the number of rings is the target attribute. The number of rings varies between 1 and 29 and we expect that two abalones with the same number of rings should also present similar values for the independent attributes sex, length, diameter, height, and so on. In other words, *the degree of dissimilarity between crustaceans computed on the independent attributes should be proportional to the dissimilarity in the dependent attribute* (i.e., the difference in the number of rings). This property is called *monotonic increasing dissimilarity* (MID).

The experimental evaluation has been performed through three different steps. Firstly, the continuous attributes have been discretized[2] since the proposed dissimilarity measures must be computed only on modal symbolic variables. Secondly, abalone data have been aggregated into symbolic objects, each of which corresponds to a range of values for the number of rings. In particular, nine PSOs have been generated by means of the DB2SO tool [15]. Thirdly, the dissimilarity measures presented in Section 2 have been applied to these PSOs.

Since the computation of some dissimilarity coefficients is indeterminate when a distribution has a zero-valued probability for some categories, the *KT-estimate* has been used to estimate the probability distribution. This estimate is based on the idea that no category of a modal symbolic variable in a PSO can be associated with a zero probability. Indeed, the KT-estimate is the following:

$$p(y) = \frac{\left(\#\,\text{times } y \text{ occurs in } \{R_1,\ldots,R_M\}\right) + 1/2}{M + (K/2)}$$

where $y$ is the category of the modal symbolic variable, $\{R_1,\ldots,R_M\}$ are set of aggregated individuals, $M$ is the number of individual in the class, and $K$ is the number of categories of the modal symbolic variable.

By ordering the nine PSOs in ascending order with respect to the number of rings, the MID property can be formally expressed by a Robinsonian dissimilarity matrix $D=(d_{kl})_{k,l=1,\ldots,9}$ between PSOs such that dissimilarities $d_{kl}$ increases when $k$ and $l$ moves away from the diagonal (with k=l). In this work, we are interested to understand which dissimilarity measures defined for PSOs returns (an approximation of) a Robinsonian dissimilarity matrix [16].

Two classes of dissimilarity measures have been considered, namely the generalized Minkowski's measure for PSOs (*PU_1*) combined with one of the following comparison functions:

- J: *J-coefficient*
- CHI2 : $\chi^2$ – *divergence*
- REN: *Rènyi's distance*
- CHER: *Chernoff's distance*
- LP: *Minkowski's $L_p$ distance*
- L: *L divergence*

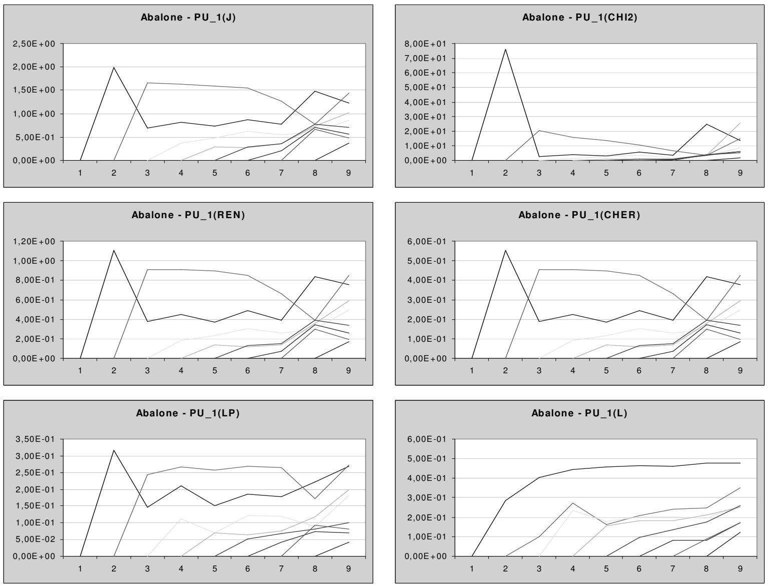
and the normalized distance function derived by aggregating $L_p$ coefficient

[2] For this process we used the Weka data mining software (http://www.cs.waikato.ac.nz/ml/weka/), which implements the *entropy-based* discretization method.

(*PU_2*). The experimental results have been obtained setting the value of the parameter *s* to 0.5, the order of power *q* to 2, the value of the parameter *p* to 2 and weights to uniform distribution.

Results are shown in Figure 1. Dissimilarities are reported along the vertical axis, while PSOs are listed along the horizontal one, in ascending order with respect to the number of rings. Each line represents the dissimilarity between a given PSO and the subsequent PSOs in the sorted list.

In [4] the same data set has been used for a comparative study of dissimilarity measures between BSOs. In that case, only some dissimilarity measures generated an approximate Robinsonian dissimilarity matrix. This result was partly expected due to the loss on information of the distribution of values when BSOs are generated. Surprisingly, similar results are also reported in the case of PSOs. In particular the dissimilarity matrix has no approximate Robinsonian value distribution in the case of the generalized Minkowski's measure (PU_1) combined with the comparison functions J, CHI2, REN, CHER, and LP. In these cases, the most atypical objects are PSO1 and PSO2. In fact, PSO1 and PSO7 are more similar than PSO1 and PSO2, as well PSO2 is more similar to PSO8 than PSO3. By combining PU_1 with REN and CHER coefficients, the dissimilarity measures show a similar behaviour. This is due to the fact that the formulation of



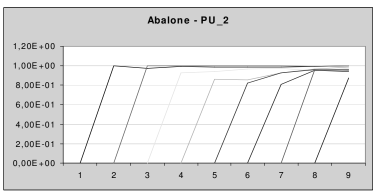




**Fig. 1** Graphs of six dissimilarity measures in Abalone data.

both coefficients derives from the Hellinger coefficient. When the dissimilarity measure is computed by means of the PU_1 aggregated with L divergence, it has a better Robinsonian approximation. Indeed, PSO4 seems to be the only slightly atypical object, since the dissimilarity between PSO2 or PSO3 and PSO4 is higher that the dissimilarity PSO2 or PSO3 and PSO5. Finally, the Robinsonian approximation also holds for PU_2. In this case all PSOs are totally dissimilar, since $d'_2(PSO_i, PSO_j) \approx 1$, for each $i \neq j$.

## 5 Conclusions

In symbolic data analysis a key role is played by the computation of dissimilarity measures. Many measures have been proposed in the literature, although a comparison that investigates their applicability to real data has never been reported. The main difficulty was due to the lack of a standard in the representation of SOs and the necessity of implementing many dissimilarity measures. The software produced by the ESPRIT Project ASSO has partially solved this problem by defining a suite of modules that enable the generation, visualization and manipulation of SOs.

In this work, a comparative study of the dissimilarity measures for PSOs is reported with reference to a particular data set for which an expected property could be defined. Interestingly enough, such a property has been observed only for some dissimilarity measures, which show very different behaviours. A more extensive experimentation is planned to confirm these initial observations.

## Acknowledgements

This work was partly supported by the European IST project no. 25161 ASSO and by the MURST project COFIN-2001 on "Methods of knowledge discovery, validation and representation of the statistical information in decision tasks".

## References

[1] Bock, H.H., Diday, E.: Symbolic Objects. In Bock, H.H, Diday, E. (eds.): *Analysis of Symbolic Data. Exploratory Methods for extracting Statistical Information from Complex Data*, Series: Studies in Classification, Data Analysis, and Knowledge Organisation, Vol. 15, Springer-Verlag, Berlin, pp. 54-77, 2000.

[2] Bock, H.H., Diday, E. (eds.): *Analysis of Symbolic Data. Exploratory Methods for Extracting Statistical Information from Complex Data*, Series: Studies in Classification, Data Analysis, and Knowledge Organisation, Vol. 15, Springer-Verlag, Berlin, 2000.

[3] Esposito, F., Malerba, D., Tamma, V.: Dissimilarity Measures for symbolic

objects. In: Bock, H.H, Diday, E. (eds.): *Analysis of Symbolic Data. Exploratory Methods for extracting Statistical Information from Complex Data*, Series: Studies in Classification, Data Analysis, and Knowledge Organisation, Vol. 15, Springer-Verlag, Berlin, pp. 165-185, 2000.

[4] Malerba, D., Esposito, F., Gioviale, V., Tamma, V.: Comparing Dissimilarity Measures for Symbolic Data Analysis. *Pre-Proceedings of ETK-NTTS 2001*, Hersonissos, Crete, vol. 1, pp. 473-481, 2001.

[5] Csiszár, I.: Information-type measures of difference of probability distributions and indirect observatons, *Studia Scient. Math. Hung*., vol. 2, pp. 299-318, 1967.

[6] Ali, S. M., Silvey S. D.: A general class of coefficient of divergence of one distribution from another, *Journal of the Royal Statistical Society*, Ser. B, vol. 2, pp. 131-142, 1966.

[7] Withers, L.: Some Inequalities Relating Different Measures of Divergence Between Two Probability Distributions, *IEEE Transactions on Information Theory*, vol. 45, n. 5, pp. 1728-1735, 1999.

[8] Kullback, S., Leibler, R.A.: On information and sufficiency. *Annals of Mathematical Statistics*, vol. 22, pp. 76-86, 1951.

[9] Beirlant,k J., Devroye, L., Györfi, L., Vajda, I.: Large deviations of divergence measures on partitions, *Journal of Statistical Planning and Inference*, vol. 93, pp. 1-16, 2001.

[10] Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by their population distributions, *Bulletin Calcutta Mathematical Society*, 35, 99-109, 1943.

[11] Toussaint, G., T.: Comments on the divergence and the Bhattacharrya distance measures in signal selection, *IEEE Transaction on Communication Technology*, Vol. COM-20, p.485, 1972.

[12] Kang, K., Sompolinsky, H.: Mutual Information of Population Codes and Distance Measures in Probability Space. *Physical Review Letters*, vol. 86, pp. 4958-4961, 2001.

[13] Rached, Z., Alajaji, F., Campbell, L.L.: Rényi's Divergence and Entropy Rates for Finite Alphabet Markov Sources. *IEEE Transactions on Information theory*, vol. 47, n. 4, pp.1553-1561, 2001.

[14] Lin, J.: Divergence Measures Based on the Shannon Entropy. *IEEE Transactions on Information theory*, 37(1):145--151, 1991.

[15] Stéphan, V., Hébrail, G., and Lechevallier, Y.: Generation of Symbolic Objects from Relational Databases. In: Bock, H.H., Diday, E. (eds.): *Analysis of Symbolic Data. Exploratory Methods for extracting Statistical Information from Complex Data*, Series: Studies in Classification, Data Analysis, and Knowledge Organisation, Vol. 15, Springer-Verlag, Berlin, pp. 78-105, 2000.

[16] Fichet, B.: Data Analysis: geometric and algebraic structures. In: Yu. Prochorov and V.V. Sazanov (eds.): *Proc. First World Congress of Bernoulli Society*, Tashkent, URSS, 1986. VNU Science Press, Utrecht, Netherlands, 1987, Vol. 2, 123-132, 1987.