

# Feature Selection Using Typical $\epsilon$ : Testors, Working on Dynamical Data

Jesús Ariel Carrasco-Ochoa<sup>1</sup>, José Ruiz-Shulcloper<sup>2</sup>,  
and Lucía Angélica De-la-Vega-Doría<sup>1</sup>

<sup>1</sup> Computer Science Department  
National Institute of Astrophysics, Optics and Electronics  
Luis Enrique Erro No. 1 Sta María Tonanzintla, Puebla, CP: 72840, Mexico  
{ariel,lucia}@inaoep.mx

<sup>2</sup> Advanced Technologies Application Center, MINBAS (Cuba)  
jshulcloper@cenatav.co.cu

**Abstract.** Typical  $\epsilon$ :testors are useful to do feature selection in supervised classification problems with mixed incomplete data, where similarity function is not the total coincidence, but it is a one threshold function. In this kind of problems, modifications on the training matrix can appear very frequently. Any modification of the training matrix can change the set of all typical  $\epsilon$ :testors, so this set must be recomputed after each modification. But, complexity of algorithms for calculating all typical  $\epsilon$ :testors of a training matrix is too high. In this paper we analyze how the set of all typical  $\epsilon$ :testors changes after modifications. An alternative method to calculate all typical  $\epsilon$ :testors of the modified training matrix is exposed. The new method's complexity is analyzed and some experimental results are shown.

## 1 Introduction

Yu. I. Zhuravlev introduced the testor concept to Pattern Recognition problems [4]. Upon which a new research line was opened [3]. Specially, there are many researches pointing to the development and application of algorithms for calculating all typical testors of a given matrix. Testors can be applied on the problem of feature selection for supervised classification; working in situations where there are qualitative and quantitative features and may be incomplete object descriptions. Zhuravlev's concept was defined for Boolean features. After, this concept was extended to any kind of features, but anyway all comparison criteria have to be Boolean. Another limitation of Zhuravlev's testor concept is that the implicit similarity function is the total coincidence, but in some problems two objects could be similar even if they are different in some features, this type of similarity functions are called one threshold similarity functions. So a new testor concept was developed for one threshold similarity functions, the  $\epsilon$ :testor concept [3].

Using typical  $\epsilon$ :testors to solve the feature selection problem, has as inconvenient the high complexity of algorithms for computing all typical  $\epsilon$ :testors, which makes impossible to use those in situations where the training matrix changes frequently. In this paper we analyze the behavior of the set of all typical  $\epsilon$ :testors when the comparison matrix changes. Also we propose an alternative method for adjusting this set after modifications, and a new incremental algorithm for computing the set of all typical

$\epsilon$ :testors. The new methods' performance is shown. Finally complexity analysis of the adjusting method is done.

## 2 Typical $\epsilon$ : Testors and Feature Selection

Into the framework of the Logical Combinatorial Pattern Recognition [1], feature selection is done using typical testors [1,2,3].

Let TM be a training matrix with  $m$  objects described using  $n$  features of any type  $(x_1, \dots, x_n)$  and grouped in  $r$  classes. Let M be a dissimilarity Boolean matrix (0=similar, 1=dissimilar), obtained from feature by feature comparisons of all the pairs of objects from TM belonging to different classes. M has  $n$  columns  $(x_1, \dots, x_n)$  and  $k$  rows  $(f_1, \dots, f_k)$ . Testors and Typical Testors are defined as follows:

**Definition 1.** A subset of features  $T$  is a testor of M if and only if when all features are eliminated, except those from  $T$ , there is not any row of M with only 0's.

**Definition 2.** A subset of features  $T$  is a typical testor of M if and only if  $T$  is a testor of M and there is not any other testor of M  $T'$  such that  $T' \subset T$ .

As it was mentioned, these definitions have as implicit similarity function the total coincidence, where two objects are similar if they are similar in all their features.

For a one threshold similarity function with threshold= $\epsilon$ , two objects are similar if they are different in less than  $\epsilon$  features. For a one threshold similarity function with threshold= $\epsilon$ ,  $\epsilon$ :Testors and Typical  $\epsilon$ :Testors are defined as:

**Definition 3.** A subset of features  $T$  is an  $\epsilon$ :testor of M if and only if when all features are eliminated, except those from  $T$ , there is not any row with less than  $\epsilon$  1's..

**Definition 4.** A subset of features  $T$  is a typical  $\epsilon$ :testor of M if and only if  $T$  is a  $\epsilon$ :testor of M and there is not any other  $\epsilon$ :testor of M  $T'$  such that  $T' \subset T$ .

Each  $\epsilon$ :testor is a set of features such that there is not a pair of objects from different classes which are similar for a one threshold similarity function with threshold= $\epsilon$ . Typical  $\epsilon$ :testors are irreducible  $\epsilon$ :testors, this is, all features are essential for class separation. Therefore, typical  $\epsilon$ :testors can be used as feature selection.

## 3 Working on Dynamical Environments

The main problem of the  $\epsilon$ :testor concept is that all algorithms for computing the set of all typical  $\epsilon$ :testors are of exponential complexity with regards to the number of features. So, using them for feature selection on problems where data change frequently (Dynamical Data) is inadequate.

For this cause, it is necessary to find a new method for adjusting the set of all typical  $\epsilon$ :testors when there are some changes on data. The search of this new method is called sensitivity analysis [5].

It is known [5] that all possible alterations to a training matrix TM can be summarized in 4 cases: delete a feature; add a feature; delete an object; add an object, or in successive compositions of them. All possible alterations on TM may have an effect on the comparison matrix M; it is shown in the table 1.

**Table 1.** Modifications on the training matrix TM and their effects on the comparison matrix M.

Alteration on TM	Effect on M
Delete a feature	The corresponding column is eliminated.
Add a feature	The corresponding column is added.
Delete an object	Rows that come from the comparisons with the deleted object are eliminated.
Add an object	Rows that come from the comparisons with the new object are added.

Clearly, possible modifications to the comparison M also can be reduced to the same 4 cases. For this reason, we only study how the set of all typical  $\epsilon$ :testors changes because of each one of this 4 types of alterations.

A consequence of the typical  $\epsilon$ :testor concept (definition 4) is the following:

Let  $t = \{x_{i_1}, \dots, x_{i_s}\}$  be a typical  $\epsilon$ :testor of M, then there are  $r < s$  rows (associated to

t)  $F_t = \{f_{i_1}, \dots, f_{i_r}\}$  such that:

1. For each feature  $x_{i_j} \in t$  there is a row  $f_{i_k} \in F_t$  such that  $|(t \setminus \{x_{i_j}\}) \cap f_{i_k}| < \epsilon + 1$
2. For each row  $f_{i_k} \in F_t$  there is a feature  $x_{i_j} \in t$  such that  $|(t \setminus \{x_{i_j}\}) \cap f_{i_k}| < \epsilon + 1$

Also, for each row  $f_{i_k} \in F_t$ ,  $F_t \setminus \{f_{i_k}\}$  does not fulfill these conditions.  $F_t$  can be not unique. We denote by  $F_t$  the set of all possible  $F_t$ .

In order to simplify the notation, we suppose that the comparison matrix is a set of rows, and each row is a set of those features which there is a 1 in the corresponding column. ( for example if  $n=5$  then (1 0 1 1 0) is the set  $\{x_1, x_3, x_4\}$  ).

The behavior of the set of all typical  $\epsilon$ :testors, when the comparison matrix M changes, is described in the following theorems (sensitivity theorems):

Let M be a dissimilarity comparison matrix whit values into  $\{0,1\}$ .

*Theorem 1.* Let  $M'$  the obtained matrix when the column  $x_{i_1}$  is eliminated from M, then  $T' = T \setminus T_{i_1}$  where:

$T'$  is the set of all typical  $\epsilon$ :testors of  $M'$

$T$  is the set of all typical  $\epsilon$ :testors of M

$T_{i_1}$  is the set of all typical  $\epsilon$ :testors of M which contains the feature  $x_{i_1}$ .

*Theorem 2.* Let  $M'$  the obtained matrix when the column  $x_{n+1}$  is added to M, then  $T' = T \cup T'_{n+1}$  where:

$T'$  is the set of all typical  $\epsilon$ :testors of  $M'$

$T$  is the set of all typical  $\epsilon$ :testors of M

$T'_{n+1}$  is the set of all typical  $\epsilon$ :testors of M which contains the feature  $x_{n+1}$ .

*Theorem 3.* Let  $M'$  the obtained matrix when the row  $f_{i_1}$  is eliminated from M, then  $T' = (T \setminus T_0) \cup T_1$  where:

$T'$  is the set of all typical  $\epsilon$ :testors of  $M'$

$T$  is the set of all typical  $\epsilon$ :testors of M

$$T_0 = \{t \in T \mid \text{there is } F_t \in F_t \text{ with } f_i \in F_t \text{ and there is no a row set } \{f_{j_1}, \dots, f_{j_s}\} \subseteq M, \\ f_i \notin \{f_{j_1}, \dots, f_{j_s}\} \text{ such that } (F_t \setminus \{f_i\}) \cup \{f_{j_1}, \dots, f_{j_s}\} \in F_t \}$$

$$T_1 = \bigcup_{M'' \in M^{(i)}} TT(M'')$$

where:

$TT(M'')$  is the set of all typical  $\epsilon$ :testors of  $M''$ ,

$M^{(i)}$  is the set of all matrices that can be constructed with all rows of  $M \setminus \{f_i\}$ , all columns where  $f_i$  has 0, and  $\epsilon$  columns where  $f_i$  has 1.

*Theorem 4.* Let  $M'$  the obtained matrix when the row  $f_{k+1}$  is added to  $M$ , then  $T' = (T \setminus T_0) \cup T_1$  where:

$T'$  is the set of all typical  $\epsilon$ :testors of  $M'$

$T$  is the set of all typical  $\epsilon$ :testors of  $M$

$$T_0 = \{t \in T \mid |t \cap f_{k+1}| \leq \epsilon\}$$

$$T_1 = \{t' \mid t' = t \cup E, t \in T_0, t \cap E = \emptyset, E \subseteq f_{k+1}, |E| = \epsilon - |t \cap f_{k+1}| + 1 \text{ and there is not any } t_0 \\ \epsilon\text{:testor of } M' \text{ with } t_0 \subset t'\}$$

Proofs for these 4 theorems can be found in [6].

### 4 Proposed Methods

Based on the sensitivity theorems it is possible to define a method for adjusting the set of all typical  $\epsilon$ :testors when a modification on the comparison matrix  $M$  occurs. For each kind of modification the adjusting method works as follows:

- If a column is eliminated, the set of all typical  $\epsilon$ :testors of the new matrix is obtained taking the set of all typical  $\epsilon$ :testors of the original matrix and discarding such testors which contain the eliminated column.
- If a column is added, the set of all typical  $\epsilon$ :testors of the new matrix is obtained taking the set of all typical  $\epsilon$ :testors of the original matrix and adding all typical  $\epsilon$ :testors that contain the added column. These new typical  $\epsilon$ :testors must be calculated with an adapted traditional algorithm, by example  $\epsilon$ :BT [3,6].
- If a row  $f_i$  is eliminated, the set of all typical  $\epsilon$ :testors of the new matrix is obtained applying the following process:

Verifying for each typical  $\epsilon$ :testor  $t$ , of the original matrix, if  $|t \cap f_i| \neq \epsilon + 1$  or  $|t \cap f_i| = \epsilon + 1$  and  $t \setminus \{x_k\}$  is not an  $\epsilon$ :testor of the new matrix, for all  $x_k \in (t \cap f_i)$ , then  $t$  is a typical  $\epsilon$ :testor of the new matrix, otherwise  $t$  is discarded. Additionally, building all different submatrices  $M''$  of the new matrix, taking all columns where  $f_i$  has 0 and  $\epsilon$  columns where  $f_i$  has 1. Calculate the set of all typical  $\epsilon$ :testors of each  $M''$ . The set of all typical  $\epsilon$ :testors of the new matrix is obtained from the union of the original set of the typical  $\epsilon$ :testors, which are not discarded, with the sets of all typical  $\epsilon$ :testors of all the matrices  $M''$ .

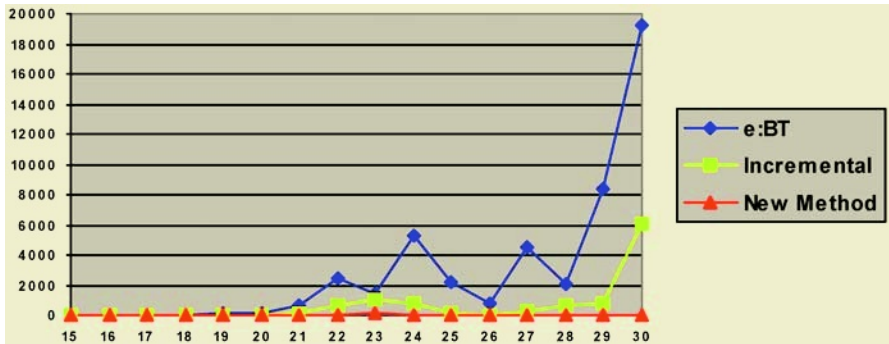
- If a row  $f_{k+1}$  is added, the set of all typical  $\epsilon$ :testors of the new matrix is obtained applying the following process to each typical  $\epsilon$ :testor of the original matrix.

Let  $t$  be a typical  $\epsilon$ :testor of the original matrix, if the added row  $f_{k+1}$  is such that  $|t \cap f_{k+1}| \geq \epsilon + 1$ , then  $t$  is a typical  $\epsilon$ :testor of the new matrix; if not,  $t$  must be substituted by the new typical  $\epsilon$ :testors obtained from  $t' = t \cup E$ , where  $E \subseteq f_{k+1}$ ,  $|E| = \epsilon - |t \cap f_{k+1}| + 1$ ,  $t \cap E = \emptyset$ , and  $t'$  is not a superset of any other  $\epsilon$ :testor.

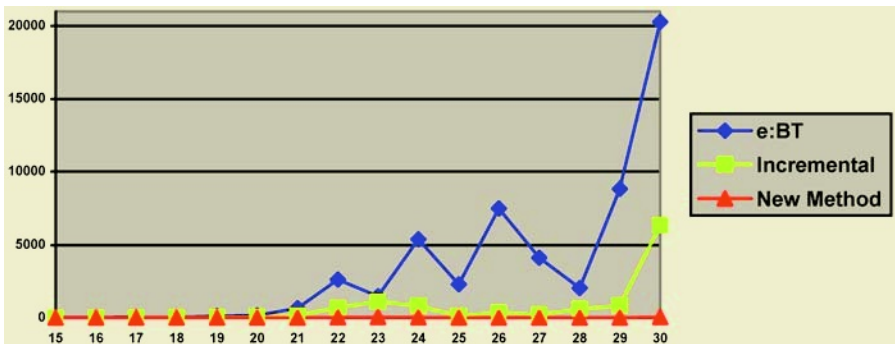
Additionally, based on the theorem 4, we define a new incremental algorithm for calculating all typical  $\epsilon$ :testors as follows: Take the first row of the comparison matrix  $M$ , all subset of exactly  $\epsilon$  features, taken from the columns where this row has 1, are the initial typical  $\epsilon$ :testors, then add all other rows applying the adjusting method.

### 5 Experimental Tests

Performance of the adjusting method was proved. Runtimes of the adjusting method were compared against runtimes of  $\epsilon$ :BT algorithm and the incremental algorithm based on theorem 4. First, some synthetic tests were done in order to show the new method performance, figures 1-4 show the results of these tests.



**Fig. 1.** Runtimes in hundredths of second for deleting a row with  $\epsilon=1$ , for matrixes from 15 to 30 columns.



**Fig. 2.** Runtimes in hundredths of second for adding a row with  $\epsilon=1$ , for matrixes from 15 to 30 columns.

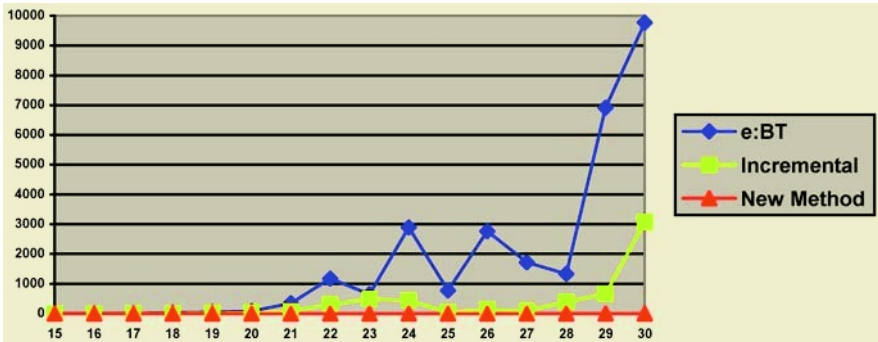


Fig. 3. Runtimes in hundredths of second for deleting a row with  $\epsilon=1$ , for matrixes from 15 to 30 columns.

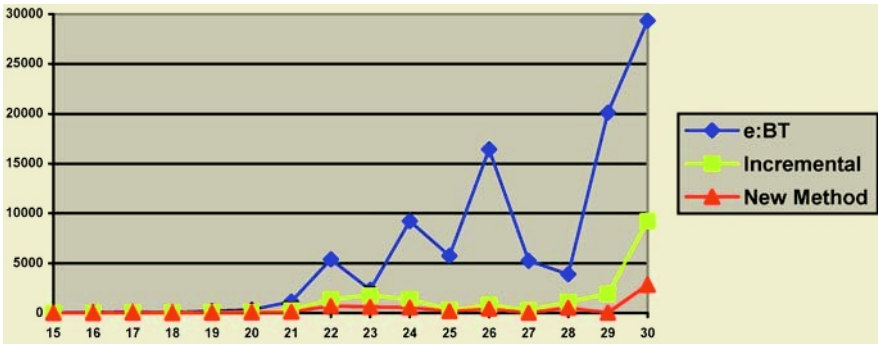


Fig. 4. Runtimes in hundredths of second for adding a column with  $\epsilon=1$ , for matrixes from 15 to 30 columns.

These synthetic tests were done with a randomly generated matrix for each different number of columns.

Additional tests were done with the “annealing” problem [7]. This problem has 798 objects described through 38 features: 6 real-valued; 3 integer-valued; and 29 nominal-valued. There are several missing values. This problem has mixed incomplete data descriptions. Comparison criteria were:

For real-valued features:

$$C_i(v_1, v_2) = \begin{cases} 1 & \text{if } |v_1 - v_2| > \sigma_i \\ 0 & \text{otherwise} \end{cases}$$

where  $\sigma_i$  is the standard deviation of the sample data for the feature  $x_i$ .

For integer-valued and nominal-valued features:

$$C_i(v_1, v_2) = \begin{cases} 1 & \text{if } v_1 \neq v_2 \\ 0 & \text{otherwise} \end{cases}$$

Using these comparison criteria, we got a basic matrix [3], from the comparison matrix, with 38 columns and 45 rows. Some tests were done with this matrix apply-

ing the adjusting method, the new incremental algorithm, and  $\epsilon$ :BT algorithm. Table 2 shows the results of these tests.

**Table 2.** Runtimes in hundredths of second for the “annealing” problem with  $\epsilon=1,2,3$ . (# of T.  $\epsilon$ :T means number of typical  $\epsilon$ :testors).

$\epsilon=1$					
Problem	Modification	Adjusting Method	Incremental	$\epsilon$ :BT	# of T. $\epsilon$ :T.
“annealing” 38x45	Original	NA	294	13290	2968
	Adding column	759	323	14298	3120
	Deleting column	<1	247	11446	2529
	Adding row	1	296	13548	2968
	Deleting row	8	272	12472	2970
$\epsilon=2$					
Problem	Modification	Adjusting Method	Incremental	$\epsilon$ :BT	# of T. $\epsilon$ :T.
“annealing” 38x45	Original	NA	16446	139532	29135
	Adding column	15956	18937	158125	32103
	Deleting column	3	13726	114613	23968
	Adding row	7	16520	142122	29135
	Deleting row	118	15927	136174	29176
$\epsilon=3$					
Problem	Modification	Adjusting Method	Incremental	$\epsilon$ :BT	# of T. $\epsilon$ :T.
“annealing” 38x45	Original	NA	793903	939564	184772
	Adding column	165847	954793	1122892	213907
	Deleting column	17	681475	734084	145979
	Adding row	44	801726	956008	184772
	Deleting row	1175	788636	916909	185167

Finally, complexity analysis of the new methods was done. The table 3 shows the complexity of the new methods for each kind of modification and for the incremental algorithm.

**Table 3.** Complexity of the new methods.

Modification	New method’s complexity
Delete a column	$w$
Add a column	$2^{n-1}$
Delete a row	$kw + \binom{n-y}{\epsilon} 2^{y+\epsilon}$
Add a row	$\binom{n-y}{\epsilon} w^2$
Incremental Algorithm	$M \binom{n-y}{\epsilon} w^2$

where  $n$  is the number of columns of the new matrix,  
 $k$  is the number of rows of the new matrix,  
 $w$  is the number of typical  $\epsilon$ :testors of the original matrix,  
 $y$  is the number of 0’s in the row added or deleted,  
 $\epsilon$  is the epsilon of the similarity function,  
 $m$  is the number of rows of the comparison matrix

## 6 Conclusions

As result of this work, it is possible to use typical  $\epsilon$ :testors for feature selection in problems where modifications appear frequently. The new method allows adjusting the set of all typical  $\epsilon$ :testors faster than applying any traditional algorithm, this can be appreciated from complexities, and it was corroborated with experimentation.

Also, a new incremental algorithm for calculating all typical  $\epsilon$ :testors were proposed. In all experiments this new incremental algorithm had a better performance than  $\epsilon$ :BT algorithm.

## Acknowledgement

This work was financially supported by CONACyT (Mexico) through project I38436-A.

## References

1. Martínez-Trinidad J. F. and Guzmán-Arenas A. "The logical combinatorial approach to pattern recognition an overview through selected works", *Pattern Recognition*, Vol 34, No. 4, pp. 741-751, 2001.
2. Ruiz-Shulcloper J. and Lazo-Cortés M. "Mathematical Algorithms for the Supervised Classification Based on Fuzzy Partial Precedence", *Mathematical and Computer Modeling*, Vol. 29, No. 4, pp. 111-119, 1999.
3. Lazo-Cortes M., Ruiz-Shulcloper J. and Alba-Cabrera E. "An overview of the evolution of the concept of testor", *Pattern Recognition*, Vol. 34, No. 4, pp. 753-762, 2001.
4. Dmitriev A.N., Zhuravlev Y.I., Krendeliev F.P. "About mathematical principles of objects and phenomena classification", *Diskretni Analiz* Vol 7, pp. 3-15, 1966. (In Russian).
5. Carrasco-Ochoa J. A. and Ruiz-Shulcloper J. "Sensitivity Problems of the Set of Typical Testors of a Boolean Matrix", *Proceedings of the III Iberoamerican Symposium on Pattern Recognition*, SIARP 98, pp. 257-266, Mexico, 1998 (in Spanish).
6. Carrasco-Ochoa J. A. "Sensitivity in Logical Combinatorial Pattern Recognition", PhD Thesis, CIC-IPN, Mexico, 2001
7. <http://www-old.ics.uci.edu/pub/machine-learning-databases/>