Feature Selection using Typical Testors applied to Estimation of Stellar Parameters
Selección de Características usando Testores Típicos aplicada a la Estimación de Parámetros Estelares

[1] José Á. Santos, [2] Ariel Carrasco y [2] José F. Martínez
Instituto Nacional de Astrofísica Óptica y Electrónica
(INAOE), México
[1] angellisse@hotmail.com
[2] {ariel, fmartinez}@inaoep.mx,

Abstract

In this paper a comparative analysis of feature selection using typical testors applied on astronomical data, is presented. The comparison is based on the classification efficiency using typical testors as feature selection method against the classification efficiency using Ramirez (2001) method, which uses genetic algorithms. The well-known *K*-nearest neighbors rule (KNN) was used as classifier. The feature selection based on typical testors was modified to be applied on a prediction problem of a real valued function. The feature selection obtained with typical testors reduces the amount of features in approximately 50% and the classification error index is better than both using the original data and Ramirez's method.
Keywords: Feature Selection; Typical Testors; Logical Combinatorial Pattern Recognition; Prediction of Stellar Parameters.

Resumen

En este artículo se presenta un análisis comparativo de la selección de características usando testores típicos aplicada a la estimación de parámetros estelares. La comparación está basada en la eficiencia de clasificación al usar testores típicos como método de selección de variables contra la eficiencia de clasificación al usar el método de Ramírez (2001), el cual usa algoritmos genéticos. Como clasificador se usó la regla de *K*-vecinos más cercanos (KNN). Se propuso una modificación al concepto de testor típico para poder aplicarlo en problemas de predicción de una función con imagen en los números reales. En los experimentos, la selección de variables con testores típicos redujo la cantidad de características aproximadamente en un 50%. El error de clasificación usando nuestro método fue menor que usando todas las características o el método de Ramírez.
Palabras Clave: Selección de características, Testores típicos, Reconocimiento lógico combinatorio de patrones, Predicción de parámetros estelares.

1 Introduction

The feature selection through typical testors is formulated in a supervised classification context with mixed incomplete data (see Lazo-Cortes *et al.*, 2001). However, the problem of estimation of stellar parameters consists in the prediction of continuous values; therefore, the feature selection method must be modified in order to be applied to this kind of problems. In the next sections, the modified feature selection method based on typical testors will be described. It will be explained how, given a selection, the prediction is made using the KNN rule. Some comparisons of feature selection using typical testors against Ramirez (2001) method on different data sets will be shown. Finally, the performance of our method on the problem of stellar parameters estimation will be compared.

2 Feature Selection Using Typical Testors

Into the framework of the Logical Combinatorial Pattern Recognition (Martínez-Trinidad *et al.*, 2001; Alba-Cabrera, 1997; Ruiz-Shulcloper, 1999), the feature selection is made using typical testors (see Lazo-Cortes *et al.*, 2001). If R is the whole set of features, a testor is defined as follows:

Definition 1. A feature subset *T* is a testor if and only if when all features are eliminated, except those from *T*, there is not any pair of similar subdescriptions in different classes.

This definition indicates us that a testor is a feature subset, which allows complete differentiation of objects from different classes. Within the set of all testors, there are some testors, which are irreducible. These kinds of testors are called typical testors. Typical testors are defined as follows:

Definition 2. A feature subset *T* is a typical testor if and only if *T* is a testor and there is not any other testor *T'* such that *T'* ⊂ *T*. This definition indicates us that a typical testor is a testor where every feature is essential, this is, if any of them is eliminated the resultant set is not a testor.

The approach based on Testor Theory was first proposed by Dimitriev et al. (1966) and the basic idea is the following: A testor is a feature subset, which does not confuse any pair of subdescriptions from different classes. Moving, from a testor to a typical testor (eliminating features, when it is possible) we get an irreducible combination of features, where each feature is essential in order to keep differences between classes. This property distinguishes each typical testor. It is natural to suppose that if a feature appears many times in different typical testors, it is more difficult to disregard it. That is, we could say it is more useful to differentiate between classes. Based on this idea, a definition of feature's weight as the relative frequency of the occurrence of each feature in the set of all typical testors is introduced. Let τ be the number of typical testors for a certain problem. Let $\tau(i)$ be the number of typical testors, which contain feature x_i. Feature's weight of x_i is given by:

$$P(x_i) = \frac{\tau(i)}{\tau} \qquad \text{for } i = 1,...,n, \text{ for } x_i \in R.$$

Once the typical testors and the features' weights are computed, fixing a minimal threshold of relevance or choosing one or some typical testors, the feature selection is made.

So far, a typical testor and the feature's weight were defined assuming that a classified sample of objects described in terms of *R* is given, but for prediction problems where there are not classes, we do not have a classified sample instead of this we have an associated value to each object. This associated value is what we want to predict for a new object. This prediction is made from the associated values of the sample objects following the next idea: If two objects have similar descriptions, they must have similar associated values.

Following this idea we introduce the next testor definition:

Definition 3. A feature subset T is a testor if and only if when all features are eliminated, except those from T, there is not any pair of similar subdescriptions with non similar associated values.

Since a testor is a feature subset such that the sample does not have any pair of similar subdescriptions with non similar associated values, so testors fulfill that if two sample objects have similar descriptions, then they have similar associated values.

Consequently, the definition of typical testor remains without change. And all expressions can be applied to compute the feature's weight.

In order to carry out the feature selection a weight for each typical testor is evaluated as follows:

$$P(T) = \frac{\sum_{x_i \in T} P(x_i)}{|T|}$$

Finally, choosing the best typical testors regarding to the typical testors weight the feature selection is made. For the case of stellar parameters estimation, in the experimental results section, the function used to decide when two descriptions and two associated values are similar will be explained.

3 K-NN Classifier

The KNN algorithm (*K*-nearest neighbors) is the simplest and most commonly used classifier (see Mitchell, 1976). In the case of stellar parameter estimation, the nearest neighbors are obtained using the standard Euclidean distance with $R^s \subset R$, the obtained selection.

$$d(O_i, O_j) = \sqrt{\sum_{r=1}^{s} (x_r(O_i) - x_r(O_j))^2}$$

The algorithm stores the training data and in order to classify a new instance O_q, KNN finds the training examples O_1, $O_2,...,O_k$ that are most similar to O_q and KNN returns the most common class among the neighbors of O_q.

When it is necessary to approximate a continuous function, it can be computed as:

$$f(O_q) = \frac{\sum_{i=1}^{K} f(O_i)}{K}$$

where $f(O_i)$ are the associated values for the *K* nearest neighbors.

An extension consists in assigning a weight w_i to each nearest neighbor based on his distance, with the instance to classify (thus neighbors that are closer to the instance to classify are weighted more heavily), if w_i is calculated as:

$$w_i = \frac{1}{d(O_q, O_i)^2}$$

the chosen class will be that with greatest voting between the *K* nearest neighbors, here the vote of each O_i is w_i for the class where it appears and 0 for the other classes.

In the case of a real-valued target function, the value of the function is calculated as:

$$f(O_q) = \frac{\sum_{i=1}^{K} w_i f(O_i)}{\sum_{i=1}^{K} w_i}$$

4 Experimental Results

After selecting the databases to use, the typical testors are computed for each one of these. The typical testors' weights are calculated and those testors with highest weight are chosen as feature selection. Then each typical testor is given, as selection, to the KNN classifier; afterwards the classification (or predictions) and the classification error index is evaluated.

In the figure 1 it is shown the process of feature selection used in this paper.
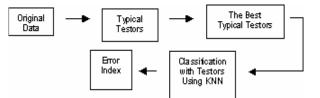

Fig 1. Process of feature selection using typical testors and classification using KNN

In all experiments, the KNN classifier with *K*=3, and the *K*-Fold Cross Validation with *K*=10 was used. In addition, the average of five tests with the classification algorithm is shown.

5 Experiments in Classification Problems

In this section, we present the tests carried out with some databases of supervised classification taken from http://www-old.ics.uci.edu/pub/machine-learning-databases UC-IRVINE repository. The databases used were Monks and Iris, the information about these databases is shown in the table 1.

The Monk databases were provided by Sebastian Thrun of the School of Computer Science, Carnegie Mellon University at October 1992. There are three Monk's problems. The domains for all Monk's problems are the same. We selected the Monk-3 database; this database has 5% of noise.

The Iris database was created by R. A Fisher, The data set contains three classes with 50 examples each one, where each class refers to a type of iris plant. One class is linearly separable from the other two. The attribute to predict is the class of iris plant.

Data	Values	Objects	Features	Classes
Monk	Integer	122	6	2
Iris	Real	150	4	3

Table 1. UC-IRVINE databases used for experimentation

For Monk-3 database the comparison criterion used to compare features' values was the equality.

In the case of Iris database, we used two different comparison criteria, for this reason, two different experimentations were done (called MBAlba and MBDS respectively). In the MBAlba case, the comparison criteria used for x_1 and x_2 considers that two values (of x_1 or x_2) are similar if the absolute variation is less or equal than 0.1. In addition, for x_3 and x_4 two values are similar if the absolute variation is less or equal than 0.25.

Data	Number of testors	Size of testors
Monk-3	1	3
Iris (MBAlba)	1	3
Iris (MBDS)	1	3

Table 2. Size and number of typical testors obtained for monk-3 and iris databases

For MBDS experimentation, the standard deviation was used to determine if two feature's values are similar. We go on selecting a feature and calculating the standard deviation for all the objects in the database. Then two values for a feature will be similar if the absolute variation is less or equal than the standard deviation for this feature. The standard deviations for the four features on Iris database are: $\sigma_1=0.8281$, $\sigma_2=0.4336$, $\sigma_3=0.7644$, and $\sigma_4=0.763$. The size and number of typical testors for Iris and Monk-3 databases experimentation are shown in the table 2.

After compute the typical testors, the testors' weights are calculated, however, in this experimentation for both Monk-3 and Iris only one typical testor was found so the weight for both typical testors is 1. In the table 3, the results obtained by KNN using each testor as selection, against the whole data (without selection) and the selection proposed by Ramirez (2001) are shown. The best results in the table are emphasized for each class.

Data	% error		
	Original data	T_1	Ramirez's method
Monk-3	13.11	<u>10.49</u>	13.93
Iris (MBAlba)	<u>3.86</u>	4.26	5.33
Iris (MBDS)	<u>3.86</u>	4.79	5.33

Table 3. Error indices for the classification of monk-3 and iris databases

6 Experiments in Stellar Parameters Estimation

The astronomical data were produced by Jones (1996) in a homogeneous catalog of 48 spectral indices for 684 stars observed at Kitt Peak National Observatory with the coudé feed instrument. The spectral indices were measured from the spectra in the wavelength regions 3820-4500A° and 4780-5450 A° by following the definition of the Lick indices (see, Worthey et al., 1994), the Rose (1994) indices and new Lick-type Balmer indices (see, Jones et al., 1995). In our experimentation, we used the indices in conjunction with physical atmospheric parameters given in the catalog. The spectral indices are widely used for interpretation of observational and physical properties of stars.

The problem consists in predicting the values for these physical atmospheric parameters from the spectral indices, so this is a prediction problem.

Some of the parameters in the catalog are missing, and then 42 stars were eliminated, so in this experimentation, for the prediction of atmospheric parameters, only 642 stars from the catalog were used. In the table 4, the characteristics of the astronomical data are shown.

Data	Values	Objects	Features
Astronomical	Real	642	48

Table 4. Properties of astronomical data

For the astronomical data we predicted three physical atmospheric parameters: effective temperature p1, surface gravity p2 and metallicity (p3).

The comparison criteria between features' values (spectral indices), for each index i, i=1,...,48, are as follows:

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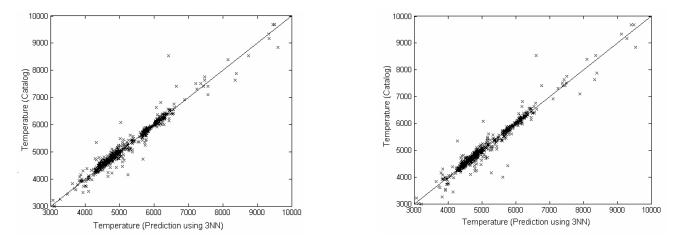

Fig. 2. Prediction of Temperature parameter made with the Jone's catalog against the prediction made with 3NN using (left side) the original set of data (48 spectral indices) and (right side) the set of data corresponding to the testor T_1 (17 spectral indices)
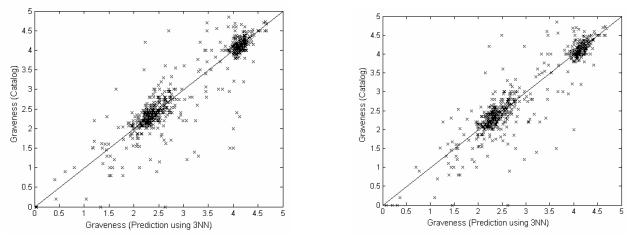
Fig. 3. Prediction of Gravity parameter made with the Jone's catalog against the prediction made with 3NN using (left side) the original set of data (48 spectral indices) and (right side) the set of data corresponding to the testor T_1 (27 spectral indices)
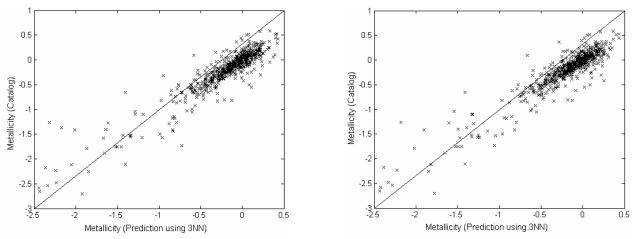

Fig. 4. Prediction of Metallicity parameter made with the Jone's catalog against the prediction made with 3NN using (left side) the original set of data (48 spectral indices) and (right side) the set of data corresponding to the testor T2 (29 spectral indices)

Where σ_i is the standard deviation in the sample for the respective index.

The typical testors were computed using the testor concept given in definition 3, the similarity function used to determine when two subdescriptions are similar is the total coincidence of the values in the subdescription. In order to determine if two associated values are similar the following expression was defined:

$$\left| p_j(O_k) - p_j(O_l) \right| \leq \sigma_j \qquad j=1,2,3$$

Where σ_j is the standard deviation for the parameter p_j in the sample. Since we have three parameters, then three different sets of typical testors were computed. In the table 5, the number of typical testors, their size and the amount of features selected by Ramirez's method are shown.

Parameter	Testors	Size	Feature in Ramirez (2001)
Temperature	1	17	8
Graveness	4	27	12
Metallicity	2	29	20

Table 5. Number and size of the typical testors obtained for astronomical data

20

For temperature parameter only one testor was found, it has 17 features, for gravity parameter 4 testor were found, each one with 27 features, and for metallicity parameter 2 testor were found both with 29 features.

Since one typical testor was found for temperature parameter, we only calculated the testors' weights for the testors of gravity and metallicity parameters.

The weight for the four testors of the gravity parameter is the same; the reason is that these testors only are different in the last two features. Therefore, each testor has a weight of 0.9629. The same situation occurs with the weights for the two testors of metallicity parameter, in this case these testors only are different in the last feature; the weight for these testors is 0.9827.

Parameter: Temperature rms error		
Original Data	T_1	Ramirez's method
182.80	106.26	124.7

Table 6. Error indices for temperature parameter.

Parameter: Gravity rms error					
Original data	T_1	T_2	T_3	T_4	Ramirez' method
0.4030	0.2180	0.2230	0.2184	0.2181	0.2600

Table 7. Error indices for gravity parameter.

Parameter: Metallicity rms error			
Original data	T_1	T_2	Ramirez' method
0.1770	0.1175	0.1173	0.1360

Table 8. Error Indices for metallicity parameter.

The tables 6, 7 and 8 show the root mean squared error (rms) indices after application of 3NN; the lowest errors are highlighted in bold. The figures 2, 3 y 4 shown the Jones's prediction against 3NN prediction considering the whole sample (without selection) and the prediction made with the selection using typical testors respectively.

7 Conclusions

We can see that in the experimentation with Monk-3 database the feature selection reduces about 50% the number of features. In addition, with the selection made with our method, a better classification than both using the whole feature set and using Ramirez (2001) selection was obtained.

In the two experimentations with Iris database, the feature selection reduces about 25% the number of features. The classification error index becomes worse in both experimentation using typical testors than using the whole feature set. Here we must highlight the increase of the classification error index when was used the deviation standard as threshold in the comparison criteria. However, we get a better classification error index than that one reported by Ramirez's method.

In general, we can see that the classification through typical testors is similar to the classification using the whole data set and in all the tests; we get a better classification than using the method reported by Ramirez (2001).

In the case of astronomical data, the feature selection reduces about 41% the number of features (from 48 to 27 or 29 features). The error indices for all three parameters are lesser than both using the original data set. and those reported by Ramirez (2001) method.

The feature selection through typical testors has been not enough spread, however we can see that the reduction in the number of features is good and the quality of classification with the selection made with typical testors is better or in some cases very similar to the classification using the whole data set.

On the other hand, the amount of features obtained by Ramirez's method is lesser than the amount obtained by typical testors; however, in the classification stage or parameters estimation the error indices using Ramirez's method are greater than using typical testors.
  Future work will attempt to apply other recent testor concepts in order to try getting lower error indices than those reported in this paper. The new concepts take account different comparison functions to establish the similarity of objects in real problems. Some examples of these concepts are Goldman (see, Alba-Cabrera, 1997) and FS Testors (see, Lazo-Cortes, 2001).

The Φ-testor is the most recent definition proposed for fuzzy environments, however there is not an algorithm to compute Φ-testors, as future work and as part of the development of Testor Theory and their application to practical problems we will develop an algorithm to compute Φ-testors.

The feature selection through typical testors has the advantage that a detailed analysis is done, but the disadvantage is that a high computational time consuming to compute all the typical testors is needed. However, when the typical testors have been computed we do not need to compute them again, therefore this feature selection can be used to classify any amount of new objects.

Acknowledgement

References

1. **Alba Cabrera E.,** "New extensions of testor concept for different types of similarity functions" Ph. D. Thesis, ICIMAF, Cuba, 1997. (In Spanish).
2. **Dmitriev A. N., Zhuravlev Y.I. and Krendeliev F.P.,** "About mathematical principles of objects and phenomena classification", *Diskretni Analiz* 7 (1966), (In Russian).
3. **Jones L. A.,** "Star populations in galaxies", Ph. D. Thesis, University of North Carolina, Chapel Hill, North Carolina, 1996.
4. **Jones L.A. and Worthey G.,** "New age indicators for old stellar populations", *Astrophys. J. Lett.*, 446(1), (1995) L31-L34.
5. **Lazo-Cortes M., Ruiz-Shulcloper J. and Alba-Cabrera E.,** "An overview of the evolution of the concept of testor", *Pattern Recognition,* 34(4), (2001), 753-762.
6. *Martínez Trinidad J. F., and Guzmán-Arenas A.,* "The logical combinatorial approach to pattern recognition an overview through selected works", *Pattern Recognition* 34(4), (2001), 741-751.
7. **Mitchell T. M.,** Machine Learning, Carnegie Mellon University, USA, 1976.
8. **Ramirez J. F., Fuentes O., and Gulati R.K.,** "Prediction of stellar atmospheric parameters using instance-based learning and genetic algorithms", Experimental Astronomy, 12 (3) 163-178, 2001.
9. **Rose J. A.,** "The integrated spectra of M32 and of 47 Tuc: a comparative study at high spectral resolution", *Astronomy J.,* 107(1), (1994) 206-229.
10. **Ruiz Shulcloper J., Guzmán Arenas A., and Martínez Trinidad J. F.,** *Logical Combinatorial Approach to Pattern Recognition. Feature Selection and Supervised Classification*, IPN, Mexico, 1999.(In Spanish).
11. *Worthey G., Faber S. M., Gonzalez J. J., and Burstein D.,* "Old stellar populations. V. Absorption feature indices for the complete Lick/IDS sample of stars", *Astrophys. J., Suppl. Ser* 94(2), (1994) 687-722.

José Ángel Santos Gordillo. *He Received the B.S. Degree in Computer System Engineering from Instituto Tecnológico de Tuxtla Gutiérrez, Chiapas, México, en 2002, the M. Sc. Degree in Computer Science from Instituto Nacional de Astrofísica Óptica y Electrónica (INAOE), México, en 2003. His current research interests include Logical Combinatorial Patter Recognition, Testor Theory, and Feature Selection.*

Jesús Ariel Carrasco Ochoa. *He was born in Mexico in 1965. He received his PhD degree in Computer Science form the Center for Computing Research of the National Polytechnic Institute (CIC-IPN), Mexico, in 2001. He works as associated researcher at the National Institute for Astrophysics, Optics and Electronics of Mexico. His current research interests include Sensitivity Analysis, Logical Combinatorial Patter Recognition, Testor Theory, Feature Selection and Clustering.*

José Francisco Martínez Trinidad. *Received the B.S. Degree in Computer Science from Benemérita Universidad Autónoma de Puebla, México in 1995, the M. Sc. Degree in Computer Science from Benemérita Universidad Autónoma de Puebla, México in 1997 and the Ph. D. Degree in Computer Science from Centro de Investigación en Computación del IPN, México in 2000. Currently is working at INAOE, México, where he is a full time researcher. His research interests include Pattern Recognition, Conceptual Clustering, Text Analysis and Intelligent Data Analysis.*