

# Two Floating Search Strategies to Compute the Support Sets System for ALVOT

Erika López-Espinoza, Jesús Ariel Carrasco-Ochoa, and José Fco. Martínez-Trinidad

Computer Science Department

National Institute of Astrophysics, Optics and Electronics

Luis Enrique Erro No. 1 Sta. María Tonantzintla, Puebla, CP 72840, México  
{danae, ariel, fmartine}@inaoep.mx

**Abstract.** In this paper, two strategies to compute the support sets system for the supervised classifier ALVOT (voting algorithms) using sequential floating selection are presented. ALVOT is a supervised classification model based on the partial precedence principle, therefore, it needs, as feature selection, a set of features subsets, this set is called support sets system. The sequential floating selection methods for feature selection find only one relevant features subset. The introduced strategies search for a set of features subsets to generate a support sets system. Both strategies are compared between them and against the feature selection method based on testor theory, which is commonly used to compute this system. Results obtained with both strategies on different databases from UCI and on the faces database from Olivetti Research Laboratory (ORL) in Cambridge are presented.

## 1 Introduction

The feature selection problem traditionally consists in searching for one subset of features which improves or maintains the efficiency of the classifier. In the case of classification methods based on partial precedence, comparisons among the objects are made through subdescriptions previously specified. These methods do not take account whole comparisons between object descriptions. Partial conclusions about the similarity among the objects are considered to reach a final conclusion. Therefore a Support Sets System (SSS) is required.

ALVOT [1] is a supervised classification model developed in the framework of the Logical Combinatorial Pattern Recognition (LCPR) [1]. This classifier is based on the partial precedence principle, therefore it requires a SSS that indicates the subdescriptions that will be considered to make partial comparisons. An important characteristic of this classifier is that it can work with descriptions in terms of quantitative and/or qualitative variables, admitting incomplete descriptions [1]. The feature selection into the LCPR is commonly carried out by mean of the testor theory [1, 3]. A subset of typical testors can be used to form the SSS, nevertheless, the algorithms to compute the set of all typical testors are of exponential complexity with regard to the number of features, thus the computation of all typical testors, in problems where the objects are described in terms of high dimensionality  $n$ -uples, becomes computationally expensive or unfeasible.

In this work, two strategies to select the SSS for ALVOT using Sequential Floating Selection (SFS) [2] are proposed. In order to evaluate the performance of the subsets of selected features, the ALVOT classification rate is used, in addition it allows us to work with mixed and incomplete data. Experiments with different databases using the proposed strategies will be shown.

## 2 Foundations

### 2.1 Sequential Floating Selection

The SFS algorithms are part of the feature selection methods that use the Wrapper strategy. In this strategy the selection is made using a classifier to evaluate the efficiency of the subsets of selected features ( $J$ ).

There are two ways to do a floating search, the Sequential Floating Forward Selection (SFFS) and the Sequential Floating Backward Selection (SBFS) [2].

The idea behind of SFFS consists in to initiate with an empty set of features, and to make the best inclusion of the feature that maximizes the classifier's efficiency, and after to make features exclusions as long as the resulting subset improves the classification efficiency compared to the subset obtained in the previous step. The algorithm finishes when it finds the subset with the desired cardinality.

In SBFS the idea is the same, but instead of initiating with an empty set of features it initiates with the whole set and exclusions followed by inclusions are done.

The algorithm of SFFS [2] is as follows.

1. *Initialization:*  $Y = \emptyset$ ,  $k = 0$

Stop when  $k$  equals the number of required features.  
(in practice it is possible begin with  $k=2$  applying two inclusions)

2. *Inclusion:* Select the most significant feature.

$$x^+ = \arg \max_{x \in X \setminus Y_k} [J(Y_k \cup \{x\})]$$

$$Y_{k+1} = Y_k \cup \{x^+\}; \quad k = k + 1$$

3. *Exclusion:* Select the least significant feature.

$$x^- = \arg \max_{x \in Y_k} [J(Y_k \setminus \{x\})]$$

If  $J(Y_k \setminus \{x^-\}) > J(Y_k)$  then

$$Y_{k-1} = Y_k \setminus \{x^-\}; \quad k = k - 1$$

go to step 3

else

go to step 2.

The SBFS algorithm can be obtained from SFFS by starting with the whole features set and substituting inclusion by exclusion and exclusion by inclusion.

### 2.2 ALVOT

ALVOT [1] is an algorithm developed into LCPR [1] to do supervised classification based on the *partial precedence principle*.

The classification is made through six stages:

1. Definition of support sets system.
2. Definition of similarity function.
3. Object evaluation function for a fix support set.
4. Class evaluation function for a fix support set.
5. Class evaluation function for the whole support sets system.
6. Solution rule.

In the first stage the SSS is defined. The SSS can be understood as any set of features subsets that indicates which parts of the objects will be compared in the classification stage. Each subset is a support set.

The similarity function is defined in the second stage. This function determines how subdescriptions will be compared and it should reflect how the comparisons among the objects are made in the real world.

Object evaluation function for a fix support set determines how much information is given by the similarity of a new object with each one of the sample objects, for a fix support set. The result of this function is called the vote given by each sample object to a new object with regard to a fix support set.

Class evaluation function for a fix support set summarizes all object evaluations for a new object within each class, for a fix support set. The result of this function is called the vote given by each class to a new object with regard to a fix support set.

Class evaluation function for the whole support sets system summarizes all class evaluations for a new object for whole the support sets system. The result of this function is called the vote given by each class to a new object with regard to the whole support sets system.

Finally the solution rule is applied, this function takes all votes for each class and decides which class or classes the new object belongs to.

### 2.3 Typical Testors

Into LCPR, feature selection is commonly carried out by mean of the testor theory [1, 3]. Typical testors or a subset of them can be used as SSS for the ALVOT classifier.

A features subset  $T$  is a testor if and only if considering the features from  $T$  there are no similar subdescriptions among objects from different classes, i.e., objects from class  $i$  are not similar to objects from class  $j$ , where  $i \neq j$ .

A testor  $T$  is a typical testor (irreducible) if and only if eliminating any feature from  $T$  the resultant set is not a testor. It means that there is not any other testor  $T'$  such that  $T' \subset T$ .

Typical testors are irreducible features combinations which allow distinguishing objects from different classes. We may think that if a feature appears in many irreducible combinations or typical testors, it will be more indispensable to distinguish classes. Based on this idea Zhuravlev formulated his feature's weight definition as the relative frequency of the occurrence of each feature in the set of all typical testors.

Let  $\tau$  be the number of typical testors in a sample and let  $\pi(i)$  be the number of typical testors where the feature  $x_i$  appears. Then the feature's weight (relevance) of  $x_i$  is as follows:

$$P(x_i) = \frac{\pi(i)}{\tau}, i = 1, \dots, n. \quad (1)$$

## 3 Proposed Strategies

As we have seen ALVOT requires a SSS to make the comparisons among the objects that will be classified.

Now we introduce two new strategies based on SFS to compute SSS for ALVOT. We only expose the algorithms for SFFS; the modifications apply also for SBFS.

### 3.1 Strategy 1

In the first strategy, the subset with the best classification efficiency for each cardinality  $i$  is searched (where  $i=1, \dots, \text{number\_features}$ ), i.e., the subset of cardinality 1 with the best classification efficiency is searched, after the best subset of cardinality 2 is searched, this process continues until the best subset of cardinality equal to the total amount of features ( $N$ ) is found, finally the best  $p$  subsets are selected as a SSS.

In this strategy the SSS may have at most one subset of cardinality  $i$  for each  $i$ .

The modified algorithm with the first strategy is as follows:

1. Initialization  $Y = \phi$ ,  $k = 0$ ,  $S = \phi$ ,  $N$ ;  
 $p$  = number of subsets to search
2. For  $i = 1$  to  $N$ 
  - 2.1. While  $Y$  don't have the cardinality  $i$ 

Inclusion:  $x^+ = \arg \max_{x \in X \setminus Y_k} [J(Y_k \cup \{x\})]$

$Y_{k+1} = Y_k \cup \{x^+\}$ ;  $k = k + 1$

Exclusion:  $x^- = \arg \max_{x \in Y_k} [J(Y_k \setminus \{x\})]$

If  $J(Y_k \setminus \{x^-\}) > J(Y_k)$  then

$Y_{k-1} = Y_k \setminus \{x^-\}$ ;  $k = k - 1$

go to step Exclusion

else

go to step Inclusion
  - 2.2.  $S = \text{Select\_the\_best\_}p\text{-subsets}(S \cup Y_i)$

### 3.2 Strategy 2

The second strategy generates a SSS selecting the best  $p$  subsets from whole evaluated subsets during the entire floating search. This strategy admits a SSS with more than one subset with the same cardinality, while in the first strategy only one subset of cardinality  $i$  is permitted for each  $i$ . The maximum cardinality of SSS in the first strategy is limited to the number of features  $n$ , while in this strategy the SSS can be  $\gg n$ .

The modified algorithm with the second strategy is as follows:

1. Initialization,  $Y = \phi$ ,  $k = 0$ ,  $S = \phi$ ;  
 Stop when  $k$  equals the number of required features.  
 (in practice it is possible to begin with  $k=2$  applying two inclusions)  
 $p$  = number of subsets to search
2. Inclusion: Select the most significant feature.
 

$x^+ = \arg \max_{x \in X \setminus Y_k} [J(Y_k \cup \{x\})]$ 

$S = \text{Select\_the\_best\_}p\text{-subsets}(S \cup (Y_k \cup \{x\}))$  para  $x \in X \setminus Y_i$

$Y_{k+1} = Y_k \cup \{x^+\}$ ;  $k = k + 1$
3. Exclusion: Select the least significant feature.
 

$x^- = \arg \max_{x \in Y_k} [J(Y_k \setminus \{x\})]$ 

$S = \text{Select\_the\_best\_}p\text{-subsets}(S \cup (Y_k \setminus \{x\}))$  where  $x \in X \setminus Y_i$

If  $J(Y_k \setminus \{x^-\}) > J(Y_k)$  then

```


$$Y_{k-1} = Y_k \setminus \{x^-\}; \quad k = k-1$$

    go to step 3
else
    go to step 2

```

## 4 Experimental Results

In this section, some experiments using both strategies are presented. The experiments were carried out on databases from [4] and on the faces database from [5].

ALVOT was used to measure the efficiency of selected subsets. In the experiments the test sample is equal to the training sample.

In the experiments the typical testors were selected based on their relevance using the equation (1). The relevance of typical testor is computed with equation (2):

$$R(\tau_j) = \frac{\sum_{x_i \in \tau_j} P(x_i)}{|\tau_j|}. \quad (2)$$

where  $\tau_j$  is a typical testor,  $|\tau_j|$  is the cardinality of this testor and  $P(x_i)$  is the feature's weight of  $x_i$  see (1).

The first experiment was carried out with the ZOO database. This database contains 101 animals grouped in 7 classes, each description is given in terms of 16 features, where 15 are Boolean and 1 is nominal. The best classification efficiency is reached using the strategy 2 with SBFS, obtaining a 100% recognition rate with SSS of 14 subsets. On the other hand, in this sample there are 32 typical testors, choosing support sets systems with 4 typical testors or more the classification efficiency is maintained. The results are shown in the figure 1.

The second experiment was carried out with the HEPATITIS database. This database has 155 objects in 2 classes described through 19 features, where 6 are numeric and 13 are Boolean. This database has incomplete descriptions (missing data). The best classification efficiency was obtained using the strategy 2 with SFFS with a SSS of 12 subsets. The number of computed typical testors was 35. The best classification efficiency was obtained using 10 testors as SSS (figure 2).

The third experiment was carried out with the FLAGS database. This database contains 193 objects in 8 classes and described with 28 features, 15 are nominal, 2 are numeric and 11 are Boolean. In this sample 1469 typical testors were found and the best classification efficiency was obtained with a SSS of 3 testors. The best classification efficiency was obtained using the strategy 2 with SFFS and a SSS of 12 subsets (figure 3).

The fourth experiment was carried out with the SPECTF Heart Data database. This database contains 80 objects grouped in 2 classes described with 22 Boolean features. 26 typical testors were found and the best classification efficiency was obtained with only one testor as SSS. The best classification efficiency is reached using the strategy 2 with SFFS and a SSS of 4 subsets (figure 4).

The last experiment was carried out with the faces database from Olivetti Research Laboratory (ORL) in Cambridge [5]. This database contains 10 different images of 40 subjects and the size of each image is 92x112. The figure 6 shows images of two subjects from ORL database.

An image may be considered as a vector of dimension  $N^2$  (size of the image), in this case a vector of dimension 10304 is equivalent to a point in a 10304-dimensional space. To work with the set of vectors from all the images is unfeasible, due to this different methods are used to reduce the dimension of an image. In this work the Principal Component Analysis (PCA) method is used [6-8]. Once the principal components of the images set are obtained, the proposed strategies are applied to obtain the SSS for ALVOT.

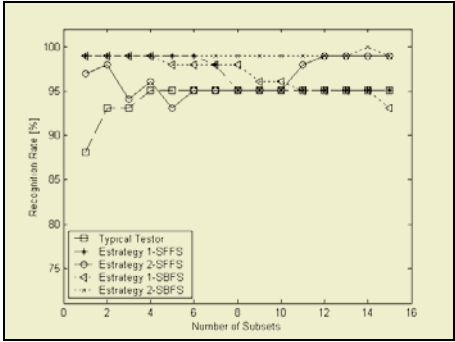


Fig. 1. Floating Methods and Typical Testors for ZOO.

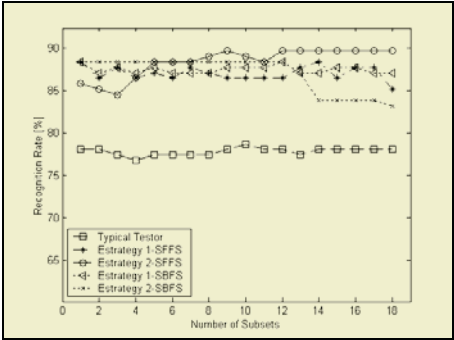


Fig. 2. Floating Methods and Typical Testors for HEPATITIS.

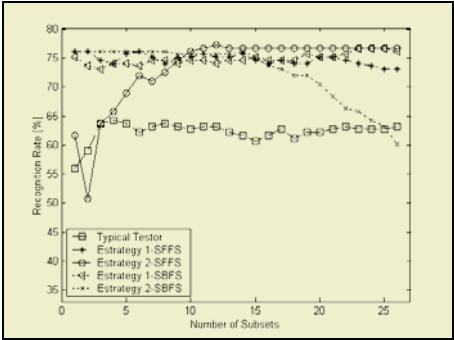
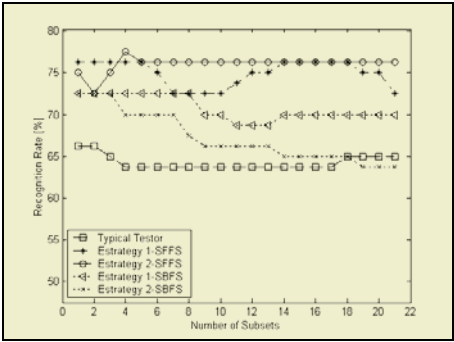
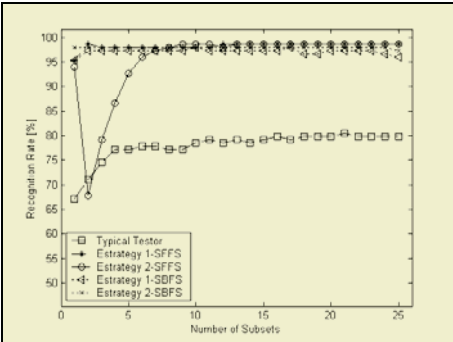


Fig. 3. Floating Methods and Typical Testors for FLAGS.

In this experimentation 150 images of 15 subjects were used from ORL database. Once the principal components were computed 26 eigenvectors which capture the 80.366% of the information were considered. In this case all features are numeric. Taking account the 26 eigenvectors 145951 typical testors were found and the best classification efficiency was obtained using a system with 21 typical testors. Both forward strategies obtained the best efficiency, nevertheless, with the second strategy this efficiency may be obtained using a SSS of cardinality smaller than the first. The classification efficiency is shown in the figure 5.



**Fig. 4.** Floating Methods and Typical Testors for SPECTF.



**Fig. 5.** Floating Methods and Typical Testors for ORL.



**Fig. 6.** Examples of subjects that exists in the ORL database.

5 Conclusions

In this paper, two strategies to compute the support sets system for the supervised classifier ALVOT using sequential floating selection were presented. In the experiments the two proposed strategies to compute the support sets system have better performance than using typical testors. The best classification efficiency for ALVOT may be obtained computing the SSS with the second strategy. Using our methods it is possible to find a SSS with lesser cardinality which increases the classification efficiency of ALVOT.

Another point to highlight is the fact that the number of typical testors is bounded exponentially, so that in some cases they can be too many to be useful in the classification stage. Contrarily, in the proposed method the size of the support sets system is one of the parameters, which allows fixing the size of the system according to the practical requirements.

Generally in the floating methods, the classifier used to evaluate the classification efficiency of selected subsets only works with numeric information. In this work, ALVOT is used to evaluate the classification efficiency of subsets, because ALVOT allowed us to carry out experiments with mixed and incomplete data.

Future work includes determining the optimum value of  $p$  (SSS's cardinality) to find the SSS that maximizes the classification efficiency of ALVOT.

## Acknowledgement

This work was financially supported by CONACyT (México) through the project J38707-A.

## References

1. J. Ruiz Shulcloper, Adolfo Guzmán Arenas y J. Fco. Martínez Trinidad, "Enfoque Lógico Combinatorio al Reconocimiento de Patrones", Editorial Politécnica, ISBN: 970-18-2384-1, 1999.
2. P.Pudil, F.J.Ferri, J.Novovièová, and J.Kittler. "Floating Search Methods for Feature Selection with Nonmonotonic Criterion Functions", In: Proceedings of the 12th International Conference on Pattern Recognition, Jerusalem, Israel, 1994. Los Alamitos, IEEE Computer Society Press, 279-283, 1994.
3. M. Lazo-Cortes, J. Ruiz-Shulcloper and E. Alba-Cabrera, "An overview of the evolution of the concept of testor", Pattern Recognition, 34(4), 2001, 753-762.
4. Machine Learning Databases, University of California, Irvne, Department of Information & Computation Science, <http://ftp.ics.uci.edu/pub/machine-learning-databases/>
5. A. L. Cambridge, Olivetti Research Laboratory face database <http://www.uk.research.att.com/facedatabase.html>
6. L. Sirovich and M. Kirby, "Low-Dimensional Procedure for the Characterization of Human Faces", Journal of the Optical Society of America A., Vol. 4, pp 519, 1987.
7. M. Kirby and L. Sirovich, "Application of the Karhunen-Loève Procedure for the Characterization of Human Faces", IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1), 1990.
8. M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience. V. 3, pp. 71-86, 1991.