Abstract: This paper proposes to use a fuzzy entropy based feature selection approach to reduce the dimensionality of DAIS hyperspectral data. To compare its performance, three other filter based feature selection approaches were used. A support vector machine was used as a classification algorithm. In order to compare various feature selection approaches with full dataset, McNemar’s test based test for non-inferiority was used. The results of the non-inferiority testing between total features and selected features by different feature selection approaches suggests the suitability of fuzzy entropy based approach in terms of classification accuracy and small number of features used to derive same level of accuracy as achieved by the use of total number of features. Results suggest inferior performance by Relief based feature selection approach with the used dataset. The classification accuracy achieved by entropy and signal to noise ration based feature selection approach was comparable to that of fuzzy entropy based approach but requires more number of selected features.

Introduction

Recent development in remote sensing systems allows measurement of radiation in hundreds of spectral bands. The increased dimensionality of such hyperspectral data provides a challenge to current classification techniques. As the number of spectral bands increases, the capability to detect more detailed classes should also increase. Further, with the increase of the number of features, it is expected that the classification accuracy will also increase. Usually the number of training samples is limited. It has been observed frequently in practice that beyond a certain point, if the number of training samples per feature is small, the addition of more dimensions leads to a worse performance in terms of classification accuracy. Hughes (1968) suggested that the basic source of the problem is the limited number of training samples. The problem becomes more serious in high dimensional cases.

In order to avoid the negative impact of high dimensionality on classifier performance some form of feature selection may be used. Feature selection is an important phase for a classification system, as the selection of feature subset is found to affects the classification results as well as providing other benefits in terms of reducing the data storage requirements and training time as well as defying the curse of dimensionality (Liu and Motoda, 1998). Feature selection methods are dependent on the properties of the input data and the classifier used. Feature selection require that a criterion be defined by which it is possible to judge the quality of each feature. A computational method is then required to search through a subset of features consisting of “best” features based upon the pre-defined criterion. The search procedure used in selecting a subset of feature may include an exhaustive search or a non-exhaustive search. In general, exhaustive search approach guarantee to find the optimal subset but in a practical
application, due to unreasonably large computational requirements, a non-exhaustive search approach is used. A number of studies on feature selection methods are reported in image classification (Jain and Zongker, 1997; Kavzoglu, 2001; Serpico and Bruzzone, 2001).

In general, feature selection methods can be divided in three different categories i.e. the filter model, wrapper model and embedded models (Guyon and Elisseeff, 2003; Pal and Foody, 2010). In the wrapper model (Kohavi and John, 1997), a feature selection method uses a classification algorithm and selects a subset of features providing best classification performance in comparison to full dataset. A major problem of wrapper approaches is their large computational requirements. Embedded methods produce a ranking of all features during classification process, which can later be used for feature selection. The filter approach is generally computationally more efficient as no classifier is employed to evaluate the performance of subset of features. Based on computational efficiency of the filter based feature selection approaches, this paper propose to use a fuzzy entropy based feature selection (Hu and Yu, 2005) for DAIS hyperspectral dataset. To compare its performance in terms of classification accuracy and the number of features used to achieve the same level of accuracy as that by full dataset, three other filter approaches viz; entropy, Relief and signal to noise ratio based approaches was used. A support vector machine was used to classify the full and reduce dataset in order to compare the performance of various feature selection approaches.

**Entropy and Fuzzy entropy approach for feature selection**

For a finite set \( X = \{x_1, x_2, \ldots, x_n\} \), \( P \) is the probability distribution on \( X \) and \( R \) is a fuzzy indiscernibility on \( X \) (Hu and Yu, 2005). Yager’s entropy is defined by:

\[
H_Y(R) = \sum_i p(x_i) \log_2 p(x_i)
\]

where \( p(x_i) \) represents probability distribution of \( x_i \).

A fuzzy entropy based feature selection approach (Hu and Yu, 2005; Hu et al. 2006) was used to reduce the dimensionality of hyperspectral data. A brief description of the proposed approach is provided below:

For a given fuzzy information system defined by \((U, A, V, f)\), where \( U \) is a finite set of objects (Hu and Yu, 2005), \( A \) is the attributes set (features or bands with hyperspectral data), i.e. \( A = \{a_1, a_2, \ldots, a_n\} \) or \( A = C \cup D \) and \( C \cap D = \emptyset \), where \( C \) is the condition attributes and \( D \) is the decision attributes. \( V_a \) is the domain of attribute \( a \). For each \( a \in A \), an information function is defined as \( f_a : U \to V_a \). If \( Q \) is a subset of attribute set \( A \), where \( Q = \{a_1, a_2, \ldots, a_k\} \) and \( F_q \) is the fuzzy relation matrix by an indiscernibility relation, \( a \in Q \). Then the significance of \( a \) is defined as

\[
\text{Significance} \ (a \mid Q-a) = H(a \mid Q-a) = H(Q) - H(Q-a)
\]

If Significance \( (a \mid Q-a) = 0 \) attribute \( a \) is considered redundant otherwise this attribute is indispensable. If \( (\forall a \in Q) : a \ is \text{indispensable} \), one can say that \( Q \) is independent. For further details about this algorithm, readers are referred to Hu and Yu (2005).
Relief algorithm

The general idea of Relief is to choose the features that can be most distinguished between classes. These are known as the relevant features. At each step of an iterative process, an instance is chosen at random from the dataset and the weight for each feature is updated according to the distance of this instance to its Near-miss and Near-hit (Kira and Rendell, 1992). Finally, this algorithm selects a subset of features whose average weight is above a given threshold. For a classification problem, a dataset \((X)\) is represented by a vector consisting of \(f\) features. An instance from the dataset will be a near-hit to \(X\), if it belongs to the close neighbourhood of \(X\) and belongs to the same class as that of \(X\). On the other hand, an instance would be called a near-miss if it belongs to the neighbourhood of \(X\) but not to the same class as that of \(X\). Relief uses a \(f\)-dimensional Euclidean distance to select near-miss and near-hit.

Signal to noise ratio approach

Signal to noise ratio based feature selection approach is a statistical method that measures effectiveness of a feature in identifying a class out of another class. Signal to noise ratio statistics is calculated by:

\[
\left(\frac{\text{mean}_{\text{class1}} - \text{mean}_{\text{class2}}}{\text{stddev}_{\text{class1}} + \text{stddev}_{\text{class2}}}\right)
\]

This approach rank all features in order to define how well a feature discriminates between two classes. In order to use this approach for multiclass classification problem, one against one approach, was used in present study.

Dataset and methodology

A Digital Airborne Imaging Spectrometer (DAIS) image acquired on 29th June 2000, of the La-Mancha area, lying to the south of Madrid, Spain, was used in present study. Only 65 of the available 72 spectral features in the optical region and area of interest comprising of 512 columns by 512 rows were used. Seven of the DAIS features suffered from severe striping problems were rejected (Pal and Foody, 2010). Due to the lack of field data for the year 2000, a reference map was generated after a field visit on 30th June 2001, which included consultations with the local farmers. Eight land cover classes (wheat, water, dry salt lake, hydrophytic vegetation, vineyards, bare soil, pasture lands and built-up area) were identified. Random sampling (Pal and Mather, 2003) was used to select the training and test dataset using the reference map. A total of 800 pixels was used for training and 3800 pixels for testing the classifiers.

Classification accuracy derived with support vector machine (SVM) classifier was used to compare the performance of different feature selection approaches used in this study. SVM were initially designed for binary classification problems. In this study, the ‘one against one’ approach and a radial basis function kernel was used with \(\gamma\) (kernel width parameter) = 2 and \(C = 5000\) (Pal and Foody, 2010). Studies suggest that feature selection has positive impacts in terms of reduced data processing time and storage requirements, and small feature set could potentially be used without any significant loss of classification accuracy (Pal and Foody, 2010). For this purpose, a test of non-inferiority (Foody, 2009; Pal and Foody, 2010), which is based on fitting the confidence intervals to the estimated difference in accuracy using McNemar’s test was
used. The approach used in this study assume that a 1.0% decline in accuracy from that achieved by using all 65 features is of no practical significance, and this value was taken to define the extent of the zone of indifference in the test. The proposed fuzzy entropy based feature selection approach uses a triangle fuzzy neighbourhood. The value of the radius of the fuzzy neighbourhood influences the number of selected features which in turns affects the classification accuracy of the classifier. After several trials, a value of 0.2 was found to work well both in terms of the number of selected features and the accuracy achieved with the dataset used.

Results

All four filter based feature selection approaches was used to determine a subset of features comprising the best features. Entropy and fuzzy entropy based approaches provides a subset of features whereas Relief and signal to noise ratio provides a ranked list of all features. In order to compare the performance of Relief and signal to noise ratio approach with full feature set, a total of top 20 selected features were used for further analysis. Table 1 provides a list of selected features obtained with all four feature selection approaches, while Table 2 provides the classification accuracies achieved with selected features using support vector machine based classification algorithm. A classification accuracy of 91.68% was achieved with 14 selected features using fuzzy entropy based approach in comparison to 91.68% and 91.61% by signal to noise ratio and entropy based feature selection approaches using a total of 20 and 17 features respectively. In comparison, Relief based approach provides an accuracy of 88.61% with a total of 20 selected features suggesting inferior performance by this approach with DAIS dataset.

Table 1. Selected features with different feature selection approaches.

<table>
<thead>
<tr>
<th>Feature selection method</th>
<th>Selected feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entropy</td>
<td>32, 51, 63, 35, 8, 49, 42, 27, 48, 64, 6, 50, 65, 11, 53, 39, 22</td>
</tr>
<tr>
<td>Fuzzy entropy</td>
<td>32, 41, 50, 6, 27, 63, 36, 49, 10, 22, 65, 51, 40, 48</td>
</tr>
<tr>
<td>Relief</td>
<td>3, 4, 2, 11, 10, 5, 8, 6, 9, 7, 12, 1, 13, 23, 22, 25, 24, 20, 31, 30</td>
</tr>
<tr>
<td>Signal to noise ratio</td>
<td>5, 7, 8, 9, 6, 10, 11, 4, 12, 3, 32, 31, 33, 30, 24, 23, 25, 29, 13, 26</td>
</tr>
</tbody>
</table>
Table 2. Classification accuracy with SVM classifier with different selected features.

<table>
<thead>
<tr>
<th>Feature selection method</th>
<th>Number of features used in classification</th>
<th>Classification accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No feature selection</td>
<td>65</td>
<td>91.76</td>
</tr>
<tr>
<td>Fuzzy entropy</td>
<td>14</td>
<td>91.68</td>
</tr>
<tr>
<td>Entropy</td>
<td>17</td>
<td>91.61</td>
</tr>
<tr>
<td>Signal to noise ratio</td>
<td>20</td>
<td>91.68</td>
</tr>
<tr>
<td>Relief</td>
<td>20</td>
<td>88.61</td>
</tr>
</tbody>
</table>

Table 3. Difference and non-inferiority test results based on 95% confidence interval on the estimated difference in accuracy from the accuracy achieved with 65 features and the feature sets selected using fuzzy entropy, entropy, signal to noise ratio and Relief based feature selection approach.

<table>
<thead>
<tr>
<th>Number of features</th>
<th>Accuracy (%)</th>
<th>Difference in accuracy (%)</th>
<th>95% confidence interval</th>
<th>Conclusion (at 0.05 level of significance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>91.76</td>
<td>0.00</td>
<td>0.000-0.000</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>91.68</td>
<td>0.36</td>
<td>0.071-0.089</td>
<td>Non-inferior</td>
</tr>
<tr>
<td>17</td>
<td>91.61</td>
<td>0.13</td>
<td>0.142-0.158</td>
<td>Non-inferior</td>
</tr>
<tr>
<td>20</td>
<td>91.68</td>
<td>0.26</td>
<td>0.071-0.089</td>
<td>Non-inferior</td>
</tr>
<tr>
<td>20</td>
<td>88.61</td>
<td>3.00</td>
<td>3.140-3.160</td>
<td>Inferior</td>
</tr>
</tbody>
</table>

Results of test of non-inferiority (Table 3) also indicates that fuzzy entropy, entropy and signal to noise ratio based feature selection approaches works well with this dataset and decline in accuracy with small number of features are not significant thus justifying the need of feature selection with the used dataset. In comparison to signal to noise ratio approach, fuzzy entropy based feature selection approach uses a smaller subset of features to achieve same level of classification accuracy, indicating the usefulness of fuzzy entropy based feature selection approach with DAIS dataset.

Conclusions

This paper discusses four filter based feature selection approaches. To compare their performance, classification accuracy derived using SVM classifier was used. Results from this study suggest that the fuzzy entropy based feature selection approach works well and provides comparable performance with 14 selected features in comparison to the all 65 features. Accuracy achieved by signal to noise ratio and entropy based approaches employing 20 and 17 features is
also comparable to that is achieved with full dataset. Further, results with 20 selected features by Relief based approach show a significant decline in classification accuracy in comparison to all 65 features.

References


