Feature Extraction using The K-Means Fast Learning Artificial Neural Network

Yin Xiang, Alex Tay Leng Phuan1
Nanyang Technological University

Abstract - The Fast Learning Artificial Neural Network is a small neural network bearing two types of parameters, The tolerance, \( \delta \) and the vigilance, \( \mu \). By exhaustively setting the combinatorial space of these parameters, it is possible to extract the data clustering behaviour to test for significance between the obtained data clusters and the actual data. If the correlation between the clustered data output and the actual data output is high, a clustering function would likely exist in the neural network that uses the prescribed parameter set. In doing so, it is possible to extract significant factors from an array of input factors and thus determine the principal factors that contribute to the particular output. Experimental results are presented to illustrate the network’s ability to extract significant factors using available test data.

1 Introduction

It is common to encounter the paradox of gaining access to a large throve of data, but yet be devoid of any useful information. It is by no means an easy task to convert such data into meaningful knowledge [2],[5],[8],[10]. This paper presents the K-Means Fast Learning Artificial Neural Network (K-FLANN) as an appropriate neural network architecture to perform data mining. Although the combinatorial parameterization of the network settings is done in the crudest of ways, it shows that the technique indeed has merit in extracting unseen relationships. An introduction of the Fast Learning Artificial Neural Network is provided to assist the understanding of the concept.

1.1 The K-FLANN Algorithm

The basic architecture of the K-FLANN is shown in Figure 1 [1],[10],[11]. It has 2 layers, the input and output layer, and a set of weight vectors connecting the 2 layers. The K-FLANN is a fully connected network.

The number of output nodes can grow according to the classification. As each new cluster is formed, a new output node is created and the weight vectors of the new node are cloned with the exemplar values. The algorithm of the K-FLANN follows [1],[10],[11].

1.2 Algorithm of K-FLANN

Initialize network with \( \mu \) between 0 and 1. Determine and set \( \delta_i \) for \( i = 1, 2, 3, \ldots, n \). The values of \( \mu \) and \( \delta \) affect the behaviors of the classification and learning process.

Present the next pattern to the input nodes. If there are no output clusters present, GOTO 6.

Determine the set of clusters that are possible matches using equation (1). If there are no output clusters GOTO 6.

Using criteria in equation (2) determine the winning cluster from the match set from Step 3. Normalize \( W_{ji} \) and \( I_i \). The following distance is calculated between the normalized versions.

When the Winner is found. Add vector to the winning cluster. If there are no more patterns, GOTO 7. Else GOTO 2.

No match found. Create a new output cluster and perform direct mapping from input vector into weight vector of new output cluster. If there are no more patterns, GOTO 7. Else GOTO 2.

Re-compute cluster center using K-means algorithm. Find the nearest vector to the cluster center in each cluster using equation (2). Place the nearest vector in each cluster to the top of the training data and GOTO 2.

Where \( \mu \) is the Vigilance Value, \( \delta \) are the Tolerance Values, \( W_{ij} \) - Weight connecting the \( i \)th input and the \( j \)th output neurons and \( D[a] = 1 \) if \( a > 0 \). Otherwise \( D[a] = 0 \).

After each cycle of training, when all exemplars have been processed, the cluster centers are updated using K-
means algorithm. This is Step 7 of the K-FLANN algorithm. Then comparison between each cluster center and patterns in respective cluster is conducted to find the nearest pattern to each cluster center. The nearest patterns are placed on the top of the training patterns and new cycle starts.

1.3 Parameter Search for \( \delta \) and \( \mu \)

The K-FLANN algorithm has the ability to cluster effectively, with consistent centroids forming regardless of variations in the data presentation sequence. However, it is important that the correct sets of \( \delta \) and \( \mu \) values are used so that the correct clustering results are created from the data [11]. The objective of this paper is to study the K-FLANN, as a data mining model to extract important and meaningful information from raw data.

A brute force combinatorial exhaustive search was used to set the tolerance and vigilance values. This algorithm tries all possible combinations of tolerance and vigilance values. This is appropriate because there is no assumed knowledge of the data and it removes the requirement of knowing the data or desired class assignments. There are other plans to improve this search methodology, but this is not discussed in the paper. Being an exhaustive search methodology, it is inevitable that the algorithm is slow, compared to other methods, since it tries all possible combinations of network parameters.

The method essentially permutes the different tolerance values in step-wise intervals. In an exemplar with \( m \) attributes and \( n \) step-wise intervals \( n^m \) modifications have to be made on the tolerance values. This means that as the number of attributes grows linearly, the time taken to complete the exhaustive search grows exponentially. Moreover, it is difficult to determine the correct resolutions of tolerance values. Small resolutions will result in polynomial growth of computation while big resolutions may miss good tolerance values. The value of \( \mu \) is also changed between 0.5 and 1.

Two sets of data are discussed in this paper, namely the Fisher's Iris data and the Wine data. Parameters of the correctly classified runs based on the brute force search were stored for further analysis. The widely publicized data was essential because the data is widely known and studied. Queries, if any, should not arise from the integrity of the data, but from the ability of the algorithm. If the blind match was successful, the results do not have to be verified because the data has already been well understood. It would then serve as a good basis to deduce the effectiveness of the algorithm.

The iris and wine data sets are well known and widely and was obtained from a web-site [17].

1.4 Combination on Input Attributes

Although input data have been preprocessed, there may also be correlated or unrelated data fields in the data set. Therefore, combination on input attributes can help choose relevant subset of the attributes as the input to the K-FLANN. After a clustering is done, The clustered output data is checked with the input data for similarities in the classes defined. This is done using a correlational analysis.

1.5 Statistical Correlation Analysis

The K-FLANN can extract information inherent in the input data. Some may be useful, while others may be meaningless and redundant. A form of statistical correlation analysis was performed to test for desired clustering after each convergence of K-FLANN training. When the K-FLANN reads training data, the correct class assignment for each data pattern is simultaneously into K-FLANN. In this way, it is possible to determine if the final cluster maintains a high percentage of correlation with the actual class assignments.

The K-FLANN then calculates the percentage of correct class assignments for each desired class of every cluster in order to remove the scale dominance of one class. Equation (3) was used as a simplistic correlation calculator.

To show that a significant correlation is found, the 2nd maximum percentage of correct class assignment is subtracted from the maximum percentage in one cluster and the correlation in the cluster is the normalized result of subtraction. Thus, the correlation of each cluster represents how far 2 desired clusters are separated and it functions as a threshold to filter the clusters with lower correlation value.

\[
\text{correlation} = \frac{\text{max} - 2\text{nd max}}{\text{max} - \text{min}} \quad (3)
\]

As shown in the equation, correlation has a range from 0 to 1. Although any correlation value ranging from 0 to 1 can be chosen, 0.9 is recommended because it provides a reasonable clustering.

2 Experiments and Results.

2.1 Iris Data Set

Fisher's paper is a classic in the field and is referenced frequently to this day. The data set contains 150 random samples of flowers from the Iris species: Setosa, Versicolor, and Virginica. From each species there are 50 observations with 4 attributes each in centimeters. The first class, Setosa, is linearly separable from the other 2, while the latter two are not linearly separable from each other. This means that Versicolor and Virginica intersect with each other and K-FLANN is not able to 100% correctly classify the 3 iris species. The following are the 4 attributes of iris data:
- Sepal Length
- Sepal Width
- Petal Length
- Petal Width

The following figures show the distribution of iris patterns with combination of 2 attributes each.

![Figure 2: Pattern Distribution of Iris: Sepal Length vs. Sepal Width](image)

As seen from the above figures, Setosa is sufficiently separated from Versicolor and Virginica, while Versicolor and Virginica intersect with each other, so that it is not possible to separate them. Furthermore, the petal length versus petal width gives better distribution and more separable iris patterns comparing with sepal length versus sepal width.

![Figure 3: Pattern Distribution of Iris: Petal Length vs. Petal Width](image)

The results of data mining exercise on the Iris data is discussed. Since there were many good clustering results, only the best is shown here. As seen from Figure 4, the K-FLANN can give very accurate results on pattern recognition and classification if the tolerance and vigilance values can be properly set. And it was noticed that Setoso was perfectly classified in most results, whereas Versicolour and Virginica were never perfectly separated. Therefore, from the results the following conclusion could be made: Versicolor and Virginica must have some intersection with each other at their boundaries.

![Figure 4: the Best Results of the mined Iris Data](image)

As mentioned previously, that’s because Setoso is linearly separable from the other 2 plant classes while the other 2 are not linearly separable from each other. K-FLANN can only perfectly classify linearly separable patterns. Therefore, pattern separability of a data set can be deduced from its classification results.

<table>
<thead>
<tr>
<th>Sepal Len</th>
<th>Sepal Wid</th>
<th>Petal Len</th>
<th>Petal Wid</th>
<th>Accu</th>
<th># of Clus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>√</td>
<td>96%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>√</td>
<td>96%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>√</td>
<td>96%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>√</td>
<td>96%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>√</td>
<td>95.30%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>√</td>
<td>92.70%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>√</td>
<td>86.70%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>√</td>
<td>85.30%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>√</td>
<td>83%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>√</td>
<td>79.30%</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>√</td>
<td>78.70%</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>√</td>
<td>72.70%</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>√</td>
<td>66.70%</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>√</td>
<td>66.70%</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>√</td>
<td>58%</td>
<td>7</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The summary of the significance table

The most accurate classification of each subset combination of iris attributes is shown in Table 1. The attributes with a tick “√” indicate the presence of the attribute to achieve the accuracy. The number of clusters is recorded in the last column.

Table 1 was sorted based on accuracy of results in descending order so as to show the combinations achieving the best accuracy. The following observations can be made from this table.
• The first 4 combinations all achieved the highest obtained accuracy, 96%.
• Petal width was included in all the first 4 combinations.
• K-FLANN could achieve 96% accuracy using only petal width according to the result of the 1st combination.
• The accuracy dropped significantly as combinations excluded petal width according to Row No. 9 to 15.

From the above observation it could be deduced that petal width was a main factor in determination of the Iris classification, and that building larger subsets using the main factor would be more likely to get better classification results. This might enable us to perform a faster search with aim instead of a blind exhaustive search used now.

2.2 Wine Data

This data was obtained from a chemical analysis of wines grown within the region of Italy, but were derived from three different cultivators. The analysis determined the quantities of 13 constituents found in each of the three types of wines, namely

- Alcohol
- Malic acid
- Ash
- Alcalinity of ash
- Magnesium
- Total phenols
- Flavanoids
- Nonflavanoid phenols
- Proanthocyanins
- Colour intensity
- Hue
- OD280/OD315 of diluted wines
- Proline

There were 178 instances of wine samples in the data set and the 3 types of wines bearing the following distribution;

- Class 1: 59 instances
- Class 2: 71 instances
- Class 3: 48 instances

Entire wine data set was used in K-FLANN clustering. Exhaustive permutations of attributes were clustered and the results were compared against the actual output class assignments of the input data set. K-FLANN clustering results exceeding 90% were considered to have a significant correlation with that specific class and to be valuable.

It took approximately 2 weeks to extract the combinations of the 5 significant attributes. There were more than one hundred significant results generated and only the best combination of attributes were considered. This data was also sorted according to accuracy of results in descending order. Most combinations had more than 90% accuracy and there were no obvious variations in between. From the significance data obtained from clustering, several attributes were observed to appear frequently. These are listed below with the number of occurrences in brackets.

- Alcohol (35 Times)
- Flavanoids (26 Times)
- Colour intensity (32 Times)
- OD280/OD315 of diluted wines (31 Times)
- Proline (26 Times)

These frequent occurrences indicate that the factors are probably significant in the determination of wine classification. To further investigate the main factors, a cut off line of 94% was set and the resulting Table 2 was obtained. The attributes are listed in numerical form to aid the charting exercise.

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th>SIGNIFIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13</td>
<td></td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>95.51%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>95.51%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>95.51%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>94.94%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>94.84%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>94.94%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>94.94%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>94.94%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>94.38%</td>
</tr>
<tr>
<td>✓ ✓ ✓ ✓✓ ✓✓</td>
<td>94.38%</td>
</tr>
</tbody>
</table>

Table 2: Analysis of Highly Significant Clusters

From it was clear that only the following 3 attributes of wine played significant roles in the wine classification.

- Flavanoids (Item 7)
- Colour intensity (Item 10)
- Proline (Item 13)

A statistical analysis was performed to find the mean accuracy and standard deviation of each combination. For example, combination of Flavanoids and colour intensity might play a part in a larger combination such as Flavanoids, colour intensity and Proline. The mean accuracy and standard deviation of Flavanoids and colour intensity could be calculated based on appearance of their existence in all combinations. The higher the mean accuracy and the less the standard deviation, the more accurate the result is. The results with mean accuracy higher than 94% were illustrated in Table 3.
This Table was sorted according the mean accuracy in descending order and the 1st column represents the number of appearances of each combination. SD indicated standard deviation in this table. Row No. 2 (combination of Flavanoids, colour intensity and Proline) showed absolute priority over others in the following 3 characteristics:

- Very frequent appearance
- High mean accuracy
- Low standard deviation

Table 3: Analysis of Mean and Standard Deviation

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>12</th>
<th>13</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>0.9466</td>
<td>0.0119</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>0.9452</td>
<td>0.0094</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>0.9458</td>
<td>0.0159</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>0.9458</td>
<td>0.0112</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>0.9419</td>
<td>0.0130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td></td>
<td>0.9419</td>
<td>0.0086</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>0.9410</td>
<td>0.0040</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>0.9410</td>
<td>0.0040</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>0.9410</td>
<td>0.0040</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>0.9410</td>
<td>0.0040</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The 1st characteristics indicated that K-FLANN heavily relied on this combination for correct classification. The 2nd and 3rd characteristics meant that it had very high classification accuracy. Therefore, the result of statistical analysis also showed that Flavanoids, colour intensity and Proline were the main factors in determination of wine classification.

As long as the main factors of wine data were determined, it was possible to building larger combinations of attributes with high accuracy using the main factors. The deduction here also proved what was obtained from the iris data set. Therefore, a faster searching algorithm may be established with further investigation in this field.

### 3 Conclusions

Although the data used were from well known sources which have been investigated by many, the emphasis of the experiments were on the technique which was used to extract the features from the data. From the Fisher Iris data, the analysis resulted in the determination of two significant factors, being petal width and petal length. This information was extracted without the need to have an understanding of the data. Similar results were obtained from the wine data, from which it was determined that only 3 factors were significant out of the 14. Further investigations are underway to determine if there is a less expensive method to search for optimal clustering parameters without having to spend on the exhaustive search strategy. It is conclusive that the method of feature extraction is feasible, but the search technique needs to be improved.

### 4 References

[17]. (http://www.ics.uci.edu/~mlearn/MLRepository)