Fuzzy Beans in Classification

Pasi Luukka

Laboratory of Applied Mathematics
Lappeenranta University of Technology
P.O. Box 20, FIN-53851 Lappeenranta, Finland

Abstract

In the article fuzzy bean based classifier is given. Three different type of structures are tested in the classifier. Fuzzy bean based classifier is supervised learning method and with proper optimization scheme promising results were covered. Differential evolution algorithm was used in optimizing the required parameters. Classifier was applied to diagnosis of liver-disorders with 73.9% classification accuracy, to Pima-indian diabetes with 77.8% accuracy, to diagnosis of breast cancer with 97.8% accuracy and to echocardiogram with 93.0% accuracy. Results can be considered good and they compare well with results reported with literature.

Key words: Fuzzy bean classifier; Fuzzy systems; Differential evolution; Learning; Liver-disorder data; Breast cancer data; Diabetes data; Echocardiogram;

1 Introduction

Classification is used in many fields. It can be a very important part of process in many different fields like e.g. medicine and manufacturing. The disease must be diagnosed before suitable treatment can be given or faulty products or materials must be removed from the pipeline. In quality control classification is essential tool nowadays. Traditionally this kind of job has been made by humans, but now one of course want to automate this kind of work as much as possible. Reasons for that are very simple. First, usually classification can be made much more faster by computers than by humans and computers are cheaper to use than humans. Also with easy classification tasks one can usually find an algorithm for computer which does not make classification errors, humans always make some errors. Some classification tasks can be also simply too difficult for humans because there can be too many parameters to measure and/or they can be heavily correlated. Sometimes the situation can also be contrary, for example humans might sense much more easier than any algorithm if some individual is violent.
In this article a fuzzy bean based classifier is introduced. Usually in classification we have several measured features and we need to build a recognition system that can take all these features into account. Also they are coupled with noise and errors so we have to find ways to build measures than can cope under this kind of uncertainty. In classification this type of uncertainty is almost always present. Fuzzy logic is a logic that allows vagueness, imprecision and uncertainty. Zadeh (1965) [1] introduced the concept of fuzzy sets and the respective theory that can be regarded as the extension of the classical set theory. Fuzzy bean [2] is giving a suitable structure to cope with the task of comparing vectors where uncertainty is present. Besides this, there are many coefficient values which needs to be fitted in general structure of fuzzy bean to the particular task at hand. Here a supervised learning method is needed and in this article we have chosen differential evolution algorithm [3], [4] to find suitable values to fuzzy bean and to ideal vectors that are representing the class best. Here supervised learning seems to be essential to the classification task at hand instead of just using identity matrix or covariance matrix and mean values. This can be clearly seen in the results.

Data sets used in this experiment were taken from a UCI-Repository of Machine Learning Database [5]. Chosen data sets were, diagnosis of liver-disorders, diabetes, breast cancer and echocardiogram. Classifier was implemented with MATLAB™-software.

2 Data Sets

Next a short description of the data sets used in this article is given. Data sets were taken from UCI machine learning data repository [5] where they are freely available. The fundamental properties of the data sets are shown in Table 1.

[Table 1 about here.]

Liver-disorder data set:
The data set was donated (by R. S. Forsyth) to the UCI machine learning data repository [5]. The problem is to predict whether or not a male patient has a liver-disorder based on blood tests and alcohol consumption. The attribute information for the liver-disorder data set is the following 1) mean corpuscular volume 2) alkaline phosphotase 3) alamine aminotransferase 4) aspartate aminotransferase, 5) gamma-glutamyl transpeptidase 6) number of half-pint equivalents of alcoholic beverages drunk per day. The first five variables are all blood tests which are thought to be sensitive to liver disorders that might arise from excessive alcohol consumption.
Breast cancer data set:
This data set was created by Dr. Wolberg and the purpose was to accurately diagnose breast masses based solely on Fine Needle Aspiration (FNA). He identified nine visually assessed characteristics of an FNA sample which he considered relevant to the diagnosis. The resulting data set is well-known as the Wisconsin Breast Cancer Data. The nine variables used to predict benign or malignant cases were: 1) Clump Thickness 2) Uniformity of Cell Size 3) Uniformity of Cell Shape 4) Marginal Adhesion 5) Single Epithelial Cell Size 6) Bare Nuclei 7) Bland Chromatin 8) Normal Nucleoli and 9) Mitoses.

PIMA Indians:
The Pima Indian data set concerns the presence or absence of diabetes among Pima Indian women living near Phoenix, Arizona. There are eight covariates: number of pregnancies; plasma glucose concentration; diastolic blood pressure (mmHg); triceps skin fold thickness (mm); serum insulin (µU/ml); body mass index (kgm\(^{-2}\)); diabetes pedigree function and age in years.

Echocardiogram:
In this data set, one sample consists of results from echocardiogram measurements of one patient who has recently suffered an acute heart attack. Measurements are taken from echocardiograms, which are ultrasound measurements of the heart itself. The goal of physicians using these measurements is to predict a patient’s chances of survival. Experimental work is being performed to determine if an echocardiogram, in conjunction with other measures, could be used to predict whether or not a patient would survive for longer than a certain time period. If this data can be classified accurately enough, it would give means for predicting future heart attacks in former heart patients.

3 Classification Procedure

3.1 Fuzzy bean classifier

The problem of classification is basically one of partitioning the attribute space into regions, one region for each category. Ideally, one would like to arrange this partitioning so that none of the decisions are ever wrong [6].

Consider we would like to classify a set \( X \) of objects into \( N \) different classes \( C_1, \ldots, C_N \) by their attributes. We suppose that \( t \) is the number of different kinds of attributes \( f_1, \ldots, f_t \) that we can measure from the objects. We suppose that the
values for the magnitude of each attribute are normalized so that they can be presented as a value between [0, 1]. Consequently, the objects we want to classify are vectors that belong to $[0, 1]^t$.

First one must determine for each class the ideal vector $v_i = (v_i(f_1), \ldots, v_i(f_t))$ that represents the class $i$ as well as possible. Here these vectors that represent the classes $i$ are optimized using differential evolution algorithm. Once the ideal vectors have been determined, then the decision to which class an arbitrarily chosen $x \in X$ belongs is made by comparing it to each ideal vector. The comparison is done here by using fuzzy beans and Bocklisch membership value function. In this study we have used two different type of fuzzy beans [2] and Bocklisch [7], [2] membership function for vectors.

For fuzzy bean, membership value between the data vector and ideal vector can be computed in form

$$\mu\langle x, v_i \rangle = \max\{[1 - (v_i - x)^TB(v_i - x)], 0\}$$  \hspace{1cm} (1)$$

where $B$ is $t \times t$ matrix and $x$ is our data vector and $v_i$ is the ideal vector for class $i$.

The decision to which class the data vector belongs is made by

$$\mu\langle x, v_i \rangle = \max_{i=1,\ldots,N} \mu\langle x, v_i \rangle .$$  \hspace{1cm} (2)$$

In other words the decision to which class the sample belongs is made according to which ideal vector the sample has the highest membership value. Notice that the simples case would be to choose matrix $B$ as identity matrix when we would have fuzzy bean as 1–euclidean distance. and if $B$ would be chosen as covariance matrix we would have similar structure as 1–mahalanobis-distance. In case of fuzzy beans $B$ is not restricted to these choices but more arbitrary values can be chosen. Clearly next question is how to determine the values in $B$. In this article it is done by using differential evolution algorithm to find suitable values for $B$. By doing it like this we are allowing more general structure to the membership function values and this should in principle allow more accurate results in cases where euclidean or mahalanobis-distance are not the best choices. One advantage of calculating the membership values in this manner is that we expect to increase the classification accuracy. About the drawbacks of doing it like this is that now we are facing the problem of finding optimal values for whole matrix $B$ as well as for the ideal vectors. This is not a problem with smaller data sets where there is not that many features but when the amount of features increases this optimization problem can come too large for finding proper values. This version of fuzzy bean (1) is in future called fuzzy bean 1 (FB1). Now since we noticed similar structure in fuzzy bean as in 1– distance another way to define membership function would be
\[
\mu(x; v_i) = \frac{1}{1 + d(x; v_i)}
\] (3)

and replacing \(d(x; v_i)\) here again with the bean structure we now get

\[
\mu(x; v_i) = \frac{1}{1 + (v_i - x)^T B(v_i - x)}
\] (4)

Here again the matrix \(B\) and ideal vectors needs to be optimized in supervised manner. This version of fuzzy bean (4) is later in the article called fuzzy bean 2 (FB2). One proposal starting also from (3) but allowing left and right side parameters, is coming from Bocklisch [7, 2] where membership function was given in form

\[
\mu(x; v_i) = \frac{1}{1 + A(x; v_i) + B(x; v_i)}
\] (5)

where

\[
A(x; v_i) = \sum_{j=1}^{t} \frac{1}{2} \left(1 + sgn(x_j - v_j)\right) \left(\frac{1}{b_{lj}} - 1\right) \left|\frac{x_j - v_j}{c_{lj}}\right|^{d_{lj}}
\]

\[
B(x; v_i) = \sum_{j=1}^{t} \frac{1}{2} \left(1 + sgn(v_j - x_j)\right) \left(\frac{1}{b_{rj}} - 1\right) \left|\frac{v_j - x_j}{c_{rj}}\right|^{d_{rj}}
\]

with \(b_{lj}, b_{rj} \in (0, 1]\) and \(c_{lj}, c_{rj}, d_{lj}, d_{rj} > 0\) and \(sgn\) simply meaning sign of the difference. Parameters are again here to be optimized. Now \(b\) and \(c\) parameters control the membership values and \(d\) parameters are so called shape parameters for the membership function hence managing the deviate from bean like shape.

The differential evolution algorithm [3, 4] here is used to optimize the ideal vectors for each class, the \(B\) matrix and in the Bocklisch case the parameters, \(b_{lj}, b_{rj}, c_{lj}, c_{rj}, d_{lj}, d_{rj}\). Optimization is done in supervised manner. Data is divided in the learning set and in the testing set. In learning set we use the classification accuracy as fitness value which is to be optimized. After this the optimized ideal vectors and \(B\)-matrix (or optimized parameter values in Bocklisch case) are used with the testing set. Next a bit more thorough description of differential evolution algorithm is given.
3.2 Differential evolution algorithm

The basic idea of evolutionary algorithms is that we create a population $V_0$ of trial solutions (vectors) for the optimization problem. Next, we combine the members of $V_0$ in a certain way and check if the combined solutions are better than the original trial solutions in $V_0$. The best solutions continue to population $V_1$, the next step of the evolution, and then the whole procedure starts again.

Differential evolution (DE) was developed around 1995 by Storn and Price [3]. Unlike genetic algorithms, DE works directly with continuous variables without the encoding and decoding of vectors. We denote the population at evolution step $n$ by set $V_n$. Unlike in GA, our DE algorithm is written for minimizing problems, so fitness function $f : \mathbb{R}^D \to [0, 1]$ is now set so that if $f(v) < f(u)$ then $v$ is a better solution for our problem than $u$.

A basic DE contains also crossover operation, but it has no real mutation operation. Instead of mutation, DE demonstrates differential variation operation. In one evolution step, we carry out the following two basic operations for all $v_n = \{v_1, \ldots, v_D\} \in V_n$:

1. **Differential variation.** The basic idea of differential variation is to add "noise vector" $n$ to vector $w \in V_n$. Both $n$ and $w$ can be chosen in many ways. Usually, $n$ is either the difference vector between two vectors in $V_n$ or a linear combination of these kinds of difference vectors. From $n$ and $w$ we form a new vector

   $$ u = w + F n, $$

   where usually parameter $F$ is constant and between 0 and 1.

2. **Crossover.** From $u = \{u_1, \ldots, u_D\}$ and $v_n$ we form a trial vector $t = \{t_1, \ldots, t_D\}$ by setting

   $$ t_i = \begin{cases} 
   u_i & \text{if } x_i < CR \\
   v_i & \text{otherwise}
   \end{cases} $$

   where $x_i \in [0, 1]$ is a uniformly distributed random variable and $CR$ is the crossover probability. Usually values $F = 0.9$ and $CR = 0.8$ work quite well. Finally the vector $v_{n+1}$ which is chosen to the next generation $V_{n+1}$ is

   $$ v_{n+1} = \begin{cases} 
   t & \text{if } f(t) < f(v) \\
   v_n & \text{otherwise}
   \end{cases} $$
w was chosen as the best member of $V_n$,

$$f(w) = \min_{v \in V} \{ f(v) \},$$  \hspace{1cm} (9)

and $n$ was the difference vector

$$n = y_1 - y_2,$$  \hspace{1cm} (10)

where $y_1 \in V_n$ and $y_2 \in V_n$ were randomly chosen vectors.

4 Classification results and comparison

In all data sets data was split in half. One half was used for training and one half for testing. This procedure was repeated randomly 30 times and mean classification accuracies were computed. In Table 2 mean classification accuracies of the given methods are listed. As can be seen from the Table 2 FB1 was given the highest mean accuracies with data sets, Liver-disorders, Breast cancer and Echocardiogram. With Pima-Indians data set highest mean accuracy was gained with FB2.

[Table 2 about here.]

Classification accuracy of FB1 is compared to other classifiers which are giving quite good results in Table 3. Compared classifiers are decision tree classifier CN2, neural network classifiers MLP and DIMLP [10] and SIM which is similarity based classifier, [8] [9]. Classification results are taken from [10], [9] or classified for this comparison (this is done with Echocardiogram data set). As can be seen with all four data sets FB1 managed to get higher mean classification accuracy than the compared classifiers. With liver-disorders data set, a mean accuracy of 73.91% was achieved with FB1. This was over 3% units higher accuracy than what was gained with other classifiers. With Breast cancer data set, a mean accuracy of 97.82% was received with FB1. This was again highest classification accuracy compared to the other four but now only 0.2% better than similarity classifier. With Pima-Indians data set, a mean accuracy of 77.83% was achieved with FB2 and FB1 gave second highest mean accuracy of 77.23%. Difference between FB1 to second best classification accuracy in Table 3 was around 2%. Echocardiogram data was classified using FB1 with mean accuracy being 93.0% which is around 3% higher accuracy than what was gained using similarity classifier. As can be seen from the comparison fuzzy beans with proper optimization scheme has shown to be classifier with respectable results.

[Table 3 about here.]

To further test whether we really managed to find better values and more general form using differential evolution algorithm we also made the classification results
so that instead of optimizing the ideal vectors and $B-$matrix we used mean values for each class as ‘ideal vectors’ and $B-$matrix was in case (1) an identity matrix, getting the distance function similar to euclidean distance case (2) covariance matrix now distance part being the Mahalanobis-distance. In Table 4 one can see the results from this comparison compared to optimized FB1 which seemed to work best to these data sets. As can be seen from the Table 4 much higher classification accuracy was gained by using fuzzy bean so that ideal vectors and $B-$matrix was optimized. The enhancement of the accuracy was over 10% in all cases except with breast cancer data set where already with the two simpler cases managed with over 96% mean classification accuracy. This experiment shows that eventhough the classification scheme is quite simple one can get good results with this type of general fuzzy bean structure and optimizing the $B-$matrix can is clearly enhancing the results and much needed instead of using simply euclidean or Mahalanobis-distance type structure.

[Table 4 about here.]

5 Discussion

In the article fuzzy bean based classifier was given. It was tested with couple different type of fuzzy bean structures and ideal class vectors. The $B-$matrix in bean structure was optimized using differential evolution algorithm. Results are tested with four different data set, liver-disorders, breast cancer, pima-indians and echocardiogram data sets. Classification results are compared to other existing results. Mean classification accuracies were quite good with liver-disorders 73.9%, breast cancer 97.8%, pima-indians 77.8% and echocardiogram 93.0%. This can be considered quite good mean accuracies and fuzzy bean based classifier has shown to be a promising classifier.

One idea with this scheme was that by optimizing the $B-$matrix one can get a more general form and this way by optimizing the suitable values for fuzzy bean a better accuracy could be gained. This seemed to happen in this case and when compared to $B$ being identity matrix or covariance matrix it was noticed that better accuracy was gained this way. On the otherhand one limiting factor to this procedure are high dimensional data sets since when measured features in data are growing also $B-$matrix is growing and eventually there will be too many parameters to be optimized using the chosen method. This may be avoided by using preprocessing schemes and reduced dimensions which is a subject of further study. Also other optimization schemes besides differential evolution algorithm e.g. particle swarm optimization could be tested.
References


**List of Tables**

1. Test data sets and their properties.  
2. Classification results for the data sets. The data set is listed in the first column, the second column gives the method used, the third column gives mean classification accuracy (in %), and variances are reported in the fourth column.  
3. Classification result comparison.  
4. Classification results for the data sets. Effects of supervised learning method is tested and compared to simply using euclidean or mahalanobis distance type structure in fuzzy bean. The data set is listed in the first column, the second column gives the method used, the third column gives mean classification accuracy (in %), and variances are reported in the fourth column.
Table 1
Test data sets and their properties.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Nb. classes</th>
<th>Dim</th>
<th>Nb. cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pima-indians</td>
<td>2</td>
<td>8</td>
<td>768</td>
</tr>
<tr>
<td>Liver disorders</td>
<td>2</td>
<td>6</td>
<td>345</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>2</td>
<td>9</td>
<td>699</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>2</td>
<td>7</td>
<td>131</td>
</tr>
</tbody>
</table>
Table 2
Classification results for the data sets. The data set is listed in the first column, the second column gives the method used, the third column gives mean classification accuracy (in %), and variances are reported in the fourth column.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Method</th>
<th>Mean acc.</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver-disorders</td>
<td>FB1</td>
<td>73.91</td>
<td>0.0424</td>
</tr>
<tr>
<td>Liver-disorders</td>
<td>FB2</td>
<td>72.83</td>
<td>0.0315</td>
</tr>
<tr>
<td>Liver-disorders</td>
<td>Bocklisch</td>
<td>70.40</td>
<td>0.0306</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>FB1</td>
<td>97.82</td>
<td>0.0044</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>FB2</td>
<td>97.54</td>
<td>0.0050</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>Bocklisch</td>
<td>96.39</td>
<td>0.0123</td>
</tr>
<tr>
<td>Pima-Indians</td>
<td>FB1</td>
<td>77.24</td>
<td>0.0157</td>
</tr>
<tr>
<td>Pima-Indians</td>
<td>FB2</td>
<td>77.83</td>
<td>0.0289</td>
</tr>
<tr>
<td>Pima-Indians</td>
<td>Bocklisch</td>
<td>77.29</td>
<td>0.0164</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>FB1</td>
<td>92.99</td>
<td>0.0487</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>FB2</td>
<td>92.43</td>
<td>0.0569</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>Bocklisch</td>
<td>92.20</td>
<td>0.0567</td>
</tr>
</tbody>
</table>
Table 3
Classification result comparison.

<table>
<thead>
<tr>
<th>Data-classifier</th>
<th>CN2</th>
<th>MLP</th>
<th>DIMLP</th>
<th>SIM</th>
<th>FB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver-disorders</td>
<td>65.5</td>
<td>70.2</td>
<td>70.1</td>
<td>66.6</td>
<td>73.9</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>93.9</td>
<td>97.3</td>
<td>96.5</td>
<td>97.6</td>
<td>97.8</td>
</tr>
<tr>
<td>Pima-Indians</td>
<td>74.2</td>
<td>75.5</td>
<td>75.4</td>
<td>75.3</td>
<td>77.2</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>68.6</td>
<td>65.7</td>
<td>68.5</td>
<td>90.1</td>
<td>93.0</td>
</tr>
</tbody>
</table>
Classification results for the data sets. Effects of supervised learning method is tested and compared to simply using euclidean or mahalanobis distance type structure in fuzzy bean. The data set is listed in the first column, the second column gives the method used, the third column gives mean classification accuracy (in %), and variances are reported in the fourth column.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Method</th>
<th>Mean acc.</th>
<th>Var</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liver-disorders</td>
<td>$FB_{euc}$</td>
<td>56.65</td>
<td>0.1198</td>
</tr>
<tr>
<td>Liver-disorders</td>
<td>$FB_{mah}$</td>
<td>54.07</td>
<td>0.0758</td>
</tr>
<tr>
<td>Liver-disorders</td>
<td>$FB_1$</td>
<td>73.91</td>
<td>0.0424</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>$FB_{euc}$</td>
<td>96.43</td>
<td>0.0047</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>$FB_{mah}$</td>
<td>96.47</td>
<td>0.0060</td>
</tr>
<tr>
<td>Breast cancer</td>
<td>$FB_1$</td>
<td>97.82</td>
<td>0.0123</td>
</tr>
<tr>
<td>Pima-Indians</td>
<td>$FB_{euc}$</td>
<td>63.40</td>
<td>0.0756</td>
</tr>
<tr>
<td>Pima-Indians</td>
<td>$FB_{mah}$</td>
<td>60.43</td>
<td>0.0331</td>
</tr>
<tr>
<td>Pima-Indians</td>
<td>$FB_1$</td>
<td>77.24</td>
<td>0.0157</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>$FB_{euc}$</td>
<td>81.92</td>
<td>0.1275</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>$FB_{mah}$</td>
<td>81.07</td>
<td>0.1012</td>
</tr>
<tr>
<td>Echocardiogram</td>
<td>$FB_1$</td>
<td>92.99</td>
<td>0.0487</td>
</tr>
</tbody>
</table>