Membership Function Modification in Fuzzy SVM using Combination of Distance Feature, Correlation, and Depth of Data

M. Rahmat Widyanto and Rini Wijayanti
Faculty of Computer Science, University of Indonesia
Depok Campus, West Java, Depok 16424, INDONESIA
E-mail: widyanto@cs.ui.ac.id

Abstract-- In this paper, we proposed the membership function modification in Fuzzy Support Vector Machine (Fuzzy SVM) based on exponential distribution function. This function is expressed as the level of training data contribution to its class model building. It is formed prior to training phase and started with checking and handling the missing values. Afterwards is the formation of fuzzy variables based on the distance feature, correlation, and the depth of data to be formalized into the exponential distribution function. The experiments results show that the proposed method has a higher accuracy about 2% compared to conventional SVM and Fuzzy SVM. This method is expected more contributed in the development of various pattern recognition applications, which more demand on higher accuracy.

Index Term-- Fuzzy, svm, feature combination.

1. INTRODUCTION

Nowadays, automatic pattern recognition to find an optimal solution is still becomes an interesting topic to be further developed. Pattern recognition, which is usually called as classification, is one of topic in computer science that maps a set of data into several classes that has been specified [Anto, 2005]. This topic has been applied in many fields, i.e. email classification (spam or not spam), disease classification based on medical diagnose, text document classification, etc.

One of classification methods that are widely used recently is Support Vector Machine (SVM). It has been proved as the third most popular method after C.45 and K-Means in International Conference on Data Mining (ICDM 2006). It has been considered as the most influenced algorithm to data mining communities [XindongWu, 2007]. This method was first introduced by Vapnik in 1992 and still gains many attention until now due to its well generalization (B. Boser, 1992) (Vapnik, 1995) (Mulier, 1998) (Vapnik V. N., 1998) (Wu, 1999) (Byun H., 2003).

In conventional SVM, each input points are considered to have the same contribution to the training model. However, this is not always true since there might be some points corrupted by noise so they are less meaningful and the machine should better to discard them. This paper aims to handle this kind of data using fuzzy approach by giving them a weight which indicate different contribution of each point to the model. The weights are determined with respect to their membership function. Such approach is usually called as Fuzzy SVM (FSVM).

Several researches have been conducted in Fuzzy SVM. Chun-Fu Lin and Sheng-De Wang proposed Fuzzy SVM to solve such problem by weighting the data training so that different input points can make different contributions to the learning of decision surface [Chun-Fu Lin, 2002]. For non-sequential dataset, they estimate the distance between each data to its class center as membership function. This method is widely accepted and used in some papers. Qin-Yu Ping used this method to solve the multi-class problem [Qin-Yu Ping, 2008]. Hao Tang also used such distance function combined with affinity variable to weigh the input points [Hao Tang, 2008]. However, this method always considers the outer data as outlier because they are given a very small membership value, nearly or exactly the same as lower bound value that has been specified. Whereas, not all outer data can be considered as outlier since they have similar characteristic with other data in the same class. Moreover, they used Mean value as the center of class, while it is easy affected by outlier.

In this paper, a new membership function assigned to each data is proposed. We use three fuzzy variables: distance, correlation, and depth of data, and formulize them into exponential distribution function. Distance variable is used to see overlapping between class, while the correlation and depth of data is used to find the level of similarity. Class center location is determined using the Hodges-Lehmann due to the presence of outliers is not affected.

The rest of this paper is organized as follows. A brief review of the theory of SVM and Fuzzy SVM will be described in Section II. The proposed FSVM will be derived in Section III. The experiments are presented in Section IV. Some concluding remarks are given in Section V.

2. SVM AND FUZZY SVM

In this section, we describe about SVM and fuzzy SVM as the basic in classification problem

2.1. SVM

Support Vector Machine is a classifier that tries to find the optimal hyperplane in input space by minimizing the average error in training data and maximize the margin between classes. Margin is the distance between hyperplane to
the nearest data in each class, which is called as support vectors. These illustration can be seen in figure 1.

![Image](https://example.com/image.png)

**Fig. 1. Optimal hyperplane**

Suppose there are \( n \) datasets \( \{x_1, ..., x_n\} \), where \( x_i \in \mathbb{R}^n \) is belongs to one of the two classes labeled by \( y_i \in \{-1, +1\} \), then each sample point must satisfied the following conditions

\[
\begin{align*}
    w \cdot x_i + b & \geq 1 \quad \text{for} \quad y_i = +1, \\
    w \cdot x_i + b & \leq -1 \quad \text{for} \quad y_i = -1,
\end{align*}
\]

where \( w \) is a normal to the hyperplane and \( b \) is a bias of the hyperplane. Therefore, the margin of separating hyperplane is

\[
\frac{1 - b - (-1 - b)}{w} = \frac{2}{|w|}. \tag{2}
\]

Since maximizing \( \frac{1}{|w|^2} \) is equal by minimizing \( |w|^2 \), then the optimal hyperplane can be find by formulating

\[
\begin{align*}
    \min_{w,b} & \quad \frac{1}{2} \|w\|^2 \\
    \text{s.t.} \quad & \quad w \cdot x_i + b - y_i \geq 1, \quad i = 1, ..., n.
\end{align*}
\]

However, in the non-separable case, we introduce nonnegative slack variables \( \xi_i \) such that the following conditions are satisfied

\[
\begin{align*}
    \min & \quad \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^{n} \xi_i \right) \\
    \text{s.t.} & \quad y_i (w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0.
\end{align*}
\]

where \( \xi_i = 0 \), if \( x_i \) is correctly classified. \( C \) is a predefined positive constant which determines the penalty on empirical error.

The problem above is a quadratic problem which quite difficult to resolve and requires a heavy computation time. Therefore, a set of Lagrange multipliers \( (\alpha) \) is introduce and the problem is converted into

\[
\begin{align*}
    \min_{w,b} & \quad \frac{1}{2} \|w\|^2 \\
    \text{s.t.} & \quad \sum_{i=1}^{n} \alpha_i y_i (w \cdot x_i + b) \geq 1 \\
    & \quad \sum_{i=1}^{n} \alpha_i = C.
\end{align*}
\]

These points with \( \alpha_i > 0 \) is called as support vector. If \( 0 < \alpha_i < C \), these points are lie on one of the margin. Whereas if \( \alpha_i = C \), these points are regarded as misclassified data.

By minimizing the \( L_p \) with respect to \( b \) \( (\frac{\partial}{\partial b} L_p(w,b,\alpha) = 0) \) and \( w \) \( (\frac{\partial}{\partial w} L_p(w,b,\alpha) = 0) \), then be obtained

\[
\begin{align*}
    \sum_{i=1}^{n} \alpha_i y_i &= 0, \\
    w &= \sum_{i=1}^{n} \alpha_i y_i x_i.
\end{align*}
\]

The \( w \) vector is often very large (possibly infinite), but the value of \( \alpha_i \) is finite. Therefore, by substituting the equation above will be obtained the dual problem

\[
L_D(\alpha) \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i \cdot x_j.
\]

The problem to find the optimal hyperplane can be formulated as follows,

\[
\begin{align*}
    \max_{\alpha} & \quad L_D(\alpha) \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i \cdot x_j, \\
    \text{s.t.} & \quad \sum_{i=1}^{n} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, \quad \forall i.
\end{align*}
\]

Non-separable case also can be solved by kernel trick. This original input space is mapped into a high-dimensional dot product space called as feature space. The mapping illustration can be seen in figure 2.

![Image](https://example.com/image.png)

**Fig. 2. Mapping by kernel trick**

Kernel functions that commonly used in pattern recognition problem are the following:

\[
\begin{align*}
    \min_{w,b} & \quad \frac{1}{2} \|w\|^2 \\
    \text{s.t.} & \quad \sum_{i=1}^{n} \alpha_i y_i (w \cdot x_i + b) \geq 1 \\
    & \quad \sum_{i=1}^{n} \alpha_i = C.
\end{align*}
\]

\[
\begin{align*}
    \sum_{i=1}^{n} \alpha_i y_i &= 0, \\
    w &= \sum_{i=1}^{n} \alpha_i y_i x_i.
\end{align*}
\]
Linear: $K(x_i, x) = x_i^T x$.
Polynomial: $K(x_i, x) = (\gamma x_i^T x + r)^p, \gamma > 0$.
Radial Basis Function (RBF):
$K(x_i, x) = \exp(-\gamma |x_i - x|^2), \gamma > 0$.
Tangent hyperbolic (sigmoid):
$K(x_i, x) = \tanh(\gamma x_i^T x + r)$.

According to (Hsu, 2004), kernel that are recommended tested for the first time is RBF kernel function because it has the same performance with linear kernel and sigmoid kernel behaviors on certain parameters and have a small range values [0,1]

### 2.2. Fuzzy SVM

Fuzzy SVM was first introduced by Shigee Abe in 2001 to resolve the unclassifiable regions problem. Then, in 2002, Fuzzy SVM used in different ways by Chun-Fu Lin to reduce the effect of outliers in model generation.

In conventional SVM, the sample points are treated equally, including outlier and each point is assumed to be belonged to one class. However, in many real-world problem, some input points may not exactly assigned to one of these classes and different points have different contribubon to the decision surface. To solve this problem, fuzzy membership is introduced by assigning a membership value $\mu_i$ to each training point $x_i$. The higher value of $\mu_i$ makes the corresponding point $x_i$ less important in the training, and vice versa.

Suppose there are $n$ dataset with associated fuzzy membership $(x_1, y_1, \mu_1), \ldots, (x_n, y_n, \mu_n)$, where $x_i \in \mathbb{R}^n$ is belongs to one of the two classes labeled by $y_i \in \{+1, -1\}$ which has fuzzy membership $t \leq \mu_i \leq 1$ with sufficient small $t > 0$. Then, the separating hyperplane can be find by formulating

$$\min \frac{1}{2} \|w\|^2 + C \left( \sum_{i=1}^{n} \mu_i \xi_i \right),$$

s.t $y_i(x_i, w + b) \geq 1 - \xi_i,$

$\xi_i \geq 0.$

Furthermore, the problem can be transormed into Lagrange function as follow

$$\max L_D \equiv \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1, j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j,$$

s.t $\sum_{i=1}^{n} \alpha_i y_i = 0, 0 \leq \alpha_i \leq C.$

In conventional SVM, there is only free parameter $C$ which is used to control the trade-off between maximizing margin and minimizing the amount of misclassification. A larger $C$ makes the training SVM less misclassification and the margin will be narrower. The decrease of $C$ makes SVM ignore more training points and get wider margin.

The greater the value of $C$, then the fewer the number of errors and the margin will be smaller. Conversely, the smaller the value of $C$ will cause the number of errors and the margin even greater.

In fuzzy SVM, the number of free parameters is equivalent to the number of training points because each training points is given a membership value to control the trade-off. The membership value is usually calculated based on the main property of dataset. If suppose we use fuzzy membership as function of the distance between the point and its class center, then the following formula can be use

$$\mu_i = \begin{cases} 
1 - |x_i - x|/(r + \delta) & \text{jika } y_i = 1, \\
1 - |x_i - x|/(r + \delta) & \text{jika } y_i = -1, 
\end{cases}$$

where $x_i$ is the mean of class +1, $x$ is the mean of class -1, and $\delta > 0$ to avoid the case $\mu_i = 0$. $r$ as the radius of the class +1 is obtained through

$$r_+ = \max_{x, y = 1} |x_i - x|,$$

whereas $r_-$ is the radius of the class -1 is obtained through

$$r_- = \max_{x, y = -1} |x_i - x|.$$

### 3. Proposed Membership Function for Fuzzy SVM

This section will describe the proposed membership function for Fuzzy SVM. The different membership functions have different effect to the algorithm (Qu, 2008).

In this paper, we introduce a new membership function as combining of distance feature, correlation and the depth of each point. Then these feature are integrated into exponential function. The distance concept is used to identify the overlapping between classes, while the correlation and the depth is used to represent the similarity between points.

In generating membership function, the distance feature and correlation is calculated relative to its class center, so its absolutely requires numerical data. Therefore, this process is begins by ensuring all features have numerical type. We carried out the missing value, which is usually express as question mark (?) in UCI Machine Learning data, by applying the imputation method. In this paper, we calculate the mean value of missing attribute in the same class to substitute the missing value. However, in large dataset, the observation of missing values can be ignored because a similar observation can be found from another data (Santosa, 2007).

Then, we begin the fuzzy membership calculation for each feature.

#### 3.1. Distance

In most paper, fuzzy membership often define as a function of distance between each sample point to its class center (Chun-Fu Lin, 2002). This is one of important basis for measuring contributions which sample point have (Hao, 2008).

In this paper, we use this concept to identify the overlapping between class. Therefore, we are not just only
computed each point to its class center, but to all class center. These concept is shown in figure 3.

![Fig. 3. The distance concept](image)

Figure 3, shows the observations to the class+, Point A is the farthest point of class+, while point B is the closest point from another class (class -) to class+ center. Thus, the distance between the class+ center (x+) and point A is called as maximum radius (r1), whereas the distance to point B is called as minimum radius (r2). If r1 > r2 means there is overlapping between classes. So, the fuzzy membership function based on distance ($\mu_1(x)$) is defined as

$$
\begin{cases}
1, & \text{if } ||x_+ - x|| < r_2, \\
1 - ||x_+ - x||/r_1, & \text{if } r_2 \leq ||x_+ - x|| \leq r_1.
\end{cases}
$$

(14)

In this paper, the class centers are calculated by Hodges-Lehmann (HL) method. This method is a robust statistical method to find the class center based on median of the average data pairs, as expressed as follow

$$HL = \text{median} \frac{x_i + x_j}{2}, 1 \leq i \leq j \leq n,$$

(15)

where n is the number of data. HL estimator is a more reliable than using the mean value. The mean is very sensitive to outliers because this function calculates the average value of an entire class. For example, if there is a point which have a wide range value compared with other data, then the average value of the class can be much different. Meanwhile, HL estimator determines the location of the center of classes based on ranges of data values in a class. The outermost data in a class is found by using the convex hull method (Eddy, 1977)

3.2. Similarity

The similarity is the second variable used in generating the fuzzy membership. According to (Badrul Sarwar, 2001), one way to measure the similarity of two data is using the Pearson correlation

$$\text{sim}(x_i, x_j) = \text{corr}(x_i, x_j)$$

$$= \frac{\sum(x_i - \bar{x}_i)(x_j - \bar{x}_j)}{\sqrt{\sum(x_i - \bar{x}_i)^2 \cdot \sum(x_j - \bar{x}_j)^2}}$$

(16)

The similarities are generally compared the data with its class center. The data with low similarity will be considered as outlier. In this paper, we also use correlation to find the similarity between each point to its class center. The fuzzy membership function based on similarity ($\mu_2(x_i)$) is defined as

$$
\begin{cases}
1 & \text{if } \text{abs}(\text{corr}(x_i, x_+)) \geq 0.9, \\
\text{abs}(\text{corr}(x_i, x_+)) & \text{if } \text{abs}(\text{corr}(x_i, x_+)) < 0.9.
\end{cases}
$$

(17)

In the above membership function, any point that has correlation coefficient more than 0.9 means having a very strong relationship (Wikipedia). Therefore, we define the degree membership for such point is 1. Otherwise, correlation will be close to 0 if there is no relationship exists.

3.3. Depth

In this paper, we use convex hull peeling to identify the depth of each point in its class. This method is executed with a simple algorithm. Suppose we are given a set S of n points. First, calculate the convex hull (H) from S and give a value of 1 at each vertex, so it remains S’ as a set in the hull. Then calculate the convex hull of set S’ and give a value of 2 (previous value + 1) at each vertex. This process is done iteratively until no more point inside a hull. The values given in each iteration express the depth of data. Furthermore, the depth value will be converted into new value which range [0,1] by the following equation

$$\mu_3(x) = \frac{\text{depth}(x)}{\text{max depth}}$$

(18)

The results of the convex hull peeling is illustrated in Figure 4.

![Fig. 4. Convex Hull Peeling illustration](image)

Briefly, the algorithm of membership functions generation can be written as the following stages:

Phase 1: Check whether there is missing value in each attribute. If missing value found, then substitute by mean value of its attributes.

Phase 2: For each class, find the class center (c) and the outermost point.

Phase 3: Find the membership degree based on distance ($\mu_i(x)$).
For each class $i$, find the minimum and maximum radius. Minimum radius is the distance from the class center $c$ to the outermost point from the other class (class $j$), whereas the maximum radius is the distance from the class center to the outermost point of the class $i$. Then, calculate the distance of each point in class $i$ to its class center. Apply the membership degree according to equation 14.

Step 4: Find the membership degree based on Pearson correlation ($\mu_2(x)$).

For each point, calculate its correlation to class center and apply the membership degree according to equation 17.

Stage 5: Find the membership degree based on the depth of data ($\mu_3(x)$) by using Convex Hull Peeling (CHP).

a. Set the depth ($d = 0$)
b. Find the convex hull $v_x$ of the class $i$
c. Increase the depth value ($d = d + 1$) and apply the value to $v_x$ ($\text{depth}(x) = d$)
d. Remove the convex hull $v_x$ from the class $i$
e. If there are still datasets in class $i$, then repeat these steps b - e. Otherwise, update the depth value ($\mu_3(x) = \text{depth}(x)/d$).

3.4. Integration the fuzzy variable into the exponential function

As a final step in generating the fuzzy membership degree is integrate the values of three variables into an exponential function

$$\mu(x) = 1 - ae^{-\alpha(\mu_1(x) + \mu_2(x) + \mu_3(x))} + \sigma,$$  \hspace{1cm} (19)

where $\alpha = 1$, $\mu_1(x)$ is distance value, $\mu_2(x)$ is correlation value, $\mu_3(x)$ is depth value, and $\sigma = 0.049787$ is the lower bound. The higher value of $\mu_1(x) + \mu_2(x) + \mu_3(x)$ makes the membership degree is also higher. This function is illustrated in figure 5.

By using this exponential function, it is possible for a point that only have a maximum value (value = 1) in one of three variables to have a high degree of membership. The following table is shows the example of membership degree generated from various of maximum value.

<table>
<thead>
<tr>
<th>$\mu_1(x)$</th>
<th>$\mu_2(x)$</th>
<th>$\mu_3(x)$</th>
<th>$\mu(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.049787</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.681908</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.681908</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0.914452</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0.681908</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0.914452</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.914452</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

4. Experiments

Experiments are made on artificial data and real data which are taken from UCI Machine Learning repository. The following table describe the information about these dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#Record</th>
<th>#Attribute</th>
<th>#Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial -1</td>
<td>107</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Artificial -2</td>
<td>176</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>Soybean</td>
<td>47</td>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>Iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Glass</td>
<td>214</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>Ecoli</td>
<td>336</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>BreastCancer</td>
<td>699</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>Mammographic</td>
<td>961</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Yeast</td>
<td>1484</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>PageBlocks</td>
<td>5473</td>
<td>10</td>
<td>5</td>
</tr>
</tbody>
</table>

The experiment is done by comparing the proposed method (EFSM), conventional SVM, and another Fuzzy SVM which proposed by CF Lin (Chung-Fu Lin, 2002). We compare the best accuracy value obtained from a series of experiment with different parameter of C (1.5, 10, 100, 1000, 5000, 10000). Because of some dataset does not have balance distribution between classes, then we use proportion 50%:50% as training and testing set in each class. In training phase, we use 10-cross validation and Radial Basis Function (RBF) as kernel with $\gamma = 0.1$. The experimental result is shown in the following table.
Based on the above experimental result, the proposed method generally gives better recognition rates than SVM and fuzzy SVM proposed by CF Lin. However, for small-sized datasets and have a balanced distribution, such as Soybean, Iris, and Wine have the same recognition rate in all three algorithms. For dataset with missing value (Mammographic and BreastCancer), CF Lin method cannot be used.

Furthermore, based on these experiment result, we will evaluate the effect of the proposed membership function to the algorithm. Generally, the proposed membership functions tend to produce a high membership degree. If a training point has a high value in one of the three variables, then this point may have a high membership degree (> 0.5). Therefore, almost the entire dataset is dominated by the membership value of more than 0.7 because the correlation variables tend to produce high value. This condition is different from the membership degree resulting by CFLin method, which tends to produce varying degrees of membership [0,1], so that dominated by values between 0.7 and 0.8. Moreover, CFLin method is always considered the outer point as outlier because it has minimum membership degree (∼ 0). While the proposed method does not always detect the outliers. The illustration of membership degree differences between SVM, FSVM (CF Lin) and the proposed FSVM can be seen in Figure 6.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>SVM Recognition Rate</th>
<th>C.F. LIN Recognition Rate</th>
<th>EFSVM Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Artificial-1</td>
<td>85.71%</td>
<td>82.14%</td>
<td>92.86%</td>
</tr>
<tr>
<td>Artificial-2</td>
<td>73.86%</td>
<td>73.86%</td>
<td>79.55%</td>
</tr>
<tr>
<td>Soybean</td>
<td>95.83%</td>
<td>95.83%</td>
<td>95.83%</td>
</tr>
<tr>
<td>Iris</td>
<td>97.33%</td>
<td>97.33%</td>
<td>97.33%</td>
</tr>
<tr>
<td>Wine</td>
<td>98.89%</td>
<td>98.89%</td>
<td>98.89%</td>
</tr>
<tr>
<td>Glass</td>
<td>66.06%</td>
<td>66.06%</td>
<td>68.81%</td>
</tr>
<tr>
<td>Ecoli</td>
<td>82.94%</td>
<td>82.35%</td>
<td>84.12%</td>
</tr>
<tr>
<td>BreastCancer</td>
<td>97.71%</td>
<td>-</td>
<td>98.57%</td>
</tr>
<tr>
<td>Mammographic</td>
<td>81.50%</td>
<td>-</td>
<td>83.58%</td>
</tr>
<tr>
<td>Yeast</td>
<td>55.84%</td>
<td>55.57%</td>
<td>56.11%</td>
</tr>
<tr>
<td>PageBlocks</td>
<td>93.21%</td>
<td>93.17%</td>
<td>93.24%</td>
</tr>
</tbody>
</table>

In the figure above, the hyperplane of CF Lin method and EFSVM only slightly different, but EFSVM can more correctly classify the points around the hyperplane.

In this paper, we also analyze the experiment result based on different size of dataset and dataset distribution. In a simple dataset (small-sized dataset and few overlapping), the membership function does not significantly affect to the level of accuracy, such as soybean, iris, and wine. However, in complex dataset, this membership function was influential in the form of decision surface. Suppose in Mammographic dataset, EFSVM hyperplane was away from the outlier, while conventional SVM makes outlier as support vectors. In Ecoli and Glass datasets, EFSVM can classify 20% of data around the hyperplane.

The dataset distribution is divided into balanced and unbalanced distribution. In balanced dataset, the recognition rate is high because all classes have enough data in the training phase. While in unbalanced dataset, recognition rate depends on the class that has a lot of data because the minor classes do not have enough data for training and it makes hard to recognize. In addition, the number of minor class is usually a bit so it does not influence the recognition rate.

5. Conclusion

In this paper, we propose a new fuzzy membership function for FSVM using combination of three variables: distance, similarity, and depth of data. The value of these variables are integrated into exponential function to get
membership degree for each training point. The proposed method tends to produce high membership degree and it gives effect to the form of decision surface. Based on the experimental result, the proposed method can be used to handle missing values and gives better recognition rate.

REFERENCES


