01 Jan 2001

Advantages of using Fuzzy Class Memberships in Self-Organizing Map and Support Vector Machines

Sunghwan Sohn

Cihan H. Dagli
Missouri University of Science and Technology, dagli@mst.edu

Recommended Citation

S. Sohn and C. H. Dagli, "Advantages of using Fuzzy Class Memberships in Self-Organizing Map and Support Vector Machines," *Proceedings of the International Joint Conference on Neural Networks, 2001*, Institute of Electrical and Electronics Engineers (IEEE), Jan 2001.
The definitive version is available at https://doi.org/10.1109/IJCNN.2001.938451

Advantages of Using Fuzzy Class Memberships in Self-Organizing Map and Support Vector Machines

Sunghwan Sohn[1], Cihan H. Dagli[2]

Smart Engineering Systems Laboratory
Department of Engineering Management
University of Missouri-Rolla, MO 65409
[1]ssohn@umr.edu, [2]dagli@umr.edu

Abstract

Self-organizing map (SOM) is naturally unsupervised learning, but if a class label is known, it can be used as the classifier. In SOM classifier, each neuron is assigned a class label based on the maximum class frequency and classified by a nearest neighbor strategy. The drawback when using this strategy is that each pattern is treated by equal importance in counting class frequency regardless of its typicalness. For this reason, the fuzzy class membership can be used instead of crisp class frequency and this fuzzy-membership-label neuron provides another perspective of a feature map. This fuzzy class membership can be also used to select training samples in support vector machines (SVM) classifier. This method allows us to reduce the training set as well as support vectors without significant loss of classification performance.

1. Introduction

Self-organizing map (SOM) has the ability to represent multidimensional data and analyze attribute relationships. The main advantage of SOM is the topological mapping – i.e., after leaning, close observations are associated to the same class in the SOM network. If a class label is known, SOM can be used as the classifier. After learning the classifier output is based on a winner-take-all method. Each neuron is assigned a class label based on the maximum class frequency obtained from the training data and each pattern is classified by a nearest neighbor strategy.

However, this class labeling considers each labeled pattern as equal importance regardless of its typicalness [1]. Thus, each neuron having the same label in the SOM network represents the same category (class) even if it has the different degree of typicalness. Also, in classification, it is difficult to judge input pattern's typicalness for a given class.

For this reason, the fuzzy set theory [2] can be used to assign class memberships to the neuron in the SOM network instead of crisp class labeling and so to provide more valuable information [3].

Support vector machines (SVM) have shown attractive potential and promising performance in classification. However, it has the limitation of speed and size in training large data set.

SVM builds the decision function with only the part of training samples such as support vectors that lie closest to the decision boundary. Thus, we believe that removing any training samples that are not relevant to support vectors - i.e., samples far away from the decision boundary - might have no effect on building the proper decision function. In this way, we can use the class membership of each sample via K-nearest neighbors to select the appropriate samples and reduce the training set. The samples having the non-crisp class membership are selected as the training set. Here, the size of the training set can be also controlled by adjusting K-nearest neighbors when assigning class memberships to training samples – i.e., the smaller K is chosen, the smaller number of training set is chosen.

In the following section 2 and 3, the overview of SOM and SVM algorithm are explained respectively. In section 4, the basic concept of fuzzy memberships is discussed. In section 5, the method of assigning fuzzy class memberships into a neuron in the SOM network and sample selection method in SVM are explained. In section 6, we present experimental results using SOM with fuzzy class memberships and SVM classification with selected samples. Finally, section 7 concludes the paper with a short discussion.

2. The Self-Organizing Map Algorithm

The SOM is to transform input patterns of arbitrary dimension into a one- or two-dimensional discrete map, and to perform this in a topologically ordered fashion [4]. The following is the summary of learning algorithm [4], [5], [6]:

0-7803-7044-9/01/$10.00 ©2001 IEEE

1. Choose small random values for the initial weight vectors of neuron j, $\mathbf{w}_j(0)$
2. Draw a pattern $\mathbf{x}$ from the input space with a certain probability
3. Find the winning neuron $i(\mathbf{x})$ at n iteration by minimum Euclidean criterion
   $i(\mathbf{x}) = \arg \min_j \|\mathbf{x}(n) - \mathbf{w}_j\|, \quad j = 1,2,\ldots,l$
4. Adjust the weight vectors of all neurons by the formula
   $\mathbf{w}_j(n+1) = \mathbf{w}_j(n) + \eta(n) h_{j,i(\mathbf{x})}(n)(\mathbf{x}(n) - \mathbf{w}_j(n))$
   where $\eta(n)$ is the learning rate, and $h_{j,i(\mathbf{x})}(n)$ is the neighborhood function centered at $i(\mathbf{x})$. Here, $\eta(n)$ and $h_{j,i(\mathbf{x})}(n)$ are varied with time during learning as indicated
5. Continue with step 2 to 4 with enough iteration for weight convergence

By doing this, weight vectors moves toward the input vectors and tends to follow the distribution of input vectors. The fuzzy class labeling method used in this experiment is explained in section 5.

3. Support Vector Machines Algorithm

The theoretical basis of SVM is an implementation of structural risk minimization using the Vapnik-Chervonenkis (VC) dimension [7]. SVM constructs a hyperplane as the decision boundary in that the margin of separation is maximized. The decision boundary is basically constructed by the inner-product kernel between support vectors and input vectors. The following is the summary of the learning algorithm [4], [8], [9], [10].

For the sample $\{(\mathbf{x}_i, d_i)\}_{i=1}^N$, where $\mathbf{x}_i$ is the input vector for the i_{th} example and d_i is the corresponding desired response

1. Calculate inner product kernel
   $\mathbf{K} = \{K(\mathbf{x}_i, \mathbf{x}_j)\}_{(i,j)=1}^N$
2. Find the Lagrange multipliers $\{\alpha_i\}_{i=1}^N$ that maximize
   $$\mathrm{Q}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(\mathbf{x}_i, \mathbf{x}_j)$$
   subject to the constrains:
   $$\sum_{i=1}^N \alpha_i d_i = 0$$
   $$0 \le \alpha_i \le C \quad i = 1,2,\ldots,N$$
3. Find support vectors such that $\{\mathbf{x}_i\}_{i=1}^N$ having nonzero α_i
4. Calculate the bias
   $$b = \text{mean}\left(\sum_{i \in sv} (d_i - y_i) \right)$$
   where y_i is output of SVM for i_{th} input
5. Find optimal hyperplane
   $$\sum_{i=1}^N \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) = 0$$

In this experiment, we used the following radial basis function as the inner product kernel

$$K(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\| \mathbf{x} - \mathbf{x}_i \|^2}{2\sigma^2} \right), \quad i = 1,2,\ldots,N$$

Although SVM have provided good performance in classification, it has the limitation in training large size training data. In this reason, we can take advantage of using the class membership as the filtering method of training data.

4. Fuzzy Memberships

Given a set of pattern vector, $\{x_1, x_2, \ldots, x_n\}$, a fuzzy c partition of these vectors represents the degree of membership of each vector in each of c classes. The followings are the characteristics of a fuzzy c partition:

$$\sum_{i=1}^c u_{ik} = 1 \qquad (1)$$

$$0 < \sum_{k=1}^n u_{ik} < n \qquad (2)$$

$$u_{ik} \in [0,1]$$

where $u_{ik} = u_i(x_k)$ for $i = 1,\ldots,c$, and $k = 1,\ldots,n$.

The advantage of fuzzy membership is that the degree of membership can be specified rather than just the binary and especially advantageous if patterns are not clearly members of one class or another [1]. The following class labeling is based on this concept and it could provide the benefit in implementing SOM classifier and selecting training samples of SVM.

5. Fuzzy Class Labeling

To properly represent the class typicalness in SOM classifier, the fuzzy class membership can be counted instead of crisp class frequency after training. Then, each neuron is assigned a class label based on the maximum fuzzy class membership.

Let $X = \{x_1, x_2, ..., x_n\}$ be the set of n labeled patterns and $W = \{w_1, w_2, ..., w_p\}$ be the set of weights in the SOM network. Let $u_i(w)$ be the membership of the neuron w in the i_{th} class. $u_i(w)$ is computed by [3]

$$u_i(w) = \frac{\sum_{x_j \in WN} u_{ij} \left(1/\|w - x_j\|^2\right)}{\sum_{x_j \in WN} \left(1/\|w - x_j\|^2\right)} \tag{3}$$

where $u_{ij} = u_i(x_j)$ for $i = 1, ..., c$, and WN is the set of patterns that match to the neuron w in the nearest neighbor manner. The similar strategy of membership assignment was successfully used in a fuzzy K-nearest neighbor algorithm [1].

In the above, the initial class membership for the labeled data, u_{ij} is assigned by a K-nearest neighbor rule. The K-nearest neighbors to each pattern x (let x be in class i) are found, and then the membership in each class is assigned by [1]

$$u_j(x) = \begin{cases} 0.51 + (n_j / K) * 0.49, & \text{if } j = i \\ (n_j / K) * 0.49, & \text{if } j \neq i \end{cases} \tag{4}$$

where n_j is the number of the neighbors belonging to the j_{th} class. This labeling can fuzzify the class membership of the input sample. If the sample is near in the boundary region, it would compensate memberships with its neighbors, but if the sample is far away from this region, it would have complete memberships of its original class. Therefore, the crisp class label is fuzzified.

The method of assigning initial class membership in equation (4) is also used in selecting proper training samples in SVM classifier (section 6.2). Since the decision function of SVM classifier is constructed by only support vectors, we can eliminate samples irrelevant to support vectors from the training set and so reduce the size of training sample. With the above method of initial class labeling the sample that is near in the boundary region would have non-crisp class memberships. Also, support vectors might have non-crisp class memberships because they are near in the decision boundary region. In this reason, we can reasonably select samples having non-crisp class memberships as the training set in SVM classifier.

6. Experiments and Discussion

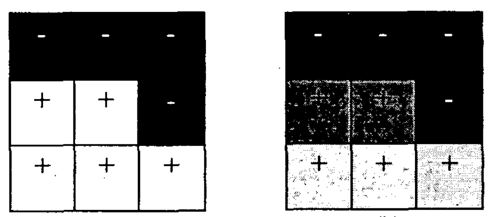
We performed two experiments. One is the SOM with fuzzy class membership (section 6.1) and another is the sample selection via class membership in SVM (section 6.2). In these experiments, we used credit approval data from UCI repository in machine learning. Credit approval data have 2 classes such as "+" and "-" with 15 features. In the data preprocessing, we transformed nominal features into integers, discard the samples containing missing values, and finally normalized the data by z-score normalization.

6.1. SOM with Fuzzy Class Memberships

Credit approval data has a lot of overlapped region in feature space. Thus, we can expect fuzzier class memberships than relatively well-separated data.

In this experiment, 70% of data was used as the training set and 30% of data was used as the test set and K=10 was used to assign initial class memberships. Both a 3-by-3 and a 4-by-4 rectangular topology were used as a lattice in feature map. In the following feature map (Figure 1 and Figure 3), the different color denotes the class label of a neuron in the SOM network and the intensity of a color denotes the degree of a class membership. The classical SOM result is also provided as the comparison.

For 3×3 rectangular lattice, the correct classification was 84.18% in both classical SOM and SOM with fuzzy class memberships. As can be seen in Figure 1. (a) classical SOM, each neuron had the same color intensity within the same class. This is because of the crisp labeling. However, in SOM with fuzzy class memberships (Figure 1. (b)), the color intensity of each neuron within the same class was different because of fuzzy class labeling – i.e., each neuron represents its own class typicalness. Especially, this trend is clear in the boundary region because of the overlapping in feature space.



(a) (b)

Figure 1: 3×3 lattice feature map. (a) Classical SOM (b) SOM with fuzzy class memberships

.04	.03	.10
.63	.61	.03
.90	.90	.86

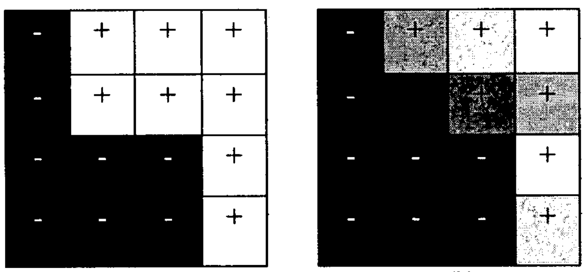
(a)

.96	.97	.90
.37	.39	.97
.10	.10	.14

(b)

Figure 2: Fuzzy class memberships of Figure 1 (b). (a) Memberships of "+" (b) Memberships of "-"

For 4×4 rectangular lattice, the correct classification was 83.67% in classical SOM and 84.69% in SOM with fuzzy class memberships. In this case, SOM with fuzzy memberships produced different class labels in the 2nd row and 2nd column neuron (Figure 3. (b)). The class label of this neuron in classical SOM was "+", but it was changed into "–" in the SOM with fuzzy memberships. The "–" class membership of this neuron is 0.57 (Figure 4. (b)) that is a little bit greater than "+". In other words, this neuron seems to have almost equal characteristics of both "+" and "–", but it has more bias on "–" based on fuzzy class memberships. From this class labeling, the correct classification was a little bit improved.



(a) (b)

Figure 3: 4×4 lattice feature map.
(a) Classical SOM (b) SOM with fuzzy class memberships

.03	.71	.85	.94
.02	**.43**	.58	.83
.03	.11	.07	.91
.11	.15	.14	.86

(a)

.97	.29	.15	.06
.98	**.57**	.42	.17
.97	.89	.93	.09
.89	.85	.86	.14

(b)

Figure 4: Fuzzy class memberships of Figure 3 (b).
(a) Memberships of "+" (b) Memberships of "-"

6.2. Sample Selection in SVM

In this experiment, we used the initial class membership in equation (4) to properly select training samples. The samples having the non-crisp class membership were selected as the training set and the samples having crisp class memberships were discarded. Here, the size of the training set could be controlled by adjusting K neighbors when assigning fuzzy class memberships into training samples.

We also used credit approval data that have 653 samples and 70% of data was used as the training set and 30% of data was used as the test set. In SVM classifier, the original SVM produced 87.04% correct classification with 457 training patterns and 379 support vectors in average of 5 random trials. The results with the different number of K neighbors are also provided in table 1:

As can be seen in table 1, SVM classifier trained with selected samples showed almost same classification performance as original SVM. In K=5, it even produced a little bit better classification performance with much smaller number of training samples. Note that the smaller K we used, the smaller number of training set we had. This is because if we choose bigger K we have more chance to have non-crisp class memberships. Also, we can notice that if K decreases, the portion of support vectors in training samples tends to increase. In other words, the smaller K, the more possibility to choose only support-vector like samples as the training set.

Table 1: Sample selection with different K
(Averaged values in 5 random trials)

K	Classification Correct (%)	# of Training Samples	# of Support Vectors
3	85.30	193	190
4	86.43	226	223
5	87.14	251	246
6	86.94	276	269
7	86.22	294	283
8	86.63	312	296
9	86.73	326	306
10	86.84	342	316

7. Conclusions

In this paper, we proposed two experiments that had advantages when using fuzzy class memberships. In SOM with fuzzy class memberships, each subcluster represented by an individual neuron could properly represent its typicalness belonging to the particular class. In credit approval data, not only we could cluster each class in a topological map but also further distinguish it based on the goodness of credit. This method is especially advantageous if the data set has a lot of overlaps. In SVM classifier, the class membership allowed us to properly select training patterns as well as to reduce support vectors. This method of sample selection is relatively simple and can speed up the training of SVM with large size of training set.

References

[1] James M. Keller, Michael R. Gary, and James A. Givens, JR, "A Fuzzy K-nearest Neighbor Algorithm," *IEEE Trans. on Systems, Man, and Cybernetics*, vol SMC-15, No. 4, July/August 1985.

[2] L. A. Zadeh, "Fuzzy Sets," *Inf. Control*, vol. 8, pp. 338-353, 1965.

[3] S. Sohn and C. H. Dagli, "Self-organizing map with fuzzy class memberships," *SPIE International Symposium on AeroSense*, vol. 4390, 2001.

[4] Simon Haykin, *Neural Networks: A Comprehensive Foundation*, Ch 9, Second Edition, Prentice Hall, 1999.
[5] T. Kohonen, "Exploration of very large databases by self-organizing maps," *International Conference on Neural Networks*, vol. I, pp. PL1-PL6, Houston, 1997.
[6] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, pp. 59-69, 1982.
[7] V.Vapnik, *Estimation of Dependences Based on Empirical Data*. New York: Springer Verlag, 1982
[8] D. P. Bertsekas, Dynamic Programming and Optical Control, vol. I and vol. II, Belmont, MA: Athenas Scientific, 1995.
[9] V. N. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.
[10] C. J .C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery* 2, pp. 121-167, 1998.