2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Support Vector Machine with SOM-based Quasi-linear Kernel for Nonlinear Classification

Yuling LIN, Yong FU and Jinglu HU

***Abstract*— This paper proposes a self-organizing maps (SOM) based kernel composition method for the quasi-linear support vector machine (SVM). The quasi-linear SVM is SVM model with quasi-linear kernel, in which the nonlinear separation hyperplane is approximated by multiple local linear models with interpolation. The basic idea underlying the proposed method is to use clustering and projection properties of SOM to partition the input space and construct a SOM based quasi-linear kernel. By effectively extracting the distribution information using SOM, the quasi-linear SVM with the SOM-based quasi-linear kernel is expected to have better performance in the cases of high-noise and high-dimension. Experiment results on synthetic datasets and real world datasets show the effectiveness of the proposed method.**

I. Introduction

Support Vector Machines (SVMs) [1] have been widely introduced in different application areas and become the state of the art. In this method one maps the data into a higher dimensional feature space and seeks an optimal separating hyperplane in this space, which can be done by the kernel trick. Nonlinear SVMs employ sophisticated kernel functions, such as polynomial functions, sigmoid function and radial basis function (RBF), etc., to fit datasets with complex decision surfaces. Unfortunately, as many other nonlinear classification methods, nonlinear kernel SVM models also face the potential over-fitting issue when facing some real world datasets with the characteristics of high-noise and high-dimensional [2], [3].

One of the ideas to overcome the overfitting problem of nonlinear SVM is to build multiple local linear classifiers because a nonlinear separation boundary can be approximately seen as an aggregation of piecewise linear boundaries. B. Chen et al. (2010) [4] proposed a quasi-linear SVM which realizes the multiple local linear model interpolated with basis function in kernel level. Thereby the quasi-linear SVM is an SVM with a composite quasi-linear kernel function. Different from conventional kernels, an interpolation basis function is designed for describing junction between multiple local linear. Thus based on the above features a quasi-linear kernel is a flexible and adjustable between the linear and nonlinear kernel function which is built by incorporating prior knowledge mined from each local training subsets.

In order to compose the quasi-linear kernel, a clustering method is needed to partition the input space along the classification boundary. The k-means clustering method is a typical one to realize the partition [4], [5]. However, the performance of the SVM is sensitive to the partition result, especially in high-noise and high-dimensional cases [6], [7], [8], [9]. Moreover, for some complicated input space, the k-means clustering method may not be suitable for the partition. Therefore it is highly motivated to develop a better way to compose the quasi-linear kernel.

The Kohonen's self-organizing map (SOM) [10] is a neural network algorithm based on unsupervised competitive learning. It has excellent capability for analyzing large and complex multivariate data. SOM attempts to address the problems of high-dimensional data and identifies the underlying patterns by reducing the dimensionality achieved through grouping of similar objects and mapping them to a low-dimensional space, usually to a two-dimensional surface also known as a topological map. Thus the properties of SOM includes clustering and projecting while trying to preserve the topological and metric relationship of the primary data space. Furthermore, visualization of complex multidimensional data is also one of the main features of the SOM. It is readily explainable, simple and easy to visualize on complex multidimensional data. Based on the above advantages, SOM has been widely applied in various engineering applications including image analysis, pattern recognition, and financial diagnosis [11], [12], [13], [14]. It has been proven to be a valuable tool in data mining and knowledge discovery in large databases (KDD) dataset [15], [16], [17], [18], [19].

This paper attempts to introduce the SOM to extract prior knowledge of the input space in order to compose a quasi-linear kernel, especially in the cases of high-noise and high-dimension. In the approach, the clustering property of SOM realizes a vector quantization (VQ) to produce a weight vector which has the same dimension as the input data. The projection property of SOM maps weight vectors to a two-dimension topological map, generating a 2-D feature vector. The close samples in the input space will be assigned with the same weight vector or similar weight vector. Thereby the close samples are mapped to the same position or to close position on the topological map also. Based on properties of SOM the distribution information of high-noise and high-dimensional dataset can be extracted effectively. Based on the trained SOM, local training subsets is obtained by simply partitioning the topological map. The information of local subsets is utilized to compose the quasi-linear kernel. Finally SVM is trained with the SOM based quasi-linear kernel.

A series of experiments are carried out to test the effectiveness of the proposed method, using a set of synthetic datasets and real world dataset. Experimental results are compared with SVMs with other kernels and k-means based quasi-

Yuling LIN, Yong FU and Jinglu HU are the Graduate School of Information, Production and Systems, Waseda University, JAPAN (email: {yuling-lin@fuji.waseda.jp. jinglu@waseda.jp.

linear SVM approaches.

The rest parts of the paper are organized as follows: Section 2 describes the quasi-linear SVM and SOM algorithm. Section 3 introduces the proposed SVM with SOM based quasi-linear kernel in detail. Section 4 describes numerical simulations to show the effectiveness of the proposed method. Finally, Section 5 presents discussions and conclusions.

II. Quasi-Linear SVM

A. SVM with quasi-linear kernel

In the case of binary classification problem, suppose we have the following labeled training data points of N samples $(x_1, y_1), \ldots, (x_k, y_k), \ldots, (x_N, y_N)$, $x_k \in R^d$ is the input vector corresponding to the kth sample labeled by $y_k \in \{-1, +1\}$ depending on its class.

The task of a classifier is to learn the discriminant or separating boundary between the two classes. Let us consider a classifier whose classification boundary can be described by

$$f_p(x) = g(x), \qquad x \in R^d \tag{1}$$

where $g(x)$ is a nonlinear function. Applying Taylor expansion to the nonlinear function $g(x)$ around the region $x = 0$,

$$f_p(x) = g(0) + \left(g'(0) + \frac{1}{2}x^T g''(0) + \ldots\right) x \tag{2}$$

and defining a vector by

$$\theta(x) = \left(g'(0) + \frac{1}{2}x^T g''(0) + \ldots\right)^T, \tag{3}$$

then we get a regression form of the classifier

$$f_p(x) = g(0) + x^T \theta(x) \tag{4}$$

where $\theta(x)$ is a vector of unknown functions of x.

By parameterizing the unknown vector $\theta(x)$ using basis function based networks

$$\begin{aligned} \theta(x) &= \sum_{j=1}^{M} \Omega_j R_j(x) + \Omega_0 \\ g(0) + x^T \Omega_0 &\Rightarrow \sum_{j=1}^{M} b_j R_j(x) + b, \end{aligned}$$

we express the classifier as one consisting of M local linear classifiers with interpolation, see Fig.1, defined by

$$f_p(x) = \sum_{j=1}^{M} (\Omega_j^T x + b_j) R_j(x) + b \tag{5}$$

where Ω_j ($j = 0, 1, ..., M$) and b_j ($j = 1, ..., M$) are the parameters of jth local linear classifier, b the constant parameter, $R_j(x)$ the basis function.

Introducing two parameter vectors $\Phi(x)$ and Θ, defined by

$$\begin{aligned} \Phi(x) &= [R_1(x), x^T R_1(x), \ldots, R_M(x), x^T R_M(x)]^T \\ \Theta &= [b_1, \Omega_1^T, \ldots, b_M, \Omega_M^T]^T. \end{aligned} \tag{6}$$

the Eq.(5) can be further expressed as:

$$f_p(x) = \Theta^T \cdot \Phi(x) + b. \tag{7}$$

Based on the procedure of a standard SVM, we apply the structural risk minimization principle to the Eq.(7), then we have a QP optimization problem, described by

$$\min_{w,b,\xi} \; \frac{1}{2}\Theta^T\Theta + C\sum_{i=1}^{N} \xi_i \tag{8}$$

$$\text{s.t.} \begin{cases} y_i[\Theta^T \cdot \Phi(x_i) + b] \geq 1 - \xi_i \\ \xi_i \geq 0, \; i = 1, 2, \ldots, N \end{cases}$$

The QP optimization problem can be computed by Lagrangian algorithm. Then the dual formulation is obtained by solving Lagrangian algorithm

$$\max_{\alpha} \mathcal{J}(\alpha) = \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \mathcal{K}(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \tag{9}$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{N} \alpha_i y_i = 0 \\ 0 \leq \alpha_i \leq C, \; i = 1, \ldots, N \end{cases}$$

where α_i is Lagrange multiplier, which can be obtained in Eq.(9). The x_i for which $\alpha_i > 0$ are called support vectors; they contribute to the geometric location of the margin hyperplane.

The composite quasi-linear kernel is defined as

$$\begin{aligned} \mathcal{K}(x_i, x_j) &= \Phi^T(x_i)\Phi(x_j) \\ &= (1 + x_i^T x_j) \sum_{l=1}^{M} \mathcal{R}_l(x_i)\mathcal{R}_l(x_j). \end{aligned} \tag{10}$$

Finally, the model $f_p(x)$ can be identified as nonlinear SVM classifier with the composite quasi-linear kernel.

$$y = \text{sign}[\sum_{i=1}^{N} \alpha_i y_i \mathcal{K}(x_i, x) + b] \tag{11}$$

B. Self-Organizing Maps (SOM)

The SOM algorithm which is an unsupervised and competitive learning algorithm was originally introduced by the Kohonen in 1982 [20]. SOM can be used as clustering tools due to that it can convert the non-linear statistical relationship of high dimensional data into simple geometric relationships on a low-dimensional display. By that way, the data points which possess similar properties are placed close to each other within the output of SOM. Fig.1 shows a visual representation of SOM model, which is made up of two layers. Let X be the number of samples in the input layer, and let $n_x * n_y$ be the number of nodes in the output layer which are arranged in a rectangular pattern with x rows and y columns, which is called topological map. Each sample in the input layer is connected to each node in the output layer. Thus, each node of topological map is associated with X weight vectors and X two-dimension (2-D) feature vectors. Let W_{ij} the weight associated with the connection between input sample i and output node j, and it has the same

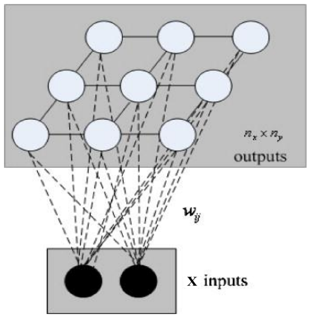


Fig. 1. A structure of SOM

dimension with input vector. The 2-D feature vector which is two dimensions is the location of the node on the topological map. SOM is also a vector quantization (VQ) algorithm. SOM is thus an algorithm that combines these two tasks: VQ and vector projection.

The VQ learning algorithm of SOM updates not only the weights of the winning node, but also those of the spatially-close nodes, based on the activation zone for each node. The training steps, including competition and weight adaptation processes, are repeated until the stopping criteria are met, which can be concluded as follows:

Step 1: Select the size and structure of the topological map. Initialize weight vectors $W_{ij}(t)$ with small random values. N is the total number of nodes in the topological map. t is the index of the learning step.

Step 2: In the competitive process, compute Euclidean distance between input vector and all nodes of topological map Eq.(12).

$$d_i = \sum_{j=1}^{K} \|X_j(t) - W_{ij}(t)\|, \quad i = 1, 2, \ldots, N \tag{12}$$

where K is the dimension of the input vector $X(t)$.

Step 3: Find the best match unit (BMU) based on the distance d_i Eq.(13).

$$d_i^* = \min_i d_i, \quad i = 1, 2, \ldots, N \tag{13}$$

Step 4: In the weight adaptation process, update the new weights at time $t+1$ of the BMU and its neighbourhood nodes defined by the activation zone.

$$W_i(t+1) = W_i(t) + \eta(t)\pi_{i,i^*}(t)(X(t) - W_i(t)) \tag{14}$$

where $\eta(t)$ is the learning rate and $\pi_{i,i^*}(t)$ is the topological neighbourhood depending on lateral distance between the node i and the BMU i^*. A Gaussian neighbourhood function centered at BMU i^* is defined by Eq.(15):

$$\pi_{i,i^*}(t) = \exp(-\frac{\|P_i - P_i^*\|}{g(t)}) \tag{15}$$

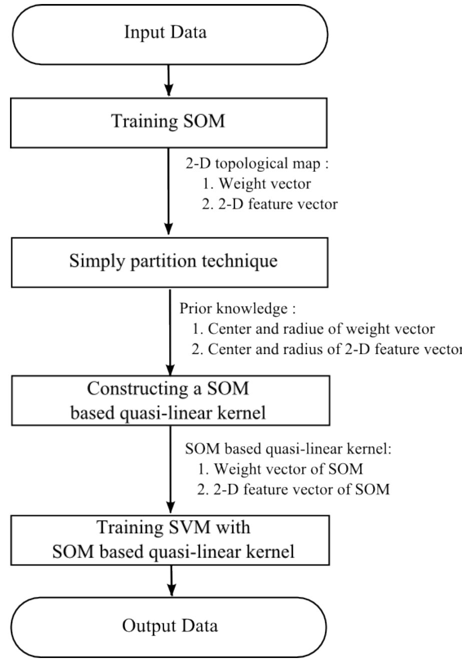


Fig. 2. General framework of the SOM based quasi-linear SVM

where P_i is the position vector of the neighbourhood node, the P_i^* is the BMU, and $g(t)$ is a parameter which is gradually decreased.

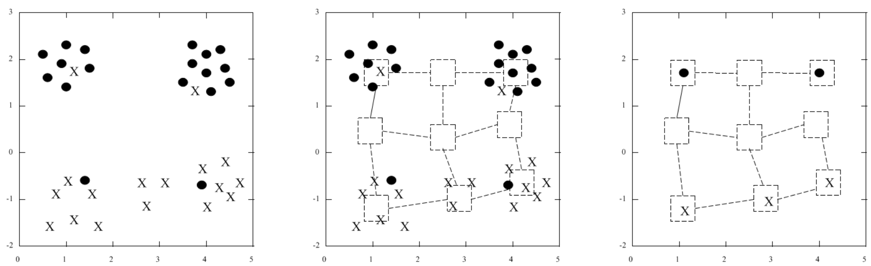
Step 5: Repeat by going to step 2 until the change in the wight vectors is less than a predetermined threshold or the maximum number of iterations is reached.

III. SOM BASED QUASI-LINEAR KERNEL

A. Framework of SVM with SOM-based quasi-linear kernel

In this paper, the SOM based quasi-linear kernel is proposed to provide robust performance of quasi-linear SVM in high-noise and high-dimensional dataset. The basic idea underlying the proposed method is to use clustering and projection property of SOM to reduce the data noise and data dimension of original input space. Based on these two properties, the SOM effectively extracts the distribution information of high-noise and high-dimensional dataset. After training SOM, a simply partition technique is used to obtain local training subsets from two-dimensional topological map, and each subset is utilized to capture prior knowledge of corresponding local linear boundary. The center and radius of each training subsets are be treated as prior knowledge. In addition, each node is associated with the weight vector and the 2-D feature vector on 2-D topological map. Thus the quasi-linear kernel is established by weight vector of SOM and 2-D feature vector of SOM respectively. Finally SVM is trained with two kinds of SOM based quasi-linear kernel respectively. The research framework is shown in Figure 2.

Process of building SOM based quasi-linear kernel is divided into three stages. There are training SOM to produce a topological map, partition dataset on the topological map and constructing the SOM based quasi-linear kernel. The detailed processing stages are introduced next subsection.



(a) (b) (c)

Fig. 3. Using clustering property of SOM to reduce noise

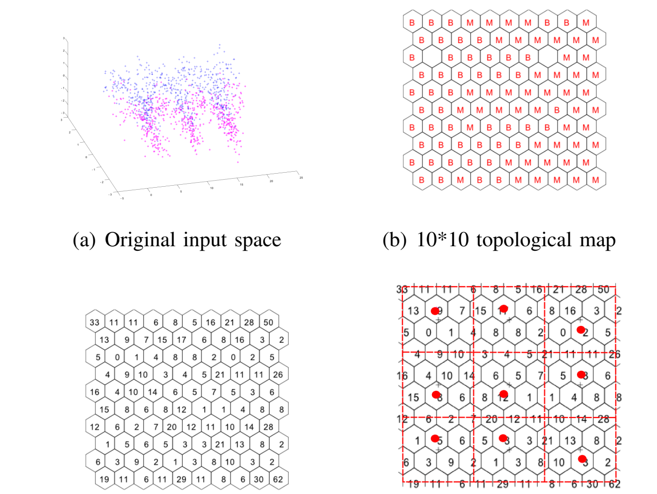
B. Composition of SOM-based quasi-linear kernel

1) Training SOM: In this study, SOM is trained for reducing the data noise and data dimension of input space. The clustering of SOM realizes a vector quantization (VQ), unlike more standard VQ algorithms, centroids are priori ordered on a two-dimension grid of nodes. Moreover during the learning phase, its implemented by competitive learning and learning algorithm uses a neighborhood function to preserve the topological properties of the input space. After learning, SOM produces a similarity graph of input space. Each node implies a cluster and it allows appearing empty node which means that not input data classify into the node. The close samples in the input space will be assigned to the same node or to close nodes on the topological map. This property can be seen as a data filtering so as to reduce noise, when setting a sufficient nodes of topological map. L.E.B. da Silva and J.A.F. Costa (2013) [21] and S. Klanke and H. Ritter (2005) [22] also applied SOM to reduce data noise.

A illustrate is given in Fig.3. There are two kinds of labels of sample which are black dots and cross. The dotted line represents a 3*3 two-dimensional topological map. Fig.3(a) exhibits a dataset with noise in the input space. Fig.3(b) shows the 3*3 two-dimensional topological map is trained to approach input space. Fig.3(c) explains the close samples in the input space are assigned to the same node or to close nodes and given a same label on the topological map. Thus the clustering property of SOM can reduce the amount of data and noise.

On the other hand, data dimension is reduced to a two dimensional topological map that is implemented by projection property of SOM. In the SOM method, each node has a weight vector and a 2-D feature vector associated with it. The 2-D feature vector is generated by projecting weight vector nonlinearly onto a two-dimensional grid of nodes. The relation of the horizontal and vertical dimensions of the grid of nodes ought to comply at least roughly with the relation of the two largest principal components of the weight vector of SOM, respectively. Therefore SOM as excellent capable for analyzing large and complex multivariate data. The Sammon's projection method [23] is used in this research.

2) Partition on the topological map: After training SOM, a simply partition technique is used to extract prior knowledge from two-dimensional topological map for constructing a composite quasi-linear kernel. In other words, it partitions training dataset to several local training subsets. Here a hit matrix of SOM is used to help partition map. Because the empty nodes should be avoided to be selected, and the node which has highest sample density near the center of local partition area should be chosen. The center and radius of each local partition area are treated as prior knowledge.



(a) Original input space (b) 10*10 topological map

(c) Hit map of SOM (d) Simple partition

Fig. 4. An example on the training SOM and partition of topological map

An example on synthetic data is used to formulate the step of training SOM and partition of topological map. As showed in Fig.4(a), a two-class dataset is constructed comprising 500 points in each class, the points are uniformly drawn from some shifted cosine signals in the three-dimensional space and perturbed with Gaussian noise. Fig.4(b) is a 10*10 topological map with B and M class labels. The hit map of SOM is showed in Fig.4(c). It represents the classification result of number of samples on each node. Fig.4(d) describes a simple partition on the topological map which is cutting into square pieces of 3 * 3.

3) Construction of the SOM-based quasi-linear kernel: According to the definition of the quasi-linear SVM described in the section 2, the quasi-linear kernel Eq.(10) is defined as inner product of an explicit nonlinear mapping. The mapping capacity of the quasi-linear kernel can be adjusted by the number of local linear classifiers M. In addition, quasi-linear kernel is built by using basis function $R_l(x)$ which is used to interpolate the piecewise linear hyperplanes, and each basis function corresponds to one local linear hyperplane. In this study, the number of local linear classifiers M and parameters of basis functions are obtained from above two steps. The Gaussian function is selected as the basis function to represent the distribution information of input space.

$$\mathcal{R}_l(x_i) = \exp(-\frac{\|x-\mu_l\|^2}{\lambda\sigma_l^2}) \tag{16}$$

where μ_l is center of the l-th data cluster, σ_l is width of the l-th data cluster, and λ is a scale parameter.

IV. Numerical Experiments

In this section, experiments are carried out to show the effectiveness of the proposed SOM based quasi-linear SVM in synthetic datasets and real world dataset classification. The results are compared with SVMs with other kernels and K-means based quasi-linear SVM.

A. Evaluation metrics

The hit ratio is used to measure the model performance, which is defined as

$$Hit\ ratio = \frac{1}{N}\sum_{i=1}^{N} H_i \tag{17}$$

where $H_i = 1$ if $MO_i = AO_i$; $H_i = 0$, otherwise. MO_i is the model output, AO_i is the actual output, and N is the number of the testing examples.

B. Research Data

1) Synthetic dataset: The two class nonlinear synthetic datasets are generated, which is normal distributed along the cosine direction. One dimension represents angle variable, the rest dimension corresponding its cosine value using different coefficients. The noise is added to sample point by normal distribution also, where $\mu = 0$ and $\delta = Z$. The noise intensity is controlled by Z. The higher disorderly data by increasing parameter of Z.

2) Financial dataset: The financial dataset is a stock market direction problem. The directions of stock price are categorized as "1" and "-1" in the research data. "1" means that the next day's stock price is higher than today's stock price, and "-1" represents that the next day's stock price is lower than today's stock price as detail refers to Y. LIN et al. (2013) [24] please.

The four nonlinear synthetic datasets and one financial dataset are applied to execute experiments, which is two classification problem. The characteristics of each dataset are showed in Table I. The 90% of the total data is utilized to train model, and test model by remainder data. Table II shows Syn2, Syn3 and Finance datasets occur over-fitting problem. Since nonlinear RBF kernel can not obtain better classification result than linear kernel.

TABLE I
Research data

Data set	Dim	Noise	Sample	Main problem
Syn 1	3	2	1000	High noise
Syn 2	100	0	1000	High dimension
Syn 3	100	3	1000	High noise and high dimension
Syn 4	3	0	1000	No noise and low dimension
Finance	53	high	900	High noise and high dimension

TABLE II
Comparing of linear kernel and RBF kernel

Kernel	Hit ratio				
	Syn 1	Syn 2	Syn 3	Syn 4	Finance
Linear	76	80	74	76	61.11
RBF	73	98	48	100	52.22

C. Experimental Results and Analysis

In the experiments, we firstly investigate the clustering property of SOM to show the effect of noise reduction by testing different size of topological map. Secondly, we compare the performance of three kinds of quasi-linear kernels to exhibit the performance of proposed SOM based quasi-linear SVM by setting different number of partition M. There are K-means based quasi-linear kernel (K-means QLK), weight vector of SOM based quasi-linear kernel (SOM-W QLK) and 2-D feature vector of SOM based quasi-linear kernel (SOM-2D QLK). In this study, the dimensions of original input space is reduced to a two-dimensional topological map. The node shape is set to hexagonal lattice. The learning rate and neighborhood range are set to 0.05 and 1.

TABLE III
Experimental results of noise reduction by clustering property of SOM

Data	Hit ratio				
	10*10	20*20	30*30	40*40	50*50
Syn 1	76	77	77	79	79
Syn 2	81	81	88	91	93
Syn 3	68	71	74	75	83
Syn 4	83	95	100	100	100
Finance	58.88	61.11	62.22	64.44	64.44

1) Experimental results of noise reduction by clustering property of SOM: In order to test effecting of clustering property of SOM with relation of noise reduction. We test different n*n size of topological map and different noise level datasets, as shown in table III. The weight vector of SOM based quasi-linear kernel is used here. Since the clustering property of SOM realizes the vector quantization (VQ) which

TABLE IV
EXPERIMENTAL RESULTS OF DIFFERENT PARTITION NUMBER

Data	kernel	Number of partition (%)							
		4	9	16	25	36	49	64	81
Syn 1	K-means QLK	75	74	74	75	75	76	73	73
	SOM-W QLK	75	75	75	75	79	79	79	80
	SOM-2D QLK	74	74	74	75	75	75	75	75
Syn 2	K-means QLK	80	79	80	82	82	81	84	86
	SOM-W QLK	80	88	89	93	91	91	92	92
	SOM-2D QLK	82	81	80	79	79	82	82	82
Syn 3	K-means QLK	74	74	74	74	74	75	70	65
	SOM-W QLK	75	75	76	75	76	77	76	79
	SOM-2D QLK	75	75	76	77	77	79	80	80
Syn 4	K-means QLK	83	98	99	100	100	100	100	100
	SOM-W QLK	82	83	84	95	95	100	99	100
	SOM-2D QLK	79	80	79	79	79	81	81	82
Finance	K-means QLK	61.11	60	60	61.11	63.33	62.22	59.17	59.17
	SOM-W QLK	54.44	57.78	62.22	64.44	64.44	64.44	66.67	66.67
	SOM-2D QLK	61.11	62.22	62.22	62.22	61.11	62.22	62.22	62.22

is without reduces dimension. Number of partition on the topological map is 36.

In Table III, it is easy to see larger topological map provides higher accuracy of the model. One reason is that excessive input item are overlapped to a small size of topological map. It can not preserves the topological and metric relationship of the original data space effective. Therefore, a larger size of map is needed for sufficient resolution. In other word, reducing data noise by setting a sufficient nodes of topological map.

2) Experimental results of SOM based quasi-linear SVM: We investigate the performance of proposed SOM based quasi-linear SVM by comparing with K-means based quasi-linear kernel in terms of testing different number of partition in four kinds of nonlinear synthetic and one financial datasets. The relative performance of the models is measured by hit ratio, which is the percentage of the total correct of trend prediction. The experiment results are shown on Table IV. The size of topological map is set 50*50.

In Table III, the SOM-W QLK and SOM-2D QLK reaches higher performance in Syn 1, Syn 3 and Finance datasets. The common characteristics of these two datasets is high noise. In other words, proposed SOM based quasi-linear SVM can effective deal with high noise data than K-means based quasi-linear SVM.

On the other hand, the K-means QLK has high accuracy than SOM-2D QLK on Syn 2 and Syn 4 datasets. These two datasets without adding noise and it not occurs over-fitting problem, see Table II. Thereby K-means based quasi-linear SVM can works will and stable on noise-free dataset.

From Syn 2 and Syn 4 datasets we also know when number of partition increases the accuracy increases too. It means the separating boundary of SVM is quite soft. Comparing with results of Table II, the performance of RBF kernel is highest among all five kernels. Thus RBF kernel has high performance on non-overfitting data. As we know RBF kernel can approaches a very soft separating boundary of SVM.

The characteristics of Syn 3 and Finance datasets are high noise and high dimensional. In Syn 3, the SOM-2D QLK has highest performance of all five kernels. In Finance dataset, SOM-W QLK obtains highest performance. Thus the proposed SOM based quasi-linear SVM can effective deal with high noise data and high dimensional dataset. It worth to have further study to verify the function of dimensional reduction of SOM in real world datasets.

V. CONCLUSIONS

In this paper, a SOM-based quasi-linear SVM is proposed to provide better performance in high-noise and high-dimensional dataset. The SOM-based quasi-linear SVM is an quasi-linear SVM with the SOM-based quasi-linear kernel function. The clustering and projection properties of SOM is used to construct a SOM-based quasi-linear kernel function. Based on these two properties, the distribution information of high-noise and high-dimensional input space is extracted effectively. After training SOM, a simply partition technique is used to catch local distribution information subsets from topological map. Then a SOM-based quasi-linear kernel function is built by local distribution information from SOM. Finally quasi-linear SVM is trained by SOM-based quasi-linear kernel function.

The proposed approach was tested on four synthetic datasets and one real world dataset. Simulation results exhibit that our proposed method produces robust performance which are outperform other comparable methods in high-noise and high-dimension datasets.

REFERENCES

[1] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.

[2] Z. Barutcuoglu, R. E. Schapire, and O. G. Troyanskaya, "Hierarchical multi-label prediction of gene function," *Bioinformatics*, vol. 22, no. 7, pp. 830–836, 2006.

[3] S. Diplaris, G. Tsoumakas, P. Mitkas, and I. Vlahavas, "Protein classification with multiple algorithms," *Proc. of the 10th Panhellenic Conference on Informatics (PCIf05)(Volos, Greece)*, pp. 448–456, 2005.

[4] B. Chen, F. Sun, and J. Hu, "Local linear multi-SVM method for gene function classification," in *Proc of the World Congress on Nature and Biologically Inspired Computing (NaBIC'10) (Kitakyushu)*, Dec. 2010, pp. 183–188.

[5] Y. Cheng and J. Hu, "Nonlinear system identification based on svr with quasi-linear kernel," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*. IEEE, 2012, pp. 1–8.

[6] B. Dash, D. Mishra, A. Rath, and M. Acharya, "A hybridized k-means clustering approach for high dimensional dataset," *International Journal of Engineering, Science and Technology*, vol. 2, no. 2, pp. 59–66, 2010.

[7] N. Asgharbeygi and A. Maleki, "Geodesic k-means clustering," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[8] V. Estivill-Castro and J. Yang, "Fast and robust general purpose clustering algorithms," in *PRICAI 2000 Topics in Artificial Intelligence*. Springer, 2000, pp. 208–218.

[9] G. H. Ball and D. J. Hall, "A clustering technique for summarizing multivariate data," *Behavioral science*, vol. 12, no. 2, pp. 153–155, 1967.

[10] T. Kohonen, *Self-organizing maps*. Springer, 2001.

[11] S. B. Katwal, J. C. Gore, R. Marois, and B. P. Rogers, "Unsupervised spatiotemporal analysis of fmri data using graph-based visualizations of self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 9, pp. 2472–2483, 2013.

[12] K. Yin and L. Gang, "Fault pattern recognition of thermodynamic system based on som," in *Electrical and Control Engineering (ICECE), 2010 International Conference on*. IEEE, 2010, pp. 3742–3745.

[13] F. E. H. Tay and L. J. Cao, "Improved financial time series forecasting by combining support vector machines with self-organizing feature map," *Intelligent Data Analysis*, vol. 5, no. 4, pp. 339–354, 2001.

[14] G. J. Deboeck, "Financial applications of self-organizing maps," *Neural Network World*, vol. 8, no. 2, pp. 213–241, 1998.

[15] J. Vesanto, "Som-based data visualization methods," *Intelligent data analysis*, vol. 3, no. 2, pp. 111–126, 1999.

[16] C. Brighenti and M. Á. Sanz-Bobi, "Auto-regressive processes explained by self-organized maps. application to the detection of abnormal behavior in industrial processes," *Neural Networks, IEEE Transactions on*, vol. 22, no. 12, pp. 2078–2090, 2011.

[17] K. Tasdemir and E. Merényi, "Exploiting data topology in visualization and clustering of self-organizing maps," *Neural Networks, IEEE Transactions on*, vol. 20, no. 4, pp. 549–562, 2009.

[18] A. Rauber, D. Merkl, and M. Dittenbach, "The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data," *Neural Networks, IEEE Transactions on*, vol. 13, no. 6, pp. 1331–1341, 2002.

[19] A. Konig, "Interactive visualization and analysis of hierarchical neural projections for data mining," *Neural Networks, IEEE Transactions on*, vol. 11, no. 3, pp. 615–624, 2000.

[20] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

[21] L. E. B. da Silva and J. A. F. Costa, "Clustering, noise reduction and visualization using features extracted from the self-organizing map," pp. 242–251, 2013.

[22] S. Klanke and H. Ritter, "Psom+: Parametrized self-organizing maps for noisy and incomplete data," in *Proceedings of the 5th Workshop on Self-Organizing Maps (WSOM 05), Paris, France*, 2005.

[23] J. W. Sammon Jr, "A nonlinear mapping for data structure analysis," *Computers, IEEE Transactions on*, vol. 100, no. 5, pp. 401–409, 1969.

[24] Y. Lin, H. Guo, and J. Hu, "An svm-based approach for stock market trend prediction," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*. IEEE, 2013, pp. 1–7.