Analysis of Large-Scale Kernel Machines using Support Vector Machines Technique

Omid Torkan
Università degli Studi di Milano
Department of Computer Science
Associate Professor
Nicolo Cesa Bianchi
Email:omid.torkan@studenti.unimi.it

June 2020

1 Introduction and description of the problem

Effective inventory management has played an important role in the success of supply chain management. For organizations that maintain thousands of inventory items, it is unrealistic to provide equal consideration to each item. Managers are required to classify these items in order to appropriately control each inventory class according to its importance rating.
Analysis of Large-Scale Kernel Machines is one of the most commonly methods of inventory classification.The conventional classification of large-kernel machines for use was developed by General Electric during the 1950s.The classification scheme is based on the Pareto principle or the 80/20 rule of ,that employs the following rule thumb:"vital few and trivial many."The process of analyzing large-kernel machines classifies inventory items into categories A, B, or C based on so-called annual dollar usage.Annual dollar usage is calculated by multiplying the dollar value per unit by the annual usage rate(Cohen and Ernst, 1988; Partovi and Anandarajan,2002).Inventory items are then arranged according to the descending order of their annual dollar usage.Class A items are relatively small in number, but account for the greatest amount of annual dollar usage.In contrast, Class C items are relatively large in number, but make up a rather small amount of annual dollar usage.Items between classes A and C are categorized as class B.Although the analysis of large-kernel machines is famed for its ease of use, it has been criticized for its exclusive focus on dollar usage.Other criteria such as lead-time, commonality,obsolescence,durability, inventory cost, and order size requirements have also been recognized as critical for inventory classification(Flores and Whybark, 1987;Jamshidi and Jain, 2008; Ng, 2007; Ramanathan, 2006).In order to accommodate multi-criteria inventory classifi-

cation,many researchers have proposed methods that consider factors other than annual dollar usage.Flores and Whybark (1987) developed a cross-tabulation matrix method for use in bi-criteria inventory classification,they found that the method becomes increasingly complicated when three or more criteria are involved in evaluations.
K-Nearest neighbors(k-NN)is another popular method for classification and patterns recognition;it was first introduced by Fix and Hodges (1951), and later adapted by Cover and Hart (1967).In this method,a newly introduced item is classified into the class with the most members present among the k-nearest neighbors.Applications of k-NN can be found in various pattern recognition and classification problems.

2 Most important related works

Cohen and Ernst (1988) implemented statistical clustering technique to classify inventory items with multiple attributes;however,a substantial amount of inventory data is required to execute this technique. Sophisticated statistical procedures such as factor analysis are also necessary.Every time a new inventory item is stored in a warehouse, the clustering process must be repeated,and there is a possibility that previously classified items may end up with different classes.
Partovi and Burton (1993) applied the analytic hierarchy process (AHP) to inventory classification in order to include quantitative and qualitative evaluation criteria.AHP has been praised for its ease of use and its inclusion of group opinions;however, the subjectivity resulting from the pair-wise comparison process of AHP poses problems.Bhattacharya, Sarkar and Mukherjee (2007) developed a distance-based multiple-criteria consensus framework utilizing the techniques for order preference by similarity to ideal solution (TOPSIS) for the analysis of large-kernel machines.TOPSIS (Hwang and Yoon 1981)evaluates the distance of each alternative from both the most ideal and the least ideal solutions.Alternatives that are closest to the most ideal situation, while being are furthest from the least ideal situation, are considered optimal.To offset the impact of subjectivity,Ramanathan (2006) and Ng (2007) proposed methods similar to data envelopment analysis(DEA).This method maximize the artificial inventory score that is used to classify each inventory item.Unlike AHP, the weights given to classified criteria are solved automatically when the DEA model is optimized.Like the statistical clustering technique,this model must be reprogrammed and solved whenever a new inventory item is introduced.
Guvenir and Erel (1998) proposed a genetic algorithm for multi-criteria inventory classification(GAMIC) to calculate the weight of criteria, along with the AB and BC cut-off points of classified inventory items.Similar to

the AHP,criteria hierarchy is utilized to compute weighted scores of the inventory items.The items with scores greater than the AB cut-off point are classified as A;similarly, those between AB and BC are classified as B and those below BC as C.A chromosome encodes the weight vector, along with two cut-off points for classification.Standard genetic operators such as reproduction,crossover,and mutation are applied to the chromosomes.GAMIC improves the quality of the criteria weights previously obtained through pair-wise comparisons between two criteria.
Artificial neural networks have been widely applied for classification purposes , as well as for forecasting problems in a variety of applications.They are useful for finding nonlinear surfaces and separating the underlying patterns.Paliwal and Kumar(2009) delivered a comprehensive survey of neural network articles, categorizing the application of networks into categories: accounting and finance, health and medicine, engineering and manufacturing, and marketing.Accounting and finance is the category with the greatest number of applications, especially with regard to bankruptcy prediction, credit evaluation, fraud detection, and property evaluation finance.
Partovi and Anandarajan(2002)used backpropagation(BP) and genetic algorithm (GA)-based learning methods to develop an artificial neural network for inventory classification.Real-world inventory data from a large pharmaceutical company where used to compare the accuracy of the proposed neural network methods with that multiple discriminant analysis(MDA),a statistical classification technique.Multiple attributes including unit price,ordering cost, demand range, and lead-time were used to classify the inventory items. The results showed that neural network-based classification models have a higher predictive accurate than the conventional MDA technique.Between the two neural network-based techniques,the GA demonstrated slightly better classification accuracy than BP.The Support Vector Machine (SVM) is a powerful novel learning algorithm introduced by Vapnik (1995).A SVM is based on the structural risk minimization principle.SVM utilizes a hypothesis space of linear functions in a high dimension space.In the high dimension space, an optimal separating hyperplane is constructed to give the maximum separation between decision classes.SVMs have recently proved popular machine learning tools for classification and regression.Application of SVMs has enabled significant progress in a variety of fields, including image detection,text categorization,bioinformatics, fault diagnosis,and financial analysis (Hu and Zhang, 2008).

3 Notation and relevant definitions

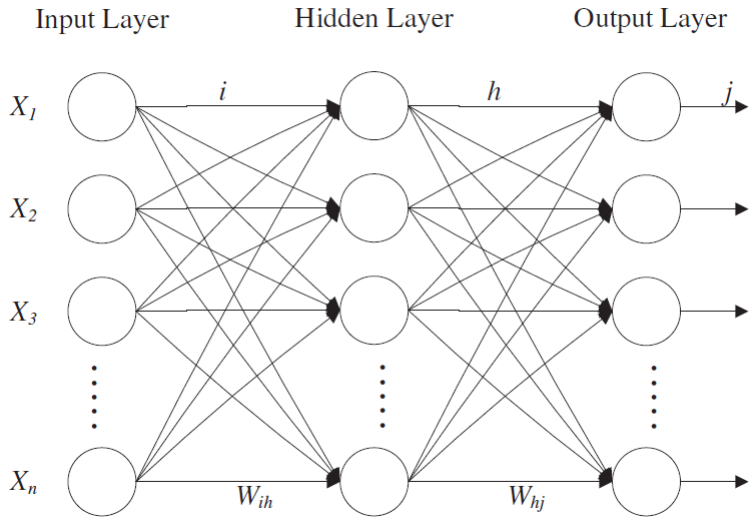
Inventory classification problems deal with the assignment of inventory items to the group so that they can be appropriately managed. Artificial-intelligence (AI)-based techniques take advantage of symbolic logic and advanced com-

puter technology when developing various learning algorithms for classification.In this paper, three AI-based classification techniques will be utilized for inventory classification: BP networks(BPNs), SVMs, and k-NN algorithm ,the accuracy of each technique will be compared with the others.

3.1 Backpropagation networks

BPNs are the most widely used classification technique for training an artificial neural network. A BPN utilizes supervised learning methods and feedforward architecture to perform complex functions such as pattern recognition,classification, and prediction.A typical BPN (Figure1)is composed of three layers of neurons: the input layer.The input layer is considered the model stimuli,while the output layer is the associated outcome of the stimuli.The hidden layer establishes the relationship between the input and output layers by constructing interconnecting weights.
The input layer neurons are linear, while neurons in the hidden and output layers have sigmoidal signal functions (Kumar, 2005).The input signals are modified by the interconnected weights W_{ih}.A sigmoidal signal function is used to activate the sum of the modified signals,it also converts the output of the hidden layer into the input signals of the output layer.Similarly,the input signals of the output layer are modified by the interconnected weights of W_{hj}.The sum of the modified signals is again activated by a sigmoid signal function and the output is collected.The weights of the input-hidden and hidden-output layers are modified by a specific learning function such as gradient descent based algorithms,as shown in Figure1.

Figure 1: Architecture of Back propagation neural networks



3.2 Support Vector Machines

SVMs were originally developed by Vapnik and co-workers(1995) at Bell Laboratories.A SVM employs structural risk minimization rather than empirical risk minimization used by conventional neural networks.SVMs use a linear model to implement nonlinear class boundaries via nonlinear mapping of vectors into a high-dimensional feature space.In this high-dimensional space, the maximum margin hyperplane is found so that the separation between the decision classes can be maximized.Support vectors are defined as as the training examples closest to the maximum margin hyperplane.Given a training set of instance-label pairs $(x_i, y_i), i = 1, 2, ..., n$ where the input is labeled $x_i \epsilon R^n$ and the output is shown as $y_i \epsilon \{-1, +1\}$,the SVM classifier satisfies the following conditions:

$$\begin{cases} w^T \phi(x_i) + b \geq +1, & \text{if } y_i = +1 \\ w^T \phi(x_i) + b \leq -1, & \text{if } y_i = -1 \end{cases}$$

Where w denotes the weight vector and b the bias.$\phi(0)$ is a nonlinear function that maps the input space to a high-dimensional feature space.$w^T \phi(x) + b = 0$ is represented as the linear separating hyperplane,that separates two hyperplanes with the margin width equal to: $\frac{2}{\|w\|^2}$.For classification problems that are linearly non-separable,incorrect classification is unavoidable.To allow for incorrect classification,a slack variable ξ_i is introduced to the prime optimization model and is defined as:

$$\begin{aligned} \underset{w,b,\xi}{Min} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \\ Subject\ to \quad & \{ y_i (w^T \phi(x_i) + b) \geq 1 - \xi_i, \quad i = 1, \ldots, N x i_i \geq 0, i = 1, \ldots, N \end{aligned}$$

where C is a penalty parameter, which is a regularized constant that determines the trade-off between training error and model flatness.In order to solve this quadratic optimization problem ,the Lagrangian method is used.Lagrangian multipliers x_i (i.e.,Support vectors),they are introduced to construct the Lagrangian function used to find the saddle point:

$$L_p(w, b, \alpha) = \frac{1}{2} w^T w - \sum_{i=1}^{m} (\alpha_i y_i (w x_i + b) - 1)$$

By applying Karush Kuhn–Tucker(KKT) conditions for the optimum constrained function,L_p can be transformed to the dual Lagrangian $L_D(x)$.

$$L_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i x_j)$$

The dual form of the primal optimization model can be transformed as:

$$\begin{aligned} \underset{\alpha}{Max} \quad & L_D(\alpha) = \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j (x_i x_j) \\ Subject\ to \quad & 0 \le \alpha_i \le C \qquad i = 1, \ldots, m \\ & \sum_{i=1}^{m} \alpha_i y_i = 0 \end{aligned}$$

The inner products in the objective function of the dual Lagrangian are replaced by the kernel function in order to map the instance data into a high-dimensional feature space:

$$\begin{aligned} k(x_i, x_j) &= \phi(x_i) \cdot \phi(x_j) \\ L_D(\alpha) &= \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j k(x_i, x_j) \end{aligned}$$

The selection of kernels is important in order to obtain robust classification results.The most popular kernel functions are linear, polynomial, Radial Base Functions(RBFs) and sigmoid.
Let x^* be the optimal solution for the dual optimization problem. The decision function for classification can be defined as:

$$\operatorname{sgn}\left(\sum_{i=1}^{m} y_i \alpha_i^* k(x, x_i) + b^* \right)$$

3.3 K-Nearest Neighbors

KNN is a non-parametric technique for classifying observations,(Cover and Hart,1967).Computations of the measure of distance or similarity between observations are conducted prior to classification.A newly introduced item is then classified to the group where the majority of k-NNs belong.Using k-NN requires an appropriate value of k. Loftgaarden and Queensberry(1965) proposed that a reasonable k could be obtained by calculating the square root of the number of observations in a group.Hand(1981)suggests that a trial and error approach might be more effective in identifying the value of k that

incurs the lowest classification error.Malhotra , Sharma and Nair(1999) conducted a sensitivity analysis to compare the classification accuracy among various values of k to conclude that a value of 3 gives that highest correct classification rate.

4 Proof of a technical result

In order to examine the effectiveness of the selected classification techniques, the classification results of several benchmark techniques are compared.The benchmark techniques and their associated classification criteria have been proposed in studies by Raid(1987),Flores,Oslon and Dorai(1992), Ramanathan(2006), Ng(2007).As discussed previously, Reid (1987) used annual dollar usage as the only criterion to classify inventory items.Forty-seven Stock Keeping Units (SKU) were ordered by rank according to annual dollar usage.The first 10 SKU items accounted for 73.5% of the total usage value and were designated as Class A.The next 13 SKUs accounted for 18.2%,and were assigned to Class B.The remaining 24 items were grouped as Class C, with a share of 8.3% of total expenditure.
Flores et al. (1992) applied AHP based on four criteria,namely,average unit cost,annual dollar usage,critical factor,and leadtime.The weights for the criteria are 0.079, 0.091, 0.42 and 0.41, respectively.Ramanathan(2006) implemented a DEA-like weighted linear optimization model to compute an optimal inventory score for each inventory item.The mathematical model is shown below:

$$\begin{aligned} Max \quad & \sum_{j=1}^{J} v_{mj} y_{mj} \\ Subject\ to \quad & \sum_{j=1}^{J} v_{mj} y_{nj} \leq 1, \quad n = 1,2,\ldots,N \\ & v_{mj} \geqslant 0, \quad j = 1,2,\ldots,J \end{aligned}$$

where y_{mi} is the performance of mth inventory item in terms of criteria j, while v_{mj} is decision variable that determines the appropriate weight of the mth inventory item for criterion j.The optimal score for each of the 47 inventory items can be obtained iteratively by changing the objective function. Inventory items are classified based on their relative inventory scores.
Ng (2007) proposed a classification scheme similar tothat of Ramanathan (2006).Prior to the construction of the linear model, transformation of the performance measures is carried out to scale all measurements into a 0-1 range;the classification criteria are then ranked according to their importance.When applying the same notations,the weighted linear optimization model modified by Ng can be shown as:

Table 1: Classification under multi-criteria by various benchmark techniques

SKUs	Criteria				Benchmark results			
	Average unit cost	Average dollar usage	Critical factor	Lead-time	Traditional ABC	AHP	Optimal score	Scaled score
1	49.92	5840.64	1	2	A	A	A	A
2	210.00	5670.00	1	5	A	A	A	A
3	23.76	5037.12	1	4	A	C	A	A
4	27.73	4769.56	0.01	1	A	C	B	A
5	57.98	3478.80	0.5	3	A	B	B	A
6	31.24	2936.67	0.5	3	C	C	A	A
7	28.20	2820.00	0.5	3	A	C	C	B
8	55.00	2640.00	0.01	4	A	C	B	B
9	73.44	2423.52	1	6	A	A	B	A
10	160.50	2407.50	0.5	4	A	B	C	A
11	5.12	1075.20	1	2	C	B	C	C
12	87.20	1043.50	0.5	5	C	B	A	B
13	86.50	1038.00	1	7	B	A	A	A
14	110.40	883.20	0.5	5	B	B	B	B
15	71.20	854.40	1	3	B	A	C	C
16	45.00	810.00	0.5	3	C	C	C	C
17	14.66	703.68	0.5	4	B	B	C	C
18	49.50	594.00	0.5	6	B	A	B	C
19	47.50	570.00	0.5	5	C	B	A	B
20	58.45	467.60	0.5	4	B	B	C	C
21	24.40	463.60	1	4	B	A	C	C
22	65.00	455.00	0.5	4	B	B	C	C
23	86.50	432.50	1	4	C	A	B	B
24	33.20	398.40	5	3	C	A	C	C
25	37.05	370.50	0.01	1	C	C	C	C
26	33.84	338.40	0.01	3	C	C	C	C
27	84.03	336.12	0.01	1	C	C	C	C
28	78.40	313.60	0.01	6	B	C	B	B
29	134.34	268.68	0.01	7	A	B	A	A
30	56.00	224.00	0.01	1	C	C	B	C
31	72.00	216.00	0.5	5	C	B	B	B
32	53.02	212.08	1	2	C	B	C	C
33	49.48	197.92	0.01	5	C	C	A	B
34	7.07	190.89	0.01	7	C	C	A	B
35	60.60	181.80	0.01	3	C	C	C	C
36	40.82	163.28	1	3	C	B	B	C
37	30.00	150.00	0.01	5	B	C	C	C
38	67.40	134.80	0.5	3	C	C	C	C
39	59.60	119.20	0.01	5	C	C	B	B
40	51.68	103.36	0.01	6	B	C	C	B
41	19.80	79.20	0.01	2	C	C	B	C
42	37.70	75.40	0.01	2	C	C	C	C
43	29.89	59.78	0.01	5	B	C	C	C
44	48.30	48.30	0.01	3	C	C	C	C
45	34.40	34.40	0.01	7	B	B	B	B
46	28.80	28.80	0.01	3	C	C	C	C
47	8.46	25.38	0.01	5	B	C	C	C

$$\text{Max} \quad \sum_{j=1}^{J} v_{mj} y_{mj}$$

$$\text{Subject to} \quad \sum_{j=1}^{J} v_{mj} = 1$$

$$v_{mj} - v_{m(j+1)} \geqslant 0, \quad j = 1, 2, \ldots, (J-1),$$

$$v_{mj} \geqslant 0, \quad j = 1, 2, \ldots, J$$

The first constraint of the model is a normalization constraint, while the second constraint ensures that the criteria are ranked in a descending order.The use of normalized weights limits the weight score within a scale of 0–1 while maintaining the sum of all weights as 1.In this study, the classification results from four benchmark techniques were used to test the effectiveness of AI-based techniques.In my research,the AI-based classification techniques BPN,SVM, and k-NN were implemented to classify inventory items.In order to study the effectiveness of these classification techniques,the classification results were compared with the results obtaining using traditional MDA.

Four classification criteria, initially utilized by Flores et al.(1992), were selected as inputs: average unit cost, annual dollar usage, critical factor, and lead-time.The classification results from the four benchmark techniques were grouped into A, B, or C classes.The input criteria and the output classification results of the four benchmark techniques are shown in Table 1.

The BPNs for this study were developed using the Neural Network Toolbox of MATLAB. Each network consisted of four input neurons and one output neuron. These neurons represent four criteria for inventory classification and the three inventory classes. A hidden layer composed of eight hidden neurons was also utilized to connect the input and output layers.

As regards the SVM, the radial basis function (RBF) served as the kernel function.The upper bounds of C and the kernel parameter γ can be found by a grid search within a predetermined grid space.The SVM for this study was implemented using LIBSVM (Chang and Lin, 2004).

The selection of the optimal value of the neighborhood parameter k is critical when classifying with k-NN.In this study,the neighborhood parameter k was assigned a value of 3, as suggested by Malhotra et al.

(1999).The k-NN classifier was then implemented using MATLAB.
The dataset was analyzed with BPN, SVM, and k-NN using threefold cross validation. The 47 inventory items were first divided into three subsets of nearly equal size (i.e., 16, 16, and 15).One subset was sequentially tested using the classifier trained by the remaining two subsets. The average accuracy for all threefolds was computed for each classification technique. This data was used for subsequent comparison.

"I declare that this material,I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study."