

OPTIMIZATION METHOD AND SYSTEM FOR SUPERVISED LEARNING UNDER TENSOR MODE

FIELD OF THE INVENTION

The present application belongs to the technical field of pattern recognition, especially to an optimization method and a system for supervised learning under tensor mode.

BACKGROUND OF THE INVENTION

With the advent of big data era, tensor expression of data has gradually become a mainstream. However, during achieving the invention, the inventor found out that the prior art still utilized vector model algorithm to process tensor data. On the basis of the concept of vector model algorithm, during a preprocessing phase, original data should be extracted (vectorization), which firstly is easy to break spatial information and inner correlation which are specific to tensor data, secondly possesses superabundant modern parameters which would easily lead to issues such as curse of dimensionality, over learning and small sample.

A plurality of tensor mode algorithms have become a trend of the era. However, solving an objective function of STM is a non-convex optimization issue, in which solving an alternative projection is required; the time complexity of the algorithm is high and a local minimum value occurs frequently.

SUMMARY OF THE INVENTION

Problem to be solved

In light of this, an embodiment of the invention provides an optimization method and a system for supervised learning under tensor mode so as to deal with issues such as curse of dimensionality, over learning and small sample occurred

when the vector mode algorithms provided by the prior art process the tensor data and overcome the shortcomings of the tensor mode algorithms of the prior art. The algorithm of the invention is aimed at dealing with limitations of the algorithms of the prior art, for example, the time complexity of the algorithms is high, and a local minimum value occurs frequently, etc.

Technical solutions

On one hand, an optimization method for supervised learning under tensor mode is provided; the method includes:

- receiving an input training tensor data set;

- introducing a within class scatter matrix into an objective function such that between class distance is maximized, at the same time, within class distance is minimized by the objective function;

- constructing an optimal frame of the objective function of an optimal projection tensor machine OPSTM subproblem;

- transforming N vector modes of quadratic programming subproblems into a multiple quadratic programming problem under a single tensor mode, and constructing an optimal frame of an objective function of an OPSTM problem;

- according to lagrangian multiplier method, obtaining a dual problem of the optimal frame of the objective function, introducing a tensor rank one decomposition into calculation of tensor transvection, and obtaining a revised dual problem;

- utilizing sequential minimal optimization SMO algorithm to solve the revised dual problem and output an alagrangian optimal combination and an offset scalar b;

- calculating a projection tensor W_* ;

- performing the rank one decomposition to the projection tensor W_* ;

- performing a back projection to a component obtained after performing the

rank one decomposition to the projection tensor $\mathbf{W}_*$;

performing rank one decomposition inverse operation to the component obtained after performing the back projection to obtain an optimal projection tensor $\mathbf{W}$ which is corresponded to the training tensor data set;

decision function construction phase: by the optimal projection tensor $\mathbf{W}$ which has been rank-one decomposed together with the offset scalar b , constructing a decision function;

application prediction phase: inputting to-be-predicted tensor data which has been rank-one decomposed into the decision function for prediction.

Furthermore, after introducing the within class scatter matrix into an objective function of an STM subproblem, through an eta coefficient η , the objection function of the quadratic programming problem of an n-th subproblem is changed into:

$$\min_{\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}} \quad \frac{1}{2} \left[\left(\|\mathbf{w}^{(n)}\|_F^2 + \eta ((\mathbf{w}^{(n)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(n)}) \right) \prod_{1 \leq i \leq N, i \neq n} \left(\|\mathbf{w}^{(i)}\|_F^2 + \eta ((\mathbf{w}^{(i)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(i)}) \right) \right] + C \sum_{m=1}^M \xi_m^{(n)}$$

wherein, $\mathbf{S}_w^{(n)}$ is an n-th order within class scatter matrix estimated after the training tensor data set is expanded along the n-th order; $\mathbf{w}^{(n)}$ is an n-th order optimal projection vector of the training tensor data, $n=1, 2, \dots, N$; C is a penalty factor; $\xi_m^{(n)}$ is a slack variable; eta coefficient η is configured to measure the importance of the within class scatter matrix.

Furthermore, the optimal frame of an objective function of an OPSTM problem is a combination of N vector modes of quadratic programming problems, which respectively corresponds to a subproblem; wherein, a quadratic programming problem of an n-th subproblem is:

$$\min_{\mathbf{w}_*^{(n)}, b^{(n)}, \xi^{(n)}} \quad \frac{1}{2} \|\mathbf{w}_*^{(n)}\|_F^2 \prod_{1 \leq i \leq N, i \neq n} \|\mathbf{w}_*^{(i)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)}$$

$$\begin{aligned} \text{S.t} \quad & y_m \left((\mathbf{w}_*^{(n)})^T \left(\mathbf{V}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}_*^{(i)} \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)} \\ & \xi_m^{(n)} \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\mathbf{w}_*^{(n)}$ is the n-th order projection vector of the training tensor data set; $\mathbf{w}_*^{(n)} = \Lambda^{(n)1/2} \mathbf{P}^{(n)T} \mathbf{w}^{(n)}$, wherein $\Lambda^{(n)}$ and $\mathbf{P}^{(n)}$ meet the equation, $\mathbf{P}^{(n)T} (\mathbf{E} + \eta \mathbf{S}_w^{(n)}) \mathbf{P}^{(n)} = \Lambda^{(n)}$; E is an identity matrix; $\mathbf{V}_m = \mathbf{X}_m \prod_{i=1}^N \times_i \left[(\mathbf{P}^{(i)} \Lambda^{(i)1/2})^{-1} \right]^T$ is tensor input data obtained after tensor input data $\mathbf{X}_m$ in the training tensor data set is projected along each order; $\mathbf{X}_i$ is an i-mode multiplication operator; $b^{(n)}$ is the n-th order offset scalar of the training tensor data set.

Furthermore, according to a formula $\|\mathbf{w}_*\|_F^2 = \prod_{n=1}^N \|\mathbf{w}_*^{(n)}\|_F^2$, and a formula $(\mathbf{w}_*^{(n)})^T (\mathbf{V}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}_*^{(i)}) = \langle \mathbf{w}_*, \mathbf{V}_m \rangle$, transforming the N vector modes of quadratic programming subproblems into the multiple quadratic programming subproblem under a single tensor mode. A constructed optimal frame of the objective function of the OPSTM problem meets that:

$$\begin{aligned} \min_{\mathbf{w}_*, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}_*\|_F^2 + C \sum_{m=1}^M \xi_m \\ \text{S.t} \quad & y_m (\langle \mathbf{w}_*, \mathbf{V}_m \rangle + b) \geq 1 - \xi_m \\ & \xi_m \geq 0 \quad m = 1, 2, \dots, M \\ & \xi_m = \max_{n=1, 2, \dots, N} \{\xi_m^{(n)}\} \end{aligned}$$

wherein, $\langle \rangle$ is a transvection operator, and

Furthermore, according to lagrangian multiplier method, an obtained dual problem of the optimal frame of the objective function is:

$$\begin{aligned}
& \max_{\alpha} \quad \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j < \mathbf{V}_i, \mathbf{V}_j > \\
& \text{S.t} \quad \sum_{m=1}^M (\alpha_m y_m) = 0 \\
& \quad 0 < \alpha_m < C \quad m = 1, 2, \dots, M;
\end{aligned}$$

introducing the tensor rank one decomposition into the calculation of the tensor transvection. An obtained revised dual problem is:

$$\begin{aligned}
& \max_{\alpha} \quad \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \sum_{p=1}^R \sum_{q=1}^R \alpha_i \alpha_j y_i y_j \prod_{n=1}^N < \mathbf{v}_{ip}^{(n)}, \mathbf{v}_{jq}^{(n)} > \\
& \text{S.t} \quad \sum_{m=1}^M (\alpha_m y_m) = 0 \\
& \quad 0 < \alpha_m < C \quad m = 1, 2, \dots, M.
\end{aligned}$$

Furthermore, calculating the projection tensor $\mathbf{W}_*$ according to a formula,

$$\mathbf{W}_* = \sum_{m=1}^M \alpha_m y_m \mathbf{V}_m.$$

On the other hand, an optimization system for supervised learning under tensor mode is provided; the system includes:

a data receiving unit configured to receive an input training tensor data set;

a within class scatter introducing unit configured to introduce a within class scatter matrix into an objective function such that between class distance is maximized, at the same time, within class distance is minimized by the objective function;

a subproblem optimal frame constructing unit configured to construct an optimal frame of an objective function of an optimal projection tensor machine OPSTM subproblem;

a problem optimal frame constructing unit configured to transform N vector modes of quadratic programming subproblems into a multiple quadratic programming problem under a single tensor mode, and construct an optimal frame

of an objective function of an OPSTM problem;

a dual problem obtaining unit configured to obtain a dual problem of the optimal frame of the objective function, introduce a tensor rank one decomposition into calculation of tensor transvection, and obtain a revised dual problem according to lagrangian multiplier method;

a dual problem solving unit configured to utilize sequential minimal optimization SMO algorithm to solve the revised dual problem and output an alagrangian optimal combination and an offset scalar b ;

a projection tensor calculating unit configured to calculate a projection tensor W_* ;

a projection tensor decomposition unit configured to perform the rank one decomposition to the projection tensor W_* ;

a back projection unit configured to perform a back projection to a component obtained after performing the rank one decomposition to the projection tensor W_* ;

an optimal projection tensor calculating unit configured to perform rank one decomposition inverse operation to the component obtained after performing the back projection to obtain an optimal projection tensor W which is corresponded to the training tensor data set;

a decision function constructing unit configured to construct a decision function construction phase and construct a decision function by the optimal projection tensor W which has been rank-one decomposed together with the offset scalar b ,

a predicting unit configured to input to-be-predicted tensor data which has been rank-one decomposed into the decision function for prediction in the application prediction phase.

Furthermore, through an eta coefficient η , after the within class scatter introducing unit introduces the within class scatter matrix into an objective function of an STM subproblem, the objection function of the quadratic

programming problem of the n-th subproblem is changed into:

$$\min_{\mathbf{w}^{(n)}, b^{(n)}, \xi_m^{(n)}} \quad \frac{1}{2} \left[\left(\|\mathbf{w}^{(n)}\|_F^2 + \eta ((\mathbf{w}^{(n)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(n)}) \right) \prod_{1 \leq i \leq N}^{i \neq n} \left(\|\mathbf{w}^{(i)}\|_F^2 + \eta ((\mathbf{w}^{(i)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(i)}) \right) \right] + C \sum_{m=1}^M \xi_m^{(n)}$$

wherein, $\mathbf{S}_w^{(n)}$ is an n-order within class scatter matrix estimated after the training tensor data set is expanded along the n-th order; $\mathbf{w}^{(n)}$ is the n-th order optimal projection vector of the training tensor data, $n=1, 2, \dots, N$; C is a penalty factor; $\xi_m^{(n)}$ is a slack variable; eta coefficient η is configured to measure the importance of the within class scatter matrix.

Furthermore, in the subproblem optimal frame constructing unit, the optimal frame of the objective function of the OPSTM problem is a combination of N vector modes of quadratic programming problems, which respectively corresponds to a subproblem; wherein, a quadratic programming problem of the n-th subproblem is:

$$\begin{aligned} \min_{\mathbf{w}_*^{(n)}, b^{(n)}, \xi_m^{(n)}} \quad & \frac{1}{2} \|\mathbf{w}_*^{(n)}\|_F^2 \prod_{1 \leq i \leq N}^{i \neq n} \|\mathbf{w}_*^{(i)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)} \\ \text{S.t} \quad & y_m \left((\mathbf{w}_*^{(n)})^T \left(\mathbf{V}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}_*^{(i)} \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)} \\ & \xi_m^{(n)} \geq 0 \quad m=1, 2, \dots, M \end{aligned}$$

wherein, $\mathbf{w}_*^{(n)}$ is the n-th order projection vector of the training tensor data set; $\mathbf{w}_*^{(n)} = \mathbf{\Lambda}^{(n)1/2} \mathbf{P}^{(n)T} \mathbf{w}^{(n)}$, wherein $\mathbf{\Lambda}^{(n)}$ and $\mathbf{P}^{(n)}$ meet the equation, $\mathbf{P}^{(n)T} (\mathbf{E} + \eta \mathbf{S}_w^{(n)}) \mathbf{P}^{(n)} = \mathbf{\Lambda}^{(n)}$; $\mathbf{E}$ is an identity matrix; $\mathbf{V}_m = \mathbf{X}_m \prod_{i=1}^N \times_i \left[(\mathbf{P}^{(i)} \mathbf{\Lambda}^{(i)1/2})^{-1} \right]^T$

is tensor input data obtained after tensor input data $\mathbf{X}_m$ in the training tensor data set is projected along each order; $\mathbf{X}_i$ is an i-mode multiplication operator; $b^{(n)}$ is the n-th order offset scalar of the training tensor data set.

Furthermore, according to a formula $\|\mathbf{W}_*\|_F^2 = \prod_{n=1}^N \|\mathbf{w}_*^{(n)}\|_F^2$, and a formula $(\mathbf{w}_*^{(n)})^T (\mathbf{V}_m \prod_{1 \leq i < N}^{i \neq n} \times_i \mathbf{w}_*^{(i)}) = \langle \mathbf{W}_*, \mathbf{V}_m \rangle$, the problem optimal frame constructing unit transforms the N vector modes of quadratic programming subproblems into the multiple quadratic programming subproblem under a single tensor mode. A constructed optimal frame of the objective function of the OPSTM problem meets that:

$$\begin{aligned} \min_{\mathbf{W}_*, b, \xi} \quad & \frac{1}{2} \|\mathbf{W}_*\|_F^2 + C \sum_{m=1}^M \xi_m \\ \text{S.t} \quad & y_m (\langle \mathbf{W}_*, \mathbf{V}_m \rangle + b) \geq 1 - \xi_m \\ & \xi_m \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\langle \rangle$ is a transvection operator, and $\xi_m = \max_{n=1, 2, \dots, N} \{\xi_m^{(n)}\}$.

Furthermore, according to lagrangian multiplier method, the dual problem solving unit obtains the dual problem of the optimal frame of the objective function, which is:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j \langle \mathbf{V}_i, \mathbf{V}_j \rangle \\ \text{S.t} \quad & \sum_{m=1}^M (\alpha_m y_m) = 0 \\ & 0 < \alpha_m < C \quad m = 1, 2, \dots, M, \end{aligned}$$

the dual problem solving unit introduces the tensor rank one decomposition into the calculation of the tensor transvection. An obtained revised dual problem is:

$$\max_{\alpha} \quad \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \sum_{p=1}^R \sum_{q=1}^R \alpha_i \alpha_j y_i y_j \prod_{n=1}^N \langle \mathbf{v}_{ip}^{(n)}, \mathbf{v}_{jq}^{(n)} \rangle$$

$$\begin{aligned} \text{S.t} \quad & \sum_{m=1}^M (\alpha_m y_m) = 0 \\ & 0 < \alpha_m < C \quad m = 1, 2, \dots, M \end{aligned}$$

Furthermore, the projection tensor calculating unit calculates the projection tensor $\mathbf{W}_*$ according to a formula, $\mathbf{W}_* = \sum_{m=1}^M \alpha_m y_m \mathbf{V}_m$.

Beneficial Effect

In the embodiments of the present invention, N vector modes of quadratic programming problems are transformed into a multiple quadratic programming problem under a single tensor mode. The transformed optimal frame of the objective function is the optimal frame of the objective function of the OPSTM problem. This can reduce the number of model parameters significantly, overcome issues such as curse of dimensionality, over learning and small sample occurred when traditional vector mode algorithms process the tensor data, which ensures high-activity processing, at the same time, highlights excellent classifying effects. Above all, the algorithms provided by the embodiments of the invention can process tensor data effectively and directly in tensor field, at the same time, possesses features of optimal classifying ability as well as strong practicability and popularization.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig.1 is an implementation flow chart of an embodiment of an optimization method for supervised learning under tensor mode of the invention;

Fig.2 is a structural block diagram of an embodiment of an optimization system for supervised learning under tensor mode of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

In order to make the purposes, technical solutions and advantages of the present invention clearer, the invention is described hereinafter in further details with reference to the drawings and embodiments. It should be understood that the specific embodiments described herein are merely for explaining the invention, but not intended for limitation.

In one embodiment of the invention, receiving an input training tensor data set; introducing a within class scatter matrix into an objective function such that between class distance is maximized, at the same time, within class distance is minimized by the objective function; constructing an optimal frame of the objective function of an optimal projection tensor machine OPSTM subproblem; constructing an optimal frame of an objective function of an OPSTM problem; according to lagrangian multiplier method, obtaining a dual problem of the optimal frame of the objective function, introducing a tensor rank one decomposition into calculation of tensor transvection, and obtaining a revised dual problem; utilizing sequential minimal optimization SMO algorithm to solve the revised dual problem and output a lagrangian optimal combination and an offset scalar; calculating a projection tensor; performing the rank one decomposition to the projection tensor; performing a back projection to a component obtained after performing the rank one decomposition to the projection tensor; performing rank one decomposition inverse operation to the component obtained after performing the back projection to obtain an optimal projection tensor W which is corresponded

to the training tensor data set; decision function construction phase: by the optimal projection tensor W which has been rank-one decomposed together with the offset scalar, constructing a decision function; application prediction phase: inputting to-be-predicted tensor data which has been rank-one decomposed into the decision function for prediction.

The implementation of the present invention will be described in detail with reference to specific embodiments:

The First Embodiment:

Fig.1 shows an implementation process of the optimization method for supervised learning under tensor mode provided by the first embodiment of the present invention. The details are as follows:

Step 101, receiving an input training tensor data set.

In the embodiment of the invention, let the training tensor data set be $\{X_m, y_m \mid m=1, 2 \dots M\}$, wherein X_m represents tensor input data, $y_m \in \{+1, -1\}$ represents a label.

Take a gray level image for example, sample points are stored in a form of a second-order tensor (matrix), and all the sample points which are in a form of a column vector comprise an input data set. In a similar way, a label set is also a column vector; furthermore, the location of each label is corresponded to the location of the corresponding sample point.

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_M \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}$$

Step 102, introducing a within class scatter matrix into an objective function such that between class distance is maximized, at the same time, within class distance is minimized by the objective function.

In the embodiment of the invention, the optimal frame of the objective

function of the support tensor machine (STM) problem is a combination of N vector modes of quadratic programming problems, which respectively correspond to a subproblem, wherein, a quadratic programming problem of the n-th subproblem is:

$$\min_{\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}} \quad \frac{1}{2} \|\mathbf{w}^{(n)}\|_F^2 \prod_{1 \leq i \leq N}^{i \neq n} \|\mathbf{w}^{(i)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)} \quad (1-1)$$

$$\text{S.t} \quad y_m \left((\mathbf{w}^{(n)})^T \left(\mathbf{X}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}^{(i)} \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)} \quad (1-2)$$

$$\xi_m^{(n)} \geq 0 \quad m = 1, 2, \dots, M \quad (1-3)$$

wherein, $\mathbf{w}^{(n)}$: the n-th order optimal projection vector of the training tensor data set, $n=1, 2, \dots, N$;

$b^{(n)}$: the n-th order offset scalar of the training tensor data set, $n=1, 2, \dots, N$;

C: a penalty factor;

$\xi_m^{(n)}$: a slack variable.

After introducing the within class scatter matrix into an objective function of an STM subproblem, through an eta coefficient η , the objection function of the quadratic programming problem of the n-th subproblem is changed into:

$$\min_{\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}} \quad \frac{1}{2} \left[\left(\|\mathbf{w}^{(n)}\|_F^2 + \eta ((\mathbf{w}^{(n)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(n)}) \right) \prod_{1 \leq i \leq N}^{i \neq n} \left(\|\mathbf{w}^{(i)}\|_F^2 + \eta ((\mathbf{w}^{(i)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(i)}) \right) \right] + C \sum_{m=1}^M \xi_m^{(n)} \quad (1-4)$$

wherein, $\mathbf{S}_w^{(n)}$ is the n-th order within class scatter matrix estimated after the training tensor data set is expanded along the n-th order; $\mathbf{w}^{(n)}$ herein possesses Fisher criterion effect, “maximum between class distance, minimum within class distance” at the n-th order of the training tensor data set; eta coefficient η is configured to measure the importance of the within class scatter.

Step 103, constructing an optimal frame of the objective function of an optimal projection tensor machine OPSTM subproblem.

In the embodiment of the invention, the optimal frame of an objective function of an optimal projection tensor machine OPSTM problem is a

combination of N vector modes of quadratic programming problems, which respectively correspond to a subproblem; wherein, a quadratic programming problem of the n-th subproblem is:

$$\min_{\mathbf{w}_*^{(n)}, b^{(n)}, \xi^{(n)}} \quad \frac{1}{2} \|\mathbf{w}_*^{(n)}\|_F^2 \prod_{1 \leq i \leq N}^{i \neq n} \|\mathbf{w}_*^{(i)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)} \quad (2-1)$$

$$\text{S.t} \quad y_m \left((\mathbf{w}_*^{(n)})^T \left(\mathbf{V}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}_*^{(i)} \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)} \quad (2-2)$$

$$\xi_m^{(n)} \geq 0 \quad m = 1, 2, \dots, M \quad (2-3)$$

wherein, $\mathbf{w}_*^{(n)}$: the n-th order projection vector of the training tensor data set; $\mathbf{w}_*^{(n)} = \mathbf{\Lambda}^{(n)1/2} \mathbf{P}^{(n)T} \mathbf{w}^{(n)}$ $n = 1, 2, \dots, N$; $\mathbf{w}^{(n)}$: the n-th order optimal projection vector of the training tensor data set of formula (1-4); $\mathbf{\Lambda}^{(n)}$ and $\mathbf{P}^{(n)}$ meet the equation, $\mathbf{P}^{(n)T} (\mathbf{E} + \eta \mathbf{S}_w^{(n)}) \mathbf{P}^{(n)} = \mathbf{\Lambda}^{(n)}$; E is an identity matrix; $\mathbf{V}_m = \mathbf{X}_m \prod_{i=1}^N \times_i \left[(\mathbf{P}^{(i)} \mathbf{\Lambda}^{(i)1/2})^{-1} \right]^T$ is tensor input data obtained after tensor input data $\mathbf{X}_m$ is projected along each order; $\mathbf{X}_i$ is an i-mode multiplication operator.

Step 104, transforming N vector modes of quadratic programming subproblems into a multiple quadratic programming problem under a single tensor mode, and constructing an optimal frame of an objective function of an OPSTM problem.

In the embodiment of the invention,

$$\begin{aligned} \|\mathbf{W}_*\|_F^2 &= \left\| \mathbf{w}_*^{(1)} \circ \mathbf{w}_*^{(2)} \circ \dots \circ \mathbf{w}_*^{(N)} \right\|_F^2 \\ &= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \dots \sum_{i_N=1}^{I_N} w_{*i_1, i_2, \dots, i_N}^2 \end{aligned}$$

$$= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} \left(w_{*i_1}^{(1)} \cdot w_{*i_2}^{(2)} \cdots w_{*i_N}^{(N)} \right)^2 \quad \text{Eq. 1}$$

$$\begin{aligned} &= \langle \mathbf{w}_*^{(1)}, \mathbf{w}_*^{(1)} \rangle \langle \mathbf{w}_*^{(2)}, \mathbf{w}_*^{(2)} \rangle \cdots \langle \mathbf{w}_*^{(N)}, \mathbf{w}_*^{(N)} \rangle \\ &= \prod_{n=1}^N \left\| \mathbf{w}_*^{(n)} \right\|_F^2 \\ &= (\mathbf{w}_*^{(n)})^T \left(\mathbf{V}_m \prod_{1 \leq i < N}^{i \neq n} \times_i \mathbf{w}_*^{(i)} \right) \\ &= \mathbf{V}_m \times_1 \mathbf{w}_*^{(1)} \times_2 \mathbf{w}_*^{(2)} \times \cdots \times_{(n-1)} \mathbf{w}_*^{(n-1)} \times_n \mathbf{w}_*^{(n)} \times_{(n+1)} \mathbf{w}_*^{(n+1)} \cdots \times_N \mathbf{w}_*^{(N)} \\ &= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} v_{m,i_1,i_2,\dots,i_N} w_{*i_1}^{(1)} w_{*i_2}^{(2)} \cdots w_{*i_N}^{(N)} \\ &= \sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} v_{m,i_1,i_2,\dots,i_N} w_{*i_1,i_2,\dots,i_N} \\ &= \langle \mathbf{W}_*, \mathbf{V}_m \rangle. \end{aligned} \quad \text{Eq. 2}$$

Wherein, $\left\| \right\|_F^2$ represents a norm and “ $\circ$ ” represents an outer product operator. According to the formulas, Eq. 1 and Eq. 2, $\left\| \mathbf{W}_* \right\|_F^2 = \prod_{n=1}^N \left\| \mathbf{w}_*^{(n)} \right\|_F^2$, and $(\mathbf{w}_*^{(n)})^T (\mathbf{V}_m \prod_{1 \leq i < N}^{i \neq n} \times_i \mathbf{w}_*^{(i)}) = \langle \mathbf{W}_*, \mathbf{V}_m \rangle$. Therefore, vector mode of quadratic programming problems of N subproblems can be transformed into a multiple quadratic programming problem under a single tensor mode, which means the optimal frame of an objective function of an OPSTM problem is:

$$\min_{\mathbf{W}_*, b, \xi} \quad \frac{1}{2} \left\| \mathbf{W}_* \right\|_F^2 + C \sum_{m=1}^M \xi_m \quad (3-1)$$

$$\text{S.t} \quad y_m (\langle \mathbf{W}_*, \mathbf{V}_m \rangle + b) \geq 1 - \xi_m \quad (3-2)$$

$$\xi_m \geq 0 \quad m = 1, 2, \dots, M \quad (3-3)$$

wherein, $\mathbf{W}^*$ is a projection tensor; $\langle \rangle$ is a transvection operator, and

$$\xi_m = \max_{n=1,2,\dots,N} \{\xi_m^{(n)}\}.$$

Through the Eq. 1 and Eq. 2, N vector modes of quadratic programming problems are transformed into a multiple quadratic programming problem under a single tensor mode. The transformed optimal frame of the objective function is the optimal frame of the objective function of the OPSTM problem. This can reduce the number of model parameters significantly, overcome issues such as curse of dimensionality, over learning and small sample occurred when vector mode algorithms process the tensor data.

Step 105, according to lagrangian multiplier method, obtaining a dual problem of the optimal frame of the objective function, introducing a tensor rank one decomposition into calculation of tensor transvection, and obtaining a revised dual problem.

In the embodiment, according to lagrangian multiplier method, a dual problem of the optimal frame [(3-1), (3-2), (3-3)] of the objection function of the OPSTM problem is obtained, wherein α_m is a lagrangian multiplier.

$$\max_{\alpha} \quad \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j \langle \mathbf{V}_i, \mathbf{V}_j \rangle \quad (4-1)$$

$$\text{S.t} \quad \sum_{m=1}^M (\alpha_m y_m) = 0 \quad (4-2)$$

$$0 < \alpha_m < C \quad m = 1, 2, \dots, M \quad (4-3)$$

A tensor CP (CANDECOMP/PARAFAC) is decomposed and introduced to the calculation of tensor transvection.

Rank one decompositions of tensor data $\mathbf{V}_i, \mathbf{V}_j$ are respectively:

$$\mathbf{V}_i = \sum_{r=1}^R \mathbf{v}_{ir}^{(1)} \circ \mathbf{v}_{ir}^{(2)} \circ \dots \circ \mathbf{v}_{ir}^{(N)} \quad \mathbf{V}_j = \sum_{r=1}^R \mathbf{v}_{jr}^{(1)} \circ \mathbf{v}_{jr}^{(2)} \circ \dots \circ \mathbf{v}_{jr}^{(N)}$$

$$\langle \mathbf{V}_i, \mathbf{V}_j \rangle = \sum_{p=1}^R \sum_{q=1}^R \prod_{n=1}^N \langle v_{ip}^{(n)}, v_{jq}^{(n)} \rangle$$

and

Therefore, the dual problem can be changed into:

$$\max_{\alpha} \quad \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \sum_{p=1}^R \sum_{q=1}^R \alpha_i \alpha_j y_i y_j \prod_{n=1}^N \langle v_{ip}^{(n)}, v_{jq}^{(n)} \rangle \quad (4-4)$$

$$\text{S.t} \quad \sum_{m=1}^M (\alpha_m y_m) = 0 \quad (4-2)$$

$$0 < \alpha_m < C \quad m = 1, 2, \dots, M \quad (4-3)$$

In the objection function (4-1) of the dual problem, a tensor rank one

decomposition auxiliary calculation, $\langle \mathbf{V}_i, \mathbf{V}_j \rangle = \sum_{p=1}^R \sum_{q=1}^R \prod_{n=1}^N \langle v_{ip}^{(n)}, v_{jq}^{(n)} \rangle$, is

introduced into the tensor transvection calculation part, which further reduces calculation complexity and storage cost, at the same time, the tensor rank one decomposition can obtain a more compact and more meaningful representation of a tensor objection, extract structural information and internal correlation of the

tensor data more effectively, and avoid a time-consuming alternative projection iterative process of the tensor mode algorithms of the prior art.

Step 106, utilizing sequential minimal optimization SMO algorithm to solve the revised dual problem and output an alagrangian optimal combination $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_M]$ and an offset scalar b.

Step 107, calculating a projection tensor $\mathbf{W}_*$.

In the embodiment of the invention, the projection tensor $\mathbf{W}_*$ is calculated

according to the formula
$$\mathbf{W}_* = \sum_{m=1}^M \alpha_m y_m \mathbf{V}_m$$
.

Step 108, performing the rank one decomposition to the projection tensor $\mathbf{W}_*$.

In the embodiment of the invention, the rank one decomposition is performed to the projection tensor $\mathbf{W}_*$ and the formula $\mathbf{W}_* = \mathbf{w}_*^{(1)} \circ \mathbf{w}_*^{(2)} \circ \dots \circ \mathbf{w}_*^{(N)}$ is obtained.

Step 109, performing a back projection to a component obtained after performing the rank one decomposition to the projection tensor $\mathbf{W}_*$.

In the embodiment of the invention, the back projection to a component is performed after performing the rank one decomposition to the projection tensor $\mathbf{W}_*$ to obtain the formula $\mathbf{w}^{(n)} = (\Lambda^{(n)1/2} \mathbf{P}^{(n)})^{-1} \mathbf{w}_*^{(n)}$, wherein $\mathbf{w}^{(n)}$ corresponds to the optimal projection vector of (1-4), and is the n-th order optimal projection vector of the training tensor data set, $n = 1, 2, \dots, N$.

Step 110, performing rank one decomposition inverse operation to the component obtained after performing the back projection to obtain an optimal projection tensor $\mathbf{W}$ which is corresponded to the training tensor data set.

In the embodiment of the invention, the components obtained after performing back projection are blended (rank one decomposition inverse operation)

into the optimal projection tensor W , $W = \mathbf{w}^{(1)} \circ \mathbf{w}^{(2)} \circ \dots \circ \mathbf{w}^{(N)}$. therefore, the optimal projection tensor W can embody Fisher criterion at each order.

Step 111, decision function construction phase: by the optimal projection tensor W which has been rank-one decomposed together with the offset scalar b , constructing a decision function.

In the embodiment of the invention, at the decision function construction phase, the rank one decomposition should be performed to the optimal projection tensor W , the decomposed optimal projection tensor W and the offset scalar b are used for constructing the decision function:

$$y(X) = \text{sign} \left[\prod_{n=1}^N \sum_{p=1}^R \sum_{q=1}^R \langle w_p^{(n)}, x_q^{(n)} \rangle + b \right].$$

Step 112, at the application prediction phase, inputting to-be-predicted tensor data which has been rank-one decomposed into the decision function for prediction.

In the embodiment of the invention, at the application prediction phase, the to-be-predicted tensor data which has been performed rank-one decomposed is input into the decision function for prediction.

Compared with the prior art, the embodiment possesses the following advantages: 1) N vector modes of quadratic programming problems are transformed into a multiple quadratic programming problem under a single tensor mode. The optimal frame of the transformed objective function is the optimal frame of the objective function of the OPSTM problem. This can reduce the number of model parameters significantly, overcome issues such as curse of dimensionality, over learning and small sample occurred when traditional vector mode algorithms process the tensor data, which ensures high-activity processing, at the same time,

highlights excellent classifying effects. Above all, the algorithms provided by the embodiments of the invention can process tensor data effectively and directly in tensor field, at the same time, possesses a feature of optimal classifying ability as well as strong practicability and popularization. 2) the within scatter matrix is introduced into the object function, which can receive and process the tensor data directly in the tensor field, and the output optimal projection tensor W can embody the Fisher criterion effect, “maximum between class distance, minimum within class distance” at each order. 3) In the objection function (4-1) of the dual problem, a tensor rank one decomposition auxiliary calculation,

$$\langle \mathbf{V}_i, \mathbf{V}_j \rangle = \sum_{p=1}^R \sum_{q=1}^R \prod_{n=1}^N \langle v_{ip}^{(n)}, v_{jq}^{(n)} \rangle$$

, is introduced into the tensor transvection calculation part, which further reduces calculation complexity and storage cost, at the same time, the tensor rank one decomposition can obtain a more compact and more meaningful representation of the tensor objection, and extract structural information and internal correlation of the tensor data more effectively, and avoid a time-consuming alternative projection iterative process.

It should be understood that the serial number of each step in this embodiment does not signify the execution sequence; the execution sequence of each step should be determined according to its function and internal logic, and should not form any limitation to the implementation process of the embodiment of the invention.

It should be understood by those skilled in the art that the whole or partial steps of the methods in each embodiment can be achieved by relevant hardware

instructed by program, and the corresponding program can be stored in a computer-readable storage medium; the storage medium can be for example ROM/RAM, magnetic disk or light disk, etc.

The Second Embodiment

Fig.2 shows a specific structural block diagram of an optimization system for supervised learning under tensor mode provided by the second embodiment of the invention. For illustration purposes, merely the part relevant to the embodiment of the invention is shown. The optimization system 2 for supervised learning under tensor mode includes: a data receiving unit 21, a within class scatter introducing unit 22, a subproblem optimal frame constructing unit 23, a problem optimal frame constructing unit 24, a dual problem obtaining unit 25, a dual problem solving unit 26, a projection tensor calculating unit 27, a projection tensor decomposition unit 28, a back projection unit 29, an optimal projection tensor calculating unit 210, a decision function constructing unit 211 and a predicting unit 212.

Wherein the data receiving unit 21 is configured to receive an input training tensor data set;

the within class scatter introducing unit 22 is configured to introduce a within class scatter matrix into an objective function such that between class distance is maximized, at the same time, within class distance is minimized by the objective function;

the subproblem optimal frame constructing unit 23 is configured to construct an optimal frame of an objective function of an optimal projection tensor machine OPSTM subproblem;

the problem optimal frame constructing unit 24 is configured to transform N

vector modes of quadratic programming subproblems into a multiple quadratic programming problem under a single tensor mode, and construct an optimal frame of an objective function of an OPSTM problem;

the dual problem obtaining unit 25 is configured to obtain a dual problem of the optimal frame of the objective function, introduce tensor rank one decomposition into calculation of tensor transvection, and obtain a revised dual problem according to lagrangian multiplier method;

the dual problem solving unit 26 is configured to utilize sequential minimal optimization SMO algorithm to solve the revised dual problem and output a lagrangian optimal combination and an offset scalar b ;

the projection tensor calculating unit 27 is configured to calculate a projection tensor W^* ;

the projection tensor decomposition unit 28 is configured to perform the rank one decomposition to the projection tensor W^* ;

the back projection unit 29 is configured to perform a back projection to a component obtained after performing the rank one decomposition to the projection tensor W^* ;

the optimal projection tensor calculating unit 210 is configured to perform rank one decomposition inverse operation to the component obtained after performing the back projection to obtain an optimal projection tensor W which is corresponded to the training tensor data set;

the decision function constructing unit 211 is configured to construct a

decision function construction phase and construct a decision function by the optimal projection tensor $\mathbf{W}$ which has been rank-one decomposed together with the offset scalar b ;

the predicting unit 212 is configured to input to-be-predicted tensor data which has been rank-one decomposed into the decision function for prediction in the application prediction phase.

Furthermore, through an eta coefficient η , after the within class scatter introducing unit 22 introduces the within class scatter matrix into an objective function of an STM subproblem, the objection function of the quadratic programming problem of the n-th subproblem is changed into:

$$\min_{\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}} \frac{1}{2} \left[\left(\|\mathbf{w}^{(n)}\|_F^2 + \eta ((\mathbf{w}^{(n)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(n)}) \right) \prod_{1 \leq i \leq N, i \neq n} \left(\|\mathbf{w}^{(i)}\|_F^2 + \eta ((\mathbf{w}^{(i)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(i)}) \right) \right] + C \sum_{m=1}^M \xi_m^{(n)}$$

wherein, $\mathbf{S}_w^{(n)}$ is an n-order within class scatter matrix estimated after the training tensor data set is expanded along the n-th order; $\mathbf{w}^{(n)}$ is the n-th order optimal projection vector of the training tensor data, $n=1, 2, \dots, N$; C is a penalty factor; $\xi_m^{(n)}$ is a slack variable; eta coefficient η is configured to measure the importance of the within class scatter matrix.

Furthermore, in the subproblem optimal frame constructing unit 23, the optimal frame of the objective function of the OPSTM problem is a combination of N vector modes of quadratic programming problems, which respectively corresponds to a subproblem; wherein a quadratic programming problem of the n-th subproblem is:

$$\min_{\mathbf{w}_*^{(n)}, b^{(n)}, \xi^{(n)}} \frac{1}{2} \|\mathbf{w}_*^{(n)}\|_F^2 \prod_{1 \leq i \leq N, i \neq n} \|\mathbf{w}_*^{(i)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)}$$

$$\begin{aligned} & y_m \left((\mathbf{w}_*^{(n)})^T \left(\mathbf{V}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}_*^{(i)} \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)} \\ \text{S.t} \quad & \xi_m^{(n)} \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\mathbf{w}_*^{(n)}$ is the n-th order projection vector of the training tensor data set;

$$\begin{aligned} \mathbf{w}_*^{(n)} &= \Lambda^{(n)1/2} \mathbf{P}^{(n)T} \mathbf{w}^{(n)}, \quad \text{wherein } \Lambda^{(n)} \text{ and } \mathbf{P}^{(n)} \text{ meet the equation,} \\ \mathbf{P}^{(n)T} (\mathbf{E} + \eta \mathbf{S}_w^{(n)}) \mathbf{P}^{(n)} &= \Lambda^{(n)}; \quad \mathbf{V}_m = \mathbf{X}_m \prod_{i=1}^N \times_i \left[(\mathbf{P}^{(i)} \Lambda^{(i)1/2})^{-1} \right]^T \end{aligned}$$

is

tensor input data obtained after tensor input data $\mathbf{X}_m$ in the training tensor data set is projected along each order; $\mathbf{X}_i$ is an i-mode multiplication operator; $b^{(n)}$ is the n-th order offset scalar of the training tensor data set.

Furthermore, according to a formula $\|\mathbf{w}_*\|_F^2 = \prod_{n=1}^N \|\mathbf{w}_*^{(n)}\|_F^2$, and a formula $(\mathbf{w}_*^{(n)})^T (\mathbf{V}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}_*^{(i)}) = \langle \mathbf{w}_*, \mathbf{V}_m \rangle$, the problem optimal frame constructing unit 24 transforms the N vector modes of quadratic programming subproblems into the multiple quadratic programming subproblem under a single tensor mode. A constructed optimal frame of the objective function of the OPSTM problem meets that:

$$\begin{aligned} \min_{\mathbf{w}_*, b, \xi} \quad & \frac{1}{2} \|\mathbf{w}_*\|_F^2 + C \sum_{m=1}^M \xi_m \\ \text{S.t} \quad & y_m (\langle \mathbf{w}_*, \mathbf{V}_m \rangle + b) \geq 1 - \xi_m \\ & \xi_m \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\langle \rangle$ is a transvection operator, and $\xi_m = \max_{n=1, 2, \dots, N} \{\xi_m^{(n)}\}$.

Furthermore, according to lagrangian multiplier method, the dual problem solving unit 26 obtains the dual problem of the optimal frame of the objective function, which is:

$$\begin{aligned}
& \max_{\alpha} \quad \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j < \mathbf{V}_i, \mathbf{V}_j > \\
& \text{S.t} \quad \sum_{m=1}^M (\alpha_m y_m) = 0 \\
& \quad 0 < \alpha_m < C \quad m = 1, 2, \dots, M,
\end{aligned}$$

the dual problem solving unit 26 introduces the tensor rank one decomposition into the calculation of the tensor transvection. An obtained revised dual problem is:

$$\begin{aligned}
& \max_{\alpha} \quad \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \sum_{p=1}^R \sum_{q=1}^R \alpha_i \alpha_j y_i y_j \prod_{n=1}^N < \mathbf{v}_{ip}^{(n)}, \mathbf{v}_{jq}^{(n)} > \\
& \text{S.t} \quad \sum_{m=1}^M (\alpha_m y_m) = 0 \\
& \quad 0 < \alpha_m < C \quad m = 1, 2, \dots, M.
\end{aligned}$$

Furthermore, the projection tensor calculating unit 27 calculates the projection tensor $\mathbf{W}_*$ according to a formula, $\mathbf{W}_* = \sum_{m=1}^M \alpha_m y_m \mathbf{V}_m$.

The optimization system for supervised learning under tensor mode provided by the embodiment of the invention can be applied to the corresponding method of the first embodiment. Please refer to the description of the first embodiment, which would not be further described herein.

Those skilled in the art should understand that the exemplary units and algorithm steps described in accompany with the embodiments disclosed in the specification can be achieved by electronic hardware, or the combination of computer software with electronic hardware. Whether these functions are executed

in a hardware manner or a software manner depends on the specific applications and design constraint conditions of the technical solutions. With respect to each specific application, a professional technician can achieve the described functions utilizing different methods, and these achievements should not be deemed as going beyond the scope of the invention.

Those skilled in the art can be clearly understood that for convenience and briefness, the specific working process of the described system, apparatus and unit can refer to the corresponding process of the above method embodiment, which would be further described herein.

It should be understood that the systems, devices and methods disclosed in several embodiments provided by the present application can be achieved in alternative ways. For example, the described device embodiments are merely schematically. For example, the division of the units is merely a division based on logic function, whereas the units can be divided in other ways in actual realization; for example, a plurality of units or components can be grouped or integrated into another system, or some features can be omitted or not executed. Furthermore, the shown or discussed mutual coupling or direct coupling or communication connection can be achieved by indirect coupling or communication connection of some interfaces, devices or units in electric, mechanical or other ways.

The units described as isolated elements can be or not be separated physically; an element shown as a unit can be or not be physical unit, which means that the element can be located in one location or distributed at multiple network units. Some or all of the units can be selected according to actual needs to achieve the purpose of the schemes of the embodiments.

Furthermore, each functional unit in each embodiment of the present invention can be integrated into a processing unit, or each unit can exist in isolation, or two or more than two units can be integrated into one unit.

If the integrated unit is achieved in software functional unit and sold or used as an independent product, the integrated unit can be stored in a computer-readable storage medium. Based on this consideration, the substantial part, or the part that is contributed to the prior art of the technical solution of the present invention, or part or all of the technical solutions can be embodied in a software product. The computer software product is stored in a storage medium, and includes several instructions configured to enable a computer device (can be a personal computer, device, network device, and so on) to execute all or some of the steps of the method of each embodiment of the present invention. The storage medium includes a U disk, a mobile hard disk, a read-only memory (ROM, Read-Only Memory), a random access memory (RAM, Random Access Memory), a disk or a light disk, and other various mediums which can store program codes.

The above contents are merely specific embodiments of the present invention, however, the protection scope of the present invention should not be limited by this. Any person skilled in the art can easily envisage alternations and displacements within the technical scope disclosed by the invention, which should also be within the protection scope of the present invention. Therefore, the protection scope of the present invention should be subjected to the protection scope of the claims.

CLAIMS

What is claimed is:

1. An optimization method for supervised learning under tensor mode, wherein the method comprise:

receiving an input training tensor data set;

introducing a within class scatter matrix into an objective function such that between class distance is maximized, at the same time, within class distance is minimized by the objective function;

constructing an optimal frame of the objective function of an optimal projection tensor machine OPSTM subproblem;

transforming N vector modes of quadratic programming subproblems into a multiple quadratic programming problem under a single tensor mode, and constructing an optimal frame of an objective function of an OPSTM problem;

according to lagrangian multiplier method, obtaining a dual problem of the optimal frame of the objective function, introducing a tensor rank one decomposition into calculation of tensor transvection, and obtaining a revised dual problem;

utilizing sequential minimal optimization SMO algorithm to solve the revised dual problem and output an alagrangian optimal combination and an offset scalar b;

calculating a projection tensor W_* ;

performing the rank one decomposition to the projection tensor W_* ;

performing a back projection to a component obtained after performing the rank one decomposition to the projection tensor W_* ;

performing rank one decomposition inverse operation to the component obtained after performing the back projection to obtain an optimal projection tensor W which is corresponded to the training tensor data set;

decision function construction phase: by the optimal projection tensor W

which has been rank-one decomposed together with the offset scalar b , constructing a decision function;

application prediction phase: inputting to-be-predicted tensor data which has been rank-one decomposed into the decision function for prediction.

2. The method of claim 1, wherein after introducing the within class scatter matrix into an objective function of an STM subproblem, through an eta coefficient η , the objection function of the quadratic programming problem of an n -th subproblem is changed into:

$$\min_{\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}} \quad \frac{1}{2} \left[\left(\|\mathbf{w}^{(n)}\|_F^2 + \eta ((\mathbf{w}^{(n)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(n)}) \right) \prod_{1 \leq i \leq N}^{i \neq n} \left(\|\mathbf{w}^{(i)}\|_F^2 + \eta ((\mathbf{w}^{(i)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(i)}) \right) \right] + C \sum_{m=1}^M \xi_m^{(n)}$$

wherein, $\mathbf{S}_w^{(n)}$ is an n -th order within class scatter matrix estimated after the training tensor data set is expanded along the n -th order; $\mathbf{w}^{(n)}$ is an n -th order optimal projection vector of the training tensor data, $n=1, 2, \dots, N$; C is a penalty factor; $\xi_m^{(n)}$ is a slack variable; eta coefficient η is configured to measure the importance of the within class scatter matrix.

3. The method of claim 2, wherein the optimal frame of an objective function of an OPSTM problem is a combination of N vector modes of quadratic programming problems, which respectively corresponds to a subproblem; wherein, a quadratic programming problem of an n -th subproblem is:

$$\begin{aligned} \min_{\mathbf{w}_*^{(n)}, b^{(n)}, \xi^{(n)}} \quad & \frac{1}{2} \|\mathbf{w}_*^{(n)}\|_F^2 \prod_{1 \leq i \leq N}^{i \neq n} \|\mathbf{w}_*^{(i)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)} \\ \text{S.t} \quad & y_m \left((\mathbf{w}_*^{(n)})^T \left(\mathbf{V}_m \prod_{1 \leq i \leq N}^{i \neq n} \times_i \mathbf{w}_*^{(i)} \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)} \\ & \xi_m^{(n)} \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\mathbf{w}_*^{(n)}$ is the n -th order projection vector of the training tensor data set;

$\mathbf{w}_*^{(n)} = \Lambda^{(n)1/2} \mathbf{P}^{(n)T} \mathbf{w}^{(n)}$, wherein $\Lambda^{(n)}$ and $\mathbf{P}^{(n)}$ meet that $\mathbf{P}^{(n)T} (\mathbf{E} + \eta \mathbf{S}_w^{(n)}) \mathbf{P}^{(n)} = \Lambda^{(n)}$; $\mathbf{E}$ is an identity matrix; $\mathbf{V}_m = \mathbf{X}_m \prod_{i=1}^N \times_i \left[(\mathbf{P}^{(i)} \Lambda^{(i)1/2})^{-1} \right]^T$ is tensor input data obtained after tensor input data $\mathbf{X}_m$ in the training tensor data set is projected along each order; $\mathbf{X}_i$ is an i-mode multiplication operator; $b^{(n)}$ is the n-th order offset scalar of the training tensor data set.

4. The method of claim 3, wherein according to a formula $\|\mathbf{W}_*\|_F^2 = \prod_{n=1}^N \|\mathbf{w}_*^{(n)}\|_F^2$, and a formula $(\mathbf{w}_*^{(n)})^T (\mathbf{V}_m \prod_{1 \leq i < N}^{i \neq n} \times_i \mathbf{w}_*^{(i)}) = \langle \mathbf{W}_*, \mathbf{V}_m \rangle$, transforming the N vector modes of quadratic programming subproblems into the multiple quadratic programming subproblem under a single tensor mode; a constructed optimal frame of the objective function of the OPSTM problem meets that:

$$\begin{aligned} \min_{\mathbf{W}_*, b, \xi} \quad & \frac{1}{2} \|\mathbf{W}_*\|_F^2 + C \sum_{m=1}^M \xi_m \\ \text{S.t} \quad & y_m (\langle \mathbf{W}_*, \mathbf{V}_m \rangle + b) \geq 1 - \xi_m \\ & \xi_m \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\langle \rangle$ is a transvection operator, and $\xi_m = \max_{n=1, 2, \dots, N} \{\xi_m^{(n)}\}$.

5. The method of claim 4, wherein according to lagrangian multiplier method, an obtained dual problem of the optimal frame of the objective function is:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j \langle \mathbf{V}_i, \mathbf{V}_j \rangle \\ \text{S.t} \quad & \sum_{m=1}^M (\alpha_m y_m) = 0 \\ & 0 < \alpha_m < C \quad m = 1, 2, \dots, M, \end{aligned}$$

introducing the tensor rank one decomposition into the calculation of the tensor transvection; an obtained revised dual problem is:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \sum_{p=1}^R \sum_{q=1}^R \alpha_i \alpha_j y_i y_j \prod_{n=1}^N \langle \mathbf{v}_{ip}^{(n)}, \mathbf{v}_{jq}^{(n)} \rangle \\ \text{S.t} \quad & \sum_{m=1}^M (\alpha_m y_m) = 0 \\ & 0 < \alpha_m < C \quad m = 1, 2, \dots, M \end{aligned}$$

6. The method of claim 5, wherein calculating the projection tensor $\mathbf{W}_*$ according to a formula, $\mathbf{W}_* = \sum_{m=1}^M \alpha_m y_m \mathbf{V}_m$.

7. An optimization system for supervised learning under tensor mode is provided, wherein the system comprises:

a data receiving unit configured to receive an input training tensor data set;

a within class scatter introducing unit configured to introduce a within class scatter matrix into an objective function such that between class distance is maximized, at the same time, within class distance is minimized by the objective function;

a subproblem optimal frame constructing unit configured to construct an optimal frame of an objective function of an optimal projection tensor machine OPSTM subproblem;

a problem optimal frame constructing unit configured to transform N vector modes of quadratic programming subproblems into a multiple quadratic programming problem under a single tensor mode, and construct an optimal frame of an objective function of an OPSTM problem;

a dual problem obtaining unit configured to obtain a dual problem of the optimal frame of the objective function, introduce a tensor rank one decomposition into calculation of tensor transvection, and obtain a revised dual problem according to lagrangian multiplier method;

a dual problem solving unit configured to utilize sequential minimal optimization SMO algorithm to solve the revised dual problem and output an alagrangian optimal combination and an offset scalar b;

a projection tensor calculating unit configured to calculate a projection tensor W_* ;

a projection tensor decomposition unit configured to perform the rank one decomposition to the projection tensor W_* ;

a back projection unit configured to perform a back projection to a component obtained after performing the rank one decomposition to the projection tensor W_* ;

an optimal projection tensor calculating unit configured to perform rank one decomposition inverse operation to the component obtained after performing the back projection to obtain an optimal projection tensor W which is corresponded to the training tensor data set;

a decision function constructing unit configured to construct a decision function construction phase and construct a decision function by the optimal projection tensor W which has been rank-one decomposed together with the offset scalar b,

a predicting unit configured to input to-be-predicted tensor data which has been rank-one decomposed into the decision function for prediction in the application prediction phase.

8. The system of claim 7, wherein after the within class scatter introducing unit introduces the within class scatter matrix into an objective function of an STM subproblem, through an eta coefficient η , the objection function of the quadratic programming problem of the n-th subproblem is changed into:

$$\min_{\mathbf{w}^{(n)}, b^{(n)}, \xi^{(n)}} \frac{1}{2} \left[\left(\left\| \mathbf{w}^{(n)} \right\|_F^2 + \eta ((\mathbf{w}^{(n)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(n)}) \right) \prod_{1 \leq i < N}^{i \neq n} \left(\left\| \mathbf{w}^{(i)} \right\|_F^2 + \eta ((\mathbf{w}^{(i)})^T \mathbf{S}_w^{(n)} \mathbf{w}^{(i)}) \right) \right] + C \sum_{m=1}^M \xi_m^{(n)}$$

wherein, $\mathbf{S}_w^{(n)}$ is an n-order within class scatter matrix estimated after the training tensor data set is expanded along the n-th order; $\mathbf{w}^{(n)}$ is the n-th order

optimal projection vector of the training tensor data, $n=1, 2, \dots, N$; C is a penalty factor; $\xi_m^{(n)}$ is a slack variable; eta coefficient η is configured to measure the importance of the within class scatter matrix.

9. The system of claim 8, wherein in the subproblem optimal frame constructing unit, the optimal frame of the objective function of the OPSTM problem is a combination of N vector modes of quadratic programming problems, which respectively corresponds to a subproblem; wherein, a quadratic programming problem of the n -th subproblem is:

$$\begin{aligned} \min_{\mathbf{w}_*^{(n)}, b^{(n)}, \xi_m^{(n)}} \quad & \frac{1}{2} \|\mathbf{w}_*^{(n)}\|_F^2 \prod_{1 \leq i \leq N, i \neq n} \|\mathbf{w}_*^{(i)}\|_F^2 + C \sum_{m=1}^M \xi_m^{(n)} \\ \text{S.t} \quad & y_m \left((\mathbf{w}_*^{(n)})^T \left(\mathbf{V}_m \prod_{1 \leq i \leq N, i \neq n} \times_i \mathbf{w}_*^{(i)} \right) + b^{(n)} \right) \geq 1 - \xi_m^{(n)} \\ & \xi_m^{(n)} \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\mathbf{w}_*^{(n)}$ is the n -th order projection vector of the training tensor data set; $\mathbf{w}_*^{(n)} = \Lambda^{(n)1/2} \mathbf{P}^{(n)T} \mathbf{w}^{(n)}$, wherein $\Lambda^{(n)}$ and $\mathbf{P}^{(n)}$ meet that $\mathbf{P}^{(n)T} (\mathbf{E} + \eta \mathbf{S}_w^{(n)}) \mathbf{P}^{(n)} = \Lambda^{(n)}$; $\mathbf{E}$ is an identity matrix; $\mathbf{V}_m = \mathbf{X}_m \prod_{i=1}^N \times_i \left[(\mathbf{P}^{(i)} \Lambda^{(i)1/2})^{-1} \right]^T$ is tensor input data obtained after tensor input data $\mathbf{X}_m$ in the training tensor data set is projected along each order; $\times_i$ is an i -mode multiplication operator; $b^{(n)}$ is the n -th order offset scalar of the training tensor data set.

10. The system of claim 9, wherein according to a formula $\|\mathbf{w}_*\|_F^2 = \prod_{n=1}^N \|\mathbf{w}_*^{(n)}\|_F^2$, and a formula $(\mathbf{w}_*^{(n)})^T (\mathbf{V}_m \prod_{1 \leq i \leq N, i \neq n} \times_i \mathbf{w}_*^{(i)}) = \langle \mathbf{w}_*, \mathbf{V}_m \rangle$, the problem optimal frame constructing unit transforms the N vector modes of quadratic programming subproblems into the multiple quadratic programming subproblem under a single

tensor mode; a constructed optimal frame of the objective function of the OPSTM problem meets that:

$$\begin{aligned} \min_{\mathbf{W}_*, b, \xi} \quad & \frac{1}{2} \|\mathbf{W}_*\|_F^2 + C \sum_{m=1}^M \xi_m \\ \text{S.t} \quad & y_m (\langle \mathbf{W}_*, \mathbf{V}_m \rangle + b) \geq 1 - \xi_m \\ & \xi_m \geq 0 \quad m = 1, 2, \dots, M \end{aligned}$$

wherein, $\langle \rangle$ is a transvection operator, and $\xi_m = \max_{n=1, 2, \dots, N} \{\xi_m^{(n)}\}$.

11. The system of claim 10, wherein according to lagrangian multiplier method, the dual problem solving unit obtains the dual problem of the optimal frame of the objective function, which is:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j \langle \mathbf{V}_i, \mathbf{V}_j \rangle \\ \text{S.t} \quad & \sum_{m=1}^M (\alpha_m y_m) = 0 \\ & 0 < \alpha_m < C \quad m = 1, 2, \dots, M ; \end{aligned}$$

the dual problem solving unit introduces the tensor rank one decomposition into the calculation of the tensor transvection; an obtained revised dual problem is:

$$\begin{aligned} \max_{\alpha} \quad & \sum_{m=1}^M \alpha_m - \frac{1}{2} \sum_{i,j=1}^M \sum_{p=1}^R \sum_{q=1}^R \alpha_i \alpha_j y_i y_j \prod_{n=1}^N \langle \mathbf{v}_{ip}^{(n)}, \mathbf{v}_{jq}^{(n)} \rangle \\ \text{S.t} \quad & \sum_{m=1}^M (\alpha_m y_m) = 0 \\ & 0 < \alpha_m < C \quad m = 1, 2, \dots, M . \end{aligned}$$

12. The system of claim 11, wherein the projection tensor calculating unit calculates the projection tensor $\mathbf{W}_*$ according to a formula, $\mathbf{W}_* = \sum_{m=1}^M \alpha_m y_m \mathbf{V}_m$.