# Sparse Least Squares Twin Support Vector Machines with Manifold-preserving Graph Reduction

i iong ie

*The School of Information Science and Engineering,*
*Ningbo University, Zhejiang 315211, China*

e words: on-parallel perplane Classi er Least Squares Twin Support Vector Machines Manifold-preserving Graph Reduction

A stract: Least squares twin support vector achines are a new non-parallel h perplane classi er in which the pri al opti i ation pro le s of twin support vector achines are odi ed in least square sense and inequalit constraints are replaced equalit constraints In classi cation pro le s enhancing the ro ustness of least squares twin support vector achines and reducing the ti e co ple it of ernel function evaluation of a new e a ple when inferring the la el of a new e a ple are ver i portant In this paper we propose a new sparse least squares twin support vector achines ased on anifold-preserving graph reduction which is an ef cient graph reduction algorith with anifold assu ption This ethod rst selects infor ative e a ples for positive e a ples and negative e a ples respectivel and then applies the for classi cation peri ental results con r the feasi ilit and effectiveness of our proposed ethod

## 1 INTRODUCTION

Support vector achines SVMs are a ver ef - cient classi cation algorith Shawe-Ta lor and Sun 2011 Vapni 1 5 Christianini and Shawe-Ta lor 2002 Riple 2002 which are ased on the princi-pled idea of structural ris ini i ation in statistical learning theor Co pared with other achine learn-ing algorith s SVMs can o tain a etter generali a-tion The are well- nown for their ro ustness good generali ation a ilit and unique glo al opti u so-lution in the case of conve pro le Recent ears witnessed e ergence of an successful non-parallel h perplane classi ers Twin support vector achines TSVM a adeva et al 2007 are a representative non-parallel h perplane classi er which ai s to gen-erate two non-parallel h perplanes such that one of the h perplanes is closer to one class and as far as possi le fro the other class Twin ounded SVM T SVM Shao et al 2011 is an i proved version of TSVM whose opti i ation pro le s are changed slightl adding a regulari ation ter with the idea of a i i ing the argin TSVM has een e tended to these learning fra ewor s such as ulti-tas learn-ing ie and Sun 2015 ulti-view learning ie and Sun 2015a ie and Sun 2014 se i-supervised learning Chen et al 2016 ulti-la el learning i et al 2012 and regression pro le Peng 2010 The two non-parallel h perplanes of TSVM are o - tained solving a pair of quadratic progra ing pro le s PPs Thus the ti e co ple it is rel-ative high Least squares twin support vector a-chines LSTSVM u ar and Gopal 200 were proposed to reduce the ti e co ple it chang-ing the constraints to a series of equalities constraints and leading to a pair of linear equations and can easil handle large datasets Man i proved vari-ants of LSTSVM have een proposed such as nowl-edge ased LSTSVM u ar et al 2010 Laplacian LSTSVM for se i-supervised classi cation Chen et al 2014 eighted LSTSVM Mu et al 2014 owever enhancing the ro ustness of LSTSVM and reducing the ti e co ple it of ernel function eval-uation of a new e a ple when inferring the la el of a new e a ple are ver i portant

ne of sparse ethods uses onl a su set of the data and focuses on the strategies of selecting the representative e a ples to for the su set These ethods lead to a signi cant reduction of the ti e co ple it Although so e ethods such as ran-do sa pling or - eans clustering can e used to reduce the si e of the graph the have no guar-antees of preserving the anifold structure or ef-fectivel re oving outliers and nois e a ples In



Xie, X.
Sparse Least Squares Twin Support Vector Machines with Manifold-preserving Graph Reduction.
DOI: 10.5220/0006690805630567
In *Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2018)*, pages 563-567
ISBN: 978-989-758-276-9

particular the - eans ethod is sensitive to outliers and ti e-consu ing when the nu er of clusters is large Manifold-preserving graph reduction Sun et al 2014 is a graph reduction algorith which can effectivel eli inate outliers and nois e - a ples In this paper a novel LSTSVM algorith ased on anifold-preserving graph reduction is proposed The e peri ental results on four datasets validated the feasi ilit and effectiveness of the proposed ethod

The re ainder of this paper proceeds as follows Section 2 reviews related wor a out LSTSVM and MPGR Section 3 thoroughl introduces our proposed SLSTSVM After reporting e peri ental results in Section 4 we give conclusions and future wor in Section 5

# 2 RELATED WORK

In this section we rie review LSTSVM and MPGR

## 2.1 LSTSVM

Given a training dataset containing $m$ e a ples elonging to classes 1 and $-1$ are represented atrices $A_+$ and $B_-$ and the si e of $A_+$ and $B_-$ are $(m_1 \times d)$ and $(m_2 \times d)$ respectivel e ne two atrices A and four vectors $v_1$ $v_2$ $e_1$ $e_2$ where $e_1$ and $e_2$ are vectors of ones of appropriate di ensions and

$$A = (A_+, e_1),\ B = (B_-, e_2),$$
$$v_1 = \begin{pmatrix} w_1 \\ b_1 \end{pmatrix},\ v_2 = \begin{pmatrix} w_2 \\ b_2 \end{pmatrix}. \qquad 1$$

The central idea of LSTSVM u ar and Gopal 200 is to see two nonparallel h perplanes

$$w_1^\top x + b_1 = 0 \text{ and } w_2^\top x + b_2 = 0 \qquad 2$$

around which the e a ples of the corresponding class get clustered The classi er is given solving the following PPs separatel

LSTSVM1

$$\underset{v_1,q_1}{\text{in}} \ \frac{1}{2}(Av_1)^\top(Av_1) + \frac{c_1}{2} q_1^\top q_1$$
$$\text{s t} \ -Bv_1 + q_1 = e_2, \qquad 3$$

LSTSVM2

$$\underset{v_2,q_2}{\text{in}} \ \frac{1}{2}(Bv_2)^\top(Bv_2) + \frac{d_1}{2} q_2^\top q_2$$
$$\text{s t} \ Av_2 + q_2 = e_1, \qquad 4$$

where $c_1$ and $d_1$ are nonnegative para eters and $q_1$ $q_2$ are slac vectors of appropriate di ensions ach of the a ove two PPs can e converted to the e plicit e pression of LSTSVM

LSTSVM1

$$\underset{v_1,q_1}{\text{in}} \ \frac{1}{2}(Av_1)^\top(Av_1) + \frac{1}{2} c_1 (e_2 + Bv_1 + q_1)^\top$$
$$(e_2 + Bv_1 + q_1), \qquad 5$$

LSTSVM2

$$\underset{v_2,q_2}{\text{in}} \ \frac{1}{2}(Bv_2)^\top(Bv_2) + \frac{1}{2} d_1 (e_1 - (Av_2 + q_2))^\top$$
$$(e_1 - (Av_2 + q_2)). \qquad 6$$

The two nonparallel h perplanes are o tained solving the following two s ste s of linear equations

$$v_1 = -(A^\top A + \frac{1}{c_1} B^\top B)^{-1} A^\top e_2,$$
$$v_2 = (B^\top B + \frac{1}{d_1} A^\top A)^{-1} B^\top e_1. \qquad 7$$

The la el of a new e a ple $x$ is deter ined the ini u of $|x^\top w_r + b_r|$ $r = 1,2$ which are the perpendicular distances of $x$ to the two h perplanes given in 2

## 2.2 MPGR

In this section we rie introduce the anifold-preserving graph reduction algorith Sun et al 2014

MPGR is an ef cient graph reduction algorith ased on the anifold assu ption A sparse graph with anifold-preserving properties eans that a point outside of it should have a high connectivit with a point to e reserved Suppose there is a graph $G$ co posed of all unla eled e a ples the anifold-preserving sparse graphs are those sparse graph candidates which have a high space connectivit with $G$ The value of space connectivit is as follows

$$\frac{1}{m-s} \sum_{i=s+1}^{m} \left( \underset{j=1,\dots,s}{\text{a}} W_{ij} \right), \qquad 8$$

where $m$ is the nu er of all vertices $s$ is the nu er of vertices to e retained and $W$ is the weight atri or su set selection of all the unla eled e a ples a point which is closer to surrounding points should e selected since it contains ore i portant infor ation This confor s to MPGR in which the e a ples with a large degree will e preferred The degree $d(p)$ is de ned as

$$d(p) = \sum_{p-q} w_{pq},$$

where $p - q$ eans that e a ple $p$ is connected with e a ple $q$ and $w_{pq}$ is their corresponding weight If two e a ples are not lin ed their weight would e ero ue to its si plicit $d(p)$ is generall considered as a criterion to construct sparse graphs A igger $d(p)$ eans the e a ple $p$ contains ore infor ation a el the e a ple $p$ is ore li el to e selected into the sparse graphs In a word the su set constructed MPGR is high representative and aintains a good glo al anifold structure of the original data distri ution This can eli inate the outlier and noise e a ples and enhance the ro ustness of the algorith

**Algorithm 1 :** Manifold-preserving Graph Reduction Algorith

1. **Input:** Graph $G(V,E,W)$ with vertices
2. $s$ is the nu er of the vertices in the desired sparse graph
3. for $z = 1,2,\ldots,s$
4. co pute the degree $d(i)(i = 1,2,\ldots,m-z+1)$
5. select the vertice $v$ with the a i u degree
6. re ove $v$ and associated edges fro $G$ add $v$ to $G_s$
7. end for
8. **Output** Manifold-preserving sparse graph $G_s$ with $s$ vertices

## 3 SLSTSVM

As entioned earlier LSTSVM generates two non-parallel h perplanes such that each h perplane is close to one class and as far as possi le fro the other Ta e the positive h perplane for e a ple so e outliers or nois e a ples in the positive e a ples a have negative effect on the o taining of the opti al positive h perplane owever MPGR can effectivel re ove outliers or nois e a ples The train e a ple reduction algorith also can speed up the LSTSVM train and testing process

The MPGR constructs a graph using the corresponding positive e a ples Initiall the candidate set contains all positive e a ples while the sought sparse set is null or each e a ple in the candidate set the MPGR calculates the degree of the corresponding verte in the graph It selects a verte with the a i u degree in the graph corresponding to the positive e a ples Then we include the data point associated with the chosen verte into the sought sparse set and re ove it fro the candidate set This step considers the representativeness criterion ue to the propert of high spatial connectivit the su set is highl representative and preserving the glo al structure of the original distri ution of training set The sparse set selection of negative e a ples is si ilar to a ove processes verall inspired the anifold-preserving principle SLSTSVM not onl can enhance the ro ustness of algorith ut also reduce the train and testing ti e

**Algorithm 2 :** Sparse Least Squares Twin Support Vector Machines

1. **Input:** Positive e a ples $A$ and negative e a ples $B$ odel para eters $c_1$ $d_1$
2. se MPGR on positive e a ples and negative e a ples to o tain the sparse su sets $T_1$ and $T_2$ corresponding to the positive e a ples and negative e a ples according to the retained percentage $r$ respectivel
3. ed the two sparse su sets $T_1$ and $T_2$ into the opti i ation of LSTSVM
4. eter ine para eters of two h perplanes solving the linear equation 7
5. **Output:** or a test e a ple $x = (x^\top\ 1)^\top$ if $|x^\top v_1| \leq |x^\top v_2|$ it is classi ed to class $+1$ otherwise class $-1$

The co putation ti e of LSTSVM with the ernel ethod is a out $O(m^3/4)$ which is the ti e of a tri inversion operation while the co putation ti e of SLSTSVM is reduced to $r^3$ ti es of the co putation ti e of LSTSVM

## 4 EXPERIMENTAL RESULTS

In this section we evaluate our proposed SLSTSVM on four real-world datasets The four datasets co e fro CI Machine Learning Repositor ionosphere classi cation handwritten digit classi cation pi a and sonar Speci c infor ation a out ionosphere and handwritten digits is listed in Ta le 1

Ta le 1 atasets

| a e | Attri utes | Instances | Classes |
|---|---|---|---|
| Ionosphere | 34 | 351 | 2 |
| andwritten digits | 64 | 2000 | 10 |

### 4.1 Ionosphere

The ionosphere dataset was collected a s ste in Goose a La rador that contains a phased arra of 16 high-frequenc antennas with a total trans itted power on the order of 6 4 ilowatts The targets were free electrons in the ionosphere Good radar returns are those showing evidence of so e t pe of structure in the ionosphere ad returns are those that do not and their signals pass through the ionosphere It includes 351 e a ples in total which are divided into 225 Good positive e a ples and 126 ad negative e a ples

In our e peri ents we capture of the data variance while reducing the di ensionalit fro 34 to 21 with PCA e use ten-fold cross-validation to select the est para eters for all involved ethods in the region $[2^{-10}, 2^{10}]$ with e ponential growth 1 and get the average classi cation accurac rates running the algorith s for ve ti es e use 300 e a ples for training and the others for testing e set the output nu er of MPGR as 10 20 30 40 100 of the 300 e a ples Linear ernel is chosen for the dataset LSTSVM with rando sa pling is used for co parison ro the e peri ental results in Ta le 2 we can nd that our ethod SLSTSVM perfor s etter than LSTSVM hen the percentage is 10 the perfor ance of SLSTSVM is alread as sa e as the one with the percentage 100 hen the percentage is 30 the perfor ance is est e conclude SLSTSVM can i prove its ro ustness compared with LSTSVM

Ta le 2 Classi cation accuracies and standard deviations on Ionosphere

| Method \ Per | LSTSVM | SLSTSVM |
|---|---|---|
| 10 | 76 08 10 78 | 82 35 5 72 |
| 20 | 80 78 5 61 | 83 53 4 72 |
| 30 | 80 3 6 04 | 85 4 4 51 |
| 40 | 78 43 10 28 | 81 57 5 65 |
| 100 | 82 35 7 6 | 82 35 7 6 |

Ta le 3 Classi cation accuracies and standard deviations on andwritten digits

| igit pair \ Method | LSTSVM | SLSTSVM |
|---|---|---|
| 0 8 | 5 20 3 0 | 6 0 1 2 |
| 3 | 7 0 1 47 | 8 10 1 08 |
| 3 5 | 6 30 1 2 | 6 0 1 3 |
| 2 8 | 6 60 1 47 | 6 70 1 82 |

## 4.2 Handwritten Digits

This dataset contains features of handwritten digits ($0 \sim$ ) e tracted fro a collection of utch utilit aps It contains 2000 e a ples 200 e a ples per class with ve views e use the view 64 arhunen-Love coef cients of each e a ple i age ecause TSVMs are designed for inar classi cation while handwritten digits dataset contains 10 classes e choose four pairs 3 5 2 8 0 8 and 3 to evaluate all involved ethods for the e peri ent Linear ernel is chosen for the dataset e use 200 e a ples for training and 200 e a ples for testing e use ten-fold cross-validation to select the est para eters for all involved ethods in the region $[2^{-10}, 2^{10}]$ with e ponential growth 1 e set the input nu er of MPGR as 10 0 100 of the 200 e a ples ro the e peri ental results in Ta le 5 we can conclude that the perfor ance of SLSTSVM is superior to the one of LSTSVM

## 4.3 Pima and Sonar

Pi a is dataset that can predict dia etes of Pi a Indians according to the incidence of edical records over 5 ears It consists of 768 e a ples and 8 attri utes Sonar is a dataset that can predict whether the o ect is a roc or a ine according to the strength of a given sonar fro different angles It contains 208 e a ples and 60 attri utes ro the e peri ental results we can conclude that SLSTSVM are superior to LSTSVM hen the percentage is 10 the perfor ance of SLSTSVM outperfor s the one with the percentage 100 e conclude SLSTSVM can i prove its ro ustness

Ta le 4 Classi cation accuracies and standard deviations on Pi a

| Per \ Method | LSTSVM | SLSTSVM |
|---|---|---|
| 10 | 54 55 5 8 | 5 85 7 18 |
| 0 | 55 73 5 27 | 57 31 6 86 |
| 100 | 55 67 5 22 | 55 67 5 22 |

Ta le 5 Classi cation accuracies and standard deviations on Sonar

| Per \ Method | LSTSVM | SLSTSVM |
|---|---|---|
| 20 | 57 78 4 75 | 60 56 10 26 |
| 0 | 62 6 6 14 | 63 8 4 72 |
| 100 | 62 22 3 0 | 62 22 3 0 |

# 5 CONCLUSION AND FUTURE WORK

In this paper we have proposed a novel sparse least squares support vector achines ased on anifold-preserving graph reduction peri ental results on ultiple real-world datasets indicate that SLSTSVM are superior to LSTSVM using rando sa pling It would e interesting for future wor to e ploit the wa which selects the infor ative and representative e a ples fro unla eled e a ples to ulti-view se i-supervised learning

## ACKNOWLEDGEMENTS

This wor is supported ing o niversit talent pro ect 421703670 as well as progra s sponsored C ong Magna und in ing o niversit It is also supported the he iang Provincial epart ent of ducation under Pro ects 801700472

## REFERENCES

Chen Shao and eng 2014 Laplacian least squares twin support vector achine for se i-supervised classi cation *Neurocomputing* 145 465 476

Chen Shao Li C and eng 2016 MLTSVM A novel twin support vector achine to ulti-la el learning *Pattern Recognition* 52 61 74

Christianini and Shawe-Ta lor 2002 *An introduction to support vector machines* Ca ridge niversit Press Ca ridge

a adeva he chandani S and Chandra 2007 Twin support vector achines for pattern classi cation *IEEE Transactions on Pattern Analysis and Machine Intelligence* 74 05 10

u ar M and Gopal M 200 Least squares twin support vector achines for pattern classi cation *Expert Systems with Applications* 36 7535 7543

u ar M he chandani R and Gopal M 2010 nowledge ased least squares twin support vector achines *Information Sciences* 180 4606 4618

Mu Li and Chen L 2014 Classi cation with noise via weighted least squares twin support vector achine *Computer Simulation* 31 288 2 2

Peng 2010 Tsvr An ef cient twin support vector achine for regression *Neural Networks* 23 365 372

i Tian and Shi 2012 Laplacian twin support vector achine for se i-supervised classi cation *Neural Networks* 35 46 53

Riple 2002 *Pattern recognition and neural networks* Ca ridge niversit Press Ca ridge

Shao hang C ang and eng 2011 I prove ents on twin support vector achines *IEEE Transactions on Neural Networks and learning systems* 1 62 68

Shawe-Ta lor and Sun S 2011 A review of opti i ation ethodologies in support vector achines *Neurocomputing*

Sun S ussain and Shawe-Ta lor 2014 Manifold-preserving graph reduction for sparse se i-supervised learning *Neurocomputing* 124 13 21

Vapni V 1 5 *The nature of statistical learning theory* Springer-Verlag ew or

ie and Sun S 2014 Multi-view laplacian twin support vector achines *Applied Intelligence* 41 105 1068

ie and Sun S 2015a Multi-view twin support vector achines *Intelligent Data Analysis* 1 701 712

ie and Sun S 2015 Multitas centroid twin support vector achines *Neurocomputing* 14 1085 10 1