Semi-supervised Learning from General Unlabeled Data

Kaizhu Huang[1], Zenglin Xu[2], Irwin King[2], Michael R. Lyu[2]

[1]Department of Engineering Mathematics
University of Bristol, Bristol, UK
K.Huang@bris.ac.uk
[2]Department of Computer Science & Engineering
The Chinese University of Hong Kong
{zlxu, king, lyu}@cse.cuhk.edu.hk

December 17, 2008
ICDM 2008, Pisa, Italy

Problems

Definition

Input: Labeled D_L & Unlabeled D_u

$D_L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, known $\mathbf{y} \in \pm 1$

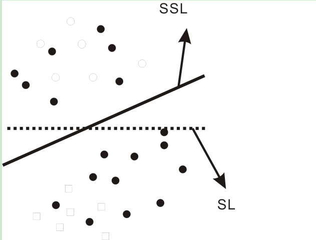
$D_U = \{(\mathbf{x}_j, z_j)\}_{j=l+1}^{l+u}$, unknown $\mathbf{z} \in \pm 1$

Output: $f : \mathcal{X} \rightarrow \pm 1$

SL: using D_L

SSL: using D_L and D_U

Illustration



Problems

Definition

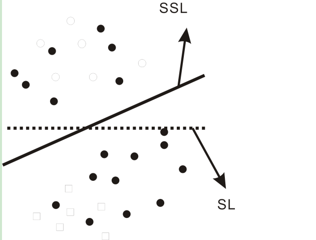
Input: Labeled D_L & Unlabeled D_u

$D_L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, known $\mathbf{y} \in \pm 1$
$D_U = \{(\mathbf{x}_j, z_j)\}_{j=l+1}^{l+u}$, unknown $\mathbf{z} \in \pm 1$

Output: $f : \mathcal{X} \rightarrow \pm 1$

SL: using D_L
SSL: using D_L and D_U

Illustration



- Remarks
    - Performance: SSL is very useful especially in the case of limited number of labeled samples
    - Assumption: Unlabeled data samples share the same set of labels as the labeled data
- Problem: Such assumption may be violated in many cases.

Problems

Definition

Input: Labeled D_L & Unlabeled D_u

$D_L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, known $\mathbf{y} \in \pm 1$

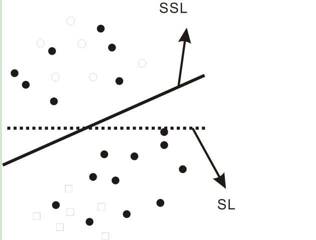
$D_U = \{(\mathbf{x}_j, z_j)\}_{j=l+1}^{l+u}$, unknown $\mathbf{z} \in \pm 1$

Output: $f : \mathcal{X} \rightarrow \pm 1$

SL: using D_L

SSL: using D_L and D_U

Illustration



- Remarks
  - Performance: SSL is very useful especially in the case of limited number of labeled samples
  - Assumption: Unlabeled data samples share the same set of labels as the labeled data
- Problem: Such assumption may be violated in many cases.

Problems

Definition

Input: Labeled D_L & Unlabeled D_u

$D_L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, known $\mathbf{y} \in \pm 1$

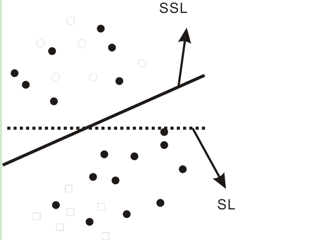
$D_U = \{(\mathbf{x}_j, z_j)\}_{j=l+1}^{l+u}$, unknown $\mathbf{z} \in \pm 1$

Output: $f : \mathcal{X} \rightarrow \pm 1$

SL: using D_L

SSL: using D_L and D_U

Illustration



- Remarks
  - Performance: SSL is very useful especially in the case of limited number of labeled samples
  - Assumption: Unlabeled data samples share the same set of labels as the labeled data
- Problem: Such assumption may be violated in many cases.

Problems

Definition

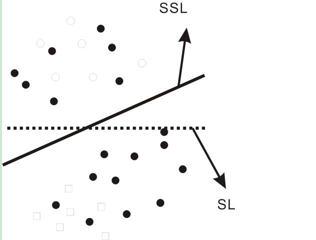
Input: Labeled D_L & Unlabeled D_u

$D_L = \{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, known $\mathbf{y} \in \pm 1$
$D_U = \{(\mathbf{x}_j, z_j)\}_{j=l+1}^{l+u}$, unknown $\mathbf{z} \in \pm 1$

Output: $f : \mathcal{X} \rightarrow \pm 1$

SL: using D_L
SSL: using D_L and D_U

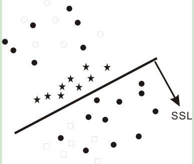
Illustration



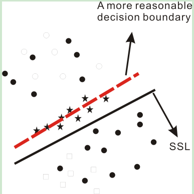
- Remarks
  - Performance: SSL is very useful especially in the case of limited number of labeled samples
  - Assumption: Unlabeled data samples share the same set of labels as the labeled data
- Problem: Such assumption may be violated in many cases.

Motivation I

General Unlabeled Data



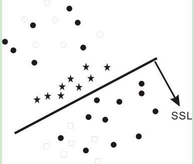
Proposed General SSL



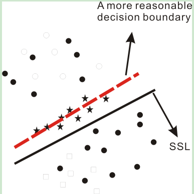
- Unlabeled data can be divided into either relevant or irrelevant data
  - relevant: either $+1$ or -1 class
  - irrelevant: neither $+1$ nor -1, denoted as the 0 class
- Margin maximization principle for decision f

Motivation I

General Unlabeled Data



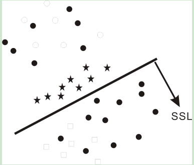
Proposed General SSL



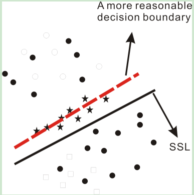
- Unlabeled data can be divided into either relevant or irrelevant data
    - relevant: either $+1$ or -1 class
    - irrelevant: neither $+1$ nor -1, denoted as the 0 class
- Margin maximization principle for decision f

Motivation I

General Unlabeled Data



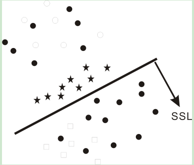
Proposed General SSL



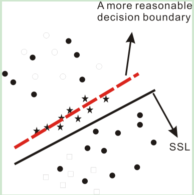
- Unlabeled data can be divided into either relevant or irrelevant data
  - relevant: either $+1$ or -1 class
  - irrelevant: neither $+1$ nor -1, denoted as the 0 class
- Margin maximization principle for decision f
  - Relevant data, i.e., $+1$ and -1 class should be pushed away from the boundary as far as possible
  - Irrelevant data, i.e., 0 class should be clustered around

Motivation I

General Unlabeled Data



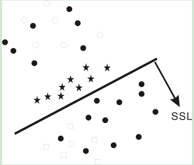
Proposed General SSL



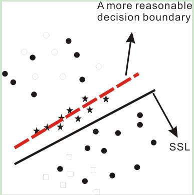
- Unlabeled data can be divided into either relevant or irrelevant data
  - relevant: either $+1$ or -1 class
  - irrelevant: neither $+1$ nor -1, denoted as the 0 class
- Margin maximization principle for decision f
  - Relevant data, i.e., $+1$ and -1 class should be pushed away from the boundary as far as possible
  - Irrelevant data i.e., 0 class should be clustered around f

Motivation I

General Unlabeled Data



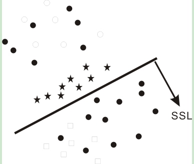
Proposed General SSL



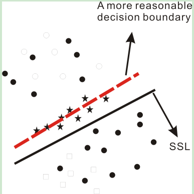
- Unlabeled data can be divided into either relevant or irrelevant data
  - relevant: either $+1$ or -1 class
  - irrelevant: neither $+1$ nor -1, denoted as the 0 class
- Margin maximization principle for decision f
  - Relevant data, i.e., $+1$ and -1 class should be pushed away from the boundary as far as possible
  - Irrelevant data i.e., 0 class should be clustered around f

Motivation I

General Unlabeled Data



Proposed General SSL

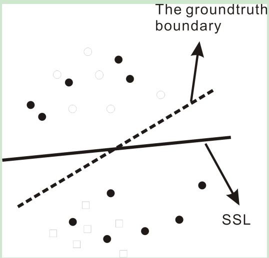


- Unlabeled data can be divided into either relevant or irrelevant data
  - relevant: either $+1$ or -1 class
  - irrelevant: neither $+1$ nor -1, denoted as the 0 class
- Margin maximization principle for decision f
  - Relevant data, i.e., $+1$ and -1 class should be pushed away from the boundary as far as possible
  - Irrelevant data i.e., 0 class should be clustered around f

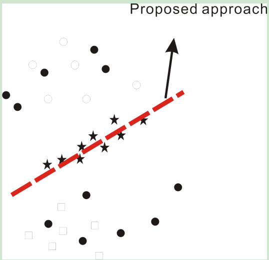
Motivation II

Irrelevant data are useful especially when both the numbers of labeled and unlabeled relevant data are limited but the unlabeled irrelevant data are sufficiently large or structured.

SSL



Proposed General SSL



Contributions

- A general SSL framework where unlabeled data do not necessarily share the same set of labels as the labeled data
- A decision boundary as well as the automatic label of unlabeled data could be learned simultaneously.
- A Semi-Definite Programming (SDP) method is proposed for solving the involved optimization problem.

Contributions

- A general SSL framework where unlabeled data do not necessarily share the same set of labels as the labeled data
- A decision boundary as well as the automatic label of unlabeled data could be learned simultaneously.
- A Semi-Definite Programming (SDP) method is proposed for solving the involved optimization problem.

Contributions

- A general SSL framework where unlabeled data do not necessarily share the same set of labels as the labeled data
- A decision boundary as well as the automatic label of unlabeled data could be learned simultaneously.
- A Semi-Definite Programming (SDP) method is proposed for solving the involved optimization problem.

Related Work

- Related Work
  - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
    using the third class of data (irrelevant) within the SL framework
  - SSL with Universum [D. Zhang et al. SDM 2008]
    using the third class of data (irrelevant) within the SSL framework
- Problem
  - Universum data (the third class) need to be indicated beforehand
  - In another word, the third class needs to be labeled beforehand

Related Work

- Related Work
  - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
    using the third class of data (irrelevant) within the SL framework
  - SSL with Universum [D. Zhang et al. SDM 2008]
    using the third class of data (irrelevant) within the SSL framework
- Problem
  - Universum data (the third class) need to be indicated beforehand
  - In another word, the third class needs to be labeled beforehand

Related Work

- Related Work
  - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
    using the third class of data (irrelevant) within the SL framework
  - SSL with Universum [D. Zhang et al. SDM 2008]
    using the third class of data (irrelevant) within the SSL framework
- Problem
  - Universum data (the third class) need to be indicated beforehand

Related Work

- Related Work
  - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
    using the third class of data (irrelevant) within the SL framework
  - SSL with Universum [D. Zhang et al. SDM 2008]
    using the third class of data (irrelevant) within the SSL framework
- Problem
  - Universum data (the third class) need to be indicated beforehand
  - In another word, the third class needs to be labeled beforehand
    —impractical in many cases

Related Work

- Related Work
    - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
      using the third class of data (irrelevant) within the SL framework
    - SSL with Universum [D. Zhang et al. SDM 2008]
      using the third class of data (irrelevant) within the SSL framework
- Problem
    - Universum data (the third class) need to be indicated beforehand
    - In another word, the third class needs to be labeled beforehand
      —impractical in many cases

Related Work

- Related Work
  - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
    using the third class of data (irrelevant) within the SL framework
  - SSL with Universum [D. Zhang et al. SDM 2008]
    using the third class of data (irrelevant) within the SSL framework
- Problem
  - Universum data (the third class) need to be indicated beforehand
  - In another word, the third class needs to be labeled beforehand
    —impractical in many cases

Related Work

- Related Work
  - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
    using the third class of data (irrelevant) within the SL framework
  - SSL with Universum [D. Zhang et al. SDM 2008]
    using the third class of data (irrelevant) within the SSL framework
- Problem
  - Universum data (the third class) need to be indicated beforehand
  - In another word, the third class needs to be labeled beforehand
    —impractical in many cases

Related Work

- Related Work
  - Universum Learning (USVM) [J. Weston et al. ICML 2007, Sinz et al NIPS 2008]
    using the third class of data (irrelevant) within the SL framework
  - SSL with Universum [D. Zhang et al. SDM 2008]
    using the third class of data (irrelevant) within the SSL framework
- Problem
  - Universum data (the third class) need to be indicated beforehand
  - In another word, the third class needs to be labeled beforehand —impractical in many cases

Model Definition (USSL)

$$
\begin{aligned}
\min_{\mathbf{w}, b, \xi, \eta, \mathbf{y}_{l+1:n}} \quad & \frac{1}{2}||\mathbf{w}||^2 + C_L \sum_{i=1}^{l} \xi_i + C_U \sum_{j=l+1}^{n} \min(\eta_j, \xi_j) \\
\text{s.t.} \quad & y_i(\mathbf{w}_i \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, \ldots, l, && (1) \\
& y_j(\mathbf{w}_j \cdot \mathbf{x}_j + b) \geq 1 - \xi_j, && (2) \\
& |\mathbf{w}_j \cdot \mathbf{x}_j + b| \leq \varepsilon + \eta_j, && (3) \\
& \eta_j \geq 0, j = l+1, \ldots, n, \xi_k \geq 0, k = 1, \ldots, n,
\end{aligned}
$$

(1) describes the loss for the labeled data.
(2) provides the loss if $\mathbf{x}_j$ is judged as the class of ± 1
(3) presents the loss if $\mathbf{x}_j$ is judged as the class of 0
The loss incurred by unlabeled $\mathbf{x}_j$ is given by the minimum loss that it is judged as the class of ± 1 or 0.

Theoretical Justification

Theorem 1

A slightly modified version of the USSL optimization is equivalent to training a standard Transductive SVM with the training points projected onto the orthogonal complement of span $\{\mathbf{z}_j - \mathbf{z}_0, \mathbf{z}_j \in \mathcal{U}\}$*, where* $\mathbf{z}_0$ *is an arbitrary element of the space spanned by the irrelevant samples denoted by* $\mathcal{U}$*.*

Remarks

- Irrelevant data do not contribute to the final accuracy directly
- It decides the subspace where the decision function is derived and consequently affect the performance.

Optimization issues

- Difficults:
  non-convex problem caused by two terms
  - $y_j \mathbf{w}_i$ — a classical problem encountered by SSL
  - $\min(\eta_j, \xi_j)$ —the new problem encountered in our General SSL
- Solution
  - Transformed to the dual space and relax yy^T as matrix $\mathbf{M}$ — similar to the traditional SSL
  - Transformed $\min(\eta_j, \xi_j)$ to Integer Programming problem, and further relaxed to Linear Programming problem

Optimization issues

- Difficults:
  non-convex problem caused by two terms
  - $y_j \mathbf{w}_i$ — a classical problem encountered by SSL
  - $\min(\eta_j, \xi_j)$ —the new problem encountered in our General SSL
- Solution
  - Transformed to the dual space and relax yy^T as matrix $\mathbf{M}$ — similar to the traditional SSL
  - Transformed $\min(\eta_j, \xi_j)$ to Integer Programming problem, and further relaxed to Linear Programming problem

Optimization issues

- Difficults:
  non-convex problem caused by two terms
  - $y_j \mathbf{w}_i$ — a classical problem encountered by SSL
  - $\min(\eta_j, \xi_j)$ —the new problem encountered in our General SSL
- Solution
  - Transformed to the dual space and relax yy^T as matrix M — similar to the traditional SSL
  - Transformed $\min(\eta_j, \xi_j)$ to Integer Programming problem and further relaxed to Linear Programming problem

Optimization issues

- Difficults:
  non-convex problem caused by two terms
  - $y_j \mathbf{w}_i$ — a classical problem encountered by SSL
  - $\min(\eta_j, \xi_j)$ —the new problem encountered in our General SSL
- Solution
  - Transformed to the dual space and relax $\mathbf{y}\mathbf{y}^T$ as matrix $\mathbf{M}$ — similar to the traditional SSL
  - Transformed $\min(\eta_j, \xi_j)$ to Integer Programming problem, and further relaxed to Linear Programming problem

Optimization issues

- Difficults:
  non-convex problem caused by two terms
  - $y_j \mathbf{w}_i$ — a classical problem encountered by SSL
  - $\min(\eta_j, \xi_j)$ —the new problem encountered in our General SSL
- Solution
  - Transformed to the dual space and relax $\mathbf{y}\mathbf{y}^T$ as matrix $\mathbf{M}$ — similar to the traditional SSL
  - Transformed $\min(\eta_j, \xi_j)$ to Integer Programming problem, and further relaxed to Linear Programming problem

Optimization issues

- Difficults:
  non-convex problem caused by two terms
  - $y_j\mathbf{w}_i$ — a classical problem encountered by SSL
  - $\min(\eta_j, \xi_j)$ —the new problem encountered in our General SSL
- Solution
  - Transformed to the dual space and relax $\mathbf{y}\mathbf{y}^T$ as matrix $\mathbf{M}$ — similar to the traditional SSL
  - Transformed $\min(\eta_j, \xi_j)$ to Integer Programming problem, and further relaxed to Linear Programming problem

Transformed to Integer Programming problem...

The optimization can be equivalently transformed to

$$\min_{\mathbf{w},b,\xi,\eta,\mathbf{y}_{l+1:n},\mathbf{d}} \frac{1}{2}||\mathbf{w}||^2 + C_L \sum_{i=1}^{l} \xi_i + C_U \sum_{j=l+1}^{n} (\eta_j + \xi_j),$$

s.t.

$$y_i(\mathbf{w}_i \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, \ldots, l \quad (4)$$

$$y_j(\mathbf{w}_j \cdot \mathbf{x}_j + b) + \xi_j + M(1 - d_j) \geq 1, \quad (5)$$

$$|\mathbf{w}_j \cdot \mathbf{x}_j + b| \leq \varepsilon + \eta_j + Md_j, \quad (6)$$

$$d_j = \{0, 1\} \quad j = l + 1, \ldots, n,$$

$$\eta_j \geq 0, j = l + 1, \ldots, n, \xi_k \geq 0, k = 1, \ldots, n.$$

where, $d_j = \begin{cases} 0 & \text{if} \quad y_j = \pm 1 \\ 1 & \text{if} \quad y_j = 0 \end{cases}$, and M is a large positive constant.

IP problem is still hard to solve.

Transformed to Integer Programming problem...

The optimization can be equivalently transformed to

$$\min_{\mathbf{w},b,\xi,\eta,\mathbf{y}_{l+1:n},\mathbf{d}} \frac{1}{2}||\mathbf{w}||^2 + C_L \sum_{i=1}^{l} \xi_i + C_U \sum_{j=l+1}^{n} (\eta_j + \xi_j),$$

s.t.

$$y_i(\mathbf{w}_i \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, i = 1, \ldots, l \quad (4)$$

$$y_j(\mathbf{w}_j \cdot \mathbf{x}_j + b) + \xi_j + M(1 - d_j) \geq 1, \quad (5)$$

$$|\mathbf{w}_j \cdot \mathbf{x}_j + b| \leq \varepsilon + \eta_j + Md_j, \quad (6)$$

$$d_j = \{0, 1\} \quad j = l + 1, \ldots, n,$$

$$\eta_j \geq 0, j = l + 1, \ldots, n, \xi_k \geq 0, k = 1, \ldots, n.$$

where, $d_j = \begin{cases} 0 & \text{if} \quad y_j = \pm 1 \\ 1 & \text{if} \quad y_j = 0 \end{cases}$, and M is a large positive constant.

IP problem is still hard to solve.

Relaxed as an SDP problem...

$$\min_{\mathbf{M},\mathbf{d},\nu,\delta,t} \quad t \quad \text{s.t.}$$

$$\begin{pmatrix} P & \mathbf{a}+\nu-B^T\delta \\ (\mathbf{a}+\nu-B^T\delta)^T & t-2\delta^T\mathbf{C} \end{pmatrix} \succeq 0,$$

$$0 \le d_j \le 1,$$

$$rank(\mathbf{M}) = 1, \mathbf{M}_{1:l,1:l} = \mathbf{y}_{1:l}\mathbf{y}_{1:l}^T.$$

where

$$P = \begin{pmatrix} \mathbf{K}\circ(\mathbf{y}\mathbf{y}^T) & \mathrm{Diag}(\mathbf{y})\mathbf{K}_{1:n,l:n} & -\mathrm{Diag}(\mathbf{y})\mathbf{K}_{1:n,l:n} \\ \mathbf{K}_{1:n,l:n}^T\mathrm{Diag}(\mathbf{y}) & \mathbf{K}_{l+1:n,l+1:n} & -\mathbf{K}_{l+1:n,l+1:n} \\ -\mathbf{K}_{1:n,l:n}^T\mathrm{Diag}(\mathbf{y}) & -\mathbf{K}_{l+1:n,l+1:n} & \mathbf{K}_{l+1:n,l+1:n} \end{pmatrix}$$

$$B = \begin{pmatrix} \mathbf{I}_{n\times n}, & \mathbf{0}_{n\times 2m} \\ \mathbf{0}_{m\times n}, & Q_{m\times 2m} \end{pmatrix}, \mathbf{a} = (\mathbf{1}_l; \mathbf{1}_m - M(\mathbf{1}-\mathbf{d}); -M\mathbf{d}; -M\mathbf{d})$$

- Similar to traditional SSL, by removing the rank-one constraint and relax $\mathbf{y}\mathbf{y}^T = \mathbf{M}$, the above problem is exactly an SDP problem.
- SDP problem can be solved by some packages such as Sedumi in polynomial time.

Relaxed as an SDP problem...

$$
\begin{aligned}
\min_{\mathbf{M},\mathbf{d},\nu,\delta,t} \quad & t \quad \text{s.t.} \\
& \begin{pmatrix} P & \mathbf{a}+\nu - B^T\delta \\ (\mathbf{a}+\nu - B^T\delta)^T & t - 2\delta^T\mathbf{C} \end{pmatrix} \succeq 0, \\
& 0 \le d_j \le 1, \\
& rank(\mathbf{M}) = 1, \mathbf{M}_{1:l,1:l} = \mathbf{y}_{1:l}\mathbf{y}_{1:l}^T.
\end{aligned}
$$

where

$$
P = \begin{pmatrix} \mathbf{K}\circ(\mathbf{y}\mathbf{y}^T) & \mathrm{Diag}(\mathbf{y})\mathbf{K}_{1:n,l:n} & -\mathrm{Diag}(\mathbf{y})\mathbf{K}_{1:n,l:n} \\ \mathbf{K}_{1:n,l:n}^T\mathrm{Diag}(\mathbf{y}) & \mathbf{K}_{l+1:n,l+1:n} & -\mathbf{K}_{l+1:n,l+1:n} \\ -\mathbf{K}_{1:n,l:n}^T\mathrm{Diag}(\mathbf{y}) & -\mathbf{K}_{l+1:n,l+1:n} & \mathbf{K}_{l+1:n,l+1:n} \end{pmatrix}
$$

$$
B = \begin{pmatrix} \mathbf{I}_{n\times n}, & \mathbf{0}_{n\times 2m} \\ \mathbf{0}_{m\times n}, & Q_{m\times 2m} \end{pmatrix}, \mathbf{a} = (\mathbf{1}_l; \mathbf{1}_m - M(\mathbf{1}-\mathbf{d}); -M\mathbf{d}; -M\mathbf{d})
$$

- Similar to traditional SSL, by removing the rank-one constraint and relax $\mathbf{y}\mathbf{y}^T = \mathbf{M}$, the above problem is exactly an SDP problem.
- SDP problem can be solved by some packages such as Sedumi in polynomial time.

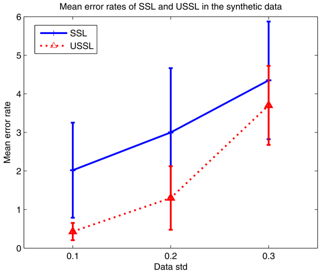
Experimental Setup

- Comparison Algorithms
  - Universum SVM: All the unlabeled data are treated as the irrelevant data
  - SSL: All the unlabeled data are treated as the relevant data
  - USSL (proposed approach): Automatically detect from the unlabeled data whether a sample is irrelevant or relevant
- Data Set
  - Toy Dataset
    Three two-dimensional Gaussian distributions, centered at $(-0.3, -0.3)$, $(0, 0)$, and $(0.3, 0.3)$ respectively, are treated as class -1, 0, and $+1$.

    5 labeled samples for each class; 10 unlabeled samples for each class ($+1$, -1, and 0)
  - MNIST and USPS (Follow [Weston et al. 07])
    5 and 8 are the relevant classes (class $+1$ and -1 respectively); the other digits as the irrelevant classes.

    20 labeled samples for 5 & 8 per class; 30 unlabeled samples for each class ($+1$, -1, and 0)

Toy Data: Accuracy

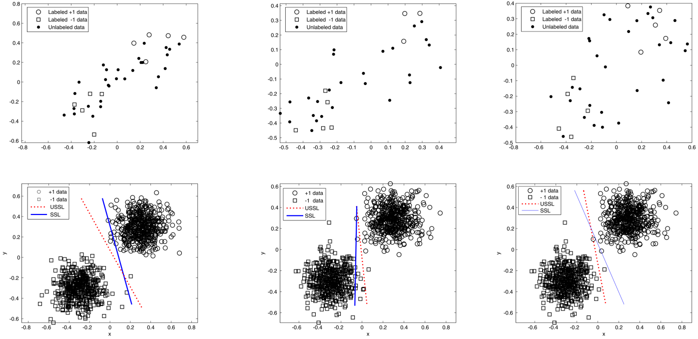


- USSL can indeed boost the performance of SSL in the data

Toy Data: Illustration I

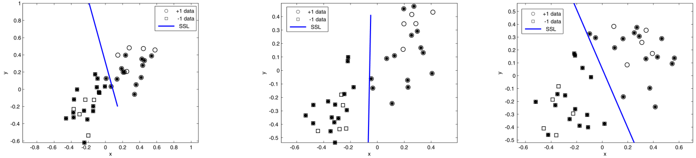
Tr. Data

Te. Data

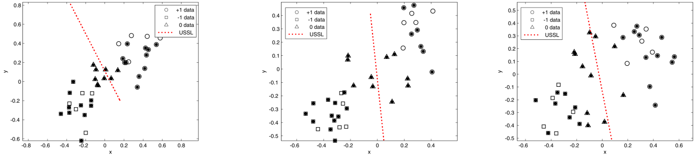


Toy Data: Illustration II

SSL:



USSL:



Experimental results on USPS data

Data set	USVM	SSL	USSL
0	67.05 ± 2.31	85.05 ± 1.94	$\mathbf{89.85 \pm 1.47}$
1	71.45 ± 1.59	83.61 ± 2.52	$\mathbf{89.23 \pm 1.89}$
2	69.50 ± 4.29	84.44 ± 2.08	$\mathbf{89.81 \pm 2.34}$
3	70.43 ± 1.68	84.75 ± 1.86	$\mathbf{89.65 \pm 2.24}$
4	65.80 ± 3.04	85.12 ± 3.91	$\mathbf{86.69 \pm 2.01}$
6	64.80 ± 2.36	78.45 ± 2.21	$\mathbf{83.70 \pm 1.90}$
7	66.93 ± 3.75	87.37 ± 2.51	$\mathbf{90.42 \pm 1.75}$
9	72.37 ± 3.42	82.86 ± 2.39	$\mathbf{85.13 \pm 2.31}$

1. USSL outperforms the other two algorithms consistently.
2. USVM treats all the data as irrelevant data and cannot benefit from unlabeled relevant data.
3. SSL treats all the data as relevant data and cannot refine the decision boundary.

Experimental results on USPS data

Data set	USVM	SSL	USSL
0	67.05 ± 2.31	85.05 ± 1.94	$\mathbf{89.85 \pm 1.47}$
1	71.45 ± 1.59	83.61 ± 2.52	$\mathbf{89.23 \pm 1.89}$
2	69.50 ± 4.29	84.44 ± 2.08	$\mathbf{89.81 \pm 2.34}$
3	70.43 ± 1.68	84.75 ± 1.86	$\mathbf{89.65 \pm 2.24}$
4	65.80 ± 3.04	85.12 ± 3.91	$\mathbf{86.69 \pm 2.01}$
6	64.80 ± 2.36	78.45 ± 2.21	$\mathbf{83.70 \pm 1.90}$
7	66.93 ± 3.75	87.37 ± 2.51	$\mathbf{90.42 \pm 1.75}$
9	72.37 ± 3.42	82.86 ± 2.39	$\mathbf{85.13 \pm 2.31}$

1. USSL outperforms the other two algorithms consistently.
2. USVM treats all the data as irrelevant data and cannot benefit from unlabeled relevant data.
3. SSL treats all the data as relevant data and cannot refine the decision boundary.

Experimental results on USPS data

Data set	USVM	SSL	USSL
0	67.05 ± 2.31	85.05 ± 1.94	$\mathbf{89.85 \pm 1.47}$
1	71.45 ± 1.59	83.61 ± 2.52	$\mathbf{89.23 \pm 1.89}$
2	69.50 ± 4.29	84.44 ± 2.08	$\mathbf{89.81 \pm 2.34}$
3	70.43 ± 1.68	84.75 ± 1.86	$\mathbf{89.65 \pm 2.24}$
4	65.80 ± 3.04	85.12 ± 3.91	$\mathbf{86.69 \pm 2.01}$
6	64.80 ± 2.36	78.45 ± 2.21	$\mathbf{83.70 \pm 1.90}$
7	66.93 ± 3.75	87.37 ± 2.51	$\mathbf{90.42 \pm 1.75}$
9	72.37 ± 3.42	82.86 ± 2.39	$\mathbf{85.13 \pm 2.31}$

1. USSL outperforms the other two algorithms consistently.
2. USVM treats all the data as irrelevant data and cannot benefit from unlabeled relevant data.
3. SSL treats all the data as relevant data and cannot refine the decision boundary.

Experimental results on MNIST data

Data Set	USVM	SSL	USSL
0	45.25 ± 2.19	53.25 ± 2.84	$\mathbf{58.25 \pm 2.11}$
1	52.77 ± 1.42	54.10 ± 2.78	$\mathbf{60.25 \pm 2.75}$
2	54.58 ± 2.67	56.92 ± 3.12	$\mathbf{57.67 \pm 2.97}$
3	55.14 ± 1.90	52.09 ± 2.30	$\mathbf{57.25 \pm 1.32}$
4	56.65 ± 1.22	57.12 ± 2.49	$\mathbf{59.25 \pm 2.10}$
6	52.75 ± 2.80	54.50 ± 2.12	$\mathbf{57.67 \pm 1.27}$
7	60.51 ± 2.12	58.09 ± 3.01	$\mathbf{68.50 \pm 2.26}$
9	59.25 ± 1.15	48.25 ± 2.64	$\mathbf{63.00 \pm 1.50}$

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?
  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class
- Q2: In what cases will the USSL be useful?
  Hints:
- Q3: Can the optimization be further speed up?
  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL.
- Q4: How do the relaxations influence the final performance?
  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?
  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class
- Q2: In what cases will the USSL be useful?
  Hints:
- Q3: Can the optimization be further speed up?
  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL
- Q4: How do the relaxations influence the final performance?
  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class

- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL

- Q3: Can the optimization be further speed up?

  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL

- Q4: How do the relaxations influence the final performance?

  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class

- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class
- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL
- Q3: Can the optimization be further speed up?

  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL
- Q4: How do the relaxations influence the final performance?

  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class

- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL

- Q3: Can the optimization be further speed up?

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class

- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL

- Q3: Can the optimization be further speed up?

  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL.

- Q4: How do the relaxations influence the final performance?

  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class

- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL

- Q3: Can the optimization be further speed up?

  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL.

- Q4: How do the relaxations influence the final performance?

  Answer: Unclear. Similar to the same issue in traditional SSL, question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class

- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL

- Q3: Can the optimization be further speed up?

  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL.

- Q4: How do the relaxations influence the final performance?

  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class
- Q2: In what cases will the USSL be useful?

  Hints:
  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL
- Q3: Can the optimization be further speed up?

  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL.
- Q4: How do the relaxations influence the final performance?

  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Discussion & Future Work

- Q1: Are Universum (class 0) data always helpful?

  Answer: NO. Universum data may hurt the performance especially when class 0 resembles one class over the other class

- Q2: In what cases will the USSL be useful?

  Hints:

  - samples of class 0 are neither like those of class $+1$ nor -1
  - the more concentrated the data of class 0, the more helpful the USSL

- Q3: Can the optimization be further speed up?

  Answer: YES. Actually, the optimization resembles the SSL optimization very much and recent progress on speeding SSL can also benefit USSL.

- Q4: How do the relaxations influence the final performance?

  Answer: Unclear. Similar to the same issue in traditional SSL, this question is still open to solve.

Conclusion

- We have proposed a general SSL framework where unlabeled data do not necessarily share the same label as the labeled data
- We can learn the decision boundary as well as the automatic label of unlabeled data simultaneously.
- We have proposed a Semi-Definite Programming (SDP) for solving the involved optimization problem.
- Experimental results show that the proposed USSL is useful in certain cases especially when the numbers of labeled & unlabeled relevant samples are both limited.