Semi-Supervised Discriminatively Regularized Classifier with Pairwise Constraints

Jijian Huang, Hui Xue*, and Yuqing Zhai

School of Computer Science and Engineering
Southeast University, Nanjing, 211189, P.R. China
{huangjijian2008,hxue,yqzhai}@seu.edu.cn

Abstract. In many real-world classifications such as video surveillance, web retrieval and image segmentation, we often encounter that class information is reflected by the pairwise constraints between data pairs rather than the usual labels for each data, which indicate whether the pairs belong to the same class or not. A common solution is combining the pairs into some new samples labeled by the constraints and then designing a smoothness-driven regularized classifier based on these samples. However, it still utilizes the limited discriminative information involved in the constraints insufficiently. In this paper, we propose a novel semi-supervised discriminatively regularized classifier (SSDRC). By introducing a new discriminative regularization term into the classifier instead of the usual smoothness-driven term, SSDRC can not only use the discriminative information more fully but also explore the local geometry of the new samples further to improve the classification performance. Experiments demonstrate the superiority of our SSDRC.

Keywords: Discriminative information, Structural information, Pairwise constraints, Semi-supervised classification.

1 Introduction

Semi-supervised learning is a class of machine learning techniques that makes use of both labeled and unlabeled data, which has achieved considerable development in theory and application [1-4]. According to different actual circumstances, semi-supervised learning usually involves two categories of class information, that is, the class label and the pairwise constraint. The class label specifies the concrete label for each datum, which is common in the traditional classification, while the pairwise constraint is defined on the data pair, which indicates that whether the pair belongs to the same class (must-link) or not (cannot-link). In many applications, the pairwise constraint is actually more general than the class label, because sometimes the true label may not be known prior, while it is easier for a user to specify whether the data pair belong to the same class or different class. Moreover, the pairwise constraints can be derived from the class label but not

* Corresponding author.

P. Anthony, M. Ishizuka, and D. Lukose (Eds.): PRICAI 2012, LNAI 7458, pp. 112–123, 2012.

vice versa. Furthermore, different from the class label, the pairwise constraints can sometimes be obtained automatically [5].

Recently, the research on semi-supervised classification with pairwise constraints has attracted more and more interests in machine learning [6,7]. However, compared to the common models, such classification problem is much harder due to the difficulties in extracting the discriminative information from the constraints. Generally, common classifiers mostly contain loss functions and regularization terms, where loss functions measure the difference between the predictions and the initial class labels. However, pairwise constraints just represent the relationship between the data pairs rather than certain labels. Consequently, the constraints cannot be incorporated into the loss functions directly, which leads to the inapplicability of most existing classifiers. Due to the particularity of pairwise constraints, Zhang and Yan proposed a method dependent on the value of pairwise constraints (OVPC), which combines the data pairs into some new samples by outer product [8] and then designs a smoothness-driven regularized classifier based on these samples. OVPC can deal with a large number of pairwise constraints and avoid local minimum problem simultaneously. Yan et al. [9] presented a unified classification framework which consists of two loss functions for labeled data and pairwise constraints respectively and a common Tikhonov regularization term to penalize the smoothness of the classifier.

Although these methods have shown much better classification performance, they still extract the prior information from the pairwise constraints insufficiently. Firstly, they use the Tikhonov term as the regularization term which only emphasizes on the smoothness of the classifier but ignores the limited discriminative information inside the constraints. Xue et al. [10] have presented that relatively speaking, the discriminability of the classifier is more important than the usual smoothness. Hence, such regularization term is obviously insufficient for classification. Secondly, they also neglect the structural information in the data. Yeung et al. [11] have indicated that a classifier should be sensitive to the structure of the data distribution. Much related research has further validated the effectivity of the structural information for classification [12-17]. The absence of such vital information in these methods undoubtedly influences the corresponding classification performance.

In this paper, we propose a new classifier with pairwise constraints called SSDRC which stands for semi-supervised discriminatively regularized classifier for binary classification problems . Inspired by OVPC, SSDRC firstly combines the pairs into new samples. Then it applies a discriminative regularization term into the classifier instead of the traditional smoothness-driven term, which directly emphasizes on the discriminability of these new samples through using two terms to measure the intra-class compactness and inter-class separability respectively. Moreover, SSDRC also introduces the local sample geometries into the construction of the two terms in order to further fuse the structural information.

The rest of the paper is organized as follows. Section 2 briefly reviews the OVPC methods. Section 3 presents the proposed SSDRC. In Section 4, the experiment analysis is given. Conclusions are drawn in Section5.

2 Classifier Dependent On the Value of Pairwise Constraints(OVPC)

Given the training data which consist of two parts, (x_1, y_1),...,(x_m, y_m) are labeled data with y_i representing the class label, and (x_{11}, x_{12}, y'_1),...,(x_{n1}, x_{n2}, y'_n) are pairwise constraints with y'_i indicating the relationship between pairs. 1 represents must-link constraint and -1 represents cannot-link constraint. That is,

$$y'_i = \begin{cases} +1, y_{i1} = y_{i2} \\ -1, y_{i1} \neq y_{i2} \end{cases}$$

OVPC firstly combines the pair (x_{i1}, x_{i2}) into a new sample by outer product [5], which equals a transformation from the original space X to a new space $\tilde{X}$. Let z_i denote the corresponding transformed sample in $\tilde{X}$. Therefore, (z_i, y'_i) is the new sample whose class label is y'_i. Then OVPC constructs a common regularized least-squares classifier in $\tilde{X}$. That is,

$$\hat{\varphi}(n, \lambda_n) = \arg\min\{\frac{1}{n}\sum_{i=1}^{n}(y'_i - \varphi^T z_i)^2 + \lambda_n \parallel \varphi \parallel^2\} \quad (1)$$

where λ_n is the regularization parameter. $\|\varphi\|^2$ is the Tikhonov regularization term which penalizes the smoothness of the classifier. Finally, for a testing datum x, OVPC applies some simple inverse transformations on the estimator $\hat{\varphi}(n, \lambda_n)$ to get the final classification result.

Though OVPC has been shown much better performance in applications such as video object classification [8,9], it still has some limitations. On the one hand, the regularization term in OVPC is still the common Tikhonov term which cannot fully mine the underlying discriminative information inside the pairwise constraints. On the other hand, OVPC also ignores the structural information of the data distribution which can be used to enhance the classification performance.

3 Semi-Supervised Discriminatively Regularized Classifier(SSDRC)

In this section, we present SSDRC which introduces a new discriminative regularization term into the classifier instead of the smoothness-driven Tikhonov term. As a result, SSDRC can not only mine the discriminative information in the pairwise constraints more sufficiently but also preserve the local geometry of the new samples (z_i, y'_i) derived from pairwise constraints.

3.1 Data Pair Transformation

Inspired by OVPC, SSDRC firstly projects the data pair in each pairwise constraint into a new space $\tilde{X}$ as a single sample. Given the pairwise constraints

(x_{11}, x_{12}, y'_1),...,(x_{n1}, x_{n2}, y'_n), where $x_i \in R^p$. Let A denote the result of outer product between the data pair in each pairwise constraint, that is,

$$A = x_{i1} \circ x_{i2} \tag{2}$$

Here, we also use the operator *vech* which returns the upper triangular elements of a symmetric matrix in order of row as the new sample z, whose length is $(p+1) \times p/2$. That is,

$$z = vech(A + A^T - diag(A)) \tag{3}$$

For example, given $A = [a_{i,j}] \in R^{3\times3}$, $z = [a_{11}, a_{12} + a_{21}, a_{13} + a_{31}, a_{22}, a_{23} + a_{32}, a_{33}]$. Consequently, the pairwise constraints (x_{11}, x_{12}, y'_1),...,(x_{n1}, x_{n2}, y'_n) in the original space X are transformed into new samples $(z_1, y'_1), ..., (z_n, y'_n)$ in the new transformed space $\tilde{X}$.

3.2 Classifier Design in the Transformed Space

In view of the limitations in OVPC, here we aim to further fuse the discriminative and structural information hidden in the new samples into the classifier. Obviously, through the data pair transformation, the semi-supervised classification problem in the original space X has been transformed to a supervised binary-class classification task in the transformed space $\tilde{X}$, which can be solved by some state-of-the-art supervised classifiers. In terms of the least-squares loss function, Xue et al. [10] have proposed a new discriminatively regularized least-squares classifier(DRLSC). Instead of the common Tikhonov regularization term, DRLSC defines a discriminative regularization term

$$R_{disreg}(f, \eta) = \eta A(f) - (1 - \eta)B(f) \tag{4}$$

where $A(f)$ and $B(f)$ are the matrices which measure the intra-class compactness and inter-class separability of the data respectively. η is the regularization parameter which controls the relative significance of $A(f)$ and $B(f)$.

Following the line of the research in DRLSC, we further introduce the discriminative regularization term into the classifier design in $\tilde{X}$. Based on the spectral theory[14], we also use two graphs, intra-class graph G_w and inter-class graph G_b with the weight matrices W_w and W_b respectively to define $A(f)$ and $B(f)$, which can characterize the local geometry of the sample distribution in order to utilize the structural information of the new samples better.

Concretely, for each sample z_i, let $ne(z_i)$ denote its k nearest neighborhood and divide $ne(z_i)$ into two non-overlapping subsets $ne_w(z_i)$ and $ne_b(z_i)$. That is,

$$ne_w(z_i) = \{z_i^j | \text{ if } y'_i = y'_j, 1 \le j \le k\}$$

$$ne_b(z_i) = \{z_i^j | \text{ if } y'_i \ne y'_j, 1 \le j \le k\}$$

Then we put edges between z_i and its neighbors, and thus obtain the intra-class graph and inter-class graph respectively. The corresponding weights are defined as follows:

$$W_{w,ij} = \begin{cases} 1, & \text{if } z_j \in ne_w(z_i) \text{ or } z_i \in ne_w(z_j); \\ 0, & \text{otherwise.} \end{cases}$$

$$W_{b,ij} = \begin{cases} 1, & \text{if } z_j \in ne_b(z_i) \text{ or } z_i \in ne_b(z_j); \\ 0, & \text{otherwise.} \end{cases}$$

The goal of SSDRC is to keep the neighboring samples of G_w as close as possible while separate the connected samples of G_b as far as possible. Thus,

$$A(f) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{w,ij} \parallel f(z_i) - f(z_j) \parallel^2$$

Similarly,

$$B(f) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{b,ij} \parallel f(z_i) - f(z_j) \parallel^2$$

Assume that the classifier has a linear form, that is,

$$f(z) = w^T z \tag{5}$$

Substitute the equation (5) into $A(f)$ and $B(f)$ and then obtain

$$\begin{aligned} A(f) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{w,ij} \parallel f(z_i) - f(z_j) \parallel^2 \\ &= w^T Z(D_w - W_w)Z^T w \\ &= w^T Z L_w Z^T w \end{aligned}$$

where D_w is a diagonal matrix and its entries $D_{w,ij} = \sum_j W_{w,ij}$, $L_w = D_w - W_w$ is the laplacian matrix of G_w.

$$\begin{aligned} B(f) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} W_{b,ij} \parallel f(z_i) - f(z_j) \parallel^2 \\ &= w^T Z(D_b - W_b)Z^T w \\ &= w^T Z L_b Z^T w \end{aligned}$$

where D_b is a diagonal matrix and its entries $D_{b,ij} = \sum_j W_{b,ij}$, $L_b = D_b - W_b$ is the laplacian matrix of G_b.

The final optimization function can be formulated as

$$\min\{\frac{1}{n}\sum_{i=1}^{n}(y_i' - f(z_i))^2 + \eta A(f) - (1-\eta)B(f)\} \tag{6}$$

that is,

$$\min\{\frac{1}{n}\sum_{i=1}^{n}(y_i' - w^T z_i)^2 + w^T Z[\eta L_w - (1-\eta)L_b]Z^T w\} \tag{7}$$

The solution of the optimization function can follow from solving a set of linear equations by embedding equality type constraints in the formulation. Interested reader can refer the literature [10] for more details.

It is worthy to point out that, although the classifier design in SSDRC is similar to DRLSC, the corresponding classifier is defined in the transformed space $\tilde{X}$ rather than the original space as that in DRLSC. As a result, for a new testing sample, SSDRC should firstly conduct the classifier in the transformed space and then get the final classifier through some additional inverse transformations rather than DRLSC that predicts the class label in the original space directly. The particular process of the inverse transformation will be given in the next subsection.

3.3 Classification in the Original Space

Here we apply the inverse operation of *vech* to the discriminative vector w obtained in the transformed space $\tilde{X}$, resulting in a $p \times p$ symmetric matrix Θ. That is,

$$\Theta = vech^{-1}(w) \tag{8}$$

For example, given $w = \{a_{11}, a_{12}, a_{13}, a_{22}, a_{23}, a_{33}\}$, $\Theta = [a_{11}, a_{12}, a_{13}; a_{12}, a_{22}, a_{23};\ a_{13}, a_{23}, a_{33}] \in R^{3\times 3}$, . Then we perform the eigen-decomposition to the symmetric matrix Θ, and obtain the largest eigenvalue s_1 and its corresponding eigenvector u_1. We select

$$\hat{\theta} = \sqrt{s_1}u_1$$

as the sign-insensitive estimator of $\hat{\beta}$, which is the discriminative vector in the original space. Here sign-insensitive means that the real sign of $\hat{\beta}$ is still unknown. The labeled sample (x_1, y_1),...,(x_m, y_m) are used to determine the correct sign of $\hat{\beta}$.

To be more specific, the real sign of $\hat{\beta}$ can be computed as

$$s(\hat{\theta}) \begin{cases} +1, \sum_{i=1}^{m} I(y_i\hat{\theta}^T x_i) \geq \lceil \frac{m}{2} \rceil; \\ -1, \text{otherwise.} \end{cases} \tag{9}$$

where $\lceil t \rceil$ is the ceil function which returns the smallest integer value that is no less than t [5]. So the real estimator of the discriminative vector in the original space is

$$\hat{\beta} = s(\hat{\theta})\hat{\theta} \tag{10}$$

For a new testing datum x, the predicted class label is

$$\tilde{y} = \hat{\beta}^T x \tag{11}$$

3.4 The Pseudo-code for SSDRC

Based on the previous analysis, we present the SSDRC method. The corresponding pseudo-code is summarized in Algorithms 1.

input : Labeled Samples$\{(x_i, y_i)\}_{i=1}^m$;
Pairwise Constraints $\{(x_{j1}, x_{j2}, y'_j)_{j=1}^n\}$;
The number k of the nearest neighbors of new sample z_i derived from the pairwise constraints;
The regularization parameters η $(0 \leq \eta \leq 1)$
output: the estimator $\hat{\beta}$ of discriminative vector

for $i \leftarrow 1$ **to** n **do**
  $A = x_{i1} \circ x_{i2}$;
  $z_i = vech(A + A^T - diag(A))$;
end
for $j \leftarrow 1$ **to** n **do**
  $z_j^k \leftarrow k$th nearest neighbor of z_j among $(z_i)_{i=1}^n$;
end
for $i \leftarrow 1$ **to** n **do**
  for $j \leftarrow 1$ **to** k **do**
    if $y'_i = y'_j$ **then** $W_{w,ij} \leftarrow 1$;
    else $W_{b,ij} \leftarrow 1$;
  end
end
$D_{w,ij} = \sum_j W_{w,ij}$; $D_{b,ij} = \sum_j W_{b,ij}$;
$L_w = D_w - W_w$; $L_b = D_b - W_b$;
$w \leftarrow$ solve the optimization function:

$$\arg\min\{\frac{1}{n}\sum_{i=1}^{n}(y'_i - w^T z_i)^2 + w^T Z[\eta L_w - (1-\eta)L_b]Z^T w\}$$

$\Theta = vech^{-1}(w)$
Compute the largest eigenvalue s_1 and its corresponding eigenvector u_1 of Θ
$\hat{\theta} = \sqrt{s_1}u_1$
if $\sum_{i=1}^m I(y_i\hat{\theta}^T x_i \geq 0) \geq \lceil \frac{m}{2} \rceil$ **then** $s(\hat{\theta}) = +1$;
else $s(\hat{\theta}) = -1$;

$$\hat{\beta} = s(\hat{\theta})\hat{\theta}$$

Algorithm 1. Pseudo-code for SSDRC

4 Experiments

In this section, we evaluate the performance of our SSDRC algorithms on the real-word classification datasets: six datasets in UCI[1] and IDA datasets[2] in comparison to some state-of-the-art algorithms shown in the Table 1. We select the supervised method RLSC as the baseline. OVPC and PKLR are two popular semi-supervised classifiers with pairwise constraints.

[1] The dataset is available from
`http://www.ics.uci.edu/ mlearn/MLRepository.html`

[2] The database is available from
`http://ida.first.fraunhofer.de/projects/bench/benchmarks.htm`

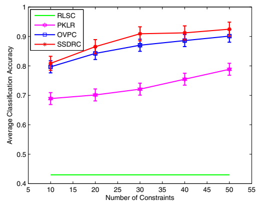
Table 1. The acronyms, full names and citations algorithms compared with the SSDRC in the experiments

Acronym	Full name	Citation
RLSC	Regularized Least Square Classifier	[18]
OVPC	On the Value of Pairwise Constraints	[8]
PKLR	Pairwise Kernel Logistic Regression	[9]

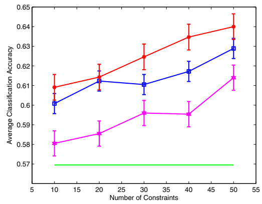
4.1 UCI Dataset

In this section we compare the relative performance of SSDRC with other three classification algorithms on six datasets in UCI, namely *Water(39, 116)*, *Sonar(60, 208)*, *Ionosphere(34, 351)*, *Wdbc*(31,569), *Pima*(9,768), *Spambase*(58,4600). These datasets are typical binary-classification datasets in UCI. To be more specific, the first element in the brackets represents the dimension while the second means the number of samples in each dataset. Notice that the scale of the dataset is increment, and we divide each dataset into two equal parts. One is for training set and the other is for testing set. In our experiment, the pairwise constraints are obtained randomly selecting pairs of instances from the training set, and creating must-link and cannot-link constraints. The number of constraints is changed from 10 to 50 at a rate of 10 increment. Moreover, In SSDRC, the number of the k nearest neighbors is selected from $\{5, 10, 15, 20\}$. Especially, when the number of pairwise constraints is 10 which is relatively less to select the large number of nearest neighbors, the value of k is selected from $\{5, 10\}$. Moreover, in PKLR, we select the liner kernel as the kernel function. The regularization parameters λ in OVPC and PKLR are selected from $\{2^{-10}, 2^{-9}, ..., 2^{9}, 2^{10}\}$, and the regularization parameter η in SSDRC is chosen in $[0, 0.1, ..., 0.9, 1]$. All the parameter selections are done by cross-validation. Since labeled samples are only used to determine the real sign of the estimator, so we only select one sample for each class. The whole process is repeated 100 runs and the average results are reported.

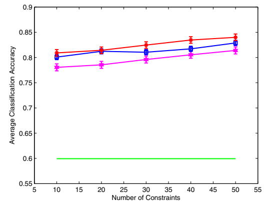
Figure 1 shows the corresponding average classification accuracies of the algorithms in the six datasets. From the figure, we can see that the accuracies of OVPC, PKLR and SSDRC are basically improved with the increase of the number of the pairwise constraints step by step, which validates the "No Free Lunch" Theorem [14], that is, with more prior information incorporated, the better classification performance we can get. In the comparison of the four algorithms, the performance of RLSC is always the worst as a straight line, since it only uses the limited labeled data, which justifies the significance of the pairwise constraints data in semi-supervised learning. Furthermore, SSDRC outperforms PKLR and OVPC at each same number of pairwise constraints in all the six datasets, especially in Water, Wdbc and Pima, with more than 10% improvement in average. Besides, the variance of experimental result in SSDRC is much less than the ones in other three algorithms on most datasets. This also demonstrates that the utilization of pairwise constraints and structural information in SSDRC is much better than PKLR and OVPC.



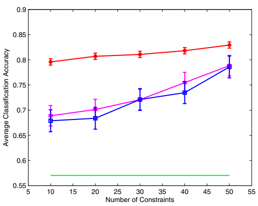
(a) Water



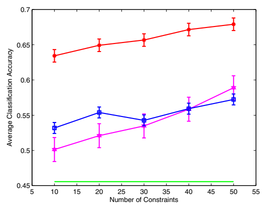
(b) Sonar



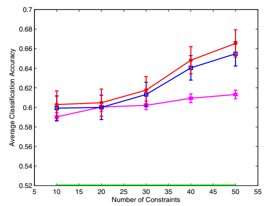
(c) Ionosphere



(d) Wdbc



(e) Pima



(f) Spambase

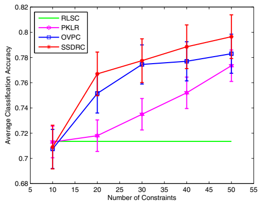
Fig. 1. Classification accuracy on UCI database with different number of constraints

Table 2. The attributes of the thirteen datasets in the IDA database

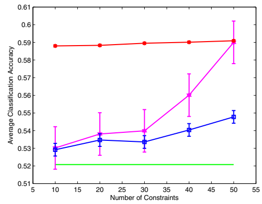
Dataset	Dimension	Training Set Size	Testing Set Size
Heart	13	170	100
Banana	2	400	4900
Breast-cancer	9	200	77
Diabetis	8	468	300
Flare-solar	9	666	400
German	20	700	300
Ringnorm	20	400	7000
Thyroid	5	140	75
Titanic	3	150	2051
Twonorm	20	400	7000
Waveform	21	400	4600
Image	18	1300	1010
Splice	60	1000	2175

4.2 IDA Database

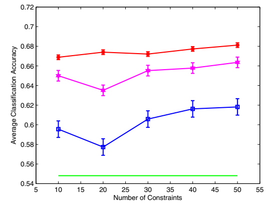
In this subsection, we further evaluate the performance of the SSDRC algorithm on the IDA database, which consists of thirteen datasets, and all of them has two classes. The training and testing sets have been offered in each dataset already. Table 2 shows the attributes of the thirteen datasets in the IDA database: the number of dataset's dimension, the size of training set and testing set respectively. The experimental settings are the same as those in the previous UCI datasets.



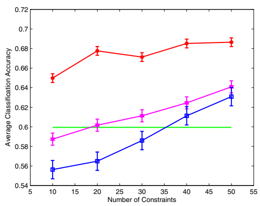
(a) Heart



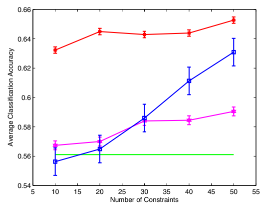
(b) Banana



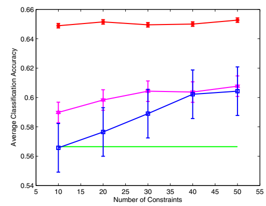
(c) Breast-cancer



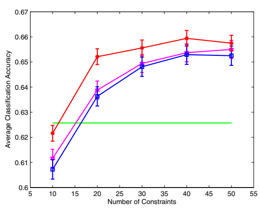
(d) Diabetis



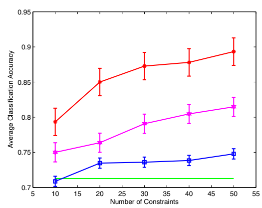
(e) Flare-solar



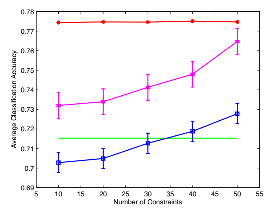
(f) German



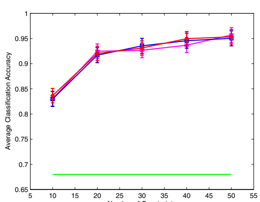
(g) Ringnorm



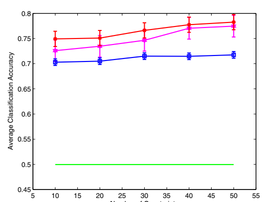
(h) Thyroid



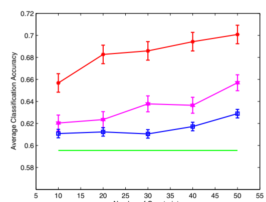
(i) Titanic



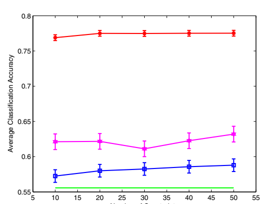
(j) Twonorm



(k) Waveform



(l) Image



(m) Splice

Fig. 2. Classification accuracy on IDA database with different number of constraints

Figure 2 shows the corresponding average classification accuracies of the four algorithms in the IDA datasets. From the figure, we can see that the SSDRC outperforms other three algorithms obviously in most datasets, especially in Banana, Splice, Titanic and German. The reason more likely lies in that different from the other algorithms, SSDRC embeds the local structure involved in data and makes use of the discriminative information in constraints more sufficiently, which results in its superior performance in the real-world classification tasks.

5 Conclusion

In this paper, we propose a novel classification method with pairwise constraints SSDRC. Different form many existing classifiers, SSDRC firstly transforms the data pairs in pairwise constraints into some new samples and then designs a discriminability-driven regularized classifier in the transformed space, which can not only fully capture the discriminative information in the constraints but also preserve the local structure of these new samples. Experimental results demonstrates that SSDRC is much better than the popular related classifiers OVPC and PKLR.

Throughout the paper, SSDRC focuses on the binary classification problems. How to extend SSDRC to the multi-class problems deserves our further work. Furthermore, the kernelization of SSDRC also needs more study.

Acknowledgements. This work was supported by National Natural Science Foundation of China (Grant No. 60905002) and the Scientific Research Startup Project for New Doctoral Faculties of Southeast University. Furthermore, the work was also supported by the Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education.

References

1. Zhu, X.: Semi-supervised learning literature survey. Technical Report, 1530, Madison: Department of Computer Sciences, University of Wisconsin (2008)
2. Chapelle, O., Schölkopf, B., Zien, A.: Semi-Supervised Learning. MIT Press, Cambridge (2006)
3. Han, X., Chen, Y., Xiang, R.: Semi-supervised and Interactive Semantic Concept Learning for Scene Recognition. In: 20th International Conference on Pattern Recognition, pp. 3045–3048 (2010)
4. Martnez-Us, A., Sotoca, J.M.: A Semi-supervised Gaussian Mixture Model for Image Segmentation. In: 20th International Conference on Pattern Recognition, pp. 2941–2944 (2010)
5. Tang, W., Zhong, S.: Pairwise constraints-guided dimensinality reduction. In: SDM 2006 Workshop on Feature Selection for Data Mining, Bethesda, MD (2006)
6. Zhang, D., Zhou, Z., Chen, S.: Semi-supervised dimensionality reduction. In: SIAM Conference on Data Mining, ICDM (2007)

7. Chen, S., Zhang, D.: Semi-supervised dimensionality reduction with pairwise constraints for hyper spectral image classification. IEEE Geoscience and Remote Semsing Letters 8(2), 369–373 (2011)
8. Zhang, J., Yan, R.: On the value of pairwise constraints in classification and consistency. In: Proceedings of the 24th International Conference on Machine Learning, pp. 1111–1118 (2007)
9. Yan, R., Zhang, J., Yang, J., Hauptmann, A.: A discriminative learning framework with pairwise constraints for video object classification. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(4), 578–593 (2006)
10. Xue, H., Chen, S., Yang, Q.: Discriminatively regularized least-squares classification. Pattern Recognition 42(1), 93–104 (2009)
11. Yeung, D.S., Wang, D., Ng, W.W.Y., Tsang, E.C.C., Zhao, X.: Structured large margin machines: sensitive to data distributions. Machine Learning 68, 171–200 (2007)
12. Xue, H., Chen, S., Yang, Q.: Structural regularized support vector machine: a framework for structural large margin classifier. IEEE Trans. on Neural Networks 22(4), 573–587 (2011)
13. Lanckriet, G.R.G., Ghaoui, L.E., Bhattacharyya, C., Jordan, M.I.: A robust minimax approach to classfication. J. Machine Learning Research 3, 555–582 (2002)
14. Huang, K., Yang, H., King, I., Lyu, M.R.: Learning large margin classifiers locally and globally. In: The 21th International Conference on Machine Learning, ICML (2004)
15. Belkin, M., Niyogi, P., Sindhwani, V.: Manifold regularization: A geometric framework for learning from examples. Department of Computer Science, University of Chicago, Tech. Rep: TR-2004-06 (2004)
16. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (2001)
17. Pan, J.J., Pan, S.J., Yin, J., Ni, L.M., Yang, Q.: Tracking mobile users in wireless networks via semi-supervised colocalization. IEEE Trans. on Pattern Analysis and Machine Intelligence 34(3), 587–600 (2012)
18. Haykin, S.: Neural Networks: A Comprehensive Foundation. Tsinghua University