

# Research Semi-Supervised Learning via Kernel Function Method

Jianfeng Liu<sup>1</sup> and Jia Lv<sup>2,\*</sup>

<sup>1</sup>Chongqing Three Gorges University, Wanzhou, Chongqing, China, 404100

<sup>2</sup>College of Computer and Information Science of Chongqing Normal University, Chongqing, China, 401331

\*Corresponding author

**Abstract**—Recently, researchers focus on Graph-based semi-supervised learning, how to construct a graph and introduce kernel function into semi-supervised learning affect the effect of the algorithm, and mostly. This paper proposed a improved graph-based method and introduce kernel function into semi-supervised learning, the experimental verification algorithm has achieved some results.

**Keywords**—semi-supervised learning; kernel function; construct graph; graph-based

## I. INTRODUCTION

Data mining technology is the core field of artificial intelligence. It is an important way to deal with large amounts of data and knowledge. There are three kinds of [1]: training data is labeled with the class label and the training sample is Learning Supervised. Only a small number of training samples are labeled, and a large number of training samples are not labeled as semi-supervised learning.

According to semi-supervised learning, the graph-based method is an important and effective learning algorithm, which is based on a complete mathematical foundation, and has improved with the development of basic mathematics.

Graph-based semi supervised learning algorithm, the focus is on the structure of the graph. Most graphs are constructed using the Laplace term, and the classification model  $\hat{f}$  is estimated by means of information. The classification model  $\hat{f}$  has two characteristics: the value  $\hat{f}$  should be equal or close to the labeled data, which is guaranteed by the loss function of the objective function. Moreover, the data points on the graph should be smooth, the assumption of manifold false assumption and clustering is satisfied.

## II. KERNEL FUNCTION

When the data dimension is too high, the graph-based method is non-separable. The kernel function can be mapped to a high dimensional space into a linear problem of low dimensional space, which can be linearly separable. Compared to the geometry of the new kernel Hilbert space, the kernel function has been introduced to the semi-supervised learning, which is based on the graph. The semi-supervised kernel objective function is given as follows, Sindhwani. V[2]:

$$f = \arg \min_{f \in H} \left\{ \frac{1}{l+u} \sum_{i=1}^{l+u} c(f, x_i, y_i) + \phi(\|f\|_H) \right\} \quad (1)$$

Through the calculation of the kernel transformation, the relationship between the kernel function of the semi-supervised kernel space and the original space is as follows:

$$\tilde{K}(x_i, y_i) = K(x_i, y_i) - K_x^T (I + MK)^{-1} MK_x \quad (2)$$

Among them,  $K_x = [K(x_1, x), \dots, K(x_{l+u}, x)]^T$ ,  $\tilde{K}(x_i, y_i)$  is Kernel function of semi-supervised kernel,  $K(x_i, y_i)$  is Kernel function of the original space. The kernel function  $K$  can be selected as the Gauss kernel function, the linear kernel function and the polynomial kernel function. Literature [3]-[5] gives a objective function as follows:

$$\arg \min \frac{1}{l} \sum_{i=1}^l C(x_i, y_i, f) + \gamma_A \|f\|^2 + \gamma_I \|f\|_I^2 \quad (3)$$

It is not difficult to find that the problem is two quadratic programming problems. The optimal solution can be obtained by solving the optimal solution as follows:

$$\min_F ((F - Y)^T D (F - Y) + \frac{\gamma_1}{n^2} F^T L F + F^T H F) \quad (4)$$

$$F = ((I - A)^T (I - A) + \frac{\gamma_1}{n^2} (D - W) + D)^{-1} D Y \quad (5)$$

The optimal solution of the objective function of the semi-supervised kernel is obtained by G S Kimeldorf [6], as follow:

$$F = \sum_{i=1}^n a_i K(a_i, \cdot) \quad (6)$$

From (5) and (6) can be known to have the following:

$$\sum_{i=1}^n \alpha_i K(a_i, \cdot) = ((I - A)^T (I - A) + \frac{\gamma_1}{n^2} (D - W) + D)^{-1} D Y \quad (7)$$

When  $F$  is solved,  $\alpha_i$  can be calculated. Any given a sample  $x_i$ , data can be computed by a linear combination of linear programming, where  $L, D$  are weight measurement of

kernel space, and satisfied:  $L = D - W$ . The nonlinear mapping  $\varphi$  transforms the data  $X$  in the original sample space  $\varphi(X)$  to the data in the kernel space, and then calculates the distance between two points in the kernel space as follows:

$$d(x_i, x_j) = d(\varphi(x_i), \varphi(x_j)) = \|\varphi(x_i) - \varphi(x_j)\| \quad (8)$$

The following distance formula can be obtained by means of kernel transformation:

$$d(x_i, x_j) = \sqrt{K_{ii} - 2K_{ij} + K_{jj}} \quad (9)$$

where  $K$  is Gauss kernel function.

By computing (7), the label  $y_i$  of data sample  $x_i$  is obtained by the following formula:

$$y_i = \sum_{i=1}^m a_i x_i \quad (10)$$

The emphasis of this algorithm is the calculation of the kernel space distance and the solution of the linear coefficients. According to the nature of the kernel, the distance of the kernel space can be transformed into Euclidean distance. The linear coefficients can be solved by solving linear programming problem.

### III. KERNEL FUNCTION

Consider global manifold and local clustering hypothesis, based on kernel the semi supervised learning algorithm, using MATLAB programming, in the process of implementation of the algorithm, the first parameter is adjusted, the algorithm is greatly affected by the parameters, the experimental parameters are mainly related to  $\gamma$  and  $\lambda$ , they are respectively as follows:

$$\gamma = [0.01; 0.04; 0.2; 0.4; 0.8; 1; 2; 4; 8; 16]$$

$$\lambda = [0.2; 0.4; 1; 4; 8; 10; 12; 16; 24; 36]$$

The parameter adjusting experiment can be obtained following experimental results as follows,  $F$ , which is shown in Figure, is already the value after the symbol function transformation,  $f$  is a real valued when is calculation.

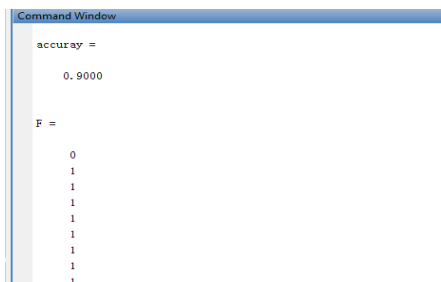


FIGURE I. :THE VALUE OF THE VALUE FOR THE CLASSIFICATION

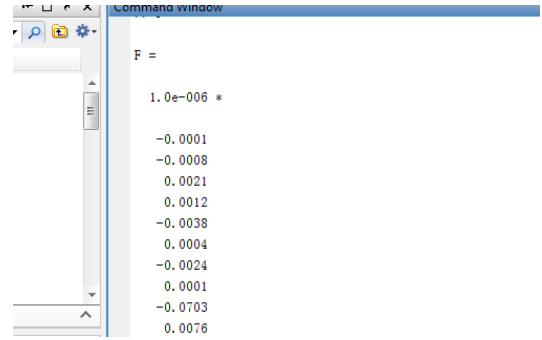


FIGURE II. THE REAL VALUE F

The classification performance of the algorithm is greatly affected by the parameters, and the classification performance of the algorithm is better than that of the algorithm. In the process of experiment, the parameters  $\lambda$  of the classification of the parameters are greatly affected, and the relationship of the correct classification rate is as follows.

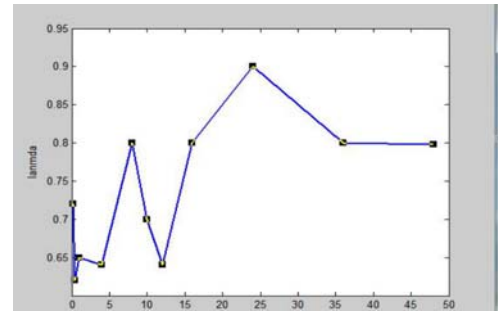


FIGURE III. CLASSIFICATION ACCURACY AND RELATIONSHIP

Comparison test. In order to verify the effectiveness of the algorithm, the algorithm based on the kernel of the algorithm is compared with the global and local semi supervised classification algorithm [7], which is based on the kernel of the KLGSS (GL\_Reg). Experiment repeated 20 times, the average classification accuracy and the standard error of the two algorithms are obtained from table1. Although there is no obvious advantage in classification accuracy, the proposed algorithm is better than the contrast algorithm in the standard error.

TABLE I. COMPARISON RESULTS: THE CORRECT RATE (CR) , THE ALGORITHM AND THE STANDARD ERROR(SE)

Date sets	GL-Reg		KGLSS	
	CR	SE	CR	SE
students	89.10	2.79	89.5	1.64

### IV. SUMMARY AND FUTURE WORK

In this paper, we study the semi-supervised learning via graph, and introduced the kernel function. In a certain range, solved the high dimensional space linearly inseparable problem, Experiments have achieved good results, the future work is to verify the more and practical data set and to study the structure of the graph.

## ACKNOWLEDGMENTS

The research work was supported by the Chongqing Natural Science Project (cstc2014jcyjA40011), Science and technology project of Chongqing Municipal Education Commission (KJ1400513).

## REFERENCE

- [1] Lippmann R. Pattern classification using neural networks. IEEE Transactions on Communications Magazine, 1989, 27(11):256-267.
- [2] Sindhwani V, Niyogi P, Belkin M. Beyond the point cloud: from transductive to semi-supervised learning. In: Proceedings of the 22nd International Conference on Machine Learning. Bonn, Germany: MIT Press, 2005. 824-831.
- [3] ZHU X J, GHAHARMANI Z, LAFFERTY J. Semi-supervised learning using Gaussian fields and harmonic functions[C]. FAWCETT T, MISHRA N. Proceedings of 20th International Conference on Machine Learning, Menlo Park: AAAI Press, 2003: b912- 919.
- [4] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples[J]. Journal of Machine Learning Research, 2006, (11): 2399-2434.
- [5] XIANG S M, NIE F P, ZHANG C S. Semi-supervised classification via local spline regression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 32(11): 2039-2053.
- [6] Kimeldorf G S, Wahba G. Some results on tchebycheffian spline functions. Journal of Mathematical Analysis and Applications, 1971, 33(1): 82-95
- [7] K. Nigam, A. McCallum, S. Thrun et al. Text classification from labeled and unlabeled documents using EM[C]. International Conference on Machine Learning, Stanford, USA, June, 2000, 39(2-3): 103-134.