# A Robust Graph Based Learning Approach to Agricultural Data Classification

Baojie Ji, Caili Su, Wanzhong Lei



# A Robust Graph based Learning Approach to Agricultural Data Classification

Baojie Ji [1] , Caili Su [2] , Wan zhong Lei [3]
[1] Mechanical & Electrical Engineering College, Hennan Agricultural University,Zhengzhou 450002, China;
[2] Institute of Environmental Science and Engineering; North China University of Water Resources and Electric Power; Zhengzhou 450011, China
[3] Department of Electrical Information Engineering, Henan Institute of Engineering,Zhengzhou 451191, China
hnkfjybj@163.com

**Abstract.** This paper proposes a novel graph based learning approach to classify agricultural datasets, in which both labeled and unlabelled data are applied to the classification procedure. In order to capture the complex distribution of data, we propose a similarity refinement approach to improve the robustness of traditional label propagation. Then the refined affinity matrix is applied to label propagation. Thus, the traditional pair-wise similarity is updated with scores using median filter of its neighbors in manifold space. And the proposed classification approach can propagate the labels from the labeled data to the whole dataset. The experiments over agricultural datasets have shown that embedding information fusion approach in manifold space is beneficial in classification.

**Keywords:** semi-supervised learning; graph based learning; label propagation

## 1 Introduction

With the development of precision agriculture, probabilistic modeling and machine learning method has attracted great research attention in the agriculture domain. Many representative classifiers have been used in agriculture classification, such as the two-dimensional multi-resolution hidden Markov models [1], support vector machine [2],[3], Bayes Point Machine [4], and Mixture Hierarchical Model [5], but the classification result can not be satisfied. In this paper, we propose novel graph based learning approach to classify agriculture dataset. The pair-wise similarities of dataset are first computed in attribute space. To improve the robustness of traditional graph based learning methods, we then refine the pair-wise similarities by embedding the neighbor information in manifold space.

The rest of the paper is organized as follows. Section 2 describes the theory of graph based learning. Section 3 introduces the proposed classification scheme. Section 4 represents the key algorithm. Section 5 reports and discusses the experimental results and Section 7 concludes the paper.

## 2 Graph-based Semi-supervised Learning

### 2.1 Graph-based Learning Theory

Semi-supervised learning (SSL)[6],[7],[8],[9],[10],[11],[12], which attempts to learn from both labeled and unlabeled data, is a promising approach recently. As a major family of SSL, graph-based learning[13] algorithms have been the focus of much recent machine leaning research. The key to which learning is the two basic assumptions: 1) neighborhood assumption: nearby points are likely to have the same label; 2) structure assumption: points on the same structure (such as a cluster or a sub-manifold) are prone to have the same label. Note that the first assumption is local, while the second one is global. The cluster assumption implies us to consider both local and global information during learning.

In graph-based learning, data points $(1,\cdots,l$ labeled points and $(l+1,\cdots,n$ unlabeled points) are arranged in a weighted undirected graph. The graph is characterized by a weight matrix $W$, whose elements $W_{ij} \geq 0$ are similarity measures between vertices $i$ and $j$, and by its initial label vector $Y_L = (y_1,\cdots,y_l), y_i \in \{1,\cdots,C\}$,that defines labels for the first $l$ points. If there is no edge linking nodes $i$ and $j$ $W_{ij} = 0$ .Other than that, applications have considerable freedom in choosing the edge set and the $W_{ij}$ weights. The commonly used weight is defined by a Gaussian kernel: $W_{ij} = \exp(-d(x_i, x_j)^2/(2\sigma)^2)$ where $d(x_i, x_j)$ is the estimated distance between feature vector $x_i$ and $x_j$ , $\sigma$ is a bandwidth hyper-parameter. Various distance measures can be used, e.g. Cosine distance, Euclidean distance or Jensen Shannon.

### 2.2 The Robustness of Graph-based Learning

Although graph based learning, which attempts to learn from both labeled and unlabeled data, is a promising approach to deal with classification, one of the key problems is the robustness. This can be easily understood from the label propagation framework, which states that predicting the labels of the unlabeled data on a graph is equivalent to propagating the labels of the labeled data along the edges to the unlabeled ones. Clearly, the existence of the bridge points on the graph will cause the labels to wrongly propagate to different classes. The two moon example in [14] shows that the bridging points can bias the final classification results severely.

## 3 The Proposed Classification Approach based on Graph-based Learning

### 3.1 Graph Construction

Suppose the dataset is $X=(x_1,\cdots,x_l,x_{l+1},\cdots,x_n)\subset R^m(1,\cdots,l$ labeled dataset and $l+1,\cdots,n$ unlabeled dataset), the initial similarity matrix $W$, is constructed as

$$W_{ij}=\exp(-d(x_i,x_j)^2/(2\sigma)^2) \quad (1)$$

Where $d(x_i,x_j)$ is the distance in feature space. L1 distance is used in our implementation. $\sigma$ is a bandwidth hyper-parameter, Note that $W_{ij}=0$ because there are no loops in the graph. We use the KNN approach to construct the graph.

To alleviate the bridge points in traditional graph based learning to improve robustness in classification, we propose a pair-wise similarity refinement scheme to refine the traditional approach, in which the average Hausdorff distance[15] of the neighborhood information of each data is used to refine the traditional pair-wise matrix $W$,

$$aveH(A,B)=\frac{\sum_{a\in Ab\in B}\min\|a-b\|+\sum_{b\in Ba\in A}\min\|b-a\|}{|A|+|B|} \quad (2)$$

Where $|\bullet|$ measures the cardinality of a set. In other words, $aveH(\bullet,\bullet)$ averages the distances between neighbor information of the data and that in the other data. So

$$\tilde{W}_{ij}=\exp(-d(A,B)^2/(2\sigma)^2 \quad (3)$$

Thus the similarity becomes the fusion information of its neighborhood instead of the initial pair-wise similarity. In this way, the bridge point problem can be alleviated to some extent. Then symmetrically normalize $\tilde{W}$ by $S=D^{-1/2}WD^{-1/2}$. where $D$ is a diagonal matrix , $D_{ii}=\sum_{j=1}^{n}\tilde{W}_{ij}$ and $S$ is the initial pair-similarity symmetrical matrix.

### 3.2 Label Propagation

After the graph has been constructed, we predict the classification labels of the unlabeled data in the dataset with the data similarities.

Suppose there are $C$ classification labels, and the classification label set is $L=\{1,\cdots,C\}$. Let $M$ be $n\times c$ matrices with nonnegative real valued entries. $F$ denote the set of label functions defined on $X$, initialize a $n\times c$ matrix $T$ with binary vectors encoding the known labels for the first $l$ rows, $F_i=\delta_C(y_i)\forall i\in\{1,2,\cdots,l\}$,in each propagation step, we let each data fuse a fraction of label information with its similar data, therefore, the label of $X$ at time $m+1$ becomes

$$F^{m+1}=aSF^m+(1-a)T \tag{4}$$

where $a$ is the parameter and $a\in(0,1)$ We will use (4) to update the labels of each image until convergence. The theorem in [14] guarantees that the sequence $\{F^m\}$ converges to

$$F^*=(1-a)(I-aS)^{-1}T \tag{5}$$

## 4 The Proposed Key Algorithm

Step1. Compute the k nearest neighbors of each data with L1 distance.

Step2. Construct the semantic similarity matrix $\tilde{W}$ as $\tilde{W}_{ij}=\exp(-d(A,B)^2/(2\sigma)^2$ , average Hausdorff distance is used in our implementation.

Step3. Symmetrically normalize $\tilde{W}$ by $S=D^{-1/2}WD^{-1/2}$ , where D is a diagonal matrix and $D_{ii}=\sum_{j=1}^{n}\tilde{W}_{ij}$ .

Step4 Do iteration according to (4) $F^{m+1}=aSF^m+(1-a)T$ until convergence.

Step5. According to (5) decide the label for each unlabeled image based on the convergent matrix of $F^*$.

## 5 Experimental Results and Discussion

### 5.1 Experimental Data Sets

To evaluate the proposed techniques, we use the UCI[15] agricultural dataset on mushroom and soybean. The first one includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms in the Agaricus and Lepiota Family.

The attribute information is 22D, such as cap-shape, cap-surface, gill-color, stalk-root and so on. The second dataset has 19 classes, there are 35 categorical attributes. The construction of each data set for our experiments is done as follows: Firstly, we randomly select 40% of the data from the each category, and put them into test set to evaluate the performance of classifier. Then, the rest are used to create training sets.

### 5.2 Performance Measures

To analyze the performance of classification, we adopt the popular F1 measure. F1 measure is combination of recall (re) and precision (pr), $F_1 = 2.re.pr/(re + pr)$. Precision means the rate of documents classified correctly among the result of classifier and recall signifies the rate of correct classified documents among them to be classified correctly. The F1 measure which is the harmonic mean of precision and recall is used in this study since it takes into account effects of both quantities.

### 5.3 Experimental Results and Discussion

The comparison between the traditional SVM method and the proposed scheme are showed in figure1 and figure2, from the result, it can be seen that the method is effective. Our approach yields a higher performance compared to traditional methods over all categories. For mushroom, our approach yields values of 86.4% average classification result, whereas the SVM yields the F1 values of 84.2%. For figure2, our approach also yields higher average classification performance of 21.2% over the SVM methods.

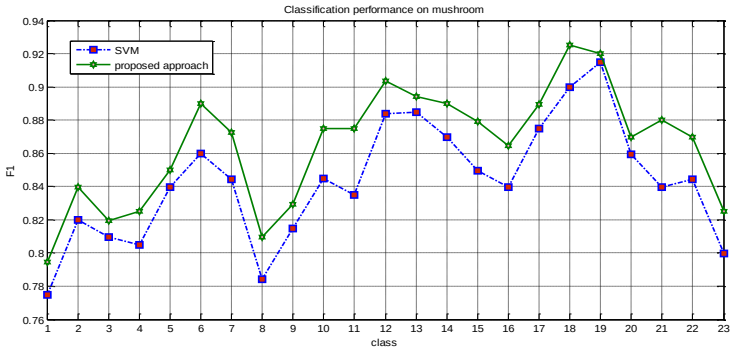


Fig.1.Comparison of classification performance on mushroom

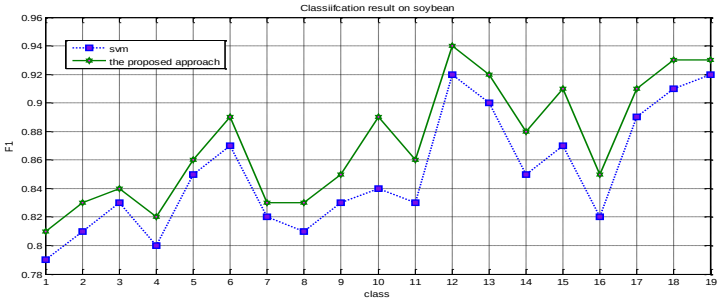


Fig.2.Comparison of classification performance on soybean

## 6 Conclusion

In this paper, a novel approach to agricultural data classification based on a novel graph based learning approach is proposed. First the initial pair-wise similarities in dataset are computed based on the distance in attribute space, then it is refined by the fusion of its neighborhood information. Finally the updated similarities are adopted to propagate the class labels from the labeled dataset to the whole dataset. Experimental results show that the proposed approach can propagate more accurate classification results and improve the robustness of the traditional scheme.

## References

1. J. Li and J. Z. Wang. Automatic Linguistic Indexing of Pictures by a Statistical Modeling Approach. IEEE Trans.Pattern Anal. Mach. Intell. 25, 9 (Sep. 2003), 1075--1088(2003)
2. C. Cusano, G. Ciocca, and R. Schettini. Image Annotationusing SVM. In Proceedings of Internet imaging IV, Vol.SPIE 5304.(2004)
3. Y. Gao, J. Fan, H. Luo, X. Xue, and R. Jain. Automaticimage annotation by incorporating feature hierarchy andboosting to scale up SVM classifiers. In Proceedings of the 14th Annual ACM international Conference on Multimedia (Santa Barbara, CA, USA, October 23 -- 27( 2006)
4. E. Chang, G. Kingshy, G. Sychay, and G. Wu. CBSA:content-based soft annotation for multimodal image retrieval using Bayes point machines. IEEE Trans. on CSVT,13(1):26--38, Jan(2003)
5. G. Carneiro and N. Vasconcelos. A database centric view of semantic image annotation and retrieval. In Proceedings of the 28th Annual international ACM SIGIR Conference on Research and Development in information Retrieval(Salvador, Brazil, August 15--19(2005)
6. Lippmann R P. Pattern classification using neural networks. IEEE Communications, 27(11): 47--64(1989)
7. Chapelle, A.Zien, B, “Scholkopf. Semi-supervised learning,” MIT press(2006)
8. Shahshahani B, Landgrebe D. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. IEEE Transactions on Geoscience and Remote Sensing, 32(5):1087--1095(1994)

9. Zhu X, Ghahramani Z, Lafferty J. Semi-supervised learning using Gaussian fields and harmonic functions.In: Proceedings of the 20th International Conference on Machine Learning,912--919(2003)
10. Zhou D,Bousquet O,Lal T N,Weston J,Schollkopf B. Learning with local and global consistency. In:Thrun S,Saul L,Scholkopf B,eds. Advances in Neural Information Processing Systems, 321--328(2004)
11. Belkin M, Niyogi P. Sindwani V. On manifold regularization. In:Proceedings of 10th International Workshop on Artificial Intelligence and Statistics,17--24(2005)
12. A Blum and S. Chawla, learning from labeled and unlabeled data using graph mincuts, proc.18th international Conf. on Machine Learning, 19--26(2001)
13. D. Zhou, J,Weston, A. Gretto, O, Bousquet, and B.Scholkpf, Ranking on data manifolds. Proc of NIPS'03,(2003)
14. Wang, Changshui Zhang. Label propagation through linear neighborhoods. IEEE Transactions on Knowledge and Data Mining, 55--67 (2008)
15. D. Zhou, J,Weston, A. Gretto, O, Bousquet, and B.Scholkpf, Ranking on data manifolds. Proc of NIPS'03,(2003)