International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 04 | April-2016 www.irjet.net p-ISSN: 2395-0072

Dynamic Recovery over Large Scale Graph-Structured Data with Subgraphs

Yamini Meduru
M.E.(CSE) Student
Department of Computer Science, VELS University, Pallavaram, Chennai

U.V.Anbazhagu
Assistant Professor
Department of Computer Science, VELS University, Pallavaram, Chennai

---

Abstract: In this paper, we propose a novel Part-level Regularized Semi-Nonnegative coding (PRSN) approach to construct a discriminative graph benefiting from the part-level graph regularizer. Specifically, with the low-rank decomposition via SNMF, our method can well uncover the global structure of the multiple subspace of the data. Meanwhile, it preserves the local intrinsic information by virtue of part-level similarity measurement. Finally, with the iteratively optimized data representation matrix *Z*, the label information can be effectively propagated to the remaining unlabeled data.

Introduction: The main objective of this paper, a novel graph construction approach for graph based learning, including data clustering and semi-supervised classification. By taking advantages of low-rank coding and sparsification constraints, we jointly learned symmetric and sparse graphs. Graphs have been widely applied in modeling the relationships and structures in real-world applications. Graph construction is the most critical part in these models, while how to construct an effective graph is still an open problem. In this paper, we propose a novel approach to graph construction based on two observations. First, by virtue of recent advances in low-rank subspace recovery, the similarity between every two samples evaluated in the low-rank code space is more robust than that in the sample space. Second, a sparse and balanced graph can greatly increase the performance of learning tasks, such as label propagation in graph based semi-supervised learning. The k-NN sparsification can provide fast solutions to constructing unbalanced sparse graphs, and b-matching constraint is a necessary route for generating balanced graphs. These observations motivate us to jointly learn the low-rank codes and balanced (or unbalanced) graph simultaneously. In particular, two non-convex models are built by incorporating k-NN constraint and b-matching constraint into the low-rank representation model, respectively. We design a majorization-minimization augmented Lagrange multiplier (MM-ALM) algorithm to solve the proposed model.

Existing System: In unsupervised and semi-supervised learning, the algorithms usually show effective performance on data that obey the smoothness, cluster or manifold assumptions.

Methodologies: Databases and settings, Problem formulation & optimization, Spectral Clustering with Graph,Semi-Supervised Classification with Graph.

Databases and settings: In this module, we randomly generate dataset and process with these data. we randomly select 50 datasets of every class in the A and B tables, 100 datasets from each class in the C database, and use all the datasets of ORL database.

Semi-Supervised Classification with Graph: We first normalize all the images to be unit-norm as shown in Algorithm 2. All methods are repeated 10 times, and each time we randomly select a subset of images for each individual to create a labeled sample set. We can observe that our two graphs obtain better performance than other graphs. Even though Ours-I graph is unbalanced, it performs better than our previous work LRCB graph that is balanced. The reason is that Ours-I, as well as Ours-II, reformulates the rank-minimization model to obtain a new similarity metric, which is the key during graph construction.

Data Collection with Part-level regularized graph: The graph regularizer helps generate a more sparse and discriminative coefficient matrix Z [21], [28]. As we know, the similar data points should also have similar coefficients Z, so the graph regularizer is designed to transfer such local intrinsic structure via a similarity matrix S. The common way to measure the similarity is via various distance metrics based on Z. However, the common way is NOT the best choice in this case. We consider part-level representation coefficient H can uncover much richer information from underlying sample structure. In order to differentiate the similarity matrix learned over part-level representation H from the conventional one, we denote the novel similarity as SP, and its corresponding Lapidarian matrix as LP.

Optimization and Complexity Analysis: In this section, we discuss the time complexity of our model. For simplicity, assume X is a $n \times n$ matrix. The time-consuming components of Algorithm 1 are the semi nonnegative matrix factorization operation in Step 1, and matrix multiplication and inverse operations in Step 2. As discussed in, the computation complexity for SNMF is $m(n2p + np2)$ for updating variable $W \in Rn\times p$, where m is the number of iterations to convergence (we set $m = 10$ in our experiment) and p is the dimension of W (we set it as the rank of Z) . The computation complexity of updating variable $H \in Rp\times n$ for SNMF is $3mn2p$. Thus, the total complexity for Step 1 is of order $(4mn2p+mnp2)$. Generally, each matrix multiplication and inverse operation take $O(n3)$. Thus, the order of time cost for Step 2 is $O(n3)$. To sum up, the time complexity of our method is $O(n3+4mn2p+mnp2)$.

Algorithm:

Input: data matrix X, parameter $\lambda 1$, $\lambda 2$, k

Initialize: $Z0 = E0 = W0 = H0 = LP0 = 0$,

$Y1,0 = Y2,0 = 0$, $\mu 0 = 10-6$, $\mu max = 106\_ = 10-4$, $\rho = 1.3$, $t = 0$.

while not converged do

1. Fix the others and update $Wt+1,Ht+1$ using Eq. (6)
2. Fix the others and update $Zt+1$ using Eq. (7)
3. Fix the others and update $LPt+1$ using k-NN.
4. Fix the others and update $Et+1$ using Eq. (8).
5. Update the multiplier $Y1,t+1$, $Y2,t+1$ by

$Y1,t+1 = Y1,t + \mu(X - XZt+1 - Et+1)$,

$Y2,t+1 = Y2,t + \mu(\mathbf{1}TnZt+1 - \mathbf{1}Tn)$,

6. Update the parameter μ by $\mu = \min(\rho\mu, \mu max)$
7. Check the convergence condition by

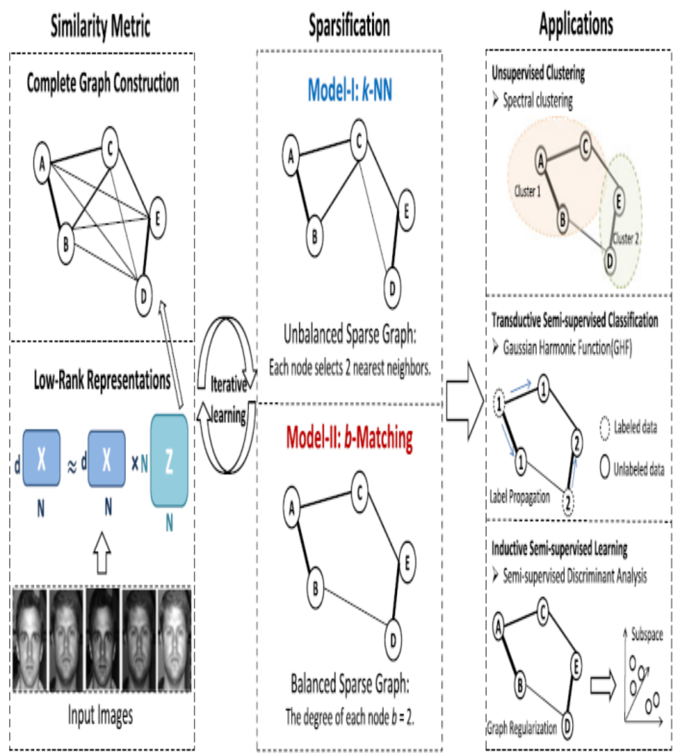
$\|X - XZ^{t+1} - E^{t+1}\|_\infty < \epsilon, \|\mathbf{1}^T_n Z^{t+1} - \mathbf{1}^T_n\|_\infty < \epsilon$.

8. $t = t + 1$.

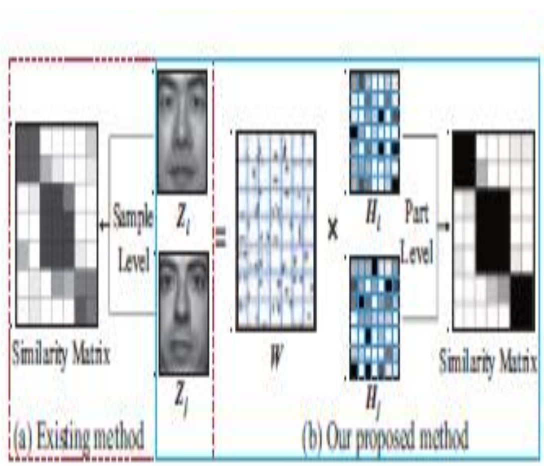
end while

Output: Z, W, H, E, LP

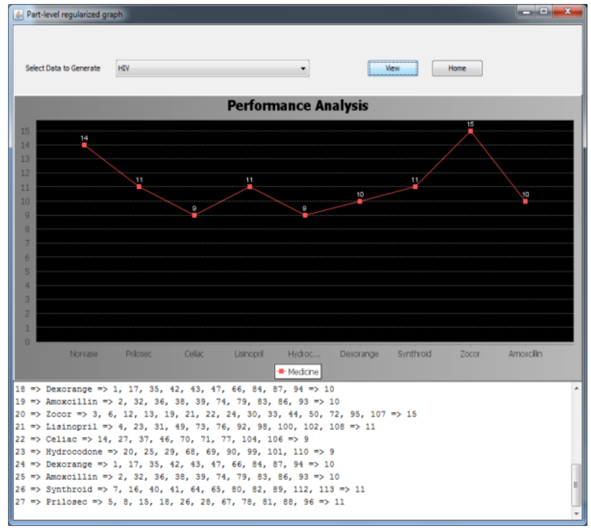
Architecture:

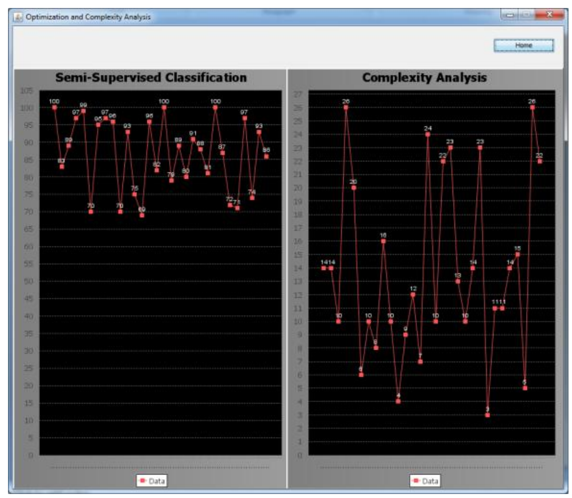


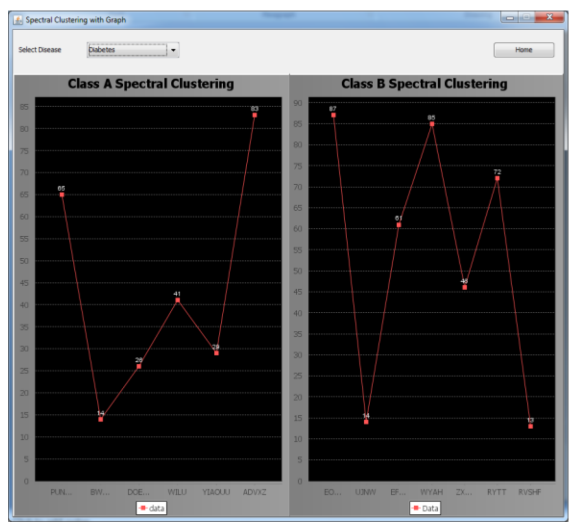
Existing method and proposed method:



Example:







Conclusion: In this paper, we have proposed a novel graph construction approach for graph based learning, including data clustering and semi-supervised classification. By taking advantages of low-rank coding and sparsification constraints (i.e., k-NN and b-matching), we jointly learned symmetric and sparse graphs. We also designed novel optimisation algorithms to solve the proposed models.

References: **1.** L. Zhuang, H. Gao, Z. Lin, Y. Ma, X. Zhang, and N. Yu, "Non-negative low rank and sparse graph for semi supervised learning," in CVPR, 2012.

2. J. Feng, Z. Lin, H. Xu, and S. Yan "Robust subspace segmentation with block-diagonal prior," in *CVPR,* 2014.

3. M. Zhao, L. Jiao, J. Feng, and T. Liu "A simplified low rank and sparse graph for semi-supervised learning," 2014.