DSL: Discriminative Subgraph Learning via Sparse Self-Representation

Lin Zhang, Petko Bogdanov

Department Of Computer Science, University at Albany-SUNY

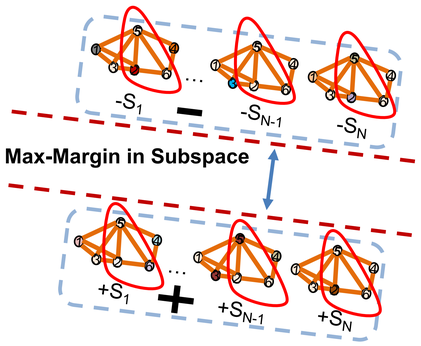
UNIVERSITY AT ALBANY
State University of New York

2019 SIAM International Conference on DATA MINING

DMG LAB

Problem Setting: Network Data Classification and Feature Selection

- Features are **nodes** in a fixed network
- **Network instance** comprised of values of nodes
- Global **instance label** represents the state of the instance
- **Goal**: Learn a discriminative connected subgraph subspace
- Applications:
  - Neuroscience: brain subnetworks which differentiate stages of a disease (e.g. Alzheimer's)
  - Gene expression: pathways in protein-protein interaction (PPI) networks altered by a disease
  - Road transportation: Subnetworks behaving most differently on weekends versus weekdays

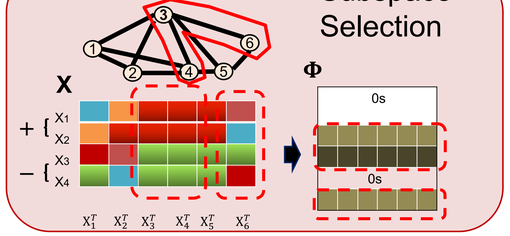


Model: Connectivity, discriminative power and sparsity

Objective:

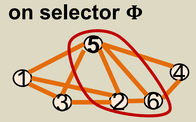
$$\underset{\mathbf{\Phi},\mathbf{w},b}{\operatorname{argmin}} \left\| \mathbf{X}^T - \mathbf{X}^T\mathbf{\Phi} \right\|_F^2 + \lambda_1 \left\| \mathbf{\Phi} \right\|_{2,1} + \lambda_2 \operatorname{tr}\left(\mathbf{\Phi}^T \mathbf{L} \mathbf{\Phi}\right) + \pi \left\{ \left\| \mathbf{w} \right\|_f + C \sum_i^n \ell\left(y_i, \mathbf{w}^T \hat{\mathbf{x}}_i + b\right) \right\}$$

Subspace Selection



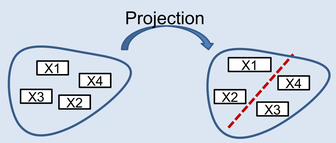
Subgraph Connectivity

Laplacian regularization on selector Φ



Discriminative power

Max-Margin in subspace



Contributions

- **Unified framework for network data classification and feature selection**
  - sparsity : predict with small-size subgraphs
  - connectivity : features form connected subgraphs, interpretable
  - discriminative power
- **High quality**: both on subgraph detection and instance classification

Solution Outline

- Alternating Optimization scheme to solve the DSL model
  - Split into two subproblems
  - Alternate between optimizing them in turn
- Split:
  1: Discriminative learning
  2: Connected subspace selection

① $\underset{\mathbf{w},b}{\operatorname{argmin}} \left\| \mathbf{w} \right\|_f + C \sum_i^n \ell\left(y_i, \mathbf{w}^T \hat{\mathbf{x}}_i + b\right)$

② $\underset{\mathbf{\Phi}}{\operatorname{argmin}} \left\| \mathbf{X}^T - \mathbf{X}^T\mathbf{\Phi} \right\|_F^2 + \lambda_1 \left\| \mathbf{\Phi} \right\|_{2,1} + \lambda_2 \operatorname{tr}\left(\mathbf{\Phi}^T \mathbf{L} \mathbf{\Phi}\right) + C^* \sum_i^n \ell\left(y_i, \mathbf{w}^T \mathbf{\Phi}^T \mathbf{x}_i + b\right)$

Algorithm 1 DSL Optimization

Input: Training data $(\mathbf{X}, y)$, and parameters $(\lambda_1, \lambda_2, C, \pi)$.
Output: The subgraph selection matrix $\mathbf{\Phi}$ and a classifier $(\mathbf{w}, b)$

1: Initialize: $\mathbf{\Phi} \leftarrow \mathbf{I}$
2: **while** $(\mathbf{w}, b)$ and $\mathbf{\Phi}$ have not converged **do**
3: $(\mathbf{w}, b) \leftarrow \operatorname{argmin}_{\mathbf{w},b} \|\mathbf{w}\|_f + C\sum_i^n \ell\left(y_i, \mathbf{w}^T \hat{\mathbf{x}}_i + b\right)$
4: **while** $\mathbf{\Phi}$ has not converged **do**
5: $\boldsymbol{\alpha} \leftarrow \operatorname{argmax}_{\boldsymbol{\alpha}} \frac{1}{2}\boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} + \mathbf{q}\boldsymbol{\alpha}, \text{s.t} 0 \le \boldsymbol{\alpha} \le C^* \mathbf{1}$
6: $D_{ii} \leftarrow (2\|\mathbf{\Phi}_i\|_2)^{-1}$ if $\|\mathbf{\Phi}_i\|_2 \neq 0$ or $D_{ii} \leftarrow 0$
7: $\mathbf{\Phi} \leftarrow \frac{(\mathbf{X}\mathbf{X}^T + \lambda_1 \mathbf{D} + \lambda_2 \mathbf{L})^{-1}}{2} (\sum_i \alpha_i y_i \mathbf{x}_i \mathbf{w}^T + 2\mathbf{X}\mathbf{X}^T)$
8: **return** $\{\mathbf{\Phi}, (\mathbf{w}, \mathbf{b})\}$

Discriminative learning

Connectivity, Subspace learning

bottleneck

O(m³)

Only D changes in iterations

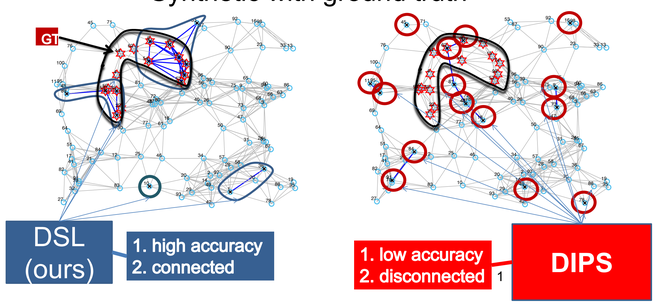
Employ Sherman-Morrison: low-rank inversion

$O(ms_X)$ s_X: umber of nonzero elements.

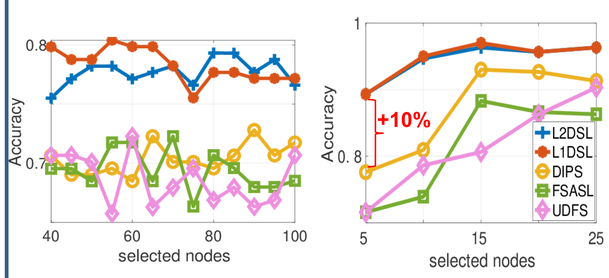
Results from Experimental Evaluation

- **Multiple datasets:** PPI, Transportation, Communications, Brain
- **Metrics**
  - Instance classification accuracy
  - Selection of known relevant subgraphs (GT) when available
- **Comparisons**: DSL(L1/L2), DIPS state-of-the-art for network data, other feature selectors (FSASL,UDFS)

Synthetic with ground truth

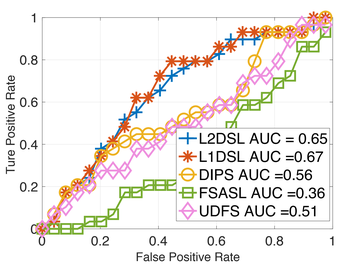


Instance Classification
Increasing number of features (subgraph)



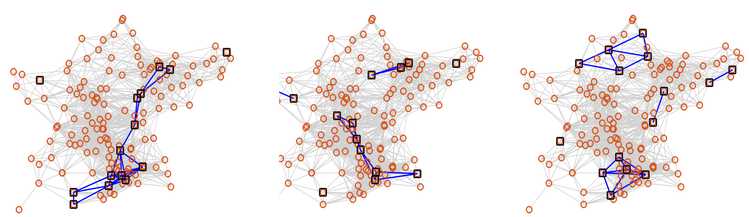
CCT Bike

Ground Truth Subgraph Detection
Liver Metastasis



Unbalanced (Random << 0.5)

Real world data – bike rentals



L1DSL (ours) DIPS UDFS

Sponsor: NSF-SCC