Semi-Supervised Clustering Based on Signed Total Variation

Peter Berger, Thomas Dittrich, and Gerald Matz

Institute of Telecommunications, Technische Universität Wien, Austria
<firstname>@nt.tuwien.ac.at

Introduction

Motivation
- Problem: semi-supervised clustering, i.e., splitting a dataset into disjoint classes under the assumption that the cluster affiliation is known for certain data points
- Assumption: nodes within a cluster are similar and nodes from different clusters are dissimilar
- Example social network: similarity links ↔ follower/friends dissimilarity links ↔ blocking or quoting behavior
- Question: how can dissimilarity information be incorporated into total variation based clustering

Contributions
- Introduce the signed total variation
- Formulate semi-supervised two-class clustering with dissimilarity based on the signed total variation
- Introduce a suitable $\ell_1$ regularization to ensure reliable clustering even when only few labels are known
- Develop a low-complexity ADMM-based algorithm

Signed Clustering

Signed Laplacian
- Negative weight $W_{ij} < 0$ models dissimilarity between $i$ and $j$
- Signed graph Laplacian: $L = D - W$ with the signed degree matrix $D = \text{diag}(d_1, \ldots, d_N)$, $d_i = \sum_j W_{ij}$
- Induced Laplacian form:
  $$x^T L x = \sum_{i,j} \left| x_i - \text{sign}(W_{ij}) x_j \right|^2 |W_{ij}|$$

Signed total variation
- This motivates the new concept of the signed total variation:
  $$\|x\|_{TV} \defeq \sum_{i,j} |x_i - \text{sign}(W_{ij}) x_j| |W_{ij}|$$
- The signed total variation $\|x\|_{TV}$ is a semi-norm and convex
- For unbalanced graphs (contains a cycle with an odd number of edges with negative weight) it is a norm

Regularity
- Problem 1: total variation minimization tends to declare (one of the) the label sets $L^+, L^-$ as clusters
- Problem 2: the signed total variation tends to assign zero values since both $|x_i + x_j|$ and $|x_i - x_j|$ can be minimized by setting $x_i = x_j = 0$
- Regularized signed total variation clustering problem
  $$\min_x \left( \|x\|_{TV} + \lambda \sum_{i \in \mathcal{L}^+} |x_i + x_j| + \mu \sum_{i \in \mathcal{L}^-} |x_i - x_j| \right)$$
  s.t. $x_i = -1$ for $i \in \mathcal{L}^-$, $x_i = 1$ for $i \in \mathcal{L}^+$
  (3)

Constrained total variation minimization
- Consider unsigned weight matrix $W$, $W_{ij} \geq 0$
- A positive weight $W_{ij} > 0$ models similarity between $i$ and $j$
- Min-cut approach determines $V^+$ and $V^- = V \setminus V^+$ via
  $$\min_{V} \sum_{i \in V} \sum_{j \notin V} W_{ij}$$
  s.t. $\mathcal{L} \subseteq V^+$, $\mathcal{L} \subseteq V^-$
  (1)

Total variation based unsigned clustering
- Consider unsigned weight matrix $W$, $W_{ij} \geq 0$
- A positive weight $W_{ij} > 0$ models similarity between $i$ and $j$
- Min-cut approach determines $V^+$ and $V^- = V \setminus V^+$ via
  $$\min_{x} \sum_{i \in V} \sum_{j \notin V} |x_i - x_j| W_{ij}$$
  s.t. $x_i = -1$ for $i \in \mathcal{L}^-$, $x_i = 1$ for $i \in \mathcal{L}^+$
  (2)

If the min-cut problem (1) has a unique solution ($V^+, V^-)$, then (2) yields the equivalent solution
$$x_i = \begin{cases} -1, & i \in V^-, \\ 1, & i \in V^+ \end{cases}$$

Algorithm 1: Signed TV clustering with parameter tuning
Input: $W$, $L^+, L^-$, $x_{\text{true}}$ (slightly smaller than 1)
Initialize: $\lambda = 0$, $\lambda^* = 0$
1. repeat
2. calculate minimizer $x$ of (3)
3. $M^- = \{ i \in N^- : x_i < 0 \}$
4. $M^+ = \{ i \in N^+ : x_i > 0 \}$
5. $x^* = \min_{i \in M^-} |x_i|$
6. $x^- = \min_{i \in M^-} |x_i|$
7. $a = 0$
8. if $M^- = \emptyset$ or $x^* < x_{\text{true}}$ then
9. increase $\lambda^*$, $a = 1$
10. end if
11. if $M^+ = \emptyset$ or $x^* < x_{\text{true}}$ then
12. increase $\lambda^*$, $a = 1$
13. end if
14. until $a = 0$
Output: $x$

Simulations

Setup
- Simulations on two-moon datasets with $N = 500$ nodes
- Coordinates of each node generated from a random angle on a center curve and Gaussian jitter (variance $\sigma^2 = 0.03$)
- Graph generated as kNN graph with $k = 10$ neighbors and Gaussian kernel for edge weights (parameter $\sigma = 0.6$)
- $M$ samples drawn randomly while ensuring at least one known label from each cluster
- $L$ randomly chosen dissimilarity edges between pairs of nodes from different clusters

Illustrative example
- Different colors represent different clusters
- Sampled nodes represented by dark colors
- Dissimilarity edges represented by dashed lines

Ground truth

Monte Carlo simulations

Error rates in percent (mean and standard deviation)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$M = 2$</th>
<th>$M = 5$</th>
<th>$M = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L = 0$</td>
<td>7.3 ± 12.5</td>
<td>4.0 ± 9.0</td>
<td>1.8 ± 5.1</td>
</tr>
<tr>
<td>$L = 5$</td>
<td>3.0 ± 9.1</td>
<td>1.2 ± 3.5</td>
<td>1.0 ± 2.4</td>
</tr>
<tr>
<td>$L = 10$</td>
<td>1.4 ± 6.0</td>
<td>0.9 ± 1.9</td>
<td>0.7 ± 0.7</td>
</tr>
<tr>
<td>$L = 15$</td>
<td>0.5 ± 5.4</td>
<td>0.36 ± 3.5</td>
<td>2.5 ± 2.1</td>
</tr>
</tbody>
</table>

Discussion
- Incorporation of dissimilarity substantially improves performance
- Total variation is directly connected to a minimum cut and therefore outperforms Laplacian based algorithms
- Most state of the art algorithms have free parameters
- Proposed algorithm has no free parameters

Funded by WWTF Grant ICT15-119