Semi-Supervised Learning on Data Streams via Temporal Label Propagation

**Introduction**

Problem:
- Data points arrive on stream
- Few are labeled (0, 1)
- Most are unlabeled
- Task: Label points on-the-fly

Examples:
- Medical signal monitoring
- Network monitoring
- Object detection on video feed

Challenges:
- Time: Label points quickly
- Memory: Stream too large to fully store
- Semi-supervised learning: Learn from labeled and unlabeled data

**Background**

Label Propagation:
- Offline algorithm for semi-supervised learning
- (Zhu, Ghahramani, Lafferty, ICML 2003)
- ICML 10-Year Classic Paper Award, 2013

Algorithm:
- Build graph on data points
- Propagate labels by minimizing electric energy

**Graph Compression**

Intermediate goal: Compress irrelevant nodes in graph
- “Irrelevant” means we do not need to classify them anymore
- E.g., old points on the stream
- Problem: Relevant labels still depend globally on irrelevant nodes
- Solution: Encode irrelevant nodes into edge weights

**Star-Mesh Transform**

- Known in electric network theory (Campbell 1922, Rosen 1924)
- Removes one irrelevant node
- Updates edge weights to encode it

**Sequential Star-Mesh Transforms**

- Input: Large graph with $n$ nodes, only $\tau < n$ are relevant
- Compress irrelevant nodes one by one by star-mesh transforms:
  - Compression:
    - Size before: $\Omega(n \log n)$ bits
    - Size after: $O(\tau^2 \log n)$ bits

**Theorem:** Label Propagation is preserved on the relevant nodes

- This works for any graph.
- However, sequential star-mesh transforms are as slow as Label Propagation on the entire original graph, so we get no improvement for offline data.
- We will use it for streaming data.

**Streaming Algorithm**

Temporal Label Propagation:
- Pick small integer parameter $\tau$
- Store the most recent $\tau$ unlabeled points on the stream
- Upon new unlabeled point arrival:
  - Compress oldest unlabeled point by star-mesh transform
  - Insert new point
  - Propagate labels

Guarantees for step $n$:
- Running time: $O(\tau^3)$
- Memory: $O(\tau^2 \log n)$
- Output: Label Propagation on a “temporal vicinity graph” that contains the entire stream seen so far:

**Experiments**

Toy data: Streaming two rings

Real data: Classifying irregular heartbeats on EKG signal
- Task: Classify atrial vs. ventricular premature contractions
  - Accuracy: 94.9% with $\tau = 5$ and 2 labeled examples per class.
  - Data source: physionet.org. Normal heartbeats were ignored.