EXPERIMENTS

Minsu Cho and Kyoung Mu Lee
Department of EECS, ASRI, Seoul National University, Seoul, Korea

INTRODUCTION

Clustering Problem

To partition a set of data into disjoint subsets based on the underlying structure of the data.

How to achieve robust clustering with a hierarchy of data at multiple scales?

MOTIVATION & IDEA

Clustering by Mode-Seeking on data

- Mode-seeking by iterative mean-update by local kernel
- Local shift without global data information
- Authority score of a node by Personalized PageRank (PPR)

Authority-shift clustering (Proposed)

How to make better shifts for data clustering?

Link analysis techniques for WWW & social networks:

- HITS (Kleinberg, 1998)
- PageRank (Page & Brin, 1998)

How to achieve

Clustering by Mode-Seeking on data

Global authority score by PageRank (Page & Brin, 1998)

A Random Walker follows the structure of the graph by the transition matrix $P$ and sometimes randomly jumps to a node by the probability distribution $v^T$.

The authority score by the steady-state distribution $v$.

Authority score of a node by Personalized PageRank (PPR)

Jumps to node $i$ of data hierarchy

Seek authority nodes with the high scores, and shift to it!

PROPOSED METHOD

High-order Authority

- Higher-order steady-state, personalized to PPR$(i)$

\[ \text{PPR}_{n+1}(i) = \text{PPR}(n)(i) \]

Random walkers starting again from the steady state of the previous steady state.

(a) Node 1 (b) Node 2 (c) Node 3 (d) Node 4 (e) Node 5 (f) Global PageRank

Higher-order PPRs propagate authorities further by linearity of PPRs, its computation reduces to a matrix product!

What is a Good Cluster?

- nth order auth node for node $i$: a node with the highest score

\[ \text{Auth}_n(i) = \arg \max_{s \in V} \text{PPR}_n(i, s) \]

Self-authoritative cluster: the set of nodes, each of which has its authority node as a member of the set.

\[ C = \{ i \in V | \text{Auth}_n(i) \in C, \forall i \in C \} \]

It is simply obtained by shifting authority nodes, finally to authority sinks.

Cluster by authority-seeking on Graphs

Hierarchical Authority-Shift

Aggregate nodes in each cluster into a supernode after every shift.

\[ \text{PPR}^{[n+1]}(a, b) \leftarrow \frac{1}{\sigma_{n+1}(a)} \sum_{a \in C} \sigma_{n}(i) \text{PPR}^{(n)}(i, a) \]

\[ \sigma_{n+1}(a) \leftarrow \sum_{a \in C} \sigma_{n}(i) \]

Reduced graph structures at higher-layer of a hierarchy

- The num of nodes reduces rapidly.
- The PPR order for the shift starts to soar with the emergence of reliable clustering.

- It improves clustering quality and convergence speed by integrating the temporal smoothness constraint that no higher-order clustering result deviates from the lower-order ones.

CONCLUSION

- Stable and robust hierarchical clustering based on link analysis
- Balance between bottom-up similarity and top-down global structure
- Authority sink as the most important node for each cluster
- Applicable to directed graphs

EXPERIMENTS

Point set clustering

- Random data by Gaussian & Uniform distributions
- Normalized to $[-1,1]$
- Graph by weight

\[ W_{ij} = \exp \left( \frac{-d(i, j)^2}{\sigma^2} \right) \]

- Comparative examples

Comparative examples

average accuracy over 100 point set generation for each pattern

Hierarchical Image Segmentation

- Conventional 8-NN graph of pixel nodes with edges weighted by RGB dist.
- Test on Berkeley image database, initial seeding technique is used.

Comparison on Berkeley DB

- Average accuracy

- Special Problem

- Higher layers for 6th pattern