A Short Survey of Recent Advances in Graph and Subgraph Matching

Shuyue Jia
July 2020
Learning Objectives

• **GM**: Construct *vertex corresponding*, which maximize *affinities*  
  • Object Function ← Lawler’s QAP  
  • Optimization Problem ← Maximize Affinity Score  

• **SubGM**: Find *all subgraphs* in the database graph  
  • Making *Index* ← Dataset Preprocessing  
  • Query processing ← GM  

• **Current Hotspots**:  
  • Multi-graph Matching  
  • Incremental GM  
  • Deep Learning based GM/SubGM  

• **Applications**:  
  • Effective and Efficient SubGM for bio networks  
  • Online Incremental Learning
Background:

1. Subgraph Matching for (large) biological networks/graphs
2. Maximum Common Subgraphs (MCS)

Survey:

1. Most *(inexact)* Graph Matching
2. Few *Subgraph Matching*
Graph Matching

Definitions

- Model vertex corresponding to maximize object function (e.g., affinity score, distance cost)
- Properties: isometric (等距), isomorphisms (同构), homeomorphisms (同胚)
- Two Steps:
  1. Build vertex corresponding (Assignment Matrix) and object function
  2. Optimize the Assignment Matrix to maximize/minimize object function

Measurements

- Which metrics we aim to optimize → Assignment Matrix
- How we define our object function → Lawler’s QAP (Most commonly used affinity score)

Keywords

- Graph Matching, Subgraph Matching,
- Multi-graph Matching, Incremental Matching, Higher-order Graph Matching
Usage Scenarios

**Graphs/Subgraphs Matching to (Multiple/Large) Domain Database Graphs**

- Subgraphs $\rightarrow$ (Large, $>10^3$ vertexes) domain Knowledge Graphs
- **Physical Science** (Chemistry, Fluid Dynamics, Astronomy, structural mechanics, and ecosystem modeling)
- **Life Science** (Biological Networks, protein-protein interaction networks, gene regulatory networks, metabolic networks, brain connectivity from alignment-free functional magnetic resonance imaging (fMRI) data)

**Computer Science Scenarios $\leftrightarrow$ A lot of scenarios**

- **Combinatorics** $\leftarrow$ Quadratic Assignment Problem (QAP) $\leftarrow$ Graph Matching
- **Pattern Recognition, Multimedia, Computer Vision**
- Image registration, understanding, extrapolation and recognition, object recognition and tracking, scene understanding and parsing, weak-perspective 3-D reconstruction, action recognition, robotics, video surveillance and person re-identification.
Current related works

Inexact Graph Matching

Object Function

Affinity Matrix based Model

Second-order

Lawler’s QAP
Two graphs’ correspondence (Gaussian Kernel)

Königsmann-Beckman’s QAP
Two graphs’ correspondence + Graphs’ separate Adjacency Matrix

Higher-order

Factored
Separate Node and Edge Affinity SOTA models commonly used

with Graphs’ Higher-order Information

Bipartite GED
(Linear Assignment)

Graph Edit Distance (GED)

Hungarian method
(Node-to-node Cost)

Graphs only first-order information

QAP-based GED
(Quadratic Assignment)

QAP Optimization

Optimization Algorithms

Tree Search with backtracking

Partial matching + heuristic estimation

Continuous Relaxation Methods

Spectrum Relaxation
Efficient but low accuracy

Faouly Stochastic Relaxation
Trade-off

Semi-definite Optimization
Higher Computation Complexity

Spectral Methods
Use of use of eigenvalues/eigenvectors

Decomposition Methods

Neural Networks
Getting popular

Genetic Algorithms

Bipartite Matching Methods

Local Properties Methods

Path-following Paradigm

Discrete Methods

Other Techniques
Current related works

Inexact Matching

Tree Search

Tsai79[154], Graharaman80[59], Shapiro81[136], Sanfeliu83[132], Tsai83[155], 
Eshera84[44, 45], Shapiro85[137], Wong90[172], Dumay92[40], Rocha94[128], 
Shasha94[138], Wang95[164], Cordella96[30], Allen97[2], Cordella97[31], 
Oflazer97[113], Haris99[65], Serratos99[135], Berretti00[9, 10], Serratos00[134], 
Berretti01[11], Llados01[95], Valiente01[158], Gregory02[63], Fernández01[48]

Continuous Optimization

Fisher73[51], Kittler89[80], Almohamad93[3], Christamas95[27], Pelillo95[119], 
Pelillo95[120], Gold96[61], Rangaranjan96[127], Bomze97[12], Wilson97[167], 
Pelillo98[121], Branca99[13], Huet99[71], Medasani99[105], Pelillo99[122, 123], 
Myers00[112], Luo01[98], Medasani01[104], Torsello01[151], Pelillo02[124], 
Van Wyk02[159, 160]

Spectral Methods

Umeyama88[157], Carcassoni01[21], Kosinov01[83], Xu01[173], 
Shokoufandeh01[144]

Kitchen79[79], Gendreau93[58], Wang94[163], Liu95[93], Depiero96[37], 
Shokry96[145], Wang97[166], El-Sonbaty98[42], Messmer98[108], 
Sugathan98[149], Fuchs99[57], Ozer99[115], Perchant99[125], 
Williams99[168], Baeza-Yates00[5], Fuchs00[56], Jagota00[74], Liu00[92], 
Sugathan00[146], De Mauro01[34], Khoo01[78], Hlaoui02[66]

Other Techniques
Current related works

- **Subgraph Matching (Search Problem)**
  - Frequent Discriminate Substructure
    - gindex
    - TreePi
    - FG-index
    - Tree+
  - Path, vertex, and Neighborhood Substructure
    - GraphGrep
    - All Paths
    - Closure-Tree
    - Pseudo subgraph isomorphism
    - TALE
    - Neighborhood Units
  - Graph Alignment
    - Not applicable when subgraph is small
### Summary

<table>
<thead>
<tr>
<th>Year</th>
<th>Author</th>
<th>Institution</th>
<th>Title</th>
<th>Scenario</th>
<th>Metrics</th>
<th>Model</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>2016</td>
<td>Junchi Yan (严骏驰)</td>
<td>上海交大</td>
<td>A Short Survey of Recent Advances in Graph Matching</td>
<td>Multiple</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2018</td>
<td>Junchi Yan (严骏驰)</td>
<td>上海交大</td>
<td>(中文期刊) 计算机视觉中图匹配研究进展: 从二图匹配迈向多图匹配 (博士论文) 图匹配问题的研究和算法设计</td>
<td>Multiple</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2015</td>
<td>Donatello Conte, Pasquale Foggia, and et al.</td>
<td>萨勒诺大学</td>
<td>Thirty Years Of Graph Matching In Pattern Recognition</td>
<td>Multiple</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2004</td>
<td>Andrei Zanfir and Cristian Sminchisescu</td>
<td>隆德大学</td>
<td>Deep Learning of Graph Matching</td>
<td>Multiple, Computer Vision in particular</td>
<td>PCK: Percentage of Correct Keypoints PCK@10 pixels, PCK@0.05, PCK@0.1</td>
<td>Neural Network (CNNs)</td>
<td>MPI-Sintel: PCK@10 pixels: 92.6% CUB: PCK@0.05: 0.86 PASCAL VOC keypoints: PCK@0.1 (class average): 40.6</td>
</tr>
<tr>
<td>2009</td>
<td>Shijie Zhang, Shirong Li, and Jiong Yang</td>
<td>凯斯西储大学</td>
<td>GADDI: Distance Index based Subgraph Matching in Biological Networks</td>
<td>Biological Networks</td>
<td>Accuracy, Precision, Recall, Query Time</td>
<td>Distance Index based method</td>
<td>-</td>
</tr>
</tbody>
</table>
To do

- Continue to survey *subgraph matching*, especially those methods for biological networks.
- Try to reproduce “Deep Learning of Graph Matching” and apply the algorithm into biological networks (graph matching).
- See whether the method is suitable for *subgraph matching*.
- Get engaged with knowledge graph and corresponding tasks.
Thanks and have a nice day!