Learning Graph Distances with Message Passing Neural Networks

Pau Riba, Andreas Fischer*, Josep Lladós, Alicia Fornés

Computer Vision Center, *Université de Fribourg

ICPR, Beijing, August 23rd, 2018.
Outline

Introduction

Related Concepts

Architecture

Experimental Validation
  Datasets
  Classification
  Retrieval

Conclusion and Future Work
Introduction
Motivation

Graph representations

1. Increasing relevance in **visual object recognition and retrieval**, beyond classical pure appearance-based approaches.
Motivation
Graph representations

1. Increasing relevance in visual object recognition and retrieval, beyond classical pure appearance-based approaches.
Motivation
Graph representations

1. Increasing relevance in **visual object recognition and retrieval**, beyond classical pure appearance-based approaches.
2. Visual object detection using graphs involves an **inexact subgraph matching formulation**.
Motivation
Graph representations

1. Increasing relevance in **visual object recognition and retrieval**, beyond classical pure appearance-based approaches.
2. Visual object detection using graphs involves an **inexact subgraph matching formulation**.
1. Increasing relevance in **visual object recognition and retrieval**, beyond classical pure appearance-based approaches.
2. Visual object detection using graphs involves an **inexact subgraph matching formulation**.
3. It is unavoidable in **large scale retrieval** (i.e. subgraph matching).
A graph is a powerful representation, both for text and graphics.
Geometric Deep Learning

Extension of Deep Learning techniques to graph/manifold structured data.
Motivation
Geometric Deep Learning

Geometric Deep Learning
Extension of Deep Learning techniques to graph/manifold structured data.

Image:
- Regular grid
- Operations well defined
- Same size → batch processing
- 8-neighbourhood

Graph:
- 4-Tuple $G = (V, E, L_V, L_E)$
- Operations not efficient
- Different size → batch processing
- Different neighbourhood
Hypothesis

Local structural node information can be learned by Geometric Deep Learning and exploited by Graph Distance algorithms.

Thus, we avoid a graph embedding that may be difficult to learn.
Related Concepts
Graph Edit Distance

Definition

Given a set of Graph Edit Operations, the Graph Edit Distance (GED) between two graphs $g_1$ and $g_2$ is defined as

$$
\text{GED}(g_1, g_2) = \min_{(e_1, \ldots, e_k) \in P(g_1, g_2)} \sum_{i=1}^{k} c(e_i)
$$

where $P(g_1, g_2)$ denotes the set of edit paths transforming $g_1$ into $g_2$ and $c(e)$ is the cost of each edit operation.
Graph Edit Distance
Approximated Techniques

Computation

Exact GED is not feasible in real applications due to its complexity. Several approximations have been proposed.

* Fischer et al., “Approximation of graph edit distance based on Hausdorff matching”.
† Riesen et al., “Approximate graph edit distance computation by means of bipartite graph matching”.
Computation

Exact GED is not feasible in real applications due to its complexity. Several approximations have been proposed.

Some approximated algorithms have been proposed.

- **Hausdorff Edit Distance (HED)** $\mathcal{O}(n_1 \cdot n_2)$
- **Bipartite Graph Matching (BP)** $\mathcal{O}((n_1 + n_2)^3)$

---

* Fischer et al., “Approximation of graph edit distance based on Hausdorff matching”.

† Riesen et al., “Approximate graph edit distance computation by means of bipartite graph matching”.

---

16 Learning Graph Distances
Pau Riba et al.
Computation

Exact GED is not feasible in real applications due to its complexity. Several approximations have been proposed.

Some approximated algorithms have been proposed.

- *Hausdorff Edit Distance (HED)* $O(n_1 \cdot n_2)$
- *Bipartite Graph Matching (BP)* $O((n_1 + n_2)^3)$

The usual *Graph Edit Operations* in the GED computation are:

- Insertion and Deletion (nodes and edges)
- Substitution (nodes and edges)

---

* Fischer et al., “Approximation of graph edit distance based on Hausdorff matching”.
† Riesen et al., “Approximate graph edit distance computation by means of bipartite graph matching”.
Geometric Deep Learning

Neural Message Passing*

Message Passing Neural Network (MPNN) is composed of 3 functions:

- Message
- Update
- Readout

* Gilmer et al., “Neural message passing for quantum chemistry”.

Learning Graph Distances

Pau Riba et al.
Geometric Deep Learning

Neural Message Passing*

Message

\[ m_{v}^{t+1} = \sum_{w \in \mathcal{N}(v)} M_t(h_v^t, h_w^t, e_{vw}) \]

* Gilmer et al., “Neural message passing for quantum chemistry”.

Learning Graph Distances
Pau Riba et al.
Geometric Deep Learning

Neural Message Passing*

Message

\[ m^{t+1}_v = \sum_{w \in \mathcal{N}(v)} M_t(h^t_v, h^t_w, e_{vw}) \]

Example:

\[ M_t(h^t_v, h^t_w, e_{vw}) = A(e_{vw})h^t_w \]

where \( A(\cdot) \) is a NN mapping to a \( d \times d \) matrix.

---

* Gilmer et al., “Neural message passing for quantum chemistry”.
Update

\[ h_{v}^{t+1} = U_{t}(h_{v}^{t}, m_{v}^{t+1}) \]

---

* Gilmer et al., “Neural message passing for quantum chemistry”. 
Geometric Deep Learning

Neural Message Passing

Update

\[ h_v^{t+1} = U_t(h_v^t, m_v^{t+1}) \]

Example:

\[ U_t(h_v^t, m_v^{t+1}) = GRU(h_v^t, m_v^{t+1}) \]

where \( GRU(\cdot, \cdot) \) is a Gated Recurrent Unit.

* Gilmer et al., “Neural message passing for quantum chemistry”.
Geometric Deep Learning

Neural Message Passing*

Readout

\[ \hat{y} = R(\{h_v^T | v \in G\}) \]

*Gilmer et al., “Neural message passing for quantum chemistry”.
**Geometric Deep Learning**

Neural Message Passing\(^*\)

**Readout**

\[
\hat{y} = R(\{h_v^T | v \in G\})
\]

**Example:**

\[
R(\{h_v^T | v \in G\}) = \sum_{v \in V} \sigma \left( i(h_v^{(T)}), h_v^0 \right) \odot \left( j(h_v^{(T)}) \right)
\]

where \(i\) and \(j\) are NN and \(\odot\) denotes element-wise multiplication.

\(^*\) Gilmer et al., “Neural message passing for quantum chemistry”.

---

**Learning Graph Distances**

Pau Riba *et al.*
Architecture
Siamese Architecture

\[ g_1 \]

\[ g_2 \]
Siamese Architecture

Learning Graph Distances
Pau Riba et al.
Siamese Architecture

Graph similarity $d(W(x_1), W(x_2))$

$D_W$

$G_W(g_1)$

$G_W(g_2)$

Update
Message

$G_W(X)$ Network branch 1

W shared (siamese)

Update
Message

Update
Message

Update
Message

$G_W(X)$ Network branch 2

$g_1$

$g_2$
Graph Similarity

- Hausdorff Distance-based Similarity

\[ H(A, B) = \max \left( \max_{a \in A} \inf_{b \in B} d(a, b), \max_{b \in B} \inf_{a \in A} d(a, b) \right) \]

- More robust distance

\[ \hat{H}(A, B) = \sum_{a \in A} \inf_{b \in B} d(a, b) + \sum_{b \in B} \inf_{a \in A} d(a, b) \]

- Proposed distance

\[ d(g_1, g_2) = \frac{\hat{H}(V_1, V_2)}{|V_1| + |V_2|} \]
Contrastive Loss

Given \( D_W = d(G_W(g_1), G_W(g_2)) \) where \( g_1 \) and \( g_2 \) are graphs and \( W \) a set of specific weights \( W \), the Loss Function is

\[
I(D_W) = \frac{1}{2} \begin{cases} 
D_W^2, & \text{if } Y = 1 \text{ (positive pair)} \\
\max(0, m - D_W)^2, & \text{if } Y = 0 \text{ (negative pair)}
\end{cases}
\]

where \( m = 1 \) is the adaptive margin.
Experimental Validation
Datasets

Letters
- Classification of Synthetic Graphs
- 15 classes
- 750 graphs per class
- 3 different distortion levels

George Washington
- Retrieval of Handwritten Words
- Several graph constructions
- 105 keywords
- 4894 instances
- HistoGraph subset for classification
Experimental Setup

Classification

- k-Nearest Neighbor Classifier
- Accuracy + Standard Deviation (5 runs)
- Tested with well-known Aproximated Graph Edit Distance algorithms
### Table: Accuracy ± Std for 5 runs.

<table>
<thead>
<tr>
<th></th>
<th>LOW</th>
<th>MED</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>BP*</td>
<td>99.73</td>
<td>94.27</td>
<td>89.87</td>
</tr>
<tr>
<td>HED†</td>
<td>97.87</td>
<td>86.93</td>
<td>79.2</td>
</tr>
<tr>
<td>Embedding‡</td>
<td>99.80</td>
<td>94.90</td>
<td>92.90</td>
</tr>
<tr>
<td>MPNN</td>
<td>95.04</td>
<td>83.20</td>
<td>72.27</td>
</tr>
<tr>
<td>Siamese MPNN</td>
<td>98.08</td>
<td>89.0136</td>
<td>74.77</td>
</tr>
<tr>
<td>Test BP</td>
<td>98.19</td>
<td>88.37</td>
<td>79.65</td>
</tr>
<tr>
<td>Test HED</td>
<td>98.00</td>
<td>89.79</td>
<td>77.07</td>
</tr>
</tbody>
</table>

* Riesen et al., “Approximate graph edit distance computation by means of bipartite graph matching”.

† Fischer et al., “Approximation of graph edit distance based on Hausdorff matching”.

‡ Gibert et al., “Graph embedding in vector spaces by node attribute statistics”.

---

Learning Graph Distances
Pau Riba et al.
## Table: Classification accuracy for the HistoGraph dataset.

<table>
<thead>
<tr>
<th>Subset</th>
<th>BP*</th>
<th>PSGE†</th>
<th>Siamese MPNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3-NN</td>
<td>5-NN</td>
<td></td>
</tr>
<tr>
<td>Keypoint</td>
<td>77.62</td>
<td>80.42</td>
<td>85.31</td>
</tr>
<tr>
<td></td>
<td>± 1.6552</td>
<td>± 0.5600</td>
<td>82.80</td>
</tr>
<tr>
<td>Projection</td>
<td>81.82</td>
<td>80.42</td>
<td>73.15</td>
</tr>
<tr>
<td></td>
<td>± 2.6014</td>
<td>± 1.5064</td>
<td>69.65</td>
</tr>
</tbody>
</table>

* Stauffer et al., “A Novel Graph Database for Handwritten Word Images”.

† Dutta et al., “Pyramidal Stochastic Graphlet Embedding for Document Pattern Classification”.
Experimental Setup

Retrieval

Mean Average Precision + Standard Deviation (5 runs)

\[ mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q}, \]

Learning Graph Distances
Pau Riba et al.
### Table: mAP from different approaches on GW dataset.

<table>
<thead>
<tr>
<th>Method</th>
<th>mAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHOC*</td>
<td>64.00</td>
</tr>
<tr>
<td>BOF HMM†</td>
<td>80.00</td>
</tr>
<tr>
<td>DTW</td>
<td></td>
</tr>
<tr>
<td>DTW’01</td>
<td>42.26</td>
</tr>
<tr>
<td>DTW’08</td>
<td>63.39</td>
</tr>
<tr>
<td>DTW’09</td>
<td>64.80</td>
</tr>
<tr>
<td>DTW’16</td>
<td>68.64</td>
</tr>
<tr>
<td>Mean Ensemble BP‡</td>
<td>69.16</td>
</tr>
<tr>
<td>Siamese MPNN</td>
<td>75.85±3.64</td>
</tr>
</tbody>
</table>

* Ghosh et al., “Query by string word spotting based on character bi-gram indexing”.
† Rothacker et al., “Segmentation-free query-by-string word spotting with bag-of-features HMMs”.
‡ Stauffer et al., “Ensembles for Graph-based Keyword Spotting in Historical Handwritten Documents”.
Conclusion and Future Work
Final thoughts

Conclusions
- Enriched graph representation, incorporating the local context
- Fast similarity measure based on the Hausdorff Distance
- It emphasises the structure
- Improvements in real applications

Future Work
- To explore uses of graph structures to model relations among several images (each image encoded as a node)
Thank you for your attention!

Pau Riba
Computer Vision Center
priba@cvc.uab.cat
www.cvc.uab.cat/people/priba