TouchUp-G: Improving Feature Representation through Graph-Centric Finetuning

Jing Zhu, Xiang Song, Vassilis Ioannidis, Danai Koutra, Christos Faloutsos

KnowledgeNLP – Aug.6 2023
Recommendation has been widely applied in online services: E-commerce, Content Sharing, Social Networking ...

Items can be: Products, News, Movies, Videos, Friends, etc.
Recommendation System as LP

One typical approach for solving this is to cast as a Link Prediction (LP) task for feature-rich graphs.

- Co-purchasing: Given item A, which item tend to be bought together with item A
- Features: Text descriptions, images of candidate items

Product Recommendation

Frequently bought together

A  +  B  +  C
Total price: $208.9
Add all three to Cart
Add all three to List

Image credit: Jiliang Tang

Jing Zhu et. al
Pretrained Models for GNNs

For feature-rich graphs: Pretrained models such as BERT is typically adopted to generate feature embeddings for each node.

Jing Zhu et. al
Why pretrained features may fail

A subgraph of Amazon Co-purchasing graph (Amazon-CP). Products have completely different visual features, but they are often bought together.

Jing Zhu et. al
Why pretrained features may fail

Pretrained Model (BERT, ViT etc.)

The graph agnostic feature embeddings prevents fully utilize the potential correlations between graph structures and node features.

Jing Zhu et. al
Research Question

How can we improve node features obtained from Pretrained Models (PMs) for downstream graph tasks such as link prediction?

Graph-Centric Finetuning on the PMs!

Jing Zhu et. al
TouchUp-G

- **General**: Can be applied to a variety of graph tasks.
- **Multi-modal**: Can be applied any pretrained models from any modality, e.g., texts, images etc.
- **Principled**: Propose a novel metric: feature homophily to measure the correlation between node features and graph structure.
- **Effective**: Outperforms baselines on 4 real datasets, with up to $2\times$ performance improvement across various tasks, metrics and modalities.
Previous Works

- Most work focuses on finetuning language models (LMs) improving node representation for node classification on text-rich graphs.

- TouchUp-G can be applied to any pretrained models from any modality, e.g., texts, images, and a variety of graph tasks, including node classification and link prediction.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Multi-modal</td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Principled</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Effective</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>?</td>
<td>✓</td>
</tr>
</tbody>
</table>

[Chien+ ’22, Zhao+ ’23]

Jing Zhu et. al
Key Intuition:
Make the node features obtained from the PMs graph-aware ($L_{\text{struct}}$).
Measure the awareness by feature homophily score.
TouchUp-G Pipeline

1. Feature Homophily $h$

- $d$
- $|V|$
- $h > 0.5$
- $h \leq 0.5$

2. Graph-Centric Finetuning

- Pretrained Model
- Baseline
- Graph Neural Networks

Jing Zhu et. al
TouchUp-G Pipeline

① Feature Homophily $h$

$$\begin{align*}
|V| & \quad d \\
\downarrow & \quad \downarrow \\
X & \quad |V| \\
\downarrow & \quad \downarrow \\
A & \quad |V|
\end{align*}$$

- $h > 0.5$
- $h \leq 0.5$

② Graph-Centric Finetuning

Jing Zhu et. al
Feature Homophily $h$

Def: The tendency of nodes with similar features to be connected to each other.

Feature Homophily is an implicit assumption shared by most GNNs. [Yang+ '21]

Goal: Quantify the correlation between features and structures for any graph with features and decide if TouchUp-G is needed.
Feature Homophily $h$

A vectorized extension of scalar assortativity.

$$h = \frac{\sum_{ij \in E} (x_i - \bar{x}) \cdot (x_j - \bar{x})}{\sqrt{\sum_{ij \in E} (x_i - \bar{x}) \cdot (x_i - \bar{x})} \cdot \sqrt{\sum_{ij \in E} (x_j - \bar{x}) \cdot (x_j - \bar{x})}}$$

$$\bar{x} = \frac{\sum_{ij \in E} (x_i + x_j)}{2|E|}$$

The Pearson correlation between the set of all head node features $x_i$ and the set of all tail node features $x_j$. Bounded in $[-1, 1]$. 

(a) $h = 1$

(b) $h = -1$
TouchUp-G Pipeline

1. Feature Homophily $h$

2. Graph-Centric Finetuning

- Pretrained Model
- Baseline
- Pretrained & TouchUp Transformer
- TouchUp-G
- Forward
- Backward
- $l_{struct}$

Graph Neural Networks

Jing Zhu et. al
Graph-centric Finetuning

\[ L_{\text{struct}} = -\frac{1}{|E|} \sum_{(u,v) \in E} \left( \log(x_u' \cdot x_v') + \log(1 - x_u' \cdot x_m') \right) \]

- For each training edge \((u, v) \in E\), randomly sample a negative node \(m\)
- \(x_u' = \max(T(x_u), 0)\), \(x_v' = \max(T(x_v), 0)\), \(x_m' = \max(T(x_m), 0)\) are the pretrained feature embeddings obtained from a pretrained model \(T\)
- Finetune \(T\) using binary cross entropy loss
Experiments

✧ **RQ1 - Effective**: How accurate is TouchUp-G?
✧ **RQ2 - Multi-modal**: Can TouchUp-G handle other modalities (like, images), in addition to text?
✧ **RQ3 - General**: Can TouchUp-G handle other downstream tasks (like node classification), besides link prediction?
✧ **RQ4 - Principled**: How well is the feature representation learnt by TouchUp-G, according to feature homophily score, compared with features obtained directly from pretrained models?
Amazon-CP: We extract the copurchasing information from the metadata in Amazon-Review. Each product’s high-resolution image are used as feature.

Books: Extract node feature is each book’s description and the links capture if a reader who recommends one book will recommend the other book.

For textual features, BERT/SciBERT is used as pretrained models.

For visual features, Vision Transformer (ViT) is used as pretrained models.
## Experiment Details

### Baselines:
- Structure embeddings: Degree, Deepwalk
- Feature embeddings: BERT+, ViT+, Ogb+, SciBERT+, DeBERTa+

### Evaluation:
- LP – report MRR, Hits@10, and Hits@1
- NC – report Acc

### GNN Backbones: SAGE, GCN, GATv2

<table>
<thead>
<tr>
<th>Name</th>
<th>Nodes</th>
<th>Edges</th>
<th>Description</th>
<th>Node Features</th>
<th>Pretrained Model</th>
<th>Downstream Task</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ogb-Arxiv [18]</td>
<td>169,343</td>
<td>1,166,243</td>
<td>Citation Network</td>
<td>Text</td>
<td>SciBERT [1]</td>
<td>NC</td>
</tr>
</tbody>
</table>
RQ1 Effective (LP)

Ogb-Products

(a) GCN

(b) SAGE

Jing Zhu et. al
RQ1 Effective (LP)

Books

(a) GCN

(b) SAGE

Jing Zhu et. al
RQ2 Multimodal(LP)

TouchUp-G works better than ViT+ both quantitatively and qualitatively. TouchUp-G correctly predicts the ground truth in top-2 predictions while ViT+ fails.

Jing Zhu et. al
RQ3 General (NC)

TouchUp-G works better than BERT+, SciBERT+.

For Ogb-Products, ogb+ mainly uses bag of words representation and works much better than BERT+, DeBERTa+.

This indicates contextualized embeddings from PMs hurts downstream graph task performance, if the contextualization is irrelevant wrt. the downstream task.

Jing Zhu et. al
### RQ4 Principled: Feature Homophily Score

All datasets exhibit a more than $2 \times$ increase in feature homophily score.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ogb-Products</td>
<td>0.223</td>
<td>-</td>
<td>-</td>
<td><strong>0.762 (3.4x)</strong></td>
</tr>
<tr>
<td>Books</td>
<td>0.137</td>
<td>-</td>
<td>-</td>
<td><strong>0.579 (4.2x)</strong></td>
</tr>
<tr>
<td>Amazon-CP</td>
<td>-</td>
<td>-</td>
<td>0.173</td>
<td><strong>0.622 (3.6x)</strong></td>
</tr>
<tr>
<td>Ogb-Arxiv</td>
<td>-</td>
<td>0.194</td>
<td>-</td>
<td><strong>0.408 (2.1x)</strong></td>
</tr>
</tbody>
</table>
We thank the following for the helpful feedback on this project.