metapath2vec: Scalable Representation Learning for Heterogeneous Networks

Yuxiao Dong, Yuchun Jo, and Ananthram Swami

MetaPath-Based Random Walks

We design meta-path-based random walks to generate paths that are able to capture both the semantic and structural correlations between different types of nodes, facilitating the transformation of heterogeneous network structures into metapath2vec's skip-gra.

Heterogeneous Network Mining Tasks

Node label inference
Community detection
Similarity search
Link prediction
...

The heterogeneous skip-gram model used in metapath2vec++ when predicting for α. Instead of one set of multinomial distributions for all types of neighborhood nodes in the output layer, it specifies one set of multinomial distributions for each type of node in α_i's neighborhood.

Network maximization in both
arg max
θ

Softmax in metapath2vec

Softmax in metapath2vec++

Objective function in metapath2vec++ (heterogeneous negative sampling)

Heterogeneous Network Data

AMiner [6]: 9 million authors, 3 million papers, 3800+ venues, and 8 categories of venues for labeling both venues & authors.

Computer Linguistics

Computer Graphics

Computer Networks

Computer Vision

Computing Systems

Databases & Info

Human Computer Interaction

Theoretical Computer Science

DBIS [5]: 5 thousand authors, 72 thousand papers, 464 venues.

Experiments: Label Prediction

Multi-class venue node classification results in AMiner data

<table>
<thead>
<tr>
<th>Method</th>
<th>Paper</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Micro-F1</th>
<th>Macro-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepWalk nodevec</td>
<td>0.0723</td>
<td>0.6378</td>
<td>0.1905</td>
<td>0.2756</td>
<td>0.3427</td>
<td>0.5913</td>
</tr>
<tr>
<td>LINE (1st+2nd)</td>
<td>0.2245</td>
<td>0.4629</td>
<td>0.7181</td>
<td>0.8473</td>
<td>0.8953</td>
<td>0.9203</td>
</tr>
<tr>
<td>PTE</td>
<td>0.1702</td>
<td>0.3888</td>
<td>0.6555</td>
<td>0.8394</td>
<td>0.8956</td>
<td>0.9210</td>
</tr>
<tr>
<td>metapath2vec</td>
<td>0.3053</td>
<td>0.5247</td>
<td>0.8023</td>
<td>0.9371</td>
<td>0.9466</td>
<td>0.9532</td>
</tr>
<tr>
<td>metapath2vec++</td>
<td>0.3090</td>
<td>0.5444</td>
<td>0.8400</td>
<td>0.9895</td>
<td>0.9958</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Micro-F1

<table>
<thead>
<tr>
<th>Method</th>
<th>Paper</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
<th>Micro-F1</th>
<th>Macro-F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepWalk nodevec</td>
<td>0.1791</td>
<td>0.2412</td>
<td>0.4860</td>
<td>0.5378</td>
<td>0.5890</td>
<td>0.6490</td>
</tr>
<tr>
<td>LINE (1st+2nd)</td>
<td>0.3000</td>
<td>0.5541</td>
<td>0.7510</td>
<td>0.8457</td>
<td>0.8950</td>
<td>0.9209</td>
</tr>
<tr>
<td>PTE</td>
<td>0.2512</td>
<td>0.4267</td>
<td>0.6870</td>
<td>0.8372</td>
<td>0.8950</td>
<td>0.9219</td>
</tr>
<tr>
<td>metapath2vec</td>
<td>0.4173</td>
<td>0.5975</td>
<td>0.8327</td>
<td>0.9311</td>
<td>0.9480</td>
<td>0.9522</td>
</tr>
<tr>
<td>metapath2vec++</td>
<td>0.4331</td>
<td>0.6192</td>
<td>0.8336</td>
<td>0.9302</td>
<td>0.9463</td>
<td>0.9581</td>
</tr>
</tbody>
</table>

Acknowledgements

Yuxiao Dong, Nitesh V. Chawla, Ananthram Swami.

metapath2vec: Scalable Representation Learning for Heterogeneous Networks.

In Proc. of 23rd SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD'17).

References