Label Incorporated Graph Neural Networks for Text Classification

Yuan Xin†, Linli Xu†§, Junliang Guo†, Jiquan Li†, Xin Sheng†, Yuanyuan Zhou†
†Anhui Province Key Laboratory of Big Data Analysis and Application,
School of Computer Science and Technology, University of Science and Technology of China
‡Information Science Center, University of Science and Technology of China
§IFLYTEK Co., Ltd.
†lwjxy@mail.ustc.edu.cn, linlixu@ustc.edu.cn, {guojunll,lijiquan,xins}@mail.ustc.edu.cn
‡yoyo@ustc.edu.cn

Yuan Xin 2021/1/10
• Background

• Our model: label-incorporated GCN

• Experiment

• Conclusion
01 Background
Background

• **What is text classification**
  • A task to annotate a given text sequence with one or multiple class labels.

• **What is graph convolutional network**
  • A Graph Convolution Network (GCN) is a multi-layer neural network that directly operates on graph structures and learns node embeddings based on their neighbors.

• **Our work:** build a label-incorporated GCN and transform text classification problem into node classification problem
The difference between existing studies and our work

- Previous works only consider the text information while building the graph, heterogeneous information such as labels is ignored.
  - Eg: TextGCN (L. Yao, C. Mao, and Y. Luo, 2019)

- We treat labels as nodes in the graph which also contains text and word nodes, and then connect labels with texts belonging to that label
02 Label-incorporated GCN
Label-incorporated GCN

- Heterogeneous Text Graph Construction
  - Word-Word Subgraph: captures word co-occurrences in local contexts
    \[
    \text{PMI}(i, j) = \log \frac{p(i, j)}{p(i)p(j)}, \quad p(i, j) = \frac{\#W(i, j)}{\#W}, p(i) = \frac{\#W(i)}{\#W}.
    \]
  - Word-Text Subgraph: captures word-co-occurrences in text level
    \[
    A_{i,j} = \text{TF-IDF}_{v_i,d_j}.
    \]
  - Text-Label Subgraph
    - Connect text with its corresponding labels, through this connection we incorporate and propagate the label information through text-label-text paths.
Label-incorporated GCN

- We send the heterogeneous graph into GCN for classification

\[ Z = \text{softmax}(\tilde{A}f(\tilde{A}XW_0)W_1), \]

- Loss function:

\[ \mathcal{L} = -\sum_{d \in V_d} Y_d \ln Z_d + \lambda \sum_{l \in V_l} Y_l \ln Z_l, \]
03 Experiment
TABLE II: Percentage test accuracy on the text classification task, some results are directly taken from previous works. Our method outperforms baselines by a significant margin.

<table>
<thead>
<tr>
<th>Model</th>
<th>R8</th>
<th>R52</th>
<th>Ohsumed</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF + LR</td>
<td>93.74 ± 0.00</td>
<td>86.95 ± 0.00</td>
<td>54.66 ± 0.00</td>
</tr>
<tr>
<td>CNN-rand</td>
<td>94.02 ± 0.57</td>
<td>85.37 ± 0.47</td>
<td>43.87 ± 1.00</td>
</tr>
<tr>
<td>CNN-non-static</td>
<td>95.71 ± 0.52</td>
<td>87.59 ± 0.48</td>
<td>58.44 ± 1.06</td>
</tr>
<tr>
<td>LSTM</td>
<td>93.68 ± 0.82</td>
<td>85.54 ± 1.13</td>
<td>41.13 ± 1.17</td>
</tr>
<tr>
<td>LSTN(pretrain)</td>
<td>96.09 ± 0.19</td>
<td>90.48 ± 0.86</td>
<td>51.10 ± 1.50</td>
</tr>
<tr>
<td>Bi-LSTM</td>
<td>96.31 ± 0.33</td>
<td>90.54 ± 0.91</td>
<td>49.27 ± 1.07</td>
</tr>
<tr>
<td>PV-DBOW</td>
<td>85.87 ± 0.10</td>
<td>78.29 ± 0.11</td>
<td>46.65 ± 0.19</td>
</tr>
<tr>
<td>PV-DM</td>
<td>52.07 ± 0.04</td>
<td>44.92 ± 0.05</td>
<td>29.50 ± 0.07</td>
</tr>
<tr>
<td>fastText</td>
<td>86.04 ± 0.24</td>
<td>71.55 ± 0.42</td>
<td>14.59 ± 0.00</td>
</tr>
<tr>
<td>fastText(bigrams)</td>
<td>82.95 ± 0.03</td>
<td>68.19 ± 0.04</td>
<td>14.59 ± 0.00</td>
</tr>
<tr>
<td>SWEM</td>
<td>95.32 ± 0.26</td>
<td>92.94 ± 0.24</td>
<td>63.12 ± 0.55</td>
</tr>
<tr>
<td>LFAM</td>
<td>93.31 ± 0.24</td>
<td>91.84 ± 0.23</td>
<td>58.58 ± 0.79</td>
</tr>
<tr>
<td>Graph-CNN-C</td>
<td>96.99 ± 0.12</td>
<td>92.75 ± 0.22</td>
<td>63.86 ± 0.53</td>
</tr>
<tr>
<td>Graph-CNN-S</td>
<td>96.80 ± 0.20</td>
<td>92.74 ± 0.24</td>
<td>62.82 ± 0.37</td>
</tr>
<tr>
<td>Graph-CNN-F</td>
<td>96.89 ± 0.06</td>
<td>93.20 ± 0.04</td>
<td>63.04 ± 0.77</td>
</tr>
<tr>
<td>TextGCN</td>
<td>97.07 ± 0.10</td>
<td>93.56 ± 0.18</td>
<td>68.36 ± 0.56</td>
</tr>
<tr>
<td><strong>our model(label-text connection)</strong></td>
<td><strong>97.37 ± 0.17</strong></td>
<td><strong>94.15 ± 0.10</strong></td>
<td><strong>69.10 ± 0.20</strong></td>
</tr>
<tr>
<td><strong>our model's variant(label-word connection)</strong></td>
<td><strong>97.10 ± 0.07</strong></td>
<td><strong>93.91 ± 0.11</strong></td>
<td><strong>68.93 ± 0.17</strong></td>
</tr>
</tbody>
</table>
04 Conclusion
Conclusion

• What we do?
  • We build a novel heterogeneous graph convolutional network for text classification by adding label nodes to the graph
  • We designed an auxiliary classification loss function of the label embeddings to enhance the interpretability of label representations.
  • Experimental results demonstrate the superior performance of the proposed model over baselines on several text classification benchmark datasets.

• Future
  • In the future, we will apply our models to other scenarios and applications leading to a more general framework.
Thank you!  Q&A

Yuan Xin

2021/1/10