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**

Background

- **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

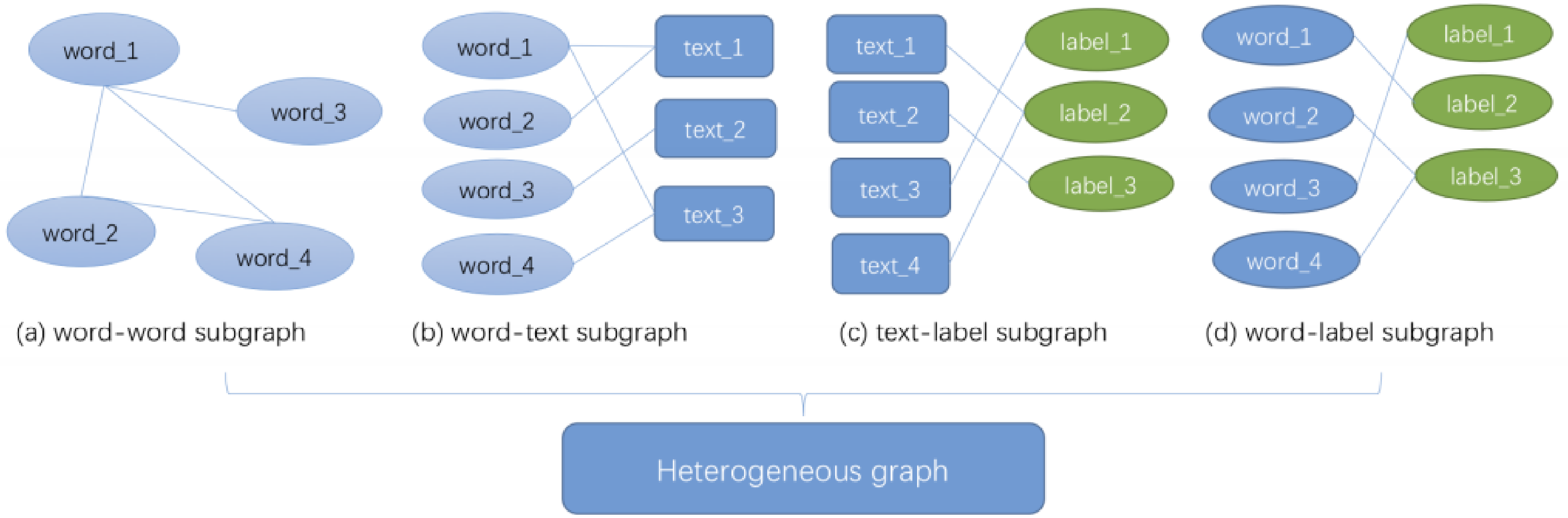
$$\mathrm{PMI}(i,j) = \log \frac{p(i,j)}{p(i)p(j)}, \qquad p(i,j) = \frac{\#W(i,j)}{\#W}, p(i) = \frac{\#W(i)}{\#W}.$$

  - 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

$$Z = \mathrm{softmax}(\tilde{A} f(\tilde{A} X W_0) W_1),$$

- Loss function: $\mathcal{L} = -(\sum_{d \in V_d} \mathrm{Y}_d \ln \mathrm{Z}_d + \lambda \sum_{l \in V_l} \mathrm{Y}_l \ln \mathrm{Z}_l),$

03 Experiment

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.

Model	**R8**	**R52**	**Ohsumed**
TF-IDF + LR	93.74 ± 0.00	86.95 ± 0.00	54.66 ± 0.00
CNN-rand	94.02 ± 0.57	85.37 ± 0.47	43.87 ± 1.00
CNN-non-static	95.71 ± 00.52	87.59 ± 00.48	58.44 ± 1.06
LSTM	93.68 ± 0.82	85.54 ± 1.13	41.13 ± 1.17
LSTN(pretrain)	96.09 ± 0.19	90.48 ± 0.86	51.10 ± 1.50
Bi-LSTM	96.31 ± 0.33	90.54 ± 0.91	49.27 ± 1.07
PV-DBOW	85.87 ± 0.10	78.29 ± 0.11	46.65 ± 0.19
PV-DM	52.07 ± 0.04	44.92 ± 0.05	29.50 ± 0.07
fastText	86.04 ± 0.24	71.55 ± 0.42	14.59 ± 0.00
fastText(bigrams)	82.95 ± 0.03	68.19 ± 0.04	14.59 ± 0.00
SWEM	95.32 ± 0.26	92.94 ± 0.24	63.12 ± 0.55
LEAM	93.31 ± 0.24	91.84 ± 0.23	58.58 ± 0.79
Graph-CNN-C	96.99 ± 0.12	92.75 ± 0.22	63.86 ± 0.53
Graph-CNN-S	96.80 ± 0.20	92.74 ± 0.24	62.82 ± 0.37
Graph-CNN-F	96.89 ± 0.06	93.20 ± 0.04	63.04 ± 0.77
TextGCN	97.07 ± 0.10	93.56 ± 0.18	68.36 ± 0.56
our model(label-text connection)	$\mathbf{97.37 \pm 0.17}$	$\mathbf{94.15 \pm 0.10}$	$\mathbf{69.10 \pm 0.20}$
our model's variant(label-word connection)	$\mathbf{97.10 \pm 0.07}$	$\mathbf{93.91 \pm 0.11}$	$\mathbf{68.93 \pm 0.17}$

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