Zero-Shot Text Classification with Semantically Extended Graph Convolutional Network

Tengfei Liu[1], Yongli Hu[1,*], Junbin Gao[2], Yanfeng Sun[1] and Baocai Yin[1,3]
[1]Beijing University of Technology,China
[2]The University of Sydney,Australia
[3]Dalian University of Technology,China

Introduction

As a challenging task of Natural Language Processing(NLP), zero-shot text classification has attracted more and more attention recently. It aims to detect classes that the model has never seen in the training set. For this purpose, a feasible way is to construct connection between the seen and unseen classes by semantic extension and classify the unseen classes by information propagation over the connection.

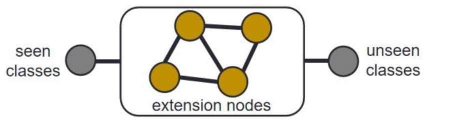

Fig 1. The main solution for the zero-shot text classification

Motivation

- Proposing a novel method for graph construction by extending each category with three types of semantic knowledge, the class description, the class hierarchy and the class knowledge in ConceptNet, to tackle the zero-shot text classification.
- The Graph Convolutional Networks is firstly introduced into the task of zero-shot text classification, to the best of our knowledge.

BackGround: Graph Convolutional Network

A graph can be formulated as $G=(V,E,A)$,where $V(|V|=n)$ is the set of graph nodes, E is the set of graph edges, and $A\in R^{n\times n}$ is the graph adjacency matrix. Given a graph with N nodes and d input features per node,we represent the input feature matrix as $X\in R^{N\times d}$ and the propagation rule is as follows:

$$H^{(l+1)}=f\left(H^{(l)},A\right)=\sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}}H^{(l)}W^{(l)}\right) \quad (1)$$

where $\tilde{A}=A+I$ is the adjacency matrix with added self-connections, I_N is the identity matrix, $\tilde{D}$ is diagonal degree matrix with $\tilde{D}_{ii}=\sum_j \tilde{A}_{ij}$.

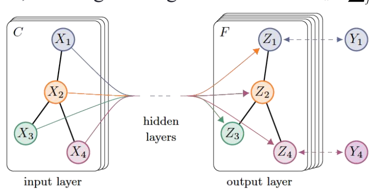

Fig 2. Graph Convolutional Network

Graph Construction via Semantic Extension

As shown in Fig 3,three types of semantic node are considered,superclasses, the nouns in the class description and related knowledge graph nodes in ConceptNet of each categories.

we only consider four relations,i.e.'/r/IsA', '/r/PartOf', '/r/AtLocation' and '/r/RelatedTo' in our semantic extension. We rank the related nodes from high to low according to the weight given in ConceptNet and only the top ten nodes as the extended nodes.

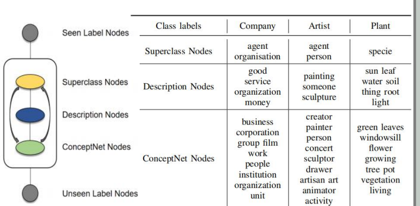

Fig 3. An illustrations of semantic category extension

Semantically Extended Graph Convolutional Network

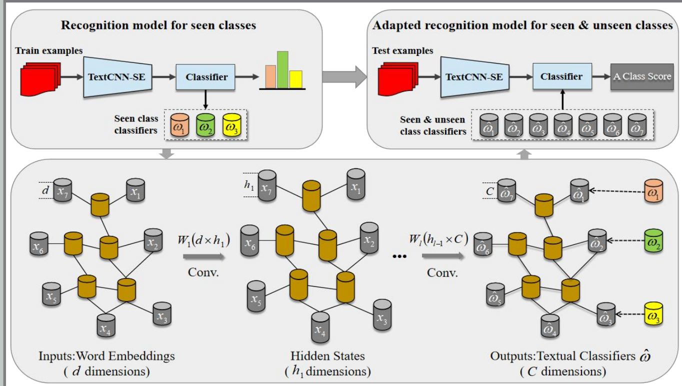

Fig 4.SEGCN consists of three stages: 1) The TextCNN-SE model is trained on the seen texts based on classification loss to obtain the seen class classifiers which are regarded as the ground-truth; 2) The GCN takes word embeddings as inputs and outputs the textual classifier of each category based on the graph constructed by semantic category extension. 3) Using the textual classifier of each category, we adapt the recognition model (top-left) into a new model (top-right) that can classify both the seen and unseen classes in a unified way.

Results and Discussion

Table 1.The classification results of different methods.We run all models 10 Times and report the mean Accuracy.Con_nodes(w/o) means that graph construction without using related nodes in conceptnet of each category.

Datasets	Dbpedia						20news					
Unseen rate	25%			50%			25%			50%		
y	seen	unseen	overall	seen	unseen	overall	seen	unseen	overall	seen	unseen	overall
Count-base	0.322	0.372	0.334	0.358	0.304	0.333	0.205	0.201	0.204	0.219	0.196	0.207
Label Similarity [28]	0.377	0.426	0.386	0.401	**0.369**	0.386	0.279	**0.287**	0.280	0.293	**0.266**	0.280
RNN Autoencoder	0.250	0.267	0.254	0.202	0.259	0.230	0.263	0.149	0.236	0.275	0.126	0.200
RNN+FC [5]	0.895	0.046	0.713	0.960	0.044	0.502	0.614	0.065	0.482	0.709	0.052	0.381
CNN+FC	**0.985**	0.204	0.818	**0.991**	0.069	0.530	0.792	0.134	0.633	0.684	0.126	0.405
ISK-ZSTC [7]	0.975	0.402	0.852	0.982	0.197	0.590	0.745	0.280	0.633	0.767	0.168	0.469
Ours (con_nodes(w/o))	0.975	**0.431**	**0.858**	0.948	0.305	**0.621**	0.823	0.192	0.669	0.797	0.162	0.481
Ours	0.983	0.269	0.830	0.972	0.224	0.598	**0.843**	0.211	**0.690**	**0.810**	0.171	**0.491**

The classification results of different methods are presented in Table 1. It is shown that our proposed model outperforms all baselines in the overall accuracy. Specially, our model outperforms the most related method ISK-ZSTC with 3.1% improvement for Dbpedia and 5.7% for 20news, which means that GCN is able to transfer knowledge well between different categories over the graph constructed by our semantic category extension.

Conclusion

We have presented an approach for zero-shot text classification by using semantic category extension. Based on the graph constructed by extension nodes and category nodes, we generate a textual classifier for each category with the special network of GCN, over which the extension nodes strengthen the knowledge transfer from seen to unseen classes. The experiments on two datasets show that the textual classifiers obtained by the GCN can improve the unseen and overall accuracy on both datasets with different unseen rates. Furthermore, we discuss the influence of different ways of semantic extension and their combinations. In the future, the advanced language understanding model, e. g. BERT can be taken into account to extract the text features.

Acknowledgments and Main Reference

National Natural Science Foundation of China under Grant No.61772048,61632006,61672071, U1811463, U19B2039, Beijing Talents Project(2017A24).

[1] P. K. Pushp and M. M. Srivastava.Train once,test anywhere: Zero-shot learning for text classification.CoRR,2017.
[2] J. Zhang, P. Lertvittayakumjorn and Y. Guo.Integrating semantic knowledge to tackle zero-shot text classification.In NAACL 2019.