E0-270 : Machine Learning
Indian Institute of Science,
Bangalore.

Zero Shot Recognition Using GCN

Ameenudeen PE, Danish Shaikh,
Deepa TM, Najath Pathiyil

April 27, 2019

Overview

Introduction

- Zero-Shot Learning: A method to predict the labels of unseen examples without the burden of collecting its training samples.
- Lets elaborate in more detail, consider the below image.

- Okapi : "zebra-striped four-legged animal with a brown torso and a deer-like face."

Problem Statement

- **Motivation**: Can we build a classifier given the above mentioned constraints?
- **Solution**: Transfer knowledge that is obtained from seen classes to describe the unseen classes.
  1. Learn a vector representation of different categories using text data and then learn a mapping between the vector representation to visual classifier directly.
  2. Use explicit knowledge bases or knowledge graphs.

Our Method

- Combine both word embeddings and knowledge graphs.
- Each node corresponds to a class and relationships among them are shown via edges.
- Word embeddings of each category is given as the node input.
- Use GCN to transfer information between the layers.

ConSE

- Convex Combination of Semantic Embeddings (ConSE) uses Method 1 approach.
- We estimate the conditional probability distribution of each training class given the samples.
- For test samples, we compute a weighted combination of the label embeddings of top 'T' training classes in the semantic space where, T is hyperparameter.

$$f(\mathbf{x}) = \frac{1}{Z} \sum_{t=1}^{T} p(\widehat{y}_0(\mathbf{x}, t) \mid \mathbf{x}) \cdot s(\widehat{y}_0(\mathbf{x}, t))$$

- Using cosine similarity to calculate the nearest class in the semantic space.

GCN for Zero-shot Learning

- C denotes the set of all classes
  - C_{te}: Testing classes.
  - C_{tr}: Training classes.
- Given a graph G = (V, E), it takes the feature matrix and Adjacency matrix of the graph as the input.
- $C_{te} \cap C_{tr} = \phi$
- Training data points: $D_{tr} = (X_i, c_i), i = 1, ...N$. $c_i \in C_{tr}$ is the class-label.
- We give 300-D semantic representation vector of all classes to the graph.
- Using GCN, we predict the class labels of C_{te} in a semi supervised manner.

Approach

- Given: A graph with N nodes and S input features per node, $X \in R^{NXS}$ denotes the feature matrix.
- Each node represents one distinct class.
- Adjacency matrix $A \in R^{NXN}$, shows connections between the classes in the knowledge graph.
- Propagation rule : $H^{i+1} = f(D^{-1}AH^{i}W^{i})$
  - H^{i}: Activations in the i^{th} layer
  - W^{i}: Trainable weight matrix for layer i.
  - D: Degree matrix.
- $H^{0} = X$
- The activation function f used is Leaky Relu.

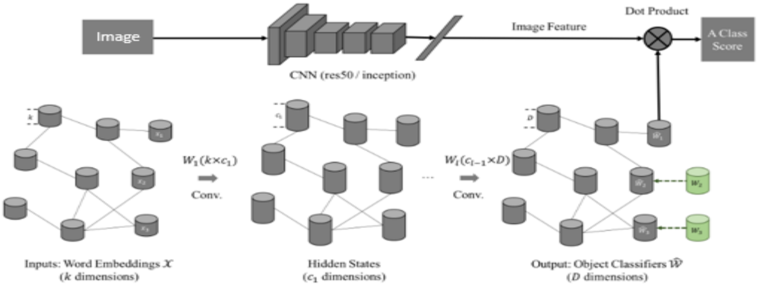


Figure: Block diagram of the approach

Training and Testing

Training :

- We use the mean square error between the predicted and ground truth weights as loss function and use this to estimate the parameters of GCN.

$$L = \frac{1}{2M}\sum_{i=1}^{M}\sum_{j=1}^{P}(W_{ij} - \hat{W}_{ij})^2$$

,
M: no. of training classes, $\hat{W}$: predicted weights, P: dimensionality of weights.
Using these parameters, classifier weights for zero shot categories are estimated.

Testing :

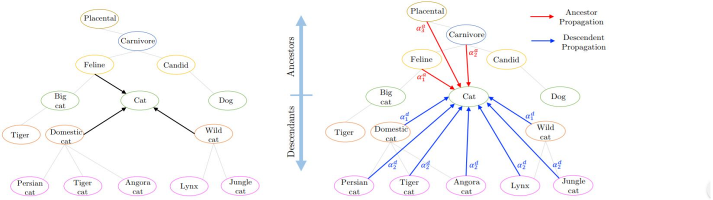
- Extract the features of testing images using pre-trained CNN.
- Cosine similarity of those features with generated classifiers.

Implementation Details

- Training : Imagenet 2012 1K dataset
- Testing.
  1. 300 classes from "2-hops".
  2. 100 random classes.
  3. AWA dataset consisting of 50 classes.
- Knowledge graph: sub-graph of the WordNet.
- Feature extraction using ResNet-50.
- Word embedding by GloVe text model.
- Average of class attributes of the word embeddings.
- 6 Convolution layers.
- Output feature dimension: 2048
- During inference phase, cosine-similarity is used between learned GCN weights and test set features from CNN.
- Softmax score.

Modified approach: Dense Graph Propagation (DGP)

- Normal graph: Dilution of knowledge because of heavy smoothing in each layer.
- Solution: Dense Graph and convert it into one layer.



(a) Graph Propagation (b) Dense Graph Propagation

Figure: GP vs. DGP for Node 'Cat'

- DGP for zero-shot learning aims to use a two phase hierarchical graph structure: namely descendant propagation and ancestor propagation.

Results and Discussions

Test Set	Model	Hit @ k %				
		1	2	5	10	20
2-hops (300 classes)	GCNZ	49.8	65.4	78.3	85.4	90.4
2-hops + 1K	GCNZ	13.7	40	61.7	72.6	80.1
Random 100 classes	GCNZ	2.1	3.6	7.9	11.8	20.2
Random 100 classes + 1K	GCNZ	1.2	2.4	3.8	8.1	13.5
100 training classes	GCNZ	55.3	70.1	79.5	84.4	89.7
100 training classes	DGP	65.53	80.09	91.75	96.12	98.53

Figure: Top K accuracy for different models in different settings.

Dataset	ConSE	GCNZ	DGP
AWA	52.67	67.92	72.18

Figure: Test Accuracy with AWA dataset.

Conclusions

- Our work shows that a knowledge graph provides supervision to learn meaningful classifiers on top of semantic embeddings.
- We also compared our results with current state-of-the-art ConSE and got significant improvements.
- A modified approach using a DGP module is also implemented and results were improved significantly.
- We also observed that the DGP model has overcome the problem of knowledge dilution.

Future Scope

- There are many hyperparameters that we did not tune due to the lack of time, like, number of layers in GCN, the CNN model to extract features, etc.
- We have given input as averaging of the class-attributes. It can be replaced with another weighted graph that gives relation of these attributes.
- We can use weighted edge between classes in the graph instead of direct connection.
- Instead of pre-trained word-embedding space, one can use some other space to get the relation between test and training samples.

References

- [1] Yongqin Xian, Christoph H. Lampert, Bernt Schiele, Zeynep Akata: "Zero-Shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly". CoRR abs/1707.00600 (2017).
- [2] Xiaolong Wang, Yufei Ye, Abhinav Gupta: "Zero-Shot Recognition via Semantic Embeddings and Knowledge Graphs". CVPR 2018: 6857-6866.
- [3] Michael Kampffmeyer, Yinbo Chen et al.: "Rethinking Knowledge Graph Propagation for Zero-Shot Learning" to appear in CVPR 2019.

Thank You!