上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

Self-introduction

Zhehao Zhang （张哲昊）

饮 水 思 源 · 爱 国 荣 校

目 录

01

Basic information

Basic information

- Zhehao Zhang (张哲昊)
- Third-year undergraduate student in Shanghai Jiao Tong University, major in Artificial Intelligence.
  GPA: 88.20/100 (Zhiyuan honor program)
- Zhiyuan honor program scholarship(2019-2022)
- Research Interest
  1. Deep learning (continual learning, domain adaptation)
  2. Natural Language Process
  3. Computer vision

Detail information can be found in my personal website!

https://zzh-sjtu.github.io/zhe_hao_Zhang.github.io/

02

Learned courses

Selected courses

- Computer vision
- Natural language processing
- Reinforcement learning
- Machine learning
- Data mining ...

Full courses can be found in my personal website or transcripts.

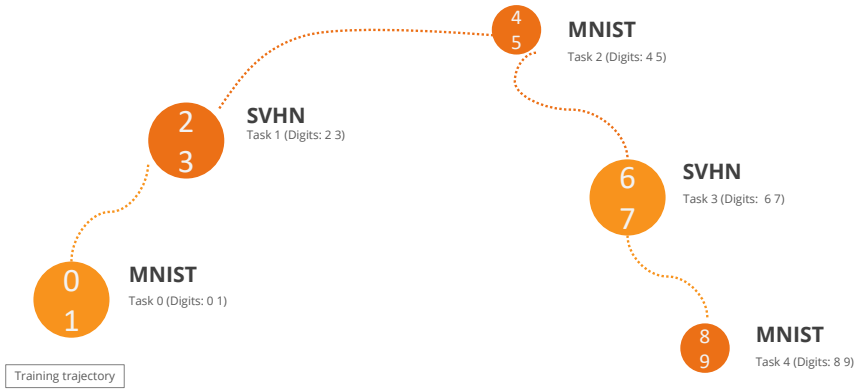
Self-study courses

CS229 CS231n CS224n
Dive into Deep Learning
Hung-yi Lee’s machine learning course ...

Course Projects

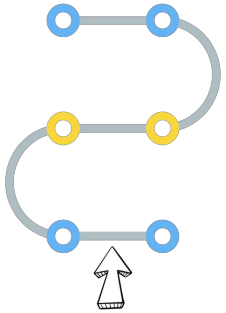
Continual learning with domain shift

Propose a new algorithm to solve the catastrophic forgetting problem with large difference between tasks in class incremental learning.



Modified DeepWalk for link prediction

Propose a new algorithm to utilize attribute information in order to do better in graph embedding for link prediction task.

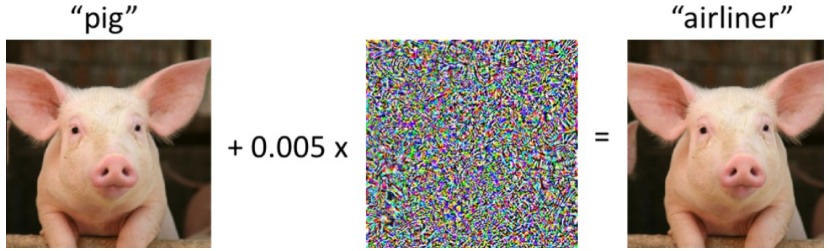


Sample weight of trajectory

$$weight(v_i, v_j) = \gamma * AS(v_i, v_j) + (1 - \gamma) * SS(v_i, v_j)$$

Adversarial Attacks

Implement several attack method for image classification of CIFAR10. Implement training method for the defend of several attacks.



Projects Reports

Continual learning with domain shift

How human actually learn?
Continual learning: A more practical setting for image classification

Zhehao Zhang
Shanghai Jiao Tong University
zzh121380@sjtu.edu.cn

Abstract

Is traditional supervised learning always suitable for image classification? Do human beings learn things through supervised learning fashion? The answer to both above questions is absolutely no. As a result, in recent years, the setting of continual learning has drawn more and more attention. In my project, I study continual learning (especially class incremental learning), analyze its challenges, implement several classical methods. Besides, I **formalize a new setting** which is more practical and more similar to the process of human learners. After that, I **propose a new method** to solve this harder problem with performance better than classic methods. Furthermore, I also **learn new things** such as Fisher Information Matrix, Knowledge distillation and Causal inference.

1 Introduction

1.1 Background and settings

In the field of image classification, supervised learning has always been the mainstream technique since the era of deep learning. According to the traditional supervised learning fashion, the input dataset with all classes is shuffled and then fed into the classification model during training. During a fixed period of time, the probability to be encountered by the model is equal for every class. As a result, the model will update its parameters to minimize the loss function of every class equally without any bias (if there is no imbalance between classes in the dataset).

However, the learning process of the traditional supervised learning fashion is quite different from that of human learners. For example, it is not practical for a person to look through all classes of objects before classifying images. On the contrary, We always learn new things through our life and these new things are impossible to show up before a certain moment. This is why the setting of continual learning is also called Lifelong learning, and incremental learning.

There are different sub-settings in the field of continual learning, such as task incremental, domain incremental, and class incremental. The difference between each of them can be found in [1], and we will only discuss the setting of class incremental. The setting of class incremental learning is clearly illustrated by Figure 1.

1.2 Challenges and goals

Just like human beings, a classification model (eg. deep neural networks) will also witness the forgetting process. As a result, the major challenge is to learn without the so-called catastrophic

https://zzh-sjtu.github.io/zhe_hao_Zhang.github.io/script/CV_final_project-4.pdf

Modified DeepWalk for link prediction

Combination of Structural and Attributed Phased DeepWalk Method for Link Prediction

Zhehao Zhang,Zilong Wang, Juntu Zhao

Abstract—Graph data mining is a very important branch of data mining. Many things in our life can be described by a graph structure, such as our social network and paper citations network. Therefore, the study of graph data is particularly important. Link prediction is a classic task in network science, which is used to predict which new links will appear in the network using the currently obtained network. However, networks with only structural similarity are not practical in our real life. On the contrary, networks with more information (attribute) about the node are widely used such as social networks. To better utilize the attribute information, we propose a new algorithm called CSAPDW for link prediction in attributed networks by making use of both the structural similarity and attribute similarity of the network. Efficient experiment results indicate that our methods perform better than others.

Index Terms—Link prediction, Attributed network, node similarity, random walk

1 Introduction

Complex networks can be used to describe many natural phenomena, such as our social network and paper citation network, etc. Nowadays graph network technology has developed into a very common and important field, in which we have encountered many important challenges such as Community detection, link prediction, etc. Link prediction is one of the very important topics. It uses past data to predict the future structure of a complex network, that is, existing unknown states and possible future states can be estimated through link prediction.

The methods used in link prediction can be mainly divided into three types: similarity-based prediction, model-based prediction, and embedded technology. Among them we use embedded technology that is a very efficient method for solving complex network analysis tasks, which maps the network to a low-dimensional space, preserving the desired main feature information. However, the embedded-based link prediction methods we have learned before mainly focus on the structural information of the graph. While in real scenarios, nodes in the network not only have structural features related to other nodes but also have their own feature information.

For example, in the classmate information table, each classmate has his own gender, student number, and dormitory number, this information should obviously also be used as important content in the interpersonal network. Through the above example, we can find that just like using the structural features of nodes, using the own characteristics of network nodes can also help improve the accuracy of link prediction.
In order to solve the above problems, we modify the deep walk algorithm random walk strategy in the attribute network. Because the original deep walk algorithm only uses the structural information of the network but does not use the point's own feature information (it is called attribute information in the attribute network). This paper uses a linear time complexity algorithm, uses local structural features and attribute features, maps the network to a low-dimensional space, and obtains node features for subsequent prediction.

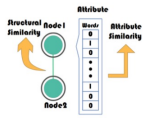


Fig. 1. Structural and attribute similarity between two nodes

At the same time, in order to further improve the accuracy of the algorithm, we also use the prediction results of each time to make loop predictions. Specific details will be described later.

In this project, we have used a lot of techniques in data mining such as node embedding and computation of similarity. Besides, graph mining is an important part of data mining. **(This paragraph is to show the relationship between our project and data mining)**

The rest of the paper is structured as follows: In Sect. 2, the introduction of related work on link prediction. In Sect. 3, The specific introduction and implementation of our algorithm. In Sect. 4, our experimental results. In Sect.5, the discussion, and in Sect.6, our conclusions.

2 Related Work

This section briefly introduces related research on link prediction and our work.

2.1 Similarity-based methods and Probabilistic model-based methods

First, similarity-based methods are the most common link prediction methods. By assigning similarity to each pair of nodes, we can judge the probability that there is an edge

https://zzh-sjtu.github.io/zhe_hao_Zhang.github.io/script/DM_link_prediction.pdf

Adversarial Attacks

模型攻击实验报告

张哲昊 519030910383
王子龙 51903091038

日期：2021 年 6 月 13 日

摘 要

本次大作业攻击任务，小组成员在 PGD baseline 的基础上对每个攻击模型，进行测试，发现可能的防御方式。同时设计其他攻击算法（如黑盒攻击）对防御模型进行攻击，验证猜想。并且在 PGD baseline 的基础上进行了多种实验尝试，找到了能够进一步提升攻击效果的方法。接下来，小组成员广泛阅读相关文献，进行理论分析，并加以借鉴最终实现了 Myattack 攻击算法，使得结果更加理想。

1 黑盒攻击尝试

1.1 PGD baseline 实验结果

PGD baseline accuracy on each model						
Method	model1	model2	model3	model4	model5	model6
No attack	0.9429	0.83020	0.80330	0.84920	0.81420	0.88260
PGD attack	0.0004	0.51320	0.64340	0.56170	0.54810	0.64340

1.2 简单黑盒攻击的实现

为了测试 model3 的防御方式，我们设计了一个简单的黑盒随机攻击算法，主要方式为将原始图片加上随机噪声。算法伪代码如下：

Algorithm 1: Simple black-box attack
Result: picture with noise
Initialization: $x_{adv} = x_{input}$ $model = model_{input}$ $y = label$
$loss = cross\_entropy(model(x), y)$;
for $i \leftarrow 0$ **to** *query_budget* **do**
  Sample noise: $\delta \sim \alpha \cdot \mathcal{N}(0, 1)$
  $add\_noise = clip(x_{adv} + \delta)$
  $x_{adv} = clip(add\_noise)$;
end
return x_{adv}

1

https://zzh-sjtu.github.io/zhe_hao_Zhang.github.io/script/attack.pdf

Course Projects

Other Projects

- Spoken Language Understanding
- SIFT implement
- Discounting algorithm for language model
- Community detection ...

Other projects can be found on my github.

https://github.com/zzh-SJTU

03

Research Experience

Continual Learning

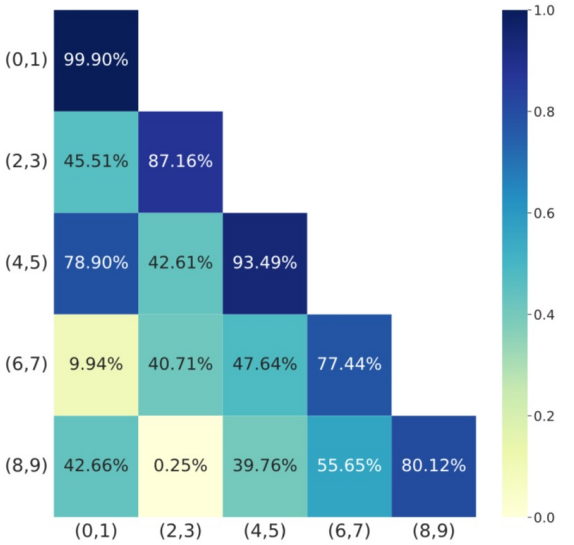
Topic: Causal NLP. Advisor: Prof. Shuai Li

Algorithm 1: Improvement of current method to work well for different dataset

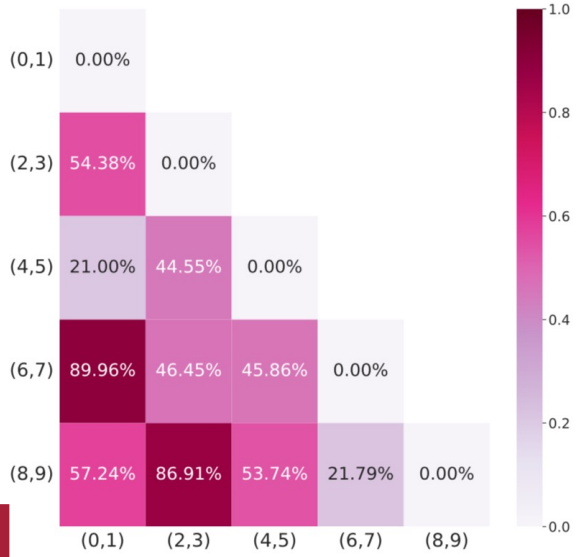
Data: Stream data from MNIST and SVHN with image x_i and label Y_i
Initialization: A classifier: f, model buffer : B. hyper parameter $\lambda_1$andλ_2
. **for** $t \leftarrow 0$ **to** *number of tasks* **do**
  if $t > 0$ **then**
    $Loss = 0$
    Take out the model f_i out from model buffer B.
    Feed the image from the current task to f_i to get the Pseudo label Y_i'.
    The total loss consists of three parts: distillation loss, Fisher information loss, new task cross entropy.

$$Loss = \sum_{i,d(i)=d(n)} \lambda_1 L_{old}(Y_i', \hat{Y_i'}) + \sum_i \lambda_2 F_i(\theta_i - \theta^*_{old,i})^2 + L_{new}(Y_t, \hat{Y_t})$$

  else
    Train the model $f(x)$ with cross entropy loss function: $L_{new}(\hat{Y_t}, Y_t)$
  Store the current trained model in the model buffer B
  Calculate Fisher information.



(a) MNIST and SVHN accuracy in My algorithm



(b) MNIST and SVHN forgetting in My algorithm

Learned things: Causal inference, knowledge distillation…

Stat NLP group intern, SUTD

Topic: NLP models and its application. Mentor: Wei Lu

Transformer

Get a deeper insight into seq2seq model such as LSTM, GRU especially Transformer. Read several recent papers about Transformer and implement it on machine translation task.

Code can be found through https://github.com/zzh-SJTU/data_pre-process_translation.

Interpretability on NN

Read several papers about interpretability of neural networks such as Neural Tangent Kernel. After that, I can better understand how neural networks can have a such powerful impact on everything.

Slides can be found through https://slides.com/zhehaozhang123/deck-0b14be

Medical image processing

Classification of kidney cancer

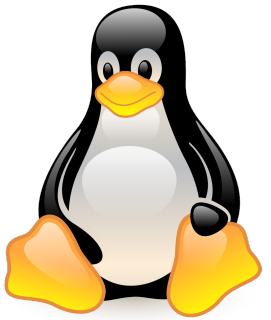
Using neural networks to classify different kinds of kidney cancer to help the doctor's diagnosis.
Deal with medical image with much higher resolution, much larger scale, more common pattern than traditional dataset in computer vision.
Start this project from image labelling to giving predictions.

04

Skills

Skills

- Programming Languages: Python, C/C++, MATLAB, linux shell (ranked by proficiency)
- Tools and Frameworks: Git, GitHub, LATEX, PyTorch, Numpy, Scikit-learn, OpenCV, huggingface, NLTK.
- English: overall band score 7.5



PyTorch

LaTeX

THANKS