2022 CS3612 机器学习

课程大作业

2022年3月24日

张洪鑫

目标和要求

- 每个人**独立**完成一个课题
- 结合课程所学知识，并且收集资料，分析总结现有工作，提出问题，设计模型，解决问题
- 不强调模型性能，重点在于**分析问题的过程**以及**设计实验**充分验证自己结论

内容

- Week.6-7:　选择课题
- Week.8-13：探究课题
- Week.14：
  - 制作一张**poster**
  - 撰写一篇**至少4页纸**的**mini paper(English)**
  - 并且**提供代码**(最好是notebook)
  - 截止时间**星期日(5.22) 24:00**
- 6月2日逸夫楼一楼或者理学院草坪展示，由评委对其打分；评分最高的三位同学颁奖

评分

- Informativeness（25%）知识性：你的论文给读者带来多少知识
- Scientific Novelty（25%）科学新颖性：你在科学方面有多少创新
- Attractiveness（25%）吸引程度：你的论文整体展示有多吸引人
- Writing（25%）可读性：论文写作是否规范，可以成为一篇学术论文的可能

Mini Paper

Format: https://nips.cc/Conferences/2022/PaperInformation/StyleFiles

- Abstract
- Introduction(Background,Motivation)
- Related work
- Method
- Experiments(baseline,ablation)
- Conclusion
- Reference

选题

- 选择10个课题之一
  - 每个课题都有具体任务和参考数据集
  - 设计新模型，在此任务下提升性能或者效率
  - 探究现有方法的局限性，设计实验，并提出解决方案
- 或自拟课题
  - 解决一个有价值或者有趣的问题
  - 对现有机器学习方法或现象提出一种解释（包括理论和经验解释）

1.Text Classification

Text classification is the task of assigning a sentence or document an appropriate category.

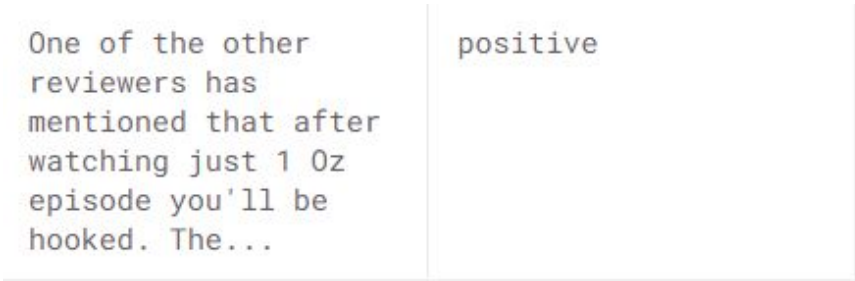
Text classification problems include **Sentiment Analysis**, **Topic Classification** and so on.

References:

https://paperswithcode.com/task/text-classification

https://paperswithcode.com/dataset/sst

https://paperswithcode.com/dataset/ag-news



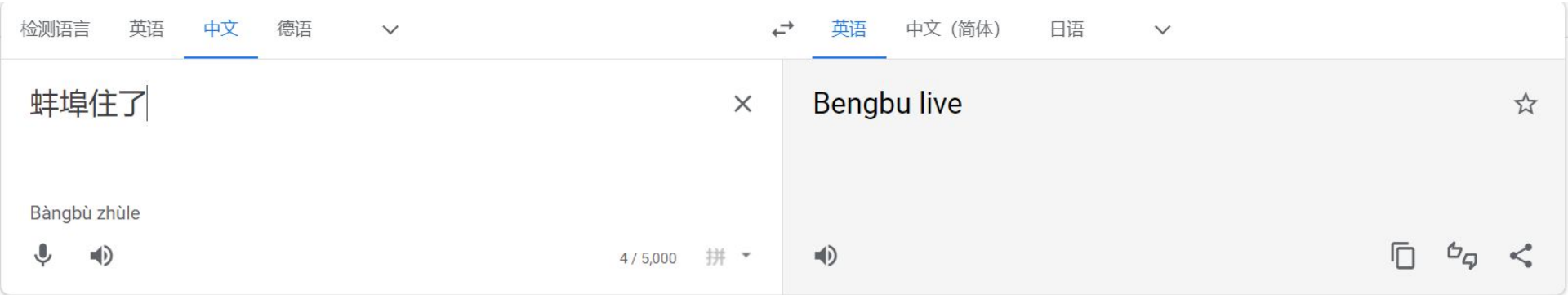
2.Machine Translation

Machine translation is the task of translating a sentence in a source language to a different target language.

References:

https://paperswithcode.com/task/machine-translation

https://paperswithcode.com/dataset/wmt-2014



3.Audio Classification

Audio is a sequence of digital signals sampled at a certain frequency.

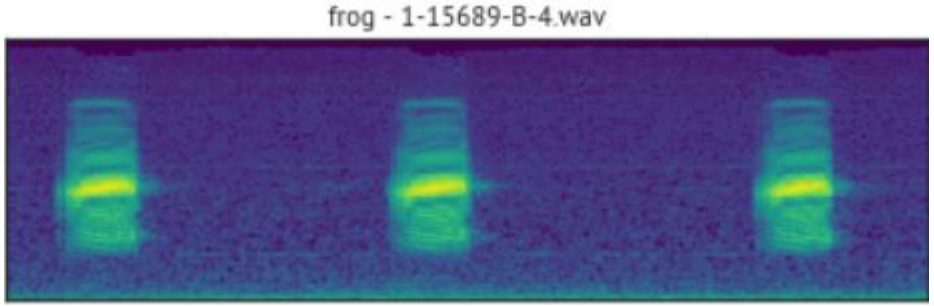
References:

https://paperswithcode.com/task/audio-classification

ESC-50

AudioSet

UrbanSound8K



4.CTR prediction

Click-through rate prediction is the task of predicting the likelihood that something on a website (such as an advertisement) will be clicked.

It’s a very valuable task in recommendation system.

References:

https://paperswithcode.com/task/click-through-rate-prediction

MovieLens 20M

avazu

Dianping

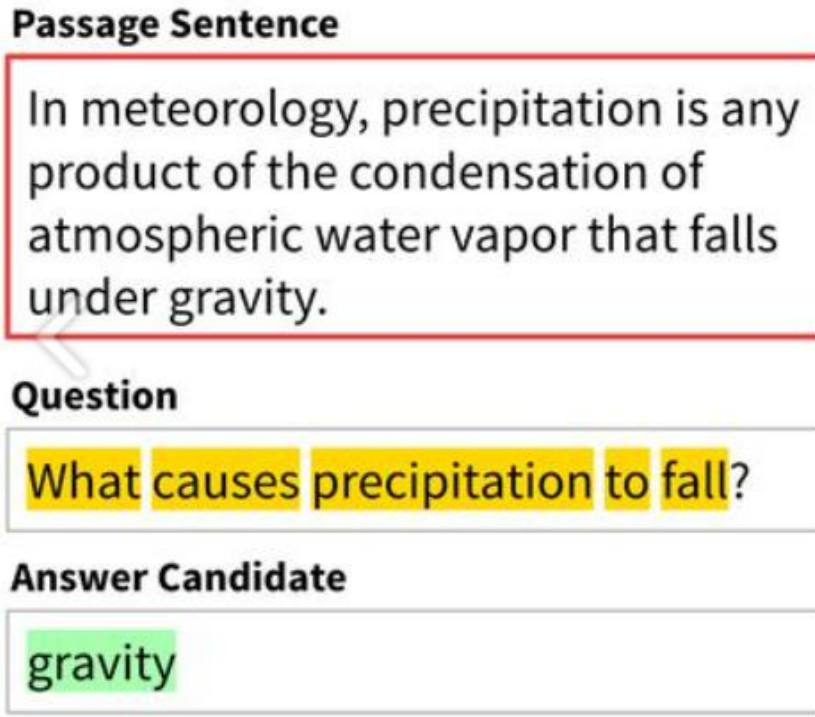
5.Question Answering

Question Answering is the task of answering questions (typically reading comprehension questions), but abstaining when presented with a question that cannot be answered based on the provided context.

References:

https://paperswithcode.com/task/few-shot-image-classification

https://paperswithcode.com/dataset/squad



6.Few-Shot Image Classification

Few-shot image classification is the task of doing image classification with only a few examples for each category.

References:

https://paperswithcode.com/task/few-shot-image-classification

Mini-Imagenet

tieredImageNet

CIFAR-FS

7.OOD Detection

Out of Distribution Detection: detecting instances that do not belong to the distribution the classifier has been trained on.

References:

https://paperswithcode.com/task/ood-detection

https://github.com/wzhouad/Contra-OOD

https://github.com/guyAmit/GLOD

8.Node Classification

Node classification models aim to predict non-existing node properties based on other node properties. Typical models used for node classification consists of a large family of graph neural networks.

References

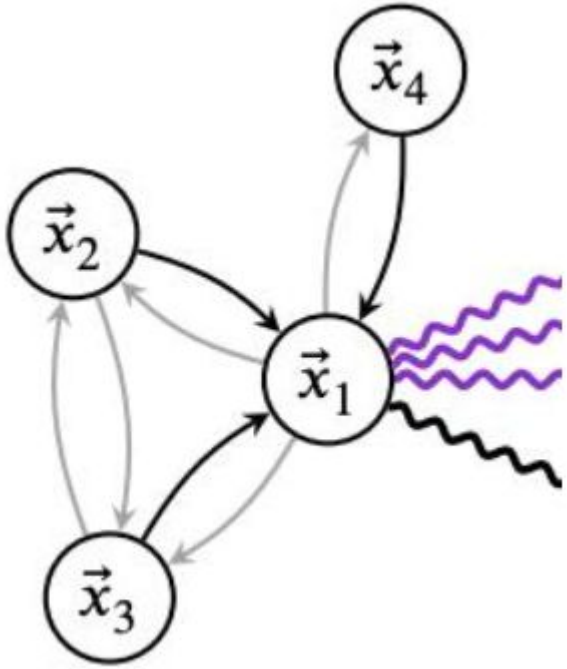
https://paperswithcode.com/task/node-classification

cora, citeseer, pubmed

Elliptic

ogbn-arxiv

DGL



9.Continual (Lifelong) Learning

Continual Learning is a concept to learn a model for a large number of tasks sequentially without forgetting knowledge obtained from the preceding tasks, where the data in the old tasks are not available any more during training new ones.

References

https://paperswithcode.com/task/continual-learning

http://yann.lecun.com/exdb/mnist/

https://paperswithcode.com/dataset/asc-til-19-tasks

10.Study of non-parametric models

Non-parametric models Including **SVM**, **KNN**, **Decision Tree** and **Gaussian Process** still matters, we should understand how they work and when to use them. Maybe even improve them.

You may study on

- SVM: different kernel functions' influence
- Decision Tree: ID3, C4.5, CART -> GBDT, XGBoost, LightGBM
- KNN: different distance functions' influence

References

https://archive.ics.uci.edu/ml/datasets/Mice+Protein+Expression

https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

https://archive.ics.uci.edu/ml/datasets/QSAR+oral+toxicity

http://yann.lecun.com/exdb/mnist/

Q&A

Thanks!