10/03/2023

Meeting

Talk about the problem

- Proposal structure
    - Introduction: to describe the problem, based on literature, optimization objectives
    - Experiment design: For time series data, to find out how to group data to training and testing
    - Take one, e.g., LSTM, and get some data to try the algorithm and understand it
    - Meeting Tuesday afternoon?

To do

- Get the proposal template
- Read papers, Aaron and Hui will send some readings on workload prediction
- Finish introduction: aim + objectives (e.g., 1) proposal a new ...algorithm, 2) to conduct experiments using real world data)
- Read >3 papers about workload predication, write a brief summary, and explain your understanding in the next meeting

21/03/2023

Done

- Read some papers

Meeting

- A paper using LSTM for prediction published in 2021
- Input: Google cluster of one month workload data, predict
- Explain how LSTM work.

To do

- Get a first draft of the proposal, in particular, to finish the overall aim and objectives
- To decide which algorithm to use
    - Transformer??? GPHH??
    - Baseline algorithms: HMM, SVR, Logistics regression, etc.
- Try any of the exiting approaches, LSTM, ARIMA, to test on the benchmark data
- To get datasets for testing, Google cluster
- To get familiar with transformer via online videos (2 – 3 weeks to learn)

- Meet next week at 4pm on Tuesday

29/03/2023

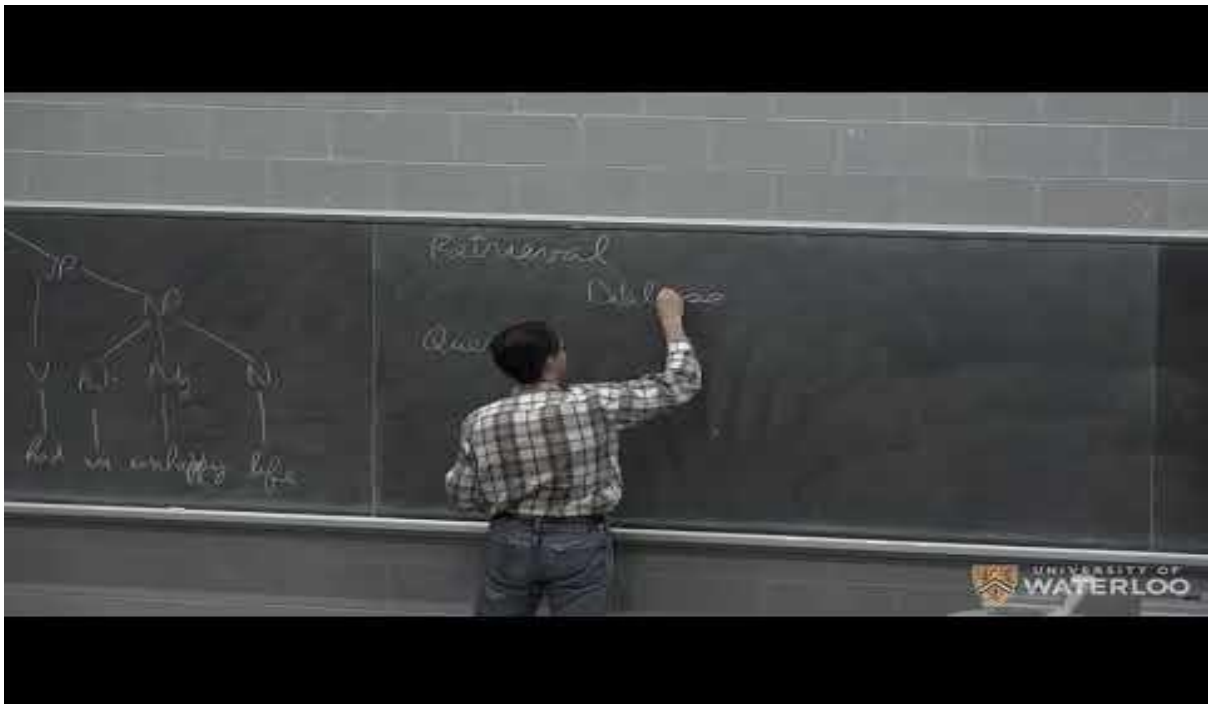
Done

- Learned transformers, like the concept
- Still need time to understand how it works
- Find code in Github, will use some datasets, that is similar to our problem, e.g., time serious data analysis, financial data analysis, weather data analysis.

Meeting

To complete the proposal, for the method part, provide some details of how we will design the transformer to do workload prediction.

CS480/680 Lecture 19: Attention and Transformer Networks



To do

Proposal due on Friday

Add a Gantt chart to schedule the project, important milestones to be included

References,

04/04/2023

- During the break, we meet on Monday 17 April at 7pm (NZ time). Before the meeting, we expect to experiment on some machine learning models such as LSTMs or transformer on some datasets.

- Before Monday 17 April, Joel will send a short report on the progress to us.
- Read three papers and write short paper summaries (update the related work chapter of the preliminary report).

17/04/2023

Not attending

25/04/2023

Done

- Have sent emails to people who use LSTM for workload prediction
- Implement a LSTM for workload prediction, use Alibaba datasets, evaluate accuracy MSE,

Meeting

- 70% data for training and 30% for testing

To do

- Try different parameters for LSTM, and check existing settings of some existing work
- Check how existing work design training and testing datasets and chose one with justifications
- Check what are the important matrix used in the literature and choose at least two for showing the performance.
- Present the results of the prediction
- Work on related section

2/05/2023

Done

- A prediction of LSTM on the Alibaba data
- The results are very noisy

Meeting

- Use another dataset which are more predictable, e.g., Google one
- Alibaba data after summing up it, find reference that also do the same pre-processing

To do

- Continue with our writing, include the discussion of the baseline and others, and which show that choosing LSTM [?] is justified.
- Record what has changed, and why?

- To start implement transformer, and test on the same dataset, Google, to get result by the end of week 9
- Use the preliminary report, which chapters

9/05/2023

Done

- Download data from Google, but it is still noisy
- Get some results from LSTM, but is not very good,

Meeting

Not yet start implementing transformer

To do

- To get the dataset ready
- To implement transformer and get some results
- An old paper: https://ieeexplore.ieee.org/abstract/document/6212065
- Meet at 4pm on Tuesday

16/05/2023

Done

- Read the paper
- Discovered that the prediction was only for one machine, and find a paper, which use 7 days workload (Google cluster) to predict the next 10, 20, 30 minutes workload.

meeting

- Is currently go through and clean up the dataset, could not find prepared data

To do

- LISM in the paper, and regenerate the results

- Keep eyes on if can find the data used by papers about workload prediction in cloud
- Try the transformer from GitHub and reproduce the result.

- Preliminary report **2 June**, a complete draft by 26 May
- Start writing the preliminary report, using the template
- Meet at 4pm on Tuesday

13/06/2023

No meeting

20/06/2023

No meeting

27/06/2023

Find code in Github predict spatial temporal data

To do

Look for more Transformer, and get paper regarding how to implement.

Space time transformer: https://github.com/QData/spacetimeformer Get one next one to try on the datasets

20/06 2023

no meeting

27/06/2023

No meeting:

4/07/2023

No meeting

11/07/2023

Done

- Downloaded a transformer in a github
https://github.com/gzerveas/mvts_transformer
- A general transformer, tried on the data on the github

Meeting

- The transformer used ts file, can take csv file
- May need to change architecture to suit the file format
-

To do

- Comparing with LSTM results
- May need to adapt the architecture to improve the performance
- Find one or more datasets to test, e.g., Alibaba datasets to test
- Think about what will be included in the method chapter, any new design to make transformer effective for workload predication
- Explain the paper in the next meeting

- Start to use the template of final report, and transform some content from the preliminary report

25/07/2023

Done

- Read a paper https://arxiv.org/pdf/2010.02803.pdf

Meeting

- Use sliding window for the transformer
- May change the mask used in the paper, and use sliding window

To do

- To start working on implementing transformer for workload prediction
- Compare with LSTM, implemented

- Making progress on the report writing.
- Meet 4pm on next Tuesday

1/08/2023

Done

- Try to get transformer work, feel like it will work soon

Meeting

.ts format, and transfer to csv file

To do

- Try to get transformer work

- Start working on the report, chapter 2-related work. (Chapter 3- method, Chapter 4-results, Chapter 5 conclusion)

- Explain how existing work made masks
- Plan:

    - Week 4: make transformer work and generate some initial results
    - Week 5-6: improving the existing transformer, by adding new mask, or using sliding window.
    - Week 7-8: continue to improve method and complete chapter 3 + 4

    - Week 9: chapter 1 + chapter 2
    - Week 10: complete a draft report

8/08/2023

Done

- Transformer is working, waiting to complete and get results

Meeting

- Start working on the final report using the template

To do

- To get results and compare with LSTM, in terms MSE
- Feedback from examiners

Overall it's OK. There is no introduction on the existing prediction methods for cloud workload and the technical challenges in this project. The motivations can be more specific. Some motivations need to be further clarified, e.g., "If cloud workflow accuracy reaches a high enough accuracy, we will be able to avoid under-provisioning resources to users." Why prediction can avoid under-provisioning? This would need another smart step of decision making and optimisation.

Introduction emphasis the importance of workload prediction. it would be good also mentioned existing work on workload prediction and their limitations. The the aim of the project is clear stated. Some discussions of how to achieve the aim would be good.

The background literature is very limited. Only two specific time-series analysis methods are reviewed. No general big picture is given. No existing methods for predicting cloud workload is reviewed and discussed.

Can the cloud workload prediction be directly treated as a time-series analysis be solved by any existing time-series analysis method? If so, what will be the challenges?

From the code in Gitlab, the development that has been done is not much. The data preprocessing looks not much (100 lines). The LSTM is not much either - most of the file is about I/O, and the LSTM was just called through one line of function. It doesn't look like a considerable development/implementation.

There are some critical thinking on the advantages of Transformers and Attention, but not much other than that.

critical discussions of existing approaches shows critical thinking. More discussions about the proposal of using transformer for workload prediction would be good.

Writing is OK - there are some errors in the writing, e.g., dependenciesli2019enhancing on page 5.

It is highly encouraged to provide deep analysis on the issues/challenges of cloud workload prediction vs general time-series analysis. Is it as simple as applying the existing time series methods, e.g., transformers, to another time series dataset? Why should you consider Transformers in particular? Is it suggested by your supervisors or you have your own valid reasons? Why only compare with LSTM? How about other time-series analysis methods?

So far the development is too simple - the current core code is just a line of calling a Python library for LSTM. I don't think it would be enough for a 489 project if the implementation of Transformers is also a line of calling. You are suggested to consider carefully what will be your main contributions in this project, and how to show your efforts in this project.

15/08/2023

Done

- Transformer is implemented and compared with LSTM

Meeting

- Hui and Aaron will be away (22 Aug – 2 September) resume meeting on 5 September

Method	LSTM	Transformer	Slide_Transfomer (proposed)
Instance 1 (10 time units)			
…			
Instance 3 (30 time unites)			

To do

- Design datasets with different sampling scales, for training and for testing, and update the report on the Evaluation chapter.
- Improve transformer with sliding windows, clearly explain what the difference from the others.
- Other ideas for improving transformer, e.g., position encoding, number of layers, the size of each layers, (may use standard loss function, MSE),

- Comparing MSE on tables
- Read and summarize at least five papers on the same problem.
- One section on Related work focusing on sliding window for WPC and for machine learning problems.
- Having code about data preparation modeled and document, to be read to submit,
- **Describe in the report how datasets (training, validation and testing), refer to the existing work on how to describe.**

- May try different **sittings (size)** and see how it affect the MSE, study the tradeoff of the settings (sensitive analysis), can put before the main results.

5/09/2023

Done

- Make transformer work
- Set up sliding window of 10 time unites to produce one time unit workload, may add one more time unite for predication

Meeting

- Need to submit an abstract by this Friday, 8 September.

To do

- To complete the abstract on overleaf and let Hui and Aaron know via Zoom
- Get a result for the current implementation and setting, and compare with LSTM
- Writing: related work -> sliding window; architecture (diagram), algorithm pseudocode (?) with highlights on new components
- Check the to do list from previous meeting and complete if have time

12/09/2023

Done

- Can use transformer to do predicting data

- Discover that using 50% of the original datasets can even get better results (accuracy)

Meeting

- To improve transformer by using feature selection
- Read a paper Arima using feature selection, would like to adapt to Transformer

To do

- Change LSTM to predict workload
- Complete transformer and compare Accuray (MSE) and show figures of the comparing algorithms.
- To design sensitivity evaluation
- If we change the number of layers (3 at the moment and would like to try 4), we need to justify why it is helpful.

19/09/2023

Done

- Implemented LSTM
- Implemented feature selection for transformer, faster to train, and results are good

Meeting

- Stop training because of overfitting, using verification datasets

- Overfitting is when verification is used,
- use early stopping,
- Feature selections are normally used for classification, not much work on time series data prediction.

To do

- To adjust some parameters, number of layers, to improve MSE

- More coding in the second half of the project
- To change the number of layers, and compare the results of the selected number of layers
- Datasets are used by exiting work, and we can directly compare the results.
- A paper: https://ieeexplore.ieee.org/document/9209730

26/09/2023

Done

- Implemented time window for prediction

Meeting

- What is the baseline method for early stopping (to avoid overfitting)

- If the model is good, reducing validation may improve the efficiency of training
- Balance between efficiency and quality of the model

To do

- To compare the new propose methods with some existing methods, e.g., LSTM, using tables on different evaluation data sets
- Complete draft by the end of week 11 the least.

3/10/2023

Done

- Parameter training
- Sliding window

Meeting

- Redraw module diagram for clarity.
- Related works should contain previous models used for the same problem, in the same context
- Background should be general content on the models. Such as what is an LSTM? And more
- Change order of report,

    - For chaper 3: do propose method, framework, data preparation then model/sliding window

    - Chaper 4 should be settings in the experiments

To do

- Finish report
- Submit artefacts on Friday

10/10/2023

Done

- Related works and part of methodology

Meeting

- Finish off report and get evaluation part complete

To do

- FInish off report