Australian National University

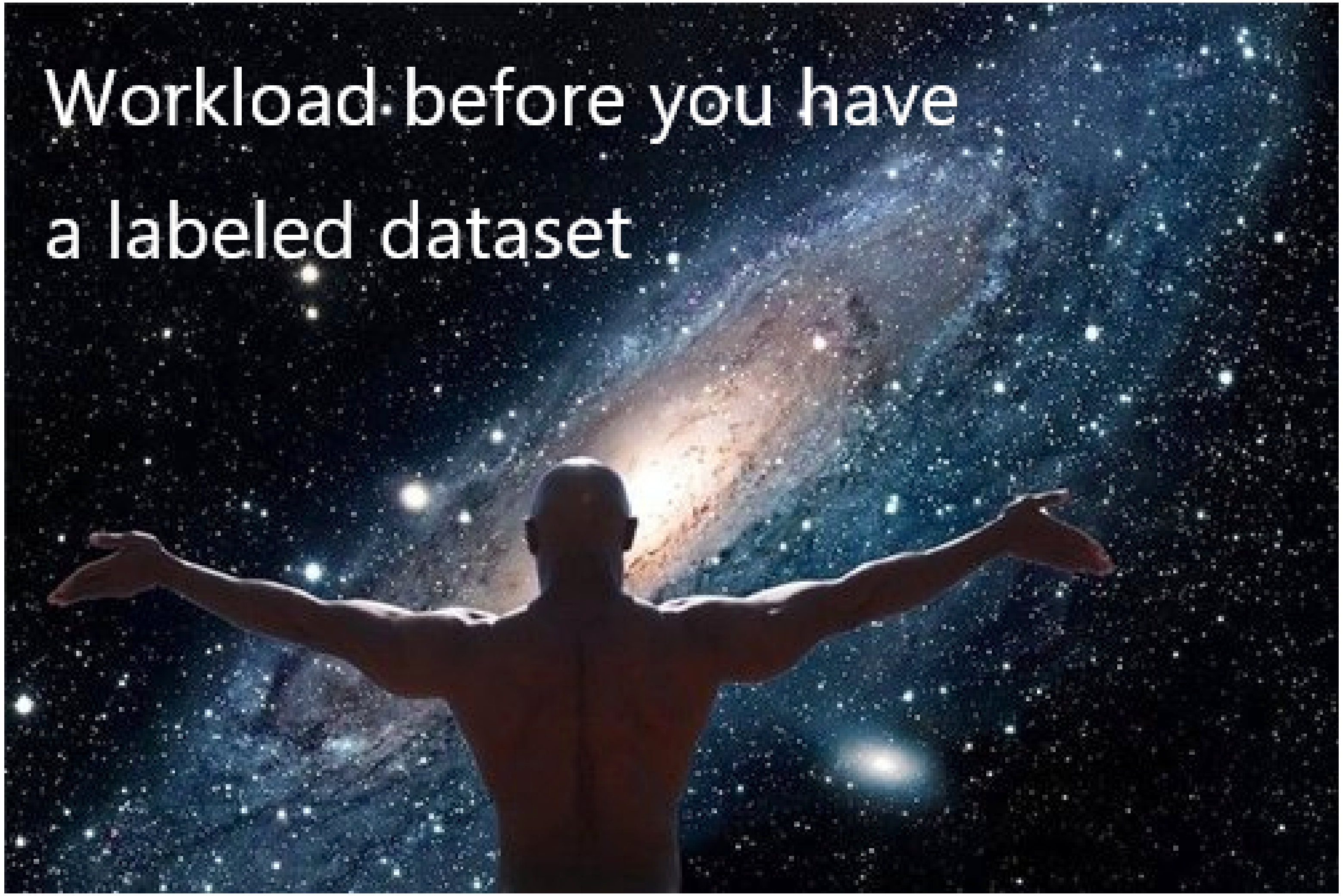
Labeling Words, Sentences, or Documents?

—— A Comparison of Sequence Labeling and Text Classification for Cohort Selection for Clinical Trials

Haozhan Sun

Supervisor: Assoc. Prof. Hanna Suominen

Motivation



The Problem

Want to reduce human-labor cost

↓

How?

↓

Want best annotation granularity

↙ ↘

Best sequence labeling

Best text classification

ID: *14*, **TYPE:** *cardiovascular*
PROFILE:

Name: *Ken Harris*
Age: *71 years*
Admission story: *Ken is suffering from arrhythmia for the first time in his life. He is feeling pretty sick but this does not seem to be too serious.*
In-patient time: *He has been at the ward for three days.*
Familiarity: *Both you and the next nurse have looked after him earlier.*

SPOKEN, FREE-FORM TEXT DOCUMENT:
WAV file (93 words, 48 seconds, 4.25 MB)
The first author typed a transcription only for this spoken document
On a bed three is Ken Harris, 71 years old under Dr Gregor. He came in with arrhythmia. He complained of chest pain this morning and ECG was and was reviewed by the team. He was given some anginine and morphine for the pain and he is still tachycardic and new meds have been ordered in the medchart. Still for pulse checks for one full minute. Still awaiting for echo this afternoon. His blood pressure is just normal though he is scoring MEWS of three for the tachycardia. Otherwise he still for monitoring.

WRITTEN, FREE-FORM TEXT DOCUMENT:
Ken harris, bed three, 71 yrs old under Dr Gregor, came in with arrhythmia. He complained of chest pain this am and ECG was done and was reviewed by the team. He was given some anginine and morphine for the pain. Still tachycardic and new meds have been ordered in the medchart. still for pulse checks for one full minute. Still awaiting echo this afternoon. His BP is just normal though he is scoring MEWS of 3 for the tachycardia. He is still for monitoring.

WRITTEN, STRUCTURED DOCUMENT:
Ken[1] harris[2], bed three[5], 71 yrs old[3] under Dr Gregor[6.1], came in with arrhythmia[7]. He[4] complained of chest pain[1] this am and ECG[2] was done[1] and was reviewed by the team[1]. He was given some anginine[1] and morphine for the pain[1]. Still tachycardic[2] and new meds[1] have been ordered in the medchart. still for pulse checks for one full minute[1]. Still awaiting echo[2] this afternoon[3]. His BP is just normal[2] though he is scoring MEWS of 3 for the tachycardia[2]. He is still for monitoring[1].

PATIENT INTRODUCTION:
1. GivenNames/Initials: *Ken*
2. LastName: *harris*
3. AgeInYears: *71 yrs old*
4. Gender: *He*
5. CurrentBed: *bed three*
6. UnderDr: 6.1. LastName: *Dr Gregor*
7. AdmissionReason/Diagnosis: *arrhythmia*

MY SHIFT:
1. Status: *chest pain*
2. OtherObservation: *tachycardic; BP is just normal; scoring MEWS of 3 for the tachycardia*

APPOINTMENTS:
1. Status: *was done; was reviewed by the team*
2. Description: *ECG; echo*
3. Time: *this afternoon*

MEDICATION:
1. Medicine: *anginine; morphine for the pain; new meds*

FUTURE CARE:
1. Goal/TaskToBeCompleted/ExpectedOutcome: *for pulse checks for one full minute; still for monitoring*

How it works?

Raw clinical record → Annotated clinical record → Train → Auto illness classification

↑ Sequence Labeling

↑ Text Classification

What to do:

1. Find the model with best performance in each step
2. Find the best tricks to fine-tune the model to the task

How to achieve?

Dataset

Cohort Selection for Clinical Trials of the 2020 N2C2 Shared-Tasks:

Predict whether or not a patient meet a list of selection criteria given the patient’s clinical records

A Kaggle challenge this year!

Sequence Labeling - Flair

"George Washington was born."

→

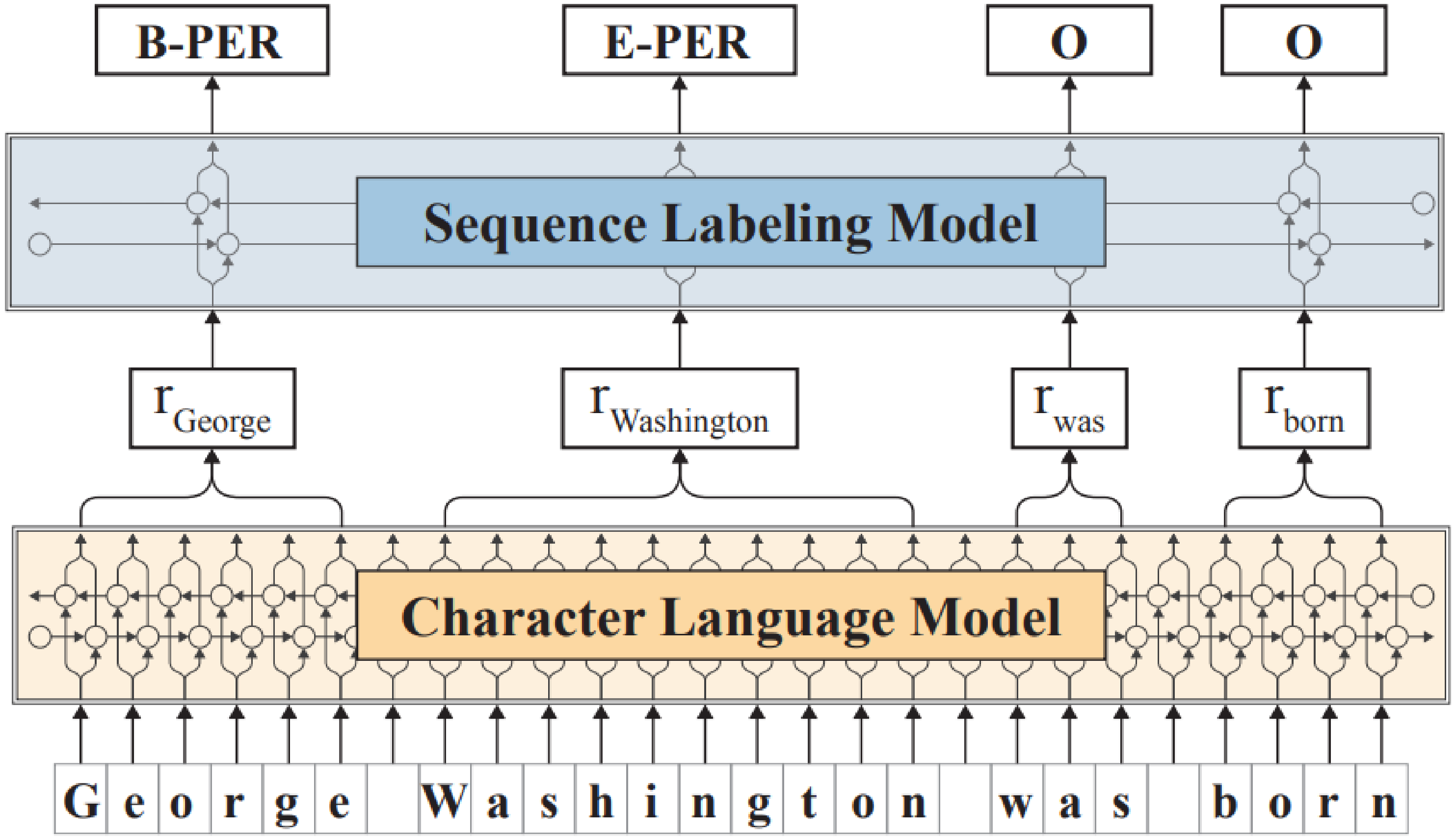


Figure 1. Akbik et al., 2019

Text Classification - ULMFiT

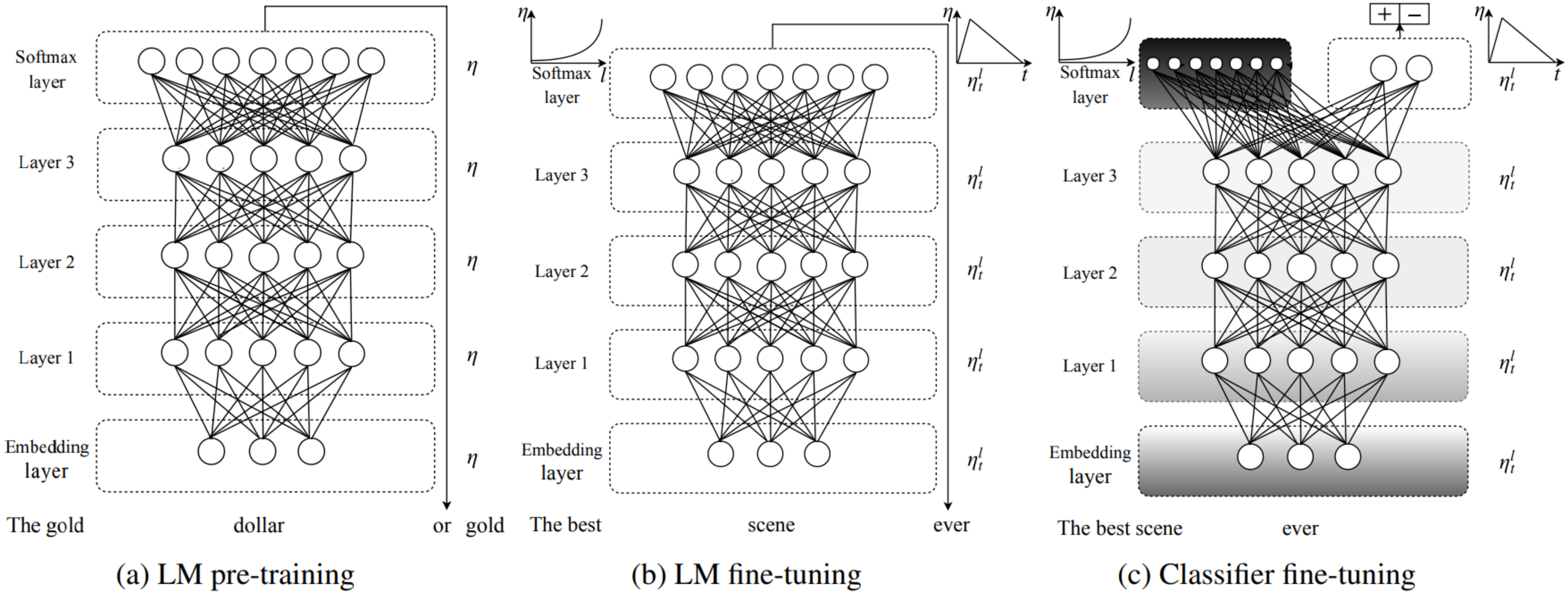


Figure 2. Howard, J. and Ruder, S., 2018

Evaluation

```
He <PatientIntroduction_Gender> is <NA> on <NA> oxygen <Medication_Medicine> 2l/nasal <Medication_Dosage> prongs <
Medication_Dosage> . <NA>
He <NA> is <NA> on <NA> oxygen <MyShift_OtherObservation> 2l/nasal <MyShift_OtherObservation> prongs <MyShift_Othe
rObservation> . <NA>
He <NA> is <NA> on <NA> oxygen <MyShift_OtherObservation> 2l/nasal <MyShift_OtherObservation> prongs <MyShift_Othe
rObservation> . <NA>
```

$$Prec_c = \frac{TP_c}{TP_c + FP_c}, \quad and \quad Rec_c = \frac{TP_c}{TP_c + FN_c},$$

$$F1_c = \frac{2Prec_c * Rec_c}{Prec_c + Rec_c}.$$

Statistical Significance Test:

$$\delta(X) = M(A, X) - M(B, X) \qquad H_0 : \delta(X) \leq 0, \quad H_a : \delta(X) > 0$$

Future Work

1. Hyper-parameter choice and language model fine-tuning tricks
2. Text classification analysis
3. Statistical significance test
4. Guideline for future researchers

Australian National University

Thank you for listening!

Q&A

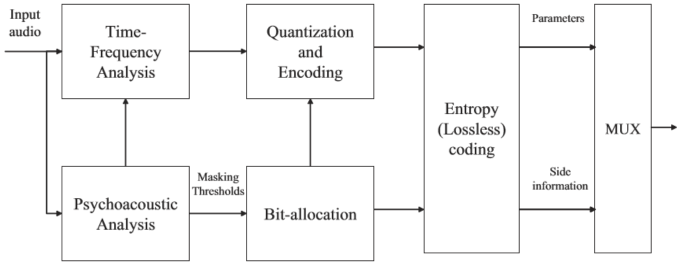
Australian National University

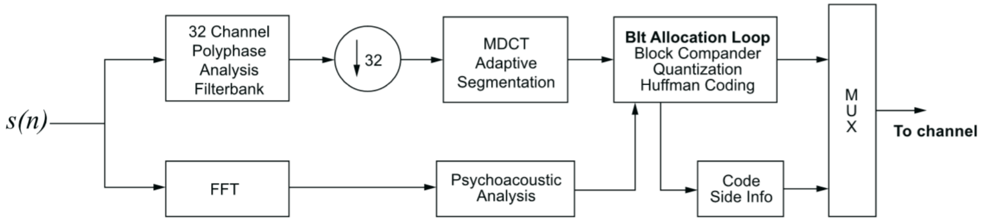
Using the Neural Networks to recover the removed data of lossy compression music

U6531709, Kunpeng Chen

Supervisor: Charles Martin

Lossless vs Lossy compression





Lossless compression
Thiagarajan and Spanias (1996)

Lossy compression (MP3)
Thiagarajan and Spanias (1996)

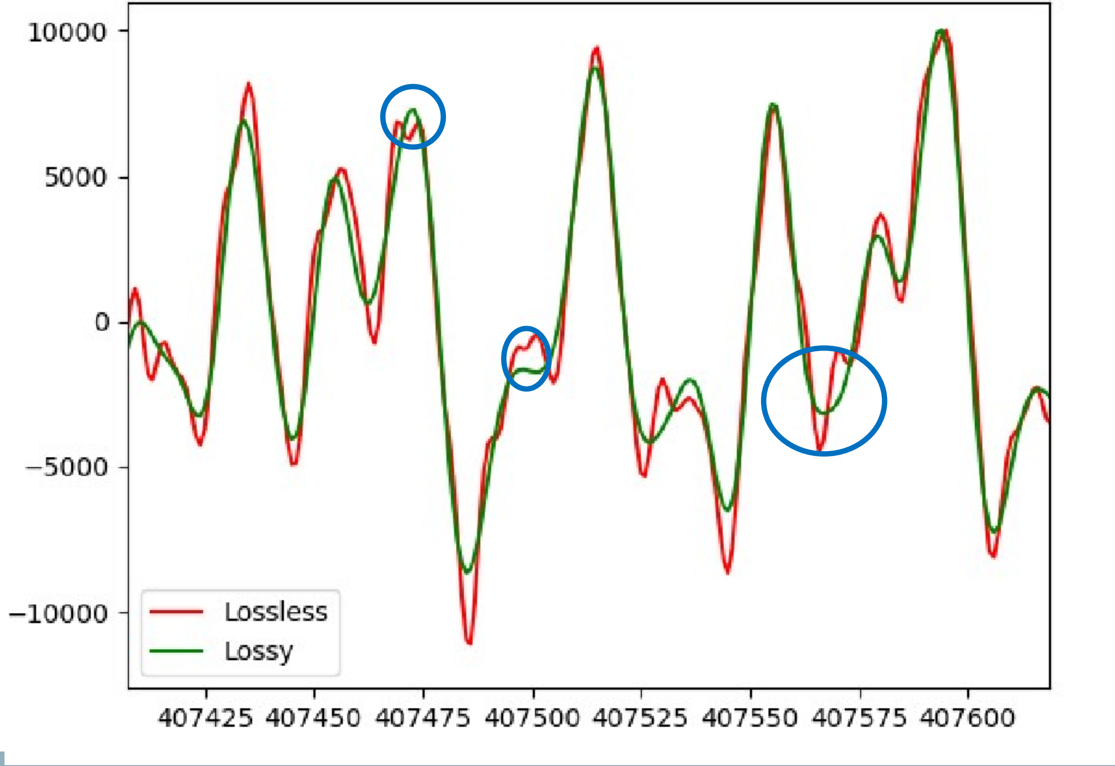
Dataset

- “**MAESTRO** (MIDI and Audio Edited for Synchronous TRacks and Organization) is a dataset composed of over 200 hours of virtuosic piano performances captured with fine alignment (~3 ms) between note labels and audio waveforms.” (Hawthorne, Stasyuk, Roberts, Simon, Huang, Dieleman, Elsen, Engel, and Eck , 2019).

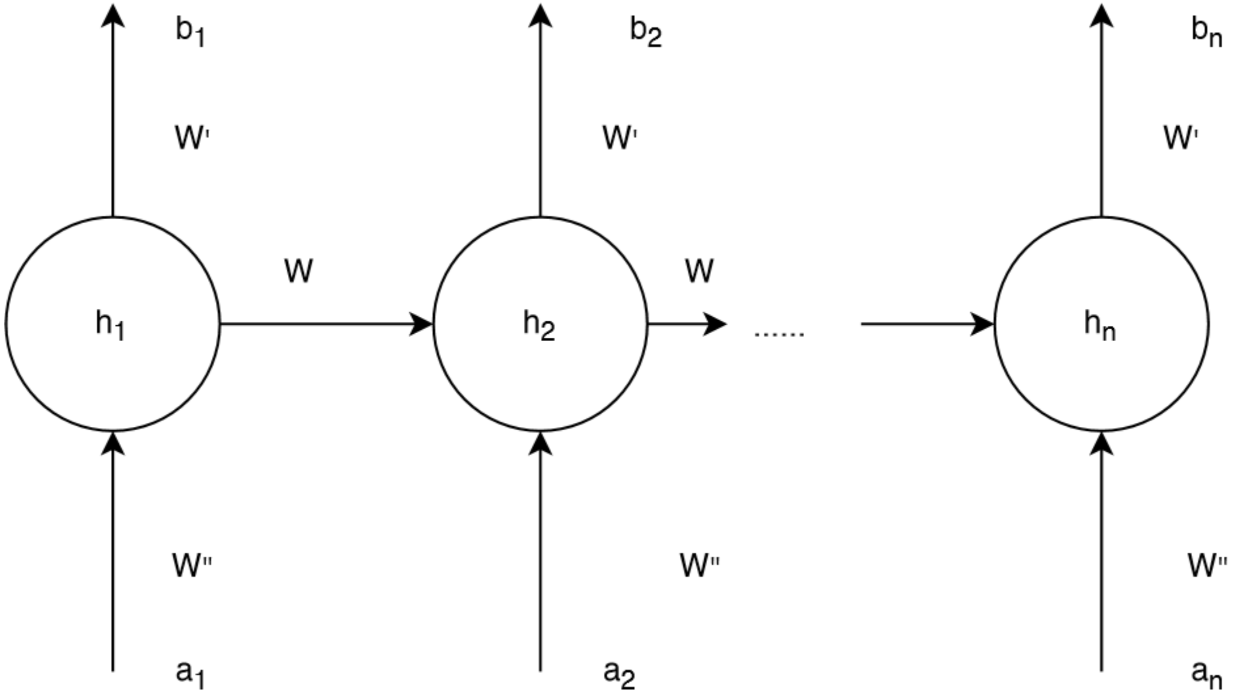
Frame’s corresponding?

- WAV (Lossless): 319 MB
- VS
- 8Kbps MP3 (Lossy): 6.65 MB
- **Solution:**
- MP3 -> WAV

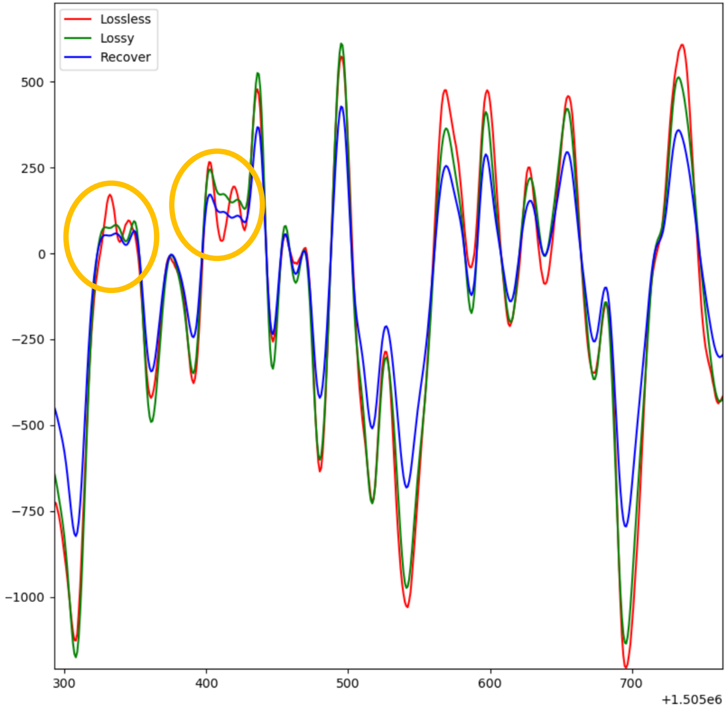
Real Problem



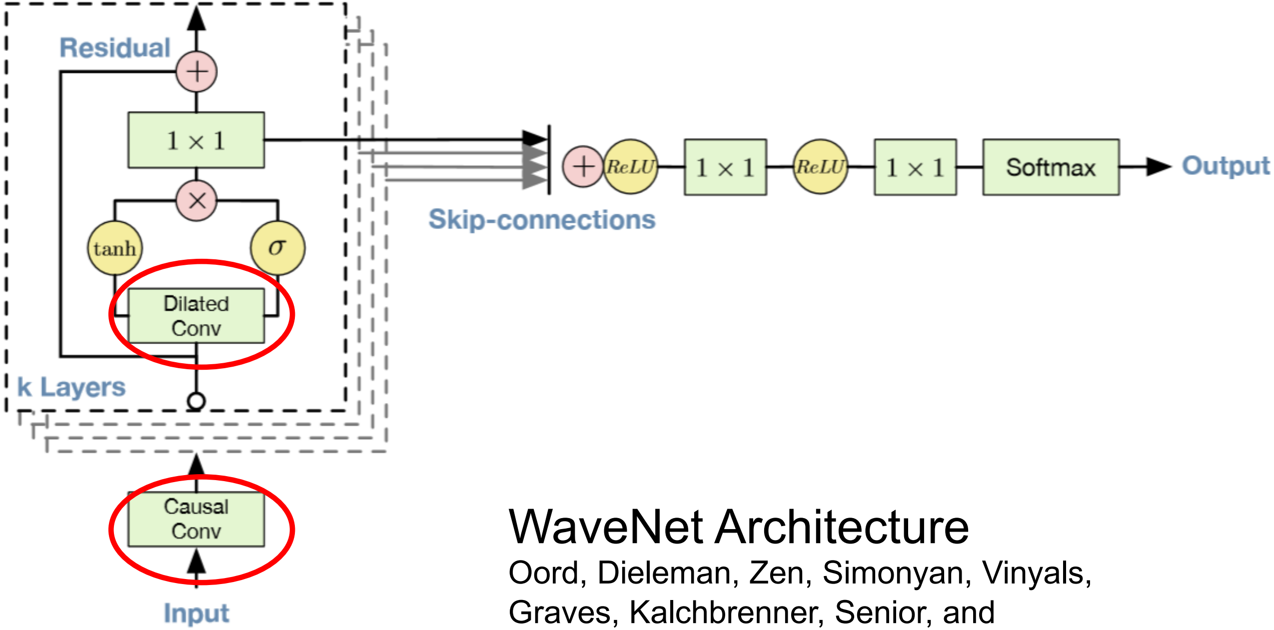
RNN



Result -- Overfitting



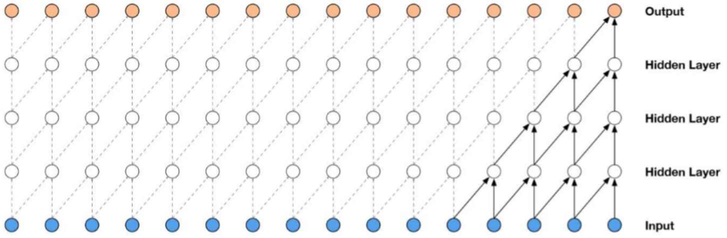
WaveNet

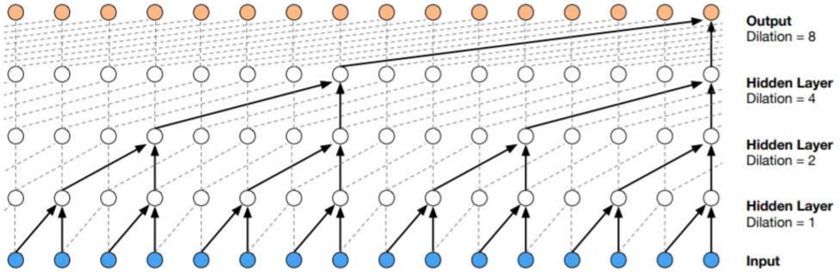


WaveNet Architecture

Oord, Dieleman, Zen, Simonyan, Vinyals, Graves, Kalchbrenner, Senior, and Kavukcuoglu (2016)

WaveNet’s main features





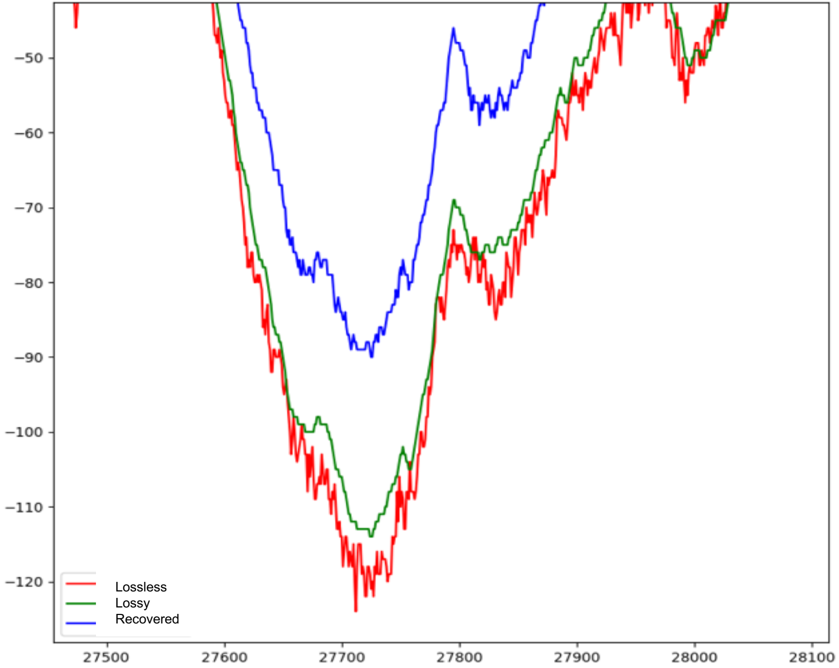
Causal convolutional layers

retrieved from Oord, Dieleman, Zen, Simonyan, Vinyals, Graves, Kalchbrenner, Senior, and Kavukcuoglu [2016]

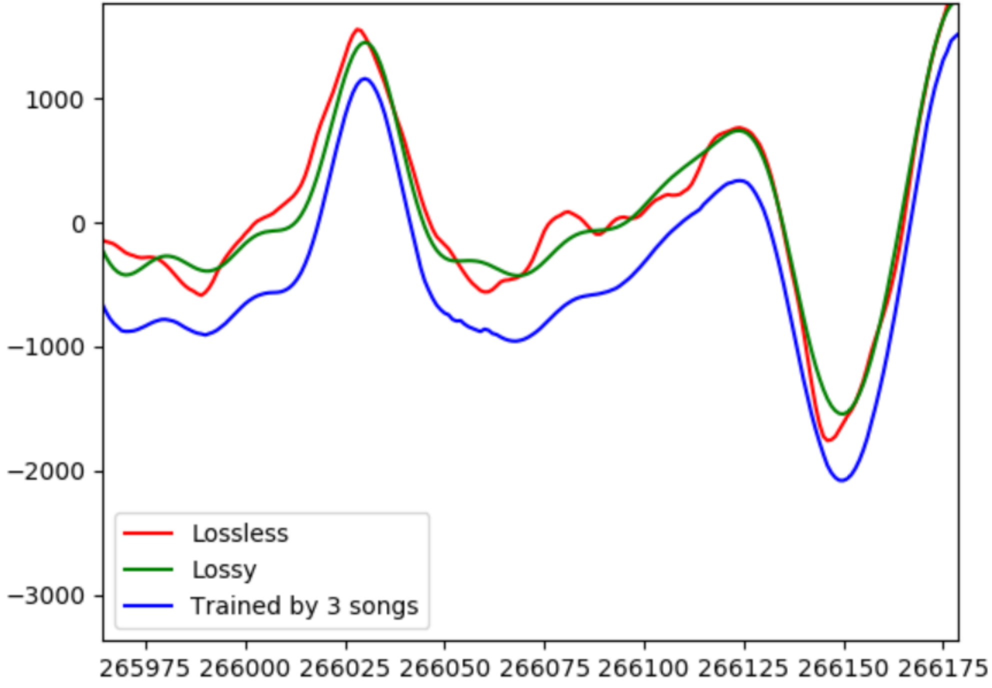
Dilated causal convolutional layers

Retrieved from Oord, Dieleman, Zen, Simonyan, Vinyals, Graves, Kalchbrenner, Senior, and Kavukcuoglu [2016]

Result -- Overfitting

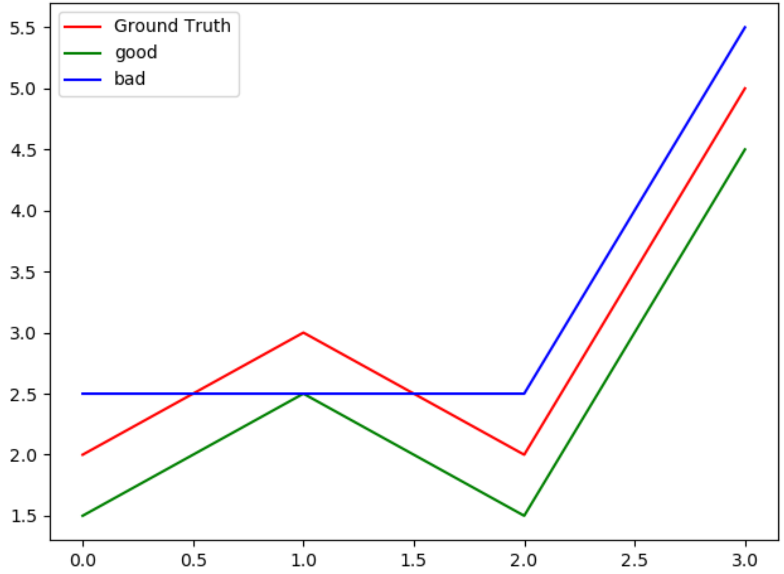


Result – Predict different song



Loss Function Issue

- MSE cannot represent audio quality



Future Work

- 1. Find more data
- 2. New Loss function.
- 3. Thesis

Reference

- Hawthorne, C.; Stasyuk, A.; Roberts, A.; Simon, I.; Huang, C.-Z.A.; Dieleman, S.; Elsen, E.; Engel, J.; and Eck, D., 2019. Enabling factorized piano music modeling and generation with the MAESTRO dataset. In *International Conference on Learning Representations*. https://openreview.net/forum?id=r1lYRjC9F7.
- Oord, A. v. d.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; and Kavukcuoglu, K., 2016. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, (2016).
- Thiagarajan, J. J. and Spanias, A., 1996. *Analysis of the MPEG-1 Layer III (MP3) Algorithm Using MATLAB*. Morgan&ClaypoolPublishers, SanRafael, Calif.

Learning-based Attributed Graph Matching

Yanxi Lu

Supervisor: Dr. Qing Wang

Outline

- Background
  - Problem Definition
  - Existing methods and motivation
- Methodology
  - Overview
  - Computing node embedding
    - Graph Neural Network
  - Matching
    - bi-directional LSTM
- Experimental Results
- Future Work

Background

- Attributed graphs

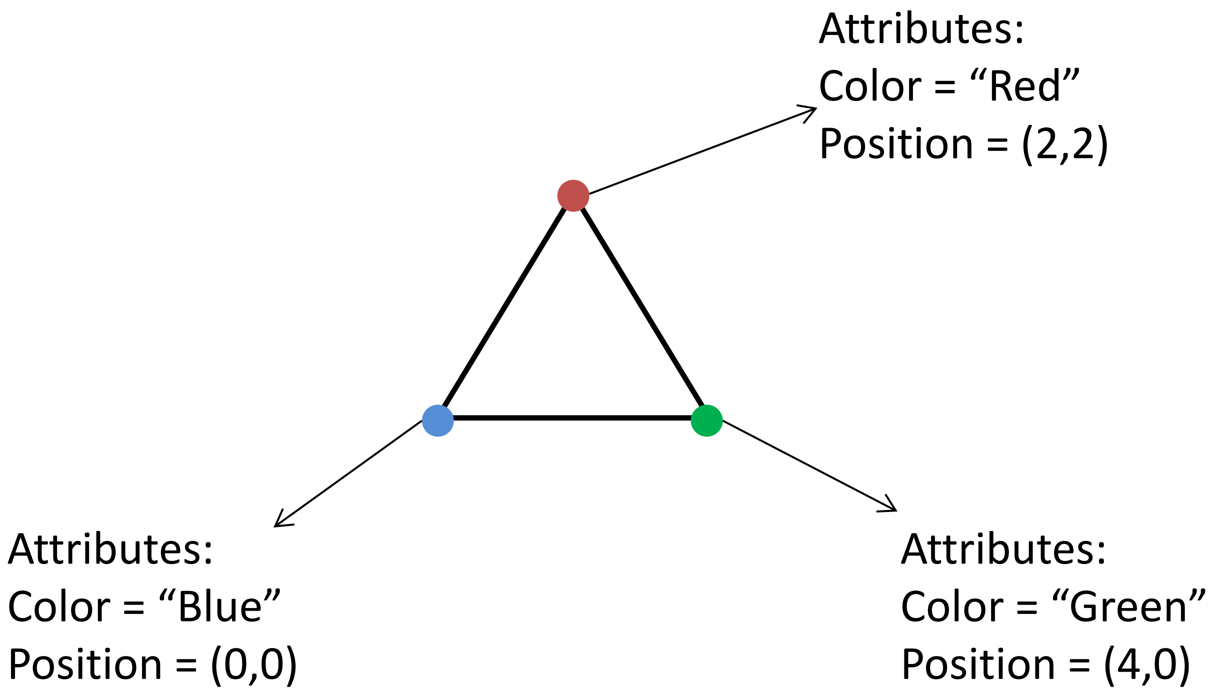


Figure 1: Example of Attributted Graph

Background

- Attributed graph Matching

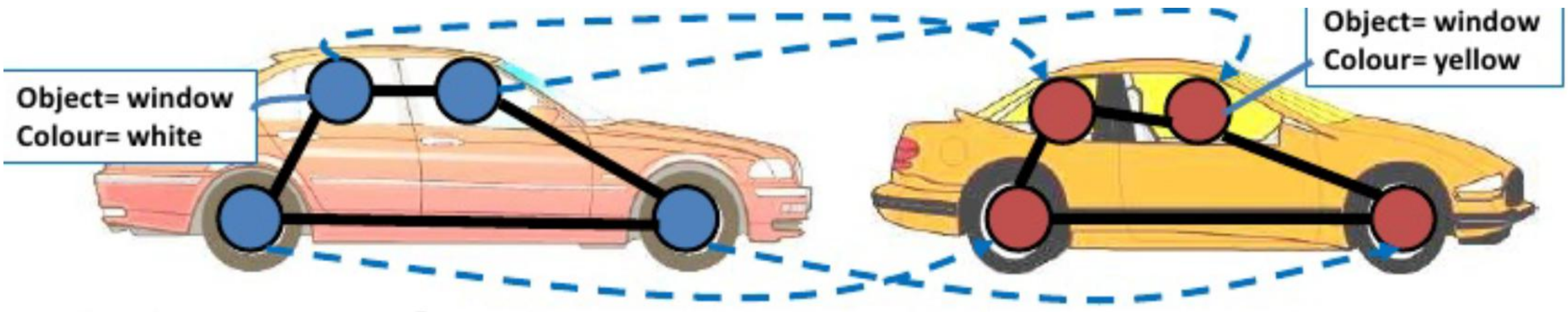


Figure 2: Example of graph matching[1]. Both the structure of the car(denoted by the structure of nodes) and the attributes are matched.

Problem Definition

- Given two attributed graphs G and G', for any connected query subgraph g of G, find top-k matching **connected** subgraphs in G'.

- Real-life Applications
  - Computer Vision
    - 2D & 3D Image analysis
    - Object Recognition
  - Graph Database Indexing and Retrieval
    - Biometric Indentification
    - Recommender System
  - NLP
    - Document matching
  - Many others

Why learning?

- NP problem - Computationally expensive, requires approximation

- Common Mathematical approach[2]

  Step 1: Compute pairwise distance(similarity) between nodes in both graph to give a distance matrix. The distance is computed using a predefined measure.

  Step 2: Match nodes using approximation algorithms

- Shortcoming of the approach above:

  - time & space complexity
  - handcrafted distance measure

Existing Learning Approaches

- Graph Neural Network(GNN)

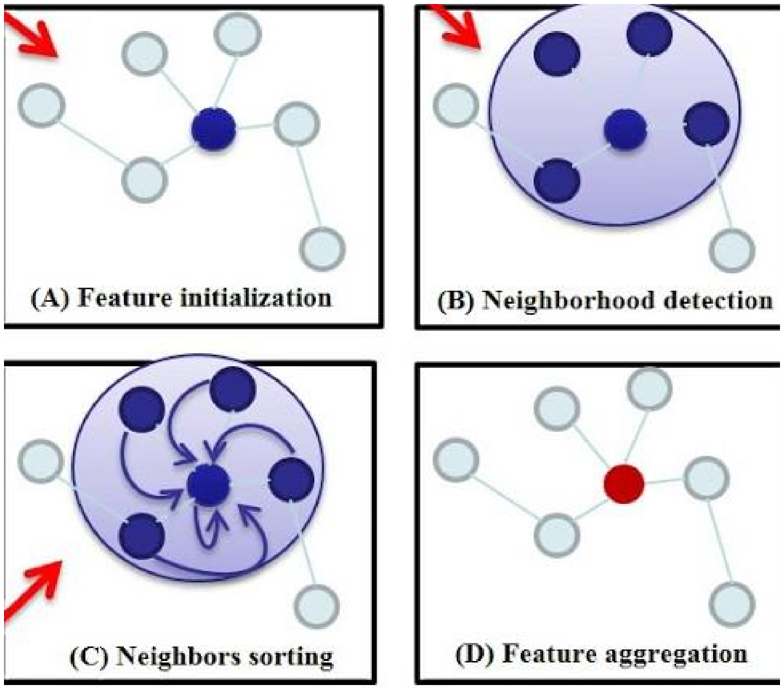


Figure 3: An illustration of 1-degree (hop) graph neural network[3]. The network works based on the principle of message passing.

Existing Learning Approaches

- GNN for graph matching

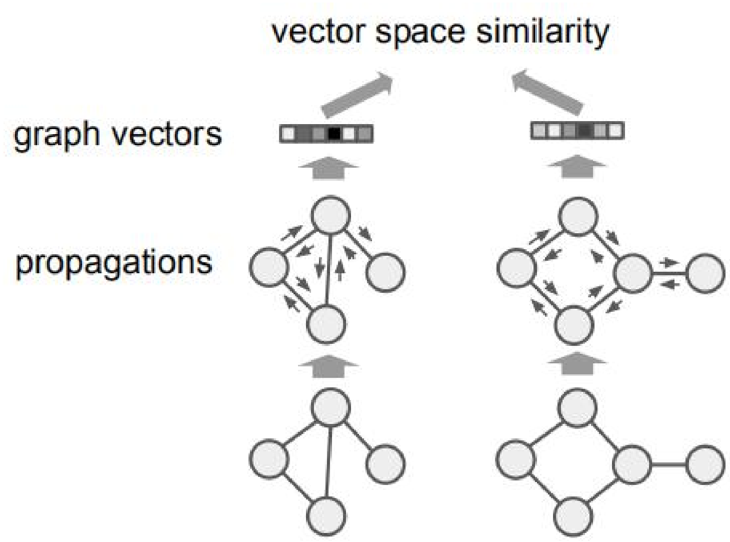


Figure 4: Graph matching using GNN[4]. GNN is used to process each graph to generate a graph level embedding, then the two embeddings are compared to compute the similarity.

Existing Learning Approaches

- Graph Matching network

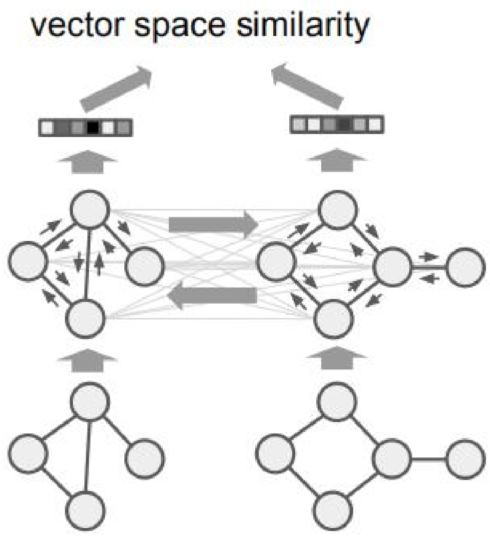


Figure 5: Graph matching network[4]. Uses attention mechanism to learn cross-graph correlation when training.

Existing Approaches - Problems

- Similarity based on graph-level representations
- Does not specify which part matches and on what basis do the two graphs match
- Generally focused on matching the attributes
  - possible to capture more structural information with deeper GNNs, however, exponentially computationally expensive
- Does not actively search for matching subgraphs, more focused on computing the similarity of given graph
- Searching for nodes based on closest embeddings often gives unconnected graphs

My Approach

- Two parts in the model
  - Embedding part
  - Matching part

- General idea: first generate node embeddings, then learn to matching the subgraphs(not individual nodes)

Synthetic Dataset

- Randomly generate 150 pairs of matching subgraphs of size 5 with attributes Color $\in$ {R, G, B}
- Matching based on attributes/structure or both
- Aggregate the first subgraph of each pair to form G
- Aggregate the second subgraph of each pair to form G'
- 4:1:1 for training data, validation data and testing data

Australian National University

Embedding part

- Use GNN

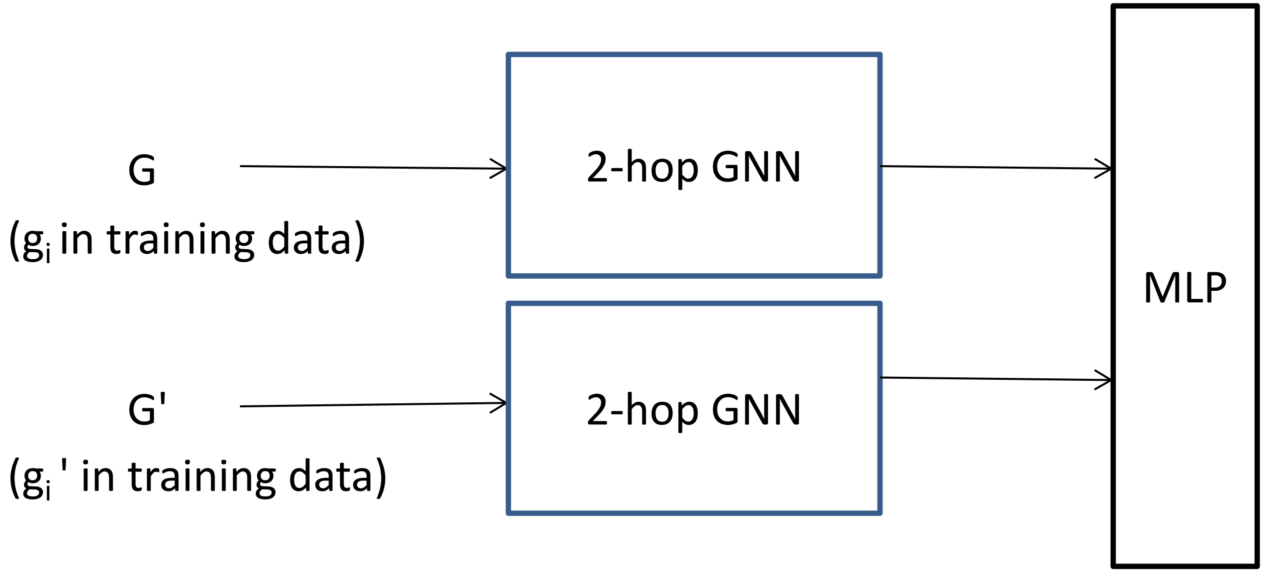


Figure 6: I use separate 2-hop GNNs for each graph. Matching subgraphs are passed into the model one pair at one time.

Embedding part

- Result: vector representation of each node in G and G'
- Loss:

  Embedding of a subgraph = concatenation of constituent node embeddings

  L_E = Euclidean distance between the embedding of each training pair

Matching part

- High-degree GNN can capture more structural information, but
  - computationally expensive
  - structural information not clearly shown in the result

- My approach:
  - inspired by NLP methods
  - In graphs, nodes also have relations
  - Treat graphs as sequence of nodes
  - Use LSTM to learn about the sequence

Matching part

- Bi-directional LSTM(BiLSTM)

input: one subgraph g of G

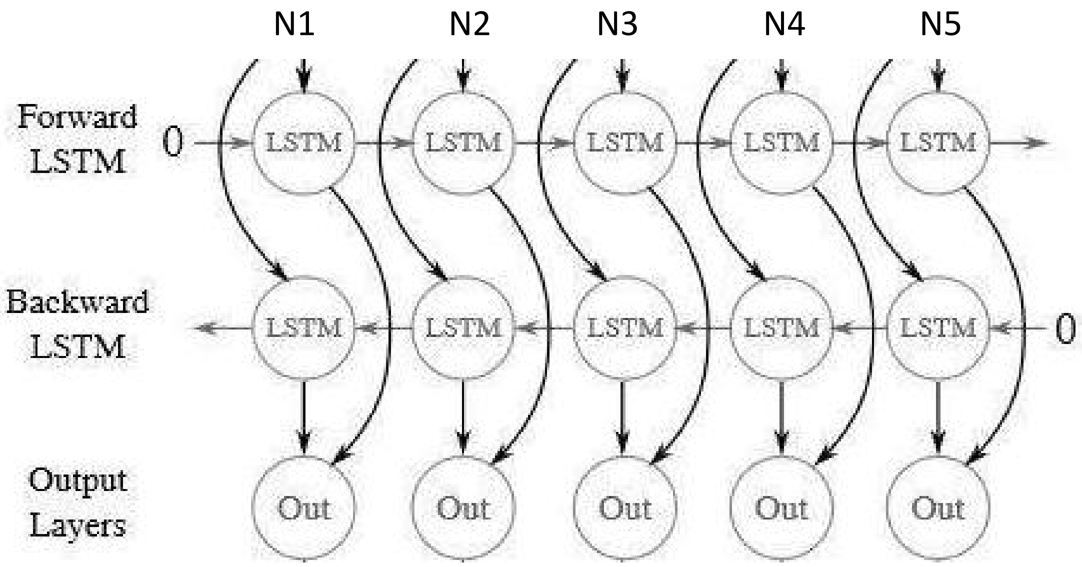


Figure 7: 1 layer Bi-directional LSTM used in the model.[5] The input is one subgraph in each pair.

Matching part

- LSTM: capable of catching long-short term dependency between nodes

- How to sequence the nodes?
  - random sequence

That's why using bi-directional LSTM. Nodes before may also have dependencies on the node after.

Matching part

- Result: vector representation of nodes, with structure information of the subgraphs considered. To get the matching subgraph g'' for g, find the closest node in G' for each node in g with the shortest euclidean distance between vectors

- Matching Loss:
  For each training pair g in G and g' in G',
  L_M = euclidean distance between g'' and g'

- Total loss:
  $L_E + \gamma L_M$ where γ is a parameter to be optimized via validation

Experimental Result

- Based on the synthetic dataset with subgraphs of 5 nodes
- ~ 20 % accuracy rate(g'' = g')
- > 80 % more than 3 nodes in g'' correctly identified
- ~ 20 % g'' found connected
- > 40 % all nodes in g' within a 2 hop neighbourhoods of g''
- Numbers are not good apparently
- relatively unresearched field, no baseline found

Future Work

- Try using GMN for the embedding part
- Further imbed the nodes into Wasserstein space to potentially improve performance.
- Try improving the matching part to increase the percentage where the matching subgraph is connected.
- Test on complex real-world datasets(Cora, etc)
- Submit a paper

Thank you!

References

[1]: Image from: https://www.slideshare.net/mobile/raul_A/attributed-graph-matching-of-planar-graphs

[2]: D.H.Kim,I.D.Yun and S.U.Lee,A New Attributed Relational Graph Matching Algorithm Using the Nested Structure of Earth Mover's Distance, 17th International Conference on Pattern Recognition, p. 48-51, 2004.

[3]. Image from: Jun Wu, Jingrui He, Jiejun Xu, https://www.kdd.org/kdd2019/

[4]. Image from: Y. Li et.al. Graph Matching Networks for Learning the Similarity of Graph Structured Objects. ICLR 2019.

[5]. Image from: https://www.sohu.com/a/283978749_717210

POMDP Approach for Tool-Use Planning

Junxian Liu, u6686392

Supervisor: Dr. Hanna Kurniawati

Content

- POMDP introduction
- My current problem
- State of art
- Model definition
- Things to do

POMDP Introduction

- A partially observable Markov decision process can be described as a tuple <S, A, T, R, Ω, O>.
- S is a finite set of states of the world;
- A is a finite set of actions;
- T is the state transition function. Given a world state and an action, there's a probability distribution over next wold state;
- R is the reward function;
- Ω is a finite set of observations the agent can get;
- O is the observation function that gives probability distribution over possible observations.

My current problem

- Generate a plan for a cutting task in high level.
- Why cutting?
  - Cutting is an interesting problem. Most things in this problem is uncertain to the agent. It's common for most of the other tasks.
- Why high level?
  - It's a gap. Now most works focus on how to identify tools, how to grab a tool and so on. But not much people tried to solve this entire problem in a high level.

State of art

- Other ways of doing planning:
  - Logic-Geometric Program
- Decide which tool to use:
  - Model-based approach: with rgbd camera, with point cloud;
  - Ontology & knowledge graph: wikidata, physical understanding;
  - Neural network approach.
- Decide where to grab:
  - Model-based approach: point cloud.
- POMDP solver:
  - OPPT;
  - Despot.

Model definition

- S (State)

  Holding x Strength x Tool1_Damage x Tool2_Damage x Tool3_Damage x Delta_Strength.

- A (Action)

  Pickup(tool), Putdown(), Cut(tool).

Model definition

- T (Transition)

Pickup (tool):

    pre: Holding nothing;
    post: Holding tool;
    probability: depends on which tool is being used.

Putdown ():

    pre: Holding tool;
    post: Holding nothing;
    probability: 1.0, always success.

Cut (tool):

    pre: Holding tool and Strength > 0;
    post: Strength -= Tool_Damage * RandomNumber in range[0, 1);
        Delta_Strength = the amound of strength reduced;
    probability: depends on which tool is being used.

Model definition

- R (Reward):

  Only cut action can get reward.

  If strength reduced, then the reward is 10;

  Otherwise, the reward is –5.

- Ω (Observations):

  Pickup / Putdown_result x the amount of Strength reduced.

Australian National University

Model definition

- O (Observation function):

  For Pickup action, if the operation success, there's a 0.9 probability to get a Success observation and a 0.1 probability to get a Fail observation. Similarly with the case when operation fails.

  Putdown action will always success, and will always get a Success observation.

  Cut action get an observation about how much Strength is reduced by the tool that the agent use for the cut. The probability is normally distributed in range [0, 11], with the reduced amount as mean.

Things to do

- Try to improve the plan with some heuristics;
- Try to execute the plan in a simulator if time permitted;
- Finish my thesis.

Thank you for listening.

Australian National University

References

- *Leslie Pack Kaelbling, Michael L. Littmanb, Anthony R. Cassandra (1998). Planning and acting in partially observable stochastic domains. Artificial Intelligence 101 (1998) 99–134.*
- *Marc Toussaint, Kelsey R. Allen, Kevin A. Smith, Joshua B. Tenenbaum. Differentiable Physics and Stable Modes for Tool-Use and Manipulation Planning.*
- *Paulo Abelha, Frank Guerin, Markus Schoeler(2016). A Model-Based Approach to Finding Substitute Tools in 3D Vision Data. 2016 IEEE International Conference on Robotics and Automation (ICRA).*
- *Lydia Fischer et al(2018). Which tool to use? Grounded reasoning in everyday environments with assistant robots. http://ceur-ws.org*

Australian National University

Seasonal Droughts Forecast in Australia

Pengbo Li, u6683977
Supervisor: Dr. Warren Jin

Outline

- Motivation
  - Severe drought in Australia
  - Drawbacks of state-of-the-art models
- State of Art
  - ANN
  - ELM
- GAM
- Result
  - Evaluation
  - Comparison
- Future Work

Motivation

- Severe drought in Australia

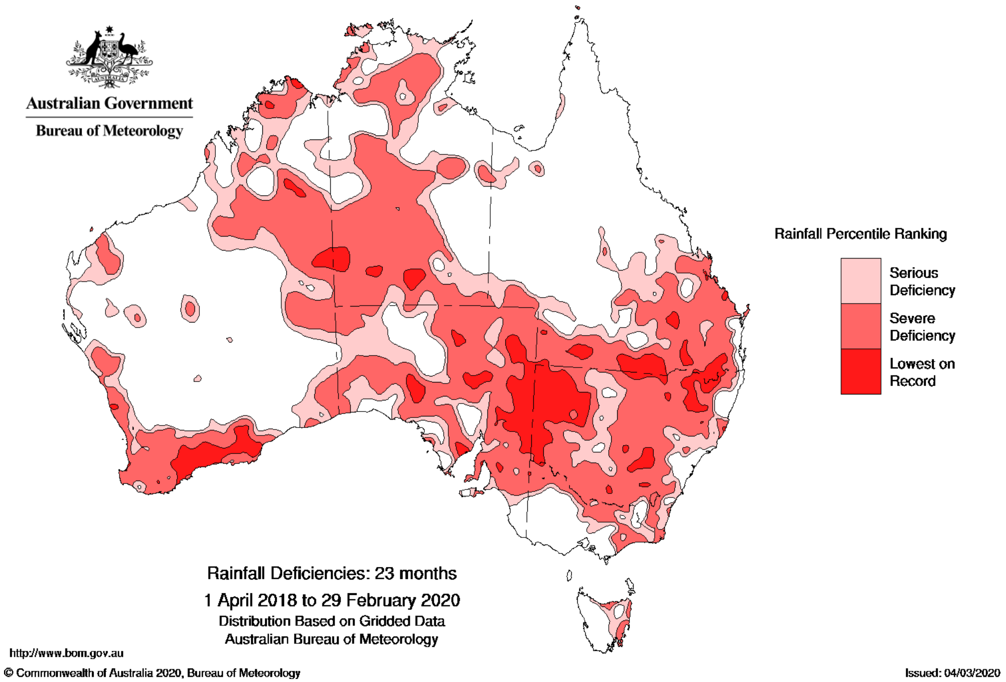
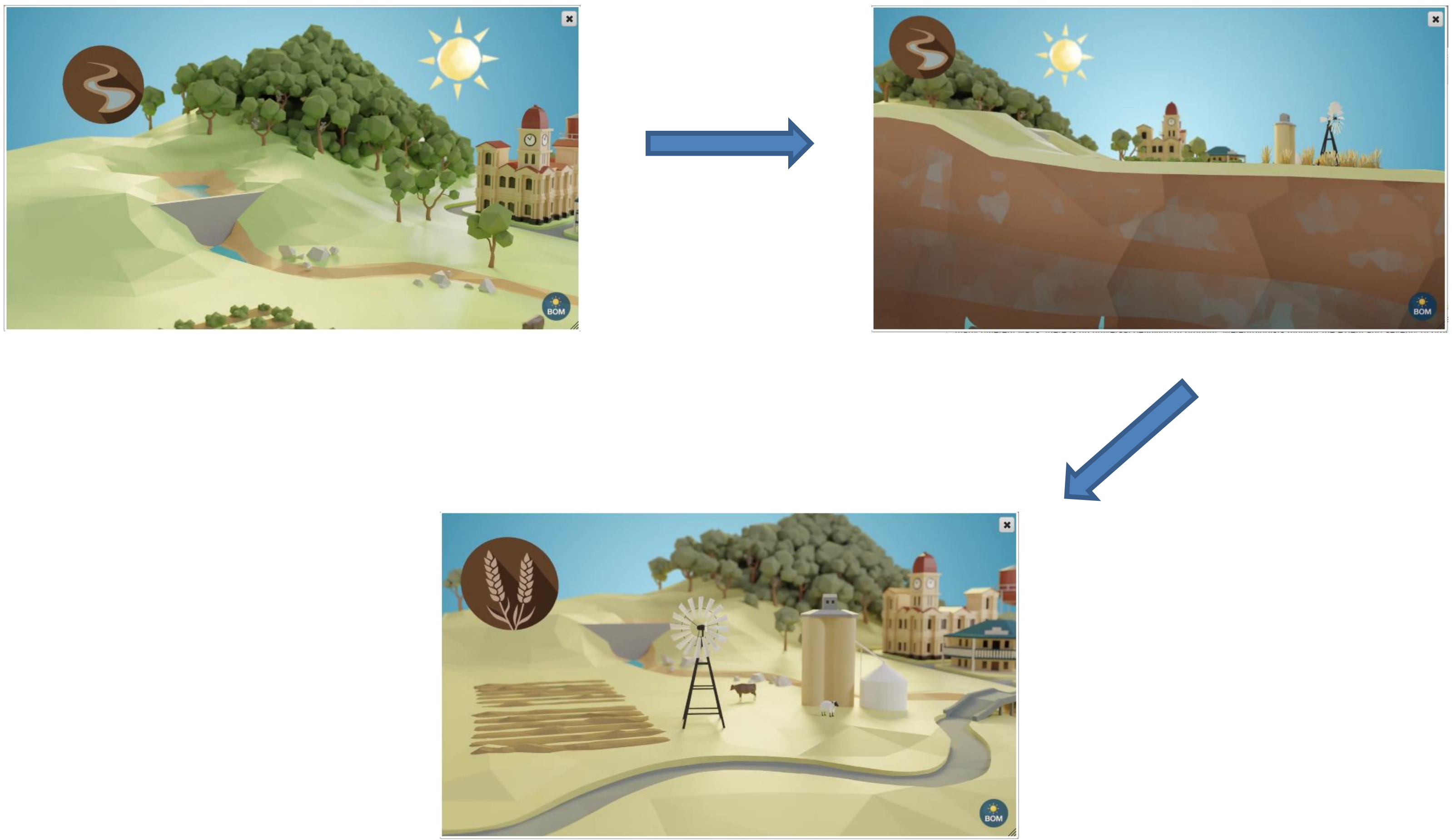


Figure 1 Rainfall deficiencies

http://www.bom.gov.au/climate/drought/archive/20200305.drought2.lr.col.gif

Motivation



State of Art

- Artificial Neural Network[1]
  - Inspired by the biological neural networks that constitute animal brains;
  - Specified by network topology, neuron characteristics, and training or learning rules;
  - Operating in parallel;
  - Massive network connection characterized by different weights
  - Less assumption-dependent

State of Art

- Extreme Learning Machine[2]
  - State-of-the-art single layer feedforward neural network;
  - Easy to use and to be tuned;
  - Better generalization ability than ANN, SVM and singular value decomposition (SVD) algorithms;

Model in this Project

- Generalized Additive Models
  - Developing from Generalized Linear Models;

$$y \sim N(\mu, \sigma^2) \quad\longrightarrow\quad y \sim N(\mu, \sigma^2)$$
$$g(\mu) = b_0 + b_1X_1 + b_2X \quad\longrightarrow\quad g(\mu) = f(X)$$

  - Smooth functions allow for nonlinear relationship
  - Different smooth classes to use

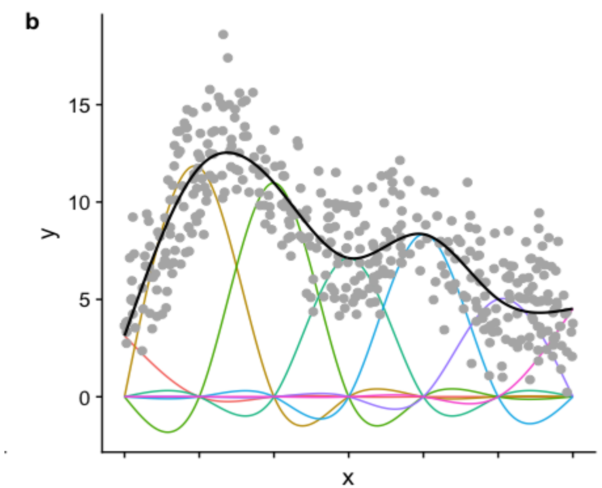


Figure 2 Basis functions add up to create the overall smooth shape

How to define the f(x)

- Find the indices interactions geographical papers
- Write and modify the formula

gam(SPEI ~ ti(soi, iod) + s(soi) + s(iod) +…)

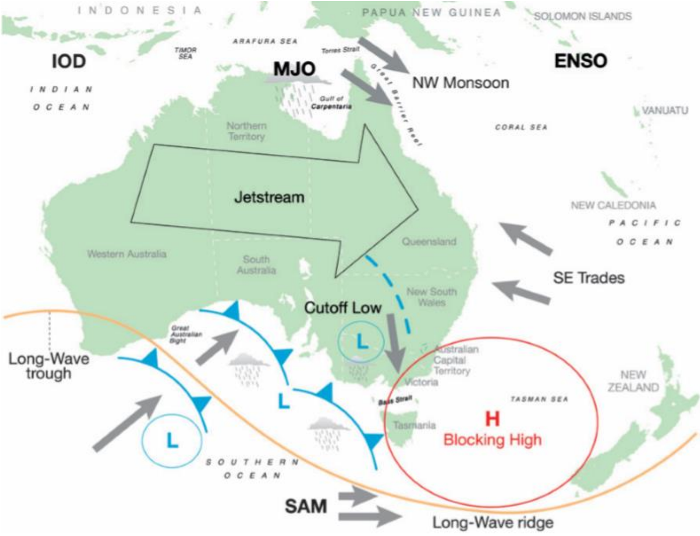


Figure 3 The climate indices influencing Australia rainfall

The interaction

- How they affect the predictor

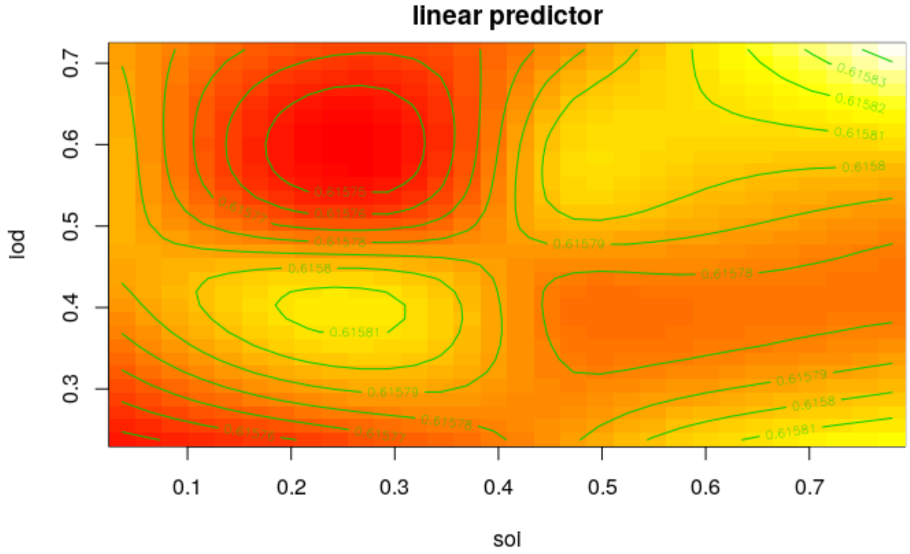


Figure 4 The heatmap of the interaction between soi and iod

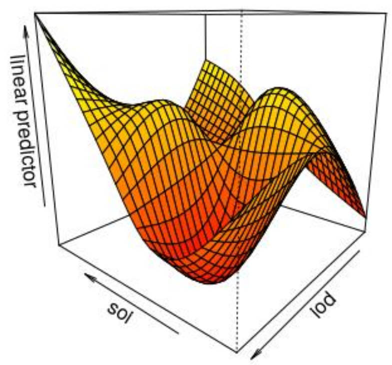


Figure 5 The 3D view of the interaction between soi and iod

Evaluation

- Nash-Sutcliffe efficiency coefficient
  - NASH = 1 - $\frac{\sum_{i=1}^{N}(SPEI_{o,i}-SPEI_{p,i})^2}{\sum_{i=1}^{N}(SPEI_{o,i}-\overline{SPEI_{o}})^2}$, $-\infty \leq NASH \leq 1$
- Root Mean Square Error
  - RMSE = $\frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(SPEI_{p,i}-SPEI_{o,i})^2}}{SPEI_{o,max}-SPEI_{o,min}}$
- Normalized Mean Absolute Error
  - NMAE = $\frac{\frac{1}{N}\sum_{i=1}^{N}\left|(SPEI_{p,i}-SPEI_{o,i})^2\right|}{SPEI_{o,max}-SPEI_{o,min}}$

Result Comparison

Model	NASH	RMSE	NMAE
ANN	0.987	0.024	0.018
ELM	0.953	0.049	0.036
GAM	**0.978**	**0.033**	**0.024**

Future Work

- Test on more stations
- Generate the formula automatically
- Increase the accuracy
- Write the thesis

Reference

[1] Deo, Ravinesh C., and Mehmet Şahin. "Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia." *Atmospheric research* 161 (2015): 65-81.

[2] Mouatadid, Soukayna, et al. "Input selection and data-driven model performance optimization to predict the Standardized Precipitation and Evaporation Index in a drought-prone region." *Atmospheric research* 212 (2018): 130-149.

Async-Sync Learning System

Hangyeul Lee

Supervisor Mr. Tom Worthington

Overview : Async vs Synchronous -learning

- Asynchronous Learning

Not occurring in same time & place

→ Independent learning

e.g.)

Forums, lecture recordings, collaborative software, quiz, etc.

Overview : Async vs Synchronous -learning

- Synchronous Learning

Does occur in same time & place!

→ Instructor - student engage in real-time

e.g.)

Livestreams, Video conference,
Live demo, *Scheduled* quiz, etc.

Problems of Async/Sync Learning

Asynchronous

- No instant feedbacks

Synchronous

- Time dependency
- Bandwidth (i.e. slow internet)
- Lack of individual attention

Problems of Async/Sync Learning

Asynchronous

- No instant feedbacks

Synchronous

- Time dependency
- Bandwidth (i.e. slow internet)
- Lack of individual attention

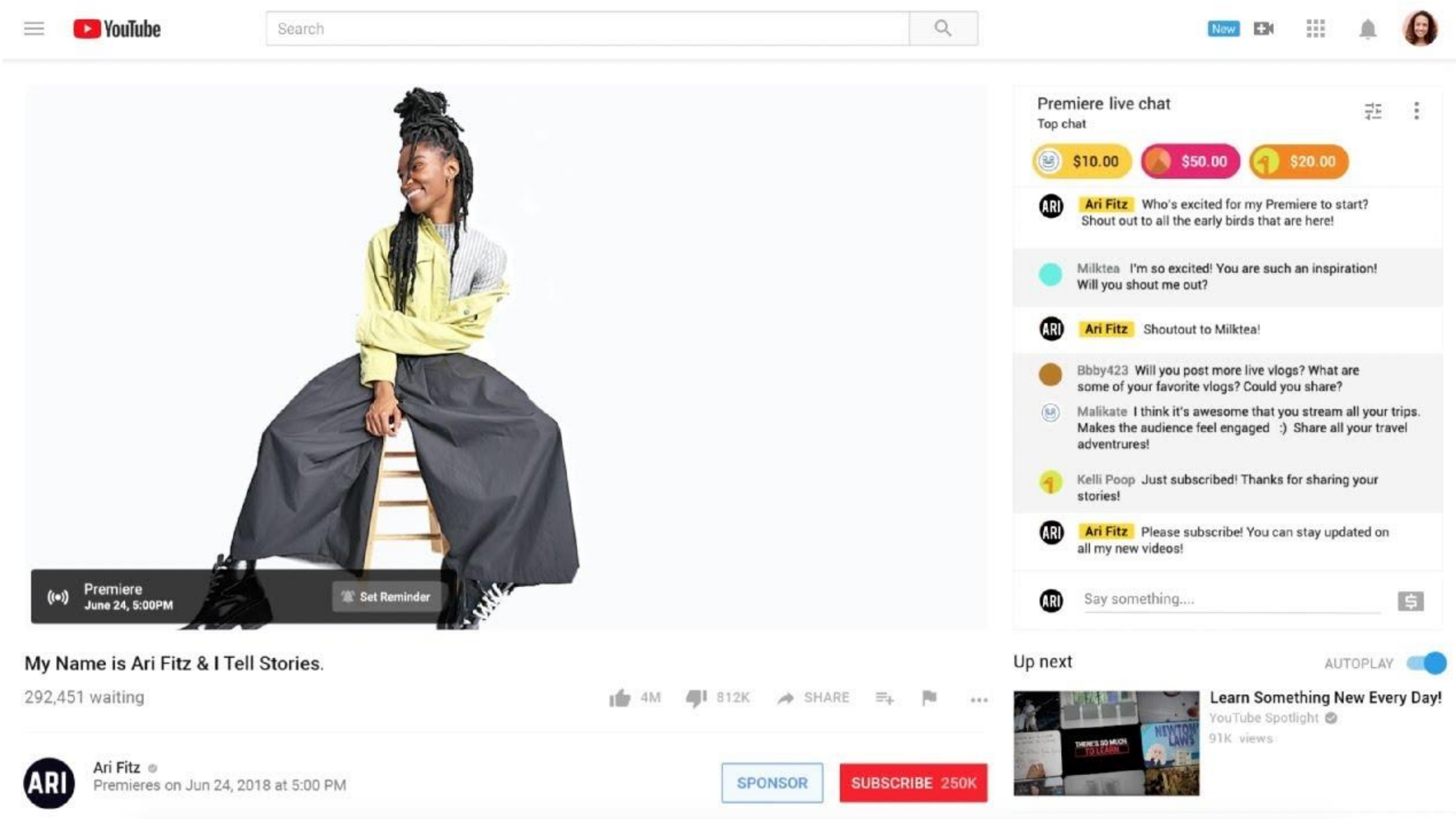
How to complement?

Synchronised Asynchronous Learning

Simple idea : Separate *communication* from *content delivery*

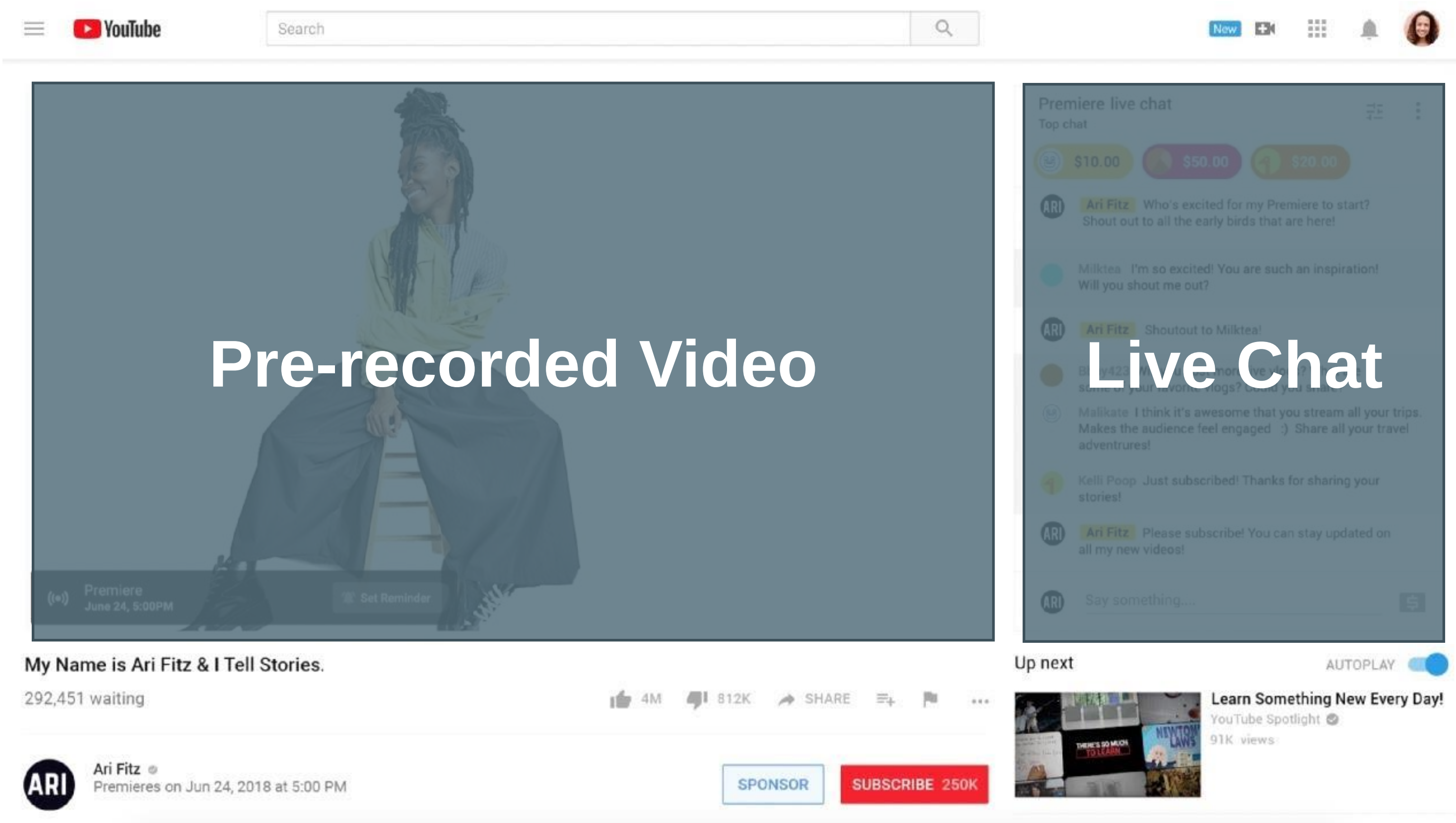
Content Delivery	Communication
Recorded video (streamed)	Text-based live chat

Synchronised Asynchronous Learning



https://support.google.com/youtube/answer/9080341

Synchronised Asynchronous Learning



https://support.google.com/youtube/answer/9080341

Problems of Async/Sync Learning

Asynchronous

- ~~No instant feedbacks~~

Synchronous

- ~~Time dependency~~
- ~~Bandwidth (i.e. slow internet)~~
- ~~Lack of individual attention~~

Problems of Async/Sync Learning

Asynchronous

- ~~No instant feedbacks~~

Synchronous

- ~~Time dependency~~
- ~~Bandwidth (i.e. slow internet)~~
- ~~Lack of individual attention~~

Problems Solved?

Measures

- **Efficacy**
  – Is this system suitable for education?
- **Robustness**
  – Can this system cope with unexpected situations?
- **Flexibility**
  – Is there any other functionality users might need?

Measures

- Build a system, make students use it for a semester, check their learning outcomes…

- Long-term researches…

→ *NOT* possible!

Transcript Analysis

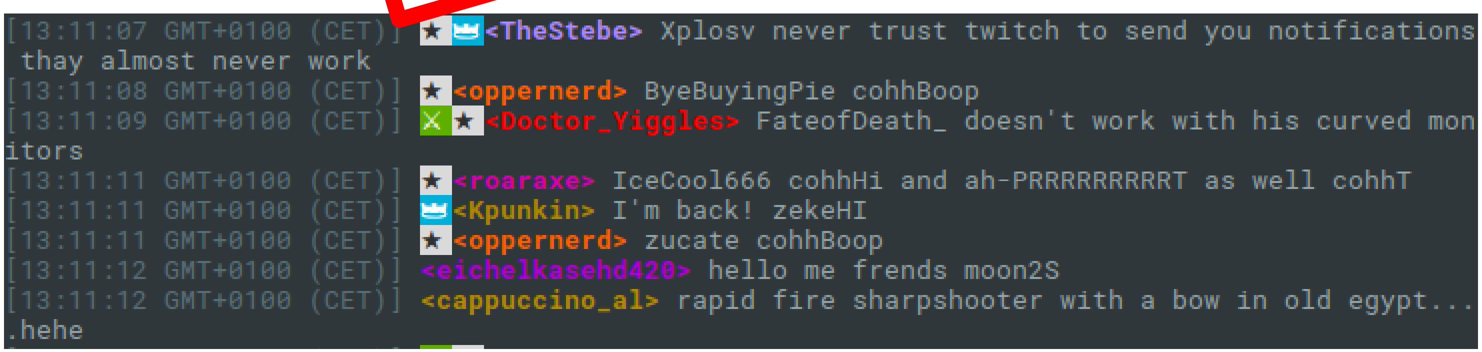
- Extract & analyse different types of interactions from live chat logs from actual lectures

- student-student, student-teacher, student-content

Transcript Analysis

- Extract & analyse different types of interactions from live chat logs from actual lectures
- student-student, st … n, student-content

Suspended



Focus Group Discussion

- 6 participants invited to online forum
- 1-week duration
- Subforums : Design, Usability, Functionality

Role	Number of participants
Education experts	2
IT experts	1
Course Coordinators	1
Students	2
Total	6

"How do you think about the concept?"
"Is this system usable in higher education?"
"What functionality will be needed in real life?"

Future Works

- Transcript analysis
- Build working prototype based on FGD
- Create design criteria which can be applied to learning systems / modules

Thank You!