Towards Natural Intelligence (NI): Technologies for Understanding Human Behavior and Activities

Ashutosh Modi

CS Katha Barta Series, NISER

Department of Computer Science and Engineering

Exploration Lab

अन्वेषण प्रयोगशाला

IIT KANPUR

Indian Institute of Technology, Kanpur

Overview of Exploration Lab

Exploration Lab

अन्वेषण प्रयोगशाला

NLP and NLU

Legal NLP

Understanding and Processing Indian Legal Texts, Legal Foundational models, Summarization, Cross-Lingual, Cross Domain Knowledge Transfer, Legal KG

Natural Language Retrieval

Retrieving information from databases via natural language queries

Biomedical NLP

NER, Relation Extraction, Clinical Trials….

Machine Unlearning

Forgetting Unwanted information in LLMs, Updating LLMs with latest facts without training

Social Reasoning in LLMs

Teaching ethics and etiquettes to LLMs

Miscellaneous

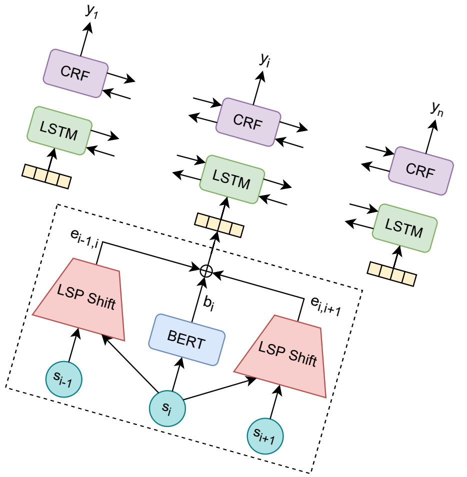
Automatic Speech Recognition for noisy, code-mixed speech

ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation
ACL, 2021

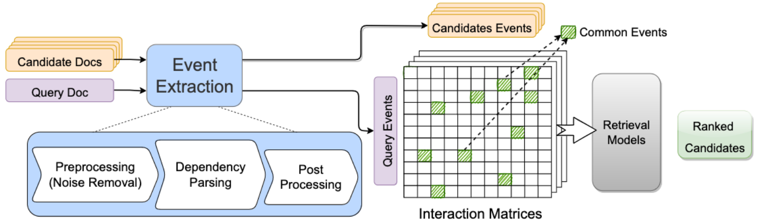
Semantic Segmentation of Legal Documents via Rhetorical Roles
NLLP, EMNLP 2022

HLDC: Hindi Legal Document Corpus
ACL Findings 2022

Corpus for automatic structuring of legal documents
LREC 2022



U-CREAT: Unsupervised Case Retrieval using Events extrAction
ACL 2023



BookSQL: A Large Scale Text-to-SQL Dataset for Accounting Domain
Under review, EACL 2023

EtiCor: Towards Analyzing LLMs for Etiquettes
EMNLP 2023

ASR for Low Resource and Multilingual Noisy Code-Mixed Speech
Interspeech, 2023

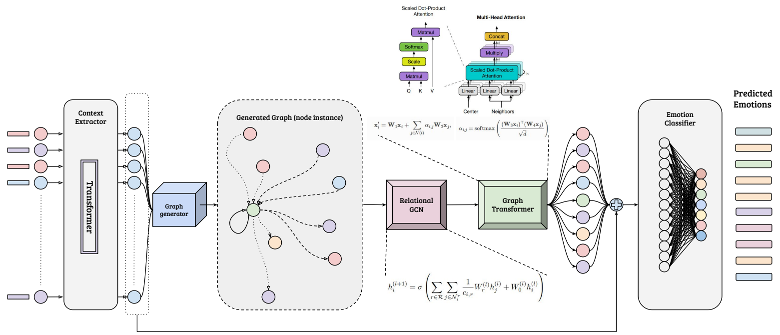
Fine Grained Emotion Prediction by Modeling Emotion Definitions
Best Student Paper Award, ACII , 2021

Adapting a Language Model for Controlled Affective Text Generation
CoLING, 2020

An End-to-End Network for Emotion-Cause Pair Extraction
WASSA, EACL, 2020

Shapes of Emotions: Multimodal Emotion Recognition in Conversati
PIM3SM, COLING 2022

Multi-Task Learning Framework for Extracting
Emotion Cause Span and Entailment in Conversations
TL4NLP, NeuRIPS 2022



COGMEN: COntextualized GNN based Multimodal Emotion recognitioN, NAACL, 2022

Modeling Human Behavior and Decision Making

Affective Computing

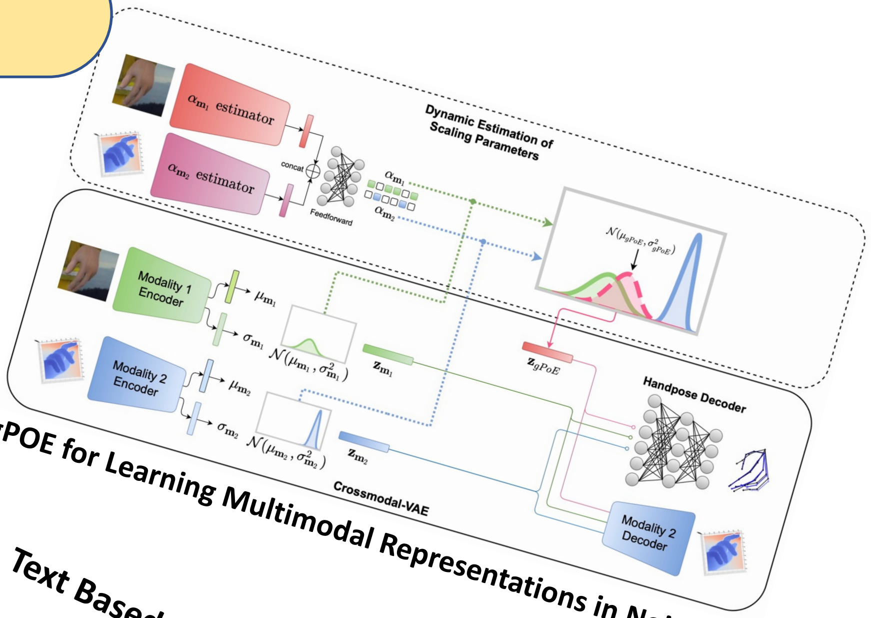
- Multimodal Representations
- Multimodal Multilingual Contextualized Affect Prediction
- Multimodal Generation
- Emotion and Decision Making: Emotion Cause Prediction

RL Worlds (Towards Embodied AI)

- Decision Making by Agents in Text Worlds
- Agents learn about real world without any explicit supervision via interactions with the environment simulating real world.

Mental Health

- Study correlation between speech, language, neuro-imaging, and **Schizophrenia symptoms**.



gPOE for Learning Multimodal Representations in Noisy Environments
ICMI, 2022

Text Based Environment For Learning Procedural Knowledge
ScriptWorld:
Outstanding Paper Award, LAREL NeuRIPS 2022
IJCAI-23, EA-AAMAS 2023,

Pre-Trained Language Models as Prior Knowledge for Playing Text Based Games
AAMAS, 2022

SERB INDIA
CONVIN
DRDO
UCL

Exploration Lab

Ashutosh Modi

AI For Social Good

Sign Language Translation and Generation

- Sign language understanding
- Linguistic Analysis
- NLP Tools for Sign Language
- Translation within sign languages and with natural language
- Generation conditioned on context and other modalities

Corpus for Indian Sign Language Recognition
EMNLP 2022

There is an imminent need for development of Sign language Technologies for promoting and protection of linguistic rights of the deaf community.

राष्ट्रीय मानव अधिकार आयोग, भारत
NATIONAL HUMAN RIGHTS COMMISSION, INDIA

ISLTranslate: Dataset for Translating Indian Sign Language
ACL Findings 2023

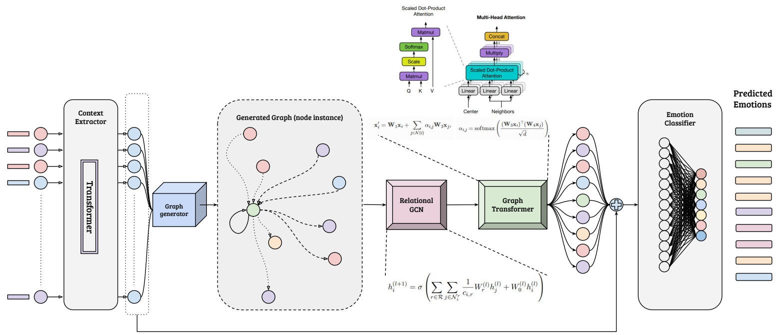
Fine Grained Emotion Prediction by Modeling Emotion Definitions
Best Student Paper Award, ACII , 2021

Adapting a Language Model for Controlled Affective Text Generation
CoLING, 2020

An End-to-End Network for Emotion-Cause Pair Extraction
WASSA, EACL, 2020

Shapes of Emotions: Multimodal Emotion Recognition in Conversati
PIM3SM, COLING 2022

Multi-Task Learning Framework for Extracting
Emotion Cause Span and Entailment in Conversations
TL4NLP, NeuRIPS 2022



COGMEN: COntextualized GNN based Multimodal Emotion recognitioN, NAACL, 2022

Modeling Human Behavior and Decision Making

Affective Computing

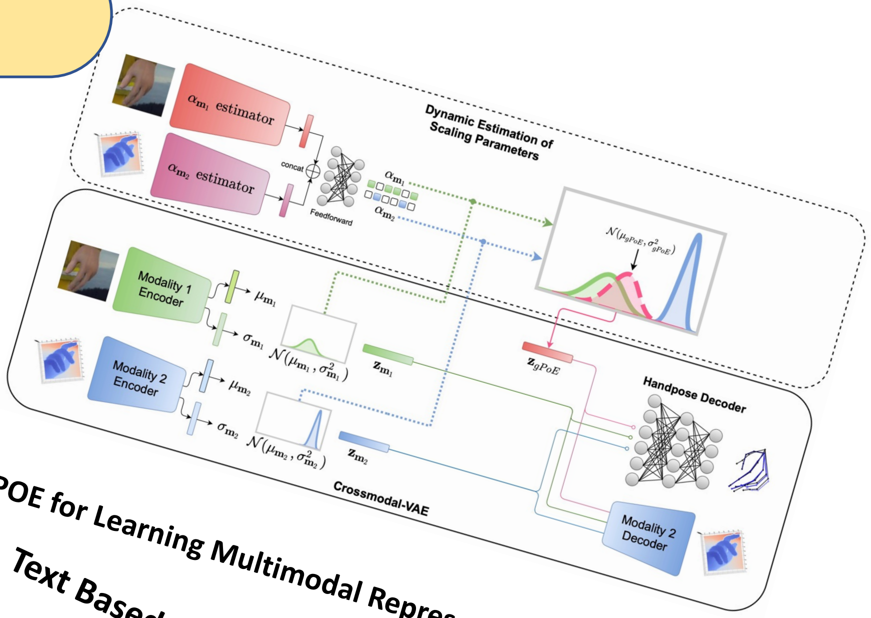
- Multimodal Representations
- Multimodal Multilingual Contextualized Affect Prediction
- Multimodal Generation
- Emotion and Decision Making: Emotion Cause Prediction

RL Worlds (Towards Embodied AI)

- Decision Making by Agents in Text Worlds
- Agents learn about real world without any explicit supervision via interactions with the environment simulating real world.

Mental Health

- Study correlation between speech, language, neuro-imaging, and **Schizophrenia symptoms**.



POE for Learning Multimodal Representations in Noisy Environments
ICMI, 2022

ScriptWorld:
Text Based Environment For Learning Procedural Knowledge
AAMAS 2023,

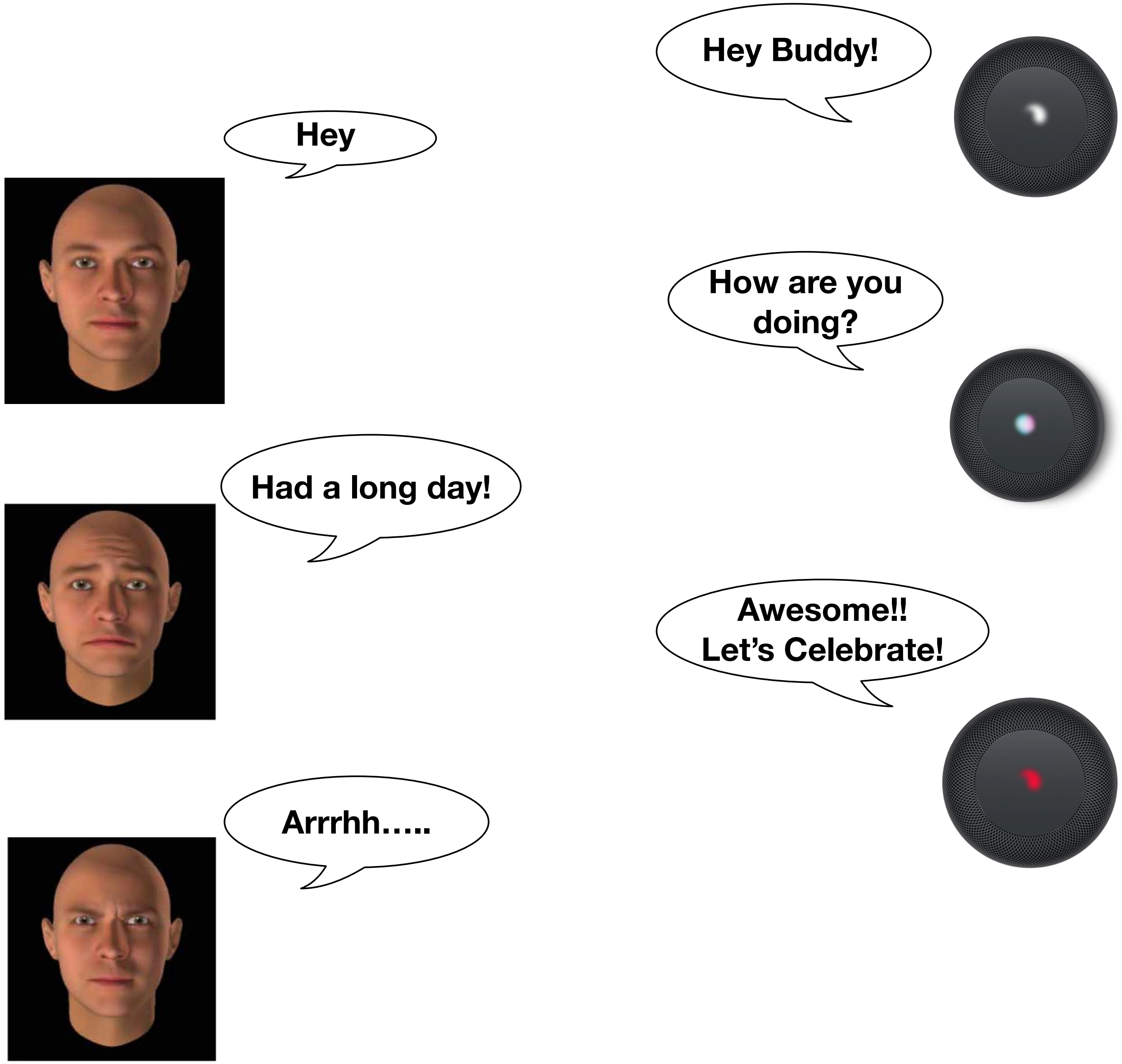
Outstanding Paper Award, LAREL NeuRIPS 2022

Pre-Trained Language Models as Prior Knowledge for Playing Text Based Games
AAMAS, 2022

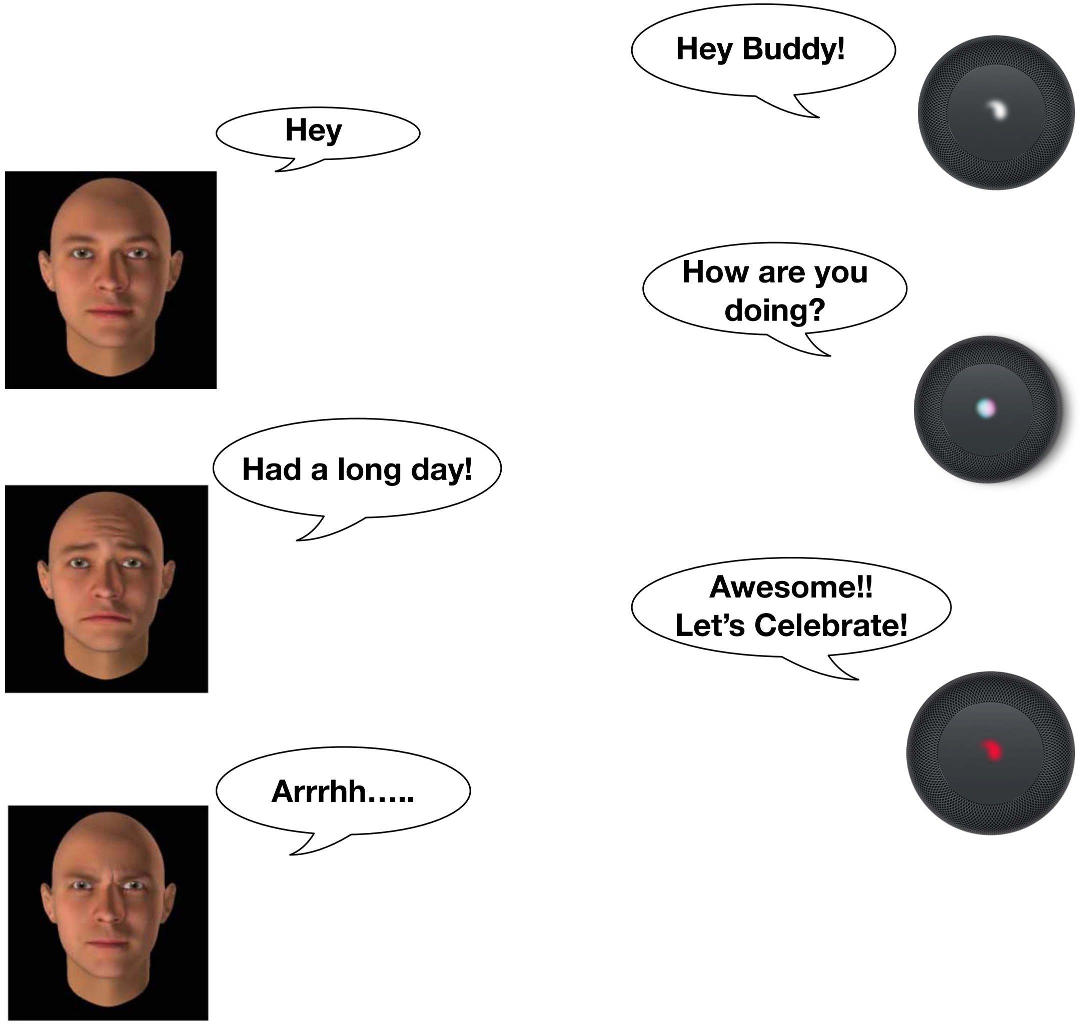
Ashutosh Modi

Why Affect (Emotion)?

Interaction with Personal Digital Assistants



Interaction with Personal Digital Assistants



Currently, machines fail to understand ***Affects*** (emotions)

Why Study Emotions?

- Emotions are universal (Ekman, 1972, 1973).
- To interact seamlessly, it is important to understand underlying emotions.
- Emotions convey information beyond surface level features in communication

Emotion is not especially different from the processes that we call thinking.

- Emotion Machine, 2007
M. Minsky (AI Pioneer)

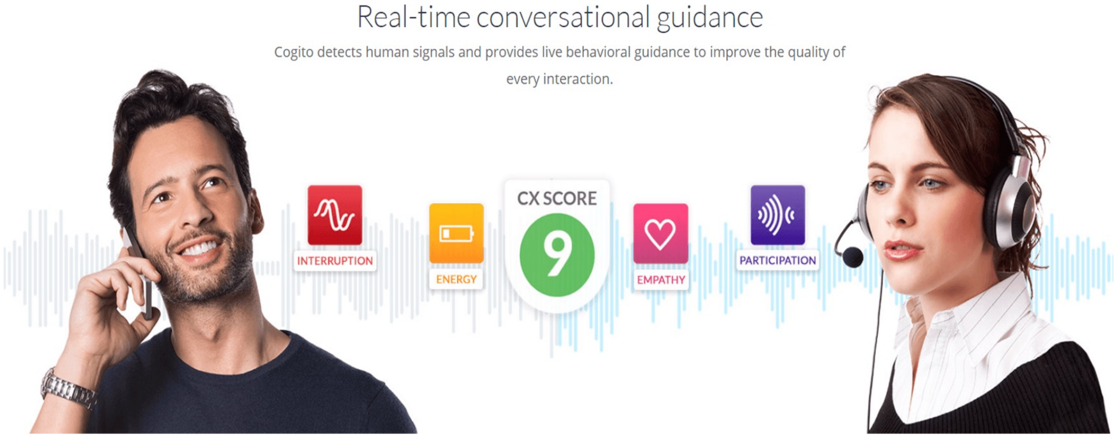
Affective Computing

- **Affective Computing**: Study and development of systems that recognize, interpret, process and simulate human feelings and emotions (R.W. Picard, MIT, 1995)
- a.k.a. Artificial Emotional Intelligence or Emotional AI

Image: https://www.scienceofpeople.com/microexpressions/

Applications

Customer Behavior Understanding

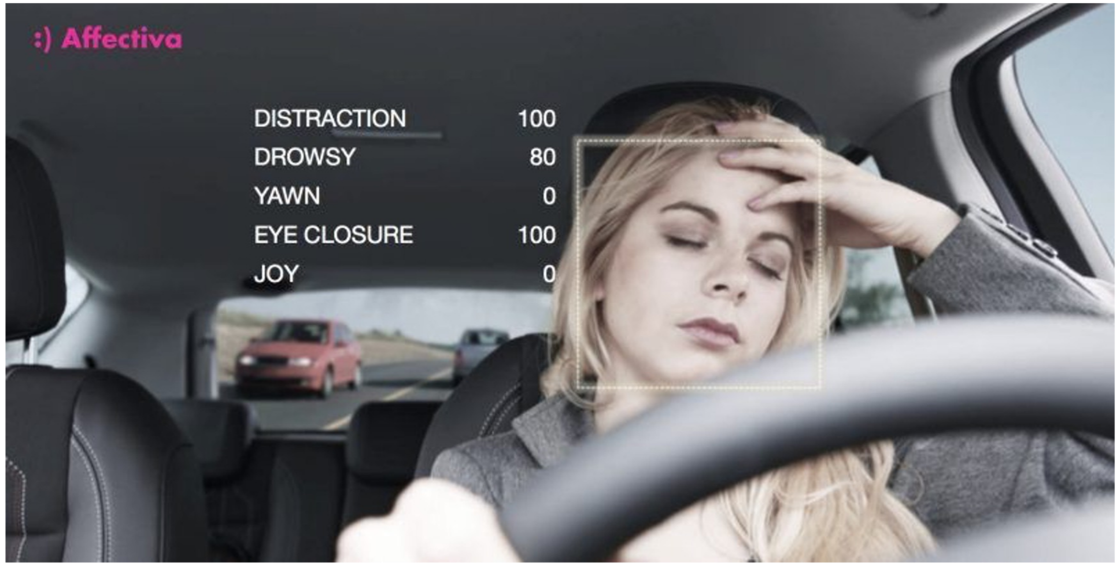


HBS

Social Media Analysis

fiverr

Vehicular Technologies



FutureCar

Audience Understanding

Image: https://www.searchenginejournal.com/content-seo-audience-understanding/386525/

Emotional Machines

- To develop machines that interact seamlessly with humans, machine should understand the emotions as well as should be able to exhibit emotions.

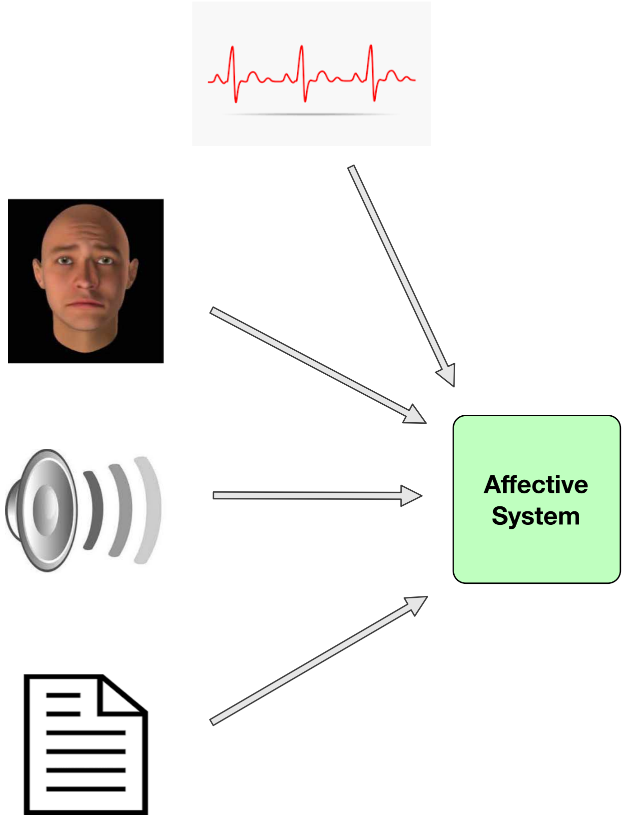
- Two class of problems need to be addressed:
  - Emotion Prediction or Recognition
  - Emotion Generation

Fine-Grained Emotion Prediction by Modeling Emotion Definitions, ACII 2020
Affect-Driven Dialog Generation, NAACL 2019
Adapting a Language Model for Controlled Affective Text Generation, CoLING 2019

Multimodal Affective Computing

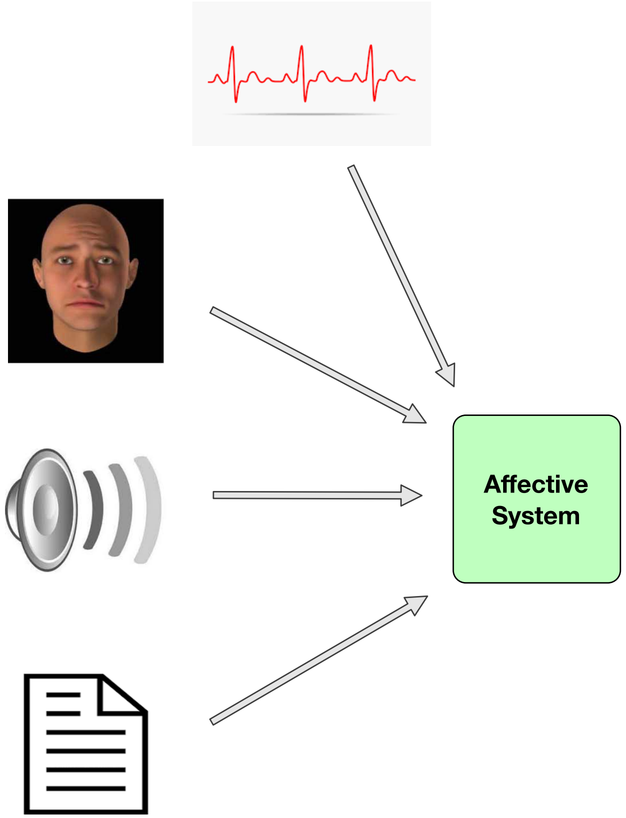
- Affect is not an isolated phenomenon, it is present across different modalities (Text, Audio, Video, Pulse Rate, Eye Movement, etc.)

- Modalities complement each other regarding the affect information.



Multimodal Affective Computing

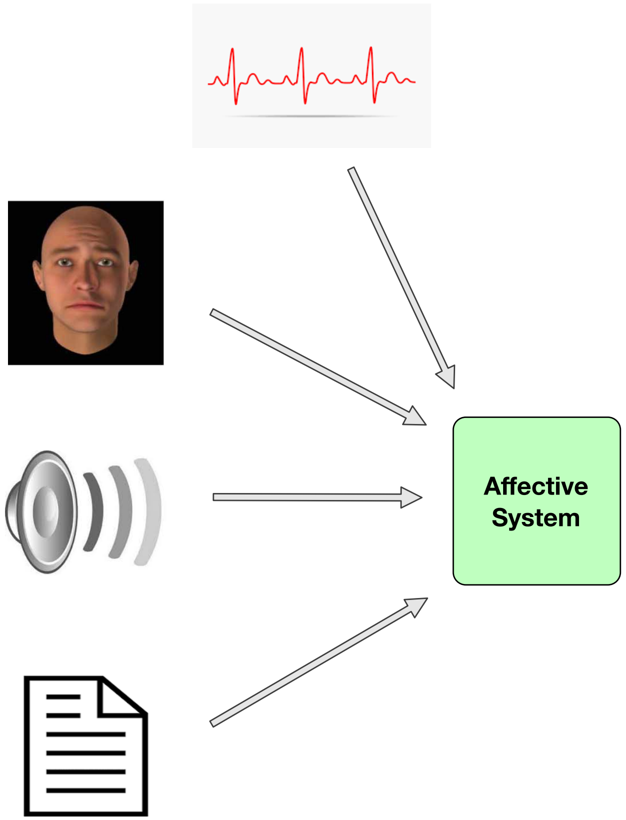
- Affect is not an isolated phenomenon, it is present across different modalities (Text, Audio, Video, Pulse Rate, Eye Movement, etc.)
- Modalities complement each other regarding the affect information.
- How does one fuse the information from different modalities?



Generalized Product-of-Experts for Learning Multimodal Representations in Noisy Environments, ICMI, 2022

Multimodal Affective Computing

- Affect is not an isolated phenomenon, it is present across different modalities (Text, Audio, Video, Pulse Rate, Eye Movement, etc.)
- Modalities complement each other regarding the affect information.
- How does one fuse the information from different modalities?
- Affect is contextualized

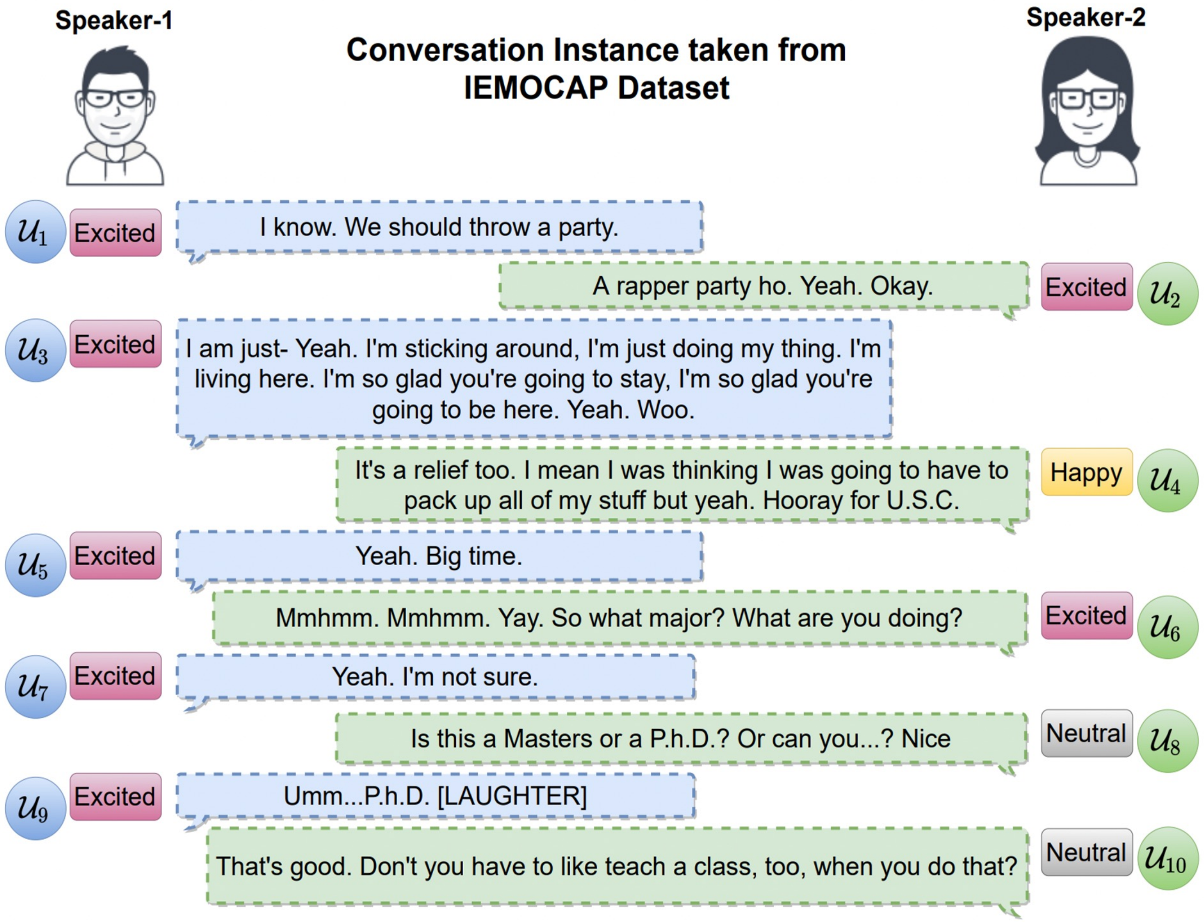


COGMEN: COntextualized GNN based Multimodal Emotion recognitioN

Abhinav Joshi, Ashwani Bhat,
Ayush Jain, Atin Vikram Singh,
Ashutosh Modi

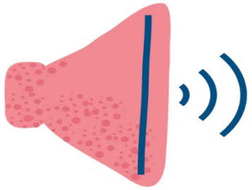
NAACL 2022

Emotion Recognition in Conversations (ERC)

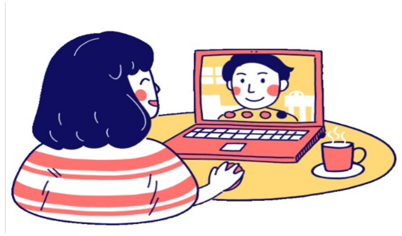


Text

Audio



Visual

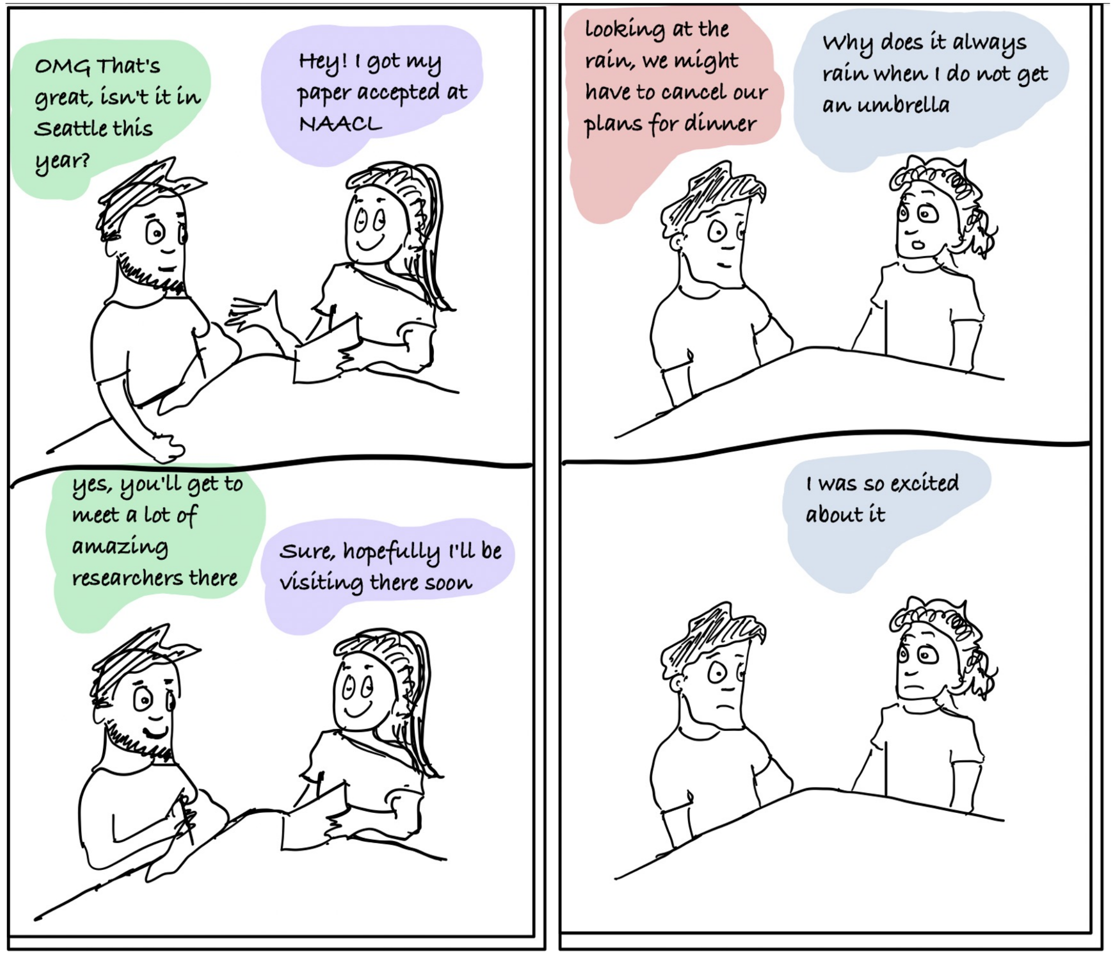


Given a multimodal conversation between different speakers, predict the emotional state of the speaker after each utterance.

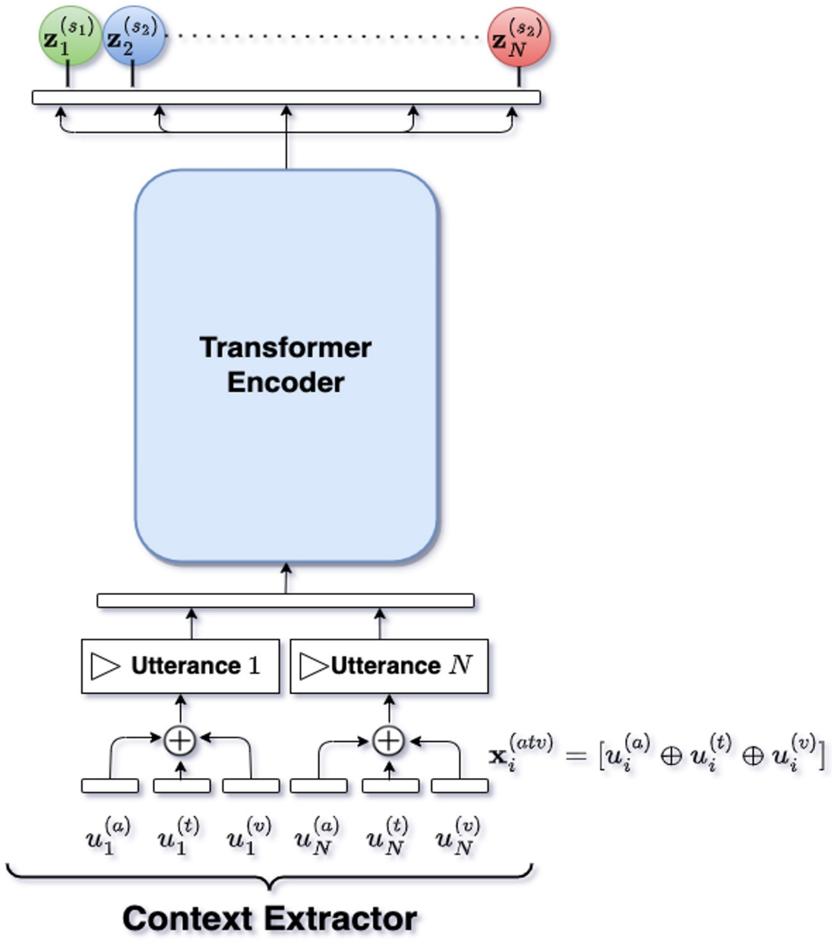
Model Intuition

Global Information:
How to capture the impact of underlying context on the emotional state of an utterance?

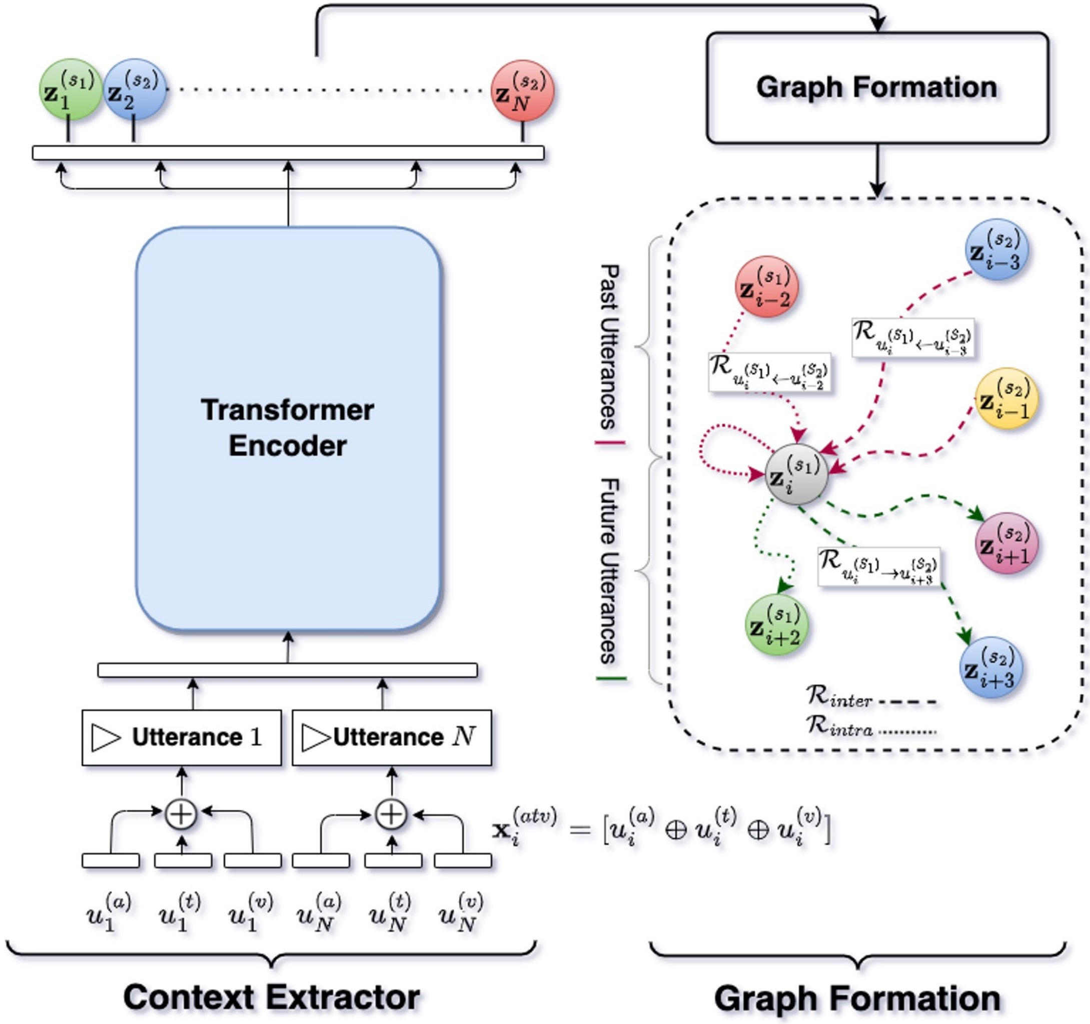
Local Information:
How to establish relations between the nearby utterances that preserve both inter-speaker and intra-speaker dependence on utterances in a dialogue?



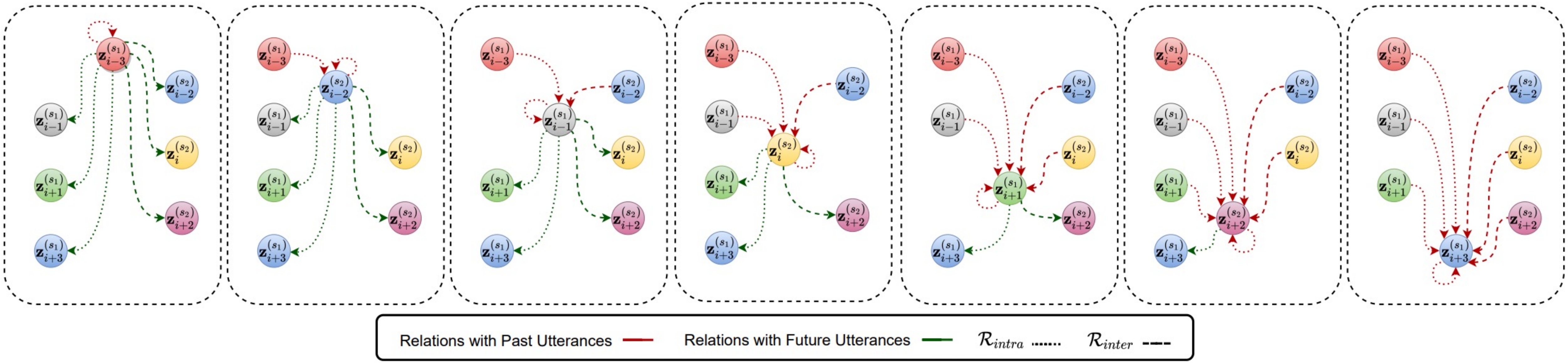
COGMEN Architecture



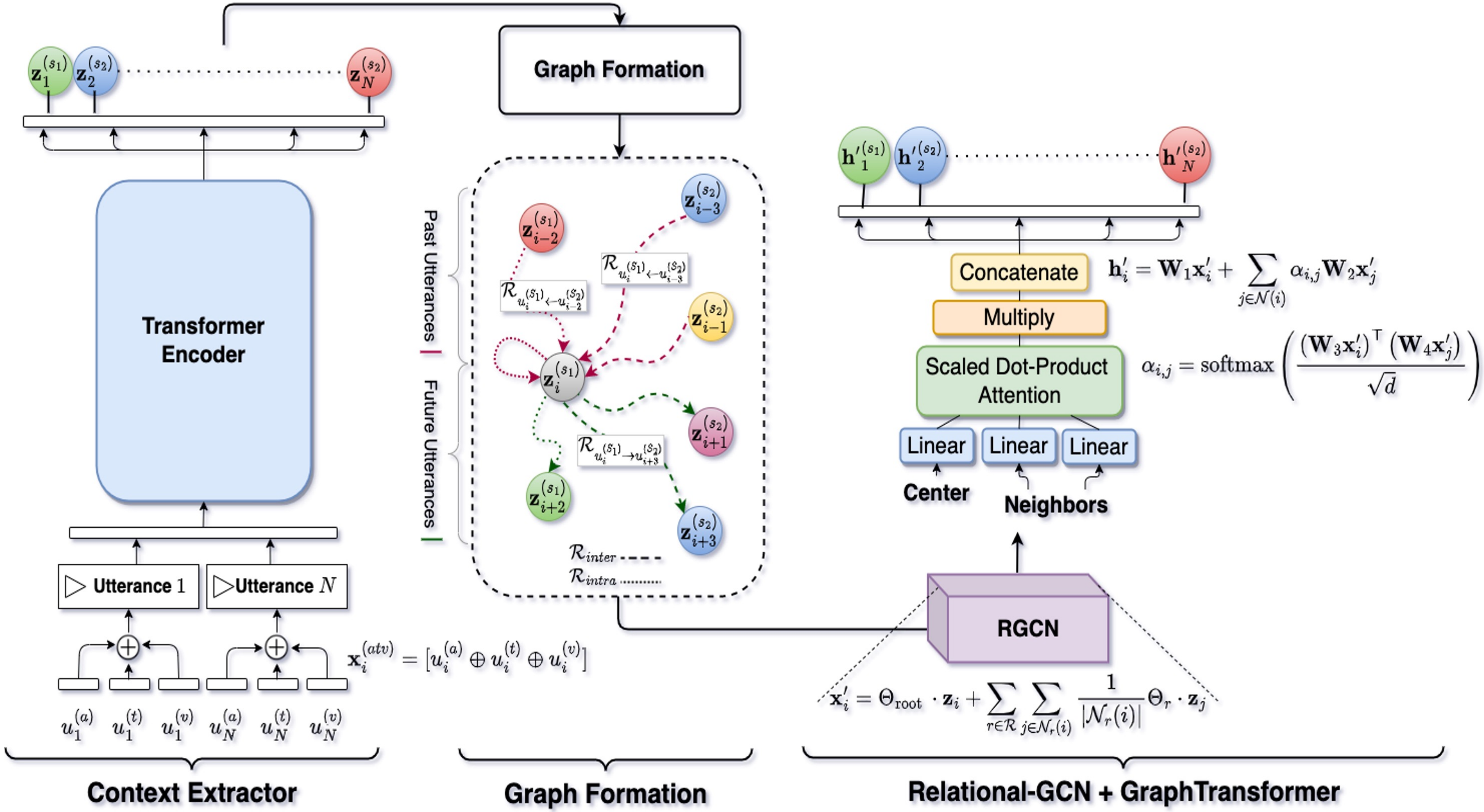
COGMEN Architecture



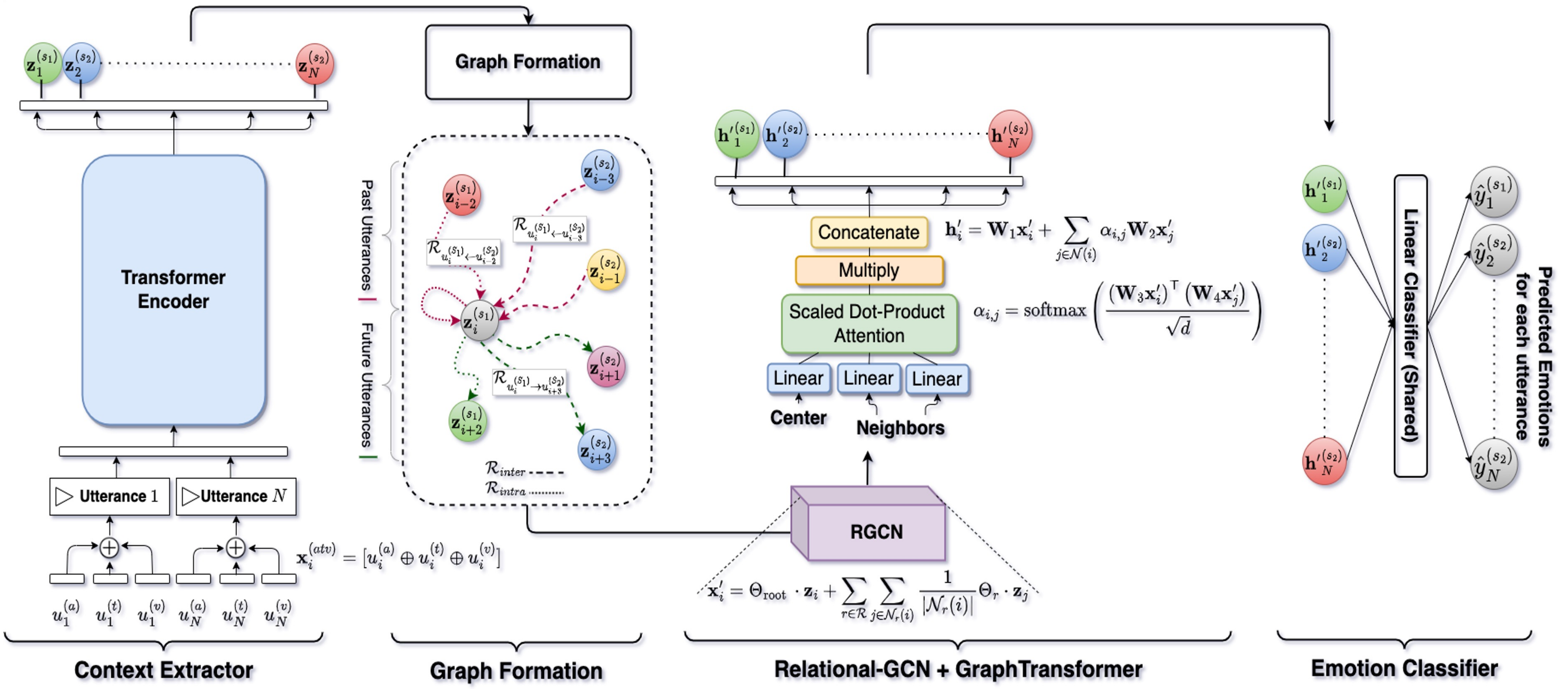
Graph Formation in COGMEN Architecture



COGMEN Architecture

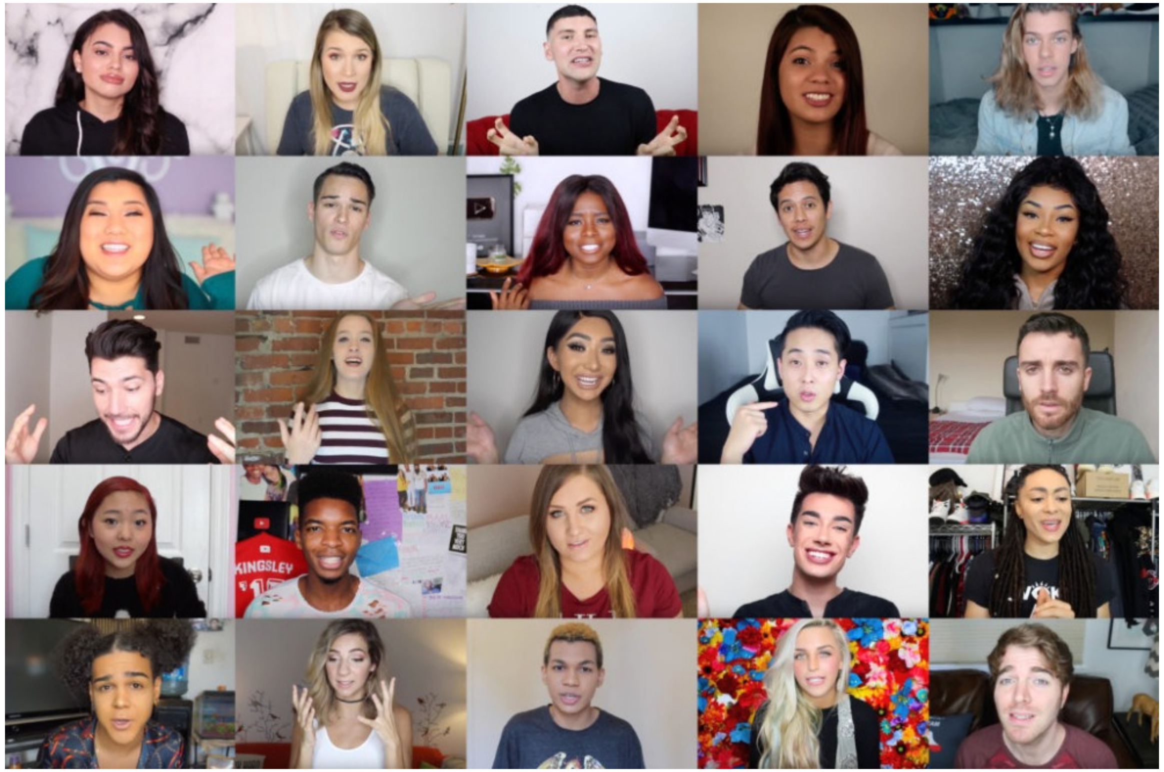


COGMEN Architecture



Multimodal Emotion Datasets

CMU-MOSEI Benchmark



IEMOCAP Benchmark

Experiments on IEMOCAP Benchmark

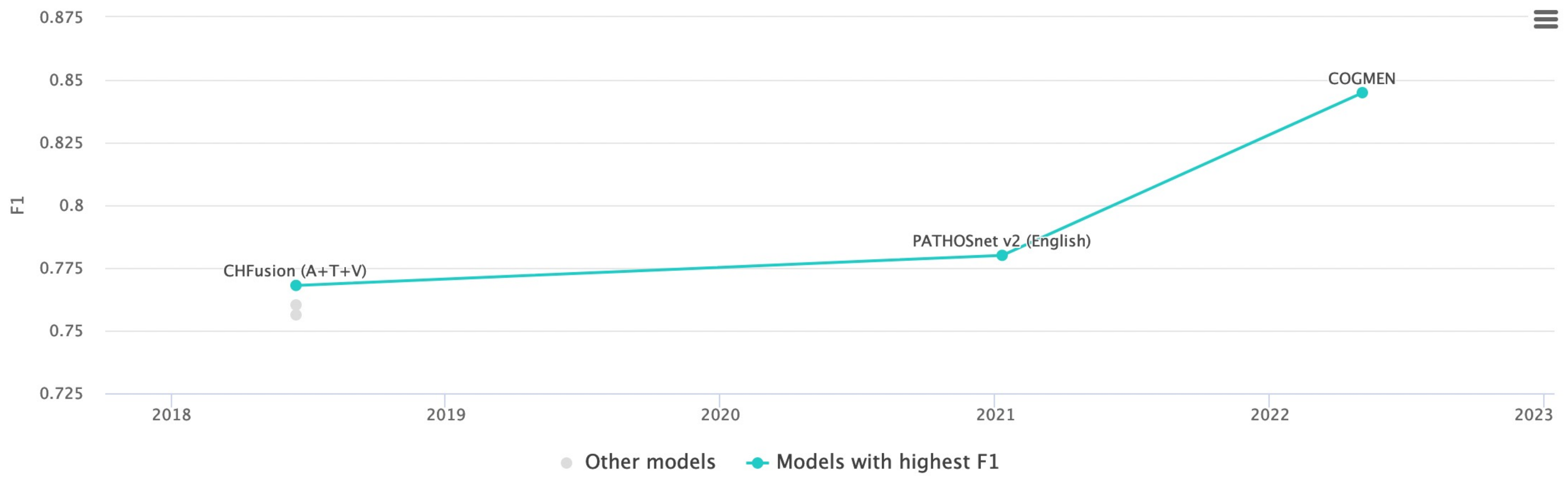
Models	IEMOCAP: Emotion Categories						Avg.	
	Happy	Sad	Neutral	Angry	Excited	Frustrated		
	F1 (%)	F1 (%)	F1 (%)	F1 (%)	F1 (%)	F1 (%)	Acc. (%)	F1 (%)
bc-LSTM	35.6	69.2	53.5	66.3	61.1	62.4	59.8	59
memnet	33	69.3	55	66.1	62.3	63	59.9	59.5
TFN	33.7	68.6	55.1	64.2	62.4	61.2	58.8	58.5
MFN	34.1	70.5	52.1	66.8	62.1	62.5	60.1	59.9
CMN	32.6	72.9	56.2	64.6	67.9	63.1	61.9	61.4
ICON	32.8	74.4	60.6	68.2	68.4	**66.2**	64	63.5
DialogueRNN	32.8	78	59.1	63.3	73.6	59.4	63.3	62.8
CAN	31.8	71.9	60.4	66.7	68.5	66.1	63.2	62.4
Af-CAN	37	72.1	60.7	**67.3**	66.5	66.1	64.6	63.7
COGMEN	**51.9**	**81.7**	**68.6**	66	**75.3**	58.2	**68.2**	**67.6**

Multimodal Emotion Recognition

Multimodal Emotion Recognition on IEMOCAP

Leaderboard   Dataset

View F1 by Date



State Of The Art Model

https://paperswithcode.com/sota/multimodal-emotion-recognition-on-iemocap

Experiments on MOSEI Benchmark

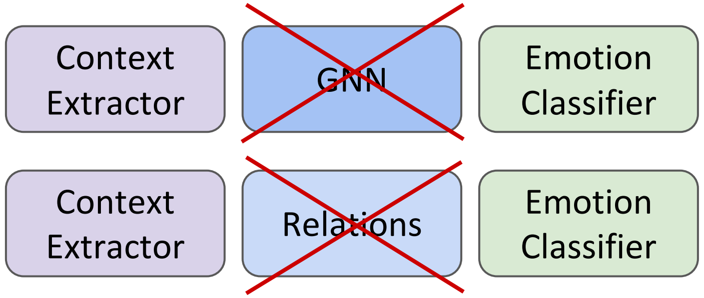
Model		Sentiment Class Accuracy(%)		Emotion Class (weighted) F1-score (%)						Multi-label Emotion Class (weighted) F1-score (%)					
Model		2 Class	7 Class	Happiness	Sadness	Angry	Fear	Disgust	Surprise	Happiness	Sadness	Angry	Fear	Disgust	Surprise
Multilogue-Net	T + A + V	82.88	**44.83**	67.84	65.34	67.03	87.79	74.91	**86.05**	70.6	70.7	74.4	86.0	83.4	87.8
TBJE	T	81.9	44.2	-	-	-	-	-	-	63.4	65.8	75.3	84.0	84.5	81.4
	A + T	82.4	43.91	65.91	70.78	70.86	87.79	82.57	86.04	65.5	67.9	76.0	**87.2**	84.5	86.1
	T + A + V	81.5	44.4	-	-	-	-	-	-	64.0	67.9	74.7	84.0	83.6	86.1
COGMEN	T	84.42	43.50	69.28	70.49	73.04	87.80	83.69	85.83	69.92	72.16	77.34	86.39	**86.00**	88.27
	A + T	**85.00**	44.31	68.39	**73.28**	74.98	88.08	**83.90**	85.35	69.62	72.67	76.93	86.39	85.35	88.21
	T + A + V	84.34	43.90	**70.42**	72.31	**76.20**	**88.17**	83.69	85.28	**72.74**	**73.90**	**78.04**	86.71	85.48	**88.37**

Comparison with Unimodal Approaches

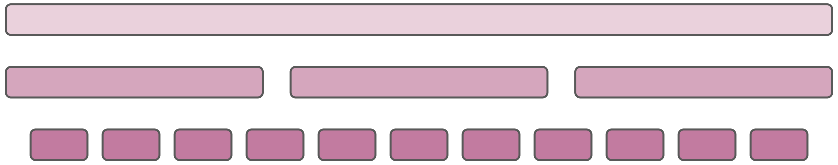
Model	Modality	F1-score (%)
4-way		
DialogueGCN	T	71.58
DialogXL	T	73.02
DAG-ERC	T	78.08
COGMEN	T	**81.55**
	A+T+V	**84.50**
6-way		
EmoBERTa	T	**68.57**
DAG-ERC	T	68.03
CESTa	T	67.10
SumAggGIN	T	66.61
DialogueCRN	T	66.20
DialogXL	T	65.94
DialogueGCN	T	64.18
COGMEN	T	66.00
	A+T+V	67.63

Importance of Local and Global Interactions

	Modalities	T	A+T	A+T+V
(6 way)	Actual	66.00	65.42	67.63
	w/o GNN	64.34 (↓1.66)	61.69 (↓3.73)	62.96 (↓4.14)
	w/o Relations	60.49 (↓5.51)	65.32 (↓0.10)	62.13 (↓5.50)
(4 way)	Actual	81.55	81.59	84.50
	w/o GNN	81.18 (↓0.37)	80.16 (↓1.43)	80.28 (↓4.22)
	w/o Relations	76.76 (↓4.79)	80.27 (↓1.32)	79.61 (↓4.88)

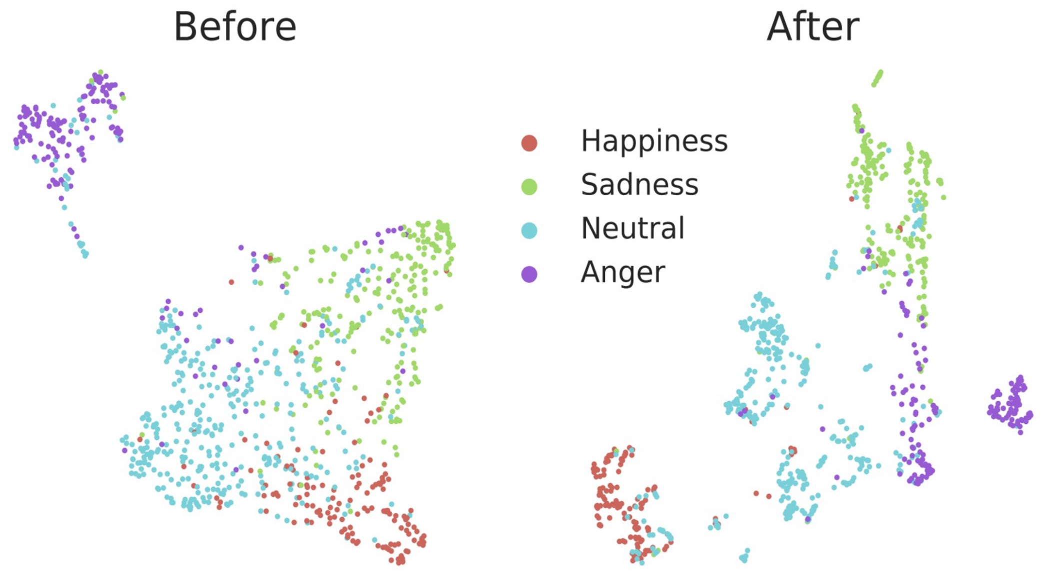


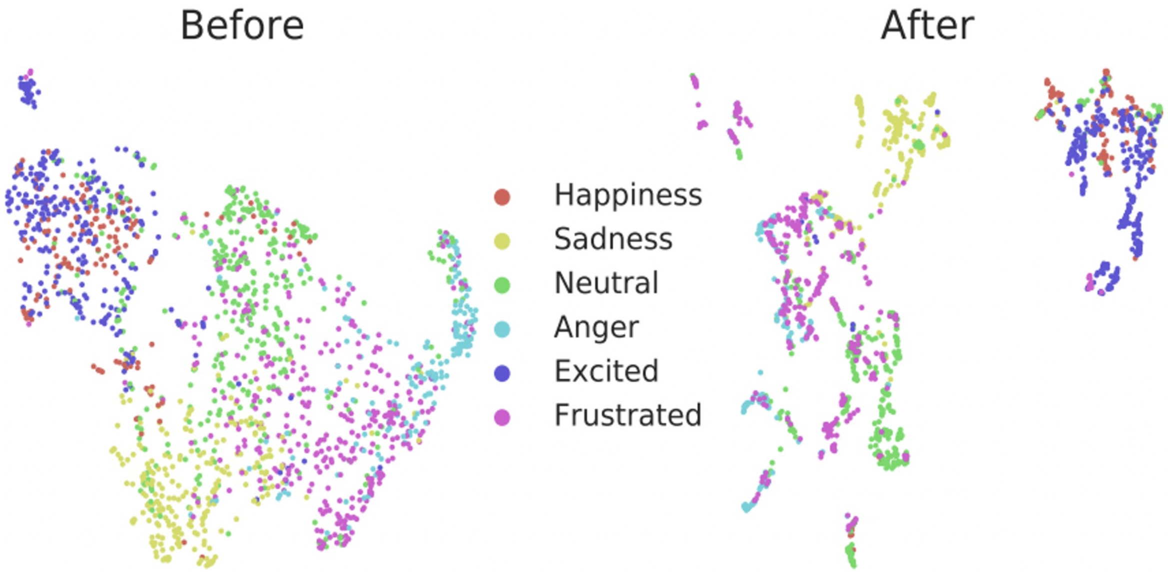
Dialogue divided into set of Utterances



# Utterances in Context	F1-score (%)
All Utterances in a dialogue	84.50
10 Utterances in a dialogue	77.43 (↓7.07)
3 Utterances in a dialogue	75.39 (↓9.11)

Effect of GNN on Multimodal Features





More details in the paper

Code Repository: https://github.com/Exploration-Lab/COGMEN

Special Thanks to Google Research India for the NAACL Travel Support

Google Research

Shapes of Emotions: Modeling Emotion Shift for Multimodal Emotion Recognition in Conversations

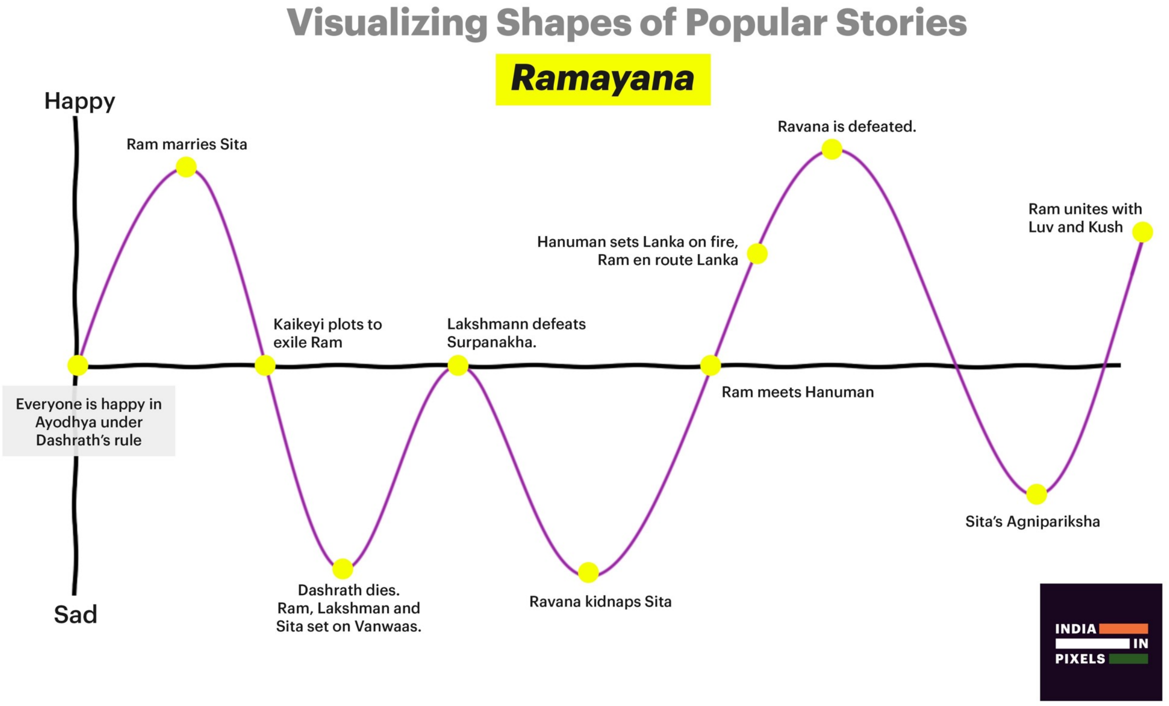
Keshav Bansal Harsh Agarwal Abhinav Joshi

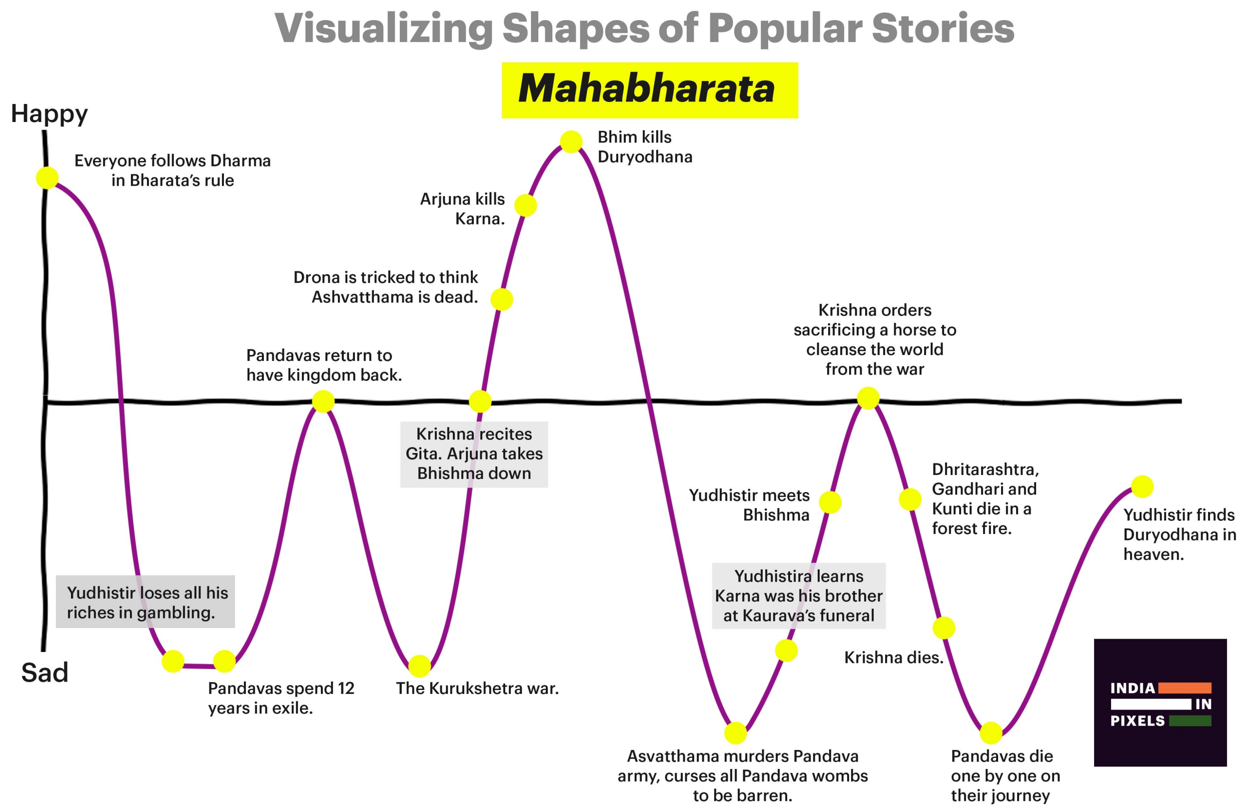
Ashutosh Modi

PIM3SM, CoLING 2022

Motivation: Shapes of Stories

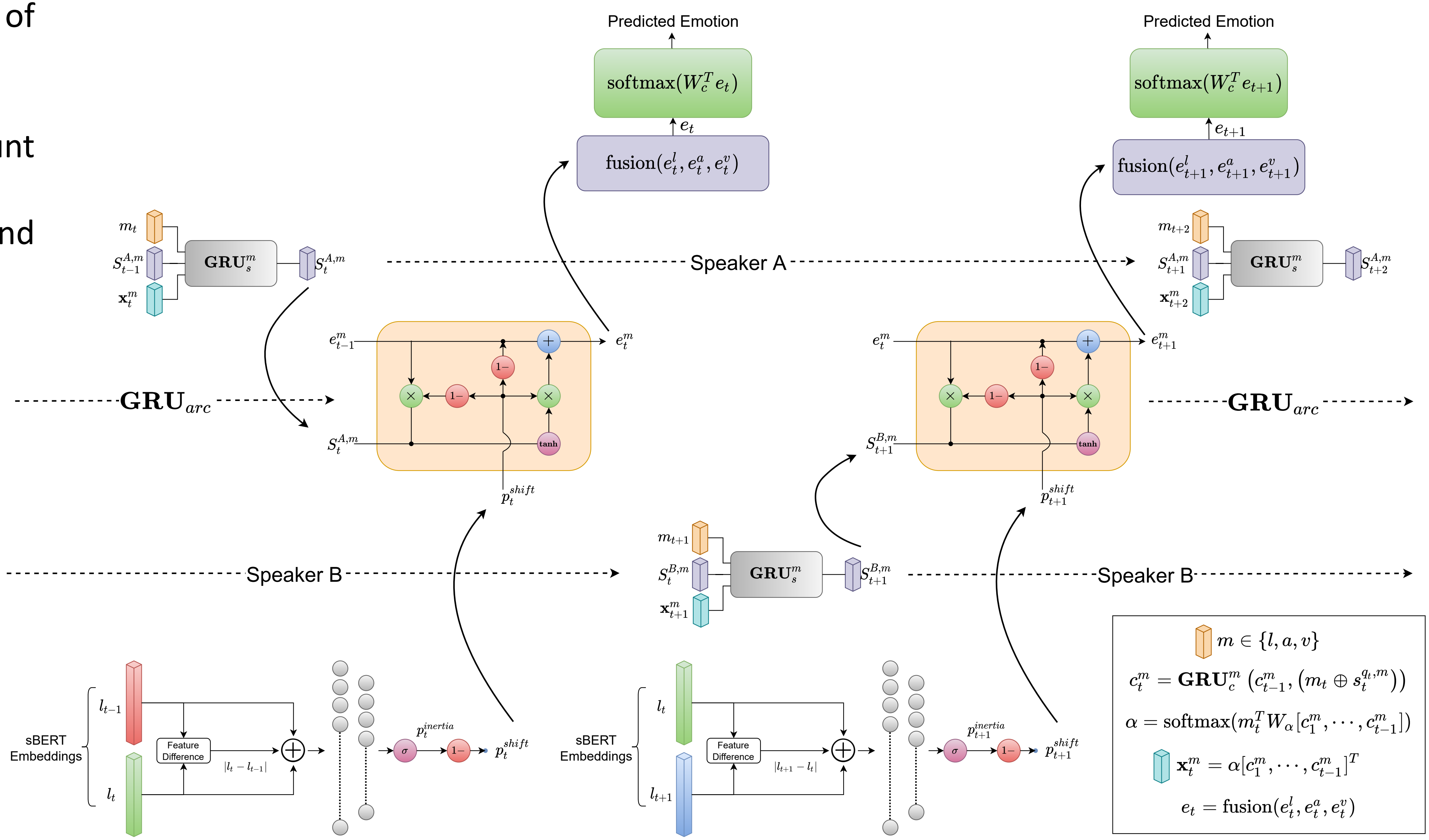
Kurt Vonnegut (Vonnegut, 1995) proposed that every story has a shape plotted by ups and downs experienced by the characters of the story. This defines the *Emotional Arc* of a story.





https://twitter.com/indiainpixels/status/1181567180829278215

- Model the ebb and flow of emotions
- Take into account speaker interactions and context



More details in the paper

Code Repository:

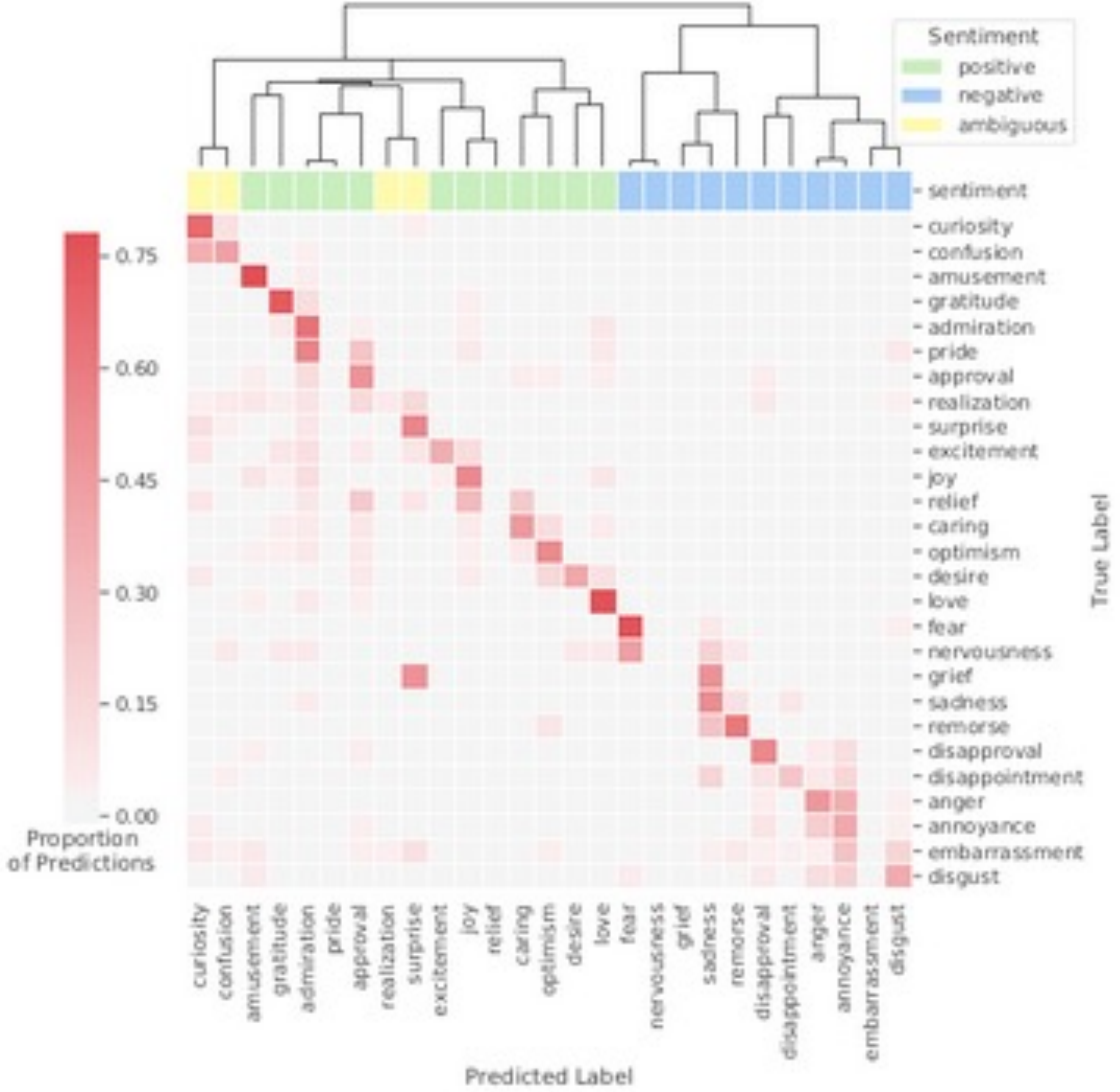
https://github.com/Exploration-Lab/multimodal-emo-prediction-with-emo-shift

Future Directions

- Structure of Emotions

Taxonomy

Embedding Emotions in Hyperbolic Spaces



Future Directions

- Structure of Emotions
- Emotion Cause Prediction

Multi-Task Learning Framework for Extracting
Emotion Cause Span and Entailment in Conversations
Ashwani Bhat and Ashutosh Modi
TL4NLP, NeurIPS, 2022
https://arxiv.org/abs/2211.03742

Future Directions

- Structure of Emotions
- Emotion Cause Prediction
- How does emotion play a role in decision making?

Future Directions

- Structure of Emotions
- Emotion Cause Prediction
- How does emotion play a role in decision making?
- Emotion AI for Indian Settings

Future Directions

- Structure of Emotions
- Emotion Cause Prediction
- How does emotion play a role in decision making?
- Emotion AI for Indian Settings
- Mental Health

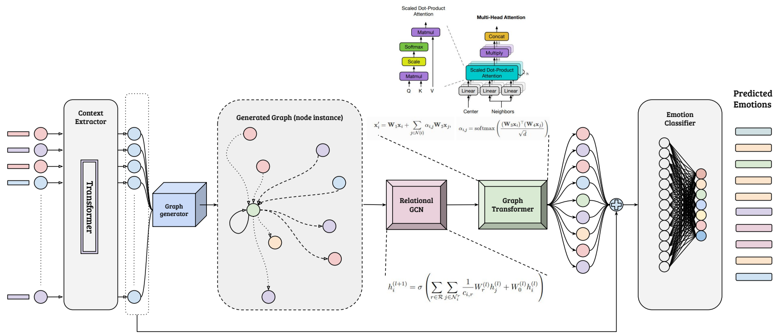
Fine Grained Emotion Prediction by Modeling Emotion Definitions
Best Student Paper Award, ACII , 2021

Adapting a Language Model for Controlled Affective Text Generation
CoLING, 2020

An End-to-End Network for Emotion-Cause Pair Extraction
WASSA, EACL, 2020

Shapes of Emotions: Multimodal Emotion Recognition in Conversati
PIM3SM, COLING 2022

Multi-Task Learning Framework for Extracting
Emotion Cause Span and Entailment in Conversations
TL4NLP, NeuRIPS 2022



COGMEN: COntextualized GNN based Multimodal Emotion recognitioN, NAACL, 2022

Modeling Human Behavior and Decision Making

Affective Computing

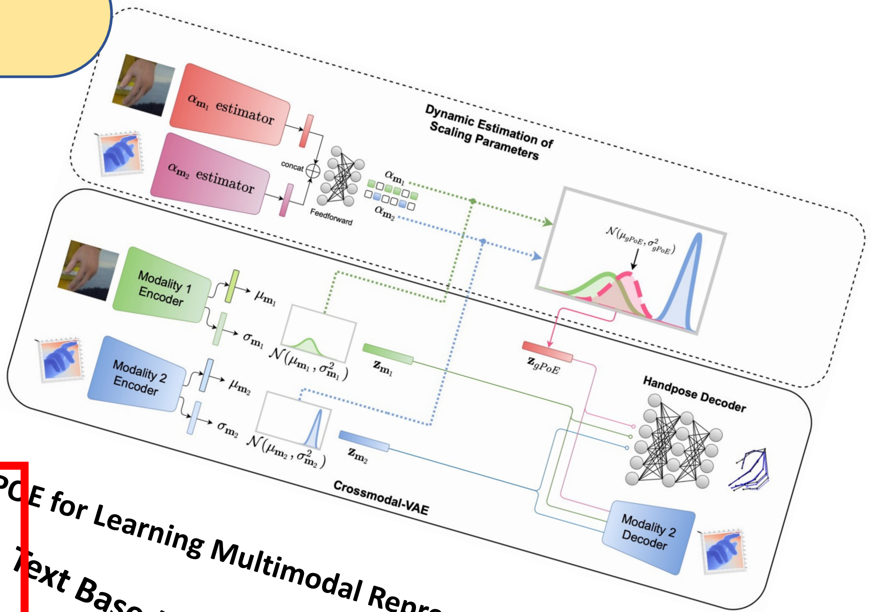
- Multimodal Representations
- Multimodal Multilingual Contextualized Affect Prediction
- Multimodal Generation
- Emotion and Decision Making: Emotion Cause Prediction

RL Worlds (Towards Embodied AI)

- Decision Making by Agents in Text Worlds
- Agents learn about real world without any explicit supervision via interactions with the environment simulating real world.

Mental Health

- Study correlation between speech, language, neuro-imaging, and **Schizophrenia symptoms**.



PoE for Learning Multimodal Representations in Noisy Environments
ICMI, 2022

ScriptWorld:
Text Based Environment For Learning Procedural Knowledge
AAMAS 2023,

Outstanding Paper Award, LAREL NeuRIPS 2022

Pre-Trained Language Models as Prior Knowledge for Playing Text Based Games
AAMAS, 2022

Ashutosh Modi

ScriptWorld: Text Based Environment For Learning Procedural Knowledge

Abhinav Joshi, Areeb Ahmad, Umang Pandey, Ashutosh Modi

LaREL, NeurIPS 2022 (Best Paper Runner-up)
AAMAS (EA) 2023
IJCAI, 2023

Teaching Daily Chores

- Can an agent learn to do the daily chores that humans do effortlessly without explicit supervision?

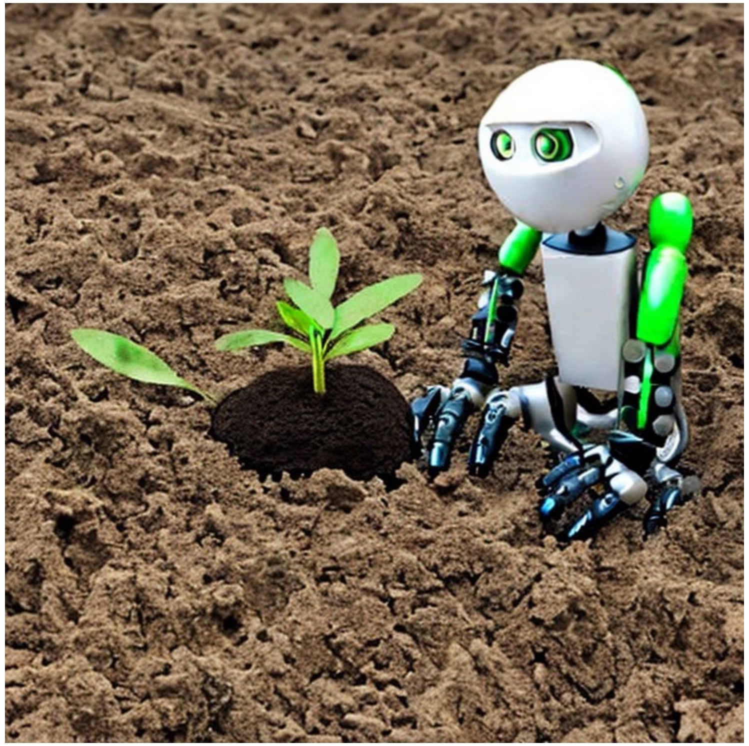


Image generated using Stable Diffusion

Teaching Daily Chores

- Can an agent learn to do the daily chores that humans do effortlessly without explicit supervision?

- Humans make use of the implicit common-sense knowledge about the world.

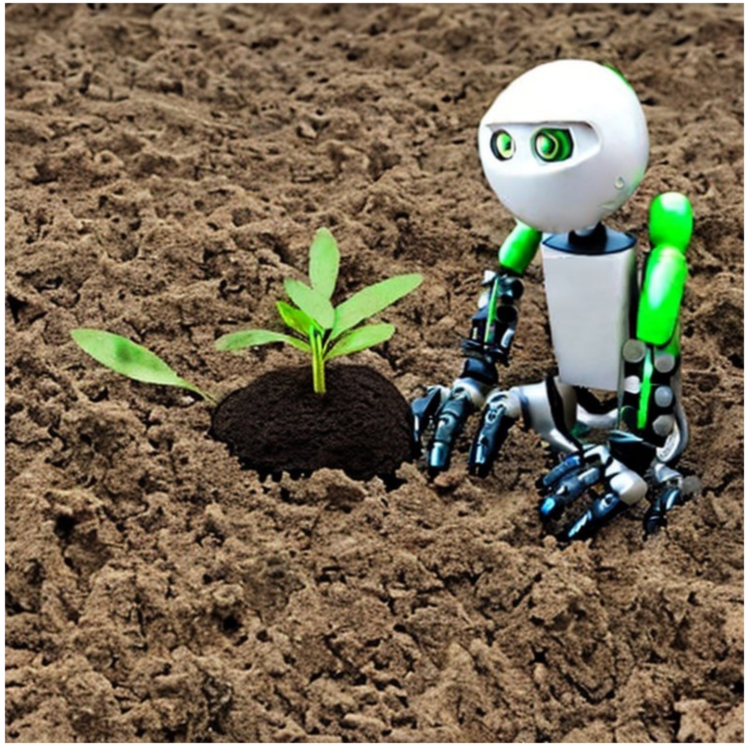


Image generated using Stable Diffusion

Teaching Daily Chores

- Can an agent learn to do the daily chores that humans do effortlessly without explicit supervision?
- Humans make use of the implicit common-sense knowledge about the world.
- What is the nature of this knowledge and how to impart it to agents?

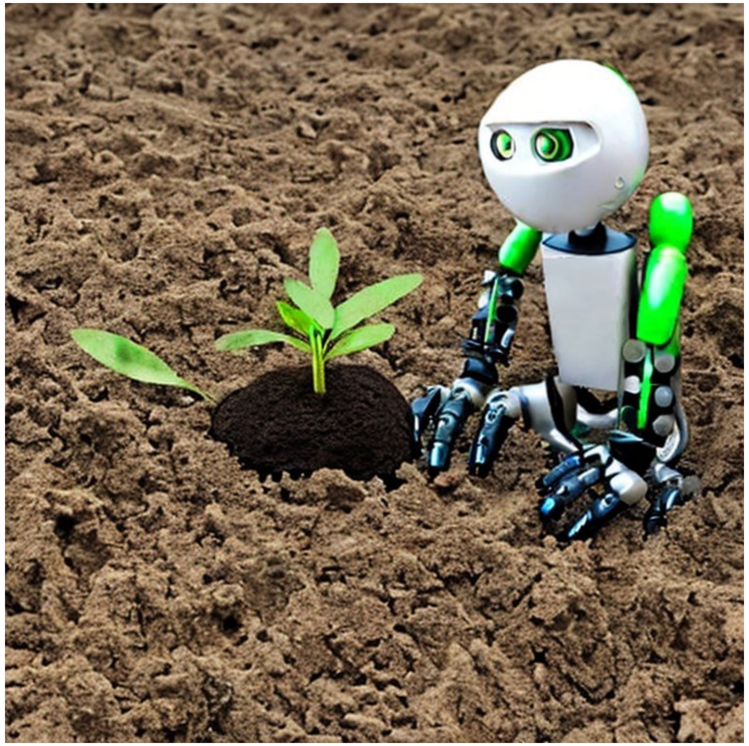


Image generated using Stable Diffusion

Scripts

Scripts are defined as sequences of actions describing stereotypical human activities, for example, cooking pasta, making coffee, etc.

(Schank and Abelson, 1975)

Washing Dishes

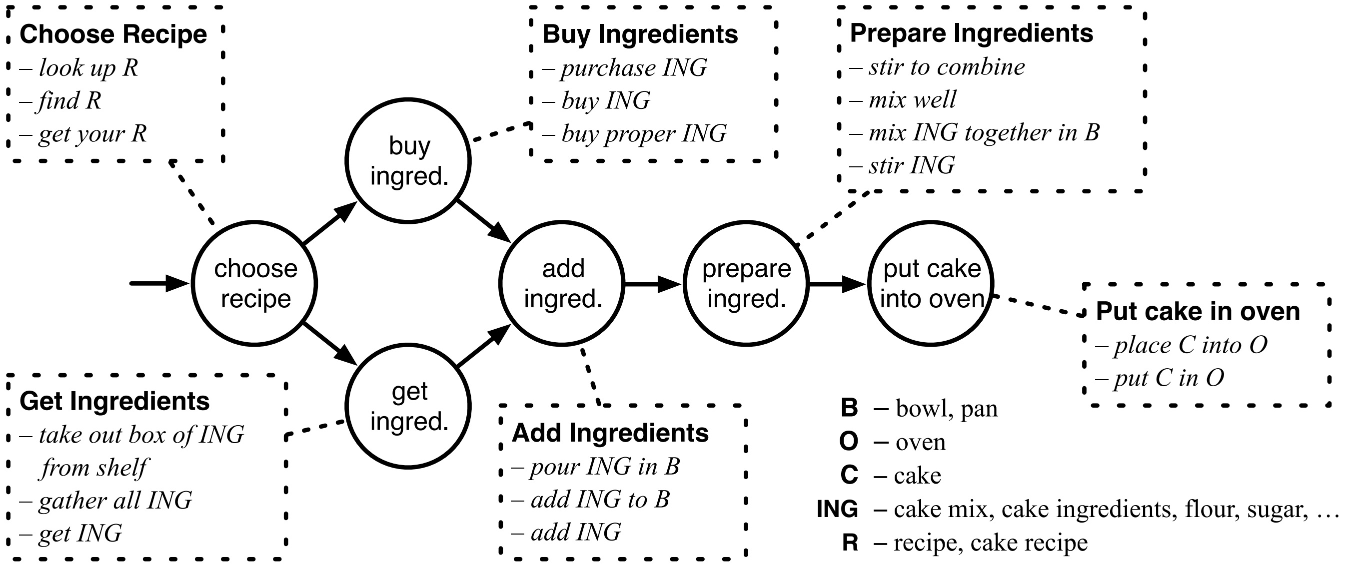
1. take dirty dishes to sink
2. run warm water into sink
3. add soap
4. scrub dishes with scrubber to remove food stains
5. rinse dishes
6. place clean dishes in rack to air dry

Event Sequence Description (ESD)

Scripts

Scripts are defined as sequences of actions describing stereotypical human activities, for example, cooking pasta, making coffee, etc.

(Schank and Abelson, 1975)



Wanzare et al., 2016

Scripts

Scripts are defined as sequences of actions describing stereotypical human activities, for example, cooking pasta, making coffee, etc.

(Schank and Abelson, 1975)

ESDs for Washing Dishes Scenario

DeScript Corpus

- A crowdsourced corpus capturing script knowledge for about 40 scenarios
- Each scenario is described by 100 participants ➔ 100 Event Sequence Descriptions (ESD)
- Semantically similar events manually aligned for 10 scenarios

Scenario
taking a bath
baking a cake
flying in an airplane
going grocery shopping
going on a train
planting a tree
riding on a bus
repairing a flat bicycle tire
borrowing a book from the library
getting a hair cut

Wanzare et al., 2016

ScriptWorld

- Text based Environment for teaching common sense (script) knowledge about the world to agents

ScriptWorld

- Text based Environment for teaching common sense (script) knowledge about the world to agents

- Design Choices
  - Complexity
  - Flexibility
  - Grounded in real world

ScriptWorld

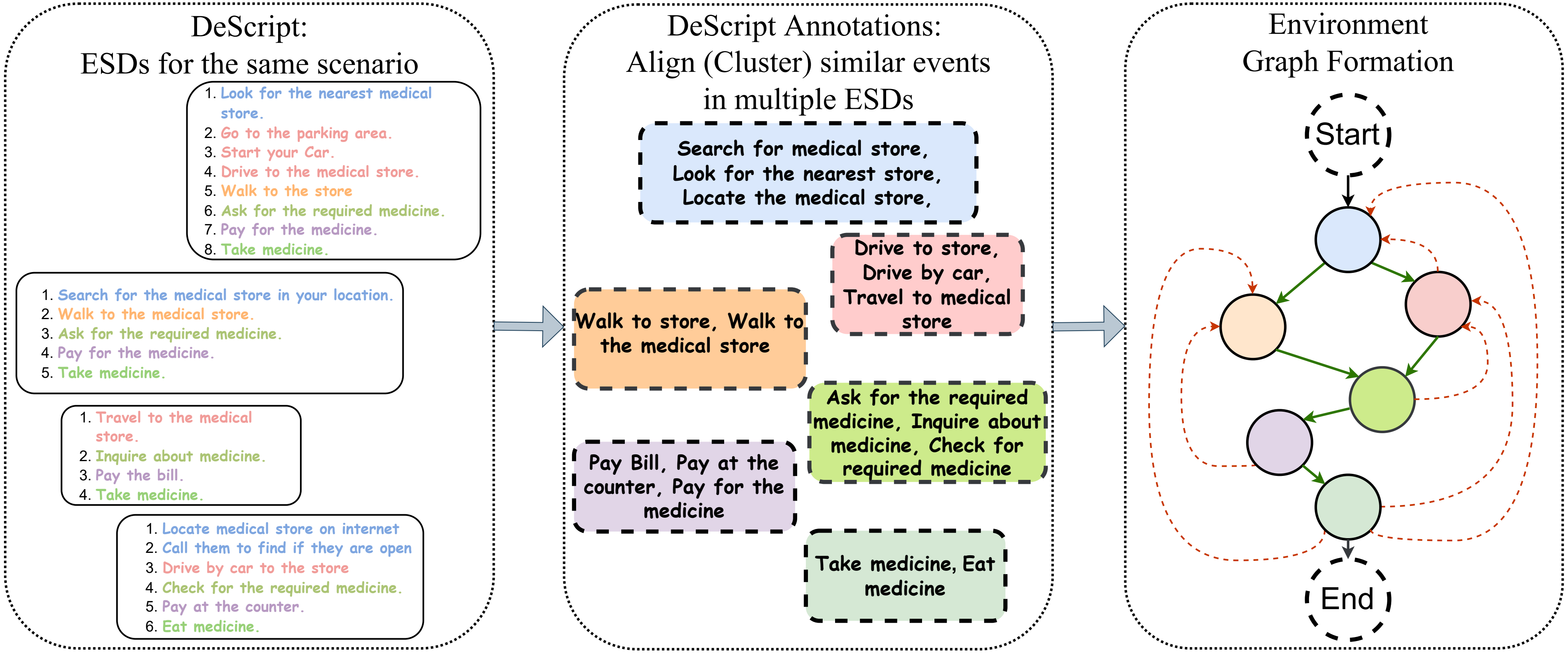
- Text based Environment for teaching common sense (script) knowledge about the world to agents

- Solving the task requires an agent to maintain a memory and to take complex sequential decisions in a dynamic environment.

ScriptWorld

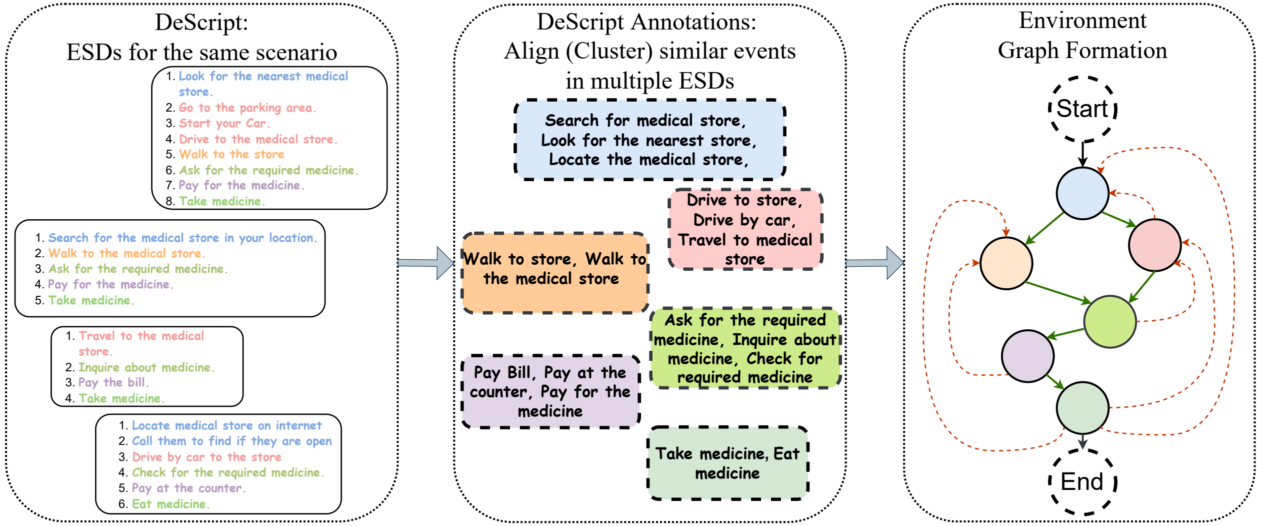
- Text based Environment for teaching common sense (script) knowledge about the world to agents
- Solving the task requires an agent to maintain a memory and to take complex sequential decisions in a dynamic environment.
- Step towards Embodied AI and towards creation of agents in the MetaVerse

ScriptWorld Creation



ScriptWorld

Scenario	Nodes	Deg.	Paths
Taking a **Bath**	525	3.7	$3.1e+27$
Baking a **Cake**	542	3.6	$4.0e+26$
Flying in an **Airplane**	528	3.6	$2.6e+30$
Going Grocery **Shopping**	544	3.7	$2.3e+26$
Going on a **Train**	427	3.7	$3.1e+21$
Planting a **Tree**	373	3.7	$1.6e+16$
Riding on a **Bus**	376	3.8	$1.0e+17$
Repairing Flat **Bicycle** Tire	402	3.4	$8.4e+18$
Borrowing Book from **Library**	397	3.7	$3.1e+19$
Getting a **Haircut**	528	3.7	$4.0e+28$



```
Point Acquired : 0
Total reward :  -1
Lives Left :  4
Percentage completion: 87.5 %
 ███████████████████████████████████████████████████████████████████████████████████            |  87.5 %
 ***************************************************************************************************
 ******************************   going grocery shopping  ******************************
 ***************************************************************************************************

HINT :  leave

 ***************************************************************************************************

ACTIONS:


 0  :  Go shopping


 1  :  Take a cart


 2  :  Leave


 3  :  Make a list of items you need at the grocery


 4  :  Place items in cart on belt for cashier to scan



Choose an Action: 2
You Chose :  Leave
Point Acquired : 10
Total reward :  9
Lives Left :  4
Percentage completion: 100.0 %
 ████████████████████████████████████████████████████████████████████████████████████████████████|  100.0 %

 ***************************************************************************************************

 ******************************************************************************** *******************

 ******************************************   Right Answer!   ******************************************
 ***************************************************************************************************
```

```
Point Acquired : 0
Total reward :  -1
Lives Left :  4
Percentage completion: 75.0 %
 ███████████████████████████████████████████████████████████████████████████                        |  75.0 %
 ***************************************************************************************************
 ******************************   going grocery shopping  ******************************
 ***************************************************************************************************

HINT : get your receipt

 ***************************************************************************************************

ACTIONS:


 0  :  Get the bill for groceries


 1  :  Make a grocery list


 2  :  Turn the car on


 3  :  Take list to store


 4  :  Drive to store



Choose an Action: 0
You Chose :  Get the bill for groceries
Point Acquired : 0
Total reward :  -1
Lives Left :  4
Percentage completion: 87.5 %
 ███████████████████████████████████████████████████████████████████████████████████████           |  87.5 %

 ***************************************************************************************************

 ******************************************************************************** *******************

 ******************************************   Right Answer!   ******************************************
 ***************************************************************************************************
```

Baseline RL Agents

DQN

A2C

PPO

RPPO

Aim:
Learn
$q_{\prod}(s, a)$ or $\prod(a \mid s)$

Baseline RL Agents

For reinforcement learning baselines, we consider pre-trained SBERT language model as a source of prior real-world knowledge, which could be used directly in RL algorithm

DQN

A2C

PPO

RPPO

Baseline RL Agents

DQN

A2C

PPO

RPPO

For reinforcement learning baselines, we consider pre-trained SBERT language model as a source of prior real-world knowledge, which could be used directly in RL algorithm

We consider a generalized scheme where a pre-trained language model is used to extract information from the observations, i.e., the available set of choices.

Baseline RL Agents

DQN

A2C

PPO

RPPO

For reinforcement learning baselines, we consider pre-trained SBERT language model as a source of prior real-world knowledge, which could be used directly in RL algorithm

We consider a generalized scheme where a pre-trained language model is used to extract information from the observations, i.e., the available set of choices.

In the generalized scheme, a pre-trained language model generates embeddings corresponding to each of the provided options

Baseline RL Agents

DQN

A2C

PPO

RPPO

For reinforcement learning baselines, we consider pre-trained SBERT language model as a source of prior real-world knowledge, which could be used directly in RL algorithm

We consider a generalized scheme where a pre-trained language model is used to extract information from the observations, i.e., the available set of choices.

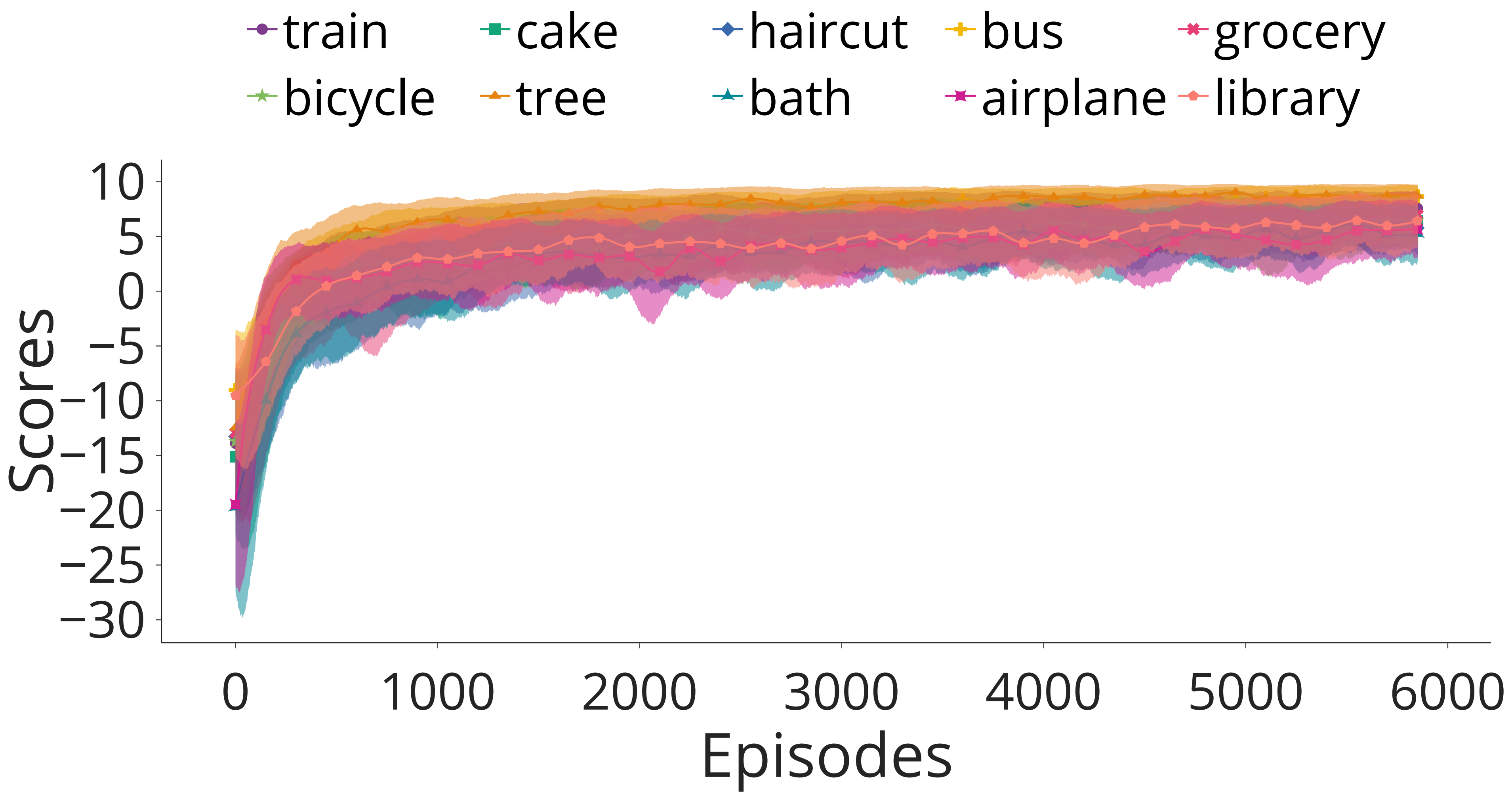
In the generalized scheme, a pre-trained language model generates embeddings corresponding to each of the provided options

The obtained embeddings are concatenated and passed as input to the learning frame-work

Agent Performance

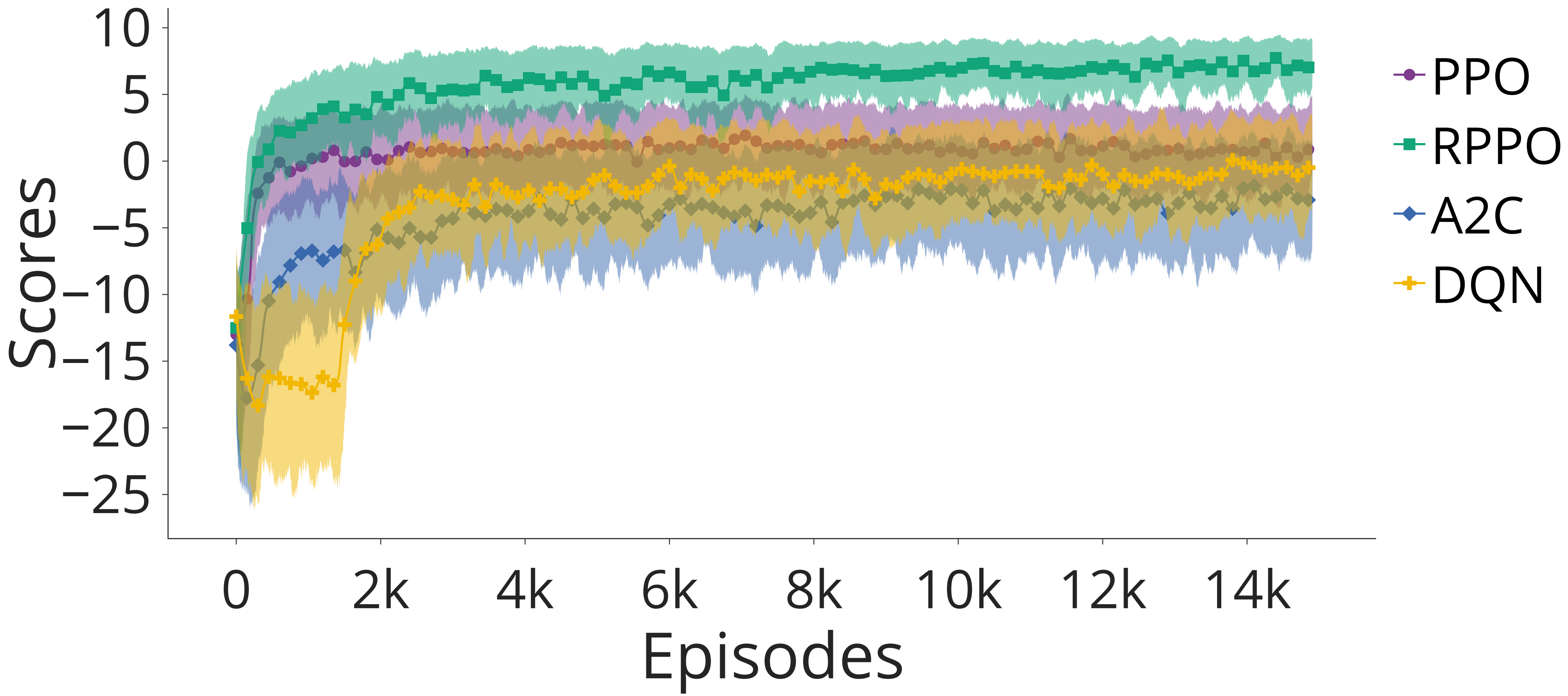
Algorithm	DQN		A2C		PPO		RPPO	
	handicap	w/o handicap	handicap	w/o handicap	handicap	w/o handicap	handicap	w/o handicap
Shopping	9.60 (± 0.62)	-7.28 (± 13.15)	9.90 (± 0.30)	-9.81 (± 14.71)	9.84 (± 0.39)	-4.78 (± 10.79)	9.71 (± 0.57)	**8.79 (± 4.15)**
Bus	8.98 (± 0.79)	-1.47 (± 11.16)	9.89 (± 0.34)	-7.37 (± 17.09)	9.93 (± 0.25)	1.50 (± 7.50)	9.97 (± 0.17)	**9.32 (± 1.24)**
Train	9.21 (± 2.07)	-3.10 (± 11.16)	9.89 (± 0.31)	-8.13 (± 14.99)	9.75 (± 0.49)	-1.13 (± 9.47)	9.56 (± 0.80)	**8.19 (± 4.70)**
Library	9.51 (± 0.68)	-1.94 (± 9.87)	9.88 (± 0.32)	-3.03 (± 9.84)	9.90 (± 0.30)	1.12 (± 7.31)	9.89 (± 0.31)	**8.41 (± 4.77)**
Haircut	9.88 (± 0.35)	-9.30 (± 12.93)	9.89 (± 0.34)	-5.87 (± 12.28)	9.85 (± 0.38)	-4.30 (± 10.84)	9.63 (± 0.64)	**6.32 (± 5.29)**
Cake	9.32 (± 0.84)	-4.13 (± 9.22)	9.48 (± 0.92)	-7.58 (± 13.18)	9.87 (± 0.34)	-4.46 (± 12.32)	9.78 (± 0.48)	**7.18 (± 4.97)**
Bicycle	9.50 (± 0.75)	0.07 (± 7.89)	9.95 (± 0.22)	-3.49 (± 12.39)	9.90 (± 0.33)	1.17 (± 6.93)	9.74 (± 0.57)	**7.85 (± 5.12)**
Tree	9.94 (± 0.24)	-0.15 (± 7.83)	9.86 (± 0.44)	-3.54 (± 12.56)	9.98 (± 0.14)	1.43 (± 7.29)	9.96 (± 0.19)	**8.88 (± 3.23)**
Airplane	9.68 (± 0.75)	-4.21 (± 12.39)	9.86 (± 0.35)	-8.66 (± 12.66)	9.86 (± 0.40)	-4.74 (± 11.08)	9.54 (± 0.73)	**6.85 (± 6.12)**
Bath	9.68 (± 0.61)	-6.49 (± 13.23)	9.75 (± 0.57)	-10.02 (± 15.95)	9.84 (± 0.37)	-5.35 (± 11.19)	9.45 (± 0.82)	**6.35 (± 5.59)**

Agent Performance



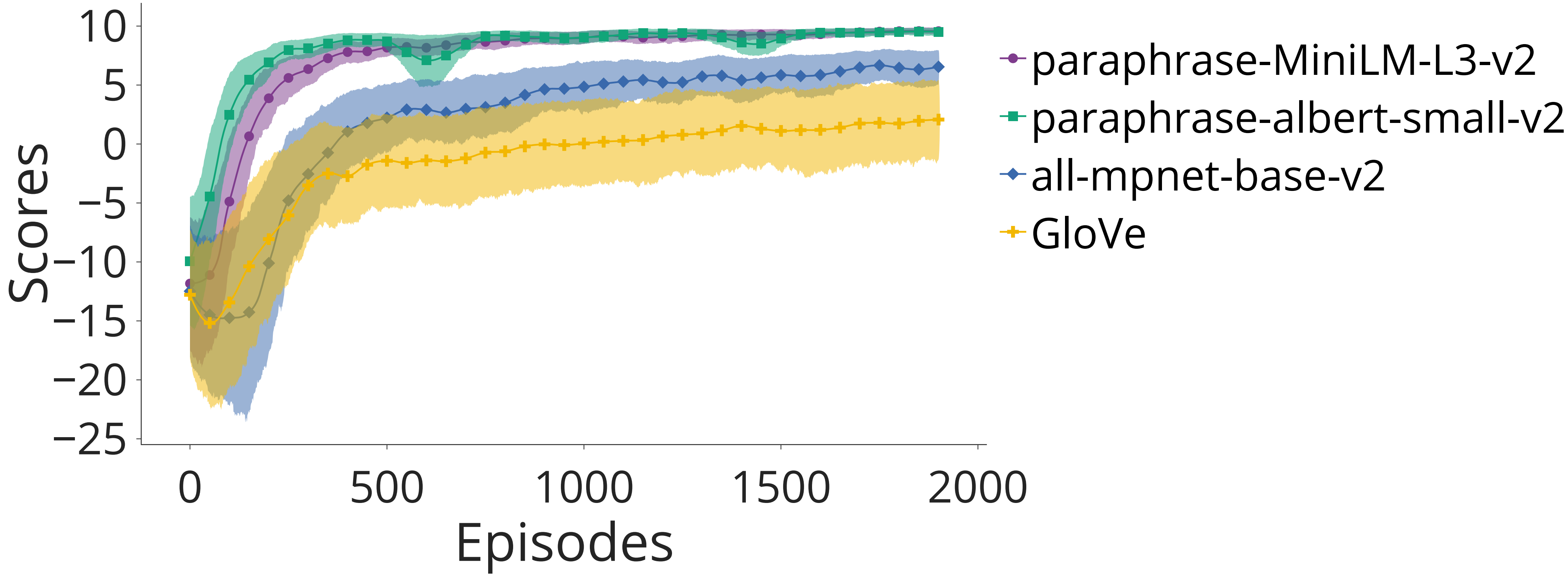
RPPO, Choices =2, without Handicap

Agent Performance



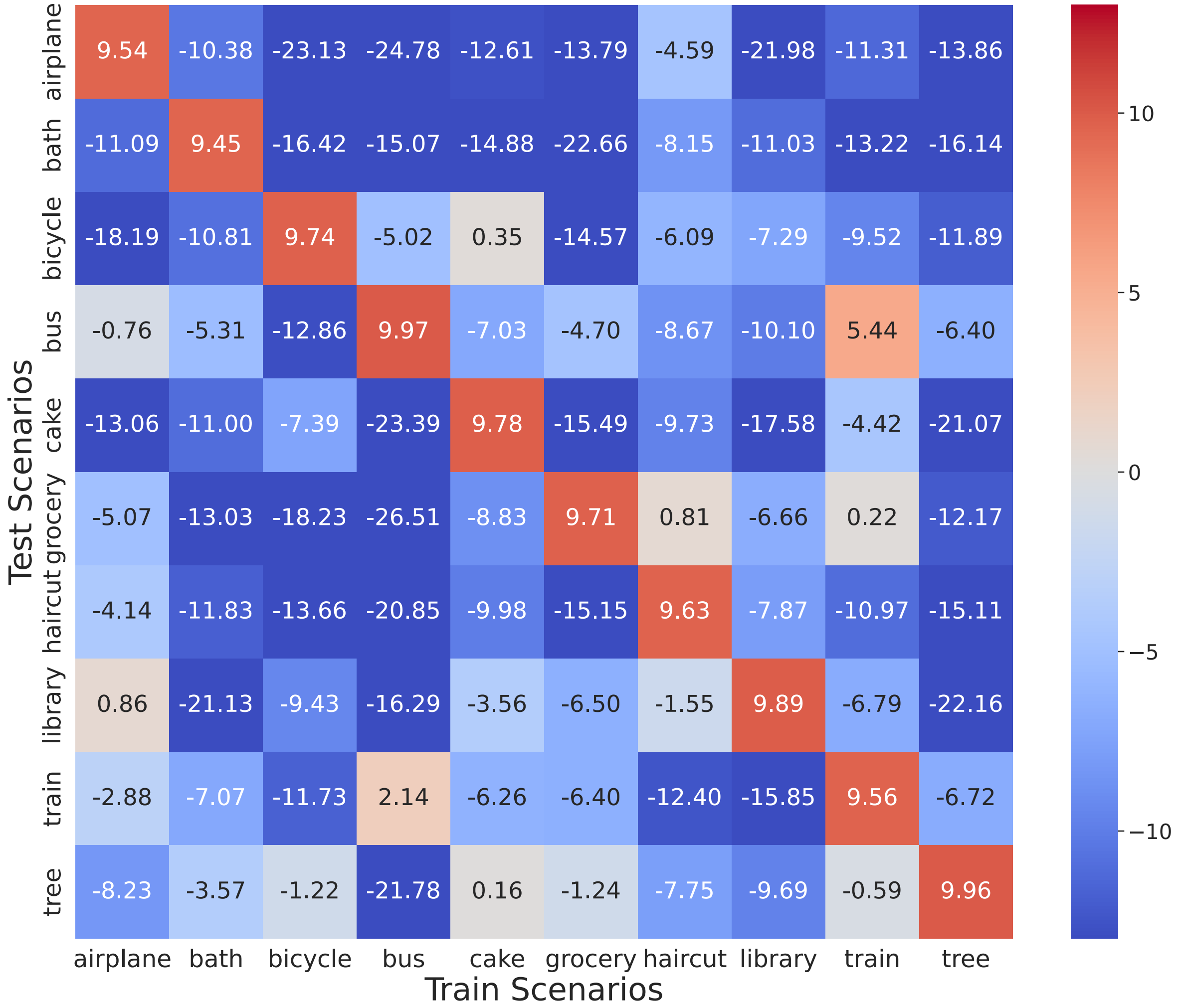
Scenario= Repairing Bike Flat Tire, Choices =2, with Handicap

Effect of Language Model



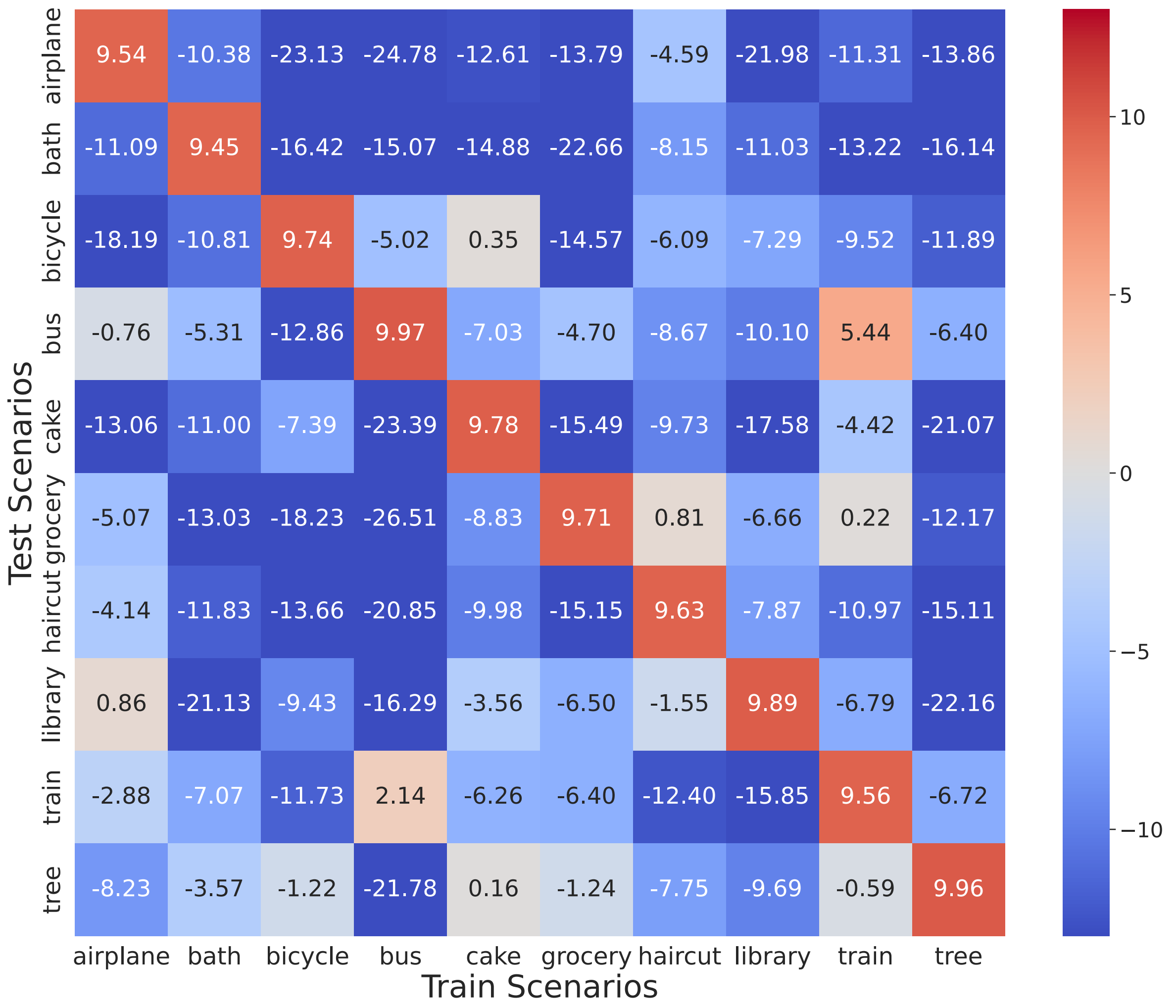
RPPO, Scenario= Repairing Bike Flat Tire, Choices =2, with Handicap

Generalization Across Scenarios

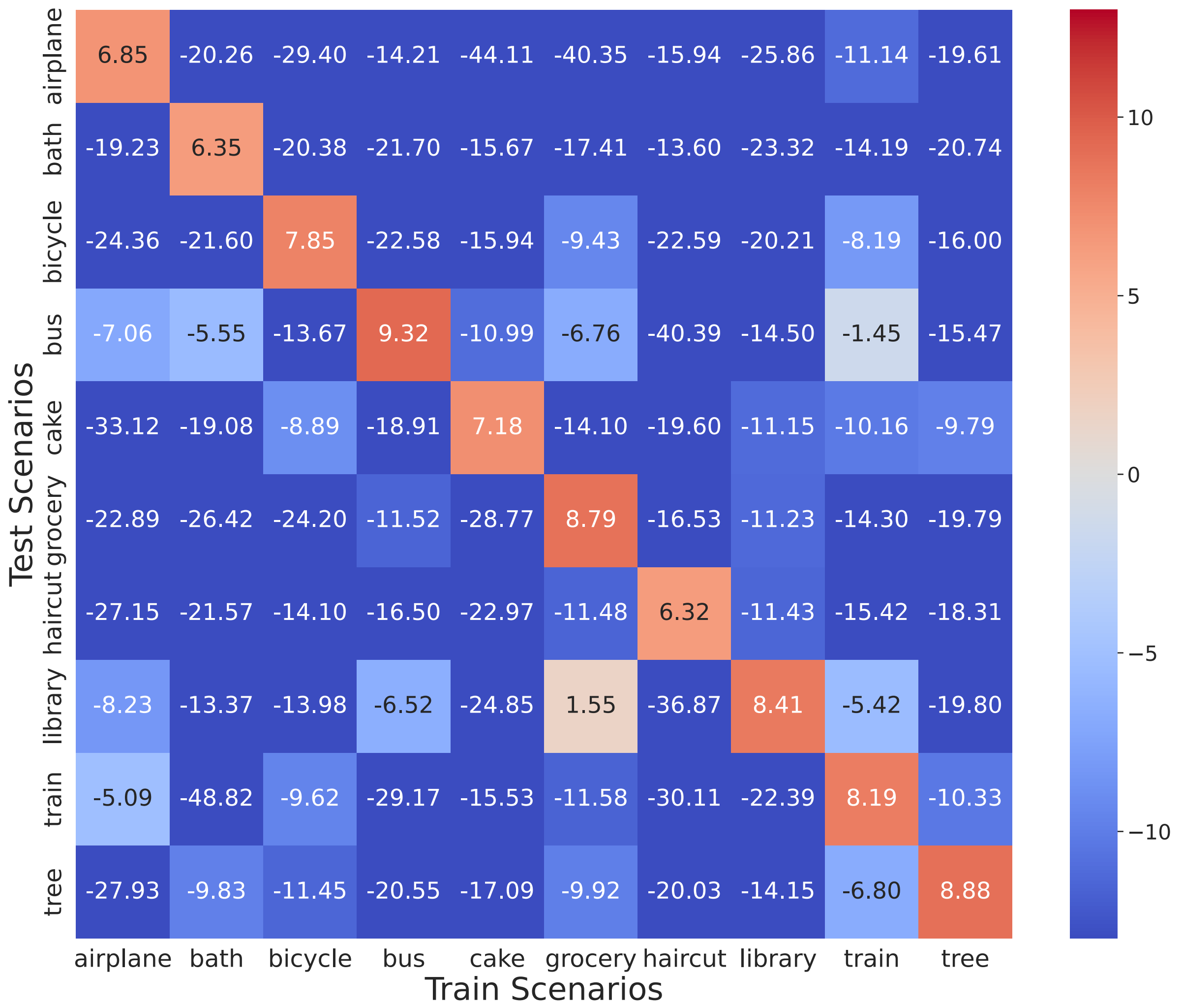


With Handicap

Generalization Across Scenarios (2 Choices)

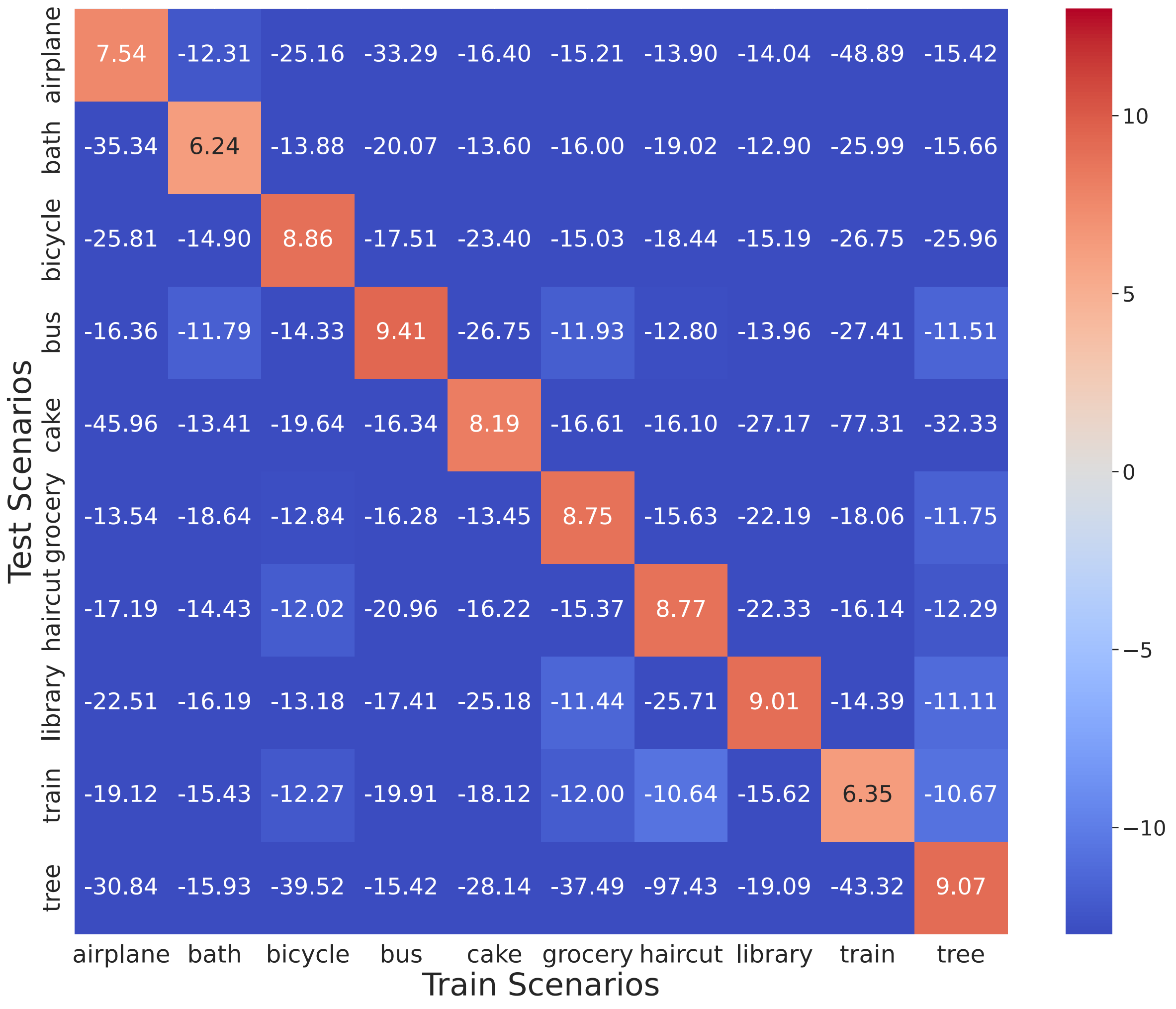


With Handicap

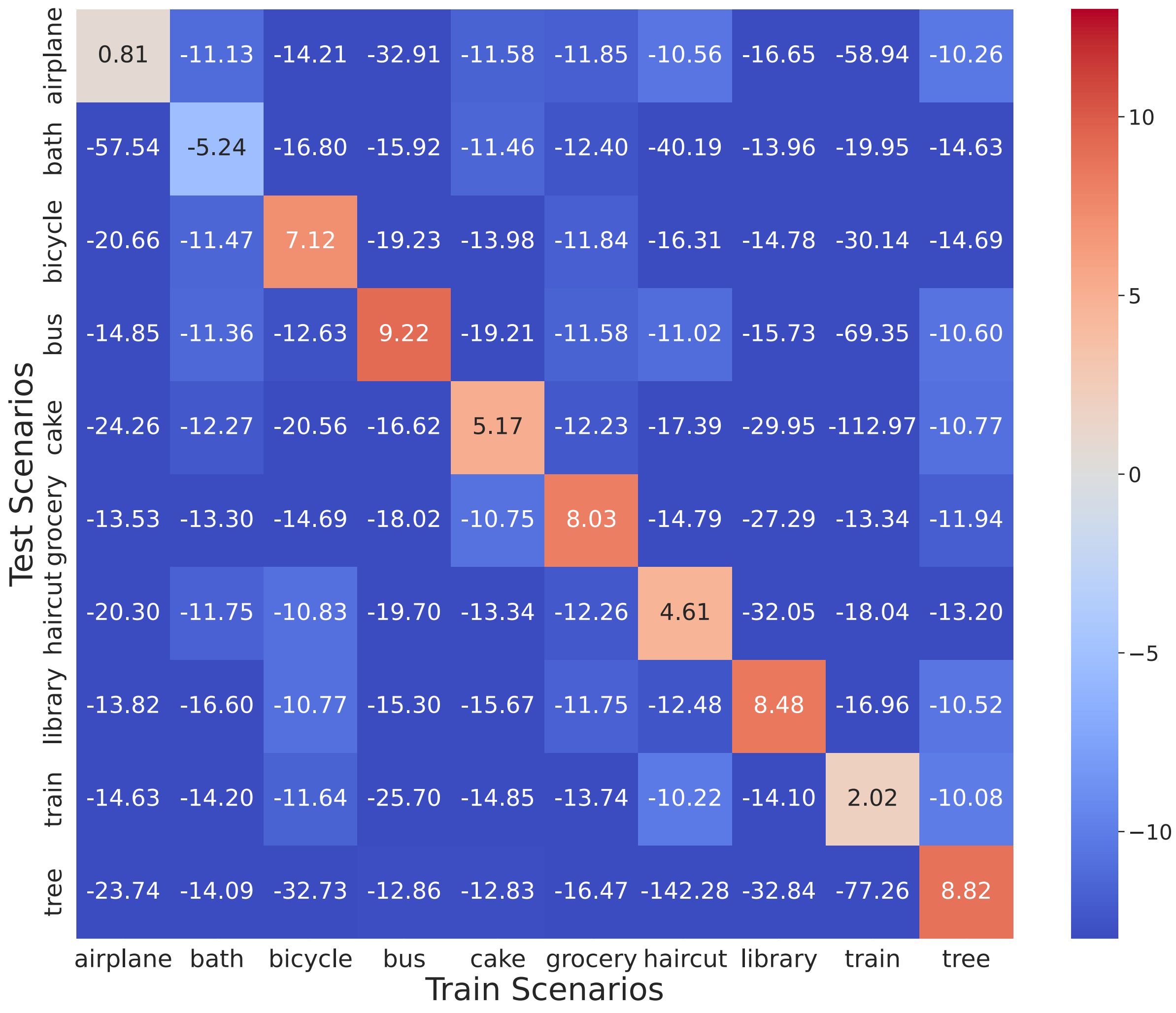


Without Handicap

Generalization Across Scenarios (5 Choices)

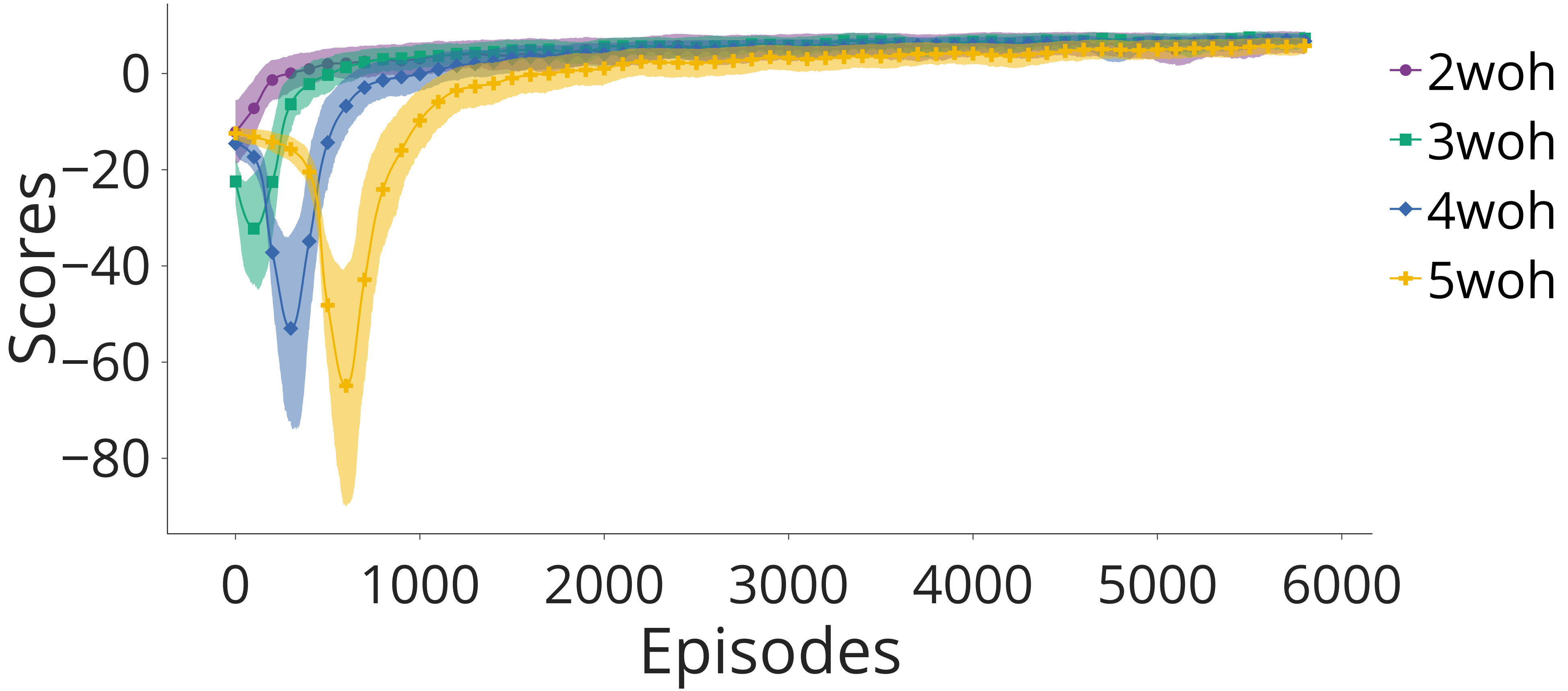


With Handicap



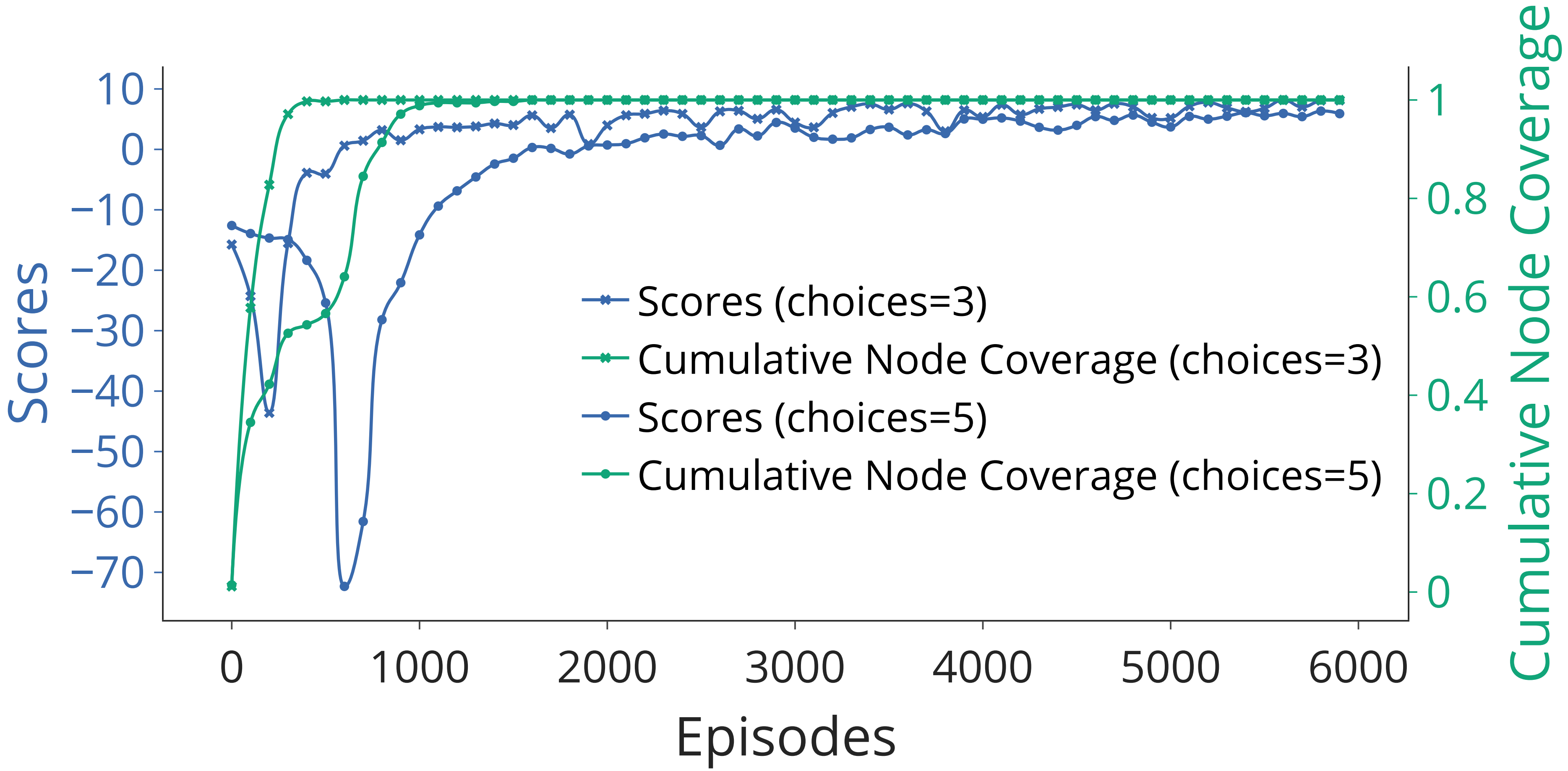
Without Handicap

Effect of Number of Choices



RPPO, Scenario= Repairing Bike Flat Tire, without Handicap

Effect of Choices



RPPO, Scenario= Repairing Bike Flat Tire, without Handicap

ScriptWorld

- ScriptWorld: An environment for teaching procedural knowledge to agents
- Prior knowledge obtained from a pre-trained language model helps to solve real-world text-based gaming environments.
- Agent are still not able to solve the environment completely
- Development of Parser based environment that allows free-form text as action
- More scenario coverage required

More details in the paper

Code Repository:

https://github.com/Exploration-Lab/ScriptWorld

Future Directions

- Multimodal Environment

Source: https://www.quantamagazine.org/ai-makes-strides-in-virtual-worlds-more-like-our-own-20220624/

Future Directions

- Multimodal Environment
- Hierarchical Learning in Agents

Future Directions

- Multimodal Environment
- Hierarchical Learning in Agents
- Self Learning Agents

NLP and NLU

Legal NLP
Understanding and Processing Indian Legal Texts, Legal Foundational models, Summarization, Cross-Lingual, Cross Domain Knowledge Transfer, Legal KG

Natural Language Retrieval
Retrieving information from databases via natural language queries

Biomedical NLP
NER, Relation Extraction, Clinical Trials….

Machine Unlearning
Forgetting Unwanted information in LLMs, Updating LLMs with latest facts without training

Social Reasoning in LLMs
Teaching ethics and etiquettes to LLMs

Miscellaneous
Automatic Speech Recognition for noisy, code-mixed speech

Modeling Human Behavior and Decision Making

Affective Computing
- Multimodal Representations
- Multimodal Multilingual Contextualized Affect Prediction
- Multimodal Generation
- Emotion and Decision Making: Emotion Cause Prediction

RL Worlds (Towards Embodied AI)
- Decision Making by Agents in Text Worlds
- Agents learn about real world without any explicit supervision via interactions with the environment simulating real world.

Mental Health
- Study correlation between speech, language, neuro-imaging, and **Schizophrenia symptoms**.

AI For Social Good

Sign Language Translation and Generation
- Sign language understanding
- Linguistic Analysis
- NLP Tools for Sign Language
- Translation within sign languages and with natural language
- Generation conditioned on context and other modalities

Contact: ashutoshm@cse.iitk.ac.in

Lexaii Google INTUIT turbotax credit karma quickbooks mailchimp Miimansa CONVIN Microsoft SERB INDIA DRDO

If you are interested in exploring the world with AI

Exploration Lab
अन्वेषण प्रयोगशाला

Openings: MSR/Ph.D./PostDoc

Contact: ashutoshm@cse.iitk.ac.in