Inducing Politeness, Empathy and Multimodal Knowledge in Conversational AI

Asif Ekbal
AI-NLP-ML Research Group
Department of Computer Science and Engineering
IIT Patna, Patna, India
Email: asif.ekbal@gmail.com, asif@iitp.ac.in

GIAN Workshop on Deep Learning Techniques on Conversational AI

IIT Patna

April 22, 2022

Slides scribed by: Mauzama, Deeksha, Arindam

Outline

- Background: Conversational AI
- Empathetic Dialogue Agents: Inducing Courteousness, Sentiment, Emotion in Conversational Agents
- Personalization in Conversational Agents
- Inducing multimodal knowledge in conversational AI
- Our experience towards building Indian Language Conversational AI
- Summary and Conclusion
- Our Research at a Glance

Conversational AI: A brief Introduction

Artificial Intelligence and Conversational Agents

- ***Artificial intelligence (AI)*** is one of the most-discussed technology topics among the researchers, consumers and enterprises today

- **Conversational AI powered by NLP and ML** has been in the centre of AI revolution during the last few years

Examples: Conversational AI Systems

Phone-based Personal Assistants

SIRI, Cortana, Google Now

Talking to your car

Communicating with robots

Clinical uses for mental health

Chatting for fun

The most simplest form of Conversational System: Chatbot

The chatbot market size is projected to grow from $2.6 billion in 2019 to $9.4 billion by 2024 at a compound annual growth rate (CAGR) of 29.7% (BusinessInsider)

Today's Chatbot: A Long way from ELIZA (1960)

- Nowadays, ***Chatbots*** have grown into a full-blown industry with constant innovations bridging the human-to-machine communication gap
  - *Going beyond simple tasks like playing a song or booking an appointment*

- Beyond knowledge-based conversational agents that match a query to a predefined set of answers

- Chatbot should mimic the dynamics of human conversations

BUT how?

Today's Chatbot: A Long way from ELIZA

- **Generating coherent and engaging responses in conversations**
  - Through Deep Language Understanding and Reasoning

- Should understand a user's need, context and mood
- Should be able to respond with *personalization, sentimental and emotional analysis*
- Balancing human-like aspects such as **specificity** and **empathy**

- **Need advanced NLP and ML Systems**
  - Beyond understanding a single sentence or taking discrete actions
  - Understanding long-form sentences in specific contexts

Empowering AI for Human-like Conversation

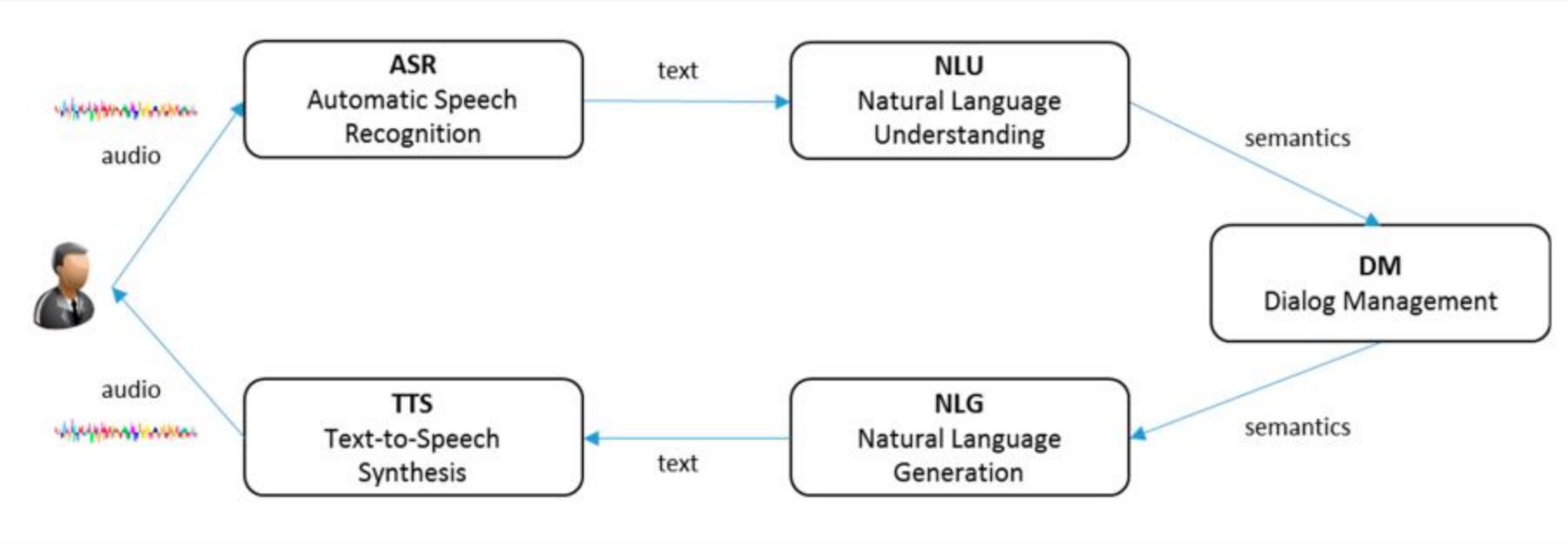
- **AI has to master the art of conversation at human level, then it has an uphill task ahead (*Facebook AI*)**
  - Consistency: *to ensure that it generates appropriate response without missteps, such as contradictions*
  - Specificity: *generating specific response*
  - **Empathy:**

    Affect-awareness (Sentiment-aware, Emotion-aware), Courteousness etc.
  - Knowledgeability: *should be able to take into account the external knowledge and facts, and generate response accordingly*
  - Multimodal understanding: *should be able to operate with text, image, audio, video etc.*

Dialogue Agents: Types

- **Open Chit-Chat Agents (Open IE)**
  - Designed for extended conversations, set up to mimic the unstructured conversational or ‘chats’ characteristic of human-human interaction
  - NOT focused on a particular task like airline reservation etc.
  - Systems often have an entertainment value, such as *Microsoft’s XiaoIce*
- **Task-oriented Dialog Agents**
  - Designed for a particular task and set up to have short conversations to get information from the user to help complete the task
  - E.g. Digital assistants like Siri, Cortana, Alexa, Google Now/Home, etc.
  - Agents can give travel directions, control home appliances, find restaurants, or help make phone calls or send texts

Modules in a Task-Oriented Conversational Agent



Natural Language Generation

- One of the key components to a dialogue system
- Goal of NLG is to generate natural language sentences given the semantics
  - Often performed in two steps, *viz.* *Content Planning* and *Sentence Realization*
- Content Planning: by the *Dialogue manager* ("*what to say*"?)
- Sentence Realization: *how to say it?*

For e.g.,

User: **Book a flight from Kolkata to Delhi.**

System: **Can you please specify the date of travel?**

Empathetic Conversational AI

I finally got promoted today at work!!

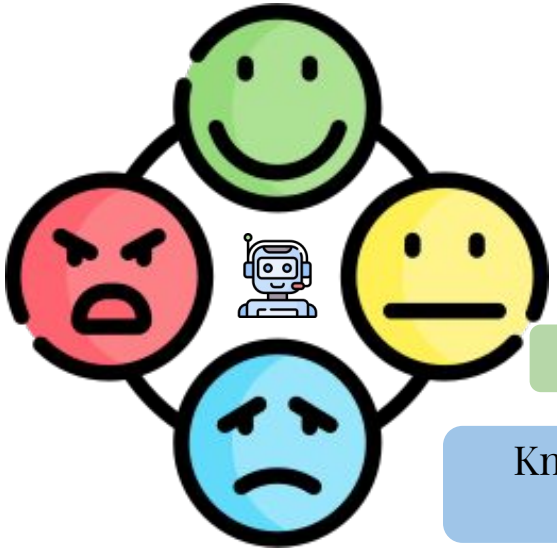
Why would anyone promote you?

Congrats!! That's great!!

Empathy

Multi-Emotion Controlled Dialogue Generation ***(AAAI 2021)***

It's amazing, I am thrilled you got promoted [**Surprise, Joy**] [**0.3**, 0.9]
Stop sulking, I am sure you will manage it [**Anger, Acceptance**] [**0.7, 0.4**]
I am sorry this could be an infection or cancer [**Sadness, Fear**] [**0.6**, 0.3]



Sentiment awareness

Knowledge grounded Conversation

Intent

Context

Mood

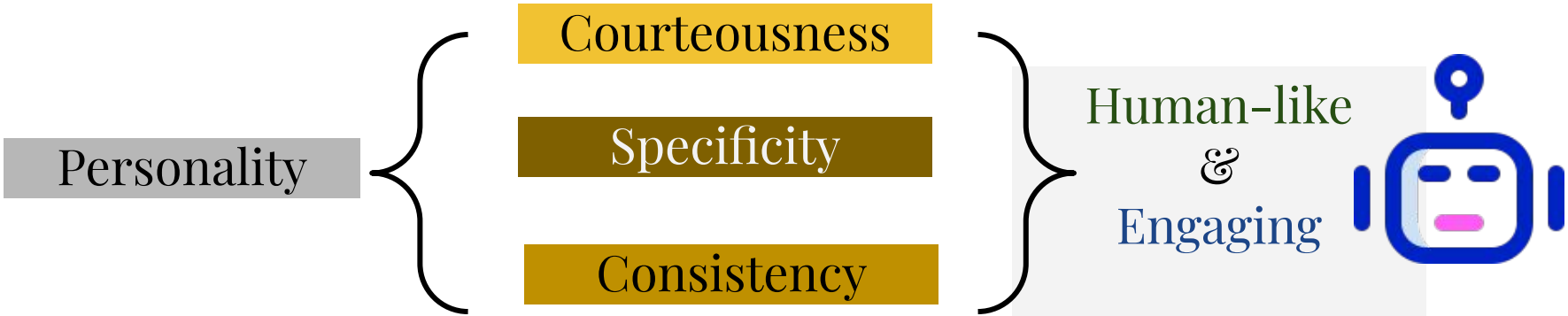
Aware & Relatable

Politeness and Personalization (HLT/NAACL 2019, INLG 2021)

We are sorry to see you are having trouble. Help has arrived. How can I help? **Apology**

Hey Christiano!! Happy to help. Can you send us a screenshot? **Greet, Assurance**

Thank you for your kind words!! Enjoy your show while flying!! **Appreciation**



Hitesh Golchha, Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya (2019). *Courteously Yours: Inducing courteous behavior in Customer Care responses using Reinforced Pointer Generator Network*. In NAACL-HLT 2019, PP. 851-860

What we solve?

To **transform** a **generic chatbot response** into a response which uses **courteous phrases** and make the users **more engaged in conversation**

Domain: Customer Care on Twitter

For any goal-oriented or open-domain systems, ***courteous response*** plays an important role in keeping the ***users engaged with the system***

Purpose is to increase user satisfaction and to build customer relations

Courtesy-The behaviour: Derived from the Politeness Theory [Brown and Levinson (1987)]

- The showing of ***politeness*** in one's attitude and behaviour towards others
- A courtesy is a polite remark or respectful act- ***very common communicative behavior***

  For example,

  Complain about a bad meal, and you are asked to leave ***BUT***,

  the common courtesy is usually an ***apology from the manager*** and, if you're **lucky,** a free **dinner**

Politeness: **Dual Nature**

- **Politeness**
  - Positive politeness
  - Negative politeness
- **Positive** and **negative** faces exist universally in human culture
- ***Positive politeness*** is expressed by satisfying *positive face* in two ways:
  - by indicating similarities amongst interactants; or
  - by expressing an appreciation of the interlocutor's self-image
- ***Negative politeness*** can also be expressed in two ways
  - by saving the interlocutor's ***face*** (either '*negative*' or '*positive*') by mitigating face threatening acts, such as ***advice-giving*** and ***disapproval***; or
  - by satisfying ***negative face*** by indicating respect for the addressee's right not to be imposed on

Politeness Strategies

- **Less Polite Strategies**
  - seeking agreement
  - joking
  - expressing optimism

- **More Polite Strategies**
  - being apologetic
  - minimizing the imposition
  - appreciating
  - thanking

Let’s see:

Some use cases of courteous behaviours in Customer Care Systems

Use-cases of Courteousness

Generic	Courteous	Behaviour
How can we help?	*Help has arrived! We are sorry to see that you are having trouble, how can we help?*	*Apology*
Can you send us a screenshot of what you're seeing?	*Hey Craig, help's here! Can you send us a screenshot of what you're seeing?*	*Greet*
Let's discuss it in GM.	*We want to help. Let's discuss it in GM.*	*Assurance*
What is happening with your internet?	*Oh no that's not good. I can help! What is happening with your internet?*	*Empathy*
Enjoy your show while flying!	*Thanks for your kind words and enjoy your show while flying!*	*Appreciation*

Resource Creation: Data Source and Attributes

- **Twitter dataset from Kaggle**
  - Interactions between customers and professional customer care agents of companies
  - Tweets have company names, anonymized user ids, time stamps, and response tweet ids

- **Pre-processing**
  - Segment the tweet into sentences
  - Remove purely ***courteous (and non-informative)*** sentences
  - Retain purely informative sentences
  - Transform the informative sentences with courteous expressions (***to remove only the courteous part from the sentence***)

Challenges: Annotation

- Identifying *different variations and styles of courteous behaviours across different companies, service providers, demographics and cultures*

- Identifying courteous behaviours in the customer care domain is not straightforward

- Hard to model the **emotion across the conversation** for effective courteous response generation
  - needs to capture the **correct emotion;** and
  - accordingly handle the customers by replying courteously

  For e.g.,

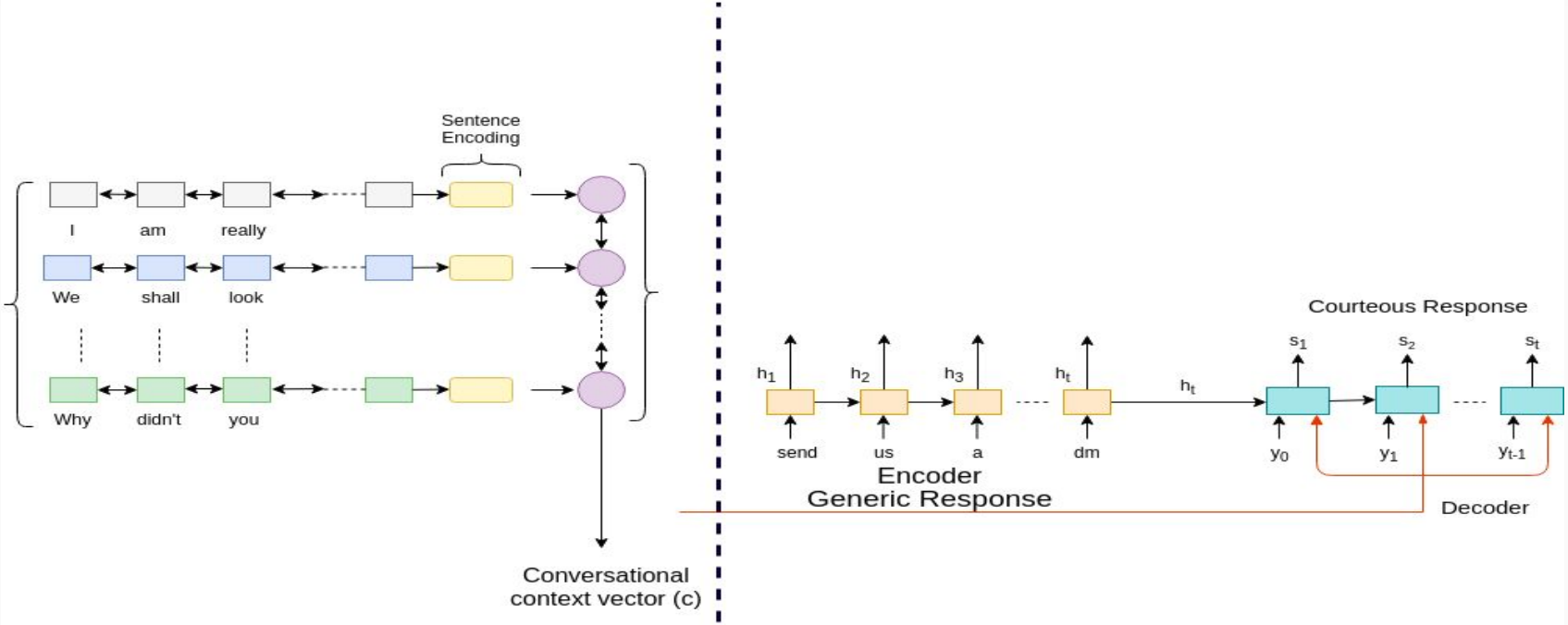
if the customer is angry*, the system needs to *pacify* and *apologize

If they are *happy*, then *appreciate*

Dataset: Statistics

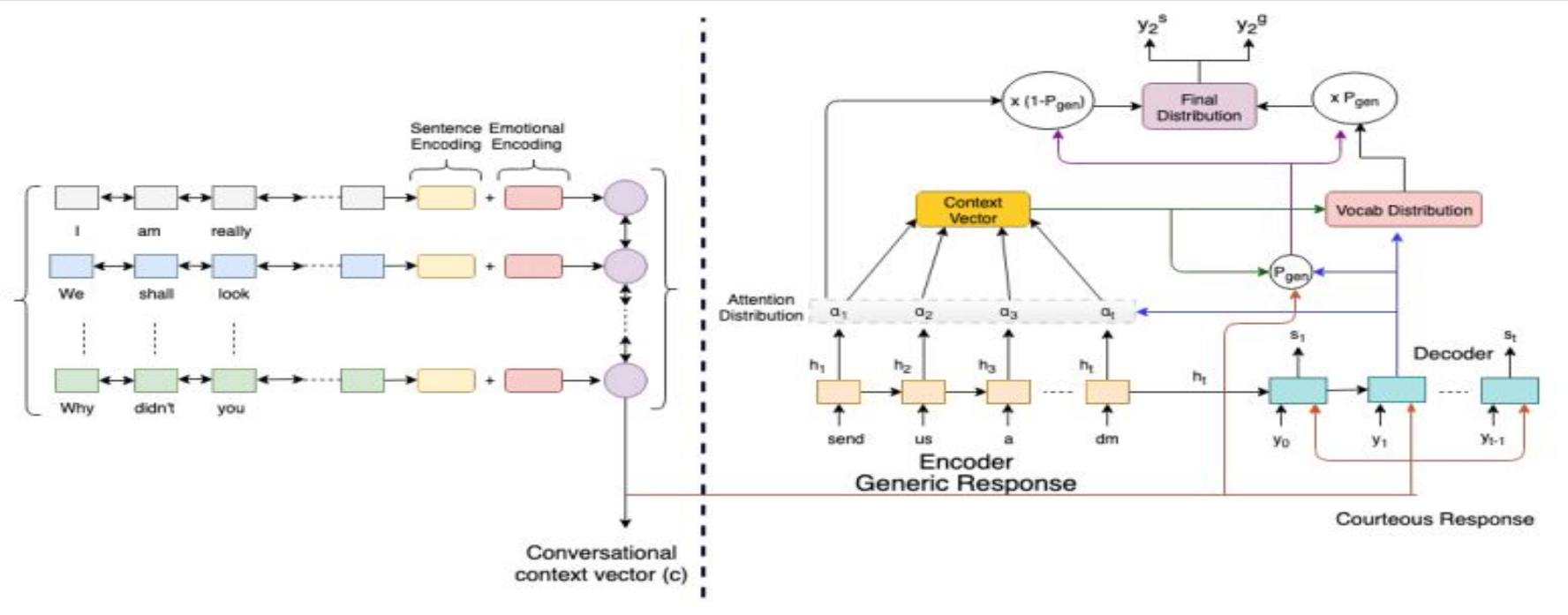
	Train	Validation	Test
# Conversations	140203	20032	40065
#Utterances	179034	25642	51238

Baseline Model



Input to the Model: Generic Response and Conversational History, Output: Courteous Response

Proposed Methodology



Inputs to the model: Conversation history (left), Generic response (centre), **Output**: Courteous response (right). The Conversation history is encoded by hierarchical Bi-LSTM to a Conversational Context vector *c*. The encoder encodes the Generic Response into hidden states h_i . Response tokens are decoded one at a time. Attention a_i, and vocabulary distributions (p_{vocab}) are computed, and combined using p_{gen} to produce output distribution. Sampling it yields y^s_i and taking its argmax yields y^g_i

Evaluation Metrics

- Content Preservation: how much of the informative content of the generic response transferred to the polite response?
- Emotional Accuracy: measures the cosine similarity of Emoji distribution between the generated and ground truth response
- Fluency: *ensures the grammatical correctness*
- Content Adequacy: ensures there is no loss of information
- Courtesy Appropriateness: Courtesy part added to the generic responses is in accordance to the conversation history
- Scoring scheme for fluency and content adequacy:
  0-incorrect or incomplete; 1: moderately correct; 2: correct
- Scoring scheme for courtesy appropriateness
  -1: in-appropriate; 0-non-courteous; 1: appropriate

Evaluation Results- Automatic

Model	BLEU	ROUGE			PPL	CP	EA
		1	2	L			
Seq2Seq	56.80	63.8	59.06	64.52	58.21	68.34	82.43
Seq2Seq + P	66.11	69.92	64.85	66.40	**42.91**	**77.67**	81.98
Seq2Seq + P + EE	68.16	72.18	67.92	71.17	43.52	76.05	85.75
Proposed Model	**69.22**	**73.56**	**69.92**	**72.37**	43.77	77.56	**86.87**

P: Pointer Generator Model; EE: Emotional embedding; PPL: Perplexity; CP: Content Preservation; EA: Emotion Accuracy

- **Observations**:
  - Model-2 is aided by copying mechanism, and hence performance is improved as it can copy portions from generic response and forward to courtesy
  - Model-3 improves the performance by 3.77 (EA) over Model-2 as it can better understand the emotional states and generate more courteous responses
  - Perplexities in Model-3 and Model-4 are more compared to Model-2: may be due to emotional embeddings that confuse generated response from the ground truth

Evaluation Results- Human

Model	F			CA			CoA		
	0	1	2	0	1	2	-1	0	1
Seq2Seq	15.70	42.50	41.80	16.21	41.69	42.10	23.71	51.08	25.21
Seq2Seq + P	14.23	42.77	43.00	15.62	39.65	44.73	22.05	39.43	38.52
Seq2Seq + P + EE	11.15	44.10	44.75	13.66	41.12	45.22	15.23	41.22	43.55
Proposed Model	10.05	44.90	44.60	13.85	38.48	47.67	14.11	41.11	44.78

F: Fluency, CA: Content Adequacy and CoA: Courtesy Appropriateness

M. Firdaus, H. Chauhan, A. Ekbal and P. Bhattacharyya (2021). *More the Merrier: Towards Multi-Emotion and Intensity Controllable Response Generation. In AAAI 2021*

What has been done in this work?

- Defining a new task: **Multiple emotion and intensity controlled dialogue generation**
- Created a large-scale Multiple Emotion and Intensity aware Multi-party Dialogue (MEIMD) dataset

- **Proposed architecture**
  - Two **novel memory-based mechanisms** to ensure the incorporation of **multiple emotions** with their corresponding **intensity** in the responses

More the Merrier: Towards Multi-Emotion and Intensity Controllable Response Generation. In AAAI 2021, **12821-12829**

Multi-Emotion Generation: *Why is relevant?*

- Utterance in a dialogue often has multiple emotions
  - **Example:** *Oh my God!!! How could you treat them in this manner! (**surprise, anger**)*

- In the absence of one of the emotions the entire meaning of the utterance is left incomplete

  Here, *"**Oh my God**"* is crucial for emphasizing the fact that the ***anger*** of the user is due to ***unawareness*** of the situation leading to ***surprise*** emotion as well

Emotion Intensity in Generation: *Why is relevant?*

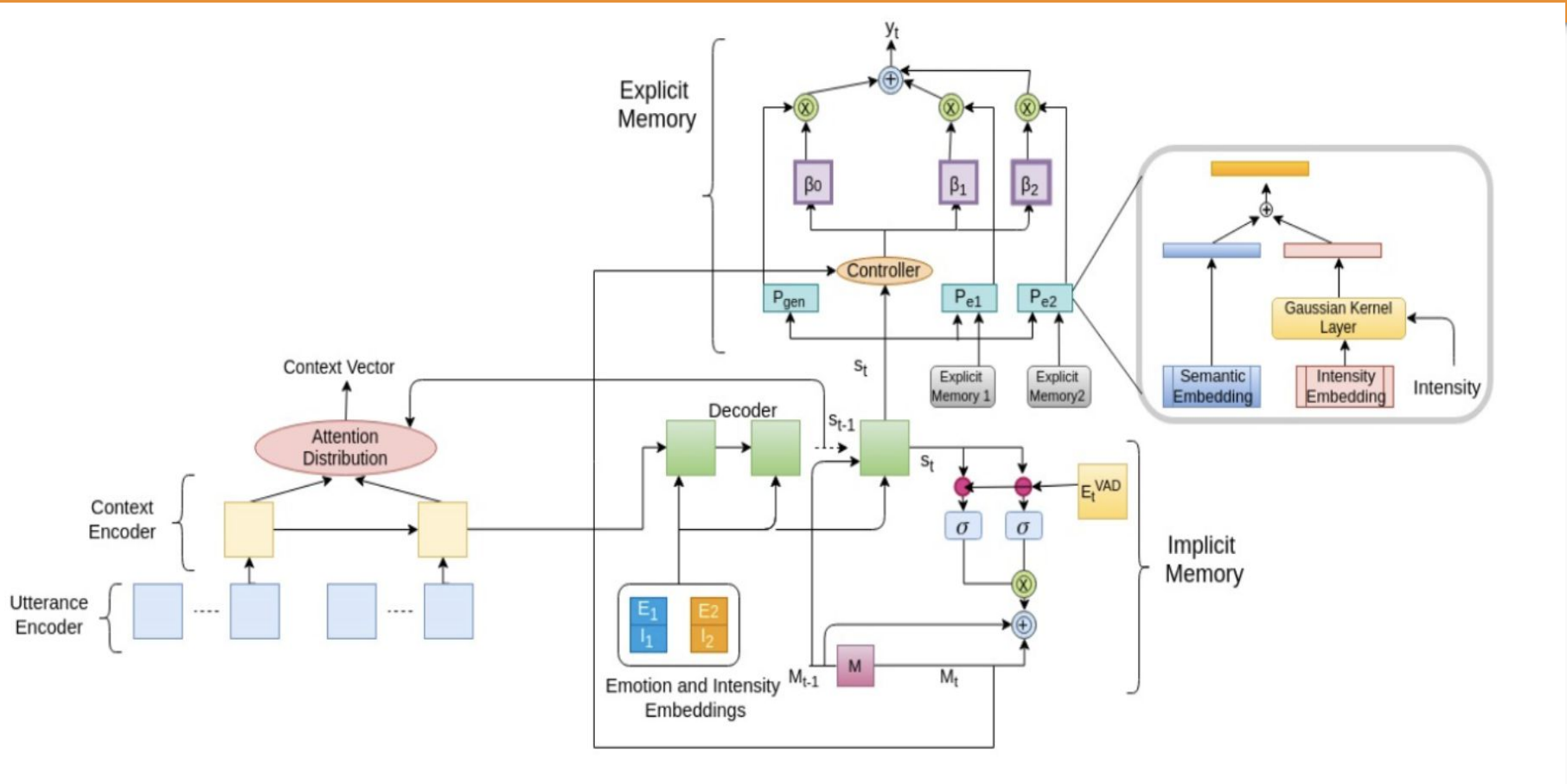
- Intensity of emotion varies, especially in case of multi-emotion generation

Example:
It's amazing, I am thrilled you got promoted
[**Surprise (0.3), Joy (0.9)**]

I am sorry this could be an infection or cancer
[***Sadness (0.6), Fear (0.3)***]

I am afraid but I know you could help me [***Acceptance (0.3), Fear (0.6)***]

Architecture



Architecture: Details

- **Utterance Encoder**: Bidirectional LSTM (BiLSTM) to encode each word of the utterance

- **Context-level Encoder**: uni-directional LSTM to model the dialog history

- **Decoder**
  - uni-directional LSTM
  - generates words sequentially conditioned on the context vector, implicit memory states of previous time step, the desired emotion embeddings, with the corresponding intensity and the previously decoded words

Dataset: MEIMD

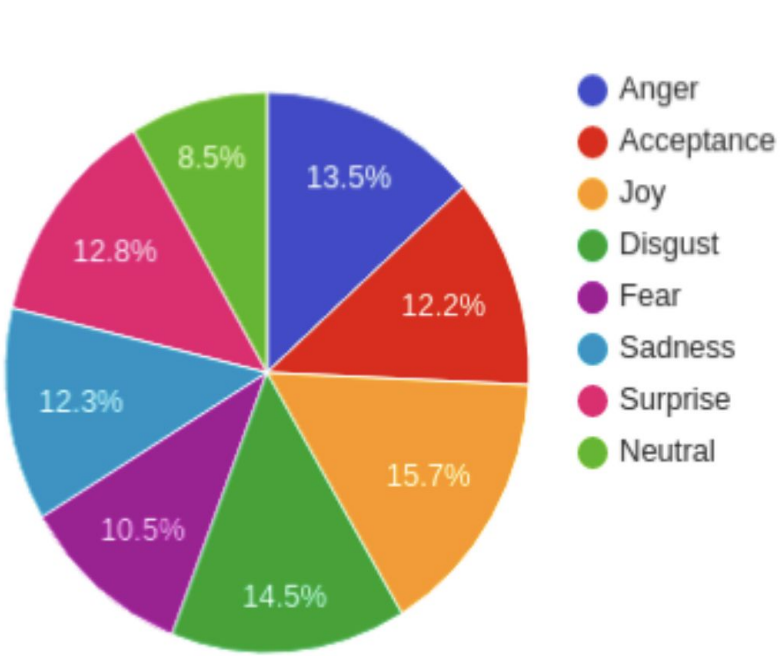
- **8 famous TV shows**: 507 episodes, spanning 456 hours
  - ➢ **Drama**: Breaking Bad, Castle, Game of Thrones, Grey's Anatomy, and House M.D.
  - ➢ **Comedy**: Friends, How I Met Your Mother and The Big Bang Theory

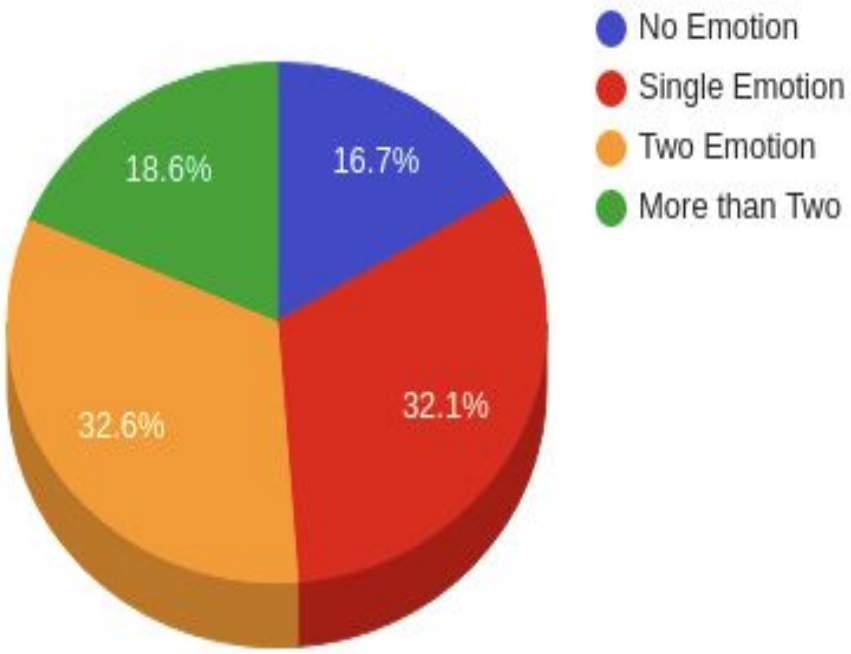
- Every utterance labelled with **emotion** (*multi-label*) and **intensity**
  - anger, acceptance, disgust, fear, joy, sadness, surprise
  - intensity: 0-3

Dataset Statistics

Show	*Genre*	*# Episodes*	*# Dialogues*	*# Utterances*	*Avg. Turns per Dialogue*	*Avg. Utterance Length*	*# Emotions per Dialogue*
Breaking Bad	Drama	62	1659	32653	20.16	14.2	3.5
Castle	Drama	105	5172	102394	21.11	13.8	4.2
Friends	Comedy	236	4228	82353	23.40	10.6	5.5
Game of Thrones	Drama	67	2263	47471	22.50	13.7	3.8
Grey's Anatomy	Drama	126	4428	86104	22.17	14.5	4.1
House M.D.	Drama	177	6476	126780	21.43	13.6	3.3
How I Met Your Mother	Comedy	208	4968	96314	22.33	12.8	5.5
The Big Bang Theory	Comedy	207	5410	86913	21.98	12.5	5.3
Total	**-**	**1188**	**34604**	**660982**	**21.88**	**13.21**	**4.4**

Emotion distribution





Evaluation Metrics

- Automatic Evaluation
  - Perplexity
  - Macro-average weighted F1 score for Emotion
  - Pearson correlation coefficient for Intensity
  - Embedding scores-based metrics (*average*, *greedy*, *extreme*)
- Manual Evaluation
  - **Fluency**: To check grammatical correctness of the response
  - **Relevance**: To ensure that the generated response is coherent with the dialog history
  - **Emotion**: To judge whether the emotional category of the generated response is consistent with the specified emotions and the dialogue history
  - **Intensity**: To check whether the degree of a particular emotion expressed in the generated response is in accordance to the intensity specified for the given emotion

Results: Automatic Evaluation

Models	***PPL***	***Embedding***			***Emotion Content***	
		Average	***Greedy***	***Extreme***	***E-F1***	***IP-Corr***
No Emotion HRED	80.7	0.491	0.360	0.371	0.39	0.26
Single Emotion HRED + Emb	75.2	0.493	0.361	0.373	0.61	-
ECM (Zhou et. al. 2018)	74.6	0.519	0.375	0.381	0.63	-
EMOTICONS (Colombo et. al. 2019)	74.3	0.523	0.381	0.385	0.63	-
EmoDS (Song et. al. 2019)	74.1	0.526	0.389	0.387	0.65	-
MEI-DG (Ours)	73.9	0.533	0.409	0.399	***0.67***	-
Single Emotion + Intensity HRED + Emb	75.2	0.493	0.361	0.373	0.63	0.44
Affect-LM (Ghosh et. al. 2017)	73.1	0.526	0.389	0.387	0.66	0.50
MEI-DG (Ours)	72.7	0.544	0.419	0.411	***0.69***	***0.57***

Models	***PPL***	***Embedding***			***Emotion Content***	
		Average	***Greedy***	***Extreme***	***E-F1***	***IP-Corr***
Multiple Emotion + Intensity HRED + Emb	73.2	0.498	0.369	0.376	0.57	0.41
HRED + IM	72.9	0.512	0.396	0.413	0.59	0.48
HRED + EM - GK	74.1	0.531	0.412	0.407	0.60	0.43
HRED + EM	73.6	0.539	0.428	0.415	0.62	0.51
MEI-DG (HRED+EM+IM)	***71.2***	***0.552***	***0.443***	***0.428***	0.66	0.54

HRED+EM-GK: model having explicit memory without the Gaussian Kernel(GK); HRED+EM: model having explicit memory with Gaussian Kernel

Observations:

- ➔ Our proposed MEI-DG framework have a lower perplexity of 71.2 than all the other baselines
- ➔ For all the metrics, our proposed framework outperforms the existing approaches significantly

Results: Human Evaluation

Models	***Fluency***	***Relevance***	***Emotion***	***Intensity***
No Emotion HRED	31.7	2.89	15.9%	13.6%
Single Emotion HRED + Emb	3.25	2.93	28.3%	-
ECM (Zhou et. al. 2018)	3.45	3.08	36.7%	-
EMOTICONS (Colombo et. al. 2019)	3.48	3.05	37.5%	-
EmoDS (Song et. al. 2019)	3.47	3.12	39.2%	-
MEI-DG (Ours)	3.49	3.13	45.1%	-
Single Emotion + Intensity HRED + Emb	3.52	3.21	32.5%	30.6%
Affect-LM (Ghosh et. al. 2017)	3.58	3.26	41.7%	34.2%
MEI-DG (Ours)	3.69	3.38	***47.3%***	***40.1%***

Models	*Fluency*	*Relevance*	*Emotion*	*Intensity*
Multiple Emotion + Intensity HRED + Emb	3.66	3.33	35.4%	32.9%
HRED + IM	3.71	3.40	40.7%	35.4%
HRED + EM - GK	3.69	3.38	38.6%	33.1%
HRED + EM	3.73	3.41	42.9%	36.8%
MEI-DG (HRED+EM+IM)	***3.82***	***3.48***	43.9%	37.2%

HRED+EM-GK: model having explicit memory without the Gaussian Kernel(GK); HRED+EM: model having explicit memory with Gaussian Kernel

Observations:

- Obtains the highest emotion and intensity score of 43.9% and 37.2%, respectively
- Single emotion and intensity-based MEI-DG model have the highest scores in terms of emotion and intensity
- Although the emotion scores are less in multiple emotion models, the fluency and relevance scores are better

Generated Examples

Conversational History	***Model***	***Emotion***	***Intensity***	***Generated Response***
It was an amazing night. You think it looked amazing?	HRED	-	-	Yes, I think so.
	MEI-DG	Joy	0.6	Yes it was really fascinating, I enjoyed it.
	MEI-DG	Joy	0.3	Yes it was a nice night.
	MEI-DG	Sadness	0.3	I found the night gloomy.
	MEI-DG	Sadness	0.9	The whole night I cried all alone
	MEI-DG	Joy, Sadness	0.3, 0.6	It was nice but I terribly missed my friend.
	MEI-DG	Joy, Sadness	0.9, 0.3	The night was really great only had a little trouble in finding cab.

Results: Automatic Evaluation

Models	PPL	Embedding			Emotion Content	
		Average	Greedy	Extreme	E-F1	IP-Corr
No Emotion HRED	80.7	0.491	0.360	0.371	0.39	0.26
Single Emotion HRED + Emb	75.2	0.493	0.361	0.373	0.61	-
ECM (Zhou et. al. 2018)	74.6	0.519	0.375	0.381	0.63	-
EMOTICONS (Colombo et. al. 2019)	74.3	0.523	0.381	0.385	0.63	-
EmoDS (Song et. al. 2019)	74.1	0.526	0.389	0.387	0.65	-
MEI-DG (Ours)	73.9	0.533	0.409	0.399	***0.67***	-
Single Emotion + Intensity HRED + Emb	75.2	0.493	0.361	0.373	0.63	0.44
Affect-LM (Ghosh et. al. 2017)	73.1	0.526	0.389	0.387	0.66	0.50
MEI-DG (Ours)	72.7	0.544	0.419	0.411	***0.69***	***0.57***

Models	***PPL***	***Embedding***			***Emotion Content***	
		Average	***Greedy***	***Extreme***	***E-F1***	***IP-Corr***
Multiple Emotion + Intensity HRED + Emb	73.2	0.498	0.369	0.376	0.57	0.41
HRED + IM	72.9	0.512	0.396	0.413	0.59	0.48
HRED + EM - GK	74.1	0.531	0.412	0.407	0.60	0.43
HRED + EM	73.6	0.539	0.428	0.415	0.62	0.51
MEI-DG (HRED+EM+IM)	***71.2***	***0.552***	***0.443***	***0.428***	0.66	0.54

HRED+EM-GK: model having explicit memory without the Gaussian Kernel(GK); HRED+EM: model having explicit memory with Gaussian Kernel

Observations:

➔ Proposed MEI-DG framework has a lower perplexity of 71.2 compared to the others

➔ Proposed framework outperforms the existing approaches significantly

Results: Human Evaluation

Models	***Fluency***	***Relevance***	***Emotion***	***Intensity***
No Emotion HRED	31.7	2.89	15.9%	13.6%
Single Emotion HRED + Emb	3.25	2.93	28.3%	-
ECM (Zhou et. al. 2018)	3.45	3.08	36.7%	-
EMOTICONS (Colombo et. al. 2019)	3.48	3.05	37.5%	-
EmoDS (Song et. al. 2019)	3.47	3.12	39.2%	-
MEI-DG (Ours)	3.49	3.13	45.1%	-
Single Emotion + Intensity HRED + Emb	3.52	3.21	32.5%	30.6%
Affect-LM (Ghosh et. al. 2017)	3.58	3.26	41.7%	34.2%
MEI-DG (Ours)	3.69	3.38	***47.3%***	***40.1%***

Models	***Fluency***	***Relevance***	***Emotion***	***Intensity***
Multiple Emotion + Intensity HRED + Emb	3.66	3.33	35.4%	32.9%
HRED + IM	3.71	3.40	40.7%	35.4%
HRED + EM - GK	3.69	3.38	38.6%	33.1%
HRED + EM	3.73	3.41	42.9%	36.8%
MEI-DG (HRED+EM+IM)	***3.82***	***3.48***	43.9%	37.2%

HRED+EM-GK: model having explicit memory without the Gaussian Kernel(GK); HRED+EM: model having explicit memory with Gaussian Kernel

Observations:

- Obtains the highest emotion and intensity score of 43.9% and 37.2%, respectively
- Single emotion and intensity-based MEI-DG model has the highest score in terms of emotion and intensity
- Although the emotion scores are less in multiple emotion models the fluency and relevance scores are better

Generated Examples

Conversational History	***Model***	***Emotion***	***Intensity***	***Generated Response***
It was an amazing night. You think it looked amazing?	HRED	-	-	Yes, I think so.
	MEI-DG	Joy	0.6	Yes it was really fascinating, I enjoyed it.
	MEI-DG	Joy	0.3	Yes it was a nice night.
	MEI-DG	Sadness	0.3	I found the night gloomy.
	MEI-DG	Sadness	0.9	The whole night I cried all alone
	MEI-DG	Joy, Sadness	0.3, 0.6	It was nice but I terribly missed my friend.
	MEI-DG	Joy, Sadness	0.9, 0.3	The night was really great only had a little trouble in finding cab.

M Firdaus, N Thangavelu, A Ekbal, P Bhattacharyya. *I enjoy writing and playing, do you: A Personalized and Emotion Grounded Dialogue Agent using Generative Adversarial Network*

IEEE Transaction on Affective Computing, 2022

Persona aware Response Generation with Emotions

- **Persona aware Emotional response generation**
  - System is able to generate emotional, specific and consistent responses

- Every individual has a **personality** and is driven by **emotions**

- **What it does?**
  - Makes the responses interactive and interesting
  - Conversation with a consistent personality helps in bringing **consistency** and **specificity** in responses
  - Facilitates building user's ***trust*** and ***confidence***
  - Infuses the **emotions in the responses** to make these more human-like (as per the ***personality information***)

An example

Persona 1	Persona 2
As a child, I won a national spelling bee. I've been published in the New Yorker magazine. I am a gourmet cook. I've perfect pitch.	I'm very athletic. I have brown hair. I love bicycling. I hate carrots.
[Person 1] Hi! I work as a gourmet cook. [Person 2] I don't like carrots. I throw them away. [Person 1] Really. But, I can sing pitch perfect. [Person 2] I also cook, and I ride my bike to work.	

- **Persona 1 and Persona 2** represent the personalized information of Person 1 and Person 2, respectively. The last row represents the ***Dialogue between Person 1 and Person 2***
- Agent maintains **unique personality**, but the conversation is more like **stating facts,** and lacks in **emotional connection**
- Emotion would make it more **engaging** and **human-like**

From the Example

The response of **Person 2** to **Person 1** could be more empathetic like

That's a great job, but I don't like carrots and throw them away

Instead of

I don't like carrots. I throw them away

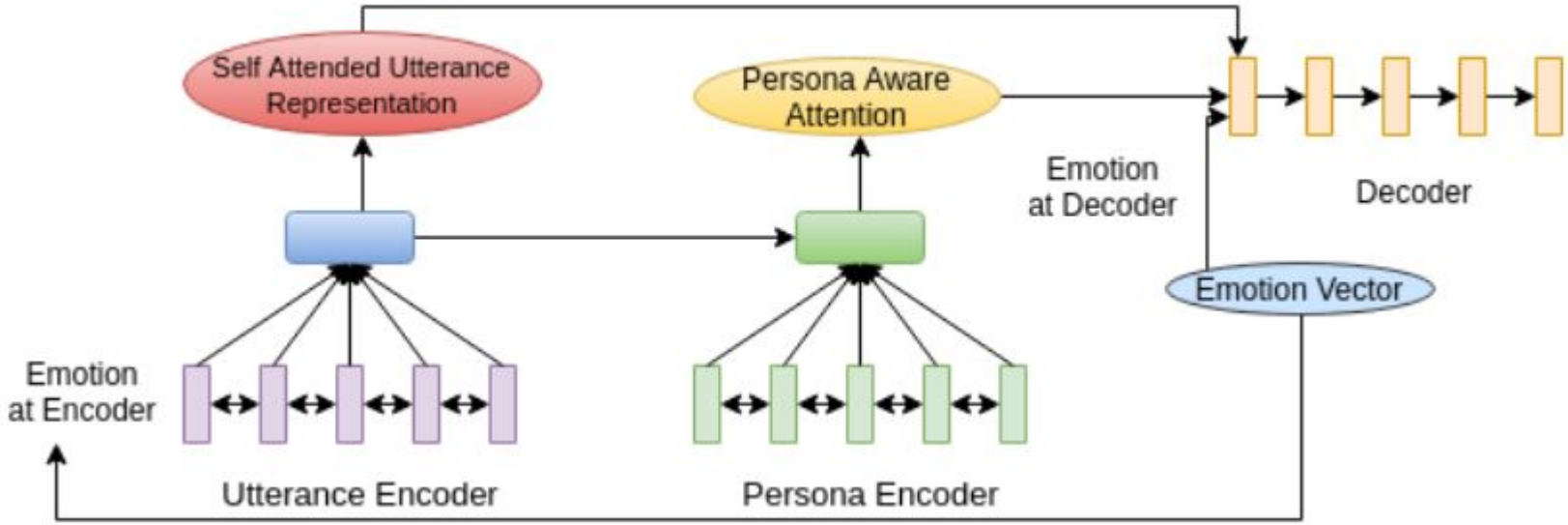
Has a happy undertone than the ground-truth response which is neutral and contains only facts about Person 2

Methodologies

- **First Approach**
  - **Encoder-decoder framework**
  - **Emotion information included directly to the decoder**

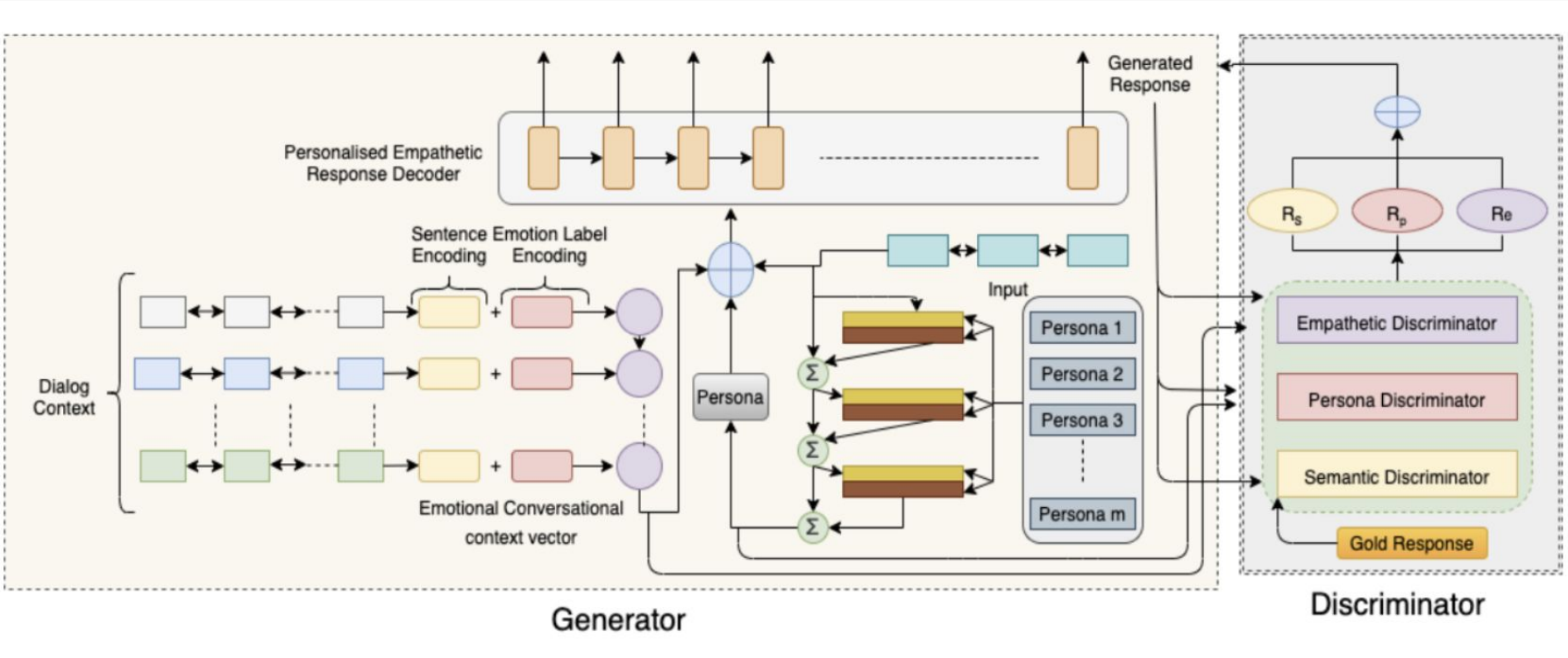
- **Second Approach**
  - **Generative adversarial network** (***GAN***)
  - **Interactive discriminators** to ensure that the responses are personalized and empathetic

The First Architecture



- **Utterance encoder and persona encoders**: Bi-GRU to encode the utterance and personas, respectively
- **Emotion vector:** appended either at the encoder side or given directly to the decoder
- **Persona-aware attention**
- **Self-attended utterance**

The Second Architecture



The generator comprises of hierarchical dialog encoder, persona memory network and decoder. While the discriminator comprises of three interactive components

Methodology

- **The generator** is a basic **encoder-decoder framework**
  - Utterance Encoder:

$$h_U = [\, BiGRU_u(w_i , h_{U,i-1}) \,.\, e_i]$$

  - Context Encoder:

$$h_{con} = GRU_c\,(h_{U,n} , h_{con,i-1})$$

  - Persona Memory Network, with two functions
    - To *encode the persona information* into a dense representation and
    - To decide the *appropriate persona to be expressed in the response*
  - Decoder

Interactive Discriminators

- Semantic Discriminator: Calculates the semantic distance between the generated response and the ground-truth response
- Persona Discriminator: Measures whether the conversational agent is capable of conversing according to the specified persona of the user
- Empathetic Discriminator: Determines whether the dialogue agent has the ability to converse in an empathetic manner

Dataset Statistics & Emotion Classification

- **ConvAI2 benchmark dataset** (***Logacheva V et al, 2020: Non-goal oriented Human-bot Dialogues***)
  - Extended version (with a new test set) of the Persona-chat dataset

Dataset Statistics	Train	Valid	Test
# Dialogues	7686	1640	1655
# Utterances	124816	19680	19860
Avg. turns per Dialogue	12.51	12.73	12.74
Avg. words in a Response	11.89	9.57	10.75
# Emotions per Dialogue	7.4	6.5	5.1
# Unique words	20322	13415	15781

Emotion Classification

- Followed a semi-supervised approach for annotating the ConvAI2 dataset with emotions
  - Use Empathetic Dialogues (EmpD) dataset of 25k conversations grounded with emotional situations
  - 32 fine-grained emotions, covering a wide range of positive and negative emotions, such as *surprised, excited, angry, joy, furious, grateful, disgusted, etc*
  - Built several classifiers
  - Best classifier used to classify ConvAI2 dataset with emotions

Model	E-F1
LSTM	37.06
CNN	34.90
Bi-LSTM	39.87
BERT	61.74
RoBERTa	59.89

Evaluation Metrics

- ★ **Automatic evaluation metrics**
  - Perplexity
  - BLEU
  - Rouge-L
  - Emotion Accuracy
- ★ **Human evaluation metrics**
  - Fluency*:* Measures the grammatical correctness of the generated response
  - Persona Consistency*:* Takes care of the fact that the response generated is in accordance with the persona information of the speaker
  - Emotion*:* Judges whether the generated response is in accordance with the desired emotions

Fluency: 0- incomplete response or else incorrect response, 1- moderately correct response, and 2- correct response

Emotion and persona consistency: 0: for the absence of emotion in the reply and the reply is inconsistent to the specified persona; and 1: for the presence of emotion in the response along with the consistency of the response with the persona information

Evaluation Results: Automatic (First Approach)

Model Description		Perplexity	BLEU-4	Rouge-L	Emotion Accuracy
Baseline Approaches	*Seq2Seq*	59.11	0.042	0.149	0.35
	Seq2Seq+ Attn	58.23	0.047	0.151	0.38
	Seq2Seq + Attn + PAA	57.60	0.088	0.154	0.42
	Seq2Seq + Attn + EE	56.87	0.092	0.157	0.58
	Seq2Seq + Attn + ED	56.39	0.096	0.158	0.61
Proposed Approaches	*Seq2Seq + Attn + PAA + EE*	55.59	0.099	0.162	0.65
	Seq2Seq + Attn + PAA + ED	***52.68***	***0.108***	***0.169***	***0.67***

- ***Seq2Seq + Attn + PAA + ED*** **outperforms all the baseline approaches**
- **Emotion information provided directly to the decoder is more helpful in comparison to the encoder**
- **Directly putting the information into the encoder causes loss of information, but when this information is directly subjected to the decoder, the emotion manifests better in the responses**

Evaluation Results: Human Evaluation (First Approach)

Model Description		Fluency			Emotion		Persona Consistency	
		0	1	2	0	1	0	1
Baseline Approaches	*Seq2Seq*	27.36	45.83	26.81	75.93	24.07	77.20	22.80
	Seq2Seq+ Attn	26.11	44.71	29.18	74.56	25.44	76.14	23.86
	Seq2Seq + Attn + PAA	23.41	42.96	33.63	73.81	26.19	51.64	48.36
	Seq2Seq + Attn + EE	24.17	43.11	32.72	59.33	40.67	70.88	29.12
	Seq2Seq + Attn + ED	23.05	42.88	34.07	57.49	42.51	70.31	29.69
Proposed Approaches	*Seq2Seq + Attn + PAA + EE*	19.64	38.65	41.71	55.72	44.28	49.85	50.15
	Seq2Seq + Attn + PAA + ED	***18.15***	***37.32***	***44.53***	***53.91***	***46.09***	***48.11***	***51.89***

Seq2Seq + Attn + PAA + ED **outperforms all the other approaches in case of all the metrics**

Evaluation Results: Automatic Metrics (Second Approach)

Model Description		Perplexity	BLEU-4	Rouge-L	Emotion Accuracy
Baseline Approaches	*Seq2Seq*	56.11	0.089	0.196	35.8
	HRED	55.63	0.096	0.201	37.6
	SeqGAN	55.61	0.098	0.203	38.1
	Seq2Seq + E + P	54.13	0.103	0.219	65.7
	HRED + E + P	54.85	0.116	0.224	66.5
Proposed Approach	***EP-GAN***	***51.92***	***0.143***	***0.266***	***71.5***
Ablation Study	*EP-GAN - SD*	53.47	0.118	0.239	67.8
	EP-GAN - ED	53.44	0.125	0.242	69.5
	EP-GAN - PD	52.39	0.129	0.249	68.3
	EP-GAN - SD + ED	52.26	0.130	0.251	70.8
	EP-GAN - SD + PD	52.12	0.135	0.257	68.8

Here, ED: Empathetic Discriminator, SD: Semantic Discriminator, PD: Persona Discriminator, E: Emotion, P: Persona. From the Table, it is clear that **EP-GAN** outperforms all the given baselines in case of all the metrics. Also, from the ablation study it is evident that *ED assists in increasing the emotional quotient* in the response while *PD incorporates the persona information* in them

Evaluation Results: Human (Second Approach)

Model Description		Fluency			Relevance			Emotion		Persona Consistency	
		0	1	2	0	1	2	0	1	0	1
Baseline Approach	*Seq2Seq*	28.71	46.25	25.04	20.71	32.63	46.66	74.18	25.82	75.63	24.37
	HRED	27.51	45.89	26.60	18.45	33.87	47.68	72.37	27.63	74.88	25.12
	SeqGAN	27.41	45.96	26.63	16.21	41.69	42.10	71.96	28.04	74.23	25.77
	Seq2Seq + E + P	22.49	42.33	35.18	15.62	39.65	44.73	56.81	43.19	51.93	48.07
	HRED + E + P	21.24	40.13	38.63	14.23	40.07	45.70	55.72	44.28	49.64	50.36
Proposed Approach	***EP-GAN***	***18.76***	***37.35***	***43.89***	***13.85***	***38.48***	***47.67***	***48.03***	***51.97***	***45.11***	***54.89***

- EP-GAN performs better than all the baselines
- Generated responses are *fluent and in accordance to the conversational history*
- The emotional content is higher simultaneously the persona information also gets incorporated in the responses

Generated O/P Examples

Dialog Input	Persona Information		Ground Truth	Generated Response
[Person 1] What do you do for a career? **[Person 2]** I like to watch kids. **[Person 1]** I actually play guitar and dolot of welding.	**Persona 1** I am primarily a meat eater. I am a guitar player. Welding is my career field. My parents don't know I am gay.	**Persona 2** I've a sweet tooth. I'm a babysitter and drive a mercedes. I'm the middle child of 3 sibling. I'm getting married in six weeks.	What do you weld? houses?	**Seq2Seq:** What is weld? <UNK> <UNK> **HRED:** What you weld? **SeqGAN:** What are you welding? **Seq2Seq+E+P:** Wow! What you weld mostly? (**surprise)** **HRED+E+P:** Really! Do you weld houses? (**surprise**) **EP-GAN:** Wow that's great, what do you weld mostly, is it houses? (**surprise**)
[Person 1] Hi! do you like turtles? **[Person 2]** I have two cats actually. **[Person 1]** I have a turtle his name is Speedy.	**Persona 1** I don't pick up my toys. I have a pet turtle. I like to play with my dolls. I'm fourteen.	**Persona 2** I love cats and have two cats. My favorite season is winter. I won a gold medal in the 1992 olympics. I've a hat collection of over 1000 hats.	I am a cat person.	**Seq2Seq:** I have cats. **HRED:** I like cats mostly. **SeqGAN:** I am into cats. **Seq2Seq+E+P:** Turtles are nice but I like cats. (**joy**) **HRED+E+P:** Nice name for a pet, but I love cats. (joy) **EP-GAN:** That is an adorable name for a turtle! but I am a cat person. (**joy)**

The EP-GAN generates responses that not only incorporates persona but also makes the response empathetic

M. Firdaus, H. Chauhan, A. Ekbal and P. Bhattacharyya (2021). **EmoSen: Generating Sentiment and Emotion Controlled Responses in a Multimodal Dialogue System (2020)**. ***IEEE Transactions on Affective Computing***

What we do here?

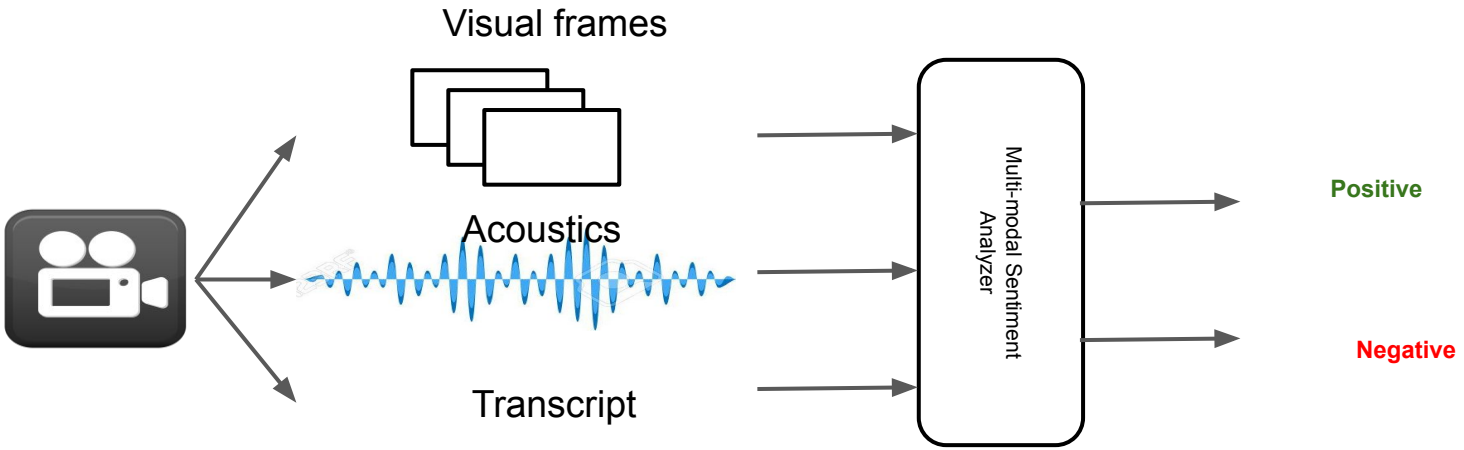
- Sentiment and Emotion are two closely related problems
  - E.g. **joy→ positive; disgust→ negative**

- Generate response with respect to underlying emotion and sentiment

- Multimodal sources (**text + video + audio**) provide better evidences for predicting sentiments and emotions

Let us see how multimodal sentiment analysis helps!

- **Aims to leverage the varieties of (often distinct) information from multiple sources for building an efficient system**



Utterance	Affect	Text	Acoustic	Visual
Thanks for putting me on hold! I have all the time in the world.	**Sentiment (Negative)**	-	Intensity, Pitch, Tone etc.	Facial expression, eyes movement etc.

Sentiment and Emotion aware Multi-modal Dialogue (SEMD) Dataset

- No existing datasets that take into account the emotion and sentiment while generating a response in a dialogue system

- **Prepare a large-scale multi-party dataset**
  - seamlessly employs multimodal information
  - incorporates sentiment and emotion in the dialogues

An Example from the SEMD dataset

Every utterance is labelled with corresponding emotion and sentiments

You know, we had all this cool stuff in basement.

(Surprise, Positive)

No no, I am paddling away.

(Disgust, Positive)

See, Yeah

(Joy, Positive)

Really, you got all this rustic crap for free.

(Anger, Negative)

Data Collection

- Source: 10 famous TV shows belonging to different genres
  - **Comedy:** Friends, The Big Bang Theory, How I Met Your Mother, The Office
  - **Drama:** House M.D., Grey's Anatomy, Castle and Game of Thrones, House of Cards, Breaking Bad
- Contains dialogues mostly from all the episodes belonging to the different seasons of the TV series, giving us a *wide variations in conversations*
  - No. of episodes: 1258, spanning 746 hours
  - Extracted all the subtitles and transcripts for every episode

Data Collection

- Segment the episodes into scenes
- Scenes further divided into short clips representing a dialogue following the heuristics
  - **Increasing order of time-stamp of the utterance in the conversation**
  - **All utterances should be in the same scene and episode**
- Extract the corresponding time-stamps of every utterance in a dialogue
- Obtain audio and visual clips from the source episodes
- Audio clips are then formatted as 16-bit PCM WAV files for further processing
- Video clips used to extract 2048D pooled features using the last block of ResNet101
- Our final *SEMD dataset comprises of textual, visual, and audio features* that bring the three crucial modalities together for effective emotion and sentiment controlled dialogue generation.

Dataset Statistics

Show	Genre	# Season	# Episodes	# Dialogues	# Utterances
Breaking Bad	Drama	5	62	1659	32653
Grey's Anatomy	Drama	15	351	14926	295496
House of Cards	Drama	6	73	2851	56416
Friends	Comedy	10	236	4228	82353
House M.D.	Drama	8	177	6476	127780
Castle	Drama	8	173	7401	146669
How I Met Your Mother	Comedy	9	208	4968	97344
The Office	Comedy	9	201	4813	94470
Game of Thrones	Drama	8	73	2263	47472
The Big Bang Theory	Comedy	12	279	5456	86024
Total	**-**	**90**	**1833**	**55041**	**1066677**

Data Annotation

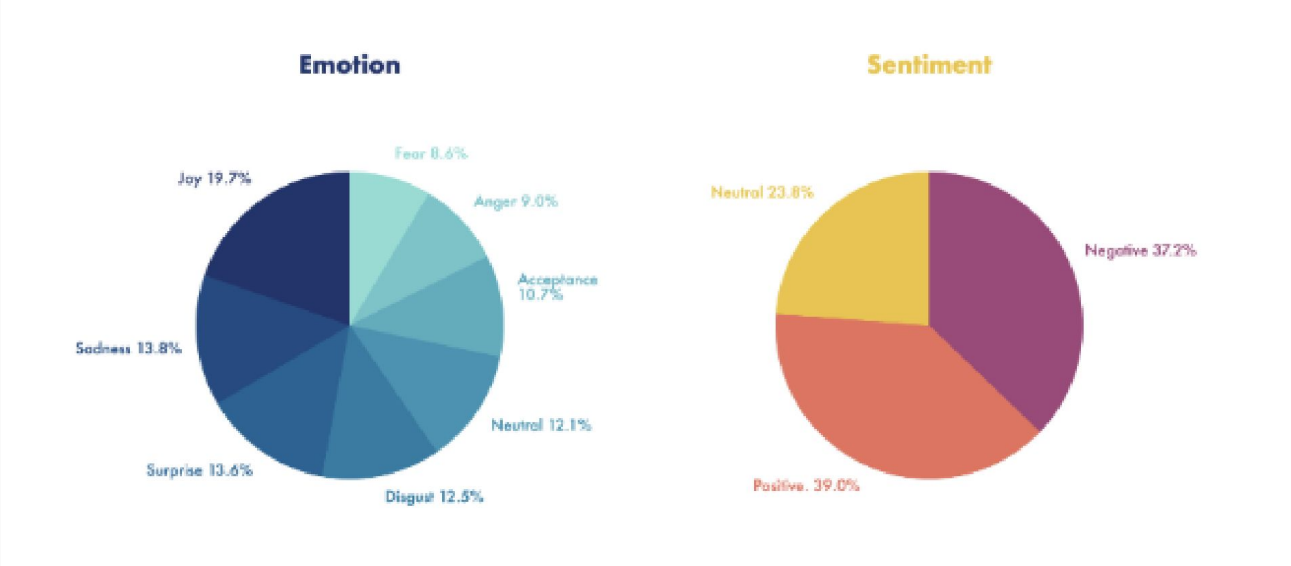
- Propose a *semi-supervised approach* for labeling SEMD with emotions and sentiments

- Manually annotate all the 10 TV series
  - Ekman's six universal emotions + 2 extra
  - Joy, Sadness, Anger, Fear, Surprise, and Disgust + Acceptance and Neutral
  - Sentiment labels: positive, negative, and neutral

Data Annotation

- From all the TV series, we take 2000 utterances, each belonging to the different seasons and episodes for the annotation with sentiment and emotion labels

- Three annotators were employed to label each utterance



Emotion and Sentiment Classifier

- Training: Validation: Test set : 7:1:2

- Feature representation
  - 128-dimension VGGish features for audio
  - 2048-dimension Resnet features for the video
  - 300-dimensional Glove embeddings for the textual representation

- Classifiers
  - CNN, Bi-LSTM for text
  - bc-LSTM, DialogRNN for audio and video

Emotion and Sentiment Classifier: Results

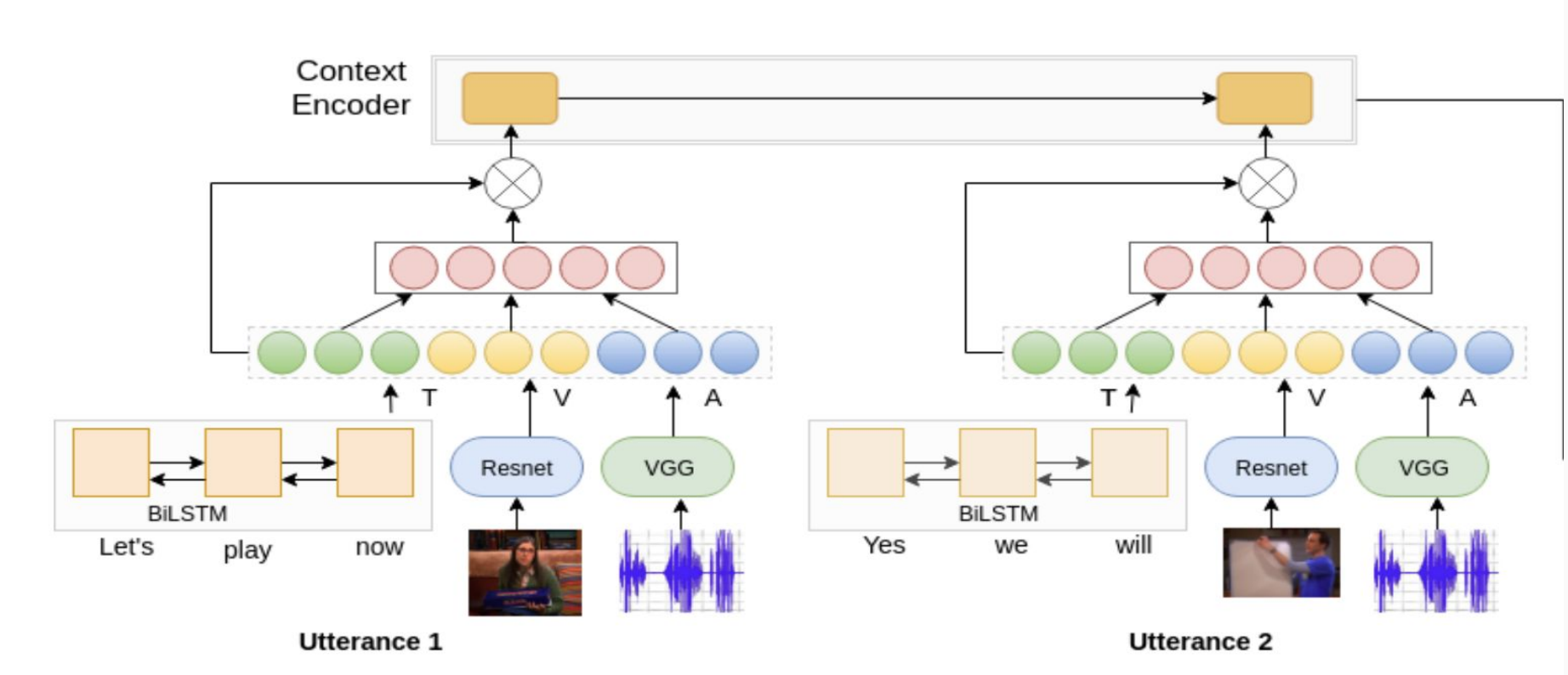
Model	E-F1	S-Acc
Bi-LSTM	54.06	63.28
CNN	54.90	63.12
bc-LSTM	58.87	65.45
DialogueRNN	60.74	66.29
DialogueRNN + BERT (ours)	62.89	68.98

- Finally, for labeling the entire dataset, we use the best-performing classifier **DialogueRNN + BERT**
- **SEMD dataset**: 55k dialogues; Average dialogue length: 20; No. of emotion labels: 8; No. of sentiment labels: 3

Final Dataset Statistics of SEMD

Statistics	Train	Valid	Test
# Modalities	[t,a,v]	[t,a,v]	[t,a,v]
# Dialogues	38527	5504	11010
# Utterances	737514	110080	219083
Avg. turns per Dialogue	21.71	21.73	21.74
Avg. words in Text utterance	6.89	6.57	6.75
# Emotions per Dialogue	5.5	4.2	5.1
# Unique Words	40322	23415	25781

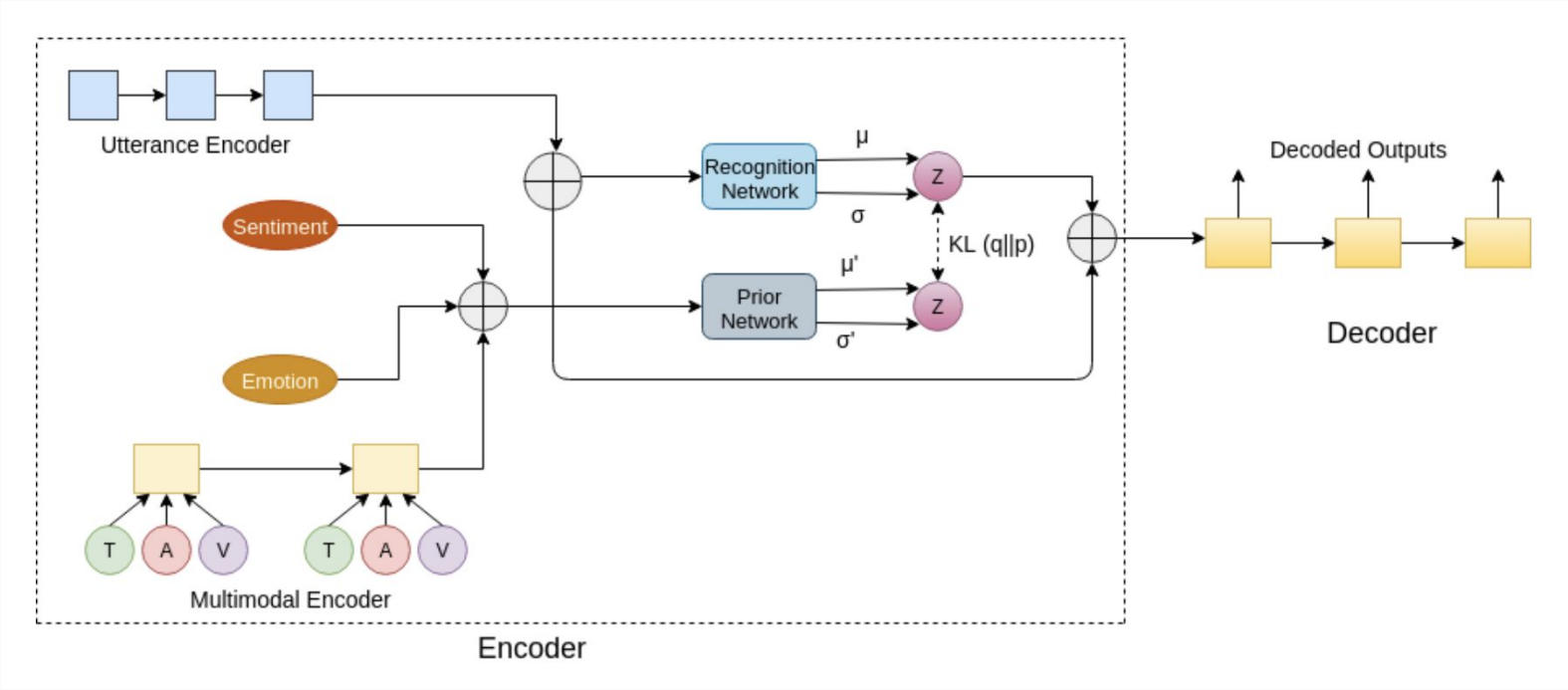
Multimodal Hierarchical Encoder with Attention



The attended utterance representation(with features from all the three modality) is passed to the context encoder.

Multimodal Conditional Variational Autoencoder (M-CVAE)

- In M-CVAE, dialog response y is generated conditioned on dialog context h_c along with the desired emotion V_e and sentiment V_s embedding and latent variable z



Evaluation Results: Automatic Evaluation Metrics

Model Description		Modality			W/o S & E			Only S		Only E		S + E		
		T	*A*	*V*	*PPL*	*SA*	*EA*	*PPL*	*SA*	*PPL*	*EA*	*PPL*	*SA*	*EA*
Amodal Baselines	*HRED*	√			73.8	0.40	0.35	72.1	0.60	71.8	0.64	70.5	0.61	0.65
	HRED		√		120.2	0.31	0.20	120.6	0.35	119.5	0.28	119.1	0.35	0.30
	HRED			√	118.4	0.31	0.22	117.9	0.34	117.5	0.29	116.8	0.33	0.30
Bimodal Baselines	*MHRED*	√	√		69.4	0.41	0.36	67.9	0.63	66.2	0.66	65.1	0.64	0.68
	MHRED	√		√	68.8	0.40	0.36	66.3	0.63	64.7	0.65	64.2	0.66	0.66
	MHRED		√	√	102.7	0.33	0.23	101.9	0.34	100.8	0.31	100.2	0.34	0.31
Trimodal Baselines	*MHRED*	√	√	√	65.8	0.43	0.37	64.1	0.65	63.2	0.68	62.1	0.68	0.71
	MHRED + Attn	√	√	√	63.1	0.44	0.38	60.8	0.66	60.1	0.68	59.4	0.68	0.72
OurProposed Approach	*M-CVAE*	√			46.4	0.42	0.40	45.8	0.72	44.7	0.69	44.1	0.78	0.74
	M-CVAE		√		95.3	0.31	0.25	93.9	0.36	93.1	0.32	92.4	0.36	0.34
	M-CVAE			√	93.5	0.33	0.27	92.7	0.34	91.8	0.33	91.4	0.36	0.35
	M-CVAE	√	√		44.9	0.42	0.41	42.9	0.74	42.0	0.72	41.2	0.77	0.76
	M-CVAE	√		√	44.2	0.43	0.41	43.7	0.73	42.6	0.72	41.8	0.78	0.75
	M-CVAE		√	√	91.2	0.35	0.27	90.5	0.35	90.1	0.35	89.4	0.36	0.35
	M-CVAE	√	√	√	43.2	0.46	0.41	41.4	0.75	40.7	0.74	37.5	0.81	0.78
	M-CVAE + Attn (EmoSen)	√	√	√	**42.7**	**0.47**	**0.43**	**39.5**	**0.77**	**38.2**	**0.75**	**35.9**	**0.83**	**0.79**
	M-CVAE + Attn + GMP (EmoSen)	√	√	√	**42.1**	**0.47**	**0.44**	**38.7**	**0.79**	**37.1**	**0.76**	**34.8**	**0.85**	**0.80**

Results of different models on SEMD dataset.

Here, T: Text, A: Audio, V: Visual features, S: Sentiment, E: Emotion, PPL: Perplexity, SA: Sentiment Accuracy,

EA: Emotion Accuracy; GMP: Gaussian Mixture Prior network

Automatic Evaluation Results

Model Description	W/o S & E		Only S		Only E		S + E	
	d-1	d-2	d-1	d-2	d-1	d-2	d-1	d-2
MHRED	0.0068	0.0157	0.0075	0.0211	0.0080	0.0281	0.0082	0.0292
MHRED + Attn	0.0072	0.0168	0.0077	0.0201	0.0077	0.0289	0.0084	0.0294
M-CVAE	0.0184	0.0423	0.0188	0.0468	0.0189	0.0434	0.0192	0.0475
M-CVAE + Attn	**0.0187**	**0.0489**	**0.0193**	**0.0505**	**0.0196**	**0.0512**	**0.0198**	**0.0520**
M-CVAE + Attn + GMP	**0.0192**	**0.0495**	**0.0198**	**0.0510**	**0.0201**	**0.0519**	**0.0203**	**0.0520**

Results on diversity of the generated responses for the baseline and proposed models. Here, d-1:distinct-1, d-2: distinct-2, S: Sentiment, E-Emotion, GMP: Gaussian Mixture Prior network

Evaluation Results: Human Evaluation Metrics

Model Description	Fluency			Sentiment		Emotion	
	0	1	2	0	1	0	1
MHRED + Attn (only S)	18.23	43.70	38.07	33.5	56.5	-	-
MHRED + Attn (only E)	16.85	40.30	42.85	-	-	42.8	57.2
MHRED + Attn (only S + E)	17.64	39.2	43.16	41.7	58.3	40.5	59.5
M-CVAE + Attn (S + E)	**14.7**	**38.5**	**46.8**	**35.4**	**64.6**	**34.6**	**65.4**

Analysis of Multimodality for Response Generation

You know, we had all this cool stuff in basement.

No no, I am paddling away.

See, Yeah

Really, you got all this rustic crap for free.

Ground Truth: Yeah, it's nice
Given S and E: Surprise, Negative
MHRED + Attn: You don't like it?
M-CVAE + Attn: I am amazed, how can you not like it

Example 1

You are unbelievable

I know.

All right Pictionary, what are the teams?

How about boys versus girls.

Ground Truth: Ohh that hardly seems fair!
Given S and E: Joy, Positive
MHRED + Attn: Wow, that's a nice idea.
M-CVAE + Attn: Great let's play, I love to draw on board.

Example 2

Generated Examples

Conversational History	Model	Emotion	Sentiment	Generated Response
Did he hear anything, at least? He was listening to his Ipod. Nothing like a little quality time with daddy. How could this happen in our own home?	MHRED	-	-	I am not sure about it.
	EmoSen	Sadness	-	I am sorry I could not hear anything.
	EmoSen	-	Positive	We like it to be at our home.
	EmoSen	Sadness	Positive	Oh no have been was looking for it, good I found it here.
	EmoSen	Surprise	-	Wow was it supposed to be like this?
	EmoSen	-	Negative	I hate to search this stuff around here at this time.
	EmoSen	Surprise	Negative	Oh my God! I am shocked by the sound coming from that room.
Call me crazy, but I don't think he was looking at Random's picture. It was like he was checking out that document. Oh, you mean, "the world's least-interesting doodle"?	MHRED	-	-	Yes you are right.
	EmoSen	Joy	-	That's really funny I love to doodle all the time.
	EmoSen	-	Positive	I am interested and want to finish these documents today.
	EmoSen	Joy	Positive	That's great I enjoyed taking the pictures of the doodles.
	EmoSen	Anger	-	Are you mad? this is not the way to say these things.
	EmoSen	-	Negative	Not a fan of these hobbies especially doodling.
	EmoSen	Anger	Negative	Don't screw up these documents or else I will kill you.
Oh no-no-no, give me some specifics. I love the specifics, the specifics were the best part! Hey, what about the scene with the kangaroo? Did-did you like that part?	MHRED	-	-	Sure I did like it.
	EmoSen	Disgust	-	That scene was horrible I almost puked.
	EmoSen	-	Positive	It was amazing I enjoyed it thoroughly.
	EmoSen	Disgust	Positive	Hey it was awfully good I never saw this coming.
	EmoSen	Fear	-	I was afraid when I saw it at first.
	EmoSen	-	Negative	Its not right I suppose it was very irritating.
	EmoSen	Fear	Negative	That movie scares the hell out of me especially the end part.
Chris says they're closing down the bar. No way! Yeah, apparently they're turning it into some kinda coffee place.	MHRED	-	-	Not heard of it yet.
	EmoSen	Surprise	-	When did this happen? I have no idea.
	EmoSen	-	Positive	Coffee is good I like it.
	EmoSen	Surprise	Positive	Wow it will be great I love coffee.
	EmoSen	Acceptance	-	That's nice to hear it will be a good place.
	EmoSen	-	Negative	I hate coffee I have never been to a cafe.
	EmoSen	Acceptance	Negative	I like this place alot but don't want it to become a cafe.

Selected Publications

1. M. Firdaus, H. Chauhan, A. Ekbal and P. Bhattacharyya (2021). More the Merrier: Towards Multi-Emotion and Intensity Controllable Response Generation. In *Proceedings of AAAI 2021*
2. D. Vershney, A. Ekbal and P. Bhattacharyya (2021). Modelling Context Emotions using Multi-task Learning for Emotion-Controlled Dialogue Generation. In *Proceedings of EACL 2021*
3. M. Firdaus, A.P Shandilya and A. Ekbal (2020). More to Diverse: Generating Diversified Responses in a Task Oriented Multimodal Dialog System. PLosONE, https://doi.org/10.1371/journal.pone.0241271.
4. M Firdaus, N Thakur, A Ekbal (2020). Aspect-Aware Response Generation for Multimodal Dialogue System. ACM Transactions on Intelligent Systems and Technology (TIST) 12 (2), 1-33.
5. M. Firdaus, N. Thakur and **A. Ekbal** (2020). MultiDM-GCN: Aspect-guided Response Generation in Multi-domain Multi-modal Dialogue System using Graph Convolutional Network. In *Proceedings of EMNLP Findings 2020*, *PP. 2318-2328*.

Selected Publications

6. Mauajama Firdaus, Shobhit Bhatnagar, Asif Ekbal, Pushpak Bhattacharyya (2018). *A Deep Learning based Multi-task Ensemble Model for Intent Detection and Slot Filling in Spoken Language Understanding*. In proceedings of 25th International Conference on Neural Information Processing (ICONIP 2018); 647-658; Siem Reap, Cambodia, 2018.

7. Hitesh Golchha, Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya; *Courteously Yours: Inducing courteous behavior in Customer Care responses using Reinforced Pointer Generator Network*; In proceedings of Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019); Minneapolis, USA; 2019.

8. Hardik Chauhan, Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya; *Ordinal and Attribute Aware Response Generation in a Multimodal Dialogue System*. In proceedings of 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019) ; Florence, Italy; 2019.

9. Mauajama Firdaus, Ankit Kumar, Asif Ekbal, Pushpak Bhattacharyya; *A Multi-Task Hierarchical Approach for Intent Detection and Slot Filling*; Knowledge Based Systems (KBS), Elsevier, 2019.

Selected Publications

1. Mauajama Firdaus, Shobhit Bhatnagar, Asif Ekbal, Pushpak Bhattacharyya; I*ntent Detection for Spoken Language Understanding Using a Deep Ensemble Model*; In proceedings of 15th Pacific Rim International Conference on Artificial Intelligence (PRICAI 2018); 629-642; Nanjing, China; 2018.
2. M. Firdaus, H. Golcha, A. Ekbal and P. Bhattacharyya (2020). A Deep Multi-task Model for Dialogue Act Classification,Intent Detection and Slot Filling. Cognitive Computation, Springer, https://doi.org/10.10007/s12559-020-09718-4
3. M. Firdaus, A. Ekbal and P. Bhattacharyya (2020). Incorporating Politeness across Languages in Customer Care Responses: Towards building a Multi-lingual Empathetic Dialogue Agent. In Proc. of LREC-2020 , Paris
4. M. Firdaus, N. Thangavelu, A. Ekbal and P. Bhattacharyya (2020). Persona aware Response Generation with Emotions. In Proc of IJCNN 2020 , UK

Summary and Conclusions

- **Empathy and Politeness:** Two important factors for building conversational AI
  - Increases human-machine conversational engagingness
  - Can yield a better context-aware, engaging and human-like responses
  - Very useful for many sectors such as retails, customer care centres, healthcare etc.

- **Presented politeness oriented dialogue generation framework**

- **Presented a new task for multi-emotion dialogue generation**

- Multimodality provides additional evidences for generating effective responses

Thank you for your attention!