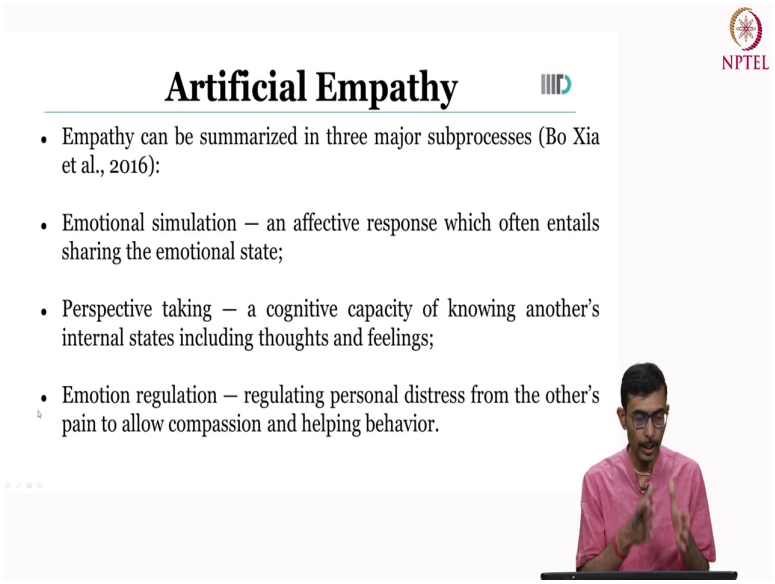
Affective Computing
Prof. Jainendra Shukla
Department of Computer Science and Engineering
Indraprastha Institute of Information Technology, Delhi

Week - 09
Lecture - 02
Development of Artificial Empathy

So, now let us having understood the basics of the empathy and now let us try to understand that how the artificial empathy can be generated.

(Refer Slide Time: 00:32)



So, for that let us first try to understand that how the empathy has been understood in the classical sense with respect to the humans. So, empathy has been summarized can be

summarized in as three major subprocesses, one is the emotional simulation, another is the perspective taking and another is the emotion regulation.

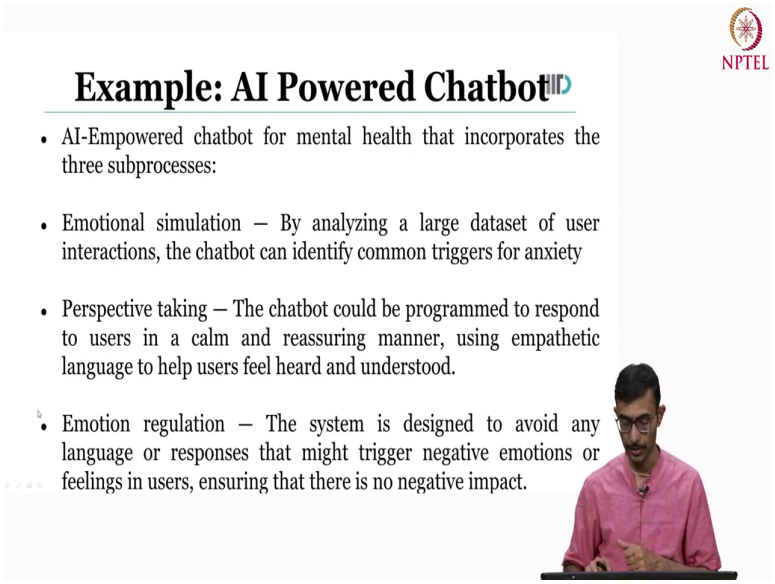
Now, what is emotional simulation? So, basically emotional simulation is all about trying to analyze and understand the emotional state of an individual. This is what we have been doing so far essentially by trying to understand the emotions of an individual in different, why different physiological and the behavioral cues.

Now, perspective taking is what there is a perspective taking? So, perspective taking is all about how can you understand that once you are going to give a response to the user, what the user is going to feel about it, how the user is going to feel about it, that is what the perspective taking.

So, what the user will feel you know how I would feel for example, if I would be in the user situation, if I would receive this particular type of response. We will see for example, and it will get very clear. And then of course, emotion regulation is all about that your aim or the aim of the empathetic responses or the interactions is that we want to provide a empathetic response, but at the same time we do not want to transfer any negative emotion to the towards the individual.

And that is what is the emotion regulation. So, traditionally we as humans this is how we approach the empathy or the empathetic interaction between each other. And having understood this thing will also help as simplify that what it means for an intelligent machines to be empathetic, ok.

(Refer Slide Time: 02:18)



So, let us see with this an example. Imagine that there is an AI empowered chatbot that you want to create, which is for the mental health issues you know where the individuals those who are having any type of mental health issue they can go and discuss and interact enhance some sort of counselling with a chatbot which is AI empowered. And now if we have to create this, we have to make this AI empowered chatbot a bit more empathetic then how will it look like?

Ok, the very first thing of course, it will have to look into the emotional simulation. So, when we say that it will have to look into the emotional simulation what it means, that it will have it needs to have the ability to identify the emotional state of an individual. And in this particular case it needs to identify the common triggers for the anxiety while it is making an interaction with the humans, right.

So, basically it needs and how will it be able to do that? Identify what are the common triggers for the anxiety or when the user is feeling anxiety. Of course, it will have it needs to have access to a large amount of data may be physiological and behavioral both the types of data.

And then it needs to have a training on that type of data and using that trained model it should be able to identify when there is a trigger for a anxiety or when the individual who is interacting with the chatbot is getting anxious or you know how it is interacting. So, that is what the emotional simulation is.

Now, of course, next thing would be that it has to understand that how the user wants to be felt like right, you know like it needs to understand that I need to make use of a language as a chatbot. I need to make use of a language which is going to help my users to feel that you know they are being heard and they are being understood. And how can it be done?

The chatbot can be programmed in advance to respond to the users in a you know very calm and reassuring manner and then that is how they are going to that is how this is going to interact with the users an intense user as going to feel good about it.

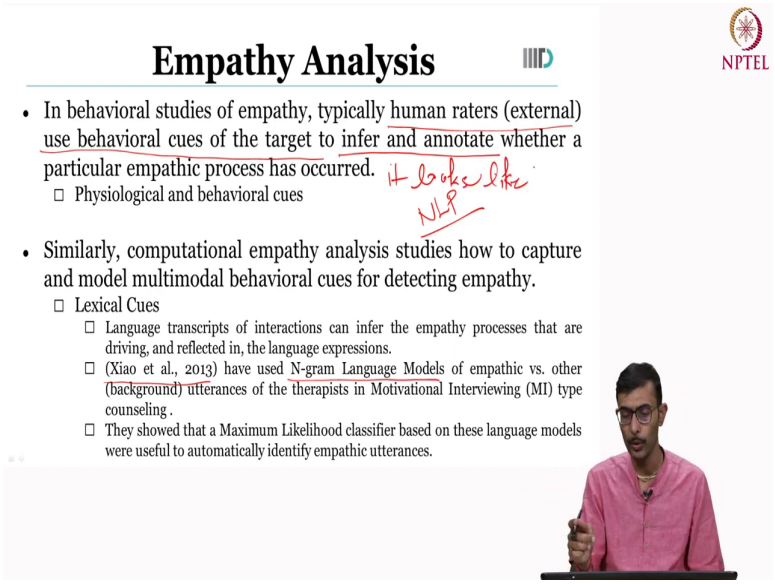
Because of course, you have to understand if you the chatbot having understood what is the emotional state of the individual if the chatbot is going to respond them in an aggressive way for example, in an aggressive manner or for example, in a higher pitch higher tone then of course, the users are not going to feel comfortable about it.

So, this is what it has to understand that you know depending upon the state of the user how should I respond so that you know they feel comfortable around it. And that is what is the perspective taking, how will they feel if I respond in particular way. And of course, then emotional regulation is ok, while all this is being done of course, the chatbot has to ensure that it should not trigger or any negative emotions or the feelings in the users. So, that there are no negative impacts around it, right.

So, for example, if the user has been talking about a particular scenario which is causing distress, it may it may not to necessarily poke you know the user to talk more about that particular scenario. Of course, keeping aside how the therapy works and all that, it has to see take all these things into account that in what are the different ways in which I am going to respond so that you know it is not necessarily going to invoke the negative emotions among the user which I definitely do not want to do it.

So, you know this is how the traditionally this is how the empathy can be understood and then this is how it can be transferred to a artificial intelligent systems.

(Refer Slide Time: 05:52)



Now, let us try to understand two major components related to the artificial empathy and the empathy generation. One is the empathy analysis and then we will next talk about the empathy simulation. So, basically empathy analysis is what? Empathy analysis is basically the

first part that we just saw, that we want to understand the emotional state of the user and we want to understand the overall the whether the interaction that is happening is empathetic or not.

Now, before we talk about the agents or the systems let us understand how traditionally it is being done or it can be done without making it automated. So, you know in the behavioral studies it is very very common to study empathy during the interactions and this could be you know empathy during the interaction for example, of a doctor with a patient and. so on so forth. And then based on this thing there are different types of training which are also provided.

So, for example, another example could be an interaction that is happening between a call centre employee and a user who has some issues who recently had some issues with credit card or something like that, right. So, you this is how you know like this kind of interactions that traditionally they have been analyzed and you know people are being trained on that to so that you know they can become a bit more empathetic while doing the interactions.

Now, if so in the absence of the AI or like without making use of intelligent systems how it was being done? And that is when you know typically what was being done here, that typically human raters or the external annotators what we call that human raters which are not part of the entire interaction they were involved in this case.

And they used to you know make use of the behavioral cues of the target which for example, in this case could be the patient or for example, could be the client, the customer who has called to the customer care they are they were they are going to look at the behavioral cues and they were trying to infer you know what is the overall state of the interaction.

Of course, if the user is continuously feeling frustrated about it maybe there is not much empathetic interaction that is happening, right. So, basically, they are external human annotators number one they are going to observe the behavioral cues of the target which could be client patient all that and having targeted that they would like to infer that ok.

Whether the empathetic interaction is being happening and at the same time they would like to annotate. For example, at what point of time the empathetic interaction was happening. And this is what an empathetic interaction looks like, right. So, this is how it has been done traditionally with the in the case of the in behavioral studies.

And when I said the behavioral cues, I mean there could be like physiological and behavioral cues both. So, for example, they for example, when they are making a call to the customer care of course, the only a modality that you have access to is of course, you guessed it right, is the audio modality.

So, of course, they can you know just monitor the audio modality of the client who has made a call to the customer care and using this audio modality they can try to understand that what is the emotionally state of an individual and at the same time you know overall whether it is being empathetic the interaction is empathetic or not.

Of course, in order to understand whether it is empathetic or not they also need to understand what is the interaction that is the entire interaction and the other party that is involved in the interaction which for example, could be the doctor or for example, could be the customer care employee who is taking the call of the client, right.

And of course, physiological cues as you rightly guessed physiological cues could be a bit cumbersome for example, while analyzing the customer care client. But for example, it could be very much feasible if the target is a patient who is interacting with the doctor.

For example, it is in a mental health counselling. Of course, we are not talking about very intrusive a physiological cues, but for example, the user can always come with a and can always be a variable sensor through which you know some physiological cues such as you know heart rate, even oxygen for that matter and things like that can be monitored and understood.

And this can be analyzed to understood that how the user is feeling and overall whether there has been an empathetic process that takes place. Perfect, now, so, you know so once this is how traditionality is being done in the behavioral studies.

Now, the idea of artificial empathy is what? That the computational empathy analysis studies that are focused on developing this artificial empathy, they want to similarly capture and model the multi-modal behavioral cues for generating the empathy. The way for example, an external annotator was doing in case of a behavioral study, right. So, this is what traditionality it has been doon, it has been done and you guessed it right that there are multiple physiological and behavioral cues.

Some of which we have already studied to understand how emotions in those cues can be understood can also be used the same type of physiological and behavioral cues can also be used to understand the empathy rather than just understanding the emotional state, they can be and the used to understand the empathy that is occurring in the entire interaction and how the user is feeling about it. And let us try to look at some of the these cues.

So, for example, one particular cue is as simple as the lexical cues. Now, lexical cues is what? Lexical cues is basically text based data. Now, you may want to ask ok, so, where there will be have access to the text data and how can text data be used to analyze the empathetic response? So, for example, for the entire interaction that is happening between a patient and a client, between a patient and a doctor, we can have access to the transcripts of the entire interaction.

And using those transcripts we can understand whether there has been any pathetic interaction throughout between the patient and the doctor. Of course, you guessed it right that we will have to use some sort of you know NLP here, some sort of you know NLP Natural Language Processing here in order to understand that what sort of features are there in this transcripts that we can analyze.

And on the top of those features, may we may have to use some machine learning models to understand whether there has been any pathetic reaction. So, let us talk about one particular study which is of quite interest in this case. So, for example, Xiao et al., in 2013, they made use of a N-gram language model to understand whether the language to understand the language transcriptions of the therapists and the clients in an motivational interviewing type counselling.

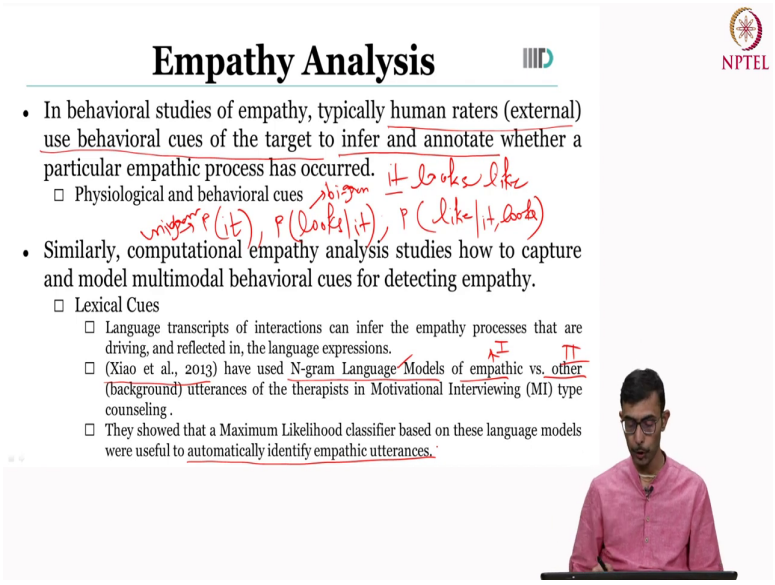
So, basically let us break this down for you. Ok, what is motivational interviewing type counselling? So, basically motivational interviewing type counselling is the type of the counselling where you know the whole job of the therapists or the doctors is to motivate the clients, motivate the patients towards certain goal or towards certain objectives and that is what is the motivational interviewing counselling is.

And of course, now in this case, what is N-gram type of modelling? So, basically N-gram type of modelling is what? N-gram type N. So, N-gram is basically you know you may have gone through it while we were talking about the emotions in the text, but basically N-gram is what? N-gram is a sequence of words.

N-gram is a sequence of N words for example, for example, if a if we were to take a simple example of you know, let us say that you know it looks like, let us write a simple example, it looks like ok, let me just remove this this word now for now, it looks like. So, ok, how many words are there? There are three different words.

So, what does it mean? How can we model it? How can we model it using N-gram? So, for example, if we are using a unigram, then maybe we are going to look at only one single statement.

(Refer Slide Time: 13:58)



So, for example, I just as simple as that probability of it, probability of occurrence of it could be something that is a unigram model. Similarly, imagine if we were to look at, sorry, if we were to look at the probability of let us say looks given that it has already occurred, then that is what you can call it as a, this is the unigram model and this is what we can call it as a bigram model.

So, this is a unigram and this is a bigram model, right. So, basically when we are looking at two words that is a bigram model, when we are looking at one word that is a unigram model. Similarly, we can also have a trigram model, you guessed it right, where for example, we want to understand what is the probability of occurring like when the model has already looked at it looks, right.

So, basically that is what is a trigram model. So, there are different characteristics of these unigram models and for depending upon what is the type of the task that you are looking at, you may want to use bigram, trigram or even higher sequence models. So, basically in this particular study, what the Xiao and his group did, that they made use of a N-gram model.

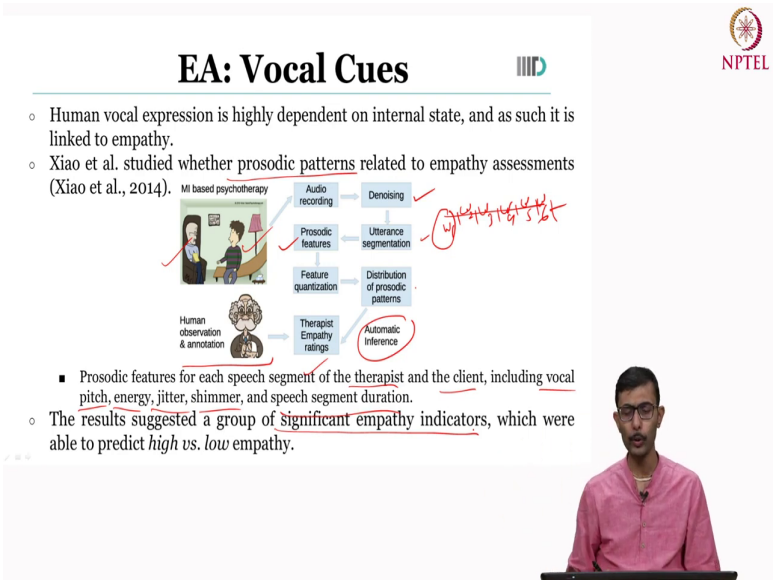
In this case, they used the trigram model to understand that, to understand that, to analyze the entire, the language transcriptions between the therapist and the client. And basically, what they did? They were able to show that just by using a maximum likelihood classifier, which is a very common type of classifier, which looks at the maximum likelihood of assigning a class.

So, for example, in this case, they were looking at a classifier, which could assign the entire transcription, whether it is belonging to a empathetic category. For example, or whether for example, it was belonging to other category empathetic or other category. So, there are two classes, right. I hope you can understand this thing. This is class 1, and maybe this is class 2, which is other, which could be non-empathetic or which is you know neutral even.

So, basically what they showed in this study, that by making use of a maximum likelihood classifier, we just looks at the maximum probability of occurring a particular, of assigning a particular class on these language models, which is you know N-gram language models that we just saw, they were able to automatically identify the empathetic utterances. And that is fascinating, actually.

So, basically, you know they were just looking at the language transcriptions, making use of the N-gram language model, applying a maximum likelihood classifier on the drop of it, and they were able to identify when there was an empathetic interaction happening, when there was no empathetic interaction happening, right.

(Refer Slide Time: 16:40)



So, for example, this is how you can make use of the lexical cues or the text based data to analyze the empathetic process, to understand the empathy, not only the emotional state, but also to understand the empathy, whether the empathetic interaction has been has happened.

Of course, you got it right, that you need to have the annotated data in order to make a classifier supervised classification work in this case also. Similarly, you know of course, you know, like test is has been is limiting to certain extent, but nevertheless, one of the most common modalities where empathy has been understood in the research, in the literature, is the vocal cues.

So, basically in the voice, and of course, you already understand that the vocal cues or the voice modalities of the humans is highly dependent on the internal state. And hence, this is a very, very good indicator of the empathetic, the empathetic state of the individual, right.

And there are there has been many different studies that has been done on this thing. For example, one study again by Xiao et al., and his group only that was done in 2014, they showed whether the prosodic patterns. So, prosodic patterns are basically again, if you recall your class of emotions in speech, then you may recall the prosodic patterns, the features of the speech.

Prosodic patterns related to the empathy assets assess, they analyze these prosodic patterns related to the empathy assessments. For example, again in the same type of setting where you know like there was a motivational interviewing based therapy happening.

So, you can look at this diagram and understand there is a therapist, of course, of course, there is a client, the therapist and the patient they are talking to each other. So, basically the whole idea was that entire audio was getting recorded. The next step, of course, audio recording was being done, that is how you obtain the vocal cues.

Of course, you will have to do some sort of denoising on the top of it, background noise or some sorts of other noise. So, then you first you recorded the audio, you recorded the entire audio, you did the denoising to remove the low frequency or the high frequency noises. Then of course, you may want to segment the utterances.

So, for example, the way we were doing trying to identify a word unigram, bigram and N-gram and so on and so forth, in the same way you want to segment the entire speech into different utterances, speech corresponding to one word, speech corresponding to two word and so on and so forth.

And then of course, you may want to, you know after the segmentation of the utterances, you may want to extract the prosodic features within those segments. So, it may happen that you

know you may have entire segment of audio like this. So, for example, time t, but then you may want to chunk in different segments and for example, where this is belonging to maybe word 1, this is word 2, this is word 3, this is word 4, this is word 5, this is word 6 and so on and so forth.

So, now and of course, this is just an example, you can definitely segment the entire audio sequence depending upon so many things in one word in two words or you know depending upon, for example, as simple as that, that you can just segment it within of with the segments of 5 seconds, 10 seconds itself a uniform segment, 10 seconds each and so on so forth.

So, once you have segments then you know for each particular segment, then what they did? They extracted the prosodic features related to the speech and of course, they did the feature, some feature quantization on the top of it, they looked at the entire distribution of the prosodic patterns in the different over the different segments.

And of course, you need to have always, as always you need to have an external annotator and external observer who can look at the entire interaction and can annotate for you that, ok, this particular segment, yes, it was empathetic, this particular segment it was not empathetic.

So, there is an external annotator or observer who provided the empathetic, empathy ratings of the therapist that, ok, whether the therapist for example, was being empathetic in this particular segment or not. And then of course, you know, once you have the empathy ratings, then you simply ran some model and then you tried to understand did the automatic inference of the entire thing.

So, that was the model that for example, Xiao et al., proposed for the vocal cues also and then you know what the, what so, ok. So, for the prosodic features for each speech segment, of course, they looked at the therapist and they looked at the client client's data as well. Now, you may want to understand, ask, ok, if they wanted to understand the therapists empathy ratings, then why did they collect the client's prosodic features, why did they look at the client's audio signal, any guesses?

Ok, I mean, so, this is quite easy to understand. Your idea is to understand what was the empathetic response on the, on the target, target here in this case is the client. You want your client to have an empathetic feeling. And unless until you analyze the emotional, the vocal state, the vocal modality of the client, you can never know that whether the therapist was, whatever the therapist was intending, whether that was being having an impact or not, right.

So, of course, you will have to look at the both the therapist and the client and then of course, the prosodic features included many different features such as, you know, vocal pitch, energy of the signal, jitter, shimmer and speech segment duration itself, for example, one simple feature in this case, ok.

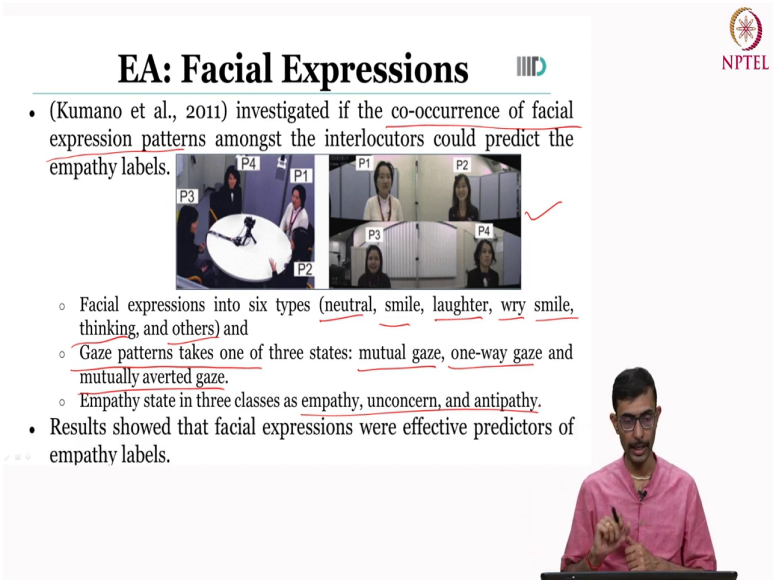
So, they looked at these entire features and what they could show with the help of results, that a group of significant empathy indicators, they were able to find a group of significant empathy indicators which were able to predict, for example, low versus high, which were able to predict, you know, what were low and what was, for example, high empathy in this case.

And I think I have some example here, if I can just understand. So, for example, they were able to understand that an increased distribution of medium length segments with high energy and high pitch was associated with a low empathy situation. And it should not be surprising because of course, you know, what it means that, if there is an segment with high energy and high pitch, may it may suggest that, you know, the therapist is making use of a louder voice and is having a raised intonation.

And because of this, the therapist response within that particular segment cannot be considered or may not have been considered as high highly empathic. So, for example, these are some of the ways in which they were able to understand and identify that how the different prosodic features they are related to the lower the high empathy, empathy empathetic responses.

And this is how not only they were able to show, that for example, using vocal cues also, you can analyze the empathetic situation of the of a particular interaction, perfect.

(Refer Slide Time: 23:23)



So, I will not talk about how that how the empathetic responses can be evaluated in all different modalities, but I guess you got the idea. So, of course, you already saw that empathetic responses can be understood with the help of the lexical analysis; you just look at the text data of a transcription of the entire interaction.

You can also look at, for example, vocal cues and understand whether the empathetic interaction has happened. Similarly, it is not a trivial to understand that it is just quite trivial to understand, for example, that the facial expressions has also, if you can analyze the facial expressions of the respondents, you can simply understand whether they have been feeling experiencing an empathetic interaction or not.

And motivated by this idea, for example, one of the first studies that Kumano et al., Kumano and his their group did in twenty 2011 was they wanted to understand if the co-occurrence of

the facial expression patterns between different individuals who were part of the same group can also be related the empathetic labels, can also be related to the empathetic responses, right.

So, for example, you can see in this particular picture. So, on the left hand side you have a high level view of the entire interaction the way it happened. Of course, this is an image from the same paper, where for example, there were four participants, they were sitting facing each other and of course, having a discussion and this is the individual participants' camera highlighting individual's a facial expressions.

So, now in this study what they try to do? Of course, to take a take a step back, they wanted to understand if the co-occurrence of the facial expression patterns means the facial expression of participant 1 for example, with participant 2, with participant 3, with participant 4 can be used to understand that there has been an empathetic response or there has been an empathetic interaction within that particular segment of course.

So, what they did? In this, they segmented the facial expressions into they categorized the facial expression into six types. So, you already know, ok they neutral, smile, laughter, wry, wry smile and thinking and of course, others. Others could be something, you know, which is not among these.

For example, which could be disgusting, shaming and all those kinds of things, right. So, basically, other feelings. So, there were the six types in which they classified the facial expressions of an individual, how to do the classification of facial expressions? I think you already know the answer to it; we have already talked about it in emotions in facial expressions.

Next, what they did? They not only looked at the facial expressions, they also looked at the gaze patterns. And you can understand that ok, if when we are looking at the facial expressions along with the gaze patterns, it means simply we are creating a multi-modal system.

And of course, in this multi-modal system, the reason why we make use of a multi-modal system is because the hypothesis that the response of a multi-modal system would be better in comparison to a uni-modal system such as making only making use of a facial expressions, ok.

So, in this case, they also looked at the gaze patterns of all the participants and then classified the gaze patterns into three categories. One is the mutual gaze, what is mutual gaze? Mutual gaze is the labelling where the participants, they were looking at each other.

For example, they were like participants 1 and 2 and maybe they were looking at each other. So, that there was a mutual gaze between these participants. Now, similarly, what is one-way gaze? So, one-way gaze is basically ok, one of the participants is looking at other participant, but other participant is not looking back at the first participant.

So, that is one way gaze. And what is mutually averted gaze? So, you guessed it right. Mutually averted gaze is basically neither of the participants are looking at each other. So, ok, nevertheless, facial expressions, six states, gaze patterns, three states, our their objective was to understand if we can look at facial expressions and gaze patterns can be predict what is the empathy label of interaction in this particular segment.

So, empathy of course, then you know in this case, they rather than using a binary classification, they made use of a three class classification problem where empathy was empathy, unconcern and antipathy actually. So, they have looked at antipathy as well as the third class. So, there were three classes. So, you got the idea, right, ok.

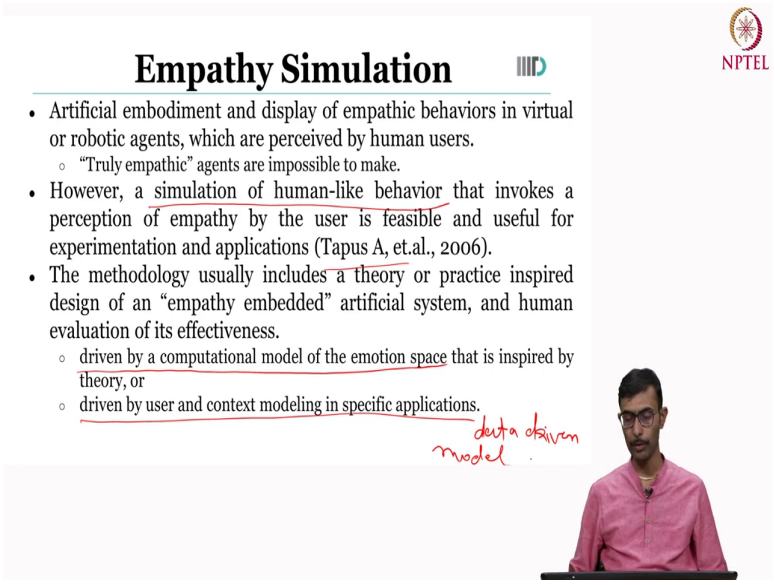
Facial expressions, gaze patterns were the data; input data, empathy state is the label. Now, let us see what they found. So, the results showed that the facial expressions were effective predictors of the empathy labels, not very hard to understand, but nevertheless, it was good to see that using an automated analysis, they were able to understand that how the empathy can be analyzed in maybe making use of facial expressions.

And I think if you look at, go ahead and read the paper, you will find that they had some interesting observations about the gaze patterns as well, and also about how and what was happening when they were mixing the facial expressions and the gaze patterns. So, nevertheless, overall, you got the idea that in order to understand the, the in order to analyze the empathy during an interaction, you can look at the physiological data; you can look at the behavioral cues.

And there are n number of examples, such as we just talked about three, which looked at the how, for example, lexical cues; how, for example, the vocal cues and how, for example, the facial expressions can be used to understand the label of empathy between during an interaction, ok.

And similarly, you can just take the idea further and maybe make use of multiple modalities, other modalities, such as the ones that we talked about during our previous classes, to understand the emotional states, you can also use them to understand what is the relation of the, this all those modalities with the emotional, with the with the empathy during a particular interaction.

(Refer Slide Time: 29:17)



So, this is how you can do the analysis of the empathy, perfect. So, now the next step is what? Next step is the empathy simulation that is closely related to the development of artificial empathy. So, once you have understood that how the how the empathy can be analyzed, of course, the next step is what?

To understand how can we have an artificial embodiment and display of these empathetic behaviors in virtual or robotic agents, which can you know display artificial empathy as it can be perceived by the human users. So, we already know that we now can see how it can be analyzed.

Now, the next time would be to put them into the virtual and the robotic agents. Of course, a bit of warning here, whatever empathetic responses these agents are going to display, they are

not going to be truly empathetic, right. They are not going to be truly empathetic, is as simple as that.

Of course, because they are not sentient beings, any virtual agent or robotic agent is not going to be a sentient being. And unless and until it is a sentient being, it can never feel the emotion and since it can never feel the emotion, it cannot display a truly empathetic empathic response.

So, I hope we understand, but nevertheless, there is hope. And the hope is that it has been shown by previous research such as for example, Tapus and (Refer Time: 30:40) Tapus and her group that we do not really need the agents to show a truly empathetic behavior.

What we really need, that even if we can have a simulation of a human life behavior by the virtual agents that can just invoke a perception of the empathy by the user, it is of course, this is going, this is feasible and this is useful for most of the experimentation and the application purposes, right. So, the idea is not to make the agents feel empathetic, but just to simulate the empathetic behavior which is a human like empathetic behavior in the agents and that is going to serve us well in the longer duration.

And that is what gives us a lot of hope actually and that is why we are talking about how should what is artificial empathy and how can we generate the empathy and of course, in the hope that this artificial empathy even though it is not true, it can really help us in making the response more empathetic and hence more interactive.

Now, it turns out there are then there has been mainly two different directions using which the previous research or the literature has been trying to do this simulation of human like behavior. So, of course, we understand truly empathetic response cannot be generated. So, we can just do a simulation of a human like behavior. So, previous research has been mainly focusing on two different methodologies to generate this simulation of a human like behavior.

So, for example, one particular one particular direction has been where it has been driven by the computational model of the emotional space. So, basically it is very very driven by the

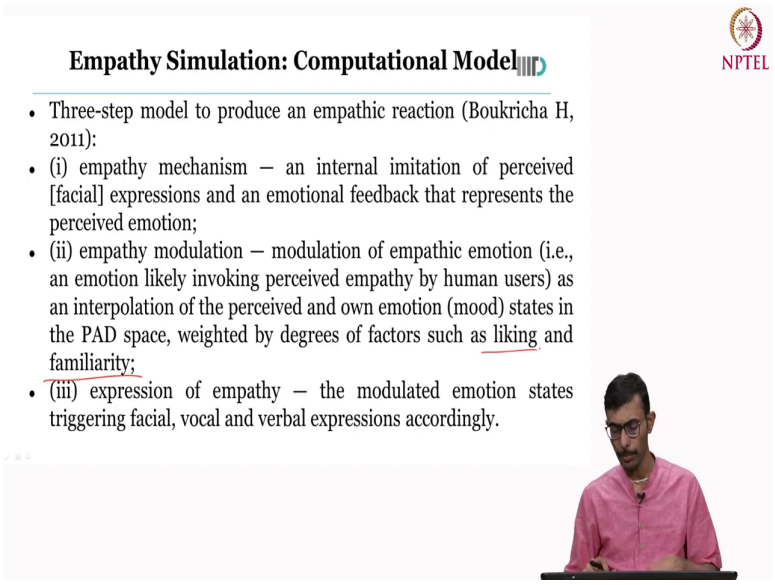
theory that is behind the generation of the empathy in humans and the idea is the broader idea is.

If we can understand the theory behind the generation of the empathy in the humans, we can somehow you know create a cognitive model or computational model of that and using that we can generate the empathy in the virtual agents or the robotic agents. Of course, it is a bit not so easy to do.

Other approach that many of the studies they have been following is it is mostly driven by the data where it is driven by the user and the context in which we want to apply the empathy to a particular application. And if we have data related to that we are just going to other the previous research has been trying to make use of that data to generate the empathetic behavior. So, we will basically look at all these kind of things you know in the different ways to understand that how can it be done, perfect.

So, now as I said there are two ways one is the computational model and one is the let us call it as a data driven model. The second one we can call it as a data driven model. So, let us try to understand both of them, how can we make use of these two models.

(Refer Slide Time: 33:45)



How the previous research has used these two different models methods to generate the empathy. So, computational model as I said before, computational model is basically we want to understand how the empathy is being generated in the humans and when can we replicate the same in the machines. So, there has been many different models that has been proposed in this sense to understand how the empathy has been generated among humans.

So, for example, one particular model that has been proposed by Boukricha and his group is of generating the empathetic reaction can be can has three steps. So, basically the first step is the empathy mechanism. So, basically this is what is the understanding of the emotional state of the user who is of the target in this case. Other is the empathy modulation.

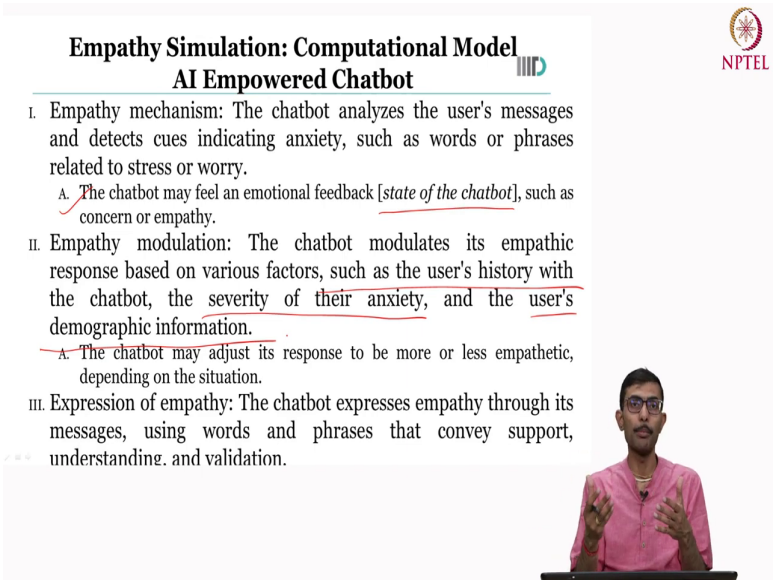
So, basically what you want to do? You want to understand that for example, as simple as that what is the level of the emotional state, what is the severity of the emotional state, what is

the for example, the relationship between the target and the for example, the other observer, other interaction other individual who is interacting?

And for example, what is the demographic information associated with the user using all these, you may there can be a modulation of the empathy, there is usually a modulation of the empathy such as and it is weighted by many different degree of factors for such as you know. Of course, if you have a sense of familiarity between the patient and the doctor, of course, there is going to be a more empathy. If there is going to be a liking between the patient and the doctor of course, there can be a bit more empathy.

So, all these aspects comes under the empathy modulation. So, first trying to understand what is the emotional state of the user, second how can you modulate, how can it be modulate and third thing is the of course, is the expression of the empathy is basically making use of the different physiological and the behavioral cues to express the empathic behavior, which of course, look like the empathetic behavior to the users in this case. And this is how traditionally one three-step model has been produced.

(Refer Slide Time: 35:55)



And let us try to see for example, how can this be applied in the case of a, if we were to use this particular model, we were to use this particular model to generate empathy, how can it be applied to again our earlier example, which is the AI empowered chatbot for the mental health purposes. So, basically what you can do? Of course, the first step is the empathy mechanism.

So, basically you want to understand and analyze what is the user's emotional state. It is simple in the case of the chatbot, which is looking which is interacting with the human, maybe of course, it has access to the users' wise data for example, we just saw. It has access maybe to the users' you know transcriptions, the entire transcription of the interaction that is happening.

So, basically it can look at those cues and it can look for specific words or the phrases, for example, it can look at prosodic features, it can look at the N-grams. Remember, we just talked about that in previous and it can understand that you know, like what is the emotional state of the user in this interaction.

Having understood what is the emotional state of the individual, then it can also you know understand that ok, that it can also adapt a particular, this is very, very important. Please pay attention to this particular point. This is very, very important. The chatbot can also adapt a particular emotional state in response to that in response to that particular emotional state of the target.

For example, if the target is feeling stress, anxious and so on so forth, then of course, you do not want your chatbot to be responding in a very jolly manner. It needs to be, it needs to look like that it is concerned about the users concern user's current state or target's current state, right. So, it needs to have a mood, we can call it as a state of the chatbot or we can call it as the mood of the chatbot.

Should be a bit of showing concern, should be a bit of empathy rather than for example, showing a very jolly behavior and smiling always, you know, like laughing for example, which is not going to be very sensitive towards this interaction. So, that is the empathy mechanism.

In next case, of course, you know, once you have understood user's emotional state, once you have understood what the chatbots mood could be according to that, next the chatbot would like to modulate its empathy response and based on so many different things. For example, if it has access to user's history with the chatbot, it may know some other context, it may know some more information and it can modulate with respect to that.

Of course, it can also take into account what is the severity of the anxiety. Maybe the individual is feeling a bit, you know, like less anxious, maybe individual is feeling highly anxious and according to that, the state of the chatbot also can be, you know, programmed or

the chatbot can, the whatever the responses that the chatbot now is going to take, can be modulated with respect to that.

And of course, as I said, it also needs to look at the user's demographic information such as for example, what is the age, what is the gender and so on so forth. Maybe they all may play a particular role in this case. And they all using all this information, the response of the chatbot can be modulated, right. Now, for example, the chatbot may just, its response to be more or less empathic depending upon the situation. If the user is very much concerned, maybe more empathetic, if the user is less concerned, maybe, you know, less empathetic so, for example.

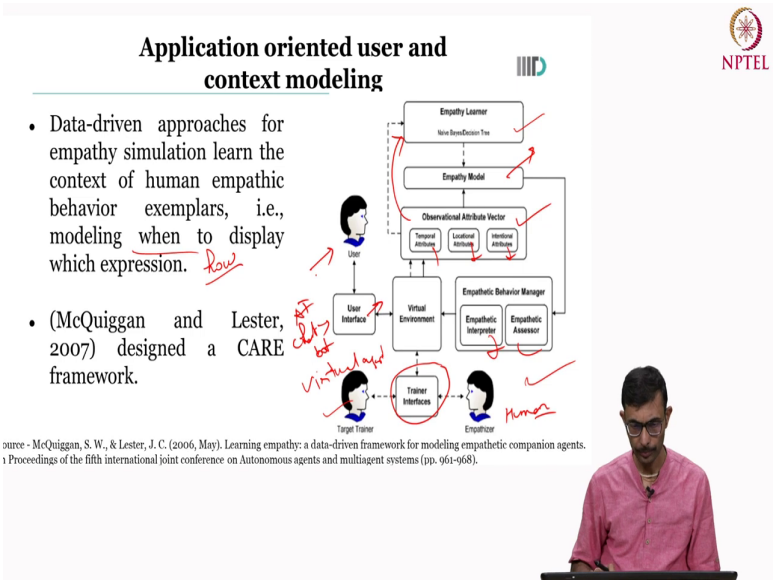
Now of course, the next thing would be having understood what is the emotional state or the empathetic state of the user and having understood what is the degree through which it is, it has to be modulated. Now, the chatbot has to express the empathy. Now, how can chatbot express empathy?

So, basically the chatbot can express empathy through its messages. For example, if it is just interacting using a chat, using words and phrases that convey support, understanding and validation. And of course, you know, like this, the same thing can be done maybe by making use of a vocal cues, voice modality. If for example, the chatbot is like voice assistant.

So, this is how the expression can be done and you rightly guessed that all the, for example, you may not want to have a very high pitch or high tone of the voice of the speaker, of that chatbot when it is interacting, when it is expressing the empathy because of course, we just saw in a few moments back that in one of the studies Xiao et al., for example, they showed that the high pitch and the high energy is associated with the less empathy.

So, for example, this is how you modulate the degree to which the empathetic response has to be shown to the target user. So, I hope. So, far it is making sense that how can the empathy can be stimulated with the help of a computational model by for example, in a AI empowered chatbot.

(Refer Slide Time: 40:55)



So, that was the first way in which you know the empathy can be generated with the help of a computational model, trying to understand how the how the humans, how the empathy is evolved in the humans, how it is expressed in the humans and having understood that we may want to replicate the same for in artificial agent which may not be so easy to do.

This is a bit difficult actually and then we will talk about what are the challenges associated with it. But then the method that is more common among the community and that has been exploited more is basically the data driven approach. So, in the data driven approach what happens that its very very specific to a particular application and it is of course, oriented towards the user and based on a particular context only. We will see more about it in order to understand.

So, basically in this case what happens? It tries to learn the context of the human empathetic behavior and it tries to understand for example, ok in a particular scenario when the humans they display emotions and how they display emotions, ok. It is not concerned that how what is the cognitive model of the generation of the empathy in humans, but it is more concerned.

Ok, Can I understand that when they are displaying emotion and of course, other thing that the I would like to understand in this case how they are displaying emotion that is it. I do not want to understand the model that is behind when and how, ok. For example, which is guiding the cognitive mechanism behind it? And that is what the data driven approach has been.

So, for example, in one particular work that is by McQuiggan and Lester in 2007, they designed a framework which they call it as a CARE framework. So, basically in this CARE framework what they did was in this particular CARE framework, they collected the behaviors of a virtual agent. In this case, for example, they collected the behavior of a virtual agent that was being manipulated by a human who was acting in an empathetic manner, right.

For example, feeling frustrated when the user was losing a particular game. And in this entire way, the entire data of being was being recorded and then entire data for example, was then further used this entire data was being recorded. And then it was further used to train for example, a particular virtual agent to display to display empathetic behavior in response to when the user was displaying the behavior and or when the user was or and accordingly how the user was displaying the that particular behavior.

So, let us try to see this with the help of this particular example. Imagine that this is your virtual agent, right this is your virtual agent which you want to train to show some empathetic behaviour. And then of course, you know then there is an for example, there is an that there is a human who is an empathizer in this case, you want to see how this human is doing and then you want to learn from this particular human.

Basically, what is going to happen in this case, there is a trainer interface and through this trainer interface, this virtual agent and the empathizer, the human are going to interact or for example, through this virtual, through this interface, the humans are going to control how the this virtual agent is going to respond.

Of course, it is completely controlled by the humans, this virtual agent's behavior for example, what is the type of the message they are going to send or for example, what is the they are going to say, everything is being controlled by this human.

Now, this virtual, there is an virtual environment which has access to all the data of this entire interaction. When I say that the virtual environment has access to the entire data, which could be you know that all the temporal attributes. For example, you know in which particular time what was happening, locational attribute, all the special attributes, special properties for example, and then what was the intentional attributes.

So, basically all sort of you know physiological and the behavioral cues that was getting observed and in what particular time they were being observed, what was the synchronization between all the cues and everything. So, all this data virtual environment was able to observe.

Now, once the virtual environment, it had access to all this particular data, what it did? It you know you can see this particular relation. So, basically using all this data, it trained for example, there is an empathy learner module, which is an intelligent module. For example, in this case, they simply made use of a Naive Bayesian decision tree. So, you may recall the Naive Bayesian decision tree from your ML classes before.

They simply made use of a Naive Bayesian decision tree, which is a very simple model actually, to train on this particular observational data, where you know for example, they had access to all the physiological and the behavioral cues and when was it happening and at the same time, what was the type of the response that was being generated by the by the by the humans and in response to what?

For example, we just saw that for example, when the user was playing a game and when the user was losing the game, then you know like the other agent or the human is going to feel the human is going to feel frustrated about it when they are losing the game. So, they are feeling frustrated and how they are displaying the frustration, maybe you know saying I am losing the game.

And when they said it and before that, what they felt after that, how they felt it. All this data is being fed to a machine learning model, which is a empathy learner model in this case. They just simply made use of a Naive Bayesian decision tree and then with the help of this thing ultimately, they were able to create a model that was being deployed in real time.

So, basically this model was what? Now, this model was they making use of this model, now they can do two things. They can this model is used to feed to this empathetic behavior, general manager, which can you know generate this empathetic behavior and it can do those two things.

Of course, it can understand, ok it can understand ok, whether the entire interaction is being empathetic or not, we just saw how? It can interpret the empathetic interactions and it can understand ok, when should I display the empathy and what is going to be the empathetic, how should I display this type of empathetic.

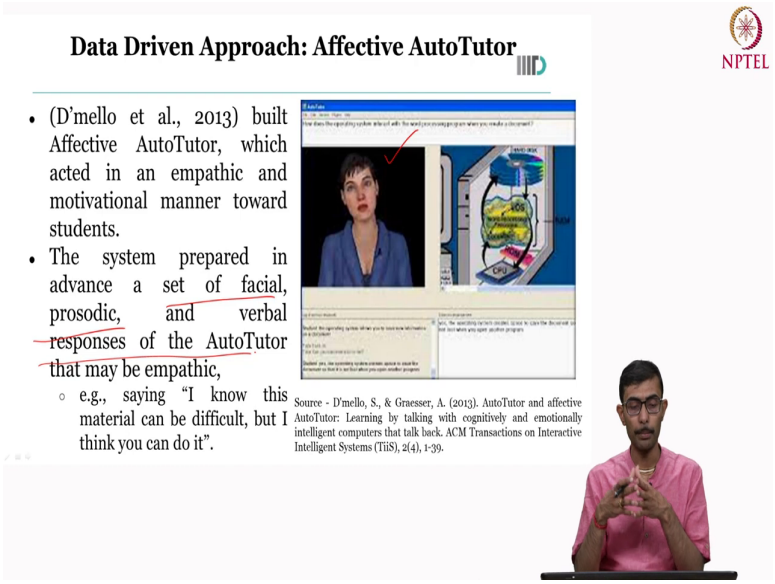
So, for example, maybe increase my pitch increase my tone, for example, or reduce my pitch, reduce my tone or for example, say that ok, oh I am not have feeling happy about it and things like that. So, basically, this is the behavior, this is these are all the things that the empathetic manager has knowledge to now, thanks to this empathetic module, which is being trained on this observational data with the help, thanks to with the help of the humans.

And then of course, you know, all this with the help of this, then you are going to make use of a user, then all this is being fed to a user interface through which a user, a new user is going to interact with. So, imagine this is your AI chatbot, imagine that this is your AI chatbot with

which a new user is now, another user is now interacting. But now since AI chatbot is being trained on the data that intern is being generated from the humans.

Now this is going to show empathy in terms of time and in terms of some expressions based on the similar data. And it to an end user, it may look like a human-like behavior or it may look like a human-like empathetic behavior right, and that was the end goal of the entire model. So, pretty interesting framework that these guys have created. Of course, for more information, I would invite you to please go through this particular paper.

(Refer Slide Time: 48:24)



Similarly, so, care framework, for example, it has not been only this one single work that has been used to generate the empathetic behavior following the data driven approach. As I said, most of the current work and some of the previous work has used the data driven approach only to generate the affective behavior, empathetic behavior.

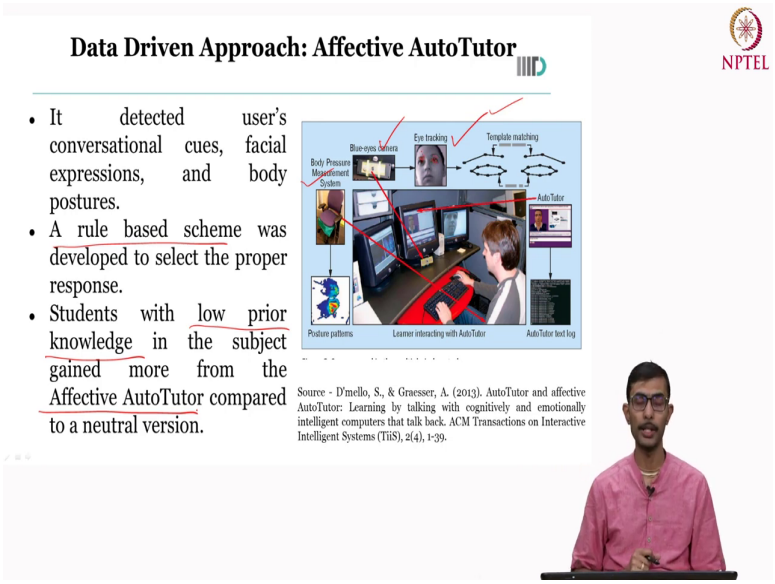
So, for example, one from D'mello and their group in 2013, they built a very nice module which is known as Effective Autotutor where they created an you know artificial virtual agent. You can see, I hope that you can see here, there is an artificial virtual agent which is the programmed, this this virtual empathetic agent to act in an empathetic and motivational manner towards students.

For example, those who are feeling frustrated about the module that is being taught and you know the course that is being taught to them in real time. And of course, the way they were doing it, as I said, it was entirely data driven. So, the system of course, prepared in advance a set of facial, prosodic and verbal responses. What sort of facial, prosodic and verbal responses?

Of course, they were able to look at you know the response of the teachers. For example, that how the teachers respond, what are the type of facial expression they make when they see their students are struggling, what is the type of the speech modulation that they do when they see their students are struggling and what is the type of the verbal response that they give you know when they see their students are struggling.

So, for example, they may say that, ok I know that you know this material maybe can be a bit difficult, but ok, do not worry, like we can do it together and let me help you understand it for example. So, all these things were programmed in advance for this affective agent.

(Refer Slide Time: 50:10)



And then of course, in order to do it in real time, they were detecting, for example, if you look at the, now I hope that this particular diagram is visible to you. So, then they were looking at many different sorts of cues of the users, such as for example, they were looking at their facial expressions, right. So, for example, they were looking at their facial expressions, they were also looking at their gaze patterns and of course, you know they were looking at their postures, body postures also you know.

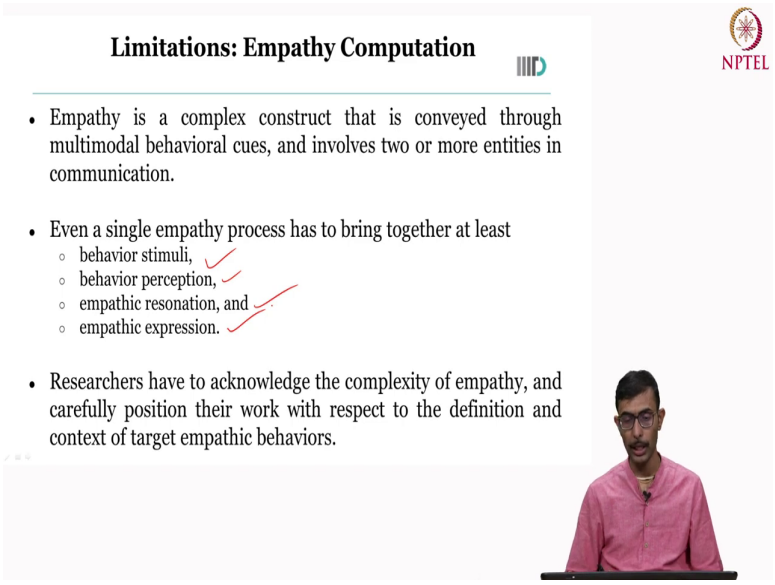
So, for example, if the seat they were where the user was sitting, they were able to monitor the movements within that chair itself and they were looking at all different types of you know body postures, facial expressions and of course, they had access to the conversational cues as well in terms of the voice data. And by observing all these thing, what they were able to do? They created a rule based scheme to select the proper response.

So, for example, as simple as that if the user is feeling frustrated, for example, if the users, for example, if the user speech the pitch and the tone or energy in the voice is looking like this, maybe the user is frustrated. So, generate this particular type of response.

So, this can if then if else kind of rule. And actually, they evaluated this model and they were able to show the results with the help of the results that the students who had the low prior knowledge in the subject, they gained more knowledge with the help of this affective AutoTutor in comparison to a neutral version of it.

So, basically since the user tutor was affective, empathetic, maybe they were able to gain more understanding of it, right. So, for example, this is another pretty nice interesting study where it was entirely data driven, they were able to look at the previous data and they were able to generate a affective tutor, which was empathetic in its response, perfect.

(Refer Slide Time: 51:58)



So, now I hope you understand that there are two different ways in which the community in in which the previous research has been able to generate the empathy in the artificial empathy. One is of course, by making use of the computational model of the empathy, other is the data driven approach.

Now, let us try to understand what is the what are some of the limitations of both these approaches. So, for the computational model of the empathy, of course, very first problem is that the empathy itself is a very complex construct. And it is being expressed through multiple multi-modal behavior cues and many of course, it involves walls at least two individuals to show an empathetic behavior, you not remember a doctor and a patient, a customer care agent and a client for example, and maybe in many cases it can be more such as for example, when there is a group conversation happening.

So, this is a very very complex scenario and hence it makes the entire job very very challenging to be being able to understand what is the group dynamics that is happening and what is the you know happening when the there is a (Refer Time: 53:03) conversation that is taking place and what is the cognitive modelling behind it.

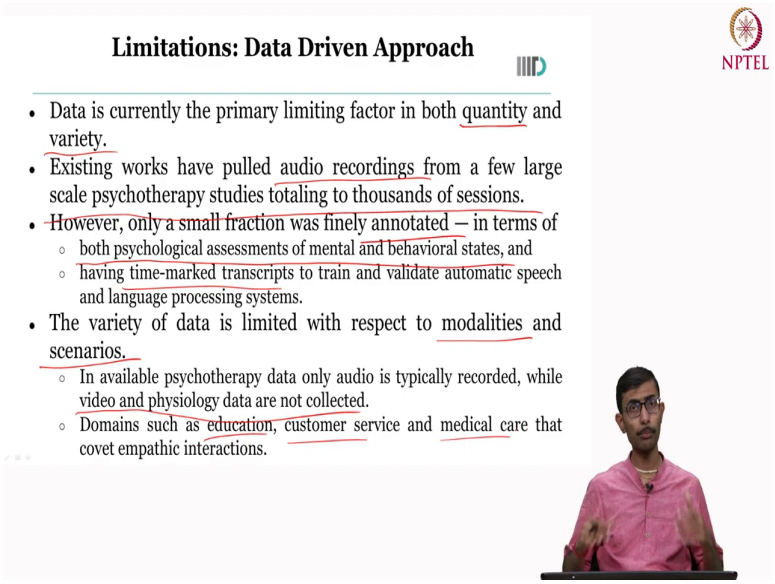
So, there is not its not has not been very easy to understand it. And in order to understand it like there are many different things that they have to understand. Of course, they have to understand what is the stimuli that generates that triggers the empathetic reaction and then how can we perceive the behavior that is there within two agents or that is there within a group.

Of course, we already see that physiological and the behavioral cues and these kind of things, but of course, what could be the most feasible and optimal way to do it. Of course, once we have understood the perception of it the behavior, then we need to understand how can we you know modulate the empathy by looking at very different factors such as for example, the demographic information, the severity for example, of the emotional state and so on so forth.

And then of course, we also need to look at that you know once we have understood this thing, how can we express the empathy. So, these are the different things in which that they need to look at in order to create an empathetic response or behaviour. And since this has been this all together, this is quite complex. Hence, you know and also we have to understand that the definition of the empathy itself is also not commonly agreed very commonly agreed.

Of course, there has been some definitions that people have been using that the researchers have been using in different context, but neither the definition of the empathy has been agreed nor there has been a very established understanding of how these empathetic is the cognitive model of the empathy. And hence, to make use of that model computational or to put it into the virtual agents, it is a bit challenging, it has been a challenging task so far.

(Refer Slide Time: 54:48)



So, that is about the computational modelling of the empathy. Now, what is the problem or the limitations of the other model which is the data driven approach? I think you may have guessed it right. With the case of the data driven approach, of course, the problem is the data itself, right. And it turns out that in case there has been two different types of limitations that the researchers have been facing. Both in terms of the quantity and both in terms of the variety or the quality.

Now, what do you mean by the quantity? So, it turns out that existing words as we talked about before also, most of the time they have been making use of the audio recordings only, that is a unimodal system to understand the empathetic reaction that is one. But those audio recordings also have been obtained only like you know very few large scale psychotherapy

studies that could be totaling to the you know thousands of sessions maybe not more than that.

And then also you have to understand there are lots of ethical issues involved around getting this kind of data. Hence, there has been not very good set of availability of this type of data and now on the top of that we do not just want the data.

You have to understand as a machine learning if you want to create a machine learning model for it, then you want the entire data to be annotated also and that is where the other problem is. That ok, first the data itself is limited and now whatever data that is available that itself is not annotated properly.

And when I said it is not annotated properly it means of course, the very first thing that we want to have is we want to have the annotation of what is the emotional state or the psychological assessments of the mental and the behavioral state of the target and as well as for example, the doctor for example, or the customer care employee. And then at the same time we also need to see that all these have to be very very time synced, right.

We need to have the time synced transcripts, we need to have the time synced speech segments in order to train and validate a proper machine learning model. So, this is where it has been fairly limited in terms of quantity. And as I said in terms of the quality also it has been fairly limited with respect to modalities and also with respect to the scenarios.

So, what do you mean by with respect to modalities? Most of the time only the audio data has been made available and is available to the community, most of the cases to analyze and understand it. And most of the time you know of course, many times it is due to the confidentiality and privacy issues as well.

Because, for example, in the case of the psychological counselling where most of the work has been done so far, the video and the physiological data is not collected actually and is not hence available. But nevertheless, other problem is that if the respect to the scenarios as well, we do not we can see that the there is the response or the impact of the empathetic reactions,

interactions can be in other scenarios as well such as for example, education, customer service and medical care as well.

And basically, in these scenarios you know we can generate more data which is time synced data which is having multiple modalities also and then where the most of the work can be done to make more efficient data driven models to generate the empathy, right.