

MODELING EMOTION AND MOTIVATION BASED AGENT FOR TUTORIAL DIALOGUE SYSTEM

Muhammad Asam^{*,1}, Syed Ali Raza^{*,2}

[#]Cybernetics Intelligence Research Lab, Government College University, Lahore, Pakistan

Corresponding author: asam_mirza@yahoo.com

Abstract- It's a challenging task to develop socially interactive agents, because the paradigm has shifted from behavioral to emotion and motivational based agents. The systems that claim to have the ability of natural language processing have some limitations in them. In this study those limitations are mentioned and a model has been proposed that tries to overcome those deficiencies found in the existing models. By doing so, the proposed model MyTutor will be able to perform tutoring in more human like way, because it will be develop by overcoming the shortcomings present in the existing systems

Keywords: socially interactive agents, Natural Language Processing, machine consciousness

1 .INTRODUCTION

In the area of intelligent tutorial dialogue systems, the creation of automated tutorial dialogue systems has appeared as an important research topic. According to many researchers, if the computer tutors are given the ability to engage in natural language dialogue, then the existing performance gap between human and computer tutors can be decreased, in order to increase student learning. Unfortunately, building a tutorial dialogue system is not at all straightforward. By examining how specific student and tutor dialogue behaviors correlate with learning, current work has begun to try practically determining how to make such tutorial dialogue systems more effective. There are two main differences in researchers which are: how they program their data of dialogue and how that data is statistically analyzed. The evaluation of the generality of specific findings is often different [1].

Most tutorial dialogue systems have shown positive results with students, but most of them presently are text based. Many evidence show that speech-based tutorial dialogue systems could be more useful. It is so because spontaneous self-explanation occurs more frequently in spoken tutoring than in text-based tutoring. And self-explanation done by students in spontaneous way helps improve learning gains during human-human tutoring. An interactive pedagogical agent that works using speech rather than text improves student learning. Whereas the absence or presence (visually) of the agent does not have any effect on performance. It is also claimed that if the emotion of a student are also recognized and responded then the success of computer tutors could be increased. It is shown that the emotional state(s) can be predicted more accurately in the speech-based conversation than in other types of conversations.

Intelligent tutor dialogue systems that are used in public schools have been proven to be more effective than classroom instructions. But a more room for betterment has been suggested by the effectiveness of both expert and novice human tutors. The results shown by the tutoring systems that are dialogue based are favorable and suggest that the tutoring systems that are dialogue based may be more effective than the systems with no dialogue. But most of the current tutoring systems use either keyboard-to-keyboard interaction or keyboard-to-speech interaction. This movement towards human-like use of natural language suggests that tutoring systems that use speech to speech communication could be more useful. The researchers have been facilitated by the

present state of speech technology to make spoken dialogue systems in various domains like travel planning and in-car route navigation. There is another advantage of spoken dialogues which is that spoken dialogue contain a higher proportion of student words to tutor words, which is proved to associate with the learning of the student. Tutorial dialogue systems are needed because it is shown by research that the student who are engaged in one-on-one instructions perform as good as the top two percent of students who get traditional classroom teaching[2].

2. LITERATURE REVIEW

Goertzel et al proposed a cognitive architecture called OpenCogBot which can be used to achieve human like general intelligence. Sophisticated methods for visual and auditory pattern inference like action planning, language comprehension and generation, commonsense uncertain reasoning and concept creation are involved in this system. These methods are combined in OpenCogBot in a very unique way. This combination of methods is done using the principle of cognitive synergy, which results in an overall cognitive architecture. In this architecture, real goals are achieved through multiple learning processes. These processes are associated with memories of multiple types. These memory types cooperatively manipulate a common network-based repository of knowledge.[3]. LITMAN et al report that previous discussion and methods in the area of tutorial dialogue across dialogue corpora variant in different aspects like domain, modality and tutor type. They provide methods for integrating their previous coding and analysis methods. They also show that their previous findings regarding learning and student dialogue behavior generalize across corpora and their methodology provides additional findings. At the end they prove that natural use to automate some of these analyses. The field of conceptual mechanics constitutes their data of spoken tutoring. This data contains the corpus of hundred dialogues. These dialogues are taken from twenty students, who were interacting with system of ITSPoke (Intelligent Tutoring Spoken dialogue systems). ITSPoke represents the speech version of Why2-Atlas system. Why2-Atlas is a text-based intelligent tutoring system. A study was performed in which the comparison of spoken and typed human and computer tutoring was performed. In this study this data was collected, which helped in evaluating the system. The data of text based tutoring was in the area of electronics and electricity. This data contained

sixty dialogues which were taken from thirty students. These students were communicating with one of 3 tutors which were humans[1]. Litman et al report a spoken dialogue system called ITSPOKE. In this system a qualitative physics problem is given to student and then student types a natural language answer of that problem. The student is then engaged in a spoken dialogue to provide feedback and to collect misconceptions. ITSPOKE is a speech-enabled version of the Why2-Atlas. Why2-Atlas is an intelligent tutoring system, which is basically text-based. Just like Wh2-Atlas in ITSPOKE, an answer in natural language is first typed by the student. This answer is given in response to a qualitative physics problem presented by the system to the student. But difference between Why2-Atlas and ITSPOKE occurs in the mode of communication with the student. In Wh2-Atlas, all the communication is text-based, whereas in ITSPOKE, the communication is done why speech, which helps in achieving more complete explanations. The speech input by the student is sent to the Sphinx2 automatic speech recognizer. Sphinx2 gets the student speech after it is converted in the digitized format. Sphinx2 then generates the best hypothesis which is received by Why2-Atlas. Why2-Atlas then performs the semantic, syntactic, discourse and domain level processing. It also handles the finite-state dialogue management. The output of the tutor, which is in textual form, is then sent to the Cepstral system, which converts it into speech from text. Then the output is played via a speaker or a headphone[2].The use of dialogue management and deep natural language techniques in two different domains are supported by some tutorial dialogue systems, which are reported by Dzikowska et al. The natural sounding feedback was produced by these systems. This feedback was produced by using generation techniques which are adapted to student performance and the dialogue history. Interprets tentative answers were used by them which are phrased as questions. Two intelligent tutoring systems are demonstrated by them. These are: BEEDIFF and BEETLE. These systems use techniques from task-oriented dialogue systems, which let them to improve the interaction. The information state update approach is used to build these systems. The information state update approach is used for dialogue management and generic components for deep natural language understanding and generation.[4]

Rose *et al* report the results of a comparative evaluation between a model teaching tutors, the Andes System, with the otherwise equivalent dialogue enhanced Atlas-Andes. Andes is model (MTT) which can help the students to solve qualitative physics problems. The focus of Andes is to help students develop good physics problems solving skills. While Andes has been successful, this is simple evidence to suggest that teaching students to solve physics problems is not all that is required to provide them with a solid ground in physics. Atlas-Andes can teach student about basic physics. It has the ability to express the different topics logic and knowledge,e.g. Newton's Laws. To encourage the significant learning of the students and to build up significant problem solving strategies about conceptual physics, the directed lines are provided in Atlas-Andes. The new version of Atlas-Andes is much better than the standard version of Atlas-Andes having an enhanced dialogue version. In the new version

interactive directed lines of reasoning are merged together. The exposure of conceptual tutoring has enhanced in the current version as interactive directed lines of reasoning as interactive directed lines of reasoning are associated with conceptual helper[5].

Various Aspects of Natural Language processing

There is some confusion in the use of the terms, so must get clear with this before proceeding further. The term "natural language" is used for a more limited set of human languages but not for real languages as they are used in regular dialogue, like we use English to talk in everyday life. According to one author "human languages allow anomalies that natural languages cannot allow." A language such as this could be needed, but a language which is claimed to be natural, limited in this way is not real, but it is actually artificial. So the phrase "natural language" should not be used in this limited sense of an artificial natural language. Also, the term "natural language processing" cannot always be used in the same sense. Rather there is a broad meaning and a narrow meaning to the term natural language processing. If we talk about the broad sense, then following things should be included: signal processing, speech recognition, context reference issues, discourse planning and generation, as well as syntactic and semantic planning and analysis. And if the narrow sense is taken into account, then only the syntactic and semantic analysis should be considered. But even if we neglect this confusion in the use of the term, the term "natural language processing" may or may not be considered same as "natural language understanding". Interpretation and generation are considered in the processing part, whereas understanding of the module is represented by the interpretation section. So if a system is said to be understanding the natural language, it should not be implied that that system can generate natural language too, only natural language can be interpreted by it. Processing represents both interpretation and production. But some authors use the term "natural language understanding" as identical to "natural language processing". And they consider interpretation and generation in this use as well. Another problem in the use of the term "natural language processing" is that it is used sometimes in the capacity of research into the working of the natural languages, and the effort in developing a computational model which carries out this working. In this sense, "natural language processing" indicates a system that can interpret and generate natural languages. There are 2 separate aims of this area of research. One of those aims is to comprehend the working of natural language processing. And the 2nd aim is to make a computer able to do the processing of natural languages.

There is another important question to be answered, which is what amount of ability should be considered as the ability to process natural languages. Because research reveals that all humans being process the natural language(s) at different levels. So this question cannot be answered precisely. But the aim should be to interpret and do conversation with other people in usual and ordinary human way. This is not an easy thing to do. Processing refers to translating into or from a natural language. This includes the interpretation and generation of the language. In order for a computer to perform conversation with other humans, it should not only

Table 1: The table also shows the advantages of the MyTutor system

	ITSPOKE	BEETLE II	Auto Tutor	Atlas-Andes	CIRCSIM-Tutor	MyTutor
Speech based system(Voice Recognition)	Yes	No	Yes	No	No	Yes
Text based system	Yes	Yes	No	Yes	Yes	Yes
Self Explanation	Yes	No	Yes	No	No	Yes
A knowledge based system	Yes	Yes	Yes	Yes	Yes	Yes
Emotion detect ion.	Yes	No	No	No	No	Yes
Auto learning	Yes	No	No	No	No	Yes
Motivation to the Learner	No	Yes	Yes	Yes	Yes	Yes
Adaptability	No	No	Yes	Yes	Yes	Yes
Emotion Generation	No	No	No	No	No	Yes

process sentences of natural language, but it should also process knowledge about the world. Generation of sentences in natural language and also giving response to comments and questions requires some common-sense knowledge. But this helps in having a decent conversation. Such common-sense knowledge is needed to understand the meaning of many sentences of natural language. Let’s consider an example, if we tell a computer “Ali had small goat”. The common sense knowledge may let the computer to think that Ali might have eaten the goat in lunch, rather than he was the owner of a goat. But this same kind of common sense knowledge could show that this meaning is not possible for “Ali had cricket bat”. So this tells us that common sense helps in knowing that goats are eaten by humans not cricket bats.

3. Analysis of Tutorial Dialogue System

These tutorial dialogue systems are analyzed during the study ITSPOKE, BEETLE II, Auto Tutor, Atlas-Andes and CIRCSIM-Tutor. In this chapter the above mentioned tutorial dialogue systems will be discussed according to the following parameters: Speech based system(Voice Recognition), Text based system, Self-Explanation, A knowledge based system, Emotion detection, Auto learning, Motivation to the Learner, Adaptability, Emotion Generation. These are the salient features which distinguish a tutorial dialogue system from other systems

4. MyTutor

In this work an intelligent tutoring system called MyTutor is presented and outlined. The discourse patterns and pedagogical strategies of a typical human tutor are implemented by MyTutor. MyTutor is designed in such a way

that it can help a large number of students of various backgrounds and fields in learning the fundamentals of their particular fields. The most distinguishing factor of MyTutor is that it can teach any subject, because it has the ability to learn new domains of knowledge with the help of its knowledge base. Whereas other intelligent tutoring systems are not highly trained in tutoring techniques and they also have very little experience on the tutoring topic. The following table presents the shortcomings of some of the major intelligent tutoring systems.

As it can be seen from the above table 1 that all of the tutoring systems have some deficiencies in them, i.e. none of them implements and caters for all of the features that are necessary for a tutor to have human-like tutoring abilities. Some of the tutors are only text-based or speech-based i.e. they do not converse with the student in either speech or text, so they do not have a vital human faculty of speech which enables humans to converse better. Some of the tutors do not detect the emotion of the student, so they are unable to understand the nature of the student completely. Similarly, some of the tutors do not have the learning ability i.e. they cannot learn new things and subjects. Also, not all of them motivate the learning if the learner needs to be motivated. It can be seen that adaptability is another major factor which is not implemented by most existing intelligent tutoring systems. So they are unable to adapt to new and different situations. Similarly, not all of the above mentioned tutors

generate emotion i.e. they do not have emotion in them, because of which they are unable to mimic a real human tutor. Therefore, from the above discussion, it can be concluded that none of the existing tutoring systems have all the necessary features to be able to teach like human tutors.

The tutor presented in this work, MyTutor aims to overcome all of the deficiencies found in existing systems. To name some it is both text-based and speech-based which makes it converse like human tutors. Apart from that it also aims to implement features

like emotion detection, auto learning, motivation to the learner, emotion generation, adaptability etc. In short, MyTutor tutoring systems overcomes all the deficiencies found in existing tutoring systems.

Each module is described as follows:

User Input:

The input from the student is handled by this module. As can be seen from the figure above that student can input both in textual and audio format. The system has the ability to handle both types of input.

The MyTutor system is presented and described below:

MyTutor:

Parser:

The output from the student input module is sent to the Parser module. The Parser module is used to extract the relevant

semantic content (the content which is relevant to the domain of the student concerned) present in the student input.

Domain Reasoning & Diagnosis Module:

The Domain Reasoning & Diagnosis Module is used to determine whether the explanation given by the student is correct or not. It accomplishes this task with the help of the knowledge base, which is the container which holds the overall knowledge of the system.

Tutorial Planner:

The Tutorial Planner module is used to select the tutorial strategy that could be used in a particular conversation. Again knowledge base module is the main player in this scenario, i.e. it gives the Tutorial Planner necessary information to adopt a particular tutorial strategy.

Response Generation:

After all the processing is done by the system, then it has to generate the response of the system, so the Response Generation Module produces the overall response of the MyTutor system. The response of the system is then presented to the student with which the system is conversing.

Dialogue Manager Module:

The Dialogue Manager Module manages the overall dialogue of the MyTutor system. In other words, it has

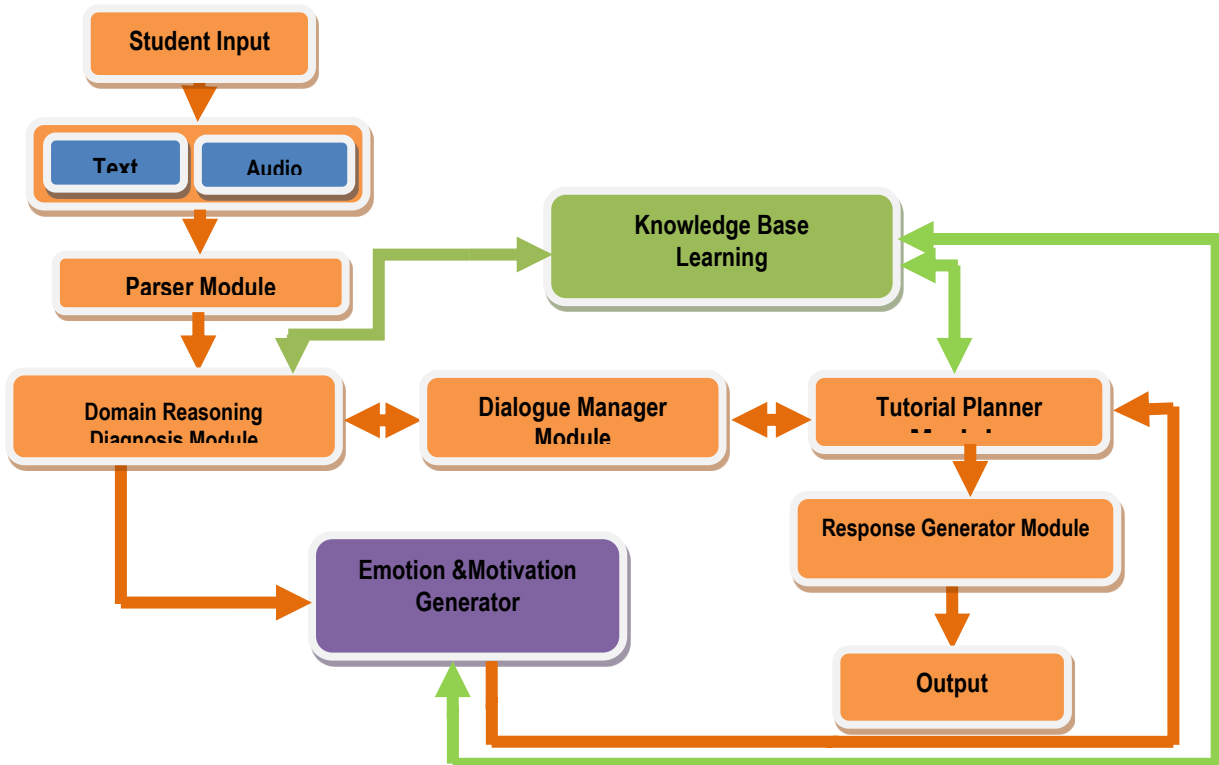


Figure: The MyTutor Architecture

the responsibility to check that whether the dialogue is happening correctly, and there are no errors and miscommunication during the dialogue. If any error or miscommunications occurs, it has the responsibility to check and correct that error or miscommunication.

Knowledge Base:

The most important module is the Knowledge Base of the system. It is the most important part of the system, because it holds the overall knowledge of the system. And it helps other modules to work properly by sharing its information with them. This module also represents the learning ability of the system to learn new domains of knowledge.

Emotion & Motivation Module:

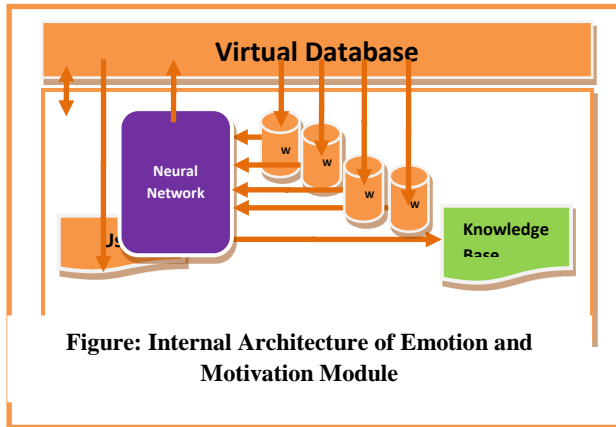


Figure: Internal Architecture of Emotion and Motivation Module

This Module is the emotion and motivation center of the system i.e. it not only detects the emotions and motivation of the user, but also generates the emotional and motivational response of the system. This makes the MyTutor system to work in the situation where the handling of the emotions is necessary or helpful in carrying the conversation.

Output:

The final result of the modules combines in the output module and this module generates overall output of the system. This module basically represents the output interface of the MyTutor system i.e. it is used to show the overall response/output of the system.

Control Flow of MyTutor:

This section of the thesis defines the basic control logic of MyTutor to communicate with the human beings.

Algorithmic steps for the MyTutor are given below.

Receive user input coming from environment.

Transfer user input to Parser module.

Perform initial perception of user

Perform Part of Speech tagging in Parser

Perform Necessary Analysis to find frequent items in the sentence to generate context

Perform Synonyms Identification for semantics understanding

Retrieve Relevant Knowledge from Knowledge base

Perform perception evaluation for identification of interactive user.

Perform matching of user profile on getting evidence of having previous experience.

On non-matching add current profile in Knowledge base for future conversations

Recall previous experience of the agent from Knowledge base to initiate valuable conversation

Generate response of query using according to extracted context using Context Free Grammar to environment. And generate responses using Knowledge base.

Store generated response in the Knowledge base for future conversations.

Goto step 1 to sense next input.

MyTutor Scenario Dialog for Topic Illustration:

This section shows a sample of conversation between student and MyTutor on the topic of concept of Microcontroller in Electronics.

MyTutor can do conversation as under:

S/N	Dialogue	Polarity
1.	STUDENT: Hello.	Positive
2.	MyTutor: Hi, What is your name?	Positive
3.	STUDENT: I am Asam.	Positive
4.	MyTutor: Asam? Are you student?	Positive
5.	STUDENT: Yes, I am student of MS.	Positive
6.	MyTutor: Hi Asam! You are here after a long time?	Positive
7.	STUDENT: oh, I was busy.	Negative
8.	MyTutor: How may I help you?	Positive
9.	STUDENT: I want to discuss some points.	Positive
10.	MyTutor: Do you want to discuss earlier topics?	Positive
11.	STUDENT: No, I have new topic of Electronics.	Negative
12.	MyTutor: Is it your major subject?	Positive
13.	STUDENT: Yes, it is really important.	Positive
14.	STUDENT: I want to ask what is microcontroller.	Positive
15.	MyTutor: "A microcontroller is a small computer on a single integrated circuit containing a processor core, memory, and programmable input/output peripherals".	Positive
16.	STUDENT: Is the microcontroller having the same functions as normal computer.	Positive
17.	MyTutor: Not 100% same, small in size, it has limited functions of normal computer processors.	Negative
18.	MyTutor: Do you remember computer processor? You have already studied it in computer classes.	Positive
19.	STUDENT: Yes I studied but it was boring subject.	Positive
20.	MyTutor: What is boring in it? Computer is an interesting subject.	Positive
21.	STUDENT: Yes, I think so.	Negative
22.	MyTutor: Why do you think so?	Positive
23.	STUDENT: (No Reply from Student)	Negative

24.	MyTutor: Do you want to ask more about microcontrollers?	Positive
25.	STUDENT: Yes, tell me where it can be used?	Positive
26.	MyTutor: It can be used in many toys, cameras and in many home appliances.	Positive
27.	STUDENT: Ok, why the size of articles is so small having the microcontroller as compare to normal computer?	Positive
28.	MyTutor: As the microcontrollers are programmed for single purpose usually. That the reason they small size.	Positive
29.	STUDENT: Ok, thank you for helping me.	Positive
30.	MyTutor: Ok, Bye.	Positive
31.	STUDENT: Bye.	Positive

6. Future Work

In the future versions of this application it would be tried to make this architecture more dynamic and robust to work in dynamic environment. For this purpose modifications would be required at different levels of robotic setup. First of all it is needed to improve the functionality of perceptual module. As now the perceptual module is just detecting very few features out of the sensed information but in future version it would be designed with a perspective of rapidly changing environment so more powerful perception would be required. At the moment another limitation of the architecture is that if robot finds itself in a confused state like it encounters with different students all of which have distinguishing features then it is not easy for the robot to accurately communicate with all of the students. On the basis of this limitation the future versions of this application will be designed with the ability to reassess its teaching strategies and decide if it wants to restructure the strategies or not. This facility will help the architecture to tackle any situation it finds itself.

REFERENCES

- [1] LITMAN, D., MOORE, J., DZIKOVSKA, M. O., & FARROW, E. (2010). Using Natural Language Processing to Analyze Tutorial Dialogue Corpora Across Domains and Modalities. *AIED*, 149-156.
- [2] LITMAN, D. J., & Silliman, S. (2004). ITSPOKE: An Intelligent Tutoring Spoken Dialogue System. *Proceeding of HLT-NAACL*, (pp. 5-8).
- [3] Goertzel, B., Garis, H. d., Pennachin, C., Geisweiller, N., Araujo, S., Pill, J., et al. (2010). OpenCogBot: Achieving Generally Intelligent Virtual Agent Control And Humanoid Robotics Via Cognitive Synergy. *Proceeding of ICAE*. Beijing.
- [4] O.Dzikovska, M., Callaway, C. B., Farrow, E., Marques-Pita, M., Matheson, C., & Moore, J. D. (2007). Adaptive Tutorial Dialogue Systems Using Deep NLP Techniques. *Association for Computational Linguistics*, 5-6.
- [5] Royce, W. W. (1987). Management the Development of Large Software Systems. *87 Proceedings of the 9th international conference on Software Engineering* (p. 10). Los Alamitos, CA, USA: IEEE Computer Society Press.