M-Path: A Conversational System for the Empathic Virtual Agent

Özge Nilay Yalçın[1] and Steve DiPaola[1]

School of Interactive Arts and Technology
Simon Fraser University
Surrey, BC, CANADA
oyalcin@sfu.ca
sdipaola@sfu.ca

Abstract. M-Path is an embodied conversational agent developed to achieve natural interaction using empathic behaviors. This paper is aimed to describe the details of the conversational management system within the M-Path framework that manages dialogue interaction with an emotional awareness. Our conversational system is equipped with a goal-directed narrative structure that adapts to the emotional reactions of the user using empathy mechanisms. We further show the implementation and a preliminary evaluation of our system in a consultation scenario, where our agent uses text-based dialogue interaction to conduct surveys.

Keywords: empathy, conversational agents, affective computing, human-computer interaction

1 Introduction

Conversation forms the basis for many of our social interactions. In recent years, artificial conversational systems have become more ubiquitous and have revolutionized the nature of human-computer interaction. Natural language based assistants are becoming increasingly popular in our daily lives to accomplish goal-driven tasks and act as social companions. Emotions often provide a feedback mechanism during conversational interactions, which makes recognizing and responding to the emotions an important part of social interactions.

Empathy, as the ability to understand and respond to the emotions of others [9], can be used as a guide to interaction. Recent examples of conversational agents in clinical psychology [11] as healthcare assistants [3] and counsellors [7] showed that being emotionally-aware could enhance the interaction by increasing the perceived usefulness, trust, and naturalness of the agent. These findings suggests, showing empathy during conversational exchanges can help to build and strengthen relationships while facilitating natural and believable interaction.

In this work, we aim to use empathic conversation strategies that use emotions as a feedback mechanism that helps inform the dialogue management. We describe the design and implementation of the dialogue framework for an empathic conversational virtual agent, M-Path. Our system is aimed to use system-initiated and user-initiated conversational strategies to guide the goal-oriented conversation while generating appropriate empathic responses. We further show a proof-of-concept implementation of our system in a psychological consultation scenario, where the goal of the agent is to successfully collect the required information and provide appropriate empathic responses. We conduct a preliminary evaluation that shows that our system is capable of providing empathic behavior.

2 M-Path: The empathic conversational agent

Our empathic conversational agent, M-Path, is aimed to create a real-time, goal-driven and empathic interaction with the user. M-Path is capable of initiating and sustaining socio-emotional interactions with the conversation partner by using different levels of empathic behavior. Our embodied agent is designed to produce synchronized verbal and non-verbal behaviors to capture the richness of natural conversational interaction.

The framework for our embodied conversational agent includes a perceptual module that processes the inputs of the system via its sensors, a decision making module, and a behavior generation module that prepares and outputs the synchronized behaviors of the agent. A detailed description of this framework can be found in previous work [14].

This paper focuses on the conversation engine, which is part of the decision making module of M-Path. Within the system, the conversation engine is responsible for initiating and maintaining a meaningful and goal-driven conversation with the interaction partner. This engine works closely with the Empathy Mechanisms module, that makes decisions on the empathic behavior of the agent during the conversation. In the following sections, we will examine in detail how an empathic dialogue can be simulated in a goal-driven environment and provide an implementation scenario to show proof-of-concept implementation of such a system.

3 Empathic Conversation Engine

The central component to achieve an empathic conversational behavior in M-Path is the conversation engine. The empathic conversation engine is designed to achieve the goal of the conversation while adapting to the emotional reactions of the user using empathy mechanisms. It consists of three major components: natural language understanding (NLU), natural language generation (NLG) and the Dialogue Manager (DM).

The Natural Language Understanding (NLU) component handles the extraction of the relevant information from the users' linguistic input and can be as simple as keyword detection to natural language modeling vector representation systems. The NLU component of our system is responsible for parsing, tagging and categorizing the linguistic input for extracting context-related information

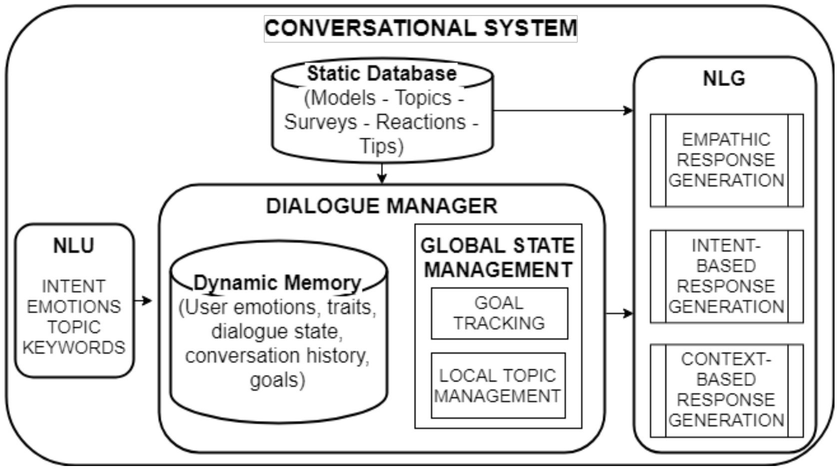

Fig. 1. The outline of the Conversation Engine of M-Path. Three main component of the system are natural language understanding (NLU), natural language generation (NLG) and the Dialogue Manager (DM).

(e.g., topic, intent) and user-related information (e.g., emotions, personality, background information). This information is then sent to the dialogue manager to decide on the response of the agent.

Table 1. Example initiation of a dialogue

SPEAKER	UTTERANCE	RESPONSE TYPE
M-PATH	Hello, I don't think we met before. My name is Alex, what is your name?	Context-based (Global Greeting)
User	It's Anthony	
M-PATH	It is very nice meeting you. Are you a student here at SFU?	Context-based (Local information gathering)
User	nice to meet you too, I am indeed	
M-PATH	That's wonderful.	Intent-based (Approval)

The Natural Language Generation (NLG) component handles the realization of the intended response, depending on the decision made by the dialogue manager. Depending on this decision, the response might be empathic, context-based or intent-based (Table 1). Intent-based responses include acknowledgment replies to the decisions or actions taken by the user. For example, if the user approves to participate in a survey, the agent will confirm that decision by "Okay" or "Thank you". Context-based responses are generated depending on the state of the conversation, both globally (goal) and locally (topic). Global state is used to determine the goal of the conversation. For example, if the user is asking a question, the goal of the conversation will be to answer that question momentarily before moving on with the main goal of the agent. The local state determines the specific topic of the utterance of the user. This is used to give topic-based answers and keep track of the hierarchical structure of the topics. For example, if the user asks "What else?", while being in the topic of "AI", the agent will continue giving examples on the topic AI.

An empathic response of the agent consists of an emotional reply to the emotions of the user, as well as a response related to the coping of the emotions. For example, if the user talks about how bad their day was, the emotional part of the empathic response could be "I am sorry to hear that" or "That is too bad". The coping response could include tips or recommendations that are less emotion-oriented and more context/topic specific. Similar to the context-based responses, these empathic responses can be local reactions to the single utterances of the user, as well as reactions to the global state of the user gathered from multiple turns of dialogue. For example, after receiving multiple negative sentiments, the agent might decide to comment on the mood of the user before saying farewell. An example for both these responses can be seen in Table 2 that shows a real interaction of the system with the user.

Table 2. Empathic reactions during conversation

SPEAKER	UTTERANCE	RESPONSE TYPE
M-PATH	How often do you feel little interest or pleasure in doing things?	Context-based
User	Very often	
M-PATH	(1)I'm sorry to hear that (2) Maybe you should try to build up a routine. You can pick up a new habit with your friends (3) How often have you been feeling down, depressed, irritable, or hopeless lately?	(1) Emotional Reaction (2) Coping Reaction (3) Context-based
User	My problem is i am not used to wake up early	
M-PATH	I understand.	Topic Fallback
User	I feel down most of the time because I do not have a friend circle in campus	
M-PATH	(1) I'm so sorry you feel like that. (2) Try not to withdraw from your social life, your family and friends. They can improve your mood a lot	(1) Emotional Reaction (2) Coping Reaction

These set of example responses are currently hand-crafted and held in a database that stores sample utterances based on context, intent and emotional reactions. Most of the contents of the static database are authored to be representative of how users might interact with the system as well as the responses added after the evaluation processes in order to be more responsive to the actual user utterances. Moreover, the Dialogue Manager (DM) also holds models that are trained based on these example utterance-response pairs. We use two main models in the DM component, where the first is TF-IDF based similarity algorithms that are used to basic Q/A. We used the sklearn library [10] in Python for the TF-IDF vectorizer model, while using NLTK Lemmatizer [8]. The second model is generated by fine-tuning the pre-trained language representation model BERT [4] for the classification of user responses based on sentiment and survey answers. The sequence and properties of each response are decided by using these models within the DM component.

The main role of the Dialogue Manager (DM) is to decide what action the agent should take at each turn in the dialogue. The DM is responsible for generating proper verbal responses according to the utterance of the interaction partner, as well as the empathic reactions. The dialogue manager operates on two different scales: local and global. The local structure handles the immediate responses in isolation while dealing with information gathered from individual dialogue turns. The global structure keeps track of the overarching goal of the system and operates on the entire interaction history using the local information.

Depending on the overarching goal of the dialogue and the local information, the dialogue manager decides on the proper actions to be taken. It keeps track of the context-related information (e.g., topic, intent) and user-related information (e.g., emotions, personality, background information) both locally and globally. This information is gathered from the NLU component and stored in the Dynamic Memory within the DM.

3.1 Global State Management

The global state tracking sub-module keeps track of the overarching goals of the agent in order to decide on which actions to take in a given dialogue turn. Every dialogue is embedded in a generalized structure of greetings/introduction, main topic flow and conclusion/closings. For example, the initiation of every conversation starts with a simple greeting, introduction and basic questions that can be used as a transition before going through the main topic. Similarly, the concluding remarks may include an overall statement about the dialogue and a farewell. The concluding statement can include an empathic response to the overall mood of the user during the conversation. An example excerpt for the introduction and conclusion processes can be seen in Table 1.

Goal Tracking The goal tracking sub-module operates while tracking the current state of the goal or multiple goals, which in turn determines how the main topic flow operates. Goals usually include gathering certain information or going through certain topics or questions within the narrative flow. These usually trigger a system-initiated conversation, where the local topic manager allows for user-initiated conversation to allow a mixed-initiative flow. For example, if the agent has the goal of gathering demographic information, the goal tracking module makes sure the conversation flow always directs back into asking demographics questions until all the information is gathered. After this goal is achieved, the agent can move to other goals or concludes the conversation.

Local Topic Manager The local topic manager determines the flow of conversation between goal-directed conversations. Many topics make for a narrative and can be added and reused in many ways. Topic modules are a container of a narrative topic and its parts and can process through to other topics. The standard topic modules consist of a cascade of Q&A nodes, that taken as a whole create a small conversation or narrative topic. Topics can be on universal issues like "mental health" or specific to the user topics like "anxiety", "depression" and can be reused in many ways (see Table 3).

Table 3. Local Topic Management

SPEAKER	UTTERANCE	RESPONSE TYPE
M-PATH	Would you be okay if I ask some questions about you?	Context-based Goal related Question
User	Are you a bot?	
M-PATH	Yes, are you?	Topic-Based Local QA Answer
User	No	
M-PATH	Okay	Topic Fallback
User	Sure you can ask me questions	
M-PATH	Sounds good, I will ask about generally how you feel.	Continue Context-based

4 Example Implementation

Such an empathic conversational agent might be useful in a variety of applications, where understanding the emotions of the user and responding emphatically is important. In order to evaluate the conversational capabilities of our agent, we picked a "counselling service intake process" as an interaction scenario. For this specific use-case, we had to determine the interaction context and develop content accordingly.

To initiate a counselling service, individuals go through a screening process to gather basic information about them. This usually includes collecting demographic data, as well as a simple assessment in the form of a survey. Based on the department of the intake, these surveys vary in terms of content and questions. For the proof-of-concept scenario, we used an intake scenario for the Psychological Counselling Service (PCS) for the students of the Simon Fraser University (SFU). The screening process for the counselling office includes demographic questions and a short survey to assess the depression and anxiety of students. Demographic questions contain basic identification information such as name, age, gender, and whether or not the user is a student at SFU. The survey that follows the demographic questions is aimed to determine the severity of the issues the student may have. The common method of delivery for this intake process is a self-administered pen-and-paper form where each survey question is answered by selecting an option out of a standard likert-scale.

For the proof-of-concept implementation of our system, we picked this screening process as the main scenario of the interaction. We developed additional content for the purposes of this interaction, including a general classification system for likert-scale surveys and inclusion of the standard questionnaire as dialogue content. The goal of our agent is set to initiate the conversation, collect demographic information, conduct the intake survey and provide proper suggestions according to the survey results.

4.1 Development of the Material

As the material for the screening process, we used the nine items in the California Patient Health Questionnaire (PHQ) to screen for the severity of depression [6]. Items include statements on the symptoms of depression, and patients are asked about the frequency of these symptoms they experienced over the last two weeks. The answers for the questionnaire are picked from the 4-item likert scale scored between 0 to 3, where 0 represents “not at all” and 3 indicating “nearly every day”. The total score for these items indicates the severity of the symptoms, which are scored between 0-27. According to the final score of the questionnaire, there are multiple interventions that can be suggested to the patient.

To integrate this survey in our system, we used each survey question as a topic within the local topic space where the goal is to make sure all these questions are answered. Additionally, we created specific empathic responses that would serve as coping mechanisms depending on the answers of each question as well as the overall score of the whole survey.

The main goals of the agent are to gather information about the demographics of the user and to conduct a survey.Both of these goals are embedded within the global state tracking system that was explained earlier. The global state tracking system holds information about which state/goal in the dialogue the agent is in by constantly updating the state of the dialogue in the dynamic memory. This is to make sure each utterance of the agent is being evaluated according to the current intent.

For the specific implementation of the screening process, our agent can use multiple dialogue strategies in order to make sure that the goal is reached. Each of these strategies are used to make sure the agent successfully directs the conversation to reach its goals. There are two main goals of the system for the counselling scenario: gather demographics and conduct the PHQ survey. The overarching behavior of the agent within this goal-directed scenario is to act empathically towards user responses.

5 Preliminary Evaluation

Although the standard method of submission for these tests are the pen-and-paper survey methods, a direct comparison with this method would not be plausible due to a number of variables that is needed to be controlled. Therefore, we evaluated the empathic screening agent to its non-empathic counterpart in a text-based interaction environment. The main difference between these two agents is their responsiveness to the emotional utterances of the user. Our hypothesis is the empathic version of the conversational agent would be perceived as more empathic, which would in turn have a positive effect on the attitude towards the interaction.

5.1 Method

Participants A total of 16 users (10 Female, 6 Male) completed the study that were between the ages 20 and 39 (M=26.65, SD=7.74). Because we were focusing on the screening process for the student consultation service at Simon Fraser University (SFU), we chose undergraduate and graduate students in SFU. Participation in this study was voluntary and was based on open invitations to a large group of students at SFU via online communication.

Materials We used an empathic and a non-empathic version of the same conversational system to be able to evaluate the empathic properties of the system. Both versions had the same goal of gathering demographic data as well as finishing the survey for the screening process. The empathic version, as described in earlier sections, was providing emotional and coping responses based on the utterances of the user. Additionally, an empathic response was given at the end of the survey based on the overall score of the survey. On the other hand, the non-empathic version was only giving acknowledgment responses to the user utterances, and a generic closing statement after the survey is concluded. This was done to make sure only the quality of the responses are different, while the quantity of the system responses are the same.

Each user evaluated the conversational agents based on their perceived empathy. The perceived empathy of the agent is evaluated by using a modified version of the Toronto empathy questionnaire [13], which is a 16-item survey that originally is used as a self-report measure. Each item on the questionnaire are scored in a 5-item likert scale (Never = 0; Rarely = 1; Sometimes = 2; Often = 3; Always = 4), where half of the items are worded negatively. Scores are summed to derive total for the perceived empathy and can be varied between -32 to +32.

In addition to the perceived empathy measures, we also evaluated the user's attitude towards interaction while focusing on the use-case as an alternative screening process. We used items from technology acceptance [5] and Godspeed [1] questionnaires, which includes statements about usability, believability, and human-likeness of the agent. We included items that focus on the preference towards the screening process and compares the agent-based interaction to the classic paper-based method ("I prefer the interaction to a paper-based survey") and human-initiated method ("I prefer the interaction to a survey conducted by a human"). We also included three items to understand the trust felt towards the agent, which was used in similar studies [7]. We used a total of three statements to evaluate the trust felt towards the agent, based on "trusting the advice agent gave", "feeling better interaction privacy" and "trusting to disclose information". The total score for trust was derived from averaging these values. A 5-point Likert scale that shows agreement with the statements with items between "Strongly Disagree" to "Strongly Agree". The high scores mean more agreement with the statements where the low scores show disagreement, where the lowest score is 0 and the highest is 4 per item.

We implemented the user interface of the dialogue agents in the Slack messaging environment, where each user was using a chat channel in order to interact with the agents by using text. For the display names for the agents, we used gender-neutral names: Alex and Joe. These names were counterbalanced between the conditions as well as the interaction order and the types of surveys. This ensured there was an equal amount of participants interacting with each possible combination of agent type, order, survey type and agent names.

Procedure We used within-subject methods, where each user is interacting with both the empathic and non-empathic versions of the conversational agent. Participants used the Slack messaging environment in standard computers in order to interact with the agents using text messaging. Participants were briefed about the context of the interaction and the procedure before the experiment. Each interaction started with an informed consent procedure.

According to the counterbalancing, each participant first interacted with one of the conversational agent (empathic or non-empathic) and took the evaluation survey about the agent after the interaction is done. After that survey, the participant went through the same process with the other conversational agent and took the evaluation survey on the second interaction. Participants had to greet the conversational agent to be able to start the conversation. Participants were assigned across conditions, while being counterbalanced in terms of the order of conditions as well as the type of survey each condition is conducting. Each subject took about 30 minutes to complete the experiment.

5.2 Results

From 16 users, only one encountered an unsuccessful interaction for both of the agents, where the goal of conducting the survey was not reached. None of the user responses were excluded from the final analysis of the results. All analysis and plotting are done using linear mixed models on R [12] with lme4 [2] package.

We performed a linear mixed effects analysis of the relationship between the perception of empathy and system type (empathic vs. non-empathic). As fixed effects, we entered the subjects into the model. Results show that perceived empathy is significantly higher in the empathic agent, relative to the non-empathic agent condition ($p = .02$).

We also examined the attitude towards the interaction. Results showed that the system type condition (empathic vs. non-empathic) significantly effects the perceived usefulness of the agent ($p = .05$). The empathic agent is found more human-like ($p < .01$) and preferred more to a human agent ($p < .01$), than the non-empathic agent. The preference of the agent over the pen-and-paper based screening process was not significantly different ($p = .2$), but high in both cases. Moreover, the results showed the system type does not have an effect on trust towards the system ($p = .41$). Table 4 shows details for the results.

Table 4. Results of the Evaluation

Variable	Empathic agent M	Empathic agent SD	Non-empathic agent M	Non-empathic agent SD	F(1,15)	p
Empathy	3.38	8.18	-1.12	7.80	6.43	.02*
Usefulness	3.06	1.00	2.56	0.96	4.29	.05*
Human-like	2.56	0.63	1.81	0.98	10.38	<.01**
Believable	2.88	1.02	2.38	0.96	5	.04*
Prefered to human	2.06	1.24	1.69	1.40	8.99	<.01**
Prefered to paper	2.88	1.36	2.62	0.96	1.36	.26
Trust	1.81	1.02	1.64	0.95	0.71	.41

6 Discussion

Results showed that the empathic dialogue capabilities that we introduced for the conversational agent resulted in an increase in the perception of empathy during the interaction in the screening process. The empathic capabilities also increase the believability and human-likeness of the conversational agent, as well as its perceived usefulness. We also see that users prefer the empathic agent more than the non-empathic counterpart in terms of its use in respect to a screening process with a human. However, we see that users would still prefer talking to a human, rather than interacting with the agent. We also saw that, counter to previous studies on empathic agents, that the empathic capabilities did not increase the perception of trust.

Further examination of the scripts created from the interaction data revealed that the interactions with the agents were not homogeneous in terms of the emotions that the participants were showing. We observed that when the participants showed more negative emotions and scored lower in the surveys, they rated the behavior of the empathic agent more positively. However, we did not control for this behavior and this phenomenon needs to be examined further.

7 Conclusion and Future Work

In this work, we proposed and implemented a dialogue system to equip empathic behaviors in a conversational agent. We evaluated the empathic capabilities of the agent in a proof of concept use-case, the screening process in a consultation scenario. We compared the conversational agent with and without the empathic behaviors to be able to capture the effect of our system. The results suggest that the inclusion of emotional and coping responses as empathic behavior in a conversational agent leads to increase in the perception of empathy, usefulness as well as human-likeness and believability of the agent. Even though the implementation only included the PHQ survey, any type of survey can be used with a minimal amount of development process. However, in sensitive circumstances, such as the depression screening process, trusting the agent seems to pose a challenge that needs to be addressed. This system is intended to be used in an embodied conversational agent, in real-time multi-modal interaction. For future work, we intend to integrate this system into our embodied agent framework and further compare the perception of the agent during face-to-face interaction.

Acknowledgements This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2019-06767].

References

1. Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.
2. Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
3. David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, pages 1061–1068, 2014.
4. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2018.
5. Marcel Heerink, Ben Krose, Vanessa Evers, and Bob Wielinga. Measuring acceptance of an assistive social robot: a suggested toolkit. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 528–533. IEEE, 2009.
6. Kurt Kroenke and Robert L Spitzer. The phq-9: a new depression diagnostic and severity measure. *Psychiatric annals*, 32(9):509–515, 2002.
7. Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS)*, 4(4):19, 2013.
8. Edward Loper and Steven Bird. Nltk: The natural language toolkit. In *In Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. Philadelphia: Association for Computational Linguistics*, 2002.
9. Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3):11, 2017.
10. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
11. Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. Embodied conversational agents in clinical psychology: a scoping review. *Journal of medical Internet research*, 19(5):e151, 2017.
12. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.
13. R Nathan Spreng, Margaret C McKinnon, Raymond A Mar, and Brian Levine. The toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of personality assessment*, 91(1):62–71, 2009.
14. Özge Nilay Yalçın. Empathy framework for embodied conversational agents. *Cognitive Systems Research Journal*, in press.