ARTICLE IN PRESS

Biologically Inspired Cognitive Architectures xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Biologically Inspired Cognitive Architectures

journal homepage: www.elsevier.com/locate/bica

Research article

A computational model of empathy for interactive agents

Özge Nilay Yalcin*, Steve DiPaola

Simon Fraser University, 250 -13450 102 Ave. Surrey, BC, Canada

ARTICLE INFO

Keywords:
Empathy
Affective computing
Conversational agents

ABSTRACT

Empathy has been defined in the scientific literature as the capacity to relate another's emotional state and assigned to a broad spectrum of cognitive and behavioral abilities. Advances in neuroscience, psychology and ethology made it possible to refine the defined functions of empathy to reach a working definition and a model of empathy. Recently, cognitive science and artificial intelligence communities made attempts to model empathy in artificial agents, which can provide means to test these models and hypotheses. A computational model of empathy not only would help to advance the technological artifacts to be more socially compatible, but also understand the empathy mechanisms, test theories, and address the ethics and morality problems the Artificial Intelligence (AI) community is facing today. In this paper, we will review the empathy research from various fields, gather the requirements for empathic capacity and construct a model of empathy that is suitable for interactive conversational agents.

Introduction

It has been argued that emotions have social and cognitive functions that is mandatory for an intelligent system (Minsky, 1991; Damasio, 2006). Empathy, as the capacity to relate another's emotional state (De Waal & Preston, 2017), is found to evoke altruistic and prosocial behavior, have a positive effect on the length of relationship, provide competence in communication and have a negative effect on aggressive behavior (Omdahl, 1995). These findings suggest that the study of empathy in artificial agents and especially in interactive systems can benefit from the development of a computational model of empathy in being able to evoke empathic emotions in the interaction partner and being able to act empathically towards the interaction partner Table 1.

Studies conducted so far demonstrated that agents with the ability of showing empathy, a complex socio-emotional behavior, lead to more trust (Leite, Castellano, Pereira, Martinho, & Paiva, 2014; Brave, Nass, & Hutchinson, 2005), have positive effect on the length of interaction (Bickmore & Picard, 2005), help reduce stress and frustration (Prendinger & Ishizuka, 2005; Burleson & Picard, 2004) and increase engagement (Coplan, 2011). Such a capability for computational systems would enhance the social interaction in educational applications, training environments, artificial companions, medical assistants and gaming applications, where initiating and maintaining a social interaction is of great importance.

As studies from neuroscience, ethology and psychology suggest, empathy is the result of complex interaction of lower and higher level cognitive processes (De Waal & Preston, 2017). Empathic capacity can be seen in a broad spectrum of behavior, which can be categorized as behaviors related to communication capacity, emotion regulation and cognitive mechanisms. A computational model of empathy must reflect the theoretical background and empirical findings. Existing models of empathy in artificial agents often fail to present a complete picture of empathic behavior or prefer to approach the problem by modeling only some components of empathic capacity, due to the complexity of the problem. However, the research outcomes so far show that it is worth studying empathy further to advance our understanding of empathy mechanisms, and to provide better interaction techniques into human-computer interaction research. In this paper, we propose a model of empathy which is suitable for conversational agents and provide an overview of techniques that can be used to implement such an agent.

Theoretical background on empathy

Since its philosophical foundations, empathy has been a research topic in several research fields such as clinical psychology, social psychology, aesthetics, ethics and neuroscience (see Coplan, 2011 for an extensive overview of empathy research). The behaviors assigned to empathy gathered in these various fields are mirroring, affective matching, sympathetic concern, consolidation, theory of mind, perspective taking and projection of self (De Waal & Preston, 2017; Coplan, 2011; Batson, 2009; Leiberg & Anders, 2006). These behaviors are usually categorized as affective or low-level empathy and cognitive or

* Corresponding author.
E-mail addresses: oyalcin@sfu.ca (Ö.N. Yalcin), sdipaola@sfu.ca (S. DiPaola).

https://doi.org/10.1016/j.bica.2018.07.010
Received 17 June 2018; Accepted 6 July 2018

Table 1
A summary of empathy components, the mechanisms involved and the empathic behavior they are linked with.

Empathy Component	Mechanisms	Empathic Behavior
Communication Competence	Emotion Recognition; Emotion Expression; Emotion Representation	Mirroring; Affective Matching
Emotion Regulation	Features of the observer (mood, arousal, personality, age, gender); Features of the relationship (liking, relationship, social status)	Empathic Concern; Consolation
Cognitive Mechanisms	Emotional Appraisal; Theory of Mind; Simulation Theory	Altruistic helping; Perspective-Taking

high-level empathy.

Affective empathy is defined as the automatic and often unconscious mimicking of other's emotional response and includes mirroring and affective matching behaviors. Cognitive and high-level empathy is considered as the ability to understand the other's emotional and mental states by using cognitive mechanisms such as perspective taking and theory of mind (ToM). There is disagreement on how these behaviors are related, and whether empathy is a discrete or continuous phenomenon (Coplan, 2011). However, research efforts suggest the levels of empathy are interconnected (De Waal & Preston, 2017; Coplan, 2011). De Waal and Preston (2017) suggests the components of empathy are built on top of each other due to evolutionary mechanisms (see Fig. 1). According to this view, perceiving the behavior of another in one's own representation is automatic in and based on the perception–action mechanism (PAM). This mechanism is the essential for affective and cognitive empathic behavior.

Different theoretical approaches to model empathy agree on a set of cognitive and behavioral capabilities required for empathic behavior, which we will call components. Understanding these main components and the underlying processes is crucial for simulating empathy in AI based interactive agents, regardless of the chosen theoretical model. These components can be categorized as communication competence, emotion regulation and cognitive mechanisms. Similar components have been suggested by Paiva and colleagues (Paiva, Leite, Boukricha, & Wachsmuth, 2017) and Boukricha (Boukricha, Wachsmuth, Carminati, & Knoeferle, 2013) as empathic responses, empathy modulation and empathy mechanism. Their categorization of empathic capacities does not include the perception of emotions as an essential communication capability, but rather as one of the mechanisms of empathy.

Fig. 1. Russian Doll Model of Empathy, adopted from De Waal and Preston (2017).

Another categorization by Davis's (Davis, 1994) refers to antecedents, processes and empathic outcomes that resemble this approach. Antecedents, refer to the biological capacities and individual differences that affect the strength of empathic response, similar to emotion regulation and modulation parameters. Processes refer to the type of cognitive or noncognitive processes that result in the empathic response, such as mimicry, association and role taking. Empathic outcomes can be intrapersonal and interpersonal outcomes that shows the empathic behavior presented. According to this view, intrapersonal outcomes include parallel and reactive outcomes as well as non-affective judgements. Interpersonal outcomes are helping behavior, aggression or social behavior.

A proposed model of empathy

Following the theoretical foundations and research on empathy from various fields, we can conclude that the empathic capacity should contain a broad spectrum of behaviors that we can categorize as communicative ability or emotion, regulation of emotions and the cognitive mechanisms related on the appraisal of the emotional situation. A model of empathy for interactive agents should include these capabilities that are linked to each other to express different levels of empathic behavior. Our proposed model is therefore composed of three-level structure, as shown in Fig. 2, is aimed to capture the link between empathic behavior and biological capabilities that are needed to satisfy them.

The hierarchical structure of the components in this model resonates with the layered representation of the Russian doll model of empathy (de Waal, 2010), where each layer is built on top of the mechanisms of the previous layer. In the following sections we will provide a description of each component.

Communication competence

The definitions and models of empathy suggest an underlying mechanism required for all levels and intensities of empathic behavior, where without the successful recognition and expression of an emotion it is impossible to show any type of empathic response towards it. The concept of emotional intelligence (Salovey & Mayer, 1990) in psychology is addressing this distinct capacity of perceiving, accessing and generating emotions to be used in assisting thought, reasoning and understanding. This includes identifying and expressing emotions in other people as well as art pieces. Similarly, Scherer (Scherer, 2010) focuses on Emotional Production and Recognition Competence while referring to distinct skills that are underlying factors for any type of emotional behavior. Preston and de Waal (Preston & De Waal, 2002) points out that neurological disorders such as autism, sociopathy and prefrontal damage suggests that the lack of empathy is linked to the disfunction of recognition and expression of emotions. De Vignemont & Singer (De Vignemont & Singer, 2006) mentions these the communication capacities and emotional repertoire of the empathizer effects the intensity of the empathic response. Therefore, it is mandatory for any computational approach of empathy to include the emotion recognition and emotion expression capabilities, which we will call communication competence. The lowest level of the hierarchy of the empathy model is the affective communication competence, which

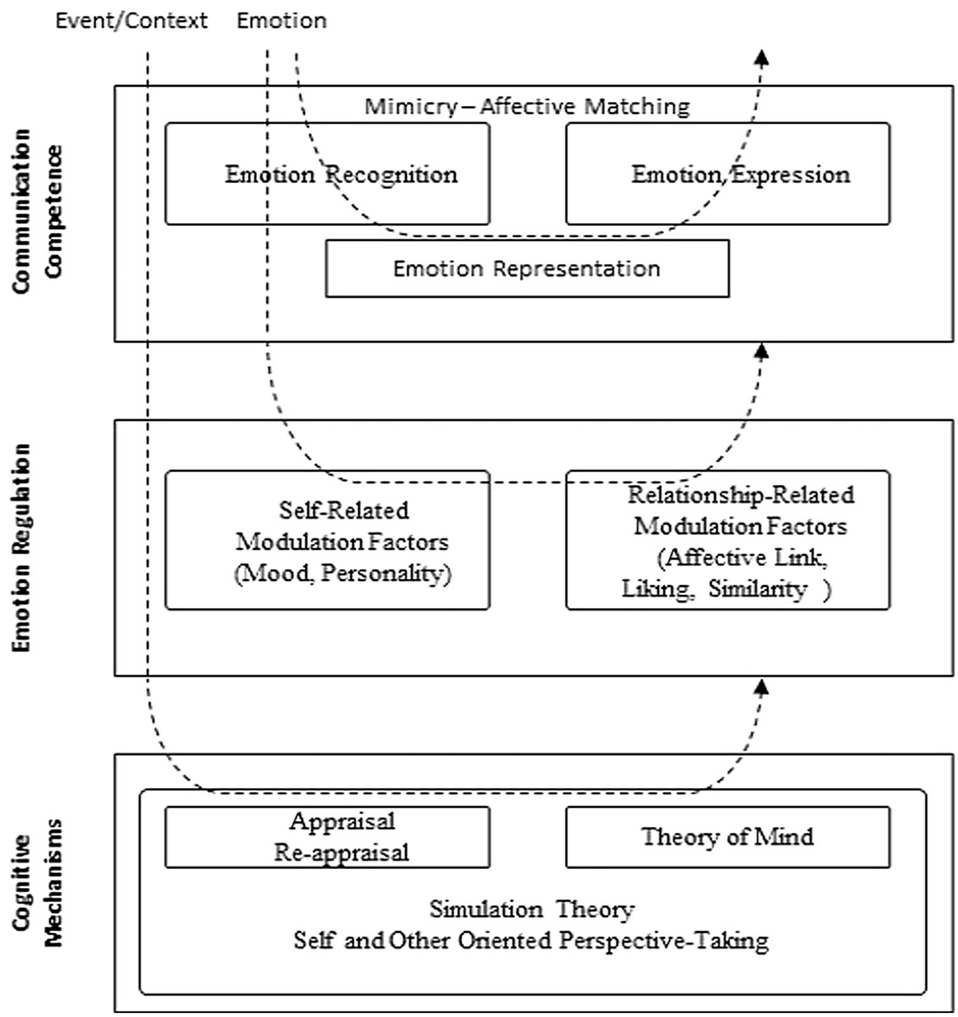

Fig. 2. The proposed model of empathy with hierarchical components.

includes emotion recognition, emotion expression and emotion representation capabilities.

Low-level or affective empathic behavior, which includes motor mimicry and affective matching arises from the affective communication competence. Mirroring behavior such as facial and motor expression, yawn contagion and emotional contagion are the low-level empathic behavior are the products of a basic perception-action mechanism that allows humans to resonate the observed actions with their own representations and actions. According to the Russian doll model of empathy by de Waal (De Waal, 2007), the motor mimicry and emotional contagion are at the foundation of any empathic behavior and are based on perception-action mechanism (PAM). PAM suggest humans can understand emotions due to the mapping of other's emotional state onto their own representations of those emotional states. Therefore, the affective communication competence requires the recognition of emotions, expression of emotions and the internal representation of those emotions within the agent.

Emotion regulation

Another main component of empathy is the regulation of the empathic emotions that is related to a range of psychological, social and cultural factors. De Waal and Preston mention that perception-action mechanisms (PAM) is not limited to perform in the low-level mimicry and emotional contagion stages that only requires automatic recognition and expression mechanisms. They argue that more complex forms of empathy require the interaction of PAM with regulatory processes such as attention, experience, as well as similarity, relatedness, familiarity and social closeness (De Waal & Preston, 2017). Scherer (Scherer, 2010) defines Emotion Regulation Competence as the capacity to modify the 'raw' emotion by a set of automatic and conscious monitoring mechanisms.

Davis (Davis, 1994) mentions antecedents as biological and individual differences that result in modulating the strength of empathic response. This model was used by de Vignemont and Singer (De Vignemont & Singer, 2006), who illustrate these modulatory functions as features of the emotions (valence, intensity, saliency, primary vs. secondary emotions), social link between target and empathizer, context, and the characteristics of the empathizer (mood, personality, gender, age). Hoffman (Hoffman, 2000) further adds familiarity bias that includes in-group, friendship and similarity biases. Coplan (Coplan, 2011), states the inhibition of these mechanisms is required for other-oriented perspective taking.

Paiva and colleagues provide a categorization of these factors (referred as empathy modulation factors) in empathy research: features of the observed emotion, relationship between the observer and the target, situation/context and features of the observer. They mention the shared concepts between the emotion research and empathy research regarding these regulatory factors and state how the OCC model of emotions introduce a set of variables that influence the intensity of the emotions (Paiva et al., 2017).

The regulation mechanisms include the self-related and other related emotion regulation factors. Self-related factors include characteristics of the empathizer that are independent on the observed emotion or situation such as mood and personality. These factors alter the expressed emotion in different timescales as well as effecting the representation of emotions, the attention mechanisms towards the observation and can have an impact on how the observed situation will be interpreted in the cognitive mechanisms. Relationship-related regulation factors are factors that are dependent on the link between the observer and the target. Liking, similarity, social link, and familiarity biases are some of these factors that effect the expression of the emotional behavior. Context is another important parameter that may affect the expressed empathic emotion.

Cognitive mechanisms

High-level or cognitive empathic behavior requires a set of cognitive capabilities such as appraisal and re-appraisal of the situation, theory of mind and mental simulation. These capabilities form the top level in the hierarchy of the empathy model and depends on the lower levels in the model. Cognitive mechanisms can be affected by the emotion regulation factors and mirroring mechanisms. Therefore, higher level empathy mechanisms such as perspective taking, and targeted helping require the inhibition of the effects of the lower level components.

In cognitive theories of emotion, appraisals are the evaluations of events happening in the environment of the agent based on the impact of that situation (Lazarus, 1991). Appraisals can be made in accordance to user's cognitive functions such as intentions, plans and goals as well as emotions and social values. The cognitive approaches to emotions (Arnold, 1960) state that, for an event to invoke emotional response, it should affect a person based on the goals of a person. The OCC (Ortony, Clore, & Collins, 1988) theory of emotion belongs to this category of emotional models, which is frequently used in virtual agent research. The agency component of this theory is very suitable for simulating human cognitive processes, which is a key for higher level/cognitive empathy. However, the appraisal mechanisms mentioned by the emotion research that ties the goals and values of the organism to cognitive function cannot be enough for high-level empathy where perspective taking is required (Coplan, 2011). Higher cognitive processes such as theory of mind, perspective taking, and projection of self is required for higher level empathy functions (Goldman, 2012).

Omdahl (Omdahl, 1995) demonstrates how cognitive appraisal of an affective situation allows people to arrive at assessments about the situation not only on self-directed perspective taking mechanisms but also on other-directed perspective taking mechanisms by re-appraisal of the situation. Following the work on appraisal theory (Scherer, 1982) and semantic network-theory (Bower, 1981), Omdahl (Omdahl, 1995) demonstrated that reasoning about other's emotions and perspective taking are related to appraisal and re-appraisal of the situation. This notion of achieving perspective taking by reappraisal is also proposed in de Vignemont & Singer (Preston & De Waal, 2002).

Empathy in conversational agents

In her extensive study on the link between cognitive appraisal theories and empathy, Omdahl (Omdahl, 1995) argues about the special need to examine verbal information in empathy research as it is the primary source of understanding the context of affective information. This idea suggests that conversational agents can be beneficial in empathy research, as they can be used to provide additional context to affective and social signals. However, as empathy refers to a range of phenomena with complex interactions, the empathy models and definitions used in conversational agent research do not always match the theoretical background.

Conversational interaction has a central role in human-to-human communication, where information from multiple sensory channels are received and sent simultaneously. Embodied conversational agents (ECAs) are agents that can interact with users with a multimodal, situated (and often anthropomorphic), and real-time interaction to emulate a similar experience of human-to-human conversational interaction (Cassell, Bickmore, Campbell, Vilhjálmsson, & Yan, 2000). There is a tendency in ECA research to refer to sympathetic (D'mello and Graesser, 2012; Looije, Neerincx, & Cnossen, 2010), emotionally supportive (Brave et al., 2005) or compassion behavior (Looije et al., 2010) as empathy, which only reflects narrow fragments of empathic behavior. Others focus on the low-level empathic functions such as mirroring (Smith, 2011) and affective matching (Lisetti, Amini, Yasavur, & Rishe, 2013) in conversational agents, that may provide a basis for higher-level phenomena. Some researchers attempt to model higher-level empathic behavior, such as reactive empathy in Davis (Davis, 1994), however do not provide any emotion-regulation or appraisal mechanisms to justify agent actions (Moridis & Economides, 2012). Even though these approaches do not reflect the full array of empathic reaction, they provide a good foundation for building up a more complex empathy model.

More complete models of empathy have also been studied in ECA research. The EMMA framework (Boukricha et al., 2013) uses a continuous representation of empathy that includes modulation mechanisms that uses mood, liking and familiarity factors to adjust the expressed empathic emotion, where empathy is defined as affective matching. Their theoretical model includes higher level empathy mechanisms however, was not implemented in their agent framework. Another example is the CARE framework (McQuiggan, Robison, Phillips, & Lester, 2008), the agent learns empathy by observing the interactions between humans in a treasure hunt game called Crystal Island. Their system can track intentions, age, gender, affective state and goal of the users. The system learns to show parallel or reactive empathy depending on the context and attributes of the target.

Furthermore, some research efforts are tackling similar problems and generate matching models of emotion without mentioning empathy. GRETA's (Clavel & Callejas, 2016) motor resonance module can be regarded as a low-level empathy component that allows for mimicry where the agent-mind takes the goals, beliefs and social relationships into account to have a long-term decision, which can be regarded as a higher-level regulation mechanism. A more complex form of empathic behavior can be found in the user adaptation applications in dialogue systems (Jokinen, 2003), where the aim is to model the user intentions and state to predict and adapt to the user behavior (Jokinen & McTear, 2009). The TRAINS-92 system (Traum, 1993) was aimed to predict user intentions to adapt the dialogue state accordingly, which can be seen as a foundation for high-level empathy.

The general challenges in implementing empathy models in ECA research are reflecting the multi-level nature of empathic behavior, integrating the context, and automatization of the behaviors (Paiva et al., 2017). However, the growing number of research efforts in empathic agent research and technological advances suggest these challenges might be overcome in near future. The fields of affective computing and social computing as well as dialogue management research provide valuable insights on how to implement empathy based on emotional communication, regulation and cognitive capacities.

Affective computing

It was argued that emotional intelligence, a set of social and emotional skills to recognize, regulate and express emotions, is crucial for a successful communication, coping with the social life and empathy (Bar-On, Tranel, Denburg, & Bechara, 2004). The ability to detect, understand and exhibit affective signals is therefore said to be indispensable for successful interaction (Goleman, 2006). These capabilities lay the foundation for any level of empathic behavior and is crucial for understanding others. The research area for building systems that can recognize and express emotions is called affective computing (Picard, 1997).

Affective computing research helped to develop effective techniques for emotion recognition and expression in artificial conversational agents, as well as testing and creating models of emotion (Gratch & Marsella, 2004) and simulating computational agents for human-computer interaction (HCI) applications Pantic, Sebe, Cohn, & Huang, 2005; Jaimes & Sebe, 2007. The body of work has proved the importance of adapting to the user's moods, preferences, attention and intentions. Interaction systems that can recognize and react to the affective states of the users are more likely to be perceived as natural, trustworthy and efficient (Pantic et al., 2005).

Affect communication with ECAs is crucial to determine user's emotions and attitudes to have an adaptive interaction; and being able to express them properly to have more meaningful and natural interaction (Bates, 1994). It was found that detection of sentimental cues during dialogue is valuable to achieve long-term interaction between humans and ECAs (Pecune, Ochs, & C. Pelachaud., 2013). Emotions can also be viewed as a control mechanism for the range of actions that the agent can choose from Wilks, Catizone, Worgan, and Turunen (2010).

Researchers use two main approaches to represent emotions: categorical emotion recognition (such as fear, anger, happiness) (Ekman, 1992) or dimensional approach (valence and arousal space) (Osgood, May, & Miron, 1975). Valence represents the pleasantness of a stimuli between positive and negative scale, where arousal shows the intensity of the emotion. Another approach on appraisal theories suggests that emotions are triggered due to the assessment of the events based on the goals and abilities of the agent (Scherer, Bänziger, & Roesch, 2010). The OCC model of emotions (Ortony et al., 1988), Scherer's appraisal model (Scherer et al., 2010) or EMA (Marsella & Gratch, 2009) are the most used appraisal models in ECA research.

Clavel and colleagues mention (Clavel, Pelachaud, & Ochs, 2013) the choice of computational and theoretical models should be related to the capabilities and goals of the embodied conversational agent. Dimensional analysis has been widely used in detecting negative-positive emotions and model agent behavior (Smith, 2011). Discrete approaches are used to detect certain emotional categories such as user frustration or attention (D'mello and Graesser., 2012) that can be beneficial to track application-specific emotions. Appraisal based models can be used to guide the dialogue, model agent behavior (El-Nasr, Yen, & Ioerger, 2000) and to model the user's affective (Jaques & Vicari, 2007).

These techniques used in emotion recognition, representation and expression can be easily used as a foundation for an empathic conversational agent and can be used to generate low-level empathic behavior and inform higher-level components of empathy. Moreover, cognitive models of emotion such as appraisal models can be used to reason about the experienced situation, react to the situation based on its appraisal and the reappraisal of the situation are fundamental for cognitive mechanisms of empathy.

Social computing

A key feature of the social communication is the detection of social signals that helps us to predict the behavior of the interactive partner and act accordingly. Social signals are defined as "a communicative or informative signal that, either directly or indirectly, provide information about social facts, namely social interactions, social emotions, social attitudes, or social relations" (55, p.4). A successful social interaction requires the detection of the correct non-verbal social signals from machine sensors, reasoning about these signals in a dynamic way and expressing social signals based on that reasoning (Mairesse & Walker, 2010). The empathic behavior can be regulated based on self-related and relationship-related social signals. In social computing research, the personality of an agent can influence the expression of behaviors and can be used to capture the personality of the target to compute the similarity between the observer and the target.

It is possible to capture the engagement, stress and interest levels of the users, which is useful for understanding user preferences and providing feedback to the system about its performance (Pentland, 2005). Moreover, recognition of the patterns in social signals can be used for modeling personality traits. The most adopted personality model in social computing research is the Big 5 personality model (Costa & MacCrae, 1992) that represents personality in openness, consciousness, extraversion, agreeableness and neuroticism dimensions. However, it is a challenge to integrate the effect of context, beliefs, goals and intentions of the agent with these social signals (Wang, Carley, Zeng, & Mao, 2007). Automated methods have been proposed for personality recognition from conversational data. A recent approach on text-based conversational models uses Speaker-Addressee model using LSTM network approach to generate consistent responses that matches personalities which is represented as a vector embedding (Li et al., 2016a). This vector representation holds the information about dialect, age, gender and other personal information that will influence the content and style.

Moreover, as a feature of cognitive mechanisms, simulation of the target's beliefs, desires and intentions within the observer is required for perspective-taking. User modeling techniques used in social computing research can provide a basis for internalizing the other's mental states within the agent. Wahlster and Kobsa (Wahlster & Kobsa, 1989) define the user model as a knowledge base that contains assumptions about the user that might be relevant to the behavior of the agent, which are separate from the rest of the agent's knowledge. Such a user model dynamically that consists of the general knowledge, beliefs and the goals of the user can allow for higher-level empathic reasoning and perspective taking. An adaptive capability for empathic conversational agent should address the issues of how to recognize user traits, deciding on which of those traits to be reasoned with and expressing consistent responses based on that decision.

Context in conversational systems

Context can provide valuable information about the cause and effect relationship between the event and the emotional effect it triggers in the target. Verbal channel can offer important cues for decoding of the situation and empathy. Linguistic context has been found to be vital in dialogue management (Pickering & Garrod, 2004), recognizing emotions (Calvo & D'Mello, 2010; Picard, 2014; Cambria, 2016), understanding social behavior (Vinciarelli et al., 2012) and for multi-modal interaction (Pantic et al., 2005; Jaimes & Sebe, 2007). Same emotional gestures might signify different meanings based on their context (Ekman, 2004). Understanding and representing context of the verbal information is therefore crucial for an empathic ECA.

In dialogue systems, the context can be extracted by using methods as simple as a keyword detection mechanism or more advanced techniques as natural language modeling and vector representation. The recent advances in deep learning techniques made it possible to create a distributed representation of a given text (Bengio, Ducharme, Vincent, & Jauvin, 2003; Mikolov, Chen, Corrado, & Dean, 2013) which can be used to understand the context of an utterance.

The traditional conversational systems follow a modular approach (Jokinen, 2009; Traum, 2017) which consists of Natural Language Understanding (NLU) module that processes the input from the user, the Dialogue Manager (DM) that updates the internal state to select a proper response given that input and Natural Language Generation (NLG) module that generates and outputs a proper utterance according to that internal state. Recently, advances in Deep neural networks (DNNs) has been very successful in many applications of NLP, speech recognition and response generation. However, these approaches still need to be integrated into ECA frameworks.

Conclusion

Empathic capacity, as the capacity to relate and react to another's emotional state, consists of emotional communication competence, emotion regulation and cognitive mechanisms that result in a broad spectrum of behavior. A computational model of empathy should embody these capabilities to present a complete array of empathic behavior. Existing models and implementations of empathy in interactive agents including conversational agents often portray an incomplete picture of empathy or represent a binary classification of empathy. We propose a hierarchical approach to model empathy in interactive agents, that can be implemented by using some of the existing techniques in affective computing, social computing and dialogue research. Although each of these research areas provide advanced techniques that cover individual factors and components of empathic capacity, successful integration of them to reflect the hierarchical multi-leveled structure of empathy requires further investigation.

References

Arnold, M. B. (1960). *Emotion and personality.* Columbia University Press.

Bar-On, R., Tranel, D., Denburg, N. L., & Bechara, A. (2004). Emotional and social intelligence. Social neuroscience: key readings, 223.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM, 37*(7), 122–125.

Batson, C. (2009). *The social neuroscience of empathy (p. The social neuroscience of empathy, Chapter 2)*. The MIT Press.

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of machine Learning Research, 3*, 1137–1155 Feb.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(2), 293–327.

Boukricha, H., Wachsmuth, I., Carminati, M. N., & Knoeferle, P. (2013). *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference*. (pp. 1–6). IEEE.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36*(2), 129.

Brave, S., Nass, C., & Hutchinson, K. (2005). Computers that care: Investigating the effects of orientation of emotion exhibited by an embodied computer agent. *International Journal of Human-Computer Studies, 62*(2), 161–178.

Burleson, W., & Picard, R. W. (2004). Affective agents: Sustaining motivation to learn through failure and a state of stuck. In Workshop on Social and Emotional Intelligence in Learning Environments.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18–37.

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems, 31*(2), 102–107.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., & Yan, H. (2000). Conversation as a system framework: Designing embodied conversational agents. *Embodied Conversational Agents,* 29–63.

Clavel, C., & Callejas, Z. (2016). Sentiment analysis: From opinion mining to human-agent interaction. *IEEE Transactions on Affective Computing, 7*(1), 74–93.

Clavel, C., Pelachaud, C., & Ochs, M. (2013). User's sentiment analysis in face-to-face human-agent interactions–prospects. In Workshop on Affective Social Signal Computing, Satellite of Interspeech. Association for Computational Linguistics, August.

Coplan, A. (2011). In *Empathy: Philosophical and psychological perspectives*(pp. 3–18). Oxford University Press.

Costa, P. T., & MacCrae, R. R. (1992). *Revised NEO personality inventory (NEO PI-R) and NEO fivefactor inventory (NEO-FFI): Professional manual.* Psychological Assessment Resources Inc.

Damasio, A. R. (2006). *Descartes' error.* Random House.

Davis, M. (1994). *Empathy: A social psychological approach.* Madison, WI): Brown and Benchmark Publishers.

De Vignemont, F., & Singer, T. (2006). The empathic brain: How, when and why? *Trends in Cognitive Sciences, 10*(10), 435–441.

De Waal, F. B. (2007). The 'Russian doll' model of empathy and imitation. On being moved: From mirror neurons to empathy, 49–69.

De Waal, F. B., & Preston, S. D. (2017). Mammalian empathy: Behavioural manifestations and neural basis. *Nature Reviews Neuroscience, 18*(8), 498.

D'mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS), 2*(4), 23.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3–4), 169–200.

Ekman, P. (2004). *Language, knowledge, and representation (pp. 39-50)*. Dordrecht: Springer.

El-Nasr, M. S., Yen, J., & Ioerger, T. R. (2000). Flame—fuzzy logic adaptive model of emotions. *Autonomous Agents and Multi-Agent Systems, 3*(3), 219–257.

Goldman, A. I. (2012). *Theory of mind (Vol. 1).* Oxford Handbooks Online http://www.oxfordhandbooks. com/view/https://doi.org//10.1093/oxfordhb/9780195309799.001.

Goleman, D. (2006). Emotional intelligence. Bantam.

Gratch, J., & Marsella, S. (2004). A domain-independent framework for modeling emotion. *Cognitive Systems Research, 5*(4), 269–306.

Hoffman, M. L. (2000). *Empathy and moral development: Implications for caring and justice.* Cambridge University Press.

Jaimes, A., & Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding, 108*(1), 116–134.

Jaques, P. A., & Vicari, R. M. (2007). A BDI approach to infer student's emotions in an intelligent learning environment. *Computers & Education, 49*(2), 360–384.

Jokinen, K. (2003). Natural interaction in spoken dialogue systems. In Proceedings of the Workshop Ontologies and Multilinguality in User Interfaces (pp. 730–734). Crete.

Jokinen, K. (2009). *Constructive dialogue modelling: Speech interaction and rational agents.* John Wiley & Sons.

Jokinen, K., & McTear, M. (2009). Spoken dialogue systems. *Synthesis Lectures on Human Language Technologies, 2*(1), 1–151.

Lazarus, R. S. (1991). *Emotion and adaptation.* Oxford University Press on Demand.

Leiberg, S., & Anders, S. (2006). *Progress in Brain Research: . 156. The multiple facets of empathy: A survey of theory and evidence* (pp. 419–440). Elsevier.

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2014). Empathic robots for long-term interaction. *International Journal of Social Robotics, 6*(3), 329–341.

Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, B. (2016a). A persona-based neural conversation model. arXiv preprint arXiv:1603.06155.

Lisetti, C., Amini, R., Yasavur, U., & Rishe, N. (2013). I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS), 4*(4), 19.

Looije, R., Neerincx, M. A., & Cnossen, F. (2010). Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. *International Journal of Human-Computer Studies, 68*(6), 386–397.

Mairesse, F., & Walker, M. A. (2010). Towards personality-based user adaptation: Psychologically informed stylistic language generation. *User Modeling and User-Adapted Interaction, 20*(3), 227–278.

Marsella, S. C., & Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Cognitive Systems Research, 10*(1), 70–90.

McQuiggan, S. W., Robison, J. L., Phillips, R., & Lester, J. C. (2008). Modeling parallel and reactive empathy in virtual agents: An inductive approach. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1 (pp. 167–174).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

Minsky, M. (1991). Society of mind: A response to four reviews. *Artificial Intelligence, 48*(3), 371–396.

Moridis, C. N., & Economides, A. A. (2012). Affective learning: Empathetic agents with emotional facial and tone of voice expressions. *IEEE Transactions on Affective Computing, 3*(3), 260–272.

Omdahl, B. L. (1995). *Cognitive appraisal, emotion, and empathy.* Lawrence Erlbaum Associates Inc.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge University Press.

Osgood, C. E., May, W. H., & Miron, M. S. (1975). *Cross-cultural universals of affective meaning (Vol. 1).* University of Illinois Press.

Paiva, A., Leite, I., Boukricha, H., & Wachsmuth, I. (2017). Empathy in virtual agents and Robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS), 7*(3), 11.

Pantic, M., Sebe, N., Cohn, J. F., & Huang, T. (2005). *Proceedings of the 13th annual ACM international conference on Multimedia*. (pp. 669–676). ACM.

Pecune, F., Ochs, M. and C. Pelachaud. (2013). A formal model of social relations for artificial companions. In EUMAS 2013, December.

Pentland, A. (2005). Socially aware, computation and communication. *Computer, 38*(3), 33–40.

Picard, R. (1997). Affective computing/Rosalind W. Picard.

Picard, R. W. (2014). *The promise of affective computing. 11*. The Oxford handbook of affective computing.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(02), 169–226.

Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based interface that addresses users' affective states. *Applied Artificial Intelligence, 19*(3–4), 267–285.

Preston, S. D., & De Waal, F. B. (2002). Empathy: Its ultimate and proximate bases. *Behavioral and Brain Sciences, 25*(1), 1–20.

Salovey, P., & Mayer, J. D. (1990). Emotional intelligence. *Imagination, Cognition and Personality, 9*(3), 185–211.

Scherer, K. (1982). Emotion as a process: Function, origin and regulation. *Social Science Information, 21*(4–5), 555–570.

Scherer, K. R. (2010). Emotion and emotional competence: Conceptual and theoretical issues for modelling agents. *Blueprint for Affective Computing,* 3–20.

Scherer, K. R., Bänziger, T., & Roesch, E. (2010). Cross-cultural universals of affective meaning (Vol. 1). In *A blueprint for affective computing: A sourcebook and manual (pp. 1–20)*Oxford University Press.

Smith, M. (2011). Empathy, expansionism, and the extended mind. *Empathy: Philosophical and psychological perspectives,* 99–117.

Traum, D. R. (1993). Mental state in the TRAINS-92 dialogue manager. In Working Notes AAAI Spring Symposium on Reasoning about Mental States: Formal Theories and Applications (pp. 143–149).

Traum, D. (2017). *Computational approaches to dialogue. 143*. The Routledge Handbook of Language and Dialogue.

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., & Schroeder, M. (2012). Bridging the gap between social animal and unsocial machine: A survey of social signal processing. *IEEE Transactions on Affective Computing, 3*(1), 69–87.

Wahlster, W., & Kobsa, A. (1989). *User models in dialog systems (pp. 4-34)*. Berlin Heidelberg: Springer.

Wang, F. Y., Carley, K. M., Zeng, D., & Mao, W. (2007). Social computing: From social informatics to social intelligence. *IEEE Intelligent Systems, 22*(2).

Wilks, Y., Catizone, R., Worgan, S., & Turunen, M. (2010). Some background on dialogue management and conversational speech for dialogue systems. *Computer Speech & Language, 25*(2), 128.