ISSN 1846-6168 (Print), ISSN 1848-5588 (Online)
https://doi.org/10.31803/tg-20220504100707

Preliminary communication
Received: 2022-05-04, Accepted: 2022-05-05

Context-Driven Method in Realization of Optimized Human-Robot Interaction

Leon Koren*, Tomislav Stipančić, Andrija Ričko, Juraj Benić

Abstract: Perceptual uncertainty and environmental volatility are among the most enduring challenges in robotic research today. Contemporary robotic systems are usually designed to work in specific and controlled domains where a total number of variables is defined. Traditional solutions therefore often result in over-constrained interaction spaces or rigid system architectures where any unexpected change can result in system failure. The focus of this work is set on achieving a constant adaptation of the system to changes through interaction. A computational mechanism based on the entropy reduction method is integrated along with the three-component control model. This model is seen as a context-to-data interpreter used to provide context-aware reasoning to the technical system. The mechanism is using a decrease in interaction uncertainties when proofs are provided to the system. In this way, the robot can choose the right interaction strategy that resolves reasoning ambiguities most efficiently.

Keywords: affective; cognitive informatics; context; emotions; human-robot interaction; knowledge representation; probabilistic reasoning; robotics; ubiquitous computing

1 INTRODUCTION

The problems associated with smart system control are recognized as being the most important open problems in the field of robotics. The paradigm shift in the development of smart systems is moving toward a realization of adaptive system skills. The relationship and role the system has within its environment are slightly becoming the focus of research today [1].

Human-robot interaction (HRI) is a very prominent and widespread research activity in which the focus is set on processes of interaction [2]. In contemporary robotics, the focus is usually set on participants where the robot control routines should cover all possible outcomes that the environment could produce. But the nature of the environment is chaotic and undefined in terms of degrees of freedom in temporal or spatial dimensions. Some researchers, therefore, tried to decrease the level of uncertainty by constraining the environment [3]. Such approaches often lead to complex, expensive, or inefficient systems that are not adaptable to the current changes occurring within the environment. In this vision, industrial robots are often placed behind the safety fences or boxes to ensure safe and efficient work. These systems are not suitable to work efficiently in an open and unconstrained world. On the other hand, living beings found the path in constant interaction and adaptation to the environment where they are placed. For example, a human brain to ensure existence is using and shaping the current thoughts, feelings, and decisions by using memories, imagination, predictions, guessed facts, etc. It seems that it is impossible to achieve such tremendous capabilities in any artificial existence today [4].

Interaction is slightly becoming the mandatory skill that every cognitive system should have to adapt to the current changes. Besides the reasoning mechanism, such systems use different sensors to track the significant changes. These sensors are used as a part of sensing modalities to analyze different information spaces including vision, sound, touch, etc. Based on the number of used sensing modalities these inputs are then fused in a multimodal approach [5].

The probabilistic nature of interaction requires control mechanisms that can work under conditions of uncertainties. Artificial Intelligence is the scientific field that provides a plethora of possibilities for the design of such systems, including many different techniques from Deep Learning or Probabilistic Computing. For example, Deep Neural Networks are efficiently used in many areas of data mining and reasoning, such as Natural Language Processing – NLP [6] or Machine Vision [7] tasks. On the other hand, some other AI techniques are very efficient in probabilistic reasoning, like Bayesian Networks [8] or Markov Hidden Models which are very suitable for reasoning in the temporal domain [9].

Interaction, in essence, is the natural process through which living beings are acquiring new information by use of senses to detect significant and meaningful changes. Some authors, especially in the field of cognitive psychology, connected this process with the creation of mental reasoning maps associated with the representation of the current thoughts or occurrences within the physical environment [10]. Within this vision, the interaction is used to decrease the reasoning error through the continuous process in the loop. For example, within the Theory of Constructed Emotions proposed by the same author, *"variations and not uniformity is the norm"* and similar occasions within the same individual can result in different responses. The partial information currently concluded or acquired by senses can instantly change the reasoning output. In this way, a single piece of information can result in recognition, or it can result in a change in perspective [11].

In this work, the authors developed a model based on the semantic representation of the environment and probabilistic reasoning. This computation model is realized as a three-component computation mechanism, containing: (i) the part for analyzing the environment based on principles of Ubiquitous Computing, (ii) the semantic representation part, and (iii) the probabilistic reasoning part. In (i), the model uses a couple of inputs to constantly collect the significant information based on a multimodal approach. In (ii), a semantic representation of the environment is developed in a form of ontology. In (iii), the probabilistic reasoning part is realized in a form of a Bayesian Network (BN). BN is used to provide a single solution for the robot. Even though it is very rare, BN can become undecidable about the most appropriate solution. Therefore, as a special part of interaction strategy, a reasoning mechanism is developed. This mechanism is based on the entropy reduction method to resolve ambiguities derived within BN by suggesting the most appropriate subroutine for execution.

This solution is implemented on an affective robot PLEA, as shown in Fig. 1.

Figure 1 PLEA affective robot head at the Art & AI Festival in Leicester, UK, 2022.

PLEA is an interactive biomimicking robot head [12]. PLEA samples its environment to reason about the feelings of the person in the interaction. PLEA then demonstrates its affection using its visual expressions.

The rest of this paper is organized as follows. Section 2 rationale about the theoretical concept of the proposed model for the robot interaction control in which the more detailed insights are provided about all components of the model. Section 3 introduces the interaction reasoning concept in which the interaction is analyzed, and the new interaction strategy is proposed. Based on the acquisition of new information during the interaction the virtual agent can resolve the reasoning ambiguities.

The physical model implementation containing all model components is discussed in Section 4. At the end of the paper, a conclusion, summarizing the main concepts and the potential future research directions is provided in Section 5.

2 MODEL FOR THE ROBOT INTERACTION CONTROL

Within the working scenario, the robot is observing two people in communication and guessing about the quality of their interaction. A reaction of the robot is expressed by appropriate facial expressions shown on the robot's face. For the information visualization part, it is used a small light projector placed within the neck part of the robot, as shown in Fig. 1. When the robot detects a break in communication between those two observed people, according to a determined occasion, the robot will use the appropriate facial response.

2.1 Information Acquisition Sensing Modalities

The model for the robot action control is made of three main parts, as shown in Fig. 2.

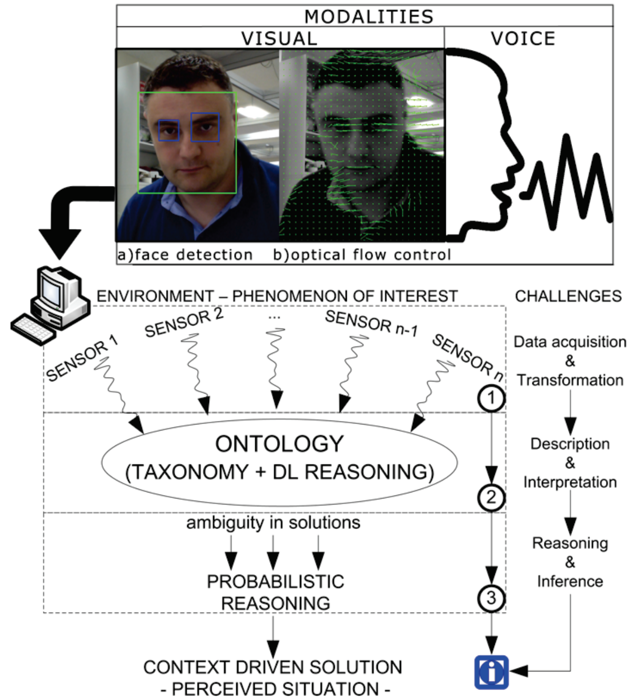


Figure 2 The model for the robot action control

The first part of the model is used for information acquisition based on the principles of Ubiquitous Computing [13]. Ubiquitous Computing is a paradigm of embedded computational capability in which smart sensors and devices are placed within some environment of interest to track significant occurrences. These occurrences are in a close relationship to the activities of the robot which are defined by control algorithms. In this way, program subroutines that control the robot's responses are intertwined with the acquired information in a meaningful way. The robot is then capable of reacting in accordance with environmental changes.

Under the visual modality, the model uses two submodalities. The first one is used for detecting the person's face. For this purpose, the model relies on ResNet neural network architecture [14]. Detection of one or two faces indicates a break in communication because people stopped communicating and started to pay attention to the robot. Combining those two submodalities indicates the level of distraction. The second submodality is used to measure the level of body movements of the person in the interaction. This information also provides insights into the inner psychophysical status of the person, including the level of excitement or nervousness. This submodality relies on the Optical Flow algorithm which is defined as the pattern of apparent motion of objects, surfaces, and edges in a visual scene caused by the relative motion between an observer and scene [15, 16]. The level of body movements is represented numerically. If the value of the sensor output is over a predefined threshold, the contextual meaning behind that value indicates the excitement or nervousness of the person.

The second sensing modality is the voice modality, which is defined similarly to the body movement submodality. When the numerical value of the sensor output reaches a predefined threshold value, the sensor will return the value 1 instead of 0. This value indicates a very loud conversation describing its nature (quarrel or great excitement of the person).

In both cases, threshold values are determined through empirical measurements with human subjects in an uncontrolled environment.

2.2 Formal Knowledge Representation

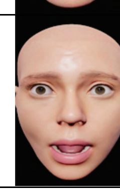
The formal knowledge about the application is stored in a form of ontology.

Table 1 Knowledge representation in a form of ontology

ID	V	$F1$	$F2$	$B1$	$B2$	Emotion Classes	Robot Response (R)
01	0	0	0	0	0	Support	01, 02
02	0	0	0	1	0	Support	01, 02
03	0	0	0	0	1	Support	01, 02
04	0	0	0	1	1	Neutral	03, 04
05	1	0	0	1	1	Neutral, Condemn	03, 04, 05
06	0	0	1	0	0	Support	01, 02
07	0	0	1	0	1	Neutral	03, 04
08	0	0	1	1	0	Neutral	03, 04
09	1	0	1	1	0	Neutral, Condemn	03, 04, 05
10	0	0	1	1	1	Neutral, Condemn	04, 05
11	1	0	1	1	1	Condemn	05, 06
12	0	1	0	0	0	Support	01, 02
13	0	1	0	0	1	Neutral	03, 04
14	1	1	0	1	0	Neutral, Condemn	03, 04, 05
15	0	1	0	0	1	Neutral	03, 04
16	0	1	0	1	0	Neutral, Condemn	03, 04, 05
17	0	1	0	1	1	Neutral, Condemn	04, 05
18	1	1	0	1	1	Condemn	05, 06
19	0	1	1	0	0	Neutral	03, 04
20	1	1	1	0	0	Neutral, Condemn	03, 04, 05
21	0	1	1	0	1	Neutral, Condemn	04, 05
22	1	1	1	0	1	Condemn	05, 06
23	0	1	1	1	0	Neutral, Condemn	04, 05
24	1	1	1	1	0	Condemn	05, 06
25	0	1	1	1	1	Condemn	05, 06
26	1	1	1	1	1	Condemn	05, 06
27	Other sensor combinations					Support	01, 02

From the Artificial Intelligence point of view, an ontology is a set of concepts and categories related to some domain of interest showing their properties and relations between them. Ontologies denote a formal representation of entities (classes) along with associated attributes (objects) and their mutual relations [17]. Ontology is written in Ontology Web Language – OWL [18], as is the case in this work, it is made of taxonomy and Descriptive Logic - DL [19]. Taxonomy is made of classes, attributes, and individuals within classes. Employing DL, taxonomy can be developed into an ontology in which different classes, attributes, or individuals are connected in meaningful ways by DL conjunctions to form DL sentences that represent knowledge. Some authors stated that certain types of contexts are important while characterizing a situation of a particular entity [13]. In this way, these sentences become contextual and can provide different results by changing the query input parameters. Ontology written in OWL is following so-called Open World Assumptions (OWA), meaning the fact that all generated responses have the same level of importance [20]. In that case, for the same combination of the input sensor values, the knowledge base will propose more than one possible solution resulting in ambiguities in reasoning, as defined in Tab. 1. The table abbreviations are as follows: V – voice detected, $F1$ – Face 1 detected, $F2$ – Face 2 detected, $B1$ – intense Body 1 movements detected, $B2$ – intense Body 2 movements detected.

Table 2 Information Visualisation

Emotion Classes	Robot Responses	Goodness	Facial Expressions
Support	$R01$		
Support	$R02$		
Neutral	$R03$	⇑ Reactions to perceived ⇓	
Neutral	$R04$		
Condemn	$R05$		
Condemn	$R06$	Badness	

Facial expressions are separated into three defined classes: *Support*, *Neutral*, and *Condemn*. Each of these classes contains two facial expressions, as shown in Tab. 2. When a certain sensor combination occurs (for example, Tab. 1, ID17), the ontology will suggest the list of the most possible solutions respecting the emotion class (for ID17 the ontology suggests 04 or 05). In the provided example the robot will remain unsure about what is the most suitable solution because both suggested solutions have the same importance. Therefore, the reasoning mechanism is updated with probabilistic reasoning in a form of a Bayesian Network (BN).

2.3 Probabilistic Reasoning

The ontology can suggest more than one solution having the same priority and significance based on the principles of Open World Assumptions. That causes ambiguities in which the robot cannot decide on the most suitable solution for the current context. Therefore, to resolve ambiguities the initial reasoning model is updated with the probabilistic reasoning part in a form of a Bayesian Network – BN [21, 22]. BNs are directed acyclic graphs containing nodes (variables) and links that connect them. Links indicate a direction of influences that parent nodes have on their children through the structure of BN. Each network variable has a Conditional Probability Table – CPT, in which the influences of ancestors are expressed. BN also has query nodes at the end of the structure, which contain beliefs or percent of certainty of the network about the most probable solutions, as shown in Fig. 3.

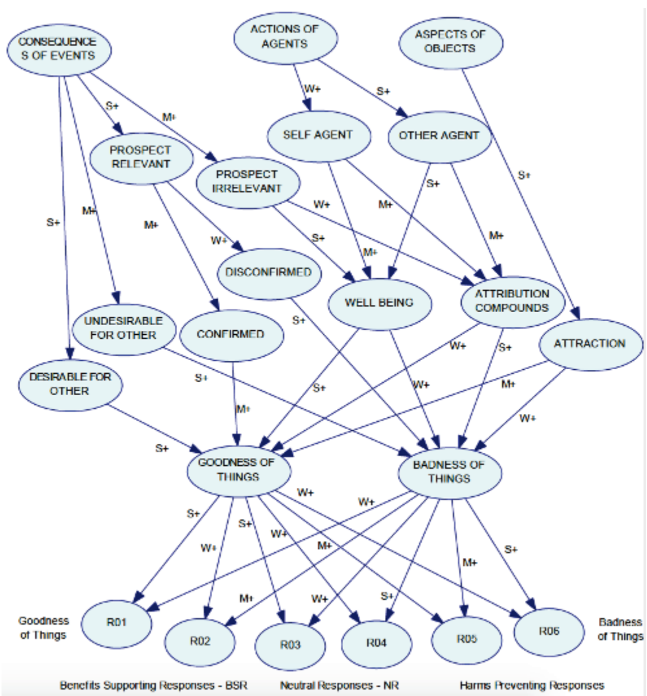


Figure 3 BN architecture used as a mechanism for reasoning under uncertainties

The presented BN is used to resolve ambiguities that appeared as a result of the ontology reasoning where for each sensor combination the system has more than one possible solution, as shown in Tab. 1.

This BN contains the OCC model of emotion that has a taxonomy of emotions in three main pillars (Consequences of Events, Actions of Agents, and Aspects of Objects), as described in [23]. The overall description of all network variables is given in Tab. 3.

Table 3 Network nodes

Node (variable)
Consequences of Events (GE): believes about the nature of events (*pleased* or *displeased*).
Actions of Agents (AA): believes about the intentions of agents (*approving* or *disapproving*).
Aspects of Objects (AO): believe in the likeness of perceived objects (*liking* or *disliking*).
Prospect Relevant (PR): expectations about some event (*hope* or *fear*).
Prospect Irrelevant (PI): expectations about some event (joy or distress).
Self-Agent (SA): feelings about the relation of the robot towards the environment (*pride* or *shame*).
Other Agent (OA): feelings of the robot towards actions of the person in interaction (*admiration* or *reproach*).
Desirable for Other (DO): reasoning about events concerning some other person (*happyFor* or *resentment*).
Undesirable for Other (UO): reasoning about events concerning some other person (*gloating* or *pity*).
Confirmed (C): feelings about some event that happened (*satisfaction* or *fearsConfirmed*).
Disconfirmed (D): feelings about some event that does not happen (*relief* or *disappointment*).
Well Being (WB): evaluation of action or event during interaction (*gratification* or *remorse*).
Attribution Compounds (AC): evaluation of action or event during interaction (*gratitude* or *anger*).
Attraction (A): evaluates the level of affection (*love* or *hate*).
The Goodness of Things (GT): aggregated effect of other nodes about the goodness of things (*good* or *neutral*).
The Badness of Things (BT): aggregated effect of other nodes about the badness of things (*bad* or *neutral*).
Query Nodes (R): 01, 02, 03, 04, 05, 06 (*true* or *false*).

Each of these nodes in BN has its Conditional Probability Table – CPT which is calculated using a hand-crafted approach based on strength of influences, as shown in Fig. 3. The procedure of BN development is described more in detail in [24, 25]. For example, at the top of the network, the "Consequence of Events" variable is positioned which has a strong positive influence on the "Desirable for Others" variable. That node, at the same time, has a strong influence on the "Goodness of Things" node which at the end of BN has a weak influence on the R02 query node. This influence is spreading from the top node through the network. The strength of influence is used to calculate CPT for each node using the Eq. (1).

$$WF = \frac{\left(TV - \dfrac{1}{NS}\right)}{NP} \tag{1}$$

WF stands for Weight Factor, *NS* for Number of States, *TV* for Threshold Value, and at the end *NP* stands for Number of Parents. The threshold is the value that is directly connected to the Strength of Influences, where S+ has a threshold value of 0.98, M+ has a threshold value of 0.8, and W+ has a value of 0.6. These values are defined arbitrarily to represent the real influences of links.

Condition probability tables for each of the nodes are calculated using the same methodology in which the previously calculated values for Weight Factors are used. The equations for calculating the values in CPT for the query node $R01$ are listed in Eqs. (2) - (9).

$$(R01 = true / GoodnessOfThongs = good \,\&\, BadnessOfThings = bad) \tag{2}$$

$$(R01 = true / GoodnessOfThongs = good \,\&\, BadnessOfThings = neutral) \tag{3}$$

$$(R01 = true / GoodnessOfThongs = neutral \,\&\, BadnessOfThings = bad) \tag{4}$$

$$(R01 = true / GoodnessOfThongs = neutral \,\&\, BadnessOfThings = neutral) \tag{5}$$

$$(R01 = false / GoodnessOfThongs = good \,\&\, BadnessOfThings = bad) \tag{6}$$

$$(R01 = false / GoodnessOfThongs = good \,\&\, BadnessOfThings = neutral) \tag{7}$$

$$(R01 = false / GoodnessOfThongs = neutral \,\&\, BadnessOfThings = bad) \tag{8}$$

$$(R01 = false / GoodnessOfThongs = neutral \,\&\, BadnessOfThings = neutral) \tag{9}$$

As seen, equations express the influences that parents have on the child node. The same procedure is repeated to calculate all other CPTs in BN. Within the BN reasoning phase, evidence can be acquired in advance which can affect the reasoning output of the network.

3 INTERACTION DRIVEN REASONING

Interaction is a very important part of communication [26, 27]. During the interaction, all participants exchange verbal and nonverbal signals that affect and drive the communication. These signals usually contain crucial information representing the proofs that shape the communication process. In the same way, BN is used as a part of the reasoning model in this work, as it contains variables (nodes) which can be affected by information acquired from the environment. As explained earlier, CPTs directly control the reasoning outputs of the network. In this way, CPT stores beliefs and influences of all parent nodes that are connected to some node. In the case that the proof about some node in the network is provided, this node is not altered by parent influences anymore. The influence of this node starts spreading through the network until it reaches the query nodes. In this way, the network can change the reasoning output during the interaction. For example, the robot could find out the specific information about some node in the network and update CPT accordingly. That could strengthen believes of the network that some query node is the right solution according to the current context. On the other hand, it is rare but possible that BN cannot decide about one solution, as shown in Fig. 4.

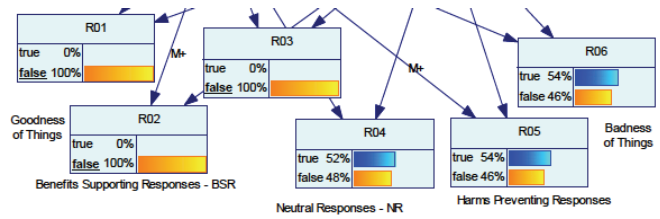


Figure 4 Ambiguities in BN reasoning

In this example, the ontology suggested $R04$, $R05$, and $R06$ as solutions (query nodes $R01$, $R02$, and $R03$ are set to false because the ontology did not suggest them as solutions). As can be seen, BN expressed the same percent of certainty in all three solutions meaning that BN cannot decide which one is the most appropriate. In this case, the robot can find the proofs through interaction which can resolve ambiguities and suggest only one solution.

Every node in BN has its influence on expressing beliefs when deciding which query node is the most suitable candidate for a final solution. In that case, some nodes in the network will generate a more significant change in percent of certainty for query nodes compared to some less influential nodes. Therefore, it could be very useful to determine the most influential nodes in the network. By knowing this information, the robot can plan its interaction to resolve all ambiguities efficiently.

The approach used in this work is based on the entropy reduction method which is closely related to sensitivity analysis [28]. By using Eq. (10), it is possible to calculate the current level of uncertainty in any node of interest.

$$H(P) = -\sum_{s \hat{\text{I}} S} P(s) * \log_2 P(s) \tag{10}$$

As a part of information science, Shannon's Entropy is a measure of uncertainties [29]. Every node in BN contains a certain level of uncertainty. By using the entropy reduction method, it is possible to determine a drop in uncertainty in query nodes when evidence or proof is provided to some particular node in the network, as shown in Eq. (11).

$$I = H(before) - H(after) = H(Q) - H\left(\frac{Q}{F}\right) \tag{11}$$

By using this approach, it is possible to calculate the entropy decrease in query nodes in relation to other nodes in the network, as shown in Fig 5.

Fig. 5 shows that the nodes CE (Consequences of Events - *pleased* or *displeased*), DO (Desirable for Other - *happyFor* or *resentment*) and PR (Prospect Relevant - *hope* or *fear*) have the greatest influence on all query nodes when evidence is provided because they cause the greatest drop in uncertainty. In this way, the robot must find out the information about these tree nodes during the interaction to resolve ambiguities. The nodes GT and BT are not involved in the graph representation as they are representing the Goodness and Badness of Things. As shown in Fig 6., these two nodes are influenced by all previous nodes and all these influences are summed in them so they have the greatest influence on the query nodes and showing them would cloud all others.

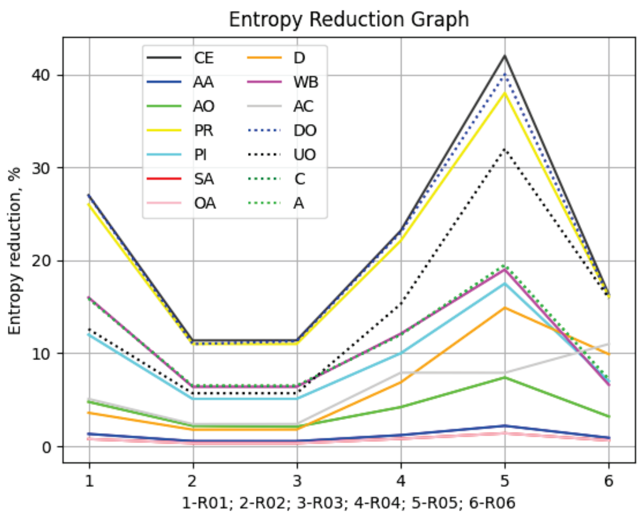


Figure 5 The graphical representation of entropy reduction

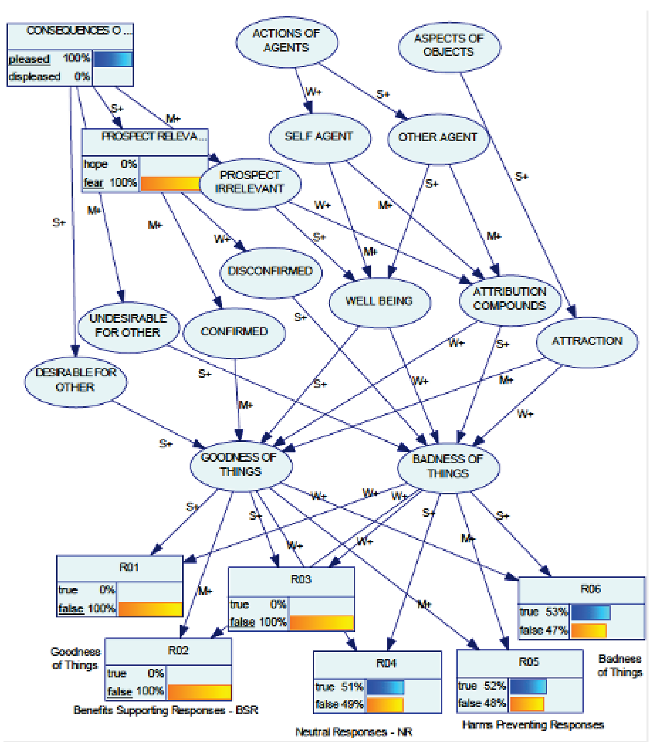


Figure 6 The BN interaction reasoning procedure

To resolve ambiguity presented in the Fig. 4 example, where the ontology suggested three possible solutions having the same percent of certainty (*R*04, *R*05, and *R*06), BN reasoning is used during the interaction. The robot can then find proofs about the three most influential nodes (CE, DO, and PR) and update its reasoning in real-time. BN representing the process of such reasoning is shown in Fig. 6. During the interaction, the robot firstly acquired the information about the CE node, and according to this information the CE node is set to "pleased". But that did not cause any change in the query nodes. Within the next step, the PR node is set to "fear" which caused the query node *R*06 to prevail with 53 %. The proposed solution is in accordance with the current context because the response *R*06 represents the reaction of the robot when the context in the environment is perceived as bad.

4 PHYSICAL MODEL IMPLEMENTATION

The model implementation is shown in Fig. 7.

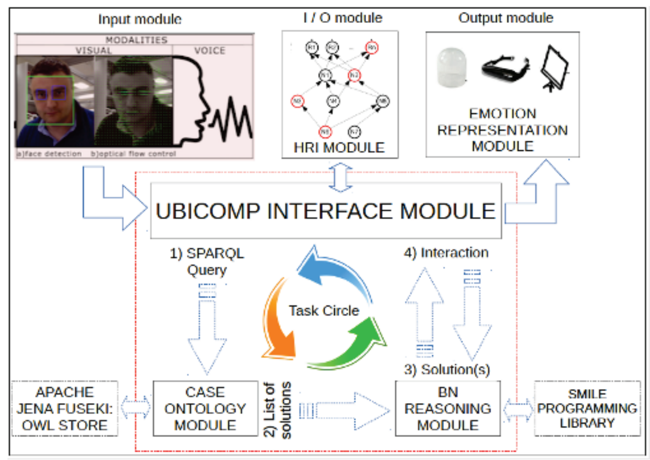


Figure 7 Physical model implementation

The very true heart of the system is the Ubicomp Interface Module which is used to receive data from the Input Module and present results using the Output Module. The Input Module contains mechanisms for multimodal acquisition and fusion of information, as explained earlier. The Ubicomp Module receives the query vector containing the information on whether the face(s) is (are) detected, or whether the detected movements are significant or not. The last information is acquired by the sound modality as the level of noise in the room.

When the query containing the status of sensing modalities is received by the ontology, the list of possible solutions is proposed, and the BN reasoning module can determine the single solution that is going to be executed. This solution is a facial expression in which the robot is expressing its current emotions. The HRI module is used if the BN reasoning module cannot find a single solution and resolve ambiguities. This module then uses the described interaction strategy to determine a single solution.

The Output Module is used to show the face of the virtual agent and can be a physical head of the robot or any other device like a Smart TV, XR glasses, or the screen of a PC.

5 DISSCUSION AND CONCLUSION

The presented methodology enabled a virtual agent to resolve the reasoning ambiguities and find the most optimal solution in relation to the current context. This methodology is a part of a research project in which the virtual agent PLEA is exchanging nonverbal signals with people in interaction. By changing the facial expressions, PLEA can drive the interaction and guess about the next steps in communication by analysing the previous facial expressions of the communicating person. The methodology that controls the robot, in this case, is based on Deep Learning models and state-of-the-art computer network technology, as the virtual agent can respond to distant connections and show up using communication interfaces wherever in the world. The virtual agent can accomplish its actions using XR or devices that operate within the real world.

The computational model presented in this work is a step forward to putting PLEA in communication with a group of people where the virtual agent can observe and analyse the interaction as a more passive member of the group. The model is in a testing phase and a couple of refinements could be implemented before this model becomes a part of the main model. For example, the interaction when PLEA is resolving ambiguities is currently realized as a chatbot agent that asks questions. The result of such an approach is a break in the communication grounding process that takes place between people in interaction whenever PLEA asks something [30, 31]. To be more natural and less intrusive, in future work the virtual agent will guess independently, without asking direct questions, to resolve ambiguities. By using new sensing modalities within the multimodal approach, PLEA will analyse hidden communication cues to resolve whether somebody is happy or not.

Another improvement and analysis can be made in providing different strength of influence values to BN links. These values are subjectively defined and represent the vision of a system designer about the nature of the agent's responses. By changing these values, the agent can express different moods during the interaction. For example, sometimes the agent can be more sensitive. The way the agent behaves can be also context-driven and can depend on the way how people interact and communicate mutually.

Acknowledgments

This work has been supported in part by the Croatian Science Foundation under the project "Affective Multimodal Interaction based on Constructed Robot Cognition—AMICORC (UIP-2020-02-7184)".

Notice

The paper was presented at MOTSP 2022 – 13th International Conference Management of Technology – Step to Sustainable Production, which took place in Primošten/Dalmatia (Croatia) on June 8–10, 2022. The paper will not be published anywhere else.

6 REFERENCES

[1] Lasota, P. A., Fong, T., & Shah, J. A. (2017). *A survey of methods for safe human-robot interaction* (Vol. 104). Delft, The Netherlands: Now Publishers. https://doi.org/10.1561/9781680832792

[2] Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., & Šabanović, S. (2020). *Human-robot interaction: An introduction*. Cambridge University Press. https://doi.org/10.1017/9781108676649

[3] Stipancic, T., Jerbic, B., & Curkovic, P. (2016). A context-aware approach in realization of socially intelligent industrial robots. *Robotics and computer-integrated manufacturing*, 37, 79-89. https://doi.org/10.1016/j.rcim.2015.07.002

[4] Rebout, N., Lone, J. C., De Marco, A., Cozzolino, R., Lemasson, A., & Thierry, B. (2021). Measuring complexity in organisms and organizations. *Royal Society Open Science, 8*(3), 200895. https://doi.org/10.1098/rsos.200895

[5] Koren, L., Stipancic, T., Ricko, A., & Orsag, L. (2022). Person Localization Model Based on a Fusion of Acoustic and Visual Inputs. *Electronics, 11*(3), 440. https://doi.org/10.3390/electronics11030440

[6] Chowdhary, K. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603-649. https://doi.org/10.1007/978-81-322-3972-7_19

[7] Golnabi, H. & Asadpour, A. (2007). Design and application of industrial machine vision systems. *Robotics and Computer-Integrated Manufacturing, 23*(6), 630-637. https://doi.org/10.1016/j.rcim.2007.02.005

[8] Stipancic, T., Jerbic, B., & Curkovic, P. (2013). Bayesian Approach to Robot Group Control. In: Ao, SI., Gelman, L. (eds) *Electrical Engineering and Intelligent Systems. Lecture Notes in Electrical Engineering*, vol 130. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-2317-1_9

[9] Li, T., Choi, M., Fu, K., & Lin, L. (2019, December). Music sequence prediction with mixture hidden Markov models. In *2019 IEEE International Conference on Big Data (Big Data)*, 6128-6132. https://doi.org/10.1109/BigData47090.2019.9005695

[10] Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Boston: Houghton Mifflin Harcourt. https://doi.org/10.7202/1064926ar

[11] Koren, L. & Stipancic, T. (2021, July). Multimodal Emotion Analysis Based on Acoustic and Linguistic Features of the Voice. In *International Conference on Human-Computer Interaction*, Springer, Cham, 301-311.

[12] Stipancic, T., Koren, L., Korade, D., & Rosenberg, D. (2021). PLEA: A social robot with teaching and interacting capabilities. *J. Pac. Rim Psychol.*, 15, 18344909211037019. https://doi.org/10.1177/18344909211037019

[13] Krumm, J. (Ed.). (2018). *Ubiquitous computing fundamentals*. CRC Press. https://doi.org/10.1201/9781420093612

[14] Ranjan, R., Patel, V. M., & Chellappa, R. (2017). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41, 121–135. https://doi.org/10.1109/tpami.2017.2781233

[15] Kemouche, M. S., Aouf, N., & Richardson, M. (2013). Object detection using Gaussian mixture-based optical flow modelling. *The Imaging Science Journal, 61*(1), 22-34. https://doi.org/10.1179/1743131x11y.0000000050

[16] Farnebäck, G. (2003). Two-Frame Motion Estimation Based on Polynomial Expansion. In: Bigun, J., Gustavsson, T. (eds) Image Analysis. *SCIA 2003. Lecture Notes in Computer Science*, vol 2749. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45103-X_50

[17] Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies, 43*(5-6), 907-928. https://doi.org/10.1006/ijhc.1995.1081

[18] McGuinness, D. L. & Van Harmelen, F. (2004). OWL web ontology language overview. W3C recommendation, 10(10), 2004.

[19] Baader, F., Horrocks, I., & Sattler, U. (2008). Description logics. *Foundations of Artificial Intelligence*, 3, 135-179. https://doi.org/10.1016/S1574-6526(07)03003-9

[20] Loyer, Y. & Straccia, U. (2005). Any-world assumptions in logic programming. *Theoretical Computer Science, 342*(2), 351-381. https://doi.org/10.1016/j.tcs.2005.04.005

[21] Marcot, B. G. & Penman, T. D. (2019). Advances in Bayesian network modelling: Integration of modelling technologies. *Environmental modelling & software*, 111, 386-393. https://doi.org/10.1016/j.envsoft.2018.09.016

[22] Daniel, B. K., Zapata-Rivera, J. D., & McCalla, G. (2003). A Bayesian computational model of social capital in virtual communities. In *Communities and technologies*, Springer, Dordrecht, 287-305. https://doi.org/10.1007/978-94-017-0115-0_15

[23] Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions Cambridge*. UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511571299

[24] Twardy, C. R., Nicholson, A. E., Korb, K. B., & McNeil, J. (2006). Epidemiological data mining of cardiovascular Bayesian networks. *Electronic Journal of Health Informatics, 1*(1), 3.

[25] Jerbic, B., Stipancic, T., & Tomasic, T. (2015, November). Robotic bodily aware interaction within human environments. In *2015 IEEE SAI Intelligent Systems Conference (IntelliSys)*, 305-314. https://doi.org/10.1109/IntelliSys.2015.7361160

[26] Scholl, W. (2013). The socio-emotional basis of human interaction and communication: How we construct our social world. *Social Science Information, 52*(1), 3-33. https://doi.org/10.1177/0539018412466607

[27] Grönroos, C. (2000). Creating a relationship dialogue: communication, interaction and value. *The marketing review, 1*(1), 5-14. https://doi.org/10.1362/1469347002523428

[28] Oakley, J. E. & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(3), 751-769. https://doi.org/10.1111/j.1467-9868.2004.05304.x

[29] Daniel, B. K., Zapata-Rivera, J. D., & McCalla, G. (2003). A Bayesian computational model of social capital in virtual communities. In *Communities and technologies*, Springer, Dordrecht, 287-305. https://doi.org/10.1007/978-94-017-0115-0_15

[30] Clark, H. & Brennan, S. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition*. Washington, DC: American Psychological Association, 127-149. https://doi.org/10.1037/10096-006

[31] Nathan, M. J., Alibali, M. W., & Church, R. B. (2017). Making and breaking common ground: how teachers use gesture to foster learning in the classroom. In *Why Gesture? How the Hands Function in Speaking, Thinking and Communicating*, eds R. Breckinridge Church, M. W. Alibali, and S. D. Kelly (Amsterdam: John Benjamins), 285-316. https://doi.org/10.1075/gs.7.14nat

Authors' contacts:

Leon Koren, mag. ing. mech.
(Corresponding author)
University of Zagreb,
Faculty of Mechanical Engineering and Naval Architecture,
Ivana Lucica 5, 10000 Zagreb, Croatia, EU
Phone: +38516168480, e-mail: leon.koren@fsb.hr

Tomislav Stipancic, Assist. Prof. Dr. Sc.
University of Zagreb,
Faculty of Mechanical Engineering and Naval Architecture,
Ivana Lucica 5, 10000 Zagreb, Croatia, EU
Phone: +38516168380, e-mail: tomislav.stipancic@fsb.hr

Andrija Ricko, univ. back. ing. mech.
University of Zagreb,
Faculty of Mechanical Engineering and Naval Architecture,
Ivana Lucica 5, 10000 Zagreb, Croatia, EU
E-mail: ar200783@stud.fsb.hr

Juraj Benic, mag. ing. mech.
University of Zagreb,
Faculty of Mechanical Engineering and Naval Architecture,
Ivana Lucica 5, 10000 Zagreb, Croatia, EU
Phone: +38516168357, e -mail: juraj.benic@fsb.hr