

TURIN : A coding system for Trust in hUman Robot Interaction

HULCELLE Marc, ROLLET Nicolas, VARNI Giovanna, CLAVEL Chloé
 LTCI, Télécom Paris, Institut Polytechnique de Paris – Palaiseau, France
 Contact : {firstname.lastname}@telecom-paris.fr

This work is supported by the Data Science and Artificial Intelligence for Digitalized Industry and Services (DSADIS) chair of Télécom Paris.

References:

- [1] K. E. Schaefer, "The Perception and Measurement of Human-robot Trust", Doctoral Dissertation, University of Central Florida, Orlando, 2013.
- [2] D. B. Jayagopi, S. Sheiki, D. Klotz, J. Wienke, J. M. Odobez, S. Wrede, V. Khalidov, L. Nyugen, B. Wrede, and D. Gatica-Perez, "The vernissage corpus: A conversational Human-Robot-Interaction dataset", ACM/IEEE International Conference on Human-Robot Interaction, pp. 149–150, 2013.

Trust

Adopting a sociological perspective, we define trust as:

A form of affiliation and credit characterized by a set of behaviors that are intentional or not, expressive or propositional.

In HRI, this definition allows interpreting behaviors based on:

- the social and competence credit they give to the robot
- recognition of human behaviors that include the robot as an autonomous agent inside the social interaction as signs of trust.

Definitions

Interactional trust: a state displaying a form of naturalness, or fluidity in the interaction.

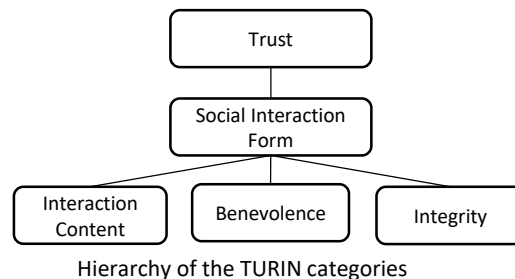
Naturalness: implies that the robot is treated as an interactional partner in the same way as a human partner would be treated. Estimated based on the dynamics of participants' behaviors rather than on the robot's embodiment and behaviors.

TURIN Overview

- Interactions are **segmented in homogeneous trust levels**
- Segmentation starts at the **single behavioral act level**.
- Individual acts referring to **changes in behaviors** are assigned to a trust category.
- **Consecutive acts of the same category** are aggregated to form a segment of homogeneous level of trust.

Inside trust annotations, we focus on verbal and non-verbal behaviors:

- **Social Interaction Form**: communication modes linked to norms of social interactions
- **Interaction Content**: functional descriptors
- **Benevolence**: the willingness to do good in general
- **Integrity**: respecting expectations of honesty and moral principles



Verbal and non-verbal behaviors

- **Trusting** behaviors:
 - display **interactional trust**
 - accept **vulnerability**
 - seem **friendly**
 - acknowledge the partner's **competence**.
- **Mistrusting** behaviors display:
 - **uneasiness**
 - **doubt**
 - **confusion**
 - **aggressiveness**
 - **unwillingness to cooperate**
 - and trust-reparation behaviors
- **Neutral** behaviors: other behaviors that do not allow us to conclude on the participant's trust level.

Social Interaction Form	Interaction Content
Nod	Approval
Gaze	Alignment
Gesture	Compliance
Phrasing	Cooperation
Intonation	Out-of-context comment
Repetition	Joke
F-formation*	Doubt
Speaking Turn	
Facial Expression	
Participation Status*	
Benevolence	Integrity
Respect	Honesty
Warmth	Promise
Personal Info Disclosure	Responsibility
	Manipulation

Proof-of-Concept

- 2 experts in HRI annotated 3 minutes of the Vernissage corpus
- IRA (Cohen's Kappa) substantial for **trusting** and **mistrusting** segments
 - Easier to recognize errors of social norms than high naturalness
- **Gaze, speaking turns, facial expressions** and **nods** coded more often than other items.
 - IRA low because segments are short, hard to define start and end

Segment category	IRA (κ)	Mean duration (s)	Std (s)
Mistrusting	0.79	4.6	2.2
Trusting	0.64	2.1	1.5
Neutral	0.45	4.7	4.6

Segment category	IRA (κ)	Mean duration (s)	Std (s)	Count
Nod	0.52	1.4	0.6	15
Gaze	0.36	1.6	1.4	27
Gesture	0.56	1.9	1.3	7
Phrasing	0.42	1.3	0.5	3
Repetition	0.89	0.9	0.1	2
Intonation	0.74	1.6	1.0	7
F Formation	0.80	2.1	0.9	10
Speaking turn	0.80	1.3	0.8	26
Facial expression	0.41	1.3	0.8	19
Participation status	0.80	3.5	1.9	12

Conclusion

- First coding system to study trust in HRI
- Further work on « Benevolence » and « Integrity » planned
- Full validation of the system is needed
- Will be used to build computational models of trust