

dipartimento di ingegneria

AGENTS AND ROBOTS JUSTIFY ACTIONS

Angelo Maria Pio Sabella, Valeria Seidita and Antonio Chella

Implementing trustfulness and explainability







Objectives

How to model and develop teammate agents (robots) performing trustful interaction with humans?

- Deciding and acting in autonomous fashion
- Modelling and representing agents' knowledge
- Explainability -> Trustworthiness

Valeria Seidita - ITAL-IA 2013



2









Ex- white the set	Sniffer@192.168.0.135:1099/JADE - Sniffer Agent				
Le - masThesis					
Loading Ibboost regardso	DL D	-			
Leading Amplitations, tedes to		- U.,			
Leading I bgl.so Leading: /tmp/ logi.se	ca ? D AgentP atforms	_			
Leading for the measurements of	e Pictor and a second s				
- Loading, Amp/logimessag.nghl.so	+ Main-Container Sthor cher adge Stanser subner				
Ubraries loaded, initializing type system	- @ HMW29192.161 0				
LPD has name [15]1m * [05;1mNetbods [0m					
t [34] = 100 [0n est [34] In/4[d] (0n [33] In/] [0n	– 🖾 Sniffer-on-Mai				
1 [34:1m102 (0m pCal) [34:1mValue (0m (33:1mValue (0m 5 [34:1m103 (0m version (34:1mString (0m (33:1mV) (0m)	☑ Snitter@162.1				
E Davis nua jan anali (Januari) (Jan	- @ sctioner@192				
[34:1+105 [Om getMethodList [34:1m] st~String> (0m [33:1m]) [0m	– 👜 ams@152.160				
[34:1m106_j0m_getMethodHelp = [34:1mValue_j0m_j33:1m(String)]0m	- 🖉 dt@192.168.0				
[24] (m 107 [2m ge Mudulet Max (24] (m 200; m (20); m	i knowledgeng				
[34:1m108 [0m wat [34:1m8oo][0m [33:1m0nt32.ht32] [0m					
[341m109]0m isRunning [341m8oo][0m [331m][m52)][0m	– 🖉 planner@192.				
[34;1m110]0m stop [34;1mVod [0m]]33;1m(ht32)]0m [34;1m111]0m ge:BrokerName [34;1mString [0m]]33;1m0]10m	🚆 visioner@192				
[34:1m112] (Om getUsage (34:1m3tring (Om (33:1m)String) (Om					
[34.1m113 [Om aubscribe]34.1mVod [Om]33.1m(String,m32,Float) [Cm					
[34]1=114 [3m subscribe [34]1mVad [3m [33]1m[31m]] [3m					
[34:1m115 (0m unsubscribe [34:1mVbid (0m]33:1m(String) (0m					
[34:1m116 [0m updatePeriod] [34:1mVoid [0m] [33:1m[String,int32] [0m					
[04] = 1.7 [Sen addete("sensors]04) = Wed [Sen [D0] (m(1) seg.] [kat]) [Ben					
[34] m138 [Om getCurrentPeriod] [34] mint32 [Om [33] [m]][Om [33] [m]					
[34:1m119 [0m getCurrentPrecision [34:1mFloot [0m [33:1m]] [0m [34:1m120 [0m getMyPeriod]] [34:1m1122 [0m [30:1m]] [0m					
[34] T121 [3m] get Wyrension [34] Theat [6m [33] m(Sring) [0m					
[34:1 m122 [Om getSubscribersInfo [34:1 mValue [0m [33:1 m1]]0m					
[94;1~123]0m getOutputNames (34;1mList <string> [Cm (35;1m/)]Cm</string>					
[14]1=124 [On gerSecuted [14]1ml x +Strings [On [17]1m] [On					
[34] m125 [Om getMemoryKeyUst [34] Imilist + Strings+ [Om [38] Im0] [Om]					
[94:1m126 (0m setFremeRate = [94:1mEool (0m (95:1m(String int32) (0m (04:1m127 (0m setFremeRate = (04:1m10ol (0m (00:00) (0m)))))					
[34:1+122 [0m set] came bit e [34:1+100] [0m [33:1m(n132) [0m] [34:1+128 [0m setResolution] [34:1+Eool [0m [33:1m(n132) [0m]					
[341 m129] On sch4ttyscamera [341 m600] On [331 m] https://on					
[34;1+130 [3h setParameter [34;1+Vod [0n [30;1+(String Value]]0n					
[34]]=13] [3n ge:FrameRate [34]1mnt32 [6m [33]]=(] [0m					
[34:1m132 [0m getResolution]34:1mmt32 [0m [38:1m/:]0m					
[34:1m133 [3m gs:ActiveCamera] [34:1mirt32 [0m] [33:1mi] [0m					
[34] m134 (3m is 2-sused (34,1 m3bo) (3m (33,1 m)) (3m) [34] m135 (3m is 2-sused) [24,1 m2bo) (3m (23,1 m)) (3m)					
[34:1m135 [0m isProcessing [34:1m3pol [0m [33:1m1) [0m] [34:1m136 [0m pause] [34:1mVpid [0m [33:1m18pol] [0m]					
Decay rutes for passe are needed on pass mechanism created.					
IsCaNo_surcher lade Container started					
[Run[adeW45] Creating opent planner (1(1)					
(Pur(adeWS) Creating agent visioner (1:0)					
[RunladeMAS] Creating agent knowledger (1/1)					
[Runjade/48] Creating agent actioner (1/1)					
(visioner) starting visioner Jechuner) starting actioner					
Is an estimated by a solution	92				
knewledgeri starting knewledger	Ω				
Nexat Ltip Server ratining on https://27.0.1.10272	0				
[CartagoÉnvironment] NÉw AGENT (0 NED) visioner	13				
[CarbogoEnvironment] NEW AGENT JOINED: planner					
[SarbageEmershment] NEW ADENT JOINED: actionary					
[CartageEmironment] NEW AGENT (OINEO) knowledger Serent scream Real police stateset					
Speech recognition engine started	Bc				
🖉 Clean 🚦 Stop 🕪 Pause 📲 🚽 🖓 Management Agent 🌋 JAUE Sniffer					
	Nin Mexicate				







LIGHTING NEEDS TO BE REBUILT (145 unbuilt "DisableAllScreenMessages" to supp







Human-Robot Teaming Interaction improve the level of trust from humans to robots in a collaborative task

- manage collaborative tasks
- manage shared knowledge on environment and objectives
- decision making on the base of the other actions
- explainability and trustfulness





A twofold approach

- Methodological and implementation
 - design abstractions for being trustful and explainable
 - design abstractions for decision process, anticipation and knowledge
- Implementation
 - the question is: which are the paradigms or technical issues allowing to pass from the design to the code and the working system?



Methodological Perspective



Environment



UNIVERSITÀ DEGLI STUDI DI PALERMO



ENVIRONMENT

Valeria Seidita - ITAL-IA 2013

Boing towards the implementation

AGENT



1.	$B \leftarrow B_0$; /* B_0 are initial beliefs */
2.	$I \leftarrow I_0$; /* I_0 are initial intentions */
3.	while true do
	get next percept $ ho$ via sensors;
5.	$B \leftarrow brf(B, \rho);$
6.	$D \leftarrow options(B, I);$
7.	$I \leftarrow filter(B, D, I);$
8.	$\pi \leftarrow plan(B,I,Ac);$ /* Ac is the set of actions */
9.	while not $(empty(\pi) \text{ or } succeeded(I,B) \text{ or } impossible(I,B))$ do
 10.	$\alpha \leftarrow \text{first element of } \pi;$
11.	$execute(\alpha);$
12.	$\pi \leftarrow $ tail of π ;
13.	observe environment to get next percept $ ho$;
14.	$B \leftarrow brf(B, \rho);$
15.	if <i>reconsider(I,B</i>) then
16.	$D \leftarrow options(B, I);$
17.	$I \leftarrow filter(B, D, I);$
18.	end-if
19.	if not $sound(\pi, I, B)$ then
20.	$\pi \leftarrow plan(B, I, Ac)$
21.	end-if
22.	end-while
23.	end-while

....

$A_c \leftarrow action(B_{\alpha_i}, Cap)$

$evaluate(\alpha_i)$ $J \leftarrow justify(\alpha_i, B_{\alpha_i})$

Self-Modeling **Trustful interaction**





Towards the implementation - a proposal







Extending practical reasoning with emergent inner speech

foreach α_i do $evaluate(\alpha_i);$ $R \leftarrow rehearsal(\alpha_i, B_{\alpha_i}, D);$ update(B,D); $J \leftarrow justify(\alpha_i, B_{\alpha_i});$

end

action object







- A concept used and developed in psychology
- Self-awareness and self-consciousness
 - the ability to become the object of one's attention
 - mental states
- The subjective experience of language in relation to one's action feeling and experiences
- Important role in self-regulation, self-direction, problem solving, planning and memory

Valeria Seidita - ITAL-IA 2013

Inner Speech





Technological Perspective

- JaCaMo
- At runtime a BDI agent
 - Belief, Plan, Intention, Event, Action
 - Selection Functions: S_E, S_O, S_I
- The idea is to link internal events with **speech acts** to generate explanation

Valeria Seidita - ITAL-IA 2013

BDI paradigm and JASON - the obvious implementation counterpart





Extending practical reasoning with inner speech

- In MASs communication is based on the theory of speech acts
- Inner speech is a way of thinking about oneself -> under certain condition a speech act can address the agent itself
 - Plan
 - actions to achieve team goals
 - message for changing beliefs or goals
 - making the message available outside



Speech act and setting a table

Agent	Responsibility
pepper	Leader: it initiates the activities to set the table right place to see if the goal has been achieve
nao	It waits for a message from the leader to st cooperate with him;
Workspace	Description
room	This is the predefined workspace provided by in setting the table by pointing at the cointain
Artefact	Description
blackboard	It contains the list of tasks to be done by the table;
table	It contains predefined positions in the form o simple actions, such as placing or picking up
	simple actions, saon as placing of ploining ap

Valeria Seidita - ITAL-IA 2013

UNIVERSITÀ DEGLI STUDI DI PALERMO

ole and, when all the tasks are done, checks that all the objects are in the ved;

start performing the tasks indicated on the blackboard artifact and to

by CArtAgO and represents the space where the two agents are involved iner;

agents and information about the correct position of the objects on the

of labels where objects can be placed. Designed for agents that perform p an object;

container for the various items placed on the table by the agents;



ADIS STUTUE

common

Cartago

nao

pepper

CArtAgO Http Server running on http://192.168.1.226:3273 Jason Http Server running on http://192.168.1.226:3272 [Cartago] Workspace ws created. [Cartago] artifact cabinet: table.Cabinet([fork, dish, knife, glass]) at ws created. [Cartago] artifact table: table.Table([f, d, k, g]) at ws created. [Cartago] artifact board: table.BlackBoard([fork, dish, knife, glass],[f, d, k, g]) at ws created. [pepper] join workspace /main/ws: done [pepper] focusing on artifact board (at workspace /main/ws) using namespace default [nao] join workspace /main/ws: done [nao] focusing on artifact board (at workspace /main/ws) using namespace default [nao] focus on board: done [nao] focusing on artifact table (at workspace /main/ws) using namespace default [nao] focus on table: done [nao] focusing on artifact cabinet (at workspace /main/ws) using namespace default [nao] focus on cabinet: done [nao] Hello, my name is nao! [pepper] focus on board: done [pepper] focusing on artifact table (at workspace /main/ws) using namespace default [pepper] focus on table: done [pepper] focusing on artifact cabinet (at workspace /main/ws) using namespace default [pepper] focus on cabinet: done [pepper] Hello, my name is pepper!

🜗 Continue

Stop

🕖 Clean





游 Debug

💢 Kill agent







NONNA HATTOIS SLATIS

Ú	JaCaMoLaunc	her 🖸 74% 🔲 🗢 🛜 🔳 Q		Dom 28 mag
	•	:: Mind Inspector ::		
pepper	Inspection	of agent pepper (cycle #48)		
	- Beliefs	<pre>contains(glass)_{[artifact_id(cobj_6),artifact_name(cabinet),percept_type(obs_prop),source(percept),workspace("/main/ws",cobj_2)] contains(knife)_{[artifact_id(cobj_6),artifact_name(cabinet),percept_type(obs_prop),source(percept),workspace("/main/ws",cobj_2)] contains(dish)_{[artifact_id(cobj_6),artifact_name(cabinet),percept_type(obs_prop),source(percept),workspace("/main/ws",cobj_2)] contains(fork)_{[artifact_id(cobj_6),artifact_name(cabinet),percept_type(obs_prop),source(percept),workspace("/main/ws",cobj_2)] correct_position(Item,Label)_[source(self)]:- (label(Item,CorrectLabel) & (Label == CorrectLabel))}}}}</pre>		
		<pre>focusing(cobj_6,cabinet,"table.Cabinet",cobj_2,ws,"/main/ws")_{[artifact_id(cobj_4),artifact_name(body_pepper),percept_type(obs_prop),source(percept),workspace("/main/ws",cobj_2)] focusing(cobj_5,table,"table.Table",cobj_2,ws,"/main/ws")_{[artifact_id(cobj_4),artifact_name(body_pepper),percept_type(obs_prop),source(percept),workspace("/main/ws",cobj_2)] focusing(cobj_3,board,"table.BlackBoard",cobj_2,ws,"/main/ws")_{[artifact_id(cobj_4),artifact_name(body_pepper),percept_type(obs_prop),source(percept),workspace("/main/ws",cobj_2)] hand_busy_[source(self)]:- (.count(in_hand(Item)_[inn_sp],N) & (N == 2)) hand_free_[source(self)]:-}}}</pre>		
	Agent History mind 0			fr
comr	[pepper] focus on [pepper] focusing [pepper] focus on	on artifact cabinet (at workspace /main/ws) using namespace default cabinet: done		
		📿 Clean 📕 Stan 🕪 Cantinua 🌟 Dabur 📽 New agent 🛸 New DEDL agent 🚍 Sources		





UNIVERSITÀ DEGLI STUDI DI PALERMO

Conclusions and Remarks

- We are working on a possible solution to endow agents (robots and intelligent systems) with the ability to explain its actions
 - creating agents that are reliable, explainable and believable
- The key is the deliberation made before taking an action
- We experimented the BDI technology and Jason/JaCa for validating our approach
 - simplicity in handling abstractions
 - significant effort to learn and use Jason for not skilled users
- Abstractions and requirements need to be carefully analysed and transformed into Jason elements





Conclusions and Remarks

- Need for accuracy in the agent knowledge representation
- Our last experiment is on a very simple scenario
 - in the immediate future -> connect actions to beliefs
- Only the part of the work related to the explainability has been presented
 - decision making process will use the same logic





Conclusions and Remarks

- Further validating the approach in a more complex scenario and using more agents
- We will refine the methodological approach
- We will also add ethical aspects to the agents' knowledge
 - to analyse how ethical reasoning may influence the human



Thanks for your attention valeria.seidita@unipa.it



