

Estudi de les funcions de confiança entre agents amb estructures i objectius cooperatius

Ferran González Barberà

Enginyeria Superior en Informàtica
Intel·ligència Artificial

Professor Consultor

David Isern Alarcón

Professor Responsable Àrea

Carles Ventura Arroyo

2 de Gener de 2018



Aquesta obra està subjecta a una llicència de [Reconeixement-NoComercial-CompartirIgual 3.0 Espanya de Creative Commons](https://creativecommons.org/licenses/by-nc-sa/3.0/es/)

Apartat de llicències

A) Respecte a la llicència de Creative Commons:

Aquesta llicència permet a qualsevol persona mesclar, adaptar i construir a partir d'aquesta obra sense finalitat comercial, sempre que se'n reconegui l'autoria i es mantingui la llicència en les noves creacions.

B) GNU Free Documentation License (GNU FDL)

Copyright © ANY Ferran González Barberà.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

C) Copyright

© (Ferran González Barberà)

Reservats tots els drets. Està prohibit la reproducció total o parcial d'aquesta obra per qualsevol mitjà o procediment, compresos la impressió, la reprografia, el microfilm, el tractament informàtic o qualsevol altre sistema, així com la distribució d'exemplars mitjançant lloguer i préstec, sense l'autorització escrita de l'autor o dels límits que autoritzi la Llei de Propietat Intel·lectual.

FITXA DEL TREBALL FINAL

Títol del treball:	<i>Estudi de les funcions de confiança entre agents amb objectius i estructures cooperatius.</i>
Nom de l'autor:	<i>Ferran Gonzalez Barberà</i>
Nom del consultor/a:	<i>David Isern Alarcón</i>
Nom del PRA:	<i>Carles Ventura Arroyo</i>
Data de lliurament (mm/aaaa):	<i>01/2018</i>
Titulació o programa:	<i>Enginyeria Superior en Informàtica.</i>
Àrea del Treball Final:	<i>Intel·ligència Artificial.</i>
Idioma del treball:	<i>Català</i>
Paraules clau	<i>Confiança, Reputació, Col·laboració</i>

Resum del Treball (màxim 250 paraules): *Amb la finalitat, context d'aplicació, metodologia, resultats i conclusions del treball*

Aquest treball acadèmic té l'objectiu d'aportar una visió conceptual, aclaridora i de resum sobre les interaccions basades en la confiança entre agents cooperatius. Tanmateix identifica els principals mecanismes més utilitzats en la bibliografia per tal de fer avaluacions de confiança i gestions en les consultes de reputació. També presenta dues solucions evolutives i repassa diverses oportunitats de desenvolupament en el camp de les relacions de confiança entre agents.

La metodologia seguida ha estat inicialment la d'analitzar la situació actual a través de la bibliografia específica, per seguidament estructurar els coneixements obtinguts i presentar-los de la manera més clara possible. Per aquest motiu aquest document no pretén tenir una màxima profunditat, donat que es perdria així la visió global que tracta d'aportar. Aquest estudi ofereix la possibilitat d'expandir coneixements en cadascun dels seus capítols mencionant les principals referències trobades en els seus respectius camps de coneixement.

En quant a la part evolutiva d'aquest document, la intenció ha estat també la d'obtenir i presentar aquelles àrees en les que hi ha marge de desenvolupament. Així es presenten dues solucions teòriques però concretes: un sistema de pissarra adaptat a l'ús d'agents cooperatius basats en la confiança, i un protocol de negociació de recompenses que possibilita als

agents obtenir millors marges de benefici i alhora obrir més el seu camp d'agents objectiu.

El resultat d'aquest treball és un document que aporta visibilitat multi-disciplinària i de mitjana profunditat sobre els progressos fets en l'àrea de coneixement de les interaccions entre agents basades en la confiança, obrint noves oportunitats d'investigació a futurs estudis.

Abstract (in English, 250 words or less):

This academic work intends to bring a conceptual, summarizing and clarifying point of view over a specific area of knowledge of Artificial Intelligence: trust-based interactions between or amongst collaborative agents. It also intends to identify the main and most used mechanisms in the available bibliography concerning calculations of trust and reputation management. Two evolutive solutions are also presented, and several areas of open development are described.

The applied methodology has initially been of analysing the current situation through the published academic works, organising then all acquired knowledge and presenting them on the most clarifying manner. For this reason this document does not intend to obtain a deep depth vision of this field, as the general view would be compromised. This work offers the reader the possibility of expanding the knowledge of all chapters by referring to relevant key documents.

Concerning the evolutive part of this work, the intention has also been the one of gathering and presenting those areas where there is a clear margin for future development and investigation. Therefore two theoretical but specific solutions are proposed: a blackboard system adapted to the use of cooperative agents based on trust relationships, and a negotiation protocol through which autonomous agents can obtain better margins of profit when outsourcing their tasks, and enhancing their possibilities when reaching a defined number of third agents.

The result of this work is a document that gives cross-functional and mid-depth visibility over the progress made in the area of knowledge of trust based agent interactions, opening new oportunities of investigation for upcoming studies.

Índex

1. Introducció.....	7
1.1 Context i justificació del Treball	7
1.2 Objectius del Treball.....	7
1.3 Enfocament i mètode seguit.....	8
1.4 Planificació del Treball.....	8
1.5 Breu sumari de productes obtinguts	9
1.6 Breu descripció dels altres capítols de la memòria	9
2. Estudi de les funcions de confiança entre agents amb estructures i objectius cooperatius.....	10
2.1 - Anàlisi de la situació actual (Fase I)	10
2.1.1 - Introducció i context	10
2.1.2 - Els sistemes multi-agent com a entorn de treball (MAS).....	11
2.1.3 - Agents autònoms pre-cooperatius.....	11
2.1.4 - La conveniència i el risc de confiar.....	12
2.1.5 - Introducció a la col·laboració entre agents	13
2.1.6 – Models de confiança i reputació	14
2.1.7 - Factors d'interès sobre els models de confiança i reputació	17
2.1.8 - Relacions i coalicions com a fruit dels jocs de confiança	19
2.2 - Redefinició conceptual i arquitectònica (Fase II)	20
2.2.1 - Introducció i context	20
2.2.2 - Propietats dels agents i sistemes multi-agent orientats a la col·laboració	21
2.2.3 - Capacitats i mecanismes per a agents i sistemes multi-agent orientats a la col·laboració.	23
2.2.4 – Arquitectura general del parell Agent i Sistema Multi-Agent.....	28
2.3 - Proposta evolutiva (Fase III)	30
2.3.1 – Propostes evolutives	31
2.3.1.1 – Protocol de negociació per a jocs de confiança cooperatius	31
2.3.1.2 – Sistemes de pissarra per a agents cooperatius.....	34
2.3.2 – Àrees obertes de recerca	38
3. Conclusions.....	40
4. Glossari	43
5. Bibliografia.....	45

Llista de figures

Figura 1: Visió completa arquitectònica d'un agent cooperatiu

Figura 2: Sistema de pissarra per a agents cooperatius

1. Introducció

1.1 Context i justificació del Treball

L'àrea de coneixement d'aquest estudi és la dels mètodes d'anàlisi de confiança i de decisió en sistemes multi-agent, en que els agents autònoms es relacionen amb altres agents o amb humans. En concret, l'estudi es focalitza en aquelles tècniques d'anàlisi de confiança i decisió que utilitzen estructures organitzatives [14] i aquelles que tenen fins cooperatius.

La majoria dels estudis i coneixements recents prenen com a criteri el propi interès i benefici del propi agent que ha de resoldre el problema. Això condueix a una limitació dels beneficis que es poden obtenir. Les tècniques organitzatives i cooperatives permeten la creació de xarxes de confiança i cooperació que permeten uns rendiments superiors i un més fàcil accés als beneficis individuals a partir de l'especialització de funcions i la cooperació entre agents.

1.2 Objectius del Treball

En un primer pas, aquest treball tracta de fer un recorregut teòric de tots els conceptes que giren entorn a la confiança entre agents autònoms, així com de les estructures i solució col·laboratives trobades en la bibliografia. Donat que es tracta d'un camp de coneixement extens i que compta amb nombrosos documents científics publicats, es prioritza tractar d'aportar una visió global per sobre d'aprofundir en cap de les categories citades. El segon capítol d'aquest document té un objectiu similar al primer, però estructurant les diferents característiques, mecanismes i solució sovint implementades per a la col·laboració entre agents autònoms.

En el capítol final, es proposen evolucions de mètodes de càlcul de confiança sobre els existents, intentant obrir noves vies d'investigació en l'àmbit dels jocs multi-atribut, amb recompensa per col·laboració, i amb suports organitzatius per a la millora de rendiments de les seves funcions.

1.3 Enfocament i mètode seguit

L'estratègia d'execució d'aquest treball s'ha traduït directament en el disseny i estructura de l'obra en tres apartats. En el primer apartat s'intenta fer, per començar, un resum progressiu dels conceptes entorn als agents, el seu entorn i els diferents tipus d'estratègia davant la necessitat de col·laborar. En un segon apartat, i utilitzant els coneixements adquirits en el primer, es tracta de desgranar quins mecanismes de col·laboració són els més efectius per a la construcció d'un agent amb òptimes capacitats cooperatives. En l'últim apartat, es fan diferents propostes evolutives per tal de dotar als agents autònoms de capacitats exteses per a la col·laboració i coordinació amb altres agents.

És clau remarcar que aquest document té un enfocament eminentment teòric. Pretèn donar una visió transversal i global sobre l'àrea de coneixement de la confiança entre agents, així com certes propostes evolutives. Les aplicacions i proves pràctiques dels mecanismes esmentats es deixen com a propostes de futures línies d'investigació.

1.4 Planificació del Treball

Part I - Anàlisi de context de la situació actual (Fase I)

I.a- Introducció i context.

I.b- Agents autònoms pre-cooperatius.

I.c- La conveniència i el risc de confiar.

I.d- Introducció a la col·laboració entre agents.

I.e- Models de confiança i reputació.

I.e- Factors d'interès sobre els models de confiança i reputació.

I.f- Relacions i coalicions.

Part II - Redefinició conceptual i arquitectònica (Fase II)

II.a- Introducció i context.

II.b- Propietats dels agents i sistemes multi-agent orientats a la col·laboració.

III.c- Capacitats i mecanismes per a agents i sistemes multi-agent orientats a la col·laboració.

III.d- Arquitectura general del parell Agent i Multi-Agent.

Part III. Proposta evolutiva

III.a- Propostes evolutives

III.b – Àrees obertes de recerca

1.5 Breu sumari de productes obtinguts

D'acord amb els objectius definits en aquest treball acadèmic, els productes entregats són els següents:

- Resum teòric conceptual del camp de coneixement de la Confiança entre Agents Autònoms, i Esquema Arquitectònic de funcions de Confiança entre Agents. Aquesta part es correspon als apartats 2.1 i 2.2.
- Proposta evolutiva i àrees de recerca sobre la mateixa àrea de coneixement. Aquesta part es correspon a l'apartat 2.3.

1.6 Breu descripció dels altres capítols de la memòria

Capítol 1: dedicat a fer un repàs exhaustiu dels conceptes que giren entorn als agents i sistemes multi-agent, per tal de donar una visió completa d'entorn i de les problemàtiques que s'hi presenten. Aquest capítol té una orientació conceptual i didàctica, que permet al lector obtenir un coneixement de les complexitats d'aquesta branca de la Intel·ligència Artificial. Cada subapartat aporta bibliografia adicional per a que el lector en pugui estendre el coneixement.

Capítol 2: orientat a descomposar i explicar les propietats, funcions i solució proposades en la bibliografia existent per tal que els agents amb funcions col·laboratives puguin interactuar entre sí.

Capítol 3: propostes de caire evolutiu aprofitant les carències detectades en l'anàlisi fet durant els capítols anteriors. Les solució d'aquest apartat són de caire teòric, pel que la programació i el testeig de les mateixes queden fora de l'abast d'aquest treball.

2. Estudi de les funcions de confiança entre agents amb estructures i objectius cooperatius

2.1 - Anàlisi de la situació actual (Fase I)

En aquest primer capítol el treball gira entorn als conceptes i bases dels mecanismes de confiança entre agents autònoms. Després d'una introducció i definició del context de treball, es tractaran punts clau com els principis de confiança i conveniència, els models de càlculs de confiança, i la formació de relacions i coalicions.

2.1.1 - Introducció i context

Aquesta anàlisi pretén explicar d'una manera planera i entenedora la necessitat i conveniència de l'existència de mecanismes de confiança en entorns oberts amb múltiples agents. També té com a objectiu l'anàlisi de mecanismes de confiança existents, a través de diverses classificacions que ens portaran a entendre les finalitats a les que cadascun d'ells responen. En l'apartat final d'aquest treball es presenta una solució evolutiva i complementària als mètodes d'avaluació de confiança ja existents.

Per tal de calibrar adequadament les expectatives del lector d'aquest document acadèmic, cal notar des d'un principi que aquest estudi, de caire eminentment teòric, barreja sovint anàlisis de diferents branques del coneixement, tan diferents com són les teories de jocs, la lògica, la psicologia social o el disseny de mecanismes econòmics [1]. Tots ells s'han pres en consideració sempre des d'una perspectiva informàtica, centrada en l'àrea de coneixement de la intel·ligència artificial. Igual que succeeix en ciències com l'aeronàutica o la biotecnologia, les solucions adoptades per les formes de vida conegudes serveixen de referència en l'estudi de les relacions multi agent. L'extrapolació i adaptació dels mètodes de confiança i reputació que ens són naturals als humans són operacions constants en la majoria d'anàlisi d'aquest camp de coneixement. Com es podrà observar en la part conclusiva d'aquest document, l'evolució proposada es basa justament en mètodes que les persones emprem sovint per avaluar el grau de confiança que altres humans ens generen.

Continuant amb aquest primer objectiu de centrar les expectatives del lector, cal aclarir que aquest estudi ni és ni pretén ser un catàleg ni una

classificació exhaustiva del mètodes de confiança proposats per altres autors. Amb aquesta finalitat s'han publicat complets estudis com el de Yu, Shen, Leung, Miao i Lesser [2].

2.1.2 - Els sistemes multi-agent com a entorn de treball (MAS)

Com a punt de partida cal definir primerament l'entorn en el que treballen els agents que són objecte d'aquest estudi. Els agents els comportaments dels quals seran estudiats es troben en entorns que anomenarem Entorns Multi-Agent (de l'acrònim MAS, "multi-agent systems"), i tenen les següents propietats:

Arbitraris: perquè les intencions dels altres agents són desconegudes.

Oberts: perquè s'hi poden incloure o excloure agents de manera lliure i no limitada.

Diversos: perquè els agents poden ser de molt diversa mena i procedència, considerant inclús interaccions humanes com agents independents.

Els entorns multi-agent poden ser en sí mateixos de diversos tipus: entorns d'arquitectura semàntica (com la "semantic web"), entorns d'interacció d'igual a igual ("peer to peer") o entorns creats específicament per a la interacció entre agents ("com els smart grids"), entre d'altres. Cadascun d'aquests entorns proporciona unes especificacions i regles de joc que els agents han de complir i respectar, però també eines per a la comunicació (protocols) i interacció (estructures de reputació, per exemple).

2.1.3 - Agents autònoms pre-cooperatius

La major part dels agents que interactuen amb tercers estan configurats i determinats per a aconseguir els seus objectius buscant únicament el benefici propi. Això equival a dir que en totes les interaccions que es produeixin, l'únic criteri a tenir en compte serà el d'obtenir el millor resultat possible al problema individual de cada agent. En altres paraules, equival a dir que la força que dirigeix les interaccions és l'interès propi per aconseguir el millor resultat possible al problema propi de cada agent.

Aquesta base de programació en els agents té certes limitacions, derivades bàsicament de dos factors:

- Manca d'un objectiu comú: cada agent mira de resoldre la seva tasca lo abans possible, sense tenir en compte factors com el

rendiment del conjunt, l'estabilitat del propi MAS o la consistència dels resultats amb altres agents.

- Estratègia de curt termini: donat que les relacions que s'estableixen són purament situacionals -per resoldre problemes concrets i puntuals-, no es tenen en compte interessos futurs que poden aportar beneficis més enllà del proper problema a resoldre.

2.1.4 - La conveniència i el risc de confiar

Els diferents objectius assignats als agents en qüestió sovint no poden ser assolits sense la interacció amb altres agents. Donat que la informació que els agents necessiten està sovint distribuïda i és heterogènia, en la majoria dels casos precisen de col·laboracions per a fer tasques de processament, d'investigació i de recaptació d'informació. També existeix un factor d'especialització dels agents, pel que sovint és més convenient, i en ocasions la única opció factible, la col·laboració entre agents per assolir els objectius buscats. Aquestes seran precisament les necessitats que portaran als agents a interactuar amb agents tercers amb un grau relatiu de confiança.

El fet de confiar tasques a un agent tercer implicarà, en qualsevol cas, un cert grau d'incertesa i en conseqüència un risc inherent. Que una part de la informació requerida o processada provingui d'un agent tercer implica un cert nivell de desconeixement sobre el seu origen o la seva fiabilitat. Alhora, un agent que sol·licita els serveis d'un altre tampoc té visibilitat ni pot preveure la qualitat dels resultats que rebrà. Per últim, una relació, encara que sigui puntual, comportarà una certa perda de control sobre el temps de finalització de la tasca, per més que s'hagi acordat un temps d'entrega.

Tot aquest nivell de desconeixement haurà de ser entès com un risc, i ser conseqüentment avaluat amb l'objectiu d'escollir i decidir quines interaccions, d'entre totes les disponibles, seran convenients o hauran de ser desestimades. Justament perquè confiar comportarà sempre un cert nivell de risc, la confiança en cooperacions no només haurà de ser gestionada apropiadament, sino també justificada [4].

És important aclarir que les avaluacions de confiança que es vagin succeïnt i les possibles interaccions que en resultin seran inicialment situacionals. Això es deu a que l'avaluador i l'avaluat podran intercanviar els rols si les necessitats s'inverteixen. A més, factors com la disponibilitat immediata de cada agent o altres que es detallaran posteriorment podran

afectar l'avaluació de confiança en cada moment, generant resultats diferents en moments diferents.

2.1.5 - Introducció a la col·laboració entre agents

Com a punt de partida per a l'anàlisi de capacitats col·laboratives, els humans poden tornar a ésser considerats com a referència (en la mateixa mesura que en la gestió de conflictes, fora de l'abast d'aquest document). Els humans ténen una capacitat inherent per al treball en equip, tant per a l'increment de l'eficiència del conjunt com per a acomplir tasques complexes. El treball en equip i la coordinació són capacitats essencials a incorporar en els agents autònoms si es pretèn que aquests puguin resoldre problemes complexos i a sobre, ho facin d'una manera eficient. Les capacitats col·laboratives esdevenen encara més necessàries quan l'entorn és ràpidament canviant [8].

Com es determina en [13], tant en el supòsit que els agents tenen una tasca comuna, com en el cas que un agent sol·licita col·laboració a un altre, cadascuna de les seves accions es basarà en un cert nivell d'assumpció respecte a que l'altre agent:

- té la capacitat de portar a terme les tasques demanades.
- té la intenció o el compromís a executar les tasques sol·licitades i retornar els resultats obtinguts.

I afegides a les dos creences explicades en [13], se'n podria afegir:

- té la capacitat d'entregar resultats objectivats, o no afectats per interessos particulars.

Mentre ens mantenim dins de l'estudi dels agents amb interessos únicament individuals (anomenats "selfish" en llengua anglesa), és una obvietat dir que aquests, per tal de minimitzar el risc de les seves interaccions, hauran d'escollir aquelles en que el risc sigui conclòs com el menor possible per un determinat tipus de tasca i en un moment també determinat. Aquesta es configura com una base de consens que sustenta tots els jocs de decisió d'agents amb interessos únicament individuals.

És en aquest punt en el que entren els models i jocs de confiança ("trust games"), que passen a ser explicats en el següent capítol.

2.1.6 – Models de confiança i reputació

Els models de confiança i reputació són mètodes d'obtenció d'informació i de càlcul que serveixen per a l'avaluació del nivell de confiança en agents tercers per a determinades funcions i entorns. És a partir d'aquests models que es dissenyen i implementen els Sistemes de Confiança i Reputació (o TRS, abreviatura en anglès de "Trust and Reputation Systems" [25]). D'aquests algorismes i de les seves classificacions se n'han publicat molts i diversos.

A continuació es presenta una classificació dels diferents tipus de sistemes de càlcul de la confiança amb els seus avantatges i inconvenients:

- a) Mètodes d'avaluació directa: a través de l'observació dels resultats d'interaccions passades en que l'agent avaluador ha estat implicat en primera persona.

Avantatges:

- la informació és de disponibilitat immediata, donat que només cal consultar els propis registres d'informació. És a dir, que l'agent avaluador no depen de cap recurs extern per a obtenir aquestes dades.
- no hi ha possibilitat que les dades estiguin subjectivades o afectades per interessos externs.
- són informacions que es poden obtenir sense un cost afegit. Cal recordar que els registres de reputació poden tenir un cost associat per a l'agent que els consulta.
- de les interaccions passades se'n podran obtenir generalment valors de confiança acurats.

Inconvenients

- l'agent avaluador només tindrà informació respecte als agents amb els que hagi tingut relació, pel que ens trobem davant d'un registre limitat d'informació. En entorns amb molta rotació d'agents aquest mètode podrà resultar poc efectiu.
- si les experiències que es consultin en els registres propis no són recents, la fiabilitat de les estimacions de confiança que en resultin podrà ser en cert grau inversament proporcional a l'antiguitat de les mateixes. En el cas dels mètodes d'avaluació directa, l'impacte d'aquest inconvenient es podrà controlar perquè també es disposarà de la data de les experiències de referència.

- b) Models d'avaluació indirecta: l'agent interessat en conèixer la confiança que pot depositar en un segon, simplement consulta altres agents per obtenir dades de l'agent avaluat. A aquests mètodes se'ls sol anomenar de reputació.

Avantatges:

- l'abast dels agents sobre els que obtenir informació serà més extens que en els mètodes d'avaluació directa, doncs es podran consultar tants agents com hi hagi en el mateix entorn o en la mateixa coalició.
- Teòricament els valors de reputació transmesos seran acurats, perquè respondran a les pròpies experiències dels agents consultats respecte a agents tercers.

Inconvenients:

- les dades de reputació cedides per agents tercers poden tenir un cert grau de subjectivació. Els agents que cedixin valors de reputació poden tenir interessos que donin valors sobre o infradimensionats de reputació.
- els valors de reputació transmesos podran tenir una certa antiguitat, que a sobre l'agent avaluador probablement no coneixerà. La utilitat d'aquests valors de reputació decreixerà si estan basats en experiències o interaccions antigues, sense que l'agent avaluador possiblement ho sàpiga.

- c) models d'avaluació socio-cognitius: aquests mètodes de càlcul es basen en presumpcions i aproximacions aplicats sobre les dades que l'agent avaluador pot aconseguir en primera persona.

Avantatges i aplicabilitat

- a través d'aquests mètodes es poden deduir valors de confiança desconeguts a partir d'informacions conegudes, pel que es dona un ús estès a les dades disponibles.
- les estimacions de confiança que es facin per aquests mètodes són independents respecte a altres agents o estructures, de manera que no poden ser subjectivades o afectades per interessos de tercers.
- Si les informacions que s'usen en aquests mètodes són actuals, les estimacions de confiança que en resultin reflectiran la situació també actual. En altres paraules, s'avaluaran els agents de cooperació només sobre característiques i dades únicament actualitzades.
- Aquests mètodes són especialment convenients en entorns amb alta volatilitat i dinamisme, on una alta rotació d'agents no permeti obtenir informacions d'experiències passades.

Inconvenients

- la fiabilitat dels càlculs de confiança fets per aquest mètodes pot ser diferent depenent de l'entorn i el nivell d'especialització dels agents analitzats.
- d) Mètodes suportats per eines organitzatives: en aquest cas l'agent avaluador consulta estructures o fonts especialment dedicades al registre de reputacions. En aquest camp es solen usar certificats de confiança emesos per les entitats registradores.

Aventatges

- Es pot assumir que els valors de reputació obtinguts tindran una fiabilitat en correspondència a l'entitat certificadora dels mateixos. En altres paraules, l'agent avaluador tindrà informació força precisa de la fiabilitat dels valors de reputació si també té en compte la reputació de la pròpia entitat de certificació.
- Donat que s'obtindran els valors de reputació d'estructures centrals, aquests valors estaran probablement basats en mostres d'experiències més grans que les d'avaluació directa o indirecta. Per tant en teoria aquests valors de reputació seran també més acurats.

Inconvenients

- Les entitats emissores de certificats de reputació hauran estat creades per corporacions o governs amb possibles interessos comercials o polítics. Per tant hi haurà un cert risc de subjectivació de les dades a tenir en compte.
- Els certificats de reputació poden tenir un cost associat. Si l'agent avaluador l'afegeix als costos estimats per a les relacions de confiança, podrà succeir que prioritzi aquells agents ja coneguts, reduint així la diversitat i la renovació dels agents amb els que col·laborar.

Una altra forma de copsar la diversitat de conceptes sobre els quals es defineixen els mètodes d'avaluació de confiança, és analitzant-ne les seves dimensions. La llista següent parteix de la que va ser publicada per Sabater i Sierra [4] i que he tractat d'explicar tal i com segueix:

- Tipus de paradigma: numèric o cognitiu.
- Segons la font d'informació: experiència directa, informació de testimonis, informació sociològica (estudiant les relacions entre altres agents), prejudicis.
- Visibilitat: Sistemes d'informació i repositoris globals (“reputation systems”), o registres privats (subjectius).
- Especificitat: com de generals o específics son els mètodes respecte a l'entorn de treball i el problema a resoldre.
- Robustesa respecte a l'entorn: com de fiable és una arquitectura respecte a diferents entorns.
- Segons el tipus d'informació intercanviada: boleana, discreta, continua, o estructures de coneixement.
- Segons el nivell d'orientació a la col·laboració.
- Preparació per a la resistència enfront als danys de reputació i altres interessos. És a dir, si el disseny del mètode d'avaluació contempla mecanismes per contrarrestar possibles danys en la reputació de l'agent.

2.1.7 - Factors d'interès sobre els models de confiança i reputació

Amb la finalitat de donar completesa a la perspectiva actual sobre els models d'interacció entre agents cooperatius, hi ha alguns problemes i factors que mereixen especial atenció. Aquests en són alguns:

- Els jocs de confiança: es tracta de models matemàtics que proven d'anticipar en quina manera els agents interactuaran entre sí per a aconseguir els objectius buscats. Aquests models es centren bàsicament en el càlcul i repartiment de dos valors: la recompensa i el cost de les relacions de confiança. A més afegeixen en les equacions conceptes de sinergia. Hi ha dos categories principals de jocs de confiança: els jocs competitius i els jocs cooperatius [6 i 24].
- Normatives i protocols en MAS: els diferents mecanismes d'avaluació poden veure's limitats per les normes establertes en cada sistema multi-agent. Cal recordar que cada MAS pot establir unes regles de joc que poden, per exemple, acotar o definir els

protocols d'interacció entre els agents. Dins d'aquests factors hi encabiríem també els protocols de comunicació: el llenguatge o llenguatges establerts per a l'ús dins un sistema multi-agent tindran un cert impacte en les pròpies relacions entre els agents, i inclús en les dades que es comparteixin.

- Estructures i agents especialitzats: un factor que pot tenir un gran impacte sobre el rendiment i els resultats dels jocs de confiança són els registres de reputació existents i proporcionats per estructures externes. Aquests impactes poden ser tan positius – aportant informació contrastada sobre reputació – com negatius – si les dades de reputació han estat subjectivades atenent a interessos no explicats. A aquests agents independents que suposadament són fiables i objectius, se'ls anomena sovint “Escrow agents” [14].
- Els jocs de congestió: una altra variable que pot impactar en els càlculs de confiança és el de la càrrega o congestió de l'agent avaluat. Dit en altres paraules, els càlculs de confiança i els registres de reputació sovint no tenen en compte la disponibilitat efectiva de cada agent consultat. També és possible que l'agent avaluat sigui humà, i que per tant la seva disponibilitat sigui parcial i la seva capacitat de resposta limitada en el temps.
- Interessos i subjectivació de les dades i resultats compartits entre agents: igual que succeeix entre els éssers humans i els nostres mitjans de comunicació, les dades intercanviades entre agents són susceptibles de tenir un cert nivell de biaix degut a interessos d'organitzacions públiques o privades.
- Estabilitat de l'entorn MAS i consistència de les observacions: els resultats que els agents puguin obtenir de les seves col·laboracions i dels valors de confiança que calculin es podran veure afectats per característiques inherents al propi entorn MAS: el nivell de rotació dels agents o l'estabilitat de les estructures organitzatives disponibles (p.e. registres de reputació).

2.1.8 - Relacions i coalicions com a fruit dels jocs de confiança

Fins ara s'han exposat les bases que fonamenten els models de confiança i reputació. Aquests models, un cop posats en pràctica, es traduiran en interaccions que tindran un major o menor nivell d'èxit, i que per tant faran que els agents tinguin més o menys probabilitat de tornar a interactuar entre sí o amb altres. Les diferents interaccions que es vagin produïnt deixaran registres a mode d'experiències, i amb elles certs càlculs de confiança i possiblement de reputació (si arriben a transmetre's a altres agents o a registres de certificació). Segons la teoria de jocs cooperatius, ben diferent a la dels jocs de confiança de benefici individual, es promourà que aquestes experiències es converteixin en relacions (entre dos agents) o en coalicions (múltiples agents). Amb aquesta finalitat en els jocs cooperatius s'afegeix un valor que no existeix en els jocs competitius: la recompensa per col·laboració. En aquest jocs ja no s'hi val només a aconseguir el millor resultat possible, sinó que també cal mirar d'establir relacions de cooperació que poden servir per a objectius a mitjà o llarg termini.

Relacions

Les relacions entre agents, igual que succeeix en els éssers humans, són el resultat de la interacció i la cooperació en situacions d'interés o necessitat entre dos agents. En certa manera les relacions es poden traduir en registres d'experiències o interaccions puntuals passades, és a dir, en informació històrica en primera persona de com les interaccions entre dos agents concrets van resultar (recordem els mecanismes d'avaluació directa). Si aquestes varen resultar positives per a ambdós agents, la probabilitat que es repeteixin serà més alta. Si aquesta tendència es creixent, els agents aniran interactuant sovint per conveniència mútua, dins d'un esquema que podrem anomenar relació.-

Coalicions

A través d'una coalició, diversos agents poden estar fortament connectats en sub-sistemes en que cada agent té una relació amb cadascun dels altres agents de la mateixa coalició [6]. El que manté els agents units en una coalició és el que s'anomena valor de recompensa. A través d'algoritmes que favoreixen la creació de coalicions, el valor de recompensa dividit o distribuït entre els agents no serà mai inferior al que cadascun dels agents podria obtenir individualment. Això provoca que un conjunt d'agents que aconsegueix amb èxit solucionar els problemes que se li plantegen, serà un avantatge per als mateixos agents que el componen.

2.2 - Redefinició conceptual i arquitectònica (Fase II)

2.2.1 - Introducció i context

En aquesta secció del treball es tractarà d'enunciar i explicar quines són les característiques, propietats i capacitats que són necessàries per a la cooperació entre agents.

En un primer lloc l'estudi es centra en les propietats que presenten els agents amb necessitats d'interacció amb altres agents, bé sigui buscant únicament el benefici propi, bé sigui amb finalitats cooperatius.

En un segon lloc ens focalitzarem en els recursos i capacitats dels agents cooperatius, intentant entrar en detall i descomposició de les mateixes.

Finalment, es presenta un esquema gràfic centrat en l'agent que treballa en un sistema multi-agent (arbitrari, obert i divers, com s'explica en la primera secció). Aquesta representació utilitza un format de mapa d'idees en que es relacionen conceptes de fàcil enteniment entorn a l'agent que volem descriure. L'esquema presentat es fa servir de guia per explicar detalladament cada característica en el següent apartat.

És clau recordar que totes les característiques que es descriuen en aquesta part del treball són per defecte aplicables tant als agents programats per mirar únicament per beneficis propis, com a aquells que tenen objectius cooperatius. Aquests últims contaran amb unes capacitats exteses que s'explicaran a part i que els permetran tenir visió i funció cooperativa.

2.2.2 - Propietats dels agents i sistemes multi-agent orientats a la col·laboració

Un model d'avaluació de confiança hauria d'estar dissenyat segons les següents característiques:

- Precisió en el llarg termini: un sistema de càlcul de confiança hauria de ser capaç de reflectir valors similars a aquells registres de reputació consolidats. També hauria de ser capaç d'aplicar un tractament totalment diferenciat als agents desconeguts comparat a aquells que disposin de mala reputació.
- Consistent: el sistema hauria de retornar valors de confiança i reputació similars en avaluar agents també similars.
- Orientat a l'actualitat: en els mètodes de càlcul de confiança i reputació haurien de tenir un pes considerable aquells comportaments i tendències que s'hagin produït recentment. D'una manera equivalent, les informacions recollides sobre interaccions produïdes en el passat haurien de tenir un pes inferior en alguna proporció a la seva antiguetat.
- Eficient: el model hauria de ser capaç de calcular els valors de confiança, o recalculat els valors de reputació, ràpidament.
- Independent i robust: el disseny hauria de considerar poder resistir atacs d'entitats externes amb interessos privats o subjectivats.
- Susceptibles d'avaluació estadística: hauria de ser possible trobar patrons i factors que defineixin el comportament d'un sistema d'avaluació de confiança.
- Privat: no hauria de ser fàcil reproduir el sistema d'avaluació de confiança sense tenir el coneixement de com s'ha desenvolupat.
- Progressiu: un petit canvi en els valors de reputació hauria de traduir-se també en un petit impacte en els valors de confiança calculats.
- Comprensible: Les implicacions d'un valor de reputació calculat per un sistema haurien de ser entenedores per sí mateixes.
- Verificable: el sistema hauria de guardar informació que pugui explicar com s'han calculat els seus valors de reputació.

- Situacionals: més enllà dels valors de reputació calculats per a cada agent, el sistema hauria de ser capaç de tenir en compte l'estat de cada agent en el moment de decidir les interaccions necessàries. Factors com la disponibilitat o la càrrega poden ser decisius en la selecció d'agents.
- Aprenentatge i millora continua: els mètodes d'avaluació de confiança haurien de ser capaços d'aprendre i millorar el seu rendiment a partir de pròpies i terceres experiències.

2.2.3 - Capacitats i mecanismes per a agents i sistemes multi-agent orientats a la col·laboració.

En aquest apartat es tracten d'enumerar i explicar les estratègies i mecanismes que fan capaçs als agents de col·laborar entre ells en interaccions basades en la confiança.

Mecanismes d'avaluació de confiança

Com s'ha descrit anteriorment, aquests algorismes tenen per objectiu fer estimacions per saber en quina mesura un agent pot confiar en altres. Aquest tipus de mecanismes estan basats només en recursos propis, entre els quals hi ha [2]:

1. Models d'avaluació directa: el més simple és el de consultar el registre d'experiències pròpies amb altres agents. Aquest seria el primer recurs a utilitzar per un agent que preten estimar la confiança en un altre: consultar la seva pròpia història d'interaccions per veure quina experiència i resultat en va treure.
2. Models socio-cognitius: en aquest cas es tracta de funcions que estimen el nivell de confiança en un agent tercer basant-se en propietats o qualsevol altra informació recollida respecte a l'agent a analitzar. És el que per als humans equivaldria al judici d'altres persones: usar les informacions observables de que disposem per valorar la confiança que hi podem dipositar.

Mecanismes de consulta de reputació

Solucions que permetran als agents consultar valors de reputació a altres agents i en registres o estructures organitzatives. [2 i 11]. Els més habituals són:

1. Testimonis de reputació: aquesta mesura d'estimació indirecta consisteix simplement en consultar a altres agents els valors de reputació de que disposen respecte a l'agent analitzat.
2. Reputació certificada [19 i 21]: en aquest cas es tracta d'entitats que emeten certificats de reputació basats en anàlisi propis o en valors de reputació contrastats per múltiples agents. Aquesta solució es presenta factible només quan l'entitat certificadora és reconeguda dins del sistema multi-agent, per tal de minimitzar els interessos i possibles

subjectivacions dels valors de reputació dels agents. Aquests principis s'han aplicat en la creació de diverses solució, bàsicament acadèmiques i enfocades a l'ús experimental. Un dels exemples més destacats és:

Certified Reputation in Trust (CRIT): model pel que juntament a cada valor de reputació s'aporta un valor adicional que quantifica la fiabilitat de la reputació aportada.

Reliability and Reputation Agent Framework (RRAF): sistema proposat per a l'estudi de la reputació certificada en entorns multi-agent competitiu.

La reputació certificada té, en els sistemes proposats, dos desavantatges principals: en primer lloc, en els diversos estudis i proves fets, no s'ha demostrat una millora substancial en el rendiment individual dels agents que utilitzen registres d'aquesta mena. Per una altra banda, normalment la consulta d'aquests registres té un cost, que pot fer que a l'agent examinador no li surti a compte pagar. Una alternativa al segon problema són els models en que es proposa que el certificat de reputació ha de ser guardat pel propi agent al qual fa referència. Un exemple d'aquest model és el següent:

Mediator-based certified reputation scheme (MCR): per aquest sistema cada agent s'encarrega de guardar i distribuir sota comanda els certificats de reputació que l'afecten. Per evitar que els agents evitessin distribuir aquells certificats de reputació negatius, es proposa la creació d'un registre amb números de sèrie per tal que tots els certificats puguin ser consultats.

Entitats de coordinació i mediació

Aquests són uns altres tipus de recursos que poden ser part d'un sistema multi-agent o d'una coalició ja formada. Faciliten el repartiment de tasques en benefici de l'eficiència del conjunt. En algunes ocasions faciliten la negociació i l'intercanvi dels termes dels contractes puntuals entre agents [6].

1. Agents coordinadors o de repartiment de tasques: un sistema multi-agent pot aportar estructures o agents tercers encarregats de repartir i equilibrar les càrregues de treball per maximitzar l'eficiència del conjunt. Les tècniques més destacades en serien:
 - Balanceig de càrregues: funcions que proven d'igualar la càrrega de processament de la manera més uniforme

possible. Aquesta distribució pot ser tant estàtica (exercida en un moment inicial o puntual del treball) o dinàmica (distribució repetida en diferents moments).

- Assignació geomètrica: aplicació de funcions de distribució uniforme utilitzant algorismes de divisió dimensional. Solen ser d'aplicació en distribució de superfícies o volums.
- Coordinació basada en rols [18]: es proposen mecanismes pels quals els agents poden determinar quins són els agents més convenients per a la col·laboració depenent de la funció d'assignada. Els rols poden ser revisats i reassignats de forma dinàmica de forma que el sistema multi-agent o la coalició que l'incorpora variarà de topologia al llarg del temps.

2. Agents mediadors: també anomenats "escrow agents", s'encarreguen de garantir que l'intercanvi de bens entre dos agents es realitza segons els termes acordats inicialment. Dos agents poden establir, per exemple, intercanviar un treball de processament a canvi d'una valoració positiva de la reputació del primer. Aquest tipus d'agents són utilitzats sovint en transaccions econòmiques a canvi de serveis. En aquests casos els agents mediadors s'asseguren que el servei o producte s'ha entregat abans de realitzar el pagament final, quantitat econòmica que s'ha demanat per avançat a l'agent client. Els agents mediadors poden també precisar rebre una comissió pel servei prestat.

Mecanismes de selecció d'agents per a la col·laboració

Escollir el millor agent o agents per a una col·laboració determinada (o inclús per a una relació consistent en un joc o programa de col·laboracions) és el pas posterior a l'avaluació de confiança i reputacions descrites fins ara. Dins d'aquests mecanismes se'n solen fer dues grans classificacions [2]:

- Algorismes que responen únicament a criteris propis: aquest tipus de mecanismes són els més simples i també sobre els quals hi ha més literatura. Són criteris per a la selecció de l'agent o agents més convenients per al benefici de l'agent que inicia la col·laboració. Això implicarà cercar en cada cas una maximització de la funció del benefici buscat però també alhora una minimització del risc derivat de la cooperació. També s'anomenen mecanismes de selecció egoïsta, per motius evidents. Com s'ha explicat en apartats anteriors, aquests programes poden tenir en compte avaluacions de

confiança pròpies o valors de reputació externs, però sempre buscant un resultat positiu únicament per al primer agent.

- Algoritmes d'interés col·lectiu: també anomenats dinàmics, tenen en compte variables i factors més complexes que els anteriors: la variabilitat de l'entorn, la càrrega d'altres agents o la cerca de millors rendiments en el mitjà o llarg terminis.

Mecanismes d'avaluació del rendiment del càlcul de confiança

Una àrea d'interès durant l'estudi de com els agents avaluen i apliquen criteris de confiança en agents tercers és la de l'avaluació del rendiment dels seus propis mecanismes. A l'hora d'estimar en quina mesura els criteris i fórmules de selecció d'agents són efectius, hi ha dues formes d'obtenir valors de referència:

- A través de simulacions: mètode més extès en que s'utilitzen entorns de prova, on es fan coexistir els agents necessaris per poder testear l'efectivitat de cada mètode.
- A través de valors pre-calculats externs: comparant valors de rendiments publicats per estudis científics i organitzacions especialitzades.

Tècniques d'aprenentatge, d'imitació i referència

Una de les capacitats que més poden marcar la diferència en el rendiment d'agents i sistemes multi-agent cooperatius és la de l'aprenentatge i l'evolució autònoma. Això és així perquè aquesta branca de l'intel·ligència artificial, com moltes altres, presenta dos reptes principals. En primer lloc, transmetre coneixement als agents autònoms és intrínsecament una tasca complexa. En segon lloc, perquè l'agent treballarà en general en un entorn canviant, obert, dinàmic i amb reptes i problemàtiques segurament també noves. Amb l'ús de tècniques d'aprenentatge, els agents implicats en processos d'avaluació de confiança seran capaços d'evolucionar a mesura que ho faci el seu entorn, i per una altra banda, podran incrementar el seu coneixement d'una manera més natural.

Una tècnica que s'aplica en els agents cooperatius és la de la imitació i demostració [8]: dotar un agent amb mecanismes d'aprenentatge per imitació aporta diverses capacitats estratègiques per a la pròpia evolució de l'agent en el curt, mitjà i llarg termini. En primer lloc, el permetrà adaptar-

se a les condicions possiblement canviants de l'entorn en el que treballi. En segon lloc, la tasca de transmissió de coneixement per part d'éssers humans serà més fàcil, a través de la demostració de les tasques a aprendre (utilitzant entorns d'entrenament en que l'humà podrà executar accions manuals de col·laboració entre agents que l'agent observador podrà imitar o replicar).

2.2.4 – Arquitectura general del parell Agent i Sistema Multi-Agent

Aquest capítol tracta de resumir esquemàtica i gràficament l'arquitectura general però alhora completa de la tupla entorn multi-agent i d'agent cooperatiu amb fins cooperatius. Els criteris de construcció aplicats a ambdós elements són els següents:

- facilitar la comunicació, interacció i negociació
- orientar-se a la diversitat i la inclusió, per a permetre que agents altament especialitzats puguin aportar el seu valor
- l'establiment de relacions obertes i productives
- el rendiment de les col·laboracions entre agents.

El model de sistema multi-agent tindria les següents característiques i propietats:

Obert: el sistema multi-agent ha d'estar totalment obert a l'entrada i sortida d'agents, per tal d'afavorir l'adaptació i el dinamisme. Per tant, el sistema multi-agent proposat ha de comptar amb:

- Registre d'entrada i sortida d'agents
- Condicions establertes i comunicades

Regulat: es defineixen condicions per a l'entrada i sortida d'agents. Aquestes condicions han de ser poc restrictives per afavorir la diversitat i la inclusió.

Organitzat: el sistema ha d'aportar les solució ja conegudes per a facilitar la interacció entre agents.

En aquest apartat es mostra un esquema gràfic que gira entorn a l'agent cooperatiu, envoltat de totes les capacitats, propietats, recursos i mecanismes que el defineixen i completen.

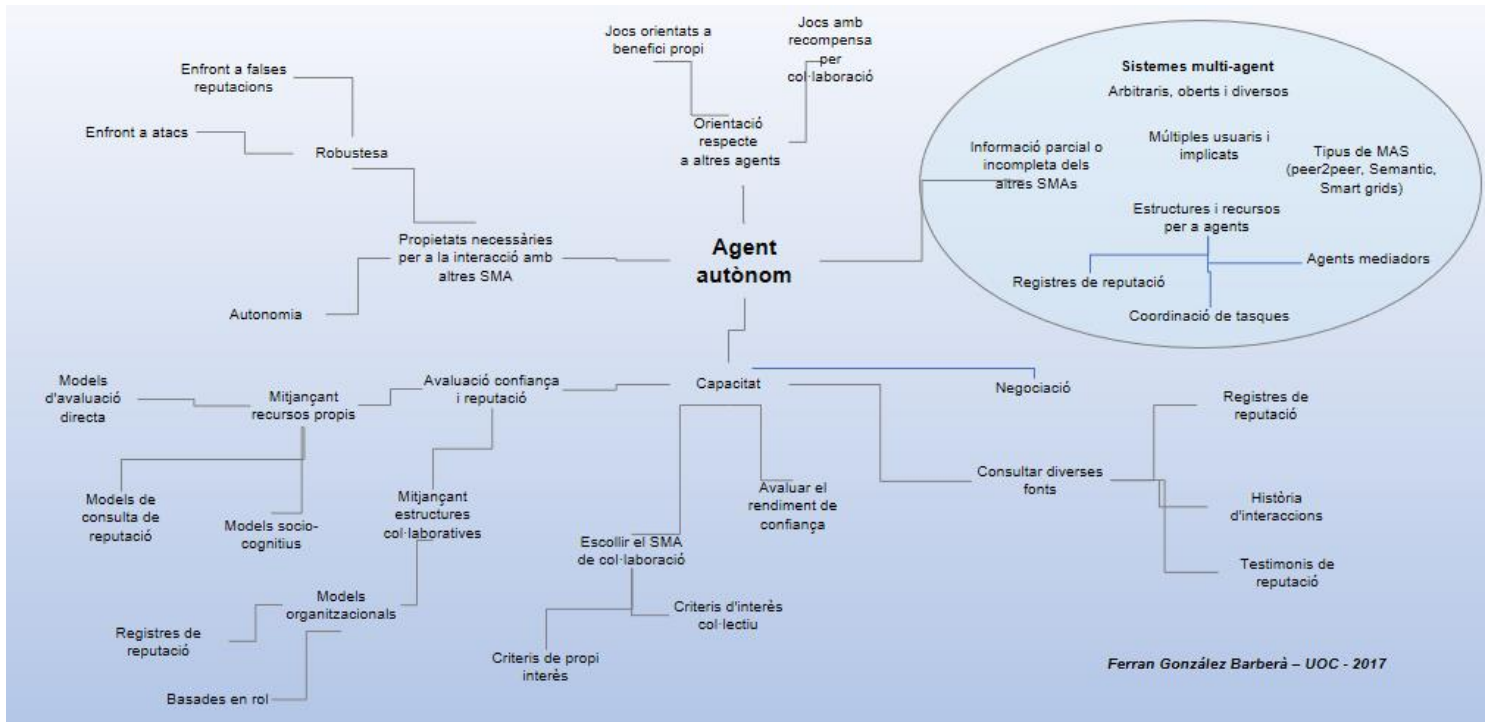


Figura 1: "Visió completa dels conceptes relacionats amb "cooperació" i "reputació" en un sistema multi-agent"

2.3 - Proposta evolutiva (Fase III)

Gràcies a les investigacions i a la corresponent literatura produïda en els darrers 15 anys entorn a aquesta branca de la Intel·ligència Artificial, comptem amb una gran completesa i diversitat d'eines a implementar i aplicar. Cal remarcar que la majoria d'elles són d'una utilitat complementària a les altres, en el sentit que moltes de les tècniques avui en dia documentades no exclouen l'aplicació conjunta de moltes altres.

Per una altra banda, en els sistemes que treballen amb agents cooperatius hi ha nombroses àrees d'investigació obertes, algunes d'ells encara poc explorades. Exemples en són els models d'avaluació socio-cognitius, el disseny de reptes per a l'avaluació d'agent tercers, o els protocols orientats a la negociació dels termes cooperatius.

D'acord amb lo exposat, les propostes que es fan a continuació van dirigides a afegir nous enfocaments i mecanismes que puguin afegir-se a qualsevol agent amb aspiracions cooperatives. El primer d'ells és un protocol de negociació de recompenses, que permet que els agents acordin com repartir-se les compensacions com a condició prèvia a començar a cooperar. Aquesta solució és innovativa perquè suposa un pas més enllà als repartiments pre-establerts proposats per les teories de jocs competitius, que també assumeixen la participació dels agents a qualsevol preu.

La segona proposta es correspon a una evolució dels sistemes de pissarra tradicionals, enfocada als agents autònoms basats en relacions de confiança. Aquest mecanisme reuneix diverses funcions útils per a aquest tipus d'agents: és alhora un registre de reputació de lliure accés i escriptura, un directori de recursos i un registre d'experiències. Incorpora també compensacions per tal de motivar als agents a l'aportació de dades. Aquesta solució suposa una innovació perquè evoluciona sistemes ja existents aportant serveis útils però de baix cost als agents que l'utilitzin.

Igual que succeeix amb la majoria dels mètodes i algorismes vistos fins ara en aquest document, els mecanismes proposats seran de més o menys utilitat depenent de la funció específica de l'agent.

2.3.1 – Propostes evolutives

En aquest capítol es presenten les dues solucions mencionades anteriorment. Cal recordar que degut al caràcter modular del disseny dels agents cooperatius, els mecanismes proposats poden ser afegits com a funcions complementàries que podran ser utilitzats en les situacions i entorns necessaris.

2.3.1.1 – Protocol de negociació per a jocs de confiança cooperatius

El cas concret de la teoria de jocs cooperatius, que prova de modelar com agents autònoms formen coal·licions i els beneficis que se poden extreure, es basa normalment en la assumpció que les recompenses seran lliurement distribuïdes entre els agents integrants (“transferable utility assumption” [6]). El fet de com es distribuïran les recompenses fruit de relacions de confiança exitoses es deixa en segon terme degut a una segona assumpció: que els agents estaran interessats en formar part (o seguir formant part) d’una coal·lició o una relació en concret independentment de la quota de recompensa que se’ls hi assigni. En realitat, el cost de cada col·laboració es contraposarà a la recompensa que se n’espera obtenir, pel que conèixer per avançat el valor exacte de la recompensa serà clau per decidir la conveniència d’una col·laboració.

També cal considerar la importància dels costos associats a l’establiment de relacions de confiança. Obviar-los pot incórrer en el risc que en més o menys mesura es troba en tots els jocs de confiança cooperatius: la dissolució o el trencament de relacions (“disbanding” en anglès).

Per solucionar els problemes exposats es proposa un protocol de negociació de recompenses entre l’agent iniciador de la col·laboració i els agents contactats o susceptibles de cooperar. Amb l’objectiu de simplificar la modelització, descriurem el problema entorn a només dos agents.

Notarem com “O” l’operació que l’agent “A₁” necessita externalitzar (que la facin altres agents). Per tal de realitzar l’operació “O”, l’agent “A₁” necessitarà d’un agent “A₂”. Notem també la recompensa inicial d’acomplir la citada operació com “R_{BASE}”, i com “R₁” i “R₂” les recompenses parcials (un cop repartides), i el cost de cada agent com “C₁” i “C₂”. Respecte al Benefici “B_n” que cada agent tregui d’una cooperació en concret, podem partir de l’equació següent:

$$B_n = R_n - C_n$$

Haurem de tenir en compte que cada agent haurà de calcular-se per ell mateix el cost de processar la part de feina repartida, incloent el cost i risc associats a la relació de confiança. En aquest càlcul de cost també s'hi afegiran les comissions que un possible agent mediador ("escrow agent") pot carregar si intervé per a assegurar-se que les transaccions s'efectuen segons l'acord.

En un pas previ a l'inici de les negociacions de recompenses, partirem de la base que l'operació és inicialment interessant per a l'agent A_1 . També podríem establir com a cas possible que la recompensa inicial no compensés a l'agent A_1 inicialment, de forma que busqués altres agents per tal que el benefici fos finalment positiu. Deixarem aquest cas a part per ser molt concret i poc probable.

$$R_1 > C_1, \text{ és a dir, } B_1 > 0$$

En qualsevol cas, durant el procés de negociació sempre s'haurà d'acomplir que la suma de recompenses repartides no pot ser superior a la recompensa inicial. En altres paraules, no es pot repartir més recompensa entre els agents de la que hi ha inicialment:

$$R_{BASE} = R_1 + R_2$$

Un pas clau per a l'agent iniciador A_1 abans d'iniciar el protocol de negociació és definir quin serà el marge de negociació. En aquest punt assumirem que l'únic objectiu de l'agent iniciador serà aconseguir un benefici lo més alt possible. Per tant, per a que la col·laboració sigui interessant per A_1 , el benefici resultant haurà de ser superior al que n'obtidria de fer la operació ell mateix, i en tot cas que aquest benefici sigui positiu. Això equival a dir que la cooperació serà convenient només si el marge (benefici) que aconseguix A_1 col·laborant amb l'agent A_2 és superior al que aconseguiria sol:

$$\begin{aligned} C_{BASE} > C_1 &\rightarrow B_{BASE} < B_1 \\ B_{BASE} &< R_1 - C_1 \\ R_1 > 0; R_2 > 0 \end{aligned}$$

El protocol de negociació en sí pot ser definit com una funció que tractarà de maximitzar el benefici a obtenir per a l'agent A_1 , i que haurà de jugar sota les següents condicions:

- Només l'agent iniciador A_1 coneix el valor B_{BASE} .
- Només l'agent iniciador A_1 coneix el valor C_{BASE} .
- L'agent A_1 no coneix estat de càrrega o costos de l'agent A_2 .
- L'agent A_2 també intentarà conseguir un benefici màxim.

No entrarem en aquest estudi en estratègies de negociació, però s'entèn que aquestes podrien ser una possible via de desenvolupament. En aquesta proposta s'intenta simplement començar oferint valors de recompensa molt baixos, i anar incrementant paulatinament fins que l'agent consultat accepti la proposta.

Un cop definit el marc de negociació, es pot començar a descriure el protocol a seguir entre els dos agents per arribar a un acord.

Fins que decisió (A_2) = positiva o $B_1 \leq B_{BASE}$ repeteix:
A₁: Calcula_proposta_recompensa $R_2 = R_2 + R_{BASE} / 10$
A₁: Proposa_tasca (O, A₂, R₂)
A₂: C₂=Avalua_cost_tasca (O)
A₂: B₂=Calcula_benefici (R₂, C₂)
A₂: Respon_conveniència (decisió)
Fi repetició

La senzillesa d'aquest algoritme respon a que hem pres un conjunt d'assumpcions per simplificar aquest exercici. Es podria adaptar a la negociació multi-lateral (amb diversos agents alhora), a valors de canvi multi-dimensionals (afegint factors de negociació com el temps d'entrega), o adaptar-se a serveis continus a proveïr en el temps.

Per una altra banda, és també cert que existeix bibliografia relacionada directament a la negociació entre agents, tot i que no està orientada a buscar els equilibris entre recompenses i costos deguts a les relacions de confiança. Una referència que investiga en la negociació entre agents de valors multi-dimensionals pot consultar-se en [26]. Una altra que resulta d'especial interès és [27], en la que cada agent analitza les incerteses subjectives de cada situació com a factor clau de cada negociació.

2.3.1.2 – Sistemes de pissarra per a agents cooperatius

La majoria de solució presentades en papers acadèmics limiten la informació a compartir entre agents autònoms a dos tipus de sistemes: els registres de reputació -on només es consulten valors numèrics referents als càlculs de confiança-, i la mateixa interacció directa entre agents. En ocasions també es proposen sistemes de pissarra per a la comunicació i coordinació entre agents, però només per a que els agents es coordinin entre ells apuntant les evolucions fetes en les tasques en les que treballen.

Aquí es proposa un sistema de pissarra orientada a agents cooperatius, que permet als agents compartir coneixement respecte a interaccions, relacions, coal·licions o qualsevol tipus d'experiències passades entre agents o amb recursos externs. A més el sistema proposat incorpora un agent moderador, que s'encarrega que s'apliquin els criteris d'ús definits més endavant.

L'acció de compartir informació per part de qualsevol agent autònom ha de ser recompensada per tal d'afegir un component de motivació. Els mètodes i mecanismes de compensació queden definits dins de la teoria de jocs cooperatius, pel que no entrarem aquí en aquest detall.

Un sistema específic de pissarra com l'aquí descrit pot ser implementat com a públic o com a privat, definint separadament drets de lectura i escriptura (per exemple, pot ser obert a l'escriptura dins del MAS al que pertany, però només de lectura per a accessos externs). A través d'aquest sistema de pissarra es poden compartir informacions respecte a:

- a. Disponibilitat i estat de serveis i recursos interns: es podrien publicar dades de l'estat de càrrega de biblioteques o serveis de cerca, l'adició de nous agents mediadors, etc.
- b. Informació respecte a recursos externs, incloses incidències, errors, noves fonts d'informació entre d'altres.
- c. Experiències respecte a cooperacions entre agents.
- d. Resultats de col·laboracions ja completades, per evitar repetir les mateixes operacions ja fetes anteriorment.

Cal recordar que els sistemes de pissarra són òptimament aplicables quan es compleixen les següents condicions, que són justament les que en general defineixen els sistemes multi-agent de caire cooperatiu:

1. Hi ha un considerable nombre d'agents. En aquest cas, a més, el nombre d'agents pot ser indeterminat i variable en el temps.
2. Les informacions que s'hi han de publicar poden ser d'interès per a diversos agents.
3. Les dades publicades poden ser clarament tipificades i concretes.
4. Els agents solen estar interessats en altres agents en concret o en serveis específics, pel que la cerca pot ser ràpida i el resultats concrets.
5. Les consultes es fan només quan es necessiten, o en altres paraules, no cal que els agents estiguin continuament informats dels canvis produïts en el sistema de pissarra. Normalment les informacions requerides pels agents seran relatives a agents o serveis en concret, i les sollicitaran en el moment en que avaluin si un agent o un servei pot ser útil per a una col·laboració puntual.

Els avantatges d'aquest sistema respecte als anteriorment documentats serien:

- Els agents poden consultar resultats d'operacions i col·laboracions anteriors, evitant duplicar esforços i costos.
- La consulta de les informacions podria fer-se sense cost per als agents sol·licitants, per tal que aquesta operació no afecti al rati cost-benefici calculat pels agents.
- L'aportació de coneixement al conjunt ha de ser premiada amb recompenses, segons els mètodes seguits en la teoria de jocs cooperatius.

A més, tenint en compte que sovint els MAS de treball són oberts, es poden definir mecanismes per tal de connectar el nostre sistema de pissarra a altres de similars. Aquest mecanisme aportaria coneixements respecte a nous agents incorporats en el MAS o serveis externs d'accés lliure. Aquest mecanisme també ha quedat reflectit en l'arquitectura proposada en aquest mateix capítol. Per ampliar informació sobre aquest punt, el document citat en [22] pot resultar útil. Es tracta d'un sistema

distribuït de pissarres interconnectades entre elles mitjançant la definició de regles entre elles.

Per a la construcció d'un sistema com aquest, caldrà definir les següents estructures:

- Polítiques d'accés tant per a la lectura com per a l'escriptura de registres.
- Definició de l'estructura de dades a emmagatzemar, amb camps i subestructures com "Agent autor", "data de registre", "data de caducitat de la informació", "Agent o servei objecte", "Valoració numèrica multi-dimensional", etc.
- Acotació dels tipus d'informació a emmagatzemar: els tipus podrien ser "Incidència", "experiència entre agents" o "Font d'informació" per exemple.
- Definició d'un sistema de compensacions per a la publicació de dades útils. Les compensacions podrien calcular-se tenint en compte el "feedback" rebut pels propis agents, entre d'altres criteris.
- Creació d'un agent moderador que segueixi les polítiques i regles definides.
- Publicació i registre d'aquest servei en els directoris del MAS de treball, o en cercadors oberts si procedeix.

L'arquitectura que es desprèn dels requeriments enunciats és la que es representa a continuació:

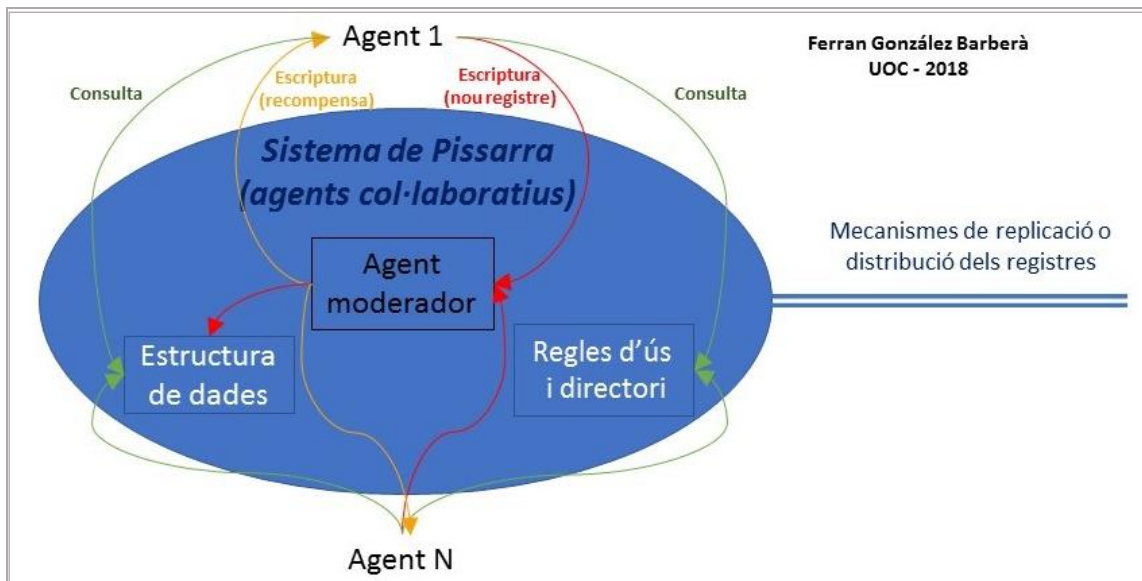


Figura 2: Sistema de pissarra per a agents cooperatius

2.3.2 – Àrees obertes de recerca

Tot i existir una extensa bibliografia relativa al camp de coneixement de la confiança entre agents autònoms, segueixen estudiant-se i essent publicats documents científics que tracten d'avançar en diversos fronts oberts. En lo següent es tracta de resumir quins reptes es presenten actualment en aquesta àrea i en alguns casos possibles vies d'investigació:

Robustesa dels sistemes de confiança i reputació

Que un TRS (sistema de confiança i reputació) sigui robust segueix sent un dels principals reptes en aquest camp [25]. Existeixen multitud de tipus d'atacs que poden usar-se per impactar, en positiu o en negatiu, els valors de confiança i reputació calculats o registrats en un sistema multi-agent (o inclús en una coal·lició). Estratègies orientades a l'alteració dels valors de reputació com els "playbooks" (agents entregant temporalment resultats d'alta qualitat obtenint així alta reputació, per a obtenir molts contractes en el que entregar baixa qualitat amb baixos costos), les conspiracions (coal·licions d'agents amb l'únic objectiu d'impactar en valors de reputació concrets) o les valoracions injustes (agents donant valors de reputació molt baixos respecte a agents tercers per perjudicar-los) són molt difícils de detectar i corregir.

Per tal d'incrementar la robustesa d'aquests sistemes, es proposen solució com ampliar el testeig d'agents i MAS a entorns reals (amb agents maliciosos), o dissenyar tests que específicament posin a prova la robustesa dels TRS.

Mètodes socio-cognitius avançats per a l'avaluació de confiança

Com s'ha descrit en l'apartat de "Capacitats i Mecanismes per a Agents i MAS", els mètodes socio-cognitius capaciten als agents autònoms per a esbrinar tanta informació com sigui possible amb l'objectiu de fer càlculs acurats de confiança respecte un segon agent. Aquests són els mètodes més complexes i que ahora permeten més desenvolupament. Un dels mecanismes pocs explorats en aquesta categoria és la del disseny de reptes enfocats a l'avaluació de confiança.

Algoritmes multi-dimensional de prioritizació de tasques

Un altre dels camps poc explorats en el terreny dels agents autònoms que treballen en entorns cooperatius és el de la prioritizació multi-dimensional de càrregues i tasques. Des d'una visió i objectiu

inicialment simplista, els agents cooperatius estan programats a aconseguir dos objectius bàsics: aconseguir per un costat els millors resultats possibles, i per una altra banda la creació de relacions (incloses coalicions) per tal d'aconseguir grups d'agents lo més efectius possibles. En aquests criteris es troba a faltar sovint l'adició de mecanismes per a la maximitació de càrregues de treball, incloent solució per al balanceig de tasques entre agents. Arquitectures de funcions com les següents queden en una àrea encara poc explorada:

- Sistemes complexes de prioritització de tasques: programació tenint en compte la prioritització de tasques amb diversos criteris alhora:
 - Tasques de primer nivell
 - Cooperacions, segons data i hora d'entrega, càrrega total...
- De màxima ocupació: donat que ha d'estar orientat a la cooperació, la programació hauria de reflectir que és millor cooperar en tasques de baix nivell de compensació, que no disposar de càrrega de treball.

Criteris avançats en la selecció d'agents per a la cooperació

Un dels objectius principals de la teoria de jocs cooperatius és la predicció de com coalicions entre agents es formaran o es trencaran. Les coalicions es formen a través de la necessitat de conjunts d'agents d'incorporar-ne de nous a partir de valors unidimensionals o multidimensionals de reputació (específics per tipus de tasca), però aquests criteris presenten certes limitacions. Valors com la inclusió (donar oportunitats a nous tipus d'agents) o la diversitat (integració d'agents amb diferents capacitats) són poc o gens presents en les estratègies de formació de coalicions, tot i que en ocasions poden representar increments de productivitat importants en les mateixes.

3. Conclusions

Conclusions respecte al camp d'estudi

La meua visió sobre el camp d'estudi dels models de confiança entre agents autònoms ha anat canviant durant el desenvolupament d'aquest treball. Es va partir d'unes poques referències bibliogràfiques que creia que contenien bona part dels progressos fets en aquesta matèria ([1], [2] i [4]), i a partir d'elles vaig estimar un ritme de treball i uns objectius que després hauria de canviar. L'estudi de mètodes de confiança i reputació entre agents té certes peculiaritats que el fan tant interessant com complex:

a) és un camp multi-disciplinar, que es creua sovint amb bases matemàtiques i financeres (p.e. teoria de jocs), bases de la psicologia (confiança i relacions), de la ètica (falses reputacions), o més purament informàtiques (protocols i estàndars de comunicació) i específiques de la intel·ligència artificial (p.e. aprenentatge automàtic).

b) permet una anàlisi des de punts de vista molt diferents, degut a la gran diversitat d'ús d'agents i sistemes multi-agent (usos financers, de recerca, comercials, etc).

c) és d'aplicació modular. És a dir, la majoria de teories i papers científics publicats proposen solució que són complementaris als altres. Això implica que els agents poden anar sumant propietats i capacitats, utilitzant mecanismes diferents en entorns i situacions diferents.

d) utilitzar la pròpia experiència humana en aquest camp no és només interessant sino que és molt útil. Totes i cadascuna de les modalitats d'estudi, amb les seves solució proposades, tenen un cert grau de similitud amb el comportament humà. Des de mètodes senzills com els d'avaluació directa (recordar experiències passades respecte a altres persones) o els de reputació indirecta (demanar a tercers quines experiències han tingut amb la persona analitzada), fins a mètodes complexes com els socio-cognitius (en els humans seria fixar-se en les aparences o informació de primera vista), s'apropen molt als nostres comportaments. De la mateixa manera, aquests mecanismes pateixen de riscos similars als que patim nosaltres: danys en la reputació, errors en les estimacions de

confiança, intents d'aconseguir millor reputació per a aconseguir beneficis subjacents, ...

Un fet remarcable en aquest camp de treball és que hi ha pocs documents publicats que tractin de donar una visió global dels avenços fets fins al moment. Hi ha un alt grau de disgregació d'informes i anàlisis, amb molts articles publicats oferint punts de vista i solució diferents, però sovint sense introduir el coneixement anterior. Aquí és on he trobat un dels reptes principals quan tractava de copçar el coneixement disponible, però també l'oportunitat d'oferir un treball amb una característica diferenciadora: donar una visió general dels progressos fets fins ara d'una manera planera i alhora estructurada.

Procés de producció d'aquest treball

Des d'un principi he hagut d'anar adaptant els objectius a mida que explorava a fons aquest camp de coneixement, pels motius que conté una bibliografia més extensa de la que m'esperava (i amb molts enfoc diferents), i que hi ha pocs estàndards en la matèria. Els canvis que he hagut de fer durant la producció d'aquest treball han estat:

a) ajustament en el repartiment de pesos del treball: inicialment tenia planejat dedicar aproximadament un terç dels esforços per cada tema (resum del coneixement, architectures i evolució), però he hagut d'adaptar-me en veure que entregar una visió completa del camp d'estudi em prendria la major part del temps. Al final ha estat un repartiment aproximat d'un 50% pel resum de coneixement, un 30% per architectures, i un 20% per evolutius i resta del document.

b) l'enfoc inicial era el de fer un treball en que les dues primeres fases permetessin treballar en un producte a produir en la tercera etapa (evolutius). El fet de veure que per una part, l'àrea de coneixement de la confiança entre agents és força extensa, i per una altra banda detectar que no hi ha cap document que sigui capaç de copçar el coneixement disponible, m'ha fet canviar d'objectiu i considerar els primers dos apartats com un entregable en sí mateix. Finalment la part d'evolutius ha quedat com a secundària, aprofitant algunes àrees per explorar detectades en els últims mesos.

c) inicialment també tenia la intenció de proposar un disseny complet d'agent col·laboratiu en la Fase III (Evolutius), però vaig descobrir que, igual que fan molts altres estudis científics, no calia repetir mecanismes que ja funcionen, sinó proposar-ne de nous i addicionals. Això també és deu a que com he mencionat abans, els

agents i els sistemes multi-agent amb fins cooperatius són en realitat una suma de solució, propietats i capacitats orientades a la cooperació amb altres agents.

Línies de treball pendents

Més enllà de les citades línies d'investigació, explicades en l'apartat 2.3.2, en aquest treball m'hagués agradat poder incloure informació relativa als atacs i riscos inherents als models de confiança i reputació. Tot i que s'expliquen en una referència que considero clau [25], hagués sigut interessant poder completar aquest treball amb una descripció i un reenforc dels problemes principals que es troben en l'aplicació de mecanismes de confiança i reputació.

També crec que s'escauria dedicar un capítol a les tècniques i solució de testeig de funcions de confiança. Sovint les metodologies de test per a agents amb relacions de confiança inclouen la creació o adaptació d'entorns anomenats "testbeds", en la que es crea un nombre determinat d'agents amb els comportaments i configuracions dessitjats, juntament amb serveis, normatives i altres estructures organitzatives relacionades amb l'objectiu del test. Les simulacions es fan normalment per tandes fins que s'assoleix una mostra prou representativa.

Per últim, crec que aquest document deixa oberta la possibilitat a aprofundir en tots els capítols que conté. Tant pot servir d'entrada introductòria a aquesta disciplina, com de punt de partida per al desenvolupament de qualsevol de les àrees que es tracten.

4. Glossari

Agent autònom: entitat de software capaç de portar a terme operacions amb un cert grau d'autonomia, utilitzant un cert grau de coneixement o representació necessaris.

Agent o sistema multi-agent cooperatiu: entorn de programari que conté agents autònoms i regula la seva actuació mitjançant l'aplicació de tècniques per a l'afavoriment de la cooperació, típicament mitjançant l'establiment i repartiment de recompenses.

Avaluació directa: en el camp específic de l'avaluació de confiança, conté aquells mecanismes que un agent disposa per analitzar interaccions passades amb un segon agent i predir el nivell de confiança aplicable a interaccions futures.

Avaluació indirecta: conté tots aquells mètodes d'avaluació de confiança que parteixen d'informacions de reputació obtingudes per part d'agents tercers.

Avaluació per mètodes socio-cognitius: mètodes d'avaluació de confiança que utilitzen dades d'arquitectura, procedència, tipologia, especialització o qualsevol altra informació relativa a un segon agent per tal d'avaluar el nivell de confiança en el mateix.

"Certified reputation in Trust" (CRIT): model de registre de reputacions en el a més del valor de confiança s'aporta un valor adicional que quantifica la fiabilitat de la reputació aportada

Coal·lició: sub-sistema dins d'un sistema multi-agent en que cada agent té una relació amb cadascún dels altres agents.

Confiança: assumpció per part d'un agent basada en que un segon agent seguirà els termes d'una relació previàment negociada, incloent inherentment un cert nivell de risc.

"Escrow agent": també anomenats agents intermediaris, són agents encarregats de fer de mitjancers en relacions regulades per un contracte, assegurant que els compromisos presos, inclòs el possible pagament de serveis, es porta a terme segons lo acordat.

Estructures organitzatives: específicament en el camp dels sistemes de confiança, són aquells recursos públics o privats que són independents respecte als agents que hi interactuen però que els aporten un o més serveis comuns.

Jocs de congestió: tipus específic de joc de confiança en el que es té en compte la càrrega o congestió de l'agent analitzat.

Jocs de confiança: models i mètodes per a l'avaluació de confiança per a la cooperació puntual o continua entre agents autònoms.

Recompensa ("Pay-off"): és el factor afegit en els jocs cooperatius per tal de motivar els agents a que cooperin per a acometre les tasques que porten a terme.

"Reliability and Reputation Agent Framework" (RRAF): sistema proposat per a l'estudi de la reputació certificada en entorns multi-agent competitiu.

Reputació: nivell calculat de confiança calculat o registrat per tercers agents respecte a un o més agents autònoms.

Reputació certificada: nivell de confiança suportat per un certificat de reputació emès per una entitat certificadora reconeguda.

"Sistemes de reputació i confiança" (TRS, Trust and Reputation systems): engloba tots aquells sistemes creats en base als models de reputació i confiança, que tenen per objectiu definir els mètodes i regles per a l'avaluació, registre i propagació de valors de confiança.

Sistema multi-agent (MAS - "Multi-agent system"): entorn o recipient de software que aporta un seguit de serveis comuns als agents que hi actuen o interactuen.

Sistema de pissarra: eina utilitzada entre dos o més agents amb l'objectiu de compartir coneixement entre ells. Tots els agents poden actualitzar-hi dades, i poden consultar-les també en qualsevol moment sota demanda.

Teoria de jocs: estudi de models matemàtics que tracten de predir les formes i conseqüències de la interacció entre agents autònoms.

Teoria de jocs competitiu: subcamp de la teoria de jocs focalitzat en aquelles interaccions entre agents que tenen el benefici propi com a únic objectiu.

Teoria de jocs cooperatius: subcamp de la teoria de jocs centrat en l'estudi de les tècniques que afavoreixen la cooperació entre agents i la formació de relacions i coalicions. També tracten de predir la formació i dissolució de relacions i els seus efectes.

Web semàntica: projecte que té com a objectiu crear un medi per a l'intercanvi d'informació continguda en Webs per a l'ús per part de les màquines. Els estàndars inclouen formats de dades i protocols per a l'intercanvi d'informacions procedents de Web.

5. Bibliografia

- [1] P. Noriega, “Coordination, Organizations, Institutions, and Norms in Multi-Agent Systems”, UAB, 07/2005.
- [2] Han Yu, Zhiqi Shen, Cyril Leung, Chunyan Mia, and Victor R. Lesser, “A Survey of Multi-Agent Trust Management Systems”, IEEE, 05/2013.
- [3] C. McLeod, “Trust”, The Stanford Encyclopedia of Philosophy, 09/2015.
- [4] Jordi Sabater, Carles Sierra, “Review on Computational Trust and Reputation Models”, CSIC, Campus UAB, 2005.
- [5] Isaac Pinyol , Jordi Sabater-Mir, “Computational trust and reputation models for open multi-agent systems: a review”, Springer Science+Business Media B.V., 07/2011.
- [6] Dariusz G. Mikulski, “Cooperative Trust Games”, Intech, 2013.
- [7] G. Xu, G. Zhang, C. Xu, B. Lin, M. Li, Y. Ren, X. Li, Z. Feng, D. Zhang, “A multi-attribute rating based trust model: improving the personalized trust modeling framework”, Springer Science+Business Media New York, 02/2015.
- [8] Saleha Raza, Sajjad Haider, “Using imitation to build collaborative agents, Artificial Intelligence Lab, IBA, 04/2016.
- [9] Pipattanasomporn, Feroze, Rahman, “Multi-agent systems in a distributed smart grid”: Design and implementation”, IEEE, 04/2009.
- [10] C. Guttman, F. Dignum, M. Georgeff, “Collaborative Agents – Research and development”, Springer, 08/2010 (Revised paper).
- [11] J. Weng, C. Miao, A. Goh, Z. Shen, and R. Gay, “Trust-based agent community for collaborative recommendation” in Proc. 5th Int. Joint Conf. Auto. Agents Multiagent Syst., 2006.
- [12] Jian, t. Baras, JS: “Trust evaluation in anarchy: a case study on autonomous networks”, IEEE Infocom, 04/2006.
- [13] T. Haus, I. Palunko, D. Tolic, S. Bogdan, F.L. Lewis, D.G.Mikulski, “Trust-based self-organising network control”, IETDL, 08/2014
- [14] M. Branner, R. Ellis, A. Macintosh, “Applications and Innovations in Intelligent Systems XI”, BCS- Springer, 2003.

- [15] D. Isern, D. Sánchez, A. Moreno, “Organizational structures supported by agent-oriented methodologies”, *J. Syst. Soft.*, vol 84, pp. 169-184, 2011.
- [16] Chris Arney, Kristin Arney, Elisha Peterson, “Models and metrics of geometric cooperation”, US Military Academy, 2010.
- [17] R.I. Whitfield a, *, A.H.B. Duffy a , I. Boyle a , S. Liu a,1 , I. McKenna b, “An integrated environment for organisational decision support”
- [18] R. Hermoso, H. Billhardt, and S. Ossowski, “Role evolution in open multiagent system as an information source for trust”, *Conf. Auto. Agents Multiagent Systems*, 2010.
- [19] T. D. Huynh, N. R. Jennings, and N. R. Shadbolt, “Certified reputation: How an agent can trust a stranger,” in *Proc. 4th Int. Joint Conf. Auto. Agents Multiagent Syst.*, 2006.
- [20] Luhmann, "Familiarity, Confidence, Trust: Problems and Alternatives," pp. 94-108, 1988.
- [21] F. Buccafurri, A. Comi, G. Lax, D. Rosaci, “Experimenting with Certified Reputation in a Competitive Multi-Agent Scenario”, University Mediterranea of Reggio Calabria, 2016.
- [22] J. Straub, “A Distributed Blackboard Approach Based Upon a Boundary Node Concept”, *Journal of Intelligent & Robotic Systems*, 2006.
- [23] Mohamed Gaha ; Michel Gagnon ; Frederic Sirois, “A modern blackboard system”, IEEE, 2011.
- [24] Y. Shoham and K. Leyton-Brown, "Teams of Selfish Agents: An Introduction to Coalitional Game Theory," in *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge: Cambridge University Press, 2009, pp. 367-391.
- [25] Audun Jøsang, Jennifer Golbeck, “Challenges for Robust Trust and Reputation Systems”, University of Maryland, USA, 2009.
- [26] A. Rosenfeld, I. Zuckerman, E. Segal-Halevi, O. Drein, S. Kraus, “NegoChat-A: a chat-based negotiation agent with bounded rationality”, Springer, 2015.
- [27] O. Marey, J. Bentahar, E. Khosrowshahi-Asl, K. Sultan, R. Dssouli, “Decision making under subjective uncertainty in argumentation-based agent negotiation”, Springer, 2015.