

Bibliografie

1. L Blaga, Despre conștiința filosofică. Ed. Făclia 1974, p. 55
2. Ibid., p. 55
3. Ibid., p. 56
4. Ibid., p. 64
5. Ibid., p. 67
6. Ibid., p. 65
7. Ibid., p. 67
8. Ibid., p. 68

UN CONCEPT DE AGENT CREDIBIL ÎN CADRUL UNEI SOCIETĂȚI BAZAT PE SISTEME DE VALORI

Braga Vasile
Universitatea Tehnică a Moldovei

***Abstract:** Lucrarea dată propune o arhitectură de agent credibil social, influențată de așa concepte socioumane cum ar fi "sistem de valori", "rol social" și "arhetip". Conceptul este bazat pe structura de agent de tip BDI[1], cu un strat adițional "Arhetip" - ABDI. Rolul stratului dat este de a susține existența unei "societăți" de agenți, care se auto-balansează în funcție de necesitățile și scopurile sale.*

Introducere. Un agent poate fi definit drept un program care poate avea scopuri, pe care încearcă să le rezolve, capacități pe care le posedă, unele cunoștințe pe care le are și e autonom pentru respectarea acțiunilor sale.

Agenții pot fi considerați următorul pas în progresul gradual al dezvoltării softului. Programarea bazată pe agenți poate fi considerată o extensie logică a etapelor precedente și poate fi văzută ca o formă specializată a unui obiect, unde nimeni din afara nu poate să schimbe starea obiectului, sau ca o formă particulară a unui component.

Semnificația cuvântului "credibil" în domeniul inteligenței artificiale diferă de cea cotidiană. Pentru mulți oameni, expresia de agent credibil înseamnă, agent care spune adevărul sau un agent în care se poate avea încredere. Însă în articolul de față nu se are în vedere acest lucru. Credibil este un termen care vine din caracterele animate. Un personaj credibil este cel care pare realist, ale cărui acțiuni au sens, care permite privitorului să creadă în el. Acest lucru nu este aceeași cu realismul. De exemplu, Bugs Bunny este un personaj credibil, dar nu un caracter realist.[2] Scopul principal al cercetării este de a permite artiștilor și

altor profesioniști implicați în producția media interactivă, cum ar fi Jocuri, Visual Novella sau Lumi Virtuale, să creeze agenți emoționali credibili. Pentru a realiza acest obiectiv, s-a realizat o cercetare dintr-o serie largă de surse, inclusiv arta, psihologia, ontologia și AI. Artele ajută ca să înțelegem mai bine problema, iar psihologia și AI oferă anumite idei despre posibilitatea rezolvării problemei[3]. Faptul că agenții credibili sunt interactivi îi deosebesc de personajele artelor tradiționale și fac crearea lor și mai dificilă. Autorul unui agent credibil nu poate controla sau cunoaște cu certitudine situațiile pe care agentul le va întâlni[4]. În acest context, puterea principală a jocurilor video ca mediu de divertisment constă în rolul activ al publicului sau al jucătorului, de aceea problema fundamentală care trebuie rezolvată este modul în care designerul poate folosi marele potențial al unui mediu interactiv pentru jucători. Unul dintre principalii factori este crearea unui mediu dinamic prin folosirea unor personaje emergente și credibile.

Agenții credibili, asemănători oamenilor, au un potențial mare pentru a fi utilizați în domenii critice de cercetare, cum ar fi războiul[5,6] și simularea mulțimilor [7]. Nu în ultimul rând, acestea reprezintă o metodă rentabilă pentru a testa teoriile comportamentului uman în diferite domenii ale cercetării, cum ar fi teatrul educațional [4,8]și psihologia[9]. Când se vorbește despre un agent credibil în industria jocurilor video, se caută o iluzie a vieții; cu alte cuvinte, o suspendare de neîncredere. În situații reale, oamenii nu sunt perfecți. Comportamentul și alegerile lor nu confirmă întotdeauna previziunile. Procesul uman de luare a deciziilor nu satisface întotdeauna o funcție de utilitate. Cel puțin, nu toți oamenii au aceleași priorități și utilitate în diferite alegeri.

Cerințe pentru agenții credibili:

Personalitate. În domeniul inteligenței artificiale, și în particular, în Jocurile de pe calculator, personalitatea se caracterizează prin calitățile sale individuale, particulare și nu cele generale. De aceea personalitatea trebuie să fie bogată în caracteristici individuale și ar trebui să i se infuzeze tot ceea ce face un personaj, astfel ei vorbesc și se mișcă în modul în care gândesc. Ceea ce face ca personajele să fie interesante sunt căile lor unice de a face lucrurile.

Emoția - Personajele își manifestă propriile emoții și răspund la emoțiile celorlalți în moduri specifice personalității.

Motivația de sine - Personajele nu reacționează doar la activitatea altora. Ei au propriile forțe și dorințe interne pe care le urmăresc dacă alții interacționează sau nu cu ei.

Schimbare - Caracterele se dezvoltă și se schimbă odată cu timpul, într-o manieră compatibilă cu personalitatea lor.

Relațiile sociale - Personajele se angajează în interacțiuni detaliate cu ceilalți într-o manieră compatibilă cu relația lor. La rândul lor, aceste relații se schimbă ca rezultat al interacțiunii.

Iluzia vieții - Aceasta este o colecție de cerințe cum ar fi: urmărirea unor obiective și acțiuni multiple, simultane, având capacități largi (de exemplu, mișcare, percepție, memorie, limbă) și reacționarea rapidă la stimulii din mediul înconjurător. Artiștii tradiționali nu menționează aceste cerințe în mod explicit, pentru că le primesc de multe ori gratuit (de la un actor uman sau ca o presupunere profundă în animație). Însă designerii de caractere interactive trebuie să se preocupe în mod explicit de crearea arhitecturilor de agenți, care susțin aceste cerințe[2].

Valorile sunt preferințe generale, principii generale de orientare și standarde acceptate pentru comportamente și rezultate. Valorile principale informează procesele de comunicare strategică și servesc drept apeluri persuasive în relațiile publice și în campaniile de management al problemelor. Valorile sunt, de asemenea, folosite pentru crearea judecăților valorice despre bine și rău, despre dorințe și plăceri, pentru a evalua legitimitatea organizațională și responsabilitatea socială corporativă. Valorile sunt factori din ce în ce mai importanți în activitățile de comunicare strategică.

Valorile și normele posedă proprietăți adaptive care organizează cunoașterea socială și facilitează interacțiunile dintre indivizi și grupuri. În acest sens, ele sunt legate de cele mai importante proprietăți ale culturii umane cumulate care coordonează viața socială și permit o comunicare semnificativă într-un grup. Cultura este înțeleasă drept o extindere a capacității umane de a organiza, stoca, prelua și transmite informații de o semnificație colectivă care depășesc capacitatea individuală de memorie și nevoile individului de informare pentru uz personal.

Valorile sunt de obicei conceptualizate ca idei care pot forma sisteme coerente și există într-o formă ierarhică în ceea ce privește stabilirea obiectivelor 1. Asociațiile de valori adoptate în comportamente acționează ca o manifestare modală a unei culturi. În acest sens, cultura poate fi considerată drept model organizat de variație a discursului asupra a ceea ce este bun; discursul colectiv asupra valorilor.

Dar totuși, valorile separate nu au sens atât de pronunțat pe cât sistemele de valori. Nu te poți baza în viața de zi cu zi pe o singură valoare cum ar fi de exemplu libertatea, deoarece chiar fiind o valoare destul de largă, ea nu ne permite să reglementăm nici chiar să dam răspuns la toate situațiile și conflictele. De exemplu, cum să se comporte în luptă/conflict. De aceea persoanele tind spre un sistem integru de valori, ce ar permite să se integreze în societate și ar răspunde "la pachet" la toate problemele care ar putea apărea. Concomitent în fiecare societate pot fi valide mai mult decât un sistem de valori, fiecare din aceste sisteme poate fi numit rol în societate sau ne vom permite noi sa-l numim arhetip.

Carl Jung a înțeles arhetipurile ca modele universale, arhaice și imagini care derivă din inconștientul colectiv și sunt corespondentul psihic al instinctului. Acestea sunt potențiale moștenite, care sunt actualizate atunci când intră în conștiință ca imagini sau se manifestă în comportament în interacțiunea cu lumea exterioară. Ele sunt forme autonome și ascunse care sunt transformate odată ce intră în conștiință și sunt date expres de particulari și de culturile lor. În psihologia Jungiană, arhetipurile sunt elemente extrem de dezvoltate ale inconștientului colectiv. Existența arhetipurilor poate fi dedusă indirect numai prin utilizarea de povestiri, artă, mituri, religii sau vise[10].

Pentru a explica tendința de a compromite libertatea personală și a se conforma unui sistem de valori oarecum străin și impus, și în același timp, statistic vorbind, absolut atrăgător, se vor aminti conceptele lui Jaques Derrida și anume "Writing and Difference" [11] Despre simboluri și conținut, Simbolul fiind doar o formă, și conținutul îi este diferit la fiecare citire și depinde de capacitățile cititorului, cred că același lucru se aplică și arhetipurilor. Fiecare persoana "umple" arhetipul spre care tinde la limita posibilităților și capacităților sale. Din acest punct de vedere fiecare instanțiere a unui rol social poate fi diferită și nu este o copie fidelă a unui original ideal de tip platonian[12]

În modelul sistemelor de valori asta ar însemna ca de exemplu Persefona - arhetipul care formează personalitatea sub influența altor oameni. O persoană de arhetip persefona poate fi destul de diferită, deoarece trăsătura principală a acestui arhetip este pasivitatea, flexibilitatea și umilința.

Dacă am încerca să cercetăm societăți mult mai simple decât cele umane, de exemplu cea a furnicilor, sau a lupilor or cimpanzeilor, vom putea observa că conceptul de arhetipuri este aplicabil peste tot.

Conceptele propuse. Se propune un nou model cognitiv de agent cu straturi noi de personalitate și scop social. MPVBCDI (Misiune, Personalitate / Identitate, Valori, Convingeri, Capacități / Abilități, Dorințe, Intenții). Arhitectura este de tip hibridă[13], care combină o bază de cunoștințe simbolice cu reguli și elemente de procesare sub - simbolica folosind așa tehnici stohastice ca rețele neuronale ori rețele bayezienne[14] pentru un aparat decizional probabilistic.

Nivele de interes pentru prezenta lucrare sunt P - Personalitate / Identitate și M-Misiune / Scop Social, la aceste niveluri putem lucra cu simularea personalității și a scopurilor sociale. Această nouă structura schimbă perspectiva în ansamblu. Problemele de reactivitate și pro-activitate sunt rezolvate cu straturile vechi BDI[1] Sistemul de Valori - pentru rezolvarea intențiilor conflictuale, identitate pentru conștientizarea personalității, de asemenea, acesta este nivelul de control indirect pentru toți factorii care pot influența deciziile noastre și Misiune, este stratul, care răspunde la întrebarea: "Care este scopul meu în această societate?", și el poate avea două aspecte: unul de conformare la cerințele societății și altul de a modifica societatea, astfel încât să includă alte

roluri, arhetipuri, care ar rezolva problemele individuale, care nu sunt luate în considerare de structura data de roluri.

Deci, avem trei straturi noi, Identitate, Misiune și Sistem de Valori, care vor răspunde pentru personalitatea și interacțiunea socială a agentului. Toate trei pot fi ușor implementate într - o abstracție pe care o vom numi Arhetip, și ea ne va permite să generalizăm aceste elemente în continuare, când avem nevoie să vorbim despre toate împreună. O notație a agentului dat va fi ABDI: Archetype Beliefs Desires Intentions.

Deoarece, Experiența agentului este un factor de influență enorm în ceea cum va fi înțeles - "citit" arhetipul, noi nu o putem scoate în afara sistemului și vom adăuga stratul istoric de experiență ABEDI, Care va fi absorbit de stratul de Beliefs, modificând-ul doar, ceea ne întoarce la modelul clasic de agent BDI cu Arhetip - ABDI.

Arhetipurile fiind roluri ale societăților de agenți - pot fi reprezentate printr-un sistem ierarhic sau ponderat de Valori și/sau o hartă comportamentală a rolului. La moment, se presupune, că comportamentul agentului este direct influențat de sistemul său de valori, deci, nu este necesară o dublare printr-un mecanism de algoritmi comportamentali. Adică, doi agenți, care au aceeași structura și același rol, vor avea comportamente similare în situații similare, dacă cunoștințele lor și experiența este similară. În acest caz, vom avea un agent credibil ideal cu toate cerințele îndeplinite [1]

Exemple și algoritmi. Se propun un șir de metaprograme, ce vor permite agenților sa-și concentreze atenția[15] și au menirea sa joace rolul de un **metaprocesor** simbolic, o parte din ele se ocupa de concentrarea atenției senzoriale, alta de selectarea reacției emoționale și o a treia parte de selectarea acțiunii din aparatul decizional.

Pentru analiza atenției, vom folosi taxonomia mecanismelor atenționale propuse de Tsotsos[18]. Elementele de atenție sunt grupate în trei clase de mecanisme de reducere a informațiilor:

1. selecție (alegeți unul din multe)
2. restricție (alegeți unele din multe)
3. suprimare (suprimați unele din multe)

O serie de metaprogramele au funcția de:

1. restricție: INTERNAL PROCESSES, CHUNK SIZE FILTER, FRAME OF REFERENCE FILTER, MANAGEMENT DIRECTION FILTER, THE WORK PREFERENCE FILTER
2. altele de suprimare: INTERNAL STATE, RELATIONSHIP FILTER, EMOTIONAL STRESS RESPONSE, TIME ORIENTATION FILTER, COMPARISON FILTER

Ca exemplu propun să modelăm rolurile în cea mai simplă, dar și extrem de eficientă societate, pe care o cunoaștem, o colonie de furnici. Într-o colonie obișnuită avem trei roluri de bază:

1. Queen (Regină)
2. Worker (Lucrător)
3. Soldier (Războinic)

Rolul Queen prin prisma metaprogramelor, poate fi reprezentat printr-un tuplu (-1,-1,-1,0,1,-1,-1,-1,0,1,1,0,-1,1,0,-1), ce e descifrat în felul următor:

1. EXTERNAL BEHAVIOR: -1 -> Introvert
2. INTERNAL PROCESSES: -1-> Intuitor
3. INTERNAL STATE: -1 -> Thinker
4. THE ADAPTIVE RESPONSE: 0, nu este important pentru rolul dat.
5. DIRECTION FILTER: 1 -> Away from.
6. CHUNK SIZE FILTER: -1 ->Global
7. ACTION FILTER: -1-> Proactive.
8. FRAME OF REFERENCE FILTER: -1-> Internal.
9. RELATIONSHIP FILTER: 0, neimportant
10. REASON (OR MODAL OPERATOR) FILTER: 1 -> Procedures.
11. MANAGEMENT DIRECTION FILTER: 1-> Others.
12. THE WORK PREFERENCE FILTER: 0 non important.
13. EMOTIONAL STRESS RESPONSE: -1 -> Dissociated.
14. TIME ORIENTATION FILTER: 1-> Future.
15. COMPARISON FILTER: 0 non important.
16. CLOSURE FILTER: -1-> Time.

În același mod definim și Worker ca (1,1,1,1,~0.5,1,-1,-1,-1,1,-1,1,0,-1,1) sau Soldier ca (1,1,1,-1,~-0.5,1,-1,1,1,-1,1,0,1,0)

Concluzie. Modelul ABDI permite să se emuleze o societate fără a avea reguli generale, înscrise într-un sistem separat de membrii săi, doar prin crearea unui șir de roluri complimentare. Stratul de arhetip se propune să fie autoindus de însăși membrii societății și să fie comportamental, astfel încât alți agenți să poată simplu analiza comportamentul oricărui vecin, să detecteze rolul acestuia, sau absența oricărui rol tradițional pentru societatea dată.

Se propune și identificarea metodelor de a premia sau marginaliza oricare agent cu un comportament dorit și respectiv nedorit.

Referințe

1. Rao AS, Georgeff MP. Modeling rational agents within a BDI-architecture. KR. 1991;91: 473–484.
2. Mateas M. An Oz-Centric Review of Interactive Drama and Believable Agents. Lecture Notes in Computer Science. 1999. pp. 297–328.
3. Reilly WS. Believable Social and Emotional Agents [Internet]. Carnegie-Mellon Univ Pittsburgh pa Dept of Computer Science; 1996. Available:

<http://www.dtic.mil/docs/citations/ADA310766>

4. Bryan Loyall A. *Believable Agents: Building Interactive Personalities*. Ph.D., School of Computer Science Carnegie Mellon University. 1997.

5. Bacon TJ, Jones P, Garrett RB, Tolk A. Integration of Psycho-Social models and methods in NATO's approach to operations; A review of NATO Research and Technology Organization (RTO) Systems Analysis Studies (SAS-074). Simulation Conference (WSC), Proceedings of the 2009 Winter. IEEE; 2009. pp. 2852–2859.

6. Schmorow D, Klein GL, Foster R, Boiney J. Applied use of socio-cultural behavior modeling and simulation: An emerging challenge for C2. dtic.mil; 2009; Available: <http://www.dtic.mil/docs/citations/ADA503085>

7. Durupinar F, Pelechano N, Allbeck JM, Gudukbay U, Badler NI. How the Ocean personality model affects the perception of crowds. *IEEE Comput Graph Appl*. 2011;31: 22–31.

8. McCoy J, Treanor M, Samuel B, Tarse B, Mateas M, Wardrip-Fruin N. *Comme Il Faut 2: A Fully Realized Model for Socially-oriented Gameplay*. Proceedings of the Intelligent Narrative Technologies III Workshop. New York, NY, USA: ACM; 2010. pp. 10:1–10:8.

9. Durupinar F, Allbeck J, Pelechano N, Badler N. Creating Crowd Variation with the OCEAN Personality Model. Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems; 2008. pp. 1217–1220.

10. Jung CG. *The Archetypes and the Collective Unconscious*. 1959. Trans RFC Hull New York: Princeton UP. 1990; Available: <http://blogs.jefftwp.org/wordpress/pvandigriff/files/2009/12/The-Archetypes-and-the-Collective-Unconscious1.docx>

11. Derrida J. *Writing and difference*. Routledge; 2001.

12. Plato, Plato. *Oxford World's Classics: Plato: Republic* [Internet]. 1994. doi:10.1093/actrade/9780199535767.book.1

13. Kotseruba I, Tsotsos JK. 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*. 2018; doi:10.1007/s10462-018-9646-y

14. Hon KC. *Dynamic bayesian networks: representation, inference and learning*. 2002; Available: <https://pdfs.semanticscholar.org/b322/eacdbe43b38cd0018f4ab800e75e706042d9.pdf>

15. James T, Woodsmall W. *Time line therapy: And the basis of personality*. Crown House Publishing Ltd; 2017.

16. Timmis J, Neal M, Thorniley J. An adaptive neuro-endocrine system for robotic systems. 2009 IEEE Workshop on Robotic Intelligence in Informationally Structured Space. 2009. pp. 129–136.

17. Samani HA, Saadatian E. *A Multidisciplinary Artificial Intelligence Model of an Affective Robot*. *Int J Adv Rob Syst*. SAGE Publications; 2012;9: 6.

18. Tsotsos JK. *A Computational Perspective on Visual Attention*. MIT Press; 2011.