Modely emocí pro autonomní agenty a umělé bytosti

Marek Kukačka

Úvod

V posledních letech (cca. od roku 1997) se otázky kolem konstrukce inteligentních systémů a umělých bytostí začínají stále častěji zaměřovat na téma umělých emocí. Různé články a projekty podporují názor, že implementace systému umělých emocí vede k vytvoření agentů úspěšnějších při řešení problémů v dynamickém a nepředvídatelném prostředí. Marvin Minsky ve svém díle *„The Society of Mind"* [4] podotkl, že „..není otázkou, zda by autonomní agenti měli mít emoce, ale zda je možné vytvořit inteligentního agenta bez emocí." Je možné, že emoce jsou podstatnou součástí lidské inteligence, pak by jejich použití bylo nezbytné pro vytvoření věrohodné umělé bytosti.

Ačkoliv skutečná podstata fungování lidských emocí je nám stále záhadou, bylo vytvořeno množství různých teorií a modelů ve snaze o co nejpřesvědčivější simulaci emocionálního chování. Některé z teorií byly s většími či menšími úspěchy implementovány, jiné zůstaly pro svou složitost ve fázi teoretické konstrukce. V této eseji popíši dva jednodušší modely použité při konstrukci emocionálních autonomních agentů a pokusím se o porovnání jejich výhod a nedostatků.

Fungování umělých emocí

Modely umělých emocí můžeme rozdělit podle několika kritérií. V některých architekturách jsou emoce použity pouze jako doplněk činnosti agenta, zatímco jinde jsou jednou z hlavních součástí řídícího systému. Dále můžeme rozlišovat mezi systémy, které jsou určeny pouze k rozpoznávání či vyjadřování emocí (např. robot Kismet, viz [5]), zatímco na druhé straně existují systémy, o kterých se dá říci, že „prožívají" emoce. Mnou vybrané modely náleží do druhé kategorie.

I přes rozdíly mezi různými modely je funkce umělých emocí v architekturách agentů ve většině případů stejná. Emoce slouží jako signály, směrující pozornost agenta na ty vnitřní nebo vnější podněty, které mají nějaký význam pro agentovy cíle, čímž těmto podnětům zajišťují přednostní zpracování. Vytvářejí tak sekundární rozhodovací systém (vedle např. reaktivního systému či deliberativního plánování, viz Sloman[1] kapitola 9), napojený na vnitřní a vnější senzory agenta, který ohodnocuje signály podle jejich významu pro hlavní úkoly agenta (tzv. *appraisal system*) a patřičným způsobem pak upravuje jeho chování.

Model emocí Cathexis

Cathexis je model umělých emocí inspirovaný Minského dílem „The Society of Mind" [4], jehož autorem je Juan D. Velásquez (viz [2]). Sestává ze dvou podsystémů: generátoru emocí a emocionálního řízení chování.

Generátor emocí je síť tzv. „proto-specialistů". Jsou to jednoduché mechanismy ne nepodobné umělým neuronům. Každý proto-specialista (dále PS) svou aktivitou ovlivňuje jednu emoci z vybrané základní množiny: vztek, strach, smutek, štěstí, znechucení a překvapení. Složitější emoce vznikají současnou aktivací většího počtu těchto jednoduchých emocí. Kromě krátkodobých emocí se silným efektem modeluje Cathexis i *nálady*, které mají slabší, ale dlouhodobější efekt.

Proto-specialisté monitorují vnitřní a vnější senzory agenta. Toto monitorování je rozděleno do čtyř kategorií: neuronové, sensomotorické, motivační a kognitivní. V závislosti na těchto vstupech se mění vnitřní potenciál PS. Každý PS má dvě prahové hodnoty, Alfa a Omega. Pokud potenciál překročí hodnotu Alfa, PS je aktivován a vysílá signál (excitační nebo inhibiční) ostatním PS, se kterými je spojen, a Systému emocionálního chování, kde aktivuje „svou" emoci s intenzitou odpovídající vnitřnímu potenciálu PS. Hodnota Omega signalizuje nasycení PS, nad ní už potenciál stoupat nemůže. Vnitřní potenciál PS při absenci vzruchů s časem klesá. Ty PS, které jsou zodpovědné za emoce, mají vyšší hodnoty Alfa a Omega než PS formující nálady.

Systém emocionálního chování (Emotional Behavior System) rozhoduje, jak bude agent navenek svůj emocionální stav vyjadřovat. Možným druhům chování je přiřazena váha podle různých činitelů, mezi které patří

emoce, nálady, vnitřní a vnější stimuly a aktuální cíle agenta. Konečné chování agenta je pak určeno podle vah systémem „*vítěz bere vše*“. Každé emocionální chování má dvě složky: *výrazovou* (expressive), sestávající z výrazu agentovy tváře, postoje a podobně, a *zkušenostní* (experiential), která se projeví mimo jiné při volbě akcí či jako vliv při plánování a výběru cílů.

Cathexis byl implementován v experimentálním počítačovém modelu dítěte „Simon the Toddler“. Uživatel mohl měnit nastavení proto-specialistů a hladiny neurotransmiterů, nebo se Simonem komunikovat pomocí několika akcí (hlazení, krmení, atd.), na které Simon reagoval změnou svého emocionálního stavu a odpovídající změnou výrazu tváře.

Teoretický model OCC

Teoretický model, jehož autory jsou Ortony, Clore a Collins a který je podle nich nazván OCC, se stal standardem pro implementaci emocí do umělých bytostí a autonomních agentů. Pro některé případy je příliš složitý a pro jiné (jako například pro tvorbu komplexnější umělé bytosti) naopak vykazuje podstatné nedostatky, ale pro svou rozšířenost a snadnou implementaci se stal nejlépe vyzkoušeným emocionálním modelem.

OCC je postaveno na předpokladu, že emoce jsou reakcí na kognitivní činnost. Proto jsou emoce v OCC modelu vyvolávány pouze v závislosti na určitých kognitivních aktivitách. Emocionální reakce je určena především tím, zda je vyvolána v souvislosti s *událostí*, *činností agenta* či *vlastností objektu*. Tímto jsou určeny tři základní třídy emocí: agent může být potěšen či nepotěšen událostí, souhlasit či nesouhlasit s činností jiného agenta, a může se mu líbit či nelíbit vlastnost objektu. Druhá kategorie je dále rozdělena podle toho, zda činnost vyvolal agent sám či jiný agent. Podobně události jsou rozlišovány podle toho, zda jsou jejich důsledky důležité pro tohoto agenta či pro ostatní, zda jsou žádoucí či nežádoucí, a podobně. Na konci této fáze *klasifikace* je určeno, které kategorie emocí budou aktivovány. OCC takto rozlišuje 22 kategorií emocí.

Dalším krokem syntézy emoce je určení její intenzity. Ta je ve vztahu k událostem nazývána *žádoucnost* (desirability) a je počítána podle důležitosti cílů, ke kterým má událost vztah. Intenzita emocí vyvolaných činností agenta je nazývána *chvalitebnost* a je určována podle nastavených standardů. Pro vlastnosti objektů se určuje *líbivost*.

Následující kroky sestávají z určení interakcí vytvářené emoce s již existujícími emočními stavy v dané kategorii v systému agenta (tj. zjištění, jak nová emoce změní již pociťované emoce, ale pouze stejného typu) a mapování nové emoce na výrazové prostředky agenta.

Srovnání

Přestože je model OCC dobrým prvním krokem pro vybavení agenta emocemi, má několik nedostatků, které brání tomu, aby se stal konečným řešením. Nijak například nespecifikuje vliv emocí na chování agenta, na rozdíl od modelu Cathexis, kde je tato vazba realizována Systémem emocionálního chování. OCC také nijak nepopisuje, jak má být řešena interakce mezi emocemi různých kategorií, kdežto v Cathexis je toto určeno vzájemným propojením proto-specialistů, pomocí nichž se mohou jednotlivé emoce posilovat či oslabovat. OCC neobsahuje dlouhodobější emocionální stavy – nálady, které jsou v Cathexis realizovány pomocí proto-specialistů s nízkým excitačním prahem. Konečně, OCC nepopisuje možné způsoby ovlivňování osobnosti agenta, která je v Cathexis určena nastavením vlastností proto-specialistů (hodnotami Alfa a Omega, inhibiční funkcí, váhou vstupů a podobně).

Na druhou stranu, model Cathexis nebyl tak široce testován jako OCC, v praxi ani experimentálně. Testování by pravděpodobně ukázalo, že základní sada emocí v Cathexis je buďto nedostatečná, nebo špatně vybraná. Základní návrh generátoru emocí v Cathexis je pravděpodobně příliš primitivní, sestával z pouze jednoho proto-specialisty pro každou emoci, zatímco složitější síť by byla flexibilnější a dokázala by lépe vyjadřovat různé emoce. U modelu Cathexis také není snadno zjistitelné, které konkrétní aspekty vnitřního či vnějšího světa vyvolaly určitou emoci, a proto by bylo obtížné umožnit agentovi přemýšlet a argumentovat o svých vlastních emocích.

Ani jeden z modelů neumožňuje modelovat složitější emocionální fenomény, jako například více protikladných emocí zároveň. Možný směr pro hledání řešení tohoto problému je zavedení systému vícedimensionální logiky, viz [6].

Závěr

Zájem o modelování emocí v umělých bytostech je relativně mladá záležitost, přesto již existuje množství teoretických návrhů i vyzkoušených počítačových modelů. OCC i Cathexis patří mezi ty jednodušší, umožňující implementovat systém umělých emocí do nepříliš komplexních autonomních agentů a umělých bytostí. Oba však mají své nedostatky, díky kterým jsou nepoužitelné pro konstrukci agenta, který by například mohl věrohodně vyjadřovat emoce při komunikaci s člověkem. Toto by mohly umožnit složitější modely, na jejichž efektivní implementaci si zřejmě budeme muset ještě počkat.

Reference:

[1] Emotional Computers - http://www.schooldays.de/ruebentemp/emeocomp/content.HTM
[2] Cathexis - http://portal.acm.org/citation.cfm?id=267808
[3] Integrating the OCC model of emotion in embodied characters – http://www.bartneck.de/work/bartneck_hf2002.pdf
[4] Societies of Mind - Minsky, M. (1987). Picador.
[5] Kismet, the robot - http://www.ai.mit.edu/projects/sociable/baby-bits.html
[6] Modelling emotions with multidimensional logic - http://student.vub.ac.be/~cgershen