

De Evolutie van Conventies in Multi-Agent Systemen

Bart De Vylder

Samenvatting

Vele conventies ontstaan geleidelijk aan, zonder dat ze centraal worden opgelegd centrale controle. In de menselijke samenleving is het meest opvallende en complexe voorbeeld ongetwijfeld onze taal, die evolueerde vanuit de noodzaak om te communiceren. Ook in artificiële multi-agent systemen zoals mobiele robots of software agents is het vaak wenselijk dat de agents een conventie kunnen bereiken op een gedistribueerde manier. Om dit mogelijk te maken, is het noodzakelijk belangrijk inzicht te krijgen in het mechanisme proces waarbij conventies tot stand komen.

In deze thesis definiëren we een theoretisch kader om dit proces nauwkeurig te kunnen onderzoeken. We maken een strikt onderscheid tussen de omschrijving van een *conventieprobleem* enerzijds en een oplossing voor dit probleem in de vorm van een agent architectuur anderzijds. Een conventieprobleem specificereert de randvoorwaarden waaraan elke agent moet voldoen. Dit omvat de ruimte van alternatieven waaruit de conventie wordt gekozen, de structuur van het multi-agent systeem en de informatie die de agents kunnen uitwisselen tijdens een interactie. Een bepaalde agent architectuur lost een conventieprobleem op als een verzameling van dergelijke agents in staat is om tot een consensus te komen in een aanvaardbare tijdsperiode. Een essentieel kenmerk van is telkens dat er meerdere evenwaardige alternatieven bestaan, maar dat alle agents uiteindelijk dezelfde keuze moeten maken.

Om inzicht te krijgen in hoe conventies tot stand komen, dient men de relatie te begrijpen tussen het gedrag van de individuele agents en de dynamica dat dit voortbrengt op het populatieniveau. Deze globale dynamica omschrijft of het de agents lukt om tot een conventie te komen en zo ja, hoe lang ze er over doen.

Dit probleem heeft veel raakvlakken met disciplines zoals o.a. populatie genetica, evolutionaire speltheorie, opinion dynamics en statistische fysica. Er bestaat echter, voor zover wij weten, nog geen algemene wiskundige theorie op vlak van de evolutie van conventies die de relatie tussen het gedrag van de individuele agents enerzijds en de macroscopische dynamica die hieruit voortvloeit anderzijds onderbouwt.

We beweren niet dat we zo een theorie aanreiken in deze thesis. We trachten echter wel de eerste stappen op dit vlak te zetten. Dit doen we door eerst en vooral een theoretisch kader te schetsen waarin verschillende bestaande problemen aangaande de evolutie van conventies passen en waarin deze ook nauwkeurig omschreven kunnen worden. We maken een strikt onderscheid tussen de omschrijving van het probleem enerzijds en een oplossing voor dit probleem in de vorm van een agent design anderzijds. Het *conventie probleem*, zoals we de omschrijving verder zullen benoemen, bepaalt de randvoorwaarden waaraan elke agent moet voldoen. Dit omvat (i) de ruimte van alternatieven waaruit de conventie wordt gekozen, (ii) het model van de interactie tussen de agents die bepaalt welke agents wanneer samenwerken en (iii) het soort van informatie die de agents uitwisselen tijdens een interactie.

Een bepaalde agent design lost een conventie probleem op wanneer een populatie van zulke agents tot een akkoord komt binnen een aannemelijke tijdsperiode

Vervolgens en aan de gegeven voorwaarde voldoet.

Wij focussen we dan ook vooral op eende klasse van conventie problemen met een globaal interactie model: elke agent kan met elke andere agent informatie uitwisselen. Voor deze klasse van problemen tonen we aan dat we kunnen voorspellen of een bepaald type agent

al dan niet een conventie zal bereiken, wat wil zeggen dat elke agent met elke andere agent kan en waarschijnlijk ook zal samenwerken. Wij beweren hier dat het gedrag van zo een agent te voorspellen is aan de hand van de eigenschappen van de agent's *response functie*. aan de hand van zijn *responsefunctie*. Deze functie response functie beschrijft omvat het gemiddelde gedrag op lange termijn dat een agent vertoont wanneer deze interacties heeft met andere agents uit van een agent wanneer die samenwerkt met een agent uit een niet-veranderende populatie.

Deze analysetechniek it resultaat passen we is toe gepast op verschillende soorten conventie problemen. Voor de meer eenvoudige conventieproblemen definiëren we algemene, voldoende eigenschappen die garanderen dat een conventie na verloop van tijd ontstaat wanneer een agent ze bezit.sommigen konden we de eigenschappen van een klasse van agents die het probleem oplossen, opstellen.

VVoor de moeiljkere meer conventieproblemen houden we ons bij het ontwerp van een agent waarvan we kunnen aantonen dat deze het probleem oplost.

ingewikkelde problemen stellen we bepaalde agents voor die binnen ons kader het probleem kunnen oplossen.

Als laaTenslotte passen tstewe ons theoretisch wordt ons kader toegepast op het probleem van de taal evolutie in van artificiële agents. Dit is een gecompliceerd zeer ingewikkeld domein waarin exacte wiskundige resultaten moeilijk rvoor het moeilijk is om exacte wiskundige resultaten te bekomen zijn. We zullen focussen daarom op een klasse van relatief eenvoudige modellen van taalevolutie die we ruwweg kunnen omschrijven als op de 'naming games'. In bepaalde gevallen brengt onze analyse convergentieproblemen aan het licht die eerder nog niet werden opgemerkt. Dit toont enerzijds aan dat het belangrijk is om computereperimenten in taalevolutie theoretisch te onderbouwen en anderzijds dat het raamwerk dat we in deze thesis ontwikkelden hier zeer geschikt voor is., een relatief eenvoudig onderdeel binnen het paradigma van de language games. We tonen dan uiteindelijk ook aan dat ons kader kan gebruikt worden om deze naming games te begrijpen. Spijtig genoeg brengt het kader ook voorheen ongekende problemen aan de oppervlakte die dienen opgelost te worden in de toekomst.