

Samenvatting

Reinforcement leertechnieken (RL) werden reeds succesvol toegepast in heel uiteenlopende onbekende omgevingen met een hoge graad van onzekerheid. Ook voor domeinen waarin meerdere agenten actief zijn, is RL een interessant paradigma. In dergelijke domeinen zijn er echter ook verschillende nieuwe problemen. De agenten handelen autonoom en hebben mogelijk tegenstrijdige doelstellingen die zij wensen te bereiken. Een voor de hand liggende oplossing is om de agenten te voorzien van de nodige informatie over de toestand van de andere agenten, samen met hun handelingen en beloningen. Dit laat de agenten toe om evenwilibriumpunten in de toestandsruimte te leren, maar brengt een hoge kost met zich mee. De agenten leren in de gecombineerde toestand-actie ruimte wat, gezien de grootte van deze ruimte, het leerproces aanzienlijk vertraagt.

In deze doctoraatsverhandeling argumenteren wij dat in situaties waar de agenten maar zelden met elkaar interageren, het interessanter is voor het leerproces om de agenten onafhankelijk van elkaar te laten leren en elkaar enkel in rekening te brengen als dit nodig blijkt. In de toestanden waar agenten elkaar niet beïnvloeden zijn de toestandovergangsfunctie en de beloningsfunctie onafhankelijk van de toestand of actie van de andere agenten in het systeem. In deze situatie kunnen de andere agenten in de omgeving genegeerd worden en een RL techniek voor één enkele agent kan toegepast worden. Wanneer deze vereiste over de onafhankelijk echter niet geldt, hebben we te maken met een multi-agent coördinatie probleem en is een multi-agent leertechniek vereist. Een belangrijke vraag hierbij is hoe gedetecteerd kan worden wanneer de agenten interageren.

In deze thesis introduceren wij nieuwe leertechnieken die dergelijke interacties kunnen detecteren en, gebruik makend van deze informatie, ofwel een onafhankelijke leertechniek gebruiken, ofwel een multi-agent techniek. Het eerste algoritme, genaamd *2Observe*, maakt gebruik van de ruimtelijke relatie die bestaat in de toestandsruimte om de verzameling van toestanden waarin agenten elkaar beïnvloeden te leren. Deze techniek is gebaseerd op generalized learning automata en kan deze relatie benaderen. De tweede techniek, dat we *CQ-learning* noemen, gebruikt de onmiddellijke beloningen die een agent krijgt om de invloed van andere agenten in bepaalde toestanden te detecteren. Door het uitvoeren van statistische test op significante verschillen tussen deze beloningen is het mogelijk om de relevante

toestandsinformatie van andere agenten in interacties te bepalen. Het laatste algoritme, *FCQ-learning*, is een uitbreiding van dit idee, maar laat toe om coördinatieproblemen te anticiperen, verschillende tijdsstappen voor deze de kop op steken, en hierop te reageren. Dit resultaat wordt bereikt door de statistische tests uit te voeren op de som van de onmiddellijke en toekomstige beloningen.

Ten slotte demonstreren we ook hoe, gebruik makend van *2Observe* en *CQ-learning*, ervaring omtrent coördinatieproblemen kan veralgemeend worden en gedeeld worden tussen agenten en omgevingen. Deze methoden zijn de eerste technieken die dergelijke kennis kunnen doorgeven in multi-agent systemen.

Abstract

Reinforcement learning has already been widely used in unknown domains with a high degree of uncertainty. Also for domains in which multiple agents are acting together it is an interesting paradigm. In these domains however several additional problems arise. Agents behave autonomously and might have conflicting goals. A straightforward approach is to allow agents to always observe the state information of the other agents, as well as their actions and rewards they receive. This allows the agents to learn to reach equilibrium points in the environment, but it comes at a high cost. Agents are learning in the joint state-action space, which considerably slows down the learning process.

In this dissertation we argue that in settings where the interactions between agents are sparse, an efficient learning approach is to allow the agents to learn individually and only take into account the other agents when necessary. In the former case, agents are not influencing each other in a particular state. Hence, the state transition function and the reward function are independent of the state and action of any other agent acting in the environment. In this case, the learning can be reduced to single agent reinforcement learning and the agent can safely ignore the other agents in the environment. In the latter case, when this independency requirement does not hold, we are dealing with a multi-agent coordination problem and a multi-agent learning approach is required. A key question is how to determine when interaction occurs.

We propose novel approaches, which are capable of learning in which states such sparse interactions, occur and based on this information use either a single agent approach or a multi-agent approach. The first algorithm, called *2Observe*, exploits spatial dependencies that exist in the joint state space to learn the set of states in which sparse interactions occur. This approach is based on generalized learning automata that can approximate these dependencies in the state space. The second algorithm, called *CQ-learning*, uses the immediate reward signal to determine the influence of other agents in certain states. By performing statistical tests on these immediate rewards, the relevant state information of other agents during sparse interactions can be determined. The last algorithm, called *FCQ-learning*, extends on this idea, but also allows to anticipate coordination issues, several timesteps before they actually occur and as such dealing with the issue in a timely fashion. This is achieved by performing the statistical tests on the sum of immediate and future rewards.

Finally, we also introduce some methods to generalize knowledge about coordination problems and demonstrate how experience can be shared between agents and environments using *2Observe* and *CQ-learning*. These methods are the first in their kind to provide knowledge transfer about coordination experience in multi-agent systems.