Voting Mechanisms in Reinfocement Learning

Jost Alemann
Institute of intelligent & cooperating systems
Otto von Guericke University
Magdeburg, Germany
jost.alemann@ovgu.de

***Abstract*—This paper aims to deliver an overview over how voting mechanisms can be incorporated in reinforcement learning. Voting mechanisms and their properties are first introduced to the reader and then explained in more detail by describing their application in related work in the field of multi-agent systems and reinforcement learning.**

***Index Terms*—voting, reinforcement learning, multi-agent systems**

I. Introduction

In a democratic society choices are not made by a dictator but by taking the preferences of the whole society into account. Therefore each member of the society is entitled to cast a vote which represents it's preferences. Votes are then evaluated and a decision is derived by a given voting scheme. Plurality where the choice with the most votes wins, might be the most common voting scheme. Still there are many different voting schemes each of which can or cannot fulfil certain properties. This lies mainly in the subject of social choice theory and therefore is only described briefly if needed in the following. The concept of considering multiple individuals' preferences by using a voting mechanism to make choices can be transferred to multi-agent reinforcement learning systems. This is motivated by the expectation that agents combine their limited perception and knowledge of the environment by deciding together. Therefore they are expected to obtain better results than agents that choose actions based on only their own perception. [1]
Since learning agents try to optimise their own reward, they might learn to act selfish [2] or even learn to use strategic voting to exploit the voting system [3] for that purpose. Thus the design of a voting mechanism incorporated in a multi-agent reinforcement learning system is non-trivial. To avoid extensive selfish behaviour in agents several constraints can be introduced by modifying the reward function or voting system. The possibility to exploit a voting system depends on the chosen voting scheme. Unfortunately Arrow's Impossibility Theorem implies that no voting scheme can be designed to be completely fair. This means there is always a way in which agents could exploit such a voting scheme by finding a certain voting strategy. Arrow's Impossibility Theorem will be explained later on in Section II.
To give an overview over voting schemes and their possible properties Section II introduces basic principles of social choice theory as far as they are required to understand the concepts described later on. Section III then shows related work to give example applications of voting mechanisms incorporated in reinforcement learning. Additionally used methods to avoid exploitation or selfish behaviour of agents will be highlighted. As a conclusion different applications of voting mechanisms in multi-agent reinforcement learning settings are briefly compared in Section IV.

II. Basic Priniciples

The following gives an introduction to basic principles of social choice theory that are needed to understand the concepts described in III.

A. Different Voting Schemes

To discuss properties of voting systems we have to introduce those systems first.

- *Plurailty vote*:
  Voting scheme in which each voter is allowed to only cast one vote and the alternative with the most votes is chosen.
- *Borda Count*:
  Voting scheme in which all alternatives are ranked by each voter. Each vote consists of the ranked list of alternatives and a corresponding score for each ranked alternative. Let n be the total number of alternatives and let r be the assigned rank of an alternative in a single vote. The score is then calculated as following:
  $score_{alternative} = n - (r_{alternative} - 1)$
  Each alternative gets a total score by adding up it's scores of each vote. This way higher ranked alternatives get a higher score per vote.

B. Properties of Voting Schemes

- *Pareto efficiency*:
  If every voter prefers option X over option Y, then the society prefers X over Y.
- *Independence of irrelevant alternatives*:
  If every voter prefers option X over option Y and option Z is removed without changing the former relation, the societies preference of X over Y also remains unchanged.
- *Non-dictatorship*:
  No single voter posseses the power to always determine the group's preference.

C. Theorems

- *Arrow's Impossibility Theorem* [4]:
  Arrow's Impossibility Theorem states that no rank-order voting scheme can fulfil the properties Pareto efficiency, independece of irrelevant alternatives and non-dictatorship at the same time. Meaning that every voting scheme can be exploited by strategic voting. [3]

III. Related Work

H. Carr, J. Pitt and A. Artikis and their work [2] from 2008 consider a multi-agent setting in which the environment has limited resources that can be requested and offered by single agents. All other agents then decide which proposals are accepted by voting to redistribute the global resources.
Because of Arrow's Impossibility Theorem and the thereby implied possibility to manipulate the voting scheme by strategic voting, each agent is considered to be capable of either responsible or selfish behaviour. The authors point out that selfish behaviour is very likely in systems without social constraints. Thus they add a reputation system consisting of each agent's voting history to the voting process. Responsible acting agents only vote for other agents with a responsible voting history. This works as a constraint for selfish behaviour because agents that always vote selfish will not be voted except by themselves and therefore receive no reward and eventually act responsible in the future.
The voting mechanism consists of two phases. In the first phase all agents vote on a threshold value τ that specifies how many votes a proposal requires to be accepted. A τ that is too low leads to many accepted requests and the global resource storage being bankrupted. On the other hand a τ that is too high leads to only few accepted requests and unsatisfied agents changing their strategy to get a reward in the future. Threshold τ is decided in two rounds. In round one all agents propose a suggestion for value τ. In round two all agents cast a vote for one of the two most common suggestions. Resource reallocations are decided in phase two of the voting mechanism. First all agents propose a request or offer to the global resource storage. For simplicity reasons the authors only consider fixed value requests of resources. Then each agent votes on which proposal to accept. The authors found that a plurality voting scheme

IV. Conclusion

References

[1] I. Partalas, I. Feneris, and I. Vlahavas, "A hybrid multiagent reinforcement learning approach using strategies and fusion," *International Journal on Artificial Intelligence Tools*, vol. 17, no. 05, pp. 945–962, 2008.

[2] H. Carr, J. Pitt, and A. Artikis, "Peer pressure as a driver of adaptation in agent societies," in *International Workshop on Engineering Societies in the Agents World*. Springer, 2008, pp. 191–207.

[3] J. Pitt, L. Kamara, M. Sergot, and A. Artikis, "Voting in multi-agent systems," *The Computer Journal*, vol. 49, no. 2, pp. 156–170, 2006.

[4] K. J. Arrow, *Social choice and individual values*. Yale university press, 2012, vol. 12.