Learning Local Neighbourhoods in Multi-Agent Systems

Self-organisation of autonomous intelligent agents through local neighbourhoods of connectivity and knowledge

NIALL CREECH, Kings College London

In systems with a large number of agents there are fundamental pressures on the centralised coordination techniques used to provide inter-communication, task orchestration, and routing of messages. As the scale of interacting components expands, we reach resource constraint plateaus, where computation, storage, or communication pathways become saturated. At these points we must decompose each agents functionality into a number of specialisms that can then be taken up by other agents, at the cost of even more orchestration communications and synchronisation to provide this distributed functionality. To provide solutions to tackle these issues we focus on distributed agent systems where reinforcement learning behaviours are constrained by resource usage limits and hence by local neighbourhood awareness rather than global system knowledge.

In this work we develop algorithms for autonomous intelligent agents within a distributed multi-agent system that enable agents to learn and repeatedly adapt a subset of state-action space while also exploiting it to achieve a goal.Through the investigation of these systems we also provide some insight and define concepts that illustrate the behaviours of agents under these conditions that prove useful to build further contributions, including the use of constrained local neighbourhoods as units of scalability in large distributed systems.

LEARNING LOCAL NEIGHBOURHOODS

The Local Neighbourhood Learning Algorithm (Local-NL) tackles the need for agents to be able to adapt their connectivity and knowledge of other agents within the system to find sub-spaces that allow completion of a composite task. Using an extension of Q-learning, agents learn utilities of the range of actions available to them. They then have the choice of taking an action that will exploit their current local neighbourhood of other agents, bringing them closer to task completion, explore the neighbourhood to optimise their requests of other agents, or reshape the neighbourhood altogether. Altering the neighbourhood topology will bring in connections to new agents that may open up more optimal actions. However, due to resource constraints this comes at the cost of the loss of knowledge of some of the previous sub-space.

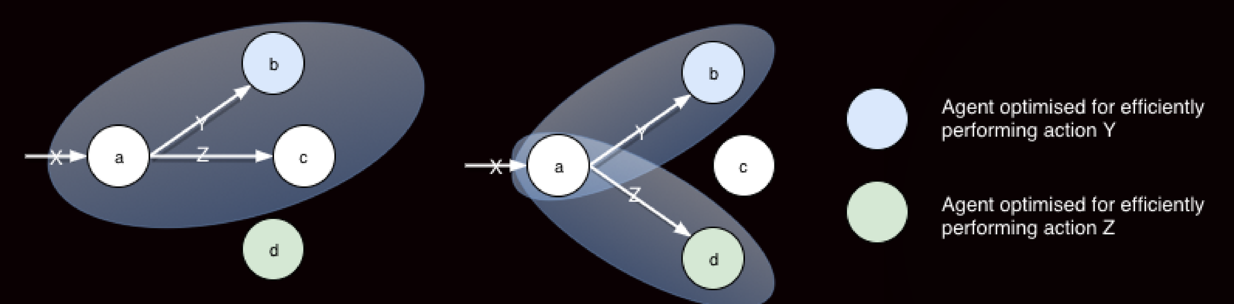

Figure 1: Local neighbourhood adaptation allows an agent to learn better connectivity and other agents to collaborate with within the system space to achieve the sub-tasks it needs them to achieve to accomplish its overall goal.

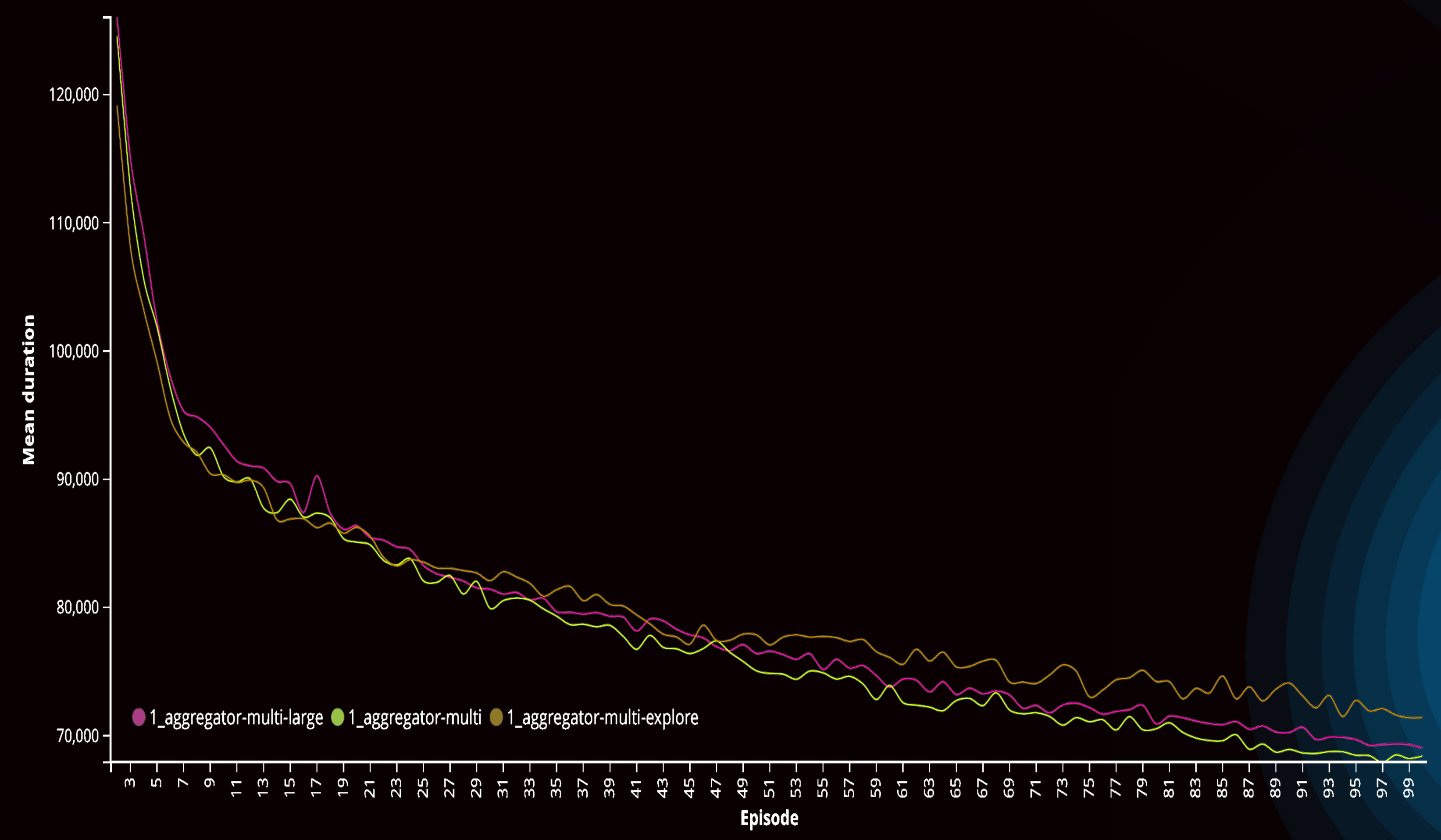

Figure 2: The average aggregating agent return times for datasets. Simulations show an approximately a 40% drop in the request durations as the agents react to rewards, explore their state-action spaces, and learn to adapt their local neighbourhoods to give increasingly performant links to the needed data.

PRIORITISATION BY FUNCTIONAL APPROXIMATION

Approaching the communication challenge from the opposite perspective, here agents learn to prioritise their responses to requests to optimse sub-task completion. The Multi-Channel Priority Optimisation through Function Approximation Algorithm (MC-POFA) allows agents to flexibly prioritise actions to satisfy responses in a way that preserves knowledge over incoming channels and shrink or grow the function approximations capacity to handle a broad spectrum of request demand. The adaptive capacity functionality is a cornerstone of the agents learning scalability, ensuring it can focus on a manageable subset of prioritisations when placed in a highly communicative environment.

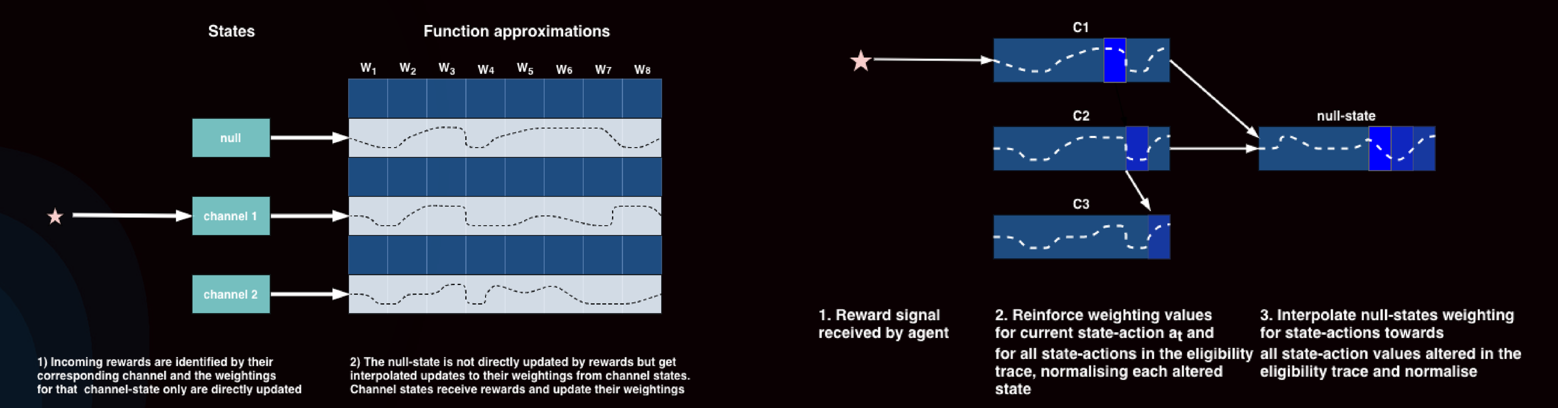

Figure 4: Use of channel-states mapped to function approximations for MC-POFA. Updates on a channels functional approximation weightings are used to update an aggregated functional approximation that is the basis for active prioritisation.

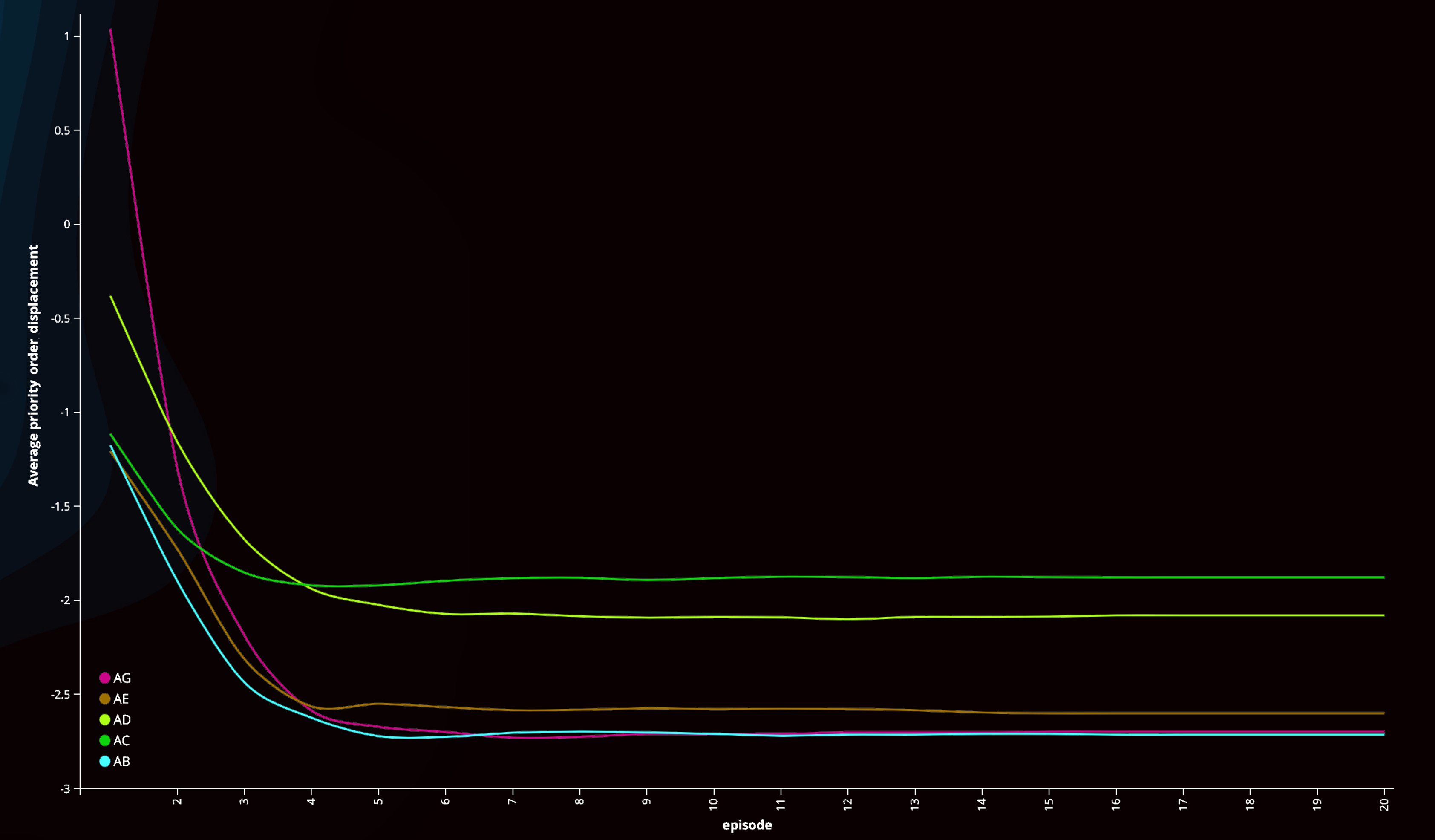

Figure 5: The prioritisation of certain data pieces is shown as the ones actively requested are increased in priority by varying amounts, dependent on the accumulated rewards obtained by the agent in responding to that particular request set.

RISK-ACTION PROBABILITIES

The Reward Trends for Risk-Action Probabilities (RT-RAP) algorithm we introduce combines a relative performance metric for an agent and a transformation function for its role to generate behaviours that dynamically alter its exploration and optimisation strategy. This allows an agent to use a comparison of its current learning policies performance against its historical reward trends to optimise and exploit subspaces of the systems state-action space without losing their flexbility to adapt in the face of variations and system disruption. This functionality is crucial to the agents ability to find an optmisation solution within the system, without it the agent will make too many changes to its connectivity and knowledge of other agents to realistically find exploitable knowledge within the system.

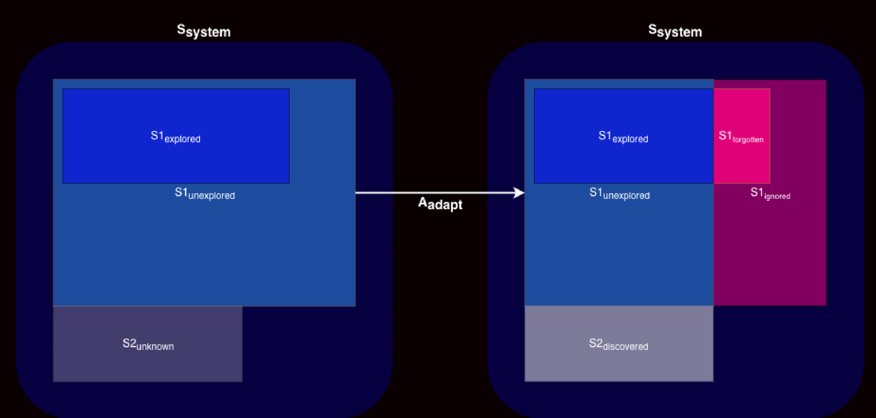

Figure 3: Adaptation actions in state-action space. Taking a state-action space altering action opens up new areas for exploration and exploitation, however, resource constraints mean that the loss of knowledge previously accumulated is inevitable.

ENVIRONMENTAL SENSOR NETWORKS

We target combining all these elements into a generic universal agent system that can be applied to real-world problems in a highly adaptive way. This will be simulated as an Environmental-Wireless Sensor Network (E-WSN) application in a harsh environment where resilience and self-organisation will be key to success and feasibility. To make the simulation realistic, we focus on long-term monitoring of radioactivity where conditions preclude human interference and will degrade agents in the field such as that of Chernobyls radioactive contamination. With a UAV deploying a large number of sensors over a disperse and remote geographical area, leading to a relatively ad-hoc, randomised placement of devices, using solar power cells to maintain enough energy to power themselves over a number of years.

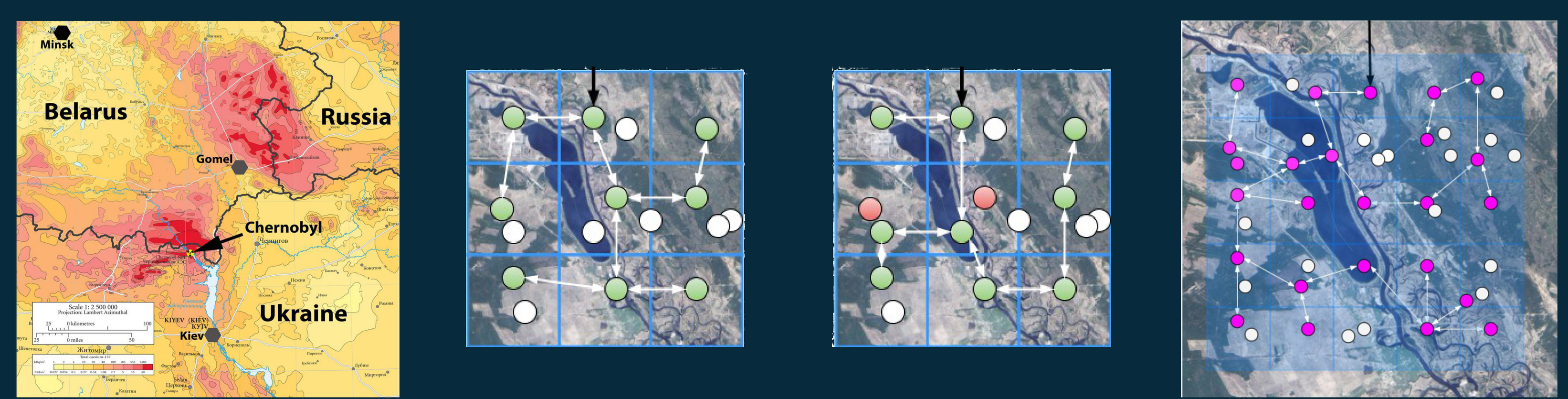

Figure 7a: The highly radioactive isotope caesium-137 has a half-life of around 30 years.

Figure 7b: Adaptation of each agents local neighbourhood as they individually react to device failures.

Figure 7c: Agents gain knowledge of the system space and form links to new agents.

ENVIRONMENT SIGNAL Q-TRANSFORMATIONS

We tackle the problem of driving behavioural change of agents dependent on how well its performing through the Relative Environmental Signals Q-Transformation Algorithm (RES-QT). At a high level this algorithm generates an agent-specific internal metric based on a combination of its rewards and the entropy of its learned knowledge over its current state-action space. The internal reward signal is compared against similar metrics collected from the agents local neighbourhood through informational communications. This then allows the transformation of an agents state-action space Q-values based on the agents view of its relative performance, driving risk-taking or conservative behaviours in response its belief in its success.

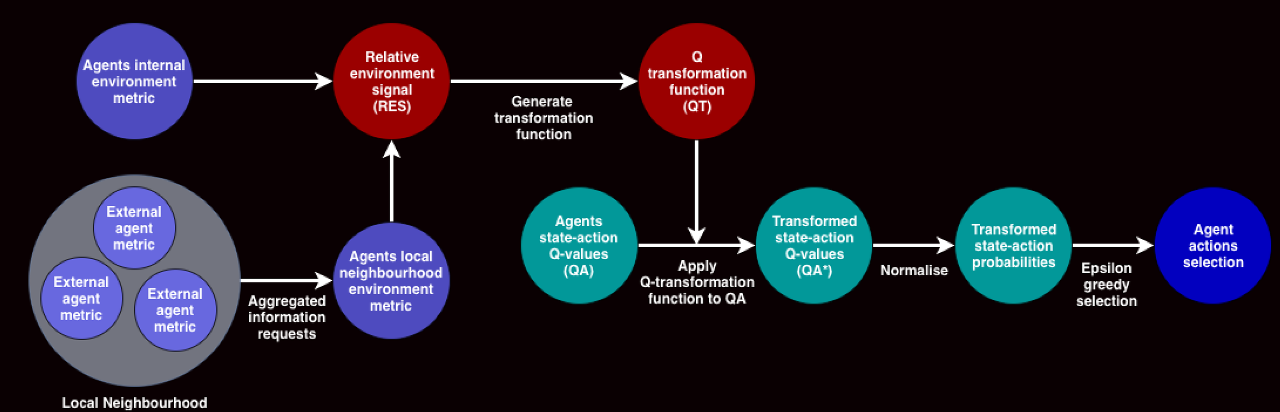

Figure 6: Relative Environmental Signals Q-Transformation Algorithm (RES-QT) uses metrics aggregated from the other agents in the local neighbourhood compared with an agents own internal metric to give the relative environment signal used to generate the Q-value transformation function that adapts the agents behaviours