

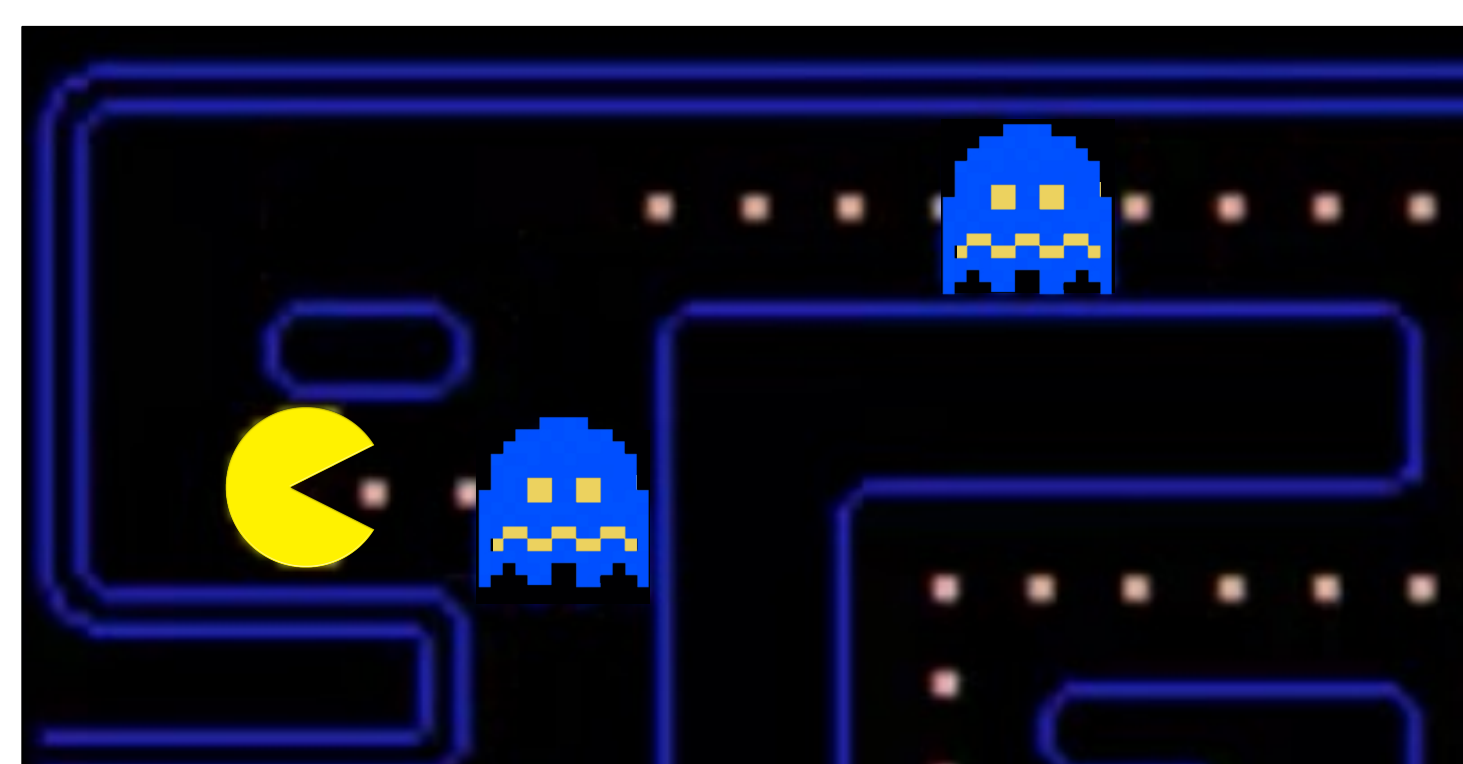
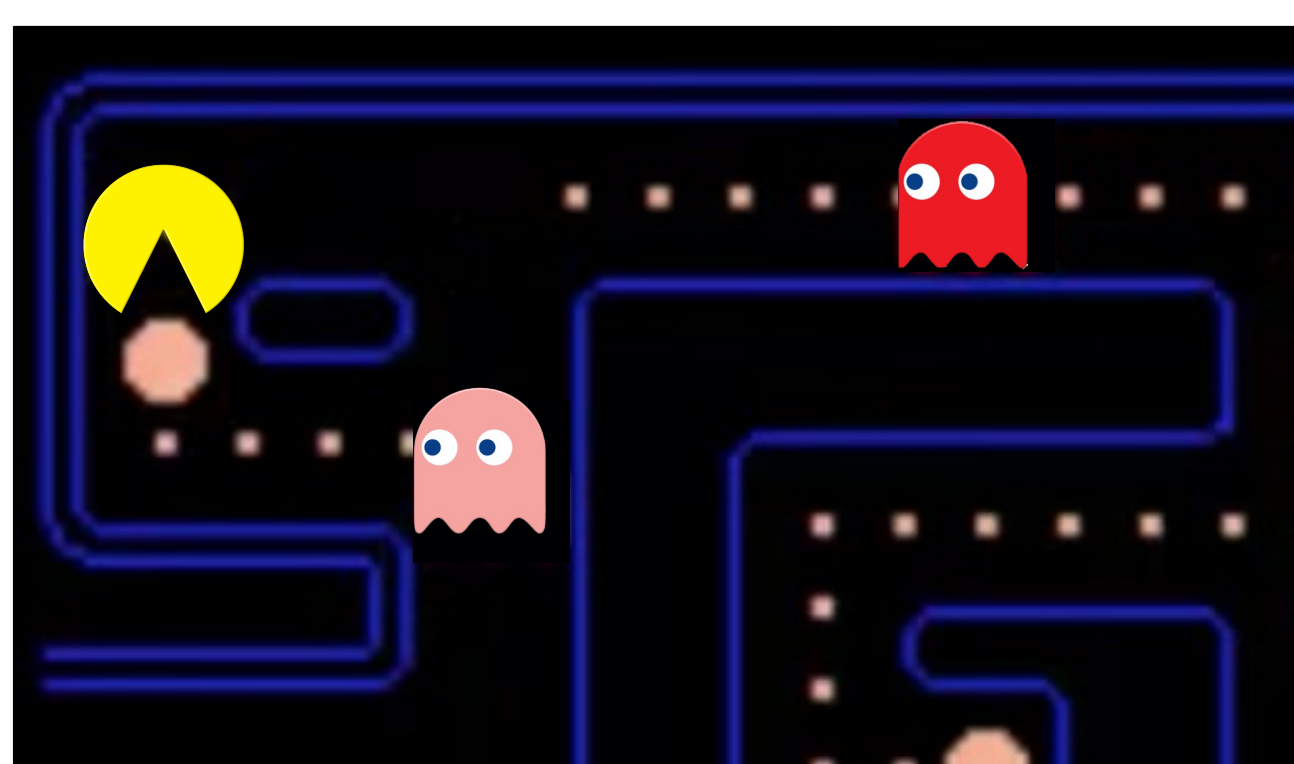
# STRATEGY EXTRACTION FOR TRANSFER IN AI AGENTS

Archana Vadakattu | vadakattua@unimelb.edu.au

Suppose you are an **AI agent** trained to play the classic arcade game, **Pacman**. Your goal is to **win the game** and to do this you will need to know how to **defeat the ghosts**.

From training, you learnt to defeat ghosts by eating the larger dots. Using this experience, the agent generalises and forms a strategy:

**"Collect an item to defeat an enemy"**



## Definition

A **STRATEGY** is a partial sequence of events - where an event is both the result of an agent's action and changes in state - to reach some predefined event of interest.

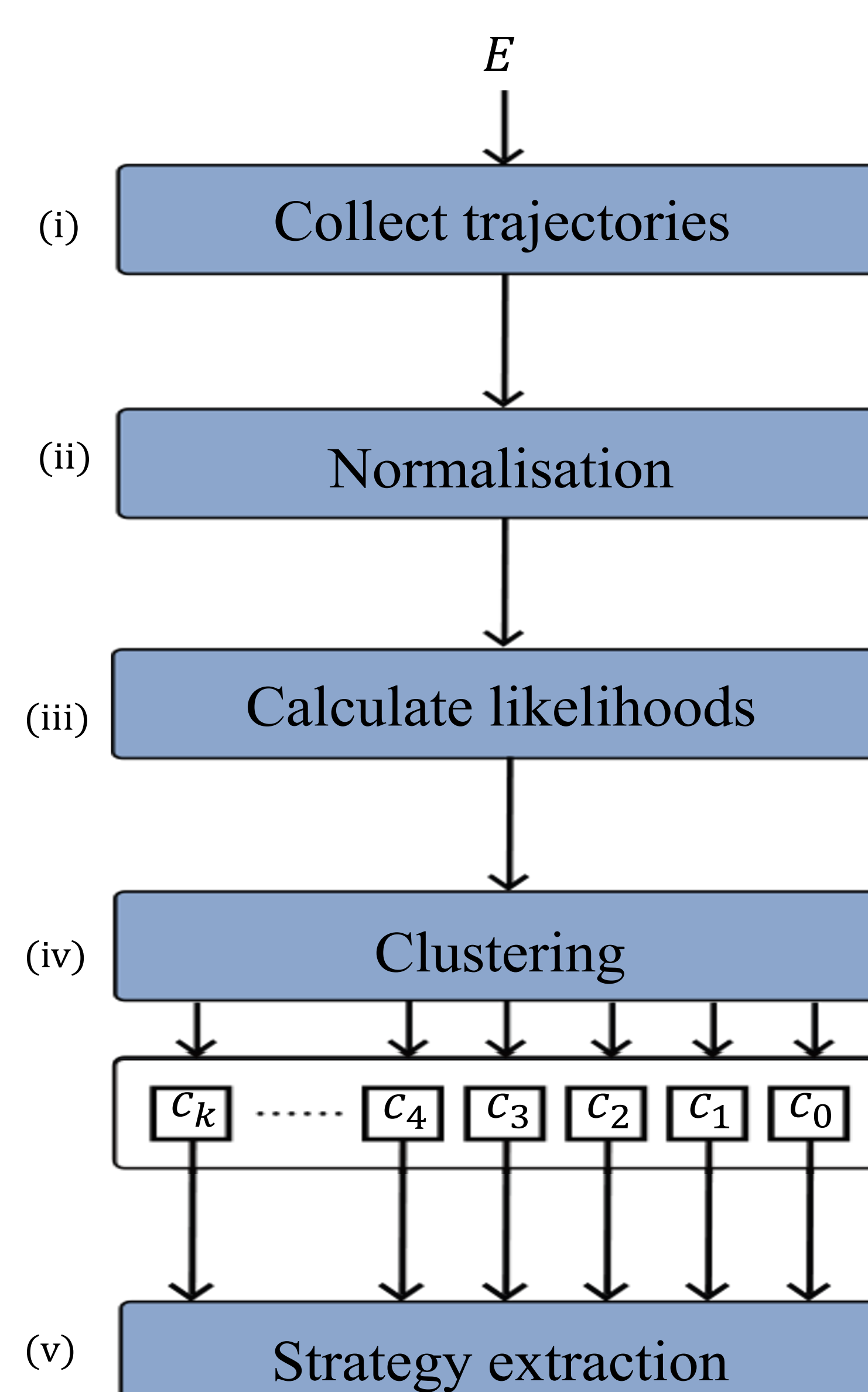
## Motivation

- Inspired by human cognition, our aim is to **extract strategies from an agent's existing knowledge** that can be applied in multiple contexts.
- Other methods find complete solutions - even if content is abstracted, they are highly unlikely to be applicable in different contexts. Our definition includes **plans with a partial ordering** and **potentially unnecessary events**.
- We envision the strategy extraction approach as the **first step towards generalised transfer**.

## Contribution

A unique approach to strategy extraction by treating the problem as a sequential pattern mining task.

## Method & Experiments



- Using sequence analysis methods to locate similar regions in sequences of event trajectories.
- Performance is evaluated on 3 custom video game environments.
- Our results, showcase the ability of this method to identify reasonable strategy candidates in different contexts.

## Future Work

- Use **generalisation techniques to support transfer** to domains with differing action and state spaces and environmental dynamics.
- Examine the extent to which strategy transfer improves **policies learnt, learning time and real time performance**.

Scan here  
for my  
personal  
website

