Cognitively-Inspired Salience Computation for Intelligent Agents

Sterling Somers
Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

Konstantinos Mitsopoulos
Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

Christian Lebiere
Carnegie Mellong University, Pittsburgh, Pennsylvania, United States

Robert Thomson
United States Military Academy , West Point, New York, United States

Abstract

We describe a method for determining feature salience of action decisions in intelligent agents based on cognitively-inspired salience. Salience is defined as the degree of influence that a factor has on a given decision. This is generated by having a cognitive model using instance-based learning theory to mirror the actions of an intelligent agent, and then determining which features most uniquely contributed to the actions of the agent. We present three examples of this salience techniques, including reinforcement learning agents based in the StarCraft II and autonomous drone domains, as well as part of a risk assessment model. A benefit of our method is that it does not rely on a specific implementation of an agent, it only requires the underlying decision feature-space. It is also capable of utilizing features at different levels of abstraction