

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Cognitively-Inspired Saliency Computation for Intelligent Agents

Permalink

<https://escholarship.org/uc/item/2h89m785>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors

Somers, Sterling

Mitsopoulos, Konstantinos

Lebiere, Christian

et al.

Publication Date

2019

Peer reviewed

Cognitively-Inspired Saliency 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 Mellon 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 saliency of action decisions in intelligent agents based on cognitively-inspired saliency. Saliency 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 saliency 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