World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:13, No:06, 2019

Cognitive Dissonance in Robots: A Computational Architecture for Emotional Influence on the Belief System

Authors : Nicolas M. Beleski, Gustavo A. G. Lugo

Abstract : Robotic agents are taking more and increasingly important roles in society. In order to make these robots and agents more autonomous and efficient, their systems have grown to be considerably complex and convoluted. This growth in complexity has led recent researchers to investigate forms to explain the AI behavior behind these systems in search for more trustworthy interactions. A current problem in explainable AI is the inner workings with the logic inference process and how to conduct a sensibility analysis of the process of valuation and alteration of beliefs. In a social HRI (human-robot interaction) setup, theory of mind is crucial to ease the intentionality gap and to achieve that we should be able to infer over observed human behaviors, such as cases of cognitive dissonance. One specific case inspired in human cognition is the role emotions play on our belief system and the effects caused when observed behavior does not match the expected outcome. In such scenarios emotions can make a person wrongly assume the antecedent P for an observed consequent Q, and as a result, incorrectly assert that P is true. This form of cognitive dissonance where an unproven cause is taken as truth induces changes in the belief base which can directly affect future decisions and actions. If we aim to be inspired by human thoughts in order to apply levels of theory of mind to these artificial agents, we must find the conditions to replicate these observable cognitive mechanisms. To achieve this, a computational architecture is proposed to model the modulation effect emotions have on the belief system and how it affects logic inference process and consequently the decision making of an agent. To validate the model, an experiment based on the prisoner's dilemma is currently under development. The hypothesis to be tested involves two main points: how emotions, modeled as internal argument strength modulators, can alter inference outcomes, and how can explainable outcomes be produced under specific forms of cognitive dissonance.

Keywords : cognitive architecture, cognitive dissonance, explainable ai, sensitivity analysis, theory of mind
Conference Title : ICSR 2019 : International Conference on Social Robotics
Conference Location : New York, United States
Conference Dates : June 04-05, 2019