Artificial Intelligence 230 (2016) 134–172

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Affect control processes: Intelligent affective interaction using a partially observable Markov decision process

Jesse Hoey [a,*], Tobias Schröder [b,c], Areej Alhothali [a]

[a] David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
[b] Centre for Theoretical Neuroscience, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada
[c] Potsdam University of Applied Sciences, Institute for Urban Futures, Kiepenheuerallee 5, 14469 Potsdam, Germany

ARTICLE INFO

Article history:
Received 2 April 2014
Received in revised form 12 August 2015
Accepted 17 September 2015
Available online 3 October 2015

Keywords:
Affect
Emotion
Sociology
Affect control theory
Markov decision process
Intelligent tutoring system
Assistive technology
Human–computer interaction

ABSTRACT

This paper describes a novel method for building affectively intelligent human-interactive agents. The method is based on a key sociological insight that has been developed and extensively verified over the last twenty years, but has yet to make an impact in artificial intelligence. The insight is that resource bounded humans will, by default, act to maintain affective consistency. Humans have culturally shared *fundamental* affective sentiments about identities, behaviours, and objects, and they act so that the *transient* affective sentiments created during interactions confirm the fundamental sentiments. Humans seek and create situations that confirm or are consistent with, and avoid and suppress situations that disconfirm or are inconsistent with, their culturally shared affective sentiments. This "*affect control principle*" has been shown to be a powerful predictor of human behaviour. In this paper, we present a probabilistic and decision-theoretic generalisation of this principle, and we demonstrate how it can be leveraged to build affectively intelligent artificial agents. The new model, called *BayesAct*, can maintain multiple hypotheses about sentiments simultaneously as a probability distribution, and can make use of an explicit utility function to make value-directed action choices. This allows the model to generate affectively intelligent interactions with people by learning about their identity, predicting their behaviours using the affect control principle, and taking actions that are simultaneously goal-directed and affect-sensitive. We demonstrate this generalisation with a set of simulations. We then show how our model can be used as an emotional "plug-in" for artificially intelligent systems that interact with humans in two different settings: an exam practice assistant (tutor) and an assistive device for persons with a cognitive disability.

1. Introduction

Designers of intelligent systems have increasingly attended to theories of human emotion, in order to build software interfaces that allow users to experience naturalistic flows of communication with the computer. This endeavour requires a comprehensive mathematical representation of the relations between affective states and actions that captures, ideally, the subtle cultural rules underlying human communication and emotional experience. In this paper, we argue that Affect Control Theory (ACT), a mathematically formalized theory of the interplays between cultural representations, interactants'

* Corresponding author.
E-mail addresses: jhoey@cs.uwaterloo.ca (J. Hoey), post@tobiasschroeder.de (T. Schröder), aalhothal@cs.uwaterloo.ca (A. Alhothali).

http://dx.doi.org/10.1016/j.artint.2015.09.004

identities,[1] and affective experience [1], is a suitable framework for developing emotionally intelligent agents. To accomplish this, we propose a probabilistic and decision theoretic generalisation of ACT, called *BayesAct*, which we argue is more flexible than the original statement of the theory for the purpose of modelling human–computer interaction. *BayesAct* is formulated as a partially observable Markov decision process or POMDP. The key contributions of this new theory are: (1) to represent sentiments as probability distributions over a continuous affective space, thereby allowing these sentiments to be dynamic and uncertain; (2) to propose a new kind of agent based on affect control theory that has the ability to learn affective identities of interactants; (3) to integrate the affective dynamics proposed by affect control theory with standard POMDP-based artificial intelligence; and (4) to introduce explicit utility functions to affect control theory that parsimoniously trade-off affective and propositional goals for a human-interactive agent. These contributions allow *BayesAct* to be used as an artificially intelligent agent: they provide the computerised *agent* with a mechanism for predicting how the affective state of an interaction will progress (based on affect control theory) and how this will modify the object of the interaction (e.g. the software application being used). The *agent* can then select its strategy of action in order to maximize the expected values of the outcomes based both on the application state and on its affective alignment with the human.

Affect control theory arises from the long tradition of symbolic interactionism that began almost three hundred years ago with the insights of Adam Smith [2] into the self as a mirror of the society in which it is embedded: the so-called looking-glass self [2]. These insights eventually led to the modern development of structural symbolic interactionism through Mead, Cooley, and Stryker [3], and culminating in Heise's affect control theory (ACT) [1], which this paper extends. Although ACT, and symbolic interactionism in general, are very well established theories in sociology, they have had little or no impact in artificial intelligence. This paper is the first to propose affect control theory as a fundamental substrate for intelligent agents, by elaborating a POMDP-based formulation of the underlying symbolic interactionist ideas. This new theory allows ACT to be used in goal-directed human-interactive systems, and thereby allows A.I. researchers to connect to over fifty years of sociological research on cultural sentiment sharing and emotional intelligence. The theory also contributes a generalisation of affect control theory that we expect will lead to novel developments in sociology, social psychology, and in the emerging field of computational social science [4].

The main contribution of this paper is therefore of a theoretical nature, which we demonstrate in simulation. We have also implemented the theory in a simple tutoring system and in an assistive technology that is designed to assist persons with dementia. We report the results of an empirical survey and demonstrative study with human participants in the case of the tutoring system. The assistive technology is further described in [5]. Therein, a prompting system delivers audio-visual cues to a person using a variety of different affective "styles". The mapping from non-verbal behaviours of the user to the "style" of prompt is defined by *BayesAct* alone.

1.1. Model overview

BayesAct is a partially observable Markov decision process (POMDP, see Fig. 1(a) and Section 2.2) model of an agent interacting with an environment. The environment is modelled, as usual in a POMDP, with a set of states, $\mathbf{X}$. A *BayesAct* agent has actions, A, available to it, and these actions change the state of the environment according to a stochastic transition function. The environment model (states) are not assumed to be observable (they are latent), but the agent has access to a set of observations $\boldsymbol{\Omega}_x$, from which it can infer the state of the environment by using Bayes' rule and a stochastic observation function that relates states to observations. Finally, a utility function, R, describes the preferences of the agent on a numerical scale. The utility function can be used by a Bayesian (sequential) decision maker to optimize decisions (action choices) in the long term.

BayesAct is modelling the case where the environment contains humans (or other *BayesAct* agents) who are partially responsible for the state dynamics. *BayesAct* therefore includes a latent *user model* as part of its state space (shown as factor $\mathbf{Y}$ in Fig. 1(a)). The user model describes the identity (see footnote 1) of the agent and of the human it is interacting with, and conditions (stochastically) the dynamics of the state.

The identities are modelled as four concurrently evolving discrete-time non-linear dynamical systems over a three dimensional continuous affective space. The three dimensions are: evaluation (how good/bad something is), potency (how strong/weak), and activity (how active/passive). The space is referred to as "EPA" space, and it has been found by sociologists to capture over 80% of the variance in affective meanings ascribed by humans across cultures and languages [6], and is in some sense "fundamental" to human emotion (see Section 2.1). It has also been used by other works in affective computing (where it is referred to as "PAD" space or Pleasure–Arousal–Dominance, see Section 2.3).

BayesAct departs from other works on affective computing because it also includes the dynamics of identities in the EPA space. These dynamics are learned from datasets of human sentiments about events, measured during decades of research by sociologists in different cultures around the world, and forming part of a sociological theory called *affect control theory* (ACT) ([1]; see Section 2.1). As the EPA space, the dynamics are found to be culturally stable and consistent [7]. The dynamics form part of the transition function (for the identities, $\mathbf{Y}$) in the POMDP (see Section 3.2). The dynamics relate an agent's stable (through time), culturally shared affective sentiments about itself and about other agents ($\mathbf{f}$), to the transient sentiments

[1] The meaning of the term *identity* differs considerably across scientific disciplines. Here, we adhere to the tradition in sociology where it essentially denotes a kind of person in a social situation.