Neurocomputing 221 (2017) 32–47

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A Situation-Aware Fear Learning (SAFEL) model for robots

Caroline Rizzi[a,*], Colin G. Johnson[a], Fabio Fabris[a], Patricia A. Vargas[b]

[a] School of Computing, University of Kent, Canterbury, UK
[b] Robotics Laboratory, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK

ARTICLE INFO

Communicated by Bin He

Keywords:
Contextual fear conditioning
Brain Emotional Learning
Temporal pattern
Affective computing
Autonomous robotics
Amygdala and hippocampus modeling

ABSTRACT

This work proposes a novel Situation-Aware FEar Learning (SAFEL) model for robots. SAFEL combines concepts of situation-aware expert systems with well-known neuroscientific findings on the brain fear-learning mechanism to allow companion robots to predict undesirable or threatening situations based on past experiences. One of the main objectives is to allow robots to learn complex temporal patterns of sensed environmental stimuli and create a representation of these patterns. This memory can be later associated with a negative or positive "emotion", analogous to fear and confidence. Experiments with a real robot demonstrated SAFEL's success in generating contextual fear conditioning behavior with predictive capabilities based on situational information.

1. Introduction

Learning to fear unpleasant or harmful stimuli from the environment is ubiquitous in nature. Fear can be defined as a brain's mechanism for automatic learning and memorization of potential threats to one's survival. It offers exceptional advantages over conscious-rational thinking during critical situations due to its involuntary and automatic responses, leading to faster decision-making and reaction in the face of danger [1,2], as well as increased focus and attention [3]. Fear learning is also an important ally for environmental adaptation as the brain constantly associates fear with newly experienced dangers. Hence, it assists animals to learn and react to the new patterns and threats of unfamiliar environments.

Fear learning supports not only survival and environmental adaptation, but also social adaptation (i.e., one's ability of adjusting its behavior to the rules of its own society). The concept of society applies to many animal species, where individuals feel an instinctive need to be accepted by others of its kind. As belonging to a community can highly increase one's chances of survival, the brain of many animal species evolved to process social rejection as an aversive environmental stimulus. Consequently, the brain triggers fear learning when an individual observes disapproval from others towards its actions.

By being real agents that inhabit the physical world and interact with human beings, autonomous robots are also susceptible to environmental threats and to social adaptation. Hence, autonomous robots could also take advantage of a mechanism inspired by fear learning. Robot companions [4–7], for instance, are gaining more space in our society as social entities and have shown a great potential for applications in many areas (e.g., healthcare [8]). However, a common issue with long-term robot companions is the rapid loss of interest from their users, who get frustrated and lose motivation over time as companions continue to perform pre-defined and repetitive behaviors [5]. This poses a challenge to the broad development and practical use of robot companions.

From the HRI (Human-Robot Interaction) point of view, robots' social interaction becomes more believable and natural as they become more adaptable and responsive to environmental cues [9,4,6]. As humans, we expect others to be able to identify environmental factors that can represent unpleasantness or danger to themselves and act accordingly. Therefore, being able to properly express fear responses could highly increase the believability of a long-term robot companion [9].

Fear learning has been a strong source of inspiration for developing more flexible and adaptive artificial intelligence [10–13]. The potential of artificial intelligence based on fear-learning models is demonstrated by its successful contribution to a variety of engineering and robotic applications [14–29]. Despite its advances, research on artificial fear-learning is still in its infancy and has several aspects with margin for improvement, among which we can highlight *situation appraisal*.

In the real world, people react not only to individual environmental stimuli (e.g. pain, smells, noises, location, light levels, etc.), but also to contextual variation over time, also known as *situation*, which is characterized by the temporal order and intensity variation of all appraised stimuli in a given period of time (e.g., being in a forest at night, with impaired visibility, and hearing animals' noises). Here, we define the emotional outcome and evaluation of a situation as *situation*

* Corresponding author.
E-mail addresses: cr519@kent.ac.uk (C. Rizzi), C.G.Johnson@kent.ac.uk (C.G. Johnson), ff79@kent.ac.uk (F. Fabris), p.a.vargas@hw.ac.uk (P.A. Vargas).

http://dx.doi.org/10.1016/j.neucom.2016.09.035
Received 18 December 2015; Received in revised form 13 July 2016; Accepted 14 September 2016
Available online 20 September 2016
0925-2312/

appraisal.

To the best of our knowledge, artificial fear-learning models proposed to date do not substantially address situation appraisal, which is a significant part of the brain's fear-learning system, and essential for a organism to predict outcomes and adapt to threats and environmental changes [30].

This paper proposes a novel hybrid computational model, named SAFEL (Situation-Aware FEar Learning), which is based on the brain's fear-learning system and incorporates the concept of situation awareness from expert systems. SAFEL builds on our fear-learning model, proposed in [31], which is inspired by three brain regions essential in fear learning: the sensory system, the amygdala and the hippocampus, along with a cognitive function of the brain known as the working memory [2]. Here, we discuss the implementation of SAFEL's hippocampus and working memory modules, which are responsible for simulating situation appraisal regarding fear. Experiments with a NAO robot demonstrate that SAFEL has successfully generated fear-conditioning behavior with predictive capabilities based on situational information.

The main contributions of this work as compared to the state of the art are:

1. Integration of a fear learning model with the concept of temporal context. SAFEL performs threat predictions based on complex temporal and contextual information. Existing fear memory models either focus in the contextual or the temporal aspect, overlooking the need of both skills for an artificial intelligent agent to properly react to real-world threatening situations.
2. SAFEL is focused on real-world applications for artificial and autonomous intelligence in robotics. Many existing fear-learning models that are inspired by the real mechanisms of the brain focus on providing a close-to-real emulation of brain functions without addressing the practical usage of the model for artificial intelligence.
3. The successful integration of a symbolic rule-based platform for situation management with a classification algorithm for memorizing and predicting threats based on complex temporal context.

This paper is organized as follows: Section 2 discusses related work. Section 3 summarizes the biological background and neuroscientific findings that have inspired SAFEL. Section 4 presents SAFEL's modeling and implementation. Experimental methodology and results are discussed in Section 5 and Section 6, respectively. The paper concludes with Section 8, and also suggests future work.

2. Previous models of contextual fear conditioning

The idea of using models of emotion for improving autonomous learning in artificial systems started with Picard's research in 1995 [32,33]. Picard's work originated one of the most recent branches of computer science: *affective computing*. According to Picard [33], affective computing tackles three aspects of artificial intelligence: (1) the ability of machines to recognize and express emotions, (2) the ability of machines to respond intelligently to human emotion, and (3) the capability of machines to regulate and utilize emotions in order to behave more intelligently and effectively. In this work, we focus on the latter aspect of affective computing, though all the three aspects are indirectly addressed.

A large range of approaches have been proposed for simulating emotions in artificial agents, such as affective space models [34,35], motivation-driven models [13], neuro-inspired models [10,12,36–38], hormonal or homeostatic systems [39–42], among others [43,44] (for a broader review on the varied approaches and challenges of affective computing, we refer the reader to [45]). Here, we are particularly interested in approaches addressing the temporal properties of context applied to fear conditioning for providing robots with fast, efficient and flexible decision-making.

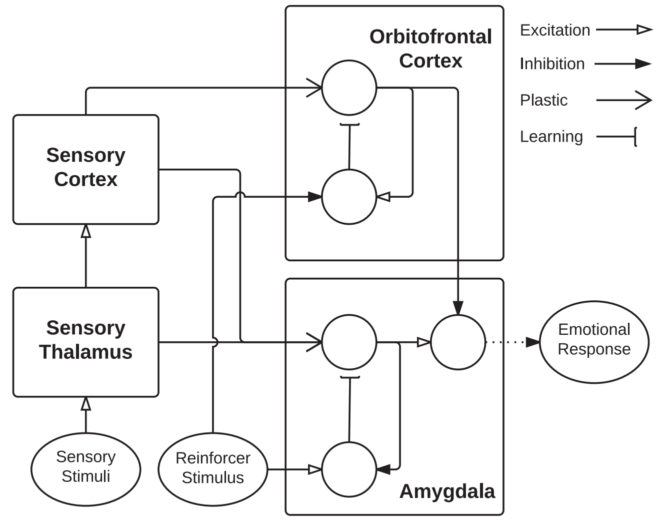

Fig. 1. Fear-learning model proposed by Morén and Balkenius [10]. Each component of their model represents an ANN. Circles represent individual ANNs internal to the respective component.

One of the most influential works in artificial fear conditioning is the brain emotional learning (BEL) model, proposed by Morén and Balkenius [10]. Their model (Fig. 1) consists of interconnected modules of artificial neural networks (ANNs) that simulate the role of neural circuitries involved in fear learning. It receives two types of inputs – environmental neutral stimuli and a reward signal – that are processed by four simulated neural regions: the thalamus, the sensory cortex, the amygdala and the orbitofrontal cortex.

The thalamus and sensory cortex simply relay input information to the orbitofrontal cortex and amygdala and, together, compose the “low and high roads” to the amygdala, respectively [2]. The sensory cortex receives information from the thalamus, which in turn receives information directly from the environment. As the thalamic pathway is shorter, it provides the amygdala with low latency information about environmental stimuli. On the other hand, information projected through the thalamic-cortical pathway takes longer to reach the amygdala, but provides a higher-level and more accurate representation of the sensed world.

The amygdala is responsible for assessing and predicting the emotional value of stimuli, based on the significance of the accompanied reward. Finally, the orbitofrontal cortex is responsible for inhibiting emotional associations of the amygdala that are no longer valid. This model has been tested for the most basic effects of classical conditioning – such as *fear acquisition*, *fear extinction*, *blocking*, *habituation* and *spontaneous recovery* – showing satisfactory results.

The BEL model was later improved in [46], with the addition of a module that simulates the contextual processing performed by the brain's hippocampal regions. BEL's hippocampus module has four main components: the *Bind subsystem*, the *Mem system*, the *Match system* and the *Context system*. The Bind subsystem is responsible for binding stimuli that are simultaneously detected. The Mem system generates expectations about stimuli manifestation at specific locations. These expectations are later compared with the actual stimuli in the Match system. Lastly, the Context system combines information from the Match and Bind systems to generate a contextual code that feeds the amygdala and orbitofrontal cortex.

With the aid of the hippocampal module, BEL is able to express fear responses based on contextual information. For example, one of the experiments performed in [46] consisted on presenting two different stimuli, CS0 and CS1, sometimes separately and sometimes together. All single presentations of either CS0 or CS1 were followed by a reinforcing signal, whereas all simultaneous presentations were followed by nothing. The model gradually learned to differentiate between