http://dx.doi.org/10.5755/j01.eee.18.10.3060 ELEKTRONIKA IR ELEKTROTECHNIKA, ISSN 1392-1215, VOL. 18, NO. 10, 2012

Cloud Interconnected Affect Reward based Automation Ambient Comfort Controller

D. Drungilas[1], A. A. Bielskis[2]
[1]*Institute of Mathematics and Informatics, Vilnius University, Akademijos St. 4, Vilnius, LT-08663, Lithuania*
[2] *Department of Electrical Engineering, University of Klaipeda, Bijunu St. 17, LT-91225 Klaipeda, Lithuania*
dorition@gmail.com

***Abstract*—The paper presents the human Affect Reward Based Automation Ambient Comfort Controller (ACARBC) as the interconnected cloud computing intelligent services that provide intelligent calculus for any instrumented interconnected environment sense and control system. The ACARBC has been modelled and, as the experimental results show, that an environmental state characteristics that create an optimum ambient comfort can be obtained by ACAR index. The ACAR index is dependent on human physiological parameters: the temperature, the ECG- electrocardiogram and the EDA-electro-dermal activity. The fuzzy logic is used to approximate the ACAR index function by defining two fuzzy inference systems: the Arousal-Valence System, and the Ambient Comfort Affect Reward (ACAR) System. The Radial Basis Neural Network is used as the main component of the ACARBC to performing of two roles - the policy structure, known as the Actor, used to select actions, and the estimated value function, known as the Critic that criticizes the actions made by the Actor. The Critic in this paper was used as a value function approximation of the continuous learning tasks of the ACARBC.**

***Index Terms*—Ambient intelligence, automatic control, emotion recognition, human computer interaction, machine learning.**

I. INTRODUCTION

The IBM vision of the smarter home enabled by cloud technology shows that such “smarter home” becomes “instrumented, interconnected and intelligent”. “Today’s Internet of people is evolving into an “Internet of things”, and by 2013, 1.2 billion connected consumer electronics devices are expected in the more than 800 million homes with broadband connections”. Compared with previous attempts to enable the “smart home,” where the intelligence was based on centralized control through a home server or gateway, the intelligence and with it the complexity in the new smarter home is moved out from the home onto the network, or more precisely the Internet cloud. By IBM, “*Instrumented* is the ability to sense and monitor changing conditions”. Instrumented devices provide increasingly detailed information and control about their own functioning and also provide information about the environment in which they operate. “*Interconnected* is the ability to communicate and interact, with people, systems and other objects”. Interconnected devices make possible remote access to information about a device and control of the device. This enables services throughout the Internet, removing complexity from the home and lowering costs for the service providers. At the same time, it supports the aggregation of information and control of devices throughout the network. This means that consumers can get a consistent view of their devices, both from home and from mobile devices. For service providers, it provides an aggregate view of customer characteristics according to criteria such as geographic location, consumption patterns, or types of service. “*Intelligent* is the ability to make decisions based on data, leading to better outcomes”. Intelligent devices support the optimization of their use, both for the individual consumer and for the service provider. For instance, “a utility can send signals to consumers’ homes to manage discretionary energy use in order to reduce peak loads. By coordinating this process throughout an entire service area, the utility can optimize the peak reduction, while saving the consumers money on their bill”.

Inspired by investigations of thermal comfort, indoor air quality and adequate illuminance by using the Predicted Mean Vote Index (PMV) [1]–[3], the human Ambient Comfort Affect Reward, the *ACAR* index is proposed for automatic quality control of heating/ventilation and lighting automation devices [4], [5]. This controller is planned to be used to improve energy savings in the sustainable environment. Specifically, it predicts the indoor lighting, heating and air quality conditions at a given time by measuring integrated ACAR index that defines ambient comfort affect to the human. Principles of development of the intelligent cloud services of Ambient Comfort Affect Reward Based heating, ventilation, and air conditioning Controller, the *ACARBC* are described in this paper.

II. MODEL OF THE AMBIENT COMFORT AFFECT REWARD BASED HEATING, VENTILATION, AND AIR CONDITIONING CONTROLLER

The developing process of the smart environment is based on automatic control which adopts environment by smart sensing of human physiological signals. The architecture of

Manuscript received March 10, 2012; accepted May 13, 2012.

the Ambient Comfort Affect Reward Based ambient comfort control system as shown in Fig. 1 can be used to find such environmental state characteristics that create an optimal comfort for people affected by this environment. The *ACARBC* consists of three main parts: *the Instrumented Interconnected Environment Sense and Control System, the Intelligent System and the Interface.*

The *Instrumented Interconnected Environment Sense and Control System* is equipped with embedded smart devices, that can sense environment state – heating, lighting, air condition and respectively control the actuators changing ambient environment microclimate. As the human is the main criterion to evaluate the comfort of ambient environment there are instrumented bio-sensors used to read physiological signals of the human like electro dermal activity (EDA), electrocardiogram (ECG) and skin temperature as these physiological signals quite credibly characterizes human emotional state [4] i.e. it is possible to evaluate the affect of the ambient environment (temperature, lighting, air quality) to the human.

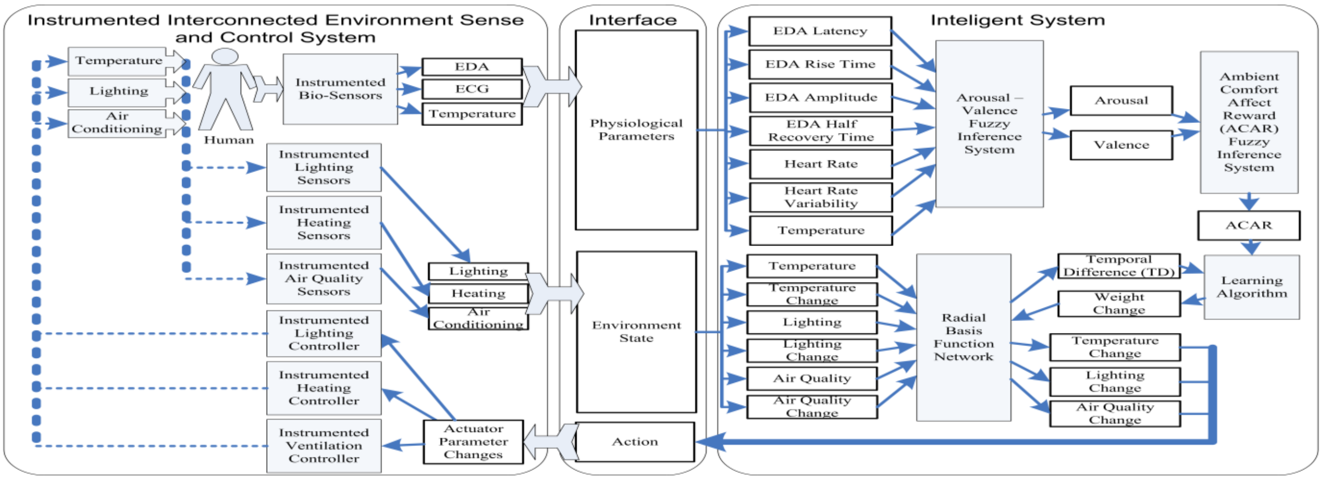


Fig. 1. The architecture of reinforcement learning based ambient comfort control system.

The *Interface* is responsible for data fluent exchange protocols between different platforms that implement *Instrumented Interconnected Environment Sense and Control System* and *Intelligent System*.

The *Intelligent System* is used for main calculations as decision support system for predicting action that changes the parameters of the environment actuators, by parameters of environmental state and human physiological signals. The reward index representing ambient comfort affect to the human is expressed as an Ambient Comfort Affect Reward, the *ACAR* index function

$$ACAR = f\{a(t,c,d), v(t,c,d)\}, \quad ACAR \in [-3,3], \qquad (1)$$

where *a* and *v* are arousal and valence functions respectively dependent on human physiological parameters: *t* – temperature, *c* – *ECG*, electrocardiogram and *d* – *EDA*, electro-dermal activity. It is shown [4], [5], that (1) type function can be approximated by neural networks, fuzzy logic or other regression methods. In this case, we use fuzzy logic to approximate (1) by defining two fuzzy inference systems: the *Arousal-Valence System*, and the *Ambient Comfort Affect Reward (ACAR) System* as shown in Fig. 1.

The Radial Basis Neural Network is the main component of *ACARBC* responsible for two roles – the policy structure, known as the *Actor*, used to select actions, and the estimated value function, known as the *Critic* that criticizes the actions made by the *Actor*. We use *Critic* as value function approximates for continuous learning tasks (like *ACARBC*), because discrete state representation of environment can be problematic. The continuous MDP can lose its Markov property if the state discretization is too coarse. As a consequence, there are states which are not distinguishable by the agent, but which have quite different effects on the agent's future. Using reinforcement learning for control tasks is a challenging problem, because we typically have continuous state and action spaces. For learning with continuous state and action space, a function approximation must be used. Linear function approximations are very popular in this problem area, because they can generalize better than discrete states and are also easy to learn at least when using local features [5]. A feature state consists of *N* features, each having an activation factor in the interval [0, 1]. Linear approximations calculate their function value by

$$f(x) = \sum_{i=1}^{N} \phi_i(x) \cdot w_i , \qquad (2)$$

where $\varphi(x)$ is the activation function and w_i is the weight of the feature *i*. Instead of keeping track of each unique state separately, we seek to find a function that approximates the state space with a small number of adjustable parameters.

Radial basis functions (RBFs) are the natural generalization of coarse coding to continuous-valued features. Rather than each feature being either *0* or *1*, it can be anything in the interval [0, 1], reflecting various degrees to which the feature is present. A typical RBF feature, *i* have a Gaussian (bell-shaped) response $\phi_s(i)$ dependent only on the distance between the state, *s* and prototypical or center state of the feature, *ci* and relative to the feature's width σ_i

$$\phi_s(i) = \exp\left(-\frac{\|s - c_i\|^2}{2\sigma_i^2}\right). \qquad (3)$$

The RBF network is a linear function approximation using RBFs for its features. The *Learning Algorithm* is used to adopt RBF network weights in order to fit *Actor* and *Critic* functions. The feature of the *Actor-Critic* learning is that the *Actor* learns the policy function and the *Critic* learns the value function using the Temporal Difference (TD*)* method simultaneously [5]. The *TD* error $\delta_{TD}(t)$ is calculated by the temporal difference of the value function between successive states in the state transition as

$$\delta_{TD}(t) = r(t) + \gamma V(t+1) - V(t), \tag{4}$$

where $r(t)$ is the external reinforcement reward signal, $0 < \gamma < 1$ denotes the discount factor that is used to determine the proportion of the delay to the future rewards. The *TD* error indicates, in fact, the goodness of the actual action. Therefore, the weight vector θ of the policy function and the value function are updated as

$$\vec{\theta}_{t+1} = \vec{\theta}_t + \alpha\delta_t \vec{e}_t, \tag{5}$$

where α is the learning rate and the eligibility trace, e can be calculated by:

$$\begin{cases} \vec{e}_0 = 0, \\ \vec{e}_{t+1} = \gamma\lambda \vec{e}_{t-1} + \nabla_{\vec{\theta}_t} V(t). \end{cases} \tag{6}$$

III. MODELLING CLOUD SERVICES FOR AMBIENT COMFORT AFFECT REWARD BASED CONTROLLER

The prediction of action by ambient environmental state and physiological parameters of the human is based on reinforcement learning and the model of *ACARBC* implemented as a cloud services is shown in Fig. 2.

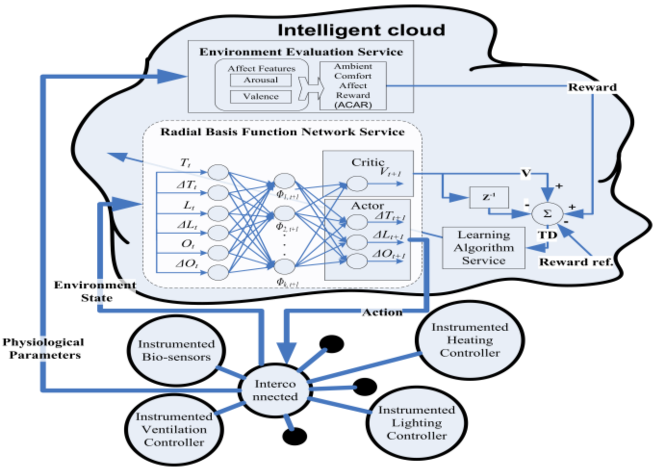


Fig. 2. Intelligent subsystem implementation as a cloud services for ambient comfort control.

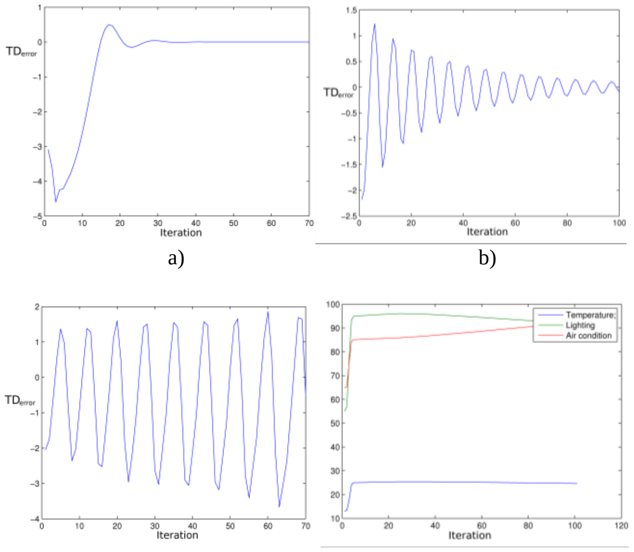
The implementation of main intelligent calculus of ACARBC as a cloud services reduces IT management cost, the complexity of performing tasks, increases efficiency of use provisioned resources, service accessibility flexibility, simplifies instrumented interconnected devices of entire ACAR system. The *ACARBC* consists of the following parts: the *Environment Evaluation Service*, the *Radial Basis Neural Network Service*, and the *Learning Algorithm Service.* Each service is implemented as described previously using fuzzy inference systems, RBF neural networks and learning algorithms. Using interface, the ACARBC services can be accessible for any instrumented interconnected environment sense and control system.

IV. EMPIRICAL RESULTS OF SIMULATION ACARBC

The main characteristic of *ACARBC* is the TD error $\delta_{TD}(t)$ dynamics. The optimum state is obtained if this parameter converged to 0 i.e. no need any change of the state. Using MatLab software tools we applied the proposed ambient comfort control algorithm to track the optimal environmental state. The pseudo code is shown in Fig. 3. The corresponding parameters (γ - discount factor, λ – decay factor, α – the learning rate) for the *ACARBC* are set for different experimental setups. The detailed simulation results are shown in Fig. 4. The Simulation results indicate that the proposed *ACARBC* can be stable ($\delta_{TD}(t)$ *– converges to 0*) using appropriate parameters (Fig. 4 (a, b)). As well as the $\delta_{TD}(t)$ –can diverge (Fig. 4(c)). As we can see in Fig. 4(d) starting from the initial state: temperature – 13, lighting – 55 and air conditioning – 65 the optimum state is obtained.

Input: $state_{t=0}$ *// the initial state*
Output: state trace
action trace
temporal difference trace
value trace
Initialize: *Δreward // the change of the reward*
$ARVA_{t=0}$ *// the initial arousal-valence value vector*
$ACAR_{t=0}$ *// the Ambient Comfort Affect Reward*
center // the centers of basis functions
σ *// the width of basis functions*
γ *// the discount factor*
λ *// the decay factor*
$e_{t=0}=0$ *// the eligibility trace*
α *// the learning rate*
v *// the RBF neural network weights of actor*
w *// the RBF neural network weights of critic*
FOR *each t^{th} observation* **DO**
FOR *each action* **DO**
Initialize: $V_j=0$
FOR *each j^{th} basis function center* **DO**
$$\phi_j^{critic} = \exp\left(-\frac{\|state_t - center_j\|^2}{2\sigma_j^2}\right)$$
$V_j=V_j+ \Phi_j^{critic} \cdot w_j$
ENDFOR
$e_{t+1}=\gamma \cdot \lambda \cdot e_t+ \Phi_j^{critic}$
$\delta_t= ACAR_t + V_{j-1} - \gamma \cdot V_j$ *// the temporal difference*
FOR *each basis function center j* **DO**
$w_j = w_j + \alpha \cdot \delta_t \cdot e_{t+1} \cdot \Delta reward$
ENDFOR
Initialize: $action_t=0$
FOR *each basis function center j* **DO**
$v_j= v_j + \alpha \cdot \delta_t \cdot e_{t+1} \cdot \Delta reward$
$$\phi_j^{actor} = \exp\left(-\frac{\|state_t - center_j\|^2}{2\sigma_j^2}\right)$$
$action_t= action_t + \Phi_j^{actor} \cdot v_j$;
ENDFOR
get new state: $state_{t+1} = state_t + action_t$
simulate Arousal Valence: $ARVA_{t+1} = FIS(state_{t+1})$
// FIS – Fuzzy Inference System
$ACAR_{t+1} = FIS(ARVA_{t+1})$;
$\Delta Reward = ACAR_{t+1}- ACAR_t$
ENDFOR

Fig. 3. The pseudo code for ambient comfort control algorithm.



a) b)

c) d)

Fig. 4. The TD error $\delta_{TD}(t)$ dynamics using corresponding controller parameters: a – γ=0.8, λ=0.5, α=0.2; b – *γ=0.8, λ=0.5, α=0.01;* c – γ=0.9, λ=0.9, α=0.02; d – The dynamics of ambient environment state using corresponding controller parameters: γ=0.8, λ=0.5, α=0.2.

V. CONCLUSIONS

In this paper, the human Affect Reward Based Automation Ambient Comfort Controller (ACARBC) is proposed as the interconnected cloud computing intelligent services that provide intelligent calculus for any instrumented interconnected environment sense and control system.

The ACARBC has been modelled and, as the experimental results show, that an environmental state characteristics that create an optimum ambient comfort can be obtained by ACAR index. The ACAR index is dependent on human physiological parameters: the temperature, the *ECG*- electrocardiogram and the *EDA*-electro-dermal activity.

For future works, it is necessary to investigate the stability of ACARBC as it has been shown that it varies for different controller parameters.

REFERENCES

[1] D. B. Hoang, L. Chen, "Mobile cloud for assistive healthcare (MoCAsH)", in *Proc. of the APSCC*, 2010, pp. 325–332.

[2] H. Zhang, 'Human thermal sensation and comfort in transient and non-uniform thermal environment", Ph.D. dissertation, UC Berkeley, 2003, p. 415.

[3] K. Dalamagkidis, D. Kolokotsab, K. Kalaitzakisc, G. S. Stavrakakis, "Reinforcement learning for energy conservation and comfort in buildings", *Building and environment*, vol. 42, pp. 2686–2698, 2007. [Online]. Available: http://dx.doi.org/10.1016/j.buildenv.2006.07.010

[4] D. Dzemydienė, A. A. Bielskis, A. Andziulis, D . Drungilas, G. Gricius, "Recognition of human emotions in reasoning algorithms of wheelchair type robots", *Informatica*, vol. 21, no. 4, pp. 521–532, 2010.

[5] M. Sedighizadeh, A. Rezazadeh, "Adaptive PID controller based on reinforcement learning for wind turbine control", in *Proc. of the World Academy of Science, Engineering and Technology*, Cairo, Egypt, vol. 27, 2008, pp. 257–262.