Ambient Lighting Controller Based on Reinforcement Learning Components of Multi-Agents

A. A. Bielskis, E. Guseinoviene
Department of Electrical Engineering, University of Klaipėda, Bijūnų str. 17, LT-91225 Klaipėda, Lithuania, e-mails: bielskis@ik.ku.lt, guseinoviene@gmail.com

D. Dzemydiene
Institute of Mathematics and Informatics, Vilnius University, Akademijos str. 4, Vilnius, LT-08663, Lithuania, e-mail: daledz@mruni.eu

D. Drungilas, G. Gricius
Institute of Mathematics and Informatics, Vilnius University, Akademijos str. 4, Vilnius, LT-08663, Lithuania, Department of Electrical Engineering, University of Klaipėda, Bijūnų str. 17, Klaipeda, Lithuania, Department of Informatics Engineering, University of Klaipėda, Bijūnų str. 17-206, Klaipeda, Lithuania, e-mails: dorition@gmail.com, gediminas@ik.ku.lt

Introduction

Inspired by investigations of thermal comfort, indoor air quality and adequate luminance by using the Predicted Mean Vote Index (PMV) [1–3], the human Ambient Lighting Affect Reward (ALAR) index is proposed for automatic quality control of lighting in the ambient assisted living environment [4, 5]. The ALAR based multi-agent ambient lighting controller is planned also to be used to improve energy savings. Specifically, it predicts the indoor RGB LED lighting conditions at a given time by measuring integrated ALAR index that defines ambient lighting affect to the human. Principles of development of the Ambient Lighting Affect Reward Based Multi-Agent Controller, the ALARBMAC are described in this paper. The ALARBMAC is planned to be applied in the process of development of Eco-social laboratory, the ESLab as a laboratory prototype of the Smart Eco-Social Apartment [4].

The ALAR based controller model

The developing process of the smart environment is based on automatic control which adopts environment by smart sensing of human physiological signals. The reinforcement learning is used to get optimal environment characteristics. The architecture of the reinforcement learning based ambient lighting controller (RLBALC) is shown in Fig. 1. Formally, the goal of the reinforcement learning based ambient lighting controller (RLBALC) is to find such environmental state characteristics that create an optimal RGB LED lighting for people affected by this environment.

The RLBALC consists of the following parts: the Environment Evaluation System, the Radial Basis Neural Network, and the Learning Algorithm. The Environment Evaluation System is used to evaluate the human comfort by sensing the affect of the following RGB LED lighting color and intensity parameters: the intensity of red ($L_r$), green ($L_g$) and blue ($L_b$) light.

---

Fig. 1. Structural-functional schema of ambient lighting controller (RLBALC)
The human comfort is expressed as an Ambient Lighting Affect Reward, the ALAR index function

\[ ALAR = f[a(t, c, d), v(t, c, d)], \text{ALAR} \in [-3, 3], \]  

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 Lighting Affect Reward (ALAR) System as shown in Fig. 2.

![Fig. 2. Block diagram of the Arousal-Valence and the Ambient Lighting Affect Reward (ALAR) fuzzy inference system](image)

The Radial Basis Neural Network is the main component of the RLBALC 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 RLBALC), 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 \([6, 7]\). A feature state consists of \( N \) features, each having an activation factor in the interval \([0, 1]\). Linear approximations calculate their function value with

\[ f(x) = \sum_{i=1}^{N} \phi_i(x) w_i, \]  

where \( \phi(x) \) 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 \( \phi_i(x) \) dependent only on the distance between the state, \( s \), and prototypical center state of the feature, \( c_i \), and relative to the feature’s width \( \sigma_i \)

\[ \phi_i(x) = \exp \left( -\frac{||x - c_i||^2}{2\sigma_i^2} \right), \]  

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 TD method simultaneously \([7]\). The TD error \( \delta_{TD}(t) \) is calculated by the temporal difference of the value function between successive states in the state transition as

\[ \delta_{TD}(t) = r(t) + \gamma V(t+1) - V(t), \]  

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 \( \theta \) of the policy function and the value function are updated as

\[ \tilde{\theta}_{t+1} = \tilde{\theta}_t + \alpha \delta_{TD}(t) \]  

where \( \alpha \) is the learning rate and the eligibility trace, \( e \) can be calculated by:

\[ \begin{align*}
\dot{e}_0 &= 0, \\
\dot{e}_{t+1} &= \gamma \dot{e}_{t} + \nabla_{\theta} V(t)
\end{align*} \]  

As an example, by empirically defining fuzzy membership functions and fuzzy rules, we get the fuzzy system surface for recognition of the ALAR index versus arousal and valence (Fig. 3).

![Fig. 3. Fuzzy system surface for recognition of the ALAR index versus arousal and valence](image)
The ambient comfort affect reward based multi-agent controller

Ambient comfort measurement and environment control multi-agent system (MAS) of Fig. 4 monitors the environment and controls the devices in the KUSLab of Fig. 8 by returning the reward of ambient comfort.

Whole smart ambient comfort control model was implemented by using multiple interacting intelligent agents. The multi-agent system (MAS) was constructed by using JACK – Java-based framework for multi-agent system development. The MAS model consists of the three main parts: 1) the environment monitoring and data collecting subsystem of Fig. 5; the Ambient Comfort Affect Reward (ACAR) index recognition subsystem of Fig. 6, and 3) the ambient comfort control subsystem of Fig. 7. The hardware part of the smart e. wellness meter, the SeWM (see Fig. 1), which is the part of the biofeedback monitoring agent, has been realized on ATMega32 microcontroller. The human body temperatures are measured by DS18B20 digital thermometers and communicate with IC over the 1-Wire bus. The microcontroller filters the ECG signal using kernel regression smoothing, the EDA signal - using Gaussian smoothing and sends these samples to the PC via USB/RS232 based interface. When new data is collected, an appropriate event occurs in the MAS to be handled by “AV_FIA” (arousal valence fuzzy inference agent) agent of Fig. 6.

The agent “AV_FIA” uses two plans, the “GetEDAParams” and the “GetECGParams” to analyze new samples and get standardized parameters (latency, rise time, amplitude, half recovery time for EDA and heart rate, heart rate variability for ECG) and then stores new data to database. The agent “AV_FIA” uses plan “calculate_AV” to calculate affect features: arousal and valence from physiological data. Then the agent ACAR_FIA, by communicating with the fuzzy inference system, gets the ACAR index parameter and generates event “ACAR” which will be handled by the “AmbiantComfort” agent (Fig. 7).

The whole MAS are shown in the Fig. 7. Agent “Environment_monitor” monitors environment, analyses new and collected environment data and if environment changes then event “New_state” occurs which is handled by two agents. Agent “AmbiantComfort” takes ACAR and new environment state and if needed, using plan
“ModifyEnvironment” generates event to the agent “EnvironmentController”.

Vision of embedding of the multi-agent-based comfort control subsystem into the ESLab

Based on the idea of the development of the Smart Eco-Social Apartment [4], the block diagram of the Eco-social laboratory, the ESLab is proposed in Fig. 8. Fig. 8 depicts a block diagram of the ESLab with integrated wind generator, PV panels, solar collector, air-to-water-air-to-air heat pump, heating-ventilation-automatic controller HVAC as well as a smart energy meter, the SEM and the Ambient Comfort Affect Reward Meter, the ACARM.

The ESLab can be dislocated on the last floor of the building of any high education institution, and it may occupy 3 - 4 rooms for arranging of sustainable laboratory and office as well as some place on the roof to deploy an outdoor unit of air-to-water-air-to-air heat pumps, solar panel units, PV panel units, and small wind generator units. The renewable energy, the Ewg from wind generator as well as the Epv from photovoltaic panels is planned to be monitored and economically distributed by the smart energy meter, the SEM. The SEM has to adaptively control all available at that moment Ewg and Epv energy flow for feeding the heat pump and electrical heater of hydro unit. By using signals from the heating-ventilation-automation controller, the HVAC, the SEM should manage the climate control by adding energy from conventional sources such as central heating system of the building if there is not enough heating power from alternating energy sources at that moment. The hardware implementation of the smart ambient comfort control system in the ESLab is shown in Fig. 9.

Fig. 6. The ACAR index recognition multi-agent subsystem

Fig. 7. Multi-agent-based comfort control subsystem

The Fig. 9 represents two embedded systems: a) the environment monitoring and control agent-based system and b) the biofeedback monitoring agent-based system. The environment monitoring and control agent consists of:
1) The central module of Fig. 9a-1 which has the Atmega32 microcontroller connected to the computer via USB/RS232 interface to controlling devices and sending data to the multi-agent system implemented in the connected computer;

2) The environment monitoring and control module of Fig. 9a-4 which has the several sensors such as temperature, light level, carbon dioxide and has ability to control connected devices such as fan, lamps, and air-conditioner;

3) The information module (Fig. 9a-3) containing the LCD which displays an information about the system status as well as the keypad for manual system control.

The biofeedback monitoring agent of Fig. 9b has the same central and information modules similar to the environment monitoring and control agent, only the software of the microcontroller is different, and it has one physiological signal sampling module of Fig. 9b-5. This agent takes ECG, EDA and temperature samples.

The ACARM, described in this paper is developed by authors of this paper and tested in electronic laboratory by students during their laboratory works in the department. The measured data helped to predict some aspects of wellness of the students during their laboratory work and exams used to improve the software of the ACARM.

Conclusions

1) A vision is introduced of sustainable eco-social laboratory, the ESLab to be used to speed up the process of development of the Smart Eco-Social Apartment recently proposed by authors of this paper [4]. The multi-agent model of the ambient comfort measurement and environment control system is proposed and recommended to be used for further development of the ESLab in this paper.

2) The human Ambient Lighting Affect Reward index, the ALAR index is proposed at the first time used for development of the Reinforcement Learning Based Ambient Comfort Controller, the RLBACC for the ESLab.

3) The proposed ALAR index is described as a function that depends on the human physiological parameters: the temperature, the ECG- electrocardiogram and the EDA-electro-dermal activity. The fuzzy logic is used to approximate the ALAR index function by defining two fuzzy inference systems: the Arousal-Valence System, and the Ambient Lighting Affect Reward (ALAR) System.

4) The goal of the RLBACC controller is to find such environmental state characteristics that create an optimal comfort for people affected by this environment. The Radial Basis Neural Network is used as the main component of the RLBACC controller to performing of two roles of the Actor and the Critic.

5) The policy structure, known as the Actor, is used to select actions, and the estimated value function, known as the Critic, is applied to criticize the actions made by the Actor. The Critic was used as a value function approximation of the continuous learning tasks of the RLBACC controller.
Acknowledgments

The authors of this paper express their sincere thanks to the authorities of the University of Klaipėda for their support making the experimental work necessary to creating of this paper as well as the Project “LED-Increasing Energy Saving through Conversion to LED lighting in public space” within the South Baltic Cross-Border Co-operation Programme 2007-2013, Subsidy contract No:WTTPB.02.02.00-94-001/09-00) for the possibility to complete a scientific research.

References


Received 2011 12 13
Accepted after revision 2012 01 14


The paper presents a vision of sustainable eco-social laboratory, the ESLab which might be used to speed up the process of development of the recently proposed by authors of the Smart Eco-Social Apartment. It is presented the multi-agent model of the ambient comfort measurement and environment control system to be used for the development of the ESLab. The human Ambient Lighting Affect Reward index, the ALAR index is proposed at the first time used for development of the Reinforcement Learning Based Ambient Comfort Controller, the RLBACC for the ESLab. The ALAR 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 ALAR index function by defining two fuzzy inference systems: the Arousal-Valence System, and the Ambient Lighting Affect Reward (ALAR) System. The goal of the RLBACC is to find such the environmental state characteristics that create an optimal comfort for people affected by this environment. The Radial Basis Neural Network is used as the main component of the RLBACC 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 RLBACC. III. 9, bibl. 7 (in English; abstracts in English and Lithuanian).


Pristatoma universiteto tipo darniosios laboratorijos ESLab vizija, kuri plėtoja neseniai autorių pasiūlyto išsamiąjį ekosocialaus būsto įgyvendinimo idėją. Pateikiamas aplinkos komforto matavimo ir aplinkos kontrolės sistemos valdiklio modelis, kuris bus panaudotas ESLab plėtotei. Straipsnyje pasiūlytas žmogaus aplinkos apšvietimo efekto paskatos AAAP (ALAR) indeksas pritaikytas kuriant paskatos mokytojų pagrįstajai aplinkos komforto valdiklį ESLab laboratorijai. AAAP (ALAR) indeksas priklauso nuo žmogaus fiziologinių parame: temperatūros, ECG (elektrokardiogramos) ir EDA (elektrinio odos aktyvumo). Neraikioji logika yra panaudota AAAP (ALAR) indekso funkcijai aproksiminui, taikant dvi neraikiškas išvedimo sistemas: susijaudinimo ir malonumo mokymosi sistemų. Kritikas tai taikoma kritikos funkcijai, kuri yra žinoma kaip kritikos, kuris kritikuoja aktoriaus padarytus veiksmus. Kritikos vietoje buvo įtrauktas pasiūlytas PMGAAP (RLBACC) mokymosi užduočių jverčio funkcijos aproksimavimo ilg. 9, bibl. 7 (anglų kalba; santraukos anglų ir lietuvių k.).