A Neurocomputational Model of Nicotine Addiction Based on Reinforcement Learning

Selin Metin[1] and N. Serap Şengör[1]

Istanbul Technical University, Electrical and Electronics Engineering Faculty, Maslak
34469, Istanbul, Turkey
selinmetin@gmail.com, sengorn@itu.edu.tr

Abstract. Continuous exposure to nicotine causes behavioral choice to be modified by dopamine to become rigid, resulting in addiction. In this work, a computational model for nicotine addiction is proposed and the proposed model captures the effect of continuous nicotine exposure in becoming addict through reinforcement learning. The computational model is composed of three subsystems each corresponding to neural substrates taking part in nicotine addiction and these subsystems are realized by nonlinear dynamical systems. Even though the model is sufficient in acquiring addiction, it needs to be further developed to give a better explanation for the process responsible in turning a random choice into a compulsive behavior.

Keywords: computational model, dynamic system, nicotine addiction, reinforcement learning

1 Introduction

The value of an experience or an action is imposed by the reward gained afterwards. An action inducing a greater reward is sensed as a better action, and thus rewarding it is repeated frequently [1]. In the case of addiction, the abusive substance (nicotine, drugs, etc.) has a greater value in the brain than other forms of reward imposing actions. It is believed that some persistent modifications in the synaptic plasticity is the cause of addiction, thus we can define addiction as a disorder in the mesolimbic system which modifies responses of rewarding actions. Mislead by overemphasized reward sensations addicts compulsively seek the object of their addiction. As the reward mechanism has persistently changed, addicts are usually not completely cured and relapse into drug use after treatment [2].

The two main approaches in explaining addiction are the opponent process theory and reward related learning [3–5]. Using reinforcement learning theory, addiction is explained as the cumulative result obtained by the administration of a drug as a positive reinforcer [5–7]. The opponent-process theory of motivation [3] is used to explain the conditioning principles leading to pleasurable and compulsive activity. According to this model, emotions are paired and when one emotion in a pair is experienced, the other is suppressed. In [8], these two approaches are considered together in deriving a computational model for nicotine addiction.

The hypothesis we considered in this work claims that nicotine addiction is a transition from impulsive behavior to compulsive behavior developed through reinforcement learning. While developing the model considering this view, neural substrates taking part in cognitive processes related to addiction as action selection and value evaluation is considered like in [9]. The proposed model is simulated with an m-file created in MATLAB.

2 The Proposed Model for Nicotine Addiction

Nicotine addiction, as with all other kinds of abusive substance addictions, develops with the malfunctioning of the reward mechanism. Nicotine effects the VTA DA signaling, which in turn modify the glutamatergic processes responsible in learning. The behavioral choices depend on the learned situations, in nicotine addiction this choice is in favor of obtaining more nicotine. Continuous exposure to nicotine causes behavioral choice modified by DA to become rigid, resulting in addiction. The proposed model captures this property through reinforcement learning which adapts a parameter that denotes the effect of VTA DA signaling on action selection.

2.1 Implementation of the Model

The model has two parts: a DA signaling module which is triggered by nicotine presence and an action-selection module (Figure 1). A-S module is a well-studied cortex-basal ganglia-thalamus dynamical system [10–12, 9]. The DA signaling module is composed of an action evaluation part, the operation of which is based on the presence of nicotine and a value assignment part which calculates the rewards assigned to the actions and expectation error. The DA signaling module drives the A-S loop with the representation of hedonic value of the previous actions.

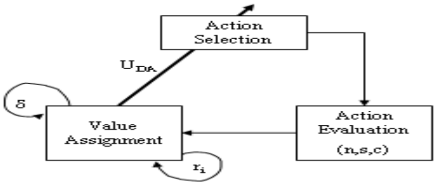

Fig. 1. The main blocks of the nicotine addiction model.

As in [8] the effect of DA is demonstrated by a DA module which is represented by a difference equation in order to model the dynamic behavior of the process (1):

$$u_{DA}(k+1) = u_{DA}(k) + \mu_{DA}(-u_{DA}(k) + s_{DA}(ri, Ni)) \tag{1}$$

The activation function s_{DA} is a sigmoidal function given as (2):

$$s_{DA}(ri, Ni) = 0.5(1 + \tanh(Ni * ri - \theta_{DA})) \tag{2}$$

Ni is nicotine uptaking represented by the product of the values of n and s signals when nicotine injection stops (Appendix 1). θ_{DA} is the threshold setting the minimum tonic DA. We took θ_{DA}=0.01. ri is the reward signal initiated by nicotine taking. μ_{DA} is the learning rate in the DA subsystem.

The action-selection module used here is acquired from [10, 9] which is expressed with the following equations where premotor (pm) and motor (m) loops model the cortex-basal ganglia-thalamus (C-BG-TH) loops (4, 5, 5).

$$\begin{aligned}
p_{pm}(k+1) &= f(\lambda p_{pm}(k) + m_{pm}(k) + W_{c_{pm}} I(k)) \\
m_{pm}(k+1) &= f(p_{pm}(k) - d_{pm}(k)) \\
r_{pm}(k+1) &= W_{r_{pm}} f(p_{pm}(k)) \\
n_{pm}(k+1) &= f(p_{pm}(k)) \\
d_{pm}(k+1) &= f(W_{d_{pm}} n_{pm}(k) - r_{pm}(k))
\end{aligned} \tag{3}$$

$$\begin{aligned}
p_{m}(k+1) &= f(\lambda p_{m}(k) + m_{m}(k) + \beta p_{pm} + \text{noise}) \\
m_{m}(k+1) &= f(p_{m}(k) - d_{m}(k)) \\
r_{m}(k+1) &= W_{r_{m}} f(p_{m}(k)) \\
n_{m}(k+1) &= f(p_{m}(k)) \\
d_{m}(k+1) &= f(W_{d_{m}} n_{m}(k) - r_{m}(k))
\end{aligned} \tag{4}$$

$$f = 0.5(1 + \tanh(a(x - 0.45))) \tag{5}$$

$W_{d_{pm/m}}$ adds the diffusive effect of subthalamic nucleus and is a symmetrical matrix. The diagonal matrix $W_{r_{pm}}$ represents the effect of ventral striatum (nucleus accumbens) on dorsal striatum (caudate nucleus and putamen). The representation of stimulus is formed by the matrix $W_{c_{pm}}$. The adaptation of weights $W_{c_{pm}}$ and $W_{r_{pm}}$ is done as below (6):

$$W_{c_{pm}}(k+1) = W_{c_{pm}}(k) + \eta_c \delta(k) p_m(k) I(k)^{'} \tag{6}$$

$$\begin{aligned}
W_{r_{pm}}(k+1) = W_{r_{pm}}(k) + &\\
&\eta_r((\overline{U_{DA}} + Ni)(U_{DA} - \theta_{w_{DA}})^{'}(p_m(k) - \theta))^{'} f(p_m(k)) r_m(k)
\end{aligned} \tag{7}$$

The factors are the phasic DA activitiy U_{DA}, running average of 10 steps denoted as in [8] by $\overline{U_{DA}}$. $W_{r_{pm}}$ is calculated only after the reward signal ri becomes greater than 0.5. Thresholds for U_{DA} and p_m, respectively, are θ_{DA} and θ, and are taken as 0.1 times their respective signal. The learning rate η is taken as 0.1. The variable δ represents the error in expectation and is calculated as (8):

$$\delta(k) = ri(k) + \mu V(k+1) - V(k) \tag{8}$$

The evaluation of the action selection based on the cortex input and the corresponding reward is given as the value signal (9):

$$V(k) = (W_v + \mathrm{base})I(k) \tag{9}$$

Here, W_v is a row vector and the term base is a row vector with identical entries. $I(k)$ corresponds to input, which in this case is the action performed as a result of (4, 5, 5), and corresponds to "smoke" or "not smoke". An expectation signal based on the value signal is generated which, together with ri, gives rise to the error δ. The error signal represents the modulating role of the neurotransmitters and modulates the behavior of dorsal striatum stream via $W_{r_{pm}}$. The error signal strengthens the representation of the input via $W_{c_{pm}}$ and updates the value of stimuli via W_v are as below (10):

$$W_v(k+1) = W_v(k) + \eta_v \delta(k) I(k)^{'} \tag{10}$$

2.2 The Simulation Results

To measure the performance of the proposed model, the response to nicotine uptaking is considered. At the beginning reward value is very small (like 0.01). Each time the selected action is smoking, ri is multiplied by 2 until ri=1.

The action selected by the A-S module is determined by calculating the solution of p_m. The value function and the error function are calculated, and using these calculations the weight matrices $W_{c_{pm}}$, $W_{r_{pm}}$, and W_v are updated. The simulation stops if the smoking action is selected successively for 20 times in a given time frame. After numerous trials, it is observed that 20 successiv smoking decisions are enough for the system to be considered as a model of an addict. Otherwise, it is decided that addiction is not established.

The parameter values used in the simulation are λ=0.5, β=0.03, a=3, μ_{DA}=0.1, η_c=0.1, η_v=0.1, η_r=0.1 and base is 0.2. The initial values of the weight matrices W_c and W_v are generated randomly with small positive real numbers. The initial value of the diagonal matrix $W_{r_{pm}}$ is ones. During the updating phase the matrix values $W_{c_{pm}}$ and $W_{r_{pm}}$ are normalized. The matrices $W_{d_{pm/m}}$ and W_{rm} are composed of 0.5's and they are constant. The noise signal is generated as a very small random number. The action outputs are coded as [1 0]' for smoking, [0 1]' for nonsmoking, and [1 1]' for indecisive behaviors.

In 20 of the 50 successive runs the model completed the task of becoming an addict. The average number of trials to become addict is 346 out of 1000 trials, with a standard deviation of 265.7671. The final matrices for a successful trial are given as follows (11):

$$W_{r_{pm}} = \begin{bmatrix} 1 \\ 0.6179 \end{bmatrix}, \quad W_{c_{pm}} = \begin{bmatrix} 0.8569 & 0.2965 \\ 0.16 & -0.4331 \end{bmatrix} \tag{11}$$

The expectation error signals for two different cases are given in Figure 2. δ remains constant if the same choice is made successively, and changes otherwise.

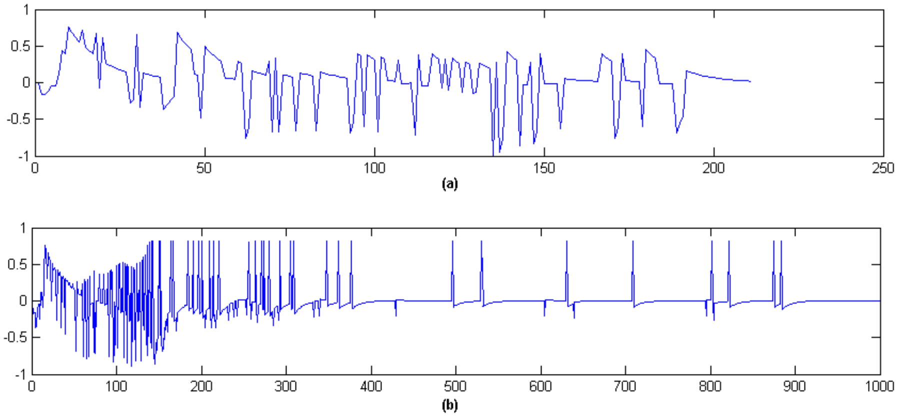

Fig. 2. Reinforcement error signal δ when addiction is a) set up b) not set up

3 Discussion and Conclusion

Our work proposes a cortico-striato-thalamic A-S circuit driven by the effects of nicotine uptaking as a model for nicotine addiction. The A-S circuit has two components: an action selection component corresponds to the dorsal stream which simulates behavioral choices, and the other component corresponds to the ventral stream which simulate the evaluation of the action choices and modulates the action selection. The A-S circuit utilizes a competitive learning which is modified with the VTA DA signaling affected by the nicotine. While the structure of the A-S circuit is interconnected nonlinear dynamical systems corresponding to premotor and motor loops, the modification is realized changing a parameter in premotor loop. While in [8], the A-S module is a winner-take-all system, in this work a dynamical system triggered by previous actions, and their evaluation is utilized for A-S. Furthermore, the n-s-c circuit used here is novel. Thus, the system proposed in [10, 9] for A-S is enhanced in this work for a more complicated process wheree reward has more importance on the overall process.

The aim in this work is to support the idea that addiction develops as a form of goal-directed behavior, and therefore the interaction of cortico-striato-thalamic action selection loops have an important role in the development of addiction.

References

1. Hyman, S.E., Malenka, R.C., Nestler, E.J.: Neural Mechanisms of Addiction: The Role of Reward-Related Learning and Memory. Annual Review of Neuroscience. **29** 565–598 (2006)
2. Spanagel, R., Heilig, M.: Addiction and Its Brain Science. Addiction **100** no.12, 1813–1822 (2005)

3. Solomon, R.L., Corbit, J.D.: An Opponent-Process Theory of Motivation The American Economic Review, Vol. 68, No. 6., 12–24 (1978)
4. Koob, G.F., Le Moal, M.: Neurobiological mechanisms for opponent motivational processes in addiction Phil. Trans. R. Soc. B **363** 3113–3123 (2008)
5. Dayan, P.: Dopamine, Reinforcement Learning, and Addiction Pharmacopsychiatry **42** (2009) 56–65 Addiction **100** no.12, 1813–1822 (2005)
6. Peele, S., Alexander, B.K.: The Meaning of Addiction Chapter 3: Theories of Addiction, http://www.peele.net/lib/moa3.html
7. Delgado, M.R.: Reward-Related Resposes in the Human Striatum. Annals of the New York Academy of Sciences, **1104** 70–88 (2007)
8. Gutkin, B.S., Dehaene, S., Changeux, J.P.: A Neurocomputational Hypothesis for Nicotine Addiction. PNAS **103** no.4, 1106–1111, (2006)
9. Metin, S., Sengor, N.S.: Dynamical System Approach in Modeling Addiction Accepted to be published in Proceedings of BICS, (2010)
10. Sengor, N.S., Karabacak, O., Steinmetz, U.: A Computational Model of Cortico-Striato-Thalamic Circuits in Goal-Directed Behavior. LNCS 5163, Proceedings of ICANN, 328-337 (2008)
11. Taylor, J.G., Taylor, N.R.: Analysis of Recurrent Cortico-Basal Ganglia-Thalamic Loops for Working Memory. Biological Cybernetics, 82, 415–432 (2000)
12. Gurney, K., Prescott, T.J., Redgrave, P.: A Computational Model of Action Selection in the Basal Ganglia I: A New Functional Anatomy. Biological Cybernetics, 84, 401–410 (2001)

Appendix 1

Initial values of n, s, and c are all 0.1. Parameters used in equations are as follows: θ_n=0.6, θ_s=0.7, θ_c=0.7, β_s=0.4, β_c=0.4. *nicotine* level is taken as 0.3 if k is less than 500, 0 otherwise. Rates used in the equations are μ=0.1, τ_n=0.25, τ_s=1, τ_c=2.

Activation functions (13):

$$\begin{aligned} \alpha_n(k) &= 0.5(1+\tanh(nicotine-\theta_n)) \\ \alpha_s(k) &= 0.5(1+\tanh(n(k)-\theta_s)) \\ \alpha_c(k) &= 0.5(1+\tanh(s(k)-\theta_c)) \\ \beta_n(k) &= 0.5(1+\tanh(c(k)-\theta_n)) \end{aligned} \tag{12}$$

Dynamic equations of n, s, and c (14):

$$\begin{aligned} n(k+1) &= n(k)+\mu*(1/\tau_n)[-\beta_n(k)c(k)+\alpha_n(k)(1-n(k)c(k))] \\ s(k+1) &= s(k)+\mu*(1/\tau_s)[-\beta_s s(k)+\alpha_s(k)(1-s(k))] \\ c(k+1) &= c(k)+\mu*(1/\tau_c)[-\beta_c c(k)+\alpha_c(k)(1-c(k))] \end{aligned} \tag{13}$$