Design of Sigmoid Activation Functions for Fuzzy Cognitive Maps via Lyapunov Stability Analysis

In Keun LEE†, Nonmember and Soon Hak KWON†(a), Member

SUMMARY Fuzzy cognitive maps (FCMs) are used to support decision-making, and the decision processes are performed by inference of FCMs. The inference greatly depends on activation functions such as sigmoid function, hyperbolic tangent function, step function, and threshold linear function. However, the sigmoid functions shown in Eq. (1)) in each decisional process must be selected. They have also emphasized that decision-making processes have been designed by experts. Therefore, we propose a method for designing sigmoid functions through Lyapunov stability analysis. We show the usefulness of the proposed method through the experimental results in inference of FCMs using the designed sigmoid functions.

1. Introduction

Fuzzy cognitive maps (FCMs) [1]–[7] are used to support decision-making, which make decisions about future implications, especially when what-if analysis is needed [2]. The decision-making operations using FCMs are performed by inference of FCMs; therefore, the results of decision-making greatly depend on the type and the inference method of FCMs. For the rational design and inference of FCMs, various FCMs (RFCM [3], RBFCM [4], and so on) have been proposed [5]. These FCMs generally use following operations for inference: max–min operation [1] and numerical matrix operation [2], [6]–[8]. In particular, during inference of FCMs using numerical matrix operation, the value of nodes in FCMs could be out of the range of [0, 1]; therefore, activation functions are used to keep the value of the nodes within the range. Various functions are used as activation functions in FCMs such as sigmoid function, hyperbolic tangent function, step function, and threshold linear function [2], [6], [7]. In particular, Bueno and Salmeron [2] have shown by experimental method that sigmoid function is effective for the activation function in a complex decisional environment. They have also emphasized that decision makers must select \( \lambda \) (a parameter determining the steepness of sigmoid functions shown in Eq. (1)) in each decisional process. However, \( \lambda \) is still determined through the subjective knowledge and opinion of experts, and a few studies have selected \( \lambda \) by experimental method. Therefore, we propose a method for designing sigmoid function not by experimental method but through Lyapunov stability analysis based on mathematical models of FCMs in [8]. In designing sigmoid functions, defined as

\[
 f(x) = a_1 + a_2 \left( 1 + e^{-\lambda(x-a_3)} \right),
\]

we only focused on \( \lambda \) here because the sigmoid functions where \( a_1 = 0, a_2 = 1, \) and \( a_3 = 0 \) are generally used for inference of FCMs using numerical matrix operation. We show the usefulness of the proposed method through the experimental results in inference of FCMs using the designed sigmoid functions.

2. Model Description and Preliminaries

For convenience, we will use the following notations throughout this letter.

Notations: \( \mathbb{N}, \mathbb{R}, \mathbb{R}^n, \) and \( \mathbb{R}^{n \times m} \) denote, respectively, the set of natural numbers, the real number space, the real \( n \)-space, and the set of real \( n \times m \) matrices. The superscript ‘\( T \)’ denotes vector and matrix transposition (i.e., if \( \mathbf{u} \in \mathbb{R}^n \) then \( \mathbf{u}^T = [u_1, \ldots, u_n] \), and if \( \mathbf{A} = [a_{ij}]_{n \times m} \in \mathbb{R}^{n \times m} \) then \( \mathbf{A}^T = [a_{ij}]_{m \times n} \), where \( 1 \leq i \leq n, 1 \leq j \leq m, \) and \( n, m \in \mathbb{N} \)). For all \( \mathbf{u} \in \mathbb{R}^n \), let \( ||\mathbf{u}\|| \) denote the Euclidean vector norm (i.e., \( ||\mathbf{u}\|| = (\mathbf{u}^T \cdot \mathbf{u})^{1/2} \)). For all \( \mathbf{A} \in \mathbb{R}^{n \times m} \), let \( ||\mathbf{A}\|| \) denote the spectral norm (i.e., \( ||\mathbf{A}\|| = (\text{maximum eigenvalue of } \mathbf{A}^T \cdot \mathbf{A})^{1/2} \)). If \( f : \mathbb{R} \rightarrow \mathbb{R} \), then \( f'(\cdot) \) is the first derivative of \( f(\cdot) \). If \( \mathbf{u} = [u_1, \ldots, u_n]_{1 \times n} \) is a state vector of a system, then \( \mathbf{u}^* = [u_1^*, \ldots, u_n^*]_{1 \times n} \) denotes the equilibrium state vector of the system.

A FCM is a directed graph showing the relations between essential components in a complex system, and the output of the system is determined by the inference of the FCM. Therefore, we represent a FCM and its inference process by the following definitions, and will use the notation in the definitions throughout this letter.

Definition 1 (Components of FCM): (Refer [8]) Suppose \( C_i \) and \( C_j \) are concepts in a FCM, and \( v_i \) and \( v_j \) are the values of \( C_i \) and \( C_j \) belonging to \( [0, 1] \), respectively, when \( i, j \in \mathbb{N} = \{1, 2, \ldots, n\} \) and \( n \in \mathbb{N} \) is the number of concepts. Then, the weight \( w_{ij} \) is defined as a real number in \( [-1, 1] \). We call the weight positive, negative and no causality from \( C_i \) to \( C_j \) if \( w_{ij} > 0 \), \( w_{ij} < 0 \), and \( w_{ij} = 0 \), respectively.

Definition 2 (Inference of FCM): (Refer [8]) For every \( i \in \mathbb{N} \) and any \( j \in \mathbb{N} \), let \( C_j \) be the causal concepts which influence a concept \( C_i \). Then, for every \( j \in \mathbb{N} \) and all iteration step \( k \geq 0 \) during inference of the FCM,

\[
y_j^{(k+1)} = f \left( r_{11}y_j^{(k)} + r_2 \sum_{l=1}^{n} (w_{lj}y_l^{(k)}) \right)
\]

\[2\]
where $\rho_1, \rho_2 \in (0, 1]$, and $v^{(k)} \in [0, 1]$ represents the value of $C_j$ at the $k$-th iteration step. Moreover, $f : \mathbb{R} \to \mathbb{R}$ is an activation function to restrict $v^{(k+1)}$ into the interval $[0, 1]$. Eq. (2) is also represented as a vector form:

$$v^{(k+1)} = f \left( \rho_1 v^{(k)} + \rho_2 w^T \cdot v^{(k)} \right)$$

(3)

where $v^{(k)} = \left[ v_1^{(k)} \; v_2^{(k)} \; \cdots \; v_n^{(k)} \right]^T$ and $w = [w_{ij}]_{n \times n}$, where $1 \leq i, j \leq n$, which are called state vector and weight matrix, respectively. Moreover, $f(v^{(k)}) = \left[ f(v_1^{(k)}) \; \cdots \; f(v_n^{(k)}) \right]^T$.

Various functions such as the sigmoid function, sinusoidal function, algebraic function, step function, and ramp function are used as the activation functions, and they are classified as unipolar and bipolar functions, or linear and nonlinear functions, or continuous and discontinuous functions [2], [7], [9], [10]. We will consider the unipolar, nonlinear, and continuous functions as the activation functions of FCMs. Therefore, we assume that the activation functions satisfy the following conditions:

**Assumption 1:** The function $f : \mathbb{R} \to \mathbb{R}$ satisfies the Lipschitz condition with a Lipschitz constant $L > 0$; i.e., $|f(u_1) - f(u_2)| \leq L|u_1 - u_2|$ for all $u_1, u_2 \in \mathbb{R}$.

**Assumption 2:** The function $f : \mathbb{R} \to \mathbb{R}$ is bounded; i.e., $0 \leq f(u) \leq M$ for every $u \in \mathbb{R}$ and any $M \in \mathbb{R}$ such that $M > 0$.

We know from Assumptions 1 and 2, and Definition 3, we can derive the following lemmas, which are essential for the stability analysis of Eq. (4).

**Lemma 1:** (Refer [9]) If $g(\cdot)$ and $y_j^{(k)}$ are the same as in Definition 3, and $L$ is a Lipschitz constant as shown in Assumption 1, then for every $j \in N$ and all $k \geq 0$, $|g(y_j^{(k)})| \leq L |y_j^{(k)}|$.

**Proof.** From Assumption 1, $|f(y_j^{(k)}) - f(y_j^{'(k)})| \leq L |y_j^{(k)} - y_j^{'(k)}|$. Here, $f(y_j^{(k)}) - f(y_j^{'(k)}) = g(y_j^{(k)}) - g(y_j^{'(k)}) = y_j^{(k)} - y_j^{'(k)}$. Therefore, we have $|g(y_j^{(k)})| \leq L |y_j^{(k)}|$.  

**Lemma 2:** If $x^*$ is an equilibrium value as shown in Definition 3 and $L$ is a Lipschitz constant as shown in Assumption 1, then for every $j \in N$, $|f(x^*_j)| \leq L$.

**Proof.** From Lemma 1, we can derive the following inequalities: For all $k \geq 0$ and every $j \in N$,

$$|f(y_j^{(k)} + x^*_j) - f(y_j^{(k)})| \leq L,$$

$$\lim_{j \to 0} \frac{|f(y_j^{(k)} + x^*_j) - f(y_j^{(k)})|}{|y_j^{(k)}|} = |f'(x^*_j)| \leq L.$$ 

Therefore, the derivative of $f(x)$ evaluated at $x = x_j^*$ is bounded by the Lipschitz constant $L$.  

We are now ready to show that Eq. (4) is globally exponentially stable. Consider the following definition:

**Definition 4 (Global exponential stability):** (Refer [10]) The vector $x^*$ equilibrium of Eq. (4) is said to be globally exponentially stable if the following condition is satisfied:

$$\|x^{(k)} - x^*\| \leq \sigma \epsilon^{r(k)} \|x^{(0)} - x^*\|$$

(6)

for all $k \geq 0$, where $0 < \epsilon < 1$, $r > 0$, and $\sigma \geq 1$.

3. Design of Sigmoid Activation Functions

We use a Lyapunov functional to analyze the stability of the equilibrium of Eq. (4).

**Theorem 1:** If $L$ is the Lipschitz constant, shown in Lemma 1, such that
0 < L < \frac{1}{\rho_1 + \rho_2 \|w\|} \tag{7}

where \(w\) is the weight matrix and \(\rho_1, \rho_2 \in (0, 1]\) as in Definition 1, then Eq. (4) is globally exponentially stable.

**Proof.** If there exists a positive definite matrix \(P = \text{diag} [p_{11}p_{22} \cdots p_{nn}]\), then the Lyapunov functional is given by \(V(\hat{y}^{(k)}) = (\hat{y}^{(k)})^T \cdot P \cdot y^{(k)}\) for all \(k \geq 0\). Let \(p_{\text{min}} = \min_{k \in \mathbb{N}} (p_k)\) and \(p_{\text{max}} = \max_{k \in \mathbb{N}} (p_k)\) (i.e., \(p_{\text{max}} \geq p_{\text{min}} > 0\)); then, we can derive the following inequality by Definition 3 and Lemma 1:

\[
\Delta V(\hat{y}^{(k)}) = V(\hat{y}^{(k+1)}) - V(\hat{y}^{(k)}) = (\hat{y}^{(k+1)})^T \cdot P \cdot y^{(k)} - (\hat{y}^{(k)})^T \cdot P \cdot y^{(k)} \\
\leq p_{\text{max}} \sum_{i=1}^n (\hat{y}_i^{(k+1)})^2 - p_{\text{min}} \sum_{i=1}^n (\hat{y}_i^{(k)})^2 \\
= p_{\text{max}} \|\hat{y}^{(k+1)}\|^2 - p_{\text{min}} \|\hat{y}^{(k)}\|^2 \\
= p_{\text{max}} [p_1 \mathfrak{g}(\hat{y}^{(k)}) + \rho_2 \mathfrak{S}^{T} \mathfrak{g}(y^{(k)})] - p_{\text{min}} \|y^{(k)}\|^2 \\
\leq p_{\text{max}} (p_1 + \rho_2 \|w\|)^2 \|\mathfrak{g}(y^{(k)})\|^2 - p_{\text{min}} \|y^{(k)}\|^2 \\
\leq p_{\text{max}} (p_1 + \rho_2 \|w\|)^2 L^2 \|y^{(k)}\|^2 - p_{\text{min}} \|y^{(k)}\|^2 \\
= (p_{\text{max}} (p_1 + \rho_2 \|w\|)^2 \|y^{(k)}\|^2 - p_{\text{min}}) \|y^{(k)}\|^2 \\
\leq \alpha \|y^{(k)}\|^2. 
\]

If \(y^{(k)}\) is a null vector (i.e., \(\|y^{(k)}\| = 0\)), then \(\Delta V(y^{(k)}) = 0\). Therefore, from stability in the sense of Lyapunov, if \(p_{\text{max}} (p_1 + \rho_2 \|w\|)^2 L^2 - p_{\text{min}} < 0\), then Eq. (4) is globally asymptotically stable. Next, we prove the global exponential stability of Eq. (4). Let \(\alpha = p_{\text{min}} / p_{\text{max}} (p_1 + \rho_2 \|w\|)^2 L^2 > 0\).

Then, \(\Delta V(y^{(k)}) \leq -\alpha \|y^{(k)}\|^2 \leq -\alpha V(y^{(k)})\). Therefore, for all \(k \geq 0\), \(V(y^{(0)}) \leq V(y^{(0)}) (1 - \alpha)^k\) (see Lemma 3), and \(p_{\text{min}} \|y^{(0)}\|^2 \leq V(y^{(0)}) (1 - \alpha)^k \leq p_{\text{max}} (1 - \alpha)^k \|y^{(0)}\|^2\).

Thus, \(\|y^{(k)}\|^2 \leq (p_{\text{max}} / p_{\text{min}})^{1/2} (1 - \alpha)^{k/2} \|y^{(0)}\|^2\).

Therefore, we have \(\|x^{(k)} - x^*\| \leq (p_{\text{max}} / p_{\text{min}})^{1/2} (1 - \alpha)^{k/2} \|x^{(0)} - x^*\|\). If \(0 < \alpha < 1\), then condition (6) is satisfied.

Moreover, \(\alpha\) can be represented with respect to \(L\) as

\[
\left(\frac{p_{\text{min}} - p_{\text{max}}}{p_{\text{max}}}\right)^{1/2} \frac{1}{\rho_1 + \rho_2 \|w\|} L < \left(\frac{p_{\text{min}}}{p_{\text{max}}}\right)^{1/2} \frac{1}{\rho_1 + \rho_2 \|w\|}.
\]

Here, the Lipschitz constant \(L\) is a positive real number (i.e., \(L \in \mathbb{R} \) where \(L > 0\)); therefore, \(p_{\text{max}} = p_{\text{min}}\).

Thus, we finally have the following condition: \(0 < L < 1 / (\rho_1 + \rho_2 \|w\|)\). This implies that the equilibrium \(x^*\) of Eq. (4) is globally exponentially stable.

**Lemma 3:** Let \(y^{(k)}\), \(\alpha\), and \(V(\cdot)\) be the same as in Theorem 1. Then the statement

\[
V(y^{(k)}) \leq V(y^{(0)}) (1 - \alpha)^k
\]

holds for all \(k \geq 0\).

**Proof.** It is simply proven by mathematical induction. □

**Example 1:** Suppose a sigmoid function where \(a_1 = 0\), \(a_2 = 1\), and \(a_3 = 0\) in Eq. (1), which is bounded by \([0, 1]\). Then the maximum of the derivative of the sigmoid function occurs when \(x = 0\). Therefore, the range of \(\lambda\) is calculated by Lemma 2 and condition (7) as follows:

\[
f'(x) = \frac{\lambda}{e^{\lambda x} + 2 + e^{-\lambda x}},
\]

\[
\left|f'(0)\right| = \frac{\lambda}{e^{\lambda} + 2 + e^{-\lambda}} = \frac{\lambda}{4} \leq L < \frac{1}{\rho_1 + \rho_2 \|w\|}.
\]

where \(w\) is a weight matrix as in Definition 1. Therefore, \(\lambda\) is limited as

\[
0 < \lambda < \frac{4}{\rho_1 + \rho_2 \|w\|}.
\]

If \(w\) and an initial state vector \(v^{(0)}\) are given by

\[
w = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}, \quad v^{(0)} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix},
\]

then the range of \(\lambda\) is calculated as \(0 < \lambda < 1.3333\) using inequality (8). When \(\lambda = 1.333\) is within the range, the system converges to \(v^* = [0.3769 \ 0.3769]^T\), as shown in Fig. 1 (a), but when \(\lambda = 5.9\) is out of the range, the system oscillates between \(v^* = [0.1844 \ 0.1844]^T\) and \(v^* = [0.2519 \ 0.2519]^T\), as shown in Fig. 1 (b).

**4. Experimental Results**

We compared the results of the inference using the sigmoid activation functions designed by proposed method and those used in [2] and [6]. We considered three kinds of sigmoid functions having different \(\lambda\) values in Eq. (1): (i) as used in [6]; (ii) within the range of \(\lambda\) calculated by the proposed method; and (iii) as used in [2] and out of the range. We let \(\rho_1 = \rho_2 = 1\) and \(M = 1\). We calculated the range of \(\lambda\) as \(0 < \lambda < 1.93359008679557\) using inequality (8) with the following weight matrix \(w\) used in [6], and performed the inference with the initial state vector \(v^{(0)}\):

\[
w = \begin{bmatrix} 0 & -0.4 & -0.25 & 0 & 0.3 \\ 0.36 & 0 & 0 & 0 & 0 \\ 0.45 & 0 & 0 & 0 & 0 \\ -0.9 & 0 & 0 & 0 & 0 \\ 0.6 & 0 & 0.3 & 0 & 0 \end{bmatrix}, \quad v^{(0)} = \begin{bmatrix} 0.4 \\ 0.7077 \\ 0.6120 \\ 0.7171 \\ 0.3 \end{bmatrix}.
\]
Figure 2 shows the state values at each iteration step of the inference process and Table 1 shows the equilibrium state values by the inference with respect to the \( \lambda \) values. As the inference process continued, as shown in Fig. 2, the state values converged to equilibrium points. We also observed that the gaps between state values in Figs. 2 (a) and 2 (b) were wider than the gaps in Fig. 2 (c). To analyze the results objectively, we calculated the standard deviations of the equilibrium state values in Table 1 and had (i) 0.0460, (ii) 0.0725, and (iii) 0.0094 as the standard deviation values. It shows us that the equilibrium state values in (i) and (ii) are wider than the values in (iii). If the results are used for system operation as done in [6] (i.e., the system operators decide how much open the valves of a chemical plant based on the equilibrium state values), we can think the results of (i) and (ii) are more useful than the result of (iii), because the equilibrium state values in (i) and (ii) are enough widened for the operator to notice the state of nodes in the FCM. Moreover, the bigger the \( \lambda \) value was, the closer the equilibrium state values to ‘1’ were as the result of (iii). This may cause that the decision-makers hardly notice the differences of the state of the nodes. That is, the result of (iii) are less suitable to make decision in the FCM comparing with the results of (i) and (ii), even though \( \lambda = 5 \) was a good parameter for decision process in [2]. Consequently, we could judge that the inference not considering the characteristic of the FCMs can make decision-making tasks difficult. Moreover, we confirmed that the designed activation function based on the characteristic of the FCM is effective to decision-making process.

5. Conclusion

We proposed a method for designing sigmoid function, i.e., determining \( \lambda \) value, through Lyapunov stability analysis, which guarantees global exponential stability of the inference of FCMs. We showed the usefulness of the proposed method by determining a proper \( \lambda \) value of sigmoid function in the experiments.

Acknowledgments

This research was supported by the Yeungnam University Research Grants in 2009.

References

[1] B. Kosko, “Fuzzy cognitive maps,” Int. J. Man-machine Stud., vol.24, pp.65–75, 1986.
[2] S. Bueno and J.L. Salmeron, “Benchmarking main activation functions in fuzzy cognitive maps,” Expert Syst. Appl., vol.36, no.3, pp.5221–5229, 2009.
[3] J. Aguilar, “A dynamic fuzzy cognitive map approach based on random neural networks,” Int. J. Comp. Cogn., vol.1, no.4, pp.91–107, 2003.
[4] M.S. Khan and S.W. Khor, “A framework for fuzzy rule-based cognitive maps,” Proc. PRICAI 2004, pp.454–463, 2004.
[5] W. Stach, L. Kurgan, and W. Pedrycz, “Higher-order fuzzy cognitive maps,” Proc. NAFIPS 2006, pp.166–171, 2006.
[6] C.D. Styrios and P.P. Groumpos, “Fuzzy cognitive map in modeling supervisory control systems,” J. Intell. Fuzzy Syst., vol.8, pp.83–98, 2000.
[7] A.K. Tsadiras, “Comparing the inference capabilities of binary, trivalent and sigmoid fuzzy cognitive maps,” Inf. Sci., vol.178, pp.3880–3894, 2008.
[8] C.D. Stylios and P.P. Groumpos, “Mathematical formulation of fuzzy cognitive maps,” Proc. 7th Mediterranean Conference on Control and Automation (MED99), pp.28–30, 1999.
[9] Z. Wang, Y. Liu, and X. Liu, “On global asymptotic stability of neural networks with discrete and distributed delays,” Phys. Lett. A, vol.345, pp.299–308, 2005.
[10] Y. Liu, Z. Wang, A. Serrano, and X. Liu, “Discrete-time recurrent neural networks with time-varying delays: Exponential stability analysis,” Phys. Lett. A, vol.362, pp.480–488, 2007.