Letter

On RNN-Based *k*-WTA Models With Time-Dependent Inputs

Mei Liu and Mingsheng Shang

Dear editor,

This letter identifies two weaknesses of state-of-the-art k-winners-take-all (k-WTA) models based on recurrent neural networks (RNNs) when considering time-dependent inputs, i.e., the lagging error and the infeasibility in finite-time convergence based on the Lipschitz continuity. Specifically, in the case of time-dependent inputs, theoretical analyses and simulations are conducted to illustrate that the lagging error is inevitable for the dual network model based on RNN. Then, a new k-WTA model aided with RNN is constructed in this letter with the ability of eliminating the lagging error. Theoretical analyses demonstrate that the finite-time convergence of the existing k-WTA models based on the Lipschitz continuity with time-dependent inputs cannot be achieved. Besides, this letter offers a feasible solution to perform k-WTA operations with desired convergent speed efficiently and precisely.

Introduction: Deemed as a competitive neural network, winner-takes-all (WTA) networks gain widespread applications in various fields. As a generalized form, k-WTA is widely applied to the modelling of a competitive frame with k denoting the number of winners [1]. For example, the k-WTA network is leveraged to explore the evolution law of individuals in social systems [2]. A dynamic thresholding k-WTA model owing relatively simple structure and faster convergent speed is presented in [3]. Moreover, an approach to complete task object tracking of multiple robots using the k-WTA strategy is designed in [4].

Recurrent neural RNN algorithms achieve great success for optimal online solutions over the recent years. For example, an RNN model with the aid of the saddle-point theorem is utilized to dispose of the non-convex optimization problem, which is applied to the identification problem of genetic regulatory networks [5]. A deep RNN model is constructed to predict the residual life of the roller by exploiting a comprehensive health indicator [6]. A gesture prediction model based on RNN is presented and verified on social robots [7]. In a nutshell, RNN-based models demonstrate the high advantages in processing optimization problems.

In view of the feasibility that the k-WTA operation can be transformed into a constrained optimization problem, many RNN algorithms are investigated for online solutions of k-WTA. A series of k-WTA models aided with the dual neural network is studied to deal with the k-WTA operations [8], [9]. The above RNN-based k-WTA schemes demonstrate their strengths with different investigation focuses. However, the lagging error is inevitable in handling k-WTA operations with time-dependent inputs for the existing RNN-based k-WTA models, which leads to an unsatisfactory performance. Qi *et al.* [10] introduce a robust k-WTA method with time-dependent inputs, which overcomes the time-lagging error for disposing k-WTA problems, while it has higher computational complexity than most of existing k-WTA models. This letter constructs a simplified RNN-based k-WTA model which is capable of eliminating the lagging error. Furthermore, theoretical analyses and simulative results are provided to verify the validity of the constructed RNN-based k-WTA model. It is worth noting that, the existing models, including the model constructed in this letter, are not feasible in terms of finite-time convergence based on the Lipschitz continuity with time-dependent inputs.

This letter analyzes weaknesses of the existing RNN-based k-WTA models employed for the k-WTA operations with time-dependent inputs. Then, a new RNN-based k-WTA model which can be used in the competitive coordination of a multi-robot system [11]−[13] is constructed with lagging errors conquered. Moreover, a feasible replacement is provided to achieve the finite-time convergence of the RNN-based k-WTA models in conducting the time-dependent k-WTA operations.

Preliminary: The mathematical expression for the k-WTA operations is formulated as

$$y_i = f(x_i) = \begin{cases} 1, & \text{if } x_i \in \{k \text{ largest elements of } \boldsymbol{x}\} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where x_i and y_i represent the ith element of the corresponding inputs $\boldsymbol{x}$ and outputs $\boldsymbol{y}$ of the system, respectively; k signifies the number of selected winners from $\boldsymbol{y}$. According to [8]−[10], (1) is converted into a quadratic programming problem

$$\begin{aligned} &\min\ \delta \boldsymbol{y}^T(t)\boldsymbol{y}(t) - \boldsymbol{x}^T(t)\boldsymbol{y}(t) \\ &\text{s.t. } \boldsymbol{\beta}^T\boldsymbol{y}(t) = k \\ &\qquad 0 \le y_i(t) \le 1,\ i = 1,2,\ldots,m \end{aligned} \tag{2}$$

where $\boldsymbol{\beta} = [1;1;\ldots;1] \in \mathbb{R}^m$ and m denotes the dimension of the k-WTA system; the superscript T is the transpose operation; $\delta \le (\bar{x}_k(t) - \bar{x}_{k+1}(t))/2$ is a positive scalar to guarantee the existence and uniqueness of convex solutions, in which $\bar{x}_k(t)$ and $\bar{x}_{k+1}(t)$ denote the kth and $(k+1)$th largest values of inputs, respectively [14].

Dual-network-based k-WTA model: The dual-network-based k-WTA model in [8] is utilized to prove the lagging error in the time-dependent case. It is described as

$$\begin{aligned} &\text{State equation} && \dot{\boldsymbol{u}}(t) = \gamma(-\mathcal{D}\boldsymbol{u}(t) + \mathcal{A}(\mathcal{D}\boldsymbol{u}(t) - \boldsymbol{u}(t) + \boldsymbol{\epsilon}) - \boldsymbol{\epsilon}) \\ &\text{Output equation} && \boldsymbol{y}(t) = \mathcal{D}\boldsymbol{u}(t) + \boldsymbol{\epsilon} \end{aligned} \tag{3}$$

where $\gamma > 0$ represents a coefficient; $\mathcal{D} = (I - \boldsymbol{\beta}\boldsymbol{\beta}^T/m)/\delta$; $\boldsymbol{u}(t) \in \mathbb{R}^m$ is the state vector; $\boldsymbol{\epsilon} = \mathcal{D}\boldsymbol{x}(t) + k\boldsymbol{\beta}/m$; $\mathcal{A}(\cdot)$ is described as

$$\mathcal{A}(e) = \begin{cases} 1, & \text{if } e > 1 \\ e, & \text{if } 0 \le e \le 1 \\ 0, & \text{if } e < 0. \end{cases}$$

Theorem 1: With time-dependent inputs, the lagging error is inevitable when performing the k-WTA operations with the dual-network-based k-WTA model (3).

Proof: In order to explore the convergence of the dual-network-based k-WTA (3), an auxiliary Lyapunov function is constructed as

$$\mathcal{L}(t) = \|\mathcal{M}\boldsymbol{\varsigma}(t)\|_2^2/2 \ge 0$$

where $\|\cdot\|_2$ denotes the Euclid norm; $\boldsymbol{\varsigma}(t) = \boldsymbol{u}(t) - \widehat{\boldsymbol{u}}(t)$ signifies the residual error with $\widehat{\boldsymbol{u}}(t)$ representing the equilibrium point at time instant t; $\mathcal{M}$ is a symmetric positive matrix and $\mathcal{M}^2 = \upsilon(I + \mathcal{D})$ with $\upsilon = 1/\gamma$. Computing the time derivative of $\mathcal{L}(t)$ leads to

$$\begin{aligned} \dot{\mathcal{L}}(t) &= \boldsymbol{\varsigma}^T(t)\mathcal{M}^2\left(\dot{\boldsymbol{u}}(t) - \dot{\widehat{\boldsymbol{u}}}(t)\right) \\ &= \boldsymbol{\varsigma}^T(t)\upsilon(I+\mathcal{D})\dot{\boldsymbol{u}}(t) - \boldsymbol{\varsigma}^T(t)\upsilon(I+\mathcal{D})\dot{\widehat{\boldsymbol{u}}}(t). \end{aligned} \tag{4}$$

Combining (4) with (3) generates

$$\dot{\mathcal{L}}(t) = \boldsymbol{\varsigma}^T(t)(I+\mathcal{D})\mathcal{S} - \boldsymbol{\varsigma}^T(t)\upsilon(I+\mathcal{D})\dot{\widehat{\boldsymbol{u}}}(t)$$

where $\mathcal{S} = -\mathcal{D}\boldsymbol{u}(t) + \mathcal{A}(\mathcal{D}\boldsymbol{u}(t) - \boldsymbol{u}(t) + \boldsymbol{\epsilon}) - \boldsymbol{\epsilon}$. Defining $\|\dot{\widehat{\boldsymbol{u}}}(t)\|_2 \le \eta$ with η expressing a positive factor, we have

Corresponding author: Mingsheng Shang.

Citation: M. Liu and M. S. Shang, "On RNN-based k-WTA models with time-dependent inputs," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 11, pp. 2034–2036, Nov. 2022.

The authors are with the Chongqing Key Laboratory of Big Data and Intelligent Computing, Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, and also with the Chongqing School, University of Chinese Academy of Sciences, Chongqing 400714, China (e-mail: liumeisysu@qq.com; msshang@cigit.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JAS.2022.105932

$$\dot{\mathcal{L}}(t) \le \boldsymbol{\varsigma}^T(t)(I+\mathcal{D})\mathcal{S} + \eta\upsilon\|\boldsymbol{\varsigma}^T(t)(I+\mathcal{D})\|_2.$$

By leveraging the known inference stated in [15] that

$$\boldsymbol{\varsigma}^T(t)\mathcal{S} + \mathcal{S}^T\mathcal{D}\boldsymbol{\varsigma}(t) \le -\|\mathcal{S}\|_2^2 - \boldsymbol{\varsigma}^T(t)\mathcal{D}\boldsymbol{\varsigma}(t) \le 0$$

it is further obtained

$$\begin{aligned}\dot{\mathcal{L}}(t) &\le -\|\mathcal{S}\|_2^2 - \boldsymbol{\varsigma}^T(t)\mathcal{D}\boldsymbol{\varsigma}(t) + \eta\upsilon\|\boldsymbol{\varsigma}^T(t)(I+\mathcal{D})\|_2 \\ &\le -\|-\mathcal{D}\boldsymbol{u}(t) + \mathcal{A}(\mathcal{D}\boldsymbol{u}(t) - \boldsymbol{u}(t) + \boldsymbol{\epsilon}) - \boldsymbol{\epsilon}\|_2^2 \\ &\quad - \boldsymbol{\varsigma}^T(t)\mathcal{D}\boldsymbol{\varsigma}(t) + \eta\upsilon\|\boldsymbol{\varsigma}^T(t)(I+\mathcal{D})\|_2. \end{aligned} \tag{5}$$

There is bound to exist a situation that a positive coefficient $\widehat{\Gamma}$ could be found to propel $\|-\mathcal{D}\boldsymbol{u}(t) + \mathcal{A}(\mathcal{D}\boldsymbol{u}(t) - \boldsymbol{u}(t) + \boldsymbol{\epsilon}) - \boldsymbol{\epsilon}\|_2^2 \ge \widehat{\Gamma}\|\boldsymbol{\varsigma}(t)\|_2^2$. Hence, (5) is reformulated as

$$\begin{aligned}\dot{\mathcal{L}}(t) &\le -\widehat{\Gamma}\|\boldsymbol{\varsigma}(t)\|_2^2 - \boldsymbol{\varsigma}^T(t)\mathcal{D}\boldsymbol{\varsigma}(t) + \eta\upsilon\|\boldsymbol{\varsigma}^T(t)(I+\mathcal{D})\|_2 \\ &\le -\widehat{\Gamma}\|\boldsymbol{\varsigma}(t)\|_2^2 + \lambda_1\|\boldsymbol{\varsigma}(t)\|_2^2 + \eta\upsilon\|\boldsymbol{\varsigma}(t)\|_2 + \eta\upsilon\lambda_2\|\boldsymbol{\varsigma}(t)\|_2 \\ &= \|\boldsymbol{\varsigma}(t)\|_2\left((\lambda_1 - \widehat{\Gamma})\|\boldsymbol{\varsigma}(t)\|_2 + \eta\upsilon(1+\lambda_2)\right) \end{aligned} \tag{6}$$

of which the minimal eigenvalue λ_1 and the maximum one λ_2 of $\mathcal{D}$ are applied in (6) for simplification. Then, the analysis of the convergence of $\boldsymbol{\varsigma}(t)$ generated by the dual-network-based *k*-WTA model (3) for performing the *k*-WTA operations exploiting the Lyapunov theory is equivalent to discussing the value of $(\lambda_1 - \widehat{\Gamma})\|\boldsymbol{\varsigma}(t)\|_2 + \eta\upsilon(1+\lambda_2)$.

1) $\lambda_1 \ge \widehat{\Gamma}$: $(\lambda_1 - \widehat{\Gamma})\|\boldsymbol{\varsigma}(t)\|_2 + \eta\upsilon(1+\lambda_2) > 0$ holds invariably for all cases of error $\boldsymbol{\varsigma}(t)$. One readily has $\dot{\mathcal{L}}(t) > 0$, which means that $\boldsymbol{\varsigma}(t)$ is of divergence.

2) $\lambda_1 < \widehat{\Gamma}$: There exist two possible conditions with

a) $(\lambda_1 - \widehat{\Gamma})\|\boldsymbol{\varsigma}(t)\|_2 + \eta\upsilon(1+\lambda_2) > 0$ for $\|\boldsymbol{\varsigma}(t)\|_2 < \eta\upsilon(1+\lambda_2)/(\widehat{\Gamma} - \lambda_1)$. One has $\dot{\mathcal{L}}(t) > 0$, which means that $\boldsymbol{\varsigma}(t)$ is of divergence until $\|\boldsymbol{\varsigma}(t)\|_2 = \eta\upsilon(1+\lambda_2)/(\widehat{\Gamma} - \lambda_1)$.

b) $(\lambda_1 - \widehat{\Gamma})\|\boldsymbol{\varsigma}(t)\|_2\eta\upsilon(1+\lambda_2) \le 0$ for $\|\boldsymbol{\varsigma}(t)\|_2 \ge \eta\upsilon(1+\lambda_2)/(\widehat{\Gamma} - \lambda_1)$. We have $\dot{\mathcal{L}}(t) \le 0$, which indicates that $\boldsymbol{\varsigma}(t)$ is of convergence until $\|\boldsymbol{\varsigma}(t)\|_2 = \eta\upsilon(1+\lambda_2)/(\widehat{\Gamma} - \lambda_1)$.

In a nutshell, for the first case with $\lambda_1 \ge \widehat{\Gamma}$, $\boldsymbol{\varsigma}(t)$ diverges all the time; for the second case with $\lambda_1 < \widehat{\Gamma}$, $\boldsymbol{\varsigma}(t)$ goes to a bound $\eta\upsilon(1+\lambda_2)/(\widehat{\Gamma} - \lambda_1)$. Hence, the dual-network-based *k*-WTA model (3) is of less efficiency in performing the *k*-WTA operations (2). ■

In the ensuing part, a new RNN-based *k*-WTA model is constructed to remedy the lagging errors of the existing models with time-dependent inputs considered.

New RNN-based *k*-WTA model: A new *k*-WTA mathematical model applies the RNN-based dynamic formula $\dot{\boldsymbol{\varphi}}(t) = -\gamma\boldsymbol{\varphi}(t)$ with $\boldsymbol{\varphi}(t) = \mathcal{W}\boldsymbol{z}(t) - \boldsymbol{\rho}(t)$ denoting the error function of the *k*-WTA system to process the *k*-WTA operations (2)

$$\mathcal{W}\dot{\boldsymbol{z}}(t) = \dot{\boldsymbol{\rho}}(t) - \dot{\mathcal{W}}\boldsymbol{z}(t) - \gamma(\mathcal{W}\boldsymbol{z}(t) - \boldsymbol{\rho}(t)) \tag{7}$$

where $\gamma > 0$; $\boldsymbol{z}(t) = [\boldsymbol{y}(t);\psi(t);\boldsymbol{\theta}(t)] \in \mathbb{R}^{m+n+1}$ are the decision variable and output of the model where $\psi(t) \in \mathbb{R}$ and $\boldsymbol{\theta}(t) \in \mathbb{R}^n$ are Lagrangian multipliers with $n = 2m$; inertia matrix $\mathcal{W} = [2\delta I_{m\times m}\boldsymbol{\beta}E^T(t); \boldsymbol{\beta}^T 00; -E(t)0I_{n\times n}] \in \mathbb{R}^{(m+n+1)\times(m+n+1)}$ with $E(t) = [I_{m\times m}; -I_{m\times m}] \in \mathbb{R}^{n\times m}$; $\rho(\mathrm{t}) = [\mathrm{x}(\mathrm{t});\mathrm{k};\rho_3(\mathrm{t})] \in \mathbb{R}^{\mathrm{m+n+1}}$ and $\boldsymbol{\rho}_3(t) = -\}(t) + \sqrt{\boldsymbol{\zeta}(t)\circ\boldsymbol{\zeta}(t) + \boldsymbol{\theta}(t)\circ\boldsymbol{\theta}(t) + \sigma}$ where $\}(t) = [\boldsymbol{\beta};0] \in \mathbb{R}^n$; $\boldsymbol{\zeta}(t) = \}(t) - E(t)\boldsymbol{y}(t)$ and $\sigma \to 0_+ \in \mathbb{R}^n$.

Theorem 2: The error function $\boldsymbol{\varphi}(t)$ of the RNN-based *k*-WTA model (7) globally and exponentially converges to 0.

Proof: According to the general solution form of the first-order differential equation, the error function solution of the *k*-WTA system is $\boldsymbol{\varphi}(t) = \boldsymbol{\varphi}(0)\exp(-\gamma t)$ with $\boldsymbol{\varphi}(0)$ denoting the initial value of the system error. It is easy to conclude that the error function of the RNN-based *k*-WTA model (7) globally and exponentially converges to 0. ■

In terms of finite-time convergence, there exist many kinds of techniques dedicated to research on it [16]. However, finite-time convergence based on the Lipschitz continuity for the *k*-WTA operations with time-dependent inputs may be impractical. The discussion on the convergent time to steady-state outputs of the RNN-based *k*-WTA model (7) is set forth below.

Theorem 3: The finite-time convergence is realized by exploiting a specific type of activation function. Nonetheless, these series of activation functions do not satisfy the Lipschitz condition and thus fail in being implemented in practice.

Proof: The function $f(\boldsymbol{x})$ is Lipschitz continuous in the domain $\boldsymbol{Q}$, if there exists $\chi > 0 \in \mathbb{R}$ such that for any $x_1, x_2 \in \boldsymbol{Q}$, one has

$$\|f(x_1) - f(x_2)\|_2 \le \chi\|x_1 - x_2\|_2. \tag{8}$$

As a consequence, the function satisfies Lipschitz continuity, assuring the existence and uniqueness of the solution to the optimization problem, and limiting the change rate of the function to a value less than the Lipschitz constant χ.

The existing activation functions utilized to achieve the finite-time convergence property in RNN-based *k*-WTA models are provided as follows:

1) The activation function presented in [16] is shown as $\mathcal{F}_{v1}(\varphi_i) = \Phi^a(\varphi_i) + \Phi^{1/a}(\varphi_i)$ with

$$\Phi^a(\varphi_i) = \begin{cases} |\varphi_i|^a, & \text{if } \varphi_i > 0 \\ 0, & \text{if } \varphi_i = 0 \\ -|\varphi_i|^a, & \text{if } \varphi_i < 0 \end{cases}$$

where the parameter $0 < a < 1$.

2) The activation function presented in [16] is shown as

$$\begin{cases} \mathcal{F}_{v2}(\varphi_i) = \exp\left(|\varphi_i|^b\right)|\varphi_i|^{1-b}\operatorname{sign}(\varphi_i)/b \\ \mathcal{F}_{v3}(\varphi_i) = \left(|\varphi_i|^{1-b} + \dfrac{1-b}{b}|\varphi_i|^{1-2b}\right)\exp(|2\varphi_i|^b\operatorname{sign}(\varphi_i)/b \end{cases}$$

where factor $0 < b < 1/2$; the function $\operatorname{sign}(\varphi_i) = 0$ for $\varphi_i = 0$ and $\operatorname{sign}(\varphi_i) = \varphi_i/|\varphi_i|$ for $\varphi_i \ne 0$.

To substantiate the Lipschitz continuity with the aid of the above-mentioned definition, it is obtained that the derivative of $\mathcal{F}_{vi}(\cdot)(i = 1,2,3)$ is related to χ. Therefore, for 1),

$$\dot{\mathcal{F}}_{v1}(\varphi_i) = \begin{cases} a{\varphi_i}^{a-1} + \frac{1}{a}{\varphi_i}^{1/a-1}, & \text{if } \varphi_i > 0 \\ 0, & \text{if } \varphi_i = 0 \\ a(-\varphi)_i^{a-1} + \frac{1}{a}(-\varphi)_i^{1/a-1}, & \text{if } \varphi_i < 0. \end{cases}$$

It is evident that for $0 < a < 1$, when the residual error φ_i tends to zero (either to 0_+ or 0_-), $\dot{\mathcal{F}}_{v1}(\varphi_i)$ would go to infinity, which manifests that no constant χ is able to be determined to satisfy the Lipschitz condition and then the activation function does not fulfil the Lipschitz continuity. Further, for 2),

$$\dot{\mathcal{F}}_{v2}(\varphi_i) = \begin{cases} \exp\left({\varphi_i}^b\right) + \frac{1-b}{b}\exp\left({\varphi_i}^b\right){\varphi_i}^{-b}, & \text{if } \varphi_i > 0 \\ 0, & \text{if } \varphi_i = 0 \\ -\exp\left((-\varphi_i)^b\right) - \frac{1-b}{b}\exp\left((-\varphi_i)^b\right)(-\varphi_i)^{-b}, & \text{if } \varphi_i < 0. \end{cases}$$

and

$$\dot{\mathcal{F}}_{v3}(\varphi_i) =$$

$$\begin{cases} \left((1-b)\varphi_i^{-b} + \frac{(1-b)(1-2b)}{b}\varphi_i^{-2b}\right)\exp\left(2\varphi_i^b\right)/b & \\ +\left(b2^b\exp(2{\varphi_i}^b) + 2^b(1-b)\varphi_i^{-b}\exp(2\varphi_i^b)\right)/b, & \text{if } \varphi_i > 0 \\ 0, & \text{if } \varphi_i = 0 \\ \left(\frac{1-b}{b}(-\varphi)_i^{-b} + \frac{(1-b)(1-2b)}{b^2}(-\varphi)_i^{-2b}\right)\exp\left((-2\varphi_i)^b\right) & \\ -2^b\exp\left((-2\varphi_i)^b\right) + 2^b\frac{1-b}{b}(-\varphi_i)^{-b}\exp\left((-2\varphi_i)^b\right), & \text{if } \varphi_i < 0. \end{cases}$$

For the above two activation functions, it is derived similarly that they do not satisfy the requirement of the Lipschitz continuity. Thus, implementing the models equipped with these types of activation functions with finite-time convergence is extremely difficult or even impossible in practice. ■

Feasible replacement to the finite-time convergence: The absolute zero error, i.e., a completely exact solution is non-existent on account of the environmental disturbance and hardware errors, and the most desirable result is to obtain an optimal solution with the expected precision in the preset time.

Generally, achieving the result that the residual error $\varphi(t)$ converges to a predefined tiny value (e.g., 10^{-6} which is larger than 0) in any short time requires a sufficiently large γ. In addition, under the condition that the proportional coefficient γ is sufficiently large, after the time period of $4/\gamma$ s, $|\varphi_i(t)|$ would be less than 1.85% of $|\varphi_i(0)|$, $\forall i \in [1, m+n+1]$. In other words, for $\gamma = 400$, $|\varphi_i(t)|$ is no more than $0.0185 \times |\varphi_i(0)|$ at $t = 0.01$ s, and less than $4.25 \times 10^{-18} \times |\varphi_i(0)|$ at

$t = 0.1$ s. This means that, when the residual error is less than $4.25 \times 10^{-18} \times |\varphi_i(0)|$ at $t = 0.1$ s, the steady-state outputs of the RNN-based k-WTA model (7) are obtained and hold at each time instant.

Experiments: A group of four sine signals as inputs of the k-WTA system are employed to demonstrate the results of the dual-network-based k-WTA model (3) and the RNN-based k-WTA model (7) with different γ. In simulations, the sinusoidal functions are set as $x(t) = -\sin(2\pi(t+0.8(i-1)))(i = 1,2,3,4)$; $\delta = 0.001$ and the selected winners $k = 2$. Evidently, in Figs. 1 (a) and 1 (b), the dual-network-based k-WTA model makes a wrong selection for the winners. Compared with Fig. 1(a), when $\gamma = 1$, results in Fig. 2(a) are consistent with the demonstration of Theorem 3 that the RNN-based k-WTA model is exponentially convergent. According to Figs. 1(c) and 1(d), the dual-network-based k-WTA model (3) shows the accurate selection of time-dependent inputs by increasing the coefficient γ. However, there are noticeable prickles and vibrations that can generate unacceptable errors and even select wrong winners. Significantly, as demonstrated in Fig. 2, with different γ, the RNN-based k-WTA model (7) is capable of accurately and effectively selecting the winners in the time-dependent k-WTA system.

Conclusions: This letter has analyzed the limitations of the existing RNN-based models for executing the k-WTA operations with time-dependent inputs considered. Then, theoretical analyses and simulative results have provided that the dual-network-based k-WTA model (3) is inefficient in eliminating the lagging error for dealing with the k-WTA operations with time-dependent inputs. In addition, the infeasibility of the RNN-based k-WTA models on the finite-time convergence based on the Lipschitz continuity has been verified through theoretical analyses, and feasible alternatives are provided. This letter has raised a new RNN-based k-WTA model (7) that makes up the deficiency of the dual-network-based k-WTA model. Theoretical analyses have shown that the feasibility of the new RNN-based k-WTA model when handling the k-WTA operations considering the time-dependent inputs.

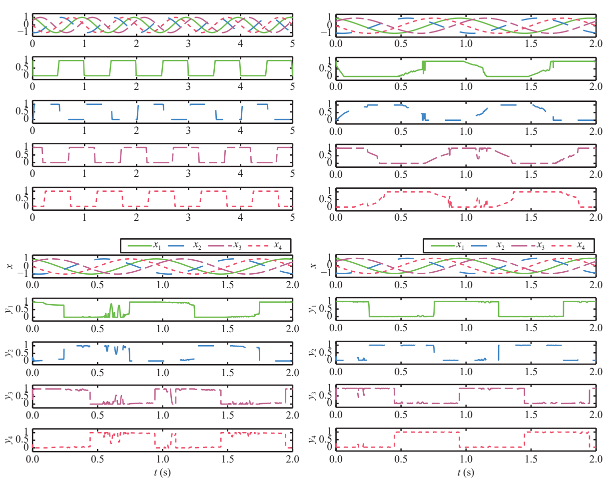

Fig. 1. Inputs and outputs of dual-network-based k-WTA model (3) with different γ. (a) $\gamma = 1$; (b) $\gamma = 10$; (c) $\gamma = 100$; (d) $\gamma = 1000$.

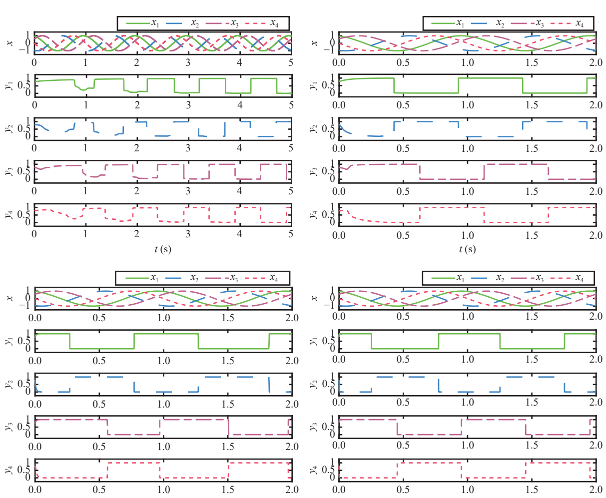

Fig. 2. Inputs and outputs of RNN-based k-WTA model (7) with different γ. (a) $\gamma = 1$; (b) $\gamma = 10$; (c) $\gamma = 100$; (d) $\gamma = 1000$.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (62072429) and the Key Cooperation Project of Chongqing Municipal Education Commission (HZ2021017, HZ2021008).

References

[1] Y. Yang, L. Liao, H. Yang, and S. Li, "An optimal control strategy for multi-UAVs target tracking and cooperative competition," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 12, pp. 1931–1947, Jan. 2020.

[2] M. Liu, S. Li, and L. Jin, "Modeling and analysis of Matthew effect under switching social networks via distributed competition," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 6, pp. 1–4, Jun. 2022.

[3] J. Yang and C. Chen, "A dynamic k-winner-take-all neural network," *IEEE Trans. Syst. Man Cybern. -Syst.* , vol. 51, no. 10, pp. 6466–6476, Oct. 2021.

[4] L. Jin and S. Li, "Distributed task allocation of multiple robots: A control perspective," *IEEE Trans. Syst. Man Cybern. – Syst.*, vol. 48, no. 5, pp. 693–701, May 2018.

[5] L. Cheng, Z. Hou, and F. Wu, "Recurrent neural network for non-smooth convex optimization problems with application to the identification of genetic regulatory networks," *IEEE Trans. Neural Netw.* , vol. 22, no. 5, pp. 714–726, May 2011.

[6] R. Jiao, K. Peng, and J. Dong, "Remaining useful life prediction for a roller in a hot strip mill based on deep recurrent neural networks," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 7, pp. 1345–1354, Jul. 2021.

[7] B. Wu, J. Zhong, and C. Yang "A visual-based gesture prediction framework applied in social robots," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 3, pp. 510–519, Mar. 2022.

[8] Q. Liu, C. Dang, and J. Cao, "A novel recurrent neural network with one neuron and finite-time convergence for k-winners-take-all operation," *IEEE Trans. Neural Netw.* , vol. 21, no. 7, pp. 1140–1148, Jul. 2010.

[9] S. Liu and J. Wang, "A simplified dual neural network for quadratic programming with its KWTA application," *IEEE Trans. Neural Netw.* , vol. 17, no. 6, pp. 1500–1510, Nov. 2006.

[10] Y. Qi, L. Jin, X. Luo, Y. Shi, and M. Liu, "Robust k-WTA network generation, analysis, and applications to multiagent coordination," *IEEE Trans. Cybern.*, 2021. DOI: 10.1109/TCYB.2021.3079457.

[11] Y. Ju, D. Ding, X. He, Q.-L. Han, and G. Wei, "Consensus control of multi-agent systems using fault-estimation-in-the-loop: Dynamic event-triggered case," *IEEE/CAA J. Autom. Sinica* , vol. 9, no. 8, pp. 1440–1451, Aug. 2022.

[12] I. Ahmad, X. Ge, and Q.-L. Han, "Decentralized dynamic event-triggered communication and active suspension control of in-wheel motor driven electric vehicles with dynamic damping," *IEEE/CAA J. Autom. Sinica,* vol. 8, no. 5, pp. 971–986, May 2021.

[13] J. Wang, Y. Hong, J. Wang, J. Xu, Y. Tang, Q. Han, and J. Kurths, "Cooperative and competitive multi-agent systems: From optimization to games," *IEEE/CAA J. Autom. Sinica,* vol. 9, no. 5, pp. 763–783, May 2022.

[14] M. Liu, X. Zhang, M. Shang, and L. Jin, "Gradient-based differential kWTA network with application to competitive coordination of multiple robots," *IEEE/CAA J. Autom. Sinica*, vol. 9, no. 8, pp. 1452–1463, Aug 2022.

[15] Y. Zhang, J. Wang, and Y. Xia, "A dual neural network for redundancy resolution of kinematically redundant manipulators subject to joint limits and joint velocity limits," *IEEE Trans. Neural Netw.*, vol. 14, no. 3, pp. 658–667, May 2003.

[16] W. Li, "A recurrent neural network with explicitly definable convergence time for solving time-variant linear matrix equations," *IEEE Trans. Ind. Informat* , vol. 14, no. 12, pp. 5289–5298, Dec. 2018.