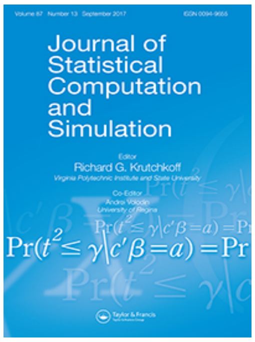

Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Some properties of nonlinear Lanchester equations with an application in military

Donghyun Kim, Hyungil Moon & Hayong Shin

JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION, 2017
VOL. 87, NO. 13, 2470–2479
https://doi.org/10.1080/00949655.2017.1296441

Some properties of nonlinear Lanchester equations with an application in military

Donghyun Kim, Hyungil Moon and Hayong Shin

Industrial and Systems Engineering, Korea Advanced Institute of Science and Technology, Daejeon, South Korea

ABSTRACT
There have been many research literature on traditional direct fire combat modelling. Recently, network centric warfare (NCW) is an active research topic, in which information plays more important role than in the traditional warfare. It can be easily agreed that the use of information affects the combat results greatly. However, it is not straightforward to measure the effect of the information, thus decision making involving the impact of information during combat is a non-trivial task. In this study, we propose a simple model for NCW modified from the original Lanchester differential equation, which can be used as a basic model for analysing characteristics of NCW. We derive some useful properties of the model in a special case. In order to solve the optimal fire allocation decision under this model in general case, we propose an algorithm based on reinforcement learning, followed by numerical examples.

ARTICLE HISTORY
Received 5 September 2016
Accepted 14 February 2017

KEYWORDS
Network centric warfare; combat modelling; direct policy search

1. Introduction

Describing the warfare is a traditional and fundamental area in the military field. Lanchester [1] suggested a foundational model with the differential equations. After his work, this domain grows constantly; see overview of combat modelling [2,3]. Especially, modelling the modern warfare including the concept of network centric warfare (NCW) or force multiplier is an emerging field.

The concept of NCW can be summarized into one phrase, 'The importance of collecting and utilizing the information'; for an overview of NCW see [4–8]. Since the information is not only intangible, but also hard to collect and utilize, the effort on developing and designing the metric for NCW is addressed in [9]. Network attack can be interpreted differently. A discontinuous shock like an electromagnetic pulse attack is modelled in [10] and a malware spread analyse with an epidemic SIR model is modelled in [11]. Not like traditional direct fire model, modelling the modern warfare becomes hard because of complexity and ambiguity. So we propose an introductory model for NCW in Section 2 with differential equations. This model is intuitive and includes all the concepts of NCW and force multiplier.

Beside, researchers are interested in the optimal decision making problems in military field as more than modelling the battle. Since the purpose of battle is to win against the enemy, the optimal fire allocation problem is natural to ask, and many studied about the weapon-target assignment problems [12–15]. Since there are many interpretations about the term NCW, there is no basic model like Lanchester [1] in the field of modern warfare. In this paper, we interpret the network power as a signal corpsman or an information processing centre that assist the attack forces to hit effectively. In other

CONTACT Hayong Shin hyshin@kaist.ac.kr

words, if the network power decreases, attrition rate of the attack forces will be weaken. So we suggest the basic model with the differential equations based on this idea. Like any other works, we focus on the optimal fire allocation within the view of blue force; what is the optimal fire proportion that maximizes the blue force at the end of battle. Our main contribution is that we proved some analytical properties of this model for a restricted case and suggested the reinforcement learning technique to solve the optimal allocation problem for complex cases.

In the next section, we present the basic model we suggest and prove some analytical properties for the linear attrition rate function. And then we propose our main contribution, the optimal fire allocation strategy using the reinforcement learning technique called the direct policy search (DPS) algorithm. We present the numerical results for several cases and conclude with the promising future works that can be done.

2. New model for NCW and properties

The key idea of NCW is the utilization of the network power. Not only the amount of information, but also to utilizing the information is very important. On the original combat models, all the fighting power or ability was denoted as some constants or independent variables which is called an attrition rate or a kill rate. These models cannot apply the concept of force multiplier or NCW because of the limitation of model structure. So we propose the introductory model for NCW as shown in Figure 1. The main difference from the previous literature is that the attrition rate is represented as a function of network power. Below are the notations that we use through this paper.

B	=	Blue force
A	=	Red force (Attack)
N	=	Red force (Network)
β_A	=	an attrition rate of B to A
β_N	=	an attrition rate of B to N
$f_\alpha(N)$	=	an attrition function of A to B
π	=	the fire allocating proportion of B to A

Since this is the introductory model, we assume only red forces have the network, N, and the blue forces does not and only A can attack B with the attrition rate $f_\alpha(N)$. We can represent Figure 1 as a

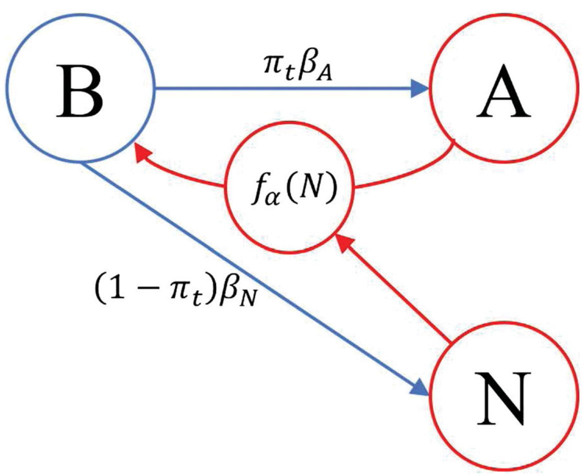

Figure 1. Basic combat model including NCW.

system of differential equations,

$$\frac{dB}{dt} = -f_\alpha(N)A, \tag{1}$$

$$\frac{dA}{dt} = -\pi_t \beta_A B, \tag{2}$$

$$\frac{dN}{dt} = -(1-\pi_t)\beta_N B. \tag{3}$$

2.1. Some useful properties if $f_\alpha(N)$ is a linear function of N

If $f_\alpha(N) = \alpha_d + (\alpha_c - \alpha_d)(N/N_0)$, which is a linear function of N, then we can derive useful properties of the optimal fire allocation. In here, α_c denotes the fully-connected attrition rate of A and α_d denotes the fully disconnected attrition rate, where $\alpha_d \le \alpha_c$. The differential Equations (1)–(3) can be rewritten as

$$\frac{dB}{dt} = -\left(\alpha_d + (\alpha_c - \alpha_d)\frac{N}{N_0}\right)A \tag{4}$$

$$\frac{dA}{dt} = -\pi_t \beta_A B \tag{5}$$

$$\frac{dN}{dt} = -(1-\pi_t)\beta_N B. \tag{6}$$

Lemma 2.1: *If $\pi_t = \pi$ for any time t, where $\pi \in [0,1]$, then one of $\pi = 0$ and $\pi = 1$ achieves the optimal fire allocation to maximize the Blue force at any time.*

Proof: If $\pi_t = \pi$ is a fixed constant over time t, this differential equation type is special type of second-order nonlinear ordinary differential equation which is called a second-order autonomous system. Let $X(t) = \int_0^t B(s)\, ds$, then Equations (4)–(6) can be rewritten as a second-order differential equation of $X(t)$:

$$X''(t) = -C_1 X^2(t) + C_2 X(t) - C_3, \tag{7}$$

where

$$C_1 = \frac{\pi(1-\pi)\beta_A\beta_N(\alpha_c - \alpha_d)}{N_0}$$

$$C_2 = \frac{(1-\pi)\beta_N(\alpha_c - \alpha_d)A_0 + \pi\beta_A\alpha_c N_0}{N_0}$$

$$C_3 = \alpha_c A_0.$$

This form of differential equations has an implicit solution form, but it's hard to use in practice even though we fix the allocation π. By simple calculation,

$$X'(t) = \sqrt{-\tfrac{2}{3}C_1 X^3(t) + C_2 X^2(t) - C_3 X(t) + C_4}, \tag{8}$$

where C_4 is an integration constant. By definition of $X(t)$, $X'(t)$ is equal to $B(t)$. Since $X(t)$ is a positive value and C_1, C_2, C_3 are non-negative values, greater C_2 and lower C_1 and C_3 make $X'(t)$ a higher value. We can easily derive the result that $\pi = 0$ or $\pi = 1$ is optimal; $\max_{\pi \in [0,1]} C_2 = \max_{\pi \in \{0,1\}} C_2$, because C_2 is a linearly weight average of two numbers with respect to π. Obviously, $C_1 = 0$ when $\pi = 0$ or 1, and C_3 is independent from π. ■

When $\pi = 0$ or 1, we can explicitly solve Equation (8), because $C_1 = 0$:

$$B(t) = \frac{1}{2}\left(B_0 - \frac{C_3}{\sqrt{C_2}}\right)\exp(\sqrt{C_2}t) + \frac{1}{2}\left(B_0 + \frac{C_3}{\sqrt{C_2}}\right)\exp(-\sqrt{C_2}t). \tag{9}$$

Now we fix the fire allocation only during some time fraction. $\pi_{\tau\to\tau'}$ denotes the fixed allocation during time $t = \tau$ to τ'.

Lemma 2.2: *For any time $\tau_1 \le \tau_2 \le \tau_3$, $\pi_{\tau_1\to\tau_2} = \pi_{\tau_2\to\tau_3}$ holds.*

Proof: Lemma 2.1 shows that $\pi = 0$ or 1 maximizes the blue force at any time. We have to decide what is the optimal fire allocation π either 0 or 1. As we mentioned earlier in proof of Lemma 2.1, C_2 should be maximized. So the optimal fire allocation π^* is

$$\pi^* = \begin{cases} 1 & \text{if } C_2|_{\pi=0} < C_2|_{\pi=1}, \\ 0 & \text{if } C_2|_{\pi=0} > C_2|_{\pi=1}, \end{cases} \tag{10}$$

where

$$C_2|_{\pi=0} = \frac{\beta_N(\alpha_c - \alpha_d)A}{N}, \tag{11}$$

$$C_2|_{\pi=1} = \beta_A\alpha_c. \tag{12}$$

If $C_2|_{\pi=0} < C_2|_{\pi=1}$ holds and we choose the fire allocation π equals to 1, that means we are shooting A first rather than N. Since $C_2|_{\pi=0}$ also decreases when A decreases, the relationship between $C_2|_{\pi=0}$ and $C_2|_{\pi=1}$ does not changes. So if we choose to shoot A first, then fire allocation does not changes.

Similarly, if $C_2|_{\pi=0} > C_2|_{\pi=1}$ holds and we choose the fire allocation π equals to 0 which is shooting N first. $C_2|_{\pi=0}$ increases N decreases, the relationship between $C_2|_{\pi=0}$ and $C_2|_{\pi=1}$ does not changes and the fire allocation also does not changes. ■

Since this is a combat modelling, each forces cannot be negative. Like the original Lanchester equation, this differential equation also does not have such boundary. So Lemmas 2.1 and 2.2 holds when B, A, and N is positive. If either A or N becomes 0, then the fire allocation becomes meaningless because there are only one type of enemy left. Therefore, we have to consider more after either A or N becomes 0 and choose the optimal fire allocation strategies.

Lemma 2.3: *There are two optimal fire allocation strategies.* (i) *Shoot only A ($\pi = 1$) until A is eliminated if $((\alpha_c - \alpha_d)/\beta_A)N_0 > ((\alpha_c + \alpha_d)/\beta_N)A_0$,* (ii) *Shoot only N ($\pi = 0$) until N is eliminated if otherwise.*

Proof: See Appendix. ■

3. The optimal fire allocation strategy via DPS

We suggest some useful properties when $f_\alpha(N)$ is a linear function of N in Section 2. If $f_\alpha(N)$ is nonlinear function of N, it is hard to derive useful properties about fire allocation like the linear case. If the fire allocation $f_\alpha(N)$ is a constant α_N until time t, we can derive an explicit solution for differential equations:

$$B(t) = B_0 \cosh(\sqrt{\pi \alpha_N \beta_A} t) - A_0 \sqrt{\frac{\alpha_N}{\pi \beta_A}} \sinh(\sqrt{\pi \alpha_N \beta_A} t), \tag{13}$$

$$A(t) = A_0 \cosh(\sqrt{\pi \alpha_N \beta_A} t) - B_0 \sqrt{\frac{\pi \beta_A}{\alpha_N}} \sinh(\sqrt{\pi \alpha_N \beta_A} t), \tag{14}$$

$$\begin{aligned} N(t) = N_0 - A_0 \frac{(1-\pi)\beta_N}{\pi \beta_A} (\cosh(\sqrt{\pi \alpha_N \beta_A} t) - 1) \\ - B_0 \frac{(1-\pi)\beta_N}{\sqrt{\pi \alpha_N \beta_A}} \sinh(\sqrt{\pi \alpha_N \beta_A} t). \end{aligned} \tag{15}$$

So if we assume α_N is calculated depends on current state $N(t)$ and stays constant in small time fraction dt, then we can change Equations (13)–(15) to a state-dependent Markovian form. $T_{3\times 3}(\pi, N(t))$ denotes the transition matrix depends on π and current state $N(t)$ with the size 3×3:

$$\begin{bmatrix} B(t+dt) \\ A(t+dt) \\ N(t+dt) \end{bmatrix} = T_{3\times 3}(\pi, N(t)) \begin{bmatrix} B(t) \\ A(t) \\ N(t) \end{bmatrix}$$

$$T_{3\times 3}(\pi, N(t)) = \begin{bmatrix} \cosh(h) & -\sqrt{\frac{\alpha_N}{\pi \beta_A}} \sinh(h) & 0 \\ -\sqrt{\frac{\pi \beta_A}{\alpha_N}} \sinh(h) & \cosh(h) & 0 \\ -\frac{(1-\pi)\beta_N}{\sqrt{\pi \alpha_N \beta_A}} \sinh(h) & \frac{(1-\pi)\beta_N}{\pi \beta_A} (\cosh(h) - 1) & 1 \end{bmatrix}, \tag{16}$$

where $h = \sqrt{\pi \alpha_N \beta_A} dt$.

The fire allocation π can be interpreted as an action at each state $[B(t), A(t), N(t)]'$. The problem seems typical case for Markov Decision Process, but since the state are continuous and the transition matrix is depends on the current state, it is hard to get an optimal action with a backward induction. Reinforcement learning is powerful tool to solve such problem with forward calculation. We use the DPS method [16] which is useful to get the policy that maximizes the expected return.

3.1. DPS by gradient ascent with logit-normal policy model

This method finds the policy model that maximizes the expected return by gradient ascent. Let $\pi(a \mid \mathbf{s}, \theta)$ is the policy parameterized by θ which is the conditional probability density of a in state $\mathbf{s}$. And h is the scenario of length T then, the expected return J for policy parameter θ is defined as

$$J(\theta) = \mathbb{E}_{p(h \mid \theta)} \left[\sum_{t=1}^{T-1} r(\mathbf{s}_t, a_t, \mathbf{s}_{t+1}) \right] = \int p(h \mid \theta) R(h) \, \mathrm{d}h, \tag{17}$$

$$p(h \mid \theta) = p(\mathbf{s}_1) \prod_{t=1}^{T-1} p(\mathbf{s}_{t+1} \mid \mathbf{s}_t, a_t) \pi(a_t \mid \mathbf{s}_t, \theta). \tag{18}$$

The purpose of this method is to the the optimal policy parameter θ^* maximizes the expected return $J(\theta)$. Gradient ascent is a good algorithm to find such parameter.

$$\theta \leftarrow \theta + \varepsilon \nabla_{\theta} J(\theta) \tag{19}$$

$\nabla_\theta J(\theta)$ denotes the gradient of expected return and expressed as

$$\nabla_\theta J(\theta) = \mathbb{E}_{p(h\,|\,\theta)}\left[\sum_{t=1}^{T-1}\nabla_\theta \log \pi(a_t\,|\,\mathbf{s}_t,\theta)R(h)\right] \tag{20}$$

and since $p(h\,|\,\theta)$ is unknown, we can rewrite using the empirical average as

$$\nabla_\theta J(\theta) = \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T-1}\nabla_\theta \log \pi(a_{t,n}\,|\,\mathbf{s}_{t,n},\theta)R(h_n), \tag{21}$$

where h_n is the scenario for nth experiment.

For choosing the policy model $\pi(a\,|\,s,\theta)$, the Gaussian policy model is a popular choice. But, for our case, an action should be bounded between 0 and 1, so Gaussian policy model is not proper model. So we suggest a *logit-normal policy model* which is

$$\pi(a\,|\,\mathbf{s},\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{a(1-a)}\exp\left(-\frac{(\text{logit}(a)-\mu^{\mathrm{T}}\phi(s))^2}{2\sigma^2}\right). \tag{22}$$

The policy gradients are

$$\nabla_\mu \pi(a\,|\,\mathbf{s},\mu,\sigma) = \frac{\text{logit}(a)-\mu^{\mathrm{T}}\phi(s)}{\sigma^2}\phi(s) \tag{23}$$

$$\nabla_\sigma \pi(a\,|\,\mathbf{s},\mu,\sigma) = \frac{(\text{logit}(a)-\mu^{\mathrm{T}}\phi(s))^2-\sigma^2}{\sigma^3}, \tag{24}$$

where $\phi(s)$ is the basis function of $\mathbf{s}$. In this research, we use $\phi(s) = [1\; s^2] = [1\; B^2\; A^2\; N^2\; BA\; BN\; AN]$. We are dealing with the interactive situations between B, A and N, therefore we use the square terms of each states and the interaction terms between each states.

3.2. Numerical experiments for the linear $f_\alpha(N)$

We prove the optimal fire allocation for the linear case $f_\alpha(N) = \alpha_d + (\alpha_c - \alpha_d)(N/N_0)$ in Section 2. We conduct three numerical experiments using the DPS algorithm and compare to the theoretical optimal values. Table 1 shows the initial parameter settings and the theoretical optimal value for each experiment settings, and also the value obtained from the DPS algorithm. In this context, the value means ‘the alive blue forces at the end of the battle’.

Figures 2–4 show the changes in the each forces, and the fire allocation over time for each experiment settings. Commonly, the algorithm does not know how to allocate the fire to maximize the blue force before learning. However, after hundreds iterations for learning, they learn almost exactly how to allocate the fire to maximize the blue force at the end of battle. As we wrote in Table 1, there are small differences between the theoretical values and obtained value from the algorithm. We use the term ‘Learning curve’ for changes in sample average value of blue force at the end of battle over time in figures. Learning curve in Figures 3 and 4 are the common shape which shows gradually increasing, but learning curve in Figure 2 is somehow different. It moves like a step function at the

Table 1. The experiment settings for the linear $f_\alpha(N)$.

Experiments	B_0	A_0	N_0	α_c	α_d	β_A	β_N	Optimal value	DPS value
1	50	60	10	0.2	0.07	0.2	0.15	11.0136	10.8543
2	50	50	20	0.2	0.04	0.2	0.15	19.2483	18.5618
3	70	80	20	0.2	0.1	0.3	0.2	24.6512	24.0241

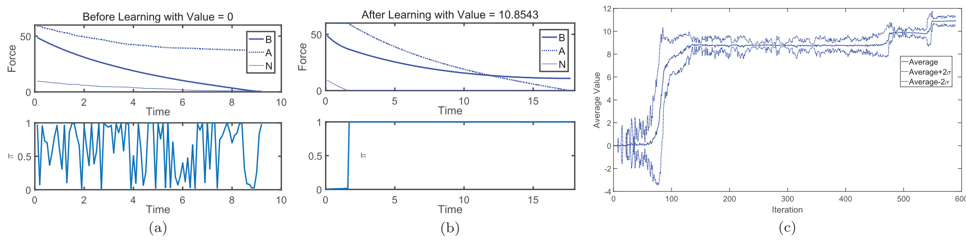

Figure 2. Changes of the each forces, the fire allocation π_t and the sample average value on linear experiment setting 1 (DPS value after 590 iterations = 10.8543). (a) Before learning, (b) after learning and (c) learning curve.

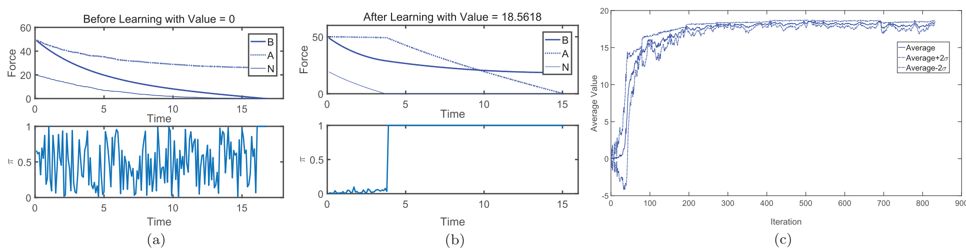

Figure 3. Changes of the each forces, the fire allocation π_t and the sample average value on linear experiment setting 2 (DPS value after 820 iterations = 18.5618). (a) Before learning, (b) after learning and (c) learning curve.

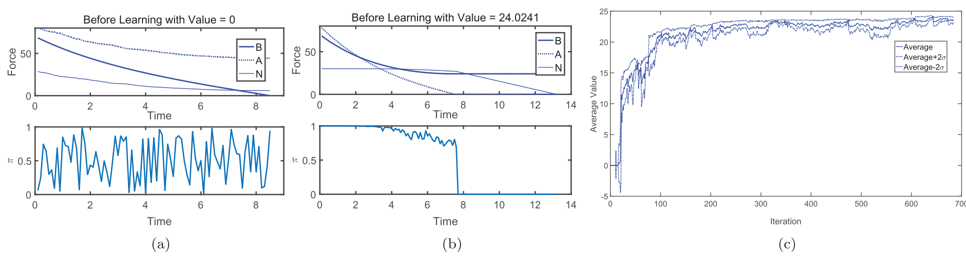

Figure 4. Changes of the each forces, the fire allocation π_t and the sample average value on linear experiment setting 2 (DPS value after 820 iterations = 24.0241). (a) Before learning, (b) after learning and (c) learning curve.

end of iterations. Since DPS algorithm is using the gradient ascent algorithm to update the policy parameter, it will easily converges to the local maximum value. We can interpret that the algorithm had been converged to the local maximum and it searched for the better value for long iterations and finally, it found the better values and moves to that point. This step-like behaviour depends on hyper-parameters in the learning algorithm which controls the exploitation and the exploration.

3.3. Numerical experiments for the nonlinear $f_\alpha(N)$

As we see in a previous subsection, the DPS algorithm helps to reach to theoretical optimal value for the linear attrition rate function. In this subsection, we conduct 2 experiments with the nonlinear attrition rate functions $f_\alpha(N)$. We do not know the theoretical optimal values like the linear case, but we can deduce from the linear case that this algorithm will reach near to the optimal value and optimal policy. Table 2 shows the 2 experiment setting for the nonlinear case. We fix the initial parameters same as experiment setting 2 in the linear case and change only the attrition rate function $f_\alpha(N)$.

The values obtained from the DPS changes dramatically compare to linear case. The optimal value of experiment setting 2 in linear case was 18.5618. But after we change the attrition rate function $f_\alpha(N)$

Table 2. Experiment settings for the nonlinear $f_\alpha(N)$.

Experiments	B_0	A_0	N_0	α_c	α_d	β_A	β_N	DPS value
Common setting	50	50	20	0.2	0.04	0.2	0.15	
1 (Logarithmic)	$f_\alpha(N) = \alpha_d + (\alpha_c - \alpha_d)\log(9\frac{N}{N_0} + 1)$							0
2 (Exponential)	$f_\alpha(N) = \alpha_d \exp(\ln(\frac{\alpha_c}{\alpha_d})(\frac{N}{N_0})^2)$							30.3291

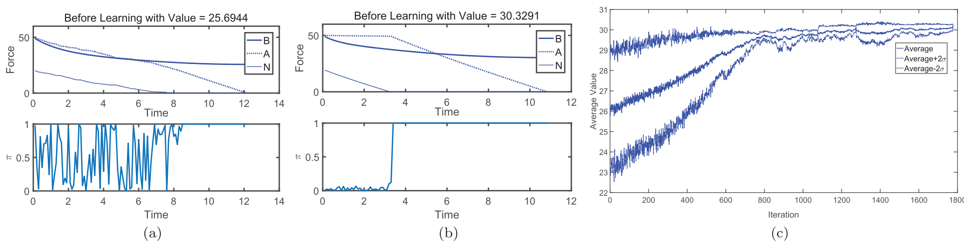

Figure 5. Changes of the each forces, the fire allocation π_t and the sample average value on nonlinear experiment setting 2 (DPS value after 1780 iterations = 30.3291). (a) Before learning, (b) after learning and (c) learning curve.

to logarithmic and exponential function, the value changes to 0 and 30.3291, so it will be important to design the attrition rate function well to predict and make the optimal decision. Figure 5 shows the same plots as previous for the exponential case. Since the logarithmic case always have value 0 for any allocation, so it does not have the optimal fire allocation like any others. We might change the definition of 'value' to '$B_T - (A_T + N_T)$; The difference of the forces at the end of the battle'. For new definition, the algorithm might lead to minimize the difference and try to find at least good strategy that maximize the casualties. This will be studied in further research.

4. Conclusion

In this paper, we suggest the new differential equation model that includes the framework of NCW and force multiplier. By solving this equation, we proved some useful properties for the special case; the linear attrition rate function case. Main contribution is that we adopt the reinforcement learning technique called the DPS and modify the model to solve the optimal fire allocation problem. We show that this algorithm works well in the linear case which has known optimal value, and conduct 2 more experiments for nonlinear case.

Our goal has been to introduce the control community about the application of military field. One next step is to conduct more experiments with the complex model, for example, a model that includes the defense assets or new attrition property like fire range. This will bring out new behaviours and explanations that can advise the commander to order actively. Another way will be considering stochastic effects in the model. In our model, we exclude the stochastic effects to clarify the situation and derive some analytical properties. We strongly expect that the stochastic term will affects the optimal fire allocation a lot and totally changes the outcome of the battle. Of course, since the direct policy algorithm is sensitive to policy model and the basis function of s, $\phi(s)$, it will be great work to stabilize the algorithm and make robust.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Defense Acquisition Program Administration and Agency for Defense Development under the contract UD140022PD, Korea.

References

[1] Lanchester FW. Aircraft in warfare: the dawn of the fourth arm. London: Constable Limited; 1916.
[2] Bracken J, Kress M, Rosenthal RE, editors. Warfare modeling. Alexandria, VA: MORS; 1995.
[3] Tolk A. Modeling effects. In: Engineering principles of combat modeling and distributed simulation. Hoboken, NJ: John Wiley & Sons; 2012 p. 145–170.
[4] Powell DS. Understanding force multipliers: the key to optimizing force capabilities in peacetime contingency operations. Army Command and General Staff Coll Fort Leavenworth KS school of Advanced Military Studies; 1990.
[5] Smith E. Effects based operations. In: Applying network-centric warfare in peace, crisis and war. Washington, DC: DoD CCRP; 2002.
[6] Smith CR, Colonel L. Network centric warfare, command, and the nature of war. In: Land warfare studies centre; 2010.
[7] Tunnell HD. Network-centric warfare and the data-information-knowledge-wisdom hierarchy. Military Rev. 2014;94(3):43–50.
[8] Tunnell HD. The US Army and network-centric warfare a thematic analysis of the literature. In: Military communications conference, MILCOM 2015–2015 IEEE. IEEE; 2015.
[9] Beck J, Moon T, Van AC, et al. Knowledge superiority parameter-a metric for network centric warfare (NCW). In: Defence science and technology organisation (Australia) defence systems analysis div. 2003.
[10] Schramm HC. Lanchester models with discontinuities: an application to networked forces. Military Oper Res. 2012;17:59–68.
[11] Schramm HC, Gaver DP. Lanchester for cyber: the mixed epidemic-combat model. Nav Res Logist. 2013;60(7):599–605.
[12] Ahuja RK, Kumar A, Jha KC, et al. Exact and heuristic algorithms for the weapon-target assignment problem. Oper Res. 2007;55(6):1136–1146.
[13] Azak M, Bayrak AE. A new approach for threat evaluation and weapon assignment problem, hybrid learning with multi-agent coordination. In: Computer and information sciences, 2008. ISCIS'08. 23rd international symposium on. IEEE; 2008.
[14] Lee Z-J, Su S-F, Lee C-Y. Efficiently solving general weapon-target assignment problem by genetic algorithms with greedy eugenics. IEEE Trans Syst Man Cybern B Cybern. 2003;33(1):113–121.
[15] Paradis S, Benaskeur A, Oxenham M, et al. Threat evaluation and weapons allocation in network-centric warfare. In: 2005 7th international conference on information fusion. Vol. 2. IEEE; 2005.
[16] Sugiyama M. Statistical reinforcement learning: modern machine learning approaches. Boca Raton, FL: CRC Press; 2015.

Appendix. Proof of Lemma 2.3

If $\pi = 0$ at first, then Equations (4)–(6) become first-order linear ODE which is,

$$\frac{\mathrm{d}B}{\mathrm{d}t} = -\left(\alpha_d + (\alpha_c - \alpha_d)\frac{N}{N_0}\right)A_0, \tag{A1}$$

$$\frac{\mathrm{d}N}{\mathrm{d}t} = -\beta_N B. \tag{A2}$$

From this ODE, we can calculate $B(t_{N=0})$ when $N(t_{N=0})$ becomes 0,

$$B(t_{N=0})^2 = B_0^2 - \frac{(\alpha_c + \alpha_d)}{\beta_N}A_0N_0. \tag{A3}$$

If N is eliminated before B is eliminated, now B concentrates the fire to A then,

$$\frac{\mathrm{d}B}{\mathrm{d}t} = -\alpha_d A, \tag{A4}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -\beta_A B. \tag{A5}$$

Table A1. The optimal fire allocation π^*.

	$B^2_{0\to1} < 0$	$B^2_{0\to1} > 0$
$B^2_{1\to0} < 0$	Blue loses	Shoot N first ($\pi^* = 1$)
$B^2_{1\to0} > 0$	Shoot A first ($\pi^* = 1$)	$B_{1\to0} > B_{0\to1}, \pi^* = 1.$ $B_{1\to0} < B_{0\to1}, \pi^* = 0.$

We can calculate $B_{0\to1}$ when A eliminates,

$$B^2_{0\to1} = B^2_0 - \frac{(\alpha_c + \alpha_d)}{\beta_N} A_0 N_0 - \frac{\alpha_d}{\beta_A} A^2_0. \tag{A6}$$

On the other hand if $\pi = 1$ at first, then Equations (4)–(6) also become first-order linear ODE which is,

$$\frac{\mathrm{d}B}{\mathrm{d}t} = -\alpha_c A, \tag{A7}$$

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -\beta_A B. \tag{A8}$$

From this ODE, we can calculate $B(t_{A=0})$ when $A(t_{A=0})$ becomes 0,

$$B(t_{A=0})^2 = B^2_0 - \frac{\alpha_c}{\beta_A} A^2_0. \tag{A9}$$

Since A is eliminated, B concentrates the fire to N. But Equation (A7) becomes 0 and it is obvious that $B_{1\to0}$ when N eliminates is same as $B(t_{A=0})$.

$$B^2_{1\to0} = B(t_{A=0})^2 = B^2_0 - \frac{\alpha_c}{\beta_A} A^2_0. \tag{A10}$$

Then we can make the optimal fire allocation chart as follows (Table A1).

Simplify the relationship between $B_{1\to0}$ and $B_{0\to1}$,

$$B_{1\to0} > B_{0\to1}$$

$$\frac{(\alpha_c + \alpha_d)}{\beta_N} N_0 > \frac{(\alpha_c - \alpha_d)}{\beta_A} A_0. \tag{A11}$$

(i) Shoot only A ($\pi = 1$) until A is eliminated if $((\alpha_c - \alpha_d)/\beta_A)N_0 > ((\alpha_c + \alpha_d)/\beta_N)A_0$, (ii) Shoot only N ($\pi = 0$) until N is eliminated if otherwise.