The Effect of the Number of Hidden Layers on The Performance of Deep Q-Network for Traveling Salesman Problem

Benzfica Hanif [a,1], Aisyah Larasati [a,2,*], Rudi Nurdiansyah [a,3], Trung Le [b,4]

[a] Department of Mechanical and Industrial Engineering, Faculty of Engineering, Universitas Negeri Malang
Jl. Semarang no. 5, Malang 65145, Indonesia
[b] Department of Industrial and Management Systems Engineering, University of South Florida
4202 E Fowler Ave, Tampa-FL 33620, USA
[1] benzfica@gmail.com; [2] aisyah.larasati.ft@um.ac.id*; [3] rudi.nurdiansyah.ft@um.ac.id; [4] tqle@usf.edu
* corresponding author

ARTICLE INFO

ABSTRACT

The Traveling Salesman Problem (TSP) effectively represents the complex distribution issues encountered by couriers, who must carefully plan a route that includes all customer addresses while minimizing the distance traveled. As the magnitude of deliveries and the range of destinations expand, the courier's responsibility becomes progressively challenging. In this particular context, the objective of our research is to expand the existing knowledge and explore the complete capabilities of Deep Q-Network (DQN) models in order to achieve the most efficient route determination. This endeavor can potentially bring about significant changes in the courier and delivery service sector. The foundation of our unique methodology relies on an empirical inquiry, utilizing a comprehensive dataset including 178 observations obtained from motorcycle-based package delivery agents. Our research is carefully planned and executed using a comprehensive factorial experimental design. This design incorporates three crucial factors: the number of hidden layers, episodes, and epochs. The hidden layer parameter is set to a singular level, while the episode parameter is configured to explore five levels, and the epoch parameter is designed to travel four levels. The evaluation of our DQN models' performance is conducted utilizing the MSE metric as a measure. This assessment is carried out at every iterative cycle, ensuring thorough scrutiny. The central focus of our research centers on the intricate connection between episodes and epochs, and their influence on MSE. The findings of our study reveal that the association between episodes, epochs, and errors is not statistically significant although different level of episodes and epochs produces slightly different level of error.

I. Introduction

Consumer behavior has been changing due to the desire for fast, safe, and efficient fulfillment of their needs, driven by the digital era. Meeting these consumer expectations requires the intervention of delivery services. During the delivery process, problems often arise in route determination. These problems occur because couriers rely on their knowledge to deliver items to customer addresses, which can lead to further complications when dealing with larger quantities of items and diverse customer addresses. The impacts of such issues include wasted delivery time, increased operational costs, and unmet delivery targets.

The Traveling Salesman Problem (TSP) involves a salesman and a set of N cities. This issue aims for the salesman to visit each city exactly once while covering the shortest possible total tour distance [1]. The solution to the TSP has been widely addressed using optimization algorithms to optimize the resources available in the distribution process. The essence of TSP is to find the shortest route involving a number of points, including returning to the starting point. As a complex mathematical problem, various heuristic methods have been developed over time to find approximate solutions [2]. The research by [3] utilized the Harris Hawk Optimization algorithm, which employs random-key

encoding to generate a tour. The research conducted by [4] uses the New Ant Colony Optimization for solving TSP, achieving high accuracy and fast computational times. The research conducted by [5] used ant colony optimization to determine TSP routes and showed that the execution time of ant colony optimization was faster in obtaining results than the execution time of exact methods.

With the growing popularity of Machine Learning (ML) and deep learning, numerous research teams have embraced Mg combinatorial optimization challenges, including the widely recognized TSP. New models and architectures for solving TSP have been progressively created using deep (reinforcement) learning, improving the performance [6]. The ML algorithm can be applied in the Deep Q-Network (DQN). Reference [7], in their study, utilized the DQN algorithm to address shipping and route issues for autonomous robots. Research conducted by [8] used DQN to solve the truck routing problem between terminals to minimize the total cost incurred. However, in recent years, machine learning advancements have been utilized to solve TSP-related problems. The deep neural network method provides significantly more robust capabilities in pattern recognition and feature representation. These algorithms can provide solutions based on performance comparisons in determining the best routes.

The TSP has a long history and finds numerous real-world applications. It aims to discover the most efficient route that includes each city exactly once and ends at the starting city [9]. Equation (1) is the objective function of TSP, denoted by *Z*, which aims to minimize the total distance traveled to determine the route. The formulation of TSP modeling represents the distance traveled from point *i* to point *j* as *Cij*. The decision variable, denoted as *Xij*, represents whether there is a travel from point *i* to point *j*.

$$Minimum\ Z = \sum_{j=1}^{n}\sum_{i=1}^{n} Cij\ Xij \tag{1}$$

Having constraint limitations, (2) and (3) constraints ensure that the selected route arrives and leaves the destination once. Travel value from point *i* to point *j* as in (4).

$$\sum_{j=1}^{n} Xij = (j = 1, 2,\ 3, \dots, N) \tag{2}$$

$$\sum_{i=1}^{n} Xij = (i = 1, 2,\ 3, \dots, N) \tag{3}$$

$$Xij = 0\ or\ 1 \tag{4}$$

Reference [10] researched the TSP using Genetic algorithms with performance assessment of the model based on the total distance traveled. In the research conducted by [11], algorithms were compared to solve the TSP to obtain an optimal route that visits each destination once and returns to the starting point. This study applies constraints in determining the optimal route based on the loss value of the model. The performance evaluation of the DQN algorithm in determining the optimal route is based on the loss value. The loss function typically employed is the mean square error (MSE). MSE represents the expected value of the squared difference between the estimated parameter and the true parameter. A lower MSE value indicates greater accuracy in describing the experimental data [12].

DQN is a multi-layered neural network with specific states to generate action values [13]. The essential components of DQN are the target network and experience replay. DQN is a combined algorithm of Q-learning and deep neural networks to consider the value of the state-action function. The advantage of using DQN is its ability to represent observation results in high-dimensional states and calculate Q-function values using a deep neural network. The target parameter used in the DQN algorithm is defined as in (5).

$$Y_t^{DQN} = R_{t+1} + \gamma \max Q\ (S_{t+1}, a\ ;\ \theta_k^-) \tag{5}$$

The Q-values are updated using the parameters present in the neural network. The updated neural network values are obtained through reverse transfer from the loss function. The loss function of DQN is defined as the squared error between the target Q-value and the estimated Q-value, as in (6).

$$Loss^{DQN} = [R_{t+1} + \gamma max Q(S_{t+1}, a; \theta_k^-) - Q)s; a; \theta)]^2 \tag{6}$$

Figure 1 illustrates the DQN training process [14]. The DQN training process is an improvement over the Q-learning algorithm that addresses the instability issues in representing the function of the non-linear network. DQN uses experience replay to process the transfer samples. At each time step *t,* the transfer samples obtained by the agent interacting with the environment are stored in the replay buffer unit. During the training process, a batch of transfer samples is randomly selected, and the stochastic gradient descent algorithm is used to update the network parameters θ.
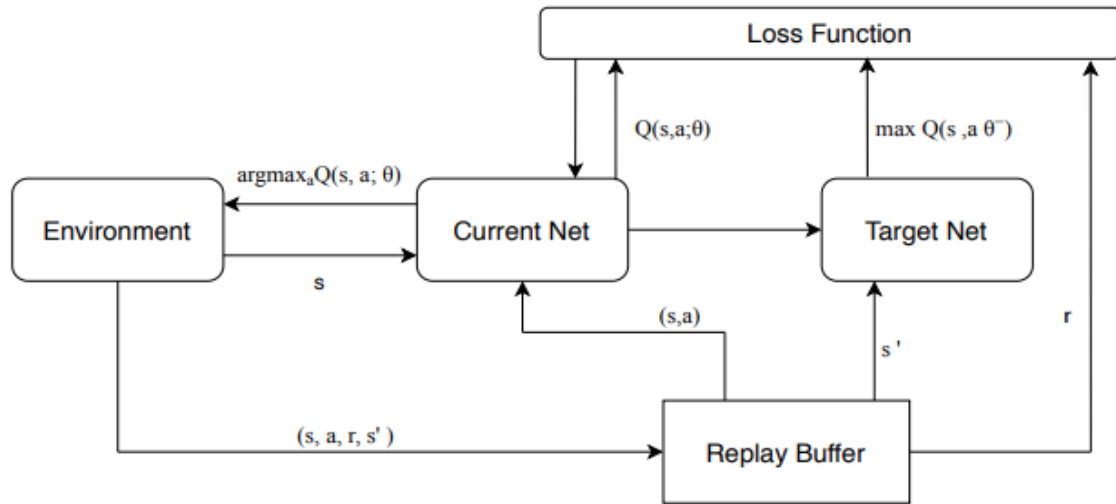

Fig. 1. Overview of the DQN training

Within artificial intelligence and optimization, there is a significant research emphasis on enhancing the effectiveness of DQNs in tackling complex combinatorial issues such as the TSP. The primary aim of this study is to examine the impact of manipulating the number of hidden layers in a DQN framework on its efficacy in addressing the TSP. The objective of this research endeavor is to discover novel insights and advancements that can significantly improve the effectiveness and precision of DQN-based solutions for this well-established topic.

II. Methods

The study procedure, as seen in Figure 2, undertakes a thorough exploration to tackle the intricate task of identifying the most favorable pathway. Fundamentally, this undertaking is supported by the powerful Deep Q-Network (DQN) algorithms, which have the potential to transform the field of route optimization significantly. The research begins by carefully selecting a comprehensive compilation of literature reviews, specifically chosen to cover essential findings regarding the challenges of route determination and references explaining the deep learning techniques utilized in this study.

Once the core knowledge base has been established, the succeeding step thoroughly examines real-world challenges in determining routes and the various strategies to address them. This comprehensive inquiry forms the foundation for making well-informed decisions, allowing the research team to develop a methodology grounded in empirical evidence and relevant in practical terms.

The data acquisition phase is important, as it involves carefully collecting a comprehensive dataset that includes crucial features such as order ID, origin, postal codes, addresses, and geographical coordinates (latitude and longitude). After the data-gathering procedure, a rigorous preprocessing protocol is implemented to examine the dataset carefully. This approach aims to remove duplicate entries and extract the fundamental attribute variables that will serve as the foundation for constructing the DQN algorithm model in the upcoming steps.

```
┌──────────┐
│  Start   │
└──────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│           Literature Review              │
└─────────────────────────────────────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│       Identification Problem Statement   │
└─────────────────────────────────────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│        Collecting Data (Observation)     │
│ Order ID, Origin, Postal, Address,       │
│         Latitude, Longitude              │
└─────────────────────────────────────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│           Data Preprocessing             │
│        Cleaning and Reduction Data       │
└─────────────────────────────────────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│        Modeling Deep Q-Network           │
└─────────────────────────────────────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│           Setting Parameter              │
└─────────────────────────────────────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│             Evaluation                   │
└─────────────────────────────────────────┘
      │
      ▼
┌─────────────────────────────────────────┐
│             Conclusion                   │
└─────────────────────────────────────────┘
      │
      ▼
┌──────────┐
│  Finish  │
└──────────┘
```
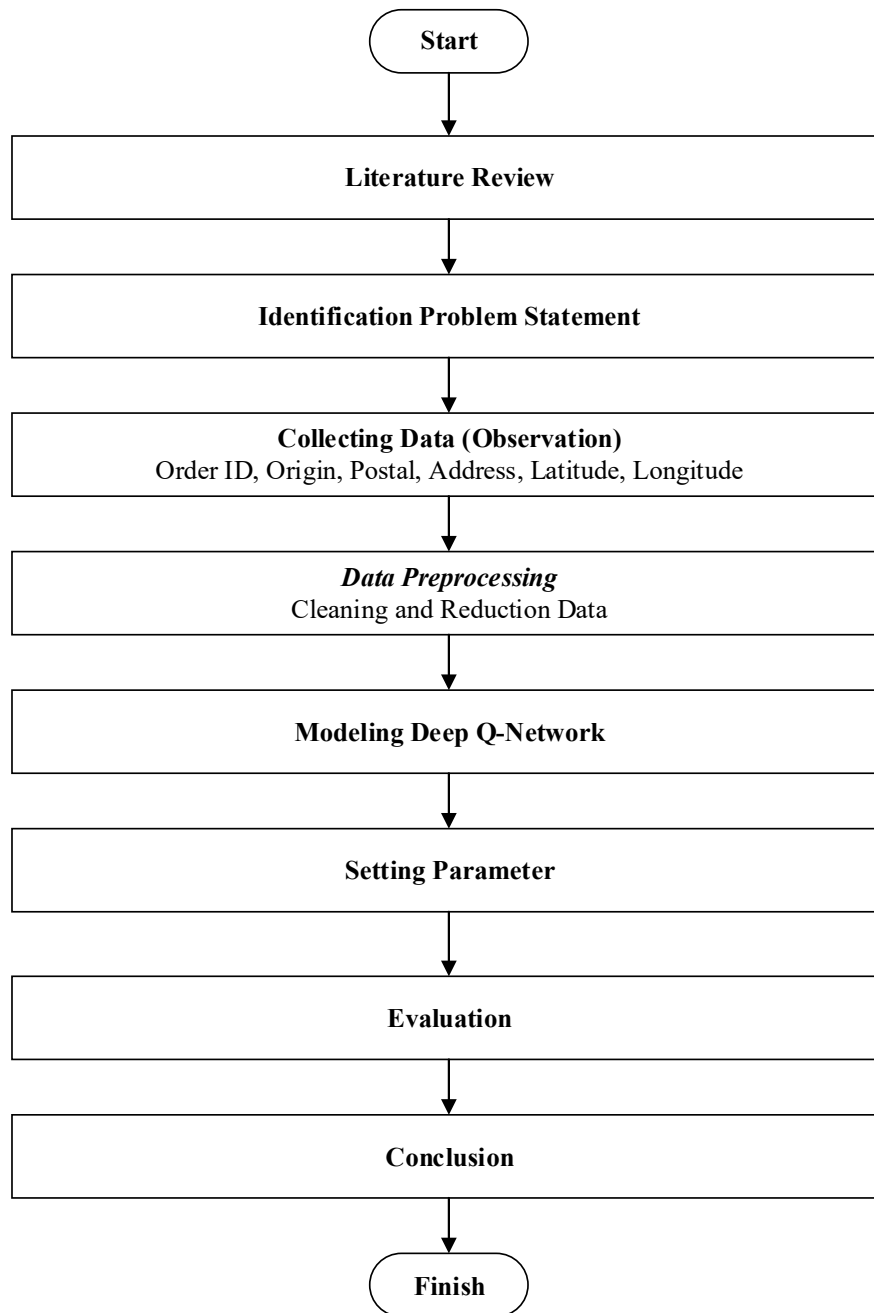
Fig. 2. Research flow

The culmination of the research process is conducting a rigorous three-factorial experiment that examines the crucial factors of hidden layers, episode configuration, and epoch settings. This study aims to systematically investigate the most practical combination of these parameters, leading to the optimization of the DQN model and paving the way for significant advancements in route optimization approaches. Within the complex interplay of theory and practice, this study provides significant advancements in deep learning-driven route determination, offering novel approaches to address urgent practical obstacles.

The dataset consists of 178 data points collected on a single day of the delivery procedure. The data is acquired during the observation process and documented within the application possessed by each courier. Data selection based on attribute variables and the subsequent cleaning procedure are necessary steps to ensure the data's integrity by removing duplicates and incomplete entries. The variables to be utilized are location, latitude, and longitude.

The study incorporates several criteria, namely the quantity of hidden layers, episodes, and epochs. The augmentation of hidden layers has enhanced precision; nevertheless, it necessitates a lengthier training period and heightens the potential for overfitting [14]. The quantity of epochs denotes a comprehensive iteration within the machine learning process, during which the model acquires knowledge from the entirety of the training dataset. In the context of neural network methodologies, the iterative nature of learning processes plays a crucial role in achieving the convergence of weight values. Given the lack of knowledge regarding the ideal number of episodes and epochs, it becomes imperative to conduct experiments using various values to attain the lowest possible loss.

Consequently, this research investigates the impact of manipulating the parameters of hidden layers, episodes, and epochs. The number of hidden layers to be tested will be limited to one. The episode will be conducted for 50, 100, 150, 200, and 250 iterations. Additionally, the epoch will be set to 1, 50, 100, and 500. This experiment yielded 20 unique combinations of hidden layers, episodes, and epochs.

The construction of the Deep Q-Network model commences with establishing the environment, wherein the initial state is determined by referencing the historical data about deliveries. The initial location of the delivery is situated at Jl Raya Sawojajar, namely at Ruko WOW No.11A. The courier, functioning as the agent, will traverse the state space region, encompassing the latitude and longitude coordinates of addresses within the given environment. The mobility of the agent is constrained to the provided location data. The agent will continue its movement till it reaches the ultimate delivery destination, located explicitly at Jl Danau Towuti Raya Blok G4 A17.

The Deep Q-Network configuration is produced using the Keras toolkit. The model is constructed using Dense or completely connected layers, each consisting of 32 neurons, and employing the Rectified Linear Unit (ReLU) activation function. The optimization function, Adam, is a widely used algorithm in machine learning that is designed to update the parameters of a model efficiently. It combines the benefits of the Adaptive Moment Estimation (Adam) and Root Mean Square Propagation (RMSProp) algorithms. Adam utilizes adaptive learning rates for each parameter, which are computed based on the first and second moments of the gradients, allowing for effective optimization of the model's parameters.

On the other hand, the loss function, Mean Square Error (MSE), is a commonly employed metric in regression tasks. It measures the average squared difference between the predicted and actual values. MSE is widely used due to its simplicity and ability to penalize more significant errors more heavily. Minimizing the Deep Q-Network can generate various outputs, including the trajectory followed, the overall distance covered, and the computed value of the loss function.

III. Result and Discussion

The research design comprises a comprehensive framework considering three essential criteria: the number of hidden layers, episodes, and the epochs utilized. The enigmatic topography of concealed strata unveils a captivating compromise - an augmentation of these strata has exhibited a distinct inclination to enhance degrees of accuracy. Nevertheless, this approach is accompanied by extended training periods and an increased likelihood of overfitting, as supported by previous research [14]. In Table 1, epochs are a guiding principle for a comprehensive exploration of machine learning. Within the vast expanse of this particular environment, the model experiences a transformative process of acquiring knowledge, integrating valuable insights derived from the entirety of the training dataset. Within the domain of neural network techniques, the iterative nature of these learning processes facilitates the complex process of weight value convergence. However, the indeterminate characteristic of determining the ideal quantity of episodes and epochs motivates us to undertake an empirical investigation, conducting experiments with various numerical values to achieve the elusive objective of minimizing loss.

Consistent with the empirical approach, our research aims to investigate the complex interaction between hidden layers, events, and epochs. The hidden layers' canvas is deliberately limited to a single layer, enabling us to isolate and examine the influence of other variables. In contrast, episodes will be carefully planned and executed, spanning a range of 50, 100, 150, 200, and 250 occurrences.

Concurrently, the epoch parameter will be strategically set to values of 1, 50, 100, and 500. The comprehensive experimental approach employed in this study yields a diverse array of 20 unique combinations, providing a comprehensive understanding of the interplay and impact of these variables on the performance of our model. In this comprehensive investigation, our objective is to decipher the most effective arrangement that facilitates improved accuracy while minimizing the potential drawbacks of overfitting. This endeavor will ultimately lead to the developing of more efficient and precise approaches for determining routes. The loss value obtained from different level parameter is shown in Table 1.

Table 1. Loss value using different level parameter

Parameter		Loss
Episode	Epoch	
50	1	7.005321025848389
50	50	3.2833027944434434e-05
50	100	0.00012886642070952803
50	500	1.189620525110513e-05
100	1	3.721908797160722e-05
100	50	1.953684477484785e-05
100	100	1.5766545402584597e-05
100	500	9.782820598047692e-06
150	1	0.0010764036560431123
150	50	1.7226924683200195e-05
150	100	1.4794331036682706e-05
150	500	0.0002136115072062239
200	1	0.009234399534761906
200	50	1.270834582101088e-05
200	100	1.1694006389006972e-05
200	500	5.954136940999888e-05
250	1	0.0018385164439678192
250	50	4.7960747906472534e-05
250	100	2.5381839805049822e-05
250	500	3.104201823589392e-05

Episode and epoch are two crucial elements that must strategically interact for the DQN model to be constructed. In order to minimize the loss function, a key performance statistic, these parameters work together to create the model's architecture. We use the time-tested Analysis of Variance (ANOVA) approach to determine the specific influence of each parameter on the final loss value. We use ANOVA as our dependable compass to guide us through the challenging environment of parameter influence. The findings of the ANOVA test performed between episode and loss are revealed in Table 2, a gold mine of insights. The computed p-value, which is just 0.438, survives statistical inspection. Even though this value is substantial, careful interpretation is required.

Table 2. ANOVA between episode and loss

	Sum of Squares	df	Mean Square	F	Sig.
Betwwen Groups	9.807	4	2.452	0.999	0.438
Within Groups	36.805	15	2.454		
Total	44.612	19			

Table 3 serves as a guiding light for understanding the complex dynamics in our DQN model, providing the results of our ANOVA study performed between epoch and loss. A significant p-value of 0.416 emerges from this tableau of statistical findings, attesting to the thorough examination of our dataset. However, when this result is compared to the revered confidence level of 0.05, a staunch bulwark of statistical rigor set at the high 95% threshold, the actual significance of this finding becomes clear.

Table 3. ANOVA between episode and loss

	Sum of Squares	df	Mean Square	F	Sig.
Betwwen Groups	7.386	3	2.462	1.004	0.416
Within Groups	39.226	16	2.462		
Total	46.612	19			

The null hypothesis, which maintains no significant epoch-driven influence on the loss value, triumphs in this delicate dance of numbers as the p-value exceeds the confidence standard. This significant resultANOVA between epoch and loss.

Based on the results obtained from the many ANOVA tests, it can be concluded that the parameters of episode and epoch have negligible or no significant impact on the loss value. However, it is essential to note that this discovery is closely linked to the sample size used in our research efforts, highlighting the need for a nuanced comprehension. As expounded upon in the literature by [14], attaining statistical significance becomes challenging when conducted within a restricted sample size. Furthermore, the significance of replications in the study's design, as highlighted by [17], should not be ignored, as they can uncover or disguise specific effects within the data. The significance of degrees of freedom, as highlighted by [18], is another crucial aspect closely connected to sample size and replication data. Although a degree of freedom value of 15 is generally considered acceptable, it is crucial to recognize that contextual limitations may hinder the specific selection of this number in some studies.

The inquiry undertaken by [19] on weather classification, which employed the backpropagation method with different numbers of hidden layers (1, 2, and 3), is a noteworthy reference point when considering past research. The results of their study suggest that modifying these parameters did not result in statistically significant improvements in accuracy values. The significance of hidden layers, which function as intermediates within the neural network, becomes prominent since they are provided with activation functions that enable the transfer and training of data across different layers of the network [20]. The selection of the ideal number of hidden layers is still a subject of debate, as evidenced by other research that has used hidden layers as parameters and obtained different accurate results. Several factors contribute to determining the appropriate number of hidden layers in a neural network. These factors include the complexity of the network design, the number of input and output units, the volume of training samples, the presence of noise in the dataset, and the intricacy of the training process [21].

Additional knowledge can be acquired from the study conducted by [22], wherein artificial neural networks were employed to mimic air pressure resulting from overpressure. The parameter of the epoch was considered in the analysis. Interestingly, their model demonstrated no substantial dependence on the epoch parameter, establishing a connection between epochs and the notion of weight convergence in machine learning algorithms. The complex interconnection between episodes and epochs becomes evident as epochs effectively serve as a higher-level loop that encompasses the episode loop. The absence of a clearly defined deterministic rule for calculating the episode's number is emphasized by the findings of [23].

Nevertheless, it is essential to acknowledge that an unexpected pattern emerged during the experimental procedure: a positive correlation was observed between the increment in the number of episodes and epochs and a noticeable reduction in the loss value. The observed inconsistency between the ANOVA-based statistical analysis, which showed no statistically significant variation in the loss value concerning episode and epoch, adds a level of intricacy to our comprehension. The presence of incongruity in the given context indicates the potential influence of random elements, such as inadequate data for conducting ANOVA testing and the lack of data replication. The complexity of quantifying the exact relationship between parameters and the loss value in the ANOVA framework presents a challenge, highlighting the significance of recognizing the interaction between statistical analysis and real-world data in pursuing comprehensive insights.

Figure 3 and Figure 4 provide aesthetically captivating representations of the research outcomes, revealing intricate patterns in the dynamics of loss values. Figure 3 presents a visual representation of the loss value, illustrating a substantial decline in loss value as the number of episode increases up to 100. However, when the episode is set to 125 and more, the changes in loss value becomes up and down depends on the number of epoch. The visual representation resembles the underlying data patterns, wherein a noticeable decline can be observed in each experimental session using different episode values. Likewise, Figure 4 presents a captivating depiction of the behavior of the loss value, illustrating a clear pattern of decrease with each successive increase in the number of epochs when

the epoch is set less than 100. However, when the epoch is set equal to or more than 100, the decrease becomes up and down depends on the number of episode. The graphical narrative presented in this analysis effectively illustrates a pattern of constant yet dynamic reductions in the loss value. This pattern mirrors the complex interplay between different epochs and their influence on the model's performance.
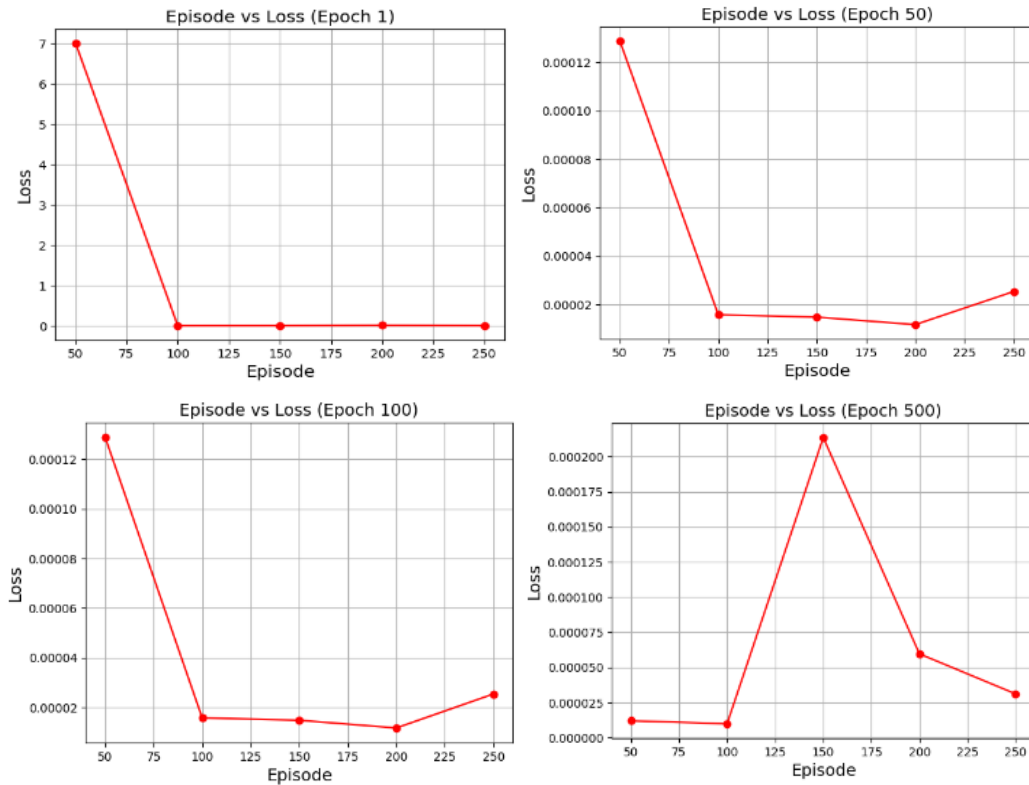

Fig. 3. Episode vs loss

Upon further examination of the modeling process, it becomes apparent that the algorithm can construct a model that accurately reflects the intricacies of the real-world situation being addressed. Upon more profound analysis of Table 1, a notable accomplishment becomes apparent - the painstaking optimization of hyperparameters, carried out for 500 epochs within the scope of 100 episodes, resulting in a meager loss value of 0.000010. The modest size of this figure serves as a strong indication of the algorithm's effectiveness, solidifying the idea that lower loss values indicate the attainment of an ideal model. If we were to represent this notable accomplishment visually, it would depict a captivating depiction of a gradual decrease in loss over 500 carefully planned periods within the span of 100 instances, serving as a vivid demonstration of the algorithm's ability to improve and enhance its performance consistently.

This study provides significant contributions to the domain of route optimisation through the utilisation of DQN models. However, there are certain limitations that can be addressed in future research. These limitations include the potential for expanding the dataset, exploring a broader range of hyperparameters, incorporating data replication techniques, adopting additional evaluation metrics, transitioning towards real-world deployment, and leveraging enhanced computational resources. These enhancements will enhance our comprehension of DQN-based route determination and its pragmatic implementations in the courier and delivery sector.
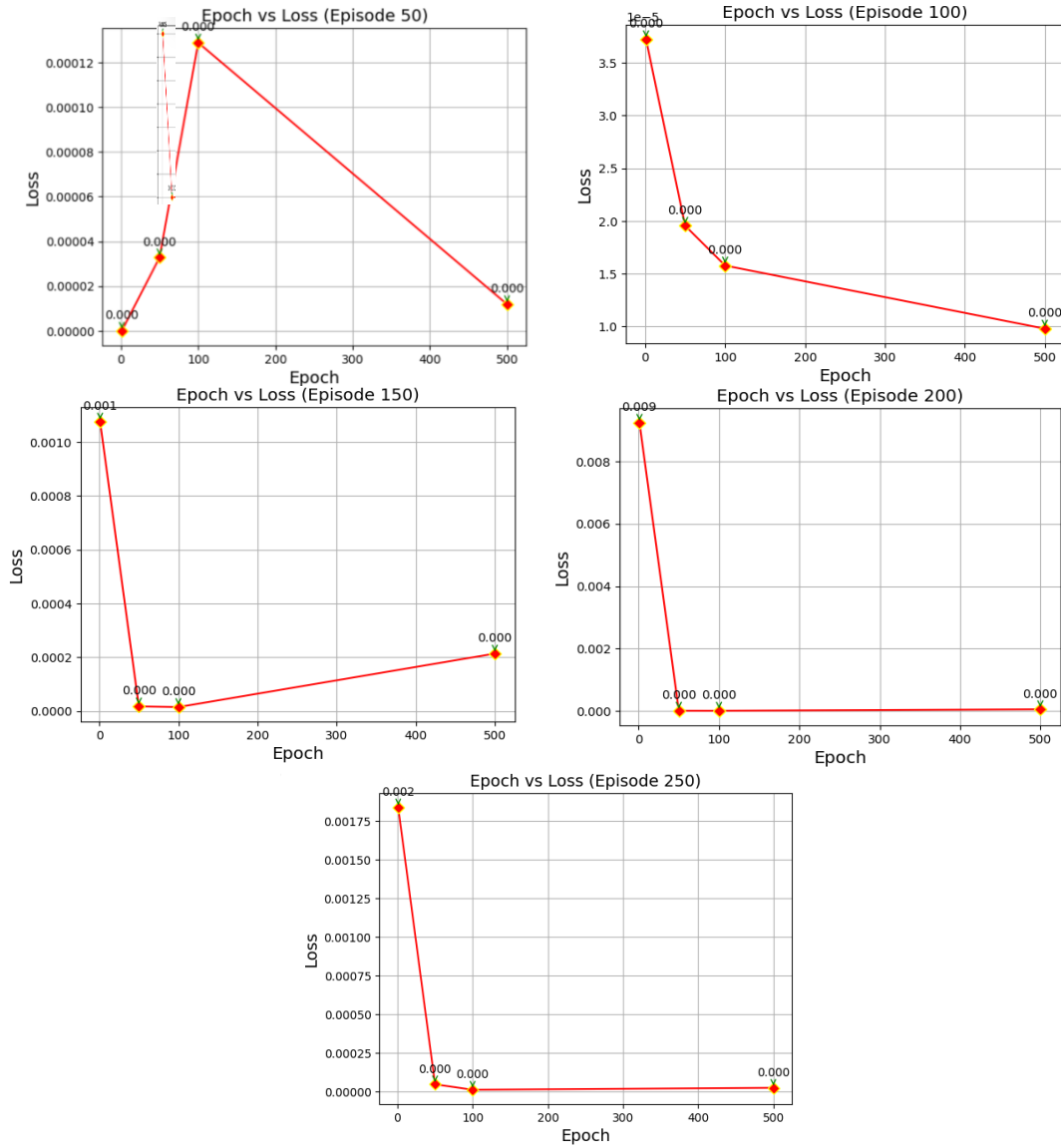
Fig. 4. Epoch vs loss

IV. Conclusions

In summary, our research endeavor has encompassed a thorough investigation into the factors that impact the efficacy of a DQN model in addressing the traveling salesman problem. The study focused on the influence of hidden layers, episodes, and epochs to elucidate their importance in optimizing the loss value. By doing a rigorous analysis of variance (ANOVA), we determined that neither episode nor epoch had a statistically significant impact on the loss value. Nevertheless, it is crucial to interpret these results in light of the limitations inherent in our sample size, the availability of replication data, and the degrees of freedom, as these factors might significantly influence the conclusions of the statistical analysis. Interestingly, although episode and epoch are statistically neutral, our visual representations in Figures 3 and 4 demonstrate a captivating storyline of a steady decline in loss value as the number of episodes and epochs increases. The observation above highlights the algorithm's proficiency in generating models from processed data, as demonstrated by the notable accomplishment of attaining a minimal loss value of 0.000010 during hyperparameter tweaking. Our research highlights the complex relationship between statistical analysis and empirical observations in practical contexts. Although statistical tests offer valuable insights, they may occasionally fail to capture the intricacies of complicated models. Therefore, to fully comprehend the issue, it is necessary to combine statistical rigor and empirical observations, which will allow us to effectively navigate the

always-changing field of deep reinforcement learning and route optimization with enhanced clarity and accuracy.

Declarations

Author contribution

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering and Informatics - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

References

[1] A. Jaradat, B. Matalkeh, and W. Diabat, "Solving Traveling Salesman Problem using Firefly algorithm and K-means Clustering," 2019 IEEE Jordan Int. Jt. Conf. Electr. Eng. Inf. Technol. JEEIT 2019 - Proc., no. September, pp. 586–589, 2019.

[2] J. N. Macgregor and T. Ormerod, "Human performance on the traveling salesman problem," Percept. Psychophys., vol. 58, no. 4, pp. 527–539, 1996.

[3] F. S. Gharehchopogh and B. Abdollahzadeh, "An efficient harris hawk optimization algorithm for solving the travelling salesman problem," Cluster Comput., vol. 25, no. 3, pp. 1981–2005, 2022.

[4] W. Gao, "New ant colony optimization algorithm for the traveling salesman problem," Int. J. Comput. Intell. Syst., vol. 13, no. 1, pp. 44–55, 2020.

[5] B. P. Silalahi, N. Fathiah, and P. T. Supriyo, "Use of Ant Colony Optimization Algorithm for Determining Traveling Salesman Problem Routes," J. Mat. "MANTIK," vol. 5, no. 2, pp. 100–111, 2019.

[6] A. François, Q. Cappart, and L.-M. Rousseau, "How to Evaluate Machine Learning Approaches for Combinatorial Optimization: Application to the Travelling Salesman Problem," 2019.

[7] M. P. Li, P. Sankaran, M. E. Kuhl, R. Ptucha, A. Ganguly, and A. Kwasinski, "Task Selection by Autonomous Mobile Robots in A Warehouse Using Deep Reinforcement Learning," Proc. - Winter Simul. Conf., vol. 2019-Decem, pp. 680–689, 2019.

[8] T. N. Adi, H. Bae, and Y. A. Iskandar, "Interterminal truck routing optimization using cooperative multiagent deep reinforcement learning," Processes, vol. 9, no. 10, 2021.

[9] S. Singh and A. Lodhi, "Study of Variation in TSP using Genetic Algorithm and Its Operator Comparison," Int. J. Soft Comput. Eng., no. 3, p. 264, 2013.

[10] H. A. Abdulkarim and I. F. Alshammari, "Comparison of Algorithms for Solving Traveling Salesman Problem," Int. J. Eng. Adv. Technol., vol. 4, no. 6, pp. 76–79, 2015.

[11] G. Ding and L. Qin, "Study on the prediction of stock price based on the associated network model of LSTM," Int. J. Mach. Learn. Cybern., vol. 11, no. 6, pp. 1307–1317, 2020.

[12] E. Xing and B. Cai, "Delivery Route Optimization Based on Deep Reinforcement Learning," Proc. - 2020 2nd Int. Conf. Mach. Learn. Big Data Bus. Intell. MLBDBI 2020, pp. 334–338, 2020.

[13] H. van Hasselt, A. Guez, and D. Silver, "Deep Reinforcement Learning with Double Q-Learning," Proc. Thirtieth AAAI Conf. Artif. Intell., pp. 2094–2100, 2016.

[14] Z. Hu, R. Beuran, and Y. Tan, "Automated Penetration Testing Using Deep Reinforcement Learning," Proc. - 5th IEEE Eur. Symp. Secur. Priv. Work. Euro S PW 2020, pp. 2–10, 2020.

[15] Y. Shen, N. Zhao, M. Xia, and X. Du, "A deep Q-learning network for ship stowage planning problem," Polish Marit. Res., vol. 24, no. S3, pp. 102–109, 2017.

[16] S. Yigit and M. Mendes, "Which effect size measure isappropriate for one-way andtwo-way anovamodels? A Monte Carlo simulation study," Revstat Stat. J., vol. 16, no. 3, pp. 295–313, 2018.

[17] R. A. Armstrong, F. Eperjesi, and B. Gilmartin, "The application of analysis of variance (ANOVA) to different experimental designs in optometry," Ophthalmic Physiol. Opt., vol. 22, no. 3, pp. 248–256, 2002.

[18] W. J. Ridgman, Experimentation in Biology., vol. 32, no. 4. London: Blackie, 1975.

[19] A. E. Verawati and A. N. S. Kiswanto, "the Effect of the Number of Hidden Layers in the Backpropagation in Case Study Weather Classification," Proxies J. Inform., vol. 2, no. 2, p. 58, 2021.

[20] M. Uzair and N. Jamil, "Effects of Hidden Layers on the Efficiency of Neural networks," Proc. - 2020 23rd IEEE Int. Multi-Topic Conf. INMIC 2020, pp. 1–6, 2020.

[21] K. G. Sheela and S. N. Deepa, "Review on methods to fix number of hidden neurons in neural networks," Math. Probl. Eng., vol. 2013, 2013.

[22] E. Tonnizam Mohamad, M. Hajihassani, D. Jahed Armaghani, and A. Marto, "Simulation of blasting-induced air overpressure by means of Artificial Neural Networks," Int. Rev. Model. Simulations, vol. 5, no. 6, pp. 2501–2506, 2012.

[23] M. Van Otterlo and M. Wiering, Reinforcement learning and markov decision processes, vol. 12, no. May. 2012.