A Study on Positional Embeddings and Ordinal Regression for Transformer-Based Bridge Deterioration Prediction Considering Component Locations

Shogo Inadomi1[0009-0009-3742-0299] and Pang-jo Chun1[0000-0002-9755-8435]

1 The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan.

inashouut@g.ecc.u-tokyo.ac.jp

**Abstract.** Conventional bridge deterioration prediction models do not account for the spatial relationships between structural components and damage locations. Since deterioration is a spatially distributed phenomenon, incorporating positional relationships can lead to more accurate predictions. The authors previously developed a Markov chain-based deterioration prediction model using a Graph Transformer architecture capable of capturing multi-scale spatial relationships among components. To further advance Transformer-based deterioration prediction, this study evaluates different positional representations and embedding methods, as well as the introduction of ordinal regression. A method that directly embeds the coordinates of components on structural diagrams was examined. This approach achieved a high precision of 81.7% in detecting damage evolution (DE precision), outperforming both the percentage prediction method and a Transformer model without positional embedding, and performing comparably to the graph-based method that uses explicit adjacency information. Furthermore, following recent practices in Transformer-based modeling, applying positional embeddings at all Transformer layers yielded even greater performance gains. Moreover, to address the ordinal nature of the five-level damage rating system, ordinal regression was introduced at the output layer. This resulted in a 7.1% improvement in DE recall compared to a simple classification model. These findings demonstrate the potential of building a highly reliable deterioration prediction model, enabling optimized inspection scheduling based on damage conditions and facilitating more efficient bridge maintenance and management.

**Keywords:** Bridge Deterioration Prediction, Transformer, Ordinal Regression.

1. Introduction

Predicting bridge deterioration is essential for determining appropriate maintenance timing. Accurate predictions enable timely interventions, which in turn minimize both the direct costs associated with repairs and the social losses caused by road closures [1]. Markov chain models have been employed for deterioration prediction, and enhancements addressing temporal patterns have been proposed including the introduction of non-homogeneous Markov chains [2].

Bridge deterioration is a phenomenon that spatially propagates across bridges and structural members. Thus, considering spatial relationships can contribute to improving prediction accuracy. Prior studies have incorporated the geographical locations of structures into prediction model to account for shared deterioration factors such as climate conditions and traffic loads [3]. In addition to such macro-scale relationships, multiple spatial scales must be considered, including micro-scale relationships such as the relative positions of individual bridge components.

Considering adjacency at the level of bridge component is crucial for capturing deterioration mechanisms. One such mechanism is the simultaneous progression of deterioration due to shared influencing factors. For instance, cracking may progress concurrently across deck slabs subjected to similar traffic loads. Another mechanism involves one type of damage triggering another, such as the progression of corrosion in areas affected by water leakage. This has been identified in bridge diagnosis, where steel corrosion is observed beside damaged drainage [4].

Despite the importance of component-level spatial relationships, methods incorporating such adjacency information remain limited, although some attempts including prediction of unobservable damage of lower components based on routine inspections [5] have been made. This study aims to develop a predictive model that explicitly incorporates spatial relationships among components and evaluate its effectiveness. To achieve this, a Transformer model [6] is introduced with specific adaptations for deterioration prediction. The Transformer model, recently used in the analysis of textual and image data, is characterized by its ability to consider the entire input data for processing information. The position embedding module within the Transformer enables the incorporation of positional relationships between data points, referred to as tokens, corresponding to each structural component. Given the multi-scale nature of spatial deterioration propagation, the Transformer, which balances global and local information processing, is considered a suitable architecture for this task.

Previous research by the authors demonstrated that representing the spatial relationships among components as a graph and applying a Graph Transformer achieved superior performance compared to methods that do not consider spatial information, such as the percentage prediction method [7], and graph neural networks (GNNs), which primarily aggregate information from neighboring nodes [8]. To further enhance Transformer-based predictive modeling, this study proposes improvements in three areas, including preprocessing of input positional data, positional modeling architecture, and output head design. Specifically, the planar coordinates of structural components from inspection diagrams are introduced for position embeddings, and adjustment are made regarding the method for incorporating positional embeddings (PEs) within Transformer layers. Furthermore, ordinal regression is introduced to explicitly account for the ordered nature of damage assessments.

Through comparative experiments, the proposed methods showed significant improvements in deterioration prediction. The coordinate-based PE achieved a DE precision of 81.7%, applying PEs at all Transformer layers further increased DE precision to 83.7% and DE recall to 43.5%, and introducing ordinal regression improved DE recall by 7.1% compared to standard classification.

1. Method

This study introduces three refinements to the prediction model, each aimed at enhancing its ability to represent spatial and ordinal structures in bridge deterioration. The following subsections describe these refinements in sequence.

* 1. Position Embedding Based on Inspection Diagram Coordinates

Spatial relationships among structural components were derived from inspection diagrams. In Japanese bridge inspection records, components are typically divided into units called "elements," and inspection data are collected at the element level. These elements are labeled with specific numbers on the diagrams (see Fig. 1), which were treated as representative spatial locations for the corresponding components in this research. The two-dimensional coordinates of each element number on the drawing were used as positional data. Although the locations of element numbers do not precisely reflect the actual physical dimensions of components, their relative positions provide a reasonable approximation of spatial adjacency. While graph-based representations [8] could offer higher fidelity in modeling such relationships, the proposed method benefits from reduced data preparation effort, as it avoids explicit graph construction.

The data processing workflow is as follows. Element numbers were extracted from diagrams using a semi-automated method that combined optical character recognition (OCR) with large language models (LLMs) [8]. Since each type of component is managed using a separate diagram, multiple diagrams were overlaid to integrate spatial information across components. Once integrated, the (x, y) coordinates of all elements were normalized using min-max scaling to be a range of [0, 1]. This normalization mitigated the effects of differences in diagram scale among bridges.

* 1. Adjustment of Position Embedding within Transformer Layers

In the original Transformer architecture [6], the position embedding is applied only once at the token embedding stage. The structure of the decoder-only Transformer is shown in Fig. 2 (left). In contrast, in Transformer-based models adapted for tasks such as graph processing (Graphormer [9]), point cloud analysis (Point Transformer [10]), and image segmentation (DETR [11]), positional information was integrated multiple times across Transformer layers (see Fig. 2 right).

For deterioration prediction, accurately capturing spatial relationships within the model is considered essential. This study introduced a modification in alignment with the latter approach. At the beginning of each Transformer layer, the operation of adding PEs on token embeddings was applied. The impact of this architectural adjustment on prediction performance was evaluated through comparison with a Transformer using a single PE.



Fig. 1. Element number diagrams in inspection records. The xy coordinates were normalized from 0 to 1 in this study.



Fig. 2. PEs employed in Transformer models. The left diagram illustrates the original architecture with a single PE, while the right diagram shows the modified architecture where PEs are incorporated at each Transformer layer.

* 1. Assessment with Ordinal Regression

Damage assessment is represented by five levels, labeled a through e, where a indicates the most intact state and e denotes the most severely damaged condition. This setup inherently introduces an ordinal relationship among the levels.

A straightforward modeling approach treats the prediction task as a standard multi-class classification problem, outputting independent probability estimates for each of the five classes. While this method can achieve strong performance when sufficient training data are available, it fails to account for the ordinal structure of the labels, potentially leading to unstable or semantically inconsistent predictions.

In deterioration prediction, avoiding large deviations from the correct class is particularly important. To address this, ordinal regression was incorporated, where the magnitude of misclassification is reflected in the loss function. This approach allows for more efficient learning with limited data and increases the likelihood of predictions that are closer to the ground truth during inference.

Specifically, to predict the five damage levels, four independent output values were produced using sigmoid functions. The four output probabilities indicate the likelihood that the predicted class falls into the following cumulative categories: b or higher (i.e., b, c, d, or e), c or higher, d or higher, and e or higher, respectively.

The loss function is defined as the sum of the binary cross entropy (BCE) loss calculated for each output probability. Formally, the total loss  is given by:

 (1)

where  is a binary indicator defined as:

 (2)

This formulation allows the loss to reflect the ordinal structure of the task, enabling the model to learn progressive classifications such as "b or higher" and "c or higher."

During inference, each of the four output probabilities was compared against a threshold of 0.5. A probability greater than 0.5 was interpreted as True, and otherwise as False. If the predicted outputs  are , for instance, the predicted class is determined to be c.

1. Experiment

The proposed models were applied to the Tokyo girder bridge dataset, which comprises 119 spans across 33 bridges. These structures include girders made of reinforced concrete, prestressed concrete, and steel, with cross-sectional shapes such as box and H girders. Inspections were conducted at five-year intervals between 2004 and 2024, resulting in 379 recorded transition instances over three to four inspection cycles per bridge.

The dataset was divided into training, validation, and test subsets. During training, model performance was monitored using the validation dataset by evaluating the loss at each epoch. The model weights corresponding to the lowest validation loss were selected for subsequent evaluation on the test dataset.

In addition to standard metrics such as loss and accuracy, damage evolution (DE) precision and recall were employed to assess the capability of capturing damage progression between consecutive inspection times. In this context, a true positive is defined as a case where both the observed and predicted damage levels increase between two consecutive inspection periods. DE precision quantifies the correctness of predicted damage progressions, while DE recall measures the proportion of actual progressions that were successfully detected.

The experimental results using diagram-based PEs are first presented. Compared to the percentage prediction method, the Transformer model incorporating positional information from diagrams demonstrated superior performance. Furthermore, both accuracy and DE precision improved relative to the Transformer model without PEs. These results underscore the advantages of incorporating spatial information for predicting structural component deterioration. Although graph representations [8] may enhance performance by leveraging precise adjacency relationships, this result reveals that even when utilizing the diagram-based positional representation, which has bias due to discrepancies between the inferred and actual member positions and lacking explicit adjacency information, performance improvements can still be achieved.

Furthermore, introducing PEs into all Transformer layers resulted in higher performance compared to the percentage prediction and the configuration with a single PE. This finding indicates that tuning the model architecture contributes to improved deterioration prediction, and that explicitly propagating positional information to deeper model layers is particularly effective.

**Table 1.** Result of introducing positional encoding.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Test Loss | Accuracy | DE Precision | DE Recall |
| Percentage Prediction | 0.0601 | 0.9764 | 0.070 | 0.2962 |
| Transformer w/o PE | 0.0437 | 0.9884 | 0.701 | 0.4019 |
| Transformer with single PE | 0.0397 | 0.9895 | 0.8174 | 0.3776 |
| Transformer with multi-time PE | **0.0386** | **0.9903** | **0.8367** | **0.4350** |

Finally, the results of the ordinal regression experiment are presented. This experiment was conducted as an extension of the Transformer model with multi-time PEs. Model evaluation employed nested cross-validation, a robust method for performance assessment, particularly when working with limited data. As demonstrated in prior studies including [12], k-fold cross-validation mitigates variance by averaging results across multiple train-test splits.

In this study, the dataset was split into three parts: training, validation, and testing. Accordingly, nested cross-validation was conducted with an outer loop of three folds and an inner loop of two folds. Specifically, three combinations of <train + validation> and <test> splits were used in the outer loop, and for each of these, two train-validation splits were applied in the inner loop, resulting in a total of six configurations. The final performance metrics were obtained by averaging across these six evaluations.

The results are shown in **Table 2**. Although a slight decrease in accuracy was observed when applying ordinal regression, the reduction was marginal and considered negligible. In contrast, both DE precision and recall improved. Notably, DE recall increased by 7.1 points, indicating a reduction in missed detections of damage progression. While the accuracy did not show substantial improvement, suggesting that precise classification remains a challenge, considering the ordinal nature of damage levels enabled more accurate detection of whether damage progression occurred.

The following points should be noted. First, the loss values are not directly comparable across classification and ordinal regression because different loss functions were employed. In addition, the numerical results presented in Tables 1 and 2 cannot be directly compared, as cross-validation was performed in only the latter.

**Table 2.** Result of introducing positional encoding.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Accuracy | DE Precision | DE Recall |
| Classification | **0.9879** | 0.5520 | 0.2272 |
| Ordinal Regression | 0.9871 | **0.5558** | **0.2980** |

1. Conclusion

This study evaluated the effectiveness of three methodologies: the incorporation of diagram-coordinate embeddings, introduction of multiple PE modules, and the application of ordinal regression. The experiments on PEs demonstrated that introducing and utilizing spatial relationships multiple times is beneficial for degradation prediction tasks, complementing the graph-based approach that use explicit adjacency information [8]. Furthermore, the adoption of ordinal regression improved the accuracy of progressive deterioration prediction in multi-stage classification tasks.

Nonetheless, several limitations should be acknowledged. First, further investigation is required to determine which specific damage progression patterns derive benefit most from each of the proposed techniques. Second, expanding the bridge dataset is crucial to improving model performance. While graph-based representations more accurately capture adjacency relationships and are thus desirable, this study suggests that graph structures are not strictly required. This finding highlights the potential for scaling up data collection through alternative spatial representations.

References

1. Saydam, D., Frangopol, D.M., Dong, Y.: Assessment of Risk Using Bridge Element Condition Ratings. J. Infrastruct. Syst. 19, 252–265 (2013).

2. Fu, G., Devaraj, D.: Methodology of Homogeneous and Non-Homogeneous Markov Chains for Modelling Bridge Element Deterioration. 90 (2008).

3. Gao, L., Yu, K., Lu, P.: Considering the Spatial Structure of the Road Network in Pavement Deterioration Modeling. Transp. Res. Rec. (2023). https://doi.org/10.1177/03611981231188373.

4. Yamane, T., Chun, P.J., Dang, J., Okatani, T.: Deep learning-based bridge damage cause estimation from multiple images using visual question answering. Struct. Infrastruct. Eng.: Maint. Manage. Life-Cycle Des. Perform. 1–14 (2024).

5. Minami, T., Fujiu, M., Takayama, J.: Analysis of Bridge Damage Discovered By the Bridge Surface Condition Using Bridge Inspection Data. J. Jpn. Soc. Civ. Eng. Ser D3 (Infrastruct. Plan. Manag.). 74, I\_339-I\_348 (2018).

6. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Adv. Neural Inf. Process. Syst. 2017-Decem, 5999–6009 (2017).

7. Yi, J., Saito, M., Sinha, K.C.: Bridge performance prediction model using the Markov chain. Transp. Res. Rec. 25–32 (1988).

8. Inadomi, S., Chun, P.J.: Spatially-Aware Markov Chain-Based Deterioration Prediction of Bridge Components Using a Graph Transformer. Comput.-aided civ. infrastruct. eng.

9. Ying, C., Cai, T., Luo, S., Zheng, S., Ke, G., He, D., Shen, Y., Liu, T.Y.: Do Transformers Really Perform Bad for Graph Representation? Adv. Neural Inf. Process. Syst. 34, 28877–28888 (2021).

10. Zhao, H., Jiang, L., Jia, J., Torr, P., Koltun, V.: Point Transformer. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE (2021). https://doi.org/10.1109/iccv48922.2021.01595.

11. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers, http://arxiv.org/abs/2005.12872, (2020).

12. Assaad, R., El-adaway, I.H.: Bridge infrastructure asset management system: Comparative computational machine learning approach for evaluating and predicting deck deterioration conditions. J. Infrastruct. Syst. 26, 04020032 (2020).