$NTAB_0$: Design Priors for AI-Augmented Generative Design of Network Tied-Arch-Bridges

Sophia V. Kuhn[1], Rafael Bischof[2], Georgios Klonaris[1], Walter Kaufmann[1] and Michael A. Kraus[1]
[1]Chair of Concrete Structures and Bridge Design, ETH Zurich, Stefano-Franscini-Platz 5, 8093 Zurich, Switzerland
[2]Swiss Data Science Center, Turnerstrasse 1, 8092 Zürich, Switzerland
E-mail(s): {sophia.kuhn,kraus}@ibk.baug.ethz.ch

Abstract: Projects in the Architecture, Engineering and Construction (AEC) industry inherit a great complexity due to a tremendous amount of design parameters, multiple objectives, and many involved stakeholders. Especially in the conceptual design stage of bridges, an in-detail analysis of many performance attributes for each design alternative is time-consuming and infeasible under the current approaches. In the industry today, therefore the initial design solution predominantly depends on the expertise of the involved team. In contrast to the status quo, this paper introduces the novel concept of bridge design prior models to predict the layout and structural properties of bridges as the (near-optimal) starting point for Generative Design. The concept of design prior models for bridges is demonstrated on network tied-arch bridges (NTAB). $NTAB_0$ is calibrated upon a curated database consisting of existing real-world NTABs and captures numeric, semantic, and topological relations between bridge properties such as materials, cross-sections or bracing systems. First, a clustering analysis is performed by applying the k-Prototype and DBSCAN algorithms. In the second step, a predictive model is trained using a gradient-boosted decision tree algorithm. A subsequent study evaluates the suitability of the algorithms to serve as sensible design priors. We found that the AI prior model $NTAB_0$ is able to suggest meaningful design parameters, assisting the designing team with an informed initial bridge design for further design space exploration and optimisation. It enables designers to make more informed decisions towards optimised bridge structures at an early design stage. The application of the AI prior model shows great potential to improve future construction projects by providing easy and fast access to the information saved in the existing structures of today.

Keywords: Conceptual Structural Design, Machine Learning, Design Automation, Algorithms

1 Introduction

In current Architecture, Engineering and Construction (AEC) practice, the conceptual design phase is often disconnected from performance evaluations of the structure with respect to safety, functionality, cost, material impacts, and the construction processes. However, this phase is most influential for a building project. Today, the initial design is mainly based on the prior knowledge of (i) the designing team, and (ii) the manual analysis of a few similar reference projects. This design practice results in lengthy and costly processes with numerous manual translations, detailing, and reiterations of rigid design models between early-stage designers and engineers or contractors of later planning stages.

Built bridge structures however represent the final result of detailed investigations during a bridge project, in which the expertise of civil engineers was applied. This fact motivates the derivation of a method for data collection and automated analysis as well as calibration of quantitative prior design models to be used in conceptual design phases. Specifically, this paper proposes to (i) start free-access and open databases on structural properties of bridges, and (ii) use modern Machine Learning (ML) algorithms for data mining and calibration of predictive prior models upon these databases to speed up early phase design and/or validate the design process choices. This idea is investigated within the *Bridge Genome Project*, cf. Fig. 1, for different bridge types. The prior models and their insights are interpreted as the "genome" of how bridges have been planned and built. The workflow and results for calibrating a design prior model upon a repository are presented for the case of network tied-arch bridges (NTAB), described by a parameter set $\theta_B \in R^d$, leading to the $NTAB_0$ prior.

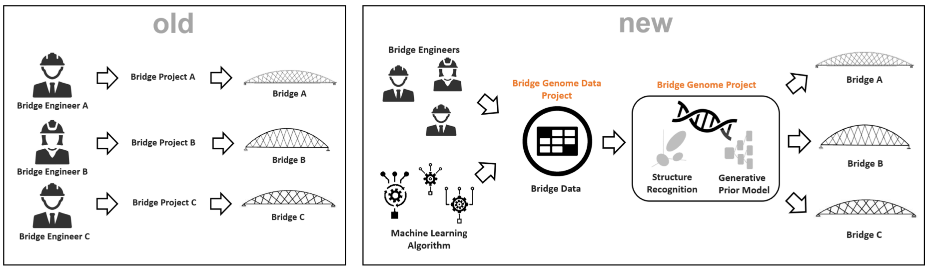

Figure 1: The Bridge Genome Project

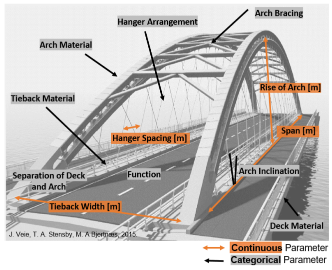

Figure 2: Bridge properties

2 Background and Related Literature

Bayesian methods are an alternative approach to traditional statistical analysis. They are employed widely for analysing and interpreting data or training Artificial Intelligence (AI) models in many fields, including data and computer science as well as civil engineering [1]. The Bayesian method computes a posterior probability about a problem based on observed data (likelihood) and prior knowledge using Bayes' theorem, cf. Sec. 3. Prior information is obtained from the results of previous experiments, simulations, domain theory, and expert assessments [2]. Data-driven learning (i.e. learning from data repositories or sets) through Machine and Deep Learning allows for new discriminative and generative models to be employed to the design task. Discriminative models learn to solve a learning task (e.g. classification, regression, clustering, dimensionality reduction) where for a new input the respective output is provided. Applying a Bayes view, a discriminative model strictly learns the posterior probabilities of the output given an input [2]. In our present work, the inferred discriminative models serve as prior models for the design task. Here the posterior probabilities of the design output for a given input (i.e. a specific bridge construction project situation) are augmented by the informative prior.

The potential of repositories, evolutionary algorithms, and parametric tools as drivers for design actions was explored for performance-based shading device design for office buildings by [3]. [4] is one of the few studies on bridge simulation data mining as an approach to extract knowledge and decision rules from a database of pre-computed models (design variants) containing simulation results. However, the greatest current hurdle for data mining employing AI in the AEC sector is the lack of systematized, free-accessible databases to learn from the successes and failures of past engineering projects. The potentials and pitfalls of data mining in the AEC sector are addressed in [5].

3 Methods

The formal capturing and quantitative incorporation of domain-knowledge in the form of priors distributions of data about a design problem works within the Bayesian framework: $\mathcal{P}(\mathcal{D}|\mathcal{M}(\theta))$, where $\mathcal{P}$ denotes a probability distribution, $\mathcal{D}$ is the data set and $\mathcal{M}(\theta)$ the parametric model with design parameters $\theta \in \mathrm{R}^{\mathrm{d}}$ respectively. Instead of employing Bayesian model averaging, we suggest to a-priori select a model $\mathcal{M}^*(\theta)$, and subsequently the observed data set is a manifestation of that model drawn according to the probability $\mathcal{P}(\mathcal{D}|\mathcal{M}^*(\theta))$. The goal of the ML algorithms then is to estimate $\mathcal{P}(\mathcal{M}^*(\theta))$. This paper applies the described approach to NTABs for calibrating the NTAB_0 prior.

3.1 Database of Network Tied-Arch Bridges (NTABs)

In the AEC Industry, the systematic documentation of structural data (such as cross-sections, material grades, bracing system types, etc.) of completed projects is currently still insufficient due to the absence of public repositories. To nevertheless have a basis for creating a prior model for bridge design, a data set was curated, which is publicly available via [6]. The database is structured in a table containing mixed-data-type properties including functionality, geometrical, and material parameters of 203 NTABs. Fig. 2 shows the important bridge features together with a distinction of the categorical resp. continuous data types. As the curated data set exhibits missing values, two pre-processing methods are followed to gain a complete data set: for the first data set, the data is filtered for complete rows (bridges). For the second, the missing values are filled utilizing a k-NN algorithm [7]. k-NN estimates missing data by computing a weighted average over the k nearest neighbors in the data set based on the euclidean distance between non-missing entries. We empirically found k = 3 to yield a good compromise between consistency and diversity in the data set.

NTABs belong to the class of (tied-)arch bridges and have been first introduced by Per Tveit in the 1950s [8]. Arches are highly efficient structures due to their ability to carry loads by compression in the case that the thrust line lies within its cross-section [9]. The horizontal forces carried by the arch are short-circuited through the deck in tension. The structure typically includes two arch ribs to which the vertical loads acting on the deck are suspended by inclined and intersecting hangers.

3.2 Machine Learning (ML) Algorithms

Our approach suggests to (i) identify structure (i.e. similarity groups) within the bridge data using clustering algorithms, and (ii) train a discriminative ML model for predicting suitable bridge parameters.

The cluster analysis is applied to the filtered data set. As the NTAB data set carries mixed data types, the k-Prototype algorithm as implemented in [10] is applied. The k-Prototype algorithm combines the well-known k-means algorithm [11] with the k-modes algorithm [12], to handle mixed data types of continuous and categorical data. To enable the identification of well separated clusters in all dimensions the equal contribution of all properties to the clustering process is achieved by (i) individually standardizing each property column to zero mean and unit variance, and (ii) by iteratively choosing a relative importance weight γ so that an even contribution of both the continuous and categorical properties is observed within the loss function of the k-prototype algorithm. Furthermore, the clustering is conducted 100 times in parallel with a new 'Cao' initialisation each time. The number of clusters k is chosen according to the 'elbow method'. In order to test the stability of the clusters induced via distance-metrics by the k-prototype algorithm, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [13] algorithm is used to provide a clustering of the data by fitting probability distributions. However, the results are not shown within this paper.

The second step is to calibrate mixed-data-type ML algorithms upon the NTAB data set as a discriminative design prior for predicting structural bridge properties for a specific project. This means, that starting from a set of fixed design parameters $\theta_{\mathrm{B}_1} \in \mathrm{R}^{\mathrm{d}_1}$, the algorithm generates meaningful suggestions for the remaining parameters $\theta_{\mathrm{B}_2} \in \mathrm{R}^{\mathrm{d}-\mathrm{d}_1}$ in a specific order, taking into account all the previous predictions that have been made. For the data set at hand, a decision tree approach is chosen as these algorithms are capable of solving both classification as well as regression tasks, which is necessary due to the mixed-type nature of the data set. Specifically, we calibrate a CatBoost [14] algorithm to the bridge data set as in a pre-study it outperformed other existing state-of-the-art implementations of gradient boosted decision trees (such as XGBoost) in terms of accuracy. This is in accordance with literature, where CatBoost was found especially efficient and accurate if categorical features are present and play an important role [15]. The model is trained on 85% of the bridges from the second imputed NTAB data set. The residual 15% of samples are used to test model performance. A stratified train-test splitting method was applied. For the regression head, `CatBoostRegressors` [14] are fitted to the training data using the *Root Mean Squared Error* (RMSE) loss function with an evaluation via the RMSE on the test data set. For the classification head, `CatBoostClassifiers` [14] are calibrated to the training set using the *Logloss* function for binary categorical properties and the *MultiClass* loss function for categorical properties with more than two categories (i.e. classes). Model evaluation was also conducted on the test data by calculating the accuracy and balanced accuracy. For both algorithm heads, hyper-parameter tuning was performed using grid search.

4 Results

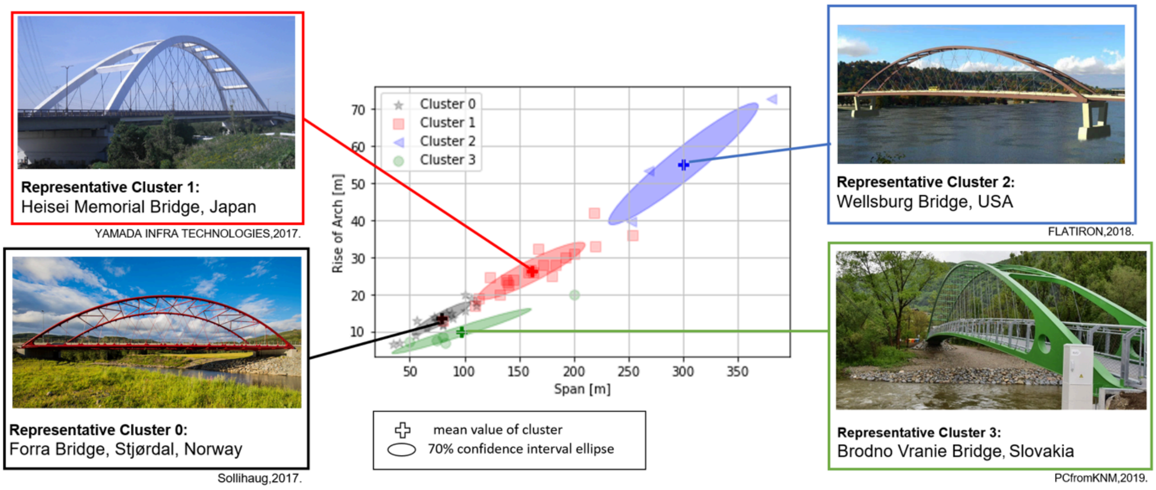

Figure 3: 2D cut scatter plot of the multi-dimensional clusters with a representative bridge for each

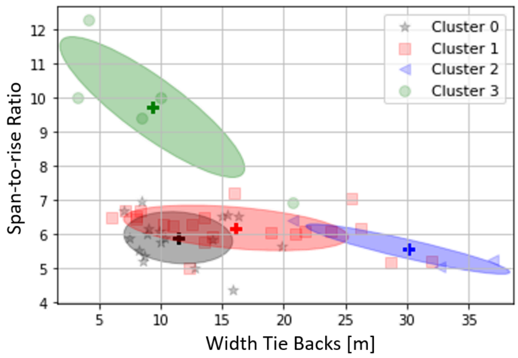

Figure 4: 2D scatter plot of the multi-dimensional clusters (+: Mean value of cluster; 70% confidence interval ellipse)

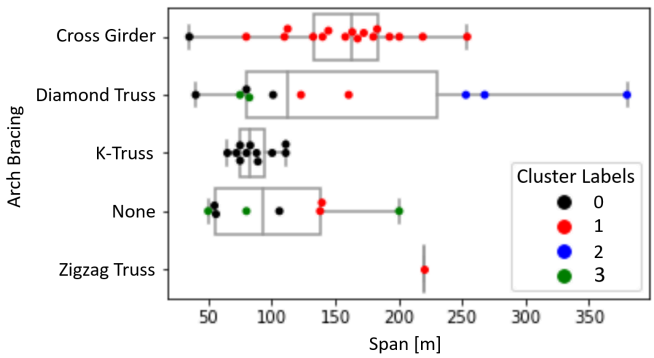

Figure 5: 2D swarm plot of the four multi-dimensional clusters identified with the k-Prototype algorithm (25%- and 75%-quantile Boxplot limits)

For the clustering analysis a reasonable number of clusters for the k-Prototype algorithm is established upon the resulting cost degradation with increasing number of clusters k, which shows a slope change ("elbow") at k = 4 for multiple γ-values. A suitable γ was found to be 4. The four multi-dimensional clusters found in the bridge data set by the k-Prototype algorithm are visualised in Figs. 3, 4, and 5. A good separation between the clusters is visible in all dimensions, where we show as an example the distributions for span to rise, width-tie-back to span-to-rise ratio, and span to arch-bracing. These are to be interpreted as 2D-cuts of the multi-dimensional cluster.

The achieved performance of the trained ML model w.r.t. individual parameter predictions are summarized in Table 1. Initially, we assume the parameters *span, tie back width* and *bridge function* as fixed. The remaining parameters are predicted in series by the ML model. For the classification heads, a high predictive performance was achieved for all parameters on the training set. On the evaluation set an accuracy of above 77% was achieved for all parameters, but a decreased balanced accuracy for

Table 1: Prediction performances of the regression - classification models of the $NTAB_0$ prior model

	On Training Set		On Evaluation Set	
Classification:	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
1. Separation Deck-Arch Structure	0.87	0.87	0.84	0.85
2. Arch Material	1.00	1.00	0.97	0.67
3. Tieback Material	0.97	0.68	0.81	0.52
4. Deck Material	0.92	0.82	0.94	0.89
Regression:	RMSE		RMSE	
5. Rise of Arch	5.81 m		15.58 m	
6. Hanger Spacing	1.2 m		16.7 m	
Classification:	Accuracy	Balanced Accuracy	Accuracy	Balanced Accuracy
7. Hanger Arrangement	0.99	0.98	0.84	0.46
8. Arch Inclination	0.89	0.89	0.94	0.94
9. Arch Bracing	0.95	0.91	0.77	0.53

arch and tieback material, hanger arrangement and arch bracing is visible. The regression models are performing well on the training set, but exhibit a substantially increased RMSE on the evaluation set.

5 Discussion

The k-Prototype appropriately solved the pattern recognition task given the difficulty of the mixed data types present. Four clusters, distinctly well-separated in all dimensions, were found, which proves the equal contribution of all bridge features (both continuous and categorical) during the clustering process. Comparing the probability distributions of the individual features within the clusters reveals interesting relations: The identified bridge clusters are clearly separated w.r.t the span, indicating strong similarities in all bridge properties for the span ranges of the clusters (Fig. 3). Another dominant distinction is the span-to-rise ratio, separating mostly pedestrian bridges from the road and rail bridges (Fig. 4). Fig. 5 displays the cross girder as the most frequently used arch bracing type, mostly used for medium-span NTABs. In addition, engineers in the past have found K-truss bracings beneficial for short-span NTABs and diamond truss bracings effective for long-span NTABs. These detected distinct patterns indicate a good chance for calibrating discriminative ML prediction models for bridge features.

Evaluation of the trained discriminative model showed that gradient boosted decision trees are suitable to capture the dependencies of all bridge parameters in the training set. A reduced balanced accuracy compared to the unbalanced accuracy is identified on the evaluation data set for the classification of some of the bridge parameters. The reason here is the strong imbalance of classes present in the data set, leading to a more performant ML classifier for the dominant classes. Improving the performance by employing skew-insensitive metrics such as DKM and Hellinger distance as splitting criteria in the construction of the decision tree [16] is omitted given the small size of the bridge data set. The limited performance of the regression algorithm on the evaluation data is a result of overfitting, which is often detected for small data sets. Another probable reason is that the available features are insufficient to find a generalised relationship. While the performance evaluation proves that the ML model provides suitable recommendations based on past bridge projects, we can also compare the recommendations to research findings on NTABs. For an input parameter vector falling in the medium-span cluster (red in Fig. 5) the model recommends a span-to-rise ratio of 6.4, which matches literature recommendations of 5.8 to 6.67 [17]. For the same example, the model advocates the cross girder due to its frequent use, while [18] identifies it to be the least cost-efficient arch bracing type. This reveals a limitation of the the presented method. It does not evaluate weather the underlying data of existing bridges is compliant with current design standards or good engineering practice w.r.t. efficiency, durability, etc.. Hence, the method can adopt systematic mistakes from past bridge projects due to its data-driven nature. Additionally, The cluster analysis and prior model calibration were performed on a small data set of 203 bridges and can therefore not claim to be entirely representative of all the already built NTABs worldwide. Consequently, gathering a larger data set of NTABs is initiated for future investigations. While the approach was applied to NTABs in the present work, the implemented framework is directly applicable to further data sets of other frequently built structure types.

6 Conclusions

Today, the conceptual bridge design remains disconnected from performance evaluations of later stages and therefore dictated by the prior knowledge of the involved experts, leading to a lengthy and costly design process characterized by many iterations. We provide a two-stage method of clustering a data set of structural bridge information and subsequent building discriminative regression resp. classification models as informative design prior. The performed k-Prototype cluster analysis detected design patterns for NTABs in the form of 4 distinct clusters. The identified structure enabled to draw useful conclusions about sensible bridge parameter choices, which can be checked for plausibility by bridge engineering experts yet also inform about hidden design patterns. The discriminative gradient boosted decision tree algorithms serve as a powerful prior model for predicting sensible parameters for a new bridge project situation based on existing bridge construction projects. The two-stage prior modelling approach is found especially useful for detecting multi-dimensional dependencies between the bridge features, which are not easily identifiable by conventional methods. Identified limitations are the risk of adopting mistakes from the past and the limited availability of data in the AEC sector.

Acknowledgements

The authors would like to thankfully acknowledge the facilities of Design++ at ETH Zürich and the funding through ETH Foundation grant No. 2020-HS-388 (provided by Kollbrunner/Rodio).

References

[1] M. A. Kraus, “Machine learning techniques for the material parameter identification of laminated glass in the intact and post-fracture state”, 2019.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2016.

[3] B. Ercan and S. T. Elias-Ozkan, “Performance-based parametric design explorations: A method for generating appropriate building components”, *Design Studies*, vol. 38, pp. 33–53, 2015.

[4] S. Burrows, B. Stein, J. Frochte, D. Wiesner, and K. Müller, “Simulation data mining for supporting bridge design”, in *Proceedings of the Ninth Australasian Data Mining Conference-Volume 121*, Citeseer, 2011, pp. 163–170.

[5] V. Ahmed, Z. Aziz, A. Tezel, and Z. Riaz, “Challenges and drivers for data mining in the aec sector”, *Engineering, Construction and Architectural Management*, 2018.

[6] A. Müller, S. Kuhn, and M. Kraus, *Scientific machine learning for structural engineering repository*, 2022. [Online]. Available: https://sciml4structeng.github.io/Repository, (accessed: 07.07.2022).

[7] G. Batista and M.-C. Monard, “A study of k-nearest neighbour as an imputation method.”, vol. 30, Jan. 2002, pp. 251–260.

[8] P. Tveit, “Considerations for design of network arches”, *Journal of Structural Engineering*, vol. 113, no. 10, pp. 2189–2207, Oct. 1, 1987, Publisher: American Society of Civil Engineers.

[9] W. Kaufmann, *Lecture Notes: Arch Bridges*, 2021.

[10] N. de Vos. “Kmodes”. (2021), [Online]. Available: https://datasolut.com/wiki/clusteranalyse/. (accessed: 10.04.2021).

[11] E. W. Forgy, “Cluster analysis of multivariate data : Efficiency versus interpretability of classifications”, *Biometrics*, vol. 21, pp. 768–769, 1965.

[12] Z. Huang, “Extensions to the k-means algorithm for clustering large data sets with categorical values”, *Data Mining and Knowledge Discovery*, pp. 283–304, 1998.

[13] Andrewngai, “Understanding dbscan algorithm and implementation from scratch”, *Towards Data Science*, 2020.

[14] Y. LLC, *Catboost*, 2019. [Online]. Available: https://github.com/catboost/catboost.

[15] A. Jha. “CatBoost – A new game of Machine Learning”. (2020), [Online]. Available: https://affine.ai/catboost-a-new-game-of-machine-learning/. (accessed: 27.06.2021).

[16] W. Daelemans and B. Goethals and K.Morik (Eds.), *Machine Learning and Knowledge Discovery in Databases - ECML PKDD 2008, Part I*. Springer Berlin Heidelberg, 2008, pp. 241–256.

[17] P. Tveit, “Systematic Thesis on Network Arches”, Adger University, 2014.

[18] F. Schanack, “Network arch bridges”, Ph.D. dissertation, University of Cantabria, 2008.