A Machine Learning Based Approach to Bridge Design Prototyping

Achyuthan Jootoo[1], David Lattanzi[2]

1) Research Assistant, Sid and Reva Dewberry Department of Civil, Environmental, and Infrastructure Engineering, George Mason University, Fairfax, Virginia, USA. Email: ajootoor@gmu.edu
2) Assistant Professor, Sid and Reva Dewberry Department of Civil, Environmental, and Infrastructure Engineering, George Mason University, Fairfax, Virginia, USA. Email: dlattanz@gmu.edu

Abstract:

Billions of dollars are spent each year on the construction of bridges. While the financial impact of sub-optimal bridge designs is difficult to determine, given the hundreds of thousands of bridges in the United States alone, it is clearly significant. The bridge design process can be broken into three stages ¬¬– (1) preliminary design to choose the bridge prototype for detailed design, (2) detailed final design, and (3) design optimization. This paper focuses on the first stage of the design process – choosing the bridge structure prototype. In the preliminary design stage, an engineer first chooses a preliminary design predominantly via his/her experience in tandem with preliminary calculations. To simulate this experience-based design process, machine learning is used to predict multiple bridge prototypes via models developed from the National Bridge Inventory data of over 600,000 bridges and seismic data from United States Geological Survey (USGS). The results of this machine learning study indicated several key design parameters that controlled most design choices, and demonstrated that such models are capable of sufficient accuracy for use in the design process. Significant variation in model performance was observed among individual states, likely due to differences in design preferences, material availability, and site-specific considerations. The next phase of this work will explore how to combine these machine learning models with advanced optimization methods in an effort to provide robust design support systems to engineers.

Keywords: Machine Learning, Bridge Design, Optimization, Evolutionary Algorithms, Data Mining

1. INTRODUCTION

All bridges constructed in the US are designed as per the AASHTO design codes and are hence deemed safe for usage at the time of construction. However, while the AASHTO code ensures a safe design, it does not guarantee an optimal one. The optimality of a design often depends on a broad variety of parameters, such as the engineer's skill and experience, which can be difficult to quantify. A suboptimal bridge is more costly to construct and potentially more expensive to maintain. This paper presents current efforts to develop a framework for design of bridge prototypes through a machine learning based methodology.

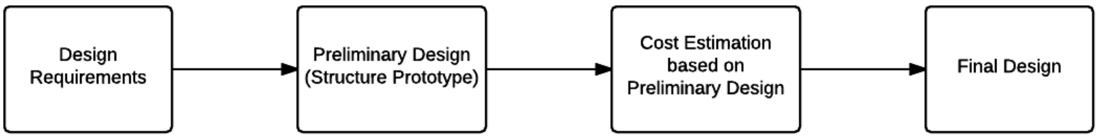

Figure 1. Stages in the design of a structure

The design of a bridge can be broken into multiple stages as illustrated in Figure 1. In the first stage, the design requirements, such as geometric constraints and usage requirements are specified. The engineer then devises a preliminary design based on these requirements, his/her experience and some structural calculations. The preliminary design or prototype is then used for a cost estimate, which is followed by final design detailing of all the components of the structure. The cost estimate and final design processes are heavily dependent on the prototyping phase, reflecting the importance of making the right choice for preliminary design.

To improve the prototyping process, machine learning can be used to augment the experiential nature of design prototyping, and to provide potentially optimal structure system types on the basis of design parameters available to an engineer. After the prototype has been predicted, an algorithm designs the bridge as per the AASHTO design criteria, with the aid of an evolutionary algorithm to provide an optimal design.

1.1 Literature Review

There have been a variety of machine learning implementations in civil engineering and in fields such as medical diagnosis. The use of machine learning for prototype prediction is similar in concept to the problem of medical diagnosis and structural health monitoring, in which the design requirements are analogous to symptoms or sensor or visual data from the bridge and the preliminary design type corresponds to the patient's disease or health of the bridge. Similar to structural design, where different design requirements result in different designs, different medical symptoms indicate different diseases.

There has been extensive research using and comparing different machine learning techniques for medical decision making and diagnosis (Harper, 2005; Kononenko, 2001; Lavrac, 1999; Potter, 2007). Past research in this domain indicates that the confidence of a physician in a machine learning based model for disease prediction is higher if he can explain the model's decision making process i.e. the model is comprehensible (Bellazi & Zupan, 2008). This aspect is very relevant for civil engineers who are more likely to trust a machine learning model if it is comprehensible. In civil engineering, research in structural health monitoring uses machine learning primarily as a means of pattern recognition (Worden & Dulieu-Barton, 2004; Worden & Manson, 2007). Machine learning has also been used often for predicting bus arrival times and designing adaptive traffic signals (Abdulhai et al., 2003; Chien et al., 2002). Machine learning has also been used in structural engineering to calculate effective length factor and for structural optimization (Hajela & Lee, 1997; Hung & Jan, 1999). A survey of the use of machine learning in structural engineering from 1989 to 2000 can be found in (Adeli, 2001).

1.2 Contributions of this work

Given prior efforts to use machine learning for health monitoring or medical diagnosis, the purpose of this study was to investigate the use of machine learning in the prototyping phase of the engineering design process. The next section describes the developed methodology and gives an introduction to the machine learning techniques applied in this research. The third section explains the data analysis and experiments conducted with detailed results. The final section presents the conclusions and proposed future work for this research.

2. METHODOLOGY

Machine Learning (ML) encompasses algorithms that use data, usually in large amounts, to develop models that can predict or classify new instances of known similar data. In the context of this research, data on the design details of previously constructed bridges in the United States was used to construct the predictive models.

2.1 Datasets Used for Analysis

This study uses the NBI dataset (FHWA, 2015a) as the primary dataset and combines it with the USGS seismic dataset (USGS, 2015). The NBI dataset provides information regarding the bridge specifications that are available to a bridge designer prior to designing a bridge. The USGS data for seismic intensity provides additional information for deciding the bridge preliminary design.

The NBI data contains over 110 attributes for each bridge of which a majority are not relevant for a bridge engineer when designing a bridge. A few attributes provided in the dataset are: bridge length, number of spans, navigational vertical clearance, maximum span length, deck structure type, sufficiency rating, material type, design type and average daily traffic. A full list of attributes of the NBI dataset can be found on the FHWA website (FHWA, 2015b). The USGS dataset provides the 2014 peak ground acceleration with a 2% probability of exceedance for the lower 48 states. Seismic data for Alaska, Hawaii and Puerto Rico was only available for 2013 during the time of the research and hence was not considered. The dataset developed for this research is publicly available (LRG, 2015).

2.2 Dataset Development

The development of this dataset consisted of two stages of cleaning of the NBI dataset followed by integration of the USGS seismic intensity dataset. The first stage removed attributes which were irrelevant for a structural engineer or unknown to an engineer during design stage such as the bridge address, latitude, longitude, county code, sufficiency rating and route number. The next stage of data cleaning was done by two methods. In the first approach, preliminary machine learning models were first constructed. A single attribute was then removed from the dataset and a new model was constructed. This method will be referred to as "leave-one-attribute-out" strategy. If removal of an attribute did not affect Bayesian Network model accuracy by more than 1%, then it was considered as unimportant for prototype design prediction. The second method of data cleaning was the chi-squared feature evaluation algorithm (Datumbox, 2015), which evaluates the correlation between an attribute and the class to be predicted. Attributes with higher correlation score are better for use in classification. For evaluating attribute suitability, the second attribute which had a drop of 30% in chi-squared score (relative to the previous attribute) was found out. All attributes, including this one, with scores lesser than that of this attribute were not considered for developing machine learning models. After these two steps of data cleaning, we are left with only 4 attributes: maximum span length, average span length, deck structure type, material type and the attribute to be predicted, design type. Then the seismic intensity corresponding to the bridge's latitude and longitude from the USGS dataset was added to the NBI data.

2.3 Machine Learning Algorithms

For design prototyping applications, an ideal machine learning algorithm should not only have a high accuracy, but also should be comprehensible by structural engineers (Bellazi & Zupan, 2008). The comprehensibility is more important from a structural design standpoint since the model might make erroneous assumptions, leading to design errors, that can only be corrected if the decision making process can be followed, understood and checked by an engineer. Keeping this in mind, two algorithms were used for the study: Decision Tree and Bayesian Network Classifier because they both result in models which can be understood by an engineer.

(1) Decision Tree

A decision tree model visualizes all the data as classifiable using a sequence of conditional statements. The representation of this model is in the form of a tree. Figure 2 shows a simplified example decision tree for bridge prototyping. The figure implies that if material is steel, and maximum span length is greater than 150ft, then the bridge is of type Stringer or Multi-beam. If the span length is less than 150ft, then a different design type is used. The details of how the decision tree is constructed cannot be explained in this manuscript and the authors refer the reader to the relevant explanations from the literature (Mitchell, 1997; Quinlan, 1986; Witten et al., 2011).

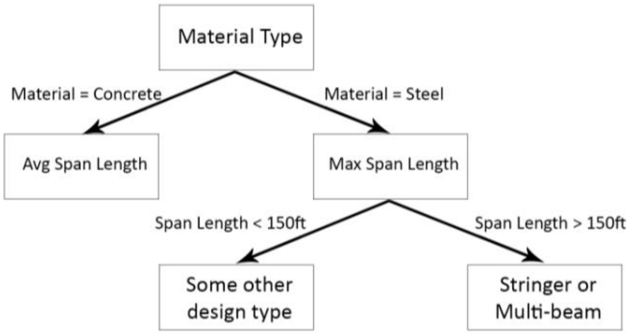

Figure 2. Example decision tree for bridge design prototyping

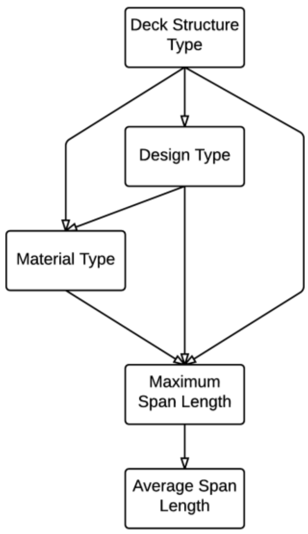

Figure 3. Example Bayesian Network for Bridge Prototyping

(2) Bayesian Network Classifier

This classifier learns a Bayesian network from the instances that are provided in the dataset. The probability of each class is determined from the network and associated probabilities learned from the instances provided to develop the network. A sample Bayesian network for predicting the design type using only the final four attributes from NBI dataset is shown in Figure 3. Details on the Bayesian Network classifier are beyond the scope of this work and the authors refer the reader to previous explanations in the literature (Mitchell, 1997;

Cheng & Greiner, 2001).

(3) Other Classifiers considered

There are many popular classifiers that can be used for developing machine learning models. Some of the most popular ones, like Artificial Neural Networks and Support Vector Machines, give excellent performance in terms of accuracy but the decision making process of the models developed are not readily comprehensible by an engineer. For this reason, other classifiers were not developed in this work.

3. EXPERIMENTS AND RESULTS

All experiments for the machine learning based data analysis were performed in Weka a java based data analysis tool developed at the University of Waikato (Weka, 2015). The experiments can be divided into four stages. The first stage of experiments, Feature Selection and Extraction, chose the best-suited features and tested the potential of new features developed by combining existing ones.

The second stage of experiments, National Model Analysis, developed models using data from all the states except for one. The model was then tested on the data of the state that was not used for training. These experiments tested if it was possible to develop a single unified model using J48 Decision Tree and Bayesian Network algorithms, along with 10-fold cross validation for predicting design type of all bridges of the US. The J48 algorithm implementation in Weka was run with a confidence factor of 0.25, minimum 2 objects per leaf and 3 folds. The Bayesian Network algorithm implemented in Weka used the K2 algorithm for searching network structures and the Simple Estimator for developing the conditional probability tables.

The third stage, Individual State Model Analysis, developed machine learning models for each state using only the data for that state. This analysis developed more fine grain models that captured the variations in bridge design choices from factors like climatic conditions, seismic effects and design code requirements. J48 Decision Tree and Bayesian Network were used with 10-fold cross validation for developing and validating models. To account for seismic hazard, seismic data from USGS was added to the dataset as an additional attribute and its impact on the accuracy of the models was examined.

3.1 Feature Selection and Feature Extraction

After the first step of initial feature selection by removing irrelevant attributes from the dataset, 17 attributes were left. Before the next stage of attribute reduction was performed, one additional attribute was extracted from the existing attributes: average span length. After this additional attribute was added, chi squared attribute selection was performed. Table 1 shows the 6 best attributes as per chi squared attribute selection scores for four representative states. In Table 1, the relative drops in chi-squared scores of the next attribute relative to the current one are shown by the symbol Δ. It was noticed that there was almost always a huge decrease in chi-squared evaluation score after the first attribute. For evaluating attribute suitability, the second attribute which had a value of Δ greater than 30% in chi-squared score was found out. All attributes with scores lesser than that of this attribute were not considered for developing machine learning models. Among the attributes, it was noted that there was a strong correlation between total length and maximum span length and total length and average span length. Next, the second round of feature selection was carried out by both the approaches outlined in Section 2.2. The first approach, the leave-one-attribute-out analysis, reduces the number of features from 17 to 4. Removal of any one of these remaining four attributes results in significant drop of over 1.5% for the Bayesian Network model. These four attributes were maximum span length, average span length, deck structure type and material type. Only Bayesian Network model is considered since it is less sensitive to noise and change in parameters than Decision Tree and hence is more robust. Since, the four attributes from chi-squared feature evaluation and the leave-out-attribute-out analysis matched very closely, only these were considered for developing the models. Note that for Kansas, Deck Structure Type was considered in the final attributes for machine learning since its removal affected the model accuracy significantly.

Table 1. List of Best 6 Attributes Based on Chi-Squared Feature Evaluation for Georgia, California and Texas

Kansas	**Georgia**	**California**	**Texas**
Average Span Length (Δ = 49.19%)	Average Span Length (Δ = 68.12%)	Max. Span Length (Δ = 36.26%)	Average Span Length (Δ = 81.56%)
Material Type (Δ = 15.51%)	Material Type (Δ = 1.46%)	Avg. Span Length (Δ = 13.36%)	Maximum Span Length (Δ = 25.13%)
Maximum Span Length (Δ = 38.57%)	Deck Structure Type (Δ = 2.47%)	Deck Structure Type (Δ = 0.56%)	Material Type (Δ = 24.28%)
Total Length (Δ = 33.97%)	Max. Span Length (Δ = 52.72%)	Material Type (Δ = 35.8%)	Total Length (Δ = 29.75%)

Deck Structure Type	Total Length	Navigational Vertical Clearance	Deck Structure Type
Number of Spans	Navigational Vertical Clearance	Total Length	Number of Spans

3.2 National Model Analysis

The accuracy of the national models (Table 2) on an average was not very high and hence could not be considered useful as a recommendation tool for engineers. The accuracies marked in bold and underlined are the best and worst accuracies using the respective machine learning algorithm. While the Bayesian Network models show a fair amount of robustness in the results with results being somewhat consistent, the Decision Tree models show a significant variance in the accuracy of models. In fact, the best accuracy using the Decision Tree algorithm is, surprisingly, for the state of California, a state which later consistently shows poor performance with both Decision Tree and Bayesian Network methods and different, but more relevant, training sets. The reason for this discrepancy in model performance could not be determined, but appears to be a statistical anomaly caused by the configuration of the developed Decision Tree. The high variance in model accuracy using the Decision Tree illustrates the noise sensitivity of this algorithm. Bayesian Networks, being less sensitive to noise show slightly more robust results. There are several possible reasons for the low accuracy of the models. There are significant variations in the design code from state to state, introducing inherent differences in design choices. Hence, when the machine learning algorithm is developed, instances with similar values of attributes but different design types result in a model with lower accuracy. This led to the next stage of experiments, where a separate model for each state was developed in order to reduce conflicting design choices resulting from differences in state codes.

Table 2. Accuracy of national models

Attributes used	Material type, Maximum span length, Average span length, Deck structure type				
State	**Decision Tree**	**Bayesian Network**	**State**	**Decision Tree**	**Bayesian Network**
CA	**88.4**	**44.7**	FL	75.8	78
GA	88	80.9	IL	72.4	81.9
MN	86.5	**85.1**	ME	76.5	74.8
MS	78.4	84	NV	70.3	75.1
OR	**39.7**	75.5	PA	54	75.7
WA	50.1	46	VA	80	74.9

Table 3. Accuracy of individual state models

Attributes used	Material type, Maximum span length, Average span length, Deck structure type				
State	**Decision Tree**	**Bayesian Network**	**State**	**Decision Tree**	**Bayesian Network**
CA	79.5	74.8	FL	70	87.1
GA	95.4	**96.3**	IL	91	92.2
MN	**95.5**	93.1	ME	84.6	81.7
MS	95.2	93.1	NV	79	75.8
OR	77.8	67	PA	70.1	69.9
WA	**67.8**	**60.2**	VA	85	88.2

3.3 Individual State Model Analysis

Individual state models were first developed using only the NBI data. Seismic data was later integrated to account for seismic conditions. Decision Trees (J48) and Bayesian Network were used to develop all models. The parameters used for Decision Trees and Bayesian Network were the same as mentioned previously in Section 3.

(1) Models using the NBI dataset only

Table 3 tabulates the classification accuracy of the J48 and BN models for a few representative states, and illustrates the variations in machine learning accuracy for each state. The best accuracy is for the state of Georgia with BN algorithm while DT gives best accuracy for Minnesota. Decreased accuracy can be seen for the seismically active states of Washington, Oregon and California. While seismic activity does decrease accuracy, it is not the only cause of loss of accuracy as can be seen from the low accuracy for Pennsylvania, Idaho and New Jersey (not shown in table). The values highlighted in bold and underline in the table are the maximum and minimum accuracies of the model.

(2) Individual state models with seismic data

Results from the experiments performed with the seismic data as an attribute are tabulated in Table 4. The results show an average improvement in model accuracy of 1.2% compared to the analysis of only NBI data. Although most states show a minor increment in model accuracy, some states like Idaho show a small decrease in model accuracy (0.3%). In a few states, the increment is significant. The state of Florida shows an improvement of 10.45% in model accuracy whereas for South Carolina, model accuracy improves by 5.05%. In the table, the values highlighted in bold and underline are the maximum and minimum accuracies of the model.

Table 4: Accuracy of individual state models including USGS seismic data

Attributes used	Maximum span length, Average span length, Deck structure type, Material Type, Seismic activity				
State	**Decision Tree**	**Bayesian Network**	**State**	**Decision Tree**	**Bayesian Network**
CA	79.9	75.3	FL	89.7	88.3
GA	**<u>96.6</u>**	**<u>96.4</u>**	IL	91.1	92.3
MN	95.4	93.3	ME	84.6	81.7
MS	95.7	94	NV	80	77.1
OR	79.3	70.3	PA	74.9	72.1
VA	88.4	88.3	WA	**<u>70.2</u>**	**<u>64.4</u>**

4. CONCLUSION AND FUTURE WORK

The average accuracy of the individual state machine learning models with seismic data was 82.9%, and indicated a significant increase in performance compared to the national models (71.7%). Upon the inclusion of USGS seismic data as an attribute, the predictive accuracy of the models improved slightly more and an average accuracy of 84.6% was obtained. The best model prediction accuracy was 96.6% for the state of Georgia using the Decision Tree algorithm and the worst prediction accuracy was 64.4% for the state of Washington using Bayesian Networks.

There are several causes for error in the model development and evaluation stage. The dataset itself can have missing entries in addition to possibly wrong entries. A bridge with a Stringer/Multibeam design could have been mistakenly classified as some other type. The models developed from data containing this type of errors would be imperfect and can misclassify other bridges. In addition to this, it should be noted that the Decision Tree algorithm is noise sensitive and minor variations in the input data, possibly due to errors in the dataset, can result in variations in its prediction. This was evidenced in the Florida state data without seismic data. The model accuracy, shown in Table 3, is 70% whereas the same state has an accuracy of 89.7% after adding seismic data. Additionally, the accuracy of the corresponding Bayesian Network models is 87.1% and 88.3%, which shows that Florida is expected to have model accuracy close to 87% with no seismic data. No other state has such a drastic improvement in accuracy, which leads to authors to conclude that the poor result was a statistical anomaly. A key limitation of this study is that machine learning algorithms perform better when they have more instances to learn from. Hence, for bridges with relatively less common design types, such as suspension bridges and bascule bridges, there are higher chances that the machine learning model is not able to learn how to predict these bridge prototypes effectively.

The next stage of work aims to use the design prototype, which is the output from the machine learning algorithm, as input for the next stage: the optimization phase. The optimization space of a structure is non-convex with numerous local minima. To perform a search of a more global nature, the authors intend to implement methods that sample the search space at multiple points, such as evolutionary algorithms and particle swarm methods. Using the structure prototype from the machine learning algorithm, multiple bridge designs would be developed. These designs would then be used to seed the optimization algorithm and come up with an optimized structural design.

REFERENCES

Abdulhai, B., Pringle, R., Karakoulas, G.J. (2003). Reinforcement Learing for true Adaptive Traffic Signal Control, *Journal of Transportation Engineering*, 129 (3), 278-285.

Adeli, H. (2001). Neural Networks in Civil Engineering: 1989-2000, *Computer-Aided Civil and Infrastructure Engineering*, 16 (2), 126-142.

Bellazzi, R., Zupan, B. (2008). Predictive data mining in clinical medicine: Current issues and guidelines. *International Journal of Medical Informatics*, 77 (2), 81-97.
Cheng, J., Greiner, R. (2001). Learning Bayesian Belief Network Classifiers: Algorithms and System. *Advances in Artificial Intelligence*, 2056, 141-151.
Chien, S.I.J., Ding, Y., Wei, C. (2002). Dynamic Bus Arrival Time Prediction with Artificial Neural Networks, *Journal of Transportation Engineering*, 128 (5), 429-438.
Datumbox. (2015). *Using Feature Selection Methods in Text Classification*. Retrieved from http://blog.datumbox.com/using-feature-selection-methods-in-text-classification/, accessed on 10 October 2015.
FHWA. (2015a). *2014 NBI ASCII Files*. Retrieved from FHWA website: http://www.fhwa.dot.gov/bridge/nbi/ascii.cfm?year=2014#del, accessed on 12 February 2015.
FHWA. (2015b). *Recording and coding guide for the Structure Inventory and Appraisal of the Nation's Bridges*. Retrieved from FHWA website: http://www.fhwa.dot.gov/bridge/mtguide.pdf, accessed on 12 February 2015.
Hajela, P., Lee, E. (1997). Topological Optimization of Rotorcraft Subfloor Structures for Crashworthiness Considerations, *Computers and Structures*, 64 (1-4), 65-76.
Harper, P. R. (2005). A review and comparison of classification algorithms for medical decision making, *Health Policy*, 71 (3), 315–331.
Hung, S.L., Jan, J.C. (1999). MS_CMAC neural network learning model in structural engineering, *Journal of Computing in Civil Engineering*, 13 (1), 1-11.
Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23 (1), pp. 89-109.
Lavrac, N. (1999). Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine*, 16 (1), 3-23.
LRG. (2015). *Lattanzi Research Group*. http://lrg.gmu.edu, accessed on 5 October 2015.
Mitchell T. M. (1997). Machine Learning (1st ed.). The McGraw-Hill Company.
Potter, R. (2007). Comparison of classification algorithms applied to breast cancer diagnosis and prognosis, advances in data mining, *7th Industrial Conference, ICDM 2007, Leipzig, Germany*, 40–49.
Quinlan J.R. (1986). Induction of Decision Trees, *Machine Learning*, 1 (1), 81-106.
USGS. (2015). *NSHM 2014 Gridded Earthquake Data*. Retrieved from USGS website: http://earthquake.usgs.gov/hazards/products/conterminous/2014/data/2014_pga2pct50yrs.dat.zip, accessed on 17 April 2015.
Weka. (2015). *Weka 3: Data Mining Software in Java*. Retrieved from http://www.cs.waikato.ac.nz/ml/weka/, accessed on 10 January 2015.
Witten, I. H., Frank, E., Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques (3rd ed.). Morgan Kaufmann.
Worden, K., Dulieu-Barton, J.M. (2004). An Overview of Intelligent Fault Detection in Systems and Structures, *Structural Health Monitoring*, 3 (1), 85-98.
Worden, K., Manson, G. (2007). The application of machine learning to structural health monitoring, *Philosophical Transactions of the Royal Society of London: Mathematical, Physical and Engineering Sciences*, 365 (1851), 515-537.