IDENTIFYING BRIDGE PERFORMANCE PATTERNS IN A BRIDGE INVENTORY DATABASE: AN ANALYTICAL INVESTIGATION

Ryan T. Schnalzer[1], Mahmoud Reda Taha[2], Molly McCuskey[1],

Stephen C. Quintana[3] and Jimmy Camp[4]

ABSTRACT

Optimal choices for new bridge designs and existing bridge maintenance strategies necessitate understanding how bridge attributes and their interactions affect the performance of bridges in different environments. Analysis of a bridge inventory database will make it possible to recognize patterns of performance and how these patterns relate to different bridge attributes. One major challenge in performing this analysis is to determine a set of bridge performance features with minimum uncertainties and consistent records in the database. This paper describes the analytical investigations performed in searching for bridge performance features and their patterns as observed from the bridge inventory database of the State of New Mexico, USA. The use of hierarchical clustering made it possible to classify the data while also creating a rule-base scheme to come to findings. This analysis showed certain material structural types to be prevalent in high or low performing groups.

KEY WORDS

Feature extractions, Hierarchical clustering, Bridge performance.

INTRODUCTION

Management of infrastructure is a complex and multi-criteria based problem. Worldwide reports are reporting on an upcoming wave of limited infrastructure efficiency and limited available resources. Moreover, infrastructure systems have to surpass their design service lives and performance expectations (Flintsch and Chen 2004). As such, there is an increasing need to employ means of artificial intelligence for modeling such complex information environments. Therefore, there is an increasing need for developing a comprehensive strategy for efficient infrastructure maintenance and for efficient decision-making. Meanwhile, many researchers demonstrated the value in using of soft-computing methods to handle data mining and inference for infrastructure management (Hsieh and Liu 2004,

[1] Undergraduate Student, Department of Civil Engineering, University of New Mexico, Albuquerque NM, 87131, Phone +1 505/277-1258, FAX 505/277-1988, ryan7@unm.edu

[2] Assistant Professor, Department of Civil Engineering, University of New Mexico, Albuquerque NM, 87131, Phone +1 505/277-1258, FAX 505/277-1988, mrtaha@unm.edu

[3] New Mexico Department of Transportation, Santa Fe, NM, 87504-1119 Phone +1 505/285-3232, FAX 505/ 285-3251, Steve.Quintana@state.nm.us

[4] New Mexico Department of Transportation, Santa Fe, NM, 87504-1119 Phone +1 505/827-5532, FAX 505/827-5345, jimmy.camp@nmshtd.state.nm.us

Flintsch and Chen 2004, Romão et al. 2004). The advantage of soft computating methods over classical methods is their ability to effectively address issues in data mining and such as imprecision, uncertainty of subjective values and in being capable of establishing sophisticated inference rule-bases that handle large number of data sets.

Different types of soft computing methods have been used in the literature to deal with the above issues in infrastructure management research. Such methods included artificial neural networks (ANN), fuzzy logic (FL) and genetic algorithms (GA), for example, were used effectively to recognize data patterns and detect relationships between bridge condition rating data and bridge parameters (Cattan and Muhammadi 1997). ANN were capable of predicting the subjective ratings based on given bridge parameters. A major drawback with ANN is its functionality as a "black box" and therefore is not useful for system inference. Fuzzy logic and fuzzy set theory were applied to numerous infrastructure research projects but has been predominant in pavement infrastructure research (Wang and Liu 1997). In particular, membership functions were used due to their ability to deal with subjectivity and partial truth. The use of fuzzy sets allowed incorporating subjective descriptors such as "poor", "good", and "excellent" (Elton and Juang 1988). Fuzzy systems were used to create a universal pavement distress evaluator (Shoukry et al. 1997) and a comprehensive index for flexible pavements (Zhang et al. 1993). Finally, in a slightly different approach, fuzzy systems were used to help interpreting image processing techniques for pavement distress evaluation (Chou et al. 1995).

The objective of this research was to establish an intelligent rule-base that can classify and pattern bridges based on their structural performance and identify what characteristics most dictate their level of performance. A case study using the New Mexico bridge inventory database is demonstrated. The subjectivity of existing performance parameters and indicators and the high dimensionality of the database made this research further challenging. We employed fuzzy set theory to establish a comprehensive structural performance index. We then demonstrate the possible use of statistical methods and hierarchical clustering methods to identify bridge performance patterns. The rest of the paper is structured as follows. First we describe the methods used to establish a sensitive performance index and the methods used for extracting the bridge performance patterns. We then demonstrate the bridge performance patterns identified. We provide representative findings and discuss how these findings can be used in future decision-making to improve overall bridge performance.

METHODS

Here we present the hierarchical clustering technique employed to identify the bridge performance patterns. We start by discussing some of the fundamental processes that were necessary prior to clustering the datasets.

DATA PRE-PROCESSING

The New Mexico National Bridge Inventory (NBI) database includes data describing approximately 4,000 bridges and culverts in New Mexico. Each bridge or culvert is described using 132 different parameters/attributes in addition to performance indices. Exemplar bridge attributes include bridge identification values, average daily traffic, detour length, structure kind, structure type, clearance measurements, structure length and many more. Data preprocessing was essential to enable computational development. The pre-processing of the data incorporated removing un-coded parameters, conversion of non-numeric data to numerical values, and data normalization. The normalization process reformatted the data to range between 0 and 1 to prevent parameter distortion that can result from numerical weights (Sutton and Reggia 1994).

PARAMETER CLASSIFICATION

The second step was to classify all bridge attributes to a few intuitive parametric categories that might help in understanding the complex relationship that relate all these parameters to bridge performance. Five parametric categories are identified:
1) *Bridge Parameters*: This category includes all physical attributes of the bridge such as geometry, materials and other characteristics used which may affect the bridge performance.
2) *Loading Parameters*: This category includes all parameters describing the nature, type, and intensity of loading on the bridge. Such parameters include Average Daily Traffic (ADT) and percentage of Trucks in Average Daily Traffic.
3) *Identification Parameters*: This category includes all parameters used to index the bridge by number or location. These parameters are very important for the database operation but might have less significance on the bridge performance.
4) *Indirect Parameters*: This category includes miscellaneous parameters that might have indirect effect on the bridge performance. Such parameters include relative humidity exposure from nearby waterways, bridge elevations (weather patterns), and other issues that do not directly relate with bridge characteristics or loading, but might have an affect on the bridge performance.

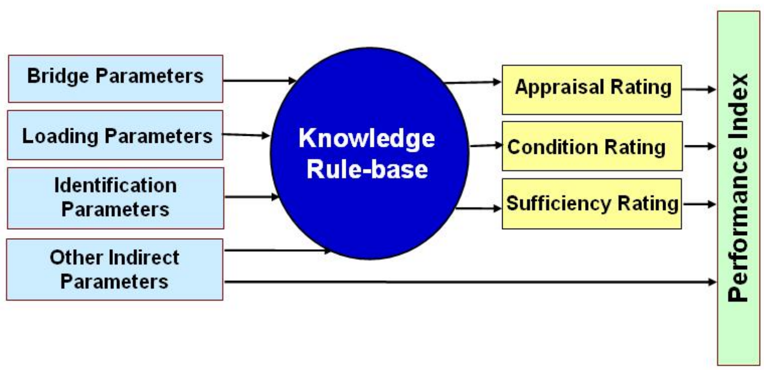


Figure 1: Schematic representation of parameter categories and bridge knowledge rule-base.

5) *Performance Parameters*: We consider this category of parameters that represents the bridge performance or in modeling terms what is known as the knowledge rule-base that relates all bridge attributes to bridge performance. The performance parameters category includes all condition ratings of bridge deck, super and sub structures, appraisal rating and sufficiency rating as well. Condition ratings are defined by the US Federal Highway Administration (FHWA) (1995). Figure 1 illustrates the proposed relationship between the five parametric categories.

ESTABLISHING A NEW PERFORMANCE INDEX

It is evident that the system shown in Figure 1 cannot be established unless a comprehensive performance metric that can accurately represent the bridge performance exists. While the FHWA uses the sufficiency rating (SR) for decision making, the SR does not accurately represent structural performance but represents a decision making index. This is because the SR does not only depend on structural performance, but on other parameters such as serviceability and functional obsolescence, level of importance and essentiality for public use. Therefore, SR might not be a good performance metric to understand the relationship between bridge performance and bridge attributes. Another alternative is to use the structural adequacy (SA), which is defined by FHWA (1995) as a satisfactory structural performance metric. However, the drawback of SA for being a structural performance metric is its lack of representation of bridge deck condition. It is well known that bridge deck conditions usually represent critical elements in bridge deck performance and therefore bridge deck conditions shall be included in any comprehensive metric for bridge performance.

Therefore, we suggest developing a comprehensive performance index denoted (PI) and is designed to represent the combined effect of the structural adequacy, the bridge deck condition and the structural evaluation. The new performance index was developed using fuzzy-set theory to incorporate the uncertainty in the bridge evaluation process for its human dependence and to account for the vagueness and ambiguity in the definitions describing many of the modeling parameters. The new PI is modeled to integrate three parameters from the inventory database including the structural adequacy (SA), the deck condition rating, and the structural evaluation. A group of fuzzy sets were first defined over the modeling domains of these three parameters. A rule-base that relates the three modeling parameters to the proposed performance index (PI) was then established using Mamdani inference. A centroid defuzzification process was used thereafter to aggregate the output of the fuzzy rules.

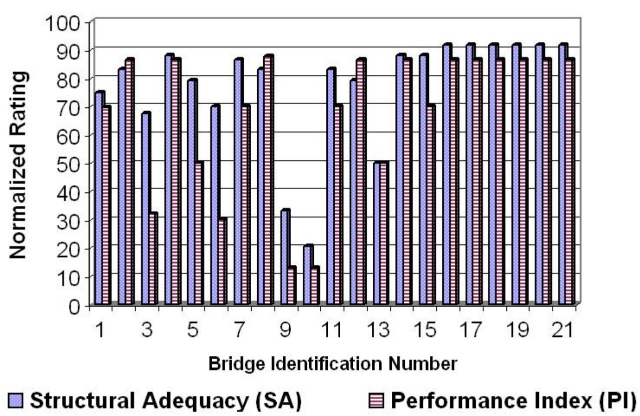


Figure 2: Comparison between SA and the new performance metric (PI)

The new performance metric (PI) ranges between zero and 100 with zero indicating poor performance and 100 indicating excellent performance. Similar metrics have been developed in the area of infrastructure management for providing a meaningful solution to complex modeling of pavement damage assessment (Shoukry et al, 1997). A pictorial comparison between the new performance metric and the structural adequacy for a few bridges from the database is shown in Figure 2. It is evident from Figure 2 that the fuzzy-based performance index agrees with the structural adequacy metric defined in the bridge inventory database.

PRELIMINARY STATISTICAL ANALYSIS

Once the comprehensive performance metric (PI) was established, a preliminary statistical analysis was performed to explore the database. The statistical analysis was conducted on random datasets of the ten different material types, and the thirteen different structure types. Each random set of data was randomly selected using the bootstrapping technique that is based on Monte Carlo simulation (Martinez and Martinez 2002). Preliminary statistical analysis helped gaining insight into the composition of the database. The preliminary analysis was followed by performing hierarchical clustering, which serves as a multidimensional classification system to group similar datasets based on a multitude of attributes. The statistical analysis was performed on datasets considering two performance criteria: the structural adequacy (SA) and the fuzzy-based performance index (PI). To reveal deterioration rates for several material types, we studied the change in structural adequacy over time by quantifying the loss in structural adequacy (LSA) for different types of materials. LSA is defined as

$$LSA = SA_{max} - SA \qquad (1)$$

where SA is the structural adequacy defined by FHWA (1995) that has a maximum value of SA_{max} reported by FHWA (1995) to be 55 and LSA is the loss in structural adequacy. Based on Equation (1), LSA ranges between 0 and 55 with zero representing a very adequate bridge and 55 representing a very poor bridge.

HIERARCHICAL CLUSTERING

Clustering techniques are used for performance pattern recognition (Duda et al. 2001). As an unsupervised method, clustering can provide an efficient and computationally inexpensive technique for feature extraction with the advantage of being reversible for inference. Moreover, hierarchical clustering furthers this analysis by sub-clustering each cluster to create a tree-like structure that summarizes the hierarchy format. This classification process not only classifies data in steps, but also creates a path that can be followed to establish a knowledge rule-base that describes the complex relationships between bridge attributes and bridge performance patterns. The objective of the hierarchical clustering was to find the commonalities between the different bridges. Common datasets were sub-clustered into sub-sets with common attributes and so on. This allowed inferring which bridge attributes helped to enhance bridge performance and which attributes might have resulted in a reduced bridge performance. Thus the hierarchical clustering established a hierarchical tree structure that allowed establishing a knowledge rule-base that relates bridge attributes to bridge performance patterns. Figure 3, shows an example case of the hierarchical clustering to infer bridge performance patterns.

K-means clustering method was used to provide the main and sub clusters in each cycle of analysis. K-means clustering has the advantage of quick convergence and being computationally inexpensive (Duda et al. 2001). This method is also referred to as nearest neighborhood clustering. For a sample set of n data samples with m features such that

$$X = \{x_1, x_2, x_3, x_4 \cdots x_n\} \tag{2}$$

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, x_{i4} \cdots x_{im}\} \tag{3}$$

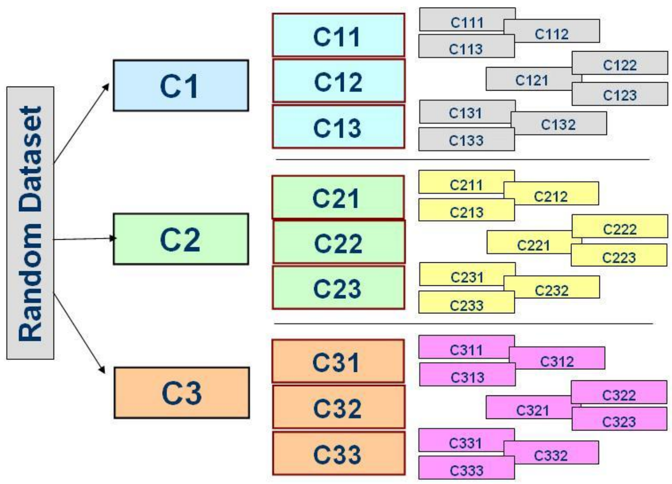


Figure 3: Tree structure based on hierarchical clustering

The K-means technique creates k clusters such that $2 < k < n$ by minimizing an objective function J based on the Euclidean distance between the data samples x_i in the cluster and clusters centers v_j. The objective function is thus defined as

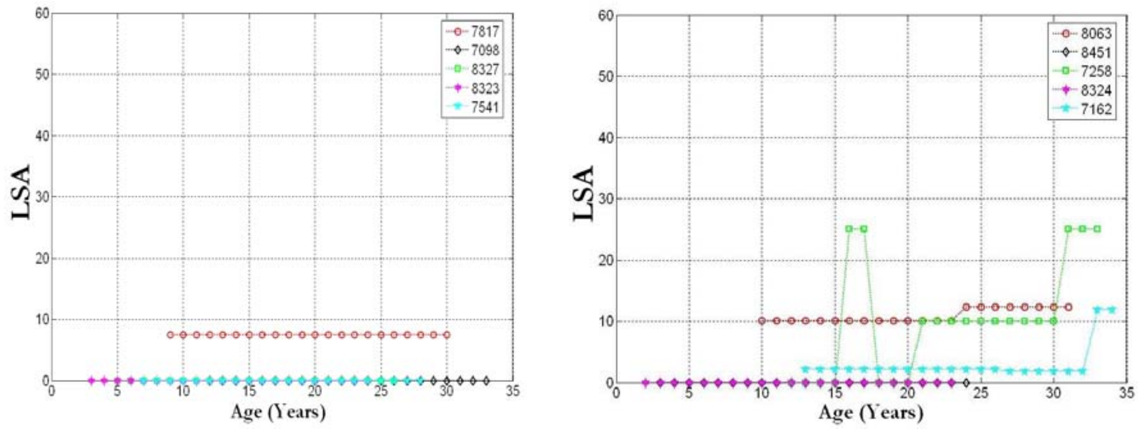
$$J(U,v) = \sum_{j=1}^{n} \sum_{i=1}^{k} \chi_{ij} \left(d_{ij}\right)^2 \tag{4}$$

Where d_{ij} is the Euclidean distance measure of m dimensional feature space between the j^{th} data sample x_{ij} and the i^{th} cluster and $\chi_{Ai}(x_j)$ is a characteristic function deciding on the belonging of the data sample x_j to the i^{th} cluster A_i as

$$\chi_{A_i}\left(x_j\right) = \begin{cases} 1, & x_j \in A_i \\ 0, & x_j \notin A_i \end{cases} \tag{5}$$

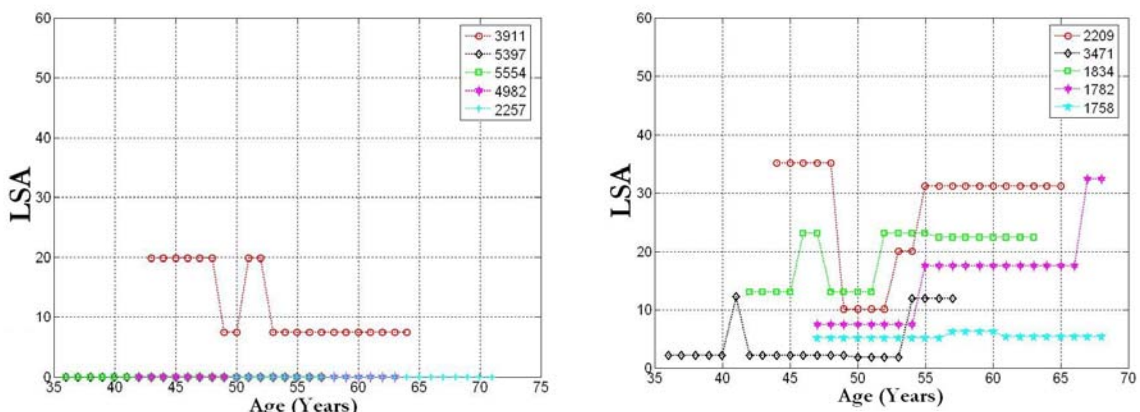
RESULTS AND DISCUSSION

While both the statistical analysis and the hierarchal clustering were performed on datasets considering two performance criteria: the structural adequacy (SA) and the fuzzy-based performance index (PI), results considering SA are only shown here for space limitations. The preliminary statistical analysis showed that the rate of performance deteriorations is much dependent on the material type. Figure 4 shows an exemplar case for young bridges (younger than 35 years of age). Figure 4 presents the observation that concrete bridges tend to have high performance (low LSA) and its performance is steady for long period of time while steel bridges tend to have performances lower than that of concrete (relatively high LSA) and tend to observe abrupt changes in performance. It was usually observed that steel bridges seemed to start showing minor deficiency after 15-20 year of service life.



[a] [b]

Figure 4: Age versus loss in structural adequacy (LSA) for young bridges (younger than 35 years of age) [a] concrete [b] steel. Legend identifies bridge numbers



[a] [b]

Figure 5: Age versus loss in structural adequacy (LSA) for old bridges (older than 35 years of age) [a] concrete [b] steel. Legend identifies bridge numbers

The previous observation is also evident for old bridges (35 years of age and older). Figure 5 shows the change in the LSA with age for old concrete and steel bridges. While only one deteriorating bridge can be observed in the concrete sample, a few deteriorating steel bridges can be observed. Similar observations were recorded during the analysis of most randomized datasets. Another interesting observation was the existence of 45 years of age mark for many concrete bridges where its deterioration rates start to increase. While Figures 4 and 5 show good examples for the effect of age on deterioration rates, they also serve as randomized samples to illustrate the effect of coding subjectivity in the database. Reverse slopes represent repair effects and subjective evaluation effects as well. While the information presented represents a general idea on bridge deterioration rates, they also prove the existence of a considerable level of uncertainty in the database records that might not be possible to handle using classical methods. An example of hierarchical clustering is presented in Figure 6. We demonstrate an example of the hierarchical path that can be followed to isolate bridge clusters with similar characteristics. This path allows establishing a knowledge rule-base that relates bridge performance indices to bridge attributes. The hierarchical path can be better understood with the aid of Figure 3 as cluster being divided into k sub-clusters in the m dimensional feature space. Figure 6 [a] presents the three major clusters that can be found in a set of bridges randomly selected from the inventory database. The major clustering feature here is age.

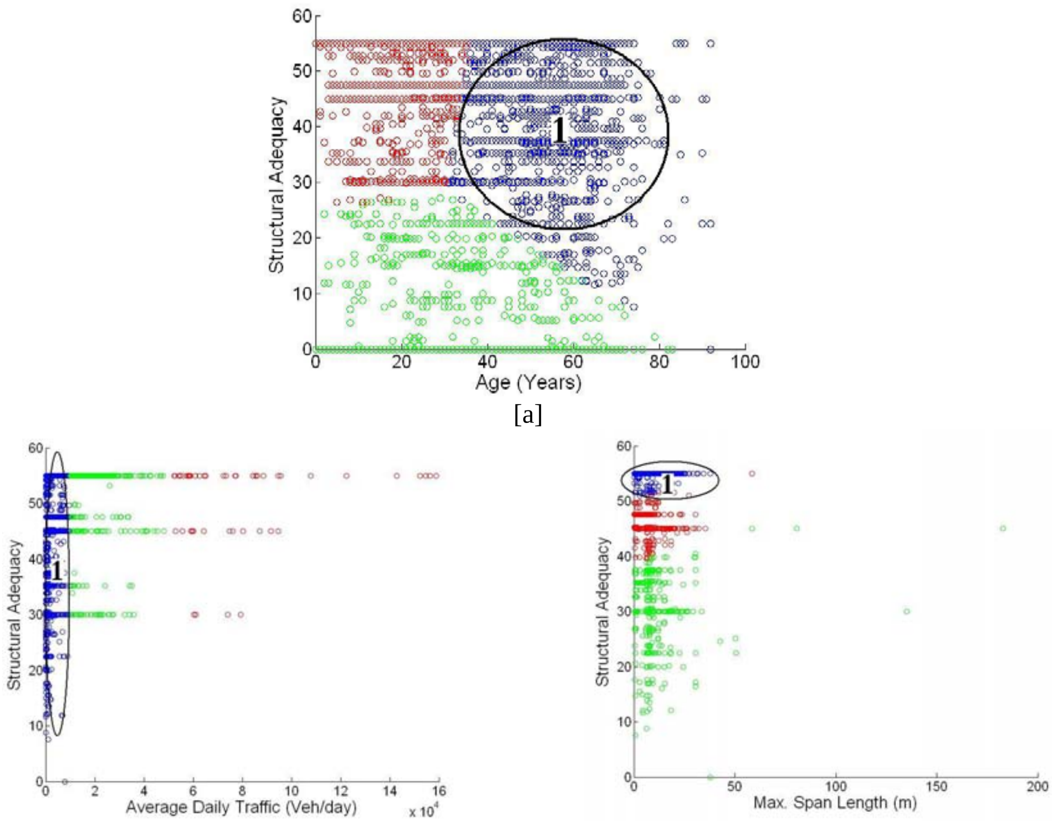
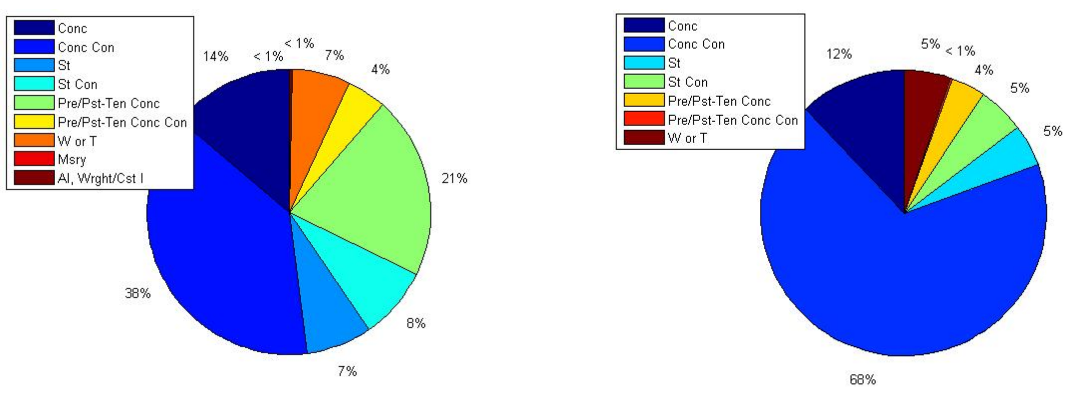


Figure 6: Hierarchical k-means clustering plots for a group of randomly selected bridges
[a] Cluster C1, C2 and C3, [b] Sub-clusters C11, C12 and C13
[c] Sub-clusters C111, C112 and C113

The next step in the hierarchical process is to isolate cluster 1, denoted C1, which in this case represents old bridges with good performance (bridges older than 40 years of age and with high structural adequacy). The data samples in cluster 1 are then sub-clustered with respect to the second feature space being the average daily traffic (ADT). Figure 6 [b] shows three new sub-clusters: C11, C12 and C13. By observing sub-cluster C11, this sub-cluster represents a group of bridges that are characterized by having a high structural adequacy, older than 40 years of age, and with low ADT. We go one step further and sub-cluster this group of bridges with respect to maximum span length. Figure 6 [c] shows three more clusters with the sub-cluster C111 representing bridges that are have high structural adequacy, older than 40 years of age, low ADT and short spans.

Analysis of these clusters and sub-clusters can reveal the fundamental bridge characteristics that might have helped in such performance. For example, analysis of material type distributions within the major clusters and sub clusters can allow us to understand if specific types of materials dictated low or high performances of bridges. Figure 7 illustrates the material type distribution within cluster C1 and sub-cluster C111. It is evident that concrete and continuous concrete bridges represent the majority of bridges in the randomized set. This was found to be true for all randomized sets in the analysis. This fact is represented in the composition of cluster C1 (Figure 7[a]) with concrete bridge forming (52%) of the bridges in the random set. Observing the sub-cluster C111, concrete bridges 80% of this sub-cluster. This means other types of materials have dropped from this cluster. Considering the fact that C111 represents relatively old bridges with high structural adequacy, low ADT and short span bridges, it becomes evident that concrete contributes to the structural performance of this class of bridges. The analysis was repeated for six times on randomized datasets to guarantee coming to general conclusions.



[a] [b]

Figure 7: Material type distribution [a] Cluster C1 [b] Sub-cluster C111

CONCLUSIONS

We demonstrated here the possible use of soft computing and clustering techniques to analyze a high dimensional bridge inventory database to reveal bridge performance patterns. A new comprehensive bridge performance index is developed and is capable of representing all the bridge performance criteria. K-means clustering was employed in a hierarchical scheme to identify the major bridge performance patterns and establish a tree-structure that allows building a knowledge rule-base that can relate the different bridge attributes to the performance criteria and their patterns. We demonstrated an exemplar case showing the identification of bridge attributes and their relation to performance using the hierarchical clustering technique. The hierarchical clustering process was performed on six randomized sets for both the structural adequacy and the fuzzy-based performance index (PI). The investigations found that material and structural types have significant influences on bridge performance. While concrete bridges showed better performance than steel ones, multi-girder bridges typically showed poor performance especially with high average daily traffic.

ACKNOWLEDGMENT

This research work is funded by New Mexico Department of Transportation. The authors greatly acknowledge this support. The special support to the first and third authors by the Student Research Associate (SRA) program as part of the Defense Threat Reduction Agency (DTRA)-University of New Mexico partnership is greatly acknowledged.

REFERENCES

Cattan, J., and Muhammadi, J. (1997). “Analysis of bridge condition rating data using neural networks.” Microcomput. Civ. Eng., 12(6), 419-429.

Chou, J., O’Neill, W.A., and Cheng, H. (1995). “Pavement distress evaluation using fuzzy logic and moment invariants.” Trans. Res. Records. 1505, TRB, pp: 39-46.

Duda, R. O., Hart, P. E., Stork, D.G. (2001), Pattern Classification. 2nd Ed. J. Wiley & Sons.

Elton, D.J., and Juang, C.H. (1988). “Asphalt pavement evaluation using fuzzy sets.” Trans. Res. Rec. 1196, TRB, pp: 1-6.

Federal Highway Administration (FHWA) (1995) Recoding and Coding Guide for the Structure Inventory and Apprisal of the Nation’s Bridges.

Flintsch W. G., Chen, Chen. (2004). “Soft Computing Applications in Infrastructure Management.” ASCE J. of Infrastructure Systems, 10(4), 157-166

Hsieh, T-Y., and Liu, H-L., (2004), “Genetic Algorithm for Optimization of Infrastructure Investment Under Time-Resource Constraints”, Computer-Aided Civil and Infrastructure Engineering, 19 (3), pp. 203-212.

Martinez, W. L., Martinez, A.R. (2002). Computational Statistics Handbook with Matlab. Chapman & Hall, 2002.

Ross, T. J. (2004). Fuzzy Logic with Engineering Applications. 2nd Edition. J. Wiley & Sons.

Romão, W., Freitas, A. A., de S. Gimenes, I. M. (2004), “Discovering interesting knowledge from a science and technology database with a genetic algorithm”, Applied Soft Computing, 4 (2), pp: 121-137

Shoukry, S. N., Martenelli, D.R., and Reigle, J.A., (1997). “Universal pavement distress evaluator based on fuzzy sets.” Trans. Res. Rec. 1592, TRB, pp:180-186.

Sutton, G.G. and Reggia J.A., (1994), “Effects of normalization constraints on competitive learning”, IEEE Transactions on Neural Networks, 5 (3), pp. 502-504.

Wang, K. and Liu, F., (1997), “Fuzzy Set-Based and Performance-Oriented Pavement Network Optimization System”, ASCE J. of Infrastructure Systems, 3 (4) pp. 154-159

Zhang, Z., Singh, N., and Hudson, W. R. (1993). “Comprehensive ranking index for flexible pavement using fuzzy sets model.” Transp. Res. Rec. 1397, TRB, pp: 96-102.