

# Localized Prediction of Continuous Target Variables Using Hierarchical Clustering

Aleksandar Lazarevic<sup>1</sup>, Ramdev Kanapady<sup>2</sup>, Chandrika Kamath<sup>3</sup>, Vipin Kumar<sup>1</sup>, Kumar Tamma<sup>2</sup>

<sup>1</sup>Dept. of Computer Science, University of Minnesota, Minneapolis, MN 55455

<sup>2</sup>Dept. of Mechanical Engineering, University of Minnesota, Minneapolis MN 55455

<sup>3</sup>Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, CA 94551

aleks@cs.umn.edu, ramdev@cs.umn.edu, kamath2@llnl.gov, kumar@cs.umn.edu, ktamma@tc.umn.edu

## Abstract

*In this paper, we propose a novel technique for the efficient prediction of multiple continuous target variables from high-dimensional and heterogeneous data sets using a hierarchical clustering approach. The proposed approach consists of three phases applied recursively: partitioning, localization and prediction. In the partitioning step, similar target variables are grouped together by a clustering algorithm. In the localization step, a classification model is used to predict which group of target variables is of particular interest. If the identified group of target variables still contains a large number of target variables, the partitioning and localization steps are repeated recursively and the identified group is further split into subgroups with more similar target variables. When the number of target variables per identified subgroup is sufficiently small, the third step predicts target variables using localized prediction models built from only those data records that correspond to the particular subgroup. Experiments performed on the problem of damage prediction in complex mechanical structures indicate that our proposed hierarchical approach is computationally more efficient and more accurate than straightforward methods of predicting each target variable individually or simultaneously using global prediction models.*

## 1. Introduction

Many large-scale data analysis problems very often involve an investigation of high dimensional and heterogeneous databases, where different prediction models are responsible for different regions [7, 9]. In addition, these large data sets sometimes have extremely large numbers of target (output) variables (order of thousands or hundred of thousands of target variables) that need to be predicted and the task of predicting all of them may seem daunting. For instance, in a manufacturing process we may want to predict various quality aspects of a product from the parameter settings used in the manufacturing. In

financial markets, given econometric variables as predictors, the goal may be to predict changes in the valuations of the stocks in large number of industry groups [2]. These scenarios are often characterized with the presence of strong correlation among target variables, and incorporating this knowledge into learning process may produce more efficient and accurate prediction models. Grouping highly correlated target variables naturally fits into the general problem of learning multiple target variables, which can be described through three prediction levels: 1) is there any change among target variables; 2) which target variables have been changed; and 3) how much these target variables have changed.

In this paper, we propose a novel technique for efficient prediction of multiple continuous target variables from high-dimensional and heterogeneous data sets using a hierarchical clustering approach. Instead of predicting target variables individually or simultaneously using global prediction models, which are built considering all data records, the proposed approach consists of three phases applied recursively: (1) partitioning of data set (2) localization of groups of interesting target variables, and (3) the prediction of target variables. In the partitioning step, similar data records are first grouped using clustering algorithm and then corresponding groups of similar target variables are identified. In the localization step, a classification model is used to predict which group of the target variables needs to be further investigated. If the identified group of target variables still contains large number of target variables, then the partitioning and localization steps are repeated recursively and the identified group is further split into subgroups with more similar target variables. When the number of target variables per identified subgroup is sufficiently small, the prediction phase takes place and the target variables are predicted using localized prediction models that are built not using all data records from entire data set, but using only those data records that correspond to the particular subgroup.

Significant research work has been done in the area of predicting multiple target variables. Some of these approaches include multitask learning [3], learning to learn

[1], curds and whey algorithm [2], clustering learning tasks [15] and learning internal representations [6]. A key feature that distinguishes our work from these existing algorithms is that our approach learns a very large number of target variables by exploring problem-specific interrelationships among them. In addition, it decomposes the complex problem of predicting multiple target variables from the global data set into many simpler ones where particular target variables are predicted using only relevant set of data records as well as relevant features.

The proposed hierarchical approach has two main advantages over existing techniques for predicting large numbers of target variables from high-dimensional and heterogeneous databases. First, it is computationally more efficient, since it uses less number of training records as well as a fewer number of features. Second, the clusters obtained at each partitioning level correspond only to small subsets of similar target variables and tend to be more homogeneous than the entire data set. Therefore, when predicting target variables, more accurate prediction models can be constructed more efficiently by using only the local data from corresponding clusters instead of the entire global data set.

The effectiveness of the proposed approach is demonstrated by applying it to scientific simulation data sets (Finite Element Analysis) for damage prediction in complex mechanical structures. Our experiments performed on the large-scale complex civilian structures indicate that the proposed method is computationally more effective and more accurate than straightforward approaches of predicting each target variable individually or simultaneously using global prediction models. An extended version of this paper is available in [8].

## 2. Problem description

Given a high-dimensional heterogeneous data set  $D$  with a large number of continuous target variables (Table 1), the problem consists of effectively and accurately predicting real value of all target variables. The typical data layout used in predicting multiple target variables is shown in Table 1, where data set  $D$  contains data records  $d_1, d_2, \dots, d_N, i = 1, \dots, N$ . Each data record  $d_i$  is described with the pair  $\{\mathbf{f}, \mathbf{E}\}$ , where  $\mathbf{f} = \{f_1, f_2, \dots, f_m\}$  is the feature set, while  $\mathbf{E} = \{E_1, E_2, \dots, E_n\}$  is the set of target variables. Each data record  $d_i$  in the data set  $D$  pertains to a specific state depending on domain knowledge.

One of the problems that exhibit these characteristics corresponds to damage detection in complex mechanical structures. This phenomenon includes localized softening or cracks in a certain neighborhood of a structural component due to high operational loads, or the presence of flaws due to manufacturing defects. In general, there are three levels of damage identification: 1. *Recognition* -

qualitative indication that damage might be present in the structure; 2. *Localization* - information about the probable position of the damage in the structure 3. *Assessment* - estimate of the extent of severity of the damage in the structure. Structural damages in mechanical structures usually result in static deformations and changes in dynamic characteristics such as natural frequencies and the mode shapes. In our work, structural damage is assumed to be associated with structural stiffness: a reduction in Young's modulus or modulus of elasticity ( $E$ ) [14].

**Table 1. A typical input to data mining model for damage detection in mechanical structures.**

Data records	Features (Frequencies)				Target variables			
	$f_1$	$f_2$	$\dots$	$f_m$	$E_1$	$E_2$	$\dots$	$E_n$
$d_1$	72.833	151.67	$\dots$	213.45	0.5E	E	$\dots$	E
$d_2$	73.45	152.56	$\dots$	213.65	0.6E	E	$\dots$	E
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
$d_N$	74.01	153.01	$\dots$	214.21	E	E	$\dots$	0.7E

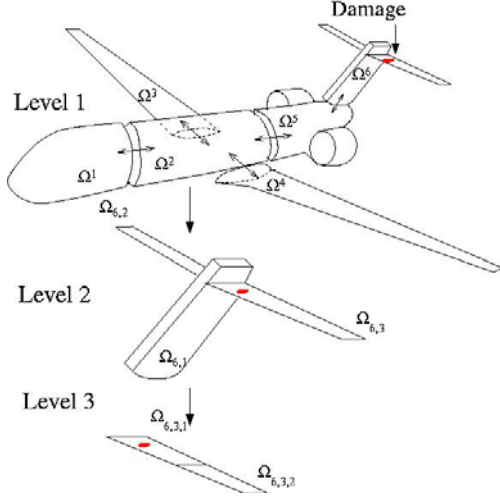
Since mechanical structures may be represented with a certain number of structural elements, the problem of predicting multiple target variables in damage detection in mechanical structures is reduced to predicting the intensity of the damage in structural elements  $\{E_1, E_2, \dots, E_n\}$  (Table 1) using dynamic characteristics (e.g., natural frequencies, mode shapes) as features  $\{f_1, f_2, \dots, f_m\}$  (Table 1).

Standard analytical techniques for this problem employ mathematical models to approximate the relationships between specific damage conditions and changes in the structural response or dynamic properties [4]. Although recent research work [12, 14] has shown that neural networks may provide a potential solution for damage prediction, these studies are restricted to very small models with a small number of target variables (order of ten).

## 3. Methodology

### 3.1. Hierarchical Partitioning

To partition target variables, there is a need to define similarity between them using domain specific knowledge. Although in this paper we focus on spatial domains that incorporate extra constrains, the proposed method is general and may be applied to any domain. In spatial domains, variables may be similar if they are spatially close to each other or if they are symmetric according to some symmetry axis. Therefore, the simplest method for partitioning target variables into similar groups is to perform *manual partitioning* of target variables by incorporating heuristics describing spatial locality of target variables (e.g. locations of target variables). For example, Figure 1 illustrates manual partitioning of the airplane structure.



**Figure 1. Manual partitioning of the airplane**

However, the *manual partitioning* approach is restricted to using only visual characteristics of mechanical structures, and many hidden characteristics inherent for the problem may not be available for sub-structuring. Therefore, our alternative approach groups similar target variables  $\{E_1, E_2, \dots, E_n\}$  using features  $\{f_1, f_2, \dots, f_m\}$  that characterize them, since the features usually indirectly incorporate the information about spatial locality and symmetry of target variables, and may incorporate any other similarities. For example, in damage detection problem, frequencies provide information about locality and symmetry of structural elements, but they also provide information about the correlation among target variables, which is not available in the *manual partitioning*. Therefore, by using particular set of features, it is possible to group similar target variables that are either close or symmetric in 3D space. In such a way, data records with similar set of features belong to one group of target variables, while data records that have different features from a specific set of features correspond to different groups. When the identified group needs to be further partitioned, different and/or additional set of features that bring additional knowledge needs to be considered (e.g., higher frequencies in damage detection problem) in order to distinguish among similar target variables already identified.

The proposed hierarchical clustering based approach for partitioning groups of target variables and predicting their values is presented in Figure 2. As input arguments, the proposed algorithm takes entire data set  $D$  (Table 1) and  $m_1$  features  $\{f_1, f_2, \dots, f_{m_1}\}$  considered at that level of partitioning. The first step in the proposed algorithm involves applying the clustering algorithm with similar specified features  $\{f_1, f_2, \dots, f_{m_1}\}$ . One of the biggest challenges of applying a clustering algorithm is to select right set of features  $\{f_1, f_2, \dots, f_{m_1}\}$  that will be used at particular level of clustering, but this choice depends on

specific domain and is typically determined using expert background knowledge. This grouping is performed using several clustering algorithms (k-way clustering algorithms with repeated bisections, direct k-way clustering, agglomerative clustering and a graph partitioning clustering approach) from the CLUTO clustering package [16]. Another challenge in clustering is to determine the optimal number of discovered clusters. The methodology that we employ here is to discover a relatively small number (between 3 and 10) of well-balanced clusters.

**Algorithm** *Partitioning*( $D, m_1$ )

- Given: set with data records  $D = \{d_1, d_2, \dots, d_N\}$ , where each  $d_i = \{f_1, \dots, f_m, E_1, \dots, E_n\}$ ,  $i = 1, \dots, N$ .
- Select the set of low natural frequencies  $\{f_1, f_2, \dots, f_{m_1}\}$ , where  $m_1 \ll m$ .
- Apply clustering algorithm on the following data set  $\{f_1, f_2, \dots, f_{m_1}\}$  and obtain  $k$  clusters  $C_1, C_2, \dots, C_k$ , where the criterion for obtaining optimal clusters was identification of well balanced clusters.
- Identify  $k$  substructures  $S_1, S_2, \dots, S_k$  such that they correspond to identified clusters  $C_1, C_2, \dots, C_k$ .
- For  $i = 1, \dots, k$ 
  - Predict the existence of damage in substructure  $S_i$  using classification model.
  - If there is the damage in the substructure  $S_i$ ,
    - If the substructure  $S_i$  is not sufficiently small:
      - Chose the set of low natural frequencies  $\{f_{m_1}, f_{m_1+1}, \dots, f_{m_2}\}$ , where  $m_1 < m_2 < m$ .
      - Call algorithm *Partitioning* ( $C_i, m_2$ ) to discover finer substructures in the substructure  $S_i$ .
      - Else (the substructure  $S_i$  is sufficiently small)
        - Apply localized regression model to predict the intensity of the damage for all the elements within the identified substructure with the damage.

**Figure 2. Partitioning algorithm for discovering groups of similar target variables, identifying the group of interest and predicting the values of target variables.**

*Identifying interesting groups of target variables.* When the clusters of data records with similar features are identified, the next step is to identify their corresponding groups of similar target variables (e.g., to identify the structures of similar structural elements in damage detection problem). Although obtained clusters usually do not contain contiguous data records, for simplicity assume that obtained clusters are presented in Figure 3.

Clusters  $C_i$ ,  $i = 1, \dots, k$  from Figure 3 contain particular data records as sets of particular features  $\{f_1, f_2, \dots, f_{m_1}\}$ . For example, cluster  $C_1$  contains data records  $d_{11}, \dots, d_{1, n_1}$ , where  $n_1$  is the number of data records in the cluster  $C_1$ . Each data record from a particular cluster  $C_i$  determines which particular target variable is set or which particular target variable is of particular interest. For instance, if data record  $d_{11}$  that belongs to the cluster  $C_1$  is

the result of damage in the structure element  $E_l$ , it is apparent that this structural element may belong to the corresponding group of elements (structure)  $S_l$ . Therefore, all the damaged structural elements that are result of data records from the cluster  $C_l$  should belong to the group of target variables (elements)  $S_l$ . However, sometimes this method will create situations when a specific target variable may belong to more than one group of target variables. In such scenario, there is a need to determine what is the most appropriate group of target variables that the particular target variable should belong to. For example, in Figure 3, the target variable  $E_l$  is set for two data records from the cluster  $C_l$  and for one data record from the cluster  $C_k$ , and it is not clear whether the target variable  $E_l$  should belong to the corresponding group  $S_l$  or to the group  $S_k$ . In order to determine the most appropriate group of target variables that the specific target variable should belong to, we considered three approaches:

- *Maximum value* – For each target variable  $E_i$ , identify the cluster  $C_f$  that contains the maximum value of target variable  $E_i$ . The considered target variable  $E_i$  then belongs to the group of target variables  $S_f$ .
- *Weighted majority* – For each target variable  $E_i$ , sum the values of specific target variables within each cluster and identify the cluster  $C_f$  with the largest computed value. The considered target variable  $E_i$  then belongs to the group of target variables  $S_f$ .
- *Overlapping* – For each target variable  $E_i$ , identify all clusters  $C_f$  for which the target variable  $E_i$  is set. The considered target variable  $E_i$  then belongs to all groups of target variables  $S_f$ . In this approach one target variable may belong to more than one group.

Data records	Features					Target Variables				
	$f_1$	$f_2$	$f_3$	...	$f_m$	$E_1$	$E_2$	$E_3$	...	$E_n$
$d_{11}$	}					*				
$d_{12}$						*				
						*				
$d_{1,n_1}$										*
...		...						...		
$d_{k1}$	}							*		
$d_{k2}$						*				
$d_{k,n_k}$										*

**Figure 3. Typical data obtained by employing the clustering algorithm (\* means that specific target variable has damage, e.g. for  $d_{11}$ ,  $E_1$  is damaged).**

### 3.2. Localization of Interesting Target Variables

When the groups of similar target variables are identified, a classification model is used to identify the group of target variables that is of specific interest (e.g., in damage detection problem, a structure with damaged elements). If

a data record  $d_{ij}$  that belongs to a cluster  $C_i$  has a target variable which is set within the subgroup  $S_i$ , the specific group of target variables  $S_i$  is assumed to be of particular interest for further analysis and the corresponding data record  $d_{ij}$  is assigned an “interesting class” label. Otherwise, if data record  $d_{ij}$  does not contain any target variables which are set within a corresponding group of target variables, the “non-interesting” class label is assigned to considered data record. For example, in damage detection, every structure that has damaged elements will have “interesting” class label, while structures that do not have damaged elements will be labelled as “non-interesting”.

This procedure is repeated for all the data records within all clusters and corresponding groups of target variables. As a result, if there are  $k$  clusters and  $k$  corresponding groups of target variables,  $k$  data sets  $DS_i$  are also created, and each of them is used to build a classifier that will predict the presence of set target variables within the group  $S_i$  (e.g. the presence of damaged elements in the structure). As classification models, we have used multi-layer (2-layered) feedforward neural network models with two output nodes representing both “non-interesting” and “interesting” classes, where the predicted class is from the output with the largest response. Our previous experimental work [9] has shown that this type of neural network is more accurate than a neural network with a single output node coding 0 as “non-interesting” and 1 as “interesting” class. The number of hidden neurons was equal to the number of input attributes. In order to reduce the number of input attributes considered in neural network model, dimensionality reduction through principal component analysis is also employed. Three learning algorithms were used: resilient propagation [11], conjugate gradient backpropagation with Powell-Beale restarts [10], and Levenberg-Marquardt [5].

### 3.3. Prediction of Target Variables

If sufficiently large number of target variables are available within identified group of particular interest, the entire procedure of partitioning the group of target variables and predicting new subgroups of interests, presented in Figure 2, is repeated recursively. Otherwise, if the group of target variables contains sufficiently small numbers of target variables, the localized regression models that are constructed using only those data records that correspond to the identified group are employed to predict the value of each belonging target variables (e.g., to predict the intensity of the damage of belonging structural elements). As local regression models, we have trained 2-layered feedforward neural network models with number of hidden neurons equal to the number of input attributes.

The proposed hierarchical approach can identify similar target variables that are either close or symmetric in 3-D space, unlike manual partitioning which only discovers

similar target variables that are close in 3-D space. Therefore, identified clusters contain more similar data records and consequently correspond to more homogeneous regions than data subsets obtained with manual partitioning thus causing better prediction performance.

#### 4. Experiments

Predicting damage in mechanical structures using data mining techniques requires feature construction and data generation steps. For more details about these steps, reader is referred to [8]. The electric transmission tower (Figure 4), studied in [13], has been chosen to demonstrate the effectiveness of our approach in damage detection. The dataset of 1,560 records, 700 natural frequencies (features) and 312 structure elements (target variables) is generated by failing a random single element by a random amount. This data set corresponds to the scenario of 5 failure states per element (312 elements  $\times$  5 = 1560). The training and testing data sets were obtained by randomly splitting generated data set into equal partitions.

We first performed manual partitioning of the electric transmission tower into four legs and a head (Figure 4). The results of applying the neural network classifiers for predicting the existence of the damage within the structures identified using manual sub-structuring approach is given in Table 2. The prediction performance of neural network classifiers is measured by computing the overall classification accuracy, as well as by computing the recall, precision and F-value for each class (since in damage detection problem target variable (element) of interest is damaged one, we consider “damage” class and. “non-damage” class). In order to alleviate the effect of neural network instability in our experiments, measures for pre-

diction accuracy for each substructure are averaged over 20 trials of the neural network learning algorithm.

From Table 2 it can be observed that the manual sub-structuring approach followed by building neural network classification models can be very accurate for predicting the presence of damage within the substructures at the first level of partitioning. The achieved accuracy was higher than 98% for four substructures, while for leg 2, the classification accuracy was slightly worse (95.8%).

When partitioning the electric transmission tower using the hierarchical clustering approach, one of major issues was to select right set of natural frequencies that will be used in clustering. Since natural frequencies that are easy to be measured in practice are usually less than 25Hz, the first step typically involves elimination of higher frequencies. The second step includes the selection of features for the first level of partitioning and according to the domain knowledge, usually considers 10-30% of the total number of features (frequencies). In order to identify the exact number of features for the first level, we use a heuristic approach that looks for the largest gap between two natural frequencies in the vicinity of some round natural frequency (e.g. the biggest gap around 0.5 Hz or 1Hz). Identifying the number of frequencies that needs to be considered at each additional level of partitioning is performed in the same manner. Thus, the first level of partitioning was performed employing the 252 lowest frequencies, since these frequencies were smaller than 0.5Hz.

All five clustering algorithms within the CLUTO package were investigated, but only the best results achieved with agglomerative approach are reported here.

We have discovered 5 well balanced clusters  $C_i, i = \overline{1,5}$  of

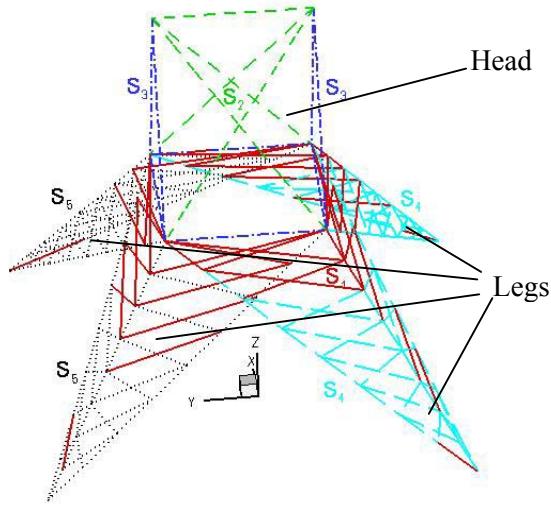
**Table 2. The manual sub-structuring based approach employing neural networks for five sub-structures by partitioning electric transmission tower.**

	Leg 1		Leg 2		Leg 3		Leg 4		Head	
<b>Accuracy</b>	<b>98.2</b>		<b>95.8</b>		<b>98.3</b>		<b>98.5</b>		<b>99.5</b>	
<b>Classes</b>	Non-damage	Damage	Non-damage	Damage	Non-damage	Damage	Non-damage	Damage	Non-damage	Damage
<b>Recall</b>	98.3	97.6	98.93	91.4	99.4	93.6	99.4	93.2	100	98.5
<b>Precision</b>	99.5	91.9	94.5	94.3	98.6	97.1	98.8	96.5	99.2	100
<b>F-value</b>	98.9	94.7	96.6	94.7	98.9	95.3	99.1	94.8	99.6	99.2

**Table 3. Classification results for predicting the presence of the damage within the substructures of electric transmission tower identified at the first level of partitioning**

	S1		S2		S3		S4		S5	
<b>Accuracy</b>	<b>98.3</b>		<b>99.8</b>		<b>99.5</b>		<b>97.7</b>		<b>99.2</b>	
<b>Classes</b>	Non-damage	Damage	Non-damage	Damage	Non-damage	Damage	Non-damage	Damage	Non-damage	Damage
<b>Recall</b>	98.5	97.5	100	97.9	99.4	100	98.0	96.9	99.3	99.2
<b>Precision</b>	99.6	91.7	99.7	100	100	97.2	99.0	93.9	99.6	98.4
<b>F-value</b>	99.1	94.5	99.9	98.9	99.7	98.6	98.5	95.4	99.4	98.8

size: 571, 167, 241, 278 and 303 data records. In identifying five substructures corresponding to the discovered five clusters, the *weighted majority* approach achieved the best results, and therefore only these results are reported here.



**Figure 4. Illustrative five sub-structures at the first level of clustering employing the algorithm Partitioning. (Figure is best viewed in color)**

Figure 4 shows five substructures  $S_i, i = \overline{1,5}$ , identified using the proposed approach, which sizes are 47, 40, 63, 76 and 86 structural elements respectively. From Figure 4 it is clear that the proposed algorithm *Partitioning* (Figure 2) is able to partition the mechanical structure into spatially close and symmetric sub-structures that are characterized by their low natural frequencies. The results of applying the neural network classifiers for predicting the existence of the damage within identified structures is given in Table 3. It is apparent from Table 3 that the proposed hierarchical partitioning approach achieved similar and even better prediction performance when identifying the presence of the damage within discovered substructures. For three substructures ( $S_1, S_2$  and  $S_3$ ) the achieved classification accuracies were slightly ( $S_1$ ) or significantly better ( $S_2, S_3$ ) for hierarchical clustering approach than for the manual partitioning approach, while for other two substructures ( $S_4, S_5$ ), the achieved classification accuracies for hierarchical partitioning approach were slightly worse than for the manual partitioning approach. It is important to note that the neural network classifiers in both cases were constructed using only the features employed at the first level of partitioning (in this case first 252 frequencies), since other features (frequencies) are irrelevant for predicting the presence of the damage at the first level.

Although the number of elements per substructures identified at the first level of hierarchical partitioning is still large, we constructed local regression models  $R_j, j$

$= \overline{1,312}$  for each element from substructures  $S_i, i = \overline{1,5}$ , using only data records from corresponding clusters  $C_i$ . We used these models then to predict the intensity of the damage of particular elements within the substructure  $S_i$ . To compare the predictive power of both manual and hierarchical partitioning, we also constructed local regression models  $M_j, j = \overline{1,312}$ , that are responsible for predicting each of the elements from substructures  $SM_i, i = \overline{1,5}$ , identified using manual approach. Finally, the global neural network regression models  $G_j, j = \overline{1,312}$ , were also built using all data records from entire data set.

For each of these models, the prediction performance was measured using the coefficient of determination defined as  $R^2 = 1 - \text{MSE}/\sigma^2$ , where  $\sigma$  is a standard deviation of the target variable.  $R^2$  value is a measure of the explained variability of the target variable, where 1 corresponds to a perfect prediction, and 0 to a trivial mean predictor. The  $R^2$  value for each element from substructures is again averaged over 10 trials of the neural network learning algorithm. The experimental results of predicting the intensity of damage in the elements within particular substructures are given in Table 4.

**Table 4. Prediction of damage intensity (given in  $R^2$  values) for those elements within structures for which hierarchical clustering approach achieves the worst and the best accuracy.**

Structure	Element	Direct (global) approach	Manual partitioning	Hierarchical clustering approach
$S_1$	$E_{15}$ (worst)	$< 0$	$\approx 0$	<b>0.154</b>
	$E_{96}$ (best)	$< 0$	0.485	<b>0.814</b>
	average*	0.004	0.182	<b>0.447</b>
$S_2$	$E_{241}$ (worst)	$< 0$	$< 0$	<b>0.140</b>
	$E_{263}$ (best)	$< 0$	0.23	<b>0.411</b>
	average*	0.03	0.172	<b>0.245</b>
$S_3$	$E_{312}$ (worst)	$< 0$	0.121	<b>0.148</b>
	$E_{209}$ (best)	$< 0$	$\approx 0$	<b>0.469</b>
	average*	0.005	0.04	<b>0.172</b>
$S_4$	$E_{102}$ (worst)	$< 0$	$\approx 0$	$\approx 0$
	$E_2$ (best)	0.243	$\approx 0$	<b>0.357</b>
	average*	0.02	0.04	<b>0.225</b>
$S_5$	$E_{207}$ (worst)	$< 0$	$\approx 0$	$\approx 0$
	$E_{195}$ (best)	$< 0$	0.218	<b>0.299</b>
	average*	0.007	0.08	<b>0.152</b>

Since the lack of space did not allow reporting prediction performance for all 312 elements, for each structure identified using the hierarchical approach, two elements were chosen such that they had the worst and the best prediction performance within the particular substructure. In addition, the average  $R^2$ -value of all the elements within particular structures is also reported. Since the  $R^2$ -

value of predicting particular elements may be less than 0 in many cases (e.g.,  $-1$  or  $-2$  for extremely poor global regression models in the direct approach), the average  $R^2$ -value for substructures is computed such that the  $R^2$ -value of those models with negative  $R^2$ -value is assigned zero value. Thus, the exceptionally bad regression models will not negatively influence the accurate ones.

It is apparent from Table 4 that both partitioning approaches produce localized regression models that are more accurate as well as require less data records for training than the global regression models used in the direct approach. In addition, the localized regression models built using hierarchical clustering approach are in most cases more accurate than localized models obtained by manual partitioning. Only in approximately 5-10% of the total number of elements, are the models obtained from manual partitioning  $M_i$  more accurate than the models  $R_i$  constructed in the hierarchical clustering approach. For example, the prediction of element  $E_{145}$  that belongs to the structure  $S_5$  is more accurate when using the localized models obtained through manual partitioning ( $R^2 = 0.245$ ) than when using the models obtained in hierarchical clustering based approach ( $R^2 = 0.165$ ).

The superior prediction performance of local regression models built through the proposed hierarchical approach compared to other regression models may be explained by the fact that local regression models  $R_j$  are constructed using more similar data records than models  $M_j$  from manual partitioning and global prediction models  $G_j$ . Specifically, the similarity of data records in clusters originates from elements that are both close and symmetric in 3-D space, while the similarity of data records in the data subsets used for building models  $M_j$  from manual partitioning arise only from elements that are close in space. Therefore, the heterogeneity of these data subsets is potentially higher than the heterogeneity of data clusters and thus, the corresponding localized models built on these data subsets are less accurate. On the other hand, a few scenarios where the models  $M_j$  built through manual partitioning are more accurate than the models  $R_j$  built through hierarchical clustering may be explained by imperfect quality of the obtained clusters. Imperfect clusters were probably achieved due to the hidden information in domain specific knowledge, which was incorporated into the clustering algorithm.

Since the number of elements per structures  $S_i$ ,  $i = \overline{1,5}$ , identified at the first level is still large, we performed additional level of partitioning using algorithm *Partitioning*, (presented in Figure 2) in order to further reduce the number of elements per substructure. Each of the clusters  $C_i$  identified at the first partitioning level was further partitioned into clusters  $C_{ij}$  using additional 198 lowest frequencies (frequencies smaller than 1.5Hz). The clustering was performed using only these 198 frequencies (fre-

quencies  $f_{253}$  to  $f_{450}$ ), since the lowest 252 frequencies used at the first level of partitioning are not useful for further identification of similar structure elements. The number of clusters identified at the second level varied between 3 and 5, and each of these clusters  $C_{ij}$  at the second level corresponded to particular substructures  $S_{ij}$ . The results of applying the neural network classifiers for predicting the existence of the damage within the substructures identified at the second level of partitioning using our proposed hierarchical clustering based approach are omitted due to lack of space. Although there was a slight decrease in overall classification accuracy when predicting the presence of the damage, especially for structures  $S_{12}$ ,  $S_{13}$  and  $S_{43}$ . The achieved prediction performance for remaining substructures (using local neural network classifiers that are also constructed using only data records from corresponding clusters  $C_i$ ) is still very high, larger than 97.5% as earlier for the first level of partitioning.

Our previous experimental results [13] have shown that the decrease in the classification performance at the second level of manual partitioning was significantly larger (20–30%) than the drop observed for the proposed hierarchical approach. The prediction of the modulus of elasticity for elements within the substructures is again performed using only the localized regression models built after the second level of hierarchical approach.

**Table 5. Prediction of damage intensity ( $R^2$  value) for those elements within structures for which hierarchical approach at the first level achieves the worst and the best accuracy**

Structure	Element	Hierarchical clustering approach (the first level)	Hierarchical clustering approach (the second level)
S <sub>1</sub>	E <sub>15</sub>	0.154	<b>0.264</b>
	E <sub>96</sub>	<b>0.814</b>	0.668
	average	<b>0.447</b>	0.424
S <sub>2</sub>	E <sub>241</sub>	0.140	<b>0.477</b>
	E <sub>263</sub>	<b>0.411</b>	0.264
	average	0.245	<b>0.294</b>
S <sub>3</sub>	E <sub>312</sub>	0.148	<b>0.243</b>
	E <sub>209</sub>	<b>0.469</b>	0.431
	average	0.172	<b>0.244</b>
S <sub>4</sub>	E <sub>102</sub>	≈ 0	<b>0.145</b>
	E <sub>2</sub>	<b>0.357</b>	0.285
	average	0.225	<b>0.234</b>
S <sub>5</sub>	E <sub>207</sub>	≈ 0	≈ 0
	E <sub>195</sub>	0.299	<b>0.334</b>
	average	0.152	<b>0.195</b>

Table 5 illustrates how the prediction of the modulus of elasticity for the same elements investigated within substructures at the first level has been changed when these substructures are further partitioned. It is apparent

that for large number of elements, further localization and the building of more specific regression models results in an improved  $R^2$  value, while for some other smaller number of elements additional localization hurt the prediction accuracy. There may be several reasons for such phenomenon. First, at the second level the number of data records that are relevant for the damage of the specific element may be smaller than the number of data records at the first level of partitioning thus causing a possible loss of valuable information or potential overfitting of neural network regression model. Second, this loss is inevitably caused by an imperfect clustering algorithm that does not necessarily assign all data records corresponding to the damage of specific element to a single cluster. On the other side, improvement in prediction performance at the second partitioning level is evidently due to the fact that there are also fewer data records that are irrelevant to the damage of a particular element and therefore, localized regression models are more accurate.

## 5. Conclusions

The paper presented a novel general framework for the efficient prediction of multiple target variables from high dimensional and heterogeneous data sets using the hierarchical clustering approach. This approach is especially effective where there is a natural relationship among target variables. A key desirable feature of this scheme is that it achieves a reduction in both data records and the number of used features, which is of great importance in applications where it is difficult to find a large number of training records, or when learning a monolithic prediction model from very large data sets is prohibitively slow. The effectiveness of the approach was demonstrated on the problem of damage detection in very large and complex mechanical structures. Our experiments indicate that the proposed approach can be successfully used to predict the presence of damage within the structure, as well as the intensity of the damage in any of the several hundred structure elements that served as target variables. Furthermore, we have shown that the proposed approach is computationally more efficient, more accurate and requires less data records than the direct approach that was earlier used for damage detection, as well as the manual partitioning approach.

## 6. Acknowledgments

This work is partially supported by the Department of Energy LLNL B512340 and by the Army High Performance Computing Research Center (AHPARC) under the auspices of the Department of the Army, Army Research Laboratory (ARL) under contract number DAAD19-01-2-0014. The content does not necessarily reflect the position or the policy of the government, and no official

endorsement should be inferred. Additional thanks are also due to S. Sandhu and M. Steinbach for related technical discussions. Access to computing facilities was provided by the AHPARC and the Minnesota Supercomputing Institute.

## 7. References

- [1] J. Baxter, Theoretical Models of Learning to Learn, In T. Mitchell, S. Thrun, Eds, *Learning to Learn*. Kluwer, Boston, 1997.
- [2] L. Breiman, J. Friedman, Predicting Multivariate Responses in Multiple Linear Regression, *Journal of the Royal Statistical Society*, Series B 59, 3—54, 1997.
- [3] R. Caruana, Multitask Learning, *Machine Learning*, 28:41-75, 1997.
- [4] H. Chen, N. Bicanoc, Assessment of Damage in Continuum Structures Based in Incomplete Modal Information, *Computers and Structures*, 74:559-570, 2000.
- [5] M. Hagan, M. Menhaj, Training feedforward networks with the Marquardt algorithm, *IEEE Transactions on Neural Networks*, 5:989-993, 1994.
- [6] N. Intrator, S. Edelman, How to Make a Low-dimensional Representation Suitable for Diverse Tasks, *Connection Science*, 8, 1996.
- [7] Jordan, M. I., Jacobs, R. A., Hierarchical Mixtures of Experts and The EM Algorithm, *Neural Computation*, 6 (2), 181 – 214, 1994.
- [8] A. Lazarevic, R. Kanapady, C. Kamath, V. Kumar, K. Tamma, Localized Prediction of Continuous Target Variables Using Hierarchical Clustering, *AHPARC Technical Report 2003-123*, [www.cs.umn.edu/~aleks/pub\\_list.htm](http://www.cs.umn.edu/~aleks/pub_list.htm)
- [9] A. Lazarevic, Z. Obradovic, Knowledge Discovery in Multiple Spatial Databases, *Journal of Neural Computing and Applications*, Vol. 10, pp. 339-350, 2002.
- [10] M. Powell, Nonconvex minimization calculations and the conjugate gradient method, In D. Griffiths Ed., *Numerical Analysis, Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 1066: 122-141, 1984.
- [11] M. Riedmiller, T. Braun, A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm, *IEEE International Conference on Neural Networks*, San Francisco, 1993.
- [12] G. J. Rix, Interpretation of Nondestructive Integrity tests using artificial neural networks, *Structure Congress 12, ASCE*, Reston VA, 1246-1351, 1994.
- [13] S. Sandhu, R. Kanapady, K. Tamma, C. Kamath, V. Kumar, A sub-structuring approach via data mining for damage prediction and estimation in complex structures, *SIAM International Conference on Data Mining*, Arlington, VA, 2002.
- [14] Z. Szewczyk, P. Hajela, Damage detection in structures based on feature sensitive neural networks, *Journal of Computation in Civil Engineering*, ASCE, 8(2):163-179, 1994.
- [15] S. Thrun, J. O'Sullivan, Clustering Learning Tasks and the Selective Cross-task Transfer of Knowledge, *Technical Report CMU-CS-95-209*, Carnegie Mellon University, Pittsburgh, 1995.
- [16] Y. Zhao and G. Karypis, Criterion Functions for Document Clustering Experiments and Analysis, *AHPARC Technical Report #01-40*, 2000.