KNOWLEDGE DISCOVERY IN RESIDENTIAL CONSTRUCTION PROJECT COST DATA

A. Boussabaine and T. Elhag

The University of Liverpool, School of Architecture and Building Engineering, Construction Cost Engineering Research Group, Liverpool L69 3BX, UK

Clients, designers, consultants and contractors are required to make business decisions. Frequently the decision must be based on supporting information that is uncertain, incomplete, or expressed in qualitative rather than quantitative terms. Estimation of construction costs always includes uncertainty, which is caused by the limited ability of decision makers to recognize the association between different cost attributes.

Knowledge discovery is a recent development in the field of database exploration. A powerful and well-investigated topic is the discovery of association rules. This paper examines the association between the cost of projects and cost attributes. The paper reports on relationships related to a particular event occurring and examines the rules, which state that, for example, when a particular cost attribute is present then the cost of construction projects increases or decreases in a certain percentage of cases. The association between the cost attributes will also be described. This provides a means to examine many possible relationships between cost attributes and final projects cost prior to any design or contractual decisions.

Keywords: cost, cost attributes, construction cost, knowledge, interdependency, residential projects.

INTRODUCTION

Clients, designers, consultants and contractors are required to make business decisions. Frequently the decision must be based on supporting information that is uncertain, incomplete, or expressed in qualitative rather than quantitative terms. Estimation of construction costs always includes uncertainty, which is caused by the limitations of construction knowledge, available data and decision makers recognition of the association between different cost attributes.

Despite the wealth of information, clients, consultants and contractors have been unable to fully capitalize on its value because information and knowledge implicit in the construction cost data is not easy to discern. Recent research commissioned by DETR to study the feasibility of developing a framework for a national construction cost database has reviewed previous and existing work relating to construction cost databases and discussed the measures and data requirement in the construction industry (Davis Langdon and Everest 1998). The report found that there is a lack of consistency between existing databases. The report also informed industry of the necessity to develop a system of industrial performance indicators. The data requirements recommended by the report do not include nor mention tools for exploring and extracting knowledge that are commonly understood by decision makers in the construction industry.

Boussabaine, A H and Elhag, T M S (1999) Knowledge discovery in residential construction project cost data. *In:* Hughes, W (Ed.), *15th Annual ARCOM Conference*, 15-17 September 1999, Liverpool John Moores University. Association of Researchers in Construction Management, Vol. 2, 489-98.

Table 1: Cost attributes and associated costs

Cost attributes		
Attribute	**Category**	**Costs**
Building function	Estate housing, nursing homes, others	Sub structure
Type of units	House flat bungalow cottage bed-sits	Super structure
Type of client	Private/developer, public/developer	Internal finishes
Location		Fittings
Duration		Services
Number of stories		External works
GFA	(m^2)	Preliminaries
Site condition	Green, demolition, normal , others	Contingencies
Site topography	Level, gentle slope ,steep slope, others	Contract sum
Ground condition	Not contaminated, contaminated, good, moderate, others	
Access to site	Highly restricted, restricted, un-restricted, others	
Work space	Highly restricted1, restricted, un-restricted , others	
Market condition	Stable, keen competitive ,unstable, others	
Tender document	BoQ, others	
Number of tenders	Issued, received	
Type of contract	JCT (intermediate), JCT (LA), JCT (private), D&B, others	
Selection	Negotiation, competitive, others	
Cost fluctuation	Fixed, firm, NEDO, formula, others	
External work	Yes, no	
Foundation	Strip pad raft, pile, concrete, slab, concrete trench, others	
Type of floors	Timber, PCC , in-situ conc., Chipboard , others	
Walls	Bricks and blocks, cladding ,timber , others	
Windows	Timber, PVC, steel, others	
Roof	Pitched, flat	
Heating	Central heating, gas heaters, elect. Heaters , either / not specified	

An increasing problem is one of superabundant information, with a huge selection of potential relevant items to sift through in order to ascertain the truly significant ones. Traditional computer-based methods are weak at helping in this area, largely because decisions are made mainly on the basis of historically recognized or perceived patterns (Hosheimer and Siebes 1996). The range of intelligence-technology computing tools which have developed rapidly in recent years is better suited to learning and recognizing patterns in data-sets with complex internal relationships (Dietrich and Wittmann 1998). Knowledge discovery techniques are increasingly used as a source for business intelligence groups using historic records to predict the level of risk when setting insurance premiums, and supermarket chains examining till receipts to develop profitable marketing campaigns (Thomas, W., and Johannes 1998). Construction cost data, is known to be very complex, difficult to model, especially with fitted to linear models (Davis Langdon & Everest), and is clearly a candidate for analysis with intelligent methods of information discovery.

This work uses an ANN, known as Kehonen Self-Organizing feature Maps (SOM), in construction cost data analysis. SOM has been particularly useful in data analysis in providing variability. The current work emphasis is on SOM as a valuable aid in representing cost data and addresses the problem of variability and interdependence between cost attributes and cost of residential projects.

HOUSING PROJECTS CONSTRUCTION COST DATA

This work concentrates on construction cost data classification, clustering (segmentation) and focuses on dependencies between the attributes that are used to

Table 2: Correlation matrix before clustering the data set

Attributes	Sub Structure	Super Structure	Internal Finishes	Fittings	Services	External Works	Prelim-inaries	Contin-gencies	Contract Sum
	Costs								
GFA	0.60	0.94	0.92	0.36	0.93	0.76	0.80	0.56	0.91
House	0.56	0.89	0.88	0.31	0.87	0.69	0.75	0.52	0.85
Duration	0.57	0.68	0.68	0.30	0.62	0.55	0.64	0.34	0.67
Location factor	0.39	0.28	0.25	0.19	0.25	0.29	0.20	0.36	0.30
Number of stories	0.31	0.30	0.30	0.11	0.25	0.20	0.26	0.08	0.27
Cost fluctuation	0.19	0.19	0.15	0.07	0.16	0.19	0.14	0.11	0.19
Type of contract	0.19	0.15	0.16	0.19	0.16	0.18	0.11	-0.01	0.17
Type of client	0.07	0.15	0.12	-0.06	0.13	0.09	0.16	0.26	0.14
Windows	0.04	0.10	0.10	0.11	0.12	0.11	0.07	0.22	0.11
Work space	0.13	0.13	0.11	0.05	0.13	0.06	0.14	0.06	0.11
Market condition	0.05	0.10	0.13	0.04	0.13	0.04	0.07	0.13	0.09
Bed sits	0.00	0.06	0.01	0.08	0.07	0.11	0.18	0.06	0.09
Type of floors	0.01	0.07	0.11	0.23	0.08	0.04	0.08	0.04	0.07
Access to site	0.11	0.09	0.11	0.06	0.12	-0.01	0.11	-0.01	0.07
Foundation	0.03	0.05	0.02	-0.05	0.07	0.08	0.00	0.02	0.06
External work	0.03	0.03	0.03	0.02	0.01	0.06	0.04	0.01	0.04
Site topography	0.09	0.00	0.01	-0.06	0.00	-0.04	0.02	-0.06	-0.01
Bungalow	-0.01	-0.05	-0.02	0.14	0.02	0.02	0.01	-0.03	-0.01
No of tenders	0.00	0.00	-0.01	-0.05	-0.01	-0.04	0.00	0.00	-0.02
Flat	-0.03	-0.03	-0.03	-0.02	-0.03	-0.03	-0.02	-0.03	-0.03
Cottage	-0.01	-0.04	-0.02	-0.03	-0.03	-0.04	-0.06	-0.07	-0.04
Ground condition	0.02	-0.02	-0.01	0.00	-0.01	-0.12	0.02	-0.08	-0.05
Site condition	0.05	-0.05	-0.07	-0.15	-0.05	-0.08	0.00	-0.06	-0.05
Heating	-0.03	-0.18	-0.18	-0.14	-0.18	-0.20	-0.20	-0.14	-0.19

describe a domain of interest, e.g., the cost of construction projects, procurement method, type of structure, etc. In this proposal the discovery of attribute dependencies are considered as three sub-tasks; discovery of redundant attributes in the representation of dependencies between condition and decision attributes (e.g., cost of super and sub-structures, etc.), identification of the most important condition attributes, and discovery of decision rules characterizing the dependency between values of condition attributes and decision attributes. The data is collected from a variety of sources and consists of the records of 200 residential projects. The cost attributes and associated construction costs are shown in Table 1.

Statistical analysis was carried out on the whole set of data before using SOM to cluster projects into neighbouring classes. The purpose of this is to see whether the interdependency features between cost attributes and construction costs change significantly from the original sample. A summary of the interdependencies between cost attributes and project costs are shown in Table 2. This table shows that there are only three attributes that have a significant linear relationship with costs.

METHODOLOGY

Construction cost data is very complicated which requires intelligent processing to get a precise view of the effects of the cost attributes on project costs. In this work SOM is used for the analysis and visualization of construction cost data. The cost attributes used in the analysis are shown in Table 1. SOM clustered the training data using the learning algorithm of a commercial ANN package.

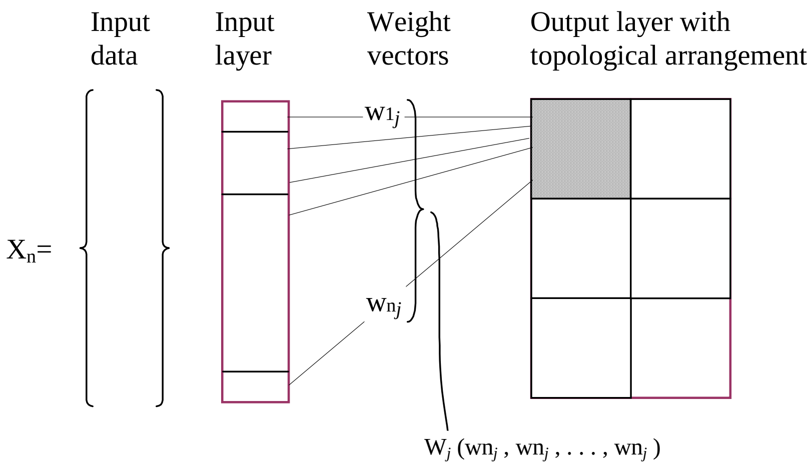

Figure 1: A 2-dimensional SOM network

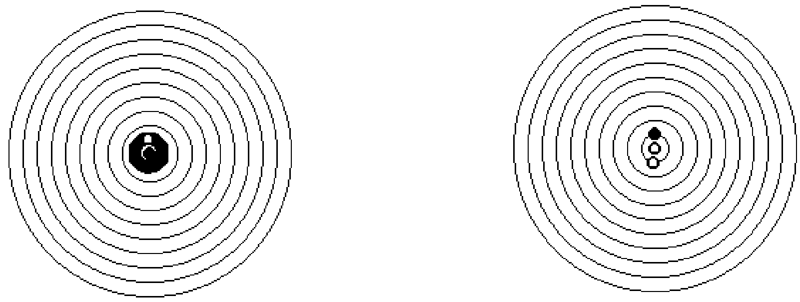

Figure 2: Result of clustering

Figure 1 illustrates the typical map of 3 by 2 two-dimensional SOM trained with data from 200 residential projects. The 2-dimesional SOM, shown in Figure 1 has been used to assist in the analysis of cost data of residential projects. SOM is an orderly mapping of a high-dimensional distribution of data onto a regular low-dimensional grid. The results produced by SOM algorithms can be considered to be similarity diagrams of data and their clusters.

After training the nodes on data, the output map comes to represent characteristic classes of projects with similar patterns. An example of an output class is shown in Figure 2. Figure 2 indicates the strength of the similarity of the patterns within the group represented by the neuron. A strong grouping will have all the patterns close to the centre. A weak group will have patterns widely distributed away from the centre. A weak grouping may indicate the need for more training or for more neurons to be added to the net to allow more groups to be created during training allowing the similarities to be better represented. The closer the pattern to the centre, the stronger the similarity between the pattern and the centre of the group represented by the neuron. The further from the centre the weaker the similarity between the pattern and the group.

INTERDEPENDENCY ANALYSIS WITH THE KOHONEN SELF-ORGANIZING FEATURE MAP

Representing the cost data into clusters or maps graphically provides the first means by which to interpret them. Graphical systems of today allow the visualization of information in any desired way. The following aspects of cost data visualization have to be considered to make the information helpful and easy to use for construction professionals.

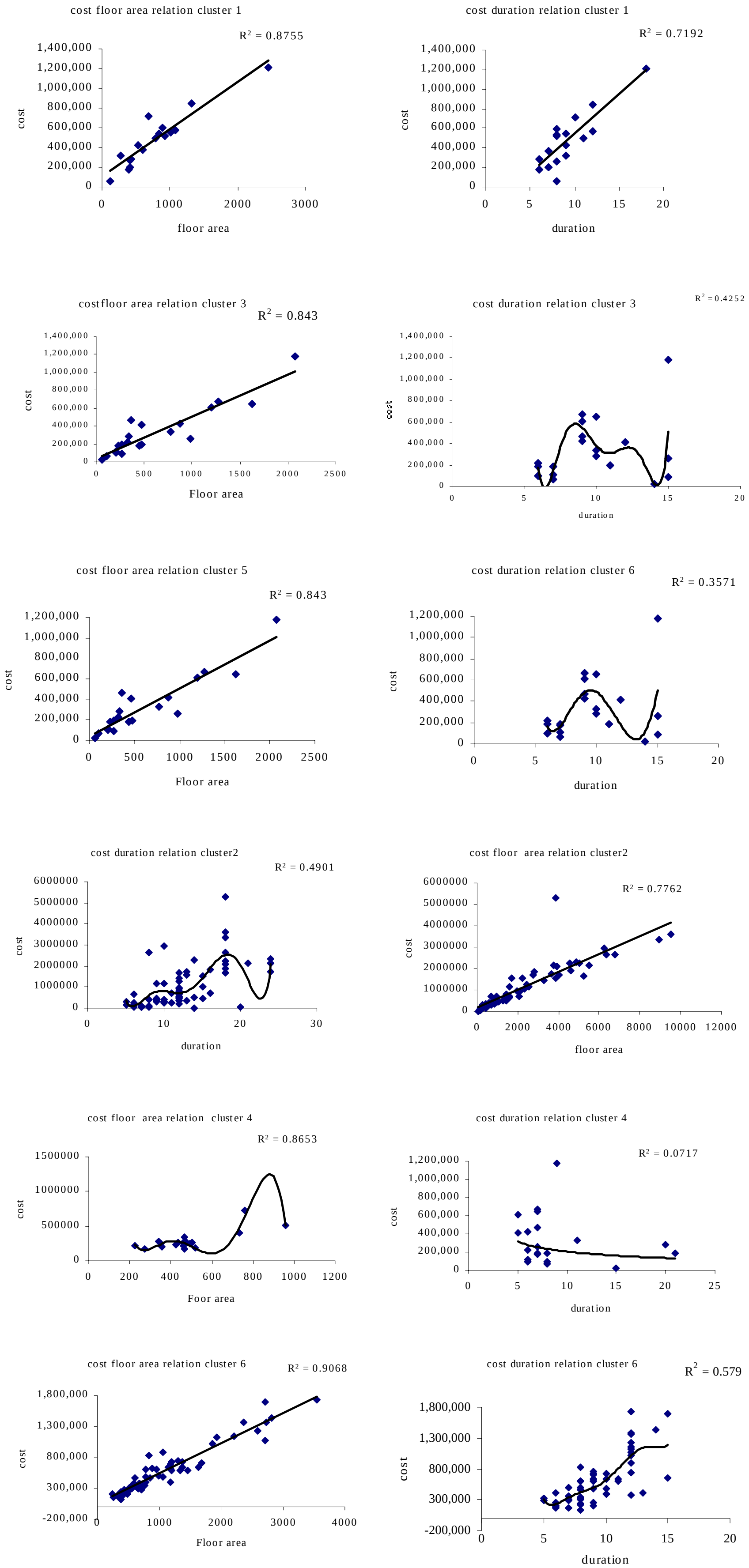

Figure 3: Cost floor area and duration relationship

- Intuitive grasp of the situation
- Classification of the actual project state in the set of all possible states
- Visualization of continuously or abrupt changing states
- Complex problems reduced to compressed information
- Assessed information and not only data values

To visualize and identify the interdependency between the feature of each generated class, correlation matrix for each cluster and scatter plots of grass floor area and duration VS costs were used in the interpretation of the classes. Other combinations of graphs were developed but there is not enough space to include and discuss them in this paper. Table 3-8.3 shows the correlation between cost attributes and combinations of 9 cost categories (see Table 1). Figure 3 illustrates the relationship between floor area, duration and contract sum of each of the 6 clusters. These figures will be discussed the next section.

Characteristics of cluster 1
Cluster 1 represents bungalows with a range of gross floor area between 119 and 2448 M2. The best fit relationship between cost and floor was found to be linear with an accuracy of 87.55%. Also the best fit for the relationship between cost and duration is linear with an accuracy of 71%. Table 3 shows that gross floor area, duration, access, location and type of floor cost attributes have a significant correlation with costs. Note that the correlation of the super and sub-structures costs with cost attributes is almost identical, whereas contingency cost has a negative negligible correlation with cost attributes. Also the correlation with fittings cost is very low.

Characteristics of cluster 2
This group contains only house projects. The gross floor area of these projects ranges from 78 to 9538 m^2 and the duration between 5 and 18 months. The best fit relationship between contract cost and floor area is linear with a 77.6% accuracy, whereas the best fit for the duration is polynomial with an accuracy of 49%. This suggests that there is a considerable noise in the relationship between cost and duration as can be seen in Figure 3. Notice in Table 4 that the order of the correlation strength between attributes and costs has changed. In this cluster, floor area, duration, contract type, cost fluctuation and windows type factors have a strong correlation with costs. Also of notice is the positive and strong correlation between attributes and contingency cost

Characteristics of cluster 3
This group of projects consists of houses with the duration ranging between 6 and 15 months and floor area between 64 and 2075 m2. The best fit between cost floor area and duration is linear and polynomial with an accuracy of 84.3% and 42% respectively. The order of magnitude of the correlation between cost and cost attributes is the same as in cluster 2.

Characteristics of cluster 4
This segment of project covers houses with multiple stories and bungalows. What is noticeable in this cluster is the non-linear relationship between cost and floor area. The best fit was found to be polynomial within an accuracy of 86.33% accuracy. The best fit between cost and duration is exponential with only 7.17% accuracy. This noise is very noticeable in the correlation between attributes and costs as shown in Figure 3. The order of the correlation strength between costs and attributes is different

Table 3: Correlation between costs and cost attributes - cluster 1

Attributes	Costs								
	Sub structure	**Super Structure**	**Prelim-inaries**	**External Works**	**Services**	**Internal Finishes**	**Fittings**	**Conting-encies**	**Contract cost**
GFA	0.83	0.90	0.85	0.81	0.86	0.79	0.52	-0.08	0.94
Bungalow	0.82	0.86	0.83	0.82	0.81	0.71	0.42	-0.08	0.91
Duration	0.82	0.73	0.76	0.80	0.78	0.67	0.50	-0.22	0.85
Access	0.50	0.36	0.39	0.41	0.50	0.56	0.45	-0.14	0.48
Location	0.38	0.20	0.23	0.43	0.30	0.30	0.33	0.15	0.36
Type of floors	0.47	0.28	0.18	0.32	0.41	0.14	0.39	-0.13	0.34
Space	0.22	0.30	0.33	0.32	0.28	0.29	-0.03	-0.17	0.32
Foundation	0.21	0.11	0.12	0.03	0.08	0.03	-0.06	0.13	0.09
Topography	0.09	-0.05	0.35	0.05	0.09	0.14	0.09	-0.37	0.08
Fluctuation	0.13	-0.11	0.00	0.17	0.09	0.20	0.23	-0.13	0.08
Contract	0.12	-0.15	0.04	0.05	0.02	-0.42	-0.15	-0.36	-0.04
No Tender	-0.08	-0.08	0.27	-0.05	-0.17	-0.25	-0.28	-0.34	-0.07
Ground cond.	-0.29	-0.15	0.33	-0.09	-0.24	-0.10	-0.46	-0.28	-0.12
Windows	-0.20	-0.04	-0.07	-0.24	-0.10	0.20	0.15	0.23	-0.12
N Tenders	-0.05	0.02	-0.50	-0.16	-0.08	0.02	0.14	0.31	-0.13
Heating	-0.13	-0.18	-0.19	-0.21	-0.25	-0.09	-0.23	0.21	-0.21
S. Condition	-0.40	-0.38	-0.34	-0.35	-0.42	-0.48	-0.42	0.03	-0.42

Table 4: Correlation between costs and cost attributes - cluster 2

Attributes	Costs								
	Sub structure	**Super Structure**	**Prelim-inaries**	**External Works**	**Services**	**Internal Finishes**	**Fittings**	**Conting-encies**	**Contract cost**
GFA	0.75	0.92	0.91	0.55	0.92	0.69	0.76	0.53	0.88
House	0.71	0.89	0.89	0.51	0.89	0.65	0.74	0.49	0.84
Duration	0.58	0.56	0.56	0.46	0.42	0.46	0.53	0.22	0.55
Contract	0.37	0.34	0.30	0.28	0.26	0.38	0.16	0.21	0.34
Fluctuation	0.35	0.28	0.22	0.22	0.23	0.30	0.20	0.15	0.29
Windows	0.20	0.23	0.24	0.18	0.27	0.30	0.15	0.39	0.27
Location	0.38	0.23	0.20	0.42	0.20	0.30	0.09	0.34	0.27
Type of Floors	0.06	0.12	0.17	0.23	0.13	0.05	0.12	0.02	0.11
Stories	0.09	0.07	0.09	0.07	0.08	0.03	0.20	-0.09	0.08
NTender	0.06	0.03	0.13	0.10	0.03	0.03	0.21	0.02	0.06
Bedsits	-0.04	0.03	-0.04	0.18	0.04	0.09	0.17	0.03	0.06
Access	0.02	0.09	0.13	0.10	0.14	-0.07	0.11	0.01	0.05
Bungalow	0.06	0.04	0.07	0.01	0.04	0.01	0.12	-0.11	0.05
Market Condition	0.02	0.05	0.09	0.11	0.10	0.00	0.00	0.09	0.04
Space	-0.01	0.06	0.03	0.07	0.08	-0.03	0.05	0.04	0.03
Topography	0.05	0.04	0.04	0.01	0.04	-0.05	0.05	-0.04	0.02
Foundation	0.02	-0.01	-0.06	-0.11	0.03	0.02	-0.09	-0.04	-0.01
N Tenders	-0.05	-0.05	-0.03	-0.05	-0.04	-0.07	0.05	-0.10	-0.05
Heating	-0.10	-0.12	-0.11	-0.14	-0.09	-0.10	-0.14	-0.10	-0.12
G Condition	-0.14	-0.12	-0.09	-0.11	-0.08	-0.25	-0.05	-0.15	-0.15
S Condition	-0.11	-0.16	-0.17	-0.34	-0.14	-0.16	-0.10	-0.18	-0.16

from the previous clusters. What is noticeable here is that ground condition and heating type attributes have a moderate positive correlation with costs whereas contingency cost has a weaker relationship with cost attributes with the exception of the client type variable.

Characteristics of cluster 5

This group of projects contains x houses. The best fit between cost and floor area is linear with an accuracy of 84.3% whereas the best cost-duration is polynomial with an accuracy of 35.71%. In this group contract type, site condition, space and access attributes have a significant correlation with costs. A notable characteristic in this cluster is the similarity of the correlation between attributes and costs with the exception of contingency cost relationship with windows and foundations attributes as shown in Table 7.

Table 5: Correlation between costs and cost attributes - cluster 3

Attributes	Costs								
	Sub structure	**Super Structure**	**Prelim-inaries**	**External Works**	**Services**	**Internal Finishes**	**Fittings**	**Conting-encies**	**Contract cost**
GFA	0.75	0.92	0.91	0.55	0.92	0.69	0.76	0.53	0.88
House	0.71	0.89	0.89	0.51	0.89	0.65	0.74	0.49	0.84
Duration	0.58	0.56	0.56	0.46	0.42	0.46	0.53	0.22	0.55
Contract	0.37	0.34	0.30	0.28	0.26	0.38	0.16	0.21	0.34
Fluctuation	0.35	0.28	0.22	0.22	0.23	0.30	0.20	0.15	0.29
Windows	0.20	0.23	0.24	0.18	0.27	0.30	0.15	0.39	0.27
Location	0.38	0.23	0.20	0.42	0.20	0.30	0.09	0.34	0.27
Type of floors	0.06	0.12	0.17	0.23	0.13	0.05	0.12	0.02	0.11
Stories	0.09	0.07	0.09	0.07	0.08	0.03	0.20	-0.09	0.08
N Tender	0.06	0.03	0.13	0.10	0.03	0.03	0.21	0.02	0.06
Bedsits	-0.04	0.03	-0.04	0.18	0.04	0.09	0.17	0.03	0.06
Access	0.02	0.09	0.13	0.10	0.14	-0.07	0.11	0.01	0.05
Bungalow	0.06	0.04	0.07	0.01	0.04	0.01	0.12	-0.11	0.05
Market condition	0.02	0.05	0.09	0.11	0.10	0.00	0.00	0.09	0.04
Space	-0.01	0.06	0.03	0.07	0.08	-0.03	0.05	0.04	0.03
Topography	0.05	0.04	0.04	0.01	0.04	-0.05	0.05	-0.04	0.02
Foundation	0.02	-0.01	-0.06	-0.11	0.03	0.02	-0.09	-0.04	-0.01
Heating	-0.10	-0.12	-0.11	-0.14	-0.09	-0.10	-0.14	-0.10	-0.12
G Condition	-0.14	-0.12	-0.09	-0.11	-0.08	-0.25	-0.05	-0.15	-0.15
S Condition	-0.11	-0.16	-0.17	-0.34	-0.14	-0.16	-0.10	-0.18	-0.16

Table 6: Correlation between costs and cost attributes - cluster 4

Attributes	Costs								
	Sub structure	**Super Structure**	**Internal Finishes**	**Fittings**	**Services**	**External Works**	**Prelim-inaries**	**Conting-encies**	**Contract cost**
Stories	0.75	0.90	0.52	0.38	0.58	0.06	0.79	0.07	0.89
GFA	0.42	0.79	0.37	0.30	0.44	0.37	0.54	0.02	0.76
Duration	0.45	0.69	0.15	0.25	0.20	-0.21	0.51	0.23	0.57
Location	0.58	0.49	0.59	0.32	0.40	-0.26	0.57	0.24	0.48
Fluctuation	0.00	0.50	-0.20	0.40	-0.08	0.27	0.19	0.37	0.41
Bungalow	0.00	0.50	-0.20	0.40	-0.08	0.27	0.19	0.37	0.41
G Condition	0.33	0.23	0.38	0.52	0.45	0.22	0.34	-0.38	0.37
Heating	0.29	0.26	0.21	0.41	0.21	0.19	0.22	0.19	0.35
Type of Floors	-0.05	0.44	-0.28	0.38	-0.10	0.25	0.16	0.27	0.35
Space	0.37	0.29	0.19	0.36	0.40	0.09	0.19	-0.34	0.32
Access	0.37	0.31	0.14	0.07	0.34	0.00	0.17	-0.26	0.29
S Condition	0.34	0.20	0.17	0.09	0.18	-0.21	0.38	-0.10	0.21
Topography	0.33	0.19	0.14	0.02	0.24	-0.20	0.21	-0.12	0.18
Windows	-0.08	0.20	-0.18	0.50	-0.06	0.07	0.00	0.62	0.13
Foundation	-0.10	-0.11	-0.10	0.02	-0.02	-0.02	-0.18	-0.20	-0.14
Client	-0.20	-0.05	-0.46	0.20	-0.51	-0.51	-0.20	0.58	-0.25
N Tender	-0.39	-0.25	-0.45	-0.50	-0.48	-0.11	-0.55	0.18	-0.42

Characteristics of cluster 6

This class contains a mixture of houses, cottages and bungalows (57 projects). The best fit relationship between cost and floor area is linear with a 90.68% accuracy whereas the relationship between cost and duration is polynomial with an accuracy of 57.9%. The noticeable aspect in this group is that contract type of floor and number of stories cost attributes have a moderate correlation with costs. Table 8 shows that the strength of correlation between costs and cost attributes is almost identical with the exception of fittings and contingency costs.

DISCUSSION

An important contribution of this work is to report a method of performing construction cost data analysis using SOM techniques to analyse the interdependency between attributes and different project costs. The work shows how construction cost

Table 7: Correlation between costs and cost attributes - cluster 5

Attributes	Costs								
	Sub structure	Super Structure	Internal Finishes	Fittings	Services	External Works	Prelim-inaries	Conting-encies	Contract cost
GFA	0.81	0.91	0.95	0.93	0.92	0.75	0.63	0.26	0.88
House	0.63	0.73	0.82	0.86	0.87	0.74	0.62	0.40	0.79
Duration	0.45	0.49	0.36	0.32	0.34	0.37	0.34	0.05	0.43
Contract	0.41	0.49	0.42	0.52	0.39	0.36	-0.02	-0.14	0.40
S Condition	0.26	0.31	0.29	0.26	0.32	0.24	0.47	0.35	0.32
Space	0.32	0.29	0.20	0.20	0.10	0.20	-0.20	-0.21	0.20
Access	0.24	0.12	0.14	0.02	-0.03	0.16	-0.19	-0.41	0.11
type of Floors	0.11	0.19	0.12	0.13	0.04	-0.13	0.06	-0.23	0.03
Windows	-0.12	-0.09	-0.09	0.22	0.13	0.05	-0.09	0.51	0.00
G Condition	0.01	0.00	-0.11	-0.14	-0.07	0.00	-0.02	-0.28	-0.03
Stories	0.05	0.14	0.05	0.08	-0.01	-0.19	0.01	-0.21	-0.03
Foundation	-0.10	0.06	-0.02	0.15	0.01	-0.05	-0.32	0.50	-0.03
Market Condition	-0.05	-0.12	-0.06	-0.11	-0.11	-0.07	0.02	-0.03	-0.08
Cottage = 4	-0.01	0.05	-0.03	0.00	-0.08	-0.23	-0.12	-0.30	-0.10
Location	-0.22	-0.08	-0.05	-0.02	0.01	-0.22	0.25	0.26	-0.11
Heating	-0.15	-0.15	-0.10	-0.20	-0.21	-0.24	0.11	-0.35	-0.18
Topography	-0.31	-0.31	-0.28	-0.30	-0.24	-0.26	0.02	-0.39	-0.28
N Tender	-0.31	-0.31	-0.20	-0.23	-0.28	-0.39	-0.10	-0.34	-0.34
Fluctuation	-0.20	-0.40	-0.50	-0.70	-0.59	-0.29	-0.35	-0.22	-0.39

Table 8: Correlation between costs and cost attributes - cluster 6

Attributes	Costs								
	Sub structure	Super Structure	Internal Finishes	Fittings	Services	External Works	Prelimin aries	Continge ncies	Contract cost
GFA	0.86	0.96	0.94	0.40	0.92	0.88	0.69	0.25	0.92
House	0.83	0.94	0.90	0.32	0.87	0.86	0.66	0.25	0.89
Duration	0.67	0.65	0.66	0.24	0.69	0.57	0.55	0.06	0.65
Contract	0.41	0.31	0.35	0.18	0.38	0.28	0.39	-0.01	0.33
Bungalow	0.27	0.21	0.30	0.71	0.29	0.27	0.40	0.14	0.29
Stories	0.30	0.18	0.22	-0.07	0.22	0.28	0.12	0.01	0.24
Type of floors	0.29	0.18	0.21	0.62	0.22	0.15	0.29	0.06	0.22
Cottage	0.19	0.04	0.11	-0.03	0.07	0.03	-0.06	-0.12	0.05
S. Condition	0.08	0.10	0.04	-0.11	0.07	0.02	0.14	0.09	0.04
Location	-0.05	0.07	0.04	0.12	0.08	-0.06	0.19	0.20	0.02
Fluctuation	0.02	-0.01	0.02	0.03	0.02	-0.01	-0.02	0.07	0.00
Space	-0.08	0.01	0.06	0.06	0.01	-0.06	0.08	-0.04	-0.02
Windows	-0.07	0.00	-0.02	0.18	-0.01	-0.16	0.00	-0.04	-0.05
Topography	0.06	-0.02	0.01	-0.10	-0.05	0.03	-0.11	-0.09	-0.07
Foundation	-0.06	-0.13	-0.14	-0.10	-0.14	-0.01	-0.10	-0.12	-0.08
G. Condition	-0.15	-0.02	0.02	0.07	0.04	-0.07	-0.05	0.04	-0.09
Access	-0.12	-0.07	-0.02	0.04	-0.09	-0.14	0.01	-0.05	-0.10
N. Tenders	-0.09	-0.04	-0.10	-0.10	-0.06	-0.06	-0.23	0.05	-0.11
Market Condition	0.06	0.05	0.09	0.00	0.01	-0.01	0.06	0.20	-0.18
Heating	-0.07	-0.21	-0.21	-0.09	-0.18	-0.25	-0.23	-0.19	-0.18

data may classified by the ANN and statistical techniques usefully represented to assist in the cost decision-making process. While the utilization of the SOM has been demonstrated in other fields, the use in construction cost data representation and in accounting for noisy cost data is both novel and important.

The main contribution of the paper is that it addresses the serious issue of noise within construction cost data and how a combination of SOM and statistical analysis techniques might usefully account for it. The SOM analysis visually highlights the problem of variability by showing how interdependencies between attributes and costs can change class across the map. The SOM not only highlights the problem of noisy data but also provides a novel way of addressing the challenge of detecting significant cost indicators; selecting the most important cost attributes; developing if-then rules; developing accurate forecasting models, developing cost performance indicators;

providing guidance on costing projects with similar characteristics and informing clients about potential cost features to be considered in the early stages of a project development.

CONCLUSION

The cost data of 200 projects were classified and segmented into different clusters. Then, statistical techniques were used to analyse the results of these clusters. The findings of these clusters were briefly analysed and described. The results show that there is a considerable noise across the clusters especially in the relationship between cost and duration of projects. The relationship between floor area and cost is linear with the exception of cluster 4, which was polynomial. It is not surprising to find the noise between costs and floor area is less than 20%. This due to the fact that many estimators use the floor area as a guide for estimating future costs. The results also showed the variability of the strength of the correlation between cost attributes and costs across the clusters

ACKNOWLEDGEMENTS

This work is financially supported by EPSRC (Grant GR/ K/85001)

REFERENCES

Davis Langdon & Everest (1998) *Framework for national construction cost* database, Final report to the Department of the Environment, Transport and the Regions.

Dietrich, E., and Thomas Wittmann (1998) The use of fuzzy logic in dependency analysis. *In: Procs of the 6th European Congress on Intelligent Techniques and Soft Computing*, Aachen, Germany, September, Vol. 2. pp 1170–1174.

Hosheimer, M. and Siebes A. (1996) *Data mining the search for knowledge in databases.* Technical Report CS-R94406, CWI, Amsterdam, Netherlands

Kohonen, T. (1995) *Self-organizing maps*. Berlin, Heidelberg: Springer.

O'Leary, D.E. (1988) Knowledge-management systems: converting and connecting. *IEEE Intelligent Systems*. May/June. pp30–33

Thomas, W. and Johannes (1998), Aspects of rule understandability in neuro-fuzzy data mining. *In: Procs 6th European Congress on Intelligent Techniques and Soft Computing*. Aachen, Germany, September, Vol. 2. pp 1165–1169.