Clustering Analysis on Questionnaire Data for Program Accreditation

George Papadourakis[#1], Fernando Maciel Barbosa[*2], Cyril Burkley*[3], Michael Hoffmann[##4], Daniel Pasquet[**5], Georgios Tsiriotis**[6], Ioannis Kiriakides[#7]

[#] *Department of Applied Informatics and Multimedia, Technological Educational Institute of Crete*
Heraklion, Crete, Greece
[1]papadour@cs.teicrete.gr
[7]kyriakidis@teicrete.gr

[*] *Faculdade de Engenharia da Universidade do Porto, INESC Porto, Porto, Portugal*
[2]fmb@fe.up.pt

**University of Limerick, Ireland*
[3]cyril.burkley@ul.ie

[##]*Universität Ulm, 89069 Ulm, Germany*
[4]michael.hoffmann@uni-ulm.de

[**]*ENSEA, 95014 Cergy, France*
[5]Daniel.Pasquet@ensea.fr

***Electrical Engineering Department, Technological Educational Institute of Kavala*
Agios Loukas, 654 04, Kavala, Greece
[6]tsirigo@teikav.edu.gr

Abstract—**The aim of the SOCRATES EIE-Surveyor project is to be a reference point for Electrical and Information Engineering in Europe, bringing together representatives from 27 out of 31 eligible countries. One of the tasks of the project is the evaluation of the accreditation processes in the participating countries. A questionnaire about the accreditation process was developed and sent to project partners in each participating country. The main areas investigated the nature of the accreditation body, the criteria, which are evaluated, the structure of the visit and the decision formulation. The results of the questionnaire, will be analyzed using clustering analysis and more precisely hierarchical, in order to compare the answers in 17 European countries and to find similarities among them. As distance measures the Euclidian metric and the City block distance will be used. Average linkage, and Ward clustering algorithms will be utilized.**

Index Terms— **accreditation, evaluation, clustering**

I. Introduction

The first moves towards formal engineering education in Europe began around the middle of the 18th century initially in France, but within a short space of time engineering schools were established in much of Europe [1]. In due course universities across the world established engineering programmes based on the European models. In mainland Europe, the duration and structure of engineering programmes were based on a programme of studies of four or five years duration and firmly grounded in mathematics and the sciences. Initially in the UK and Ireland programmes were generally of three years duration. The structure in the UK has evolved into a four years Masters of Engineering degree programme, while in Ireland, a four year Bachelor degree has been in place for nearly 50 years.

In June 1999 the Bologna Declaration [2] was published and its overall objective was the establishment of a European area of higher education in which student mobility would be facilitated and enabled. A follow-up conference in Prague [3] highlighted the important role that quality assurance systems play in ensuring high quality standards and in facilitating the comparability of qualifications throughout Europe. Prior to this the recognition or accreditation of qualifications was done largely on a national basis and, within individual countries, recognition or accreditation of programmes of study could take place at either institutional, national or the professional level. However, since the Bologna Declaration, the need for European wide recognition and accreditation of higher education programmes and their relationship to quality assurance are at present the subject of many discussions and activities in Europe [4].

Under the auspices of FEANI, the European Federation of National Engineering Associations, a group of national associations involved in accreditation submitted a proposal to the European Commission to set up the EUR-ACE (EURopean ACcredited Engineer) project with the objectives of ensuring consistency between existing national engineering accreditation systems and establishing a European "quality label" for accredited programmes [5]. Following the successful completion of the EUR-ACE project, the partners established ENAEE (European Network for Accreditation of Engineering Education) to develop policies and procedures whereby professional accreditation agencies in Europe will be authorised to add the EUR-ACE label to their accreditation [6].

An earlier Thematic Network project THEIERE [7], conducted a preliminary study of existing accreditation procedures in the field of electrical and information engineering across a range of universities in Europe. This work has been extended in the EIE Surveyor Thematic Network project [8], which by means of a questionnaire has collected material on the existing accreditation processes and procedures in Europe. Some early results of this work have already been presented [9] and this paper presents a detailed analysis of the Questionnaire Data obtained.

II. Questionnaire

A. Context of the questionnaire

The starting point of the accreditation task in Surveyor project was the EUR-ACE project (European accreditation of European Engineering and graduates) [5], which was a consortium of 14 partners, supported by the European Commission. The objectives of the EUR-ACE project were (i) to ensure consistency between existing national engineering accreditation systems, (ii) establish a European "quality label" for accreditated programmes and (iii) assist with the establishment of accreditation in European countries where it does not yet exist, thus improving the quality of engineering education, facilitating transnational recognition and mobility of engineering graduates.

The aim of the EIE Surveyor task was to see how the EUR-ACE results could apply to the field of electrical and communication engineering.

The EUR-ACE project evaluated the various factors that should be taken into consideration when assessing an engineering programme. These have been used as the guideline for constructing the questionnaire. They are gathered into six domains.

1) General information and curriculum

The general points concerning the curriculum are:

- Identification of educational goals
- Profile of the programme
- Duration, workload, ECTS

A difference must be made between the duration of courses, tutorials and practical works and the actual workload which includes the personal unsupervised study time of the students.

- Teaching methods
- Programme structure
- Programme content
- Number and duration of internships or workplacements

The internships may be in academic laboratories or in industry.

2) Professors and academic staff

- Teaching staff (number, specialisation, qualification)

The ratio between professors and other academic staff is considered.

- Academic staff – student ratio
- Technical and support staff
- Research activities of staff
- Professional activities and consultancy

3) Admission and educational standards

- Admission requirements

Students may be admitted to the programme on the basis of a general national or state examination or by a selective entrance examination.

- Assessments of demand for the programme
- Assessments of student performance
- Student performance

The performance must be evaluated according to ECTS criteria. The distributions of the results among the different grades may be evaluated.

- Graduate employment opportunities

4) Quality assurance measures and development

- Quality assurance measures
- Plans for the future development of the programme

5) Institutional context

- General requirements (organizing, management,…)
- Cooperation with Higher Educational Institutions
- Industry cooperation
- Finances
- Facilities

6) Internationalisation

- Study abroad opportunities
- International co-operations

The international co-operations between two institutions consist of student and teachers mobility.

- Foreign language requirements and education
- Subject or specific classes taught in foreign languages

Many institutions propose some courses in English and a few have a full curriculum in English.

B. Questionnaire content

It was considered important that the questionnaire evaluated how the EUR-ACE criteria are considered by the institutions during the accreditation process. The goal was to have some complementary information specific to EIE field.

The questionnaire was constructed so that it could be completed in a quick and straight forward manner. It was sent to one partner in each participating country. Where a country has several accreditation bodies, several questionnaires were sent. The questions were divided into four sections.

1) Accreditation body

- Is accreditation compulsory to deliver engineering degrees in EIE?
- Is the accreditation awarded by the government, the university, a professional body or some other agency?
- Is the accreditation awarded to a programme, a department or the whole institution?
- Does the accreditation body include faculty, employers, engineers in industry?
- Does the accreditation process include quality assurance measures?

2) Parameters measured

A number of different parameters can be considered during the accreditation process. For each of them the questionnaire asked whether it is evaluated and if documentation is provided in advance or during the visit. The parameters listed in the questionnaire were:

- Curricula
- Examination papers
- Student examination scripts
- Projects reports and thesis
- Students' performance
- Employment of graduates
- Academic staff
- Recruitment
- Research activities
- Collaboration with industry
- Facilities

3) Evaluation visit

In general the accreditation body sends a visiting panel in the institution to be reviewed. In order evaluate the visiting process, the following questions were asked:

- What is the frequency of the visits?
- What is the size of the visiting panel?
- What is the composition of the visiting panel (academics, industrial, others)?
- What is the duration of the visit?
- Whom does the panel meet during the visit?
  - students
  - academic staff
  - technical staff
  - administrative staff
  - employers
  - graduates

4) Conclusions

On the completion of the visit, the visiting panel in general gives a verbal presentation of their findings to the staff in the institution visited. Subsequently a report is written which includes a recommendation on the accreditation. In order to evaluate how the conclusions are processed the followed questions were asked.

- To whom do the review panel report (government, university, professional body, agency)?
- Who makes the final decision (government, university, professional body, agency)?
- What are the different possible decisions?
  - full accreditation
  - accreditation for reduced period of time
  - no accreditation
  - additional non-compulsory recommendations

At the end of the questionnaire, the participants were also invited to add any further comments they may wish to make.

C. First analysis

A first analysis shows that some countries have not yet introduced a formal accreditation process. These countries are generally in a transition situation in relation to introducing the Bologna process. The accreditation process, ECTS and the quality assurance measures will probably be introduced at the same time.

In some other countries several accreditation bodies exist depending on the region (in Germany according to the Länder) or the nature of the institution (in France between universities and Grandes Ecoles). It also appears that the accreditation for masters and PhD degrees is not yet compulsory everywhere.

Other issues regarding the accreditation process that are also being considered include the payment of the expenses in relation to the accreditation process. This point is important in the countries where the accreditation process is not paid by government. Finally the issue of whether industrial placement is compulsory and for how long must it last is being reviewed.

III. Clustering and Data Encoding

Cluster analysis or clustering is the classification of objects (patterns) into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset are similar according to some defined distance measure. Central to all of the goals of cluster analysis is the notion of degree of similarity (or dissimilarity) between the individual patterns being clustered. Data clustering is a common technique for statistical data analysis. The patterns are given in the form of feature vectors containing elements that describe in numeric form objects or events.

In this study the objective is to discover similarities among countries so each questionnaire answered is a distinct pattern. The feature vector for each questionnaire is formed by encoding numerically the answers to the questionnaire using various techniques.

TABLE 1
ENCODING OF QUESTIONNAIRE

		Desciption	Encoding					Inputs
1		**The accreditation body**						
	a	yes	1					1
		no	0					
	b	Government	1	0	0	0	0	5
		University	0	1	0	0	0	
		Professional body	0	0	1	0	0	
		Independent body	0	0	0	1	0	
		International agency	0	0	0	0	1	
	c	a programme	0,2					1
		a department	0,5					
		the whole institution	1					
	d							
	e	faculty	1	0	0	0	0	5
		employers	0	1	0	0	0	
		engineers in industry	0	0	1	0	0	
		accreditation body	0	0	0	1	0	
		Others	0	0	0	0	1	
	f	yes	1					1
		no	0					
	g	International	1	0	0			3
		European	0	1	0			
		EUR-ACE	0	0	1			
	h	Yes	1					1
		No	0					
2		**The Criteria**						
	a	11 choices accumulated In 3 columns (0-1)						3
3		**The evaluation visit**						
	a	2 years	0.25					1
		5 years	0.63					
		Other (normalized 8 y=1)						
	b	1	0.14					1
		2	0.29					
		3	0.43					
		Other (normalized 7=1)						
	c	Academics	1	0	0			3
		Industrial	0	1	0			
		Other	0	0	1			
	d	1 day	0.25					1
		2 days	0.50					
		3 days	0.75					
		longer	1.00					
	e	6 choices accumulated(0-1)						1
4		**The conclusion**						
	a	Government	1	0	0	0	0	5
		University	0	1	0	0	0	
		Professional body	0	0	1	0	0	
		Independent body	0	0	0	1	0	
		International agency	0	0	0	0	1	
	b	Government	1	0	0	0	0	5
		University	0	1	0	0	0	
		Professional body	0	0	1	0	0	
		Independent body	0	0	0	1	0	
		International agency	0	0	0	0	1	
	c	full accreditation	1.00					1
		reduced accreditation	0.50					
		no accreditation	0.00					
		additional non-compulsory recommendations	0.70					

Table 1 illustrates the encoding utilized. Each multiple choice question is encoded as a binary input or a real number between 0-1. In most cases, for example questions 1b), 1g) etc, the use 1-of-C coding is utilized. The number of input is determined by the possible choices of a question. Each choice is given the value zero except for the one corresponding to the correct one, which is given the value one. In other cases, such as questions 1c), 3a), 3b) a real number between 0-1 can represent the answer, and only 1 input is needed. Finally, in multiple choice questions, where the answers could be several categories, such as questions 2a) and 3 e) the total number of the selected categories is accumulated and normalized between 0-1. There was one question 1d) where the encoding was not possible and it was not used as input. In total, as shown in Table 1, 38 inputs formed the feature vector for each questionnaire.

Using the above encoding scheme the feature vector of each questionnaire was formed. However, there were several problems encountered with missing answers in the questionnaires. According to the nature of the question different strategies were used to resolve the problem. One of the most common problems encountered was the missing answers to a yes, no question, such as 1a), 1f), etc. In this case the value 0.5 was used. In questions such as 1c), 3a), 3b) where a real number between 0-1 can represent the answer, and only 1 input is needed, the value 0 is reserved for no answer.

Finally, analyzing the answers to al the questionnaires, in question 4a) nobody choose “international agency”, so this input was eliminated. Similarly, in question 4b) the inputs for choices “University” and “International Agency” were eliminated.

Data clustering algorithms can be hierarchical or partitional [10]. Hierarchical algorithms find successive clusters using previously established clusters, whereas partitional algorithms determine all clusters at once. Hierarchical algorithms can be agglomerative ("bottom-up") or divisive ("top-down"). Agglomerative algorithms begin with each element as a separate cluster and merge them into successively larger clusters. Divisive algorithms begin with the whole set and proceed to divide it into successively smaller clusters. Hierarchical clustering may be represented by a two dimensional diagram known as dendrogram which illustrates the fusions or divisions made at each successive stage of analysis. An example of such a dendrogram is shown in Fig. 1.

An important step in any clustering is to select a distance measure, which will determine how the similarity of two elements is calculated. The most common distance measure, which will be used in this paper, is the Euclidean distance. The Euclidean distance between feature vectors $\underline{x}$ and $\underline{y}$ is given by:

$$d_E\left(\underline{x},\underline{y}\right)=\sqrt{\sum_i\left(x_i-y_i\right)^2} \tag{1}$$

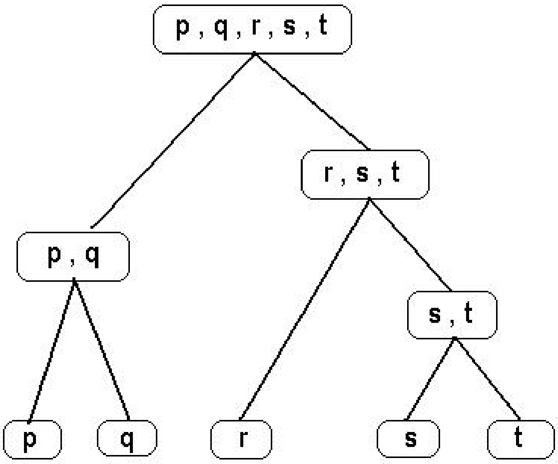

Fig. 1. Example of dendogram

Another popular distance measure, which is utilized on integer values, and suitable to the data presented in this paper, is the City block distance also known as Manhatan distance or Taxi distance. The City block distance between feature vectors x and y is given by:

$$D_C\left(\underline{x},\underline{y}\right)=\sum_i \left|x_i-y_i\right| \tag{2}$$

In this paper agglomerative hierarchical clustering is utilized using the Euclidean and City block distance. At each particular stage the method joins together the two clusters which are closest together (most similar). Differences between methods arise because of the different ways of defining distance (or similarity) between clusters.

One of the most common agglomerative hierarchical clustering methods is **Average linkage**. The distance between two clusters is defined as the average distances between a point in one cluster and a point in the other cluster.

Ward's hierarchical clustering method minimizes the loss associated with each cluster. At each step in the analysis, among all pairs of clusters, it merges the pair that produces the smallest **squared error** for the resulting set of clusters, resulting in minimum increase in information loss. Information loss is defined by Ward in terms of an error squared error criterion. The squared error for a cluster is the sum of the squared distances in each element from the cluster mean. The squared error is thus equal to the total variance of the cluster times the number of elements in the cluster. The squared error for a set of clusters is defined to be the sum of squared errors for the individual clusters.

Each agglomeration occurs at a greater distance between clusters than the previous agglomeration, and one can decide to stop clustering either when the clusters are too far apart to be merged (distance criterion) or when there is a sufficiently small number of clusters (number criterion).

IV. Clustering Results

Hierarchical clustering was performed on the encoded data of the questionnaires in order to discover similarities among countries concerning accreditation procedures. The best results, in the Mean Square Error Sense, were the average linkage algorithm and the Ward's algorithm utilizing either Euclidean or City block distance.

The results of clustering using the average linkage algorithm and City block distance are shown in Fig. 2.

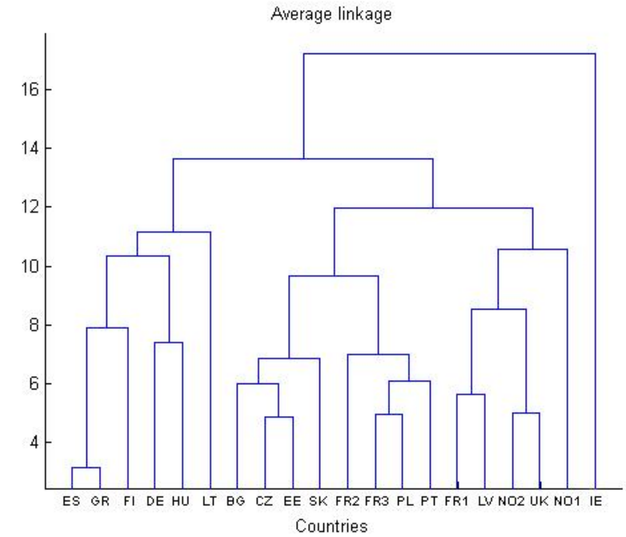

Fig. 2. Results of average linkage algorithm using City block distance

The results of clustering using the average linkage algorithm and Euclidean distance are shown in Fig. 3.

The results of clustering using Ward's algorithm and City block distance are shown in Fig. 4.

The results of clustering using Ward's algorithm and Euclidean distance are shown in Fig. 5.

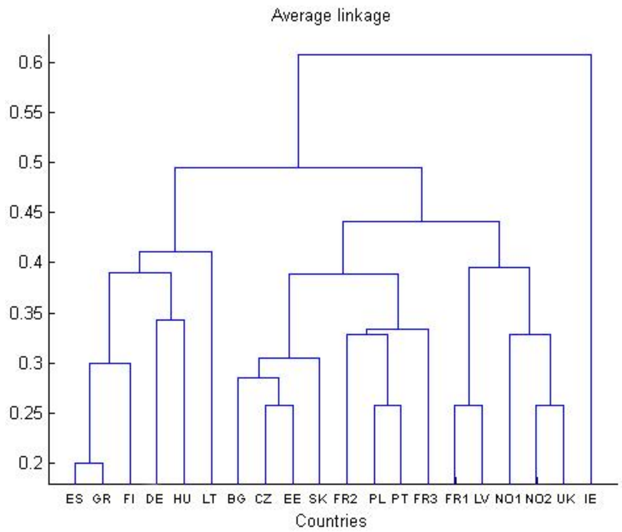

Fig. 3. Results of average linkage algorithm using Euclidean distance.

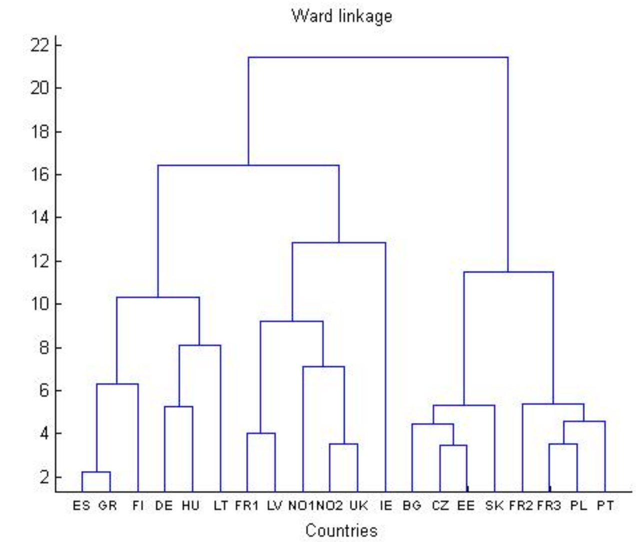

Fig. 4. Results of Ward's algorithm using City block distance.

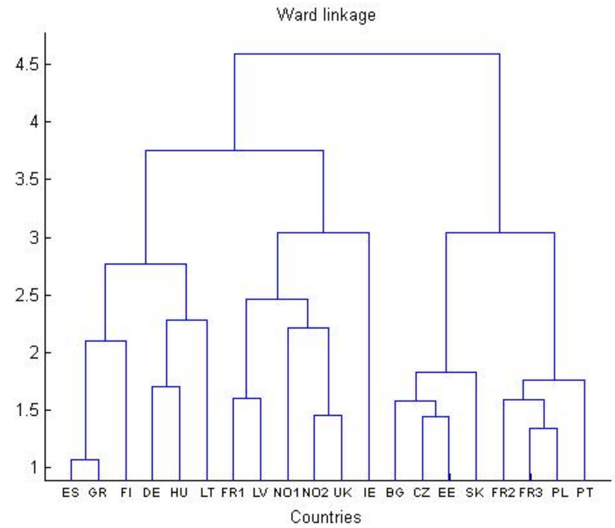

Fig. 5. Results of Ward's algorithm using Euclidean distance.

Analyzing the results we can see that results were similar in all cases and we note the formation of three clusters:

- Spain, Greece, Finland, Germany, Hungary and Lithuania.
- Bulgaria, Czech Republic, Estonia, Slovakia, France(2 and 3), Poland and Portugal.
- France(1), Latvia, Norway(1 and 2), United Kingdom and Ireland.

The algorithms utilized, Average and Ward's, gave the same results for both distance measures, City block and Euclidean. The only difference between the two algorithms is the assignment of Ireland. The average algorithm shows that Ireland is different from all the other countries while Ward's algorithm assigns Ireland to cluster 3. One final notice is that the clustering results show the questionnaires in France to belong to two different clusters. This result can be explained by the fact that accreditation in France is done by different agencies and the results reflect the different approach to accreditation among these agencies.

V. Conclusion

In this paper the analysis of a survey about the accreditation process in 17 European countries was performed.

A first analysis shows that some countries have not yet introduced a formal accreditation process. These countries are generally in a transition situation in relation to introducing the Bologna process. The accreditation process, ECTS and the quality assurance measures will probably be introduced at the same time.

In some other countries several accreditation bodies exist depending on the region (in Germany according to the Länder) or the nature of the institution (in France between universities and Grandes Ecoles).

Cluster analysis showed the formation of three groups:

- Spain, Greece, Finland, Germany, Hungary and Lithuania.
- Bulgaria, Czech Republic, Estonia, Slovakia, France(2 and 3), Poland and Portugal.
- France(1), Latvia, Norway(1 and 2), United Kingdom and Ireland.

References

[1] James Dooge, "Engineering Education and Training", Ronald Cox (ed) p.36, Engineering Ireland, Cork, Collins Press, 2006.

[2] "The European Higher Education Area", Joint Declaration of the European Ministers of Education, Bologna, June 1999.

[3] Conference of the European Ministers for Education, Prague 2001.

[4] Giuliano Augusti "Accreditation of Engineering programmes: European perspectives and challenges in a global context", European Journal of Engineering education, Vol. 32, No. 3, pp.1-11, June 2007.

[5] EUR-ACE Project 2004, details on the FEANI website, www.feani.org/EUR_ACE.

[6] Giuliano Augusti "Bologna Process, quality and accreditation of engineering educational programmes in Europe: the EUR-ACE label, the ENAEE network", JABEE Symposium/Workshop "Global Trend and Perspectives on the Quality Assurance of Engineering Education", Tokyo, Japan, December 2006.

[7] THEIERE Project 2002-4, details on EAEEIE website, www.eaeeie.org/theiere.

[8] EIE Surveyor Project 2005-8, details on EAEEIE website, www.eaeeie.org/EIESurveyor.

[9] FM Barbosa, C Burkley, M Hoffmann and JM Thiriet, "Accreditation of higher education in EIE in Europe", 18th International EAEEIE Conference, Prague, July 2007.

[10] E. Gose, R. Johnsonbaught, S. Jost, *Pattern recognition and image analysis*, Prentice Hall, 1996.