A METHODOLOGY FOR DETERMINING THE CLUSTER OF A NEW PROJECT

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

AYBENİZ YİĞİT

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

JUNE 2005

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Canan ÖZGEN
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science.

Prof. Dr. Çağlar GÜVEN
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Canan SEPİL
Co-Supervisor

Prof. Dr. Ömer SAATÇİOĞLU
Supervisor

Examining Committee Members

Assoc. Prof. Dr. Gülser KÖKSAL (METU, IE)

Prof. Dr. Ömer SAATÇİOĞLU (METU, IE)

Assoc. Prof. Dr. Canan SEPİL (METU, IE)

Assoc.Prof. Dr. Yasemin SERİN (METU, IE)

Elif BAKTIR (MS, ASELSAN)

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Aybeniz, Yiğit

Signature :

ABSTRACT

A METHODOLOGY FOR DETERMINING THE CLUSTER OF A NEW PROJECT

Yiğit, Aybeniz
M.S., Department of Industrial Engineering
Supervisor: Prof. Dr. Ömer Saatçioğlu
Co-Supervisor: Assoc.Prof. Dr. Canan Sepil

June 2005, 105 pages

By definition, all projects are unique; however R&D projects have specific characteristics that make them harder to manage. The project management methodology applied to R&D projects may show differences due to the categorization of them. But if there exists a categorization of projects, one can analyze the properties of the project classes and then manage similar projects similarly.

In this study, the R&D projects of a main military electronics company of Turkey, are analyzed. Tunç (2004) has developed a methodology for clustering the projects of this electronics company. Continuing from his studies, a methodology for determining the class of a new project of this electronics company is developed.

For defining the projects in a project space, a Project Identification Card (PIC) is developed. The measurement scale of the PIC is constructed by using the absolute measurement Analytic Hierarchy Process. A clustering Tabu Search algorithm is generated for using in the sensitivity analyses of the clusters to projects. And a methodology for determining the cluster of a new project is developed.

Keywords: Analytic Hierarchy Process, clustering R&D projects, absolute measurement AHP, Tabu search, sensitivity analysis, project identification card, classification of a new project.

ÖZ

YENİ BİR PROJENİN SINIFININ BELİRLENMESİ İÇİN BİR METODOLOJİ

Yiğit, Aybeniz

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Ömer Saatçioğlu

Ortak Tez Yöneticisi: Assoc.Prof. Dr. Canan Sepil

Haziran 2005, 105 sayfa

Proje kelimesi tanım olarak her projenin tek olduğunu tariflemektedir. Fakat araştırma ve geliştirme projelerinin kendilerine ait özellikleri yönetimlerini zorlaştırmaktadır. Araştırma ve geliştirme projelerine uygulanan proje yönetim metodolojisi, projelerin kategorilerine göre değişiklikler gösterebilmektedir. Eğer projelerin bir sınıflandırması yapılabilirse, proje sınıflarının özelliklerine göre benzer projeler benzer şekilde yönetilebilir.

Bu çalışmada, Türkiye'deki bir askeri elektronik şirketinin araştırma ve geliştirme projeleri incelenmiştir. Tunç (2004) aynı şirketin projelerinin sınıflandırılması için bir metot geliştirmiştir. Bu çalışmalardan devam ederek, yeni bir projenin sınıfının belirlenmesi için bir metodoloji geliştirilmiştir.

Projeleri bir proje uzayında tanımlamak için, bir Proje Kimlik Bilgisi kartı (PKB) geliştirilmiştir. Geliştirilen PKB'nin ölçüm cetveli, tam ölçümlü Analitik Hiyerarşi Metodu kullanılarak oluşturulmuştur. Proje sınıflarının projelere göre hassasiyetlerini analiz etmek için bir sınıflandırma Tabu arama algoritması geliştirilmiştir. Ayrıca yeni bir projenin sınıfının belirlenmesi için bir metodoloji oluşturulmuştur.

Anahtar Kelimler: Analitik Hiyerarşi Metodu, araştırma ve geliştirme projelerinin sınıflandırılması, tam ölçümlü Analitik Hiyerarşi Metodu, Tabu arama, hassaslık analizi, Proje Kimlik Bilgisi kartı, yeni bir projenin sınıflandırılması.

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor Prof. Dr. Ömer Saatçioğlu for his patience, encouragements, guidance, and invaluable suggestions throughout this study. I am indebted to my thesis co-advisors Assoc. Dr. Canan Sepil and Assoc. Dr. Yasemin Serin for their guidance, patience, motivations, helps and invaluable contributions.

I would like to express my gratitude and sincere thanks to my manager, Ms. Elif Baktır for her motivation, guidance and patience throughout all my master studies. I would also like to thank to her, for her invaluable advises during my course registrations, for the hours in which we studied together to determine a thesis subject. It would be impossible for me to complete my thesis studies without her understanding, motivation and helps.

I would like to thank to my dear colleague Mehmet Tunç for his support, suggestions, and his valuable thesis study, which forms the beginning point of my thesis studies.

Moreover, I would like to thank to my old colleagues, Mustafa Onur Kabul and Feyzullah Aykut Özsoy for their motivations, suggestions, friendship and belief in me in the beginning phase of my thesis studies.

I would like to thank to my colleagues, Taner Gülez, İnci Yüksel and Gözde Çakıcı for their helps, support, friendship during my work time and for the times we were together.

Finally, I would like to express my gratitude to my mother İclal Yiğit, my father Ekrem Yiğit, my brothers; Zafer Yiğit and Tamer Yiğit, their wifes; Şükran Yiğit and Ayça Yiğit, and my nephews; Pelin Yiğit and Ümitcan Yiğit, for their understanding, motivation, support and most importantly for their love.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Today's rapid changes in market needs and the acceleration in the speed of technology have increased the importance of R&D activities. One can observe R&D activities in almost all industries, e.g. pharmaceutical industry, defense industry. Although the scopes of products in these different industries may differ, the common characteristics of R&D activities such as uncertainty, risk and the importance of managing innovation and the concept of the speed to market are highlighted in all industries.

The defense industry has all these common characteristics of the R&D. But what differs the defense industry from the others is the nature of their products. The military products are complex technical systems.

Having long research and development cycles (usually five years or more), requiring the coordinated work of many engineers from different disciplines are the main properties of complex technical projects. Considering these properties, the importance of project management methodology applied for these projects, become obvious.

Project management methodology has many knowledge areas, defined in the Guide to the Project Management Body of Knowledge of Project Management Institute (2000). These knowledge areas are project integration management, project scope management, project time management, project cost management, project quality management, project human resource management, project communications management, project risk management and project procurement management.

By definition, all projects are unique; however R&D projects have specific characteristics that make them harder to manage. The project management methodology applied to R&D projects may show differences due to the categorization of them. But if there exists a categorization of projects, one can analyze the properties of the project classes and then manage similar projects similarly.

In this study, the R&D projects of one of the main military electronics company of Turkey, are analyzed. Tunç (2004) has studied the clustering of the R&D projects of this electronics company using AHP. He defined the projects in a five dimensional space using five features and compared the projects pairwisely according to each feature. The project clusters are the main output of his study.

Continuing from his studies, a methodology for determining the class of a new project of this electronics company is studied in this thesis. After forming the project clusters there comes a new question of how to determine the class of a new project.

Tunç's approach requires a large number of pairwise comparisons. For the case of 14 projects and 5 project features, the number of pairwise comparisons is (91x5)=455. Each time a new project comes, this number increases, making impossible to evaluate all the projects by a decision maker consistently. In order to determine the feature values of a new project, the decision maker should know all the previously clustered projects characteristics. Also the feature values of the previously clustered projects and the clustering structure all together change each time a new project is accepted, making his approach impractical and not reliably applicable in the long run.

In order to develop a dynamic classification scheme, a Project Identification Card (PIC) is introduced. This card has the project features of Tunç (2004) as the main dimensions. The scale of this PIC is constructed by using the absolute measurement AHP method. The clustering structure of this PIC is analyzed and an approach for determining the class of a new project is proposed. Considering the need for defining an objective function for clustering and analyzing the changes in the objective function under different circumstances, a Tabu search clustering algorithm is developed and used for clustering the projects.

This chapter includes the general information about the current project management state of this electronics company, the need for project classification, the purpose of the thesis study and the outline of the thesis.

1.1 THE CURRENT PROJECT MANAGEMENT STATE IN THIS ELECTRONICS COMPANY

The projects are performed in a multi-project environment with multidiscipline project teams. Before signing the contract of a project, business development activities are performed. In these business development activities a core development team is constructed and the initial concept design studies are performed. Initial concept design, total cost, total labor hour requirement, and the work breakdown structure of a project are the main outputs of these business development activities.

The project management activities of the projects are performed by using customer requirements, contract obligations, project specific characteristics and expert knowledge about past projects. The project leaders and technical leaders are chosen according to their experiences in similar projects. But there are no formal definitions of what criteria set determines the similarity of projects, formal definitions of project groups and similar project sets. Also there are no formal definitions of project characteristics or project management highlights for similar projects that can be grouped, and a method for assigning a new project to a project group.

The need for project classification is realized by the managers of this electronics company and there are on-going project classification studies. Five features are defined to be used for classifying the projects. These features are also used by Tunç (2004).

1.2 THE NEED FOR PROJECT CLASSIFICATION

The need for project classification arises from two points of views. The first one is the project management view. Project management activities like project quality management, resource management, risk management, team formation should be performed according to project specific properties. For example, for a new design project, more importance should be given to the risk management concept than for a continuous improvement project.

The second one is the project performance measurement view for evaluation of projects. In performance measurement literature the determination of project success factors and the metrics that are used to monitor and to evaluate the project success is widely studied. Many measurement frameworks and metrics are proposed. The importance of defining project success factors correctly, selecting and developing the metrics that are used to evaluate the project in terms of its success factors and monitoring the metrics throughout the project is emphasized. The success factors of projects are different from each other, since each project has different characteristics and similar projects should have similar success factors. By grouping projects, it will be easier to define the right success factors and the metrics.

In performance measurement there are three important questions to be answered;

1. *What to measure?:* Determining the success factors of the project. To evaluate the project as *a successful project,* from which points of views the project should be evaluated?
2. *How to measure?:* Selecting and developing the metrics that will be measured and used for evaluating the success factors of a project. Which measures are to be used for evaluating the success factors of a project?
3. *How the measurement results are evaluated?:* Constructing the definitions of good and bad performance. With what the measurement results are to be compared?

The most vital part of performance measurement is the interpretation of the measurement results. Same measurement results of two different projects do not mean that the two projects are performing the same according to this measurement view. The same measurement result can be good for one project, whereas the same measurement result can be bad for the other project. Also the importance of the same measurement can be different for these two projects. Therefore for evaluating the measurement results correctly, two important information regarding project specific properties is needed. The first one is the importance of the measurement for evaluating the project success and the second one is the value to which the measurement result is to be compared with for determining the success level of the project.

Regarding project management and performance measurement views there come the need for grouping similar projects and treating the projects in the same group similarly. After the project groups are obtained, the projects within the same group should be analyzed together and two important types of information (identification type information and measurement type information) about a project group should be developed.

Complexity levels, resource requirements, project team experience level, and budgets of projects can be the examples of identification type information about project groups.

Statistics about engineering change orders, requirements development, documentation, and system/subsystem integration test results can be the examples of measurement type information about project groups.

The main steps of project classification studies for this electronics company can be listed as follows:

i. Developing project clusters.
ii. Constructing a method for determining the characteristics of a project in a structured way.
iii. Constructing the characteristic properties of project clusters by analyzing the projects in the same class.
iv. Constructing a method for assigning a new project to the developed project clusters.
v. Developing project management rules specific to each project cluster and managing similar projects similarly.
vi. Analyzing the measurement data of the projects in the same project cluster and developing cluster standards.
vii. Conducting performance measurement of projects in the same project cluster by comparing their measurement results with its project cluster standards for evaluation of projects.

After determining the project clusters, the new problem that should be solved is determining the class of a new project. The developed method for solving this problem should satisfy the following requirements:

i. It should be easy enough to apply every time a new project comes.
ii. It should not cause the reforming of the developed project clusters upon arrival of a new project.

iii. Constructing the feature values of the new project should not necessitate the expertise in all the past projects, for an expert.
iv. It should provide a method for quantifying the features of a new project in a structured way.

In this study, a methodology for determining a meaningful class for a new project, taking into consideration the above stated requirements is studied.

The developed method has three main parts. The first part is the quantification method of features of a new project; the second one is analyzing the application of the developed quantification method, and the third part is the construction of a classification method by using the developed quantification method.

1.3 OUTLINE OF THE THESIS

In Chapter 2, the literature about cluster analysis and classification, quality assessment in clustering, project classification, Analytic Hierarchy Process Method, the AHP clustering, Vector Space Formulation of the AHP (VAHP) clustering are mentioned. Since Tunç's work is taken as a starting point for this study, a detailed description of his work is given in Chapter 2.

In Chapter 3, the proposed approach for representing the projects in a multi dimensional space, the development of Project Identification Card, the mathematical model of clustering the projects, Tabu search clustering algorithm, sensitivity of the clusters to the projects, and the proposed methodology for determining the cluster of a new project and its application are mentioned. Chapter 4 covers the conclusions derived out from this study and the recommendations about the future studies.

CHAPTER 2

LITERATURE SURVEY

The literature survey includes the studies in the cluster analysis and classification, the quality assessment in clustering, project classification, Analytic Hierarchy Process (AHP) Method, the absolute measurement in AHP, the AHP and the clustering, Vector Space Formulation of the AHP (VAHP) based clustering and Tunç's work.

2.1 CLUSTER ANALYSIS AND CLASSIFICATION

Cluster analysis term is most commonly used for techniques those are used to separate data into consistent groups (Everitt, 1974). Cluster analysis is commonly applied in many engineering and scientific disciplines such as pattern recognition, data mining, biology, psychology, medicine and marketing.

Everitt (1974) provides following definitions of a cluster:

1) "A cluster is a group of contiguous elements of a statistical population; for example, a group of people living in a single house, a consecutive run of observations in an order series, or a set of adjacent plots in a field."
2) "A cluster is a set of entities which are alike, and entities from different clusters are not alike."
3) "Clusters may be described as continuous regions of a multidimensional space containing a relatively high density of points, separated from other such regions by regions containing a relatively low density of points."

Cluster analysis has many objectives; according to Ball (1971) the objectives of cluster analysis are as follows:

i. Finding a true typology,
ii. Model fitting,
iii. Prediction based on groups,
iv. Hypothesis testing,
v. Data exploration,
vi. Hypothesis generating,
vii. Data reduction.

Many different terms are being used for cluster analysis in different scientific fields, such as unsupervised learning in pattern recognition, numerical taxonomy in biology and ecology, typology in social sciences and partition in graph theory. (Theodoridis and Koutroubas, 1999)

Unsupervised learning term is used for cluster analysis, since there are no predefined classes in clustering and before beginning to cluster data there is not any knowledge about the desirable types of relationships among data and the classes. (Berry and Linoff, 1996) On the other hand, in classification there are predefined classes and the problem turns into assigning the data into these predefined classes. (Fayyad et al., 1996)

The major components of a clustering activity are stated by Jain and Dubes (1988):

i. Pattern representation including deciding the number of clusters, the number and types of features,
ii. Definition of a pattern proximity measure (a distance measure) appropriate to the data,
iii. Clustering by applying a clustering technique,
iv. Data abstraction is forming simple and compact representation of the data set,

v. Assessment of the output is evaluating the clustering scheme's compatibility with the data structure.

2.1.1 Clustering Techniques

A taxonomy of clustering approaches stated by Jain and Dubes (1988) is shown in Figure 2.1.

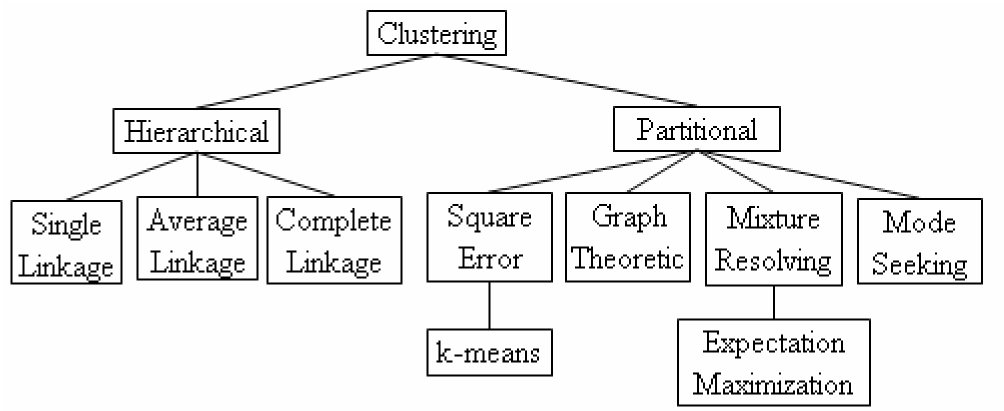

Figure 2.1 A taxonomy of clustering approaches

Apart from the taxonomy shown in Figure 2.1, each clustering technique is affected by some other issues that affect all of them regardless of their placement in the taxonomy, Jain et al. (1999).

<u>Agglomerative vs. divisive:</u>

An agglomerative clustering technique begins with each data point in a different cluster, and successively merges the data points into clusters until a stopping criterion is met. A divisive clustering technique begins with all data points in a single cluster, and performs splitting data points to clusters until a stopping criterion is met.

Monothetic vs. polythetic:

This aspect is related to the sequential or simultaneous use of features. Monothetic techniques use the features sequentially to obtain the clustering structure. In polythetic techniques all features are considered for clustering.

Hard (crisp) vs. fuzzy:

A hard clustering technique clusters each data point to a single cluster, whereas a fuzzy clustering technique assigns degrees of membership to a cluster for each data point. Hard clustering can be found with different names in literature. For example Halkidi et al., (2001) uses crisp clustering definition for hard clustering.

Hierarchical Clustering:

Hierarchical clustering techniques can be divided into two categories, agglomerative methods and divisive methods. Agglomerative techniques combine the N data points into groups and divisive techniques partition the N data points into smaller partitions, (Everitt, 1974).

The results of a hierarchical clustering technique are visualized by the use of a special type of a tree structure which is called *dendrogram*. A dendrogram consists of layers of nodes that represent a cluster. Lines connect nodes represent clusters which are nested into one another, (Jain and Dubes, 1988). A sample example of a dendrogram is shown in Figure 2.2.

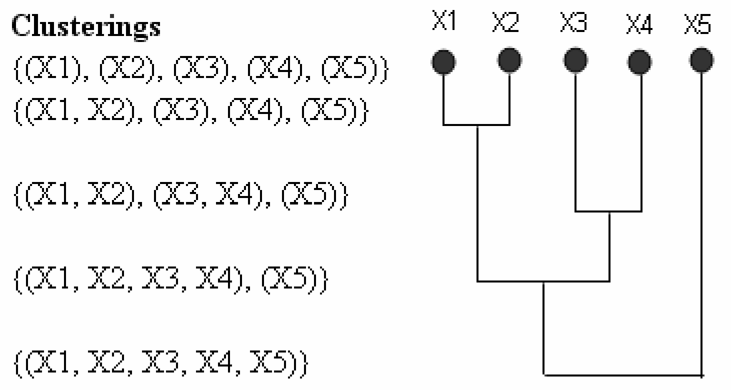

Figure 2.2 An example representation of a dendrogram

Most hierarchical clustering techniques are variants of the single linkage, average linkage, complete linkage and the minimum variance algorithms. Single linkage, average linkage and complete linkage algorithms are the most popular of these. Single linkage method considers the distance between the two clusters as the minimum of the distances between all pairs of data points from the two clusters (one data point from the first cluster and the other from the second one). Average linkage method considers the distance between two clusters as the average distance between all pairs of data points from the two clusters. Complete linkage method considers the distance between two clusters as the maximum of the distances between all pairs of data points from the two clusters, (Jain et al., 1999).

k-means Clustering:

k-means clustering algorithm is a partitioning method. It partitions the data points into k mutually exclusive clusters. k-means clustering algorithm produces clusters by optimizing a criterion function. The algorithm starts with a random initial partition and keeps reassigning the data points to clusters based on the similarity between the data point and the cluster centroids until a stopping criterion is met.

The most commonly used criterion function is the squared error criterion. The squared error for clustering L objects having H features to K clusters is;

$$e^2(H,L)=\sum_{j=1}^{K}\sum_{i=1}^{N_j}\left\|x_i^j-c_j\right\|^2$$

where x_i^j is the i^{th} data point belonging to the j^{th} cluster, N_j is the number of data points in the j^{th} cluster and c_j is the centroid of the j^{th} cluster.

Main steps of the k-means clustering algorithm are as follows:

i. Choose k cluster centres randomly,
ii. Assign each data point to the closest cluster center,
iii. Recompute the cluster centers using the current clustering structure,
iv. If a stopping criterion is not met go to step ii, otherwise stop, (Jain et al., 1999).

One of the major drawbacks of k-means algorithm is its sensitivity to the selection of the initial cluster centroids. The results of clustering structure obtained by k-means algorithm depend on these initially chosen points and due to this reason the clusters could be struck in outliers, (Lodha, et. al. , 2005). In order to overcome this drawback, the statistical analysis and graphics software STATISTICA proposes three methods for selecting the initial cluster centroids.

They are;

i. Sorting distances between data points and taking the cluster centroids at constant intervals from these sorted data points,
ii. Choosing the cluster centroids so as to maximize the initial inter-cluster distances,
iii. Choosing the data points for forming the initial cluster centroids, by this method the decision maker has full control over the choice of initial configuration.

2.1.2 Fisher's Classification Method

In Fisher's classification method, in order to determine the cluster of a new data point the case of assigning the data point to all clusters is evaluated by calculating the cost of misclassification. The new data point is classified to a cluster that has the minimum value of the expected cost of misclassification.

Let $f_i(x)$ be the density associated with the cluster π_i, i = 1, 2, …, g for the case of g clusters.

Let p_i be the prior probability of cluster π_i, i = 1, 2, …, g and $c(k \mid i)$ be the cost of allocating a data point to cluster π_k when, in fact, it belongs to cluster π_i, for k, i = 1, 2, …, g.

For $k = i$, $c(i \mid i) = 0$.

Let R_k be the set of x's classified to cluster π_k and

$P(k \mid i) = P($classifying the data point to cluster $\pi_k \mid \pi_i) = \int_{R_k} f_i(x)dx$

for k, i = 1, 2, …, g with $P(i \mid i) = 1 - \sum_{\substack{k=1 \\ k \neq i}}^{g} P(k \mid i)$.

The conditional expected cost of misclassifying the data point x from cluster π_1 to cluster π_2, or to cluster π_3, …, or to cluster π_g is;

$$\text{ECM}(1) = \sum_{k=2}^{g} P(k \mid 1)\ c(k \mid 1).$$

The data point x will be assigned to a cluster π_k that has the minimum expected cost of misclassification, (Johnson and Wichern, 2002).

2.2 QUALITY ASSESSMENT IN CLUSTER ANALYSIS

One of the main objectives of clustering is to group similar objects into the same group and separate different objects into different groups. Therefore to decide the quality of a clustering scheme, the results should be evaluated with respect to these criteria. In literature, two criteria are proposed for evaluating the quality of a clustering scheme. (Halkidi et al., 2000)

They are:

1) *Compactness*, the objects in the same cluster should be as close to each other as possible. This criteria evaluates, "how similar the objects in the same cluster are",
2) *Separation*, the clusters should be away from each other. This criterion evaluates, "how well separated the clusters from each other are". There are three ways to measure the distance between two clusters (Berry and Linoff, 1996):
    i. *Single linkage,* measures the distance between the closest members of two clusters,
    ii. *Complete linkage*, measures the distance between the farthest members of two clusters,
    iii. *Comparison of centroids*, measures the distance between the centroids of two clusters.

Clustering algorithms can be divided into two categories according to how the clustering algorithm deals with the uncertainty in terms of cluster overlapping. They are crisp clustering and fuzzy clustering, (Halkidi et al., 2001).

The cluster quality indexes are different for fuzzy and crisp clustering cases. For the case of clustering R&D projects, it is suitable to analyze the cluster quality indexes for crisp clustering, since in our case it is desired to have each project belonging to only one project cluster.

2.2.1 Crisp Clustering Quality Measures

Three indexes those are explained in Halkidi et al. (2001) are explained in this part.

Dunn Index:

This index tries to evaluate the compact and well separated clusters. It considers the distances between clusters and the within cluster distances.

The Dunn Index formula is as follows:

- $$D_{nc} = \min_{i=1,\dots nc}\left\{ \min_{j=i+1,\dots nc}\left\{ \frac{d(c_i,c_j)}{\max\limits_{k=1,\dots nc} diam(c_k)} \right\}\right\}$$

  where

  - nc is the total number of clusters.
  - $d(c_i,c_j) = \min\limits_{x\in c_i, y\in c_j} d(x,y)$ is the dissimilarity function between two clusters c_i and c_j. This function finds the minimum distance between the elements of two different clusters. Squared Euclidean distance can be used for defining this distance.
  - $diam(c_k) = \max\limits_{x,y\in c_k} d(x,y)$ is the diameter of cluster c_k. This function finds the distance between the farthest elements of a cluster.

If the clustering scheme has compact and well separated clusters, then it is expected that the intra-cluster distances are small (smaller values of $diam(c_k)$) and the inter-cluster distances are large (larger values of $d(c_i, c_j)$). Therefore the large values of Dunn index indicate that the clustering scheme has compact and well separated clusters.

The Davies-Bouldin (DB) Index:

Davies-Bouldin index measures the similarity between clusters.

- $\mathrm{R_{ij}} = \dfrac{(\mathrm{s_i} + s_j)}{d_{ij}}$, where
  - $\mathrm{s_i}$ and $\mathrm{s_j}$ indicates the dispersion of clusters i and j respectively.
  - d_{ij} is a dissimilarity measure between clusters i and j.

The $\mathrm{R_{ij}}$ index is calculated for each cluster by comparing it with other clusters pairwisely. Milligan et al., (1985) used average within cluster distance as the measure of the dispersion of a cluster ($\mathrm{s_i}$ and $\mathrm{s_j}$ values), and the distance between cluster centroids is used as a measure of the dissimilarity between two clusters.

Then the maximum $\mathrm{R_{ij}}$ values obtained for each cluster are averaged to obtain the DB index.

The DB Index formula for nc clusters is as follows:

- $\mathrm{DB_{nc}} = \dfrac{1}{nc}\sum_{i=1}^{nc} R_i$, where
  - $R_i = \max_{i=1,\ldots,nc,\, i\neq j} R_{ij}$, $i = 1,\ldots,nc$

SD Measure:

The SD measure has two parts; its first part measures the scattering of the clusters and the second part measures the separation of clusters.

Some definitions for explaining the SD measure are as follows:

- $\sigma(X)$, *is the variance vector of data set* and its p^{th} dimension is (Halkidi et al., 2000):

  $$\sigma_x^p = \frac{1}{n}\sum_{k=1}^{n}\left(x_k^p - \bar{x}^p\right)^2 \text{, where}$$

  - x_k^p is the p^{th} dimension of the k^{th} object,
  - n is the total number of objects and
  - $\bar{x}^p = \frac{1}{n}\sum_{k=1}^{n} x_k^p$

- σ_{Vi} *is the variance vector of cluster* i and its p^{th} dimension is (Halkidi et al., 2000):

  $$\sigma_{Vi}^p = \frac{\sum_{k=1}^{n_i}\left(x_k^p - v_i^p\right)^2}{n_i} \text{, where}$$

  - x_k^p is the p^{th} dimension of the k^{th} object,
  - v_i^p is the p^{th} dimension of the centroid of cluster i,
  - n_i is the number of objects in cluster i.

- $Scat(c)$ *is the average scattering for clusters (intra-cluster distance)* and is computed as:

  $$Scat(c) = \frac{\frac{1}{c}\sum_{i=1}^{c}\left\|\sigma(Vi)\right\|}{\left\|\sigma(X)\right\|} \text{, where}$$

- c is the number of clusters,
- $\sigma(Vi)$ is the variance vector of the centroid of cluster i .

- *$Disc(c)$ is the total separation between clusters (inter-cluster distance)* and its formula is as follows:

$$Disc(c) = \frac{D_{\max}}{D_{\min}} \sum_{k=1}^{c} \left(\sum_{z=1}^{c} \left\| V_k - V_z \right\| \right)^{-1}$$, where

- c is the number of clusters.
- V_k and V_z are the vectors of centroids of clusters k and z respectively.
- $D_{\max} = \max \left\| V_i - V_j \right\| \ \forall\ i\ ,\ j \in \{1, 2, 3, ..., c\}$, the maximum distance between cluster centroids.
- $D_{\min} = \min \left\| V_i - V_j \right\| \ \forall\ i\ ,\ j \in \{1, 2, 3, ..., c\}$, the minimum distance between cluster centroids.

The SD measure is as follows:

- $SD(c) = a * Scat(c) + Disc(c)$, where
  - a is a weighting factor equals to the maximum number of clusters that is obtained from decision makers.

Silhouette Value:

Silhouette value measures the degree of similarity of an object to the objects in its cluster compared to objects in other clusters. For an object i , which is in cluster A , the silhouette value is calculated as follows:

$$
\begin{aligned}
s(i) &= 1 - \frac{a(i)}{b(i)} && \text{if } a(i) < b(i) \\
&= 0 && \text{if } a(i) = b(i) \\
&= \frac{b(i)}{a(i)} - 1 && \text{if } a(i) > b(i)
\end{aligned}
$$

The formula for $s(i)$ can be expressed as follows:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

- $a(i)$: average dissimilarity of object $i \in A$ to all other objects of cluster A.
- $d(i,C)$: average dissimilarity of object $i \in A$ to all objects in cluster C, $C \neq A$.
- $b(i)$: $\min_{C \neq A} d(i,C)$ the cluster B which this minimum is obtained, that is $d(i,B) = b(i)$, is called the neighbor of object i. Cluster B can be interpreted as the second-best cluster choice for object i other than the cluster A.

From the above definitions:

$$-1 \leq s(i) \leq 1 \text{ for each object } i.$$

The proposed interpretation, from Kaufmann and Rousseeuw (1989), of the average silhouette value for the entire data set is given in Table 2.1.

Table 2.1 Subjective interpretation of the average silhouette value for the entire data set

Average Silhouette Value	Proposed Interpretation
0.71 - 1.00	A strong structure has been found
0.51 - 0.70	A reasonable structure has been found
0.26 - 0.50	The structure is weak and could be artificial; please try additional methods on this data set
$\leq$ 0.25	No substantial structure has been found

2.3 PROJECT CLASSIFICATION

Shenhar et al. (1998) stated that project success factors are not universal for all projects. The difficulties observed during the identification of success factors of projects occur due to not tailoring the project management rules according to project specific characteristics. In studies of project success factors, commonly it is assumed that all projects are similar, whereas different types of projects should be evaluated by different sets of success factors since the success factors of projects mainly depend upon the project properties.

Shenhar (2001) stated that project management should be tailored to project specific characteristics. The common assumption in project management literature stating that all projects are fundamentally similar and "one size fits all" is not correct. Different types of projects should be managed in different ways. Although most organizations are using implicitly different strategies for different projects, there are no clear definitions of project groups, before beginning to a project, no decision is given about which project group this project belongs to and the management rules of a new project are not determined by taking into consideration the structure of project groups. In order to be more successful, the organizations should apply a formal project classification. The selection of project leaders, project team members, and many decisions regarding project management should be made according to the specific properties of project groups. Shenhar (2001) proposed a framework, within the projects are classified into four levels of technological uncertainty, and into three levels of system complexity.

The concept of managing different projects differently can also be seen in the staged version of Capability Maturity Model Integration (CMMI^{SM}) Version 1.1[1].

The staged version of CMMI includes the following statements;

i. Establish organizational process assets
ii. Establish life-cycle model descriptions
iii. Establish tailoring criteria and guidelines.

The second and the third statements involve construction of the project life-cycle models, criteria for selecting life-cycle models for a project, and documenting the tailoring guidelines for the organization's set of standard processes.

Construction of project life-cycles can be interpreted as project classification, criteria for selecting life-cycle models for a project can be interpreted as determining the project class of a new project. The tailoring of the organization's standard processes can be interpreted as managing the project according to its project class properties.

[1] Capability Maturity Models (CMMs) are used as a guide to use while developing a process. They contain basic elements needed for effective processes. These elements are derived from the concepts developed by Crosby, Deming, Juran and Humphrey.

2.4 ANALYTIC HIERARCHY PROCESS

The Analytic Hierarchy Process (AHP) is a multi-criteria decision making method, which is introduced by Saaty in the 1970's. The AHP is a decision support tool that can be used to solve complex decision problems, since AHP can deal with problems having criteria expressed in different units or when the related data is difficult to be quantified. By using AHP the objectives of the decision, the criteria for evaluating the alternatives and the alternatives are represented hierarchically. The evaluation data are derived by performing a set of pairwise comparisons, (Triantaphyllou and Mann, 1995).

The standard form of a decision schema, (Zahedi, 1986) is shown in Figure 2.3.

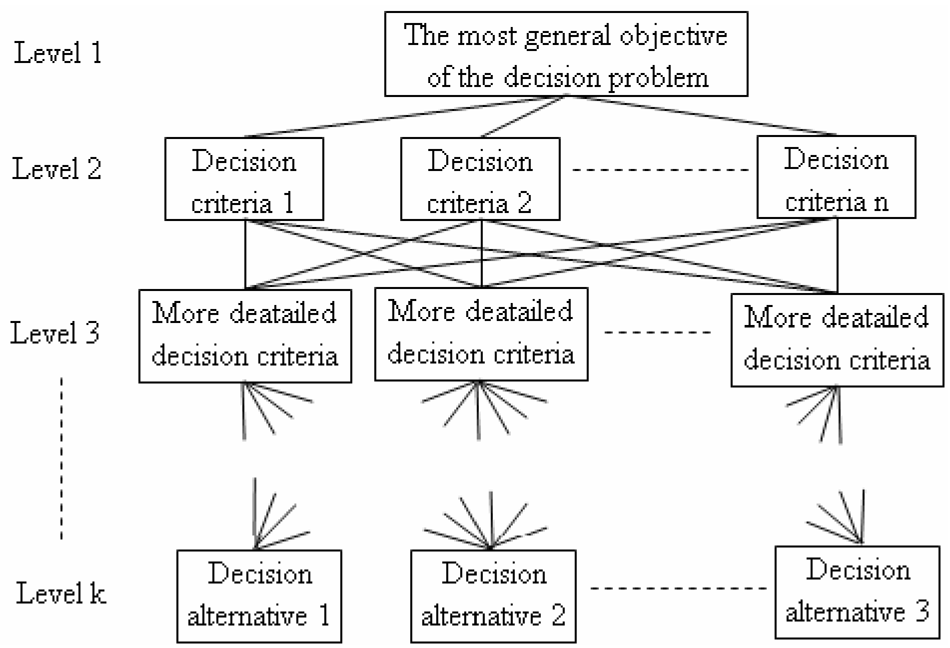

Figure 2.3 The standard form of decision schema in AHP with k level hierarchy

2.4.1 The Axioms of the AHP

The axioms of the AHP, stated by Saaty, (1986) are as follows:

1. The reciprocal property: When making paired comparisons, both members of the pair to judge the relative value, should be considered. If one of the compared members is judged to be x times more preferable than another, then the other member is automatically $1/x$ times more preferable than the first one.
2. Homogeneity: Homogeneity is essential for comparing similar things, since large errors can happen when comparing widely disparate elements. When the disparity is large, the elements are placed in separate groups of comparable size giving rise to the idea of levels and their decomposition.
3. Independence: While making preferences, the evaluation of the criteria are assumed to be independent of the properties of the alternatives.
4. Expectations: The constructed hierarchic model should represent the ideas of the decision makers. All alternatives and all criteria should be represented in the hierarchy.

2.4.2 The Steps of the AHP

The Analytic Hierarchy Process (AHP) has eight steps which are explained by Saaty, (1982).

These steps are;

1. The definition of the problem and establishing what should be known.
2. Structuring the hierarchy of the problem by defining the main goal, the sub goals and the list of alternatives.
3. Performing pairwise comparisons of alternatives for each lower level sub goals.
4. Constructing the preference matrices by assigning the reciprocals to each pairwise comparison.

5. Evaluation of the consistency, after completing the pairwise comparisons.
6. Repeating steps 3, 4, and 5 for all sub goals at all levels of the hierarchy.
7. Finding the weights of the criteria and the alternatives by using the hierarchic structure.
8. Calculating the consistency of the entire hierarchy.

2.4.3 Normalization in the AHP

The normalization in conventional AHP is performed in one-dimension. A is an n x n preference matrix constructed through pairwise comparisons of n objects,

$$A = \begin{pmatrix} a_{11} & a_{12}. & . & .a_{1n} \\ a_{21} & a_{22}. & . & .a_{2n} \\ . & & & \\ . & & & \\ . & & & \\ a_{n1} & a_{n2}. & . & .a_{nn} \end{pmatrix} \tag{2.1}$$

where a_{ij}'s are the judgments or the relative importance of alternative i to alternative j.

By the first axiom of AHP - the reciprocal property,

$$a_{ij} = \frac{1}{a_{ji}} \tag{2.2}$$

If R is the right principle eigenvector of A then

$$A\ R = \lambda_{\max}\ R \tag{2.3}$$

where $\lambda_{\max}$ is the maximum eigenvalue of A.

Let R_i be an entry of R, then the normalized relative properties w_i of the objects can be found by

$$w_i = \frac{R_i}{\sum R_i} \tag{2.4}$$

where $\sum w_i = 1$ and $\lambda_{\max}$ is the maximum eigenvalue of the eigenvector W, (Zahir, 1999).

In the case of having fully consistent values for a_{ij} then

$$a_{ij} = a_{ik} . a_{kj} \qquad (i,\ j,\ k = 1,2,3,\ldots,n) \qquad (2.5),$$

If A is consistent then $a_{ij} = \frac{w_i}{w_j}$, where w_i and w_j denote the actual values of i and j respectively.

Let $W = (w_1, w_2, ..., w_n)$ then using $a_{ij} = \frac{w_i}{w_j}$,

$$\sum_{j=1}^{n} a_{ij} w_j = \sum_{j=1}^{n} w_i = n w_i \qquad \text{where } i = 1, 2, \ldots, n \tag{2.6}$$

stating differently

$$A\,W = n\,W \qquad \text{where } W = (w_1, w_2, ..., w_n) \tag{2.7}$$

Equation 2.7 states that n is an eigenvalue of A with W as the corresponding eigenvector.

2.4.4 Relative and Absolute Measurement Methods

In AHP each considered alternative is assessed based on predefined criteria. The numeric evaluation results of alternatives are called the weights of alternatives.

AHP provides two ways for calculating the weights of alternatives:

1. Relative measurement
2. Absolute measurement

Relative measurement is generally used in new learning situations, whereas the absolute measurement method is used on standardized problems. (Saaty, 1986)

In relative measurement method, alternatives are evaluated pairwisely according to a criterion. In absolute measurement a pre-established scale developed through experience, is used to evaluate the alternatives.

The characteristic properties of absolute measurement method are as follows:

1. Addition or deletion of an alternative does not change the rankings of alternatives, (no rank reversal) (Saaty, 1986).
2. The weights of the alternatives do not depend on the number and the priorities of the alternatives.
3. The number of judgments necessary to determine the weight of an alternative is less than the one in relative measurement. In relative measurement method AHP, $n(n\text{-}1)/2$ pairwise comparisons (where n is the number of alternatives) should be performed for each sub goal of the hierarchy.
4. It has the capability to rate large number of alternatives. (Forman, 1996)
5. Absolute measurement method looks more like a traditional evaluation methodology; therefore the decision maker may find it more familiar. (Forman, 1996)

The absolute measurement method is used for college admission decisions, personnel evaluation, resource allocation; it is also applicable when there are rules or regulations that are prohibiting the comparison of an alternative against another. (Forman, 1996)

2.5 ANALYTIC HIERARCHY PROCESS AND CLUSTERING

Ben-Arieh and Triantaphyllou (1992) applied AHP clustering in group technology for clustering parts which have similar features. They established the quantitative feature values of parts by using AHP and by using the obtained results; they applied two different clustering approaches.

The first method is matrix-based clustering. The weighted feature values of each part are used for forming a matrix, and the clustering is performed by using this part-feature matrix. In matrix-based clustering all the feature values of parts are used as the input parameters.

The second method is aggregate-value clustering. In aggregate-value clustering, each part is represented by a unique value that is obtained by aggregating the feature values of a part. This value is calculated by using the following formulation for each part i,

$$y_i = \sum_{j=1}^{M} w_j a_{ij}$$

In the above formula, the M is the total number of features, w_j is the weight of feature j, and a_{ij} is the weight of j th feature of part i.

2.6 VECTOR SPACE FORMULATION OF THE AHP (VAHP) AND CLUSTERING

Zahir, (1999) has proposed a Vector Space Formulation of the AHP (VAHP) using Euclidean normalization. The VAHP has the same decision hierarchy as the conventional AHP and uses the same eigenvector method. All the axioms of AHP are also valid for VAHP.

Clustering algorithms generally uses Euclidean distances. Zahir (1999) stated that using the numerical data that is obtained from AHP method in clustering methods can cause problems, since AHP method uses summation normalization. Zahir (1999) states that VAHP based approach is more meaningful in using with clustering techniques that uses Euclidean distances, since the data points obtained from VAHP satisfies Euclidean normalization.

VAHP uses the same preference matrix-A of conventional AHP, whereas it converts it to another form. In VAHP there is a preference operator called P which is obtained by taking the square roots of each element of A, where A is the conventional preference matrix of AHP.

If A is consistent, P is also consistent and $\lambda_{\max} = n$ for both. The eigenvector W of A is normalized to satisfy $\sum w_i = 1$, whereas the eigenvector v of P is normalized to satisfy $\sum_{i=1}^{n} v_i^2 = 1$. Therefore we have $v_i = \sqrt{w_i}$ or $w_i = v_i^2$. In this structure w_i is interpreted as the relative priority for A, v_i^2 is interpreted as the relative priority for P.

We further normalize v_i's using ideal mode AHP normalization as:

$$v_i^{\ N} = \frac{v_i}{\max v_i}$$

Ideal mode AHP normalization is first proposed by Belton and Gear (1983) and then accepted by Saaty. Belton and Gear (1983) observed that the AHP might reverse the ranking of alternatives when an alternative identical to one of the already existing alternatives is introduced. In order to overcome this situation, Belton and Gear (1983) proposed that each column of the AHP preference vector to be divided by the maximum entry of that column.

2.7 THE METHODOLOGY APPLIED IN TUNÇ'S WORK

Tunç (2004) has clustered R&D projects of the electronics company. For clustering the projects, a five dimensional project feature vector is established and used. The feature vector consists of one objective and four subjective dimensions. Only the amount of resource (labor) dimension is objective. The other four dimensions are subjective.

These dimensions are;

1. Technological Uncertainty,
2. Platform Type,
3. Work & Test Environment,
4. System Scope,
5. Amount of Resource (Labor).

Technological Uncertainty: This dimension requires the evaluation of two factors. The first one is the technology level that a project requires. The second one is the maturity level of that technology in the company at the time of project initiation. Therefore the technological uncertainty of a project can not be determined without considering the company's past experiences and the knowledge levels in the related technological areas.

Platform Type: This dimension is used for defining the working environment of the products. Platform type of a project determines the quality standards that the product should obey and affects the difficulty. Generally the platform types are divided into two main groups, the commercial platforms and the military platforms. Stationary, mobile, on-the move platforms are some examples.

Work & Test Environment: According to the project specific characteristics, the work & test environment of projects may differ. Some of them only require the facilities of this electronics company, whereas some require military area or a development facility that does not belong to this electronics company. Because of this reasons this dimension affects mostly the efficiency of the project.

System Scope: Complex military products are composed of sub systems, components, and parts that function together. The hierarchical structure of the system of products affects the design and the managerial aspects of each.

Amount of Resource (Labor): Human resource is an important input of an R&D project, and it gives a general understanding about the size of a project. The total R&D labor hour spent for completed projects, and the total estimated R&D labor hour for new projects are used for evaluating this dimension.

First the dimensions of the constructed project feature vector are evaluated by pairwise comparisons in terms of their impact on clustering the projects, then the projects are evaluated by pairwise comparisons in terms of these dimensions. The evaluations were performed by the four managers from the different R&D departments of this electronics company. The hierarchy of the constructed AHP model is in Figure 2.4.

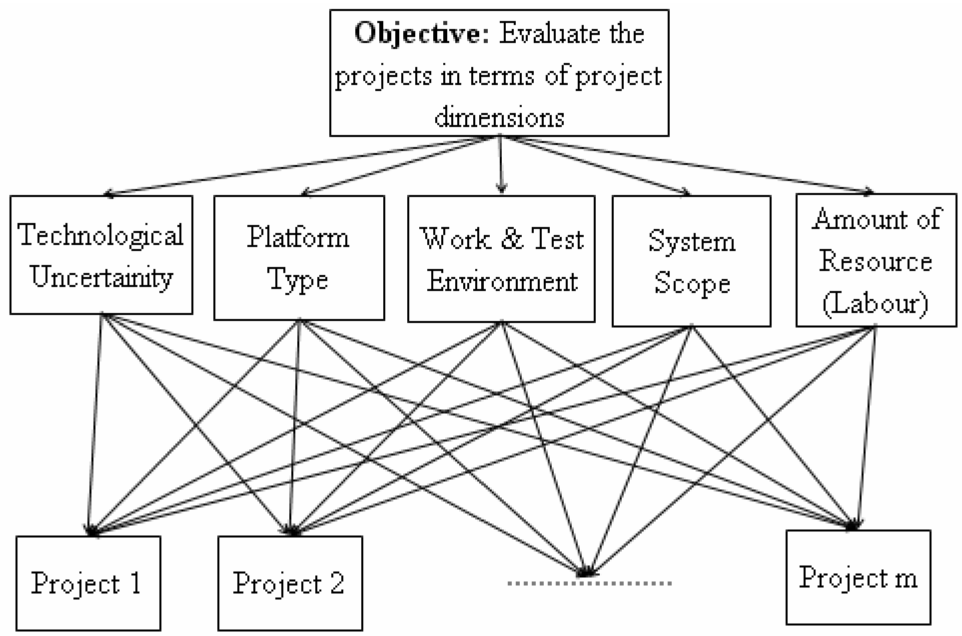

Figure 2.4 The hierarchy of the AHP model constructed by Tunç (2004)

The steps followed for clustering the projects are listed as follows:

1. Defining the projects as data points in a multi dimensional space.

   i. *Constructing the dimensions of the projects:* The dimensions are constructed in order to define the projects in a project space. Five dimensions are constructed by analyzing the literature about high-tech projects and obtaining the opinions of project technical managers and R&D department managers of this electronics company.

   ii. *The pairwise comparison of dimensions:* The decision makers were asked to evaluate the dimensions according to their importance on the clustering of the projects. Each decision maker made pairwise comparisons of the dimensions, and the group average is taken as the dimension weight.

   iii. *The pairwise comparison of projects according to the dimensions:* For each dimension (dimensions other than amount of resource dimension, numeric values are used for the amount of resource dimension) all projects are compared by decision makers according to their complexity levels.

2. Applying the clustering algorithms and evaluating the results.

   i. *Clustering the projects according to their feature vectors:* The results of the projects feature vectors are used for clustering. Two different approaches are used while using the project feature vectors. The first one is the matrix based clustering that use the project feature matrix as an input and the second one is the aggregate-value clustering that use the aggregate project values obtained by multiplying the project feature vector with the related project features' weights. Different clustering methods are applied by using the matrix based and aggregate value clustering methods.

ii. *Evaluation of the clustering results:* Several clusters are obtained and each of them is evaluated by using cluster quality indexes. The methods that use the matrix constructed by project feature vectors (not the aggregate values for each project) is selected. Since this result was evaluated by decision makers as the one representing the reality best and it includes five groups of projects.

Tunç (2004) used several different methods for clustering the projects. These methods can be grouped as AHP based clustering and VAHP based clustering.

Both the k-means and the hierarchical clustering algorithms are used for the methods listed below. The methods that applied by Tunç (2004) are:

1. AHP based clustering
   i. Matrix based clustering
   ii. Aggregate-value clustering

2. VAHP based clustering

The results of the following methods are the same and the clustering scheme of these methods are evaluated as the one representing the true clustering structure best by decision makers:

i. AHP - Matrix based clustering by using k-means clustering algorithm
ii. VAHP based clustering by using Hierarchical clustering algorithm
iii. VAHP based clustering by using k-means clustering algorithm

The main outputs of Tunç (2004) are;

i. AHP weights of the five features, and projects' weights for subjective and objective features,
ii. VAHP weights of the five features, and projects' weights for subjective and objective features,

iii. the project clusters.

These outputs are represented in the Table 2.2, Table 2.3 and Table 2.4.

Table 2.2 The AHP weights of the features and projects' weights for features obtained by Tunç (2004)

	Technological Uncertainty	Platform Type	Work and Test Environment	System Scope	Amount of Resource (Labor)
Features' Weight	0,41	0,12	0,2	0,14	0,13
Project 1	0,0555	0,0777	0,2287	0,1010	0,0200
Project 2	0,0830	1,0000	0,4400	0,3473	0,1400
Project 3	0,0709	0,9356	0,4045	0,1098	0,0800
Project 4	0,1978	0,1922	0,1635	0,2611	0,0700
Project 5	0,1741	0,1514	0,0685	0,1235	0,0500
Project 6	0,0946	0,1544	0,1574	0,0954	0,0100
Project 7	0,8074	0,9730	1,0000	0,6161	1,0000
Project 8	0,3646	0,2906	0,6380	1,0000	0,4200
Project 9	0,1109	0,2665	0,1999	0,1513	0,0900
Project 10	0,2287	0,0737	0,7037	0,1912	0,3100
Project 11	0,4170	0,2856	0,2519	0,2862	0,4700
Project 12	0,0649	0,1598	0,0647	0,0947	0,1600
Project 13	0,6298	0,1514	0,1167	0,1513	0,1400
Project 14	1,0000	0,1996	0,2878	0,1697	0,3400

Table 2.3 The VAHP weights of the features and projects' weights for features obtained by Tunç (2004)

	Technological Uncertainty	Platform Type	Work and Test Environment	System Scope	Amount of Resource (Labor)
Features' Weight	0,6411	0,3451	0,4515	0,3733	0,3559
Project 1	0,2357	0,2783	0,4808	0,3192	0,1300
Project 2	0,2915	1,0000	0,6664	0,5870	0,3800
Project 3	0,2692	0,9678	0,6387	0,3324	0,2900
Project 4	0,4496	0,4378	0,4043	0,5094	0,2600
Project 5	0,4210	0,3897	0,2619	0,3517	0,2200
Project 6	0,3113	0,3934	0,3982	0,3101	0,0900
Project 7	0,9065	0,9871	1,0000	0,7807	1,0000
Project 8	0,6095	0,5394	0,8022	1,0000	0,6500
Project 9	0,3364	0,5165	0,4496	0,3895	0,3000
Project 10	0,4826	0,2797	0,8419	0,4383	0,5600
Project 11	0,6532	0,5343	0,5041	0,5334	0,6800
Project 12	0,2560	0,3996	0,2551	0,3090	0,4000
Project 13	0,7995	0,3899	0,3432	0,3903	0,3800
Project 14	1,0000	0,4453	0,5382	0,4128	0,5800

Table 2.4 The project clusters that are obtained by Tunç (2004)

Cluster Number	Project
1	Project 1
2	Project 2
2	Project 3
1	Project 4
1	Project 5
1	Project 6
3	Project 7
4	Project 8
1	Project 9
4	Project 10
4	Project 11
1	Project 12
5	Project 13
5	Project 14

CHAPTER 3

PROPOSED METHODOLOGY

This chapter includes the proposed methodology for representing the projects in a multi dimensional space, the Project Identification Card, the absolute measurement method of AHP applied for constructing the numerical scale of Project Identification Card, mathematical model of the clustering projects problem, the Tabu Search clustering algorithm, sensitivity analysis of clusters to each project and the proposed classification method of a new project.

In this study a Project Identification Card (PIC) is developed for defining the projects. By using the developed PIC, the projects are clustered. The developed PIC is recommended for use to define a new project, and a methodology for determining the cluster of a new project is developed. In this electronics company generally a project comes ones a year. Because of this reason and due to the nature of our problem it is more suitable to consider the case of a new incoming project rather than the case of coming projects in lots, since projects come one by one.

The methodology of this thesis study is given in Figure 3.1.

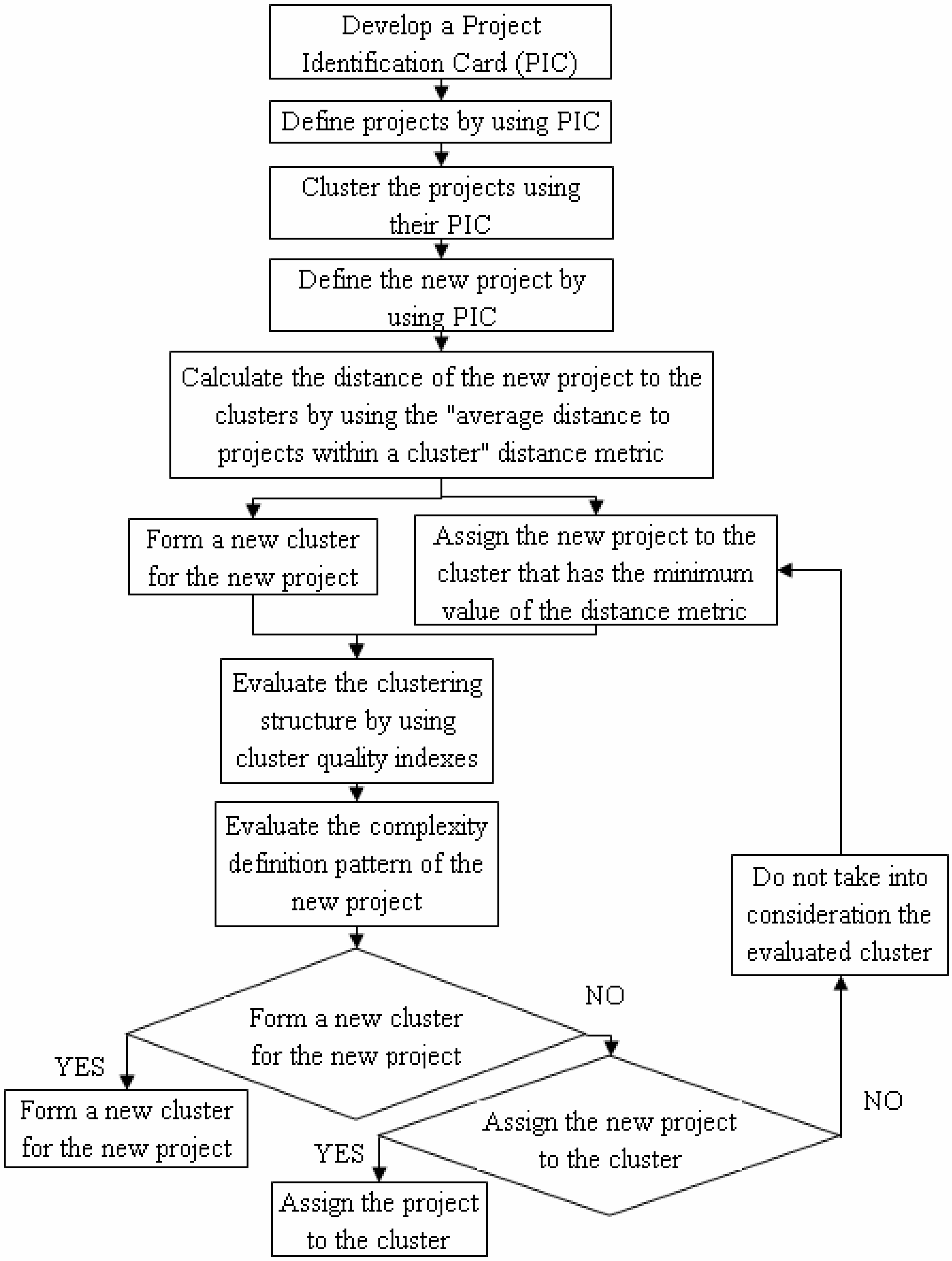

Figure 3.1 The flow chart of the developed methodology

3.1 THE PROPOSED APPROACH FOR REPRESENTING THE PROJECTS IN A MULTI DIMENSIONAL SPACE

The main input of any clustering or classification method is some representation of objects that are either to be clustered or to be classified, in some pre-defined space. In order to identify the projects in a project space, a template, which is called Project Identification Card (PIC) is constructed. The constructed PIC is used for describing the projects according to their characteristic properties and representing them in a multi dimensional project space.

Tunç (2004) obtained the project dimension values by comparing all the projects pairwisely for each project dimension. Whereas in our case, the project dimension values are obtained by using the constructed PIC. The project dimensions developed by Tunç (2004) forms the main dimensions of the PIC, and complexity definitions for each dimension are developed. In this method only the complexity definitions of a project dimension are compared pairwisely for each project dimension.

Tunç (2004) obtained the values of project dimensions from the results of pairwise comparisons. Whereas in our case the pairwise comparisons are used for developing the numerical scale of PIC. The project dimensions are determined by defining a project for each project dimension by choosing one of the developed complexity definitions. Then the project's dimension value is obtained by using the numerical value related to that complexity definition. Therefore the decision maker can define a project with less effort, since the number of pairwise comparisons is less than the number of Tunç (2004). The decision maker does not have to compare the project with all the other projects for each project dimension, which requires a lot of time and a lot of concentration.

3.1.1 Project Identification Card

In this study, following Tunç (2004), the projects are represented in a five dimensional project space, using the same dimensions of his study. Project Identification Card (PIC), is a template that is used to define the characteristic properties of projects, constructed by detailing these five dimensions.

The main dimensions of PIC are;

i. Technological Uncertainty,
ii. Platform Type,
iii. Work & Test Environment,
iv. System Scope,
v. Amount of Resource (Labor).

One of the main purposes of PIC is to provide the decision maker a measurement tool for evaluating the complexity of a project based on the main project dimensions. In order to evaluate a project's complexity based on a dimension, some information about the definitions of complexity level of a dimension should be given to the decision maker for consideration. Therefore complexity definitions for each dimension are constructed. By choosing a complexity definition for each dimension the decision maker can represent each project categorically.

In PIC, every dimension is broken down into its complexity definitions. The complexity definitions of each dimension are arranged from the least complex one to the most complex one. In PIC, a project can be represented by only one complexity definition for each dimension. The general structure of PIC is given in Table 3.1. In Table 3.1, "Complexity Definition 1" stands for the least complex case of a dimension and the "Complexity Definition n" stands for the most complex case of a dimension, when there are n complexity definitions.

Table 3.1 The general structure of the Project Identification Card

PROJECT IDENTIFICATION CARD	
Technological Uncertainty Dimension	Complexity Definition n
	⋮
	⋮
	Complexity Definition 1
Platform Type Dimension	Complexity Definition n
	⋮
	⋮
	Complexity Definition 1
Work & Test Environment Dimension	Complexity Definition n
	⋮
	⋮
	Complexity Definition 1
System Scope Dimension	Complexity Definition n
	⋮
	⋮
	Complexity Definition 1
Amount of Resource (Labour) Dimension	Complexity Definition n
	⋮
	⋮
	Complexity Definition 1

3.1.2 The Construction of the Measurement Scale of Complexity Levels of Dimensions of PIC

Absolute measurement method of Analytical Hierarchy Process is used for constructing the numerical values of each complexity definition of PIC. Firstly the complexity definitions of each dimension of PIC are constructed. Then these complexity definitions are compared according to their complexity levels pairwisely. While comparing the complexity definitions of each dimension "how much more complex a project having complexity definition d_1 for dimension d, than a project having complexity definition d_n?" type questions are posed.

The hierarchy of the AHP model used for constructing the numerical scale of PIC is given in Figure 3.2.

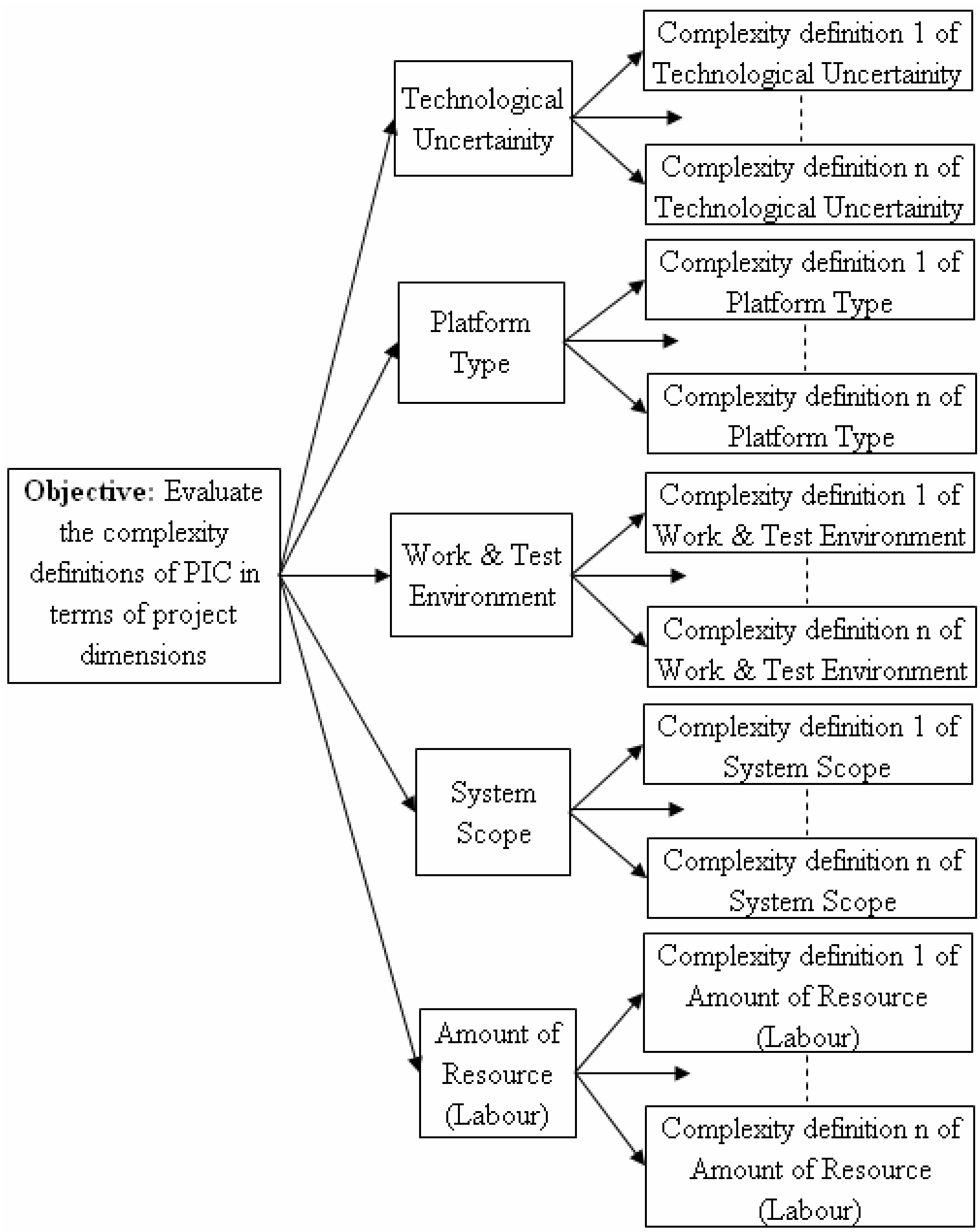

Figure 3.2 The hierarchy of the AHP model constructed for developing the numerical scale of PIC

After completing the pairwise comparisons, the numeric values of each complexity definition is constructed by applying VAHP method, eigenvalue method and ideal mode AHP normalization. By performing these methods a numerical value is constructed for each complexity definition of the dimensions of PIC.

3.2 CONSTRUCTION OF THE COMPLEXITY DEFINITIONS OF PIC

In order to determine the numerical values for the complexity definitions of each dimension of the PIC, absolute measurement method of Analytic Hierarchy Process is used. The results of this method are analyzed by three different Project Identification Card (PIC) structures. In these three different PIC's; three, four and five different complexity definitions are constructed for the main dimensions of PIC, respectively.

Absolute measurement method of AHP requires two inputs from the decision maker:

1. Evaluation of the complexity definitions of a dimension pairwisely,
2. Deciding the complexity definition of each dimension of a project.

For example, in the case when the PIC's technological uncertainty dimension has three complexity definitions (design type I, design type II, design type III), the decision maker first performs three pairwise comparisons for quantifying these three definitions. These comparisons can be described by answering questions like "How much more difficult is *a design type I project* when it is compared with *a design type II project?*". Then the decision maker determines the complexity definitions of each dimension of a project.

In the following sections, the numerical values of projects' main dimensions and the clustering structure obtained by using the outputs of these numerical outputs of PIC's are evaluated for the cases when the PIC's main dimensions has three, four and five complexity definitions. The pairwise comparisons of the complexity definitions of project dimensions stated in the following sections are approved by the Manager of the Engineering Planning Department of this electronics company.

3.2.1 The PIC Structure with Three Complexity Definitions

In this structure each dimension of PIC has three complexity definitions and each project is represented by these definitions in Table 3.2.

The developed complexity definitions for "Technological Uncertainty" dimension are;

i. continuous design,
ii. innovative design and
iii. breakthrough design.

Continuous design projects cover second generation products that are developed by modifying a previously developed product. Innovative design projects are new design projects. Breakthrough design projects are also new design projects but what makes them different from innovative design projects is their level of complexity. Breakthrough design projects may require knowledge that should be developed by extensive research and their uncertainty and risk levels are also higher than the innovative design projects. The established definitions are taken from Moody, et. al. (1997).

The developed complexity definitions for "Platform Type" dimension are;

i. stationary,
ii. mobile and
iii. on-the move.

The products that have stationary platform can function stationary and needs another product for moving from one place to another. Mobile products can function stationary but they have necessary infrastructure for moving. On-the move products can both function stationary and while moving. Also they have necessary infrastructure for moving.

The developed complexity definitions for “Work & Test Environment” dimension are;

i. zone 1,
ii. zone 2 and
iii. zone 3.

The work & test environment of the projects are evaluated and because of two reasons a more descriptive definitions could not be established. These reasons are; the secrecy level of these kind of information and it is very difficult to combine the work & test environment specifications for projects under some definitions. Zone 1 defines the projects having the least complex work & test environment specifications and zone 3 defines the projects having the most complex work & test environment specifications.

The developed complexity definitions for “System Scope” dimension are;

i. assembly,
ii. system and
iii. array.

Assembly products are a collection of components and modules combined into a single unit, either as a subsystem of a larger system or a stand-alone product performing for a single function. System products are a collection of subsystems and interactive elements that perform a wide range of functions. Array products are a large widespread collection of systems functioning together to achieve a common purpose, Shenhar (2001).

The developed complexity definitions for “Amount of Resource (Labor)” dimension are;

i. systems requiring less than X man-hours,
ii. systems requiring less than 4X man-hours and but more than X man-hours, and

iii. systems requiring more than 4X man-hours.

These definitions are developed by analyzing the dispersion of the considered 14 projects' total R&D labor hours.

The considered 14 projects are taken as the base projects for identifying the clustering structure of the projects. The clusters of them will not be changed and they will not be taken out the clustering structure later on. An incoming project will only be classified to the obtained project clusters.

The evaluation of the complexity definitions can be performed by the managers of the R&D departments of this electronics company, the project managers and project technical managers. Either all the decision makers can construct the pairwise comparison matrix by consensus or the group aggregate matrix can be constructed.

Aczel and Saaty (1983) showed that using the geometric mean of the individual judgments to obtain the group judgment for each pairwise comparison is the uniquely appropriate way for aggregating the judgments in the AHP. Also geometric mean preserves the reciprocal property. The aggregate preference matrix can be used as the conventional preference matrix of AHP.

Let $a^k{}_{ij}$ be the relative importance of alternative i to alternative j for k th decision maker. Then the aggregate importance of alternative i to alternative j, a_{ij} can be computed as:

$$a_{ij} = (a_{ij}^1 \ x \ a_{ij}^2 \ x \ ....x \ a_{ij}^M)^{1/M} \text{ where there are } M \text{ decision makers.}$$

Table 3.2 The structure of three level PIC and the projects' complexity definitions

PROJECT IDENTIFICATION CARD		Project 1	Project 2	Project 3	Project 4	Project 5	Project 6	Project 7	Project 8	Project 9	Project 10	Project 11	Project 12	Project 13	Project 14
Technological Uncertainty	Breakthrough design							x							x
	Innovative design	x	x	x	x	x			x	x	x	x	x	x	
	Continuous design						x								
Paltform Type	On-the Move		x	x				x	x						x
	Mobile	x			x					x					
	Stationary					x	x				x	x	x	x	
Work & Test Environment	Zone 3							x	x		x				
	Zone 2	x	x	x	x		x			x		x			x
	Zone 1					x							x	x	
System Scope	Array							x	x			x			
	System		x	x	x						x			x	x
	Assembly	x				x	x			x			x		
Amount of Resource (Labour)	More than 4X man-hours							x	x		x	x			x
	Less than 4X man-hours but more than X man-hours		x							x			x	x	
	Less than X man-hours	x		x	x	x	x								

In order to quantify these definitions the following steps are followed:

1. Complexity definitions for each project dimension are compared pairwisely.
2. The preference matrix for the complexity definitions of project dimensions is constructed (by using the reciprocal principle).
3. Square roots of each element in the preference matrix are taken (VAHP).
4. By applying the eigenvalue method, the priority vectors are obtained.
5. The ideal-mode AHP normalization is applied to the obtained priority vectors.

In order to explain this method, the quantification of the complexity definitions of "Technological Uncertainty" dimension for the case it has three complexity definitions (continuous, innovative, breakthrough) is explained according to the stated steps in the following part.

1. *Complexity definitions for each project dimension are compared pairwisely.*

Saaty's 1-9 scale is used for the pairwise comparisons as given in Table 3.3.

The complexity definitions of "Technological Uncertainty" dimension;

i. Breakthrough design,
ii. Innovative design, and
iii. Continuous design complexity definitions are compared pairwisely.

For comparing breakthrough design vs. innovative design the answer of the following question is evaluated. "How much more difficult is *a breakthrough design project* when it is compared with *an innovative design project?*". The answer of this comparison is "*a breakthrough design project is essentially or strongly important than an* innovative *design project".* According to this comparison and using Saaty's 1-9 scale "5" is given for breakthrough design complexity definition when it is compared with innovative design.

The other complexity definitions are compared pairwisely accordingly and the following results are obtained.

Breakthrough design vs. continuous design (9 is given for breakthrough design)
Continuous design vs. innovative design (6 is given for innovative design)

Table 3.3 Saaty's 1-9 scale

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Weak importance of one over another	Experience and judgment slightly favor one activity over another
5	Essential or strong importance	Experience and judgment strongly favor one activity over another
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
9	Absolute importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values between adjacent scale values	When compromise is needed

2. *The preference matrix for the complexity definitions of project dimensions is constructed (by using the reciprocal principle).*

By using the results obtained in the previous step the pairwise comparison matrix of the complexity definitions of "Technological Uncertainty" dimension is constructed as follows:

	Breakthrough	Innovative	Continuous
Breakthrough	1	5	9
Innovative	*1/5*	1	6
Continuous	*1/9*	*1/6*	1

3. *In order to apply the VAHP, square roots are taken.*

The VAHP applied pairwise comparison matrix of the complexity definitions of "Technological Uncertainty" dimension is:

	Breakthrough	Innovative	Continuous
Breakthrough	1	2,2361	3
Innovative	0,4472	1	2,4495
Continuous	0,3333	0,4082	1

4. *By applying the eigenvalue method, the priority vectors are obtained.*

The priority vector of the complexity definitions of "Technological Uncertainty" dimension is:

Breakthrough	0,8534
Innovative	0,4664
Continuous	0,2327

5. *The ideal-mode AHP normalization is applied to the priority vectors. All the elements are divided by the maximum element of the priority vector.*

The obtained v_i^N values of the complexity definitions of "Technological Uncertainty" dimension is:

Breakthrough	1,0000
Innovative	0,5465
Continuous	0,2727

All the values for the complexity definitions of other project dimensions are obtained similarly. The v_i^N values for the complexity definitions for three level PIC are given in Table 3.4.

Table 3.4 The v_i^N values for the complexity definitions of three level PIC

Dimensions of PIC	**Complexity Definitions**	v_i^N **values**
Technological Uncertainty	Breakthrough design	1,0000
	Innovative design	0,5465
	Continuous design	0,2727
Platform Type	On-the Move	1,0000
	Mobile	0,5143
	Stationary	0,2646
Work & Test Environment	Zone 3	1,0000
	Zone 2	0,4966
	Zone 1	0,3184
System Scope	Array	1,0000
	System	0,5210
	Assembly	0,3035
Amount of Resource (Labor)	More than 4X man-hours	1,0000
	Less than 4X man-hours but more than X man-hours	0,5014
	Less than X man-hours	0,2513

For each dimension of a project, the project's complexity definition's v_i^N value is multiplied with the related weight of that dimension. The VAHP weight values obtained by Tunç (2004) are used as the dimension weights. These weight values are given in Table 3.5. The projects' weighted dimension values for the three level PIC case are in Table 3.6. The clustering structure obtained by using three level PIC is given in Figure 3.3.

The projects are clustered by using these weighted dimension values. k-means algorithm that uses squared Euclidean distances is used for clustering and the required cluster number is given as three, four and five. All three cases are evaluated by using average Silhouette values of the resulting clustering structures. It is observed that the clustering structure that has five clusters, gives more compact and separate clusters. When the cluster number is three, one cluster contains nearly all the projects and other clusters contain two or three projects. The clustering structure that has three clusters is not as separate as the ones when there are four and five clusters. And the clustering structure that has four clusters is not as compact as the case when there are five clusters. Because of these reasons it is decided to separate the project into five clusters.

Table 3.5 The projects' feature VAHP weight values obtained by Tunç (2004)

Project Dimension	**VAHP weights**
Technological Uncertainty	0,6411
Platform Type	0,3451
Work & Test Environment	0,4515
System Scope	0,3733
Amount of Resource (Labor)	0,3559

Table 3.6 The weighted project dimension values obtained by using 3 level PIC

	Technological Uncertainty	Platform Type	Work & Test Environment	System Scope	Amount of Resource (Labor)
Project 1	0,3504	0,1775	0,2242	0,1133	0,0894
Project 2	0,3504	0,3451	0,2242	0,1945	0,1784
Project 3	0,3504	0,3451	0,2242	0,1945	0,0894
Project 4	0,3504	0,1775	0,2242	0,1945	0,0894
Project 5	0,3504	0,0913	0,1437	0,1133	0,0894
Project 6	0,1748	0,0913	0,2242	0,1133	0,0894
Project 7	0,6411	0,3451	0,4515	0,3733	0,3559
Project 8	0,3504	0,3451	0,4515	0,3733	0,3559
Project 9	0,3504	0,1775	0,2242	0,1133	0,1784
Project 10	0,3504	0,0913	0,4515	0,1945	0,3559
Project 11	0,3504	0,0913	0,2242	0,3733	0,3559
Project 12	0,3504	0,0913	0,1437	0,1133	0,1784
Project 13	0,3504	0,0913	0,1437	0,1945	0,1784
Project 14	0,6411	0,3451	0,2242	0,1945	0,3559

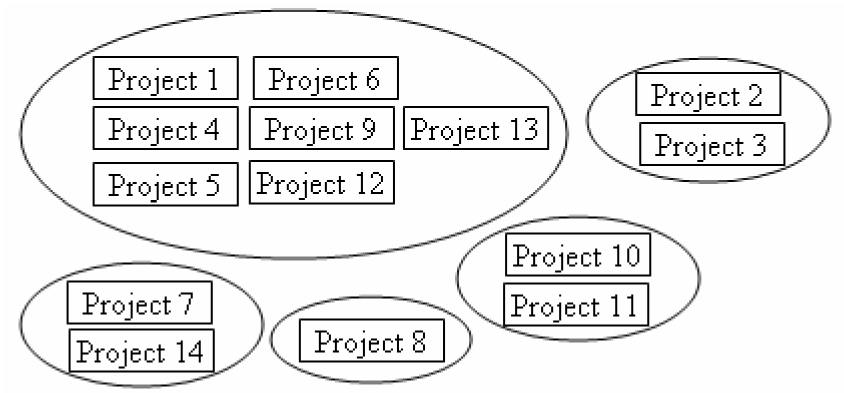

Figure 3.3 The clustering structure obtained by using three level PIC

3.2.2 The PIC Structure with Four Complexity Definitions

In this structure each dimension of PIC has four complexity definitions and each project is represented by these definitions in Table 3.7. By applying the same quantification method described for the three level PIC case, the v_i^N values are obtained for complexity definitions of four level PIC given in Table 3.8.

The complexity definitions developed for three level PIC are detailed by adding a new complexity definition to each project dimension, since some complexity definitions cover more projects than the other complexity definitions and it is better to evaluate the complexity levels of the projects within these complexity definitions separately.

Innovative design complexity definition of "Technological Uncertainty" dimension is broken down in two parts; innovative design-1 and innovative design-2. Innovative design-2 complexity definition represents projects that are more complex according to "Technological Uncertainty" dimension than the innovative design-1 projects.

Stationary complexity definition of "Platform Type" dimension is broken down in two parts; stationary-1 and stationary-2. Stationary-2 complexity definition represents projects that are more complex according to “Platform Type” dimension than projects having stationary-1 complexity definition.

Zone-2 complexity definition of “Work & Test Environment” dimension is broken down in two parts, and the new complexity definitions turned into zone-1, zone-2, zone-3 and zone-4. Zone 1 defines the projects having the least complex work & test environment specifications and zone 4 defines the projects having the most complex work & test environment specifications.

System complexity definition of “System Scope” dimension is broken down in two parts; system-1 and system-2. System-2 complexity definition represents projects that are more complex according to “System Scope” dimension than projects having system-1 complexity definition.

Complexity definitions of “Amount of Resource (Labor)” dimension are turned into;

i. systems requiring less than X/2 man-hours,
ii. systems requiring less than X man-hours and but more than X/2 man-hours,
iii. systems requiring less than 4X man-hours and but more than X man-hours,
iv. systems requiring more than 4X man-hours.

Table 3.7 The structure of four level PIC and the projects' complexity definitions

PROJECT IDENTIFICATION CARD		Project 1	Project 2	Project 3	Project 4	Project 5	Project 6	Project 7	Project 8	Project 9	Project 10	Project 11	Project 12	Project 13	Project 14
Technological Uncertainty	Breakthrough design							x							x
	Innovative design-2								x		x	x		x	
	Innovative design-1	x	x	x	x	x				x			x		
	Continuous design						x								
Platform Type	On-the Move		x	x				x	x						x
	Mobile	x			x					x					
	Stationary-2										x	x			
	Stationary-1					x	x						x	x	
Work & Test Environment	Zone 4							x	x		x				
	Zone 3		x	x											
	Zone 2	x			x		x			x		x			x
	Zone 1					x							x	x	
System Scope	Array							x	x			x			
	System-2		x		x						x				
	System-1			x										x	x
	Assembly	x				x	x			x			x		
Amount of Resource (Labor)	More than 4X man-hours							x	x		x	x			x
	Less than 4X man-hours but more than X man-hours		x							x			x	x	
	Less than X/2 man-hours but more than X man-hours			x	x	x									
	Less than X/2 man-hours	x					x								

Table 3.8 The v_i^N values for the complexity definitions of four level PIC

Dimensions of PIC	**Complexity Definitions**	v_i^N **values**
Technological Uncertainty	Breakthrough design	1,0000
	Innovative design-2	0,6882
	Innovative design-1	0,4621
	Continuous design	0,2554
Platform Type	On-the Move	1,0000
	Mobile	0,5540
	Stationary-2	0,3444
	Stationary-1	0,2469
Work & Test Environment	Zone 4	1,0000
	Zone 3	0,6067
	Zone 2	0,4289
	Zone 1	0,3195
System Scope	Array	1,0000
	System-2	0,5632
	System-1	0,4350
	Assembly	0,2933
Amount of Resource (Labor)	More than 4X man-hours	1,0000
	Less than 4X man-hours but more than X man-hours	0,5889
	Less than X man-hours but more than X/2 man-hours	0,3460
	Less than X/2 man-hours	0,2298

The projects' weighted dimension values for the four level PIC case are in Table 3.9. The clustering structure obtained by using four level PIC is given in Figure 3.4.

Table 3.9 The weighted project dimension values obtained by using four level PIC

	Technological Uncertainty	Platform Type	Work & Test Environment	System Scope	Amount of Resource (Labor)
Project 1	0,2963	0,1912	0,1937	0,1095	0,0818
Project 2	0,2963	0,3451	0,2739	0,2103	0,2096
Project 3	0,2963	0,3451	0,2739	0,1624	0,1231
Project 4	0,2963	0,1912	0,1937	0,2103	0,1231
Project 5	0,2963	0,0852	0,1442	0,1095	0,1231
Project 6	0,1637	0,0852	0,1937	0,1095	0,0818
Project 7	0,6411	0,3451	0,4515	0,3733	0,3559
Project 8	0,4412	0,3451	0,4515	0,3733	0,3559
Project 9	0,2963	0,1912	0,1937	0,1095	0,2096
Project 10	0,4412	0,1189	0,4515	0,2103	0,3559
Project 11	0,4412	0,1189	0,1937	0,3733	0,3559
Project 12	0,2963	0,0852	0,1442	0,1095	0,2096
Project 13	0,4412	0,0852	0,1442	0,1624	0,2096
Project 14	0,6411	0,3451	0,1937	0,1624	0,3559

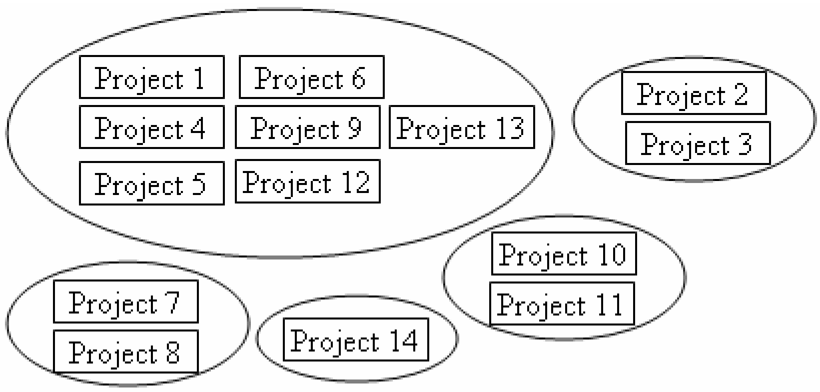

Figure 3.4 The clustering structure obtained by using four level PIC

3.2.3 The PIC Structure with Five Complexity Definitions

In this structure each dimension of PIC has five complexity definitions and each project is represented by these definitions in Table 3.10. By applying the same quantification method described for the three level PIC case, the numerical values are obtained for complexity definitions of five level PIC, Table 3.11.

The complexity definitions developed for four level PIC are detailed by adding a new complexity definition to each project dimension.

Innovative design-1 complexity definition of "Technological Uncertainty" dimension is broken down in two parts; and the new complexity definitions turned into innovative design-1, innovative design-2 and innovative design-3. The complexity level increases from innovative design-1 projects to innovative design-3 projects.

On-the move complexity definition of "Platform Type" dimension is broken down in two parts; on-the move-1 and on-the move-2. On-the move-2 complexity definition represents projects that are more complex according to "Platform Type" dimension than projects having on-the move-1 complexity definition.

Zone-4 complexity definition of "Work & Test Environment" dimension is broken down in two parts, and the new complexity definitions turned into zone-1, zone-2, zone-3, zone-4 and zone-5. Zone 1 defines the projects having the least complex work & test environment specifications and zone 5 defines the projects having the most complex work & test environment specifications.

Array complexity definition of "System Scope" dimension is broken down in two parts; array-1 and array-2. Array-2 complexity definition represents projects that are more complex according to "System Scope" dimension than projects having array-1 complexity definition.

"Systems requiring more than 4X man-hours" complexity definition of "Amount of Resource (Labor)" dimension is broken down in two parts;

i. Systems requiring less than 8X man-hours but more than 4X man-hours, and
ii. Systems requiring more than 8X man-hours.

Table 3.10 The structure of five level PIC and the projects' complexity definitions

PROJECT IDENTIFICATION CARD		Project 1	Project 2	Project 3	Project 4	Project 5	Project 6	Project 7	Project 8	Project 9	Project 10	Project 11	Project 12	Project 13	Project 14
Technological Uncertainty	Breakthrough design							x							x
	Innovative design-3								x		x	x		x	
	Innovative design-2				x	x									
	Innovative design-1	x	x	x						x			x		
	Continuous design						x								
Platform Type	On-the Move-2		x	x				x							x
	On-the Move-1								x						
	Mobile	x			x					x					
	Stationary-2										x	x			
	Stationary-1					x	x						x	x	
Work & Test Environment	Zone 5							x							
	Zone 4								x		x				
	Zone 3		x	x											
	Zone 2	x			x		x			x		x			x
	Zone 1					x							x	x	
System Scope	Array-2								x						
	Array-1							x				x			
	System-2		x		x						x				
	System-1			x										x	x
	Assembly	x				x	x			x			x		
Amount of Resource (Labor)	More than 8X man-hours							x							
	Less than 8X man-hours but more than 4X man-hours								x		x	x			x
	Less than 4X man-hours but more than X man-hours		x							x			x	x	
	Less than X man-hours but more than X/2 man-hours			x	x	x									
	Less than X/2 man-hours	x					x								

Table 3.11 The v_i^N values for the complexity definitions of five level PIC

Dimensions of PIC	**Complexity Definitions**	v_i^N **values**
Technological Uncertainty	Breakthrough design	1,0000
	Innovative design-3	0,7602
	Innovative design-2	0,5357
	Innovative design-1	0,3473
	Continuous design	0,2207
Platform Type	On-the Move-2	1,0000
	On-the Move-1	0,6405
	Mobile	0,4075
	Stationary-2	0,2735
	Stationary-1	0,2090
Work & Test Environment	Zone 5	1,0000
	Zone 4	0,8184
	Zone 3	0,5285
	Zone 2	0,3910
	Zone 1	0,2995
System Scope	Array-2	1,0000
	Array-1	0,7538
	System-2	0,5035
	System-1	0,3900
	Assembly	0,2756
Amount of Resource (Labor)	More than 8X man-hours	1,0000
	Less than 8X man-hours but more than 4X man-hours	0,7016
	Less than 4X man-hours but more than X man-hours	0,4459
	Less than X man-hours but more than X/2 man-hours	0,2829
	Less than X/2 man-hours	0,2056

The projects' weighted dimension values for the four level PIC case are in Table 3.12. The clustering structure obtained by using four level PIC is given in Figure 3.5.

Table 3.12 The weighted project dimension values obtained by using five level PIC

	Technological Uncertainty	Platform Type	Work & Test Environment	System Scope	Amount of Resource (Labor)
Project 1	0,2227	0,1406	0,1765	0,1029	0,0732
Project 2	0,2227	0,3451	0,2386	0,1880	0,1587
Project 3	0,2227	0,3451	0,2386	0,1456	0,1007
Project 4	0,3435	0,1406	0,1765	0,1880	0,1007
Project 5	0,3435	0,0721	0,1352	0,1029	0,1007
Project 6	0,1415	0,0721	0,1765	0,1029	0,0732
Project 7	0,6411	0,3451	0,4515	0,2814	0,3559
Project 8	0,4874	0,2211	0,3695	0,3733	0,2497
Project 9	0,2227	0,1406	0,1765	0,1029	0,1587
Project 10	0,4874	0,0944	0,3695	0,1880	0,2497
Project 11	0,4874	0,0944	0,1765	0,2814	0,2497
Project 12	0,2227	0,0721	0,1352	0,1029	0,1587
Project 13	0,4874	0,0721	0,1352	0,1456	0,1587
Project 14	0,6411	0,2211	0,1765	0,1456	0,2497

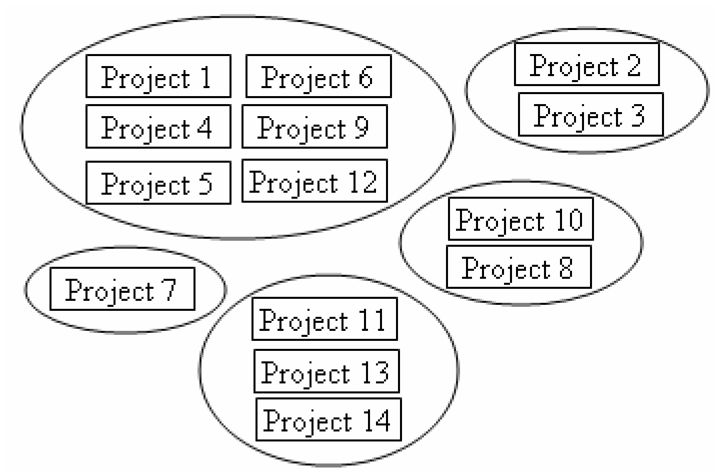

Figure 3.5 The clustering structure obtained by using five level PIC

The comparison of the clustering structure obtained by using five level PIC and the clustering structure obtained by Tunç (2004) are in Figure 3.6.

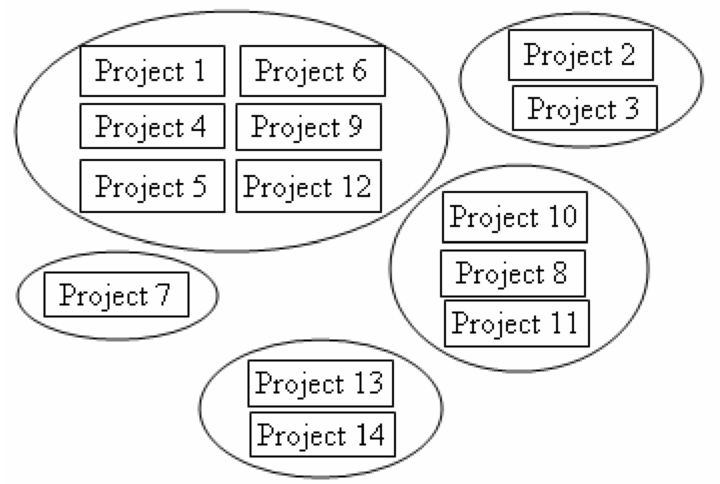

Figure 3.6 The clustering structure obtained by Tunç (2004)

The main difference between these two clustering structures is the placement of the Project 11. Project 11 is clustered together with Project 8 and Project 10 in the clustering structure obtained by Tunç (2004), whereas it is clustered with Project 13 and Project 14 with using five level PIC. Due to the project specific characteristics Project 11 has common properties with Project 8, Project 10 and Project 13, Project 14. Also these two clustering structures are evaluated by the manager of Engineering Planning Department of this electronics company as acceptable.

Tunç (2004) used relative measurement AHP method for quantifying the project dimensions and used the obtained results for clustering. In this study a Project Identification Card is constructed for defining projects categorically. The measurement scale of this PIC is developed by absolute measurement AHP method.

Relative measurement AHP method requires lots of pairwise comparisons and expert knowledge about all the considered projects' characteristics; whereas the absolute measurement method can be performed with much less pairwise comparisons and there is no need for the decision maker's knowledge about all considered projects. Therefore the approach used in this study is much more dynamic than the approach used by Tunç (2004).

The clustering structures obtained by using three, four and five level PICs are compared by using cluster quality indexes that are explained in Chapter 2. The comparison results are given in Table 3.13.

Table 3.13 The comparison of the clustering structures obtained by using three, four and five level PICs

PIC Type	Average Silhouette Value	Dunn Index	Davies Bouldin Index	SD Measure
Evaluation Criteria for Quality Measures	*larger the better*	*larger the better*	*smaller the better*	*smaller the better*
3 Level PIC	0,5415	0,5796	**0,6846**	4,7256
4 Level PIC	0,5665	0,5734	0,7206	**4,7223**
5 Level PIC	**0,6028**	**0,6676**	0,7199	4,8540

The five level PIC structure is evaluated as the best one for defining the projects among the three and four level PICs. Since;

i. The clustering structure obtained by using five level PIC is evaluated as more meaningful than the clustering structures obtained by three and four level PICs. Because it places Project 7 to a unique cluster. Project 7 is a special project that is complex according to each project dimension and should be evaluated and clustered apart from the other projects. And five level PIC satisfies this condition and the clustering of the other projects obtained by it, is also meaningful.
ii. The number of complexity definitions of five level PIC is more than the other PICs, giving the decision maker the opportunity of distinguishing the projects according to the project dimensions more easily.
iii. According to the results given in Table 3.13, the five level PIC is superior according to the average silhouette value and Dunn index. The values of

the Davies Bouldin index and the SD measure values are also not so bad for five level PIC.

Because of these stated reasons, among the three, four and five level PICs the five level PIC is chosen to use for describing the projects.

3.3 THE MATHEMATICAL MODEL OF CLUSTERING THE PROJECTS

A mathematical model of clustering that is suited the case of project clustering is constructed in order to;

i. determine an objective function for clustering the projects,
ii. analyze the changes in the clustering structure and changes in the objective function when different project sets are used (sensitivity analyses of the clusters to each project).

Abbreviations of the model:

The project dimensions are abbreviated as follows:

- TU: Technological Uncertainty,
- PT: Platform Type,
- WT: Work and Test Environment,
- SS: System Scope,
- AR: Amount of Resource (labor).

Indices:

- j : project indices j =1,…, 14
- i : project cluster indices i =1,…, 5
- k : project dimension indices k =1,…, 5 (1: TU, 2: PT, 3:WT, 4:SS, 5:AR)

Parameters:

- P_{jk} : dimension value of project j for dimension k

Variables:

- y_{ji} : $\begin{cases} 1 \text{ if project } j \text{ belongs to cluster } i \\ 0 \text{ other wise} \end{cases}$
- C_{ik} : Centroid coordinate of cluster i for dimension k
- $Intra_i$: Average intra cluster distance of cluster i

The Model:

The objective function of the model is minimizing the squared Euclidean intra cluster distances. Al-Sultan (1995) and Rao (1971) also used the within cluster squared Euclidean distances as the objective function that should be minimized.

Objective Function:

Min $\quad Z = \sum_{i=1}^{5} Intra_i$

Constraints:

1) $\sum_{i=1}^{5}\sum_{j=1}^{14} y_{ji} = 14$ → Totally 14 projects are to be assigned to project clusters.

2) $\sum_{j=1}^{14} y_{ji} \geq 1 \quad \forall\ i$ → None of the project clusters should be empty.

3) $\sum_{i=1}^{5} y_{ji} = 1 \ \forall\ j$ → Each project should be assigned to only one cluster.

4) $C_{ik} = \left(\dfrac{\sum_{j=1}^{14} \left(y_{ji} * P_{jk} \right)}{\sum_{j=1}^{14} y_{ji}} \right)$ $\forall\ i$, $\forall\ k$ → The centroid coordinate of cluster i for k th dimension.

5) $Intra_i = \dfrac{\sum_{j=1}^{14} \left(\sum_{k=1}^{5} \left(C_{ik} * y_{ji} - P_{jk} * y_{ji} \right)^2 \right)}{\sum_{j=1}^{14} y_{ji}}$ $\forall\ i$ → Average squared intra cluster distance of cluster i.

The constructed model tried to be solved with using the optimization program GAMS. Since the clustering model is a mixed integer non-linear model, the DICOPT solver of GAMS was used. But the used version of DICOPT solver was demo-version and the clustering model exceeded the limits of it and a solution could not be obtained.

Then the binary variables are represented as continuous variables in the model and new constraints for satisfying the 0-1 conditions are added to the model. The CONOPT solver of GAMS was used, whereas the solution always struck into the local optimum points and no satisfactory solution could be obtained.

Because of these reasons it is decided to construct a heuristic model that can be used for solving the clustering mathematical model. Tabu Search algorithm is used for modeling the clustering projects problem.

3.4 TABU SEARCH ALGORITHM FOR CLUSTERING THE PROJECTS

The Tabu search algorithm for clustering the projects is constructed by using MATLAB program. The clustering Tabu search algorithm uses the project dimension values obtained by five level PIC. In order to begin with a good initial point, the Euclidean distances between the projects are considered in the constructed algorithm. Firstly all the pairwise squared Euclidean distances between projects are calculated. The algorithm chooses randomly five projects and assigns each of them to a unique cluster. Then the unassigned nine projects are assigned to the nearest project that has been selected as the first members of the clusters, according to the squared Euclidean distances between them.

The inputs and the main steps of the algorithm are explained in the following sections.

3.4.1 Inputs and Parameters of the Tabu Search Algorithm

Inputs related to clustering:

i. Projects matrix: This is a project by dimension values matrix
ii. k: the number of clusters

Inputs related to Tabu search heuristic:

i. ITMAX: Maximum number of iterations.
ii. MTLS: Maximum tabu list length
iii. Pr: Probability threshold, this parameter is used for generating new trial solutions
iv. NTS: Number of trial solutions

Tabu Search Parameters:

Ac: Current solution coding. This is a 1x14 matrix; each element of this matrix shows the clusters of the projects. For example for the case of 14 projects and 5 clusters, the array in Figure 3.7 shows which project belongs to which cluster.

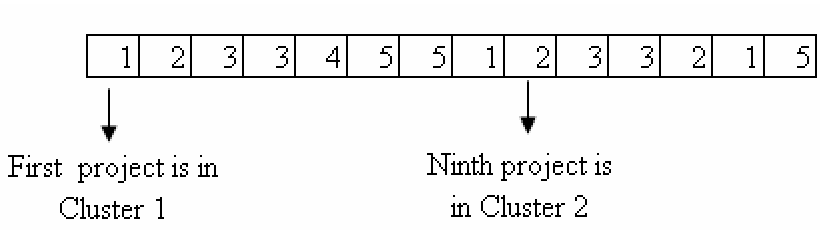

Figure 3.7 The coding structure of the current solution Ac

Ab: Best solution coding. The coding structure is the same as the one in Ac.

At: Trial solutions coding. The coding structure is the same as the one in Ac, whereas At is a NTS by 14 matrix.

Jc: Current objective function value (1x1 matrix)

Jb: Best objective function value (1x1 matrix)

Jt: Trial objective functions value (1xNTS matrix)

i: i is used as project indices

3.4.2 The Main Steps of the Tabu Search Algorithm

Step 1 - Initialization Step:

i. An initial solution Ac is found and the Jc is the corresponding objective function value (Our objective is the squared Euclidean intra cluster distances. Al-Sultan (1995) and Rao (1971) used the objective functions as the within cluster squared Euclidean distances).

ii. Let Ab = Ac and Jb = Jc

iii. Let iter_num = 1 → increase iteration number

iv. Insert Ac to the tabu list and let TLL = 1 → TLL is the Tabu list length

Step 2 - Using Ac, generating NTS trial solutions:

i. By using Ac and a probability threshold value Pr, for each project generate a random number such that $R \sim u(0,1)$. If $R < Pr$, then At(i) = Ac; otherwise generate a random integer (1 to 5: since in our case there are 5 clusters) for At(i), but this random number should be different from Ac.

ii. Evaluate the objective functions of trial solutions Jt's.

Step 3 - Selecting the current solution (Ac):

i. Order Jt's in an ascending order; Jt(1) < Jt(2) <….< Jt(NTS), since our objective is to minimize the within cluster distances.

ii. If Jt(1) is not tabu or if it is tabu but Jt(1) < Jb then Ac = At(1) and Jc= Jt(1), and go to step 4; otherwise, find the solution that has the best objective function value and that is not tabu then go to step 4. If all trial solutions (At(1), At(2), …., At(NTS)) are tabu, then go to step 2.

Step 4 - Updating tabu list and program parameters:

i. Insert Ac at the bottom of the tabu list and increase the tabu list length (TLL = TLL + 1).

ii. If TLL = MTLS + 1 then delete the first element in the tabu list and let TLL = TLL - 1.

iii. If Jc < Jb, then let Ab = Ac and Jb = Jc, if a better solution than the best solution is found change the best solution related parameters.

iv. If iter_num = ITMAX, then stop, otherwise go to step 2.

3.4.3 The Program Flow

The program flow contains two parts the initialization step, where all the beginning parameters of the model are set and then there comes the iteration loop which contains the steps 2, 3 and 4.

Initialization step:

i. Calculate the pairwise squared Euclidean distances of projects.

ii. Choose randomly 5 projects as the first members of the 5 clusters.

iii. According to the squared Euclidean distances of the other 9 unassigned projects, assign them to the nearest project that has been selected as the first member of the clusters. (after applying i, ii, iii, the Ac current solution will be obtained)

iv. Calculate the objective function value Jc of the Ac.

v. Let Ab = Ac, Jb = Jc, iter_num = 1

vi. Insert Ac to the tabu list, and let TLL = 1

<u>Iteration Loop:</u>

While iter_num < (ITMAX + 1)

For 1 to NTS → for every trial solution

While an empty cluster exists

For 1 to 14 (for every project)

Randomly assign projects to clusters

End

End

End

For 1 to NTS → for every trial solution

calculate objective functions Jt's

End

Sort Jt's → in an ascending order

If Jt(1) < Jb

Let Ac = Ab and Jc = Jb

Else

For 1 to NTS → for every trial solution, beginning from the one having the smallest obj. function value

If the trial solution is not tabu

Let Ac = At and Jc = Jt

Break the program here and exit the for loop and continue the program from the following If statement

End

End

End

If (any Jt < Jb is found) or (not all the solutions are tabu)

Let TLL = TLL + 1

Insert Ac to the bottom of the tabu list

If TLL = MTLS + 2

Delete the first element in tabu list and let TLL = TLL-1

End

If Jc < Jb then

Let Ab = Ac and Jb = Jc

End

If iter_num is not equal to ITERMAX + 1

Let iter_num = iter_num + 1

End

End

End

3.4.4 The Tabu Search Clustering Algorithm Results

For different combinations of Tabu search input parameters (different probability threshold values, different number of trial solutions – NTS values and different number of iterations), by using the five level PIC project dimension values, the Tabu search algorithm was run and the following results were obtained. The best objective function values obtained in these runs are given in Table 3.14 Table 3.15 and Table 3.16. The objective functions which are bold and italic in these tables gave the same clustering structure as the k-means clustering algorithm**.**

Table 3.14 Tabu search clustering results obtained when Pr=0.70

Tabu Search Results	Pr=0.70	
	MTLS=15	
Iteration numbers	NTS=15	NTS=20
250	0,1620	0,1740
500	0,1546	0,1507
750	0,1403	0,1545
1000	0,1674	***0,1299***

Table 3.15 Tabu search clustering results obtained when Pr=0.90

Tabu Search Results	Pr=0.90	
	MTLS=15	
Iteration numbers	NTS=15	NTS=20
250	0,1366	***0,1299***
500	0,1374	***0,1299***
750	0,1366	0,1376
1000	***0,1299***	***0,1299***

Table 3.16 Tabu search clustering results obtained when Pr=0.95

Tabu Search Results	Pr=0.95	
	MTLS=15	
Iteration numbers	NTS=15	NTS=20
250	0,1467	0,1366
500	0,1366	***0,1299***
750	0,1516	***0,1299***
1000	0,1384	***0,1299***

According to the trial Tabu search clustering results, as the probability threshold value, number of trial solutions (NTS) value and the number of iterations increases the Tabu Search algorithm finds a better solution. And the Tabu Search input parameters given in Table 3.17 are found to be the best input parameters for obtaining a meaningful clustering structure for the project clustering case.

Table 3.17 Best Tabu search parameter values for clustering the R&D projects case

Probability Threshold	0.95
Maximum Tabu List Length	15
Maximum Number of Iterations	1000
Number of Trial Solutions	20

3.5 SENSITIVITY OF CLUSTERS TO EACH PROJECT

In order to examine the changes in clustering structure in the absence of any project – (sensitivity analysis according to the absence and existence of a project), the projects are re-clustered for all the cases, in which a project is not included in the projects' list. And for each project the obtained clustering structures are compared with the clustering structure that has 14 projects.

In order to re-cluster the projects, the Tabu search clustering algorithm is used. The Tabu search best parameter values those are previously obtained from the trial runs are used (Table 3.17). The objective functions of the clustering structures obtained when clustering 13 projects and clustering 14 projects are given in Table 3.18. The objective function value of the Tabu search clustering algorithm is minimizing the squared intra-cluster distance.

As it can be seen from Table 3.18, the existence of each project causes an increase in the objective function value. By analyzing the percentage of these changes the projects that constitute the main structure of clustering and the projects that do not have so much effect on the clustering structure can be found.

Table 3.18 The objective function values of clustering 13 projects and 14 projects

Omitted Project	Objective function values of clustering without the omitted project (13 Projects Case)	Objective function values of clustering with the omitted project (14 Projects Case)	Improvement of the objective function values from the 14 Projects Case to the 13 Projects Case
Project 1	0.1254	0.1299	3%
Project 2	0.1273	0.1299	2%
Project 3	0.1273	0.1299	2%
Project 4	0.1114	0.1299	14%
Project 5	0.1166	0.1299	10%
Project 6	0.1123	0.1299	14%
Project 7	0.0967	0.1299	26%
Project 8	0.1038	0.1299	20%
Project 9	0.1244	0.1299	4%
Project 10	0.1042	0.1299	20%
Project 11	0.1101	0.1299	15%
Project 12	0.1237	0.1299	5%
Project 13	0.1097	0.1299	16%
Project 14	0.0967	0.1299	26%

Re-clustering the projects by omitting each project one at a time, the following similarities to and the deviations from the clustering structure obtained by clustering 14 projects case are obtained:

1. In the absence of either Project 7, or Project 8, or Project 10, or Project 14 the objective function of the clustering structure that is obtained by clustering 14 projects increases by more than 20%. Therefore it can be concluded that these projects' existence causes distortions in the clustering structure.

2. In the absence of some projects, only the omitted project move away from its cluster and all the other projects stay within their clusters. Therefore it can be concluded that the current clustering structure will not be affected by the existence or absence of these projects. These projects are Project 1,

Project 2, Project 3, Project 4, Project 5, Project 6, Project 9, Project 11, Project 12 and Project 14. Whereas in the absence of Project 7, Project 8, Project 10 and Project 14 the clustering structure changes. The comparison of the changes in the clustering structure in the absence of these projects are given in Figure 3.8, Figure 3.9, Figure 3.10 and Figure 3.11.

3. In the absence of Project 7; Project 14 formed a new one-member-cluster, and the clusters of all the other projects stayed same.

4. In the absence of Project 8; Project 14 formed a new one-member-cluster, and Project 10, Project 11, Project 13 are formed another cluster and the clusters of all the other projects stayed same.

5. In the absence of Project 10; Project 7 and Project 8 formed a new cluster, Project 14 formed a new one-member-cluster and the clusters of all the other projects stayed same.

6. In the absence of Project 13; Project 14 formed a new one-member-cluster, and Project 8, Project 10 and Project 11 are formed another cluster and the clusters of all the other projects stayed same.

In all the clustered 14 projects, some have special properties that can be used to explain the above results. Because of these special properties these projects occupy a different place in the five dimensional project space. Project 7 is a special project that is complex according to each project dimension and should be evaluated and clustered apart from the other projects. Project 14 is complex according to “Technological Uncertainty”, “Platform Type” and “Amount of Resource (Labor)” dimensions and not so complex according to “Work & Test Environment” and “System Scope” dimensions. Since Project 14 has a different complexity definition pattern from the other projects, the cluster of it is affected by other projects absence. Project 8 is a complex commercial project that does not have military characteristics, but has special characteristics that distinguishes it from the other projects. Project 10 is complex according to “Work & Test Environment” and “Amount of Resource (Labor)” dimensions and not so complex according to other dimensions.

By considering the above stated information, the following implications can be made.

i. Project 7, Project 8, Project 10, and Project 13 constitute the main structure of the clustering and the other projects do not effect the clustering as much as them.
ii. Project 14 tends to form a new cluster, but the existence of Project 7, Project 8, Project 10 and Project 13 prevents Project 14 from forming a new cluster.
iii. Project 8 and Project 10 tend to be in the same cluster. Because of this reason in the absence of one of them the other one enters into another clusters.

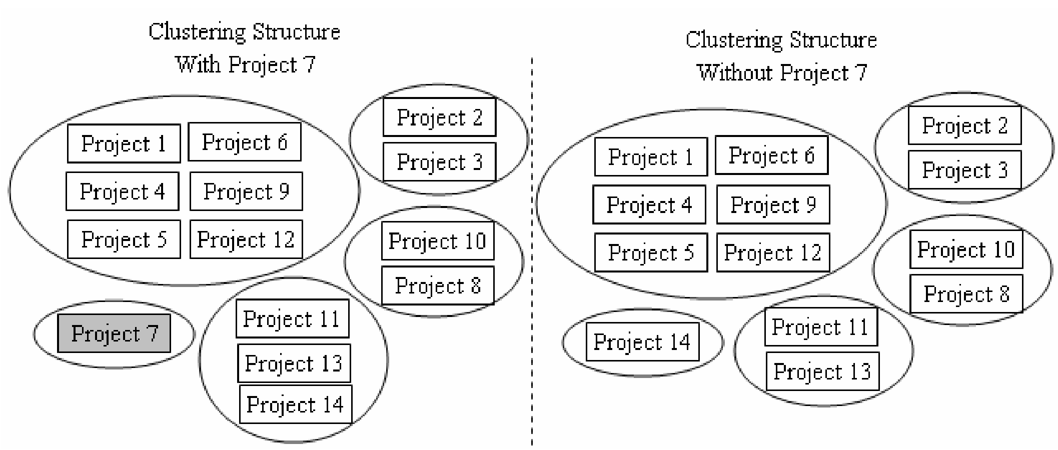

Figure 3.8 The clustering structure obtained by using all 14 projects and the clustering structure obtained by clustering projects other than Project 7

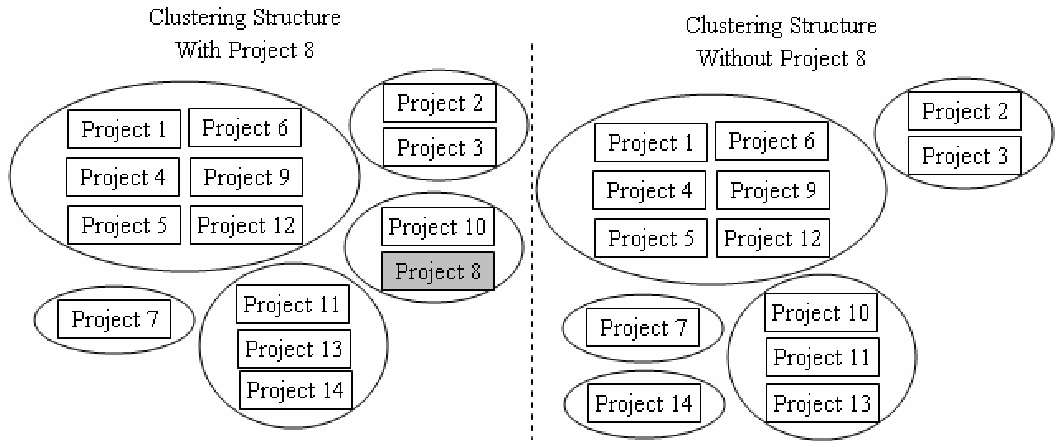

Figure 3.9 The clustering structure obtained by using all 14 projects and the clustering structure obtained by clustering projects other than Project 8

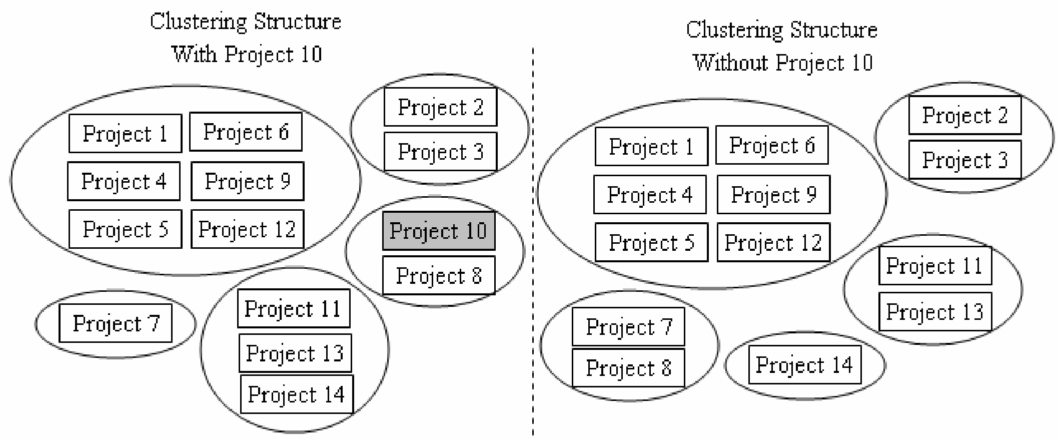

Figure 3.10 The clustering structure obtained by using all 14 projects and the clustering structure obtained by clustering projects other than Project 10

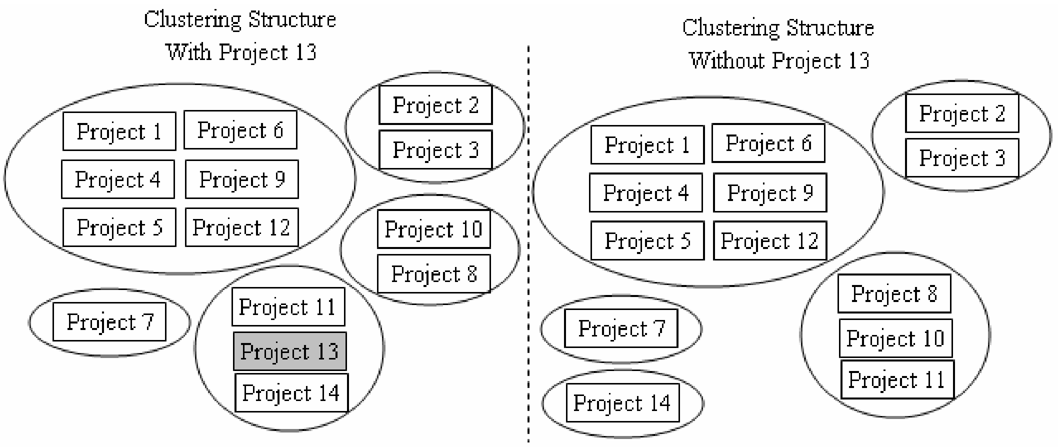

Figure 3.11 The clustering structure obtained by using all 14 projects and the clustering structure obtained by clustering projects other than Project 13

3.6 CLASSIFICATION OF A NEW PROJECT

After having constructing the project clusters, determining the cluster of a new project turns into a classification problem, since the project clusters are formed. Before deciding on the cluster of a new project, the project should be represented numerically by using the same measures as the previously clustered projects. The PIC is used for this purpose. The complexity definitions of the new project are decided by the decision maker and the numerical values corresponding to the complexity definitions are used as the dimension values of this project.

By this method the following benefits are obtained:

i. The new project is represented using the same dimensions and same measurement scale of the previously clustered projects,
ii. The decision maker can construct the numerical values for each dimension of PIC, by using categorical information about each dimension of the project,
iii. Representing the new project numerically for a decision maker, does not depend on his knowledge and experience of all previously clustered projects,

iv. After the PIC is constructed, this method does not require pairwise comparisons, and the time it takes to construct the numerical values of a project for each dimension is much more less than the relative measurement AHP method.

Once the new project is represented by numerically in five dimensional project space, according to some distance metric the new project should be assigned to a project class. In this study two different distance metrics are evaluated and one of them is suggested for further use.

3.6.1 The Application of Classification of an Incoming Project

In order to examine any classification method, the question of what will happen when an incoming project is tried to be classified should be evaluated. In this study the considered 14 base projects are all considered separately like an incoming project, and the cases in which they are tried to be classified are evaluated project by project.

Using the dimension values obtained by five level PIC, all projects are treated separately and classified to the project clusters using two different distance metrics. The first classification method assigns an incoming project to the existing clusters by considering the Euclidean distance between the incoming project and the centroids of the existing clusters. The second method assigns an incoming project to the clusters according to the average distance of the incoming project to the projects within the same cluster. Both methods assign the incoming project to a cluster according to the minimum of the analyzed distances.

Let Project i among 14 base projects be the incoming project, where $i=1, 2,...,14$. The objective function and the average silhouette values of the clustering structures

i. without Project i,

ii. if Project i is classified to an existing project cluster, and

iii. if Project i forms a new project cluster are given in Table 3.19 for the distance metric of "minimum distance to cluster centroids" and Table 3.20 for the distance metric of "minimum average distance to projects within a cluster". Fisher's classification method requires defining costs of misclassifying data points and defining the densities of the clusters. Whereas in our proposed classification method there is no need for defining misclassification costs for data points and densities for the clusters. Due to the nature of our problem we deal with less data points and this makes it harder to define densities for each cluster. Determining the cluster of a new data point according to some predetermined distance metrics is more suitable for our case.

Table 3.19 The classification results of "minimum distance to cluster centroids" distance metric

	Without Project i		If Project i is assigned to an existing cluster		If Project i forms a new cluster	
	Average Silhouette Values	Objective Function	Average Silhouette Values	Objective Function	Average Silhouette Values	Objective Function
Project 1	0,5824	0,1254	0,6029	0,1299	0,2374	0,1254
Project 2	0,5773	0,1273	0,6029	0,1299	0,6075	0,1273
Project 3	0,5873	0,1273	0,6029	0,1299	0,6075	0,1273
Project 4	0,6246	0,1114	0,6029	0,1299	0,4383	0,1114
Project 5	0,6036	0,1166	0,6029	0,1299	0,3698	0,1166
Project 6	0,5969	0,1123	0,6029	0,1299	0,4302	0,1123
Project 7	0,5942	0,1299	0,5804	0,1810	0,6029	0,1299
Project 8	0,6684	0,1047	0,6029	0,1299	0,6921	0,1047
Project 9	0,5879	0,1244	0,6029	0,1299	0,2434	0,1244
Project 10	0,7075	0,1047	*0,6306*	*0,1403*	0,6921	0,1047
Project 11	0,6286	0,1101	0,6029	0,1299	0,5408	0,1101
Project 12	0,5784	0,1237	0,6029	0,1299	0,2702	0,1237
Project 13	0,6018	0,1113	0,6029	0,1299	0,4831	0,1113
Project 14	0,6157	0,0967	0,6029	0,1299	0,6354	0,0967

Table 3.20 The classification results of “minimum average distance to projects within a cluster” distance metric

	Without Project i		If Project i is assigned to an existing cluster		If Project i forms a new cluster	
	Average Silhouette Values	Objective Function	Average Silhouette Values	Objective Function	Average Silhouette Values	Objective Function
Project 1	0,5824	0,1254	0,6029	0,1299	0,2374	0,1254
Project 2	0,5773	0,1273	0,6029	0,1299	0,6075	0,1273
Project 3	0,5873	0,1273	0,6029	0,1299	0,6075	0,1273
Project 4	0,6246	0,1114	0,6029	0,1299	0,4383	0,1114
Project 5	0,6036	0,1166	0,6029	0,1299	0,3698	0,1166
Project 6	0,5969	0,1123	0,6029	0,1299	0,4302	0,1123
Project 7	0,5942	0,1299	0,5804	0,181	0,6029	0,1299
Project 8	0,6684	0,1047	0,6029	0,1299	0,6921	0,1047
Project 9	0,5879	0,1244	0,6029	0,1299	0,2434	0,1244
Project 10	0,7075	0,1047	0,6029	0,1299	0,6921	0,1047
Project 11	0,6286	0,1101	0,6029	0,1299	0,5408	0,1101
Project 12	0,5784	0,1237	0,6029	0,1299	0,2702	0,1237
Project 13	0,6018	0,1113	0,6029	0,1299	0,4831	0,1113
Project 14	0,6157	0,0967	0,6029	0,1299	0,6354	0,0967

The first method classifies all projects to their previous clusters, except Project 10. It classifies Project 10 to the cluster that contains Project 11, Project 13, and Project 14. The classification of Project 10 to this cluster is not a desirable result.

The second method takes into consideration all the distances between the incoming project and the previously clustered projects, and classifies all the projects to their previous clusters. Since the second method considers all the distances between the incoming project and the projects in project clusters, it classifies an incoming project better than the first method.

For deciding the changes in the clustering structure in cases of classifying the incoming project to an existing cluster or forming a new cluster for it, is evaluated by analyzing the percentage of increase in the average silhouette values of the clustering structure of these cases. Table 3.21 shows these values for each project.

Table 3.21 Percentage increase in average silhouette values of the clustering structures for projects

	Without Project i	If Project i is assigned to an existing cluster		If Project i forms a new cluster	
	First case	Second case		Third case	
	Average Silhouette Values	Average Silhouette Values	% increse from first case	Average Silhouette Values	% increse from second case
Project 1	0,5824	0,6029	0,0352	0,2374	-0,6062
Project 2	0,5773	0,6029	0,0443	0,6075	**0,0076**
Project 3	0,5873	0,6029	0,0266	0,6075	**0,0076**
Project 4	0,6246	0,6029	-0,0347	0,4383	-0,2730
Project 5	0,6036	0,6029	-0,0012	0,3698	-0,3866
Project 6	0,5969	0,6029	0,0101	0,4302	-0,2864
Project 7	0,5942	0,5804	-0,0232	0,6029	**0,0388**
Project 8	0,6684	0,6029	-0,0980	0,6921	**0,1480**
Project 9	0,5879	0,6029	0,0255	0,2434	-0,5963
Project 10	0,7075	0,6029	-0,1478	0,6921	**0,1480**
Project 11	0,6286	0,6029	-0,0409	0,5408	-0,1030
Project 12	0,5784	0,6029	0,0424	0,2702	-0,5518
Project 13	0,6018	0,6029	0,0018	0,4831	-0,1987
Project 14	0,6157	0,6029	-0,0208	0,6354	**0,0539**

According to the percentage increase values represented in Table 3.21, a threshold value for determining the forming of a new cluster or not, could not be established. Since the least percentage increase belongs to Project 7 and it is desirable for Project 7 to form a new cluster and not desirable for Project 2, Project 3, Project 8 and Project 14 to form a new cluster.

For Project 8 and Project 10, the percentage increase in the objective function from second case (the case of assigning Project i to an existing cluster) to third case (the case that Project i forms a new cluster) is higher than the other projects. If six were a desirable cluster number then Project 8 and Project 10 would form new clusters. But dividing 14 projects into 6 clusters is not desirable and evaluating Project 8 and Project 10 in the same cluster is not an unacceptable situation, therefore the case of forming new clusters for Project 8 and Project 10 is not a desirable result.

According to the analyses of classification of projects, the classification of an incoming project should be performed by using the distance metric that uses "the average distance to projects within a cluster". For an incoming project this metric for each cluster should be calculated and the incoming project should be assigned to the cluster that has the minimum value of this metric.

Also the conditions of assigning the incoming project to an existing cluster and forming a new cluster for it should be evaluated. If the objective function of the clustering structure obtained by forming a new cluster for the incoming project is better than assigning it to an existing cluster than the decision of the forming of a new cluster should be given by;

i. consulting the decision makers (the decision maker can be either the project manager or a manager of one of the R&D departments) and
ii. evaluating the complexity definition pattern of the incoming project, if the incoming project shows a very different pattern than the existing projects then a cluster can be formed for it.

CHAPTER 4

CONCLUSION AND RECOMMENDATIONS

In this thesis, a methodology for classifying a new project after establishing the project clusters is studied. A template which is called Project Identification Card – PIC, for quantifying the projects according to project dimensions (technological uncertainty, platform type, work and test environment, system scope, and amount of resource-labor) is developed. Complexity definitions suitable for these dimensions are defined. Absolute measurement method of Analytical Hierarchy Process is used for constructing the numerical scale of the PIC. The clustering structure obtained by using this developed PIC is analyzed, and a classification scheme for a new project is proposed.

While constructing the PIC, the resolution level of the complexity definitions are decided by analyzing different PIC structures with three, four and five level complexity definitions. Since the PIC that has five complexity definitions for each dimension, gives decision maker more choices for describing a project and the clustering structure obtained by using this PIC is evaluated more desirable than the ones obtained by using three and four level PICs, it is decided to form five complexity definitions for each project dimension.

The clustering of the projects is modeled mathematically, for defining an objective function and for analyzing the changes in the clustering structure and the objective function while conducting sensitivity analyses of projects' existence and absence. For this purpose a Tabu search clustering projects algorithm is constructed. The best parameter setting of the algorithm is found by trying alternative parameter sets. The obtained parameter setting is used while conducting the sensitivity analyses and while determining the class of a new project.

The sensitivity analysis of projects' existence and absence analysis showed that some projects are the main building blocks of the clustering structure whereas some projects are not as important as the others in shaping the clustering structure. The existence or absence of latter ones does not change the clustering structure whereas the existence and absence of former ones changes the clustering structure.

Project 7, Project 8, Project 10 and Project 13 are the projects those establish the main structure of the current clustering structure. They shape the current clustering structure, and their existence and absence affects the clustering structure. The existence and absence of Project 1, Project 2, Project 3, Project 4, Project 5, Project 6, Project 9, Project 11, Project 12, and Project 14 does not affect the current clustering structure. The projects that have influence on the clustering structure have special characteristic properties. For example Project 7 is a complex project according to all project dimensions.

Two different distance metrics are analyzed and the classification method that assigns a new project to the project clusters according to the minimum "average distance between the new project and the projects in a cluster" distance metric is recommended for use, since it classifies all the current projects to their clusters. Our proposed classification method is more suitable to our case than the Fisher's classification method. In our method there is no need for defining the costs of misclassification of projects and densities of the clusters. Also the number of data points in our case is not adequate to define the densities for the clusters.

The decision about whether a new project should be classified to the existing clusters or it should form a new cluster should be given by analyzing the complexity definition pattern of the new project, rather than analyzing the changes in silhouette values. Since a threshold value for determining such a decision can not be obtained from the classification results of 14 projects. And by analyzing the classification results of Project 7, it is concluded that Project 7 forms a new cluster because of its complexity definition pattern. Therefore, if a new project has a different complexity definition pattern than the previous projects, the forming of a new cluster should be evaluated. Also the forming of a new cluster decision should be given by the approval of the decision makers.

By analyzing the projects within the same complexity definition, detailed explanations for developed complexity definitions can be formed as a future work. For each project cluster, a definition that can be used to define that cluster can be developed by analyzing the projects within the same cluster.

The managers of the R&D departments of this electronics company, the project managers and project technical managers should involve in the process of developing the numerical scale of Project Identification Card (PIC). Either the group preference matrix can be constructed by aggregating the individual preferences by applying geometric mean to each pairwise comparison or a unique group preference matrix can be constructed by consensus.

The Delphi method can be applied for developing the numerical scale of the PIC. First pairwise comparison results of all the decision makers that involve in the decision process can be collected. After analyzing the obtained results, group aggregate results and individual results and the related numerical scale of the PIC for both cases can be sent to decision makers and the numerical scale can be developed in a meeting in which all the decision makers will be attend.

The importance of deciding the right complexity definition for a new project can also be studied as a future work. For each project dimension the accuracy of choosing the complexity definitions can be analyzed. The sensitivities of project dimensions, complexity definitions and the importance of the dimension weights on these sensitivities can be analyzed.

The project specific data like engineering change orders, documentation statistics, and resource (labor, budget, material…) usage schemes can be analyzed for each cluster. The data structures within a cluster, between clusters and the data structures according to the project dimensions can be analyzed. Taking into consideration the analyses results, the project management activities should be tailored according to the underlying data structures and main characteristics of projects and project clusters.

REFERENCES

Aczel, J., Saaty, T. L., 1983, “Procedures for synthesizing ratio judgments”, Journal of Mathematical Psychology, 27, 93-102.

Al-Sultan, K., 1995, “A Tabu search approach to the clustering problem”, Pattern Recognition, Vol. 28, No. 9, 1443-1451.

Ball, G., H., 1971. Classification Analysis, Stanford Research Institute, SRI Project 5533.

Belton, V., Gear, T., 1983, “On a short-coming of Saaty’s method of analytic hierarchies”, Omega, 228-230.

Ben-Arieh, D., Triantaphyllou, E., 1992, “Quantifying data for group technology with weighted fuzzy features”, International Journal of Production Research, Vol. 30, No. 6, 1285-1299.

Berry, M., J., Linoff, G., 1996. Data mining techniques for marketing, Sales and Customer Support. John Wiley & Sons, Inc.

Capability Maturity Model Integration (CMMI^{sm}), 2001. Version 1.1 Staged Representation. (CMMI-SE/SW, V1.1).

Everitt, B., 1974. Cluster Analysis. John Wiley & Sons, Inc., New York.

Fayyad, U., M., Gregory Piatesky-Shapiro, Padhraic Smuth and Ramasamy Uthurusamy, 1996. Advances in Knowledge Discovery and Data Mining. AAAI Press.

Forman, E. H., 1996. The analytic hierarchy process and Expert Choice. McLean, VA.

Glover, F., Laguna, M., 1997. Tabu Search, Kluwer Academic Publishers, USA.

Golden, B.L., Wasil, E.A., Harker, P.T., 1989. The Analytic Hierarchy Process Applications and Studies, Springer-Verlag.

Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2001, “Clustering algorithms and validity measures”, Proceedings of SSDBM Conference, Virginia, USA.

Halkidi, M., Vazirgiannis, M., 2000, “Quality assessment in the clustering process”, Proceedings of PDD, Lyon, France.

Jain, A. K., Dubes., R.C., 1988. Algorithms for clustering data. Prentice Hall, New Jersey.

Jain, A. K., Murty, M. N., Flynn, P. J., 1999, “Data Clustering: A Review”, ACM Computing Surveys, Vol. 31, No. 3, 265-323.

Johnson, R., A., Wichern, D., W., 2002. Applied Multivariate Statistical Analysis, Prentice Hall.

Kaufman., L., Rousseuw., 1989. Finding Groups in Data: An Introduction to Cluster Analysis, Jhon Wiley & Sons, Inc.

Lodha, S., K., Kumar, K., Kumar, A., 2005, “Semi-Automatic Roof Reconstruction from Aerial Lidar Data Using k-means With Refined Seeding”, ASPRS 2005 Annual Conference Baltimore, Maryland.

Milligan, G.W., Cooper, M.C., 1985, “An examination of procedures for determining the number of clusters in a data set”, Psychometrika, Vol. 50, No. 2, 159-179.

Moody, J. A., Chapman, W. L., Van Voorhees, F., D., Bahill, A. T., 1997. Metrics and Case Studies for Evaluating Engineering Designs, Prentice Hall PTR.

Project Management Institute, 2000. A Guide to the Project Management Body of Knowledge, Newtown Square, Pensylvania, USA.

Rao, M. R., 1971, “Cluster Analysis and Mathematical Programming”, Journal of American Statistical Association, Vol. 66, No. 335, 622-626.

Rayward-Smith, V.J., Osman, I. H., Reeves, C. R., Smith, G.D., 1996. Modern Heuristic Search Methods, John Wiley & Sons, England.

Saaty, Thomas L., 1980. The Analytic Hierarchy Process, McGraw-Hill.

Saaty, Thomas L., 1982, “Priority Setting in Complex Problems”, Proceedings of the Fifth International Conference on Multiple Criteria Decision Making, Mons, Belgium.

Saaty, Thomas L., 1986, “Absolute and relative measurement with the AHP, the most livable cities in the United States”, Socio-Economic Planning Sciences, Vol. 20, No. 6, 327-331.

Saaty, Thomas L., 1994, “How to make a decision: The Analytic Hierarchy Process”, Interfaces, Vol.24, No. 6, 19-43.

Sen, P., Yang, J.B., 1998. Multiple Criteria Decision Support in Engineering Design, Springer-Verlag London.

Shenhar, A.,J., 2001, “One size does not fit all projects: exploring classical contingency domains”, Management Science, Vol. 47, No.3, 394-414.

Shenhar, A.,J., Dvir, D., Lipovetsky, S., Tishler, A., 1998, “In search of project classification: a non-universal approach to project success factors”, Research Policy, Vol 27, 915-935.

STATISTICA (the statistical analysis and graphics software) help manual, 2005.

Theodoridis, S., Koutroubas, K., 1999. Pattern Recognition. Academic Press.

Triantaphyllou E., Mann S. H., 1995, “Using the analytic hierarchy process for decision making in engineering applications: Some challenges”, International Journal of Industrial Engineering, Vol. 2, No. 1, 35-44.

Tunç M., 2004, “An application of clustering R&D projects by using the analytic hierarchy process”-Thesis, Bilkent University.

Zahir, S., 1999, “Clusters in a group: decision making in the vector space formulation of the analytic hierarchy process”, European Journal of Operational Research, 112, 620-634.

Zahir, S., 1999, “Geometry of decision making and the vector space formulation of the analytic hierarchy process”, European Journal of Operational Research, 112, 373-396.

Zahir, S., 2002, “Using the analytic hierarchy process for quantifying and classifying objects with multiple attributes,” Infor 40 (2), 149-172.

APPENDIX A

INPUT DATA: PAIRWISE COMPARISONS

Table A.1 Pairwise comparison matrix of the complexity definitions of "Technological Uncertainty" dimension for three level Project Identification Card

	Breakthrough	Innovative	Continuous
Breakthrough	1	5	9
Innovative	*1/5*	1	6
Continuous	*1/9*	*1/6*	1

Table A.2 Pairwise comparison matrix of the complexity definitions of "Platform Type" dimension for three level Project Identification Card

	On-the Move	Mobile	Stationary
On-the Move	1	6	9
Mobile	*1/6*	1	6
Stationary	*1/9*	*1/6*	1

Table A.3 Pairwise comparison matrix of the complexity definitions of "Work & Test Environment" dimension for three level Project Identification Card

	Zone 3	Zone 2	Zone 1
Zone 3	1	5	8
Zone 2	*1/5*	1	3
Zone 1	*1/8*	*1/3*	1

Table A.4 Pairwise comparison matrix of the complexity definitions of “System Scope” dimension for three level Project Identification Card

	Array	System	Assembly
Array	1	5	8
System	*1/5*	1	4
Assembly	*1/8*	*1/4*	1

Table A.5 Pairwise comparison matrix of the complexity definitions of “Amount of Resource (Labor)” dimension for three level Project Identification Card

	A	B	C
A	1	7	9
B	*1/7*	1	7
C	*1/9*	*1/7*	1

A: More than 4X man-hours

B: Less than 4X man-hours but more than X man-hours

C: Less than X man-hours

Table A.6 Pairwise comparison matrix of the complexity definitions of “Technological Uncertainty” dimension for four level Project Identification Card

	Breakthrough	Innovative-2	Innovative-1	Continuous
Breakthrough	1	3	6	9
Innovative-2	*1/3*	1	4	6
Innovative-1	*1/6*	*1/4*	1	7
Continuous	*1/9*	*1/6*	1/7	1

Table A.7 Pairwise comparison matrix of the complexity definitions of "Platform Type" dimension for four level Project Identification Card

	On-the Move	Mobile	Stationary-2	Stationary-1
On-the Move	1	6	8	9
Mobile	*1/6*	1	4	6
Stationary-2	*1/8*	*1/4*	1	3
Stationary-1	*1/9*	*1/6*	*1/3*	1

Table A.8 Pairwise comparison matrix of the complexity definitions of "Work & Test Environment" dimension for four level Project Identification Card

	Zone 4	Zone 3	Zone 2	Zone 1
Zone 4	1	3	6	8
Zone 3	*1/3*	1	2	4
Zone 2	*1/6*	*1/2*	1	2
Zone 1	*1/8*	*1/4*	*1/2*	1

Table A.9 Pairwise comparison matrix of the complexity definitions of "System Scope" dimension for four level Project Identification Card

	Array	System-2	System-1	Assembly
Array	1	4	6	8
System-2	*1/4*	1	2	4
System-1	*1/6*	*1/2*	1	3
Assembly	*1/8*	*1/4*	*1/3*	1

Table A.10 Pairwise comparison matrix of the complexity definitions of “Amount of Resource (Labor)” dimension for four level Project Identification Card

	A	B	C	D
A	1	6	8	9
B	*1/6*	1	5	8
C	*1/8*	*1/5*	1	4
D	*1/9*	*1/8*	1/4	1

A: More than 4X man-hours

B: Less than 4X man-hours but more than X man-hours

C: Less than X/2 man-hours but more than X man-hours

D: Less than X/2 man-hours

Table A.11 Pairwise comparison matrix of the complexity definitions of “Technological Uncertainty” dimension for five level Project Identification Card

	A	B	C	D	E
A	1	3	5	8	9
B	*1/3*	1	4	6	9
C	*1/5*	*1/4*	1	5	8
D	*1/8*	*1/6*	*1/5*	1	6
E	*1/9*	*1/9*	*1/8*	*1/6*	1

A: Breakthrough

B: Innovative-3

C: Innovative-2

D: Innovative-1

E: Continuous

Table A.12 Pairwise comparison matrix of the complexity definitions of "Platform Type" dimension for five level Project Identification Card

	A	B	C	D	E
A	1	6	8	9	9
B	*1/6*	1	5	7	9
C	*1/8*	*1/5*	1	4	6
D	*1/9*	*1/7*	*1/4*	1	3
E	*1/9*	*1/9*	*1/6*	*1/3*	1

A: On-the Move-2

B: On-the Move-1

C: Mobile

D: Stationary-2

E: Stationary-1

Table A.13 Pairwise comparison matrix of the complexity definitions of "Work & Test Environment" dimension for five level Project Identification Card

	A	B	C	D	E
A	1	2	4	6	8
B	*1/2*	1	3	5	7
C	*1/4*	*1/3*	1	2	4
D	*1/6*	*1/5*	*1/2*	1	2
E	*1/8*	*1/7*	*1/4*	*1/2*	1

A: Zone 5

B: Zone 4

C: Zone 3

D: Zone 2

E: Zone 1

Table A.14 Pairwise comparison matrix of the complexity definitions of "System Scope" dimension for five level Project Identification Card

	A	B	C	D	E
A	1	3	4	6	8
B	*1/3*	1	3	5	7
C	*1/4*	*1/3*	1	2	4
D	*1/6*	*1/5*	*1/2*	1	3
E	*1/8*	*1/7*	*1/4*	*1/3*	1

A: Array-2

B: Array-1

C: System-2

D: System-1

E: Assembly

Table A.15 Pairwise comparison matrix of the complexity definitions of "Amount of Resource (Labor)" dimension for five level Project Identification Card

	A	B	C	D	E
A	1	5	7	9	9
B	*1/5*	1	6	8	9
C	*1/7*	*1/6*	1	5	8
D	*1/9*	*1/8*	*1/5*	1	4
E	*1/9*	*1/9*	*1/8*	*1/4*	1

A: More than 8X man-hours

B: Less than 8X man-hours but more than 4X man-hours

C: Less than 4X man-hours but more than X man-hours

D: Less than X man-hours but more than X/2 man-hours

E: Less than X/2 man-hours

APPENDIX B

THE CODE OF TABU SEARCH CLUSTERING ALGORITHM

The following code is written by using MATLAB 6 Release 13.

```
Function [AC,AB,JC,JB] =
TabuSearchClusterIntra(projects,k,ITMAX,MTLS,P,NTS)
% projects is the project by feature matrix
% k is the number of clusters
% ITMAX is the number of tabu search iterations
% MTLS is the maximum tabu list length
% p is the probability threshold
% NTS is the number of trial solutions
%**************************************************************
% for finding the number of projects
   findprnr = size(projects);
   nr_projects = findprnr(1,1);
%**************************************************************
% deciding the initial point
   % calculating the pair wise distances of projects
   pd = pdist(projects);
   pdmatris = squareform(pd);

   % select randomly 5 points for clusters
   CL1ROOT = unidrnd(nr_projects,1,1);
   CL2ROOT = unidrnd(nr_projects,1,1);
   while (CL2ROOT==CL1ROOT)
      CL2ROOT = unidrnd(nr_projects,1,1);
```

```
    end
    CL3ROOT = unidrnd(nr_projects,1,1);
    while ((CL2ROOT==CL3ROOT)|(CL1ROOT==CL3ROOT))
        CL3ROOT = unidrnd(nr_projects,1,1);
    end
    CL4ROOT = unidrnd(nr_projects,1,1);
    while
((CL2ROOT==CL4ROOT)|(CL1ROOT==CL4ROOT)|(CL4ROOT==CL3ROOT
))
        CL4ROOT = unidrnd(nr_projects,1,1);
    end
    CL5ROOT = unidrnd(nr_projects,1,1);
    while
((CL5ROOT==CL1ROOT)|(CL5ROOT==CL2ROOT)|(CL5ROOT==CL3ROOT
)|(CL5ROOT==CL4ROOT))
        CL5ROOT = unidrnd(nr_projects,1,1);
    end
    AC = zeros(1,10);
    AC(1,CL1ROOT) = 1;
    AC(1,CL2ROOT) = 2;
    AC(1,CL3ROOT) = 3;
    AC(1,CL4ROOT) = 4;
    AC(1,CL5ROOT) = 5;
    % assign other 9 projects to the 5 project clusters
    for i=1:nr_projects
        ay = pdmatris(i,CL1ROOT);
        by = pdmatris(i,CL2ROOT);
        cy = pdmatris(i,CL3ROOT);
        dy = pdmatris(i,CL4ROOT);
        ey = pdmatris(i,CL5ROOT);
        if (min ([ay by cy dy ey]) == ay)
```

```
            AC(1,i) = 1;
        elseif (min ([ay by cy dy ey]) == by)
            AC(1,i) = 2;
        elseif (min ([ay by cy dy ey]) == cy)
            AC(1,i) = 3;
        elseif (min ([ay by cy dy ey]) == dy)
            AC(1,i) = 4;
        else
            AC(1,i) = 5;
        end
    end
%*****************************************************************
% calculating the initial objective function value
    % calculating the cluster centers
    sumcenter = zeros(5,5);
    clustnum = zeros(1,5);
    for i=1:nr_projects % for our 14 projects
        for c=1:k % for our 5 clusters
            if (AC(1,i) == c)
                clustnum(1,c) = clustnum(1,c) + 1;
                for j=1:5 % for our 5 features
                    sumcenter(c,j) = projects(i,j) + sumcenter(c,j);
                end
            end
        end
    end
    for c=1:k % for each cluster of a project
        for j=1:5 % for each feature of a project
        centroid (c,j) = sumcenter (c,j) / clustnum(1,c);
        end
    end
```

```
   % calculating the objective function for the current clustering scheme
   sumdist_intra = zeros(1,k);
   for c=1:k
      for i=1:nr_projects
         if (AC(1,i) == c)
            for j=1:5 % for project features
               sumdist_intra(1,c) = sumdist_intra(1,c) + (projects(i,j)-
centroid(c,j)).^2;
            end
         end
      end
   end

   JC = sum (sumdist_intra);
%*****************************************************************
% setting the beginning program parameters
AB = AC;
JB = JC;
iter_num = 1;
TABU=AC; % insert the first element into the tabu list
TLL = 1;
%*****************************************************************
while iter_num <= ITMAX
   % generating NTS (number of trial solutions) by using AC
   for t=1:NTS % for each number of trial solutions
      clu1 = 0;
      clu2 = 0;
      clu3 = 0;
      clu4 = 0;
      clu5 = 0;  % for preventing empty cluster occurence
```

```
    while ((clu1==0)|(clu2==0)|(clu3==0)|(clu4==0)|(clu5==0)) % for every t if
an empty cluster is formed repeat this loop
        for i=1:nr_projects
            rand_num = random('Uniform',0,1); %for indicating the random
numbers generated for each project of NTS
            if (rand_num < P)
                AT(t,i) = AC(1,i);
            else
                AT(t,i) = unidrnd(k,1,1);
                while (AT(t,i) == AC(1,i))
                    AT(t,i) = unidrnd(k,1,1);
                end
            end
        end
        clu1 = 0;
        clu2 = 0;
        clu3 = 0;
        clu4 = 0;
        clu5 = 0;
        for i=1:nr_projects
            if (AT(t,i) == 1)
                clu1 = clu1 + 1;
            end
            if (AT(t,i) == 2)
                clu2 = clu2 + 1;
            end
            if (AT(t,i) == 3)
                clu3 = clu3 + 1;
            end
            if (AT(t,i) == 4)
                clu4 = clu4 + 1;
```

```
            end
            if (AT(t,i) == 5)
                clu5 = clu5 + 1;
            end
        end
    end
end
%**************************************************************
% calculating objective function values of NTS (number of trial solutions
% calculating the cluster centers of NTS's
for t=1:NTS
    sumcenter = zeros(5,5);
    clustnum = zeros(1,5);
    for i=1:nr_projects % for our 14 projetcs
        for c=1:k % for our 5 clusters
            if (AT(t,i) == c)
                clustnum(1,c) = clustnum(1,c) + 1;
                for j=1:5 % for our 5 features
                    sumcenter(c,j) = projects(i,j) + sumcenter(c,j);
                end
            end
        end
    end

    for c=1:k % for each cluster of a project
        for j=1:5 % for each feature of a project
            centroid (c,j) = sumcenter (c,j) / clustnum(1,c);
        end
    end
    % calculating the objective function for the current clustering scheme
    sumdist_intra = zeros(1,k);
```

```
      for c=1:k
         for i=1:nr_projects
            if (AT(t,i) == c)
               for j=1:5 % for project features
                  sumdist_intra(1,c) = sumdist_intra(1,c) + (projects(i,j)-
centroid(c,j)).^2;
               end
            end
         end
      end
      JT(1,t) = sum (sumdist_intra);
   end
%**************************************************************
   % sorting the objective functions and AT's of NTS's
   [JTSORT,IX] = sort(JT);
   if (JTSORT(1,1)<JB)
      trial=IX(1,1);
      for i=1:nr_projects % if a better solution is found then change the current
solution
         AC(1,i) = AT(trial,i);
         JC = JT(1,trial);
      end  % if the model ends here and do not enters the else statement it will
continue from 4th step
      nottabubest = 0; % for counting how many times turns the for-loop that tries
to find not tabu-best solution
   else
      nottabubest = 0; % for counting how many times turns the for-loop that tries
to find not tabu-best solution
      tabudegil = zeros(1,NTS);
      for t=1:NTS
         nottabubest = nottabubest + 1;
```

```
        ab = IX(1,t);
        for z=1:TLL % for tabu list length
          protabudegil = 0;
          for i=1:nr_projects
            if (AT(ab,i)~=TABU(z,i))    % find the trial solution that has the
smallest obj func value and that is not tabu
              protabudegil = protabudegil +1;
            end
          end
          if protabudegil ~= 0 % if at least one of the elements of At and TABU
are different
                              %then the 2 solutions are different
            tabudegil(1,t) = tabudegil(1,t) + 1;
          end
        end
        if (tabudegil(1,t) == TLL)
          for i=1:nr_projects
            AC(1,i) = AT(ab,i);
            JC = JT(ab);
          end
        end
        if (tabudegil(1,t) == TLL) % if the model ends here it will continue from
4th step
          break
        end
      end
    end

    if not(nottabubest == NTS) % if not all the solutions are tabu
      %if all the trial solutions are tabu then the following statements
```

```
        % will not be done and the model will continue from 2nd step without
increasing the iteration number
        % and without updating the tabu list, AC, JC
        TLL = TLL+1; % tabu list length is increased by one
        TABU(TLL,:)=AC; % insert the AC current solution at the bottom of the
tabu list
        if (TLL == (MTLS + 1)) % only increasing the plus value here from 1 to 2 is
enough for the adjustment of tabu list length
            TABU(1,:)=[]; % if tabu list length is equal to TLL then delete the first
element in the tabu list
            TLL = TLL - 1;
        end
        if JB>JC % if a solution that is better than the best one is found change the
best solution settings
            for i=1:nr_projects
                AB(1,i) = AC(1,i);
                JB = JC;
            end
        end
        iter_num = iter_num + 1
    end
end
% for writing the final values of AC,JC,AB,JB
AC
JC
AB
JB
clear
```