Chapter Tow
Literature Review

2.1 Criticism on Data Collection of Household Health Surveys

Household surveys performed by World Health Organisation (WHO) are often criticized for missing observations (Aalto 2000; Almeida et al. 2001; Williams 2000) and for the methodology of data collection (Blendon et al. 2001a; Navarro 2001; Almeida et al. 2001). Biasness information is also be found in such criticism is including used few data for many imputations, used limited number of questions out of a large number of questions for indexing, inadequacy of the sample to represent the population, inherent flaws in method, and majority of key informants being the people of WHO (Williams 2005).

2.1.1 Relative weight of responsiveness

The framework of health surveys is often criticised by the commentators from regional consultants (WHO Regional Office for the Eastern Mediterranean 2001) of having non-responsiveness, as it is relatively an important factor in household surveys.

2.1.2 Domain weights

In aggregating an overall responsiveness index, the relative weights of the seven domains were criticized by several commentators (WHO Regional Office for Africa 2001 WHO)

2.1.3 Broader Health System and the Responsiveness

Regional consultants (WHO Regional Office for South-East Asia 2001) have been commented the inability of survey responsiveness to represent the health system with its broader boundaries including the services of health promotion and protection like public access to information (Oswaldo Cruz Foundation 2000; Travassos 2001; Ugá et al. 2001).

2.1.4 Sources of information

Whether the users of a health system can better judge it or the key informants? , This question has been arised by critiques (Blendon et al. 2001a). Canvassing the information from population crucially requires measure of responsiveness as well as satisfaction (Blendon et al. 2001b).

2.1.5 Translation, validity and reliability

Translation of the concept of cross-culture validity and responsiveness is described in the question raised by Aalto (2000). Issues raised by several regional consultatants included the cross-cultural validation of the household health surveys (WHO Regional Office for the Western Pacific 2001). In comparison with the modules owe to the abstractness of involved concepts, translation responsiveness might be a slightly more difficult problem (Almeida et al. 200; WHO Regional Office for the Western Pacific 2001).

The responsiveness of key informant instrument has been criticized by critiques and participants in regional consultatants for the availability of standard instrument psychometric data (Aalto 2000; WHO Regional Office for South-East Asia 2001). Aalto (2000) and the regional consultatants WPRO and SEARO indicated that any subsequent responsiveness questionnaire instruments should capture this type of data (WHO Regional Office for the Western Pacific 2001).

2.1.6 Universality of domains

Extensive critiques on the choice of domain included the need to provide convincing rationale (Aalto 2000). WHO ensured the domain of survey question to caution cross-cultural validation commending the changes in household surveys?

Critiques and the participants of some regional consultatants still have been raised the issue of relevance of the domains in different cultural settings (Williams 2000; WHO Regional Office for Africa 2001).

2.1.7 Non-users

The indicators to measure the experience of standards used by WHO are being limited to those people actually use the health system and hence the ground for responsiveness has been criticized in a commentary of the Brazilian Ministry of Health (Oswaldo Cruz Foundation 2000).

2.2 Suggestions for Analysing Survey Data

Kish (1965) suggested that analysis of survey data is based on assumption if it ignores the weights and the sample design. The weights for estimating relationships between variables, rates, or means might be safely ignores if the sample design is capable of generating sample of equal probability. Kish (1965) called these designs *epsem* designs and stated that at near final or the final stage of the design, *epsem* can be designed even with complex multi-stage samples. Even with the initially *epsem* design, unequal weights can be created by the adjustments for non-response.

Linear statistics can be used to calculate unbiased linear estimation of population that can provide the inverse of probability of selection for each observational unit and design based weights for the theoretical case of surveys from all sample members with complete response (Horvitz and Thompson, 1952). The observations of non-response that are being dropped from the analysis without taking any other action lead to biase estimation of household surveys in practice. The biasness due to non-response is now been reducing using the continually developing techniques. Among the earliest techniques proposed, one simplest one was given by Rubin (1987) that reduces the difference between the parameters of population for non-respondents and respondents to a negligible value through partitioning the sample into weighting classes. Folsom and Singh (2000) evolved the calibration method of Deville and Särndal (1992) for weigh adjustment of post-stratification, for non-response, or for both. Calibration methods simultaneously control the weighted sample distribution in several dimensions.

The non-linear estimation is consistent in the trivial sense that is would be exactly equal the value of comparable finite population if the size of sample was increased to the finite size of population but the non-linear estimates are not unbiased for small samples (Cochran, 1977). The sample size can be allowed to increase without limit if it is allowed to consider the population of finite size as arising from a hypothetical population of infinite size. In this case, as the sample size increases, probability of the non-linear estimation converges to the parameter of super-population and thus the consistency of the model can be claimed (Skinner, Holt, and Smith, 1989).

2.3 Missing Data Treatment

Every real world study frequently face of complication of missing data. From chances to the design, numerous factors result in missing data. The situations leading to missing data often occur as some participants in a study try to protect confidentiality and purposely excise information; to provide values of some subjects may decline, and some variables may not be collected from all subjects. A potentially biased as well as inefficient method of complete case in which the subjects with missing observations are dropped, is frequently used despite of its disadvantages. Many researchers have been trying to find the efficient and appropriate method for analysing the data with missing values. A large number of values with missing data might lead by the a few missing data points in each covariate thus many real-world settings require the models that incorporate predictors observed partially.

Multiple imputations have been primarily focused in real-world settings in the comprehensive overviews provided by Allison (2002), Schafer (1997), and Little and Rubin (2002). Though the work is somewhat dated, maximum likelihood approach of Ibrahim (1990) was included in the hierarchy of approaches to deal with the missing predictors described by Little (1992). A general introduction of incompleteness on only one variable is provided by Raghunathan (2004) and Meng (2000) and through the application on cancer data set, Ibrahim et al. (2005) reviewed recent developments in a comprehensive fashion. A comprehensive reading list in form of annotated bibliography is provided online (Carpenter 2006a).

Here the literature review of the analysis of a data set with missing values is done through updating the prior review of Horton and Lipsitz (2001). The discussion will cover the compromises, approaches, assumptions of modelling within current implications. The general focus is on the methods that are used to resolve the issues in analysis that arise when some observations are missing from the data set and to deal with the complications arising in cluster analysis (Jansen et al. 2006; Robins, Rotnitzky, and Zhao 1995; Laird 1988).

2.3.1 “Ad-hoc” Methods

Missing data are found to be addressed by a series of “ad-hoc” methods. One approach for continuous involves creation of a new variable that indicates the missing data, recoding missing observations to some common value, and then including the interaction between these variables and the variables themselves in the model. Creation of new variables for missing data holds for categorical variable too. These ad-hoc approaches are not recommended due to the potential induction of biasness (Greenland and Finkle 1995; Jones 1996). Another approach is to drop those subjects from the analysis, which lack information for many variables. This approach is also not attractive as it often results in unnecessary large standard errors, consequent biasness, and exclusion of important variables. Two other non-recommended methods that have large variability and potential of inducing biasness are found in the work of Jansen et al. (2006), Cook, Zeng, and Yi (2004), and Carpenter et al. (2004). These methods impute missing values through using the last observed value (also, known as, last observation carried forward LOCF or last value carried forward LVCF) for longitudinal analysis and the average of observed values, that is mean imputation.

2.3.2 Multiple Imputations

Rubin (1976) describes the reasons for using three-step approach, multiple imputations, in estimation of models with incomplete data. First reason is the uncertainty about the non-response model reflected by the creation of plausible values for missing values. Missing observations are then imputed or filled out by these plausible values. A number of completed data set is created through this process repeatedly. Second reason is the availability of complete data methods for analysing the data sets. The last but not least reason is the handling of uncertainty regarding the imputation allowing by combined results.

A public survey data was the first setting to use the method of multiple imputations. Inclusion of detailed and confidential information in a model can be created as auxiliary information, which is unsuitable to include in the public data set. Hence, in survey data settings, multiple imputations remains ideally suited (Rubin, 1996). Each of that data sets can be analysed through the utilization of existing software provided the complete data sets. However, in a setting where a single person is the imputer and the analyst, multiple imputations is more commonly used (Allison 2000; Barnard and Meng 1999; van Buuren et al. 1999; Schafer 1999; Rubin 1996; Glynn et al. 1993; Rubin 1999).

The potential of bias arises from the misspecification of the model; hence, the suitable specification of the model of imputation is the key issue for an analyst. Estimation of multivariate mode only needs the variance-covariance matrix and mean vector therefore this computationally traceable model has been used very often. Biasness in result and complications in analysis often occur when some of the variables are not Gaussian, in such situation multiple imputation is used (Allison 2005; Horton, Lipsitz, and Parzen 2003). Complication in joint distribution due to missing values in multiple continuous and categorical variables is a salient reason for using multiple imputations. However, the model for analysis must not be richer than the one used for imputation (Little and Rubin 2002). Following is the description of a number of methods that were found in the literature reviewed.

In addition to the aforementioned imputation methods, which replace each missing value with one value, the multiple imputations (MI) by Rubin (1986) replaces each missing value with a set of plausible values that represent the uncertainty of the correct value. The key steps can be concluded as follows (Allison (2000)):

 first, impute missing values using an appropriate model that incorporates Random variation;

 Second, repeat this process for M times (e.g., 3 to 5 times), producing M “Complete” data sets;

 third, perform the desired analysis on each data set using standard complete data methods;

 fourth, average the values of the parameter estimation across the M samples to produce a single point estimation.

2.3.4 Conditional Gaussian

Schafer (1997) improved the Conditional Gaussian approach of imputation for both discrete and continuous missing values. Cases of continuous variables assume a multivariate normal distribution and cases of discrete variables assume a log-linear model (Bishop, Fienberg, and Holland 1975). In real world of multiple categorical variables, a proliferation of parameters can be lead by the fit of this general location model as saturated multinomial with shared covariance and separate means (Olkin and Tate 1961). This resulted in the need of simplified log-linear model in practice. S-Plus missing data library and Schafer's mix program (assuming a form of monotonicity) has been implementing this approach.

2.3.5 Chained Equations

Chain equations are used in an alternative variable-by-variable approach (van Buuren et al. 2006; Raghunathan et al. 2001; van Buuren et al. 1999). Other variables are involved as predictors in the separate specification of each variable in this imputation model. An imputation is generated for the missing variable at each stage of the algorithm then the next variable is imputed using the previous imputed value. The process reaches convergence at last after the repetition of the Gibbs sampling procedure to impute the missing values. Multiple imputations are generated using separate chains. Predictive matching (where the value from one of the nearest set of observed value in the data set is taken by the imputed variables) or a linear regression model is involved in the model for continuous variables. For categorical variables, polytomous models are needed and logistic regression can be fit for dichotomous variables. AregImpute (for R and S-Plus), IVEware (for SAS or standalone), ICE (for Stata), or MICE library (for R and S-Plus) can provide the implementations of the chained equation approach.

Raghunathan et al. (2001) describes the problem with the approach of chained equation approach as its inability to converge to a sensible stationary distribution where multivariate distributions and separate variables are not compatible though van Buuren et al. (2006) obtained reasonable imputations in a series of studies on simulation even with incompatible separate models. Further establishment of the validity of this approach needs additional work.

2.3.6 Methods for Monotone Data sets

SAS PROC MI handles data sets with monotone missing structure implementing a number of approaches. A value randomly from such a set of observed values whose values are closest to predicte value can be imputed using the method of predictive mean matching. Imputation of a categorical variable with more than two levels complicates this method while the method remains straightforward in imputation of a continuous random variable. The observations with different numbers of missing values are processed in this approach in an ascending order of number of missing values. In application to missing predictor models, biased results were found with the predictive mean matching approaches, which warned the analysts from using this approach (Allison 2000). Missing values can be imputed in the similar way using propensity score or regression models.

2.3.7 Issues with Imputation

Raghunathan et al. (2001) found accountable limiting of the imputed values (e.g. plausible value of years of smoking for non-smokers is only zero). Similarly problem of limiting arise when variables require transformations, or when there is a certain range defined for missing values (e.g. a five point Likert scale may have values between three and four). Calculating the standard errors of the maximum likelihood estimation is another complication. SPSS, the S-Plus missing data library, and LogXact version 7 address these complications and provide the implementations of maximum likelihood (von Hippel 2004).

2.3.8 Methods of Weighting

The approach of weighting methods used to account missing predictor data (Carpenter et al. 2006; Horton and Lipsitz 1999; Horton et al. 2001; Xie and Paik 1997; Robins et al. 1995). Complete cases use weights in this approach, which are actually the probabilities, obtained through fitting a model for the probability of amusingness. Software such as SAS, SUDAAN, SPSS, or Stata that allows for weights can be used to fit weighting approaches. Carpenter et al. (2006) and Ibrahim et al. (2005) detailed this approach for a single missing predictor but for multiple missing non-monotone variables, this approach becomes considerably less tractable.

The general formula for a sample design weight is arithmetically very simple, it is one divided by the probability of selection for the survey design. However, these are usually scaled, so we define the weight as proportional to this number. For example, if there are three adults in a given household the resulting sample design weight for the single interviewed adult will be proportional to (1/(1/3)), i.e. proportional to three. In a one adult household, the weight will be simple proportional to 1/1, i.e. proportional to one. In other words, the influence of the former respondent is being increased threefold relative to the influence of the latter respondent to exactly compensate for the fact the former respondent was three times less likely to be included in the sample.

2.3.9 Bayesian Approaches

Posterior distribution sampling of interest is involved in Bayesian framework. Bayesian methods have been more generally applied while multiple imputations were obtained within a Bayesian framework. The close relationship between MI and ML methods and the Bayesian approach estimates the covariates with a prior distribution as described Ibrahim et al. (2005). Estimation of relationships requires model with a package like WinBugs and specific coding of prior distributions partly due to the flexibility of these methods. Carpenter (2006b) and Carpenter, Pocock, and Lamm (2002) provide the models of missing data and relatively straightforward coding.

2.4 Cluster Analysis

2.4.1 What is a cluster?

A formal definition of cluster is hard to give despite of the easy visual recognition of clusters from a two-dimensional view. Many authors with the contribution in the literature of clustering address the lack of formal and universal definition of cluster. However, giving one definition is regarded as an intractable problem by the authors (Aldenderfer and Blashfield 1984; Cormack 1971; Everitt, Landau, and Leese 2001; Kaufman and Rousseeuw 1990; Tan, Steinbach, and Kumar 2005; Xu et al. 2005). The weakly defined notion of a cluster depends on the application (Tan, Steinbach, and Kumar 2005). The definition is also affected by the goal of cluster analysis. There are different sizes and shapes of clusters depending on the application. Moreover, for dependency on the resolution, one is looking in the data (global versus local); even the number of inherent clusters in the data is not unambiguous (Tan, Steinbach, and Kumar 2005; Jan and Dubes 1988).

Typically, strong internal similarities are possessing in data description in terms of clusters yielded from clustering methods (Duda and Hart 1973). External isolation (separation) and internal cohesion (homogeneity) are often used to define cluster. Hence the definition of cluster is a set of objects dissimilar to the objects in the other clusters but similar to the objects within the same cluster (Han and Kamber 2001). Following three definitions of cluster are found to be best described in the literature reviewed (Jain and Dubes 1988; Aldenderfer and Blashfield 1984). Regions of low density of points are separated from regions of relatively high density of points of a multidimensional space and such connected regions are called clusters. Cluster is such an aggregation of points where distance between the points included in a cluster is less than the distance between the points within the cluster and the points outside the cluster. Cluster forms set alike entities in a way that the entities across clusters are unlike.

Certain properties like separation, shape, dimension, variance, and density are used to compare the clusters even though the cluster is an application dependent concept (Aldenderfer and Blashfield 1984). In comparison with other areas of space, a region of compact and tight high-density points is called cluster. Small degree of variance or dispersion is meant by the tightness and compactness. The shape of cluster is a priori known, determined by the clustering criteria, and used algorithm. Distance between clusters and the degree of possible cluster overlap is defined through separation. Each point of the clusters are assigned a degree of membership in fuzzy clustering thus overlapping clusters are producing (Baraldi and Blonda 1999). Each data point is assigned to only one cluster and thus separated clusters are produced in the traditional partitioning clustering methods, such as hierarchical methods and K-Means. Variables define a cluster in its dimensions and radius is possible to be determined if the cluster has around shape. Assigning relations or universal values of these measureable features of any cluster is not possible. Perhaps size and shape are the most problematic features.

2.4.2 Main Elements of Cluster Analysis

The successful completion of tasks presumes a large number of correct choices and decisions from several alternatives despite of the simple idea behind cluster analysis. Before attaining the results, at least nine major elements appear in cluster analysis (Anderberg 1973). The list with strategy of missing data and data presentation, because the current real-world data set contains missing values as well, include Interpretation of results, Number of clusters, Computer and algorithms implementation (and their reliability, e.g., convergence), Choice of missing data strategy, Choice of clustering criterion (objective function), Choice of dissimilarity measures, Normalization of variables, What to cluster: variables or data units, Choice of variables, Choice of objects, and Data presentation (Karkka, Inen, and Yramo 2004; Little and Rubin 1987). According to Jain, Duin, and Mao (2000), importance of strategies used in cluster validity, normalization, data representation, and data collection is same as that of the cluster strategy itself. Importance of choice of the best (dis)similarity measure is greater than the importance of choice of clustering algorithms (Hastei, Tibshirani, and Friedman 2001). Result interpretation and estimation of the number of clusters are closely related to the validity of resulting cluster solution e.g. a kind of validation technique, visual exploration of the obtained solution (Jain and Dubes 1988; Aldenderfer and Blashfield 1984).

2.4.3 Missing Data in Cluster Analysis

In describing the two alternative approaches of handling missing values; marginalization where missing values are ignored and imputation where estimated values are used to fill in missing values; Green et al. (2001) did not consider imputation as a reliable approach in comparison with actually observed data. In evaluation of different imputation methods on biological data, Troyanskaya et al. (2001) clearly stated:

> “However, it is important to exercise caution when drawing critical biological conclusions from data that is partially imputed. [. . .] [E]estimated data should be flagged where possible [. . .] to avoid drawing unwarranted conclusions.”

There is no mechanism to indicate the less reliability of imputed values still data imputation is common despite of this warning. Troyanskaya et al. (2001) described other sophisticated approaches to handle missing values including inferring the missing values’ feature based on observed features and similarities between the missing observation and known observation in the data set; modelling the selected and observed values according to the true distribution; and replacing the missing values with the actual mean value. Ghahramani and Jordan (1994) processed data with missing value and presented modification in EM algorithm. Values for the missing features, data cluster assignments, and maximum likelihood parameters are simultaneously estimated in this model. Due to the lack of full reliability, each of these approaches suffers from a disability of discounting imputed values.

Marginalisation is also sometimes considered as a better solution because no new data values are created. Supervised methods such as Hidden Markov Models (Vizinho et al., 1999) or neural (Tresp et al., 1995) networks are focused in most of the previous work in marginalisation. Wagstaff et al., (2001) proposed a set of hard constraints that were guaranteed satisfied by the output produced by a variant of k-means. Hard constraints show whether the group of certain items should or should not be formed while soft constraints shows the strength of grouping. Both the soft and hard constraints satisfy in the clustering of data with missing values.

Exploratory data analysis often includes cluster analysis. Covariance, sample mean, and other classical second order statistics are insufficient to describe the internal structure of complex data sets that almost every exploratory study does. MacQueen (1967) stated that instead of only a computational process, that identifies definitive and unique grouping for the data, investigators obtain the quantitative and qualitative understanding of a large size multivariate data set through clustering applications. Later the clustering application became a core method of knowledge discovery and data mining due to the summarizing, descriptive, and unsupervised nature of data clustering. New clustering algorithms developed due to the increasing number of large multidimensional data collections especially during the last decade (Han and Kamber 2001; Hand, Mannila, and Smyth 2001; Tan, Steinbach, and Kumar 2005).

Task of clustering was performed manually until the use of computers become common. Since different individuals see things in different ways, "human clustering" is not likely an inconsistent procedure though perceiving groups visually from a two or three-dimensional data set is easy. Between different individuals, the direction and level one is looking at the data or the measure of similarity are not consistent. For example, classification of people can be done forming a number of groups according to annual alcohol consumption or the economical status etc. The same individual will not be necessarily captured by this grouping (Everitt, Landau, and Leese 2001). The user's background (culture, profession, education, position etc.) impact the direction in which he or she is looking in the data set (Jain and Dubes 1988). The literature proposes various definitions of cluster analysis. Different aspects of the methodology are emphasized by the different ways in which the definitions slightly differ. When nothing is known about the category structure, the task of identifying natural groups in a data set is done in cluster analysis (Anderberg 1973). Division of a set of elements based on their similarity over a number of variables into a small number of relative homogenous groups is the main target of cluster analysis where similarity across all objects provides the basis for groupings of variables (Bailey 1975). Hence grouping of variables or objects, or even both can be done through cluster analysis (Everitt, Landau, and Leese 2001). As one representative and summarising variable can be substitute in place of correlating variables, reducing the number of variables is another aspect of cluster analysis. According to Jain et al. (1999), cluster analysis can be defined from the perspective of statistical pattern recognition as the organisation of collection of patterns based on similarity into clusters. The patterns are usually represented as a point in multidimensional space or as a vector of measurements. From the perspective of statistics, Hastie et al. (2001) describes the task of cluster analysis as making dissimilarities within the groups in which observations are partitioned smaller than the dissimilarities across groups. From the point of view of data mining, useful, meaningful, or both types of groups of data are formed as the result of cluster analysis (Tan et al. 2005). Initial settings are served in useful clusters for some methods such as regression methods or principle component analysis (PCA) where summarising the data set beforehand is useful while meaningful cluster in a data set captures the natural structure.

One of the key assumptions in cluster analysis is the unknown structure of target data set as emphasized in the first definition. This assumption of clustering (unsupervised classification) majority differ it from classification (supervised classification). The object collections with unknown class labels are focused in cluster analysis unlike classification where a priori knowledge of category structures is available. Process of cluster formation is unaffected of information about the data sources such as class labels which influence the interpretation of results (Jain et al. 1999). During the configuration of correct number of clusters or initial parameters, the understanding of domain is often of great use. Multidimensionality of the data objects (records, observations etc.) is stressed in the second and third definitions. The difficulty that a human being faces in grouping of objects that possess three or more variables without automated methods emphasizes the importance of previous notion. The notion of similarity is addressed in most of the aforementioned definitions naturally. Choosing an appropriate measure of similarity is one of the most influential tasks of cluster analysis, as similarity is one of the key issues of cluster analysis. Problem of selecting a measure of similarity depends on the data. Anderberg (1973) used the degree of "natural association" instead of talking about "similarity". Cluster analysis, based on aforementioned definitions, can be described as analysing a multidimensional data set with an unknown structure and choosing a measure of similarity to determine a (small) number of meaningful variables or objects. Here, meaningful refer to the description given by Tan et al. (2005).

However, up to three-dimensional data cluster, method of visual perception is suitable; computers become indispensible as visual perception turns into a complex task in a space of three or more dimensions. Even for the same data, different algorithms due to the inconsistency of a human classifier form different groupings. Therefore, there is no universally best algorithm for clustering (Aldenderfer and Blashfield 1984, Jain, Duin, and Mao, 2000). The best understanding of data set can be obtained when several cluster algorithms are tried (Jain et al. 1999). A major part of applications can be covered using a set of six clustering algorithms (MONA, DIANA, AGNES, FANNY, CLARA, and PAM) proposed by Kaufman and Rousseeuw (1990). MONA is a divisive algorithm that uses a single variable to carry out the separation of objects into groups. DIANA is a continuous process of splitting each cluster into two groups after putting all the objects into one cluster. AGNES is an agglomerative method of clustering that uses successive fusion of clusters to produce a tree-like cluster hierarchy put in inverse order with respect of DIANA. The fuzzy clustering method FANNY gives the clusters a degree of membership for all the objects. CALRA reduced the computational cost by sub-sampling the data set and uses medoid points to partition the data set. PAM produces a given number of disjoint clusters and divides the data set using a partition-based K-medoid method. The internal structure of any data set can be overviewed using the above-mentioned set of methods. One may utilize different visualisation techniques such as multidimensional scaling, MDS and PCA as interpretation of result is a human process (Hand, Mannila, and Smyth 2001). Information related to any problem like priori domain knowledge is integrated to the clusters after the interpretation.

Contributions made by engineers (Jain and Dubes 1988), social scientists (Bailey 1975), statisticians (Fraley and Raftery 2002), biologists (Jiang, Tang, and Zhang 2004; Tseng and Wong 2005) and psychologists (Milligan and Cooper 1985) show that the development of clustering methods is interdisciplinary. Typological analysis, botryology, automatic classification, numerical taxonomy and various other names for cluster analysis have emerged naturally (Kaufman and Rousseeuw 1990). Also data partition (Roberts, Everson, and Rezek 1999), data segmentation (Hand, Mannila, and Smyth 2001), and unsupervised classification (Jain, Murty, and Flynn 1999; Tan, Steinbach, and Kumar 2005) are used as synonyms for data clustering. Development of cluster methods also contributed in a great amount by knowledge discovery and data mining after it constituted its own separate scientific discipline and have grown further off the other original fields. For large data sets, computational efficient algorithms has been specially focused (Han and Kamber 2001; Hand, Mannila, and Smyth 2001; Tan, Steinbach, and Kumar 2005; Dunham 2003). On different disciplines, the same methods are often invented with different names perhaps due to the interdisciplinary nature of cluster analysis. To mention just a few, decision sciences, policy, information, engineering, earth sciences, social and behavioural sciences, medical sciences, life sciences, biological sciences and many other fields dealing with a huge amount of cluster applications (Anderberg 1973; Jain and Dubes 1988; Kaufman and Rousseeuw 1990; Jain, Murty, and Flynn 1999; Everitt, Landau, and Leese 2001; Xu et al. 2005). The importance of data clustering as a key technique of statistics (Dillon and Goldstein 1984; Hastie, Tibshirani, and Friedman 2001), pattern recognition (Duda and Hart 1973; Duda, Hart, and Stork 2001; Fukunaga 1972; Jain, Duin and Mao 2000; Tou and Gonzalez 1974), and knowledge discovery and data mining (Tan , Steinbach, and Kumar 2005; Hand, Mannila and Smyth 2001; Han and Kamber 2001; Grabmeier and Rudolph 2002; Ghosh 2003; Berkhin 2002) is thus emphasized. A very wide range of clustering applications include P2P-networks (Ramaswamy, Gedik and Liu 2005), identification of international conflicts (peace science application) Wolfsom, Madjd-Sadjadi, an dJames 2004), archeological applications 3, identification of subtypes of schizophrenia (Helmes and Landmark 2003) , optimal placement of radioports in cellular networks (Abolhassani, Salt, and Dodds 2004), classification of unknown radar emitters from received radar pulse samples (Liu et al. 2005), grouping customers of similar behaviour in marketing research (Berry and Linoff 2000), software modules and procedures (Maitra 2001; Zhong, Khoshgoftaar, and Seliya 2004) etc and some quite exotic examples form an almost endless list.