December 2002

The Use of Fuzzy Clustering to Examine End User Segmentation

Vikram Sethi
University of Texas at Arlington,

Chittibabu Govindarajulu
Drexel University

Ruth King
University of Illinois at Urbana-Champaign

Mark Eakin
University of Texas at Arlington

THE USE OF FUZZY CLUSTERING TO EXAMINE END USER SEGMENTATION

Vikram Sethi
College of Business Administration
University of Texas at Arlington
Arlington, TX USA
vikram@world.std.com

Chittibabu Govindarajulu
LeBow College of Business
Drexel University
Philadelphia, PA USA
chitti@drexel.edu

Ruth C. King
College of Business
University of Illinois at Urbana-Champaign
Champaign, IL USA
rc-king@cba.uiuc.edu

Mark Eakin
College of Business Administration
University of Texas at Arlington
Arlington, TX USA
eakin@uta.edu

Abstract

A key research theme in the field of end-user computing (EUC) is learning more about end users and their needs and designing support strategies for assisting, managing, and controlling end-user activities. Typology maps, such as those by Rockart and Flannery (1983), have been used to categorize end users into different groups based on criteria such as the skill and sophistication of EUC activity. In most such studies, users self-select themselves into one of several groups based on generic definitions provided in the study. As in most pure taxonomies, a user becomes a part of one and only one group. In this study,we provide an alternative analytical mechanism to self-selection and unitary membership: the use of fuzzy clustering, which allows for gradual membership in different groups, with membership values indicating the probability or degree of membership within a specific group or cluster. Cotterman and Kumar's (1989) classification scheme is operationalized while allowing for overlapping membership probabilities into one of three clusters: user-developer-controller (UDC), user-developer (UD), and user (U). Further, the study also examines the proposition that end users vary in their use of available support sources and also in the type of support that they require. The expectation that different categories of users have different support needs is examined at three levels: cluster level, individual level, and cluster-conjunctive level.

1 INTRODUCTION

A key research issue in the field of end-user computing (EUC) concerns learning more about those individuals who develop and use their own software applications. As a part of increasing this understanding, various authors have tried to profile end users based on characteristics such as their background, source of their computer education, and their experience (Rivard and Huff 1985). Other studies develop typologies of end users and classify end users into different categories such as "nonprogramming users" and "IS programmers" (Rockart and Flannery 1983).

Developing an understanding of end users, whether through descriptive studies or by using a classification schema, is important to knowing who they are and what type of work they do, and therefore to proactively designing EUC strategies to better fit the profile of end users. Types of end users have been shown to differ significantly with respect to system usage and dependency (Schiffman et al. 1992). A relationship between the types of end users and the nature of support they require has also been noted by Alavi et al. (1988).

In most studies, end users are classified into *one* of several categories (unitary membership) by describing different categories and asking respondents to indicate which one category best describes them (self-selection) (Barker and Wright 1997; Schiffman et al. 1982). One potential problem with self-selection is the lack of a uniform measurement standard across all respondents. An alternative to the self-selection mechanism is operationalization of the specific dimensions of a typology and using procedures such as cluster analysis to classify end users. However, data-analytic approaches (i.e., cluster analysis) that are applied to understanding user populations generally do not reflect the conditions where end users may be performing a variety of tasks (e.g., data access and application development) but to differing degrees. This has been demonstrated well in the field of marketing when examining consumer-oriented market structures (Wedel and Steenkamp 1991).

In the current study, we present an alternative approach to self-selection and cluster analysis: fuzzy clustering. We adopt the classification dimensions described by Cotterman and Kumar (1989)—operation, development, and control—and describe a procedure whereby end users may be shown to have differing memberships in each of the fuzzy clusters developed based on the three dimensions. Through the use of fuzzy clustering, membership (or non-membership) of a subject in one (non-overlapping) or multiple (overlapping) clusters is replaced by gradual membership, indicating the probability or the degree of membership of an end-user in a cluster. We propose that this classification method more accurately reflects a situation where a user may be involved in each of the three end-user activities but to varying degrees.

Another stream of research in the EUC field is the examination of the support needs of end users. A core set of studies in this domain is targeted at understanding specific areas in which end users need support. (Carr and Rainer 1990; Mirani and King 1994). These studies have led to the speculation that the nature of support needed within an organization should vary based on the end-user profile. For example, Rockart and Flannery (1983) concluded in their study that there should be "strongly, differentiated education, training, and support for the quite different classes of users." However, there has been no empirical examination of this proposition.

Our study is, thus, motivated by the need to address the above gaps in end-user research and to better understand end-user segmentation. We adopt the classification dimensions proposed by Cotterman and Kumar and examine whether their three dimensions can be used to develop meaningful end-user profiles. In addition, we examine whether increased end-user activity is associated with higher support needs. The fuzzy cluster methodology allows us to examine differences across end users at three different levels: individual level, cluster level, and cluster-conjunctive level. Finally, we also extend previous studies related to end-user support by including four different support sources in our analysis.

We first review literature related to end-user taxonomies and end-user support needs. Next, we describe the analytic procedures used for fuzzy clustering. We then apply this procedure to data collected from 196 end users and examine specific hypotheses related to support needs of end users and their use of alternative sources of support.

2 LITERATURE REVIEW

2.1 End-User Taxonomies

Brancheau and Brown (1993) and Powell and Moore (2002) reviewed the growth of research in end-user computing and note that a number of early studies focused on providing descriptive data on the background of end users and proposing end-user typologies. These include reports of differences in user computer education (Pyburn 1983) and experience (Rivard and Huff 1985). In one of the first classification schemes, McLean (1979) divided users into three groups: data processing (DP) professionals, DP amateurs, and non-DP trained users. Rockart and Flannery (1983) conducted several field studies and categorized individuals from "nonprogramming users," who only access computers with software provided by others, to "IS programmers," who develop sophisticated applications. Schiffman et al. (1992) confirmed that Rockart and Flannery's end-user types are empirically valid although they asked respondents to self-select themselves into one of the end-user types. Other studies have also relied on self-selection based on definitions provided to respondents (Brancheau et al. 1985; Mirani and King 1994).

The "user cube" typology of end users was developed by Cotterman and Kumar (1989). The user cube classifies users based on the dimensions of operation, development, and control. In this framework, *operation* is the initiation and termination of system operations, monitoring, or use of software, and the execution of manual functions necessary for the operation of an information system. *Development* is the performance of any or all tasks of the system development process. Finally, the dimension of *control* is the decision-making authority to acquire, deploy, and use the resources needed to develop and operate the computer-based information system.

2.2 EUC Support Areas and Sources of Support

Different areas in which end users require support have been well studied in the literature by either reviewing services that end users require or by surveying services that are offered by various support sources such as information centers and help desks.

In one of the first studies that surveyed end users to determine support needs, Mirani and King created a nine-factor instrument of which eight relate to specific support needs. Other studies have surveyed information centers and help desks to determine services provided to end users. Wetherbe and Leitheiser (1985) interviewed 25 information center managers and determined that most offered services such as office automation support, electronic mail support, user group support, departmental needs assessment, data access coordination, user-DP liaison, and software assistance.

In summary, studies that have examined end-user needs have either focused on the information center as the provider of services or have ignored who provides such services, concentrating instead on determining specific support needs. Recently, however, Govindarajalu and Reithel (1998) expanded the list of support sources to include local MIS staff, informal support, and external support. In this context, *local MIS staff* exclusively support end users of a specific department and often report to the functional department manager. *Informal support* consists of technical help or advice from peers, friends, and lead users. *External support* includes help from vendors, technical Web-sites, and other online sources. Based on a survey of 80 end users, this study classified support requirements into software, hardware, data, functional, purchase, and training. More importantly, they found that support from peers/colleagues was used extensively, followed closely by support from information centers, local MIS staff, and external vendors.

An overlap of end-user typologies with support needs has also emerged in a few studies. Mirani and King tested the nomological validity of their end-user measure by testing whether all types of end users (based on Rockart and Flannery's categories) perceived similar needs for support. As mentioned previously, users self-selected themselves into one of several categories. An ANOVA indicated that the perceived support needs of different end-user types were not the same. The extent of support needed increased from the nonprogramming user to command-level users, from command-level users to end-user programmers, and from end-user programmers to functional support personnel. This led the researchers to support Rockart and Flannery's conclusion that "diversity among end users calls for strongly differentiated education, training, and support for the quite different class of users." (1983, p. 778).

In summary, existing studies related to end-user classification methods have relied mostly on unitary self-selection based on definitions provided to respondents. In addition, the primary focus of these studies is support provided by information centers even though there is increasing evidence that support sources have become more varied. Further, while there is speculation that increasing end-user activity is associated with higher support needs, there is a lack of empirical support for it. In the current study, we examine whether an alternative classification scheme can be reasonably used for the purpose of end-user classification. Multiple sources of end-user support are studied and the hypothesis that higher end-user activity is associated with higher support needs is examined for multiple support sources.

3 OVERALL METHOD SELECTION

In this study, we apply fuzzy cluster analysis to assess the degree of membership of each user in the hypothesized clusters proposed by Cotterman and Kumar (1989). A brief overview of the fuzzy cluster procedure appears below.

The problem of cluster analysis is well known; its goal is to find the best partition of n entities into c classes. Generally, there are two techniques for cluster analysis: hard or crisp clustering and fuzzy clustering. In hard or crisp clusters, each entity belongs to only one cluster, and thus the membership functions are zero-one vectors (Duda and Hart 1973). In fuzzy clusters, however, the condition of exclusive belongingness for entities is relaxed, and the membership becomes fuzzy expressing the degree of membership of an entity to a cluster. Cluster prototypes are usually defined as weighted averages of the corresponding entities. The fuzzy c partition is defined in such a way that the membership of an entity to a cluster expresses a part of the cluster's prototype reflected in the entity. Thus, an entity may bear 60 percent of a prototype A and 40 percent of a prototype B, which simultaneously expresses the entity's membership to the respective clusters. The prototypes are considered as offered by the knowledge domain. In this case, the prototypes are the three end-user functions: operations, development, and control.

3.1 The Fuzzy c-Means Algorithm

The fuzzy *c*-means (FCM) algorithm (Bezdek 1981), one of the most widely used methods in fuzzy clustering, is summarized below.

Let X = $\{x_1,...,x_n\}$ be a set of given data, where each data point x_k ($k=1,...,n$) is a vector in $\Re^p$, U_{cn} be a set of real $c \times n$ matrices, and c be an integer, $2 \leq c < n$. Then, the fuzzy *c*-partition space for X is the set:

$$M_{fcn} = \{U \in U_{cn} : u_{ik} \in [0,1]$$

$$\sum_{i=1}^{c} u_{ik} = 1, 0 < \sum_{k=1}^{n} u_{ik} < n \} \tag{1}$$

where u_{ik} is the membership value of x_k in cluster $i(i=1,...,c)$. The aim of the FCM algorithm is to find an optimal fuzzy c-partition and corresponding prototypes minimizing the objective function:

$$J_m(U,V;X) = \sum_{k=1}^{m} \sum_{i=1}^{c} (u_{ik})^m \|x_k - v_i\|^2. \tag{2}$$

In (2), V = $(v_1, v_2, ..., v_c)$ is a matrix of unknown centers (prototypes) $v_i \in \Re^p$, $\|\cdot\|$ is the Euclidean norm, and the weighting exponent m in $[1,\propto]$ is a constant that influences the membership value. To minimize criterion J_m, under the fuzzy constraints given in (1), the FCM algorithm is defined as an alternating minimization algorithm. First, values for c, m, and ε, a small positive constant, are chosen; then, a fuzzy *c*-partition U^0 is randomly generated and iteration number t is set to 0. In the second step, the membership values $u_{ik}^{(t)}$ are used to calculate the cluster centers $v^{(t)}(i = 1,\ldots,c)$ as follows:

$$v_i^{(t)} = \frac{\sum_{k=1}^{n} \left(u_{ik}^{t}\right)^m x_k}{\sum_{k=1}^{n} \left(u_{ik}^{(t)}\right)^m}. \tag{3}$$

Given the new cluster center $v_i^{(t)}$, the membership values $u_{ik}^{(t)}$ are updated:

$$u_{ik}^{(t+1)} = \left[\sum_{j=1}^{c} \left(\frac{\left\|x_k - v_i^{(t)}\right\|^2}{\left\|x_k - v_j^{(t)}\right\|^2} \right)^{\frac{2}{m-1}} \right]^{-1}. \tag{4}$$

The process stops when $\left|U^{(t+1)} - U^{(t)}\right| \leq \varepsilon$, or a predefined number of iterations is reached. U^0 may be randomly generated, or a hard cluster solution may be used to initialize membership.

One of the problems with fuzzy clusters, which also occurs in hard clustering, is that of determining the number of clusters. Several cluster validity functions have been developed for this interpretation, for example, the compactness and separation validity function:

$$S(c) = \sum_{k=1}^{n} \sum_{j=1}^{c} (u_{ik})^m \left(\left\| x_k - v_i \right\|^2 - \left\| v_i - \bar{x} \right\|^2 \right)$$

The smaller the value of S(c), the better the compactness and separation between the clustering groups of in-cluster samples. Therefore, one goal of the clustering algorithm is to minimize the value of S(c).

3.2 Data Collection, Measurement, and Procedures

Data was collected from 196 end users through a questionnaire survey sent to a random sample of 250 employees in a large Midwestern city in the United States. The original sample was selected from the alumni database at the business school of one of the researchers

3.2.1 End-User Computing Dimensions

The empirical use of Cotterman and Kumar's (1989) taxonomy is based on Govindarajulu (2002). The operations dimension was measured by an average of the extent to which a respondent used different types of applications—level-1, level-2, and level-3 applications. Development was measured by a four-item scale, which assesses the extent to which end users are involved in development tasks, for example, specification of end-user requirements, design of applications, actual programming of applications, and systems implementation. Control was measured by a five-item scale which assesses the authority of the end user to acquire hardware, software, initiate, manage, and implement end-user applications, collect, store, and use data for end-user applications, and assign, acquire, and staff development of end-user applications.

3.2.2 EUC Support Needs and Sources of Support

Based on Govindarajulu and Reithel (1998), 36 separate support needs were identified with four sources on which end users rely for support: information centers, local MIS staff, and informal and external sources of support. Respondents were asked to report the extent to which they relied on each support source for each of the 36 support needs.

4 RESULTS

The four development items loaded to a single factor with an alpha value of 0.92. The five control items also loaded to a single factor with an alpha value of 0.91. Scores on operation, development, and control were standardized.

U^0 was initialized with results from a hard cluster analysis. An SAS IML algorithm (although other algorithms may be used) was written to implement the alternating optimization algorithm described earlier. Data was clustered into 2, 3, 4, and 5 clusters using a fuzzy exponent value of 1.3. The algorithm converged after 20, 21, 25, and 30 iterations respectively with an ε value of 0.0001. The values of several validity functionals—partition coefficient, non-fuzzy index, and the fuzzy performance index—are calculated for 2, 3, 4, and 5 cluster solutions. There is a consistent indication from all validity functions of a three-cluster solution.

The cluster centers for the three clusters are shown in Table 1. Scores of operation, development, and control are high in cluster 2, because of which it is interpreted as the user-developer-controller (UDC) cluster. Due to the high score of operation in cluster 3 as compared to development and control, it is labeled as the user (U) cluster. Cluster 1 has relatively high scores for both the operation and development categories. It is interpreted as the user-developer (UD) group.

In addition to the cluster centers, cluster membership is calculated for each observation. Thus respondent 1 can best be represented as falling 46.7 percent in the UDC category, 35.9 percent in the UDC category, and 17.4 percent in the U category. However, a crisp cluster solution may be imposed on the data by classifying each observation into one cluster. For example, given the data above, respondent 1 would be classified into the cluster 1 (membership in clusters 1 changed to 1) and membership reduced to zero in other clusters.

Table 1. Fuzzy Cluster Centers

	User	Developer	Controller
Cluster 1	4.2	3.9	3.0
Cluster 2	5.3	6.0	5.7
Cluster 3	3.4	1.1	1.3

In order to obtain measures of end-user support needs from various sources, an average value of the support required over all four sources was calculated. These support needs were subjected to a factor analysis which resulted in three factors interpreted as training needs (13 items), data support (eight items), and general support (six items). For each factor, four scales were computed, one each for the four support sources. Thus, for the training needs factor, four scales were computed: support required from the information center (TDIC), from local MIS (TDMIS), from information sources (TDINF), and from external sources (TDEXT). Similarly for the data factor, four scales (DIC, DMIS, DINF, and DEXT) were computed and four scales (GSIC, GSMIS, GSINF, and GSEXT) were calculated for the general support factor.

5 EXAMINING THE VARIATION IN SUPPORT NEEDS FROM DIFFERENT SUPPORT SOURCES

In general, our expectation is that users with a high level of EUC activity will demonstrate higher support needs from all of the available support sources. In classical cluster analysis, this expectation can be tested by means of a formal hypothesis that compares the level of support for each user need across different clusters. In the case of fuzzy clusters, however, more detailed information is available that allows us to form three levels of assessment.

The first level of analysis is at the *cluster level*. As in the classical case of crisp cluster analysis, the above expectation may be stated at the cluster level. In this case, fuzzy cluster membership can be converted to crisp cluster membership and the formal hypothesis is as follows:

> **HA_1: The extent of reliance on support sources is higher for the UDC group as compared to the UD and the U groups; and**
>
> **HB_1: The extent of reliance on support sources is higher for the UD group as compared to the U group.**

The above hypothesis was tested by imposing a crisp structure on the fuzzy cluster memberships and performing ANOVA analyses on the support scales. The results of the ANOVA are shown in Table 2. The first part of the table shows the average values of training, data, and general support from all four sources for the three clusters. Scheffe's tests show that, in general, the UDC group relies more heavily on all support sources. Next, the average value of support (training, data, and general) required from each source (IC, local MIS, informal, and external) is calculated for each cluster. These are shown in the middle part of the table. Once again, we find that the UDC group relies more heavily on support from all sources, as compared with the UD and U groups. The last section of the table shows the averages for all support need from all support sources across clusters. These also show significant differences and are higher for the UDC group.

The second level of analysis is the *individual level*. Since membership data is available at the individual level, the expectation of higher support needs may be stated as follows:

> **HA_1: Increasing membership in the user category is associated with increasing use of support sources.**
>
> **HB_1: Increasing membership in the user-developer category is associated with an increasing use of support sources.**
>
> **HC_1: Increasing membership in the user-developer-controller category is associated with an increasing use of support sources.**

Table 2. Average Support Need Across User Type

	User	User-Developer	User-Developer-Controller	F	P
Training Support	3.04	3.49	3.84	7.08	0.00
Data Support	3.53	3.87	3.97	2.26	0.10
General Support	3.39	3.85	4.21	8.21	0.00
Training–IC	2.89	3.46	3.71	3.92	0.02
Training–Local	3.78	4.16	4.40	2.27	0.10
Training–Informal	2.85	3.07	3.74	6.03	0.00
Training–External	2.42	2.73	3.58	7.60	0.00
Data–IC	3.46	4.24	4.08	3.39	0.03
Data–Local	4.11	4.68	4.64	1.98	0.14
Data–Informal	3.21	3.27	3.59	1.10	0.33
Data–External	2.66	2.48	3.15	1.99	0.14
General–IC	3.10	3.81	4.19	6.89	0.00
General–Local	4.17	4.71	4.64	1.64	0.19
General–Informal	3.16	3.38	3.91	3.80	0.02
General–External	2.53	2.80	3.53	5.08	0.00
Total IC	9.55	11.52	119.96	4.55	0.01
Total Local	12.19	13.57	13.65	1.79	0.18
Total Informal	9.21	9.69	11.19	3.65	0.02
Total External	7.43	7.74	10.28	5.80	0.00

Table 3 shows the correlation matrix of fuzzy memberships with support needs. The table includes overall support needs for training, data, and general support. Next, each training need is separated by its source, and finally, total support from each of the sources is shown. The correlations are positive and highly significant in the case of the user and user-developer-controller categories. Correlations greater than 0.18 are significant at the 0.01 level.

The third level of analysis is termed the *cluster-conjunctive* level. It may be desirable to examine whether the relationship between the use of support services and membership in the three clusters is linear. In this situation, membership values are not independent (since they sum to 1). However, by applying the basic principles of mixture design (Cornell 1979), we can obtain a measure of the influence of each membership value singly or in combination. Since this analysis occurs at the level of cluster combination, it is termed the *cluster-conjunctive* level.

We assume that the measured response, use of support sources, is dependent only on the relative proportions of membership in the different clusters. The difference between the cluster-conjunctive analysis and the independent-variable analysis is the nonindependence of membership values. Applying the constraints of nonindependence to a standard polynomial yields the four canonical forms: linear, quadratic, full cubic, and special cubic. The terms in the canonical polynomial models have simple interpretations: the parameter β_i represents the expected response to the pure mixture $x_i = 1, x_j = 0, j \neq i$. The portion $\sum_{i=1}^{q} \beta_i x_i$ of each of the models is called the linear blending portion, and if the blending of the components is strictly additive, then the linear model is an appropriate representation of the surface. When there is curvature in the mixture surface due to nonlinear blending between pairs of components, the parameter β_{ij} represents deviations of the surface. Higher order terms (such as $\beta_{123}x_1x_2x_3$) describe additional perturbations of the response surface beyond those described by first- and second-order terms.

A sequential model fitting strategy is adopted to determine which of the above is the most complete model. The process relies on experimental error variance being available from replicated observations. First a linear model is fitted and tested for lack of fit. If the lack of fit is significant, a quadratic model is next tested. If the lack of fit is still significant, we attempt to fit the special cubic model, and then the full cubic model. We applied this strategy using fuzzy membership data and the extent of use of support sources. In each case, the linear model (for all support needs) shows a nonsignificant lack of fit, implying there is no reason to doubt a linear relationship between membership values. The results are shown in Table 4.

Table 3. Correlation Matrix of Membership With Support Needs

	User	User-Developer	User-Developer-Controller
Training Support	0.28*	0.01	0.25*
Data Support	0.16*	0.05	0.10
General Support	0.30*	0.03	0.25*
Training–IC	0.19*	0.00	0.19*
Training–Local	0.22*	0.02	0.17
Training–Informal	0.23*	0.04	0.25*
Training–External	0.27*	0.11	0.33*
Data–IC	0.21*	0.10	0.10
Data–Local	0.18*	0.07	0.10
Data–Informal	0.06	0.03	0.08
Data–External	0.07	0.13	0.17
General–IC	0.27*	0.01	0.24*
General–Local	0.18*	0.08	0.07
General–Informal	0.20*	0.01	0.19*
General–External	0.25*	0.05	0.26
Total IC	0.23*	0.03	0.18*
Total Local	0.19*	0.06	0.11
Total Informal	0.17	0.03	0.18*
Total External	0.21*	0.14	0.30*

*Correlations are significant at the 0.01 level

Table 4. Results for the Sequential Model Fitting Exercise

Model Form	ANOVA Summary Statistics				F(*p*)
	Source	df	SS	MS	
Training Needs—IC					
Linear	Regression	2	27.07	13.5	
$R^2 = 0.04$	Residual	187	650.6		
	Lack of fit	177	628.5	3.55	1.60 (0.20)
	Pure error	10	22.1		
Training Needs—Local					
Linear	Regression	2	37.2	18.6	
$R^2 = 0.04$	Residual	187	847.02		
	Lack of fit	177	817.20	4.61	1.55 (0.22)
	Pure error	10	29.81		
Training Needs—Informal					
Linear	Regression	2	39.6	19.81	
$R^2 = 0.04$	Residual	187	567.2		
	Lack of fit	177	527.3	3.33	2.97 (0.78)
	Pure error	10	39.89		
Training Needs—External					
Linear	Regression	2	50.3	3.52	
$R^2 = 0.04$	Residual	187	659.7		
	Lack of fit	177	603.83	3.41	0.61 (0.93)
	Pure error	10	55.93		

Model Form	ANOVA Summary Statistics				F(*p*)
	Source	df	SS	MS	
Data Needs—IC					
Linear	Regression	2	6.2	3.1	
$R^2 = 0.04$	Residual	187	728.1		
	Lack of fit	177	691.50	3.90	1.07 (0.50)
	Pure error	10	36.63		
Data Needs—Local					
Linear	Regression	2	39.19	19.58	
$R^2 = 0.04$	Residual	187	938.31		
	Lack of fit	177	896.92	5.06	1.22 (0.38)
	Pure error	10	41.38		
Data Needs—Informal					
Linear	Regression	2	7.21	3.60	
$R^2 = 0.04$	Residual	187	610.29		
	Lack of fit	177	563.65	3.38	0.68 (0.84)
	Pure error	10	46.64		
Data Needs—External					
Linear	Regression	2	29.31	14.65	
$R^2 = 0.04$	Residual	187	629.02		
	Lack of fit	177	573.81	3.24	0.59 (0.91)
	Pure error	10	55.21		
General Needs—IC					
Linear	Regression	2	31.56	15.78	
$R^2 = 0.04$	Residual	187	703.10		
	Lack of fit	177	672.11	3.79	1.23 (0.38)
	Pure error	10	30.99		
General Needs—Local					
Linear	Regression	2	20.52	10.26	
$R^2 = 0.04$	Residual	187	965.13		
	Lack of fit	177	927.47	5.23	1.39 (0.29)
	Pure error	10	37.66		
General Needs—Informal					
Linear	Regression	2	36.93	18.47	
$R^2 = 0.04$	Residual	187	617.95		
	Lack of fit	177	573.87	3.24	0.74 (0.79)
	Pure error	10	44.08		
General Needs—External					
Linear	Regression	2	42.00	21.00	
$R^2 = 0.04$	Residual	187	680.06		
	Lack of fit	177	612.23	3.45	0.51 (0.95)
	Pure error	10	67.83		

6 CONCLUSIONS

We propose an alternative approach to end-user classification based on a framework first proposed by Cotterman and Kumar (1989) and use fuzzy clustering as the classification method to segment respondents into three classes of end users. A crisp cluster solution could be used to segment users into different groups; however, in our view, allowing individual users to hold memberships in all three clusters represents a more realistic picture of end-user work. It creates a more dynamic an end-user profile whereby instead of categorizes an end-user into one category, a more complete picture emerges based on all end-user tasks. This method also allows us to track changes in end-user work over time since we can easily examine how membership values change, which may occur as gradual transition whereas a complete hard cluster change may only occur over longer time periods. In addition, we expect the expertise, intent and roles of individuals to evolve over time. This leads to the task of organizing support services to move end users to "higher levels." An analysis based on fuzzy clustering should allow organizations to control this aspect of their end-user strategy.

Finally, this segmentation also generates additional insights into the work of end users by allowing us to examine differences at three different levels: the cluster level, the individual level, and the cluster-conjunctive level. Membership data further allows us to examine the relationship between individuals in specific clusters and their use of support.

In another departure from traditional end-user studies, we examine four sources from which end users obtain support. In the past, much of the focus has been on support from information centers. While information centers remain a critical resource for end users, survey studies (Powell and Moore 2002) have found an increasing use of diverse support sources. Therefore, in this study, we include support from local MIS staff, informal sources, and external sources. We find that end users rely on all four sources for support in three areas: training, data, and general support.

Several studies in the past (Mirani and King 1994) have raised questions related to the differential need for support based on end-user profiles. Yet, there has been limited examination of the very basic question: Is the extent of end-user activity related to support use? In this study, we relate support use from different sources to membership probabilities and find that respondents with higher user-developer-controller activity tend to use higher support from all sources.

Thus, we find support for the argument that an end-user profile should drive appropriate end-user support strategies. Organizations where end users are involved to a greater extent in all end-user activities should make more support sources available to end users and cover a broader range of support issues. In contrast, organizations where the end-user profile suggests higher use activity may consider reliance on single support sources.

7 REFERENCES

Alavi, M., Nelson, R. R., and Weiss, I. R. "Managing End-User Computing as a Value-Added Resource," *Journal of Information Systems Management* (5:3), Summer 1988, pp. 26-35.

Barker, R. M., and Wright, A. L. "End User Computing Levels, Job Motivation and User Perceptions of Computing Outcomes: A Field Investigation," in *Proceedings of the ACM SIGCPR Conference*. New York: ACM Press, 1997, pp. 224-233.

Bezdek, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms.* New York: Plenum Press, 1981.

Brancheau, J. C., and Brown, C. "The Management of End-User Computing: Status and Directions," *ACM Computing Surveys* (25:4), 1993, pp. 437-482.

Brancheau, J. C., Vogel, D. R., and Wetherbee, J. C. "An Investigation of the Information Center From the User's Perspective," *Data Base* (17), 1985, pp. 4-17.

Carr, Y. D., and Rainer, R. "MIS's Role in Determining Information Center Services," *Journal of Information Systems Management* (7:2), 1990, pp. 33-40.

Cornell, J. A. *Experiments With Mixtures.* New York: John Wiley and Sons, 1979.

Cotterman, W. W., and Kumar, K. "User Cube: A Taxonomy of End Users," *Communications of the ACM* (32:11), 1989, pp. 1313-1320.

Duda, R. O., and Hart, P. E. *Pattern Classification and Scene Analysis.* New York: John Wiley and Sons, Inc., 1973.

Govindarajulu, C. "The Status of HelpDesk Support," *Communications of the ACM* (45:1), January, 2002, pp. 97-100.

Govindarajulu, C., and Reithel, B. "Beyond the Information Center: An Instrument to Measure Support from Multiple Sources," *Information and Management* (33:5), 1998, pp. 241-250.

McLean, E. R. "End Users as Developers," *MIS Quarterly* (3:4), December, 1979, pp. 37-46.

Mirani, R., and King, W. R. "Impacts of End-User and Information Center Characteristics on End-User Computing Support," *Journal of Management Information Systems* (11:1), 1994, pp. 141-168.

Powell, A., and Moore, J. E. "The Focus of Research in End User Computing: Where Have We Come Since the 1980s?," *Journal of End User Computing* (13:3), 2002, pp. 1-56.

Pyburn, P. J. "Managing Personal Computer Use: The Role of Corporate Management Information Systems," *Journal of Management Information Systems* (3:3), 1983, pp. 49-70.

Rivard, S., and Huff, S. L. "An Empirical Study of Users As Application Developers," *Information and Management* (8), 1985, pp. 89-102.

Rockart, J. F., and Flannery, L. S. "The Management of End User Computing," *Communicationsof the ACM* (26:10), 1983, pp. 776-784.

Schiffman, S. J., Meile, L. C., and Igbaria, M. "An Examination of End-User Types," *Information and Management* (22), 1992, pp. 207-215.

Wedel, M., and Steenkamp, J. E. M. "A Clusterwise Regression Method for Simultaneous Fuzzy Market Structuring and Benefit Segmentation," *Journal of Marketing Research* (28), November 1991, pp. 385-396.

Wetherbe, J. C., and Leitheiser, R. L. "Information Centers: A Survey of Services," *Journal of Information Systems Management* (2:3), 1985, pp. 3-10.