A Bi-metric and Fuzzy c-means Based Intelligent Stakeholder Quantification System for Value-based Software

Muhammad Imran BABAR a,1, Masitah GHAZALI a, Dayang N.A. JAWAWI a

a Department of Software Engineering, UTM, 81310 Johor, Malaysia.

Abstract: Requirements engineering (RE) deals with the software requirements elicitation, validation and verification processes and implementation of elicited requirements. The RE process, for value-based software (VBS) development, is one of the highly complex and non-linear processes which mainly depends on the key stakeholders. VBS systems are extremely different from traditional systems due to their association with financial streams. There is a need of value-based RE practices. Stakeholders’ analysis plays a vital role in the selection of critical stakeholders for RE phase in VBS development. Different stakeholders’ identification and quantification (SIQ) approaches are presented by researchers in order to identify and quantify the stakeholders. The existing approaches are not uniform in terms of process activities and different stakeholders’ attributes. Non-uniformity results in higher time consumption and higher complexity. The existing SIQ approaches are not suitable for current value-based RE practices. Hence, this research presents an intelligent decision support system for the VBS SIQP using stakeholder metrics and fuzzy logic. The data is finally divided into three clusters in order to select the highly prioritized stakeholders for VBS systems.

Keywords: software requirements engineering, value-based software, stakeholders, decision support system, fuzzy c means, clustering.

Introduction

The very term value-based is associated with value-based software engineering (VBSE). The value-based software (VBS) systems are the part of VBSE and are developed based on economic theory. Boehm states the term VBSE as “the explicit concern with value concerns in the application of science and mathematics by which properties of computer software are made useful to the people” [1]. The value of the VBS system is measured in terms of its economic worth or market leverage. The value of a VBS system is also taken into account based on its human services [2]. The economic aspect of the VBS systems distinguishes them from other traditional software applications. In VBS development, the value-based approaches are applied in order to understand the different economic aspects of the VBS system. “The value-based approach to software development integrates value considerations into current and emerging software engineering principles and practices, while developing an overall framework in which these techniques compatible reinforce each other” [3]. In VBS

1 Corresponding Author: Department of Software Engineering, UTM, 81310 Johor, Malaysia; Email: ibmuhammad2@live.utm.my
systems, an innovative idea is introduced for realization. Requirements engineering (RE) is a set of core software engineering (SE) practices related to software requirements. RE is a sub-domain of SE. In VBS development the value-based RE (VBRE) approaches are adopted. Brooks has said “the hardest single part of building a software system is deciding precisely what to build... Therefore, the most important function that the software builder performs for the client is the iterative extraction and refinement of the product requirements” [4]. Different researchers have defined RE differently. As stated by Zave “requirements engineering is the branch of software engineering concerned with the real-world goals for, functions of, and constraints on software systems. It is also concerned with the relationship of these factors to precise specifications of software behavior, and to their evolution over time and across software families” [5]. A requirement is written in a textual form, is elicited by the stakeholders and developers, and their interrelationships are managed manually [6]. Keeping in view the current business scenarios the stakeholders are seeking innovative business solutions for different software applications. Innovation is the cause of complexity in the design and development of innovative and value added systems. The complexity is the result of unclear user requirements and objectives. The complexity of the requirements can be determined through functional and non-functional aspects which are termed as function requirements (FRs) and the non-FRs (NFRs) or quality features [7]. Due to this trend of innovation different innovative RE techniques are considered as critical tools in order to remove ambiguity in the requirements [8]. Requirements elicitation phase (REP) is considered as highly significant in RE [4, 9]. RE is a challenging discipline in software development in order to make a software successful [10] and is associated with stakeholders’ identification, requirements elicitation and then implementation of the elicited requirements [11].

The success of a VBS system is directly associated with a set of key FRs and NFRs. If a software application or system meets the basic requirements criteria then the software project is successful and vice versa. The key set of valuable requirements can only be obtained from key stakeholders after applying core VBRE practices during different VBRE processes. The well managed requirements have an immense effect on the quality of the final software product [12, 13]. The well managed requirements can only be obtained after segregation of critical stakeholders from the whole universe in the domain of a given software project.

Tom Gilb defines a stakeholder as “any person or organizational group with an interest in, or ability to affect, the system or its environment” [14]. Stakeholders are key players in the RE process and possess an effective influence on the success of the software project. In order to make the project successful the functional features of the software must be in align with the critical needs or requirements of the users. The key functional aspects in case of VBS development can only be achieved by involvement of critical stakeholders in the RE process. Stakeholders usually belong to different cultures and domains. Hence, it is necessary to analyse and prioritize the stakeholders for professional software development [11, 15]. The users or stakeholders can only be satisfied if the system meets all the key needs of the stakeholders [16, 17]. A software system of good quality can only be achieved if the requirements will be of high value. Babar et al. state that “the software quality is measured based on the performance of the system, the services provided by the system and the acceptance of a system by the intended community” [18]. Different stakeholders interpret needs differently which makes the decision process complicated. Hence, it is required to find out and to add key
stakeholders in the software development life cycle and they may play a vital role in adding value to the software project.

Rest of the paper is divided into 5 main sections. Section 1 is about related works in which the current literature about the research is analysed thoroughly. Section 2 is about problem formulation in which a solution is given for the stakeholder problem. Section 3 is a detailed description of the fuzzy c-means and proposed system. Section 4 describes the results. Section 5 concludes the whole study.

1. Related Works

Researchers have presented different SIQ approaches for identification of critical stakeholders. The existing SIQ approaches use different stakeholders’ analysis processes and attributes for their identification and quantification. One of the most prominent theories in stakeholders’ analysis is Mitchell’s theory. Mitchell has divided the stakeholders into eight major categories based on stakeholders’ attributes of power, legitimacy and urgency. The reported categories of stakeholders in this research are dominant stakeholders, discretionary stakeholders, dormant stakeholders, dependent stakeholders, demanding stakeholders, non-stakeholders, definitive stakeholders and dangerous stakeholders [19]. According to Mitchell’s stakeholders’ theory, the relationship of a stakeholder can be with any attribute or all of the three attributes of power, legitimacy and urgency [19]. “A group has power to the extent it has access to coercive, utilitarian or normative means for imposing its will in the relationship. Legitimacy is a social good and more over-arching than individual self-perception and is shared amongst groups, communities or cultures” [20]. Figure 1 depicts the Mitchell’s Stakeholder Salience Theory.

![Mitchell's Stakeholder Salience Theory](image)

Pacheco and Garcia state “there is still no Stakeholder Identification Process (SIP) framework or uniform description” [21-23]. Among the current SIQ approaches, it is very difficult to decide that which approach is best for a given scenario. Among all stakeholder analysis approaches there are few approaches, may be one or two, which focus on prioritization of stakeholders while remaining focus on their identification. Using current approaches the identification and selection of valuable stakeholders is very difficult. The stakeholders’ analysis approaches identify stakeholders based on their relationships, roles and influence [15, 19, 24, 25] without telling the details at process or activity level. They provide a very high level picture of the business entities.
There are some approaches that do not take into account the aspects of relationships, roles and influence [26, 27].

The methods used in [28], for identification of stakeholders in an agile environment, are Freeman’s method [29, 30] and Mitchell’s model [19]. Freeman’s method has divided stakeholders into two groups at a higher level of abstraction that are primary stakeholders and secondary stakeholders. In [31, 32] an approach is presented based on roles and types for identification and selection of stakeholders. An approach is proposed for identification and selection of stakeholders at inter-organizational level. The stakeholder’s features which are considered in this approach are function, knowledge abilities, geographical position and level of the hierarchy. Though the approach is very powerful but the issue is how one will measure the level of competency among different stakeholders using these four features. A problem which is detected after implementation of the approach is associated with the profiles of different stakeholders. The four features are used to make the profiles of different stakeholders, but there are different stakeholders who have the same profiles with different ambitions or the FRs. The approach is proposed and implemented in order to find out all possible stakeholders for a given software system due to which lot of time is consumed. The approach is not cost efficient due to higher time consumptions. The selection of stakeholders is not based on their prioritization or quantification.

PisoSIA® (Stakeholder Identification and Analysis) approach is an extension in the existing approach called PISO® (Process Improvement for Strategic Objectives) [33]. The only suggestion which is given in PISO® framework is associated with the importance of stakeholders and their analysis. The extension in the PISO® framework is only supporting the stakeholders’ identification and analysis and the impact of these identified stakeholders on the quality of the system. PisoSIA® is used to identify stakeholders after incorporating a change in an information system. Then the impact of that change on current stakeholders is analyzed or measured. The incorporation of change helps in identification of some new stakeholders or business entities who may add some value to the system. For stakeholder identification, Mitchell’s model is integrated with PISO® that helps in the adoption of stakeholders’ attributes for their selection. The focus of both PISO® and PisoSIA® is stakeholders’ engagement without giving any consideration to their importance. The effectiveness of PisoSIA® approach is associated with the early findings if the early findings are correct then the approach will be effective or vice versa. In the case of PisoSIA® the early findings are incorrect due to which the actual efficiency of the technique is difficult to calculate.

A research called “ERP-implementation project from a stakeholder perspective” is conducted in [20]. In this research, Mitchell’s model is used that helps in the classification of stakeholders at higher level of abstraction. Just like PisoSIA® a change is incorporated in order to identify new stakeholders and to measure the impact of change on existing stakeholders. There is no novel contribution regarding SIQP that may help in the selection of valuable stakeholders for VBS systems. A change is incorporated in an ERP system and the impact of this change is analyzed on the stakeholders. SIQP is not the focus of this research directly. The “stakeholder identification <process> precedes any other RE activity: we must first determine who they are and how important they are” [34]. The different features are presented by Glinz and Wieringa for identification of stakeholders that include interest in the system, must manage, introduce, operate, or maintain the system, involvement in the development, business responsibility, financial interest, constrain the system as regulators and negatively affected by the system. The stakeholders are quantified into
three major categories that are critical, major and minor. For stakeholder identification and classification the process level details are not given. So it is difficult to adopt the model when a project or product requires an agile environment in terms of its execution, implementation or development.

A framework of stakeholder identification and selection is given in [35] and is comprised of three stages that are identification, filtering and prioritization. The framework is a conceptual overview of the necessary aspects that must be considered essential during SIQP and to make RE process effective. The proposed framework provides a very high level picture and “the framework may not be conclusive as it needs to be confirmed and refined further”. The framework is generic and is not yet implemented in real scenarios. There are some other approaches that do not focus the SIQP itself rather they discuss the stakeholders in a casual way without due consideration. The overview of these approaches is given here.

In HyDRA [36] the too much stress is given on the viewpoints of the users in order to gain information about multiple requirements’ resources and requirements’ traceability links. There is no special consideration of stakeholders as such. In QSARA (Qualitative Systematic Approach to Requirements Analysis) the major stress is on the “stakeholders <who> have little knowledge or understanding of the domain or the initial document is poorly structured” [37]. The approach is helpful in identification of resource allocation and interdependencies. It also helps analysts to construct a complete description of system features. The approach lacks in providing a solid SIQP and focuses on the stakeholders who lack in domain knowledge.

The Activity Theory for Requirements Elicitation (ATRE) focuses on contextualized activities of the stakeholders and it defines general activities that are strongly based on the knowledge and skills of the stakeholders. ATRE is based on the concept of social property which presents knowledge from social sciences that can be useful in gaining new insights into the human context in a given software system. However the elaboration of social activities is very difficult because “they require the collaboration of heterogeneous teams on wide topics” [38]. The approach directly involves the stakeholders in order to negotiate the requirements without providing any mechanism for SIQ. Same is the case with the research performed by Kasirun and Salim that only focuses on the involvement of stakeholders during requirements elicitation based on the aspects like activity, environment and support of the tool [39]. SIQP is not the focus of this research. The research study conducted by Woolridge et al. is also based on major stakeholder categories and the SIQP is based on these high level categories. The categories are financial supporters, customers, internal stakeholders, external stakeholders, special interest stakeholders, and influencer stakeholders [40]. In this research, the major stress is on the risk imposed by the stakeholders and the impact of stakeholder risk is calculated.

Williams et al. have presented a study in which a list of enterprise stakeholders is given who may involve in the commercial development. The framework is being proposed for enterprise software development coordination in which the stakeholders are identified based on “their role and their name (e.g. Bob Smith: director of marketing)”. This framework helps in the enhancement of “collaboration across enterprises engaged in software development projects” [41]. For existing work practices, it is not a hard and fast rule.

A comprehensive literature review is conducted on the said problem of stakeholders for VBS development in [18]. Babar et al. have forced that there is a dire need of a new framework, based on stakeholders’ metrics for VBS development due to
their distinguished nature. However, there are two types of the SIQP problems that are process problems and technical problems. The four major process problems of the SIQP are highly complex, non-uniform, inconsistent, and time-consuming. The technical problems of the SIQP are non-existence of metrics, lack of low level details and use of limited stakeholders’ attributes. In this research, fuzzy logic based intelligent system in order to identify and quantify the stakeholders of VBS systems.

2. Problem Formulation

In this section, the stakeholder problem is formulated. Currently, the datasets about stakeholders do not exist in any database so for data the metrics are proposed based on their skills and interest in a given project. The measurement of the two key metrics is highly important in knowing the worth of each entity for the RE process. In order to calculate the value of these key metrics the different stakeholders’ aspects are used as input. The industry experts assign the values to these attributes based on an ordinal or ranking scale. These metrics are used to quantify the stakeholders initially in a manual way. The computed data values by the experts are used in the proposed fuzzy c-means based intelligent system for the SIQP. The values of these two metrics or factors are based on stakeholders’ aspects which are described here.

\[
\text{Factor} \rightarrow F \\
\text{Aspect} \rightarrow T
\]

**Definition:** Here F is a notation for stakeholders’ factor or metric and T represents the stakeholders’ attribute or aspect. The stakeholders’ aspects are evaluated on a ranking scale of 0 to 5.

The stakeholders’ attributes considered for the two metrics are included after long discussions with industry professionals. Hence, we have two types of datasets the stakeholders’ skill factor dataset and the interest factor dataset.

2.1 Stakeholders’ Skill Factor (FSS) Dataset

This dataset is based on the stakeholders’ skill factor. The stakeholder’s aspects which are considered for the skill factor in this research are experience (TEX), managerial abilities (TMA), domain knowledge (TDK), domain training (TDT), and self-esteem (TSE).

\[
F_{SS} = 0.2 (T_{EX} + T_{MA} + T_{DK} + T_{DT} + T_{SE}) + 0.2
\]

\[
F_{SS} = 0.2 \left( \sum_{i=1}^{n} F_{SSI} \right) + 0.2 \ldots eq - 1
\]

In the equation 1, \( i \) is an element or aspect of the \( F_{SS} \) and the total number of aspects is represented by \( n \). The \( F_{SS} \) is calculated using values of aspects that are in the range of 0 to 5 and \( F_{SS} \) is in the range of 0.2 to 5.2. Table 1 shows a partial dataset of \( F_{SS} \) metric or feature. The values to each aspect are assigned by the industry experts during the evaluation of the proposed intelligent system.
### Table 1. $F_{SS}$ partial dataset sample

<table>
<thead>
<tr>
<th>$T_{EX}$</th>
<th>$T_{MA}$</th>
<th>$T_{DK}$</th>
<th>$T_{DT}$</th>
<th>$T_{SE}$</th>
<th>$\sum_{i=1}^{n} F_{SS i}$</th>
<th>$F_{SS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>3.4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>13</td>
<td>2.8</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>16</td>
<td>3.4</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>12</td>
<td>2.6</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>11</td>
<td>2.4</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>1.8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>2.2</td>
</tr>
</tbody>
</table>

### 2.2 Stakeholders’ Interest Factor ($F_{SIT}$) Dataset

This dataset is based on the stakeholders’ interest factor. The stakeholder’s aspects which are considered for the interest factor in this research are domain scope knowledge ($T_{DSK}$), business knowledge ($T_{BK}$), and objectivity ($T_{OB}$). These three aspects help in judging the interest of a stakeholder in a given project.

$$F_{SIT} = 0.2 (T_{DSK} + T_{BK} + T_{OB}) + 0.2$$

$$F_{SIT} = 0.2 \left( \sum_{i=1}^{n} F_{SIT i} \right) + 0.2 ... eq - 2$$

In the equation 2, $i$ is an element or aspect of $F_{SIT}$ and the total number of aspects is represented by $n$. The $F_{SIT}$ is calculated using values of aspects that are in the range of 0 to 5 and $F_{SS}$ is in the range of 0.2 to 3.2. Table 2 shows the partial dataset of $F_{SIT}$ metric.

### Table 2. $F_{SIT}$ partial dataset sample

<table>
<thead>
<tr>
<th>$T_{DSK}$</th>
<th>$T_{BK}$</th>
<th>$T_{OB}$</th>
<th>$\sum_{i=1}^{n} F_{SIT i}$</th>
<th>$F_{SIT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>8</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3</td>
<td>10</td>
<td>2.2</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2</td>
<td>9</td>
<td>2.2</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>7</td>
<td>1.6</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1.4</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>2</td>
<td>8</td>
<td>1.8</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>4</td>
<td>11</td>
<td>2.4</td>
</tr>
</tbody>
</table>

After calculating the values of $F_{SIT}$ and $F_{SS}$, it becomes difficult to determine the inclusion and exclusion criterion for stakeholders. The system in terms of decisions is highly complex and the data is highly fuzzy. It becomes difficult to determine that which stakeholder must be a part of RE process and which entity should not be considered. Hence, the only possible way is to apply some applied heuristic approach in order to find out the solution for the said stakeholders’ problem. In this research, the
fuzzy c-means algorithm is applied for data clustering in order to define the threshold or inclusion and exclusion criterion for stakeholders in RE phase.

3. Fuzzy oriented intelligent system

In order to make the SIQP efficient the values of stakeholders’ metrics are used as features in soft computing techniques. There are several data clustering techniques, however, the data in the current research belongs to a fuzzy problem instead of the hard problem. Hence, based on fuzzy and hard problem the clustering can be divided into fuzzy clustering and hard clustering. In clustering the data possess two properties i.e. homogeneity and heterogeneity. In homogeneity, the objects of one class or cluster are similar to each other while in heterogeneity the objects in one class are dissimilar to objects of other classes [42, 43]. The data points in fuzzy clusters hold a degree of fuzzification which shows their links with different clusters instead of showing their association with a single cluster.

3.1 Fuzzy C-Means Method

In this research, an existing clustering approach called fuzzy c-means (FCM) is used [44]. Previously the FCM is applied in the domain of software requirements prioritization [45]. The FCM is a base for the fuzzy extensions. An improved version of the FCM is widely used for pattern recognition [46]. FCM is used to classify the objects in a dataset into different clusters based on the key features or attributes. The FCM computes the right position of an object in the dataset and places it in its respective cluster in the space of two or more clusters. In the improved version of FCM, the fuzzification parameter \( m \) is introduced and is in the range of \([1, n]\). The degree of fuzzification is determined by \( m \) in each cluster. The commonly used value for \( m \) is 2. The initialization of the FCM starts from a guess in order to find out the centroid and these centroids are termed as distance mean of each cluster. The distance matrix is created using Euclidean distance formula that is given in equation 3.

\[
d(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2} \quad ... \text{eq} - 3
\]

A membership rank is assigned to every object in each cluster. The centroid and membership ranks in each cluster are updated in an iterative fashion and the position of the centroids is adjusted in each cluster of the dataset. The centroid is adjusted based on an objective function which measures the distance of every object from the centroid and is weighted by membership rank of the object in a cluster. Equation 4 is used in order to update the new centroids of the clusters and equation 5 computes the membership rank during iterative optimization process of the FCM.

\[
C_j = \frac{\sum_i [\mu_j(x_i)]^m x_i}{\sum_i [\mu_j(x_i)]^m} \quad ... \text{eq} - 4
\]
Equation 6 represents the cost function or objective function used in the FCM.

\[ J = \sum_{i=1}^{N} \sum_{j=1}^{C} \left( \frac{1}{C} \sum_{k=1}^{C} \frac{1}{\|x_i - c_j\|^m} \right) \| x_j - V_i \|^2 \]  

\[ \text{eq} \quad 6 \]

\( U_{ij} \) is the membership of \( x_j \) in the \( i^{th} \) cluster, \( V_i \) is the center of the \( i^{th} \) cluster, the bars \( \| \ldots \| \) represent the norm metric and \( m \) constant is associated with the degree of fuzzification. The data values closer to the centroid are assigned with higher values of membership and minimize the cost function while the data objects far away from the centroid are assigned with lower membership values. The probability of association of a given data object with a cluster is shown with the membership function. This probability is associated with the distance of one object in a cluster from its respective centroid.

Figure 2. Proposed Intelligent System

Figure 2 describes the proposed intelligent system for stakeholders’ identification and quantification. The interface is used as a means for input to the system. The input is of two types which are stakeholders’ parameters and fuzzy c-means parameters. The stakeholders’ parameters are given to the system in order to compute the values of \( F_{SS} \) and \( F_{ST} \). The computed values of the \( F_{SS} \) and the \( F_{ST} \) are stored in a database or in a text file. The \( F_{SS} \) and the \( F_{ST} \) are serving as stakeholders’ features in this scenario that are given as data input to the FCM. The second type of input to the system is associated with the fuzzy c-means parameters. The FCM parameters are required number of clusters \( N \), the value of exponent or fuzzification parameter \( E \), \( I \) is for the total number
of iterations and the parameter of initial centroid is also defined. After initialization, the centroids are computed and membership values $U_{ij}$ are modified. If the value of Euclidean norm $||...||$ is less than $\varepsilon$ the algorithm is going to stop otherwise the centroids are updated again using equation 3 and the loop continues until the matrix norm value is not less than $\varepsilon$. Here, $\varepsilon$ is the predefined accuracy or convergence criterion. If the threshold value of $\varepsilon$ is achieved the system will show the results in the form of different clusters of the data otherwise the algorithmic loop will continue in order to achieve the optimal results of the problem.

4. Results

The experimental results are based on the value of fuzzification parameter $m$ also called as exponent. The data in case of stakeholders’ quantification is based on two key features which are skill set and interest. The fuzzification in the data induced when a given object in the matrix is going to achieve high skill value and low interest value and vice versa. Hence, in a large data set it is extremely difficult to find out the most valuable stakeholders based on these two key features. For stakeholders’ quantification, both interest and skill features are extremely important so it is difficult to neglect them. Table 3 is a representation of the partial dataset, for the objects of the universe, which shows the values of $F_{SIT}$ and $F_{SS}$ for some objects. The current problem domain is a 2-D problem with two parameters.

Table 3. Partial dataset sample

<table>
<thead>
<tr>
<th>Object</th>
<th>$F_{SIT}$</th>
<th>$F_{SS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stakeholder 21</td>
<td>1.2</td>
<td>3.4</td>
</tr>
<tr>
<td>Stakeholder 22</td>
<td>1.8</td>
<td>1.8</td>
</tr>
<tr>
<td>Stakeholder 23</td>
<td>0.8</td>
<td>3.2</td>
</tr>
<tr>
<td>Stakeholder 24</td>
<td>2.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Stakeholder 25</td>
<td>1.0</td>
<td>1.8</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

The 2 value-based projects online car showroom and restaurant management system are chosen for experimentation in order to evaluate the proposed stakeholder quantification approach. Two teams comprised of 4 people have evaluated the $F_{SIT}$ and $F_{SS}$ features of the stakeholders. The stakeholder factors are evaluated based on the key aspects considered in these two factors. The final values of $F_{SIT}$ and $F_{SS}$ are given as an input to FCM in order to get an optimized solution for stakeholders’ quantification problem. The total population of the universe is 144 for both projects.

Table 4 describes the different values of the parameters used in fuzzy c-means clustering in this research. $N$ is the total number of stakeholder samples, $d$ represents the dimensionality, $m$ is the fuzzification parameter, $n$ is the total number of iterations, $k$ represents clusters and $\text{min}\_\text{improvement}$ is the performance or convergence criterion. The results are evaluated based on the parameters of $m$ and $\text{min}\_\text{improvement}$ by applying the variations of the both. The value of fuzzification parameter $m$ is applied in a descending order in order to know the effect on the overall clustering of the data.
Table 4. Fuzzy C-Means Parameters

<table>
<thead>
<tr>
<th>Stakeholder</th>
<th>N</th>
<th>d</th>
<th>m</th>
<th>n</th>
<th>k</th>
<th>min improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>144</td>
<td>2</td>
<td>2.5</td>
<td>100</td>
<td>2</td>
<td>10e-5</td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>2</td>
<td>10e-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.5</td>
<td>2</td>
<td>10e-005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2</td>
<td>10e-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0</td>
<td>2</td>
<td>10e-005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>2</td>
<td>10e-5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>2</td>
<td>10e-005</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exponent 2.5
Min Improvement 10e-5

Exponent 2.0
Min Improvement 10e-5

Exponent 1.5
Min Improvement 10e-5

Figure 3. FCM Results
Figure 3 shows the different results under different FCM parameters and the value of $k$ is 2. It is observed that after application of different parameters the significant effect is on the iteration count. The performance of FCM is highly optimized for the stakeholder problem with its default parameters. The default parameters of FCM are shown in bold in Table 3. In this research, the results are chosen with exponent ($m$) value of 2.0 and min improvement value of 10e-5. The results in the Figure 3 show that there is no prominent difference in clustering of the stakeholder universe.

Figure 4 shows the selected stakeholders for the two projects. The * and o represent the center of gravity or centroids. The clusters are divided into four coordinates in order to further explore their proximity with the two features of $FSIT$ and $FSS$. This thing makes for us to select the stakeholders conveniently. The horizontal and vertical lines serve as a boundary between critical and non-critical stakeholders. Hence, based on skill and interest factors of stakeholders it is decided that those stakeholders lying on the horizontal and vertical lines should be included in the RE process. However, the stakeholders belonging to the same cluster with blue entities having higher skill but less interest and vice versa should not be a part of RE process initially. The stakeholders with these 4 values are rejected initially, but their priority is second in the RE process.

However, from the data clusters shown in Figure 4 it is difficult to define the inclusion and exclusion criterion based on two clusters only. The inclusion or exclusion based on two clusters induces a sense of biasness. Maybe there will be some human error or induced biasness at this stage. In order to avoid such a situation there is a need to divide the data into more than 2 clusters. The 2 cluster approach also induces fuzziness for some data values. The division of data into 3 or more clusters will help to lessen the fuzzy situation. In the Figure 5, data is divided into 3 clusters.

Figure 5 is a representation of the data clusters for $FSIT$ and $FSS$ values of the stakeholders with $k = 3$. The signs +, o and * are the centroids of the clusters. The
cluster with centroid + is of high priority and all the entities are considered as vital for RE process. The cluster with centroid o is also vital for RE and the entities with higher interest and lesser skills and the entities with lesser interest and higher skills may be extremely important for value-based software development. The entities with centroid * are discarded at initial stage and they are not important for RE process. The stakeholders’ values shown in red are the key entities that must be included in the initial RE phase of VBS development. However, the few stakeholders of second highly prioritized cluster ‘o’ may also be selected based on some key criteria in order to elicit more requirements of the system under development.

5. Conclusion

Stakeholders are the key entities in requirements elicitation phase of the RE process. So far different SIQ approaches are proposed by researchers in this domain but unfortunately all focus on diverse domains using different methodologies. These approaches are not suitable for VBS systems due to their peculiar nature. Hence, in this research a bi-metric fuzzy c-means based approach is presented in order to resolve the plight of stakeholders of a VBS system. The stakeholders are selected based on the proposed system and the system is found promising in evaluation of the key stakeholders for the given VBS systems. This research has few limitations. Firstly, the approach is not applied on larger projects. Secondly, there is a need to apply the proposed system in a global software development environment where stakeholders are distributed across the globe. However, the future research is based on a robust intelligent VBS system development by applying heuristic approaches like neural networks, genetic algorithm and clustering methods for projects with the large number of stakeholders in order to cope with the issue of scalability.
Acknowledgement

The authors would like to thank profoundly to the Ministry of Education (MOE) Malaysia Vote No. 05H75 for financial support and not forgotten the Software Engineering Research Group (SERG) and EReTSEL lab members for their help. The authors also give special thanks to Professor Siti Maryam Shamsuddin for her kind support in this research.

References


