Contextual Clustering
Grouping and classifying problems in a defined context

Christopher J Manning[a] **and Fred DJ Bowden**[a]

[a] *Joint Operations Analysis Division*
Defence Science and Technology Organisation
Email: *christopher.manning@dsto.defence.gov.au*

Abstract: Contextual Clustering is a new Operations Research technique designed to objectively group issues within a defined context. It does not rely solely on subject matter expert opinion and provides an audit trail and logical path to the final results. Contextual Clustering enables analysts and decision makers to gain a better understanding of the problem space by considering it as a whole and then defining themes for further investigation whilst removing pre-conceived ideas and biases.

There were three key drivers in its development the most important being the need to construct a defined context for all participants to ensure that responses are based on the specified (new) concept. This should encourage participants to think beyond current practices and assist them to break current day paradigms. The next was the desire for a quick turn around time (approximately one to two hours) with limited resources, to enable feedback and further discussion with participants based on the findings during the same activity or workshop. The final key driver was the need to conduct an objective and rigorous analysis that would provide a firm basis for any future investigations. In addition to these drivers the following constraints were applicable; a small sample size (limited number of participants), a large quantity of complex inter-related issues for consideration, the unavailability of training data and the desire to conduct additional sensitivity analysis. As a result the Contextual Clustering technique was created with inspiration from a number of techniques and could be described as a "meta-method" or flexible Operational Research approach that enables the use of the most appropriate algorithms and techniques as required. The basic steps of Contextual Clustering are: (1) determine the characteristics that define the context; (2) identify issues for clustering; (3) capture relationship data; (4) apply appropriate clustering algorithm; (5) graphically represent the clusters to apply labels and enable discussion; and (6) validation/confirmation of clusters.

Contextual Clustering should be considered when conducting analysis of issues related to new concepts or contexts where an objective and rigorous analysis is required. It allows analysts and decision makers to gain a better understanding of the problem space by considering it as a whole and then defining major areas for further investigation. As a result Contextual Clustering is an especially useful technique during the problem definition and scoping stages of larger problems, as it groups issues into themes enabling a range of issues to be studied simultaneously rather than individually. The main aim of Contextual Clustering is to ensure that participants consider issues within the specified context, removing pre-conceived ideas and biases compelling participants to think about the problem differently and thus enabling current day paradigms to be broken. However, it should be implemented as part of an overarching schema including group discussion or the application of a subjective method. This will confirm immediate results and in the long term, help to validate the Contextual Clustering technique as a process for exploration of new concepts or contexts.

Keywords: *Grouping, classification, context; category; operations research; operational analysis*

1. INTRODUCTION

Contextual Clustering is a new technique developed to group or categorise issues within a specific context. There were three key drivers in its development the most important being the need to construct a defined context for all participants to ensure that responses are based on the specified (new) concept. This should encourage participants to think beyond current practices and assist them to break current day paradigms. The next was the desire for a quick turn around time (approximately one to two hours) with limited resources, to enable feedback and further discussion with participants based on the findings during the same activity or workshop. The final key driver was the need to conduct an objective and rigorous analysis that would provide a firm basis for any future investigations. In addition to these drivers the following constraints were applicable; a small sample size (limited number of participants), a large quantity of complex inter-related issues for consideration, the unavailability of training data and the desire to conduct additional sensitivity analysis. These drivers and constraints combined made existing clustering techniques and subjective methods utilising facilitated group discussion unsuitable on their own.

As a result the Contextual Clustering technique was created with inspiration from a number of techniques including Concept Mapping (Jackson and Trochim 2002) and Multi-Dimensional Scaling (Young 1985) using a graphical representation developed through the use of a suitable algorithm such as Spring Embedding (Fruchtermann and Reingold 1991, Kamada and Kawai 1989) or hybrid techniques (Morrison et al, 2003 and Jourdan and Melancon, 2004) to represent the data. An optional step during Contextual Clustering is to conduct a subjective grouping using facilitated discussion, such as Affinity Diagramming, also known as the Post-It Note Technique (Gaffney, 1999, Straker, 1996 and Holtzblatt *et al*, 2005), after the data has been captured while the clusters are being calculated and compare the results. Additional features of Contextual Clustering are the ability to control the weightings and influence the number and size of groups (without pre-setting values) as well as the ability to use existing clustering/statistical packages where appropriate.

This paper describes the steps involved with Contextual Clustering and demonstrates its use through a practical example and compares results with a subjective grouping technique. It concludes with recommended applications for Contextual Clustering.

2. CONTEXTUAL CLUSTERING

The steps in Conceptual Clustering are:

1. Determine the characteristics that define the (new) context
2. Identify Issues
3. Capture data on the relationships between the issues and characteristics
    a. Gather data from the participants on which characteristics related to each issue
    b. Gather data on the primary, secondary and tertiary characteristics of each issue
4. Determine the strength of the relationships between the characteristics
5. Graphically represent the issues in relation to the characteristics
6. Cluster the issues
7. Label the clusters
8. Validation/Confirmation of clusters
    a. Analysis of the relationship of the cluster centroid to the characteristics
    b. Optional comparison to a subjective method

These steps are described in more detail below.

2.1 Step 1: Determine Characteristics

The purpose of determining the characteristics is to define the concept and describe the context in a clear and unambiguous way that can be easily understood. It is likely that a large amount of thought as well as time and effort will have been put into generating a new concept and as a result it will probably be quite complicated. Even if a concept document has been written it may still be difficult for participants to come to a mutual understanding in the short period of time available to digest all the information. It is important that the characteristics are defined correctly and can be clearly articulated to participants as they will have a significant impact on the clusters produced. It is recognised that the characteristics are unlikely to be completely orthogonal and will have complex multi-dimensional relationships.

For these reasons a similar approach to that used in a Delphi process (Brown 1968) is employed, where the characteristics are determined by considering a statement that encompasses the context. This allows the characteristics to be distilled by a group of SMEs. An example statement is below.

'If I see X in the workplace then the future operating concept is being utilised'

In this manner the SME considerations of the statement or the "X" become the characteristics that describe the concept or context. This approach could be easily applied to other contexts, through modification of the statement. Previously developed or existing "known" key values of the concept could also be used as characteristics. However a diverse group of SMEs should review both the characteristics and the concept to ensure that the entire concept space is adequately represented.

2.2 Step 2: Identify Issues

Contextual Clustering requires as an input a well defined list of clearly articulated issues with subordinate detail so that all participants can easily comprehend what was intended when each issue was originally raised. There is a clear advantage for at least some of the participants of Contextual Clustering to have been involved in the development of the issues, as this allows other participants to ask for clarification prior to starting the clustering process. It is also possible to take a set of predefined issues into Contextual Clustering with new participants. However in both cases it is important that all participants have a common understanding of all the issues. To achieve this time should be allocated for facilitated group discussion and if required the development of rich pictures (Checkland 1981) or the conduct of other activities to refine the issues.

2.3 Step 3: Capturing the Relationships between Issues and Characteristics

The data used to determine the relationships between characteristics is captured through the conduct of a survey, as this allows detailed information to be elicited from subject matter experts in a short amount of time. This should be done directly after a review of both the issues and the characteristics so that each participant is familiar with all aspects of the survey. The survey collects each participant's belief on the connectedness of issues to characteristics. An individual survey method is suggested rather than a group approach so that each participant is not influenced by others and group dynamics don't adversely affect this step (Morgan, 1996).

For each issue, the survey captures the characteristics each individual participant believes were most related to the issue. A priority of 1, 2 or 3 is assigned according to which characteristics were most related to the issue, or simply a 'Y' for any other characteristics that were related. Thus both exhaustive and critical data sets are captured in a single survey process.

Table 1 shows part of an example completed survey form, with characteristic listed down the left hand side and the issue numbers across the top. Any number of characteristics and many issues can be part of this process. For this example only 8 issues and 12 characteristics are considered. What Table 1 shows is that for issue 7 this participant considered the primary characteristic to be Characteristic 6, followed in priority order by Characteristic 9 and Characteristic 10, while also being somewhat related to Characteristics 1, 3, 4, 7, 11 and 12.

Table 1: Part of a Sample Characteristic (rows) and Issue (columns) Survey

		Issues								
		1	2	3	4	5	6	7	8	...
Characteristics	1	3	Y	Y			Y	Y		
	2				Y				Y	
	3		Y	Y	2	2	3	Y		
	4		Y	Y	Y	Y	Y	Y		
	5				1					
	6	1	1	2	Y	1	1	1	Y	
	7	2	2	1	Y	Y	Y	Y	2	
	8								Y	
	9	Y	3	3	Y	Y	Y	2	1	
	10	Y	Y	Y	3	3	2	3	3	
	11	Y	Y	Y	Y	Y	Y	Y	Y	
	12	Y		Y	Y	Y	Y	Y	Y	
	...									

2.4 Step 4: Relationships between Characteristics

The context will be multi-dimensional in terms of the characteristics that define it. Therefore, it is necessary to determine the relationship between the characteristics within the context space. This is done by considering how any two characteristics relate to the issues. That is, the strength of the relationships between any two characteristics is

determined by how often the two characteristics are seen to have the same relationship with the same issues. Hence, if a given pair of characteristics appeared together for many of the issues they are considered to be closely related. For example in Table 1, characteristics 4 and 5 appear together only in issue 4 so from this data they would be seen to have a weak similarity. Also in Table 1, characteristics 11 and 12 appear together seven times, and thus have a strong relationship.

The strength of the relationships between the characteristics is calculated for each participant and then combined, using an appropriate technique, to give the overall strength. It may also be possible to use the absence of characteristics to the same issues to investigate the relationships in a similar way.

Although the context dimension is most likely undefined, the orthogonality of the characteristics in terms of the issues is a useful way of checking if the characteristics defined are robust and diverse and that they represent the dimensions appropriately. This is examined in more detail as a part of step 8

2.5   Step 5: Graphically represent the issues in relation to the characteristics

A simple 2D visual representation of the data was desired for ease of presentation to SMEs for discussion. Given that the characteristics are unlikely to be completely orthogonal and the fact that the concept space is multi-dimensional a 2D representation of the characteristics is created before the issues are plotted. Reducing the dimensionality, plotting and then clustering (step 6) allows discussion of the clusters and simplifies sensitivity analysis of different clustering approaches.

As mentioned above the first stage in graphically representing the issues in relation to the characteristics is to create a 2D representation of the characteristics. This is to ensure those issues relating to the same characteristics will be plotted near each other. If this is not done there is a chance unrelated issues could cluster together.

For example, suppose that the characteristics were placed arbitrarily in a circular formation as shown in Figure 1 where the characteristics are shown as large circles A to E and issues small squares 1 to 6. Any issues related to a pair of characteristics next to each other (for example, characteristics A and D) would be distributed around the edge of the circle. However, any issues related to characteristics on opposite sides of the circle would tend to cluster towards the centre. This would cluster unrelated issues together in the centre. It would not be possible to distinguish between these issues during the clustering, thus creating clusters that would not have any significance. In the example given, issue 3 relates to characteristics B and C, whilst issue 5 relates to characteristics A and E. Due to the poor choice of layout for the characteristics they cluster together.
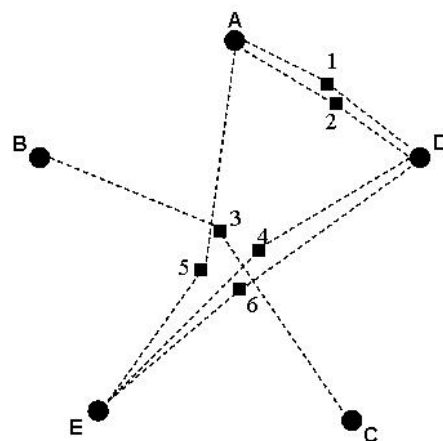

Figure 1: Example of Characteristics layout

Thus, the placement of the characteristics with respect to each other is important in order to distinguish the clusters.

There are a number of possible algorithms and methods available to achieve this representation including Multi-Dimensional Scaling (Young, 1985), Spring Embedding (Kamada and Kawai, 1989 and Fruchtermann and Reingold, 1991) and hybrid techniques (Morrison et al, 2003 and Jourdan and Melancon, 2004). For lower order numbers of dimensions both Multi-Dimensional Scaling and Spring Embedding are suitable, but as the number of dimensions increases the time taken to compute and the number of iterations required increases. So for higher order numbers of dimensions, hybrid techniques, although slightly more complicated, are more effective.

Once an appropriate two-dimensional characteristic representation has been developed, the issues can then be plotted in relation to this projection of the characteristic space. To do this, only the top three characteristics are used. Even with the care taken in positioning of the characteristics there is still a chance of incorrect placement of issues as illustrated by Figure 1. Thus, when considering any issues positioned between characteristics or groups of characteristics you need to look at why they are in that position. This is aided by considering how groupings change with different weightings. Once more it is also useful to be able to represent pictorially the issues and their related characteristics to ensure these types of issues are identified.

As each participant conducted the survey, they selected a Primary, Secondary and Tertiary characteristic for each issue. This data is aggregated using a weighted sum over all the participants to get the overall Primary, Secondary and Tertiary characteristics. If the characteristics and issues are not independent of each other, then a weighted sum

will not be the appropriate aggregation technique and the reader is referred to Yoon and Hwang, 1995 and Williams *et al*, 2010 for alternative approaches.

Initially the data for each issue is entered in one table so that all characteristic comparisons for a given issue are captured in one place. The two axes for this table are the participants and characteristics. Table 2 shows an example of this data with 12 characteristics and nine participants. The last four rows in this table show how the data aggregation is conducted. The first three of these four rows count the number of Primary, Secondary and Tertiary (1, 2 and 3) selections for each characteristic over all the participants. The final row then gives a weighted value calculated using:

$$WS = Sp * W_1 + Ss * W_2 + St * W_3 \qquad (1)$$

Where WS is the Weighted Sum, Sp Ss and St are the sums of the primary, secondary and tertiary priority selections respectively, W_1 is the primary weighting, W_2 is the secondary weighting and W_3 is the tertiary weighting, noting that $W_1 + W_2 + W_3 = 1$.

The selected weightings should be determined by thorough discussion with the participants of the difference between the three priorities and then subjected to sensitivity analysis. For example a weighting of {4/7, 2/7, 1/7} might be chosen to reflect the relative importance of each of the preferences with the perception of the participants' ratings of: first being much more important than both the second and third and the second more important than the third i.e. geometrically declining importance. Sensitivity analysis should be conducted over a range of weightings including extremes, such as {1, 0, 0} where only the primary characteristic is considered or {1/3, 1/3, 1/3} where all are considered equal, as well as finer adjustments. Also by considering how the issues move and clusters change throughout the conduct of the sensitivity analysis it is possible to confirm that the characteristics have been placed correctly and avoid incorrect issue clustering as illustrated in Figure 1 A graphical representation of the issues and their related characteristics can be particularly useful during this sensitivity analysis for ease of identifying potential problems.

The three characteristics with the highest weighted values are selected to be the overall Primary, Secondary and Tertiary characteristics for an issue. In cases where two or more characteristics have the same weighted value, the number of 1s, 2s, 3s and Ys is used, in that order, to discriminate between them. (That is, the characteristic with the most 1s is selected, then if it is still tied the one with the most 2s, and so on). An example is shown in Table 2.

Table 2: Example Characteristic Selection

	Issue X	Characteristics				
		A	**B**	**C**	**D**	**E**
	P1	1	Y	Y	2	3
	P2	1	0	2	3	0
	P3	2	0	Y	3	1
	P4	Y	1	3	Y	2
Participants	**P5**	2	1	3	Y	Y
	P6	Y	1	3	2	Y
	P7	Y	Y	2	3	1
	P8	Y	Y	2	1	3
	P9	Y	Y	1	2	3
Sum of Priorities	**Sp**	2	3	1	1	2
	Ss	2	0	3	3	1
	St	0	0	3	3	3
Weighted Sum		180	180	180	180	180

Applying the selection process described above, for issue X the primary, secondary and tertiary characteristics are B, A, and E respectively. Since characteristics A to E all have the same total score, the total number of 1's is compared and we find that characteristic B has three 1's whilst characteristics A and E both have two 1's. Hence characteristic B is the primary characteristic and characteristics A and E will be the secondary or tertiary characteristics determined by the number of 2's, 3's and Y's if required. Characteristic A has two 2's and characteristic E has one 2. Therefore, A is the secondary and E is the tertiary characteristic.

Once the aggregated priorities are established each issue is plotted on the two-dimensional characteristic map using the formula:

$$R = \frac{W_1 X_1 + W_2 X_2 + W_3 X_3}{3}$$

(2)

Where R is the resultant location, W_1 is the primary weighting, W_2 is the secondary weighting and W_3 is the tertiary weighting, noting that $W_1 + W_2 + W_3 = 1$, X_1, X_2 and X_3 are x and y locations of Primary, Secondary and Tertiary characteristics respectively.

Again the weighting values can be varied as part of the clustering process to reflect the relative importance of each of the preferences with the perception of the group's ratings as described above.

2.6 Step 6: Clustering the Issues

Having positioned the issues so that those relating to the same characteristics are in similar regions of the two-dimensional space, the next step in the Contextual Clustering technique is to identify the clusters of the issues. Note that although this is a key step, the problem addressed by the Contextual Clustering technique is much broader than just the algorithm used to cluster the data. The clustering of the issues is just one step in the over all technique.

There are many different clustering algorithms to choose from each with a different approach and outcome. Kaufman and Rousseeuw 1990 provide guidance on considerations for this choice, but in essence the most appropriate clustering algorithm should be selected, based on the expected data, to be used as part of the Contextual Clustering technique. Some well known algorithms are the K-means (MacQueen 1967 and Hartigan, 1975), the Fuzzy C-means (Bezdek, 1981), the Single Linkage or Nearest Neighbour (Sneath 1957), Self-Organising Maps (Kohonen, 1997) and the mixture of Gaussians (Everitt *et al*, 2001). It should be noted that the selected algorithm needs to be flexible enough to allow the weights and number of clusters to be varied as described previously to enable sensitivity analysis to be undertaken.

2.7 Step 7: Label the clusters

Once the visualisations are complete and sensitivity analysis for appropriate weightings has been done, a workshop is conducted in order to name each visual cluster on the basis of the common theme that links all issues within it. This is a group activity where the participants look at the issues that make up a cluster and give the cluster an appropriate name using group consensus. This enables a quick and easy reference to the cluster and the resultant theme and is also an opportunity for participants to provide feedback on the clusters giving insights into their usefulness.

For example consider a cluster containing three issues and give it a name based on the detail for those issues contained within it. Lets say that the issues are "milk not put back in the Fridge", "lid left off the sugar and ants keep getting in" and "coffee/tea spilled all over the bench" the cluster could be named "dirty kitchen" based on the theme that people making tea/coffee are not cleaning up after themselves and leaving the kitchen in a mess.

2.8 Step 8: Validation/Confirmation of clusters

Part A: Relating the cluster centroids to the characteristics

This step looks at the position of the centroid of each cluster in relation to the characteristics and gives an indication of which characteristics a particular cluster relates to. Hence providing an indication of the validity and coverage of the characteristics generated to represent the context and ensuring that the two-dimensional representation of those characteristics does not create any unbalanced clusters.

In order to do this, the centroid of each cluster is calculated for each ratio using the plotted positions of the issues for each of the relevant characteristic weighting sets. This centroid is calculated as the average of the positions of each of the elements of the cluster. This centroid is then used to determine the closest characteristic to each cluster. Where there is more than one characteristic shown as being close to the centre of a cluster it means that multiple characteristics are of similar importance.

Part B: Optional comparison to subjective method

A useful optional step in the Contextual Clustering technique is to use a subjective grouping technique such as Affinity Diagramming, also known as the Post-It Note Technique (Gaffney, 1999, Straker, 1996 and Holtzblatt *et al*, 2005). It is recommended that this is conducted after the data capture surveys while the contextual clusters are being calculated. By conducting the subjective technique after the individual data capture the participants will begin this group activity having already put considerable thought into each issue and what it relates to in the defined context.

The resultant groups from the subjective method can then be compared to the clusters and any similarities or differences investigated and discussed with participants. The real value of this is that combining Contextual Clustering with a subjective method enables an increased level of confidence in the set of clusters produced. This is because the individual assessments made as a part of Contextual Clustering and the whole Contextual Clustering technique itself are supported by the group judgment conducted as a part of the subjective method. Thus, confirming and validating the Contextual Clustering technique. The overall combination of objective (both quantitative and qualitative data) and subjective information as well as the use of both group and individual processes takes advantage of the strengths of the approaches whilst minimising their weaknesses.

3. PRACTICAL APPLICATION

Contextual Clustering was used to determine and prioritise the major themes for further exploration during experimentation and force development for the Australian Army. The following describes how Contextual Clustering was applied in this case.

The first step in Contextual Clustering, determination of the context's characteristics, was completed by considering the Army's concept document for future operations, using the following statement.

'If I see X on the battlefield then the future army operating concept is being utilised'

This resulted in 12 characteristics being identified. An existing list of 117 issues was taken into Step 2. The existing issues were a combination of concerns that were identified during feedback from previous activities and responses to specific questions asked in an effort to further explore the concept. Some of the participants had been involved in the previous activities and so had already taken part in the development of the issues. For the benefit of those who had not seen the list of issues before, the group discussed and then refined the issues list, adding detail where necessary, in order to develop a common and thorough understanding of the issues. This step generated an issues database containing 39 high level issues with some previous issues identified as sub-issues within a high level issue. For clarity only the high level issues were taken forward for Contextual Clustering and a brief heading/statement to capture the essence of the high level issues was created for quick reference with underlying detail and sub-issues available if required.

The data captured on the relationships between the issues and characteristics and the strength of the relationships between the characteristics as required for Step 3 was completed using an online survey. Using the results of this survey, the individual participant's characteristic relationship strengths were summed across all the participants and normalised to give the overall strength to complete Step 4.

The graphical representation (Step 5) was completed using the program Pajek (Batagelj and Mrvar, 1998) utilising Spring Embedding. This was chosen due to the low number of dimensions (12). In essence, this program took the 12 orthogonal axes of characteristics in the 12-dimensional space and projected them onto the two-dimensional plane, whilst retaining some graphical relationship between the characteristics. The weightings selected for use during Spring Embedding were 60, 30 and 10 to represent each individual's first, second and third preferences respectively. These rating were chosen to reflect the relative importance of each of the preferences with the perception of the groups' ratings of: first being much more important than both the second and third and the second more important than the third. Sensitivity analysis showed that when plotting the issues there was a minimal difference between weightings of 70, 20, 10 and 50, 30, 20 when compared to 60, 30, 10.

These issues were then clustered using a process similar to that used in the K-means (MacQueen 1967 and Hartigan, 1975), and Fuzzy C-means (Bezdek, 1981) clustering algorithms. This process was chosen as it is possible for the data to belong to more than one cluster which was important as the nature of the issues list suggests that points could belong to multiple clusters. This process was also flexible enough to allow the number of clusters or the weightings used to create the clusters to be changed quickly and easily in order to determine the impact this has on clusters. Note that although Step 6, Clustering the Issues, is a key step, the problem addressed by the Contextual Clustering technique is much broader than just the algorithm used to cluster the data. The clustering of the issues is just one step in the over all technique. This process will now be described for illustrative purposes.

The inputs to the clustering process were a list of two-dimensional points (expressed as (x, y) coordinates) and a variable h that represents the radius of a cluster. The output was a list of clusters, where each cluster is a list of the points that fall within h of one another. The centre of each cluster is the midpoint of all the points within that cluster. Initially this cluster list is empty. As each point is read in, it is checked against the list of clusters. If the point does not fall within h of an existing cluster centre, a new cluster is added with only that point. Conversely, if the point does fall within h of a cluster centre, it is added to the list and a new centre point is calculated. Points may belong to more than one cluster, and thus the clusters may overlap.

During the analysis it was found that after conducting the clustering for a number of weights and clustering radii, a clear clustering pattern could not be identified. To overcome this, a combination of three standard weightings was used, $\{1, 0, 0\}$, $\{1/3, 1/3, 1/3\}$ and $\{4/7, 2/7, 1/7\}$. These three sets of weightings represent the cases where: only the Primary characteristic is considered; where the priority of all characteristics is equal; and where each characteristic is considered of decreasing importance with a geometric decline. Hence, these weightings provide an insight to the influence of the different characteristics on the cluster an issue was placed in. For each combination, a set of clusters was identified using some subjective visual inspection. The key consideration was how the different issues were clustered for different weights. An issue was determined to be part of a cluster if it was contained within that cluster in at least two combinations of weights. These issues formed the initial base clusters that were then added to. The remaining issues were then clustered dependent on how they moved in relation to the existing clusters for other different weights, with the primary always equal to or greater than the other characteristics and the tertiary always less than or equal to the secondary characteristic. An example of clustering the remaining issues is shown in Figure 2.
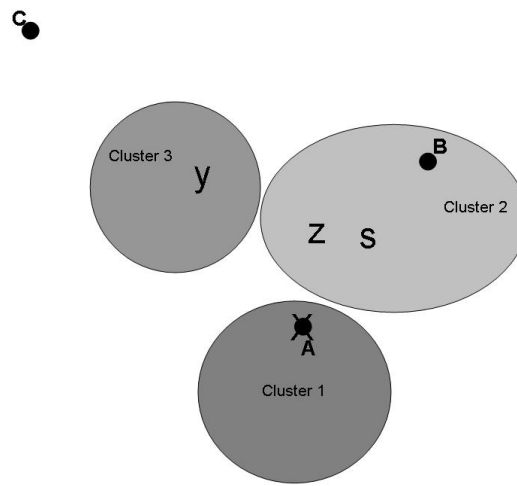

Figure 2 Remaining Issues Clustering Example

In Figure 2, the points A, B and C represent the primary, secondary and tertiary characteristics respectively of a remaining issue. Clusters 1 to 3 are existing clusters determined as indicated above. The location of the issue on the plot for the three standard weightings is shown by x, y and z representing weightings of $\{1, 0, 0\}$, $\{1/3, 1/3, 1/3\}$ and $\{4/7, 2/7, 1/7\}$. A range of other different weightings are considered, for example with a weighting of $\{1/2, 1/2, 0\}$, shown as S, the issue would be within Cluster 2. Hence, this issue would be added to Cluster 2 as when the issue moves as a result of a range of different weightings it is contained within Cluster 2 more than the other clusters.

The clusters were then labelled during a facilitated group discussion as per Step 7 of Contextual Clustering. This process highlighted a previously un-recognised group of issues that will require further investigation due to the level of uncertainty associated with their new and emerging nature. The participants considered this to be a significant discovery.

The final step in Contextual Clustering is to validate/confirm the clusters. The first part of this as described in Step 8a is to analyse the relationship of the cluster centroid to the characteristics. In this case there were no particular dominant characteristics, showing that the characteristics developed captured the basis of the context, thus creating a balanced framework to build upon. Also, with more than one characteristic being close to all but two clusters, it is evident that the spring embedding was successful with no central groupings formed. The second part is a comparison between the contextual clusters and subjective themes. As expected, there was similarity between these groupings. However, the participants felt that the contextual clusters were further removed from traditional stove-piped specifications. As a result, the contextual clusters can be described more independently and provide more tangible topics to be explored. Additionally the participants actually preferred the contextual clusters indicating that the resultant topics are further removed from the current paradigm, and thus are more likely to meet the requirements for the development of the future operating concept.

4. CONCLUSIONS & RECOMMENDATIONS

Contextual Clustering is a flexible Operational Research approach that enables the use of the most appropriate algorithms and techniques as required. It should be considered when conducting analysis of issues related to new

concepts or contexts where an objective and rigorous analysis is required. Contextual Clustering allows analysts and decision makers to gain a better understand of the problem space by considering it as a whole and then defining major areas for further investigation. As a result it is an especially useful technique during the problem definition and scoping stages of larger problems, as it groups issues into themes enabling a range of issues to be studied simultaneously rather than individually. The main aim of Contextual Clustering is to ensure that participants' consider issues within the specified context, removing pre-conceived ideas and biases compelling participants to think about the problem differently and thus enabling current day paradigms to be broken. Contextual Clustering provides an audit trail and logical path to the final result including reference back to the context of interest, with confidence provided by the ability to conduct sensitivity analysis.

However, Contextual Clustering should continue to be implemented as part of an overarching schema including group discussion or the application of a subjective method. This will confirm immediate results and in the long term, help to validate the Contextual Clustering technique as a process for exploration of new concepts or contexts. This will also exploit the strengths of both group and individual techniques whilst minimising each of their weaknesses.

ACKNOWLEDGEMENTS

REFERENCES

Akerelrea, C and Zimmerman, D (2002), A group card sorting methodology for developing informational web sites, *Proceedings of the 2002 IEEE Professional Communications Conference* 437-445.

Batagelj, V and Mrvar, A (1998), Pajek - Program for Large Network Analysis, *Connections21* 2, 47-57.

Bezdek, JC (1981), *Pattern Recognition with Fuzzy Objective Function Algorithms* Plenum Press, New York.

Bowden, FDJ, Bowley, D, Coutts, A, Williams, P and Baldwin, M (2007) , The Role of AEF in Army's Objective Force Development Process, *proceedings of the Land Warfare Conference*, Brisbane, 24-27 October.

Brown, B (1968), *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts* RAND Corporation, California.

Checkland, P (1981) *Systems Thinking, Systems Practice*, Wiley, Chichester.

Darlington, RB (1997), Factor Analysis, http://comp9.psych.cornell.edu/Darlington/factor.htm accessed 11 October 2010.

Everitt, B, Landau, S and Leese, M (2001) *Cluster Analysis, fourth edition* Arnold, London.

Fruchtermann, TMJ and Reingold, EM (1991), Graph drawing by force-directed placement, *Software - Practice and Experience* 21, 1129-1164.

Gaffney, G (1999), What is Affinity Diagramming?,

http://www.infodesign.com.au/usabilityresources/affinitydiagramming accessed 11 October 2010.

Härdle, W and Simar, L (2003), *Applied Multivariate Statistical Analysis*, Springer-Verlag, Berlin.

Hartigan, J (1975), *Clustering Algorithms*, John Wiley & Sons, New York.

Holtzblatt, K, Wendell, JB Wood, S (2005), *Rapid Contextual Design: a how-to guide to key techniques for user centred design*, Elsevier/Morgan Kaufmann, Boston.

Jackson, KM and Trochim, WMK (2002), Concept Mappings in an Alternative Approach for the Analysis of Open-Ended Survey Responses, *Organizational Research Methods*, 5, 4, 307-336.

Jourdan, F and Melancon, G (2004), Multiscale hybrid MDS, *International Conference on Information Visualization*, 338-393.

Kamada, T and Kawai, S (1989), An Algorithm for Drawing General Undirected Graphs, *Information Processing Letters*, 31, 7-15.

Kaufman, L and Rousseeuw, PJ (1990), *Finding Groups in Data: An Introduction to Cluster Analysis,* John Wiley and Sons, New York.

Kohonen, T (1997), *Self-Organizing Maps*, Springer-Verlag, New York.

MacQueen, J (1967), Some Methods for Classification and Analysis of Multvariate Observations, *5th Berkeley Symposium Mathematics Statistics and Probability*, 1, 281-297.

Martin, S (1999), Cluster Analysis for Web Site Organization, http://www.internettg.org/newsletter/dec99/cluster_analysis.html accessed 11 October 2010.

Morgan, DL (1996), Focus Groups, *Annual Review of Sociology*, 22, 129 – 142.

Morrison, A, Ross, G and Chalmers, M (2003), Fast multidimensional scaling through sampling, springs and interpolation, *Information Visualization,* 68-77.

Sneath, PHA (1957), *Numerical Taxonomy,* W.H. Freeman, San Francisco.

Straker, D (1996), *Rapid Problem Solving with Post-it Notes*, Gower Publishing, Aldershot Hampshire.

Tabachnick, BG and Fidell, LS 1989, *Using Multivariate Statistics,* 2nd edn, Harper Collins Publications, New York.

Wasserman, S and Faust, K (1999), *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York.

Young, FW 1985, *Multidimensional Scaling, Encyclopaedia of Statistical Sciences,* vol. 5*,* John Wiley & Sons, New York.

Yoon, K and Hwang, CL 1995, *Multi Attribute Decision Making: An Introduction*, Sage, London.

Williams, P, Erbacher, P and Bowden, FDJ 2010, 'The Surface-Weighted Options Ranking Technique', *presented at The 21st International Conference on Multiple Criteria Decision Making*, Jyvaskyla Finland, June 2010