A Combined Multidimensional Scaling and Hierarchical Clustering View for the Exploratory Analysis of Multidimensional Data

Paul Craig*[a], Néna Roa-Seïler[a, b]
[a] Universidad Tecnológica de la Mixteca, Carretera a Acatlima Km. 2.5, Huajuapan de León, Oaxaca, México C.P. 69000; [b] Edinburgh Napier University, 42 Colinton Rd, Edinburgh, United Kingdom, EH10 5BT

ABSTRACT

This paper describes a novel information visualization technique that combines multidimensional scaling and hierarchical clustering to support the exploratory analysis of multidimensional data. The technique displays the results of multidimensional scaling using a scatter plot where the proximity of any two items' representations is approximate to their similarity according to a Euclidean distance metric. The results of hierarchical clustering are overlaid onto this view by drawing smoothed outlines around each nested cluster. The difference in similarity between successive cluster combinations is used to colour code clusters and make stronger natural clusters more prominent in the display. When a cluster or group of items is selected, multidimensional scaling and hierarchical clustering are re-applied to a filtered subset of the data, and animation is used to smooth the transition between successive filtered views. As a case study we demonstrate the technique being used to analyse survey data relating to the appropriateness of different phrases to different emotionally charged situations.

Keywords: Information Visualization; Multi-dimensional Scaling; Hierarchical Clustering

1. INTRODUCTION

By far the most widely used of set of guidelines for the design of information visualization interfaces is the *Visual Information-Seeking Mantra*[1]. The mantra summarizes several design guidelines as *Overview first, zoom and filter, then details on demand* and is supported by task by data type taxonomy compiled to guide researchers to new opportunities and help categorize prototypes and techniques. The taxonomy contains seven data types; 1 dimensional, 2 dimensional, 3 dimensional, temporal, multi-dimensional, tree-structure and network. The general idea of the taxonomy is that visualization techniques can be easily adapted for use with similarly classified data regardless of the specific application area. One of the most common data classifications is that of multidimensional data which includes data with multiple non-spatial attributes. Examples of multi-dimensional visualizations are those that deal with stock market statistics[2, 3], microarray data analysis[4-6] and survey data such as the phrase appropriateness survey data described in the case-study for this paper. Some common tasks associated with multidimensional data are to scale the data to two dimensions in order to have an overview or divide the data into clusters to facilitate navigation. This paper describes a new technique which combines multidimensional scaling and hierarchical clustering in a single visualization in order to realize both of these objectives.

2. RELATED WORK

Clustering[7] is the most popular operation employed in the analysis of multidimensional data. While the term clustering is often more specifically used to describe the procedure of applying algorithmic methods to partition data into subsets (clusters) of items, it can also be more broadly defined to include any procedure that provides a visual representation of the results from which groupings can be interpreted. These include principle component analysis [8], self-organizing maps[9, 10] and multidimensional scaling[11-14].

A conceptualization common to most forms of clustering is the notion of a multidimensional space (also known as n-dimensional or attribute space). Here, items are thought of as having a position in n-dimensional space according to their attribute values. Each distinct data attribute is thought of being a dimension (n attributes equates to n dimensions) with each item's position determined by its attribute values. A related concept is that of similarity or dissimilarity. Dissimilarity is commonly calculated as the Euclidean distance between items in n-dimensional space (see Eqn. 1).

Similarity is calculated as the inverse of dissimilarity. One basic use of similarity measures is the creation of a similarity matrix. Here both rows and columns correspond to the full list of items while items reflect inter-item similarities.

$$d(i, j) = \sqrt{\sum_{x=1}^{n} (i_x - j_x)^2} \tag{1}$$

2.1 Multidimensional scaling

Multidimensional scaling is a common approach to the visualization of multidimensional data which represents items as points placed in a 2d or 3d display space so that their proximity corresponds to their calculated similarity. As the similarity matrix of all item-item similarities exists in a higher dimensional space than the two or three dimensional display space it is impossible to make proximity proportional to similarity in all cases. Instead, multidimensional scaling produces a representation where the proximity of any two items is only approximate to their calculated similarity. Multidimensional scaling is typically used in areas such as genomics[11-14] where there is a need to analyze large scale data or marketing where the data tends to have a lot of dimensions.

2.2 Hierarchical Clustering

Another approach to the analysis of multidimensional data is clustering where the data is partitioned into clusters to provide an overview or facilitate navigation. The most common form of cluster analysis is hierarchical clustering[15]. This is an algorithmic method which partitions the data to produce nested clusters adhering to a hierarchical structure. At the base of the hierarchy are individual data-items while at the top is a single super-cluster containing all items. Higher levels of the hierarchy include clusters with higher numbers of items. All clusters are conglomerates of lower level clusters with no two clusters on the same level including the same item. Hierarchical clustering algorithms are either divisive (constructed from the top layer down) or, more commonly, agglomerative (constructed from the bottom layer up).

The basic process of agglomerative clustering begins by considering each item as its own cluster. After this, the most similar pair of clusters are merged. This repeats until all items become merged into a single cluster. The clusters formed at each iteration of the cycle can be combined to classify items at different levels of a hierarchical tree structure.

When a cluster contains multiple items there are different methods of measuring the distance from that cluster to any other cluster. These include; single linkage which rules that the similarity between two clusters is the maximum similarity between any item in the first cluster and any item in the second, complete linkage which is the opposite of single linkage and rules that the similarity between two clusters is the minimum similarity between any item in the first cluster and any item in the second, and average linkage (sometimes referred to as un-weighted pair-group average linkage) which rules that the similarity between two clusters is the average of all the similarities between item in the first cluster and items in the second.

These methods have their various advantages and disadvantages. For example, single linkage does well at detecting long chained clusters of items but suffers from the disadvantage that the close similarity of two items will force the amalgamation of two, otherwise dissimilar, clusters. Complete linkage avoids this disadvantage but is unable to detect chained clusters. Average linkage is the most popular option. Despite being more computationally expensive it avoids undesirable phenomena caused by items with outlying patterns of expression dominating the output.

The most common visual representation of hierarchically clustered data uses the results to produce a combined heat-map/dendrogram display[15, 16] (Figure 1 bottom left). In the heat-map (also known as a color mosaic), attribute values are color-coded and displayed in a grid with rows corresponding to items and columns corresponding to attributes. Items are ordered so that the groups defined by hierarchical clustering are unbroken. The dendrogram is a type of binary tree which is attached to the left-hand-side of the color-mosaic to illustrate the groupings defined.

2.3 Multiple Coordinated Views

An information visualization interface need not be restricted to a single view of the data. Indeed, the majority of interfaces include two or more views relating to the same data. These views are often *linked* [17] so that interaction with one view will transform the data representation in another. Here multiple combined views have a variety of functions [18]. Normally, however, they can be considered as different representations that give the user a better understanding of the underlying data[18], or allow the user to select and manipulate objects more easily[19]. Different views either represent

different mappings of the same data or different subsets of the data. In the former case the different representations can serve different user requirements and combine to allow the user to form a better understanding of the data. When views relate to different subsets of the data, an overview might be linked to a detail view to provide context, allowing the user to determine which subset of the data they are looking at.

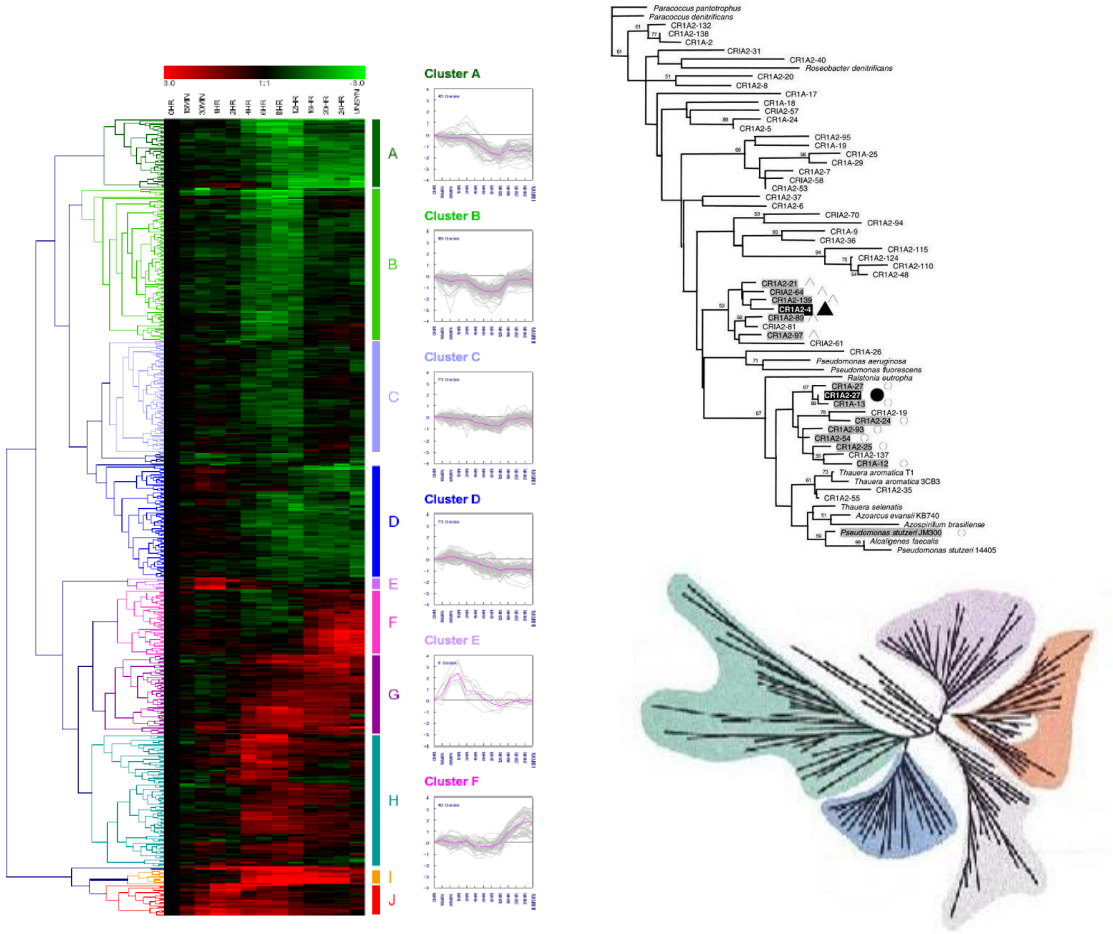

Figure 1. Views of Hierarchical Clustering Results; Left, heat-map/dendrogram visualization of gene expression (taken from[19]); Top right, a horizontal distance tree representations of hierarchical clustering results with symbols representing gene classifications (taken from [21]); Bottom right, the radial distance tree representations of hierarchical clustering results with five main visual clusters color coded (taken from [22]).

3. COMBINED MULTIDIMENSIONAL SCALING AND HIERARCHICAL CLUSTERING VIEWS

The combined multidimensional scaling and hierarchical clustering view is shown in figure 2. This uses a combination of techniques to reveal patterns in multidimensional data. Firstly, multidimensional scaling is used to position items as points in the two dimensional display space. Hierarchical clustering is then used to outline natural groupings. We decided to combine these two particular methods since both provide us with different types of overview with neither relying on the data having any particular characteristic other than natural groupings. MDS is more flexible allowing us to observe general patterns, partial clusters and outliers while the fixed clusters of hierarchical clustering make it easier for us to draw conclusions since the clusters are derived algorithmically rather than being products of (fallible) human perception. We use the average linkage method to quantify the similarity of two clusters to avoid chaining and sensitivity to outliers. The grouping outlines generated by the hierarchical clustering are drawn as shapes that surround items at a fixed radius. These are smoothed to remove concave internal angles to simplify the shapes reducing clutter and

improving general legibility[4]. Outlines are also colour coded (from white to green) according to the strength of the cluster. The strength of a cluster is calculated as the difference in similarity between successive combination similarities. This means that if a cluster is dissimilar to its siblings and its children are similar to each other, it can be judged to be a stronger cluster and appears to be more prominent in the display. Figure 2 shows the view operating with survey data where 10 distinct users were asked to make appropriateness judgements (72,638 in all) for 134 distinct terms and the spectrum of 32 emotions proposed by Plutchik[20]. The data can be seen to contain three prominent natural clusters of emotions. These are negative emotions, positive emotions and very positive emotions (Ecstasy and Joy).

Labels are attached to each item in the cluster view. When no items are highlighted all labels have equal weighting but when any number of items are highlighted, those labels are darkened and the unselected labels are greyed out. This allows the labels to be used for exploring the data actively by moving the mouse or passively by shifting focus. Items can be highlighted in the cluster view by moving the mouse cursor close to a point, moving the mouse cursor inside a cluster outline (Figure 2 lower left) or using an excentric labelling[21, 22] tool to highlight items within a fixed radius. If two cluster or more outlines overlap, moving the mouse cursor inside the overlapping area highlights all the overlapping clusters.

Groups of items can be selected in the cluster view by either clicking on a highlighted cluster, dragging a freeform shape around a group of emotions or dragging a box around emotions. Once a subgroup of items are selected, both multidimensional scaling and hierarchical clustering are re-applied for the filtered data set. This allows the user to view sub-sets of the data with more detail. Figure 2 lower right shows the cluster view zoomed into the positive emotion cluster to reveal two interesting nested clusters. When the view changes, animation is used to smooth the transition[23] with newly selected items fading in, de-selected items fading out and all other items moving gradually to their new positions.

3.1 Cluster Explorer

Figure 3 shows the Cluster Explorer application which was developed to test the combined multidimensional scaling hierarchical clustering view by allowing it to be used to support the analysis of phrase appropriateness data. The application consists of three main views: the multidimensional scaling hierarchical clustering view, a matrix view that shows the appropriateness for every term/emotion pair, and a detail view that shows the most appropriate terms for selected emotions or groups of emotions.

The matrix view of the application shows the average appropriateness rating for every term-emotion pair in the data set. Each row represents an emotion and each column represents a term. Each cell is color-coded on a scale from white to green in order to communicate the percentage of users that judged the term of the column to be appropriate to the emotion of the row. Since there is not enough space to label rows and columns, emotion names, term names or appropriateness ratings, these are displayed using tool-tip text that is activated when the mouse cursor passes over a cell. Moving the mouse over a cell also highlights the selected emotion and term in all other views of the application. When an emotion or set of emotions are selected to be highlighted (in this or any other view) the relevant rows are coloured using a different scale (from white to red). In figure 3 we see the term ‘surprising’ highlighted with the emotion *Surprise*.

In ordered to reveal natural groupings, both rows and columns of the matrix view are ordered by similarity. This is done using hierarchical clustering to organize terms or emotions into nested clusters with clusters at the same level arranged so that the most similar clusters are juxtaposed. Terms are clustered according to the responses for different emotions. Emotions are clustered according to the responses for different terms. As with the multidimensional scaling hierarchical clustering view, we use the average linkage method to avoid chaining and sensitivity to outliers.

The application detail view shows term statistics for highlighted emotion or emotion groupings. Here terms are ordered according to average appropriateness and colour coded (white to green) according to their appropriateness for the selected emotion(s) in relation to the dataset as a whole. Here we found that some, more neutral, terms such as ‘can you repeat that’ are highly placed for most groups but are rarely highlighted in the colour coding since they are common to all groups. Other terms are specific to certain emotion or emotion groupings. For example, the term ´that’s great’ is judged more than four times more appropriate for the *strong positive* emotions that for the data-set as a whole. We can also observe similarly high ratios for pairings like ‘how awful’ and the emotion *Awe*, ‘wow!’ and *Amazement* etc. The metric used for relative appropriateness is the fold change from data-set appropriateness.

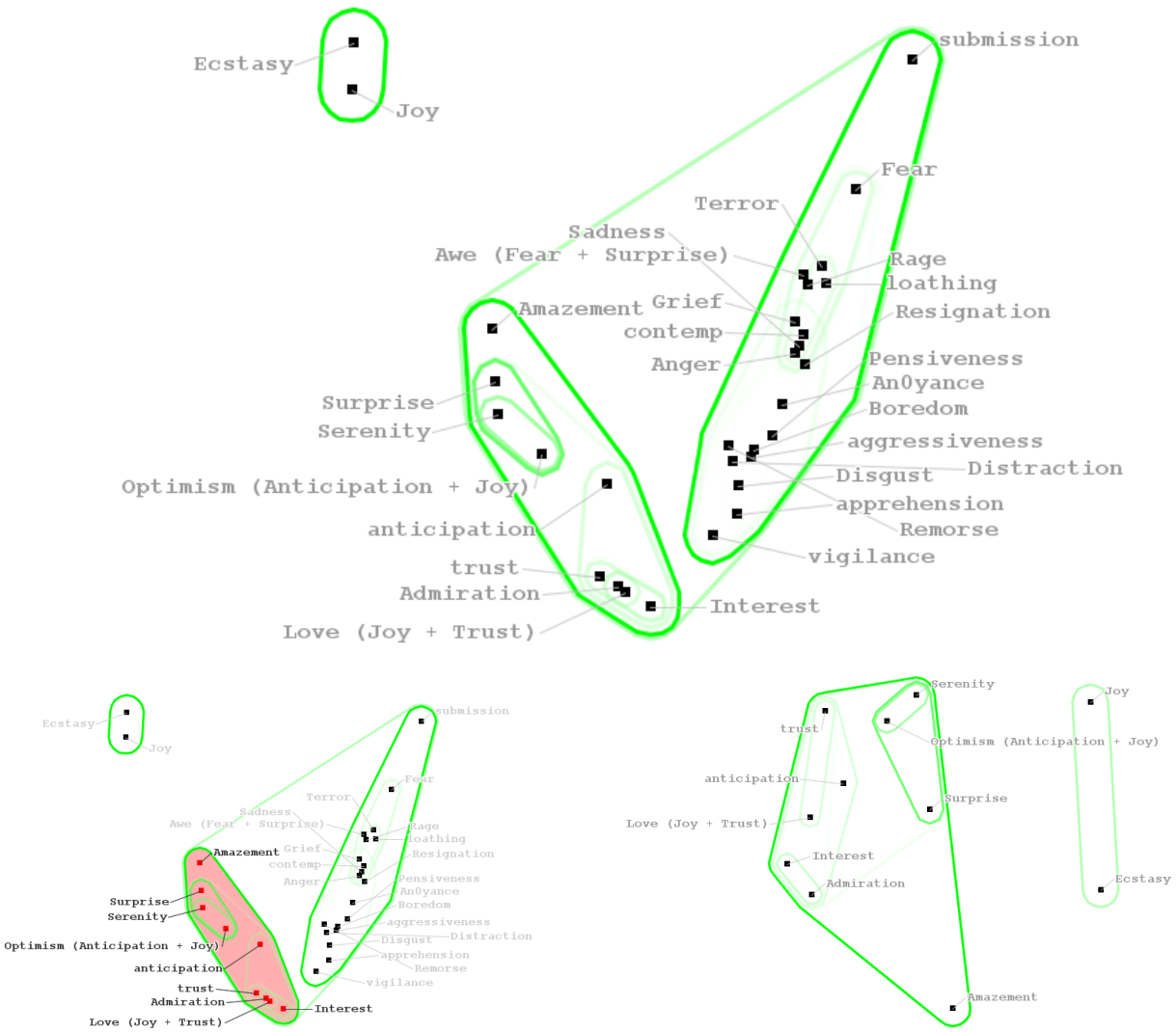

Figure 2. Combined Multidimensional Scaling and Hierarchical Clustering View: overview (upper), highlighting a cluster (lower left) and a filtered cluster (lower right).

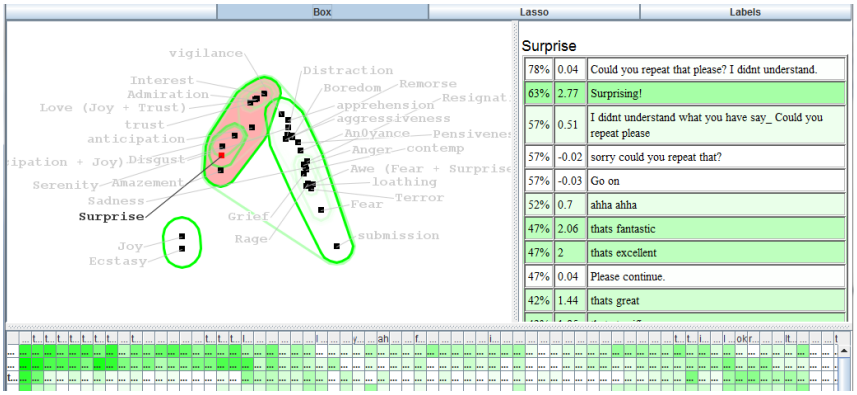

Figure 3: The cluster explorer application interface; multidimensional scaling hierarchical clustering view (upper left), detail view (upper right) and matrix view (lower).

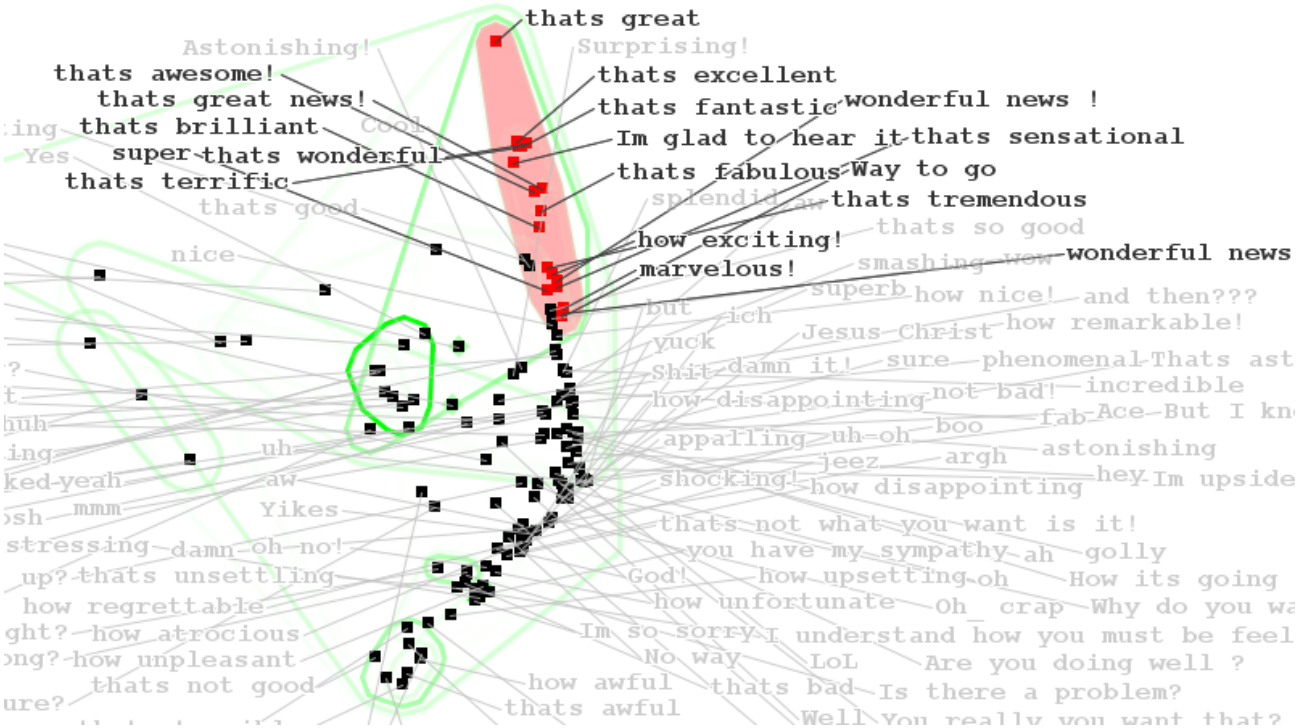

Figure 4: A clustering of phrases according to their appropriateness for different emotions.

The cluster explorer application also allows the user to choose whether columns or rows of their data are used as items or attributes when clustering their data. For example, in figure 2 we see the emotion-term survey data clustered with emotions as items and phrases as attributes, figure 4 shows an alternate method of clustering with phrases as items and emotions as attributes. In this particular view of the data (in figure 4 with positive phrases highlighted) we can see a continuum of positive phrases. Further exploration of the clustered phrases revealed a similar continuum of negative phrases and a small strong natural cluster of affirmative neutral phrases ("ok", "that's ok", "right" etc).

4. EVALUATION

The combined multidimensional scaling hierarchical clustering view was evaluated with two expert users analyzing survey data reporting the appropriateness of phrases to respond to a variety of emotionally charged situations. While these users had not previously viewed the data in other applications they were familiar with a variety of statistical and visualization techniques that could be used for the analysis of this type of data (including multidimensional scaling, hierarchical clustering and various other types of clustering). After a brief training session, lasting less than five minutes, the users where asked to operate the Cluster Explorer application (see section 3.1) in order to try to find interesting patterns in their data. Whilst using the tool the users where encouraged to "think aloud"[24] so that information relating to their use of the tool and its usability could be recorded. Each evaluation session was followed up with an informal interview that allowed us to confirm our understanding of the "think aloud" stage results and gather more un-structured information relating to the general usability and utility of the tool (i.e. general opinions and feelings about the technique).

The results of the evaluation were generally positive. The experts found that they were able to familiarize themselves with the cluster view of the data relatively quickly, after less than two minutes. After this short period of time the users were able to perceive general patterns in the data and navigate, by selecting clusters, to find more subtle patterns. Here, the animation used to smooth the transition between views helped the user keep track of items and patterns between different views without being disorientated by a sudden change in the view. The users also felt that both aspects of the combined clustering view were useful and that it benefited them to have them combined in a single view. The multidimensional scaling layout of items allowed the users to perceive general patterns and outliers in the data and the

hierarchical clustering overlay allowed them to have a better view of natural groupings. Hierarchical clustering also made it easier to draw conclusions since the clusters were derived algorithmically rather than being products of human perception. Combining multidimensional scaling and hierarchical clustering into a single overview was considered to lessen the cognitive overhead that the use of multiple linked views might have incurred.

The users were able to find a number of interesting patterns in their data using the new technique. These included the identification of interesting natural clusters, nested clusters, outliers and general trends. Significant clusters included the three main clusters of negative, positive and strong positive emotions. The positive emotion cluster also contained two nested clusters of progressively stronger positive emotions. Zooming into the negative emotions cluster revealed a number of outlying negative emotions each with their own individual small sets of characteristic appropriate phrases ('that's a shame' for sadness, 'uh-oh' for vigilance, 'that's terrible' for terror, 'how unpleasant' for disgust etc). Clustering phrases showed a continuum of negative through neutral to positive phrases. Using these results the users where able to draw conclusions that were significant with regard to their overall understanding of the data and, ultimately, their understanding of how people react to phrases as responses to emotional dialogue.

While the experts were unsure as to whether the same patterns could or could not be revealed using other techniques, they were convinced that the combined multidimensional scaling hierarchical clustering technique had some major advantages over other techniques. Firstly, they considered that this technique would be easier to use since it allowed them to combine different types of overview in a single view. They believed this would save them time and allow them to reveal types of pattern that they would not necessarily expect to find before they begun their analysis (since several aspects of the data are apparent in the initial overview). Another perceived advantage was that results would be easier to share since a number of different data characteristics could be communicated in a single screenshot. They also felt that the interactive nature of the visualization gave them more freedom to explore different subsets of the data. While other tools allowed them to filter to view different subsets of the data, this process was often found to be cumbersome and destroyed the 'flow' of the discovery process.

5. CONCLUSIONS

We have developed a novel visualization technique that combines multidimensional scaling and hierarchical clustering to support the analysis of multidimensional data. Both of these techniques act to provide us with different types of overview with neither relying on the data having any particular characteristic other than natural groupings. Multidimensional scaling is more flexible allowing us to observe partial clusters and outliers while the fixed clusters of hierarchical clustering make it easier for us to draw conclusions since the clusters are derived algorithmically rather than being products of (fallible) human perception. We evaluated the application by linking it to matrix views and text detail views in an application designed to support the exploratory analysis of term appropriateness survey data. This has allowed us to easily uncover a number of distinct emotion groupings and identify the terms that are appropriate for these emotions as well as the terms that are more uniquely appropriate. While it is likely that we may have eventually been able to find some of these patterns using standard statistical techniques, we feel that the information visualization approach we have used has allowed us to find the patterns faster, allowed us to reveal types of patterns we would not expect to occur and, overall, made the data more accessible. As future work we plan to extend the cluster explorer tool to enable it to work with other types of survey data. This will also allow us to perform a more extensive user evaluation to quantify the tool’s usability and overall utility.

ACKNOWLEDGEMENTS

This work would not be possible if it were not for the kind and generous support of the Teacher Enhancement Program (PROMEP) of the Mexican government Secretaria of Public Education. The authors would also like to acknowledge Mireya Silva Jiménez and David Díaz Pardo de Vera for the help they provided during the evaluation of the software developed.

REFERENCES

[1] B. Shneiderman, “The eyes have it: A task by data type taxonomy for information visualizations,” Proceedings IEEE Visual Languages, 336-343 (1996).
[2] W.-A. Jungmeister, and D. Turo, [Adapting Treemaps to Stock Portfolio Visualization] University of Maryland, (1992).
[3] H. Siirtola, and K.-J. Räihä, “Interacting with parallel coordinates,” Interacting with Computers, 18(6), 1278-1309 (2006).
[4] P. Craig, A. Cannon, R. Kukla *et al.*, [MaTSE: The Microarray Time-Series Explorer] IEEE, Seatle, US(2012).
[5] K. Aas, [Microarray Data Mining: A Survey] Norski Regnesentral, Oslo, Norway(2001).
[6] G. Kerr, H. Ruskin, M. Crane *et al.*, “Techniques for clustering gene expression data,” Computers in Biology and Medicine, 38(3), 283-293 (2008).
[7] J. Quackenbush, “Computational Analysis of Microarray Data,” Nature Reviews Genetics, 2(6), 418- 427 (2001).
[8] M. Kendall, [Multivariate Analysis] Charles Griffin&Co, (1975).
[9] T. Kohonen, “The Self-Organizing Map,” Proceedings of the IEEE, 78(9), 1464-1480 (1990).
[10] P. Tamayo, D. Slonim, J. Mesirov *et al.*, “Interpreting patterns of gene expression with self-organizing maps.,” Proceedings of the National Academy of Sciences of the United States of America, 96(6), 2907-2912 (1999).
[11] J. Luo, D. J. Duggan, Y. Chen *et al.*, “Human Prostate Cancer and Benign Prostatic Hyperplasia: Molecular Dissection by Gene Expression Profiling,” Cancer Research(61), 4683-4688 (2001).
[12] M. Bittner, P. M. X, Y. C. X *et al.*, “Molecular Classification of Cutaneous Malignant Melanoma by Gene Expression Profiling,” Nature, 406(6795), 536-540 (2000).
[13] A. A. Jazaeri, C. J. Yee, C. Sotiriou *et al.*, “Gene expression profiles of BRCA1-linked, BRCA2-linked, and sporadic ovarian cancers,” J Natl Cancer Inst.(94), 990-1000 (2002).
[14] R. Simon, and A. P. Lam, [BRB ArrayTools] National Cancer Institute Biometric Research Branch Division of Cancer Treatment and Diagnosis (linus.nci.nih.gov), (2005).
[15] M. B. Eisen, P. T. Spellman, P. O. Brown *et al.*, “Cluster analysis and display of genome-wide expression patterns.,” Proc. Natl. Acad. Sci. USA, 95(25), 14863-14868 (1998).
[16] A. Sturn, J. Quackenbush, and Z. Trajanoski, “Genesis: cluster analysis of microarray data,” Bioinformatics, 18(1), 207-208 (2000).
[17] A. Buja, J. A. McDonald, J. Michalak *et al.*, "Interactive Data Visualization Using Focusing and Linking." 156-163.
[18] J. C. Roberts, "Multiple-View and Multiform Visualization." 3960, 176-185.
[19] J. C. Roberts, "Issues of Dataflow and View Presentation in Multiple View Visualization." 177-183.
[20] R. Plutchik, and H. Kellerman, [Theories of emotion] Academic Press, (1980).
[21] E. Bertini, M. Rigamonti, and D. Lalanne, “Extended Excentric Labeling,” Computer Graphics Forum, 28(3), 927-934 (2009).
[22] J. Fekete, and C. Plaisant, "Excentric Labeling: Dynamic Neighborhood Labeling for Data Visualization." 512 - 519.
[23] P. Craig, and J. Kennedy, [Concept Relationship Editor: A visual interface to support the assertion of synonymy relationships between taxonomic classifications] Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, San Jose, CA(2008).
[24] C. H. Lewis, [Using the "Thinking Aloud" Method In Cognitive Interface Design], (1982).