A Particle Visualization Framework for Clustering and Anomaly Detection

Ian Davidson and Matthew Ward, {iand, mattward}@sgi.com

MineSet Group, SGI

1300 Crittenden Ln, MS 876

Mountain View, CA, 94043-1351

Abstract

Non-hierarchical clustering is used extensively in data mining for tasks such as segmentation and anomaly detection. Visualizing the output of a clustering exercise can help assimilate the extensive information produced and can provide insights that are not evident from text descriptions. We propose a general framework for visualizing clustering results that can be specialized for specific clustering tasks. We describe our three-dimensional information visualization in which the clustered data points are represented as particles that are affected by gravitational forces. We map the cluster centers into a three-dimensional space while attempting to maintain their respective distances that occur in the instance space. This results in clusters that are similar being adjacent and those that are different being far apart. We then place the particles amongst the centers according to the gravitational effect exerted by the centers. The gravitational effect of the cluster centers on a particle is given by the particle's degree of belonging to the clusters. We show that visualizations based on the general framework convey much of the information that a clustering solution contains. However, in some situations we wish to emphasize specific information that is important for the task at hand. We describe a variation of our framework that is particularly useful for anomaly detection and suggest that other variations maybe useful for different tasks that make use of clustering. The variations consist of changing the visualization parameters such as the gravitational law used.

1 Introduction

Non-hierarchical clustering, also called unsupervised or intrinsic classification, has a long history in numerical taxonomy [1] and machine learning [2]. Clustering attempts to find sub-groups within a set of s data points (objects/observations) each described by d attributes. Clustering involves determining the number of clusters/groups (k), generating a description for each cluster (which can be used to determine membership), and assigning each data point to one or more of these clusters. Soft clustering involves assigning a proportion of the data point (partial assignment) to every cluster, whilst hard clustering assigns the data point to only one cluster (exclusive assignment). In both types of clustering the information provided by a solution is great and visualization of the results can help provide insight into the structure found. For example, how well defined are the clusters, how different are they from each other, what is their size, and do the data points belong strongly to the cluster or only marginally.

Typical problems that clustering addresses include segmentation and anomaly detection. Segmentation involves dividing the data points into regions that are as dissimilar to each other as possible for purposes such as target marketing a particular cluster. Consider a cross-sell data set where the attributes consist of demographic information and a binary attribute that denotes whether that household cross-sold or not. One may attempt to cluster the data points solely on the demographic information and determine how the distribution of the binary cross-sold attribute varies over the clusters. The descriptions of clusters that contain above population average proportions of individuals who cross-sell can be used in target marketing. Anomaly detection attempts to identify those observations that don't "belong" and are interesting and should be investigated. Clustering is used to define normality or typical occurrences in the data set; subsequently some criterion is used to identify those anomalous data points that do not belong strongly to any one cluster. Typical applications are finding flaws in manufactured material [3], detecting money laundering, network intrusion [4], and data cleaning. Each application of clustering places an emphasis on different parts of the solution. For example in segmentation we are particularly interested in how different the clusters are from each other and their size, whilst in anomaly detection whether the data point belongs strongly or only marginally to its most likely cluster is most important.

In this paper we describe a general particle framework that can capture the information that a clustering solution provides. In section two we introduce hard and soft clustering. Our visualization framework is not tied to any particular clustering algorithm; instead it only requires the original data set used in the clustering with k appended columns that contain the degree of belonging that the observation has for each and every cluster. The degrees of belonging must sum to one. The normalized observation posterior or likelihood probabilities obtained from a mixture model fit this requirement, as would the normalized distance measures obtained from K-Means clustering results. We then compare our approach to other recent cluster visualization approaches. Most existing cluster visualization work is quite different from ours as it focuses on hierarchical clustering and using visualization to help form the clustering solution, rather than displaying a solution previously found. We believe that three aspects of our visualization differentiate our work. The placement of cluster centers in a three-dimensional space that reflects their distance in the instance space. The visualization of clustering results where observations are partially assigned to clusters. The principled placement of the data points as particles between the cluster centers to reflect the gravitational pull (as given by the degrees of belonging) the cluster centers exert on the particles. Our data point placement strategy is computationally very efficient and allows displaying clustering results of thousands of records with wait times of seconds.

In the subsequent sections we describe our general particle visualization framework in detail and then a specialization for anomaly detection. Our general framework defines how to place the cluster centers in a three-dimensional space and how to place the data points amongst the cluster centers according to an attractive gravitational law. We illustrate that different gravitational laws (where mass and distance affect the degree of force) can be created that emphasize different aspects of the visualization. We construct a special gravitational law suitable for anomaly detection. We illustrate the anomaly detection visualization on the UCI [15] churn data set that contains information on the calling statistics of customers. Each record corresponds to one customer and includes information such as their calling plan, total number of minutes spent making international, evening and day time calls. We conclude by summarizing our approach and describing potential extensions to our framework.

2 Clustering Algorithms

The Expectation Maximization (EM) algorithm [5] and the K-Means clustering algorithm [6] are two popular techniques for searching the model space. Both attempt to find the best model, though it is well known that the definition of "best" varies between the two. Typically EM is associated with soft clustering and K-Means with hard clustering.

The EM algorithm in the classical inference setting attempts to find the model with maximum likelihood estimate (MLE) or in a Bayesian setting the model with the maximum a posteriori (MAP) estimate. The K-Means algorithm aims to find the model with the minimum distortion within each cluster for all clusters. Both algorithms consist of two fundamental steps:

1) *The Observation assignment* step, where the observations are assigned to clusters based on cluster descriptions.

2) *The Cluster re-estimation* step, where the cluster descriptions are recalculated from observations assigned to them.

The two steps are repeated until convergence to a point estimator is achieved.

In the first step of the K-Means algorithm the observations are assigned *exclusively* to the "closest" cluster as defined by some distance metric. Euclidean distance is often used. In the EM algorithm an observation is assigned *partially* to each cluster, the portion of the observation assigned depending on how probable (or likely) it is that the cluster generated the observation.

In the second step both algorithms use the attribute values of the observations assigned to a cluster to recalculate cluster descriptions. As K-Means uses exclusive assignment we recompute the estimates from only those observations that are assigned to the cluster. However, in the EM algorithm the cluster descriptions are calculated from every observation, weighted by the proportion of the observation assigned to the cluster.

After each algorithm has converged to its respective local optima, EM produces for each observation a probability of belonging to each and every cluster. K-Means produces a set partition of the observations but a degree of membership for an observation that is analogous to probabilities can be produced by considering the distance to an individual cluster and the sum of the distances to each and every cluster.

The input to our visualization of clustering results is the original data set with *k* appended columns that contain the observations degree of membership to each cluster which must sum to one. In all our results presented in this discourse we use EM clustering in a maximum likelihood context.

3 History of Visualization of Clustering Results

In this section we discuss relevant work in the field of visualizing clustering results. We begin by highlighting what differenciates our work and then summarize relevant previous work. Much of the work in the cluster visualization field has been for visualizing *hierarchical* (often-called agglomerative) clustering [7]. Hierarchical clustering typically begins with all of the data-points placed on a line, at the same level. Then the very similar data points are combined/fused pair-wise to obtain the next level and so forth until some stopping criterion is reached. This leads to a tree like structure as shown in Figure 1 that is sometimes referred to as a clustering tree [11].

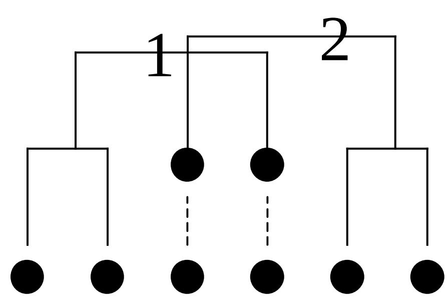


Figure 1: A functional representation of the hierarchical clustering of six data points into two clusters.

Hierarchical clustering is typically bottom-up clustering using exclusive assignment. Since our visualization is for clustering where partial assignments are calculated, it is not directly comparable to previous hierarchical clustering visualization unless fractional assignments were somehow calculated and used in the visualization, which to our knowledge has not been attempted. However, several pieces of prior work in hierarchical cluster visualization use the idea of placing observations around the cluster centers at a distance equal to the degree of belonging. What differentiates our work from previous clustering visualization work is:

a) We attempt to visualize an already established clustering solution, rather than providing decision support to help in forming the clustering solution as others have [11].

b) Our visualization is specifically for non-hierarchical clustering with partial assignment and therefore we need to place our observations subject to the constraint of multiple degrees of belonging not just one.

c) We place our cluster centers on the canvas so that similar clusters are adjacent and different clusters are far away

d) Our principled use of attractive laws (gravitational in our case, though others could be used) to model the placement of observations around the cluster centers.

e) The scalability of our approach to visualize large data sets.

We now survey the related work, which we divide into interactive and non-interactive. In one example of interactive cluster visualization, the author [8] uses a scatterplot view of cluster centers that are obtained by hierarchical clustering, with interactive control of the splitting criteria to increase or decrease the number of visible clusters. Clusters can be selected and their contents viewed in alternate visualizations (parallel coordinates, scatterplot matrices, and box plots) to assess the distribution of cluster members.

In another example of interactive visualization that employs parallel coordinates, the authors [9] present a technique for visual presentation and interactive navigation through data that has been hierarchically clustered. Each cluster is depicted by its center, extents, and population. In parallel coordinates, there is one vertical axis for each dimension, and a data point is represented by a polyline crossing each axis at a position proportional to its value in that dimension. The authors augment this display by mapping the cluster extents for each dimension into a colored band surrounding the center line, with opacity at the center proportional to the cluster population and tapering off to zero at the boundaries. These extent bands can be scaled to reduce overlap. Color is used to depict the location within the hierarchy of each cluster, so that sibling clusters will have similar color. A navigational tool, called a structure-based brush, allows users to drill-down or roll-up at selective locations in the hierarchical structure. The authors have since extended this work to incorporate other multivariate visualization techniques besides parallel coordinates.

There has been some work in non-hierarchical hard clustering visualization in the document clustering field [10]. A document set is clustered, and each cluster is assigned a fixed size region on the display. Documents, represented as points, are placed at a distance proportional to their similarity to their cluster center. A cosine rule is used to avoid overlaps between two points with equal distance to their center. Topical words from each cluster are used as labels. Though we have not presented it in this paper, we use the notion of sign-posting clusters with their names that are derived from what differentiates the cluster from the population (or even other clusters).

In work that uses some similar techniques to our own, Sprenger et. al. [11] use a spring embedding system to position *data* relative to their similarities. They then use ellipsoidal gaussian fields (blobs) to define spaces of influence for each point, and these fields are combined to form clusters. Implicit surfaces, a mechanism for defining an isosurface in a continuous field, are used to visualize the clusters, with the user controlling the level of interest in the cluster hierarchy and the isovalue used to generate the surface. Thus a cluster may have a complex shape, depending on the distribution of its membership.

Leuski and Allan [12] propose value adding to the results of an information retrieval query by placing the documents (represented as spheres) in two or three-dimensional space according to their degree of similarity. This is in principle similar to the way we place cluster centers in a three dimensional space. The idea could be applied to all observations in the data set, however this would ignore the results found by the clustering algorithms and be computationally very expensive.

4 Clustering Visualized as Particles Affected by Gravitational Forces

Clustering is inherently density and mass estimation in an instance space. Consider the two-dimensional (i.e. two variables) clustering problem shown in Figure 2. Each observation occupies a point in the two dimensional instance space. The general aim of clustering is to find sub-regions of the instance space where many observations occur. How these sub-regions are described can vary. In K-Means clustering they are described solely by the single central point of the cluster, whilst in EM mixture modeling the sub-region is defined by a complete parameterized probabilistic model that can be considered a generation mechanism for the observations in that cluster. However, no matter how the sub-regions are described, they are selected because the concentrations of observations within them are higher than normal. In short, the clusters represent sub-regions where dense collections of observations occur.

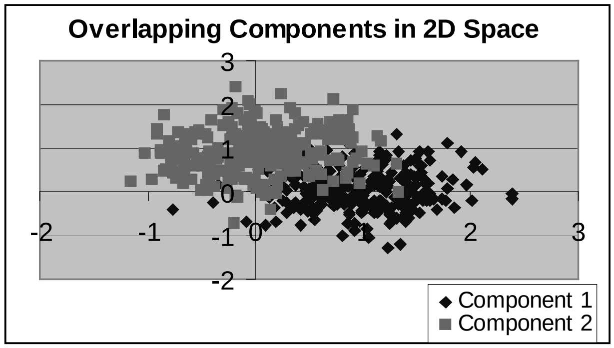


Figure 2: 500 Observations from two clusters. Component/Cluster 1 (Gaussian (1,0.5), Gaussian(0,0.5)), Component/Cluster 2 (Gaussian (0,0.5), Gaussian (1,0.5))

Density is the property of how often observations occur in a unit of space. Mass is a measurement of the total number of observations in some region. The clusters represent regions in the instance space where the mass of observations is great. Newton's law of gravity states that any particle of mass attracts any other with a force varying directly as the product of the masses and inversely as the square of the distance between them. If we consider the observations as being particles with mass then a natural graphical view of clustering is as particles being pushed and pulled by the gravitational forces exerted on each other. We wish our visualization to be a snapshot of the particles as they would be positioned if gravitational attractive forces were applied to them. One could attempt to model simultaneously the effect of the particles on each other, but this would be a computationally extensive problem**.** Instead, we can make the problem tractable by iterating over each particle and consolidating the mass of all the remaining particles into the cluster centers, which then exert a force on the particle.

We will now describe in principle the general ideas behind our particle visualization of clustering results. The various options that can change the visualization significantly are then described. Our canvas will be a three-dimensional space, though a two dimensional space would suffice for problems with a small number of clusters and observations.

We first map the concentrated areas of mass (the cluster centers) onto the canvas. Because the observations may consist of many columns (say m) beyond three dimensions we must apply a dimension reduction technique. We do this by attempting to place the cluster centers in the three dimensional space so as to preserve the distance spacing that occurs in the original m dimensional space. We achieve this by measuring the Kullback-Leibler distances between the cluster centers and using Multi-Dimension Scaling (MDS) [13] to place the k cluster centers in a cube whose diagonal length is equal to 1. The clusters that are close in the original space are adjacent and those that are different are far apart. Functionally MDS takes as input a k by k matrix (D_{Matrix}) that contains normalized distances (that sum to one) between k

points. MDS attempts to create a layout of the points in the cube so that distances between points (as measured in the cube) are close to the matrix distances (D_{Matrix}). This is achieved by a combinatorial optimization technique where the *k* points are initially randomly placed and then moved until a good solution is found. From the positions of the points in the cube a distance matrix can be calculated which we will call (D_{Cube}). The objective function that the optimization task is trying to minimize is $|D_{Matrix} - D_{Cube}|$. However, this may be a complicated function with many local minima, hence the MDS approach attempts to find a good local optimum. We use the Kullback-Leibler (KL) distance as a distance measure between clusters to generate D_{Matrix}. As the KL distance is not symmetrical, when we refer to the KL distance between say *a* and *b* we will be considering the mean value: (KLD(*a*,*b*) + KLD(*b*,*a*)) / 2. We use a simulated annealing [14] approach with multiple random restarts to find a good local optimum.

Next we need to place the observations around the cluster center so as to reflect the gravitational pull on the particles by the cluster centers. Each particle belongs to every cluster with some degree of membership. To place the particle around the centers we adopt the general approach of placing the particle at intermediate positions around each and every cluster center as in Figure 3. Then we place the particle in its final position by averaging the coordinates of the intermediate positions.

Figure 4 illustrates how we place the particles in their intermediate positions. It should be remembered that distance and degree of membership are inversely related. Therefore, if the degree of membership is high, then as the distance is proportional to 1 - the degree of membership, the distance is small. Firstly, we place the observation near cluster *A* at a distance proportional to $(1 - P(D \mid \theta_A))$. This provides a loci of points (a circle) at which the observation may be placed. To determine which point on the circle it occupies, we consider the "pull" the other cluster centers have on the observation. The overall pull on the observation is achieved by considering that the other cluster centers exert a force on the observation. The magnitude of an individual cluster's force on an observation is equal to how likely the observation belongs to that cluster. The direction of the force is given by a straight line between the center of the cluster *A* and the other cluster's center.

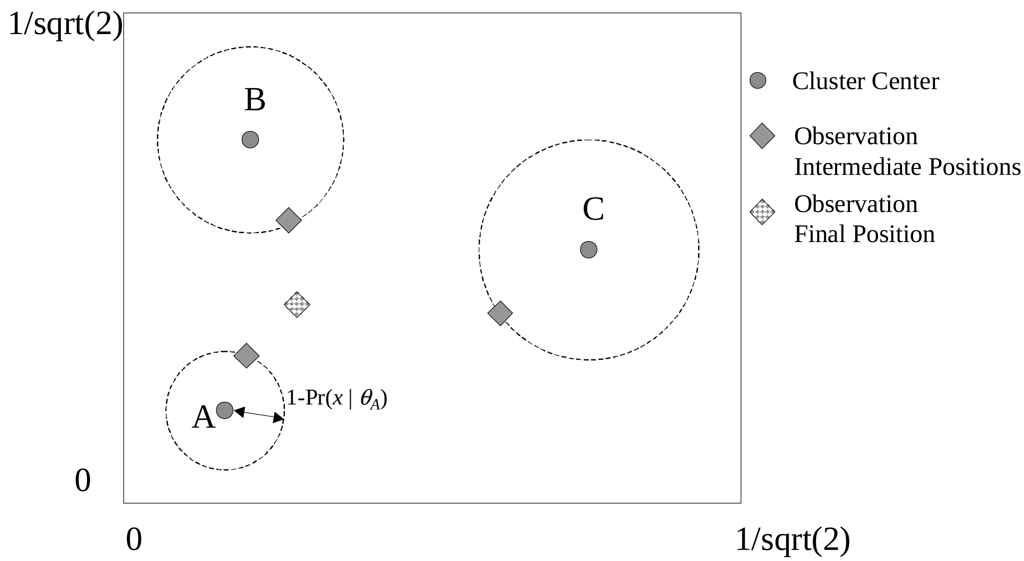


Figure 3: Placement of an observation, x, in two-dimensional space (for clarity), where $\Pr(x \mid \theta_A) > \Pr(x \mid \theta_B) > \Pr(x \mid \theta_C)$. The final positioning of x is given by averaging the coordinates of the intermediate positions. See Figure 4 to determine how the particles are placed in their intermediate positions.

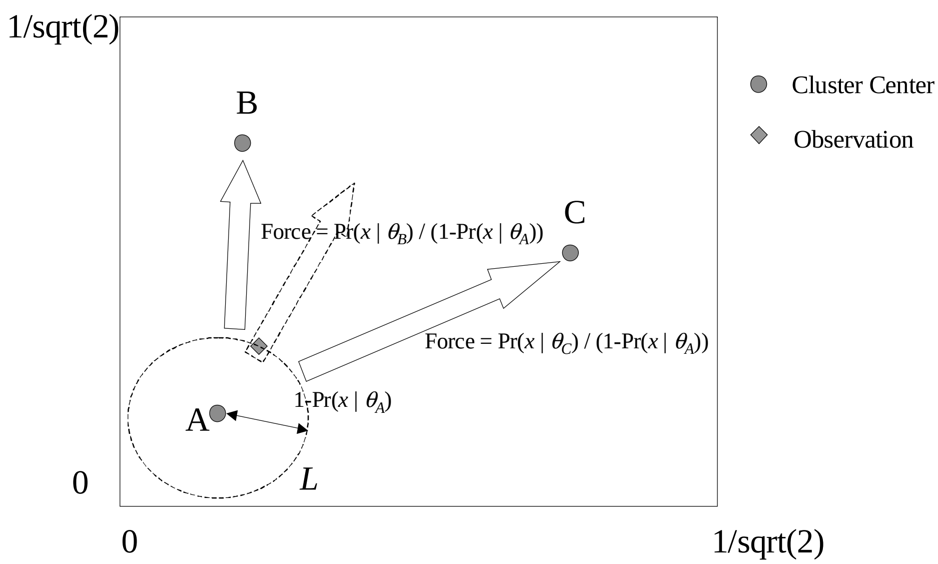


Figure 4: How an observation is placed around a cluster center where force is directly proportional to degree of membership. Placement of an observation, x, in two-dimensional space (for clarity), where $\Pr(x \mid \theta_A) > \Pr(x \mid \theta_B) > \Pr(x \mid \theta_C)$. The positioning of x in the locus, L is given by the direction of the resultant vector. The resultant vector's direction is the weighted (by the observation's likelihoods) sum of the component vectors. The direction of the component vectors is given by the direction between the two cluster centers.

The above is the general approach to our particle visualization. Place the cluster (density) centers in a three-dimensional space to best preserve the spacing in the original *m* dimensional space. Next place the observations (particles) around the centers so as to reflect the gravitational pull on the particles as represented by the degree of belonging that a particle has for each cluster. The results of applying this general framework are shown in Figure 5. Note that the centers of the clusters are dense regions of observations, that belong strongly to that cluster, whilst the outer edges of the clusters represent particles that do not belong strongly to any one particular cluster. Throughout this paper we demonstrate our clustering results using the UCI churn data set [15] with the *state, churned* and *telephone area code* variables removed. This data set consists of 5000 customers (records) calling behavior described by information such as their voice mail plan, call lengths and usage patterns. Typically it takes under ten seconds to generate the visualization of this data set on a Pentium III 500 MHz machine.

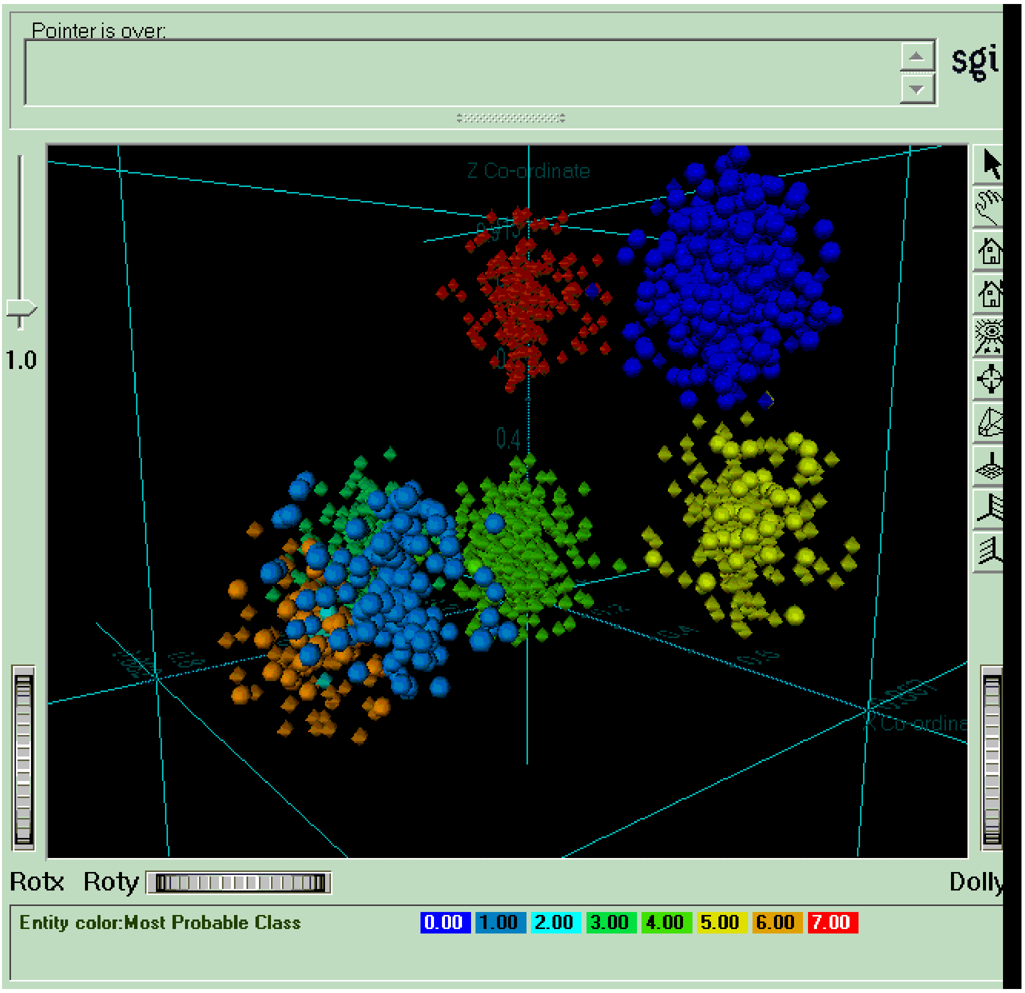


Figure 5: Particle visualization (using the general framework) of the UCI Churn data set with eight color-coded clusters. In the law of gravitational effect being used, force is directly proportional to mass and inversely proportional to the distance.

There are many available parameters that can influence the visualization. We shall discuss some of these parameters and their effects. The gravitational law that affects the particles is a key parameter. Newton's law of gravity is only one law of attraction; other laws can emphasize more or less the role of distance and mass. Newton's law states that a particle of mass attracts any other with a force varying *directly* as the *product* of the masses and *inversely* as the *square* of the distance between them. In our general framework we have implemented a gravitational law where the force is inversely proportional to the *linear* distance. This results in a more even distribution of particles over the canvas. An inverse law (like Newton's) would result in tighter clumps of observations. Newton's law could easily be achieved by using the intermediate particle placement scheme shown in Figure 6; the results can be seen in Figure 7.

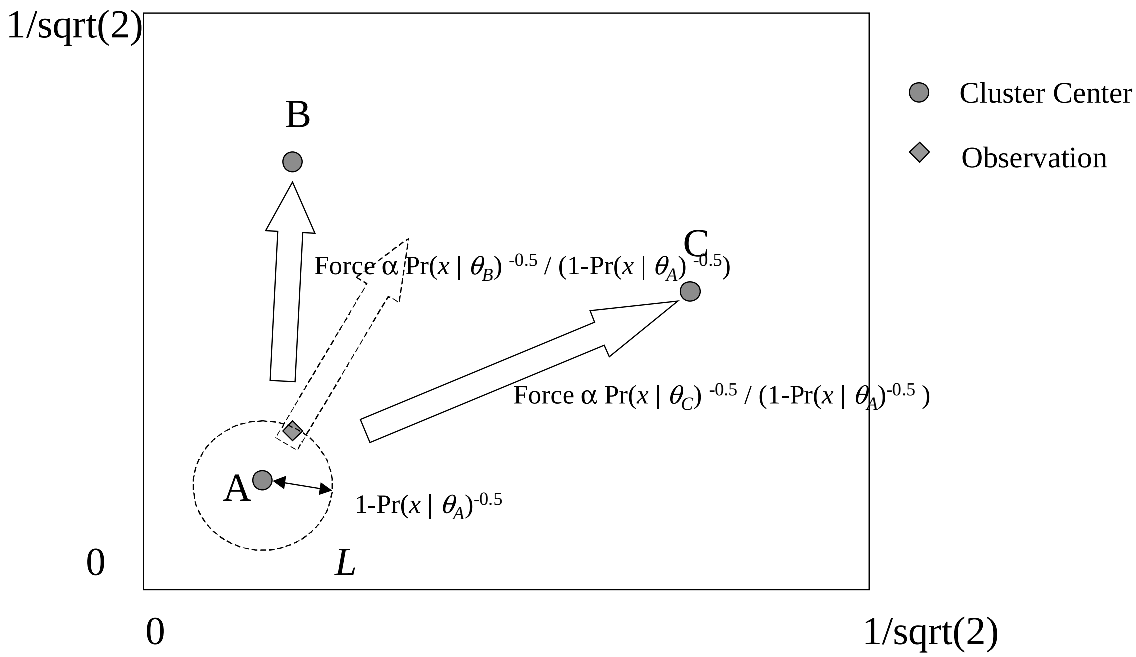


Figure 6: How an observation is placed around a cluster center where force is proportional to the degree of membership squared. Placement of an observation, x, in two-dimensional space (for clarity), where $\Pr(x \mid \theta_A) > \Pr(x \mid \theta_B) > \Pr(x \mid \theta_C)$. The force calculations are now for an attractive law that is inversely proportional to the distance squared.

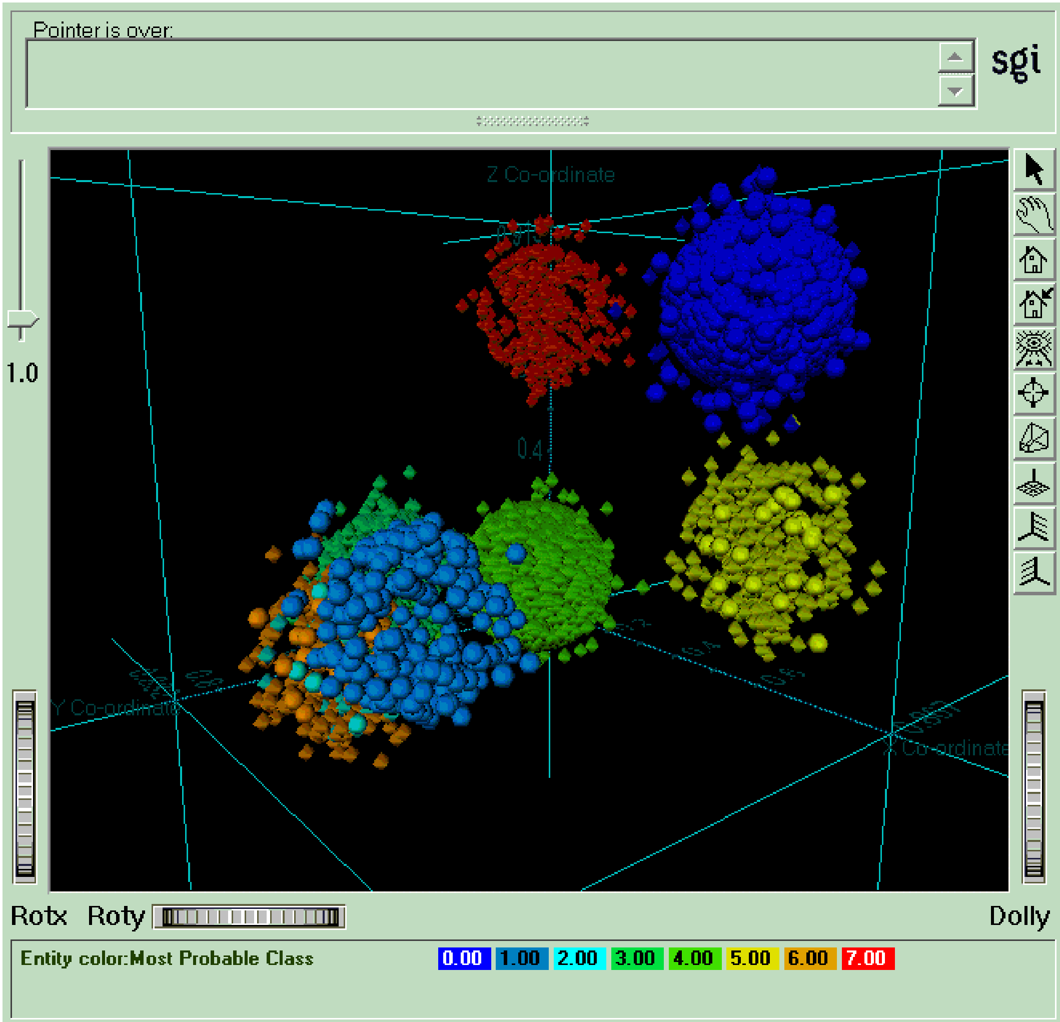


Figure 7: Particle visualization (using the general framework) of the UCI Churn data set with eight color-coded clusters. In the law of gravitational effect used force is inversely proportional to distance squared and directly proportional to mass. Note that the clusters are now more densely packed than in Figure 5.

Similarly in our general framework we choose to implement a gravitational law where mass does affect the size of the force. For EM mixture modeling this naturally occurs as the radius of placement for a particle around cluster A is $\Pr(x \mid \theta_A)$, and $\Pr(x \mid \theta_A)$ is the product of the relative weight of the cluster and the probability of the particle given the cluster's parameters. This results in a more compact placement of observations around larger clusters and sparser placement around smaller clusters, which is an efficient use of the canvas. By using weighted K-Means [16] clustering a similar affect can be achieved.

5 A Variant for Anomaly Detection

5.1 Introduction to Anomaly Detection

"What does the data tell us?", is the general question that data mining, machine learning and statistical analysis attempt to answer. More specific questions involve determining what can we *predict* from the data and how can we *summarize* and *generalize* the data. Anomaly detection asks questions with a different aim. Given a set of data we wish to ascertain *what* observations don't "belong" and *which* are interesting and should be investigated. Some researchers have postulated that anomaly detection is a separate class of knowledge discovery task along with dependency detection, class identification and class description [17].

Anomaly detection has been used in many different contexts: detection of unusual images from still surveillance images [18], identifying unusual organic compounds [19], data cleaning, [20] and identifying flaws in manufactured materials [21]. In most applications the basic steps remain the same:

1) Identify normality by calculating some "signature" of the data.

2) Determine some metric to calculate an observation's degree of deviation from the signature.

3) Set some criteria/threshold which, if exceeded by an observation's metric measurement, means the observation is anomalous.

The signature of the data consists of identifying regularities in the data. In clustering-based anomaly detection the signature is the clusters found in the data. The criteria to determine if an observation is anomalous is typically application specific. In some domains observations where only one variable deviates from the signature are called anomalous, whilst in others a systematic deviation across all variables is tolerable. A typical criterion in clustering-based anomaly detection is the *minimum degree of membership*. That is, if an observation does not belong to any one cluster with a degree greater than some threshold, then it is deemed anomalous. The metric used to measure the degree of anomalousness is specific to the problem formulation (modeling technique). In a probabilistic problem formulation it may be the maximum likelihood or posterior probability of an observation, or in K-Means clustering some function of the distance between the observations and the remaining observations or the distance between an observation and the cluster centers.

5.2 Visualization of Anomalies

Our aim is to convey information that aids the user in exploring anomalous observations. To do this we need to convey what observations are anomalous and why. We explain our variation of the general framework that achieves this.

The first step to placing the cluster centers in the three-dimensional space is identical to our general framework. However, for the particle placement we adopt a specialization of the idea in our general framework. Rather than placing the particles in intermediate positions around each and every cluster center, we make the final position of the particle the position the particle takes around its most likely/probable cluster. This accentuates the degree of belonginess of each particle to a cluster. This has the effect of severely reducing the secondary gravitational effects on the particle. We introduce another specialization, which we call the radius of gravitational effect. The radius of gravitational effect places a sphere around the cluster centers. Those observations that fall within the sphere are not affected by any other cluster's gravitational pull, whilst those that fall outside the sphere are affected. This has the desirable effect of clearly identifying those observations that belong very strongly to a cluster. Usually the radius of gravitational effect is equal to (1- *minimum degree of membership*), so that extra-cluster gravitational forces only affect those observations that are anomalous. Observations that fall inside the sphere are randomly placed on the surface of a sphere whose radius is proportional to their degree of membership. Figure 8 shows our visualization of anomalies for the UCI churn data set. In all there are eight clusters, each can be saliently described by what differentiates the cluster statistics from the population statistics as shown in Table 1.

Cluster ID	Cluster Description	Size
1	voice mail plan:Different & number voicemail messages:Very High & total evening minutes:Low & total evening charge:Low & total international minutes:High & total international charge:High	405
2	voice mail plan:Different & number voicemail messages:Very High & total international minutes:High & total international charge:High	244
3	voice mail plan:Different & number voicemail messages:Very High & total international minutes:Very High & total international charge:Very High	355
4	voice mail plan:Different & number voicemail messages:Very High	305
5	total international minutes:Very Low & total international calls:Low & total international charge:Very Low	38
6	Very similar to population statistics	2259
7	total evening minutes:Very High & total evening charge:Very High & total night minutes:High & total night charge:High	468
8	total evening minutes: High & total evening charge: High & total night minutes:High & total night charge:High	926

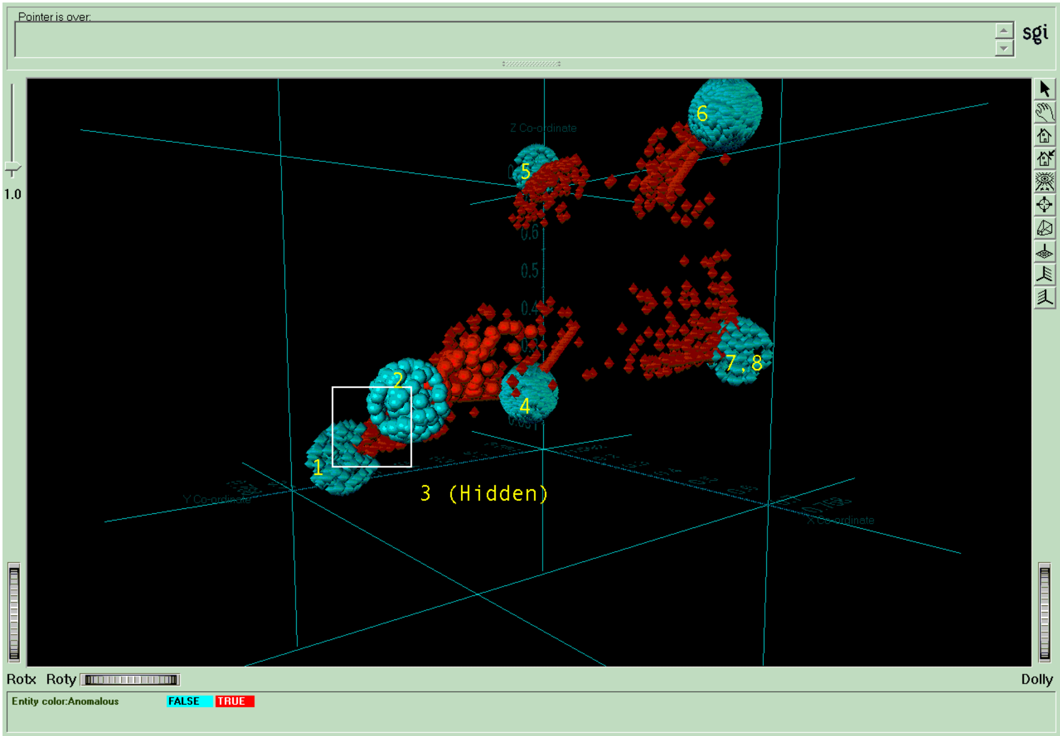
Table 1: A description of what differentiates each cluster found in the UCI Churn data set from the population. A cluster's column statistics are deemed to be significantly different from the population statistics if its average Kullback-Leibler distance is greater than 0.5. Significantly different nominal attributes are only described as different, whilst continuous variables whose means differ by more than one population standard deviation are termed High/Low and those more than two population standard deviations away are termed Very High/Very Low. The clusters are listed in the same order as their left to right appearance order in Figure 8. Cluster numbers are included for reference between the table and figure.

The clusters that are close to each other are more similar than those clusters far apart. Those observations that belong strongly to a cluster occur towards its central region. Anomalies (that do not belong strongly to any one cluster) tend to be *between* the cluster centers. The farther away an observation is from any of the cluster centers, the more anomalous it is. Observations in the central region are the most anomalous because they are being pulled towards all clusters. A patent application is being submitted on specific details of the visualization.

From Table 1, it can be seen that clusters one, two and three are very similar and are hence placed very close to each other in Figure 8. Cluster four is similar to this trio but does not have similar call usage patterns which is why it is not closer to the trio. Clusters seven and eight are very similar and from the perspective of our snap shot they appear to be one cluster. The bulk of the observations are in cluster six which has statistics very similar to the population. It is placed in a position that is far from the remaining clusters.

Many anomalies lie between clusters one, two and three, as can be seen in Figure 8b. This trio of clusters represents individuals with voice plans that are different from the population and have a high number of voice mail messages. Each cluster specializes the generic profile by different international call usage. The anomalies in this region are those observations that meet this general profile but also have high usage of other types of calls such as daytime calls.

It can be seen from Figure 8 that there are different types of anomalies. Consider the top right hand corner of Figure 8a. Most of the anomalies in this region belong fairly strongly to one cluster but are anomalous because they could belong to one other cluster. We can determine this other cluster by the cluster center the anomalous observations are "pointing" towards. This type of anomaly is different from the type contained in the center of the figure, which are essentially those observations that belong rather weakly to many clusters. If we were to measure the entropy (degree of disorder) in the likelihoods of the observations, we would find that the latter type of anomalies has a higher entropy that the first type.



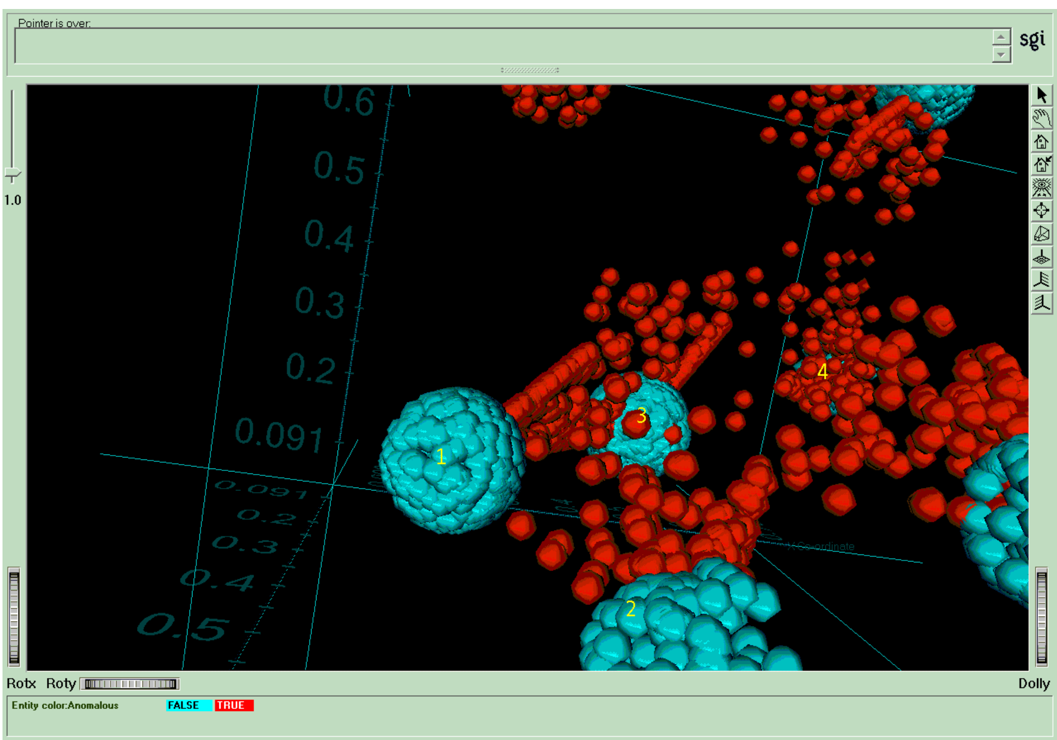


Figure 8a and b): Visualization of anomalies generated from the UCI Churn data. Figure a) shows anomalies versus normal observations with anomalies colored darker. The subsection highlighted in the a) is blown up in b). Clusters 7 and 8 are extremely close together, and from the perspective of the snap shot provided effectively look like one cluster.

6 Conclusion and Further work

We have presented a general framework that enables the visualization of clustering results where the observations that are clustered are represented as particles affected by gravitational forces. To our knowledge, visualizing clustering results as particles affected by gravitational forces is unique and naturally fits well with the non-hierarchical clustering philosophy. In our framework, we place the cluster centers in a three-dimensional space such that similar clusters are adjacent to each other and different clusters are far apart. We then place the observations amongst these centers to reflect the degree of belonging that each observation has for the clusters. This is a very computationally efficient approach and we were able to perform the calculations for a 5000 observation data set in less than ten seconds on a desktop machine. Our information visualization can present a lot of information and we propose a specialization for anomaly detection that focuses on information useful for that purpose. We believe that other specializations of our general framework can be created for other uses of clustering, such as segmentation.

A novel and natural extension to our ideas would be to incorporate motion into the visualization. This could convey further information depending on the motion of the particles around the cluster centers. Our idea is to essentially model the particles as being affected by the gravitational laws of attraction, one could modify this idea to follow other attraction/repulsion laws such as those found in the fields of electricity and magnetism. Finally, we plan to conduct a user study to help in suggesting improvements particularly for the anomaly detection visualization.

7 Acknowledgements

Thanks for Marc Hansen and Jose Renteria of the MineSet Visualization team for their thoughts on this work and comments on the paper.

8 References

[1] G. Dunn, B.S.Everitt, *An Introduction to Mathematical Taxonomy*. Cambridge University Press 1982.

[2] R.S. Michalski, R. Stepp, ”Learning from Observation: Conceptual Clustering,” in R.S. Michalski, J.G. Carbonell, T.M. Mitchell editors, *Machine Learning: An Artificial Intelligence Approach*, CA: Morgan Kaufmann, 1983.

[3] J. G. Campbell, C. Fraley, F. Murtagh and A. E. Raftery, *Linear Flaw Detection in Woven Textiles using Model-Based Clustering*, Pattern Recognition Letters: 18(1997):1539-1548

[4] T. Fawcett, and F. Provost, Activity Monitoring: Noticing Interesting Changes in Behavior, in KDD'99, ACM Press, 1999.

[5] A.P. Dempster, et al, Maximum Likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B*, Vol 39 pages 1-39, 1977.

[6] J. MacQueen, Some Methods for classification and analysis of multivariate observations, Proceedings of the Fifty Berkeley Symposium on Mathematics, Statistics and Probability, volume 1, pages 281-296, 1967.

[7] T. Clifford, W. Stephenson, An introduction to numerical classification, Academic Press (1975)

[8] G. J. Wills, "An interactive view for hierarchical clustering", Proceeding of Information Visualization '98, pp. 26 – 31, 1998.

[9] Y. Fua, M. Ward, and E. Rundensteiner, ``Hierarchical parallel coordinates for exploration of large datasets,'' Proc. Visualization '99, pp. 43-50, 1999.

[10] M. Song, ``BiblioMapper: a cluster-based information visualization technique,'' Proc. Information Visualization '98, pp. 130 - 136, 1998.

[11] T. Sprenger, R. Brunella, and M. Gross, ``H-BLOB: a hierarchical visual clustering method using implicit surfaces,'' Proc. Visualization 2000, pp. 61-68.

[12] A. Leuski, J. Allen, "Lighthouse: Showing the Way to Relevant Information", Proceedings of Information Visualization 2000, pages 125-129, 2000.

[13] C. Bentley and M. Ward, "Animating Multidimensional scaling to visualize N-Dimensional data sets", Proceedings of Information Visualization 1996, pages 72-73, 1996.

[14] L. Van Laarhoven, Theoretical and Computational Aspects of Simulated Annealing, CWI Tract 1988

[15] C.J. Merz and P.M. Murphy, UCI Repository of machine learning databases [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Department of Information and Computer Science, 1998.

[16] M. Kearns, Y. Mansour, A. Ng, "An information-theoretic analysis of hard and soft assignment methods for clustering," Proc. 13th Conference on Uncertainty in Artificial Intelligence, 1997.

[17] E. Knorr, and R.Ng, “Finding Intensional Knowledge of Distance-Based Outliers”, *Proceedings of the 25th VLDB Conference, Edinburgh, Scotland, 1999*.

[18] L. Lee, R. Romano, and G. Stein. "Monitoring Activities from Multiple Video Streams: Establishing a Common Coordinate Frame.", to appear in IEEE PAMI Special Section on Video Surveillance and Monitoring (2000).

[19] G. C. Osbourn, J. W. Bartholomew, A. J. Ricco, and G. C., “A New Approach to Sensor Array Analysis Applied to Volatile Organic Compound Detection: Visual Empirical Region of Influence (VERI) Pattern Recognition”, http://www.sandia.gov/imrl/XVisionScience/Xchempap.htm

[20] B.D. Wright, S. Rebrik and K.D. Miller, Spike-sorting of Tetrode Recordings in Cat LGN and Visual Cortex: Cross-channel Correlations, Thresholds, and Automatic Methods, Society for Neuroscience 28th Annual Meeting November 7-12, 1998, Los Angeles

[21] J. G. Campbell, C. Fraley, F. Murtagh and A. E. Raftery, *Linear Flaw Detection in Woven Textiles using Model-Based Clustering*, Pattern Recognition Letters: 18(1997):1539-1548