Single Cluster Visualization to Optimize Air Traffic Management

Frank Rehm[1], Frank Klawonn[2], and Rudolf Kruse[3]

[1] German Aerospace Center
Lilienthalplatz 7
38108 Braunschweig, Germany
frank.rehm@dlr.de
[2] University of Applied Sciences Braunschweig/Wolfenbüttel
Salzdahlumer Strasse 46/48
38302 Wolfenbüttel, Germany
f.klawonn@fh-wolfenbuettel.de
[3] Otto-von-Guericke-University of Magdeburg
Universitätsplatz 2
39106 Magdeburg, Germany
kruse@iws.cs.uni-magdeburg.de

Abstract. In this paper we present an application of single cluster visualization (SCV) a technique to visualize single clusters of high-dimensional data. This method maps a single cluster to the plane trying to preserve the relative distances of feature vectors to the corresponding prototype vector. Thus, fuzzy clustering results representing relative distances in the form of a partition matrix as well as hard clustering partitions can be visualized with this technique. The resulting two-dimensional scatter plot illustrates the compactness of a certain cluster and the need of additional prototypes as well. In this work, we will demonstrate the visualization method on a practical application.

1 Introduction

Evaluation of clustering partitions turned out to be challenging. Common prototype-based clustering algorithms minimize an objective function (Bezdek (1981)). These methods always fit the clusters to the data, even if the cluster structure is not adequate for the problem. Evaluating a partition by the value of the objective function is thusly not meaningful. To cope with this problem many validity measures are developed to analyze clustering partitions (Davies and Bouldin (1979), Dunn (1974), Höppner et al. (1999), Rubens (1992), Windham (1981)).

Unfortunately, many of the measures condense the whole issue to a single value which is associated with certain loss of information. Aggravating the situation, different measures provide contradictory results for the same partition. In the recent years visual techniques are developed to enlighten encoded information in the partition matrix, a matrix that holds membership degrees of feature vectors to prototype vectors, or to visualize clusters on low-dimensional mappings (Abonyi and Babuska (2004), Hathaway and Bezdek (2003), Huband et al. (2005), Klawonn et al. (2003)).

In this paper we apply single cluster visualization (SCV), a recently proposed method (Rehm et al. (2006)), on a practical example. SCV is used to visualize partitions of weather data that will be needed to predict flight durations at Frankfurt Airport. The rest of the paper is organized as follows. In section 2, we will briefly describe fuzzy clustering as a common representative for prototype-based clustering. Section 3 recalls the visualization technique. Results on the practival data will be given in section 4. Finally we conclude with section 5.

2 Clustering

Clustering techniques aim at finding a suitable partition for a given data set. Prototype-based clustering methods, like k-means (for crips clustering) or fuzzy c-means (for fuzzy clustering), represent clusters by means of centre (or prototype) vectors. A partition matrix U describes a partitioning extensively holding for every feature vector $\mathbf{x}_j$ a single membership degree u_{ij} to a certain prototype vector $\mathbf{v}_i$. For crisp clustering techniques a partition matrix can be easily computed after the partitioning process. Fuzzy clustering algorithms provide membership degrees directly as one part of the clustering result. In this section, we briefly describe fuzzy c-means as a common representative for fuzzy clustering.

Fuzzy c-means aims at minimizing an objective function J that describes the sum of weighted distances d_{ij} between c prototypes vectors $\mathbf{v}_i$ and n feature vectors $\mathbf{x}_j$ of the data set $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ in the feature space $\mathbb{R}^p$

$$J = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}. \tag{1}$$

By means of the fuzzifier $m \in (1, \infty]$ one can control how much the clusters overlap. In order to avoid the trivial solution assigning no data to any cluster by setting all u_{ij} to zero and avoiding empty clusters, the following constraints are required:

$$u_{ij} \in [0, 1] \qquad 1 \le i \le c, \;\; 1 \le j \le n \tag{2}$$

$$\sum_{i=1}^{c} u_{ij} = 1 \qquad 1 \le j \le n \tag{3}$$

$$0 < \sum_{j=1}^{n} u_{ij} < n \qquad 1 \leq i \leq c. \tag{4}$$

The Euclidian norm

$$d_{ij} = d^2(\mathbf{v}_i, \mathbf{x}_j) = (\mathbf{x}_j - \mathbf{v}_i)^T(\mathbf{x}_j - \mathbf{v}_i)$$

is used for fuzzy c-means as distance measure. Other distance measures can be applied resulting in clustering techniques which can adopt different cluster shapes (Gath and Geva (1989), Gustafson and Kessel (1979)). The minimization of the functional (1) represents a nonlinear optimization problem that is usually solved by means of Lagrange multipliers, applying an alternating optimization scheme (Bezdek (1980)). This optimization scheme considers alternatingly one of the parameter sets, either the membership degrees

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{1}{m-1}}} \tag{5}$$

or the prototype parameters

$$\mathbf{v}_i = \frac{\sum_{j=1}^{n}(u_{ij})^m \mathbf{x}_j}{\sum_{j=1}^{n}(u_{ij})^m} \tag{6}$$

as fixed, while the other parameter set is optimized according to equations (5) and (6), respectively, until the algorithm finally converges.

3 Single Cluster Visualization

We apply in this paper a recently presented method to visualize clustering results of high-dimensional data on the plane (Rehm et al. (2006)). The main concepts of Single Cluster Visualization (SCV) are the visualization of the data set from the perspective of a certain cluster while preserving the fuzzy membership degrees when mapping the data onto the plane. Thus, the challenge is to determine representative distances of feature vectors to respective cluster prototypes that preserve the fuzzy membership degrees approximately.

As already mentioned, membership degrees describe the feature vector's gradual membership to a certain cluster and can be easily determined for any prototype based clustering technique using equation (5). Preserving membership degrees instead of preserving original distances allows a very efficient computation of meaningful transformations.

In general, membership degrees cannot be preserved exactly when dimensionality reduction is carried out. A helpful step to preserve membership degrees approximately is the adaptation of the noise clustering idea (Davé (1991), Davé and Krishnapuram (1997)). Noise clustering is based on the introduction of an additional cluster – the noise cluster – that is supposed to contain all noisy or outlying vectors. Noise is defined over a fixed distance, the so-called noise distance, and denoted by δ. The prototype $\mathbf{v}_c$ of such a noise cluster is rather virtual and thusly has no parameters. The clustering scheme differs only in that point from fuzzy c-means, that membership degrees to the noise cluster will be obtained considering $d_{cj} = \delta^2$ in equation (5). Note, the noise clustering aspect is only used here as a dodge to reduce the number of variables and not for outlier detection. A second step required for our visualisation purposes is a relaxation in that point that mainly the membership degrees to the two mostly competing prototypes will be regarded. All other prototypes will be regarded as one composed noise cluster.

When mapping a clustering result onto the plane two coordinates are needed for each data point. To achieve this, the usual computation of membership degrees according to equation (5) is considered which provides a simple connection between membership degrees and distances:

$$\frac{u_{ij}}{u_{\ell j}} = \frac{\frac{1}{\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{1}{m-1}}}}{\frac{1}{\sum_{k=1}^{c}\left(\frac{d_{\ell j}}{d_{kj}}\right)^{\frac{1}{m-1}}}} = \left(\frac{d_{\ell j}}{d_{ij}}\right)^{\frac{1}{m-1}} . \tag{7}$$

Cluster i is defined to be the cluster that is to be visualized. Note, the complete data set will be plotted by SCV, however, from the perspective of cluster i. With cluster ℓ we regard a second cluster, which is a virtual cluster, since it contains all feature vectors with the highest membership degree apart from u_{ij}. Membership degrees to this cluster will be denoted by $u_{\ell j}$. In this sense, this second cluster indicates for each data object, how much cluster i has to compete for this data object with another cluster. Data objects that are assigned with a high membership degree to cluster i will be placed close to projected cluster centre of cluster i, whereas data objects assigned to a high degree to any other cluster will be placed near the centre of the virtual (combined) cluster ℓ. Moreover, a noise cluster is defined covering all other clusters aside from i and ℓ. Since original distances are not preserved but representative distances by means of membership degrees, cluster i and cluster ℓ can be initially placed on arbitrary positions on the plane. To place prototype $\mathbf{v}_i$ at $(0,0)$ and prototype $\mathbf{v}_\ell$ at $(1,0)$ as proposed in (Rehm et al. (2006)) suggests to define the noise distance by $\delta = 1$. Keeping in mind that only two clusters plus the noise cluster are considered, we have $u_{noisej} = 1 - u_{ij} - u_{\ell j}$. According to equation (7) this leads to

$$\frac{u_{ij}}{u_{noisej}} = \left(\frac{1}{\hat{d}_{ij}}\right)^{\frac{1}{m-1}} . \tag{8}$$

The distance between cluster i and feature vector $\mathbf{x}_j$ on the plane is denoted by $\hat{d}_{ij}$ to emphasize the fact that it is not dealt with original distances any more but with representative distances with respect to the according membership degrees. Solving equation (8) for $\hat{d}_{ij}$ one obtains

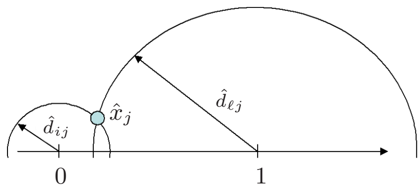


Fig. 1. Placement of $\hat{x}_j$ in the plane

$$\hat{d}_{ij} = \left(\frac{u_{noisej}}{u_{ij}}\right)^{m-1} . \tag{9}$$

Analogously, one obtains for the second cluster ℓ

$$\hat{d}_{\ell j} = \left(\frac{u_{noisej}}{u_{\ell j}}\right)^{m-1} . \tag{10}$$

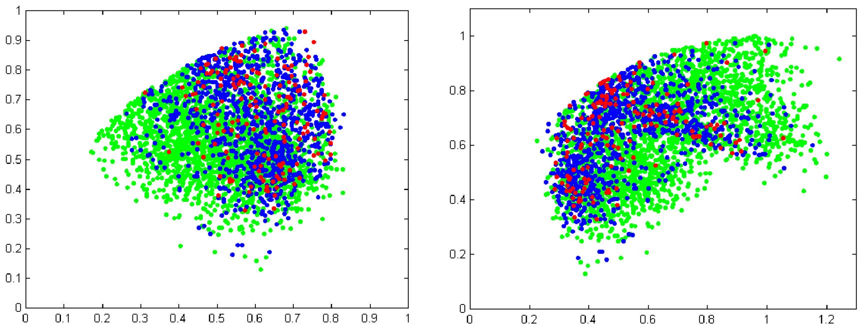
Figure 1 illustrates this approach. With equation (9) one can compute the representative distance of each feature vector x_j to the cluster i, so that it is possible to draw a circle around $(0, 0)$, the position of cluster i, as one hint for the feature vector's position in the plane. With the distance to the other cluster $(1, 0)$ that we get from equation (10), one could draw another circle around the cluster centre of cluster ℓ. The intersection point of these two circles would be the position of the new feature vector $\hat{x}_j$ in the plane.

Demonstrative examples are given in (Rehm et al. (2006)). In the next section, we discuss the results of an application of SCV on a problem that arises when predicting aircraft delay as a function of weather.

4 Results

In this section, we discuss the results of SCV applying it to a weather data set that is extensively described in (Rehm et al. (2005)). The data describes some weather factors captured by various sensors present at Frankfurt Airport. An ensemble of the different weather factors at one point in time, such as atmospheric pressure, temperature, wind speed, precipitation, height of cloud layers, etc., forms a weather report. Such a report is usually released every thirty minutes. In case of rapidly changing weather the frequency is increased.

In addition to the weather data flight durations of arriving aircraft are available. More precisely, we consider flight durations in the Terminal Management Area (TMA) - a controlled airspace over the airport - and classify them into short, medium and long flights. This traffic data as well as the weather data is available for one year. Earlier studies analyzed the same data set to predict flight durations of arriving aircraft (Rehm (2004), Rehm et al. (2005)). A variety of methods were applied and some weather factors affecting flight duration could be discovered. However, these predictions were always



(a) Cluster of short flight durations (b) Cluster of mixed flight durations

Fig. 2. Visualization of weather clusters

associated with a wide variance. The visualizations shown on figure 2 reveal the reason for that.

Figure 2(a) shows one cluster of a three-cluster-partition. Note, that the axis labels are intentionally omitted here, since the features of the new 2-dimensional feature space are combined attributes of the original space. All feature vectors left from 0.5 on the x-axis have their highest membership degree to the cluster we try to visualize here. On the first sight it is visible that no compact cluster could be found. The changeover from cluster i to cluster ℓ is quite fluent. According to the visualization no clear border between cluster i and another cluster can be drawn. A second cluster, depicted in figure 2(b), represents flights of all three categories. Estimating flight durations based on the cluster's average flight duration produces in comparable cases a considerable variance and poor predictions accordingly. Borders between cluster i and cluster ℓ cannot be decided.

These visualizations reveal that flight durations can be partly classified using weather data as figure 2(a) evinces. However, the whole data set should not be analyzed only by partitioning methods. Some regions in the feature space seem to be more complicated and flight duration categories cannot be separated linearly. Recent studies applying support vector machines could improve prediction quality and underline our assumptions (Lesot et al. (2006)).

5 Conclusions

We presented in this paper the application of SCV – an efficient technique to map clustering results of high-dimensional data onto the plane. We showed results on a complex weather data set that is used to predict aircraft delay. Earlier studies have already shown that delay prediction using this weather data is fairly complicated. The results in this work reveal reasons for that and give hints how to overcome the problem.

References

ABONYI, J. and BABUSKA, R. (2004): FUZZSAM - Visualization of Fuzzy Clustering Results by modified Sammon Mapping. *Proceedings of the IEEE International Conference on Fuzzy Systems, 365–370.*

BEZDEK, J.C. (1980): A Convergence Theorem for the Fuzzy ISODATA Clustering Algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 2, 1–8.*

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

DAVE, R.N. (1991): Characterization and Detection of Noise in Clustering, *Pattern Recognition Letters, 12, 657–664.*

DAVE, R.N. and KRISHNAPURAM, R. (1997): Robust Clustering Methods: A Unified View, *IEEE Transactions on Fuzzy Systems, 5, 270–293.*

DAVIES, D.L. and BOULDIN, W. (1979): A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1, 224–227.*

DUNN, J.C. (1974): Well Separated Clusters and Optimal Fuzzy Partitions. *Journal of Cybernetics, 4, 95–104.*

GATH, I. and GEVA, A.B. (1989): Unsupervised Optimal Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 773–781.*

GUSTAFSON, D.E. and KESSEL, W.C. (1979): Fuzzy Clustering with a Fuzzy Covariance Matrix. *Proceedings of the IEEE Conference on Decision and Control*, San Diego, 761–766.

HATHAWAY, R.J. and BEZDEK, J.C. (2003): Visual Cluster Validity for Prototype Generator Clustering Models. *Pattern Recognition Letters, 24, 9–10.*

HÖPPNER, F., KLAWONN, F., KRUSE, R. and RUNKLER, T. (1999): *Fuzzy Cluster Analysis.* John Wiley & Sons, Chichester.

HUBAND, J.M., BEZDEK, J.C. and HATHAWAY, R.J. (2005): bigVAT:Visual Assessment of Cluster Tendency for Large Data Sets. *Pattern Recognition Letters, 38, 1875–1886.*

KLAWONN, F., CHEKHTMAN, V. and JANZ, E. (2003): Visual Inspection of Fuzzy Clustering Results. In: Benitez, J., Cordon, O., Hoffmann, F. and Roy, R. (Eds.): *Advances in Soft Computing: Engineering Design and Manufacturing.* Springer, London, 65–76.

REHM, F. (2004): Prediction of Aircraft Delays as a Function of Weather. *2nd WakeNet2-Europe Workshop on Capacity Gains as Function of Weather and Weather Prediction Capabilities*, Langen, Germany.

REHM, F., KLAWONN, F. and KRUSE, R. (2005): Learning Methods for Air Traffic Management. In: Godo, L. (Ed.): *ECSQARU 2005, LNAI, 3571.* Springer, Berlin-Heidelberg, 992–1001.

REHM, F., KLAWONN, F. and KRUSE, R. (2006) Visualization of Single Clusters. *LNCS, 4029*, Springer, Berlin-Heidelberg, 663–671.

LESOT, M.J., REHM, F., KLAWONN, F. and KRUSE, R. (2006). Prediction of Aircraft Flight Duration. *Proceedings of the 11th IFAC Symposium on Control in Transportation Systems, Delft, 107–112.*

RUBENS, M. (1992): Fuzzy Clustering Algorithms and their Cluster Validity. *European Journal of Operational Research, 10, 294–301.*

WINDHAM, M.P. (1981): Cluster Validity for Fuzzy Clustering Algorithms. *Fuzzy Sets and Systems, 5, 177–185.*