Modern Data Visualization for Air Traffic Management

Frank Rehm
German Aerospace Center
Lilienthalplatz 7
38108 Braunschweig
Germany
frank.rehm@dlr.de

Frank Klawonn
University of Applied Sciences
Braunschweig/Wolfenbüttel
Salzdahlumer Str. 46/48
38302 Wolfenbüttel
Germany
f.klawonn@fh-wolfenbuettel.de

Georg Ruß
University of Magdeburg
Universitätsplatz 2
39106 Magdeburg
Germany
russ@iws.cs.uni-magdeburg.de

Rudolf Kruse
University of Magdeburg
Universitätsplatz 2
39106 Magdeburg
Germany
kruse@iws.cs.uni-magdeburg.de

***Abstract*— Air traffic at airports is affected by various factors. The capacity of an airport and the demand at a certain point in time are serious parameters that account for a big extent to aircraft delay and related variables. It has been proven that weather is another important impact in this regard. Although weather cannot be controlled, the knowledge of how weather affects the air traffic at an airport can be very helpful to optimize air traffic management. Data mining promises to gain that knowledge.**

Usually, the very first step in data mining is data visualization. In this paper we discuss two new visualization techniques that allow to visualize aviation data and weather data in order to contribute to the optimization process. These modern multi-dimensional scaling techniques provide mappings of high-dimensional data to low-dimensional feature spaces. We will show some results on a practical application of a major European airport.

I. Introduction

During the recent years a continuous world wide growth of air traffic could be observed, neglecting a temporary decrease due to terrorist acts in 2001. Likewise traffic forecast estimates an increase of about 4% yearly for Europe [6]. Of course, the airport's capacity is mainly constrained by its traffic demand. Earlier studies have shown that weather is another major factor that affects the performance of an airport [1], [3], [5], [7].

Since weather is a comprehensive description of multiple single phenomena like temperature, height of cloud layers, wind speed, etc., data sets, describing the combination of weather and traffic at an airport, are fairly complex and thus high-dimensional when projecting them into a feature space.

Commonly, visualization is used as a part of data preprocessing when trying to determine appropriate data mining methods. In order to visualize high-dimensional data, sophisticated techniques need to be applied that permit to display complex data on limited projection media as computer monitors or printed hard copies.

Multi-dimensional scaling [2] is a family of methods that seek to present the important structure of complex data on a reduced number of dimensions. In this paper we present two recent techniques of this kind and apply them on the practical example of aviation and weather data.

The rest of the paper is organized as follows: in section II we discuss two recent multi-dimensional scaling approaches. In section III we briefly describe the application and the data. In section IV results on aviation and weather data will be shown followed by some concluding remarks in section V.

II. Multi-dimensional scaling

Multi-dimensional scaling (MDS) aims at arranging n data objects $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^p$ into a low-dimensional (typically a two- or three-dimensional) feature space while preserving pairwise similarity of objects approximately. Similarity is usually defined by a distance measure, e.g. the Euclidian distance. In this section we present two recent approaches that provide multi-dimensional scaling based on density preservation and on the approximate preservation of pairwise angles between data objects.

A. MDS_{polar}

Different from conventional multi-dimensional scaling where pairwise distances between data are considered, MDS_{polar} generates 2-dimensional layouts preserving angles between data approximately. For this purpose, a representation in polar coordinates $Y = \{(l_1, \varphi_1), \ldots, (l_n, \varphi_n)\}$ is used for the target feature space, which allows to preserve the length l_k of the original vector x_k on the one hand and demands the optimization of the angle φ_k on the other hand. The solution of MDS_{polar} is defined to be optimal, if all angles φ_i and φ_k between pairs of data objects in the projected data set Y coincide as well as possible with the angles ψ_{ik} in the original feature space X.

The use of polar coordinates in the target space has two major advantages: the number of parameters to be optimized is reduced and generalization can be achieved, which means that new points can be added without recalculation of the entire mapping.

Due to length preservation, data vectors of different length will be mapped far away from each other. This already guarantees a roughly correct placement of the feature vectors in the target space. In order to differentiate dissimilar data of similar length, the approximation of the respective angles have

to be considered. Since dissimilarity of objects of different length is inherently reflected in the mapping, efficiency can be gained when angle arrangements are mainly constrained to objects of similar length. For vectors having a significant difference in length, angle arrangements can be neglected. The minimization of the error that has to be taken into account when mapping high-dimensional data onto the plane while approximating pairwise angles, is formalized by the following objective function:

$$E = \sum_{k=2}^{n} \sum_{i=1}^{k-1} w_{ik}(\varphi_i - \varphi_k - a_{ik}\psi_{ik})^2. \quad (1)$$

Function (1) contains some variables that are not introduced so far. Although, original angles ψ_{ik} will always satisfy $0 \leq \psi_{ik} \leq 180°$, the order of the minuends φ_i and φ_k can have an influence on the sign of the resulting angle. Therefore, the angle between y_i and y_k might perfectly match the angle ψ_{ik}, $\varphi_i - \varphi_k$ can either be ψ_{ik} or $-\psi_{ik}$. The straight forward approach, putting the term $\varphi_i - \varphi_k$ into brackets and taking the absolute or the quadratic value, unfortunately, yields either a function that is not differentiable entirely or whose derivatives describe a system of non-linear equations, for that no analytical solution can be provided. In [8] it is explained in detail that we can take the freedom to choose whether we want the term $\varphi_i - \varphi_k$ or the term $\varphi_k - \varphi_i$ to minimize function (1). Since we have $(\varphi_k - \varphi_i - \psi_{ik})^2 = (-(\varphi_k - \varphi_i - \psi_{ik}))^2 = (\varphi_i - \varphi_k + \psi_{ik})^2$, instead of exchanging the order of φ_i and φ_k, we can choose the sign of ψ_{ik}. The function reflects the respective sign of ψ_{ik} in form of the a_{ik}-values, with $a_{ik} = \{-1, 1\}$.

Parameter w_{ik} refers to the concept of weighting the error of angle adjustments. This weighting parameter should be controlled in that way that mainly angles of data objects with similar vector length will be optimized. Thus, the weight of vectors with similar length should be larger than the weight of vectors with significantly different length. This has the effect that an optimization procedure focuses on these angles since they produce large errors. The optimization procedure can be improved when omitting the adjustment of weights for pairs of vectors, whose mapping is sufficiently differentiated by the consideration of their length.

As a fist step to minimize E we obtain the following partial derivative

$$\frac{\partial E}{\partial \varphi_k} = -2 \sum_{i=1}^{k-1} w_{ik}(\varphi_i - \varphi_k - \psi_{ik}) \quad (2)$$

and set it equal to zero in order to fulfill the necesary condition for a minimum. Solving $\partial E / \partial \varphi_k$ for φ_k we get

$$\varphi_k = \frac{\sum_{i=1}^{k-1} w_{ik}(\varphi_i - a_{ik}\psi_{ik})}{\sum_{i=1}^{k-1} w_{ik}}. \quad (3)$$

For an efficient implementation it is useful to sort the original feature vectors by means of their length and using a decreasing weighting function that depends on the dissimilarity of the data pair. Then an efficient computation scheme can be defined that reduces computational costs drastically.

Algorithm 1 Greedy MDS$_{polar}$

Given the data set $X = \{x_1, x_2, \ldots, x_n\}$
Let $\Psi_{n \times n}$ be a matrix with the pairwise angles ψ_{ij} between all (x_i, x_j)
$\varphi_1 = 0$
for $k = 2$ to n **do**
  $a_{ik} = 1$ for all $i = 1 \ldots k-1$
  for $i = 1$ to $k-1$ **do**
    w_{ik} = `getWeight(`l_i, l_k`)`
    if $w_{ik} > 0$ **then**
      $\varphi_k = \sum_{i=1}^{k-1} w_{ik}(\varphi_i - a_{ik}\psi_{ik}) / \sum_{i=1}^{k-1} w_{ik}$
      $e_k = \sum_{j=1}^{k-1} w_{ik}(\varphi_j - \varphi_k - a_{jk}\psi_{jk})^2$
      $t = \varphi_k$
      $a_{ik} = -1$
      $\varphi_k = \sum_{i=1}^{k-1} w_{ik}(\varphi_i - a_{ik}\psi_{ik}) / \sum_{i=1}^{k-1} w_{ik}$
      $f_k = \sum_{j=1}^{k-1} w_{ik}(\varphi_j - \varphi_k - a_{jk}\psi_{jk})^2$
      if $e_k < f_k$ **then**
        $a_{ik} = 1$
        $\varphi_k = t$
      end if
    else
      break //breaks the inner **for**-loop
    end if
  end for
end for

Algorithm 1 describes the optimization procedure schematically. The function `getWeight()` returns a weight according to a user defined weighting function, e.g.

$$\texttt{getWeight}(l_i, l_k) \quad = \quad \begin{cases} 1, & \text{if } |l_i - l_k| \leq \vartheta \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where ϑ is a user defined threshold. This simple binary weighting function returns either 1 for similar data vectors or 0 for data vectors that are sufficiently differentiated by means of their vector lengths. Many other weighting functions are feasible, e.g. restricting the maximum number of non-zero weights for each k is suggestive. This refers to binning-strategies that allow a conservative estimation of maximum computational costs.

Once one weight w_{ik} became zero, the inner **for**-loop will be interrupted and the next k will be proceeded. This approach is efficient and reasonable, since sorting by vector length will arrange the data set X in such a way that, once one w_{ik} became zero, all succeeding weights will also be zero for the same k. Initially, all signs for ψ_{ik} are positive, i.e. $a_{ik} = 1$. Algorithm 1 greedily sets a negative sign to ψ_{ik} when $e_k \geq f_k$, i.e. when $-\psi_{ik}$ actually minimizes E.

B. Density-based MDS

The approach of density-based MDS has been recently proposed in [9]. The idea behind is to reflect density variations of high-dimensional data on low-dimensional mappings. For this purpose, a multivariate Gaussian distribution is defined for each data object in the original space that represents the data point's potential energy. Adding these potentials yields multi-dimensional mountains. Summits of the mountains represent dense data regions, valleys represent sparse data areas. Density-based MDS aims at transforming the original data and finding a low-dimensional layout whose potentials coincide as well as possible with the mountains of the original.

Formally the problem is described as follows. Given the data set $X = \{x_1, \ldots, x_n\} \subset \mathbb{R}^p$ we seek for the mapped data set $Y = \{y_1, \ldots, y_n\} \subset \mathbb{R}^k$ with $k = 2$ or $k = 3$ with the following potential for x_i:

$$f_i(x) \;=\; \frac{1}{c} \exp\left(-\frac{1}{2} \sum_{t=1}^{p} \left(\frac{x^{(t)} - x_i^{(t)}}{\sigma} \right)^2 \right) \tag{5}$$

with

$$c \;=\; \frac{1}{\sigma^p \sqrt{(2\pi)^p}}.$$

By $x^{(t)}$ and $x_i^{(t)}$ the t^{th} attribute of data object x or x_i is denoted. Function f_i simply describes the density of a p-dimensional Gaussian distribution with mean value x_i and variance σ^2 in each dimension. The parameter σ is fixed for the entire procedure. If σ is small, then the potentials do rarely overlap. For very large σ the potential landscape will be blurred completely with little variance in height. Therefore, it is useful to define σ according to the diameter d of the data space, the average distance between data points, the number n of data and the dimensionality p.

For the target data space, the procedure tries to map the feature vectors $Y = \{y_1, \ldots, y_n\} \subset \mathbb{R}^k$ such that the potentials

$$g_i(y) \;=\; \frac{1}{\tilde{c}} \exp\left(-\frac{1}{2} \sum_{t=1}^{k} \left(\frac{y^{(t)} - y_i^{(t)}}{\tilde{\sigma}} \right)^2 \right) \tag{6}$$

with $\tilde{c} = \frac{1}{\tilde{\sigma}^k \sqrt{(2\pi)^k}}$ coincide at least in these points with those in the original space. For this, $\tilde{\sigma}$ should be chosen similarly to σ.

Normally, the diameter d in the target space should be approximately of the same extent as for the original space. Otherwise, the area (or the volume) of the target space will be much smaller compared to the hyper volume of the original space ($k \ll p$). This means that the density in the target space is also higher for the same size of the data set. Thus, $\tilde{\sigma}$ should be chosen smaller than σ. Further, it should be assured that the maximum height of the single potentials in the original space and in the target space match, i.e. the respective maxima of the Gaussian distributions should be:

$$f_i(x_i) \;\approx\; g_i(y_i).$$

Normally this will not be the case. Therefore a constant a is introduced:

$$a \cdot f_i(x_i) \;=\; g_i(y_i)$$

which can be derived from equations (5) and (6):

$$a \;=\; \frac{\sigma^p}{\tilde{\sigma}^k} \sqrt{(2\pi)^{p-k}}.$$

From this one can formulate the objective function. The summarized modified potential in the original space is $\sum_{j=1}^{n} a \cdot f_j(x_i)$ at x_i and $\sum_{j=1}^{n} g_j(y_i)$ at y_i in the target space. Since both potentials should be equal, the objective function can be defined as follows:

$$\begin{aligned} E_{density} \;&=\; \sum_{i=1}^{n} \left(\sum_{j=1}^{n} g_j(y_i) \;-\; \sum_{j=1}^{n} a \cdot f_j(x_i) \right)^2 \\ &=\; \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \left(g_j(y_i) - a \cdot f_j(x_i) \right) \right)^2 . \end{aligned} \tag{7}$$

The gradient for each component s is:

$$\frac{\partial E_{density}}{\partial y_{ls}} \;=\; 2 \sum_{i=1}^{n} \sum_{j=1}^{n} \left(g_j(y_i) - a \cdot f_j(x_i) \right) \cdot \frac{\partial}{\partial y_{ls}} g_j(y_i). \tag{8}$$

$\frac{\partial}{\partial y_{ls}} g_j(y_i)$ is only zero when we have $l = i$ or $l = j$. For both cases we obtain the derivatives from equation (8):

$$\frac{\partial}{\partial y_{ls}} g_l(y_i) = \frac{1}{\tilde{c}} \exp\left(-\frac{1}{2} \sum_{t=1}^{k} \left(\frac{y_i^{(t)} - y_l^{(t)}}{\tilde{\sigma}} \right)^2 \right) \cdot \frac{y_i^{(s)} - y_l^{(s)}}{\tilde{\sigma}}$$

$$\frac{\partial}{\partial y_{ls}} g_j(y_l) = -\frac{1}{\tilde{c}} \exp\left(-\frac{1}{2} \sum_{t=1}^{k} \left(\frac{y_l^{(t)} - y_j^{(t)}}{\tilde{\sigma}} \right)^2 \right) \cdot \frac{y_l^{(s)} - y_j^{(s)}}{\tilde{\sigma}}.$$

For $i = j = l$ we have $\frac{\partial}{\partial y_{ls}} g_j(y_i) = 0$. Finally the gradient is:

$$\begin{aligned} \frac{\partial E_{density}}{\partial y_{ls}} = \frac{2}{\tilde{c}} \sum_{i=1}^{n} \Bigg(& \left(g_l(y_i) - a \cdot f_l(x_i) \right) \\ & \cdot \exp\left(-\frac{1}{2} \sum_{t=1}^{k} \left(\frac{y_i^{(t)} - y_l^{(t)}}{\tilde{\sigma}} \right)^2 \right) \cdot \frac{y_i^{(s)} - y_l^{(s)}}{\tilde{\sigma}} \\ & - \left(g_i(y_l) - a \cdot f_i(x_l) \right) \\ & \cdot \exp\left(-\frac{1}{2} \sum_{t=1}^{k} \left(\frac{y_l^{(t)} - y_i^{(t)}}{\tilde{\sigma}} \right)^2 \right) \cdot \frac{y_l^{(s)} - y_i^{(s)}}{\tilde{\sigma}} \Bigg). \end{aligned} \tag{9}$$

The density-based approach should be combined with conventional MDS like Sammon's mapping [11]. Thus, the Sammon gradient E_{sammon} and the density gradient $E_{density}$ can be combined through linear combination:

$$E = \alpha \frac{\partial E_{sammon}}{\partial y_l} + \beta \frac{\partial E_{density}}{\partial y_l}. \tag{10}$$

The parameters α and β can be considered as learning rates or weights to control the impact of the respective mapping strategy. Thus, higher weights α for the Sammon gradient favour distance-based mappings and larger values β for the density gradient favour the density approach.

III. Application

In this section we briefly describe the data to analyze. Subject of the study is to show the influence of weather in the vicinity of Frankfurt Airport on the travel time of approaching aircraft. Therefore, two combined data sets will be considered: a weather data set that describes the weather situation at the airport and an aviation data set that comprises arrival times of all approaching aircraft at the airport.

A. Weather Data

The weather data originate from the ATIS[1] weather data set. Different sensors present at the airport capture several weather characteristics and form a weather report. Such a report is released every thirty minutes (in case of rapidly changing weather, the frequency is increased). Each weather report contains information such as temperature, air pressure, wind speed and precipitation information, e.g. the presence of snow, rain or hail.

B. Traffic Data

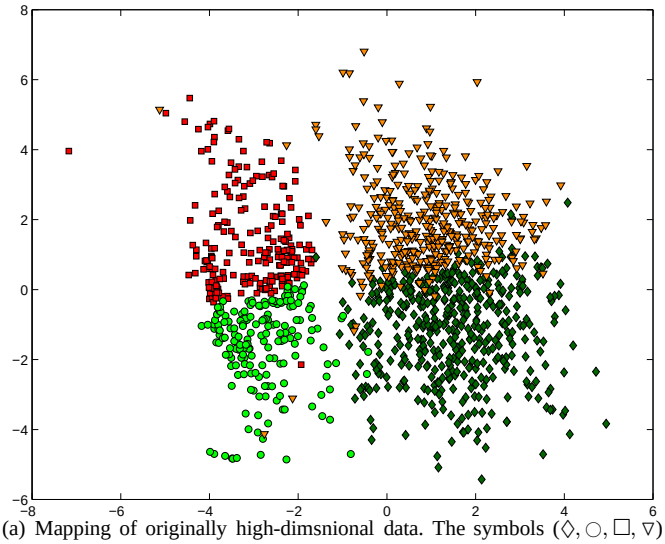
In addition to the weather data set, information about the traffic is available, through a data set that contains the arrival times of all aircraft at Frankfurt Airport for the observed time period. Since the variation in the travel time is of interest that is caused by weather factors in the vicinity of the airport, the point in time of the aircraft's entrance in the airport vicinity – the terminal area (TMA) – and the time when the corresponding aircraft is landing are considered. The difference between these two times corresponds to the observed travel time in the TMA.

Many research has been done on this subject[4], [7], [10]. In this paper we will analyze a sample of the data set with the objective to get more insight into a specific scenario. Subject of this study will be the analysis of the impact of traffic demand during extreme weather situations. For this purpose we build a data set that contains all weather reports denoting either low visibility range or increased wind speed, both, in association with low traffic demand and high traffic demand, respectively. This corresponds to 1205 data which are labeled according to three travel time classes (short, medium, long). 23% of this data account for short flights, 42% account for flights with medium travel times and 35% for flights with long flight durations in the TMA[2].

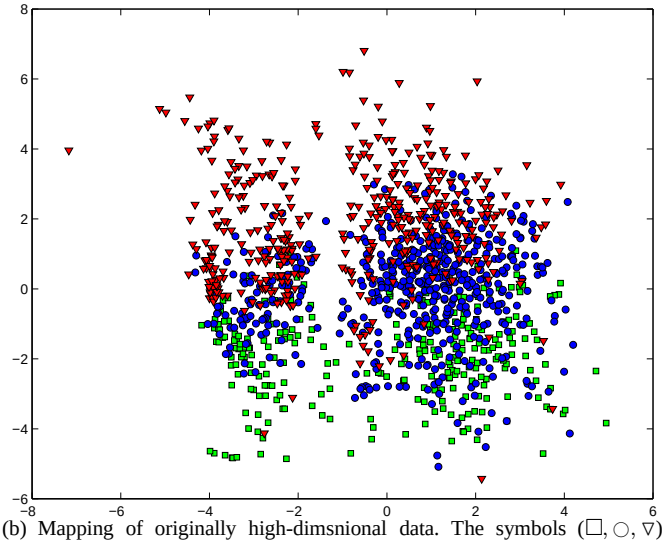
With both techniques that we have described above, we will visualize these different scenarios. It will show whether the impact of traffic demand or weather, combined or separately, is reflected in the data.

[1]ATIS (Automatic Terminal Information Service) is a continuous broadcast of recorded information in airports. ATIS broadcasts contain essential weather information but also the active runway and other information required by the pilots.

[2]For comparison, in the full data we have 45% short flights, 41% medium flight durations and 14% long flights in the TMA.



(a) Mapping of originally high-dimsnional data. The symbols ($\Diamond, \bigcirc, \square, \triangledown$) denote different weather situations under various traffic demand.



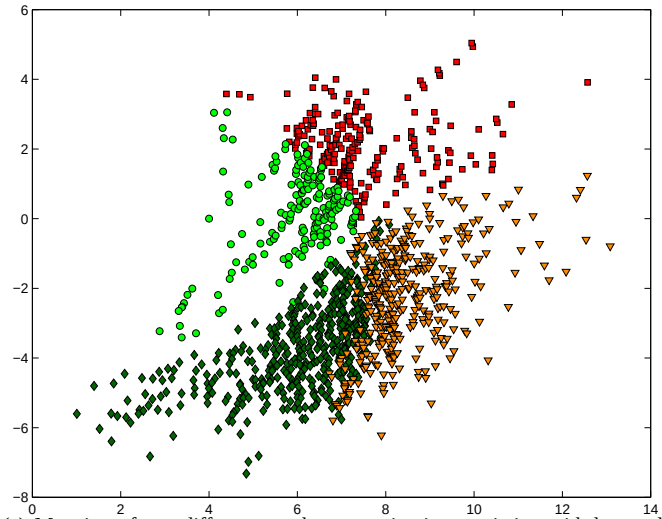
(b) Mapping of originally high-dimsnional data. The symbols ($\square, \bigcirc, \triangledown$) denote different flight durations.

Fig. 1. Sammon's Mapping of the combined aviation and weather data set, describing two weather scenarios in association with low and high traffic demand.

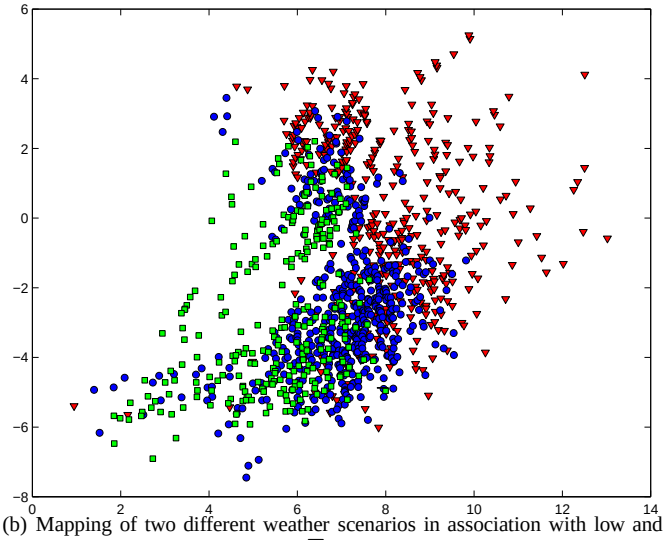
IV. Results

In this section we will show some results of both multi-dimensional scaling techniques on the combined aviation and weather data set.

As a reference, figures 1(a) and 1(a) show different interpretations of one mapping of the data gained with Sammon's mapping. The symbols in figure 1(a) can be read as follows. Small circles ($\bigcirc$) represent weather reports with low visibility and low traffic demand. Small boxes ($\square$) refer to low visibility und high traffic demand. Weather, implying increased wind speed with low traffic demand, is visualized by diamond symbols ($\Diamond$). Small triangles ($\triangledown$) represent weather reports with increased wind speed and high traffic demand. The scattering of the feature vectors clearly shows that two

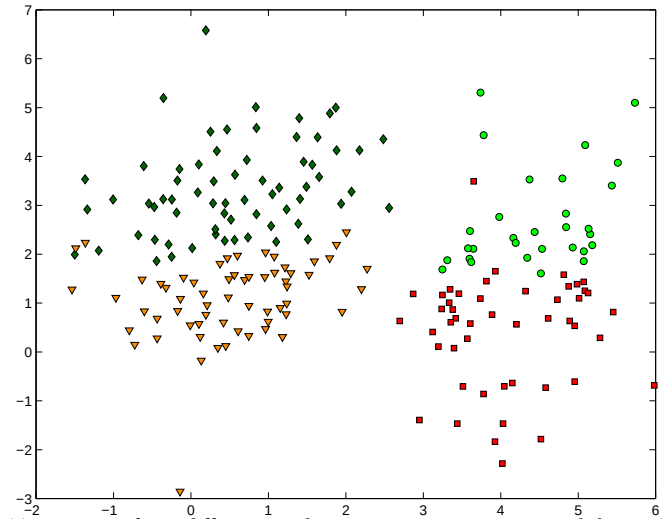


(a) Mapping of two different weather scenarios in association with low and high traffic demand. The symbols ($\Diamond, \bigcirc, \square, \triangledown$) denote different weather situations under various traffic demand.

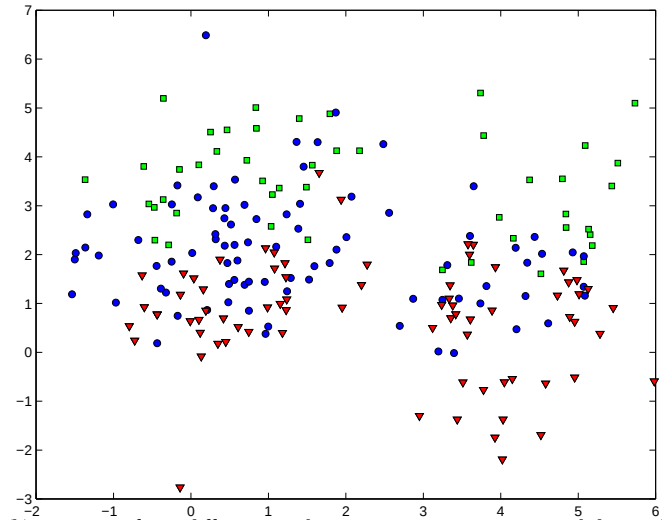


(b) Mapping of two different weather scenarios in association with low and high traffic demand. The symbols ($\square, \bigcirc, \triangledown$) denote different flight durations

Fig. 2. Mapping of the combined aviation and weather data set with MDS_{polar}.



(a) Mapping of two different weather scenarios in association with low and high traffic demand. The symbols ($\Diamond, \bigcirc, \square, \triangledown$) denote different weather situations under various traffic demand.



(b) Mapping of two different weather scenarios in association with low and high traffic demand. The symbols ($\square, \bigcirc, \triangledown$) denote different flight durations

Fig. 3. Mapping of the combined aviation and weather data set with density-based MDS.

weather clusters are formed. Likewise, different traffic demand is clearly reflected in the mapping. While the upper parts of both figures mainly represent high traffic demand, lower traffic demand is mostly spread in the lower part of both figures. Figure 1(b) shows the same mapping but symbols refer to another interpretation. In this figure, small boxes ($\square$) represent weather reports where short travel times could be achieved. Medium travel times are visualized by small circles ($\bigcirc$) and longer travel times by triangles ($\triangledown$), respectively. Obviously, most data points account for medium and long travel times, but also data, comprising short travel times, can be found. Comparing to the travel time distribution of the full data set, where short flight durations prevail a shift to long travel times can be observed. This fact reflects the impact of weather. Further, the correlation of increasing traffic demand and longer travel times is clearly reflected in the mapping.

Figures 2(a) and 2(b) show a mapping of the data using MDS_{polar}. The main characteristics of the reference mapping (see figure 1) can also be found here. In the lower left part of the mapping some outlying medium and long flight durations can be observed. Obviously, another factor that is not reflected by the weather data affects flight duration. A binary weighting function yielding a bin-size of 200 is used with MDS_{polar} to obtain the mapping. As for all mappings that are obtained by multi-dimensional scaling, both axes do not represent either of the original parameters, but a combination of all. Therefore, axes labeling is omitted intentionally. Mappings are rotation invariant. Both interpretations of the mapping, again, show

that two weather clusters are formed as for the reference mapping, and travel time increases for demanding traffic. Modifications on the greedy algorithm in such a way, that the inner **for**-loop will be repeated several times, allow a drastic reduction of the bin-size yielding equal quality mappings to lower computational costs.

Density-based multi-dimensional scaling leads to mappings shown in figures 3(a) and 3(b). This technique is fairly expensive regarding computational costs and therefore only applicable to smaller data sets. For this reason we applied it to a sample of the combined aviation and weather data set. Also this mapping shows that weather and flight duration discriminate the data noticeably.

V. Conclusion

In this paper, we have described two recent techniques that provide visualization of high-dimensional data. As a practical application, we have shown mappings of real high-dimensional weather and aviation data. The results that were obtained by MDS_{polar} and density-based MDS clearly show the impact of weather on arriving aircraft, but also the impact of traffic demand during critical weather situations. Subject of future research should be a speedup of density-based MDS on the one hand and the consideration of different weather scenarios on the other hand.

References

[1] S.S. Allan, J.A. Beesley, J.E. Evans and S.G. Gaddy, "Analysis of Delay Causality at Newark International Airport," *4th USA/Europe Air Traffic Management R&D Seminar*, Santa Fe, 2001.

[2] I. Borg and P. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer, Berlin, 1997.

[3] J.E. Evans, S. Allan and M. Robinson, "Quantifying Delay Reduction Benefits for Aviation Convective Weather Decision Support Systems," *Proceedings of the 11th Conference on Aviation, Range, and Aerospace Meteorology*, Hyannis, 2004.

[4] M-J. Lesot, F. Rehm, F. Klawonn and R. Kruse, "Prediction of Aircraft Flight Duration," *Proceedings of the 11th IFAC Symposium on Control in Transportation Systems*, Delft, 2006, pp. 107–112.

[5] Z. Nazeri and L. Zhang, "Mining aviation data to understand impacts of severe weather on airspace system performance," *Proceedings of the International Conference on Information Technology: Coding and Computing*, 2002, pp. 518–523.

[6] *Performance review commission: Performance Review Report*, Eurocontrol, Brussels, 2005.

[7] F. Rehm and F. Klawonn, "Learning Methods for Air Traffic Management," *Lecture Notes in Computer Science 3571*, Springer, 2005, pp. 992–1001.

[8] F. Rehm, F. Klawonn and R. Kruse, "MDS_{polar} - A new Approach for Dimension Reduction to Visualize High Dimensional Data," in A.F. Famili, J.N. Kok, J.M. Pea, A. Siebes and A.J. Feelders, Eds. *Proceedings of the 6th International Symposium on Intelligent Data Analysis (IDA 2005), Lecture Notes in Computer Science 3646*, Springer, 2005, pp. 316–327.

[9] F. Rehm, F. Klawonn and R. Kruse, "Density-based Multidimensional Scaling," (submitted for publication).

[10] F. Rehm and F. Klawonn, "Improving Air Traffic Management with Visual Data Analysis," *Proceedings of the AIAA Infotech@Aerospace 2007 Conference*, Rohnert Park (CA), 2007.

[11] J.W. Sammon, "A nonlinear mapping for data structure analysis," *IEEE Transactions on Computers, vol. 18*, 1969, pp. 401–409.