Clutter reduction in Parallel Coordinates using locality-aware clustering

Anže Starič and Janez Demšar and Blaž Zupan

January 29, 2014

1 Introduction

Parallel coordinates are a way of presenting multivariate data in a 2D plot. It is a commonly used visualization, but does not scale well to large datasets (¿1000 data points), as large number of lines results in too much visual clutter that obscures the trends we are looking for.

Several approaches have been proposed for reducing clutter. Most of them are based on aggregating data and later drawing the aggregates as densities [5], wavelets [10], polygons [1, 9] or bars with varying opacity [2, 4].

Many authors [4, 7, 9, 1] use clustering as an aggregating function splitting data into multiple subsets and displaying properties of those subsets on the parallel coordinates plot. These methods usually use hierarchical [4] clustering when they need multiple levels of detail, or k-means clustering because of its speed. These clustering algorithms optimize clusters globally and do not care for the ordering of the features in the visualization.

We propose a novel clutter reducing technique using locality aware clustering. It is based on the Gaussian Mixture Models which are learned using expectation maximization. Learning process uses information about the order of attributes in the visualization to produce clusters that are optimized for the given projection. The resulting clusters have smaller variance and are better separated in the parallel coordinates plot than those obtained with the ordinary clustering methods.

2 Related work

Multiple existing methods use clustering to reduce clutter.

Fua et.al [4] use hierarchical clustering to provide multiple representations with different level of detail. Clusters are displayed as bands with opaque means and transparency linearly decreasing towards the cluster edge. This emphasizes the cluster width, but does not provide information about the examples within the cluster.

Johansson et. al [7] crate high-precision textures for clusters and outliers and combine those with transfer function to create visualizations. Changing the transfer function allows the user to put emphasis on different aspects of data. Different statistical properties of the data can be displayed using visualization.

Novotny [9] uses a modified version of k-means clustering and displays clusters as polygons covering all examples in the cluster. Polygons are sorted by size before they are displayed to optimize the visibility of dense clusters.

Andrienko et. al [1] propose two methods of visualizing distribution of attribute values within clusters. First approach splits intersection of cluster with each axis into quantiles and then draws lines connecting them, thus showing dense parts as narrow bands within the cluster. Another approach are ellipse plots that can be shown behind each axis and also visualize axis value distribution.

Our approach is similar to [4, 7, 9] as it also uses clusters to reduce clutter. Like [2], it uses probability based clustering that assigns example a probability of belonging to a specific cluster. Clusters are drawn with polygon like in [9, 1]. We borrow drawing quantiles in clusters from [1] to provide additional information about value distribution inside clusters.

3 Expectation maximization

A Gaussian mixture model (GMM) is a parametric statistical model that assumes that the data originates from a weighted sum of several Gaussian sources. More formally, a GMM is given by $p(x|\Theta) = \sum_{j=1}^{k} \alpha_j p(x|\theta_j)$, where α_j denotes the weight of each Gaussian, θ_j its respective parameters and k denotes the number of Gaussian sources in GMM. Expectation maximization (algorithm 1) is a widely used method for estimating parameter set of the model (Θ) using unlabeled data [3]. In the algorithms, we optimize GMM for dataset X, which has m examples, and each of the Gaussians is parametrized by mean μ, covariance matrix Σ and its prior ϕ.

TODO: definicija preisana iz enega clanka, povzemi po svoje

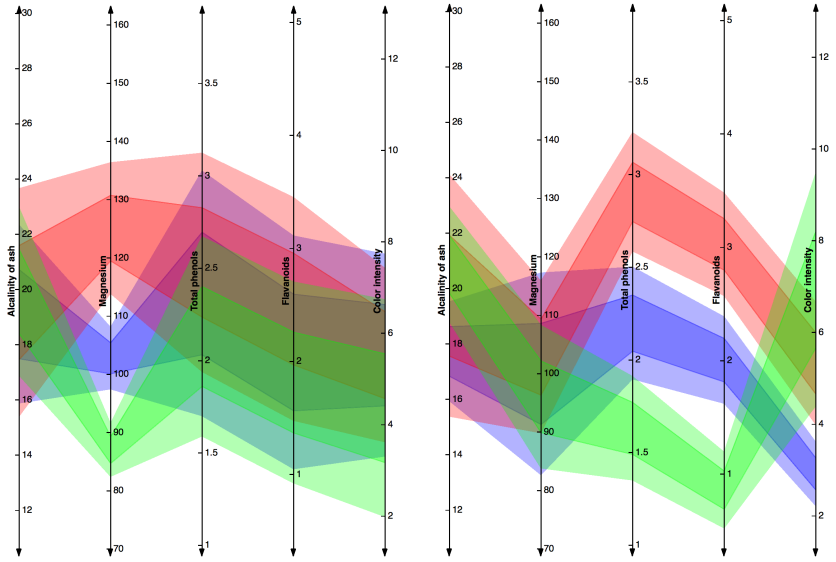
Algorithm 1 Standard EM GMM

function EM_GMM(max_steps, X)
  $initialize(\mu, \Sigma, \phi)$
  for $step$ **in** $1..max\_step$ **do**
    for each $i \leftarrow 1..m$, $j \leftarrow 1..k$ **do**
      $w_j^{(i)} \leftarrow p(z^{(i)} = j|x^{(i)}; \phi, \mu, \Sigma)$
    end for

    $\phi_j \leftarrow \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)}$
    $\mu_j \leftarrow \frac{\sum_{i=1}^{m} w_j^{(i)} x^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}}$
    $\Sigma_j \leftarrow \frac{\sum_{i=1}^{m} w_j^{(i)} (x^{(i)} - \mu_j)(x^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} w_j^{(i)}}$
  end for
end function

4 Locality aware clustering

Learning a mixture of Gaussian models on multidimensional vectors constructs a model that takes all features into account. Clusters fit the data well, but when



(a) k-means clusters (b) locality aware clusters

Figure 1: Comparison of clusters produced with k-means on the left (a) and locality aware clustering on the right (b). Our method finds clusters that are narrower and better separated in maximum number of axes.

displayed on a parallel coordinates display they overlap.

Learning a mixture of Gaussians for each feature independently leads to clusters that fit value distribution for each feature well, but now examples that belong to the same cluster on one feature connect to multiple clusters on the other feature which results in noise.

Our method (algorithm 2) optimizes each component of the parameters independently, but considers local neighborhood when calculating updates. When drawn on the parallel coordinates plot, computed clusters are better separated and produce less clutter compared to the clusters computed with k-means.

5 Evaluation

We define our evaluation metric as ....

Algorithm 2 Modified EM GMM

function OUR_EM_GMM(max_steps, $window\_size$, X)
  $initialize(\mu, \Sigma, \phi)$
  for $step$ **in** $1..max\_step$ **do**
    for f **in** $features$ **do**
      $XS \leftarrow select\_features(f, window\_size, X)$
      for each $i \leftarrow 1..m$, $j \leftarrow 1..k$ **do**
        $w_j^{(i)} \leftarrow p(z^{(i)} = j | xs^{(i)}; \phi, \mu, \Sigma)$
      end for

      $\phi_j \leftarrow \frac{1}{m} \sum_{i=1}^{m} w_j^{(i)}$
      $\mu_{j,f} \leftarrow \frac{\sum_{i=1}^{m} w_j^{(i)} xs^{(i)}}{\sum_{i=1}^{m} w_j^{(i)}}$
      $\Sigma_{j,f} \leftarrow \frac{\sum_{i=1}^{m} w_j^{(i)} (xs^{(i)} - \mu_j)(xs^{(i)} - \mu_j)^T}{\sum_{i=1}^{m} w_j^{(i)}}$
    end for
  end for
end function

adult.tab	466722.67	327341.18
adult_sample.tab	286367.89	358062.77
anneal.tab	1013.93	861.37
breast-cancer-wisconsin-cont.tab	129.06	127.62
brown-selected.tab	40.64	21.65
bupa.tab	447.15	431.49
crx.tab	499.98	295.23
echocardiogram.tab	176.93	62.95
emotions.tab	289.08	354.48
glass.tab	26.66	6.77
heart_disease.tab	454.14	102.28
horse-colic.tab	354.36	81.46
horse-colic_learn.tab	335.56	81.27
horse-colic_test.tab	198.24	191.44
housing.tab	785.60	794.86
imports-85.tab	3871.48	3040.34
ionosphere.tab	120.04	49.30
vehicle.tab	1260.12	715.01
water-treatment.tab	37078.71	50445.88
wdbc.tab	2293.66	2119.69
wine.tab	116.19	108.52
yeast-class-RPR.tab	40.64	21.65

References

[1] G. Andrienko and N. Andrienko. Parallel coordinates for exploring properties of subsets. In *Procedings of the Second International Conference on Coordinated and Multiple Views in Exploratory Visualizations* (2004), pp. 93-104.

[2] M. Berthold and L. Hall. Visualizing fuzzy points in parallel coordinates. *IEEE Transactions on Fuzzy Systems* 11(3), (2005), pp. 369-374.

[3] A. P. Dempster, N.M. Laird, and D.B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* JRSSB, 39:1-38, 1997

[4] Y. H. Fua, M. O. Ward and E. A. Rundensteiner. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of IEEE Visualization* (1999), pp. 43-50.

[5] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), (2009) pp. 1531-1538.

[6] J. Johansson, P. Ljung and M. Cooper. Depth cues and density in temporal parallel coordinates. In *Proceedings of the Eurographics/IEEE-VGTC Symposium on Visualization* (2007), pp. 35-42.

[7] J. Johansson, P. Ljung, M. Jern and M. Cooper. Revealing structure within clustered parallel coordinates displays. In *Proceedings of the IEEE Symposium on Visualization* (2005), pp. 125-132.

[8] M. Novotny and H. Hauser Outlier-preserving Focus+Context visualization in parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), (2006), pp. 893-900.

[9] M. Novotny. Visually effective information visualization of large data. In *Proceedings of the 8th Central European Seminar on Computer Graphics* (2004), pp. 41-48.

[10] R. Rosenbaum, J. Zhi and B. Hamann. Progressive parallel coordinates, In *Proceedings of the IEEE Pacific Visualization Symposium* (2012), pp. 25-32.