User-Centered Feature Space Transformation

1 Introduction

Most algorithms and computational tools require as input a feature-based representation of the data. Feature-spaces are usually generated by applying feature extraction mechanisms to raw data so as to identify attributes that characterize data instances. Feature extraction methods, however, are prone to generate irrelevant attributes that affect the accuracy of techniques operating on the data. Automatic and supervised feature (or attribute) selection methods are commonly used to filter out irrelevant attributes. As the name suggests, automatic feature selection methods avoid user intervention altogether, accomplishing the relevance analysis of attributes based primarily on statistical measures. In contrast, supervised schemes rely on training sets to perform the attribute selection, delegating to the user the task of picking out good representatives for the training process. Visualization-assisted interactive methods have become a trend in the context or automatic and supervised attribute manipulation. However, those methods do not provide intuitive mechanisms for the manipulation of data attributes.

We propose an interactive methodology to provide simple and intuitive mechanisms for attribute manipulation while allowing users to figure out changes in neighborhood structures during the interaction. The proposed technique relies on a combination of dimensionality reduction (DR) [3] and orthogonal linear mappings [1] to enable interactive feature space manipulation. Specifically, changes made by the user in the visual space are mapped back to the feature space so as to modify the distance between a subset of instances. Modified distances are then used to construct local linear mappings that transform the feature space according to user's guidance. Each local mapping is defined as an affine transformation obtained as the solution of an orthogonal minimization problem formulated in terms of the user-driven data. In order to modify the feature space according to user intervention, we propose a transfer function that maps the distance between control points (which can be manually positioned by the user) in the visual space to distances in the feature space. Data instances corresponding to control points in the feature space are then rearranged so as to cope with the new distances. The remaining instances are displaced in the feature space according to the position of their closest control points. A local orthogonal affine transformation is built to perform such displacement.

The main contributions of this work is: A novel mathematical and computational approach for transforming feature spaces. This new approach relies on force-based scheme and local affine transformations both operating in a high-dimensional feature space.

2 Proposed Methodology

The proposed methodology begins by selecting representative samples from a dataset (in our experiments these samples are randomly selected). Those representative samples are then projected into the visual space through a distance preserving technique that enables to visualize the neighborhood relation of the samples. The user can manipulate the provided layout by changing the position of representative samples, modifying thus the neighborhood structures in the visual space. After user manipulation, neighborhood structures in the visual space and in the feature space are not in agreement anymore. In order to restore the concordance between them, the set of samples are displaced in the feature space so as to minimize the difference between distances in both spaces. The final transformation of the feature space is performed by a family of local affine mappings built from the new position of representative samples.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

The feature space transformation is driven by user manipulation of sample points in the visual space. Interaction is typically initiated with a small set of samples to avoid visual clutter and reduce user effort. However, important structures and clusters may not be properly captured when using a few samples. If the manipulation of the initial set of samples does not result in the expected outcome, an user can successively add new samples to interact with.

3 Results

In order to show the effectiveness of our feature space manipulation approach we perform transformations in 7 distinct datasets. Our first experiment shows that the silhouette coefficient [2] of each dataset improves considerably after a few interaction cycles. Table 1, shows the original silhouette (second column) of each dataset used in our experiments. We use those values of silhouette as a basis for quantifying the effectiveness of our approach in improving cohesion and separation. Starting with $\sqrt{n}/2$ samples, which correspond, on average, to less than 1% of the number of instances in the datasets, the user manipulates the samples in the visual space so as to visually group instances belonging to the same class. Notice from the third column in Table 1 that silhouette improves in most cases after the first interaction cycle, that is, after the user to rearrange the initial samples in the visual space. In fact, expressive improvements can be seen in the shuttle and caltech datasets. Third and fourth columns of Table 1 show silhouette values after the second and third in teraction cycles , where $\sqrt{n}/2$ new samples were added in each cycle.

Table 1: Silhouette of the datasets after one, two and three interaction cycles.

Dataset	original	$\sqrt{n}/2$	$\sqrt{n}$	$3\sqrt{n}/2$
spam	0.0445	0.0504	0.1374	0.2328
wdbc	0.3412	0.4219	0.5131	0.5919
segmentation	0.2410	0.1763	0.3107	0.3036
shuttle	0.2879	0.5127	0.5775	0.6156
caltech	0.1190	0.2762	0.3364	0.3812
imageclef	0.0305	0.0960	0.1231	0.1228
msrcorid	0.1020	0.1750	0.2668	0.3361

The usefulness of our approach is shown in an image retrieval interactive attribute manipulation application. For the sake of comparison between the original and the transformed feature space, Figure 1(a) and Figure 1(b) present the 25 most similar images returned for the target image highlighted by a red border. For the query, 17 and 3 irrelevant images are returned using the original and the transformed space, respectively. Therefore, the precision increases from 72% to 96%, attesting the effectiveness of our approach. Figure 1(c) shows the precision versus recall plot. The precision obtained with the transformed space is higher and stabilizes in high levels when more images are recovered while it decreases when images are directly retrieved from the original space.

4 Conclusion

In this work we have proposed a novel approach for feature space transformation based on user manipulation of representative samples. Representative samples are mapped to the visual space via a DR technique. The projected data can then be manipulated to create groups of interest from which local transformations are defined. Our experiments have shown that the proposed approach improves considerably the cohesion and separation of groups. The proposed methodology aims at adding the user on the loop of data analysis, classification and clustering without overwhelming him/her with complex interactive interfaces.

References

[1] P. Joia, D. Coimbra, J. A. Cuminato, F. V. Paulovich, and L. G. Nonato. Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2563–2571, Dec 2011.

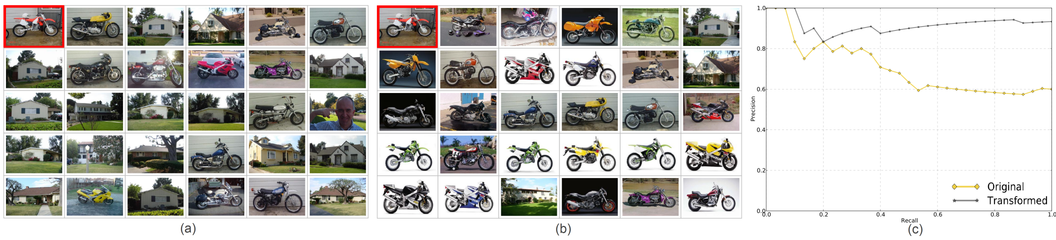

Figure 1: Images retrieved and the precision versus recall plot considering the original and transformed space. The precision worsens considerably if compared with the transformed space.

[2] Peter Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20(1):53–65, November 1987.

[3] D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. Visual interaction with dimensionality reduction: A structured literature analysis. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):241–250, Jan 2017.