Simultaneous Interaction with Dimension Reduction and Clustering Projections

John Wenskovitch, Michelle Dowling, Chris North
Virginia Tech
jw87@vt.edu,dowlingm@vt.edu,north@cs.vt.edu

ABSTRACT

Direct manipulation interactions on projections are often incorporated in visual analytics applications. These interactions enable analysts to provide feedback to the system, demonstrating relationships that the analyst wishes to find within the projection. However, determining the precise intent of the analyst is a challenge; when an analyst interacts with a projection, the system could infer a variety of possible interpretations. In this work, we explore interaction design considerations for the simultaneous use of dimension reduction and clustering algorithms to address this challenge.

CCS CONCEPTS

• **Human-centered computing → Visualization**; **Visual analytics**; *Visualization design and evaluation methods*;

KEYWORDS

Dimension reduction, clustering, interaction, visual analytics

ACM Reference Format:
John Wenskovitch, Michelle Dowling, Chris North. 2019. Simultaneous Interaction with Dimension Reduction and Clustering Projections. In *24th International Conference on Intelligent User Interfaces (IUI '19 Companion), March 17–20, 2019, Marina del Rey, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308557.3308718

1 INTRODUCTION

"What respect to what" was described as a usability issue with direct manipulations in interactive projections by Self et al [9]. This usability issue revolves around interpreting the analyst's intent accurately. That is, when the analyst moves an observation to a new position, what is that movement in relation to? Interactive projections are a popular feature in visual analytics applications [1–3, 5–8, 11]. As such, resolving this "with respect to what" problem is increasingly important in order to capture the intent of the analyst.

In previous work, we evaluated the design possibilities for the creation of projections that feature dimension reduction and clustering algorithms [10], and we proposed a cluster membership solution to "with respect to what," utilizing interactive clustering reassignment to communicate similarity relationships in the projection [11].

IUI '19 Companion, March 17–20, 2019, Marina del Rey, CA, USA

ACM ISBN 978-1-4503-6673-1/19/03.
https://doi.org/10.1145/3308557.3308718

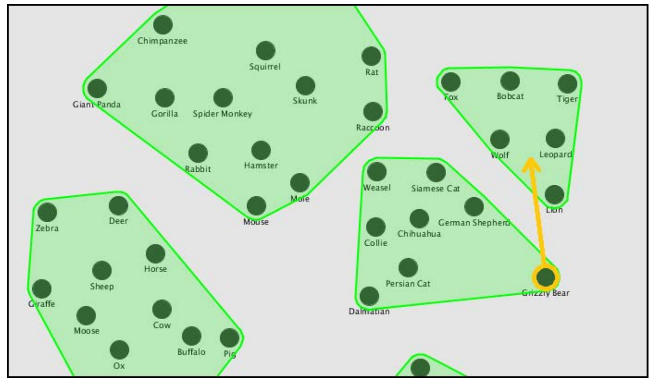

Figure 1: An analyst repositions the Grizzly Bear observation within the projection, indicated by the orange arrow.

This use of clustering is a natural choice, as implicit clusters often form in projections that display similarity relationships.

However, ambiguity in the interpretation of these interactions does still exist after explicit clustering has been introduced, as described next in our motivating example. In this work, our goal is to begin to explore the interaction space for the simultaneous use of dimension reduction and clustering algorithms, particularly in interactive projections that feature a learning component.

2 MOTIVATING EXAMPLE

To motivate our discussion of this interaction space, consider the example shown in Figure 1. Here, an analyst is provided with a dimension-reduced projection of an animal dataset, positioned according to their attribute relationships. A clustering algorithm then groups the observations into discrete categories. After viewing the projection, the analyst repositions the Grizzly Bear, changing both its position and cluster assignment. With this simple interaction, the analyst could be trying to convey a number of possible intents.

The analyst may be considering only relationships between the animals in the projection, such as with respect to the starting position of the interaction ("the Grizzly Bear is not similar to the animals near the source") or the ending position of the observation. The analyst may also be limiting their comparison to a single observation, communicating a relationship with respect to just the closest observation ("the Grizzly Bear is most similar to the Lion").

Alternatively, the analyst may have mapped semantic meaning onto the cluster groupings in the projection, attempting to communicate a membership assignment update based on those groups ("the Grizzly Bear is a better fit in the Predators cluster than in the Pets cluster"). Such relationships could incorporate both the source and the target cluster, or perhaps a case where the target is irrelevant ("the Grizzly Bear appears to be an outlier in the Pets cluster and belongs elsewhere") or the source is irrelevant.

3 INTERACTION DESIGN CONSIDERATIONS

The preceding example suggests the following dimensions to consider when interpreting the intent of an interaction:

Interaction Target: The interaction could be applied to the observations, the clusters, or both.

Cardinality: The interaction could be applied to a variety of cardinalities: the nearest observation, the nearest *n* observations, all observations within a cluster, or all observations in the projection.

With Respect To What: Is the important relationship relative to other observations in the projection at the source of the interaction, the destination of the interaction, or both?

Level of Thinking: When performing the interaction, is the analyst is thinking high- or low-dimensionally? In other words, is the analyst merely altering the projection, or are they considering all properties of a group of observations?

Visual Design: Is the intent of the interaction influenced by the way that observations and clusters are encoded in the visualization? For example, using a boundary to delineate cluster membership may imply that dragging an observation across the boundary leads to a reclassification.

4 CAPTURING INTERACTIONS WITH A DATA FLOW REPRESENTATION

In the case of a visualization system that incorporates dimension reduction and clustering algorithms into the same interface, each of these algorithms represent a separate model in a multi-model pipeline sequence [1, 4]. The order of these models in the sequence can therefore change both the meaning and the behavior of the visualization. For the motivating example, running the dimension reduction computation before the clustering computation implies that a dataset is reduced from the high-dimensional space to the low-dimensional space, after which the clustering algorithm is performed on the low-dimensional data.

The intricacies of interactions can be demonstrated by showing how data flows through the pipeline. For example, the pipelines displayed in Figure 2 shows three different possible models for the motivating example. The interaction directions modeled by these pipelines reflect the system interpretation of a direct manipulation interaction. In the first, a Clustering Model detects a change in cluster membership, followed by learning distances with a Dimension Reduction Model, indicating the importance of the cluster membership change to the interaction. In the second, the model order is swapped, and the distance computation occurs prior to the cluster membership computation, indicating the importance of the distance change to the interaction. In the third, the greyed-out interaction computation of the Dimension Reduction model indicates that only cluster membership alterations are considered by this system. Still further system and interaction designs could be supported by altering the flow of data in the pipeline.

ACKNOWLEDGMENTS

This work was funded by NSF grant IIS-1447416.

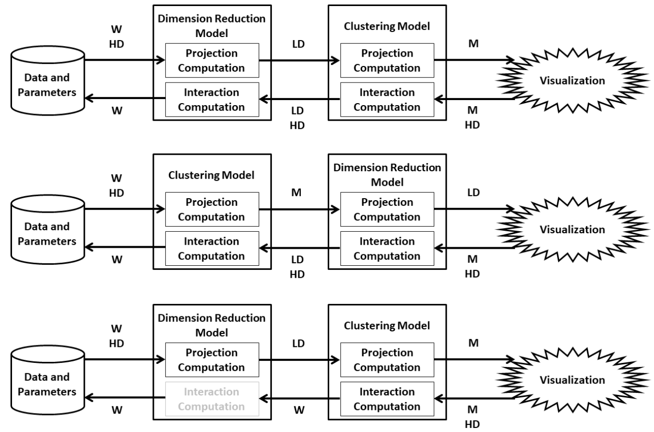

Figure 2: Three representations of data flow based on the motivating example. W=dimension weights, HD=high-dimensional data, LD=low-dimensional data, M=cluster membership.

REFERENCES

[1] L. Bradel, C. North, L. House, and S. Leman. 2014. Multi-model semantic interaction for text analytics. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 163–172. https://doi.org/10.1109/VAST.2014.7042492

[2] E. T. Brown, J. Liu, C. E. Brodley, and R. Chang. 2012. Dis-function: Learning distance functions interactively. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 83–92. https://doi.org/10.1109/VAST.2012.6400486

[3] Michelle Dowling, John Wenskovitch, J.T. Fry, Scotland Leman, Leanna House, and Chris North. 2019. SIRIUS: Dual, Symmetric, Interactive Dimension Reductions. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan 2019), 172–182. https://doi.org/10.1109/TVCG.2018.2865047

[4] Michelle Dowling, John Wenskovitch, Peter Hauck, Adam Binford, Nicholas Polys, and Chris North. 2018. A Bidirectional Pipeline for Semantic Interaction. In *Proceedings of the Workshop on Machine Learning from User Interaction for Visualization and Analytics (IEEE VIS 2018)*. 11.

[5] Alex Endert, Patrick Fiaux, and Chris North. 2012. Semantic interaction for sensemaking: inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2879–2888.

[6] Paulo Joia, Danilo Coimbra, Jose A Cuminato, Fernando V Paulovich, and Luis G Nonato. 2011. Local affine multidimensional projection. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2563–2571.

[7] G. M. H. Mamani, F. M. Fatore, L. G. Nonato, and F. V. Paulovich. 2013. User-driven Feature Space Transformation. *Computer Graphics Forum* 32, 3pt3 (2013), 291–299. https://doi.org/10.1111/cgf.12116

[8] F.V. Paulovich, D.M. Eler, J. Poco, C.P. Botha, R. Minghim, and L.G. Nonato. 2011. Piecewise Laplacian-based Projection for Interactive Data Exploration and Organization. *Computer Graphics Forum* 30, 3 (2011), 1091–1100. https://doi.org/10.1111/j.1467-8659.2011.01958.x

[9] Jessica Zeitz Self, Radha Krishnan Vinayagam, J. T. Fry, and Chris North. 2016. Bridging the Gap Between User Intention and Model Parameters for Human-in-the-loop Data Analytics. In *Proceedings of the Workshop on Human-In-the-Loop Data Analytics (HILDA '16)*. ACM, New York, NY, USA, Article 3, 6 pages. https://doi.org/10.1145/2939502.2939505

[10] John Wenskovitch, Ian Crandell, Naren Ramakrishnan, Leanna House, Scotland Leman, and Chris North. 2018. Towards a Systematic Combination of Dimension Reduction and Clustering in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics Proceedings of the Visual Analytics Science and Technology 2017* 24, 1 (Jan 2018), 131–141. https://doi.org/10.1109/TVCG.2017.2745258

[11] John Wenskovitch and Chris North. 2017. Observation-Level Interaction with Clustering and Dimension Reduction Algorithms. In *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics (HILDA'17)*. ACM, New York, NY, USA, Article 14, 6 pages. https://doi.org/10.1145/3077257.3077259