Interpretation and Hierarchical methods for Dimensionality Reduction

Wilson E. Marcílio-Jr[§], Danilo M. Eler
Department of Mathematics and Computer Science
São Paulo State University
Presidente Prudente, São Paulo, Brazil
Emails: {wilson.marcilio, danilo.eler}@unesp.br

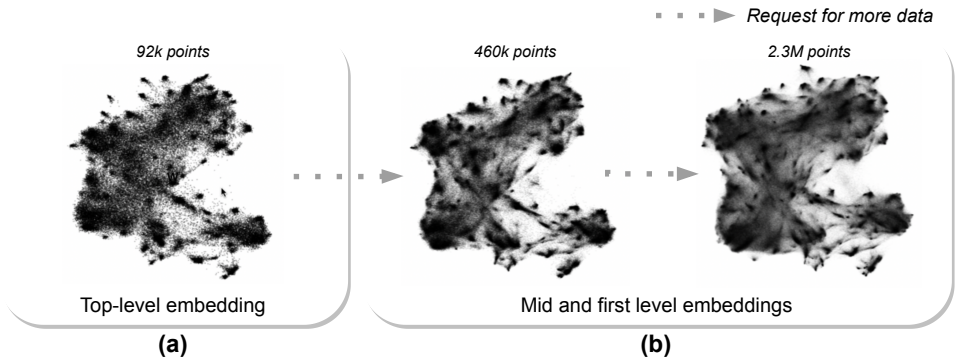

Fig. 1: 2.3 million papers projected using HUMAP. (a) Top-level embedding presented to the user; (b) Second and first levels visualized according to information demand.

***Abstract*—High-dimensional data analysis is a ubiquitous task in practical and research activities. Dimensionality Reduction (DR) techniques are usually employed as they map high-dimensional data to lower spaces and allow for knowledge discovery. This thesis focuses on the interpretability and representation aspects of non-linear DR approaches' output, such as t-SNE and UMAP. That is, we propose methods for interpreting and hierarchically learning embeddings. To accomplish these goals, the following main research activities were carried out, representing separate but interconnected works: (1) a sampling method in visual space ($\mathbb{R}^2$) that can preserve class boundary structures while keeping outliers visible; (2) a technique for understanding cluster formation by leveraging statistical tests on the feature values after dimensionality reduction; (3) we advance the state-of-the-art by adapting SHAP to explain cluster formation after dimensionality reduction; (4) a novel hierarchical DR technique that employs an adaptive kernel for global/local neighborhood learning while preserving context across embeddings.**

I. Introduction

Dimensionality Reduction (DR) techniques are widely used in a wide range of scientific fields. Embeddings encoded by scatterplots aid in the analysis of high-dimensional datasets from different scientific domains—for example, see the 2.3 million papers [1] embedded in $\mathbb{R}^2$ using HUMAP [2]. However, non-linear techniques make it difficult to understand embeddings according to feature contributions for embedding formation. Scatterplot representations of embeddings become visually cluttered as the dataset size increases, deceiving understanding of the relationship among data points—a limitation that can be addressed by summarizing the data during analysis using sampling [3]–[5] or visual summarization approaches. Interpretation techniques can also improve the interpretability of embeddings. While many approaches have been proposed for explaining and interpreting the decisions of common machine learning models [6], dimensionality reduction interpretation has been performed with enriching layout techniques [7], in which visualizations of feature distributions and other attribute information aid analysts in making sense of clusters and other patterns perceived in the visual space ($\mathbb{R}^2$)—simple statistics, on the other hand, cannot fully provide the differentiation among the clusters. Other approaches' limitations can be summarized as prohibitive runtime execution of high-dimension datasets [8] or the inability to extend the analysis for different DR techniques [9].

Finally, traditional DR techniques operate at a single level of detail, representing the relationship between data points in a single low-dimensional scatterplot representation—which limits analysis when important information is prioritized and details are added as needed. This limitation can be circumvented using hierarchical DR (HDR) techniques, which project subsets of the input dataset and allow the user to select where to focus during the analysis. The idea behind HDR techniques is to project dominant structures 1 (**a**) and let users decide when to refine for more information (**b**). The main problems with previous HDR techniques are their inability to scale [10], [11] and the non-preservation of the mental map [12].

To address these discussed research gaps, we formulated the

[§]This work relates to a Ph.D. thesis.

following hypothesis.

A. Main Hypothesis

Hierarchical exploration and interpretation of embeddings improve the analysis power of DR xtechniques by assisting in identifying and emphasizing various facets present in data. While mental map preservation in hierarchical embeddings allows for context preservation throughout exploration, dimensionality reduction interpretation explains feature contributions to layout formation in the visual space.

II. Related Works

A. Sampling for Dimensionality Reduction

Well-known sampling techniques such as Knuth's [4] and Reservoir's [3] rely on probabilistic mechanisms to extract samples that depict the dominant structures of a dataset. On the other hand, deterministic methods, like CSM [5], select representatives by computing correlation indices through matrix decomposition. In summary, the idea when sampling for DR is to maintain the relationship seen in the whole dataset. HSNE [12] technique selects dominant structures using Monte Carlo sampling on a Finite Markov Chain. Nguyen and Song [13] adopt centrality cluster-based sampling approaches, which employ centrality measures to obtain more informed samples than random sampling methods.

B. Embedding Interpretation

The literature presents studies that support the analysis of DR results via the inspection of global information [5], [14]–[17], emphasizing the importance of features imposed on the embedded space. These contributions, however, do not highlight the clusters' distinct characteristics. ccPCA [8] uses contrastive PCA [18] to find the unique characteristics for each cluster—its main limitation stems from PCA, which is the prohibitively run-time execution time for high-dimensional datasets. ContraVis [9] is only applicable to textual data and cannot assist in the interpretation of DR layouts because it is already a dimensionality reduction approach.

Studies on feature values [15], [17], [19], for example, do not account for the dimensionality reduction process and instead concentrate on the reduced low-dimensional space. Other works [5], [14] retain feature importance through the principal components (PC). The PCs produce biased results for classes with high variation [5], and their inability to focus on local information [8] hinders cluster-oriented analysis.

C. Hierarchical Dimensionality Reduction

DR techniques have traditionally operated at a single level of detail. These layouts, however, may conceal important data characteristics, such as differences between samples within a cluster, which provide an overview of the structures and benefit analysis when focusing on important information first.

Hierarchical DR approaches try to reduce this problem by providing embeddings according to information demand. HiPP [10] creates an overview of the dataset and guides users through further analysis by using nodes. It uses LSP [20] as backbone, which does not scale well for datasets with thousands of points. HSNE [12] can successfully depict manifolds and deal with thousands of data points/dimensions. However, as the analyst descends the hierarchy, HSNE fails to preserve the user's mental map. Multiscale PHATE [11] is able to handle massive datasets describing continuous phenomena but does not work well with higher dimensions.

III. Methods

A. SADIRE[1]

To deal with visual clutter of embeddings in $\mathbb{R}^2$, we propose a sampling technique called SADIRE (SAmpling from DImensionality REduction).

SADIRE [21] considers an embedding $X \subset \mathbb{R}^2$. It selects candidates for the representative set using a grid structure on the projection plane. Then, it eliminates redundant data instances from the sampled set.

The first stage of the sampling strategy involves sampling from an $\mathbb{R}^2$ projection. SADIRE divides the projected space by generating a grid with cells of size $k \times k$ (note that k is measured in pixels). The sampled set is then created by using only the cells with at least one data point, from which a number $m < n$ of data points can be retrieved, where n is the number of data points in each cell. While our explanation and experiments are geared toward $m = 1$, keep in mind that using $m > 1$ conveys the density of the clusters.

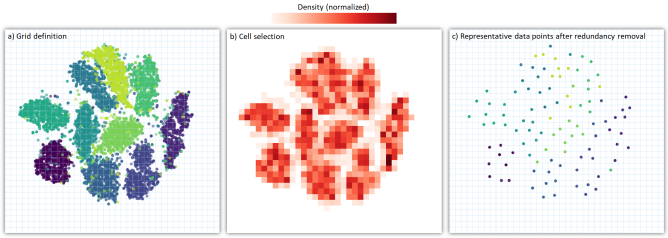

Fig. 2: Defining the grid and selecting samples. (a) A grid with cells of size k by k is used to divide the $\mathbb{R}^2$ space; (b) Only cells with at least one data point are retrieved; (c) The resulting representative set after redundancy removal.

Second, windowing process is used in the second main step of the algorithm to reduce redundancy from neighboring cells. For a window size of $\alpha = 3$, the density information stored in the first step is used to keep useful information about the embedding structures while reducing redundancy. Starting with the densest cell, all of its neighboring cells are eliminated from the possibility in order to be included in the resulting representative set. This process is repeated for all active cells from the most to the least densest that were not removed.

SADIRE is the backbone for a exploration approach called ExplorerTree [22], published in the Big Data Research journal.

[1]https://github.com/wilson_jr/SADIRE

B. cExpression

The first contribution for helping with the interpretation of DR results considers the features that are most distinctive for clusters. The technique, called *cExpression* [23], consists of comparing value distributions of different clusters using t-test to determine if the differences between these distributions are significant enough to describe clusters. Given a dataset X, let X_f^c and $X_f^{c'}$ be the feature f values for data observations in the cluster c and the feature f values for data observations **not** in the cluster c (c')—c and c' are visual space clusters from which users want to find unique characteristics. For each feature (f) and cluster (c), were compare X_f^c and $X_f^{c'}$ by extracting t-statistic and p-value. Then, these two numbers give the information about how well f describes c.

Fig. 3 depicts the idea. The distributions (**b**) of values for the cluster and feature are generated using the dataset (**a**)—red encodes the feature distribution of data samples within the cluster, while gray encodes the distribution of data samples outside the cluster. The t-test is used in these distributions (**c**) to determine the probability (p-value, **d**) to observe the t-score when the distributions are similar.

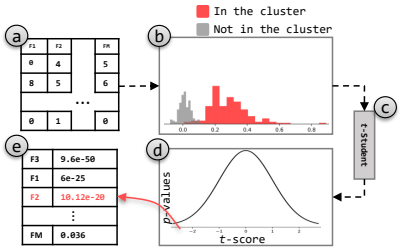

Fig. 3: Comparing feature distributions with t-test.

Fig. 4 (left) depicts a UMAP projection of the Vertebral dataset, which includes 310 data instances described by six biomechanical features derived from the shape and orientation of the pelvis and lumbar spine: *pelvic incidence*, *pelvic tilt*, *lumbar lordosis angle*, *sacral slope*, *pelvic radius*, *d. of spondylolisthesis*. The patients of interest have spondylolisthesis, a spine disorder in which a bone slides forward over the bone below it. To investigate cExpression on this task, we look at the distribution of values for the key features throughout the projection. The two most important ones retrieved by cExpression are shown in Fig. 4 (right). Note how *degree of spondylolisthesis* is a very distinctive feature for the class.

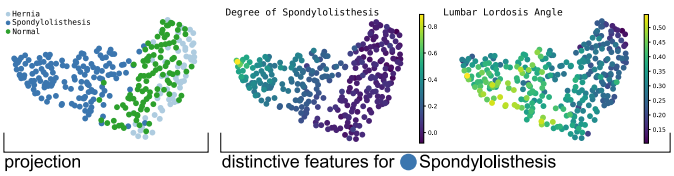

Fig. 4: UMAP projection of the *Vertebral Column* dataset and two distinctive features for the Spondylolisthesis class.

1) Visualizing the feature importance: We visualize the importance of the features for a cluster as their position over a horizontal axis while providing distribution plots for each visualized feature. The visualization design is shown in Fig. 5, where the color hue for the distribution plots corresponds to an arbitrary cluster ●.

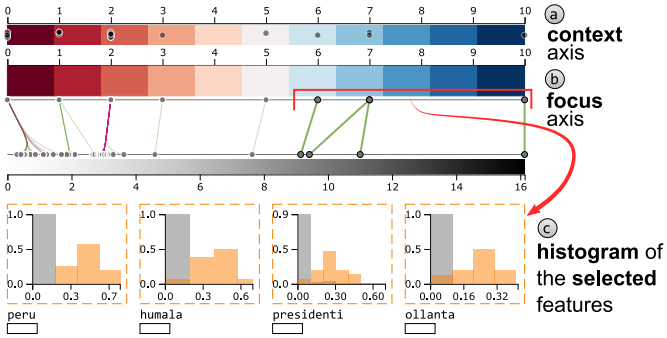

Fig. 5: A visual summary of the most distinctive features for a cluster.

The information presented on axes (**a**) and (**b**) in Fig. 5 is the same, namely the number of decimal places after zero for each p-value. The gray circles in the context axis (**a**) represent the analyzed features, and their horizontal position encodes the p-values. A red-to-blue color scale assists users in determining feature confidence—blueish colors encode greater confidence (or lower p-values). We encode the difference between distributions (t-scores) using another axis in addition to the confidence in the difference represented by the p-values. The relationship between these statistical variables is depicted by line segments, and the color represents the t-score signal (pink for negative and green for positive).

Note that the visualization of Fig 5 is highly interactive, and we invite the readers to read the paper [23] for more details.

C. ClusterShapley[2]

One step further in interpreting DR results can be understanding the contribution for cluster formation for embedding in $\mathbb{R}^2$. ClusterShapley [24], our DR interpretation approach, is based on explaining the contribution of feature values for the cluster cohesion through the use of Shapley values [25].

1) Shapley estimation: Shapley values are estimated for each data sample after the cluster definition. That is, we define a model f that returns the prediction similarities for a data sample x. We measure the distance from data samples x to each cluster centroid to return the prediction similarities for a data sample x. The distances are then converted into similarities using L1 normalization—lower values indicate proximity in the visual space and thus cluster cohesion.

The estimated Shapley values correspond to the contribution of each feature to the embedding. That is, a feature with high absolute Shapley value contributes significantly to the cluster formation of the projected dataset. Each data point

[2]https://github.com/wilson_jr/ClusterShapley

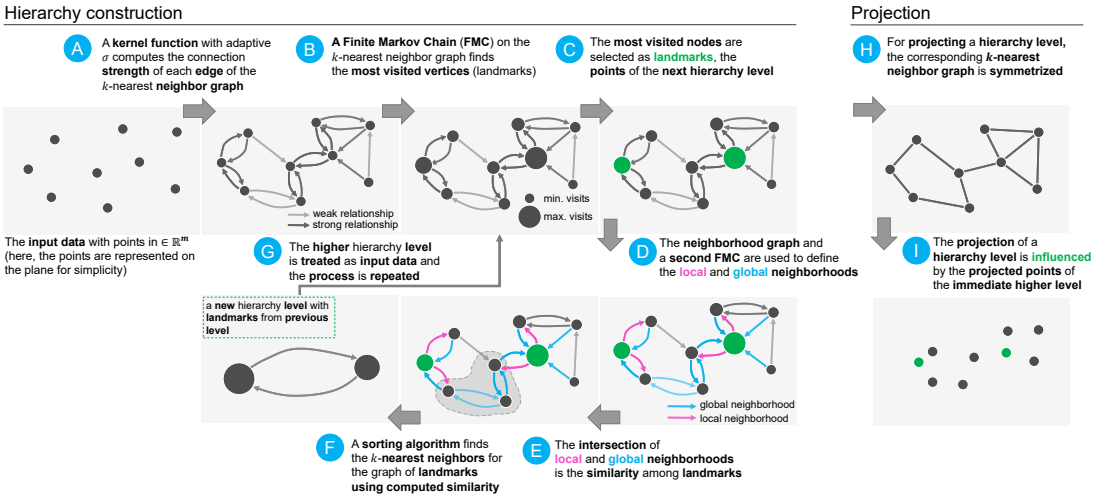

Fig. 6: The hierarchy is constructed from bottom to top. First, a k-nn graph approximates the connection strength among data points (A). Then, random walks are used to find landmarks (B), the points of superior levels (C). The representation neighborhood—local and global relationships (D)—is used for computing the similarity among landmarks (F). In the embedding (H), we use the influence of superior levels for the mental map preservation (I).

used for Shapley value estimation in this case contains the corresponding Shapley value. Negative Shapley values indicate that the feature contributes to cluster formation, while positive Shapley values indicate that the feature does not.

2) Visual Component for Interpretation: Fig. 7 shows dot plots of the four most important features ordered by their influence to the cluster. It is worth noting that the more the absolute Shapley values (encoded as horizontal position) deviate from zero, the more influence the feature has in characterizing the cluster.

The coordination between the scatterplot and the dot plot assists users in determining which features contributed to cluster cohesion and which features contributed to cluster dispersal during the detailed inspection. The selected cluster in Fig. 7 very cohesive due to the contribution of *petal length* (see negative Shapley values), that is, data points with lower values for this feature are very distinctive from the others.

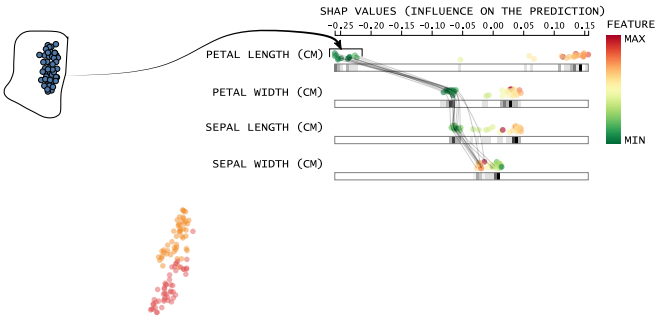

Fig. 7: Each line segment represents a point in the scatterplot.

D. HUMAP[3]

The fourth contribution advances the state of the art of hierarchical embeddings by providing means for large scale and mental map-preserving hierarchical dimensionality reduction.

The method, called HUMAP [2], is composed of two major components: hierarchy construction and projection (Fig. 6). In the former, we use a similarity measure among landmarks to impose a hierarchical structure on the dataset. In the latter, we embed the hierarchy levels based on the user's demand for more detailed information. All of the steps (A-G) of these components are depicted in Fig. 6.

The first step in building a hierarchy from bottom to top is to use a kernel function to determine the connection strengths (local affinities) of a k-nearest neighbor graph of data points in the high-dimensional space $\mathbb{R}^m$ (step A). Then, as in previous work [12], we use the Finite Markov Chain (FMC) to find the most visited nodes (step B), which are the superior level landmarks (step C). The FMC procedure is also used to build a neighborhood structure for high hierarchy levels (level > 1) and to encode local and global neighborhood information for each landmark (steps (D) and (E)). Finally, using the computed similarity to define a new hierarchy level (step F), a new neighborhood graph is created (step G).

The neighborhood graph is first symmetrized (step H) before projecting hierarchy levels. Except for the top hierarchy level, the low-level projection is influenced by the low-dimensional positions (I) of data points in higher levels for mental map preservation. This is performed through the modification the Stochastic Gradient Descent algorithm, where we limit the movement of already projected data points and influence the projection of new points.

[3]https://github.com/wilson_jr/humap

Thus, instead of presenting the embedding with the whole dataset, HUMAP provides an overview as in Fig. 1 (**a**). Users can request for more data for certain clusters or for the whole projection, as shown for the middle and right embeddings.

IV. Results

A. SADIRE

To evaluate SADIRE, we employ the concept of representativeness [26] given by the metrics redundancy and coverage. Here, we report the results for the *Coverage* metric against CSM, Knuth, and Reservouir upon several datasets—see the paper [21] for the embedding results and the other metrics.

B. Coverage

To model the *Coverage* of a sampled set, we take the mean and std. deviation of the distance among sampled points and non-sampled points. Thus, lower mean and std. deviation are better and encode the information that a non-representative point is close to a sampled point.

Fig. 8 shows the results for coverage metric. For a few cases, SADIRE presents (for *fiber* and *photographers-36k*, for example) higher results than *Reservoir* and *Knuth*'s. This is due to the fact that the majority of the points selected from them are concentrated in denser areas—mean of lower distances gives lower values. However, the standard deviation shows that a few points are too far from a representative (sampled point), which implies in bad results for *Knuth* and *Reservoir* techniques. *CSM* results are related to its formulation, in which there is no way to control where to focus for extracting sample points.

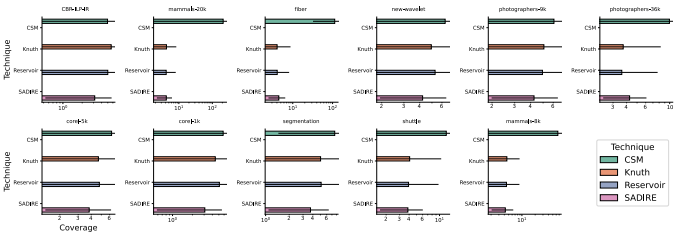

Fig. 8: *Coverage* metric to assess if a sampled set is able to represent all data structures.

C. cExpression

To investigate more about the class of patients with Spondylolisthesis (■) using cExpression, the selection of the features with high confidence on their distinction aspects (that is, features whose p-values are lower than 10^{-5}) shows that the distribution plots present the values skewed to the right, as illustrated in Fig. 9. The distribution plots also show how data instances of patients with Spondylolisthesis present higher values for *D. Spondylolisthesis*, *L. Angle*, *P. Incidence*, and *S. Slope*. According to [27], these features are prone to be higher in patients with developmental spondylolisthesis.

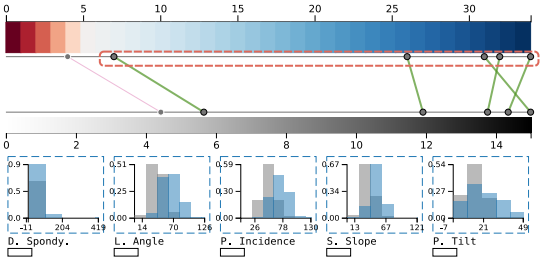

Fig. 9: The class of patients with Spondylolisthesis shows greater values for all of the features, except for Pelvic Radius.

D. ClusterShapley

Continuing with the same dataset, we now employ ClusterShapley to understand the contributions of the dataset for the embedding in $\mathbb{R}^2$. Fig. 10 shows the projected instances using t-SNE, colored based on the ground truth classes: class 2 ■ for normal patients, class 0 ■ for patients with Hernia, and class 1 ■ for patients with Spondylolisthesis. Again, there is separation of class 1 ■ among the remaining data points.

The dot plot shows that the degree of *spondylolisthesis* is a determinant factor for the presence of Spondylolisthesis. By coordinating the scatterplot and the dot plot for class 1 ■, we can see that its data points have higher values for the *degree of spondylolisthesis* with negative Shapley values—which contribute to cluster formation. The other three most important features also contributed to clustering cohesion.

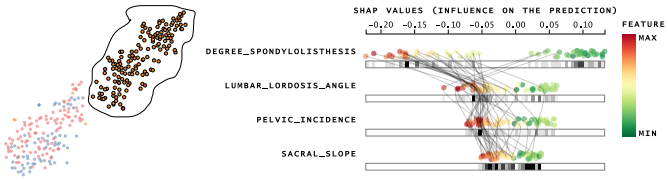

Fig. 10: Correlation between feature values and their contribution to cluster formation.

E. HUMAP

We compared HUMAP against HSNE, Multiscale PHATE and UMAP upon various evaluation metrics and different datasets. Here, we report the two most important results.

1) DEMaP: To analyze how well the techniques convey manifolds, clusters, and other high-dimensional space structures, we use the DEMaP metric [28]. Fig. 11 shows that HUMAP presented better results than HSNE and UMAP on level 2 and HSNE-GPU on level 0 for the MNIST and FMNIST datasets. For the mammals dataset, HUMAP shows higher values when embedding the whole dataset, providing evidence of our technique's stability across hierarchical levels.

2) Mental map preservation: To quantify the mental map preservation, we use Procrustes Analysis on subsequent hierarchy levels. That is, for the hierarchy levels i and $i+1$, we retrieved the points present in both i and $i+1$ and used them to compute disparity between the sets of points. Fig. 12

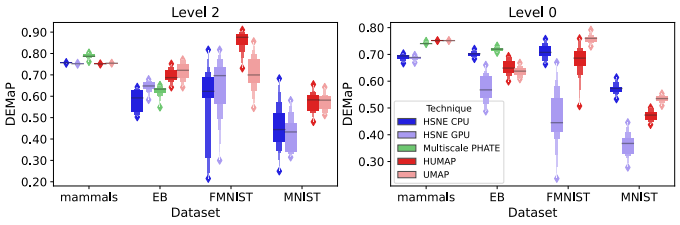

Fig. 11: Evaluation of techniques' ability to represent complex structures such as clusters, manifolds, and other relationships.

shows that HUMAP is superior over the other techniques at preserving the mental map (the lower disparity the better).

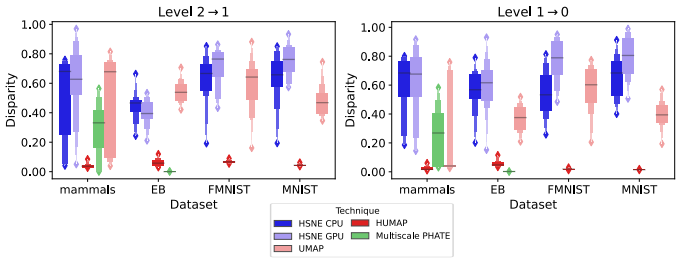

Fig. 12: Mental map preservation measured with Procrustes Analysis.

V. Final Remarks

In this section, we discuss practical implications of our work and summarize this doctorate's scientific output.

A. Practical Implications

- **cExpression**: a method for differentiating cluster formation. While this concept is broadly applicable to cell-type annotation [29], the cExpression visualization component enhances the analysis' visual scalability. Annotation and class definition for textual datasets is an important practical application of cExpression.
- **ClusterShapley**: we detail a technique that provides addictive explanations for the organization of the embeded layout after dimensionality reduction. ClusterShapley can also be used to understand the separation ability of deep learning models after projecting bottleneck layers. Finally, ClusterShapley has also been used in the material science domain [30].
- **HUMAP**: we present a novel hierarchical dimensionality reduction technique. HUMAP provides an overview of the dataset, making analysis easier and emphasizing intracluster characteristics that are difficult to see using traditional projections. The discovery of cell types and subtypes is a direct application of this characteristic [31].

B. Scientific production

1) First-authored publications:

1. *SADIRE: a context-preserving sampling technique for dimensionality reduction visualizations*. WE Marcilio-Jr, DM Eler. **Journal of Visualization** 23 (6), 999-1013.
2. *Visual analytics of COVID-19 dissemination in São Paulo state, Brazil*. WE Marcílio-Jr, DM Eler, RE Garcia, RCM Correia, RMB Rodrigues. **Journal of Biomedical Informatics** 117, 103753.
3. *Explaining dimensionality reduction results using Shapley values*. WE Marcílio-Jr, DM Eler. **Expert Systems with Applications** 178, 115020.
4. *Contrastive analysis for scatterplot-based representations of dimensionality reduction*. WE Marcílio-Jr, DM Eler, RE Garcia. **Computers & Graphics** 101, 46-58.
5. *A hybrid visualization approach to perform analysis of feature spaces*. WE Marcílio-Jr, DM Eler, RE Garcia, RCM Correia, LF Silva. 17th **ITNG**.
6. *From explanations to feature selection: assessing shap values as feature selection mechanism*. WE Marcílio-Jr, DM Eler. 2020 33rd **SIBGRAPI**.
7. *Model-agnostic interpretation by visualization of feature perturbations*. WE Marcílio-Jr, DM Eler, F Breve. **arXiv preprint** arXiv:2101.10502.
8. *HUMAP: Hierarchical Uniform Manifold Approximation and Projection*. WE Marcílio-Jr, DM Eler, FV Paulovich, RM Martins. a**rXiv preprint** arXiv:2106.07718.
9. *Overlap Removal of Dimensionality Reduction Scatterplot Layouts*. GM Hilasaca, WE Marcílio-Jr, DM Eler, RM Martins, FV Paulovich. **arXiv preprint** arXiv:1903.06262.

2) Co-authored publications:

10. *Visual approach to support analysis of optimum-path forest classifier*. DM Eler, MP Batista, RE Garcia, DR Pereira, WE Marcílio-Jr, 2019 8th **BRACIS**, 777-782.
11. *A class-based evaluation approach to assess multidimensional projections*. J Teixeira, **WE Marcílio-Jr**, DM Eler, A Artero, B Brandoli. 2020 24th **International Conference Information Visualisation** (IV), 174-181.
12. *Entropy-Based Filter Selection in CNNs Applied to Text Classification*. R Bezerra de Menezes Rodrigues, **WE Marcílio-Jr**, DM Eler. **BRACIS**, 497-510.
13. *PlaceProfile: Employing Visual and Cluster Analysis to Profile Regions based on Points of Interest*. R Christófano, **WE Marcílio-Jr**, DM Eler. Proceedings Of The 23rd **Intern. Conf. On Enterprise Information Systems**.
14. *Analyzing Accessibility Reviews Associated with Visual Disabilities or Eye Conditions* ADA Oliveira, PSH Dos Santos, **WE Marcílio-Jr**, W Aljedaani, DM Eler, MM Eler. Proceedings of the 2023 **CHI Conference on Human Factors in Computing Systems – Honorable Mention**.

Acknowledgements

This work was supported by the agencies: *Fundação de Ampara à Pesquisa do Estado de São Paulo (FAPESP)* [#2018/17881-3, #2018/25755-8]; and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)*, process #88887.487331/2020-00.

References

[1] "Macrocosm.so," https://alex.macrocosm.so/download, accessed: July 7, 2023.

[2] W. E. Marcílio-Jr, D. M. Eler, F. V. Paulovich, and R. M. Martins, "Humap: Hierarchical uniform manifold approximation and projection," *ArXiv e-prints*, Feb. 2021.

[3] J. S. Vitter, "Random sampling with a reservoir," *ACM Trans. Math. Softw.*, vol. 11, no. 1, pp. 37–57, Mar. 1985.

[4] D. E. Knuth, *The Art of Computer Programming, Volume 2 (3rd Ed.): Seminumerical Algorithms*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1997.

[5] P. Joia, F. Petronetto, and L. G. Nonato, "Uncovering representative groups in multidimensional projections," *Computer Graphics Forum*, vol. 34, pp. 281–290, 2015.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should I trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 2016, pp. 1135–1144.

[7] L. G. Nonato and M. Aupetit, "Multidimensional projection for visual analytics: Linking techniques with distortions, tasks, and layout enrichment," *IEEE Transactions on Visualization and Computer Graphics*, p. 1, 2018.

[8] T. Fujiwara, O.-H. Kwon, and K.-L. Ma, "Supporting analysis of dimensionality reduction results with contrastive learning," *IEEE Trans. Vis. and Comp. Graph.*, vol. 26, pp. 45–55, 2019.

[9] T. Le and L. Akoglu, "Contravis: Contrastive and visual topic modeling for comparing document collections," in *The World Wide Web Conference*, ser. WWW '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 928–938.

[10] F. V. Paulovich and R. Minghim, "Hipp: A novel hierarchical point placement strategy and its application to the exploration of document collections," *IEEE Transactions on Visulization and Computer Graphics*, vol. 14, no. 6, pp. 1229–1236, 2008.

[11] M. e. a. Kuchroo, "Multiscale phate exploration of sars-cov-2 data reveals multimodal signatures of disease," *bioRxiv*, 2020.

[12] N. Pezzotti, T. Höllt, B. Lelieveldt, E. Eisemann, and A. Vilanova, "Hierarchical stochastic neighbor embedding," *Comput. Graph. Forum*, vol. 35, no. 3, pp. 21–30, Jun. 2016.

[13] T. T. Nguyen and I. Song, "Centrality clustering-based sampling for big data visualization," *International Joint Conference on Neural Networks (IJCNN)*, pp. 24–29, 2016.

[14] C. Turkay, A. Lundervold, A. J. Lundervold, and H. Hauser, "Representative factor generation for the interactive visual analysis of high-dimensional data," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2621–2630, 2012.

[15] D. B. Coimbra, R. M. Martins, T. T. Neves, A. C. Telea, and F. V. Paulovich, "Explaining three-dimensional dimensionality reduction plots," *Information Visualization*, vol. 15, no. 2, pp. 154–172, 2016.

[16] J. Stahnke, M. Dörk, B. Müller, and A. Thom, "Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions," *IEEE Trans. on Vis. and Comp. Graph.*, vol. 22, pp. 629–638, 2016.

[17] L. de Carvalho Pagliosa, P. A. Pagliosa, and L. G. Nonato, "Understanding attribute variability in multidimensional projections," in *29th Conf. Graphics, Patterns and Images, SIBGRAPI, 2016, Sao Paulo, Brazil, October 4-7, 2016*, 2016, pp. 297–304.

[18] H. Abdi and D. Valentin, "Multiple correspondence analysis," *Encyclopedia of Measurement and Statistics*, pp. 651–657, 2007.

[19] W. Marcilio-Jr, D. Eler, R. Garcia, R. Correia, and L. F. Silva, "A hybrid visualization approach to perform analysis of feature spaces," *International Conference on Information Technology–New Generations*, vol. 1134, 2020.

[20] F. V. Paulovich, L. G. Nonato, M. Rosane, and H. Levkowitz, "Least square projection: A fast high-precision multidimensional projection technique and its application to document mapping," *IEEE Transactions on Visulization and Computer Graphics*, vol. 3, pp. 564–575, 2008.

[21] W. E. Marcílio-Jr and D. M. Eler, "Sadire: a context-preserving sampling technique for dimensionality reduction visualizations," *Journal of Visualization*, vol. 23, pp. 999–1013, 2020.

[22] W. E. Marcílio-Jr, D. M. Eler, F. V. Paulovich, J. F. Rodrigues-Jr, and A. O. Artero, "Explorertree: A focus+context exploration approach for 2d embeddings," *Big Data Research*, vol. 25, p. 100239, 2021.

[23] "Contrastive analysis for scatterplot-based representations of dimensionality reduction," *Computers & Graphics*, vol. 101, pp. 46–58, 2021.

[24] W. E. Marcílio-Jr and D. M. Eler, "Explaining dimensionality reduction results using shapley values," *Expert Systems with Applications*, vol. 178, p. 115020, 2021.

[25] L. Shapley, "A value for n-person games, vol ii of contributions to the theory of games," 1953.

[26] B. Ma, Q. Wei, and G. Chen, "A combined measure for representative information retrieval in enterprise information systems," *Journal of Enterprise Information Management*, vol. 24, pp. 310–321, nov 2011.

[27] H. Labelle, P. Roussouly, E. Berthonnaud, J. Dimnet, and M. O'Brien, "The importance of spino-pelvic balance in l5-s1 developmental spondylolisthesis: A review of pertinent radiologic measurements," *Spine*, vol. 30, pp. 27–34, 2005.

[28] K. Moon, D. van Dijk, and Z. e. a. Wang, "Visualizing structure and transitions in high-dimensional biological data," *Nat Biotechnol*, pp. 1482–1492, 2019.

[29] M. D. Luecken and F. Theis, "Current best practices in single-cell rna-seq analysis: a tutorial," *Molecular Systems Biology*, vol. 15, 2019.

[30] Y. Pan, H. Hou, and Y. Zhao, "A rapid and interpretable feature screening workflow for high-entropy alloys," Available at SSRN: https://ssrn.com/abstract=4464727, 2023.

[31] D. Lähnemann, J. Köster, and E. e. a. Szczurek, "Eleven grand challenges in single-cell data science," *Genome Biol*, vol. 31, 2020.