ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/336777807

Towards Enhancing RadViz Analysis and Interpretation

Conference Paper · October 2019

DOI: 10.1109/VISUAL.2019.8933775

CITATIONS
0

READS
137

5 authors, including:

Marco Angelini
Sapienza University of Rome
53 PUBLICATIONS **259** CITATIONS

SEE PROFILE

Graziano Blasilli
Sapienza University of Rome
9 PUBLICATIONS **19** CITATIONS

SEE PROFILE

Simone Lenti
Sapienza University of Rome
13 PUBLICATIONS **28** CITATIONS

SEE PROFILE

Alessia Palleschi
Sapienza University of Rome
3 PUBLICATIONS **1** CITATION

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project PROMISE View project

Project Progressive Visual Analytics View project

All content following this page was uploaded by Marco Angelini on 24 October 2019.

The user has requested enhancement of the downloaded file.

Towards Enhancing RadViz Analysis and Interpretation

Marco Angelini* Sapienza University of Rome

Graziano Blasilli* Sapienza University of Rome

Simone Lenti* Sapienza University of Rome

Alessia Palleschi* Sapienza University of Rome

Giuseppe Santucci* Sapienza University of Rome

Abstract

RadViz plots are commonly used to represent multidimensional data because they use the familiar notion of 2D points for encoding data elements, displaying the original data dimensions that act as springs for setting the x and y coordinates. However, this intuitive approach implies several drawbacks and produces misleading visualizations that can confuse the user, even while analyzing a single data point. The paper attacks this problem following the well known idea of changing the order of the dimensions and introducing ancillary visualizations to mitigate some of RadViz drawbacks. In particular, the paper defines the notion of *point optimal disposition* of the dimensions for a single data point, generalizes this concept to a set of data points, and proposes effective heuristics for dealing with the intractable problem of exploring all the $\frac{(n-1)!}{2}$ dispositions of the dimensions along the RadViz circumference. Additional views, visual quality metrics, and a circular grid superimposed on the RadViz complement the attribute reordering strategy and provide a better understanding of the actual plot of the data elements.

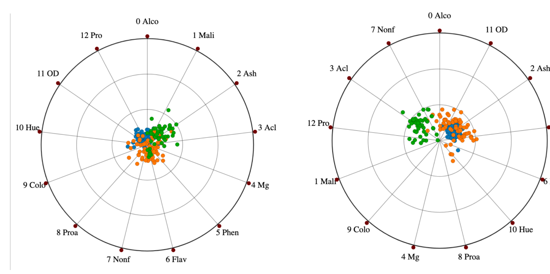
Keywords: RadViz, dimension arrangement, multiple views.

1 Introduction

The interactive analysis of high-dimensional data includes the usage of visualizations able to display multidimensional elements on a 2D plane. Some solutions map the data on poly-lines connected to data values, such as parallel coordinates [15] or radar-charts [10]; other solutions map multidimensional elements to points, loosing a clear connection with the original data dimensions, e.g., PCA [27] (linear mapping, preserving distances, not showing original dimensions), t-sne [17] (non linear mapping, altering distances, not showing original dimensions), RadViz [14] (non linear mapping, altering distances, showing original dimensions). RadViz and parallel coordinates exhibit opposite pros and cons (see, e.g., [2]) and are quite popular because they display the original dimensions. In this paper we focus on RadViz, aiming at mitigating some problems it introduces:

1. Clumping: points accumulate close to the center, see, e.g., [7], because, as the number of dimensions increases, it is likely that points have approximately equal (normalized) values at the dimensions which lie on the opposite sides of the circumference and that push them towards the center, see Figure 1a;
2. Many to one plotting: it happens that points that are far away in the multidimensional space are represented very close for a given disposition. As an example, the 5D points $P1 = \{a: 0.8, b: 0.03, c: 0.8, d: 0.08, e: 0.13, f: 0.18\}$ and $P3 = \{a: 0.11, b: 0.49, c: 0.11, d: 0.08, e: 0.13, f: 0.25\}$ are plotted very close using the disposition $< a,b,c,d,e,f >$, see Figure 2. It is worth noting that the presented case is totally different by the popular example in which the two points are merely proportional, like points $\{a: 0.3, b: 0.6, c: 0.9\}$ and $\{a: 0.1, c: 0.2, d: 0.3\}$: these two points are plotted always on the same position, regardless the chosen disposition;
3. Effectiveness of a point position in representing its normalized values: plotting quality is highly influenced by the order of the dimensions and some dispositions are more effective than others (see Figure 3). Dealing with this problem, at the level of single points, is a specific contribution of the paper: most of the available proposals only explore dimensions dispositions considering all the dataset points. Moreover, the heuristic we have devised to estimate the optimal disposition is used also to estimate the (average) optimal disposition of an arbitrary subset S of points, e.g., all dataset points (*global optimal disposition*, see Figure 1b) or points belonging to a cluster (*local optimal disposition*).

*e-mail: {angelini, blasilli, lenti, palleschi, santucci}@diag.uniroma1.it

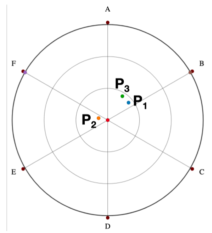


(a) Random disposition (b) Global optimal disposition

Figure 1: (a) the clumping problem displaying the n-dimensional Wine dataset (n=13). All the data, but one point, is within $1/3$ of the RadViz radius, using less than 10% of the area. (b) the same data after changing the disposition of the dimensions using our heuristic.

Figure 1 shows an example of the clumping problem and how our heuristic allows for mitigating this effect changing the dimensions arrangement along the circumference.

Figure 2 shows how some points are normalized and plotted on RadViz. It is possible to observe the non linearity of the min-max normalization: e.g., point P1 has two quite different values for dimensions A (12) and C (5.9) but after the normalization these dimensions show the same value (0.8). While it is impossible to deal with such issues without changing the standard definition of RadViz, our proposal mitigates the problem using a coordinated view (see Section 4) allowing the user to explore the original dimension values: mouse-hovering on P1, P2, and P3 generates the images at the bottom of Figure 2: the orange lines show the spring forces acting on the points using the normalized values and the numerical values on the circumference reflect the original data values. That supports a better understanding of the actual points position. As an example, the normalized dimensions of P1 exhibit very high values for A and C while its actual position suggests a prominent influence of B, as correctly happens for P3. This example highlights another situation that pushes points close together even if their normalized values are quite different and not proportional. Moreover, P2 is very close to the center, because it has similar values for B, D, and F that

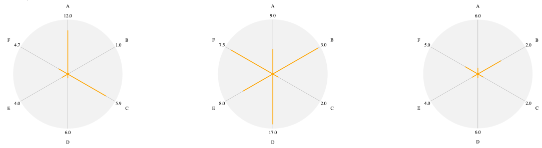


	A	B	C	D	E	F
P1	12	1	5.9	6	4	4.7
P2	9	3	2	17	8	7.5
P3	6	2	2	6	4	5
P4	13.72	3.1	7	18	11	8
P5	5	0.94	1.36	5	3	4

Original data

	A	B	C	D	E	F
P1	0.80	0.03	0.80	0.08	0.13	0.18
P2	0.46	0.95	0.11	0.92	0.63	0.88
P3	0.11	0.49	0.11	0.08	0.13	0.25
P4	1.00	1.00	1.00	1.00	1.00	1.00
P5	0.00	0.00	0.00	0.00	0.00	0.00

Normalized data

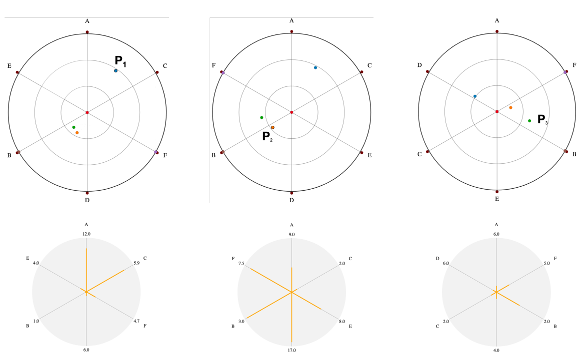


(a) P1 (b) P2 (c) P3

Figure 2: The standard non linear RadViz min-max normalization ($\frac{X-x_{min}}{X_{max}-X_{min}}$) and a RadViz plotting a synthetic dataset with a random dimensions disposition. P1, P2, and P3 are displayed in a non-optimal way as demonstrated by our ancillary views (a,b,c), that show the spring forces and the original data values. P4 (all maximum values) and P5 (all minimum values) coincide on the center and it is not possible to separate them changing the dimensions arrangement.

are symmetrically arranged along the circumference; such values compensate each other (see Figure 2b) and the point is plotted very close to the center (i.e., the cause of the clumping problem). P4 and P5 overlap on the center. The confusing position of P1 pushed us to define the *point optimal disposition* of the dimensions for a single point as the disposition (not necessarily unique) that maximizes the distance of the point from the center (see Section 3). Figure 3 shows the point optimal dispositions for P1, P2, and P3; obviously, it is likely that the optimum for a single point is not the optimum for other points. Computing the optimum by inspecting all dispositions is an intractable problem and we have devised a suitable heuristic for estimating it. The heuristic is also used for:

- characterizing the correctness of the points position in a given disposition (see Section 3), providing, at point level, a visual indicator of the RadViz plotting quality;
- developing the basis of a second heuristic that estimates the disposition optimizing the *mean* distance of a subset of points, increasing their overall quality and mitigating the clumping problem and the many to one plotting cases that are not caused by numerical proportionality.



(a) P1 optimal disposition (b) P2 optimal disposition (c) P3 optimal disposition

Figure 3: Optimal disposition for P1, P2, and P3. The distance from the center is maximized and the points are very close to the dimensions exhibiting the highest (normalized) values.

Summarizing, the contributions of the paper are the following:

- the definition of a visual quality metric at the level of single points, based on the notion of optimal disposition, and its usage as a local and global quality indicator;
- the definition and experimentation of heuristics able to approximate the intractable optimal disposition computation, effectively reducing the clumping problem, better positioning points, and improving the overall RadViz quality;
- ancillary visualizations and widgets, bridging the gap between the original multidimensional values and the normalized ones;
- a demonstrative full working prototype, implementing all the aforementioned contributions.

2 Related Work

RadViz [13] is a visualization solution useful when coping with multidimensional data, and several variations have been proposed according to different tasks, ranging from domain specific analysis [18] to classification [21], and data exploration. Several proposals attempt to improve the readability and interpretability of RadViz, often focusing on dimensions rearrangements or enrichment of the visualization.

Existing approaches about dimensions reordering focus on the global optimization of a quality indicator or similarity among dimensions, as in the work of Ankerst et al. [1]. Di Caro et al. [9] investigated the relation between RadViz dimensions disposition and visualization quality, optimizing the Davies-Bouldin index [8] for cluster separation measurement; the same holds for VizRank [16] that defines a usefulness measure for data projections based on cluster separation. Our proposal, instead, focuses only on data values, being independent from any additional information on the data like, e.g., clustering: we define the optimal disposition of dimensions for a point and we build on that to estimate local and global optimal disposition optimizing the effectiveness of representing points with respect to their values.

Several approaches coped with the interpretation problems of RadViz by proposing modifications to its classic formulation: among them, Cheng et al. [6] propose RadViz Deluxe, a set of techniques optimizing the layout resulting from RadViz: while the approach shares with our solution the goal of enhancing the comprehension of RadViz it modifies the original RadViz definition by introducing distortion, while our technique leaves intact the original shape of the RadViz. The work by Nováková and Štepanková [20] identifies drawbacks of RadViz and proposes two modifications to the classic approach demonstrating their validity. As the work by Zhou et al. [28], both introduce artificial dimensions in order to improve the layout or finding the best dimensions arrangement, effectively introducing distortion in the form of added dimensions. Ono et al. in [21] propose Concentric RadViz, an enhanced visual representation of RadViz to better support classification tasks using concentric circles for representing groups of dimensions and a filtering scheme based on sigmoidal weighting to reduce cluttering and ambiguity in the visualization. While the approach mitigates some drawbacks of RadViz, it is dependant on dimensions grouping and more focused on allowing data exploration with respect to our approach. Sharko et al. proposed Vectorized RadViz [24], a variation that extends the number of dimensions through data flattening to increase the power of RadViz by enhancing the flexibility of dimensions layout; again this is in contrast to our approach that does not alter the classic RadViz formulation. Pagliosa and Telea proposed RadViz++ [22], a visual augmentation of RadViz to improve its comprehension and scalability; while sharing tasks with this work, our approach is applicable without modifying the classic representation of RadViz as RadViz++ does. Finally, Wang et al. proposed PolarViz [26], a focus + context distortion approach to manipulate the radial distribution of data points, with the goal of mitigating the clumping of projected data points towards the origin; differently from that, our approach does not introduce any distortion while aiming at the clumping reduction.

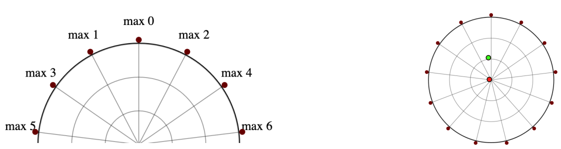
3 The Point Optimal Disposition Heuristic

RadViz projects data points from a multidimensional space into a bi-dimensional space using a non linear min-max normalization and the practical consequences of this approach (see Section 1) are: a) the position of a point depends on the dimensions disposition and b) similar normalized values may represents different values in the original multidimensional space.

To mitigate the first problem, we have investigated how to alter the disposition of the dimensions to move a point to a more informative position, i.e., to find the *point optimal disposition*. The main idea is that the further away a point is from the center, the more informative its position is, being the point closer to the attributes having the highest values.

We define the optimal disposition for a single point as the disposition (not necessarily unique) that maximizes the distance of the point from the center, and we have designed a heuristic able to approximate such an optimum without enumerating all the $\frac{(n-1)!}{2}$ distinct dispositions of the dimensions. Given a point $p = (x_0, x_1, \ldots, x_{n-1})$ constituted by n normalized values of the original dimensions we approximate the optimal disposition as follows: we sort the normalized values in descending order $O = (v_0, v_1, \ldots, v_{n-1})$ and we list the dimensions accordingly $(max_0, max_1, \ldots, max_{k-1})$ with $k = n$ and being max_i the dimension corresponding to the i value in the ordered list O. Then we rearrange the dimensions in the following way $(\ldots, max_3, max_1, max_0, max_2, max_4, \ldots)$: the maximum in the center of the disposition, the second maximum on the left, the third maximum on the right, and so on (see Figure 4a). The idea is to arrange as close as possible the dimensions that exhibit the highest values to reduce the cases in which two attributes with high values are symmetrically arranged and pushes the point towards the center.

Figure 4 shows the disposition generated by our heuristic.



(a) (b)

Figure 4: (a) The heuristic arrangement of the dimensions; (b) the point original position (red) and the optimal position (green).

The heuristic can be applied to a) improve the position of a single point, approximating its optimal disposition, or b) to improve the dimensions arrangement for a set of points S, devising a disposition that maximize the *mean* of the distances from the center of the S points, with the goal of mitigating the clumping problem and improving the overall quality of the positions of the S points, or c) for providing a quality indicator at point level.

To generalize the approach to a subset S of points, we consider the S points percentile i_{th} of all dimensions:

1. we compute the i_{th} percentile for each dimension;
2. we sort such n percentiles and we use our ordering heuristic to devise the optimal disposition of point $p = (i_{th0}, i_{th1}, \ldots, i_{thn})$;
3. we compute the mean distance of the S points from the center;
4. we iterate the procedure on k different percentile values (e.g., $k[1 \ldots 100]$) and we select the permutation that maximizes this mean; the corresponding percentile is called the best percentile. An example of the application of the heuristic to all the points is visible in Figure 1b.

The computational complexity of the heuristic is dominated by $dimensions \cdot dataElements$ (step 3, computing the mean distance). We tested it on a $13 \cdot 10{,}000$ case, getting a response time of 30 seconds, on a single 2.7GHz Intel i7 core. Some preliminary experiments using random sampling produced a response time of 2 seconds with a mean degradation of the accuracy of about 1.5%.

To provide a quality indicator we compare, for each point $p = (x_p, y_p)$, the actual point distance from the center with its value in the point optimal disposition $p* = (x_{p*}, y_{p*})$, see Figure 4b. Using these two values, we compute the *distance proportion* (*DP*) between them, the higher the better:

$$DP(p) = \frac{\sqrt{x_p^2 + y_p^2}}{\sqrt{x_{p*}^2 + y_{p*}^2}} \tag{1}$$

Function 1 defines a visual quality metric (see, e.g., [4]), i.e., it refers to the quality of the visualization and not to a data property. We use it to show how much the current position is far from the optimal one.

4 The Prototype

Together with the analytics exposed in Section 3, we propose a set of ancillary visualizations and widgets that help the user in better interpreting a RadViz under investigation and we present them in a demonstrative prototype called RadWidget, that can be found at `http://awareserver.dis.uniroma1.it/radwidget/`.

The prototype (see Figure 5) is composed of two main parts: the RadViz explorer environment (left) and the data inspector environment (right): the former shows the RadViz and allows to trigger our heuristics to mitigate the drawbacks 1 and 3 described in Section 1, while the latter contains the visual tools showing the multidimensional nature of the points to mitigate the drawback 2.

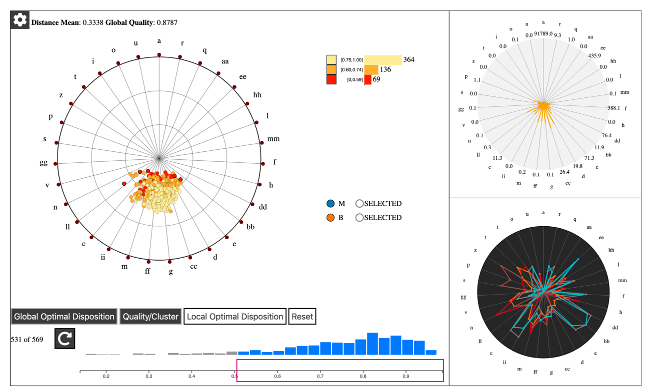


Figure 5: The RadWidget prototype

The RadViz explorer accommodates the main representation of the RadViz, constituted by a reference grid and draggable anchor points, allowing to manually change the order of the dimensions. The reference grid provides for a better reading of the points placement, (e.g., better estimating the distance of a point from the center or its placement with respect to a dimension). After a dataset is loaded, its dimensions are arranged in their default order and the points are plotted on the RadViz. The analyst can select the global optimal disposition through the relative button or the point optimal disposition by right-clicking a point. If the dataset is labeled, its classes are listed on the right of the RadViz and can be selected to compute their local optimal disposition.

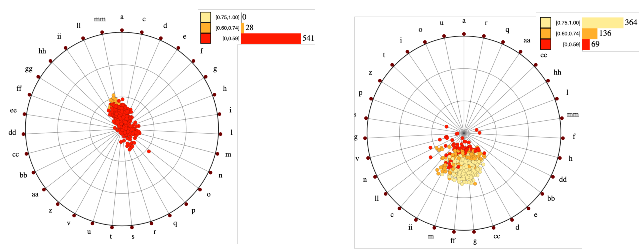
The distance proportion (Function 1) is computed for each point and the values are used to bin points in 3 categories: well-positioned ($DP \geq 0.75$, mapped to red), average-positioned ($0.6 \leq DP < 0.75$, mapped to orange) and bad-positioned ($DP < 0.6$, mapped to yellow) points. This information is summarized with an histogram on the right of the RadViz and is encoded on the points on demand; it is useful to have an overview of the quality of the dimensions disposition identifying areas where the points are well or bad positioned (see Figure 6).

To further mitigate the clumping problem we have implemented a CrossWidget slider [5] (at the bottom of the RadViz), that shows the distribution of the points quality to support a more informed filtering operation.

On the top of the RadViz the mean distance of the points and the proportion of well-positioned points (global quality) are reported to provide high level indicators of the quality of the current disposition.

The interaction with the RadViz points triggers the data inspector environment. Mouse-hovering a point displays in the SpringChart (upper part) its original values. The point is shown on the center of the circular layout and for each dimension an orange radial segment proportional to its normalized dimension value is represented, while the label shows the original value. Furthermore, a radar-chart plot (lower part) allows to compare or visualize multiple points at the same time.

5 Experimental Evaluation



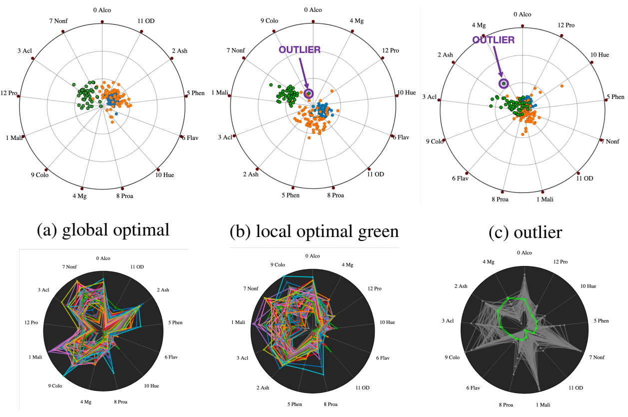
(a) Random disposition

(b) Global optimal disposition

Figure 6: (a) The default Breast Cancer Wisconsin Diagnostic dataset disposition. (b) The global optimal disposition.

We have explored the global optimal disposition behaviour using the Breast Cancer Wisconsin Diagnostic (BCWD) dataset [25] and the Wine dataset [12]. The BCWD dataset is composed by 569 instances, 32 dimensions, and 2 clusters. Figure 6 shows a comparison between the original dimensions arrangement (Figure 6a, mean distance = 0.16, global quality = 0.05, clumping and clutter phenomena) and the global optimal disposition (Figure 6b, mean distance = 0.33, global quality = 0.88, reduction of clumping and clutter). The original dimensions arrangement presents 0 well-positioned, 28 average-positioned and 541 bad-positioned points; using our heuristic 364 points result well-positioned (63% improvement) with bad-positioned points reduced to 69 (about 10% of the dataset) and 136 average-positioned.

The Wine dataset contains 178 instances, 13 dimensions, and 3 classes. The global optimal disposition changes the mean distance from 0.12 to 0.20, and the global quality from 0.06 to 0.61, improving the plotting quality, see Figure 1. In particular, the application of the global optimal disposition heuristic raises the quality by more than 16% considering only the well-positioned and 54% considering also the average-positioned points, and reduces by almost 60% the number of bad-positioned points. Finally, we have performed a broader experimental evaluation, generating 4400 random dispositions scrambling row subsets and column positions of 4 different datasets (BCWD, Wine, Iris, Diabetes [11]) and estimating for each of them the global optimal disposition. The distance mean proportion between the original random disposition



(a) global optimal

(b) local optimal green

(c) outlier

(d)

(e)

(f)

Figure 7: The global optimal disposition (a) generates a global improvement, but the dimensions for the green cluster (d) show a wrong relevance of dimension 12 (Proline). A local optimal disposition based on the green cluster (b) better characterizes the most relevant dimensions (e) and points out an outlier. An optimal disposition for the outlier (purple circle in (c)) shows its peculiarity.

and the estimated global optimal one shows an average improvement of 212%, median 185% with a maximum of 700%. Only in 11 cases the distance mean of the random disposition was slightly better (about 3%) than the one generated by the heuristic.

The local optimal disposition, instead, aims at characterizing subsets of instances, highlighting their most dominant dimensions. Figure 7a shows the global optimal disposition for the Wine dataset. Instances of the green cluster are grouped towards the dimensions 12 (Proline), 3 (Acl), 7 (Nonflavanoid) suggesting that they are the most relevant dimensions for it. However, the green cluster distribution in Figure 7d shows that all the instances have low values for the dimension 12 (Proline). This is made evident by the local optimization for the green cluster visible in Figure 7b that further separate it from the other clusters towards the dimensions 1 (Malic) acid, 7 (Nonflavanoid), 9 (Malic acid), i.e., the dimensions with the highest values on average as visible in Figure 7e, except for an outlier. The point optimal disposition of the outlier (see Figure 7c) shows that the outlier has higher values of 2 (Ash), 4 (Mg), and 0 (Alcohol), see Figure 7f.

6 Conclusion

The paper presented a combination of metrics, heuristics, and widgets aiming at mitigating some of the RadViz drawbacks. In particular, the paper investigated the idea of point optimal disposition, focusing on the effectiveness of single point position devising a suitable heuristic able to estimate such intractable optimum. Building on such definition and heuristic, the paper generalizes the point optimal disposition concept, providing means for better positioning set of points. Ancillary visualizations and visual quality metrics allow for more complex analysis.

We plan to improve the heuristics we have defined in the paper, increasing both their scalability and their effectiveness. In particular we want to investigate adaptive and efficient random sampling based solutions (see, e.g., [3, 19]) and progressive solutions (see, e.g., [23]).

Moreover, we are planning a formal user study, to validate the system also from the end user perspective.

REFERENCES

[1] M. Ankerst, S. Berchtold, and D. A. Keim. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In *Proceedings IEEE Symposium on Information Visualization (Cat. No.98TB100258)*, pp. 52–60, Oct 1998. doi: 10.1109/INFVIS.1998.729559

[2] E. Bertini, L. Dell'Aquila, and G. Santucci. Springview: Cooperation of radviz and parallel coordinates for view optimization and clutter reduction. In *Coordinated and Multiple Views in Exploratory Visualization (CMV'05)*, pp. 22–29. IEEE, 2005.

[3] E. Bertini and G. Santucci. By chance is not enough: preserving relative density through nonuniform sampling. In *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.*, pp. 622–629. IEEE, 2004.

[4] E. Bertini and G. Santucci. Visual quality metrics. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1–5. ACM, 2006.

[5] G. Blasilli, S. Lenti, and A. Palleschi. CrossWidget: a D3.js Plugin to Support Multiple Scented Cross Filtering Activities. In J. Madeiras Pereira and R. G. Raidou, eds., *EuroVis 2019 - Posters*. The Eurographics Association, 2019. doi: 10.2312/eurp.20191131

[6] S. Cheng, W. Xu, and K. Mueller. Radviz deluxe: An attribute-aware display for multivariate data. *Processes*, 5(4), 2017. doi: 10.3390/pr5040075

[7] K. Daniels, G. Grinstein, A. Russell, and M. Glidden. Properties of normalized radial visualizations. *Information Visualization*, 11(4):273–300, 2012.

[8] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, April 1979. doi: 10.1109/TPAMI.1979.4766909

[9] L. Di Caro, V. Frias-Martinez, and E. Frias-Martinez. Analyzing the role of dimension arrangement for data visualization in radviz. In M. J. Zaki, J. X. Yu, B. Ravindran, and V. Pudi, eds., *Advances in Knowledge Discovery and Data Mining*, pp. 125–132. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[10] G. M. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *IEEE transactions on visualization and computer graphics*, 15(5):759–776, 2009.

[11] D. Dua and C. Graff. UCI machine learning repository, 2017.

[12] M. Forina, S. Lanteri, C. Armanino, et al. Parvus-an extendible package for data exploration, classification and correlation, institute of pharmaceutical and food analysis and technologies, via brigata salerno, 16147 genoa, italy (1988). *Av. Loss Av. O set Av. Hit-Rate*, 1991.

[13] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley. Dna visual and analytic data mining. In *Proceedings. Visualization '97 (Cat. No. 97CB36155)*, pp. 437–441, Oct 1997. doi: 10.1109/VISUAL.1997.663916

[14] P. Hoffman, G. Grinstein, and D. Pinkney. Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In *NPIVM '99: Proc. Workshop on New Paradigms in Information Visualization and Manipulation*, pp. 9–16. ACM Press, 1999. doi: 10.1145/331770.331775

[15] A. Inselberg. The plane with parallel coordinates. *The visual computer*, 1(2):69–91, 1985.

[16] G. Leban, B. Zupan, G. Vidmar, and I. Bratko. Vizrank: Data visualization guided by machine learning. *Data Mining and Knowledge Discovery*, 13(2):119–136, Sep 2006. doi: 10.1007/s10618-005-0031-5

[17] L. v. d. Maaten and G. Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[18] A. Martínez, L. Sánchez, and I. Couso. Engine health monitoring for engine fleets using fuzzy radviz. In *2013 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 1–8, July 2013. doi: 10.1109/FUZZ-IEEE.2013.6622420

[19] G. McIntyre. A method for unbiased selective sampling, using ranked sets. *Australian Journal of Agricultural Research*, 3(4):385–390, 1952.

[20] L. Nováková and O. Štepanková. Radviz and identification of clusters in multidimensional data. In *2009 13th International Conference Information Visualisation*, pp. 104–109, July 2009. doi: 10.1109/IV.2009.103

[21] J. H. P. Ono, F. Sikansi, D. C. CorrÃła, F. V. Paulovich, A. Paiva, and L. G. Nonato. Concentric radviz: Visual exploration of multitask classification. In *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, pp. 165–172, Aug 2015. doi: 10.1109/SIBGRAPI.2015.38

[22] L. d. C. Pagliosa and A. C. Telea. Radviz++: Improvements on radial-based visualizations. *Informatics*, 6(2), 2019. doi: 10.3390/informatics6020016

[23] H. Schulz, M. Angelini, G. Santucci, and H. Schumann. An enhanced visualization process model for incremental visualization. *IEEE transactions on visualization and computer graphics*, 22(7):1830–1842, 2016.

[24] J. Sharko, G. Grinstein, and K. A. Marx. Vectorized radviz and its application to multiple cluster datasets. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1444–1427, Nov 2008. doi: 10.1109/TVCG.2008.173

[25] O. L. M. W. Nick Street, W. H. Wolberg. Nuclear feature extraction for breast tumor diagnosis, 1993. doi: 10.1117/12.148698

[26] Y. C. Wang, Q. Zhang, F. Lin, C. K. Goh, and H. S. Seah. Polarviz: a discriminating visualization and visual analytics tool for high-dimensional data. *The Visual Computer*, May 2018. doi: 10.1007/s00371-018-1558-y

[27] S. Wold, K. Esbensen, and P. Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.

[28] F. Zhou, Wei Huang, Juncai Li, Yezi Huang, Yang Shi, and Ying Zhao. Extending dimensions in radviz based on mean shift. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 111–115, April 2015. doi: 10.1109/PACIFICVIS.2015.7156365

View publication stats