

Method for Improving RadViz's Navigation Function Based on Focusing and Filtering

Hyunwoo Han*

Taerin Yoon†

Hyoji Ha‡

Juwon Hong§

Kyungwon Lee¶

*†‡Life Media Interdisciplinary Program
Ajou University

§¶Department of Digital Media
Ajou University

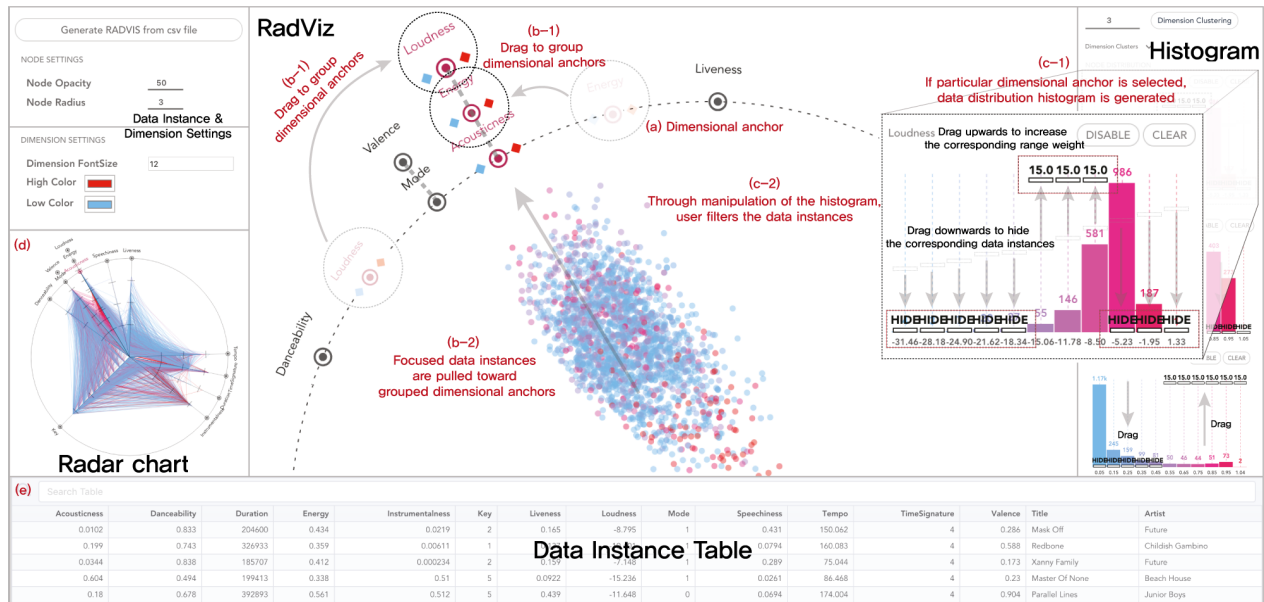


Figure 1: Improved visualization system of RadViz. The dimension of the data is indicated by the dimensional anchor(a). And desired data instances can be focused through a process(b-1, b-2) of grouping by dragging similar dimensional anchors. When user selects a specific dimension, Data distribution histogram(c-1) is generated. Through manipulation of it(c-2), user can navigate data instances by filtering. Also, multi-variable values of the selected instances can be verified through the Radar chart(d) and the table(e). (<https://radviz-focus.hanu.io/>)

ABSTRACT

For user to analyze large data set with various attributes as desired, rearranging data based on purpose is vital. This study proposes a method to navigate data with a desired purpose of user by improving RadViz visualization through focusing and filtering. To help understand visualization, user studies were conducted using the music data of Spotify. As a result, our system is discovered to be effective at classifying large data set of music with attributes and navigating desired music efficiently.

Keywords: Multi-task classification, Information Visualization, RadViz.

Index Terms: Human-centered computing [Visualization toolkits]

1 INTRODUCTION

Visualization of RadViz projected onto a plane is based on spring-force analogy. However, as the number of dimensions increases, it becomes difficult to find the meaningful dimensional anchors that

affect the location of the data because the arrangement of the dimensional anchors is not optimal[1]. Furthermore, even if the aforementioned issue has overcome, the data instances have a tendency to merge into the center of the RadViz when the dimensional anchors have the same weighting system based on the spring-force analogy. Consequently, these problems creates difficulties in navigating data. To solve the problems, this research proposes a RadViz model that can focus on the desired data and filter that for analyzing purposes. Improved RadViz not only enables user to selectively navigate large data set but also provides guidelines to navigate data that they desire. Furthermore, this visualization makes it easy to interpret over-all data instances' cluster by setting dimensional anchors considered with correlation and making intervals of dimensional anchors adjustable. This function has the advantage of enabling effective visualization interactions even when quantity of data instance increases, or the number of dimensional anchors shown over RadViz increases.

2 APPLICATION

2.1 Method

The visualization system proposed in this study can analyze any multidimensional CSV data. It offers anchor clustering based on k-means algorithm by automatically analyzing the correlation between dimensions. The mapping of each data instance differs depending on the arrangement and weight of dimensions.

*e-mail: ainatsumi@ajou.ac.kr

†e-mail: yoontr@ajou.ac.kr

‡e-mail: hjha0508@ajou.ac.kr

§e-mail: penguin373@ajou.ac.kr

¶e-mail: kwlee@ajou.ac.kr

Instance rearrangement formula is as follows.

$$DIP_{(n)} = \left(\sum_d^{Dimensions} DAP_{(d)} \cdot P_{(d)(n)} \right) / DIP_{max} \cdot R \quad (1)$$

$DIP = DataInstancePosition,$
 $DAP = DimensionalAnchorPosition,$
 $P = Springpowerofdimensionvalue$

The location of the n th instance is the sum of the position of the dimension multiplied by the weights set for the value that an instance has for each dimension. Then, to distribute data instances widely across the screen, divide location of each instances by the $DataInstancePosition(max)$ at maximum distance from the central point of the RadViz. Afterward, it is to be multiplied by the radius R of the total RadViz.

2.2 Visualization and Functions, Interaction

This study used RadViz visualization with focusing and filtering to determine the characteristics of data instances based on a user's preference (Figure 1). Focusing is grouping similar anchors into one axis, which means that instances with higher values of the dimension belonging to the group are pulled toward it[2]. The arrangement of dimensional anchors can be assembled as user wants and if dimensional anchor is judged unnecessary, it can be deleted. Filtering is possible by manipulating the histogram that shows how the instances are distributed based on each dimension. This histogram allows a user to specify the range and change its weight or exclude relatively unimportant data instances. For example, if the user increases the weight value of a specific range of the anchors selected by the user, only the instances that fall within that range will be pulled strongly toward the anchor resulting change in the instance distribution. Figure 1 displays how the system of this research interacts. Even if the quantity of data instance is massive or the number of dimensional anchor which data has is huge, this visualization system on the basis of filtering and focusing is constructed to precisely distinguish characteristics of data and to modify efficiently. If user uploads CSV file of multi-dimension data, visualization about that data generates. User can use interaction of moving dimensional anchors and modifying histogram to distinguish data according to properties. Also, with 'Dimension clustering' function to adjust interval and dispersion of dimensional anchors, system allows to focus on data instances with similar attributes. Furthermore, by using a Radar chart, we can verify the multivariable values of data instances. And Data instance table makes it simple to find out actual value of data instances that appear in RadViz. Besides, there is incidental feature that allows to set the size, transparency and color of the RadViz's data instances. Figure 2 shows the process of obtaining the data output by utilizing the function of each visualization. By manipulating the dimensional anchors or weight, the distribution of instances can be improved to find only those data instances that meet the desired conditions.

3 CASE STUDY

To aid understanding, we developed a music data navigate scenario by using RadViz with focusing and filtering. We used a music dataset that included 15 dimensions such as *Acousticness* and *Danceability*. Let us assume that a user is looking for acoustic music with middle value of *Energy* and *Loudness*. Based on the anchor clustering result that the system proposed, dimensional anchors can be grouped referring to the user's discretion over one axis (Figure 2(b)). So, the user grouped *Energy* and *Loudness* with *Acousticness* to find customized music with accuracy (Figure 2(c)). As the grouping progresses, the music data instances will gain more strength toward the group of anchors. Furthermore, by clicking on each anchor, the data distribution histogram can be confirmed, and

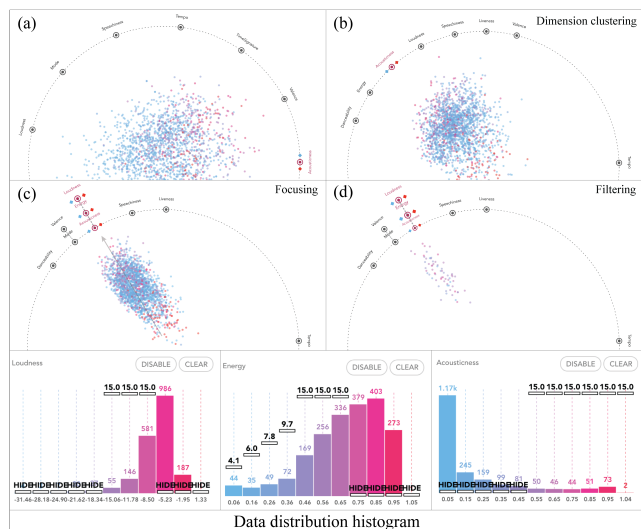


Figure 2: The process by which users adjust their RadViz to search for data instances for their desired purposes. (a) RadViz as conventional method, data instances are distributed in the dimensional anchors arranged at equal intervals (b) After Dimension clustering based on Dimension correlation (c) In order to focus, dimensional anchors are grouped (d) RadViz when the weight value of the desired range is increased, and the remaining range is hidden through the data distribution histogram for filtering.

the weight of range for each dimension can be manipulated. To search for acoustic music, the weight of the upper range is to be increased and the weight of the lower range is to be decreased to hide *Acousticness* dimension histogram. Even if high-valued music of *Acousticness* are picked out according to the previous procedure, there are still music that does not meet users desired condition. To find music that has medium value of *Energy* and *Loudness*, the weight of the middle range in the histogram of both the dimensions should be increased while the other ranges weights are decreased to hide and filter data (Figure 2(d)). In RadViz with focusing and filtering, just by choosing the data instances that a user desires will show its value distribution in the Radar chart, and at the bottom, the actual value of the dataset can be confirmed. As a result, by using our visualization system with focusing and filtering, user can experience a more efficient and certain way to navigate music suitable to the user from a massive music data set.

4 CONCLUSION

In this research, the RadViz data cohesiveness problem is improved to enable user to navigate in various ways in more detail. In this system, by proposing an improved RadViz with focusing and filtering, dispersion of data instances changes based on user interaction, and it is possible to search for instances that meet the conditions. It means user can accurately navigate visualized data. Because this system can hold not only quantitative data but also categorized data, it is expected that the system will be capable of embracing the various datasets.

REFERENCES

- [1] Piazzentin J. H. Ono et al. Concentric RadViz: Visual Exploration of Multi-Task Classification. *Sibgrapi 2015 (28th Conference on Graphics Patterns and Images)*, 165-172, 2015.
- [2] A. Buja, J. A. McDonald, J. Michalik, and W. Stuetzle. Interactive Data Visualization Using Focusing and Linking. *In Proceedings of the 2Nd Conference on Visualization 91, VIS 91*, pages 156163, Los Alamitos, CA, USA, 1991. IEEE Computer Society Press.