2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

Reordering Dimensions for Radial Visualization of Multidimensional Data - A Genetic Algorithms Approach

Huynh Thi Thanh Binh[1], Tran Van Long[2], Nguyen Xuan Hoai[3], Nguyen Duc Anh[4], Pham Manh Truong[5]

***Abstract*— In this paper, we propose a Genetic Algorithm (GA) for solving the problem of dimensional ordering in Radial Visualization (Radviz) systems. The Radviz is a non-linear projection of high-dimensional data set onto two dimensional space. The order of dimension anchors in the Radviz system is crucial for the visualization quality. We conducted experiments on five common data sets and compare the quality of solutions found by GA and those found by the other well-known methods. The experimental results show that the solutions found by GA for these tested data sets are very competitive having better cost values than almost all solutions found by other methods. This suggests that GA could be a good approach to solve the problem of dimensional ordering in Radviz.**

I. INTRODUCTION

Today′s scientific and business applications produce large data sets with increasing complexity and dimensionality. While information is growing in an exponential way, data is ubiquitous in our world. Data should contain some kind of valuable information that can possibly explore the human knowledge. However, extracting the meaningful information in large scale data is a difficult task.

Information visualization techniques have been proven to be of high value in gaining insight into these large data sets. The aim of information visualization is to use the computer-based interactive visual representations of abstract and non-physically based data to amplify human cognition. It aims at helping users to detect effectively and explore the expected, as well as discovering the unexpected, to gain insight into the data [6].

A major challenge for information visualization is how to present multidimensional data to analysts. Data visualization methods often employ a map from multidimensional data into lower-dimensional visual space. The reason is that visual space representation is composed of two or three spatial coordinates and a limited number of visual factors such as color, texture, etc. However, when the dimensionality of the data is high, usually from tens to hundreds, the mapping from multidimensional data space into visual space incurs information loss. This leads to the ultimate question in information visualization [6]. How to project from a multidimensional data space into a low-dimensional space that could best preserve the characteristics of the data?

[1] Hanoi University of Science and Technology, Hanoi, Vietnam, email: binh.huynhthithanh@hust.edu.vn; [2] University of Transport and Communications, Hanoi, Vietnam, email: vtran@utc.edu.vn; [3] Hanoi University, Hanoi, Vietnam, email: nxhoai@gmail.com; [4] Hanoi University of Science and Technology, Hanoi, Vietnam, email: nguyenanh.nda@gmail.com; [5] Hanoi University of Science and Technology, Hanoi, Vietnam, email: manhtruong0106@gmail.com.

One of the challenge to obtain good data projections in the current visualization techniques is the problem of dimensional ordering. As stated by Mihael Ankerst et al. in [3]: "The order and arrangement of data dimensions is crucial for the effectiveness of a large number of visualization techniques such as parallel coordinates, scatterplot matrix, recursive pattern, and many others". Moreover, they suggest that the data dimensions of higher similarity should be place close together in the visual space.

The Radviz technique is one of the most common visualization techniques used in medical analysis [11], [12], [14]. Finding the optimal order of data dimensions in Radviz is known to be an NP-complete problem [3]. Although there have been a number of proposed methods for solving the dimensional ordering problem in Radviz [7], [14], most of them are exhaustive or greedy local search in the space of all permutations of data dimensions. These methods are usually only tested on some data sets with small number of dimensions. In this paper, we propose the use of GA for solving the problem of dimensional ordering for Radviz and compare with other methods on the data sets in both low and high dimensional spaces.

The rest of the paper is organized as follows. Section II gives the essential ideas of Raviz and the cost functions for the problem of dimensional ordering. In section III, we give a brief overview of the approaches used to solve the problem of dimensional ordering for a number of information visualization methods. The details of our method for solving the dimensional ordering in Radviz are presented in section IV. Section V describe our experiments, results, and discussions. The paper concludes with section VI, where some possible extensions for the work in this paper are highlighted.

II. RADVIZ AND ITS DIMENSIONAL ORDERING PROBLEMS

The Radviz was first introduced by Patrick Hoffman et al. in [11], [12], and it could be regarded as an effectiveness of non-linear dimensionality reduction method. Radviz directly maps multidimensional data points into a visual space based on an equibalance of a spring system.

Figure 1 shows a simple example of Radviz for visualizing a multidimensional point in a six dimensional space. Each dimension of the data set presents by an anchor point on the circumference of the unit circle. The dimensional anchors are usually placed evenly on the unit circle. Imagine that each dimensional anchor is attached to a spring, the stiffness of each springs equals the value of the dimension corresponding to the dimensional anchor. The other end of

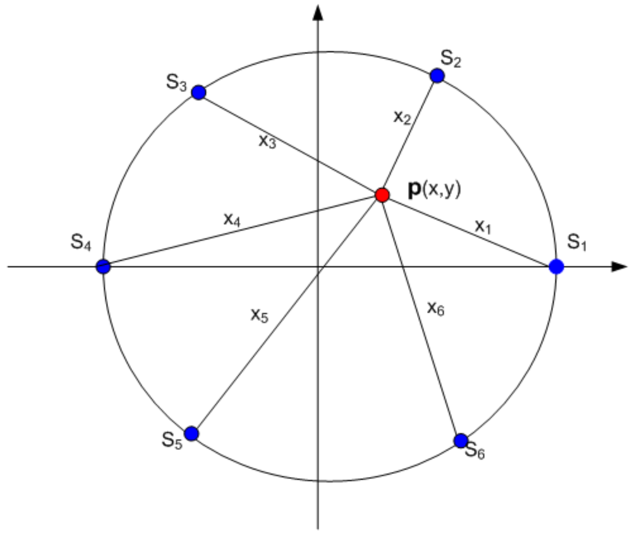


Fig. 1: Radviz visualize a multidimensional point with six attributes.

each spring is attached to a point in the visual space. The location of this point ensures the equibalance of the spring systems. For a multidimensional point $\mathbf{x} = (x_1, x_2, \ldots, x_p)$ in the p-dimensional space, the dimensional anchors $S_i, i = 1, 2, \ldots, p$ can be easily calculated by as:

$$S_i = \Big(\cos\frac{2\pi(i-1)}{p}, \sin\frac{2\pi(i-1)}{p}\Big), i = 1, 2, \ldots, p.$$

For the spring system to be equibalance, we must have

$$\sum_{i=1}^{p}(S_i - p)x_i = 0.$$

Thus, the multidimensional point $\mathbf{x}$ is presented by the point $p(x, y)$ with

$$x = \frac{\sum_{i=1}^{p} x_i \cos\frac{2\pi(i-1)}{p}}{\sum_{i=1}^{p} x_i}, y = \frac{\sum_{i=1}^{p} x_i \sin\frac{2\pi(i-1)}{p}}{\sum_{i=1}^{p} x_i}.$$

The important properties of the Radviz method are described in [8]:

- If a multidimensional point with all p coordinates have the same value, the data point lies exactly in the origin of the graph. Points with approximately equal dimensional values (after normalization) lie close to the center. Points with similar dimensional values, whose dimensions anchors are opposite each other on the circle lie near the center.
- If the point is a unit vector point, it lies exactly at the fixed point on the edge of the circle where the spring for that dimension is fixed. Points which have one or two coordinate values significantly greater than the others lie closer to the dimensional anchors (fixed points) of those dimensions.
- The position of a point depends on the layout of the particular dimensional anchors around the circle.
- Many points can be mapped to the same position. This mapping represents a non-linear transformation of the data that preserves certain symmetries.
- The Radviz method maps each data record to a point in a multidimensional data set that is within the convex hull of the dimensional anchors.

These important properties of the Radviz help to visualize the entire of data dimensions in a single visual form. Radviz is particularly useful for detecting clusters and outliers in multidimensional data [2]. However, the performance of Radviz is very sensitive to the order of data dimensions. A good dimension order satisfies the condition that similar data dimensions should be close to each other, and the optimal dimension order is the one that minimize the sum of similarity of successive dimensions in the order. The problem of finding the optimal order for Radviz (and many other visualization techniques) could be seen as a form of Travelling Salesman Problem (TSP) [3], where each dimension represent a city to visit and the similarity between the two successive dimensions in the order resemblances the distance between the two cities.

In details, the formalization of the dimensional ordering problem is given as follows [3]. Given a data set $X = \{x_i = (x_{i1}, x_{i2}, \ldots, x_{ip}), 1 \leq i \leq n\}$ that contains n observations, each observation contains p attributes. We use the metric to measure the dimension similarity as in [5], in which the similarity matrix of the $i-$th and the $j-$th attributes is defined as:

$$s_{ij} = 1 - \frac{1}{n}\Big[\sum_{k=1}^{n}\Big|\frac{x_{ki} - \min_i}{\max_i - \min_i} - \frac{x_{kj} - \min_j}{\max_j - \min_j}\Big|\Big], \quad (1)$$

where $\min_i = \min\{x_{ki}, 1 \leq k \leq n\}$, $\max_i = \max\{x_{ki}, 1 \leq k \leq n\}$. We denote the matrix of similarity as $S = (s_{ij})_{p\times p}$, where s_{ij} is the similarity between the $i-$th and the $j-$th dimensions (data attributes). The problem of dimensional ordering is to find an optimal permutation π of dimensional arrangement $\{1, 2, \ldots, p\}$ such that minimizes the cost function

$$C = \sum_{i=1}^{p} s_{\pi_i, \pi_{i+1}}, \quad (2)$$

where $\pi_{p+1} = \pi_1$. The Radviz visualization of the multidimensional data point $x = (x_1, x_2, \ldots, x_p)$ that corresponding to the permutation π is given as follows:

$$x = \frac{\sum_{i=1}^{p} x_{\pi_i} \cos\frac{2\pi(i-1)}{p}}{\sum_{i=1}^{p} x_i}, y = \frac{\sum_{i=1}^{p} x_{\pi_i} \sin\frac{2\pi(i-1)}{p}}{\sum_{i=1}^{p} x_i}. \quad (3)$$

III. Related Work

Finding a good mapping of multidimensional data into a visual space is a challenging problem. Such a good mapping is usually data dependent. In this section, we briefly describe the linear method related visualization techniques (other than

Radviz) and their methods used to solve the problem of dimensional ordering.

The Radviz visualization was firstly introduced in [11], [12]. Radviz visualization maps multidimensional data onto a two-dimensional space in a non-linear fashion. The dimensions of the data set are represented by points that are placed evenly on the unit circle. This is a compact representation, but it suffers from overlapping points. In this paper, the authors mention that the ordering of the position of dimensional anchors can reduced the overlapping.

Gregor Lean et al. proposed the VizRank method [14] based on Radviz. The VizRank is an algorithm for optimal subset selection of data dimensions. The authors used k-nearest neighbour classifiers to rank the score of the visualizations based on to the degree of class separation. However, the computational time of VizRank increases exponentially with the size of subset of data dimensions. The number of all possible Radviz projections for selection of q data dimensions of the p data dimensional space is $\dfrac{p!}{(p-q)!}$.

Vectorized Radviz (VRV) is another variant method of Radviz proposed by John Sharko et al. [18]. The VRV is used to represent cluster ensembles. The VRV visualization is used in exploring different clustering results by projecting data records on a vectorized cluster space. This approach proves to be useful in validating the clusters when many different clusterings for the same data set exist. The vectorization process flattens the cluster dimensions by creating a new dimension for each cluster in each cluster set. The RadViz visualization then has a separate dimensional anchor for each cluster in each cluster set. Di Caro et al. [7] analyzed the dimension arrangement for Radviz. In this paper, the authors mention that the Radviz suffers from several problems as overlapping of data points in the visual space, visual clutter, and NP-completeness of the ordering data dimensions. Two objective functions proposed that name as independent dimensional anchors and Radviz-dependent dimensional anchors. The best dimensional arrangement was obtained by exhaustively exploring all permutation. This limits Radviz's applications to only small dimensional data sets.

The star coordinates [13] arrange the coordinate axes on a circle placed on a two-dimensional plane with axes having their origin at the center of the circle and an arrangement exposing equal angles between adjacent axes. A multidimensional point is mapped to a 2D point by summing the unit vectors of each coordinate multiplied with the coordinate of the multidimensional point. The star coordinates support some interactive techniques as: scaling, rotation, marking and range selection. To alleviate the problem of overlapping clusters, Tran Van Long [15] used class preserving projections as a method for minimizing the overlap of centroid clusters or maximizing the distance all pairwise of the centroid clusters. The author also proposed a method to identify the region of clusters by enclosing all the projection points in the visual space by a compact and connected region.

The parallel coordinates is another popular visualization techniques for multidimensional data visualization [10]. The parallel coordinates also suffers from the problem of ordering axes dimensions. There have been a number of proposed approaches for solving this problem. However, most of these approaches could only find local optimal solutions to the problems.

To solve the problems of dimensional ordering in these aforementioned visualization techniques, there have been a number of proposed approaches in the literature. They are mainly based on greedy and local search in the space of dimension permutation. In [4], [5], The authors proposed a greedy algorithm named Similarity-Based Attribute Arrangement (SBAA) to find the dimensional arrangement with high dimensional similarity. The SBAA algorithm is a heuristic approach that starts with the two most similar attributes and find the closest data dimension with the current growing ordered sequence of dimensions. The procedure is repeated until all dimensions are included in the ordered sequence. Georgia Albuquerque et al. [1] derived a weighted sort procedure for solving the problem of dimensional ordering, in which the dimensions are ordered by weights, where the weight of the jth attribute (data dimension) is estimated as follow:

$$D_j = \sum_{i=1}^{p} \frac{n-i}{n} s_{ij}, \tag{4}$$

Weighted sort (WSORT) was then applied for solving the problem of dimensional ordering in the scatterplot matrix and the parallel coordinates. The reported best method so far (called AGENS in this paper) for solving the problem of dimensional ordering in visualization systems was proposed in [19]. Here, they used hierarchical clustering techniques to aid the search for the optimal ordering of data dimensions. AGENS was tested for the parallel coordinates visualization techniques.

These aforementioned algorithms could also be adapted for Radviz. However, they are only greedy and local search, which could find only local optimal in the space of permutations of data dimension. Therefore, in this paper, we propose the use of GAs for solving the problem of dimensional ordering in Radviz and compare with these greedy and local search approaches.

IV. Methods

As mentioned in section II, the problem of dimensional ordering in Radviz (and many other visualization techniques) could be casted as the TSP problem, where the task is to find an permutation of dimensions (cities) that is optimal with respect to the cost function given in equation 3. GA has long been used to solve optimization problems on the space of permutations and, in particular, the TSP problem with great success [9], [16]. Unlike greedy and local search approaches mentioned in section III, GA could act as a global optimizer and to our best knowledge, GA has not been used to solve the problem of dimensional ordering for Radviz.

In this paper, we use the standard and simple GA for permutations as in [9], [16]. To test the effects of different GA crossovers on permutations, 4 well-known GA crossovers that act on permutations are used and they are: Partially-mapped crossover (PMX), Cycle crossover (CX), and Order Crossovers (version 1 and 2 - OX1 and OX2). The descriptions of these crossovers could be found in [9], [16], [17]. The mutation operator is the swapping mutation (i.e choosing two random dimensions and exchange their positions in the permutation).

V. Experiments and Results

To test the performance of GA on the problem of dimensional ordering for Radviz, we conducted experiments on 5 data sets with dimensions ranging from 10 to 50 and compare the quality of solutions found by GA with those found by the methods mentioned in section III and the original dimensional order of the data.

A. Experiment Setting

DATA	Original	AGENS	WSORT	SBAA	GA
DB50	46.57	**47.31**	46.28	47.19	47.27
DB40	26.06	28.57	26.59	28.34	**28.61**
DB20	12.57	15.28	12.52	14.99	**15.34**
DB13	9.80	10.71	10.02	10.55	**10.74**
DB10	6.43	7.84	6.65	7.84	**7.98**

TABLE I: Cost values of solutions found by different methods

Table I summarizes the 5 data sets (DB10, DB13, DB20, DB40, Db50) used in our experiment. The DB10 data set is a synthetic data set that includes 480 data points with ten attributes[1]. The DB13 data set is a real data set that include 178 data points with 13 attributes[2]. The DB20 data set is another synthetic data set in [4] that contains 14831 observations and 20 attributes. The DB40 data set is a synthetic data set that creates from DB20 data set by adding 20 uniform random attributes. The last data set name DB50 is a real data set that contains 78823 data points with 50 attributes [3].

For GA, we tested four different well-known crossovers CX, PMX, OX1, and OX2 as mentioned in section IV. The population size was set as 50 and the number of generation was 200000 for each run. The crossover and mutation (using swapping) probabilities were 0.8 and 0.1 respectively. For each data set, 50 runs were conducted. All the runs were conducted on a machine with Intel Core I5, speeding at 2.4 GHz., having 4 GB Ram. and Windows 7 Professional operating system.

B. Cost Value Results and Discussion

Table I depicts the solutions with best cost function values found by GA and other tested methods (mentioned in section III). Here, "Original" means the natural order of the data set (no application of any optimization methods for dimension rearrangements). The results show that GA outperformed almost all other methods on the tested data sets, with an exception on DB50, where AGENS was only slightly better than GA.

Table II shows the cost values (in terms of maximum, mean, and standard deviation values) of solutions found by GA with 4 different crossovers. The results clearly show that GA with CX and PMX have similar performance on all 5 tested data sets with PMX was slightly better than CX on the data sets with high dimensions (DB40 and DB50). These crossovers performed better than OX (OX1 and OX2), which, to us, is a surprise as it has been shown that OX preserves the order better than PMX and CX (which usually shuffle good orders found by the evolutionary process so far) [16], [17].

C. Visualization Results

In this subsection we use Radviz to visualize multidimensional data set based on the best dimension orders found be the systems in the previous subsection. For each data set, we present five figures from (a) Original data dimensions, (b) AGENS algorithm, (c) WSORT algorithm, (d) SBAA algorithm, and (e) our GA.

Figure 2 shows the Radviz visualization techniques for the DB10 data set. This data set includes 480 data points with ten attributes, the data set is divided into 14 clusters. Each cluster in encoded with a different color. Figure 2a shows the data visualization with the natural data dimension ordering of the data set. In this figure we can see many clusters are overlapped. This reflexes the fact that its cost value in table I is the lowest. Figures 2b, 2c, 2d, 2e shows the data visualization with the dimensional order found by AGENS, WSORT, SBAA, and GA respectively. It could be seen in the Figures that the visualization quality of Figure 2e is the best and slightly better than Figure 2b and 2d.

Figure 3 shows four variants of the dimensional order for the DB13 data set. The one with the lowest cost function value is shown in Figure 3a, whereas the one with highest cost function value is depicted in Figure 3e. The best Radviz visualizations for this data set are the Figures 3e.

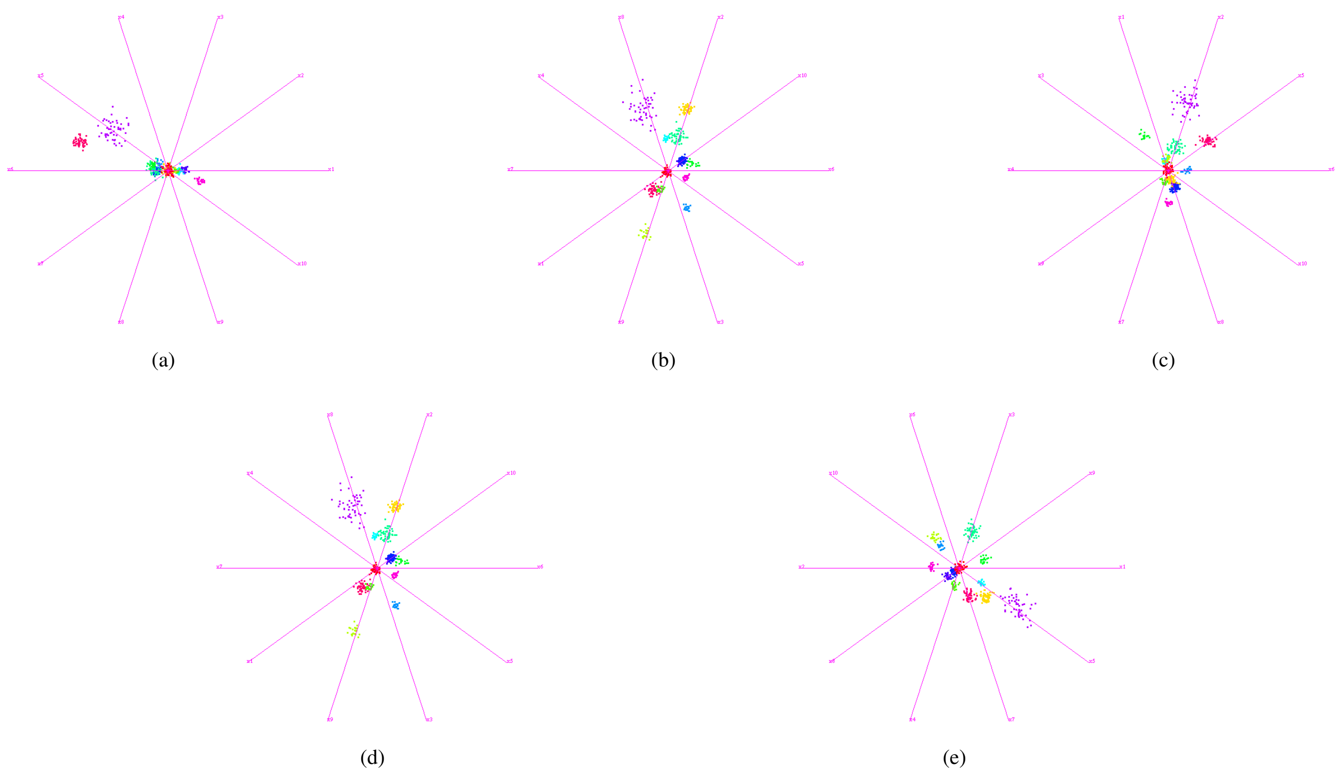
Figure 4 shows the Radviz visualizations for the DB20 data set. Among all methods to find the optimal order of data dimensions, the solution found by GA achieved the highest function cost value, and the best visual quality visualizations for this data set are showed in Figure 4e. On the DB40 data set (Figure 5), the result is similar, but it could be seen from the Figure that all visualizations could not show a good separation of data clusters.

Figure 6 shows the Radviz visualizations for DB50, a large-scale data set with high dimensions. The highest cost function value was obtained by AGENS, which is only slightly better than the cost function value of the solution found by GA. For this high-dimensional data set the visual clutter shows in the Radviz visualizations. It could be seen from the Figure that all visualizations could not show a good separation of data clusters.

[1]http://archive.igbmc.fr/projets/fcm/
[2]http://archive.ics.uci.edu/ml/datasets/Wine
[3]http://www.csie.ntu.edu.tw/ cjlin/libsvm/

Method	Statistics	DB50	DB40	DB20	DB13	DB10
	Max	47.253059	28.611128	15.339989	10.736052	7.984727
	Avg	47.211964	28.483503	15.314731	10.736040	7.984735
CX	Std	0.026529	0.071375	0.022903	0.000010	0.000009
	Max	47.265308	28.625044	15.339989	10.736052	7.984727
	Avg	47.219307	28.548325	15.328526	10.736040	7.984736
PMX	Std	0.029464	0.048375	0.015481	0.000011	0.000010
	Max	47.066906	28.368116	15.232319	10.736052	7.984727
	Avg	46.988663	27.658743	14.752338	10.728061	7.984726
OX1	Std	0.030158	0.058970	0.053262	0.007677	0.000000
	Max	46.565895	27.783726	14.890115	10.736052	7.984727
	Avg	46.473843	28.260378	15.062360	10.736053	7.984726
OX2	Std	0.037865	0.064201	0.103345	0.000002	0.000000

TABLE II: Cost values of solutions found by GA with different crossovers



(a) (b) (c) (d) (e)

Fig. 2: Radviz visualization of the DB10 data set. 2a Original. 2b AGENS. 2c WSORT. 2d SBAA. 2e GA.

VI. Conclusions

In this paper, we investigate the use of GA for solving the dimensional ordering problem in Radial Visualization (Radviz) systems. We conducted experiments on 5 common data sets with dimensions ranging from 10 to 50 to test the effectiveness of GA. The experimental results indicate that GA could find good dimensional orders for Radviz visualizations achieving hight cost function values. Among 4 different GA crossovers, PMX was shown to be the best for the problem. However, we have found that a high cost function value is sometimes not corresponding to the best visual quality (as our GA has found solutions with best cost function values but not always produced best visualizations). This, we hypothesize that, due to the noise inherent in the data which might also be amplified by the fitness (cost) function.

At this early stage of study, we have only used basic and elementary GA crossovers that work on permutations. In the near future, we are planning to test more sophisticated crossovers on permutations especially in combination with greedy and local search techniques to better solve the problem of dimensional ordering in Radviz. Other meta-heuristic techniques which have been used to solve the TSP problem well (such as Ant Colony Optimization) will be tried and

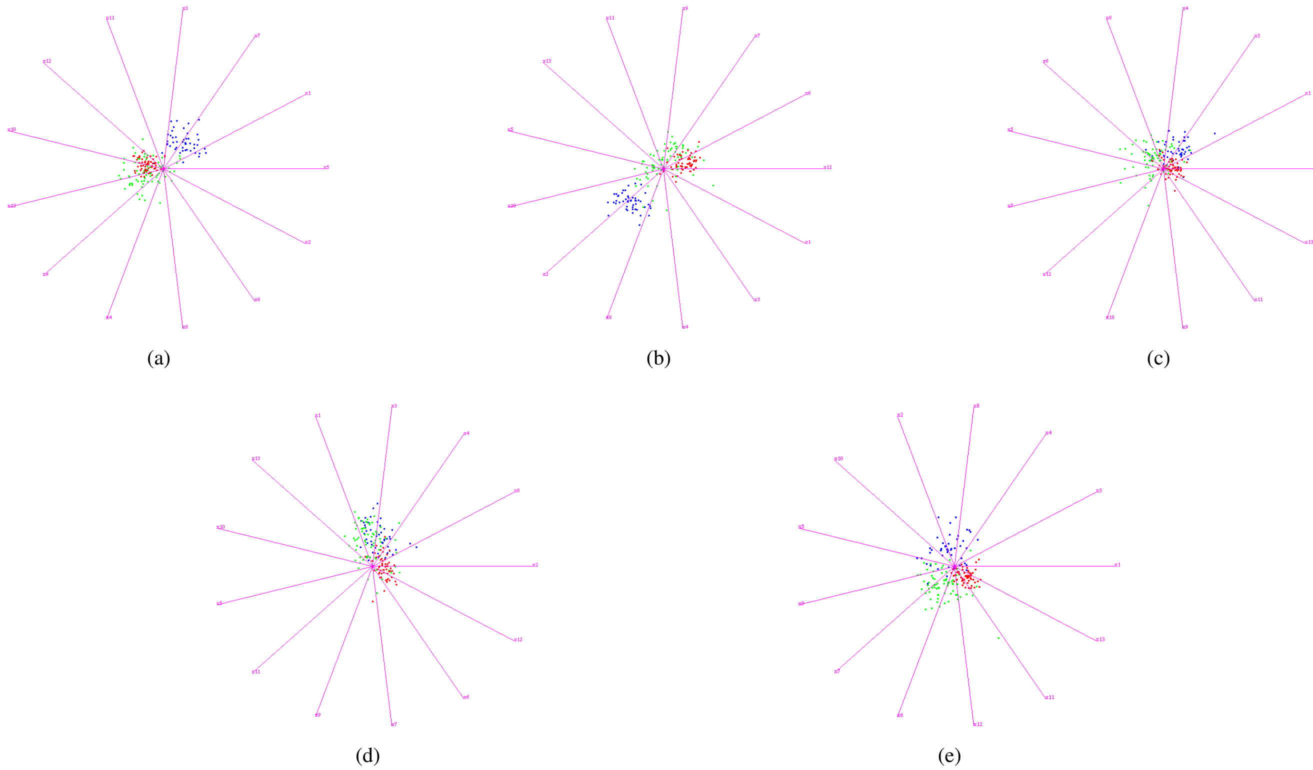


Fig. 3: Radviz visualization the DB13 data set. 3a Original. 3b AGENS. 3c WSORT. 3d SBAA. 3e GA.

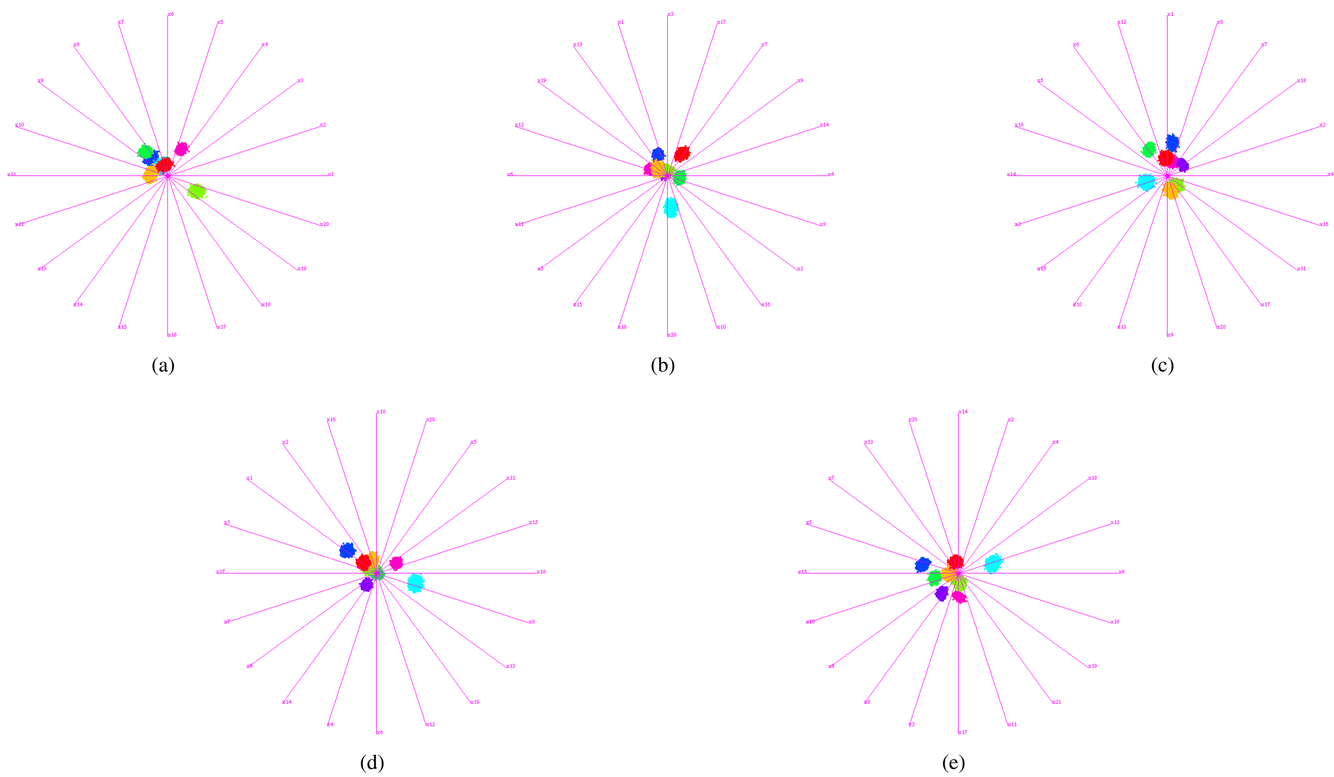
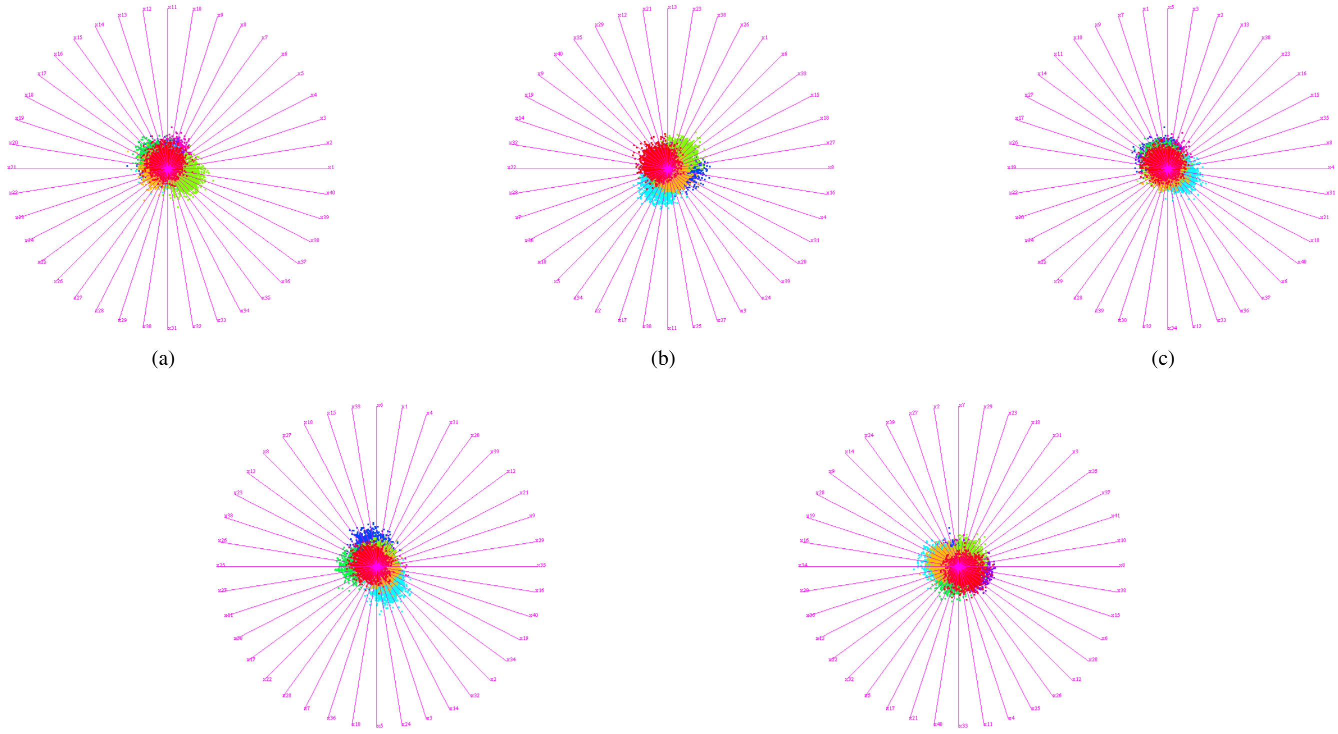
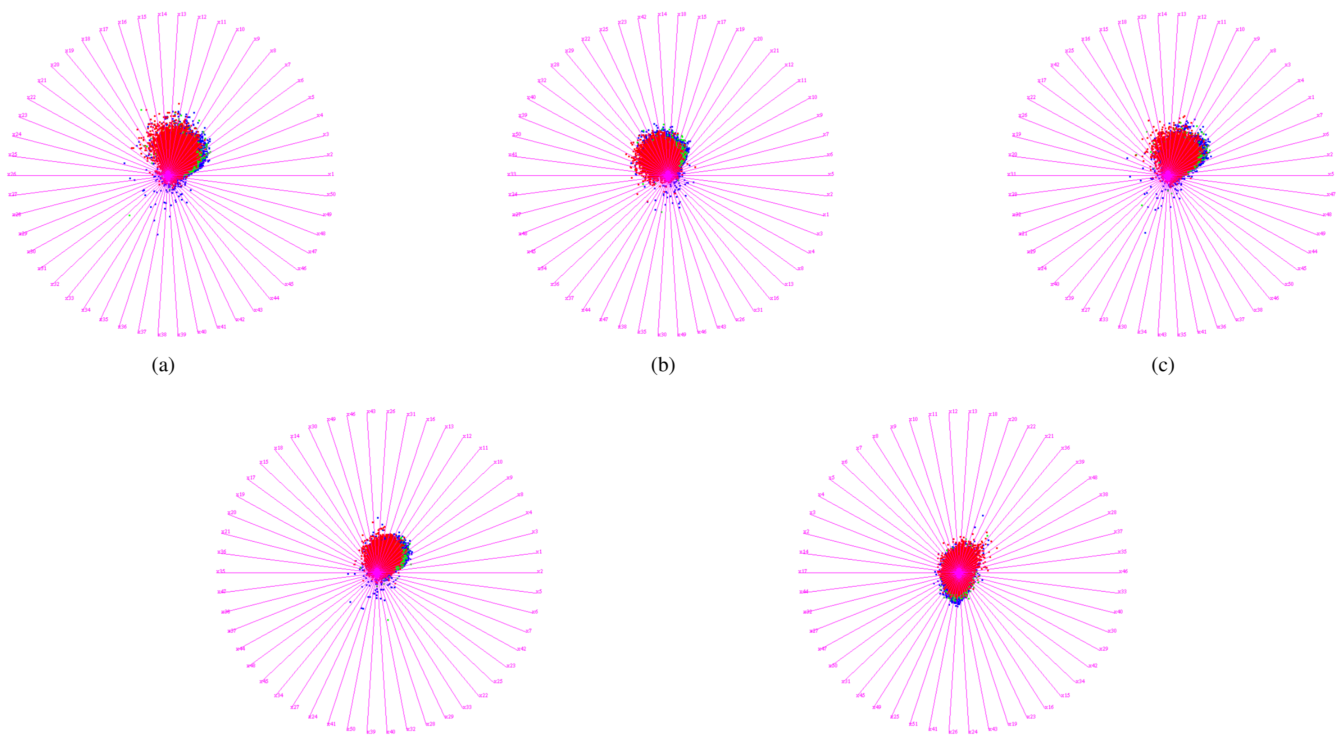


Fig. 4: Radviz visualization the DB20 data set. 4a Original. 4b AGENS. 4c WSORT. 4e GA.



(a) (b) (c) (d) (e)

Fig. 5: Radviz visualization of the DB40 data set. 5a Original. 5b AGENS. 5c WSORT. 5d SBAA. 5e GA.



(a) (b) (c) (d) (e)

Fig. 6: Radviz visualization of the DB50 data set. 6a Original. 6b AGENS. 6c WSORT. 6d SBAA. 6e GA.

more data sets with higher data dimensions will also be tested to validate our approach. More importantly, we are planning to improve the cost function to cope with noisy data.

VII. Acknowledgements

The paper is financially supported by The Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.01-2012.04.

References

[1] Georgia Albuquerque, Martin Eisemann, Dirk J. Lehmann, Holger Theisel, and Marcus A. Magnor. "Quality-Based Visualization Matrices." In Proceedings of Vision Modeling and Visualization (VMV), Braunschweig, Germany, pp. 341-349, 2009.

[2] Georgia Albuquerque, Martin Eisemann, Dirk J. Lehmann, Holger Theisel, and Marcus Magnor. "Improving the visual analysis of high-dimensional datasets using quality measures." The IEEE Conference on Visual Analytics Science and Technology (VAST) 2010, pp. 19-26, 2010.

[3] Mihael Ankerst, Stefan Berchtold, and Daniel A. Keim. "Similarity clustering of dimensions for an enhanced visualization of multidimensional data." The IEEE Conference on Information Visualization (IV) 1998, pp. 52-60, 1998.

[4] Almir Olivette Artero, and Maria Cristina Ferreira de Oliveira. "Viz3d: Effective exploratory visualization of large multidimensional data sets." In The 17th Symposium on Computer Graphics and Image Processing 2004, Brazilian, pp. 340-347, 2004.

[5] Almir Olivette Artero, Maria Cristina Ferreira de Oliveira, and Haim Levkowitz. "Enhanced high dimensional data visualization through dimension reduction and attribute arrangement." The Tenth International Conference on Information Visualization 2006, pp. 707-712, 2006.

[6] Stuart K. Card, Jock D. Mackinlay, and Ben Schneiderman, eds. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999.

[7] Luigi Di Caro, Vanessa Frias-Martinez, and Enrique Frias-Martinez. "Analyzing the role of dimension arrangement for data visualization in radviz," *Advances in Knowledge Discovery and Data Mining*, Springer Berlin Heidelberg, pp. 125-132, 2010.

[8] Karen Daniels, Georges Grinstein, Adam Russell, and Mason Glidden. "Properties of normalized radial visualizations." Information Visualization 11, No. 4, pp. 273-300, 2012.

[9] D.E. Goldberg, *Genetic Algorithm in Search, Optimization, ans Machine Learning*, Addison-Wesley Pubs., 1989.

[10] Julian Heinrich, and Daniel Weiskopf. "State of the Art of Parallel Coordinates." In Eurographics 2013-State of the Art Reports, The Eurographics Association, pp. 95-116, 2013.

[11] Patrick Hoffman, Georges Grinstein, Kenneth Marx, Ivo Grosse, and Eugene Stanley. "DNA visual and analytic data mining." In Proceedings of the IEEE Information Visualization Symposium 1997, pp. 437-441, 1997.

[12] Patrick Hoffman, Georges Grinstein, and David Pinkney. "Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations." In Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM internation conference on Information and knowledge management, pp. 9-16, 1999.

[13] Eser Kandogan. "Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions." In Proceedings of the IEEE Information Visualization Symposium 2000, vol. 650, 2000.

[14] Gregor Leban, Bla Zupan, Gaj Vidmar, and Ivan Bratko. "Vizrank: Data visualization guided by machine learning." Data Mining and Knowledge Discovery 13, No. 2, pp. 119-136, 2006.

[15] Van Long Tran: "Laplacian Star Coordinates for Visualizing Multidimensional Data", The 10th IEEE RIVF International Conference on Computing and Communication Technologies, RIVF 2013, pp. 259-264, 2013.

[16] Melanie Mitchell, *An introduction to Genetic Algorithm*, MIT Press, 1999.

[17] Franz Rothlauf, *Representations for Genetic and Evolutionary Algorithms*, Springer-Verlag, 2006.

[18] John Sharko, Georges Grinstein, and Kenneth A. Marx. "Vectorized radviz and its application to multiple cluster datasets." IEEE Transactions on Visualization and Computer Graphics, No. 6, pp. 1444-1427, 2008.

[19] Jing Wang, Wei Peng, Matthew O. Ward, and Elke A. Rundensteiner. "Interactive hierarchical dimension ordering, spacing and filtering for exploration of high dimensional datasets." IEEE Symposium on Information Visualization 2003 (INFOVIS 2003), pp. 105-112, 2003.