SEMARAK ILMU
PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

Random Dimension Manipulation for Efficient High-Dimensional Data Clustering

Ummu Hani' Hair Zaki[1], Izyan Izzati Kamsani[1,*], Roliana Ibrahim[1], Norzehan Sakamat[2], Eser Kandogan[3]

[1] Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[2] Faculty of Information Technology and Quantitative Science, Universiti Teknologi Mara, Kampus Dengkil, 43800 Dengkil, Selangor
[3] Megagon Labs, 444 Castro St #900, Mountain View, CA 94041, United States

ARTICLE INFO

Article history:
Received 24 August 2023
Received in revised form 20 February 2024
Accepted 15 August 2024
Available online 2 September 2024

Keywords:
High-Dimensional Data; Clustering, Correlation, Dimensions Dependencies, Dimension Arrangement; Dimension Scaling; Random Dimension Manipulation; Star Coordinates; Visualization

ABSTRACT

High-dimensional data is collected from various sources, fields, and applications such as medicine, science, business and more to provide helpful information to others. Unfortunately, the complexity of high-dimensional data has made it difficult to interpret and understand. As a result, sophisticated data analysis is required to extract knowledge and information from it. This can be illustrated through a visualization presentation. However, overlap between data can occur during visualization as data increases. Indirectly, it can cause a cluttered visual presentation. As a result, it affects the visual perception of high-dimensional data patterns. High-dimensional data can be deeply explored using dimension arrangement and scaling to overcome it. The arrangement of dimensions is essential since the relationship between these dimensions can influence the existence of an efficient cluster. This dimension is arranged based on the correlation value. The dimension that is more related will be placed next to each other. While performing clusters, dimensions will be scaled in or out. These features are available through Star Coordinate (SC) technique. This paper aims to conduct an exploratory data analysis in the SC environment where users can visualize and interact in a low-dimensional data visualization space. This paper demonstrates data dimensions manipulation's importance in structuring the projected space layout using two data sets. As a conclusion, formation of clusters was crucial and manipulation of data dimensions were essential to structure the projected space layout. The proposed approach has helped users find significant cluster formations by randomizing the scaling and order of dimensions.

1. Introduction

Multivariate, high-dimensional, and multidimensional data are all synonyms [1]. The generation of multidimensional data is increasingly common in a wide range of research fields, such as healthcare [2], transportation [3,4], weaponry system simulation [5] and agriculture [6]. The

* *Corresponding author.*
E-mail address: izyanizzati@utm.my

https://doi.org/10.37934/araset.51.1.129140

complexity of these data has made them vital to many fields. Hence sophisticated data analysis method is required to extract knowledge from it.

Hence sophisticated data analysis method to overcome related issues concerning high dimensional data or visual analysis have been gravely considered in light of the useful and beneficial knowledge that can be derived from interacting with the high dimensional data.

A common issue when working with high-dimensional data is the well-known curse of dimensionality [7]. This is due to the overlapping data when data is gets larger. The curse of dimensionality can be overcome by undergoing pre-processing techniques. This technique can reduce dimensionality [8]. Reducing the dimensionality is essential to avoid clutter visualization presentation [9]. In detail, the critical dimensions will be visualized, while the less important ones will be ignored and detached from the visualization.

Another method of extracting meaningful information is recognizing user interactions for visual analysis. Therefore, it facilitates decision-making or concludes something. In another way, adding interactive visualizations can enable and guide data analysis processes and perform various tasks such as exploration, feature selection, and clustering. Thus, exploratory data analysis can act as a method designed to get an overview of data. Various techniques are used to generate a visual representation of high-dimensional data.

Exploring high-dimensional data is difficult [10]. Large numbers of dimensions can cause clutter in projecting the visualization. Thus, it is difficult for users to discover and interpret the data. Dimension manipulation can significantly affect visualization expression. It is believed that the relationships among adjacent dimensions are more visible than those among dimensions positioned far from each other.

In the SC environment, dimensional arrangement and dimensional scaling assist the visual discovery of patterns within the data. Different arrangements of dimensions and scaling features are believed to have different cluster appearance effects [11]. Previous studies on dimension manipulation suggested that properly manipulating attribute dimensions is paramount to uncovering patterns in SC layouts, and the order of the dimension plays an essential role in this context [12]. Users can gain the perspective they choose with this feature. For example, Yang *et al.,* [13] proposed an interactive hierarchical dimension ordering, spacing, and filtering approach based on dimension hierarchies derived from similarities among dimensions, called DOSFA. They used a bottom-up agglomerate dimension clustering algorithm.

While Artero *et al.,* [14] introduced a technique called SBAA (Similarity-Based Attribute Arrangement), the data dimension highly correlated with others is removed from a visualization display, either automatically or by a user activation. This method helps to reduce visualization clutter and assists the user in finding out the most relevant data attributes for a particular task. Effective dimension manipulation, such as dimension arrangement and scaling, is critical for visually exploring such data sets. High-dimensional visualization techniques like Parallel Coordinates and SC directly show the relationships between dimensions by how the dimensions are arranged and scaled. They facilitate the visual discovery of patterns within the data. Mainly, the evaluation of the effectiveness of the dimension arrangement from the perspective of visual information is typically carried out using human intervention. However, to overcome that, Caro *et al.,* [15] used the Davies-Bouldin index for cluster analysis to compute the visual quality of the information being plotted by Radviz.

Besides, Suematsu *et al.,* [16] present a framework that displays an arrangement of low-dimensional data visualization spaces. This technique takes the dimensions of the input data sets into groups of lower dimensions based on their correlations or other relationships. Dimensions with a lower value will be visualized as independent rectangular spaces. In addition, Stahnke *et al.,* [17] contributed to the concept of probing. It is a method that allows the viewers to see the approximation errors, question the positioning of the dimension on why the dimension is placed far apart or close together, compare them to each other, and visualize the influence of data dimensions on the projection space.

Furthermore, Ameur *et al.,* [18] proposed an approach for visual mining association rules (classification) from data or clusters based on dimensions reordering using an adapted parallel set. They used a parallel set [19] as a Knowledge Discovery in Databases (KDD) tool to visualize, explore, interpret, and extract association rules from data and clusters. However, with their manner, i.e., displaying the clusters and rules simultaneously unless other, they use visualization techniques to visualize the results of the association rules mining process. Also, Abdul Khalid *et al.,* [11] assisted novice users in manipulating interactive SC features to obtain better cluster visualization for user understanding. Thus, it assists the users in other cluster exploration – between dimensions angle in 360 degrees-based correlation value and within an equal angle.

In another study, He *et al.,* [3] contributed to solving trajectory data problems for mining trucks. Because the substantial analysis of trajectory data may involve dimensions that exceed space, they found that larger dimensions can cause difficulties in analyzing the internal correlation between dimensions. Hence, they developed a general-purpose modelling framework for high-dimensional trajectory databases, which can gain helpful knowledge about mining trucks. The solution is based on the SC environment and is implemented in a scalable spatial database system centred on actual mining truck data. Talukder & Deb [20] have demonstrated a new visualization approach (PaletteStarViz), where this approach uses an SC environment to address the issue of high-dimensional Pareto-optimal data visualization to aid a multi-criteria decision-making (MCDM) task.

Zhu *et al.,* [21] developed a methodology for three-dimensional (3D) radial visualization (RadViz) of high-dimensional data. Previously, classical two-dimensional (2D) RadViz visualized multivariate data in the 2D plane. This method maps every observation to a point inside the unit circle. Then RadViz3D allocates anchor points uniformly on the 3D unit sphere. They proved that this uniform distribution provides the best visualization with minimal artificial visual correlation for data with uncorrelated variables.

Cheng *et al.,* [22] have addressed the problem of identifying interfaces and boundary nodes in clustered social networks. This node can be identified as the interface between unclear clusters, or their boundaries cannot be easily assigned to any cluster. For example, identifying these nodes is relevant in marketing applications. Unfortunately, this identification task is less well studied than other network analysis tasks such as clustering, central node identification and motif detection. They approach this task by deriving new geometric features from network structures that naturally lend themselves to an interactive visual approach to identify interfaces and boundary nodes.

Above all the dimensional arrangement and dimensional scaling practices, this paper aims to represent an exploratory data analysis in an SC environment where users can visualize and interact in low-dimensional data visualization spaces. This paper also focuses on assisting analysts in examining clusters using SC representations to understand the details of the data. As well as to understand the significance of data dimensions in structuring the layout in the projected space. This work is organized as follows: Section 2 presents the related works; Section 3 explains the methodology; Section 4 shows the dimension manipulation result; Section 5 discusses the result obtained; the work is concluded in Section 6.

2. Methodology

Figure 1 shows the workflow of the random dimension manipulation experiment. The experiment has five phases, producing a crucial result for the next phase.

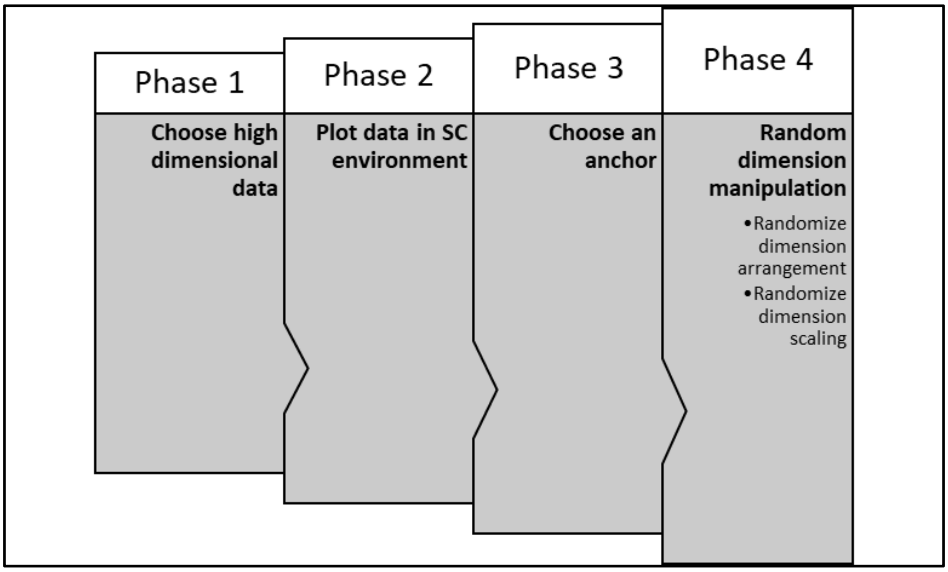


Fig. 1. Workflow of random dimension manipulation in SC environment

2.1 Phase 1: Choose High Dimensional Data

Two data sets are utilized in this experiment. Firstly, the credit card data set [23] summarizes the usage behavior of about 9,000 active credit card holders during six months with 18 behavioral dimensions. Secondly, the diabetes data set [24] is used to diagnostically predict whether a patient has diabetes based on specific diagnostic measurements included in the data set. Table 1 and Table 2 list the dimension description for both data sets.

Table 1
Dimension description of the credit card data set

Dimensions	Description
cust_id	Identification of credit card holder (categorical)
balance	Balance amount left in their account to make purchases
balance_frequency	How frequently the balance is updated, a score between 0 and 1 (1 = frequently updated, 0 = not frequently updated)
purchases	Number of purchases made from the account
oneoff_purchases	Maximum purchase amount did in one-go
installments_purchases	Amount of purchase done in instalments
cash_advance	Cash in advance given by the user
purchases_frequency	How frequently the purchases are being made, a score between 0 and 1 (1 = frequently purchased, 0 = not frequently purchased)
oneoff_purchases_frequency	How frequently purchases are happening in one go (1 = frequently purchased, 0 = not frequently purchased)
purchases_installments_frequency	How frequently purchases in instalments are being made (1 = frequently done, 0 = not frequently done)
cash_advance_frequency	How frequently is the cash in advance being paid
cash_advance_trx	Number of transactions made with "*Cash in Advanced*"
purchases_trx	Number of purchase transactions made

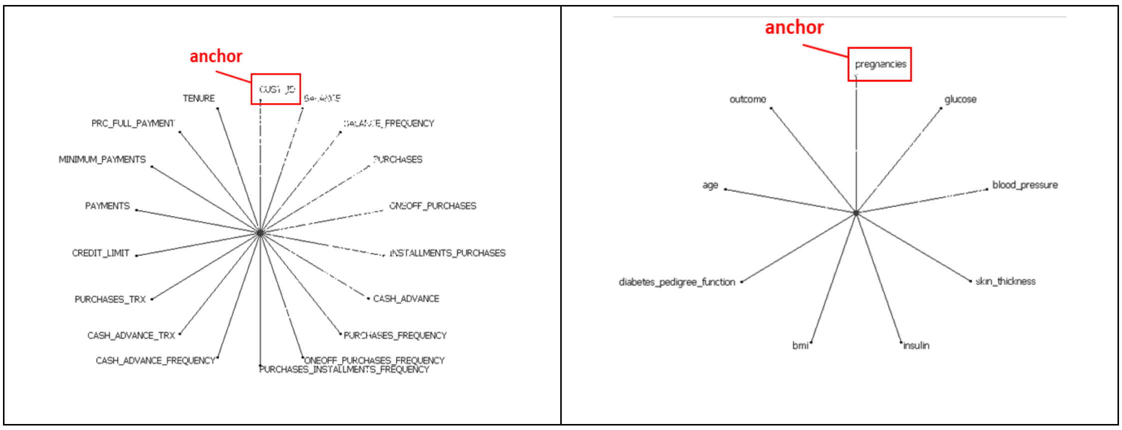
Dimensions	Description
credit_limit	Limit of credit card for user
payments	Amount of payment done by the user
minimum_payments	Minimum number of payments made by the user
prc_full_payment	Per cent of full payment paid by the user
tenure	Tenure of credit card service for user

Table 2
Dimension description of the diabetes data set

Dimensions	Description
pregnancies	Number of times pregnant
glucose	Plasma glucose concentration 2 hours in an oral glucose tolerance test
blood_pressure	Diastolic blood pressure (mmHg)
skin_thickness	Triceps skin fold thickness (mm)
insulin	2-Hour serum insulin (mIU/L)
BMI	Body mass index (weight in kg/(height in m)^2)
diabetes_pedigree_function	diabetes pedigree function
age	Age of patient
outcome	Class variable (0 or 1) 268 of 768 are 1, the others are 0

2.2 Phase 2: Plot Data in SC Environment

In the SC environment, the data was then plotted. Without first selecting which relevant data dimension should be displayed, all data were visualized in this phase. Each dimension's default position was determined based on the order of data dimensions in raw data. The default position is organized based on the code in ascending order, as illustrated in Figure 2(a) and 2(b) for the credit card and diabetes data set.



(**a**) (**b**)

Fig. 2. Default anchor for the (**a**) credit card and (**b**) diabetes data set

2.3 Phase 3: Choose an Anchor

During the visualization, a particular dimension was randomly selected as an anchor. Two elements should be considered during the selection: independent and dependent dimensions. The anchor serves as an independent dimension. The non-anchor dimensions are called dependent dimensions. The anchor can be alternated to examine the effects on the dependent dimensions. The interchange of independent dimensions and the arrangement of dependent dimensions affects the appearance of cluster visualization. The arrangement of dependent dimensions was tested and measured in Phase 4.

2.4 Phase 4: Random Dimensions Manipulation

Phase 4 includes randomized dimension arrangement and scaling. The task information is described below.

2.4.1 Random dimension arrangement

Dimensions were randomly rearranged to exhibit cluster patterns. This study was not interested in the rationale behind placing the dimensions adjacent to one another. It essentially focused on identifying the clusters. Dimensions can be changed randomly based on the user's needs to produce distinct clusters. Finally, each data dimension's angle needs to be distributed equally. Section 4 shows the results obtained from random dimension manipulation using the selected data set (card credit and diabetes data set).

2.4.2 Random dimension scaling

After dimensions were randomly arranged, the random dimension scaling experiment was conducted. A clear cluster presentation from overlapping data visualization can be obtained through dimension scaling. As previously indicated, users have limited knowledge of using the SC technique. As a result, users randomly resized the dimensions to improve visibility and make clusters more obvious. A scale was applied to a particular dimension. It was accomplished by adjusting each dimension's length. The chosen dimensions can be lengthened (expanding) or shortened (collapsing) following any respected dimension. As a result, it can aid users in comprehending many cluster trends that emerge during visualization.

3. Results

Results were obtained using two (2) different anchors for the credit card data set, two (2) different anchors for the diabetes data set and Figures 3 and 4. Both figures are randomly plotted with all dimensions using the same angle in the SC environment. Dimensions are plotted in sequence from the data.

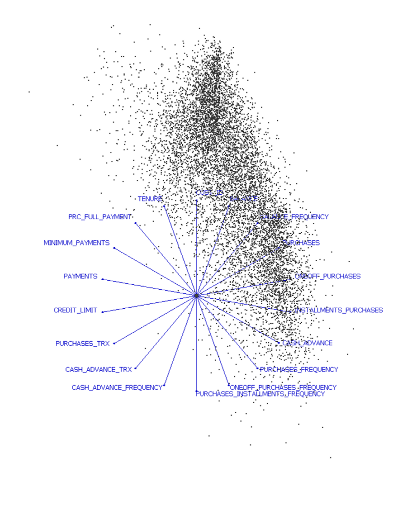
3.1 Random Dimension Arrangement

Figures 3 and 4 show that both figures are randomly plotted with all dimensions using the same angle in the SC environment. Dimensions are plotted in sequence from the data. The same data set is shown in Figure 3(a) and 3(b), but with different axes and anchor dimension arrangement. Angles for every data dimension in this experiment were distributed equally, clockwise and in a circle arrangement. This experiment aimed to identify any cluttered and overlapped data distribution. Another aim was to identify if interpreting high-dimensional data could be affected by different dimension arrangements in a data pattern distribution.

The relationship between dimensions was not very clearly portrayed, as shown in Figure 3. As a result, it was challenging to determine the relationship between the dimensions and extract relevant information. Figure 3(a) shows the data distribution when the dimensions are arranged in the default arrangement. The positions of data dimensions are placed regardless of how closely connected data dimensions are. Simply displaying the default dimension arrangement is not that revealing. An excellent first step is to use other dimensions as the anchor to explore more about the relationship between other data dimensions.

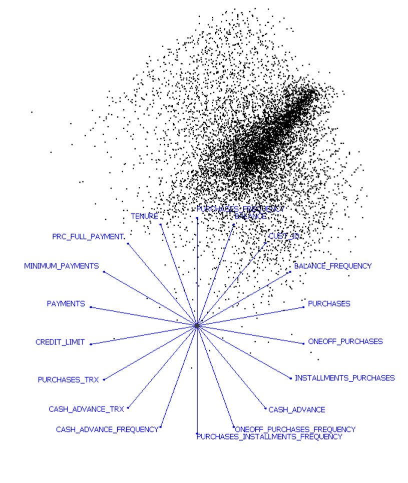
A new cluster can be recognized as the balance dimension presented as an anchor in Figure 3(b). In Figure 3(b), deciding purchases frequency dimension as an anchor, a cluster emerges clearly when data distribution becomes more compressed at the upper SC plot. The data were plotted using a trial-and-error procedure based on other dimensions to obtain valuable information. The data distribution pattern becomes more expansive as the anchor changes.

Without knowing the rationale for arranging the orders of data dimensions, Figure 3(c) still expresses the appearance of clusters. Packed cluster appearances are revealed at the upper side of the SC plot, as seen in Figure 3(d). The appearance was created using the bare minimum of prior knowledge of data dimension arrangement. Unfortunately, crucial data still cannot be understood in Figure 3(e) – 3(f) because there are no apparent patterns in the data distribution.

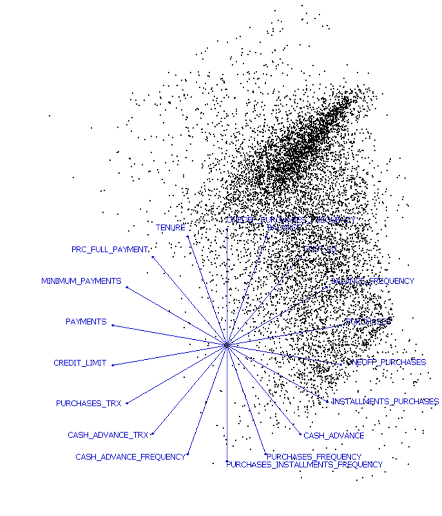


(a) cust id

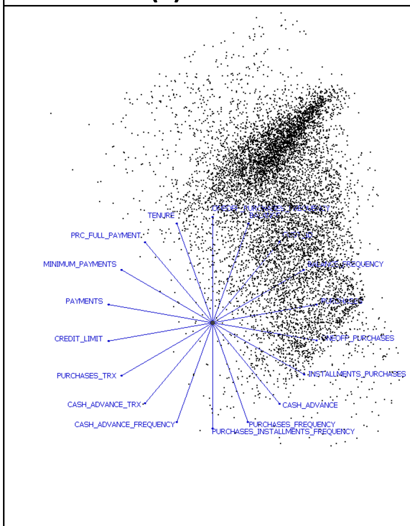
c



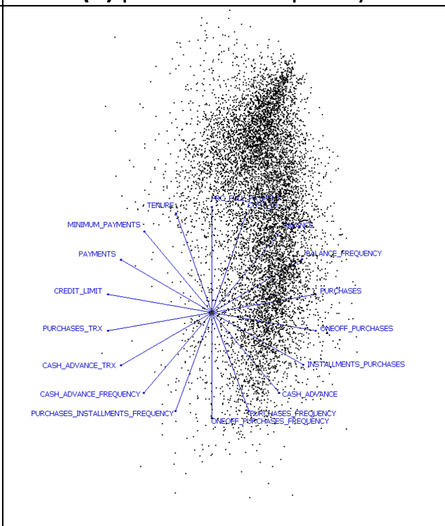
(b) purchases frequency



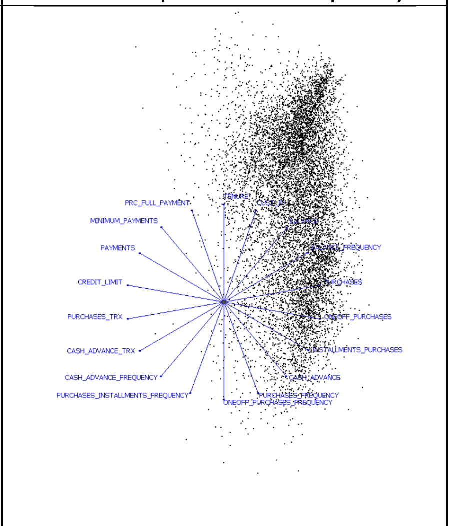
© oneoff purchases frequency



(d) oneoff purchases frequency



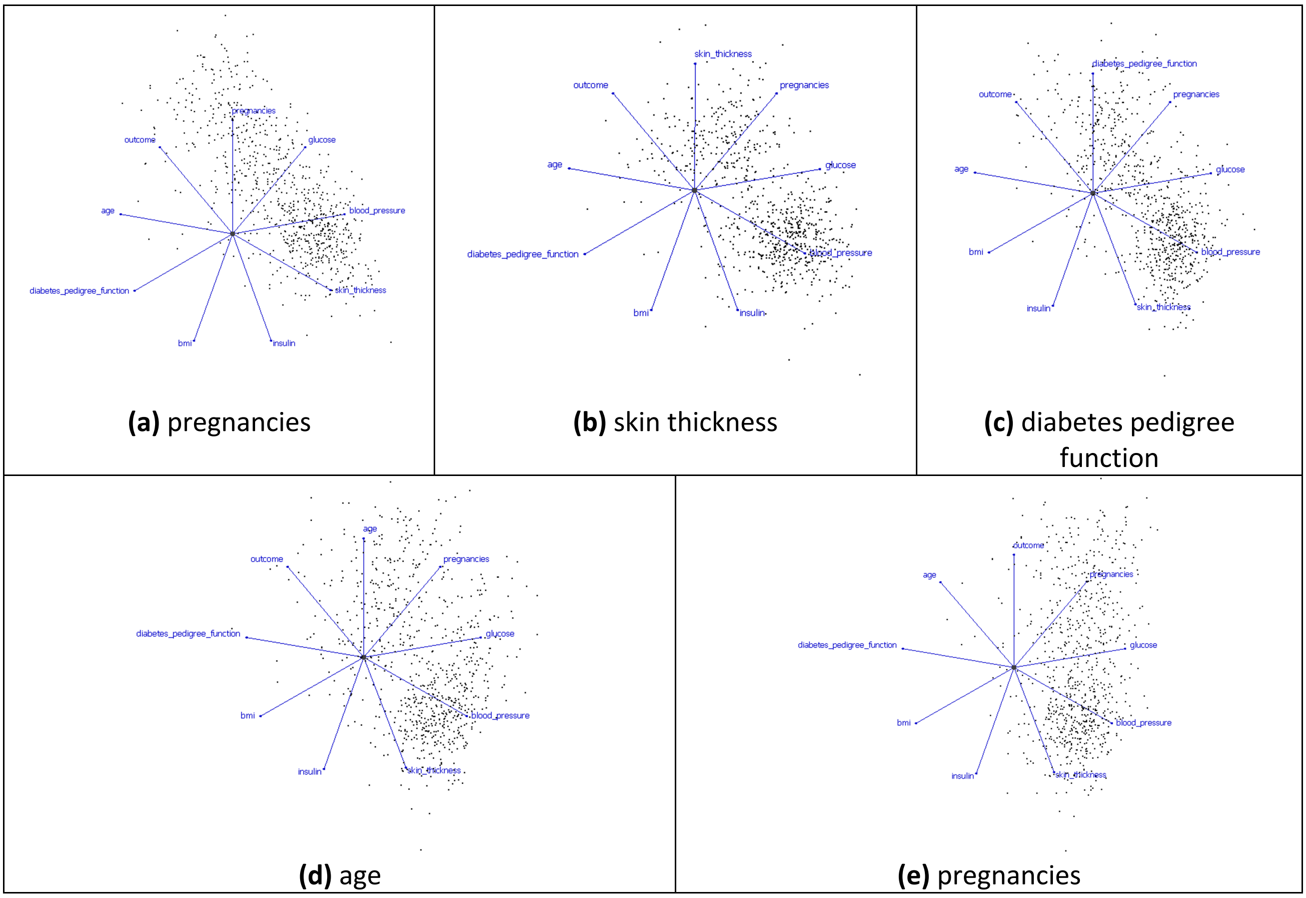
(g) prc full payments



(h) tenure

Fig. 3. Random arrangement with 135ifferent anchors using the same angle in SC for credit card data set

The dimensions arrangement for the diabetes dataset is shown in Figure 4. As usual, the first arrangement was plotted based on the default dimension order from the raw dataset, with the pregnancies dimension as an anchor. Another dimension was selected as an anchor to see how the other cluster emergence differed from Figure 4(a).



(a) pregnancies **(b)** skin thickness **(c)** diabetes pedigree function **(d)** age **(e)** pregnancies

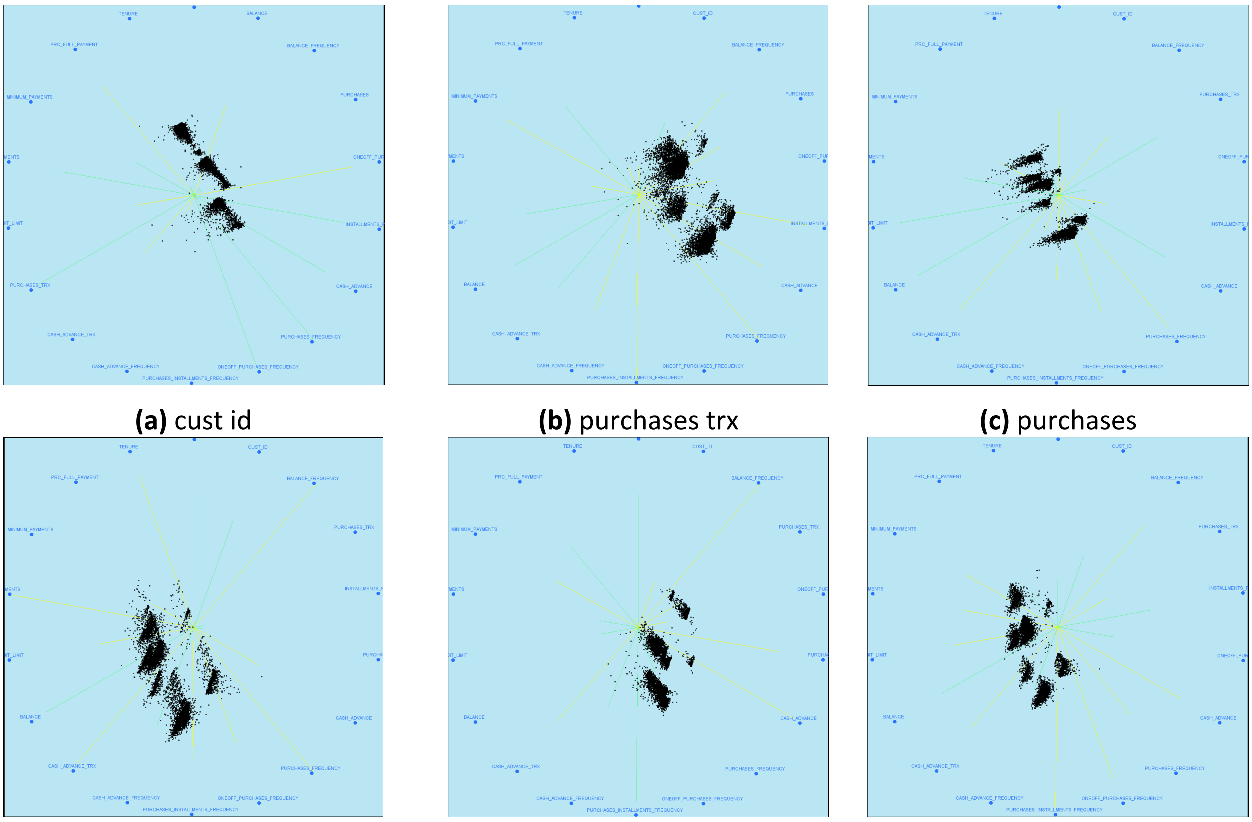
Fig. 4. Random arrangement with different anchors using the same angle in SC for the diabetes data set

In Figure 4(b), two light clusters have been observed at the SC plot's upper center and bottom right. While the diabetes pedigree function acts as an anchor in Figure 4(c), data distribution becomes more expensive, and two light clusters emerge. In Figure 4(d), the data distribution becomes crowded at the upper and right sides of the SC plot, with a dense cluster at the right bottom has been discovered. As in Figure 4(e), the data distribution moves to the SC plot's right side, and no apparent cluster emerges.

Indeed, customizing dimension arrangement through a time-consuming trial-and-error approach is a challenging task. It took a long time to use the trial-and-error procedure and obtain accurate cluster results. The proposed method was therefore put into practice to get over this issue. However, it was difficult for users with little background in the SC technique to comprehend this method, and it was time-consuming to study to master the cluster patterns.

3.2 Random Dimension Scaling

In this part, the result of dimension scaling will be described. Figures 5 and 6 illustrate dimension scaling for the credit card dataset and diabetes dataset, respectively. The green axis represents positive values, while the axis with yellow color represents negative values. The scaling feature was used for various anchors, as indicated by the red arrows. The green color dimensions represent the positive value from 0 to 1, and the yellow represents the negative value from -1 to 0 [25].

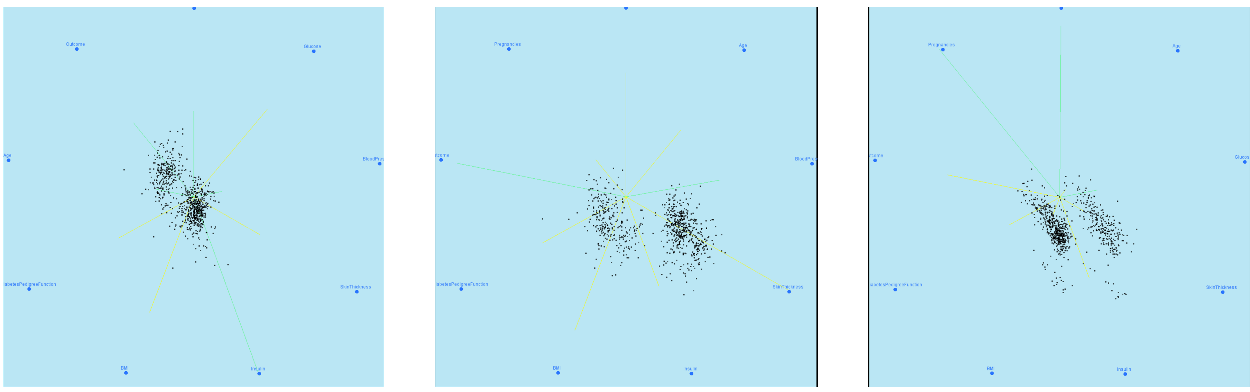


(a) cust id **(b)** purchases trx **(c)** purchases

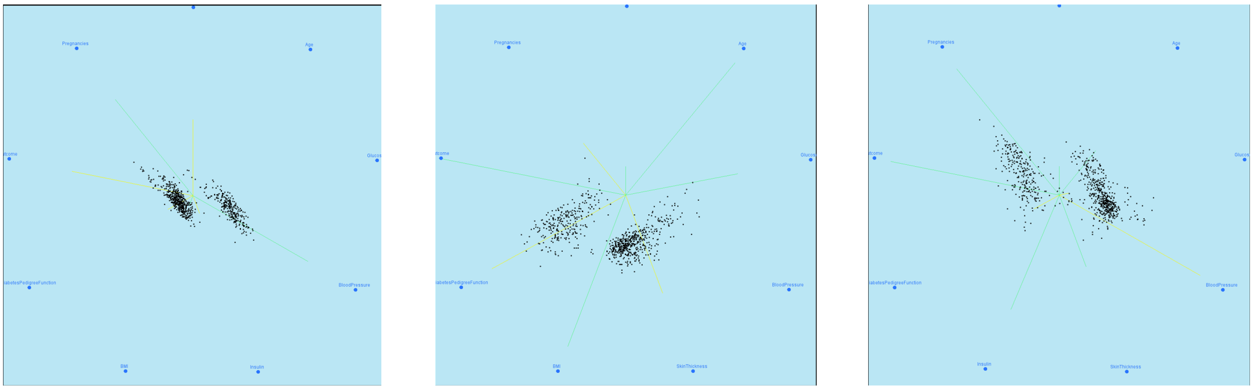
(d) oneoff purchases **(e)** cash advance **(f)** purchases frequency

Fig. 5. Random dimension scaling with different anchors and data dimensions using the same angle in SC for credit card dataset

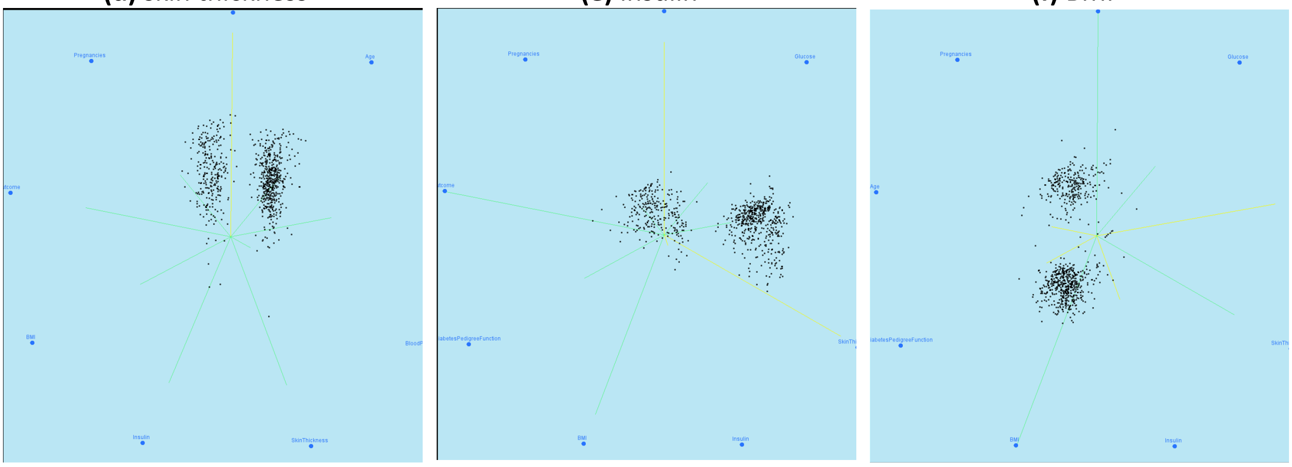
Figure 5(a), the dimension arrangement is organized in default order as in the raw dataset sequence. Other dimensions have been scaled in various lengths. From this illustration, it showed that four significant clusters have appeared. As for Figure 5(b) – 5(f), the different clusters emerged with numerous data distribution formations. Each visualization represents a different dimension as its anchor. Due to the high dimensional data in the credit card dataset, with 18 dimensions, every detail's name will not be stated. The most important thing here is that this experiment managed to represent data clusters, and the user can gain important information from it.



(a) pregnancies **(b)** glucose **(c)** blood pressure



(d) skin thickness **(e)** insulin **(f)** BMI



(g) diabetes pedigree function **(h)** age **(i)** outcome

Fig. 6. Random dimension scaling with different anchors and data dimensions using the same angle in SC for diabetes dataset

The initial dimension scaling attempt is depicted in Figure 6(a). Using the pregnancies dimension as an anchor, we can get two clusters by scaling down several dimensions such as diabetes pedigree function, age, skin thickness, blood pressure, and pregnancies. Moreover, we are scaling up specific dimensions like BMI, insulin, glucose, and outcome. A smaller (shorter) axis suggests a lesser contribution to visualization. Compared to other data dimensions, diabetes pedigree function, age, skin thickness, blood pressure, and pregnancies have a lower contribution to visualization in Figure 6(a). However, it is unclear to what extent the dimensions of this data suggest that they have minimal impact on cluster performances.

Throughout Figure 6(b) – 6(i), there is apparent cluster formation. Some anchors are being scaled down, for example, skin thickness, insulin, and BMI. Furthermore, specific anchors such as glucose, blood pressure, diabetes pedigree function, age, and outcome are being scaled up. Based on this result in Figure 6(b) – 6(i), within this experiment scope, it can be suggested that skin thickness, insulin, and BMI do not have a higher impact on diabetes diagnosis. On the other hand, glucose, blood pressure, diabetes pedigree function, age, and outcome relate to diabetes diagnosis to predict whether a patient has diabetes.

4. Conclusions

This paper discovered that data are scaled through trial and error if clusters appear without knowing how much each data dimension contributed to the final visualization. It was complicated and an arduous task to get reliable clusters results using the trial-and-error method. Nevertheless, it provided users with much needed experience and knowledge especially on the importance of dimension manipulation for high- dimensional data. Interactive operations in the SC environment aim to explore and visualize the dataset in different projection planes. Users can often get much more information from dynamic visualization, especially in an SC environment. This paper presents random dimension manipulation for visualizing clusters using high-dimensional data. The contribution of this work is to enable users to interact with the SC technique using random dimension arrangement and scaling features. It allows a clear cluster presentation from overlapped data visualization by randomly alternating different dimensions as an anchor and randomly adjusting dimension axes' length (scale down or scale up). However, to improve high dimensional data visualization, it is suggested that correlation measurement is conducted to aid dimension axes manipulation and enable search for axes configurations that maximize the visual cluster.

Acknowledgement

This work was funded by the Ministry of Higher Education under Fundamental Research Grant Scheme (FRGS/1/2021/ICT03/UTM/02/2).

References

[1] Zhou, Liang, and Daniel Weiskopf. "Indexed-points parallel coordinates visualization of multivariate correlations." *IEEE transactions on visualization and computer graphics* 24, no. 6 (2017): 1997-2010.
[2] Mohedano-Munoz, Miguel A., S. Alique-García, Manuel Rubio-Sánchez, Laura Raya, and Alberto Sánchez. "Interactive visual clustering and classification based on dimensionality reduction mappings: A case study for analyzing patients with dermatologic conditions." *Expert Systems with Applications* 171 (2021): 114605. https://doi.org/10.1016/j.eswa.2021.114605
[3] He, Jing, Haonan Chen, Lingyu Wang, and Yebin Zou. "i-tStar (3D): Three-Dimensional Interactive Trajectory Star Coordinates." In *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pp. 747-752. IEEE, 2020. https://doi.org/10.1109/ICTC49870.2020.9289363
[4] He, Jing, Lingxiao Li, and Xin Wang. "i-tStar: Interactive Trajectory Star Coordinates." In *2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture*, pp. 2044-2047. 2021. https://doi.org/10.1145/3495018.3501046
[5] Zheng, Lizhen, Lu Gao, and Tonggang Yu. "Visualization analysis method based on multidimensional scale transformation for indicators on simulation data of weapon system." In *Journal of Physics: Conference Series*, vol. 1983, no. 1, p. 012081. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1983/1/012081
[6] López, Iván Darío, Apolinar Figueroa, and Juan Carlos Corrales. "Multi-dimensional data preparation: A process to support vulnerability analysis and climate change adaptation." *IEEE Access* 8 (2020): 87228-87242. https://doi.org/10.1109/ACCESS.2020.2992255
[7] Molchanov, Vladimir, and Lars Linsen. "Shape-preserving star coordinates." *IEEE Transactions on Visualization and Computer Graphics* 25, no. 1 (2018): 449-458. https://doi.org/10.1109/TVCG.2018.2865118
[8] Genender-Feltheimer, Amy. "Visualizing high dimensional and big data." *Procedia Computer Science* 140 (2018): 112-121. https://doi.org/10.1016/j.procs.2018.10.308
[9] Zhou, Zhiguang, Yuming Ma, Yong Zhang, Yanan Liu, Yuhua Liu, Lin Zhang, and Shengchun Deng. "Context-aware visual abstraction of crowded parallel coordinates." *Neurocomputing* 459 (2021): 23-34. https://doi.org/10.1016/j.neucom.2021.05.005
[10] Self, Jessica Zeitz, Michelle Dowling, John Wenskovitch, Ian Crandell, Ming Wang, Leanna House, Scotland Leman, and Chris North. "Observation-level and parametric interaction for high-dimensional data analysis." *ACM Transactions on Interactive Intelligent Systems (TiiS)* 8, no. 2 (2018): 1-36. https://doi.org/10.1145/3158230
[11] Khalid, Noor Elaiza Abdul, Saadiah Yahya, and Izyan Izzati Kamsani. "Dimension Manipulation Based 360 Degrees to Excavate Clusters in Star Coordinate." In *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)*, pp. 1-6. IEEE, 2019. https://doi.org/10.1109/CAIS.2019.8769498
[12] Feng, Kang, Yunhai Wang, Ying Zhao, Chi-Wing Fu, Zhanglin Cheng, and Baoquan Chen. "Cluster aware star coordinates." *Journal of Visual Languages & Computing* 44 (2018): 28-38. https://doi.org/10.1016/j.jvlc.2017.11.003
[13] Yang, Jing, Wei Peng, Matthew O. Ward, and Elke A. Rundensteiner. "Interactive hierarchical dimension ordering,

spacing and filtering for exploration of high dimensional datasets." In *IEEE Symposium on Information Visualization 2003 (IEEE Cat. No. 03TH8714)*, pp. 105-112. IEEE, 2003. https://doi.org/10.1109/INFVIS.2003.1249015
[14] Artero, Almir Olivette, Maria Cristina Ferreira de Oliveira, and Haim Levkowitz. "Enhanced high dimensional data visualization through dimension reduction and attribute arrangement." In *Tenth International Conference on Information Visualisation (IV'06)*, pp. 707-712. IEEE, 2006. https://doi.org/10.1109/IV.2006.49
[15] Di Caro, Luigi, Vanessa Frias-Martinez, and Enrique Frias-Martinez. "Analyzing the role of dimension arrangement for data visualization in radviz." In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 125-132. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. https://doi.org/10.1007/978-3-642-13672-6_13
[16] Suematsu, Haruka, Zheng Yunzhu, Takayuki Itoh, Ryohei Fujimaki, Satoshi Morinaga, and Yoshinobu Kawahara. "Arrangement of low-dimensional parallel coordinate plots for high-dimensional data visualization." In *2013 17th International Conference on Information Visualisation*, pp. 59-65. IEEE, 2013. https://doi.org/10.1109/IV.2013.7
[17] Stahnke, Julian, Marian Dörk, Boris Müller, and Andreas Thom. "Probing projections: Interaction techniques for interpreting arrangements and errors of dimensionality reductions." *IEEE transactions on visualization and computer graphics* 22, no. 1 (2015): 629-638. https://doi.org/10.1109/TVCG.2015.2467717
[18] Ameur, Khadidja, Nadjia Benblidia, and Saliha Oukid Khouas. "Dimensions reordering for visual mining of association rules using parallel set." *International Journal of Data Analysis Techniques and Strategies* 8, no. 4 (2016): 296-315. https://doi.org/10.1504/IJDATS.2016.081362
[19] Kosara, Robert, Fabian Bendix, and Helwig Hauser. "Parallel sets: Interactive exploration and visual analysis of categorical data." *IEEE transactions on visualization and computer graphics* 12, no. 4 (2006): 558-568. https://doi.org/10.1109/TVCG.2006.76
[20] Talukder, AKM Khaled Ahsan, and Kalyanmoy Deb. "PaletteViz: A visualization method for functional understanding of high-dimensional Pareto-optimal data-sets to aid multi-criteria decision making." *IEEE Computational Intelligence Magazine* 15, no. 2 (2020): 36-48. https://doi.org/10.1109/MCI.2020.2976184
[21] Zhu, Yifan, Fan Dai, and Ranjan Maitra. "Fully Three-Dimensional Radial Visualization." *Journal of Computational and Graphical Statistics* 31, no. 3 (2022): 935-944. https://doi.org/10.1080/10618600.2021.2020129
[22] Cheng, Shenghui, Joachim Giesen, Tianyi Huang, Philipp Lucas, and Klaus Mueller. "Identifying the skeptics and the undecided through visual cluster analysis of local network geometry." *Visual Informatics* 6, no. 3 (2022): 11-22. https://doi.org/10.1016/j.visinf.2022.07.002
[23] A. Bhasin, "Credit Card Dataset for Clustering," *Kaggle*, 2018. https://www.kaggle.com/datasets/arjunbhasin2013/ccdata (accessed Dec. 29, 2022).
[24] Smith, Jack W., James E. Everhart, W. C. Dickson, William C. Knowler, and Robert Scott Johannes. "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus." In *Proceedings of the annual symposium on computer application in medical care*, p. 261. American Medical Informatics Association, 1988.
[25] Chen, Keke, and Ling Liu. "VISTA: Validating and refining clusters via visualization." *Information Visualization* 3, no. 4 (2004): 257-270. https://doi.org/10.1057/palgrave.ivs.9500076