Pacific Graphics (2022)
Y. Yang, A. D. Parakkat, B. Deng, and S.T. Noh (Editors)

Short Paper

DARC: A Visual Analytics System for Multivariate Applicant Data Aggregation, Reasoning and Comparison

Yihan Hou, Yu Liu, He Wang, Zhichao Zhang, Yue Li, Hai-Ning Liang, and Lingyun Yu †

Xi'an Jiaotong-Liverpool University

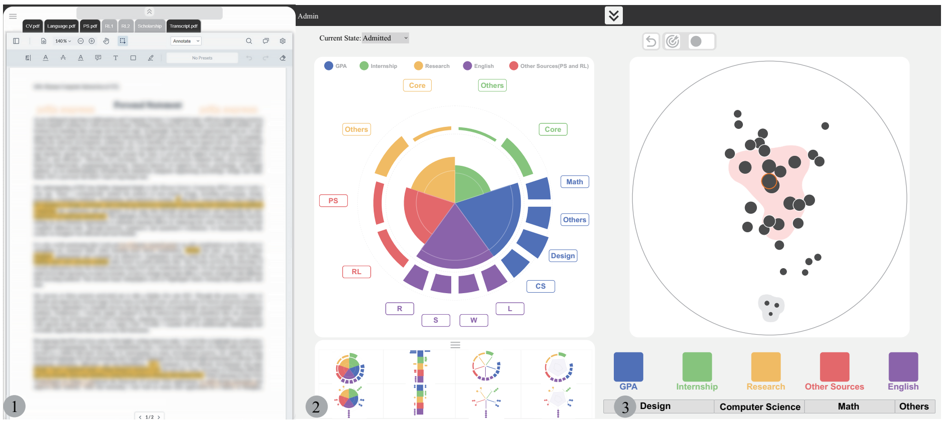

Figure 1: *DARC: an visual analytics system for multivariate data: (1) Reasoning View, (2) Analysis View, and (3) Comparison View*

Abstract

People often make decisions based on their comprehensive understanding of various materials, judgement of reasons, and comparison among choices. For instance, when hiring committees review multivariate applicant data, they need to consider and compare different aspects of the applicants' materials. However, the amount and complexity of multivariate data increase the difficulty to analyze the data, extract the most salient information, and then rapidly form opinions based on the extracted information. Thus, a fast and comprehensive understanding of multivariate data sets is a pressing need in many fields, such as business and education. In this work, we had in-depth interviews with stakeholders and characterized user requirements involved in data-driven decision making in reviewing school applications. Based on these requirements, we propose DARC, a visual analytics system for facilitating decision making on multivariate applicant data. Through the system, users are supported to gain insights of the multivariate data, picture an overview of all data cases, and retrieve original data in a quick and intuitive manner. The effectiveness of DARC is validated through observational user evaluations and interviews.

1. Introduction

Multivariate data is encountered in many fields, ranging from business to science [YYH*20, ALdK*21]. Extracting information from massive and complex data is not easy, since information is presented in different forms, such as tables, charts, images and text. These information may be hard to interpret, especially for non-experts. However, people often need to make decisions based on their comprehensive interpretation of the extracted information and overall understanding of the issue. Moreover, when making decisions about required actions, the current situation needs to be assessed, and possible solutions are required to be evaluated and compared. Thus, for many domains, it is a crucial but laborious task for decision makers to extract precise and sufficient information from multivariate data, and make decisions within strict time constraints.

For instance, recruiting students is not a trivial task. Usually, many application documents need to be reviewed, such as Curriculum Vitae, personal statements, transcripts, and English proficiency. In the recruitment, Programme Directors assess candidates' applications based on the following requirements: the previous scores, and whether they satisfy the entry requirements; the suitability of study background; the research experience; and the candidates' study motivation for the applied programme. Even with

† Corresponding author (Email: Lingyun.Yu@xjtlu.edu.cn)

DOI: 10.2312/pg.20221248

modern retrieval and analytics systems, extracting useful information from the multivariate data for decision can be time consuming.

Visualization aims to use visual representations to amplify human cognition and to help users gain insights more effectively. Multivariate data visualization is increasingly important in many working scenarios when researchers need to interpret different data features, understand distributions, patterns and structures, and extract critical information in the data. The needs of multivariate data-driven decision making motivated us to develop a visual analytics system for presenting and exploring multivariate data. However, developing such an analytics system poses three challenges:

Aggregation & Representation. Extracting and displaying salient information from multivariate and complex data is not easy. There have been many multi-dimensional visualization techniques, and glyph-based visualization is an established one for presenting multivariate data [War08]. However, designing and generating a glyph-based visualization can be challenging since it requires training and practice to create effective visualizations.

Linked view & Reasoning. When multivariate data is presented in multiple views, linked view becomes very essential, especially when decisions are made based on considerations from different aspects [Rob07]. In our work, we intend to support users to analyze applicants' data, and compare the applications from different aspects. To this end, our system should support users to specify their reviewing criteria and, importantly, provide the possibility to trace conclusions back to the corresponding evidence.

Exploration & Comparison. Data exploration is an effective way for people to find solutions from visualizations [SKBE16]. Decision making is a complicated process, especially when users need to form opinions based on multivariate data [CCH*14]. Users need to be supported to manipulate data attributes and requirements thus enabling them to compare and filter data based on multiple criteria.

In this research, we develop a visual analytics system to support people interpreting, analyzing and comparing multivariate applicant data. We focus on the student recruitment and work closely with target users to identify the design requirements. We evaluate our system through four expert evaluations and interviews.

2. Related Work

Multivariate Data Visualization. Many multivariate data visualization techniques have been proposed, including geometric projection, pixel-oriented techniques [KK96], icon-based techniques [Tuf86], and graph-based techniques. Glyph-based visualization is widely used as a form of presentation and integration of multivariate data [LPZ*21]. People can perceive the data characteristics easily since it capitalizes on the human sensitivity to graphical shapes [BKC*13]. Thus, several novel glyph designs have been presented with a noticeable impact on various analytics applications [KLT*20]. These glyph-visualizations are designed based on the data features and exploration tasks.

Multi-criteria Decision Making. Multivariate data visualizations are believed to effectively support multi-criteria decision making [KIL20]. Many visual analytics systems have been proposed to facilitate multi-criteria decision making in various field, such as SkyLens [ZWC*17] for sports and life quality, LineUp [GLG*13] for food nutrition and QS university ranking, HomeFinder [WZB*18] for real estate, and MetricsVis [ZKS*19] for employees analysis. However, most of the systems above do not support retrieving raw data, which is either numerical or non-numerical. Evaluating and comparing non-numerical data are challenging since standard quantification rule is hard to establish. For instance, it is difficult to quantifies student's study motivation and research experience mentioned in the application materials. When no standard rule can be applied in these tasks, it becomes important to support users to find the key information efficiently and make their own judgements. Tracing original documents can provide evidence to make decisions more justifiable. In summary, decision making can greatly benefit from a visual analytics system which can enables multivariate data visualization, interactive sorting and exploration. Considering the particular issues in exploring multivariate applicant data, we propose DARC, a visual analytics system for extracting information from multivariate data, comparing different data, and supporting decision making.

3. User Interview and Task Analysis

To better identify specific requirements in multivariate data analysis for the need of decision making, we focus on a specific use case — Master student recruitment. We conducted in-depth interviews with domain experts to identify specific requirements. In this section, we present the review process of the students' recruitment and our findings from the interviews. We further discuss design requirements that will guide the design of the visual analytics system.

3.1. Interview and review process

To identify specific actions performed in students' recruitment, we conducted in-depth interviews with three Programme Directors (PD) and two Graduate School staff from the local University. All of them reported extensive experience in reviewing students' applications and in selecting suitable candidates. In the interview, we discussed the current application system, the application materials, the reviewing process, and their considerations of applications.

The Graduate School staff mentioned that they usually have to review various types of original materials to find specific information, for instance, GPA, English scores (Listening, Writing, Reading, Speaking), the ranking of undergraduate university and major, etc. When reviewing applications, although they know what information they are looking for, it is still difficult to find since this information is scattered throughout various documents. Moreover, being non-domain experts, they are not able to recognize the relevant modules to the Master programmes. Therefore, they can only provide initial recommendations based on the overall GPA. However, being domain experts, PDs also felt difficult to extract useful information (e.g. working experience) from the documents. PDs also mentioned that the same modules might be named differently by two universities. In the interview, we noticed that PDs have different criteria for judging candidates' applications. All PDs mentioned that, even though the applications have been sorted out, evaluated, and recommended by Graduate School staff, they would still check the original documents to search for specific information. Thus, the whole admission practice is labor-intensive and inefficient.

3.2. Design Requirements

Based on the interviews, we carefully analyzed design requirements. We followed a nested model of visual design and validation [AKOSN14] to refine these requirements in multiple iterations.

Aggregation & Representation:

A1. Extract salient information effectively. Users need an effective way to extract information from multivariate data (e.g. tabular and text data). To this end, relevant information should be extracted from the original document. For instance, candidates' project experience in the PS should be recognised and highlighted.

A2. Explore multivariate data visually. The relevant information should be quantified and visualized effectively. An overview of aggregated information gives users an overview of the data. To enable users to interpret multiple features in one view, we need to find a proper mapping between data dimensions and visual elements. Moreover, design alternatives should be provided for users to choose the most appropriate visualization.

A3. Adjust weights interactively. The decision makers have their own criteria for the judgments. Thus, the system should support them to adjust the attribute weights in the data representation. To enable users to judge the effect of sorting rule modifications, adjustments should be immediately reflected in the visualization.

Reasoning & Linked view:

R1. Support linked view interactions. Relevant information is necessary for data analysis and decision making. Users would like to examine different attributes from different perspectives. For instance, PDs often check detailed information from various documents while comparing different applicants.

R2. Provide evidence. Decisions are often made based on complex data variables. Since multivariate data are quantified and aggregated, it is important to make the reasons for the quantification clear and evident. Thus, the system needs to provide and highlight key information that needs to be considered.

Comparison & Exploration:

C1. Provide an overview of all data. An important step in decision making process is to compare all candidates/solutions. For instance, all experts in the interviews agreed that it would be more efficient if they knew the average level of applicants. The proposed system should support the visualization of a brief overview of all data cases for in-depth analysis.

C2. Manipulate multivariate information. Users should be supported to explore and examine how multiple attributes contribute to the final ranking, since it is important to build a comprehensive understanding of the overall context.

4. Data Processing

This section introduces the multivariate data and the data processing algorithm used for extracting information. In general, the system extracts key information from tabular data and text data. Tables in the application documents are mainly stored in two formats: PDF/WORD format and image format. The tables in PDF/WORD format contain table boundaries, text and their coordinates on the page. Thus, we can identify the table structure through the line boundaries and text coordinates, and further reconstruct tables to facilitate subsequent analysis tasks. For the tables in image format, we use optical character recognition technology (PaddleOCR [YM19]) to identify the tables and extract key information (course names, grades) in the tables. The system supports text analysis, such as similarity analysis and key phrase detection, to extract essential information from text data.

Key sentence detection. TextRank [MT04] is used to grade each sentence after obtaining the word embedding of the text representation. Critical sentences are awarded higher scores and highlighted in the text. However, using a fixed word vector such as TextRank may also cause the loss of contextual information, which may potentially create misunderstandings. To aovid this, our system supports users to highlight and add new keywords manually in the text. The newly added words will be combined with the original embedded words, based on which, we re-rank sentences as follows:

$$\alpha \times \mathit{TextRank\_Score} + (1-\alpha) \times \mathit{NewKeywords\_Score}$$

α decides the weight of TextRank's score. The NewKeywords score is derived from a newly highlighted text. Currently, α is set to 0.8, representing that 80% of the score is calculated based on the words from TextRank and 20% is based on newly added words.

Similarity analysis. After the keywords are obtained, the system maps words to vectors and performs similarity analysis by measuring the distance between two word vectors. More specifically, we use the classic word embedding model GLoVe [PSM14] to combine the advantages of the LSA model [DDF*90] and the word2vec model [GL14]. Then, we use a global corpus to construct context-based word vectors, which enables the system to measure the cosine or Euclidean distance between two vectors directly without additional model. More specifically, we apply the similarity analysis to our use case as follows. Initially, the applicant's project experience is extracted from the personal statement through keyword searching. The keywords are then analyzed on the similarity based on the programme specification. After that, we can obtain a quantitative score about the applicant's project experience. Users can highlight and add new keywords, the score will be re-computed.

5. DARC: Multivariate Data Visual Analytics System

Base on the design requirements, we propose DARC, a visual analytics system for multivariate visualization (Fig. 1). Our system aims to support users in interpreting, analyzing, and comparing multivariate data in three linked views. The working process involves the following steps: 1) the salient information is extracted from multivariate applicant data. The information is encoded by a circular glyph visualization which presents applicant's information in the *Analysis view*. 2) users can select an attribute in the glyph-based visualization, and the relevant information is highlighted in the *Reasoning view*; 3) users can click on a particular applicant point in the *Comparison view*. The detail of the applicant's data will be shown in the *Analysis view*.

5.1. Analysis View

The Analysis view is motivated by the requirements for extracting and visualizing multivariate information (**A1**), and especially, supporting users to explore multivariate data (**A2**). We adopt a glyph-based visual representation, a circular glyph with two layers, to present multivariate data (Fig. 1(2)). The circular glyph contains two layers. The interior layer presents the main attributes of multivariate data, for instance, internship experience, GPA, and English grades. The exterior layer presents sub-attributes under main attributes, such as the grades of the four skills in IELTS (English language test) Colors are chosen based on the related main attributes. In general, to support users in choosing a desirable representation for applicant data **A2**, we provide four design alternatives (rose diagram, bar chart, radar chart and radial bar chart) for the interior and two design alternatives (lines and circles) for the exterior layer (Fig. 2a). These design alternatives are designed based on the literature on visualization techniques [FIBK16]. The glyphs encode the following information:

Attributes: Each attribute is represented as a radiating spoke in rose diagram, a bar in bar chart, a point in radar chart or a bar in radial bar chart. Color is used to indicate different attributes/sub-attributes. These graphical representations are the typical methods used for presenting multivariate information.

Value: Attribute value is encoded by the radius of radiating spoke in rose diagram, the length of rectangular bar in bar chart and radial bar chart, or the relative position of point in radar chart.

Weights: To meet the design requirements of weight adjustment (**A3**) and visual exploration (**A2**), the system assists users to adjust the attribute weights. For the rose diagram and the bar chart in the interior layer, users can control the weight settings by adjusting: 1) the angle of radiating spokes, or 2) the height of rectangular bars. The glyph visualization will be updated accordingly.

Eight design combinations are shown in Fig. 1. Reference lines in the interior layer are provided to indicate the average and maximum level of each attribute.

5.2. Reasoning View

The Reasoning view (Fig.1(1)) is used to trace "evidence" for the aggregated multivariate visualization from the original data (**R2**). In our use case, the Reasoning view shows all submitted application documents. The system automatically computes, extracts and highlights key information. Next to it, the Analysis view shows a circular glyph to present scores of different categories. Users can trace the "evidence" of computed scores from the Reasoning view by clicking on the corresponding visual elements in the glyph visualization. Through this, users can gain a rapid understanding of the overall quality of the applicant and the reasons behind the scores.

5.3. Comparison View

The Comparison view (Fig. 1(3)) is designed to assist users to explore and compare applications (**C1** and **C2**). In this view, users can rapidly gain an idea of all applicants data through the points distribution and further compare their details through user interactions. The view consists of three main parts, as shown in Fig. 1(3): the applicant points, the category blocks, and the feature tags area.

Inspired by Dust & Magnet [YMSJ05], we develop a "Best-Following" visualization for facilitating the ease of understanding relationships between multivariate applicant data and decision-relevant factors. A solid point represents an applicant. The point's size shows the overall score of the applicant data, and the light grey circle is used to enclose them (Fig. 1(3)). To meet the design requirements **A2**, **C1** and **C2**, users are supported to interact with the attribute blocks and the points for a better understanding of the impact of different attributes on the dataset. The distance between points and blocks encodes applicant's overall value or the value of a specific category. The overall value is the weighted sum of all attribute values. When attribute block is not activated, the point's position encodes the overall value of a applicant. The points with high values are located at the centre and lower ones are located outward. Points with the same values have the same distance from the centre. Thus, users can find the applicants with good scores in the centre of the Comparison view. Moreover, the applications can be also sorted out based on specific values. If users drag attributes blocks into the view, the selected blocks are enabled. The applicant points will start moving according to their scores of the selected attributes. Points with large values of a specific attribute move close to the attribute block. For the ease of understanding, we only enable users to enable up to two attribute blocks. The user interactions and points' movements are introduced in Sec. 5.4.

5.4. Interactions

Several interaction techniques have been provided for analyzing multivariate applicant data in the Comparing view.

Measurement. For decision-making, it is intuitive and helpful if users can use a tool to measure the degree of relevance of multiple candidate data and compare them. To this end, we employ a ruler-like tool to measure the lengths between points and the attribute block. As shown in Fig. 2b (Measurement), a straight line connects the selected points and the attribute block. The numbers showing above lines indicate the degree of relevance of the points (candidate) to the selected block (attribute). A big number indicates a big attribute value (high relevance).

Distribution. For supporting users to gain a rough judgement about the applicant data, we design two zones showing around points to indicate suitable candidates area (pink) and unsuitable candidates area (grey) (Fig. 2b (Distribution)). These two zones are used for two reasons: 1) if a new applicant point is added in the view, users can have a rough idea about it compared to other points; 2) outliers can be noticed. For instance, a point is located in the pink zone, indicating that the applicant is in the suitable area. However, if the decision is made differently, then the point is enclosed by a white border. This feature is important to student recruitment since it can catch users' attention to the outliers which may involve wrong judgment. Users can identify the candidate data based on a specific feature by clicking on the buttons provided below. For instance, users can check the students of a specific major.

Movement. To understand the degree of relevance of specific attributes, users can drag attribute blocks into the view and observe the points' movement as well as the distance between points and blocks. After an attribute block is dragged into the view, the points start moving toward the attribute block. The points stop at a place determined by the relevance value of the attribute (Fig. 2b (Movement)). For instance, when the attribute "GPA" is dragged into the view, the points (students) who have high GPA move quickly towards the attribute block and stop at a place which indicates their grades. The points with the highest GPA will be located closest to the block, while the ones with the lower grades will move away.

When the user drags the second attribute block into the view, the positions of the points will be re-computed: 1) the points with large values of two attributes will move to the attribute blocks; 2) the points which have large value of only one attribute will move to the side (close to the attribute block which has large value). Following the previous example, if the user drags the second attribute block "English" into the view, the points move according to the values of the two attributes. The points (students) who have high scores at both English and GPA move to the centre (being close to both blocks). The students who are only good at English will move

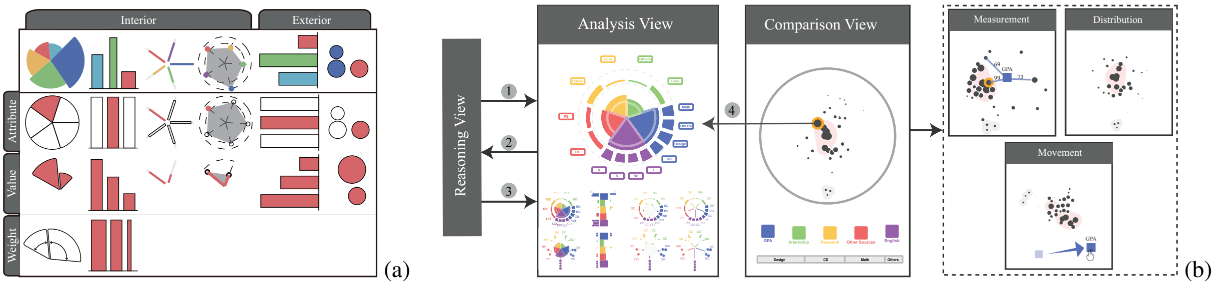

Figure 2: *(a) Design alternatives for two layers of the glyphs. (b) The system provides linked views to support users check and compare applications, including: 1. Data aggregation, 2. Reasoning, 3. Weight adjustment, 4. Finding details.*

to the side of the "English" block (being away from the "GPA" block). The moving speed is decided by the degree of relevance to the attribute blocks. The bigger the relevance is, the faster the point moves. Finally, all points stop at the estimated positions according to the degree of relevance of the considered attributes. Thus, by observing the movements, users can understand the impact of the attributes on the points. The force $F(B_j,P_i)$, which represents the movement of ith point related with jth block, is calculated as:

$$F(B_j,P_i) = W(B_j) \cdot \frac{D(B_j,P_i) - S(B_j,P_i)}{|D(B_j,P_i) - S(B_j,P_i)|} \cdot \left(\frac{R_{max} - S(B_j,P_i)}{R_{min}} + 1 \right) \quad (1)$$

$D(B_j,P_i)$ is the distance between the jth block (B_j) and the ith point (P_i), $W(B_j)$ is the weight of the attribute B_j, and R_{min} and R_{max} represent the minimum and maximum distances of the points from the block, respectively. The linear scale $S(B_j,P_i)$ of the distance between P_i and B_j can be defined as follows:

$$S(B_j,P_i) = R_{max} - \frac{V(B_j,P_i) - V_{min}(B_j)}{V_{max}(B_j) - V_{min}(B_j)}(R_{max} - R_{min}) \quad (2)$$

Here, $V(B_j,P_i)$ is the attribute value B_j of P_i, and $V_{min}(B_j)$ and $V_{max}(B_j)$ represent the minimum/maximum attribute value of B_j.

5.5. Linked Views

When decisions need to be made based on multivariate information, it is important that decisions can be traced back to reasons, for instance, the specific data. Thus, the three views in the system are linked. As shown in Fig. 2b, users can select an attribute in the glyph-based visualization in the Analysis view, the highlighted information is presented in the Reasoning view. When users select a point (candidate) in the Comparison view, the detail of the candidate data is shown in the Analysis view. Once the setting of weights is adjusted, the visualization in the Analysis view is updated.

5.6. Implementation

DARC is a browser-based application. The user interface is implemented in Javascript conjunction with D3.js and the functions of text recognition, keywords extraction and similarity analysis are developed with python. PaddleOCR [YM19] is used for table recognition, TextRank [MT04] is used for key sentence extraction and the GLoVe model [PSM14] is used for similarity analysis.

6. Evaluation

We conducted semi-structured expert interviews with four domain experts in the local University: an expert in case studies (EC), a staff member working in the Graduate Office (EG), and two Programme Directors (PD1 and PD2) who also participated in the previous interview at the design stage. PD2 is a co-author of the paper.

Procedure. The interviews started with a demonstration of the system using the student recruitment data of the previous years. After the experts got familiar with the system, they can explore the system by themselves. We encouraged them to tell their findings and ask questions if they encountered any problems. We recorded the interview and took notes of the feedback for later analysis. In the end, we asked several questions about the visual design, learnability, interface features and interaction design. The expert interview was conducted individually.

System effectiveness. The experts were particularly impressed by the integrated solution for assessing candidates' applications. EG mentioned, "I would have to open every document to search for this information in the past". They highly appreciated the glyph-based visualization for showing multivariate information, the two PDs mentioned that "the glyph visualization provides an overview of the applicants' scores (A1, A2). "Although some information might be missing, I can always check more details since the original data was provided (R2)". They also found that the weighting adjustments can fit their evaluation criteria into the system effectively (A3). The experts commented that the three linked views in the system could help to trace all aspects of the visualization back to the original information highlighted in the application materials (R1). All the experts agreed that the "Best-Following" design is an innovative approach for revealing more information (C2). An interesting point came from PD2, who recently started working as a Master Programme Director. He highly appreciated the Analysis view which provides indications about the average level from last year's recruitment. He mentioned that the visual representations and interactions of the points and attributes in the Comparison view make it easy to compare the levels of all applicants (C1).

Usability. Based on the feedback from the domain experts, they all thought DARC was easy to start. As EG mentioned, the graphical presentation would be easy for anyone who can use a computer and would be more effective with a period of proficiency with this type of graphical system. All the experts involved also expressed a strong desire to replace or supplement their current system with DARC. PD2 said he would also be keen to recommend it to other teachers in the school who need to use multivariate data.

Generalizability. The experts mentioned that DARC can be useful in various scenarios, for instance, when making decisions for house purchase or for university selections. EC suggested that the system could be used to facilitate analysis and decision making for the assessment of teachers' performance, the evaluation of declared research projects and the design and evaluation of curricula.

Preference. The experts showed a consistent preference for 8 glyphs, which preference ranking is 1) radar chart; 2) rose diagram; 3) bar chart; 4) radial bar chart. For the exterior layer, bar chart was preferred since it allows for a quick visual comparison of the data.

7. Discussion and Conclusion

DRAC makes it easy to make decisions based on multivariate data. Users can quickly access key information from multiple files, trace back to the original data, and compare similar choices. Regarding scalability, we found that aggregation, comparison, and reasoning are three fundamental and essential tasks in decision-making on multi-attribute data. We believe that the three views proposed in our system can effectively support the entire process of decision. This model can be extended to more application scenarios include multidimensional evaluation of curriculum design, assessment of research grant submitted projects, and multidimensional rating of vehicles. Such tasks often lack well-defined perfect answers and therefore require human-machine collaboration for rapid acquisition of key information, comparison and, if necessary, traceability of the original data. Referring to the limitation, text analysis is currently beyond our work's focus. In the future, more advanced text recognition and extraction techniques are required in our system.

In summary, we proposed an visual analytics system to aid users in analyzing multivariate applicant data comprehensively. We interviewed the stakeholders and identified design requirements for the system, including data aggregation and representation, reasoning and linked view, and data comparison and exploration. Based on the findings, we proposed the visual analytics system which integrates three linked views: Reasoning view, Analysis view and Comparison view. User interactions are provided to assist users to explore and compare candidates' data. We will extend our system for other multivariate data analytics scenarios.

Acknowledgements: The work was supported by XJTLU Research Development Fund RDF-19-02-11.

References

[AKOSN14] Al-Kassab J., Ouertani Z. M., Schiuma G., Neely A.: Information visualization to support management decisions. *International Journal of Information Technology & Decision Making 13*, 02 (2014), 407–428. 2

[ALdK*21] Alminagorta O., Loewen C. J., de Kerckhove D. T., Jackson D. A., Chu C.: Exploratory analysis of multivariate data: Applications of parallel coordinates in ecology. *Ecological Informatics 64* (2021), 101361. 1

[BKC*13] Borgo R., Kehrer J., Chung D. H., Maguire E., Laramee R. S., Hauser H., Ward M., Chen M.: Glyph-based visualization: Foundations, design guidelines, techniques and applications. In *Eurographics (State of the Art Reports)* (2013), pp. 39–63. 2

[CCH*14] Conati C., Carenini G., Hoque E., Steichen B., Toker D.: Evaluating the impact of user characteristics and different layouts on an interactive visualization for decision making. In *Computer Graphics Forum* (2014), vol. 33, Wiley Online Library, pp. 371–380. 2

[DDF*90] Deerwester S., Dumais S. T., Furnas G. W., Landauer T. K., Harshman R.: Indexing by latent semantic analysis. *Journal of the American society for information science 41*, 6 (1990), 391–407. 3

[FIBK16] Fuchs J., Isenberg P., Bezerianos A., Keim D.: A systematic review of experimental studies on data glyphs. *IEEE transactions on visualization and computer graphics 23*, 7 (2016), 1863–1879. 4

[GL14] Goldberg Y., Levy O.: word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *CoRR abs/1402.3722* (2014). 3

[GLG*13] Gratzl S., Lex A., Gehlenborg N., Pfister H., Streit M.: Lineup: Visual analysis of multi-attribute rankings. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2277–2286. 2

[KIL20] The role of decision makers' use of visualizations in project portfolio decision making. *International Journal of Project Management 38*, 5 (2020), 267–277. 2

[KK96] Keim D., Kriegel H.-P.: Visualization techniques for mining large databases: a comparison. *IEEE Transactions on Knowledge and Data Engineering 8*, 6 (1996), 923–938. 2

[KLT*20] Kui X., Lv H., Tang Z., Zhou H., Yang W., Li J., Guo J., Xia J.: Tvseer: A visual analytics system for television ratings. *Visual Informatics 4*, 3 (2020), 1–11. 2

[LPZ*21] Liu S., Peng D., Zhu H., Wen X., Zhang X., Zhou Z., Zhu M.: Muluba: multi-level visual analytics of user behaviors for improving online shopping advertising. *Journal of Visualization 24*, 6 (2021), 1287–1301. 2

[MT04] Mihalcea R., Tarau P.: Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (2004), pp. 404–411. 3, 5

[PSM14] Pennington J., Socher R., Manning C. D.: Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (2014), pp. 1532–1543. 3, 5

[Rob07] Roberts J. C.: State of the art: Coordinated & multiple views in exploratory visualization. In *Fifth international conference on coordinated and multiple views in exploratory visualization (CMV 2007)* (2007), IEEE, pp. 61–71. 2

[SKBE16] Saket B., Kim H., Brown E. T., Endert A.: Visualization by demonstration: An interaction paradigm for visual data exploration. *IEEE transactions on visualization and computer graphics 23*, 1 (2016), 331–340. 2

[Tuf86] Tufte E. R.: *The Visual Display of Quantitative Information*. Graphics Press, USA, 1986. 2

[War08] Ward M. O.: Multivariate data glyphs: Principles and practice. In *Handbook of data visualization*. Springer, 2008, pp. 179–198. 2

[WZB*18] Weng D., Zhu H., Bao J., Zheng Y., Wu Y.: Homefinder revisited: Finding ideal homes with reachability-centric multi-criteria decision making. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), CHI '18, p. 1–12. 2

[YM19] Yanjun Ma Dianhai Yu T. W. H. W.: Paddlepaddle: An open-source deep learning platform from industrial practice. *Frontiers of Data and Domputing 1*, 1 (2019), 105. 3, 5

[YMSJ05] Yi J. S., Melton R., Stasko J., Jacko J.: Dust & magnet: Multivariate information visualization using a magnet metaphor. *Information Visualization 4* (06 2005), 239–256. 4

[YYH*20] Yu S., Yang D., Hao Y., Lian M., Zang Y.: Visual analysis of merchandise sales trend based on online transaction log. *International journal of pattern recognition and artificial intelligence 34*, 11 (2020), 2059036. 1

[ZKS*19] Zhao J., Karimzadeh M., Snyder L. S., Surakitbanharn C., Qian Z. C., Ebert D. S.: Metricsvis: A visual analytics system for evaluating employee performance in public safety agencies. *IEEE transactions on visualization and computer graphics 26*, 1 (2019), 1193–1203. 2

[ZWC*17] Zhao X., Wu Y., Cui W., Du X., Chen Y., Wang Y., Lee D. L., Qu H.: Skylens: Visual analysis of skyline on multi-dimensional data. *IEEE transactions on visualization and computer graphics 24*, 1 (2017), 246–255. 2