Multivariate Cube for Representing Multivariable Data in Visual Analytics

Hong Thi Nguyen[1], Anh Van Thi Tran[2], Tuyet Anh Thi Nguyen[4], Luc Tan Vo[3], and Phuoc Vinh Tran[4(✉)]

[1] University of Information Technology (UIT), Vietnam National University - HCMC, Hochiminh City, Vietnam
hongnguyenl6ll@gmail.com
[2] Hochiminh City Economics College, Ho Chi Minh City, Vietnam
ttvanh26@gmail.com
[3] Eastern International University, Binhduong, Vietnam
luc.vo@eiu.edu.vn
[4] Thudaumot University (TDMU), Binhduong, Vietnam
{tuyetnta,Phuoctv}@tdmu.edu.vn, Phuoc.gis@gmail.com

Abstract. The data visualization enables users to contribute their knowledge and experience to the analysis of data stored in storages or resulted from collecting systems in real time. Visual techniques displaying data table as 2D or 3D charts, pies, lines, and so on, do not completely enable to explore multivariable data. Multivariate cube is modified from parallel coordinates by rotating the reference axis to the direction perpendicular to parallel coordinates plane. Multivariate cube represents multivariable data to enable users to answer elementary tasks in visual analytics by associating a point with its references on axes of 3-dimensional coordinates. Multivariate cube represents visually multivariable data to enable users to answer synoptic tasks in visual analytics by viewing the variation of data along the reference axis for each variable, or viewing the correlation between variables on the plane being perpendicular to and moving along the reference axis. Multivariate cube is illustrated in this paper with two case studies for visual analytics, the evaluation of learning outcomes of a program of higher education and the happenings of a disease.

Keywords: Visual analytics · Multivariate cube · Multivariable data · Data visualization

1 Introduction

Techniques of visualization display data tables as charts, pies, lines, and so on, on two-dimensional screen. Many visual softwares supply with 2D and 3D charts to display data tables as columns arranged in a cube but not enable more tools of flexible interaction to explore information or analyze data. Parallel coordinates [1–4] represent data variables on parallel axes on a plane but not share reference variables to enable to answer tasks in visual analytics. Meanwhile, visual analytics demands to show multivariable data with shared reference variables to answer analytical tasks, specially

P. Cong Vinh et al. (Eds.): ICCASA 2016, LNICST 193, pp. 91–100, 2017.
DOI: 10.1007/978-3-319-56357-2_10

synoptic questions. It also needs tools of complete and flexible interaction to discover new knowledge from data and estimate the correlation between data variables.

The main idea of this study is to apply multivariate cube which is modified from parallel coordinates for representing data sets of multi-attributes. In that, each of attribute is considered as a data variable, the key reference variable is chosen to represent on the reference axis perpendicular to the plane of parallel coordinates. The attribute axis is an axis on parallel coordinates plane and perpendicular to axes of parallel coordinates. The axes of parallel coordinates are arranged along the attribute axis, where each axis indicates the values of an attribute, termed characteristic axis. The reference axis, the attribute axis, and characteristic axes constitute a cube to represent characteristics referring to a reference variable, termed multivariate cube. The multivariate cube displaying visually characteristics as data variables enables user to answer analytical tasks including elementary and synoptic questions [5].

This paper is structured as follows. The next section presents the constitution of multivariate cube from parallel coordinates by rotating the reference axis to the direction perpendicular to parallel coordinates plane. The third section is the first case study that applies multivariate cube for representing the statistical evaluation by recruiters, alumni, last-year students, and lecturers on learning outcomes of information system program by TDMU to enable users to analyze the educational state of the program at TDMU. The fourth section is the second case study that applies multivariate cube for representing and analyzing visually the data concerning hand, foot, and mouth disease in Binhduong province during 2012–2014. The last section summarizes main contents of the paper.

2 Multivariate Cube for Representing Multivariable Data

2.1 Related Works

A data table is structured as a 2-dimensional matrix of columns and rows. Each column represents an attribute as a set of values of a variable and each row is a data tuple representing a relation between variables. Data variables shown in table may be divided into reference variables and characteristics [5]. Data visualization is studied as a conversion of a data variable of table into a visual variable as an axis or an colored axis (Fig. 1).

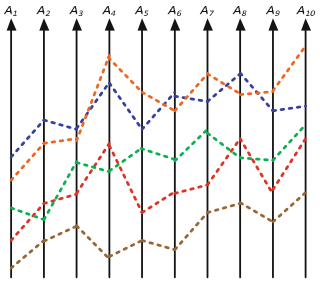

Fig. 1. The parallel coordinates.

Parallel coordinates composed of parallel axes on a plane are utilized to represent a data table, where each axis shows data of a column considered as a data variable, polylines connect the values on axes, which associate with each other on the same row of data table [1–3]. The parallel coordinates does not limit the number of variables as well as the number of rows of data table. However, parallel coordinates do not define reference axes as well as not enable completely to answer questions relating the variation of a variable corresponding to an interval of reference variable or questions concerning the correlation between variables.

Space-time cube is a cube of 3-dimensional orthogonal coordinates, where 2 axes indicate ground locations and another axis indicates time [6]. Space-time cube enables to represent and analyze visually space-time and moving objects [7–10]. However, it does not enable to represent objects of multivariable data. Therefore, some authors have tried to expand space-time cube to represent movement data including space-time location and attributes [11, 12].

2.2 Multivariate Cube Constituted from Parallel Coordinates

Multivariate cube is constituted from parallel coordinates as follows. Parallel coordinates are composed of parallel axes on a plane, each of which is utilized to represent a data variable. The axis representing the key reference variable, termed reference axis, is rotated to the direction perpendicular to the plane of parallel axes. Other parallel axes, termed characteristic axes, are arranged on the characteristic plane and along the attribute axis perpendicular to them. The orthogonal axes of reference, characteristic, and attribute form multivariate cube. In the cube, a characteristic value which associates with a point on the characteristic axis and a point on the reference axis is displayed as a bar or a point. The plane formed by the reference axis and a characteristic axis displays visually the variation of the characteristic with respect to the reference variable within displayed reference interval. The planes perpendicular to the reference axis at various reference values show the relation and correlation between characteristics at the reference values (Fig. 2).

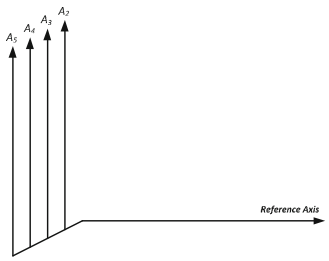

Fig. 2. Multivariate cube constituted from parallel coordinates.

3 Case Study 1: The Evaluation of Higher Education Program of Information System at Thu Dau Mot University

Thu Dau Mot University (TDMU) is a provincial university located in Thu Dau Mot city belonging Binhduong province, Vietnam, which borders on Hochiminh City in the north. The TDMU provides with diverse programs for undergraduate and graduate students. Expected as a big center providing with human resources for the provinces in the southeast of Vietnam, the TDMU not only provides students with academic knowledge but also skill and attitude (Fig. 3). The TDMU always enhances educational qualification by evaluating repeatedly student's competence with respect to the demand of society. The information system is one of programs made evaluation in academic year 2015–2016.

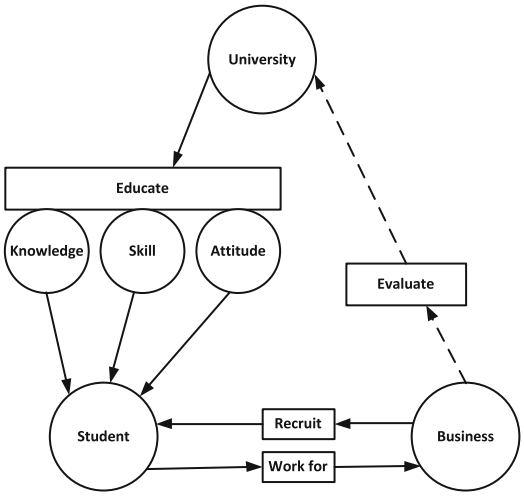

Fig. 3. The triad of higher education "University – Student – Business". The triad represents the relation between university, student, and business, where university educates students on knowledge, skill, and attitude to meet the need of recruiter and business, meanwhile business recruits students graduated from university and evaluates the response rate of university with respect to its demand (Source: "*Evaluating the competence of students graduated from information technology programs in Hochiminh City*." Master dissertation by Anh Van Thi Tran).

Based on the triad of higher education, the school of information technology belonging to the TDMU carried out a survey of the program of information system with interviewees who are business and recruiters considered as receivers the results of the program, alumni and last-year students as results of the program, and lecturers as executers of the program. Each learning outcome on knowledge, skill, and attitude is evaluated on five levels. The evaluation results are summarized in three tables, the

Table 1. The statistics on the cognition of importance of learning outcomes (Source: Survey by Tuyet Anh Thi Nguyen in a study funded by TDMU in 2016)

Learning outcome, symbol of learning outcome	Business & recruiters	Alumni	Last-year students	Lecturers
Knowledge of basic science, KBS	3	2.61	2.96	3.61
Knowledge of basic engineering, KBE	3.47	3.4	3.67	3.88
Knowledge of advanced engineering, KAE	3.4	3.46	3.91	3.95
Knowledge of supplement, KOS	3.41	3.22	3.42	3.32
Capacity of reasoning, analyzing, and solving problems, RAS	3.88	3.42	3.66	3.08
Researching and exploring knowledge, REK	3.69	3.53	3.02	3.66
Thought of systematic level, TSL	3.88	2.97	3.09	3.55
Team work, TWK	3.88	4	3.37	3.76
Communication, COM	3.68	3.49	3.56	3.08
English-based communication, EBC	4.11	3.69	4.2	2.97
Society and environment, SAE	3.23	3.1	2.4	2.74
Job context and business, JCB	3.04	3.63	3.53	2.75
Constituting idea and managing technical system, MTS	3.04	3.63	3.53	2.75
Designing system of information technology, SIT	3.4	3.3	3.5	2.54
Implementation, IMP	3.48	3.02	2.71	2.71
Operating and maintaining system of information technology, OAM	3.21	2.85	3.46	2.59
Attitude, idea, and education, AIE	4.22	3.92	3.56	4.09
Morality and justice, MAJ	3.93	3.71	3.39	3.88

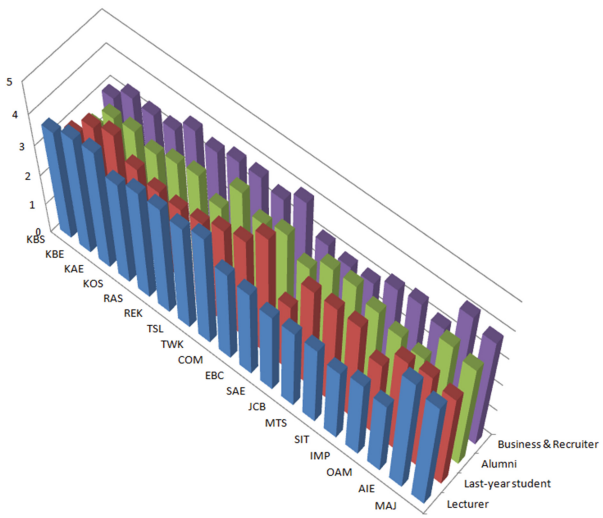

Fig. 4. The multivariate cube representing visually data of the Table 1 on the interviewees' cognition of importance of learning outcomes.

cognition of the importance of learning outcomes, the expectation of the achievement of learning outcomes from students, and the possibility of students' response of learning outcomes. The first two tables are utilized in this paper to illustrate the application of multivariate cube for visual analytics in evaluating higher education.

The surveyed data on information system program of TDMU in the academic year 2014–2015 are represented on the multivariate cubes at Figs. 4 and 5. Various tools of data visualization softwares such as show/hide, rotation, slide bar, zoom, selection, and so on, enable users to analyze the results evaluating learning outcomes by different groups of interviewees. The multivariate cubes show generally that the elementary questions relating interviewee and learning outcome can be answered easily by referring to corresponding axes. The cubes also shows on learning outcome axis the difference in evaluating different learning outcomes by an interviewee. The cubes provide users with the correlation of cognitions of the importance (Fig. 4) as well as the correlation of expectations of the achievement from students (Fig. 5) between recruiters, last-year students, alumni, and lecturers for each learning outcome.

Table 2. The statistics on the expectation of achievement of learning outcomes from students (Source: Survey by Tuyet Anh Thi Nguyen in a study funded by TDMU in 2016).

Learning outcome, symbol of learning outcome	Business & recruiter	Alumni	Last-year students	Lecturers
Knowledge of basic science, KBS	4.141	3.694	3.866	4.305
Knowledge of basic engineering, KBE	4.141	3.694	3.866	4.305
Knowledge of advanced engineering, KAE	4.313	4.019	4.081	4.583
Knowledge of supplement, KOS	4.177	3.944	3.657	4.068
Capacity of reasoning, analyzing, and solving problems, RAS	4.281	4.083	3.614	4.234
Researching and exploring knowledge, REK	4.406	3.972	3.651	4.344
Thought of systematic level, TSL	4.219	3.917	4.219	4.094
Team work, TWK	4.55	4.356	3.956	4.288
Communication, COM	4.112	3.689	4.201	2.969
English-based communication, EBC	3.875	4.333	4.485	3.412
Society and environment, SAE	4	3.714	4.04	3.17
Job context and business, JCB	4.25	3.762	4.025	3.223
Constituting idea and managing technical system, MTS	4.25	3.762	4.025	3.223
Designing system of information technology, SIT	4.063	3.444	4.091	3.177
Implementation, IMP	4.05	3.622	4.065	2.75
Operating and maintaining system of information technology, OAM	4.208	3.833	4.081	2.927
Attitude, idea, and education, AIE	4.625	4.238	3.756	4.429
Morality and justice, MAJ	4.6	4.378	3.847	4.463

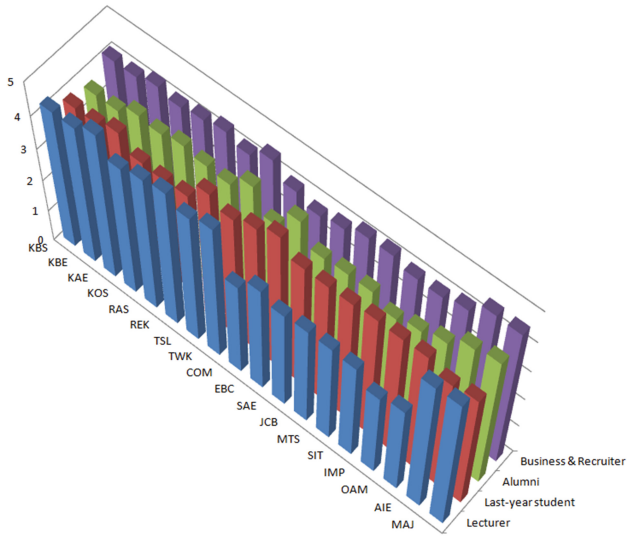

Fig. 5. The multivariate cube representing visually the data Table 2 showing the statistics on the interviewees' expectation of achievement of learning outcome from students.

The multivariate cube of Fig. 4 represents visually the data of the Table 1. The cube shows the very low correlation between the evaluations by recruiter and evaluations by lecturer on the importance of several learning outcomes as Capacity of reasoning, analyzing, and solving problems (RAS), Communication (COM), English-based communication (EBC), Society and Environment (SAE), Job context and business (JCB), Constituting idea and managing technical system (MTS), Designing system of information technology (SIT), Implementation (IMP), Operating and maintaining system of information technology (OAM), and so on.

Similarly, the multivariate cube of Fig. 5 represents the data of the Table 2. The cube displays the very low correlation between the expectations of recruiters and the expectations of lecturers on achievement from students for several learning outcomes as Capacity of reasoning, analyzing, and solving problems (RAS), Team work (TWK), Communication (COM), English-based communication (EBC), Society and Environment (SAE), Job context and business (JCB), Constituting idea and managing technical system (MTS), Designing system of information technology (SIT), Implementation (IMP), Operating and maintaining system of information technology (OAM), and so on.

The visual analysis on the Figs. 4 and 5 results in the necessity for making or adjusting the policy to improve the educational qualification of the university. Technically, an extended multivariate cube can be studied to represent and analyze the development as well as the improvement of the evaluation on information system program when the survey is carried out in several academic years.

4 Case Study 2: The Happenings of Hand, Foot, Mouth Disease in Binh Duong Province

Binhduong province borders on northern Hochiminh City, Vietnam. It is about 2,700 km^2 large with over 1.7 million habitants. In the study of the disease of hand, foot, and mouth in the whole Binhduong province, the related attributes are composed of the total of patients, the total rainfall, the average humidity, and the average temperature. The attributes change over time and are recorded every week during 2012–2014. In that, the total of patients is recorded in the whole Binhduong province and its administrative units as Thudaumot, Thuanan, Dian, Bencat, Dautieng, Phugiao, Tanuyen according to the administrative map in 2013; the total rainfall is recorded at 7 stations in the whole province; the average humidity and temperature are recorded at one station in the whole province.

The visual representation of data enables people to make visual analysis to understand the happenings as well as the outbreaking of the disease. The data variables related to the disease as the total of patients, the total rainfall, the average humidity, and the average temperature are converted to visual variables, where time variable is converted to the time axis, each data variable related to the disease is converted to an axis used as a planar variable associating with a color variable. Each disease data variable changing over time is represented on a 2D-coordinate system, where an axis represents time and another associating with a color represents the variable concerning the disease. The 2D-coordinate systems representing the variables related to the disease may be arranged on parallel planes so that they share the time axis. The combination of 2D-coordinate systems constitutes a multivariate cube representing the data variables related to the disease (Fig. 6). In the study, the intensity of color is used to show the value weight of variable over time to perceive more easily the correlation between disease data variables.

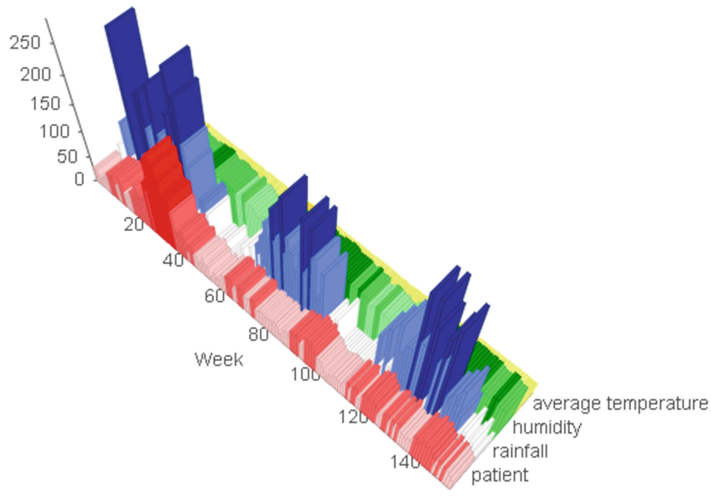

Fig. 6. The multivariate cube representing visually the data of disease of hand, foot, and mouth in Binhduong province during 2012–2014. (Data provided by Binhduong Medical Center for Providing Against Possible Contingencies)

The interactive tools of software as time selection, attribute selection, rotation, zoom, and so on provide user with different viewings to extract information for analytical tasks. Viewing along the time axis of the Fig. 6, user can cognize the developments of the disease. Viewing on the plane perpendicular to the time axis at a time point, user may percept the relation between the total of patients and rainfall, humidity, temperature. In that, the correlation between the total of patients and rainfall or humidity is high, meanwhile the correlation between the total of patients and temperature is very low. The correlation shows that hand, foot, and mouth disease happens when the rainfall and humidity increase and temperature does not affect the development of the disease. This result enable epidemiologists to build scenarios of warning the disease.

5 Conclusion

The multivariate cube is modified from parallel coordinates, where one of axes is rotated to the direction perpendicular to the plane of other axes. The multivariate cube is suitable for representing data sets of several attributes for visual analytics. This paper presents two case studies applying multivariate cube for analyzing the data of evaluating educational result of information system program at Thu Dau Mot University, and the data of hand, foot, and mouth disease provided by Binhduong Medical Center for Providing Against Possible Contingencies. Viewing data displayed on the multivariate cube for educational evaluation of information system program, TDMU's University Board can cognize that the thought of lecturers does not meet the demand of business and recruiters with respect to several learning outcomes. Viewing data shown on the multivariate cube for hand, foot, and mouth disease, epidemiologist can detect the law developing the disease.

References

1. Inselberg, A.: The plane with parallel coordinates. Visual Comput. **1**, 69–91 (1985)
2. Inselberg, A., Dimsdale, B.: Parallel coordinates: a tool for visualizing multi-dimensional geometry. In: IEEE, pp. 361–378 (1990)
3. Andrienko, G., Andrienko, A.: Parallel coordinates for exploring properties of subsets. Presented at the International Conference on Coordinated & Multiple Views in Exploratory Visualization (CMV 2004), pp. 93–104 (2004)
4. Inselberg, A.: Parallel Coordinates: Visual Multidimensional Geometry and Its Applications. Springer, New York (2008)
5. Andrienko, N., Andrienko, G.: Exploratary Analysis of Spatial and Temporal Data - A Systematic Approach. Springer, Heidelberg (2006)
6. Hagerstrand, T.: What about people in regional science? Presented at the Ninth European Congress of Regional Science Association (1970)
7. Tran, P.V., Nguyen, H.T.: Visualization cube for tracking moving object. Presented at the Computer Science and Information Technology, Information and Electronics Engineering (2011)

8. Nguyen, H.T., Duong, C.K.T., Bui, T.T., Tran, P.V.: Visualization of spatio-temporal data of bus trips. Presented at the IEEE 2012 International Conference on Control, Automation and Information Science, ICCAIS 2012, Hochiminh City, Vietnam (2012)
9. Tran, P.V., Tran, T.V., Nguyen, H.T.: Visualization-based tracking system using mobile device. In: Sobecki, J., Boonjing, V., Chittayasothorn, S. (eds.) Advanced Approaches to Intelligent Information and Database Systems. SCI, vol. 551, pp. 345–354. Springer, Cham (2014). doi:10.1007/978-3-319-05503-9_34
10. Bach, B., Dragicevic, P., Archambault, D., Hurter, C., Carpendale, S.: A review of temporal data visualizations based on space-time cube operations. In: Eurographics Conference on Visualization (EuroVis), pp. 1–19 (2014)
11. Tran, P.V., Nguyen, H.T., Tran, T.V.: Approaching multi-dimensional cube for visualization-based epidemic warning system - dengue fever. Presented at the 8th International Conference on Ubiquitous Information Management and Communication, ACM IMCOM 2014, Siem Reap, Cambodia (2014)
12. Nguyen, H.T., Tran, T.V., Tran, P.V., Dang, H.: Multivariate cube for visualization of weather data. Presented at the IEEE 2013 International Conference on Control, Automation and Information Science, ICCAIS 2013, Nha Trang, Vietnam (2013)