EconomicVis: Visual Analytics for Financial Health, Employment and Similar Life Patterns Mining - IEEE VAST Challenge 2022 MC3 Award for Strong Support for Visualization-Derived Insight

Ting Liu, Fudan University
Huiting Wang, Fudan University
Ziyue Lin, Fudan University
Yuxiao Li, Fudan University
Xuexi Wang, Fudan University
Yue Wang, Fudan University
Siming Chen*, Fudan University

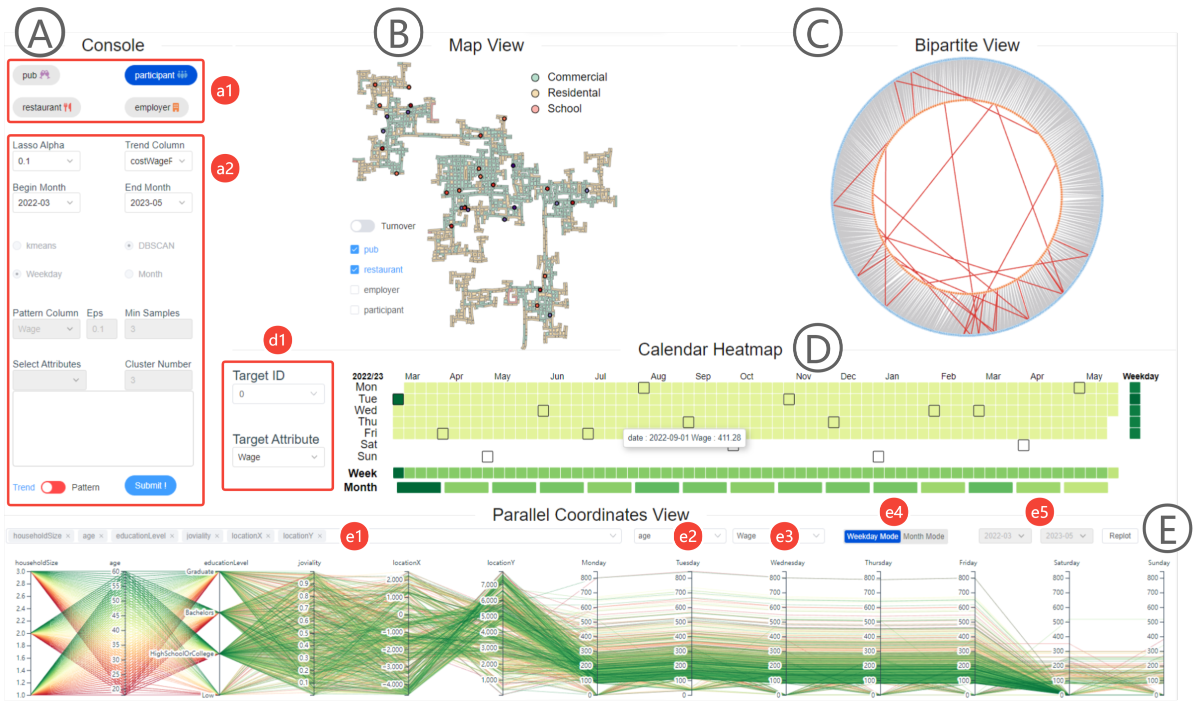

Figure 1: System interface. (A) Console. (B) Map View. (C) Bipartite View. (D) Calendar Heatmap (E) Parallel Coordinates View.

Abstract

In this paper, we design a visual analytics system to display the financial health of a city. The system contains four views, that is, map view, bipartite view, calendar heatmap, and parallel coordinates view. Combined with the console, the interactive system provides a spatial perception of economic volumes of 4 entities, i.e., participants, employers, pubs, and restaurants, display of freely combinable temporal and non-temporal attributes, and algorithm-driven mining of temporal trends and similar patterns.

1 Introduction

The VAST Challenge 2022 Mini-Challenge 3 provides 15 months of check-in, income and consumption data, and several demographic variables for 1011 volunteers in the city, requiring the participants to depict the financial health of the city. We define the keys to this task as time series visualization [1], trend analysis, and similar pattern mining [4].

Our solution focuses on the complementary nature of algorithms and visual analytics. The system consists of 4 views, the map view to show the spatial distribution of economic development, the bipartite view to show employment and separation relationships, the calendar heatmap to show time series at different granularities, and the parallel coordinates view for both high-dimensional data and time series visualization, combined with the console to support clustering.

*e-mail: simingchen@fudan.edu.cn

2 Visualization System Design

Our system is mainly implemented by Vue.js, D3.js, and Python. The system contains a console (Fig. 1.A) and four views, i.e., map view (Fig. 1.B) , bipartite view (Fig. 1.C) , calendar heatmap (Fig. 1.D) , parallel coordinates view (Fig. 1.E) .

Map View: The map is used to provide a spatial perception of the economic situation of the city. There are two modes: **default** mode and **turnover** mode. The four colors represent the four types of economic entities of the city. In the default mode, the color shades encode the economic volume of the entities. For bars and restaurants, the economic volume is the total revenue of the whole period. For participants, the economic volume is the total wages received for

15 months , and for employers, the economic volume represents the total wages paid. In the turnover mode, the size of the circle encodes the number of employees who left the company/employer.

Bipartite View: The bipartite view is used to show the employment relationship and reveal the turnover (red line) . The circles in the inner ring indicate employers and the circles in the outer ring indicate participants. The line indicates the employment relationship.

Calendar Heatmap: The calendar heatmap is used to show trends and patterns of an entity at daily granularity, weekly granularity, monthly granularity, and within a week. Users can select specific individuals and attributes (Fig. 1.d1) they want to explore. The first day of each month is highlighted with a black border to reveal the patterns of consumption on these days.

Parallel Coordinates View: The parallel coordinates view is the core view of the system. For each entity, there are both non-time series attributes such as age and education level, and time series attributes such as salary, food, etc. For non-time series attributes, parallel coordinate plots are a common tool for high-dimensional data visualization. For time series visualization, there are line graphs, density graphs, stream graphs, and other visualization forms. Compared to these tools, parallel coordinate plots have the advantage of both small space occupation and range brushing. Therefore, the parallel coordinate chart is the best tool to visualize both types of variables simultaneously. For non-time series attributes, we provide the function of dimension-free combination (Fig. 1.e1) and color-coded attributes (Fig. 1.e2) . For time series attributes, attribute switching (Fig. 1.e3) and time range filtering (Fig. 1.e5) are available, as well as two modes (Fig. 1.e4) : **weekday** mode and **month** mode. Weekday mode takes the average value for each day of a week and is used to present the cyclical nature of city life. Month mode sums up each month and is used to present the long-term trends of economic variables.

Console: The console is closely linked to other views. Entity switching (Fig. 1.a1) , time trend analysis and similar pattern clustering (Fig. 1.a2) can be performed. In trend analysis, users can choose the time interval for analysis, trend column, and parameter alpha for Lasso regression [3]. After clicking the submit button, the backend will regress each time series with natural number series and classify the results into 3 labels according to the coefficients, 1 for the upward trend, 0 for no trend, and -1 for the downward trend. The results will be updated in the parallel coordinates view. In similar pattern clustering, the user can perform k-means clustering for non-time series attributes or DBSCAN [2] clustering for time series attributes. The console allows users to select the clustering attributes and the number of clusters while using k-means clustering. The DBSCAN clustering is based on the similarity of time series, i.e., the correlation coefficient. Users can select the time series to be analyzed (pattern column) and adjust the parameters of the DBSCAN algorithm, ε (eps) and the minimum number of samples required to form a dense region (min samples).

3 Case

3.1 How Are People Changing Jobs

Through map view, bipartite view, and calendar heatmap, users can observe how people change their jobs in the city. From the map view (Fig. 2.A), it can be seen that the areas in the city where departures occur more frequently are the central (a1) and the eastern (a2) , while companies in the southern (a3) are more stable and there are no employee departures. The red line in Fig. 2.B clearly shows all separations throughout the whole period. By hovering the mouse over the outer circles, you can see all the companies where the employees have worked. In this example, the employee with id 40 has worked for companies 420, 862, and 1281. We can further verify this through the calendar heatmap (Fig. 2.C) . The three colors in the calendar heatmap suggest three different companies.

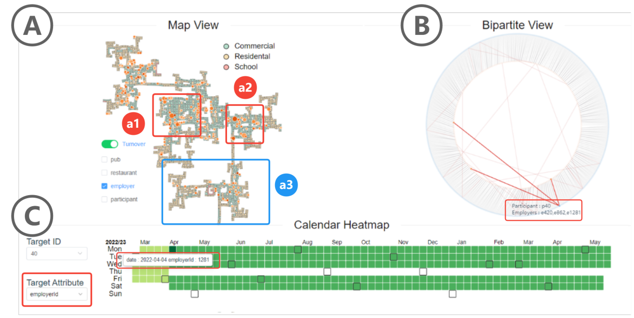

Figure 2: Three views to explore turnover in the city.

3.2 Similar Life Patterns

The system is also useful to find similar life patterns. From the calendar heatmap, we can find that the residents' life is cyclical. For example, the participant with id 840 does not work on Tuesdays and Wednesdays (Fig. 3.B1) , while it can be found that he spends more on food (Fig. 3.B2) and recreation (Fig. 3.B3) on these two days. Using DBSCAN clustering in the console (Fig. 3.A), we can see that the system clusters 21 work patterns (Fig. 3.C1) , 15 food spending patterns (Fig. 3.C2) , and 9 leisure spending patterns (Fig. 3.C3) , respectively, with the pattern column set to wage, food, and recreation.

Figure 3: Algorithm-assisted similar life pattern mining.

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. Visualization of time-oriented data. In *Human-Computer Interaction Series*, 2011.

[2] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, p. 226–231. AAAI Press, 1996.

[3] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

[4] T. Warren Liao. Clustering of time series data—a survey. *Pattern Recognition*, 38(11):1857–1874, 2005. doi: 10.1016/j.patcog.2005.01.025