

Data Mining Meets HCI: Data and Visual Analytics of Frequent Patterns

Carson K. Leung¹(✉), Christopher L. Carmichael¹, Yaroslav Hayduk^{1,2},
Fan Jiang¹, Vadim V. Kononov¹, and Adam G.M. Pazdor¹

¹ University of Manitoba, Winnipeg, MB, Canada
kleung@cs.umanitoba.ca

² Université de Neuchâtel, Neuchâtel, Switzerland

Abstract. As a popular data mining tasks, frequent pattern mining discovers implicit, previously unknown and potentially useful knowledge in the form of sets of frequently co-occurring items or events. Many existing data mining algorithms return to users with long textual lists of frequent patterns, which may not be easily comprehensible. As a picture is worth a thousand words, having a visual means for humans to interact with computers would be beneficial. This is when human-computer interaction (HCI) research meets data mining research. In particular, the popular HCI task of *data and result visualization* could help data miners to visualize the original data and to analyze the mined results (in the form of frequent patterns). In this paper, we present a few systems for data and visual analytics of frequent patterns, which integrate (i) data analytics and mining with (ii) data and result visualization.

1 Introduction and Related Works

Over the past two decades, many *frequent pattern mining* algorithms [1] have been developed for *data analytics* [3]. These algorithms usually produce long textual lists of frequent patterns, which may not be easily comprehensible. As a picture is worth a thousand words, a visual representation (i) matches the power of the human visual and cognitive system, and (ii) enables human to interact with computers effectively. This is when *human-computer interaction (HCI)* meets data mining. Specifically, HCI researches the design and usage of computer technology, with a focus on the interfaces between humans and computers. As a popular HCI tasks, *data and result visualization* could help data miners or data analysts to (i) visualize the original data and (ii) analyze the mined results (i.e., frequent patterns). This leads to *visual analytics* [2], which is the science of analytical reasoning supported by interactive visual interfaces.

Over the past two decades, several visualizers have been developed. Many of them (e.g., VisDB [6]) were designed for visualizing data only. Some were built for visualizing results of data mining tasks such as cluster analysis or anomaly detection. In the next section, we present and summarize some visualizers that have been developed for visual analytics of frequent patterns, which integrate (i) data analytics and mining with (ii) data and result visualization. Note that a

challenge of visualizing frequent patterns is the ability to show the patterns and their prefix-extension relationships (e.g., $\{a\}$ and $\{a, b\}$ are prefixes of $\{a, b, c\}$, whereas $\{a, b, c, d\}$ and $\{a, b, c, e\}$ are extensions of $\{a, b, c\}$). Another challenge is the ability to show the frequency of each pattern.

2 Frequent Pattern Visualizers

FISViz [8] visualizes frequent k -itemsets (i.e., patterns consisting of k items) as polylines connecting k nodes in a two-dimensional space with (x, y) -coordinates, in which domain items are listed on the x -axis and frequency values are indicated by the y -axis. The x -locations of all nodes in the polyline indicate the domain items contained in a frequent pattern Z , and the y -location of the rightmost node of a polyline for Z indicates the frequency of Z . Hence, prefix-extension relationships can be observed by traversing along the polylines. See Fig. 1(a). In addition, to facilitate exploration of data and mining results, FISViz also provides users with interactive detail-on-demand features. When the mouse hovers on a polyline connecting two nodes u and v , FISViz shows a list of itemsets containing both u and v . Similarly, when the mouse hovers over a node, FISViz shows a list of all patterns contained in all polylines starting or ending at this node.

As polylines in FISViz can be bent and crossed over each other, it may not be easy to distinguish one polyline from another. To solve this problem, **WiFISViz** [9] and **FpVAT** [7] were designed. As shown in Fig. 1(b), **WiFISViz** uses two half-screens to visualize frequent patterns. Both half-screens are wiring-type diagrams (i.e., orthogonal graphs), which represent frequent patterns as horizontal lines connecting k nodes in a two-dimensional space (where the x -axis lists all the domain items). The left half-screen provides the frequency information by using the y -location of the horizontal line to indicate the frequency of the frequent pattern. The right half-screen lists all frequent patterns in the form of a trie.

FpVAT [7] also uses wiring-type diagrams to visualize frequent patterns. However, FpVAT shows all the frequent patterns and their frequencies on the same full-screen. See Fig. 1(c).

The above three visualizers show *all* frequent patterns. When handling very large datasets, the number of frequent patterns to be displayed can be huge due to pattern explosion. To improve this situation, **CloseViz** [5] extends WiFISViz and FpVAT by providing users with explicit and easily-visible information among the *closed* patterns, which greatly reduces the number of displayed patterns without losing any frequency information. Note that a frequent pattern Z is *closed* if no superset of Z has the same frequency as Z . As shown in Fig. 1(d), CloseViz represents closed patterns as horizontal lines in a two-dimensional graph.

The above four visualizers show frequent patterns from a single database instance. However, there are situations in which users may be interested in differences between the results returned from two database instances. For example, a store manager may be interested in finding out the difference between popular sets of merchandise items sold in the summer and in the winter in order to detect the (temporal) changes in frequencies of the mined frequent patterns as

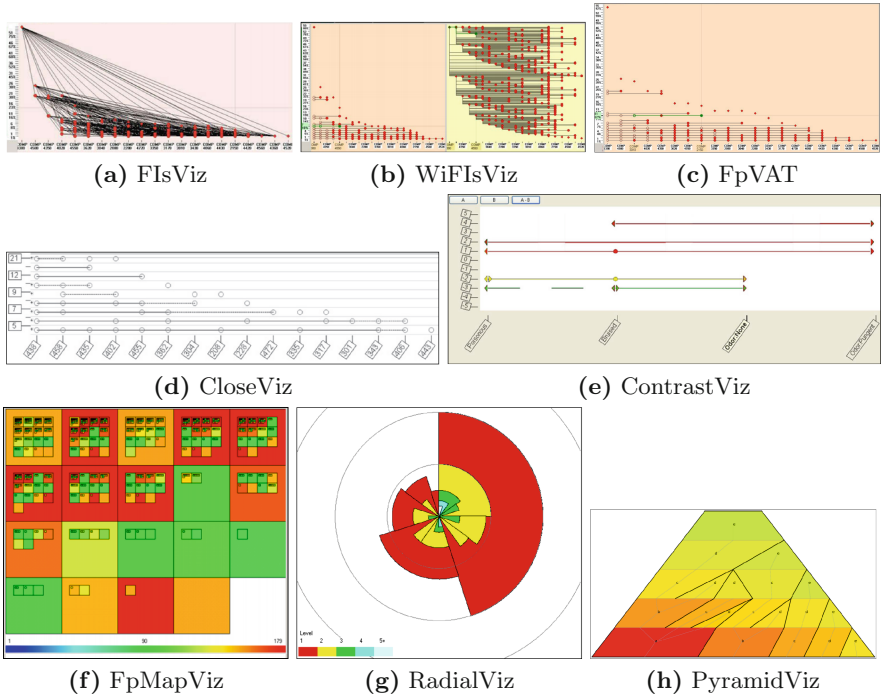


Fig. 1. Frequent pattern visualizers

well as their trends from one database instance to another. Similarly, a regional manager may want to find out the (spatial) difference between the popular sets of merchandise items sold in two different locations. To handle with these real-life situations, **ContrastViz** [4] extends WiFIsViz and FpVAT by helping users to visually contrast two collections of frequent patterns. As shown in Fig. 1(e), ContrastViz visualizes and analyzes all the frequent patterns, their frequencies, as well as changes in frequencies.

Instead of polylines or wiring-type diagrams (i.e., orthogonal graphs), FpMapViz [11], RadialViz [10] and PyramidViz [12] use alternative design with emphasis on showing the prefix-extension relationships among the frequent patterns. For example, inspired by the tree map representation of hierarchical information, **FpMapViz** represents frequent patterns as squares in a hierarchical fashion so that extensions of a frequent pattern Z are embedded within squares representing the prefixes of Z . The colour of the square representing Z indicates the frequency range of Z . See Fig. 1(f).

As shown in Fig. 1(g), **RadialViz** [10] also visualizes frequent patterns but in a radial layout, which leads to a benefit of being *orientation-free*. As such, the legibility of the represented frequent patterns is not be impacted by the orientation. Hence, RadialViz is ideal for the collaborative environment (cf. traditional two-dimensional rectangular space, which favors the viewer who visualizes data

or mining results at the up-right position but not favors those on the opposite side or the left/right sides). Moreover, RadialViz also represents frequent patterns in a hierarchical fashion so that extensions of a frequent pattern Z are embedded within sectors representing the prefixes of Z . The frequency of Z is represented by the radius of the sector representing Z .

Recently, **PyramidViz** [12] visualizes frequent patterns in a tree or building block layout. As shown in Fig. 1(h), the frequent 1-itemsets are located at the bottom of the pyramid, whereas frequent patterns of higher cardinalities are located near the top of the pyramid. Moreover, frequent patterns are represented in a hierarchical fashion so that the building blocks representing the extensions of a frequent pattern Z are put on top of the blocks representing the prefixes of Z . The colour of the block representing Z indicates the frequency range of Z .

3 Conclusions

This paper presents instances when data mining meets HCI, with focus on data and visual analytics of frequent patterns by describing eight frequent pattern visualizers: FIsViz, WiFIsViz, FpVAT, CloseViz, ContrastViz, FpMapViz, RadialViz, and PyramidViz. As ongoing work in the current era of big data, we are extending existing visualizers to support big data visualization. We are also broadening our study by including alternative frequent pattern visualizers.

Acknowledgement. This project is partially supported by NSERC (Canada).

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: VLDB 1994, pp. 487–499 (1994)
2. Andrienko, N.V., Andrienko, G.L., Fuchs, G., Jankowski, P.: Visual analytics methodology for scalable and privacy-respectful discovery of place semantics from episodic mobility data. In: Bifet, A., May, M., Zadrozny, B., Gavaldà, R., Pedreschi, D., Bonchi, F., Cardoso, J., Spiliopoulou, M. (eds.) ECML PKDD 2015 Part III. LNCS, vol. 9286, pp. 254–258. Springer, Heidelberg (2015)
3. Börner, M., Rhode, W., Ruhe, T., Morik, K.: Discovering neutrinos through data analytics. In: Bifet, A., May, M., Zadrozny, B., Gavaldà, R., Pedreschi, D., Bonchi, F., Cardoso, J., Spiliopoulou, M. (eds.) ECML PKDD 2015. LNCS, vol. 9286, pp. 208–212. Springer, Heidelberg (2015)
4. Carmichael, C.L., Hayduk, Y., Leung, C.K.: Visually contrast two collections of frequent patterns. In: IEEE ICDM Workshops 2011, pp. 1128–1135 (2011)
5. Carmichael, C.L., Leung, C.K.: CloseViz: visualizing useful patterns. In: ACM SIGKDD Workshop on UP 2010, pp. 17–26 (2010)
6. Keim, D.A., Kriegel, H.-P.: Visualization techniques for mining large databases: a comparison. IEEE TKDE **8**(6), 923–938 (1996)
7. Leung, C.K., Carmichael, C.L.: FpVAT: a visual analytic tool for supporting frequent pattern mining. ACM SIGKDD Explor. **11**(2), 39–48 (2009)

8. Leung, C.K., Irani, P.P., Carmichael, C.L.: FIsViz: a frequent itemset visualizer. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 644–652. Springer, Heidelberg (2008)
9. Leung, C.K., Irani, P.P., Carmichael, C.L.: WiFIsViz: effective visualization of frequent itemsets. In: IEEE ICDM, pp. 875–880 (2008)
10. Leung, C.K., Jiang, F.: RadialViz: an orientation-free frequent pattern visualizer. In: Tan, P.-N., Chawla, S., Ho, C.K., Bailey, J. (eds.) PAKDD 2012, Part II. LNCS, vol. 7302, pp. 322–334. Springer, Heidelberg (2012)
11. Leung, C.K., Jiang, F., Irani, P.P.: FpMapViz: a space-filling visualization for frequent patterns. In: IEEE ICDM Workshops, pp. 804–811 (2011)
12. Leung, C.K., Kononov, V.V., Pazdor, A.G.M.: PyramidViz: visual analytics and big data visualization of frequent patterns. In: IEEE DASC/DataCom 2016 (2016)