

Visualización comparativa: un enfoque desde la visualización de información para la comparación de datos

Dana K. Urribarri^{1,2,3}, Martín Larrea^{1,2,3}, Leandro Luque^{1,3}, Matías Selzer^{1,2,3}

¹Laboratorio de I+D en Visualización y Computación Gráfica (VyGLab) (UNS-CIC PBA)

²Dpto. de Ciencias e Ingeniería de la Computación, Universidad Nacional del Sur (DCIC-UNS)

³Instituto de Ciencias e Ingeniería de la Computación (UNS-CONICET)

{dku, mll, leandro.luque, matias.selzer}@cs.uns.edu.ar

RESUMEN

La comparación es una tarea importante en el análisis de datos, en la cual dos o más conjuntos de datos deben analizarse buscando similitudes o diferencias. La visualización cumple un rol importante en el proceso de entender, analizar y comunicar datos. La comparación implica un desafío para la visualización, dado que se debe combinar tanto factores de los objetos individuales como de la relación entre ellos. Mientras que la visualización de información tradicionalmente se ha centrado en examinar objetos o conjuntos de datos individuales, muchos trabajos han resaltado la importancia de desarrollar herramientas de visualización que soporten explícitamente la tarea de comparación.

La finalidad de esta línea de investigación es proponer y desarrollar técnicas de visualización que faciliten la comparación de grandes conjuntos de datos y/o la comparación de múltiples conjuntos de datos, teniendo en cuenta: (1) el compromiso entre la complejidad del diseño y la eficiencia en la realización de la tarea; (2) la diversidad de tareas que pueden considerarse como comparación de datos. La simpleza o complejidad de la tarea de comparación dependerá de la naturaleza de los datos y de la relación existente entre ellos.

Palabras clave: Visualización de información; Visualización comparativa; Grandes conjuntos de datos

CONTEXTO

Esta línea de investigación continúa la presentada en el proyecto PGI 24/ZN39 financiado por la Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Argentina, y finalizado el año pasado; actualmente, el proyecto en el cual se inserta esta investigación se encuentra en evaluación por la misma secretaría.

El trabajo se lleva a cabo en el Laboratorio de Investigación y Desarrollo en Visualización y Computación Gráfica (VyGLab) perteneciente al Departamento de Ciencias e Ingeniería de la Computación, de la Universidad Nacional del Sur y centro asociado CIC-PBA. Los trabajos realizados bajo esta línea involucran a docentes investigadores, becarios doctorales y alumnos de grado.

1. INTRODUCCIÓN

La sociedad está atravesando la *era de los datos*: no solo se generan constantemente enormes cantidades de datos, sino que además han crecido las expectativas sobre su análisis [28]. Con una cantidad cada vez mayor de recursos informáticos, de dispositivos de detección disponibles y de conectividad, la

capacidad para recopilar y generar muchos conjuntos de datos complejos sigue creciendo. La digitalización y el contenido digital se produce en tres lugares: el núcleo (los centros de datos tradicionales y en la nube), el borde (la infraestructura como torres de telefonía móvil) y los dispositivos finales (PC, teléfonos inteligentes, y dispositivos IoT). La suma de todos estos datos, ya sean creados, capturados o replicados, está experimentando un enorme crecimiento [25]. Con el auge de IoT y las redes de sensores como herramienta de monitoreo, a los datos digitales más tradicionales se le suman los datos de una amplia gama de dominios, como atención médica, energía, transporte, ciudades inteligentes, automatización de edificios, agricultura, o industria 4.0. Más aún, estos son conjuntos de datos diversos (texto, imagen, video, números), voluminosos (gran cantidad de ítems), altamente dimensionales (cada ítem está descrito por un gran número de dimensiones) y, en muchos dominios, provenientes de un flujo de datos. En este escenario, las tecnologías de ciencias de datos y, en particular, las herramientas de análisis visual juegan un papel importante en la generación de conocimiento valioso para apoyar la toma de decisiones críticas [24].

La *comparación* es una tarea importante en el análisis de datos, en la cual dos o más conjuntos de datos deben analizarse buscando similitudes o diferencias; aparece a lo largo de áreas de aplicación tan diversas como epidemiología [14], oceanografía [32], biología [18], meteorología [4, 16], ciencias forestales [19], etc. y, a su vez, involucra diversos tipos de datos, tales como imágenes [9], grafos [1, 20], superficies [12], series de tiempo [6], etc. El mayor desafío de la comparación es que no sólo debe combinar factores de los objetos individuales sino también de la relación entre ellos.

La *visualización* cumple un rol importante en el proceso de entender, analizar y comunicar datos. Como disciplina, busca desarrollar métodos, herramientas y representaciones visuales que faciliten el análisis de datos y permitan extraer conclusiones de manera rápida y eficiente sobre las posibles relaciones existentes en los datos. Esto la hace eficaz cuando la gran cantidad de ítems y/o su dimensionalidad dificulta el análisis y entendimiento de los datos crudos [3]. Mientras que la visualización de información tradicionalmente se ha centrado en examinar objetos o conjuntos de datos individuales, muchos trabajos han resaltado la importancia de desarrollar herramientas de visualización que soporten explícitamente la tarea de comparación [8, 7, 10, 17, 21, 22, 23, 27, 26].

El *diseño* de visualizaciones en general y de visualizaciones comparativas en particular no está exento de desafíos. La naturaleza de los datos (datos multidimensionales, datos temporales, datos espaciales, categóricos, etc.) y la tarea de comparación que se desea realizar son determinantes a la hora de diseñar una visualización de datos. Para lograr un diseño de calidad hay que tener en cuenta varios aspectos, por ejemplo, la escalabilidad visual de la técnica [5], la incertidumbre de los datos [2, 11], la percepción humana [10, 22], los requisitos de la especificación y el correcto funcionamiento de la visualización y sus interacciones [13, 33, 15].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo de esta investigación consiste en estudiar y desarrollar técnicas de visualización, representaciones visuales y herramientas que asistan en la comparación de conjuntos de datos y se centra en dos ejes principales:

- (1) La visualización comparativa de series de tiempo con características de los datos que requieran de un enfoque especial (por ejemplo, datos con atributos categóricos o multidimensionales),
- (2) Visualizaciones comparativas que permitan la comparación de más de dos conjuntos de datos.

3. RESULTADOS OBTENIDOS y ESPERADOS

Sobre los ejes presentados se han obtenido resultados parciales. Se ha trabajado en una técnica de visualización comparativa de series de tiempo. Esta técnica se ha empleado tanto en la comparación de datos provenientes de movimientos de karate [26,27] como en la comparación de datos meteorológicos [28,29]. En la continuación del proyecto se espera avanzar con la implementación de una versión genérica de la técnica que nos permita explorar la viabilidad de esta técnica en la comparación de series de tiempo con datos de otros tipos, por ejemplo, categóricos.

4. FORMACIÓN DE RECURSOS HUMANOS

En lo que concierne a la formación de recursos humanos, se incentiva la incorporación de alumnos de grado o de posgrado que deseen trabajar en temas relacionados.

5. BIBLIOGRAFÍA

- [1] K. Andrews, M. Wohlfahrt, and G. Wurzinger. Visual Graph Comparison. In 2009 13th International Conference Information Visualisation, volume 00, pages 62–67, jul 2009.
- [2] G. P. Bonneau, H. C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and state-of-the-art of uncertainty visualization, 2014.
- [3] W. S. Cleveland. Visualizing Data. Hobart Press, 1993.
- [4] A. Dasgupta, J. Poco, Y. Wei, R. Cook, E. Bertini, and C. T. Silva. Bridging theory with practice: An exploratory study of visualization use and design for climate model comparison. *IEEE Transactions on Visualization and Computer Graphics*, 21(9):996–1014, sep 2015.
- [5] S. G. Eick and A. F. Karr. Visual scalability. *Journal of Computational and Graphical Statistics*, 11(1):22–43, mar 2002.
- [6] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg. Evaluation of alternative glyph designs for time series data in a small multiple setting. *ACM*, apr 2013.
- [7] M. Gleicher. Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):413–423, 2018.
- [8] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts. Visual comparison for information visualization. *Information Visualization*, 10(4):289–309, 2011.
- [9] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. High-Resolution Global Maps of 21st-

- Century Forest Cover Change. *Science*, 342(6160):850–853, nov 2013.
- [10] N. Jardine, B. D. Ondov, N. Elmqvist, and S. Franconeri. The Perceptual Proxies of Visual Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):1012–1021, jan 2020.
- [11] A. Kamal, P. Dhakal, A. Y. Javaid, V. K. Devabhaktuni, D. Kaur, J. Zaiantz, and R. Marinier. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization*, may 2021.
- [12] K. Kim, J. V. Carlis, and D. F. Keefe. Comparison techniques utilized in spatial 3d and 4d data visualizations: A survey and future directions. *Computers & Graphics*, 67:138–147, oct 2017.
- [13] R. Kirby and C. Silva. The need for verifiable visualization. *IEEE Computer Graphics and Applications*, 28(5):78–83, sep 2008.
- [14] T. Kreuz. Comparative visualization of epidemiological data during various stages of a pandemic. Feb. 2021.
- [15] M. L. Larrea. Black-box testing technique for information visualization. Sequencing constraints with low-level interactions. *Journal of Computer Science and Technology*, 17(01):p. 37–48, Apr. 2017.
- [16] M. L. Larrea and D. K. Urribarri. Visualization Technique for Comparison of Time-based Large Data Sets. In *Conference on Cloud Computing Conference, Big Data & Emerging Topics (JCC-BD&ET)*, volume 1444 of *Communications in Computer and Information Science*, pages 179–187. Springer International Publishing, 8 2021.
- [17] S. LYi, J. Jo, and J. Seo. Comparative layouts revisited: Design space, guidelines, and future directions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1525–1535, feb 2021.
- [18] S. Nagpal, K. D. Baksi, B. K. Kuntal, and S. S. Mande. NetConfer: A web application for comparative analysis of multiple biological networks. *BMC Biology*, 18(1), 2020.
- [19] J. W. Nam, K. McCullough, J. Tveite, M. M. Espinosa, C. H. Perry, B. T. Wilson, and D. F. Keefe. Worlds-in-Wedges: Combining Worlds-in-Miniature and Portals to Support Comparative Immersive Visualization of Forestry Data. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 747–755, mar 2019.
- [20] Y. Naragino and K. Misue. Visualization Techniques for the Comparative Analysis of Weighted Free Trees. In *2017 21st International Conference Information Visualisation (IV)*, pages 45–51, jul 2017.
- [21] B. Ondov, N. Jardine, N. Elmqvist, and S. Franconeri. Face to Face: Evaluating Visual Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):861–871, jan 2019.
- [22] B. D. Ondov. Revealing Perceptual Proxies in Comparative Data Visualization. PhD thesis, 2021.
- [23] H.-G. Pagendarm and F. H. Post. Comparative Visualization: Approaches and Examples. *Visualization in Scientific Computing* (H. Göbel, H.

- Müller and B. Urban), (February):95–108, 1995.
- [24] A. Protopsaltis, P. Sarigiannidis, D. Margounakis, and A. Lytos. Data visualization in internet of things: Tools, methodologies, and challenges. In Proceedings of the 15th International Conference on Availability, Reliability and Security, ARES '20, New York, NY, USA, 2020. Association for Computing Machinery.
- [25] D. Reinsel, J. Gantz, and J. Rydning. Data Age 2025, 2018. Accedido 2-noviembre, 2023.
- [26] J. Schmidt. Scalable Comparative Visualization. Visual Analysis of Local Features in Different Dataset Ensembles. PhD thesis, Faculty of Informatics (TU Wien), 2016.
- [27] A. Srinivasan, M. Brehmer, B. Lee, and S. M. Drucker. What's the Difference?: Evaluating Variants of Multi-Series Bar Charts for Visual Comparison Tasks. In ACM Conference on Human Factors in Computing Systems (CHI), pages 1–10. ACM, Jan. 2018.
- [28] J. J. Thomas and K. A. Cook. Illuminating the Path. IEEE, 2005.
- [29] D. K. Urribarri and M. L. Larrea. A visualization technique to assist in the comparison of large meteorological datasets. Computers & Graphics, mar 2022.
- [30] D. K. Urribarri, M. L. Larrea, S. M. Castro, and E. Puppo. Visualization to compare karate motion captures. In P. Pesado and M. Arroyo, editors, XXV Congreso Argentino de Ciencias de la Computación. CACIC 2019. Libro de Actas., pages 446–455, Oct. 2019.
- [31] D. K. Urribarri, M. L. Larrea, S. M. Castro, and E. Puppo. Overview+detail visual comparison of karate motion captures. In P. Pesado and M. Arroyo, editors, Computer Science – CACIC 2019, volume 1184 of Communications in Computer and Information Science, pages 139–154. Springer International Publishing, Cham, 5 2020.
- [32] M. Yano, T. Itoh, Y. Tanaka, D. Matsuoka, and F. Araki. A comparative visualization tool for ocean data analysis based on mode water regions. Journal of Visualization, 23(2):313–329, 2020.
- [33] X. Yuan, X. Zhang, and A. Trofimovsky. Testing visualization on the use of information systems. In Proceedings of the third symposium on Information interaction in context. ACM, aug 2010.