

VISUALIZACIÓN COMPARATIVA: ESTUDIO Y DISEÑO DE VISUALIZACIONES ORIENTADAS A LA COMPARACIÓN DE CONJUNTO DE DATOS

Dana K. Urribarri, Martín Larrea, M. Luján Ganuza, Silvia M. Castro

Laboratorio de I+D en Visualización y Computación Gráfica

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

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

Laboratorio de I+D en Visualización y Computación Gráfica,

(UNS-CIC Prov. de Buenos Aires)

{dku, mll, mlg, smc}@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, que cumple un rol importante en el proceso de entender, analizar y comunicar datos en general, también puede contribuir en la comparación de datos en particular. Estas comparaciones implican 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 sistemas han demostrado la importancia de desarrollar herramientas de visualización que soporten explícitamente la tarea de comparación. Si visualizar un único objeto o conjunto de datos complejo ya es una tarea desafiante, realizar una comparación suma desafíos nuevos.

La finalidad de este proyecto es proponer y desarrollar técnicas de visualización que faciliten el proceso de comparación de grandes conjuntos de datos y/o el proceso de comparación de múltiples conjuntos de datos.

Palabras Clave: *Visualización de información; Visualización comparativa; Grandes conjuntos de datos.*

CONTEXTO

Esta línea de investigación está inserta en el proyecto PGI 24/ZN39 financiado por la Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Argentina.

El proyecto 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 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 [1]. Aparece a lo largo de áreas de aplicación tan diversas como, por ejemplo, epidemiología [2], la oceanografía [3], la biología [4], la meteorología [5,6], y la ingeniería forestal [7] entre muchas otras, e involucra una amplia gama de tipos de datos, incluidos gráficos, tablas, superficies, series de tiempo, etc.

La visualización, que cumple un rol importante en el proceso de entender, analizar y

comunicar datos en general. En particular, la visualización de información puede dar soporte a la comparación de datos sin abordarla explícitamente. Sin embargo, las técnicas diseñadas explícitamente para facilitar la comparación son las que brindan mejor soporte al usuario [8].

La visualización comparativa plantea visualizar datos de dos o más fuentes diferentes para evidenciar similitudes y diferencias [9]. Pueden distinguirse tres tipos de objetivos en la comparación: comprender las similitudes y diferencias entre varios conjuntos de datos, comparar diferentes características de un conjunto de datos particular y comparar diferentes tipos de visualizaciones vinculadas. Además, pueden distinguirse tres grandes categorías de diseño de las visualizaciones comparativas [10]: la yuxtaposición que presenta los objetos a comparar por separado pero cercanos entre sí (en tiempo o en espacio), la superposición que presenta múltiples objetos en el mismo sistema de coordenadas (uno arriba del otro) y la representación explícita que codifica directamente las conexiones entre los objetos visuales. Finalmente, estas tres categorías pueden combinarse para formar categorías híbridas. Se han estudiado ventajas y desventajas de estas estrategias y se han esbozado guías de diseño [11].

Dentro de la visualización de datos se utilizan varias técnicas para la comparación, incluso las más tradicionales y simples [12]. La visualización *side-by-side* (una yuxtaposición espacial) es una de las estrategias más usadas. Si bien esta técnica es adecuada para la comparación, genera una alta carga cognitiva en el usuario y no escala bien ni con el tamaño ni con la cantidad de conjuntos de datos a comparar. El uso de esta estrategia para visualizar varios conjuntos de datos se denomina *small multiples* [13]. La superposición, a su vez, presenta problemas como la oclusión y la distorsión. Dentro de esta clasificación, los diagramas de espagueti son

muy utilizados aunque estos no escalan bien ni con la dimensionalidad ni con la cantidad de conjuntos a comparar. En este último caso, la acumulación de líneas en el área de visualización genera problemas de oclusión e interferencia en la información. Las visualizaciones comparativas 3D, considerando todas sus ventajas y desventajas, también son ampliamente utilizadas y necesarias [14].

El diseño de visualizaciones comparativas comparte los mismos desafíos que la visualización en general, añadidos aquellos inherentes a la propia tarea. Dado que la tarea de comparación puede cambiar drásticamente los objetivos de la visualización, las estrategias aprendidas de la visualización en general no siempre son apropiadas para este caso; por ejemplo, se ha estudiado el impacto de la tarea de comparación en la percepción [15,16,17]. Un buen diseño debe considerar el compromiso entre la complejidad del diseño y la eficiencia en la realización de la tarea [10], sin dejar de ser expresivo (i.e. reflejar los datos de manera óptima), efectivo (i.e. soportar la tarea de manera óptima) y apropiado (i.e. adaptarse al tipo de dato a visualizar). Los desafíos de diseño que la visualización comparativa comparte con la visualización en general son la escalabilidad visual, la incertidumbre de los datos y la percepción humana. Por otro lado, es importante contar con herramientas para medir y evaluar la calidad de las visualizaciones.

La escalabilidad visual [18] es la capacidad de las técnicas de visualización para mostrar eficazmente grandes conjuntos de datos, en términos de número o dimensión de los elementos individuales de estos. Y, dado que los problemas de comparación escalan no solo con la complejidad de los conjuntos de datos (cantidad de elemento y dimensionalidad) sino también con el número de conjuntos a comparar, la escalabilidad representa un desafío importante en la visualización comparativa [19]. Tres factores pueden afectar

a la escalabilidad visual: la cantidad de datos o su complejidad, la cantidad de conjuntos de datos a comparar, y la cantidad o la complejidad de las relaciones existentes entre los conjuntos.

La mayoría de los trabajos se centran en comparaciones 1 a 1, por lo que en estos casos la cantidad de datos o la complejidad de los datos de cada conjunto es el factor determinante en la escalabilidad visual. Por otro lado, tanto realizar comparaciones de varios conjuntos de datos a partir de estrategias 1 a 1 como encontrar representaciones visuales adecuadas para hacerlo no es sencillo, independientemente de la cantidad de datos en los conjuntos.

Otro gran desafío en visualización es la incertidumbre de los datos [20,21]. A pesar de la incertidumbre asociada a casi todos los conjuntos de datos, la visualización de información a menudo ha asumido que la información es en su mayoría precisa. Sin embargo, ignorar la incertidumbre asociada a los datos puede ser engañoso para los analistas y llevar a decisiones erróneas. La incertidumbre es un concepto complicado que tiene múltiples facetas y en la literatura existen múltiples definiciones. En general, hace referencia a que la información puede ser incompleta, inconsistente o errónea y, por lo tanto, el resultado final no es fiable; además, también pueden influir la falta de claridad, precisión o conocimiento de los usuarios. En el contexto de la visualización comparativa, la incertidumbre de los datos puede influenciar el proceso de comparación.

La medición es una parte integral de la ciencia moderna, ya que proporciona los medios fundamentales para la evaluación, la comparación y la predicción [22]. Las medidas pueden ser cuantitativas o cualitativas, y uno de sus principales objetivos es proporcionar medios objetivos para razonar sobre las visualizaciones y su eficacia [23,24,25].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación se centran en proponer y desarrollar técnicas de visualización que faciliten el proceso de comparación de grandes conjuntos de datos y/o el proceso de comparación de múltiples conjuntos de datos. En particular, se continuará trabajando en diseño de visualizaciones para la comparación de datos temporales multidimensionales [26,27,28,29]. Paralelamente, para las técnicas de visualización comparativa, se trabaja en el diseño de medidas objetivas para de evaluación de las técnicas, en el estudio de diferentes estrategias de visualización comparativa para diferentes tipos de datos, en el impacto de la cantidad de datos o la cantidad de conjuntos de datos a comparar en la escalabilidad visual de las técnicas y en la capacidad de las técnicas para representar la incertidumbre en los datos y en la comparación.

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 y basada en Dynamic Time Warping. 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 contar con la implementación de una versión que 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] Schulz, H., Nocke, T., Heitzler, M., & Schumann, H. (2013). A Design Space of Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2366–2375. [10.1109/TVCG.2013.120](https://doi.org/10.1109/TVCG.2013.120)
- [2] Kreuz, T. (2021). *Comparative visualization of epidemiological data during various stages of a pandemic*. <https://arxiv.org/abs/2102.11315>
- [3] Yano, M., Itoh, T., Tanaka, Y., Matsuoka, D., & Araki, F. (2020). A comparative visualization tool for ocean data analysis based on mode water regions. *Journal of Visualization*, 23(2), 313–329. [10.1007/s12650-020-00629-y](https://doi.org/10.1007/s12650-020-00629-y)
- [4] Nagpal, S., Bakshi, K. D., Kuntal, B. K., & Mande, S. S. (2020). NetConfer: A web application for comparative analysis of multiple biological networks. *BMC Biology*, 18(1), Article 1. [10.1186/s12915-020-00781-9](https://doi.org/10.1186/s12915-020-00781-9)
- [5] Dasgupta, A., Poco, J., Wei, Y., Cook, R., Bertini, E., & Silva, C. T. (2015). *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. [10.1109/tvcg.2015.2413774](https://doi.org/10.1109/tvcg.2015.2413774)
- [6] Rautenhaus, M., Böttinger, M., Siemen, S., Hoffman, R., Kirby, R. M., Mirzargar, M., Röber, N., & Westermann, R. (2018). *Visualization in Meteorology—A Survey of Techniques and Tools for Data Analysis Tasks*. *IEEE Transactions on Visualization and Computer Graphics*, 24, 3268–3296. [10.1109/TVCG.2017.2779501](https://doi.org/10.1109/TVCG.2017.2779501)
- [7] Nam, J. W., McCullough, K., Tveite, J., Espinosa, M. M., Perry, C. H., Wilson, B. T., & Keefe, D. F. (2019). Worlds-in-Wedges: Combining Worlds-in-Miniature and Portals to Support Comparative Immersive Visualization of Forestry Data. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 747–755. [10.1109/VR.2019.8797871](https://doi.org/10.1109/VR.2019.8797871)
- [8] Gleicher, M. (2018). Considerations for Visualizing Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 413–423. [10.1109/TVCG.2017.2744199](https://doi.org/10.1109/TVCG.2017.2744199)
- [9] Pagendarm, H.-G., & Post, F. H. (1995). Comparative Visualization: Approaches and Examples. *Visualization in Scientific Computing (H. Göbel, H. Müller and B. Urban)*, February, 95–108. <https://elib.dlr.de/37368/>
- [10] Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, 10(4), 289–309. [10.1177/1473871611416549](https://doi.org/10.1177/1473871611416549)
- [11] LYi, S., Jo, J., & Seo, J. (2021). Comparative Layouts Revisited: Design Space, Guidelines, and Future Directions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 1525–1535. [10.1109/tvcg.2020.3030419](https://doi.org/10.1109/tvcg.2020.3030419)
- [12] Srinivasan, A., Brehmer, M., Lee, B., & Drucker, S. M. (2018, January). What's the Difference? *ACM Conference on Human Factors in Computing Systems (CHI)*. [10.1145/3173574.3173878](https://doi.org/10.1145/3173574.3173878)
- [13] Fuchs, J., Fischer, F., Mansmann, F., Bertini, E., & Isenberg, P. (2013, April). *Evaluation of alternative glyph designs for time series data in a small multiple setting*. [10.1145/2470654.2466443](https://doi.org/10.1145/2470654.2466443)
- [14] Kim, K., Carlis, J. V., & Keefe, D. F. (2017). Comparison techniques utilized in spatial 3D and 4D data visualizations: A survey and future directions. *Computers & Graphics*, 67, 138–147. [10.1016/j.cag.2017.05.005](https://doi.org/10.1016/j.cag.2017.05.005)
- [15] Jardine, N., Ondov, B. D., Elmqvist, N., & Franconeri, S. (2020). The Perceptual

- Proxies of Visual Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 26(1), 1012–1021. [10.1109/TVCG.2019.2934786](https://doi.org/10.1109/TVCG.2019.2934786)
- [16] Ondov, B., Jardine, N., Elmqvist, N., & Franconeri, S. (2019). Face to Face: Evaluating Visual Comparison. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 861–871. [10.1109/TVCG.2018.2864884](https://doi.org/10.1109/TVCG.2018.2864884)
- [17] Ondov, B. D. (2021). *Revealing Perceptual Proxies in Comparative Data Visualization*. PhD Thesis. [Digital Repository at the University of Maryland]. [10.13016/RYUV-KZ1E](https://doi.org/10.13016/RYUV-KZ1E)
- [18] Eick, S. G., & Karr, A. F. (2002). Visual Scalability. *Journal of Computational and Graphical Statistics*, 11(1), 22–43. [10.1198/106186002317375604](https://doi.org/10.1198/106186002317375604)
- [19] Schmidt, J. (2016). *Scalable Comparative Visualization. Visual Analysis of Local Features in Different Dataset Ensembles*. PhD Thesis. [Faculty of Informatics (TU Wien)]. <https://diglib.eg.org/handle/10.2312/2631090>
- [20] Bonneau, G. P., Hege, H. C., Johnson, C. R., Oliveira, M. M., Potter, K., Rheingans, P., & Schultz, T. (2014). Overview and state-of-the-art of uncertainty visualization. In *Mathematics and Visualization* (Vol. 37, pp. 3–27). [10.1007/978-1-4471-6497-5_1](https://doi.org/10.1007/978-1-4471-6497-5_1)
- [21] Kamal, A., Dhakal, P., Javaid, A. Y., Devabhaktuni, V. K., Kaur, D., Zaiantz, J., & Marinier, R. (2021). Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization*. [10.1007/s12650-021-00755-1](https://doi.org/10.1007/s12650-021-00755-1)
- [22] Bolte, F., & Bruckner, S. (2020). Measures in Visualization Space. In *Foundations of Data Visualization* (pp. 39–59). Springer International Publishing. [10.1007/978-3-030-34444-3_3](https://doi.org/10.1007/978-3-030-34444-3_3)
- [23] Urribarri, D. K., & Castro, S. M. (2017). Prediction of data visibility in two-dimensional scatterplots. *Information Visualization*, 16(2), 113–125. [10.1177/1473871616638892](https://doi.org/10.1177/1473871616638892)
- [24] Castro, S. M., Larrea, M. L., Urribarri, D. K., Ganuza, M. L., & Escarza, S. (2018). *Métricas, técnicas y semántica para la visualización de datos*. XX Workshop De Investigadores En Ciencias De La Computación (WICC 2018). <http://sedici.unlp.edu.ar/handle/10915/6745>
- [25] Urribarri, D., Larrea, M., & Castro, S. (2019). Automatic suggestions to improve the quality of scatterplots during its creation. *Journal of Computer Science and Technology*, 19(2), 100–109. [10.24215/16666038.19.e10](https://doi.org/10.24215/16666038.19.e10)
- [26] Urribarri, D. K., Larrea, M. L., Castro, S. M., & Puppo, E. (2019). Visualization to compare karate motion captures. *XXV Congreso Argentino De Ciencias De La Computación (CACIC 2019, Universidad Nacional De Río Cuarto)*. <http://sedici.unlp.edu.ar/handle/10915/90359>
- [27] Urribarri, D. K., Larrea, M. L., Castro, S. M., & Puppo, E. (2020). Overview+Detail Visual Comparison of Karate Motion Captures. In P. Pesado & M. Arroyo, *Communications in Computer and Information Science: Vol. 1184 CCIS* (pp. 139–154). Springer, Cham. [10.1007/978-3-030-48325-8_10](https://doi.org/10.1007/978-3-030-48325-8_10)
- [28] Larrea, M. L. & Urribarri, D. K. (2021). Visualization Technique for Comparison of Time-based Large Data Sets. *Conference on Cloud Computing Conference, Big Data & Emerging Topics (JCC-BD&ET), Springer International Publishing, 1444*, 179-187. [10.1007/978-3-030-84825-5_13](https://doi.org/10.1007/978-3-030-84825-5_13)
- [29] Urribarri, D. K. & Larrea, M. L. (2022). A visualization technique to assist in the comparison of large meteorological datasets *Computers & Graphics, Elsevier BV*. [10.1016/j.cag.2022.02.011](https://doi.org/10.1016/j.cag.2022.02.011)