

Scatter Plot 3D

Maximiliano J. Escudero

María Luján Ganuza

Dayanna Wilberger

Sergio R. Martig

Departamento de Ciencias e Ingeniería de la Computación

VyGLab

Laboratorio de Investigación en Visualización y Computación Gráfica

Universidad Nacional del Sur

Avenida Alem 1253

Argentina, CP 8000, Bahía Blanca, Buenos Aires

{mje, mlg, srm}@cs.uns.edu.ar

Resumen

Las computadoras, Internet, y la Web generan continuamente una cantidad de datos que crece dramáticamente. Ante esta cantidad abrumadora de información surge la necesidad de transformar los datos en información que sea útil para quien debe usarla. De la necesidad de estudiar y obtener técnicas efectivas para comunicar información a través de imágenes nace la Visualización. Una herramienta de Visualización ampliamente utilizada para estudiar las relaciones entre los diferentes atributos de los items presentes en un conjunto de datos es el Scatter Plot. Un Scatter Plot 2D consiste en una representación gráfica utilizada en estadística para mostrar y comparar dos o más conjuntos de datos en cantidades finitas, teniendo una coordenada en el eje horizontal (X) y otra en el eje vertical (Y), presentando la información a partir de una distribución bivariada.

Resulta importante poder representar items de datos definidos a partir de tuplas cuya dimensionalidad sea superior a dos, en cuyo caso surge la necesidad de contar con un mapeo visual mas rico para poder soportar los atributos de los items de datos. En base a lo expuesto resulta útil diseñar una técnica de Scatter Plot 3D capaz de representar una mayor cantidad de valores por dato haciendo, por un lado, uso de la tercera dimensión para mapear un valor al tercer eje y enfatizando la explotación de las propiedades gráficas del elemento de representación utilizado. Se propone entonces el uso de Glifos como elemento de representación en un Scatter Plot 3D, aumentando potencialmente la dimensionalidad del mapeo visual.

Keywords: Visualización de información, Scatter Plot, Glifos

1. INTRODUCCIÓN

Si bien el objetivo común de la tarea de Visualización es la obtención de representaciones visuales interactivas con el propósito de la adquisición y el uso del conocimiento; según la naturaleza y características de la información a visualizar podemos hablar de tres tipos de visualización: Visualización científica, Visualización de Software y Visualización de información. En este artículo nos focalizaremos en la Visualización de Información en general, y en la técnica conocida como Scatter Plot en particular.

La Visualización de información se define como la visualización de datos abstractos, no basados en lo físico. Estos datos abstractos raramente serán representados por tuplas de dimensionalidad menor a 3, sino que generalmente se tratará de items de datos representados por tuplas de mayor dimensionalidad. Además, la ausencia de un sustrato espacial inherente a los datos a visualizar presenta

todo un desafío a la hora de abordar este tipo de visualizaciones; ya que el usuario requerirá representaciones visuales que le permitan estudiar interactivamente patrones y relaciones entre los distintos valores que pueden tomar las distintas tuplas.

El diagrama Scatter permite en cierto grado visualizar relaciones entre distintos datos, y es una de las siete herramientas básicas de control de calidad, el cual incluye: histograma, diagrama de Pareto, check sheet, diagrama de control, diagrama de causa-y-efecto y diagrama de flujo. En base a todo lo expuesto surge la necesidad de diseñar una técnica de Scatter Plot capaz de representar simultáneamente una mayor cantidad de valores por dato.

1.1. Scatter Plot 2D

Entre las técnicas más populares de visualización se encuentra el Scatter Plot 2D, que visualiza datos multidimensionales mapeando dos dimensiones a las coordenadas X e Y, y mapeando otras dimensiones a atributos visuales como color, tamaño, intensidad o forma [6].

En un estudio experimental, cada dato es representado por un punto en el espacio bidimensional, con el objetivo de estudiar la posible relación entre dichas variables. Entre las relaciones más conocidas se encuentran la lineal y la cuadrática, las cuales nos pueden ayudar a identificar patrones o clusters en los datos.

Como herramienta de visualización, el Scatter Plot 2D presenta varias ventajas, que se mencionan a continuación:

- Una rápida observación del gráfico revela algunas características del espacio de información analizado. Proveen patrones que pueden contener información importante y ayudar a focalizar la atención en áreas que necesitan especial consideración.
- Una rápida inspección de los gráficos permite detectar velozmente los valores extremos y agrupamientos de puntos.
- No requiere que el usuario especifique variables dependientes e independientes. Cualquier tipo de variable puede ser planteada independientemente del eje que se le asigne.

En contrapartida, esta herramienta presenta ciertas limitaciones:

- Los Scatter Plots se exploran desde el exterior hacia el interior. La facilidad de navegación a través a los elementos visuales depende enormemente de la cantidad y complejidad de los mismos. Para poder ver un gran número de elementos es necesario reducir su complejidad.
- En general, en un Scatter Plot simple, los puntos (o elemento de representación) se representan independientemente unos de otros, lo cual podría implicar que no se detecte visiblemente una conexión entre ellos.
- En un Scatter Plot, la posición, el color y el tamaño del elemento gráfico se basan completamente en los valores de los atributos de los datos. Si un ítem de datos está ausente o perdido (no tiene ningún valor asociado), se le podría asignar un valor por defecto, pero puede llevar a conclusiones erróneas), ya que no existe nada en el gráfico que indique la existencia de valores de datos ausentes.

A continuación analizaremos la técnica Scatter Plot aplicada a 3 dimensiones.

2. SCATTER PLOT 3D

Los Scatter Plots 3D corresponden a una extensión conceptualmente simple de los Scatter Plot 2D [7] [1]. Los Scatter Plots 3D visualizan una relación entre tres o más variables, explotando la tercera dimensión, representando las coordenadas X,Y y Z en el espacio tridimensional. Eventualmente se podrían agregar más coordenadas al modelo, llevándolo al espacio n-dimensional.

En versiones más complejas, los scatter plots 3D pueden incluir atributos específicos para los elementos de representación según el dato representado (color, tamaño, orientación, forma, etc), guías (líneas de referencia desde los datos hasta algún punto de referencia) y combinaciones de datos scatter con objetos adicionales como superficies de ajuste. Una aplicación común de los Scatter Plots 3D es mostrar tanto los datos relevados de manera experimental como aquellos de ajuste teórico, a fin de poder determinar los puntos de concordancia de aquellos que no lo son.

En la figura 1(b) se puede observar un Scatter Plot en tres dimensiones que hace uso del tamaño variable de las esferas para mapear un atributo adicional.

Los Scatter Plots 3D presentan ciertas ventajas y limitaciones respecto a los modelos en 2D. A continuación se detallan algunas de ellas.

2.1. Ventajas

- En un Scatter Plot 3D mantener la misma densidad de puntos que en un Scatter Plot 2D implica aumentar el número de datos experimentales a mostrar (mayor espacio muestral). Si mantengo el número de puntos del scatter plot inicial (2D), existe mayor discriminación de las relaciones existentes entre variables, dado que se incorpora una característica más de los datos.
- El uso de visualizaciones de volúmen de Scatter Plots 3D con glifos para representar los datos brinda la posibilidad de utilizar técnicas de generación procedural de formas [5]. Estas técnicas permiten aumentar el número de dimensiones de los datos a mostrar explotando la forma de los glifos, aprovechado así la habilidad pre-atentiva del sistema visual humano para discriminar formas.
- Si se logra un mapeo de atributos eficiente, y se proveen las interacciones necesarias para la navegación y consulta, el poder expresivo de un Scatter Plot 3D podría permitir el rápido análisis de relaciones complejas entre múltiples variables.

2.2. Limitaciones

- No es recomendable abusar de la multidimensionalidad si no es absolutamente necesario y el resultado no es visualmente ilustrativo.
- Trasladar representaciones de información en 2-dimensiones a 3-dimensiones no es una tarea simple, ya que la dimensión extra afecta enormemente el modo en que la información puede ser presentada e interpretada. La visualización debe realizar un uso eficiente de la nueva dimensión y evitar que la nueva representación sea mal interpretada por el usuario como consecuencia de un mapeo inapropiado.
- Es necesario tener especial consideración respecto de la percepción de la distancia espacial. El tamaño de los objetos puede provocar que el usuario no perciba la perspectiva correcta de la información mostrada; es decir, se dificulta la discriminación entre las distintas profundidades de los objetos, por lo cual resulta necesario proveer las interacciones adecuadas.

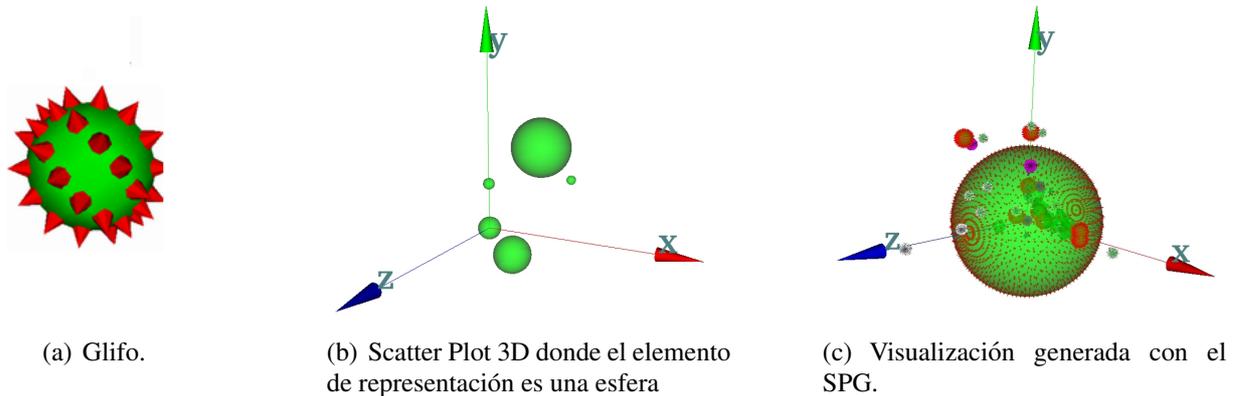


Figura 1: Scatter Plot 3D con esferas y glifos.

- Una desventaja proveniente del uso de objetos tridimensionales es la oclusión, que se da cuando un objeto tapa a otro ocultándolo en el espacio muestral. Este tipo de problema se presenta principalmente cuando la densidad de items de datos a mostrar es grande, o cuando simplemente un objeto muy grande se posiciona por delante de objetos más pequeños.

3. SCATTER PLOT 3D SPG

En base a todo lo expuesto surgió la necesidad de diseñar una técnica de Scatter Plot 3D capaz de visualizar datos representados por tuplas de gran dimensionalidad.

Para lograr esto se estudio la posibilidad de introducir los Glifos como elementos de representación dentro de la técnica Scatter Plot. Los glifos [Figura 1(a)] son objetos gráficos en tres dimensiones que representan uno o más valores en una única locación del espacio. Son formas o imágenes generadas por mapeo de componentes de datos a atributos gráficos. Un único glifo puede utilizarse para representar muchas propiedades o atributos de los datos en un punto dado del espacio, de forma tal que resultan ser herramientas muy útiles a la hora de interpretar una gran cantidad de información a primera vista.

Se está desarrollando una aplicación, llamada SPG, que permite visualizar un conjunto de datos representados por tuplas de dimensionalidad mayor a 3 mediante el uso de la técnica Scatter Plot aplicada a tres dimensiones utilizando glifos como elemento de representación. Dicha aplicación permitió en primera instancia visualizar items de datos multidimensionales. El SPG toma un conjunto de datos de entrada que puede residir en un archivo o ser generado proceduralmente y genera una visualización en 3D basada en glifos, tal como puede observarse en la figura 1(c). En una primera instancia, el SPG permite mapear datos a 6 atributos gráficos, con posibilidad de extender el mapeo a 8 atributos si se agregan ciertas interacciones sobre la vista que permitan revelar detalles que de otro modo no serían claramente perceptibles por el usuario.

Una técnica como la propuesta exige el diseño de las interacciones adecuadas para asegurar la provisión de información contextual y detalle bajo demanda.

4. AGRADECIMIENTOS

El presente trabajo fue parcialmente financiado por PGI 24/N015 y PGI 24/ZN12, Secretaría General de Ciencia y Tecnología, Universidad Nacional del Sur, Bahía Blanca, Argentina.

5. TRABAJO FUTURO

Analizar el poder expresivo de la herramienta visualizando distintos conjuntos de datos y efectuando distintos mapeos de los atributos de datos a atributos gráficos, estudiando los patrones resultantes en base a la información provista por la visualización.

Se definirán e implementarán nuevas interacciones a nivel de usuario, en caso que se consideren necesarias, con el fin de proveer herramientas que ayuden a explorar y estudiar los datos mostrados, y de proveer información contextual y detalle bajo demanda.

Se estudiará particularmente el problema de la oclusión, el cual se da en el caso que dos o más items de datos fueran mapeados a glifos situados en locaciones espaciales muy cercanas. Esto da lugar a visualizaciones donde algunos glifos quedan ocultos detrás de otro, dado su inferior tamaño o su ubicación en la pantalla relativa al usuario. Buscando distintas soluciones al problema y analizando ventajas y desventajas de las mismas.

Por último, se hará especial hincapié en la performance de la aplicación desarrollada, con el objetivo de lograr una aplicación interactiva y eficaz para los usuarios.

REFERENCIAS

- [1] B. G. Becker. Volume rendering for relational data. pages 87–91.
- [2] Benjamin B. Bederson and Ben Shneiderman. *The Craft of Information Visualization: Readings and Reflections*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.
- [3] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, editors. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.
- [4] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical Methods for Data Analysis*. Wadsworth Internat. Group, 1983.
- [5] D. S. Ebert, R. M. Rohrer, C. D. Shaw, P. Panda, J. M. Kukla, and D. A. Roberts. Procedural shape generation for multi-dimensional data visualization. *Computers & Graphics*, 24:375–384, 2000.
- [6] J. Fekete and C. Plaisant. Interactive information visualization of a million items proceedings of ieeec symposium on information visualization, 2002.
- [7] R. Kosara, G. Sahling, and H. Hauser. Linking scientific and information visualization with interactive 3d scatterplots, 2004.
- [8] Sergio Martig, Silvia Castro, Pablo Fillottrani, and Elsa Estévez. Un modelo unificado de visualización. In *Proceedings 9º Congreso Argentino de Ciencias de la Computación.*, pages 881–892, 2003.
- [9] P. Rheingans and C. Landreth. Perceptual principles for effective visualizations, 1995.
- [10] Bernice E. Rogowitz, Lloyd A. Treinish, and Steve Bryson. How not to lie with visualization. *Comput. Phys.*, 10(3):268–273, 1996.
- [11] Ware, Colin. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers, 2000.