

Universidad Tecnológica de Chihuahua
Tecnologías de la Información



**Universidad Tecnológica
de Chihuahua**

Reporte de investigación de técnicas
de visualización

Alumno:

Jatzel Israel Cruz Castruita

Grupo:

IDGS91N

Materia:

Extracción de Conocimiento en Bases de Datos

Docente:

Enrique Mascote

Introducción.....	3
Técnicas de visualización y representación de la información.....	4
Relación entre visualización y extracción del conocimiento.....	6
Conclusión.....	8
Referencias.....	9

Introducción

En la actualidad, el volumen de datos generados por organizaciones, dispositivos y usuarios crece de manera exponencial, lo que convierte a la información en uno de los recursos más valiosos para la toma de decisiones. Sin embargo, la utilidad de estos datos depende de la capacidad para analizarlos, interpretarlos y transformarlos en conocimiento significativo. En este contexto, las técnicas de visualización y representación de la información desempeñan un papel fundamental dentro del proceso de extracción de conocimiento. La visualización no solo facilita la comprensión inicial de los datos, sino que también permite identificar patrones, relaciones, tendencias y anomalías que serían difíciles de detectar mediante métodos exclusivamente numéricos o tabulares.

Este documento aborda la importancia de estas técnicas como herramientas esenciales en las distintas fases del proceso de descubrimiento de conocimiento en bases de datos (KDD). A través de gráficos, mapas, diagramas y representaciones multidimensionales, la visualización contribuye tanto a la exploración preliminar de los datos como a la interpretación final de los resultados obtenidos por los algoritmos de minería de datos. La investigación presentada busca analizar y explicar cómo estas técnicas influyen en la calidad del conocimiento extraído y de qué manera se integran eficazmente en un flujo de análisis de datos moderno.

Técnicas de visualización y representación de la información

Visualizaciones descriptivas

Histogramas

- Muestran cómo se distribuyen los valores de una variable (por ejemplo, edades, ingresos, alturas).
- Ayudan a ver si los datos tienen forma normal, si están sesgados o si hay grupos naturales.

¿Para qué sirven en minería de datos?

Antes de aplicar algoritmos, ayudan a entender la forma de los datos y decidir transformaciones normalización y escalamiento.

Diagramas de Caja (Boxplots)

- Resumen de una variable con su valor mínimo, máximo, mediana y cuartiles.
- Detectan outliers claramente.

Relación con extracción de conocimiento:

Si hay muchos valores atípicos, pueden afectar a algoritmos como K-means; un boxplot te alerta de eso.

Gráficas de dispersión (Scatter plots)

- Comparan dos variables y muestran si existe algún tipo de relación (lineal, cuadrática, agrupamientos).
- Son esenciales para observar patrones naturales.

Ejemplo:

Si ves que los puntos se agrupan en 3 zonas, podrías pensar que K-means debe usar $K=3$.

Visualizaciones multivariadas

Mapas de calor (heatmaps)

- Representan matrices de correlación o valores en colores.
- Permiten ver rápidamente qué variables están correlacionadas.

Importancia en KDD:

Variables altamente correlacionadas pueden ser redundantes; esto guía la reducción de dimensionalidad.

Matriz de dispersión (pairplot)

- Crea múltiples scatter plots para varias variables a la vez.
- Permite ver interacciones entre muchas variables.

Ejemplo:

En el dataset Iris, ayuda a ver qué medidas de la flor separan las especies.

Visualizaciones para modelos de Machine Learning

Gráficas de PCA (mapas 2D o 3D)

El PCA reduce la dimensionalidad a 2 o 3 componentes y los grafica.

¿Qué revela?

- Si las clases o grupos están separados,
- Si un algoritmo como clustering puede encontrar patrones,
- Si los datos tienen estructura.

Ejemplo:

Si los puntos forman nubes claramente separadas, K-means funcionará bien.

Visualización de clusters

Después de aplicar K-means u otro método, cada punto se colorea según su clúster.

¿Qué permite ver?

Si los grupos están bien separados o mezclados.

Si la elección de K fue correcta.

Si hay clústeres naturales no detectados antes.

Relación entre visualización y extracción del conocimiento

Fase 1: Comprensión del negocio y de los datos

La visualización permite:

- Detectar patrones iniciales,
- Entender la variabilidad de los datos,
- Conocer el tipo de relaciones entre variables.

Fase 2: Preparación de los datos

La limpieza de datos depende de ver:

- Outliers (boxplots),
- Faltantes (mapas de calor),
- Distribuciones anómalas (histogramas),
- Relación entre variables (pairplots).

Fase 3: Minería de datos (algoritmos)

Las visualizaciones permiten:

- Evaluar si los clústeres obtenidos tienen sentido,
- Ver si el PCA capturó la varianza importante,
- Validar modelos supervisados (curvas ROC, etc.).

Fase 4: Interpretación y evaluación del modelo

Aquí la visualización es fundamental para convertir resultados matemáticos en ideas útiles.

Ejemplos:

- Un mapa de calor de correlaciones revela qué variables influyen más.
- Una gráfica de PCA muestra si el modelo aprendió correctamente.
- La gráfica de clusters confirma si hay separaciones claras.

Fase 5: Presentación del conocimiento

La visualización es la forma final de comunicar conocimiento a quienes toman decisiones.

Ejemplo:

Un directivo no quiere ver matrices numéricas, pero una gráfica que muestra que las ventas caen en ciertos meses sí.

Conclusión

Al investigar sobre las técnicas de visualización y su relación con el proceso de extracción de conocimiento, comprendí que la visualización no es solamente una etapa final para presentar resultados, sino una herramienta esencial en todo el ciclo de análisis de datos. Descubrí que representar la información de manera gráfica me permite entender mejor la estructura de los datos, detectar patrones, identificar valores atípicos y reconocer relaciones que no son evidentes a simple vista. También entendí que estas representaciones influyen directamente en la toma de decisiones durante la preparación de datos y la selección de algoritmos, ya que ayudan a evaluar la calidad y relevancia de los resultados obtenidos.

Finalmente, aprendí que la visualización es un puente entre los datos y el conocimiento: convierte información compleja en algo accesible, interpretable y útil. Esta investigación reforzó mi comprensión de que, en cualquier proyecto de análisis de datos, las técnicas de visualización no solo apoyan el descubrimiento de conocimiento, sino que también permiten comunicar de forma clara y significativa.

Referencias

Chen, C., Härdle, W. K., & Unwin, A. (2008). *Handbook of data visualization*. Springer.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37–54.

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*. Springer.

Han, J., Pei, J., & Kamber, M. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: Strategies for effective human-computer interaction* (5th ed.). Addison-Wesley.

Tufte, E. R. (2001). *The visual display of quantitative information* (2nd ed.). Graphics Press.

Ware, C. (2013). *Information visualization: Perception for design* (3rd ed.). Morgan Kaufmann.