

VisDecode: Extracción de decisiones de diseño en visualizaciones con modelos pixel-to-text

Martín A. Sinnona¹, Viviana Siless¹, Emmanuel Iarussi^{1,2}

¹ Universidad Torcuato Di Tella

² Consejo Nacional de Investigaciones Científicas y Técnicas

Abstract. En este trabajo introducimos VisDecode, un framework para extraer decisiones de diseño a partir de visualizaciones. Haciendo uso de imágenes rasterizadas de visualizaciones genéricas (gráficos de barra, líneas, dispersión), nuestro algoritmo identifica atributos perceptuales y los relaciona con los datos. Entrenamos una red neuronal profunda con un dataset sintético que consiste en pares de visualizaciones-decisiones de diseño. Estas visualizaciones fueron rasterizadas a partir de tablas de datos, generadas aleatoriamente por modelos de lenguaje, que incluyen variables cuantitativas y categóricas que se asemejan a datos del mundo real. Luego del entrenamiento, VisDecode es capaz de extraer decisiones de diseño de gráficos provenientes de internet. Nuestro objetivo a mediano plazo es poder mejorar las interfaces de visualización para lograr mejores prácticas de diseño.

Keywords: Visualización de datos · Extracción de atributos · Lenguaje visual.

1 Introducción

La comprensión y extracción automática de características de visualizaciones es un desafío fundamental en el área de visualización de datos. Las visualizaciones desempeñan un papel crucial en la comunicación de información derivada de conjuntos de datos complejos. Sin embargo, la interpretación precisa de un gráfico puede ser difícil debido a las decisiones de diseño que influyen en la percepción del usuario. Los atributos visuales como colores, formas, posición, entre otros, tienen un impacto significativo en cómo se percibe la información presentada. Por ejemplo, la elección de colores puede afectar la legibilidad y la interpretación de los datos, mientras que la posición y el tamaño de los elementos pueden influir en la atención y en la comprensión de la relación entre diferentes partes de una visualización.

Por otro lado, están surgiendo modelos que permiten recuperar los datos que producen un determinado gráfico, por ejemplo, Matcha [1]. No obstante, no existen modelos que permitan simultáneamente destilar las decisiones de diseño de estos gráficos.

En respuesta a este problema, presentamos VisDecode, un framework que permite extraer automáticamente decisiones de diseño, en particular, de gráficos estadísticos, y mediante Matcha [1] obtiene la tabla de datos asociada al gráfico. Eventualmente, este conjunto de decisiones de diseño y datos, permitiría la reconstrucción total de los gráficos, la sugerencia automatizada de buenas prácticas de diseño, entre otras tareas que motivan a este proyecto.

VisDecode nace de Matcha [1], el cual entrenamos para nuestra tarea específica de extraer decisiones de diseño a partir de gráficos. Se entrenó utilizando un dataset sintético de pares de visualizaciones - decisiones de diseño, el cual fue generado a partir de una tabla de datos aleatoria provista por GPT-3.5 con diversas variables cuantitativas y categóricas.

2 Extracción de atributos

VisDecode es capaz de extraer, únicamente a partir de la imagen de un gráfico, distintos atributos que forman parte de las decisiones tales como tipo de gráfico (barras, línea, dispersión), tipo de variables (cuantitativas, temporales, nominales u ordinales [3]), y nombres de las variables. Estos atributos son proporcionados en formato XML, adicionalmente relacionando estos con las dimensiones de datos correspondientes (ver Fig. 1). Para la extracción de los nombres de las variables empleamos un modulo extra que se encarga del OCR.

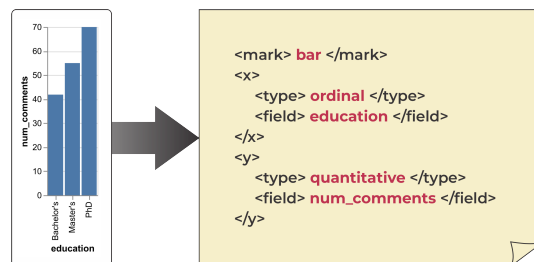


Fig. 1. A la izquierda, una visualización aleatoria extraída del set de testing. A la derecha, decisiones de diseño en formato XML dadas por VisDecode.

3 Dataset sintético

En nuestra labor inicial, no encontramos disponible un dataset masivo con pares de visualizaciones - decisiones de diseño a partir del cual realizar el entrenamiento. Por lo tanto, generamos proceduralmente dicho dataset a partir de una tabla de datos aleatorios proporcionada por GPT-3.5. Esta tabla contiene en total 45 variables de tipo cuantitativas, temporales, nominales y ordinales, que se asemejan a datos del mundo real, y 24 valores distintos para cada variable.

A partir de la tabla, tomamos todos los pares de variables que tenían sentido entre sí y rasterizamos las imágenes haciendo uso de Vega-lite. Alrededor de 18k imágenes fueron generadas mediante este proceso (ver Fig. 2). Sin embargo, dada la naturaleza procedural de este método, es posible generar tantas visualizaciones como queramos añadiendo variaciones en su diseño.

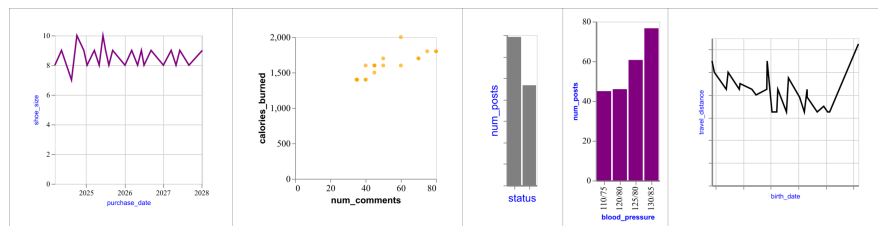


Fig. 2. Algunas de las visualizaciones generadas a partir de la tabla de datos aleatoria, con variaciones en tamaños, fuentes y colores.

4 Configuración experimental

Haciendo uso del dataset sintético de imágenes - decisiones de diseño, con 5000 instancias, entrenamos Matcha [1] de forma supervisada durante 300 épocas hasta lograr una buena performance en el set de validación, de 500 instancias (10% val - 80% train). El proceso se llevó a cabo en una NVIDIA A100 de 80GB durante 8 horas aproximadamente, con Adam optimizer.

5 Resultados preliminares

Evaluamos VisDecode sobre nuestro dataset de testing (VisDecode Test), un dataset construido a partir de gráficos de internet (VisDecode Web) y PlotQA[2]. Se midió la eficacia al extraer los atributos de tipo de gráfico, "mark type", (barras, líneas o dispersión) y tipo de variables, "x-y types", (cuantitativa, temporal, nominal u ordinal) a partir de un accuracy que cuantifica el porcentaje de aciertos por sobre el total de clases disponibles (tres para el tipo de gráfico y cuatro para los tipos de variables).

Podemos notar que a la hora de extraer el tipo de gráfico (mark type) los resultados son consistentes a través de los distintos datasets. Sin embargo, cuando se trata de extraer los tipos de variables (x-y types) vemos que el accuracy disminuye para visualizaciones desconocidas. Con respecto a esto, vemos que extraer el tipo de variables únicamente a partir de información visual no es una tarea trivial, ya que, en muchos casos puede ser ambiguo y quizás se requiera apoyo del módulo de OCR para este fin.

Table 1. Accuracy de extracción de los distintos atributos. base es MatCha entrenado con 1k instancias del dataset propio sin variabilidad en los estilos visuales. Se compara con MatCha entrenado con variabilidad en estilos visuales.

Accuracy	VisDecode Test	VisDecode Web	PlotQA
Mark type acc. (base)	0.41	0.40	0.35
Mark type acc.	1.00	1.00	0.98
X-Y types acc. (base)	0.42	0.40	0.35
X-Y types acc.	0.90	0.70	0.68

Por otro lado, vimos que respecto a la generación del texto de la estructura que acompaña a los atributos no se presentaron dificultades algunas, y es por esto que no se juzgó este proceso.

6 Conclusiones

VisDecode resulta útil en la extracción de decisiones de diseño a partir de la imagen de un gráfico y muestra una buena generalización para el atributo "mark type". Respecto a "x-y types" aún quedan mejoras pendientes para lograr una mayor generalización con imágenes desconocidas. Este trabajo representa un paso inicial hacia la extracción masiva de decisiones de diseño a partir de visualizaciones. Su potencial radica en su capacidad para fomentar la comprensión y mejorar el diseño de visualizaciones, lo que resulta en representaciones visuales más efectivas.

Objetivos futuros

Para potenciar aún más su utilidad y alcance, es fundamental continuar refinando y perfeccionando su capacidad para generalizar y extraer información de manera precisa a partir de una amplia variedad de imágenes de gráficos. Este proceso de mejora garantizará que VisDecode se convierta en una herramienta valiosa para diseñadores y analistas de visualizaciones de datos.

Mirando hacia el futuro, el objetivo para VisDecode es transformarlo en una herramienta que no solo extraiga decisiones de diseño de visualizaciones, sino que también sugiera y mejore automáticamente las visualizaciones basada en los datos, decisiones existentes, y conocimiento adquirido a partir del análisis a gran escala de otras visuales existentes.

References

1. Liu, F., Piccinno, F., Krichene, S., Pang, C., Lee, K., Joshi, M., ... & Eisenschlos, J. M. (2022). Matcha: Enhancing visual language pretraining with math reasoning and chart derendering.
2. Methani, N., Ganguly, P., Khapra, M. M., & Kumar, P. (2020). Plotqa: Reasoning over scientific plots. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1527-1536).
3. Tamara Munzner. Visualization Analysis and Design. A K Peters Visualization Series