What Would They Say? Predicting User's Comments in Pinterest

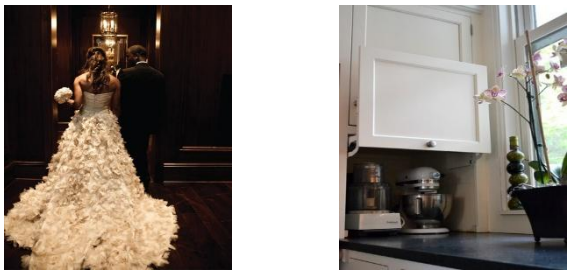
J. C. Gomez, T. Tommasi, S. Zoghbi and M. F. Moens

***Abstract*— When we refer to an image that attracts our attention, it is natural to mention not only what is literally depicted in the image, but also the sentiments, thoughts and opinions that it invokes in ourselves. In this work we deviate from the standard mainstream tasks of associating tags or keywords to an image, or generating content image descriptions, and we introduce the novel task of automatically generate user comments for an image. We present a new dataset collected from the social media Pinterest and we propose a strategy based on building joint textual and visual user models, tailored to the specificity of the mentioned task. We conduct an extensive experimental analysis of our approach on both qualitative and quantitative terms, which allows to assess the value of the proposed approach and shows its encouraging results against several existing image-to-text methods.**

***Keywords*— Multimodal Clustering, Pinterest, Social Media, User Generated Content, Deep-Learning Representation.**

I. INTRODUCCIÓN

EL USO de imágenes digitales en nuestra vida diaria ha cambiado mucho en los últimos años. Por un lado nos hemos convertido en proveedores prolíficos de imágenes gracias a la creciente popularidad de las cámaras digitales y los teléfonos inteligentes. Por el otro, también nos hemos convertido en consumidores ávidos y activos: casi cualquier sitio web está enriquecido con fotos o imágenes, además de que cuando navegamos en la Web las imágenes son las que comúnmente atraen más nuestra atención. Las redes sociales como Facebook and Pinterest han contribuido a impulsar estas tendencias, como lo confirma el hecho de que más de 300 millones de fotos por día se han subido a Facebook desde 2012 [1]. A su vez, Pinterest permite a sus usuarios crear tableros de marcadores visuales llamados *pins*. En estos tableros cada usuario puede colectar y guardar las imágenes que se encuentra en línea con el propósito de compartir contenido, planear viajes, coleccionar recetas, etc. Las imágenes que coleccionan los usuarios en Pinterest son comúnmente publicadas en sus tableros junto con comentarios cortos en forma de texto. A diferencia de una anotación con etiquetas de imágenes, o de una descripción detallada del contenido visual de las mismas, los comentarios que los usuarios ponen en este sitio suelen expresar emociones, opiniones o pensamientos, en conjunto con una descripción superficial de lo que aparece en las imágenes. Tales comentarios puedan dar una pista sobre los intereses personales de los usuarios, su forma de pensar y su estilo de escribir o expresarse (ver Fig. 1). Aunque explorar toda esta información puede tener un rol clave en diversas tareas como la creación de modelos de usuarios, análisis de sentimientos y publicidad personalizada, hasta donde sabemos este tipo de comentarios específicos aún no ha sido completamente estudiado.



Love this train — *Hinged cabinet. Love this to hide appliances*

Figura 1. Ejemplo de dos imágenes en Pinterest con comentarios personales de los usuarios.

En este trabajo queremos dar un paso más en el análisis del contenido generado por el usuario con tres contribuciones principales: 1) Introducimos una nueva tarea: predecir automáticamente los comentarios de los usuarios para sus imágenes; 2) presentamos una nueva colección de datos (consistente en más de 70,000 imágenes junto con los comentarios de los usuarios); y 3) realizamos un estudio experimental extensivo para la tarea.

El resto del artículo está organizado de la siguiente manera: en la sección 2 hacemos una breve revisión de la literatura; en la sección 3 describimos la tarea propuesta y la colección de datos utilizada para los experimentos; en la sección 4 presentamos y discutimos nuestro enfoque; en la sección 5 mostramos los resultados experimentales; y en la sección 6 concluimos el artículo con una discusión general y posibles líneas futuras de investigación.

II. TRABAJO RELACIONADO

Existen dos líneas principales en la literatura relacionada con asociar texto a imágenes. La primera se enfoca en asignar palabras clave (o etiquetas) a una imagen; la segunda en proveer una descripción completa de la imagen. En ambos casos el objetivo final es reconocer el contenido de la imagen en términos de los objetos [7] o de la escena [5] mostrados. Para ello, los trabajos previos utilizan principalmente estrategias basadas en contenido, las cuales predicen el texto para las imágenes ya sea al entrenar un modelo con la relación entre texto e imágenes [16][17], o al propagar las etiquetas a través de un método de vecinos más cercanos [4][15]. En estos estudios, el análisis se realiza generalmente en colecciones de

J. C. Gomez, KU Leuven, Belgium, jcgcarranza@gmail.com
T. Tommasi, UNC Chapel Hill, USA, ttommasi@cs.unc.edu
S. Zoghbi, KU Leuven, Belgium, susana.zoghbi@cs.kuleuven.be
S. Moens, KU Leuven, Belgium, sien.moens@cs.kuleuven.be