

# Ordenamiento de imágenes Web de acuerdo a su relevancia utilizando un enfoque de fusión multimodal

## Reporte final

Ricardo Omar Chávez García

Instituto Nacional de Astrofísica Óptica y Electrónica, 72840 Luis Enrique Erro # 1, Tonantzintla, Puebla, México

**Resumen** En la actualidad los resultados que se obtienen del buscador de imágenes Google Images<sup>1</sup> están basados solamente en el texto asociado a las imágenes, lo cual no ofrece una garantía de que las imágenes en las primeras posiciones sean en verdad relevantes. En este trabajo se presenta un sistema que ordena por relevancia las imágenes obtenidas directamente del buscador Google Images. Para ello se propone un enfoque de fusión multimodal que, para representar una imagen, utiliza el histograma de color en RGB como característica visual y una bolsa de palabras, obtenida del texto relacionado con la imagen, como característica textual. Posteriormente se hace uso de un clasificador Naive Bayes o una red Bayesiana TAN para obtener el ordenamiento de las imágenes al realizar su clasificación en las clases *relevante* o *no-relevante*. Para evaluar el sistema propuesto se utilizaron 4 corpus de 100 imágenes(etiquetadas manualmente) recuperadas con Google Images al introducir los términos de búsqueda “*table, airplane, bicycle y watch*” y dos métodos de selección de ejemplos de entrenamiento, el primero, un método automático basado en la posición de las imágenes, y el segundo, un método basado en la retroalimentación del usuario. Para medir el desempeño del sistema propuesto se utilizaron las medidas de precisión y recuerdo por clase y un porcentaje de instancias clasificadas correctamente como medida global. Los resultados de los experimentos indican que, al utilizar el sistema y enfoque propuesto, para algunos términos de búsqueda, se logra una mejora en el orden de relevancia de las imágenes respecto a la lista de imágenes recuperadas con Google Images.

## 1. Introducción

Para realizar la recuperación de imágenes, el buscador de Google Images toma en cuenta un conjunto de términos de búsqueda que posteriormente compara con las características textuales de las imágenes en la Web. Estas características textuales son palabras extraídas del URL de la imagen, del URL de la página Web que la contiene, del texto al rededor de la imagen, del texto dentro de

---

<sup>1</sup> <http://images.google.com/imghp?hl=en>

las etiquetas HTML que definen a la imagen(etiquetas  $\langle DIV \rangle$ ,  $\langle IMG \rangle$ ,  $\langle A \rangle$ ), así como del texto en los atributos de estas etiquetas. Debido a que sólo se consideran las características textuales para la recuperación, no todas las imágenes recuperadas corresponden con los términos de búsqueda, y como consecuencia no existe un orden de relevancia en los resultados, este problema provoca que, en ocasiones, el usuario tenga que navegar por varias páginas de resultados antes de encontrar la imagen relevante deseada.

El problema de no recuperar todas las imágenes relevantes de acuerdo a los términos de búsqueda está siendo abordado por la tarea de recuperación de imágenes por contenido(Content-Based Image Retrieval)[2], la cual también toma en cuenta las características visuales de las imágenes para la recuperación. Para tratar de solucionar la problemática de no tener un orden de relevancia en un conjunto de imágenes recuperadas se han propuesto enfoques que utilizan, como discriminante, solamente a las características visuales[1], o enfoques multimodales que toman en cuenta también a las características textuales[6] para clasificar a las imágenes, pero que consideran una etapa de selección de imágenes previa al ordenamiento.

En este trabajo se propone utilizar un enfoque de fusión multimodal para representar a las imágenes mediante características visuales y textuales para luego clasificarlas, de acuerdo a ejemplos positivos y negativos dados, en imágenes *relevantes* y *no-relevantes*, utilizando para ello un clasificador Naive Bayes o una red Bayesiana TAN. Como característica visual se utilizó el histograma de color en RGB de la imagen, ya que se ha demostrado empíricamente[4, 3] que, aunque es una de las características más básicas de una imagen, es muy útil para caracterizar un grupo de imágenes comunes. Para las características textuales se utilizó un enfoque de bolsa de palabras, pesadas de forma binaria, que son extraídas de la página Web que referencia directamente a cada imagen. Las palabras se extrajeron de los mismos lugares, en el documento Web, de donde Google Images extrae sus palabras clave y que fueron mencionados con anterioridad.

Debido a que el orden de relevancia de las imágenes está dado por el usuario, evaluar los sistemas de ordenamiento de imágenes se vuelve una tarea difícil, y por ello, muchos de los trabajos que se han propuesto evalúan sus enfoques con corpus que no son obtenidos directamente de un motor de búsqueda Web[4] o consideran una etapa de selección de imágenes previa[1] al ordenamiento. A este problema se le suma la falta de una métrica estándar para medir el desempeño de un sistema de ordenamiento de imágenes recuperadas de la Web. En este trabajo se utiliza la salida directa del buscador Google Images para realizar el ordenamiento, no se toma en cuenta ninguna etapa de pre-selección de imágenes. Para medir el desempeño del sistema se utilizaron medidas como el recuerdo y la precisión por clase, y el porcentaje de instancias clasificadas correctamente.

Para clasificar a las imágenes se eligió el clasificador Naive Bayes y la red Bayesiana TAN, ya que se quiere analizar el comportamiento del enfoque propuesto con clasificadores utilizados comúnmente en la tarea de clasificación. Además, se quiere observar cuál es el desempeño de los clasificadores con enfoque probabilistas cuando se utilizan características como las propuestas en este trabajo.

Para evaluar el sistema se utilizaron las primeras 100 imágenes de 4 consultas, de un solo término de búsqueda, realizadas en el buscador Google Images. Los términos de búsqueda utilizados fueron “*table, airplane, bicycle y watch*”. De las imágenes obtenidas se utilizaron 20 como entrenamiento para construir el clasificador y 80 para realizar las pruebas y las predicciones. Para la selección de ejemplos de entrenamiento positivos y negativos se utilizaron dos métodos: el primero es automático y está basado en la posición de los ejemplos en la lista de imágenes recuperadas, y el segundo está basado en la selección de imágenes hecha por el usuario. Las 80 imágenes de cada uno de los 4 corpus de prueba utilizados para la evaluación fueron etiquetados manualmente. Se consideró que una imagen es relevante si el término de búsqueda concuerda con el objeto principal de la imagen y el texto de la página Web que la referencia provee información adicional sobre el término de búsqueda.

El proceso utilizado para realizar el ordenamiento consta de 5 etapas: recuperación de imágenes, obtención de características, generación de vectores de características, construcción-evaluación de clasificadores y predicción de instancias. En la etapa de recuperación de imágenes se envía una consulta, de un sólo término de búsqueda, a Google Images para obtener las primeras 100 imágenes recuperadas con sus respectivas páginas Web. La etapa de obtención de características extrae el histograma de color en RGB y la bolsa de palabras de cada una de las imágenes. En la etapa de generación de vectores de características se construyen los corpus de entrenamiento y prueba con las características del paso anterior. En la etapa de construcción-evaluación de clasificadores se construye el clasificador NaiveBayes y la red bayesiana TAN para el corpus de entrenamiento y se evalúan utilizando el corpus de prueba. En la etapa de predicción se utiliza el modelo creado en el paso anterior para predecir la clase de cada una de las instancias del corpus de prueba, al final de esta etapa se obtiene el ordenamiento de las imágenes recuperadas.

El resto del reporte está organizado como sigue. La sección 2 muestra un panorama general del área de ordenamiento de imágenes Web recuperadas. En la sección 3 se detalla el proceso utilizado para generar el ordenamiento de las imágenes. En la sección 4 se describen los corpus utilizados así como los experimentos realizados y un análisis de los resultados obtenidos. Para finalizar, en la sección 5 se exponen las conclusiones del trabajo realizado así como algunas propuestas de trabajos a futuro.

## **2. Trabajo relacionado**

Los trabajos relacionados con el área de ordenamiento de imágenes están muy enfocados a utilizar algún tipo de retroalimentación hecha por el usuario, este enfoque hace uso de información que el usuario proporciona después de haber realizado una consulta y haber obtenido las imágenes recuperadas, con el fin de mejorar el orden de la recuperación. Se puede dividir este enfoque en 3 métodos principales de retroalimentación:

1. Por consulta(*query*). Se selecciona una imagen, del conjunto de imágenes recuperadas, como imagen de consulta o imagen *patrón* para ordenar el resto de imágenes de acuerdo su parecido.
2. Por selección de imágenes relevantes. Se selecciona un conjunto de imágenes relevantes para ordenar el resto de imágenes de acuerdo a las características comunes de las imágenes seleccionadas.
3. Por categorización. Se relacionan imágenes recuperadas con categorías previamente establecidas (de forma dinámica o estática) para clasificar el resto de imágenes en sus categorías correspondientes.

Los métodos antes mencionados utilizan alguna medida de similitud entre ejemplos para encontrar un grado de relevancia entre las imágenes de consulta y las imágenes restantes.

El enfoque automático de selección de imágenes no es muy explotado en el área debido a la poca relación entre la variedad en el grado de relevancia de las imágenes recuperadas con su posición en la lista de recuperación. Tampoco se explotan las ventajas del aprendizaje automático para la clasificación de ejemplos debido al bajo número de ejemplos de aprendizaje y a la representación a veces no conveniente de las imágenes.

### 3. Metodología y Desarrollo

La metodología que se siguió para obtener el ordenamiento de las imágenes está ilustrada en la Figura 1, y se compone de 5 pasos principales:

1. Recuperación de imágenes
2. Obtención de características
3. Generación de vectores de características
4. Construcción-evaluación de clasificadores
5. Predicción de instancias

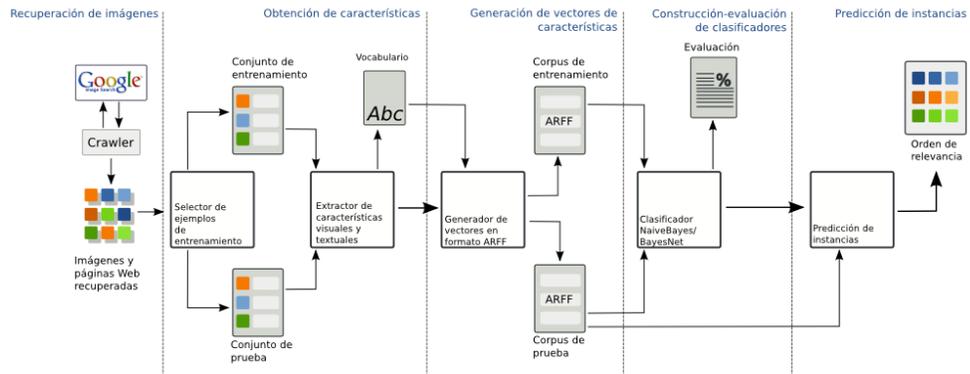
Todo el procedimiento a excepción la primera parte de la etapa de recuperación de imágenes, se realizó fuera de línea con el fin de tener un mejor control sobre las características visuales de las imágenes. Para desarrollar el sistema propuesto se utilizó como plataforma principal y como herramienta para la gestión de las características visuales al entorno de trabajo Matlab<sup>2</sup>, el software de minería de datos WEKA<sup>3</sup> se utilizó para construir los clasificadores y fue necesario crear algunas rutinas en lenguaje Java para realizar las predicciones de las instancias. Cada una de las etapas del proceso se describen en las siguientes sub-secciones.

#### 3.1. Recuperación de imágenes

Para obtener las primeras 100 imágenes y sus respectivos documentos Web, del conjunto de imágenes recuperadas por Google Images, y debido a que no

<sup>2</sup> <http://www.mathworks.com/products/matlab/?BB=1>

<sup>3</sup> <http://www.cs.waikato.ac.nz/ml/weka/>



**Figura 1.** Metodología utilizada para realizar el ordenamiento de las imágenes

existe un conjunto de rutinas para recuperación ofrecidas por Google<sup>4</sup>, se construyó un *crawler* escrito en Perl que permite descargar las primeras 100 imágenes y las páginas Web recuperadas por Google Images. Para iniciar la recuperación de objetos se envía a Google Images un término de búsqueda y se comienza la descarga de cada una de las imágenes y páginas necesarias. Todas las imágenes y sus páginas Web son almacenadas localmente para su posterior procesamiento.

### 3.2. Obtención de características

Una vez que los objetos son almacenados localmente, se dividen en dos conjuntos: uno de entrenamiento y uno de prueba, según el enfoque de selección de ejemplos utilizado.

Si se utiliza el enfoque de selección automática de ejemplos entrenamiento, el conjunto total queda dividido como sigue:

- Entrenamiento: Los ejemplos positivos son las primeras 10 imágenes recuperadas y los negativos las últimas 10.
- Prueba: Las imágenes de la 11 a la 90 son consideradas como conjunto de prueba y son etiquetadas manualmente.

Si por el contrario se utiliza el enfoque de selección de ejemplos de entrenamiento hecha por el usuario, el conjunto total queda dividido como sigue:

- Entrenamiento: Los ejemplos positivos son las 10 imágenes seleccionadas como positivas y los negativos las 10 imágenes seleccionadas como negativas, no importando la posición de las imágenes en el total de imágenes recuperadas.

<sup>4</sup> <http://www.google.com>

- Prueba: Las imágenes que no fueron seleccionadas como positivas o negativas por el usuario son consideradas como conjunto de prueba y son etiquetadas manualmente.

En cualquiera de los dos enfoques, y de manera ideal, ya que a veces las imágenes ya no están disponibles en la Web, se tiene un conjunto de entrenamiento con 20 ejemplos y un conjunto de prueba con 80 ejemplos.

Cuando los 2 conjuntos necesarios están identificados se procede a extraer las características que los describirán, primero se extraen del conjunto de entrenamiento y luego del conjunto de prueba. Para ello se utiliza el siguiente procedimiento:

#### 1. Para cada objeto

- 1) Obtener el histograma de color
  - 1) Descomponer la imagen en sus tres canales de color R, G y B
  - 2) Obtener el histograma de color de cada uno de los canales
  - 3) Concatenar los histogramas para obtener un histograma general
  - 4) Almacenar el histograma
- 2) Obtener la bolsa de palabras
  - 1) Eliminar las etiquetas HTML del documento Web
  - 2) Identificar los lugares de dónde se extraerán las palabras
  - 3) Extraer todas las palabras diferentes de los lugares identificados
  - 4) Almacenar la bolsa de palabras
  - 5) Aumentar el vocabulario

El paso de aumentar el vocabulario no se lleva a cabo cuando se extraen las características del conjunto de prueba. Toda la extracción de características se lleva a cabo utilizando rutinas escritas en el lenguaje de programación de Matlab y rutinas escritas en lenguaje Java.

### 3.3. Generación de vectores de características

Al terminar la etapa de obtención de características se tienen dos conjuntos de atributos y el vocabulario del conjunto de entrenamiento. Para construir los vectores de características que formarán el corpus de entrenamiento y el corpus prueba se realiza el siguiente procedimiento para cada conjunto:

#### 1. Para cada objeto del conjunto

- 1) Almacenar en el vector de características el histograma de color en RGB tomando cada valor como un atributo
- 2) Almacenar en el vector de características un atributo por cada palabra del vocabulario, marcando con 1 aquellas que aparecen en la bolsa de palabras de la imagen y 0 a aquellas que no

El pesado utilizado para los atributos textuales es binario, 1 indica que sí aparece la palabra en la página Web de la imagen y 0 que no. Los corpus de entrenamiento y prueba quedan escritos en formato ARFF (Attribute-Relation File Format)<sup>5</sup>, el cual es interpretado por el software de minería de datos WEKA que será utilizado para procedimientos posteriores.

<sup>5</sup> <http://www.cs.waikato.ac.nz/ml/weka/arff.html>

### 3.4. Construcción-evaluación de clasificadores

Al llegar a este paso ya se tienen construidos (en formato ARFF) los corpus de entrenamiento y de prueba, y con ellos se procede a construir dos clasificadores. El primero es un clasificador Naive Bayes, el cual considera independencia condicional de los atributos dada la clase y que se puede representar como una red bayesiana de estructura simple. El segundo es una red Bayesiana TAN, que es una extensión de una red Bayesiana simple con la consideración de que existen dependencias entre los atributos y que se representa con una estructura de árbol[5]. Para la construcción de los clasificadores se utilizaron rutinas escritas en lenguaje Java que hacen uso de objetos y métodos del conjunto de clases ofrecidas por WEKA.

Para construir cada clasificador se utilizó el mismo corpus de entrenamiento, y se evaluaron los resultados utilizando el mismo corpus de prueba. El modelo construido por cada clasificador fue almacenado para su utilización en la etapa de predicción de instancias.

### 3.5. Predicción de instancias

Una vez que se tiene construido el modelo, ya sea con Naive Bayes o con redes Bayesianas, se utiliza el siguiente procedimiento para obtener las predicciones.

1. Recuperar el modelo correspondiente
2. Para cada instancia en el corpus de prueba
  - 1) Utilizar el modelo para obtener la predicción de la clase
  - 2) Almacenar la predicción

El procedimiento para obtener las predicciones fue escrito en una rutina en lenguaje Java. Una vez que se tienen almacenadas las predicciones, estas son mostradas para visualizar el ordenamiento por relevancia.

## 4. Experimentos y resultados

Los términos utilizados para evaluar el sistema fueron “*table, airplane, bicycle y watch*”, para cada término se recuperaron las primeras 100 imágenes obtenidas por Google Images además de las páginas Web que las contenía. Se eligieron esos términos de búsqueda porque sus resultados ofrecen una gran variedad de características:

1. *table*. Los resultados son ambíguos y hay una cantidad similar de ejemplos relevantes y no relevantes.
2. *airplane*. Las imágenes relevantes son pocas y tienen variedad de presentaciones.
3. *bicycle*. Hay una cantidad similar de imágenes relevantes y no relevantes, y las imágenes relevantes tienen diferentes presentaciones.
4. *watch*. La gran mayoría de las imágenes son relevantes y tienen diferentes presentaciones.

Los términos de búsqueda están en idioma inglés porque existe mas variedad de imágenes en la recuperación en este idioma que en idioma español.

En el Cuadro 1 se muestra un resumen de los corpus de prueba utilizados para evaluar el sistema. Se generaron 2 corpus de prueba para cada término, el primero corresponde al método de selección automática de ejemplos de entrenamiento y el segundo al enfoque de selección de ejemplos de entrenamiento hecha por el usuario. Se puede observar que los totales no siempre suman 80 o no son iguales en los dos experimentos, esto se debe a que no todas las imágenes que Google Images recupera están disponibles en su URL original, porque Google Images utiliza un indización para realizar la recuperación.

**Cuadro 1.** Resumen de los corpus de prueba utilizados para la evaluación, donde X = ejemplos del enfoque automático / Y = ejemplos del enfoque de selección hecha por el usuario

término	ejemplos positivos	ejemplos negativos	totales
table	35/40	45/40	80/80
airplane	21/19	58/59	79/78
bicycle	38/33	45/46	79/79
watch	64/70	15/9	79/79

Los corpus de entrenamiento del método de selección hecha por el usuario constan de 10 ejemplos positivos y 10 ejemplos negativos. Para el caso del enfoque de selección automática el número de ejemplos positivos y negativos varía debido a la no disponibilidad de las imágenes recuperadas.

Se realizaron dos tipos de experimentos, el primero utilizando el método automático de selección de ejemplos de entrenamiento y el segundo utilizando el enfoque de selección hecha por el usuario. Para cada experimento se generó un clasificador Naive Bayes y una red Bayesiana TAN para cada uno de los cuatro términos de búsqueda, y se evaluaron estos clasificadores utilizando los corpus de prueba respectivos. Los resultados de los experimentos se muestran en el Cuadro 2 y el Cuadro 3. Las medidas utilizadas para la evaluación del sistema son el recuerdo y precisión por clase y el porcentaje de instancias clasificadas correctamente. Se utilizaron las medidas de recuerdo y precisión por clase debido a el diferente número de ejemplos positivos y negativos para algunos términos de búsqueda. Se utilizó el porcentaje de instancias clasificadas correctamente para tener una medida global del desempeño del sistema propuesto.

La Figura 2 muestra el ordenamiento obtenido al utilizar el sistema propuesto sobre las imágenes recuperadas con el término de búsqueda *table*, utilizando el método de selección automática de ejemplos de entrenamiento y el clasificador Naive Bayes para el ordenamiento. Con esta configuración se obtuvo el mejor resultado en cuanto el porcentaje de instancias clasificadas correctamente.

El Cuadro 2 muestra los resultados que se obtuvieron al considerar como ejemplos de entrenamiento a las primeras y las últimas 10 imágenes, del total de imágenes, como ejemplos positivos y negativos respectivamente.

**Cuadro 2.** Resultados obtenidos del experimento con selección automática de ejemplos de entrenamiento. Donde  $X = \text{clase relevante} / Y = \text{clase no-relevante}$

término	Naive Bayes			BayesNet		
	precisión	recuerdo	instancias clasificadas correctamente(%)	precisión	recuerdo	instancias clasificadas correctamente(%)
table	0.704/0.698	0.543/0.822	70.00	0.500/0.596	0.400/0.689	56.25
airplane	0.333/0.800	0.619/ 0.552	56.96	0.225/0.692	0.429/ 0.466	45.57
bicycle	0.360/0.463	0.237/0.610	43.04	0.379/0.460	0.289/0.561	43.04
watch	0.840/0.204	0.328/0.733	40.51	0.800/0.167	0.688/0.267	60.76

**Cuadro 3.** Resultados obtenidos del experimento con selección de ejemplos de entrenamiento hecha por el usuario. Donde  $X = \text{clase relevante} / Y = \text{clase no-relevante}$

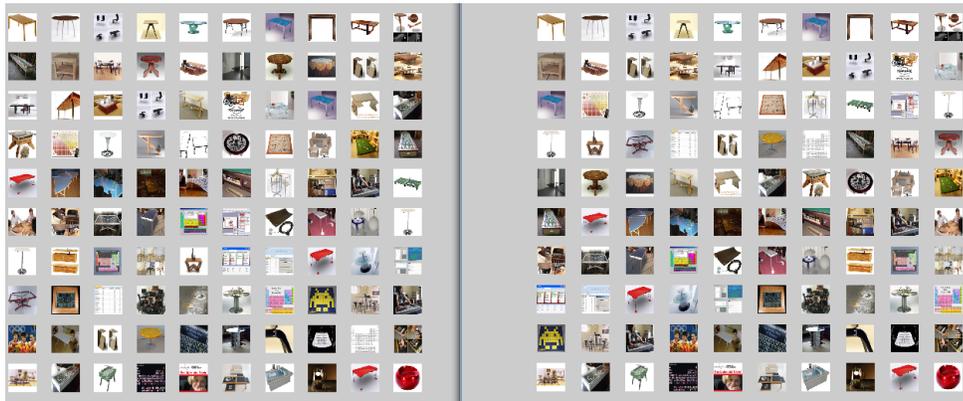
término	Naive Bayes			BayesNet		
	precisión	recuerdo	instancias clasificadas correctamente(%)	precisión	recuerdo	instancias clasificadas correctamente(%)
table	0.640/0.564	0.400/0.775	58.75	0.514/0.511	0.450/0.575	51.25
airplane	0.333/0.789	0.368/0.763	66.67	0.438/0.806	0.368/0.847	73.08
bicycle	0.559/0.689	0.576/0.674	63.29	0.476/0.649	0.606/0.522	55.70
watch	0.923/0.121	0.171/0.889	25.32	0.897/0.125	0.500/0.556	50.63

Se puede observar que para los términos de búsqueda *table* y *watch* este enfoque funciona de manera aceptable. Los resultados altos obtenidos al utilizar este enfoque se deben a que en las 100 imágenes recuperadas están bien definidos los ejemplos relevantes en las primeras 10 posiciones y/o los ejemplos no relevantes en las últimas 10 posiciones, y que los atributos utilizados para describirlos en verdad contribuyeron a diferenciarlos del resto.

Para los términos *table* y *watch* se puede observar que de las imágenes clasificadas como relevantes, un alto número en verdad son relevantes, lo que lleva a un valor de precisión alto, esto se debe a la naturaleza de los corpus de prueba, ya que en ellos existen un número considerable de ejemplos relevantes con características visuales y textuales similares. El término *watch* ofrece un resultado peculiar al utilizar Naive Bayes para el ordenamiento, ya que obtiene una alta precisión y un bajo recuerdo, lo cual se debe a la gran cantidad de ejemplos relevantes que hay en el corpus de prueba y que estos ejemplos son de gran variedad visual y textual.

En general, utilizar el clasificador Naive Bayes con el enfoque de selección de ejemplos automático ofrece mejores resultados en el ordenamiento de las imágenes recuperadas, que utilizar una red Bayesiana TAN.

El Cuadro 3 resume los resultados obtenidos al utilizar el método de selección de ejemplos de entrenamiento hecha por el usuario. Se puede observar que utilizar este método mejora la clasificación de imágenes para los términos *airplane* y *bicycle* ya que para estos dos términos los ejemplos relevantes tienen una gran variedad de presentaciones, y al solicitar que el usuario seleccione solo a una de



**Figura 2.** Comparación entre la salida de Google Images(lado izquierdo) al utilizar el término de búsqueda *table* y el ordenamiento obtenido por el sistema propuesto(lado derecho) al utilizar el método automático de selección de de ejemplos de entrenamiento y el clasificador Naive Bayes

estas presentaciones significa que la clasificación se realizará tomando en cuenta solo a un tipo de imagen relevante.

Para el caso de el término *watch* se obtuvo una disminución considerable en el porcentaje de instancias clasificadas correctamente al utilizar Naive Bayes, esto se debe a que a pesar de que la mayoría de las imágenes clasificadas como relevantes en verdad lo eran, no se pudo recuperar un buen número del total de imágenes relevantes, porque estas contaban con características visuales y/o textuales distintas. La disminución no fue tan notoria al utilizar la red Bayesiana porque para construirla se consideran relaciones entre atributos que evitaron un descenso mayor en la clasificación.

Los resultados mas altos(términos *airplane* y *bicycle*), en comparación con el enfoque automático, se obtuvieron al utilizar una red Bayesiana, debido a que, otra vez mas, las relaciones entre los atributos de las imágenes seleccionadas por el usuario permitieron identificar con mas precisión y recuerdo a las imágenes relevantes y no-relevantes.

## 5. Conclusiones y trabajo futuro

En este trabajo se propuso un sistema de ordenamiento por relevancia de las imágenes recuperadas directamente de Google Images. Para este fin, se propuso representar a las imágenes por su histograma de color en RGB así como por su bolsa de palabras relacionadas y se utilizaron tanto un clasificador Naive Bayes como una red Bayesiana TAN para realizar el ordenamiento de las imágenes en dos clases: *relevante* y *no-relevante*. Los resultados de los experimentos realizados sobre cuatro términos de búsqueda muestran que hay una mejora en el ordena-

miento de las imágenes obtenidas por el sistema propuesto sobre las imágenes obtenidas por Google Images. Es importante concluir que el nivel de mejora en el ordenamiento obtenido, para el método propuesto, está ligado a la naturaleza del término de búsqueda, a la variedad de las imágenes relevantes recuperadas, y al método utilizado para la selección de ejemplos de entrenamiento.

Como trabajo a futuro se propone extender el conjunto de características visuales a características de forma, de textura, de posición, y mas características de color. Si se incluye también, como característica textual, a el fragmento de texto en la página Web que haga mención de la imagen podría mejorar la representación de las imágenes relevantes.

Al utilizar el enfoque automático de selección de ejemplos de entrenamiento propuesto, se obtuvieron buenos resultados para algunos términos, pero para otros una selección basada en una heurística pudo haber ofrecido mejores resultados de acuerdo a la naturaleza de las imágenes relevantes.

Los resultados mostraron que utilizar una misma metodología para todos los términos de búsqueda no siempre es la mejor solución. El utilizar una metodología que varíe algunas de sus etapas según la naturaleza del término de búsqueda permitirá explotar mejor algunas tecnicas de clasificación.

Existen clasificadores que son muy utilizados para tareas específicas de recuperación de textos y de recuperación de imágenes de acuerdo al contenido, es por ello que se propone utilizar un ensamble de clasificadores para realizar el ordenamiento. Este ensamble puede ser secuencial, es decir que la salida de un clasificador sea la entrada de otro, por votación de clasificadores sobre las mismas instancias, o un enfoque en el que un clasificador especializado en texto clasifique las imágenes basándose en las características textuales y un clasificador especializado en imágenes clasifique de acuerdo a las características visuales para después realizar un consenso en cuanto a la clasificación de cada instancia.

## Referencias

- [1] CUI, J., WEN, F., AND TANG, X. Real time google and live image search re-ranking. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia* (New York, NY, USA, 2008), ACM, pp. 729–732.
- [2] DATTA, R., JOSHI, D., LI, J., AND WANG, J. Z. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.* *40*, 2 (2008), 1–60.
- [3] DESELAERS, T., KEYSERS, D., AND NEY, H. Features for image retrieval: An experimental comparison. *Information Retrieval* *11* (03/2008 2008), 77–107.
- [4] DESELAERS, T., PAREDES, R., VIDAL, E., AND NEY, H. Learning weighted distances for relevance feedback in image retrieval. In *International Conference on Pattern Recognition 2008* (Tampa, Florida, USA, 08/12/2008 2008).
- [5] SIERRA, ARAUJO, B. *Aprendizaje Automático : Conceptos básicos y Avanzados*. Pearson, Mexico, 2006.
- [6] ZHU, Q., YEH, M.-C., AND CHENG, K.-T. Multimodal fusion using learned text concepts for image categorization. In *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia* (New York, NY, USA, 2006), ACM, pp. 211–220.