Aprendizaje basado en la interacción de usuarios para búsqueda y recuperación de imágenes

R. Pedraza Jiménez; J. Cid Sueiro; J. Arenas García; A. Guerrero Curieses; C. Tarín Sauer;
H. Molina Bulla; A. Navia Vázquez; y A. Figueiras Vidal
{rpedraza, jcid, jarenas, alicia, tarin, h.molina, navia, arfv}@tsc.uc3m.es
Departamento de Teoría de la Señal y Comunicaciones
Universidad Carlos III de Madrid

Abstract

Se propone, al amparo del proyecto "Nuevos Algoritmos para la Gestión Eficiente de Contenidos Multimedia en Redes de Comunicaciones Móviles" (NAGEC), un nuevo mecanismo para la búsqueda y recuperación de imágenes basado en realimentación de relevancia. La arquitectura propuesta se compone de una red neuronal y un tesauro. La red neuronal extrae de las imágenes dos parámetros: textura y color. El tesauro recoge las relaciones semánticas existentes entre los términos descriptores de las imágenes de la base de datos VisTex. Ambos componentes se relacionan mediante un modelo de realimentación de relevancia que, a través de las interacciones del usuario con el tesauro durante el proceso de búsqueda, permite a la red aprender relaciones semánticas inherentes a las imágenes.

Palabras clave: Recuperación de información, Realimentación de relevancia, Mapa semántico, Red bayesiana.

1. Introducción

La gestión eficiente de contenidos multimedia pasa por resolver, al menos, dos grandes problemas de procesado de señal: parametrización (indexación) de los registros multimedia, y operaciones de categorización, agrupamiento y estimación de relevancia de dichos registros. Como la exigencia de calidad de servicio en estos nuevos escenarios será alta, se propone, al amparo del proyecto "Nuevos Algoritmos para la Gestión Eficiente de Contenidos Multimedia en Redes de Comunicaciones Móviles" (NAGEC[1]) [1], un nuevo mecanismo para la búsqueda y recuperación de imágenes basado en realimentación de relevancia, y generación de mapas de conocimiento obtenidos mediante la interacción con los usuarios.

En los últimos años, los Sistemas de Recuperación de Información Basados en Contenidos (CBIR) han experimentado un enorme desarrollo. Estos sistemas se centran principalmente en la extracción de parámetros como la textura, el color, etc. [2], que pueden obtenerse de forma automática. Sin embargo, existen algunos inconvenientes: (1) baja velocidad del proceso de búsqueda, pues al realizarse ésta exclusivamente con parámetros de bajo nivel difícilmente se alcanza una solución en las primeras iteraciones; (2) bajo nivel de consistencia, a la hora de medir las similitudes, en comparación con la percepción humana; (3) y falta de una manera coherente y completa de expresar las necesidades de información del usuario [3] y de generar patrones de petición.

Las limitaciones de estos sistemas se deben a que utilizan sólo parámetros de bajo nivel. Es necesario por tanto utilizar junto a estos, otros parámetros de mayor nivel, los semánticos, que, aunque más complicados de implementar, expresan de mejor forma la percepción subjetiva del usuario.

Siguiendo esta filosofía, se ha diseñado una arquitectura con capacidad para aprender relaciones semánticas inherentes a las imágenes de una base de datos, mediante el análisis del proceso de búsqueda de los usuarios, que se realizará con la ayuda de un mapa conceptual (un tesauro [4], donde las relaciones jerárquicas, asociativas, y de relación están perfectamente definidas) y un mapa semántico (una red bayesiana [5]) que relaciona los términos entre si para facilitar el aprendizaje de la máquina.

Este artículo presenta un sistema de recuperación de imágenes basado en texturas para visualizar imágenes de una base de datos. Los componentes de salida del sistema incluyen la extracción de características de la textura, parametrización de los registros, el desarrollo de un mapa conceptual (un tesauro de texturas para una rápida búsqueda e indización) y desarrollo de un mapa semántico que permita al sistema aprender las relaciones semánticas inherentes a las imágenes.

2. Descripción Técnica

El funcionamiento del sistema es el siguiente: el usuario realiza una consulta en una base de datos, indexada automáticamente utilizando un selector de características que extrae parámetros de color y textura conforme a MPEG7[6].

Esta consulta se parametriza utilizando un clasificador binario, y se genera una matriz con las características de las imágenes de la base de datos.

[1] Este trabajo ha sido financiado por la CICYT bajo proyecto TIC2002-03713.

Fig. 1: Indexado automático mediante un selector de características (parámetros MPEG-7)

Ambos, matriz y consulta parametrizada, son procesados por un clasificador, que da como salida un vector de decisiones, que determina qué registros son relevantes a la consulta. Las imágenes se ordenan y muestran al usuario, que decidirá si los resultados de la consulta son apropiados (Fig. 2).

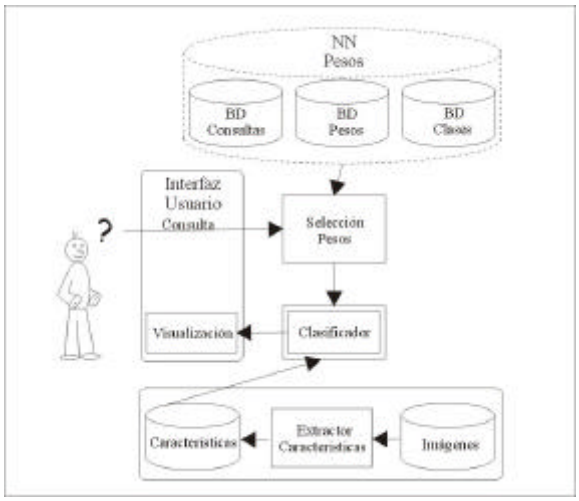

Fig. 2: Sistema de recuperación

En caso negativo, el usuario indicará al sistema cuáles de las imágenes mostradas se ajustan más al resultado esperado (Fig. 3). Esta información la recibe un algoritmo de aprendizaje que, junto con la información que le suministra el clasificador y la base de datos de características, modificará los pesos de la consulta. Este proceso se repetirá hasta que el usuario obtenga los resultados deseados.

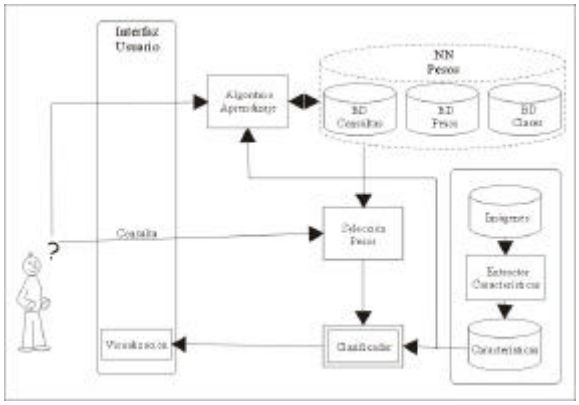

Fig. 3: Realimentación de relevancia

El proceso se verá agilizado por la existencia de un mapa conceptual (un tesauro) que mediante el uso de reglas de asociación permite: (1) ante una nueva consulta mostrar, en la primera iteración, imágenes de todos los grupos semánticos, lo que reduce el número de iteraciones necesarias para alcanzar un resultado; (2) detectar las consultas que guarden relación con otras realizadas anteriormente, lo que acelerará también el proceso de búsqueda.

Los registros analizados se han extraído de *VisTex* [7], una colección de imágenes de texturas creada por el MIT, que se caracteriza por presentar condiciones de iluminación y de perspectiva del mundo real y no en condiciones de laboratorio.

2.1. Modelo de usuario

Aun no se dispone de tecnologías que permitan relacionar eficazmente elementos (patrón de búsqueda y registros) que pertenecen a espacios de representación muy diferentes. Así ocurre en nuestro caso, donde comparamos texto e imágenes. Por ello, se precisa flexibilizar el proceso de búsqueda explotando al máximo la interacción con el usuario.

Este sistema de búsqueda, basado en realimentación de relevancia, permite al usuario un acceso cómodo y flexible a los registros, mediante la interacción con el sistema utilizando un tesauro, único vocabulario permitido por el sistema para realizar las consultas.

2.2. Mapa conceptual

Con la intención de combinar los métodos máquina de clasificación con la información procedente de las interacciones de los usuarios, se ha elaborado un tesauro [8] [9]. Esta herramienta permite generalizar mejor los conceptos de la búsqueda (es decir, acercar al máximo lo que el usuario tiene en mente y lo que se le presenta como resultado), así como optimizar los accesos a la información aumentando la precisión de los resultados. Se compone de una base léxica estructurada en: relaciones jerárquicas (genéricas y específicas), relaciones asociativas, relaciones de equivalencia, y relaciones de definición; que proporcionan al usuario un modo sencillo y coherente de expresar sus necesidades de información.

Se caracteriza porque:
1. Es dependiente de dominio.
2. Proporciona un árbol de indización eficiente, a la vez que mantiene, e incluso mejora, los resultados de la recuperación en términos de percepción humana.
3. La representación de las palabras claves en el tesauro puede ser utilizada como una sencilla información para ayudar a los usuarios a visualizar la base de datos.

Se compone de 406 términos, de los cuales 102 tienen consideración de “descriptores” y 304 de “no

descriptores", todos ellos agrupados en torno a ocho familias semánticas.

La extracción del vocabulario se ha realizado en dos fases: una deductiva y otra inductiva. En la fase deductiva un especialista ha extraído un conjunto de términos a partir de la visualización de las imágenes. En la fase inductiva, se han buscado términos sinónimos y relacionados con los ya obtenidos con la ayuda de distintas fuentes de información.

Para el correcto mantenimiento del tesauro, se está diseñando un sistema que, de forma automática, genere estadísticas sobre la utilización de sus términos. De este modo conoceremos el número de veces que un término es utilizado en la recuperación, el porcentaje que representa ese empleo en el conjunto de vocabulario, y la variación de empleo de dos términos relacionados.

2.3. Parametrización

La parametrización de los registros es una de las piedras angulares de un sistema de gestión de imágenes, y un requisito previo al desarrollo de tareas de alto nivel.

Se ha utilizado una arquitectura neuronal que considera dos parámetros de bajo nivel: las características de la textura, representadas por los histogramas de repuesta a los filtros de Gabor de diferentes frecuencias y direcciones; y un histograma de color basado en el espacio HSV. Se ha entrenado esta red siguiendo un modelo de usuario basado en realimentación de relevancia. A través de las interaciones del usuario, se pretende que la red aprenda relaciones semánticas (parámetros de alto nivel) inherentes a las imágenes de la base de datos (descritas en el tesauro).

2.4. Mapa semántico

Los datos de relevancia constituyen en nuestro caso etiquetas útiles para el aprendizaje supervisado. Con el objeto de aprovechar lo aprendido de unas consultas para responder otras, el sistema que se presenta dispone, a priori, de un mapa semántico que articula las relaciones del vocabulario del tesauro.

El mapa semántico puede definir relaciones del tipo: el término (categoría) C1 es parte de C2; o C2 y C5 son similares. Dicho mapa consiste en una red bayesiana que relaciona términos probabilísticamente.

Asimismo, la relación entre documentos, categorías y parámetros del clasificador se modela probabilísticamente. A título ilustrativo, el modelo mas sencillo se compone de:

x: un documento.
c: la categoría de un documento.
w: los parámetros del clasificador que relaciona x con su probabilidad de pertenecer a la categoría c.

Estas componentes se suponen relacionadas a través de la relación probabilística:

$$p(x, c, w) = p(c \mid x, w)\, p(x)\, p(w)$$

caracterizado por la estructura de red bayesiana de la figura 4.

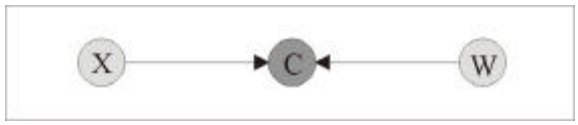

Fig. 4: Modelo bayesiano

No obstante, una vez que el usuario interactúa con el sistema de recuperación, los registros pasan a estar etiquetados. Entonces los componentes del sistema son:

$\mathbf{x_k}$: una colección de documentos.
$\mathbf{c_k}$: las categorías de los documentos $\{x_k\}$.
w: los parámetros del clasificador que relaciona x con su probabilidad de pertenecer a la categoría c.

Y su modelo:

$$p(x, c, x_k, c_k, w) = p(c \mid x, w)\, p(c \mid x_k, w)\, p(x)\, p(x_k)\, p(w)$$

Para simplificar la inferencia, supondremos w determinístico (Fig. 5), dado por:

$$W = \arg\max_w \log p(c_k \mid x_k, w)$$

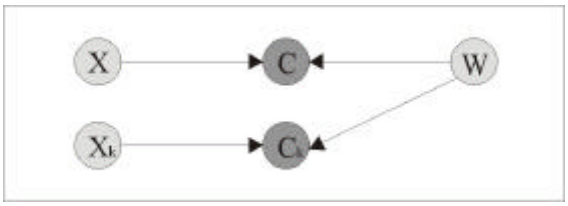

Fig. 5: Modelo bayesiano con registros etiquetados

El modelo para dos categorías (Fig. 6) se compone de los siguientes elementos:

x: un documento.
$\mathbf{c_1}$, $\mathbf{c_2}$: indicadores de pertenencia a las categorías 1 y 2.
$\mathbf{w_1}$, $\mathbf{w_2}$: parámetros (desconocidos) que relacionan un documento con su probabilidad de pertenecer a las categorías 1 y 2.

Y se define como:

$$p(x, c_1, c_2, w_1, w_2) = p(c_1 \mid x, w_1)\, p(c_2 \mid x, c_1, w_2)\, p(x)\, p(w_1)\, p(w_2)\, p(x)$$

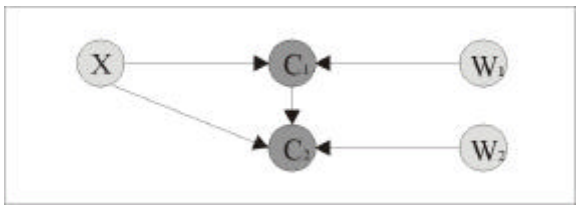

Fig. 6: Modelo para dos categorías

Si aplicamos este modelo con ejemplos etiquetados (Fig. 7) obtenemos:
$p(x, c_1, c_2, w_1, w_2) = p(c_1 \mid x, w_1)\ p(c_2 \mid x, c_1, w_2)\ p(x)\ p(w_1)\ p(w_2)\ p(x)$

Y se demuestra que los documentos etiquetados respecto a c_{1k} sólo afectan a la estimación de w_1, pero no influyen sobre w_2.

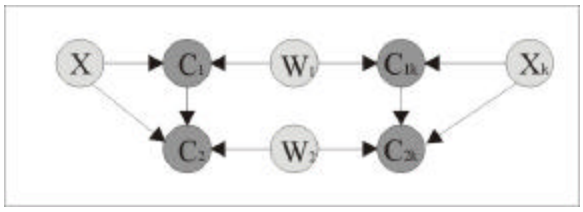

Fig. 7: Modelo con registros etiquetados

Pero ¿que ocurre cuando se aplica este modelo con varias categorías? (Fig. 8). Los componentes serían:

- **x, x_k**: un documento (de test) y una colección de documentos etiquetados en alguna de las dos categorías.
- **c_1, c_{1k}**: variables indicadoras de pertenencia del documento o de los documentos de la colección a la categoría 1.
- **c_2, c_{2k}**: idem.
- **w**: los parámetros del clasificador que relaciona x con su probabilidad de pertenecer a la categoría c.

Este modelo se define como:
$p(x, c, x_k, c_k, w) = p(c \mid x, w)\ p(c \mid x_k, w)\ p(x)\ p(x_k)\ p(w)$

Y el aprendizaje se calcula:

$$W = \arg\max_w \log p(c_k \mid x_k, w)$$

Lo que equivale a entrenamiento por minimización de entropía cruzada.

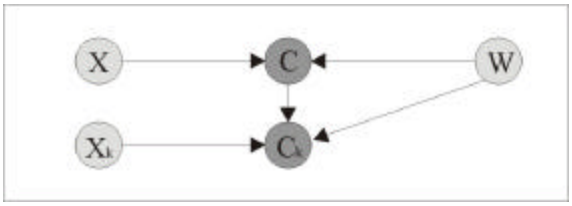

Fig. 8: Modelo para varias categorías

Cada vez que se formula una consulta, todas las NN's relativas a conceptos relacionados intervienen

No obstante este modelo supone que la relación entre categorías es conocida (es decir, se conoce $p(c_2 \mid c_1)$. Cuando esto no es así (Fig. 9), es necesario aprender también esta relación:

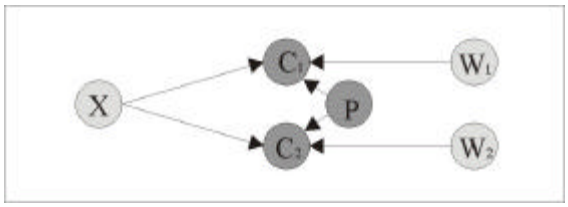

Fig. 9: Modelo para mapa conceptual desconocido.

Por último a esto habría que añadir los registros (Fig. 10):

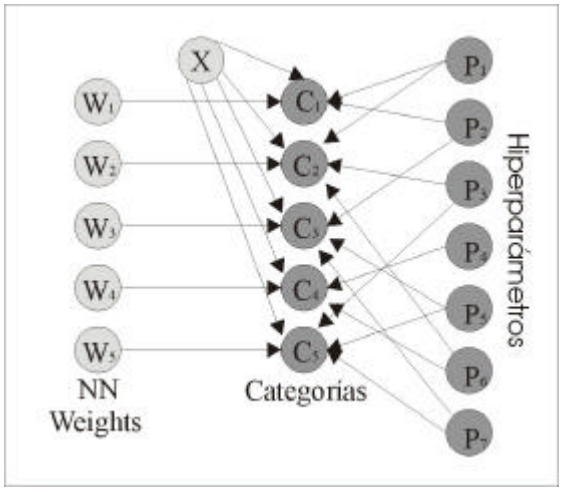

Fig. 10: Modelo general para N categorías.

Con la intención de combinar los métodos máquina de clasificación con la información procedente de las interacciones de los usuarios, se ha elaborado un tesauro [5] [6]. Esta herramienta semántica permite generalizar mejor los conceptos de la búsqueda (es decir, acercar al máximo lo que el usuario tiene en mente y lo que se le presenta como resultado), así como optimizar los accesos a la información aumentando la precisión de los resultados. Se compone de una base léxica estructurada en: relaciones jerárquicas (genéricas y específicas), relaciones asociativas, relaciones de equivalencia, y relaciones de definición; que proporcionan al usuario un modo sencillo y coherente de expresar sus necesidades de información. Además, permite el control de los accidentes lingüísticos (es decir, la superación de los problemas de sinonimia y polisemia) en la recuperación.

Nuestro tesauro se compone de 406 términos, de los cuales 102 tienen consideración de "descriptores" y 304 de "no descriptores", todos ellos agrupados en torno a ocho familias semánticas.

Este tesauro facilita el proceso de indización a la vez que mantiene un buen resultado en la recuperación.

La extracción del vocabulario se ha realizado en dos fases: una deductiva y otra inductiva. En la fase deductiva un especialista ha extraído un conjunto de términos a partir de la visualización de las imágenes. En la fase inductiva, se han buscado términos sinónimos y relacionados con los ya obtenidos con la ayuda de distintas fuentes de información.

Para el correcto mantenimiento del tesauro, se está diseñando un sistema que, de forma automática, genere estadísticas sobre la utilización de sus términos. De este modo conoceremos el número de veces que un término es utilizado en la recuperación, el porcentaje que representa ese empleo en el conjunto de vocabulario, y la variación de empleo de dos términos relacionados.

3. Conclusiones

Se ha propuesto un nuevo mecanismo para la mejora del proceso de búsqueda y recuperación de imágenes. Los primeros ensayos apuntan a que, la combinación del procesado de los parámetros de bajo nivel textura y color, junto con el análisis de la interacción del usuario con el sistema (a través de un tesauro), mejora la eficiencia del acceso a la información: aumenta la velocidad de acceso a la misma, pues el sistema se inicializa con imágenes seleccionadas de los diferentes grupos semánticos que conforman el tesauro; aumenta el nivel de consistencia de la similitud entre la percepción humana y máquina, ya que, mediante la interacción con el usuario la máquina aprende las relaciones semánticas establecidas en el tesauro (fruto de la percepción humana); y se establece un modo coherente de expresar las necesidades de los usuarios mediante una potente herramienta de recuperación de información.

Agradecimientos

Este artículo ha sido parcialmente financiado por el proyecto CICYT TIC 2002-03713.

Referencias

[1] Nuevos Algoritmos para la Gestión Eficiente de Contenidos Multimedia en Redes de Comunicaciones Móviles. Proyecto CICYT TIC 2002-03713. Investigador Principal: Aníbal R. Figueiras Vidal.

[2] H. Müller, W. Müller, and D. McG Squire, "Learning Feature Weights from User Behavior in Content-Based Image Retrieval", Proc. of MDM/KDD2000 Workshop on Multimedia Data Mining, Boston, MA, USA, August 20-23, 2000.

[3] Z. Pecenovic, M. N. Do, S. Ayer and M. Vetterli, "New Methods for Image Retrieval", Proc. of ICPS'98 Congress on Exploring New Tracks in Imaging, Antwerp, Belgium, September 1998, pp. 242-246.

[4] Aitchison, J., & Gilchrist, A. & Bawden, D. (2000). "Thesaurus construction and use: a practical manual" (4th ed.). Chicago, IL, USA: Fitzroy Dearborn.

[5] F.V. Jensen. "Bayesian Networks and Decision Graphs". Springer, New York, 2001.

[6] MPEG: http://www.cselt.it/mpeg
MPEG-7 Industry Forum: http://www.mpeg7.org

[7] Base de datos VisTex: http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html

[8] Wei-Ying Ma and B. S. Manjunath, "A Texture Thesaurus for Browsing Large Aerial Photographs", Journal of the American Society for Information Science, 49(7), 1998, pp. 633-648.

[9] J. Ruiz-del-Solar and M. Jochmann, "On Detemining Human Description of Textures", Proc. of the 12th Scandinavian Conf. on Image Analysis SCIA 2001 , Bergen, Norway, June 2001, pp. 288 294