

MÉTODOS DE EDUCACIÓN EN INGENIERÍA DE SISTEMAS INTELIGENTES

Enrique Calot, Ramón García-Martínez

Programa de Doctorado en Ingeniería, Facultad de Ingeniería, Universidad de Buenos Aires.
Laboratorio de Sistemas Inteligentes, Facultad de Ingeniería, Universidad de Buenos Aires.

ecalot@fi.uba.ar, rgarciamar@fi.uba.ar

CONTEXTO

El Proyecto articula líneas de trabajo del Proyecto “Aplicaciones de Explotación de Información basada en Sistemas Inteligentes”, con financiamiento de la Secretaría de Ciencia y Técnica de la Universidad de Buenos Aires (UBACYT 2008-2010 código I012) y acreditado por Resolución Rector-UBA N° 576/08 con radicación en el Laboratorio de Sistemas Inteligentes de la Facultad de Ingeniería de la Universidad de Buenos Aires.

RESUMEN

Enmarcado dentro del procesamiento de la información y la minería de datos se encuentra el problema del reconocimiento de patrones. Un patrón es un conjunto de características comunes que pueden ser localizadas en una región, admitiéndose un cierto error o tolerancia. Estas regiones pueden ser cadenas de texto, de ADN o imágenes entre otras cosas.

El desarrollo de algoritmos computacionales que permiten buscar sobre estas regiones representa un área en crecimiento. Pese a esto, la incorporación de sistemas inteligentes a la misma todavía requiere mucho desarrollo.

Esta línea de investigación pretende incorporar estas técnicas a algoritmos de este tipo; definir y enmarcar los problemas existentes y buscar soluciones a los mismos dentro de este campo.

Palabras clave: Reconocimiento de patrones, sistemas inteligentes, búsqueda difusa

1. INTRODUCCION

Muchas veces se tiende a modelizar los problemas mediante sistemas matemáticos rígidos y sin error. En algunas áreas, incluso, puede estar mal vista la consideración de un modelo que admita error y luego la búsqueda de su minimización.

Es importante reconocer que todo lo que es procesado por la inteligencia natural admite internamente errores y que la inteligencia artificial, en parte, como su emulación, también lo hace.

La utilización de algoritmos difusos de búsqueda sobre cadenas de texto da de una utilidad inmediata en campos crecientes como lo son la educación del

conocimiento y la biología computacional [Navarro. 2001].

Nuestro interés radica en la creación y luego aplicación de algoritmos inteligentes y/o difusos basados en actuales algoritmos convencionales.

La utilidad de algoritmos de este tipo se ve ejemplificada en los algoritmos de búsqueda sobre cadenas de caracteres. Éstos suelen ser una comparación booleana “si/no”, mientras que se utiliza muy poco las comparaciones por distancia entre el patrón y la posición en el texto a comparar.

Otro ejemplo es la búsqueda sobre imágenes, donde lo que se busca es un patrón común, como puede ser una figura o propiedad, sobre un arreglo de píxeles de 2 y hasta 3 dimensiones.

Existe un enorme número de algoritmos destinados a resolver este tipo de problemas, muchos utilizan distintas nociones de distancias entre el patrón y la región de búsqueda admitiendo errores y basándose su minimización.

La aplicación de estos algoritmos tiene mucha importancia en el área de minería de datos ya que los bancos de datos pueden contener errores humanos – denominados ruido – en la carga y no siempre se los puede asumir correctos. Este ruido no solo puede estar en las relaciones de identificadores y claves, sino también en los campos de texto y/o imágenes. En el proceso de minería de datos muchas veces es necesaria la comparación basada en este tipo de campos y eso conlleva a la creación de algoritmos que sean capaces de calcular distancias entre datos o bien encontrar patrones comunes

2. LINEAS DE INVESTIGACION y DESARROLLO

Dentro de lo que a la minería de datos representa, el tema a investigar es el reconocimiento de patrones difusos o en entornos ruidosos, es decir, donde la clasificación admite un porcentaje de error.

Dentro de este tema se propone la línea de investigación que combina las técnicas utilizadas en sistemas inteligentes con la búsqueda y reconocimiento de patrones en arreglos tanto lineales (cadenas de textos) como multidimensionales (imágenes o volúmenes).

Sobre esta línea se ha publicado un trabajo introductorio previo sobre educación de información

[Calot, *et al.* 2008] mediante *repertory grids* y otro sobre reconocimiento de patrones con redes neuronales *back propagation* aplicados a sistemas médicos –en este caso reconocimiento de lesiones mamarias– [Calot *et al.*, 2009].

3. RESULTADOS OBTENIDOS/ESPERADOS

El proyecto se encuentra en su fase inicial desde marzo de 2009 y prevén las siguientes tareas para ser realizadas en sus primeros dos años:

- Refinamiento e investigación documental orientada a la identificación de trabajos previos vinculados a la utilización de sistemas inteligentes y/o difusos en la detección de patrones.
- Identificación de las posibles futuras líneas de trabajo de los mismos que puedan ser realizadas mediante sistemas inteligentes.
- Identificación de las últimas técnicas utilizadas en el área de la detección de patrones.
- Identificación de casos de estudio aceptados por la comunidad científica en general para ser utilizados en pruebas de concepto y validación del proyecto.
- Identificación y delimitación de problemas vinculados a la búsqueda de patrones mediante estas técnicas
- Exploración y propuesta de nuevos algoritmos que resuelvan estos problemas.
- Validación de pruebas de concepto. Cálculo del orden y estimación de la velocidad de estos algoritmos y contraste de los mismos frente a otros.
- Creación de ejemplos de aplicación que prueben su utilidad concreta.

4. FORMACION DE RECURSOS HUMANOS

En el marco de este proyecto se esta desarrollando una tesis doctoral y tres tesis de ingeniería informática.

5. BIBLIOGRAFIA

- BAEZA-YATES, R., NAVARRO, G. 1996. *A faster algorithm for approximate string matching*. In Dan Hirschberg and Gene Myers, editors, *Combinatorial Pattern Matching (CPM'96)*, LNCS 1075, pages 1-23, Irvine, CA.
- BISHOP, C. M. 2005. *Neural Networks for Pattern Recognition*. Oxford University Press. ISBN 0198538642, 9780198538646.
- CALOT, E., BRITOS, P., GARCÍA-MARTÍNEZ, R. 2008. *Optimizing Relationships Information in Repertory Grids*. En *Artificial Intelligence and Practice II*, Max Bramer Ed. (Boston: Springer), IFIP Series, 276: 163-172. ISBN 978-0387-09694-0.
- CALOT, E., MERLINO, H., RANCAN, C., GARCÍA-MARTÍNEZ, R. 2009. *Tumor Classification on Mammographies Based on BPNN and Sobel Filter*. Serie: *Studies in Computational Intelligence*, Springer-Verlag (en prensa).

LINHARTA, C., SHAMIR, R. 2009. *Faster pattern matching with character classes using prime number encoding*. *Journal of Computer and System Sciences*. Volume 75, Issue 3, May 2009, Pages 155-162.

MYERS, G. 1999. *A fast bit-vector algorithm for approximate string matching based on dynamic programming*. *J. ACM* 46, 3, 395-415. Earlier version in *Proceedings of CPM'98 (LNCS, vol. 1448)*.

NAVARRO, G. 2001. *A guided tour to approximate string matching*. *ACM Computing Surveys (CSUR) archive* 33(1), pp 31-88.

UKKONEN, E., WOOD, D. 1993. *Approximate string matching with suffix automata*. *Algorithmica* 10, 353-364. Preliminary version in Rep. A-1990-4, Dept. of Computer Science, Univ. of Helsinki, Apr. 1990.

ZOBEL, J., DART, P. 1996. *Phonetic string matching: lessons from information retrieval*. In *Proceedings of the 19th ACM International Conference on Information Retrieval (SIGIR '96)*, 166-172.