

Grupo de Reconocimiento de Formas y Tecnología del Lenguaje (PRHLT)

En este artículo se describen las actividades de investigación del grupo de Reconocimiento de Formas y Tecnología del Lenguaje (Pattern Recognition and Human Language Technology- PRHLT) del Instituto Tecnológico de Informática. Se presentan someramente las tecnologías que abarca el Reconocimiento de Formas y las principales aplicaciones de estas metodologías. Finalmente se pasa revista a una serie de aplicaciones recientemente desarrolladas por el grupo PRHLT en sus dos líneas principales de actuación: Procesado de Imágenes y Procesado del Lenguaje.

Introducción

El grupo PRHLT desarrolla sus actividades de investigación y desarrollo en el ámbito del Reconocimiento de Formas (RF).

El RF es una disciplina bien establecida en las ciencias y en las ingenierías desde hace más de 20 años, con aplicaciones fuertemente implantadas en prácticamente todos los sectores productivos. La principal base metodológica del RF es la estadística. Desde este enfoque, entre las principales tecnologías que abarca el RF cabe mencionar los métodos de Extracción, Selección y Transformación de Características, técnicas de Clasificación, tanto Supervisada como No-Supervisada ("Clustering"), métodos de Interpretación, aproximaciones Basadas en Distancias, etc. En este marco, las principales actividades del grupo PRHLT se pueden agrupar en dos líneas:

- Procesado de Imágenes. Técnicas de análisis y reconocimiento de imágenes y visión por computador.

Aplicaciones: OCR, análisis de documentos, identificación de huellas dactilares, sistemas de ayuda a los discapacitados, detección automática de defectos en estampados textiles, identificación de rostros humanos, detección de cáncer prostático, etc.

- Procesado del Lenguaje, Reconocimiento Automático del Habla y Comprensión del Lenguaje. Traducción automática de texto y voz en dominios limitados. Traducción predictiva interactiva.

Aplicaciones: servicios de información telefónica, dispositivos controlados por la voz, etc., traducción de textos técnicos, traducción simultánea (voz) en servicios de hotel, etc. Sistemas de ayuda a la traducción de calidad de textos generales.

El PRHLT mantiene una colaboración estrecha con diversos grupos de investigación y empresas de España y otros países de la Unión Europea en las áreas mencionadas arriba. Asimismo, el PRHLT participa activamente en diversos proyectos y contratos de I+D en estas áreas.

En las siguientes secciones se presentarán brevemente algunas de las aplicaciones más interesantes recientemente desarrolladas por el PRHLT. Para mayor información se puede visitar <http://prhlt.iti.es>, y en particular, <http://prhlt.iti.es/demos/demos.htm> y <http://prhlt.iti.es/proyectosGrupo.htm>.

Procesado de Imágenes

Reconocimiento Automático de Huellas Digitales

La identificación basada en huellas digitales es la más antigua y la más utilizada en muchas aplicaciones de identificación de individuos. El carácter individual de la huella digital puede ser representado por un patrón de valles y crestas así como por unos puntos característicos denominados "minucias", que son los finales y las bifurcaciones de las crestas.

Las técnicas de comparación de huellas digitales pueden ser clasificadas en dos categorías: basadas en minucias y basadas en correlaciones. Las basadas en minucias primero encuentran las minucias para posteriormente establecer sus posiciones relativas dentro de la huella. Las técnicas basadas en correlaciones requieren la localización precisa de un punto de referencia y están afectadas por translaciones y rotaciones de la imagen. La primera técnica es más robusta a las translaciones y rotaciones de la imagen de la huella pero suele ser muy difícil conseguir una buena detección de las minucias a partir de imágenes ruidosas. Para ello en nuestro sistema de Identificación de Huellas Dactilares se aplica una batería de técnicas de preproceso de imagen y se obtiene un conjunto aceptable de minucias. Una vez las minucias han sido detectadas se necesita una medida de similitud entre dos huellas basada en sus respectivos conjuntos de minucias. La invarianza a la translación, escalado y rotación se consigue aplicando todas estas transformaciones al conjunto de minucias asociado. El mejor emparejamiento de ambos conjuntos es seleccionado y la similitud de ambas huellas viene dada en términos de similitud de ambos conjuntos de minucias.

Reconocimiento de Rostros

Gracias al uso de las técnicas de preproceso más recientes y al empleo de un innovador esquema de representación de características locales, se han desarrollado diversas aplicaciones para la identificación de rostros humanos.

La tarea de reconocimiento de rostros se divide en tres etapas bien diferenciadas: preproceso de la imagen original, extracción de vectores de características y clasificación.

El primer objetivo del preproceso es la selección del conjunto de píxeles pertenecientes a las regiones de la cara con mayor información discriminativa. Esto se consigue mediante el cálculo de la varianza local en una pequeña ventana para cada uno de los píxeles y la selección de aquellos que superan un umbral global.

En la siguiente etapa se implementa una técnica novedosa de extracción de características. Una imagen se representa mediante un conjunto numeroso de vectores de características correspondientes a los píxeles seleccionados en la etapa de preproceso. Cada píxel seleccionado genera un vector de valores de gris a partir de la concatenación de las filas de una pequeña ventana centrada en el mismo. Los valores de gris pertenecen a la imagen que se obtiene al aplicar la operación de gradiente sobre la original. Evidentemente, todos los vectores extraídos de la misma imagen son etiquetados con el mismo identificador. Los vectores así obtenidos son sometidos a un proceso de reducción de dimensionalidad mediante la conocida técnica de Análisis de Componentes Principales (PCA), con lo que se consigue una representación más compacta de la información contenida en cada ventana procesada.

Finalmente, se utiliza un proceso de clasificación que podría ser incluido en la familia de técnicas conocida como "direct voting scheme". Para ello, todos los vectores de características extraídos de una imagen de "test" son clasificados mediante la regla de los k-Vecinos Más Próximos. La clase ganadora de cada vector clasificado incrementa la cuenta general de "votos" para esa clase, de manera que aquella clase que finalmente más votos acumula tras la votación de todos los vectores pertenecientes a la misma imagen de "test" es la proporcionada como resultado de la clasificación de la imagen actual.

Reconocimiento Óptico de Formularios Impresos y Manuscritos

Nuestro sistema de reconocimiento de formularios impresos emplea algoritmos avanzados de OCR (reconocimiento óptico de caracteres) para la extracción de información alfanumérica de los campos de un formulario. En este caso, los algoritmos de OCR han sido entrenados únicamente con caracteres impresos, pero el reconocimiento de caracteres manuscritos también es posible.

Nuestros sistemas de OCR para caracteres manuscritos aislados extraen automáticamente estos caracteres (alfabéticos o numéricos) de los campos manuscritos de los formularios. El uso de modelos particularizados, aprendidos automáticamente a partir de muestras, permite que el sistema pueda trabajar con cualquier lengua y cualquier tipo de alfabeto. Además, se puede aplicar un postproceso lingüístico a los resultados del reconocimiento para aumentar la precisión cuando el texto a reconocer presenta restricciones léxicas o sintácticas conocidas.

Reconocimiento de Texto Manuscrito

El "Reconocimiento de Texto Manuscrito Continuo" (RTMC) está demostrando ser una tarea de gran desafío en Reconocimiento de

Formas. Aunque el texto está básicamente compuesto de caracteres, las aproximaciones tradicionales de reconocimiento de caracteres aislados (OCR) generalmente fracasan en la tarea de RTMC. Sin lugar a dudas, esto ocurre a causa de la imposibilidad material de segmentar de manera fiable un texto continuo en sus caracteres individuales. Sin embargo, los seres humanos realizan estas tareas de segmentación y reconocimiento de una manera natural y sin aparente esfuerzo. La precisión se alcanza gracias a un "reconocimiento retardado" hasta tener suficiente información para satisfacer los niveles más altos de percepción y lograr así una comprensión de (parte de) lo escrito. Como "subproducto" se consigue reconocer las palabras constituyentes, los caracteres que las componen y la correspondiente segmentación implícita. Parece claro que esta inherente habilidad humana viene dada por una fuerte inter-cooperación entre diferentes niveles de conocimiento: morfológico, léxico, sintáctico y semántico.

Esta es la misma situación que aparece en el campo del Reconocimiento del Habla. En este campo, las técnicas existentes de mayor éxito están basadas en la inter-cooperación de las mencionadas fuentes de conocimiento para conseguir un reconocimiento global. Por esta razón, el sistema aquí presentado se basa en adaptar tecnologías propias del RH para su uso en RTMC.

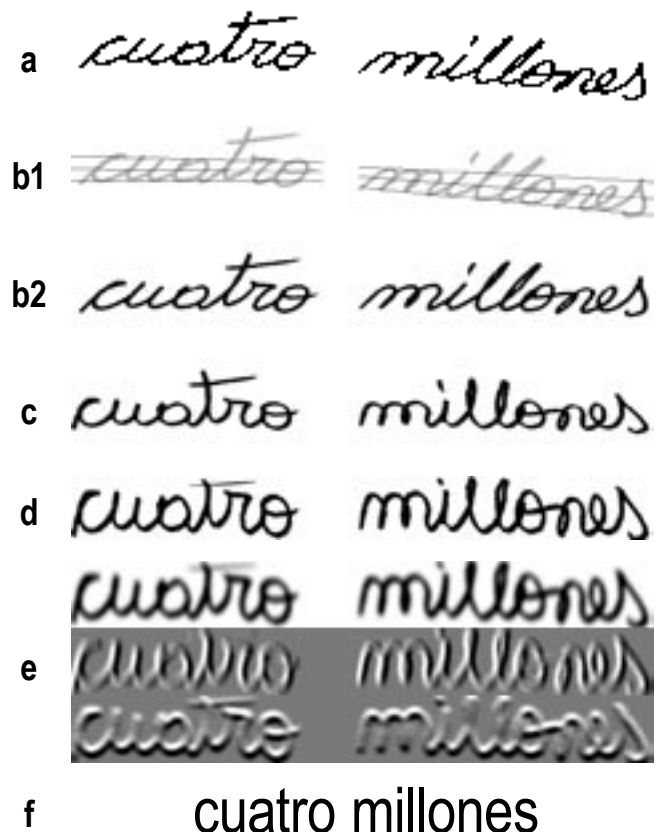


Figura 1: Proceso de reconocimiento de texto manuscrito.

La figura 1 describe el proceso de reconocimiento completo. Los paneles a-d) muestran las sucesivas etapas de preproceso llevadas a cabo con imagen original para la normalización de los atributos de estilo de la escritura. El panel e) es la representación final de la señal en forma de secuencia de vectores de niveles de gris y gradientes horizontales y verticales. Esta representación es directamente

procesada por el reconocedor global, el cual ofrece la hipótesis de reconocimiento (perfectamente correcta, en este caso) que se muestra en el panel f).

Procesado del lenguaje

Comprensión del habla

En un planteamiento tradicional, la comprensión del habla se realiza en dos etapas: una fase de reconocimiento de la elocución pronunciada y una fase de comprensión propiamente dicha de la frase reconocida. El sistema desarrollado en el grupo PRHLT adopta una aproximación alternativa en la que se integran estas dos etapas.

una figura determinada está presente en el área de dibujo, o borrar determinadas figuras. La figura 2 muestra un panel con el estado de este prototipo cuando ha recibido la orden "Se añade un círculo grande y oscuro encima del triángulo mediano". A partir de la señal vocal resultante de la pronunciación de esta frase (panel central), el sistema obtiene directamente la fórmula lógica "La(x) & Da(x) & C(x) & M(z) & T(z) & A(x,z) & Ad(x)"; es decir, x es "Large", "Dark", "Circle" y "Medium"; z es "Medium" y "Triangle"; x esta(rá) "Above" z; x se debe añadir ("Add"). Como subproducto, también se obtiene la frase reconocida que, en este caso, coincide exactamente con la que se había pronunciado.



Figura 2: Panel del prototipo MLA.

Esto se hace mediante un transductor de estados finitos que actúa como modelo de lenguaje para el reconocimiento del habla y, simultáneamente, obtiene una forma lógica asociada a la frase pronunciada. Esta forma lógica refleja el significado de la frase. Los sistemas de estas características pueden llevar a cabo con éxito la comprensión del habla en tareas de dominio restringido. Siguiendo estas ideas hemos implementado un prototipo para una tarea denominada Miniature Language Acquisition (MLA). En esta tarea se permite al usuario dibujar círculos, triángulos o cuadrados en un área de dibujo mediante órdenes orales en lenguaje natural. El sistema permite especificar el tamaño y color de las figuras y su posición relativa a otras figuras. También se puede consultar si

Centralita Automática Dirigida por Voz

La tarea básica de una operadora de centralita telefónica es responder a una llamada telefónica en la cual se demandan algunos servicios, como por ejemplo conectar con una extensión de teléfono local. El principal objetivo al implementar una centralita automática dirigida por voz es gestionar los servicios que una operadora humana puede ofrecer. El estado del arte de las técnicas de Reconocimiento Automático del Habla y Comprensión del Lenguaje permite implementar una centralita automática dirigida por voz. Cuando se trabaja en dominios limitados, un sistema simple puede complementar

de una manera eficiente el trabajo de una operadora humana. Nuestra centralita automática dirigida por la voz es un sistema de reconocimiento de voz continua. Esta característica permite al usuario hablar utilizando frases en lenguaje natural. Además, está basado en una plataforma hardware de bajo coste. El sistema ofrece dos servicios: conectar con el correspondiente número de teléfono de personas de la organización y proporcionar su número de teléfono. Si el usuario no da suficiente información, el sistema automáticamente le solicita datos más específicos. El escenario típico es el siguiente: alguien llama a la centralita automática; esta procesa la entrada oral, decodifica su significado y lleva a cabo la operación solicitada. La centralita telefónica tiene que devolver los resultados de la petición al usuario o recabar más información de este. Es posible que exista más de una persona en la organización con el nombre inicialmente indicado por el usuario. Entonces, el sistema pregunta al usuario por un nombre más específico. Finalmente, el sistema pasa la llamada a la persona solicitada.

Traducción del habla

Hoy en día, los sistemas de reconocimiento automático del habla con mayor éxito se basan en redes de estados finitos estocásticas. La traducción del habla puede abordarse de una manera similar al reconocimiento del habla. Los transductores de estados finitos estocásticos, que son una particularización de las redes de estados finitos estocásticas, han demostrado ser muy adecuados como modelos de traducción. Los sistemas de traducción del habla más comunes en la actualidad se basan en una arquitectura serie, compuesta por un sistema de reconocimiento del habla y seguidamente por un sistema de traducción de texto, lingüístico o estadístico. El uso de redes de estados finitos estocásticas es realmente adecuado para arquitecturas totalmente integradas, donde los modelos acústicos se integran en el modelo de traducción de manera similar a como se hace en reconocimiento de habla. Esta es una de las características innovadoras de nuestro sistema. Uno de los objetivos principales del proyecto EuTrans fue el desarrollo de sistemas de traducción para tareas de dominio restringido con entrada oral.

Traducción predictiva interactiva

La tecnología disponible en la actualidad solo permite abordar adecuadamente aplicaciones de traducción automática en dominios limitados. Cuando el dominio de aplicación se generaliza (o incluso se extiende a una lengua completa), la traducción completamente

automática y de calidad no parece posible de momento. Sin embargo sí es posible desarrollar sistemas útiles de traducción asistida. Entre los paradigmas bajo los que se pueden desarrollar sistemas de este tipo se encuentra la Traducción Predictiva Interactiva (TPI). Bajo esta denominación se engloban técnicas de traducción asistida en las que el sistema trata de predecir lo que un traductor humano escribiría. Para ello se usan dos fuentes de información: una frase o párrafo en el lenguaje origen que se desea traducir y algún fragmento de la traducción de esta frase a la lengua destino, que ha sido ya validada por el traductor humano. Con estas informaciones, el sistema predice (una parte de) el resto de la traducción, la cual es nuevamente validada y/o corregida por el traductor humano antes de que sea re-utilizada por el sistema para realizar nuevas y más precisas predicciones. En el ITI se está actualmente llevando a cabo un importante proyecto sobre TPI: el TransType2 (TT2). En TT2 se pretende desarrollar sistemas de traducción asistida que permitan hacer frente a la creciente demanda de traducciones de alta calidad. Se desarrollarán seis versiones diferentes del sistema para la traducción entre inglés y francés, castellano y alemán. La solución propuesta por TT2 se basa en la incorporación de un motor de Traducción Automática dentro de un entorno de TPI. De esta manera, el sistema combina las ventajas de dos paradigmas: por una parte la traducción asistida, en la cual el traductor humano asegura una salida de alta calidad. Por otra parte, la traducción automática, en la que la máquina proporciona la eficiencia necesaria para lograr un aumento de productividad. Para la implementación del motor de traducción empleado en TT2, se utilizan transductores estocásticos de estados finitos, que han demostrado su adecuación para traducción automática en aplicaciones de dominio limitado. Son interesantes por su simplicidad y por la posibilidad de inferir modelos automáticamente a partir de corpus de entrenamiento bilingües. Permiten una búsqueda muy eficiente sobre nuevos datos de test tanto en modo completamente automático como predictivo. Además, se pueden emplear técnicas híbridas de estados finitos y traducción estadística para producir transductores más precisos. En este caso el aprendizaje se basa en utilizar pares de entrenamiento alineados a nivel de palabra mediante técnicas estadísticas.

Autor: Enrique Vidal

Para más información sobre el Grupo de Reconocimiento de Formas y Tecnología del Lenguaje:
grftl@iti.upv.es