
RESUMEN

Las interfaces persona-máquina de entrada de datos, independientemente de cuál sea su naturaleza: texto manuscrito, voz, teclado, etc., están sujetas a una gran variedad de errores. La combinación de determinadas fuentes de información, producida durante el propio proceso de introducción de datos, puede contribuir de manera significativa en la reducción de dichos errores. En la tesis que aquí se presenta, dicha fusión de la información se va a abordar mediante el uso de autómatas transductores de estados finitos (WFST) y la operación de composición. Esta metodología permite el modelado de cada una de las fuentes de información de manera independiente. El resultado final de la composición de éstos autómatas genera un único autómata que integra todas las fuentes de información previamente modeladas. Como ejemplo de uso de la metodología propuesta, se presentan diferentes aplicaciones relativas a la corrección simbólica en interfaces persona-máquina, tanto desde un enfoque automático como desde el enfoque de la ayuda a la interacción. Los resultados se presentan bajo la perspectiva de un postproceso de cadenas provenientes de un proceso de digitalización de formularios.

Por otra parte, disponer de la posibilidad de discernir las cadenas correctamente postprocesadas de las incorrectas es un tema de enorme importancia. Es por ello que se plantea un método general de estimación del error frente a un coste de transformación de las cadenas de entrada que permite establecer un umbral dinámico en función de dicho coste y un parámetro propuesto por el usuario final: el error final asumible.

En esta tesis se presenta inicialmente una aplicación real de corrección de cadenas procedentes de un clasificador OCR en una tarea de digitalización de formularios. Estas cadenas, proceden de un clasificador con cierta probabilidad de error, lo que implica la posibilidad de que alguno de los caracteres pertenecientes a una palabra sea erróneo, produciendo finalmente palabras incorrectas. Esto plantea la necesidad de introducir algún tipo de postproceso que mejore dichas cadenas. Para implementar dicho postproceso, se tienen en cuenta todas las evidencias disponibles en un momento dado. En el caso propuesto aquí serán los caracteres reconocidos por el propio clasificador con su probabilidad a posteriori, la matriz de confusión entre símbolos y el modelo de lenguaje finalmente aceptado. Cada una de estas evidencias

se modela de manera independiente en forma de un WFST. Una vez modeladas se fusionan mediante la operación de composición de autómatas en un único autómata integrado. A partir de este autómata, se selecciona el camino que maximiza la probabilidad y que corresponde con la cadena perteneciente al lenguaje más cercana a la hipótesis OCR según la matriz de confusión entre símbolos. El sistema final ofrecerá dos resultados diferentes: por una parte la cadena corregida y por otra el coste de transformación de dicha corrección.

Dado que el postproceso descrito está sujeto a la generación de posibles transcripciones erróneas, en muchos casos puede ser conveniente acotar el error final producido. Ello conduce a la aplicación de políticas de rechazo de hipótesis en base a un umbral de confianza (o, equivalentemente, un umbral sobre el coste de transformación). En esta tesis se presenta un método adaptativo de estimación del umbral de rechazo que permite estimarlo para obtener un determinado porcentaje de error en un lote de cadenas de un lenguaje (muestra) que presenta diversas ventajas. Por un lado, es independiente de la distribución de los costes de transformación de dichas muestras. Por otro lado, permite al usuario establecer el umbral de una manera familiar y ventajosa, como es fijando la tasa de error deseada de la muestra. Para todo ello, en primer lugar, y para un lenguaje dado, se define un modelo que estima la probabilidad de error asociada a aceptar cadenas con un coste de transformación determinado. A continuación, se expone el procedimiento que lleva a cabo la estimación del umbral de rechazo de manera adaptativa con el objetivo de alcanzar la tasa de error predefinida para un lote de test. Además, se propone una aproximación para la obtención del modelo anterior cuando no se dispone de hipótesis OCR reales y supervisadas en la etapa de aprendizaje. El capítulo se acompaña de experimentos cuyos resultados demuestran la utilidad del método propuesto.

Seguidamente y enlazando en cierta forma con la búsqueda de un incremento de productividad en una posible validación de las cadenas, previamente rechazadas por el sistema a través del método de estimación del error anteriormente expuesto, se presenta un método de interacción persona-máquina multimodal e interactivo que fusiona la información anterior junto al prefijo introducido, por el propio usuario, durante dicho proceso de validación, haciendo uso para ello de los WFST y la operación de composición de autómatas. La búsqueda de la cadena más probable, para cada nueva interacción ofrecida por el usuario, en el autómata compuesto que aquí se expone, muestra un claro incremento de la productividad, al requerirle un número menor de pulsaciones de teclado en la obtención de la cadena correcta. Para finalizar, se muestra otra interfaz multimodal e interactiva tolerante a fallos, mediante la fusión de diferentes fuentes de información junto a un modelo de error relacionado con las posibles confusiones producidas debido a la disposición de las teclas de un teclado. Para ello, se hace uso también de WFST para su modelado. La aplicación mostrada en este caso está relacionada con la introducción de un destino en un dispositivo GPS y en ella se considera, tanto la información de los destinos próximos a un lugar concreto, como la información relativa al prefijo introducido y los errores que pueden aparecer debido a la propia disposición de las teclas en el dispositivo de entrada.