

RESUMEN

En los últimos años los modelos ocultos de Markov (HMMs) han tenido un papel destacado en el campo del reconocimiento de texto manuscrito off-line (HTR). Al igual que ocurre en el reconocimiento automático del habla (ASR), los HMMs se utilizan para modelar la probabilidad de una secuencia de observaciones dada su transcripción. En cambio, al contrario de lo que ocurre en ASR, en HTR no hay un conjunto de características estándar que esté siendo utilizado por la mayoría de sistemas existentes. En esta tesis proponemos utilizar directamente como características los píxeles originales binarizados, conjuntamente con modelos específicamente diseñados para tratar con datos binarios. Concretamente, proponemos el uso de Bernoulli HMMs (BHMMs), es decir, HMMs convencionales donde las distribuciones de (mixturas de) gaussianas son reemplazadas por (mixturas de) funciones de probabilidad de Bernoulli. El objetivo es doble: por un lado esto nos permite modelizar la naturaleza binaria de las imágenes con texto (letras/fuentes) utilizando BHMMs. Por otro lado, esto garantiza que no se perderá información discriminante en el proceso de extracción de características (la mayoría de las bases de datos de HTR disponibles pueden ser fácilmente binarizadas sin una pérdida relevante de información).

En esta tesis, toda la teoría de HMMs necesaria para desarrollar un toolkit basado en HMMs es revisada y adaptada para el caso de los BHMMs. Concretamente, empezamos definiendo un clasificador sencillo basado en BHMMs con funciones de probabilidad de Bernoulli en los estados, y acabamos con un reconocedor de HTR continuo con mixturas de Bernoulli embebidas. Respecto a las características, proponemos un proceso sencillo de extracción de características binarias con poca pérdida de información. Todas las imágenes son escaladas y binarizadas para que se puedan reinterpretar fácilmente como secuencias de vectores de características binarias. Se proponen dos extensiones a este método básico de extracción de características: el uso de una ventana deslizante para capturar mejor el contexto, y un método de reposicionamiento para tratar mejor las distorsiones verticales. Se han obtenido resultados competitivos cuando los BHMMs y los métodos propuestos se han aplicado a bases de datos conocidas en el ámbito del HTR. En particular, hemos quedado primeros en la competición *Arabic Handwriting Recognition Competition* organizada en la *12th International Conference on Frontiers in Handwriting Recognition (ICFHR 2010)*, y en

la competición *Arabic Recognition Competition: Multi-font Multi-size Digitally Represented Text* organizada en la *11th International Conference on Document Analysis and Recognition (ICDAR 2011)*.

En la parte final de esta tesis proponemos un método para entrenar clasificadores basados en BHMMs haciendo uso de criterios de entrenamiento discriminativos, en vez del criterio Maximum Likelihood Estimation (MLE). Concretamente, empezamos proponiendo un clasificador log-lineal para datos binarios basado en un clasificador de BHMMs. La estimación de los parámetros de este modelo se puede hacer utilizando cualquier criterio de entrenamiento discriminativo para modelos log-lineales. En particular, presentamos las fórmulas para diversos criterios basados en el criterio MMI. Para acabar, demostramos la equivalencia entre ambos clasificadores, y por lo tanto, demostramos que el entrenamiento discriminativo de un clasificador de BHMMs puede llevarse a cabo obteniendo el modelo log-lineal equivalente. Los resultados reportados demuestran claramente que los BHMMs discriminativos mejoran a los BHMMs convencionales.