# RESUMEN

Este trabajo se centra en problemas que cumplen dos propiedades: En primer lugar, problemas que pueden representarse (al menos aproximadamente) en términos de secuencias unidimensionales. En segundo lugar, la resolución de estos problemas implica la descomposición de la secuencia observada en segmentos que se pueden clasificar en un conjunto finito de unidades. Las tareas de segmentación y de clasificación necesarias para obtener las secuencias de unidades que mejor explican la señal observada están tan intrínsecamente interrelacionadas (hecho conocido como "La paradoja de Sayre") que deben realizarse de manera conjunta. El reconocimiento automático del habla (ASR) y de la escritura manuscrita (HTR) son ejemplos de este tipo de tareas.

Para poder realizar aportaciones en un campo de investigación tan maduro como éste nos hemos inspirado por lo que algunos autores denominan "La trilogía exitosa", término que expresa la sinergia obtenida cuando se tienen en cuenta estos tres puntos:

1. un buen formalismo, que dé lugar a buenos algoritmos;
2. un diseño e implementación ingeniosos y eficientes, que saquen provecho de las características del hardware;
3. no descuidar el "saber hacer" de la tarea, un buen preproceso y el ajuste adecuado de los diversos parámetros.

Describimos y estudiamos "modelos generativos en dos etapas" sin reordenamientos (TSGMs). Estos modelos incluyen no sólo los ampliamente utilizados modelos ocultos de Markov (HMM), sino también los modelos segmentales (SMs).

Es fácil obtener un decodificador de "dos pasos" simplemente considerando a la inversa un TSGM introduciendo no determinismo cuando sea necesario: en primer lugar, se genera un grafo acíclico dirigido (DAG) que es utilizado, en un segundo paso, y conjuntamente con un modelo de lenguaje (LM), para obtener las secuencias de unidades que mejor explican la señal observada. El ampliamente conocido decodificador de "un paso" (adecuado para el caso habitual en que se combinan HMMs con LMs regulares tales como n-gramas) es un caso particular, con lo que se presenta una comparativa entre ambos decodificadores, así como con aproximaciones alternativas.

Se describe una formalización del proceso de decodificación basada en ecuaciones de lenguajes y semianillos. En ella se propone el uso de redes de transición recurrente (RTNs) como forma normal de gramáticas de contexto libre (CFGs) y se utiliza el paradigma de análisis por medio de composición/intersección de manera que el análisis de lenguajes incontextuales queda como una ligera extensión del análisis de lenguajes regulares. Este análisis también nos ha llevado a un algoritmo de composición de transductores que permite el uso de RTNs y que no necesita recurrir al concepto de composición de filtros incluso en presencia de transiciones nulas y con semianillos no idempotentes.

En relación a los LMs, se propone una extensa revisión y algunas contribuciones menores mayormente relacionadas con su interfaz, con la representación de algunos LMs y con la evaluación de LMs basados en redes neuronales (NNLMs).

También se ha realizado una revisión de SMs que incluye SMs basados en la combinación de modelos generativos y discriminativos, así como un esquema general de tipos de emisión de tramas y de SMs.

Se proponen varias versiones especializadas del algoritmo de Viterbi para modelos de léxico que sacan provecho de determinadas topologías de los HMMs. Estos algoritmos permiten estados activos sin, por tanto, tener que recurrir a estructuras de datos de tipo diccionario (como tablas de dispersión) y son capaces de sacar provecho de la *caché* por el tipo de acceso a los datos.

Se ha diseñado e implementado una arquitectura de flujo de datos o "*dataflow*" para obtener de modo muy flexible diversos tipos de reconocedor a partir de un pequeño conjunto de piezas básicas. Sus componentes se pueden describir de un modo "reactivo" y un novedoso protocolo de serialización de DAGs permite crear generadores de DAGs capaces de emitirlos a medida que se generan, así como decodificadores que pueden procesarlos mientras se van recibiendo.

Describimos generadores de DAGs que pueden tener en cuenta restricciones sobre la segmentación, utilizar modelos segmentales no limitados a HMMs, hacer uso de los decodificadores especializados propuestos en este trabajo y utilizar un transductor de control que permite el uso de unidades dependientes del contexto.

Los decodificadores de DAGs hacen uso de un interfaz bastante general de LMs que ha sido extendido para permitir el uso de RTNs.

Otro tipo de decodificador ampliamente utilizado es el conocido como "un paso", que resulta muy adecuado cuando nos limitamos a HMMs y a LMs de estados finitos (como los n-gramas). Se proponen también mejoras para este tipo de decodificador combinando las características de los algoritmos especializados para léxicos y la interfaz de LM en modo "*bunch*" de modo que obtenemos un decodificador de tipo un paso que evita cualquier tipo de búsqueda explícita de tipo diccionario ("*hashing*" o similar) y que permite conseguir una buena paralelización.

La parte experimental de este trabajo está muy centrada en reconocimiento de escritura en diversas modalidades de adquisición (*offline*, *bimodal*). Hemos propuesto técnicas novedosas para el preproceso de escritura manuscrita *offline* que evita el uso de heurísticos geométricos prefiriendo utilizar, en su lugar, técnicas de aprendizaje automático (redes neuronales). Con este preproceso y utilizando HMMs hibridados con redes neuronales hemos conseguido, para la base de datos IAM, algunos de los mejores resultados publicados. Entre los experimentos relacionados con este trabajo podemos mencionar el uso de información de sobre-segmentación, aproximaciones sin restricción de un léxico, experimentos con datos bimodales o la combinación de HMMs híbridos con reconocedores de tipo holístico.