Resumen

El problema conocido como *de secuencia a secuencia* consiste en transformar una secuencia de entrada en una secuencia de salida. Bajo esta perspectiva se puede atacar una amplia cantidad de problemas, entre los cuales destacan la traducción automática, el reconocimiento automático del habla o la descripción automática de objetos multimedia. En los últimos años, la aplicación de redes neuronales profundas ha revolucionado esta disciplina, y se han logrado avances notables. Sin embargo, y a pesar de estas mejoras, los sistemas automáticos todavía producen predicciones que distan mucho de ser perfectas. Para obtener predicciones de gran calidad, los sistemas automáticos se utilizan con la supervisión de un ser humano, quien corrige los errores. Esta forma de trabajar es muy común en la industria de la traducción. Esta tesis se centra principalmente en el problema de la traducción del lenguaje natural, el cual se ataca usando modelos enteramente neuronales. Nuestro objetivo principal es desarrollar sistemas de traducción neuronal más eficientes. Para ello, nuestras contribuciones se asientan sobre dos pilares fundamentales: cómo utilizar el sistema de una forma más eficiente y cómo aprovechar datos generados durante la fase de explotación del mismo.

En el primer caso, aplicamos el marco teórico conocido como predicción interactiva a la traducción automática neuronal. Este proceso consiste en integrar usuario y sistema en un proceso de corrección cooperativo, con el objetivo de reducir el esfuerzo humano empleado en obtener traducciones de alta calidad. Desarrollamos distintos protocolos de interacción para dicha tecnología, aplicando interacción basada en prefijos y en segmentos. Estos protocolos se implementan básicamente modificando el proceso de búsqueda del sistema. Además, ideamos mecanismos para obtener una interacción con el sistema más precisa, pero manteniendo la velocidad de generación del mismo. Llevamos a cabo una extensa experimentación, que muestra el potencial de

estas técnicas: superamos el estado del arte anterior, obtenido mediante tecnologías clásicas, por un gran margen y observamos que nuestros sistemas reaccionan mejor a las interacciones humanas.

A continuación, estudiamos cómo mejorar un sistema neuronal mediante los datos generados como subproducto de este proceso de corrección. Para ello, nos basamos en dos paradigmas del aprendizaje automático: el aprendizaje muestra a muestra y el aprendizaje activo. En el primer caso, el sistema se actualiza al vuelo, inmediatamente después de que el usuario corrige una frase. Por lo tanto, el sistema aprende de una manera continua a partir de correcciones, evitando cometer errores previos y especializándose en un usuario o dominio concretos. Evaluamos estos sistemas en una gran cantidad de situaciones y dominios diferentes, que demuestran el potencial que tienen los sistemas adaptativos. También llevamos a cabo una evaluación humana, con traductores profesionales. Éstos quedaron muy satisfechos con el sistema adaptativo. Además, fueron más eficientes cuando lo usaron, si lo comparamos con el uso de un sistema estático. En lo referente al segundo paradigma, el aprendizaje activo, lo aplicamos para el escenario en el que se deban traducir grandes cantidades de frases, siendo inviable la supervisión de todas ellas. En este caso, el sistema selecciona aquellas muestras que vale la pena supervisar, traduciendo el resto automáticamente. Aplicando este protocolo, redujimos de aproximadamente un cuarto el esfuerzo humano necesario para llegar a cierta calidad de traducción. Además, también superamos el estado del arte anterior por un margen considerable.

Finalmente, atacamos el complejo problema de la descripción de objetos multimedia, siguiendo la misma perspectiva de secuencia a secuencia. Este problema consiste en describir en lenguaje natural un objeto visual, es decir, una imagen o un vídeo. Comenzamos con la tarea de descripción de vídeos pertenecientes a un dominio general, la cual atacamos de forma similar al problema de la traducción. A continuación, nos movemos a un caso más específico: la descripción de eventos a partir de imágenes egocéntricas, capturadas a lo largo de un día. Como estos eventos son consecutivos, buscamos extraer relaciones entre ellos para generar descripciones más informadas. Para ello, desarrollamos un sistema capaz de analizar un mayor contexto, para considerar los eventos previos mientras analiza el actual. Los resultados muestran que el modelo con contexto extendido genera descripciones de mayor calidad que el modelo básico. Por último, aplicamos la predicción interactiva a estos sistemas de descripción multimodal. De la misma forma que en el caso de la traducción automática, este protocolo disminuye la cantidad de esfuerzo necesario para corregir las salidas de un sistema automático.