

COMPRESIÓN DE DATOS E IMÁGENES

JULIAN PFEIFLE

El objetivo fundamental de la compresión de datos es representar ficheros informáticos de una manera que ocupen mucho menos sitio que los originales. Por ejemplo, se podría representar un fichero que contiene el texto *Soy largo largo largo . . .*, con 1025 repeticiones de la palabra *largo*, de manera más eficiente como *Soy (1) largo (1025)*.

A pesar de que hoy en día el espacio de disco asequible para almacenar información es más abundante que nunca, existen muchas situaciones prácticas en las cuales es interesante la compresión de datos. Por ejemplo, a menudo hay que enviar datos a través de un canal con poca capacidad, como puede ocurrir en una casa con una conexión lenta de internet, o en la antena de una sonda espacial muy lejos del sol, que no dispone de mucha energía para enviar sus datos.

Se distinguen dos tipos de compresión de datos: en la *compresión sin pérdida* es importante poder reconstruir el original exactamente, sin cambiar ni un bit. Estos algoritmos se usan, entre otros, en programas de compresión como *gzip* o WinZip, o el formato de imágenes GIF. Por otro lado, en la *compresión con pérdidas* se permite que los datos originales solamente se puedan reconstruir de manera parcial, a cambio de una mejor tasa de compresión. Un ejemplo muy conocido es el formato JPEG para almacenar imágenes. Estos algoritmos funcionan bien porque se aprovechan de las limitaciones fisiológicas de los sentidos humanos para eliminar información de los datos que de todos modos no puede ser apreciada.

El trabajo puede incorporar uno o más de los siguientes apartados:

- Estudiar con detalle uno o más algoritmos de compresión sin pérdida. Unos candidatos podrían ser el algoritmo LZW, usado en el formato de imágenes GIF, o la codificación Huffman, usada en el programa *gzip*.
- Entender los principios geométricos y algorítmicos subyacentes del algoritmo JPEG de compresión con pérdidas. En particular, se llegaría a describir cómo una imagen se considera como un vector en un “espacio vectorial”, y cómo comprimir la imagen equivale a escoger una base del espacio en la cual pocos coeficientes de la imagen son grandes, y la mayoría se puede descartar.
- Se puede llegar a describir el nuevo paradigma de *Compressed Sensing*, que fue inventado en 2004 y que no deja de remover el campo, extendiéndose a ciencias como la geografía, la biología, la medicina, etc. Una vez entendido cómo funciona JPEG, la idea subyacente es bien sencilla, si bien los detalles requieren una alta sofisticación matemática. Sin embargo, existen numerosos tutoriales en la web que facilitan la tarea.

UNIVERSITAT POLITÈCNICA DE CATALUNYA
E-mail address: julian.pfeifle@upc.edu