Criptografía

ETSInf - UPV

Práctica 3

Criptoanálisis polialfabético

Curso 2013-14

1. Introducción

El objeto de esta práctica es trabajar con un sistema de cifrado por sustitución polialfabética, el sistema Vigenèrè. En la práctica se propone implementar las primeras técnicas que se diseñaron para atacar este cifrado.

El trabajo a realizar por parte del alumno consiste en descifrar el texto cifrados proporcionado. Los textos estarán disponibles en Poliformat, y del mismo modo que en la anterior práctica, corresponden a textos castellanos que consideran exclusivamente los 27 símbolos en mayúsculas del alfabeto castellano. El objetivo es, primero encontrar la clave de cifrado que permiten descifrar el criptograma, y posteriormente obtener el texto descifrado, bien en forma de fichero o impresos.

Los ficheros de datos tienen la extensión nb y contienen un texto cifrado cuya forma es una lista de los códigos ASCII de las letras que componen el criptograma. Si se elige trabajar con Mathematica, se pueden leer como sigue:

```
SetDirectory["directorio"];
fichero=OpenRead["nombre de fichero"];
InputStream[nombre de fichero];
c=Read[fichero];
Close[fichero];
```

Después de ejecutar estas instrucciones la variable c contiene la lista de códigos. Para evitar problemas con el alfabeto, éste se puede construir como sigue:

```
alfa="ABCDEFGHIJKLMNÑOPQRSTUVWXYZ";
alfabeto=Characters[alfa];
```

Para trabajar con la lista de letras hay que hacer lo que sigue:

```
c1=FromCharacterCode[c];
c2=Characters[c1];
```

La primera instrucción produce un string de las letras correspondientes a los códigos contenidos en la variable c. La segunda instrucción convierte el string en lista de Mathematica.

2. Criptoanálisis Vigenèrè

El criptoanálisis de estos sistemas se basa en la obtención de la longitud de la clave (número de alfabetos utilizados en el proceso de cifrado). Para ello consideraremos dos aproximaciones distintas: El método de Kasiski y el cálculo del índice de coincidencia.

2.1. Kasiski

El método de Kasiski se fundamenta en que un grupo de símbolos que aparezca k veces en un texto, será cifrado (en el peor de los casos) k/n veces con el mismo alfabeto, donde n denota el número de alfabetos.

De este modo, si se detectan un serie de segmentos repetidos en un criptograma, considerando las distancias entre las respectivas repeticiones $d_1, d_2, ..., d_k$, el periodo (longitud de la clave) es divisor del máximo común divisor de las distancias. En la medida en que los segmentos considerados tengan mayor tamaño, el cálculo realizado será más fiable.

Ejercicio 1:

> Diseña un módulo *Mathematica* que reciba dos listas de símbolos correspondientes a un criptograma y un segmento del criptograma, y devuelva la secuencia de posiciones donde es posible encontrar el segmento en el criptograma.

2.2. Cálculo del indice de coincidencia

Se ha visto durante el curso que, dado un texto cifrado, puede considerarse una medida de dispersión de las frecuencias de los símbolos del mensaje respecto a una distribución uniforme. Teniendo en cuenta un alfabeto del español con 27 símbolos:

$$MD = \sum_{i=0}^{26} \left(p_i - \frac{1}{n} \right)^2 = \sum_{i=0}^{26} \left(p_i^2 - \frac{2p_i}{n} + \frac{1}{n^2} \right) = \sum_{i=0}^{26} \left(p_i^2 \right) - 0{,}037$$

Ya se ha comentado que el primer término se denomina índice de coincidencia (IC), y en un texto donde los símbolos presentan una distribución própia del castellano:

$$IC = \sum_{i=0}^{26} \left(p_i^2 \right) = 0{,}072$$

Recordamos que el IC obtiene una medida de la probabilidad de que dos símbolos tomados al azar de un texto cifrado sean iguales. Una estimación del IC puede hacerse considerando la frecuencia de los símbolos en el criptograma:

$$IC \simeq \frac{\sum_{i=0}^{26} f_i(f_i-1)}{N(N-1)}$$

donde f_i denota el número de ocurrencias del caracter i-esimo en un criptograma de N símbolos. En una práctica previa se ha propuesto la implementación de un módulo que realice este cálculo.

2.3. Criptoanálisis de un cifrado Vigenèrè

Como se ha mencionado en esta práctica, el primer paso de un ataque a un criptograma Vigenèrè consiste en la obtención de la longitud de la clave. La consideración de varios segmentos repetidos en el criptograma permite aplicar el método de Kasiski y obtener una, o varias, hipótesis de la longitud de la clave.

El cálculo del IC permite, bien confirmar en cierta forma la hipótesis obtenida, en caso que sea única, bien seleccionar una entre las distintas hipótesis obtenidas.

El proceso que se propone consiste en la división del criptograma en una serie de textos que contengan los símbolos cuya posición es congruente módulo un cierto valor k. El cálculo del IC para cada uno de estos subtextos permite seleccionar el valor k que más se aproxima a un IC de un mensaje no cifrado en español (que como se ha visto se aproxima a 0.072).

Como ejemplo, considerando:

$$\begin{array}{ll} k=1 & IC = \{0{,}0471363\} \\ k=2 & IC = \{0{,}0479231, 0{,}0463764\} \\ k=3 & IC = \{0{,}0739754, 0{,}0753505, 0{,}072086\} \\ k=4 & IC = \{0{,}0470939, 0{,}046151, 0{,}0487458, 0{,}0464554\} \end{array}$$

puede verse como considerando tres subtextos se obtienen tres IC próximos al correspondiente a un mensaje no cifrado.

Ejercicio 2:

Diseña un módulo *Mathematica* que reciba una lista de símbolos correspondientes a un criptograma y un valor entero k. El módulo

deberá devolver una lista con k subtextos conteniendo cada uno de ellos los símbolos cuya posición es congruente módulo k.

Por ejemplo, dado $k = 3$ y el siguiente texto:

$$ESTOESUNEJEMPLODETEXTONOCIFRADO$$

el módulo deberá devolver los textos:

$$\{E, O, U, J, P, D, E, O, C, R, O\}$$
$$\{S, E, N, E, L, E, X, N, I, A\}$$
$$\{T, S, E, M, O, T, T, O, F, D\}$$

Una vez obtenido la longitud de la clave, para descifrar el mensaje basta considerar cada uno de los subtextos como un cifrado por desplazamiento, descifrar cada uno de ellos, y recomponer el mensaje original.

Ejercicio 3:

Diseña un módulo *Mathematica* que reciba una lista de textos (listas de símbolos) recomponga el mensaje, devolviendo una lista que contenga los símbolos ordenados teniendo en cuenta, primero su posición en la lista, considerando, y, para aquellos símbolos que ocupan la misma posición, la lista a la que pertenecen.

Por ejemplo, dado:

$$\{E, O, U, J, P\}$$
$$\{S, E, N, E, L\}$$
$$\{T, S, E, M, O\}$$

el módulo devolverá el siguiente texto:

$$ESTOESUNEJEMPLO$$