MA4402-1 Simulación estocástica: T. y L.
Profesor: Joaquín Fontbona T.
Auxiliares: Catalina Lizana, Álvaro Márquez y Matías Ortiz.

Descifrando Mensajes Codificados Usando MCMC

Integrantes: Sebastián Flores, Matías Neto.

Abstract: Implementamos el algoritmo de Simmulated Annealing para decodificar mensajes cifrados por sustitución simple. Aplicamos algunas de mejoras propuestas por [2], evaluamos su rendimiento en términos de la precisión del mensaje decodificado resultante, y comparamos con respecto al método de *Stochastic Hill-Climbing* propuesto en [1], y le agregamos backtracking. Bajo las condiciones de prueba que propusimos, Simmulated Annealing obtiene resultados similares a los obtenidos con el benchmark.

1. Introducción

En la criptografía clásica, el cifrado por sustitución simple consiste en codificar un mensaje reemplazando un símbolo el alfabeto usual por otro, siguiendo una clave que determina el mapeo entre un alfabeto y otro. Restringiéndose al alfabeto latino, las claves de sustitución simple consisten en permutaciones del mismo alfabeto.

Existen pocos métodos sistemáticos para atacar mensajes codificados de esta manera, y las heurísticas que se le aplican se apoyan, por ejemplo, del análisis frecuencial de los símbolos que aparecen en el mensaje. En [3] se describe un algoritmo que bebe de esta idea para aplicar Markov Chain Monte Carlo.

Específicamente, consideramos el grafo G cuyos vértices son las posibles permutaciones del alfabeto latino y decimos que dos permutaciones son vecinas si se puede obtener una permutando dos símbolos de la otra. Por ejemplo, `AZERTYUIOPMLKJHGFDSQWXCVBN` es vecina de `NZERTYUIOPMLKJHGFDSQWXCVBA` pues se intercambian sólo el primer y último símbolo.

Sea M la matriz indexada por las letras del alfabeto tal que

$$M_{AB} = \text{Probabilidad de que a una A siga una B}$$

tomando como referencia el idioma castellano.

Dado un texto cifrado, y para todo vértice f del grafo de claves, definimos su plausibilidad como

$$\text{Tri-Pl}(f) = \prod_i M_{f(s_i)f(s_{i+1})f(s_{i+2})}$$

donde i recorre los símbolos del mensaje codificado. El objetivo es encontrar la clave que maximice la plausibilidad para el mensaje dado vía Simmulated Annealing

2. Metodología

Implementamos en `python` el algoritmo de Simmulated Annealing para resolver dos textos cifrados de juguete, de 100 y 1800 caracteres aproximadamente. Incorporamos algunas mejoras al método propuestas por [2] y comparamos su rendimiento en términos de la precisión de la propuesta de decodificación del mensaje.

A modo de benchmark, utilizamos también el método de *Stochastic Hill-Climbing* propuesto en [1], añadimos una modificación del método agregando backtracking y comparamos la precisión del *output* de cada método a tiempo de cálculo equivalente.

Encontramos que el método de Simmulated Annealing tienen precisiones similares al benchmark.

Referencias

[1] Luka Bulatović et al. "Automated cryptanalysis of substitution cipher using hill climbing with well designed heuristic function". En: *Mathematica Montisnigri* 44 (2019), págs. 135-143. DOI: 10.20948/mathmon-2019-44-11.

[2] Jian Chen y Jeffrey S Rosenthal. "Decrypting classical cipher text using Markov chain Monte Carlo". En: *Statistics and Computing* 22.2 (2012), págs. 397-413.

[3] Persi Diaconis. "The Markov Chain Monte Carlo revolution". En: *Bulletin of the American Mathematical Society* 46.2 (2009), págs. 179-205.