MCMC Review

Guillermina Montanari

11/15/2016

La revolución de MCMC (Marcov Chain Monte Carlo)

MCMC es un método de simulación que ha tomado mucha fuerza apoyada sobre todo por el avance de procesamiento computacional. Bajo este contexto el autor introduce el método *Random Walk* que se realiza en el sistema de interés para resolver muchos problemas científicos. Este objeto matemático, describe un **camino**, que consiste en una sucesión de pasos aleatorios. Al hablar de **caminos** podemos citar como ejemplos el movimiento de una molécula viajando en un líquido o gas; el camino que recorre un animal salvaje en busca de comida e incluso problemas que en principio no parecen tan claros en su aplicación, sin embargo, han dado muy buenos resultados, como en el campo de la criptografía.

Criptografia

Para un caso de servicios de consultoría que se ofrecen en el Departamento de Estadística de la Universidad de Stanford, llegó un caso de la mano de un Psicólogo de la cárcel del estado, que tenía varios ejemplos de mensajes "cifrados" que los presidiarios se enviaban unos a otros. Un ejemplo es el siguiente:

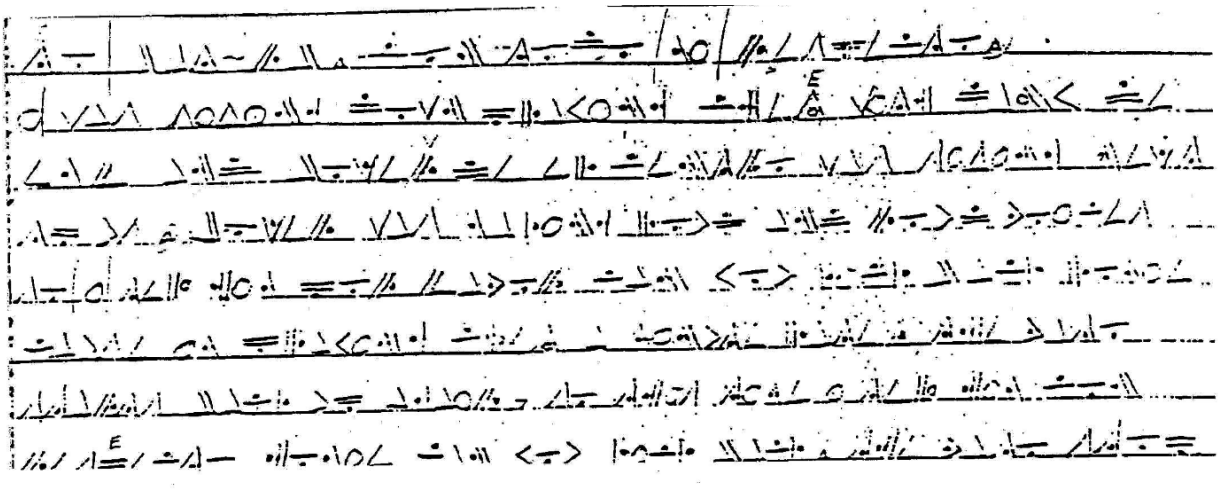

Figure 1: Mensaje Encriptado

El problema de de-codificación, implica una **función** que si conoce el algoritmo de cifrado es capaz de descifrarlo al lenguaje de interés.

$f : (código) -> alfabeto$

Este problema se resuelve en forma estándard de la siguiente manera:

- Se usan estadisticas del lenguaje escrito, en este caso era inglés (un documento rico en el lenguaje objetivo)
- Se usan diferentes alternativas de la función f
- Se prueban
- Se observa si el mensaje obtenido tiene sentido

En este caso, se usó como referencia del lenguaje el texto de *La Guerra y la Paz* y se tomaron las transiciones de primer orden. Estas transiciones se refieren a la proporción de símbolos de texto consecutivos, de un símbolo x al siguiente y. Estas relaciones entre dos símbolos, en un texto completo, dan origen a una matriz $M(x, y)$ que otorgan combinaciones de opciones plausibles de símbolos y si el anterio es x. La plausibilidad, se obtiene a través de probar estas funciones con diferentes enfoques matemáticos y las que obtienen los mayores valores son buenas candidatas para desencriptar.

Sin querer involucrar demasiados conceptos matemáticos vamos a establer el siguiente enfoque matemático a modo de ejemplo:

$PI(f) = M(f(si), f(si + 1))$

Donde si se ejecuta sobre símbolos consecutivos del mensaje encriptado para obtener los valores de plausibilidad.

El valor de MCMC

Los pasos de usar un f y buscar el valor que le asigna dado un símbolo dado se encuentra en un espacio muy grande de opciones posibles que tomaría mucho tiempo descubrir. La pregunta aquí es ¿porqué **MCMC** es exitoso?

Este problema se encaró maximizando la función f ejecutando el algoritmo de **MCMC** usando la transposición de los símbolos enla funcion f además de la evaluación directa de los mismos. (más unos trucos internos que involucran a una moneda pero exceden el objetivo de este documento, aunque se puede revisar en las referencias que vienen abajo)

Por ejemplo, dado este texto:

ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS
NOBLER IN THE MIND TO SUFFER THE SLINGS AND ARROWS OF OUTRAGEOUS
FORTUNE OR TO TAKE ARMS AGAINST A SEA OF TROUBLES AND BY OPPOSING END

Figure 2: Texto Objetivo

Se lo desarmó por completo y esa fué la entrada para el algoritmo de **MCMC**. Las diferentes simulaciones de las funciones y su consecuente opción como la función f que mejor desencriptara el mensaje dieron las siguientes salidas:

```
 100 ER ENOHDLAE OHDLO UOZEOUNORU O UOZEO HD OITO HEOQSET IUROFHE HENO ITORUZAEN
 200 ES ELOHRNDE OHRNO UOVEOULOSU O UOVEO HR OITO HEOQAET IUSOPHE HELO ITOSUVDEL
 300 ES ELOHANDE OHANO UOVEOULOSU O UOVEO HA OITO HEOQRET IUSOFHE HELO ITOSUVDEL
 400 ES ELOHINME OHINO UOVEOULOSU O UOVEO HI OATO HEOQRET AUSOWHE HELO ATOSUVMEL
 500 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
 600 ES ELOHINME OHINO UODEOULOSU O UODEO HI OATO HEOQRET AUSOWHE HELO ATOSUDMEL
 900 ES ELOHANME OHANO UODEOULOSU O UODEO HA OITO HEOQRET IUSOWHE HELO ITOSUDMEL
1000 IS ILOHANMI OHANO RODIORLOSR O RODIO HA OETO HIOQUIT ERSOWHI HILO ETOSRDMIL
1100 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1200 ISTILOHANMITOHANOT ODIO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1300 ISTILOHARMITOHAROT ODIO LOS TOT ODIOTHATOENOTHIOQUINTE SOWHITHILOTENOS DMIL
1400 ISTILOHAMRITOHAMOT OFIO LOS TOT OFIOTHATOENOTHIOQUINTE SOWHITHILOTENOS FRIL
1600 ESTEL HAMRET HAM TO CE OL SOT TO CE THAT IN THE QUENTIOS WHETHEL TIN SOCREL
1700 ESTEL HAMRET HAM TO BE OL SOT TO BE THAT IN THE QUENTIOS WHETHEL TIN SOBREL
1800 ESTER HAMLET HAM TO BE OR SOT TO BE THAT IN THE QUENTIOS WHETHER TIN SOBLER
1900 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
2000 ENTER HAMLET HAM TO BE OR NOT TO BE THAT IS THE QUESTION WHETHER TIS NOBLER
```

Lo interesante de este algoritmo es que al inicio no tiene ningún sentido, pero a medida que las simulaciones avanzan, y teniendo en cuenta la capacidad de procesamiento actual, en poco tiempo - 2000 simulaciones - el mensaje ya es muy parecido al original. Regresando al ejemplo original de la cárcel, el mensaje desncriptado fué el siguiente:

```
  to bat-rb. con todo mi respeto. i was sitting down playing chess with
danny de emf and boxer de el centro was sitting next to us. boxer was
making loud and loud voices so i tell him por favor can you kick back
homie cause im playing chess a minute later the vato starts back up again
so this time i tell him con respecto homie can you kick back.  the vato
stop for a minute and he starts up again so i tell him check this out shut
the f**k up cause im tired of your voice and if you got a problem with it
we can go to celda and handle it. i really felt disrespected thats why i
told him. anyways after i tell him that the next thing I know that vato
slashes me and leaves. dy the time i figure im hit i try to get away but
the c.o. is walking in my direction and he gets me right dy a celda. so i
go to the hole. when im in the hole my home boys hit doxer so now "b" is
also in the hole. while im in the hole im getting schoold wrong and
```

Figure 3: Mensaje desencriptado “limpiado un poco por la mano del hombre”

Conclusiones del autor

- Lo que más le gustó es que el ejemplo es real.
- El algoritmo logró descifrar el mensaje
- Una nota curiosa es que no sólo pudo descifrar inglés sino que también español y vocabulario propio de la cárcel. (Me imagino que al modelo se lo “nutrió” con textos que contenían español y jerga de la cárcel)

Conclusiones mías:

A través de las clases mi mente a comenzado a modificarse. Antes me era difícil encontrar una relación entre las matemáticas y el mundo que me rodea. (perdón la brutez). Cada nuevo tema - lo son para mi -, abre en mi mente estas - nuevas - relaciones entre las matemáticas y mis procesos diarios.

Regresando a este texto en particular, la posibilidad que abre la idea de que a través de las simulaciones de modelos matemáticas se puedan entender temas que en principio parecerían no tener nada que ver con las matemáticas se me hace fascinante.

Los procesos humanos, hablando de aquellos tan complejos como el lenguaje, por citar uno, que pueden ser "entendidos"" y reproducidos en modelos matemáticos, me hace pensar que quizás hasta nuestra conciencia funciona de la misma manera. Procesos matemáticos que determinan quienes somos, como pensamos, como sentimos. Procesos que aprenden de otros procesos, como la sociedad, la moral, la cultura ... Serán todos procesos que pueden reducirse a modelos matemáticos? Claro, lo que pienso después es hasta que punto son modificables y a criterio de quien. Que maravilla y que miedo! Se me va perdiendo de a poco mi sentido de "ser único" e "irrepetible". Quizás solo sea parte del proceso de aprendizaje.

Espero mi conclusión no sea muy fumada para el objetivo de este trabajo.

Guillermina Montanari - 101421

Figure 4: yo"

Referencias

1. "The Markov Chain Monte Carlo Revolution"" https://math.uchicago.edu/~shmuel/Network-course-readings/MCMCRev.pdf
2. Wikipedia: "Random Walk" https://en.wikipedia.org/wiki/Random_walk