



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

Actividad 8: *Markov Chain*

Tópicos Avanzados en Inteligencia de Máquina - IIC 3695

Profesor : Karim Pichara Baksai.

Ayudantes : Ignacio Becker, Francisco Pérez Galarce, Matías Vergara

Fecha : 25 de Abril de 2019

La actividad tiene por objetivo crear y aplicar una cadena de markov básica para la generación de texto. Concretamente, se utilizará un conjunto de noticias de diarios nacionales para crear la cadena y posteriormente se simularán frases a partir de esta.

Para el desarrollo de la actividad recomendamos seguir estos pasos:

- Importar el archivo de noticias (*corpus.txt*) y preprocesarlo, esto quiere decir eliminar caracteres especiales y dejarlo en una estructura de datos adecuada.
- A partir del *corpus*, para cada palabra genere un *contexto* posible. Por ejemplo, usted puede decidir que todas las palabras que aparecen posterior a una palabra dada representan su *contexto*. En esta parte recomendamos usar un diccionario cuyas *keys* sean todas las palabras del *corpus* y los valores asociados a cada *key* sean las distintas palabras que le pueden seguir a la palabra que hace de *key*. Sugerimos usar *listas* para almacenar los candidatos a palabra siguiente.
- Para generar una frase, primero considere que el estado inicial (*palabra*) será aleatorio, luego el estado siguiente posible está restringido al *contexto* de dicha palabra (distribuye uniforme dentro del contexto).
- Genere 20 frases de noticias basados en el método anteriormente descrito. Comente al respecto, ¿Qué supuesto hay en este modelamiento? ¿Qué modificaciones cree que mejorarían la calidad de la frases?.