

Entregable 1-A

Estimación de emisiones de CO2

Nosotros elegimos este proyecto, ya que en nuestra ciudad los problemas con la emisión de gases contaminantes es una situación crítica que requiere especial atención, es por ello que nos surgió un particular interés por aplicar las técnicas de búsqueda de conocimiento sobre el caso que se nos ofreció a partir de información de las partículas del aire dentro de una región de North Jutland, Dinamarca

Algoritmo EM

Decidimos utilizar EM dada la naturaleza del problema, este algoritmo permite encontrar los valores restantes para un nuevo conjunto de datos **Z** dado un conocimiento previo (conjunto de entrenamiento) aproximando mediante iteraciones una función de máxima verosimilitud hasta llegar a un punto de convergencia.[1]

El algoritmo se inicializa dándole pesos aleatorios a los vectores del conjunto Z, después, se aproxima reduciendo la función de máxima verosimilitud hasta converger en un punto donde abremos encontrado la función para los parametros del modelo que mejor nos aproximan el comportamiento.

A pesar de que este algoritmo puede atorarse en mínimos locales, el comportamiento aleatorio de las partículas en el aire nos permite tomarlo como mejor candidato sin tener problemas.

Conjunto de datos

El conjunto de datos usado será el que nos proporcionaron dentro del documento, consiste de 4 componentes, un ID para las aristas de la grafica del mapa dentro de OpenMap, la longitud, y latitud, además de la altura de la partícula, así, tenemos un mapa 3D de las partículas de aire, para el tratamiento de los datos, necesitaremos reducir los datos y no tomar en cuenta el ID de OpenMap para hacer el analisis de los puntos con solo 3 dimensiones y aplicar el algoritmo de EM.[2]

Detalles técnicos

El algoritmo se implementará sobre spark en Scala, trabajaremos sobre un solo nodo dado el reducido volumen de los datos, pero en una maquina multicore (usaremos una 16 nucleos y 30 GB de RAM para generar el modelo) [3]

[1] <http://cs229.stanford.edu/notes/cs229-notes8.pdf>

[2] [https://archive.ics.uci.edu/ml/datasets/3D+Road+Network+\(North+Jutland,+Denmark\)](https://archive.ics.uci.edu/ml/datasets/3D+Road+Network+(North+Jutland,+Denmark))

[3] <https://github.com/amplab/spark-ec2>