

Un Tour en Helicóptero por las Matemáticas Aplicadas a la Industria en Colombia (con dos aterrizajes)

Alvaro J. Riascos Villegas
University of los Andes y Quantil

Septiembre 27, 2017

Contents

- 1 **Introducción**
- 2 Detección de minería ilegal usando imágenes satelitales
- 3 Minería de datos médicos
- 4 Mercadeo basado en datos
- 5 Aterrizaje: Predicción del crimen
- 6 Aterrizaje: Minería de texto
- 7 Políticas Públicas

Introducción

- Presento los ejemplos más significativos de aplicaciones exitosas de las Matemáticas Aplicadas en la academia e industria: **una visión muy personal** como profesor de UNIANDES y co-Director de Quantil.
- Estas son aplicaciones a:
 - 1 Detección de minería ilegal.
 - 2 Minería de datos médicos.
 - 3 Mercadeo basado en datos.
 - 4 Predicción del crimen.
 - 5 Minería de texto.
 - 6 Políticas públicas.

Introducción

- Presento los ejemplos más significativos de aplicaciones exitosas de las Matemáticas Aplicadas en la academia e industria: **una visión muy personal** como profesor de UNIANDES y co-Director de Quantil.
- Estas son aplicaciones a:
 - 1 Detección de minería ilegal.
 - 2 Minería de datos médicos.
 - 3 Mercadeo basado en datos.
 - 4 Predicción del crimen.
 - 5 Minería de texto.
 - 6 Políticas públicas.

Introducción

- Presento los ejemplos más significativos de aplicaciones exitosas de las Matemáticas Aplicadas en la academia e industria: **una visión muy personal** como profesor de UNIANDES y co-Director de Quantil.
- Estas son aplicaciones a:
 - 1 Detección de minería ilegal.
 - 2 Minería de datos médicos.
 - 3 Mercadeo basado en datos.
 - 4 Predicción del crimen.
 - 5 Minería de texto.
 - 6 Políticas públicas.

Introducción

- Presento los ejemplos más significativos de aplicaciones exitosas de las Matemáticas Aplicadas en la academia e industria: **una visión muy personal** como profesor de UNIANDES y co-Director de Quantil.
- Estas son aplicaciones a:
 - 1 Detección de minería ilegal.
 - 2 Minería de datos médicos.
 - 3 Mercadeo basado en datos.
 - 4 Predicción del crimen.
 - 5 Minería de texto.
 - 6 Políticas públicas.

Introducción

- Presento los ejemplos más significativos de aplicaciones exitosas de las Matemáticas Aplicadas en la academia e industria: **una visión muy personal** como profesor de UNIANDES y co-Director de Quantil.
- Estas son aplicaciones a:
 - 1 Detección de minería ilegal.
 - 2 Minería de datos médicos.
 - 3 Mercadeo basado en datos.
 - 4 Predicción del crimen.
 - 5 Minería de texto.
 - 6 Políticas públicas.

Introducción

- Presento los ejemplos más significativos de aplicaciones exitosas de las Matemáticas Aplicadas en la academia e industria: **una visión muy personal** como profesor de UNIANDES y co-Director de Quantil.
- Estas son aplicaciones a:
 - 1 Detección de minería ilegal.
 - 2 Minería de datos médicos.
 - 3 Mercadeo basado en datos.
 - 4 Predicción del crimen.
 - 5 Minería de texto.
 - 6 Políticas públicas.

Introducción

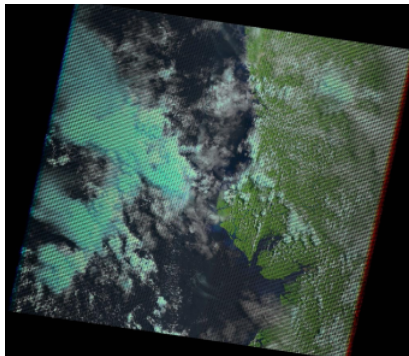
- Presento los ejemplos más significativos de aplicaciones exitosas de las Matemáticas Aplicadas en la academia e industria: **una visión muy personal** como profesor de UNIANDES y co-Director de Quantil.
- Estas son aplicaciones a:
 - 1 Detección de minería ilegal.
 - 2 Minería de datos médicos.
 - 3 Mercadeo basado en datos.
 - 4 Predicción del crimen.
 - 5 Minería de texto.
 - 6 Políticas públicas.

Contents

- 1 Introducción
- 2 Detección de minería ilegal usando imágenes satelitales
- 3 Minería de datos médicos
- 4 Mercadeo basado en datos
- 5 Aterrizaje: Predicción del crimen
- 6 Aterrizaje: Minería de texto
- 7 Políticas Públicas

Detección de minería ilegal usando imágenes satelitales

- Cortesía de Santiago Saavedra (Universidad del Rosario)

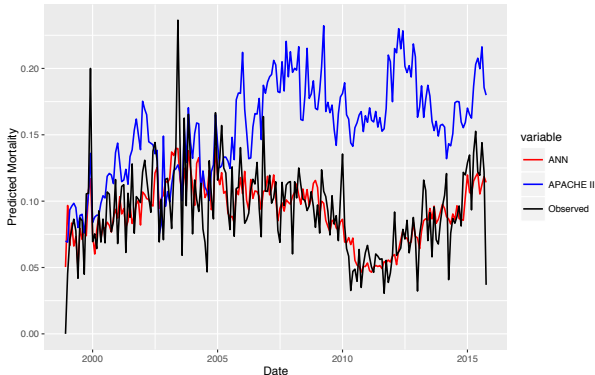


Contents

- 1 Introducción
- 2 Detección de minería ilegal usando imágenes satelitales
- 3 Minería de datos médicos**
- 4 Mercadeo basado en datos
- 5 Aterrizaje: Predicción del crimen
- 6 Aterrizaje: Minería de texto
- 7 Políticas Públicas

Minería de datos médicos

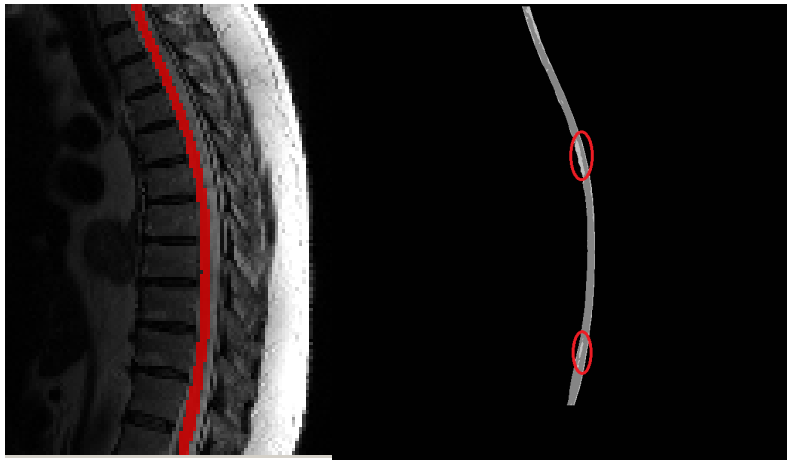
- Cortesía Fundación Valle del Lili (FVL).



Gráfica: Cortesía de Johns Hopkins University

(a) Spinal Cord

(b) Spinal Cord Injury

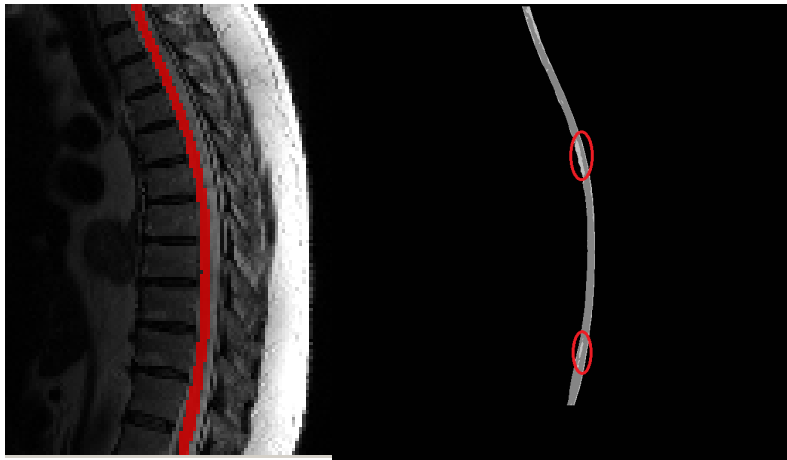


- Detección de signos vitales sin contacto: Uniandes, Harvard, ICESI, EVI

Gráfica: Cortesía de Johns Hopkins University

(a) Spinal Cord

(b) Spinal Cord Injury



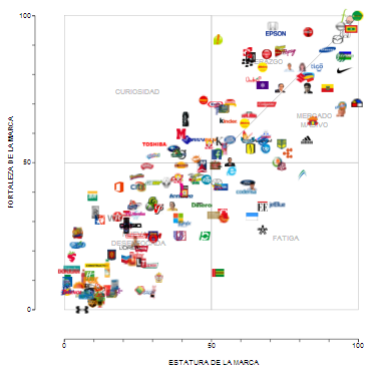
- Detección de signos vitales sin contacto: Uniandes, Harvard, ICESI, EVI

Contents

- 1 Introducción
- 2 Detección de minería ilegal usando imágenes satelitales
- 3 Minería de datos médicos
- 4 Mercadeo basado en datos**
- 5 Aterrizaje: Predicción del crimen
- 6 Aterrizaje: Minería de texto
- 7 Políticas Públicas

Mercadeo basado en datos

- Cortesía: Barbara & Frick y Young & Rubican.
- Minería de redes sociales (twitter).



Contents

- 1 Introducción
- 2 Detección de minería ilegal usando imágenes satelitales
- 3 Minería de datos médicos
- 4 Mercadeo basado en datos
- 5 Aterrizaje: Predicción del crimen**
- 6 Aterrizaje: Minería de texto
- 7 Políticas Públicas

Modelo Espacio - Temporal

- Este es un modelo basado en la clasificación de los eventos como antecedentes y réplicas.
- Es el estado del arte en modelos de predicción del crimen.
- Está motivado por teorías de contagio espacial y réplicas temporales.

Modelo Espacio - Temporal: Motivación

Mohler et al.: Self-Exciting Point Process Modeling of Crime

101

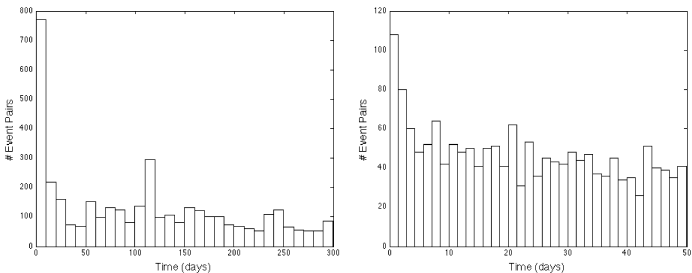


Figure 1. On the left, histogram of times (less than 300 days) between Southern California earthquake events of magnitude 3.0 or greater separated by 110 kilometers or less. On the right, histogram of times (less than 50 days) between burglary events separated by 200 meters or less.

Modelo Espacio - Temporal: Motivación

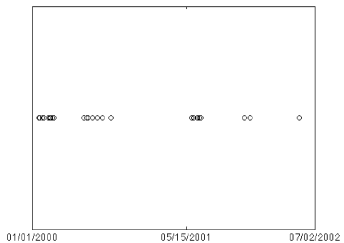


Figure 2. Times of violent crimes between two rivalry gangs in Los Angeles.

- Se considera un modelo de la intensidad espacio temporal del crimen de la forma:

$$\lambda(t, x, y) = \mu(t, x, y) + \sum_{k:t_k < t} g(t - t_k, x - x_k, y - y_k) \quad (1)$$

- Sea $\mu(t, x, y) = \nu(t)\mu(x, y)$

- Suponiendo que el modelo es correcto la probabilidad de que un evento i sea un evento antecedente es:

$$\pi_{ii} = \frac{\nu(t_i)\mu(x_i, y_i)}{\lambda(t_i, x_i, y_i)} \quad (2)$$

- La probabilidad de que el evento j cause el evento i (evento réplica) es:

$$\pi_{ji} = \frac{g(t_i - t_j, x_i - x_j, y_i - y_j)}{\lambda(t_i, x_i, y_i)} \quad (3)$$

- Sea P la matriz de probabilidades (p_{ji}) (obsérvese que la suma de columnas da 1).
- Elija una probabilidad inicial P_0 .
- Muestrear de P_0 puntos de transfondo y réplica:
 $(t_k, x_k, y_k, p_{kk})_{k=1, \dots, N}$ y $(t_i - t_j, x_i - x_j, y_i - y_j, p_{ji})_{i > j}$ y usando esta muestra hacer una estimación inicial de:
 - 1 μ_1 de μ .
 - 2 g_1 de g .
 - 3 ν_1 de νusando KDE con ancho de banda variable.
- Estimar inicialmente λ_1 usando que las columnas de P suman 1.

- En cada iteración n , ajustamos simulados con P_n los datos $g_n(t_i - t_j, x_i - x_j, y_i - y_j, p_{ji})$ de g . Ahora ajustamos una un kernel del siguiente tipo:

$$g_n(t, x, y) = \frac{1}{N} \sum_{i=1}^{N_0} \frac{1}{\sigma_x \sigma_y \sigma_t (2\pi)^{\frac{3}{2}} D_i^3} \times \exp\left(-\frac{(x - x_i)^2}{2\sigma_x^2 D_i^2} - \frac{(y - y_i)^2}{2\sigma_y^2 D_i^2} - \frac{(t - t_i)^2}{2\sigma_t^2 D_i^2}\right)$$

- Usamos un procedimiento similar para estimar ν y μ . En el primer caso un Kernel Gaussiano univariado y el segundo caso un Kernel Gaussiano bivariado.

- Actualizar P_1 .
- Repetir los pasos anteriores hasta que la matriz P no cambie mucho.
- El número de probabilidades a estimar en cada iteración es del orden de N^2 .
- Esto es computacionalmente costoso (típicamente N del orden de 1,000 o 30,000 - Bogota en un año).

Validación

- Utilizamos el *Precision Accuracy Index*

$$PAI = \frac{\text{Hit Rate}}{\text{Percentage of Area}}$$

$$\text{Hit Rate} = \frac{\text{Crimes predicted in Hotspots}}{\text{Total Crimes}}$$

$$\text{Percentage Area} = \frac{\text{Area of Hotspots}}{\text{Total Area}}$$

- Sin embargo en modelos muy granulares no es una buena medida.

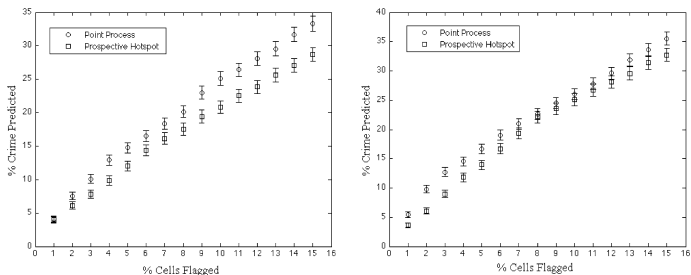


Figure 6. Forecasting strategy comparison. Average daily percentage of crimes predicted plotted against percentage of cells flagged for 2005 burglary using 200 m by 200 m cells. Error bars correspond to the standard error. Prospective hotspot cutoff parameters are 400 meters and 8 weeks (left) and optimal parameters (right) are 200 meters and 39 weeks. Spatial background intensity $\mu(x, y)$ smoothing bandwidth for the point process is 300 meters (left) selected by cross validation and 130 meters (right) selected to optimize the number of crimes predicted.

- Se cuenta con datos de crímenes en Bogotá del 16 de abril al 30 de junio de 2017: 16.402 datos.
- Para esta validación se usaron datos de la localidad de Santa Fé, delimitada por latitudes entre $[4.571, 4.629]$ y longitudes > -74.091 .

- Se cuenta con datos de crímenes en Bogotá del 16 de abril al 30 de junio de 2017: 16.402 datos.
- Para esta validación se usaron datos de la localidad de Santa Fé, delimitada por latitudes entre $[4.571, 4.629]$ y longitudes > -74.091 .

- 1.676 ($\approx 10\%$) crímenes en la localidad de Santa Fé en el período tratado.
- Se entrenó el modelo de crimen con datos entre 1 y 7 semanas y se validó con las 3 semanas posteriores al entrenamiento, en todos los casos, del 10 al 30 de junio de 2017 (407 crímenes).

- 1.676 ($\approx 10\%$) crímenes en la localidad de Santa Fé en el período tratado.
- Se entrenó el modelo de crimen con datos entre 1 y 7 semanas y se validó con las 3 semanas posteriores al entrenamiento, en todos los casos, del 10 al 30 de junio de 2017 (407 crímenes).

- Se divide Bogotá (1.547km^2) en 10,946 celdas de $\approx 145\text{m}^2$ cada una; 1.019 celdas en la localidad de Santa F.
- Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

- Se divide Bogotá (1.547km^2) en 10,946 celdas de $\approx 145\text{m}^2$ cada una; 1.019 celdas en la localidad de Santa F.
- Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

- Se divide Bogotá (1.547km^2) en 10,946 celdas de $\approx 145\text{m}^2$ cada una; 1.019 celdas en la localidad de Santa F.
- Se entrena el modelo con el correspondiente número de semanas y se predicen los puntos calientes para cada turno de 8 horas, definidos como el 10% de las celdas con mayor probabilidad de crimen.
- Se investigan cuántos crímenes de los datos de validación ocurrieron en los puntos calientes predichos por el modelo.

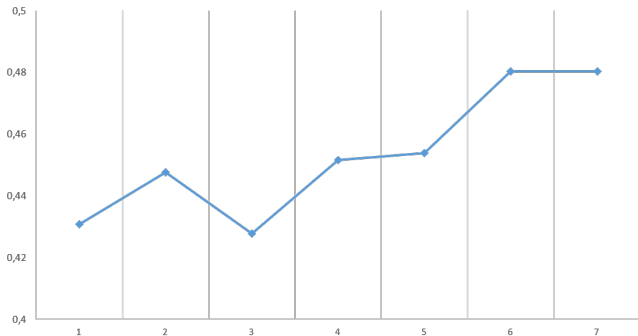
Validación: Hit Rate

- La siguiente tabla registra el Hit Rate del modelo al ser entrenado con datos entre 1 y 7 semanas. El Hit Rate se calcula para cada uno de los 21 turnos de 8 horas dentro de cada semana.

Entrenamiento	1	2	3	4	5	6	7
Semana 1	0,49	0,47	0,44	0,44	0,46	0,44	0,44
Semana 2	0,41	0,40	0,40	0,43	0,43	0,46	0,46
Semana 3	0,39	0,48	0,44	0,49	0,48	0,54	0,54
Promedio	0,43	0,45	0,43	0,45	0,45	0,48	0,48

Validación: Hit Rate

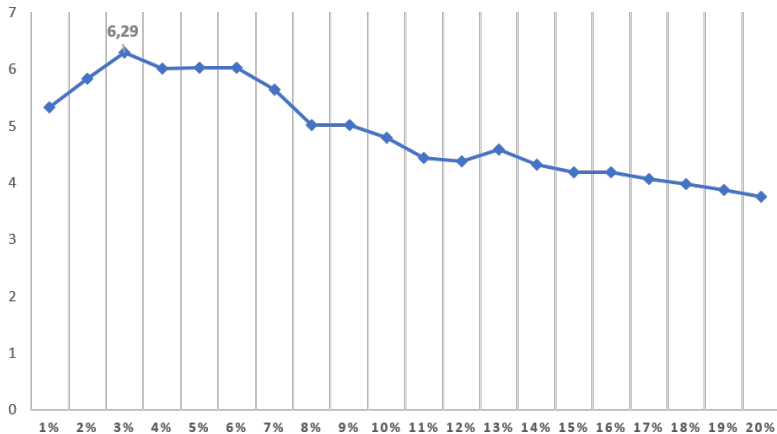
HIT RATE PROMEDIO POR SEMANAS DE ENTRENAMIENTO



Al evaluar la eficiencia del modelo entrenado con 7 semanas, se encontró:

Cobertura	1%	5%	10%	15%	20%
Semana 1	1,53	5,04	4,4	4,21	3,92
Semana 2	4,33	6,14	4,62	3,98	3,64
Semana 3	10,4	6,92	5,43	4,37	3,7
Promedio	5,33	6,02	4,8	4,19	3,75

PAI PROMEDIO SEGÚN PORCENTAJE DE COBERTURA



- Se puede apreciar que entre mayor sea el número de datos de entrenamiento, el modelo tiene mayor capacidad predictiva.
- Este método de validación se puede utilizar también para calibrar otros parámetros del modelo: tipo de kernel y anchos de banda, pesos otorgados a los datos de entrenamiento, entre otros.

- Se puede apreciar que entre mayor sea el número de datos de entrenamiento, el modelo tiene mayor capacidad predictiva.
- Este método de validación se puede utilizar también para calibrar otros parámetros del modelo: tipo de kernel y anchos de banda, pesos otorgados a los datos de entrenamiento, entre otros.

Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

Prediccin	bw fijo	bw variable
Semana 1	0,44	0,57
Semana 2	0,46	0,59
Semana 3	0,54	0,62
Promedio	0,48	0,59

Hit Rate con 7 semanas de entrenamiento y 10% de cobertura de puntos calientes:

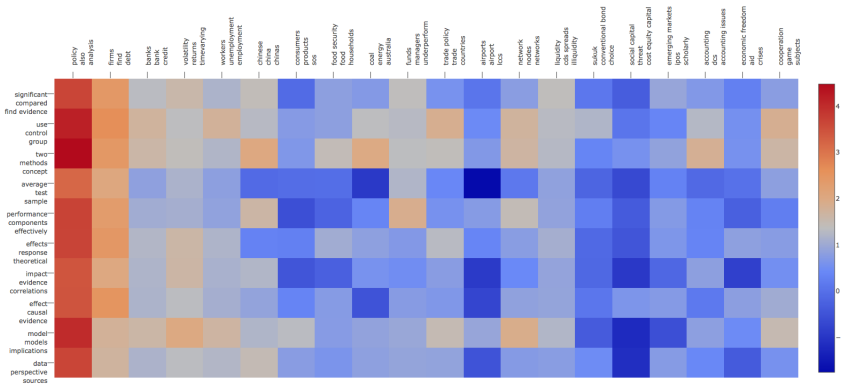
Prediccin	bw fijo	bw variable	KDE
Semana 1	0,44	0,57	0,42
Semana 2	0,46	0,59	0,44
Semana 3	0,54	0,62	0,53
Promedio	0,48	0,59	0,46

Contents

- 1 Introducción
- 2 Detección de minería ilegal usando imágenes satelitales
- 3 Minería de datos médicos
- 4 Mercadeo basado en datos
- 5 Aterrizaje: Predicción del crimen
- 6 Aterrizaje: Minería de texto**
- 7 Políticas Públicas

DetECCIÓN DE MINERÍA ILEGAL USANDO IMÁGENES SATELITALES
 MINERÍA DE DATOS MÉDICOS
 MERCADERO BASADO EN DATOS
 ATERRIZAJE: PREDICCIÓN DEL CRIMEN
 ATERRIZAJE: MINERÍA DE TEXTO
 POLÍTICAS PÚBLICAS

GAP MAPS



Definition

Un vocabulario V es una sucesión finita de elementos diferentes (v_1, \dots, v_n) . n es llamado el tamaño del vocabulario V . v_1, \dots, v_n son llamados los términos del vocabulario V .

Definition

Un documento sobre un vocabulario V es una sucesión $d := (w_1, \dots, w_n)$ con $w_i \in V, \forall i \in \{1, \dots, n\}$. n es llamado la longitud del documento d .

Definition

Sea V un vocabulario, v_i un término y $d = (w_1, \dots, w_n)$ un documento sobre este vocabulario.

$$c_i(d) := \sum_{j=1}^n \mathbb{I}(v_i = w_j)$$

es la función frecuencia del término v_i

Definition

El vector de frecuencia de términos de un documento d sobre un vocabulario V es el vector

$$f(d) := (c_1(d), \dots, c_n(d))$$

Definition

Un Corpus $C = (d_1, \dots, d_m)$ es una sucesión de documentos sobre un vocabulario V . n es llamado el tamaño del Corpus.

Definition

Dado un Corpus $C = (d_1, \dots, d_m)$ sobre un vocabulario $V = (v_1, \dots, v_n)$ se define la matriz de términos y documentos

$$[M_{ij}] = f(d_j)_i, \quad 1 \leq i \leq n, \quad 1 \leq j \leq m$$

Esto quiere decir que la entrada en la posición i, j corresponde a la frecuencia con que ocurre el término v_i dentro del documento d_j .

GAP MAPS: Latent Dirichlet Allocation (LDA)

- LDA es un modelo probabilístico generativo de un corpus.
- La idea es que los documentos son representados como mezclas aleatorias sobre tópicos latentes, donde cada tópico es caracterizado como una distribución sobre palabras.

GAP MAPS: Latent Dirichlet Allocation (LDA)

- LDA asume el siguiente proceso generativo para cada documento en un corpus :
 - 1 Escoja $N \sim \text{Poisson}(\xi)$.
 - 2 Escoja $\theta \sim \text{Dirichlet}(\alpha)$.
 - 3 Para cada una de las N palabras w_n :
 - Escoja un t3pico $z_n \sim \text{Multinomial}(\theta)$.
 - Escoja una palabra w_n de $p(w_n | z_n, \beta)$, una distribuci3n de probabilidad multinomial condicionada en el t3pico z_n .

GAP MAPS: Latent Dirichlet Allocation (LDA)

- Las probabilidades de cada palabra están parametrizadas por una matrix β de dimensiones $k \times V$, donde $\beta_{ij} = p(w^j = 1 \mid z^i = 1)$. Estos son unos valores fijos a ser estimados.

GAP MAPS: Latent Dirichlet Allocation (LDA)

- Dados los parámetros α y β , la función de probabilidad conjunta de una mezcla de tópicos θ , un conjunto de N tópicos z , y un conjunto de N palabras w está dado por:

$$p(\theta, z, w \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^N p(z_n \mid \theta) p(w_n \mid z_n, \beta), \quad (4)$$

donde $p(z_n \mid \theta)$ es θ_i para el único i tal que $z_n^i = 1$.

- Integrando sobre θ y sumando sobre z , obtenemos la distribución marginal de un documento:

$$p(w \mid \alpha, \beta) = \int p(\theta \mid \alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \right) d\theta. \quad (5)$$

GAP MAPS: Latent Dirichlet Allocation (LDA)

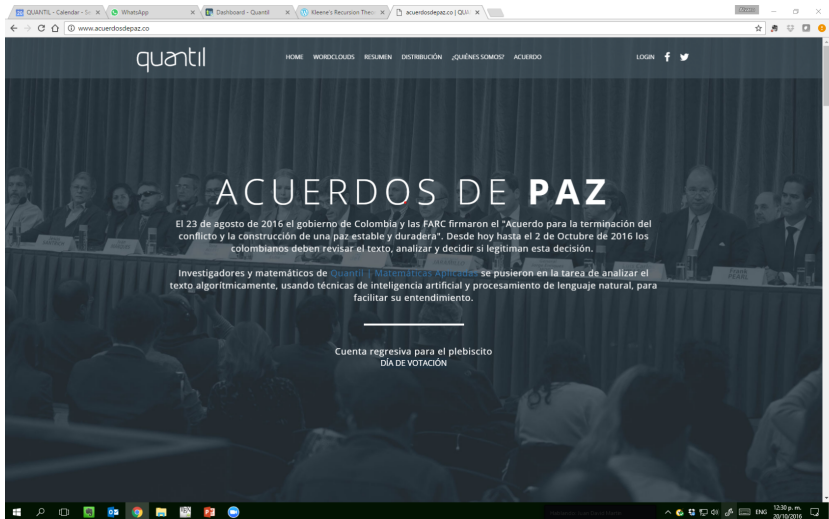
- Por último, tomando el producto de las probabilidades marginales de los documentos, obtenemos la probabilidad de un corpus:

$$p(c | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d$$

- Esta distribución es intractable, en general. A pesar de esto diferentes algoritmos de inferencia aproximada pueden ser considerados para LDA, como *Variational Approximation* y *Markov Chain Monte Carlo*.

Automatic summarization

- Cortesía Quantil (proyecto interno).



The screenshot shows a web browser window displaying the website 'www.acuerdosdepaz.co'. The page features a dark background with a photograph of a group of people seated at a long table, likely during a press conference or meeting. The main heading is 'ACUERDOS DE PAZ' in large, white, sans-serif capital letters. Below the heading, there is a paragraph of text in Spanish: 'El 23 de agosto de 2016 el gobierno de Colombia y las FARC firmaron el "Acuerdo para la terminación del conflicto y la construcción de una paz estable y duradera". Desde hoy hasta el 2 de Octubre de 2016 los colombianos deben revisar el texto, analizar y decidir si legitiman esta decisión.' A second paragraph follows: 'Investigadores y matemáticos de [Quantil](#) y [Matemáticas Aplicadas](#) se pusieron en la tarea de analizar el texto algorítmicamente, usando técnicas de inteligencia artificial y procesamiento de lenguaje natural, para facilitar su entendimiento.' Below this text is a horizontal line and a countdown timer: 'Cuenta regresiva para el plebiscito' followed by 'DÍA DE VOTACIÓN'. The website's navigation menu at the top includes 'HOME', 'WORDCLOUDS', 'RESUMEN', 'DISTRIBUCIÓN', '¿QUÉ NOS SOMOS?', and 'ACUERDO'. There are also social media icons for Facebook and Twitter, and a 'LOGIN' button. The browser's address bar shows the URL 'www.acuerdosdepaz.co'. The Windows taskbar at the bottom indicates the time is 12:30 p.m. on 20/10/2016.

Contents

- 1 Introducción
- 2 Detección de minería ilegal usando imágenes satelitales
- 3 Minería de datos médicos
- 4 Mercadeo basado en datos
- 5 Aterrizaje: Predicción del crimen
- 6 Aterrizaje: Minería de texto
- 7 Políticas Públicas**

Health Records

- Colombia has a mandatory competitive health insurance system.
- Payments are determined using information provided by HMOs.
- Health records: 400 million data items per year, 20 million individuals.
- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

Health Records

- Colombia has a mandatory competitive health insurance system.
- Payments are determined using information provided by HMOs.
- Health records: 400 million data items per year, 20 million individuals.
- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

Health Records

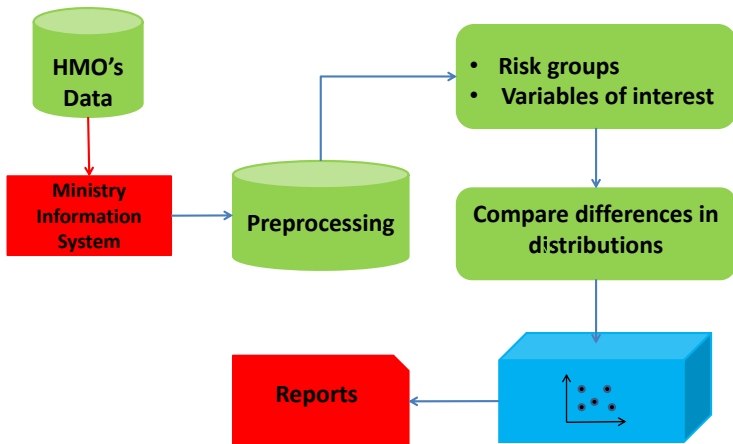
- Colombia has a mandatory competitive health insurance system.
- Payments are determined using information provided by HMOs.
- Health records: 400 million data items per year, 20 million individuals.
- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

Health Records

- Colombia has a mandatory competitive health insurance system.
- Payments are determined using information provided by HMOs.
- Health records: 400 million data items per year, 20 million individuals.
- Potential problems: Risk adjustment, solvency, fraud, data manipulation, misreport, errors, etc.

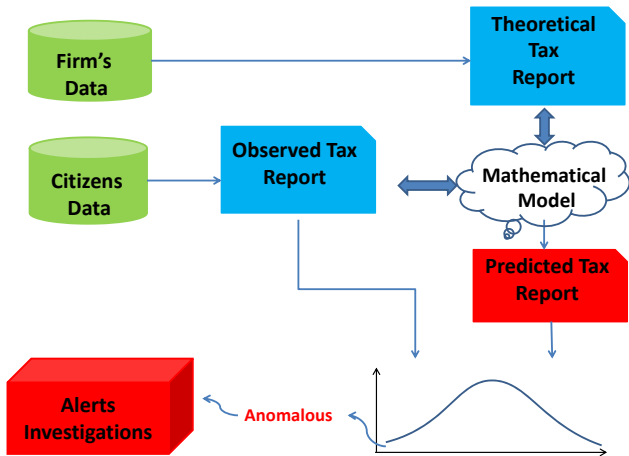
Health Records: Anomaly detection

- Courtesy of Ministry of Health and Social Protection.



Tax Evasion

- Problems: Detect likely cases of tax evasion.



Money Laundry

- Problem: Detect anomalous transactions that may be suspicious of money laundry.
- Confidentiality agreements forbids me of explaining further details.

Money Laundry

- Problem: Detect anomalous transactions that may be suspicious of money laundry.
- Confidentiality agreements forbids me of explaining further details.

Lawsuits Against the State

- Courtesy of National Agency for Judicial State Defence (ANDJE).

The screenshot displays the website of the National Agency for Judicial State Defence (ANDJE) in Spanish. The main heading is "ANDJE Probabilidad de que se otorgue una demanda contra el Estado". The interface includes a sidebar with various filters and a main content area showing a probability of 86.2%.

Filters (Left Sidebar):

- Resultado del primer fallo: Otro
- Subtipo 1: Falta judicial
- Se concede la apelación: Otro
- Subtipo 2: Pensión
- Nombre del Magistrado del tribunal: Otro
- Nombre del juez: Otro
- Numero del juzgado: 4
- El demandante es: Civil
- El demandado presenta apelación: No
- El demandado cambia de apoderado: No
- El demandante cambia de apoderado: No
- Numero del tribunal: Otro
- Entidad demandada: Caja de Retiro Militar
- Tiempo entre la admisión de la demanda y la primera sentencia, en días: 100
- Tiempo entre los hechos y la admisión de la demanda, en días: 200
- Duración total del proceso, en días: 1800
- Numero de pruebas presentadas por el demandante: 10

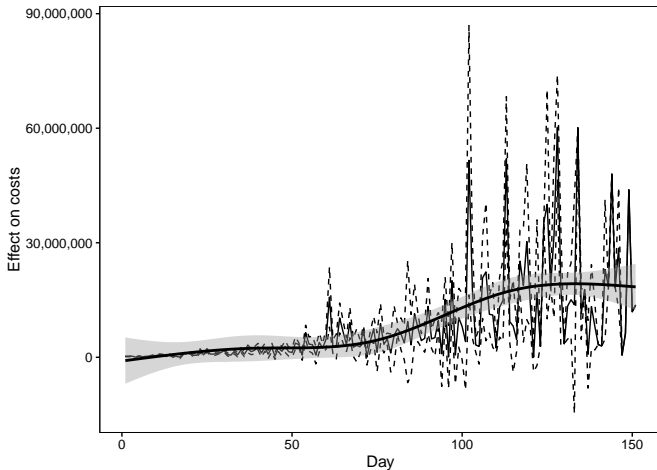
Main Content Area:

- Comenzó el proceso: Durante el proceso
- Probabilidad ajustada al comienzo del proceso: 86.2 %
- Logo of the Agencia Nacional de Defensa Jurídica del Estado

The browser's address bar shows the URL: `https://sparkstudio.com/quantil/sas12/ANDJE/`. The Windows taskbar at the bottom shows several open applications, including a file explorer, a Word document, and a presentation.

Predicting unnecessary hospitalizations

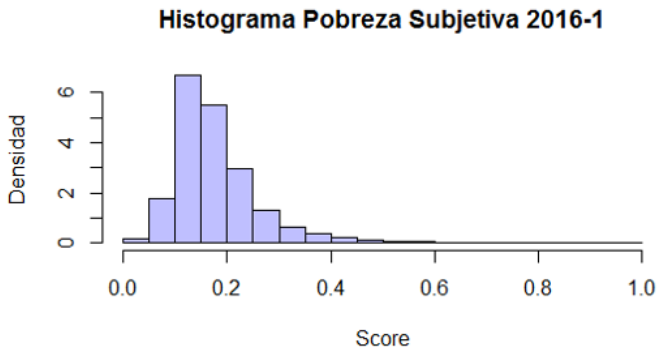
- Courtesy of Ministry of Health and Social Protection.



- Courtesy of Ministry of Health and Social Protection.

Subjective Poverty

- Courtesy of National Department of Statistics (DANE).



ariascos@uniandes.edu.co