

Una Perspectiva sobre la Minería de Datos

Alvaro J. Riascos Villegas
Universidad de los Andes y Quantil

Junio de 2018

Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Estamos viviendo una revolución científica?
- 6 Retos y reflexiones sobre la especie humana

Introducción

- ‘In god we trust, all others bring data’
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ($10^6 GB = 10^3 TB$) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.

Introducción

- 'In god we trust, all others bring data'
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ($10^6 GB = 10^3 TB$) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.

Introducción

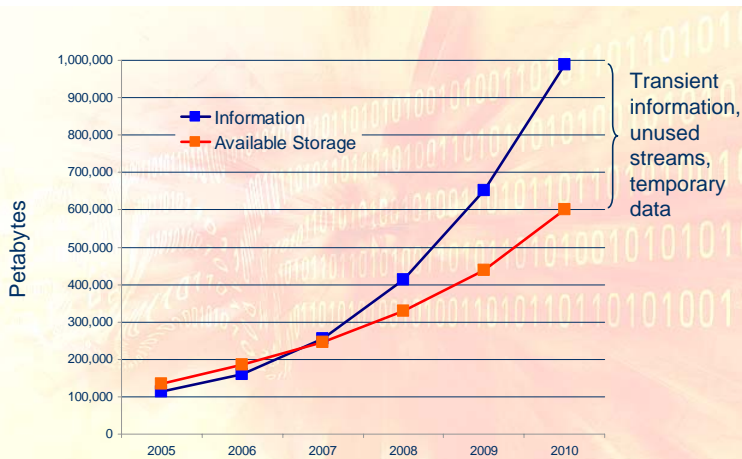
- ‘In god we trust, all others bring data’
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ($10^6 GB = 10^3 TB$) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.

Introducción

- ‘In god we trust, all others bring data’
- Con la llegada de la era de la computación, el mundo moderno ha experimentado un crecimiento en la cantidad de información disponible sin precedentes en la historia.
- El proyecto de Radio Telescópios de Australia, el acelerador de partículas CERN, etc. pueden generar varios petabytes ($10^6 GB = 10^3 TB$) de información diaria.
- La producción de información anual impresa, filmada, óptica y magnética requiere de 1,5 billones de GB. Esto es equivalente a 250 MB per capita.

- Una base de datos como la base de suficiencia que anualmente procesa el Ministerio de Salud tiene aproximadamente 460 millones de registros.

Introducción

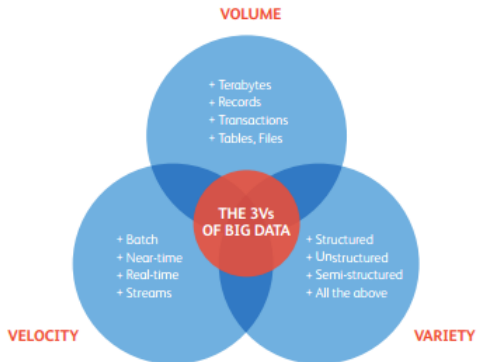


Introducción

Table 1: Worldwide production of original content, stored digitally using standard compression methods, in terabytes circa 1999.

Storage Medium	Type of Content	Terabytes/Year, Upper Estimate	Terabytes/Year, Lower Estimate	Growth Rate, %
<u>Paper</u>	Books	8	1	2
	Newspapers	25	2	-2
	Periodicals	12	1	2
	Office documents	195	19	2
	Subtotal:	240	23	2
<u>Film</u>	Photographs	410,000	41,000	5
	Cinema	16	16	3
	X-Rays	17,200	17,200	2
	Subtotal:	427,216	58,216	4
<u>Optical</u>	Music CDs	58	6	3
	Data CDs	3	3	2
	DVDs	22	22	100
	Subtotal:	83	31	70
<u>Magnetic</u>	Camcorder Tape	300,000	300,000	5
	PC Disk Drives	766,000	7,660	100
	Departmental Servers	460,000	161,000	100
	Enterprise Servers	167,000	108,550	100
	Subtotal:	1,693,000	577,210	55
TOTAL:		2,120,539	635,480	50

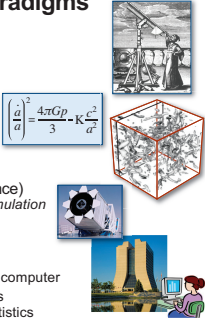
- Lo que toda esta información tiene en común:



- Una nueva forma de hacer ciencia.

Science Paradigms

- Thousand years ago:
science was **empirical**
describing natural phenomena
- Last few hundred years:
theoretical branch
using models, generalizations
- Last few decades:
a **computational** branch
simulating complex phenomena
- Today: **data exploration** (eScience)
unify theory, experiment, and simulation
 - Data captured by instruments or generated by simulator
 - Processed by software
 - Information/knowledge stored in computer
 - Scientist analyzes database/files using data management and statistics



The complex block contains several images: a historical figure with a telescope (top right), a mathematical equation $\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$ (middle), a 3D data visualization (middle right), a satellite dish (bottom left), and a modern scientist at a computer (bottom right).

- Tomado de *The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research*

Introducción

- El proceso típico de minería de datos.

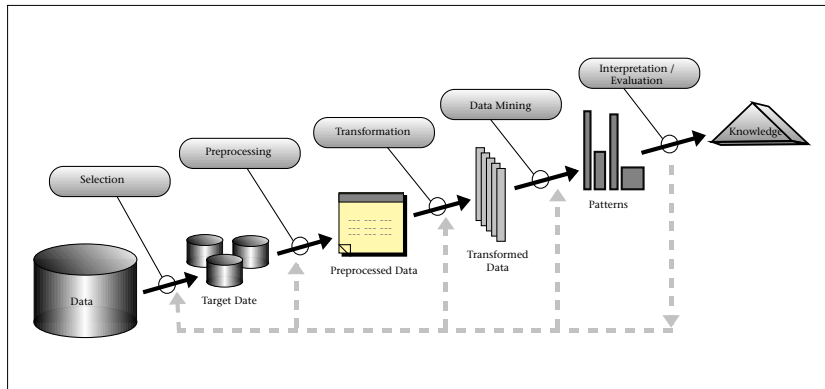
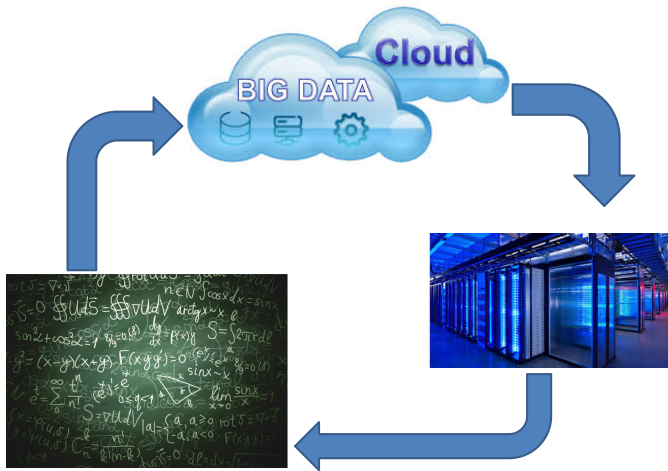


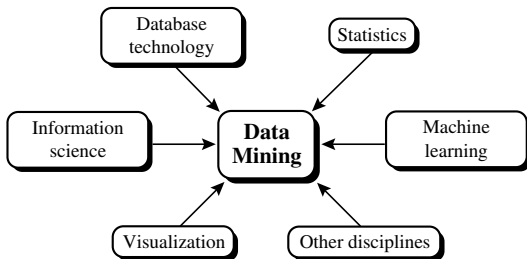
Figure 1. An Overview of the Steps That Compose the KDD Process.

- KDD: *Knowledge Discovery in Datasets*

Introducción: ¿En qué consiste esta revolución?



- Una disciplina interdisciplinaria.



Contenido

- 1 Introducción
- 2 Matemáticas en acción**
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Estamos viviendo una revolución científica?
- 6 Retos y reflexiones sobre la especie humana

Matemáticas en acción

- Un psicólogo de una prisión en California se acercó al Departamento de Estadística de Stanford con esto:

Handwritten mathematical calculations in Arabic script, showing a sequence of numbers and operations. The text is written on lined paper and includes various mathematical symbols and numbers, such as fractions and decimal points, arranged in several lines.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para f es muy grande.
- Para juzgar la plausibilidad de que una f en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo x seguía el símbolo y .
- Interpretamos esto como la probabilidad de transición de x a y en el idioma Inglés (i.e. una cadena de Markov). Denotamos por $M(x, y)$ esta frecuencia.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para f es muy grande.
- Para juzgar la plausibilidad de que una f en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo x seguía el símbolo y .
- Interpretamos esto como la probabilidad de transición de x a y en el idioma Inglés (i.e. una cadena de Markov). Denotamos por $M(x, y)$ esta frecuencia.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para f es muy grande.
- Para juzgar la plausibilidad de que una f en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo x seguía el símbolo y .
- Interpretamos esto como la probabilidad de transición de x a y en el idioma Inglés (i.e. una cadena de Markov). Denotamos por $M(x, y)$ esta frecuencia.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para f es muy grande.
- Para juzgar la plausibilidad de que una f en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo x seguía el símbolo y .
- Interpretamos esto como la probabilidad de transición de x a y en el idioma Inglés (i.e. una cadena de Markov). Denotamos por $M(x, y)$ esta frecuencia.

- Los estudiantes supusieron que el mensaje se podía decodificar de la forma:

$$f : \{\text{Símbolos del mensaje}\} \rightarrow \{\text{Alfabeto tradicional}\}$$

- El número de posibilidades para f es muy grande.
- Para juzgar la plausibilidad de que una f en particular podemos definir una medida de rendimiento.
- Tomaron un libro estándar (e.g. One Hundred Year of Solitude) y calcularon la frecuencia con la que después del símbolo x seguía el símbolo y .
- Interpretamos esto como la probabilidad de transición de x a y en el idioma Inglés (i.e. una cadena de Markov). Denotamos por $M(x, y)$ esta frecuencia.

- Ahora resolvemos este problema:

$$\max_{f \in \mathfrak{F}} \prod M(f(s_i), f(s_{i+1})) \quad (1)$$

- Este problema se resuelve usando al algoritmo de Metropolis - Hasting (un truco muy creativo!)
- Ahora, por qué esto funcionaría? Para ganar un poco de confianza en el modelo veamos el resultado en un caso conocido.

- Ahora resolvemos este problema:

$$\max_{f \in \mathfrak{F}} \prod M(f(s_i), f(s_{i+1})) \quad (1)$$

- Este problema se resuelve usando al algoritmo de Metropolis - Hasting (un truco muy creativo!)
- Ahora, por qué esto funcionaría? Para ganar un poco de confianza en el modelo veamos el resultado en un caso conocido.

- Ahora resolvemos este problema:

$$\max_{f \in \mathfrak{F}} \prod M(f(s_i), f(s_{i+1})) \quad (1)$$

- Este problema se resuelve usando al algoritmo de Metropolis - Hasting (un truco muy creativo!)
- Ahora, por qué esto funcionaría? Para ganar un poco de confianza en el modelo veamos el resultado en un caso conocido.

- Revolver todos los símbolos de este párrafo de Hamlet:

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

- Metropolis Hasting despues de n iteraciones:

```
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 IUSOPHE 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 RODIORLOS R O RODIO HA OETO HIOQUIT ERSOWHI HILO ETOSRDMIL
1100 ISTILOHANMITOHANOT ODILO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1200 ISTILOHANMITOHANOT ODILO LOS TOT ODIOTHATOEROTHIOQUIRTE SOWHITHILOTEROS DMIL
1300 ISTILOHARMITOHAROT ODILO LOS TOT ODIOTHATOENOTHIOQUINTE SOWHITHILOTENOS DMIL
1400 ISTILOHAMRITOHAMOT OFILO LOS TOT OFIOTHATOENOTHIOQUINTE SOWHITHILOTENOS FRIL
1600 ESTEL HAMRET HAM TO CE OL SOT TO CE THAT IN THE QUENTIOS WHEHEL TIN SOCREL
1700 ESTEL HAMRET HAM TO BE OL SOT TO BE THAT IN THE QUENTIOS WHEHEL 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
```

- Resultado decodificación mensaje presos:

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

Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas**
- 4 Computación y sus límites
- 5 ¿Estamos viviendo una revolución científica?
- 6 Retos y reflexiones sobre la especie humana

Límites al aprendizaje de máquinas

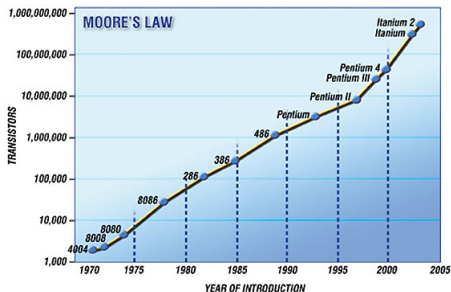
- Teorema de imposibilidad (no hay almuerzo gratis): No hay forma de aprender sin hacer alguna hipótesis sobre la estructura del fenómeno que se quiere estudiar (una teoría!).

Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites**
- 5 ¿Estamos viviendo una revolución científica?
- 6 Retos y reflexiones sobre la especie humana

Computación y sus límites

- Pocas dudas sobre la capacidad de calcular de los computadores.
- Moore's law



- Sin embargo, existen algunos límites que se derivan de nuestro conocimiento actual de la física y las matemáticas (Lloyd, S. (2000). The Ultimate Physical Limits to Computation):
 - 1 La capacidad de almacenamiento del Universo visible es: 10^{92} bits.
 - 2 Número de operaciones realizadas desde el Big Bang: 10^{122} .
 - 3 La totalidad del Universo si fuera una computadora cuántica no podría encontrar una cadena de 500 bits en tiempo polinomial para toda cadena posible de 500 bits.
 - 4 Esto probablemente también traería consecuencias enormes en la industria (vease The Future of Physics)

- Sin embargo, existen algunos límites que se derivan de nuestro conocimiento actual de la física y las matemáticas (Lloyd, S. (2000). The Ultimate Physical Limits to Computation):
 - 1 La capacidad de almacenamiento del Universo visible es: 10^{92} bits.
 - 2 Número de operaciones realizadas desde el Big Bang: 10^{122} .
 - 3 La totalidad del Universo si fuera una computadora cuántica no podría encontrar una cadena de 500 bits en tiempo polinomial para toda cadena posible de 500 bits.
 - 4 Esto probablemente también traería consecuencias enormes en la industria (vease The Future of Physics)

- Sin embargo, existen algunos límites que se derivan de nuestro conocimiento actual de la física y las matemáticas (Lloyd, S. (2000). The Ultimate Physical Limits to Computation):
 - 1 La capacidad de almacenamiento del Universo visible es: 10^{92} bits.
 - 2 Número de operaciones realizadas desde el Big Bang: 10^{122} .
 - 3 La totalidad del Universo si fuera una computadora cuántica no podría encontrar una cadena de 500 bits en tiempo polinomial para toda cadena posible de 500 bits.
 - 4 Esto probablemente también traería consecuencias enormes en la industria (vease The Future of Physics)

- Sin embargo, existen algunos límites que se derivan de nuestro conocimiento actual de la física y las matemáticas (Lloyd, S. (2000). The Ultimate Physical Limits to Computation):
 - 1 La capacidad de almacenamiento del Universo visible es: 10^{92} bits.
 - 2 Número de operaciones realizadas desde el Big Bang: 10^{122} .
 - 3 La totalidad del Universo si fuera una computadora cuántica no podría encontrar una cadena de 500 bits en tiempo polinomial para toda cadena posible de 500 bits.
 - 4 Esto probablemente también traería consecuencias enormes en la industria (vease The Future of Physics)

- Sin embargo, existen algunos límites que se derivan de nuestro conocimiento actual de la física y las matemáticas (Lloyd, S. (2000). The Ultimate Physical Limits to Computation):
 - 1 La capacidad de almacenamiento del Universo visible es: 10^{92} bits.
 - 2 Número de operaciones realizadas desde el Big Bang: 10^{122} .
 - 3 La totalidad del Universo si fuera una computadora cuántica no podría encontrar una cadena de 500 bits en tiempo polinomial para toda cadena posible de 500 bits.
 - 4 Esto probablemente también traería consecuencias enormes en la industria (vease The Future of Physics)

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Computación y sus límites

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

- Juego de Rabin (1957). Dos jugadores A y B , y un conjunto $W \subset \mathbb{N}^3$.
- Reglas: En tres rondas, A y B escogen tres números naturales.
- El resultado final es una tripla (a, b, c) .
- Si $(a, b, c) \in W$ gana B .
- ¿Qué sabemos sobre este juego?
 - 1 Es determinado (inducción hacia atrás).
 - 2 Existe W decidible tal que el juego es determinado en favor de B .
 - 3 B no puede calcular (con ningún computador incluso hipotético) su estrategia ganadora!
- La teoría es imprescindible, no basta con capacidad de cómputo.

Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Estamos viviendo una revolución científica?**
- 6 Retos y reflexiones sobre la especie humana

¿Estamos viviendo una revolución científica?

- ¿ Es el fin de la teoría?
 - Anderson: The End of Theory: The Data Deluge Makes the Scientific Method Obsolete
 - Napoletani, Panza, Struppa: Agnostic Science. Towards a Philosophy of Data Analysis.
 - Big data: A revolution that will transform how we live work and think.
- Chomsky es aún más radical: Esta nueva ciencia de moler grandes volúmenes de datos no es ciencia.

¿Estamos viviendo una revolución empresarial?

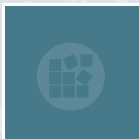
Archetype of disruption	Domains that could be disrupted
Business models enabled by orthogonal data	<ul style="list-style-type: none">▪ Insurance▪ Health care▪ Human capital/talent
Hyperscale, real-time matching	<ul style="list-style-type: none">▪ Transportation and logistics▪ Automotive▪ Smart cities and infrastructure
Radical personalization	<ul style="list-style-type: none">▪ Health care▪ Retail▪ Media▪ Education
Massive data integration capabilities	<ul style="list-style-type: none">▪ Banking▪ Insurance▪ Public sector▪ Human capital/talent
Data-driven discovery	<ul style="list-style-type: none">▪ Life sciences and pharmaceuticals▪ Material sciences▪ Technology
Enhanced decision making	<ul style="list-style-type: none">▪ Smart cities▪ Health care▪ Insurance▪ Human capital/talent

¿Estamos viviendo una revolución empresarial?

Machine learning has broad applicability in many common work activities

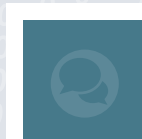
Percent of work activities that require:

Recognizing
known patterns



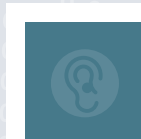
99%

Generating
natural language



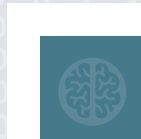
79%

Understanding
natural language



76%

Enhanced sensory
perception



59%

Optimizing and
planning



33%

MCKINSEY GLOBAL INSTITUTE

McKinsey&Company

¿Estamos viviendo una revolución empresarial?

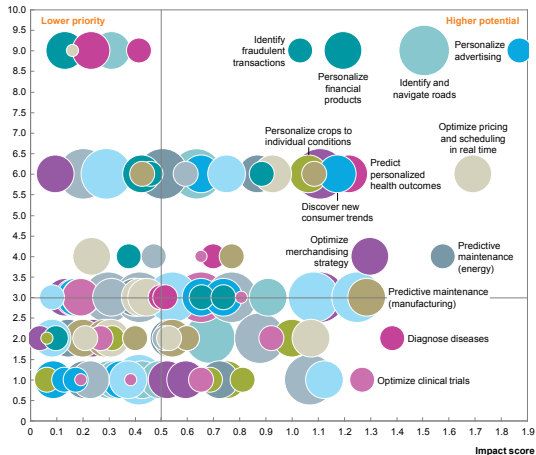
Machine learning has broad potential across industries and use cases

○ Size of bubble indicates variety of data (number of data types)



Volume

Breadth and frequency of data



SOURCE: McKinsey Global Institute analysis

¿Estamos viviendo una revolución empresarial?

- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX
- Un algoritmo de deep learning (AlphaGo) le gana por la primera vez a un experto en Go.
- Véase informe de McKinsey Global Institute: The age of analytics.

¿Estamos viviendo una revolución empresarial?

- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX
- Un algoritmo de deep learning (AlphaGo) le gana por la primera vez a un experto en Go.
- Véase informe de McKinsey Global Institute: The age of analytics.

¿Estamos viviendo una revolución empresarial?

- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX
- Un algoritmo de deep learning (AlphaGo) le gana por la primera vez a un experto en Go.
- Véase informe de McKinsey Global Institute: The age of analytics.

¿Estamos viviendo una revolución empresarial?

- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX
- Un algoritmo de deep learning (AlphaGo) le gana por la primera vez a un experto en Go.
- Véase informe de McKinsey Global Institute: The age of analytics.

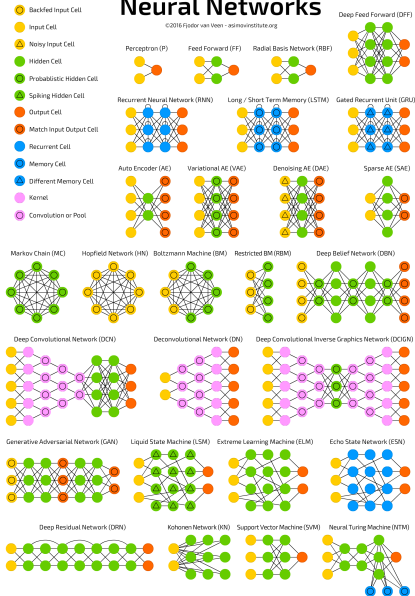
¿Estamos viviendo una revolución empresarial?

- Farecast (microsoft pagó \$115 USD): Predicciones de cuando comprar tiquetes con base en 200 billones de ofertas en Internet.
- Cada hora se cargan a facebook más de 100 millones de fotos. Los modelos de reconocimiento de imágenes utilizan los tags como muestras de entrenamiento.
- Google entrena su corrector de texto cada que hacemos click en: Quizás quisiste decir: XXX
- Un algoritmo de deep learning (AlphaGo) le gana por la primera vez a un experto en Go.
- Véase informe de McKinsey Global Institute: The age of analytics.

Aprendizaje Profundo (*Deep Learning*)

A mostly complete chart of Neural Networks

©2016 Fjodor van Veen - asensei.vrstate.org



La revolución empresarial: *deep learning*

- Una red típica tiene entre 5 y 7 capas ocultas.
- 1000 neuronas por capa.
- Es necesario estimar miles de parámetros.
- Se necesitan 1000 horas de grabación.
- Un super computador se demora dos semanas estimando.

La revolución empresarial: *deep learning*

- www.clarifai.com

Aprendizaje Profundo: Inteligencia básica AlphaZero

Game	White	Black	Win	Draw	Loss
Chess	<i>AlphaZero</i>	<i>Stockfish</i>	25	25	0
	<i>Stockfish</i>	<i>AlphaZero</i>	3	47	0
Shogi	<i>AlphaZero</i>	<i>Elmo</i>	43	2	5
	<i>Elmo</i>	<i>AlphaZero</i>	47	0	3
Go	<i>AlphaZero</i>	<i>AG0 3-day</i>	31	–	19
	<i>AG0 3-day</i>	<i>AlphaZero</i>	29	–	21

Table 1: Tournament evaluation of *AlphaZero* in chess, shogi, and Go, as games won, drawn or lost from *AlphaZero*'s perspective, in 100 game matches against *Stockfish*, *Elmo*, and the previously published *AlphaGo Zero* after 3 days of training. Each program was given 1 minute of thinking time per move.

- Extracción de tópicos:

<http://seixas.quantil.co/gapmaps/api/site/>

- Resúmenes automáticos: <http://www.acuerdosdepaz.co/>

- Extracción de tópicos:
<http://seixas.quantil.co/gapmaps/api/site/>
- Resúmenes automáticos: <http://www.acuerdosdepaz.co/>

- Lo que hace particularmente interesante esta revolución empresarial es que está surgiendo de la mano de la academia y los empresarios.

Contenido

- 1 Introducción
- 2 Matemáticas en acción
- 3 Límites al aprendizaje de máquinas
- 4 Computación y sus límites
- 5 ¿Estamos viviendo una revolución científica?
- 6 Retos y reflexiones sobre la especie humana

Retos

- La construcción de robots autónomos está a la vuelta de la esquina, con los rasgos más básicos de los seres vivos:
 - Robots con un propósito.
 - Que se puedan reproducir (su código).
 - Adquirir energía.
 - Autonomía, etc.

Retos

- La construcción de robots autónomos está a la vuelta de la esquina, con los rasgos más básicos de los seres vivos:
 - Robots con un propósito.
 - Que se puedan reproducir (su código).
 - Adquirir energía.
 - Autonomía, etc.

Retos

- La construcción de robots autónomos está a la vuelta de la esquina, con los rasgos más básicos de los seres vivos:
 - Robots con un propósito.
 - Que se puedan reproducir (su código).
 - Adquirir energía.
 - Autonomía, etc.

Retos

- La construcción de robots autónomos está a la vuelta de la esquina, con los rasgos más básicos de los seres vivos:
 - Robots con un propósito.
 - Que se puedan reproducir (su código).
 - Adquirir energía.
 - Autonomía, etc.

Retos

- La construcción de robots autónomos está a la vuelta de la esquina, con los rasgos más básicos de los seres vivos:
 - Robots con un propósito.
 - Que se puedan reproducir (su código).
 - Adquirir energía.
 - Autonomía, etc.

Retos

- La construcción de robots autónomos está a la vuelta de la esquina, con los rasgos más básicos de los seres vivos:
 - Robots con un propósito.
 - Que se puedan reproducir (su código).
 - Adquirir energía.
 - Autonomía, etc.

- Es indispensable programarles valores morales.
- Privacidad y discriminación.
- La posibilidad de hackear todos los carros autónomos de un ciudad y causar una gran catástrofe.
- Los retos de la teoría de la competencia en la era de la información.
- Noticias falsas.
- El empleo y la brecha digital.

- Es indispensable programarles valores morales.
- Privacidad y discriminación.
- La posibilidad de hackear todos los carros autónomos de un ciudad y causar una gran catástrofe.
- Los retos de la teoría de la competencia en la era de la información.
- Noticias falsas.
- El empleo y la brecha digital.

- Es indispensable programarles valores morales.
- Privacidad y discriminación.
- La posibilidad de hackear todos los carros autónomos de un ciudad y causar una gran catástrofe.
- Los retos de la teoría de la competencia en la era de la información.
- Noticias falsas.
- El empleo y la brecha digital.

- Es indispensable programarles valores morales.
- Privacidad y discriminación.
- La posibilidad de hackear todos los carros autónomos de un ciudad y causar una gran catástrofe.
- Los retos de la teoría de la competencia en la era de la información.
- Noticias falsas.
- El empleo y la brecha digital.

- Es indispensable programarles valores morales.
- Privacidad y discriminación.
- La posibilidad de hackear todos los carros autónomos de un ciudad y causar una gran catástrofe.
- Los retos de la teoría de la competencia en la era de la información.
- Noticias falsas.
- El empleo y la brecha digital.

- Es indispensable programarles valores morales.
- Privacidad y discriminación.
- La posibilidad de hackear todos los carros autónomos de un ciudad y causar una gran catástrofe.
- Los retos de la teoría de la competencia en la era de la información.
- Noticias falsas.
- El empleo y la brecha digital.

- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steven Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*
- Muchos estan preocupados: Elon Musk (Paypal, Tesla, etc), Centre for the Study of Existential Risk.

- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steven Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*
- Muchos estan preocupados: Elon Musk (Paypal, Tesla, etc), Centre for the Study of Existential Risk.

- Steve Omohundro (Parc Forums): Es necesario programarles valores y principios morales a los robots!
- Steven Hawking: *Success in creating AI will be the biggest event in human history. Unfortunately it might also be the last, unless we learn how to avoid the risks.*
- Muchos estan preocupados: Elon Musk (Paypal, Tesla, etc), Centre for the Study of Existential Risk.

- Center for the study of existential risks:
<https://www.cser.ac.uk/>
- La religión de los datos: Homo Deus. Una breve historia del mañana. Yuval Noah Harari.

- Center for the study of existential risks:
<https://www.cser.ac.uk/>
- La religión de los datos: Homo Deus. Una breve historia del mañana. Yuval Noah Harari.