

a16 Informática Biomédica

a16-00 Artificial Intelligence in Medicine

Prof. Víctor Maojo

Las herramientas de AI no toman una decisión final, porque las compañías que las crean no pueden asumir los errores

Por ello la responsabilidad recae sobre el médico

Turing propone en los 50's máquinas que pueden pensar.

Turing test → todavía ninguna máquina la ha pasado

- Cybernetics - 1943 - 1960
 - Cybernetics - 1943 - 60
 - Shannon
 - Reales genios
- 1956 → MIT, Stanford, Mellon
 - Dartmouth college
 - Principio de IA
 - Querían hacer algo distinto, asustados por cybernetics
- Cyborg → palabra griega Cybernos, timonel.
 - Cómo controlar los barcos
 - 17 doctorado
 - 18 profesor de MIT
 - Coincidió con gente de Harvard
 - que trabajaba con sífilis
 - problema en la médula espinal
 - en 20 o 50 años, se vuelven locos, tienen temblores y mueren
 - Las reglas que controlan los impulsos en la médula son similares a un barco
 - Impulsos eléctricos con feedback
 - Qué mueve el cuerpo? → energy
 - No energía, sino bio información
 - Tareas consideradas inteligentes

- reconocimiento visual de patrones
 - razonamiento
 - manejo de memoria
 - decisión
 - vision
 - sentido común
 - ...
- Herverd Simon : en 1981 , dentro de 25 años habrá
 - No hacer predicciones dentro de tu espacio de tiempo de vida. Hazlas para después de morir.
- Tipos de AI
 - Symbolic AI
 - Mathematics, Logic → Mind
 - Conciencia está en un área concreta
 - Knowledge Engineering
 - **Knowledge**
 - Connectonist AI
 - Otra rama de AI
 - La inteligencia es la consecuencia de la conexión de 10^{12} neuronas
 - Cada neurona se conecta con otras 10K neuronas
 - ML, DL, Data Mining
 - **DATA**
- 1956 estuvieron trabajando en
 - ajedrez
 - lógica
- Ordenadores en hospitales 1970
- 2 áreas
 - Artificial Intelligence
 - aviones no aletean las alas, no es eficiente
 - Tareas similares hechas por personas
 - No intenta imitar el cerebro, solo el
 - Cognitive Science
 - **Sí intenta imitar cómo funciona un cerebro en un ordenador**
 - No es construir sistemas, también es el estudio de cómo funciona el cerebro y como procesamos información

- Pavlov : Conductismo: Estímulo → Respuesta
 - Miedo a tener miedo
- 1960 → razonamiento en términos computacionales. Términos lógicos
 - ¿Cómo el doctor?
 - Cómo toma una decisión
 - Recolecta de datos
 - Detecta Patrones
 - Maneja probabilidades
- Ordenadores
 - ENIAC , universidad de pennsylvania
 - Homer Warnier - Reed Gardner, university of Utah
 - Akron Children's Hospital Ohio

The new England journal of medicine

Estudiante de harvard, porque la

Podemos construir un ordenador que sustituya a un doctor en algunas tareas

1960 → Star Trek Dr McCoy

Pasan el sistema y da el diagnóstico y el tratamiento. Inspiró a mucha gente

Medica reasoning

| Los médicos piensan diferente a los ingenieros

- Ideas clásicas
 - ojo clínico como un arte
 - 2+2 no son 4 en medicina
 - fórmulas complejas con muchas variables que desconocemos
 - gran memoria e intuición especial
- Razonamiento médico es complejo , basado en manejo de información y puede ser estudiado y modelado
 - **Diagnosis y razonamiento médico. Es una clasificación , identificar a un paciente con una enfermedad en concreto.**

- Predicción
- Monitorización
- Planear
- **Clasificación**
- Crear una lista de **hipótesis** y aplicar **razonamiento lógico**
 - Deducción
 - Lógica formal, si $p \rightarrow q$, p entonces q
 - Inducción
 - scientific research
 - recolectamos observaciones
 - tomamos una teoría
 - no asegura que es cierto
 - siempre podemos encontrar casos que no siguen la regla/teoría
 - Boot floats
 - Reglas que extraemos de lo que ha pasado
 - si probamos todas las
 -
 - Abductive
- Sistemas de IA

Woody Allen

- La úlcera gástrica estaba relacionada con estrés
 - El tratamiento era psicológico
 - y medicinal
- Todos los datos apuntaban a que era el estrés
- Un tío leyó muchos papers de úlceras gástricas
 - vio que era una bacteria
 - él cogió un poco de la bacteria y se la comió, comprobando que tenía una bacteria. *Helicobacter pylori*. Luego tomó antibióticos y se curó 15 días antes.
 - Rohelgen - mano de su mujer 15 minutos.
 - probó la vacuna Con el hijo del jardinero
 - Analizó Pubmed (Public med line)
 - 25M referencias de papers
 - PMC \rightarrow PubMedCenter tiene 1M de papers gratis
 -
- Estudio de ondas cerebrales

- cientos de letras
- identifica y contar las veces que ha escuchado la letra d
- ERP → Even Relation Potentials
- P300
- extraversión - correlación
- Misma correlación con introversión
- **Correlación no implica causa**
-
- WHO - World Hospital Organization
 - Gente que bebe diet coke, tienen 3 veces más ictus (Correlación)
 - la gente que bebe coca light suele ser obesa (No implica causa)

| Correlación no implica causa

Proceso de diagnóstico

- Generación de hipótesis
 - Número de H
- refinamiento de hipótesis
 - eliminar H
 - para ello necesitan información
 - tiene su conocimiento y necesita seguir un **proceso deductivo lógica - Proceso hipotético-deductivo**
 - pide pruebas médica, radiografías, etc
- test de diagnóstico
- razonamiento causal
- verificación de diagnóstico

Razonamiento

AI - symbolic vs connectionist

Basado en lógicas

podemos representar

- Comparado con sistemas lógicos (premisas, reglas de inferencias y conclusiones)
- Razonamiento condicional
 - Si $a \rightarrow B$
 - categorizaciones :

Tipos de Knowledge

Johnson-Laird, profesor de Yale.

representación interna más poderosa , inconsciente que puede ser traducida a lenguaje natural.

Con experiencia , el razonamiento de procesado son automatizados :

1. Interpretado
2. Compilado
3. Automático

No se puede extraer todo el conocimiento de una persona.

Porque muchos procesos (después de 15 años) son automáticos

Muchos procesos son subconscientes

Por tanto no se puede construir un ordenador que extraiga la información y trabaje como esa persona

Los tiradores del viejo oeste disparan después , cuando tu reaccionas , los movimientos son más rápidos y más precisos.

- Parte cortical → Consciente
- subcortical → automático, inconsciente
 - crearon la canción de yesterday y en un sueño
-

Desarrollo de la experiencia

Diferencias cognitiva entre principiantes y expertos

Los expertos son capaces de identificar más de 100K patrones visuales

- Los razonamientos estratégicos son específicos a determinados campos
 - Nuestra forma de pensar es distinta a los biólogos, abogados, políticos
 - Los políticos no necesitan pensar
 - Diferentes estrategias
 - El área más difícil para computer science es Médicos

Niveles de expertise

Los médicos expertos, reconocerán patrones de forma muy rápida

Un principiante, deducirá, esto está cerca de este nervio,

| Las personas con experiencia saben lo que hacer de forma muy rápida

Heurísticas en medicina

1. Representación: probabilidad de un evento es estimado comparado este evento con una población
2. Availability: la probabilidad de un evento es recordado por los easiness para recordarlo
3. Anchoring y ajuste: físicos carry out una explotación inicial y más tarde , ellos ajustan de acuerdo a las características ella encuentra en el paciente hasta alcanzar una conclusión final

Los médicos suelen recordar cosas que han leído hace poco tiempo.

| Los errores médicos. El sistema médico es una de las causas más altas de muerte

a16-03 Artificial Intelligence in Medicine

Victor Maojo

2019-10-17

Sistemas expertos

| Esta era la AI para medicina durante muchos años

Los doctores acumulan su conocimiento a través de libros, journals, práctica, mentores.

- **IA**
 - Un Sistema experto intenta expresar este conocimiento
 - CDS → Computer Decisión Support System
- Causas de problemas:
 - Síntomas: subjetivo al paciente , dolor de cabeza, ...
 - Signos: algo objetivo que podemos medir

Fisiológico → Es el comportamiento natural del cuerpo

Patología → Problema, enfermedad

El doctor reconoce patrones visuales

- **IA**
 - Image processing usando redes neuronales
 - Razonamiento basado en causas

El problema es que no tenemos ningún sistema que pueda

Por qué los sistemas expertos han emergido?

- Expertos no pueden fácilmente manejar conceptos probabilísticos in la mayoría de areas
 - para construir sistemas que hablan el mismo lenguaje natural de los usuarios
- Sistemas capaces de explicar por qué toma esta decisión, no caja negras
- Extraer los métodos de conocimiento y razonamiento de los mejores especialistas
- Necesitamos representar

Características de sistemas expertos

- Separación entre el conocimiento del problema. **Base de conocimiento - Motor de inferencia**
 - conocimiento ↔ control)
 - Representar conocimiento
- Explicar conclusiones encontradas en el sistema
- usar el proceso de razonamiento similar a esos hechos por el experto
- mecanismos para inferir el grado de certeza (probabilidad) de las conclusiones alcanzadas
- Interfaz amigable, con la habilidad de comunicarse con el usuario

- **Knowledge Engineering** (*Joshua Lederberg*)

- Nobel a los 32 años

DENDRAL

- sistema capaz de elucidar la estructura molecular de algunas moléculas químicas.

Stanford (MYCIN)

Casimir Kulikowski, Ed Feigenbaum y Ted Shortliffe → Creadores de los primeros sistemas expertos. CASNET, DENDRAL y MYCIN

MYCIN

CASNET realmente fue el primero, pero nadie lo utilizó. Demasiado nuevo.

- Sistema basado en reglas, para las selecciones de terapia in enfermedades infecciosas (septicemia and meningitis)

- IF (premisas) THEN (ACCIÓN CONSECUENTE)

- IF (fiebre , moqueo, dolor de cabeza) THEN “tienes alta probabilidad de tener gripe”

- Uso de Heurísticas
 - Método para cortar y analizar una sola parte del conocimiento del sistema o de una persona.
- Nueva forma de tratar la certeza - Factor de certeza (Cos)
- KB Knowledge Base con una interfaz muy amigable (para aquel tiempo)
- Un éxito tecnológico enorme, pero nunca fue usado en una rutina clínica

Era un experimento y no una solución final para los doctores

Trabajaron con doctores, pero eran científicos no doctores que trabajaban en el hospital todos los días

Los doctores reales, dijeron que MYCIN no aportaba valor. Hacía lo mismo que ellos pero en mucho más tiempo

CFs

El CF-model is basado en la teoría de la probabilidad.

$P(Y|X)$

MYCIN pregunta al doctor, por una terminal las premisas para poder inferir el camino a seguir en el árbol de decisión

Fue desarrollado en LISP

Estaba basado en el mecanismo de backtracking.

El conocimiento es representado en un árbol, representando las reglas del sistema

Tenía más de 400 reglas

MYCIN's Goal Rule =

INTERNIST (People , Myers, Miller)

Especialista de todo. Cuando

La idea era crear un oráculo que pudiera manejar miles de reglas

Costó X años en construir la base de conocimiento

Estaban aprendiendo en cómo las personas pesaban.

Fue comercial , pero solo como referencia.

- Puemet: Bibliografía 26M de revistas. gen p53 + cancer

CASNET

Glaucoma, presión de líquidos en el ojo

CASNET tiene 3 capas

- Observaciones (MYCIN se quedaba en este nivel + algunas heurísticas)
- Pathophysiological States
 - conocimiento más concreto
- Categoría de enfermedades

Basado en una red de causa efecto

El problema principal (en medicina) es Incertidumbre

No sabemos como funcionan las enfermedades, el organismo, etc

Sistema experto en ingeniería →

en ciencia natural → no podemos cambiar los mecanismos, solo descubrir

En la siguiente clase → Limitations of first expert system I

Crear una ontología para la clase de Victor. Solos o en grupos de dos, todavía por decidir

a16-05 technical

David Pérez del Rey

technical

Data types

En hospitales hay miles, bases de datos, réplicas, vistas, cargas el modelo

HL7 messages → XML → Fire (Json)

H8 →

Loin → laboratorio

En todos los hospitales de

Nunca podemos trabajar con datos de pacientes.

Se trabaja con catálogos

Frecuencia

- live.trinetx.com
- 10M de pacientes
- 21 hospitales que donan estos datos
- en europa nunca lo harían
- behçet's disease , buscar icd-10-
 - para conseguir el nombre importante
- puedes ver demographics
- **Feasibility** = Factibilidad
 - clinical research department
 - pregnant male
 -
- Side selection

En el hospital hay 5 o 6 personas que leen los informes y anotan / codifican la información.

- NLP ejemplos
 - para reclutamiento, puede funcionar, porque al final lo tiene que aceptar el doctor
 - TNM status , Tumor status {1,2,3}

1. Están abiertos a dejar la herramienta para trabajar

2. Ver un ejemplo real

No es big-data, real-world data

Evaluation

- 10 oct → intro and real platform example
- 24 oct - ML ranking
- 31 oct - lab coding
- 7 nov - search engine hands-on
- 21 nov - group presentation

Trabajos 50% de la nota

- 10% → Lab coding (in class / presentation)
- 10% search engine (in class / presentation)
- 30% ML Ranking (report / presentation)

a16-06 Real World Data - Secondary Use of Clinical Data

- David Pérez del Rey
- 20191024
- `02 Secondary-use-of-clinical-data.pdf`

- 10.10 intro real platform example
- 24.10 ml ranking
- 31.10 lab coding
- 7.9 Search engine hands-on
- 21.9 Group presentation

50% mark

- 10% lab coding
- 10% search engine
- 30% ML Ranking (Report and presentation)

ML Ranking

Autosugerencia, ordenar los resultados según la aproximación.

Si buscamos Diabetes → sale una lista ordenada

3 Aproximaciones

- 1.
- 2.
- 3.

LOINC

Get Started - LOINC

Terms

- | !! OJO
- **LOINC:** Logical Observation Identifiers Names and Codes

a16-07 - LOINC hands-on

- David Perez del Rey
- 20191031
-

Get Started - LOINC

Lista de LOINC

a16-08 Search Engine hands-on

- David Perez del Rey
- 20191107
-

Web-based Models

- Search engines IR engines based on models
- ...

Other challenges in Web Search Engines

Typical crawler-indexer architecture

Users → Interface (google) → Query Engine → index ← indexer ← crawler

2019 hay infinitas páginas + páginas dinámicas.

User interface

-
- SERP Heat Map →
- SEM → apuestas por aparecer en las partes calientes de la web

Recomendaciones prácticas para SEO

- Todos los términos deben aparecer en la página
- Localizar correctamente esos términos en las etiquetas html (h1, title , ...)
- Están juntos?
- Cuantas veces están mencionados → Keyword density.
 - una densidad
- **Title Tag**
- Link popularity
- keyword density
- content related to your niche

-
- dinámico
- sitemap
- Submitting articles with links (community manager)
- Responsive website

Too many links on a page > 50-100

repeat titles and descriptions

keyword stuffing

all pages must be in 4-8 links de profundidad

no desviar la popularidad a páginas legales (Terms of services)

Search rating quality programmers

Gold standard creation

10K people around the world contracted

Search quality rating guidelines

PQ Rating and explanation

| Gente contratada para Anotar el corpus de las búsquedas

Práctica

- 10% of the assignment
- Elegir una empresa pequeña en el area de AI y Health
 - Auditoría SEO del sitio seleccionado comparado a 1-2 competidores
- Analyze SEO related issues
 - basic
 - descriptive titles?
 - Descriptive URLs?
 - Indexed Keywords
 - Other issues
 - Manual inspection of website keywords (compare with competitors)
 - Technology - PageSpeed Insights (compare with competitors and propose actions)
 - Potential Keywords - Google Trends (limited data in Spain)
- Send 1-2 page report

Seo Tools

- General
 - Google webmaster
 - bing webmaster
 - seo

Video youtu.be/EI3IZFGERbM

Comparar una de estas dos empresas NLP para health data:

- iomed.es
- savanamed.com
-

Biomedical Informatics Assignment 3

a16-09 Ontologías.

- Victor Maojo
- 20191114
- ☐

Práctica

Ontologías

Causal relations in ABEL (Patil)

Lecciones aprendidas PERFEX (Emory-Georgia Tech)

Sistema experto ampliado con redes neuronales para diagnóstico y enseñanza de imágenes SPECT Comercializado en la década de 1990

Sistema experto para emergencias / catástrofes

1988 → 1996

Triage Muerto, Inmediato (minutos), Postergado (horas), Minor

1. Recogido de datos
 1. Extraer información
 2. Quién debe estar presente?
 3. Qué interfaz?
2. Protocolos para toma de decisiones
 1. árboles binarios (true false)
- 3.

Cambio de tendencia

Sistemas basados en
Conocimiento → Datos

Sistemas de soporte a la decisión, Ontologías

Explosión de datos desde:

- WEB (Internet)
 - 1993: mosaic, netscape,
-

Human Genome Project - Information Archive 1990 - 2003

Medicina basada en la evidencia, personalizada, de precisión

Big Data en medicina

imágenes radiológicas

omics

microbiota

la web

weareables

Radiología

Mamografía,

La agencia que más dinero invierte en la investigación DARPA, porque las radiografías se parecen a los mapas de los drones.

- Anatomía patológica: cáncer de próstata y mama
- Oftalmología: retina de diabéticos, glaucoma
- Dermatología : Melanoma , en móviles (Visual DX)

Watson (miguel garcía) : Sistema hecho para jugar jeopardy (programa de concurso de tv)

NLP para extraer información

Data driven system Medicine

- 12 de Octubre: Rheumatoid Arthritis

Ontologías:

Gallo → Animal, VIH, Tenor, Pez

1990 Cambio de (KB) Conocimiento a Datos

Ontologías referencias para poder construir bases de conocimiento generalizadas.

Representar conceptos en el mundo.

1. Reglas
2. Relaciones semánticas

Ontología = Término filosófico que representa qué es, qué existe en el universo?

Ontología = especificación explícita de una conceptualización [Gruber93]

Vocabulario común de términos

significado de términos

consenso: entendimiento compartido entre personas y máquinas. No se busca la mejor conceptualización.

- LIGHT : Conceptos / Clases - Relaciones - propiedades / flota
 - Vocabularios, Taxonomías
- HEAVY:
 - Estructuras lógicas
 - axiomas
 - constraints

- inferencias
- restricciones
- Clases: entidades del mundo
- Propiedades y relaciones entre los conceptos
 - taxonomía: generalización entre conceptos
 - Relación, rol o atributo: función *de* , *tiene* actividad, localización , tamaño

Herramientas

XML → OWL (Ontology Web Language → DFR → (protégé

Siguiente día se explicará el assignment.

Colección de conceptos

- Relaciones
 - is_a
 - part_of

En biología hay una herencia que va hacia arriba : Emergencia

H₂O → Hidrógeno + Oxígeno (Sinergias)

Herencia.

a16-10 Presentaciones ML

- David Perez del Rey; Gema
- 20191107
-

Presentación 1: ML Ranking pointwise approach

Ignacio arroyo, Daniel Bermejo, Roman Bats

STAGED LOGISTIC REGRESSION

relevance score

5 features

- DAF: Document absolute frequency
- DRF: Document Relative Frequency
- CNAF: Common name absolute
- CNRF: common name relative frequency
- CF: component frequency

Staged logistic regression

Presentación 2: ML Ranking

David de la hoz, Salvador y Thomas

Score function evaluation.

- how
- did you

even though the informatics won't substitute doctors , only could help them to do better diagnosis

Presentación 3: Clicking through

Alfonso Hernandez silva , Pedro Hernández Vegas y Guillaume Baldi

introduction

problem

objectives

training

ranking

application code

conclusions

problem

create autonomous system that learns how to rank data

SVM: Support vector machine

compare the relevance of data in a specific query

how we decide which data is relevant

Clickthrough method

how get information relevance from these data

how do we retrieve data about relevance without explicit feedback

how do we clasificase data

objectives

- create an automatic system that retrieves information given a query
 - so it makes a ranking

Training

Clickthrough

triplets (query, ranking, documents)

Comparison of pairs of documents

8 features: positions and cosines

query → title, component , code,

conclusions

algorithm to improve search engine performance minimizing log-files

tractable for large number of features and queries

Presentación 4: AdaRank

Daniel Rodriguez y Elena Sea Noblejas

Int

Meth

Results

Conclusions

References

Int

Con

results accuracy was low than we expected

adarank need rich features and more extensive training set

low quantity of positive search results

- Para entrenar adarank: query y ranked list
- features: incluir más componentes
-

Práctica 5

QAF: Query absolute frequency – $QRF = Q$

DAF: Document absolute frequency

IDF: N/n_{tj}

RFAD: relative frequency of the term in all documents $TNTC/TNAT$

Práctica 6: Staged Logistic Regression

Logistic inference model

NUESTRO

1. Objetivo
 1. Crear un buscador: Information retrieval
2. comparar con el buscador real de LOINC
 1. expansion de query
 2. topic model? tf-idf? buscar en el contenido de los papers
 3. tesauro
 4. n-grams to auto fix words
3. qué es clickthrough?
4. dataset y features
 1. gráfica con fechas de submissions
5. Ranking - demo

6. Conclusiones

1. que partes son difíciles
2. posibles mejoras
 1. juntar varios algoritmos
 2. primero decir cuales son relevantes y luego calcular el ranking por clickthrough

7.