MACHINE LEARNING

Universidad Carlos III de Madrid

AI

Outline

Machine Learning

- Definition
  - programs that improve automatically based on the experience according to some measure

Machine Learning

- Definition
  - programs that improve automatically based on the experience according to some measure
- Objectives
  - learn new knowledge
    given sensory information predict whether traffic will be fluid, dense or congested in the next hour
  - improve the behavior of a system
    determine a control policy for traffic signals

Machine Learning

- Definition
  - programs that improve automatically based on the experience according to some measure
- Objectives
  - learn new knowledge

    given sensory information predict whether traffic will be fluid, dense or congested in the next hour
  - improve the behavior of a system

    determine a control policy for traffic signals
- Motivation
  - software development is a bottleneck (ill-defined tasks)

Machine Learning

- Definition
  - programs that improve automatically based on the experience according to some measure
- Objectives
  - learn new knowledge

    given sensory information predict whether traffic will be fluid, dense or congested in the next hour
  - improve the behavior of a system

    determine a control policy for traffic signals
- Motivation
  - software development is a bottleneck (ill-defined tasks)
  - information society generates huge quantities of data

Machine Learning

- Definition
  - programs that improve automatically based on the experience according to some measure
- Objectives
  - learn new knowledge

    given sensory information predict whether traffic will be fluid, dense or congested in the next hour
  - improve the behavior of a system

    determine a control policy for traffic signals
- Motivation
  - software development is a bottleneck (ill-defined tasks)
  - information society generates huge quantities of data
  - the comprehensibility of the output is very important

Machine Learning

- Definition
  - programs that improve automatically based on the experience according to some measure
- Objectives
  - learn new knowledge

    given sensory information predict whether traffic will be fluid, dense or congested in the next hour
  - improve the behavior of a system

    determine a control policy for traffic signals
- Motivation
  - software development is a bottleneck (ill-defined tasks)
  - information society generates huge quantities of data
  - the comprehensibility of the output is very important
  - personalization

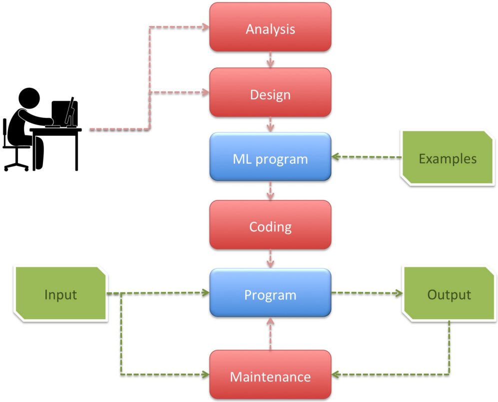
Machine Learning

- Definition
  - programs that improve automatically based on the experience according to some measure
- Objectives
  - learn new knowledge

    given sensory information predict whether traffic will be fluid, dense or congested in the next hour
  - improve the behavior of a system

    determine a control policy for traffic signals
- Motivation
  - software development is a bottleneck (ill-defined tasks)
  - information society generates huge quantities of data
  - the comprehensibility of the output is very important
  - personalization
  - adaptability

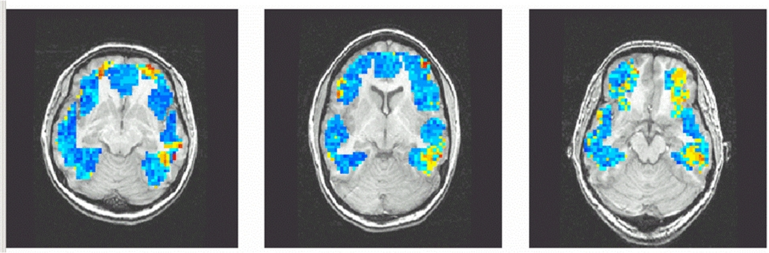
Programming by machine learning



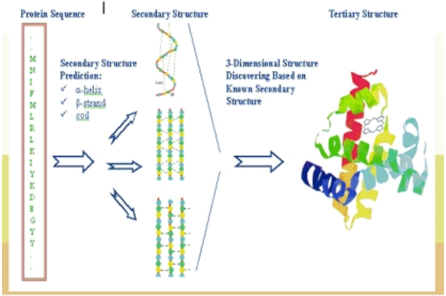
Some examples

- Medicine: how to diagnose, medicate
- Domotics: when to switch-off lights, open shades
- Smart cities: how to efficiently regulate energy consumption
- Bank: whom to loan, when to accept a credit card charge
- Marketing: what profile of people buy beers
- Personalization: what kinds of music do you listen to
- Robots: how to program a robot to grasp any object
- Astronomy: does an image contain something interesting
- Image understanding: does a given image include my wife

Analysis of brain images



Prediction of protein secondary structure



Data mining $\rightarrow$ Big data

- Data mining
  - What is the profile of costumers that spend more than X?
  - Should we invest on X tomorrow?
  - What book should we recommend after reading Y?
  - Should we allow a charge on a credit card?
  - What numbers do people call?
  - What do people buy together with beers?

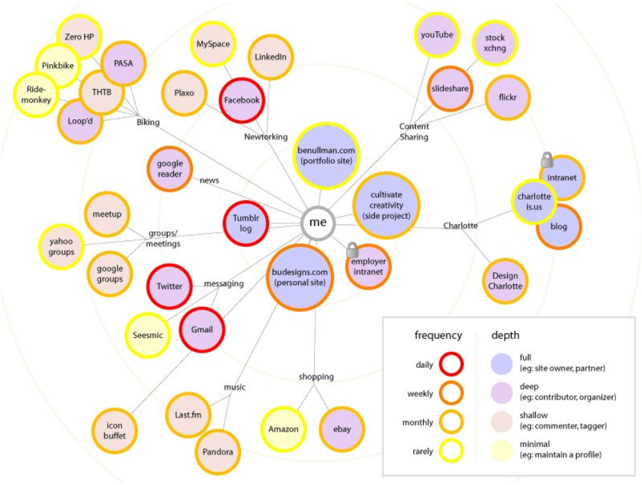
Data mining $\rightarrow$ Big data

- Data mining
  - What is the profile of costumers that spend more than X?
  - Should we invest on X tomorrow?
  - What book should we recommend after reading Y?
  - Should we allow a charge on a credit card?
  - What numbers do people call?
  - What do people buy together with beers?
- Big data
  - What public transport services people use over one year?
  - How to control traffic lights to reduce energy consumption?
  - What do images with traffic congestions have in common?
  - How does traffic behave at all city junctions over time?
  - What are uncommon traffic patterns from GPS data?

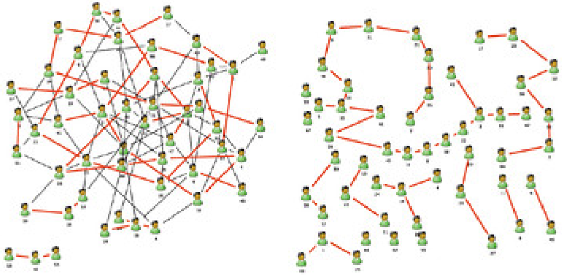
Text mining

EN
Google
ES
twitter
Donald J. Trump
@realDonaldTrump
Follow
The United States must greatly strengthen and expand its nuclear capability until such time as the world comes to its senses regarding nukes
11:50 AM - 22 Dec 2016
5,244
15,542
googleteam
GOOGLE LOTTERY WINNER! CONTAC
From: googleteam To:
Subject: GOOGLE LOTTERY WINNER! CONTACT YOUR AGENT TO CLAIM YOUR PRIZE.
GOOGLE LOTTERY INTERNATIONAL
INTERNATIONAL PROMOTION / PRIZE AWARD .
(WE ENCOURAGE GLOBALIZATION)
FROM: THE LOTTERY COORDINATOR,
GOOGLE B.V. 44 9459 PE.
RESULTS FOR CATEGORY "A" DRAWS
CONGRATULATIONS!
SPAM

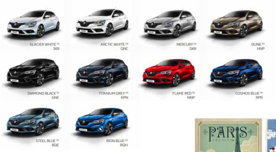
Social mining



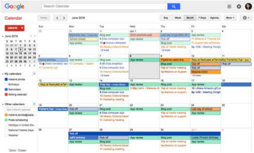
Social mining

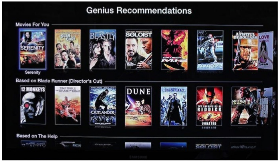


Customization



Customization






Título del libro: LA MONTAÑA MAGICA
Autor: MANN, THOMAS

5º EDICIÓN. Castrop es el protagonista de esta historia situada en sanatorio de Davos, Suiza. A través de su historia el autor nos transmite los problemas espirituales y sociales de su época. Una obra maestra.

Comentarios: 2 // Valoración: 10,7

Título del libro: HISTORIAS DE CRONOPIOS Y DE FAMAS
Autor: CORTAZAR, JULIO

Fiel a su tradicional cita anual, la editorial Aguilar ofrece a sus lectores un nuevo título de la Colección Crisolín, auténtica joya bibliográfica que se caracteriza por aunar en su pequeño formato la indiscutible altura literaria del autor seleccionado y el exquisito acabado. En esta ocasión, se trata de uno de los clásicos de la literatura: ...

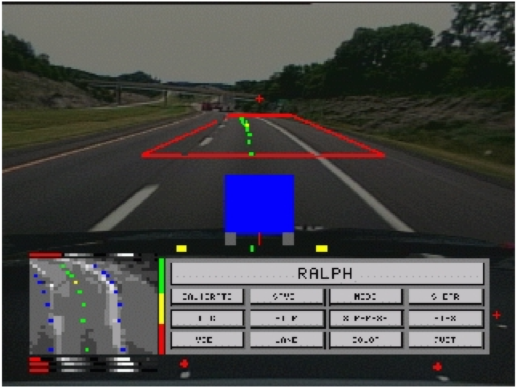
Comentarios: 1 // Valoración: 10,3

Título del libro: LA LLAMADA DE CTHULHU (JUEGO DE ROL DE HORROR)
Autor: LOVECRAFT, H.P.

Clasificado como un juego de horror está basado en el mundo creado por el escritor norteamericano Howard Phillips Lovecraft y un círculo de autores que se intercambiaban relatos entre ellos aproximadamente desde 1920 a 1930. H.P. Lovecraft creó un horror propio, apartado de los típicos terrores del romanticismo. Su mitología incluye gran canti...

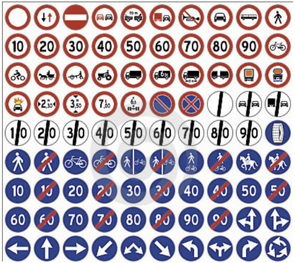
Comentarios: 1 // Valoración: 10,0

RALPH. Learning to drive



Sensors. Image recognition

Sensors. Image recognition



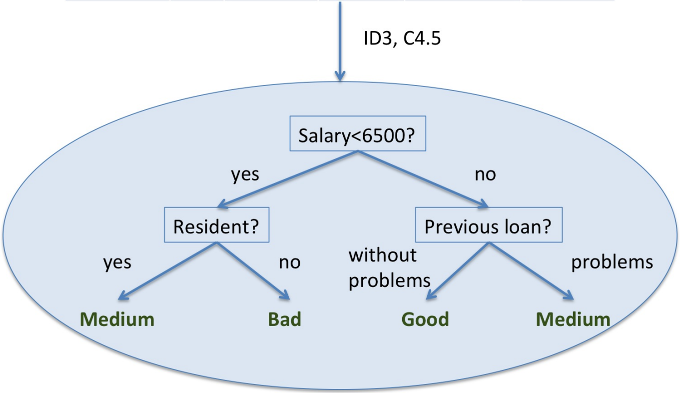
Some companies that use it

Google
twitter
Microsoft
amazon
facebook
VISA
Carrefour
NETFLIX
Santander

Outline

Decision trees

Instances	Age	Salary	...	Postal code	Class
Instance 1	30	6000	...	28911	Good
Instance 2	45	8000	...	28001	Bad
...	...	...	...	...	...
Instance E	23	1200	...	08002	Medium



Rules

Instances	Age	Salary	...	Postal code	Class
Instance 1	30	6000	...	28911	Good
Instance 2	45	8000	...	28001	Bad
...	...	...	...	...	...
Instance E	23	1200	...	08002	Medium

C4.5, AQ, CN2

If Salary<6500=yes AND Resident=yes
Then Class=Medium
If Salary<6500=yes AND Resident=no
Then Class=Bad
If Salary<6500=no AND Previous Loan=yes
Then Class=Good
If Salary<6500=no AND Previous Loan=no
Then Class=Medium

Bayesian classifier

Instances	Age	Salary	...	Postal code	Class
Instance 1	30	6000	...	28911	Good
Instance 2	45	8000	...	28001	Bad
...	...	...	...	...	...
Instance E	23	1200	...	08002	Medium

Naïve Bayes

Class	P(Class)
Good	0.6
Bad	0.2
Medium	0.2

Resident	Class	P(Resident=V| Class)
yes	Good	0.7
yes	Bad	0.1
yes	Medium	0.2
no	Good	0.3
no	Bad	0.3
no	Medium	0.4

Instance-based learning

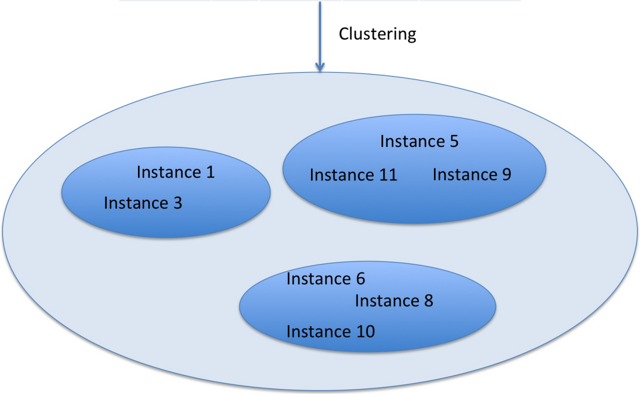
Instances	Age	Salary	...	Postal code	Class
Instance 1	30	6000	...	28911	Good
Instance 2	45	8000	...	28001	Bad
...	...	...	...	...	...
Instance E	23	1200	...	08002	Medium

IBL

Instances	Age	Salary	...	Postal code	Class
Instance 1	30	6000	...	28911	Good
Instance 2	45	8000	...	28001	Bad
...	...	...	...	...	...
Instance E	23	1200	...	08002	Medium

Clustering

Instances	Age	Salary	...	Postal code
Instance 1	30	6000	...	28911
Instance 2	45	8000	...	28001
...	...	...	...	...
Instance E	23	1200	...	08002



Outline

Bayesian classifier. Naïve Bayes

$$p(C \mid X) = \frac{p(X \mid C)p(C)}{p(X)}$$

- Class random variable: C
- Example random variable: X
- "A priori": $p(C)$
- Likelihood: $p(X \mid C)$
- Evidence: $p(X)$
- "A posteriori": $p(C \mid X)$

Bayesian classifier. Naïve Bayes

$$p(C \mid X) = \frac{p(X \mid C)p(C)}{p(X)}$$

- Class random variable: C
- Example random variable: X
- "A priori": $p(C)$
- Likelihood: $p(X \mid C)$
- Evidence: $p(X)$
- "A posteriori": $p(C \mid X)$
- In case of a particular class and example:

$$p(C = c_k \mid X = x_i) = \frac{p(X = x_i \mid C = c_k)p(C = c_k)}{p(X = x_i)}$$

Naïve Bayes

- Output of classifier

$$\begin{aligned}\text{class}(x_i) &= \arg\max_{c_k \in C} p(C = c_k \mid X = x_i) \\ &= \arg\max_{c_k \in C} \frac{p(X = x_i \mid C = c_k)p(C = c_k)}{p(X = x_i)}\end{aligned}$$

- Normalization factor is constant with respect to $C = c_k$:

$$\text{class}(x_i) = \arg\max_{c_k \in C} p(X = x_i \mid C = c_k)p(C = c_k)$$

Naïve Bayes assumption

- Each instance is a vector with values for n attributes

$$x_i = \langle x_i(1), x_i(2), \ldots, x_i(n) \rangle$$

- It can also be seen as values of a set of random variables (attributes)

$$x_i = \langle A_1 = x_i(1), A_2 = x_i(2), \ldots, A_n = x_i(n) \rangle$$

Naïve Bayes assumption

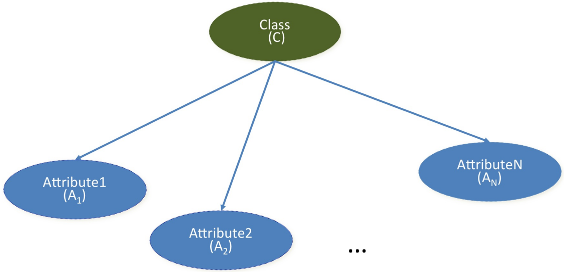
- Each instance is a vector with values for n attributes

$$x_i = \langle x_i(1), x_i(2), \ldots, x_i(n) \rangle$$

- It can also be seen as values of a set of random variables (attributes)

$$x_i = \langle A_1 = x_i(1), A_2 = x_i(2), \ldots, A_n = x_i(n) \rangle$$

- Naïve Bayes assumption: all attribute values are independent



Attributes independence

- Thus,

$$
\begin{aligned}
p(X = x_i \mid C = c_k) &= p(A_1 = x_i(1), \ldots, A_n = x_i(n) \mid C = c_k) \\
&= \prod_{j=1}^{n} p(A_j = x_i(j) \mid C = c_k)
\end{aligned}
$$

Attributes independence

- Thus,

$$\begin{aligned} p(X = x_i \mid C = c_k) &= p(A_1 = x_i(1), \ldots, A_n = x_i(n) \mid C = c_k) \\ &= \prod_{j=1}^{n} p(A_j = x_i(j) \mid C = c_k) \end{aligned}$$

- And

$$\begin{aligned} \text{class}(x_i) &= \arg\max_{c_k \in C} p(X = x_i \mid C = c_k) p(C = c_k) \\ &= \arg\max_{c_k \in C} p(C = c_k) \prod_{j=1}^{n} p(A_j = x_i(j) \mid C = c_k) \end{aligned}$$

Generating the classifier

- For each class $c_k \in C$
  - $p(C = c_k) = \frac{\text{number examples of class } c_k}{\text{total number of examples}}$
  - For each attribute A_j and value a_{jl}, compute $p(A_j = a_{jl} \mid C = c_k)$:
    - if A_j is discrete, $p(A_j = a_{jl} \mid C = c_k) = \frac{\text{number examples of class } c_k \text{ that } A_j = a_{jl}}{\text{number of examples of class } c_k}$
    - if A_j is continuous, compute μ_{jk} (average) y σ_{jk} (standard deviation)

Classification

- Given a new example x_i, classify it as:

$$\text{class}(x_i) = \arg\max_{c_k \in C} p(C = c_k) \prod_{j=1}^{n} p(A_j = x_i(j) \mid C = c_k)$$

- If A_j is discrete, $p(A_j = x_i(j) \mid C = c_k)$ has been stored for all j, values of j and k
- If A_j is continuous,

$$p(A_j = x_i(j) \mid C = c_k) = \frac{1}{\sqrt{2\pi\sigma_{jk}^2}} e^{-\frac{1}{2}\frac{(x_i(j)-\mu_{jk})^2}{\sigma_{jk}^2}}$$

Classification

- Given a new example x_i, classify it as:

$$\text{class}(x_i) = \arg\max_{c_k \in C} p(C = c_k) \prod_{j=1}^{n} p(A_j = x_i(j) \mid C = c_k)$$

- If A_j is discrete, $p(A_j = x_i(j) \mid C = c_k)$ has been stored for all j, values of j and k
- If A_j is continuous,

$$p(A_j = x_i(j) \mid C = c_k) = \frac{1}{\sqrt{2\pi\sigma_{jk}^2}} e^{-\frac{1}{2}\frac{(x_i(j)-\mu_{jk})^2}{\sigma_{jk}^2}}$$

- Correction
  if any $p(A_j = x_i(j) \mid C = c_k) = 0$
  then $p(A_j = x_i(j) \mid C = c_k) = \epsilon, 0 < \epsilon < 1$

Example

Age	Income	Fixed job	Client	Class
Young	Small	No	No	No
Adult	None	Yes	Yes	No
Adult	Medium	No	No	Yes
Young	Small	No	No	No
Young	High	Yes	Yes	Yes
Adult	Medium	Yes	Yes	Yes
Young	Small	Yes	Yes	Yes
Adult	Small	No	Yes	No

Example

- Class *a priori* probability
  - $P(\text{Yes}) = \frac{4}{8} = 0.5$
  - $P(\text{No}) = \frac{4}{8} = 0.5$
- Conditional probabilities

	Class	
Age	Yes	No
Young	$\frac{2}{4} = 0.5$	$\frac{2}{4} = 0.5$
Adult	$\frac{2}{4} = 0.5$	$\frac{2}{4} = 0.5$

	Class	
Income	Yes	No
None	$\frac{0}{4} = 0.0$	$\frac{1}{4} = 0.25$
Small	$\frac{1}{4} = 0.25$	$\frac{3}{4} = 0.75$
Medium	$\frac{2}{4} = 0.5$	$\frac{0}{4} = 0.0$
High	$\frac{1}{4} = 0.25$	$\frac{0}{4} = 0.0$

Fixed	Class	
job	Yes	No
Yes	$\frac{3}{4} = 0.75$	$\frac{1}{4} = 0.25$
No	$\frac{1}{4} = 0.25$	$\frac{3}{4} = 0.75$

	Class	
Client	Yes	No
Yes	$\frac{3}{4} = 0.75$	$\frac{2}{4} = 0.5$
No	$\frac{1}{4} = 0.25$	$\frac{2}{4} = 0.5$

Classification

- Given X_i to classify

Age	Income	Fixed job	Client
Adult	High	Yes	No

- what is its most probable class? ($\epsilon = 0.01$)

Classification

- Given X_i to classify

Age	Income	Fixed job	Client
Adult	High	Yes	No

- what is its most probable class? ($\epsilon = 0.01$)

$$
\begin{aligned}
p(C = \text{Yes} \mid X_i) = p(C = \text{Yes}) \times & p(\text{Age=Adult} \mid C = \text{Yes}) \times \\
& p(\text{Income=High} \mid C = \text{Yes}) \times \\
& p(\text{Fixed job=Yes} \mid C = \text{Yes}) \times \\
& p(\text{Client=No} \mid C = \text{Yes}) = \\
& 0.5 \times 0.5 \times 0.25 \times 0.75 \times 0.25 = 0.0117
\end{aligned}
$$

$$
\begin{aligned}
p(C = \text{No} \mid X_i) = p(C = \text{No}) \times & p(\text{Age=Adult} \mid C = \text{No}) \times \\
& p(\text{Income=High} \mid C = \text{No}) \times \\
& p(\text{Fixed job=Yes} \mid C = \text{No}) \times \\
& p(\text{Client=No} \mid C = \text{No}) = \\
& 0.5 \times 0.5 \times \epsilon \times 0.25 \times 0.5 = 0.0003
\end{aligned}
$$

- Most probable class: Yes