Probabilità e logica: Strumenti per la costruzione di una cultura dell'incertezza

Hykel Hosni

Bari - 6 ottobre 2018

Una sfida urgente

Opporre alla pretesa cultura dell'oggettività una cultura dell'incertezza

La "cultura" dell'oggettività è una combinazione vaga e spesso imperscrutabile di:

- big data / new data
- algoritmi di ranking e classificazione (machine learning / analytics)
- internet delle cose

Una sfida urgente

Opporre alla pretesa cultura dell'oggettività una cultura dell'incertezza

La "cultura" dell'oggettività è una combinazione vaga e spesso imperscrutabile di:

- big data / new data
- algoritmi di ranking e classificazione (machine learning / analytics)
- internet delle cose

Punti di attenzione

- deresponsabilizzazione delle decisioni (rating economico-finanziario, diagnosi automatiche, polizia predittiva)
- moltiplicazione degli effetti negativi dei nostri pregiudizi (cognitivi)
- confusione tra "la migliore analisi disponibile" e "la verità"

Il sogno di applicare il metodo di deduzione logica a ragionamento e all'argomentazione parte dalla Casa della Saggezza di Baghdad e arriva all'Internet delle Cose, passando per Leibniz e Turing:

Il sogno di applicare il metodo di deduzione logica a ragionamento e all'argomentazione parte dalla Casa della Saggezza di Baghdad e arriva all'Internet delle Cose, passando per Leibniz e Turing:

M. Al-Khwarizmi (780-850)
Il ragionamento matematico è calcolo (il metodo dell' *algebra*)

Il sogno di applicare il metodo di deduzione logica a ragionamento e all'argomentazione parte dalla Casa della Saggezza di Baghdad e arriva all'Internet delle Cose, passando per Leibniz e Turing:

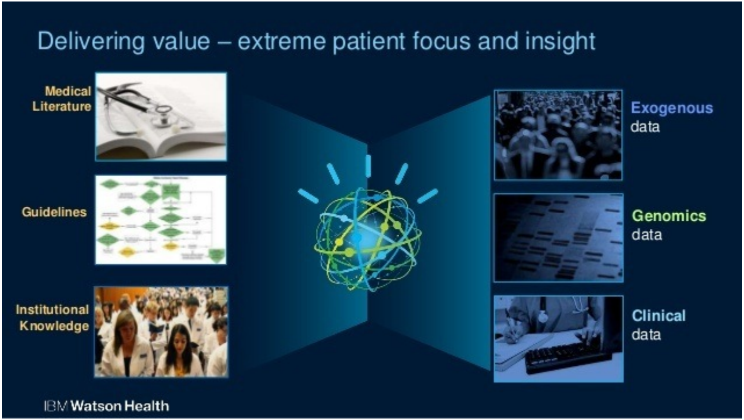
G. W. LEIBNIZ (1646–1716)

Attraverso l'uso di un linguaggio universale (characteristica universalis) ogni argomento può essere deciso mediante deduzione logica

Il sogno di applicare il metodo di deduzione logica a ragionamento e all'argomentazione parte dalla Casa della Saggezza di Baghdad e arriva all'Internet delle Cose, passando per Leibniz e Turing:

A. M. Turing (1912 - 1954)
Non c'è ragione di credere che le macchine (cioè i calcolatori digitali in grado di eseguire procedure algoritmiche) non possano pensare

Il sogno di applicare il metodo di deduzione logica a ragionamento e all'argomentazione parte dalla Casa della Saggezza di Baghdad e arriva all'Internet delle Cose, passando per Leibniz e Turing:



IBM Watson Health (2013-) `http://www.ibm.com/watson/health/`

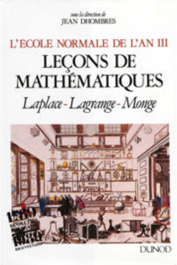
Il sogno di Leibniz e il pericolo dell' "oggettività degli algoritmi"

*[Q]uando sorgeranno controversie fra due filosofi, **non sarà più necessaria una discussione**, come [non lo è] fra due calcolatori. Sarà sufficiente, infatti, che essi prendano in mano le penne, si siedano di fronte agli abachi e (se così piace, su invito di un amico) si dicano l'un l'altro: Calculemus!*

G.W. Leibniz
(1646 – 1716)

1. La cultura dell'incertezza
2. Logica e probabilità "zero"
3. Perché la probabilità non è oggettiva

Enfin, on donnera les principes de la théorie des probabilités. **Dans un temps où tous les citoyens sont appelés à décider du sort de leurs semblables**, il leur importe de connaître une science qui fait apprécier, aussi exactement qu'il est possible, *la probabilité des témoignages, et celle qui résulte des circonstances dont les faits sont accompagnés* : il importe surtout de leur apprendre à se *défier des aperçus même les plus vraisemblables*; et rien n'est plus propre à cet objet que la théorie des probabilités, dont souvent les résultats rigoureux sont contraires à ces aperçus. D'ailleurs, les nombreuses applications de cette théorie, aux naissances, aux mortalités, aux élections et aux assurances, applications qu'il est avantageux de perfectionner et d'étendre à d'autres objets, la rendent *une des parties les plus utiles des connaissances humaines* (Laplace, Lagrange, Monge, 1795.)



[Disponibile in libero accesso su `https://books.openedition.org/editionsulm/455`]

Il valore culturale della probabilità

A. Renyi, *A diary on Information Theory*, Wiley, 1984 (pp 77-78)

WHY SHOULD PROBABILITY THEORY BE TAUGHT?
First, it may seem that this question can be answered adequately only if we are considering a certain type and level of education. In my opinion, there are still some general points which can be made anyway [...]:
A) Probability theory should be taught because of the important role it can play in the development of the students' ability to think.
B) It should be taught because of its usefulness in everyday life, in science and technology, etc.
C) It is important, indeed indispensable in mathematical education. I will now expand upon these points.

A. Renyi, *A diary on Information Theory*, Wiley, 1984 (pp 77-78)

[elaborando sulle capacità di ragionamento] The study of probability even strengthens students' character. For example, it increases their courage when they understand that certain failures are due to chance so that a set-back is not a sufficient reason for giving up. Primitive people have a tendency to be very superstitious: if something goes wrong they try to attribute it to somebody's maliciousness even if such is not the case. The reason for this is that they do not understand the notion of chance. The study of probability theory can help to erase these remnants of magical thinking from the Stone Age, to make people more understanding toward fellow human beings and to help them to find their place in society.

Ma c'è molto di più

Conoscenze elementari di logica e di probabilità sono sufficienti per dimostrare che la probabilità di un evento incerto è calcolabile in modo univoco soltanto in condizioni così particolari da non essere praticamente mai soddisfatte nei problemi di interesse individuale e sociale

Notazione

- $\mathcal{L} = \{p_1, \ldots, p_n\}$: le *variabili proposizionali*;
- $\mathcal{C} = \{\neg, \wedge, \vee, \rightarrow\}$: i *connettivi proposizionali*
- $\mathcal{SL} = \{\theta, \phi, \ldots\}$: l'insieme degli *enunciati di* L costruito ricorsivamente come segue:
  - Le variabili proposizionali sono enunciati
  - se θ, ϕ sono enunciati, allora $\neg\theta, (\theta \wedge \phi), (\theta \vee \phi), (\theta \rightarrow \phi)$ sono enunciati di L.
- Γ, Δ: insiemi di enunciati su L

Tabelle booleane

Una *valutazione proposizionale (classica)* su $\mathcal{L}$ è una funzione

$$v : \mathcal{L} \longrightarrow \{0, 1\}$$

tale che

p	q	$\neg p$	$p \wedge q$	$p \vee q$	$p \to q$
1	1	0	1	1	1
1	0	0	0	1	0
0	1	1	0	1	1
0	0	1	0	0	1

Tabella: Il significato dei connettivi proposizionali

Le valutazioni si estendono univocamente a $\mathcal{SL}$

Poiché l'insieme degli enunciati $\mathcal{SL}$ è costruito **ricorsivamente** da $\mathcal{L}$ mediante i connettivi $\mathbb{C} = \{\neg, \vee, \wedge, \rightarrow\}$ abbiamo

Il principio di composizionalità

Il valore di verità di un enunciato composto è una funzione fissata del valore di verità dei suoi componenti.

Esempio

Siano $\theta, \phi_1, \phi_2 \in SL$ e supponiamo che $\theta = (\phi_1 \wedge \phi_2)$. Esiste per composizionalità una $f : \{0,1\} \times \{0,1\} \rightarrow \{0,1\}$ tale che

$$v(\theta) = f_\wedge(v(\phi_1), v(\phi_2)) = \begin{cases} 1, & \text{se } v(\phi_1) = v(\phi_2) = 1; \\ 0, & \text{altrimenti} \end{cases}$$

per esempio

$$f_\wedge(p, q) = \min\{v(\phi_1), v(\phi_2)\}$$

Calculemus!

Diciamo che un argomento

$$\overbrace{\gamma_1, \gamma_2, \dots, \gamma_n}^{\text{premesse}} \quad \therefore \quad \overbrace{\theta}^{\text{conclusione}}$$

è corretto se e solo se

$$\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_n\} \models \theta$$

cioè se e solo se

$$\forall v \in \mathbb{V}, \text{ se } v(\Gamma) = 1 \text{ allora } v(\theta) = 1.$$

Le tabelle booleane forniscono un metodo algoritmico per decidere la correttezza di ogni argomento espresso nel linguaggio della logica proposizionale classica.

Funzioni di probabilità

Una funzione di probabilità su $\mathcal{L}$ è una mappa $P : \mathcal{SL} \to [0,1]$ tale che

(P1) se $\models \theta$ allora $P(\theta) = 1$

(P2) se $\theta \models \neg\phi$ allora $P(\theta \vee \phi) = P(\theta) + P(\phi)$.

Conseguenze immediate

1. $P(\neg\theta) = 1 - P(\theta)$
2. $\not\models \theta \Rightarrow P(\theta) = 0$
3. $\theta \models \phi \Rightarrow P(\theta) \leq P(\phi)$
4. $\theta \equiv \phi \Rightarrow P(\theta) = P(\phi)$
5. $P(\theta \vee \phi) = P(\theta) + P(\phi) - P(\theta \wedge \phi)$.

Teorema di Rappresentazione (Renyi)

Theorem

1. *Sia $P : SL \to [0,1]$ una funzione di probabilità su un linguaggio finito. Allora i valori di P sono completamente determinati dai valori che questa ha sull'insieme delle congiunzioni massimali di letterali o atomi $AT^L = \{\alpha_1, \ldots, \alpha_J\}$, rappresentati dal vettore*

$$\langle P(\alpha_1), P(\alpha_2), \ldots, P(\alpha_J) \rangle \in \mathbb{D}^L = \{\vec{a} \in \mathbb{R}^J \mid \vec{a} \geq 0, \sum_{i=1}^{J} a_i = 1\}.$$

2. *Viceversa, fissato $\vec{a} = \langle a_1, \ldots, a_J \rangle \in \mathbb{D}^L$ sia $P' : SL \to [0,1]$ definita da*

$$P'(\theta) = \sum_{\alpha_i : \alpha_i \models \theta} a_i.$$

P' così costruita è una funzione di probabilità.

Esempio

Sia $\mathcal{L} = \{p, q\}$. L'insieme degli atomi su $\mathcal{L}$ è il seguente

$$AT^L = \{\neg p \wedge \neg q, \neg p \wedge q, p \wedge \neg q, p \wedge q\}$$

Sia $\mu : AT^L \to [0,1]$ la *distribuzione uniforme su* AT^L, i.e.

$$\mu(\neg p \wedge \neg q) = \mu(\neg p \wedge q) = \mu(p \wedge \neg q) = \mu(p \wedge q) = 1/4$$

Quindi, per logica e probabilità elementare abbiamo

$$P(p \to q) = \mu(\neg p \wedge \neg q) + \mu(\neg p \wedge q) + \mu(p \wedge q) = 3/4$$

In pratica

Per linguaggi finiti, tutta l'informazione rilevante relativa a una funzione di probabilità è contenuta nella distribuzione della massa unitaria sugli *atomi* dell'algebra di Lindenbaum

La buona notizia

Se disponiamo della distribuzione di probabilità sugli atomi possiamo calcolare univocamente la probabilità di tutti gli enunciati $\theta \in SL$.

La cattiva notizia

Quando abbiamo a che fare con problemi interessanti dal punto di vista teorico e pratico non disponiamo praticamente mai di una distribuzione sugli atomi.

Calculemus probabilistico?

Per $\theta, \phi \in \mathcal{SL}$ supponiamo di conoscere $P(\theta)$ e $P(\phi)$. Questi valori sono sufficienti a determinare univocamente $P(\theta \wedge \phi)$?

Per $\theta, \phi \in \mathcal{SL}$ supponiamo di conoscere $P(\theta)$ e $P(\phi)$. Questi valori sono sufficienti a determinare univocamente $P(\theta \wedge \phi)$?

No!

Siano $L = \{p, q\}$ e $P(p) = P(q) = 1/2$. Possiamo costruire due funzioni di probabilità su L tali che

1. $P_1(p) = P_1(q) = P_2(p) = P_2(q) = P(p) = P(q) = 1/2$ e
2. $P_1(p \wedge q) \neq P_2(p \wedge q)$

Per farlo fissiamo

1. μ_1: la distribuzione uniforme su AT^L. In questo caso otteniamo $P_1(p \wedge q) = 1/4$;
2. $\mu_2(\neg p \wedge \neg q) = \mu_2(p \wedge q) = 0$ e $\mu_2(\neg p \wedge q) = \mu_2(p \wedge \neg q) = 1/2$. In questo caso otteniamo $P_2(p \wedge q) = 0$

Non è un difetto, ma una caratteristica!

Supponiamo che

1. $Bel : SL \to [0,1]$ sia una misura di incertezza tale che $Bel(\theta) = Bel(\neg\theta) > 0$
2. per ogni $\theta, \phi, \in SL$, esista $F_\wedge : [0,1] \times [0,1] \to [0,1]$ tale che

$$Bel(\theta \wedge \phi) = F_\wedge(Bel(\theta), Bel(\phi))$$

Allora

$$Bel(\theta \wedge \theta) = F_\wedge(Bel(\theta), Bel(\theta)) = F_\wedge(Bel(\theta), Bel(\neg\theta)) = Bel(\theta \wedge \neg\theta)$$

Non è un difetto, ma una caratteristica!

Supponiamo che

1. $Bel : SL \to [0,1]$ sia una misura di incertezza tale che $Bel(\theta) = Bel(\neg\theta) > 0$
2. per ogni $\theta, \phi, \in SL$, esista $F_\wedge : [0,1] \times [0,1] \to [0,1]$ tale che

$$Bel(\theta \wedge \phi) = F_\wedge(Bel(\theta), Bel(\phi))$$

Allora

$$Bel(\theta \wedge \theta) = F_\wedge(Bel(\theta), Bel(\theta)) = F_\wedge(Bel(\theta), Bel(\neg\theta)) = Bel(\theta \wedge \neg\theta)$$

Decisamente strano per una misura di incertezza ragionevole!

Riassumendo

- la quantificazione dell'incertezza è determinata univocamente soltanto in casi molto speciali – *quasi mai quando si esce dall'ambito delle scienze naturali*

Cultura dell'incertezza

Non affidarsi a un metodo che dia l'unica risposta "oggettiva" e quindi l'unica corretta perché questa molto probabilmente non esiste.

Riassumendo

- la quantificazione dell'incertezza è determinata univocamente soltanto in casi molto speciali – *quasi mai quando si esce dall'ambito delle scienze naturali*
- casi di particolare impatto sociale: diagnosi mediche, calcolo del rischio in finanza, previsione dei crimini, ecc. – *problemi per cui molti vorrebbero affidarsi con cieco entusiasmo agli imperscrutabili metodi di deep learning e ai metodi di analisi automatica di "big data"*

Cultura dell'incertezza

Non affidarsi a un metodo che dia l'unica risposta "oggettiva" e quindi l'unica corretta perché questa molto probabilmente non esiste.

Riassumendo

- la quantificazione dell'incertezza è determinata univocamente soltanto in casi molto speciali – *quasi mai quando si esce dall'ambito delle scienze naturali*
- casi di particolare impatto sociale: diagnosi mediche, calcolo del rischio in finanza, previsione dei crimini, ecc. – *problemi per cui molti vorrebbero affidarsi con cieco entusiasmo agli imperscrutabili metodi di deep learning e ai metodi di analisi automatica di "big data"*
- **esiste una pluralità di valutazioni probabilistiche logicamente coerenti**, ognuna delle quali può essere ben motivata dal contesto

Cultura dell'incertezza

Non affidarsi a un metodo che dia l'unica risposta "oggettiva" e quindi l'unica corretta perché questa molto probabilmente non esiste.

Sull'oggettività della logica (e degli algoritmi)

B. de Finetti (1972) Periodico di matematiche

La logica è come un palo, utile in quanto può impedire alla pianta del pensiero di crescere storta. Ma, come un palo non è una pianta né il possibile surrogato di una pianta, così la logica non è il pensiero né una specie di surrogato del pensiero.

Qualche lettura

- C. O'Neil *Weapons of Math Destruction*, Crown Publishing House, 2016
- H. Fry *Hello World*, Norton 2018
- H. Hosni. *Probabilità: Come smettere di preoccuparsi e imparare ad amare l'incertezza*, Carocci 2018
- H. Hosni, A. Vulpiani "La scienza delle previsioni e l'arte di costruire i modelli", *Lettera Matematica Pristem* 104 pp 21–29, 2018.
- M. Giaquinta, H. Hosni. "La Matematica Nelle Scienze Sociali", *Lettera Matematica Pristem* 93: 17–27, 2015