
Sepel Editrice
BANCHE DATI PUBBLICHE E PRIVATE ALLA PROVA DEL DEEP LEARNING
DIEGO GIORIO
XX E PRIVACY
ROMA 5 NOVEMBRE 2016

# UN SOFTWARE GARANTISTA

- Prenotiamo le vacanze via Internet.
- Le carte di credito non funzionano.
- Arrivando in albergo troveremo la camera?
- Sì, perché Booking ci stima molto ed ha garantito per noi!
- In realtà il rating di Booking è dovuto ai viaggi di lavoro.
- Apparentemente spendiamo in viaggi molto più del normale.
- Facile da spiegare ad una persona, difficile che un calcolatore lo preveda nell'algoritmo.

# MA SE I DATI SONO CONDIVISI?

- Certo Booking non è interessato alla provenienza ultima del denaro.
- Anche con un pieno accesso ai dati ed un algoritmo sofisticato non cambierebbe il rating.
- Ma se i dati (e/o la loro analisi) fossero condivisi fra altri Enti ed altri fornitori?
- Dovrei spiegare all'Agenzia delle Entrate perché abbiamo speso così tanto?
- Sicuramente molti algoritmi rileverebbero un'anomalia.

# LE RACCOLTE SFUGGONO AL CONTROLLO

- Stiamo costruendo sistemi di raccolta sempre più diffusi.
- Gli algoritmi d'analisi sono sempre più sofisticati.
- Ci si espone però al rischio di fraintendimenti ed errori.
- Problemi che diventano più gravi quanto più i sistemi divengono diffusi ed automatici.
- Devono essere automatici, essendo impossibile analizzare col contributo umano tutti questi dati.
- Avere molti dati non significa avere tante informazioni utili.

# SOLO I SOFTWARE POSSONO ANALIZZARE

- Le nostre città sono piene di telecamere, impossibile un controllo umano h24.
- Si visionano a posteriori, ad evento già accaduto.
- Oppure si sviluppano software di analisi automatica.
- Ma saranno in grado di distinguere lo stress di un rapinatore dall'arrabbiatura per un taglio di capelli errato?
- Il comportamento e lo stress sono molto soggettivi.
- Il nervosismo si riduce con l'esperienza.
- Un computer calcola un punteggio, ragiona in termini di soglia.

# ANALISI EVOLUTE

- Facebook ha censurato una foto storica di un bambino nudo.
- Logico che questo genere di immagini sia bandito.
- Notevole che un software riesca a riconoscerle.
- Impossibile che un sistema automatico capisca da solo il valore storico ed emotivo e disponga un'eccezione.

CORRIERE DELLA SERA

TECNOLOGIA / SOCIAL

SOCIAL MEDIA

**Facebook censura la foto simbolo del Vietnam**

Era stata postata da uno scrittore, il cui profilo è stato sospeso. Un giornale norvegese scrive una lettera di denuncia in prima pagina a Zuckerberg

di Alessio Lana

# ANALISI ESTESE

- Come ha fatto FB a riconoscere la foto?
- I metodi classici consentono i riconoscimenti ottici mediante scomposizioni vettoriali.
- I metodi futuristici si basano su reti neurali ed AI.
- Il *deep learning* si pone in una posizione intermedia.
- Come si può far capire ad un computer cos'è una chiave?
- Si installa un software di base e si fanno passare migliaia di immagini, alcune delle quali contengono chiavi, altre no.
- Con l'esperienza, come un umano, il computer impara.




# DEEP LEARNING

- Sinora abbiamo esaminato i casi di valutazioni puntuali.
- Booking mi considera un buon cliente, ma non è interessato al mio comportamento al di fuori dei viaggi.
- Un sistema di sicurezza mi controlla da quando entro nella metro a quando esco, non mi segue oltre.
- Un sistema più completo (invadente), sarebbe più efficace?
  - Osservare che entro nella linea blu senza un motivo impedirebbe un attentato perché si capisce che sto facendo un sopralluogo?
  - Oppure si indagherebbe un ragazzo timido che non osa suonare ad una ragazza?
- Una raccolta dati estesa ed un'analisi continua combatte il terrorismo?

# TERRORISMO ED ALTRI CRIMINI

- Peraltro il discorso non può limitarsi al mero terrorismo.
- Vero è che più di altri delitti si nutre di Internet:
  - Reclutamento
  - Auto-indottrinamento
  - Scambio di istruzioni
  - Finanziamento
- Vero è che il suo scopo precipuo è di suscitare un impatto emotivo.
- Però in certi quartieri si rischia di essere rapinati, in casa si rischia il furto, nelle zone controllate dalla criminalità organizzata il problema non è il jihadista.

# VALORE DI UNA RACCOLTA ESTESA

- Torniamo all'utilità della raccolta dati.
- Lo stesso direttore della nostra intelligence* ammette che spesso i segnali non sono stati utilizzati.
- Molti terroristi che hanno commesso attentati erano già segnalati o sorvegliati.
- Però ritiene utile il registro europeo dei passeggeri aerei.
- Alcuni terroristi avrebbero l'abitudine di fare in modo di sedersi accanto sullo stesso volo.
- Ammesso che sia vero, non potrebbero ora utilizzare un pullman o un traghetto?

*http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/4612330

# PREVENZIONE?

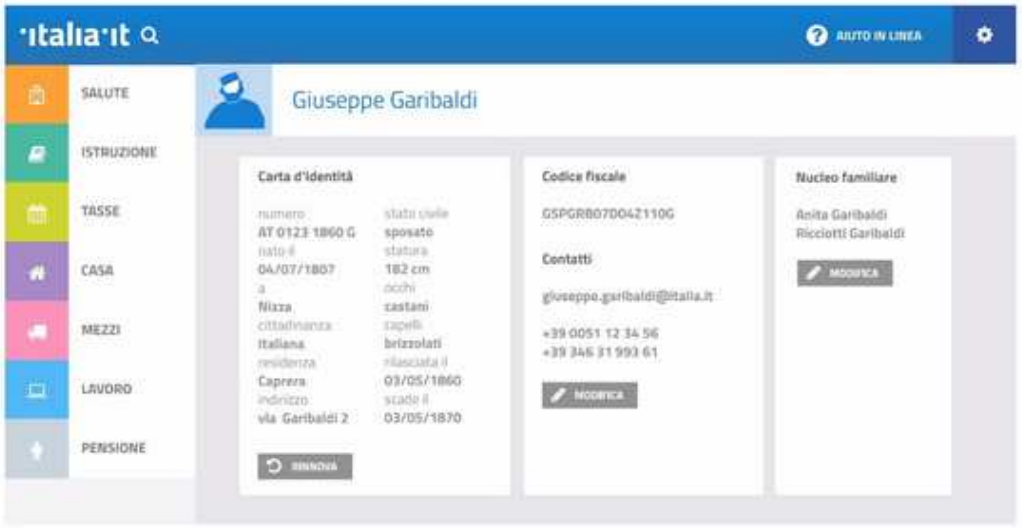
- La mera raccolta dati non serve a nulla.
- Un'analisi superficiale serve a poco.
- Un'analisi profonda, una tecnica di *deep learning* potrebbe servire a **prevenire** il reato?
- Probabilmente no:
  - Definire un comportamento come sospetto non è facile.
  - Individuare chi guarda una stazione con l'occhio del soldato non è banale.
  - L'attentatore di Nizza abitava lì, non compiva sopralluoghi.
  - Inoltre c'è la fastidiosa limitazione propria delle democrazie, che impedisce di arrestare qualcuno prima che abbia commesso un reato.

# BANCHE DATI GLOBALI

- Sistemi di analisi evoluta sono più utili in altri contesti.
- L'incrocio di varie banche dati facilita la caccia all'evasore.
- Lo studio demografico comparato consente di individuare tendenze a breve, medio e lungo termine.
- Un primo passo lo sta compiendo con l'ANPR.
- Unita a SPID e ad altri servizi, porterà ad una disponibilità di dati senza precedenti:

# BANCHE DATI GLOBALI

## Servizi ".italia.it"




Ideata dalla Presidenza del Consiglio dei Ministri, **"Italia Login"** raccoglie i servizi erogati dalle PP.AA. italiane in un ecosistema all'interno del quale ogni cittadino italiano ha un profilo civico online dal quale potrà accedere alle informazioni e ai servizi pubblici che lo riguardano.

# BANCHE DATI GLOBALI

- Sono presenti molte informazioni, ed altre saranno aggiunte:
  - Leva
  - Elettorale
  - Stato civile
  - ...
- Sicuramente sarà molto comodo avere tutti i dati radunati.
- Sarà anche pericoloso in caso di accesso abusivo.
- Ma sarà utile per prevenire terrorismo e criminalità?

# LA DITTATURA DEGLI ALGORITMI

- Ritorniamo alle considerazioni precedenti.
- Avere molti dati è inutile se non vengono elaborati correttamente.
- Tecniche di deep learning potrebbero aggiungere ad una grande capacità di calcolo una dose di intuito.
- Ma di quanto aumenterebbe il vantaggio strategico?
- Vogliamo vivere in una società dove ogni comportamento viene osservato ed analizzato?
- Quali sono i criteri di analisi?

# ALMENO MASSIMIZZIAMO I VANTAGGI

- All'inizio ho rilevato che i parametri usati da Booking per attribuirmi il rating posso solo immaginarli.
- Non sarei sorpreso se sistemi molto complessi, come Google o Facebook, sfuggissero agli stessi programmatori.
- Tecniche di analisi dati alternative alle attuali possono dare vantaggi e svantaggi.
- L'importante sarà unire lo stakanovismo dei calcolatori all'intelligenza umana.
- Il rischio è di unire invece la stupidità dei computer all'inaffidabilità umana.

# CONTROLLARE I CONTROLLORI

- Rischiamo di trovarci con una società pesantemente controllata.
- I sistemi che valutano il nostro comportamento, anche con il lodevole intento di prevenire il crimine, giudicano secondo parametri non del tutto conosciuti e di dubbia affidabilità.
- In passate edizioni avevo sostenuto l'uso dell'informatica ai fini di giustizia.
- Avevo però sostenuto anche la necessità che dette applicazioni vengano impiegate in modo controllato.
- E, sarò retrogrado, ma preferirei che il controllo ultimo avvenga ad opera di un essere umano.
- Il che è peraltro prescritto dal nuovo Regolamento europeo per il trattamento dei dati personali*.

* Punto (71) delle premesse

# SEMPRE UTILE LA RICERCA

- Non voglio dire che la ricerca debba rinunciare a sviluppare i nuovi metodi ed il nuovo modo di approcciare le analisi.
- Anche perché c'è sempre la possibilità che il nemico faccia lo stesso, ovvero che un terrorista applichi tecniche di *deep learning* per individuare autocivetta, agenti sotto copertura e simili.
- La conoscenza dei sistemi può certamente dare qualche vantaggio strategico.
- Neppure voglio dire che questo tipo di analisi verrà davvero utilizzata dallo Stato, stavo solo ipotizzando un futuro possibile, neppure troppo futuribile.

# TRASPARENZA E CONTROLLO DEMOCRATICO

Voglio dire che, come tutto ciò che riguarda i rapporti fra cittadini e Pubbliche Amministrazioni, questi sistemi devono essere usati solo se effettivamente utili ed efficaci, se non c'è un altro sistema meno invasivo, e soprattutto se possono essere impiegati in modo trasparente, **sotto il controllo democratico e popolare**.


Sepel Editrice
GRAZIE PER LA VOSTRA ATTENZIONE
Giorio Diego
XX e-privacy
Roma 5 novembre 2016
Più si impara, più il mondo cambia