

# **“Diversità, algoritmi di raccomandazione e PSM: salvaguardare gli obiettivi del servizio pubblico multimediale”**

Co.Re.Com Emilia-Romagna  
Aprile 2021

*TRASCRIZIONE del SEMINARIO tenutosi a Bologna il 26 Novembre 2019*

Martedì 26 novembre 2019 il Corecom Emilia-Romagna, in collaborazione con l'Ufficio Studi della RAI, ha organizzato un seminario sul tema: *Diversità, algoritmi di raccomandazione e PSM: salvaguardare gli obiettivi del servizio pubblico multimediale.*

Nei prossimi anni assisteremo ad una crescita esponenziale dei consumi audiovisivi non lineari ed il servizio pubblico del nostro paese ha già gettato le basi per essere protagonista in questo scenario. Nella fruizione di contenuti audiovisivi online i sistemi di raccomandazione sono uno strumento indispensabile, capace di rendere più efficaci le ricerche e proporre a dare visibilità ai prodotti più adatti al singolo utente. Tuttavia, gli algoritmi che verranno implementati dai PSM dovrebbero tenere in considerazione il ruolo specifico del servizio pubblico multimediale.

Hanno partecipato alla discussione, moderati da Stefano Cuppi (Presidente Corecom ER) e da Andrea Montanari (Direttore Ufficio Studi RAI):

**FRANCESCO RICCI** (Computer science, Università degli studi di Bolzano) Introduzione ai *Recommender Systems*

**PAOLO CREMONESI** (Sistemi Di Elaborazione Delle Informazioni, Politecnico di Milano) Algoritmi di raccomandazione nel VOD

**FABIO BASSAN** (Università degli Studi Roma Tre) Il servizio pubblico multimediale nella Concessione e nei contratti di servizio

**ERIK LAMBERT** (Esperto settore Media) I PSM Europei e l'evoluzione dei sistemi di raccomandazione

**FRANCESCO SILIATO** (Studio FRASI) I consumi lineari e non lineari: scenari, *trend* e misurazione

**PIERO DE CHIARA** (Forum Diversità e Diseguaglianze) Missioni strategiche e finanziamento delle aziende pubbliche nell'epoca dell'intelligenza artificiale

**GIANLUCA VISALLI** (Product Development Rai Digital)

**ANTONIO SASSANO** (Presidente della Fondazione Ugo Bordoni)

**FLAVIA BARCA** (Esperta settore Media) "Il pregiudizio delle macchine" ovvero come gli algoritmi riproducono stereotipi e discriminazioni di genere

**ALBERTO MORELLO** (Centro Ricerche Rai Innovazione Tecnologica e Sperimentazione RAI) - La collaborazione RAI nel progetto PEACH (EBU)

**MARINA CAPORALE** (Università degli Studi di Bologna) PSM, algoritmi e trasparenza

**GUALTIERO MAZZI** (Presidente Corecom Veneto) La tutela delle diversità culturali e

territoriali nel Servizio Pubblico Multimediale

**MARIO FRULLONE** (Direttore delle ricerche Fondazione Ugo Bordoni)

**PASQUALE CAIANIELLO** (Computer science, Università degli studi dell'Aquila)

**ADDARIO MARIA GIOVANNA** (Componente del CO.RE.COM. Emilia Romagna)

Editor: MASSIMO PARRUCCI (Corecom Emilia-Romagna)

## **“Diversità, algoritmi di raccomandazione e PSM: salvaguardare gli obiettivi del servizio pubblico multimediale”**

### **Elenco degli interventi**

<b>STEFANO CUPPI</b>	<b>5</b>
<b>FRANCESCO RICCI</b>	<b>5</b>
<b>PAOLO CREMONESI</b>	<b>16</b>
<b>STEFANO CUPPI</b>	<b>24</b>
<b>FABIO BASSAN</b>	<b>24</b>
<b>STEFANO CUPPI</b>	<b>28</b>
<b>ERIK LAMBERT</b>	<b>28</b>
<b>FRANCESCO SILIATO</b>	<b>32</b>
<b>ANDREA MONTANARI</b>	<b>36</b>
<b>PIERO DE CHIARA</b>	<b>38</b>
<b>ANDREA MONTANARI</b>	<b>45</b>
<b>GIANLUCA VISALLI</b>	<b>46</b>
<b>GINO ALBERICO</b>	<b>48</b>
<b>MARINA CAPORALE</b>	<b>51</b>
<b>FLAVIA BARCA</b>	<b>55</b>
<b>GUALTIERO MAZZI</b>	<b>56</b>
<b>MARIO FRULLONE</b>	<b>58</b>
<b>PASQUALE CAIANIELLO</b>	<b>59</b>
<b>ANDREA MONTANARI</b>	<b>60</b>

## STEFANO CUPPI

L'idea di discutere sulle impostazioni dei *Recommender System* del Servizio Pubblico Multimediale venne avanzata da Gian Paolo Tagliavia (allora CDO RAI) nel corso di un seminario Agcom nella tarda primavera del 2018. Successivamente Tagliavia è stato promosso amministratore delegato di RAI Pubblicità; io però ho coltivato questa idea poiché l'avevo trovata molto interessante; nel frattempo sono diventato Presidente del Co.Re.Com. Emilia Romagna e in questa veste ho cominciato a tormentare il neonato Ufficio studi della Rai sul tema. Il direttore Montanari ha accettato di discuterne, e lo ringrazio per questo, e quindi abbiamo di comune accordo cercato di selezionare le personalità che in questo Paese a vario titolo noi riteniamo essere le più competenti nei vari aspetti che toccano questo momento di transizione verso la tv non lineare del servizio pubblico e devo dire che sono molto contento perché tutti coloro che abbiamo contattato hanno accettato l'invito.

La discussione di oggi è completamente libera; non ci sono tesi precostituite e abbiamo organizzato i lavori secondo una modalità molto semplice.

Il primo intervento sarà del professor Francesco Ricci, che è uno dei massimi esperti di algoritmi di raccomandazione nel nostro Paese, al quale ho chiesto di fare una breve rassegna su che cosa sono gli algoritmi di raccomandazione, per poi passare a un approfondimento specifico da parte del professor Cremonese sugli algoritmi di raccomandazione nel video on demand poi seguiranno gli altri interventi; do quindi la parola al professor Francesco Ricci

## FRANCESCO RICCI

Professore di informatica alla Libera Università di Bolzano della quale è attualmente preside; specialista dei *Recommender Systems*, ha pubblicato un *Handbook* di un migliaio di pagine proprio su questa categoria di algoritmi; membro dell'*Association for Computing Machinery* (ACM) che gestisce la principale Conferenza internazionale nel settore

- Perché servono i Recommender Systems: il paradosso della scelta

Ci troviamo di fronte ad un'abbondanza di scelte sia nel retail ed anche nei multimedia. E adesso ci sono nuovi settori in cui è importante per le persone prendere delle decisioni su temi anche molto complessi. Si pensa spesso che il fatto che ci siano più scelte ci permetta di vivere meglio: possiamo scegliere le cose che sono più adatte a noi. Ma in realtà tutti gli studi che sono stati fatti nell'ambito della psicologia hanno dimostrato che avere più scelte non rende le persone più felici ma anzi ha un costo: il costo della decisione, che è un costo che poi deve essere scontato

sulla decisione che viene presa.

I *recommender systems* sono nati appositamente per ridurre questo costo.

- L'origine degli *Recommender Systems*

Dal punto di vista informatico fino a qualche anno fa, lo scenario di ricerca delle informazioni tipico era quello dei motori di ricerca: una persona ha una richiesta di informazione fa delle 'query', delle domande più o meno strutturate. Sui motori di ricerca vengono riportati dei risultati, e in questo processo, tipicamente ciclico, le persone vengono a raffinare la loro ricerca per avere dei risultati più interessanti.

Questo tende un po' scomparire, cioè i nuovi sistemi - ci sono gli esempi di Instagram di Facebook - cercano di sfruttare un approccio diverso ed usare appunto una serie di segnali che vengono dall'interazione dell'uomo con la macchina per annullare la necessità di fare delle query esplicite, e per ritornare informazioni personalizzate per l'utente.

L'impatto è grossissimo perché poi dipendiamo da questa selezione da parte degli algoritmi. Ma l'approccio è efficiente: dalle statistiche sull'uso dei motori di ricerca si vede che Google è arrivato ad un regime stabile, il numero delle query non aumenta - anzi sta tendendo a diminuire - proprio perché l'informazione viene raccolta in altro un altro modo e da qui nasce appunto l'esigenza di avere questi sistemi di raccomandazione.

I sistemi in principio dovrebbero fornire delle informazioni rispetto a degli item che sono considerati utili per l'utente. Questa idea di quello che è utile è quello che poi caratterizza algoritmi diversi e loro ottimizzazione. Si cerca di dare informazioni rilevanti ma anche informazioni adatte al momento in cui la persona sta facendo la ricerca. Dovrebbero migliorare in qualche maniera la nostra qualità della vita e dovrebbero essere delle scelte in qualche maniera razionali.

- Canali di raccomandazione diversificati

Tra gli esempi più classici c'è Amazon, dove il motore di ricerca è una delle componenti principali dello sviluppo della piattaforma. Ovviamente Amazon fornisce altri servizi, ma nell'ambito dell'e-commerce c'è un gruppo notevole di ingegneri che lavora solo sul meccanismo di raccomandazione. Il suo sistema di raccomandazione è strutturato in una maniera molto particolare a "strips" (strisce) dove ogni strip è di fatto un canale di raccomandazione differente perché usa un'euristica, un algoritmo differente. Questo è diventato un pattern abbastanza noto che anche Rai Play ha imitato.

- Il ruolo dell'interfaccia grafica e il *longterm implanting*

Anche Netflix è arrivato alla conclusione che la ricerca sugli algoritmi ha raggiunto il massimo del possibile (il limite) e che il problema in questo momento è proprio

quello dell'interfaccia grafica e come rendere questa interfaccia più efficace. Personalizzare addirittura l'interfaccia - cioè la grafica - ad ogni singolo cliente è un problema legato ad un'altra questione: non è più un problema di trovare l'item che va bene per la persona, perché ormai con gli algoritmi siamo arrivati all'osso, ma i problemi più grossi sono il "*longterm implanting*".

Si cercare di trovare un modo per suggerire item che in termini di ritorno negli anni o nei mesi possa tenere il cliente vicino alla piattaforma, e che quindi impedisca al cliente di andare su altre piattaforme. Questo ovviamente è molto più complicato e su questo la ricerca è ancora molto indietro.

- Elicitazione delle preferenze e raccomandazioni

Vi parlerò invece di come troviamo quei singoli libri e filmati che vogliamo visualizzare in queste strips. Ci sono tre fasi: primo c'è una fase di elicitazione della preferenza o del comportamento, poi segue una seconda fase dove sono questi cosiddetti algoritmi di predizione che producono predizioni su qual libro o film potrebbe piacere, e poi segue una terza fase dove l'output di questi algoritmi viene utilizzato per presentare queste raccomandazioni. Questo sono le tre fasi fondamentali per l'accettazione delle *preference*. Ci sono vari approcci e varie modalità per cercare di predire gli item rilevanti per l'utente

Prendo sempre l'esempio della musica: mi piace Brahms e mi piace Bach, ma mi piace anche qualcos'altro.

L'uso dei *rating* è tipico di Amazon: gli item che sono stati acquistati vengono valutati in una scala da 1 a 5, ma ci sono anche altre forme di valutazione che adesso si affermano. Per esempio, pensate a Facebook, dove di fatto esprimiamo delle valutazioni con delle emozioni, dando un contenuto emozionale a quello che stiamo vedendo con i like, ecc. L'altra cosa che viene utilizzata è la pura osservazione del comportamento, quindi non sono neanche delle valutazioni esplicite ma sono semplicemente il fatto che uno ha fatto una certa azione sulla piattaforma.

Questi feedback ovviamente sono molto più difficili da interpretare ed è molto più complicato da quelli ricostruire le preferenze dell'utente. Sono "sporchi", cioè noi possiamo arrivare ad una pagina e ritenere la pagina poco interessante, ma la piattaforma web riconosce comunque che tu hai visitato una certa pagina - se questa pagina poi è interessante o meno, non siamo in grado di saperlo.

- Problemi di campionamento

Un altro dei grossi problemi che vorrei enfatizzare è il fatto che questi feedback non sono campionati "random", cioè non è che noi prendiamo un insieme di prodotti e tiriamo fuori questi prodotti in maniera random per chiedere alle persone cosa ne pensi di questo e cosa ne pensi di quest'altro. Questo ci permetterebbe di avere una valutazione "un-biased" delle preferenze delle persone. In realtà il campionamento

è fatto totalmente non random perché le persone vanno nella piattaforma, qualunque sia la piattaforma, e valutano tipicamente solo le cose che cercano, che sono interessati e che sanno che esistono. Quindi c'è un bias legato al fatto che le persone valutano tipicamente soprattutto le cose che interessano e le cose di cui conoscono l'esistenza. Su ciò di cui non conoscono l'esistenza, il sistema non conosce le preferenze dell'utente e questo è uno dei più dei grossi limiti poi per l'algoritmo che deve fare la predizione, perché si basa su quello che ha appurato dell'utente e l'utente potrebbe avere avuto un comportamento sub-ottimo. Non ha mostrato alla macchina veramente che cosa vorrebbe ed è questo il motivo per cui adesso si cerca anche di combinare osservazioni con dati più espliciti, in cui la persona può dire che cosa è interessato a fare. Quindi si riesce un po' ad entrare in quel territorio che non è stato sondato.

- Costruzione del modello predittivo e come evitare le “filter bubble”

La parte invece che ha più enfasi nella ricerca è la parte di costruzione del modello predittivo.

Anche Amazon soffre del fatto che conoscere il film che è stato acquisito è molto parziale e legato solamente a quello che l'utente ha realmente visitato. Ed è anche questa la causa principale del fatto che si formano queste Filter Bubble di cui si sente tanto parlare: non è tanto prodotto dall'algoritmo - questa è la mia opinione - ma è prodotto dal fatto che l'algoritmo si basa su dati che sono legati al comportamento delle persone, dati che sono necessariamente limitati dalle conoscenze delle persone. Non è dato alle persone la possibilità di guardare a 360 gradi le offerte, perché le offerte ovviamente sono milioni e quindi il sistema comunque deve fare un filtraggio.

Adesso ci sono dei metodi che si chiamano “N-armed bandit” - non so se ne avete sentito parlare - che sono algoritmi che tentano di fare più esplorazione proprio da questo punto di vista, cioè che tendono a suggerire ogni tanto anche delle cose totalmente fuori dal modello previsionale, proprio per far sorgere qualcosa che gli utenti non hanno ancora detto. Ovviamente, questo va fatto con molta attenzione perché se iniziasse a mostrare all'utente Francesco Ricci le scarpette da ballo, pensando che forse potrebbe essere interessato, è probabile che faccia degli errori. Però è possibile anche che questa persona sia interessato alle scarpette da ballo...

- Come calibrare questa esplorazione

Calibrare questa “exploration” è un problema molto grosso, per quanto esploro quanto solo quello che so già. Di fatto quindi siamo un po' di nuovo in un problema classico: quanto devo sfruttare della conoscenza che ho già? I dati di cui i sistemi classici hanno disposizione sono abbastanza semplici, un identificativo per l'utente o per l'item.

In questo caso si potrebbe forse anche usare alcuni dati di contesto come la data, il

luogo, altre condizioni contestuali come il media, il “device” utilizzato. Poi qualche forma di feedback, lo “score” (il rating) che potrebbe essere da uno a cinque.

Movie rating data							
Training data				Test data			
user	movie	date	score	user	movie	date	score
1	21	5/7/02	1	1	62	1/6/05	?
1	213	8/2/04	5	1	96	9/13/04	?
2	345	3/6/01	4	2	7	8/18/05	?
2	123	5/1/05	4	2	3	11/22/05	?
2	768	7/15/02	3	3	47	6/13/02	?
3	76	1/22/01	5	3	15	8/12/01	?
4	45	8/3/00	4	4	41	9/1/00	?
5	568	9/10/05	1	4	28	8/27/05	?
5	342	3/5/03	2	5	93	4/4/05	?
5	234	12/28/00	2	5	74	7/16/03	?
6	76	8/11/02	5	6	69	2/14/04	?
6	56	6/15/03	4	6	83	10/3/03	?

Questi dati sono l’input fondamentale per il sistema di raccomandazione che cerca di prevedere osservazioni a venire. Dovete tener presente che il sistema viene addestrato in questo modo per capire se l’algoritmo funziona bene o male. Si fa prendono questi dati dividendoli a metà: metà di questi dati vengono utilizzati per addestrare il modello e l’altra metà viene utilizzata per capire se il modello è corretto, se la previsione corrisponde con quello che è veramente osservato.

Movie rating data							
Training data				Test data			
user	movie	date	score	user	movie	date	score
1	21	5/7/02	1	1	62	1/6/05	?
1	213	8/2/04	5	1	96	9/13/04	?
2	345	3/6/01	4	2	7	8/18/05	?
2	123	5/1/05	4	2	3	11/22/05	?
2	768	7/15/02	3	3	47	6/13/02	?
3	76	1/22/01	5	3	15	8/12/01	?
4	45	8/3/00	4	4	41	9/1/00	?
5	568	9/10/05	1	4	28	8/27/05	?
5	342	3/5/03	2	5	93	4/4/05	?
5	234	12/28/00	2	5	74	7/16/03	?
6	76	8/11/02	5	6	69	2/14/04	?
6	56	6/15/03	4	6	83	10/3/03	?

- La difficoltà di correggere un modello basandosi sul passato

Se no, sulla base di questo il modello viene corretto e qui di nuovo il problema che vi dicevo prima: se addestrato e correggo il modello sulla base di quello che ho appreso in passato della persona, faccio interpolare del modello per quello che non ha

dimostrato di essere interessata, ma non sono mai in grado di uscire da questa “bubble” se non riesco a trovare altri metodi.

Anche se non so

se c'è una fase di costruzione del vestito su misura dentro l'obiettivo o se c'è un pezzo della funzione obiettivo che invece cerca di non costruire il vestito su misura, non ne sono certo come questi due diversi pesi per la fase di generalizzazione interagiscono, anche quando i dati sono derivati da una persona singola.

È chiaro che se ci sono delle tecniche per tentare di generalizzare, ma tutti ovviamente presentano difficoltà, anche perché il problema, quando si parla di machine learning, è mal posto.

È impossibile dai dati costruire un'unica soluzione, cioè i dati non sono sufficienti per definire tutti i parametri di questi modelli. Rimane il fatto che i metodi di valutazione che vengono utilizzati in questo momento tendono a dichiarare un modello come corretto se riesce a predire correttamente quello che le persone hanno già fatto.

Questa è la mia grande critica, ed è una critica che è abbastanza discussa. Il problema è che i recommender systems vengono testati sui dati che sono già stati acquisiti perché, a meno che non fai un test online - quindi prende il modello e dice questo è il modello della persona che ho appreso, dopodiché metto la persona davanti alla raccomandazione e dico valuta quello che ti viene raccomandato - non sono in grado di capire quanto è in grado di veramente di generalizzare; questa è la prima cosa.

- Che vogliamo veramente raccomandare?

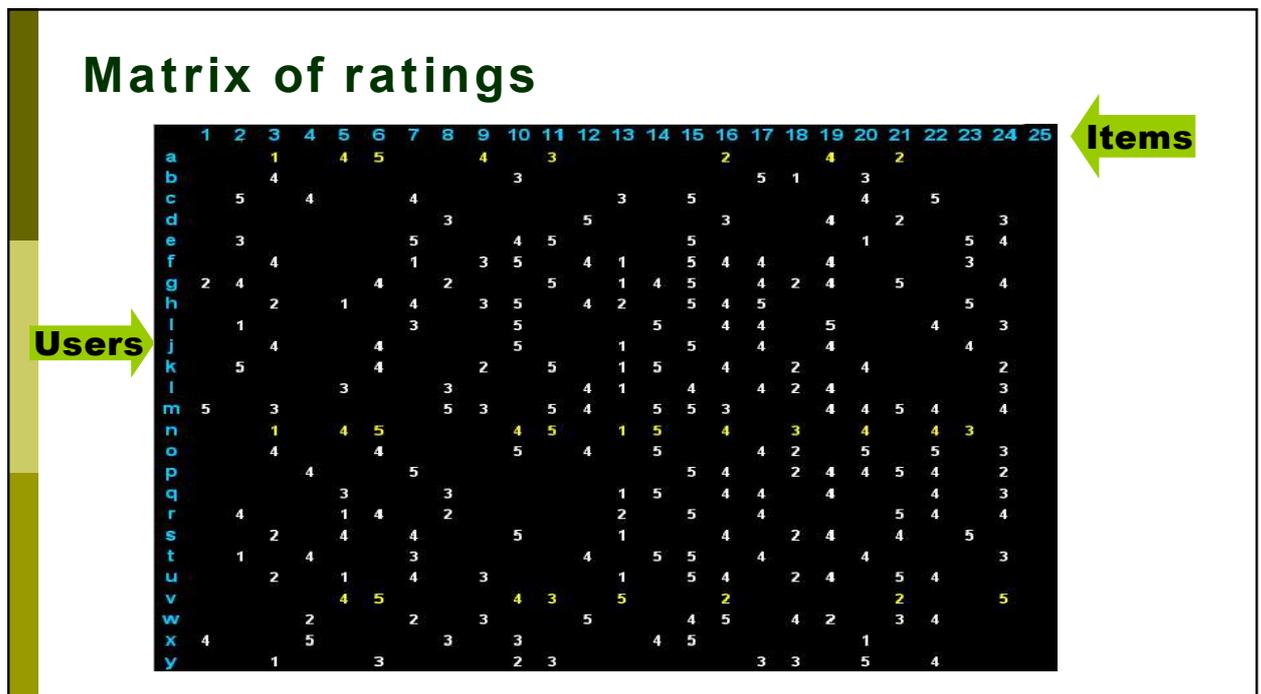
La seconda cosa è che cosa vogliamo raccomandare: vogliamo raccomandare delle cose, degli item che le persone riconoscono come interessanti o vogliamo raccomandare delle cose che pensiamo siano interessanti per loro? Già qui il problema è diverso perché quello che io so che è interessante è di nuovo la mia sfera di conoscenza e se io gli volessi suggerire qualcosa che penso che dovrebbe essere interessante per lui non lo riconoscerà nemmeno che è interessante, ma si dovrà fidare della raccomandazione del sistema. Questo tipo di cose ovviamente non possono essere fatte con il “fine tuning” dell'algoritmo e devono essere fatte studiando il comportamento delle persone quando interagiscono con questi algoritmi. Bisogna entrare in un'idea più sofisticata della funzione da ottimizzare e quindi non è un problema solo di data mining.

Enfatizzo un po' la mia critica al data mining perché questo è il pattern che ora è

molto diffuso; chiunque piglia in mano queste cose si legge un algoritmo, lo implementa e pensa di aver fatto il recommender. Di fatto anche chi fa il recommender, vi ho detto prima di Netflix, ritiene che siamo arrivati al limite cioè non è un problema adesso di trovare l'algoritmo che va a fare invece di 90 per cento e 90.4 per cento di accuratezza; bisogna capire un po' meglio qual è il problema che vogliamo risolvere dal punto di vista del dell'utente.

Questo è anche quello che osserviamo molto spesso di questi suggerimenti, ci suggeriscono delle cose che possono essere totalmente sbagliate; questo dipende dal fatto che usano un certo tipo di forma di generalizzazione, perché ovviamente sui movimenti generalizzano, quindi non raccomandano esattamente quello che noi faremo; hanno un modo di generalizzare che è interno al meccanismo di apprendimento. Lì si potrebbe dire che c'è un algoritmo migliore o peggiore che generalizza in maniera diversa; però fondamentalmente vengono testati ottimizzando l'accuratezza nel prevedere quello che le persone hanno già fatto; però adesso non so quanto entrare in questi dettagli.

- Alla base, l'uso di matrice



Quello che convergo e che utilizzano molti di questi argomenti sono matrici di questo genere. Matrici in cui abbiamo gli utenti sulle righe abbiamo item sulle colonne e immaginate che nell'incrocio c'è la valutazione o l'osservazione di un utente rispetto a quell'item; quindi questo per esempio potrebbe essere appunto

per l'item **10** alla valutazione **4** per l'utente **e**, e da queste matrici, che sono tipicamente sparse, nel senso che qui le vedete belle riempite di numeri ma sono tipicamente praticamente tutte nere e forse ogni tanto compare un numerino, se si volesse fare la ricostruzione della matrice si dovrebbe riempire questa matrice con tutti i numeri possibili, cioè riuscire a mettere per ognuno di questi incroci utente e item un numero che corrisponde a quella predizione di questo di quanto potrebbe essere interessante quel item (il numero corrisponde a una valutazione). Nostro sistema di Netflix ha le stelle per queste cose, ma è un lavoro.

Se la valutazione non è esplicita, si indica semplicemente il fatto che ho visitato la voce, e la valutazione è tipicamente 1 o niente. Si può anche estrarre una valutazione implicita – se mi sono permesso di vedere quel film c'è già una griglia di gioco, di pesi e dei comportamenti di consumo rilevati, però il rating esplicito ha un peso più alto. Si combinano i rating espliciti con i segnali che provengono dall'utente, si usa il tempo per scalare l'importanza delle cose vecchie, tutto quello che voi potete immaginare state pur sicuri che è stato fatto.

- Il nemico è il “noise”

Ma tutti questi segnali impliciti sono molto molto “noisy” sono molto disturbati. Ci sono ovviamente adesso algoritmi che cercano di usare solamente questi segnali impliciti: il fatto che tu abbia fatto un accesso a una pagina, per quanto tempo, se hai ascoltato quella musica. Ormai sono stati scritti migliaia di articoli su questi algoritmi e tutto quello che poteva essere utilizzato è stato utilizzato.

Il problema fondamentale è che per prima cosa questi algoritmi generalizzano in maniera scorretta perché non hanno abbastanza dati per sfruttare tutti questi segnali, e la seconda cosa è che le persone sono molto dinamiche e gli interessi delle persone cambiano. Tutti questi algoritmi tendono a pensare che tutto quello che è stato il passato si ripercuota nel futuro; non sono in grado veramente di prevedere che io oggi sono interessato a “Un passo dal cielo” e domani sarò interessato a qualcos'altro. Quindi tenete presente che queste limitazioni non sono risolvibili da algoritmi chiaramente, perché è un problema intrinseco di quello che stimiamo.

- Algoritmi molteplici e fonte dati molteplici

È vero che sono tutti quelli che hanno impiegato due o più algoritmi. Hanno cercato di allargare la loro base di rilevazione e di misurazione dei comportamenti, p.e. Google o Facebook. In realtà misurano cose a naso, veramente. L'esempio principale è misurare col suo video su una parte di e-commerce e aggiungere i comportamenti in ambiti diversi per ricostruire i profili.

Una delle tecniche per risolvere questi problemi sono quei sistemi liberi che tentano di sfruttare informazioni che vengono da varie fonti, o anche algoritmi diversi

combinati insieme, perché gli algoritmi sono più bravi in certe situazioni o argomenti, e sono più bravi combinati insieme. Di nuovo si riesce a risolvere alcuni di questi problemi ma poi bisogna però ottimizzare l'algoritmo diciamo al "meta level" che combina gli algoritmi, ecc.

Si va un po' a periodi: diciamo fino a qualche anno fa tutti gli algoritmi che stavano in materia "factorisation" erano quelli un po' poco utilizzati, tutti quelli appunto che tendono a fare la decomposizione di queste matrici sparse in due matrici dense. E posso con tecniche sia algebriche sia di apprendimento decomporre questa matrice in due matrici di più piccole dimensioni, però dense, ci sono varie variazioni per fare questa decomposizione. Fino a qualche anno fa questo era un po' il leitmotiv; prima ancora erano gli algoritmi tipo di "collaborative filtering" più classi che tendono ad osservare le correlazioni tra righe o colonne di queste. Adesso le neural network stanno diventando il non plus ultra e questo diventa ancora più complicato perché è un'architettura molto complessa, fatta da diversi componenti lineari e non lineari che sommano tutte queste informazioni e spara fuori un output che viene di nuovo calibrato, tipicamente quando si hanno a disposizione tantissimi dati per molti utenti.

In realtà non abbiamo tanti dati. Pensate che ne so dall'utente che entra in RaiPlay: cerchiamo di entrare un po' - allora quanti film ha riguardato, quando entra, tipi. Molti degli utenti, la stragrande maggioranza di utenti di queste piattaforme sono cold user, sono utenti freddi di cui non si sa nulla e in cui se noi utilizziamo l'algoritmo che è stato pensato per funzionare invece per Spotify - i suoi utenti che hanno centinaia di visite e centinaia di rating - fanno tipicamente niente. E quindi ci sono anche dei meccanismi di "fall back" se non funziona quello personalizzato, vado dentro quello non personalizzato. Nessuno è stupido se quando si fanno le soluzioni tipicamente per il cold user, tutti questi argomenti qui non servono praticamente a nulla, soprattutto perché richiedono enormi quantità di dati per essere addestrati.

Allora al di là delle considerazioni generali, vorrei parlare dei dati da Facebook per arricchire la profilazione degli utenti. Ovviamente gli algoritmi consumano i dati e appunto se non siano abbastanza autorizziamo nel momento in cui ci ispiriamo a questa in ospitati da Facebook e lavoro sul profilo Facebook.

Però non dobbiamo immaginare che in effetti utilizzi chissà che dati di Facebook, perché di nuovo utilizzare dati eterogenei all'interno dell'algoritmo non è poi così banale, e quindi Netflix usa fundamentalmente i dati a disposizione che colleziona dall'uso della propria piattaforma.

- L'accuratezza tecnica non è un obiettivo commerciale

Ovviamente tutto si potrebbe utilizzare a prescindere anche dei problemi ingegneristici, gestire un servizio che utilizza dati che vengono da n parti o gestire un algoritmo molto complesso che richiede fase di addestramento potrebbe non essere

vantaggioso, tenendo presente il fatto che vi ho appena detto che ci sono degli argomenti molto sofisticati che aumentano dell'1 per cento l'accuratezza teorica. Ma non è detto che questo si trasformi in un reale percezione dell'utente di un migliore algoritmo; quindi infatti tutti i gestori di queste piattaforme arrivano a fare un bilanciamento. Quando una decina d'anni fa Netflix ha lanciato una competizione offrendo un milione di dollari a chi ha ridotto del 10 per cento l'errore, questo algoritmo è stato prodotto e sono stati dati il milione, poi non hanno usato questo algoritmo perché era talmente complesso che non produceva un reale vantaggio rispetto a un argomento molto molto più semplice.

C'è un caso particolare per usare dati esterni, con un campione diverso, nel senso che usa dati di Facebook e di Twitter per fare la linea che si chiama "Trending". Perché permette di reperire le cose che sono dentro delle domande e perché è più veloce dell'uso interno al sistema. Fanno molte altre cose sulla profilazione, fermo che nel primo caso molto particolare, e si parliamo di video si deve paragonare un po' come funziona YouTube o Vevo.

- L'importanza della presentazione delle raccomandazioni

Volevo per chiudere almeno accennare quello che nella ricerca viene un po' sottovalutato ma di fatto è un'altra cosa molto importante: come da dopo la parte di predizione si decide che cosa mostrare, che cosa hanno mostrato, perché lì a quel punto. C'è da fare una selezione e da decidere come presentare le informazioni, la selezione, perché non tutto quello che potrebbe essere raccomandato può essere raccomandato. Bisogna mostrare questi prodotti in qualche maniera, e lì c'è anche molta innovazione da fare. Lo dicevo prima, questi "strips" (le strisce) di e-business, usati anche di Netflix, magari troveremo un modo diverso per presentare le informazioni, e anche le raccomandazioni di Facebook, che hanno un approccio totalmente diverso, tutti questi post sono raccomandazioni perché io non ho la visibilità su tutti i post, quindi lì c'è molto da fare e lì devo dire che entrano - dal mio punto di vista - delle ricerche anche un po' più interessanti dove si combinano gli algoritmi con la psicologia e con l'economia, per cercare di capire qual è la percezione dell'utente.

Una cosa che volevo dirvi, e chiudo con questo, è che ci sono due concetti da tener presente. La raccomandazione è una cosa, cioè un item viene raccomandato e quindi l'utente ne viene rinforzata in lui il desiderio o l'interesse per quell'altro. Ma c'è un altro aspetto che noi chiamiamo la "awareness", cioè le persone devono essere rese consapevoli che c'è un certo numero di prodotti che possono essere consumati. Queste due cose sono diverse nel senso che a questo meccanismo qua dell'estrema aumenta notevolmente la "awareness", cioè la consapevolezza che c'è una quantità di item che possono essere utilizzati. Stiamo facendo delle ricerche proprio su questo, cioè cercando di capire come la "awareness" e la raccomandazione che giocano

entrambe per determinare la diversità delle scelte delle persone, e più che la raccomandazione da sola determinano la diversità delle scelte. Vuol dire che quindi tutto il catalogo viene effettivamente utilizzato anche in maniera uniforme, e la “governance” è quindi rendere consapevoli gli utenti che esistono dei prodotti che sono effettivamente disponibili. Questa secondo me è la parte anche più rilevante per quello che noi dobbiamo fare perché il nostro obiettivo è più aumentare questa consapevolezza, l’estensione e la diversità delle informazioni per l’utente.

- Algoritmi molteplici e trasparenza

Nel caso in si usa diversi algoritmi di raccomandazione uno si sposta in un terreno completamente diverso: si trova oggi con traffico - certo - ma e sono in realtà connessi perché vediamo che ognuna di queste strip sono in qualche misura delle raccomandazioni. Però il trucco è quello di moltiplicare il numero delle raccomandazioni che possono essere consumate senza ovviamente esagerare! Se noi avessimo la classica lista invece della quotazione di Amazon, quella precedente, più di tanto le persone non scendono in questa lista e la “awareness” e comunque è molto più bassa. Quindi il la differenza - se volete leggere in questo senso - qua sono due sistemi di raccomandazione, l’una che dice “raccomando questo” e l’altra invece che sarebbe l’organizzazione di tutta questa striscia.

Qui è tutta una raccomandazione che è basata sui miei comportamenti, che però è meno esplicita. Diciamo che qui una non rende esplicito il meccanismo di filtering; mentre è reso esplicito come questo film The Ring viene raccomandato, perché almeno si dice questo te lo raccomando perché hai dimostrato interesse per quello o questo, te lo raccomando perché sono le ultime prodotti aggiunti. Quindi da questo punto di vista, invece aumenta anche la trasparenza. Che poi questo che loro dicono se quello che viene fatto ovviamente nessuno lo può sapere perché ovviamente di nuovo quanti sono gli altri che sono aggiunti al catalogo non certo quelli che ci vengono ma molti di più e di nuovo ci è stata fatta una selezione algoritmica.

Come uscire da qui?

L’unico modo reale è quello di fare la ricerca, cioè se tu fai la ricerca e sai cosa vuoi, lo trovi perché immette le parole chiave. D’altronde anche se il titolo che state cercando è conosciuto, la risposta non è certa, poiché mettendo il nome, è solo uno a confronto delle 7000 categorie su quale si fanno le ricerche.

## PAOLO CREMONESI

Professore al Politecnico di Milano, si occupa da una decina d'anni di sistemi di raccomandazione proprio per la tv e video on demand. Una decina di anni fa ha col Politecnico creato una spin-off - una start-up - che si chiama ContentWise e che fa sistemi di raccomandazione per la tv lineare. È stato un sistema usato in RAI.

Cercherò di dare degli spunti su quali sono le peculiarità del mondo del video on demand e in particolare come si valutano la bontà delle raccomandazioni. Questo è un aspetto importante nella scelta di un sistema di raccomandazione da implementare in una piattaforma.

- Peculiarità del mondo del video on demand

### Application scenarios

- VoD recommender systems should be able to address a number of different scenarios in which traditional systems fail
  - A. Context: react to user intentions
  - B. Diversity: promote a wide range of different items
  - C. Long-tail: surprise users with niche content
  - D. When and where: screen size and position
  - E. New items: react to short-lived items (live TV, news, ...)
  - F. Mise-en-scene

Qui ho elencato solo alcune delle tantissime peculiarità della raccomandazione nel mondo del video on demand e alcune di queste sono magari comuni anche da altri domini.

Faccio un esempio: il contesto. Quando accendo la televisione non è solo necessario capire quali sono i miei gusti, ma magari voglio accendere la televisione perché voglio rilassarmi, oppure perché sono con mia figlia e quindi voglio guardare qualcosa insieme a lei, oppure perché sono con mia moglie e voglio vedere una cosa romantica. Quindi c'è un aspetto molto sfidante: è quello di capire il perché voglio guardare la televisione.

L'altro aspetto importante è quello - non comodo - di capire qual è lo scopo del mio sistema di raccomandazione. Se pensiamo ai classici Netflix, lo scopo è chiaro: voglio tenere la persona sulla piattaforma di video on demand il più a lungo possibile.

Ma anche nel mondo della tv lineare voglio aumentare i miei dati di ascolto per ampliare la raccolta pubblicitaria. Ma possiamo anche ragionare e decidere che

questo non è il mio obiettivo principale. È importante capire quale sia perché poi quando andrò a realizzare o implementare un sistema di raccomandazione, devo essere in grado di misurarne l'efficacia.

- La sfida dell'efficacia per un obiettivo dato

Alla fine tutti i sistemi di raccomandazione sono in grado di imparare il comportamento dell'utente. Per sapere quanto questo si discosti o meno da quella che era la mia intenzione, si deve poter misurare la mia intenzione. Mia intenzione potrebbe anche essere "voglio aumentare la conoscenza di italiani per quanto riguarda la storia della Repubblica", ma se questa è l'intenzione devo essere in grado di misurarne l'efficacia: senza questo, il fatto che funziona o non funziona qualsiasi sistema di raccomandazioni diventa totalmente aleatorio

Se il mio obiettivo principale è quello di mantenere la persona più a lungo sulla piattaforma - che è quello che viene fatto spessissimo da questi sistemi - posso andare a cercare l'approccio più facile, quello di suggerire dei contenuti blockbuster e li metto nel catalogo; mi compro i diritti del Trono di Spade, lo metto nel mio catalogo, faccio un "push" editoriale di questo contenuto. È difficile che sbagli perché ho dei dati di ascolto, dati presi dai social e da soggetti esterni che mi dicono che è una soluzione vincente. Però torna quella che viene chiamata la Filter Bubble, ho costruito una bolla che mi garantisce sicurezza nel fatto che tengo la persona sulla piattaforma, ma non sto aiutando l'utente a scoprire cose che l'avrebbero sorpreso, quella che viene chiamata la serendipity. Quindi è importante comunque che un buon sistema di raccomandazione sia anche capace di rompere questa bolla.

Non è facile però, pensate appunto lo scenario del fatto che voglio suggerirei al mio utente un documentario di storia: so che quell'utente è refrattario i documentari, proprio non li vuole guardare, ma io, sistema di intelligenza artificiale, sono convinto di aver trovato un documentario che all'utente piace e propongo nella mia bella strisciolina delle cose, una locandina dove c'è scritto documentario di Alberto Angela sulla storia dell'antica Roma. Io utente lo vedo, dico no no no, lo salto proprio e quindi io, sistema intelligente, devo, come diceva il professore, spiegare all'utente perché vi consiglio questo contenuto. Devo motivarlo.

Vi faccio un esempio su cosa sta facendo Netflix in questa direzione. Arriva a personalizzare le singole locandine che sono create dinamicamente e in modo tale da convincere l'utente con i colori, con la scelta dell'immagine, con quello che è scritto, dove è posizionata la scritta. Anche i trailer sono personalizzati perché si deve motivare l'utente.

- Il dispositivo

Altre peculiarità del mondo della tv lineare del video on demand è il dispositivo con

che io utilizzo per fruire dei miei contenuti. La televisione e eventualmente il telefonino sono l'interfaccia con cui interagiscono gli utenti. Interfaccia che ha una serie di limitazioni nel caso della televisione: il telecomando non è la tastiera o il mouse quindi mi permette solo un'interazione limitata. Uno dei motivi per cui la scelta dei caroselli o delle strisce funziona è perché con le frecce del telecomando riesco a scorrere l'elenco-palinessto in maniera ragionevole. Però a quel punto diventa importante come creo il carosello, il palinessto.

- La corta vita dei contenuti

Un altro problema tipico del mondo della video on demand e della tv lineare è il fatto che molti contenuti sono effimeri o sono entrati da recente nel catalogo.

Pensate a un particolare evento, una diretta: questo contenuto è interessante per l'utente soltanto per una finestra di tempo limitata, e quando questo contenuto è tipicamente nuovo. Questo significa che nel momento che entra nel mio catalogo, non ho uno storico di opinioni dirette o indirette di altri utenti che hanno guardato questo contenuto. Quindi molti algoritmi che si basano su questo storico di interazioni fanno fatica ad operare. È sfidante anche da questo punto di vista. Insomma ci sono delle sfide di ricerca in questo ambito su come andare ad affrontare questo tipo di contenuti.

- Necessità di una interazione veloce

L'altra peculiarità del mondo del video on demand rispetto ai sistemi di raccomandazione degli e-commerce tradizionali è la velocità di interazione. L'utente quando usa la televisione è abituato a un sistema che risponde immediatamente, quindi non posso permettermi di introdurre algoritmi troppo sofisticati di machine learning, che magari richiedono un secondo. Un secondo nel mondo della tv è troppo come tempo di interazione; quindi questa è una priorità che rende sfidante la raccomandazione.

- Chi guarda?

L'altro problema è il fatto che io non so chi c'è davanti al televisore. Il televisore è un elettrodomestico condiviso da tanti membri della famiglia: quando mi baso sulle interazioni, sullo stesso device, il sistema traccia cosa ho visto io, cosa visto mia moglie, cosa ha visto mia figlia, gli amici di mia figlia, ma è un problema e non è banale riuscire a distinguere quali utente ha fornito quali interazioni. Questo espone anche dei rischi, perché magari guardo dei contenuti che non sono appropriati una bambina o per un adolescente. Il sistema capisce male e fa delle raccomandazioni a ai bambini che non sono adatti ai bambini.

- Le debolezze del rating

Mi aggancio anche alcune delle tematiche che sono uscite nella discussione. In precedenza si parlava di rating, cioè del fatto che io possa dare un'opinione esplicita: nella maggior parte delle piattaforme di video di on demand, il percentuale di rating che gli utenti danno è irrisoria; un sistema di raccomandazione per il VOD non può basarsi solo sui rating, anzi nella maggior parte dei casi non vengono nemmeno considerati proprio per la scarsità di questi dati. Quindi l'unica informazione a cui posso accedere sono le interazioni. Prima si chiedeva quali tipi di interazione vengono presi in considerazione? Come giustamente diceva Francesco Ricci, sono stati provati tutti i tipi di interazione, ma più interazioni vado a mettere in input al mio sistema di raccomandazione e più informazioni ho ma anche tanto rumore.

Faccio un esempio banale : se guardo la scheda descrittiva di un contenuto VOD, come lo interpreto questo segnale? Mi interessa e poi non ho guardato il film documentario che sia stato perché non avevo tempo, ma che invece lo voglio guardare più avanti, o lo interpretiamo come ho letto la trama e ho deciso che non mi piace? Non è facile dare un'interpretazione al fatto che io abbia guardato o no! O la faccio al contrario: ho guardato un contenuto video dall'inizio alla fine, ma ero sveglio o mi sono addormentato davanti alla tv?. Può anche accadere questo: se sto guardando il primo episodio di una serie tv poi mi sono addormentato e il sistema è andato avanti, mi ha fatto vedere venti episodi, ma magari era così noioso che in realtà non ho apprezzato il contenuto.

Mi meraviglia che la televisione a un certo punto si fermava e faceva uno share molto più alto di 30 certo perché tutti si addormentavano.

### Niche: Editorial vs Personalized Content

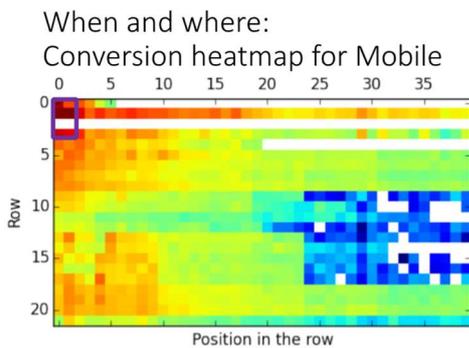
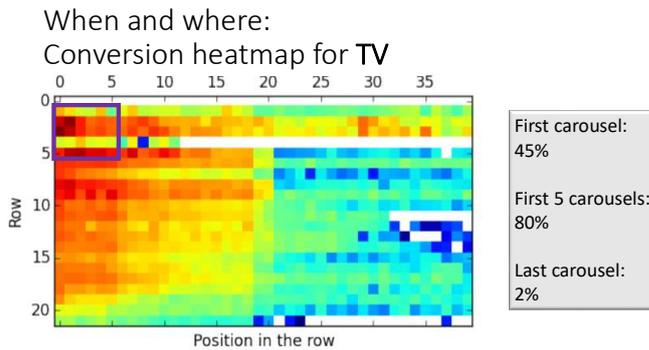
- VoD: 10% "blockbuster" items generate more than 80% of play events
- 90% long-tail items generate only 20% of play events
- The best personalized recommendations generate on average 10% play events wrt editorial strategies
- On long-tail items, personalization generates 10x play events wrt editorial strategies

Volevo darvi degli esempi degli dei del pericolo dell'effetto bolla, cioè dell'effetto dei dati numerici. Se vado a suggerire con un sistema raccomandazione i contenuti quelli che chiamo blockbuster, quelli più popolari in termini di capacità di fare ascolto, vado sul sicuro. Se ho 10 per cento di continuity blockbuster su una piattaforma, contano per l'80 per cento delle visioni. Se il mio obiettivo è quello, è inutile usare questi sofisticati sistemi. È sfidante invece andare il fare andare gli utenti su contenuti più di nicchia. Contenuti di nicchia però sono contenuti che per definizione non hanno

avuto molte visioni, e quindi è difficile andare a capire quale contenuto di nicchia è potenzialmente interessante per un utente.

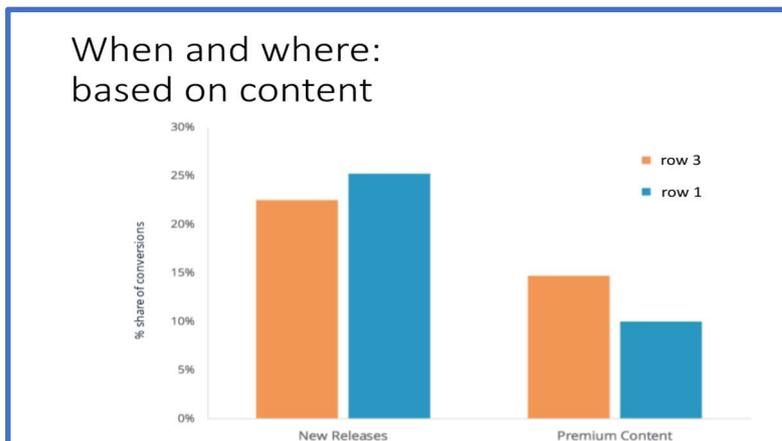
- L'importanza della posizione

qua volevo far vedere un esempio su come la posizione di un contenuto all'interno dello schermo della tv influenzi la visione. Il rettangolino viola rappresenta la dimensione della TV:



Col telecomando posso scorrere a destra posso scorrere in basso quindi ciascun quadratino colorato rappresenta un contenuto che il sistema di raccomandazione ha predetto, organizzato per caroselli e per stralcio. Ci può essere i raccomandati per te ogni riga, può essere una raccomandazione generica o può essere una connotazione suddivisa per generi. Nella cultura occidentale, scorriamo tutto dall'alto verso il basso e da sinistra verso destra. E quindi la posizione ha un'influenza fondamentale nello stimolare un utente a fare una visione.

Però la cosa strana è questa: sono sempre dati di esempio in cui qui abbiamo preso dei contenuti, li abbiamo spostati dalla riga 1 alla riga 3 e della riga 3 alla riga 1. Mi aspettavo quando si sposta qualcosa nella riga uno ne aumenti le visioni: mi han detto la riga 1 in alto è quella che tira più visione. Non è però sempre esatto, è più vero per inediti perché i contenuti blockbuster si auto promuovono da soli.



Per gli inediti, se li ho rispostiati in alto ottengo l'effetto voluto dall'attirare l'attenzione su un qualche cosa di nuovo, nel sorprendere l'utente. Quindi il posizionare all'interno della schermata ha un'influenza perché tendiamo a leggere con un certo ordine, ma il cosa mettiamo li ha un impatto importante. È tutto un ambito di esplorazione interessante.

- Le difficoltà del training e del testing

Nel mondo dei sistemi di raccomandazione, i algoritmi usati "imparano" a fare le raccomandazioni basandosi su dati vecchi, storici, cioè usando i dati di due anni fa per prevedere cosa l'utente ha visto l'anno dopo. Ma sono tutte cose vecchie e quindi l'algoritmo più di tanto non riesce a imparare. Il comportamento futuro non lo riesce a imparare se non faccio un test on line, quindi devo mettere il sistema di raccomandazione in produzione monitorando il comportamento, e poi intervenire per migliorare il sistema. Quindi devo misurare in qualche modo la qualità delle raccomandazioni.

Come faccio questa cosa? Lo posso fare con una tecnica che si chiama A/B testing che suddivide i miei utenti in categorie e a queste diverse categorie di utenti somministro raccomandazioni diverse. Poi decido a mano quali variante del mio sistema raccomandazione è meglio sulla base del comportamento degli utenti. L'A/B testing è fondamentale, purtroppo sono pochissimi i gli operatori del mondo VOD che fanno A/B testing,

Netflix si, sempre.

Ma anche nomi grossi come negli Stati Uniti AT&T che è comunque un operatore di telecomunicazioni, non fa il testing per quanto riguarda le raccomandazioni dei propri utenti. Questo ha un impatto drammatico sulla capacità di fare il tuning dell'algoritmo di raccomandazione.

- Necessità d'un intervento umano

Comunque il "tuning" dell'algoritmo richiede che ci sia qualcuno che decida e la

faccia a mano, anche dopo l'A/B testing.

Lo step successivo è quello di usare delle tecniche di raccomandazione che si chiamano di "reinforced learning on line", cioè l'algoritmo viene allenato prima su dei dati storici vecchi viene poi messo in produzione. A questo punto l'algoritmo osserva il comportamento degli utenti, raccomanda qualcosa, vede se l'utente ha interagito con il contenuto che è stato raccomandato - ha letto la scheda, ha guardato il trailer, ha guardato il contenuto - e l'algoritmo automaticamente si adatta e quindi impara online. Questi sono gli approcci migliori perché l'algoritmo si corregge man mano sulla sua capacità di prevedere e di consigliare in maniera opportuna l'utente.

L'A/B testing non si può fare su dati storici: come diceva Francesco Ricci purtroppo i dati storici non sono campionati in modo uniforme. Quindi, faccio l'esempio di nuovo del documentario, se l'utente Paolo Cremonesi non ha mai guardato un documentario di storia, nessun algoritmo andrà mai a raccomandarli un documentario di storia, perché propone qualcosa quando ha visto la reazione positiva. Dunque devo esplorare e provare a Paolo Cremonesi a raccomandare un documentario di storia e vedere come reagisce.

E questo sul passato non lo posso fare; quindi devo assumere un piccolo rischio nell'esplorare uno spazio più ampio.

- Quale è la qualità d'una raccomandazione?

Definire la qualità della raccomandazione dipende ovviamente da quali sono gli obiettivi dell'operatore. Nella maggior parte dei laboratori, la qualità vuol dire convincere l'utente a visualizzare un qualsiasi contenuto, oppure quel contenuto che mi costa di meno in termini di diritti, oppure su un contenuto premium - quindi a pagamento - che mi massimizza le "revenue".

La definizione di qualità è complicata in un ambito di servizio pubblico, e può essere ancora resa più complessa perché magari vuole stimolare i miei utenti a visualizzare produzioni italiane, produzioni europee ad esempio. Quindi la definizione di qualità è diversa a seconda dell'operatore e può non essere semplice. Ma è importante definire se è in grado di misurarla, altrimenti la valutazione di mio algoritmo è non totalmente aleatorio, ma grandemente aleatorio.

Volevo per finire lasciare due provocazioni:

Una delle frontiere della ricerca nel mondo "recommender" per andare a fare raccomandazioni che siano sempre più grandi, per sorprendere l'utente, è quello di entrare nel fare. Sono algoritmi che guardino il contenuto del video quindi che non si basino semplicemente su "chi ha guardato questo film ha guardato anche quest'altro film" oppure "hai visto tanti film con Tom Cruise e quindi ti raccomando film con Tom Cruise", ma entro all'interno del contenuto video e vado al di là di

quella che è la trama il genere. Cito sempre come esempio Romeo e Giulietta, cioè la storia è quella; ci sono 40 produzioni video totalmente diverse basate sulla trama di Romeo e Giulietta, alcune possono piacere altri non possono piacere. Cos'è alla fine che fa la differenza è lo stile di regia, come e quando è stato girato il film, ma non è nemmeno solo l'attore, per cui che so la luce che il regista ha usato, piuttosto che lo stile di movimento della telecamera, piuttosto che la profondità dei colori.

Ecco questo è uno di quegli scenari in cui gli algoritmi a cui accennava prima Francesco, di deep learning, quindi la difficile intelligence artificiale: su questi temi sono degli scenari in cui non solo può essere utile, è l'unica strada percorribile.

Negli scenari più tradizionale invece - naturalmente è solo il mio suggerimento - è evitare le soluzioni di algoritmi complessi. Paradossalmente sono in dubbio se utilizzare un algoritmo complesso, sofisticato di machine learning. Un algoritmo semplice è quello che mediamente va meglio.

Per cui se a un certo punto c'era stata una domanda su quali sono gli algoritmi più suggeriti o più di moda? Sono quelli sofisticati! Ma attenzione non è detto che è una moda nel mondo della ricerca. Ovviamente sei attirato sì la tua ricerca non è interdisciplinare, e non di raccomandazione, ma su quesiti di raccomandazione. Ma se la ricerca è algoritmi di raccomandazione e se uno lavora in modo interdisciplinare, ha una visione diversa della cosa: cioè che in alcuni punti il sofisticato, il complesso mediamente funziona peggio. Che cosa vuol dire complesso, che vuol dire funziona peggio, e che cosa vuol dire un algoritmo semplice? Non è così chiaro, la soglia dipende da noi, dipende dall'obiettivo. Ed usare un cannone per uccidere una zanzara dipende da qual è il problema che vogliamo affrontare.

Ecco, più o meno con questo simile, concludo.

## STEFANO CUPPI

A questo punto il programma prevede un salto di lato perché agli scienziati ed agli ingegneri comunque in ambito di servizio pubblico deve essere in qualche modo codificato in documenti tipicamente sono la concessione ed i contratti di servizio ora la l'attuale concessione ed il contratto di servizio sono stati pensati e maturati alcuni anni fa quando questi argomenti ancora non erano così pressante anzi sostanzialmente erano diciamo ridotti erano intuiti dal fatto che a un certo punto è comparsa la parola multimediale che non piace sicuramente a nessuno di noi, ma nessuno trova di meglio, e magari sarebbe interessante capire come sostituire il termine, il servizio pubblico radio televisivo in Italia è diventato servizio pubblico radio televisivo multimediale a partire dalla concessione, dall'ultima concessione decennale. Darei la parola al professor Bassan che su questo argomento ha lavorato moltissimo e ci può raccontare a che punto siamo su questi aspetti.

## FABIO BASSAN

Sono ordinario a Roma di materie giuridiche e mi sono occupato sempre di temi come finanza, regolazione mercati, un po' di comunicazione.

- Concessione di servizio pubblico RAI e multimedialità

Vorrei tornare a bomba per introdurre i temi giuridici e molto sinteticamente, su quello che dicevi tu Stefano all'inizio. Quando Tagliavia diceva ah ma Raiplay, il servizio pubblico c'è. È certo che il servizio pubblico non ha imparato bene perché quando abbiamo parlato di quando insomma 3 anni fa si è discussa la concessione e poi il contratto di servizio all'epoca si parlava di dati, non di algoritmi di raccomandazione;

Perché in tre anni ovviamente cambia il mondo, però c'era un piccolo nucleo di interlocutori che ha avvertito rispetto ad una generalità non solo nel mondo politico ma anche nel mercato, che certi temi non li gestiva perché il meccanismo era quello inerziale che conosciamo tutti bene nel settore delle comunicazioni.

Finché non c'è un cambiamento si rimane com'è e poi quando arriva l'onda anomala speriamo non faccia troppo male. Questo è il motivo per cui non è stato consentito di inserire in questi testi giuridici qualche elemento che poi adesso crea delle opportunità, poi saranno tolte o meno. Insomma vedremo perché il termine multimediale quella è stata una battaglia vera, ricorderete nessuno sapeva cosa

volesse significare però tutti avevano con contezza che una qualche riforma in quella dinamica andasse fatta.

Se voi scorrete la concessione del contratto di servizio, non lo faccio ma vi consiglio di farlo, perché trovate ogni tanto dei suggerimenti e così delle raccomandazioni che rendono la concessione secondo me il contratto molto attuali: dalla promozione all'innovazione tecnologica, l'educazione digitale la trasmissione su tutti quanti le piattaforme quindi l'obbligo per RAI di trasmettere i contenuti o la maggior parte dei contenuti fino al 90 per cento su tutte quante le piattaforme, vi ritrovate una serie di elementi molto all'avanguardia ma mai come una norma che prevede che RAI deve rispettare il divieto assoluto di utilizzare metodologie e tecnologie capaci di manipolare in maniera non riconoscibile allo spettatore il contenuto delle informazioni. Questa norma vuol dire no algoritmi di raccomandazione nell'informazione.

Adesso arrivo anche al tema della trasparenza, te lo rileggo, rispetto del divieto assoluto di utilizzare metodologie e tecnologie capaci di mani manipolare in maniera non riconoscibile lo spettatore del contenuto delle informazioni. Qual è il tema: il tema era quello che ha si discuteva in quel periodo che poi la Commissione europea ha risolto con la sanzione miliardaria comminata a Google: cioè della manipolazione dell'informazione e quindi gli algoritmi di raccomandazione che venivano utilizzati da Google e in base ai quali - sulla base del principio che come avete detto- nessuno va oltre la seconda pagina delle liste che vengono suggerite e le prime le liste delle prime pagine di Google erano chiaramente indirizzate verso imprese che o erano inserzionisti e comunque con accordi con Google.

Di fatto la rappresentazione di Google della realtà è una cosa molto diversa dalla realtà e questa è il problema della Commissione quando ha sanzionato. È stato che da un lato le sanzioni nel diritto antitrust hanno il limite di essere tardive. Quindi vanno su un mercato che ormai è passato, perché l'intervento era appunto su un'attività di tre anni prima. E quindi era scarsamente incisivo sull'evoluzione del settore, e dall'altro lato e che detto cosa non si doveva fare. Era molto difficile per la Commissione ma per chiunque altro che non fosse Google. indicare cosa si sarebbe dovuto fare perché evidentemente era un tema di asimmetria informativa. In questo caso è assoluta.

E non è neanche tanto il tema di trasparenza - perché come sappiamo la trasparenza del dato non vuol dir nulla se insieme al dato è legato poi l'utente - nel senso che per quanto riguarda la gestione algoritmi, che, siccome gli algoritmi si evolvono in continuazione, andare a incidere su un certo algoritmo è poco rilevante, così come andare a richiedere la trasparenza del dato.

Perché quel dato senza la capacità di gestirlo, senza la capacità di analizzarlo, senza la capacità di utilizzare algoritmi, che però si basano su quel volume su quella

quantità di utenti e per quello che possono essere così raffinati, non ha nessun significato certo.

Però i criteri della trasparenza dell'algoritmo è sempre una sua perfezione. Perché non si arriva fino a un certo livello di trasparenza fino a un livello che ha. Per l'operatore che usa l'algoritmo è poco rilevante nel senso che comunque gli consente di utilizzare l'algoritmo nel modo che ritiene e noi ci stiamo concentrando sul piano regolatorio e probabilmente si andrà per sempre in errore.

E quindi il tema anche della nuova Commissione andrà probabilmente in questa direzione, perché a mettere insieme concorrenza e innovazione digitale evidentemente vuol dire questo, e quindi vuol dire creare un nuovo sistema che consenta alla Commissione, in quanto autorità europea di concorrenza, di avere strumenti nuovi. Però se questo è il progetto è chiaramente perdente perché comunque abbiamo la concorrenza che guarda al passato e può guardare al futuro solamente in pochi casi, ad esempio quello delle concentrazioni, però quando è avvenuto questo nel recente passato la Commissione si è distratta perché insomma Facebook + Whatsapp è chiaramente un errore e quindi il problema è il limite che da un punto di vista giuridico noi abbiamo nel regolamentare.

Best practice piuttosto che benchmark

Qual è la soluzione che la concessione per quanto riguarda il servizio pubblico ma anche poi la regolazione su questi temi sta cercando di individuare? Non c'è una soluzione ma ci sono due linee direttrici principali.

La prima - che è indicata da questa e da altre norme secondo me nella concessione - è la seguente: noi non siamo in grado di dire a Google, Facebook, Amazon, Netflix o quanti altri cosa debbono fare perché è il mercato bellezza e quindi entro certi limiti sono loro che decidono l'organizzazione interna in modo equo per un mercato.

- Best practice piuttosto che benchmark

Quello che possiamo fare però noi è creare dei benchmark a cui poi il mercato automaticamente tende. Perché dico questo? Perché la creazione del benchmark - che è uno strumento tipico regolatorio - che consenta di rendere la regolazione attuale perché non deve aspettare di completare l'iter normativo a livello nazionale europeo per essere applicata. Quello che avviene in genere sui mercati è questo: si crea un benchmark sul mercato, le autorità di regolazione trasformano il benchmark in una "best practice", la trasformano e la portano alla Commissione europea in quanto consulente.

La Commissione o adotta un atto esecutivo laddove possibile, oppure una proposta normativa e la norma torna nel nostro ordinamento. Il vantaggio di questi strumenti, che a volte sono anche un po' non dico casuali ma insomma, messi in questo modo nelle norme è che la norma è vincolante, è auto vincolante da subito e non quando

ha completato il circolo. Soprattutto dà alle autorità la possibilità di intervenire sul mercato sulla base dei benchmark, prima tra tutte l'Autorità per le comunicazioni, perché tra tutte quante è l'autorità che unica può individuare mercati nuovi rispetto a quelli già esistenti e su quei mercati poi individuare le posizioni dominanti gli abusi, ecc.

- Ma anche il mercato

Certo il perimetro è ancora quello della comunicazione elettronica e quindi bisogna adottare atti che abbiano un po' di coraggio.

Evidentemente la potenzialità che l'altra linea direttrice - e qui ve lo dico proprio in estrema sintesi - è quella di mercato, cioè creare di fronte a queste dinamiche non tanto dei benchmark quanto dei soggetti che possono competere con questi. L'esempio che vi faccio è quello che si sta discutendo adesso a livello europeo, perché essendo attuale lo vediamo tutti, e cioè quello della gestione dei dati che devono essere stoccati in Europa piuttosto che negli Stati Uniti. È un tema che sta appassionando molto la Germania e la Francia, ed è un tema però su cui anche noi potremmo dire la nostra, avendo una società - non la nuova società che trasporta le informazioni come può essere una rete di telecomunicazioni - ma una società che gestisce i dati. Questo è un progetto che si sta sviluppando anche in Italia e che potrebbe consentire di creare una soluzione alternativa.

- Il servizio pubblico come modello

Quindi in conclusione quello che voglio rappresentare è che per quanto riguarda il servizio pubblico era facile creare un modello a sé, che era un modello che andasse in una direzione giustificata dal fatto che appunto il servizio pubblico anche multimediale; al di là del servizio pubblico però questa strada può essere seguita da un punto di vista metodologico per creare una un'azione non tanto di contrasto ma un'azione che possa essere di sviluppo alternativo perché altrimenti la realtà che noi abbiamo non è quella che viviamo, e la rappresentazione che ci viene fornita sarà sempre più profilata.

## STEFANO CUPPI

Ringrazio il professor Bassan nel frattempo è arrivata la mia collega Componente del Co.Re.Com. avvocatessa Maria Giovanna Addario; aumenta il tasso di giuristi presenti. Do la parola a Erik Lambert per un intervento di rassegna sulle esperienze europee in atto

## ERIK LAMBERT

Sono stato direttore delle strategie tecnologiche del gruppo Canal+ - che includeva il studio Universal a Los Angeles - e dal 2003 lavoro come consulente.

In relazione alla materia che ci interessa oggi, sono stato direttore tecnico di HBO Nordic, un servizio di VOD, per suo avvio nel 2011-2012 e dal 2014 all'anno scorso sono stato consigliere prezzo StarzPlay Arabia, un servizio di VOD basato a Dubai. E sono compiaciuto del fatto che sul mercato MENA, StarzPlay è market leader davanti a Netflix (37% verso 25% di quote). Ovviamente per questi servizi di VOD, i motori di raccomandazione sono stati uno dei temi.

Il presidente Cuppi mi ha chiesto di fare una veloce rassegna delle attività dei media di servizio pubblico europei in materia.

È importante di ricordare che i servizi VOD del servizio pubblico nascono come servizi di "catch-up", a partire dall'iPlayer in 2007. Il catch-up è basato sul palinsesto, che svolge il ruolo di "recommendation system", ed è ovviamente basato su delle scelte editoriali.

Questa preminenza delle scelte editoriali è stata a lungo ed è ancora spesso una caratteristica dei servizi VOD dei MSP, anche dovuto ai limiti regolamentari e di disponibilità dei diritti a loro imposti. Ma ci sono state evoluzioni, con prima l'introduzione, oltre al palinsesto della settimana, di funzionalità di "search" su un archivio più largo, di proposte editoriali ("selected for you") sull'archivio e il catch-up, l'indicazione dei "last chance to see": tutto questi sono raccomandazioni generiche, le stesse per tutti. Qualche MSP ha introdotto raccomandazioni collaborative, ma generiche, con per esempio proposte basate su "hot on social media".

Anche qualche elemento di personalizzazione è emerso: "continue watching", contenuti simili a quello che è stato guardato - ma con un numero limitato di categorie -, o anche la possibilità di indicare/taggare contenuti preferiti.

Nel mondo commerciale, invece, la personalizzazione della raccomandazione è

andata avanti velocemente, basata su algoritmi e data mining, spinto tra altri fattori della competizione avviata da Netflix nel 2009. Come ricordato del professore Cremonesi, questi sistemi si prevalgono principalmente di due modalità per fare loro proposte all'utente: un modo "collaborativo" basato sulla similitudine con altri utenti, e un modo "contenuto" basato sulla similitudine dei contenuti.

Netflix, YouTube e gli altri sono testimonial dell'efficacia di questi nuovi algoritmi: il numero di click aumenta, come il tempo passato sui loro siti, elementi fondamentali per la vendita di pubblicità e/o la lotta contro il "churn" per l'abbonamento. Ma gli stessi algoritmi hanno anche effetti perversi, come la creazione di "filter bubbles" e di "echo chambers", e richiedono la tracciabilità dell'utente.

Usare gli stessi algoritmi genera problemi specifici per i media di servizio pubblico: per seguire l'utente e raccogliere i dati necessari alla parte collaborativa, è necessario la loro registrazione, e spesso anche ampliare la conoscenza dell'utente legando questa registrazione all'attività sui social, ecc. - che pone un problema sul tipo di rapporto che il servizio pubblico vuole avere con suo utente ; per la parte "contenuti" è necessaria un "tagging" una etichettatura avanzata - Netflix ha una squadra di 30 persone per questo...

Infine, questi algoritmi hanno una finalità strettamente commerciale (vendere pubblicità, mantenere e sviluppare il numero di abbonati), mentre il servizio pubblico ha obiettivi appunto di servizio pubblico come la coesione sociale, la diversità e il pluralismo. Questo significa che le proposte di questi algoritmi devono essere modificati con dei "controls" (al senso informatico) per renderle compatibile con questi obiettivi.

Il sistema Peach sviluppato all'interno dell'EBU fa parte di questa generazione di motori di raccomandazione. NB: Peach non è un prodotto, ma una base per sviluppo/adequamento in house.

La tabella seguente presenta lo stato dell'arte per i MSP europei

PSM	Curation	Content	Collabor.	Controls	Fornitore
ARD Mediathek	x	x	x	Diversity	In house
BBC iPlayer	x	x	x	Serendipity	Vendor

DR TV	x				Vendor
NPO Start	x	x			In house
NRK TV	x	x	social		In house
RAI Play	x	x	x	Diversity, EU	Vendor
RTBF Auvio		x	x		In house
SVT Play	x	x			In house

Gli sviluppi recenti di algoritmi basati sull'intelligenza artificiale sta spingendo una nuova generazione di sistemi di raccomandazione, sempre basati sull'uso di big data, ma con più flessibilità e più granularità del semplice "data mining". Questi sistemi sono "objective driven" e dipendono della definizione delle funzione da minimizzare, i.e. della scelta di KPI e di metrics da usare.

I servizi commerciali oltre all'"use time" e all'"engagement" sviluppano degli "proxies" (quantità misurabile) per ottenere un miglioramento della redditività, nuovi abbonati, una efficienza maggiore per i investimenti nei programmi ("cost-efficiency").

Per i media di servizio pubblico, questa evoluzione potrebbe permettere di mettere avanti obiettivi specifici, come :

- Soddisfazione a medio/lungo termine, piuttosto che il click immediato
- Neutralità, diversità, pluralismo (varietà degli punto di vista, varietà delle fonte, ecc.)
- Coesione sociale

Tutto questo dipende comunque dell'accumulo e dell'uso di molti dati. E dunque del rapporto che i MSP possono/vogliono/debbono trattenere con loro utenti, non solo ascoltatori o clienti, e i dati personali che li rappresentano.

Quel rapporto deve essere congruo con i valori dei media di servizio anche con le necessità pratiche. Un giusto equilibrio tra:

- Ascolto, ma con carattere distintivo
- Efficienza nell'uso dei fondi pubblici
- Diversità, inclusione e pluralismo
- Trasparenza
- "Agency" dell'utente
- Privacy
- Indipendenza
- Problemi specifici legati all'informazione.

Davanti a questa evoluzione e all'importanza dell'uso dei dati che i MSP non possono trascurare, certi studiosi (cf. Sørensen, Hutchinson 2018, "Algorithms and Public Service Media" in Gregory Ferrell Lowe, Hilde Van den Bulck, Karen Donders (eds.) *Public Service Media in the Networked Society*. Göteborg: Nordicom.) hanno suggerito di ripensare la missione dei MSP anche come un servizio pubblico di curatela dei dati personali.

## FRANCESCO SILIATO

Tra il 2007 e il 2014 il consumo di televisione lineare da televisore cresce anno dopo anno. Si passa dal 15,95% di rating del 2007 al 18,08% del 2014.

Il 2019 si chiude con una quota sulla popolazione di consumatori di televisione da televisore inferiore al 17 per cento. Lo scenario elaborato dallo Studio Frasi indica un'ulteriore diminuzione del consumo di televisione lineare da televisore nei prossimi anni, stimando che comunque gli ascolti nel giorno medio non scendano sotto i nove milioni. Del resto gli ascolti sono stati superiori ai dieci milioni soltanto tra il 2011 e il 2016 e nel 2018. Tra il 2011 e il 2019 la popolazione italiana è cresciuta di 2,1 milioni di individui e il peso degli over 65 è salito dal 21% al 25,5 per cento.

Elaborando le stime anche sull'andamento della popolazione e il suo invecchiamento, basate sui dati ISTAT, si prevede che il rating non scenda sotto il 16% nel giorno medio fino al 2025.

Tra i prossimi anni due, il 2020 e il 2022, vedranno crescere di circa mezzo punto l'ascolto complessivo. Il 2020 per via degli Europei di Calcio con la nazionale italiana qualificata e per le Olimpiadi di Tokio.

I giochi olimpici, che si disputeranno tra il 22 luglio e il 9 agosto, saranno in buona parte penalizzate dalle sette ore di differenza del fuso orario. La maggior parte delle gare avverrà nottetempo. Per via della differenza di orario le Olimpiadi saranno un'ulteriore importante trampolino di lancio per i collegamenti on line. L'utilizzo delle applicazioni coinvolgerà anche parte di pubblici non ancora abituati a ricorrere ad Internet per seguire i programmi preferiti. Discovery ha i diritti di trasmissione, Eurosport, parte del gruppo, trasmetterà tutto sulla propria piattaforma *Eurosport Player*, Auditel digitale misura già il consumo televisivo di url e app del gruppo Discovery e si potrà vedere quanto le Olimpiadi notturne influenzeranno la crescita diurna e serale del consumo di televisione via Internet. Sta al servizio pubblico proseguire su Rai Play la sperimentazione del 4k, approfittando degli eventi sportivi di rilievo.

### LE ETA' DELLA TELEVISIONE

La popolazione invecchia e questo potrebbe far pensare ad un sempre maggior consumo di televisione. Sarà così, ma solo in parte, vanno considerati due fattori rilevanti.

Gli anziani di domani sono i giovani di oggi, con i loro consumi di intrattenimento e le abitudini di impiego del tempo libero, ma non riusciranno a trasportarli identici nel futuro. Avranno mogli, figli, un lavoro, una carriera da affrontare e problemi da

risolvere. Sarà difficile trovare il tempo necessario per un *Binge watching*, ma il ricorso al consumo di televisione rimarrà. Non sarà solo televisione lineare, via via che il tempo passa sarà comunque sempre più da televisore piuttosto che da *device*. Il divano, la vista e l'udito spingeranno a mettersi comodi seduti su un bel divano davanti ad un grande televisore connesso ad internet. E, come sostiene l'ISTAT, la popolazione italiana sarà ancora più anziana di quanto non lo sia già.

Tra il 2004 e il 2019 il peso sulla popolazione italiana di chi ha tra i 25 e i 34 anni è sceso di cinque punti percentuali, dal 16,4% all'11,4 per cento. Quello degli over 65 è salito dal 18,95% al 23,15%

Nelle fasce d'età intermedie, tra i 35 e i 55 anni, l'uso di Smartphone, Tablet e PC crescerà. Della Smart Tv si sa, nel 2022 gli apparecchi televisivi saranno, per amore o per forza, tutti Smart, connessi ad internet e anche in wi-fi. Già oggi 37 milioni di individui risultano all'indagine di base Auditel realizzata dalla IPSOS vivere in una abitazione dotata di broadband. La Smart tv è galeotta invita al tradimento della televisione lineare, con le sue app sulla schermata principale vere tentazioni di fuga dalla routine della linearità.

- Il passare del consumo lineare

Il secondo fattore da considerare riguarda gli over 65 di oggi. Si tratta di individui molto diversi da chi quell'età l'aveva anche solo dieci anni fa, e le loro modalità di consumo sono molto diverse dagli over di un tempo. Le nostre stime indicano comunque per gli over 65 un incremento, ma non così elevato come ci si aspetterebbe. È invece accentuata la caduta del consumo di tv lineare da parte della fascia d'età 15-19 anni, (-4,5 punti percentuali), questo segmento tra il 2005 e il 2025 decreterà di 8,5 punti percentuali la presenza davanti ad un televisore per seguire l'offerta lineare dei broadcaster. Ma già tra i 25 ed i 34 anni la differenza si farà sentire e sarà inferiore di un paio di punti percentuali rispetto al 2019. Una esistenza poco lineare porta ad un consumo poco lineare, anche di televisione

E' noto che il "popolo della televisione" è più femminile, anziano e meno istruito della popolazione italiana nel suo complesso. Questi pesi continueranno a differenziarsi, ma una parte degli individui perduti dalla tv lineare saranno recuperati dalle offerte televisive operanti sulla rete.

Per quanto riguarda le generaliste RAI la loro discesa sarà inferiore a quella delle altre reti generaliste. A favore della RAI pesa, tra l'altro, la produzione di pubblici mediamente più avanti con gli anni delle altre reti.

I neocanali nativi digitali non riusciranno a coprire le perdite delle reti generaliste, che ammontano a centinaia di migliaia di individui l'anno, mentre i canali nativi digitali ne guadagnano alcune decine di migliaia.

In sintesi, dvbt-2 o no che sia, il tracciato dell'ascolto di televisione lineare è in discesa. Una discesa più o meno accentuata a seconda delle età dei pubblici e della popolazione italiana.

- Ma crescita del consumo di televisione

Tuttavia il consumo di televisione crescerà, crescerà nel frammento, nello spezzone popolare, nel *device* più comodo utilizzato nel momento più comodo. Come certificato dalla ricerca di base Auditel e dal Censis, in Italia il numero degli smartphone ha superato quello dei televisori, e l'uso di smartphone per seguire video, anche provenienti da broadcaster, farà comunque crescere il consumo di televisione, seppure nelle nuove forme di consumo e nelle nuove dinamiche. E comunque si tratta di un consumo ancora maggiore di quanto riesca a misurare oggi Auditel, fino a quando You Tube e gli altri social dediti ai video non verranno misurati.

Gli algoritmi saranno essenziali, sia quelli di raccomandazione diretta negli stessi siti e app dei broadcaster, che quelli pilotati dai social. Social eterodiretti dagli algoritmi o per appartenenza a tribù o per inserzioni commerciali che gli stessi editori pianificheranno per spostare l'attenzione sui loro programmi indirizzati ad un determinato target, ormai divenuto micro. Le lezioni della pubblicità non si dimenticano certo con l'arrivo degli algoritmi, anzi si perfezionano. Analisi ed elaborazioni dello Studio Frasi mostrano del resto come già oggi i pubblici siano divisi da politiche e culture divisive. I pubblici delle reti generaliste Mediaset sono stanziali su quelle reti, vivono in quel territorio e solo eccezionalmente si spostano sul servizio pubblico o su altre reti. Giusto per qualche importante partita della nazionale, il Festival di Sanremo, ma già il commissario Montalbano non li vede protagonisti come il resto della popolazione.

- L'ALGORITMO DELLA SEPARAZIONE

L'algoritmo di raccomandazione è un algoritmo di separazione. Funzionale al microtargeting, alla logica commerciale della separazione di individui in target.

Il ruolo del servizio pubblico dovrà differenziarsi. Non microtarget ma popolazione, non separazione ma coesione sociale. Algoritmi al servizio della crescita della società, dei singoli individui che la compongono.

Probabilmente non c'è mai stato, comunque non di recente, un periodo in cui il servizio pubblico sia stato altrettanto indispensabile, necessario alla crescita collettiva, alla consapevolezza di ciascuno su modi e tempi del vivere. Il compito di riunire ciò che i social dividono non può che essere svolto dai servizi pubblici dell'informazione, dell'intrattenimento e della cultura. Forse oggi sotto attacco proprio per il ruolo potenziale che potrebbero svolgere. Senza servizio pubblico la separazione degli individui, delle donne, degli uomini e degli stessi bambini in tribù commercialmente e culturalmente omogenei non incontrerebbe ostacoli, e non è detto che li incontri. Non è infatti semplice per i servizi pubblici europei perseguire obiettivi controtendenza rispetto alla cultura e agli interessi commerciali e politici della separazione a tutti i costi. D'altra parte *divide et impera* non è esattamente una novità! Si avrà la forza e la temperanza per passare dall'algorithm della separazione a quello della coesione?

#### FRAMMENTI DI TELEVISIONE

La digitalizzazione che consente l'offerta via internet della programmazione televisiva, induce a scomporla, a frammentare, anche in questo caso a "separare", un programma in tanti spezzoni. I dati Auditel dedicati alla misurazione del consumo di tv attraverso siti e applicazioni dimostrano quanto sia comune per i broadcaster fare ricorso ai frammenti nell'offrire i propri prodotti. La frammentazione dell'offerta è affine alla frammentazione, alla separazione, dei pubblici, ma non è necessariamente coincidente. Se però un intervento, una frase, un'immagine, persino una battuta di spirito vengono decontestualizzati, allora la manipolazione diventa un rischio concreto.

La questione è complessa e va affrontata con prudenza, mentre la frammentazione della società in target e microtarget è atto funzionale alla separazione, la divisione di un programma non sempre lo è.

C'è anche da dire che la cultura del frammento non è la cultura di tutti, ma l'abitudine cui sono esposti i più giovani li porta a considerarla quella "giusta". Salvo poi dedicare ore ad un videogioco.

#### ALGORITMI DI TRASPARENZA

La propensione di un servizio pubblico deve essere la trasparenza, anche nel progettare gli algoritmi di raccomandazione. Il GDPR, le stringenti norme sulla privacy, penalizzano i broadcaster europei e i servizi pubblici, mentre gli altri fornitori di contenuti audiovisivi possono raccogliere più informazioni e dati e incrociarli tra loro. Gli algoritmi di Google-You Tube hanno una enorme potenza di individuazione di consumi e sentiment di ciascuno e un altrettanto enorme opacità sull'uso e l'abuso di questi dati.

## ANDREA MONTANARI

Dove vogliamo cominciare.?

Stefano mi ha imposto con la forza di prendere il suo posto immeritadamente e l'unico rilievo che devo fare, che abbia dovuto tormentare me e l'Ufficio studi: non è vero, abbiamo accettato molto volentieri, anzi per dirla tutta non abbiamo neanche dovuto accettare – Stefano - perché è una discussione che abbiamo fatto crescere insieme e che ci vede come Ufficio studi della RAI estremamente interessati, partecipi, curiosi di capire quale può essere un percorso che conduce diciamo alla soluzione.

Non credo sia possibile insomma imboccare la strada giusta di fronte alla sfida comedice il felice titolo della giornata di oggi e di salvaguardare gli obiettivi di servizio pubblico multimediale tenendo conto in primo luogo appunto di diversità e algoritmi di raccomandazione. È ovvio che io non entrerò nel merito perché non ho una competenza specifica.

Qui c'è con noi Gino Alberico dal CRIT che parlerà con ben altra competenza e anche Gianluca Visalli che è di RAI Digital che parlerà con altrettanta competenza. La nostra presenza qui credo voglia dire che è possibile compiere almeno una delle tappe importanti di un percorso di questo genere. È quella della trasparenza, quella dell'apertura al dialogo e quella del tentativo comune per trovare una strada - provo a dire una cosa che non è tecnicamente legata alla esplicitazione di questo piuttosto che quel algoritmo, questa o quella soluzione al problema - ma è in qualche modo una cosa di "politica", cioè quello che è il nostro atteggiamento di cui bisogna tenere conto.

- Salvaguardare i valori distintivi del servizio pubblico

Anche stamane è uscita a un certo momento della nostra discussione quando abbiamo parlato di Auditel e della situazione odierna, io credo che dobbiamo lavorare molto seriamente ad affrontare il problema della salvaguardia del nucleo di valori distintivi di un servizio pubblico nella transizione verso l'era digitale che vede strumenti come gli algoritmi di intelligenza artificiale al centro. Lo dobbiamo fare con profonda convinzione e la contingenza storica ci consente di farlo proprio mentre nasce o è in via di consolidamento il nuovo sviluppo tecnologico, ma dobbiamo evitare a mio giudizio di ammazzare il bambino in fasce. Dobbiamo stare molto attenti a evitare che adesso, la dico con un termine che non piacerà a qualcuno ma insomma voi capite che cosa intendo dire, un'"ansia regolatoria" sacrosanta in astratto ci porti a restringere il campo di azione, restringere eccessivamente il campo di azione dei vari strumenti che consentiranno alla RAI una presenza digitale

importante.

- Universalità del messaggio

Il primo requisito al di là del segmento della dimensione digitale, il primo requisito di un servizio pubblico insieme alla sua capacità di creare coesione sociale, inclusione, proporre diversità e attraverso la diversità stimolare capacità critica, ma il primo requisito insieme a tutto questo è l'universalità del messaggio del servizio.

Non è dato un servizio pubblico nel contesto europeo che assomiglia alla PBS americana, che è un luogo straordinario per chi come me ha fatto il giornalista tutta la vita. Devo dire che PBS è l'Eden, è una meraviglia, quindi massimo rispetto e umile deferenza. Però non si può in nessun modo dire che sia confacente al criterio dell'universalità, perché è vista dall'1,2% dei telespettatori americani più qualche straniero tra le quali ci sono io, immagino anche tu. Quindi adesso noi abbiamo - e lo vedete anche in queste settimane - un tentativo della RAI di potenziare il segmento digitale con Raiplay. Credo che questo tentativo si debba dispiegare con sempre maggiore coerenza, con maggiore finanziamento, con maggiore capacità, nel prossimo futuro. Credo che dobbiamo fare crescere insieme a questo una consapevolezza degli obblighi di servizio pubblico. Credo che dobbiamo evitare appunto un eccesso regolatorio, perché sennò faremmo un'operazione come quella dei Padri Pellegrini che partono per andare col Mayflower va be'. Loro hanno fatto l'America, e nessuno potrà dire che è una realtà piccola però ecco non servirebbe a nulla la purezza del sangue ristretta a un francobollo.

- Portare il massimo di servizio pubblico

Dobbiamo portare il massimo di trasparenza, il massimo di servizio pubblico nel massimo di diffusione dei nostri strumenti digitali. Credo in questo modo di dire questo, e penso ciò che noi dobbiamo fare di grande interesse per questo lavoro grandissimo. Sarà oggetto anche di una nostra assunzione di responsabilità. In prima persona Stefano. abbiamo detto di provare nei primi mesi del prossimo anno a convocare a Roma un'ulteriore sessione di lavoro con la natura di questo allargata, visto che saremo noi a organizzarla, a una rappresentanza un pochino più nutrita di operatori industriali. Quindi insieme ai colleghi degli altri servizi pubblici anche. Vediamo se può essere opportuno anche di altre realtà, non di servizio pubblico, perché trasparenza e dialogo sono sicuramente la base di questo ragionamento.

Adesso non vi è ammorbo più e credo debba intervenire Piero de Chiara

## PIERO DE CHIARA

Il tema che avete scelto, gli algoritmi di raccomandazione del servizio pubblico, può essere l'inizio di un percorso che apre una finestra su un mondo molto ampio.

L'industria della comunicazione è sempre più fondata su dati, algoritmi e sulla ricerca scientifica applicata al loro utilizzo. L'indirizzo della ricerca applicata dipende dalla funzione obiettivo capire e quindi occorre chiedersi in quale misura gli obiettivi degli algoritmi di un'impresa di servizio pubblico debbano essere diversi da quelli delle imprese commerciali.

- Per i privati, valorizzazione del capitale

Per le imprese private della comunicazione (in particolare per le principali piattaforme) oggi la funzione obiettivo principale è la valorizzazione del capitale, a sua volta funzione non tanto del profitto, quanto del valore delle azioni. Negli altri settori i capitali finanziari cercano rendimenti a breve; nella comunicazione, capitali molto più ingenti inseguono mercati futuri, investendo sulle imprese che dichiarano e dimostrano di avere più dati e di saperli elaborare meglio. Gli abbonati e i contatti non sono solo ricavi, ma produttori di nuovi dati e nuova capacità attrattiva di capitali.

Non entro nella discussione se si più efficace l'algoritmo di Netflix o il magazzino di Disney o il modello TikTok basato su come mantenere l'*engagement* con segmenti di 15, 30 secondi. In modi molto diversi tutti stanno utilizzando in gran parte dei sistemi non supervisionati, in cui, data la funzione obiettivo, le macchine che cercano le correlazioni significative senza intervento umano.

- Per il pubblico, obiettivi diversi

La differenza tra algoritmi privati e pubblici non è che i secondi debbano sempre essere supervisionati -semmai devono essere in maggior misura spiegabili- ma è che si pongono obiettivi diversi.

Gli obiettivi di servizio pubblico non sono il profitto, né la valorizzazione del capitale. I KPI (indicatori di performance) principali non sono i ricavi da abbonamento pubblicitario e neanche il tempo di visione dei telespettatori.

I servizi pubblici europei, compresa la Rai, sono quasi interamente finanziati con denaro dei contribuenti il cui obiettivo non può essere fare quello che i privati fanno.

Non basta che il servizio pubblico abbia obiettivi diversi dai privati, occorre anche questi obiettivi siano misurabili e collegabili ai finanziamenti pubblici, non solo per

una esigenza gestionale, ma anche perché solo così il servizio pubblico può essere un agente che influenza la ricerca e sviluppo nel campo decisivo della cosiddetta intelligenza artificiale.

- Obiettivi alternativi possibili

Tre esempi di obiettivi alternativi e misurabili. Si tratta solo di esempi perché la scelta degli obiettivi non è solo un mestiere di esperti, è un lavoro politico. Gli esempi sono quindi mie opinioni - da sottoporre a discussione o da sostituire con altri obiettivi definiti politicamente - su missioni di pubblica utilità che il privato non è in grado di garantire e rispetto alle quali rischia anzi di creare danni sociali.

1. Il primo obiettivo in ordine di importanza e anche di maturità nella analisi è **la coesione sociale**. Questa età dell'oro della comunicazione commerciale sta determinando un problema sociale; è programmaticamente fondata su una segregazione dei pubblici: i giovani non vedono gli stessi programmi del degli anziani; le persone istruite ignorano i programmi di cui si nutrono le persone non istruite; i poveri non vedono quello che vedono i ricchi. I media, che nella fase della scarsità analogica garantivano nel bene e nel male una narrativa comune e costruivano le nazioni, oggi svolgono la funzione inversa: disgregano le società nazionali e ricostruiscono segmenti globali che non si identificano in storie comuni, non si conoscono e non si parlano. Con la digitalizzazione di tutti i media e gli sviluppi degli algoritmi di profilazione e indirizzamento il processo accelera continuamente.

Questa frantumazione sociale, questi target o bolle come vengono oggi chiamate, sono sicuramente un vantaggio per la comunicazione commerciale, perché parlare solo a un gruppo ben profilato e i cui gusti sono precisi e prevedibili, aumenta i ricavi e costa meno sul piano produttivo. Il privato fa il suo lavoro e non si può chiedergliene un altro (se non una maggiore trasparenza sulle logiche che utilizza per selezionare i target e i messaggi).

Ma stiamo parlando di un settore in cui l'intervento pubblico è forte, circa un quarto dell'intero mercato, se si somma il finanziamento alla Rai in forma di canone e altro e gli incentivi alla produzione privata. Questa presenza dello stato (comune a tutti i paesi europei) è figlia legittima dell'epoca in cui si dovevano costruire o ri-costruire le nazioni, ciascuna intorno a racconti condivisi; una presenza pubblica ancora più importante, anche se più difficile, oggi che si tratta di controbilanciare una entropia sociale che ha solide basi industriali, finanziarie e scientifiche. Si tratta di ri-utilizzare l'intervento pubblico e la ricerca pubblica che già ci sono per costruire un contro agente, una interferenza, un sistema diverso che anziché favorire le bolle favorisca

la coesione sociale (o, se vogliamo nominarla da destra: la coesione nazionale). Non a caso il vigente contratto di servizio Stato-Rai contiene un impegno esplicito per la Rai a studiare, proporre e implementare una misurazione dell'indice della coesione sociale.

Rimando all'intervento del prof. Siliato la spiegazione di un modello di misura della coesione sociale (e nazionale), che combina dati campionari auditel con dati puntuali generati dalle reti. In rozza sintesi: è il contrario della funzione obiettivo su cui lavorano gli algoritmi di raccomandazione privati, che scovano i target. E' un modello che invece premia e i programmi che sono in grado di rivolgersi a persone diverse per età, reddito, livello di istruzione.

Queste misure sono già oggi possibili con dati auditel, che, per quanto campionari e sotto-utilizzati, sono in Italia forse la indagine statistica permanente più granulare e dinamica. Con l'incrocio con i dati puntuali e comportamentali disponibili in rete e con la crescente connessione alla rete dei televisori, è possibile andare molto oltre le categorie socio-demografiche.

La profilazione comportamentale è un terreno delicato, che necessita di garanzie sul lato privacy e trasparenza sul lato nudget (i.e. small nudge); ciò non significa che debba essere lasciata solo alla logica commerciale. Oggi Google e Facebook usano e spesso vendono questo tipo di profilazione, a fini commerciali spesso dannosi, ma talvolta utili. Spesso li condividono con i governi ai fini di sicurezza. Più raramente per altri fini di pubblico interesse.

Grazie alle tracce disseminate in rete, oggi sappiamo con approssimazione sempre migliore, quali persone leggono gli oroscopi e quali le notizie scientifiche, quali sono tendenzialmente simpatetici con l'immigrazione e quali ne sono soprattutto spaventati, chi è più turbato dai delinquenti che la fanno franca e chi più dagli innocenti in carcere e tanto tanto altro. Nella comunicazione commerciale conviene dare a ciascuno un racconto confermativo del suo sentimento, ciò che accentua la polarizzazione e inibisce il confronto di argomenti.

Il servizio pubblico deve quindi fare il contrario, cioè dare a persone diverse, uno materiale comune di discussione.

2. La seconda funzione obiettivo di un servizio pubblico della comunicazione è la **circolazione nel mondo di opere italiane** (ed europee). È una missione rilevante, perché il *soft power* delle nazioni si basa non solo sulla diplomazia, l'import/export, ma in misura crescente sulla circolazione dei diversi punti di vista e valori nazionali, incorporati nelle news e ancor più nelle opere artistiche, letterarie e soprattutto sceneggiate.

La definizione algoritmica di questa funzione obiettivo è più acerba rispetto a quella della coesione sociale.

Sulla base della attuale definizione normativa, confusa e datata, le opere italiane rappresentano meno del 2% del mercato mondiale in termini di fatturato, mentre non sono utilizzati, anche se reperibili, dati in tempi di esposizione. E' una quota in rapida diminuzione, che si è dimezzata in venti anni; l'Italia sta diventando invisibile, uno stereotipo del passato.

E' un mercato sempre più difficile. Nel mondo esistono oltre 30 canali all news in inglese, due terzi dei quali realizzati da servizi pubblici di paesi non anglofoni. Il costo medio delle serie tv destinate al mercato internazionale è salito a 5 milioni per ora prodotta (con sempre più frequenti punte sopra i 10 milioni). In una industria sempre più globale, la definizione normativa spesso comprende produttori multinazionali, con una filiale italiana utile quasi solo per ottenere contributi; altre volte ignora rilevanti apporti di creatività italiana a opere internazionali. Eppure sarebbe già oggi possibile tracciare ore e redditi da lavoro nazionale e non solo i fatturati ma anche i tempi di visione realizzati all'estero, paese per paese. Certo per mettere in campo queste metriche occorre un po' di ricerca ed elaborazione dati pubblica.

Solo sulla base di queste metriche è poi possibile definire obiettivi e incentivi (anche solo misurare meglio l'efficacia dei cospicui incentivi previsti dalla legge Cinema).

3. Il terzo obiettivo è il più discusso dai politici e dagli autori, ma anche il più acerbo e complicato, persino nella sua definizione: qualcuno lo chiama pluralismo (spesso calcolato come tempi di notizia e parola per soggetti politici), altri diversità (per generi), io preferisco il termine **varietà** sul lato del consumo e non solo su quello dell'offerta.

Comunque e ovunque è un obiettivo complicato da misurare; non impossibile, con la crescente e granulare massa dei dati a disposizione. Spagnoli e britannici stanno molto lavorando sulla varietà geografica: non solo quanti programmi BBC sono prodotti non a Londra o Madrid, ma quanto sono poi visti in tutto il regno. Anche in Italia sarebbe questo un tema interessante, misurabile e incentivabile.

Sul lato del consumo esistono vari modelli di varietà, che si possono misurare con algoritmi *content related*, o meglio con algoritmi collaborativi di tipo nuovo. Per solo accennare quello che può essere un filone di ricerca tutto di sviluppare, è verosimile che come sotto-prodotto degli algoritmi di coesione

sociale sia possibile misurare il tasso di varietà nella dieta dei telespettatori, usando come tag il passaggio degli spettatori da un cluster all'altro. Ma certo occorre molta lucidità e trasparenza di progettazione, ricerca, accesso ai dati, potenza di calcolo.

Prima di venire alle questioni fondamentali successive agli obiettivi e cioè i KPI, gli incentivi, i nudget e come si finanziano questi progetti di ricerca, è bene sottolineare due punti comuni a tutti gli obiettivi: il primo è che i nuovi obiettivi sostituiscono ma non eliminano anzi incorporano la metrica oggi prevalente, che è oggi quella degli ascolti; perché è evidente che senza una **massa critica di ascolti** non può esserci un impatto significativo su nessun obiettivo sociale. Con programmi di nicchia non si realizza né coesione sociale, né *soft power*, né varietà lato consumo. Il servizio pubblico resta quindi concorrente dei privati per la conquista degli ascolti, ma solo come KPI relativo a diverse funzioni obiettivo.

Il secondo punto comune ai tre obiettivi è che devono essere automaticamente collegati agli introiti da canone, così come i ricavi pubblicitari sono automaticamente collegati alle metriche commerciali. Nella fase sperimentale questo collegamento può essere parziale, ma comunque sufficiente determinare una torsione agli incentivi interni all'azienda; a tendere deve sostituire l'attuale modello di finanziamento pubblico, che peraltro non è più neanche fondato su una discutibile separazione contabile, ma è ormai esplicitamente una decisione arbitraria del Governo, con danni per l'indipendenza e la gestibilità dell'azienda.

Salto, per motivi di tempo e per la sua complessità e delicatezza, il tema del passaggio da obiettivi misurabili a azioni o nudget la cui efficacia sia prevedibile; anche questo tema ormai maturo, come dimostra questo stesso seminario sugli algoritmi di raccomandazione, ma certo moltiplicatore di costi di ricerca e potenza di calcolo applicabile. In conclusione non posso però sorvolare sul tema dei costi, per non regalare ai conservatori il loro alibi preferito: già, **chi paga?**

Con quali risorse umane e finanziarie è possibile avviare un lavoro che abbia un impatto significativo e determini una interferenza rispetto alla ricerca e sviluppo di sistemi oggi indirizzati verso obiettivi commerciali, dove lavorano decine di migliaia di ricercatori (si pensi a Google o Huawei) e si muovono capitali di decine di miliardi?

In questo contesto la Rai è certo una azienda piccola e vecchia. Anche se un turn over accelerato del personale (finanziabile con la cessione di RaiWay al concorrente El Tower, ormai anch'esso pubblico) consentisse di assumere

alcune migliaia di giovani skillati, si tratterebbe comunque di un'azienda troppo piccola per giocare questa partita. Ma se, dopo una discussione partecipata, siamo in grado di individuare obiettivi di pubblico interesse misurabili (quelli che ho accennato o altri) gran parte dei costi sono già sostenuti dal pubblico e le risorse possono essere date in comodato condizionato ai risultati.

Anche qui pochi esempi. Un fattore di costo decisivo di questa partita è la potenza di calcolo. Persino più importante dei dati è la possibilità di spendere molti soldi per fare miliardi di calcoli in pochi secondi. I consumi energetici destinati a varie forme di intelligenza artificiale assorbono oltre il 2% dei costi energetici complessivi e continuano a crescere. Questa voce di costo sta diventando la principale soglia di ingresso che i grandi privati usano per ridurre la concorrenza potenziale.

Uno dei pochi interventi significativi sul piano economico che la Commissione europea ha messo in campo nelle recenti decisioni sull'intelligenza artificiale è un investimento massiccio su 4 super calcolatori in Europa di cui uno a Bologna. I **super calcolatori finanziati con fondi pubblici europei**, non devono solo ridurre i costi di calcolo per le imprese europee, ma soprattutto devono essere a disposizione di ricerche e sviluppo di pubblico interesse: anche della RAI, se è in grado di formulare obiettivi calcolabili.

Una secondo esempio di pagamento pubblico in natura è **l'accesso ai dati**. Non ci si deve illudere che i dati posseduti dalla RAI o anche altri enti pubblici che pure sono interessanti di per sé, siano una massa critica sufficiente. I dati posseduti dai privati sono molto molto più rilevanti. Il fatto che tali dati prodotti dai clienti non siano utilizzati solo per migliorare il servizio offerto, ma siano soprattutto usati e venduti dalle piattaforme private anche per altri scopi, sulla base di contratti discutibili ai sensi del GDPR è una leva potente di regolazione e contenziosi legale. Secondo alcuni si deve e si può mettere in discussione il diritto proprietario dei dati, ma è bene sapere che sarà un percorso lungo e conflittuale.

Più rapido e agevole può essere un recente sviluppo degli algoritmi di interrogazione delle basi dati: i cosiddetti "algoritmi nomadi" che visitano le basi dati senza trasferirle o copiarle, ma per interrogarle esclusivamente ai fini di risposte a domande specifiche. Se un'autorità indipendente certifica che il sistema di interrogazione non può essere utilizzato per altri fini e che quella domanda ha un interesse pubblico superiore all'eventuale deprezzamento del valore del dato privato, l'accesso può essere consentito, per via collaborativa e se necessario per via legale.

Netflix o Amazon opporranno una strenua resistenza prima di consentire l'accesso ai dati comportamentali che usano per alimentare il loro decantato sistema di raccomandazione; ma dati che non usano e non vendono, che so? quanti cinesi o indiani vedono serie televisive italiane, da quali scene sono stati catturati o respinti, questi dati hanno scarso valore per lei e grande valore per l'industria audiovisiva italiana. In altri settori, quali la sanità o i trasporti, si possono identificare molte domande di pubblico interesse che sono rilevanti per fini di pubblico interesse e danneggiano poco o nulla i valori contabili di Google o Facebook.

Ancor più potenti sono gli strumenti pubblici utilizzabili nei confronti delle imprese private titolari di appalti, concessioni o autorizzazioni, ivi compresi tutti coloro che stanno disseminando il territorio di antenne, reti e sensori. Se solo pensiamo al 5G e alle sue applicazioni, il prossimo decennio sarà decisivo, perché moltiplicherà di due ordini di grandezza le basi dati oggi disponibili.

Certo questi apporti pubblici in natura, in forma di dati, potenza di calcolo, **ricerca applicata**, che si aggiungono a quelli in forma di canone, non possono essere incondizionati. Per riprendere un'espressione sviluppata bene dal prof. Bassan il prossimo contratto di servizio Stato-Rai dovrebbe avere in parte la forma di *smart contract* nel senso che noi contribuenti mettiamo dei soldi in una azienda pubblica in proporzione al raggiungimento di obiettivi misurabili. Anche se gli obiettivi devono essere diversi, in questo le imprese pubbliche e private si assomigliano e funzionano dando priorità a ciò che genera ricavi o altri valori misurabili. Lo *smart contract* può disciplinare non solo un finanziamento non più arbitrario, ma anche l'accesso alla potenza di calcolo, ai dati, alla ricerca pubblica sugli algoritmi applicati ai sistemi di comunicazione.

Si può obiettare diventare un servizio pubblico della comunicazione con questa ambizione è un percorso troppo difficile per la Rai. Qualcuno, a ragione, ha proposto un centro di ricerche e sviluppo a livello europeo sulla intelligenza artificiale, con ruolo analogo a quello che il Cern di Ginevra svolge nella fisica di base, ma esteso anche allo sviluppo.

Ma in attesa che nascano grandi investimenti pubblici europei oggi le imprese pubbliche nazionali possono avviare sperimentazioni importanti.

Eni, Enel, Leonardo, Poste sono certo più grandi della Rai, ma oltre ad essere condizionabili dagli azionisti privati, non hanno competenze sugli impatti

sociali della produzione e distribuzione dei contenuti.

Infine, ma non da ultimo, la Rai è, nel bene e nel male, una metafora dell'intervento pubblico, delle sue glorie passate, dell'involutione lottizzatoria e dello smarrimento attuale. Tre cicli durati ciascuno venticinque anni e di ciascuno dei quali la Rai ha rappresentato l'immagine di vizi e virtù più esposta alla pubblica opinione.

Ritrovare missioni distintive di medio-lungo periodo è un problema comune a tutte le imprese pubbliche. Sarebbe giusto e utile che la impresa Rai rappresentasse anche l'annuncio di questo nuovo ciclo dell'intervento pubblico.

## ANDREA MONTANARI

Se si facesse quello che tu dici non ci sarebbe questo dibattito demenziale per cui ad ogni elezione regionale o una tornata di nomine o un'elezione politica viene qualcuno che ti dice ti tolgo il 10 per cento.

Ma non sono capogruppo, anzi tramite i soldi se vuoi fare il servizio quindi la proposta che tu fai avrebbe l'evidente e positivo risultato di dare certezza.

Neanche dimenticare una cosa che abbiamo subito l'Italia ha un'autonomia perché poi se nella mia autonomia però esistono due condizioni che oggi se faccio una valutazione empirica di quello che accade in Italia. In Europa non possono vedere che a funziona così da nessuna parte : cioè nessuno dei grandi servizi pubblici europei prende fondi pubblici in cambio di una misurazione della sua compliance sugli obiettivi. È un dato di fatto che non accade, che poi sia auspicabile, che eventualmente accada, è un punto di vista giusto. Ma non esiste da nessuna parte.

Non dimentichiamo, perché è un aspetto che viene tralasciato, che la RAI percepisce oltre un miliardo e 700 milioni del canone, ho visto che questo miliardo deve essere 30 miliardi 8 e 50 del gettito teorico del canone bollette. Noi prendiamo - lo sapete tutti perché vi occupate di questa roba - una quantità infinitesima delle risorse che prendono i servizi pubblici per i quali noi ci compariamo spesso. Prendiamo poco più della metà dei francesi , prendiamo un terzo degli inglesi e un quarto dei tedeschi. Quindi il discorso delle risorse nel suo complesso andrebbe coniugato con l'estensione del credito produttivo - una delle strutture non c'entra niente con quello di cui parliamo oggi - a una delle più profonde strutture della RAI. Il fatto che noi produciamo troppo rispetto alle risorse che abbiamo è una follia totale, che è indubbio che se ne producesse un poco di meno potremmo puntare ad avere più

qualità anche nei termini di rispetto della riduzione della capacità trasmissiva del digitale terrestre.

Proseguiamo con la nostra lista quindi dopo Piero vedo Gianluca Visalli

## **GIANLUCA VISALLI**

Prima di addentrarmi nel tema del workshop credo sia utile presentare la mia esperienza: da poco più di 3 anni mi occupo all'interno della Direzione Digital Rai dello sviluppo del prodotto e dell'offerta di RaiPlay, in particolare negli ultimi due anni ho seguito l'introduzione del motore di recommendation sulla piattaforma.

Ed inoltre permettetemi prima di iniziare il mio intervento di dichiarare il mio senso di impreparazione rispetto al livello di questa discussione che, per fortuna, posso giustificare proprio perché l'obiettivo di oggi è quello di avviare un discorso sull'Algoritmo di Servizio Pubblico a partire dalla costruzione di una definizione condivisa.

In questo senso ritengo che possa essere utile partire dalla descrizione di quello che stiamo facendo nella Direzione Digital Rai a partire proprio dallo spunto lanciato nel 2018 da Gian Paolo Tagliavia, che è stato gentilmente ricordato in apertura dal nostro ospite, e che è stato ovviamente trasferito anche alla struttura che allora dirigeva e che, dalla fine del 2017, era impegnata ad introdurre su RaiPlay un sistema di recommendation che contribuisse all'individuazione dei contenuti da proporre agli utenti registrati. La piattaforma impiegata su RaiPlay è un prodotto di mercato, sviluppata da una società terza, che è stata integrata nei sistemi di gestione dei nostri contenuti.

Su questo fronte è utile specificare che le proposte del motore di recommendation sono rivolte agli utenti che abbiano accettato esplicitamente la personalizzazione dei contenuti al momento della registrazione: si tratta, per dare un riferimento, di circa il 75% dei quasi 13M di utenti registrati di RaiPlay. Esiste quindi un 25% degli utenti cui vengono presentate solo le selezioni editoriali definite dal team editoriale che cura la piattaforma, mentre al restante 75% vengono proposte selezioni editoriali integrate dalle proposte definite da un algoritmo automatico.

La quantità di queste proposte ha vissuto due fasi, legate allo sviluppo della piattaforma: nella fase iniziale, dal lancio del sistema fino alla fine del mese scorso, gli spazi affidati al motore di recommendation erano limitati a poche "fasce" sulla Home Page ed alla proposta di contenuti correlati a quello che gli utenti hanno scelto di guardare sulla "spalla" della pagina di fruizione. A partire dall'inizio di questo mese,

con la messa on line della nuova versione di RaiPlay, i contenuti selezionati dall'algoritmo sono presenti in quasi tutte le "fasce" presentate sulla Home Page anche se ancora non in maniera integrale.

Tornando allo spunto di Tagliavia, da cui siamo partiti, sono due le direzioni nelle quali ci siamo mossi per provare a costruire un "algoritmo di servizio pubblico": accrescere il numero e la rilevanza dei contenuti di produzione italiana tra quelli che il motore di recommendation identifica come potenzialmente graditi al singolo utente e rendere trasparente, come peraltro abbiamo sempre cercato di fare in tutta la nostra attività editoriale anche attraverso la titolazione delle nostre proposte editoriali, le regole ed i criteri che guidano il funzionamento dell'algoritmo per quelle specifiche selezioni di contenuti.

Si tratta a mio parere di un aspetto quest'ultimo particolarmente importante, anche se è stata in questo momento sospesa per gestire la complessità della fase di transizione che stiamo affrontando, perché riteniamo fondamentale in un'ottica di trasparenza e responsabilità non solo comunicare chiaramente che i contenuti proposti sono definiti da un sistema automatico ma anche esplicitare i criteri ed il funzionamento in base al quale questa selezione viene compiuta. In particolare riteniamo molto importante da parte nostra aver introdotto questo elemento anche per il ruolo che la Rai deve svolgere anche per proporsi come benchmark sul mercato e cercare di affermare questa forma di trasparenza e responsabilità anche agli altri operatori ma anche stimolando l'adozione di una regolamentazione di questo tipo da parte dei soggetti istituzionali.

Per concludere credo sia importante aggiungere un altro elemento delle nostre politiche editoriali che pur essendo, per così dire, pre-algoritmo ha sicuramente un impatto sulla missione di servizio pubblico applicata ad una piattaforma over-the-top perché condiziona e modifica il funzionamento del sistema di selezione dei contenuti basato sui comportamenti di consumo. Come probabilmente avrete notato infatti la selezione dei contenuti sulla nostra Home Page comprende necessariamente tutte i 9 generi di contenuti nei quali è articolato il nostro catalogo a prescindere dai risultati in termini di consumi che questi generi sviluppano, inclusi quindi generi come: Musica e Teatro, Documentari e contenuti delle Teche Rai. Queste selezioni sono in parte gestite dall'algoritmo e quindi possono essere basate sui comportamenti di consumo pregressi, ma in ogni caso questo tipo di contenuti devono essere proposti. Si tratta evidentemente di una scelta estremamente qualificante il nostro ruolo che ci distingue profondamente da qualsiasi altro operatore del mercato.

## GINO ALBERICO

Due parole sulle iniziative internazionali relativamente ai sistemi di raccomandazione.

In ambito EBU (l'unione dei broadcaster pubblici europei) in particolare nell'area Technology & Innovation è stata avviata qualche anno fa un'iniziativa per lo studio e lo sviluppo di una piattaforma di raccomandazione per i broadcaster denominata PEACH, acronimo che sta per "Personalisation for **EACH**".

Tale iniziativa cerca di dare una risposta al tema algoritmi per "motori di raccomandazione" per i broadcaster, ossia cosa possiamo fare in quanto broadcaster?

Ovviamente gli obiettivi sono comuni a molti ambiti applicativi, ossia quelli di riuscire a dare:

- il giusto contenuto,
- quando serve,
- alla persona giusta
- ed anche, possibilmente, sul dispositivo giusto.

Nella mattinata si son sentiti diversi interventi su cui in buona parte mi ritrovo, e che in parte sono ripetuti nelle mie slide.

Sicuramente un tema che citato molto spesso in tutti gli interventi è quello del tempo che gli utenti passano sui nostri contenuti. Ora, il valore di un servizio pubblico si può misurare secondo diverse direzioni, ad esempio produrre contenuti vasti, ricchi, diversificati (e lo facciamo già), distribuire i contenuti sulle diverse piattaforme (e lo facciamo già). Quindi per aumentare il valore della nostra offerta si può lavorare sulla "rilevanza" dei contenuti offerti che può avere come conseguenza anche il "tempo trascorso" dagli utenti sui nostri contenuti. L'obiettivo però non è meramente quello di aumentare il tempo trascorso (questo può essere la tipica funzione obiettivo degli operatori commerciali) ma far sì che, se non voglio essere irrilevante, si possano anche massimizzare altre funzioni. E' chiaro che il tempo speso è comunque una delle almeno due dimensioni su cui concentrare gli miei sforzi (perché se i contenuti poi non li guarda nessuno non serve a molto produrli e distribuirli) però questi sforzi oltre ad essere orientati a attrarre gli utenti a passare del tempo sui miei contenuti, devono far sì che in qualunque momento un utente entri nella mia offerta sia indirizzato a qualche cosa che è di suo interesse: non posso permettermi di dire "magari questo utente poi torna domani...": devo interessarlo oggi e riuscire a fare in modo che ci sia qualcosa che scatti, qualcosa per cui trova interessante ciò che sto

proponendo oggi. Diversamente abbiamo perso un'occasione, in quanto, ovviamente, se un utente passa più tempo su contenuti degli altri competitor ci sono meno chance che aumenti la percezione di fiducia in quello che sto offrendo.

Parlando della piattaforma EBU-PEACH: in realtà è costituita da due parti: una sperimentale e una operativa. La piattaforma ha tutte le tecnologie che servono a raccogliere le informazioni dai browser e dalle App con il consumo degli utenti. Su queste informazioni è possibile provare ad elaborare dei modelli basandosi su un mestiere tipico che è quello del "Data Scientist", ossia a fronte dei dati di cui dispongo come organizzo gli esperimenti: ossia costruisco degli algoritmi, produco un output (contenuti suggeriti) che invio e presento agli utenti. Ovviamente, la funzione che non può mancare come si vede dalla chart, è quella di "monitoraggio" ossia poter definire grandezze che posso poi andare a misurare (da un panel di utenza ad esempio) per valutare il successo o meno dell'esperimento che è stato pianificato (se l'algoritmo oggetto del test ha funzionato bene o male rispetto alle attese).

Perché PEACH, se ci sono tanti recommendation engine o sistemi commerciali già pronti?

Perché non usare il sistema di raccomandazione di Amazon, che mi raccomanda che cosa comprare avendo io già comprato una maglietta, un paio di scarpe e un televisore, ecc. Forse uso algoritmi che probabilmente sono gli stessi, ma magari nel settore dei media può essere necessario qualcosa di diverso.

Intanto, magari non è nemmeno necessaria tutta quella complessità richiesta da una piattaforma di e-commerce "general purpose" ed invece, magari sono necessarie funzioni che in quel sistema non sono presenti. Ecco perché PEACH potrebbe essere un "algoritmo" che i broadcaster adattano alle proprie necessità specifiche del mondo dei contenuti, dei media.

Inoltre, il fatto di essere broadcaster pubblici magari richiede modalità di comportamento/funzionamento che differiscono da quelle di operatori "verticali" (come Netflix per intenderci) che rappresentano solo una frazione del mondo dei media.

I broadcaster pubblici possono avere contenuti tipicamente molto più diversificati: ad esempio come RAI abbiamo anche diverse tipologie di "brand": abbiamo i canali generalisti, i canali tematici, i canali per bambini/ragazzi quindi contenuti destinati a diversi target. Inoltre, sappiamo che il prodotto televisivo ha una sua storia, un ciclo di vita quindi nasce anche l'esigenza di voler realizzare una piattaforma con dati "in house" su cui mantenere il controllo ed evitare di trovarsi con una rappresentazione dei dati legata alla tecnologia di uno specifico "vendor".

L'aspetto di controllo su dati e algoritmi è importante anche per poter garantire ai miei utenti di trattare i dati con la dovuta attenzione. Quindi ad esempio non sarebbe accettabile tanto facilmente il fatto che la mia linea "editoriale" ossia ciò che vado a proporre all'utente sia decisa all'interno di una "black box" nella quale poi non conosco esattamente quali parametri andare ad aggiustare affinché cambi il suo comportamento (o anche solo di cui io possa sapere quale è il comportamento).

Gli algoritmi.

Visto che se ne è già parlato, non mi dilungo sugli algoritmi, ma mi limito a dire che all'interno della piattaforma PEACH si possono sperimentare un po' tutti gli algoritmi o combinazioni di essi, lanciando esperimenti su data set limitati e analizzando il risultato.

sta dicendo adesso che arrivano sono 4 5 tv pubbliche europee quello ovviamente è costruito per esser folk tollerante scalabile con grandi numeri e quindi è un servizio che in qualche modo poi va anche si paga perché si paga il consumo del l'utilizzo di apparecchiature che hanno un costo e dicevo Al momento sulla piattaforma sono disponibili 4 algoritmi di base (trending, collaborative filtering, content based, diversity) che sono poi quelli soliti, noti a tutti.

TRENDING: non ha alcun bisogno di dati di consumo dell'utente, semplicemente indica cosa è stato più visto nelle ultime ore (o giorni, o in un lasso di tempo definito): in realtà non è poi una vera personalizzazione, però è facile da realizzare.

COLLABORATIVE FILTERING: un qualche cosa che si basa sulla per similitudine, per cui se tu come altri utenti hai visto due contenuti A e B, e questi altri utenti hanno poi scelto il contenuto C allora forse anche tu potresti essere interessato al contenuto C. In questo approccio ci può essere un problema all'avvio: ossia quando all'inizio non ho ancora dati di quell'utente (l'utente che è appena arrivato sulla mia piattaforma) provo a suggerire dei contenuti "random"

CONTENT BASED: faccio una profilazione dei contenuti (occorre però una meta-datazione ben fatta) e propongo suggerimenti per similitudine, non basandomi sulla storia di altri utenti. Il risultato può apparire meno "random" per si rischia di confinare l'utente in una "bolla".

DIVERSITY: cerco di evitare le bolle mettendomi come regola il fatto di proporre qualche contenuto che devia un po' rispetto ai migliori "candidati" della lista di raccomandazioni (evito ad esempio di proporre solo gli episodi multipli di un certo show, oppure solo certi "generi", ecc.).

Si possono poi introdurre delle "BUSINESS RULES" ossia imporre dei vincoli all'output

ragionando appunto sul fatto che si decide per esempio di non raccomandare più di “x” puntate di un certo show. O viceversa, se un utente chiede comunque (al di là della raccomandazione iniziale) di vedere 10 puntate di un certo programma assumo che quel programma gli piace così tanto che gli propongo anche l’undicesima puntata.

Altre regole potrebbero tradursi in “blacklist”, ossia filtrare certi contenuti (ad esempio “meteo”) anche se usati nella fase di training dell’algoritmo), oppure abbassare il punteggio di programmi “troppo vecchi”.

Quale sia la miglior combinazione di algoritmi e di “business rules” non è facile da stabilire, dipende da quel che vogliamo ottenere, da quali sono i nostri obiettivi: l’unico modo per stabilirlo è sperimentare ed essere in grado di misurare i risultati degli esperimenti.

Ad esempio i colleghi della TV portoghesi usano la piattaforma PEACH solo sulla homepage: per il momento un approccio molto semplice. I tedeschi di BR invece lo utilizzano a tappeto su tutta l’offerta della loro Mediateca: però dispongono di diversi “data scientist” in grado di definire, realizzare ed analizzare diversi esperimenti.

Sicuramente, ciò che è molto importante è poter sperimentare su dati reali, con utenti reali. Questo però, inquadrato in un ambito più ampio che è quello della protezione dei dati, potrebbe ridurre la possibilità di scambiare con altre realtà i dati raccolti negli esperimenti per mettere a fattor comune i risultati.

## MARINA CAPORALE

Sintesi dell’intervento di Marina Caporale

### ALCUNE PREMESSE E DOMANDE INIZIALI

Natura del SPM: per quale pubblico? Per quale dispositivo? La differenza tra chi paga il canone e chi no è una scriminante esattamente negli stessi termini previsti per il SPRTV?

Modalità di accesso ai contenuti online e identificabilità dell’utente. Contenuti online e minori.

Massimizzazione della permanenza degli utenti come obiettivo degli algoritmi di raccomandazione, sì, ma su quali contenuti?

Informare, educare, intrattenere: il prof. Bassan ha richiamato il principio di lealtà e il divieto di manipolare lo spettatore e l'informazione, da cui si desume il divieto di utilizzare algoritmi di raccomandazione nell'informazione. Ma questa definizione, come noto, non è così semplice, ha sfumature sempre più articolate. Basti solo pensare alla progressiva espansione del cosiddetto infotainment, che pregiudica in modo lampante la distinzione tra informazione e intrattenimento. E la stessa cosa vale anche per le attività volte ad "educare" l'utente. L'algoritmo non può "informare" e nemmeno "educare"?

Algoritmi di raccomandazione in generale sono tarati sempre più sui singoli utenti, individualizzazione del servizio, microtargettizzazione e non sui concetti di pubblico vasto, coesione e pluralismo cui il servizio pubblico è volto.

Servizi di media audiovisivi e radiofonici, algoritmi di raccomandazione, SPM e responsabilità editoriale...

Per questi orizzonti quali garanti? Commissione Rai, Agcom, Antitrust, Privacy...

#### NATURA DELLA RAI E ACCESSO ALLE INFORMAZIONI (E ALLA STRUTTURA DELL'ALGORITMO)

RAI e "natura sostanziale di ente pubblico" (Cass. SS. UU. n. 27092 del 2009).

RAI e assoggettabilità alla normativa in materia di accesso documentale (l. 241/1990) e trasparenza (d.lgs 33/2013), in particolare è soggetta alla disciplina in materia di accesso civico e accesso civico generalizzato. Vedi ANAC "Nuove linee guida per l'attuazione della normativa in materia di prevenzione della corruzione e trasparenza da parte delle società e degli enti di diritto privato controllati e partecipati dalle pubbliche amministrazioni e degli enti pubblici economici" (ANAC del. 2017/113) che prevede l'assoggettabilità a tutti gli obblighi di trasparenza e anticorruzione e relativi limiti. **COMINCIA A FORMARSI UNA CERTA GIURISPRUDENZA AMMINISTRATIVA SUGLI ALGORITMI UTILIZZATI DALLE PUBBLICHE AMMINISTRAZIONI E SULLA LORO CONOSCIBILITA'**. Tali sentenze, da riferirsi a soggetti (pubbliche amministrazioni) e attività (decisioni amministrative), affermano alcuni principi che possono ritenersi applicabili anche ad eventuali algoritmi usati dalla società concessionaria di servizio pubblico, in ragione del servizio espletato e dei principi e obblighi di trasparenza cui la RAI è soggetta.

Dopo le sentt. Tar Lazio sez. III bis sent. 6606/2019 e 994/2018, nonché Cons. Stato 2270/2019, la più recente Cons. Stato, Sez. VI, sent. n. 8472/2019 individua **tre principi** per l'esame e l'uso degli strumenti informatici, derivanti anche dall'interpretazione del Reg. Ue 679/2016 in materia di trattamento dei dati personali (che ha modificato il d.lgs 196/2003, "Codice della privacy"):

il **principio di conoscibilità**, per cui ognuno ha diritto a conoscere l'esistenza di processi decisionali automatizzati che lo riguardino ed in questo caso a ricevere informazioni significative sulla logica utilizzata;

**principio di non esclusività della decisione algoritmica:** nel caso in cui una decisione automatizzata "produca effetti giuridici che riguardano o che incidano significativamente su una persona", questa ha diritto a che tale decisione non sia basata unicamente su tale processo automatizzato (art. 22 Reg.);

**non discriminazione algoritmica**, secondo cui è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione.

Assume quindi pieno rilievo l'algoritmo di raccomandazione quale trattamento automatizzato dei dati personali, dimensione fin qui lasciata sullo sfondo, per cui in linea di principio non è ammissibile una decisione basata esclusivamente su un trattamento automatizzato di dati ma solo se: 1) produce effetti giuridici; 2) incide in modo significativo sulla persona dell'utente; 3) la decisione è basata interamente (solely) sul trattamento automatizzato dei dati.

#### SPUNTI CONCLUSIVI

L'algoritmo di raccomandazione del SPM, finalizzato in modo più marcato a prolungare la permanenza dell'utente sulle pagine dedicate, perseguendo quindi in senso lato un fine più commerciale, non sembra rientrare nelle ipotesi di divieto di trattamento automatizzato previste dal Reg. Ue 679/2016. Poiché comunque assoggettata a una possibile richiesta di accesso, la Rai dovrebbe in caso dimostrare di potere rendere pienamente conoscibili i meccanismi e le logiche algoritmiche utilizzate e che queste siano coerenti con i principi generali del servizio pubblico e i vincoli derivanti dalla concessione<sup>1</sup> (vedi art. 1, c. 1 concessione) e dal contratto di servizio.

Quest'ultimo in particolare afferma che la Rai assicura un'offerta di servizio pubblico improntata ai seguenti principi: "...rendere disponibile e comprensibile – nella

---

<sup>1</sup> [...] "1. La concessione ha per oggetto il servizio pubblico radiofonico, televisivo e multimediale da intendersi come servizio di interesse generale, consistente nell'attività di produzione e diffusione su tutte le piattaforme distributive di contenuti audiovisivi e multimediali diretti, anche attraverso l'utilizzo delle nuove tecnologie, a garantire un'informazione completa e imparziale, nonché a favorire l'istruzione, la crescita civile, la facoltà di giudizio e di critica, il progresso e la coesione sociale promuovere la lingua italiana, la cultura, la creatività e l'educazione ambientale, salvaguardare l'identità nazionale e assicurare prestazioni di utilità sociale: Art. 3, c. 1, lett. l) la completa digitalizzazione, la conservazione e la promozione degli archivi storici radiofonici e televisivi, garantendo anche attraverso il web il più ampio accesso gratuito del pubblico agli stessi; ..."

molteplicità delle forme divulgative – su differenti piattaforme, una pluralità di contenuti, di diversi formati e generi, che rispettino i principi dell'imparzialità, dell'indipendenza e del pluralismo, riferito a tutte le diverse condizioni e opzioni sociali, culturali e politiche, **affinché ciascuno possa autonomamente formarsi opinioni e idee e partecipare in modo attivo e consapevole alla vita del Paese**, così da garantire l'apprendimento e lo sviluppo del senso critico, civile ed etico della collettività nazionale anche all'estero, nel rispetto del diritto e del dovere di cronaca, della verità dei fatti e del diritto ad essere informati..." (art. 2, c. 1, lett. a).

Non c'è nulla nelle specifiche riguardanti l'offerta multimediale che precisi la portata applicativa di questi principi (in particolare non l'art. 5 del Contratto di servizio, dedicato all'offerta multimediale, e nemmeno gli obblighi specifici declinati, sempre con riferimento all'offerta multimediale, dall'art. 25, c. 2, lett. c).

A parte la sovversione strutturale che la logica di servizio pubblico perseguirebbe con l'algoritmo, di un servizio tarato sulla singola persona e non sul pubblico generalmente inteso, il limite più solido su cui riflettere, in generale verso la prestazione di servizi non lineari e in particolare per quelli prestati attraverso algoritmi di raccomandazione, rimane quello della autonoma formazione di opinioni e idee da parte dell'utente, che ne sarebbe evidentemente intaccata.

## FLAVIA BARCA

Possono aggiungere qualche spunto utile alla discussione alcuni dati riguardanti il rapporto fra algoritmi e discriminazioni di genere, un campo sul quale la ricerca è solo agli inizi.

Il primo riguarda il numero di donne che definiscono gli algoritmi: il World Economic Forum nel 2018 stima che la quota di donne occupate nell'intelligenza artificiale non superi il 22 per cento. I criteri usati nell'istruzione degli algoritmi rispondono a un punto di vista maschile e questo ha molteplici conseguenze a cascata.

Nel 2014 Amazon ha dovuto abbandonare un sistema sperimentale di recruiting costruito sulla base di algoritmi perché si è accorto che questi ultimi preferivano candidati maschi. La ragione è che l'assenza dal lavoro per lunghi periodi, tipica del lavoro femminile spesso interrotto per la maternità, era identificato come una caratteristica negativa e quindi influiva negativamente sul giudizio finale.

Inoltre l'algoritmo, istruito sul fatto che le donne sono tradizionalmente escluse dai ruoli apicali – elemento che segna i loro curricula - tende inevitabilmente a riproporre quel "bias".

Raggiungere la parità, quindi, non può essere frutto di un esercizio matematico, di una cd "neutralità dell'algoritmo", perché tutto sta nelle modalità con cui l'algoritmo viene istruito, con cui vengono istruiti i metadati.

Il rischio dell'algoritmo è quello di nascondere una mancanza di presa di responsabilità, per esempio nelle scelte editoriali di un telegiornale.

Secondo spunto: due ricercatori del MIT hanno studiato i software di riconoscimento facciale di Microsoft e Facebook dimostrando come l'algoritmo funziona meglio per la tipologia "uomo bianco". Per gli uomini bianchi il riconoscimento ha un errore dell'1 per cento, nel caso di donne nere del 31 per cento. Il motivo è che i dati immessi per istruire l'algoritmo provengono soprattutto da visi maschili, prevalentemente bianchi.

Lo stesso avviene per il riconoscimento vocale in cui la voce maschile ha il 70% di probabilità in più di essere compresa. Inizialmente si diceva che questo avvenisse perché le donne "parlavano male", poi anche in questo caso si è verificato che il dato dipende dal fatto che le macchine sono istruite con voci principalmente maschili.

Per chiudere i gap e stereotipi riproposti nel mondo dell'intelligenza artificiale sono molti i rimedi che si stanno studiando oltre, naturalmente, a quello di ristabilire un equilibrio nelle cabine decisionali, nei vertici di comando delle aziende, pubbliche e private che siano. Qui voglio solo evidenziare che, assieme al tema della trasparenza,

sicuramente rilevante, cruciale è quello dell'empowerment: il patto con l'utente deve avere alla base un processo di costruzione di competenze che è molto più articolato e complesso di una infinita lista di diritti e doveri da scovare in un angolo remoto del sito o dell'APP di turno.

## **GUALTIERO MAZZI**

Buonasera a tutti; ringrazio Stefano Cuppi dell'invito qui perché mi ha dato modo soprattutto di imparare un sacco di cose e cercherò di portare un mio piccolo contributo su quello che mi è stato chiesto di intervenire e quindi su come si possono tutelare le diversità culturali territoriali. Ho sentito i vostri interventi con molta attenzione e una cosa che mi ha particolarmente colpito sono i due i temi fondamentali: uno quello dei contenuti e uno quello degli algoritmi che portano a proporre i contenuti.

Negli ultimi interventi si è parlato soprattutto degli algoritmi e quindi come coinvolgere la grande massa di utenti e tenerli concentrati sul tuo programma, e come il servizio pubblico deve porsi rispetto alle aziende private e quindi ai soggetti privati per adempiere a quelle che sono le missioni del contratto di servizio. Ringrazio tutti perché il tema è molto interessante e sono sicuro che merita molti approfondimenti

Anch'io ritengo come voi che il servizio pubblico abbia una mission più completa rispetto alle aziende private Ritengo che un'azienda, una società come la RAI debba avere una certezza dei contributi e dei fondi perché un'azienda non può programmare le proprie attività se i propri introiti sono legati a fattori estemporanei di anno in anno e deve avere una certezza di programmazione e di intervento per poter anche costruire formare e produrre.

Credo che nel costruire e nel dare offerte formative. Mi è piaciuto molto l'intervento che ha fatto De Chiara, prima perché tiene conto anche delle diversità culturali e le specificità che il nostro Bel Paese ha. Nel mio ruolo di presidente Corecom Veneto, più volte sono andato a parlare, a confrontarmi anche con il direttore di RAI 3 Veneto Le sedi regionali in questi ultimi anni sono state spogliate di competenze, di produzioni, di personale e di servizi. Allora quando si parla di cultura, del territorio e di contenuti del territorio, credo che i contenuti li possono fare sia il servizio pubblico che gli operatori privati. (...)

Lo Stato dalla liberalizzazione delle frequenze ha incassato 6 miliardi e mezzo: quanto di questi soldi investe nel servizio pubblico, quanto ne mette a disposizione

per creare contenuti, quanti di questi soldi magari li darà anche alle Sedi regionali della Rai ed alle televisioni locali per fare quei prodotti locali che sono molto interessanti. In una Regione come nel Veneto sono molto anche appetibili, perché anch'io quando guardo la televisione guardo la televisione nazionale mi informo, ma non passa giorno che non guardo anche l'informazione e la comunicazione che mi viene dai televisioni locali.

Ora la mia regione è fortunata perché ha un sacco di televisioni e che nonostante la moria di questi ultimi anni, comunque manterrà una distribuzione anche capillare. Ci sono alcune emittenti che trasmettono in tutto il territorio del Veneto anche in Friuli e nel nord est, ma altre regioni non hanno questa capacità. Chi l'aveva la sta perdendo e quindi nel pensare anche a qualcosa che va a beneficio delle emittenti locali che poi svolgono un servizio pubblico anche loro in una qualche forma. Io penso che le sedi della RAI devono essere un po' riviste e in un programma generale dove la RAI come azienda deve avere la certezza anche di introiti per poter programmare i propri interventi. Anche le sedi regionali possono avere un ruolo importante per fornire i contenuti che poi gli algoritmi lavoreranno perché ci sia una copertura anche di offerta generale. Ma mi è piaciuto anche molto l'intervento e prima che mi ha fatto fare un passettino in più e cioè in un in una sfera di televisione sempre più interattiva. Perché possiamo fare videogiochi, possiamo scegliere il programma, scegliere l'ora, scegliere tutto. Sarebbe importante poter anche intervenire in un prossimo futuro e condizionando anche le scelte degli algoritmi perché inizialmente sì io vengo scelto magari su dei criteri che inizialmente su quello che è la mia "non conoscenza", ma poi nel momento in cui ho fornito i miei dati in trasparenza, per quali so come vengo scelto e come vengono scelti i programmi, dovrebbero consentire anche a me di poter agire direttamente su questo. Nei social ci sono dei sistemi di sicurezza e dei sistemi che mi consentono di andare a modificare i miei profili, di mettere delle norme di sicurezza su quello che interagiscono con me: quindi non vedo perché in un'evoluzione anche della televisione non possa essere interattivo. Anche questo tipo di servizio c'è già sui social. Con il Corecom Veneto ho lavorato con due università venete, la prima l'Università di Padova, per temi legati al cyber bullismo, la *web reputation*, e con la Università di Verona ho da poco una bozza di lavoro. Una delle evoluzioni della mia Regione potrebbe essere anche quella di intervenire un pochino di più nella programmazione nei contenuti delle emittenti locali e dell'emittente RAI regionale. Se andiamo a guardare, e questo me l'ha proprio fornito come elemento l'Università di Verona che ha detto, guardate che è la legge: le legislazioni regionali hanno delle competenze in materia di televisione locale RAI, però raramente l'hanno sfruttata tutta e si limitano per lo più ai momenti informativi regionali. Ma in realtà potrebbero incidere molto di più, solo che non lo fanno o ancora non lo vogliono fare. Allora potrebbero addirittura arrivare a condividere anche dei contenuti.

C'è un passaggio della legge nel quale sono attribuiti alle regioni delle competenze in ordine al rilascio delle autorizzazioni per fornitore di contenuti o fornitori di servizi interattivi associati o di servizi di accesso condizionato, destinati alla diffusione in ambito rispettivamente regionale e provinciale. Cioè ci sono delle competenze ancora che le regioni non sfruttano e questo sarà un tema che è nell'aria e nella Regione del Veneto lo vorrei sviluppare.

Però ecco tornando al tema di oggi, credo che la RAI abbia una missione molto importante. È una società che dobbiamo valorizzare e valorizzare non tenendo conto della moda del momento, del partito politico che ha maggiore spendita rispetto a quello che ne ha meno, e non fluttuare avanti indietro così senza meta. Deve avere un contratto di servizio che contenga sì i punti che ci sono adesso, perché mi piacciono, li trovo giusti. Ma la mission pubblica ha qualcosa di diverso, deve contemperare più esigenze, però deve avere anche una certezza di bilancio, di risorse, perché sennò qualsiasi società pubblica o privata non può funzionare se non ha dei paletti ben chiari, dei paletti ben fissi, non può rendere al meglio. E sono d'accordo di sfruttare al massimo le potenzialità anche italiane delle nostre università, tenendo conto che i nostri prodotti, ne sono convintissimo, non sono meno di quelli degli altri perché se riusciamo a fare con meno risorse più programmi ed essere competitivi, credo che noi italiani ancora all'estero possiamo insegnare tanto.

## MARIO FRULLONE

Dopo aver ringraziato Stefano Cuppi per l'iniziativa che secondo me è stata almeno per quanto riguarda molto utile e interessante volevo provare a sintetizzare, dal mio punto di vista, cosa ho capito.

Certamente ho notato una chiara divaricazione tra gli obiettivi dei sistemi di raccomandazione diciamo commerciali da quelli del servizio pubblico; credo che il discorso che ha fatto Piero De Chiara sia stata veramente molto convincente, ampio e devo dire con la visione strategica che conoscevo essere nelle sue corde. Molto interessante, però che pone anche il problema dei rischi di segregazione, delle bolle. Probabilmente mette un po' in discussione che ci sia proprio necessità di fare un sistema di raccomandazione per il servizio pubblico, e se non sia invece intendendo con questo il rigetto della personalizzazione dell'offerta del servizio pubblico.

Stamattina abbiamo sentito il professor Ricci che ci ha parlato dei 170 ingegneri di Amazon che si occupano di sistemi di raccomandazione e abbiamo visto che

facciamo fatica a metterne uno solo. I tedeschi ne mettono 2, in Italia non riusciamo a metterne uno. Quindi mentre stiamo parlando, le nazioni del mondo e stanno andando a un ritmo acceso. Forse siamo un po' polarizzati dal 5 G e dagli investimenti che vediamo profilarsi per il 5 g, per le iniziative e l'ampliamento delle funzioni di operatori cloud nelle nostre infrastrutture, il dispiegamento di capacità di calcolo - si parla di 600.000 punti di calcolo distribuiti sul territorio. Quindi proprio un cambio di paradigma nel quale probabilmente la domanda che occorre porsi è qual è il ruolo delle dimensioni di un servizio pubblico confrontato con quello che sta succedendo intorno.

Parlavi del dell'arrivo di Disney e tutti quanti; noi pensiamo che Amazon se entrasse davvero nel settore sarebbe ancora più devastante. Siliato ha fatto vedere la riduzione della fruizione del servizio lineare, quindi il dubbio che mi porto è quello relativo alle ambizioni che può avere un servizio pubblico multimediale, che è un po' il caposaldo delle televisioni generaliste, è quello che tiene. Con tutti in bocca al lupo a RaiPlay, non so se ce la farà a reggere l'impatto di questi altri soggetti che invece arrivano con altri tipi di capacità di investimento. Per cui è molto interessante il discorso del gap che abbiamo fatto tra le competenze universitarie e applicazioni nell'ambito dei broadcaster, però ho paura che questo sarà colmato dai 170 ingegneri di Amazon...

## PASQUALE CAIANIELLO

Non ho molto da dire sono venuto come uditore con l'università non so se ci sono dei problemi siamo d'accordo c'è una ormai io libero ne ho trovati due di gap da colmare; uno è sul concetto di diversità, che non è chiarissimo. Che cosa è la diversità?

L'altro è la Explorability che è un'altra questione che mi sembra che è sorta, che fanno gli algoritmi? Spieghiamo, diamo la possibilità di capire perché c'è una quella decisione che arriva lì. Questa ricerca molto attuale in campo intelligenza artificiale del complesso, però è importante che ci sia un dialogo con gli addetti scientifici e degli addetti al lavoro, perché non è chiarissimo che cosa uno vuole che si spieghi. Nello spiegare il funzionamento a livello molecolare così come a livello psicologico di un individuo - e chiaramente lì son diverse ispirazioni - auspicherei un ulteriore dialogo (magari non adesso che dobbiamo prendere il treno).

## ANDREA MONTANARI

Va bene. Posso solo dire che come accennavo, sì confermo l'intenzione di proseguire questo binario e quindi continuare questo lavoro in una sessione nuova. Cercheremo di capire tirando un po' le somme di questa giornata di lavoro odierna, capire quale può essere il passo successivo. A mio giudizio una delle cose che mi colpiscono è che in tanti ambiti si registra un ritardo o comunque una marginalità del dibattito, ma anche una fatica del dibattito in Italia rispetto ai servizi pubblici.

Questo è un tema dal quale stiamo partendo in contemporanea con altre realtà che riteniamo solitamente più virtuose di noi. Gianluca purtroppo è dovuto andare via prima ma mi ha raccontato di essere stato a un workshop a Praga proprio questo weekend, nel quale si è parlato anche di questi temi ed è germinale ovunque la discussione. Questa è una cosa secondo me molto molto positiva e probabilmente sarebbe interessante nella prossima sessione avere molti rappresentanti della industria per capire come la vedono, che cosa pensano e che cosa pensano i servizi pubblici negli altri Paesi. A qual punto gli esempi più virtuosi? Ho letto che nella nuova versione di iPlayer - mi pare che sia la per radiofonica della BBC – o che si van all'app BBC Sounds, che si vanta di essere molto all'insegna della diversity, non so che cosa esattamente questo voglia dire. Probabilmente vuol dire che se io normalmente ascolto i Led Zeppelin mi fanno sentire Bach. Comunque voglio dire è molto interessante da capire per passare per portare a questo confronto sul terreno concreto delle esperienze dell'industria. Per il resto è stato ovviamente un dibattito estremamente interessante, vi ringrazio molto perché è stata un'occasione - certo vado via senza aver ascoltato Sassano, che siccome ascolto ogni tanto tanto suo fratello ed è pirotecnico - mi aspettavo di sentire cose altrettanto fino a tecnica.

Grazie a tutti e arrivederci