

Leggi questo articolo, una tua amica lo ha trovato interessante.

Un'introduzione alle opportunità e criticità dei *recommender system* per la personalizzazione dei contenuti audiovisivi.

Paolo Casagrande, Sabino Metta
Rai - Centro Ricerche e Innovazione Tecnologica

1. COSA C'È STASERA IN TV?

Quante volte, tornati a casa e magari in procinto di sederci a tavola, lo abbiamo domandato ai nostri familiari? E quante volte ci siamo messi a sfogliare (sì, su carta) la guida TV per cercare di trovare il programma di nostro gradimento o che mettesse tutti d'accordo? Ecco. In un futuro neanche troppo lontano, questa semplice domanda potrebbe cambiare radicalmente significato. E potrebbe essere formulata in maniera differente: "Cosa ci propone stasera la TV?"

In realtà, in molti casi questo scenario è già il presente. A proporci la migliore serie televisiva oppure il paio di scarpe di cui proprio non potevamo fare a meno, ora ci pensa una *macchina*. Per essere precisi un *Recommender System (RS)*, un *Sistema di Raccomandazione*. Non si tratta di un replicante di "Blade Runner" o dell'onnipresente "Grande Fratello" di Orwell: il suggerimento arriva da un algoritmo che tenta di intercettare le nostre preferenze.

I **RS** rappresentano una soluzione tecnologica di cui molti di noi fanno quotidianamente esperienza senza esserne consapevoli. Di più. I **RS** sono sempre più richiesti al fine di migliorare la qualità della nostra abituale fruizione di contenuti multimediali e di servizi digitali in genere.

L'articolo presenta brevemente l'evoluzione del panorama radiotelevisivo avvenuta negli ultimi anni e descrive il concetto di Recommender System e delle sue più importanti declinazioni tecnologiche, sottolineandone potenzialità, limiti e criticità.

L'avvento di internet e la proliferazione dei servizi digitali hanno determinato un radicale cambiamento della modalità di fruizione televisiva. La miriade di prodotti, servizi e contenuti disponibili su Internet ha abituato i suoi utenti all'interattività, alla possibilità di ricercare e scegliere quando e quale contenuto multimediale fruire, trasformando il tradizionale 'ascoltatore' televisivo in un attivo 'consumatore' in grado di esprimere i propri gusti e le proprie preferenze. Paradossalmente, la grande quantità di contenuti disponibili ha creato un sovraccarico informativo e reso centrale il problema di trovare i più adatti all'utente.

I Recommender System rappresentano la risposta tecnologica alla domanda di rilevanza. In questo contesto, il broadcaster ha l'opportunità di far evolvere il proprio paradigma comunicativo: oltre a diffondere i propri contenuti, può proporre all'utente i più rilevanti.

Prima di trarne qualunque conclusione, cerchiamo di capire meglio quale è il contesto in cui sono nati. Approfondiamo, senza entrare troppo nei dettagli, le loro principali caratteristiche. E cerchiamo di analizzare le loro potenzialità e criticità.

2. COME È CAMBIATA LA TELEVISIONE NEGLI ULTIMI ANNI

Facendo un salto indietro di qualche decennio, la televisione nasce come un oggetto, costituito da uno schermo e degli altoparlanti, in grado di sintonizzarsi attraverso una antenna al segnale trasmesso da un emittente televisivo, il *broadcaster*. Una volta acquisiti i diritti, il *broadcaster* aveva ed ha tuttora il compito di trasmettere via etere i programmi televisivi e radiofonici ^{Nota 1}: la *radio* e la *TV lineare*, così come l'abbiamo sempre conosciuta.

Oggi l'oggetto *televisore* appare decisamente diverso e più sfaccettato rispetto a quello di una volta. A parte le caratteristiche estetiche, molti televisori in vendita sono ormai dotati di una funzionalità tipica dei computer: la possibilità, cioè, di collegarsi ad Internet. Ed è così che sorgono i servizi *on-demand* in grado di erogare contenuti audiovisivi mettendoli a disposizione dell'utente, su richiesta: la cosiddetta *TV non lineare*. Al televisore tradizionale si aggiungono oggi altri dispositivi, come *tablet* e *smartphone*, spesso utilizzati come televisore o come *companion screen*, cioè in aggiunta ad esso.

Questa metamorfosi compiuta dalla televisione porta con sé una prima importante conseguenza: il palinsesto lineare viene potenzialmente ignorato dall'utente che decide autonomamente quale contenuto selezionare. E questo avviene mentre la

varietà di contenuti disponibili, tra programmi di cultura, serie televisive, rubriche, film è in continua crescita. Tra milioni di siti Internet e migliaia di canali televisivi e stazioni radio, tra tutte le informazioni ed i servizi disponibili ed accessibili, l'ascoltatore è sovraccaricato di informazioni e fa sempre più fatica ad orientarsi: il cosiddetto *information overload*. Di qui la necessità di restituire all'utente quanto sia per lui davvero rilevante.

Pertanto se da un lato è difficile pronosticare quale sarà domani il ruolo della TV, dall'altro ciascun editore è posto con sempre più urgenza di fronte alla sfida di accontentare in maniera puntuale i bisogni dei suoi utenti. Oltre a proporre all'utente nuovi contenuti, l'editore è chiamato a mettere a punto un *servizio di raccomandazione* in grado di proporre al singolo utente il contenuto giusto al momento opportuno e di personalizzare così la sua esperienza di fruizione. È quindi chiaro che tali *sistemi di raccomandazione* permettono la personalizzazione della singola esperienza di fruizione dell'utente.

Nel caso dei servizi di fornitura di contenuti multimediali, quello che l'utente desidera ricevere non è, ad esempio, una serie televisiva scelta a caso. Ma esattamente quella che stava aspettando e che poi deciderà di guardare. Come fa il servizio di raccomandazione ad *indovinare* i gusti dell'utente? A proporgli proprio quella serie televisiva che stava aspettando e non un'altra per lui meno interessante?

Proviamo in queste poche pagine a descrivere brevemente le funzionalità, complessità e criticità dei *Recommender System*.

3. UTENTI E CONTENUTI

Al fine di suggerire all'utente il contenuto per lui di maggiore rilevanza, i **RS** utilizzano *informazioni* sugli utenti e sui contenuti (*metadati*) e opportuni *algoritmi* per elaborare tali informazioni.

Per quanto riguarda i contenuti, la loro descrizione è affidata ai cosiddetti metadati in grado di catturare

Nota 1 - non entreremo nei dettagli relativi agli obblighi di legge per le emittenti televisive (vedi ad esempio le fasce orarie di protezione dei minori, l'affollamento pubblicitario, l'informazione sportiva). Per approfondimenti rimandiamo alle direttive ed alle regolamentazioni dell'AGCOM e al Regolamento Generale sulla Protezione dei Dati (GDPR 2016/679) dell'Unione Europea.

le peculiarità del contenuto stesso ed in questo modo di renderlo il più possibile visibile all'utente particolarmente interessato. Nel caso ad esempio di un film, i metadati possono essere rappresentati semplicemente dal genere (commedia, fantasy, documentario) oppure da rappresentazioni più precise, come titolo, cast, regista, anno di uscita, scenografia. La generazione di tali metadati rappresenta per un editore una attività molto onerosa e tutt'altro che scontata.

La raccolta e l'analisi dei dati relativi agli utenti, al loro contesto ed alle interazioni degli utenti con i contenuti, prende il nome di *profilazione dell'utente* ed ha lo scopo di descrivere, con un certo livello di affidabilità, il comportamento degli utenti in generale. Le pratiche di profilazione legate all'uso di Internet, dello *streaming musicale* e della *TV smart* possono essere estremamente sofisticate. Cerchiamo di approfondire meglio le ragioni.

Oggi giorno la realtà in cui viviamo è assolutamente iper-connessa. E per affermarlo non dobbiamo guardare cosa succedeva un secolo fa. Da quando sono nati *Internet* ed il *World Wide Web* (quest'ultimo nel 1991), il mondo sembra avere cambiato marcia. In maniera rapida e fluida qualunque utente può fruire un video musicale, creare un filmato e dividerlo, ascoltare una canzone, acquistare un prodotto (un paio di scarpe, un divano), usufruire di un determinato servizio digitale (pagare con un bonifico bancario, consultare le previsioni meteorologiche) o lasciare un commento su un *social network*. Qualunque utente può creare informazioni, accedervi, salvarle e dividerle.

È stato stimato che ogni giorno vengano scambiate 294 miliardi di e-mails e 20 miliardi di messaggi, mentre su *Facebook* vengano pubblicate 250 milioni di fotografie [15]. Secondo Cisco, la multinazionale specializzata nella fornitura di apparati di rete, attualmente il traffico dati globale su rete mobile si aggira sui 10 exabyte ^{Nota 2} mensili [16].

Nota 2 - Un exabyte equivale ad un miliardo di gigabyte, cioè 10^{18} byte

A produrre continuamente informazioni e dati non ci sono solo gli utenti che ascoltano una canzone, acquistano una lampadina su Amazon o esprimono il loro stato d'animo su un social network. Ogni giorno enormi quantità di dati vengono prodotti e condivisi anche dalle macchine, vedi ad esempio i sensori che raccolgono informazioni climatiche, immagini satellitari, fotografie e video, segnali GPS. Molti di questi sensori risiedono nei dispositivi mobili e connessi di cui quotidianamente ciascun utente fa uso. E la mole di tali dati, sia pur prodotta dalle macchine, contribuisce a fornire ulteriori informazioni sugli utenti. In particolare sul contesto in cui gli utenti vengono a trovarsi.

Tutti questi dati vengono generati in maniera continua da molteplici fonti, con formati diversi crescendo a velocità impressionanti. Lo scenario appena descritto è quello che oggi giorno viene definito con il termine **Big Data**: dati che vengono prodotti velocemente, in grande quantità, da fonti eterogenee e difficili da trattare con tecniche tradizionali.

La gestione dei **Big Data** rappresenta una sfida concettuale e tecnologica. In presenza di **Big Data** non è assolutamente conveniente fare affidamento sulle metodologie e tecnologie impiegate dalla cosiddetta *business intelligence* che opera su insiemi di dati limitati, puliti ed omogenei. I **Big Data** richiedono l'utilizzo di strumenti matematici e statistici più sofisticati e complessi. Si passa da una statistica *descrittiva* ad una *inferenziale*, da *modelli lineari* a *modelli non-lineari*.

Per trattare queste moli di dati, accanto alle tradizionali tecniche di *machine-learning* ne sono state messe a punto altre con l'obiettivo di analizzare flussi continui di dati [14]. Ai database tradizionali, quelli relazionali (**RDBMS**), vengono spesso preferiti quei sistemi software che non memorizzano i dati con campi uniformi (i cosiddetti **DB NoSQL**, tra cui database a grafo, orientati al documento, etc.). Al fine di gestire l'enorme mole di dati si passa ad elaborazioni in parallelo in grado di distribuire e poi recuperare correttamente le operazioni di calcolo effettuate su un certo numero di *nodi* (unità di elaborazione).

4. PRINCIPI DI FUNZIONAMENTO DEI RECOMMENDER SYSTEMS

Cercando di riassumere quanto descritto finora, i **Recommender Systems (RS)** comprendono tecniche e strumenti software in grado di suggerire all'utente gli oggetti più rilevanti tra i tanti a disposizione, e sono stati creati con lo scopo di aiutare a ridurre il sovraccarico informativo[19]. In un **RS**, il contenuto più rilevante per un determinato utente viene inferito sulla base dei dati a disposizione ed attraverso l'implementazione di specifiche logiche.

Nella sua forma più semplice un *sistema di raccomandazione* produce per ogni utente una lista di oggetti con relativa valutazione dell'utilità (*ranking*). Un oggetto, sia materiale sia intellettuale, rappresenta un qualunque contenuto multimediale e/o servizio digitale che possa essere suggerito: ad esempio il prodotto di un supermercato, un video, un programma radiofonico, una canzone, una notizia.

I diversi sistemi di raccomandazione possono essere classificati sulla base della specifica logica e dello specifico algoritmo attraverso i quali la lista finale di oggetti viene prodotta. Partendo dalla classificazione proposta da Burke [1], possiamo raggruppare i **RS** nelle sei seguenti categorie: *Collaborative*, *Content-based*, *Demographic*, *Knowledge-based*, *Community-based* e *Hybrid*. Riportiamo qui di seguito una loro sommaria descrizione aggiungendo la categoria *Editorial* che rappresenta la tradizionale raccomandazione fornita dall'editore.

Editorial: i suggerimenti vengono scelti e creati dall'editore del canale televisivo manualmente e valendosi della competenza di professionisti esperti. Chiaramente tale tipo di raccomandazione dipende dalla cultura e dalla particolare lingua (o dalle particolari lingue) adottate dal Paese di riferimento. I media diffusivi lineari come la TV digitale tradizionale si basano esclusivamente su questo tipo di raccomandazioni.

Collaborative: i suggerimenti per uno specifico utente vengono generati sulla base di giudizi o valutazioni (*ranking*) che altri utenti, in passato, hanno dato a determinati oggetti a disposizione. Dopo avere rappresentato tali informazioni (*utenti, oggetti, ranking*) attraverso una matrice, è necessario definire e quantificare una relazione di somiglianza o distanza tra gli utenti o tra gli oggetti che verranno raccomandati. Ad esempio, due oggetti **O** e **O'** sono definiti simili se sono stati scelti da molti utenti. Allora la raccomandazione potrebbe essere quella di suggerire **O'** a chi precedentemente ha scelto **O**. Un approccio detto *item-item*. Oppure, due utenti **U** e **U'** sono definiti simili se ad esempio hanno scelto categorie di oggetti simili. In questo caso la raccomandazione potrebbe essere quella di suggerire ad **U** un oggetto che in passato è stato scelto da **U'**. Un approccio detto *user-user*. Un esempio di *Collaborative Filtering* più sofisticato basato su fattorizzazione matriciale è riportato in Appendice A.

Content-based: il **RS** suggerisce ad uno specifico utente una lista di oggetti i quali risultano essere simili ad altri oggetti che quello stesso utente ha scelto in passato. In questa categoria, vengono confrontati gli oggetti stessi o le loro descrizioni, senza prendere in considerazione le azioni e/o il ranking che altri utenti potrebbero avere dato. Questo approccio presuppone la definizione appropriata di similitudine o distanza tra due o più oggetti. Ad esempio, assumendo una relazione di similitudine tra i programmi "Report" e "Presa diretta" (entrambi rappresentano programmi contenenti un giornalismo di inchiesta), poiché l'utente in passato ha guardato "Report", allora per l'utente potrebbe essere rilevante ricevere la raccomandazione del programma "Presa diretta".

Demographic: il suggerimento del *recommender system* si basa sulla caratterizzazione demografica dell'utente, ad esempio l'età, il sesso, la residenza. Si basa sull'assunto che segmenti demografici diversi richiedano prodotti diversi.

Knowledge-based: i suggerimenti si basano su regole derivate da una conoscenza esplicita dell'utente o degli oggetti.

Community-based o Social Recommender Systems: riassunto in [2], i suggerimenti tengono conto delle scelte e delle valutazioni degli amici dell'utente sui social networks, partendo dall'evidenza che tendiamo a tenere in maggiore considerazione i suggerimenti degli amici rispetto a raccomandazioni anonime.

Hybrid: le tecniche descritte sopra vengono combinate per ottenere suggerimenti più precisi cercando di far fronte alle debolezze di ciascun metodo. Molti RS commerciali sono di tipo ibrido.

Negli ultimi anni l'informazione relativa allo specifico *contesto* in cui si trova l'utente ha assunto sempre più importanza nella creazione del suggerimento. Per questa ragione, alle categorie sopra descritte si è affiancata la categoria **Context-Aware Recommender Systems (CARS)**, vedi [3]. Il *contesto* è definito da svariati fattori quali profilo, stato emotivo, attività, posizione geografica, condizioni atmosferiche, e qualsiasi altra cosa concorra alla definizione dello stato dell'utente, dell'oggetto o dell'interazione tra essi. Un esempio di fattore di contesto è il *tempo*: l'orario può influire sulla propensione di un dato utente a fare attività sportiva, il giorno della settimana può influire sulla sua scelta di fare una gita fuoriporta (più probabile nel weekend) e la stagione estiva sulla possibilità di acquistare un gelato. Altri fattori importanti sono ad esempio l'attività dell'utente, la sua posizione, l'eventuale compagnia.

La personalizzazione dei contenuti non è però l'unico fattore da tenere in considerazione. Anche i ricercatori di **Netflix** ammettono che accanto ai video raccomandati per il singolo utente, un criterio fondamentale è quello della popolarità dei video proposti [4]: la personalizzazione quindi è completata dalla presenza dei video più popolari, in modo simile per tutti gli utenti.

5. ALCUNI ESEMPI: NETFLIX, SPOTIFY, AMAZON

A titolo esemplificativo, proviamo a descrivere in questa sezione alcuni servizi basati su tecniche di raccomandazione e di cui probabilmente ciascuno di noi ha già fatto esperienza: **Netflix**, per il *video streaming*, **Spotify**, per lo *streaming di musica*, **Amazon.com**, per l'*e-commerce*.

5.1 NETFLIX

Netflix, il popolare servizio Internet di video streaming, utilizza i dati sulle valutazioni degli utenti ma anche data e ora di visione, il luogo dove il video è stato trovato nel sistema ed il dispositivo di visione (smartphone? Smart-tv?). Circa l'80% delle ore di visione sono influenzate dall'implementazione degli algoritmi di raccomandazione [4]. Il punto di accesso principale per l'utente, e quindi il principale luogo dove creare l'esperienza di raccomandazione, è la homepage del sito. Tale homepage è composta da molte sezioni (righe di video) e ciascuna beneficia di algoritmi diversi: non esiste quindi un unico schema di raccomandazione dei contenuti. Esempi di sezioni sono le *righe di genere*, che espongono suggerimenti di video personalizzati per l'utente secondo un genere (utilizzando un *Personalized Video Ranker*), i *Top picks*, che presentano le migliori raccomandazioni indipendentemente dal genere (secondo un *Top-N Video Ranker*) e così via.

Netflix ha esplorato in maniera sistematica le tecniche di raccomandazione fin da quando il suo maggiore obiettivo era vendere DVD, arrivando a mettere in palio nel 2006 un premio, il *Netflix Prize*, di un milione di dollari a chi avesse migliorato l'errore quadratico medio delle previsioni di almeno il 10% rispetto all'algoritmo allora in uso (competizione vinta nel 2009 dal team *BellKor's Pragmatic Chaos*).

5.2 SPOTIFY

L'azienda svedese **Spotify** ha lanciato nel 2008 un servizio streaming di musica divenuto negli anni molto popolare e arrivato a 100 milioni di utenti attivi nel 2016 (di cui 40 milioni a pagamento). Le playlist di canzoni suggerite dal servizio vengono

create con un **RS** basato principalmente su algoritmi di *collaborative filtering*. Il sistema è però in costante evoluzione per migliorare l'esperienza degli utenti.

5.3 AMAZON.COM

Dalle informazioni presenti sul web si evince che **Amazon.com**, uno dei più popolari servizi di commercio elettronico del mondo, utilizza un algoritmo di raccomandazione basato sulla *similarità item-item*. Nelle *raccomandazioni item-item*, generalmente, si calcola prima la similarità tra ogni coppia di prodotti **A** e **B**, partendo dagli utenti che hanno acquistato entrambi i prodotti. Se un utente acquista il prodotto **A**, gli vengono consigliati i prodotti **B** più simili ad **A**.

6. INIZIATIVE EUROPEE

6.1 EBU RecSys

La *European Broadcasting Union (EBU)* ha riconosciuto il ruolo chiave delle raccomandazioni nella personalizzazione dei contenuti radiotelevisivi, ed ha creato il gruppo **EBU RecSys**. Il gruppo si occupa di facilitare lo scambio di informazioni tra Media Company pubbliche che utilizzano tecnologie di raccomandazione, condividendo linee guida e conoscenza su aspetti tecnici e scientifici, legati alla normativa e alla standardizzazione. **EBU RecSys**

organizza inoltre lo sviluppo e l'evoluzione di un motore di raccomandazione reale, sviluppato partendo da componenti open source.

Attualmente, i servizi Radiotelevisivi pubblici europei che partecipano all'iniziativa stanno seguendo lo stesso approccio del gruppo nello sviluppo di soluzioni di raccomandazione: sviluppare all'interno dell'azienda il *Recommender System* che soddisfi i requisiti voluti.

Al fine di personalizzare l'esperienza di uno specifico utente è necessario innanzitutto riconoscere in maniera univoca l'originalità dell'utente stesso. A tal proposito l'autenticazione, o semplicemente il riconoscimento di sessioni con mezzi simili alle cookies dei browser web, rappresenta un punto chiave. Alcuni servizi permettono già di usare i cosiddetti *social login*, che semplificano le operazioni di autenticazione associando l'identità di un utente alle informazioni di base di un social network (ad es. Facebook o Twitter come avviene in [Rai Play](#)).

Il riconoscimento dell'utente su dispositivi connessi con input limitato, come le radio o le TV, è più laborioso. Una possibile soluzione è rappresentata dallo standard **Cross-Platform Authentication** [10] il quale risulta essere compatibile con il protocollo *OAuth 2.0* utilizzato dai principali social network per associare in modo semplice un utente ad un dispositivo.

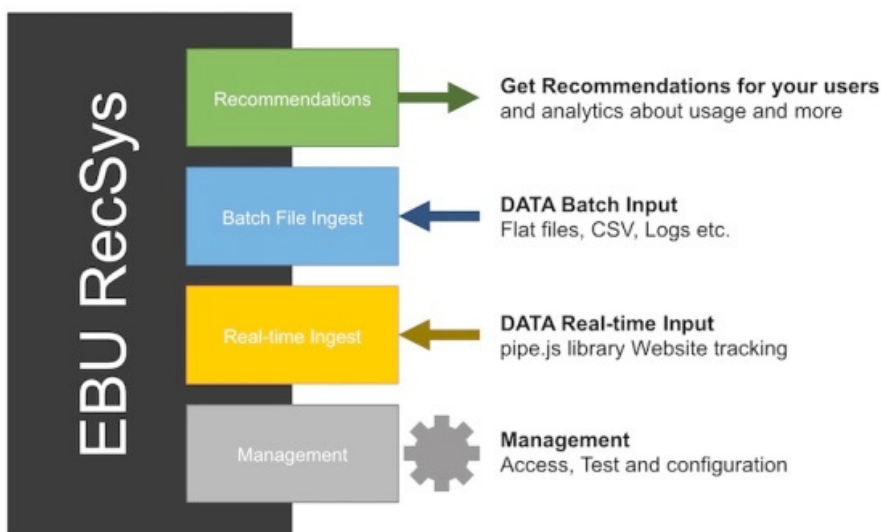


Fig. 1 – Il gruppo **EBU RecSys** sviluppa anche un raccomandatore per test ed uso da parte dei collaboratori

6.2 UN PROTOTIPO: LA HYBRID CONTENT RADIO

La **Hybrid Content Radio (HCR)** è un esempio di servizio audio personalizzato basato su raccomandazioni, ideato da Rai ed altri broadcaster pubblici della EBU [6]. HCR consiste nella personalizzazione del palinsesto di un servizio radio broadcast o streaming tradizionale, mediante la sostituzione di contenuti audio lineari con clip audio personalizzati, suggeriti in modo dipendente dal contesto. Il RS, fornendo suggerimenti sui contenuti, permette in altre parole di personalizzare il servizio di radio lineare.

Il prototipo realizzato consente di passare dall'ascolto live ad una lista di podcast suggeriti, sia manualmente (tasto skip), sia automaticamente, in modo proattivo, quando le condizioni di contesto siano appropriate [7][8]. La possibilità di fornire raccomandazioni in modo proattivo, quindi senza una richiesta esplicita dell'utente, è un ambito di ricerca aperto per i sistemi di raccomandazione.

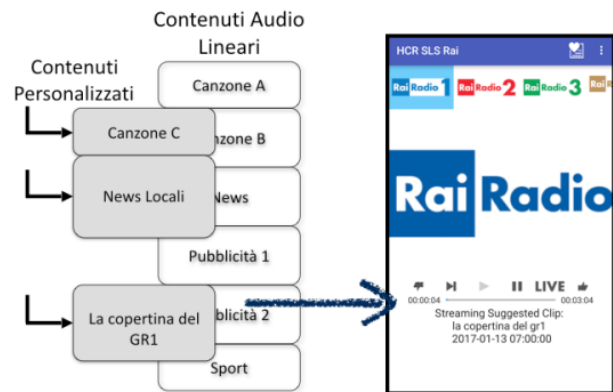


Fig. 2 – **Hybrid Content Radio**: la radio lineare viene personalizzata con contenuti raccomandati, dipendenti dal contesto.

7. LIMITI DEI RECOMMENDER SYSTEMS

L'efficacia degli algoritmi nell'aiutare l'utente nella scelta di un oggetto o di un servizio è soggetta a limitazioni che dipendono dal tipo di algoritmo, dal contesto e dal tipo di suggerimento che si vuole ottenere. Non esiste un unico algoritmo che risolva in modo ottimale ogni tipo di problema. Di seguito sono elencati alcuni limiti e sfide da superare.

Cold Start: alcune tecniche di raccomandazione necessitano di un insieme di dati iniziali per fornire suggerimenti validi. Ad esempio, perché un algoritmo di *Collaborative Filtering* dia risultati sensati è necessario che un oggetto sia già stato valutato da diversi utenti, e, analogamente, che un utente abbia già valutato alcuni oggetti. Ad esempio, il servizio social "GoodReads", che si focalizza sui libri, impone all'utente di valutare almeno 20 libri prima di fornire un suggerimento. Un algoritmo *Content-based* potrà, invece, fornire un suggerimento valido anche per un oggetto nuovo, a patto che a questo siano associati metadati adeguati.

Discovery: una lista di prodotti suggeriti da un RS dovrebbe consentire all'utente di esplorare anche prodotti nuovi e diversi tra loro, secondo un concetto di *discovery* o scoperta. Una eccessiva somiglianza tra i prodotti è in questo caso poco utile e spesso l'utente desidera anche esplorare nuovi tipi di prodotto rispetto ai soliti.

Filter Bubbles (bolle di filtraggio): è un concetto introdotto esaurientemente da Eli Pariser [9], che riguarda le informazioni che gli algoritmi lasciano arrivare all'utente. L'utente è come se fosse confinato in una bolla immaginaria costruita dagli algoritmi, bolla che permette il passaggio di alcune informazioni soltanto. In questo modo l'utente rischia di crearsi una visione molto parziale di molti fatti. Un esempio è il servizio social Facebook, che riordina, filtra e consiglia le informazioni anche in base alla storia dell'utente. Oltre alla rilevanza, altri fattori devono quindi essere tenuti in considerazione: ad esempio l'importanza in contesti economici e sociali più ampi o la possibilità di dare voce ad altri punti di vista anche scomodi.

Privacy: il *Recommender System* fornisce suggerimenti rilevanti per l'utente estraendo così un ordine dal caos e dalla complessità delle informazioni a disposizione. L'efficacia di questo processo dipende dalla quantità e dalla precisione dei dati personali messi a disposizione dall'utente. Il problema fondamentale in questo caso è garantire la trasparenza delle informazioni raccolte, dando la possibilità all'utente di accedere ad esse o di ritirarle. E poi, è assolutamente importante garantire la sicurezza di tali informazioni assicurando che esse non possano essere utilizzate da altri soggetti o per scopi diversi da quelli consentiti. La creazione di meccanismi standard per la memorizzazione, lo scambio e l'accesso alle informazioni dell'utente è un punto di grande interesse ancora aperto su cui anche MPEG sta lavorando (si veda [11]).

L'**Unione Europea** ha anche emanato nel 2016 una direttiva ed un regolamento sulla protezione dei dati personali che avranno valore negli stati membri a partire dai prossimi anni (si vedano [12] e [13]) e che richiedono trasparenza nel trattamento dei dati generati dagli utenti.

Fig. 3 – I limiti dovuti alle *filter bubbles* descritte da Pariser sono una delle sfide che gli algoritmi di raccomandazione devono superare. (foto di Clint Mason, flickr, CC BY 2.0)



8. CONCLUSIONI

In questo articolo abbiamo cercato di dare un'idea di cosa siano i *Recommender System*, descrivendo brevemente il loro funzionamento e citando alcuni esempi di aziende che adottano tali sistemi all'interno dei loro servizi. Abbiamo descritto la relazione tra **RS** e personalizzazione di un servizio, e come, in un mondo in cui le informazioni e i prodotti si moltiplicano, i sistemi di raccomandazione siano uno strumento necessario, capace di rendere più efficaci le ricerche ed ottenere prodotti più adatti a noi.

L'utilizzo appropriato dei **RS** deve tenere in adeguata considerazione la nostra privacy ed evitare la creazione di *filter bubbles* che ci confinino in un mondo chiuso, rendendoci troppo sicuri della completezza dell'informazione che stiamo ricevendo. Per questo è importante che gli algoritmi implementati tengano in considerazione anche aspetti che vanno oltre la rilevanza personale, e in questo il ruolo dell'editore è di primaria importanza.

Come molti altri prodotti dell'ingegno umano, anche i **RS** presentano grandi vantaggi ed enormi opportunità per il futuro, ma il buon senso e la trasparenza nella gestione delle informazioni sono fondamentali per mitigarne le criticità.

APPENDICE A

UN ESEMPIO DI COLLABORATIVE FILTERING OTTENUTO CON LA FATTORIZZAZIONE MATRICIALE SVD

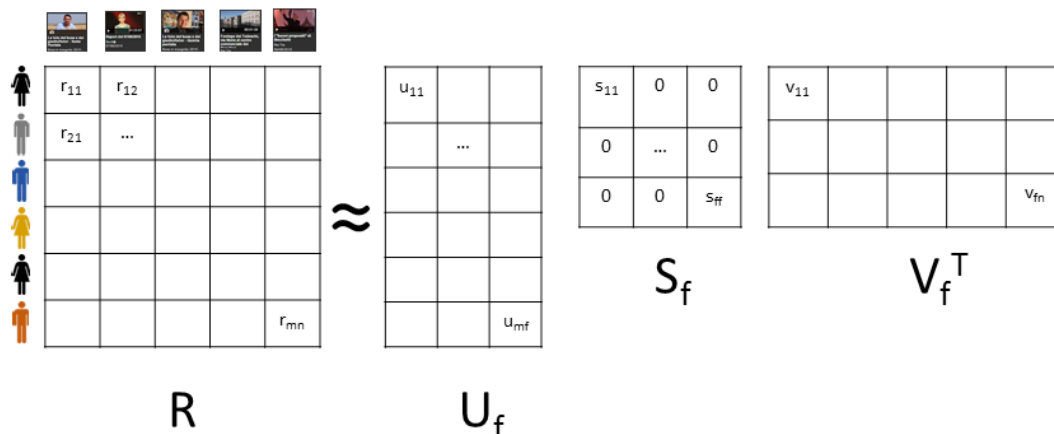
Diamo concretezza alle tecniche individuate descrivendo un esempio classico di Collaborative Filtering. Una tecnica molto efficace nella predizione delle scelte degli utenti è la **Fattorizzazione Matriciale**. Qui verrà presentata una tecnica chiamata **Singular Value Decomposition** (si veda per dettagli [5]). Sono stati presentati successivamente miglioramenti e tecniche più efficaci rispetto ad essa (si veda [18] e [19]) ma presenta il vantaggio di descrivere intuitivamente i concetti alla base della fattorizzazione. Il punto di partenza è una matrice **M** le cui righe rappresentano gli **m** utenti, e le cui colonne rappresentano gli **n** oggetti. La matrice conterrà i giudizi che gli utenti hanno dato agli oggetti (ad esempio un voto da 1 a 5).

Definiamo uno spazio di *fattori latenti*: nel caso di programmi televisivi può essere immaginato come una serie di caratteristiche del programma, come ad esempio se è un documentario, ben interpretato, con protagonisti affidabili, con attori famosi e così via. Esprimiamo poi gli utenti **m** con una matrice **U** in base all'importanza che assegnano a questi fattori latenti, e gli oggetti con una matrice **V** in base alla loro rappresentabilità con i fattori latenti. La fattorizzazione consiste nell'espressione della *matrice R* utilizzando una *matrice diagonale S* e due *matrici* di fattori latenti **U** e **V** in questo modo: $R=USV^T$. La

matrice S è una matrice diagonale $g \times g$ (con g righe e g colonne) i cui elementi sono i valori singolari della *matrice R* e g è il suo *rango*. Possiamo ridurre la complessità usando solo gli f valori singolari di **S** più significativi, con $f < g$: ciò porta ad una approssimazione $R_f = U_f S_f V_f^T$ che si dimostra essere la migliore di *rango f* e che riesce a catturare i fattori latenti più importanti rendendo i calcoli più veloci e riducendo la dimensione delle matrici, si veda figura 4.

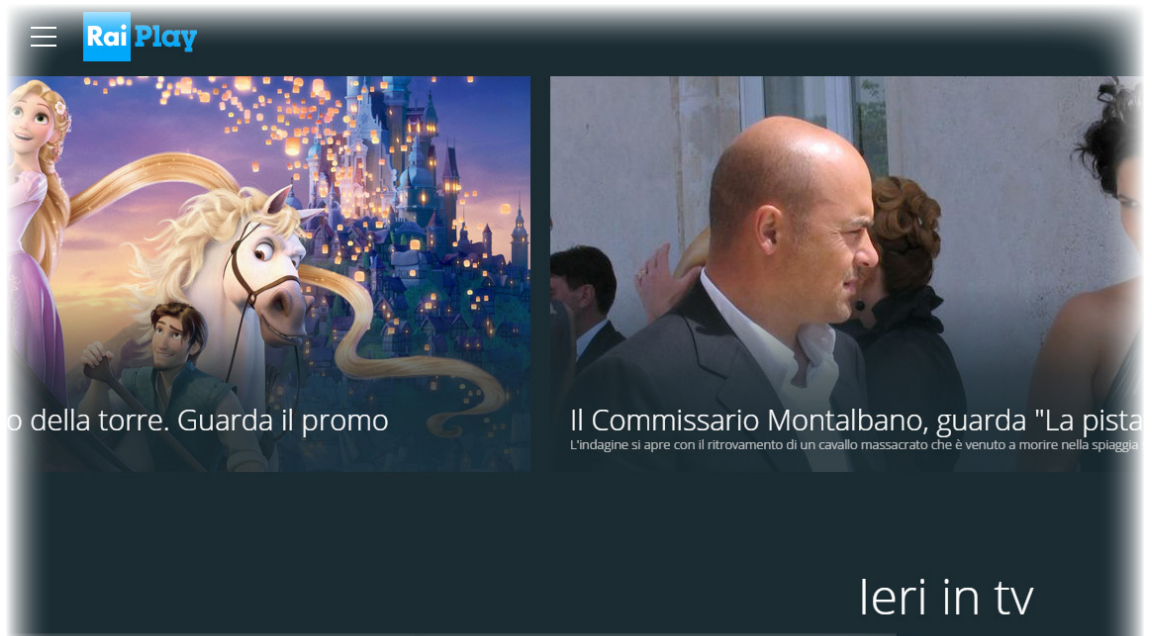
Si dimostra che un numero maggiore di fattori (fino a g) aumenta la precisione della rappresentazione, a discapito della velocità di calcolo. Le matrici **U**, **V** e **S** vengono calcolate con tecniche di approssimazione. In questo modo esprimiamo utenti e oggetti come combinazione di *fattori latenti* e utilizziamo questa rappresentazione per calcolare velocemente la vicinanza di un oggetto da raccomandare ad un utente. La predizione di una valutazione r_{ui} , che verrà usata per la raccomandazione, sarà del tipo $r_{ui} = \mu + U_f \sqrt{S_f}(u) \sqrt{S_f} V_f^T(i)$ dove (u) indica la riga corrispondente all'utente u e (i) indica la colonna dell'oggetto i e μ è la media di tutti i giudizi. Si può poi tenere in conto il discostamento dalla media dei giudizi di utenti e oggetti specifici, e aggiungere i termini b_i e b_u , discostamenti dalla media per l'utente u e l'oggetto i . Si noti che la tecnica descritta non assegna un significato ai *fattori latenti*: potrebbero avere un significato intuitivo, legato ad esempio al genere del programma, ma non è necessariamente così: sono fattori utili a creare un modello approssimato di utenti e oggetti.




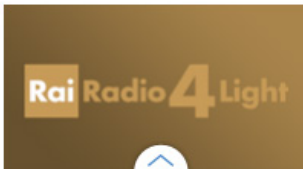
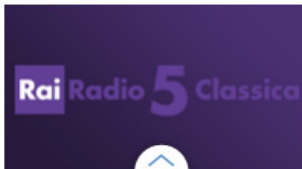




Fig. 4 – Una rappresentazione delle matrici utilizzate nella fattorizzazione matriciale SVD. L'approssimazione è dovuta al fatto che viene scelto $f < g = \text{rank}(R)$



BIBLIOGRAFIA

- [1] R. Burke, [Hybrid recommender systems: Survey and Experiments](#), in "User Modeling and User-Adapted interaction", vol. 12, Numero 4, Novembre 2002, pp. 331-370.
- [2] F. Ricci, L. Rokach e B. Shapira, [Recommender systems: introduction and challenges](#), in "Recommender Systems Handbook", Springer US, 2015, pp. 1-34.
- [3] G. Adomavicius, B. Mobasher, F. Ricci e A. Tuzhilin, [Context-Aware Recommender Systems](#), in "AI Magazine", vol. 32, Numero 3, Autunno 2011, pp. 67-80.
- [4] C. A. Gomez-Uribe e N. Hunt, [The Netflix Recommender System: Algorithms, Business Value, and Innovation](#), in "ACM Transaction on Management Information Systems", vol. 6, Numero 4, Gennaio 2016, Articolo 13.
- [5] B. Sarwar, G. Karypis, J. Konstan e J. Riedl, [Application of dimensionality reduction in recommender systems – a case study](#), ACM WebKDD-2000 Workshop, 2000
- [6] P. Casagrande, A. Erk, S. O'Halpin, D. Born e W. Huijten, [A Framework for a Context-Based Hybrid Content Radio](#), in "Proceedings of 2015 International Broadcasting Convention (IBC)", 2015.
- [7] M. Braunhofer, F. Ricci, B. Lamche e W. Worndl, [A context-aware model for proactive recommender systems in the tourism domain](#), in "MobileHCI '15 Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct", Agosto 2015, pp. 1070-1075
- [8] P. Casagrande, M. L. Sapino e K. S. Candan, [Context-Aware Proactive Personalization of Linear Audio Content](#), in "Proceedings of the 20th International Conference on Extending Database Technology (EDBT)", Marzo 2017, pp. 574-577
- [9] E. Pariser, [The filter bubble: What the Internet is hiding from you](#), Penguin UK, 2011.
- [10] ETSI TS 103 407 - V1.1.1 (2016-04), [Cross Platform Authentication for limited input hybrid consumer equipment](#)
- [11] S. Metta, P. Casagrande, A. Messina, M. Montagnuolo e F. Russo, [Leveraging MPEG-21 user description for interoperable recommender systems](#), in "Proceedings of the 31st Annual ACM Symposium on Applied Computing (SAC'16)", pp. 1072-1074, 2016.
- [12] [DIRETTIVA \(UE\) 2016/680 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO del 27 aprile 2016 relativa alla protezione delle persone fisiche con riguardo al trattamento dei dati personali da parte delle autorità competenti a fini di prevenzione, indagine, accertamento e perseguimento di reati o esecuzione di sanzioni penali, nonché alla libera circolazione di tali dati e che abroga la decisione quadro 2008/977/GAI del Consiglio](#)
- [13] [REGOLAMENTO \(UE\) 2016/679 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO, 27 aprile 2016 relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE \(regolamento generale sulla protezione dei dati\)](#)
- [14] J. Leskovec, A. Rajaraman e J. Ullman, [Mining of Massive Datasets](#), Cambridge University Press, 2014
- [15] M. Rasetti, [I Big Data: rivoluzione tra scienza e conoscenza](#), in "Oxygen", Numero 22, Febbraio 2014, pp. 110-113
- [16] [Cloud and Mobile Network Traffic Forecast - VNI](#), Cisco, ultimo accesso 1 Febbraio 2017
- [17] Commissione Europea, [Building a European Data Economy](#)
- [18] Y. Koren, R. Bell e C. Volinsky, [Matrix factorization techniques for recommender systems](#), in "IEEE Computer", vol. 42, Numero 8, Agosto 2009, pp. 42-49
- [19] F. Ricci, L. Rokach e B. Shapira, [Recommender Systems Handbook](#), Springer, 2015.



<p>RADIO 1</p>  <p>10:05 Radio1 music club</p>	<p>RADIO 2</p>  <p>10:00 Radio2 come voi</p>	<p>RADIO 3</p>  <p>10:00 Tutta la citta' ne parla</p>
<p>RADIO 4 LIGHT</p>  <p>10:21 Radio4light leggerezza</p>	<p>RADIO 5 CLASSICA</p>  <p>Poulenc: sonata per cl e pf</p>	<p>ISORADIO</p>  <p>10:06 Musica traffico e pubblica utilita'</p>
<p>RADIO 6 TECA</p>  <p>TECA MUNDIAL</p>	<p>RADIO 7 LIVE</p>  <p>Rai Radio 7 Live</p>	<p>RADIO 8 OPERA</p>  <p>Rai Radio 8 Opera</p>

