

Myrror: una piattaforma per Holistic User Modeling e Quantified Self

Cataldo Musto, Giovanni Semeraro, Marco de Gemmis, Pasquale Lops

Dip. di Informatica - Università degli Studi di Bari

{cataldo.musto, giovanni.semeraro, marco.degemmis, pasquale.lops}@uniba.it

Abstract

In questo contributo presentiamo Myrror, una piattaforma di *holistic user modeling* finalizzata a costruire un *profilo* dell'utente che codifichi in un'unica rappresentazione tutte le diverse sfaccettature che descrivono un individuo, come i suoi interessi, le sue attività, le sue abitudini, il suo umore e così via.

Tale rappresentazione è basata sulle informazioni che l'individuo diffonde attraverso i social networks e che gestisce attraverso i propri dispositivi personali. Nello specifico, Myrror acquisisce dati personali provenienti da diverse sorgenti come Twitter, Facebook, Instagram, smartphone Android e dispositivi FitBit. Successivamente, tali informazioni vengono aggregate ed elaborate al fine di inferire caratteristiche di alto livello dell'individuo e popolare così il suo modello utente con informazioni utili. I profili costruiti mediante la piattaforma sono resi disponibili all'utente finale e a servizi di terze parti. Nel primo caso, i dati vengono mostrati attraverso una interfaccia utente. Nel secondo caso, i profili sono esposti a sviluppatori esterni al fine di sviluppare servizi personalizzati basati su *holistic user models*. In entrambi i casi, l'utente ha il totale controllo sulle informazioni che la piattaforma acquisisce ed elabora, in piena trasparenza e rispetto della privacy.

1 Introduzione

I recenti sviluppi nell'area dei *social networks* e del *quantified self* [Swan, 2013] hanno alimentato la ricerca nell'area della personalizzazione, poichè numerose informazioni relative ai singoli individui possono essere ora acquisite analizzando i social media (es., i post scritti dall'utente, le sue connessioni sociali, etc.) e acquisendo dati fisici e fisiologici (es., i passi percorsi ogni giorno) disponibili su dispositivi come FitBit. Queste informazioni sono estremamente utili per costruire dei profili utente più ricchi e completi, che vadano a codificare le diverse sfaccettature della persona in un unico profilo.

Per questo motivo, in questo contributo introduciamo il concetto di *Holistic User Model* (modello di profilazione olistico), inteso come il processo che porta alla costruzione di

un'unica identità digitale dell'individuo che vada ad aggregare ed elaborare tutte le informazioni eterogenee relative all'utente che sono disponibili *on-line* (i suoi acquisti, contenuti generati sui social networks, la propria rete sociale, etc.), e nel *mondo reale* (dati GPS, attività quotidiane, pasti, attività).

La nostra piattaforma acquisisce dati grezzi relativi all'utente, effettua elaborazioni e aggregazioni su questi dati utilizzando tecniche di machine learning ed elaborazione del linguaggio naturale e crea infine una rappresentazione dell'individuo che chiamiamo *profilo olistico*. L'elemento caratterizzante della nostra piattaforma è data dal fatto che il processo di profilazione avviene in totale trasparenza, dando all'utente il totale controllo dell'intero processo di profilazione come suggerito dalle recenti indicazioni relative alla recente General Data Protection Regulation (GDPR) ¹. Il nostro framework dà la possibilità all'utente di decidere esplicitamente quali dati codificare all'interno del suo profilo e quali informazioni del profilo stesso possono essere rese disponibili ad applicazioni di terze parti, in totale controllo e nel pieno rispetto della privacy dell'individuo.

2 Holistic User Models

Nella nostra visione, un profilo olistico è una rappresentazione complessa dell'individuo che aggrega tutte le informazioni relative alla persona che sono disponibili in Rete e che sono gestite da suoi dispositivi personali. Il modello concettuale di profilazione implementato nella piattaforma prevede la rappresentazione dell'individuo attraverso otto diverse sfaccettature: *dati anagrafici, preferenze, emozioni, elementi cognitivi, attività, connessioni sociali, dati sulla salute*.

I dati anagrafici riguardano le informazioni personali dell'utente come il nome, il cognome, l'anno di nascita, il peso e così via. Le preferenze codificano tutto ciò che riguarda gli interessi e i gusti della persona. Le emozioni modellano l'andamento dell'umore e delle emozioni dell'individuo nel corso del tempo. Gli elementi cognitivi raccolgono informazioni legati a personalità, empatia e capacità di apprendimento. Le attività descrivono tutto ciò che riguarda i comportamenti della persona (dove lavora, in quali posti si reca di seguito, se è sedentario o meno). Le connessioni sociali codificano le relazioni tra l'utente e altri individui all'interno delle proprie

¹<https://tinyurl.com/gdpr-reg>

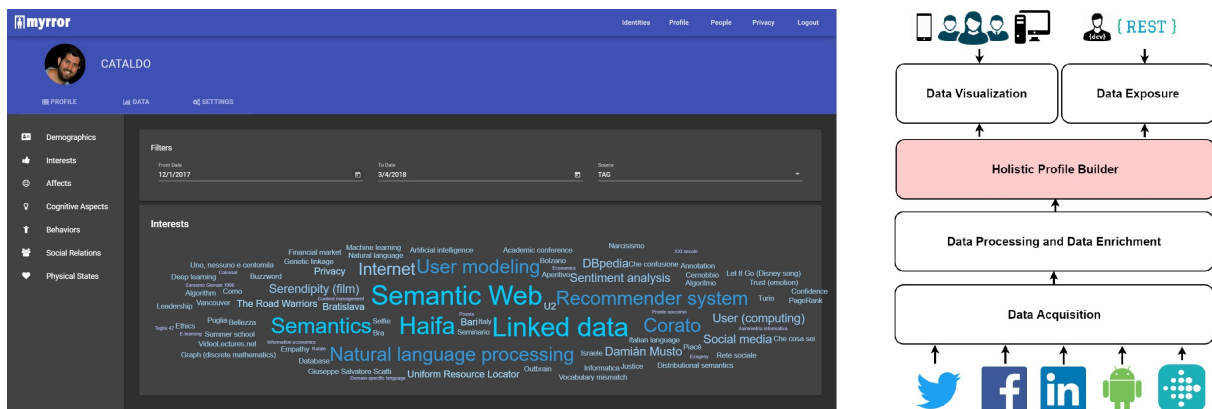


Figura 1: Visualizzazione Dati Relativa agli Interessi dell'Individuo (a sinistra) e Organizzazione Generale della Piattaforma (a destra).

reti sociali mentre infine i dati relativi alla salute memorizzano tutto ciò che riguarda gli aspetti fisiologici della persona (qualità del sonno, battito cardiaco, etc.)

3 Descrizione della Piattaforma

Come mostrato in Figura 1, la nostra piattaforma è organizzata utilizzando la consueta architettura a livelli caratterizzata da un *data acquisition layer*, un *data processing and enrichment layer*, un *holistic profile builder* e un livello finale per *data visualization and data exposure*. Per motivi di spazio, non ci è possibile fornire ulteriori dettagli sul processo di profilazione.

In ogni caso, l'obiettivo delle componenti implementate nel primo livello è quello di acquisire dati grezzi relativi agli utenti dalle singole sorgenti dati. Nel primo rilascio della piattaforma è possibile estrarre informazioni da *social networks* come Twitter, Facebook, Instagram e LinkedIn, da dispositivi Fitbit e da smartphone Android. A seguire, tutti i dati grezzi sono elaborati e aggregati al fine di modellare le sfaccettature di alto livello del profilo. Nel caso dei dati di natura testuale, tutti i post o i Tweet scritti dall'individuo sono elaborati attraverso una pipeline di algoritmi di Natural Language Processing, Sentiment Analysis e Topic Modeling algorithms, che rispecchia quella implementata in [Musto *et al.*, 2015]. Dall'altra parte, i dati personali sono elaborati attraverso tecniche di machine learning e servono a inferire caratteristiche dell'utente (es., utente *sedentario* o *attivo*) attraverso l'analisi dei dati grezzi.

Nel terzo passaggio, tutti i dati estratti in precedenza sono elaborati e arricchiti per popolare le diverse *sfaccettature* del nostro modello di profilazione olistica. Ciascuna sfaccettatura può essere popolata in modo esplicito, utilizzando dati già disponibili (es., il nome o il cognome della persona) o in modo implicito, utilizzando degli algoritmi per acquisire tali informazioni (ad esempio, gli interessi o la *personalità* possono essere inferiti a partire dal testo, o le *attività* possono essere inferite a partire dai dati GPS).

Per concludere, tali profili olistici sono resi disponibili agli utenti finali e a sviluppatori esterni. Nel primo caso, le informazioni codificate nei profili sono mostrate all'utente attraverso una interfaccia grafica (vedi Figura 1). Uno screen recording che mostra le principali funzionalità della piatta-

forma è anche disponibile online². I nostri profili sono resi infine disponibili a servizi di terze parti, che possono accedere alle informazioni gestite dalla piattaforma. È importante notare che i servizi esterni possono accedere solo alle sfaccettature e agli elementi del profilo che l'individuo ha etichettato come *pubblici*. Applicazioni esterne non possono accedere né ai dati di basso livello acquisiti dalla piattaforma né ad alcune sfaccettature del profilo che l'utente ha esplicitamente deciso di tenere private.

4 Conclusioni e Sviluppi Futuri

In questo lavoro abbiamo presentato una piattaforma che permette la creazione di *profili utente olistici*. Questa rappresentazione dell'individuo codifica numerosi aspetti della persona, permettendo così di costruire dei modelli di profilazione ampi e significativi che possono essere utilizzati per numerosi task di personalizzazione. Ad esempio, tale piattaforma potrebbe essere utilizzata in scenari avanzati per la raccomandazione di cibo o di ristoranti, in grado di gestire allo stesso tempo informazioni sulle nostre preferenze in ambito di cibo, le nostre attività (sono stato sedentario questa settimana?), i nostri obiettivi (sono a dieta?) e anche il nostro umore, così da permettere suggerimenti più completi e soddisfacenti.

Nei prossimi mesi tale piattaforma sarà utilizzata in due casi d'uso con utenti reali, finalizzata a valutare l'efficacia della piattaforma e della strategia di profilazione.

Ringraziamenti: questo lavoro è stato supportato dal progetto 'Semantic Holistic User Modeling for Personalized Access to Digital Services and Content' finanziato dalla Regione Puglia nell'ambito del programma FutureInResearch.

Riferimenti bibliografici

- [Musto *et al.*, 2015] C. Musto, G. Semeraro, P. Lops, e M. de Gemmis. Crowdpulse: A framework for real-time semantic analysis of social streams. *Information Systems*, 54:127–146, 2015.
- [Swan, 2013] M. Swan. The Quantified Self: Fundamental disruption in big data science and biological discovery. *Big Data*, 1(2):85–99, 2013.

²<https://youtu.be/3YRlcUhNZnQ>