

Twitter Vigilance: Modelli e Strumenti per l'Analisi e lo Studio di Dati Social Media ed il Monitoraggio in Real Time

Daniele Cenni, Paolo Nesi, Gianni Pantaleo, Irene Paoli, Imad Zaza

Università di Firenze, Dipartimento di Ingegneria dell'Informazione, DISIT Lab (Distributed [Systems and internet | Data Intelligence and] Technologies Lab)

Abstract. Le tecniche e gli strumenti di analisi dei Social Media stanno diventando sempre più importanti per la previsione di eventi e tendenze, per la diagnosi precoce tramite il monitoraggio sociale e l'uso degli utenti come sensori. In questo contesto, Twitter.com è uno dei canali più interessanti per la sua diffusione e le dinamiche di risposta veloce. Dall'Aprile 2015 DISIT lab ha attivato lo strumento Twitter Vigilance per permettere a ricercatori ed analisti di effettuare analisi e ricerche su dati derivati da Twitter riferiti ad aree tematiche diverse. In questi ultimi 12 mesi, tramite Twitter Vigilance sono state sviluppate moltissime analisi negli ambiti: ambiente e meteo, disastro ambientale e resilienza, farmacologia, servizi smart city, turismo, cultura, intrattenimento e TV, grandi eventi, etc. In questo articolo si presenta una sintesi della soluzione Twitter Vigilance, con i suoi strumenti, ed alcune informazioni che possono essere utili per comprenderne i meccanismi.

Keywords. Social media, Twitter Vigilance, Natural Language Processing, Sentiment Analysis, Real Time

Introduzione

L'uso di strumenti che analizzano i dati provenienti da social network e/o blog è oramai diffuso. Fra le social network più diffuse come Facebook, Twitter, G+, etc., ve ne sono alcune più o meno adatte a poter essere utilizzate per fini di ricerca e di analisi. Fra queste Twitter è una delle più interessanti per le sue caratteristiche di apertura e velocità di reazione della sua utenza. In letteratura, soluzioni basate sull'analisi di dati provenienti da Twitter sono state utilizzate per: il rilevamento dell'arrivo di nuove droghe sul mercato, l'identificazione precoce di eventi e disastri, per la definizione di modelli e soluzioni di capaci di effettuare delle previsioni, per l'analisi dell'apprezzamento di prodotti e persone (in termini di sentiment negativo, positivo, neutro, etc.), per lo studio della risposta ad eventi di intrattenimento televisivo, per la stima delle dimensioni della folla e/o per le predizioni del numero di persone coinvolte in grandi eventi, per la predizione degli andamenti in borsa, ecc. In sostanza, alcuni dati estratti da Twitter, opportunamente elaborati, possono essere sfruttati per calcolare metriche e definire modelli matemati-

ci specifici che possono essere utilizzati come strumenti di previsione, diagnosi precoce e per l'analisi della risposta sociale. Ovviamente sono risultati che possono avere un grande valore, oppure un valore limitato dipendentemente dalla correlazione fra la massa delle persone e l'utenza di Twitter.

La maggior parte delle metriche basate su dati provenienti da Twitter si fondano sul conteggio del numero di tweet, del numero di retweet, del numero di follower/amici, il numero di commenti, le relazioni fra utenti, e molti altri parametri che possono essere ottenuti con svariate, e più e o meno complesse elaborazioni. Su questa base, DISIT lab dell'Università degli studi di Firenze ha sviluppato la famiglia di strumenti Twitter Vigilance che oramai sono attivi 24 ore su 24 dall'aprile 2015. In questo periodo sono stati raccolti e analizzati oltre 220 milioni di Tweet per scopi di ricerca.

1. Servizi accessibili per gli utenti di Twitter Vigilance

Twitter Vigilance colleziona in modo automati-

co i dati e su questi effettua operazioni di data mining del contenuto. In accordo alla terminologia di Twitter Vigilance, l'utente può creare dei "Canali" di ascolto, dove ogni Canale di Twitter Vigilance può essere configurato per monitorare un gruppo di chiavi di ricerca su Twitter.com con una sintassi espressiva ed efficace. Dall'interfaccia utente è possibile ottenere direttamente l'andamento di alcune metriche e molte altre informazioni e andamenti. Alcuni dei Canali di Twitter Vigilance sono accessibili e a disposizione del pubblico tramite la pagina web <http://www.disit.org/tv/>. Inoltre, la pagina di riferimento per informazioni e news su Twitter Vigilance è <http://www.disit.org/6693>. Twitter Vigilance fornisce una serie di strumenti di analisi e di soluzioni di base ed avanzati per il controllo di metriche basate su dati che provengono da Twitter. In particolare, Twitter Vigilance è in uso in svariati contesti come ad esempio, nel: Progetto Sii-Mobility Smart City Nazionale <http://www.sii-mobility.org> per lo studio degli aspetti di mobilità e trasporti; per la valutazione della qualità del servizio, per lo studio di eventi; Progetto RESOLUTE H2020 <http://www.resolve-eu.org> per gli aspetti di resilienza, la valutazione della risposta a eventi critici in città e/o lo studio di modelli per la diagnosi precoce; Progetto REPLICATE H2020 per il monitoraggio della comunicazione relativa a servizi innovativi in città a supporto della Control Room della città; Corso di Master in Big Data Analytics and Technologies for Management, MABIDA, <http://www.disit.org/mabida> a supporto delle sperimentazioni; e in svariati progetti di più piccole dimensioni.

Gli strumenti della soluzione integrata Twitter Vigilance sono accessibili via WEB e sono adatti per lo studio, la ricerca ed il monitoraggio di social media via Twitter (Figura 1). In particolare, i loro punti accesso sono:

- Twitter Vigilance main tool: <http://www.disit.org/tv/>
- Real Time Twitter Vigilance: <http://www.disit.org/rttv/>
- Twitter Vigilance Advanced Search faci-

lity based on SOLR: <http://tvsolr.disit.org/search/?collection=1>

2. Conclusioni

La famiglia di strumenti di Twitter Vigilance offre soluzioni per l'automazione di analisi del canale social media Twitter e permette di effettuare la diagnosi precoce di problemi, lo sviluppo di modelli predittivi, lo studio dei social media, in vari domini, ecc. Le varie sperimentazioni hanno permesso alla piattaforma Twitter Vigilance di crescere e diventare uno strumento efficace e semplice da utilizzare. Questi risultati si fondano su più di un anno di rilevamento e analisi. Gli strumenti presentati non sono la risposta definitiva all'analisi del social media o di dati Twitter ma permettono di collezionare i dati in automatico, produrre delle deduzioni semplici anche in automatico, produrre indagini semplici in modo pressoché standardizzato, automatizzato ed assistito, e allo stesso tempo lasciano spazio per lo sviluppo di analisi dettagliate e approfondite dei dati collezionati, capire se vi sono margini per la definizione di modelli più complessi e raffinati, anche con il supporto del team Twitter Vigilance del DISIT lab, se necessario.

Riferimenti bibliografici

Grasso V., Zaza I., Zabini F., Pantaleo G., Nesi

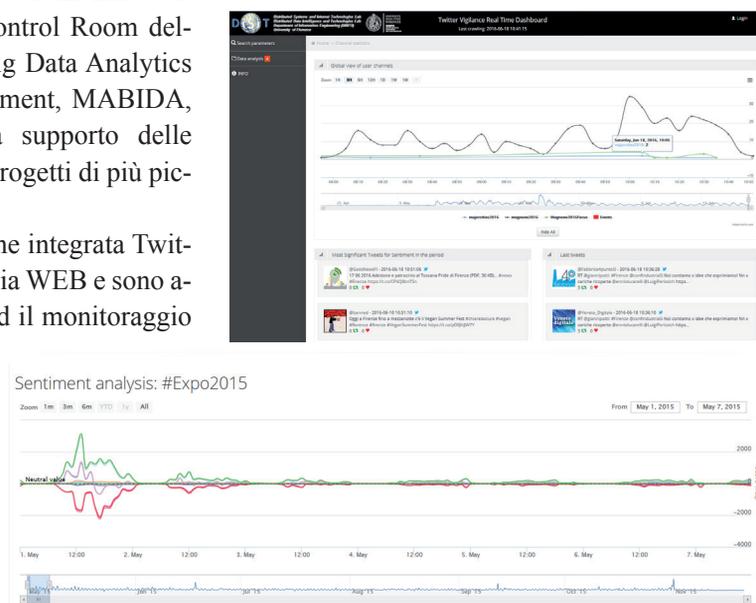


Fig. 1 Real Time Twitter Vigilance (in alto), una vista sulla Sentiment Analysis (in basso)

P., Crisci A. (2016), Weather events identification in social media streams: tools to detect their evidence in Twitter. PeerJ Preprints 4:e2241v1 <https://doi.org/10.7287/peerj.preprints.2241v1>.

Valentina Grasso, Alfonso Crisci, Alice Cavaliere, Simone Menabeni, Paolo Nesi, (2015), Un dialogo costruito anche grazie a Twitter, in Il consumo di suolo: strumenti per un dialogo, Laura Cremonini Ed., Istituto di Biometeorologia IBIMET-CNR, Italia, Bologna, ISBN 978889559724.

Alfonso Crisci, Valentina Grasso, Simone Menabeni, Paolo Nesi, Gianni Pantaleo, (2016) "Predicting Number of Visitors and TV programme Audience by Using Twitter Based Metrics", submitted.

Paolo Nesi, Alice Cavaliere, Gianni Pantaleo (University of Florence), Alfonso Crisci (IBIMET-CNR), Valentina Grasso (Lamma Consortium), Simone Menabeni (University of Florence) Monitoring Public Attention on Environment Issues with Twitter Vigilance.

Daniele Cenni

daniele.cenni@unifi.it



Si è laureato in Ingegneria Informatica e ha conseguito il Dottorato di Ricerca in Telematica e Società dell'Informazione presso l'Università degli Studi di Firenze. È Assegnista di Ricerca presso DISIT Lab (DINFO) dell'Università degli Studi di Firenze. I suoi principali interessi di ricerca includono l'Information Retrieval e l'analisi del comportamento utente in ambito Social Network e Smart City

Paolo Nesi

paolo.nesi@unifi.it



È chair del DISIT Lab dell'Università degli Studi di Firenze, ed ordinario di sistemi di elaborazione. Ha coordinato svariati progetti di ricerca ed innovazione, della commissione

europea, nazionali e regionali. Attualmente coordina la linea di ricerca Km4City e i progetti Sii-Mobility MIUR Smart City nazionale mobilità e trasporti, RESOLUTE H2020 sulla resilienza nelle infrastrutture critiche e dei sistemi di trasporto, mentre DISIT lab contribuisce in vari altri progetti su big data, Smart city, Smart cloud, industria 4.0.

Gianni Pantaleo

gianni.pantaleo@unifi.it



Ha conseguito la laurea in Ingegneria delle Telecomunicazioni nel 2007 e il dottorato di ricerca in Telematica e Società dell'Informazione nel 2012 presso l'Università di Firenze. Lavora presso il DISIT Lab (Dipartimento di Ingegneria dell'Informazione dell'Università di Firenze); i suoi principali interessi riguardano l'Information Retrieval ed Extraction, Natural Language Processing e Semantic Computing.

Irene Paoli

irene.paoli@unifi.it



Si è laureata in Statistica presso l'Università degli Studi di Firenze. Attualmente è dottoranda di ricerca in Ingegneria dell'Informazione, e lavora presso DISIT lab, DINFO, Università degli Studi di Firenze. Si interessa principalmente di analisi statistiche, modelli predittivi e algoritmi machine Learning. Il suo principale ambito di ricerca presso DISIT lab sono analisi statistiche in ambito social media.

Imad Zaza

imad.zaza@unifi.it



Si è laureato in informatica presso l'Università degli studi di Firenze. Attualmente è dottorando di ricerca in ingegneria dell'informazione ciclo XXX. Lavora presso DISIT lab, DINFO, Università degli studi Firenze, si interessa di analisi di social media, big data, sistemi distribuiti, modellazione di sistemi di interlocking, ontologie. Ha partecipato e/o partecipa ai seguenti progetti internazionali di ricerca: Trace-it, RAISSS, Sii-Mobility, RESOLUTE.