

81

In particolare, si è osservato che soluzioni basate su NB-IoT risultano avere prestazioni migliori rispetto alle altre soluzioni, ma contestualmente anche la soluzione WmBus sperimenta buoni livelli di connettività. Ciò è dovuto al fatto che tale soluzione è stata espressamente sviluppata per il *metering* e quindi in grado di offrire buone prestazioni in termini di copertura anche in situazioni di difficile propagazione. Va ricordato però che per applicazioni differenti dal *metering*, dove si richiede un traffico dati maggiore e requisiti di QoS non trascurabili, come per le applicazioni *automotive*, tale soluzione potrebbe non risultare più appropriata.

Infine, le analisi effettuate hanno confermato che le zone che sperimentano una ridotta connettività sono le zone suburbane o rurali. In tali zone, laddove la connettività dovesse risultare troppo bassa, l'installazione di un nuovo impianto basato su soluzioni proprietarie risulterebbe una soluzione più percorribile rispetto all'installazione di un nuovo sito di un operatore dedicato essenzialmente ad abilitare un traffico M2M/IoT. Ciò è dovuto al fatto che la logica degli operatori TLC per l'installazione di nuovi impianti privilegia zone densamente popolate (zone bianche) e, di conseguenza, non considera zone reputate a fallimento di mercato (zone grigie, nere). Pertanto, laddove la soluzione licenziata potrebbe non apparire una soluzione percorribile, potrebbe essere invece introdotta una soluzione proprietaria i cui costi dell'installazione risultano essere decisamente più contenuti. Ciò vale solo per applicazioni massive e non per le applicazioni "critical", dove il requisito di latenza risulta fondamentale e non ottenibile con soluzioni proprietarie. Le soluzioni su banda licenziata, invece, permettono di raggiungere prestazioni anche in termini di QoS laddove l'applicazione lo richieda. Pertanto, per applicazioni future con stringenti requisiti di ritardo come l'*automotive*, l'evoluzione della tecnologia IoT su reti 5G risulta essere l'unica soluzione percorribile. Tale indicazione viene confermata anche dalle proposte del 3GPP per le future release, dove verranno forniti miglioramenti funzionali proprio per i requisiti di disponibilità *real-time* dell'informazione al fine di dare così le prime indicazioni sulle specifiche dell'Accesso Radio di soluzioni IoT basate sulla rete 5G.

Sviluppi futuri

- Possibili sperimentazioni considerando bande dedicate per applicazioni M2M nella rete 5G
- Valutazioni delle prestazioni a seconda dei requisiti delle differenti applicazioni.

Progetti correlati

- Progetto "RSE QoS - Quality of Service for Smart Grid Services", in convenzione con RSE
- Progetto "Mappatura M2M IOT - SMART CITIES", in convenzione con Agcom (Delibera n. 626/16/CONS, ex Delibera n. 211/16/CONS abrogata).

OUTPUT SCIENTIFICI

PUBBLICAZIONI

Atti di Conferenze Internazionali

- Persia S., Carciofi C., Faccioli M., "NB-IoT and LoRA Connectivity Analysis for M2M/IoT Smart Grid Applications", *AEIT 2017*, Cagliari, Italy, September 2017.

Riviste nazionali

- Persia S., Rea L., "Analisi delle Soluzioni LTE-MTC e NB-IoT per Applicazioni Smart Grid", "L'Elettrotecnica" - Speciale "Focus ICT: tecnologie abilitanti e architetture di rete", aprile 2017.

AZIONI COST

- Persia S., Carciofi C., Faccioli M., “NB-IoT and LoRA Connectivity Analysis for M2M/IoT Smart Grid Applications”, invited to CA15104 Inclusive Radio Communication Networks for 5G and beyond (IRACON), gennaio 2018.

PARTECIPAZIONI A GRUPPI DI NORMATIVA TECNICA, NETWORK NAZIONALI E INTERNAZIONALI

- Tavolo Mise con operatori italiani della LoRA Alliance per discussione esiti tavoli internazionali CEPT, in particolare ETSI WG FM su normativa di riferimento SDR (70-03) e come adeguarla per tecnologia LoRA (ottobre 2017).

ATTIVITÀ DI REVIEWING PER CONVEGNI E RIVISTE

- Reviewing per IEEE Communications Magazine (dicembre 2017).

EVENTI

- Convegno Anfov: “Rapporto ANFoV 2017 sull’IoT infrastrutture, architetture e proposte commerciali” - Sessione Persia S. e Milan A., “Tecnologie di comunicazione dei dati nello Smart metering” (novembre 2017).

LABORATORI E SVILUPPO SOFTWARE

- Collaborazione Unidata presso i loro laboratori per sperimentazione tecnologia proprietaria LoRA.

RETI**PROGRAMMA DI RICERCA
NET NEUTRALITY**

Evoluzione del tema della net neutrality in merito all'accesso ad Internet (fisso e mobile) e in merito ai processi di instradamento dell'informazione.

Finalità

Il principio della tutela della net neutrality in Europa viene fortemente avvalorato dalle linee guida emanate dal BEREC (Body of European Regulators for Electronic Communications) che recepiscono quanto stabilito nel Regolamento Europeo 2015/2120 in merito a:

- pratiche commerciali (*zero rating*)
- trasparenza (sui contratti e sui siti web)
- pratiche di *traffic management*
- qualità del servizio (QoS) in funzione delle tematiche di net neutrality
- *specialized services*, sui servizi di accesso ad Internet, indipendentemente dalla tecnologia di rete usata e dal terminale utilizzato.

Il BEREC dà quindi mandato alle singole Autorità Nazionali di vigilare sul rispetto delle regole di equità e non-discriminazione del traffico nell'accesso ad Internet, a tutela dei diritti degli utenti.

Le linee guida potrebbero sembrare in contrasto con i requisiti stabiliti per la Quality of Services nelle reti di quinta generazione. Infatti, se per ottenere un livello adeguato di qualità dei servizi "garantito" sempre e ovunque, l'ottimizzazione delle risorse di rete in base al servizio appare la strada più immediata, d'altra parte così facendo non verrebbe rispettato il principio della net neutrality.

Il tema è di particolare rilevanza sia per quanto riguarda la rete di accesso, sia per quanto riguarda l'instradamento: un *routing* che si basa sull'informazione trasportata non rispetta, in prima battuta, il principio di neutralità della rete.

In tale contesto si inseriscono i concetti di *specialised services*, per i quali l'ottimizzazione è necessaria per soddisfare i requisiti di qualità, e di *network slicing*, nell'ambito del quale deve essere garantito e preservato il principio di neutralità della rete.

Il Programma di ricerca persegue l'obiettivo di valutare i criteri in base ai quali è possibile verificare il rispetto dei principi di neutralità della rete, avendo come riferimento quanto definito dalle direttive esistenti (BoR (16) 127) e possibili future modifiche.

Tra gli obiettivi del Programma vi è una sperimentazione in campo mirante a studiare le modalità di verifica delle politiche di QoS e di *traffic management*.

Ricadute di politica industriale

Dal quadro Europeo (Regolamento e Linee guida BEREC) emerge un'indicazione chiara nella direzione della tutela della neutralità della rete. Quindi, nella progettazione della rete di quinta generazione, che sarà una rete di reti con forte combinazione e integrazione di diverse tecnologie, è opportuno considerare fin da subito il tema della net neutrality e le specifiche tecniche per la sua tutela.

Attività di ricerca 2017

Nel 2017, l'attività è stata incentrata sullo studio delle Linee Guida e dei documenti tecnici pubblicati nel corso dell'anno (*Net Neutrality Regulatory Assessment Methodology* e *Net Neutrality measurement tool specification*), al fine di comprendere come il Regolamento possa essere recepito mentre si progetta la rete 5G e di studiare le modalità di applicazione della net neutrality sulla rete di accesso fissa e mobile nazionale. Infatti, l'obiettivo di coniugare gli obiettivi prestazionali della rete 5G con il rispetto della net neutrality comporta necessariamente grossi investimenti sulle reti fisse e mobili.

A livello tecnico, la tutela della neutralità della rete passa per la verifica dell'assenza di pratiche di *traffic management* e tramite la misurazione di *Quality of Service*. In particolare, tra specifici contenuti, applicazioni, servizi o categorie di servizi non devono sussistere: *blocking, slowing down, alteration, restriction, interference with, degradation, discrimination*. Allo stesso tempo, devono essere garantiti sulla rete fissa e dichiarati all'utente finale specifici livelli di *network performance*, determinabili tramite la misurazione di: *throughput, latency, packet loss, jitter*.

Tramite la misurazione delle prestazioni della rete di accesso, il Regolatore può valutare i requisiti necessari per consentire la presenza di *specialised services*, verificandone poi il rispetto dei suddetti requisiti. Il BEREC, infatti, contempla la possibilità che servizi specializzati possano avere nella rete dei trattamenti di tipo differente, purché non siano perpetrati a scapito di altri. Quindi, tali *specialised services* sono ammessi se la capacità della rete è sufficiente a fornirli in aggiunta a tutti i servizi di accesso a Internet.

Il giusto compromesso potrebbe essere spostato nella definizione di regole di neutralità sulle famiglie di servizi, verificando e salvaguardando i principi di *fairness*. In altre parole, potrebbe essere opportuno inserire le regole di neutralità all'interno del *network slicing*, o più in generale nell'ambito dei singoli *verticals*.

A livello di studio, comparare le misure di *network performance* in presenza e in assenza di *specialised services* su due o più linee con medesime condizioni fisiche (ad es. medesima distanza dal DSLAM) e medesimi *end-point* (destinatario e sorgente) permette di avere un riscontro sulla verifica della variazione della QoS.

Allo stesso modo, si è visto come la comparazione possa essere fatta sia considerando un servizio, o un'applicazione, misurandone la qualità su linee differenti con le stesse condizioni fisiche, sia misurando la qualità di una categoria di servizi su una medesima linea.

Con riferimento alla qualità dei servizi, il BEREC considera il *web browsing* e il *video streaming*. Per tali misure, siano esse di *network performance* o di qualità dei servizi, si è fatto riferimento allo standard ETSI ES 202 765 parte 4.

Quindi è stata progettata e sviluppata un'applicazione che permettesse la misura della qualità della rete di accesso a partire da terminali mobili, la misura della qualità del servizio di *web browsing*, misurando il tempo per scaricare la pagina ETSI standard di riferimento (utile per individuare eventuali *slowing down*), e la misura della qualità di un servizio video, tanto su una rete fissa con collegamento in wi-fi quanto su una rete mobile.

Lo scopo dello studio è stato di comprendere quale tipologia di architettura permetta una misurazione sia puntuale che comparativa e come configurare i test opportuni per raggiungere un risultato statisticamente valido.

In futuro lo studio dovrebbe essere incentrato sulla misurazione delle prestazioni della rete di accesso e della qualità dei servizi, in assenza e in presenza del *network slicing*, dando forte rilievo anche al parametro di affidabilità del servizio.

Sviluppi futuri

Tra i possibili sviluppi futuri, oltre all'individuazione delle metriche che le singole *network slice* dovranno rispettare, vi è la possibile individuazione di tecniche di istradamento e algoritmi adattativi che permettono un uso efficiente delle risorse di rete e garantiscono un trattamento *fair* dei servizi che appartengono alla stessa famiglia.

Progetti correlati

- Progetto “Misura Internet” (Delibera Agcom 244/08/CSP e s.m.i.).
- Progetto “Misura Interne Mobile” (Delibera Agcom 154/12/CONS e 580/15/CONS).

PUBBLICAZIONI

Riviste Internazionali

- Costantini L., Mammi E., Teodori M., Attanasio V., “Polynomial regression models to explain the relationship between network and service key performance indicators”, *IET Networks*, Vol. 6, Issue 5, September 2017, pp. 125-132.

PARTECIPAZIONI A GRUPPI DI NORMATIVA TECNICA, NETWORK NAZIONALI E INTERNAZIONALI

- Partecipazione Comitato Tecnico Scientifico del NaMeX.

LABORATORI E SVILUPPO SOFTWARE

- Sperimentazioni nell'ambito del progetto 5GEzE platform.
- Applicazione android per la misurazione di *network performance* e qualità dei servizi.
- Strumento web per la misurazione di *network performance*.

TESI DI LAUREA

- Marini F., “Valutazione della Quality of Service per la trasmissione di servizi su reti 5G”.
- Calderone D., “Realizzazione interfaccia Ne.Me.Sys con React”.
- Mathew J.G., “Realizzazione di uno strumento web per la misura della qualità di accesso ad Internet”.

86

ATTIVITÀ FUB 2017

DATI**PROGRAMMA DI RICERCA****BIG DATA**

Studio di tecnologie abilitanti e metodologie di Data Analytics per grandi volumi di dati in ambiente di calcolo distribuito.

Finalità

La Fondazione possiede un'infrastruttura tecnologica Software sui Big Data acquisita mediante i Progetti in convenzione con ISCOM ("TV++", "SNOOPT", "BigDOT"), e con Almagest. L'esperienza è stata maturata nel corso degli anni attraverso una rilevante attività di ricerca metodologica sui Big Data, diverse collaborazioni con il mondo universitario, scientifico nazionale e internazionale (IASI-CNR, Glasgow University, Yahoo!Labs, Università di Lugano, Università di Tor Vergata) e la partecipazione ai Comitati di Programma e Tecnici delle più rilevanti Conferenze Internazionali (ACM SIGIR, ACM CIKM, ACM ICTIR, ICWSM, ECIR, IIR, IPM, DART). Le attività di ricerca si avvalgono dell'infrastruttura hardware dell'ISCOM costituita da due cluster di 8 macchine, ciascuno con una memoria complessiva di 320GB.

Le attività di ricerca e sviluppo hanno riguardato:

- L'integrazione nella Piattaforma in uso di metodologie e tecnologie abilitanti di Data Analytics in ambiente di programmazione di tipo MapReduce, quali Spark Python e SparkR (Progetto "BigDOT").
- Analisi di scalabilità e comparazione delle piattaforme abilitanti di Big Data in modalità batch e streaming (Progetto "BigDOT").
- Clustering massivo di dati e ricerca di oggetti simili molto complessi, ad esempio usando metodologie del tipo *Locality Sensitive Hashing* (LSH) che permettono tempi lineari di elaborazione rispetto al volume dei dati. In particolare, è stata sviluppata una tecnica basata su LSH utilizzata per la classificazione massiva di profili e di documenti delle reti sociali. Tale tecnica permette di costruire delle rappresentazioni compatte (firme digitali) dei documenti e degli oggetti complessi, conservando però la proprietà di similarità tra gli oggetti. L'implementazione è stata realizzata su Hadoop e Spark (Progetto "BigDataDOCS").

Le attività svolte sono state:

- Uso di tecniche di clustering massivo sui nodi di una rete in modo tale da rendere scalare la visualizzazione dell'intera rete su porzioni omogenee della stessa.
- Scoperta, analisi e visualizzazione di cluster di nodi affini per contenuto o per flussi di interazione, in una rete di comunicazione.
- Individuazione delle spam-farm e bot-farm in Twitter.

Tali attività hanno avuto come output un sistema integrato di filtraggio, indicizzazione, consultazione, classificazione e navigazione delle reti sociali, e per grandi collezioni.

Ricadute di politica industriale

Secondo il rapporto dell'ultimo World Economic Forum, le tecnologie dei Big Data saranno *il primo motore del cambiamento* dell'ICT, a sua volta al centro della quarta rivoluzione tecnologica e industriale dei prossimi anni (http://www3.weforum.org/docs/WEF_Future_of_Jobs.pdf). Secondo il rapporto Assinform del 2017, il mercato ICT si baserà sullo sviluppo del mobile, dei social, dell'economia della condivisione e dell'IoT. Nei prossimi anni, il 30% degli investimenti delle grandi organizzazioni riguarderanno le applicazioni dei Big Data a tutta la "filiera" del dato (dalla creazione alla fruizione).

Le grandi organizzazioni, sia pubbliche sia private, dovranno sfruttare tutte le possibilità offerte dall'utilizzo delle informazioni estratte dai Big Data, rinnovare i processi e favorire l'innovazione e la crescita economica del Paese. Questi studi trovano campi di applicazione già affermati, per esempio negli Stati Uniti e in Gran Bretagna, per l'elaborazione dei dati delle reti energetiche e di telecomunicazione, dati fiscali, dati medici, dati legali, dati provenienti dalle azioni di contrasto alle frodi e di rilevamento real-time di eventi critici.

Attività di ricerca 2017

Uso e analisi della metodologia LSH. L'applicazione delle tecniche di LSH riguarda una vastissima classe di applicazioni. Ad esempio Shazam utilizza LSH per il recupero dei brani musicali, Google per la tecnica di aggregazione dei risultati simili, i sistemi di recupero delle impronte digitali utilizzano LSH. Abbiamo inoltre dimostrato che questa può essere utilizzata come tecnica di calcolo del numero di elementi distinti in un flusso informativo (quanti sono gli utenti distinti che stanno postando su Facebook o su Twitter in questo momento?). Oltre a essere estremamente efficiente a fornire il clustering massivo dei dati, LSH è dunque una tecnica utilissima per indicizzare e recuperare rapidamente dati molto complessi ossia dati che possiedono tante dimensioni descrittive degli oggetti (features). LSH dunque può trovare innumerevoli applicazioni per l'IoT perché è capace di ridurre la descrizione di oggetti complessi in modo estremamente sintetico, pur mantenendo la semplicità del recupero e il grado di similarità con gli altri oggetti.

Applicazioni. Il metodo LSH è stato applicato al calcolo del diametro di grandi grafi. Si è calcolata l'impronta digitale del vicinato dei nodi ad arbitraria distanza e si è calcolato per ciascun nodo il grado di sovrapposizione di ogni vicinato con la dimensione del grafo intero. Grazie a questa stima si è potuto stimare con grandissima precisione il numero totale di coppie raggiungibili in un grafo (ad esempio 649 miliardi di coppie per il grafo di Amazon).

Clustering di Malware. L'obiettivo è quello di individuare cluster di malware mediante condivisione di pattern comuni anche a basso livello (istruzioni a livello di codice negli script o estraendo le stringhe). Quest'analisi preliminare ha lo scopo di classificare i malware affini utile per una pronta individuazione della tipologia di malware (Progetto "BigDataDOCS").

Sviluppi futuri

La selezione di un modello appropriato per l'analisi dei dati su larga scala è un fattore di criticità. L'estrazione di informazioni utili da grandi quantità di dati richiede l'uso di algoritmi di analisi facilmente scalabili al fine di produrre risultati tempestivi o in tempo reale. Tuttavia, gli algoritmi attuali sono inefficienti per i Big Data, molti non si prestano a essere implementabili in modalità di programmazione distribuita di tipo MapReduce. Pertanto, le tecnologie di analisi dei Big Data, seppur necessarie per permettere la scalabilità degli algoritmi standard, non sono sufficienti a garantirne l'implementazione o l'efficienza. La velocità del flusso di dati provenienti da diverse fonti di dati e il confronto con le informazioni storiche richiederanno una forte tempestività nella risposta.

Secondo il rapporto Gartner 2017, relativo all'evoluzione delle piattaforme di Big Data, presto ogni organizzazione sarà in grado di accedere a piattaforme di Big Data o in modalità self-service e user-friendly o attraverso una propria piattaforma di Data Analytics. *Il vantaggio tecnologico per i fornitori di piattaforma di Big Data e per gli utenti stessi sarà pertanto la possibilità di individuare e definire modelli altamente performanti.* Ciò richiede la capacità dei ricercatori di risolvere le questioni legate all'efficienza degli algoritmi e alla qualità degli stessi, ma conservando una *visione predittiva e la padronanza dell'uso di piattaforme avanzate* e open source di Cloud Computing e di Data Analytics.

In quest'ottica il Programma di ricerca focalizza la propria attenzione sia sull'acquisizione e l'aggiornamento delle piattaforme in termini di scalabilità e di efficacia delle librerie, sia sulla semplificazione ed efficienza degli algoritmi dedicati ai dati su larga scala.

Progetti correlati

- “BigData Docs”
- “BigDot”

OUTPUT SCIENTIFICI

PUBBLICAZIONI

Atti di Conferenze Internazionali

- Amati G.B., Angelini S., Gambosi G., Rossi G., Vocca P., “Estimation of distance-based metrics for very large graphs with MinHash Signatures”, *IEEE International Conference on Big Data (IEEE BigData 2017)*, 11-14 December 2017, Boston, MA, USA.

Riviste internazionali

- Amati G.B., Angelini S., Gambosi G., Rossi G., Vocca P., “The analysis of influential users evolution in microblogging social networks”, to appear in Volume 233 of the *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* series.

DATI**PROGRAMMA DI RICERCA****WEB INTELLIGENCE**

Tecnologie intelligenti per l'analisi e lo sviluppo di sistemi e servizi basati sul web..

Finalità

La Web Intelligence si occupa dell'analisi e della creazione di sistemi, servizi e processi innovativi basati sul web, mediante l'utilizzo e l'integrazione di tre tecnologie principali: intelligenza artificiale, scienza delle informazioni e scienza delle reti. La Fondazione ha una consolidata esperienza in questi settori nei quali ha fornito, anche più recentemente, rilevanti contributi scientifici, in particolare per quanto riguarda Web mining, analisi dei grafi, motori di ricerca semantici, personalizzazione e privacy online. L'obiettivo generale di questa scheda è di mantenere e accrescere le competenze acquisite nelle tecnologie intelligenti per il web, sia per lo svolgimento ottimale delle attività previste in progetti già avviati o in fase di avvio (segnatamente quelle riguardanti anticontraffazione e analisi offerte online), sia per l'acquisizione di nuove commesse, sia a fini di ricerca scientifica. Nel 2017 le ricerche, descritte più dettagliatamente nel seguito e svolte in parte con l'Università La Sapienza, si sono concentrate sui seguenti argomenti:

- Rilevazione e misurazione automatica di siti di commercio elettronico contraffattori
- Indici di privacy per user generated content
- Riconoscimento automatico degli argomenti principali ("topic") di un testo.

Ricadute di politica industriale

La gamma di applicazioni della Web Intelligence è assai ampia e spesso di diretto interesse sia per il mondo industriale sia per la Pubblica Amministrazione. Essa pertanto trova spazio in progetti di ricerca nazionali e internazionali. Si possono citare, fra le altre: linked open data, identità digitale, anticontraffazione online, antiterrorismo online, e-commerce, pubblicità computazionale, Web privacy, data privacy, monitoraggio del brand. Oltre alle specifiche applicazioni, la Web Intelligence può essere vista in generale come una tecnologia abilitante per accrescere il valore commerciale e sociale di un generico bene o servizio online.

Attività di ricerca 2017**Rilevazione e misurazione automatica di siti di commercio elettronico contraffattori**

È stata sviluppata una metodologia innovativa per l'apprendimento automatico di siti contraffattori presenti nei risultati delle ricerche web relative a marchi commerciali. La metodologia è completa-

mente automatica e consiste di tre stadi principali in cascata : (1) interrogazione dei motori di ricerca con chiavi contenenti il nome dei marchi e acquisizione dei risultati, (2) separazione dei risultati in siti di commercio elettronico o di altro tipo, (3) valutazione della fidatezza (nel senso di rispetto delle norme sul copyright) dei siti di commercio elettronico identificati al passo precedente. Le fasi (2) e (3) vengono eseguite mediante due classificatori distinti addestrati con una serie di indicatori specifici estratti automaticamente dai siti web da valutare. La metodologia, implementata in un sistema denominato RI.SI.CO. (Ricerca Siti Contraffattori), ha evidenziato un'accuratezza predittiva elevata sia nel riconoscimento di siti di commercio elettronico sia nella valutazione della loro fidatezza. Utilizzando gli output di RI.SI.CO. è possibile costruire mappe di contraffazione che consentono di valutare il tasso di contraffazione dei singoli marchi relativamente a quelli concorrenti, in un medesimo settore merceologico. Il sistema è stato successivamente utilizzato per condurre una prima valutazione sperimentale su larga scala della fidatezza dei risultati dei motori di ricerca relativamente al settore delle calzature, dalla quale è emerso che una frazione significativa dei siti reperiti (circa il 15%) è illegale [1]. È stata inoltre svolta una valutazione comparata delle prestazioni di RI.SI.CO. e di quelle di sistemi alternativi per misurare la fidatezza dei siti di commercio elettronico [2]. Il sistema RI.SI.CO. si è caratterizzato per un bilanciamento ottimale di precisione e copertura.

Indici di privacy per user generated content

È stato definito e parzialmente implementato e valutato un indice globale di privacy per dati testuali anonimizzati, con particolare riferimento ai log delle interrogazioni inviate ai motori di ricerca [3]. L'indice si basa sull'integrazione di modelli statistici che misurano la riduzione del contenuto informativo dei testi e di tecniche semantiche di riconoscimento automatico delle informazioni sensibili contenute nei testi stessi. L'indice proposto può essere applicato, in linea di principio, ad altri tipi di dati testuali presenti sul web, quali ad esempio i post sulle reti sociali, le cartelle cliniche e i messaggi di posta elettronica.

Riconoscimento automatico degli argomenti principali ("topic") di un testo

Il metodo sviluppato accetta come input un testo qualsiasi (ad esempio una pagina web o un articolo di giornale) e restituisce gli articoli di Wikipedia che sono globalmente pertinenti al testo stesso [4]. Mentre le tecniche esistenti per l'identificazione delle topic sono prevalentemente locali ed estrattive - cioè lavorano a livello di singole frasi e richiedono un "matching" quasi esatto fra le sequenze di parole contenute nel testo di input e quelle degli articoli Wikipedia - il metodo in oggetto è globale ed astrattivo. Per eseguire la mappatura testo-topic, il testo di input viene convertito in un'interrogazione a concetti idonea ad essere elaborata dal motore di ricerca Wikipedia, i cui risultati vengono successivamente filtrati utilizzando una serie di indicatori semantici presenti nella rappresentazione Wikipedia dei risultati stessi. Questa attività è stata svolta in collaborazione con il gruppo di ricerca del Prof. Roberto Navigli de La Sapienza.

Sviluppi futuri

Con questo Programma di ricerca, FUB si propone di diventare un punto di riferimento nazionale per le attività di analisi, monitoraggio e contrasto alla manipolazione illecita dei contenuti e delle informazioni personali sul web, con particolare riferimento ad anticontraffazione, antipirateria, tutela del diritto d'autore e rispetto della privacy.

Progetti correlati

- Banda 700 - Convenzione Mises: analisi di conformità di offerta televisori online.
- UIBM-ICI: rilevazione automatica di siti di commercio elettronico contraffattori.

OUTPUT SCIENTIFICI

PUBBLICAZIONI

Atti di Conferenze Internazionali

- Carpineto C., Romano G. (2017), "Learning to detect and measure fake ecommerce websites in search-engine results", Proceedings of 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI'17), Leipzig, Germany, 23-26 August 2017, pp. 403-410.
- Carpineto C., Romano G., Lo Re D. (2017), "Using information retrieval to evaluate trustworthiness assessment of eshops", Proceedings of the 8th Italian Information Retrieval Workshop (IIR'17), Lugano, Switzerland, 5-7 June 2017, pp. 1-8.
- Carpineto C., Romano G., "A semantic privacy index for text data", (submitted).
- Carpineto C., Navigli R., Romano G., "Abstractive Global Text Wikification through Empowered Information Retrieval", (submitted).

PRESIDENZA E PARTECIPAZIONI A COMITATI DI PROGRAMMA DI CONFERENZE

- 2017 IEEE/WIC/ACM International Conference on Web Intelligence (WI'17), Leipzig, Germany, 23-26 August, 2017.
- 39th European Conference on Information Retrieval (ECIR 2017), Aberdeen, Scotland, UK, 9-13 April, 2017.
- 14th International Conference on Formal Concept Analysis (ICFCA 2017), Rennes, France, 12-16 June, 2017.
- 8th Italian Information Retrieval Workshop (IIR 2017), Lugano, Switzerland, 5-7 June, 2017.

ATTIVITÀ DI REVIEWING PER CONVEGNI E RIVISTE

- Artificial Intelligence Journal
- ACM Transactions on the Web (TWEB)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- Journal of the American Society for Information Science and Technology (JASIST)

LABORATORI E SVILUPPO SOFTWARE

- Implementazione del sistema RI.SI.CO., disponibile come applicazione web protetta da password.

DATI**PROGRAMMA DI RICERCA****MULTIMEDIA SYSTEMS & APPLICATIONS**

Sviluppo su architetture di calcolo contemporanee di metodi efficienti di classificazione e individuazione di immagini. Progetto di sistemi web e Mobile multimediali per servizi informativi interattivi, basati su diverse rappresentazioni dei dati.

Finalità

Le città contemporanee sono popolate da un crescente numero di sistemi intrinsecamente multimediali (wearable device, sensor networks, home robot, on board computer), che richiedono tecnologie intelligenti e flessibili per la comprensione, catalogazione, visualizzazione e diffusione dei contenuti. In tale contesto, le comunità e i singoli utenti sono coinvolti in molteplici esperienze multimediali, multimodali e multi-dispositivo, che generano grandi quantità di dati multimediali e richiedono sistemi in grado di accedere successivamente agli stessi dati per la loro gestione, uso e riuso.

La necessità di sistemi intelligenti per l'elaborazione di big multimedia data sta motivando l'interesse della comunità scientifica verso le metodologie di *object detection* e *pattern recognition* basate sull'analisi statistica delle features di basso livello mediante tecniche di apprendimento automatico supervisionate e non supervisionate. Particolare attenzione è stata prestata alle reti neurali convoluzionali (CNN) e al deep neural learning (DNL), tecnologie che si sono affermate recentemente grazie al progressivo potenziamento delle GPU per l'elaborazione dei dati. Queste tecniche si rivelano un potente strumento per il riconoscimento di oggetti a sostegno dei processi decisionali complessi, effettuati dai sistemi di intelligenza artificiale.

Le attività sul Multimedia svolte dalla FUB nel corso del 2017 hanno teso ad approfondire e rafforzare le competenze sui temi della progettazione e sviluppo di sistemi intelligenti per l'analisi, la gestione e l'elaborazione del dato multimediale. L'attività di studio e approfondimento scientifico, in particolare, si è focalizzata sulle seguenti tematiche:

- analisi, segmentazione, rappresentazione e indicizzazione di informazioni multimediali;
- riconoscimento e classificazione di pattern audiovisivi all'interno di dati multimediali grezzi sulla base del loro contenuto semantico;
- metodologie di deep neural learning (DNL) per l'analisi di rappresentazioni complesse di specifici pattern visuali al fine di migliorare le prestazioni dei classificatori poco profondi, come le SVM;
- studio e sperimentazione delle tecniche di progettazione di sistemi multimediali interattivi valutati con l'utente (ad esempio, applicazioni multimediali interattive personalizzabili rispetto alle attività e ai servizi richiesti dall'utente, sistemi di navigazione di cataloghi museali, sistemi di realtà virtuale e di intrattenimento).

Ricadute di politica industriale

Le applicazioni multimediali interattive sono attualmente uno dei driver che maggiormente muovono la crescita del mercato tecnologico, spaziando dall'entertainment, all'education, alla domotica. Le tecniche di *object detection* e *pattern recognition* hanno molteplici campi di applicazione, dal riconoscimento del pedone nei sistemi di guida assistita alla diagnostica medica, dal riconoscimento biometrico per il controllo di accesso alle aree protette ai sistemi di sicurezza contro atti di teppismo e terrorismo ovvero ai sistemi di sorveglianza per il contrasto delle attività illecite.

Attività di ricerca 2017

Riconoscimento del logo TV basato sull'algoritmo nearest neighbors per la classificazione dei video.

La ricerca e identificazione di specifici pattern e figure in un'immagine o video è una procedura critica in una vasta gamma di applicazioni di visione artificiale quali l'identificazione biometrica, il riconoscimento di oggetti o forme in un archivio, il monitoraggio di aree sensibili. La ricerca esaustiva della patch e la sua segmentazione dal background tramite la scansione progressiva dell'intera immagine è un'operazione computazionalmente dispendiosa che può richiedere anche diversi minuti. Tuttavia i tempi di esecuzione possono essere ridotti significativamente se si codifica il pattern da ricercare come punto di uno spazio ad alta dimensione e si esegue la ricerca sulle patch in cui è stata decomposta l'immagine seguendo l'approccio *nearest neighbourhood*.

Nell'ambito di questo Programma di ricerca, con riferimento alla problematica della tutela dei diritti di proprietà intellettuale, il problema dell'image recognition è stato contestualizzato all'individuazione e classificazione dei loghi dei broadcaster TV nei video condivisi sulle piattaforme di video sharing. A questo riguardo l'attività svolta ha riguardato:

- la definizione di uno specifico crawler per l'esplorazione e il download dei keyframes dei video presenti sulle piattaforme di video sharing;
- la definizione e implementazione di una procedura per la localizzazione delle regioni contenenti il logo TV sulla base della sua persistenza temporale su più keyframes;
- la definizione e implementazione di una metrica per la classificazione del logo in base alla sua caratterizzazione con i punti salienti nel dominio wavelet di Laguerre Gauss secondo l'algoritmo Nearest neighbourhood;
- approfondimento delle tecniche di *visual deep learning* per l'image recognition con riferimento alle *Convolutional Neural Network* (CNN) addestrate seguendo l'approccio del *transfer learning*.

Machine learning per l'audio understanding

Nell'ambito dello studio dei campi applicativi delle tecniche di machine learning, è proseguita l'attività collegata al Progetto "Octave" per la classificazione dei rumori ambientali. L'obiettivo è stato quello di definire e implementare una procedura per l'identificazione dell'impronta audio generata dai droni ai fini del rilevamento degli aereomobili a pilotaggio remoto (UAV) nelle aree protette. Le tecniche di estrazione e aggregazione delle features temporali e spettrali dei segmenti audio sono state specializzate e ottimizzate per migliorare le prestazioni del sistema di classificazione basato su un set di SVM di tipo RBF opportunamente addestrate.

La valutazione delle prestazioni del sistema di identificazione dell'impronta di rumore dei droni è stata effettuata considerando un dataset contenente cinque diverse tipologie di rumori ambientali (drone flying, nature daytime, street with traffic, train passing, crowd) estratti dal web implementando uno specifico scraper audio.

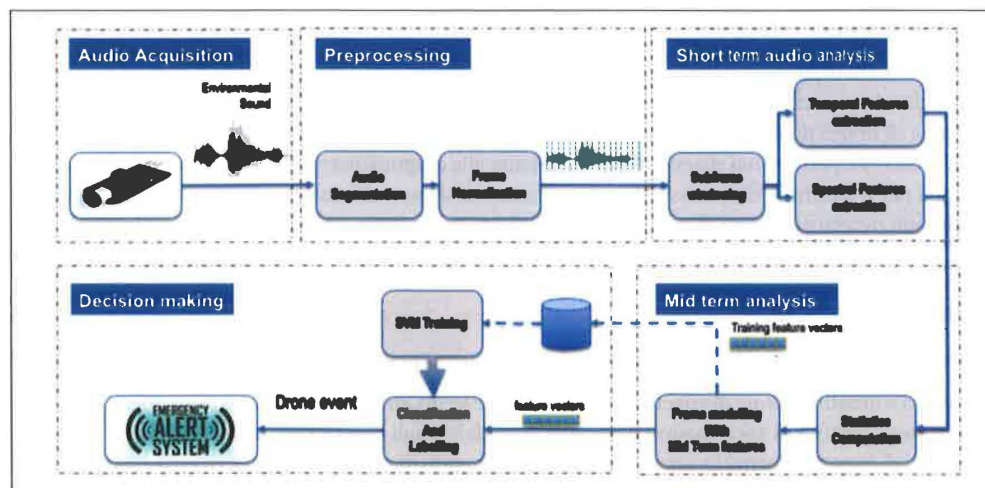


Figura 1. Framework per l'identificazione dell'impronta di rumore dei droni.

Cascading classifiers per l'estrazione di informazioni testuali da immagini naturali

Studi recenti hanno evidenziato come le nuove frontiere del commercio illegale si siano spostate dall'utilizzo di domini web tradizionali all'uso di grandi raccolte di immagini, collocate in gradi portali di condivisione online, in spazi di pubblico accesso o parzialmente riservati. All'interno delle immagini vengono solitamente inseriti i dati di contatto del venditore attraverso livelli di steganografia via via più complessi. Nella casistica individuata è ricorrente l'inserimento all'interno delle immagini di qrcode e di testi con i social contacts o i link ai siti del web-shop. Si tratta di una metodologia per rendere difficile il tracciamento del venditore e la quantificazione della dimensione dei panieri di prodotti in vendita.

Con riferimento a questi nuovi modelli di utilizzo delle immagini per la promozione di prodotti contraffatti in violazione del diritto di autore, il lavoro svolto nel Programma di ricerca si è focalizzato su:

- analisi e sperimentazione di descrittori locali delle features strutturali di un'immagine (LBP, IIAAR, HOG, BRISK ecc.) in grado di essere discriminativi, robusti, compatti e veloci da computare per il supporto del riconoscimento di oggetti, la registrazione di immagini e la ricerca nei database utilizzando anche dispositivi mobili;
- definizione e sperimentazione in python di procedure automatiche per la localizzazione e decodifica in tempo reale delle regioni delle immagini contenenti informazioni testuali e QR code secondo una strategia di classificazione supervisionata di tipo boosted basata sull'architettura di Viola Jones (cascading classifiers);
- addestramento e validazione dei classificatori su dataset opportunamente costruiti tenendo conto dello specifico problema applicativo in esame (QR code e testi di diversa dimensione sovrapposti ad immagini naturali in modo automatico senza il controllo del background);
- definizione e implementazione di un sistema di crawling per il download automatico di cataloghi di immagini dai portali di photo sharing (Instagram e Yupoo) per la creazione dei dataset di training. Nel dettaglio questa specifica attività ha prodotto come output:
 - Libreria software per l'interfacciamento automatico con Yupoo
 - Libreria software per lo scraping di immagini da un catalogo online
 - Realizzazione di dataset di immagini usate per veicolare informazioni a livello diverso di complessità e granularità, organizzate come segue:
 - dataset 1: 30K immagini, dimensione totale 3.5 GB
 - dataset 2: 10K immagini, dimensione totale 750 MB

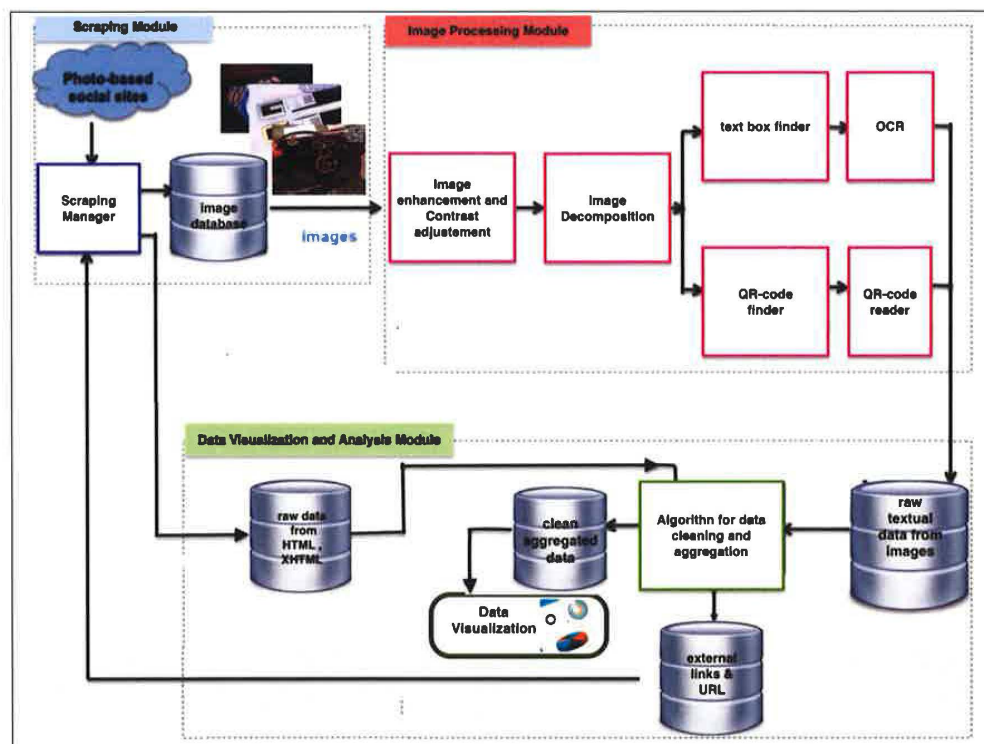


Figura 2. Framework per il riconoscimento delle informazioni testuali dalle immagini

- dataset 3: 2K immagini, dimensione totale 320 MB
- dataset 4: 250 immagini, dimensione totale 23 MB
- dataset 5: 270 immagini, dimensione totale 40 MB

Sistemi multimediali su dispositivi mobili

Con riferimento al Progetto “APP iVittoriale” sono state aggiornate e consolidate le competenze necessarie alla progettazione e sviluppo di applicazioni e sistemi multimediali per dispositivi mobili con l’obiettivo di supportare servizi informativi interattivi per la valorizzazione dei Beni Culturali. Nello specifico sono state esplorate le peculiarità che caratterizzano e distinguono l’ambito mobile rispetto a quello desktop, per quanto concerne il supporto alla comunicazione, e le modalità di rappresentazione dei dati al fine di identificare le soluzioni d’interfaccia utente (layout, view, menu, activity, XML, navigation bar, ecc) più appropriate per l’implementazione di applicazioni e servizi usabili, interattivi e multimodali nell’attuale contesto di mercato. L’obiettivo è stato di supportare la distribuzione e la fruizione dei contenuti enfatizzando i diversi contesti informativi (testo, immagini fisse, time-lapse, video, time ecc.), i nessi e la personalizzazione dei percorsi. Questo è stato reso possibile dall’impiego nella progettazione delle applicazioni di indici lineari e di mappa che offrono una lettura dinamica e articolata delle informazioni in base alle specifiche esigenze e preferenze dell’utente.

Sviluppi futuri

La penetrazione globale di Internet e la rapida diffusione dei dispositivi mobili, nonché la proliferazione di telecamere di sorveglianza hanno portato ad una crescita esponenziale delle quantità di dati multimediali. Questo crea grandi opportunità e sfide per i sistemi di analisi automatica dei dati finalizzati all’individuazione di conoscenze utili. Fornire soluzioni per processare multimedia big data per comprendere il loro significato semantico e identificare specifici pattern visuali presenta un grado di difficoltà elevato. Nei prossimi anni, lo sviluppo di queste tematiche affronterà l’integrazione delle tecniche di *machine learning* con i metodi tradizionali di visione artificiale per risolvere il problema

96

ATTIVITÀ FUB 2017

della comprensione delle scene con elaborazioni in tempo reale. Sforzi significativi saranno rivolti a migliorare l'estrazione delle features, la loro fusione, sincronizzazione e correlazione per migliorare le prestazioni dei sistemi di data mining multimediale in specifici campi applicativi.

Progetti correlati

Le attività svolte nell'ambito del Programma di Ricerca sono state sviluppate con riferimento ai seguenti Progetti attivi nel 2017:

- Horizon 2020 "OCTAVE -Objective Control for TAlker VErification";
- Fondazione il Vittoriale - "APP iVittoriale" - Applicazione multimediale interattiva per dispositivi mobili per la valorizzazione e fruizione dei beni culturali della Fondazione il Vittoriale degli italiani.

Proposte di progetto

Nel corso dell'anno, i ricercatori afferenti al Programma di ricerca hanno contribuito alla formalizzazione della proposta di Progetto europeo "Private Europe - Reinterpretations of the past through amateur audio-visual heritage" (CALL: H2020-SC6-CULT-COOP-2017-two-stage), e costruito un contesto di studi di fattibilità che hanno portato ai Progetti per i Beni Culturali "iVittoriale" (2017) e "iAventino2" (2018).

OUTPUT SCIENTIFICI

PUBBLICAZIONI

- Mangiatordi F., Bernardini A., Pallotti E., Capodiferro L., "Brand detection framework in LG wavelet domain", *Electronic Imaging Symposium 2017*, Image Processing: Algorithms and Systems XV, Society for Imaging Science and Technology, Burlingame, California, USA, 29 January - 2 February 2017, pp. 10-15, DOI: <https://doi.org/10.2352/ISSN.2470-1173.2017.13.IPAS-199>.
- Bernardini A., Mangiatordi F., Pallotti E., Capodiferro L., "Drone detection by acoustic signature identification", *Electronic Imaging Symposium 2017*, Imaging and Multimedia Analytics in a Web and Mobile World 2017, Society for Imaging Science and Technology, Burlingame, California, USA, 29 January - 2 February 2017, pp. 60-6, DOI: <https://doi.org/10.2352/ISSN.2470-1173.2017.10.IMAWM-168>.
- Mangiatordi F., Bernardini A., Pallotti E., Capodiferro L., "Multimedia analytics platform for profiling keywords embedded in photo catalogues", accettato per la pubblicazione a *Electronic Imaging Symposium 2018*, Imaging and Multimedia Analytics in a Web and Mobile World 2018, Society for Imaging Science and Technology, 28 January - 2 February 2018, Burlingame, California USA, http://www.imaging.org/site/IST/IST/Conferences/EI/EL_2018/Conference/C_IMAWM.aspx

AGGIORNAMENTO PROFESSIONALE

I ricercatori coinvolti nel Programma di ricerca hanno aggiornato le proprie competenze partecipando a seminari e corsi di elearning. Nel dettaglio :

- Bernardini : **OSINT (Open Source Intelligence) retrieval and manipulation training** at NCFTA(National Cyber-Forensics & Training Alliance) in Pittsburgh, USA, Marzo 2017.
- Mangiatordi: **Computer Vision A-Z™: Learn OpenCV, GANs and Cutting Edge AI** by Rajeev Ratan from the University of Edinburgh on Udemy. Data certificazione novembre 2017.
- Mangiatordi: **Computer Vision Intro™ OpenCV3 in Python & Machine Learning** by Hadelin de Ponteves, Kirill Eremenko, SuperDataScience Team on Udemy. Data certificazione novembre 2017.
- Mangiatordi e Pallotti: **Executive Data Science Specialization** by John Hopkins University On Coursera. Data certificazione ottobre 2017.