



Università degli Studi di Milano– Bicocca

Dipartimento di Informatica, Sistemistica e Comunicazione

LABORATORI DEL DIPARTIMENTO DI INFORMATICA, SISTEMISTICA E COMUNICAZIONE

ESSERE

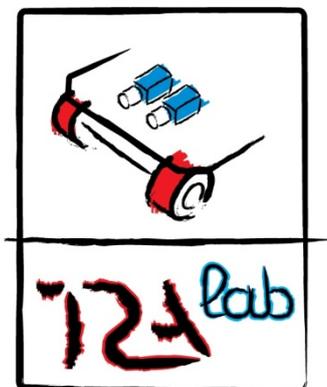
Evoluzione dei Sistemi Software e Reverse Engineering

Responsabile: Francesca Arcelli Fontana

www.essere.disco.unimib.it

Le attività di ricerca nel periodo 2011-2013 hanno riguardato strumenti e tecniche per la reverse engineering come supporto alla software maintenance, software evolution, e program comprehension, e in particolare: 1) Il riconoscimento dei design pattern nel codice come supporto alla comprensione di un sistema, delle scelte progettuali e della sua evoluzione. 2) La definizione e riconoscimento di micro strutture nel codice come indizi di buona o cattiva programmazione o progettazione. 3) L'utilizzo delle micro strutture per raffinare i risultati prodotti da tool sviluppati per il riconoscimento di design pattern nel codice. 4) Riconoscimento di code smell nel codice come sintomi di problemi da rimuovere attraverso opportune tecniche di refactoring. 5) Prioritizzazione dei code smell da rimuovere nel codice in base alla loro gravità, misurata attraverso il loro impatto su numerose metriche di qualità del software.

La ricerca si qualificherà nei prossimi anni attraverso partecipazione a progetti di ricerca regionali, europei o internazionali; pubblicazioni sulle principali riviste e conferenze internazionali, collaborazioni con l'industria sui seguenti ambiti di ricerca: a) reverse engineering, b) software quality assessment, c) mantenimento del software.



IRA Lab, Informatics and Robotics for Automation Laboratoires

Responsabile: Domenico Giorgio Sorrenti

www.ira.disco.unimib.it

L'attività di ricerca nel periodo 2011-2013 ha riguardato la modellizzazione 3D di ambienti basata su camere (VSLAM) e altri sensori e la navigazione di veicoli autonomi in mappa nota, per guida autonoma, incluso lo sviluppo di componenti elettronici e meccanici. Per quanto riguarda VSLAM l'attività di ricerca ha riguardato la parametrizzazione di punti 3D per consentire l'inizializzazione senza ritardo del filtro che ne effettua la stima. Per quanto riguarda la navigazione per guida autonoma urbana importante attività è stata la localizzazione 6DoF in mappe 3D, al fine di arrivare ad una stima della posa più accurata di quella ottenibile con un sistema 3DoF.

La ricerca si qualificherà nei prossimi anni attraverso partecipazione a progetti di ricerca regionali, europei o internazionali; pubblicazioni sulle principali riviste e conferenze internazionali, collaborazioni con l'industria nell'ambito della Machine perception, con particolare attenzione alle tematiche di a) guida autonoma su strada urbana b) guida autonoma in ambito indoor.

Responsabile: Raimondo Schettini

www.ivl.disco.unimib.it

Le principali attività di ricerca del periodo 2011-2013 hanno riguardato la definizione e sviluppo di 1) algoritmi di elaborazione e analisi di immagini per prodotti di largo consumo, quali telefoni cellulari, fotocamere, webcam, monitor, stampanti a colori; 2) algoritmi, metodi, modelli e sistemi per la gestione e riproduzione del colore; 3) algoritmi, metodi e modelli atti a quantificare la qualità percepita di immagini e video; 4) algoritmi di indicizzazione e ricerca per i sistemi di archiviazione di informazioni multimediali; 5) algoritmi e metodologie di riconoscimento per la videosorveglianza e la biometria dei volti 2D e 3D; 6) diagnostica per immagini, controllo qualità, visualizzazione e fruizione di dati scientifici; 7) Human-Computer Interaction basata su tecniche di visione artificiale, 2D e 3D Augmented Reality.

La ricerca si qualificherà nei prossimi anni attraverso partecipazione a progetti di ricerca regionali, europei o internazionali; pubblicazioni sulle principali riviste e conferenze internazionali, collaborazioni con l'industria sui seguenti ambiti di ricerca: a) Color Imaging, b) Image e Video Processing, c) Analisi e Classificazione, d) Visual Information Systems, e) Qualità dell'immagine, f) Interazione Uomo macchina, g) Face Biometrics.



LTA, Laboratorio di Test e Analisi del Software

Responsabile: Mauro Pezzè

www.lta.disco.unimib.it

Le principali attività di ricerca svolte nel periodo 2011-2013 sono relative alle tecniche di analisi statica e dinamica del software, e loro integrazione; generazione automatica di casi di test ad elevata copertura del software; diagnosi e correzione automatica dei difetti software; automazione delle attività di testing black box; sistemi autonomi ed auto adattativi.

La ricerca si qualificherà nei prossimi anni attraverso partecipazione a progetti di ricerca regionali, europei o internazionali; pubblicazioni sulle principali riviste e conferenze internazionali, collaborazioni con l'industria sui seguenti ambiti di ricerca: a) Self-healing sistemi software b) Analisi e verifica di software statico c) Analisi di software dinamico e debugging automatico d) Testing orientato ad oggetti e software component-based, e) Black-box testing di interfacce utente.



OPENIT, Opensource Tools and IT Security

Responsabili: Claudio Ferretti, Alberto Leporati

www.openit.disco.unimib.it

L'attività di ricerca ha riguardato tutti gli aspetti legati all'utilizzo innovativo del codice sorgente di sistemi e applicazioni, proponendo in particolare strumenti avanzati per il debugging dei malfunzionamenti di sistemi informatici, ricerca automatizzata di vulnerabilità nelle applicazioni, tecnologie per la sicurezza su sistemi e reti. Tali attività sono svolte anche in concorso con altre sottoaree di ricerca presenti nel Dipartimento e in particolare su analisi statica del software, e telecomunicazioni per la sicurezza sulle reti

La ricerca si qualificherà nei prossimi anni attraverso partecipazione a progetti di ricerca regionali, europei o internazionali; pubblicazioni sulle principali riviste e conferenze internazionali, collaborazioni con l'industria sui seguenti ambiti di ricerca a) Crittografia applicata; protocolli sicuri per reti; b) Analisi statica di codice sorgente; Security checkers; Astrazione e rappresentazione di sistemi software.



ArcSAL, Software Architecture Laboratory

Responsabile: Francesco Tisato

www.sal.disco.unimib.it

L'attività di ricerca ha riguardato le architetture software, con l'obiettivo principale di definire, sviluppare e sperimentare modelli architeturali di riferimento, metodologie e strumenti orientati a risolvere problematiche legate a domini applicativi emergenti. L'attività si è articolata in: ricerca di base, orientata alla definizione di modelli architeturali innovativi; progetti pilota, orientati alla applicazione dei modelli generali e alla sperimentazione di tecnologie emergenti attraverso lo sviluppo di piattaforme prototipali; progetti applicativi, in cui le conoscenze acquisite nella ricerca di base e nei progetti pilota sono trasferite nella soluzione di problemi concreti, attraverso la collaborazione organica con realtà produttive e utenti finali. Le aree di ricerca affrontate nel periodo 2011-2013 possono essere ricondotte alle seguenti keyword: time awareness, space awareness, self awareness e multimodality. Time awareness indica la capacità di un sistema di ragionare su aspetti temporali legati all'informazione di dominio e/o percepire il "flusso" del tempo reale. Space awareness indica la capacità di un sistema di ragionare su locazioni spaziali, sia fisiche che logiche, senza dover conoscere esplicitamente, e quindi gestire, la struttura dello spazio. Self awareness indica la capacità di un sistema di osservare e controllare se stesso in termini di QoS, comportamento, risorse. Multimodality indica l'abilità di supportare tipi di flussi informativi eterogenei fra componenti software. Le quattro aree di ricerca trovano la loro concretizzazione in specifici domini applicativi. Ad esempio, la time awareness può essere applicata in sistemi di videosorveglianza, la space awareness in sistemi di monitoraggio e controllo ambientale, la self-awareness e la multimodality in sistemi che ricadono sotto il termine di ambient intelligence.

La ricerca si qualificherà nei prossimi anni attraverso partecipazione a progetti di ricerca regionali, europei o internazionali; pubblicazioni sulle principali riviste e conferenze internazionali, collaborazioni con l'industria sui seguenti ambiti di ricerca: a) ingegneria del software b) architetture software, c) sistemi embedded, d) sistemi adattivi.

TangoLab

Technology Assistance to NGOs

TANGOLAB,
Technology Assistance to Non-Governmental
Organizations

Responsabile: Roberto Polillo

www.rpolillo.it/tangolab-3

La ricerca ha lo scopo di realizzare e diffondere esperienze di costruzione di applicazioni web innovative, particolarmente adatte per dare supporto alle organizzazioni non profit. Essa è finalizzata allo studio, realizzazione e valutazione di strumenti innovativi di informatica e telecomunicazioni che possano aiutare le organizzazioni non profit a svolgere con più efficacia il loro lavoro utilizzando componenti software open-source e i servizi di cloud computing, ottimizzando così la progettazione di nuovo software. Le motivazioni di questo approccio derivano dalla profonda trasformazione che l'informatica sta subendo a causa dell'evoluzione delle tecnologie Internet che mettono a disposizione: a) diversi tipi di social media, telefonia IP, sistemi di videoconferenza, Web-radio, Web-TV e numerosi strumenti di cooperazione e collaborazione online; b) Cloud computing che abbassa drasticamente le barriere di accesso ai servizi ICT e aumenta la loro sostenibilità; c) Riduzione dei costi di realizzazione delle applicazioni utilizzando tecniche di mashup, web-services, ecc., la composizione modulare di moduli open-source e la progettazione collaborativa (crowdsourcing).

I risultati delle attività di ricerca sono stati sperimentati nell'ambito di un progetto di cooperazione sulla salute infantile con il Nicaragua (Managua).



TELCOLAB, Laboratorio di Telecomunicazioni

Responsabile: Riccardo Melen

www.telco.disco.unimib.it

Le attività di ricerca nel periodo 2011-2013 hanno riguardato lo studio di reti private virtuali che siano sicure anche dal punto di vista dell'analisi statistica del traffico; analisi delle prestazioni di queste soluzioni al variare della topologia di comunicazione e del numero di dispositivi connessi; applicazioni della tecnologia NFC, studio di architetture per supportare efficacemente applicazioni di sicurezza sfruttando le opportunità di integrazione fra le informazioni provenienti da differenti domini (come ad esempio i sistemi informativi e di rete, la videosorveglianza, il controllo degli accessi, i security alert e altri tipi di informazioni disponibili in rete).

La ricerca si qualificherà nei prossimi anni attraverso partecipazione a progetti di ricerca regionali, europei o internazionali; pubblicazioni sulle principali riviste e conferenze internazionali, collaborazioni con l'industria negli ambiti di ricerca sopra menzionati

Responsabile: Gabriella Pasi

www.ir.disco.unimib.it

L'attività di ricerca nell'area dell' Information Retrieval è stata principalmente finalizzata alla definizione di modelli e tecniche innovative di IR, con particolare attenzione alle tematiche della gestione di documenti strutturati, della personalizzazione e della modellazione del contesto per migliorare l'attività di reperimento di informazioni: sia sulla base di interrogazioni dell'utente (motori di ricerca) sia di filtraggio di informazioni (Information Filtering) per la costruzione e l'utilizzo di profili utente allo scopo di identificare in modo automatico informazioni utili all'utente.

Le principali tematiche di ricerca sviluppate durante il triennio 2011-2013 sono state le seguenti:

- A) definizione di tecniche per la personalizzazione della ricerca che considerino il contesto dell'utente;
- B) definizione, implementazione e valutazione di approcci alla contestualizzazione della ricerca di informazioni in grado di stimare la rilevanza di documenti sulla base di molteplici criteri, quali: la pertinenza delle informazioni, l'affidabilità della sorgente delle informazioni, l'aderenza agli interessi dell'utente e il contesto nel quale l'utente sta svolgendo l'attività di ricerca;
- C) definizione e implementazione di motori di ricerca di documenti in formato XML: si è implementata un'estensione formale dei linguaggi di interrogazione XPath e XQuery, chiamata Flexy, all'interno del sistema open source Basex (XML engine);
- D) definizione di tecniche di estrazione di conoscenza da Wikipedia e di tecniche di analisi di informazioni su social media;
- E) definizione di approcci concettuali e personalizzati all'IR, mediante la definizione di profili utente come "ontologie personali".

Le attività di ricerca previste come sviluppo delle tematiche di ricerca illustrate sopra riguardano la definizione di modelli del linguaggio utilizzato dagli utenti nei cosiddetti "user-generated contents" per definire nuovi approcci alla personalizzazione della ricerca e alla raccomandazione di informazioni. Si intende inoltre proseguire l'attività finalizzata alla modellazione del contesto e a tecniche di "knowledge/information extraction".



ITIS. Innovative Technologies for Interactions and Services

Responsabile: Flavio De Paoli

www.itis.disco.unimib.it

Le principali attività di ricerca riguardano:

A) la definizione di modelli e linguaggi per la descrizione di processi di lavoro e organizzativi, di requisiti di qualità dei servizi, di informazioni di contesto e di open data; e per la progettazione di workstation in un contesto cooperativo, di sistemi di ambient intelligence; e la valutazione della qualità dei dati e dell'informazione.

B) la progettazione e lo sviluppo di: sistemi e tecnologie peer-to-peer per il supporto alla collaborazione e al coordinamento e per la gestione e condivisione della conoscenza; sistemi di classificazione (semi)automatica di documenti e servizi mediante ontologie e tracciamento del comportamento; tecnologie per l'interazione e interfacce uomo-macchina multimodali per sistemi collaborativi; piattaforme per la pubblicazione di dati aperti di qualità.

C) la definizione di metodologie e dei relativi framework tecnologici per la Service Science con un approccio multidisciplinare che comprende lo sviluppo di tecniche e strumenti per l'analisi e il miglioramento della qualità dei dati, la descrizione di proprietà non funzionali dei servizi (costo, performance, disponibilità del servizio, sicurezza, trust, licenze d'uso, ecc.), la selezione e ordinamento di Web services; la valutazione di compatibilità e composizione di contratti associati a servizi.

Le principali aree applicative coinvolte sono: context-aware computing, eGovernment, sistemi collaborativi flessibili adattabili al contesto e personalizzabili, sistemi informativi per le organizzazioni, social computing, ubiquitous e mobile computing, web services, qualità dei dati e dei servizi, repositories di schemi concettuali e ontologie, gestione dei Beni Culturali.

La ricerca del laboratorio dovrà essere maggiormente qualificata in termini di relazioni internazionali, di acquisizione di progetti europei, nazionali e regionali, e di miglioramento della qualità della produzione scientifica. L'attività di ricerca si prefigge di consolidare i metodi, strumenti e prototipi sviluppati all'interno dei progetti attivi nel triennio 2011-2013 sulla base di nuove funzionalità, di nuovi domini applicativi e della integrazione con tecnologie complementari.

Responsabile: Sara Manzoni

www.lintar.disco.unimib.it/space/start

Gli ambiti in cui si focalizzano le competenze in questa sottoarea sono: l'Intelligenza Artificiale Distribuita e l'Ingegneria della Conoscenza. Nello specifico, le attività svolte nel triennio riguardano prevalentemente la messa a punto di modelli di sistemi complessi e di strumenti per il loro studio con l'obiettivo, da un lato di contribuire allo studio del comportamento e delle proprietà del sistema stesso, e dall'altro, poterne generalizzare e trasferire ad altro ambito e contesto applicativo i risultati.

Contesto preferenziale per le attività del periodo 2011-2013 è stato il comportamento dinamico di gruppi e singoli individui in situazioni di affollamento (crowd, folla). La folla è un sistema dinamico complesso con proprietà e comportamenti emergenti e fornisce numerosi spunti per lo sviluppo e la sperimentazione di metodi e strumenti, sia estensione di approcci esistenti che completamente originali. I risultati di tali ricerche, oltre ad avere implicazioni di natura applicativa per la progettazione e pianificazione di ambienti, trovano riscontro in specifici ambiti delle scienze sociali (e.g. psicologia comportamentale, percezione ambientale) e possono contribuire alla realizzazione di sistemi distribuiti di monitoraggio e controllo (smart environments). Inoltre sono state svolte attività legate alla progettazione e sviluppo di sistemi basati sulla conoscenza (Knowledge Based Systems) per il supporto di attività decisionali knowledge-intensive, in particolare con riferimento ad attività di progettazione di prodotto e di processo produttivo, oltre che di organizzazione di contenuti e documenti in basi di conoscenza.

Per questa sottoarea si intende incrementare le collaborazioni scientifiche e le relazioni con l'attività didattica, sia a livello nazionale che internazionale, con la partecipazione a progetti di ricerca soprattutto a livello internazionale, di miglioramento della qualità della produzione scientifica. L'attività di ricerca intende consolidare tecniche e metodi sviluppati, anche per estendere il campo di applicazione all'analisi automatizzata delle dinamiche di folla.

Responsabile: Giancarlo Mauri

www.bimib.disco.unimib.it

L'attività di ricerca in questo settore si focalizza su diversi aspetti della Biologia Computazionale, e della Systems Biology. Comprende inoltre lo studio di modelli di calcolo bioispirati (algoritmi genetici ed evolutivi, sistemi a membrane e sistemi di splicing) e il disegno, implementazione e sperimentazione di algoritmi per problemi combinatori su grafi, sequenze ed alberi che hanno principalmente applicazione in genomica.

Le ricerche nel 2011-2013 della sottosezione "Systems Biology" si sono concentrate sui seguenti temi. (1) Ricostruzione di modelli di sistemi biologici a partire da misure sperimentali in serie temporale e no. In particolare si sono sviluppati degli algoritmi per la ricostruzione di modelli di accumulazione di alterazioni genetiche nel cancro a partire da misure "cross-sectional". (2) Sviluppo ed analisi di modelli di simulazione multi-scala su lattice (Cellular Potts Model), con particolare riferimento all'analisi di modelli per la "differenziazione stocastica" di cellule staminali e della loro progenie.

L'attività di ricerca della sottosezione "Algoritmica Sperimentale" hanno riguardato lo studio di algoritmi di approssimazione per problemi combinatori su sequenze e sull'anonimizzazione di dati e problemi di consensus clustering, lo sviluppo e sperimentazione di algoritmi per la ricostruzione di alberi filogenetici in presenza di eventi di ricombinazione o mutazioni ricorrenti.

Le ricerche nell'ambito della "Biologia Computazionale" hanno riguardato principalmente lo sviluppo di algoritmi per la ricostruzione del grafo delle isoforme e di siti di splicing a partire da dati di sequenziamento del trascrittoma, ed infine algoritmi per l'inferenza di aplotipi in presenza di pedigree. Si è prodotto software utilizzato dai biologi per l'analisi di dati biologici.

La sottosezione di Calcolo Naturale ha svolto ricerche sulla potenza e sulla efficienza computazionale di diverse varianti di sistemi a membrane (P systems) e delle corrispondenti classi di complessità, nonché sulla caratterizzazione del potere computazionale dei sistemi di splicing. Le principali competenze e aree tematiche coinvolte sono: modelli per il Calcolo Molecolare, sistemi di splicing, P-systems; algoritmi genetici ed evolutivi. Simulazione multiscala e multicellulare di modelli di sistemi biologici. Biologia Sistemica. Disegno e sperimentazione di algoritmi di approssimazione per problemi combinatori in Biologia Computazionale (inferenza di aplotipi, confronto e ricostruzione di alberi filogenetici, clusterizzazione di dati di espressione genica da microarray, analisi di dati di sequenziamento e algoritmi di allineamento).

Le ricerche in questo settore saranno orientate a potenziare le collaborazioni nazionali e internazionali realizzate mediante la partecipazione a progetti comuni di ricerca. In particolare, si intende rafforzare la parte relativa allo sviluppo e rilascio di software open-source utilizzabile dalla comunità scientifica bioinformatica, ad esempio sulla piattaforma BioConductor.



MC3, Modelli della Concorrenza, Comunicazione e Computazione

Responsabile: Lucia Pomello

<http://www.mc3.disco.unimib.it>

Le attività di ricerca di questo settore si sono concentrate su due linee principali: (1) strutture algebriche per lo studio di sistemi concorrenti; (2) metodi e modelli per lo sviluppo e la verifica di sistemi concorrenti e distribuiti.

La prima linea di ricerca riguarda i fondamenti della teoria della concorrenza e si è focalizzata sullo studio di strutture algebriche su ordini parziali, utilizzati per modellare il comportamento di sistemi concorrenti, ottenute da operatori di chiusura. In particolare, risultati ottenuti in precedenza in quest'ambito sono stati estesi a casi più generali e sono state studiate le relazioni di tali strutture algebriche con le logiche per la concorrenza e con la logica quantistica.

Gli sviluppi di questa linea di ricerca comprendono ulteriori approfondimenti degli aspetti logici e algebrici sul piano teorico, e l'esame di possibili applicazioni alla specificazione e analisi di sistemi distribuiti.

La seconda linea di ricerca mira alla definizione di strumenti efficaci di disegno e analisi di sistemi distribuiti e concorrenti, basati sulle nozioni di modularità, raffinamento e composizionalità. E' stata definita una nuova nozione di morfismo per reti di Petri che supporta un'operazione di raffinamento e di astrazione. Inoltre è stata definita un'operazione di composizione che corrisponde ad una sincronizzazione parziale tra componenti e sono state studiate le proprietà strutturali e di comportamento preservate da tale operazione. Le tecniche proposte sono state applicate e discusse su un caso di studio relativo al modello di un protocollo distribuito.

Verranno indagate ulteriori nozioni di morfismo su particolari classi di reti di Petri a supporto di tali operazioni di astrazione/raffinamento e composizione.

Verrà studiata la formalizzazione del flusso di informazione tra componenti di un sistema, considerando in particolare nozioni di osservabilità e di non-interferenza, con l'obiettivo di applicare questa formalizzazione a problemi relativi alla sicurezza nei sistemi distribuiti e alla modellazione del flusso di informazione tra componenti.

I risultati ottenuti saranno approfonditi e sviluppati anche al fine di produrre ulteriori pubblicazioni scientifiche.

Su questi temi si intende rafforzare le collaborazioni internazionali e nazionali già intraprese, in particolare con la sede di Cachan dell'ENS francese e con l'IRISA di Rennes sempre in Francia.

Responsabili: Davide Ciucci, Alberto Dennunzio

www.syslab.disco.unimib.it

Le ricerche riguardano principalmente il trattamento formale di Sistemi Complessi e Incerti, loro modellazione e implementazioni.

Le ricerche nell'ambito del trattamento formale di Sistemi Complessi riguardano modelli paradigmatici per la descrizione e simulazioni di Sistemi Complessi e mirano a comprendere a fondo le proprietà principali di tali modelli anche con l'obiettivo di permetterne un adeguato utilizzo nelle applicazioni. I risultati ottenuti hanno riguardato principalmente le proprietà dinamiche, le questioni di decidibilità e gli aspetti computazionali relativi principalmente a: Automi Cellulari con estensioni alla dimensione 2; Automi Cellulari non-uniformi; Automi Cellulari asincroni.

Inoltre, si è studiata la possibilità di utilizzare gli Automi Cellulari per calcolare funzioni e primitive crittografiche, attraverso la definizione di generatori di numeri pseudocasuali e schemi per la condivisione di segreti tramite Automi Cellulari unidimensionali bipermutivi.

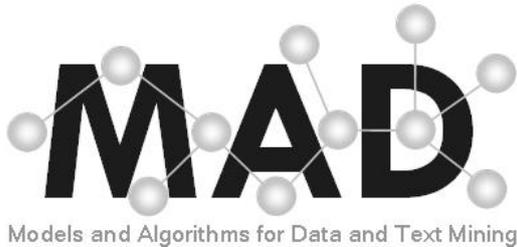
Si intendono estendere le indagini condotte in dimensione qualunque ed approfondire i casi più generali di modelli non-uniformi e asincroni. Inoltre si prevede di considerare modelli (diversi dagli Automi Cellulari) per lo studio di sistemi complessi nell'ambito biochimico. Per quanto riguarda l'utilizzo degli Automi Cellulari in crittografia, si intende applicare tecniche di ricerca ed euristiche derivanti dall'ambito degli Algoritmi Genetici e degli Algoritmi Evolutivi per individuare funzioni booleane dotate di buone proprietà crittografiche, che possano essere utilizzate come regole locali di Automi Cellulari unidimensionali bipermutivi.

La linea di ricerca sull'incertezza e sulla logica mira principalmente al trattamento formale di sistemi in presenza di incertezza mediante tecniche basate su rough sets, fuzzy sets e logica.

I risultati hanno riguardato le logiche a tre valori di verità: semantica, loro dipendenze, relazione con logiche modali. Inoltre, si è continuato con lo studio di aspetti dinamici nei rough sets allo scopo di usarli quando i dati evolvono nel tempo. Infine, si sono applicate tecniche basate su fuzzy sets e granular computing alla rappresentazione di conoscenza tramite ontologie.

Si prevede di continuare il lavoro sulle logiche a tre valori e sul loro rapporto con altri paradigmi di rappresentazione dell'incertezza. Un'ulteriore linea di ricerca riguarda il confronto tra rough set e formal concept analysis.

Infine, è nostro obiettivo proseguire e rafforzare ulteriormente le collaborazioni internazionali già esistenti sulle tematiche sopra descritte e di organizzare workshop e special session a conferenze internazionali sulle tematiche di ricerca del gruppo.



MAD, Models and Algorithms for Data and Text Mining

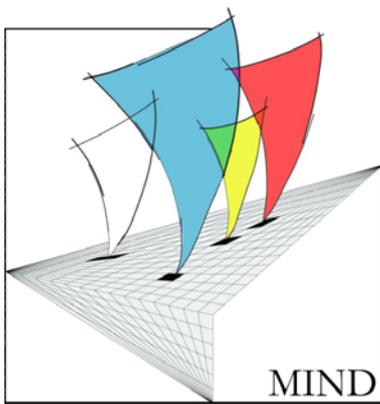
Responsabile: Fabio Stella

www.mad.disco.unimib.it

Le principali linee di ricerca riguardano: (1) continuous time Bayesian networks per la rappresentazione di sistemi dinamici a tempo continuo; (2) topic models per l'estrazione automatica di concetti a partire dal testo.

La prima linea di ricerca si è focalizzata nel periodo 2011-2013 sullo studio di algoritmi di inferenza ed apprendimento strutturale a partire da dati temporali e/o sequenziali. In tale contesto sono stati studiati, sviluppati ed implementati algoritmi esatti ed approssimati per l'inferenza. Sono stati inoltre progettati ed implementati algoritmi esatti per l'apprendimento strutturale e parametrico delle continuous time Bayesian networks. Questa linea ha visto anche la ideazione di un nuovo insieme di modelli di classificazione supervisionata per dati temporali; continuous time naive Bayes classifier, tree augmented naive Bayes classifier e K-augmented continuous time Bayesian network classifier. Gli algoritmi sviluppati sono stati applicati a problemi provenienti da diversi domini; finanza, medicina e biologia. Nell'immediato futuro gli sforzi verranno indirizzati all'estensione del modello delle continuous time Bayesian networks al caso non stazionario, vale a dire al caso in cui le relazioni di causalità sono libere di mutare nel corso del tempo. Un tale modello dovrebbe consentire di superare alcune limitazioni che rendono attualmente le continuous time Bayesian networks non applicabili in domini particolarmente importanti. Le collaborazioni scientifiche saranno con l'Università di Aalborg (DK) e con l'Università di Almeria (SP).

La seconda linea di ricerca è finalizzata ad estrarre in modo automatico i concetti presenti in un corpus di documenti. Nello specifico viene studiato il paradigma dei modelli probabilistici generativi come meccanismo di compilazione di un documento. In tale ambito è stato studiato un problema di particolare rilevanza quale quello della validazione dei concetti estratti a partire da un corpus di documenti. Infatti, i modelli in questione soffrono di patologie tipiche dei modelli di apprendimento non supervisionato, tendendo a sovra-complicare i concetti estratti. I risultati teorici ottenuti in questa linea di ricerca sono stati applicati a dati reali tramite una implementazione software realizzata in collaborazione con l'Università di Monterrey (Messico). Lo sviluppo di questa linea di ricerca passa nell'immediato futuro dallo studiare misure di concordanza dei concetti estratti e dalla individuazione di procedure atte a selezionare il numero ottimale di concetti da estrarre in relazione alla complessità presente nel corpus di documenti disponibile. Queste attività di ricerca verranno condotte in collaborazione con il gruppo di ricerca del Prof. Markus Zanker dell'Università Alpen Adria di Klagenfurt.



MIND, Models in decision making and data @analysis

Responsabili: Francesco Archetti, Vincenzina Messina

www.mind.disco.unimib.it

L'attività di ricerca nel periodo 2011-2013 ha riguardato principalmente lo sviluppo di modelli probabilistici di classificazione e previsione in diversi ambiti applicativi. Particolare attenzione viene rivolta ai modelli relazionali, volti a migliorare le performance degli algoritmi esistenti attraverso il superamento della condizione di indipendenza dei dati tipica dei modelli di machine learning. Tali modelli sono in grado di rappresentare l'incertezza ma tengono altresì in considerazione le relazioni esistenti tra gli oggetti e/o gli attributi degli oggetti per migliorare l'inferenza sui dati e trovano applicazione in tutti quegli ambiti dove gli oggetti da classificare fanno parte di una rete di oggetti in relazione, o dove il modello dei dati può essere rappresentato da un modello relazionale. I primi risultati sono stati ottenuti nelle seguenti applicazioni:

- a) eJustice: clustering di documenti in ambito eJustice per la creazione di sommari multimediali e modelli di information extraction per template filling automatico.
- b) Analisi dei sistemi biologici: caratterizzazione di linee cellulari e modelli di previsione di risposta al farmaco basati sull'analisi dell'espressione genica.
- c) Ambient Intelligence: sviluppo di metodi di apprendimento automatico per la previsione e la diagnostica in sistemi di videosorveglianza e in sistemi multisensoriali e in particolare Wireless mobile Sensor Networks.
- d) Sentiment Analysis: caratterizzazione di feature linguistiche all'interno dei principali social media e definizione di modelli supervisionati e semi-supervisionati relazionali per la previsione del sentiment all'interno di brevi messaggi testuali

Particolare attenzione è stata rivolta ai seguenti temi di ricerca:

- Clustering: sono state studiate diverse varianti per la formulazione di problemi di programmazione lineare intera vincolata, in grado di includere informazioni di tipo relazionale sottoforma di vincoli "soft".
- Classificazione: il modello di apprendimento Conditional Random Fields è stato esteso al fine di includere informazioni di tipo relazionale in fase di inferenza. Al tradizionale algoritmo di Viterbi, è stata associata una formulazione del problema di inferenza come Programmazione Lineare Intera su grafi in grado di includere le informazioni relazionali anche in questo caso sottoforma di vincoli "soft". E' stato inoltre affrontato il problema del rilassamento di tali vincoli mediante un approccio Lagrangiano e la definizione di una variante dell'algoritmo del sub-gradiente.
- Ottimizzazione: un ulteriore tema di ricerca riguarda lo studio e sviluppo di modelli matematici di ottimizzazione per problemi di decisione. In particolare i modelli sviluppati sono stati applicati a problemi di trasporto aereo e ferroviario.