

CURRICULUM VITAE ET STUDIORUM

di Luigi Pontieri

1. Posizione attuale

Luigi Pontieri è ricercatore presso l'Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR-CNR), sito in Rende (CS), con contratto di lavoro a tempo indeterminato.

2. Iter formativo e posizioni ricoperte

Luigi Pontieri ha conseguito la laurea in Ingegneria Informatica il 25 luglio 1996 presso l'Università degli Studi della Calabria, con il punteggio di 110/110 e lode, e l'abilitazione alla professione di Ingegnere nella seconda sessione dell'anno 1996.

Dal 10 ottobre 1996 al 20 maggio 1997 ha partecipato alle attività del gruppo di ricerca coordinato dal Prof. Luigi Palopoli, presso il Dipartimento di Elettronica Informatica e Sistemistica dell'Università degli Studi della Calabria. Dal 15 maggio 1997 al 15 ottobre 1997 è stato collaboratore di ricerca esterno presso l'Istituto per la Sistemistica e l'Informatica del Consiglio Nazionale delle Ricerche (ISI-CNR). Dal 1 ottobre 1997 al 5 aprile 2001 ha svolto attività di ricerca presso il Dipartimento di Elettronica Informatica e Sistemistica dell'Università degli Studi della Calabria e presso l'Istituto per la Sistemistica e l'Informatica del CNR in qualità di studente del XIII° ciclo del corso di Dottorato di Ricerca in Ingegneria dei Sistemi ed Informatica dell'Università degli Studi della Calabria, usufruendo di una borsa di studio finanziata dall'istituto ISI-CNR.

Il 5 aprile 2001 ha conseguito il titolo di Dottore di Ricerca in Ingegneria dei Sistemi ed Informatica presso l'Università degli Studi della Calabria, discutendo la tesi "*Strumenti e metodi per l'analisi di sistemi informativi complessi ed eterogenei*".

Il 29 gennaio 2001, a seguito di concorso pubblico, è stato assunto con contratto di lavoro a tempo indeterminato, con profilo di Ricercatore, presso l'Istituto per la Sistemistica e l'Informatica del Consiglio Nazionale delle Ricerche (ISI-CNR), sito in Rende (CS) ed oggi parte integrante dell'Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR-CNR).

3. Attività di ricerca scientifica

Luigi Pontieri ha svolto la propria attività di ricerca scientifica sia presso il Dipartimento di Elettronica Informatica e Sistemistica dell'Università degli Studi della Calabria sia presso l'Istituto per la Sistemistica e

l'Informatica (ISI-CNR), e successivamente presso l'Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR-CNR), del Consiglio Nazionale della Ricerche.

In particolare dal 10 ottobre 1996 al 20 maggio 1997 ha svolto attività di ricerca in collaborazione con il gruppo di ricerca coordinato dal Prof. Luigi Palopoli, presso il Dipartimento di Elettronica Informatica e Sistemistica dell'Università degli Studi della Calabria.

Dal 15 maggio 1997 al 15 ottobre 1997 è stato collaboratore di ricerca esterno presso l'Istituto per la Sistemistica e l'Informatica del Consiglio Nazionale delle Ricerche (ISI-CNR).

Da ottobre 1997 al 4 aprile 2001 ha svolto attività di ricerca presso il Dipartimento di Elettronica Informatica e Sistemistica dell'Università degli Studi della Calabria e presso l'Istituto per la Sistemistica e l'Informatica del CNR in qualità di studente del XIII° ciclo del corso di Dottorato di Ricerca in Ingegneria dei Sistemi ed Informatica dell'Università degli Studi della Calabria, usufruendo una borsa di studio finanziata dall'ISI-CNR.

Dal 29 gennaio 2001 svolge attività di ricerca presso l'istituto ISI-CNR di Rende, dal 15 giugno 2002 parte dell'istituto ICAR-CNR, in qualità di ricercatore con contratto a tempo indeterminato.

La sua attività di ricerca si è incentrata sui seguenti temi:

- Data Integration and Warehousing
- Data Compression
- Knowledge Discovery
- Business Process Analysis

Le sezioni seguenti descrivono, in modo succinto, le attività di ricerca e didattiche svolte da Luigi Pontieri. Il documento si conclude con un elenco dei principali lavori scientifici pubblicati (o in fase di pubblicazione).

3.1. Data Integration and Warehousing

Molto spesso i sistemi informativi di un'organizzazione sono costituiti da numerosi archivi di dati concepiti in maniera disorganica, sviluppati secondo approcci completamente differenti, con una conseguente diseconomicità nella gestione delle informazioni. In tale scenario risulta di notevole interesse la progettazione di nuovi sistemi capaci di gestire ed elaborare le informazioni provenienti da basi di dati eterogenee e distribuite. Il numero e le dimensioni delle sorgenti informative tipicamente coinvolte rendono, inoltre, auspicabile la disponibilità metodi e strumenti semi-automatici che supportino tale attività di progettazione. I processi di integrazione e di astrazione delle basi di dati, fondamentali per la progettazione e la realizzazione di un Sistema Informativo Cooperativo, sono considerati in [1], dove sono illustrate due nuove tecniche semi-automatiche per l'integrazione e l'astrazione, basate sull'esistenza di un insieme di proprietà che rappresentano relazioni semantiche fra oggetti degli schemi delle sorgenti informative di base.

In [4] è presentata un'evoluzione degli algoritmi di integrazione e di astrazione ed il loro impiego nel processo di costruzione di un repository dei dati, di notevole importanza per la realizzazione dei mediatori di un Sistema Informativo Cooperativo; nello stesso lavoro si descrive, inoltre, l'applicazione dell'intera metodologia ad un caso reale di notevole complessità.

Il problema dell'integrazione delle istanze contenute in basi di dati eterogenee è affrontato in [9] e [11], dove si introduce un operatore per l'integrazione dei dati ed una tecnica, basata sulla programmazione logica disgiuntiva, per la gestione dei conflitti presenti nella vista integrata e per la produzione di risposte consistenti. In [10] è presentata una variante del suddetto operatore di integrazione, definita per la costruzione di viste integrate materializzate, le cui istanze sono arricchite con l'indicazione delle sorgenti informative da cui derivano e della sequenza secondo cui queste sono state combinate per produrre la vista integrata.

Un ulteriore problema affrontato è consistito nell'identificazione di istanze con rappresentazioni eterogenee ma corrispondenti alla stessa entità del mondo reale. In particolare, in [28], è stato definito un approccio incrementale per la scoperta di dati "duplicati" che affronta tale problema in un'ottica di clustering, cercando di individuare cluster di istanze con alta similarità intra-cluster e bassa similarità inter-cluster. L'approccio trae vantaggio sull'impiego di una struttura di indicizzazione basata su tecniche hash, al fine di rintracciare per ogni nuova istanza tutte quelle che appartengono allo stesso cluster (e quindi rappresentano, con alta probabilità, la stessa entità del mondo reale) fra quelle già considerate. L'efficacia di tale schema di indicizzazione si basa sulla sua capacità di associare due oggetti allo stesso bucket con una probabilità proporzionale al loro grado di mutua similarità.

Un secondo filone di ricerca ha riguardato la definizione di metodi per la costruzione di Data Warehouse, ovvero archivi informativi integrati ed omogenei, in grado di consentire un accesso efficiente e flessibile al contenuto delle varie sorgenti informative. I dati provenienti dalle diverse fonti informative vengono materializzati nel Data Warehouse, in modo che l'estrazione delle informazioni venga effettuata sui dati materializzati e non su quelli originari. La costruzione di un Data Warehouse necessita di un complesso processo di riconciliazione e integrazione, della progettazione di un modello unico di rappresentazione dei dati e della generazione di meccanismi di mapping delle informazioni dalle sorgenti informative di partenza al modello del Data Warehouse.

Constatata l'inadeguatezza delle architetture tradizionali per la realizzazione di Data Warehouse a relativi ad un numero notevole di basi di dati di grandi dimensioni, in [4] si propone una nuova architettura, basata su una struttura gerarchica a diversi livelli di astrazione, che consente lo sviluppo di nuove metodologie per la sintesi dello schema dei dati riconciliati. In [3] viene descritto un approccio semi-automatico che consente di ottenere un Data Warehouse avente l'architettura citata precedentemente e che si fonda sull'estrazione di conoscenza di intensionale ed estensionale dalle basi di dati operazionali.

La rilevanza del problema dell'integrazione è stata amplificata dallo sviluppo del Web, che rende disponibile l'accesso ad una enorme varietà di sorgenti informative. Queste sorgenti possono essere fortemente eterogenee, non solo per quanto attiene i modelli dei dati, i linguaggi di interrogazione e di manipolazione ed i sistemi di gestione, ma anche rispetto al formato: alcune possono essere ben strutturate (e.g., basi di dati relazionali), altre semi-strutturate (e.g., documenti XML) e altre ancora possono essere totalmente prive di struttura (e.g., testo). Alla luce di tali osservazioni, in [14] è stato definito un approccio per ottenere l'integrazione di sorgenti informative con formati differenti, consistente di due componenti fortemente correlate ed integrate, che si occupano, rispettivamente, dell'integrazione intensionale e di quella estensionale e che utilizzano, rispettivamente, un modello concettuale ed uno logico, simili e coesi, per la rappresentazione di dati semi-strutturati. L'approccio è capace di gestire l'incompletezza delle informazioni (valori nulli o sconosciuti), produrre risposte consistenti a interrogazioni su dati eventualmente inconsistenti ed ottenere informazioni sul contenuto delle sorgenti informative originarie.

Un problema rilevante nella progettazione di un data warehouse riguarda la definizione di data mart destinati a particolari processi di analisi OLAP, normalmente basati sul paradigma multidimensionali. La difficoltà di tale attività di progettazione è dovuta alla dicotomia interna dei processi di analisi multidimensionali, dove l'alta dinamicità dei requisiti di analisi si scontra con la rigidità della rappresentazione multidimensionali dei dati che restringe fortemente il tipo di interrogazioni e di analisi cui i dati possono essere sottoposti. Pertanto, spesso accade che la variazione del fabbisogno informativo induca alla definizione di un nuovo data mart. D'altro canto, molti data mart potrebbero contenere informazioni molto simili, o lo stesso insieme di informazioni magari visti da punti di vista differenti. In tale contesto si colloca il lavoro presentato in [17] che intende favorire lo sviluppo incrementale e sinergico di data mart multidimensionali, introducendo un nuovo modello per la rappresentazione di reti di dati multidimensionali, dove i vari fatti possono essere definiti a partire da altri preesistenti mediante opportune operazioni di trasformazioni sulle misure e sulle dimensioni. In [19], inoltre, è definito un insieme di strumenti per la definizione di reti di fatti multidimensionali ed è presentato un sistema software in grado di supportare il progettista nelle varie fasi del processo di sviluppo di un data warehouse.

3.2. Data Compression

La disponibilità di informazioni riorganizzate ed integrate nella forma di un Data Warehouse costituisce un terreno fertile sia per l'applicazione di tecniche OLAP (On Line Analytical Processing), che consentono di accorpare e consultare in maniera rapida informazioni altrimenti di difficile interpretazione e sia per l'esecuzione dei più sofisticati e complessi processi di KDD e Data Mining. L'impiego, sinergico ed integrato, delle metodologie e delle tecniche maturate nei campi del Data Warehousing, dell'OLAP e del Data Mining si prospetta come un approccio adeguato per l'analisi di sistemi informativi costituiti da sorgenti distribuite ed eterogenee e per uno sfruttamento adeguato delle loro potenzialità informative.

In tale contesto l'attività di ricerca di Luigi Pontieri si è incentrata sul problema della valutazione efficiente di query di tipo OLAP su Data Warehouses di grandi dimensioni, considerando soprattutto range query con funzioni d'aggregazione sugli attributi di misura di un datacube. L'obiettivo è la riduzione dei tempi di risposta e della quantità di memoria utilizzata per la rappresentazione dei dati, considerando la possibilità di restituire risposte approssimate. In particolare, in [12] e [37] si introducono delle nuove tecniche per la rappresentazione compressa delle informazioni contenute nel datacube. Adottando un approccio simile a quello usato per la costruzione di istogrammi, si partizionano i dati in blocchi disgiunti memorizzando i valori aggregati degli attributi di misura per ogni blocco. Questo tipo di rappresentazione è arricchito memorizzando una quantità ridotta di informazione addizionale in grado di descrivere, in modo approssimato, la distribuzione reale dei valori all'interno del blocco. L'applicazione di tale tecnica a vari metodi classici di partizionamento, definiti per la costruzione di istogrammi, produce notevoli miglioramenti dell'accuratezza nella stima dei risultati delle query.

3.3. Knowledge Discovery

La Gestione della Conoscenza (Knowledge Management, KM) rappresenta un settore di fondamentale importanza per l'analisi e lo sviluppo dei moderni sistemi informativi. Le notevoli trasformazioni che hanno interessato il panorama socio-economico e tecnologico negli ultimi anni hanno infatti spinto molte organizzazioni a dotarsi di sistemi informativi fortemente basati sulla gestione della conoscenza correlata alle proprie attività di business, al fine di accrescere il proprio livello di competitività, consolidando il know-how acquisito e migliorando l'efficienza e l'efficacia dei processi decisionali e strategici.

Il modello organizzativo dalla "learning organization" (cioè una struttura capace di acquisire, organizzare e trasferire conoscenza, modificando il proprio comportamento sulla base della propria conoscenza sulla realtà) presuppone l'esistenza di ambienti in grado di supportare il ciclo di vita della conoscenza, offrendo strumenti per acquisire, modellare, condividere, estrarre e pubblicare conoscenza.

Infine in [21] è presentato un sistema, denominato ICAR Knowledge Portal, per la gestione della conoscenza all'interno dell'istituto ICAR-CNR. Questo sistema è stato progettato come un ambiente web-based capace di supportare la creazione, la condivisione, l'accesso e la pubblicazione di conoscenza riguardante l'organizzazione dell'istituto e varie attività istituzionali che si svolgono al suo interno. In particolare, esso consente ai ricercatori di svolgere facilmente operazioni quali: creare un sotto-portale personale o per un gruppo di lavoro, creare ed organizzare knowledge item (come news, pubblicazioni, link, annunci, progetti, etc.), condividere tali item all'interno di gruppi di lavoro, ricercare un documento specifico o navigare attraverso un insieme di documenti correlati (ontology-driven browsing).

Gli approcci di Knowledge Discovery si propongono di estrarre conoscenza da grandi quantità di dati mediante tecniche di Data Mining. Queste, consentono di sfruttare le informazioni disponibili al fine di

scoprire l'esistenza di regolarità o di indurre veri e propri modelli generalizzati, in grado di offrire una rappresentazione intensionale e sintetica dei dati e/o effettuare predizioni.

A tal riguardo, in [6] si considera l'utilizzo combinato di tecniche di induzione di alberi di decisione e di tecniche di clustering di tipo bayesiano nella classificazione di dati di tipo economico-finanziario. L'obiettivo è di ottenere un'appropriata partizione dei dati ed una loro descrizione sintetica e facilmente interpretabile. In [5] è illustrato un sistema per l'applicazione semi-automatica dell'approccio precedentemente descritto, in grado di supportare l'intero processo di classificazione, integrando in un unico ambiente di lavoro vari strumenti per l'elaborazione delle informazioni di base e per l'estrazione e la gestione di nuova conoscenza. Infine in [7] e [8] si propone l'uso delle suddette tecniche di induzione in combinazione con strumenti deduttivi, utilizzati per interrogare, valutare e analizzare i modelli indotti, al fine di produrre conoscenza di alto livello, utile per la comprensione e l'analisi dei dati.

Il notevole sviluppo di Internet e, in particolare, del WEB, cui si sta assistendo negli ultimi anni, ha dato origine ad una pressante richiesta di metodi e strumenti per analizzare l'enorme quantità di informazione disponibile in tale contesto. Un approccio interessante a questo problema consiste nel riutilizzo e nella ridefinizione di tecniche di data mining indirizzate con riferimento alle particolari tipologie di dati presenti su WEB. In [16], [18] e [18] si definiscono alcune tecniche per il calcolo della similarità strutturale fra documenti in formato XML, standard universalmente accettato per lo scambio di informazioni in Internet. L'approccio si basa, innanzitutto, sulla codifica di un documento XML sotto forma di un'opportuna serie numerica, capace di cogliere gli aspetti caratteristici della sua struttura. La similarità fra documenti è, quindi, valutata confrontando le trasformate delle loro codifiche, calcolate mediante l'applicazione della Trasformata di Fourier Discreta (algoritmo FFT). Un'applicazione rilevante di tale metodo è il clustering e la classificazione di documenti HTML, esaminata in [27] e [32], utilizzabile per l'integrazione e la ristrutturazione di siti web.

Un altro tema di ricerca concerne il problema dell'*high-order co-clustering*, dove più domini eterogenei devono essere partizionati simultaneamente, in base a dati sulle correlazioni fra valori di domini differenti, tipicamente rappresentati in forma di tabelle di contingenza. La maggior parte delle proposte presenti in letteratura si limita al caso di due soli domini (bi-clustering) e che nei pochissimi approcci ideati per un numero maggiore di domini, il problema è affrontato minimizzando una combinazione lineare delle funzioni obiettivo relative a singoli sotto-problemi di bi-clustering. Ciò richiede che ad ogni tabella di contingenza sia associato un peso, che ne esprime l'affidabilità e la rilevanza dei dati in essa contenuti ai fini del co-clustering; nonostante la scelta di tali pesi non sia un compito semplice, e influisca fortemente sui risultati del co-clustering, allo stato dell'arte non esiste nessun metodo automatico per il loro tuning. Un importante tentativo per alleviare questo problema è costituito dal lavoro presentato in [34], dove viene introdotto un framework information-theoretic per il problema del high-order co-clustering, nel caso di correlazioni fra i domini strutturate secondo una "stella", con un dominio centrale correlato con tutti i rimanenti domini.

Piuttosto che ottimizzare un'unica funzione obiettivo globale (con qualche schema di pesatura esplicita) viene introdotta un'opportuna nozione di accordo (“agreement”) fra i vari sottoproblemi di co-clustering bidimensionale, che garantisce un livello minimo di approssimazione della soluzione ottima (locale) per ogni singola funzione obiettivo. Nello stesso lavoro è stato definito un algoritmo il calcolo efficiente di un co-clustering, che segue tale nozione di accordo.

Al fine di definire un approccio efficiente e scalabile al problema del multi-way co-clustering sopra menzionato, in [52] è stata presentata una formulazione parallela e distribuita degli schemi algoritmici introdotti in [34], nel paradigma di programmazione Map-Reduce. Al fine di rendere l'approccio anche più generale e versatile, la perdita di informazione relativa dei vari sotto-problemi di bi-clustering è espressa in termini di divergenze di Bregman, che generalizzano ed unificano le misure impiegate negli approcci tradizionali. L'approccio è stato implementato in un sistema Java, basato sulla piattaforma open-source Hadoop, ed installato su un cluster computer. Una serie di test su dataset di benchmark ha infine permesso di dimostrarne l'efficacia e la scalabilità.

3.4. Business Process Analysis

Un altro tema affrontato consiste nello studio di tecniche di supporto all'analisi ed all'ottimizzazione di processi organizzativi.

Il principale tema trattato in tale contesto riguarda il problema del *process mining*, i.e., la scoperta di un modello di workflow per un processo sulla base di un log contenente dati di esecuzioni passate del processo.

In [23], [24] e [34] viene illustrato un efficiente approccio al process mining basato su un clustering strutturale delle tracce contenute nel log, che consente di identificare differenti modalità di utilizzo del processo. L'obiettivo finale è quello di estrarre una rappresentazione accurata e compatta del processo, costituita da un insieme di schemi di workflow alternativi, uno per ogni scenario di uso, che sostituisce l'unico schema di workflow che sarebbe scoperto dalle tecniche di process mining tradizionali. Inoltre, in [34] viene introdotto un framework teorico e un'analisi di interessanti proprietà computazionali.

Tale approccio è stato esteso in [36] mediante l'introduzione di un approccio per il clustering delle esecuzioni che considera informazioni relative alle performance del processo, oltre che ai dati “strutturali” sulle attività eseguite e sulla loro temporizzazione. L'idea principale consiste essenzialmente nel formalizzare il problema come problema di co-clustering, in cui il dominio delle tracce è correlato ai vari domini numerici che corrispondono, ognuno, ad una particolare metrica di performance.

In [25] le tecniche di process mining sono impiegate come componenti di un framework metodologico per la progettazione e all'analisi di processi, basata sull'utilizzo di opportune ontologie per la descrizione dei modelli di processo. In [29] si descrive l'applicazione delle tecniche di process mining per supportare la gestione di workflow distribuiti, ottenuti come composizione di servizi messi a disposizione in una Griglia computazionale (*Grid Workflow*). In tale scenario la conoscenza viene sfruttata per supportare efficacemente

diverse fasi del ciclo vita di un workflow, comprendenti (re)design, matchmaking, scheduling e performance monitoring. Nello stesso lavoro viene, inoltre, presentata un'architettura, orientata ai servizi, di sistema di workflow management, per l'implementazione dell'approccio.

In [12] sono stati analizzate varie tecniche di mining per processi workflow, nella prospettiva più generale del problema del mining di grafi in presenza di vincoli.

In [43] è stato definito un approccio per la scoperta di outlier all'interno di un log di processo dato, basato sul clustering delle tracce e sull'identificazione di pattern strutturali.

In [53] è stato affrontato il problema di indurre un modello di processo multi-prospettiva da un log di esecuzione di un processo, traendo vantaggio dalle informazioni (solitamente presenti in tali log) su aspetti non strutturali dell'esecuzione delle attività (e.g., parametri, esecutori, metriche di performance). In particolare, la rappresentazione del processo, ottenuta come risultato, combina un insieme di modelli descrittivi workflow-oriented e un modello predittivo centrato su proprietà non strutturali del processo. A tale scopo, oltre a scoprire un insieme di gruppi di esecuzioni strutturalmente omogenei, corrispondenti a diversi scenari di esecuzione del processo, l'approccio induce un albero di decisione che esprime la dipendenza dei vari scenari di esecuzione rispetto agli aspetti non-strutturali. Allo scopo di migliorare la qualità dei cluster scoperti, l'approccio è stato concepito in modo da partizionare le istanze sulla base di pattern strutturali graph-like, facendo attenzione alla presenza di tracce di esecuzione anomale o errate (individuate mediante un metodo di outlier-detection). Inoltre, per quanto riguarda l'induzione dell'albero di decisione, sono stati estesi i metodi classici in modo da preferire, nella scelta degli attributi, le proprietà (non strutturali) del processo che tendono a essere note nelle prime fasi di esecuzione di un'istanza. In tal modo, il modello di classificazione scoperto potrà essere utile a run-time, per prevedere efficacemente e tempestivamente la struttura di esecuzione di un'istanza del processo in corso di esecuzione.

In [51] è stato infine sviluppato un framework di supporto all'analisi di processi collaborativi spontanei nell'ambito di aziende virtuali, volto a fornire meccanismi per acquisire, consolidare e condividere informazioni al fine di assicurare un adeguato livello di interoperabilità semantica. In particolare, per ottenere un'applicazione versatile ed efficace delle tecniche di process mining sui log di esecuzione di tali processi, è stato sviluppato un metodo di pre-processing che permette di ristrutturare le informazioni in una forma workflow-oriented, sfruttando la disponibilità conoscenza sul contesto organizzativo ed applicativo per analizzare gli stessi log secondo differenti prospettive di analisi, ed a differenti livelli di astrazione.

In **Error! Reference source not found.** è stato definito un framework constraint-based per il process discovery, in cui la scoperta di un grafo di dipendenze fra attività è formulata come un problema di Constraint Satisfaction o di Constraint Satisfaction Optimization, in grado di incamerare eventuale conoscenza di background sulle dinamiche del processo, codificata sotto forma di vincoli di precedenza. Per inciso, la scoperta di un grafo di dipendenze fra attività ricorre come passo fondamentale nella maggioranza

degli approcci di process discovery proposti in letteratura, preliminare all'estrazione di modelli di esecuzione più espressivi (e.g., Petri-net).

Oltre ad aver studiato teoricamente la complessità computazionale dei due tipi di problemi sopra citati, in [3] sono state definite delle soluzioni algoritmiche per entrambi, ed è stato sviluppato un sistema prototipale che implementa l'intero approccio sfruttando una piattaforma consolidata di constraint programming. E' bene osservare che, l'approccio proposto in [3] costituisce un importante avanzamento rispetto allo stato dell'arte nel campo del process mining, in cui gli algoritmi di process discovery, oltre a trascurare l'eventuale presenza di conoscenza a-priori, solitamente affidano l'individuazione delle dipendenze fra attività a metodi euristici ad hoc (di tipo greedy) e si fondano sull'assunzione che i dati di log disponibili siano completi (i.e., coprano tutto i possibili pattern di comportamento del processo che li ha generati) – rischiando, così, di incorrere in risultati di scarsa qualità, nel caso di processi con un elevato grado di concorrenza e di log affetti da “rumore”.

4. Attività Didattica

Nell'anno accademico 2010-2011 è stato docente del corso di *Tecniche di Programmazione* del corso di Laurea in Ingegneria Gestionale della Facoltà di Ingegneria dell'Università degli Studi della Calabria.

Durante l'anno 2010 ha svolto attività di docenza per il modulo “B2.2 - Workflow e Processi”, nell'ambito del corso di formazione post-lauream “Azione di Formazione per OpenKnowTech, Laboratorio di Tecnologie Open Source per la Integrazione, Gestione e Distribuzione di Dati, Processi e Conoscenze”, tenutosi presso l'Università della Calabria.

Negli anni 2008 e 2009 è stato docente dei corsi “Process Mining I” e “Complex Data Mining e Workflow Mining”, per due master di primo livello, svolti presso l'Università di Pisa e presso l'Università LUISS Guido Carli di Roma.

Nell'anno accademico 2007/2008 è stato docente dei corsi di *Sistemi di Elaborazione in Rete* della Facoltà di Ingegneria dell'Università degli Studi della Calabria, e di *Sistemi Ipermediali e Information Retrieval* della Facoltà di Lettere dell'Università degli Studi della Calabria.

Nell'anno accademico 2006/2007 è stato docente dei corsi di *Sistemi di Elaborazione in Rete* della Facoltà di Ingegneria dell'Università degli Studi della Calabria, e di *Sistemi Ipermediali e Information Retrieval* della Facoltà di Lettere dell'Università degli Studi della Calabria.

Nell'anno accademico 2005/2006 è stato docente dei corsi di *Introduzione all'Informatica*, *Sistemi di Elaborazione in Rete* della Facoltà di Ingegneria dell'Università degli Studi della Calabria, e di *Sistemi Ipermediali e Information Retrieval* della Facoltà di Lettere dell'Università degli Studi della Calabria.

Nell'anno accademico 2004/2005 è stato docente dei corsi di *Introduzione all'Informatica*, *Sistemi di Elaborazione in Rete* della Facoltà di Ingegneria dell'Università degli Studi della Calabria, e di *Sistemi Ipermediali e Information Retrieval* della Facoltà di Lettere dell'Università degli Studi della Calabria.

Nell'anno 2005 è stato anche docente al master del progetto di formazione “La gestione delle funzioni aziendali nell'era dell'e-business” (correlato al progetto di ricerca PON Agroindustria n° 12970, *Management of Integrated Enterprise*) per i seguenti moduli di insegnamento: *Progettazione di Sistemi Informatici, Data Warehouse, Integrazione di Dati, Linguaggi per Applicazioni Web*.

Per l'anno accademico 2003/2004 è stato docente dei corsi di *Introduzione all'Informatica, Laboratorio di Progettazione e Sviluppo di Sistemi Informatici* e *Sistemi di Elaborazione in Rete* della Facoltà di Ingegneria dell'Università degli Studi della Calabria.

Nei mesi di giugno e luglio 2003, è stato docente dei corsi di *Informatica Applicata: Data Mining e Text Mining* per il Corso di Alta Formazione in “Knowledge Manager”, organizzato dal Centro di Eccellenza in Economia e Gestione della Conoscenza dell'Università degli Studi della Calabria.

Durante l'anno accademico 2002/2003, è stato docente a contratto per i corsi di *Introduzione all'Informatica* e di *Laboratorio di Progettazione e Sviluppo di Sistemi Informatici* della Facoltà di Ingegneria dell'Università degli Studi della Calabria.

Durante l'anno accademico 2001/2002, è stato docente a contratto per il corso di *Laboratorio di Programmazione* del corso di Laurea in Ingegneria Informatica della Facoltà di Ingegneria, e per i corsi di *Fondamenti di Informatica I* e *Fondamenti di Informatica III* del corso di Laurea in Scienze del Servizio Sociale della Facoltà di Scienze Politiche, presso l'Università degli Studi della Calabria.

Durante l'anno accademico 2000/2001, è stato docente a contratto per il corso di recupero di *Fondamenti di Informatica* del corso di Laurea in Ingegneria Informatica, presso la Facoltà di Ingegneria dell'Università degli Studi della Calabria.

Durante gli anni accademici 1997/1998, 1998/1999, 1999/2000 e 2000/2001, ha svolto le esercitazioni per il corso di Fondamenti di Informatica II dei corsi di Laurea in Ingegneria Informatica ed in Ingegneria Gestionale. Durante l'anno accademico 1998/1999, ha svolto le esercitazioni per i moduli A e B di Fondamenti di Informatica II del corso di Diploma Universitario in Ingegneria Informatica, presso la Facoltà di Ingegneria dell'Università degli Studi della Calabria.

Durante gli anni accademici 1997/1998, 1998/1999 e 1999/2000 ha svolto cicli di seminari per il corso di Basi di Dati del Corso di Laurea in Ingegneria Informatica della Facoltà di Ingegneria dell'Università degli Studi della Calabria, su “Tecniche di Data Mining per l'estrazione di conoscenza da basi di dati”.

Durante gli anni accademici 1998/1999, 1999/2000, 2000/2001, 2001/2002, 2002/2003 e 2003/2004 è stato correlatore di varie tesi di laurea in Ingegneria Informatica, presso l'Università degli Studi della Calabria.

Elenco dei lavori scientifici

- [1]. L. Palopoli, L. Pontieri, D. Ursino: “Automatic and semantic techniques for scheme integration and scheme abstraction”. In *Proc. of the 10th International Conference on Database and Expert Systems Applications* (DEXA 1999), Firenze, 1999: pp. 511-520.
- [2]. L. Palopoli, L. Pontieri, D. Saccà, D. Ursino: “Progettazione semi-automatica di data warehouse di grandi dimensioni”. In *Atti del 7° Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD 1999, Como, 1999), pp. 3-17.
- [3]. L. Palopoli, L. Pontieri, G. Terracina, D. Ursino: “Semi-automatic construction of a Data Warehouse from numerous large databases”. In *Proc. of the 6th International Conference on Re-Technologies for Information Systems* (RETIS 2000), Zurich, Switzerland, 2000: pp. 55-75.
- [4]. L. Palopoli, L. Pontieri, G. Terracina, D. Ursino: “Intensional and extensional integration and abstraction of heterogeneous databases”. In *Data and Knowledge Engineering* (DKE), 35 (3), Elsevier Science: pp. 201-237 (2000).
- [5]. G. Dattilo, E. Masciari, L. Pontieri: “Un sistema per la classificazione dei bilanci aziendali”. In *Atti dell’8° Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD 2000), L’Aquila, 2000: pp. 257-270.
- [6]. G. Dattilo, S. Greco, E. Masciari, L. Pontieri: “A hybrid technique for data mining on balance-sheets”. In *Proc. of the 2nd International Conference on Data Warehousing and Knowledge Discovery* (DAWAK 2000), Greenwich, Great Britain, 2000: pp. 419-424.
- [7]. S. Greco, E. Masciari, L. Pontieri: “Combining Different Data Mining Techniques to Improve Data Analysis”. In *Proc. of the 4th International Conference on Flexible Query Answering Systems* (FQAS 2000), Warsaw, Poland, October 25-28, 2000: pp. 455-464.
- [8]. S. Greco, E. Masciari, L. Pontieri: “Combining inductive and deductive tools for data analysis”. In *AI Communications*, 14(2), IOS Press: pp. 69-82 (2001).
- [9]. S. Greco, L. Pontieri, E. Zumpano: “Integrating and Managing Conflicting Data”. In *Proc. of the 4th International Andrei Ershov Memorial Conference "Perspectives of System Informatics"* (Novosibirsk, Russia, July 2-6, 2001 2001), *Lecture Notes in Computer Science series* (2244), Springer-Verlag, pp. 349-362 (2002).
- [10]. L. Pontieri, E. Zumpano: “Integrating possibly conflicting data”, pubblicato in *Atti del 9° Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD 2001), Venezia, 2001: pp. 241-248.
- [11]. S. Greco, L. Pontieri, E. Zumpano: “A technique for information system integration”. In *Proc. of the 1st International Conference on Information Systems Technology and its Applications* (ISTA 2001), Kharkiv, Ukraine, June 13-15, 2001, *Lecture Notes in Informatics*, GI-Edition: pp. 75-84.
- [12]. F. Buccafurri, L. Pontieri, D. Rosaci, D. Saccà: “Improving Range Query Estimation on Histograms”. In *Proc. of the 18th International Conference on Data Engineering* (ICDE 2002), San Jose, California, USA, 2002, IEEE Computer Society: pp. 628-638.
- [13]. E. Masciari, L. Pontieri: “Tecniche di Data Mining per l’analisi di bilanci aziendali”. In *Rivista di Informatica*, 31 (3), AICA, Italy, pp. 245-266 (2001).
- [14]. L. Pontieri, D. Ursino, E. Zumpano: “An approach for synergically carrying out intensional and extensional integration”. In *Proc. of the 14th International Conference on Advanced Information Systems Engineering* (CAISE 2002, Toronto, Canada, 2002), *Lecture Notes in Computer Science series* (2348), Springer-Verlag, 2002: pp. 752-756.
- [15]. L. Palopoli, L. Pontieri, G. Terracina, D. Ursino: “A novel three-level architecture for large Data Warehouses”. In *Journal of System Architecture*, 47 (11), Elsevier Science, 2002: pp. 937-958.
- [16]. S. Flesca, G. Manco, E. Masciari, L. Pontieri, A. Pugliese: “Fast Detection of XML Structural Similarity”. In *Atti del 10° Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD 2002), Portoferraio, 2002: pp. 193-207.

- [17]. P. Naggar, L. Pontieri, M. Pupo, G. Terracina, E. Virardi: "A model and a framework for defining networks of multidimensional data marts". In *Atti del 10° Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD 2002): pp. 119-126.
- [18]. S. Flesca, G. Manco, E. Masciari, L. Pontieri, A. Pugliese: "Detecting Structural Similarities between XML Documents". In *Proc. of the 5th World Wide Web and Databases Workshop* (WebDB 2002), Madison, Wisconsin, 2002: pp. 61-66.
- [19]. P. Naggar, L. Pontieri, M. Pupo, G. Terracina, E. Virardi: "A model and a toolkit for supporting incremental Data Warehouse construction". In *Proc. of 13th International Conference on Database and Expert Systems Applications* (DEXA 2002), Aix-en-Provence, France, 2002: pp. 123-132.
- [20]. L. Pontieri, D. Ursino, E. Zumpano: "An approach for the extensional integration of data sources heterogeneous representation formats". In *Data and Knowledge Engineering* (DKE), 43(5), Elsevier Science, pp. 291-331 (2003).
- [21]. M. Ettore, L. Pontieri, M. Ruffolo, P. Rullo, D. Saccà: A Prototypal Environment for Collaborative Work within a Research Organization. In *Proc. of the 14th International Workshop on Database and Expert Systems Applications* (DEXA Workshop 2003), Prague, Czech Republic, 2003: pp. 274-279.
- [22]. S. Flesca, G. Manco, E. Masciari, L. Pontieri, A. Pugliese: Fast Detection of XML Structural Similarity. In *IEEE Transactions on Knowledge and Data Engineering*, 17 (2), pp. 160-175, IEEE Computer Society, (2005).
- [23]. G. Greco, A. Guzzo, L. Pontieri, D. Saccà: Mining Expressive Process Models by Clustering Workflow Traces. In *Proc. of 8th Pacific-Asia Conference* (PAKDD 2004), Sydney, Australia, May 26-28, 2004: pp. 52-62.
- [24]. G. Greco, A. Guzzo, L. Pontieri, D. Saccà: On the Mining of Complex Workflow Schemas. In *Proc. of 12th Italian Symposium on Advanced Database Systems* (SEBD 2004), S. Margherita di Pula, Italy, June 21-23, 2004: pp. 118-129.
- [25]. G. Greco, A. Guzzo, L. Pontieri, D. Saccà: An Ontology-Driven Process Modelling Framework. In *Proc. 15th Int. Conf. on Database and Expert Systems Applications* (DEXA 2004), Zaragoza, Spain, August 30-September 3, 2004, *Lecture Notes in Computer Science* 3180, Springer, 2004: pp. 13-23.
- [26]. G. Greco, A. Guzzo, L. Pontieri: Mining Hierarchies of Models: From Abstract Views to Concrete Specifications. In *Proc. 3rd Int. Conf on Business Process Management* (BPM 2005), Nancy, France, September 2005. *Lecture Notes in Computer Science*, 3649, pp. 32-47.
- [27]. S. Flesca, G. Manco, E. Masciari, L. Pontieri, A. Pugliese: Exploiting Structural Similarity For Effective Web Information Extraction, in *Foundations of Semistructured Data*, Dagstuhl Seminar Proceedings (Nr. 05061), IBFI, Germany, 2005, <http://drops.dagstuhl.de/opus/volltexte/2005/230/>.
- [28]. G. Manco, E. Cesario, F. Folino, L. Pontieri: An Incremental Clustering Scheme for Duplicate Detection in Large Databases. In *Proc of 9th Int. Database Applications & Engineering Symposium* (IDEAS 2005), Montreal, Canada, July 25 - 27, 2005.
- [29]. A. Congiusta, G. Greco, A. Guzzo, G. Manco, L. Pontieri, D. Saccà, D. Talia: A Data Mining-based Framework for Grid Workflow Management. In *Proc. of 5th Intl. Conference on Quality Software* (QSIC 2005), Melbourne, Australia, September 19-20, 2005 (presented at the 1st Intl. Workshop on Grid and Peer-to-Peer based Workflows, associated with the conference). Edited by Kay-Yuan Cai, Atsushi Ohnishi, M.F. Lau, IEEE Computer Society, ISBN0-7695-2472-9. pp 349-356.
- [30]. G. Greco, A. Guzzo, G. Manco, L. Pontieri, D. Saccà: Mining Constrained Graphs: The Case of Workflow Systems. In *Constraint-Based Mining, Lecture Notes in Artificial Intelligence*, 3848, pp. 155-171, Springer-Verlag, 2005.
- [31]. F. Folino, G. Manco, L. Pontieri: Effective Incremental Clustering for Duplicate Detection in Large Databases, In *Atti del 14° Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD 2006), Portonovo, Ancona, 2006, pp. 93-104.

- [32]. S. Flesca, G. Manco, E. Masciari, L. Pontieri, A. Pugliese: “Exploiting structural similarity for effective Web information extraction”. In *Data and Knowledge Engineering (DKE)*, 60 (1), Elsevier Science: pp. 222-234 (2007).
- [33]. F. Folino, G. Manco, L. Pontieri: Effective Incremental Clustering for Duplicate Detection in Large Databases. In Proc of 10th Int. Database Applications & Engineering Symposium (IDEAS 2006): 45-52, 11-14 December 2006, Delhi, India.
- [34]. G. Greco, A. Guzzo, L. Pontieri, D. Saccà: Discovering Expressive Process Models by Clustering Log Traces. In *IEEE Transactions on Knowledge and Data Engineering*, 18 (8), pp. 1010-1027, IEEE Computer Society (2006).
- [35]. A. D. Chiaravalloti, G. Greco, A. Guzzo, L. Pontieri: An Information-Theoretic Framework for High-Order Co-Clustering of Heterogeneous Objects . In Proc. of the 17th European Conference on Machine Learning (ECML 2006), Berlin, Germany, September 18-22, 2006, *Lecture Notes in Computer Science*, 4212, Springer 2006: pp. 598-605.
- [36]. A. D. Chiaravalloti, G. Greco, A. Guzzo, L. Pontieri: An Information-Theoretic Framework for Process Structure and Data Mining. In Proc. of 8th Intl. Conf. on Data Warehousing and Knowledge Discovery (DaWaK 2006), Krakow, Poland, September 4-8, 2006, *Lecture Notes in Computer Science 4081*, Springer 2006: pp. 248-259.
- [37]. F. Buccafurri, G. Lax, D. Saccà, L. Pontieri, D. Rosaci: Enhancing histograms by tree-like bucket indices. *The International Journal on Very Large Data Bases (VLDB Journal)* 17(5): 1041-1061. Springer Berlin/Heidelberg (2008).
- [38]. G. Greco, A. Guzzo, L. Pontieri: An Information-Theoretic Framework for Process Structure and Data Mining. *International Journal of Data Warehousing and Data Mining*, 3 (4), pp. 99-119, IGI Publishing (2007).
- [39]. G. Greco, A. Guzzo, L. Pontieri, D. Saccà: Integrating Advanced Data Mining Techniques into an Open-Source Framework for the Analysis of Process Logs. In Proc. of 1^a Conferenza Italiana sul Software Libero, 11-13 Maggio 2007, Cosenza, Italia. Hacklab Cosenza & Università della Calabria & Free Software Foundation Europe (2007), pp. 177-189.
- [40]. F. Folino, G. Greco, A. Gualtieri, A. Guzzo, L. Pontieri: Knowledge Discovery and Classification of Cooperation Processes for Internetworked Enterprises. In Proc. of the IV Italian Conference of the Italian Chapter of AIS (itAIS 2007), Venezia, Italy, 2007, pp. 327-334.
- [41]. G. Greco, A. Guzzo, L. Pontieri: An Information-Theoretic Framework for High-Order Co-Clustering of Heterogeneous Objects. In *Atti del 15° Convegno Nazionale su Sistemi Evoluti per Basi di Dati (SEBD 2007)*, Torre Canne di Fasano, Italy, 17-20 Giugno 2007, *Lecture Notes in Computer Science 4994 (ISBN 978-3-540-68122-9)*, Springer, 2008, pp. 397-404.
- [42]. F. Folino, G. Greco, A. Gualtieri, A. Guzzo, L. Pontieri: Knowledge Discovery and Classification of Cooperation Processes for Internetworked Enterprises. In: A. D’Atri, M. De Marco, N. Casalino. *Interdisciplinary Aspects of Information Systems Studies (ISBN 978-3-7908-2009-6)*, Physica-Verlag, Germany, 2008, pp. 327-334.
- [43]. L. Ghionna, G. Greco, A. Guzzo, L. Pontieri: Outlier Detection Techniques for Process Mining Applications. In Proc. of the 17th Intl Symposium on Foundations of Intelligent Systems (ISMIS 2008), Toronto, Canada, May 20-23, 2008: pp. 150-159
- [44]. F. Folino, G. Greco, A. Guzzo, L. Pontieri: Discovering Multi-perspective Process Models. In Proc. of the 10th Intl Conference on Enterprise Information Systems (ICEIS 2008), 12-16 June 2008, Barcelona, Spain. Volume AIDSS (ISBN 978-989-8111-37-1): pp. 70-77.
- [45]. G. Greco, A. Guzzo, L. Pontieri: Mining Taxonomies of Process Models. *Data & Knowledge Engineering*, 67 (1), Elsevier Science: pp. 74-102 (2008).
- [46]. S. Basta, F. Folino, A. Gualtieri, M- A. Mastratisi, L. Pontieri: A Knowledge-Based Framework for Supporting and Analysing Loosely Structured Collaborative Processes. In P. Atzeni, A. Caplinskas,

- H. Jaakkola (Eds.): *Advances in Databases and Information Systems, Proceedings of the 12th East European Conference, ADBIS 2008*, September 5-9, 2008, Pori, Finland. Tampere University of Technology (ISBN 978-952-15-2014-3), 2008, pp. 140-145.
- [47]. L. Ghionna, G. Greco, A. Guzzo, L. Pontieri: *Outlier Detection Techniques for Process Mining Applications*. In *Proc. of the 16th Italian Symposium on Advanced Database Systems (SEBD 2008)*, 22-25 June 2008, Mondello, PA, Italy: pp. 263-270.
- [48]. F. Folino, G. Greco, A. Guzzo, L. Pontieri. *Discovering Multi-perspective Process Models: The Case of Loosely-Structured Processes*. *Enterprise Information Systems, 10th Intl Conference, ICEIS 2008, Revised Selected Papers, LNBIP 19*: pp. 130–143. Springer-Verlag Berlin Heidelberg (2009).
- [49]. G. Greco, A. Guzzo, L. Pontieri. *Co-Clustering Multiple Heterogeneous Domains: Linear Combinations and Agreements*. *IEEE Transactions on Knowledge and Data Engineering* 22(12): 1649-1663 (2010).
- [50]. F. Folino, G. Greco, A. Guzzo, L. Pontieri: *Discovering expressive process models from noised log data*. In *Proc. of the 13th International Database Engineering and Applications Symposium (IDEAS 2009)*, 16-18 September 2009, Cetraro, Italy: pp. 162-172.
- [51]. A. Cuzzocrea, F. Folino, L. Pontieri: *Effective analysis of flexible collaboration processes by way of abstraction and mining techniques*. In *Proc. of the 12th Intl Conference on Enterprise Information Systems (ICEIS 2008)*, 8 - 12 June, 2010, Funchal, Madeira, Portugal.
- [52]. F. Folino, G. Greco, A. Guzzo, L. Pontieri: *Scalable parallel co-clustering over multiple heterogeneous data types*. In *Proc. of the 2010 Intl. Conf. on High Performance Computing & Simulation, HPCS 2010*, June 28 - July 2, 2010, Caen, France. *IEEE 2010 (ISBN 978-1-4244-6828-7: 529-535)*: pp. 529-535.
- [53]. F. Folino, G. Greco, A. Guzzo, L. Pontieri. *Mining usage scenarios in business processes: Outlier-aware discovery and run-time prediction*. *Data and Knowledge Engineering*, Nr. 70 (Vol. 12): pp. 1005-1029 (2011).
- [54]. G. Greco, A. Guzzo, L. Pontieri. *Process Discovery with Precedence Constraints*. In *Proceedings of the biennial European Conference on Artificial Intelligence (ECAI 2012)*, Montpellier, France, August 27-31, 2012: pp. 366-371.
- [55]. F. Folino, M. Guarascio, L. Pontieri. *Discovering Context-Aware Models for Predicting Business Process Performances*. In *Proceedings of the 20th International Conference on Cooperative Information Systems (CoopIS 2012)*, 10-14 Sept 2012, Roma, Italy: pp. 287-304.
- [56]. F. Folino, M. Guarascio, L. Pontieri. *Context-Aware Prediction on Business Process Executions*, in *Proceedings of the first Workshop on New Frontiers in Mining Complex Patterns (NFMCP 2012)*, Bristol UK, September 24, 2012 (in conjunction with ECML-PKDD 2012), LNAI 7765, pp. 215–229, Springer, 2013.

Rende, 23 luglio 2013

In fede
