

**Curriculum dell'Attività Scientifica  
Didattica e Professionale**  
di *Fabrizio Angiulli*

## Contents

<b>1</b>	<b>Dati personali</b>	<b>2</b>
<b>2</b>	<b>Note biografiche, formazione e posizioni ricoperte</b>	<b>2</b>
<b>3</b>	<b>Attività di ricerca</b>	<b>3</b>
3.1	Breve cronologia dell'attività di ricerca . . . . .	3
3.2	Riassunto dei principali contributi scientifici . . . . .	4
<b>4</b>	<b>Organizzazione, revisione e partecipazione a riviste e conferenze</b>	<b>15</b>
4.1	Invited talks e tutorials . . . . .	15
4.2	Partecipazione a editorial board di riviste . . . . .	15
4.3	Partecipazione a comitati organizzatori di conferenze scientifiche . . . . .	15
4.4	Partecipazione a comitati organizzatori e/o di programma di conferenze scientifiche . . . . .	15
4.5	Attività di revisione per riviste . . . . .	17
4.6	Attività di revisione per conferenze . . . . .	18
<b>5</b>	<b>Progetti di ricerca ed innovazione</b>	<b>20</b>
5.1	Attività di revisione di progetti di ricerca ed innovazione . . . . .	20
5.2	Responsabilità di progetti di ricerca ed innovazione . . . . .	20
5.3	Partecipazione a progetti di ricerca ed innovazione . . . . .	21
<b>6</b>	<b>Attività didattiche</b>	<b>24</b>
6.1	Corsi Universitari . . . . .	24
6.2	Partecipazione a Dottorati di Ricerca e Supervisione di studenti di Dottorato di ricerca . . . . .	28
6.3	Relatore di tesi di Laurea . . . . .	29
<b>7</b>	<b>Attività di servizio nell'Università e negli Enti di Ricerca</b>	<b>30</b>
<b>8</b>	<b>Elenco delle Pubblicazioni</b>	<b>32</b>

## 1 Dati personali

**Data e luogo di nascita:** 24 febbraio 1970, Torino, Italia

**Nazionalità:** italiana

**Posizione rivestita:** Professore Universitario di ruolo di I fascia di Sistemi di Elaborazione delle Informazioni presso l'Università della Calabria, Italia

**Email:** fabrizio.angiulli@unical.it

## 2 Note biografiche, formazione e posizioni ricoperte

- Fabrizio Angiulli è nato a Torino il 24 febbraio 1970.
- Il 25 maggio 1993 ha conseguito il diploma della *Scuola Diretta a Fini Speciali in Informatica* con lode presso l'Università degli Studi della Calabria, discutendo la tesi dal titolo “Il Sistema Iris per la Gestione di Basi di Dati ad Oggetti”.
- Dal gennaio 1993 al dicembre 1994 è stato Borsista di Ricerca presso l'*Istituto per la Sistemistica e l'Informatica del Consiglio Nazionale delle Ricerche (ISI-CNR)*, essendo risultato vincitore di una *borsa di studio di ricerca biennale*.
- Il 23 febbraio 1996 ha conseguito il Diploma Universitario in Ingegneria Informatica e Automatica con lode presso l'Università della Calabria, discutendo la tesi dal titolo “Il Sistema GPRS: un'Architettura per l'Integrazione di Regole Attive e Deduttive”.
- Il 20 ottobre 1999 ha conseguito la Laurea in Ingegneria Informatica con lode presso l'Università della Calabria, discutendo la tesi dal titolo “Metaquerying e sue Varianti per l'Estrazione di Conoscenza da Basi di Dati”.
- Il 14 gennaio 2000 ha ottenuto l'abilitazione all'esercizio della professione di Ingegnere.
- Nel novembre 2000 ha vinto un concorso pubblico per titoli ed esami per un posto di Collaboratore Tecnico degli Enti di Ricerca, codice settore CS196/1, assegnato all'Istituto per la Sistemistica e l'Informatica del Consiglio Nazionale delle Ricerche (ISI-CNR, poi ICAR-CNR “Istituto di Calcolo e Reti ad Alte Prestazioni” dopo la riorganizzazione del CNR). Ha preso servizio in data 24 gennaio 2001 ed ha ricoperto tale posizione sino al 31 agosto 2006.
- Nel mese di luglio 2006 è risultato vincitore di un concorso per un posto di Ricercatore Universitario, settore scientifico-disciplinare ING-INF/05 – Sistemi di elaborazione delle informazioni, presso la Facoltà di Ingegneria dell'Università della Calabria (UNICAL).
- Il primo settembre 2006 ha preso servizio come Ricercatore Universitario presso la stessa Facoltà. A decorrere dal primo settembre 2009 è stato immesso nella fascia dei Ricercatori Universitari Confermati a Tempo Pieno per il SSD ING-INF/05.
- A partire dal primo settembre 2006 afferisce al DIMES (Dipartimento di Informatica, Modellistica, Elettronica e Sistemistica) dell'Università della Calabria, precedentemente DEIS (Dipartimento di Elettronica, Informatica e Sistemistica).
- Il 13 luglio 2010 è risultato idoneo alla copertura di un posto di Professore Universitario di ruolo di II fascia per il SSD ING-INF/05 al termine di una valutazione comparativa indetta presso la Facoltà di Scienze MM.FF.NN. dell'Università degli Studi della Basilicata.
- Il 27 dicembre 2011 è stato dichiarato candidato “più qualificato” a svolgere le funzioni didattiche e scientifiche di Professore Universitario di ruolo di II fascia per il settore concorsuale “09/H1 – Sistemi di elaborazione delle informazioni”, SSD ING-INF/05, al termine di una valutazione comparativa indetta ai sensi dell'art. 18, comma 1, della Legge 30 dicembre 2010, n. 240, presso la Facoltà di Ingegneria dell'UNICAL.

- In data 30 dicembre 2011 ha preso servizio come Professore Universitario di ruolo di II fascia presso la Facoltà di Ingegneria dell'UNICAL.
- In data 3 dicembre 2013 ha conseguito l'Abilitazione Scientifica Nazionale alla funzione di Professore Universitario di I fascia per il Settore Concorsuale "09/H1 – Sistemi di elaborazione delle informazioni", al termine dei lavori della Commissione giudicatrice della procedura indetta con D.D. MIUR n. 222 del 20 luglio 2012 – GURI n. 58 del 27/07/2012. Il giudizio collegiale è stato unanime nell'esprimere valutazione *eccellente* sia della qualità delle pubblicazioni scientifiche che della loro collocazione editoriale.
- In data 31 dicembre 2018 ha preso servizio come Professore Universitario di ruolo di I fascia presso il Dipartimento DIMES dell'UNICAL.
- È Senior Member della IEEE Computer Society (Institute of Electrical and Electronics Engineers) a partire dal 2012, membro della ACM (Association for Computing Machinery) a partire dal 2011, membro del MLDM.it (Italian Working group on Machine Learning and Data Mining) a partire dal 2012 e membro del CVPL (Associazione Italiana per la ricerca in Computer Vision, Pattern recognition e machine Learning, ex GIRPR–Gruppo Italiano Ricercatori in Pattern Recognition) a partire dal 2016.

## 3 Attività di ricerca

### 3.1 Breve cronologia dell'attività di ricerca

I suoi principali interessi di ricerca si collocano nell'ambito del data mining e machine learning, della analisi e gestione di grandi quantità di dati e dell'intelligenza artificiale. È interessato allo sviluppo di metodologie efficaci ed algoritmi efficienti per l'analisi e la gestione dei dati ed allo studio delle proprietà computazionali di tecniche sviluppate in tali contesti.

Durante il periodo di svolgimento della tesi di Laurea in Ingegneria Informatica presso il Dipartimento di Elettronica Informatica e Sistemistica dell'Università della Calabria si è occupato di tecniche di data mining per basi di dati, nello specifico della tecnica del metaquerying. Particolare attenzione è stata rivolta alla caratterizzazione della complessità computazionale di tale tecnica. L'attività della sua tesi ha meritato la pubblicazione negli atti dell'International Symposium on Database Systems ([103]), la conferenza di elezione del settore delle basi dati, oltre che in ulteriori sedi di rilievo (si vedano a tal proposito i lavori [104, 99, 39]).

Durante la sua permanenza presso l'Istituto di Calcolo e Reti ad Alte Prestazioni (2001–2006) ha continuato ad interessarsi di tecniche di data mining. In particolare, i lavori [102, 38, 34], aventi per oggetto la caratterizzazione computazionale dell'induzione di regole associative ed algoritmi efficienti per il metaquerying, costituiscono la naturale continuazione di ricerche iniziate precedentemente. Nel contempo i suoi interessi hanno abbracciato altri settori del data mining e delle basi di dati e di conoscenza, quali l'outlier detection in database ([101, 90, 89, 37, 33]) ed in basi di conoscenza ([98, 97, 96, 92, 91, 94, 84]), la classificazione ([86, 88]), il clustering ([93, 87, 83, 82, 28]) e le closest pairs ([100, 95, 35, 36]).

A partire dal suo arrivo presso l'Università della Calabria (2006–oggi) ha proseguito l'attività di investigazione di tecniche innovative di outlier detection in basi di dati ([76, 77, 73, 74, 69, 25, 26, 67, 68, 23, 16, 11]) e basi di conoscenza ([80, 29, 27, 21, 22, 62]), nonché di tecniche innovative di classificazione ([78, 32, 31, 30, 75, 81, 79, 72, 71, 70, 24, 20, 5]). Inoltre, ha rivolto la sua attenzione al progetto di metodi paralleli/distribuiti per l'analisi dei dati ([78, 30, 71, 64, 17]), alla problematica dell'outlier explanation ([85, 25, 18]), ai data streams ([77, 66, 23, 65, 63]), all'analisi e gestione di dati incerti ([19, 16]), alle tecniche di indicizzazione in spazi metrici ([19]), al calcolo dei modelli minimali di CNF ([13, 107]), alla cyber security ([52, ?, 8, 7]), alla bioinformatica ([108]) ed allo studio delle proprietà degli spazi ad alta dimensionalità ([6, 47, 11]).

Gran parte della sua produzione scientifica appare in riviste di comprovato prestigio internazionale — tra cui Journal of Machine Learning Research, ACM Transactions on Knowledge Discovery

from Data, ACM Transactions on Database Systems, ACM Transactions on Computational Logic, ACM Transactions on Internet Technology, IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Neural Networks, IEEE Transactions on Parallel and Distributed Systems, Artificial Intelligence Journal, Data Mining and Knowledge Discovery, Theoretical Computer Science, ed altre ancora — ed in conferenze internazionali di eccellenza nei rispettivi settori — tra cui Symposium on Principles of Database Systems (PODS), International Joint Conference on Artificial Intelligence (IJCAI), AAAI Conference on Artificial Intelligence (AAAI), International Conference on Machine Learning (ICML), International Conference on Data Mining (ICDM), International Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), ed altre ancora.

Diversi lavori scientifici sono a nome singolo ed in numerosi casi svolge il ruolo di coordinatore della linea di ricerca. In tutti i casi, il contributo personale ai lavori di cui è coautore è almeno pari a quello degli altri autori.

## 3.2 Riassunto dei principali contributi scientifici

### Studio del fenomeno della concentrazione delle distanze e dell'hubness

Il termine *curse of dimensionality* è ampiamente utilizzato per riferirsi ad una serie di fenomeni che vengono osservati al crescere della dimensionalità, ovvero del numero di attributi o features, dei dati. Tali fenomeni determinano molto spesso un scadimento delle prestazioni degli algoritmi di analisi e gestione dei dati. In particolare, la *concentrazione delle distanze* è il fenomeno per cui la differenza relativa tra massima e la minima distanza tra coppie di oggetti tende ad assottigliarsi col crescere della dimensionalità. Questo fenomeno pone grossi problemi, ad esempio, agli algoritmi di indicizzazione e ricerca dei nearest neighbors e agli algoritmi di stima della densità. Sebbene sia noto che la citata differenza sia destinata a svanire al tendere all'infinito della dimensionalità intrinseca, in letteratura sono presenti rare caratterizzazioni delle distribuzioni delle distanze per dimensionalità alte, ma comunque finite.

In questo scenario, in [6] viene derivata la forma chiusa della distribuzione delle distanze tra coppie di oggetti (punti nello spazio Euclideo), tra un oggetto dato ed ogni altro oggetto e la distanza attesa dal  $k$ -esimo nearest neighbor, nel caso di dati a dimensionalità intrinsecamente elevata generati da vettori casuali aventi distribuzione arbitraria. Viene inoltre studiata la distribuzione del numero atteso di nearest neighbors  $\epsilon$ -approssimati, utile a caratterizzare la cosiddetta dimensionalità critica ovvero la dimensionalità per cui i vicini  $\epsilon$ -approssimati coincidono con la quasi totalità dei punti.

Un altro importante fenomeno legato alla curse of dimensionality è quello dell'*hubness*. Questo fenomeno può essere descritto come segue: all'aumentare della dimensionalità si verifica l'emergenza di *hub*, ovvero di un numero ristretto punti che risultano essere vicini (nearest neighbors) di tutti gli altri punti, mentre questi ultimi punti, che rappresentano la quasi totalità dei dati, risultano essere tutti distanti tra loro. A tal proposito, in [6] viene derivata la distribuzione del numero  $N_k$  di *reverse nearest neighbors*, ovvero del numero di punti che risultano avere il punto dato tra i loro  $k$  nearest neighbors, nel caso di dati a dimensionalità intrinsecamente elevata. La derivazione della forma chiusa di  $N_k$  è un problema aperto da tempo e la sua conoscenza fornisce la completa interpretabilità del fenomeno dell'hubness.

Il fenomeno dell'hubness è ben presente nelle reti sociali in cui, tipicamente un numero esiguo di nodi risulta essere "popolare", ovvero presenta un numero elevato di archi incidenti. La coda della distribuzione del numero di archi incidenti in una rete sociale viene normalmente modellata attraverso una *power law*. In [6] viene sviluppata un'analisi su reti sociali reali, quali il *WWW*, *Twitter* e *arXiv*, che rivela che la forma chiusa della funzione  $N_k$  ottenuta è adatta a fornire un modello per la distribuzione degli archi entranti delle reti sociali non solo sulla coda della distribuzione, ma sull'intero dominio.

Il *Concentration Free Outlier Factor* (CFOF) introdotto in [47, ?] è una misura che mira a superare il problema della concentrazione delle distanze nell'ambito della stima di densità e dell'identificazione delle anomalie, il cui comportamento è legato alla funzione  $N_k$ . In particolare,

dato un parametro  $\varrho \in [0, 1]$  che rappresenta una frazione della popolazione dei dati, il CFOF associato ad un generico punto  $x$  è definito come  $\min\{k/n : N_k(x) \geq n\varrho\}$ , ovvero come il più piccolo valore per il parametro  $k$  per cui  $x$  presenta almeno  $n\varrho$  nearest neighbors. L'intuizione è che i punti isolati richiedono valori di  $k$  più elevati per essere selezionati come vicini da una frazione della popolazione di pari entità. Al contrario di tutte le nozioni di outlier note, CFOF non esibisce alcuna concentrazione al crescere della dimensionalità. Sfruttando la forma chiusa della funzione  $N_k$  derivata in [6] è possibile mostrare formalmente che i CFOF outlier sono pochi e separati dagli inlier anche in spazi ad elevata dimensionalità intrinseca, mentre l'uso diretto di  $N_k$  per isolare gli outlier è soggetto ad un elevato numero di falsi positivi.

### Similarity indexing and searching

In [19] viene considerato il problema del calcolo efficiente di range queries su oggetti incerti in spazi metrici. In questo studio, un oggetto incerto è un oggetto il cui valore viene modellato da una funzione di densità di probabilità. Il contributo principale del lavoro è di aver fornito la prima tecnica per indicizzare oggetti incerti in spazi metrici generici. Questo è stato possibile attraverso la generalizzazione della disuguaglianza triangolare al contesto probabilistico, al fine di poterla utilizzare come condizione per il taglio dello spazio di ricerca, e la conseguente introduzione di una tecnica, chiamata UP-index (Uncertain Pivot index), che la sfrutta per velocizzare il calcolo della range query. La tecnica UP-index utilizza il concetto di pivot, o punto cardine, per indicizzare i dati. I risultati sperimentali hanno evidenziato che la tecnica UP-index è più efficiente delle tecniche allo stato dell'arte, anche se specificatamente progettate per lo spazio Euclideo.

Viene inoltre introdotto un criterio per misurare la qualità di un insieme di pivot e viene, quindi, studiato il problema di selezionare un buon insieme di pivot in accordo al criterio introdotto. L'analisi di complessità rivela che selezionare l'insieme ottimo di pivot è un problema NP-hard sia in uno spazio metrico generico che nello spazio Euclideo. L'analisi condotta vale anche nel caso in cui gli oggetti siano certi. Lo studio della complessità di questo importante problema non era stata precedentemente affrontata in letteratura, nemmeno nel caso certo.

In [60] viene presentato l'algoritmo PPP, acronimo di *Principal directions-based Pivot Placement*, per la generazione di un insieme di pivot in grado di migliorare le prestazioni del passo di *candidate selection* degli algoritmi di indexing basati su pivot. Al fine di ottimizzare il potere discriminante dei pivot, l'algoritmo seleziona un numero fissato di direzioni dello spazio che massimizzano il numero di coppie di oggetti ad esse allineate.

### Outlier detection

**Nuove definizioni ed algoritmi per distance-based outliers.** I *distance-based outliers* rappresentano una una delle principali famiglie di tecniche per l'individuazione non supervisionata di anomalie in grandi quantità di dati. Queste tecniche hanno l'interessante proprietà di rappresentare una generalizzazione di alcuni test statistici ed, inoltre, non richiedono la conoscenza della distribuzione dei dati. Tuttavia i primi algoritmi presentati richiedevano tempo quadratico nel numero di oggetti oppure tempo esponenziale nel numero di attributi, che risultano inaccettabili nel caso cui si debbano analizzare grandi quantità di dati.

I lavori [37, 101] presentano una nuova definizione di outlier di tipo distance-based ed un algoritmo, chiamato *HilOut*, progettato per individuare i top  $n$  outlier, vale a dire gli  $n$  oggetti che più deviano dalla normalità, in dataset multidimensionali (decine o centinaia di attributi) ed estesi (centinaia di migliaia o milioni di oggetti). La definizione presentata ha il vantaggio di tenere in conto le fluttuazioni di densità dei dati nell'intorno di ogni oggetto. L'algoritmo *HilOut* si basa sull'utilizzo delle *curve space-filling* e garantisce una soluzione approssimata con costo temporale al più quadratico nel numero di attributi e lineare nel numero di oggetti del dataset. Durante la sua esecuzione, l'algoritmo è in grado di stabilire se la soluzione approssimata corrente coincide con la soluzione esatta e conseguentemente di terminare in anticipo rispetto al caso peggiore. I risultati sperimentali evidenziano che nei casi pratici l'algoritmo calcola sempre la soluzione esatta. Inoltre, l'analisi di scalabilità evidenzia che l'algoritmo scala quasi linearmente.

La definizione introdotta in [101] è oggi riconosciuta dalla comunità del data mining come una delle definizioni di outliers classiche e viene confrontata, di fatto, con ogni nuova proposta di nozione di anomalia. In accordo alle misure bibliometriche dell'editore Springer, il lavoro [101] è l'articolo di gran lunga più scaricato e citato dell'edizione 2002 della conferenza PKDD, con oltre il 25% delle citazioni totali. Gli algoritmi di outlier detection descritti in [37, 101] sono stati inclusi (algoritmi `HilOut` e `KNNWeightOutlier`) nella suite di data mining ELKI "Environment for Developing KDD-Applications Supported by Index-Structures".<sup>1</sup>

**Rappresentazioni compresse per outliers.** In [33, 89, 90] viene presentato un metodo di identificazione di outlier avente due obiettivi: trovare i top  $n$  outlier di un dataset molto grande ed a molte dimensioni, e apprendere un modello sugli oggetti contenuti nel dataset, modello che può essere in seguito utilizzato per classificare nuovi oggetti come outliers o meno. Il modello proposto, chiamato *Solving Set*, è un sottoinsieme del dataset avente la proprietà che le distanze tra gli oggetti nel solving set e gli oggetti nel dataset sono sufficienti per stabilire quali siano i top  $n$  outlier. In questo lavoro è dimostrato che calcolare un solving set di cardinalità minima è un problema NP-hard e viene presentato un algoritmo che calcola un solving set, in generale, non minimo in tempo al più quadratico. I risultati sperimentali descritti mostrano che il solving set calcolato dall'algoritmo è una piccola frazione del dataset e che la qualità della classificazione del solving set è confrontabile con quella del dataset. L'accuratezza della predizione di outlier ottenuta usando tutto il dataset come modello è confrontata con l'accuratezza della predizione di outlier ottenuta usando il solving set, mostrando che il solving set non solo riduce i tempi di risposta, ma anche che ha capacità predittive paragonabili a quelle ottenibili utilizzando il dataset.

**Algoritmi paralleli/distribuiti e per Graphics Processing Unit.** In [64, 17] vengono presentati algoritmi distribuiti per il calcolo dei distance-based outliers. Tali algoritmi sfruttano la strategia introdotta in [33] al fine di ridurre notevolmente il numero di distanze da calcolare e bilanciare il carico di lavoro tra i nodi. Si tratta della strategia distribuita più efficiente introdotta in letteratura per il calcolo degli outliers definiti in [37].

In [10, 57, 58, 61] vengono descritti degli algoritmi GPU-based per il calcolo dei distance-based outliers. Tali algoritmi estendono l'approccio SolvingSet per il calcolo degli outlier allo scenario sia centralizzato che distribuito in presenza di processori many-core. Le diverse strategie presentate consentono di ottenere speedups di due ordini di grandezza rispetto ai corrispondenti algoritmi sequenziali.

**Algoritmi scalabili per dati in memoria secondaria.** In [26, 73, 76] viene presentato un algoritmo di outlier detection, denominato DOLPHIN, basato sulla definizione classica di distance-based outlier, espressamente progettato per lavorare con dataset residenti in memoria secondaria. Il costo di I/O dell'algoritmo è trascurabile, in quanto esegue solo due letture sequenziali del file in cui è memorizzato il dataset. Inoltre, viene teoricamente dimostrato che per combinazioni ben fondate dei parametri il costo spaziale corrisponde ad una piccola percentuale della dimensione del dataset (inferiore al 5%) mentre il costo temporale, misurato in termini di numero di distanze calcolate, scala linearmente rispetto al numero totale di oggetti. DOLPHIN è stato confrontato sperimentalmente con gli algoritmi stato dell'arte esibendo prestazioni nettamente superiori a questi sia dal punto di vista pratico, che da quello teorico.

**Nuove definizioni e algoritmi per data streams.** In [77, 23] viene presentato l'algoritmo STORM per la scoperta di distance-based outliers in data streams. L'algoritmo assume il modello a finestra scorrevole e, su richiesta dell'utente, restituisce gli outliers contenuti nella finestra corrente. Nello specifico, vengono introdotte tre varianti dell'algoritmo STORM, denominate STORM1, STORM2 e STORM3. L'algoritmo STORM1 restituisce la risposta esatta, ma ha necessità di allocare in memoria tutti gli oggetti della finestra corrente. Gli algoritmi STORM2 e STORM3 hanno invece un costo spaziale inferiore, in quanto mantengono in memoria solo una frazione limitata degli oggetti della finestra corrente, ma restituiscono una risposta approssimata, sebbene con garanzia statistica di accuratezza. Questi due algoritmi sfruttano il nuovo concetto di *safe*

---

<sup>1</sup>ELKI è disponibile all'url <https://elki-project.github.io>.

*inlier* per ridurre notevolmente la quantità di memoria necessaria per individuare gli outlier nella finestra corrente. Un safe inlier è un oggetto che non potrà diventare outlier in ogni evoluzione futura dello stream.

STORM (nelle versioni **ExactSTORM** e **ApproxSTORM**) è stato incluso in MOA “Machine Learning for Streams”, il più popolare ambiente open source per il data stream mining.<sup>2</sup>

**Determinazione di outlier in dati incerti.** In [?, 59] viene presentata la nuova definizione di *uncertain distance-based outlier*, che è un oggetto avente bassa probabilità di occorrere in una regione densamente popolata dei dati. Viene inoltre presentato un algoritmo articolato su tre fasi in grado di determinare in maniera efficiente gli outlier. La prima fase effettua una stima dei valori da utilizzare per i parametri della definizione, la seconda fase determina l’insieme di candidati outlier utilizzando un algoritmo per dati certi applicato alla distanza massima tra coppie di oggetti incerti e la terza fase determina gli outlier a partire dall’insieme dei candidati outlier sfruttando delle proprietà basate su *lower bound* permettono di ridurre il numero di oggetti vicini da considerare durante il calcolo dell’*outlierness*.

**Nuove definizioni di outlier.** In [11] viene presentata una nuova nozione di anomalia, chiamata GOF (Gradient Outlier Factor), che ha la proprietà di comportarsi come un test statistico nel caso in cui i dati provengano da una singola distribuzione, ma che è al contempo in grado di generalizzarne il comportamento nel caso in cui i dati provengano da distribuzioni con diverse caratteristiche. Vengono derivate le proprietà formali per cui GOF si comporta come un test statistico su una distribuzione nota e viene mostrato che queste proprietà sono proprie di importanti distribuzioni note. Viene inoltre fornita una interpretazione probabilistica della nuova nozione, basata sul concetto di densità della densità dei dati.

**Determinazione di anomale in reti sociali.** In [49] viene descritta una nuova tecnica per l’individuazione di anomalie in reti sociali in presenza di informazioni temporali riguardanti l’istante di creazione delle connessioni. La tecnica introduce un insieme di sottografi temporali che sono associati a specifici comportamenti e segnala come anomali i nodi tali che la numerosità dei sottografi temporali cui partecipano devia rispetto al comportamento normale dettato dalla corrispondente distribuzione all’interno della rete.

In occasione delle celebrazioni per il 25° *anniversario dell’Associazione Italiana per l’Intelligenza Artificiale*, Fabrizio Angiulli è stato invitato a contribuire, con i risultati delle sue ricerche in ambito outlier detection, ad un *numero speciale della rivista Intelligenza Artificiale*. Il numero in questione ripercorre i contributi più rilevanti della Ricerca Italiana all’avanzamento della disciplina [15].

## Outlier Explanation

In diverse situazioni l’individuo eccezionale è già noto e ciò che si vuole determinare sono le proprietà che lo rendono eccezionale rispetto ad una popolazione di riferimento. Ad esempio, si supponga di analizzare i dati clinici di un paziente malato e di avere a disposizione un archivio di dati clinici relativi ad una popolazione di persone sane. In questo contesto, uno strumento automatico in grado di segnalare i parametri che maggiormente caratterizzano il paziente malato rispetto alla popolazione sana risulta essere di grande aiuto per l’analista.

**Proprietà anomale di una singolo oggetto per dati categorici.** In [85, 25] viene introdotto il seguente problema: data una popolazione memorizzata in una base di dati relazionale ed un oggetto *obj* (tupla) della popolazione, che è stato segnalato come eccezionale, determinare una coppia ( $\mathcal{O}$ , in alternativa, un numero fissato di coppie) di insiemi di attributi ( $E, S$ ), tali che  $E$  rappresenta una spiegazione dell’eccezionalità di *obj* ed  $S$  rappresenta invece la proprietà che rende *obj* eccezionale.

In particolare, la spiegazione  $E$  dell’eccezionalità di *obj* ha lo scopo di isolare una sottopopolazione omogenea e rilevante dei dati di cui *obj* fa parte e rispetto alla quale *obj* è eccezionale. Tale

---

<sup>2</sup>MOA è disponibile all’url <https://moa.cms.waikato.ac.nz/>.

sottopopolazione è data nello specifico da tutte le tuple che assumono lo stesso valore assunto da  $obj$  sull'insieme di attributi  $E$ . La sottopopolazione è *rilevante* solo se è sufficientemente ampia e, cioè, composta da una frazione significativa della popolazione.

La proprietà  $S$ , invece, è un insieme di attributi che caratterizza  $obj$  rispetto alla sottopopolazione individuata da  $E$ . In particolare, affinché  $S$  possa considerarsi eccezionale, le coppie attributo-valore formate combinando gli attributi in  $S$  con i valori assunti da  $obj$  su tali attributi devono discostarsi in maniera significativa rispetto alla sottopopolazione di riferimento.

Viene quindi definita una misura per quantificare l'eccezionalità di un insieme di coppie attributo-valore (dove il valore è quello assunto dall'oggetto anomalo sotto indagine) rispetto alla distribuzione dei valori dell'insieme di attributi all'interno della popolazione di riferimento.

La misura proposta è basata sul concetto di frequenza relativa. In particolare, il punto di vista adottato è che una proprietà, o insieme di attributi, è anormale per un oggetto se la combinazione di valori che l'oggetto assume su tali attributi è poco frequente rispetto alla distribuzione delle combinazioni di valori assunti sugli stessi attributi da tutti gli oggetti della collezione. Tale misura ha l'interessante proprietà di essere collegata all'indice di Gini, utilizzato in statistica per misurare l'eterogeneità di una distribuzione statistica.

I contributi vengono completati dallo studio della complessità computazionale del problema di scoprire le proprietà eccezionali più rilevanti (ovvero che massimizzano il valore di eccezionalità), problema la cui complessità è confinata in  $F\Sigma_2^P$ , dalla determinazione di upper bounds per il valore di eccezionalità di una proprietà e dal progetto di un algoritmo di ricerca per il calcolo delle proprietà eccezionali più rilevanti.

**Proprietà anomale di gruppi di oggetti.** In [18] viene considerato il problema di determinare gli attributi, o proprietà, che caratterizzano un gruppo di individui anomali (outliers) rispetto al resto della popolazione (inliers). A tale scopo viene introdotto il concetto di exceptionality score, il cui obiettivo è misurare quanto i valori assunti dagli outliers sulla proprietà oggetto d'indagine mal si adattano ad essere generati dalla distribuzione di probabilità soggiacente i valori assunti dagli inliers sullo stesso attributo. Nel caso di proprietà numeriche, l'exceptionality score è definito come un randomization test basato sull'indice chi quadrato di Pearson, mentre per le proprietà numeriche si fa uso del test Cramér-von-Mises. Gli exceptionality score così definiti sono particolarmente adatti per confrontare una popolazione rara (formata da un numero relativamente esiguo di individui) con una popolazione di grande dimensione.

**Proprietà anomale di un singolo oggetto per dati numerici.** In [9] viene considerato il problema di determinare gli attributi che caratterizzano un singolo outlier rispetto al resto della popolazione in presenza di attributi numerici. A questo scopo viene introdotta una nuova misura di outlieriness che quantifica la differenza tra la densità di probabilità associata all'oggetto outlier e quella associata al resto della popolazione. Al fine di applicare la misura a dati numerici, viene introdotto un algoritmo efficiente, di costo  $O(n \log n)$ , per la ricostruzione della densità di probabilità dei dati basato su Kernel Density Estimation.

## Outlier Detection in Basi di Conoscenza

La letteratura riguardante l'outlier detection è principalmente collegata alla statistica, al machine learning ed al data mining e quindi, la quasi totalità degli approcci presentati in letteratura lavora con dati che possono essere organizzati in una singola tabella relazionale, spesso con tutti gli attributi di tipo numerico, ed è richiesta la definizione di una distanza tra ogni coppia di righe della tabella.

Le tecniche di outlier detection proposte nei lavori [91, 92, 94, 96, 97, 98, 29, 27, 22, 21, 12] hanno invece a che fare con basi di conoscenza complesse, che includono dati di tipo relazionale, ma in generale anche forme di conoscenza semanticamente più ricche, come assiomi, regole ed altro; diverse e complesse relazioni tra gli oggetti della teoria sono quindi disponibili ed utilizzabili.

**Outlier detection in default theories.** I lavori [98, 97, 96, 27] utilizzano la Default Logic come linguaggio per la rappresentazione della conoscenza di riferimento. La Default Logic è un formalismo basato sulla logica che è stato introdotto per ragionare in presenza di informazione



incompleta. Essa prevede un particolare tipo di regole logiche, dette *defaults*, che consentono di descrivere “come vanno le cose normalmente”.

Una teoria di default proposizionale  $\Delta$  è una coppia  $(D, W)$ , dove  $D$  è un insieme di defaults e  $W$  è una teoria logica proposizionale. Un *outlier*  $L$  in  $\Delta$  è definito come un insieme di osservazioni (o letterali) in  $W$  per cui esiste un secondo insieme di osservazioni  $S$  in  $W$ , detto *witness*, tale che (i)  $(D, W \setminus S) \models \neg S$  e (ii)  $(D, W \setminus (S \cup L)) \not\models \neg S$ . Intuitivamente, la prima condizione stabilisce che  $S$  rappresenta una proprietà eccezionale, mentre la seconda condizione stabilisce che  $S$  è una proprietà di  $L$ . L’osservazione  $L$  è quindi anomala (l’outlier), mentre  $S$  è la giustificazione della sua eccezionalità.

Nei lavori [98, 97, 96, 27, 22, 68, 14] vengono analizzate le proprietà computazionali di diversi problemi di outlier detection nel contesto delle principali classi di teorie di default proposizionali. In particolare, viene dimostrato che decidere se una teoria di default contenga o meno almeno un outlier è un problema  $\Sigma_3^P$ -completo, mentre decidere se un insieme di osservazioni è un outlier minimale è un problema  $D_3^P$ -completo. Vengono inoltre individuate classi di teorie di default per cui tale problema è risolubile in tempo polinomiale. In [12, 62] viene tracciata la frontiera di trattabilità dei problemi di outlier detection su default theories, individuando come frammento trattabile quello delle *acyclic normal unary propositional default theories*. Viene, infine, affrontato il problema della caratterizzazione dell’espressività delle normal unary default theories, dimostrando che queste teorie sono in grado di esprimere almeno tutte le queries in NL.

**Outlier detection in programmi logici.** La Programmazione Logica è uno tra i formalismi per la rappresentazione della conoscenza maggiormente diffusi e risulta, quindi, di grande interesse studiare le tecniche di outlier detection anche nel contesto dei programmi logici. In [94] la nozione di outlier introdotta nell’ambito della Default Logic viene estesa al caso in cui la conoscenza è rappresentata mediante il formalismo della Programmazione Logica Disgiuntiva sotto semantica Answer Set. In [92] viene preso in considerazione il contesto della della Programmazione Logica sotto Closed World Assumption e viene quindi presentato uno studio preliminare della complessità computazione del problema per programmi logici proposizionali. In [91] viene studiata la *data complexity* del problema dell’outlier detection per programmi logici generali. In [29] lo studio della complessità computazionale del framework introdotto in [92] viene completato. Inoltre, viene descritta un riscrittura  $\mathcal{L}$  che trasforma un programma logico  $P$  in un programma logico disgiuntivo  $\mathcal{L}(P)$  tale che esiste una corrispondenza tra gli outlier del programma  $P$  ed i modelli del programma  $\mathcal{L}(P)$ . Tale riscrittura fornisce quindi un metodo effettivo per risolvere i problemi di outlier detection utilizzando i vari sistemi di programmazione logica disponibili, come mostrato in [80], dove viene descritto un prototipo di sistema di outlier detection per programmi logici.

In [21, 84] la nozione di anomalia sopra illustrata viene applicata al contesto in cui un agente interagisce con un mondo dinamico e parzialmente noto attraverso dei sensori, al fine di individuare la presenza di eventuali sensori inaffidabili. A tale scopo, l’agente sfrutta la sua conoscenza del mondo per rilevare eventuali anomalie nei dati rilevati da un certo sensore, applicando la sua specifica nozione di outlier in basi di conoscenza. Il lavoro considera infine la complessità di diversi problemi di ricerca di anomalie nei dati rilevati dai sensori.

**Outlier Detection mediante Inductive Logic Programming.** In [14, 68] viene considerato il problema di individuare osservazioni anomale all’interno di un insieme di esempi positivi e negativi, che rappresentano un concetto, in presenza di una background knowledge composta da regole logiche del primo ordine. A tal scopo viene introdotta una nozione di *compliance* di un insieme di osservazioni, che misura quanto queste influiscono negativamente sulla generalizzazione di una ipotesi generata mediante tecniche di Inductive Logic Programming. Sfruttando tale nozione è possibile distinguere tra osservazioni *irregular*, *anomalous* e *outlier*.

## Classificazione

La *nearest neighbor rule* è una delle tecniche di classificazione più popolari. Tale tecnica classifica un nuovo oggetto assegnandolo alla classe dell’oggetto del training set – l’insieme di oggetti di esempio di cui si conosce la classe di appartenenza – che risulta essere ad esso più vicino in

accordo con una metrica prestabilita (ad esempio la distanza euclidea). Tale tecnica offre delle buone garanzie teoriche in termini di qualità della classificazione ed è semplice da implementare, poichè non richiede una fase di apprendimento, ma piuttosto di memorizzare l'intero training set. Per contro, la fase di classificazione della nearest neighbor rule può risultare molto onerosa, in quanto durante questa fase deve essere calcolata la distanza tra l'oggetto da classificare ed ogni altro oggetto del training set. L'uso di strutture per l'indicizzazione di oggetti multidimensionali può alleviare il costo della classificazione, ma è noto che quando la dimensionalità dei dati cresce le prestazioni di tali strutture possono degenerare in una scansione sequenziale di tutti i dati indicizzati.

Al fine di ridurre il tempo richiesto durante fase di classificazione ed anche la quantità di memoria richiesta per archiviare il training set, in letteratura sono state proposte diverse tecniche il cui scopo è quello ridurre il numero di oggetti che compongono il training set ed al contempo mantenere inalterata la qualità della classificazione. Tra queste, particolarmente utilizzate sono le tecniche *training set consistenti*, le quali si prefiggono di selezionare un sottoinsieme  $S$  del training set originale  $T$ , avente la proprietà di classificare correttamente gli oggetti in  $T \setminus S$  mediante la nearest neighbor rule.

In letteratura sono stati introdotti diversi metodi per calcolare un sottoinsieme training set consistente. Tra questi, la Condensed Nearest Neighbor rule (CNN), la Modified CNN rule (MCNN) e la NNSRM (NN Structural Risk Minimization). Questi algoritmi hanno una complessità temporale superquadratica o cubica e poca attenzione è stata rivolta al miglioramento delle loro prestazioni al fine di renderne pratica l'applicazione a training set composti da centinaia di migliaia o milioni di oggetti.

In [88, 31] viene presentato un nuovo algoritmo, chiamato FCNN (Fast CNN), per il calcolo di un sottoinsieme training set consistente. L'algoritmo FCNN ha alcune proprietà interessanti: il sottoinsieme calcolato non dipende dall'ordine in cui gli oggetti sono elaborati, mentre altri metodi sono sensibili a tale ordine, converge in poche iterazioni, seleziona oggetti del training set prossimi al confine tra le classi ed ha complessità al più quadratica. Inoltre, la struttura dell'algoritmo permette di sfruttare efficacemente la distanza triangolare al fine di ridurre notevolmente il numero di distanze da calcolare. FCNN è stato applicato a training set di grandi dimensioni e confrontato con i metodi esistenti, risultando più veloce di questi anche di alcuni ordini di grandezza, pur calcolando un sottoinsieme di dimensione paragonabile o più piccolo e garantendo la stessa accuratezza di classificazione. Inoltre è risultato essere di tre ordini di grandezza più veloce degli algoritmi di apprendimento di tipo *hybrid instance-based* sui datasets MNIST e MIT Face, pur calcolando un modello di accuratezza confrontabile. L'algoritmo è inoltre stato applicato con buoni risultati al problema della predizione della struttura quaternaria di proteine [75] e come passo di pre-elaborazione per le Support Vector Machines [81, 79, 24].

In [30, 78] viene presentato l'algoritmo PFCNN. PFCNN è il primo algoritmo distribuito per il calcolo di un sottoinsieme training set consistente. L'algoritmo scala quasi linearmente rispetto al numero di processori sia dal punto di vista teorico che dal punto di vista pratico. Alcune varianti dell'algoritmo, specifiche per gli ambienti di griglia, sono state introdotte in [71].

In [5, 51] viene presentata la tecnica  $\alpha$ -FCNN. Tale tecnica applica il principio del Pessimistic Error Estimate (PEE) per determinare il sottoinsieme del training set che fornisce il migliore trade-off tra la dimensione e l'attesa generalizzazione in fase di testing. Diverse sono state le strategie introdotte, TRNOPT, MAXOPT, PACOPT e aPACOPT, che sono state confrontate con algoritmi ibridi allo stato dell'arte. Il confronto rivela che  $\alpha$ -FCNN è più efficiente di tali tecniche ed in grado di determinare un sottoinsieme di dimensione inferiore o comparabile a quella delle tecniche migliori in termini di riduzione dei dati.

L'algoritmo FCNN è stato incluso (algoritmo FCNN-C) nella suite open source di knowledge data discovery KEEL "Knowledge Extraction based on Evolutionary Learning".<sup>3</sup>

*Classificazione di dati incerti.* In [16] viene considerato il problema di classificare data incerti, intesi come oggetti il cui valore è modellato mediante una funzione di densità di probabilità. A questo scopo viene introdotta la Uncertain Nearest Neighbor rule (UNN), ovvero la generalizzazione della

---

<sup>3</sup>Disponibile all'url <http://www.keel.es/>.

regola di classificazione nearest neighbor deterministica allo scenario in cui gli oggetti da classificare sono incerti. La regola UNN si basa sul concetto di nearest neighbor class, piuttosto che su quello di oggetto nearest neighbor. La nearest neighbor class dell'oggetto di test è la classe che massimizza la probabilità di fornire il suo nearest neighbor. Viene dimostrato che il concetto di nearest neighbor class cattura la vera semantica della nearest neighbor rule applicata ai dati incerti e da cui risulta la sua superiorità rispetto all'applicazione diretta di distanze incerte per determinare i nearest neighbor dell'oggetto test al fine della sua classificazione. Nel lavoro viene quindi presentato un metodo efficiente per effettuare la classificazione nearest neighbor di dati incerti.

### One-class classification

La *classificazione ad una classe* è una tecnica di classificazione il cui obiettivo è quello di separare gli oggetti di una classe fissata da tutti gli altri oggetti dello spazio. Il problema da risolvere è il seguente: dato un insieme di oggetti, detto anche *insieme di riferimento*, appartenenti ad un certo spazio, trovare una descrizione di tali oggetti, ovvero una regola che partiziona lo spazio in una *regione di accettazione*, contenente gli oggetti appartenenti alla classe rappresentata dall'insieme di riferimento, ed in un *regione di rifiuto*, contenente tutti gli altri oggetti.

In [86, 32] viene presentata la regola di classificazione ad una classe chiamata NNDD (Nearest Neighbor Domain Description). Tale regola utilizza le distanze dell'oggetto da classificare dai suoi  $k$  vicini nell'insieme di riferimento. Viene inoltre studiato l'effetto di utilizzare come insieme di riferimento del classificatore un sottoinsieme dell'insieme di riferimento originale. A tal proposito, viene introdotto il concetto di *sottoinsieme di riferimento consistente* rispetto alla regola NNDD e dimostrato che trovare il sottoinsieme di riferimento consistente di cardinalità minima è un problema intrattabile. Viene quindi introdotto un algoritmo greedy, denominato CNNDD, per il calcolo di un sottoinsieme di riferimento consistente. Tale algoritmo viene confrontato con metodi già noti in letteratura, in particolare NN-d,  $k$ -center e one-class SVM, analizzando vantaggi e svantaggi della classificazione ad una classe mediante CNNDD.

In [70, 20] viene presentato un metodo di classificazione ad una classe, chiamato CPDD, basato sulla scelta accurata di alcuni *prototipi*. Tale metodo ha accuratezza comparabile a quella di CNNDD, ma utilizza un sottoinsieme di riferimento più piccolo. Viene mostrato che calcolare un sottoinsieme di riferimento di cardinalità minima per il classificatore CPDD è in generale non approssimabile con un fattore costante e viene quindi introdotto un algoritmo avente un fattore di approssimazione logaritmico.

### Regole Associative

L'induzione di *regole associative* è una tecnica di data mining introdotta nell'ambito della cosiddetta *market basket analysis*. Si consideri un supermercato che gestisce una grande collezione di articoli. Una *transazione* è l'insieme di articoli che un cliente consegna alla cassa per l'acquisto. Poiché ogni transazione è registrata in una base di dati, il supermercato dispone di una enorme collezione di transazioni che è possibile analizzare per ottenere informazioni utili al processo di *decision making* (cosa vendere, dove collocare la merce, ecc.)

Una regola associativa è una implicazione del tipo  $X \Rightarrow Y$ , dove sia  $X$  che  $Y$  sono insiemi di oggetti, ed indica che l'occorrenza di una tupla contenente tutti gli oggetti nell'insieme  $X$  implica l'occorrenza nella stessa tupla di tutti gli oggetti nell'insieme  $Y$ . La validità di ogni regola associativa è misurata mediante l'uso di *indici*, vale a dire, funzioni che ad ogni regola associano un valore reale compreso tra 0 ed 1, il quale specifica la misura con cui la regola esprime conoscenza valida nella base di dati in esame. Gli indici più interessanti ed utilizzati sono *supporto*, *confidenza*, *gain* e *laplace*.

L'interesse dimostrato dalla comunità scientifica per le tecniche di induzione di regole associative è molto grande. Numerosi sono gli algoritmi efficienti per l'induzione di regole associative messi a punto e diverse varianti del problema di base sono state proposte e studiate. Oltre alle regole sopra descritte, dette *booleane*, sono state introdotte regole *categoriche* e *numeriche*.

Dallo studio degli approcci adottati dai vari autori, si evince che il problema dell'induzione di regole associative è un problema di difficile soluzione. Tuttavia, tranne che per alcune analisi piuttosto limitate, non esisteva una caratterizzazione della complessità computazionale di tale problema e delle sue varianti.

In [102, 38] questo aspetto viene affrontato in maniera organica. In particolare, dopo aver definito una forma generalizzata di regola associativa, che include come suoi casi particolari sia le regole booleane, che le regole categoriche e quantitative, viene studiata la complessità dell'induzione di regole associative valide rispetto alle misure più utilizzate in letteratura, vale a dire supporto, confidenza,  $\theta$ -gain e  $h$ -laplace.

I risultati presentati dimostrano formalmente che in generale questi problemi sono intrattabili (NP-completi), ad eccezione di alcuni interessanti casi particolari che sono stati descritti con precisione e che risultano essere trattabili (in particolare, risolubili, a seconda del caso specifico in esame, in L, TC<sup>0</sup> o AC<sup>0</sup>).

### Metaquerying

La *metaquerying* è una tecnologia di data mining che permette di scoprire dipendenze nascoste tra più relazioni di una base di dati. Le metaquery possono essere considerate una generalizzazione delle regole associative, in quanto le prime permettono di scoprire dipendenze *proposizionali* nascoste tra gli attributi della stessa relazione, mentre col metaquerying è possibile rilevare dipendenze al *primo ordine* che coinvolgono più tabelle.

Una *metaquery* è una implicazione del tipo  $T \leftarrow L_1, \dots, L_m$ , dove  $T$  ed ogni  $L_i$  sono *schemi di letterali*, vale a dire espressioni del tipo  $Q(Y_1, \dots, Y_n)$ , dove  $Q$  è una variabile di predicato ed ogni  $Y_i$  è una variabile ordinaria. Una *risposta* ad una metaquery è una regola logica ottenuta sostituendo nomi di relazioni della base di dati alle variabili di predicato. Analogamente a quanto accade per le regole associative, la validità delle regole indotte come risposta ad una metaquery è misurata per mezzo di *indici*. I più interessanti in questo contesto sono *supporto*, *confidenza* e *cover*.

Nonostante le metaquery siano state utilizzate con successo in applicazioni reali, erano noti solo risultati preliminari circa la loro complessità computazionale. Il problema della induzione di metaquery e la complessità computazionale associata alla risoluzione di metaquery viene affrontato e risolto in [103, 39], in relazione a diverse varianti di metaquerying, che includono tutte le semantiche definite in letteratura. In particolare, del problema viene studiata sia la *combined-complexity*, cioè la complessità del problema quando sia la query che la base di dati sono considerate componenti l'input, che la *data complexity*, cioè la complessità del problema quando solo la query viene considerata costituente l'input.

I risultati riportati mostrano, per quanto riguarda la combined complexity, che questi problemi sono intrattabili (NP-completi o addirittura NP<sup>PP</sup>-completi), tranne che in un interessante caso particolare, la cui combined complexity è LOGCFL-completa. Per quanto riguarda la data complexity, viene provato che, in generale, essa è in TC<sup>0</sup>, ma è confinata in AC<sup>0</sup> in alcuni casi particolari e interessanti.

In [34, 104] viene affrontato il problema del calcolo delle istanziazioni di una metaquery consistenti rispetto ad uno schema di base di dati. Innanzitutto, viene dimostrato che calcolare il numero di istanziazioni consistenti è un problema #P-completo, quindi tanto difficile quanto contare il numero di sostituzioni che soddisfano una query congiuntiva. Dopodichè, viene definito il concetto di *ridondanza* tra istanziazioni. Intuitivamente, due istanziazioni sono ridondanti se sono sintatticamente equivalenti, ovvero se rappresentano di fatto la stessa informazione. Gli algoritmi per l'istanziamento di metaquery presentati in letteratura calcolano tutte le istanziazioni di una metaquery, anche quelle ridondanti. Tuttavia, le istanziazioni ridondanti rappresentano un aggravio, sia per i sistemi di valutazione di metaquery, in quanto ogni istanziazione calcolata dovrà poi essere valutata sulla base di dati, sia per l'utente finale, che si troverà a dover analizzare più volte la stessa informazione. Nel lavoro citato viene quindi presentato un algoritmo per l'enumerazione di tutte e sole le istanziazioni non ridondanti di una metaquery.

## Clustering e closest pairs

**Clustering.** Per clustering si intende il partizionamento di un insieme di oggetti in gruppi omogenei, detti clusters. Si tratta di uno strumento per l'analisi dei dati che ha innumerevoli applicazioni ed esistono numerosi algoritmi di clustering, sviluppati nell'ambito di diverse discipline, quali machine learning, pattern recognition, data mining, statistica, biologia ed altre.

In [93] viene presentato l'algoritmo di clustering denominato DESCRy, un metodo che combina campionamento e clustering agglomerativo al fine di lavorare con grandi collezioni di dati a bassa dimensionalità. La complessità del metodo è lineare nel numero di oggetti da elaborare.

In [83] viene presentato l'algoritmo di clustering denominato OUTCLUST, un metodo basato su densità in grado di isolare clusters di forma arbitraria e di diversa densità anche in presenza di outliers, laddove la maggior parte degli algoritmi della sua stessa categoria proposti fino ad allora non erano in grado di calcolare clusters di differente densità.

In [87, 28] viene presentato un algoritmo di biclustering applicato a matrici di espressione genetica. L'algoritmo, chiamato *Random Walk Biclustering*, è un metodo di ricerca locale che segue una strategia greedy con la possibilità di eseguire mosse che peggiorino momentaneamente il valore della funzione obiettivo al fine di evitare di rimanere intrappolato in minimi locali. In [82] viene presentato l'algoritmo *BinCoC*, che estende la strategia sopra descritta al caso in cui i dati sono di natura binaria.

**Closest pairs.** Dato un insieme di punti di  $\mathbb{R}^d$ , il problema delle *k-Closest-Pairs* consiste nel determinare le *k* coppie di punti tra loro più vicini. Tale problema, introdotto nell'ambito della Geometria Computazionale, ha acquisito un rilevante interesse da parte della comunità delle Basi di Dati. In particolare un obiettivo è quello di mettere a punto tecniche di indicizzazione per la sua soluzione efficiente. I migliori algoritmi esatti per questo problema dipendono esponenzialmente da *d*. In [100, 36] viene descritto un algoritmo per la soluzione approssimata in tempo polinomiale del problema delle *k-Closest-Pairs* in spazi a molte dimensioni (vale a dire, per *d* grande) basato sull'uso delle curve space-filling. In [95, 35] tale algoritmo viene esteso al caso in cui i punti di ogni coppia devono appartenere a due insiemi distinti (*top-k similarity join*).

## Calcolo dei modelli minimali di teorie CNF

In [13, 55] viene affrontato il *problema del calcolo dei modelli minimali di teorie proposizionali in Conjunctive Normal Form (CNF)*. Tale problema è in generale intrattabile. Viene studiato il frammento delle teorie Head-set Elementary Free (HEF) positive dimostrando che, a dispetto dell'intrattabilità del problema del loro riconoscimento, il calcolo di un loro modello minimale può essere completato in tempo polinomiale. In particolare, l'algoritmo IGEA, acronimo di *Incomplete Generalized Elimination Algorithm*, garantisce la restituzione di un modello minimale se la teoria in input è HEF. *Questo risultato amplia la frontiera di trattabilità del problema del calcolo di modelli minimali di teorie CNF positive.*

In [48, 50, 107] viene presentata una tecnica di decomposizione basata sulla struttura del grafo delle dipendenze di una teoria CNF positiva per la determinazione dei modelli minimali. Il costo della tecnica è dominato dal costo associato alla componente connessa massimale. In ogni caso, se tutte le componenti connesse di questa decomposizione appartengono a determinati frammenti, ovvero sono HCF/HEF o di dimensione superiormente limitata da una costante, l'algoritmo risulta essere polinomiale.

## Cyber Security

In [7, 52, 8, ?] vengono presentate diverse tecniche di intrusion detection basate su anomalie. In particolare, in [52, 8] viene introdotta la tecnica PCKAD (Packet Chunk Anomaly Detector). PCKAD è una tecnica basata su *n*-grammi, ovvero analizza frammenti di traffico costituiti da sequenze sovrapposte di *n* caratteri/byte consecutivi, progettata per monitorare il traffico di rete di livello applicativo al fine di identificare attacchi basati sui contenuti. Al fine di creare un modello di traffico legittimo, la quasi totalità delle tecniche della suddetta famiglia effettuano un'analisi

frequenziale degli  $n$ -grammi estratti dal traffico senza attacchi. L'elemento innovativo di PCkAD consiste nell'implementazione di un'analisi di tipo spazio-frequenziale basata sulla suddivisione del traffico in blocchi di una certa dimensione, detti *chunk*.

In [8] viene dimostrato che la strategia messa in atto da PCkAD rende l'algoritmo robusto a tecniche di offuscamento di attacchi, note come *blending*, che consistono nell'aggiunta di byte al traffico malevolo allo scopo di simulare la distribuzione frequenziale tipica del traffico legittimo. Nel caso di PCkAD, per l'attaccante determinare il traffico in grado di camuffare la minaccia è in generale un problema NP-hard. Il confronto con tecniche di intrusion detection allo stato dell'arte che fanno uso di  $n$ -grammi, tra cui PAYL e *Spectrogram*, ha evidenziato prestazioni superiori per PCkAD.

In [?] viene considerato il problema della compressione degli  $n$ -grammi utilizzati da PCkAD e da tecniche della stessa famiglia per modellare i comportamenti normali. L'obiettivo è quello di ottenere riduzioni di costo spaziale e temporale. La compressione avviene mediante una clustering ibrido di tipo geometrico e frequenziale. Il livello di compressione è variabile ed il livello ottimale viene scelto in base all'analisi del traffico normale. Nella fase di testing, l'anomalia del traffico viene determinata in base alla copertura dei byte con gli  $n$ -grammi del modello compresso. I risultati sperimentali evidenziano che l'accuratezza del modello compresso è comunque superiore a quella di diverse tecniche basate su  $n$ -grammi che non adottano compressione.

In [7] viene descritto un framework per costruire sistemi ibridi per il rilevamento di attacchi. Il framework, in grado di combinare diverse tecniche di rilevamento, basate su firme oppure su anomalie, è stato sperimentato nel contesto dell'intrusion detection, con l'ausilio della tecnica PCkAD, e della steganalisi.

## **Bioinformatica**

Le interazioni tra RNA endogeni competitivi (ceRNA) sono riconosciute come un meccanismo chiave della differenziazione cellulare, dello sviluppo dei tessuti e delle malattie. In particolare, essi agiscono come spugne nei confronti del mRNA giocando potenzialmente un ruolo onco-geno/soppressore in diversi tipi di cancro. In [108] viene presentata una tecnica in grado di individuare il sopracitato effetto spugna dall'analisi di grandi dataset di espressione genica. La tecnica si basa sull'introduzione di una nuova nozione di similarità tra una popolazione ed una sua sottopopolazione e di una nuova misura del potere discriminante di una spugna.

## 4 Organizzazione, revisione e partecipazione a riviste e conferenze

### 4.1 Invited talks e tutorials

- Ha tenuto un talk su “Outlier detection: efficient algorithms and definitions for specific settings” all’interno della *Invited session* su “Outlier and anomaly detection in modern data settings” organizzata da Robert Serfling (University of Texas and President of the Conference of Texas Statisticians). La sessione, in partecipazione con Valentina Colla (Scuola Superiore Sant’Anna di Pisa), Emmanuel Müller (Karlsruhe Institute of Technology) e Roland Fried (Dortmund University), si è tenuta nell’ambito 7th International Conference of the ERCIM WG on Computational and Methodological Statistics, University of Pisa, 6-8 December, 2014.
- Ha tenuto un *Invited tutorial* su “Outlier detection: tasks and techniques” al *21st Italian Symposium on Advanced Database Systems (SEBD)*, Roccella Jonica, June 30th–July 3rd, 2013.

### 4.2 Partecipazione a editorial board di riviste

- A partire da agosto 2012 è *Associate Editor* per l’area *Data Mining and Knowledge Discovery* della rivista internazionale *AI Communications* (AICom), IOS Press. *AI Communications* è la rivista europea di Intelligenza Artificiale ed è legata all’ECCAI (European Coordinating Committee for Artificial Intelligence).

•

### 4.3 Partecipazione a comitati organizzatori di conferenze scientifiche

- Local Organization Chair della *3rd International Conference on Model & Data Engineering (MEDI)*, September 25-27, 2014, Amantea (CS), Italy.

### 4.4 Partecipazione a comitati organizzatori e/o di programma di conferenze scientifiche

#### Partecipazione a comitati organizzatori di conferenze scientifiche

- Local Organization Chair della *3rd International Conference on Model & Data Engineering (MEDI)*, September 25-27, 2014, Amantea (CS), Italy.

#### Area Chair o Senior Program Committee Member

È od è stato Area Chair o Senior Program Committee Member delle seguenti conferenze:

- **Area Chair** of the *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery (ECMLPKDD)*, 2019
- **Area Chair** of the *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery (ECMLPKDD)*, 2020
- **Area Chair** of the *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery (ECMLPKDD)*, 2021
- **Senior Program Committee Member** of the IJCAI 2020

### Program Committee Member

È od è stato membro del comitato di programma delle seguenti conferenze:

- 2005:** WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2006:** AIDM: *International Workshop on Integrating AI and Data Mining*, in conjunction with the Australian Joint Conference on Artificial Intelligence — SAC: *ACM Symposium on Applied Computing* Track on Data Mining — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2007:** AIDM: *International Workshop on Integrating AI and Data Mining*, in conjunction with the Australian Joint Conference on Artificial Intelligence — SAC: *ACM Symposium on Applied Computing* Track on Data Mining — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2008:** ICWAPR: *International Conference on Wavelet Analysis and Pattern Recognition* — IFIP-AI: *IFIP International Conference on Artificial Intelligence* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2009:** ICTAI: *International Conference on Tools with Artificial Intelligence* — IDA: *International Symposium on Intelligent Data Analysis* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining.
- 2010:** ADBIS: *East-European Conference on Advances in Databases and Information Systems*, Track on Warehousing and OLAPing Complex, Spatial and Spatio-Temporal Data — IDA: *International Symposium on Intelligent Data Analysis* — IFIP-AI: *IFIP International Conference on Artificial Intelligence* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining.
- 2011:** AAAI: *AAAI Conference on Artificial Intelligence* — EIDWT: *International Conference on Emerging Intelligent Data and Web Technologies* — MCP: *AI\*IA International Workshop on Mining Complex Patterns* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining.
- 2012:** AAAI: *AAAI Conference on Artificial Intelligence* — IDA: *International Symposium on Intelligent Data Analysis* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2013:** DATA ANALYTICS: *International Conference on Data Analytics* — IDA: *International Symposium on Intelligent Data Analysis* — IJCAI: *International Joint Conference on Artificial Intelligence* — NBiS: *International Conference on Network-Based Information Systems* — ODD: *ACM KDD Workshop on Outlier Detection and Description* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2014:** IDA: *International Symposium on Intelligent Data Analysis* — ODD: *ACM KDD Workshop on Outlier Detection and Description* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2015:** ICDM: *International Conference on Data Mining*, PhD Forum — IDA: *International Symposium on Intelligent Data Analysis* — IJCAI: *International Joint Conference on Artificial Intelligence* — ODD: *ACM KDD Workshop on Outlier Detection and Description* — SAC: *ACM Symposium on Applied Computing*, Track on Data Mining — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.



- 2016:** ECML-PKDD: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery* — IDA: *International Symposium on Intelligent Data Analysis* — IJCAI: *International Joint Conference on Artificial Intelligence* — ODD: *ACM KDD Workshop on Outlier Detection and Description* — SAC: *ACM Symposium on Applied Computing, Track on Data Mining* — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2017:** ECML-PKDD: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery* — IDA: *International Symposium on Intelligent Data Analysis* — IJCAI: *International Joint Conference on Artificial Intelligence* — SAC: *ACM Symposium on Applied Computing, Track on Data Mining* — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2018:** ECML-PKDD: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery* — IDA: *International Symposium on Intelligent Data Analysis* — ODD: *ACM KDD Workshop on Outlier Detection and Description* — SAC: *ACM Symposium on Applied Computing, Track on Data Mining* — WI: *IEEE/WIC/ACM International Conference on Web Intelligence*.
- 2019:** ECML-PKDD: *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery* — ADF: *2nd ACM KDD Workshop on Anomaly Detection in Finance* — SAC: *ACM Symposium on Applied Computing, Track on Data Mining* — AIXIA: *18th International Conference of the Italian Association for Artificial Intelligence*.

#### 4.5 Attività di revisione per riviste

Ha partecipato o partecipa all'attività di selezione degli articoli proposti per la pubblicazione a riviste nazionali ed internazionali, tra cui:

- *ACM Transactions on Knowledge Discovery in Data* (TKDD);
- *AI Communications*, IOS Press (AICom);
- *Artificial Intelligence*, Elsevier (AIJ);
- *Artificial Intelligence in Medicine*, Elsevier (AIIM);
- *Data & Knowledge Engineering*, Elsevier (DKE);
- *Data Mining and Knowledge Discovery*, Springer (DAMI);
- *IEEE Transactions on Knowledge and Data Engineering* (TKDE);
- *IEEE Transactions on Neural Networks* (TNN);
- *IEEE Transactions on Neural Networks and Learning Systems* (TNNLS);
- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (TPAMI);
- *IEEE Transactions on Systems, Man and Cybernetics - Part B* (SMCB);
- *Information Sciences*, Elsevier (INS);
- *Information Systems*, Elsevier (IS);
- *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific (IJPRAI);
- *International Journal on Artificial Intelligence Tools*, World Scientific (IJAIT);
- *Knowledge and Information Systems*, Springer (KAIS);

- *Journal of Computer Science and Technology*, Springer (JCST);
- *Journal of Machine Learning Research* (JMLR);
- *Journal of Symbolic Computation*, Elsevier (JSC);
- *Journal of the Zhejiang University - Computers & Electronics* (JZUS-C),
- *Pattern Recognition*, Elsevier (PR);
- *Pattern Recognition Letters*, Elsevier (PRL);
- *Machine Learning*, Springer (MACH);
- *Mathematical Problems in Engineering*, Hindawi;
- *Neurocomputing*, Elsevier (NEUROCOM);
- *Rivista di Informatica*, AICA;
- *Statistical Analysis and Data Mining*, John Wiley & Sons (SAM);
- *The Computer Journal*, Oxford Journals (COMPJ);
- *The Open Artificial Intelligence Journal*, Bentham Open (TOAIJ);
- *The VLDB Journal*, Springer (VLDB J.);
- *Transportation Research - Part C*, Elsevier (TRC);

#### 4.6 Attività di revisione per conferenze

Ha partecipato o partecipa all'attività di selezione degli articoli proposti per la pubblicazione a diversi congressi nazionali ed internazionali, tra cui:

- AAAI: *Conf. of the American Association for Artificial Intelligence*;
- ADMA: *Int. Conf. on Advanced Data Mining and Applications*;
- AIDM: *Int. Workshop on Integrating AI and Data Mining*;
- BIO-LOGICAL: *Workshop on Logic-based approaches in Bioinformatics - Int. Conf. AI\*IA*;
- DASFAA: *Database Systems For Advanced Applications*;
- DBKDA: *Int. Conf. on Advances in Databases, Knowledge, and Data Applications*;
- DEXA: *Int. Conf. on Database and Expert Systems Applications*;
- DBPL: *Int. Symposium on Database Programming Languages*;
- DaWak: *Int. Conf. on Data Warehousing and Knowledge Discovery*;
- DBA: *Int. Conf. on Databases and Applications*;
- DS: *Int. Conf. on Discovery Science*;
- ECSQARU: *European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*;
- EDBT: *Int. Conf. on Extending Database Technology*;
- IAT: *IEEE/WIC/ACM Int. Conf. on Intelligent Agent Technology*;

- ICDM: *IEEE Int. Conf. on Data Mining*;
- ICDT: *Int. Conf. on Database Theory*;
- ICTAI: *Int. Conf. on Tools with Artificial Intelligence*;
- IDA: *Int. Symp. on Intelligent Data Analysis*;
- IDEAS: *Int. Database Engineering & Application Symposium*;
- IFIP-AI: *IFIP Int. Conf. on Artificial Intelligence*;
- IJCAI: *Int. Joint Conf. on Artificial Intelligence*;
- ITCC: *IEEE Int. Conf. on Information Technology: Coding and Computing*;
- JELIA: *European Conf. on Logics in Artificial Intelligence*;
- PKDD: *European Conf. on Principles and Practice of Knowledge Discovery in Databases*;
- PODS: *ACM Symposium on Principles of Database Systems*;
- SAC: *ACM Symposium on Applied Computing*;
- SEBD: *Convegno Nazionale su Sistemi Evoluti per Basi di Dati*;
- SDM: *SIAM Int. Conf. on Data Mining*;
- SSDBM: *Int. Conf. on Scientific and Statistical Database Management*;
- SUM: *Int. Conf. on Scalable Uncertainty Management*;
- WI: *IEEE/WIC/ACM Int. Conf. on Web Intelligence*.

## 5 Progetti di ricerca ed innovazione

### 5.1 Attività di revisione di progetti di ricerca ed innovazione

- Nell'anno 2016 ha svolto l'incarico di revisore per conto della *The Netherlands Organisation for Scientific Research* (NWO) delle proposte di progetto "Vidi Innovational Research Incentives Scheme", il cui bando è aperto a ricercatori nazionali ed esteri che abbiano conseguito un PhD in Physical Sciences nel corso degli ultimi 8 anni (importo massimo finanziabile su singolo progetto € 800.000).
- Nell'anno 2017 ha svolto l'incarico di valutatore dei progetti interdisciplinari d'ateneo per il macro settore Physical Sciences and Engineering–European Research Council (ERC PE) presentati nell'ambito delle selezioni delle proposte di progetto sui Fondi di Ateneo per la Ricerca (FAR) dell'Università di Modena e Reggio Emilia (UNIMORE); importo totale stanziato per il finanziamento € 650.000. L'incarico è stato affidato attraverso una selezione pubblica di valutatori esperti internazionali di chiara fama indetta da UNIMORE.

### 5.2 Responsabilità di progetti di ricerca ed innovazione

- **Rialto Plus – Una piattaforma per il Business Analytics** (*luglio 2009 – dicembre 2010; finanziato dalla Regione Calabria; ricerca e sviluppo tecnologico; responsabile scientifico*). Nell'ambito del progetto PIA-Calabria (Piano d'Innovazione Aziendale) è stato responsabile scientifico del tema "Outlier Detection". Il progetto è stato sviluppato dalla società Exeura srl e dal Dipartimento di Elettronica, Informatica e Sistemistica dell'Università della Calabria (DEIS-UNICAL) come partner di ricerca. Il progetto ha realizzato un prototipo evoluto di sistema di *Knowledge Discovery* con applicazioni nel campo della *Business Intelligence*. In particolare, il tema "Outlier Detection" ha previsto attività di ricerca sviluppatesi lungo le seguenti tre linee: algoritmi efficienti per l'outlier detection, outlier detection in flussi di dati, scoperta di caratteristiche eccezionali degli outlier.
- **Parallel Methods for Uncertain Data Retrieval** (*settembre 2010; finanziato da NVIDIA Co.; ricerca di base; responsabile scientifico*). Finanziato da NVIDIA nell'ambito dell'Academic Partnership Program con una donazione di attrezzature consistente in schede GPU many-core di ultima generazione. Il progetto ha riguardato la messa a punto di metodi per la ricerca ed il mining di dati incerti efficientemente parallelizzabili su architettura many-core.
- **EX MURST 60%** (*gennaio 2011 – dicembre 2013; finanziato dal MIUR; ricerca di base; responsabile scientifico – principal investigator*). È stato responsabile scientifico del progetto di ricerca dal titolo "Tecniche per la Gestione ed il Mining di Dati Incerti" finanziato con fondi EX MURST 60%. Il gruppo di ricerca del progetto si compone di 9 unità di personale strutturato, a tempo determinato ed in formazione dell'Università della Calabria. Il progetto si propone di mettere a punto tecniche innovative per la classificazione e la ricerca di anomalie in dati incerti.
- **IDEAS** (*gennaio 2010 – giugno 2012; finanziato dal Ministero dello Sviluppo Economico, MISE; ricerca e trasferimento tecnologico; responsabile di unità*). Il Fondo per l'Innovazione Tecnologica (FIT) ha come finalità quella di promuovere e diffondere l'innovazione basata sulle tecnologie dell'informazione e della comunicazione (ICT) nelle Piccole e Medie Imprese. È stato responsabile dell'unità UNICAL impegnata nel progetto FIT "IDEAS" che si propone di ottenere una suite di strumenti utili per la georeferenziazione e per la classificazione di forme all'interno di immagini con possibili applicazioni in ambito catastale.
- **FRAME** (*giugno 2011 – maggio 2014; finanziato dal MIUR; ricerca industriale; responsabile di unità*). È stato responsabile dell'unità DEIS-UNICAL impegnata nel progetto PON01\_02477 "FRAME – Un FRAmework flessibile ed espandibile, fondato su MEtologie e strumenti basati sulla conoscenza, per il consolidamento e la gestione ottimizzata dei sistemi informativi

complessi”. I soggetti proponenti sono SIEMENS IT Solutions and Services SpA., Exeura srl, ICAR-CNR, Dip. di Matematica (DIMAT) e Dipartimento di Elettronica, Informatica e Sistemistica (DEIS) dell’Università della Calabria (UNICAL). Il progetto FRAME mira ad ottenere un framework basato su metodologie, tecniche e servizi innovativi, mettendo a disposizione un’architettura modulare e flessibile basata su alcuni strati applicativi tra cui, di particolare interesse per la struttura operativa DEIS-UNICAL, la componente specializzata nella gestione del ciclo di vita degli asset digitali (DAM – Digital Asset Management) in grado di supportare “tagging” e “meta-datazione” semi-automatica di contenuti digitali.

- **SIARTE** (novembre 2017 – biennale; finanziato dalla Regione Calabria; ricerca industriale; responsabile della componente accademica). È responsabile dell’unità DIMES-UNICAL impegnata nel progetto POR FESR/FSE 2014-2020 SIARTE “Sistema Integrato Acquisizione ed Analisi dei dati di Riscossione dei Tributi da parte degli Enti pubblici”. I soggetti proponenti sono Syntax srl e l’Università della Calabria (DIMES-UNICAL). L’obiettivo del progetto è la definizione e realizzazione di un sistema integrato per l’acquisizione e l’analisi di dati inerenti la riscossione dei tributi da parte di Enti Pubblici, al fine di produrre innovazione digitale nell’erogazione dei servizi di riscossione, un ambiente dotato di *business intelligence* capace di gestire *big data* e di usare e produrre *open linked data*.

### 5.3 Partecipazione a progetti di ricerca ed innovazione

- **CONTACT/ADAPT** (novembre 1999 – maggio 2000; finanziato dall’Unione Europea; trasferimento tecnologico; componente del gruppo di progetto). Nell’ambito del Progetto Europeo CONTACT/ADAPT (Cooperative Network for Technology transfer ACtions and Training) è stato componente del gruppo di progetto presso il Dipartimento di Elettronica, Informatica e Sistemistica dell’Università della Calabria.
- **Modelli e tecniche per l’integrazione, l’interrogazione ed il metaquerying in ambiente data warehouse** (triennio 2000 – 2002; finanziato dal MIUR; ricerca di rilevante interesse nazionale; componente del gruppo di progetto). Ha partecipato al progetto PRIN/CONFIN’00 – “D2I: Modelli e tecniche per l’integrazione, l’interrogazione ed il metaquerying in ambiente data warehouse”, finanziato dal MIUR per il triennio 2000 – 2002.
- **Tecniche di induzione di regole, metaquerying ed estrazione di pattern strutturati su database biologici** (novembre 2003 – ottobre 2004; finanziato dal MIUR; ricerca di rilevante interesse nazionale; componente del gruppo di progetto). È stato componente dell’unità di ricerca del Dipartimento di Elettronica, Informatica e Sistemistica dell’Università della Calabria impegnata nel progetto biennale PRIN/COFIN’03 – “Tecniche di induzione di regole, metaquerying ed estrazione di pattern strutturati su database biologici”, finanziato dal MIUR.
- **Piattaforme abilitanti per griglie computazionali ad elevate prestazioni orientate ad organizzazioni virtuali scalabili** (triennio 2003 – 2005; finanziato dal MIUR; ricerca di base; componente del gruppo di progetto). È stato componente del gruppo di ricerca dell’Istituto di Calcolo e Reti ad Alte Prestazioni del C.N.R., sede di Cosenza, impegnata in nel progetto FIRB di durata triennale.
- **Ambienti intelligenti ed adattativi di sostegno all’e-government e all’e-learning nelle pubbliche amministrazioni locali** (periodo 2000 – 2006; finanziato dall’Unione Europea; ricerca industriale; componente del gruppo di progetto e responsabile di task). Realizzato con il cofinanziamento dell’Unione europea attraverso il Fondo europeo di sviluppo regionale nell’ambito del PON Ricerca 2000-2006. Tale progetto è stato selezionato sulla base della regolarità procedurale e finanziaria delle attività e delle indicazioni in merito al contenuto tecnologico di valutatori scientifici. Il progetto, nato da una collaborazione fra la società IFM e l’Università della Calabria, ha sviluppato soluzioni tecnologiche per alcuni

obiettivi di e-government ed e-democracy delle pubbliche amministrazioni locali. Nell'ambito del progetto è stato responsabile del task di Data Cleaning.

- **AutoMa** (gennaio 2007 – giugno 2009; finanziato dalla Regione Calabria e dal MIUR; ricerca applicata; componente del gruppo di progetto). È stato componente del gruppo di ricerca del progetto “AutoMa”. Obiettivo del progetto è lo sviluppo di un insieme di tool integrati a supporto della logistica automobilistica, con particolare riferimento al sito di trans-shipment del Porto di Gioia Tauro.
- **OKT – OpenKnowTech** (maggio 2009 – giugno 2011; finanziato dall’Unione Europea; ricerca di valenza strategica nazionale; componente del gruppo di progetto e responsabile di task). È stato componente del progetto OpenKnowTech (OKT) – Laboratorio di Tecnologie per l’Integrazione, Gestione e Distribuzione di Dati, Processi e Conoscenze. OKT è un Laboratorio Pubblico-Privato che ha l’obiettivo di sviluppare e favorire la diffusione di software Open Source. Partner di OKT sono l’Università della Calabria, l’Università di Pisa, il Politecnico di Torino, gli istituti ICAR, ISTI e IEIT del Consiglio Nazionale delle Ricerche (CNR) e le compagnie private EUNICS, EXEURA, ORANGEE, IDTECH ed HERZUM SW. Nell’ambito del progetto è stato componente dell’unità del DEIS-UNICAL coinvolta nell’OR6 “Applicazioni di logistica” per cui è stato responsabile del task della progettazione e sviluppo dei moduli software per la gestione e l’analisi di tracce di esecuzione di processi logistici.
- **Geo-Analytical System** (2011 – 2012; finanziato dalla Regione Calabria; ricerca e sviluppo tecnologico; componente del gruppo di progetto). È stato componente dell’unità di ricerca del DEIS-UNICAL impegnata nel progetto PIA-Calabria “Geo-Analytical System – Strumenti di business analysis e reportistica avanzata di tipo geo-referenziato”.
- **TETRIS** (aprile 2012 – maggio 2013; finanziato dal MIUR; ricerca industriale; partecipante al progetto). Nell’ambito del progetto PON “TETRIS: Servizi innovativi Open Source su TETRA” ha partecipato alle attività di formazione di esperti nella progettazione e sviluppo di servizi e software open source.
- **Integration for Business Intelligence** (giugno 2011 – novembre 2012; finanziato dalla Regione Calabria; ricerca e sviluppo tecnologico; componente del gruppo di progetto). Nell’ambito del Piano di Sviluppo Interaziendale – Contratti di Investimento Industria Artigianato Servizi della Regione Calabria, proposto dal Centro di Competenza ICT-SUD, è stato componente del Progetto di Ricerca e Sviluppo Tecnologico “E: Integration for Business Intelligence”. Obiettivo del Piano è la realizzazione dell’ecosistema digitale di ICT-SUD, un paradigma emergente per l’innovazione economica e tecnologica. Il progetto “E” comprende l’Università della Calabria quale partner di ricerca ed i seguenti 5 partner industriali: Armat srl, Almaviva SUD SpA, DLVSYSTEM srl, Herzum Società Unipersonale srl, Methodi srl. Il Progetto intende realizzare una piattaforma per l’integrazione di dati e servizi finalizzata alla Business Intelligence.
- **BA2Know** (gennaio 2014 – novembre 2015; finanziato dal MIUR; ricerca industriale; componente del gruppo di progetto). È stato componente dell’unità di ricerca del DIMES-UNICAL impegnata nel progetto PON03PE/SILAB in cui sono state affrontate tematiche di Business Analytics.
- **H-INT** (giugno 2014 – giugno 2016; finanziato dalla Regione Calabria; ricerca industriale; componente del gruppo di progetto). È stato componente dell’unità di ricerca del DIMES-UNICAL impegnata nel progetto H-INT “Health Integration” che ha riguardato l’integrazione di sistemi informativi medici attraverso l’utilizzo di standard internazionali per la rappresentazione e lo scambio di dati sanitari.
- **Cybersecurity** (gennaio 2015 – giugno 2016; finanziato dal MIUR; ricerca industriale; componente del gruppo di progetto). È stato componente dell’unità di ricerca del DIMES-UNICAL impegnata nel progetto PON “Cybersecurity” che ha coinvolto diversi partners

accademici ed industriali, tra cui Poste Italiane, all'interno di un distretto tecnologico su cyber security localizzato nel territorio cosentino.

- **SELINA** (novembre 2017 – biennale; finanziato dalla Regione Calabria; ricerca industriale; componente del gruppo di progetto). Partecipa al progetto POR FESR/FSE 2014-2020 SELINA “Smart Electronic invoices Accounting”. I soggetti proponenti sono Calìo Informatica srl, UNICZ e UNICAL. L'obiettivo del progetto è l'innovazione dei processi di contabilizzazione mediante l'impiego di tecniche di machine learning.
- **IDSERVICE** (inizio gennaio 2017; finanziato dal MIUR; ricerca industriale; componente del gruppo di progetto). È componente dell'unità di ricerca del DIMES-UNICAL impegnata nel progetto PON “IDSERVICE”. Le tematiche del progetto sono l'identità digitale e la gestione sicura dei processi aziendali attraverso l'utilizzo di tecnologie Blockchain.
- **SPIDASEC** (novembre 2017 – biennale; finanziato dalla Regione Calabria; ricerca industriale; responsabile della componente accademica). È stato responsabile fino a gennaio 2018 dell'unità DIMES-UNICAL impegnata nel progetto POR FESR/FSE 2014-2020 SPIDASEC “SPID Advanced Security”. I soggetti proponenti sono Consulthink spa, Coremuniti srl, Catenate srl, l'ICAR-CNR e l'Università della Calabria (DIMES-UNICAL). L'obiettivo del progetto è il miglioramento della sicurezza generale di SPID e del trust tra le parti consentendo una più diffusa adozione del modello SPID, una maggiore tutela degli utenti finali, la semplificazione dell'interazione tra l'utente ed il sistema e la diffusione di nuovi servizi digitali.

## 6 Attività didattiche

### 6.1 Corsi Universitari

Nel corso degli ultimi circa 15 anni è stato docente di numerosi corsi universitari (circa 40 edizioni) presso l'Università della Calabria (UNICAL), l'Università degli Studi "Magna Graecia" (UNICZ) e l'Università degli Studi "Mediterranea" (UNIRC), riguardanti un ampio spettro di tematiche nel settore dell'ingegneria informatica, quali, *Introduzione all'Informatica*, *Fondamenti di Informatica*, *Calcolatori Elettronici*, *Calcolatori Elettronici II*, *Architetture e Programmazione dei Sistemi di Elaborazione*, *Sistemi Informativi e Basi di Dati*, *Intelligenza Artificiale*, *Analisi di Immagini* e *Concetti Evoluti di Data Mining*.

Più in dettaglio, i principali corsi in cui Fabrizio Angiulli ha svolto attività di docente e/o esercitatore ufficiale presso l'Università sono di seguito dettagliati, suddivisi per anno accademico.

- **2003-04:**

- È stato docente del corso di *Architettura dei Calcolatori Elettronici* (CFU: 5), C.d.L. Specialistica in Ingegneria delle Telecomunicazioni, Facoltà di Ingegneria, Università "Mediterranea" di Reggio Calabria ed ha presieduto la relativa commissione d'esame;
- È stato docente del corso di *Introduzione all'Informatica* (CFU: 2), C.d.L. in Ingegneria, Università della Calabria ed ha presieduto la relativa commissione d'esame;
- È stato esercitatore ufficiale del corso di *Calcolatori Elettronici* (CFU: 5), C.d.L. in Ingegneria Informatica, Università della Calabria ed è stato componente della relativa commissione d'esame;
- È stato esercitatore ufficiale del corso di *Architetture Innovative di Sistemi di Elaborazione* (CFU: 5), C.d.L. Specialistica in Ingegneria Informatica, Facoltà di Ingegneria, Università della Calabria ed è stato componente della relativa commissione d'esame.

- **2004-05:**

- È stato docente del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e Biomedica, Università "Magna Graecia" di Catanzaro ed ha presieduto la relativa commissione d'esame;
- È stato docente del corso di *Calcolatori Elettronici* (CFU: 5), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame;
- È stato docente del corso di *Introduzione all'Informatica* (CFU: 2), C.d.L. in Ingegneria, Università della Calabria ed ha presieduto la relativa commissione d'esame;
- È stato esercitatore ufficiale del corso di *Intelligenza Artificiale* (CFU: 5), C.d.L. Specialistica in Ingegneria Informatica ed è stato componente della relativa commissione d'esame.

- **2005-06:**

- È stato docente del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e dei Sistemi Sanitari, Università "Magna Graecia" di Catanzaro ed ha presieduto la relativa commissione d'esame.
- È stato docente del corso di *Calcolatori Elettronici* (CFU: 5), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- È stato docente del corso di *Introduzione all'Informatica* (CFU: 2), C.d.L. in Ingegneria, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- È stato esercitatore ufficiale del corso di *Intelligenza Artificiale* (CFU: 5), C.d.L. Specialistica in Ingegneria Informatica, Università della Calabria ed è stato componente della relativa commissione d'esame.

- **2006-07:**



- È stato docente del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e dei Sistemi Sanitari, Università “Magna Græcia” di Catanzaro ed ha presieduto la relativa commissione d’esame.
- È stato docente del corso di *Fondamenti di Informatica* (CFU: 4), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona ed ha presieduto la relativa commissione d’esame.
- È stato docente del corso di *Sistemi Informativi e Basi di Dati* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona ed ha presieduto la relativa commissione d’esame.
- È stato docente del corso di *Calcolatori Elettronici* (CFU: 5), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d’esame.

• **2007-08:**

- È stato docente del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e dei Sistemi Sanitari, Università “Magna Græcia” di Catanzaro ed ha presieduto la relativa commissione d’esame.
- È stato docente del corso di *Fondamenti di Informatica* (CFU: 4), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona ed ha presieduto la relativa commissione d’esame.
- È stato docente del corso di *Sistemi Informativi e Basi di Dati* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona ed ha presieduto la relativa commissione d’esame.

• **2008-09:**

- È stato docente del corso di *Fondamenti di Informatica* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona, ed ha presieduto la relativa commissione d’esame.
- È stato docente del corso di *Sistemi Informativi e Basi di Dati* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona, ed ha presieduto la relativa commissione d’esame.
- Ha presieduto la commissione d’esame del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e dei Sistemi Sanitari, Università “Magna Græcia” di Catanzaro.

• **2009-10:**

- È stato docente del corso di *Fondamenti di Informatica* (CFU: 6), C.d.L. in Ingegneria, Università della Calabria, ed ha presieduto la relativa commissione d’esame.
- Ha presieduto la commissione d’esame del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e dei Sistemi Sanitari, Università “Magna Græcia” di Catanzaro.
- Ha presieduto la commissione d’esame del corso di *Fondamenti di Informatica* D.M. 509 (CFU: 4) e D.M. 270 (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona.
- Ha presieduto la commissione d’esame del corso di *Sistemi Informativi e Basi di Dati* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona.

• **2010-11:**

- È stato docente del corso di *Calcolatori Elettronici II* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d’esame.

- È stato docente del corso di *Concetti Evoluti di Data Mining* (CFU: 2), Corso di Dottorato di Ricerca in Ingegneria dei Sistemi ed Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- Ha presieduto la commissione d'esame del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e dei Sistemi Sanitari, Università "Magna Græcia" di Catanzaro.
- Ha presieduto la commissione d'esame del corso di *Fondamenti di Informatica* D.M. 509 (CFU: 4) e D.M. 270 (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona.
- Ha presieduto la commissione d'esame del corso di *Sistemi Informativi e Basi di Dati* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria Sede di Crotona.

• **2011-12:**

- È stato docente del corso di *Calcolatori Elettronici II* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- È stato docente del corso di *Concetti Evoluti di Data Mining* (CFU: 2), Corso di Dottorato di Ricerca in Ingegneria dei Sistemi ed Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- Ha presieduto la commissione d'esame del corso di *Intelligenza Artificiale* (CFU: 8), C.d.L. Specialistica in Ingegneria Informatica e dei Sistemi Sanitari, Università "Magna Græcia" di Catanzaro.
- Ha presieduto la commissione d'esame del corso di *Fondamenti di Informatica* D.M. 509 (CFU: 4) e D.M. 270 (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona.
- Ha presieduto la commissione d'esame del corso di *Sistemi Informativi e Basi di Dati* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona.

• **2012-13:**

- È stato docente del corso di *Fondamenti di Informatica* (CFU: 6), C.d.L. in Ingegneria, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- È stato docente del corso di *Calcolatori Elettronici II* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- È stato docente del corso di *Concetti Evoluti di Data Mining* (CFU: 2), Corso di Dottorato di Ricerca in Ingegneria dei Sistemi ed Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- Ha presieduto la commissione d'esame del corso di *Fondamenti di Informatica* D.M. 509 (CFU: 4) e D.M. 270 (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona.
- Ha presieduto la commissione d'esame del corso di *Sistemi Informativi e Basi di Dati* (CFU: 6), C.d.L. in Ingegneria Gestionale, Università della Calabria, Sede di Crotona.

• **2013-14:**

- È stato docente del corso di *Fondamenti di Informatica* (CFU: 6), C.d.L. in Ingegneria Elettronica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- È stato docente del corso di *Calcolatori Elettronici II* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.

- È stato docente del corso di *Concetti Evoluti di Data Mining* (CFU: 2), Corso di Dottorato di Ricerca in Information and Communication Engineering for Pervasive Intelligent Environments, Università della Calabria ed ha presieduto la relativa commissione d'esame.
- **2014-15:**
  - È stato docente del corso di *Fondamenti di Informatica* (CFU: 6), C.d.L. in Ingegneria Elettronica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
  - È stato docente del corso di *Calcolatori Elettronici II* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
  - È stato docente del modulo di *Analisi di Immagini* nell'ambito del corso di *Informatica Grafica e Analisi di Immagini* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria.
- **2015-16:**
  - È stato docente del corso di *Reti Logiche e Calcolatori* (CFU: 9), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
  - È stato docente del corso di *Architetture e Programmazione dei Sistemi di Elaborazione* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, ed ha presieduto la relativa commissione d'esame.
  - È stato docente del modulo di *Analisi di Immagini* nell'ambito del corso di *Informatica Grafica e Analisi di Immagini* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria.
- **2016-17:**
  - È stato docente del corso di *Reti Logiche e Calcolatori* (CFU: 9), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
  - È stato docente del corso di *Architetture e Programmazione dei Sistemi di Elaborazione* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, ed ha presieduto la relativa commissione d'esame.
- **2017-18:**
  - È stato docente del corso di *Reti Logiche e Calcolatori* (CFU: 9), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
  - È stato docente del corso di *Architetture e Programmazione dei Sistemi di Elaborazione* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, ed ha presieduto la relativa commissione d'esame.
- **2018-19:**
  - È stato docente del corso di *Reti Logiche e Calcolatori* (CFU: 9), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.
  - È stato docente del corso di *Architetture e Programmazione dei Sistemi di Elaborazione* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, ed ha presieduto la relativa commissione d'esame.
- **2019-20:**
  - È stato docente del corso di *Reti Logiche e Calcolatori* (CFU: 9), C.d.L. in Ingegneria Informatica, Università della Calabria ed ha presieduto la relativa commissione d'esame.

- È stato docente del corso di *Machine e Deep Learning* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, ed ha presieduto la relativa commissione d’esame.
- È stato docente del corso di *Architetture e Programmazione dei Sistemi di Elaborazione* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, ed ha presieduto la relativa commissione d’esame.

• **2020-21:**

- È docente del corso di *Reti Logiche e Calcolatori* (CFU: 9), C.d.L. in Ingegneria Informatica, Università della Calabria e presiede la relativa commissione d’esame.
- È docente del corso di *Machine e Deep Learning* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, e presiede la relativa commissione d’esame.
- È docente del corso di *Architetture e Programmazione dei Sistemi di Elaborazione* (CFU: 6), C.d.L. Magistrale in Ingegneria Informatica, Università della Calabria, e presiede la relativa commissione d’esame.

## 6.2 Partecipazione a Dottorati di Ricerca e Supervisione di studenti di Dottorato di ricerca

È od è stato membro del Collegio dei Docenti dei seguenti corsi di Dottorato:

- *Ingegneria dei Sistemi e Informatica*, UNICAL: a.a. dal 2008/09 al 2012/13 (cicli XXIV-XXVIII);
- *Information and Communication Engineering for Pervasive Intelligent Environments*, UNICAL: a.a. 2013/14 (ciclo XXIX);
- *Information and Communication Technologies*, UNICAL: a.a. dal 2014/15 ad oggi. È attualmente componente del Collegio effettivo composto da 16 membri sulla base dei quali viene effettuata la valutazione del corso di dottorato dalla parte dell’ANVUR.

È od è stato supervisore dei seguenti studenti di Dottorato di Ricerca:

- *Fabio Fassetti* (XX ciclo, Corso di Dottorato di Ricerca in *Ingegneria dei Sistemi e Informatica*, UNICAL): ha approfondito tematiche di scoperta di anomalie e proprietà anomale in basi di dati e data streams. Ha conseguito il titolo di Dottore di Ricerca discutendo la tesi “*Discovering Exceptional Individuals and Properties in Data*”. A partire da dicembre 2012, Fabio Fassetti ricopre una posizione di Ricercatore Universitario presso il DIMES-UNICAL; nel 2017 ha conseguito l’Abilitazione Scientifica Nazionale alla funzione di Professore Universitario di II fascia;
- *Estela Narvaez* (XXIX ciclo, Corso di Dottorato di Ricerca in *Information and Communication Engineering for Pervasive Intelligent Environments*, UNICAL): borsista di stato estero sovvenzionato dal Senescyt – Secretaria Nacional de Ciencia y Tecnologia – dell’Ecuador. ha approfondito tematiche di compressione nell’ambito della classificazione e della ricerca di anomalie in strutture a rete. Ha conseguito il titolo di Dottore di Ricerca discutendo la tesi “*Data Mining Techniques for Large and Complex Data*”. A partire da aprile 2018, Estela Narvaez ricopre una posizione di Professore presso la Facoltà di Ingegneria dell’Universidad Nacional de Chimborazo (UNACH), Riobamba, Ecuador.
- *Luciano Argento* (XXX ciclo, Corso di Dottorato di Ricerca in *Information and Communication Technologies*, UNICAL): ha approfondito tematiche di Cybersecurity mediante strumenti di apprendimento automatico con particolare riferimento a tecniche innovative di intrusion detection. Titolo della tesi: “*Anomalies in Cyber Security: Detection, Prevention and Simulation Approaches*”. A partire da dicembre 2017, Luciano Argento ricopre una posizione di Assegnista di Ricerca presso il DIMES-UNICAL;

- *Luca Ferragina* (XXXV ciclo, Corso di Dottorato di Ricerca in *Information and Communication Technologies*, UNICAL): sta approfondendo tematiche di rappresentazione dei dati per la scoperta di anomalie mediante strumenti di apprendimento automatico con particolare riferimento al deep learning.
- *Simona Nisticò* (XXXVI ciclo, Corso di Dottorato di Ricerca in *Information and Communication Technologies*, UNICAL): sta approfondendo tematiche di eXplainable Artificial Intelligence (XAI) nell'ambito del machine learning e del deep learning.

### **6.3 Relatore di tesi di Laurea**

A partire dall'anno 2001 è stato relatore o co-relatore di numerose tesi di Laurea e Laurea Magistrale in Ingegneria Informatica e Tutor Accademico di diversi studenti tirocinanti presso l'Università della Calabria e le aziende del settore ICT dislocate sul territorio Calabrese convenzionate con l'UNICAL.

## 7 Attività di servizio nell'Università e negli Enti di Ricerca

Di seguito sono riportate le attività di servizio che Fabrizio Angiulli ha svolto presso l'Università e gli Enti di Ricerca.

- Membro aggregato per il settore Ingegneria dell'Informazione della Seconda Sessione 2007 degli Esami di Stato c/o UNICAL per l'abilitazione all'esercizio delle professioni di Ingegnere (vecchio ordinamento e nuovo ordinamento sezione A) ed Ingegnere Junior (nuovo ordinamento sezione B).
- Nel mese di ottobre del 2007 ha svolto l'incarico di Segretario della Commissione per l'elezione dei Presidenti dei seguenti Corsi di Laurea dell'UNICAL: C.d.L. Specialistica in Ingegneria Informatica, C.d.L. Specialistica in Ingegneria dell'Automazione, C.d.L. Specialistica in Ingegneria delle Telecomunicazioni e C.d.L. Triennale e Specialistica in Ingegneria Elettronica.
- Ha ricevuto dal C.d.L. in Ingegneria Informatica dell'UNICAL l'incarico di gestire, per le competenze del C.d.L., l'iniziativa "AICA Premia le Matricole in Informatica" promossa da AICA in collaborazione con GRIN (Gruppo di Informatica, l'Associazione Italiana dei Docenti Universitari di Informatica) e GII (Gruppo di Ingegneria Informatica, l'Associazione Italiana dei Docenti Universitari di Ingegneria Informatica), per (1) l'edizione 2007/08 (incarico ricevuto nel mese di febbraio 2008) e (2) l'edizione 2008/09 (incarico ricevuto nel mese di gennaio 2009).
- Nel mese di maggio 2008 è stato nominato dalla Facoltà di Ingegneria dell'UNICAL membro della Commissione di gestione dell'iniziativa "Percorsi azzeramento competenze", promossa dall'Assessorato Istruzione, Alta Formazione e Ricerca della Regione Calabria. Nell'ambito di tale commissione è stato responsabile dell'organizzazione dell'erogazione dei corsi di azzeramento di informatica destinati agli studenti di Ingegneria immatricolati nell'a.a. 2008/09.
- Nel mese di febbraio 2009 è stato nominato membro effettivo della Commissione Elettorale per la Votazione del Coordinatore del Collegio di Dottorato di Ricerca in Ingegneria dei Sistemi ed Informatica dell'UNICAL, nel corso della quale ha svolto il ruolo di segretario.
- Per l'a.a. 2009/10 ha svolto il ruolo di Coadiutore del Coordinatore del Collegio di Dottorato di Ricerca in Ingegneria dei Sistemi ed Informatica dell'UNICAL.
- Per l'a.a. 2009/10 è stato nominato referente per l'organizzazione del materiale in lingua inglese del Corso di Laurea in Ingegneria Informatica e del Corso di Laurea Specialistica in Ingegneria Informatica dell'UNICAL.
- A partire dal 2009 è stato Presidente o membro effettivo di numerose commissioni per l'assegnazione di Assegni di Ricerca, borse di studio e contratti di co.co.co. per il SSD ING-INF/05 presso l'Università della Calabria.
- Nel mese di maggio 2011 ha prestato supporto alla manifestazione d'orientamento d'ateneo "Campus live", durante la quale studenti di diversi istituti calabresi hanno visitato l'Università della Calabria.
- È stato membro effettivo di Commissione giudicatrice della selezione, per titoli e colloquio, per l'assunzione, con contratto di lavoro a tempo determinato, di unità di personale laureato, con profilo di Ricercatore, III livello professionale, presso l'Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche: (1) nel corso del 2012 e (2) nel corso del 2013.
- È stato membro effettivo della Commissione giudicatrice per l'ammissione al Corso di formazione per "Esperto di Tecnologie per la Valorizzazione del Turismo" bandito dal DIMES-UNICAL nell'ambito del PON "Ricerca e Competitività 2007-2013", i cui lavori si sono tenuti da luglio a settembre 2013.

- Per l'a.a. 2013/14 è stato nominato dal CCdS in Ingegneria Informatica dell'UNICAL membro effettivo della Commissione per la verifica dei contenuti e carico didattico degli insegnamenti della laurea magistrale.
- In data 17 gennaio 2014 è stato nominato membro effettivo della Commissione giudicatrice dei titoli per la conferma in ruolo dei Professori Associati del S.S.D. ING-INF/05 giudicati idonei nella procedura di valutazione comparativa per il medesimo settore bandita il 26/06/2008 presso l'Università di BERGAMO – Facoltà di Ingegneria. I lavori della commissione si sono tenuti da aprile a settembre 2014.
- Membro effettivo della Commissione Giudicatrice dell'esame di ammissione al Corso di Dottorato di Ricerca XXXI ciclo in Information and Communication Technologies presso l'UNICAL, i cui lavori si sono svolti nel mese di ottobre 2015.
- Membro effettivo della Commissione piani di studio del Corso di Dottorato di Ricerca in Information and Communication Technologies presso l'UNICAL per l'a.a. 2016/17.
- Presidente di sedute di Laurea in Ingegneria Informatica del DIMES-UNICAL nel corso (1) dell'a.a. 2015/16, (2) dell'a.a. 2016/17 e (3) dell'a.a. 2017/18.
- Presidente della Commissione per l'ammissione al Corso di Laurea Magistrale in Ingegneria Informatica del DIMES-UNICAL per l'a.a. 2013/14 (lavori da settembre a dicembre 2013), l'a.a. 2014/15 (lavori da settembre a dicembre 2014), l'a.a. 2015/16 (lavori da settembre a dicembre 2015) l'a.a. 2016/17 (lavori da settembre a dicembre 2016), l'a.a. 2018/19 (lavori da settembre a dicembre 2018). l'a.a. 2019/20 (lavori da settembre a dicembre 2019) e l'a.a. 2020/21 (lavori da settembre a dicembre 2020).
- Membro effettivo della Commissione per l'ammissione al II e III anno del Corso di Dottorato di Ricerca in Information and Communication Technologies presso l'UNICAL, i cui lavori si sono svolti nel mese di ottobre 2017.
- Presidente e membro effettivo della Commissione per il conferimento di incarichi di insegnamento mediante affidamento e contratto di diritto privato per il SSD ING-INF/05 del DIMES-UNICAL per l'a.a. 2015/16 (lavori svolti nel periodo settembre-ottobre 2015 e di cui ha presieduto diverse riunioni), l'a.a. 2016/17 (lavori svolti nel corso dell'intero anno accademico e di cui ha presieduto diverse riunioni) l'a.a. 2017/18 con nomina a presidente della commissione, l'a.a. 2019/20 con nomina a presidente della commissione e l'a.a. 2020/21 con nomina a presidente della commissione.
- Membro effettivo della Commissione Qualità del CdS in Ingegneria Informatica dell'UNICAL, per il supporto al processo di Assicurazione della Qualità della Didattica, per l'a.a. 2013/14, l'a.a. 2014/15, l'a.a. 2015/16, l'a.a. 2016/17 e l'a.a. 2017/18.
- Nel mese di marzo 2017 è stato nominato dal Rettore dell'Università della Calabria (D.R. n. 304 del 14/03/2017) componente del Comitato Tecnico Scientifico (CTS) del Centro Information and Communication Technology (ICT) d'Ateneo. Nel corso dell'incarico, ricoperto fino al mese di agosto 2017, si è occupato di problematiche inerenti i sistemi informativi dell'Ateneo, con particolare riferimento all'attivazione e configurazione dei sistemi ESSE3 (Servizi e Segreteria Studenti) e IRIS (Institutional Research Information System) di CINECA presso l'UNICAL.
- È stato membro effettivo della Commissione che in data 15 novembre 2017 ha conferito la Laurea Magistrale Honoris Causa in Ingegneria Informatica del DIMES-UNICAL al Prof. Ronald Fagin.
- È stato membro effettivo della Commissione che in data 28 maggio 2019 ha conferito il Dottorato di Ricerca Honoris Causa in Information and Communication Technologies al Prof. Carlo Zaniolo

## 8 Elenco delle Pubblicazioni

### Riviste Internazionali

- [1] G. Pang, F. Angiulli, M. Cucuringu, H. Liu. Guest Editorial: Non-IID Outlier Detection in Complex Contexts. Special Issue of *IEEE Intelligent Systems (IS)*, 2021.
- [2] F. Angiulli, F. Fassetti. Uncertain Outlier Detection with Arbitrarily Shaped Data Objects. *Journal of Intelligent Information Systems (JIIS)*, Springer, published online: 25 October 2020.
- [3] F. Angiulli, S. Basta, S. Lodi, C. Sartori. Reducing distance computations for distance-based outliers. *Expert Systems with Applications (ESWA)*, Elsevier, 147:113215, June 2020.
- [4] F. Angiulli. CFOF: A Concentration Free Measure for Anomaly Detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(1):Article 4, January 2020.
- [5] F. Angiulli, E. Narvaez. Pruning Strategies for Nearest Neighbor Competence Preservation Learners. *Neurocomputing (NEUCOM)*, Elsevier, 308:8-20, September 2018.
- [6] F. Angiulli. On the Behavior of Intrinsically High-Dimensional Spaces: Distances, Direct and Reverse Nearest Neighbors, and Hubness. *Journal of Machine Learning Research (JMLR)*, 18(170):1-60, 2018.
- [7] F. Angiulli, L. Argento, A. Furfaro, A. Parise. A Hierarchical Hybrid Framework for Modelling Anomalous Behaviours. *Simulation Modelling Practice and Theory (SIMPAT)*, Elsevier, 82:103-115, March 2018.
- [8] F. Angiulli, L. Argento, A. Furfaro. Exploiting content spatial distribution to improve detection of intrusions. *ACM Transactions on Internet Technology (TOIT)*, 18(2):Article 25, February 2018.
- [9] F. Angiulli, F. Fassetti, G. Manco, L. Palopoli. Outlying Property Detection with Numerical Attributes. *Data Mining and Knowledge Discovery (DAMI)*, Springer, 31(1):134-163, 2017.
- [10] F. Angiulli, S. Basta, S. Lodi, C. Sartori. GPU Strategies for Distance-based Outlier Detection. *IEEE Transactions on Parallel and Distributed Systems (TPDS)*, 27(11):3256-3268, 2016.
- [11] F. Angiulli, F. Fassetti. Towards generalizing the unification with statistical outliers: the Gradient Outlier Factor measure. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(3):Article 27, January 2016.
- [12] F. Angiulli, R. Ben-Eliyahu-Zohary, L. Palopoli. Restricted default theories: expressive power and outlier detection tasks. *Theoretical Computer Science (TCS)*, 564:107-130, January 2015.
- [13] F. Angiulli, R. Ben-Eliyahu - Zohary, F. Fassetti, L. Palopoli. On the Tractability of Minimal Model Computation for Some CNF Theories. *Artificial Intelligence (AIJ)*, Elsevier, 210:56-77, May 2014.
- [14] F. Angiulli, F. Fassetti. Exploiting Domain Knowledge to Detect Outliers. *Data Mining and Knowledge Discovery (DAMI)*, Springer, 28(2):519-568, March 2014.
- [15] N. Di Mauro, P. Frasconi, F. Angiulli, D. Bacciu, M. de Gemmis, F. Esposito, N. Fanizzi, S. Ferilli, M. Gori, F. A. Lisi, P. Lops, D. Malerba, A. Micheli, M. Pelillo, F. Ricci, F. Riguzzi, L. Saitta, G. Semeraro. Italian Machine Learning and Data Mining research: The last years. *Intelligenza Artificiale*, Special Issue Celebrating 25 years of the Italian Association for Artificial Intelligence, 7(2):77-89, November 2013.
- [16] F. Angiulli, F. Fassetti. Nearest Neighbor-Based Classification of Uncertain Data. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 7(1):Article 1, March 2013.



- [17] F. Angiulli, S. Basta, S. Lodi, C. Sartori. Distributed Strategies for Mining Outliers in Large Data Sets. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 25(7):1520-1532, July 2013.
- [18] F. Angiulli, F. Fassetti, L. Palopoli. Discovering Characterizations of the Behavior of Outlier Sub-Populations. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 25(6):1280-1292, June 2013.
- [19] F. Angiulli, F. Fassetti. Indexing Uncertain Data in General Metric Spaces. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 24(9):1640-1657, September 2012.
- [20] F. Angiulli. Prototype-based Domain Description for One-Class Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 34(6):1131-1144, June 2012.
- [21] F. Angiulli, G. Greco, L. Palopoli. Detecting and Repairing Anomalous Evolutions in Noisy Environments. *Annals of Mathematics and Artificial Intelligence (AMAI)*, Springer, 60(3-4):179-228, December 2010.
- [22] F. Angiulli, R. Ben-Eliyahu - Zohary, L. Palopoli. Outlier detection for simple default theories. *Artificial Intelligence (AIJ)*, Elsevier, 174(15):1247-1253, October 2010.
- [23] F. Angiulli, F. Fassetti. Distance-based outlier queries in data streams: the novel task and algorithms. *Data Mining and Knowledge Discovery – Special Issue on Outlier Detection for Knowledge Discovery (DAMI)*, Springer, 20(2):290-324, March 2010.
- [24] F. Angiulli, A. Astorino. Scaling Up Support Vector Machines using Nearest Neighbor Condensation. *IEEE Transactions on Neural Networks (TNN)*, 21(2):351-357, February 2010.
- [25] F. Angiulli, F. Fassetti, L. Palopoli. Detecting Outlying Properties of Exceptional Objects. *ACM Transactions on Database Systems (TODS)*, 34(1):Article 7, April 2009.
- [26] F. Angiulli, F. Fassetti. DOLPHIN: an Efficient Algorithm for Mining Distance-Based Outliers in Very Large Datasets. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1):Article 4, March 2009.
- [27] F. Angiulli, R. Ben-Eliyahu - Zohary, L. Palopoli. Outlier Detection using Default Reasoning. *Artificial Intelligence (AIJ)*, Elsevier, 172(16-17):1837-1872, November 2008.
- [28] F. Angiulli, E. Cesario, C. Pizzuti. Random Walk Biclustering for Microarray Data. *Information Sciences (INS)*, Elsevier, 178(6):1479-1497, March 2008.
- [29] F. Angiulli, G. Greco, L. Palopoli. Outlier Detection by Logic Programming. *ACM Transactions on Computational Logic (TOCL)*, 9(1):Article No. 7, December 2007.
- [30] F. Angiulli, G. Folino. Distributed Nearest Neighbor Based Condensation of Very Large datasets. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 19(12):1593-1606, December 2007.
- [31] F. Angiulli. Fast Nearest Neighbor Condensation for Large dataset Classification. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 19(11):1450-1464, November 2007.
- [32] F. Angiulli. Condensed Nearest Neighbor Data Domain Description. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 29(10):1746-1758, October 2007.
- [33] F. Angiulli, S. Basta, C. Pizzuti. Distance-Based Detection and Prediction of Outliers. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 18(2):145-160, February 2006.

- [34] F. Angiulli. Enumerating Consistent Metaquery Instantiations. *AICommunications*, IOS Press, 18(2):117-141, 2005.
- [35] F. Angiulli, C. Pizzuti. An Approximate Algorithm for Top- $k$  Closest Pairs Join Query in Large High Dimensional Data. *Data & Knowledge Engineering (DKE)*, Elsevier, 53(3):263-281, June 2005.
- [36] F. Angiulli, C. Pizzuti. Approximate  $k$ -Closest-Pairs in Large High-Dimensional datasets. *Journal of Mathematical Modelling and Algorithms (JMMA)*, Kluwer, 4(2):149-179, June 2005.
- [37] F. Angiulli, C. Pizzuti. Outlier Mining in Large High-Dimensional datasets. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 17(2):203-215, February 2005.
- [38] F. Angiulli, G. Ianni, L. Palopoli. On the Complexity of Inducing Categorical and Quantitative Association Rules. *Theoretical Computer Science (TCS)*, Elsevier, 314(1-2):217-249, 25 February 2004.
- [39] F. Angiulli, R. Ben-Eliyahu-Zohary, G. Ianni, L. Palopoli. Computational Properties of Metaquerying Problems. *ACM Transactions on Computational Logic (TOCL)*, 4(2):149-180, April 2003.
- [40] F. Angiulli, L. Palopoli, R. Torlone. The GPR System: An Architecture for Integrating Active and Deductive Rules for Complex Database Objects. *Theory and Practice of Object Systems (TAPOS)*, John Wiley & Sons, 3(4):285-301, 1997.

## Conferenze e Capitoli di Libro

- [41] F. Angiulli, F. Fassetti, L. Ferragina. Improving deep unsupervised anomaly detection by exploiting VAE latent space distribution. In *Proceedings of the 23rd International Conference on Discovery Science (DS)*, Thessaloniki, Greece, October 19-21, 2020.
- [42] F. Angiulli, F. Fassetti, L. Palopoli, C. Serrao. A Density Estimation Approach for Detecting and Explaining Exceptional Values in Categorical Data. In *Proceedings of the 22nd International Conference on Discovery Science (DS)*, Split, Croatia, October 28-30, 2019.
- [43] A. Furfaro, L. Argento, D. Saccà, F. Angiulli, F. Fassetti. An Infrastructure for Service Accountability based on Digital Identity and Blockchain 3.0. In *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM)*, Paris, France, April 2-May 2, 2019.
- [44] F. Angiulli, F. Fassetti, A. Furfaro, A. Piccolo, D. Saccà. Achieving Service Accountability through Blockchain and Digital Identity. In *Proceedings of the 30th International Conference on Advanced Information Systems Engineering Forum (CAiSE)*, Tallin, Estonia, June 11-15, 2018.
- [45] F. Angiulli. Data Mining: Outlier Detection. In *Encyclopedia of Bioinformatics and Computational Biology*, Elsevier, 2018.
- [46] F. Angiulli, F. Fassetti, L. Palopoli and D. Ursino. A Tour from Regularities to Exceptions. In *A comprehensive guide through the Database research over the last 25 years*, Series *Studies in Big Data*, vol. 31, p. 307-322, Springer, 2018.
- [47] F. Angiulli. Concentration Free Outlier Detection. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD)*, Skopje, Macedonia, September 18-22, 2017. Series *Lecture Notes in Computer Science*, Springer, vol. 10534, p. 3-19.

- [48] F. Angiulli, R. Ben-Eliyahu-Zohary, F. Fassetti and L. Palopoli. Modular Construction of Minimal Models. *Proceedings of the 14th International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR)*, Hanasaari Conference Center, Espoo, Finland, July 3-6, 2017. Series *Lecture Notes in Computer Science*, Springer, vol. 10377, p. 43-48.
- [49] F. Angiulli, F. Fassetti, E. Narvaez. Anomaly Detection in Networks with Temporal Information. *Proceedings of the 19th International Conference on Discovery Science (DS)*, Bari, Italy, October 19-21, 2016. Series *Lecture Notes in Computer Science*, Springer, vol. 10558, p. 359-375.
- [50] F. Angiulli, R. Ben-Eliyahu-Zohary, F. Fassetti and L. Palopoli. Decomposing Minimal Models. *Proceedings of the International Workshop on Knowledge-based techniques for problem solving and reasoning (KnowProS)*, in conjunction with the *International Joint Conference on Artificial Intelligence*, New York, U.S.A., July 10, 2016.
- [51] F. Angiulli, E. Narvaez. Pruning Nearest Neighbor Competence Preservation Learners. *Proceedings of the 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, Vietri sul Mare, Italy, November 9-11, 2015.
- [52] F. Angiulli, L. Argento, A. Furfaro. Exploiting n-gram location for intrusion detection. *Proceedings of the 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, Vietri sul Mare, Italy, November 9-11, 2015.
- [53] S. Lodi, F. Angiulli, S. Basta, D. Luiselli, L. Pagani, C. Sartori. First Application of Distance-based Outlier Approach to Detect Highly Differentiated Genomic Regions across Human Populations. In *Mathematical Models in Biology, Bringing Mathematics to Life*, V. Zazzu, M.B. Ferraro, M.R. Guarracino Eds., ISBN 978-3-319-23496-0, Springer International Publishing, 2015.
- [54] F. Angiulli, S. Basta, S. Lodi, G. Moro, C. Sartori. Mining Big Data with High Performance Computing Solutions. *Proceedings of the 10th Scientific Meeting of the Classification and Data Analysis Group of the Italian Statistical Society (CLADAG)*, Santa Margherita di Pula, Cagliari, October 8-10, 2015. ISBN 978-88-8467-749-9.
- [55] F. Angiulli, R. Ben-Eliyahu-Zohary, L. Palopoli. Exploiting the HEF property to check model minimality. *Atti del 23rd Convegno Nazionale su Sistemi Evoluti per Basi di Dati (SEBD)*, Gaeta, June 14-17, 2015.
- [56] F. Angiulli. Outlier detection: efficient algorithms and definitions for specific settings. *Proceedings of the 7th International Conference of the ERCIM WG on Computational and Methodological Statistics (ERCIM)*, University of Pisa, Italy, Dicembre 6-8, 2014.
- [57] F. Angiulli, S. Basta, S. Lodi, C. Sartori. Many-core Distance-based Outlier Detection. *Proceedings of the Third Italian Workshop on Machine Learning and Data Mining (MLDM.it)* in conjunction with *XIII AI\*IA Symposium on Artificial Intelligence*, 2015.
- [58] F. Angiulli, S. Basta, S. Lodi, C. Sartori. Accelerating outlier detection with intra- and inter-node parallelism. *Proceedings of the International Conference on High Performance Computing & Simulation. (HPCS)*, Bologna, Italy, July 21-25, 2014.
- [59] F. Angiulli, F. Fassetti. Outlier Detection with Arbitrary Probability Functions. *Proceedings of the thirteenth International Conference on Advances in Artificial Intelligence (AI\*IA)*, Turin, Italy, December 4-6, 2013. Series *Lecture Notes in Artificial Intelligence*, Springer, vol. 8249, p. 421-432.
- [60] F. Angiulli, F. Fassetti. Principal Directions-based Pivot Placement. *Proceedings of the International Conference on Similarity Search and Applications (SISAP)*, A Coruna, Spain, October 2-4, 2013. Series *Lecture Notes in Computer Science*, Springer, vol. 8199, p. 85-90.

- [61] F. Angiulli, S. Basta, S. Lodi, C. Sartori. Fast Outlier Detection using a GPU. *Proceedings of the International Conference on High Performance Computing & Simulation*. (HPCS), Helsinki, Finland, July 1-5, 2013.
- [62] F. Angiulli, R. Ben-Eliyahu-Zohary, L. Palopoli. Tractable Strong Outlier Identification. *Proceedings of the 9th International Workshop on Nonmonotonic Reasoning, Action and Change* (NRAC), in conjunction with the *International Joint Conference on Artificial Intelligence*, Barcelona, Spain, July 16-22, 2011.
- [63] F. Angiulli, E. Masciari. Effectively Monitoring RFID based systems. *Proceedings of the 14th East-European Conference on Advances in Databases and Information Systems* (ADBIS), Novi Sad, September 20-24, 2010. Series *Lecture Notes in Computer Science*, Springer, vol. 6295, p. 31-45.
- [64] F. Angiulli, S. Basta, S. Lodi, C. Sartori. A Distributed Approach to Detect Outliers in Very Large Data Sets. *Proceedings of the 16th European Conference on Parallel Processing* (EuroPar), Ischia, Naples, Italy, August 31 - September 3, 2010. Series *Lecture Notes in Computer Science*, Springer, vol. 6271, p. 329-340.
- [65] F. Angiulli, E. Masciari. SMART: Simple Monitoring enterprise Activities by RFID Tags. *Proceedings of the 22nd International Conference on Advanced Information Systems Engineering Forum* (CAiSE), Hammamet, Tunisia, June 7-11, 2010.
- [66] F. Angiulli, F. Fassetti. Finding Distance-Based Outliers in Subspaces through both Positive and Negative Examples. *Proceedings of the International Conference on Agents and Artificial Intelligence* (ICAART), Valencia, Spain, January 22-24, 2010.
- [67] F. Angiulli, F. Fassetti, L. Palopoli, D. Trimboli. Detection of Discriminating Rules. *Proceedings of the International Conference on Agents and Artificial Intelligence* (ICAART), Valencia, Spain, January 22-24, 2010.
- [68] F. Angiulli, F. Fassetti. Outlier Detection using Inductive Logic Programming. *Proceedings of the IEEE International Conference on Data Mining* (ICDM), Miami, Florida, December 6-9, 2009.
- [69] F. Angiulli. Outlier Detection Techniques for Data Mining. In *Encyclopedia of Data Warehousing and Mining – 2nd Edition*, IGI Publishing, 2009.
- [70] F. Angiulli. Prototype-based Domain Description. *Proceedings of the 18th European Conference on Artificial Intelligence* (ECAI), Patras, Greece, July 21-25, 2008.
- [71] F. Angiulli, G. Folino. A Grid-based Architecture for Nearest Neighbor Based Condensation of Huge Datasets. *Proceedings of the ACM/IEEE International Symposium on High Performance Distributed Computing Conference & Co-Located Workshops* (HPDC), Boston, MA, USA, June 23-27, 2008.
- [72] F. Angiulli, S. Basta. Optimal Subset Selection for Classification through SAT Encodings. *Proceedings of the IFIP Conference on Artificial Intelligence* (IFIP AI), Milan, Italy, September 7-10, 2008.
- [73] F. Angiulli, F. Fassetti. An Efficient Method for Outlier Detection. *Atti del Sedicesimo Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD), Palermo, 2008.
- [74] F. Angiulli, F. Fassetti, L. Palopoli. An Unsupervised Outlier Detection Approach for Cleaning String Data Entries. *Atti del Sedicesimo Convegno Nazionale su Sistemi Evoluti per Basi di Dati* (SEBD), Palermo, 2008.

- [75] F. Angiulli, V. Fionda, S. Rombo. Protein Data Condensation for Effective Quaternary Structure Classification. *Proceedings of the 8th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*, Birmingham, UK, December 16-19, 2007. Series *Lecture Notes in Computer Science*, Springer, vol. 4881, p. 810-820.
- [76] F. Angiulli, F. Fassetti. Very Efficient Mining of Distance-Based Outliers. *Proceedings of the ACM Sixteenth Conference on Information and Knowledge Management (CIKM)*, Lisboa, Portugal, November 6-9, 2007.
- [77] F. Angiulli, F. Fassetti. Detecting Distance-Based Outliers in Streams of Data. *Proceedings of the ACM Sixteenth Conference on Information and Knowledge Management (CIKM)*, Lisboa, Portugal, November 6-9, 2007.
- [78] F. Angiulli, G. Folino. Efficient Distributed Data Condensation for Nearest Neighbor Classification. *Proceedings of the 13th European Conference on Parallel and Distributed Computing (Euro-Par)*, Rennes, France, August 28-31, 2007. Series *Lecture Notes in Computer Science*, Springer, vol. 4641, p. 338-347.
- [79] F. Angiulli, A. Astorino. Applying SVMs to large datasets through Nearest Neighbor Condensation. In *Proceedings of the 2nd Conference on Optimization Methods & Software (EUROPT-OMS)*, Prague, Czech Republic, July 4-7, 2007.
- [80] F. Angiulli, G. Greco, L. Palopoli. The LP-OD System: Logic Programming meets Outlier Detection. *Proceedings of the Ninth International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR)*, Tempe, AZ, May 14-17, 2007. Series *Lecture Notes in Computer Science*, Springer, vol. 4483, p. 254-259.
- [81] F. Angiulli, A. Astorino. Classification problems via SVM and Nearest Neighbor Condensation. *Atti della XXXVIII Annual Conference of the Italian Operations Research Society Optimization and Decision Sciences (AIRO)*, Genova, Italy, September 5-8, 2007.
- [82] F. Angiulli, E. Cesario, C. Pizzuti. A Greedy Search Approach to Co-clustering Sparse Binary Matrices. *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, Washington D.C., November 13-15, 2006.
- [83] F. Angiulli. Clustering by Exceptions. *Proceedings of the 21th National Conference on Artificial Intelligence (AAAI)*, Boston, Massachusetts, July 16-20, 2006.
- [84] F. Angiulli, G. Greco, L. Palopoli. Detecting and Repairing Anomalous Evolutions in Noisy Environments: Logic Programming Formalization and Complexity Results. *Proceedings of the IFIP Conference on Artificial Intelligence (IFIP AI)*, Santiago, Chile, August 21-24, 2006.
- [85] F. Angiulli, F. Fassetti, L. Palopoli. Un Metodo per la Scoperta di Proprietà Inattese. *Atti del Quattordicesimo Convegno su Sistemi Evoluti per Basi di Dati (SEBD)*, Portonovo, Ancona, 2006.
- [86] F. Angiulli. Condensed Nearest Neighbor Data Domain Description. *Proceedings of the 6th International Symposium on Intelligent Data Analysis (IDA)*, Madrid, Spain, September 8-10, 2005. Series *Lecture Notes in Computer Science*, Springer, vol. 3646, p. 12-26.
- [87] F. Angiulli, C. Pizzuti. Gene Expression Biclustering using Random Walk Strategies. *Proceedings of the 7th International Conference on Data Warehousing and Knowledge Discovery (DaWaK)*, Copenhagen, Denmark, August 22-26, 2005. Series *Lecture Notes in Computer Science*, Springer, vol. 3589, p. 509-519.
- [88] F. Angiulli. Fast Condensed Nearest Neighbor Rule. *Proceedings of the 22nd International Conference on Machine Learning (ICML)*, Bonn, Germany, August 7-11, 2005.

- [89] F. Angiulli, S. Basta, C. Pizzuti. Detection and Prediction of Distance-Based Outliers. *Proceedings of the 20th ACM Symposium on Applied Computing (SAC)*, Santa Fe, New Mexico, March 13-17, 2005.
- [90] F. Angiulli, S. Basta, C. Pizzuti. Improving Prediction of Distance-Based Outliers. *Proceedings of the 7th International Conference on Discovery Science (DS)*, Padova, Italy, October 2-5, 2004. Series *Lecture Notes in Artificial Intelligence*, Springer, vol. 3245, p. 89-100.
- [91] F. Angiulli, G. Greco, L. Palopoli. Detecting Outliers via Logical Theories and its Data Complexity. *Proceedings of the 7th International Conference on Discovery Science (DS)*, Padova, Italy, October 2-5, 2004. Series *Lecture Notes in Artificial Intelligence*, Springer, vol. 3245, p. 101-113.
- [92] F. Angiulli, G. Greco, L. Palopoli. Discovering Anomalies in Evidential Knowledge by Logic Programming. *Proceedings of the 9th European Conference on Logics in Artificial Intelligence (JELIA)*, Lisbon, Portugal, September 27-30, 2004. Series *Lecture Notes in Computer Science*, Springer, vol. 3229, p. 578-590.
- [93] F. Angiulli, C. Pizzuti, M. Ruffolo. DESCRy: A Grid and Density Based Clustering Algorithm for Very Large datasets. *Proceedings of the 5th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL)*, Exeter, UK, August 25-27, 2004. Series *Lecture Notes in Computer Science*, Springer, vol. 3177, p. 203-210.
- [94] F. Angiulli, R. Ben-Eliyahu-Zohary, L. Palopoli. Outlier Detection using Disjunctive Logic Programming. *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI)*, Valencia, Spain, August 22-27, 2004. FAIA, Vol. 110, Pages 416-419, IOS Press, The Netherlands.
- [95] F. Angiulli, C. Pizzuti. Top- $k$  Closest Pairs Join Query: An Approximate Algorithm for Large High Dimensional Data. *Proceedings of the 8th IEEE International Database Engineering & Applications Symposium (IDEAS)*, July 2004, Coimbra, Portugal. IEEE Computer Society, California, USA.
- [96] F. Angiulli, R. Ben-Eliyahu-Zohary, L. Palopoli. Outlier Detection using Default Logic. *Answer Set Programming: Advances in Theory and Implementation (ASP)*, Messina, Italy, September 26-28, 2003.
- [97] F. Angiulli, R. Ben-Eliyahu-Zohary. Exploiting Default Logic to Detect Outliers. *Proceedings of the APPIA-GULP-PRODE Joint Conference on Declarative Programming*, September 2003, Reggio Calabria, Italy.
- [98] F. Angiulli, R. Ben-Eliyahu-Zohary, L. Palopoli. Outlier Detection using Default Logic. *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, August 2003, Acapulco, Mexico.
- [99] F. Angiulli, T. Catarci, P. Ciaccia, G. Ianni, S. Kimani, S. Lodi, M. Patella, G. Santucci, C. Sartori. An Integrated Data Mining and Data Presentation Tool. *Proceedings of the Third International Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields*, September 25-27, 2002, Bologna, Italy.
- [100] F. Angiulli, C. Pizzuti. Approximate  $k$ -Closest Pairs with Space-Filling Curves. *Proceedings of the International Conference on Data Warehouse and Knowledge Discovery (DaWaK)*, Aix-en-Provence, France, September 2002. Series *Lecture Notes in Computer Science*, Springer, vol. 2454, p. 124-134.
- [101] F. Angiulli, C. Pizzuti. Fast Outlier Detection in High Dimensional Spaces. *Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, Helsinki, Finland, August 2002. Series *Lecture Notes in Computer Science*, Springer, vol. 2431, p. 15-26.

- [102] F. Angiulli, G. Ianni, L. Palopoli. On the complexity of mining association rules. *Atti del Nono Convegno su Sistemi Evoluti per Basi di Dati (SEBD)*, Venezia, 27-29 Giugno, 2001.
- [103] F. Angiulli, R. Ben-Eliyahu-Zohary, G. Ianni, L. Palopoli. Computational Properties of Metaquerying Problems. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, May 15-17, 2000, Dallas, Texas, USA, pages 237-244.
- [104] F. Angiulli, G.B. Ianni, L. Palopoli. Metaquerying: Proprietà e Tecniche di Implementazione. *Atti dell'Ottavo Convegno su Sistemi Evoluti per Basi di Dati (SEBD)*, L'Aquila, Giugno 26-28, 2000.

## Manoscritti

- [105] F. Angiulli, F. Fassetti, L. Palopoli, C. Serrao. A density estimation approach for detecting and explaining exceptional values in categorical data. *Manoscritto* sottomesso alla rivista internazionale *Machine Learning*, accepted under minor revision, November 2020.
- [106] F. Angiulli, F. Fassetti, L. Ferragina. *LatentOut*: an unsupervised deep anomaly detection approach exploiting latent space distribution. *Manoscritto* sottomesso alla rivista internazionale *Machine Learning*, March 2021.
- [107] F. Angiulli, R. Ben-Eliyahu-Zohary, F. Fassetti, L. Palopoli. Graph-Based Construction of Minimal Models. *Manoscritto* sottomesso alla rivista internazionale *Artificial Intelligence*, under major revision, 2020.
- [108] F. Angiulli, T. Colombo, F. Fassetti, A. Furfaro, P. Paci. Exploiting sponge effects in RNA expression data for discriminating healthy and unhealthy tissues. *Manoscritto* sottomesso ad una rivista internazionale, 2020.
- [109] F. Angiulli, F. Fassetti, S. Nisticò. Local Interpretable Classifier Explanations with Self-generated Semantic Features. *Manoscritto* sottomesso alla conferenza internazionale *ECMLPKDD*, 2021.
- [110] F. Angiulli, F. Fassetti, C. Serrao. Detecting Outliers in Networks via Stochastic Block Models. *Manoscritto* sottomesso alla conferenza internazionale *DEXA*, 2021.
- [111] F. Angiulli, F. Fassetti, C. Serrao. ODCA: an Outlier Detection Approach to Deal with Correlated Attributes. *Manoscritto* sottomesso alla conferenza internazionale *DaWaK*, 2021.

*Autorizzo il trattamento dei miei dati personali contenuti nel presente curriculum vitae ai sensi del D. Lgs. 196/2003 e dell'art. 13 GDPR 679/16.*

FIRMA