

Curriculum vitae e attività di ricerca di Giorgio Valentini

Giorgio Valentini si è laureato con lode in Biologia all'Università di Genova nel 1981.

Dal 1982 al 1999 è stato docente di biologia e chimica in diversi istituti superiori. Nel 1999 si è laureato con lode in Informatica presso l'Università di Genova. Nello stesso anno ha ottenuto una borsa di ricerca dell'INFM (Istituto Nazionale di Fisica della Materia). Nel 2003 ha conseguito il dottorato in Informatica all'Università di Genova. Nel 2001 ha usufruito di una borsa di studio presso il dipartimento di Informatica della Oregon State University.

Dal 2003 al 2004 è stato assegnista presso il DSI, Dipartimento di Scienze dell'Informazione dell'Università degli Studi di Milano. Dal gennaio 2005 all'ottobre 2010 è stato ricercatore presso il DSI. Da novembre 2010 è professore associato nel medesimo dipartimento dell'Università degli Studi di Milano.

Attività di ricerca

L'attività di ricerca si articola in due aree principali, *bioinformatica* e *apprendimento automatico*.

Schema delle linee di ricerca

- I. Bioinformatica
 - A) Analisi, sviluppo ed applicazione in ambito bioinformatico di metodi di apprendimento automatico non supervisionato
 - A1. *Metodi basati sull'analisi della stabilità per la valutazione dell'affidabilità dei cluster individuati in dati bio-molecolari complessi*
 - A2. *Metodi di ensemble clustering per la ricerca di pattern in dati bio-molecolari*
 - B) Analisi sviluppo ed applicazione in ambito bioinformatico di metodi di apprendimento automatico supervisionato e semi-supervisionato
 - B1. *Analisi e sviluppo di metodi di ensemble supervisionati per il supporto alla diagnosi bio-molecolare.*
 - B2. *Integrazione di metodi di feature extraction e feature selection e di dati bio-molecolari complessi per la classificazione supervisionata di geni co-espressi*
 - B3. *Classificazione funzionale di geni e proteine su scala genomica*
 - B4. *Integrazione di sorgenti multiple di dati per la predizione delle funzioni geniche*
 - B5. *Analisi semi-supervisionata di reti biologiche.*
 - B6. *Predizione della localizzazione subcellulare delle proteine*
 - C) Altre attività di ricerca in ambito bioinformatico
 - C1. *Modellazione biologicamente motivata dei profili di espressione.*
 - C2. *Metodi per la selezione dei geni "Gene Ontology driven".*
 - C3. *Analisi di dati di DNA microarray per lo studio della leucemia mieloide*
- II. Apprendimento automatico
 - A) Analisi e progettazione di metodi di ensemble e di metodi di machine learning
 - A1. *Metodi di ensemble a codici a correzione d'errore per la classificazione multiclasse.*
 - A2. *Metodi di ensemble basati sulla scomposizione dell'errore in bias e varianza*
 - A3. *Metodi di ensemble supervisionati basati su proiezioni randomizzate*
 - A4. *Metodi di ensemble gerarchici multiclasse, multi-etichetta e multi-path*
 - A5. *Metodi di ensemble clustering*
 - A6. *Metodi di label propagation per l'analisi semi-supervisionata di grafi*
 - B) Sviluppo di librerie software di machine learning

Una breve descrizione delle principali linee di ricerca è disponibile dalle pagine personali:
<http://homes.dsi.unimi.it/~valenti/research.activity.html>

Progetti di ricerca.

Ha contribuito ai progetti di ricerca nazionali MIUR COFIN-PRIN *Tecniche di apprendimento automatico in ambito bioinformatico per l'analisi e la modellazione di dati di espressione genica funzionale e strutturale* (coordinatore Paolo Frasconi) e *Linguaggi formali ed automi: metodi, modelli ed applicazioni* (coordinatore Alberto Bertoni).

Ha collaborato con Tom Dietterich (Computer Science Dept., Oregon State University) per lo sviluppo di metodi di ensemble ECOC e metodi di ensemble basati sull'analisi bias-varianza, nell'ambito del progetto NSF "*From Bits to Information: Statistical Learning Technologies for Digital Information Management Search*" (Brown University, M.I.T, Oregon State University, and University of Illinois at Urbana-Champaign).

Ha fatto parte della rete di eccellenza *Pattern Analysis, Statistical Modelling and Computational Learning (PASCAL)* nell'ambito del VI Programma Quadro dell'Unione Europea ed attualmente è membro dell'unità milanese della rete di eccellenza *PASCAL2*, nell'ambito del VII Programma Quadro dell'Unione Europea.

Partecipa al progetto PUR 2009 "Metodi automatici per l'analisi di pattern in ambito biomedico" finanziato dall'Università degli Studi di Milano.

E' membro associato del CNR – IEIIT (Istituto di Elettronica e Ingegneria delle Telecomunicazioni) per il progetto "*Machine learning techniques for modelling and growing up*".

E' responsabile dell'unità bioinformatica del DSI per l'analisi di dati di DNA microarray nell'ambito di un progetto comune con il Dipartimento di Biologia e Genetica di Medicina (Università degli Studi di Milano) e con l'Ospedale Niguarda di Milano per la diagnosi e terapia bio-molecolare delle leucemie mieloidi.

Collabora a progetti di ricerca con il Dipartimento di Scienze Biomolecolari e Biotecnologie dell'Università degli Studi di Milano. Collabora inoltre alle attività didattiche e di ricerca della rete Nazionale di Bioinformatica Oncologica.

E' responsabile del laboratorio di bioinformatica del DSI, e coordina l'attività di ricerca in bioinformatica all'interno del medesimo dipartimento.

Collabora con il Dipartimento di Bioinformatica del Centro de Investigacion Principe Felipe di Valencia (ricerca di geni correlati a patologie tumorali con metodi di selezione che incorporano conoscenza biologica a priori) e con il dipartimento di Informatica della Aristotle University of Thessaloniki (metodi di classificazione multilabel e gerarchici ed integrazione di dati bio-molecolari complessi per la predizione delle funzioni geniche).

Organizzazione di conferenze

Ha partecipato a comitati scientifici di diverse conferenze internazionali, ed ha organizzato conferenze internazionali e workshop sui metodi ensemble ed in ambito bioinformatico. In particolare, *CIBB 2007*, *Fourth International Conference on Computational Intelligence Methods for Bioinformatics and Biostatistics* a Portofino e *SUEMA 2007*, *International Workshop on Supervised and Unsupervised Ensemble Methods and Their Application* (nell'ambito di IbPRIA2007) a Girona (Spagna).

Ha inoltre organizzato nel luglio del 2008 la seconda edizione di *SUEMA*, nell'ambito di *ECAI 2008* (European Conference on Artificial Intelligence) a Patrasso (Grecia), con il supporto della rete di eccellenza europea *Pascal2*. Nel 2009 ha contribuito all'organizzazione del workshop NIPS "Learning from Multiple Sources with Applications to Robotics" a Whistler (Canada). E' inoltre membro del comitato scientifico di diverse conferenze e workshop internazionali nell'ambito della bioinformatica e dell'apprendimento automatico.

Nel 2010 ha organizzato la terza edizione di *SUEMA*, nell'ambito di *ECML 2010* (European Conference on Machine Learning) a Barcellona (Spagna), con il supporto della rete di eccellenza europea *Pascal2*.

Attività editoriale

Fa parte dell'editorial board dell'*Open Applied Informatics* e dell'*Open Medical Informatics Journal*. E' guest editor di *Artificial Intelligence in Medicine* e svolge attività di reviewer per le principali riviste internazionali di bioinformatica (*Bioinformatics*, *BMC Bioinformatics*, *Briefings in Bioinformatics*, *IEEE ACM Transactions on Computational Biology and Bioinformatics* ed altre) e per diverse riviste di apprendimento automatico, tra cui *Journal of Machine Learning Research*, *Machine Learning*, *IEEE Transactions on PAMI*, *IEEE Transactions on Neural Networks* e *Neurocomputing*. Ha curato l'edizione di un numero speciale di *Artificial Intelligence in Medicine* dedicato ai metodi di intelligenza computazionale in bioinformatica, e di due volumi delle *Lectures Notes in Computational Intelligence* dedicato ai metodi di ensemble ed alle loro applicazioni. E' autore o co-autore di 29 articoli pubblicati in riviste internazionali con peer-review e di più di altri 50 articoli, fra atti di conferenze internazionali e contributi a volumi collettivi (ad es. *Lecture Notes in Computer Science*).

Didattica

Attualmente (a.a. 2009/10) e' docente del corso "Bioinformatica" per la laurea magistrale in Informatica, e del corso "Metodi Bioinformatici" per la laurea magistrale in Biotecnologie Molecolari e Bioinformatica presso la Facoltà di Scienze MFN dell' Università degli Studi di Milano.

Corsi precedenti.

Presso la Facoltà di Scienze MFN, Università degli Studi di Milano:

- *Bioinformatica*: dal 2005/06 al 2008/09 (laurea magistrale in Informatica)
- *Linguaggi di programmazione per la Bioinformatica*: dal 2004/05 al 2008/09 (laurea magistrale in Genomica Funzionale e Bioinformatica)
- *Informatica applicata ai processi biotecnologici*: 2008/09 per la laurea in Biotecnologie Industriale e Ambientali.
- *Algoritmi per la Bioinformatica*: 2003/04 e 2004/05 (laurea magistrale in Genomica Funzionale e Bioinformatica)
- *Laboratorio di Bioinformatica*: 2004/05 (laurea in Biologia).

Presso la Facoltà di Medicina e Chirurgia, Università degli Studi di Milano:

- *Informatica*: 2007/08 (lauree in Fisioterapia e Dietistica)
- *Informatica*: 2006/07 (lauree in Podologia e Igiene Dentale)
- *Informatica*: 2004/05 e 2005/06 (laurea in Educazione Professionale)

Presso la Facoltà di Scienze MFN, Università degli Studi Genova

Bioinformatica Funzionale 1: 2002/03 e 2003/04 (laurea in Fisica)

Esercitazioni di Reti Neurali 1: 2001/02 (laurea in Scienze dell'Informazione)

Laboratorio di Programmazione di Sistema: 2001/02 (laurea in Scienze dell'Informazione)

Esercitazioni di Reti Neurali 2: 2000/01 (laurea in Scienze dell'Informazione)

Ha seguito come relatore o correlatore numerose tesi di laurea e laurea magistrale in Informatica e 4 tesi di dottorato in Informatica. Nel 2007 ha tenuto lezioni di bioinformatica per la Scuola internazionale di dottorato in Bioinformatica dell'Università di Brno nella Repubblica Ceca. Nel 2008 ha tenuto diversi lezioni di bioinformatica per la Scuola di Dottorato dell'Università di Palermo. Ha tenuto inoltre lezioni all' International School on Neural Nets "E.R. Caianiello" di Vietri. Dal 2008 è coordinatore del corso "Linguaggio R per la bioinformatica" per la Rete Nazionale di Bioinformatica Oncologica. E' stato invitato come relatore alla Quinta Conferenza Internazionale di Bioinformatica e Biostatistica CIBB 2008, e ha tenuto seminari di argomento bioinformatico in diverse università e centri di ricerca italiani e stranieri. Ha partecipato come relatore invitato al workshop internazionale "Statistical approaches and validation in clustering:

mixture models and nonparametric methods”, organizzato nell’ambito della rete di eccellenza *Pascal* all’Università di Caen in Francia.

Affiliazione ad associazioni nazionali ed internazionali.

E' membro dell'ISCB (International Society of Computational Biology), di BITS (Società Italiana di Bioinformatica), di INNS (International Neural Networks Society), e della SIREN (Società Italiana di Reti Neurali).

E' membro associato del CNR – IEIIT (Istituto di Elettronica e Ingegneria delle Telecomunicazioni).