Dipartimento di Informatica
Largo Bruno Pontecorvo, 3
56127 Pisa, Italy
Stanza 343 DN

Email: nicotra@di.unipi.it
Web: http://www.di.unipi.it/~nicotra
Tel: (+39) 050 2213143
Fax: (+39) 050 2212713

Luca Nicotra
Curriculum Vitae

February 2007

Personal

Date of birth: August 12, 1982

Place of birth: Monfalcone (GO), Italy

Nationality: Italian

Academic Education

Ph.D.. Computer Science (Gen 2007)
University of Pisa, Italy

M.S. Computer Science (Feb 2005 - Oct 2006)
University of Pisa, Italy
Graduated (with honors) *summa cum laude*
Thesis supervisors: Alessio Micheli and Antonina Starita
Thesis title: *Generative Kernel Functions for Structured Data*

B.S. Computer Science (Oct 2001 -Feb 2005)
University of Pisa, Italy
Graduated (with honors) *summa cum laude*

Visiting Student (Dec 2003-Jul 2004)
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
School of Computer Science
Enrolled as undergraduate student at the Computer Science School, attending graduate courses from the Center for Automated Learning and Discovery (now Machine Learning Department), Computer Science Department, Language Technologies Institute and Robotics Institute

Scientific Employment

Graduate Student Member (01/2005-Present)

Computational Intelligence and Machine Learning Group
Department of Computer Science, University of Pisa, Italy
Development of the software framework Structlab for Machine Learning and Applied Statistics on structured domains.
Supervisors: Alessio Micheli and Antonina Starita
Web reference: *http://structlab.sourceforge.net* and *http://ciml.di.unipi.it*

Software Engineer (05/2006-09/2006)

Department of Statistics and Applied Mathematics
University of Pisa, Italy
Implementation of a simplex based algorithm for solving subclasses of nonlinear problems through sensitivity analysis procedures
Supervisors: Laura Carosi and Laura Martein

Programmer Analyst (06/2004-10/2004)

Alberta Ingenuity Center for Machine Learning
Edmonton, Alberta, Canada
Development and implementation of a conditional random field model, using a min-cut algorithm for inference and a conjugate-gradient approach for parameter training for image segmentation and tumor growth prediction.
Supervisor: Russel Greiner
Web reference:
http://kingman.cs.ualberta.ca/research/projects/content/projects_template.php?num=11

Intern (02/2004-06/2004)

Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
Space invariant feature extraction from 3D brain volumes, aligned through midsagittal plane extraction, using and extending the Insight Toolkit for Image Segmentation and Registration.
Supervisors: Yanxi Liu and Leonid Teverovskiy

Research activities and interests

My main research interest is machine learning, in particular the design and analysis of systems that automatically adapt based on experience using mathematical models. During my studies and research experiences I had the opportunity to work on a sufficiently wide spectrum of methods, with particular emphasis on statistical models for time sequences and hierarchically structured data, such as Hidden Markov Models, Conditional Random Fields and Hidden Recursive Networks.

My principal area of research has been the integration of these models with Kernel Methods, through the study and development of Generative Kernel Functions, which try to combine the modeling ability of generative models with the good predictive performances of discriminative approaches.

I am mainly interested in applications on real-world structured domains, with some experience in Image Analysis, Bioinformatics and Cheminformatics.

Software Systems of Scientific Relevance

Structlab I developed and I am now extending a machine learning and applied statistics software library for structured domains, which aims to be an easy to extend framework for learning experiments with structured data, and provides a toolbox of generative and discriminative learning methods, together with tools for loading, preprocessing, cross validating, and visualization. Structlab is accompanied by a graphical user interface which allows to setup, in a visual and intuitive way, elaborate machine learning experiments.
Web reference: http://structlab.sourceforge.net

Publications

[1] Nicotra, L., Micheli, A., Starita, A. (2004), *Fisher Kernel for Tree Structured Data*, Proceedings of the IEEE International Joint Conference of Neural Networks, 1917-1922

[2] Nicotra, L., Micheli, A., Starita, A. (2007), *Generative Kernels for Gene Function Prediction through Phylogenetic Tree Models of Evolution* (submitted)

Master Thesis

Generative Kernel Functions for Structured Data (2006)

This thesis explores ways of combining probabilistic models and kernel methods. A class of *generative kernel functions* is presented defining embeddings of the input domain based on probabilistic models of the data generating process and then combining these models in order to define a similarity measure on the domain. Among the presented classes of kernels, some are extensions of previously defined approaches to more structured domains, while some other are completely new formulations, in particular the class of *relative probability kernels*. The performances of generative kernels are tested on various benchmarks, comprising a set of simulated data, a classification problem of biological sequences and two domains of molecules modeled as trees: a QSPR (Quantitative Structure Property Relationship) analysis problem on a class of alkanes and a QSAR (Quantitative Structure Activity Relationship) analysis problem on a benzodiazepines class.

Selected Coursework

Mathematics and Statistics
Mathematical Analysis, Algebra, Languages and Methods of Mathematics, Numerical Calculus, Computational Mathematics, Physical Modeling, Probability and Statistics, Operations Research

Machine Learning Statistical Approaches for Learning and Discovery (CALD at CMU), Machine Learning, Machine Learning Theory (CALD at CMU), Information Theory, Neural Networks I, Neural Networks II, Data Mining Techniques, Intelligent Systems I, Intelligent Systems II,

Applied Machine Learning
Computational Genomics and System Biology (CALD at CMU), Bioinformatics, Natural Language Processing, Learning to Turn Words into Data: Information Extraction & Integration (CMU), Methods in Medical Image Analysis (Robotics Institute at CMU)

Computer Skills

Programming C++ (2000-2006, expert), R/Splus (2004-2006, intermediate), Java (2001-2005, intermediate), Matlab (2003-2005, beginner), C#/.NET (2006, beginner), Perl (2004, beginner), Fortran (2004, beginner), C (2002-2003, beginner), Python (2006, beginner), Sql (2003, beginner), Ocaml (2003-2004, beginner)

Operating systems
Linux/Unix (2000-2006, expert), Microsoft Windows (1999-2006, intermediate)

Software Toolkits Root (Data analysis, C++), ITK (Image Segmentation and Registration, C++), Torch (Machine Learning Library, C++), Libsvm (Support vector machine, C), Clp (Linear programming solver, C++), Bayesian Network Toolkit (Graphical models, Matlab), Intel Probabilistic Networks Library (graphical models, C++), Weka (Machine learning and data mining, Java), Boost (general purpose libraries, C++)

Programming Tools
Eclipse, Emacs and Microsoft Visual Studio (IDE), Dart (testing tool), Memproof and Valgrind (memory profilers), Cvs and Subversion (versioning tools)

Scientific Formatting Tools and Languages
LaTeX/Postscript, Lyx, Gnuplot

Graphical User Interfaces Programming
wxWidgets (2006, beginner) and Gtk (2006, beginner)

Web Programming
Experience in delivery of content management based websites for many organizations (especially Drupal based), HTML

Non-technical Skills

Statistical Data Analysis and Visualization,
Numerical Optimization and Scientific Writing

Natural Languages

Italian, mother language

English fluent

German intermediate

Latin basic

Academic Honors

Graduation (M.S.) with Honors in Computer Science
University of Pisa, 2006

Graduation (B.S.) with Honors in Computer Science
University of Pisa, 2005

Grants and Scholarships

Internationalization of University Program Scholarship
Italian Ministry of Education, University and Research, 2004

Studentship and Full tuition support (2001-2006)
University of Pisa

Membership in Scientific Associations

IEEE Computer Society
since 2002

Associazione Italiana Intelligenza Artificiale
since 2002

Association for Computing Machinery (ACM)
since 2003

IEEE Computational Intelligence Society
since 2004

IEEE Information Theory Society
since 2006

American Association for Artificial Intelligence (AAAI)
since 2004

American Statistical Association (ASA)
since 2006

References

Available upon request.