Honors thesis topics – Masters in Computer Science, academic year 2016-2017

Tiziana Calamoneri

Topic: Efficient Algorithms for Biology

Many biological problems can be solved only with enumerative techniques, but standard ones are too inefficient. It is hence necessary to use the Branch and Bound techniques, allowing people to solve these problems in practice. Studies in this context consist in deepening the known algorithmic approaches for biological problems such as DNA Restriction Mapping or Regulatory Motif Finding.

Topic: Graph Algorithms to solve Problems in Biology

Some biological problems (such as sequencing) can be solved by modeling them as graph problems. Studies in this context consist in deepening the known models and solutions.

Topic: Computational Geometry Algorithms

Computational geometry is the branch of Algorithmics dealing with efficient solution of geometric problems. Typically, computational geometry algorithms require sophisticated data structures. Studies in this context consist in studying some notable examples of computational geometry algorithms with a particular attention to their computational complexity.

--

Lorenzo Carlucci

Topic: Effective content and logical strength of Hindman's Finite Sums Theorem.
Type: BSc/MSc

The following theorem is known as Hindman's Finite Sums Theorem: if the positive integers are colored in two colors then there exists an infinite set such that all finite non-empty sums of distinct elements chosen from it have the same color. The effective content and the logical strength of this important theorem of Combinatorics are still a mystery thirty years after the first investigations. It is known that there exist computable colorings of the integers such that every solution to Hindman's Theorem computes the Halting Set, i.e. the first Turing Jump. On the other hand the only known upper bound is that every computable coloring admits a solution to Hindman's Theorem that is computable in the first infinite Turing Jump, i.e., the degree of unsolvability of first-order arithmetical truth.

In Reverse Mathematics terms, Hindman's Theorem is provable in the system ACA0+ and implies the system ACA0 (ACA stands for Arithmetical Comprehension Axiom). Thus there is an "infinite" gap between the upper and the lower bound.
Recent results have shown that the lower bound is already valid for the restricted version of Hindman's Theorem where we only require that there exists an infinite set such that all its non-empty sums of at most three elements have the same color.

No upper bound other than the upper bound on the unrestricted version of the theorem is known to hold for the latter restricted version. It is an open question in Combinatorics whether Hindman's Theorem restricted to sums of at most two elements can be proved without proving the full theorem.

The present project consists in a guided study of the fundamental techniques to attack open problems in this area at the intersection of infinite Combinatorics, Reverse Mathematics and Computability Theory.

Area: Mathematical Logic, Computability Theory.

Topic: **Cognitively efficient learning strategies.**

Type: BSc and MSc.

Gold-style Computational Learning Theory (a.k.a. Theory of Inductive Inference) allows to model real-world learning strategies studied in Cognitive Science and Experimental Psychology and to answer questions concerning the necessity of such strategies for successfully completing some learning task. An important example is U-shaped learning. It has been observed that in a wide variety of key learning tasks performed by humans the learning curve is not monotonic but has a "U-shape": an early phase of good performance is followed by a phase of bad performance and the latter is in turn followed by a new phase of good performance. In contrast to other mathematical learning paradigms, Gold's model allows to prove rigorously that some appropriately modeled learning strategy - such as U-shaped learning - is redundant or instead necessary for full learning power. In the case of U-shaped learning it has been proved that this apparently unreasonable strategy is indeed necessary to achieve full learning power in many significant cases. Many lines of investigation and many problems in this area are still open. The proposed project consists in a guided study of fundamental results and techniques in Gold-style Learning Theory in order to attack existing open problems, and to model and study new cognitively relevant learning strategy.

Area: Learning Theory, Machine Learning, Computability Theory.

---

**Nicola Galesi**

Topic: **Convex Optimization and Complexity Theory**

The project introduces to the study of integer programming relaxation methods up to semidefinite program methods. The goal is to understand how these methods are related to main questions in Complexity Theory and specifically to the complexity of logical systems which uses proof methods related to Lovasz-Schrriver, Sherali-Adams and Lasserre hierarchies.

---
Luigi Vincenzo Mancini  
Availabilty: for 1 student  
Topic: **Blockchain Technologies**

Study the structure of blockchain and its applications, such as Virtual currencies (Bitcoin and derivatives) and smart contracts (e.g., Ethereum). Analysis of possible future extensions and new applications that can benefit from the use of blockchain.

Eligibility: students of the Laurea magistrale (master) in Computer science  
Availability: for 1 student

**Topic: Secure Machine Learning**

Study of machine learning techniques such as neural networks and deep learning, from the point of view of their security and privacy. Today, modern applications based on the use of machine learning techniques are applied also to sensitive data, such as: medical data, clinical, photos, audio and video recordings. How can you design neural networks privacy-preserving that are accurate and at the same time secure? How can you design neural networks that are resistant to external manipulation, and are able to preserve the privacy of the training data?

--

Toni Mancini  
Topic: **Artificial Intelligence for Personalised In Silico Medicine**

Designing a new pharmaceutical drug typically costs billions of Euros, and the entire path from research to market often takes many (>10) years, with several obstructions which often provoke failure of the entire process.

In particular, often a drug which performs well in laboratory ("in vitro", or on isolated tissues, i.e. "ex vivo") fails the subsequent testing phases ("clinical trial") on animals and human patients ("in vivo").

Such advanced drug testing phases may last several years and often are extremely expensive, as the involvements of animals and human patients requires great care about safety and security.

A late failure of the verification process for a new drug (for example, during the "in vivo" phase) implies a huge economic loss for the drug developers (for example, a pharmaceutical industry). This leads to the fact that pharma companies are often rather conservative in starting research activities on radically innovative drugs.

One of the most revolutionary research directions in medicine and pharmacology consists in defining and exploiting mathematical models of the human physiology (Virtual Physiological Human”, VPH) in order to design and verify safety and efficacy of new drugs and new treatment protocols "in silico" (i.e., by means of computer simulation), before starting expensive, risky and invasive "in vivo" clinical trials on animals and humans.

The main objective of a VPH model is to capture all biologically correct behaviours, that is all behaviours that could occur in nature. By combining clinical data on real patients (e.g., from
blood samples) and sophisticated computational techniques from Artificial Intelligence, such models can be individualised in order to make them able to simulate the behaviour of any given patient and his/her personal reaction to a set of drugs. The possibility to individualise VPH models also allows "in silico" design of individualised pharmacological therapies, i.e., therapies that maximise their performance on a certain given human patient at the same time minimising expected risk and severity of negative side-effects.

During the Honour Programme, students will learn methods based on Artificial Intelligence in the context of "in silico medicine" in order to design and verify safety and efficacy of individualised treatment protocols, exploiting complex VPH models.

For more information, please visit: 

---

**Federico Mari**
Topic: **Formal Methods for Embedded Systems Control**

Smartphones, smart ovens and fridges, automotive driving computers, space satellites components, mechanical arms for surgery aid. These are only few examples of embedded systems in modern society.

An embedded system is a mechanical or electronic system managed by a control software running on a microprocessor, with real-time computing constraints. It is estimated that 98% of microprocessors are used in embedded systems.

Embedded systems being devoted to specific operations on small hardware, one of the greatest challenges in this research field consists in designing control software with small dimensions and with high performance at the same time.

During his excellence path, the student will be driven through the use of Formal Methods for design and develop algorithms and tools aiming at automatically generating control software starting from mathematical models of different kinds of embedded systems.

---

**Gaia Maselli**
Topic: **Wireless & battery-less: a joystick example**

The student will be introduced to the network protocols and technologies useful to realize control devices that are both wireless and battery less, as for example a video game joystick.

---
Igor Melatti  
**Topic:** Algorithms and tools for Smart Cities and Smart Grids

One of the main challenges for our society's future is to exploit the ever increasing information about residential users electrical consumption, in order to cut electrical energy costs payed by utilities (such as Distributed System Operators and energy retailers). In turn, utilities will lower down residential users energy bills, in order to justify installation of the needed software (which is actually what we will develop in this honour programme) and hardware (batteries, sensors, dedicated microcomputers, etc; we will use existing hardware). In order to have residential users actually accepting the new setting, it is also necessary to guarantee their privacy (e.g., if in a smart home all appliances are able to communicate their energy consumption, then such data must be communicated to utilities only in an aggregated form). Finally, costs reduction must involve not only the economic saving, but also the environmental aspect, e.g., by provably reducing CO2 emissions. We also want to design and implement software which only works inside a single house, without communicating with the utility at all. Summing up, this honour programme will teach how to devise, design and effectively implement new algorithms (typically based on model checking techniques) achieving the above described goals.

--

**Roberto Navigli**  
**Topic:** Video games with a purpose and gamification

The student will study and develop techniques and algorithms for gamification integrated into video games with a purpose aimed at improving the quality of annotations, validations and information, data and knowledge gathering.

**Topic:** Neural networks for machine translation

The student will study, apply and improve the most advanced neural network architectures aimed at machine translation between arbitrary languages.

--

**Alessandro Panconesi**

**Availability:** Master Computer Science, for 1 student  
**Topic:** Probabilistic algorithms

Probabilistic algorithms are at the core of contemporary research in algorithms and computational complexity.

The student will be introduced to some of the basic techniques to design and analyse probabilistic algorithms and lower bounds, by studying some notable examples taken from the recent literature.

--

**Enrico Tronci**
**Topic: Model Based System Engineering**

Requirements validation and system verification are among the most critical and expensive steps in the design of Cyber-Physical systems, that is of systems comprising software, hardware and communication components.

The modern approach to V&V (Verification and Validation) rests on the use of mathematical models (Model Based System Engineering, MBSE) to define the dynamics of the system to be verified and on the use of "Hardware In the Loop Simulation" (HILS) to carry out V&V activities using a simulator running the system model.

The goal of the proposed activity is to expose the student to state-of-the-art methods and tools for MBSE for Cyber-Physical systems, both from a methodological as well as practical point of views.

--

**Paola Velardi**

**Topic: Systems for semantic recommendation**

Many internet businesses strongly base their success in the ability to engage users and provide them with personalized recommendations. The automated extraction of users’ interests from social data is therefore becoming an essential part of today’s web services. State of the art recommenders, however, face with at least one of the following problems: sparsity, overspecialization and domain-dependency. In addition to that, preferences for specific items can be volatile, thus requiring frequent updates or mechanisms to cope with interest drifts.

We propose the student to analyze the problem of designing a semantic recommender, based on an unsupervised recommending strategy based on mapping users’ interests onto a number of Wikipedia documents and then inducing a taxonomy of his interests. User’s interests are either extracted from a user’s profile or from his/her messages, or induced from his actions and social relations, or a mixture of all these strategies. While the strategies to induce ”primitive” preferences depend on the information available in specific domains, mapping preferences to Wikipedia articles and building the taxonomy are domain-independent steps.

Students may work on specific aspects of this project: among which, named entity extraction from user messages and Twitter lists; mapping of celebrity Twitter accounts to Wikipedia; extraction of users “semantic fingerprint” from Wikipedia categories; creation of learning and test set and the design of performance evaluation strategies for the recommender, and more.

--

**Daniele Venturi**

**Title: Non-malleable codes**

Description: Non-malleable codes encode a given message in such a way that mauling attempts with a codeword (within a certain class of allowed tampering functions) have the effect that decoding a modified codeword yields either the original message or a completely unrelated value. Such codes are interesting on their own right, but also have several applications to cryptography.
During this honor program the student will try to tackle open research questions in this context, e.g. studying the relationship between different flavors of non-malleable codes, constructing new codes for larger tampering families, and exploring new cryptographic applications.

References:
http://eprint.iacr.org/2014/173
http://eprint.iacr.org/2013/702

Title: **Leakage and Tamper Resilient Cryptography**

Description: The security of modern cryptographic algorithms is typically analyzed under the assumption that an adversary has neither partial knowledge nor she can modify the underlying secrets. Unfortunately, several realistic attacks (so called leakage and tampering attacks) do not obey this assumption, which creates a gap between theoretical cryptography and the real world.

During this honor program, the student will design new cryptographic primitives with provable guarantees against leakage and tampering attacks.

References:
http://eprint.iacr.org/2015/517
http://eprint.iacr.org/2016/529