Combinatorial optimization and computational biology

INRA – CR2 position

Statitics and Algorithms for Biology team

Context: the contribution of algorithmics and combinatorial optimization to the analysis of milecular biology data has increased tremendously over the last ten years. The increase in the sizes of sequencing or genotyping data has lead to the development of dedicated software tools based on increasingly sophisticated data structures and algorithms, but able to face amount of data that were not thinkable few years before. In order to develop such tools, one must cross sharp knowledge of algorithms and optimization together with a good knowledge of biology, developped through interaction with biologists.

For more than ten years, the SaAB team strives to maintain itself on the research front of algorithmic development in combinatorial optimization using weighted constraint networks while providing biologists with software tools able to directly assist them in solving generic biological data analysis problems, with increasingly large amounts of data. This ambition has abre his frits and we benefit from an excellent international image in the constraint programming community as well as on the biological side where INRA andd international colleagues use different peices of software to handle their increasingly huge data sets in genomic and genetic.

On the computer science side, our team has provided the community with a new family of lower bounds for weighted constraint satsifaction and satisfiability which has also attracted the attention of people working in the area of stochastic graphical models. This family of lower bounds has been implemented in our homemade optimization software, toulbar2, has been extended to handle large domains in order to tackle problems with millions of values such as they appear when domain represent sequence positions (eg. in known RNA gene localization problem). We have also combined this family of lower bounds with dedicated algorithms exploiting problem structures, allowing toulbar2 to anlayze very large and complex animal pedigrees (leading to a software package called MendelSoft).

More indirectly, a large expertise in combinatorial optimization allows to inject and tune existing technology in the bioinformatics area to efficiently solve generic problems. This has been the case with genetic and radiated hybrid mapping, in the CarthaGene software (which exploits a very fast implementation of the Lin-Kernighan heuristics for the TSP) or for gene prediction with the EuGene package (which incorporates both dedicated shortest path with constraints algorithms and stochastic optimization algorithms for parameter estimation). These software are used routinely by geneticists or bioinformatics colleagues.

The arrival of very high throuput data acquisition (Next generation sequencing) raises new questions and new challenges. We are looking for a talented and dynamic scientist that is interested in improving the combinatorial optimization technology as well as identifying generic bioinformtics problems where it can be used and put in practice.

The successful candidate : with a PhD in computer science, operations reserach or applied mathematics, the candidate has a good background knowledge in discrete algorithms and combinatorial optimization with a more in depth specialization in one area of combinatorial optimization (constraint programming, linear programming...), showing ability to have a sustainable scientific activity in the area. He/she shows a strong interest in implemented algorithms that can be experimented with and applied to different sort of problems. He or she is rady to interact with biologists, for modelling new problems and offering generic answers through dedicated software. Beyond combinatorial optimization, knowledge in combinatorial pattern matching, stochastics graphical models and a background in biology or bioinformatics are welcome.

Beyond these, the capacity to develop software, with a good autonomy and efficiency, in interaction with other computer scientists, mathematicians and biologists is important. A good level in English (written and spoken), the desire to rapidly be fluent in French are, ultimately, required..

Partners: On the computer science level, the local area offers several labs with whom the candidate will be able to interact (beyond the team level). In Constraint Programming, we already collaborate with different local labs (G. Verfaillie -CERT, H. Fargier, MC. Cooper, UPS). More widely, Toulouse has different labs working in Operations Research (LAAS CNRS, ENSEIHT, Paul Sabatier University...). The team also has collaborations with other teams in France (inside ANR – French NSF – funded projects), Spain, Italy , USA…

In terms of bioinformatics, the junior scientist will be able to benefit during his first years how the existing background knowledege in genetics and genomic (protein and RNA genes, interaction networks...) and of the collaboration network with molecular biologists, geneticists both inside INRA, France and beyond.