

An Overview of Haplotyping Via Perfect Phylogeny: Theory, Algorithms and Programs

Dan Gusfield
Department of Computer Science
University of California, Davis
Davis, CA 95616, USA

Abstract

The next high-priority phase of human genomics will involve the development of a full Haplotype Map of the human genome. It will be used in large-scale screens of populations to associate specific haplotypes with specific complex genetic-influenced diseases. A key, perhaps bottleneck, problem is to computationally determine haplotype pairs from genotype data. An approach to this problem based on viewing it in the context of perfect phylogeny was introduced in 2002 along with an efficient solution. A slower (in worst case) variation of that method was later implemented. Two simpler methods for the perfect phylogeny approach that are also slower (in worst case) than the first algorithm were later developed. We have implemented and tested all three of these approaches in order to compare and explain the practical efficiencies of the three methods.

In this talk I will introduce the haplotyping problem and the three algorithms for its solution that have been implemented. I will discuss two empirical observations: a strong phase-transition in the frequency of obtaining a unique solution as a function of the number of individuals in the input; and results of using the method to find non-overlapping intervals where the haplotyping solution is highly reliable, as a function of the level of recombination in the data. Finally, I will discuss the biological basis for the size of these tests.

Phase Transitions, Backbones and Heuristic Search

Weixiong Zhang
Computer Science Department
Washington University
St. Louis, MO 63130, USA

Abstract

A phase transition refers to such a phenomenon of a system (or combinatorial problem) in which some global properties change rapidly and dramatically when a control parameter crosses a critical value. A simple example of phase transition is water changing phases from solid (ice) to liquid (water) to gas (steam) as the temperature increases. The backbone of a problem is the fraction of variables that have fixed values among all solutions, which reflects the constrainedness of the problem. The concepts of phase transitions and backbones can be used to characterize typical-case properties of combinatorial problems. In this talk, I will give an overview of the research on understanding combinatorial optimization problems using phase transitions and backbones and on developing new heuristic search methods that exploit such intrinsic properties of difficult problems, including the Traveling Salesman Problem and Boolean satisfiability. I will emphasize my own contributions to these active research topics in the talk.

Artificial Intelligence in Financial Applications

Benjamin W. Wah
Department of Electrical and Computer Engineering
and the Coordinated Science Laboratory
University of Illinois, Urbana-Champaign
Urbana, IL 61801, USA

Abstract

In this presentation we discuss the role of artificial intelligence in time-series predictions, including those of market-trend data and stock quotes. The time series in these applications is characterized by stationary and non-stationary data that may depend on non-quantifiable non-numeric measures, as well as recent information that may be significant in predicting near-term trends. A good predictor should involve not only predictions of non-stationary and stationary data but also the abstraction and integration of non-numeric information. After surveying various techniques for trend predictions and intelligent mining of financial data, we discuss the use of intelligent agents for the abstraction of non-numeric information, the decomposition of non-stationary time series into multiple stationary time series, and the prediction of trends using artificial neural networks. Finally, we show some interesting prediction results.

Kernel Methods for Pattern Analysis

Nello Cristianini
Statistics Department
University of California
Davis, CA, USA

Abstract

Kernel-based learning methods are a class of algorithms for pattern analysis, whose best known element is the Support Vector Machine, based on the idea of replacing inner products by Mercer Kernels. This substitution enables them to efficiently operate in very high-dimensional spaces, implementing the following simple idea: first embed the data into some vector space, and then detect linear relations in such a space. If the kernel is suitably chosen, points that are “similar” for the task at hand, will be mapped to nearby positions, and “different” points will be far apart.

The extraordinary impact that this approach is having in the field of machine learning is due to many reasons. Firstly because they often lead to convex optimization problems in cases where neural networks have been used in the past. Secondly because they establish a connection with the theory of Reproducing Kernel Hilbert Spaces, and finally because kernels have been recently developed to deal with data as diverse as sequences, graphs, images, text documents. And kernel-based algorithms have been developed to deal with problems as diverse as classification, novelty detection, ranking, regression, canonical correlation analysis, etc.

Results from Statistical Learning Theory are routinely used in the design of such algorithms, and applications are most commonly found in bioinformatics and text analysis. In this talk we will review the main ideas as well as the most recent research directions.