Chapter I

From Evolution to Immune to Swarm to ...? A Simple Introduction to Modern Heuristics

Hussein A. Abbass
University of New South Wales, Australia

The definition of heuristic search has evolved over the last two decades. With the continuous success of modern heuristics in solving many combinatorial problems, it is imperative to scrutinize the success of these methods applied to data mining. This book provides a repository for the applications of heuristics to data mining. In this chapter, however, we present a textbook-like simple introduction to heuristics. It is apparent that the limited space of this chapter will not be enough to elucidate each of the discussed techniques. Notwithstanding, our emphasis will be conceptual. We will familiarize the reader with the different heuristics effortlessly, together with a list of references that should allow the researcher to find his/her own way in this large area of research. The heuristics that will be covered in this chapter are simulated annealing (SA), tabu search (TS), genetic algorithms (GA), immune systems (IS), and ant colony optimization (ACO).
INTRODUCTION

Problem solving is the core of many disciplines. To solve a problem properly, we need first to represent it. Problem representation is a critical step in problem solving as it can help in finding good solutions quickly and it can make it almost impossible not to find a solution at all.

In practice, there are many different ways to represent a problem. For example, operations research (OR) is a field that represents a problem quantitatively. In artificial intelligence (AI), a problem is usually represented by a graph, whether this graph is a network, tree, or any other graph representation. In computer science and engineering, tools such as system charts are used to assist in the problem representation. In general, deciding on an appropriate representation of a problem influences the choice of the appropriate approach to solve it. Therefore, we need somehow to choose the problem solving approach before representing the problem. However, it is often difficult to decide on the problem solving approach before completing the representation. For example, we may choose to represent a problem using an optimization model, then we find out that this is not suitable because there are some qualitative aspects that also need to be captured in our representation.

Once a problem is represented, the need arises for a search algorithm to explore the different alternatives (solutions) to solve the problem and to choose one or more good possible solutions. If there are no means of evaluating the solutions’ quality, we are usually just interested in finding any solution. If there is a criterion that we can use to differentiate between different solutions, we are usually interested in finding the best or optimal solution. Two types of optimality are generally distinguished: local and global. A local optimal solution is the best solution found within a region (neighborhood) of the search space, but not necessarily the best solution in the overall search space. A global optimal solution is the best solution in the overall search space.

To formally define these concepts, we need first to introduce one of the definitions of a neighborhood. A neighborhood $B_δ(x)$ in the search space $Θ(X)$ defined on $X ⊆ \mathbb{R}^n$ and centered on a solution $x$ is defined by the Euclidean distance $δ$; that is $B_δ(x) = \{x \in \mathbb{R}^n \mid ||x – x_0|| < δ, δ > 0\}$.

Now, we can define local and global optimality as follows:

**Definition 1: Local optimality** A solution $x \in θ(X)$ is said to be a local minimum of the problem iff $\exists δ > 0$ such that $f(x) ≤ f(x) \forall x \in (B_δ(x) \cap θ(X))$.

**Definition 2: Global optimality** A solution $x \in θ(X)$ is said to be a global minimum of the problem iff $\exists δ > 0$ such that $f(x) ≤ f(x) \forall x \in θ(X)$.

Finding a global optimal solution in most real-life applications is difficult. The number of alternatives that exist in the search space is usually enormous and cannot be searched in a reasonable amount of time. However, we are usually interested in good enough solutions—or what we will call from now on, satisfactory solutions. To search for a local, global, or satisfactory solution, we need to use a search mechanism.

Search is an important field of research, not only because it serves all
Semi-Automatic Design of Spatial Data Cubes from Simulation Model Results
Hadj Mahboubi, Sandro Bimonte, Guillaume Defuant, Jean-Pierre Chanet and François Pinet (2013). *International Journal of Data Warehousing and Mining* (pp. 70-95).
www.igi-global.com/article/semi-automatic-design-spatial-data/75616?camid=4v1a

RCUBE: Parallel Multi-Dimensional ROLAP Indexing
www.igi-global.com/article/rcube-parallel-multi-dimensional-rolap/1810?camid=4v1a