Combinatorial Optimization

Operations research

Operations Research (OR) is the application of scientific methods to arrive at the optimal solutions to the problems. It helps making better decisions. Further, the term operational analysis is used in the British military as an intrinsic part of capability development, management and assurance.
A **model** is a logical representation of a problem. It shows the relationships between the different variables in the problem.

Approach of Operations Research

1) Definition of the problem.
2) Construction of the model.
3) Solution of the model.
4) Validation of the model.
5) Implementation of the final results.

It is often considered to be a sub-field of applied mathematics. The terms management science and decision science are sometimes used as synonyms. Employing techniques from other mathematical sciences, such as mathematical modeling, statistical analysis, and mathematical optimization, operations research arrives at optimal or near-optimal solutions to complex decision-making problems.

The main goal: Operations research is often concerned with determining the extreme values of some real-world **objective: the maximum (of profit, performance, or yield) or minimum (of loss, risk, or cost).**

Operational research (OR) encompasses a wide range of problem-solving techniques and methods applied in the pursuit of improved decision-making and efficiency, such as simulation, mathematical optimization, queueing theory and other stochastic-process models, Markov decision processes, econometric methods, data envelopment analysis, neural networks, expert systems, and decision analysis. Nearly all of these techniques involve the construction of mathematical models that attempt to describe the system.

In the decades after the two world wars, the tools of operations research were more widely applied to problems in business, industry and society. Since that time, operational research has expanded into a field widely used in industries

ranging from petrochemicals to airlines, finance, logistics, and government, moving to a focus on the development of mathematical models that can be used to analyze and optimize complex systems, and has become an area of active academic and industrial research.

Module Information:

Module Outline

Loosely speaking, a combinatorial optimization problem is any optimization problem that has a **finite number of feasible solutions**. It is also referred to as Discrete Optimisation. In this module we shall be mainly concerned with (mixed) Integer Programming (**IP/MIP**).

Part 1: (Mixed) Integer Programming Models

• General IP/MIP problems
• Knapsack problem
• Assignment problem and Facility Location problem
• Set covering/packing/partitioning problem
• The travelling salesman (TSP) problem
• Binary Variables and Conjunctive Normal Form
• Modelling of nonlinear problems/logical conditions with binary variables
• Pre-processing approaches and model improvement

Part 2: Solution methods for IP/MIP

• LP relaxation and Convex Hull
• Special purpose algorithms and their applications
• Convex Hull and Linear Combinations
• The Branch-and-Bound (B&B) Method
• The Cutting Plane Algorithm
• Heuristics, Metaheuristics, and Nature-Inspired Algorithms
• Simulated Annealing, Tabu search Evolutionary Algorithms,
Genetic algorithm
- Representation
- Genetic Operators
- Stopping Criteria

• Application to Integer Linear Programming- The case of The travelling salesman (TSP) problem
• Matlab implementation

Recommended Reading:

• V. Chv`atal, Linear Programming. W.H. Freeman and Co., New York (1984).
• H. P. Williams, Model Building in Mathematical Programming (second edition). Wiley Interscience (1985).
• G.L. Nemhauser & L.A. Wolsey, Integer and Combinatorial Optimization. Wiley Interscience (1988).
• Wayne L. Winston, Operations Research: Applications and Algorithms (2004).
•Reeves, Colin, and Jonathan E. Rowe. *Genetic algorithms: principles and perspectives: a guide to GA theory*. Vol. 20. Springer Science & Business Media, 2002.
• Rothlauf, Franz, and Franz Rothlauf. *Representations for genetic and evolutionary algorithms*. Springer Berlin Heidelberg, 2006.
• Michalewicz, Zbigniew. *Genetic algorithms+ data structures= evolution programs*. Springer Science & Business Media, 2013.
• Lindfield, George, and John Penny. *Introduction to nature-inspired optimization*. Academic Press, 2017.
• Taha, Hamdy A. Operations research: an introduction. Pearson Education India, 2013.
• Arora, Rajesh Kumar. Optimization: algorithms and applications. CRC press, 2015.

الرجاء قبل ان نبدأ بالموضوع يحتاج الطالب مراجعة موضوع البرمجة الخطية لمعرفة كيفية بناء النموذج الخطي ومعرفة بعض طرق الحل كطريقة الحل التمثيل البياني.

(Mixed) Integer Programming Models

1.1 From LP to (M)ILP

هذا مثال لنموذج خطي للاعداد الصحيحة (والمختلطة ونقصد بها: المتغيرات اعداد صحيحة ومستمرة معا)

Example: A craftsman, Giapetto, is asked to make two types of wooden toys during next week: soldiers and trains. A soldier sells for 23 and uses 10 units of raw materials. A train sells for 21 and uses 9 units of raw materials. The manufacture of wooden soldiers and trains require certain working hours of Giapetto's labour which are 2 and 3 respectively. Next week, Giapetto can only obtain 200 units of raw material and a maximum of 50 labour hours. Assume that demand for soldiers is unlimited but at most 12 trains can be sold. Formulate a mathematical model to help Giapetto maximize his revenue for the week.

SOLUTION

Step 1: Definition of the Decision Variables نعرف متغيرات القرار

Let x_1 be the number of wooden soldiers produced next week;
Let x_2 be the number of wooden trains produced next week.

Step 2: Define the Constraints القيود **نعرف**

Raw material constraint:

$$10\,x_1 + 9x_2 \leq 200$$

Labour hour constraint:

$$2x_1 + 3x_2 \leq 50$$

Maximum demand for trains:

$$x_2 \leq 12$$

Boundary conditions:

$$x_1, x_2 \geq 0$$

Step 3: Define the Objective Function نعرف دالة الهدف

$$max \;\; 23x_1 + 21x_2$$

Let f (x_1, x_2) be Giapetto's revenue: Hence, the full optimization problem is
عليه مسألة التحسين تكتب كاملة كما يلي

$$\begin{aligned} max \;\; & 23x_1 + 21x_2 \\ s.t \;\; & 10x_1 + 9x_2 \leq 200 \quad (1) \\ & 2x_1 + 3x_2 \leq 50 \quad (2) \\ & x_2 \leq 12 \quad (3) \\ & x_1, x_2 \geq 0 \end{aligned}$$

We can easily solve this **graphically**. To find the intersection point which is optimal, we set
ممكن حل المسألة بطريقة ارسم البياني وذلك برسم الخط المستقيم لكل قيد وناخذ نقطة التقاطع للخطوط ك حل للمسالة. نجد ان الحل الامثل هو 435.83 لكن هذه لا يمكن لانه لا يمكن بيع 0.5 قطار او 0.33 جندي. لهذا سوف نقرب الحل ونحسب دالة الهدف

$$\begin{aligned} 10x_1 + 9x_2 &= 200 \\ 2x_1 + 3x_2 &= 50 \end{aligned}$$

and we see that the optimal occurs at

$(x_1^*, x_2^*) = (\frac{25}{2}, \frac{25}{3}) \approx (12.50, 8.33)$

$$f\,(x_1^*, x_2^*) = 23 \times \frac{25}{2} + 21 \times \frac{25}{3} \approx 435.83$$

Is it optimal? Can Giapetto sell 0.5 trains and 0.33 soldiers to customers? Actually to guarantee that Giapetto produces only whole toys, this problem should be an Integer Problem where both its variables are restricted to be positive integers:

$$x_1, x_2 \in Z^+ \cup \{0\}$$

This integral restriction excludes above solutions.

Now the question is **how to obtain optimal integer solutions**? Can we just round the continuous values up or down by one to decide the optimal integer solution? Let us check in turn.

$\underline{x}^{(1)} = (13, 9)$, constraint (1) yields 130+81=211 > 200, **infeasible**

$\underline{x}^{(2)} = (13, 8)$, constraint (1) yields 130+72=202 > 200, **infeasible**

$\underline{x}^{(3)} = (12, 9)$, constraint (1) yields 120+81=201 > 200, **infeasible**

اعلاه نقاط الحل غير مقبولة لانه القيود لا تتحقق . نجرب النقطة التالية حيث تحقق القيود, لكن هل هي احسن حل؟

$\underline{x}^{(4)} = (12, 8)$, $f(12, 8) = 23 \times 12 + 21 \times 8 = 444$

However, is it always true that the optimal integer solution cannot occur at any other points, e.g. $\underline{x}^{(5)} = (13, 7)$?

- constraint (1) yields: $130 + 63 = 193 < 200$, satisfied;
- constraint (2) yields: $26 + 21 = 47 < 50$, satisfied;
- constraint (3) yields: $x_2 = 7 < 12$, satisfied;
- $f(13, 7) = 23 \times 13 + 21 \times 7 = 446$

Actually, the optimal solution for this problem is $\underline{x}^{(6)} = (20, 0)$. It looks nothing like the continuous optimal solution $\underline{x}^*$.

> To prove optimality, we need to guarantee that there is no other integer combinations of x_1 and x_2, which gives a better objective value.

Generally speaking, solving a discrete optimization problem is much harder than solving its continuous analogy (نظيرتها). They are referred to as ***NP-Hard*** in the Jargon of Complexity Theory.

What we have solved by omitting all integer constraints on variables is widely known as the **LP relaxation** of the original IP, which (will be seen later) plays a key role in the solution of IPs.

However, in most case we cannot simply round the optimal solution for LP relaxation to the nearest integers to achieve the integer optimal solution.

1.2 Facility Location Problems

Example: A manufacturing firm is planning a new factory in a rural part of the country. A total of 100 employees are to be hired from the 5 surrounding villages. The following table shows the number of (equally) qualified workers available in each village and the village location coordinates (in miles).

	Village				
	1	2	3	4	5
Available workers	15	20	40	30	70
E-W coordinate	0	10	6	1	2
N-S coordinate	0	1	8	9	3

Because the company has promised to pay its employees their travel expenses, the planners want to choose a factory site that minimizes total employee travel distance. (Assume that travel distance is proportional to straight-line distance from village to factory site.) Formulate a mathematical model to help the company locate its factory.

SOLUTION

Step 1: Definition of the Decision Variables

Let x_1 be the EW-coordinate of the warehouse;
Let x_2 be the NS-coordinate of the warehouse;
Let y_i be the number of employees to be hired from village i, $where\ \ i = 1, 2, \ldots, 5$.

Step 2: Define the Constraints
A total of 100 workers are needed:

$$y_1 + y_2 + y_3 + y_4 + y_5 = 100\,.$$

Workers from every village cannot exceed its maximum available number:

$$y_1 \leq 15;\ y_2 \leq 20;\ y_3 \leq 40;\ y_4 \leq 30;\ y_5 \leq 70$$

$$y_1\,, y_2\,, y_3\,, y_4\,, y_5 \in Z^{+}\cup\{0\}$$

Step 3: Define the Objective Function (ناخذ المسافة من القرية الى المعمل)

$$f(x_1, x_2) \equiv y_1\sqrt{(x_1-0)^2+(x_2-0)^2} + y_2\sqrt{(x_1-10)^2+(x_2-1)^2} + y_3\sqrt{(x_1-6)^2+(x_2-8)^2} + y_4\sqrt{(x_1-1)^2+(x_2-9)^2} + y_5\sqrt{(x_1-2)^2+(x_2-3)^2}$$

This is a Nonlinear, Mixed-Integer Programming problem.

A more generalised (but slightly different) version of such problems, which is known as the **Facility Location Problem**, mainly concerns selecting from a discrete set of potential locations to locate a facility to serve clients economically.
مسألة تعيين مرفق حيوي هي مسألة اختيار المكان المناسب لبناء المرفق لخدمة اكبر عدد من الزبائن

- $N = 1, 2, \ldots, n$ be the set of potential locations;
- $I = 1, 2, \ldots, m$ the set of clients;
- c_j the cost of placing a facility at location j, $j \in N$;
- h_{ij} the cost of satisfying demand of client i, $i \in I$ from the facility at j.

The problem is to select a subset of locations at which to place facilities, and how to assign clients to these selected locations so as to minimize the cost of satisfying demands.

Uncapacitated Model
ونقصد به ان النموذج يحقق اي مطلب للزبائن دون تقيد

Step 1: Definition of the Decision Variables
نعرف متغير القرار أكس على انه 1 اذا بنينا المرفق في موقع معين و 0 اذا لم يبنى. ونعرف واي على انه طلب الزبون الذي حصل على طلبه من المرفق المبني في هذا الموقع

Let $x_j = \begin{cases} 1, & if\ a\ facility\ is\ placed\ at\ location\ j \\ 0, & otherwise \end{cases}$,$\forall\ j \in N.$

Let $y_{ij} \in [0, 1]$ be the fraction of demand of client *i* that is to be satisfied from facility at j.

Step 2: Define the Constraints
Demand at client *i* has to be satisfied (condition that each client's demand must be satisfied): نكتب القيد كما يلي اذا تحقق مطلب الزبون

$$\sum_{j \in N} y_{ij} = 1, \quad \forall i \in I$$

Client *i* cannot be served from facility at *j* unless a facility is placed there:
لا يتحقق مطلب الزبون اذا لم يبنى المرفق

$$y_{ij} \le x_j\ ,\ \ \forall i \in I,\ \ \forall j \in N.$$

Step 3: Define the Objective Function

دالة الهدف هي تصغير تكلفة المسألة وهي حاصل جمع تكلفة بناء الموقع وتكلفة مطلب الزبون

$$\min \quad \sum_{j\in N} c_j\, x_j + \sum_{i\in I}\sum_{j\in N} h_{ij} y_{ij}$$

Therefore, the entire optimization problem is written as

$$\begin{aligned} \min \quad & \sum_{j\in N} c_j\, x_j + \sum_{i\in I}\sum_{j\in N} h_{ij} y_{ij} \\ s.j. \quad & \sum_{j\in N} y_{ij} = 1,\ \ \forall i \in I \quad \text{(demand conditions)} \\ & y_{ij} \le x_j\ ,\ \ \forall i \in I,\ \ \forall j \in N\ \ (service\ conditions) \\ & x \in \{0,1\}^n \\ & y \in [0,1]^{m\times n} \end{aligned}$$

This is an MIP of the *uncapacitated* **Facility Location Problem**, as it assumes that the facility at j can satisfy any amount of demand from any number of clients. المتغير اكس عدد صحيح والمتغير واي متغير مستمر

Capacitated Model ونقصد به ان النموذج يحقق اي مطلب للزبائن لكن ضمن حد معين

Now suppose that the facility at location j has a maximum capacity μ_j and the demand from client i is d_i. Then the total demand to be satisfied by facility location j cannot exceed its capacity:

$$\sum_{i\in I} y_{ij} d_i \le \mu_j x_j\ ,\quad \forall j \in N$$

Replacing the service condition in uncapacitated model:

$$\begin{aligned} \min \quad & \sum_{j\in N} c_j\, x_j + \sum_{i\in I}\sum_{j\in N} h_{ij} y_{ij} \\ s.j. \quad & \sum_{j\in N} y_{ij} = 1,\ \ \forall i \in I \qquad \text{(demand conditions)} \\ & \sum_{i\in I} y_{ij} d_i \le \mu_j x_j\ ,\quad \forall j \in N\ \ (service\ conditions) \\ & x \in \{0,1\}^n \\ & y \in [0,1]^{m\times n} \end{aligned}$$

Remarks

• There are two kinds of decisions to make in a **Facility Location Problem**: which facilities to open; how to choose facilities to meet customer demands. Therefore, general **Facility Location Problem** is a mixed-integer (linear) programming problem.

• **Facility Location Problem** is closely related to network design and network flows.

• **The capacitated Facility Location model** has even less constraints than the uncapacitated one. However, this does not mean it is easier to solve, as *In the context of IP, it is sometimes beneficial of having fewer variables and/or more constraints*.

• Why we write here $\sum_{j \in N} y_{ij} = 1$ not $\sum_{j \in N} y_{ij} = D_j$? All fractional demands of customer *i* add up to 1. They are satisfied from different j's (الزبون ياخذ من كل مكان شيء). This is because we consider y_{ij} as fraction of D_j. But if y_{ij} is just what *i* gets from facility *j* then it is just a demand of *i* satisfied from facility *j* (الزبون ياخذ كل احتياجانه من مكان واحد). You can call it y_{ij} and the total demand of customer *i* is D_j. Hence the summation. Page 40 from applied integer programming book.