

Optimized Delivery Route using Genetic Algorithm: Cost cutting for e-commerce

Karthik

30 November 2015

Travelling Salesman Problem (TSP) is a well known and extensively studied NP-hard [1] combinatorial optimization problem. Even though its a NP-hard problem, there are many heuristic algorithms available which works well with few thousand cities. And a wide variety of research has been conducted to explore TSP's application in planning, logistics and manufacturing of microchips.

This article will focus its attention on the application of TSP in Logistics for e-commerce industry. We will first discuss briefly about the growth of e-commerce in recent years and the importance of having a really efficient logistics for cost cutting along with better customer experience. And then we would discuss about TSP, Genetic Algorithm (GA) and will see "how to find the optimal route for a Travelling Salesman Problem using GA in R". We are not really trying to re-invent the wheel here but would rather show through a classic example of solving TSP using GA to pave way for building a thought process towards making the delivering process in e-commerce industry more efficient and cost effective.

1. e-commerce Industry, its Growth and logistics

In this section, we would focus our attention on one of the fastest growing markets for e-commerce, India. If you see this growth, it could be attributed to, first, the uphill rise of mobile user base and second, the penetration of internet. The ease of purchasing things from the comfort of your home and so many variety, has simply lured customers towards the e-commerce portals. And to add onto this delight, the prices and discounts are unimaginable if one would decide to purchase from brick-n-mortar shops.

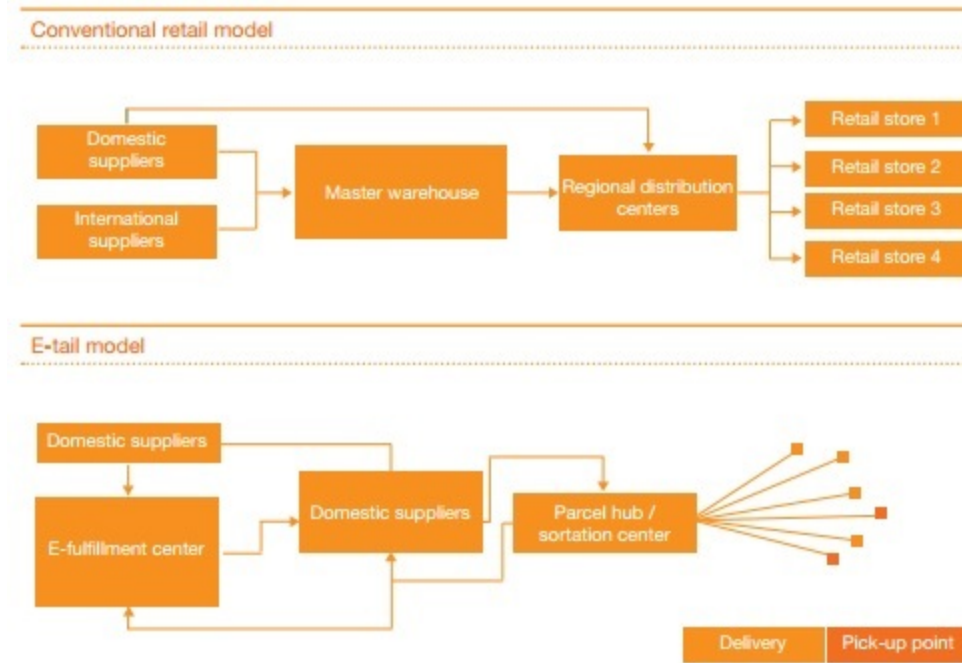
In one of the reports from AVENDUS, titled "India's mobile Internet :: The revolution has begun"⁴ (http://www.avendus.com/media/1366/avendus_india_mobile_internet.pdf), shows how the world has changed for consumers with internet and mobile usage. Below image from the report shows the trend very neatly,

Category	Old model	New model
News and information	Delayed reporters and cameramen Local / regional / national	Twitter, Reddit: Real-time, citizen reporting, global reach
Note taking	Pencil and notepad	Evernote: Always synced, multi-device, searchable
Files and folders	Print copies, storage cabinets and colour tabs	Dropbox, Google Drive, SugarSync: Always synced, always accessible, multi-device searchable digital files
Magazines	Printed copies	Flipboard and Currents: More content, always upto date, personalized, access everywhere, share
Education	Classrooms / Lectures / Reading materials	Khan academy, Coursera: Interactive, online and accessible by anyone anywhere anytime
Digital goods	Dedicated shopping sites / Fulfillment infrastructure	eBay, Amazon: Showrooming, shopping on the go

The report also provides some factual data showing how “India is going Mobile”. Below is the list of facts for your quick reference from the report.

- India has more than 160 million Internet users, of which 86 million access Internet using their mobile devices
- In the last 3-4 years, the number of users who access the Internet through a 3G connection has grown to round 22 million, To put things in perspective, compare this with the 15 million fixed line broadband connections accrued over the last 17 years
- There are over 36 million smartphone users as against c. 60 million PC users
- 9% of overall Internet pageviews in India come from mobile devices
- Over 40% of searches on Google originate from mobile device
- 30% of Facebook users in India are mobile-only Internet users and 30% of new registrations are coming through mobile
- LinkedIn ranks India among its top 4 growth markets for mobile usage

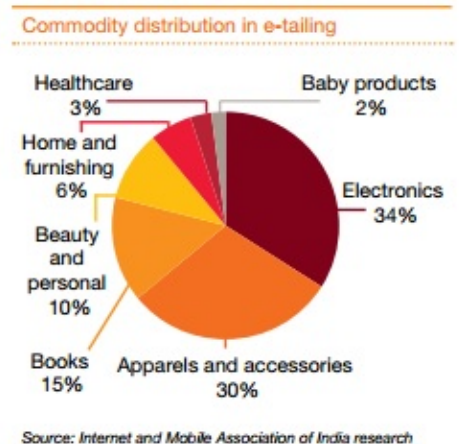
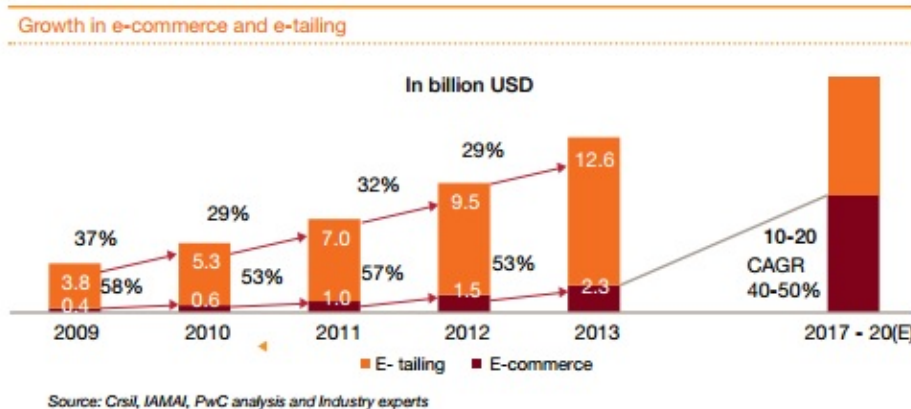
In another report, from PwC titled “Evolution of e-commerce in India : Creating the bricks behind the clicks”, which kind of takes a deep view on the e-commerce landscape in India and how its changing the convetional mode of business. Below illustration helps understand the difference between the conventional retail model and E-tail model



Further to quote from the report the below section, titled “E-commerce logistics models: A radical shift from regular logistics”,

The strong emergence of e-commerce will place an enormous pressure on the supporting logistics functions. The proposition of e-commerce to the customer is in offering an almost infinite variety of choices spread over an enormous geographical area. Firms cannot compete solely based on sheer volumes in today's ever-evolving, information symmetric and globalised world of e-commerce. Instead, the realm of competition has shifted to delivering to ever-shortening delivery timeliness, both consistently and predictably. Negligible or zero delivery prices, doorstep delivery, traceability solutions and convenient reverse logistics have become the most important elements of differentiation for providers. While the current logistics challenges relating to manufacturing and distribution of consumer products and organised retail are well-known, the demands of e-commerce raise the associated complexities to a different level. E-commerce retailers are well aware of these challenges and are cognizant of the need to invest in capital and operational assets.

And finally, the below picture shows how the e-commerce industry has been growing over the years and some projections for the future as well:



So, with this rapid growth, you can see how important the logistics planning could become for e-commerce players in order to be more efficient and cost effective.

2. Travelling Salesman Problem (TSP)

So, now with that overdose of e-commerce business model and its growth, now lets to get little more techincal and try to understand the Travelling Salesman Problem followed by Genetic Algorithm in the next section. These two ideas form the basis for showing you how we can build a simple model to plan the logistics.

We are not going to re-invent the wheel here, as the problem of TSP is already well defined and discussed. So, I will provide the below few lines to tell you what the TSP problem is all about and two useful links with the idea of TSP explained in detail.

The traveling salesman problem consists of a salesman and a set of cities. The salesman has to visit each one of the cities starting from a certain one (e.g. the hometown) and returning to the same city. The challenge of the problem is that the traveling salesman wants to minimize the total length of the trip.

1. Wiki - The best resource to learn anything !! (https://simple.wikipedia.org/wiki/Travelling_salesman_problem)
2. The Traveling Salesman Problem (<http://www.csd.uoc.gr/~hy583/papers/ch11.pdf>)

3. Genetic Algorithm

Genetic Algorithm is a powerful evolutionary strategies inspired by the basic principles of biological evolution.

The below flow chart from the paper GA: A Package for Genetic Algorithms in R (www.jstatsoft.org/article/view/v053i04/v53i04.pdf), explains the idea in a simple manner.

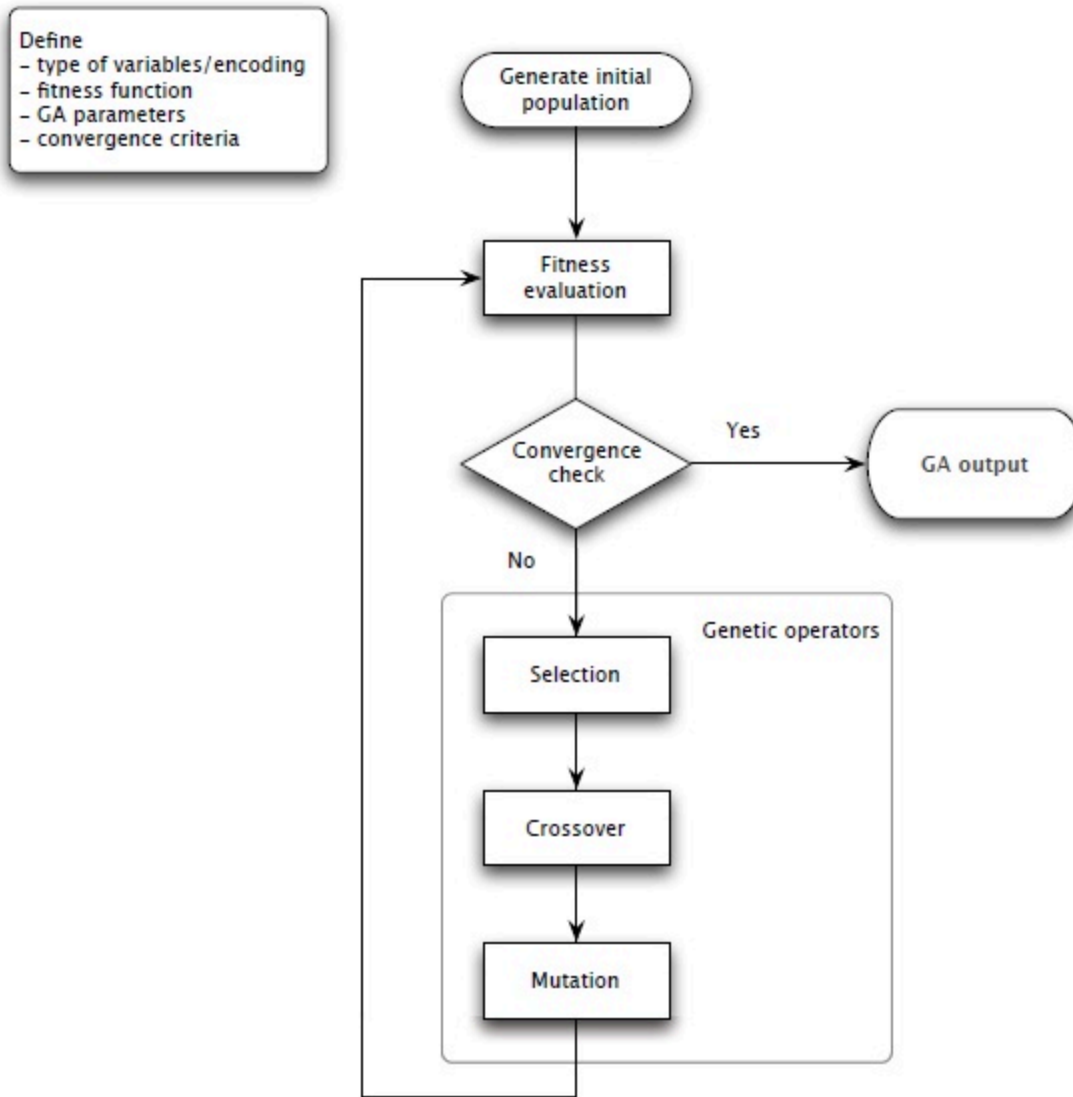


Figure 1: Flow-chart of a genetic algorithm.

The flow-chart of a typical genetic algorithm. A user must first define the type of variables and their encoding for the problem at hand. Then the fitness function is defined, which is often simply the objective function to be optimized. More generally, it can be any function which assigns a value of relative merit to an individual. Genetic operators, such as crossover and mutation, are applied stochastically at each step of the evolution process, so their probabilities of occurrence must be set. Finally, convergence criteria must be supplied.

With all these ideas, we are now ready to do some practicals.

4. Solving TSP using Genetic Algorithm in R

So far, we have seen how building an efficient logistics solution in e-commerce can bring about a huge cost benefit and we also discussed the theoretical aspects of TSP & GA. Now, it's time for some action, I have picked up an example from the paper "GA: A Package for Genetic Algorithms in R" [2]. Given a matrix with distances between every pair of source and destination, the below R code will produce the shortest path and subsequently put the results in a simple directed graph.

First, let's load the library 'GA' and a sample data called "eurodist", which gives distances between various cities in Europe (if you don't have this library, simply run `install.packages("GA")`)

```
library(GA)
data("eurodist", package = "datasets")
D <- as.matrix(eurodist)
```

Now, since our objective is to find the shortest path for our delivery boy, the fitness function should maximize the reciprocal of the tour length, which will in turn minimize the tour length. So, below two function are defined for the purpose.

```
#Function to calculate tour length

tourLength <- function(tour, distMatrix) {
  tour <- c(tour, tour[1])
  route <- embed(tour, 2)[,2:1]
  sum(distMatrix[route])
}

#Fitness function to be maximized

tspFitness <- function(tour, ...) 1/tourLength(tour, ...)
```

Finally, here comes are super hero function built in the “GA” package which actually computes the shortest path.

```
GA <- ga(type = "permutation", fitness = tspFitness, distMatrix = D,
        min = 1, max = attr(eurodist, "Size"), popSize = 50, maxiter = 5000,
        run = 500, pmutation = 0.2)
```

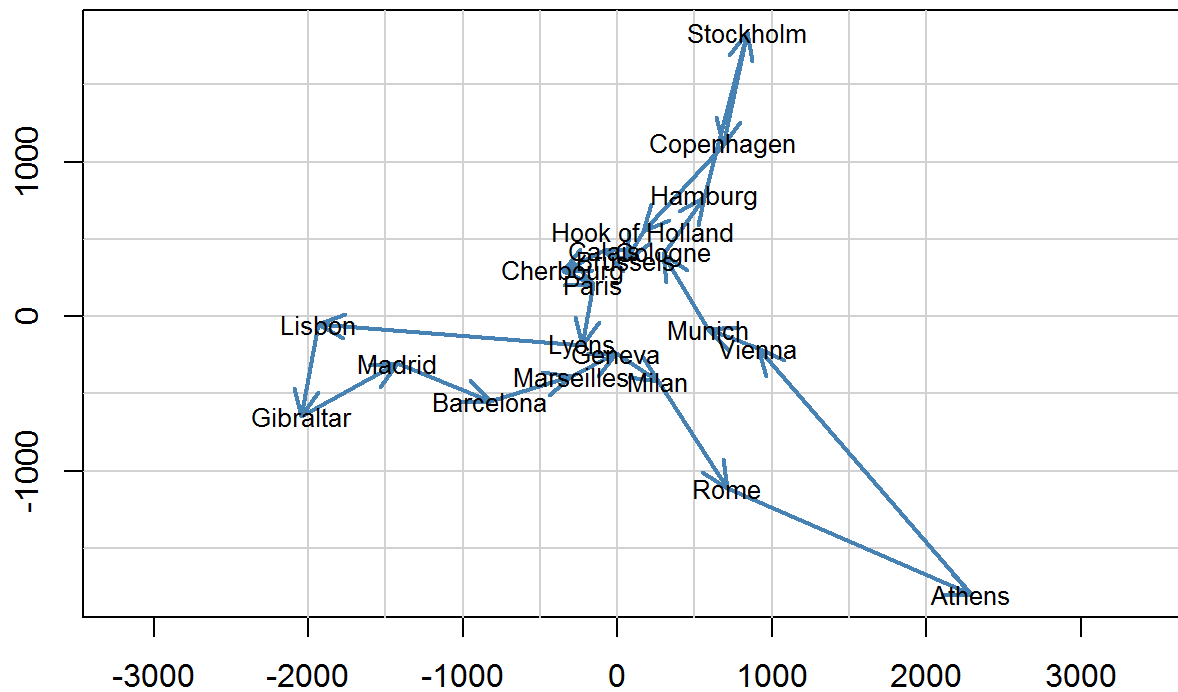
Lets look at the results and a visualization to see how the cities looks like in a graph,

```
summary(GA)
```

```
## +-----+
## |          Genetic Algorithm          |
## +-----+
##
## GA settings:
## Type                = permutation
## Population size      = 50
## Number of generations = 5000
## Elitism              = 2
## Crossover probability = 0.8
## Mutation probability = 0.2
##
## GA results:
## Iterations           = 856
## Fitness function value = 7.786949e-05
## Solutions            =
##      x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12 x13 x14 x15 x16 x17 x18 x19
## [1,] 20  7 11  3  4  5 18 13 12  9 14  2 15  8 16 19  1 21 17
## [2,]  6 10 20  7 11  3  4  5 18 13 12  9 14  2 15  8 16 19  1
## [3,] 10  6 17 21  1 19 16  8 15  2 14  9 12 13 18  5  4  3 11
##      x20 x21
## [1,]   6 10
## [2,]  21 17
## [3,]   7 20
```

#Visualization

```
mds <- cmdscale(eurodist)
x <- mds[, 1]
y <- -mds[, 2]
plot(x, y, type = "n", asp = 1, xlab = "", ylab = "")
abline(h = pretty(range(x), 10), v = pretty(range(y), 10),
       col = "light gray")
tour <- GA@solution[1, ]
tour <- c(tour, tour[1])
n <- length(tour)
arrows(x[tour[-n]], y[tour[-n]], x[tour[-1]], y[tour[-1]],
       length = 0.15, angle = 25, col = "steelblue", lwd = 2)
text(x, y, labels(eurodist), cex=0.8)
```



5. Few more factors to be considered in the Model

So, the simple model we have seen hasn't actually considered many factors which might play a big role in getting this example industry adaptable. There could be so many variables like :

- Number of orders per city and locality
- Number of available delivery person from a distribution center
- Size/Type of package to be distributed (This will decide whether you want a bike or car for delivery)

And many more ...

Finally, I will end this on a note saying that, in real world, it might be a case that, the simple TSP plus GA based model we have built will not serve our purpose at all but then our idea behind this post was to give a starting point which will generate the spark for the curious minds to explore further on these line. Hope the post was useful.