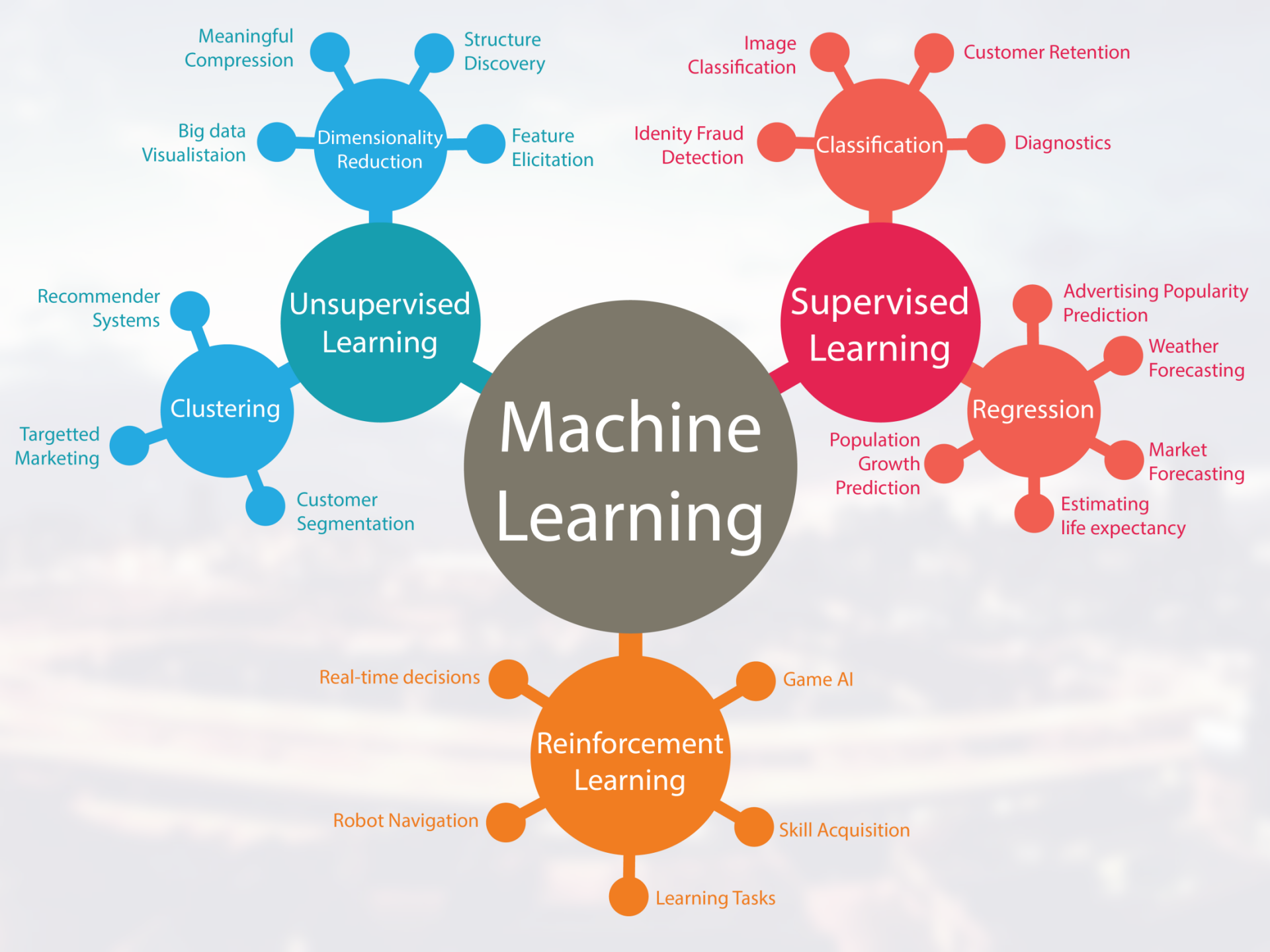

isazi
Advanced Algorithms and Optimisation
R Users Meetup
8 November 2016

43+ Languages and Tools
(98% Open Source)

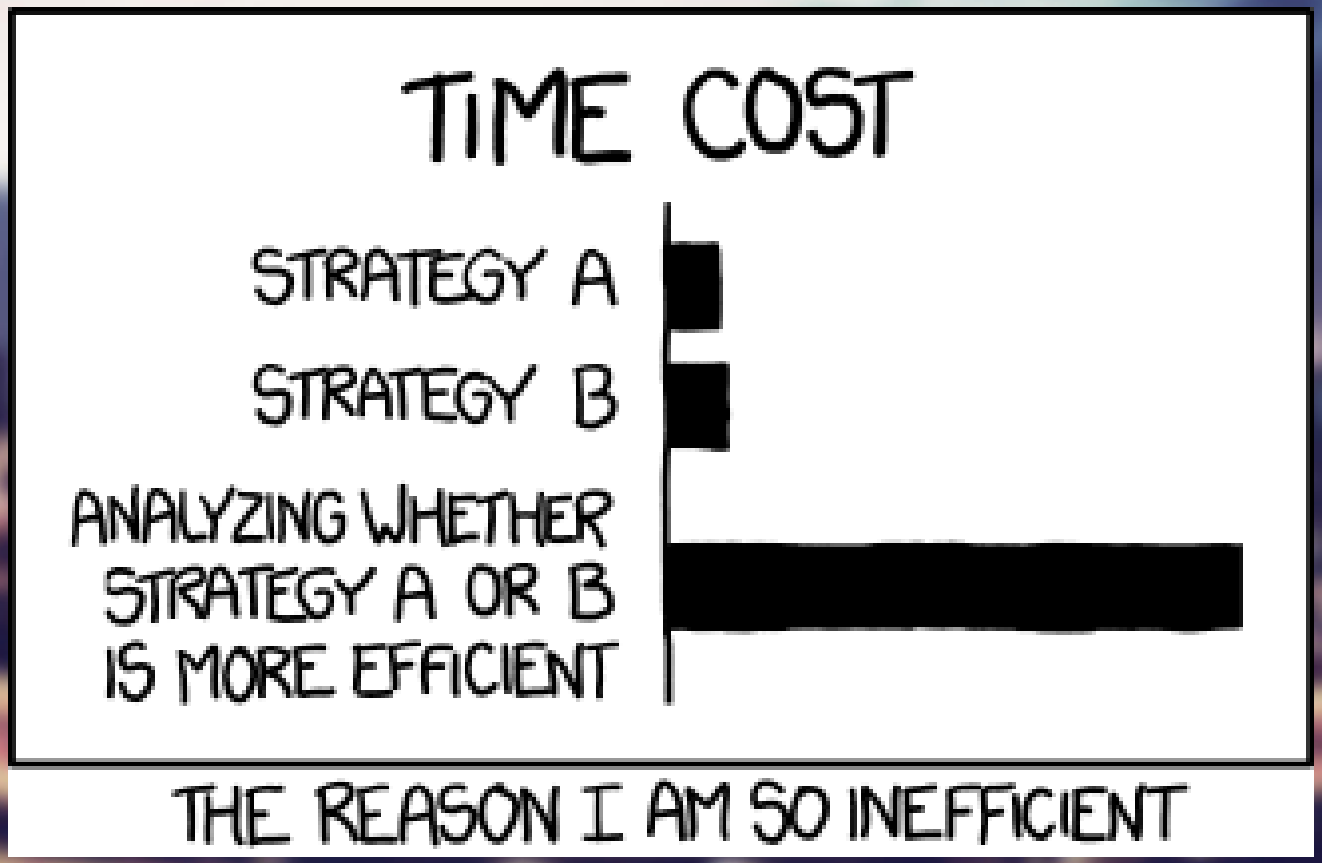

Machine Learning
Unsupervised Learning
Dimensionality Reduction
Meaningful Compression
Structure Discovery
Big data Visualistaion
Feature Elicitation
Clustering
Recommender Systems
Targetted Marketing
Customer Segmentation
Supervised Learning
Classification
Image Classification
Customer Retention
Idenity Fraud Detection
Diagnostics
Regression
Advertising Popularity Prediction
Weather Forecasting
Population Growth Prediction
Market Forecasting
Estimating life expectancy
Reinforcement Learning
Real-time decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

ClassMaster

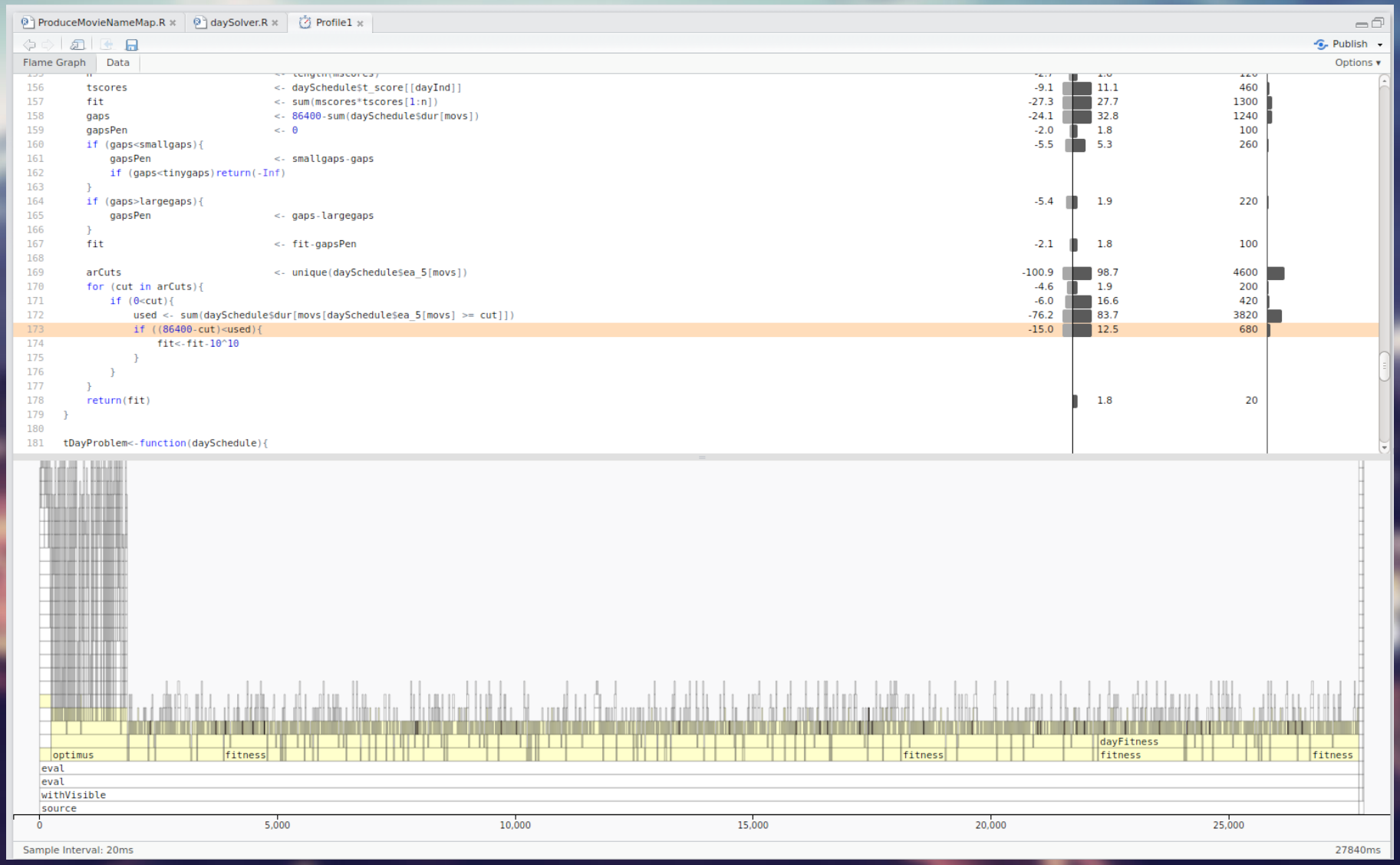
Precision Medicine

Telematics

Profiler


TIME COST
STRATEGY A
STRATEGY B
ANALYZING WHETHER STRATEGY A OR B IS MORE EFFICIENT
THE REASON I AM SO INEFFICIENT

Profiler


ProduceMovieNameMap.R
daySolver.R
Profile1
Publish
Flame Graph
Data
Options
tscores <- daySchedule$t_score[[dayInd]] -9.1 11.1 460
fit <- sum(mscores*tscores[1:n]) -27.3 27.7 1300
gaps <- 86400-sum(daySchedule$dur[movs]) -24.1 32.8 1240
gapsPen <- 0 -2.0 1.8 100
if (gaps<smallgaps){ -5.5 5.3 260
gapsPen <- smallgaps-gaps
if (gaps<tinygaps)return(-Inf)
}
if (gaps>largegaps){ -5.4 1.9 220
gapsPen <- gaps-largegaps
}
fit <- fit-gapsPen -2.1 1.8 100
arCuts <- unique(daySchedule$ea_5[movs]) -100.9 98.7 4600
for (cut in arCuts){ -4.6 1.9 200
if (0<cut){ -6.0 16.6 420
used <- sum(daySchedule$dur[movs[daySchedule$ea_5[movs] >= cut]]) -76.2 83.7 3820
if ((86400-cut)<used){ -15.0 12.5 680
fit<-fit-10^10
}
}
}
return(fit) 1.8 20
}
tDayProblem<-function(daySchedule){
optimus
fitness
fitness
dayFitness
fitness
fitness
eval
eval
withVisible
source
0
5,000
10,000
15,000
20,000
25,000
Sample Interval: 20ms
27840ms

R vs C++

VS

R

C++

RCPP

http://adv-r.had.co.nz/Rcpp.html

Non-linear Random Effects Model

- Objective = fit a model to data which describes a response which evolves with time since some event.
- Type of response = biomarker reading which is normally distributed around some signal which evolves with time from an event.
- Modelling = each subject has his/her own signal curve providing expected reading as a function of time since event.

Non-linear Random Effects Model

- MDUT : $\Omega_T = \int_0^T P_X(t)dt$
- $\Omega_T = \Omega(\beta, \Sigma, a, X, T) =$

$$\frac{1}{(2\pi)^{\wedge}\frac{q}{2}|\Sigma|^{0.5}} \int_{t\in[0,T]} \int_{b_i \in R^q} \Phi\left(\frac{X - f(t,\beta,b_i)}{a_0 + a_1|f(t,\beta,b_i)|}\right) \exp\left(-\frac{1}{2} b_i^T \Sigma^{\{--1\}} b_i\right) db_i dt$$

Where, Φ is the CDF for a std normal random variable, X is specified measurement threshold and T is a time cut-off.

Data & Model Inputs

- Individual Measurements in the form of data frame
  - Subject_ID (i)
  - Time of measurement (t_{ij})
  - Actual Measurement (x_{ij})
- Event Time distributions and parameter values
- Functional form of the signal curve $f(t, \beta, b_i)$
- Specification of nested model