Predicting task completion times

Maksim Gubin

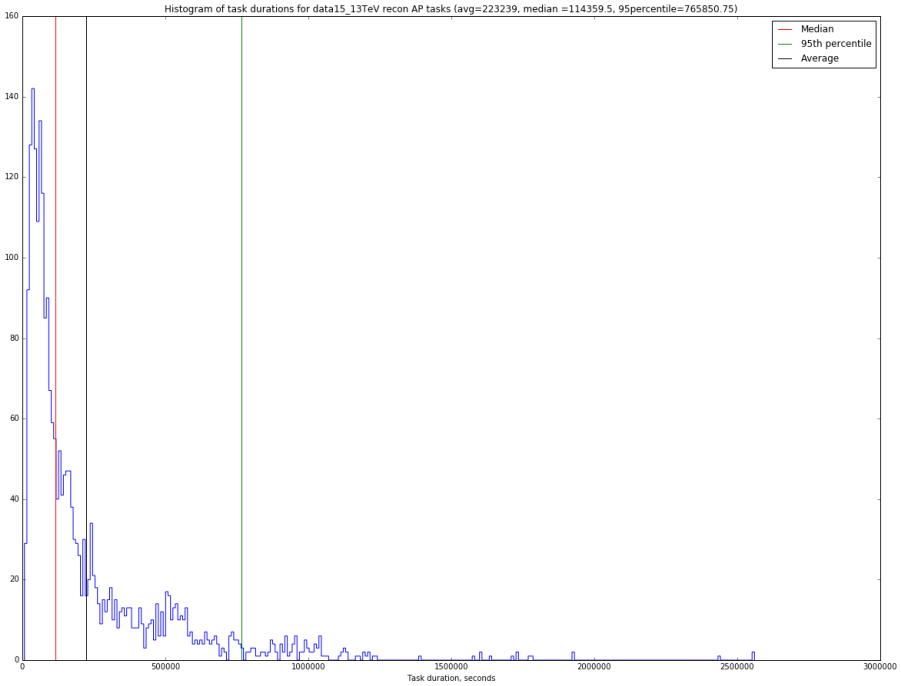
Current approach:

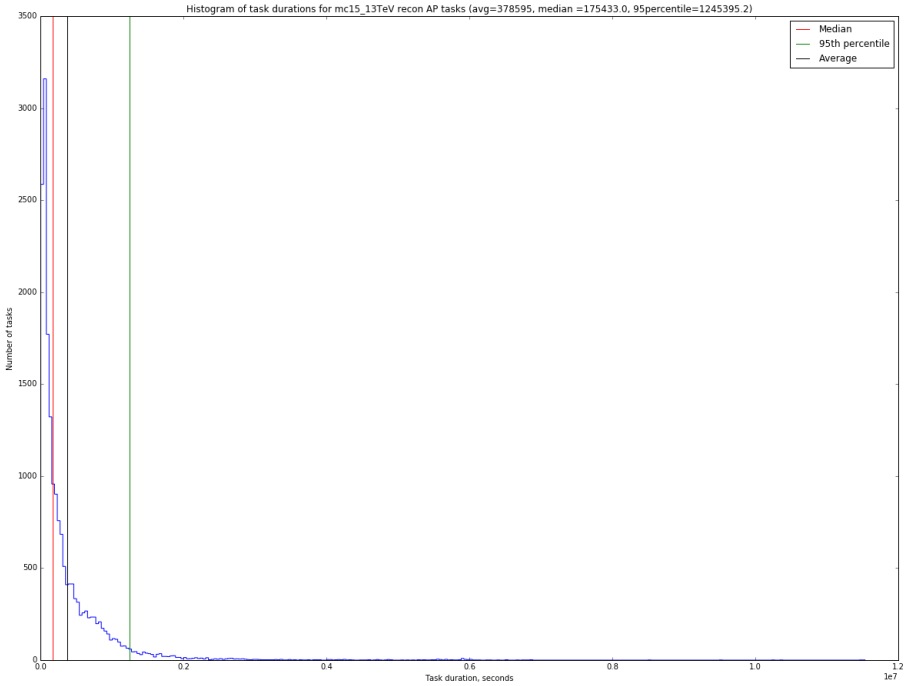
“Cold” predictions:

TTC = 95th percentile of historical task durations for a group of tasks.

Tasks are grouped by Project, Type and Provenance.

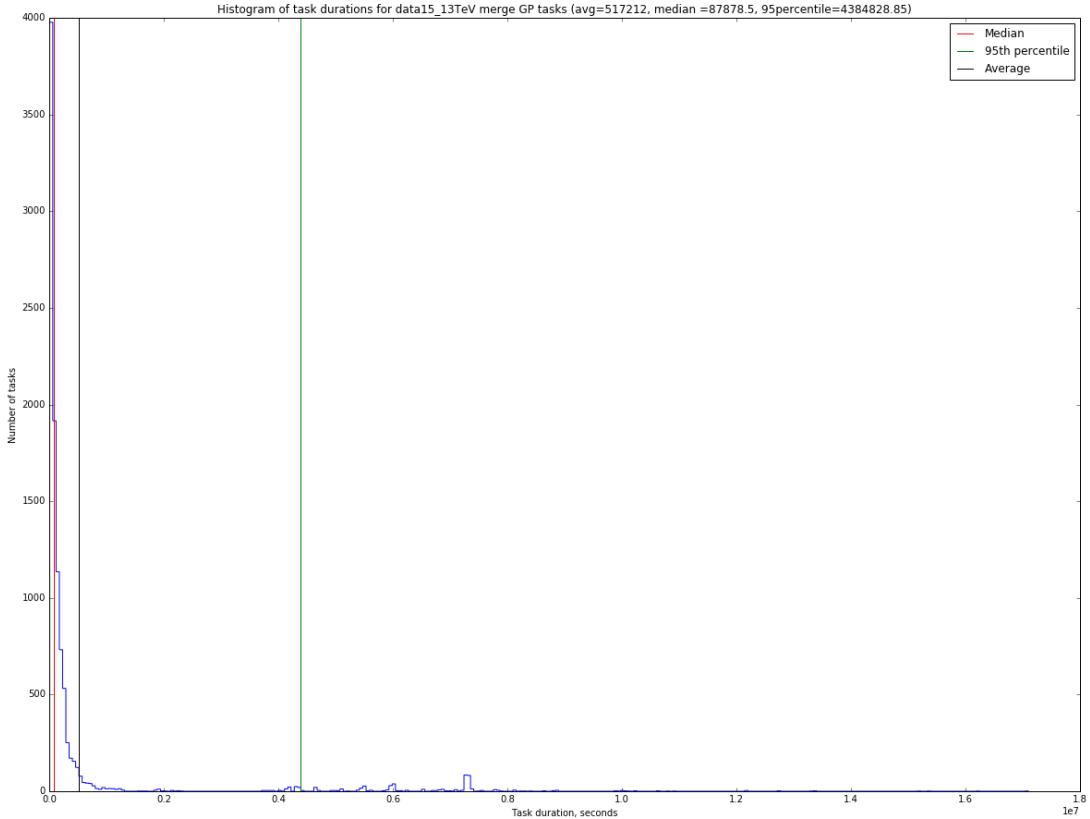
TTC prediction is already implemented; currently it is in testing phase before refining “cold” predictions.


Histogram of task durations for data15_13TeV recon AP tasks (avg=223239, median =114359.5, 95percentile=765850.75)
Median
95th percentile
Average
Task duration, seconds


Histogram of task durations for mc15_13TeV recon AP tasks (avg=378595, median =175433.0, 95percentile=1245395.2)
Median
95th percentile
Average
Number of tasks
Task duration, seconds
1e7

Task duration distribution

Project	Type	Provenance	TTC (days)
data15_13TeV	merge	GP	50.8
data15_13TeV	recon	GP	2.2
data15_13TeV	recon	AP	8.9
data15_13TeV	merge	AP	4.9
data15_5TeV	merge	GP	10.1
data15_5TeV	merge	AP	2.2
data15_5TeV	recon	AP	2



Proposal to improve TTC prediction (step 1)

Task profile: function $N_i(t)$ that returns number (or fraction) of jobs for a task group i that are completed at time t.

Task execution rate $R=N/N_i(t)$, where N = number of completed jobs of a task at time t.

Corrected time to completion, CTTC = TTC * R

Proposal to improve TTC prediction (step 2)

Use machine learning to define function TTC(T, G), where T is task parameters, and G is Grid state at the moment of task definition.

T should include Project, Type, Provenance, number of events, etc.

G should include number of running jobs, queue size, priority distribution for queued and running jobs.