

Final Quarter - 2017

The motivation is to treat productivity as an optimization problem and systematically solve it. I am trying to treat it the way industry treats revenue. Here are the factors I am considering for now:

- 1) Number of hours spent working - This is a linear thing.
- 2) Number of papers being read per month - This has the potential to grow exponentially.
- 3) Number of git commits being made - This has potential to grow non linearly.
- 4) Number of experiments being run (per project) - This should be more or less depending on the importance of the project.
- 5) I don't want to really track this, but number of submissions made to peer reviewed conferences. This might actually end up being the most important in the long run but I don't see a good way to measure this periodically for now.

There might be interesting ways to analyse these dimensions. For example, if the number of git commits are less for a project but the number of experiments are more, it is okay. But if both are low and the importance of the project is high, it means there is a scope for improvement. These may not directly correspond to productivity in the strict sense from the perspective of an academic researcher. However, I believe these account for the effort being put in. It is easy to become complacent or one dimensional while working on something for long time. These are to avoid the same.

1) Number of hours

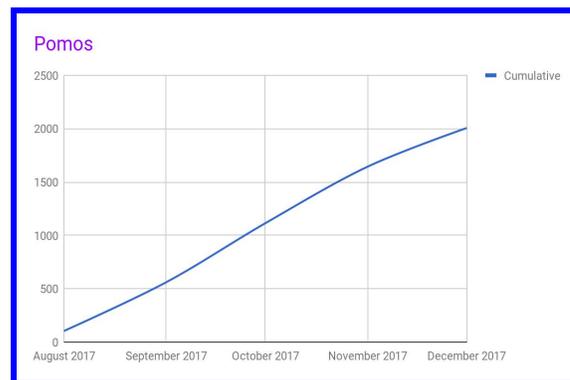


Figure 01: Plot showing the progression of pomos. One pomos equals 45 minutes of time. We would want to consistently beat 2000 every semester irrespective of the quarter.

According to the famous study by Malcom Gladwell, an individual needs to put 10000 hours of deliberate practice to be in the top 5% of a domain. Considering that an average PhD is about 5 years, this means one needs to put approximately 2000 hours per year. Ofcourse, this needs to be much more because I think a PhD needs more than just being a (happily) selfish 'learner'. From the discussions I have with career consultants, roughly 10% of the time while pursuing PhD should also be spent in 'teaching/ reaching out' as well. This means one needs to put a

non linear amount of effort per semester in *learning* - perhaps more in the early phases. I am tracking the amount of time I put using the 'pomodoro' technique and the progression so far can be seen in figure 01.

2) Number of papers being read

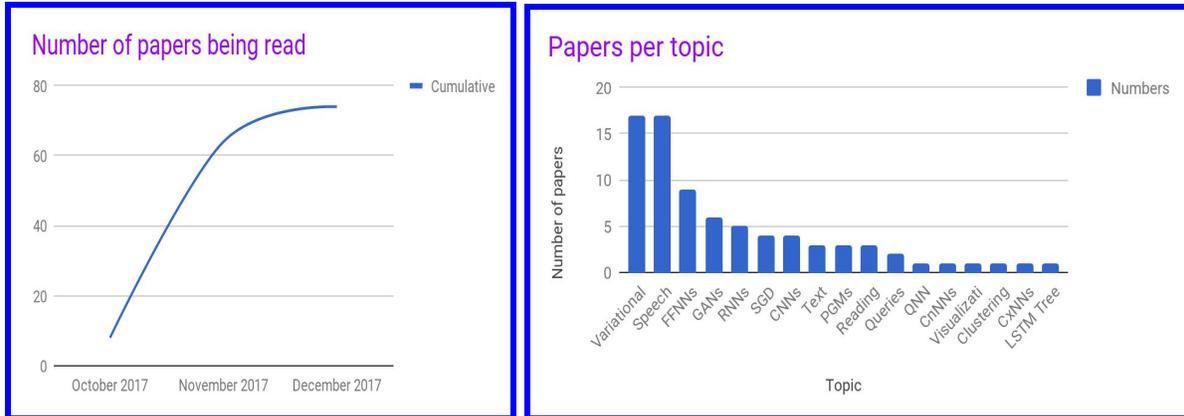


Figure 02: (left) Plot showing the progression of number of papers being read. It should ideally curve upward. (right) Plot showing the papers read per topic.

This must be self explanatory. I think going forward I must separately track the domain related (speech and NLP) vs other (AI, ML) vs miscellaneous (crypto) categories. I have started a [Google group](#) to track this. I believe this is one thing which has a huge potential to grow exponentially. If there are N students in the group and each adds a summary of what they read in the group, we can multiply the number of papers being read. This also provides different perspective. However, the group has not yet taken off. As of today, there are 17 different threads each tracking papers from disparate topics. This can be seen in figure 02. I think its better to track this per week rather than per month.

3) Number of git commits being made

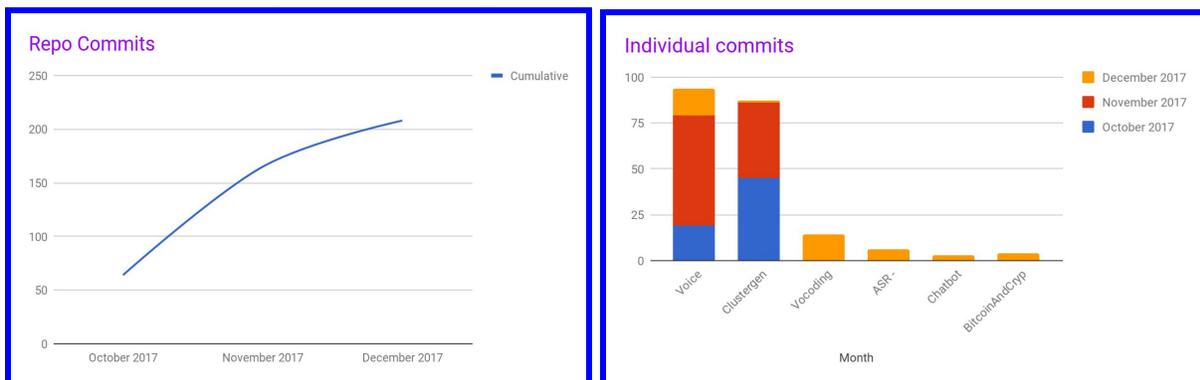


Figure 03: (left) Plot showing the progression of number of git commits being made. I am not sure if it should ideally curve upward. (right) Plot showing the commits read per repo.

Again self explanatory. I think going forward I should demarcate the repos by category. For example, the repo *cluster_gen_steroids* is really an infrastructure repo which holds the base code that is used in most projects I do. So it will have small but significant updates as months go along. On the other hand a live project repo like *Voice Conversion* will easily surpass the commits made in the infrastructure repo because of the sheer amount of energy spent fine tuning.

4) Number of experiments being run

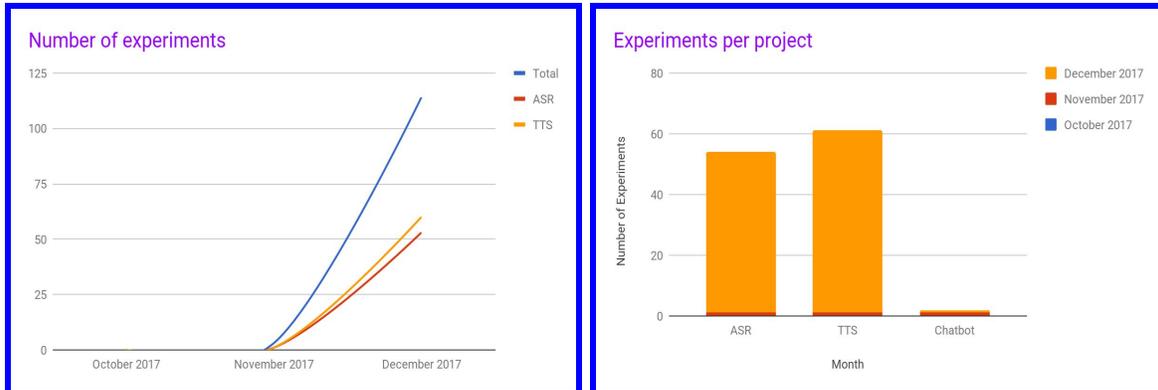


Figure 04: (left) Plot showing the progression of number of experiments.(right) Number of experiments per project.

I have woken up late to this but I think it is important to track the number of experiments being run as well. We typically work with the smartest people on the planet. Eric Schmidt calls such people smart creatives. This means three things: (1) None of the ideas are foolish. (2) None of the ideas are easy because if they were, these smart creatives would have coded them during weekend. (3) There is a reason we stumbled upon that idea. This therefore requires us to do a large number of experiments at scale and vet the idea comprehensively. In other words, we want to (smartly) do all the possible experiments and then do more by morphing the problem at hand. Given that AI fairness is going to be important going forward, I think we should also be able to recreate any experiment at will. I am currently building an infrastructure for this. I am thinking mongoDB + docker + a search mechanism. I am tracking the current experiments being run in different projects in figure 04. A graph like this easily highlights if we are neglecting something.