Using Random Fourier Features with Random Forests

Deliverable 3: Budget and Sustainability

Albert Ribes Marzá

March 19, 2018

Contents

1 Self-assessment on sustainability

When carrying out an engineering project, the first attitude is the best. I know very well the definition of sustainability, its three dimensions, and I am conscious that a project should try to balance the three to the extent possible.

I am aware of the damage that can be done to the planet with a decision in a project, and I appreciate a project for being respectful with the environment, even if I don't notice directly the benefits. I also have a good image of the current situation in this aspect in the world.

Socially, I care about social justice, equity, diversity and transparency. I know that a project only is good if it contributes to the benefit of the citizens, and I know the social consequences that a project I develop could have.

I consider that the purpose of a project should never be to earn money, but see it as a mean, and something needed to get it off the ground.

On the other hand, I don't have the means to get real information and data regarding the environment. I neither know how to evaluate the recycling process a product will have after it useful life, and I am not able to do correct calculations about the pollution of a product in the environment.

I also can't know the social consequences that mi products are having, nor if they are really being used for what I expected. Regarding colaborarive work, I know the tools and techniques to carry projects out, but in practice I don't know how to use it well, or I don't make use of it in the best way.

With the economical dimension it happens a similar thing. I know the theory, but in practice I don't know the means to recognize if a project is producing the expected goods.

2 Analysis of the sustainability of the project

2.1 Environmental dimension

2.1.1 PPP

During the production process of the project there will be needed power supply for charging the computer and general business resources such as light, heating, etc.

The computer consumption is 10 W, and although it is not always the same, I will approximate the business resources as being 0.24kW, so the total consumption is 0.25 kW. As the project is planned to last 240 hours, the amount of power spent will be $0.25kW \times 240h = 60kWh$.

The project doesn't generate residues.

2.1.2 Useful life

As this is project is a very theoretical study, it is hard to say what will be the environmental impact it will have during its useful life. If the conclusion of the project is satisfactory, it may help to develop more efficient machine learning algorithms which will need less energy spends and thus decrease the energy consumption.

Task	Time (h)	Money Spent (€)
Backgroung Approximation	50	1500
Decide the kernel	6	180
Decide Dimensionality	4	120
Decide the changes	10	300
Implement Fourier mapping	10	300
Get familiar with the module	20	600
Modify the module	20	600
Debug the code	15	450
Find testing datasets	5	150
Accuracy tests	10	300
Time tests	10	300
Study the results	10	300
Repeat parts after testing	20	600
Composition of the document	30	900
Total	**240**	**7200**

Table 1: It is considered a salary of 30 € / hour

2.1.3 Risks

The worst it could happen is that the conclusions reached with the project are useless and they do not help to improve efficiency in the machine learning algorithms.

2.2 Economic dimension

2.2.1 Budget

Direct costs

All the software used in this project will be free, and thus they don't increase the cost of the project.

The cost of the laptop will be specified in the "Depreciation" section.

The workforce of the project consists of a single person working for 240 hours. Assuming a salary of 30 €/hour, the labour costs will be 240 hours × 30€/hour = 7200€.

Table 1 shows the direct costs for each task in the project, and the costs of the whole project are summarized in table 2

Indirect costs

As the workspace of the project will be the facilities of the FIB, transport service will be needed. I will use the public transport, so the cost will be 150 €.

Depreciation

The cost of the computer needs to be depreciated. I expect it to have a lifetime of 7500 hours, and the project will spend 240, so the cost is a 3.2% of the total price of the computer. As it was 800 €, the cost of the project is 3.2% × 800 = 25.6 €.

Subject	Type	Amount (€)
Workforce	Direct	7200
Transport	Indirect	150
Power	Fixed	Undefined*
Computer	Depreciation	25.6
Total		7375.6

Table 2: Total costs of the project
* It is not possible to calculate that quantity since I will be using the facilities of the FIB

Unforeseen contingencies

The computer I plan to use to develop the project could have a breakdown. If this happens, the reparation or even replacement will increase the cost of the project.

In order to avoid losing the work done, all the data will be securely stored in GitHub, where it is very unlikely to be lost.

2.2.2 Assessment

I expected that the costs for a project as "simple" as this one would be much fewer, since it involves very little extra costs apart from the salary of the worker. I find the cost is very high given the case that good results are not guaranteed, and I've learned that it is very important to first study the costs of a project, since the conclusions could not be intuitive.

2.3 Social dimension

2.3.1 PPP

This project will teach me how to correctly plan and develop a good working methodology. As this is the fists big project I have to do, it will be very useful to see if only having good programming skills is enough to make projects get on. It will also teach me the easiest failure points in a big project, and will allow me to avoid them in more important projects in the future.

2.3.2 Useful life

Right now the algorithms used for training the Random Forest use some special kind of bagging which allows each node of the trees to use just a partition of the whole data. It is possible in this project that we find a better way to do the bagging allowing each node to see the entire dataset without loosing performance, or even improving it.

It is possible that this project reaches satisfactory results, and it is possible that the method is proven not to be useful. In the first case it will help future studies to train a better learner, and in the second one, it will make other researchers know that this is not a good method, so they will not need to invest resources in performing the same study.

Better learning algorithms are needed nowadays, so there exists a real need of this project and many others to try to develop them.

2.3.3 Risks

As the aim of this study is to provide knowledge, there are not real risks for this project. The worst it could happen is that the conclusions reached with the project are useless.