Plot Exercise in TSRT04 (Data traffic): Forecast of mobile data usage

2022-03-10

This is one of the plot exercises in the course TSRT04. To pass the course you need to solve one of the plot exercises and present it on one of the lab sessions that offers examination. The examination will be in English, so please write your code and comments in that language. The exercise should be done in pairs, or individually. It is not allowed to share or show MATLAB code or notes to other students, since the exercise is part of the examination. It is, however, okay to discuss the exercise orally; for example, to share good advice!

1 Background

The cellular networks were originally built so that we could talk to each other using mobile phones. Nowadays the voice calls and text messages (SMS) are just a small part of the traffic in the cellular networks. It is wireless data traffic from popular Internet services that drives the evolution; for example, social networks, web sites, and video streaming. The increasing data traffic requires a continuous development of the cellular networks with more base stations (on masts and roof tops of buildings) and new technologies (e.g., 4G, 5G). Hence, it seems fair and transparent that it should be the amount of data usage that determines the monthly cost of cellular subscriptions now and in the future—instead of how many phone calls that we make.

2 Description of the data set

In this exercise, we will study a forecast that the company Cisco has made on how the data traffic evolves in Sweden:
`www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html`

The study describes how many petabyte (10^{15} byte) of data that was sent in the cellular networks during 2013 and gives a forecast of how it will change over the next five years. The traffic is divided into three categories: "video", "file transfer" and "web and other". The data set is available in the file `datatraffic.mat` that can be downloaded from the course homepage.

Load the data set into MATLAB with the command

```
» load datatraffic
```

The years are given in the vector `years`, while the annual amounts of data traffic is given in the matrix `traffic`. The first column contains the amount of data traffic (per year) due to video streaming, the second column is the amount of data from file transfers (per year), and the third column shows the same thing but for "web and other". All numbers are in petabyte (see above).

3 Investigating and visualizing the data set

The purpose of this exercise is to investigate the data set by plotting the data in different ways. Follow the following steps:

1. Decide if you will write your code as a script, as different functions, or as a combination of both. If you write a script it might be good to begin with `clear` (which empties the workspace) and `close all` (which closes all open figures/graphs).

2. Plot the amount of data traffic for different types of data, as a function of time. Show everything in the same figure by using `plot`. Use `xlabel, ylabel, title`, and `legend` to describe what is shown in the figure. Think about what kind of data traffic that is dominating and is driving the traffic growth (both when it comes to the amount of data and the slope of the curves).

3. This type of data sets is often presented using bar charts. Plot the same thing by using `bar`. Add descriptive text in the same way as above.

4. There are different types of bar charts. Check the documentation to see how one chooses between "grouped" where every type of traffic has its own bar and "stacked" where the data types are put on top of each other. Use the command `subplot` to plot these types of bar charts in two different subfigures; for example, underneath each other.

5. Return to the line plot that you made with `plot`. What do these numbers mean for an average person in Sweden? Assume that there are 9.7 million Swedes and that this number will stay fixed over all the years considered in the data set. Change the scaling of `traffic` so that it gives the data traffic for an "average Swede". It might be good to also change the scaling from petabyte (10^{15} byte) to gigabyte (10^9 byte). Check if the numbers seem reasonable and if they describe your own data usage.

6. Think about how the line plot would look like in a printout. There are two important things to keep in mind: 1) Use "strong" colors, such as black, red, and blue. Green and particularly yellow is often hard to see in printouts. 2) Make sure that it is possible to separate the curves also in a gray scale printouts. This requires different line styles (e.g., solid, dotted, dashdotted and dashed). You can use `plot` to select color and line style.

7. Add a curve in the line plot that shows the total data traffic. It is easier to read a graph if the boxed description (created using `legend`) lists the curves in the same order as the curves are located in the graph. Make sure that this is the case.

8. Save your figures, both in the MATLAB-format `.fig` and in the following four common image formats: `.jpg`, `.png`, `.pdf`, and `.eps`. Open the saved figures and compare the image qualities. Zoom in on each image and see which formats that always are sharp (so called vector graphics) and which ones become unsharp when zooming. Then you should import the figures into a word processing program (e.g., Word or Open Office). Which formats are most suitable for reports and similar use? Return to MATLAB and click-drag on the lower right corner of the figure to change its size and dimension on the screen. Save the file once more and check if your changes carried over to the saved file or not.

9. Preview how the figures will look like in gray-scale. This can be done by choosing "Print Preview" in the menu "File" of a figure. You can switch between color and gray scale in the tab "Color". If it is not possible to identify each curve in gray-scale, you need to make necessary changes and make a new preview.

To pass this exercise you need to show and demonstrate the script and/or functions that you have created to solve the different tasks. You should have written comments in the code and made sure that the plots are self-explaining.

4 More background on the increasing data traffic

The forecast from Cisco (that the data set originates from) has made a large impact, in particular, among telecom operators and researchers that are planning the future cellular networks. The forecast shows a stable increase in wireless data traffic with 35% per year. To understand what this means, we let $t = 0$ denote our starting year (e.g., 2013 in the forecast). If the amount of traffic is denoted by d (unit: byte/year) at our starting year, then the amount of traffic at year t after the starting year is approximately equal to

$$d \cdot 1.35^t. \tag{1}$$

Note that this is an exponential increase over the years. Each time you see an exponential rate of increase it is reasonable to question how long it may continue—if there are any physical limitations that sooner or later will prevent the increase.

One might think that it is our modern IT society (with smart phones and tablets) that has created this exponential increase of wireless data traffic. But it has actually been like this since the beginning of wireless communication. Martin Cooper, who was a pioneer on cellular communications at Motorola, formalized in the nineties what is known as *Cooper's law*. It says that the amount of wireless communication has doubled every 2.5 year. This rule-of-thumb law has been quite accurate for the whole 20th century; thus, the traffic has increased by a factor 1 million over the last 45 years. If this is turned into an annual increase, $2^{1/2.5} = 1.32$, we see that the rule-of-thumb suggests an increase of 32 % per year. This is very close to Cisco's forecast for Sweden. In other words, the exponential increase is nothing new, but it is exciting to see where it will lead us in the future.

If you like to, you can read more about Cooper's law:

`http://www.arraycomm.com/technology/coopers-law/`

5 Reference

Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2013–2018. URL: `http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html`