MARCH 2018

ISSUE 04

DATA STREAK

A MONTHLY DIGEST ON ALL THINGS DATA

UpGrad

PROFESSOR'S DESK

By Prof. Sadagopan

Director at

International Institute of Information Technology Bangalore

Undoubtedly 'analytics' is a buzzword today. Every student wants a course on analytics and its variants – Data Science, Machine Learning, Artificial Intelligence and Big Data Analytics. Every company says they are building 'analytics' and 'machine intelligence' practices as new verticals. Every manager wants to recruit 'data scientists' and the media hypes it by talking of 'acute shortage' of professionals and the promise of 'x,y' million jobs right now! **Are analytics new 'nirvana'? Is it all so sudden? Not really. Let me elaborate in this blog...**

Analysis and analysts have been around for a long time. The best known and best paid of them all are **Wall Street Analysts** who analyse stock market behaviour and come up with recommendations for financial institutions like Investment Banks, Fund Managers and Stockbrokers. An analyst is considered good, if his/her recommendations lead to profitable decisions, consistently over a period of time. The brightest youngsters – often with a PhD in Mathematics, Physics or Computer Science from Ivy League Institutions like MIT– choose to become analysts. The number of such successful Wall Street Analysts is at best a few thousands. Of course, there are many other analyst-oriented organizations like Gartner which keep producing high value and expensive reports that evaluate a whole range of IT products and services. Such reports are subscribed to by CXO's who use them to decide on IT purchases. There are many other analysts like industry analysts, election analysts, policy analysts and security analysts.

All of them generally are specialists and their numbers are in hundreds to thousands for each industry/government segments or countries/geographies.

Analysts being specialists, make significant investments in understanding their domain through education (often a PhD) and training; they collect data from authentic sources, often investing heavily. They use very sophisticated tools (statistical analysis, optimization and simulation) using software that are generally expensive; such sophisticated software needs significant computing power. Finally, the results of such in- depth analysis must be communicated exceptionally well using writing and presentation tools which often need experts who must be deployed at high costs. All in all, at every stage, there are costs involved. Naturally, 'analysis' has remained the exclusive privilege enjoyed by CEO's and Government/Policy/Institutional heads.

The internet, WWW & software bots suddenly made significant amounts of data available to everyone 24/7 at nearly zero costs.

Most organizations that keep the data found the WWW to be the best place to store it. Cloud computing made the data collection and storage asset light (not needing computer hardware at every place). Mobile devices, smart phones and their universal availability (nearly four billion users) have made data almost freely available, on a 24/7 basis at the point of consumption. Analysis of data is also practically free of cost that too at an 'anytime, anywhere' basis, thanks to enormous computing power on the cloud and visualization tools on the mobile devices. In other words, data is available to practically everyone and analysis is available to even lay users, thereby democratizing analysis from analysts to end-users!

Elements of such democratization started in niche areas. For example; networking companies for nearly two decades have been providing network analytics – packets transmitted, network speed, packet loss, high throughput nodes etc. With the arrival of web-servers for every organization web-server analytics started to be available nearly a decade back – nodes accessing the web-server, time spent on the server, number of times a web-server was unreachable ("404" error), pages that were frequently accessed, etc. Perhaps the best-known tool in this area is Google Analytics that has been improving over the years to the point that marketing folks can use the data to analyse consumer behaviour!

What is happening in the past couple of years is full-scale democratization of analytics. That explains the explosive growth, almost into hype of interest in analytics that I referred to earlier. The next decade, if not several decades, will truly be the 'age of analytics'. This is perhaps the reason why Harvard Business Review, October 2012 termed analytics as 'the sexiest job of the 21st Century!

Happy reading!

Manikanta Yadunanda
Member Of Technical Staff 2
Adobe

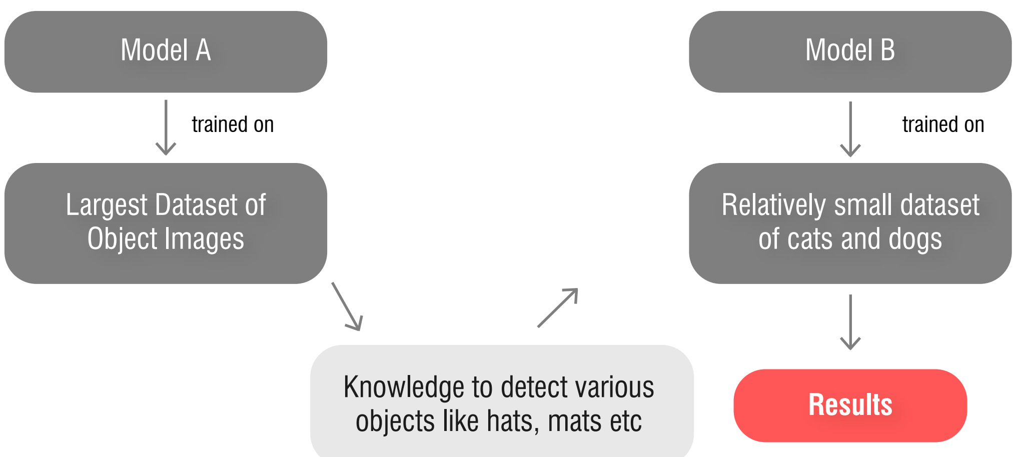
TRANSFER LEARNING
USING DIFFERENTIAL LEARNING RATES

So what is transfer learning?

It is the process of using the knowledge learned in one process/activity and applying it to a different task. Let us take a small example, a player who is good at carrom can apply that knowledge in learning how to play a game of pool.

What about our machine learning/deep learning perspective…

The same applies to our machine learning world. Let's have a look at the below picture.



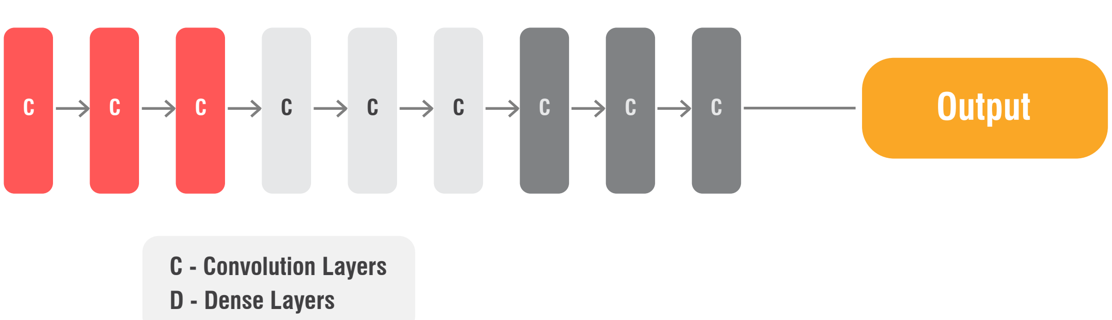
Let's say, model A's task is to identify 1000 kinds of objects like hats, cats, mats and we have such a trained model at our hand. Now let's suppose we want to create a model B to detect a cat/dog classifier. Even if we have a small dataset we can use the knowledge of model A during the training of model B and produce state of the art results.

But why should one use transfer learning?

Whenever we want to solve a unique problem using machine learning the chances are high that we might not find enough data for our model. Training with lesser data will not give you good results. And even if we have large data there is a possibility of not having enough resources like GPU to obtain high-quality results. So transfer learning addresses these problems by already using the knowledge in the form of a pre-trained model which someone has created with large datasets and resources.

Ok, can you tell us how to do it?

Let's understand it by using a sample network diagram of CNN. Although in practice the networks are large, complex and will contain various other units.



The most general way of doing it is by modifying the dense layers such that the network output suits our task at hand and train only the newly added layers. This works decently when the tasks are related to each other and you have a small amount of data. For example; if we are using a pre-trained model which already knows how to detect cats, this approach will work if we want to create a cat/dog classifier with a small amount of data.

Second approach is to even include the convolution blocks closer to dense blocks (blue ones in the diagram) in training. This is more ideal if we have a medium size dataset and the tasks are not so tightly related.

The third approach is to use the pre-trained model but to train all the layers with the dataset. This works well but requires a relatively large dataset and GPU resources. There are a couple of cool tricks if we are taking this option which we are going to cover below.

Manikanta Yadunanda's
Blog Link

Click here to visit his blog!

Mixing first and third approach

If we decide to use the third approach, it is better to apply the first approach for some epochs of training to bring up the newly added layer weights to a better point. Then we can unfreeze all the layers and follow our third approach.

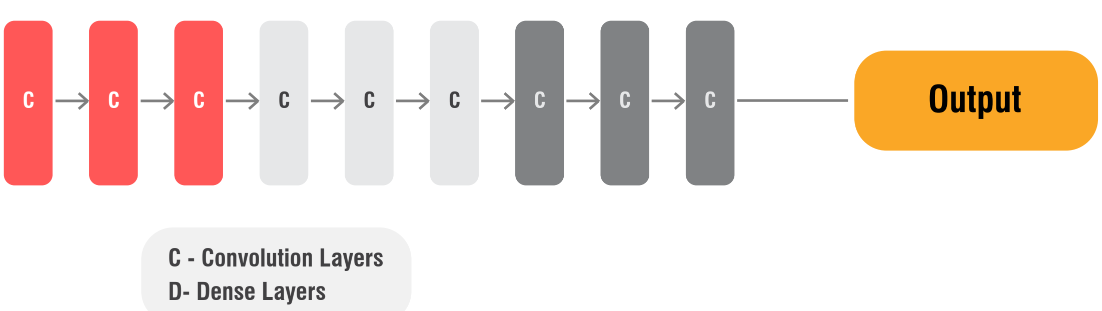
Differential learning rates

The phrase 'Differential learning rates' means to have different learning rates for different parts of the network during our training. The idea is to divide the layers into various layer groups and set different learning rate for each group so that we get ideal results. In simple terms, we control the rate at which weights change for each part of our network during training.

Whenever we want to solve a unique problem using machine learning the chances are high that we might not find enough data for our model. Training with fewer data will result in not so good results.

Why? How does it help?

If we consider the third approach above, there is a small but significant point to be noticed. To understand it let's go back to our sample CNN figure.



In general, layers in red learn generic features like edges, shapes and the middle blue layers learn specific details with respect to the dataset on which it is trained.

Given the above statement, it's not a good idea to change the learned weights on the initial layers too much because they are already good at what they are supposed to do (detecting the features like edges etc). Middle layers will have knowledge of the complex features that might help our task to some extent if we slightly modify them. So, we want to finetune them a little.

Differential learning rates help us in this regard. We can now imagine sample network in three-layer groups (red, blue and green) and set different learning rates. The initial red layers will have small learning rate as we don't want to disturb them much, the middle blue layers will have learning rate higher than initial layers and the final green layers will be having the highest learning rate that's optimal.

How much lesser or more the learning rates for initial and middle layers will be, depends on the data correlation between the pre-trained model and our required model. For example, if the task is to create a dog/cat classifier and our pre-trained model is already good at recognizing cats, then we can use learning rates of less magnitude. But if our task is to create some model on satellite imagery/medical imagery then we will have learning rates of slightly higher magnitude.

Please note that most of the deep learning libraries currently do not support differential learning rates.

Conclusion

Depending on the task at hand and resources one should choose an appropriate approach in transfer learning. In general, if we have a good amount of data and resources, transfer learning with differential learning rates will yield better results.

Youtube Videos to check out

[Linear Regression - Machine Learning Fun and Easy](#)
[Why you should love Statistics | Alan Smith](#)

DO YOU STILL THINK IN HOURS WHILE ESTIMATING USER STORIES ?

A user story should not be given an estimate by keeping only hours in mind. It is very common to map story points with hours (days) and estimate a user story. People in their back of their mind calculate, how many hours they need to complete this work, compare that with points or other unit of estimation used within the team and then quote their answers. I have observed the use of basis like 1 point = 1 day of work at quite a few workplaces.

Anurag Singhal

Business Analyst
Computer Sciences Corp

In this article I'll try to detach hours with any unit used for Agile estimation. Human effort to complete a task is never fixed even if you have done the same thing a hundred times. Speed of a photocopy machine can be measured in X copies per minute or a paper shredder can cut Y papers per minute. A soda maker may produce about 3 litres of black soda per minute.

But can you say how many/much you would be able to produce using these machines. I guess the answers will vary by person, but no one can say a fix number of output. And if I ask will it be an easy/not easy/difficult task to use this machine for a minute? This question would certainly be answerable for everyone.

Over a period of time people who found it difficult may find using these machines an easy task but giving a fixed number or quantity you can produce in a minute would still result in some hiccups. Things are not certain in this world unless they are a fact. It is close to impossible to estimate the size of a story, task or requirement in number of minutes or hours. One can only say how comfortable they are with the work, primarily less or more. Estimating as how less or more one is comfortable with the work is called relative complexity.

There is a story that goes like this….
As a <random person> I want to leave Source A to reach Destination B using my car so that I can attend an ABC conference at 11 am.
Before we talk about estimations, let's look at few pointers:

1. Destination A is very popular place in the city and it is unlikely that anybody belongs to the city have not heard or not been to that place.
2. Distance between A and B is fixed via a preferred route. Although drivers are open to change route as per their convenience.
3. Time to leave is not fixed from destination A. Only condition is that one should be able to attend conference ABC at 11 am (The value).
4. Also one has to choose mode of communication is by car only.

Now, think about how would you decide that attending conference at 11 AM is easy or difficult for you.

- Your familiarity with the route (Any drive in the past to the destination)
- Traffic conditions
- Your average driving speed
- Any unknowns which might add to driving time

If you can judge the time to leave for the destination as per your comfort level based on the above 4 factors (or may be more), you can surely estimate a user story without using hours/days unit.

Anurag Singhal's
Blog Link

Click here to visit his blog!

Q1. What was your primary motivation to learn Data Science?
In recent times many researchers have stated that, data science is critical to the success of any kind of business in the present market/ society scenario like Ecommerce, Banking, Retail, Real Estate, Pharma and many more industries. Therefore, I presumed there is a sustainable career growth in data science and thus decided to do the course and make the most out of the career opportunity in this domain.

Q2. Why did you choose Upgrade to Learn data science?
I chose UpGrad because the curriculum here is designed by industry experts, giving us very structured information. This one-year PGDDA is a collaboration between IIITB and UpGrad.
During this program we covered all the tools and techniques that were useful.

Q3. How did you come to know about UpGrad?
I saw a few advertisments on Google and Facebook. Initially I wasn't sure about the prerequisites and how much this course could add value to my career. I will not deny that at that time there was some apprehension in my mind, as I had already completed one year in a specific industry/domain. Yet I went ahead filled in the enquiry form and requested for a call back. After a quick discussion with an UpGrad Counsellor, it started seeming much better than before.

Q4. What impact did the Data Analytics have on your career?
Though I have been doing data analytics and data mining work at my existing job, this program gave me a structured approach to the process and outcome of it.

Q5. What extra steps outside the program learning did you take to escalate this career transition?
I have learnt R and Python basics before enrolling into this program and would suggest the same for the enthusiast learners to take full advantage of the course.

Q6. Would you recommend the program to any of your friends/colleagues? Why?
I would definitely recommend this program. The course contents in the program are very clear and precise and well structured. Though it does not require in-person presence at the university campus, it is very close to classroom sessions. The discussion forums where everybody is there to help you for doubts is absolutely amazing.

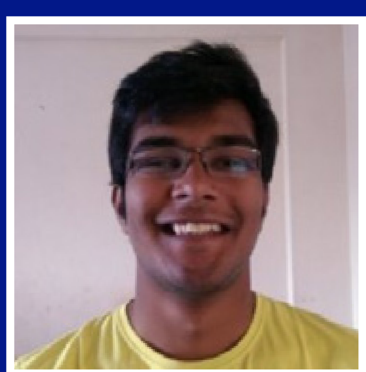
STUDENT SUCCESS STORY

Narendra Prasad
Manager Analytics
Inntti Alliance

More Student Testimonials

[Gaurav Choudhury's Testimonial]
[Srungeer Simha's Testimonial]

RECENT CAREER TRANSITIONS



Sagar Seth
Cohort 2

Associate Engineer → Techical Consultant

RYZKLYTIX → SAS THE POWER TO KNOW.

Kabilan Karunakaran
Cohort 2

Senior Analyst → ML Engineer

Alchemy → SPi Global

Arun Naudiyal
Cohort 2

Senior System Engineer → Consultant (ML/Text Analytics)

Infosys → Deloitte.

Sai Krishnan
Cohort 2

Global Leader - Finance BI COE → Director & Finance Global Process Owner

Honeywell → Honeywell

Akash Kumar
Cohort 2

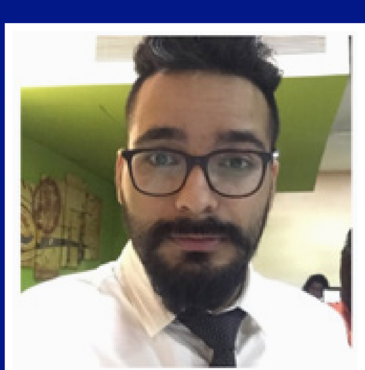
Software Engineer
CGI

Data Engineer
DXC.technology

Aman Avilash
Cohort 3

Assistant Manager (Operations)
SHOPCLUES.COM

Senior Business Analyst
OLA



Vipan Goja
Cohort 3

Senior Manager
Coverfox

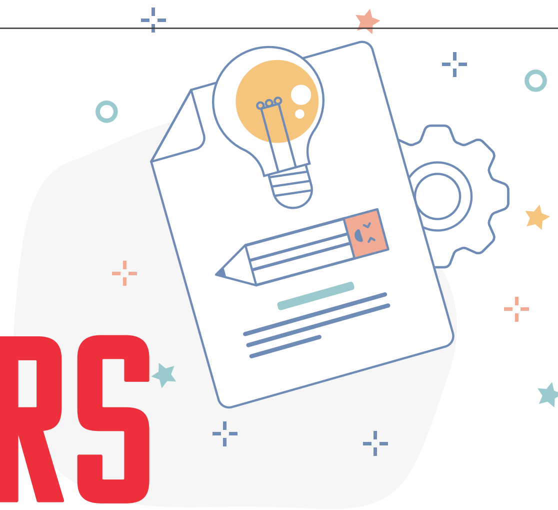
Product Analyst
wooplr

Krishnakanth C
Cohort 3

Senior Systems Engineer
Infosys

Technology Analyst
Infosys

DATA TICKLERS



BRAIN TEASERS

1 There are 8 batteries, but only 4 of them work. You have to use them for a flashlight which needs only 2 working batteries. To guarantee that the flashlight is turned on, what is the minimum number of battery pairs you need to test?

2 A birthday cake has to be equally divided into 8 equal pieces in exactly 3 cuts. Determine the way to make this division possible.

CLICK HERE TO SEND US YOUR FEEDBACK

FIND US HERE:

To share your stories/articles/blogs,
write to us at **pgdds@upgrad.com**

UpGrad Education Private Limited, Nishuvi, 75, Dr. Annie Besant Road, Worli, Mumbai – 400018.