**REPORT**

**DATA SCIENCE: Industrial Perspective**

**TOPIC 1:** Digital Disruption with Augmented Data science is Essential for Energy 4.0

**Research Area:** Application of machine learning and artificial intelligence in oil and gas industry

**Methods Proposed:**

**Energy 4.0:**

1)It has been seen crystal clear that the digital technology has a tremendous influence on business and society.

2)With time it has been seen that digital transformation is regarded as the "fourth industrial revolution", characterized by the convergence of technologies that blur the boundaries between the physical, digital and biological realms, such as artificial intelligence, robotics and

autonomous vehicles.

**Importance of data:**

1) The performance of electronic devices is enhanced due to

increment in data processing capabilities. It is desirable for oil and gas industries to use computing power for production and exploration.

2)Hydrocarbon exploration is riddled with risk. The explorationist need to identify subsurface prospects accurately for drilling and exploitation of hydrocarbon.

3)In the early 21st century limited 2D seismic data were considered to pinpoint the drilling locations based on subsurface mapping. Since it is riddled with risk the chance of success was 1:7. With time more data was acquired in each of the lease curved out for exploration.

4)This large volume of data was termed as big data which was stored in Terabytes of memory space with the advancement in acquisition, processing and interpretation of seismic and well data. These big data was analysed using the machine learning concept.

5)The objective behind use of big data and applicability of machine learning is to improve the signal to noise ratio during acquisition and processing.

**Conclusion:**

1) Machine learning has the potential of unequivocally changing the numerous critical actions made every day by administrators and engineers in the oil and gas sector.

2) Many such solutions utilizing ANN, ALM, supervised

learning, fuzzy logic, linear regression and PCA could be enforced to counteract various difficulties found in oil and gas industries and helps in maturing for profitable strategies.

3)In the forthcoming years, the increase of machine learning utilization may begin to expand rapidly, as well as its value will also be significantly utilised throughout the oil and gas industries.

**TOPIC 2:** Perceptual QoE-Optimal Resource Allocation for Adaptive Video Streaming

**Research Area:**

1)Enhancing the user’s QoE has gained a lot of attention in a

variety of over-the-top (OTT) multimedia services in the recent times.

2)This is primarily because higher user QoEs have been

shown to translate to better revenues for service providers

Maintaining an ‘acceptable’ user QoE is of paramount importance for content providers such as Netflix, YouTube, Hulu etc.

3)However, providing higher QoE and maintaining it throughout the streaming session is challenging as the end-to-end network conditions are time-varying. Hence, there is a need for QoE based service provisioning and network design for video delivery.

**Methods Proposed:**

1)Several works propose methodologies and provide guidelines for the QoE based design of networks for

multimedia streaming and services.

2) However, most of these works employ network-level quality-of-service (QoS) metrics such as average throughput, average packet delay etc. for quantifying the QoE.

3)The QoE is calculated in terms of QoS parameters

using a model that describes an exponential relation between

the QoS and the QoE.

4)The work provides a nonlinear formula to map the QoS metrics such as bitrate, frame rate, and packet loss rate to estimate MOS (QoE).

5)A survey of QoE considerations for adaptive streaming and guidelines for rate adaptation based on prior works is presented.

6)A subjective study has been conducted to identify the

impact of adaptation parameters on QoE.

**Conclusion:**

1) a QoE based algorithm for resource allocation in cellular networks for adaptive video streaming.

2)The proposed algorithm exploits the strength

of continuous perceptual QoE prediction models for performing optimal allocation of resources in the network. 3)The average rebuffering time using the proposed method is significantly reduced owing to the availability of the user’s perceptual QoE information at the eNodeB, the central resource allocator.

**TOPIC 3:** Modern Approaches to No Reference Image Quality Assessment

**Research Area:** Application of machine learning and artificial intelligence in Image quality Assessment

**Methods Proposed:**

1) A reference-free image distortion map generating algorithm for spatially localizing distortions in a natural scene.

2) No reference image quality assessment (NRIQA) algorithms derived from the generated distortion map. We use a convolutional autoencoder (CAE) for distortion map generation. We rely on distortion maps generated by the SSIM image quality assessment (IQA) algorithm as the “ground truth” for training the CAE. We train the CAE on a synthetically generated dataset composed of pristine images.

3) Image quality assessment plays a very important role in

the image signal processing (ISP) pipeline.

4)This claim is supported by the observation that camera quality is one of the prime features considered when a new device is bought.

5)Since most natural image and video content is meant for

human consumption, subjective opinion on quality is very

important.

6)Subjective opinion is quantified using mean opinion

score (MOS) and the difference in mean opinion score

(DMOS). However, a subjective study is not always feasible in many practical applications because it is expensive and time consuming.

7)An alternative approach is to design an objective

metric which automatically predicts the perceptual quality

of the image such that it correlates well with subjective opinion.

**Conclusion:**

1) We presented a novel approach to generate image distortion

maps in the no reference setting using CAE and demonstrated

its application to NRIQA.

2) The proposed approach is capable of localizing the distortions of test images in addition to delivering

competitive NRIQA performance.

3)NRIQA techniques are crucial in designing and monitoring various vision related systems like automatic camera parameters tuning to maximize the image quality while acquisition.

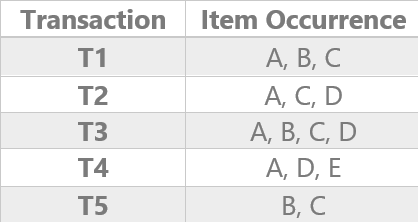
**TOPIC 4:** Frequent Pattern Mining

**Research Area:** Application of machine learning and artificial intelligence in Data mining

**Methods Proposed:**

1)Frequent Pattern is a pattern which appears frequently in a data set. By identifying frequent patterns we can observe strongly correlated items together and easily identify similar characteristics, associations among them. By doing frequent pattern mining, it leads to further analysis like clustering, classification and other data mining tasks.

2) **Support**: how often a given rule appears in the database being mined



**Confidence**: the number of times a given rule turns out to be true in practice





Example:

One of possible Association Rule is A => D

Total no of Transactions(N) = 5

Frequency(A, D) = > Total no of instances together A with D is 3

Frequency(A) => Total no of occurrence in A

Support = 3 / 5

Confidence = 3 / 4

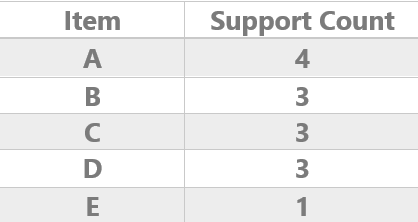
After getting a clear idea about the two terms Support and Confidence, we can move to frequent pattern mining. Frequent pattern mining, there are 2 categories to be considered,

1. Mining frequent pattern with candidate generation
2. Mining frequent pattern without candidate generation

In this article, we are focusing on Mining frequent patterns with candidate generation with Apriori Algorithm which is popularly used for Association mining. Let’s understand Apriori Algorithm with an example and it will help you to understand the concept behind it in a clear manner. Let’s consider the sample data set mentioned above and assume that the **minimum support=2**.

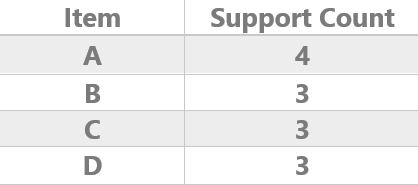
**1.Generate Candidate set 1, do the first scan and generate One item set**

In this stage, we get the sample data set and take each individual’s count and make frequent item set 1(K = 1).



Candidate set 1

“Candidate set 1” figure shows you the individual item Support count. Hence the minimum support is 2 and based on that, item E will remove from the Candidate set 1 as an infrequent item (disqualified).

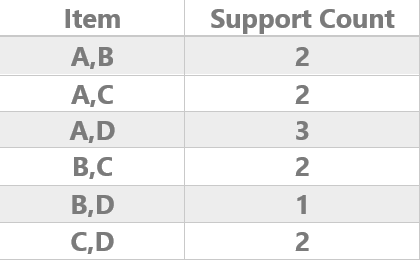


Frequent itemset from the first scan

Frequent itemset based on the minimum support value will be shown under the figure “Frequent item set from the first scan” as the “**One item set**”.

**2. Generate Candidate set 2, do the second scan and generate Second item set**

Through this step, you create frequent set 2 (K =2) and takes each of their Support counts.



Candidate set 2

“Candidate set 2” figure generates through joining Candidate set 1 and takes the frequency of related occurrences. Hence the minimum support is 2, Itemset B, D will be removed from Candidate set 2 as an infrequent item set.



Frequent item set from the second scan

“Frequent item set from the second scan” is the frequent itemset based on the minimum support value and it will generate the “**Second item set**”.

**TOPIC 5:** Generating High-Resolution Climate Prediction through Generative Adversarial Network

**Research Area:** Application of machine learning and artificial intelligence in climatology

**Methods Proposed:**

1)Regional climate prediction from large-scale climate output information. Inspired by image super resolution, we propose to apply the super-resolution models to downscaling technology. 2)However, some unpleasant artifacts always appear in the high-resolution climate images generated by exiting super-resolution models.

3)To further eliminate these unpleasant artifacts, we innovatively apply the super-resolution generative adversarial network (SRGAN) to climate prediction.

4)SRGAN adopts a perceptual loss function which consists of an adversarial loss and a content loss, which can recover more high-frequency details in the generated high-resolution images.

5)Besides, we propose a method to fuse climate data with related meteorological factors to generate climate images, this measure can improve the accuracy of climate prediction.

6)Finally, extensive experimental results on climate datasets show that SRGAN performs better than most super-resolution approaches in climate prediction.

**Conclusion:**

1) The RB(residual block) as the basic network-building unit in the generator network, and we can use RB to extract more feature maps from the input images.

2)They adopt a perceptual loss to optimize the SRGAN model, instead of the pixel-wise loss, and our algorithm using this loss can recover more detailed textures in the final generated high-resolution images.

3)Experimental results also illustrate that our method can generate high-quality climate images, and perform the best in climate prediction.

Topic 6: Potential of Metrological data for Industrial R&D

Research Area: Artificial Intelligence and Machine learning

Method Proposed

Real time accessibility and historical information about weather and climate are considered essential commodities in modern societies. For example, in the Indian context, seasonal forecast are extensively used to decide crop choice, irrigation and harvest date during the monsoon season. City as well as village level planning also nowadays use weather information. In the context of climate change, weather and climate information have become even more valuable. It is now a norm that city governments and international companies that must design long term climate change adoption plans.

In the lecture, first, there was discussion about natural disasters such as flood, drought etc. And the effects of these disasters and the solution to cope up with them. As the research takes knowledge and investment, we discussed that the governments of all the countries are contributing more and more in this field. For example, the climate finding was $574 billion in 2017-18 which was about 24% more than the previous year. Also, the Indian government has a budget of about ₹2,870 crore.

Then we further proceed to discuss about some important terminologies, the weather distribution of India as well as of the world. There is lots of fluid dynamics concepts involved in the analysis along with the probability and statistics. We then discuss about some of the examples of involving supervised and unsupervised learning like, rainfall dynamics, downscaling problem I.e, downscaling is close to super resolution technique used in image processing, understanding extratropical cyclone, cloud formation (an example of supervised learning), monsoon prediction, etc.

**Challenges:** In the weather prediction we have to deal with a large amount of data, we have to analyse in in very less time which involves a lot of computational cost.

**Future Scope:** In future, we have plans to utilise low-cost Internet of Things (IoT) devices, such as temperature and humidity sensors, in collecting weather data from different parts of a city. The use of different sensors could increase the number of local features in the training dataset. This data, along with the weather station data, will further improve the performance of our prediction models.

**Conclusion:** As machine learning advances, more models start integrating, and going with it accuracy becomes more better and forecasting will become increasingly accurate. More scope and potential of global nowcasting, which is relatively new addition to weather forecasting. As smart system penetration grows worldwide, more people will gain access to accurate, hyperlocal weather forecasting as well.