UPES

UNIVERSITY OF PETROLEUM & ENERGY STUDIES

School of Computer Science

Application of Machine Learning in Industries

Assignment 2

Submitted By:

Dhruv Singhal

500075346

R177219074

AIML B3

Submitted To:

Mr. Goutam Datta

Asst. Professor

SOCS

UPES

Q1 • How ML techniques can be helpful in transportation sector? Discuss in brief.

Machine learning had great applicability in the transport industry. In recent years, ML techniques have become a part of smart transportation. Through deep learning, ML explored the complex interactions of roads, highways, traffic, environmental elements, crashes, and so on. ML has also great potential in daily traffic management and the collection of traffic data.

The biggest answer to this question can be the example of self-driving cars which uses AI & ML to operate on their own. Based on their training in the different situations, they learn or memorizes the commands in different situations. Various tracking apps like google maps also uses ML to tell its user the expected time, best route and optimal route for their journey.

• Discuss in detail the different ML techniques that are used in legend based automatic drug discovery process.

ML Algorithms Used in Drug Discovery ML algorithms have significantly advanced drug discovery. Pharmaceutical companies have greatly benefited from the utilization of various ML algorithms in drug discovery. ML algorithms have been used to develop various models for predicting chemical, biological, and physical characteristics of compounds in drug discovery . ML algorithms can be incorporated in all steps of the process of drug discovery. For example, ML algorithms have been used to find a new use of drugs, predict drug-protein interactions, discover drug efficacy, ensure safety biomarkers, and optimize the bioactivity of molecules . ML algorithms that have been widely used in drug discovery, which include: Random Forest (RF). Naive Bayesian (NB), and support vector machine (SVM) as well as other methods.

Q2 Compare Statistical and Neural Machine Translation system in detail. Also mention some of popular tools that can be used in statistical and neural based translation system design.

Statistical Machine Translation Tool:

SMT uses predictive algorithms to teach a computer how to translate text. These models are created or learned from bilingual text corpora and used to create the most probable output, based on different bilingual examples.

One such tool which uses SMT is Moses.

Moses: Moses is a statistical machine translation system that allows you to automatically train translational models for any language pair. All you need is a collection of translated texts.

SMT uses target language data to act as a type of filter so that low scoring phrases are not selected for the final output

SMT is a lot more tolerant of data quality issues.

SMT will still give some semblance of quality when the translation input is out-of-domain.

Neural Machine Translation Tool:

NMT is the newest method of MT and is said to create much more accurate translations than SMT. NMT uses deep learning techniques to teach itself to translate text based on existing statistical models. It makes for faster translations than the statistical method and has the ability to create higher quality output.

One such tool which uses NMT is OpenNMT.

OpenNMT is a generic deep learning framework mainly specialized in sequence-to-sequence model covering a variety of tasks such as machine translation, summarization, text-to-speech, speech recognition, Yoon Kim, from the Harvard NLP group publishes the project seq2seq-attn that lays the foundation of the OpenNMT initiative.

NMT does not have the same language model concept and instead builds a sort of language model from the bilingual training data.

NMT requires higher quality training data than SMT.

NMT will simply give up when content is too far out-of-domain and just produce gibberish that is not relevant at all.

Q3 Write a short on the following :

• Application of ML based techniques in manufacturing industry for fault assessment.

The machine learning (ML) field has deeply impacted the manufacturing industry in the context of the industry 4.0 paradigm. The industry 4.0 paradigm encourages the usage of smart sensors, devices, and machines, to enable smart factories that continuously collect data pertaining to production. ML techniques enable the generation of actionable intelligence by processing the collected data to increase manufacturing efficiency without significantly changing the required resources. Additionally, the ability of ML techniques to provide predictive insights has enabled discerning complex manufacturing patterns and offers a pathway for an intelligent decision support system in a variety of manufacturing tasks such as intelligent and continuous inspection, predictive maintenance, quality improvement, process optimisation, supply chain management, and task scheduling.

The manufacturing industry has become more and more important to realize intelligent manufacturing. The researches in industrial fault diagnosis and prognosis of PHM (Prognostics Health Management) Most of the industrial data are time series with multiple features collected from different sensors and a good feature presentation is very important for accurate results. In order to extract features automatically and get accurate results, we propose an end-to-end method, CNN-LSTM, for industrial data based on deep learning and transfer learning, where CNN extracts features automatically and LSTM analyses the new feature sequences from CNN. We evaluate our method on the dataset of wind turbines blade-icing problem and get good results compared with other models. It shows we propose an effective end-to-end method for industrial fault diagnosis and prognosis.

• Application of ML based techniques in retail industry and in Oil and Gas sector(Exploration)

One of the most noticeable impacts of machine learning in oil & gas focused industries is how it transforms discovery processes. Applications employing machine learning in oil & gas enable computers to quickly and accurately analyse huge amounts of data. This includes being able to sift precisely through signals and noise in seismic data.

After this information has been gathered and analysed modern software applications can construct accurate geological models. This allows operatives to predict, accurately, what is beneath the surface before drilling has begun. Known as reservoir modelling this also allows users to predict how formations will react to certain drilling techniques. By using machine learning in oil & gas exploration in this way we will know where and how to drill. A combination of algorithms and fuzzy logic will verify these models.

As well as verifying models, this also allows for accurate predictions to be made when information is incomplete or deemed unreliable. This application allows for a model to be constructed and for virtual drilling will take place. Allowing engineers to pinpoint the best route through a rock formation. This process can occur long before any physical drilling equipment even arrives on site. Knowing exactly what is beneath the ground, and how the ground will react to drilling, we can avoid further flops like McMoRan's well Machine learning in oil & gas, as well as AI and case based reasoning, will save time and money. It will also help operatives to learn lessons from past successes and mistakes. The more this technology is used, and the more information it is given, the more accurate the models will become

• Evaluation of Machine Translation using BLEU (Students are advised to include their finding in parallel corpus they have used in the MT project)

Bilingual Evaluation Understudy Score, or BLEU for short, is a metric for evaluating a generated sentence to a reference sentence. A perfect match results in a score of 1.0, whereas a perfect

mismatch results in a score of 0.0..The score was developed for evaluating the predictions made by automatic machine translation systems. It is not perfect, but does offer 5 compelling benefits:

- **It is quick and inexpensive to calculate.**
- **It is easy to understand.**
- **It is language independent.**
- **It correlates highly with human evaluation.**
- **It has been widely adopted.**

```
from nltk.translate.bleu_score import sentence_bleu
ref = [
    'this is moonlight'.split(),
    'Look, this is moonlight'.split(),
    'moonlight it is'.split()
]
test = 'it is moonlight'.split()
print('BLEU score for test-> {}'.format(sentence_bleu(ref, test)))

test01 = 'it is cat and moonlight'.split()
print('BLEU score for test01-> {}'.format(sentence_bleu(ref, test01)))
```

```
BLEU score for test-> 1.0
BLEU score for test01-> 0.6223329772884784


/opt/conda/lib/python3.7/site-packages/nltk/translate/bleu_score.py:490: UserWarning:
Corpus/Sentence contains 0 counts of 3-gram overlaps.
BLEU scores might be undesirable; use SmoothingFunction().
  warnings.warn(_msg)
```