Part B

Introduction to AI

Naturalist intelligence — see sample paper

Applications of AI

- Machine learning: Ability to make changes to improve itself upon getting new info.
- AI like SIRI, Alexa ... etc
- Robotics
- Image & voice recognition
- NLP
- self-driving cars

#Learn difference between AI and normal machine

(
- Both can work independtly but AI involves has no human involvement
)

Deeplearning:

- Subunit of machine learning
- Uses artificial neural network for predicting outcomes (Just like neural network of human brain)

Domains of AI

(
- Data science: Studying the types of data
- Computer vision
- NLP: — NLU — ability to understand language
  - NLG — Text planning
    - Sentence planning
    - Text realisation

Date: / /

Black box: Certain actions taken by AI that has no explanation

Types of learning

Reinforcement learning: Explores data then answers (like how humans are taught)

Supervised learning: * · AI knows questions & answers
· AI al system able to identify patterns the way we want it to.

Unsupervised learning: AI is only given input/question/data & no answers
· AI is allowed to find patterns in it and predict outcomes

Transfer learning: Still continues to explore more even after finding answer to input/question

Turing test: Checks if robots are indistinguishable or not ie can a robot/AI behave like human test

Components of hood AI policy

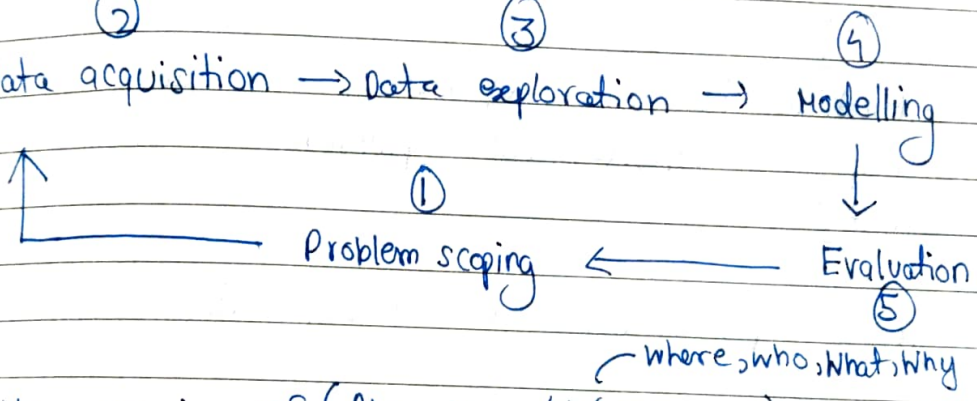
- Transparent
- AI must have right to collect data
- Consumers have choice to leave the system
- Purpose of data collected is limited
- Data must be deleted upon consumer's request

AI project cycle

IT project cycle : (linear)

Design → Develop → Test → Deploy

AI project cycle : (cyclic)



① Problem scoping: 4S (Also see 4W canvas) - Ans: How

- Identifying the problem (Stakeholders (victim), Ethical concerns)
  - then Imp # ↓
    - set goals (# see all sustainable developmental goals Pg No 141)
    - plan how to achieve goals

② Data acquisition

- Types of data
  - Structure / unstructured data / (Structural classification) — Semi-structured data
  - Numerical / text data (basic classification)
    - Discrete — ex: 4, 12, 15
    - continuous — ex: 9.1 kg, 4 m^2 s . . . etc

paper

Date:

· Big Data

(So big (metaphorically) that standard databases like oracle, SQL server, MS-Access are unable to capture, manage & process it

- Includes structure, unstructured, semi-structured data
- Varies from petabyte to zettabyte

It's importance

- Machine learning depends on big data
- helps us understand patterns made by AI

Three parameters of BIG Data

- 3v's

- Velocity: Big data mostly ~~contains high volume unstructured~~ contain data
- Volume: High volume, mostly unstructured data
- Velocity: Data can be received at high speed, because of Internet of things → IoT
- Variety: Contains lot of data (Even structured & semi) that require lot of processing

∴ Main objective in Data acquisition is to collect the required data

~~Data exploration~~:

Important milestones in Big Data

- Emergance of big data sets
  (Nice question)
- Framework that can manage BIG data are Hadoop, NoSQL

- IoT development

3) Data exploration:
- To find patterns and remove unwanted data

Types of Weak AI systems

- Heuristics/rule based:
  (AI functions on user based rules

- Brute force/Decision trees: Uses decision trees for analysing/finding different outcomes.

- Neural Networks: Mimics humain brain to some extent not fully (Deep learning (subset of machine learning functions through this)

Data Visualisation - Charts
                   (ppt's
                   (Graphs
  } Learn/Explain how they help in visual communication

Data visualisation tools (Can come in MCQ) or fill in the blanks)

- Excel
- Microsoft Power BI
- Google data studio
- Fusion charts

(4) Modelling

∴ Make models for the AI system for it to function

Modelling techniques

- Linear regression & logistic regression
- Decision tree
- Deep neural networks

} easiest to remember

AI learning system

Rule based / Heuristics
- limited by design

Learning based
- Supervised
- Unsupervised
- Transfer
- Reinforcement
  - has no limits, yet to be explored
  - Unlimited ability to stimulate true AI intelligence

Reinforcement learning:

- Positive reinforcement: on correct decisions
- Negative reinforcement: on incorrect decision

This way learner the most and assist more in producing outcomes

Drawback — takes / consume lot of time

Supervised learning — takes least time, accuracy is achieved but limited by design

5 Evaluation

- Check problems in the models, solve the problem

Helps in

- Enhancing quality of model
- " " " research
- * Allow correction of mistakes

6 Deployment

Projet is tested for the purpose it was made for

Natural Language Processing

- Domain of AI responsible for understanding and generating language

Functions

- Translations of language
- Grammarly, ~~Word~~ Microsoft word - for accuracy of texts

- Personal assistants like siri, Cortana ... etc

Processes involved in NLP

① Lexical Analysis (Text or paragraphs, sentences)
- Analysing structure of words / texts
- lexicon refers to collection of words
- Division of text into paragraphs, sentences, words (Main function)

↓

Syntactic Analysis (Parsing) - (Grammatical analysis) ~~make correct~~ grammatical ie find relationship between words

↓

Semantic Analysis - checks/ identifies meaning of text

↓

Discourse integration - Tries to integrate / join meaning of all the sentences together

↓

↓

Pragmantic analysis — message is re-interpreted ie AI system ries to understand it's meaning again as a whole

Applications of NLP

- Machine translation : Help interaction between machine & human
- Automatic summarisation : Summarises whole text to smaller text
- Text classification — organised text /date
- Virtual assistants — interact with human,
- Sentiment analysis — Can detect sentiment (human emotion by understanding text), (can get easily fooled

4W Canvas

4W's	Explanation/Description
1. What ?	— questions to be asked & answered
2. Where	
3. Why	
4. Who	
5. How	solution of problems

Chatbots → Turing test

- Bot that can talk / hold conversations with humans
- A software that 'pretends' to be human

Popular Chatbots & how they function

Elbot

- Fred Roberts - developer
- Holds conversations in entertaining way

Mitsuku

- Steve worswick
- Won Loebnar Prize
  - Given to most human human like chatbot,
- has ability to sense mood to some extent by analysing users writing style / tone.

Rose

- Bruce wilcox
- got Recognised the most human like chatbot in the first Alan Turing test
- won Loebnar prize in 2014 & 2015

Date: / /

Clever bot : · Web based chatbot
· learns from human input
· Does not use pre-programmed responses

Jabber wacky

· Rollo carpenter
· Aim was to create human like chatbot that can hold humorous conversations

Types of Chat bots

Scripted: · easy to make
· pre programmed responses
· Easily identifiable that not human

Smart Chat bots : Hard to make
· No pre-programmed responses
· Usually hard to identify if chatbot or human
· learns from input/interactions
· Can improve themselves

Human language vs Programming language

· Computers understand only through the language of numbers (programming language), cannot understand human language

- Different syntax (arrangement), same meaning, example

In Human language:
- The same word can also be noun, verb & adjective
- Humans can understand easily

In programming language
3+7 = 7+3
↳ Computer gives different results on diff version of python
- Hard for AI

- Same word multiple meanings

- Computers are sometimes unable to distinguish between ~~pork~~ nouns & verbs ie they interpret differently

#For example: She will park the car there so that we can play in the park

↳ The AI in computer must be able to differentiate both the meaning of park here

Perfect Syntax, No meaning

- Sentence has perfect syntax, but no meaning

↳ humans can understand that such sentences have no meaning but AI cannot

// - AI relies on identification & analysis of syntax, it does not understand the data/text itself

Data Processing

Text Normalisation

(Normalise text to lower level)

Corpus – whole documents / texts together

Steps of Text Normalisation

1) Sentence Tokenisation / segmentation

- Breaking bigger sentences to smaller sentence

2) work tokenisation / segmentation

- Sentences are broken down to words / segmented (for anglysing grammar roles)

3) Text Lemmatisation & Stemming

Stemming: • Remove prefixes/suffixes of words to find original form of word/base form.

• Not always successful

ex: Troub ling → Troubl (Ing is removed) → Unsuccessful

Lemmatisation • Remover suffixes, and obtains proper base word from dictionary.

for example: Troubled → Trouble

lemma – Base word

~~Stopw~~ · Removing stopwords, special characters & numbers

· Stopwords – a, an, the, as, for, and.

· Converting text to common case like (world, WORLD)

Bag of words

· Computers don't understand raw data/input, it needs to be processed ie converted in vectors of numbers

Bag of Data Algorithm

· ~~Text~~ normalisation
· Create Dictionary (~~vector~~): list out unique words
· Create document vectors: find frequency of each word
· ~~Cr~~ · Create document vectors for all documents.

Date: / /

Evaluation

- Final checking of AI model before deploying

Model Evaluation Terminologies

- In order to understand effectiveness of model, it is necessary if prediction are correct or not. To test it

Prediction	Output	
Yes	Yes	→ True paritive
No	No	→ True negetive
Yes	No	→ False paritive
No	Yes	→ False negative

Date: / /

Confusion matrise / errormatrise

- Table used to define performance effectiveness of model by analysing set of test data.

Matrise		Reality	
		Yes	NO
Prediction	Yes	Trus positive	False positive
	NO	False negative	True negative

Evaluation methods

1) $\text{Accuracy} = \left(\frac{\text{Correct prediction} \times 100}{\text{Total cases}}\right)\%$

$= \left(\frac{TP + TN}{TP + TN + FN + FP} \times 100\right)\%$

2) $\text{Precision} = \left(\frac{\text{True positives}}{\text{All positive predictions}} \times 100\right)\%$

$= \left(\frac{TP}{(TP + FP)} \times 100\right)\%$

3) $\text{Recall} = \frac{TP}{TP + FN} = \left(\frac{\text{True Positives}}{\text{True positives} + \text{False negatives}} \times 100\right)\%$

$F_1 \text{ score} = 2 \times \left(\frac{\text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}\right)$