WINTER SCHOOL 2023

DAY 1

Inaugural Session

The Winter School inauguration session started at 10.00 am with Director's welcome address followed by Inaugural ceremony by Dr. Elizabeth Sherly, distinguished Professor at Digital University of Kerala.

After the inaugural session, the keynote address on Natural Language Processing was taken by Dr. Elizabeth Sherly. She started the session by mentioning the lineage of 4 waves in NLP. She briefly explained the first wave with the concept of rationalism started in the 1950s where Alan Turing proposed the Turing test to evaluate a computer's ability to exhibit intelligent behavior indistinguishable from that of a human.The test is based on natural language conversations between a human and a computer designed to generate human-like responses.

She elaborated the first wave which started in the period 1960s to 1980s Rationalism:Rule based NLP system where efforts started to extract knowledge of language in human mind.The name coined due to wide acceptance of Naum Chomsky, the father of modern linguistics, for applying innate language structure and postulating the genetic inheritance, hardwired in the brain at birth that to endeavor to design hand-crafted as rule based NLP system.A rationalist approach design handcrafted rules and reasoning mechanism in to intelligent NLP systems.

Notable conversational systems like ELIZA(First conversational chat) and some expert systems like MYCIN(Medical diagnosis chat) based on a complex set of handcrafted rules, but confined to narrow domain problems were discussed.

She then mentioned that Angla Bharti -1 is primarily a rule based system & Anu Bharathi - II uses EBMT as the basic paradigm for translation. Both of these systems are hybridized with varying degrees of hybridisation of different paradigms.Angla MT system , a rule based machine translation system is for translating english to 8 Indian languages including Hindi,Malayalam,Bengali, Urdu, Punjabi, Tamil, Assamese,Nepali)

She concluded the First wave by mentioning its limitations like how these rule based translation systems require manual development of linguistic rules, which can be costly & which often do not generalize to other languages.Require linguistic expertise to develop systems and hence they play vital roles in computational linguistics.

Then she discussed various Machine Translation Systems like:

1.KBMT-Knowledge based Machine translation which concentrated on the development of knowledge intensive morphological syntactic and semantic information for lexicon.

2.EBMT-Example based machine translation road coverage of many different linguistic phenomena, but lacks deeper knowledge about the translation domain.

3.SMT-Statistics-based machine translation which treats translation as a machine learning problem. This means that we apply a learning algorithm to a large body of previously translated text, known variously as

- Word based
- Phrase based
- Syntax based

4.Machine Learning based- where artificial intelligence to automatically translate text from one language to another without human involvement or unsupervised.

Then she started discussing the Second wave-Empiricism which started in the early 1980s,the MT system, where speech recognition systems became popular and there is the lack of abilities to learn from data and uncertainty in reasoning.

Empiricism was characterized by exploration of data corpora(textual data) and machine learning.

The statistical MT revolutionized NLP applications in a more meaningful manner and the approach developed during the era(1990).

Later she explained the Statistical MTS which is reintroduced in 1980's and early 1990s.Started with (Hidden Markov Model) HMM with morphological analysis.Since late 1990s, discriminative models with statistical Machine translation systems include maximum entropy mode, support vector machines, expectation maximization,conditional random fields etc was a breakthrough with promising results.In the empirical era, again core NLP applications of machine translation with bilingual training of data by understanding surface level translation knowledge and spoken language understanding using Neural Network models.

Then Popular SMT Systems were discussed.She talked about Anagalabharti (1991) : First of its kind development by IIT Kanpur aided translation project on translation from English to Indian languages.

Pseudo target is generated which is applicable to a group of Indian languages.Set of rules acquired through corpus analysis to identify the plausible constituents.Uses transfer and the interlingual approach.

The following Translations tools were also discussed.

Mantra(1990) is a translation tool devised for English to Hindi in a precise domain. Used Tree Adjoining Grammar (TAG) technique.

Ambusaraka project (1995-2000) built to transfer sentences from Telugu, Kannada, Bengali,Punjabi and Marati to Hindi.

Anuvadak is an English to Hindi translation tool used in domains such as official,formal, agricultural and linguistics.

Shiva & Shakti machine translation system (2005)

Two systems that translate from English to Hindi

Siva-example based MTS and another follows a hybrid approach by combining rule and statistical approach.

Indian Language MTS (2006)(Sampark)(ILMT)

Is a Translation system for Indian Languages and it can translate any language to any language and is a bidirectional system for a pair of languages.

Indian Language Corpora Initiative(2009) (ICLI) headed by JNU New Delhi under TDIL of Deity,GOI to build a common corpora language in 17 Indian languages.

Then she explained the Third wave -Neural Machine Translation which uses encoder-decoder model paradigm

Then she briefly explained about the word Order & Word Alignment of Malayalam language.

She said that malayalam is a highly agglutinative and inflectional language

Main difference in the order.English & Malayalam is Subject-Verb-Object(SVO) where as Malayalam is Subject-Object-Verb (SVO)

eg. John(Subject) Wrote(Verb) Book(Object)

John (Subject) oru book(Object) ezhuthi(Verb)

Word Alignment aims to identify the correspondence in parallel sentences.It is a vital component of statistical machine translation (SMT) systems.Word alignments were thus introduced as a set of hidden variables.

a=a1,a2.....am.

eg: He was born on 4 december 1970

അവൻ 4 ഡിസംബർ 1970 ജനിച്ചു

A= {(1,1) (2,5) (3,6) (4,7) (5,2)}

Then she explained the Neural Machine Translation. She told that in the age of deep learning, NMT has almost entirely supplanted statistical machine translation word alignment as an explicit task is no longer necessary.Encoder - decoder model paradigm commonly used in present day NMT.In NMT, concept of alignment motivated the development of the attention mechanism.

Most prominently underlies the transformer the current neural approach to NLP problems.

Then she discussed the deep Natural Language Processing .Deep learning algorithms attempt to learn (multiple levels) of representation and an output.

From 'raw' inputs,combine ideas & goals of NLP & use representation on learning & deep learning methods to solve them and several big improvements in recent years in levels: speech, morphology,syntax,semantics and applications like machine translation, sentiment analysis and question answering.

Then she explained about the various limitations in the field.

Like the Black box like machine learning algorithms.Requirement of large amounts of data.Lack of common sense.Lacks understanding about underlying details about input data.Less powerful, beyond classification and lack of global generalizations and also requires more combinations of different technologies.

According to her, in the fourth wave -future wave,rules are devised by the experts as did in the rationalist era with deep learning may provide better interpretation and with explainable AI and the intersentential relationship can be made stronger. Explainable artificial intelligence (AI) is a set of processes & methods that allows human users to comprehend & trust the results output created by Machine learning algorithm.

Then she discussed the Brain Computer Interface (BCI) which is a future wave technology which can promise better modeling of memory & exploitation of knowledge for text & speech understanding.It is the technology that enables computers to be controlled directly with brain activity. A person with paralysis can control a computer just by thinking to do an action.She explained the brain computer interface process with a typical BCI architecture.Finally she

explained the term cognition and the TOP 10 hot architecture and concluded the session by clearing the doubts of participants.

Afternoon Session: 2:00 pm

INTRODUCTION TO NATURAL LANGUAGE PROCESSING (NLP)

By Dr. RAJEEV R. R. (Project Head (e-Gov & Consultancy),ICFOSS)

The session started at 2:00 pm. Dr. Rajeev R.R. started the session by mentioning that **understanding of the research problem and language features** is essential before conducting a research in NLP. He then discussed the difficulties in creating a chat gpt-like AI model for malayalam language because of the structure, ambiguity and context of Malayalam language. He also discussed the peculiarities of the family of languages, stating Dravidian languages and concluded that the transliteration is difficult even among different languages of the same family.

Dr.Rajeev handled an interactive session on ‘the significance of NLP’ mentioning first language acquisition, environment dependency and its relation to an individual’s thinking ability and made the session more interesting by citing the story of Birbal. He then discussed the significance of malayalam language computing specifying the agglutinative nature in word formation and difficult character formats. He quoted the works of R.E. Asher, Herman Gundert, A.R. Rajaraja Varma and Dr.Ravishankar while mentioning the scarcity of works in Malayalam syntax and semantics.

The session included description of definitions of basic terms -natural language, language technology, natural language processing and computational linguistics. Dr. Rajeev pointed out Rule-based NLP and Unstructured NLP citing examples and explained Human aided Machine Translation, Hybrid Machine translation and prediction based on language models like Hidden Markov Models. He made the audience ponder over the challenge of how to perform natural language processing without compromising the ‘naturalness’.He referenced the ASCII code for English and unicode for malayalam language during the lecture.

The session began after the evening tea-break at 4.00 p.m. with the topic ‘where does NLP fit in the CS taxonomy’. The session had a discussion on topics - relation of language technology with

multimedia & multimodality technologies, knowledge technologies, speech technologies and text technology. Dr. Rajeev briefed the topics- computational linguistics versus natural language processing and the relation of computational linguistics to other disciplines like Machine learning, Human Computer Interaction, Information Retrieval, Theory of Computation, Psychology, Philosophy of Language, Linguistics, Electrical Engineering (Optical Character Recognition) etc. . He explained about the linguistic levels of analysis pertaining to speech, written languages and gestures. He explained from top to bottom about phonology, morphology, syntax and semantics of written language. He minutely detailed the 'morphophonemic change' and 'morphosyntactic change' of malayalam language. The intriguing session ended at 5.20 p.m. with a discussion on morphological analyzer, number markers and gender neutrality of words in malayalam language with Dr. Rajeev referencing his publication work on morphological analyzer.

DAY 2

Resource person: Dr. Satheesh Kumar

Session:Forenoon

Time : Morning 10:00 to 12:00

Started with a motivation class by Research Associate Meharuniza. She took class about 2 -min rule,POMODORO Technique and EISENHOWER MATRIX

TOKENIZATION:

Tokenization is the process of replacing sensitive data with unique identification symbols that retain all the essential information about the data without compromising its security.

Examples of tokenization:

Tokenization technology can, in theory, be used with sensitive data of all kinds, including bank transactions, medical records, criminal records, vehicle driver information, loan applications, stock trading and voter registrationMost of the popular Natural Language Processing (NLP) libraries have their own sentence tokenizers. Some of them take a rule-based approach while

others take a neural network-based approach. The former identify exceptions based on rules. Abbreviations or quoted text are a major exception. In a neural network-based approach, the model can be a simple Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), etc. Packages such as NLTK and spaCy take a rule-based/algorithmic approach. FlairNLP doesn't have its own tokenizer but has an integrated *segtok* package, which is a rule-based tokenizer. StanfordNLP has a neural pipeline with LSTM as a tokenizer. .

Malayalam Sentence Tokenization Problem:

- .Malayalam abbreviations are typically lengthier than English ones. For example: B.Sc is a three-character abbreviation for Bachelor of Science but the same abbreviation in Malayalam (ബി.എസ് സി) has seven characters.
- Some Indic languages have their own punctuation to end the sentence. For example: Hindi has पूर्ण विराम (।).
- Older Malayalam texts may not carry any end punctuation at all.

Search for a Tokenization solution:

Exploring the possibility of adding Malayalam sentence segmentation support for NLP libraries. It was a multi-step process. More on it below.Data collection and preparation:

It takes lots of data to improve any NLP package.

Stages of NLP:

1.Moprphological Analysis:

Morphological Analysis is the initial step in NLP. It entails recognizing and analyzing word structures.It is the process of reaking down a text file into paragraphs, phrases, and words.It refers to the study of text at the level of individual words. It searches for morphemes, which are the smallest units of a word.

2.Syntax analysis or parsing:

Syntactic or Syntax analysis is a technique for checking grammar, arranging words, and displaying relationships between them. It entails examining the syntax of the words in the phrase and arranging them in a way that demonstrates the relationship between them. Syntax analysis guarantees that the structure of a particular piece of text is proper. It tries to parse the sentence in order to ensure that the grammar is correct at the sentence level. A syntax analyzer assigns POS tags based on the sentence structure given the probable POS created in the preceding stage.

3.Semantic analysis:

Semantic analysis is the process of looking for meaning in a statement. It concentrates mostly on the literal meaning of words, phrases, and sentences is the main focus. It also deals with putting words together to form sentences. It extracts the text's exact meaning or dictionary definition

4.Discourse Integration:

The term "discourse integration" refers to a feeling of context. The meaning of any sentence is determined by the meaning of the sentence immediately preceding it. In addition, it establishes the meaning of the sentence that follows .The sentences that come before it play a role in discourse integration

5.Pragmatic analysis:

The fifth and final phase of NLP is pragmatic analysis. The overall communicative and social content, as well as its impact on interpretation, are the focus of pragmatic analysis. Pragmatic Analysis uses a set of rules that describe cooperative dialogues to help you find the intended result.

There are 2 type of morphological analysis-inflectional and word formation. Word formation divides into derivation and compounding.Inflectional morphemes are those that serve a grammatical function, such as the plural -s or the past tense -ed. Derivational morphemes operate more directly on the meaning of a word.Derivational morphology again divides into class maintaining and class changing.Compounding divides into compound nouns, compound verbs,compound adjectives.

REVOLUTION IN THE FIELD OF MORPHOLOGICAL PARSING:

It is the process of determining the morphemes from which a given word is constructed. It must be able to distinguish between orthographic rules and morphological rules.

1.Cut and paste method

2.Finite stage and technology:

It is used for automation recognition and generation of words form.if a malayalam word has 234 inflections we can draw in finit stage.it is used in morphology

3.Two level model:

It is the first practical system incorporating the idea of finite state transducer.

MALAYALAM MORPHOLOGICAL PARSING:

Malayalam is a Drravidian language. It is so agglutinativeach root word can combine with multple morphemes. It does not show subject agreement in its verbal inflection. It is verb final relatively free word order and morphologivally rich language.

SANDHI RULES:

Sandhi is classified into 4

1.a letter is lost or dropped (ലോപം)

2. a letter is added (ആഗമം)

3.a letter is doubled (ദിത്വം)

4.a letter is substituted or transformed (ആദേശം)

PARADIGM APPROACH:

It is practised by ANUSARIKA GROUP OF RESEARCHERS to build an aid to indian language to HINDI AND VICEVERSA. NOUN PARADIGMS:32 types VERB:34 types e.g adi -adichu adi is one paradigm

POS TAGGING:

Tagging is a kind of classification that may be defined as the automatic assignment of description to the tokens. Here the descriptor is called tag, which may represent one of the part-of-speech, semantic information and so on. Part-of-Speech (PoS) tagging, then it may be defined as the process of assigning one of the parts of speech to the given word. It is generally called POS tagging. In simple words, we can say that POS tagging is a task of labelling each word in a sentence with its appropriate part of speech. We already know that parts of speech include nouns, verb, adverbs, adjectives, pronouns, conjunction and their sub-categories.

Most of the POS tagging falls under Rule Base POS tagging, Stochastic POS tagging and Transformation based tagging.

AFTER NOON:

TIME: 2.00 PM TO 5.00 PM

BASIC STATISTICAL CONCEPT FOR DATASCIENCE:STATISTICAL PROOF

Data Science is emerging now a days. Using linear regression we can solve maximum problems.

While starting data science,

1. We know the concept.
2. We shoud need a software to calculate
3. We how to interpret the result

Everything is data. How to get data?

1. Surveys and Questionnaires.
2. Interviews.
3. Observations.
4. Records and Documents.

5 Focus Groups.

DATASCIENCE:

C.f.jeff wu coined the term datascience'.He advocated that statistics be renamed data science and statisticians as data scientists. It is hot new field to development. There are qualitative and quatitative data in data science we convert qualitaive to quantitative

DATA ANALYTICS: DERIVING INFORMATION FROM DATA

e.g. wisdom-knowledge-information-data

Data should be correct. If you are giving a bad data you wont predict good result. If there is a null data we should fill with average data.

SURVEY RESULT:

Data is never clean. you will spent more time in data cleaning.

BIG DATA:

The definition of big data is data that contains greater variety, arriving in increasing volumes and with more velocity.

DATA MINING(KNOWLEDGE DISCOVERY IN DATABASES)

It is also known as Knowledge discovary. Data cleaning is the most important part -industry will take about 99 % time for cleaning the data. It means mine information out of data. Average of data is very important in data analysis. Gather information from data(large data).

APPLICATIONS

1. BANKING

2.CUSTOMER RELATIONSHIP MANAGEMENT

3.TARGETED MARKETING

4.FRAUD DETECTION

5.MANUFACTURING AND PRODUCTION

6.MEDICINE

7.MOLECULAR/PHARMACEUTICAL

8.SCIENTIFIC DATA ANALYSIS

9.WEBSITE/STORE DESIGN AND PROMOTION

Some general observation of data science

RELATIONSHIP WITH OTHER FIELD:

It overlaps with ML, statistics some basic operations in data mining are predictive:- regression,classification,collaberative filtering.

Some basic operation :regression-- depends upon matters, classification - catagorise-classify

Predictive analytics

1 Bayesian classifierstatistics - If it predict correctly it is called deterministic system.If we cannot predict outcome it is called stochastic system. It perform probability prediction. Its "attributes are independent of the class.

Conditional probability

Conditional probability is known as the possibility of an event or outcome happening, based on the existence of a previous event or outcome. It is calculated by multiplying the probability of the preceding event by the renewed probability of the succeeding, or conditional,

$$P(A|B) = N(A \cap B)/N(B)$$

Or

$$P(B|A) = N(A \cap B)/N(A)$$

Conditional Probability and Bayes Theorem

Bayes' theorem defines the probability of occurrence of an event associated with any condition. It is considered for the case of conditional probability. Also, this is known as the formula for the likelihood of "causes".

$$P(A|B) = P(B|A)\ P(A)/P(B)$$

Day 3

Resource Person: Dr Umesh P

Session : Forenoon

Time: Morning 10:00 to 11:00

Started with Data model and overall view of machine Learning and relevance of math in Machine Learning.

Shows the workflow of ML model and predictive model.

Difference between Supervised and unsupervised learning.

Different Machine Learning Algorithms

- Un-supervised
- Supervised
- Clustering
- Regression
- Reinforcement Learning

Discussed the relevance of above mentioned algorithms according to the problem.

Various areas these are applicable

Why Math?

- How to identify the right algorithm?,choose the right parameters and validation strategies.
- How to identify overfitting or underfitting?

ML is written in Mathematics

Understand the algorithm in a better way math is required

Data Collection,Represent,Prepare,Model,Search,Visualize

Branches of math relevant in Machine Learning

- Linear algebra,calculus,probability,graph theory

LINEAR ALGEBRA

Understanding through graph using a straight line and line equation.

Linear Algebra is a particular kind of geometry which connects two values.

- **Eg- y=2x+3**

3 is the intercept which meets the y axis, how it defines, substitute the values for x and predict y value. Y is independent and x is the dependent variable.

Domain

What all values can take 'x'.

Example:-

When two equations get the values of variables.So each situation converts to equations which are similar,sometimes the situations are not matched and the two lines are not intersecting.

Explain n-dimensional equations._When it comes to linear algebra the equations are n dimensional.

aA1x1+a2x2+.....+anxn = Bn

We can solve the equation by

- Situations -> equations –> Matrix
- Row reduction can be done and can solve the equation

Tensor = multidimensional array

Vector matrix tensor

How algebra and geometry is connected

- **Norm of a vector - A vector norm is a function that maps a vector to a scalar value**

Given n vectors u1,u2,u3,....un from Cm and n scalars a1,a2,a3,..,an their linear combination is the vector.

a1u1+a2u2+a3u3=,....,anun

Span of vectors a*[1 +b*[0

O] 1]

They are called basis.

Lineraly independent sets

Given a set of vectors, you can determine if they are linearly independent by writing the vectors as the columns of the matrix A, and solving Ax = 0

Linear combinations: multiplying matrices by scalars, and by adding them together.

Distance measures :-

-Euclidean

-Minikowski Distance

-Standardized Euclidean distance

-Cosine Similarity

Projection of vector

Projection of x on y= ||x||.||y||.cosq

||x||

Orthogonal projection

-Can u develop equations by connecting data points on a line?

the rejection of a vector from a plane is its orthogonal projection on a straight line which is orthogonal to that plane. Both are vectors. The first is parallel to the plane, the second is orthogonal.

- Hyper plane ,(W)^T X =0

In every plane of n dimension the hyperplane has n-1 dimension

For 2d plane it has 1 d hyper plane

For 3d plane it has 2d hyper plane

- Line ,y = ax+b

Transformation matrix

- A matrix that transforms one vector into another vector by the process of matrix multiplication.
- We can rotate a matrix by changing it with angle
- Can change the position by multiplying it with vectors
- Determinant determines the area of the vector
- Matrices is responsible for the transformation

ax=lamda(x),were lambda is the eigen value

Eigen value - a scalar associated with a linear set of equations which, when multiplied by a nonzero vector, equals to the vector obtained by transformation operating on the vector.it is the no of elongation of vectors for matrix transformation.

Eigenvectors - special vectors that maintain their direction despite bullying by the matrices (in mathematics, we call this a matrix transformation). These vectors elongate or shrink/squish depending on the intensity of bullying (transformation) by a factor of eigenvalue.

Singular value decomposition

- SVD is a technique to decompose a matrix into several component matrices, exposing many of the useful and interesting properties of the original matrix.
- Square root of eigen values is singular values

Singular value decomposition is done not only for square matrix but also for other matrix

- Eigenvalues of a Matrix and its Transpose are the Samedpdks

Step 1 Find A transpose *A which makes it square matrix

Step 2 finding eigen values and its singular value

Step 3 finding eigen vectors

Step 4 normalizing it

PCA

PCA is the main area where eigen values are very useful.PCA is used for dimensionality reduction which can be done through different steps mentioned above by singular value decomposition.

Take covariance of each element and then eigen vectors

Larger eigenvalue ⇒ more important eigen vectors

Evaluating an ML problem through google colab

- Distances Eu Manhattan etc.ipynb

Solving an eigen value problem

|A-lambda (I) |=0

Solving for eigen vector

Substituting the values of lambda in equation

A-lambda (I)* X=0 , then solve for x

Calculus

Slopes gives the rate of change .Slope decides the growth rate ,greater the slope greater the growth rate.

Slope = m = (y2 -y1)/(x2-x1)

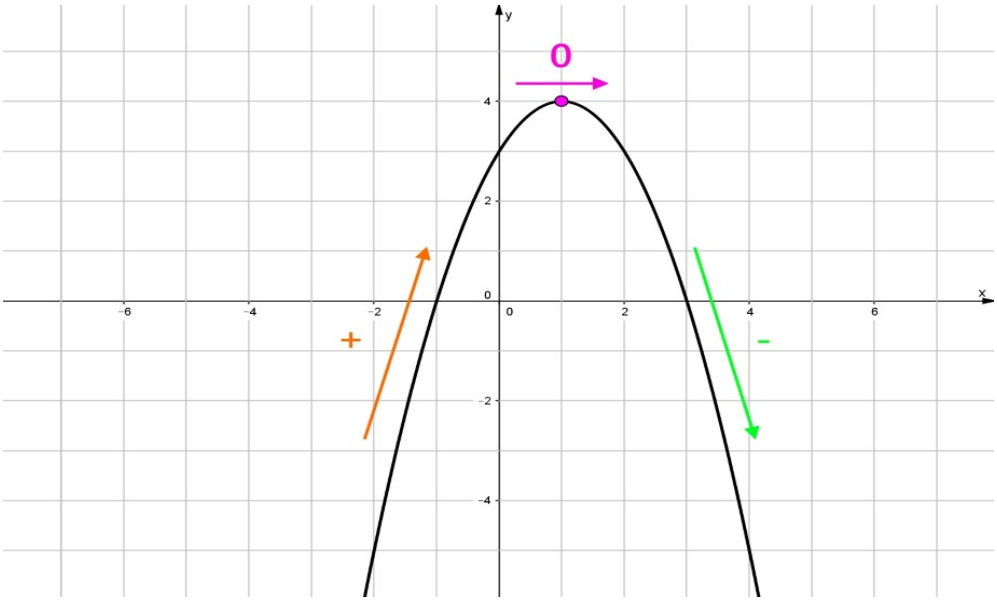
d/dx is the notation generally used to represent change with respect to x.

Derivative and slope

Positive slope

Slope = 0

Negative slope



Slope =dy/dx

Higher order derivatives

- Higher-order derivative means the derivatives other than the first derivative
- Repeated process of taking derivatives of derivatives

Partial Derivatives

If function is a surface we have to find rate of change of a certain feature only remaining feature have to be fixed constant.

Taylor Series

-It is a series that is used to create an estimate of what a function looks like.

-Expands to power series

$$
\begin{aligned}
e^x &= f(0)\frac{x^0}{0!} + f'(0)\frac{x^1}{1!} + f''(0)\frac{x^2}{2!} + f'''(0)\frac{x^3}{3!} + f^{(4)}(0)\frac{x^4}{4!} + \cdots \\
&= \frac{x^0}{0!} + \frac{x^1}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \cdots \\
&= \sum_{n=0}^{\infty} \frac{x^n}{n!}
\end{aligned}
$$

Derivative and Gradient

In one variable functions you only need the derivative to know how a function is changing. In multivariable functions there are multiple directions that a function can change in and therefore it is no longer need to have only 1 derivative. But you can define a new idea, the gradient to describe how the function changes in multiple dimensions. We can think of a gradient as a kind of multidimensional derivative.

Chain rule in partial derivatives

- partial derivatives with respect to all the independent variables

Simple chain rule. If u = u(x, y) and the two independent variables x and y are each a function of just one. other variable t so that x = x(t) and y = y(t), then to find du/dt we write down the. differential of u.

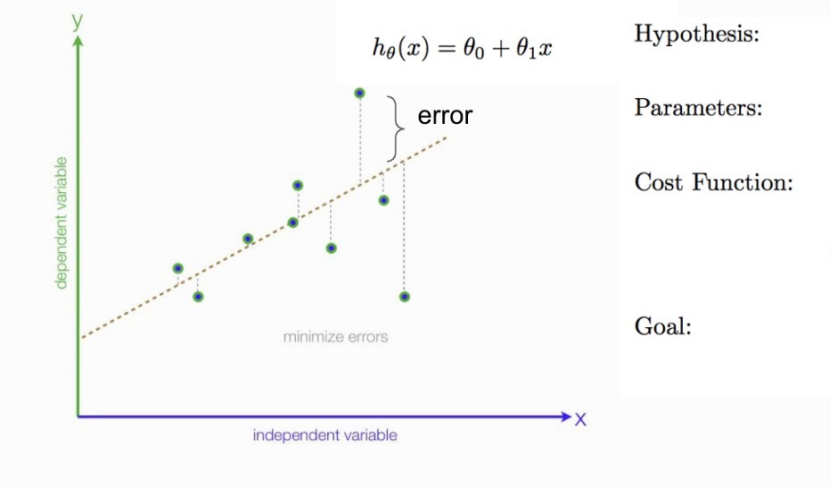
Eg:-

U = f(x,y) = x^2 y

x = t^2, y = t^3

Gradient descent

It is used to find the local minima and local maximum in a function



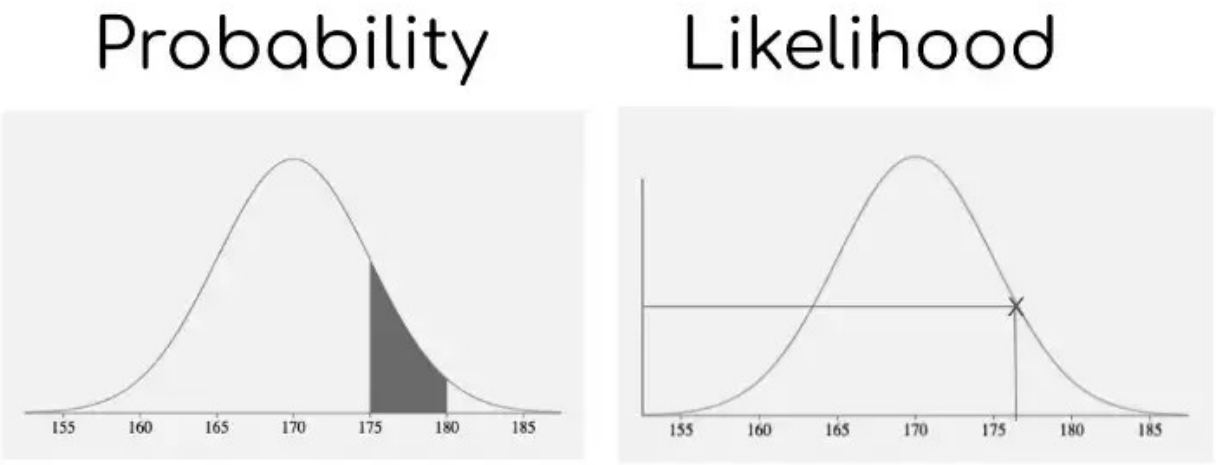
Random Variable

- A random variable is a variable whose value is determined by the outcome of a random event.
- The values of a random variable can be continuous or discrete, and the distribution of a random variable can be described by its probability distribution function.

Conditional Probability

- Conditional probability is the probability of an event occurring given that another event has already occurred.
- It is a measure of the likelihood of an event happening, based on the knowledge of some additional information

-



- Probability is a measure of the likelihood of an event occurring, expressed as a number between 0 and 1, where 0 indicates an impossible event and 1 indicates a certain event.

- Likelihood, on the other hand, is a function that describes the relationship between the parameters of a model and the data. It is used to assess how well a particular set of parameters fits the observed data.
- probability describes the uncertainty of an event happening, while likelihood describes how well a model fits the data.

Day-4

Resource Person: Dr Pattabhi R K Rao(Au-KBC)

Session : Forenoon

Time: Morning 10:00 to 11:00

Subject:NMT(Neural Machine Translation)

Statistical machine translation

- Build using statistical translation models
- Statistical translation models building is a quick process
- Mainly 3 components
- * word alignment

- * phrase table creation
- * Target language monolingual language model

Factored SMT

- Surface
- Lemma
- POS
- Morph

Evolution of NMT

- S2S Model
- S2S Model with LSTM
- S2S Model with GRU

S2S model

- The model is composed if an encoder and a decoder

Advantages

- Lack of latent alignment segmentation and sparsity associated with them

Only ReLu is used at activation level

LSTM

- Cell state and hidden state are there
- Cell state is used for memory
- Gated Recurrent units
- Computationally easier than LSTM
- NO output gate
- GRU:2 GATES, 3 LSTM gates
- Basics of neural network

Neural network is a set of connected input/output units,where each connection has a weight associated with it.

Here x1 & x2 are normalized attribute value of data, y is the output of the neuron ,i.e the class label. Weights are multiplied with attribute value. The neuron receives and output function is y=f(x.w+b), b-bias

Activation function

- Mathematical function which map whole predicted values to probabilities.
  Sigmoid function
  Hyperbolic tangent function
  Rectified linear units(ReLu)

Cost function

- These function are used to optimize the output
- Either min or max the output function
- Gradient descent
  - Computes MSE
  - Indication of MSE by adjusting of weights

Feed forward neural network
Acyclic
Weights are not adjusted.

RNN
Output from the previous stage is used as input.
It is cyclic.

LSTM
GRU
Combining input and cell state to update gate
Encoder and decoder architecture

SMT is a machine translation paradigm where translations are generated based on statistical models, whose parameters are derived from the analysis of bilingual text corpora (text bodies) – a source text of translated material and a target text of untranslated material.

Statistical machine translation starts with a very large data set of approved previous translations. This is known as a corpus (corpora is plural) of texts that is then used to automatically deduce a statistical model of translation. This model is then applied to untranslated target texts to make a probability-driven match to suggest a reasonable translation.

A bilingual corpus

A collection of parallel texts, used to train the translation model. If we had an endlessly huge corpus, translating a sentence would just be a lookup task. However, our resources are limited, and the majority of the sentences we will be required to translate will be unfamiliar. They will, however, be made up of terms that we have seen previously (even if some phrases are as short as one word). For example, “in this exercise we shall,” “size of the state space.

NMT

- Neural Machine Translation (NMT) is an approach to automated translation that uses machine learning to translate text from one language into another.
- A division of computational linguistics, NMT relies on artificial neural networks, themselves modeled on the human brain, to predict the likelihood of certain sequences of words. The MT algorithm is an example of deep learning, users can train NMT engines to recognise source and target connections using large datasets.
- As connections between words are strengthened or weakened through training on the datasets, the machine observes these correlations and adapts to predict and increase the likelihood of correct translations.

How NMT Works

- NMT functions by predicting how likely certain word sequences are, based on patterns observed in the datasets used for its training.

- A key mechanism is vector representation. In NMT, words are transcribed into vectors, each with a unique magnitude and direction, in a process of encoding and decoding.
- The engine analyzes the source text input, encodes it into vectors, then decodes it into target text by predicting the likely correct translation.
- Although it still incorporates probabilistic models, NMT differs from other statistical-based MT by assessing input as a whole, rather than separating it into sub-components.
- It can thus identify connections between words, which are strengthened or weakened depending on which words occur together.
- As a result, a trained NMT engine could ultimately make choices based on the strength of this contextual information to produce accurate, quality outputs.
- NMT engines rely on training materials to learn to produce quality output.
- It follows that the performance of the engine depends on the quality of the datasets used for training.
- This in turn raises questions about the quality of MT output, an issue that is particularly salient for medical and pharmaceutical translations where patient safety is of utmost importance.
- We look at quality standards in NMT more in-depth in the section below.

Challenges in NMT

- Low resource Language
- Morphologically Rich Language
- Technique to mitigate the low resource problem in NMT
    - Increasing the data using Back Translation
    - Phrase table injection- combining SMT and NMT
    - Leveraging the pretrained model - bert,glove,Robert
    - Combining the corpus

Break the word forms into subwords,so that the overall vocabulary size is reduced.
Training with linguistic features such as lemma-tag strategy.

Day 5: Morning Session

Time: 10.00 am to 1 PM

Resource Person 1: Dr.Sobha Lalitha Devi

Designation: Program Director,AUKBC Research Centre,Chennai

Resource Person 2 : Dr. T Pattabhi R K Rao

Designation: Project Scientist, AUKBC Research Centre, Chennai

Resource Person 3 : Dr.R Vijay Sundar Ram

Designation: Project Scientist,AUKBC Research Centre, Chennai

BLEU Matrix

Bilingual evaluation understanding,for checking the similarity of text input and text output.It can be get from scarablue library.

Meteor

Mathematical evaluation of machine translation

Dr Sobha Started the session at 11.30 am on “Shallow Parsing for NMT”.

-Pos tagger is a disambiguiter of morph analyser

- contextual grammatical information
- Anaphora Resolution
- Named Entity Identifier
- Word sense Disambiguation
- Muiltiword Identifier
- The semantic role identifier

Tokenizing is the process of tokenizing the text, white spaces are there.

- Algorithms- byte pair encoding
    - Zipping algorithm

Parsing is classified into 2 types .They are Shallow parsing and Full Parsing.

Shallow parsing

- Take only certain structures
- It is not tree like structure
- Consist of pos,chunking,clause boundary

Deep parsing

Deep parsing will give a complete syntactic structure to a sentence

Shallow parser architecture

Part of speech tagging

- English tagset available: 45 tag penn tree bank tagset,61 tag c5 tagset,146 tag c7 tagsetbetc
- Malayalam tagset:BIS tagset

Named entity recognition

NER refers to the recognition and classification of proper names in texts and their classification into predefined categories.

Event is not in NER

Things with vector and direction are not NER

Rigid Designators: It designates the same thing in all possible worlds in which the same thing exists and doesnot designate anything else in those possible worlds in which that same thing does not exist.

Designations are not in NER theoretically but it depends practically.

Day: 6

Project - ICFOSS Team

Project No.	Project Name	Mentor
1	NER identification	Dr. Prajisha, Selvaraj, Jibin
2	Malayalam- English Machine Translation NMT	Arun, Alaka, Meharuniza
3	Tamil-Malayalam Machine Translation	Navaneeth, Shinu, Abisha
4	Optical character recognition in any languages	Sabeerali
5	Hindi-Malayalam Machine Translation	Abhijith Balan
6	English-Malayalam Machine Translation	Arun, Alaka, Meharuniza

Day: 7

Resource person: Dr. Gopakumar G

Session: Forenoon

Topic : Data Analysis Using Neural Network

Different types of Data Representation

Table Data

Text Data

Image Date

Signal Data

Sequence Data

Graph Data

Data generation in Machine Learning Model

Data generated by some output of some experiments

Noise

Noise similarly refers to unwanted behaviors within the data that provide a low signal-to-noise ratio. Essentially, data = signal + noise. Most can be prevented by understanding its causes and correcting them.
Noise affects the accuracy of data that we extracted.

Population and Samples

Population refers to the entire group of individuals about whom you wish to draw conclusions.
The sample refers to the group of people from which you will be collecting data.

Generalization

Generalization is the extension of sample-based research results to the whole population; also it implies that such study's findings may be applied to another similar predicament. It is important because as in above mentioned example the researcher uses individuals as a sample from the whole population.

Two Phases Of Machine Learning Model

Building the Model*
Evaluation of Model*

Building an ML Model requires splitting of data into two sets, such as ‘training set’, ‘testing set’ and ‘validation set’.

Training Dataset

Training data is used to teach prediction models that use machine learning algorithms how to extract features that are relevant to specific goals. For supervised models, the training data is labeled. The data used to train unsupervised models is not labeled.

Training data sets fine tunes the best parameter. It may be complemented by subsequent sets of data called validation and testing sets.

Test Set

A Test set in machine learning is a secondary data set that is used to test a machine learning program after it has been trained on an initial training data set. “Data set is Plotted in the vector form and each represent in the form of samples and n dimensions”.

Vector Dataset

Vector data is what most people think of when they consider spatial data. Data in this format consists of points, lines. individual points stored as coordinate pairs that indicate a physical location in the world.

Supervised Learning

Supervised learning is the type of machine learning in which machines are trained using well "labeled" training data, and on the basis of that data, machines predict the output. The labeled data means some input data is already tagged with the correct output.

Unsupervised Learning

Unsupervised learning is a type of machine learning in which models are trained using an unlabeled dataset and are allowed to act on that data without any supervision.

Reinforcement Learning

Reinforcement learning is a machine learning training method based on rewarding desired behaviors and/or punishing undesired ones. In general, a reinforcement learning agent is able to perceive and interpret its environment, take actions and learn through trial and error.

“Agent” is an independent entity that interacts with its environment and learn from the feedbacks.

Regression

A regression problem is when the output variable is a real or continuous value, such as “salary” or “weight”. The output is in the form of Numerical values.

Classification

A classification model attempts to draw some conclusion from observed values. Given one or more inputs a classification model will try to predict the value of one or more outcomes.

Data Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. When creating a machine learning project, it is not always the case that we come across clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put it in a formatted way to ensure the quality.

Normalization

Data normalization is a technique used in data mining to transform the values of a dataset into a common scale. [0-1]

Binary Classification

Binary classification is a supervised learning algorithm that categorizes new observations into one of two classes.

Loss Function

A cost function or Error function is a measure of how inaccurate the model is in estimating the connection between X and Y. This is usually stated as a difference or separation between the expected and actual values. The term 'loss' in machine learning refers to the difference between the anticipated and actual value.

Loss function = Predicted - Target value

0 - correct prediction

Optimization

We have the concept of loss, which tells us how poorly the model is performing at that current instant. Now we need to use this loss to train our network such that it performs better. Essentially what we need to do is to take the loss and try to minimize it, because a lower loss means our model is going to perform better. The process of minimizing (or maximizing) any mathematical expression is called optimization.

Two types of Optimization

1. Maximization
2. Minimization

Optimizers are algorithms or methods used to change the attributes of the neural network such as **weights** and **learning rate** to reduce the losses. Optimizers are used to solve optimization problems by minimizing the function.

The curve that result in minimum errors is the best curve

Underfitting: It means that your model makes accurate, but initially incorrect predictions. In this case, train error is large and val/test error is large too.

Overfitting: It means that your model makes not accurate predictions. In this case, train error is very small and val/test error is large.

Linearly Separable Problem

The simplest case of a linearly separable decision problem is one consisting of two sets of points (patterns) in a 2-d vector space that belong to different classes, where the two classes can be separated by a straight line. Decision boundary are the boundary where we used to make a classification.

Linear Discriminant Function

It is a dimensionality reduction technique that is commonly used for supervised classification problems. It is used for modeling differences in groups i.e. separating two or more classes. It is used to project the features in higher dimension space into a lower dimension space.

Approaches:

Generative: Approaches estimate the discriminant function by first estimating the probability distribution of the data belonging to each class

Discriminative: Approaches estimate the discriminant function explicitly without assuming a probability distribution

Hyper Plane

Hyperplanes are decision boundaries that help classify the data points. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is just a line.

****Afternoon Section****

Perceptron Model

Perceptron is a building block of an Artificial Neural Network. It is Machine Learning algorithm for supervised learning of various binary classification tasks. Further, Perceptron is also understood as an Artificial Neuron or neural network unit that helps to detect certain input data computations in business intelligence.

Artificial Neural Network Layers

Artificial Neural network is typically organized in layers. Layers are being made up of many interconnected ‘nodes’ which contain an ‘activation function’. A neural network may contain the following 3 layers:

a. Input layer

The purpose of the input layer is to receive as input the values of the explanatory attributes for each observation.

b. Hidden layer

The Hidden layers apply given transformations to the input values inside the network.The values entering a hidden node multiplied by weights, a set of predetermined numbers stored in the program. The weighted inputs are then added to produce a single number.

c.Output layer

The hidden layers then link to an ‘output layer‘. Output layer receives connections from hidden layers or from input layer. It returns an output value that corresponds to the prediction of the response variable.

Loss function

Loss functions are one of the most important aspects of neural networks, as they (along with the optimization functions) are directly responsible for fitting the model to the given training data. It's a method of evaluating how well your algorithm models your dataset. If your predictions are totally off, your loss function will output a higher number. If they're pretty good, it'll output a lower number.

Squared Mean Error

One of the most popular loss functions, MSE finds the average of the squared differences between the target and the predicted outputs.

Squared loss is a loss function that can be used in the learning setting in which we are predicting a real-valued variable y given an input variable x.

Mean Absolute Error

MAE finds the average of the absolute differences between the target and the predicted outputs.

Binary Cross-Entopy

This is the loss function used in binary classification models — where the model takes in an input and has to classify it into one of two pre-set categories.

Activation Function

The activation function decides whether a neuron should be activated or not by calculating the weighted sum and further adding bias to it. The purpose of the activation function is to introduce non-linearity into the output of a neuron.

Sigmoid Function

Usually used in output layer of a binary classification, where result is either 0 or 1, as value for sigmoid function lies between 0 and 1 only so, result can be predicted easily to be ***1*** if value is greater than **0.5** and ***0*** otherwise.

$$A = 1/(1 + e\text{-}x)$$

RELU Function

It stands for *Rectified linear unit*. It is the most widely used activation function. Chiefly implemented in *hidden layers* of Neural network.

$$A(x) = \max(0,x)$$

It gives an output x if x is positive and 0 otherwise.

Softmax Function

The softmax function is also a type of sigmoid function but is handy when we are trying to handle multi- class classification problems.

Learning Using Iterative Optimization

It uses predictive modeling from the domain of machine learning to automatically focus search on those areas likely to give greatest performance.

Starts with a random initial weight values.

- Global Optima: It is to find the globally best solution of (possibly nonlinear) models, in the (possible or known) presence of multiple local optima.
- Local Optima: Local optimization involves finding the optimal solution for a specific region of the search space, or the global optima for problems with no local optima.

Optimizers

Optimizers are algorithms or methods used to change the attributes of your neural network such as weights and learning rates in order to reduce the losses.Optimization algorithm or strategies are responsible for reducing the losses and to provide the most accurate results possible.

Gradient Descent

Gradient descent is an iterative machine learning optimization algorithm to reduce the cost function. This will help models to make accurate prediction.

Type of Gradient Descent

1. Batch Gradient Descent
2. Stochastic Gradient Descent
3. Mini batc Gradient Descent

Batch Gradient Descent

We use the entire dataset to compute the gradient of the cost function for each iteration of the gradient descent and then update the weights.

Stochastic Gradient Descent

It’s a variant of Gradient Descent. It tries to update the model’s parameters more frequently. In this, the model parameters are altered after the computation of loss on each training example.

Mini batch Gradient Descent

Mini-batch gradient is a variation of gradient descent where the batch size consists of more than one and less than the total dataset. Mini batch gradient descent is widely used and converges faster and is more stable.

Day 8

Resource Person : Dhanya L K

Session: Forenoon

Topic: Introduction to NMT

Session:

- Introduction
- Machine Translation
- How to develop MT system?
- Methodology
- NMT
- RNN, Types of RNN
- Limitations of RNN
- LSTM
- Encoder-Decoder
- Transfer Models

Introduction

What is Natural Language Processing?

Natural language processing (NLP) combines computational linguistics, machine learning, and deep learning models to process human language. Computational linguistics. Computational linguistics is the science of understanding and constructing human language models with computers and software tools.

Two methods are available for NLP, Machine Learning and Deep Learning.

Machine Learning Vs Deep Learning

AI is basically the study of training machines to mimic a human brain and its thinking capabilities. AI focuses on 3 major aspects(skills): learning, reasoning, and self-correction to obtain the maximum efficiency possible.

Machine Learning is basically the study/process which provides the system to learn automatically on its own through experiences it had and improve accordingly without being

explicitly programmed. ML is an application or subset of AI. ML focuses on the development of programs so that it can access data to use it for itself. The entire process makes observations on data to identify the possible patterns being formed and make better future decisions as per the examples provided to them. The major aim of ML is to allow the systems to learn by themselves through experience without any kind of human intervention or assistance.

Deep Learning is basically a sub-part of the broader family of Machine Learning which makes use of Neural Networks(similar to the neurons working in our brain) to mimic human brain-like behavior. DL algorithms focus on information processing patterns mechanism to possibly identify the patterns just like our human brain does and classifies the information accordingly. DL works on larger sets of data when compared to ML and the prediction mechanism is self-administered by machines.

In ML & DL, method of training is different. In DL, some common steps are there for writing program. Text data is converted into numerical data (numbers). Whatever will be the input, we have to convert to numbers. That process is called Vectorization / feature extraction.

When you get a dataset, think about preprocessing. Preprocessing will differ from problem to problem. We can use libraries from python, for example nltk or keras for preprocessing. After preprocessing, we have to convert it to vector form. The techniques used for vectorization are One-hot encoding, Word2Vec, Glove,tfidf etc. The work quality depends on techniques. Use the better technique which is applicable for our project. Then give this output to any Deep Learning algorithm.

Machine Translation -The central problem of Machine translation is to bridge the language divergence.

- Dictionary based MT
- Rule based MT
- Statistical MT
- Neural MT

How to develop MT system?

The main task are Corpus collection, Preprocessing, Word embedding, Model creation and Performance evaluation. Here, dataset collection and preprocessing have more priority.

Dataset

Stanford NLP, Tab delimited Bilingual Sentence pairs,manythings.org etc. will be helpful for collecting datasets. Dataset may be structured or unstructured. Make your dataset balanced.For balancing oversampling and undersampling methods are available.

Preprocessing

Libraries like Nltk, keras etc. can be used.

Word Embedding

Tf-idf,glove,word2vec,one-hot encoding etc. can be used.

Neural machine translation (NMT)

Neural machine translation (NMT) is an approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

Basics of Neural network.

How to activate neurons?

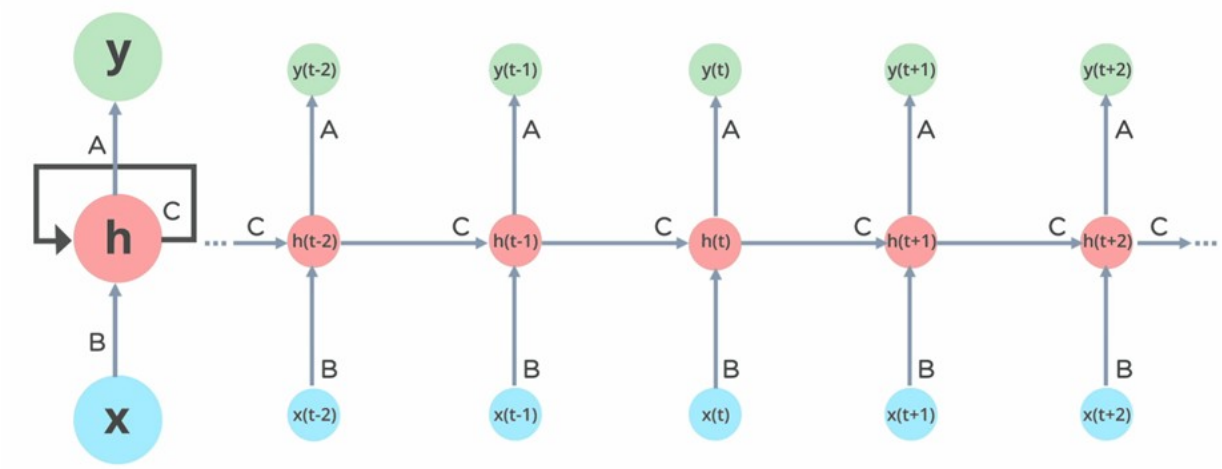
ANN-Feed forward,back propagation networks.

Recurrent Neural Network (RNN)

Recurrent Neural Network(RNN) is a type of Neural Network where the output from the previous step are fed as input to the current step. In traditional neural networks, all the inputs and outputs are independent of each other, but in cases like when it is required to predict the next word of a sentence, the previous words are required and hence there is a need to remember the previous words. Thus RNN came into existence, which solved this issue with the help of a

Hidden Layer. The main and most important feature of RNN is Hidden state, which remembers some information about a sequence.

RNN have a “memory” which remembers all information about what has been calculated. It uses the same parameters for each input as it performs the same task on all the inputs or hidden layers to produce the output. This reduces the complexity of parameters, unlike other neural networks.



Types of RNNs

One to one, one to many,many to one,many to many.

Limitations of RNN :

1. RNN training is a difficult process.
2. If it is using tanh or ReLu like activation function, it wouldn’t be able to handle very lengthy sequences.
3. The Vanishing or Exploding Gradient problem in RNN.

Long short-term memory (LSTM)

To overcome the drawbacks encounters in RNN the scientist made an invention called “LONG SHORT TERM MEMORY”. LSTM is the child of RNN where it can store long-term information and overcome the drawback of vanishing gradient.

1. **Forget Gate**

It is responsible for keeping the information or forgetting it so the sigmoid activation function is applied to it the output will be ranging from 0-1 if it is 0(forget the information) or 1(keep the information).

$$f_t = \sigma(W_f.[h_t - 1, x_t] + b_f)$$

2. **Input Gate**

It is responsible for expressing the importance of new information carried by the input here we will be applying two activation functions sigmoid and tanh. Sigmoid functionality is to keep or discard the information whereas tanh function is to subtract information from the cell state or to add new information to the cell state.

$$i_t = \sigma(W_i.[h_t - 1, x_t] + bi)$$
$$C_t = \tanh(W_C.[h_t - 1, x_t] + b_C)$$

3. **Output gate**

Based on the information we have gained we give our sentence as the fill-in-the-blank to the output gate based on the information it perceived will output the result

$$O_t = \sigma(W_o[h_t - 1, x_t] + b_o$$
$$h_t = o_t * \tanh(C_t)$$

BLEU Score

BLEU (Bilingual Evaluation Understudy) is a score used to evaluate the translations performed by a machine translator. It is calculated by comparing the n-grams of machine-translated sentences to the n-gram of human-translated sentences. Usually, it has been observed that the BLEU score decreases as the sentence length increases. This, however, might vary depending upon the model used for translation.

$$BLEUScore = BP * exp(\sum_{n=1}^{4} \frac{1}{n} P_n)$$

Afternoon Session

Resource Person :- Dr.Satheesh Kumar

Topic: R-Programming

- Session :- Hands-on of R Programming
- Linear regression
- Logistic regression
- Support Vector Machine
- Clustering - k-means.

Day: 9

Resource person: Dr. Manu Madhav

Session: NMT Implementation

- Started with basics of RNN
- Language modeling
- Basic idea
- Examples
- How language model can extend to machine translation

Issues with CNN and FFN

Sequence learning Problem

-sequence prediction, sequence generation, sequence recognition, and sequential decision making.

Sequence learning problem examples
Sequence learning problem modelling

RNN

Recurrent Neural Networks (RNNs) are used in machine translation because they are particularly well-suited for processing sequences of data, such as the words in a sentence.
By using RNNs, a machine translation system can learn to recognize patterns and dependencies between words in a sentence and generate translations that take these factors into account.
meaning of a word can be heavily influenced by the words that come before and after it.

- How we accounts for the dependency between inputs?

$y1 = f(xi)$
$y2 = f(x2,x1)$
$y3 = f(x3,x2,x1)$

Training RNN (backpropogation through time)

- ➔ Consider all the valuable dependencies
- ➔ Output we are predicting is a discrete value
- ➔ Natural way to represent discrete variables is to regard the o/p as giving the unnormalized log probabilities of each possible valuevof the discrete variable (softmax).

Problems with RNN

- Exploding gradients and vanishing gradients
- Memorizing the history

- Solution: LSTM

Language and cognition as probabilistic phenomena

- Human cognition is probabilistic and that language must therefore be probabilistic too since it is an integral part of cognition.

Language Modelling (basics)

- A language model is a probability distribution over strings on an alphabet.
- predicting the likelihood of a sequence of words occurring in a language.
- The goal of language modeling is to capture the statistical regularities of a language, such as word frequency and word order, and to use this information to generate coherent and fluent sentences.

Language Model - Other Applications

- Spelling mistake correction
- Speech recognition
- Machine Translation

-> Probability of a sentence is the joint probability of the words

$P(W) = P(w1,w2,w3,.....wn)$

$P(\text{the cat sat on the mat}) = P(\text{the})P(\text{cat}|\text{the})P(\text{sat}|\text{the cat}).....P(\text{mat} \mid \text{the cat sat on the})$

- **Markov Assumption**

$P(Wn+1 \mid W1….Wn) = P(Wn+1 \mid Wn)$

- Generally we can say it depends only on previous K words (K order markov model)

Markov Assumption and N gram model

Markov assumption is that the probability of the next event in a sequence depends only on the current state, and not on any previous states.

The n-gram model is a type of language model that is based on the Markov assumption. An n-gram is a contiguous sequence of n words from a sentence. For example, a 2-gram (or bigram) is a sequence of two adjacent words in a sentence. The n-gram model estimates the probability of a word given its n-1 previous words, based on the frequency of these n-grams in a training corpus.

For example,
suppose we have a sentence "The cat sat on the mat". The 2-gram model estimates the probability of each word given the previous word. So, the probability of "cat" given "the" is the number of times "the cat" appears in the training corpus divided by the number of times "the" appears in the training corpus. Similarly, the probability of "sat" given "cat" is the number of times "cat sat" appears in the training corpus divided by the number of times "cat" appears in the training corpus.

Simple N-Gram Model

How to estimate N-gram Probability

Preplexity

Preplexity is the most common intrinsic evaluation metric for Ngram language models

NMT (encoding)

Afternoon Session:- Hands-on NMT

Day: 10

Resource person: Mr. Deepu C

Session : Text Analytics

Text Analytics

Data types

Three types

- Unstructured
- Semistructured
- Structured

Unstructured and Semisectured data is more.

Data analytics (DA) is the process of examining data sets in order to find trends and draw conclusions about the information they contain. Data Analytics is the process where making data structured. The data is used in Text analytics is Real time data. That should be unstructured.
We get a lot of information from dealing with datas. We use Algorithms(Set of Instruction).

Text analytics uses

Sentiment analysis

Topic modelling

NER

Event extraction.

Why Python in Machine Translation?

Python is a programming language that distinguishes itself from other programming languages by its flexibility, simplicity, and reliable tools required to create modern software.

NLP Libraries

- NLTK
- Gensim
- Stanford core NLp

- SpaCy
- Text Bob.
- These are used in Text pre-processing.

Data Gathering

- API
- Database
- Web scrape
- Product reviews
- Social media
- Online surveys

Preparation of Data

- Tokenization
- POS
- NER
- Lemmatization and stemming
- Stop word removal.

Hands-on Data scraping.

Afternoon Session:- Project

DAY- 11

RESOURCE PERSON : Dr RENU S

Session : Forenoon

Topic : ISL

SIGN LANGUAGE

It is the communication medium for hearing impaired people. Using facial expressions, body movements, and gestures, sign languages provide a three-dimensional representation of thoughts and feelings.

Variants of Sign Language

American Sign Language (ASL), British Sign Language (BSL), Argentinian Sign Language (LSA), Indian Sign Language (ISL)

Sign Translation Research

Two Vertices of Sign Language Translation research:

Sign-to-Text (Video to Text\Speech) - converting a video image or a dynamic video sequence, then finding the appropriate text or speech.

Text- to-Sign - Translating text to its corresponding sign language.

Sign-To-Text

It focuses only on finger movements

1. Glove Based
2. Video\Image Processing

Text-To-Sign

1. Notation
2. Avatar Videos
3. Sigml player

Indian Sign Language

Although it is not widely recognised, Indian Sign Language is the preferred indication language in India.ISL is a complete language with defined grammar rules.

Text-to-Indian Sign Language Translation System

Text Preprocessing

Unlike English and other Indian languages, Indian Sign Language has its own set of grammar rules.

The three basic processes used to translate text into its ISL are phrase reordering, stemming, and elimination.

Elimination

Eliminating unused words, connection words, etc. are performed in this stage.

Never use linking verbs (am, is, are,etc) and gerunds (-ing).

Input text

After Elimination

I am Renu

Me\I Renu

The rose is red

Rose Red

He is a teacher

He teacher

We are friends

We friends

Stemming

ISL always uses root words and stemming means converting words into their root form.

Input text

After Stemming

Ramu is running

Ramu run

She is singing

She sing

They played Cricket

They play Cricket

She wrote a book

She write book

Phrase Reordering

Phrase reordering is a common format in ISL.

‘adjective+noun’ combination it will be signed as ‘noun+adjective ‘

Input text-

After Reordering

Red Rose

Rose Red

Beautiful Girl

Girl Beautiful

. WH- questions are always at the end.

Phrase Reordering

Input text-

After Reordering

What is your name?

Your name what?

What is the time?

Time what?

‘NOT’ is always at the end.

Phrase Reordering

Input text

After Reordering

I don't have any children

I children no

I don't know how to cook.

cook know not

ISL text follows ‘subject +object+verb ’ structure.

Preprocessing Challenges

1. Dinner, for instance, is indicated as "Night" and "Food".
2. ISL also makes use of the sentence's context.
3. For example the sentence "He listened to what I said” is signed as “He listens to me”.

Notations System

As a visual-spatial language, sign language is not capable of writing like other spoken languages.A textual representation of sign language has been established by scholars who are passionate about sign translation. Notations were used to describe these representations. The population of hearing-impaired people is unfamiliar with the notation systems.

These notation systems help to represent a 3D language in a written format.

Some of the commonly available notation systems are:

- Bebian Notation,
- Stokoe Notation
- Gloss Notation
- Hamburg Notation System (HamNoSys)
- SignWriting (SW)
- Si5s
- SignFont
- SignScript

- SLIPA

A standard phonetic notation method was required for the text to Indian Sign Language translation system in order to translate a three-dimensional language into a two-dimensional space.

HamNoSys Notation System

In order to translate a text into its equivalent visuals, the Institute of German Sign Language at the University of Hamburg created a notation system in 1980 that incorporates more than 200 symbols.

The HamNo notation has six parts. The first two parts Symmetry operator and Non-manual Features(NMF)are the optional features.

Hand-held ISL expressions can be divided into four.

- hand shapes
- hand orientation
- hand location
- hand movement.

BASIC HAND SHAPES

- hamfist
- hamflathand
- hamfinger2
- hamfinger23
- hamfinger23spread
- hamfinger2345
- hampinch12
- hampinchall
- hampinch12open
- hamcee12
- hamceeall

- Hamceeopen

To adapt and create additional symbols, such as basic hand forms, diacritics, symbols for fingers, and finger bends, the hand shape can be separated into four main categories.All the hand shapes are a combination of these different

Basic Hand shapes- Diacritics

- Hamfingerstraightmod
- hamthumboutmod
- hamthumbacrossmod
- hamthumbopenmod

Basic Hand shapes- Finger bend

- Hamdoublehooked
- Hamfingerbendmod
- Hamfingerhookmod
- Hamdoublebent

Basic Hand shapes- Symbol for Finger

- Hampinky
- Hamthumb
- Hamindexfinger
- Hammiddlefinger
- Hamringfinger

Basic Hand shapes- Finger Parts

- Hamfingerside
- Hamfingertip
- Hamfingernail
- Hamfingerpad
- Hamfingermidjoint

- Hamfingerbase

Direction of Forefinger

The direction of the extended fingers points in the same general direction as the vector that extends from the wrist along the back of the hand and continues in that same general direction. HamNoSys has 26 possible direction values.

Orientation of Palm

The palm's orientation with respect to the hand's shaft is referred to as palm orientation. The palm's orientation can take up to eight different values for each extended finger direction.For example, if the EFD is forward, palm orientation may be up, down, right, left, and four orientation intermediate to these.