Model Interpretation

**Topic Modeling**

Customer satisfaction is the key term in service industry and customer satisfaction us defined as a person’s feelings of pleasure resulting from perceived services from the service provider. Previous research and analysis have revealed that customer satisfaction has strong relationship with future spending in the respective service sectors. In tourism industry, the quality of visitors’ relationship with a destination has tremendous impact on the experience they have perceived beforehand. Service providers are relentlessly working to understand customers expectation in a more meaningful and effective way using customer reviews on specific services, in our case, thousands of hotel reviews. Topic modeling is not helping only to grab the poor area to improve but also identify the issue to overcome in future effort.

Topic modelling is to recognize words from the topics exist in the texts. Topic modelling is very useful and effective as extracting words from a text is time consuming and much more complex than extracting topics present in the texts. For instance, say, we have 10,000 documents and 1,000 words in each document. So, process takes multiple threads equivalent to thousands or even millions of threads. But, if we can divide the documents containing certain number topics based on the proportion on the documents, say 10 topics, that will be much easier to process with very a smaller number of threads.

Topic modelling is an unsupervised approach of extracting the topics by detecting the pattern like clustering. We can extract different topics from documents. This is done by extracting the pattern of word cluster and frequency of words in the documents. So based on this pattern, we divide documents in different topics. This type of modelling is very much useful when there are thousands of documents, and we are trying to get the information based on specific needs. This will take unusual amount time if we try to achieve manually. But topic modelling can satisfy the need in very little time.

Topic modelling is done using Latent Dirichlet Allocation, one of the popular modelling techniques. Topic modelling refers to identifying topics that best describes the documents. These topics only will be needed only once required; hence it is called latent.

**Pseudo code for LDA:**

For all topics in [1,k] do

Sample mixture components k ~ Dir(Beta)

End for

For all documents m in [1,m] do

Sample mixture proportion m ~ Dir(Alpha)

Sample documents length Nm ~ Poiss()

For all words n in [1, Nm] do

Sample topics index Zm,n ~ Mult(theta)

Sample term fro words wm,n ~ Mult()

**Algorithm**

1. Assume there are *k* topics across all the documents
2. Distribute these *k* topics across document *m*(this distribution is known as α and can be symmetric or asymmetric, more on this later) by assigning each word a topic.
3. For each word *w*in document *m*, assume its topic is wrong but every other word is assigned the correct topic.
4. Probabilistically assign word *w*a topic based on two things:  
   - what topics are in document *m*  
   *-*how many timesword *w*has been assigned a particular topic across all of the documents (this distribution is called*β*, more on this later)
5. Repeat this process several times for each document and you’re done!

## The Model:

Diagram

Description automatically generated

Above is what is known as a plate diagram of an LDA model where:

α is the per-document topic distributions,

β is the per-topic word distribution,

θ is the topic distribution for document m,

φ is the word distribution for topic k,

z is the topic for the n-th word in document m, and

w is the specific word

Smoothed LDA from <https://en.wikipedia.org/wiki/Latent_Dirichlet_allocation>

**Workflow for Topic Modeling**

**Diagram

Description automatically generated**

**Text Classification**

Text classification is one of the most common tasks in Natural Language processing (NLP). Text Classification is a common NLP task that assigns a label or class of text. Text classification can be described as a machine learning technique to classify the text into a specific category. These categories depend on the type of task they perform. Examples are spam detection, topic labeling, sentiment analysis, review rating, and intent detection.

A Long Short-Term Memory (LSTM) network is a particular type of recurrent network that works slightly better in practice, owing to its more powerful update equation and some appealing back propagation dynamics. Computers which can learn algorithms to map input sequences to output sequences The LSTM units give the network memory cells with read, write and reset operations. During training, the network can learn when it should remember data and when it should throw it away. LSTM is well-suited to learn from experience to classify, process and predict time series when there are very long-time lags of unknown size between important events.

**LSTM Diagram**

The following is the LSTM diagram at the t-time step.

Diagram

Description automatically generated

Xt = Input vector at the t-time.

Ht−1 = Previous Hidden state.

Ct−1 = Previous Memory state.

Ht = Current Hidden state.

Ct = Current Memori state.

[\*] = multiplication operation.

[+] = addition operation.

So, the input of each LSTM module is Xt (current input), Ht−1, and Ct−1. then the output is Ht, and Ct.

**Workflow for Text Classification**

**Diagram

Description automatically generated**

**LSTM Model Workflow**

**Timeline

Description automatically generated**

**Sentiment Analysis**

Sentiment Analysis is a technique to predict the feelings and emotions reflected by a word or group of words. Natural Language Processing is the ability of computer to understand human language. For this , we have to extract feature from a text to analyses the similarities of the text. Using machine learning we can train a model to process large amount of data. But the problem is, in NLP we have to dealt with raw text which is not simply achievable by using raw text only. We need to extract the feature from the raw text to train the model using Machine Learning algorithm. Machine Learning Algorithm will then process the list of features and classify the text or to produce the output of the text based on list of features. In NLP, we have to process the raw text and make it usable for Machine Learning Algorithm to process the text data or human language in meaningful way. Feature extraction process means to extract and produce feature representations that are appropriate for the type of NLP tasks we are trying to accomplish and the type of model we are planning to use. in NLP, feature extraction helps to reduce the amount of redundant data from dataset. In sentimental analysis, we have to preprocess the text and make the unstructured data in a form that can be used for classification. In NLP, we need feature extraction techniques to convert the raw text into a matrix or vector of features.

Sentiment analysis is a natural language processing techniques to determine whether customer review is positive, negative, or neutral. Sentiment analysis is mostly performed on text to help the business monitor its products or services based on customer feedback. Sentiment analysis helps to measure the public opinion on their services which is in turn open the window for improvement for their services and products. Sentiment analysis can be divided into five categories as below:

1. Emotion Analysis
2. Multilingual sentiment analysis
3. Aspect based
4. Intent analysis
5. Fine-grained Sentiment analysis

The overall benefits of sentiment analysis include:

1. Sorting Data at scale: Big data will be a big problem if we don’t have required tools for analysis. It will be impossible for human being to handle enormous amount of data generated every second with the rise of technological advancement. Sentiment analysis with deep neural network algorithm makes it possible and easier to classify large amount of raw data.
2. Real time analysis : Business entity can learn and analysis real time input of customer review and make necessary steps to overcome any bottleneck along the way to improve quality of service and products as per customer need.
3. New market : Sentiment analysis could open the new opportunity based on the feedback of the customer either positive or negative.

We use transformer model for sentiment analysis for our project. Transformers architecture is one of the most influential mechanisms in deep learning and NLP. Transformers consist of encoder architecture with the positional encoding, inputs, input embeddings, and a block containing neural networks. Transformers architecture is a type of deep learning architecture that learns text representations using self-attention mechanism. Transformers are the state-of-the-art NLP models that achieve best performance on many NLP benchmarks. The high-level implementation of transformer models for text classifications are:

a. Importing transformers library

b. Importing the dataset

c. Tokenize the dataset

d. Training the model and making predictions.

e. Evaluate the model

f. Making prediction of test dataset.

For tasks in which the text classes are relatively few, the best performance on text classification can be achieved using pretrained Transformers models like BERT, XLNet, and RoBERT. But transformer models scale quadratically with input sequence length and linearly with the number of classes.

**Transformer Model**

Diagram

Description automatically generated

**BERT Model**

BERT learns contextual relation between words using transformer. Transformer architecture consists of encoder and decoder stack, but BERT uses only encoder part. Encoder learns what is English, what is grammar, more importantly what is context? BERT is stack of encoder and BERT is used to learn language translation, question-answering, sentiment analysis, text summarization etc. All these problems require good understanding of language. Training of BERT is done in two phases. In first phase, BERT model understands the language and context. In second phase, BERT model will solve the problem as it already knew the language in the pretraining phase.

BERT learns language by training two unsupervised tasks simultaneously. One is Masked Language Modelling(MLM) and Next Sentence Prediction(NSP). BERT understands bidirectional context of the sentence in MLM task in pretraining phase. In NSP, BERT takes two sentences and determines whether second sentence follows the first sentence. In this way, BERT understands the context across different sentences themselves. Now, BERT has a very good understanding of language. In fine-tuning, phase BERT delves with specific problem to solve, in our case, that is sentiment analysis.

Diagram

Description automatically generated

Overall Workflow for Sentiment Analysis:

**Diagram

Description automatically generated**