

# *NLP Models (BERT, XLNET, ROBERTA) and Effective Model Deployment Strategies (Generalization, Fine Tuning, Compression, Pruning..) A Study paper on Transfer Learning in NLP*

G.Priyadharshini<sup>1</sup>, Dr.S.Bhuvaneswari<sup>2</sup>

<sup>1</sup>Research Scholar, Department of Computer Science, Pondicherry University, Karaikal Campus-609605

<sup>2</sup>Head & Professor, Department of Computer Science, Pondicherry University, Karaikal Campus-609605

Email: priyavinsundar@gmail.com

**Abstract:** *Transfer learning is a technique where Knowledge transfer occurs from a trained model for one task to another related model having different task. Natural Language processing explains, produces and makes computers to understand Human Language in an efficient manner. BERT, RoBERTa and XLNet are pre-trained models which use Transfer learning technique. These Models are used in Natural Language processing for understanding of context in a Human Language. These pre- trained models use the Model Deployment strategies for their effective pre-training.*

*Main objective of this study is to analyses how well the pre-trained models used in context understanding works and a little bit improvements in each models in NLP and also strategies applied to effectively deploy a model. In this paper the pre-trained models used in context understanding and Model Deployment strategies for these models are explained.*

**KEYWORDS:** *BERT, RoBERTa, XLNet, Generalization, fine-tuning, compression, pruning, Expansion, Data Augmentation*

## I. INTRODUCTION

Transfer learning reuses the capacity of knowledge learned from one task to another related task. Natural Language Processing is used in making computers understand, produce, and explain the Human language. It contains variety of algorithms, techniques and methods to interpret human Languages. There are several kinds of Transfer Learning Techniques available and they are

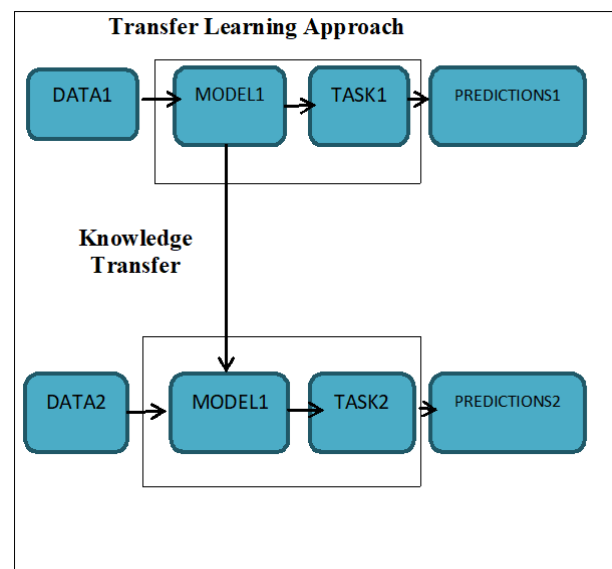


Fig: 1 Traditional Transfer Learning Approach

**A. Instance Transfer:**

It is the transfer of particular instance form one task to another. It is used for effective improvement of Target task

**B. Transductive Transfer:**

In this type of Learning the model is trained with zero or one samples of data. Examples Zero shot Learning, One shot learning.

**C. Self-Taught Learning:**

Self-Taught Learning learns from previous task input and transfers that knowledge to the next task.

**D. Multi Task Learning:**

It takes multiple views or explanations of same data.

**II. RELATED WORK**

Transfer Learning is a Solution to cold start problem which occurs when a new model requires much data but lacks prior training data [1]. Deep learning needs large data for its effective performance, for instance in medical image analysis it does not have enough data to make predictions, so transfer learning is used in Deep Learning Models with knowledge learned from Large Data sets and apply them in small Datasets.[2]. Recent Transfer Learning Models are used to detect context in a text by making some changes in configuration to the existing model [3]. Sentiment analysis is a subdivision of text mining which uses NLP and other techniques to classify emotions in text and to do this task various methods are applied such as Traditional methods ,Deep Learning methods ,Transfer Learning Methods and concluded Transfer learning is a best technique for sentiment analysis[4]. Transfer learning and Deep Ensemble Neural network are applied to find plant leaf disease detection and transfer learning is used to fine tune the pre trained Model[5]. Transfer Learning is used for sequence Generation in

a low resource downstream task and it takes multiple sources including multi source translation, multi document summarization, automatic post editing[6]. Sentimental analysis contains different kinds of data such as cross domain ,Multi modal, cross lingual, small scale and the emotional state is extracted using lexicon based approach and Machine Learning Based approach[7]. Sentiment analysis with short term and small scale data classification is also done with help of Transfer learning[8]. Transfer Learning is used in Zero shot Learning in this rich get richer problem occurs where the well performing dataset is used more for training than average performing dataset[9]. Image cyber bullying problem can be detected using Transfer and deep learning techniques[10].

**III. PRE TRAINED MODELS OF TRANSFER LEARNING IN NATURAL LANGUAGE PROCESSING****A. BERT :( Bidirectional Encoder Representations from Transformers)**

It is an NLP Technique that is efficient in understanding the context in a text. These models are trained on vast amount of Text Data and can understand Bidirectional context of words in a Sentence and can understand words based on the words that are before and after them. It is used in classification, Question Answering, Named Entity Recognition etc.,

**B. XL Net:**

It is an NLP Technique which contains Transformation Based Technique plus concept of permutation Based Architecture. XLNet is aimed to address the Limitations of BERT. It is also called “Permutation Language Modeling” because it allows the model to investigate

all permutation of words in a sentence enable it to learn Bidirectional context of words without restriction of justified reading. It perform more effective than BERT in all applications Like Classification, Question Answering Etc...

### C. RoBERTa:

It is one of the Leading model of NLP and a Variation of BERT Model which gives optimized performance than BERT. It can perform task like Classification, Sentiment Analysis, and Language Understanding etc.,

## IV. MODEL DEPLOYMENT STRATEGIES

### A. Fine Tuning:

It trains a pre-trained model trained on a large dataset with a smaller, task specific dataset. It is a transfer learning Technique with transformer based models like BERT, GPT, and RoBERTa.

### B. Generalization:

It is a technique in which the model ability to perform well on a new dataset is tested and improved. It involves aspects such as Task Similarity, Amount and Quality of data, Model Capacity and Complexity, Regularization Technique, Hyper Parameter Tuning, Domain Adaptation Tuning.

### C. Compression:

BERT, RoBERTa and XLNet are exponentially large models and use millions of parameters during pre-training. The pre-trained models can be fine-tuned by applying compression techniques to make them smaller and faster.

### D. Pruning:

It modifies the model architecture and the strategies are

- **Head pruning:** Removes Less Important Heads for a specific task.
- **Weight Pruning:** Removes unnecessary weight from an architecture.
- **Layer Pruning:** Removes full layer of transformer without affecting the architecture and the goal of pruning is to make the model more efficient, faster and requires less memory and improves prediction accuracy.

### E. Expansion:

It is the modification or addition of layers within Neural Network architecture during fine tuning process to adapt the model for a specific task and also for performing new task. Modifications such as adding layers, changing layer style, modifying output layer, fine tuning parameters can be done.

### F. Data Augmentation:

Data Augmentation technique can be used to make an unbalanced dataset to balanced dataset .In any classification technique a balanced dataset generates clear and accurate decisions.

## V. RESULT ANALYSIS

Analysis of the best pre-trained model among the three models is performed with the help of the table given below. It compares the features the pre-trained models such as BERT, XLNet , RoBERTa are discussed and the features include architecture, Training objectives, Tokenization, Training Data, Training Efficiency, Training Steps and comparative analysis is also discussed.

recognition, machine translation, and more.

**Future Directions:**

**A. Fine Tuning:**

Refining fine-tuning techniques and developing specialized models for specific domains resulting in enhanced performance on niche tasks.

**B. Multimodal Learning:**

Future research may explore models capable of understanding and generating both text and images, paving the way for more comprehensive AI systems that can analyse and interpret information from multiple modalities.

**C. Ethical Considerations:**

Ethical considerations surrounding bias, fairness, and interpretability are gaining prominence.

FEATURE	BERT	XLNET	ROBERTA
Architecture	Transformer	Transformer with Permutation Language Model	Transformer
Training Objective	Masked Language Model (MLM)	Autoregressive Permutation Language Model	Masked Language Model (MLM)
Tokenization	WordPiece	SentencePiece	SentencePiece
Training Data	BooksCorpus, English Wikipedia	BooksCorpus, English Wikipedia	BookCorpus, English Wikipedia, CC-News
Training Efficiency	Slower due to bidirectional context	Slower due to autoregressive context	Faster due to dynamic masking
Training Steps	1 million steps	800,000 steps	500,000 steps

**Table 1-Pretrained Models Analysis**

From the above comparison it is clear that each model is unique and strong in a particular feature so the choice to choose the models among the above belongs to the particular use case.

**VI. CONCLUSION AND FUTURE SCOPE**

Transfer learning, particularly in the context of pre-trained language models like BERT, RoBERTa, and others, NLP tasks more effective. These models, trained on numerous amounts of different text data, have exposed superior performance across a numerous natural language understanding tasks, such as sentiment analysis, named entity

**REFERENCES**

1. Tao Wu, EllieKa-In Chio, Heng-Tze Cheng, YuDu, Steffen Rendle, Dima Kuzmin, Ritesh Agarwal, Li Zhang, John Anderson, Sarvjeet Singh,...” Zero-Shot Heterogeneous Transfer Learning from Recommender Systems to Cold-Start Search Retrieval”-Google Inc -2020.
2. Hsu Mon Lei Aung, Charnchai Pluempitiwiriawej, Kazuhiko Hamamoto and Somkiat Wangsiripitak, “Multimodal Biometrics Recognition Using a Deep Convolutional Neural Network with Transfer Learning in Surveillance Videos”, 2022.

3. Yifei Zhao<sup>1</sup>, “Overview of Deep and transfer Learning Methods for Sentiment Analysis”, EEMAI,2023.
4. Karthick Prasad Gunasekaran , “Exploring Sentiment Analysis Techniques in Natural Language Processing: A Comprehensive Review”, June 2022.
5. Kian Long Tan, Chin Poo Lee, (Senior Member, IEEE), Kian Ming Lim, (Senior Member, IEEE), and Kalaiarasi Sonai Muthu Anbananthen, “Sentiment Analysis with Ensemble Hybrid Deep”, July 2022.
6. Shelley Gupta, Archana Singh and Vivek Kumar, “Emoji, Text, and Sentiment Polarity Detection Using Natural Language Processing”, MDPI-31 Mar 2023.
7. Nilaa Raghunathan and Saravanakumar Kandasamy,” Challenges and Issues in Sentiment Analysis: A Comprehensive Survey”, VIT India, July 2023.
8. Daniil Homskiy, Narek Maloyan, “DN at SemEval-2023 Task12: Low-Resource Language Text Classification via Multilingual Pretrained Language Model Fine-tuning”, June 2020.
9. A Transfer-Learning Scheme for Semi-Supervised Few-Shot and Zero-shot learning”, CFV, 2020.
10. Amar Almomani, Mohamad Alaudhamn, Mohammad Azmi Al Betar,” International Journal of Cognitive Computing in Engineering”-KeAi Chinese Roots Global Impact-5, 14-26-2024.