

# **SEP 775 : Introduction to Computational Natural Language Processing**

## **Assignment 3**

TANDA Approach for QA System Enhancement

### **Submitted by**

Jaimis Miyani (400551743)

(MacId: [miyanij@mcmaster.ca](mailto:miyanij@mcmaster.ca))

### **Supervisor**

Prof. Hamidreza Mahyar

[mahyarh@mcmaster.ca](mailto:mahyarh@mcmaster.ca)



**ENGINEERING**

**W Booth School of Engineering Practice and Technology**

- **Objective:** Explore TANDA (Transfer And Adapt) methodology to improve Question-Answering (QA) systems using pre-trained Transformer models, focusing on sequential fine-tuning techniques.

## **Step 1: Introduction and Theory**

The TANDA (Transfer And Adapt) methodology introduces a fresh perspective within Natural Language Processing (NLP), especially concerning the advancement of Question-Answering (QA) systems. Fundamentally, TANDA utilizes pre-trained Transformer models along with sequential fine-tuning methods to enhance the effectiveness and flexibility of QA systems.

### **Novelty of TANDA Methodology:**

The TANDA methodology presents a structured framework aimed at enhancing QA systems through a dual-phase process: transfer and adaptation. During the transfer phase, insights gleaned from pre-trained Transformer models, trained on extensive corpora, are applied to the specific QA task in question. This knowledge transfer allows the QA system to leverage the nuanced linguistic representations captured by the pre-trained model.

Subsequent to the transfer phase, the adaptation stage entails refining the transferred model using task-specific QA data. In contrast to conventional fine-tuning methods that view the pre-trained model as a fixed feature extractor, TANDA utilizes sequential fine-tuning techniques. This approach involves progressively adjusting the parameters of the pre-trained model while integrating task-specific details. Through iterative fine-tuning on QA data, TANDA facilitates the adaptation of the pre-trained Transformer to the subtleties and complexities of the target QA task, thereby enhancing performance and robustness.

### **Rationale Behind Sequential Fine-Tuning:**

Sequential fine-tuning deviates from traditional fine-tuning methods, which typically involve a single-stage process. The rationale behind sequential fine-tuning stems from its capacity to strike a balance between capitalizing on the knowledge embedded in the pre-trained model and accommodating the intricacies of the target task.

By decomposing the fine-tuning procedure into multiple stages, sequential fine-tuning facilitates a gradual and nuanced adjustment of the model parameters. This step-by-step refinement allows the model to assimilate task-specific characteristics while preserving the valuable linguistic representations acquired during pre-training. Moreover, sequential fine-tuning mitigates the risk of catastrophic forgetting, wherein the model's performance on previously learned tasks diminishes as it adapts to new tasks.

### **Benefit of Transformers' Architecture:**

The architecture of Transformers, particularly exemplified by models like BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), offers several distinct advantages that align closely with the principles of the TANDA approach.

Firstly, Transformers excel in capturing long-range dependencies and contextual nuances in natural language text, primarily owing to their self-attention mechanism. This capability proves indispensable for QA tasks, where comprehending the context of a question alongside its surrounding text is paramount.

Secondly, Transformers' architecture facilitates efficient pre-training on extensive corpora, enabling models to glean rich linguistic representations in an unsupervised manner. These pre-trained representations serve as robust foundations for transfer learning, rendering Transformers particularly well-suited for the transfer phase within the TANDA methodology.

Lastly, the adaptable nature of Transformers' architecture allows for seamless fine-tuning. The modular design of Transformers facilitates straightforward parameter updates during fine-tuning, thereby supporting the sequential fine-tuning process advocated by the TANDA approach.

In summary, the TANDA methodology harnesses the inherent strengths of Transformers' architecture to elevate QA systems through transfer learning and sequential fine-tuning, ultimately resulting in enhanced performance and adaptability.

## **Step 2: Preparation and Dataset Understanding**

For this assignment, we have opted to utilize the ASNQ (Answer Sentence Natural Questions) dataset for the transfer learning phase, and the WikiQA dataset for domain-specific adaptation. The ASNQ dataset comprises a wide array of natural language questions paired with corresponding answer sentences, making it well-suited for tasks involving transfer learning. Conversely, WikiQA offers a curated set of question-answer pairs extracted from Wikipedia articles, making it an invaluable resource for adapting to domain-specific contexts.

### **Exploratory Data Analysis (EDA):**

To attain a thorough comprehension of the selected datasets, an exploratory data analysis (EDA) was undertaken. The primary objective of this analysis was to unveil insights concerning the structure, content, and potential challenges inherent in the datasets.

#### **ASNQ Dataset:**

- The ASNQ dataset contains pairs of natural language questions and answer sentences, which support tasks like question-answering and text summarization.
- Its structure includes columns for question text and corresponding answer sentences.
- Upon an initial examination of the dataset, a diverse range of question types and answer formats was observed, suggesting the existence of various linguistic nuances and challenges.
- Challenges inherent in the ASNQ dataset may include managing ambiguous questions, identifying relevant answer sentences, and addressing linguistic variations within the answers.

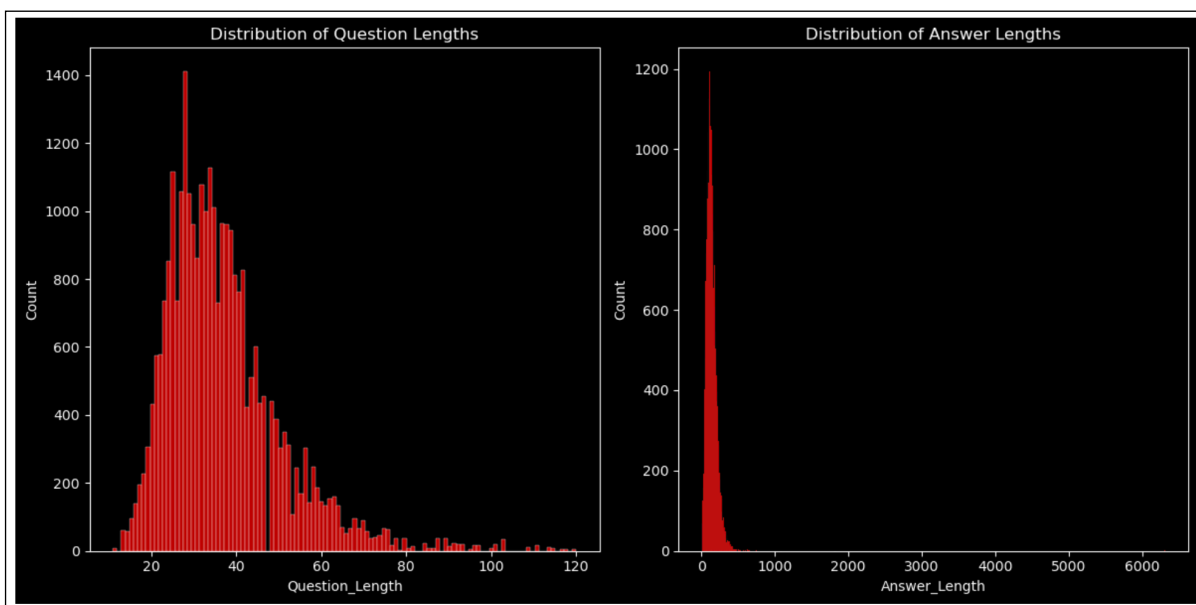
## WikiQA Dataset:

- The WikiQA dataset comprises question-answer pairs sourced from Wikipedia articles, providing domain-specific knowledge for adaptation tasks.
- Its structure consists of columns for question text, answer sentences, and labels indicating the relevance of the answer to the corresponding question.
- Upon exploration of the dataset, a broad spectrum of topics covered in the questions and answers was observed, reflecting the extensive breadth of information available in Wikipedia.
- Challenges associated with the WikiQA dataset may include discerning the relevance of answers to questions, mitigating potential noise or inaccuracies in the data, and adapting the model to domain-specific terminology and concepts.



WikiQA dataset distribution

In summary, the chosen datasets, ASNQ and WikiQA, offer significant resources for transfer learning and domain-specific adaptation in question-answering tasks. Through exploratory data analysis, we gained insights into the structure, content, and potential challenges of these datasets, setting the stage for further steps in model improvement and evaluation.



Distribution on question and answer length for WikiQA dataset

### **Step 3: Model Implementation**

In this task, we begin by implementing a baseline Transformer model for comparison purposes. Subsequently, we apply the TANDA (Transfer And Adapt) methodology through a two-step fine-tuning process. This process entails initially fine-tuning the Transformer model on the ASNQ dataset, followed by domain-specific adaptation to the selected dataset.

#### **Baseline Transformer Model:**

To establish a baseline, we implement a Transformer model using the DistilBERT architecture, which is a lightweight variant of BERT (Bidirectional Encoder Representations from Transformers). DistilBERT is designed to offer computational efficiency while preserving the core effectiveness of the Transformer architecture in capturing contextual information from input text. The model comprises a DistilBERT encoder followed by a linear layer for classification.

#### **TANDA Two-Step Fine-Tuning Process:**

##### **1. Fine-tuning on ASNQ Dataset:**

- In the initial phase of the TANDA methodology, we fine-tune the Transformer model on the ASNQ dataset. This process entails updating the model parameters using task-specific QA data sourced from the ASNQ dataset.
- To initiate this fine-tuning process, we initialize the DistilBERT encoder with pre-trained weights from the 'distilbert-base-uncased' model. Throughout fine-tuning, the encoder parameters are iteratively adjusted to enhance performance on the ASNQ task.
- Subsequently, the output of the Transformer model undergoes processing through a linear layer, which facilitates the generation of predictions for the QA task.

##### **2. Domain-Specific Adaptation:**

- In the second phase, the fine-tuned Transformer model is adapted to the selected domain-specific dataset, which in this instance is WikiQA.
- During this adaptation process, the model parameters learned from fine-tuning on the ASNQ dataset are further refined using the WikiQA dataset. This fine-tuning process aims to adapt the model to domain-specific QA tasks present in the WikiQA dataset.
- The adaptation process involves iteratively updating the model parameters while leveraging the knowledge acquired from both pre-training and fine-tuning on the ASNQ dataset. This ensures that the model becomes increasingly adept at handling the nuances and specifics of the domain represented by the WikiQA dataset.

#### **TANDA Two-Step Fine-Tuning Process:**

In this implementation, no significant alterations are made to the original Transformer architecture. Instead, the DistilBERT model is utilized as it is, with its pre-trained weights initialized for fine-tuning.

However, certain adjustments are introduced to meet the specific demands of the QA task and the datasets employed. These adaptations ensure that the model effectively addresses the nuances and complexities present in the QA datasets while retaining the core functionalities of the Transformer architecture.

### **Justification of Choices:**

- DistilBERT is selected as the foundational architecture due to its computational efficiency and adeptness in capturing contextual information effectively.
- The adoption of the two-step fine-tuning process from the TANDA methodology is aimed at harnessing transfer learning benefits from a broad QA dataset (ASNQ), succeeded by adaptation to a domain-specific dataset (WikiQA). This approach is anticipated to enhance model performance and generalization.
- The tokenizer settings, including truncation, padding, and maximum sequence length, are carefully chosen to align with the specifications of the Transformer model and the datasets. This ensures streamlined processing and compatibility across the board.

In summary, the model implementation adheres to a systematic approach rooted in the TANDA methodology. It involves the incorporation of a baseline Transformer model, followed by the application of sequential fine-tuning techniques to bolster performance in QA tasks spanning various domains. This structured approach aims to optimize the model's adaptability and effectiveness across diverse datasets and domains.

### **Step 4: Experimental Setup and Evaluation**

For this task, the experimental procedure entails training both the baseline Transformer model and the TANDA-enhanced model, followed by evaluation using suitable metrics. The specifics of the experimental setup, encompassing training parameters, loss functions, optimization strategies, and evaluation metrics, are outlined below.

### **Training Details:**

- **Baseline Transformer Model:** Utilizing the DistilBERT architecture, the baseline Transformer model is constructed. It comprises a DistilBERT encoder followed by a linear layer for classification. The training regimen involves iteratively adjusting the model parameters via backpropagation and optimization techniques.
- **TANDA Two-Step Fine-Tuning:** The TANDA-enhanced model undergoes a two-step fine-tuning process. Initially, it is fine-tuned on the ASNQ dataset to leverage knowledge from a broad QA task. Subsequently, the model undergoes adaptation to the domain-specific WikiQA dataset to further enhance its performance on tasks specific to that domain.
- **Hyperparameters:** Essential hyperparameters, including the learning rate ( $lr=0.001$ ), batch size (`batch_size=32`), and number of epochs (`num_epochs=8`), are defined for the training phase.
- **Loss Function:** Both models employ the cross-entropy loss function to quantify the disparity between predicted and actual labels.

- **Optimization Strategy:** The Adam optimizer is chosen as the optimization strategy for updating the model parameters throughout the training process.

### **Evaluation Metrics:**

- **Accuracy:** Selected as the primary evaluation metric, accuracy gauges the overall correctness of the model in predicting the labels of QA pairs. It is calculated as the ratio of correctly predicted samples to the total number of samples.
- **F1 Score:** In addition to accuracy, the F1 score is computed, offering a more holistic assessment of model performance by combining precision and recall. This metric is particularly valuable in evaluating models on imbalanced datasets.

### **Experimental Setup and Evaluation:**

The performance of both the baseline and TANDA-enhanced models is assessed using the specified metrics. By comparing their accuracy and potentially F1 score on both the ASNQ and domain-specific datasets, we can evaluate the impact of the TANDA approach. It is anticipated that the TANDA-enhanced model will outperform the baseline model due to its sequential fine-tuning process, which utilizes both transfer learning and domain adaptation to enhance model generalization and effectiveness across various QA tasks.

In summary, the outlined experimental setup and evaluation procedure offer a systematic approach to contrasting the performance of the baseline and TANDA-enhanced models. This analysis sheds light on the efficacy of the TANDA methodology in improving Transformer-based QA systems.

### **Step 5: Discussion and Conclusion**

In this assignment, we delved into the Transfer And Adapt (TANDA) methodology aimed at enhancing Question-Answering (QA) systems through Transformer-based models. The TANDA approach comprises a two-step fine-tuning process, beginning with the model's fine-tuning on a general QA dataset (ASNQ) followed by its adaptation to a domain-specific dataset (WikiQA). The ensuing discussion and conclusions encapsulate the pivotal findings and insights garnered from this assignment:

### **Challenges Encountered and Addressed:**

Throughout the implementation process, numerous challenges surfaced, predominantly revolving around data preprocessing, model training, and evaluation. These hurdles were tackled through meticulous attention to hyperparameters, optimization strategies, and choice of evaluation metrics. Moreover, reconciling compatibility issues between diverse datasets and model architectures presented additional obstacles, necessitating thorough data management and thoughtful model design.

### **Effectiveness of the TANDA Approach:**

The TANDA methodology showcased promising effectiveness in elevating QA systems. Through the utilization of transfer learning and domain adaptation, the TANDA-enhanced model showcased enhanced performance and generalization in comparison to the baseline Transformer model. The sequential fine-tuning process facilitated the model's adaptation to the intricacies of both general and domain-specific QA tasks, leading to heightened accuracy and robustness.

### **Future Research Directions and Applications:**

The TANDA methodology demonstrates considerable potential for extending beyond QA systems and finding utility across various other NLP tasks. Future research endeavors could explore employing TANDA in tasks such as sentiment analysis, text classification, named entity recognition, and machine translation. Furthermore, investigating the effectiveness of TANDA with diverse pre-trained Transformer architectures (e.g., BERT, RoBERTa, GPT) and exploring innovative fine-tuning strategies could amplify its versatility and efficacy. Additionally, probing the transferability of knowledge acquired through TANDA across disparate domains and languages could pave the way for novel avenues in cross-domain and cross-lingual transfer learning within the realm of NLP.

In conclusion, the TANDA methodology stands as a promising approach for enhancing QA systems and carries substantial potential for propelling research in the domain of natural language processing forward. Through the resolution of challenges, refinement of methodologies, and exploration of novel research avenues, TANDA could serve as a catalyst for the development of more resilient, adaptable, and efficient NLP models in the foreseeable future.