Aryan Jain
2019101056

IRE Assignment-3

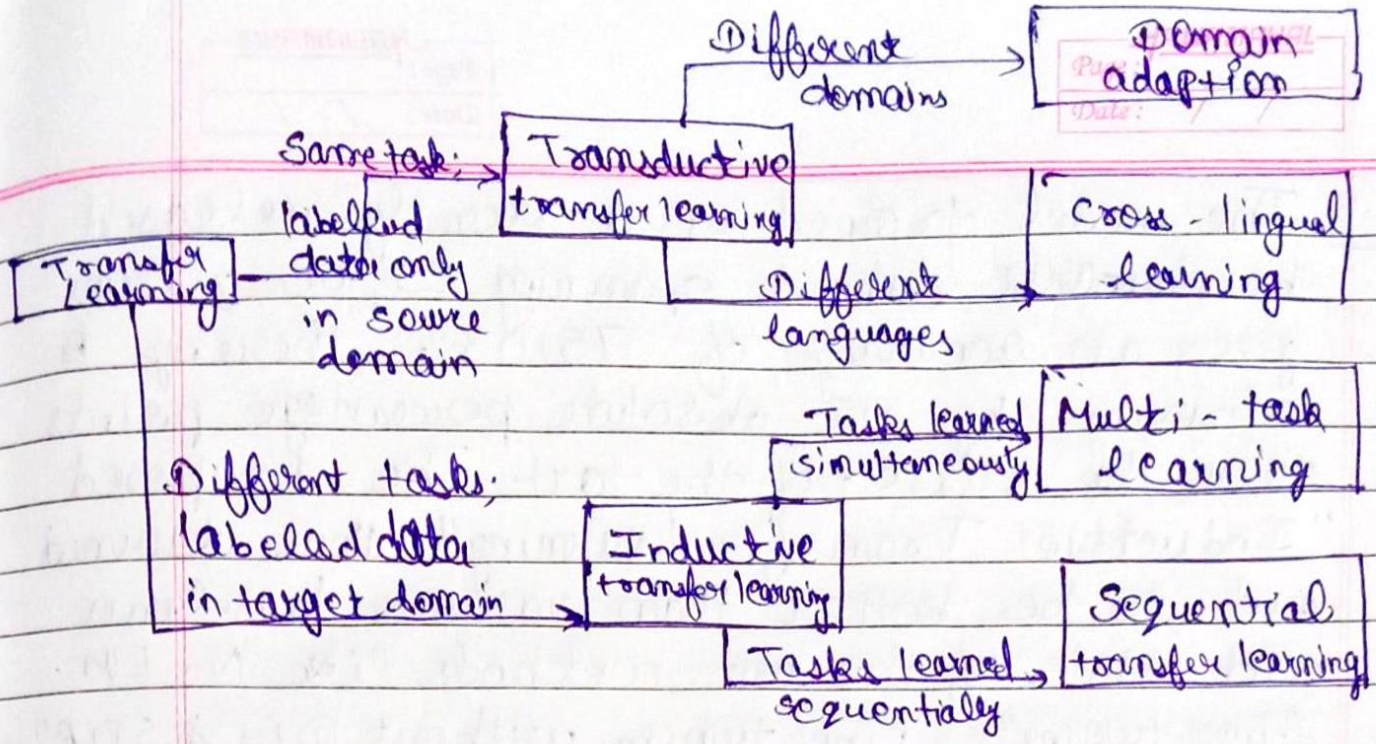
Ans.1 The problem that the authors are trying to solve is to identify whether a search query is a well-formed natural language question or not using inductive transfer learning technique. (by fine-tuning a pre trained language model).

Authors founded out solving this problem essential because recently there has been a significant rise in the number of verbose queries by the users. often times such queries are not well-formed and the lack of well-formedness in the query might adversely impact the downstream pipeline which processes these queries. A well-formed natural language question as a search query aids heavily in reducing errors in downstream tasks and further helps in improved query understanding.

Ans.2 Transfer learning is a means to extract knowledge from a source setting and apply it to a different target setting. There are different types of transfer learning common in current NLP. These can be roughly classified along 3 dimensions based on

a) whether the source and target setting deal with the same task.
b) the nature of the source and target domains.
c) the order in which the tasks are learned.

...continued



In the research paper, the authors used the inductive transfer learning to identify well-formed natural language questions.

Ans. 4 The dataset used was the Google's query well-formedness dataset which was created by crowdsourcing well-formedness annotations for 25,100 queries from the Paralex corpus, each labeled with a rating (between 0 and 1) of the query being well formed. The compiled dataset contains well-formed questions as queries along with typical constructs of search queries.

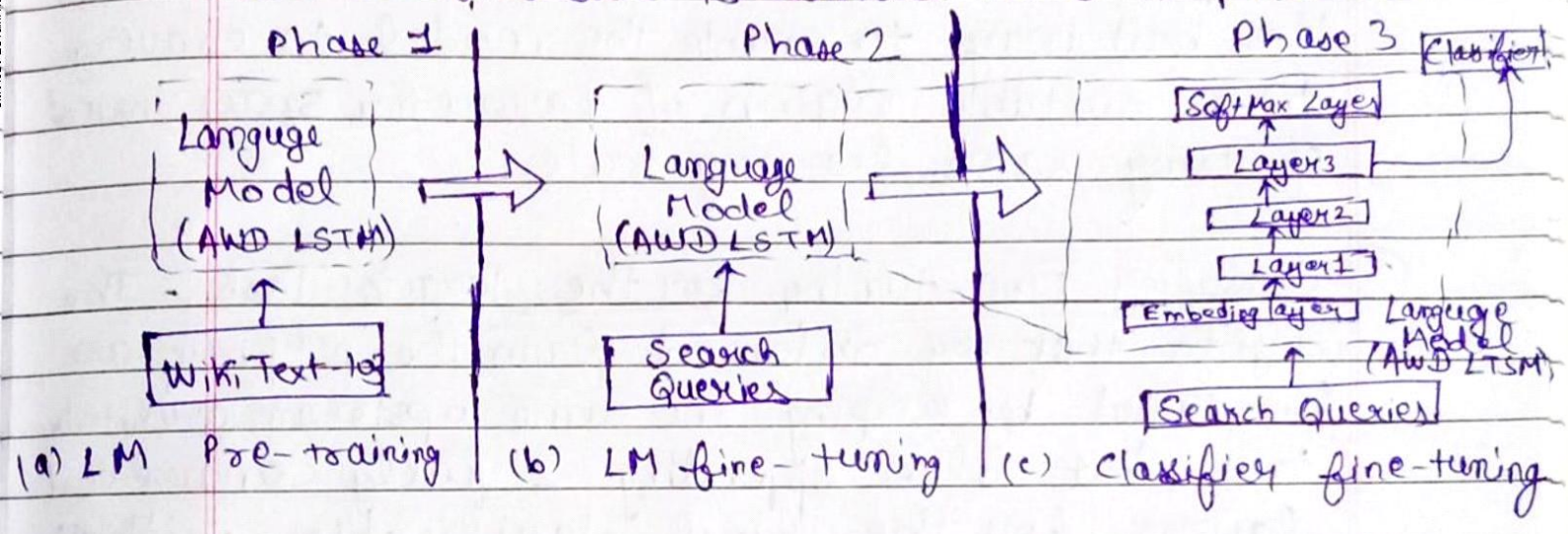
A query is annotated as well-formed if the supplied query is "grammatical" in nature, has "perfect spellings" and is an "explicit question". For each search query, the average of five scores (over each of the annotator's ratings) is calculated and then documented as the final rating R which indicates the degree of well-formedness. If $R \geq 0.8$, then it is ~~expected~~ considered to be well-formed. (as per authors

Ans 5 The model trained on a recently released benchmark dataset spanning 25,100 queries gives an accuracy of 75.03%, thereby improving by ~5 absolute percentage points over the state-of-the-art. The proposed "Inductive Transfer Learning" thus turned out to be better than various baselines and state-of-the-art methods like: "No LM fine tuning", "Fine tuning without DFT & STLR" etc. Also the paper showed that not fine-tuning the LM (Language Model) on the target task resulted in worst performance versus fine ~~tuning~~ tuning. Using DFT (Discriminative Fine Tuning) and STLR (Slanted Triangular Learning Rates) turned out to be beneficial. (a)
~~Thus, the idea of using~~ As the model performed with good accuracy (also beating the baselines) so we can say that the "idea of using inductive transfer learning by fine-tuning language models aids in identifying whether search queries are well-formed natural language questions

Ans 3 For a given query q, the proposed architecture intend to learn the label C which describes whether the query is a well-formed natural language question or not. The task is classified as binary classification task where $C=1$ implies well formed and vice versa for other case.
From some research, it was known that if language models are fine-tuned correctly, then they would not overfit to small datasets. So

the neural architecture ULMFiT model is adapted for inductive transfer learning approach. Also the AWD-LSTMs were shown to be effective in learning lower-perplexity language models, so the authors used it for their model.

The architecture is divided into 3 phases:-



① General Domain Pretraining:- The first phase involves pretraining a language model on a huge English corpus. This helps the model to learn the general language dependencies and is the 1st step before fine-tuning which targets specific data.

② Language Model Fine-Tuning for the Target Task:- Here the model is tuned in an un-supervised manner on a task-specific data. The fine-tuning involves discriminative fine-tuning (DFT) and slanted triangular learning rates (STLR) to combat the catastrophic forgetting language models.

DFT:- It involves different learning rates for all the layers of the AWD-LSTM, because

each of the layers represent a different kind of information, therefore they must be fine-tuned to different extents.

STLR:- The learning rate first increases and then linearly decays as the no of training samples increases in STLR. This is considered the best way to enable the model to converge to a suitable region of parameter space instead of using same learning rate.

(3) Classifier Fine-tuning for the Target Task:- The weights that are obtained from the 2nd phase are fine-tuned by keeping the same upstream architecture, but also appending 2 fully connected layers for the final classification with the last layer predicting the well-formed rating. "Gradual unfreezing" is one such heuristic for this task where all layers are not fine-tuned at the same time, instead the model is gradually unfrozen starting from the last year, as it contains the least general knowledge.

Ans 7 We have used binary classification for our model (Adam Optimizer), but for our case, we need to output into 6 classes and also giving the inputs into 6 classes (instead of the >0.8 rule). The 6 classes as per the question would be [0.0, 0.2, 0.4, 0.6, 0.8, 1.0]. We will use "~~multi-task~~" "multi-class task" classification and fine-tune accordingly to obtain good accuracy & F1 score.

The otherway round could be to treat the above problem as a regression problem and then rounding off the values to nearest of the allowed 6 class.

Stacking of different regression models can generate fantastic results.