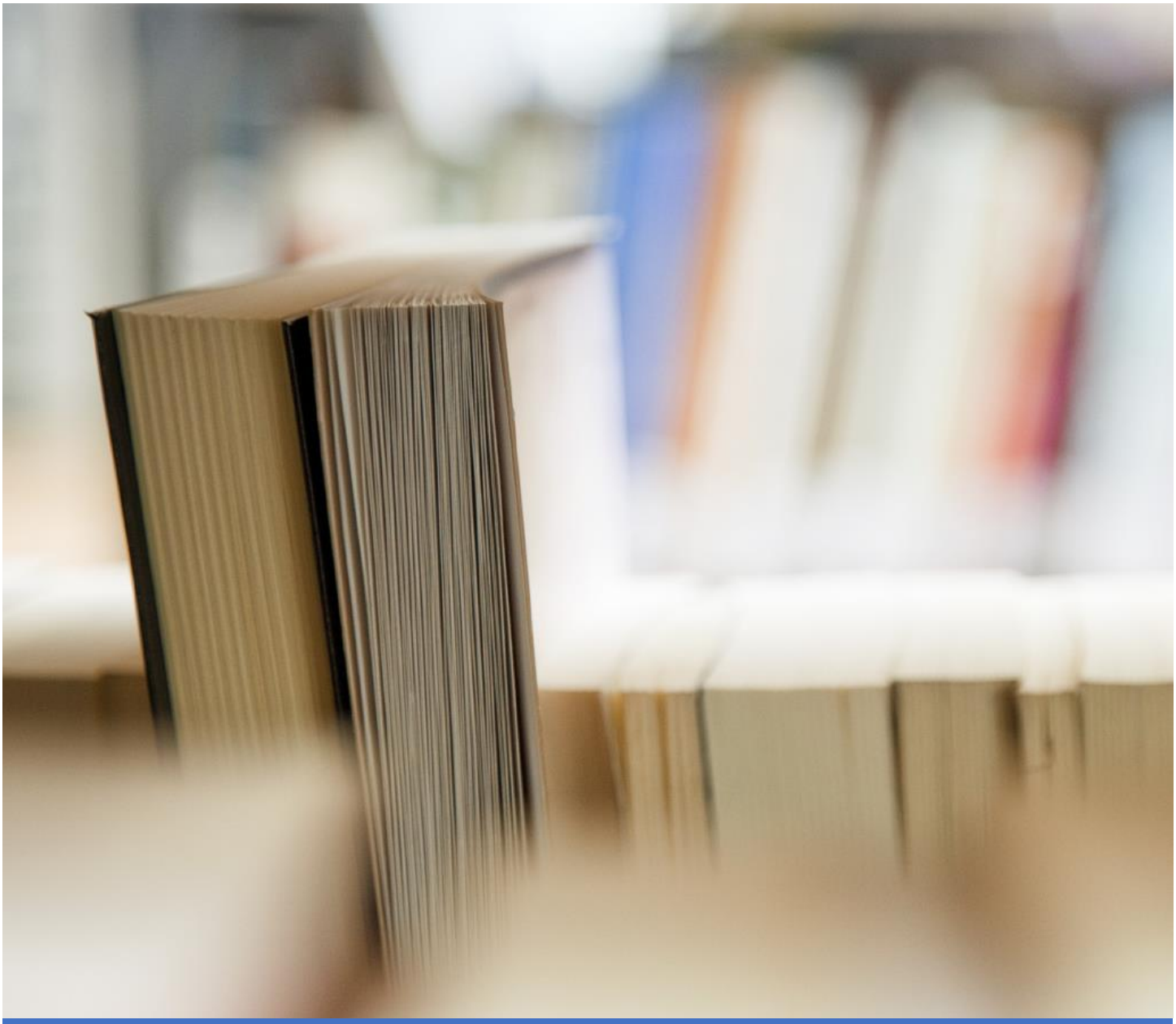

Dataset3 Project Report

Phase 2

Kiran Shanthappa

10/21/21

CSE 576

1 Contents

1. OBJECTIVE

Generate variants of existing natural instruction tasks by generating multiple alternate definitions, examples, and instances.

2. PREFACE

What you will be given: A collection of instruction tasks. GitHub Task List

What you need to generate: Multiple variants of each of the instruction tasks.

3. Existing Task List

Below are the existing task list provided.

No	Task file	Dataset name	Original Classification
1	task387_semeval_2018_task3_irony_classification.json	semeval	irony classification
2	task391_causal_relationship.json	causal	relationship
3	task392_inverse_causal_relationship.json	inverse	relationship
4	task393_plausible_result_generation.json	plausible	generation
5	task394_persianqa_question_generation.json	persianqa	generation
6	task395_persianqa_answer_generation.json	persianqa	generation
7	task396_persianqa_classification.json	persianqa	classification
8	task397_semeval_2018_task1_tweet_anger_detection.json	semeval	anger detection
9	task398_semeval_2018_task1_tweet_joy_detection.json	semeval	joy detection
10	task399_semeval_2018_task1_tweet_sadness_detection.json	semeval	sadness detection
11	task400_paws_paraphrase_classification.json	paws	classification
12	task401_numeric_fused_head_reference.json	numeric	head reference
13	task403_creak_commonsense_inference.json	creak	inference
14	task418_persent_title_generation.json	persent	generation
15	task419_persent_answer_generation.json	persent	generation
16	task420_persent_document_sentiment_classification.json	persent	sentiment classification
17	task421_persent_sentence_sentiment_classification.json	persent	sentiment classification
18	task422_persent_sentence_sentiment_verification.json	persent	sentiment verification
19	task423_persent_document_sentiment_verification.json	persent	sentiment verification

20	task424_hindienglish_corpora_hi_en_translation.json	hindienglish	en translation
21	task425_hindienglish_corpora_en_hi_translation.json	hindienglish	hi translation
22	task426_hindienglish_corpora_hi-en_classification.json	hindienglish	hi-en classification
23	task427_hindienglish_corpora_hi-en_language_identification.json	hindienglish	language identification

4. Variant Tasks

4.1 task387_semeval_2018_task3_irony_classification.json

"Definition": "In this task you are given a tweet that contains some form of irony. You must classify the type of irony the tweet has. Label the tweets (\"polarity\",\"situational\",\"other\") based on the irony they have. Situational irony happens when a situation fails to meet some expectations, Label these instances as \"situational\". polarity irony happens when irony is achieved by inverting the intended sentence, Label these instances as \"polarity\". There are other kinds of ironies that are neither polarity nor situational, Label these instances as \"other\".

4.1.1 "Alternate Definition 1":

"For the given irony in the sentence, classify them as "polarity", "situational", and "other". If a situation does not meet the expectation, then it is classified as "situation" irony. If irony is achieved by inverting the intended sentence then it is classified as "polarity". If a sentence cannot be classified neither as situational or polarity irony then it is classified as "other". "

4.1.2 "Alternate Definition 2":

"Categorise a sentence a "situational" irony if the sentence is not meeting some expectations or else as "polarity" irony if inverting the sentence causes the irony. Meeting neither of the category of irony is categorised as "other" irony."

4.1.3 "Alternate Definition 3":

"You must classify the type of irony the tweet has. Label the tweets "scenario", "contradiction", and "other" based on the irony they have. Scenario irony happens when a situation fails to meet some expectations, Label these instances as "scenario". Contradiction irony happens when irony is achieved by inverting the intended sentence, Label these instances as "contradiction". There are other kinds of ironies that are neither polarity nor situational, Label these instances as "other"."

4.2 task391_causal_relationship.json

"Definition": "In this task, you will be given two sentences separated by \", so\". You should decide whether the first sentence can be the cause of the second sentence. If you can see a possible causation, answer with \"plausible\", otherwise answer with \"not plausible\"."

4.2.1 "Alternate Definition 1":

"For the given statement try to identify the action it's possible reason for doing or taking that action. The reason is usually separated by ",so" from the action. The action is present as first part of the sentence

before “,so”. If you can identify the reason then it is called as “plausible” otherwise, classify it as “not plausible”.”

4.2.2 “Alternate Definition 2”:

“Separate the sentence into 2 parts based on the quoted “,so”. If first sentence is the cause of second sentence, then there a relationship between the 2 parts. Hence, classify this type of sentence as “plausible” otherwise as “not plausible”.”

4.2.3 “Alternate Definition 3”:

“Identify the cause from the given sentence with the separator as “,so”. If able to identify first sentence as cause then this is categorised as “plausible”, in other cases categorise the sentence as “not plausible”.”

4.3 task392_inverse_causal_relationship.json

"Definition": "In this task, you will be given two sentences separated by \"because\". You should decide whether the first sentence can be the result of the second sentence. If you can see a possible causation, answer with \"plausible\", otherwise answer with \"not plausible\"."

4.3.1 “Alternate Definition 1”:

“You are given a sentence and they are separated by “because”. Identify whether the first sentence can be because of the result of the second sentence. If able to identify first sentence as cause then this is categorised as “plausible”, in other cases categorise the sentence as “not plausible”.”

4.3.2 “Alternate Definition 2”:

“Separate the sentence into 2 parts based on the quoted “because”. If first sentence is the cause of second sentence, then there a relationship between the 2 parts. Hence, classify this type of sentence as “plausible” otherwise as “not plausible”.”

4.4 task393_plausible_result_generation.json

"Definition": "In this task, you will be given a sentence, followed by \", so\". You should complete the given text with another sentence, such that the whole text is plausible. The sentence you enter must be a likely result of the given sentence."

4.4.1 Missing Task

Create multiple choice for completion of a sentence.

4.4.2 “Alternate Definition 1”:

“For the given sentence followed by “so”, complete the sentence with another plausible sentence. The new sentence should be a probable cause.”

4.4.3 “Alternate Definition 2”:

“Complete the sentence with a plausible sentence which is effect of the given sentence.”

4.5 task394_persianqa_question_generation.json

"Definition": "In this task, you will be shown a Persian passage. You need to write a Persian question for the passage. Your question should be answerable based on the passage and only have one correct answer."

4.5.1 “Alternate Definition 1”:

“For the Persian passage given, create a question based on the passage and the answer should be present inside the passage.”

4.5.2 “Alternate Definition 2”:

“Form a question based on the passage given and the answer for the question should lie in the passage.”

4.5.3 “Alternate Definition 3”:

“You are given a passage in Persian. A question should be created based on the passage given alone and answer should always be present inside the Persian passage.”

4.6 task395_persianqa_answer_generation.json

"Definition": "In this task, you will be shown a Persian passage and question. You need to write a answer for the question. Try to keep your answers as short as possible."

4.6.1 Missing task

Translate Persian passage to English.

4.6.2 “Alternate Definition 1”:

“For the given passage and question in Persian, please find the answer for the question within the passage only.”

4.6.3 “Alternate Definition 2”:

“A question is given based on the passage in the Persian. Try to answer the question which is present in the passage.”

4.6.4 “Alternate Definition 3”:

“Answer the question based on passage given.”

4.7 task396_persianqa_classification.json

"Definition": "In this task, you will be shown a Persian passage and a question, and you have to determine whether the question is answerable based on the passage or not. If the question is answerable, choose the \"True\" label, and if not select \"False\""

4.7.1 “Alternate Definition 1”:

“Identify whether the given question can be answered based on the passage. If can be answered then classify it as true or else as false.”

4.7.2 “Alternate Definition 2”:

“If the question can answered based on passage then it is categories the question as true otherwise as false”

4.7.3 “Alternate Definition 3”:

“If you cannot answer the question based on the passage alone then it is classified as false otherwise as true.”

4.8 task397_semeval_2018_task1_tweet_anger_detection.json

"Definition": "In this task you are given a tweet. You must judge whether the author of the tweet is angry or not. Label the instances as \"Angry\" or \"Not angry\" based on your judgment."

4.8.1 "Alternate Definition 1":

"Check sentiment of the given statement is angry or not. Classify the statement as angry or not angry."

4.8.2 "Alternate Definition 2":

"Check if the given tweet is angry or not. Classify tweet as angry or not angry."

4.8.3 "Alternate Definition 3":

"Analyse the sentiment of the tweet as angry or not angry."

4.9 task398_semeval_2018_task1_tweet_joy_detection.json

"Definition": "In this task you are given a tweet. You must judge whether the author of the tweet is happy or not. Label the instances as \"Happy\" or \"Not happy\" based on your judgment."

4.9.1 "Alternate Definition 1":

"Check sentiment of the given statement is happy or not. Classify the statement as happy or not happy."

4.9.2 "Alternate Definition 2":

"Check if the given tweet is happy or not. Classify tweet as happy or not happy."

4.9.3 "Alternate Definition 3":

"Analyse the sentiment of the tweet as happy or not happy."

4.10 task399_semeval_2018_task1_tweet_sadness_detection.json

"Definition": "In this task you are given a tweet. You must judge whether the author of the tweet is sad or not. Label the instances as \"Sad\" or \"Not sad\" based on your judgment."

4.10.1 "Alternate Definition 1":

"Check sentiment of the given statement is sadness or not. Classify the statement as sadness or not sadness."

4.10.2 "Alternate Definition 2":

"Check if the given tweet is sadness or not. Classify tweet as sadness or not happy."

4.10.3 "Alternate Definition 3":

"Analyse the sentiment of the tweet as sadness or not sadness."

4.11 task400_paws_paraphrase_classification.json

"Definition": "In this task you are given a sentence pair from wikipedia that have high lexical overlap. If the sentences have the same meaning and are just paraphrases of each other label them as

\"Paraphrase\" , if not label them as \"Not paraphrase\". The two sentences are seperated by a new line character."

4.11.1 "Alternate Definition 1":

"You are given 2 sentences separated by the newline character \n. If the first sentence is paraphrase of the second sentence, then classify it as paraphrase otherwise classify it as not paraphrase."

4.11.2 "Alternate Definition 2":

"For the two-sentence given and separated by \n, analyse whether the second sentence is a paraphrase of first sentence. If the first sentence is paraphrase of the second sentence, then classify it as paraphrase otherwise classify it as not paraphrase."

4.11.3 "Alternate Definition 3":

"Identify whether second sentence is provide same inference or is a paraphrase of the first sentence. The two sentences are separated by \n."

4.12 task403_creak_commonsense_inference.json

"Definition": "In this task you are given a statement and an explanation giving you further knowledge about an entity in the statement. You must judge whether the statement is true or false based on the explanation. Label an instance as \"True\" if the explanation confirms the statement or doesn't disprove it. Label an instance as \"False\" if the explanation disproves the statement. The statement and explanation are separated by a newline character."

4.12.1 "Alternate Definition 1":

"You are given s statement followed by an explanation separated by \n. Decide if the first statement is correct or wrong fact based on the second statement. If the statement is correct, then label it as true otherwise label it as false. "

4.12.2 "Alternate Definition 2":

"A two part sentence is given with a separator \n in between. Analyse and provide your judgement whether the first statement factually correct based on the second statement. If is it factually correct then instance is inferred as true otherwise as false."

4.12.3 "Alternate Definition 3":

"A two part sentence is given with a separator \n in between. With the common knowledge from the second sentence confirm whether first sentence is true if facts in both the sentences match or doesn't contradicts. Otherwise classify the instance as false."

4.13 task418_persent_title_generation.json

"Definition": "Given a document, generate a short title of the document. The title should convey the main idea/event/topic about which the document is being written."

4.13.1 "Alternate Definition 1":

"Generate an apt title for the given document. The title should convey the major point of discussion."

4.13.2 “Alternate Definition 2”:

“A right title for a paragraph always should convey the main idea or event from the paragraph. Provide the title for the given paragraph.”

4.13.3 “Alternate Definition 3”:

“Provide the main idea or event as the title for the given paragraph.”

4.14 task419_persent_answer_generation.json

"Definition": "Given a document, find the main entity about whom the author is writing "

4.14.1 “Alternate Definition 1”:

“For the given passage, identify the main character author is discussing about.”

4.14.2 “Alternate Definition 2”:

“You are given a passage. Carefully analyse and mention the personality who the author is writing about.”

4.14.3 “Alternate Definition 3”:

“Passage on a person is given to you. Identify that main person the passage talks about.”

4.15 task420_persent_document_sentiment_classification.json

"Definition": "Given a document and an entity the task is to select the authors sentiment towards the enity. Sentiments can be Positive, Neutral and Negative. Select Positive if the article expresses a positive view towards the given entity or praises its quality or skills. Select Neutral if the document expresses no clear view towards the entity or has equal amounts of positive and negative statements or expressing some fact/quote by someone else. Select Negative if the article expresses a negative view towards like harsh remarks, criticizing entities action/decision etc."

4.15.1 “Alternate Definition 1”:

“Identify whether the paragraph about an entity is positive if only happy, joyful and praises are provided. If any harsh comments, or criticism is provided then classify the paragraph as negative. If paragraph has both positive and negative comments, then it is a neutral statement.”

4.15.2 “Alternate Definition 2”:

Classify the para into positive, negative, and neutral sentiment. Negative sentiment is when para is showing harsh language or criticizing an action or entity. Positive sentiment is when para is showing good deeds or action, enthusiastic or happy comments on an entity. Neutral if para has both positive and negative sentiments.

4.15.3 “Alternate Definition 3”:

“Sentiment of the document has to be understood and classified as positive, neutral, and negative. You can consider a neutral sentiment when document presents same no. of positive and negative statements. If you find more joyous statements, happy mood from the document then it is considered as the positive. If document only presents the criticism then it is considered as negative.”

4.16 task421_persent_sentence_sentiment_classification.json

"Definition": "Given a sentence and an entity, the task is to select the authors sentiment towards the enity. Sentiments can be Positive, Neutral and Negative. Select Positive if the sentence expresses a positive view towards the given entity or praises its quality or skills. Select Neutral if the sentence expresses no clear view towards the entity or has equal amounts of positive and negative statements or expressing some fact/quote by someone else. Select Negative if the sentence expresses a negative view towards like harsh remarks, criticizing entities action/decision etc."

NOTE: duplicate of the task420. There is no difference apart from the word sentence and document.

4.17 task422_persent_sentence_sentiment_verification.json

"Definition": "Given a sentence, an entity and its sentiment towards the entity, verify if it is the correct sentiment towards the entity. Answer should be yes or no."

NOTE: duplicate of the task420. There is no difference apart from positive and negative changed to "yes" or "no".

4.18 task423_persent_document_sentiment_verification.json

"Definition": "Given a document, an entity and its sentiment towards the entity, verify if it is the correct sentiment towards the entity. Answer should be yes or no."

NOTE: duplicate of the task420. There is no difference apart from the word sentence and document.

4.19 task425_hindienglish_corpora_en_hi_translation.json

"Definition": "In this task, you are given a sentence in the English language and your task is to convert it into the Hindi language. In translation, keep numbers as it is."

4.19.1 "Alternate Definition 1":

4.19.2 "Alternate Definition 2":

4.19.3 "Alternate Definition 3":

4.20 task426_hindienglish_corpora_hi-en_classification.json

"Definition": "In this task, you are given a sentence in the Hindi language and a corresponding English translation of the Hindi sentence. Your task is to generate a label \"Yes\" if the translation is correct, otherwise generate label \"No\". In the translation, English sentence should preserve the number as it is and it should be in sentence case (capitalize only the first word of each sentence and noun)."

4.20.1 "Alternate Definition 1":

"Translate the paragraph from English to Hindi by keeping the numbers as same."

4.20.2 "Alternate Definition 2":

"Produce a translation in hindi for the given text in English. Make sure to keep the number as it is."

4.20.3 "Alternate Definition 3":

"English text has to be translated into hindi. Make sure to keep the content same except for the the number given in the passage."

4.21 task427_hindienglish_corpora_hi-en_language_identification.json

"Definition": "In this task, you are given a sentence which is either in the Hindi language or English language. You task is to identify the language of input sentence. Input sentence can be in Hindi or English language only and also it cannot have two languages at a time."

4.21.1 "Alternate Definition 1":

"Judge whether the translation from hindi to English is accurate by mentioning correct translation as yes and incorrect translation as no. Here number in the paragraph should be same after translation. A translation is correct if the implied meaning of the sentence is same in both languages. "

4.21.2 "Alternate Definition 2":

"You have to whether the text in hindi and English produce same meaning, purpose and context. If so then it is marked as yes otherwise as no. Make sure that numbers are same hindi and English otherwise mark it as no."

4.21.3 "Alternate Definition 3":

"Text in english produce same meaning as in text given in hindi. Here number is kept same between both languages. If both of the previous rules are broken then mark the instance as no otherise as yes."

4.22 Approach

The task definition transformation can be achieved through various methods such as paraphrasing, back-translation, syntactic sentiment transformation, etc. The most successful transformation with many variants can be achieved better through paraphrasing.

There are many paraphrasing methods. In this project following paraphrasing technique is utilized due to its efficiency and simplicity, when compared with other techniques.

1. Parrot Paraphrasing
   - Parrot is a paraphrase-based utterance augmentation framework purpose built to accelerate training NLU models. A paraphrase framework is more than just a paraphrasing model.

Ref: Data Professor and Prithivraj Damodaran
https://github.com/PrithivirajDamodaran/Parrot_Paraphraser

Colab code uploaded to GitHub -
https://github.com/ZoroKiran/CSE576HW/blob/main/Dataset3_parrot_paraphrase.ipynb

2. Pegasus Paraphrase
   - PEGASUS is an acronym for Pre-training with Extracted Gap-sentences for Abstractive SUmmarization Sequence-to-sequence models. PEGASUS model used herein is from Huggingface's transformers library

     Ref: Data Professor
     https://github.com/dataprofessor/python/blob/main/transformer_pegasus_paraphrase.ipynb

     Colab code uploaded to GitHub -
     https://github.com/ZoroKiran/CSE576HW/blob/main/dataset3_paraphrase_kiran.ipynb

5. Missing Task

6. References

Ref: Data Professor and Prithivraj Damodaran
https://github.com/PrithivirajDamodaran/Parrot_Paraphraser

Ref: Data Professor
https://github.com/dataprofessor/python/blob/main/transformer_pegasus_paraphrase.ipynb