“I’d rather just go to bed”: Understanding Indirect Answers

1. Summarize the paper.

The paper titled “I’d rather just go to bed”: Understanding Indirect Answers addresses the challenge of interpreting indirect responses to yes/no questions in dialogues. The authors present a large-scale English corpus named Circa, containing 34,268 pairs of polar questions and indirect answers. These pairs include a variety of responses such as yes/no, uncertain, middle-ground, and conditional answers. The dataset was created via crowdsourcing to capture natural language usage. The paper introduces BERT-based models to classify these question-answer pairs into meaning categories. The models achieved 82-88% accuracy for a 4-class distinction and 74-85% for a 6-class distinction. While transfer learning from entailment tasks showed some promise, the models still fall short of robust performance for practical dialogue systems.

2. How many models are trained to get the numbers of one row under the matched setting? What do you train and test with?

In the matched setting, one model is trained to get the numbers of one row. It's trained and tested using 60% of the Circa corpus as training data, 20% as development data, and the remaining 20% as test data.

3. How many models are trained to get the numbers of one row under the unmatched setting? What do you train and test with?

In the unmatched setting, ten models are trained to get the numbers of one row. Each model is trained using a leave-one-out approach across the 10 scenarios. During testing, a different scenario is left out in each of the 10 runs, and the remaining nine scenarios are used for training. The reported numbers are the average performance across these 10 models.

4. Explain (your own words) what do you need to do to replicate the results highlighted in red in the screenshot. Then, run the experiment to replicate the results. Write a short report.

Steps to replicate the experiment:

- Data Preparation:
  Download the Circa dataset from Hugging Face Datasets. Split the data into training, validation, and test sets. And then select the training dataset to replicate the highlighted in red in the screenshot.

- Model Setup:
  Use the BertForSequenceClassification model from the Transformers library by Hugging Face.
  Configure the model for 4-class classification.

- Training:
  Fine-tune the model on the training set using the following hyperparameters:
  Learning rate: 2e-5
  Number of epochs: 3
  Batch size: 16

- Evaluation:
  Evaluate the model on the test set.
  Use classification_report from sklearn.metrics to calculate accuracy, precision, recall, and F1 scores.

BERT Model Training and Evaluation Report:

Dataset: Circa dataset, comprising various features like context, questions, answers, and labels.
Model: BERT (Bidirectional Encoder Representations from Transformers).
Training, Testing and Development Dataset split into 60-20-20 for training and validation. BERT fine-tuned using AdamW optimizer and linear scheduler for three epochs.
Evaluation: Model evaluated on the validation set, achieving an accuracy of 77.95%.
Conclusion: BERT model shows promise for sequence classification on Circa dataset.

Difficulties encountered:
Replicating the code posed challenges mainly in preprocessing. Adapting the data for BERT required meticulous tokenization and encoding, ensuring alignment with the model's input requirements. Handling label filtering and ensuring compatibility with BERT's format posed initial hurdles, demanding careful attention to detail and alignment with the model's specifications.

RESULTS:

Model	**Matched Setting**				
	Accuracy	**Test F-Score**			
	Test	**Yes**	**No**	**C.Yes**	**Mid**
BERT finetuned on Question or on Answer					
BERT-YN (Question only)	55.9	64.7	50.7	2.5	5.5
BERT-YN (Answer only)	81.7	84.0	80.4	89.0	23.5
BERT finetuned on Question + Answer					
BERT YN	87.5	89.5	87.8	88.5	33.0