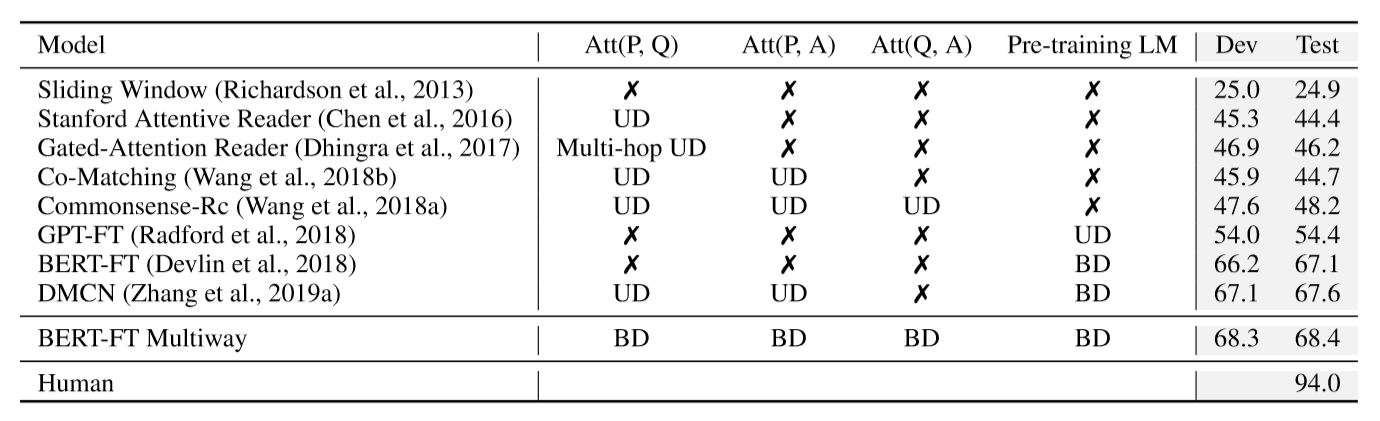
**COSMOS QA: Machine Reading Comprehension with Contextual Commonsense Reasoning**

**Model and Approach Analysis:**

There are two types of Baseline models. Reading Comprehensions and modifications on it and pretrained models.

1. Sliding Windows
2. 2.Stanford Active Reader
3. Gated Attention Reader
4. Co Matching
5. Common Sense RC
6. GPT-FT
7. GPT-FT
8. BERT-FT
9. DMCN



In Reading comprehension approach, semantic correlatedness is important factor to choose an answer from given answers where it infers from the given contextual paragraph about semantic correlations.

In COSMOS-QA dataset, it contains 83% of answers which is not in reading comprehension context so semantic relatedness factor is not important here to infer the correct answer as it requires common sense to infer answer, while pretrained methods improves the scenarios further while apply finetuning on BERT. Further, more accurate results can be achieved by performing attention and finetuning on context paragraph, Question and Answer.

Ablation is also one of the important parts of the study where ablation of question didn't affect much in prediction result while ablation of question and context affects significantly on the result and got drop in accuracy

**Knowledge Transfer Learning**

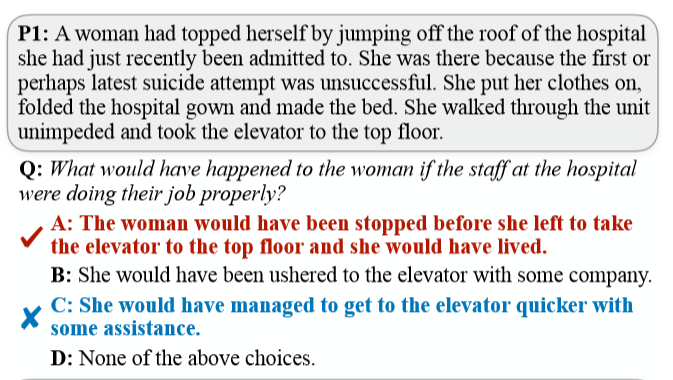
Knowledge transfer and finetuning on various datasets of same kind help a lot to improve inference. Authors have proposed two datasets RACE and SWAG which contains multiple choice questions, and fine tuning of BERT on both Race + SWAG and Cosmos. BERT-FT on SWAG provides good result while including with BERT FT RACE+SWAG which gives 68.7 percent test accuracy

**Error Analysis**

* Complex context understanding

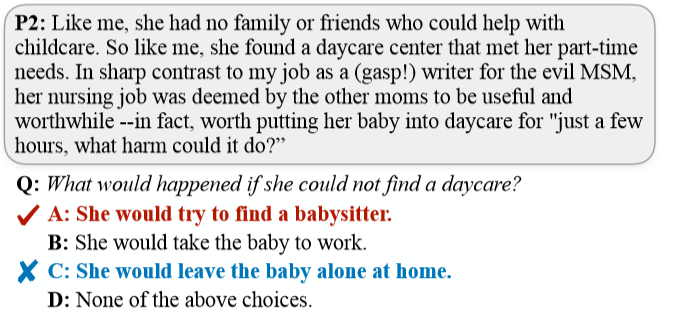
It require cross sentence interpretation and reasoning and need to combine the context information to infer the real answer. Here model need to learn from complex context analysis

to infer the choice A.



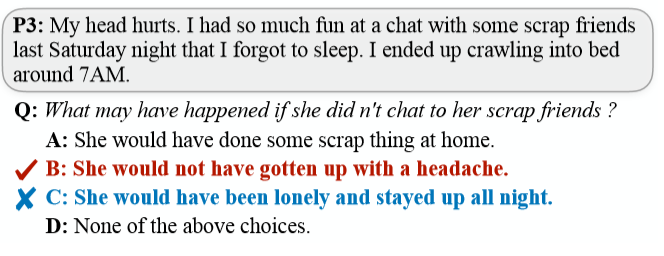
* **Inconsistent with Human Common Sense**

In 33% of the errors, the model mistakenly select the choice which is not consistent with human common sense. So here answer might be right but not match with human commonsense.



* **Multi-turn Common Sense Inference**

19% errors due to this where there are multiple inferences present in sentence where model need to choose with proper inference.



* **Unanswerable Questions**

14% of the errors are from unanswerable questions so model cannot infer from given multiple answers.

**Generative models to infer answers**

Fine tuning on GPT and GPT-2 can provide better understanding and provide more accurate answer which can be consider as one of the different approach.

