Quiz_for_day06

1. Why did TDIDF model in TVQA paper show good performance? (2.5 points)

<from paper>
Since our questions are raised by people watching the videos, it is natural for them to ask questions about specific and unique objects/locations/etc., mentioned in the subtitle. Thus, it is not surprising that **TFIDF** based similarity between answer and subtitle performs so well.

<답>

비디오를 보는 사람들이 질문을 제기하기 때문에 자막에 언급 된 구체적이고 고유한 객체 / 위치 등에 관한 질문을하기 때문입니다.

2. What is the main difference between TVQA and TVQA+? (In terms of model) (2.5 points)

<from paper>
While the TVQA dataset provides a novel question format and temporal공 annotations, it lacks spatial grounding information, i.e., bounding boxes of the concepts (objects and people) mentioned in the QA pair. We hypothesize object annotations could provide additional useful training source for models to gain a deeper understanding of visual information in TVQA. Therefore, to complement the original TVQA dataset, we collect frame-wise bounding boxes for visual concepts mentioned in the questions and correct answers.
...
The TVQA+ datasetis unique as it contains three different annotations: questionanswering, temporal localization, and spatial localization

TVQA TASK
QA = Question Answering, TL = Temporal Localization

TVQA공+ TASK
QA = Question Answering, TL = Temporal Localization, SL = Spatial Localization.

<답>

Spatial Localization task를 위한 310.8K bounding box 제공