Assignment 08/06 – 서승우

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

Question들이 비디오를 본 사람들에 의해서 제시되었기 때문에, 주로 유일하거나, 구체적인 장소, 물체 등에 관한 것들이 많아서 자막만 보더라도 충분히 유추해낼 수 있는 질문들이다.

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

이전까지의 모델들은 질문에 대답만 하는 형식이었으나, TVQA에서는 질문에 대답하는 것 뿐만 아니라 대답과 연관있는 spatiotemporal한 정보까지 예측하도록 하였음. Explicit Attention supervision loss를 통해서 spatial한 정보를 맞게 보고 있는지 확인하였고, Span prediction으로 제대로된 시간에서 참고하여 대답했는지 확인하였다.