

성능 개선 방법론

검색 알고리즘 개선 4가지 방안

4. Solar Embedding 모델

- 질문과 문서를 모두 임베딩처리
- 질문별로 문서와 매치하여 스코어를 확인
- TOPK 3~7를 받아옴
- Elasticsearch로 맵핑하여 검색(진행되지 못함)

Result

- 성능향상에는 영향이 있었음
- 기존 제출 점수보다는 높지 못함
- 문장 단위가 아닌 문서 단위 임베딩 차이 학습함

```
stages_answers = []
for j in range(0, 220):
    if j == chat[0]-1:
        del chat[0]
        continue
    else:
        question_embedding = (question_embeddings 10)
        best_i = 0
        second_i = 0
        third_i = 0
        best_score = 0
        second_score = 0
        third_score = 0
        for i in range(0, 4272): #0 ~ 4272
            print(j+1, "번 질문", i+1, "번 문서 확인", end = "\n")

            document_embedding = ast.literal_eval(document_embeddings[i])
            similarity = np.dot(np.array(question_embedding), np.array(document_embedding))

            if best_score < similarity:
                third_score = second_score
                second_score = best_score
                best_score = similarity
                third_i = second_i
                second_i = best_i
                best_i = i

            elif best_score == similarity:
                print("같은 값이 존재합니다.", similarity)
                best_score = similarity
                best_i = i
            else :
                continue
        tmp = str(str(best_i+1) * ", " * str(second_i+1) + ", " + str(third_i+1))
        stages_answers.append(tmp)
```