

## 1. 내가 수강한 이유는?

AI 관련 지식을 좀더 깊이 공부해보고 싶었고

특히, 기업에서 제공해주는 과제들을 통해 실전 경험을 체험해보고 싶었습니다.

## 2. [NLP] Text Classification과 Dialogue Generation on Persona-Chat 정리 해보기

### 2-1. Text Classification

-task 설명

텍스트 문서나 문장을 분류하는 task입니다.

이 task 안에도 감정분류, 뉴스분류 등등 다양한 task들이 존재합니다.

-데이터 세트

**AG News**라는 데이터셋이 대표적이며 많이 사용하여 다양한 분류 문제들을 해결할 수 있습니다.

데이터 구성으로 **World, Sports, Business, Sci/Tech** 4개의 카테고리 뉴스들로 이루어져 있으며

**제목과 본문으로 30,000개의 train데이터와 1,900개의 test**로 이루어진 데이터입니다.

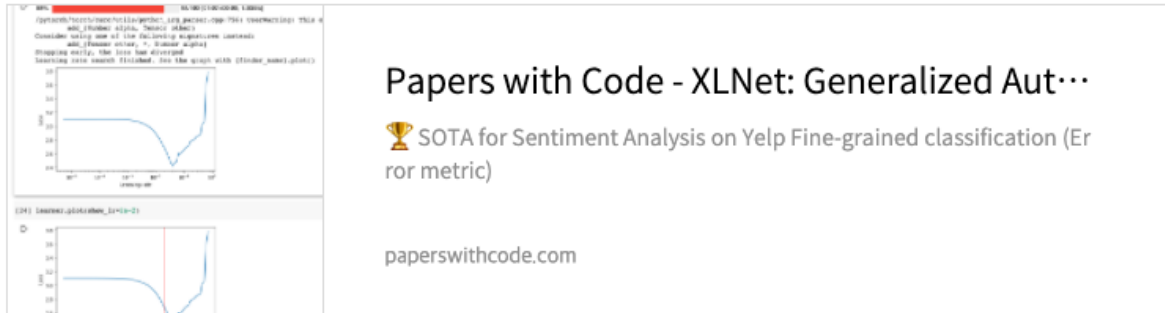
<https://paperswithcode.com/dataset/ag-news>

<p>ic: Tech</p> <p>a: r PC May Be Less Secure Than You Thi</p> <p>escription: st users think their computer is safe fr ware and spyware--but they're wrong vey conducted by Internet serv vider America Online found that cent of home computers were infected</p>	<h3>Papers with Code - AG News Dataset</h3> <p>AG News (AG's News Corpus) is a subdataset of AG's corpus of news ar ticles constructed by assembling titles and description fields of artic...</p> <p><a href="https://paperswithcode.com">paperswithcode.com</a></p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

-sota모델

XLNet: Generalized Autoregressive Pretraining for Language Understanding(2019)가 가장 대표적인 모델이라고 합니다.

<https://paperswithcode.com/paper/xlnet-generalized-autoregressive-pretraining>



## 2-2.Dialogue Generation on Persona-Chat

-task 설명

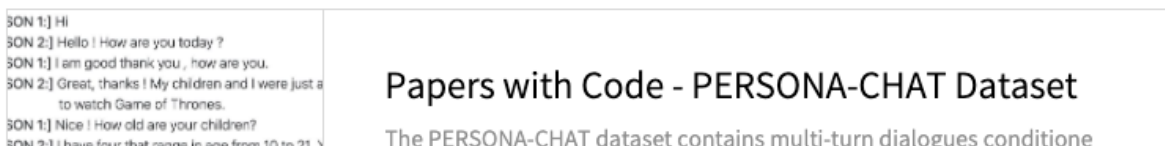
애플의 siri와 네이버의 클로바처럼 사람과 인공지능 챗봇이 페르소나가 있는 대화를 가능하게 합니다.

-데이터 세트

PERSONA-CHAT이 대표적인 데이터 세트로 페르소나가 존재하는 여러 문단의 2명의 대화 형식의 데이터가 있습니다.

train, valid, test 총 3개의 데이터가 나누어져 있으며 각각 8393, 1000, 968개의 대화가 각각 저장되어 있습니다.

<https://paperswithcode.com/dataset/persona-chat-1>



SON 1:] I don't have Cheetos at the moment.  
 SON 2:] That just means you get to keep all the pop  
 for yourself.  
 SON 1:] And Cheetos at the moment!  
 SON 2:] Good choice. Do you watch Game of Thron  
 SON 1:] No, I do not have much time for TV.  
 SON 2:] I usually spend my time painting; but, I love  
 show.

d on personas. The dataset consists of 8939 complete dialogues for ...

[paperswithcode.com](https://paperswithcode.com)

-sota 모델

대표적인 모델로 **P<sup>2</sup> Bot(2020)** 과 **TransferTransfo(2019)**이 있습니다.

<https://paperswithcode.com/paper/you-impress-me-dialogue-generation-via-mutual>



## Papers with Code - You Impress Me: Dialog...

1 code implementation in PyTorch. Despite the continuing efforts to i  
 mprove the engagingness and consistency of chit-chat dialogue sys...

[paperswithcode.com](https://paperswithcode.com)

<https://paperswithcode.com/paper/transfertransfo-a-transfer-learning-approach>

speaker2]	您	好	[speaker1]
+			
speaker2]	[speaker2]	[speaker2]	[speaker1]
+			
4	5	6	7

## Papers with Code - TransferTransfo: A Tra...

#2 best model for Dialogue Generation on Persona-Chat (Avg F1 metri  
 c)

[paperswithcode.com](https://paperswithcode.com)

## 손지아

모델 리서치에 있어서 핵심적인 내용을 간결하게 작성해주셔서 이해하기 수월했습니다. 특히, sota 모델 키워드들을 적어주신 것이 인상깊습니다.

---

## 정태호

NLP가 음성 데이터들과도 같이 처리할 수 있다는 점이 흥미로웠습니다. 또한, 깊고 간결하게 모델과 데이터를 소개해주셔서 좋았습니다.

---

## 현승환

영화 데이터 관련하여서 자세히 설명해주셔서 좋았습니다. 추가적으로 모델과 데이터 관련한 출처를 올려주시면 더 좋을 것 같아요!