NLP
Project & Study

조민제(11기/경제학과 18)
이지현(11기/언어학과 16)
최정윤(12기/영어영문학과 16)

Contents

1) Paper Review

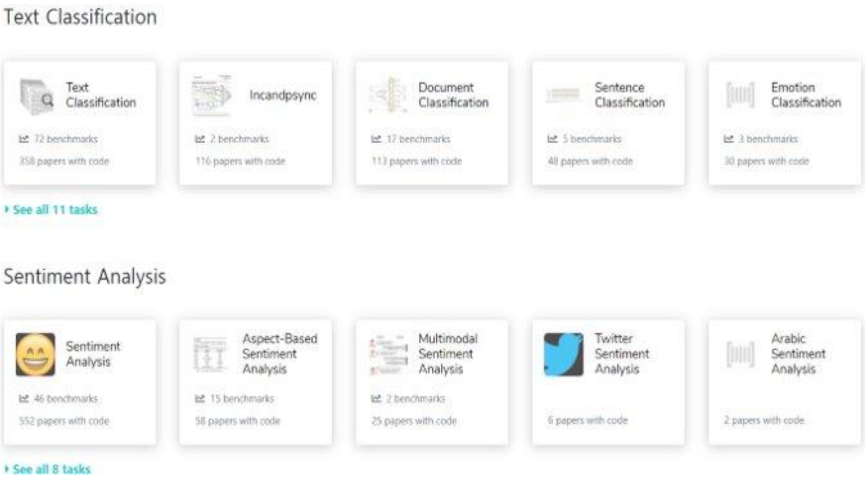
2) Pytorch

3) NLP projects

Paper Review

- NLP 관련 최신 논문 및 코드 리뷰(https://paperswithcode.com)

- Topics: Classification, Multi-label Classification, Sentiment Analysis



Text Classification

Text Classification
72 benchmarks
358 papers with code

Incandpsync
2 benchmarks
116 papers with code

Document Classification
17 benchmarks
113 papers with code

Sentence Classification
5 benchmarks
48 papers with code

Emotion Classification
3 benchmarks
30 papers with code

‣ See all 11 tasks

Sentiment Analysis

Sentiment Analysis
46 benchmarks
552 papers with code

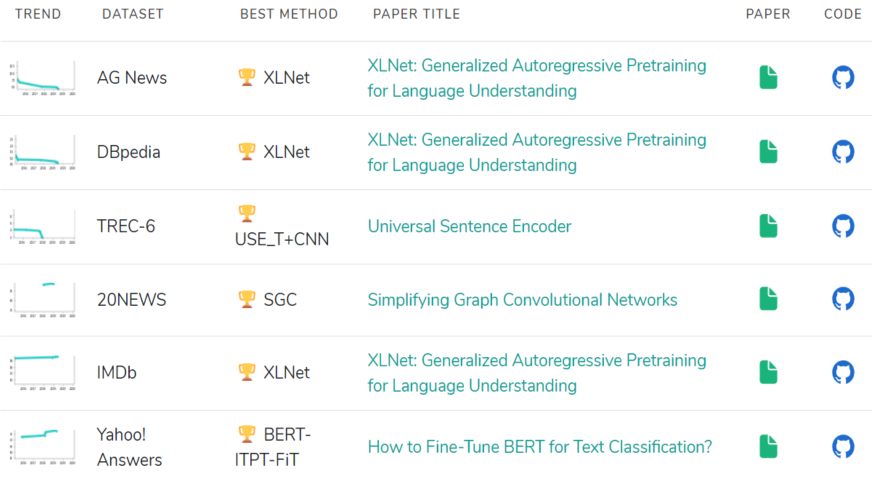
Aspect-Based Sentiment Analysis
15 benchmarks
58 papers with code

Multimodal Sentiment Analysis
2 benchmarks
25 papers with code

Twitter Sentiment Analysis
6 papers with code

Arabic Sentiment Analysis
2 papers with code

‣ See all 8 tasks



TREND	DATASET	BEST METHOD	PAPER TITLE	PAPER	CODE
	AG News	XLNet	XLNet: Generalized Autoregressive Pretraining for Language Understanding		
	DBpedia	XLNet	XLNet: Generalized Autoregressive Pretraining for Language Understanding		
	TREC-6	USE_T+CNN	Universal Sentence Encoder		
	20NEWS	SGC	Simplifying Graph Convolutional Networks		
	IMDb	XLNet	XLNet: Generalized Autoregressive Pretraining for Language Understanding		
	Yahoo! Answers	BERT-ITPT-FiT	How to Fine-Tune BERT for Text Classification?		

Pytorch

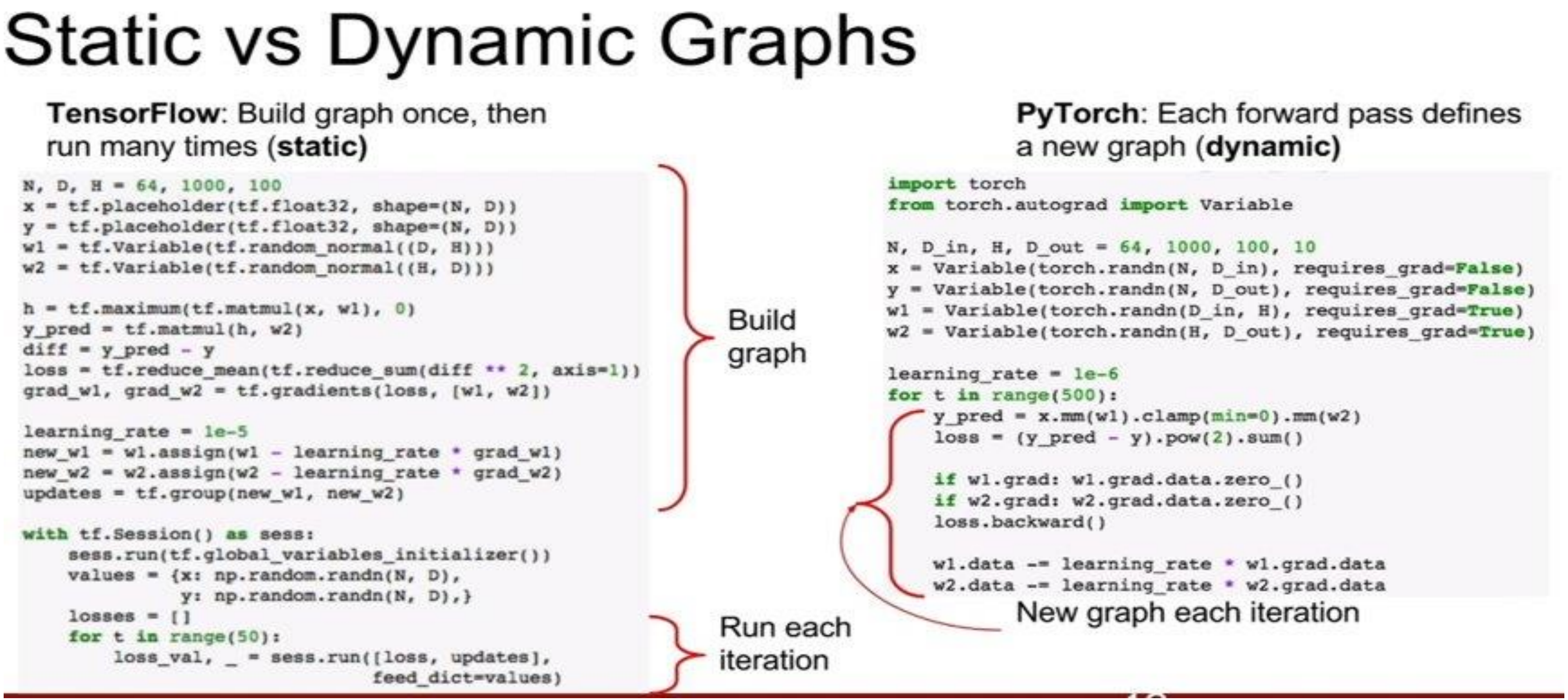
- 딥러닝을 위한 필수 프레임워크인 Pytorch 입문.

- Tensorflow(Define-and-Run) vs Pytorch(Define-by-Run)

Idea Factory Intensive Program #2
딥러닝
홀로서기
#28
이론강의 PyTorch실습 코드리뷰

Me passionately talking about why Pytorch is the superior Deep Learning framework despite no one listening

Pytorch



Static vs Dynamic Graphs

TensorFlow: Build graph once, then run many times (**static**)

```
N, D, H = 64, 1000, 100
x = tf.placeholder(tf.float32, shape=(N, D))
y = tf.placeholder(tf.float32, shape=(N, D))
w1 = tf.Variable(tf.random_normal((D, H)))
w2 = tf.Variable(tf.random_normal((H, D)))

h = tf.maximum(tf.matmul(x, w1), 0)
y_pred = tf.matmul(h, w2)
diff = y_pred - y
loss = tf.reduce_mean(tf.reduce_sum(diff ** 2, axis=1))
grad_w1, grad_w2 = tf.gradients(loss, [w1, w2])

learning_rate = 1e-5
new_w1 = w1.assign(w1 - learning_rate * grad_w1)
new_w2 = w2.assign(w2 - learning_rate * grad_w2)
updates = tf.group(new_w1, new_w2)

with tf.Session() as sess:
    sess.run(tf.global_variables_initializer())
    values = {x: np.random.randn(N, D),
              y: np.random.randn(N, D),}
    losses = []
    for t in range(50):
        loss_val, _ = sess.run([loss, updates],
                               feed_dict=values)
```

Build graph

Run each iteration

PyTorch: Each forward pass defines a new graph (**dynamic**)

```
import torch
from torch.autograd import Variable

N, D_in, H, D_out = 64, 1000, 100, 10
x = Variable(torch.randn(N, D_in), requires_grad=False)
y = Variable(torch.randn(N, D_out), requires_grad=False)
w1 = Variable(torch.randn(D_in, H), requires_grad=True)
w2 = Variable(torch.randn(H, D_out), requires_grad=True)

learning_rate = 1e-6
for t in range(500):
    y_pred = x.mm(w1).clamp(min=0).mm(w2)
    loss = (y_pred - y).pow(2).sum()

    if w1.grad: w1.grad.data.zero_()
    if w2.grad: w2.grad.data.zero_()
    loss.backward()

    w1.data -= learning_rate * w1.grad.data
    w2.data -= learning_rate * w2.grad.data
```

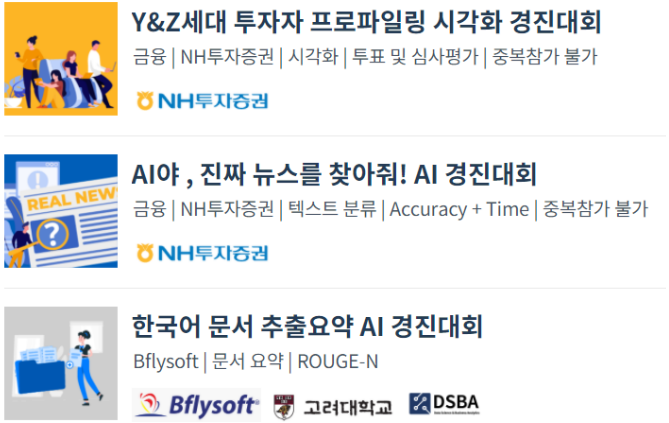
New graph each iteration

NLP projects

- 학기 중 최소 1개의 공모전에 참가 예정

- 겨울 방학에 NLP 스터디를 이어가며 다른 장기 프로젝트에 도전

https://www.mzcechackathon.co.kr
감정 대화 분류
MZ CEC
Classification of Emotional Conversation
해커톤 대회
딥러닝 기술 기반 감정 대화 분류
2020.11.9(월) ~ 11.20(금)
온라인 11.19(목) ~ 11.20(금)
NIA 한국정보화진흥원
MEDIAZEN



Y&Z세대 투자자 프로파일링 시각화 경진대회
금융 | NH투자증권 | 시각화 | 투표 및 심사평가 | 중복참가 불가
NH투자증권

AI야 , 진짜 뉴스를 찾아줘! AI 경진대회
금융 | NH투자증권 | 텍스트 분류 | Accuracy + Time | 중복참가 불가
NH투자증권

한국어 문서 추출요약 AI 경진대회
Bflysoft | 문서 요약 | ROUGE-N
Bflysoft 고려대학교 DSBA

Thank you