

Project 말따라미



Vol. 1

빅데이터(답러닝) 활용 AI 설계 8반

강태형

멀티캠퍼스



목차

CONTENTS

01

말따라미?

02

주제 선정 이유

03

구현 계획

04

Vol 2, 3...

말따라미?

01

말따라미

말따라미?

말 + 따라하다 + 미

→ 말을 따라하는, 그러나 조금은 ‘다르게’ 따라하는 모델!

말따라미?

01

Vol.1 : 성경 데이터 이용

INPUT DATA : 현대어 말뭉치

TARGET DATA : 고어 말뭉치

이제 내가 새로운 계명을
너희에게 준다.
서로 사랑하여라.
내가 너희를 사랑한 것처럼
너희도 서로 사랑하여라.



새 계명을 너희에게 주노니
서로 사랑하라
내가 너희를 사랑한 것 같이
너희도 서로 사랑하라

학습!

말따라미?

01

Vol.1 : 성경 데이터 이용

NEW INPUT DATA

NEW TARGET DATA(OUTPUT)

오늘 점심 메뉴는 어떻게
좋을까?



<성경 '처럼' 말하는 문장>

Inference!

주제 선정 이유

자연어 처리 모델을 구현하고 싶은데,

번역은 이미 다 되어있고

챗봇은 너무 박셀 거 같다!

좀 재미있는 거 없나?

주제 선정 이유

1. 재미있다
2. 학습 데이터가 존재한다
3. 실제 서비스가 가능하다
4. 자연어 처리 프로젝트를 해보고 싶었다(한글)
5. Transformer Network를 사용한다

구현 계획

- Data Preprocessing

Kharii 형태소 분석기 기반 Tokenizing

- Modeling

KERAS / TENSORFLOW 기반 자연어 처리

END-TO-END Network 구축

- Deploy

Django Web Framework

Flask server를 이용한 Telegram chatbot service

Vol 2, 3...

Vol 2 : 북한말 말따라미

Vol 3 : 사투리 말따라미

- 현재는 두 언어 다 쉽게 학습 코퍼스를 구할 수 없는 상황
- 북한말 코퍼스 : 정부주도 하 구축 사업 진행중
- 사투리 코퍼스 : STT opensource를 활용하여 학습 데이터 추출
- 다양한 시대와 공간의 말을 따라하다 → 말따라미!

00

마지막

번외 : 사람 생각이 다 거기서 거기다


이미 2019. 09에 seq2seq model 기반 prototype 완성

위 주제로 Final Project를 할 생각을
진작에 가지고 있었으나...

00

말따라미

번외 : 사람 생각이 다 거기서 거기다

**김성현**
인기 게시물 작성 멤버 · 10월 15일 오후 6:31

일하기 싫어서 재미로 학습해본 seq2seq 모델을 공유합니다 😊
텍스트를 입력하면 성경 말투(?)로 변환해줘요 ㅋㅋㅋㅋㅋㅋ

<https://github.com/MrBananaHuman/TextToBible>

[57] prediction('오늘은 빨리 퇴근 하고 싶다.')

Bible: 오늘 은 속히 도망 하 기 를 원하 더라

[59] prediction('오늘 저녁엔 치킨을 먹을까?')

Bible: 오늘 저녁 에 대하 여 어찌 하 여 마음 을 사모 하 느니라

[61] prediction('생각보다 성능이 그렇게 좋지는 않다')




Bible: 전능 하 니 자 의 전능자 보다 낮 지 아니하 였 느니라

[62] prediction('내일 날씨는 매우 맑고, 시원하겠습니다.')


Bible: 내일 날 이 심히 크 고 끝날 에 빛 이 청결 하 며


[64] prediction('올 한해가 두 달 밖에 남지 않았다는 사실을 믿을 수가 없습니다.')


Bible: 지체 하 지 아니하 니 것 을 우리 가 가 밀 지 아니하 였 노라

 Jeong Yeon Lia Nam, 외 436명

댓글 32개 공유 104회

 좋아요

 댓글

 공유하기



번외 : 사람 생각이 다 거기서 거기다

00

따라미

딥 러닝 네트워크 공부 차원

다양한 버전 제공

성능 향상

을 염두에 두고 개발할 예정



감사합니다.

THANK YOU.

멀티캠퍼스