

Deep Learning Chat System 개발

도전!

도전!



도전!

Copy (Many)Code Line by Line

while(understand==false) Read reference, blog, paper..

print(Log), print(Log), print(Log) ...

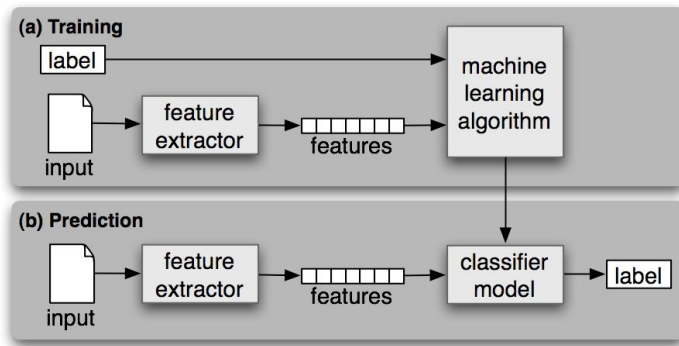


개발!

개발!

Intend Inference

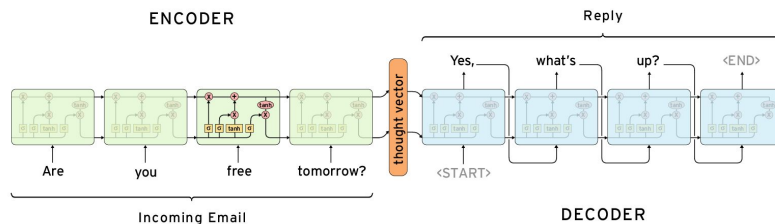
- Gensim
- NaiveBayesClassifier
- Term-existence



딥러닝

Seq2Seq

- TensorFlow
- RNN, Attention, GRU, Buckets,



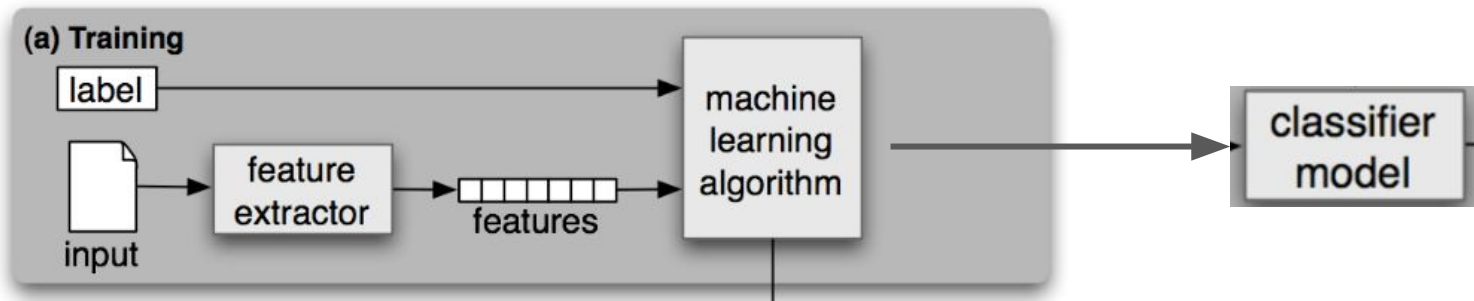
개발!

Intend Inference - 학습

사진 검색	여름에 찍은 사진만 골라서 보여줘	SEARCH	30개
	어제 찍은 사진 뭐 있어?	SEARCH	
	꽃이 나온 사진만 보여줘	SEARCH	
	바다가 나온 사진	SEARCH	
	내가 나온 사진들만 보자	SEARCH	

사진 전송	철리에게 카카오톡으로 보내줘	SEND	30개
	소피한테 방금 분류한 사진들 전송	SEND	
	이 사진들 파이한테 보내줘	SEND	
	앞으로 이런 사진들 매주마다 닉한테 전송해줘	SEND	

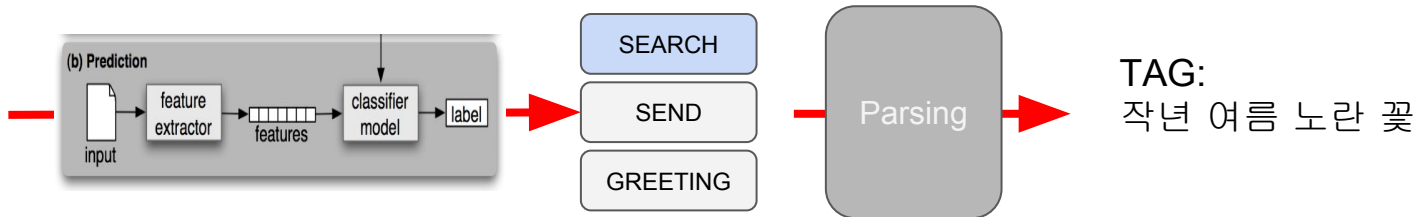
(a) Training



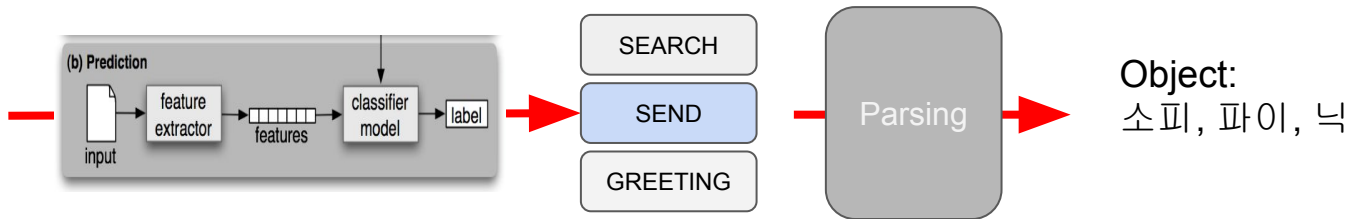
개발!

Intend Inference - 예측

Q: “작년 여름에 찍은 노란
꽃
사진 보여줘”



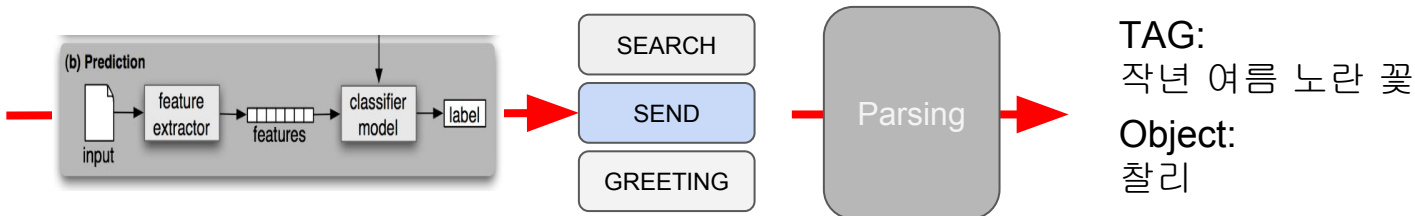
Q: “소피하고 파이하고
닉한테 보내줘”



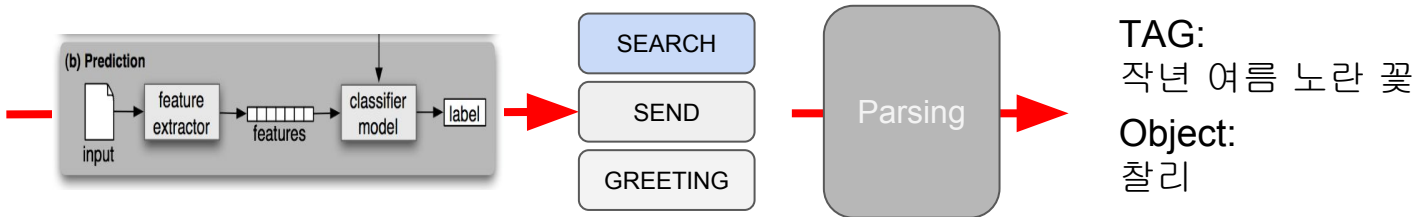
개발!

Intend Inference - 예측(예외)

“작년 여름에 찍은 노란 꽃
사진 찰리한테 보내줘”



“작년 여름에 찍은 노란 꽃
사진 찰리한테 보내줘”

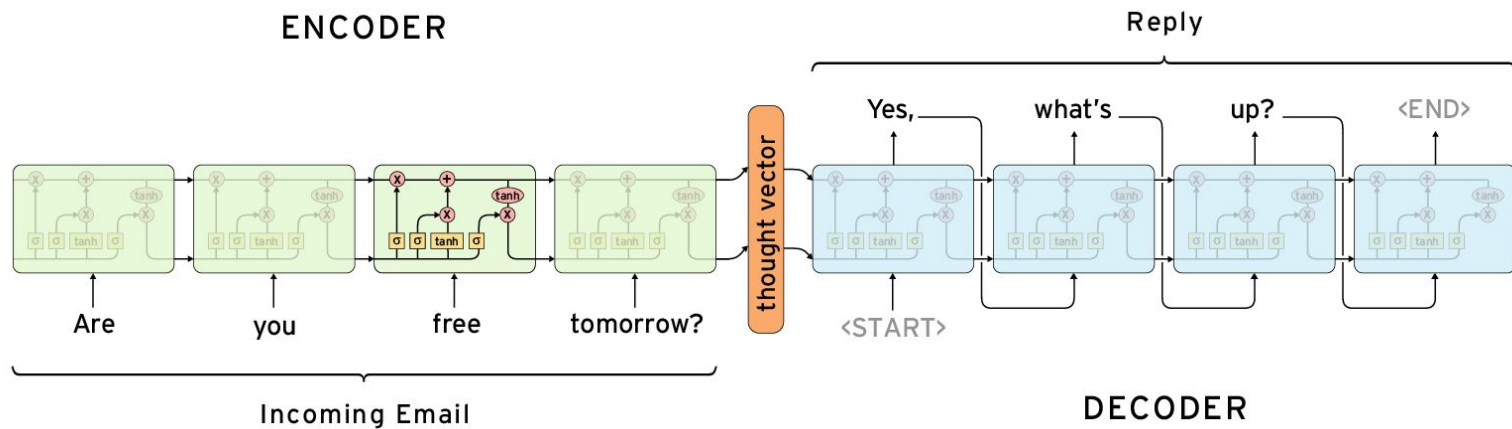


“안녕”



개발!

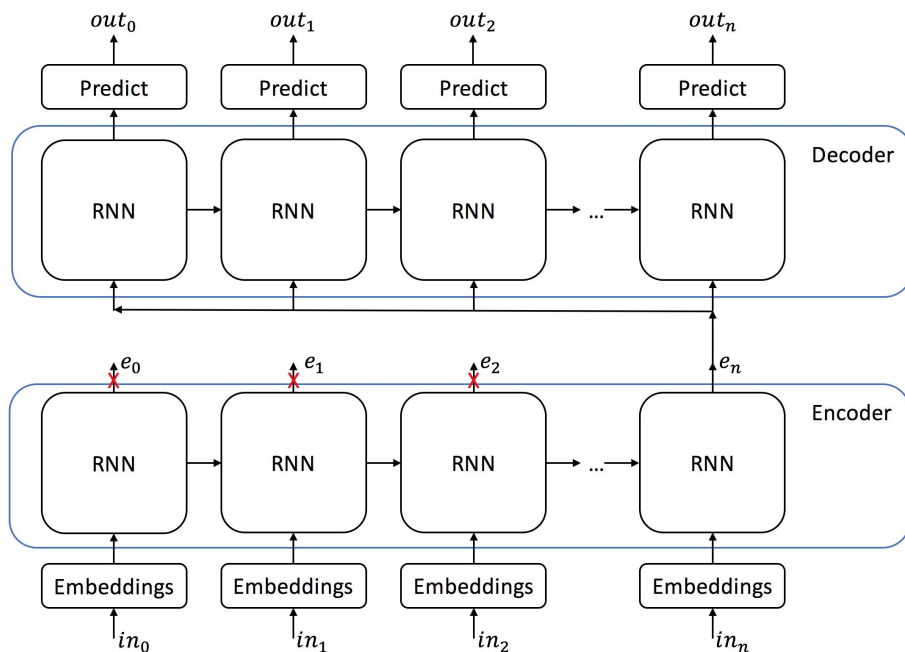
Seq2Seq - 설명



개발!

Seq2Seq - 설명

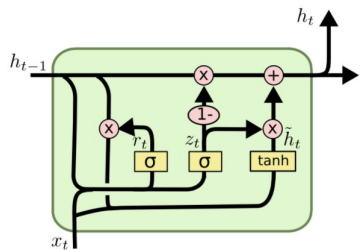
With Attention:



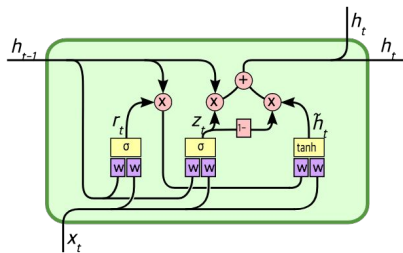
개발!

Seq2Seq - 설명

Cell:



LSTM



GRU

Bucket:

$[(8, 10), (12, 14), (16, 19)]$

Sampled Softmax, Projection ... 저도 잘 이해가..

개발!

Seq2Seq - 학습

일반적으로

- Closed Model에서도 최소 10000쌍 이상..

Cornell Movie-Dialogs Corpus

- **220,579** conversational exchanges between
- **10,292** pairs of movie characters
- **9,035** characters from 617 movies
- **304,713** total utterances
- Very well-formatted (almost perfect)

3.2 데이터 수집

시나리오의 데이터를 수집하기 위해서 간단한 대화를 할 수 있는 모델을 만들 필요가 있었다. 이를 위해서 30명 이상의 사람으로부터 10,000여개의 대화 쌍을 수집하였다. 최대한 중복되는 대화 쌍을 줄이고, 나올 수 있는 대화를 다양하게 하기 위해 각 사람들은 6개의 세부 시나리오 중 하나에 할당되어 300쌍, 전체 시나리오에서 300쌍을 만들도록 했다. 각 대화 쌍은 아이가 먼저 질문이나 이야기를 꺼내고 모델 및 로봇이 그에 대해 대답하도록 구성되어 있다. 이 데이터를 가지고 기초 모델을 만든 후, 랩 투어, 학회 등에서 시연을 하며 기록된 로그 데이터를 통해 계속해서 새로운 데이터를 수집하였다.

우리는?

- 데이터가 없음.. 이상한 데이터 100쌍정도가 다임 ㅠㅠ

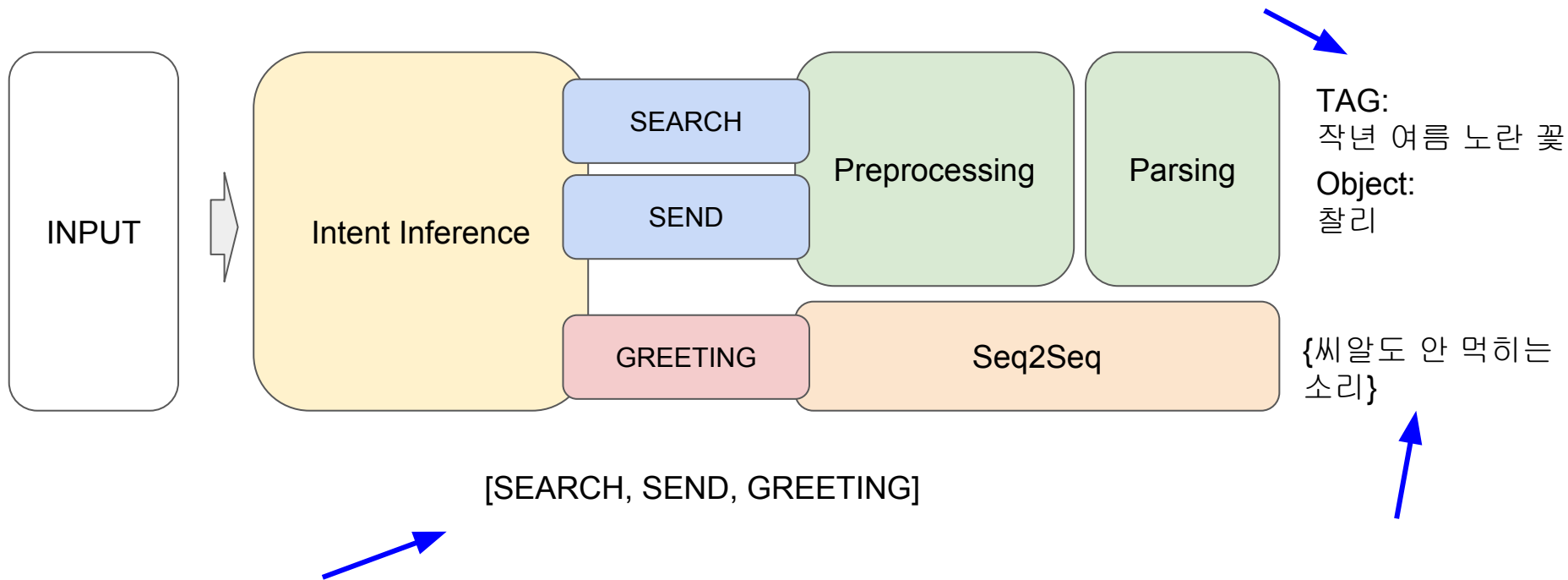
AI 란 무엇입니까?	인공 지능은 생각하는 기계를 만드는 데 전념하는 공학 및 과학의 한 분야입니다.			
어떤 언어를 쓰셨습니까?	파이썬			
너는 데이터처럼 들린다.	예, 나는 지휘관 데이터의 인격적 인성에 영감을 받았습니다.			
다시요 이고 어떤 식별이 나	그게 내 이름이야			

개발!

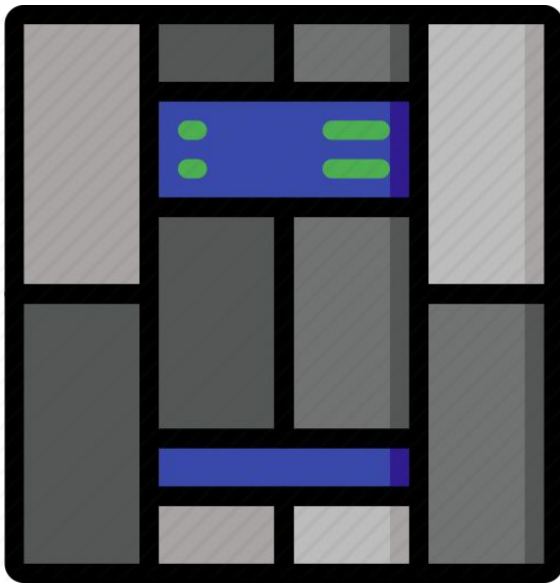
Seq2Seq - 예측



결과물!!



결과물!!



TARS: <http://13.124.109.145:5000/TARS/>
(Text Analyze Recognition System)