

Image Captioning in Korean Language

멀티미디어개론 기말프로젝트 Proposal

자연어처리연구실

허광호

Image Captioning in Korean Language

• 물체 인식 + 문장 생성



입력 이미지

생성된 문장

- a group of girls are playing in a soccer game .
- kids are playing in the game on the soccer ball .
- two men trying to kick a soccer ball .
- a group of boys playing in a soccer game .
- a group of girls are playing a soccer ball .



일반적인 물체 인식과 비교했을 때 어려운 점은
정적인 이미지에서 움직임을 나타내는 kick, play와 같은 동사를 추출하고
자연어의 문법과 world knowledge를 고려하여 문장을 생성해야 함

Background

- 다양한 학회에서 많이 연구되고 있음
- MSCOCO Dataset
 - 123,287 Images
 - 80 object categories
 - 5 captions per image.
 - 2015 COCO Captioning Challenge

M1	Percentage of captions that are evaluated as better or equal to human caption.
M2	Percentage of captions that pass the Turing Test.
M3	Average correctness of the captions on a scale 1-5 (incorrect - correct).
M4	Average amount of detail of the captions on a scale 1-5 (lack of details - very detailed).
M5	Percentage of captions that are similar to human description.

평가 기준

출처: <http://mscoco.org/dataset/#captions-leaderboard>

	M1	M2	M3	M4	M5	date
Human ^[5]	0.638	0.675	4.836	3.428	0.352	2015-03-23
Google ^[4]	0.273	0.317	4.107	2.742	0.233	2015-05-29
MSR ^[11]	0.268	0.322	4.137	2.662	0.234	2015-04-08
Montreal/Toronto ^[10]	0.262	0.272	3.932	2.832	0.197	2015-05-14
MSR Captivator ^[12]	0.25	0.301	4.149	2.565	0.233	2015-05-28
Berkeley LRCN ^[2]	0.246	0.268	3.924	2.786	0.204	2015-04-25
m-RNN ^[6]	0.223	0.252	3.897	2.595	0.202	2015-05-30
Nearest Neighbor ^[13]	0.216	0.255	3.801	2.716	0.196	2015-05-15
PicSOM ^[15]	0.202	0.25	3.965	2.552	0.182	2015-05-26
Brno University ^[3]	0.194	0.213	3.079	3.482	0.154	2015-05-29
m-RNN (Baidu/ UCLA) ^[7]	0.19	0.241	3.831	2.548	0.195	2015-05-26
MIL ^[8]	0.168	0.197	3.349	2.915	0.159	2015-05-29
MLBL ^[9]	0.167	0.196	3.659	2.42	0.156	2015-04-10
NeuralTalk ^[14]	0.166	0.192	3.436	2.742	0.147	2015-04-15
ACVT ^[1]	0.154	0.19	3.516	2.599	0.155	2015-05-26
Tsinghua Bigeye ^[17]	0.1	0.146	3.51	2.163	0.116	2015-04-23
Random ^[16]	0.007	0.02	1.084	3.247	0.013	2015-05-29

2015 Captioning Challenge 결과

목표

- 전제적인 목표
 - 이미지를 입력했을 때 그럴듯한 한국어 문장이 생성되도록 함.
- 1차목표
 - 2015 COCO Challenge에 참가했던 팀 모델을 재구현
 - 재구현한 모델이 생성한 영어문장을 한국어 문장으로 변환
 - 번역기 사용
- 2차목표
 - MSCOCO 데이터의 Caption을 한국어로 변환하여 모델 재훈련
 - 번역기와 형태소분석기 사용
 - 한국어 문장 생성 모듈을 개선
 - 생성된 한국어 문장을 (문법적 오류) 교정

Progress (1)

- Image Captioning Model Training

- Google's "ShowAndTell" Model

- 이미지 모델 (CNN)

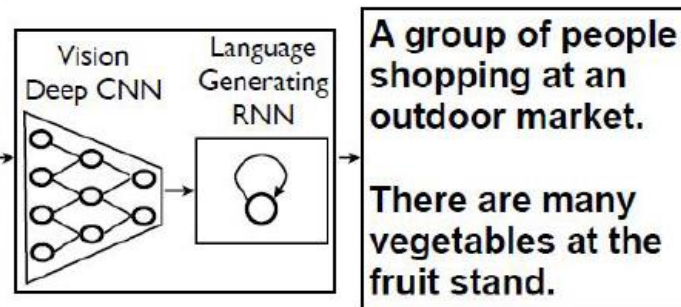
- GoogleLeNet 모델 + Batch Normalization 기법
- ImageNet 데이터 Top-5 에러율 4.8%

- 캡션 생성 모델 (RNN)

- LSTM (Long-short Term Memory)

- MSCOCO 데이터 사용

- 이미지 – 8.2만 장
- 캡션 – 41만 문장 (이미지 당 5 문장)

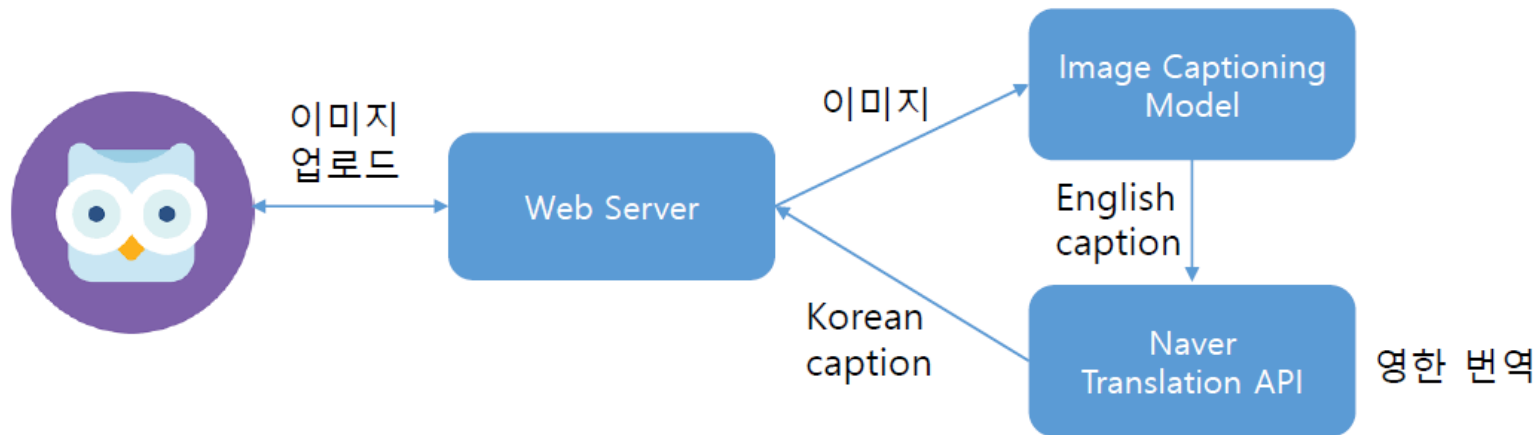


Show and Tell: Lessons learned from the 2015 MSCOCO Image Captioning Challenge

Progress (2)

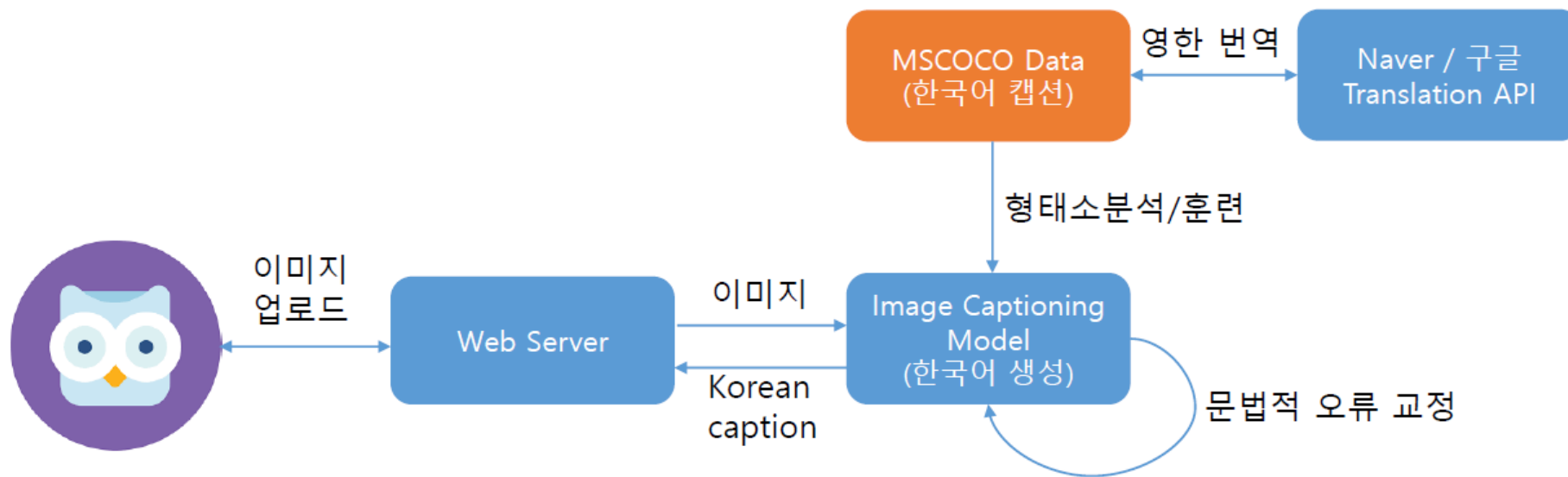
- Naver Papago API 연동
 - Generated caption (English) → 영한 번역 → Korean caption
- Demo 웹 서버
 - 이미지 업로드
 - Captioning model 연동

시스템 구조



Todos

- MSCOCO 데이터의 Caption을 한국어로 변환하여 모델 재훈련
 - 번역기와 형태소분석기 사용
- 한국어 문장 생성 모듈을 개선
- 생성된 한국어 문장을 (문법적 오류) 교정



최종 작업내역

- A) 이미지 캡션 생성 모델
 - 이미지 업로드 + 영문 캡션 생성
 - 한국어 캡션 생성 (네이버 Papago, Google Translate)
- B) MSCOCO 코퍼스 영문 캡션 → 한국어 캡션으로 번역
 - 구글 번역기 사용 - (전체 41만 중 현재 17만 완료)
- C) 한국어 캡션 태깅 툴
 - 번역된 한국어 캡션 선택/평가
 - 새 캡션 작성/저장

A) 이미지 캡션 생성 모델

파일 선택 선택된 파일 없음

업로드



이미지 분석

Analyzed Image Captions

- a large passenger jet flying through a blue sky .
- a large passenger jet flying through a cloudy sky .
- a large passenger jet flying through the blue sky .



- 큰 여객기는 푸른 하늘을 날고 있다.
- 많은 승객을 태운 비행기 구름 낀 하늘을 날고 있다.
- 큰 여객기가 푸른 하늘을 날고 있다.



- 푸른 하늘을 통해 비행하는 대형 여객기.
- 흐린 하늘을 통해 비행하는 대형 여객기.
- 푸른 하늘을 통해 비행하는 대형 여객기.

B) MSCOCO 코퍼스 영문 캡션 번역

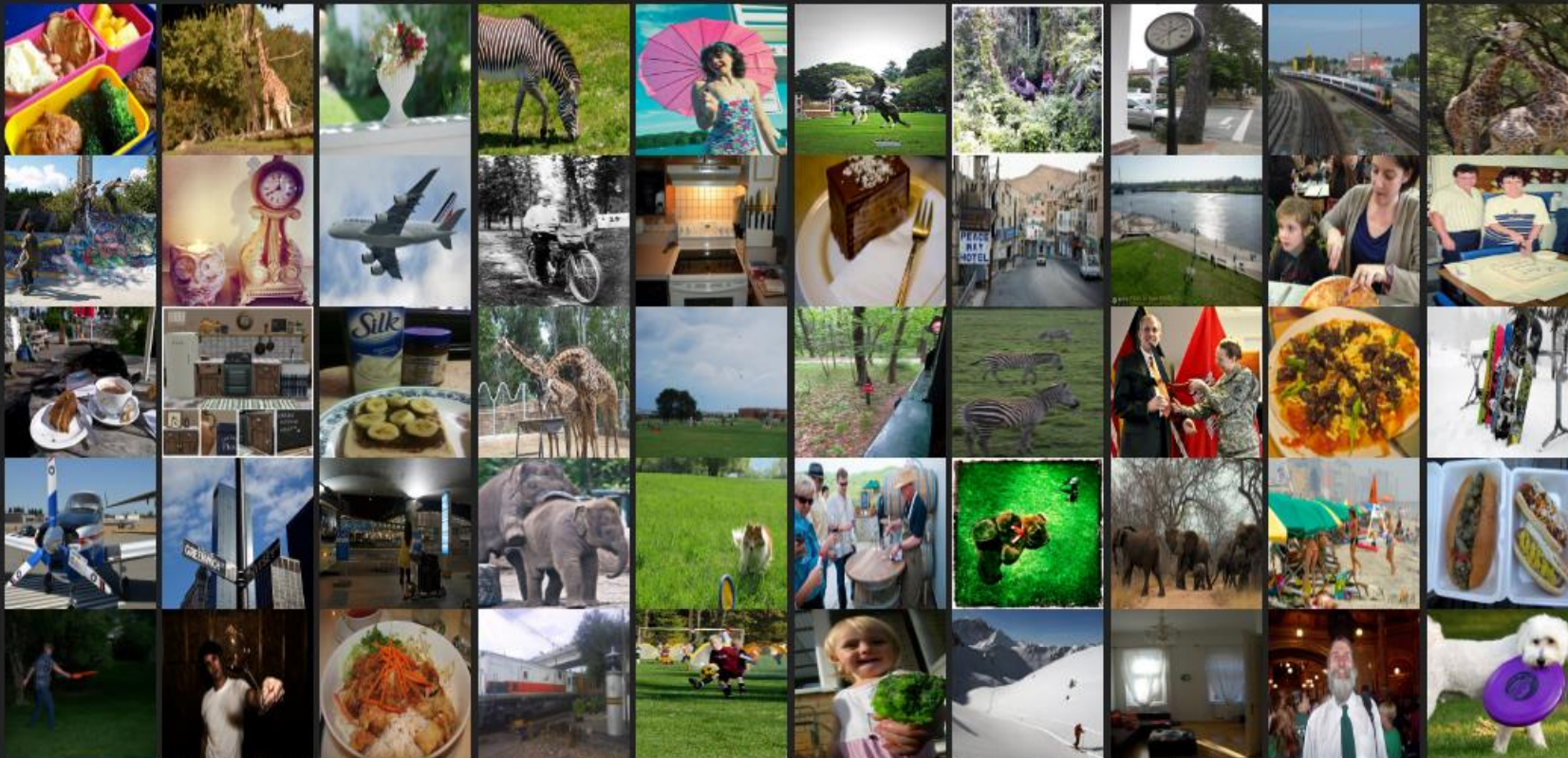
- Google-translate-token 사용
 - <https://www.npmjs.com/package/google-translate-token>
 - 번역 text에 대한 인증키 생성함
 - \$20/1백만 chars → 무료..
- Parallel 자동 번역
 - 전체 41만 중 현재 17만 완료

```
177159 Bananas out from of store that appear freshly picked and ripening.
177160 a bunch of bananas are leaning against a wall
177161 two guys are talking to each other in front of a building
177162 a close up of a street sign with a building in the background
177163 A city intersection has several signs on a pole.
177164 A large fancy old building with a number of direction signs on a po
177165 Two men standing on a street corner of a European city.
177166 A herd of sheep are inside of the pen
177167 a couple of sheep that are in side a fence
```

```
177159 바나나는 신선하게 고른 익은 것으로 보이는 가게에서 나옵니다.
177160 바나나 잔뜩 벽에 기대어있다.
177161 두 사람이 건물 앞에서 서로 이야기하고있다.
177162 가까이 백그라운드에서 건물 거리 표지판의 최대
177163 도시 교차로에는 기둥에 여러 표지판이 있습니다.
177164 포 그라운드에서 기둥에 방향 징후와 대형 멋진 오래 된 건물.
177165 유럽 도시의 거리 모퉁이에 서있는 두 남자.
177166 양 무리는 펜 안에 있습니다.
177167 담장 옆에있는 두 마리의 양
177168
```

C) 한국어 캡션 태깅 툴 - 이미지 갤러리

MSCOCO 한국어 캡션 수집 툴



C) 한국어 캡션 태깅 툴 - 수집 화면

MSCOCO 한국어 캡션 수집

COCO_train2014_000000001323.jpg



MSCOCO 영문 캡션

- An orange cat standing on the top of a brown horse.
- An orange cat standing on the back of a dark horse.
- a cat stands on top of a horse
- A cat standing on top of a black horse.
- A cat sands proudly and full of valor a top a horse

자동번역된 한국어 캡션 (구글번역)

- 갈색 말의 위에 서있는 오렌지 고양이.
- 어두운 말의 뒤에 서있는 오렌지 고양이.
- 말 위에 고양이가 서있다.
- 검은 말의 위에 서있는 고양이.
- 고양이는 자랑스럽게 모래와 용기로 가득 차있다.

새 캡션을 만들어주세요.

저장