

Now Meal.

식재료 이미지 딥러닝 인식 기반 음식 추천 서비스

냉장고 파먹기

shimmerlee13@gmail.com



목차 **a table of contents**

- 1** 서비스 소개
- 2** 개발과정
- 3** 상세구현내용
- 4** 차후 과제



Part 1,

서비스 소개

Introduce **Now Meal**

지금 먹는 건강한 한 끼

식재료 이미지 딥러닝 인식 기반 레시피 추천 서비스

! POINT !

사용자 보유 재료 이미지 인식

+

사용자의 취향, 재료/요리 히스토리 반영 추천



주제 Background

직접 요리하는 인구 증가, 요리 관련 모바일 이용자수 증가, 포스트 코로나로 인해 집에서 요리하는 트렌드 확산



요리인구 증가

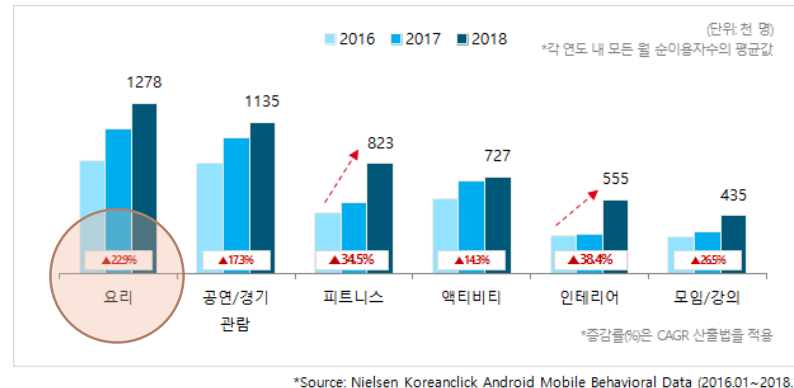
[아시아경제 최신헤 기자] 1인 가구가 지난해 600만 명을 돌파하며 강력한 소비주체로 급부상하는 가운데, 혼술·혼밥 등 이른바 '1인 소비' 맞춤형 상품이 트렌드로 자리잡고 있다. 지금까지 혼밥과 관련된 대표적인 상품으로 배달 서비스, 가정간편식(HMR) 등이 손꼽혔다면, 최근에는 건강한 식생활을 위해 **직접 요리해서 먹는 1인 가구가 점차 증가**하며 식재료 시장에서 소포장 제품이 인기다.

(출처) 아시아경제, 요리하는 1인가구 증가...식재료도 '소포장'이 대세
<https://www.asiae.co.kr/article/2020072611421421833>



워라벨

'워라벨' 트렌드에 따른 취미 및 관심사 전문 모바일 서비스 이용행태 분석



(출처) 닉슨코리아클릭, '워라벨' 트렌드에 따른 취미 및 관심사 전문 모바일 서비스 이용행태 분석
http://www.koreanclick.com/insights/newsletter_view.html?code=topic&id=511&page=1



포스트코로나

그러나 **코로나19 확산**세로 '홈족(Home族)'이 대세가 되면서 1인 가구 식생활 문제의 전환기가 찾아왔다.

집에서 생활하는 시간이 늘면서 1인 가구 역시 집안에서 식사를 해결해야 했기 때문이다. '귀찮아서', '요리를 못해서' 등의 이유로 **외면했던 주방에 1인 가구가 들어서게 된 것**이다.

(출처) 1코노미뉴스, 포스트 코로나시대, 1인 가구 '식생활 풍경' 달라지나
<http://www.1conomynews.co.kr/news/articleView.html?idxno=12083>

타 서비스 탐색 & 비교

음식 추천의 단계를 1) 레시피 나열, 2) 재료 단순 매칭, 3) 개인화 서비스로 구분하였으며 **Now Meal**은 개인화 서비스까지 구현 목표



밥타임

재료 기반 단순 매칭 서비스
개인화 없이 소장 재료 기준 단순 조리 목록
을 검색해서 보여줌
재료를 수기로 관리해야 하는 불편함



스마트 냉장고

이미지 인식으로 재료를 인식
레시피 추천과 장보기 서비스
레시피 추천도 단순 매칭 서비스 기반



Now Meal

사용자 취향 기반 음식 추천
대량의 레시피 추천 보다는 지금의 재료 상
황과 날씨 상황에 따라 지금 만들어 맛있는
한끼를 먹을 음식 한가지를 추천

서비스 기획

음식 추천 서비스를 린캔버스 기반으로 작성

Problem <ul style="list-style-type: none">- 오늘 뭐 먹지- 요리법을 모름- 요리 선택의 시간 소요 (레시피 선택, 조리법 검색, 조리법 습득 등)	Solution <ul style="list-style-type: none">* 요리 추천<ol style="list-style-type: none">1. 재료기반 요리추천2. 개인화 요리추천	Unique Value Proposition <p>집에 있는 음식을 관리해 준다는 개념이 포인트</p> <ul style="list-style-type: none">- 30분 내 요리 해 먹을 수 있는 메뉴를 알려준다.- 집밥 메뉴, 장 리스트 고민은 그만	Unfair Advantage <p>실제 서비스 활성화 시 2차, 3차 기대효과</p>	Customer Segments <ul style="list-style-type: none">- 요리에 흥미를 가지나 재료에 따른 요리 선택이 필요한 사람
	Key Metrics <ul style="list-style-type: none">- 식재료 이미지 인식- 취향별 추천		Channels <ul style="list-style-type: none">- SNS- Mobile	
Cost Structure (개발 관련 한정) <ul style="list-style-type: none">- DB(레시피) : 크롤링, 자료구조정돈, DB 구축- 식재료 이미지 확보 및 학습- 개인화 관련 요리분류법 자료확보, 개인화 학습 알고리즘 확인/선택			Revenue Streams <ul style="list-style-type: none">- 레시피에서 부족한 재료를 마켓으로 연결 후 사용자가 해당 재료를 구매 시 수익 할당	

서비스 전체 구조도 - 계획안

재료 관리를 시작으로 직접 요리할 수 있는 레시피 제공, 이 과정을 꾸준하게 이어 나갈 수 있는 서비스 내용까지 계획



프로젝트 구현 목표

해당 프로젝트에서는 **Main 기술**이 되는 1) 이미지 인식 시스템, 2) 레시피 추천 시스템 구축 중심으로 개발





Part 2,

개발 과정

개발 일정

기획, 데이터 수집/전처리, 태그 설정, DB/모델 설계, 개발 및 테스트의 과정으로 진행

8/29 ~ 9/19

주제 선정 및 기획 회의

주제 선정/린캔버스 작성/ 모델 탐색

10/10 ~ 10/31

태그 설정 및 검증

음식별 태그 설정
태그 검증 후 수정

10/24 ~ 10/31

DB 구축

ERD 설계
테이블 생성 및 데이터 입력

10/31 ~ 11/14

레시피 추천 시스템 구현

콘텐츠 기반 추천 시스템
개발

11/21

테스트

샘플 데이터로 이미지 인식 및 추천 확인

9/19 ~ 10/24

데이터 수집 및 전처리

레시피 데이터 수집/전처리
이미지 데이터 수집/전처리

10/17 ~ 10/24

세부 기획

업무흐름도, 요구사항명세서
아키텍처 정의서, ERD 작성

9/26 ~ 11/7

이미지 훈련/인식/분류

4가지 모델 실험
InceptionV3 모델 중심 훈련

11/18 ~ 11/21

웹퍼블리싱

Flask 기반 DB connection 및
이미지 인식, 추천 시스템 연결

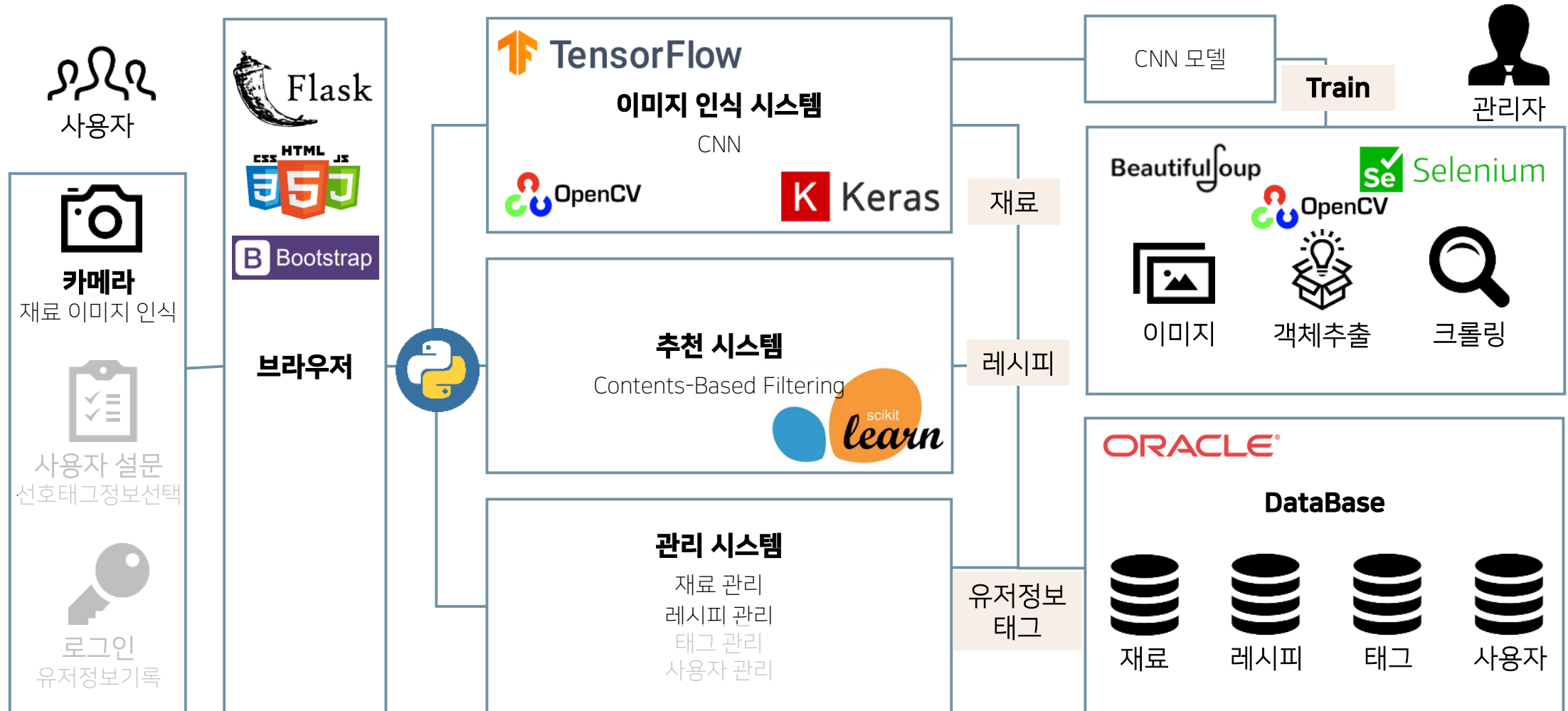
요구사항 명세서

프로젝트 구현 목표를 기준으로 요구사항 명세서 작성

No	분류	요구사항명	요구사항 내용
1	냉장고 재료 관리	식재료 보여주기	냉장고 이미지에 입력한 식재료를 시각화하여 보여준다. [식재료 DB] 레코드 이미지로 출력
2	냉장고 재료 관리	재료 입력	- 사진 촬영 이미지 인식 혹은 수기를 통해 재료 내용을 입력하도록 한다. 가공식품은 바코드 인식으로 가능할지? (재료명, 입고날짜, 재료량) - (이미지분석) 재료 입력 버튼을 누르면 식재료 명을 인식하도록 카메라 창이 뜬다. 하단에 사진 불러오기, 직접입력 메뉴도 띄운다.
3	냉장고 재료 관리	재료 관리	재료 수정/삭제: 기본화면의 재료 이미지를 누르면 삭제 혹은 재료량 변경을 할 수 있도록 한다.
4	레시피 추천	레시피 추천 기본화면	기본화면을 아래로 내리면 식재료 DB에서 있는 재료명을 바탕으로 할 수 있는 음식 2가지를 (이미지, 간략소개) 보여 준다. (1) 재료를 많이 보유한 음식 우선 (사용자 식재료 DB에서 레시피 재료정보 DB의 레시피 코드 count가 많은 순) (2) 사용자의 취향 태그를 보유한 식재료로 만들 수 있는 음식(2가지 정도)을 보여준다. (3) 비오는날, 추운날, 더운날 등에 따른 국물 음식, 찌개류, 찬음식류 등을 추천한다. (4) 사용자가 선택하는 음식의 종류를 배열해 주기적(1주 정도)으로 다른 제일 먹지 않았던 음식 종류를 추천 (5) '추가로 구입하면 할 수 있어요' 설명과 함께 '취향태그'에 맞는 음식 랜덤 추천 & 구입할 재료를 추천(재료 리스트 출력)
5	레시피 추천	레시피 상세내용	레시피 추천 기본화면'에서 음식을 클릭하면 조리할 상세 내용을 보여준다. (레시피 과정정보 출력)
6	추천 정보 관리	실제 실행 조사	주기적인 시간에 (저녁 8시) 올리도록 팝업창/알림으로 추천한 음식을 직접 요리했는지 선택하도록 한다. (선택지: 추천1, 추천2, 추천3, 다른 음식(입력), 외식) 사용자가 선택한 음식의 취향태그를 앞으로의 추천에 반영한다.
7	추천 정보 관리	사용자 취향/행동 조사	선호하는 음식 이미지 선택(취향 태그 반영), 집에서 주로 먹는 음식 선택(취향 태그 반영), 함께 사는 인원 선택(분량 반영)

아키텍처 정의서

이미지 인식 시스템, 추천 시스템, 관리 시스템을 Python Flask로 웹 구현



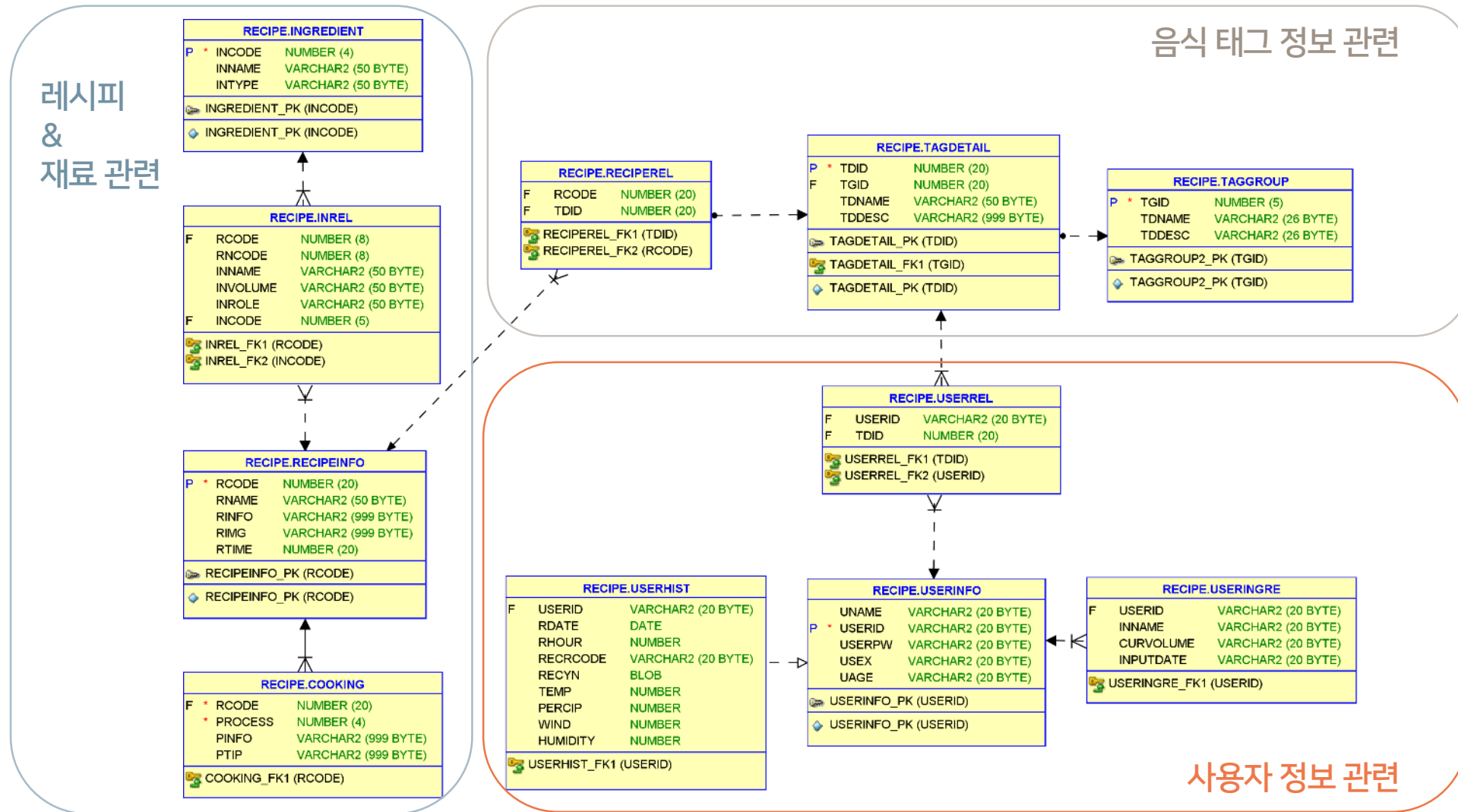
업무 개념도

각 영역의 세부 구조를 업무 개념도로 정리



ERD

각 데이터는 Relation으로 연결되도록 하여 관리가 용이하도록 함



Part 3,

상세 구현 내용

- * 이미지 인식 시스템
- * 추천 시스템



Q.

냉장고에 어떤 재료가 있나요?

이미지 인식 개발 HISTORY

레시피 데이터 탐색, 식재료 데이터 표준화, 음식 태그 설정, 추천 알고리즘 탐색, 구현 순으로 진행

1. 레시피 데이터 탐색 & 수집

- 공공데이터 포털(레시피 538개)
- 캐글 레시피 데이터 수집 (레시피 12455개)
- 한국 영양학회 Can 5.0 등

* 우리나라 식재료 위주이기에 공공데이터 선택

3. 이미지 데이터 셋 수집

- 각 재료별 구글 크롤링
- 당근/양파/대파/무/홍고추

5. 모델 선정

(InceptionV3 / VGG16 / ResNet50 / 셀프 모델)

- 식재료 이미지 인식에 높은 정확도를 보이는 주요 모델
+ 셀프 모델 선정

2. 식재료 데이터 표준화 & 샘플 식재료 선정

(공공데이터 표준화 작업) * 공공데이터 표준화 미처리 문제

- 재료명 표준화 및 재료 분류명 추가
- DB Normalize
- DB 구축

(샘플 식재료 선정)

- 주재료(야채/육류 등) 중에서 레시피에 많이 사용되는 재료 추출
- 형태가 명확한 식재료 위주로 샘플 리스트 선정

4. 이미지 전처리

- 사이즈변환: 128 * 128

(이미지 장수 부족)

- 관련 이미지 추가 수집
- 객체 추출, 데이터 증강

(검색 이미지 퀄리티 문제)

- 배경이 있거나 형태가 다른 사진 이미지 선별 작업

6. Fine-Tuning

- InceptionV3 3 Case
- ResNet50 3 Case
- VGG16 1 Case
- CNN 셀프 모델 2 Case

레시피 데이터 수집

DB 구축을 위한 레시피 데이터 탐색, 최종적으로 공공데이터포털 자료 수집

공공데이터포털 오픈API : 우리 농산물을 이용한 요리정보 제공

- 집에서 먹는 한 끼 음식이므로 우리나라 농산물 기반 레시피 데이터 선택

DATA공공데이터포털
GO . KR

데이터찾기

국가데이터맵

데이터요청

데이터활용

정보공유

이용안내

목록등록관리시스템

로그인

회원가입

사이트맵

오픈API 상세

JSON+XML 레시피 기본정보

우리 농산물을 이용해서 만드는 요리정보로 요리기본정보, 레시피, 주요 재료 정보로 구성되어 있어 다양한 요리레시피 정보를 체계적으로 제공

URL 복사

오픈API 및

담당자 문의

OpenAPI 정보

분류체계	보건 - 식품의약품안전	제공기관	농림수산식품교육문화정보원
관리부서명	지식융합본부 빅데이터실	관리부서 전화번호	044-861-8753
API 유형	LINK	데이터포맷	JSON+XML
활용신청	1238	키워드	안심레시피, 우리농수산물 요리정보, 한식요리레시피
등록	2017-01-04	수정	2017-01-09
URL	http://data.mafra.go.kr/opaedata/data/index/OpenDataDetail.do?data_id=201508270000000000464&filter_ty=O&getBack=Reset_id=&data_nm=&instt_id=&cl_code=&sharetn=		
비밀부과유무	무로		
이용허락범위	이용허락범위 제한 없음		
참고문서			

Recipe Ingredients Dataset	
Use Recipe Ingredients to Categorize the Cuisine	
Kaggle • updated 4 years ago (Version 1)	
Data	Tasks
Download (2 MB)	New Notebook
Usability 7.5	Tags: cooking and recipes
Description	
Context	
If you're in Northern California, you'll be walking past the inevitable bushels of leafy greens, spiked with dark purple kale and the bright pinks and yellows of chard. Across the world in South Korea, mounds of bright red kimchi greet you, while the smell of the sea draws your attention to squids squirming nearby. India's market is perhaps the most colorful, awash in the rich hues and aromas of dozens of spices: turmeric, star anise, poppy seeds, and garam masala as far as the eye can see. Some of our strongest geographic and cultural associations are tied to a region's local foods.	
This dataset was featured in our completed playground competition entitled <i>What's Cooking?</i> The objective of the competition was to predict the category of a dish's cuisine given a list of its ingredients.	
Content	
The data are stored in JSON format.	
• train.json - the training set containing recipes id, type of cuisine, and list of ingredients	
• test.json - the test set containing recipes id, and list of ingredients	

영양정보위원회	
CAN 5.0 (Web ver.) 일반용	
Home > 영양정보위원회 > CAN 프로그램 > CAN 5.0 (Web ver.) 일반용	
활동 및 사업소개	
CAN 프로그램	
• CAN 5.0 (Web ver.) 전문가용 vs 일반용 차이점	
• CAN-Pro 5.0 (Web ver.) 전문가용	
• CAN 5.0 (Web ver.) 일반용	
식품&음식 DB	
CAN 5.0 (Web ver.) 일반용이란:	
CAN 5.0 (Web ver.) 일반용은 개인의 영양 섭취 상태를 평가하는 프로그램입니다. 본 소프트웨어의 개발 목적은 식품과 영양 분야에 관심이 있는 일반인들에게 건강 향상을 위한 영양관리 도구로 활용할 수 있도록 도움을 주는 데 있습니다.	
식품·음식 영양소 합량 DB와 CAN 프로그램은 각각 저작권 등록된 학회의 자산입니다. 저작권 등록은 식품·음식 DB를 저작권자의 허가 없이 무단 추출 및 이용하는 것을 방지하고, CAN 프로그램 이용자의 입력자료에 대한 외부 유출 등의 심각한 문제 발생을 예방하려는 취지에서 진행하였습니다. (등록번호: D-2016-000021, D-2016-000022, C-2016-000555)	

공공데이터 포털

Size : 9638

Fomat : csv

Kaggle recipe Dataset

Size : 4806

Fomat : csv

한국 영양학회 Can 5.0

Size : 1,784

Fomat : csv

레시피 데이터 표준화 & 샘플 식재료 선정

표준화 되어 있지 않은 문제 발생 : 유사재료명을 통일, 순위 추출을 위한 종류 구분 후 샘플 후보 추출

원재료명	변경재료명	분류	종류 분류
비트 간 것	간 비트	주재료	채소
고사리나물	고사리	주재료	채소
근대잎	근대	주재료	채소
생굴	굴	주재료	수산물
김밥용김	김밥김	주재료	수산물
김치잎	김치잎	주재료	가공유제품
낙지다리	낙지	주재료	수산물
청동호박	늪은호박	주재료	채소
쌈다시마	다시마	주재료	수산물
다시물	다시마국물	주재료	양념
다진식파	다진파	주재료	양념
다진쪽파	다진파	주재료	양념
달걀	계란	주재료	육류
달걀	계란	주재료	육류
달걀노른자	계란노른자	주재료	육류
닭	닭고기	주재료	육류
닭살	닭고기	주재료	육류
대구살	대구	주재료	수산물
생대구	대구	주재료	수산물
굵은파	대파	주재료	채소
대파 대	대파	주재료	채소
도라지나물	도라지	주재료	채소
동태살	동태	주재료	수산물
돼지고기	돼지고기	주재료	육류
마른새우	말린새우	주재료	수산물
배춧잎	배추잎	주재료	채소

표준 표기 통일

오탈자

분류 기준 참고 : 밥타임

순위	재료명	재료타입	레시피 count
1	양파	채소	148
2	대파	채소	130
3	당근	채소	107
4	통마늘	채소	98
5	홍고추	채소	98
6	계란	육류	92
7	소고기	육류	91
8	청고추	채소	73
9	무	채소	61
10	통생강	채소	61
11	실파	채소	59
12	오이	채소	59
13	표고버섯	채소	57
14	두부	가공유제품	56
15	밀가루	곡물건과류	55
16	미나리	채소	40
17	돼지고기	육류	39
18	청피망	채소	38
19	쌀	곡물건과류	37
20	감자	채소	36
21	다시마	채소	34
22	오징어	수산물	34
23	새우	수산물	34
24	팽이버섯	채소	28
25	깻잎	채소	28
26	참쌀	곡물건과류	25

재료명 표준화 & 종류 분류 추가

이미지 샘플 재료 후보 추출

이미지 데이터 셋 수집

레시피 데이터에 주로 사용되는 재료를 중심으로 구글 크롤링 진행



구글 이미지 크롤링

샘플 데이터

당근, 계란, 양파, 무, 홍고추

학습 데이터 셋

$128 * 128 * 3$ (80%) → 1728장

검증 데이터 셋

$128 * 128 * 3$ (20%) → 432장

Data Augmentation (데이터 증강)

1) 가로비율변환, 대칭변환, 기울이기, 회전(random~30%)
2) 증강 전 : 1728장 → 증강 후 : **6912장**

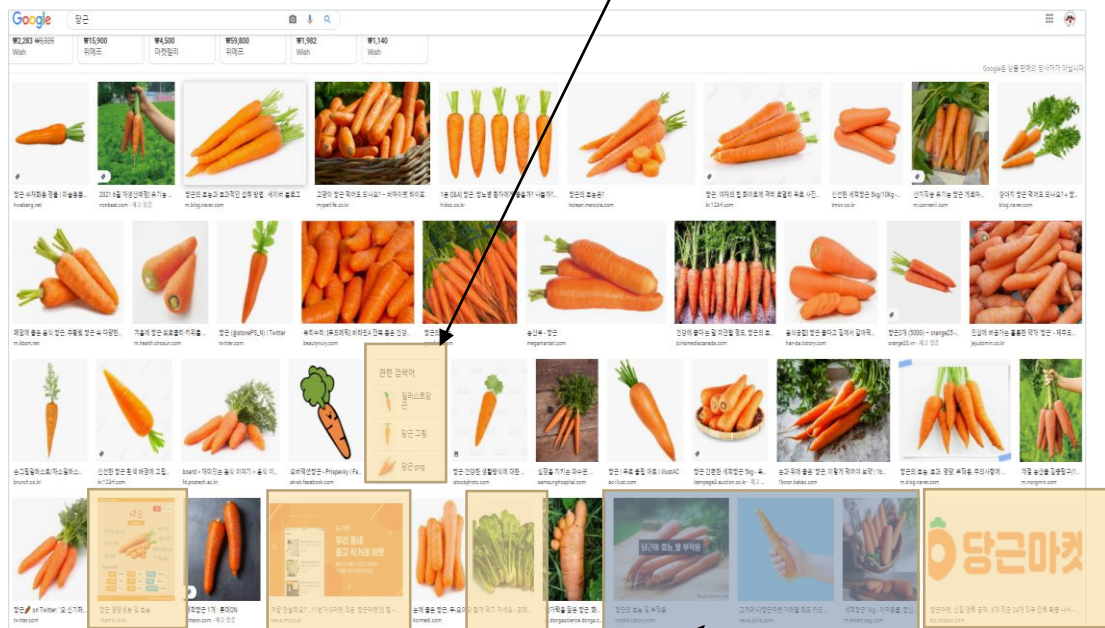


이미지 데이터 전처리

크롤링 수집 후 학습 불가 이미지 전처리

전처리 전

1. 관련도 낮은 이미지 삭제 (텍스트, 편집이미지 등)



2. 인식이 대상이 많은 배경 이미지 삭제

전처리 후



배경 없는 당근 이미지 셋 완성

이미지 인식 모델 선정

InceptionV3, VGG16, ResNet50 모델 전이학습과 자체 CNN까지 4가지 모델을 비교

InceptionV3

1000개의 클래스를 잘 분류하는 학습모델
적은 학습 데이터 가능

Optimizer	Adam, SGD
Metrics	accuracy
Loss	categorical_crossentropy
Model layer	311 layers
평균 학습 시간	35(min)
Epoch	50

InceptionV3

VGG16

기본적인 모델

Optimizer	Adam, SGD
Metrics	accuracy
Loss	categorical_crossentropy
Model layer	19 layers
평균 학습 시간	50(min)
Epoch	50

VGG16

ResNet50

이전 학습 패턴 상기

Optimizer	Adam, SGD
Metrics	accuracy
Loss	categorical_crossentropy
Model layer	176 layers
평균 학습 시간	45 (min)
Epoch	50

ResNet50

CNN Self Model

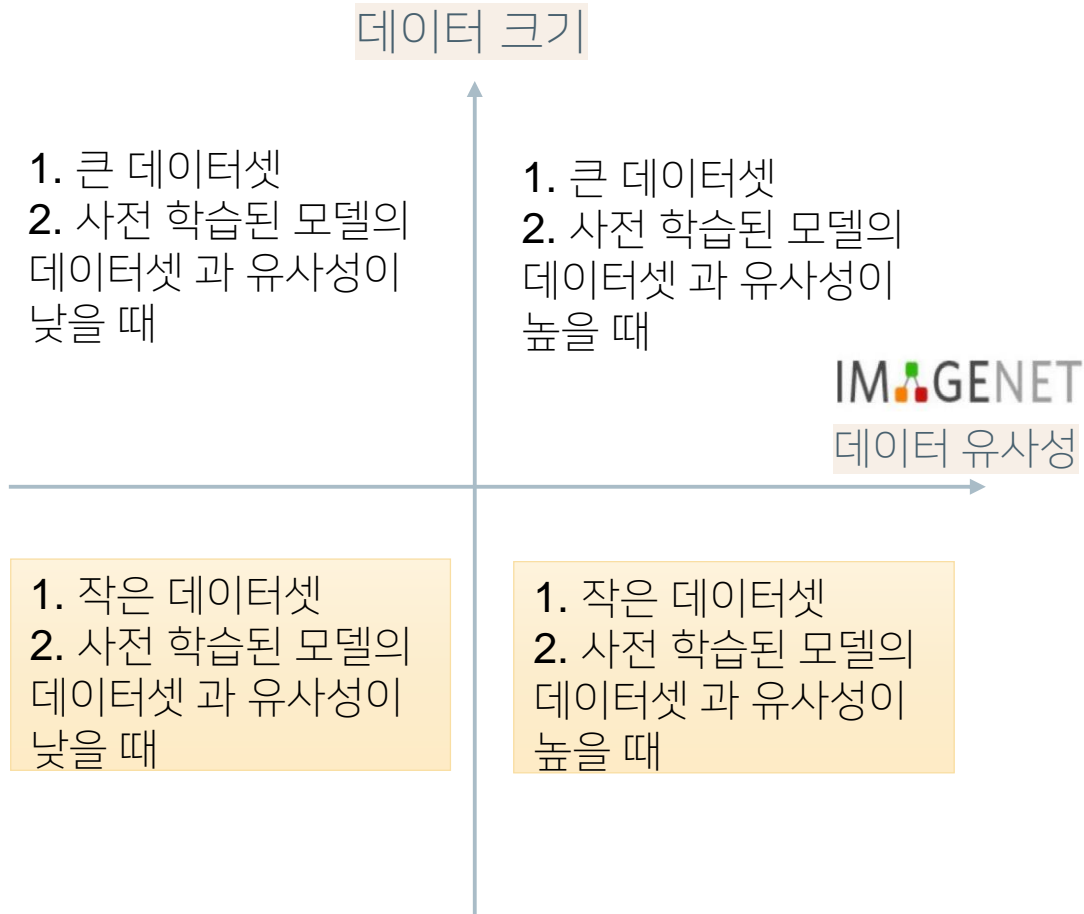
직접 설계한 모델

Optimizer	Adam
Metrics	accuracy
Loss	categorical_crossentropy
Model layer	26 layers
평균 학습 시간	40(min)
Epoch	50

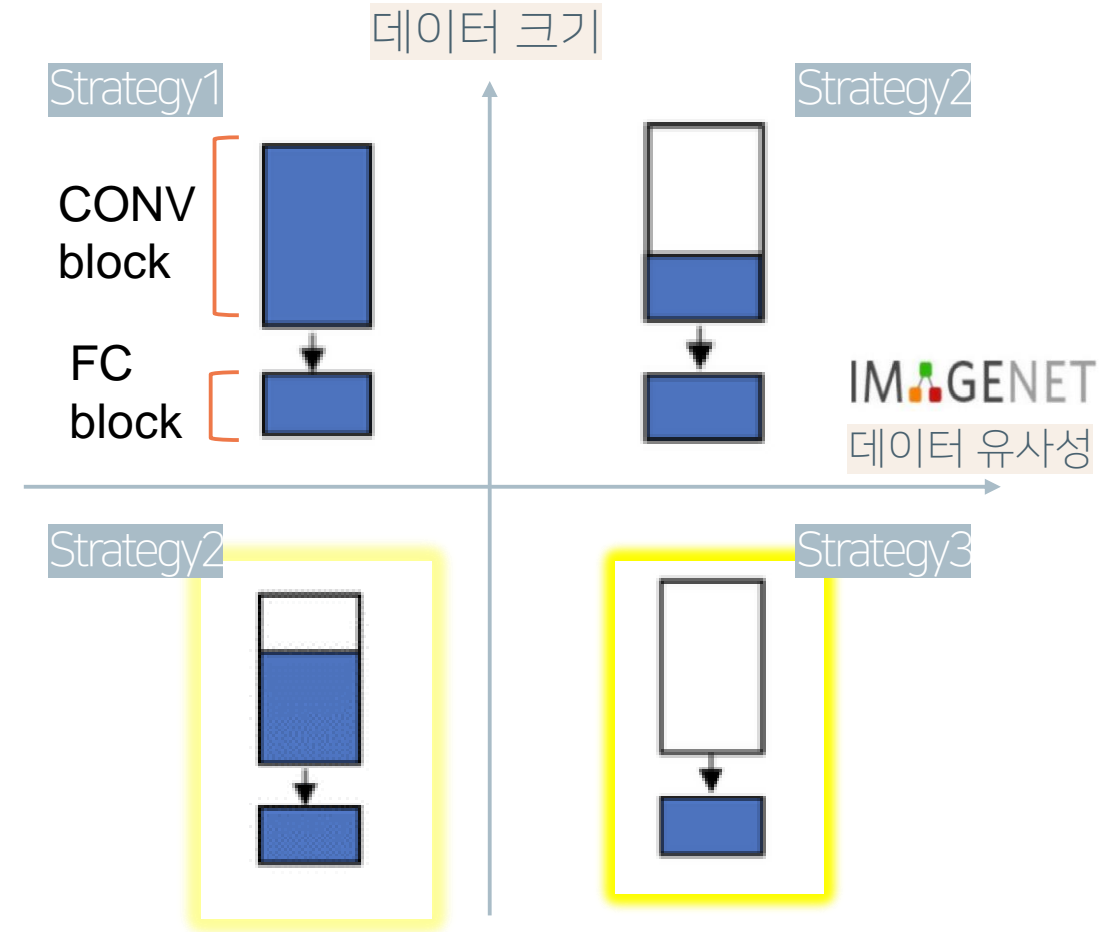
Self
Model

Fine Tuning 전략

데이터 크기와 데이터 유사성을 기준으로 좋은 결과를 낼 Fine-Tuning 을 선정



데이터 크기 - 유사성 그래프

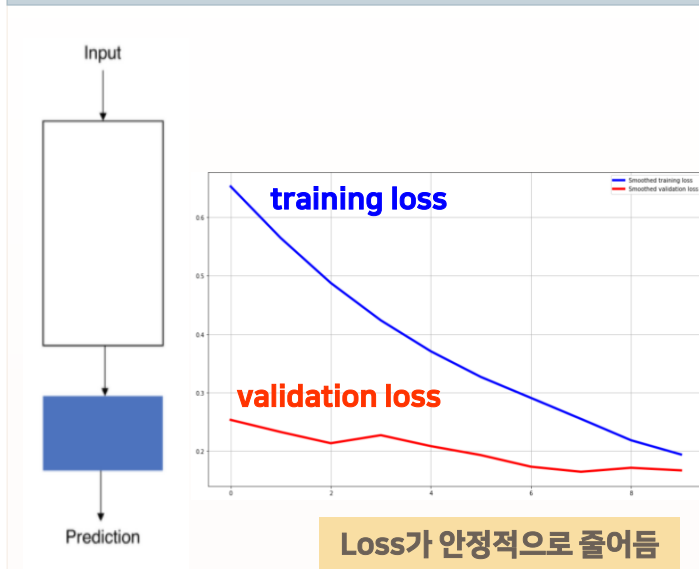


각 상황에 따른 Fine-Tuning 방법

InceptionV3 Fine Tuning 결과

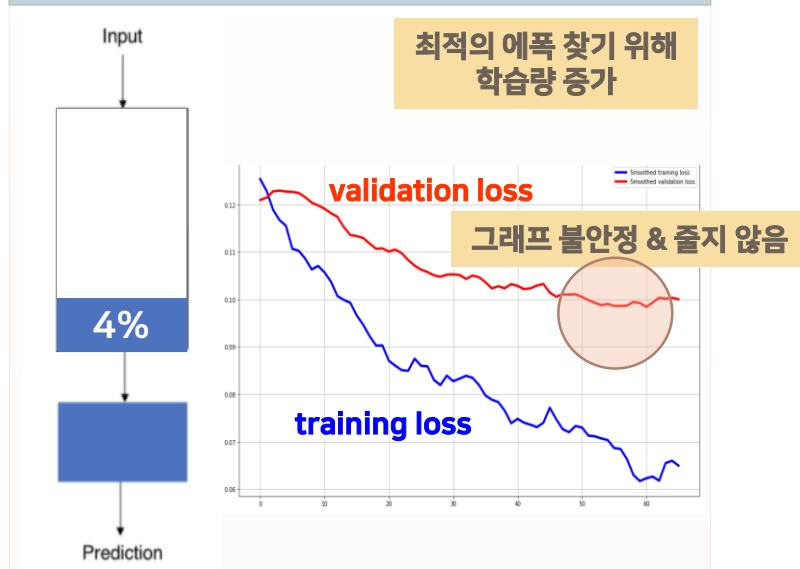
[FC Block] [CONV 상위 4% 추가] [CONV 상위 17% 추가] 케이스로 성능 비교

(실험 1) FC Block 학습



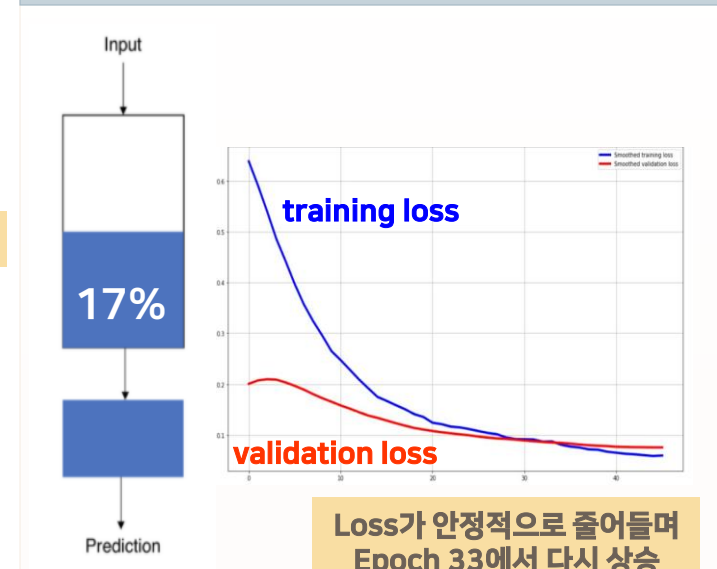
* epoch	50
* Fine tuning	Strategy3
* Early Stopping	patience = 3
* Check point	model-inception-04.hdf5
* Val_Acc	0.960
* Training Time	5(min)

(실험 2) FC Block+Conv Block 상위 4% 학습



* epoch	200
* Fine tuning	Strategy2
* Early Stopping	patience = 20
* Check point	model-inception-46.hdf5
* Val_Acc	0.972
* Training Time	40(min)

(실험 3) FC Block+Conv Block 상위 17% 학습



* epoch	50
* Fine tuning	Strategy2
* Early Stopping	patience = 5
* Check point	model-inception-33.hdf5
* Val_Acc	0.976
* Training Time	62(min)

ResNet50 Fine Tuning 결과

전반적으로 validation loss가 불안정, 오버피팅되어 선정에서 제외

(실험 4) FC Block 학습



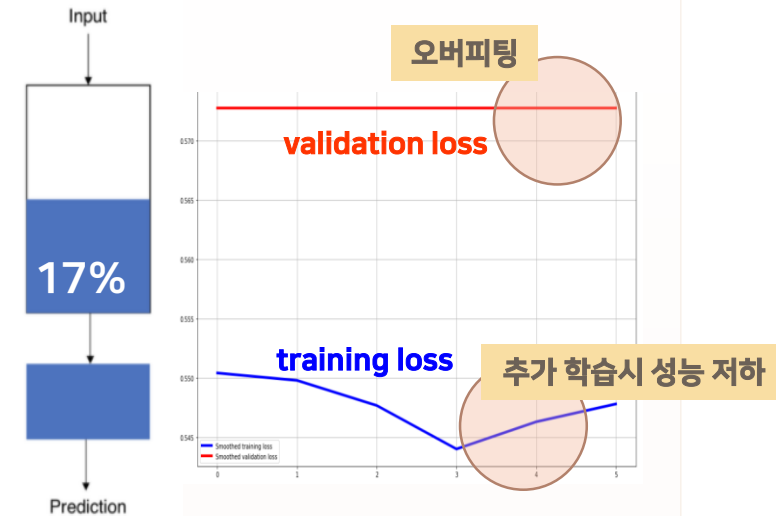
* epoch	80
* Fine tuning	Strategy3
* Early Stopping	patience = 10
* Check point	model-resnet-27.hdf5
* Val_Acc	0.851
* Training Time	34(min)

(실험 5) FC Block 학습



* epoch	250
* Fine tuning	Strategy3
* Early Stopping	patience = 25
* Check point	model-resnet-60.hdf5
* Val_Acc	0.871
* Training Time	76(min)

(실험 6) FC Block+Conv Block 상위 17% 학습

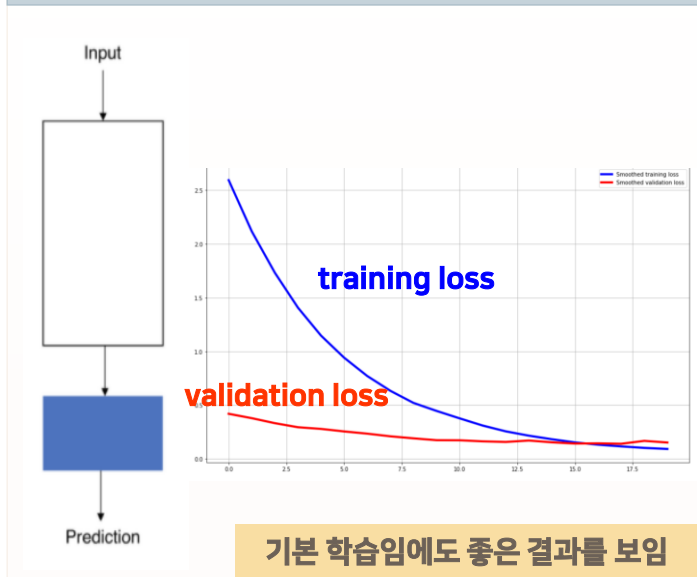


* epoch	50
* Fine tuning	Strategy2
* Early Stopping	patience = 5
* Check point	model-resnet-1.hdf5
* Val_Acc	0.820
* Training Time	21(min)

VGG 16 Fine Tuning & CNN Self Model 결과

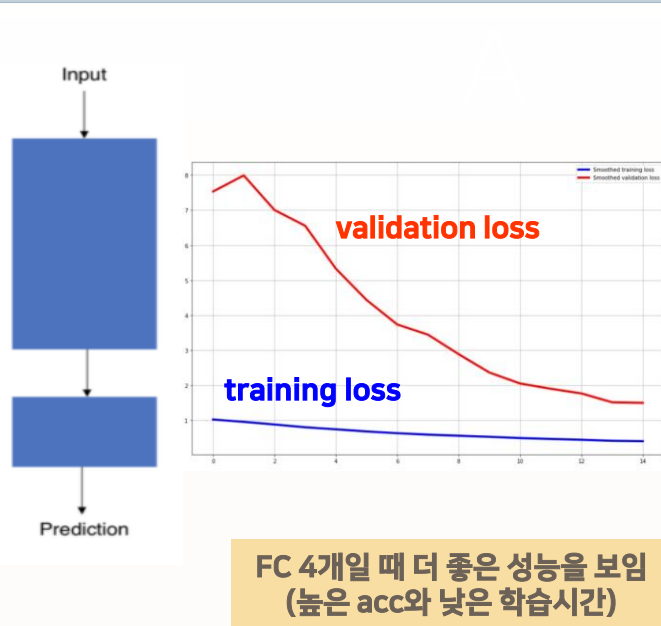
VGG16이 기본모델임에도 높은 accuracy, Self Model은 FC를 4개일 때 좋은 결과가 나옴

(실험 7) VGG 16 FC Block 학습



* epoch	50
* Fine tuning	Strategy3
* Early Stopping	patience = 5
* Check point	model-vgg16-15.hdf5
* Val_Acc	0.964
* Training Time	53(min)

(실험 8) CNN 전체 학습 (FC4개)



* epoch	50
* Early Stopping	patience = 5
* Check point	model-vgg-10.hdf5
* Val_Acc	0.914
* Training Time	33(min)

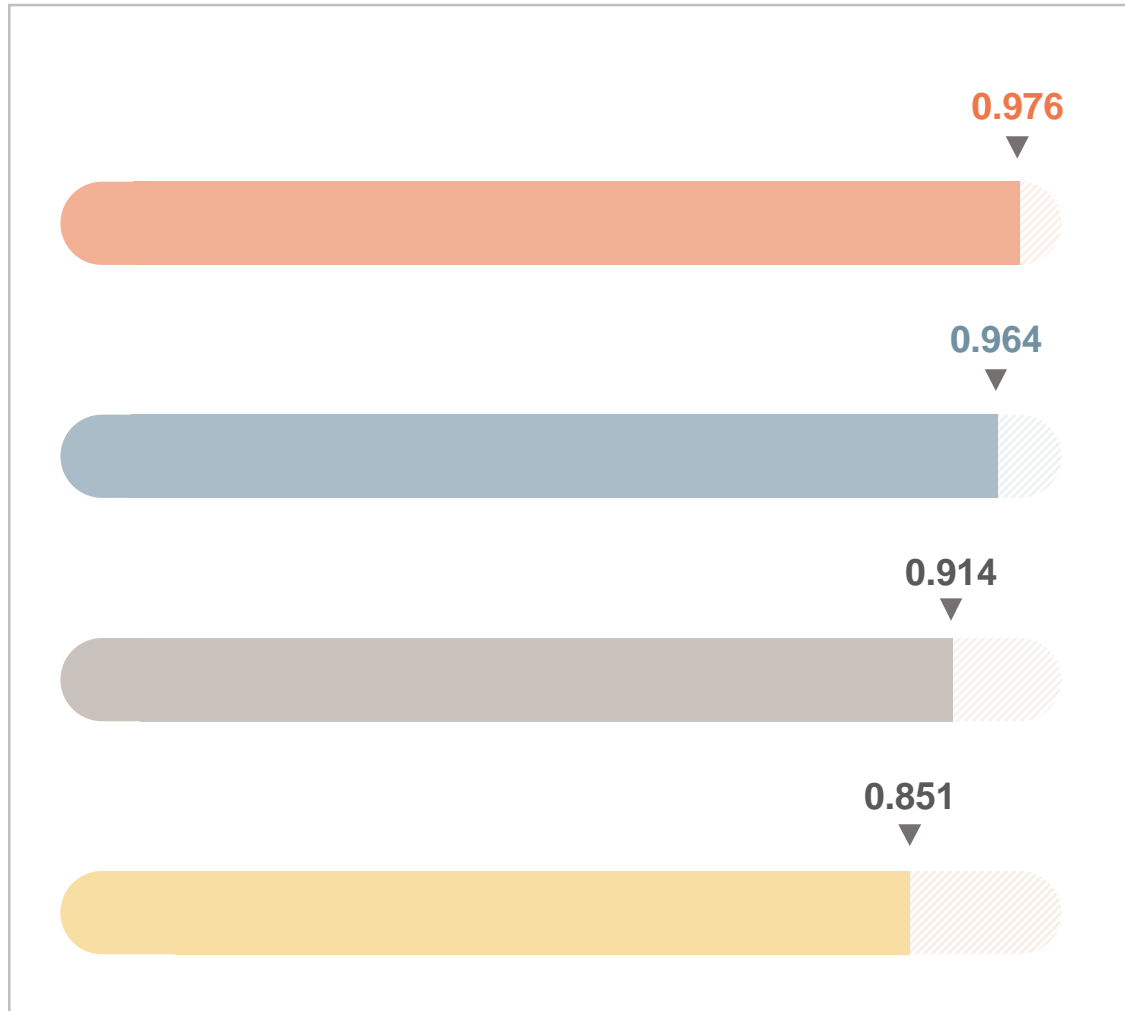
(실험 9) CNN 전체 학습 (FC3개)



* epoch	50
* Early Stopping	patience = 5
* Check point	model-our_model-10.hdf5
* Val_Acc	0.863
* Training Time	46(min)

모델 훈련 결과 비교

InceptionV3 FC Block + Conv Block 상위 17% 학습 모델이 가장 높은 결과를 내어 채택



Inception
V3

* Val_Acc : 0.976
* Training Time : 62 (min)

VGG16

* Val_Acc : 0.964
* Training Time : 53 (min)

CNN
Self Model

* Val_Acc : 0.914
* Training Time : 33 (min)

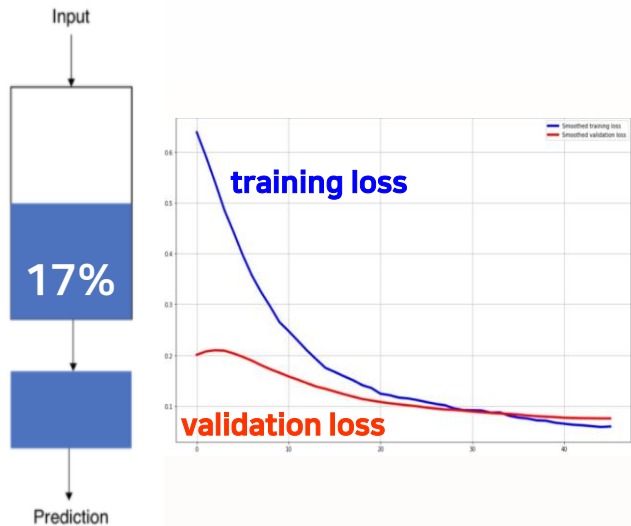
ResNet50

* Val_Acc : 0.851
* Training Time : 34 (min)
* 불안정한 validation loss

모델 선정 결과

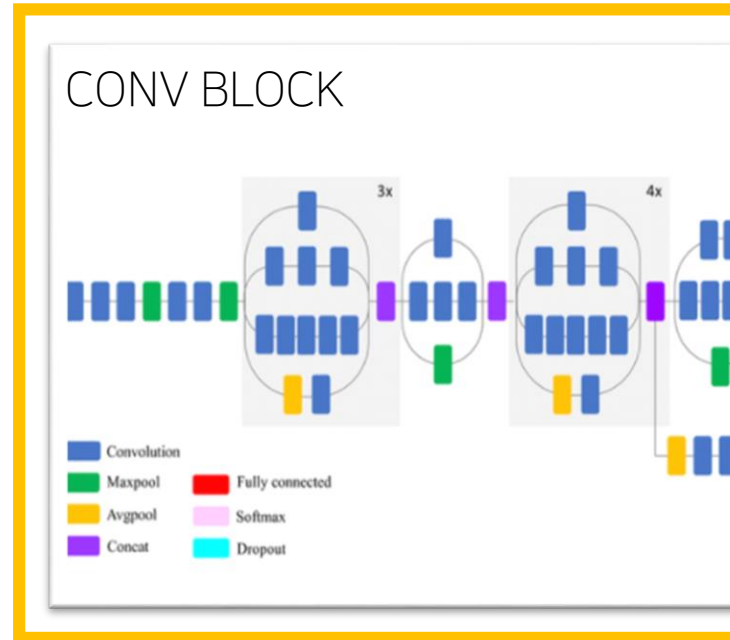
InceptionV3 FC Block + Conv Block 상위 17% 모델 선정

InceptionV3 FC + Conv 상위 17% 학습

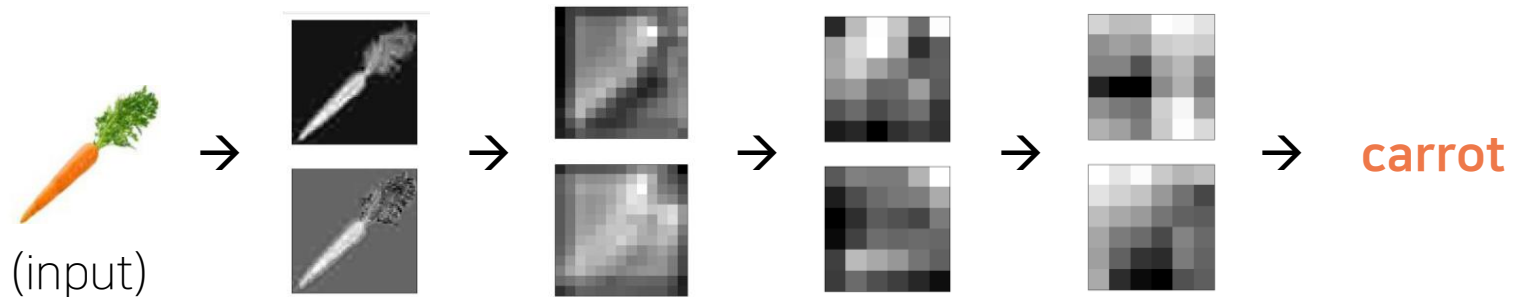
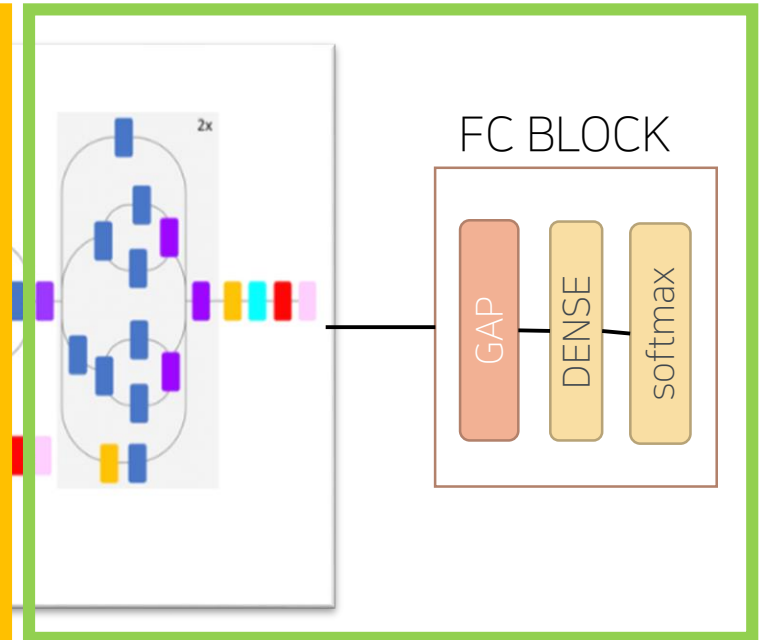


* epoch 50
* Fine tuning Strategy2
* Early Stopping patience = 5
* Check point model-inception-33.hdf5
* Test_Acc 0.976
* Training Time 62 (min)

Freeze



FINE TUNING



이미지 인식 시스템 시연



좋아하는 맛을 입력하세요

#

추천 시스템 개발 HISTORY

레시피 데이터 탐색, 식재료 데이터 표준화, 음식 태그 설정, 추천 알고리즘 탐색, 구현 순으로 진행

1. 레시피 데이터 탐색 & 수집

- 공공데이터 포털(레시피 538개)
- 캐글 레시피 데이터 수집 (레시피 12455개)
- 한국 영양학회 Can 5.0 등

* 우리나라 식재료 위주이기에 공공데이터 선택

2. 식재료 데이터 표준화

(공공데이터 표준화 작업) * 공공데이터 표준화 미처리 문제

- 재료명 표준화 및 재료 분류명 추가
- DB Normalize
- DB 구축

5. 구현

(Contents-based Filtering)

- 사용자 재료 기반 후보 레시피 추출
- 사용자 태그와 레시피 태그 비교 선호 군집 추출
- 해당 군집 음식 리스트 랜덤 추천

3. 음식 태그 작업 및 검증

(1차 정리) [레시피기본정보] 요약 컬럼 기준으로 태그 정리

(1차 검증) 군집 분석 -> 분산도와 밀집도가 높음 -> 사용 불가

(2차 정리) 참고 자료로 기준 태그 설정 후 음식별 태깅 작업

- 차후 확장성을 위해 5가지로 분류

: 종류/상황/날씨/방법/맛 (세부태그 총 72개)

(2차 검증) Accuracy : 0.963, 실루엣 계수 : 8개로 최적 결과

4. 추천 알고리즘 개발

(추천 모델 탐색)

- Collaborative / Contents-based / Hybrid Filtering
- Cold Start 문제 -> 콘텐츠 기반 필터링 모델로 개발
- 차후 Hybrid Filtering으로 확장 목표
- Tensorflow Recommenders

레시피 데이터 수집

DB 구축을 위한 레시피 목록을 공공데이터포털에서 수집

공공데이터포털 오픈API : 우리 농산물을 이용한 요리정보 제공

- 집에서 먹는 한 끼 음식이므로 우리나라 농산물 기반 레시피 데이터 선택

제공 : 농림수산식품교육문화정보원

DATA공공데이터포털GO . KR

데이터찾기국가데이터맵데이터요청데이터활용정보공유이용안내

목록등록관리시스템로그인회원가입사이트맵

오픈API 상세

JSON+XML레시피 기본정보

우리 농산물을 이용해서 만드는 요리정보로 요리기본정보, 레시피, 주요 재료 정보로 구성되어 있어 다양한 요리레시피 정보를 체계적으로 제공

1

0

관심

오픈신고 및
담당자 문의

OpenAPI 정보

분류체계	보건 - 식품의약품안전	제공기관	농림수산식품교육문화정보원
관리부서명	지식융합본부 빅데이터실	관리부서 전화번호	044-861-8753
API 유형	LINK	데이터포맷	JSON+XML
활용신청	1238	키워드	안심레시피, 우리농수산물 요리정보, 한식요리레시피
등록	2017-01-04	수정	2017-01-09
URL	http://data.mafra.go.kr/opaendata/indexOpenDataDetail.do?data_id=20150827000000000466&filter_ty=O&getBack=G&sort_id=&data_nm=&instt_id=&cl_code=&shareYn=		
비용부과유무	무로		
이용허락범위	이용허락범위 제한 없음		
참고문서			

어떤 음식인지

DATA공공데이터포털GO . KR

데이터찾기국가데이터맵데이터요청데이터활용정보공유이용안내

목록등록관리시스템로그인회원가입사이트맵

오픈API 상세

JSON+XML레시피 재료정보

우리 농산물을 이용해서 만드는 요리정보로 요리기본정보, 레시피, 주요 재료 정보로 구성되어 있어 다양한 요리레시피 정보를 체계적으로 제공

1

0

관심

오픈신고 및
담당자 문의

OpenAPI 정보

분류체계	보건 - 식품의약품안전	제공기관	농림수산식품교육문화정보원
관리부서명	지식융합본부 빅데이터실	관리부서 전화번호	044-861-8753
API 유형	LINK	데이터포맷	JSON+XML
활용신청	3332	키워드	안심레시피, 우리농수산물 요리정보, 한식요리레시피
등록	2017-01-04	수정	2017-01-09
URL	http://data.mafra.go.kr/opaendata/indexOpenDataDetail.do?data_id=20150827000000000466&filter_ty=O&getBack=G&sort_id=&data_nm=&instt_id=&cl_code=&shareYn=		
비용부과유무	무로		
이용허락범위	이용허락범위 제한 없음		
참고문서			

어떤 재료가 필요한지

DATA공공데이터포털GO . KR

데이터찾기국가데이터맵데이터요청데이터활용정보공유이용안내

목록등록관리시스템로그인회원가입사이트맵

오픈API 상세

JSON+XML레시피 과정정보

우리 농산물을 이용해서 만드는 요리정보로 요리기본정보, 레시피, 주요 재료 정보로 구성되어 있어 다양한 요리레시피 정보를 체계적으로 제공

1

0

관심

오픈신고 및
담당자 문의

OpenAPI 정보

분류체계	보건 - 식품의약품안전	제공기관	농림수산식품교육문화정보원
관리부서명	지식융합본부 빅데이터실	관리부서 전화번호	044-861-8753
API 유형	LINK	데이터포맷	JSON+XML
활용신청	423	키워드	안심레시피, 우리농수산물 요리정보, 한식요리레시피
등록	2017-01-04	수정	2017-01-09
URL	http://data.mafra.go.kr/opaendata/indexOpenDataDetail.do?data_id=20150827000000000466&filter_ty=O&getBack=G&sort_id=&data_nm=&instt_id=&cl_code=&shareYn=		
비용부과유무	무로		
이용허락범위	이용허락범위 제한 없음		
참고문서			

어떻게 만드는지

레시피 기본정보

Size : 538 * 15

Fomat : csv

레시피 재료정보

Size : 6105 * 6

Fomat : csv

레시피 과정정보

Size : 3023 * 5

Fomat : csv

33

음식 태그 작업

추천 시스템에 활용할 레시피 태그 설정 작업



레시피 태그 설정

추천을 위해 음식별 특징을 태그로 설정
음식 종류/방법/상황/날씨/맛 기준
총 65개 태그 설정

선호하는 음식종류

면요리, 탕요리, 한입요리 등 먹고 싶은 음식의 종류로 추천을 받을 수 있음



식사상황

간단히 먹는 저녁 한끼인지, 가족 친구들과 근사하게 차려먹는 음식인지



맛의취향

매운 맛, 새콤한 맛, 고소한 맛 등 선호하는 음식의 맛을 기준으로 음식 추천



레시피 태그 목록 (총 65개)

종류별

반찬 메인요리 양념/소스/육수 차/음료 베이킹 해물 수프 디저트 국/탕/찌개 면/만두 밥/죽 퓨전/양식 샐러드 간식

방법별

갈기 끓이기 무침 볶음 부침 불없이 조리 오븐요리 튀김 찜

상황별

일상 별미 술안주/야식 다이어트 간편요리 고급 영양식 신선 한입요리 간단아침 어린이 분식 든든 해장 밥도둑

날씨/기온

여름 겨울 비 쌀쌀 서늘

맛별

[미각]

새콤 달콤 상큼 매콤 얼큰 짭짤 씹쓸

[향]

향긋 구수 고소 알싸

[촉각]

쫄깃 바삭 아삭 촉촉 걸쭉 부드러움

[기타]

담백 기름진 감칠맛 깔끔 시원

참고 : 밥타임 어플, 관련논문 박현선(2016)『음식 맛 표현 감각어 의미 연구』

1차 태그 분류

간략소개 데이터를 기반으로 태깅작업, 파이썬 활용 각 태그 더미변수화

레시피 이름	간략(요약) 소개	맛			종류		방법	기온		상황		유형분류	음식분류 1	음식분류 2	음식분류 3	재료별 분류	조리시간	난이도	분량
까르보나라 (한국식)	전통식과는 다르	담백	고소	기름진	면/만두	퓨전/양식	메인요리	볶음		별미	일상	이탈리아	만두	면류		밀가루	30분	초보환영	1인분
술어샐러드	신선한 채소와 술	아삭	새콤		샐러드	해물		볶음이조리		다이어트		한식	나물	생채	샐러드	어류/패류	30분	초보환영	4인분
잉어찜	잉어에 고명을 얹	감칠맛	고소		메인요리	해물		찜		든든	고급	한식	찜			어류/패류	30분	보통	4인분
메기매운탕	메기와 각종 채소	얼큰	매콤		메인요리	국/탕/찌개		끓이기		든든	고급	한식	찌개	전골	스튜	어류/패류	60분	보통	4인분
비빔냉면	더운 여름, 시원하	새콤	달콤		메인요리	면/만두		끓이기	여름	간편요리	본식	한식	만두	면류		밀가루	50분	보통	4인분
미소된장국	단백한 국물맛이	담백			국/탕/찌개			끓이기		간편아침	일상	일본	국			콩류	30분	초보환영	2인분
콩치김치찌개	김치찌개가 지겨	얼큰	시원		국/탕/찌개			끓이기		일상	든든	한식	찌개	전골	스튜	채소류	60분	보통	4인분
떡만두국	따끈한 국물이 생	쫄깃	부드러움		메인요리	면/만두	국/탕/찌개	끓이기	쌀쌀	든든	명절	한식	만두	면류		가공식품류	30분	보통	2인분
채소파클	새콤 달콤 아삭한	새콤	달콤	아삭	반찬			볶음이조리		간편요리		서양	밀반찬	김치		채소류	20분	초보환영	2인분
발효초요구르트	발효초를 넣어 더	상콤	새콤	달콤	디저트	간식	차/음료	볶음이조리		간편요리	어린이	서양	음료			밀가루	10분	초보환영	4인분
쫄면	쫄깃하고 새콤달	쫄깃	새콤	달콤	메인요리	면/만두		끓이기		본식	간편요리	한식	만두	면류		밀가루	40분	보통	4인분
김치우동	칼칼한 국물이 땀	칼칼	시원		메인요리	면/만두		끓이기	겨울	본식	간편요리	일본	만두	면류		밀가루	30분	초보환영	2인분
해물순두부찌개	부드럽게 넣어가	부드러움	시원	담백	국/탕/찌개	해물		끓이기		일상	별미	한식	찌개	전골	스튜	콩류	30분	보통	4인분
해물매운탕	매콤하면서 시원	매콤	시원		국/탕/찌개	해물		끓이기		해장		한식	찌개	전골	스튜	어류/패류	60분	보통	4인분
새송이쌈장구이	쫄깃한 새송이를	쫄깃	고소		반찬			볶음		술안주/아식	한입요리	한식	구이			버섯류	30분	보통	2인분

▽
▽

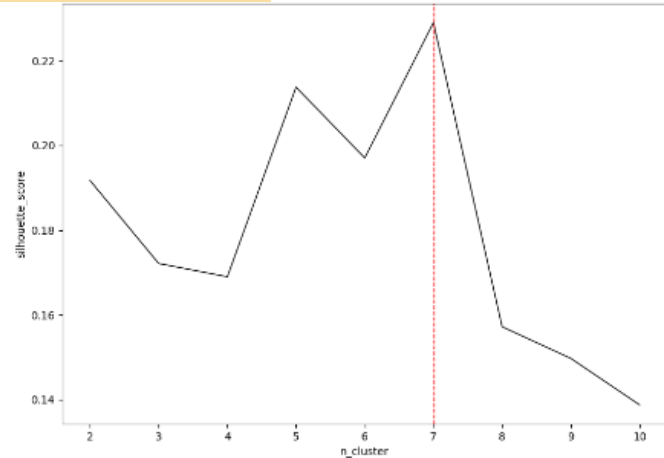
	간단아침	간식	간편요리	고급	고소	구수
RNAME						
가는파잡채	0	0	0	0	1	0
가래떡꼬치	0	1	0	0	0	0
가지그라탕	0	0	0	0	1	0
가지김치	0	0	0	0	0	0
가지된장찜	0	0	0	0	1	1

1. 위 데이터를 토대로 110개 태그의 원 핫 데이터 생성
2. 군집화 실시 결과 7개의 최적화된 군집
3. 독립성 검정(군집-태그, 태그-태그, 군집-군집)

1차 태그 검증

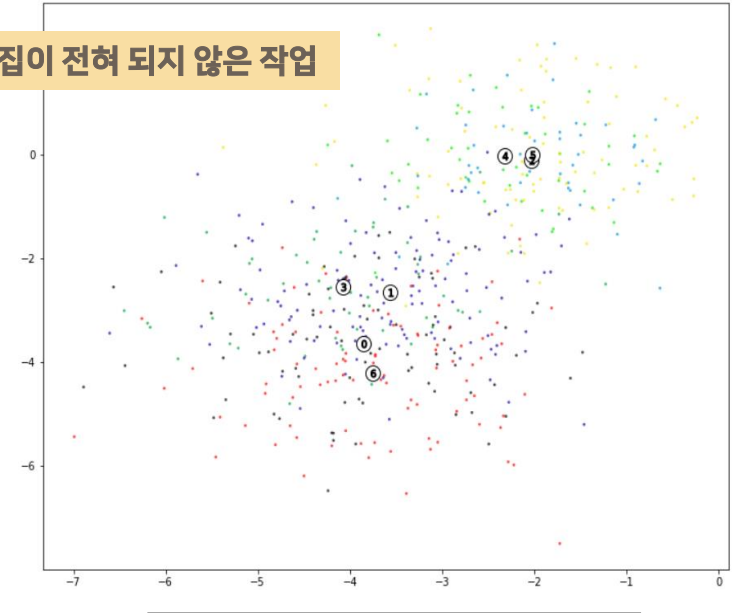
110개의 태그 기반 Kmeans 클러스터링 실시, 활용할 수 없는 태그임을 확인

실루엣계수 7에서 최적 군집



Silhouette coefficient visualization

1차 태그는 군집이 전혀 되지 않은 작업



Cluster visualization

검정

H0 : 레시피와 태그 간 관계는 독립이다.

- Chi-square statistic : 1.4288
- Degrees : 654(109 x 6)
- P-value : 0.999

P-value > 0.05 이므로 H0 기각X 즉, 레시피와 태그는 관계가 없다.

- H0 : 레시피 군집에 따라 태그가 같다.(레시피와 태그간 관계가 없다.)
- 레시피 군집에 따라 태그가 의미없이 붙어있다.
- 7개 군집 독립성 검정 결과 : 군집 간 독립.
- 110개 태그 독립성 검정 결과 : 태그 간 독립.

2차 태그 분류

태깅작업, 파이썬 활용 각 태그 더미변수화

레시피 코드	레시피 이름	맛			종류		방법		기온	상황	
195453	콩비지동그랑땡	고소	담백	기름진	반찬		부침			일상	
195428	누드김밥	고소	담백		밥/죽		불없이조리	볶음		한입요리	
180363	쪽파 새우강회	달콤	새콤		반찬		무침	불없이조리		술안주/야식	간편요리
180344	카레토마토달걀볶음밥	달콤	짭짤	고소	밥/죽		볶음			별미	
180330	식빵고구마파이	달콤	고소		베이킹		오븐요리			어린이	
120476	배추만두	아삭	고소		면/만두	간식	찜			별미	
120463	당근잎 감자전	향긋	기름진	고소	간식		부침	갈기	비	간편요리	
120441	단호박 고등어조림	달콤	짭짤	담백	반찬		끓이기			일상	
120419	나물 월남쌈	담백	깔끔		메인요리		불없이조리			별미	
120401	애호박 구이	담백	고소	기름진	반찬		부침			일상	
120379	오이보트카나페	상큼	아삭	고소	퓨전/양식	간식	불없이조리			한입요리	별미
120360	가지두부스테이크	기름진	담백	부드러움	메인요리	퓨전/양식	부침			고급	별미
120344	깻잎장아찌	짭짤	새콤	매콤	반찬		불없이조리			일상	

관련 자료를 참고하여 태그 목록 작성 후
목록 기준으로 레시피 태그를 재설정

레시피 태그 목록 (총65개)

종류별
반찬 메인요리 양념/소스/육수 차/음료 베이킹 해물 수프 디저트
국/탕/찌개 면/만두 밥/죽 퓨전/양식 샐러드 간식

방법별
갈기 끓이기 무침 볶음 부침 불없이조리 오븐요리 튀김 찜

상황별
일상 별미 술안주/야식 다이어트 간편요리 고급 영양식 신선
한입요리 간단아침 어린이 분식 든든 해장 밥도둑

날씨/기온
여름 겨울 비 쌀쌀 서늘

맛별
[미각]
새콤 달콤 상큼 매콤 얼큰 짭짤 씹쓸

[향]
향긋 구수 고소 알싸

[촉각]
쫄깃 바삭 아삭 촉촉 걸쭉 부드러움

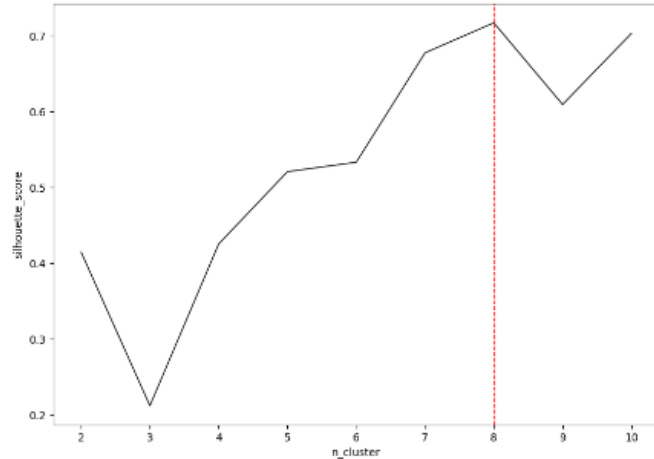
[기타]
담백 기름진 감칠맛 깔끔 시원

참고 : 밥타임 어플, 관련논문 박현선(2016)『 음식 맛 표현 감각어 의미 연구』

2차 태그 검증

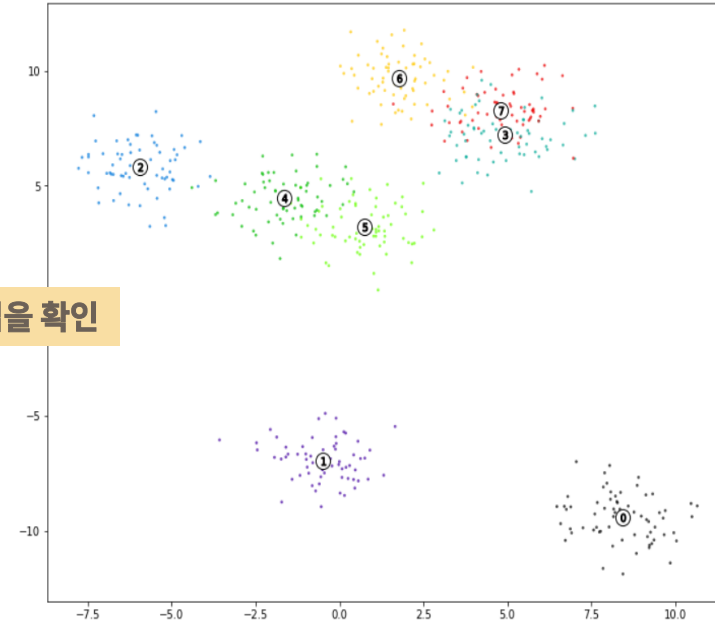
재 분류한 65개의 태그 기반 Kmeans 클러스터링 실시, 태그가 유의미함을 확인

실루엣계수 8에서 최적 군집



Silhouette coefficient visualization

8개의 군집을 확인



Cluster visualization

검정

H0 : 레시피와 태그 간 관계는 독립이다.

- Chi-square statistic : 5140.32
- Degrees : 448(64 x 7)
- P-value : $2e-16$

P-value < 0.05 이므로 H0 기각 즉, 레시피와 태그는 관계가 있다.

- H0 : 레시피 군집에 따라 태그가 같다.(레시피와 태그간 관계가 없다.)
- P-value : 0.00 이므로 귀무가설 기각, 레시피 군집에 따라 태그가 다르다. 즉, 태깅 작업이 잘 되었음을 뜻함.
- 8개 군집 독립성 검정 결과 : 군집 간 독립.
- 65개 태그 독립성 검정 결과 : 태그 간 독립.



추천 알고리즘 탐색

협업 필터링(Collaborative filtering)

- : 기존 사용자 행동 정보를 분석
비슷한 성향의 사용자들이 좋아했던 항목을 추천
- * 콜드 스타트(Cold Start) 문제
기존 자료를 활용하는 조건, 새로운 항목 추가 시 추천 곤란

콘텐츠 기반 필터링(Content-based filtering) ★

- : 항목을 분석한 프로파일(item profile)
& 사용자의 선호도 추출(user profile)의 유사성 계산
군집분석, ANN, tf-idf 등의 기술 사용

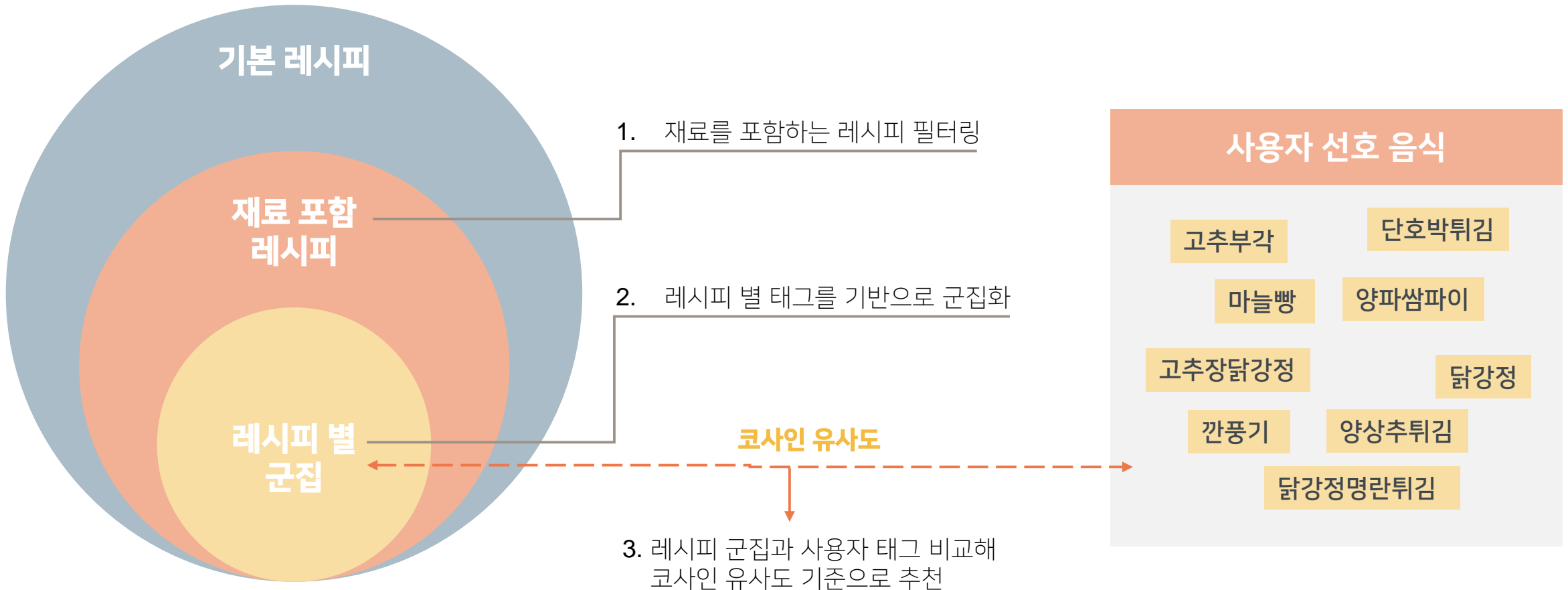
* 레시피 음식의 태그 & 사용자의 선호도 태그 활용

* 내용 자체를 분석하여 콜드 스타트 문제 해결

콘텐츠 기반 필터링(CBF)

태그 기반 레시피 코사인 유사도 비교

태그 기반 레시피 군집분석 알고리즘



콘텐츠 기반 필터링(CBF)

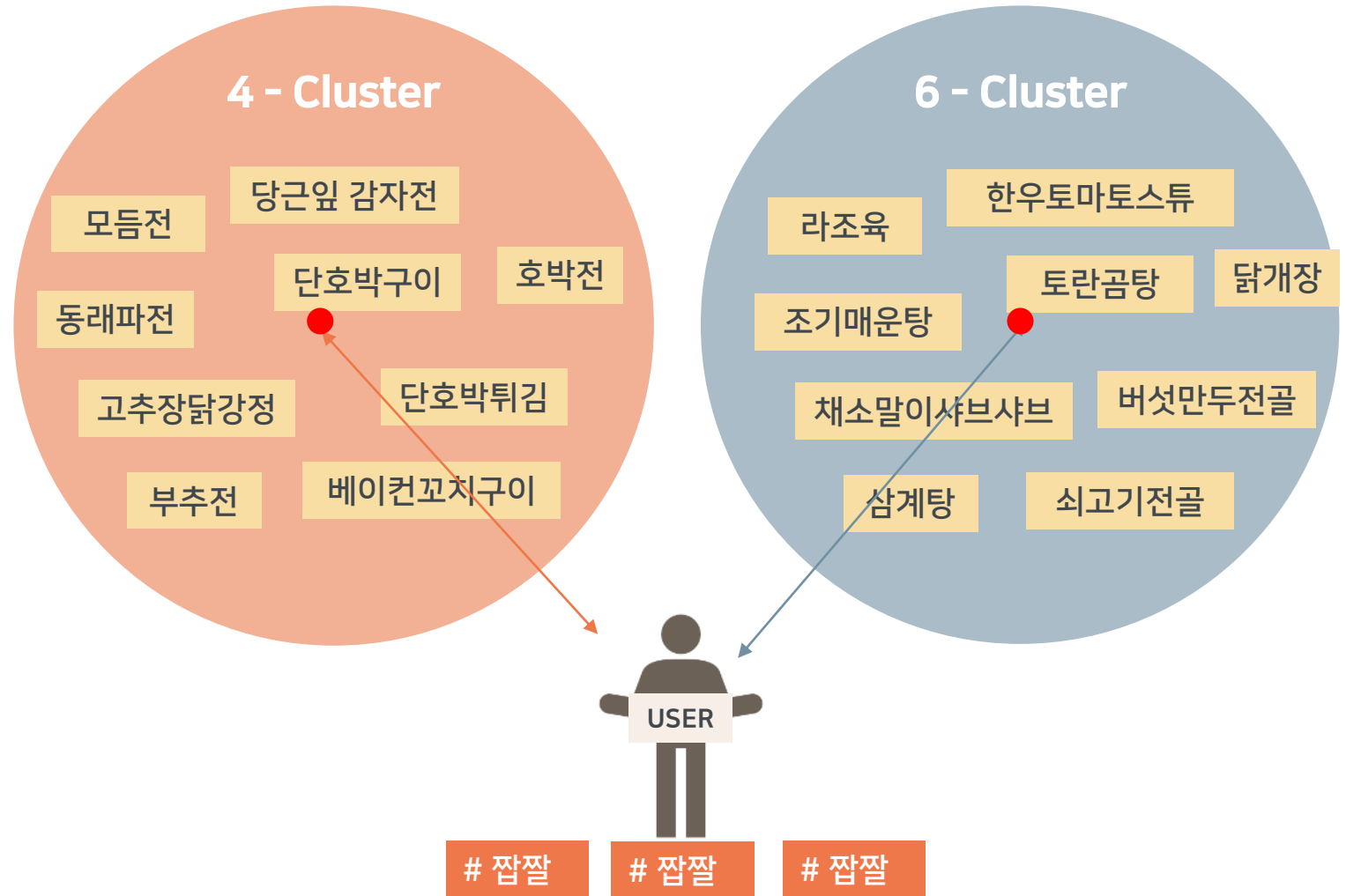
태그 기반 레시피 코사인 유사도 비교 세부 과정

1. User 가 가진 태그를 기반으로 각 군집 센터와 의 65차원 공간상의 Distance 계산

>> Distance가 가장 가까운Cluster 로 매칭

2. 군집 내에서 User를 포함한 행렬 코사인 유사 도 계산 (계산 척도는 원핫으로 이루어진 태그 데이터)

>> 코사인 유사도 정렬 후 상위 10개 음식 추천





Now Meal 예시 구현





Part 4,

차후 과제

추가 서비스 구현 내용


웹구현으로 남겨두었던 서비스 차후 개발 필요

냉장고 재료 관리	이미지 인식 입력/관리	식재료 보여주기		
레시피 추천	취향태그 기반 추천	필요한 재료 구입 추천	날씨 반영 추천	재료 히스토리 반영
추천 정보 관리	실제 실행 조사	사용자 취향/행동 조사		
서비스	장바구니 리스트	장 / 배달 서비스	로그인	검색창
개인화	요리했던 음식 기록	Gamification	로그인 활동 기록	
커뮤니티	레시피 공유	요리 유튜브 채널		

인식 시스템 확장

재료 입력 방법 보완

(1) 영수증 인식

-  Google Cloud 이미지 인식 API 활용 개발 구현
- 영수증 촬영 → 재료명/날짜/수량 text 추출 → DB 저장
- 텍스트 표준화 할 세부 개발 필요

(2) 바코드 인식

- 포장에 있는 재료 인식 (두부, 콩나물, 가공식품 등)
- 포장지에 있는 바코드를 인식해 재료명 저장하는 개발 필요

주식회사 한강식자재마트
사업자번호: 327-87-01642
대표자: 서영록
주소: 서울 강동구 고덕로83길 28 191-5
번지 지하층 1-12호 (고덕동, 백두쇼핑센터)
전화번호: 02-427-9922

판매일: 20-11-23 21:13, 월요일 계산대: 001

NO.	상품명	단가	수량	금액
001	대추방울토마토 (빨강/노랑)	2,980	1	2,980 #
002	홍당근	1,560	1	1,560 #

(=) 면세상품 : 4,540
합계 : 4,540
신용카드지불 : 4,540

고객:
발: 은포인트:
적립포인트:

카드신용승인
가맹점명: 주식회사 한강식자재마트
카드번호: [redacted]
카드명: [redacted]
전표매입: 4,540 (일시불)
승인금액: 76200808, 전표No: 211319
승인번호: Smartro-리더기
지출 VAN: 00 정상승인거래 신용승인
정상승인

거래NO: 1123002182 계산원: 관리자 (001)





추천 시스템 확장

오늘 날씨 추천 반영 (cf. spotify 음악 추천 시스템)

- 크롤링으로 네이버 현재 날씨 추출 구현
- 현재 날씨 데이터를 추천 시스템에 반영 추후 필요

메뉴 다양화

- 현재 수집한 한식 레시피의 표준화, 다양화 부족
- 냉장고 파먹기를 할 수 있는 일상 혹은 간편, 건강한 레시피 추가 관리 필요

딥러닝 기반 추천 확장

- 항목간 유사성 단순 비교를 벗어나 자료 안에 잠재노드를 사용
- Tensorflow recommenders

감사합니다





Reference.

■ 기획 및 레시피 데이터

- 아시아경제, 요리하는 1인가구 증가...식재료도 '소포장'이 대세
<https://www.asiae.co.kr/article/2020072611421421833>
- 닉슨코리아클릭, '워라벨' 트렌드에 따른 취미 및 관심사 전문 모바일 서비스 이용
행태 분석
http://www.koreandclick.com/insights/newsletter_view.html?code=topic&d=511&page=1
- 1코노미뉴스, 포스트 코로나시대, 1인 가구 '식생활 풍경' 달라지나
<http://www.1conomynews.co.kr/news/articleView.html?idxno=12083>
- 냉파족을 위한 모바일 레시피 '냉장고 속 비밀'
<https://www.venturesquare.net/744832>
- 냉장고를 부탁해 - 머신러닝 알고리즘을 활용한 각종 분석
https://jsideas.net/fridge_please_final/
- KKIDI 음식구분표 <https://www.khidi.or.kr/board?menuId=MENU01559>
- 한국영양학회 <http://www.kns.or.kr/Center/CanPro5n.asp>
- 공공데이터포털 <https://www.data.go.kr/data/15057205/openapi.do0>

■ CNN 모델 개발

- Transfer Learning | 학습된 모델을 새로운 프로젝트에 적용하기
<https://jeinalog.tistory.com/13>

■ 개인화 추천 서비스 개발

- Bodhisattwa Prasad Majumder, Shuyang Li, Jianmo Ni, Julian McAuley(2019),
『**Generating Personalized Recipes from Historical User Preferences**』,
University of California, San Diego
- 박현선(2016), 『**음식 맛 표현 감각어 의미 연구**』, 건국대학교 대학원 석사학위 논문
- 서봉원(2016), 『**콘텐츠 추천 알고리즘의 진화**』, 한국콘텐츠진흥원
- 삼성반도체이야기, 백발백중! 취향저격수 '추천 알고리즘'의 비밀
<https://www.samsungsemiconstory.com/2265>
- 당근마켓, 딥러닝 개인화 추천 <https://medium.com/daangn/9C-1eda682c2e8c>
- TensorFlow recommenders API
https://www.tensorflow.org/recommenders/examples/basic_retrieval
- 밥타임 <http://app.bobtime.net/>