

## <Food Trend Project>

컴퓨터공학부 202001581 신소정

주제 - '식품 트렌드(유행)의 기한'은 얼마나 되며,  
'최고점을 찍은 후 트렌드의 변화 모양'은 어떻게 정의할 수 있는가?

- 주제 선정 배경

흔히들 '철 지난 유행이다.', '유행의 변화가 빠르다.'라 한다.

그렇다면 '식품 유행'의 기한은 얼마나 되며, 어떻게 변화할까? 또한, 그 유행의 기한은 어떻게 정의할 수 있는가? 기한에 따라 미래 유행의 경향도 예측할 수 있지 않을까?

- 정의

유행(流行) 또는 트렌드(영어: trend) : 한 사회의 어느 시점에서 특정 생각, 표현 방식, 제품 등이 그 사회에 침투 또는 확산해 나가는 과정에 있는 상태이다.

- 가정

1. 유행은 특정 단어의 사용량과 밀접한 관계를 갖는다.
2. 유행이란 다수의 사람들이 구두, 서적, Portal 등 다양한 분야에서 사용되는 특정 단어들이 모집합이 되나, Portal 이라는 부분집합만으로도 충분한 설명이 가능하다.

- 가설

유행하는 특정단어의 검색량 그래프는 최고점을 기준으로 자연지수함수(Exponential Function)모양을 띈다. 즉,  $F(x) = \exp(-m \cdot x)$  그래프로 피팅된다.

이 때, 자연지수함수( $F(x)$ )의 1 시상수( $1/m$ )를 유행이 지난 시점으로 본다. 1 시상수는 시간이 지날수록 짧아진다. (유행의 속도가 빨라진다.)

- 빅데이터 분석

<Target 층>

유행에 민감한 10 대~20 대까지 모든 성별의 연령을 타겟층으로 삼는다.

<Data set>

Naver 통합 검색어 트렌드 api 이용 - 특정단어에 대한 네이버 통합검색의 검색 추이 데이터를 JSON 형식의 데이터로 전달한다. (하루 호출 한도: 1000 회)

- '식품' 카테고리에서 '유행'의 정의

- ➔ 검색의 빈도수가 3년이상 상위권(20위 안)에 위치한 일상적인 식품단어는 '유행'이라는 정의에서 제외한다. (예, 닭가슴살, 생수 등 <그림1> 레드라인 처리)
- ➔ 계절에 따라 사용량이 많아지는 단어들을 제외하기 위해 키워드의 표본을 1년 간 검색 량이 많은 단어 집합으로 정한다.

<그림 1>

2017 ~ 2018	2018 ~ 2019	2019~ 2020	2020~ 2021	2021~ 2022	2022~ 2023
<b>식품 인기검색어</b> 2017.10. ~ 2018.09. <b>TOP 500</b> 1 미숫가루 2 <del>단가습살</del> 3 곤약젤리 4 머랭쿠키 5 마카롱 6 갈라만시 7 중국당면 8 송주불냉면 9 <del>생수</del> 10 양배추즙 11 <del>햇반</del> 12 칼로리몬스터 13 로이스초콜렛 14 코하쿠로 15 닥터리브곤약젤리 16 팜빙보바 17 유산균 18 <del>코구마</del> 19 gm초록이분홍이 20 아몬드브리즈	<b>식품 인기검색어</b> 2018.10. ~ 2019.09. <b>TOP 500</b> 1 <del>단가습살</del> 2 타피오카펠 3 마카롱 4 <del>프로대유청</del> 5 <del>타이아르트도시락</del> 6 코호젤리 7 하드코어버닝티 8 곤약젤리 9 중국당면 10 지구젤리 11 송주불냉면 12 <del>햇반</del> 13 물티저스 14 칼로리몬스터 15 유산균 16 <del>단백질보충제</del> 17 호박즙 18 서양당국 19 임페프로도시락 20 양배추즙	<b>식품 인기검색어</b> 2019.10. ~ 2020.09. <b>TOP 500</b> 1 <del>단가습살</del> 2 물티저스 3 <del>프로대유청</del> 4 <del>크릭오거트</del> 5 <del>타이아르트도시락</del> 6 유산균 7 일리캡슐 8 <del>코구마</del> 9 타피오카펠 10 <del>단백질쉐이크</del> 11 <del>햇반</del> 12 마카롱 13 사세버팔로윙 14 <del>생수</del> 15 <del>단백질보충제</del> 16 칼로리몬스터 17 지구젤리 18 곤약젤리 19 <del>프로틴</del> 20 연어	<b>식품 인기검색어</b> 2020.10. ~ 2021.09. <b>TOP 500</b> 1 <del>단가습살</del> 2 <del>크릭요거트</del> 3 고복침초코츄러스 4 오트밀미니바이트 5 <del>프로대유청</del> 6 <del>카콜랜드프로틴바</del> 7 <del>타이아르트도시락</del> 8 유산균 9 <del>오메가3</del> 10 <del>생수</del> 11 <del>단백질쉐이크</del> 12 어드벤처캘린더 13 밀키트 14 <del>카콜랜드크릭오거트</del> 15 <del>프로틴</del> 16 <del>코구마</del> 17 크로와상생지 18 <del>햇반</del> 19 그레놀라 20 크리스탈라이트	<b>식품 인기검색어</b> 2021.10. ~ 2022.09. <b>TOP 500</b> 1 포켓몬빵 2 <del>단가습살</del> 3 <del>삼립포켓몬빵</del> 4 <del>크릭요거트</del> 5 <del>카콜랜드프로틴바</del> 6 <del>프로대유청</del> 7 약과 8 어드벤처캘린더 9 <del>프로틴</del> 10 <del>오메가3</del> 11 <del>타이아르트도시락</del> 12 유산균 13 오소몰이문 14 아르기닌 15 <del>단백질쉐이크</del> 16 밀크씨슬 17 <del>삼다수</del> 18 콜라겐 19 밀키트 20 콩부차	<b>식품 인기검색어</b> 2022.10. ~ 2023.09. <b>TOP 500</b> 1 <del>단가습살</del> 2 <del>오메가3</del> 3 양배추즙 4 약과 5 아르기닌 6 유산균 7 <del>크릭요거트</del> 8 오소몰이문 9 포켓몬빵 10 비오틴 11 <del>코구마</del> 12 밀크씨슬 13 글루타치온 14 곤약젤리 15 <del>프로틴</del> 16 콜라겐 17 효소 18 사과 19 <del>단백질보충제</del> 20 <del>타이아르트도시락</del>

<표 1. 기간별 추출 키워드>

기간	추출 키워드	기간	추출 키워드
2017.10. ~ 2018.09.	중국당면	2020.10. ~ 2021.09.	고복침초코츄러스
	곤약젤리		오트밀미니바이트
	머랭쿠키		어드벤처캘린더
	마카롱		밀키트
	갈라만시		크로와상생지
2018.10. ~ 2019.09.	타피오카펠	2021.10. ~ 2022.09.	포켓몬빵
	마카롱		약과
	코호젤리		어드벤처캘린더
	하드코어버닝티		오소몰이문
	곤약젤리		아르기닌
2019.10. ~ 2020.09.	물티저스	2022.10. ~ 2023.09.	양배추즙
	일리캡슐		약과
	타피오카펠		아르기닌
	마카롱		오소몰이문
	사세버팔로윙		포켓몬빵

- '유행' 정의에 부합하는 데이터 추출 (python)

➔ <https://datalab.naver.com/shoppingInsight/sCategory.naver> 에서 2017년도~2023년도 사이 '식품' 카테고리에 해당하는 단어 중 <그림 1>을 따라 '유행'이라 칭 할 수 있는 키워드를 <표1>과 같이 년도별로 5개씩 추출한다.

➔ 2017년도~2023년도 내 총 30개의 키워드의 날짜 별 검색 량의 상대적 비율을 각각 csv화 한다. 이때, 10대~20대 남, 여가 모바일로 <표1>의 키워드를 검색한 빈도 수를 추출하고 구간별 결과에서 가장 큰 값을 100으로 설정한다.

1\_2021\_2022

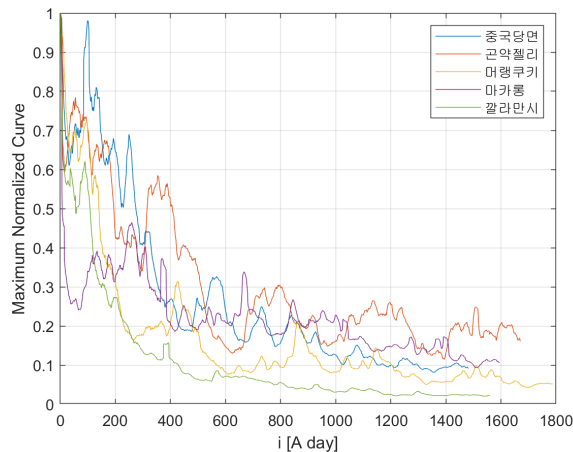
date			result1						
	0	2017-10-25	0.02483	1597	2022-03-10	84.73315			
	1	2017-10-26	0.02384	1598	2022-03-11	89.04983			
	2	2017-10-27	0.02483	1599	2022-03-12	100.0	2159	2023-09-23	0.15498
	3	2017-10-28	0.02483	1600	2022-03-13	98.62998	2160	2023-09-24	0.16491
				1601	2022-03-14	83.34227	2161	2023-09-25	0.18578

- 데이터 시각화 (matlab)

키워드 별 빈도수를 matlab을 활용하여 꺾은 선 그래프로 나타내어 추이를 확인한다.

- 1) 연간 검색량 상위 키워드를 그래프로 3개씩 표기한다.
- 2) 2022년도 이후의 데이터는 최고점 이후로 1년이상의 데이터가 존재하지 않으므로 '식품' 카테고리에서 '유행'의 정의에 부합하지 않아 본 연구에서 제외한다.

A. 그룹별 변화 추이: <그래프1>의 결과처럼 트렌드의 최고점을 찍은 후 그래프 모양은 자연지수 함수 형태와 비슷하다.



<그래프 1>

## B. A의 분석을 토대로 Least Square Method를 사용하여 Exp 함수 피팅한다.

```
%% Process 1 : Crop from max to end
[data_max, data_maxIdx] = max(data);
% data_proc1 = data(data_maxIdx:data_maxIdx+ 0);
% data_proc1 = data(1000:2162);
data_proc1 = data(data_maxIdx:data_maxIdx+80);

%% Outlier data manual removal
data_proc1(1) = []; % Remove outlier -> max outlier
% data_proc1(244) = []; % outlier

%% Process 2 : Normalize
maxVal = max(data_proc1);
minVal = min(data_proc1);

data_proc2 = (data_proc1 - minVal)/(maxVal - minVal) + 0.001;

%% Process 3 : Low pass filtering (movmean)
data_proc3 = movmean(data_proc2, 20);

%% The hypothesis is  $F(x) = \exp(-m \cdot x)$ 
% taking log on both sides

data_temp = data_proc2;
Y = log(data_temp);
X = [0:(length(data_temp)-1)]';

m = pinv(X)*Y;

%% Figure
figure;
plot(data_proc2, "r-"); hold on;

plot(data_proc3, "g-", 'LineWidth', 2); hold on;

t = 0:(length(data_temp)-1);
plot(exp(t*m), 'b-', 'LineWidth', 2)
```

### 1. %% Process 1: Crop from max to end

- 1-A. 키워드의 최대값을 가진 시점부터 600 일 이후까지의 데이터를 도시한다.
  - 1-B. 키워드의 최대값을 가진 시점부터 300 일 이후까지의 데이터를 도시한다.
  - 1-C. 키워드의 최대값을 가진 시점부터 200 일 이후까지의 데이터를 도시한다.
- A -> C 로 갈 수록 유행단어의 기한이 짧아진다.  
그래프 안정화 이후의 데이터 비율이 극대화되지 않도록 그래프를 피팅한다.

### 2. %% Outlier data manual removal

- 2-A. max 값(시작점) 제거: 최고점 값이 너무 클 경우, 이후의 변화를 잘 관찰하기 위함이다.
- 2-B. max 값(시작점) 제거 X: 최고점 이후 값의 변화가 보이지 않을 때 관찰하기 위함이다.

### 3. %% Process 2: Normalize

min\_Max 정규화를 통해 데이터의 최소값은 0, 최대값은 1로 정규화한다.

### 4. %% Process 3: Low pass filtering (movmean)

일일 데이터의 값들은 노이즈 하여 비교가 어렵기에 movmean(20일 평균)을 사용하여 정규화한다.

### 5. %% The hypothesis is $F(x) = \exp(-m \cdot x)$

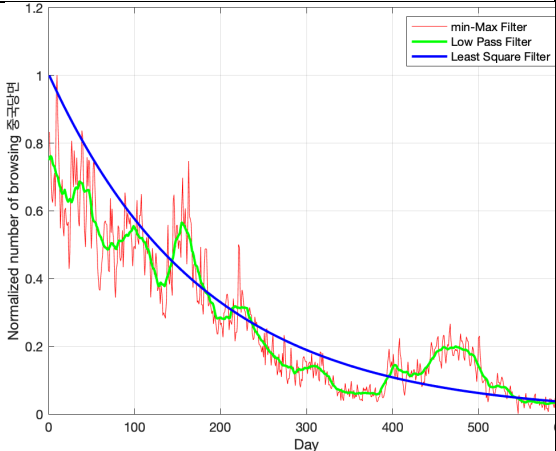

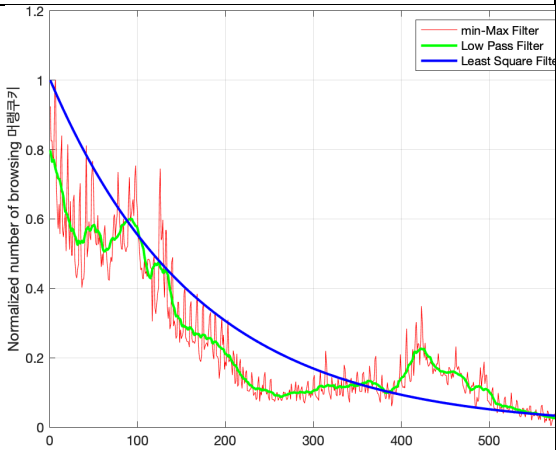
$Y = \exp(-m \cdot X)$  가설함수 피팅을 위해 양변에 log를 취하면 다음과 같은 수식이 정의된다.  
 $-m = \log(Y) \cdot \text{pinv}(X)$  (X, Y는 행렬. X = day, Y = day 별 키워드 검색량)

### 6. %% Figure

Redline -> 3.의 Process2 정규화까지 완료한 결과

Greenline -> 4.의 Process3 정규화까지 완료한 결과

Blueline -> Redline(3.의 Process2 정규화)의 결과를 가설함수 피팅한 결과

년도	데이터 추출범위	데이터 추출범위
2017-2018	(MaxIdx + 1) : 600	
1_중국당면  <u>1 시상수</u> <b>179.6</b>		
2_곤약젤리  <u>1 시상수</u> <b>264.9</b>		
3_머랭쿠키  <u>1 시상수</u> <b>155.0</b>		

년도	데이터 추출범위	데이터 추출범위
2018-2019	(MaxIdx + 1) : 600	(MaxIdx + 1) : 300
1_타피오카 필  <u>1 시상수</u> <b>154.8</b>		
2_마카롱  <u>1 시상수</u> <b>128.3</b>		
3_교호젤리  <u>1 시상수</u> <b>63.9</b>		

년도	데이터 추출범위	데이터 추출범위
2019-2020	(MaxIdx + 1) : 600	(MaxIdx + 1) : 200
1_몰티저스  <u>1 시상수</u> <b>74.7</b>		
	(MaxIdx) : 600	
2_일리카슬  <u>1 시상수</u> <b>205</b>		
	(MaxIdx + 1) : 600	(MaxIdx + 1) : 300
3_타피오카 펠  <u>1 시상수</u> <b>154.8</b>		

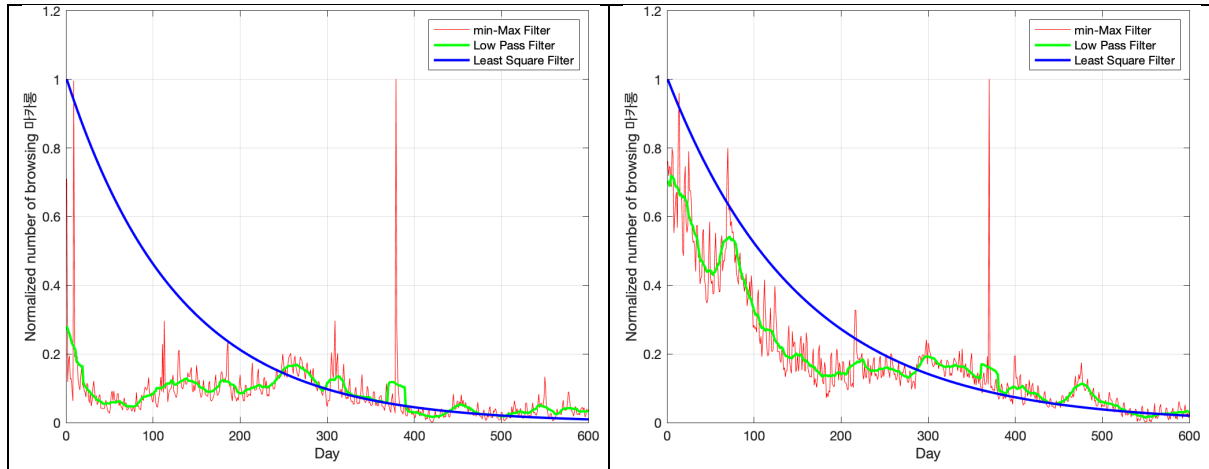
년도	데이터 추출범위	데이터 추출범위
2020-2021	(MaxIdx + 1) : 600	(MaxIdx+1) : 200
1_꼬북칩 초코츄리스  <u>1 시상수</u> <b>47.0</b>		
2_오토밀 미니바이트  <u>1 시상수</u> <b>161.2</b>		
	(MaxIdx) : 600	(MaxIdx) : 200
3_어드벤처 캘린더  <u>1 시상수</u> <b>26.7</b>		



년도	데이터 추출범위	데이터 추출범위
2021-2022	(MaxIdx + 1) : 600	(MaxIdx + 1) : 200
1_포켓몬빵  <u>1 시상수</u> 36.3		
	(MaxIdx) : 600	(MaxIdx) : 200
2_약과  <u>1 시상수</u> 66.0		
3_어드벤처 캘린더  <u>1 시상수</u> 26.7		

### <예외 1: 마카롱, 타피오카펄>

특정단어가 2 년 연속으로 검색량 상위권에 위치한 경우, 유행이 2 년간 꾸준하다고 말하기는 어렵다.

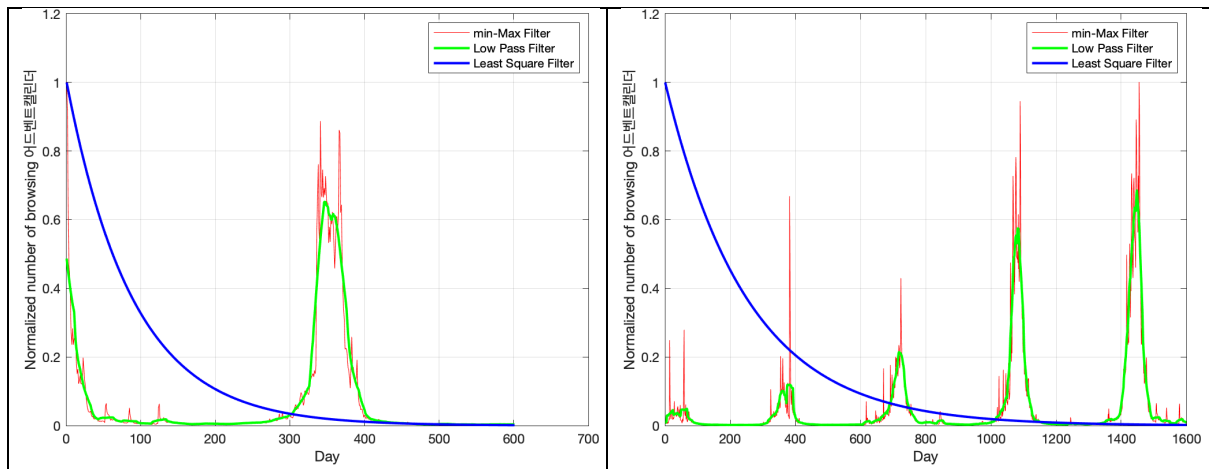


<마카롱>

<타피오카펄>

위 두 경우는 검색량이 가장 많은 날 이외에 검색량이 많은 하루(튀는 점)로 2 년 연속 유행단어 상위권에 위치한다.

### <예외 2: 어드벤처캘린더>



<최고점 이후로 600 일간의 그래프 추이>

<2017-2022 년도 그래프 추이>

어드벤처캘린더와 같이 2 년 연속으로 검색량 상위권에 위치하고, 일 년을 주기로 자연지수함수가 반복되는 경우는 2020-2021 년도 이후로 유행을 따르다가 계절의 영향을 받는 키워드가 된 경우이다.

### <결론 1>

유행하는 특정단어의 검색량 그래프는 최고점을 기준으로 자연지수함수(Exponential Function)모양을 띈다. 즉,  $F(x) = \exp(-m \cdot x)$  그래프로 피팅된다. 이 때, 키워드의 검색량 최고점을 찍은 시점으로부터 자연함수의 1 시상수가 지난 시점이 유행이 지난 시점이다.

### <결론 2>

해(year)가 지날수록 년도 별 상위 키워드 3 개의 1 시상수 평균이 낮아지는 경향을 보이므로 유행의 기한이 짧아진다.

년도	1 시상수		
2017-2018	179.6	264.9	155.0
2018-2019	154.8	128.3	63.9
2019-2020	74.7	205	154.8
2020-2021	47.0	161	-
2021-2022	36.3	66.0	-

\*유행단어로 칠 수 없는 어드벤트캘린더는 1 시상수 평균에서 제외하므로 -로 표기

년도별 1 시상수 평균 변화	299.8 -> 115.6 -> 144.8 -> 104 -> 51.1
-----------------	--