행복은 장바구니를 타고

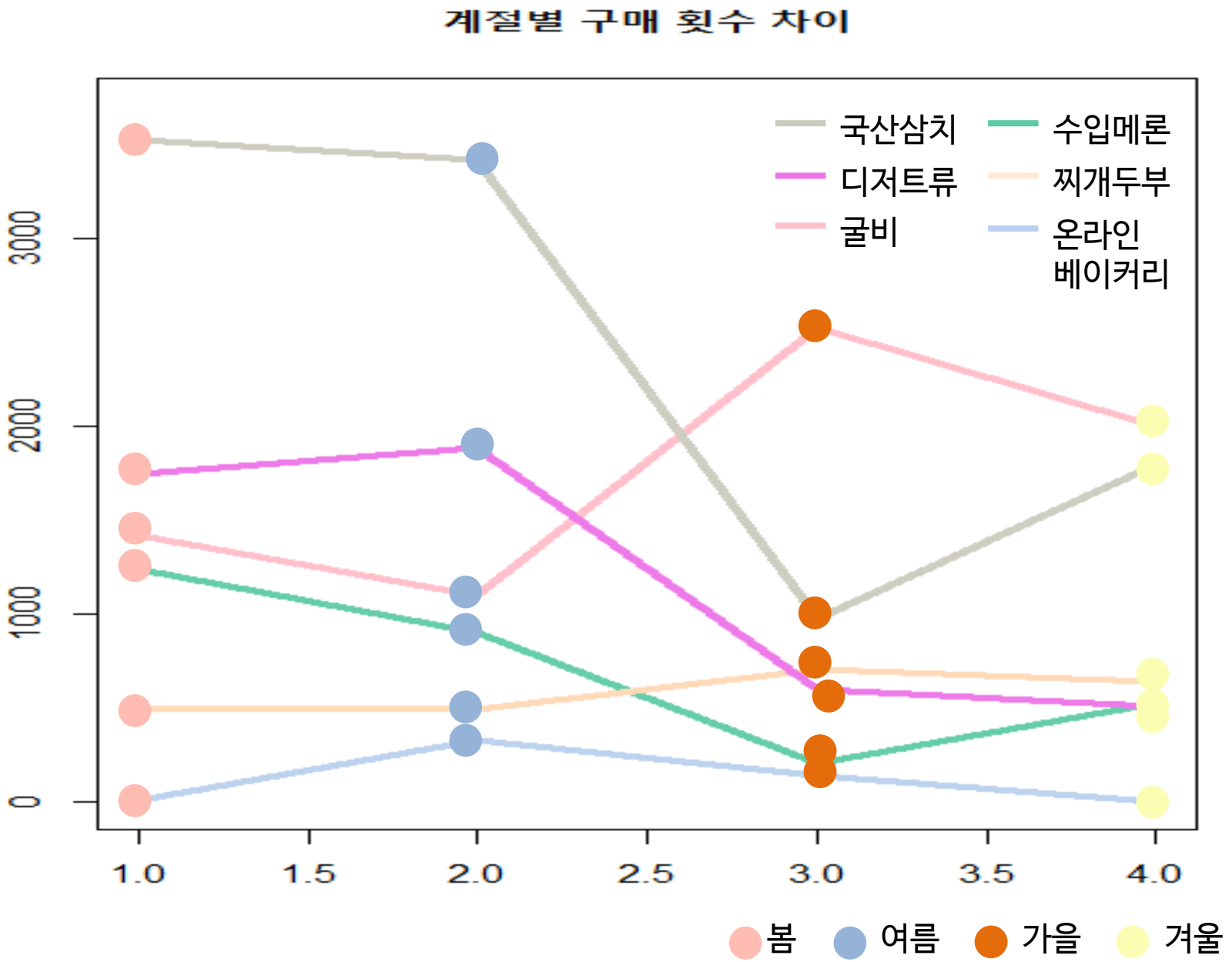
상어초밥팀
박이삭 이다경 김서연

INDEX.

1 데이터 전처리

겨울 데이터 제작

- 계절별로 구매물품에 차이가 있을 거라 생각.
- 실제 분석 결과 오른쪽 그림처럼 **계절별 물품 구매횟수 차이** 확인
- 이에 따라 겨울(12월, 1월 2월)에 물건을 구매한 사람들에 대한 **"겨울 데이터"** 생성

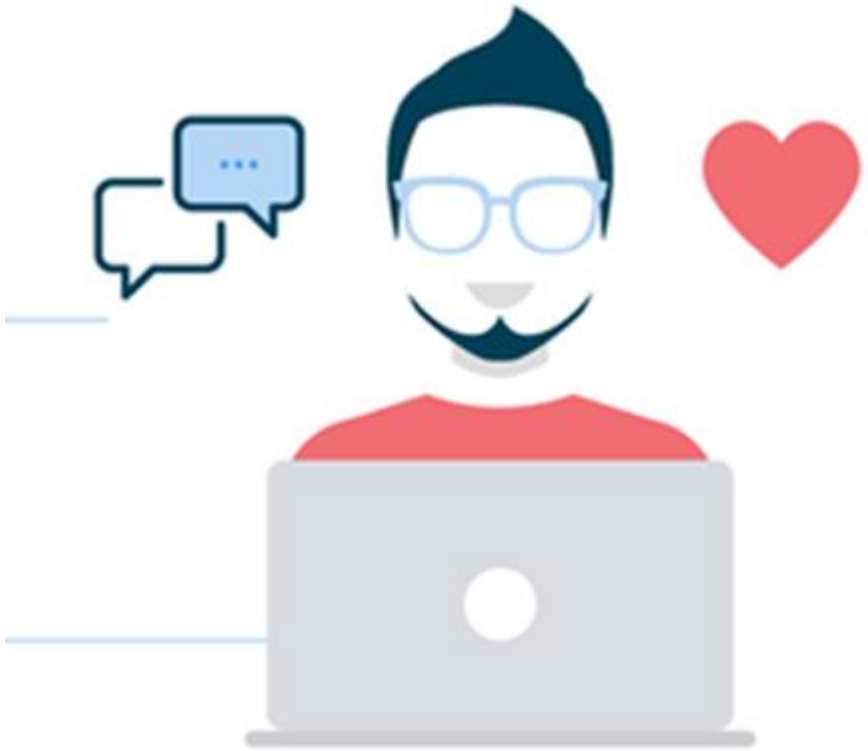


1 데이터 전처리

겨울 데이터 제작 : 계절별 구매물품 차이 검정

물품 구매 횟수에 계절별로 통계적 차이가 있음을 확인하기 위해 ANOVA 검정 실시

```
             Df Sum Sq Mean Sq F value   Pr(>F)
group         3     26   8.535    11.4 1.83e-07 ***
Residuals 17540  13132   0.749
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



- 봄, 여름, 가을, 겨울
  → DF = 4-1 = 3

- 귀무가설 : Mean(봄) = Mean(여름) = Mean(가을) = Mean(겨울)

- P-value =1.83e-07 = 0.000000183 ⟨⟨ 0.05
  이므로 **계절별 차이가 있음을 확인**

1 데이터 전처리

분석하기 앞서 생성한 데이터

분석하기 앞서 생성한 데이터

- 물건기반협업필터링 & Word2Vec 알고리즘에 적용

구매횟수 데이터

고객별 해당 물품을 몇번 구매했는지
나타내는 데이터

구매목록 데이터

각 고객이 2년(겨울)동안 구매한 구매
목록을 순서대로 나열한 데이터

1-1 데이터 전처리

분석하기 앞서 생성한 데이터

1. 구매횟수 데이터 : 고객별 해당 물품을 몇번 구매했는지 나타내는 데이터

고객번호	A010101	A010102	A010103	A010104	A010105	A010106	A010201	…
3071	0	10	3	250	0	7	0	…
4074	1	2	0	1	0	0	6	…
10719	70	0	53	22	0	0	1	…
…	…	…	…	…	…	…	…	…
4825	0	0	0	0	1	0	2	…
17357	0	0	28	0	0			
18761	0	1	0	0	0			

겨울에 하나이상 구매한 고객 : 19372명
19372명의 총 물품목록 : 4137개

→ 고객별 해당 물품 구매 횟수 매트릭스 생성
*겨울에 구매한 물품이 없는 고객 제거

1-2 데이터 전처리

분석하기 앞서 생성한 데이터
2. 구매목록 데이터 : 각 고객이 2년(겨울)동안 구매한 구매목록을 순서대로 나열한 데이터

고개번호	구매1	구매2	구매3	구매4	…	구매486	구매487	구매488
17218	B150401	B160101	B160201	B180301	…	B300601	B090402	B550601
17674	B050901	B150101	B050311	B050701	…	B180204	B610202	B380504
14388	B100306	B430101	B540301	B340402	…	B180303	B140607	B140601
…	…	…	…	…	…	…	…	…
2975	A011003	B520103			…			
7111	B340404				…			
17454	B340103	B100306	B160201	B480202	…			

겨울에 하나 이상 구매한 고객 : 19372명
19372명의 사람 중 최다 구매 횟수 : 488회

→ 각 고객별 겨울기간 동안 구매한 상품을 나열한 구매목록 매트릭스 생성

2 모델 개발 및 분석 과정

물품기반 협업필터링 & Word2Vec

모델 개발 과정

1.물품기반 협업필터링

내가 구매한 물품들과
유사한 물품 추천

2. Word2Vec

고객과
거리가 가까운
물품 추천

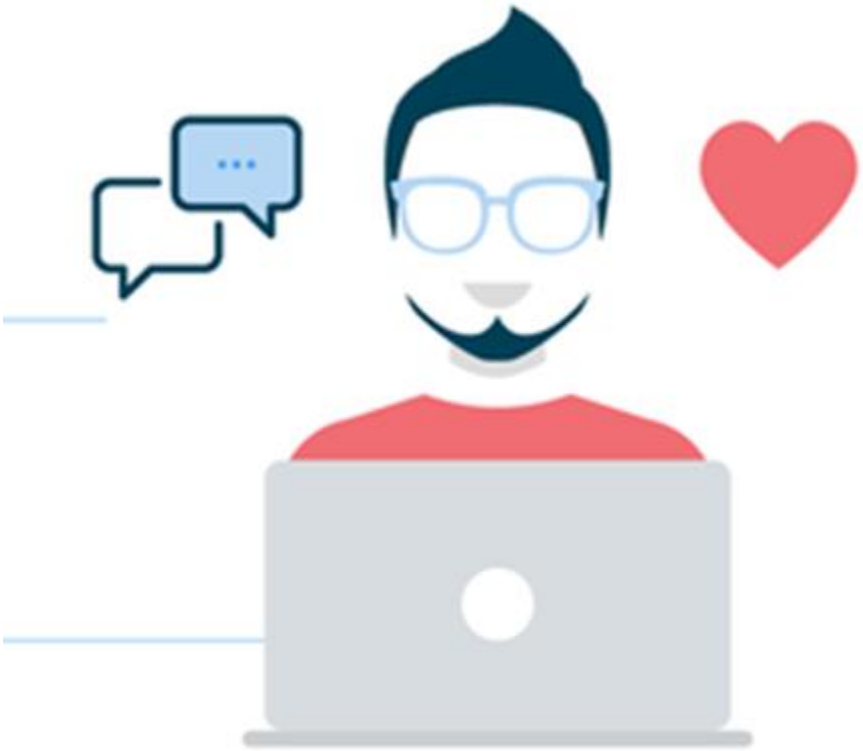
1. 물품기반 협업필터링

2-1 물품기반 협업필터링

물품기반 협업필터링의 정의

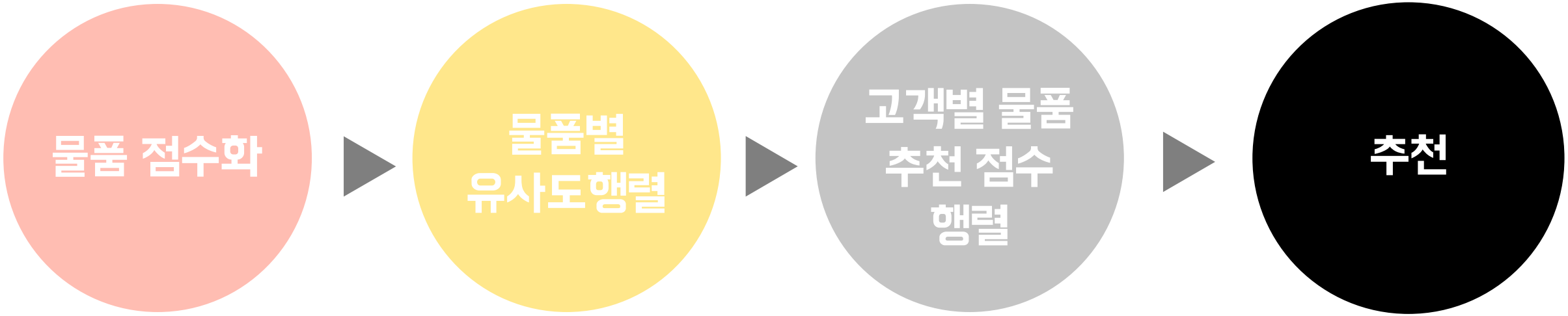
협업 필터링이란?
고객들의 프로파일정보를 활용하여 목표고객이 높게 평가할 것으로 예상되는 서비스나 아이템을 추천하는 기법.
(출처 : 장르별 협업필터링을 이용한 영화 추천 시스템의 성능 향상, 이재식.박석두, 2007년 12월)

물품기반 협업필터링(item-based collaborative filtering) 이란?
물품에 대한 선호도를 기반으로 물품들사이의 유사도를 구하여, 고객에게 물품을 추천해주는 협업필터링의 한 기법.



2-1 물품기반 협업필터링

적용 방법, 과정



고객별 물품에 대한
선호도 점수화

물품별 거리를 측정하여
유사도 행렬 생성

고객별 물품 추천
점수 행렬 생성

최종 물품 추천

2-1 물품기반 협업필터링

적용 방법, 과정
1. 물품 점수화 : 고객별 물품에 대한 선호도 점수화

물품 점수화

고객 \ 물품	A010101	A010102	…	D080302	D080401
3071	3.30	1.64	…	5.01	3.34
4707	0	0	…	6.68	10.7
…	…	…	…	…	…
18901	1.65	3.36	…	10	2.64
18154	0	1.68	…	10	100.5

물품 점수화

- 물품기반 협업필터링을 진행하기 위해서는 고객별 **'물품에 대한 선호도'** 필요
- 구매횟수 데이터에서 **TF-IDF** 기법을 이용하여 물품점수화. 다음장에 예를 들어 설명
- 0점부터 765.1점까지

2-1 물품기반 협업필터링

적용 방법, 과정
1. 물품 점수화 : TF-IDF란?

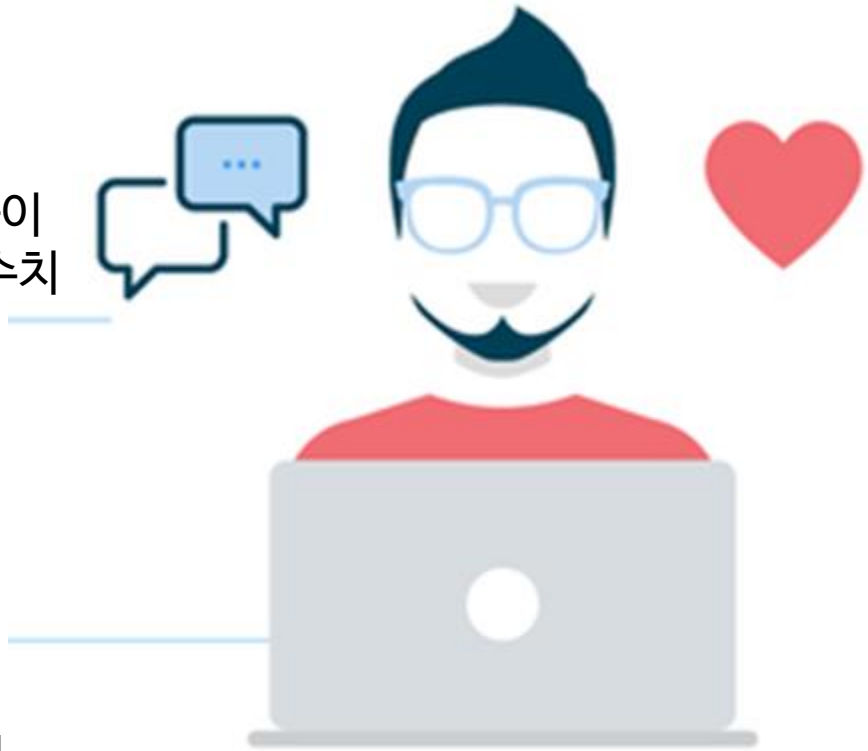
물품 점수화

TF-IDF란?
정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치

$$TF\text{-}IDF = TF/DF$$

TF : Term Frequency
DF : Document Frequency

*IDF = 1/DF



다음장에 예를 들어 설명

2-1 물품기반 협업필터링

적용 방법, 과정
1. 물품 점수화 : TF-IDF란?

물품 점수화

문서 = 고객한명한명의 구매목록
단어 = 구매물품들

TF : 특정단어(term)가 하나의 문서(document)내에 나타난 빈도
DF : 전체 문서군(global document) 중에서 특정단어(term)를 포함하는 문서 빈도

TF-IDF 예시

고객번호	구매목록1	구매목록2	구매목록3	구매목록4
1	사과	배	사과	배
2	자전거	껌	빵	딸기
3	사과	딸기	맥주	오렌지
4	배	사과	술	쉐이빙크림
5	사과	떡	오렌지	껌
6	전기면도기	쉐이빙크림	맥주	자전거
7	사과	맥주	떡	딸기

〈구매목록 예시〉

전체 문서 수(전체 고객 수) = 7

사과의 DF
= (전체 고객 중 사과를 포함하고 있는 고객 수)
= 5 (1번, 3번, 4번, 5번, 7번고객)

배의 DF
= (전체 고객 중 배를 포함하고 있는 고객 수)
= 2 (1번, 4번고객)

2-1 물품기반 협업필터링

적용 방법, 과정
1. 물품 점수화 : TF-IDF란?

물품 점수화

고객번호	구매목록1	구매목록2	구매목록3	구매목록4
1	사과	배	사과	배
2	자전거	껌	빵	딸기
3	사과	딸기	맥주	오렌지
4	배	사과	술	쉐이빙크림
5	사과	떡	오렌지	껌
6	전기면도기	쉐이빙크림	맥주	자전거
7	사과	맥주	떡	딸기

〈구매목록 예시〉

1번 고객에 대한 설명

사과의 TF
= (고객의 구매목록 중 사과의 개수)
= 2

사과에 대한 TF-IDF
= TF/DF
= 2/5 = 0.4

배의 TF
= (고객의 구매목록 중 배의 개수)
= 2

배에 대한 TF-IDF
= TF/DF
= 2/2 = 1

➡ 1번고객의 사과 점수 < 배 점수

2-1 물품기반 협업필터링

적용 방법, 과정
1. 물품 점수화 : TF-IDF란?

물품 점수화

고객번호	구매목록1	구매목록2	구매목록3	구매목록4
1	사과	배	사과	배
2	자전거	꿀	빵	딸기
3	사과	딸기	맥주	오렌지
4	배	사과	술	쉐이빙크림
5	사과	오렌지	콘	껌
6	전기면도기	쉐이빙크림	맥주	자전거
7	사과	맥주	떡	물

〈구매목록 예시〉

특정 고객에게는 많이 등장하나,
전체적으로 적게 등장하는 물품일수록
TF-IDF 大

1번 고객에 대한 설명

사과의 TF
= (고객의 구매목록 중 사과의 개수)
= 2

사과에 대한 TF-IDF
= TF/DF
= 2/5 = 0.4

배의 TF
= (고객의 구매목록 중 배의 개수)
= 2

배에 대한 TF-IDF
= TF/DF
= 2/2 = 1

➡ **1번고객의 사과 점수 < 배 점수**

2-1 물품기반 협업필터링

적용 방법, 과정
1. 물품 점수화 : TF-IDF란?

물품 점수화

고객번호	구매목록1	구매목록2	구매목록3	구매목록4
1	사과	배	사과	배
2	전기거			기
3	사과	딸기	맥주	오렌지
4	배	사과	술	쉐이빙크림
5	사과	떡	오렌지	껌
6	전기면도기	쉐이빙크림	맥주	비누
7	사과	맥주	떡	딸기

〈구매목록 예시〉

1번 고객에 대한 설명

사과의 TF
= (고객의 구매목록 중 사과의 개수)
= 2

TF/DF
= 2/2 = 1

배의 TF
= (고객의 구매목록 중 배의 개수)
= 2

TF/DF
= 2/2 = 1

→ 한 고객이 두 물품을 같은 빈도수로 구매하여도,
TF-IDF값이 클수록
해당 물품의 중요성 증가

여러 고객에게 일반적으로 등장하지 않고
소수의 고객에게만 등장하는 정도

➡ 1번고객의 사과 점수 < 배 점수

2-1 물품기반 협업필터링

적용 방법, 과정
2. 물품별 유사도 행렬 : 물품별 거리를 측정하여 유사도 행렬 생성

물품별 유사도 행렬

	A010101	A010102	…	D080302	D080401
A010101	1	0.8124	…	0.003	0.006
A010102	0.8124	1	…	0	0.007
…	…	…	…	…	…
D080302	0.003	0	…	1	0.308
D080401	0.006	0.007	…	0.308	1

다음장에 예를 들어 설명

물품별 유사도 행렬

- 물품을 점수화한 데이터를 바탕으로 각 물건별 **유사도 매트릭스** 생성.
- “코사인유사도”와 “유클리디안 거리” 중 **코사인 유사도**를 이용하여 유사도 계산
23페이지 참조
- 물품 x와 y의 코사인 유사도 :

$$\text{cosSimilarity}(x, y) = (x \cdot y) / (||x|| * ||y||)$$

2-1 물품기반 협업필터링

적용 방법, 과정
2. 물품별 유사도 행렬 : 물품 유사도 계산 예시

물품별 유사도 행렬

A010101 & A010102 물품의 유사도 예시

물품 A010101와 A010102의 코사인 유사도 :
cosSimilarity(A010101, A010102)
= (A010101 · A010102) / (||A010101|| * ||A010102||)

$$= (1x0 + 9x5 + 4x10 + 5x7 + 3x8) / \sqrt{(1^2 + 9^2 + 4^2 + 5^2 + 3^2) \times (0^2 + 5^2 + 10^2 + 7^2 + 8^2)}$$

$$= 0.8124$$

	A010101	A010102
A010101	1	0.8124

〈물품별 유사도 행렬〉

물품 \ 고객	1번고객	2번고객	3번고객	4번고객	5번고객
A010101	1	9	4	5	3
A010102	0	5	10	7	8

〈물품 점수 행렬 –물품 x 고객〉

2-1 물품기반 협업필터링

적용 방법, 과정
3. 고객별 물품 추천 점수 행렬 :
고객별 물품 추천 점수 행렬 생성

고객별 물품 추천 점수 행렬

고객 \ 물품	A010101	A010102	…	D080302	D080401
3071	0.028	0.031	…	0.031	0.0317
4704	0.04	0.0506	…	0.049	0.05
…	…	…	…	…	…
18901	0.00178	0.0016	…	0.0013	0.0015
18154	0.0002	0.00019	…	0.00019	0.00018

3071번 고객에 A010101물품을 추천할 점수

고객별 물품추천 점수행렬

- 앞서 만든 유사도 행렬을 바탕으로
  고객별 모든 물품에 추천 점수 유도
  다음장에 예를 들어 설명
- 점수가 가장 높은 **3개의 물품 추천**

2-1 물품기반 협업필터링

적용 방법, 과정
3. 고객별 물품 추천 점수 행렬 : 점수행렬 계산 식

고객별 물품 추천 점수 행렬

점수행렬 계산 식

4137개의 모든 아이템 : item(1), item(2), …, item(4137)

유저 i가 구매한 M(i)개의 아이템 : record(i, 1), record(i, 2), …, record(i, M(i))

유저 i에 대한 아이템 x의 추천 점수 score(i, x) :
score(i, x) = (itemsSimilarity(x, record(i, 1)) + … + itemsSimilarity(x, record(i, M(i)))) / (itemsSimilarity(x, item(1)) + … + itemsSimilarity(x, item(K)))
(출처 : Code Sprint 2015 Round 2 - 협업 필터링으로 유저 기호 예측하기,김재겸, 2015년 9월)

다음장에 예를 들어 설명

2-1 물품기반 협업필터링

적용 방법, 과정
3. 고객별 물품 추천 점수 행렬 : 점수행렬 계산 예

고객별 물품 추천 점수 행렬

점수행렬 계산 예

〈1번 고객의 구매목록 데이터〉

고객번호	구매목록1	구매목록2	구매목록3
1	사과	배	자전거

〈물품별 유사도 행렬〉

	사과	배	…	껌	자전거	합
사과	1	0.78	…	0.53	0.006	**79.24**
배	0.89	1	…	0.308	0.007	
…	…	…	…	…	…	
껌	0.53	0.308	…	1	0.003	
자전거	0.006	0.007	…	0.003	1	

사과, 배, 자전거를 구매한 1번 고객의 사과 추천 점수는?

score(1,**사과**)
= ((**사과**, 사과 유사도) + (**사과**, 배 유사도)
+ (**사과**, 자전거 유사도))
/ (사과와 모든 물품들의 유사도 합)

= (1 + 0.78 + 0.006) / 79.24 = 0.0225

고객 \ 물품	사과
1	0.0225

〈물품 추천 점수 행렬〉

2-1 물품기반 협업필터링

적용 방법, 과정
3. 고객별 물품 추천 점수 행렬 : 점수행렬 계산 예

고객별 물품 추천 점수 행렬

점수행렬 계산 예

〈1번 고객의 구매목록 데이터〉

고객번호	구매목록1	구매목록2	구매목록3
1	사과	배	자전거

〈물품별 유사도 행렬〉

	사과	배	…	껌	자전거	합
사과	1	0.78	…	0.53	0.006	
배	0.89	1	…	0.308	0.007	
…	…	…	…	…	…	
껌	0.53	0.308	…	1	0.003	**83**
자전거	0.006	0.007	…	0.003	1	

사과, 배, 자전거를 구매한 1번 고객의 껌 추천 점수는?

score(1,**껌**)
= ((**껌**, 사과 유사도) + (**껌**, 배 유사도)
+ (**껌**, 자전거 유사도))
/ (껌과 모든 물품들의 유사도 합)

= (0.53 + 0.308 + 0.003) / 83 = 0.01

고객 \ 물품	사과	껌
1	0.0225	0.01

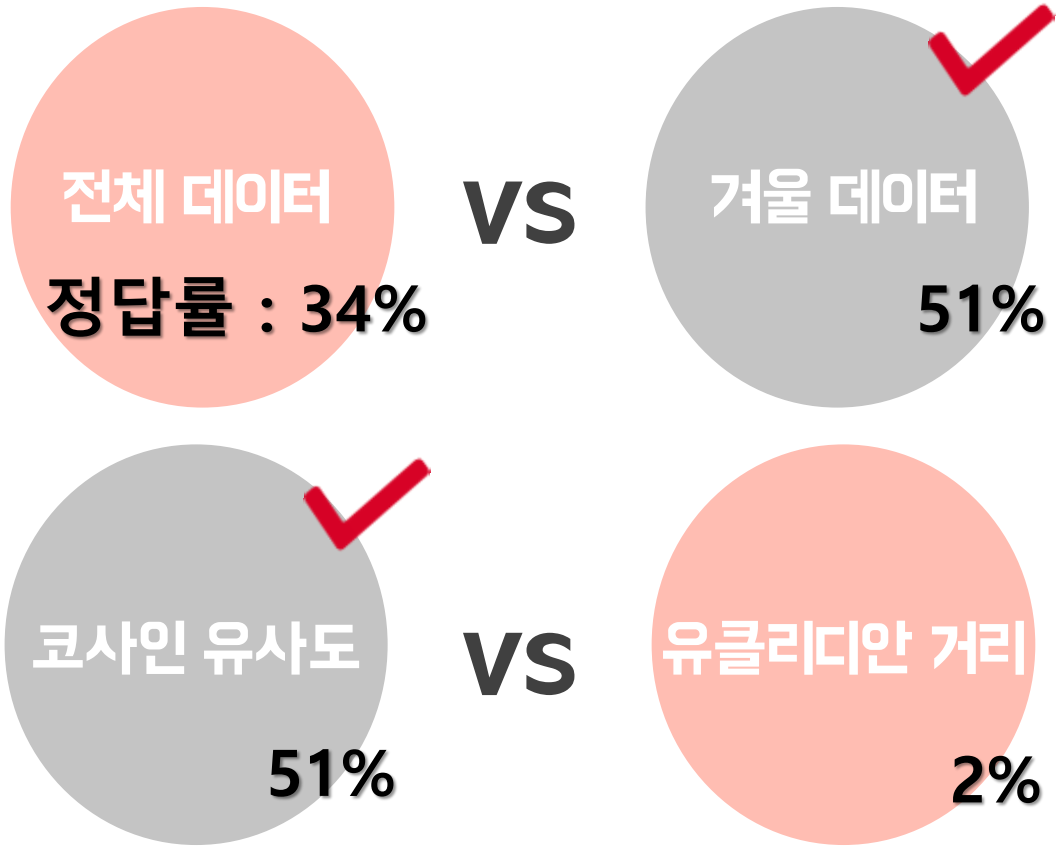
〈물품 추천 점수 행렬〉

2-1 물품기반 협업필터링

적용 방법, 과정
여러 모델 학습 후 검증

2014년 데이터로 모델 학습 , 2015년 구매 예측
→ 정답률 확인 후 비교

모델 학습 & 검증



겨울 구매 데이터로
물품간 코사인 유사도를 구하여
최종 모델 생성

2. Word 2 Vec

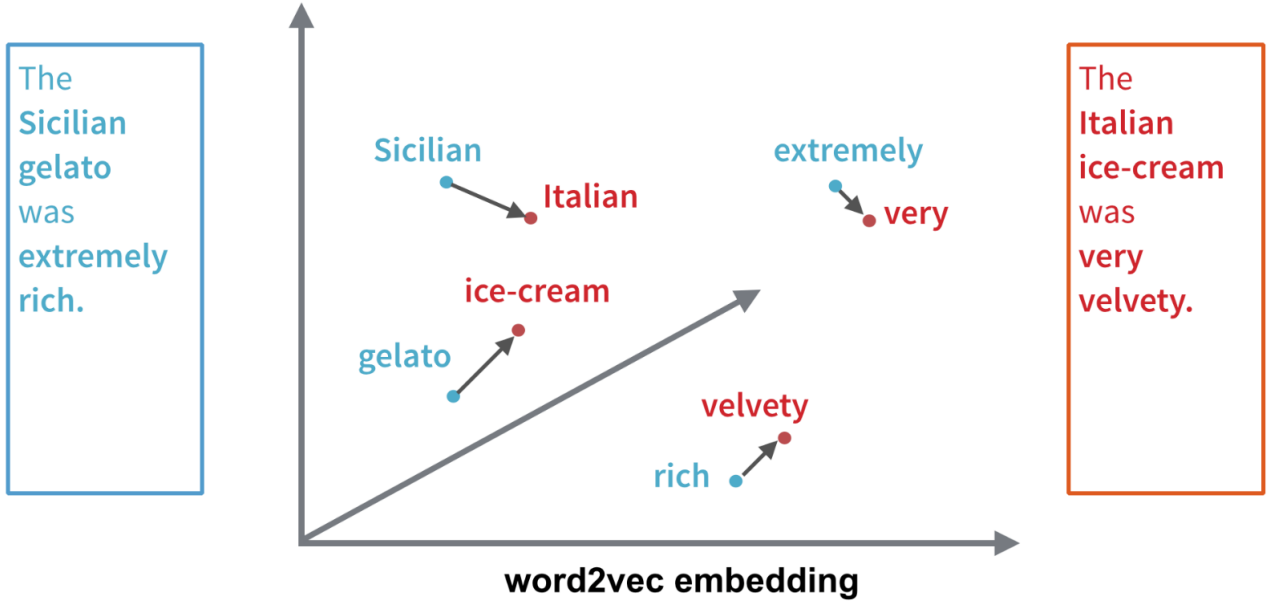
2-2 Word2Vec

Word2Vec의 정의

Word2Vec이란?

- 구글의 연구원들이 “Efficient Estimation of Word Representations in Vector Space”라는 논문에서 제안하여 구현된 알고리즘
- 인공신경망의 한 종류로, **단어를 벡터로 표현**

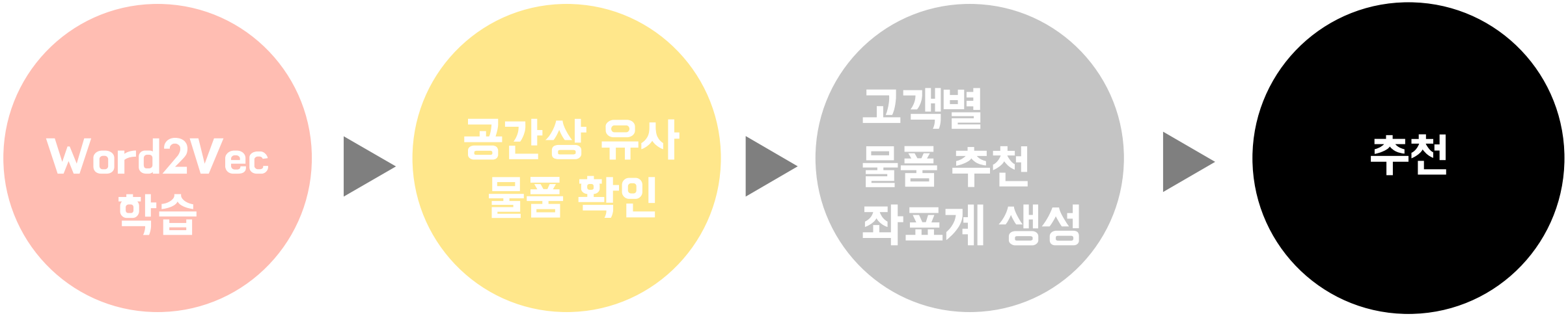
- 텍스트 문서를 학습시키는 과정을 통하여 비슷한 의미를 갖고 있는 단어들끼리 더 가까운 벡터를 갖고, 상반된 의미를 갖고 있는 경우 더 멀리 떨어지도록 학습



(출처: ICT Portal Media, 〈Word2Vec, 자연어 기계학습의 혁명적 진화, 2014년11월20일)

2-2 Word2Vec

Word2Vec의 적용방법, 과정



고객별 구매 물품
Word2Vec알고리즘
학습

물품별 공간상 거리를
측정하여
유사한 물품 확인

고객별 물품
추천 좌표계 생성
→ 고객과 물품간 거리 확인

고객과 가까운
물품부터 추천

2-2 Word2Vec

Word2Vec의 적용방법, 과정

1. Word2Vec 학습 :
   고객별 구매 물품 Word2Vec 알고리즘 학습

Word2Vec 학습

- 한 문장 속에서 단어를 찾아 벡터로 표현하는 Word2Vec의 원리를 응용

- **'고객들이 구입한 물품들은 서로 연관성이 있다'**

문장 ▶ 구매목록

한 고객이 구매한 전체 구매목록을 하나의 문장으로 인식

단어 ▶ 구매물품

각 구매물품들을 단어로 보고 학습함

2-2 Word2Vec

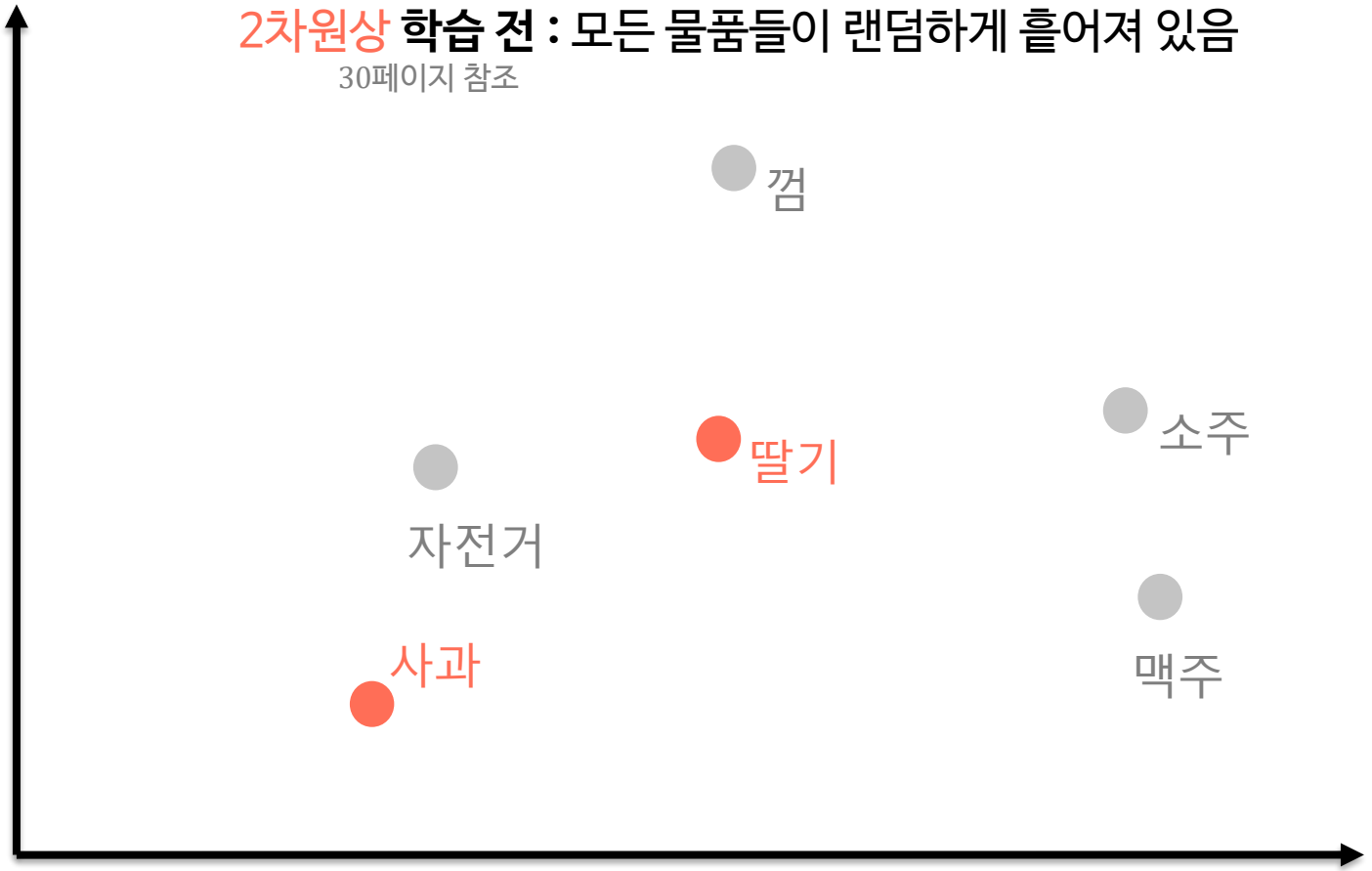
Word2Vec의 적용방법, 과정
1. Word2Vec 학습 : 학습 효과

Word2Vec 학습

고객 번호	1
구매 목록	사과 ✓
	딸기 ✓
	자전거
	맥주

2차원상 **학습 전 :** 모든 물품들이 랜덤하게 흩어져 있음

30페이지 참조



고객 번호	2
구매 목록	사과 ✓
	소주
	딸기 ✓
	껌

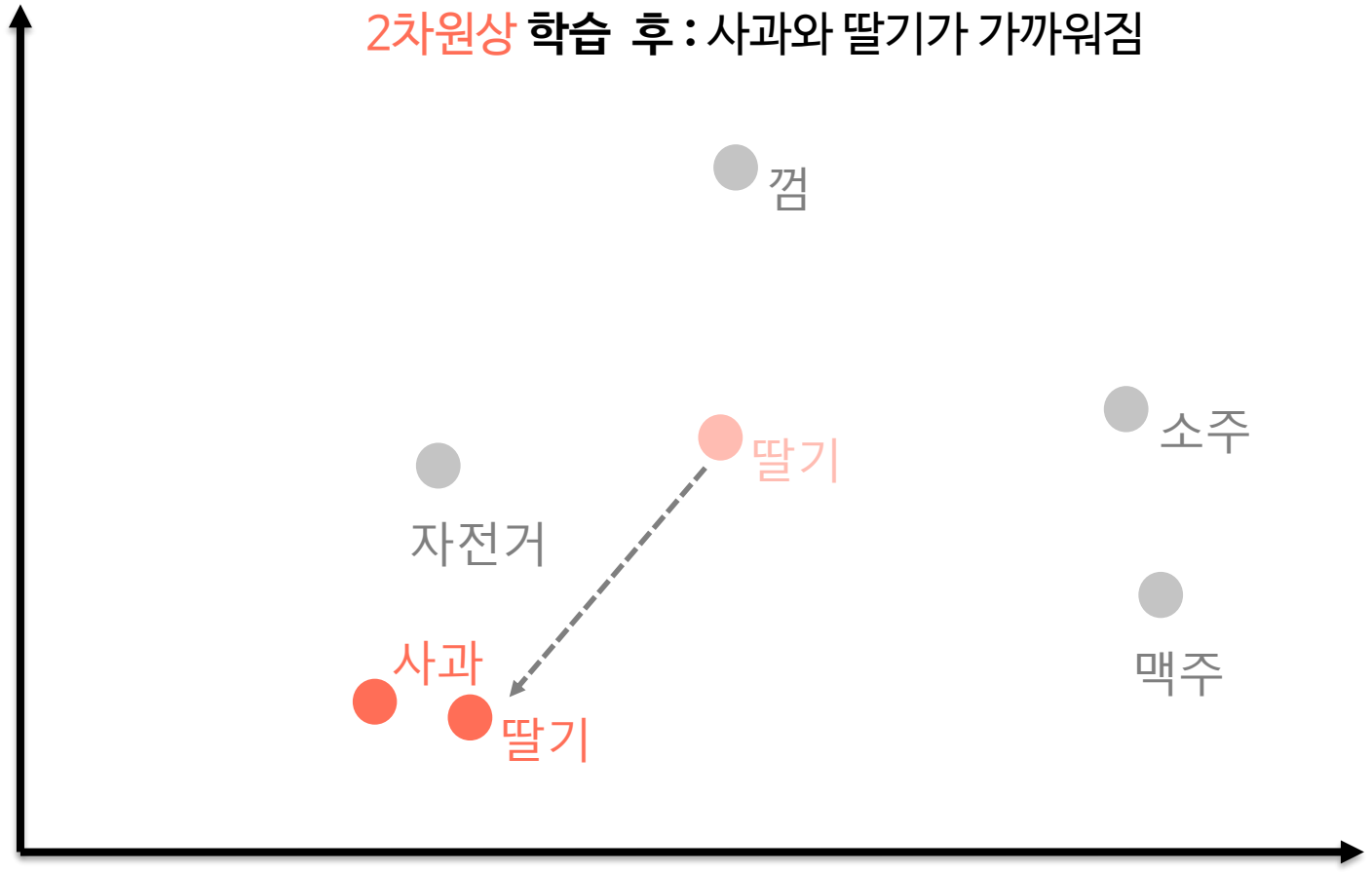
2-2 Word2Vec

Word2Vec의 적용방법, 과정
1. Word2Vec 학습 : 학습 효과

Word2Vec 학습

고객 번호	1
구매 목록	사과 ✓
	딸기 ✓
	자전거
	맥주

2차원상 **학습 후 :** 사과와 딸기가 가까워짐



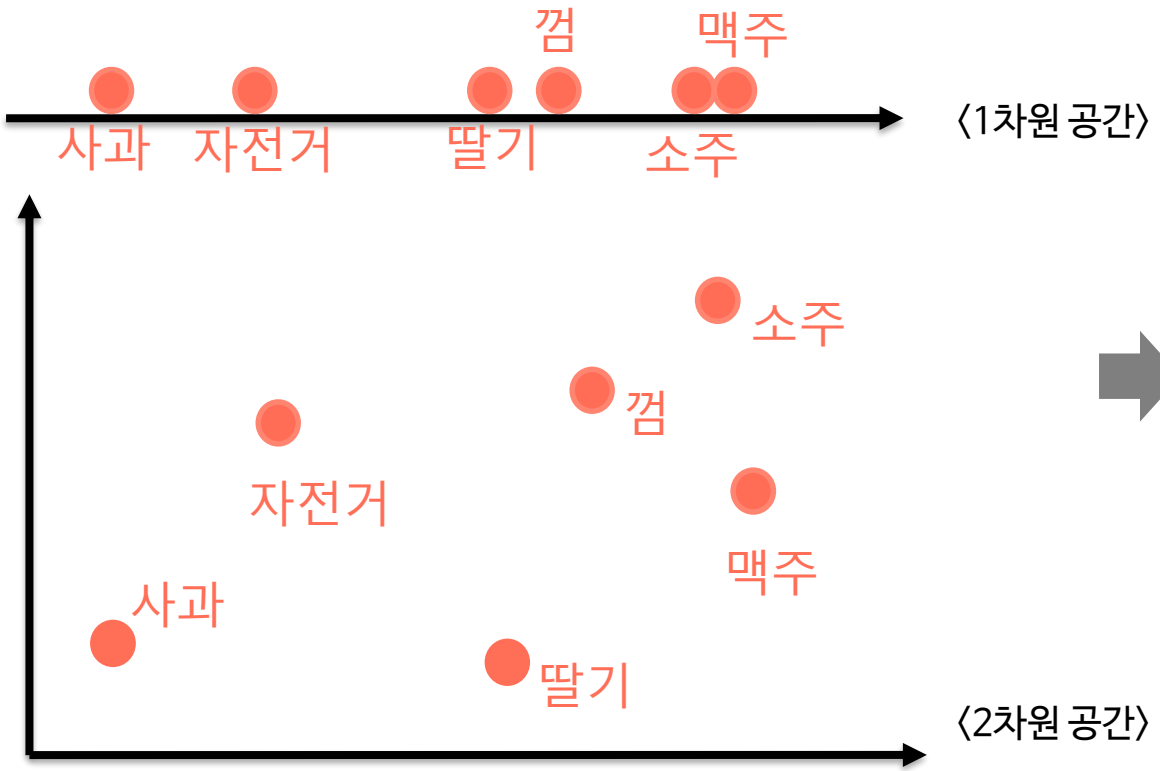
고객 번호	2
구매 목록	사과 ✓
	소주
	딸기 ✓
	껌

2-2 Word2Vec

Word2Vec의 적용방법, 과정
1. Word2Vec 학습 : 차원이란?

Word2Vec 학습

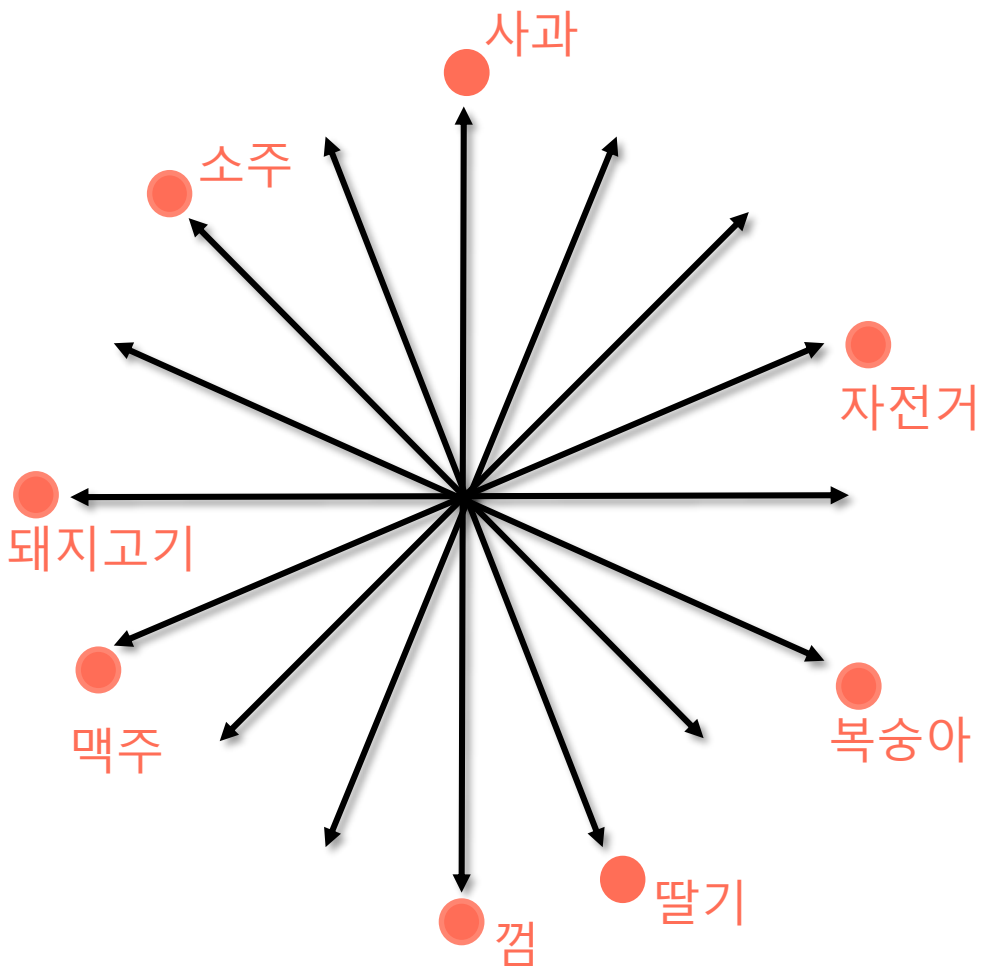
차원이란?
- Word2Vec 알고리즘을 학습시킬 공간의 차원



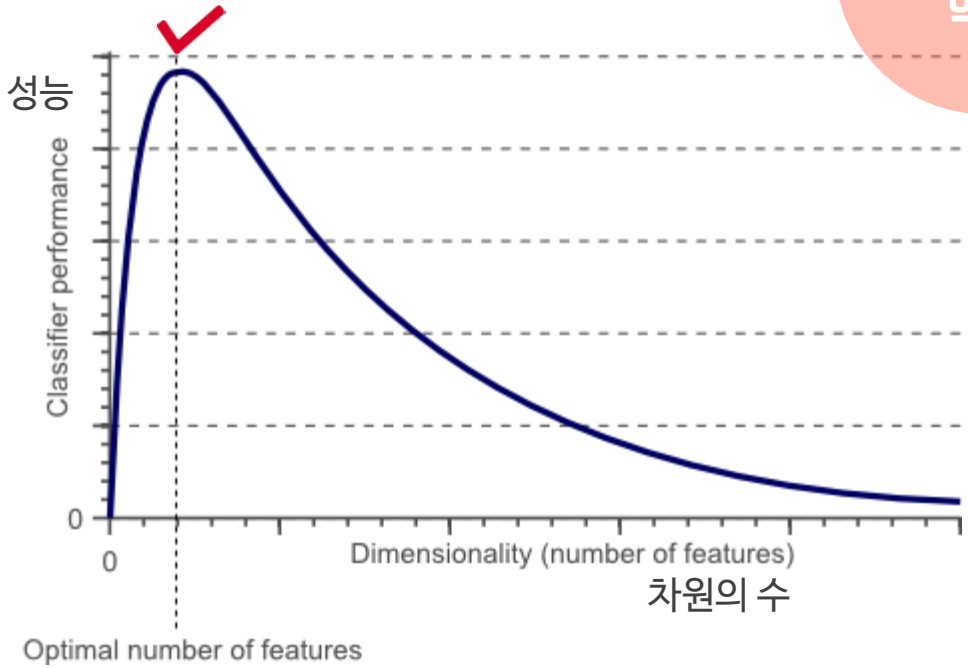
차원이 늘어나면 축이 늘어나므로
더 정확한 좌표 측정 가능

2-2 Word2Vec

Word2Vec의 적용방법, 과정
1. Word2Vec 학습 : 차원이란?



Word2Vec 학습

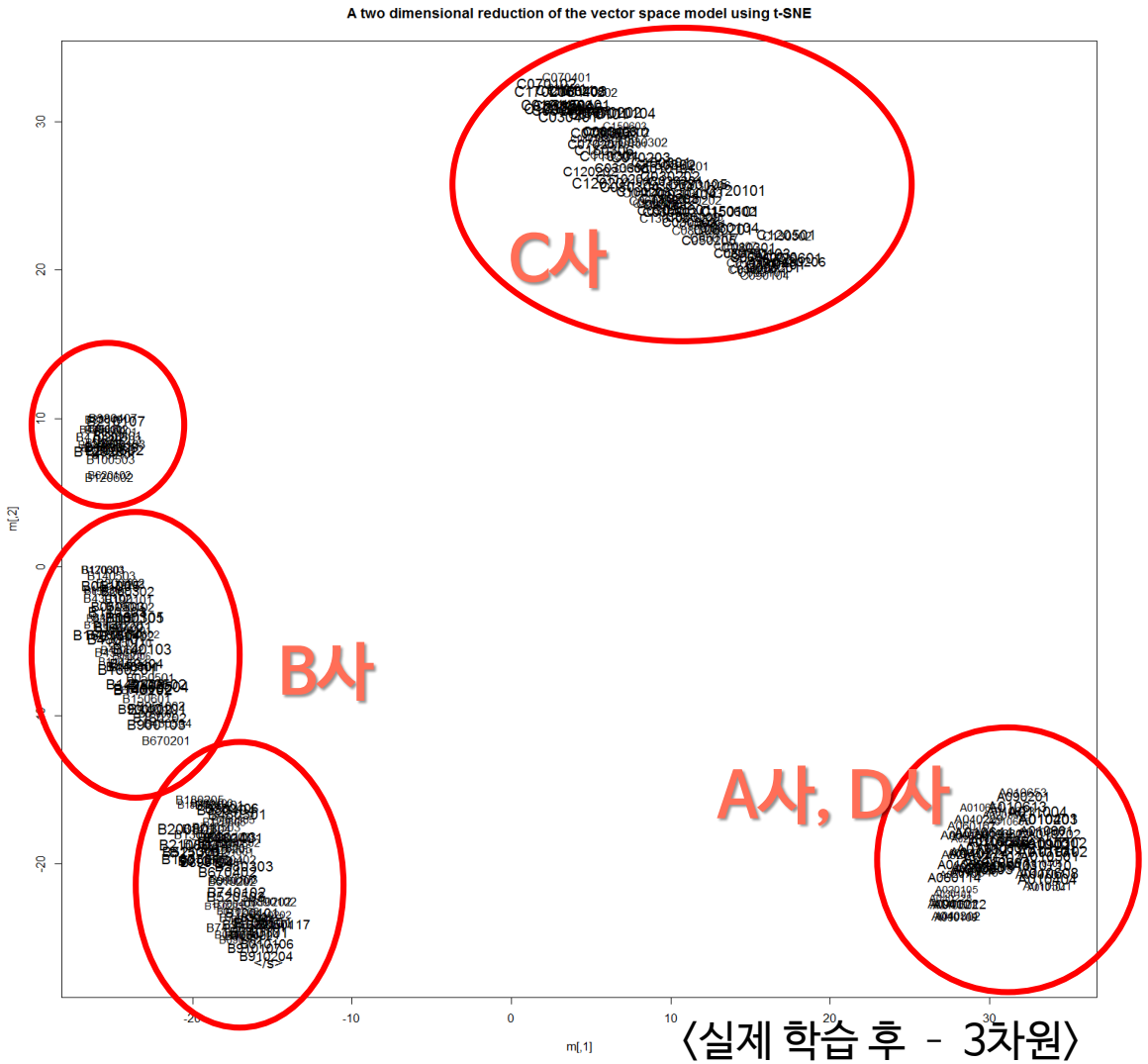


(출처: Computer vision for dummies, 〈The Curse of Dimensionality in classification〉, 2014년4월16일)

하지만 차원이 일정이상 증가하면 데이터의 대부분이 꼭지에 분포하게 되어 정확한 측정 불가

2-2 Word2Vec

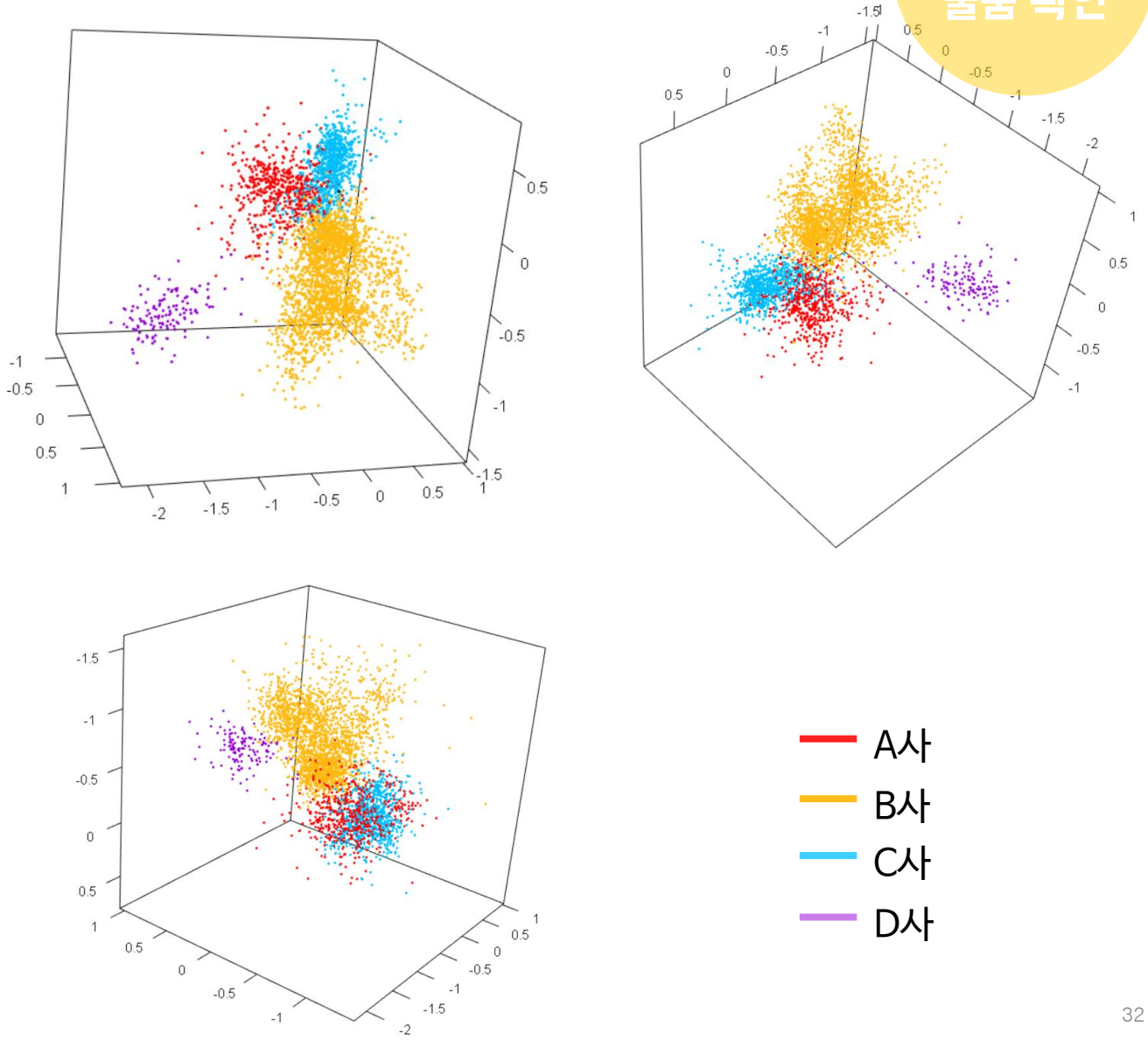
Word2Vec의 적용방법, 과정
1. Word2Vec 학습 : 차원이란?



〈실제 학습 후 - 3차원〉

〈실제 학습 결과 – 3차원〉

공간상 유사한 물품 확인

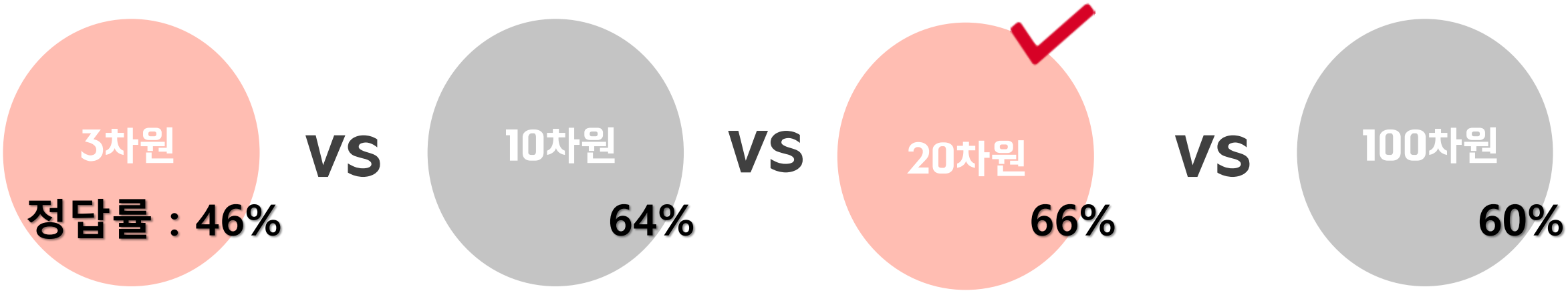


2-2 Word2Vec

Word2Vec의 적용방법, 과정
1. Word2Vec 학습 : 차원에 따른 모델 학습 & 검증

2014년 데이터로 모델 학습 , 2015년 구매 예측
→ 정답률 확인 후 비교

차원에 따른 모델 학습 & 검증



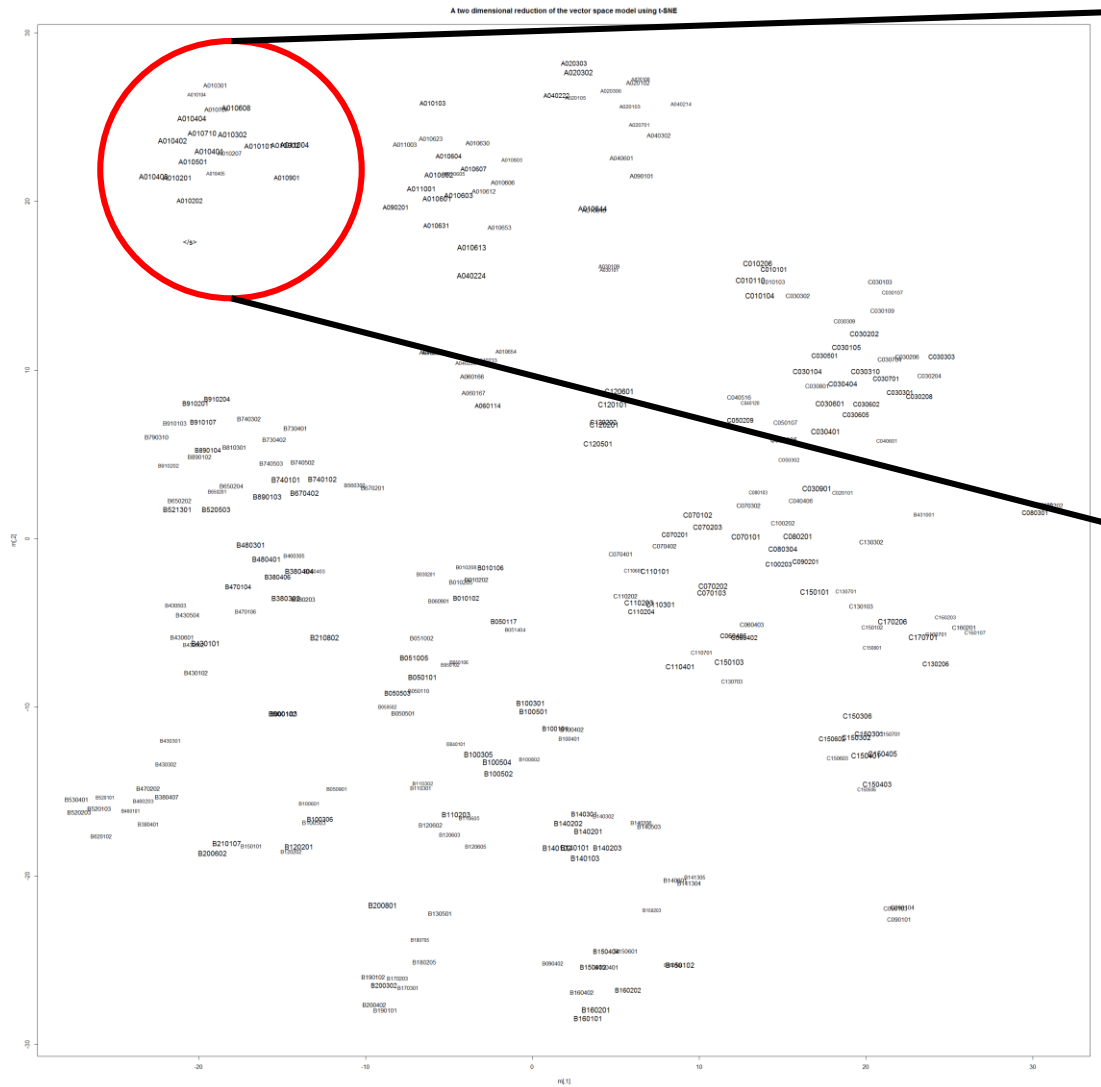
20차원 공간에 Word2Vec 알고리즘 학습시킨 모델 사용

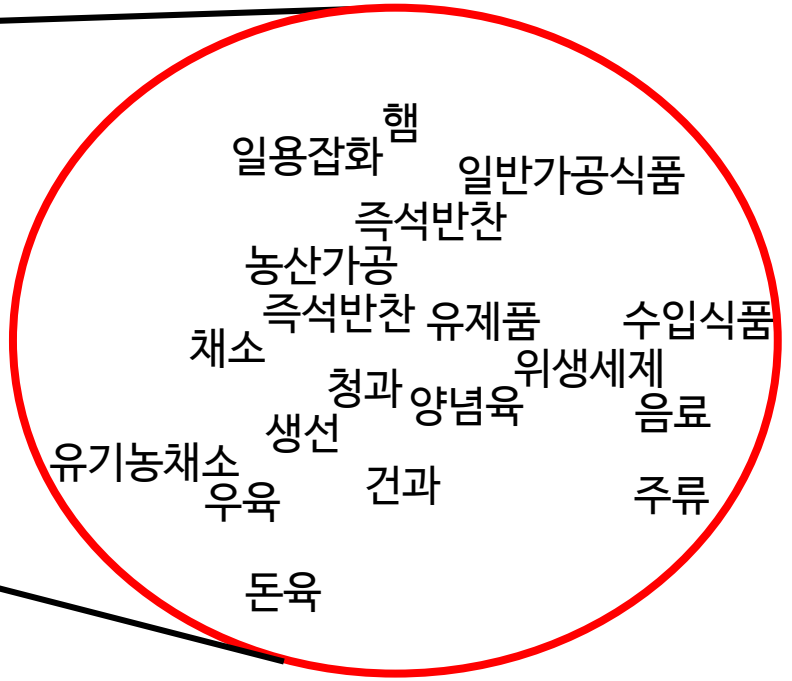
2-2 Word2Vec

Word2Vec의 적용방법, 과정
2. 공간상 유사 물품 확인 :
물품별 공간상 거리를 측정하여 유사한 물품 확인

공간상 유사한 물품 확인

<실제 학습 결과 – 20차원>





- A제휴사 물품
- 대분류가 모두 “01”로 동일
- 대부분 식품군

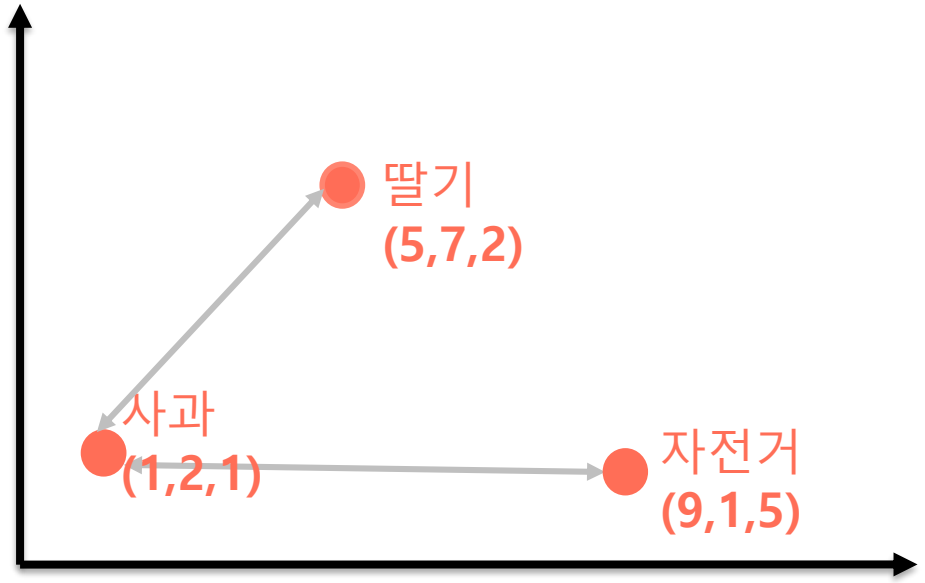
〈실제 학습 후 - 20차원〉

2-2 Word2Vec

Word2Vec의 적용방법, 과정
2. 공간상 유사 물품 확인 :
물품별 공간상 거리를 측정하여 유사한 물품 확인

공간상 유사한 물품 확인

Word2Vec 학습 결과로 물품별 공간상 거리를 측정하여 유사한 물품 확인



<3차원 학습 결과 예시>

사과 - 딸기 거리

= 1−(사과, 딸기 코사인 유사도)

$= 1 - (((1 \times 5) + (2 \times 7) + (1 \times 2))/\sqrt{(1^2 + 2^2 + 1^2) \times (5^2 + 7^2 + 2^2)})$

= 1−0.97 = 0.03

사과 - 자전거 거리

= 1−(사과, 자전거 코사인 유사도)

$= 1 - (((1 \times 9) + (2 \times 1) + (1 \times 5))/\sqrt{(1^2 + 2^2 + 1^2) \times (9^2 + 1^2 + 5^2)})$

= 1−0.63 = 0.37

사과 - 딸기의 유사도 > 사과 - 자전거의 유사도

2-2 Word2Vec

Word2Vec의 적용방법, 과정

3. 고객별 물품 추천 좌표계 :
   고객별 물품 추천 좌표계 생성 → 고객과 물품간 거리 확인

고객별
물품 추천
좌표계 생성

〈3차원 추천 좌표계〉

나의 점수

사과
복숭아
밤

PB돼지고기
미국산 돼지고기
PB수입소고기

콜라
이온음료
혼합탄산

물품 유사도

앞서 확인한 유사 물품 활용

중요성

고객과 거리가 가까운 물품 추천

2-2 Word2Vec

Word2Vec의 적용방법, 과정
3. 고객별 물품 추천 좌표계 : X축

고객별
물품 추천
좌표계 생성

최고점수 → 최저점수 → 나의 점수

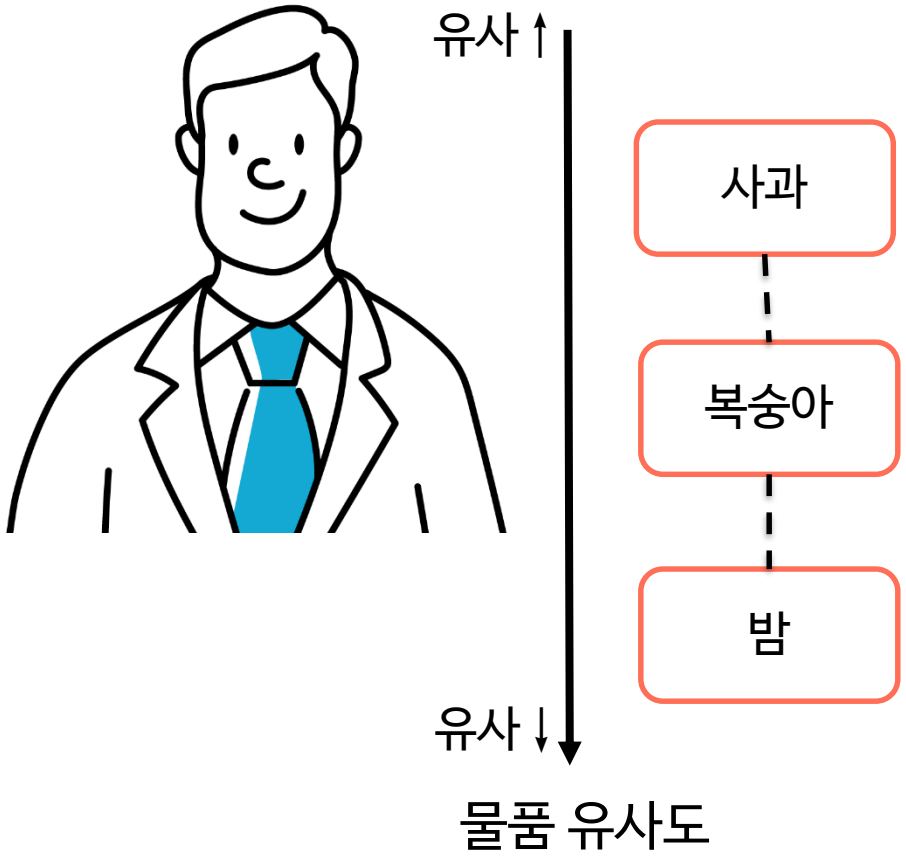
사과 | PB돼지고기 | 콜라

X축 : 고객이 겨울동안 구매한 i개의 물품 중
TF-IDF점수가 높은 상위 $\sqrt{i}$ (i개를 $\sqrt{i}$로 줄여 불필요한 좌표를 줄임)개 나열
(최고점수가 원점에 가깝도록)

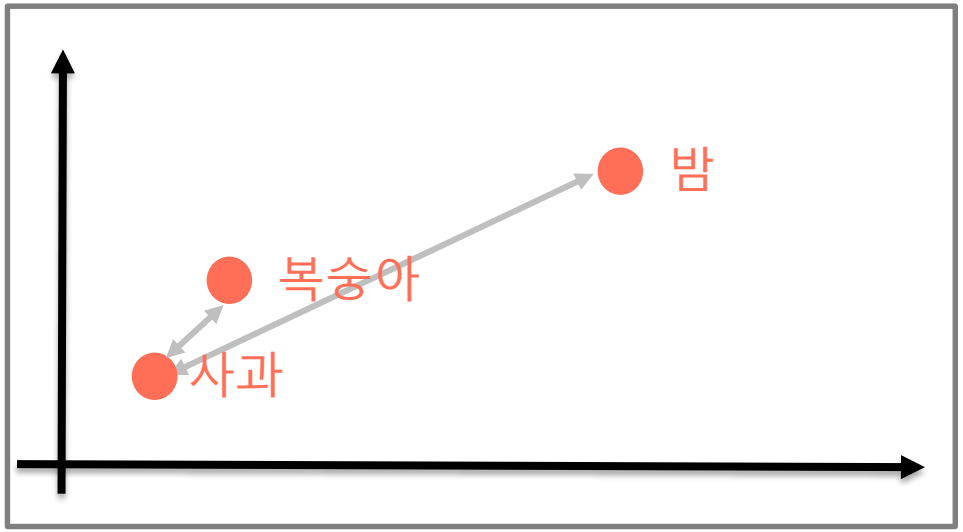
2-2 Word2Vec

Word2Vec의 적용방법, 과정
3. 고객별 물품 추천 좌표계 : Y축

고객별
물품 추천
좌표계 생성



Y축 : Word2Vec 알고리즘에 의해 확인한
물품별 공간상 거리를 활용하여
구매한 물품들과 가까운 순서대로 나열
(구매한 물품과 유사할수록 원점에 가깝도록)



사과, 복숭아의 거리 < 사과, 밤의 거리

2-2 Word2Vec

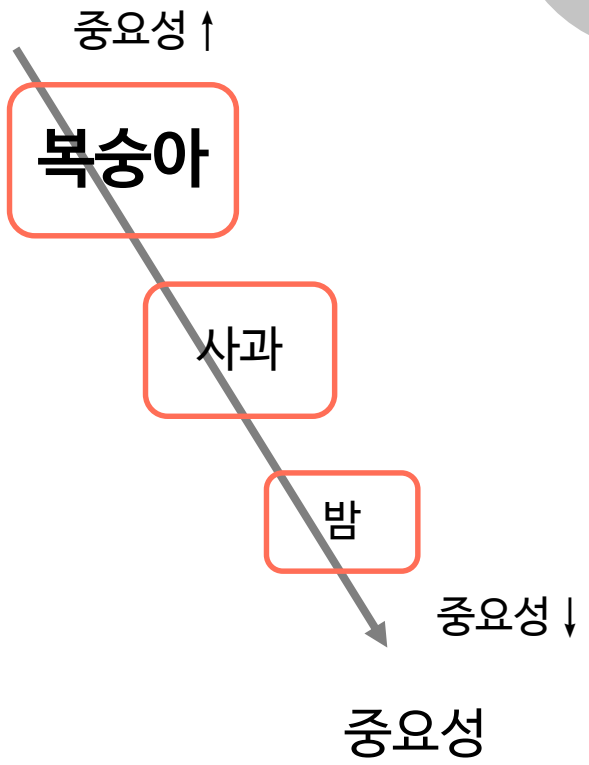
Word2Vec의 적용방법, 과정
3. 고객별 물품 추천 좌표계 : Z축

고객별
물품 추천
좌표계 생성

Z축 : 물품별 TF-IDF 값을 모두 더하여 높은 순서대로 나열
(높은 점수가 원점에 가깝도록)

	1번고객	2번고객	3번고객	4번고객	5번고객	합
사과	1	9	4	5	3	22
복숭아	0	5	10	7	8	**30**
밤	2	0	9	0	6	17

〈물품 점수 행렬 -물품 x 고객〉

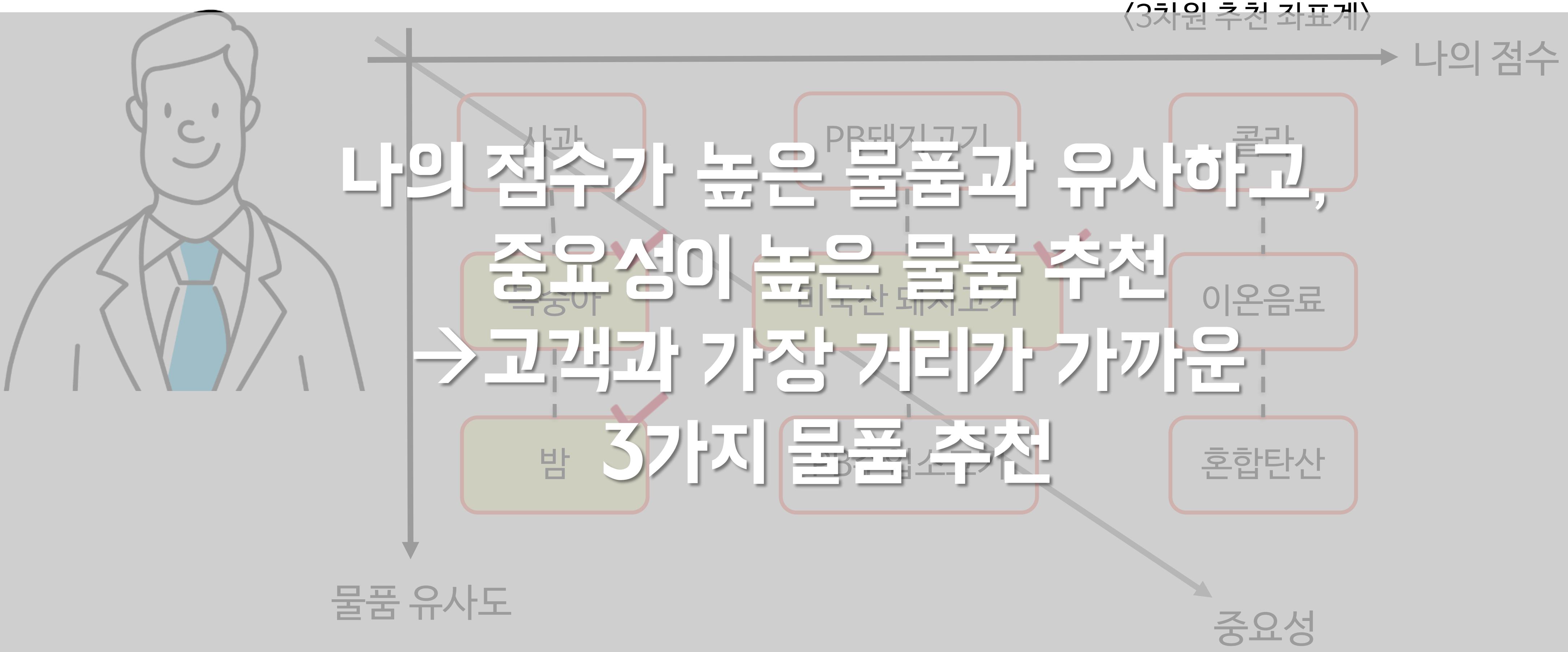


2-2 Word2Vec

Word2Vec의 적용방법, 과정
3. 고객별 물품 추천 좌표계 :
고객별 물품 추천 좌표계 생성 → 고객과 물품간 거리 확인

고객별 물품 추천 좌표계 생성

〈3차원 추천 좌표계〉



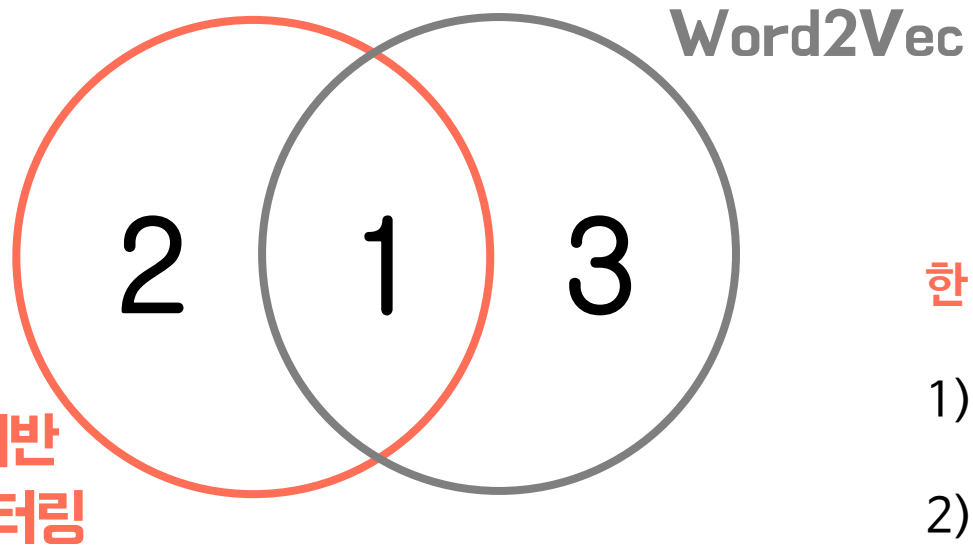
나의 점수가 높은 물품과 유사하고,
중요성이 높은 물품 추천
→고객과 가장 거리가 가까운
3가지 물품 추천

고객과 거리가 가까운 물품 추천

3. 최종 추천 방법

3 최종 추천 방법

물품기반 협업필터링 & Word2Vec 에서 최종 물품 추천



한 고객에게 총 3개의 물품 추천

1) 물품기반 협업필터링과 Word2Vec 모두에게 나온 추천물품

2) 물품기반 협업필터링에서만 나온, 높은 점수의 물품

3) Word2Vec에서만 나온, 고객과 거리가 가까운 물품

- 물품기반 협업필터링과 Word2Vec 에서 공통된 부분이 없다면, 더 높은 정답률을 가졌던 Word2Vec(66%) 에서 두개, 물품기반 협업필터링(51%) 에서 하나 추천

4. R package

4 R package

R 에서 사용한 package

data.table	“구매상품TR” 과 같이 용량이 큰 데이터를 불러오기 위해 사용한 패키지
complier	반복문의 실행 속도를 빠르게 하기 위해 사용한 패키지
stringr	문자열 처리를 위해 사용한 패키지
microbenchmark	function이 다 실행 되는데 시간이 얼마나 걸리는지 알려주는 패키지
ggplot2, rgl	시각화를 위해 사용한 패키지
devtools	git-hub 에서 wordVectors 패키지를 다운받을 수 있도록 도와주는 패키지
wordVectors	Word2Vec 알고리즘에 사용한 패키지

감사합니다