Sentimental Analysis of Product Review in Online to Estimate the Consumer's Opinion:

Focus on iPhone Series of Apple Inc. and Galaxy Series of Samsung Electronics Co.,

Unstructured Data Analysis Final Presentation
2017-06-21

MIS SOORAN KAM(감수란)
IME MINJUNG LEE (이민정)
IME YOONSANG JO (조윤상)

INDEX

01 Introduction

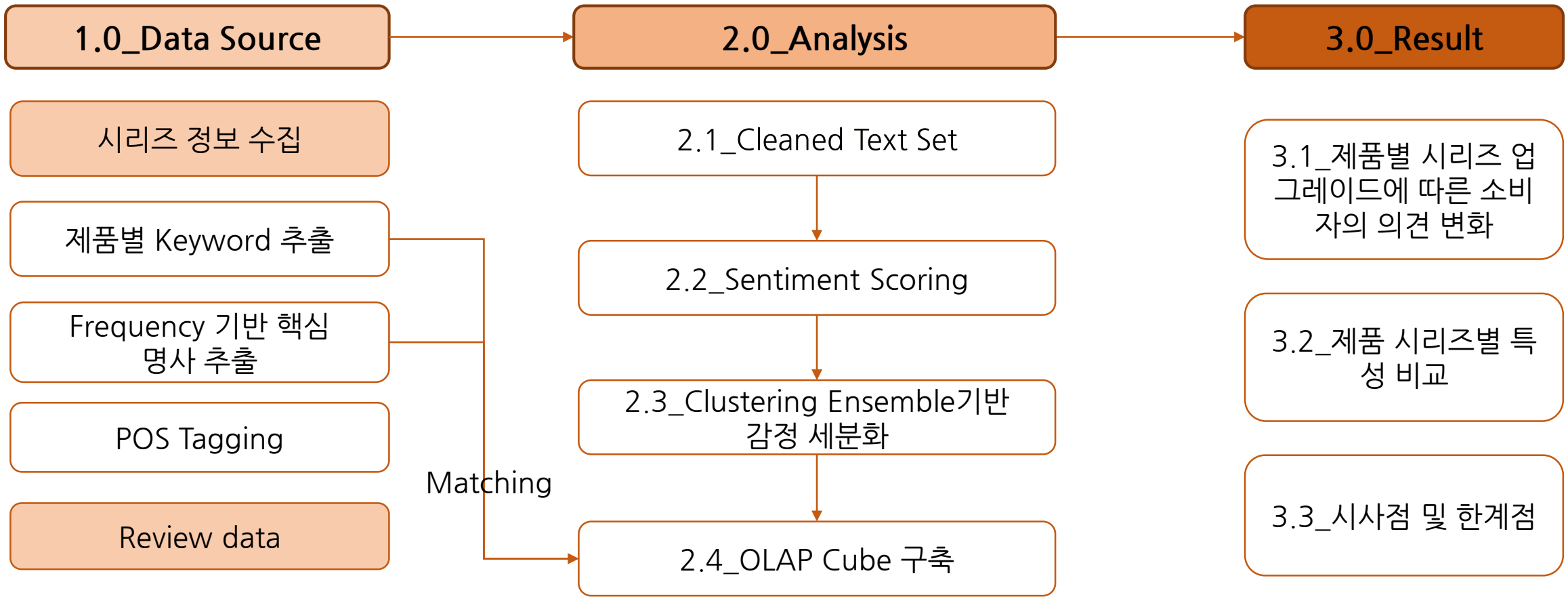
- **연구 목표**

  - ✓ Q1: 대표 전자제품(스마트폰)을 선정하여 온라인에서의 소비자 의견 감성분석을 통해 제품의 시리즈 발전이 소비자의 의견이 어떻게 영향을 미치는가, 이는 업그레이드 기술과 소비자의 의견이 흐름을 같이 하는가

  - ✓ Q2: 유관 산업에서 지속적인 시리즈를 출시하는 회사들이 더욱 빠르게 소비자의 의견을 반영하여 제품의 품질을 높이는 사업방향을 제시할 수 있는가

01 Introduction

- **Project Framework**



02 Data & Preprocessing

KOREA UNIVERSITY

- **Data Collection**

  - ✓ 초반에 트위터 데이터 추출: 최근 10일 데이터만 수집가능한 한계
  - ✓ 온라인 리뷰 전문 사이트 검색하여 휴대폰 시리즈별 리뷰 Web Scraping 수행 (Data Source: GSMArena)
  - ✓ iPhone 시리즈: 매년 9월 출시 / Galaxy S 시리즈: 매년 3월 출시
  - ✓ N년 9월 출시 iPhone VS N+1년 3월 출시 Galaxy Data 선정 및 비교
  - ✓ Edge, S+, + 시리즈 제외

갤럭시 시리즈		아이폰 시리즈	
갤럭시 S6 : (2015.03)	8,920 개	아이폰 6 : (2014.09)	9,568 개
갤럭시 S7 : (2016.03)	4,380 개	아이폰 6S : (2015.09)	2,497 개
갤럭시 S8 : (2017.03)	2,887 개	아이폰 7 : (2016.09)	3,177 개

02 Data & Preprocessing

KOREA UNIVERSITY

- **리뷰 기준 추출 단어 (공식 웹사이트 시리즈 제품별 주요 업그레이드 부분)**

✓ 삼성 갤럭시 S시리즈

	디자인	하드웨어	기타 기능 및 요소
갤럭시 S6	1. 엣지 디자인 2. 그립감	1. 무선충전 2. 퀵충전 3. LTE 속도 4. 고성능 CPU	1. 카메라 밝은 렌즈 2. 피사체 추적 3. OIS 4. 삼성페이 5. 음향 / 사운드
갤럭시 S7	1. 실용적 디자인 2. 곡선 3. 그립감	1. 방수 2. 속도 향상 3. 메모리 확장 4. 베터리 수명	1. 카메라 렌즈 초점 2. 모션파노라마 3. 후면 카메라 4. 삼성페이 보안 5. 지문인식
갤럭시 S8	1. 베젤리스디자인 2. 곡선 디자인 3. 컬러	1. 전력소모량 (CPU) 2. 빠른 속도 3. 확장메모리 4. 고속충전	1. 사운드 향상 2. 인공지능 3. 접근성 4. 액세서리 5. 홍채 인식 6. 전면 오토포커스 카메라

02 Data & Preprocessing

KOREA UNIVERSITY

- **리뷰 기준 추출 단어 (공식 웹사이트 시리즈 제품별 주요 업그레이드 부분)**

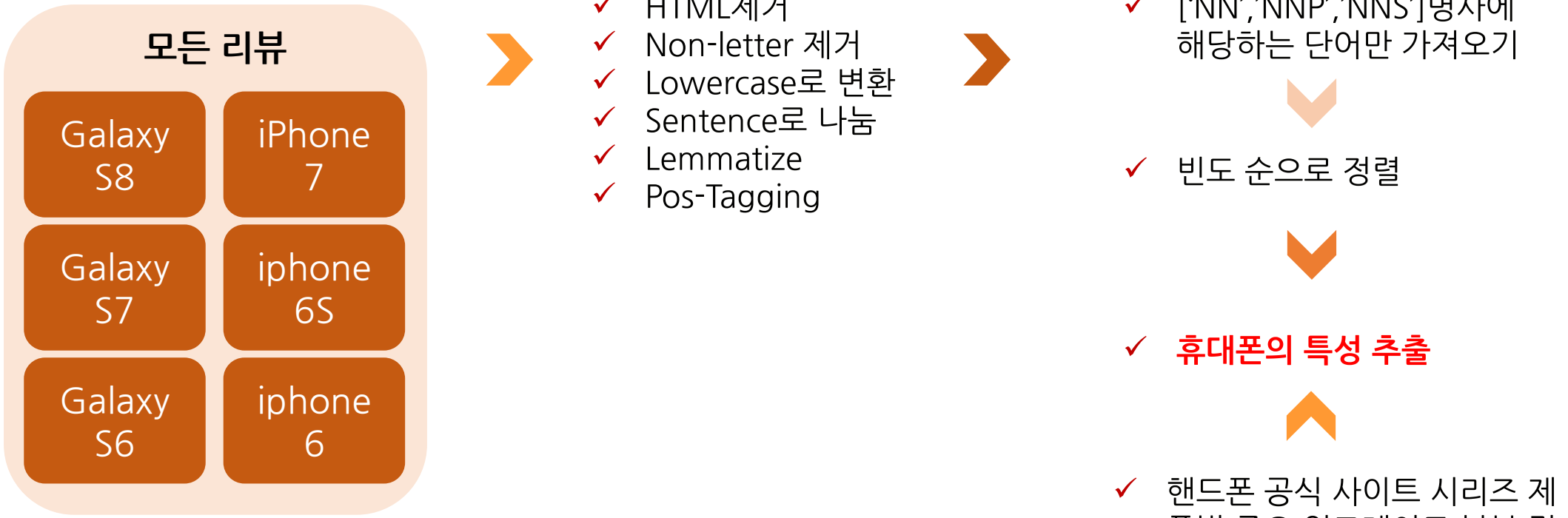
✓ 애플 아이폰 시리즈

	디자인	하드웨어	기타 기능
아이폰 6	1. 옆면의 전원버튼 2. 얇아진 두께 3. 카메라 돌출 4. 커진 화면	1. 레티나 HD 디스플레이 2. 빨라진 A8 칩 3. 배터리 수명	1. 카메라 자동초점 2. 애플페이 3. 접근성 4. 네트워크 업그레이드
아이폰 6S	1. 컬러 2. 사이즈	1. Touch ID 지문인식 2. 인터넷 / 전원/ 배터리 3. 칩 (AM9) 4. 디스플레이 해상도 5. IOS 소프트웨어 6. 영상통화	1. 카메라 화소 2. 얼굴 인식 3. 시리 4. iTunes 5. 접근성
아이폰 7	1. 새로운 컬러 2. Unibody 디자인 3. 촉감	1. 새로운 홈버튼 Touch ID 2. 배터리 수명 3. IOS 10 4. 칩(AM10) / CPU 속도 5. 3D Touch 6. 생활방수	1. 스테레오 스피커 2. Wi-Fi 및 이동통신 3. 선 없는 air pods 4. 카메라 화소 5. 전면카메라

02 Data & Preprocessing

KOREA UNIVERSITY

- **Preprocessing**



02 Data & Preprocessing

KOREA UNIVERSITY

- 휴대폰 특성 단어 추출

Battery
Camera
Screen
Ram
Device
Display
Design
Price
Quality
Memory
Size

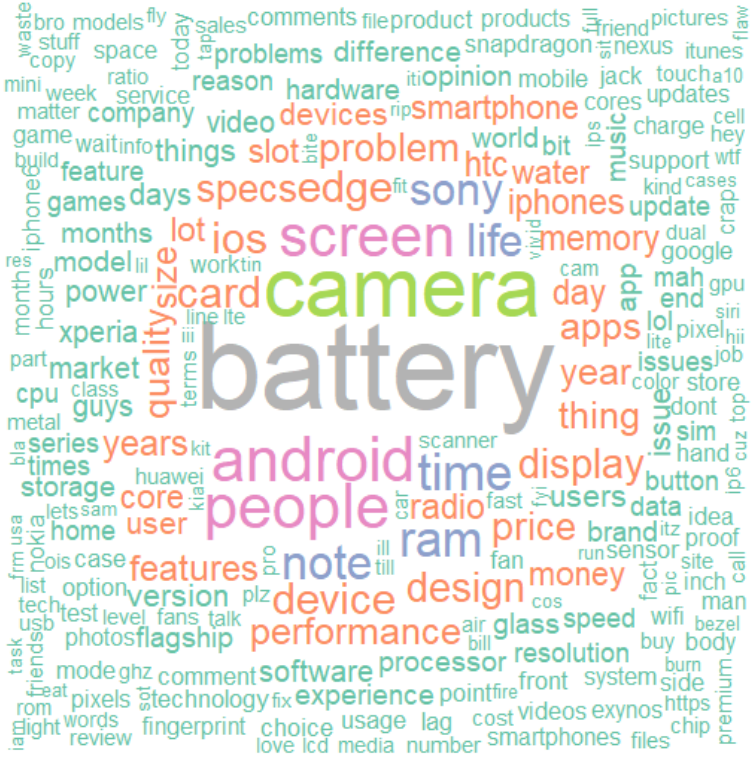
Water
Glass
Hardware
Speed
Sim
Sensor
Update
Cpu
Fan
Proof

Inch
Pixel
Fingerprint
Color
Internet
Wireless
Warranty
Sound
Security
Photo

Microsd
Chipset
Waterproof
Cam
Finger
Flash
Headphone
Megapixel
Brightness
Heating
bluetooth

02 Data & Preprocessing

- Word2Vec 사용 특성 단어 Similarity 보기



Battery	Similarity
Battry	0.8358
Baterry	0.8290
Bettery	0.8160
Batery	0.8107
Batt	0.8080
Span	0.8057
Batter	0.8025
Anker	0.7856
Decrease	0.7726
Expectancy	0.7685
Batttery	0.7480
Batterylife	0.7461

Camera	Similarity
Lense	0.8448
Photography	0.8405
Canon	0.8296
Megapixel	0.8260
Cemra	0.8161
Britecell	0.8145
Cam	0.8127
Shooter	0.8122
Capture	0.8079
Focus	0.8049
Lowlight	0.8030
movement	0.7996

03 Text Analysis Method

KOREA UNIVERSITY

- **Sentimental Analysis**
  - 감정스코어를 도출하는 알고리즘 5가지를 사용 및 정규화 (-1 < <+1)후 Clustering Ensemble 수행

Text Cleaning → **Sentiment Scoring** → **Clustering Ensemble** → **OLAP Cube Analysis**

- ‘bn’, @, 등 웹 스크래핑 후 필요없는 문자 제거
- Cleaned dataset

	Date	Text			
Doc1					

- 알고리즘 1->score1
- 알고리즘 2->score2
- 알고리즘 3->score3
- 알고리즘 4->score4
- 알고리즘 5->score5

-알고리즘 참조
*Bing Liu, Minqing Hu and Junsheng Cheng. "Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of the 14th International World Wide Web conference (WWW-2005), May 10-14, 2005, Chiba, Japan.
**Minqing Hu and Bing Liu. "Mining and Summarizing Customer Reviews." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004), Aug 22-25, 2004, Seattle, Washington, USA. See: http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#lexicon
***Saif Mohammad and Peter Turney. "Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon." In Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, June 2010, LA, California. See: http://saifmohammad.com/WebPages/lexicons.html
****Finn Arup Nielsen. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs", Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts':Big things come in small packages 718 in CEUR Workshop Proceedings : 93-98. 2011 May. http://arxiv.org/abs/1103.2903. See: http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010

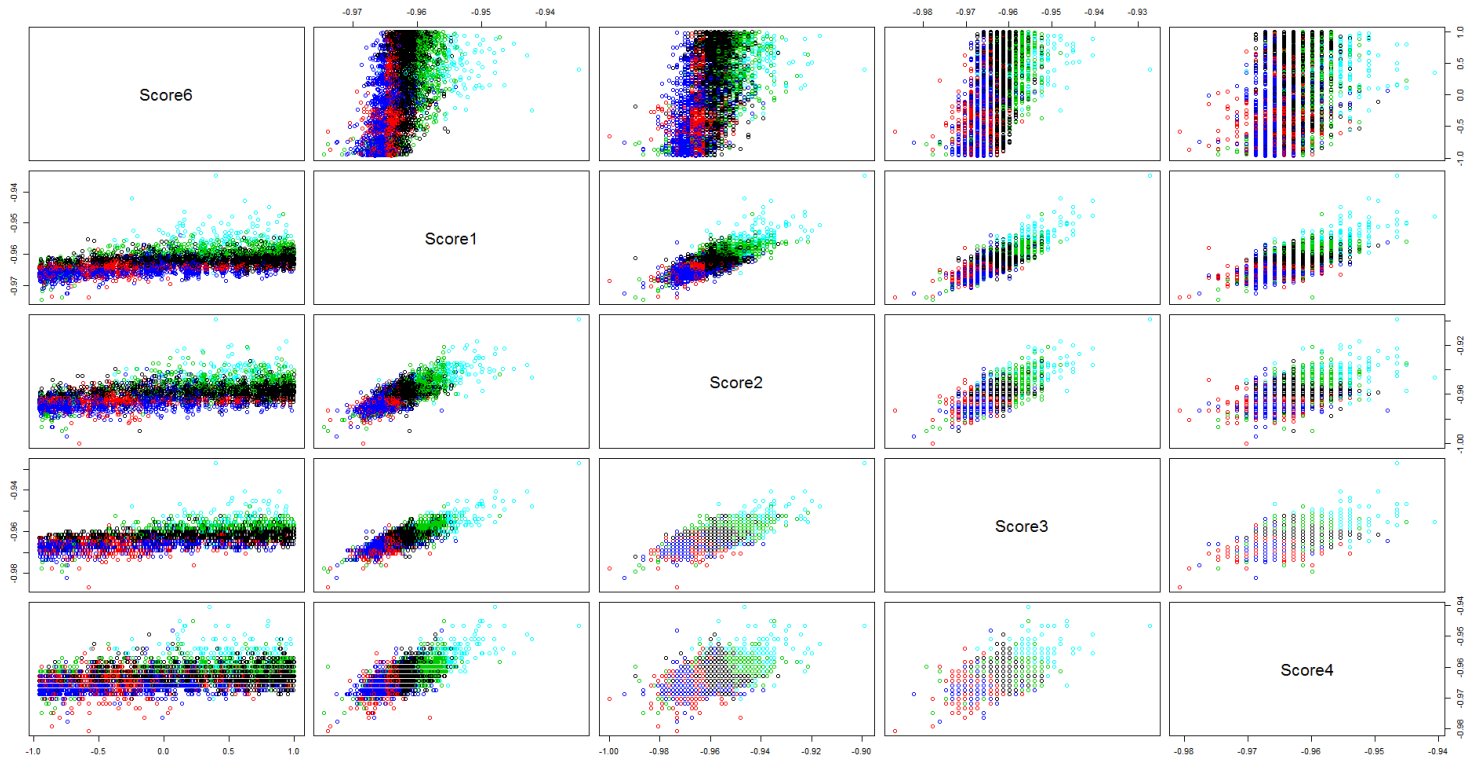
- 알고리즘 별 K-means군집화 및 ensemble(method = ‘SE’)
- K=5, 5개의 군집 도출
- 데이터 통합
- OLAP Cube 구축

- Key별 분석 진행 Key:Date, Cluster, Model, keywords

	Date	Text			
Doc1					

03 Text Analysis Method

- Clustering Ensemble 수행



Cluster	Doc 개수
1	20,596 (65%)
2	3,302 (10%)
3	4,291 (13%)
4	2,483 (8%)
5	757 (2%)

03 Text Analysis Method

KOREA UNIVERSITY

- 분석 데이터 20 example

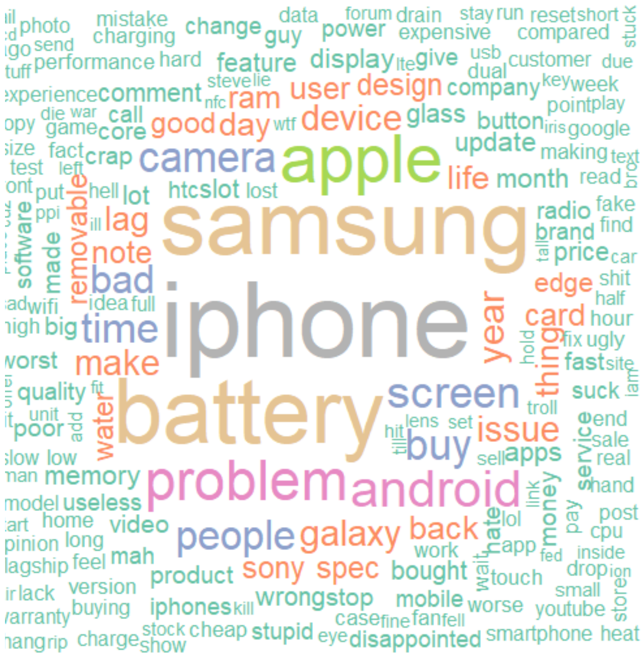
No.	Model	ID	Date	score1_ scaled	score2_ scaled	score3_ scaled	score4_ scaled	score5_ scaled	Cluster
1	i6	1	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1
2	i6	2	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1
3	i6	3	2014-05-22	-0.5	-0.37	-0.15	-0.33	-0.78	1
4	i6	4	2014-05-22	-0.46	-0.32	-0.25	-0.2	-0.99	1
5	i6	5	2014-05-22	-0.5	-0.37	-0.15	-0.33	-0.78	1
6	i6	6	2014-05-22	-0.54	-0.48	-0.35	-0.13	-1	2
7	i6	7	2014-05-22	-0.46	-0.21	-0.2	-0.2	-0.99	1
8	i6	8	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1
9	i6	9	2014-05-22	-0.57	-0.4	-0.3	-0.33	-1	2
10	i6	10	2014-05-22	-0.54	-0.32	-0.25	-0.33	-0.99	1
11	i6	11	2014-05-22	-0.52	-0.34	-0.25	-0.2	-0.99	1
12	i6	12	2014-05-22	-0.58	-0.42	-0.4	-0.33	-1	2
13	i6	13	2014-05-22	-0.46	-0.34	-0.2	-0.33	-0.99	1
14	i6	14	2014-05-22	-0.57	-0.34	-0.25	-0.27	-0.99	1
15	i6	15	2014-05-22	-0.51	-0.34	-0.25	-0.27	-0.99	1
16	i6	16	2014-05-22	-0.52	-0.34	-0.25	-0.2	-0.99	1
17	i6	17	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1
18	i6	18	2014-05-22	-0.95	-0.64	-0.7	-0.93	-1	2
19	i6	19	2014-05-22	-0.51	-0.34	-0.25	-0.2	-0.88	1
20	i6	20	2014-05-22	-0.47	-0.26	-0.2	-0.13	-0.99	3

04 Discussion

- 분석결과1
- Cluster 별 WordCloud

Cluster1

Cluster2

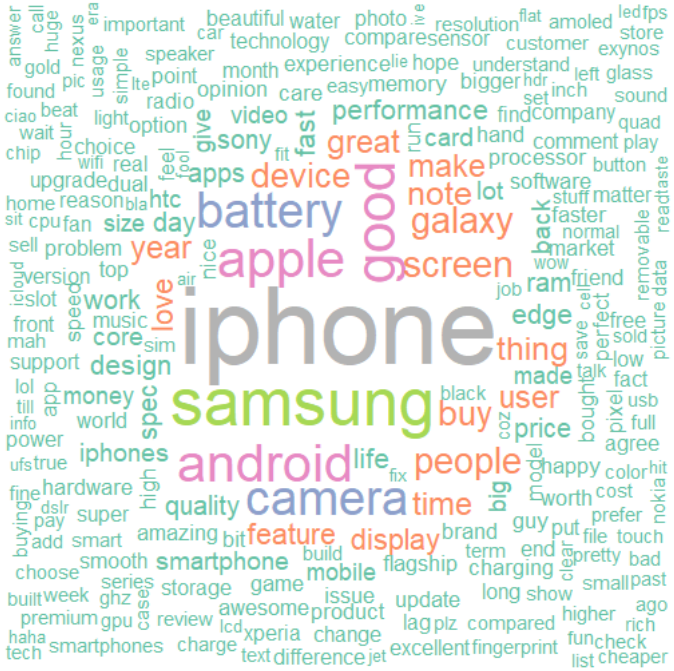


Cluster 3

04 Discussion

- 분석결과1
- Cluster 별 WordCloud

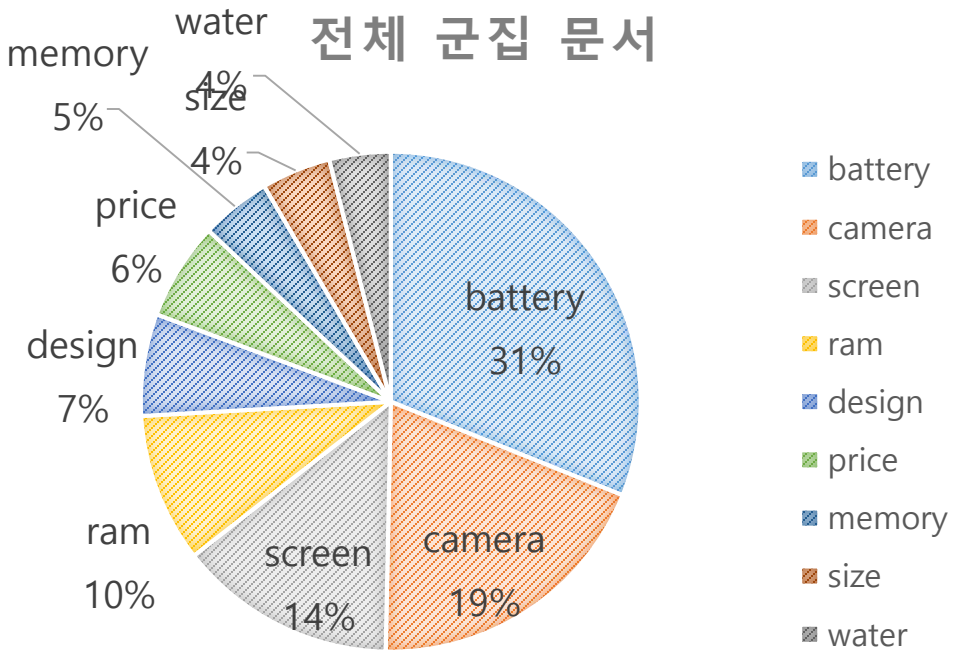
Cluster 4

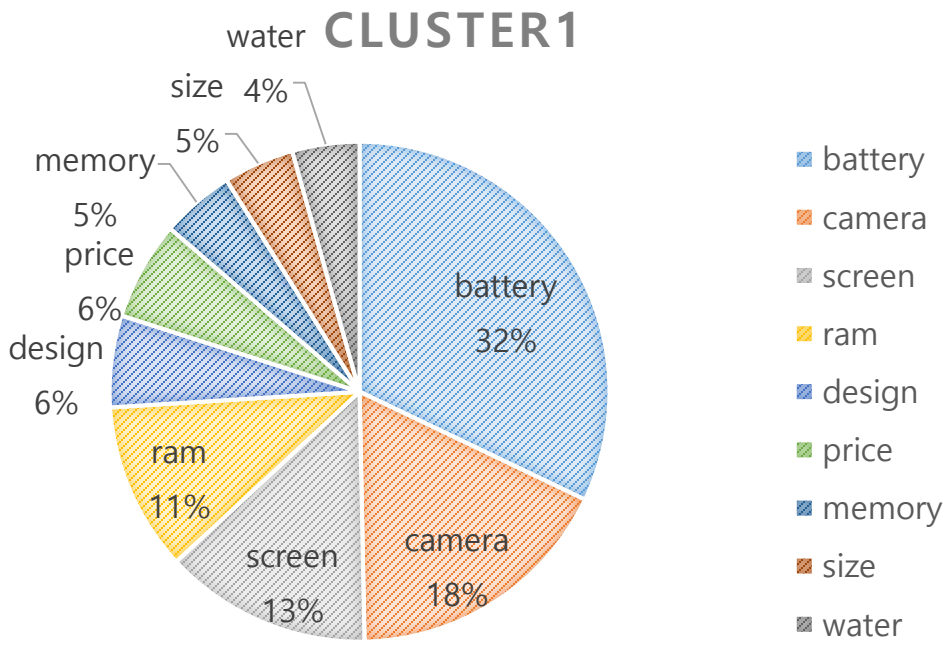


Cluster 5

04 Discussion

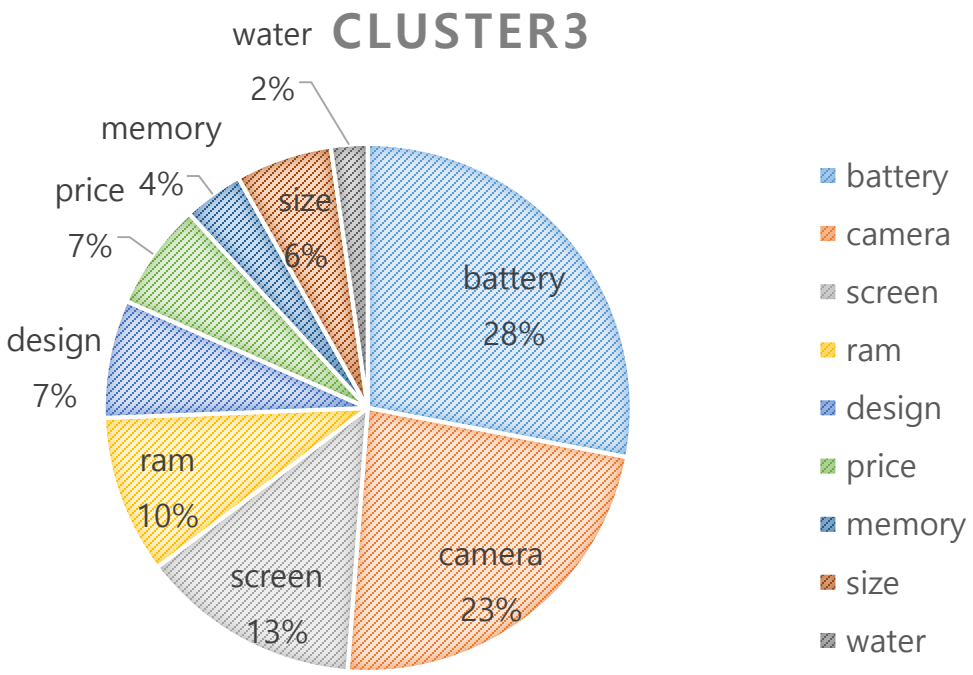
- **분석결과2 -** 5개 군집별 감성분석 단어 그래프

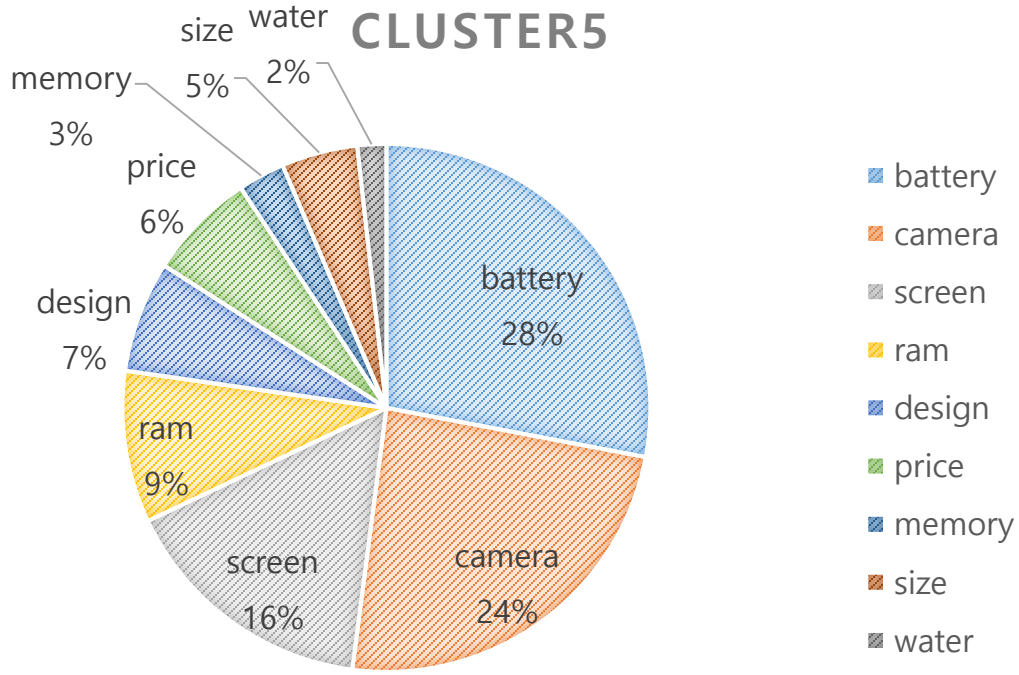




04 Discussion

- **분석결과2 -** 5개 군집별 감성분석 단어 그래프





- **전체문서에서 보이는 휴대폰 특성의 비율의 패턴이 모든 군집에서 유사하게 나타남.**

04 Discussion

KOREA UNIVERSITY

- 분석결과3
- Clustering간 시리즈별 비율

	Galaxy S6	Galaxy S7	Galaxy S8
Cluster1	5,787(65%)	2,919(67%)	1,927(67%)
Cluster2	1,077(12%)	450(10%)	300(10%)
Cluster3	1,137(13%)	640(15%)	375(13%)
Cluster4	718(8%)	294(7%)	206(7%)
Cluster5	201(2%)	77(2%)	79(3%)
합	8,920	4,380	2,887

	iPhone 6	iPhone 6S	iPhone 7
Cluster1	6,204(65%)	1,602(64%)	2,157(68%)
Cluster2	951(10%)	212(8%)	312(10%)
Cluster3	1,323(14%)	407(16%)	409(13%)
Cluster4	833(9%)	196(8%)	236(7%)
Cluster5	257(3%)	809(3)	63(2%)
합	9,568	2,497	3,177

04 Discussion

- 분석결과4

battery	I6	I6S	I7	S6	S7	S8
1	602	268	287	1236	410	368
2	155	59	75	390	124	108
3	150	79	52	230	105	81
4	145	40	52	247	81	54
5	64	20	19	73	23	30
합	1,116	466	485	2,176	743	641
	0.1166	0.1866	0.1526	0.2439	0.1696	0.2220

총문서	I6	I6S	I7	S6	S7	S8
1	6,204	1,602	2,157	5,787	2,919	1,927
2	951	212	312	1,077	450	300
3	1,323	407	409	1,137	640	375
4	833	196	236	718	294	206
5	257	80	63	201	77	79
합	9,568	2,497	3,177	8,920	4,380	2,887

04 Discussion

KOREA UNIVERSITY

- **분석결과5**

  - 특정단어 'battery', 'water', 'camera' : 비율로 나타낸 뒤 모델의 비율의 합이 1이 되도록 조정

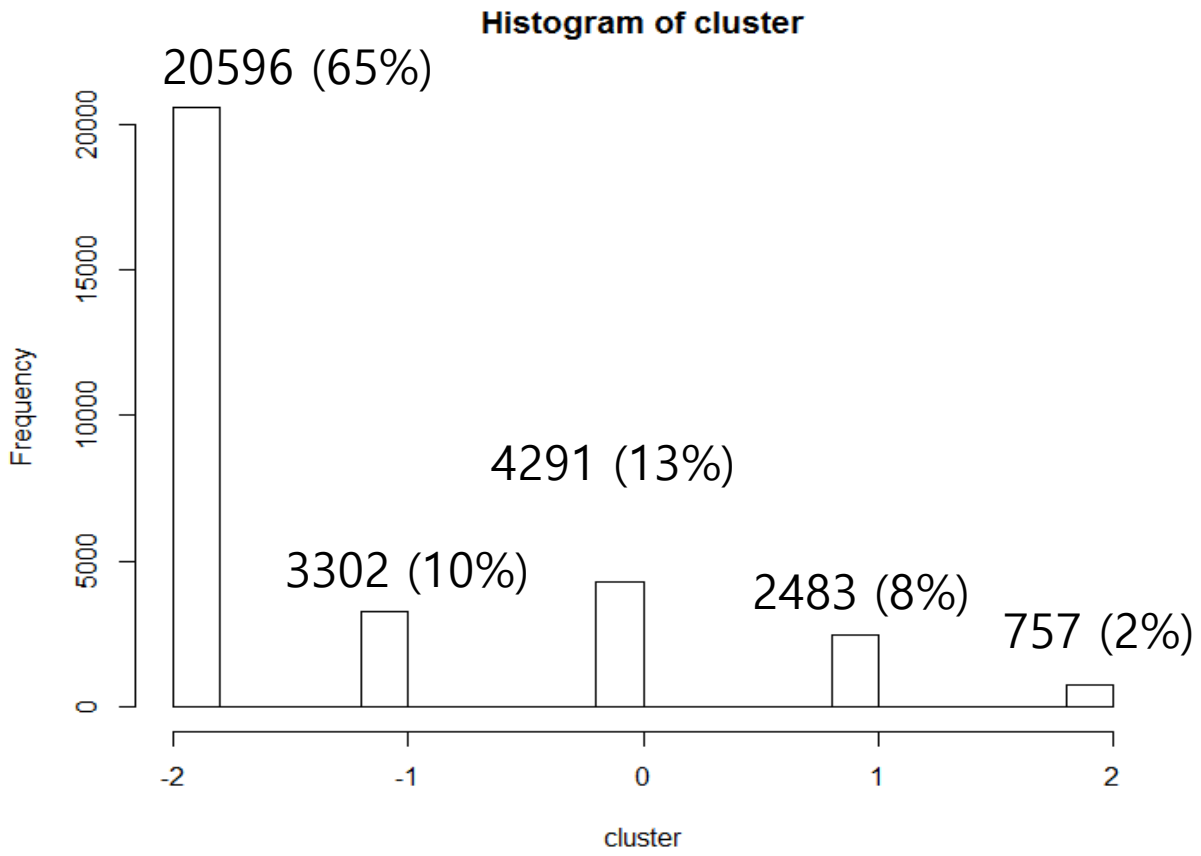
battery	I6	I6S	I7	S6	S7	S8
1	0.0970	0.1673	0.1331	0.2136	0.1405	0.1910
2	0.1630	0.2783	0.2404	0.3621	0.2756	0.36
3	0.1134	0.1941	0.1271	0.2023	0.1641	0.216
4	0.1741	0.2041	0.2203	0.3440	0.2755	0.2621
5	0.2490	0.25	0.3016	0.3632	0.2987	0.3797

battery	I6	I6S	I7	S6	S7	S8
1	12%	15%	13%	14%	12%	14%
2	20%	25%	24%	24%	24%	26%
3	14%	18%	12%	14%	14%	25%
4	22%	19%	22%	23%	24%	19%
5	31%	23%	29%	24%	26%	27%

water	I6	I6S	I7	S6	S7	S8
1	15%	12%	13%	16%	10%	3%
2	29%	39%	23%	30%	26%	28%
3	10%	8%	12%	12%	14%	12%
4	20%	27%	37%	23%	22%	28%
5	26%	14%	15%	19%	28%	29%

camera	I6	I6S	I7	S6	S7	S8
1	12%	10%	10%	9%	10%	10%
2	14%	10%	14%	10%	12%	15%
3	15%	14%	15%	17%	17%	13%
4	31%	29%	27%	30%	29%	24%
5	28%	38%	35%	34%	32%	38%

04 Discussion



- Result : Clustering Ensemble
  - ✓ Histogram : Right skewed

04 Discussion

KOREA UNIVERSITY

- 분석 데이터 20 example (추가 변경)

No.	Model	ID	Date	score1_ scaled	score2_ scaled	score3_ scaled	score4_ scaled	score5_ scaled	Cluster	avg_score
1	i6	1	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1	-0.474
2	i6	2	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1	-0.474
3	i6	3	2014-05-22	-0.5	-0.37	-0.15	-0.33	-0.78	1	-0.426
4	i6	4	2014-05-22	-0.46	-0.32	-0.25	-0.2	-0.99	1	-0.444
5	i6	5	2014-05-22	-0.5	-0.37	-0.15	-0.33	-0.78	1	-0.426
6	i6	6	2014-05-22	-0.54	-0.48	-0.35	-0.13	-1	2	-0.5
7	i6	7	2014-05-22	-0.46	-0.21	-0.2	-0.2	-0.99	1	-0.412
8	i6	8	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1	-0.474
9	i6	9	2014-05-22	-0.57	-0.4	-0.3	-0.33	-1	2	-0.52
10	i6	10	2014-05-22	-0.54	-0.32	-0.25	-0.33	-0.99	1	-0.486
11	i6	11	2014-05-22	-0.52	-0.34	-0.25	-0.2	-0.99	1	-0.46
12	i6	12	2014-05-22	-0.58	-0.42	-0.4	-0.33	-1	2	-0.546
13	i6	13	2014-05-22	-0.46	-0.34	-0.2	-0.33	-0.99	1	-0.464
14	i6	14	2014-05-22	-0.57	-0.34	-0.25	-0.27	-0.99	1	-0.484
15	i6	15	2014-05-22	-0.51	-0.34	-0.25	-0.27	-0.99	1	-0.472
16	i6	16	2014-05-22	-0.52	-0.34	-0.25	-0.2	-0.99	1	-0.46
17	i6	17	2014-05-22	-0.52	-0.34	-0.25	-0.27	-0.99	1	-0.474
18	i6	18	2014-05-22	-0.95	-0.64	-0.7	-0.93	-1	2	-0.844
19	i6	19	2014-05-22	-0.51	-0.34	-0.25	-0.2	-0.88	1	-0.436
20	i6	20	2014-05-22	-0.47	-0.26	-0.2	-0.13	-0.99	3	-0.41

04 Discussion

- 분석결과6
- 모델별 기능키워드별 감정 스코어

	디자인	하드웨어	기타 기능 및 요소
갤럭시 S6	1. 엣지 디자인 2. 그립감	1. 무선충전 2. 퀵충전 3. **LTE 속도** 4. 고성능 CPU	1. 카메라 밝은 렌즈 2. 피사체 추적 3. OIS 4. 삼성페이 5. **음향 / 사운드**
갤럭시 S7	1. 실용적 디자인 2. 곡선 3. 그립감	1. 방수 2. **속도 향상** 3. 메모리 확장 4. 베터리 수명	1. 카메라 렌즈 초점 2. 모션파노라마 3. 후면 카메라 4. 삼성페이 보안 5. 지문인식
갤럭시 S8	1. 베젤리스디자인 2. 곡선 디자인 3. **컬러/컬러**	1. 전력소모량 (CPU) 2. **빠른 속도** 3. 확장메모리 4. 고속충전	1. **사운드 향상** 2. 인공지능 3. 접근성 4. 액세서리 5. 홍채 인식 6. 전면 오토포커스 카메라

keyword	s6	s7	s8
battery	0.4037	0.4112	0.4231
camera	0.3995	0.4094	0.5032
screen	0.4026	0.4118	0.3922
ram	0.4051	0.4200	0.345
device	0.4131	0.4194	0.4998
display	0.4033	0.4115	0.4804
design	0.4166	0.4025	0.3702
price	0.4009	0.4132	0.4287
quality	0.4142	0.4083	0.5669
memory	0.4080	0.4007	0.4431
size	0.4120	0.4148	0.4954
water	0.3985	0.4093	0.1382
glass	0.4122	0.4007	0.5024
hardware	0.3912	0.4070	0.325
speed	**0.3947**	**0.4338**	**0.4132**
sim	0.3952	0.3948	0.5668
⋮			
fingerprint	0.3897	0.3950	0.5441
color	**0.4084**	**0.4375**	**0.1615**
internet	0.4231	0.4119	0.3752
wireless	0.3765	0.4064	0.6634
warranty	0.3976	0.4036	0.597
sound	0.4011	0.4044	**0.4657**
security	0.4050	0.4253	0.1282
photo	0.4185	0.4211	0.3403
microsd	0.4012	0.4053	0.5078
chipset	0.3935	0.4206	0.1906
waterproof	0.4036	0.4078	0
cam	0.4114	0.4242	0.6279
finger	0.3811	0.4151	0.5234
flash	0.3929	0.4495	0.5833
headphone	0.3731	0.4328	0.3965
colour	0.4033	0.4000	0.5082
megapixel	**0.3362**	**0.4177**	**0.7053**
brightness	0.4024	0.4312	0.1807
heating	0.3965	0.4048	1
bluetooth	0.4522	0.4186	0.4916

04 Discussion

- **분석결과6**
- **모델별 기능 키워드별 감정 스코어**

	디자인	하드웨어	기타 기능
아이폰 6	1. 옆면의 전원버튼 2. 얇아진 두께 3. 카메라 돌출 4. 커진 화면	1. 레티나 HD 디스플레이 **2. 빨라진 A8 칩** **3. 배터리 수명**	1. 카메라 자동초점 2. 애플페이 3. 접근성 4. 네트워크 업그레이드
아이폰 6S	**1. 컬러** **2. 사이즈**	1. Touch ID 지문인식 2. 인터넷 / 전원/ 배터리 3. 칩 (AM9) 4. 디스플레이 해상도 5. IOS 소프트웨어 6. 영상통화	1. 카메라 화소 2. 얼굴 인식 3. 시리 4. iTunes 5. 접근성
아이폰 7	**1. 새로운 컬러** 2. Unibody 디자인 3. 촉감	1. 새로운 홈버튼 Touch ID **2. 배터리 수명** 3. IOS 10 4. 칩(AM10) / CPU 속도 5. 3D Touch **6. 생활방수**	1. 스테레오 스피커 2. Wi-Fi 및 이동통신 **3. wireless air pods** **4. 카메라 화소** 5. 전면카메라

keyword	i6	i6s	i7
battery	**0.3675**	**0.3340**	**0.5094**
camera	0.3849	0.3631	0.5149
screen	0.3376	0.3909	0.4004
ram	0.3317	0.3799	0.6334
device	0.3805	0.4169	0.6651
display	0.3018	0.3401	0.3878
design	0.4069	0.3509	0.6094
price	0.3504	0.4215	0.4886
quality	0.3580	0.3754	0.4162
memory	0.4029	0.2727	0.9878
size	0.4282	0.3656	0.6031
water	0.5597	0.3829	0.3703
glass	0.1441	0.3311	0.5497
hardware	0.4048	0.3653	1.0000
speed	0.4030	0.3280	0.6238
sim	0.3091	0.3997	0.2104
sensor	0.3386	0.4386	0.1285
update	0.3213	0.3504	0.3444
cpu	0.4935	0.3041	0.2755
fan	0.3430	0.4660	0.3248
proof	0.3461	0.3578	0.2730
inch	0.4198	0.2849	0.5435
pixel	0.3552	0.4904	0.0000
fingerprint	0.3732	0.3016	0.1903
color	0.2774	0.2927	0.6875
internet	0.3830	0.4035	0.9069
wireless	**0.3494**	**0.4177**	**0.6688**
warranty	0.1479	0.3508	0.8024
sound	0.2798	0.5032	0.7476
security	0.4239	0.1837	0.2938
photo	0.3018	0.3260	0.1296
microsd	0.3215	0.7569	0.7825
chipset	0.3897	1.0000	0.4345
waterproof	**0.4290**	**0.2860**	**0.7463**
cam	0.4435	0.3718	0.0878
finger	0.3053	0.5045	0.5304
flash	0.0636	0.3855	0.2481
headphone	0.4749	0.5361	0.6788
colour	0.3800	0.2647	0.4452
megapixel	0.3310	0.0000	0.5614
brightness	0.1263	0.2982	0.4072
heating	0.3438	0.4669	0.8798

05 Conclusions

KOREA UNIVERSITY

- **Research Question에 대한 해석**

- Q1: 제품의 시리즈가 출시됨에 따라 새로운 제품의 업그레이드된 기술이 소비자의 의견에 크게 영향을 미치지 않는다고 보여진다.

- Q2: 이러한 결과를 통해 스마트폰 시장 및 기타 유관산업에서 지속적으로 업그레이드된 시리즈를 출시하는 데 있어 소비자의 실제 인식을 반영하여 취약점을 보완하고 좀 더 소비자의 의견을 반영한 기술개발이 이루어 질 수 있도록 시사점을 줄 수 있다.

05 Conclusions

KOREA UNIVERSITY

- **한계점 및 향후 계획**
  - 데이터의 개수
  - 완벽한 전처리의 어려움
  - 스마트폰이 아닌 다른 제품에 대한 일반화의 어려움
  - 임의의 k갯수 5개, 모델별 변화를 구체적으로 보려면 군집세분화가 필요함.
    →Parameter search
  - 감정 스코어링 알고리즘 뿐만 아니라 여러 클러스터링 알고리즘을 적용시켜볼 필요가 있음.
    → Hierachical Ensemble and so on.
  - 범용적 감정 단어사전이 아닌, 휴대폰 리뷰 분석에 맞는 감정단어셋 우리의 감정단어사전 기반 감정 스코어링 알고리즘 적용이 필요함
    → 감정단어 사전 안에서, 휴대폰 리뷰에서 가장 많이 나타나는 단어별 스코어 조정
    → 감정단어 사전에 없는, 휴대폰 리뷰데이터에서만 존재하는 단어들도 semi supervised 기반 스코어
  - 군집결과를 포함하여 최종적으로 키워드별, 휴대폰 모델별 감정군집의 변화가 유의미한지 검정할 필요가 있음.
    → 군집 타당성은?
    → 유의성 검정은?

- **의의**
  - 소비자 리뷰의 중요성 제고
  - 다양한 분야의 텍스트마이닝 가능성

감사합니다.

THANK YOU!