

소셜미디어 데이터 마이닝을 활용한 스마트 스피커 제품에 대한 기회 분석



박영재
건국대학교 산업공학과
dudwo1128@gmail.com



요약

- Amazon echo, Googlehome 등으로 대표되는 스마트 스피커는 무선 스피커에 인공지능 비서를 접목한 결과물
- 스마트 스피커는 대표적인 Technology-Push 제품으로, 제품에 대한 기술에 비해 상대적으로 놓치기 쉬운 고객의 니즈를 파악하고 기회를 분석하여 제품 개발에 도입하는 것이 중요함
- 시장에서 소셜 미디어의 영향력은 매우 크며, 제품 사용자들의 정보 교환 창구의 역할을 함
- 대중의 관심이 고조되고 시장이 본격적으로 형성되는 시기에 자사와 경쟁 제품에 대한 고객의 니즈 비교 분석은 시장 선두주자로서 우위 선점에 큰 도움을 줄 것으로 생각됨
- 제품 간의 비교가 이루어진다면 보다 구체적으로 특정 제품 대비 자사 제품의 개선 전략을 수립하는데 유용한 정보를 제공
- 소셜 미디어 데이터 마이닝을 이용한 제품 개발 기회 분석(Jeong, Yoon, and Lee, 2017)의 방법을 통해 스마트 스피커 경쟁 제품의 소비자 리뷰로 부터 도출된 기회를 비교
- 제품 간의 경쟁을 고려하여 이머징 제품의 기술 주도형(Technology-push) 개발 경향을 보조할 수 있는 시장 주도형(Market-pull) 제품 개발 기회를 포착하고자 함
- 본 발표는 1장 연구 배경 및 목적, 2장 이론적 배경, 3장 연구 방법, 4장 결과 분석, 5장 결론 및 향후 과제로 구성



1. 연구 배경 및 목적

1. 1 연구 배경

- 아마존 에코, 구글홈 등으로 대표되는 스마트 스피커는 무선스피커에 인공지능 비서를 접목
- 스마트 스피커 시장규모는 2015년 3억 6천만 달러 (약 4천 300억 원)에서 2020년 20억 달러 (약 2조 3천억원)로 6배 이상 커질 것으로 예측 (김영대, 2017)
- 스마트 스피커와 같은 이머징 제품의 경우에는 기술주도형 (Technology-push)으로 개발되는 경향을 지님 (Haines et al. 2007)
- 대중의 관심이 고조되고 시장이 본격적으로 형성되는 시기에 자사와 경쟁 제품에 대한 고객의 니즈 비교 분석은 시장 선두주자로서 우위 선점에 큰 도움을 줄 것으로 생각됨



1. 2 연구 목적

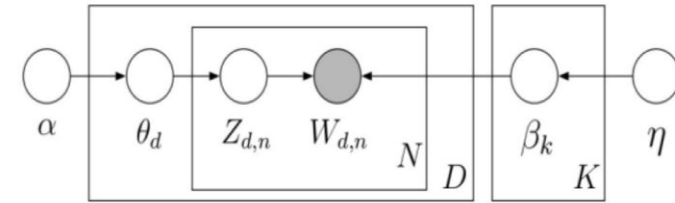
- 소셜 미디어 데이터 마이닝을 이용한 제품 개발 기회 분석(Jeong, Yoon, and Lee, 2017)의 방법을 통해 스마트 스피커 경쟁 제품의 소비자 리뷰로 부터 도출된 기회를 비교
- 경쟁을 고려하여 이머징 제품의 기술 주도형(Technology-push) 개발 경향을 보조할 수 있는 시장 주도형 (Market-pull) 제품 개발 기회를 포착하고자 함



2. 이론적 배경

2.1 토픽 모델링

- 기계 학습과 자연어 처리 분야에서 문서 집합의 주제를 파악하기 위해 사용하는 방법
- LDA(Latent Dirichlet Allocation)
 - 토픽 모델링 중에서 가장 많이 사용
 - Overfitting 문제가 없으며, 새로운 문서에 대해서도 쉽게 일반화가 가능함



K – total number of topics
 β_k – topic, a distribution over the vocabulary
 D – total number of documents
 θ_d – per-document topic proportions
 N – total number of words in a document (it fact, it should be N_d)
 $Z_{d,n}$ – per-word topic assignment
 $W_{d,n}$ – observed word
 α, η – Dirichlet parameters

- Several **inference algorithms** are available (e.g. sampling based)
- A few **extensions** to LDA were created:
 - Bigram Topic Model

<그림.1> LDA topic modeling

2.2 감성 분석

- 주어진 텍스트에서 긍정, 부정 혹은 중립을 파악하는 방법
- IBM Watson : Alchemy API
 - 키워드에 대한 긍정/ 부정을 판단
 - 1과 -1사이의 점수를 부과



Positive



Neutral



Negative

<그림.2> 감성 분석



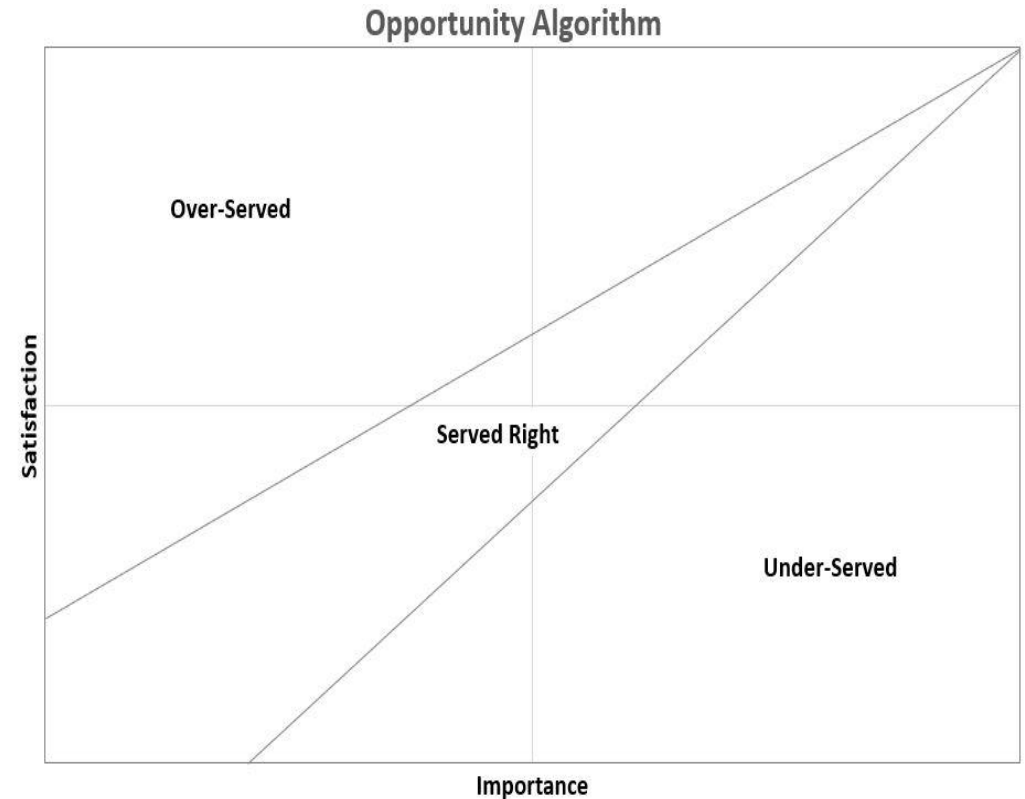
2. 이론적 배경

2. 3 Opportunity algorithm

- 제품/서비스의 잠재성 있는 Opportunity 식별 방법
- 중요도와 만족도의 값을 하나의 matrix에 결합 측정
- 다양한 고객 needs로부터 높은 기회를 가진 고객 needs 식별

Concept of opportunity algorithm

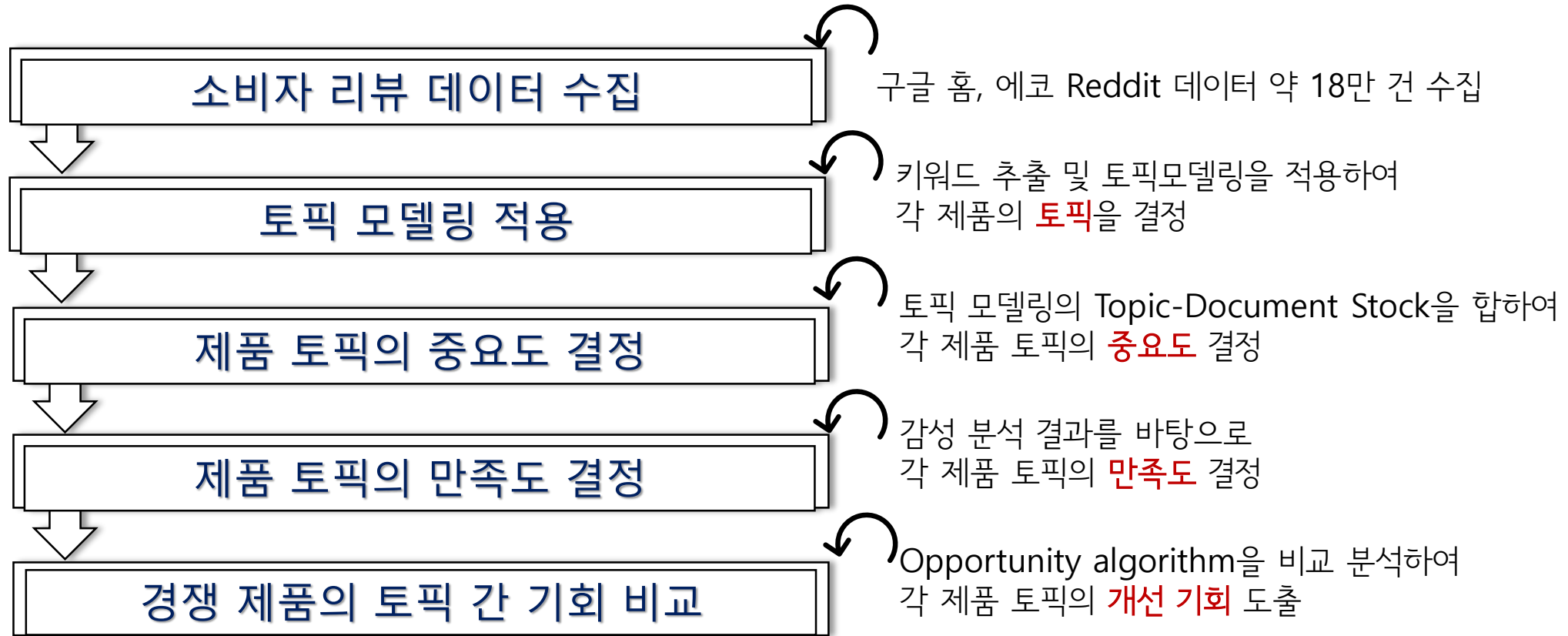
- Opportunity =
Importance + max(Importance-Satisfaction, 0)
- 중요도와 만족도의 값에 따라서 기회를
3가지 영역으로 구분할 수 있음
- Under-Served 영역에 해당하는 토픽들을
개선의 기회가 있는 토픽으로 판단



<그림.3> Opportunity algorithm landscape



3. 연구 방법





3. 연구 방법

3.1 소비자 리뷰 데이터 수집

- 소셜 뉴스 웹사이트인 Reddit을 활용
- 구글 Big Query를 활용하여 데이터 수집
- 작성자, 댓글 데이터 수집
- Echo 116,464건 vs. Google home 77,740건

3.2 토픽 모델링 적용

- 수집된 리뷰의 키워드를 전처리한 후 LDA topic modeling 적용
- 각 토픽에 해당하는 키워드의 확률 값을 고려하여 토픽을 Labeling함
- Google home과 Echo 모두 email, music, voice, lamp등의 키워드를 포함
- 그에 맞게 Phone Pairing, Music Streaming, Voice Command 등으로 토픽을 Label함



created_utc	subreddit_id	parent_id	body
1501761254	t5_3enp4	t1_dl3sj2a	Six, and she went on to list them.
1502453120	t5_3enp4	t1_dlg67c	I have had the shortcuts for approx two weeks now but not the services
1503587971	t5_3enp4	t1_dm2cvyj	I don't have that listed. Is there a way to install it?
1503750205	t5_3enp4	t1_dm4e69f	This is a great feature.
1504197743	t5_3enp4	t1_dmdmas9	You want to know the <u>time from MY watch?</u> TOO BAD.

<그림.5> 소비자 리뷰 데이터 수집

Topic	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword
<i>Voice Control</i>	spotify	alarm	speaker	bedroom
<i>Echo Connect</i>	phone	wall	sync	internet
<i>Light Control</i>	light	ifttt	night	lamp
<i>Echo Plus</i>	song	setup	echo plus	email
<i>Echo Dot</i>	echo dot	alexa app	routine	alexa
<i>Audio/Video Remote Control</i>	account	remote	word	stream
<i>Notification</i>	tunein	notification	voice	audio
<i>Smart Plug</i>	timer	command	plug	alexa

<표.1> 제품 토픽 결정

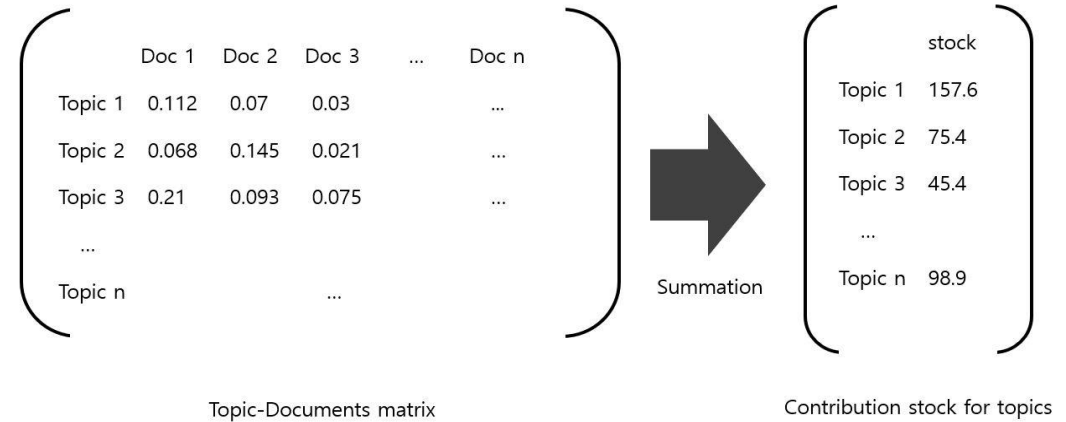
예시) 표.1의 Smart Plug 토픽의 경우 상위 키워드인 timer, command, plug의 확률 값을 종합적으로 고려하여 labeling



3. 연구 방법

3.3 제품 토픽의 중요도 결정

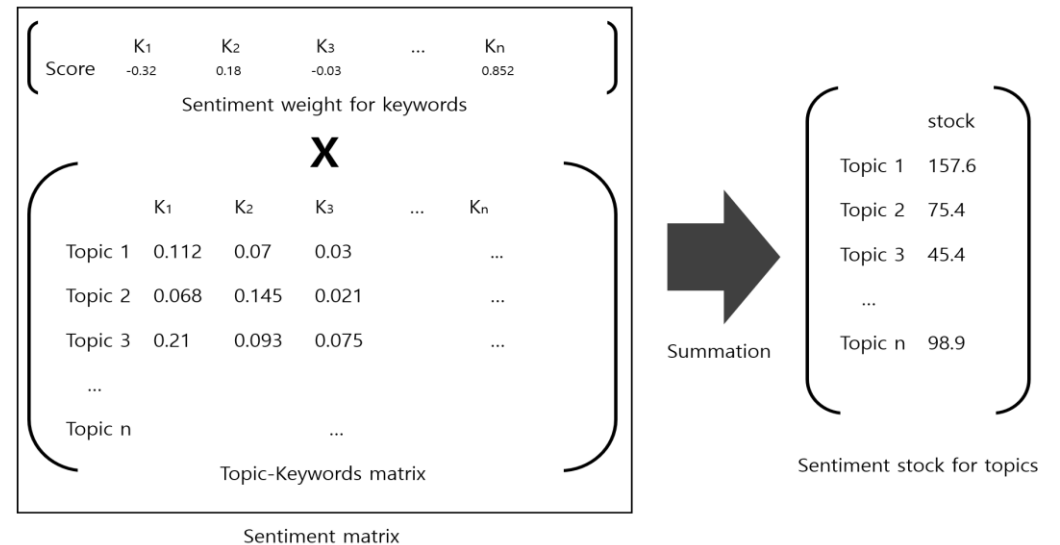
- 토픽에 해당하는 문서가 많다는 것은 해당 토픽에 대한 고객의 관심도가 높다는 것을 의미
- 고객의 관심을 기업의 입장에서 중요한 것으로 해석
- 각각의 문서가 토픽에 포함될 확률의 합을 해당 토픽의 중요도로 간주



<그림.6> 제품 토픽의 중요도 결정

3.4 제품 토픽의 만족도 결정

- 토픽에 해당하는 키워드 분포확률과, 각 키워드의 감성점수 활용
- 감성점수는 -1과 1사이의 점수 분포를 가짐
- 감성점수의 값이 1에 가까울수록 만족, -1에 가까울수록 불만족으로 간주
- 각각의 문서가 토픽에 포함될 확률에 감성 점수를 곱한 값의 합을 해당 토픽의 만족도로 간주



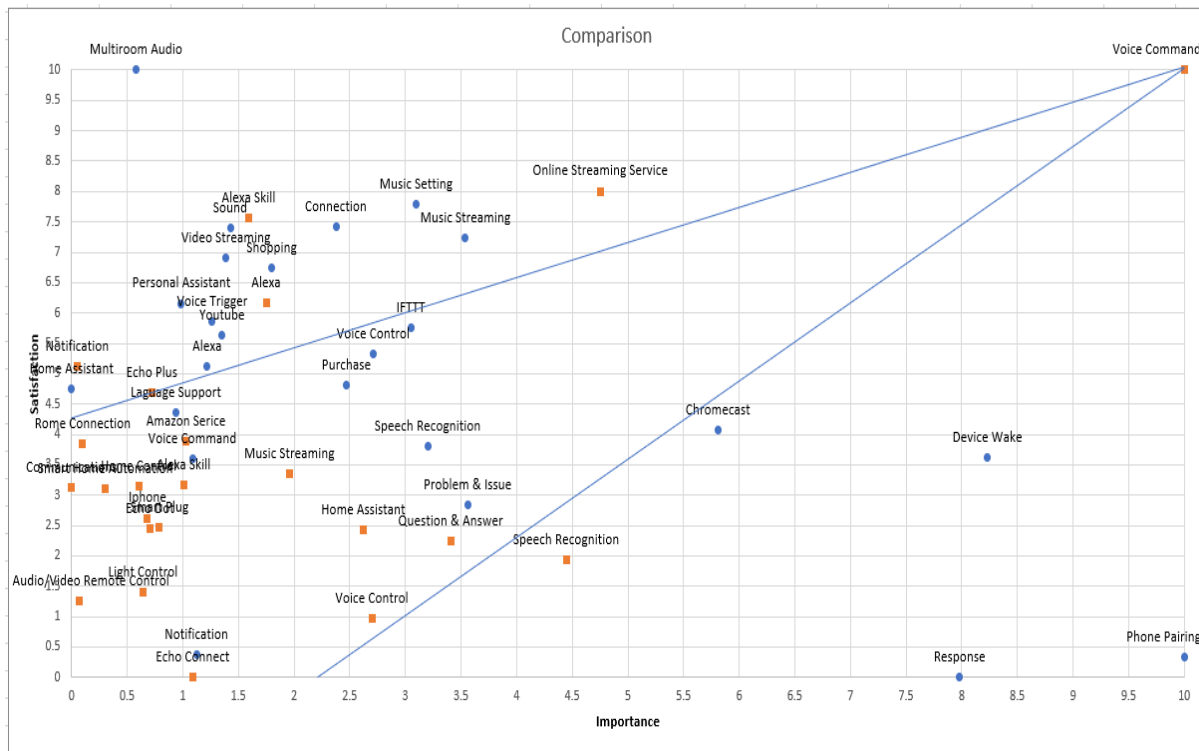
<그림.7> 제품 토픽의 만족도 결정



3. 연구 방법

3.5 경쟁 제품의 토픽 간 기회 비교

- 계산된 각 제품의 중요도, 만족도에 Opportunity algorithm을 적용하여 경쟁 제품의 기회를 비교



<그림.8> 경쟁 제품의 기회 비교 landscape

Topic	Importance	Satisfaction
Sound	1.43	7.40
Personal Assistant	0.99	6.16
Youtube	1.35	5.64
Chromecast	5.81	4.07
Home Assistant	0.00	4.75
Speech Recognition	3.20	3.81
IFTTT	3.05	5.76
Alexa	1.22	5.11
Voice Control	2.71	5.33
Problem & Issue	3.56	2.85
Device Wake	8.23	3.61
Music Setting	3.10	7.79
Multiroom Audio	0.58	10.00
Phone Pairing	10.00	0.33
Response	7.98	0.00
Connection	2.38	7.41
Music Streaming	3.54	7.23
Purchase	2.47	4.80
Voice Trigger	1.26	5.86
Voice Command	1.09	3.59
Laguage Support	0.94	4.37
Video Streaming	1.39	6.90
Notification	1.12	0.37
Shopping	1.80	6.75

<표.2> GoogleHome의
중요도와 만족도

Topic	Importance	Satisfaction
Voice Control	2.70	0.96
Echo Connect	1.09	0.00
Light Control	0.64	1.41
Echo Plus	0.72	4.69
Echo Dot	0.70	2.45
Audio/Video Remote Control	0.07	1.26
Notification	0.05	5.12
Smart Plug	0.79	2.46
Smart Home	0.30	3.11
Automation	0.30	3.11
Alexa	1.75	6.17
Alexa Skill	1.01	3.17
Communications	0.00	3.13
Home Control	0.60	3.16
Question & Answer	3.41	2.24
Amazon Service	1.03	3.88
Alexa Skill	1.59	7.57
Home Assistant	2.62	2.42
Music Streaming	1.96	3.35
Voice Command	10.00	10.00
Online Streaming	4.75	8.00
Service	0.10	3.84
Rome Connection	0.68	2.62
lphone	0.68	2.62
Speech Recognition	4.45	1.93

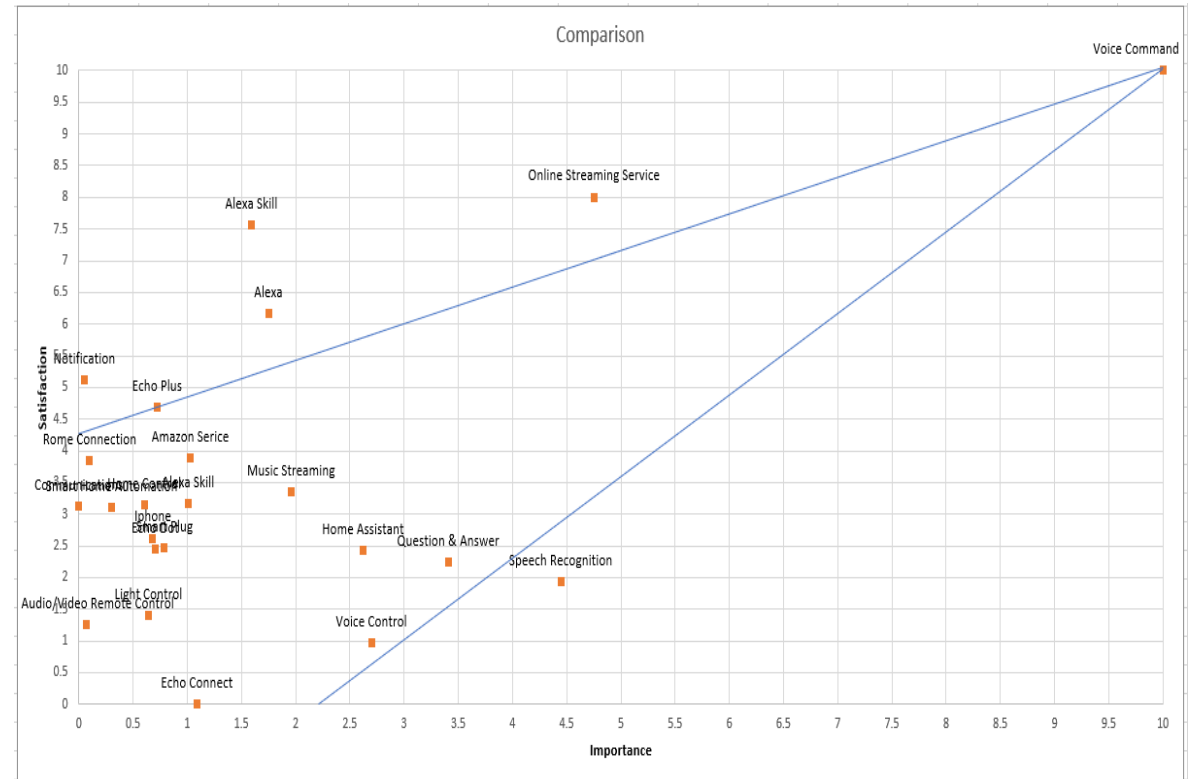
<표.3> Amazon Echo의
중요도와 만족도



4. 결과 분석

4. 1 Amazon Echo

- 소비자들은 Voice Command를 상당히 중요하게 생각하며, Echo는 이를 만족시키고 있음
- 하지만, Voice Command가 다른 토픽들에 비해 상대적으로 매우 중요하게 인식되고 있음
- Echo의 토픽은 상당수가 Served-Right 영역에 분포하며, 소비자가 관점의 중요도 대비 만족도를 충족시킴
- GoogleHome의 Home assistant 토픽이 Over-Serve영역에 위치함
- Echo의 Home assistant 토픽은 Served-Right영역에 위치하므로 해당 토픽은 파괴적 혁신을 일으킬 수 있음



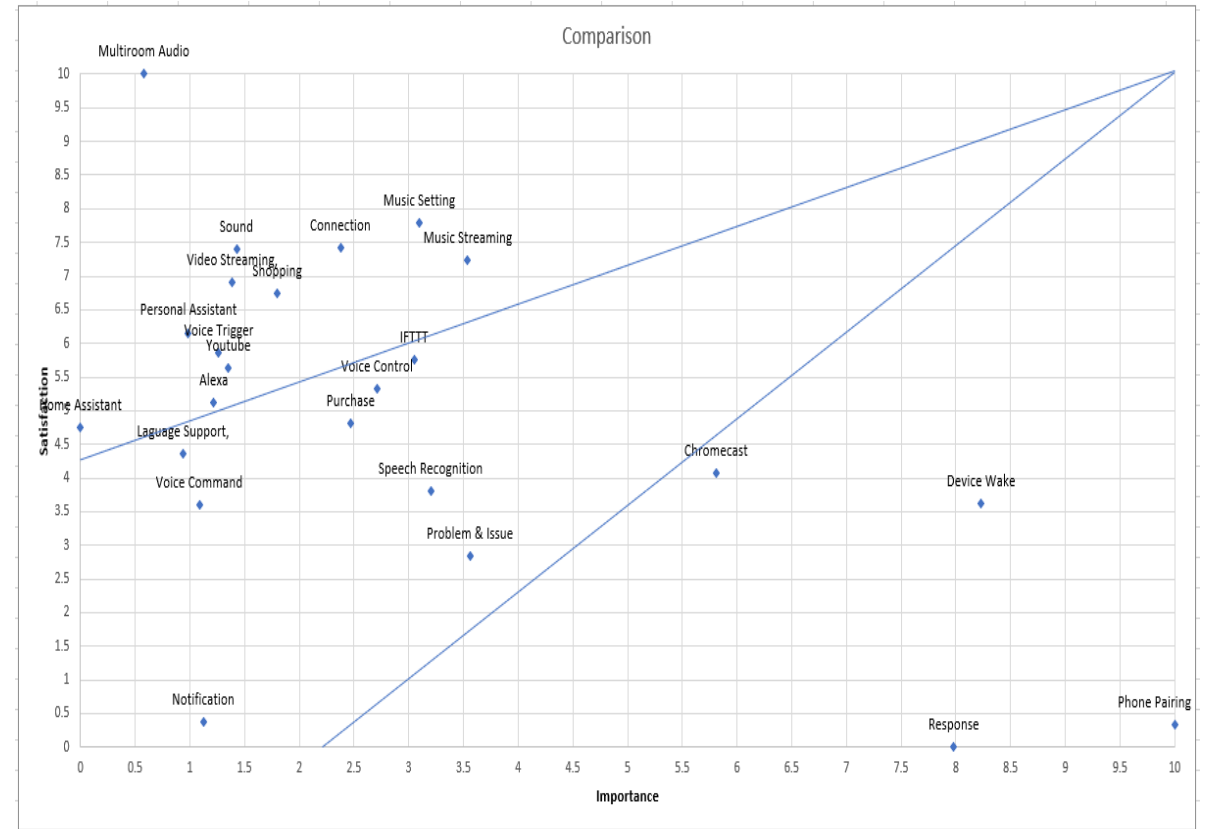
<그림.9> Amazon Echo의 기회 landscape



4. 결과 분석

4. 2 GoogleHome

- 상당 수의 토픽이 Over-Served 영역에 분포함
- Under-Served 영역에 Chromecast, Device Wake, Phone Pairing, Response 토픽이 위치하고 있으며, 이는 개선 기회임
- 특히 Chromecast 토픽은 Google Home의 독자적인 특징이고 아직 충분히 만족되지는 않았으나 소비자들에게 중요하게 인지되고 있음
- Echo와 Google Home에 동시에 출현한 토픽인 Voice Command의 경우 Echo에서는 중요도 10을 가지만, Google Home에서는 2.7 부근의 중요도를 가짐
- 평균적인 Topic-Document 확률 값은 유사하나, Google Home에서 Voice Command Topic의 중요도 값이 낮음
- 따라서, Echo의 소비자들이 Voice Command를 독보적으로 중요하게 생각하는 것에 반해, Google Home 소비자들은 Voice Command 이외의 다른 토픽들 역시 중요하게 생각하고 있음



<그림.10> GoogleHome의 기회 landscape



5. 결론 및 향후 과제

5. 1 결론

- Amazon echo, GoogleHome 등으로 대표되는 스마트 스피커는 무선 스피커에 인공지능 비서를 접목한 결과물
- 대중의 관심이 고조되고 시장이 본격적으로 형성되는 시기에 자사와 경쟁 제품에 대한 고객의 니즈 비교 분석은 시장 선두주자로서 우위 선점에 큰 도움을 줄 것으로 생각됨
- 소셜 미디어 데이터 마이닝을 이용한 제품 개발 기회 분석(Jeong, Yoon, and Lee, 2017)의 방법을 통해 스마트 스피커 경쟁 제품의 소비자 리뷰로 부터 도출된 기회를 비교함
- Amazon echo의 제품 토픽은 대다수가 served-right영역에 분포하며, 이는 해당 제품의 토픽이 소비자 관점의 중요도 대비 충분한 만족도를 가짐을 의미함
- Google home의 경우 over-served영역과 under-served영역에 다수의 토픽이 분포, 이는 해당 제품의 토픽이 다수의 개선 가능성을 내포하고 있음을 의미함
- 제품 간의 비교를 통해 보다 구체적으로 특정 제품 대비 자사 제품의 개선 전략을 수립하는데 유용한 정보를 제공할 수 있을 것으로 기대함
- 충분한 양의 리뷰를 수집할 수 있는 다양한 제품에 적용이 가능할 것이며, 세 개 이상의 제품에 적용하여 소비자 관점의 기회를 포착하는 데 도움을 줄 것으로 기대함

5. 2 향후 과제

- 본 연구는 제품 간의 경쟁을 고려하여 이머징 제품의 시장 주도형(Market-pull) 기회를 포착하였으나 한계가 존재함
- 전체 수집기간을 분할하지 않고 분석하여 시간에 따라 변화하는 소비자의 니즈를 반영하기 어려움
- 제품별로 각각 다른 데이터를 활용하기 때문에, 절대적인 비교가 어려움
- 이는 향후 과제에서 해결해야 할 문제