

---

# Steam Game Recommendation based on Multi-channel Multi-model Analyses

---

2018. 06. 14

2017020554 안지영  
2017021205 송재승  
2017021198 최지은  
2018020511 김문수



# Contents

## 1. Introduction - Steam

## 2. Modeling

- ① Clustering
- ② Topic Modeling
- ③ Multi-model feature learning

## 3. Conclusion

## 4. Q & A



## Introduction - Steam



# Introduction - Steam

- 다양한 장르의 게임을 구매할 수 있는 온라인 게임 유통 시스템
- 전세계 게임 유저들과 정보를 교환할 수 있는 세계 최대의 게임 커뮤니티



# Introduction – Data description

- 23221games → 19201 games (Preprocessing)
- 2040670 review → Positive : 1525466 reviews Negative: 515204 reviews

## Meta data

Game ID

Name

Recent positive ratio

The number of positive

Publisher

User defined tags

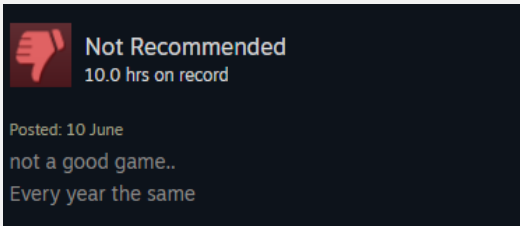
Developer

## Reviews

### Recommend reviews



### Not Recommend reviews

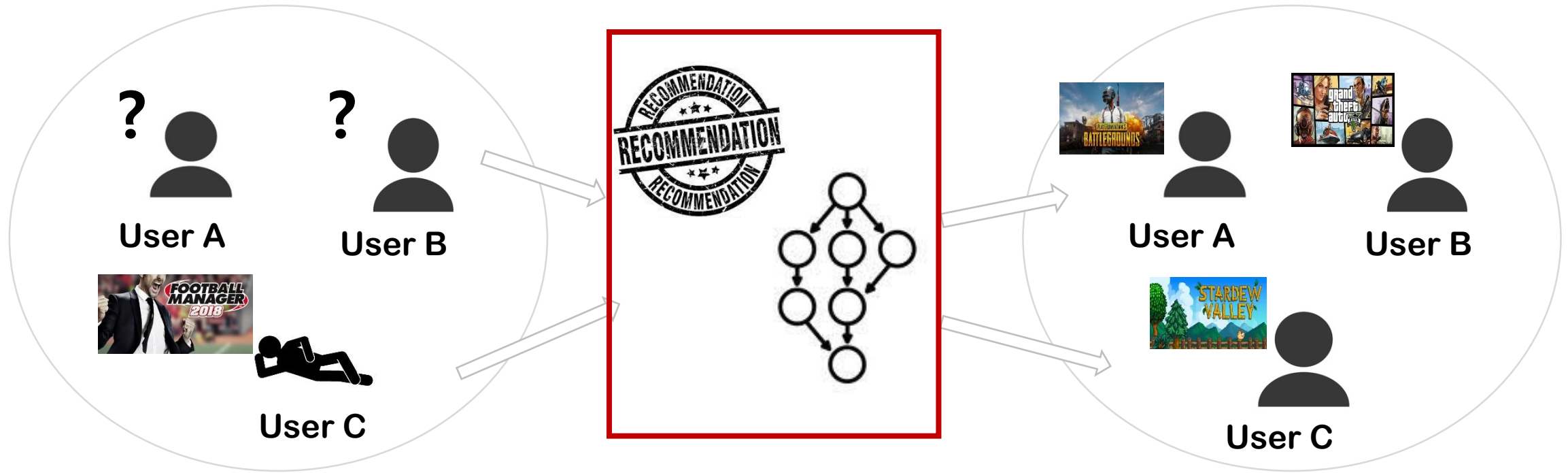


## Game Image

Logo image

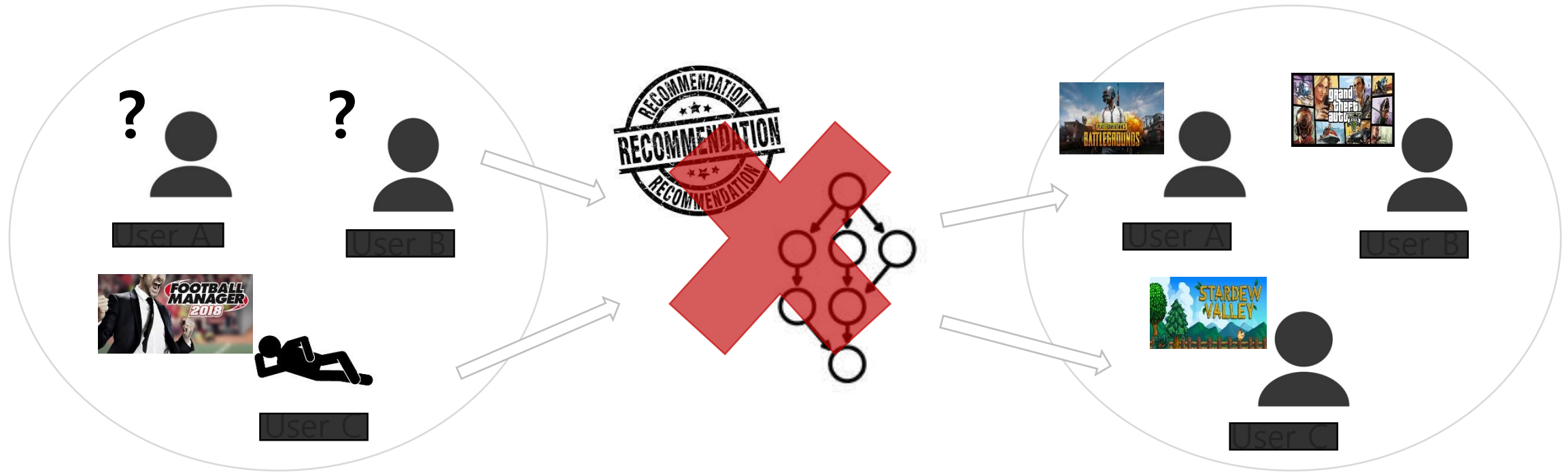


# Introduction – Goals



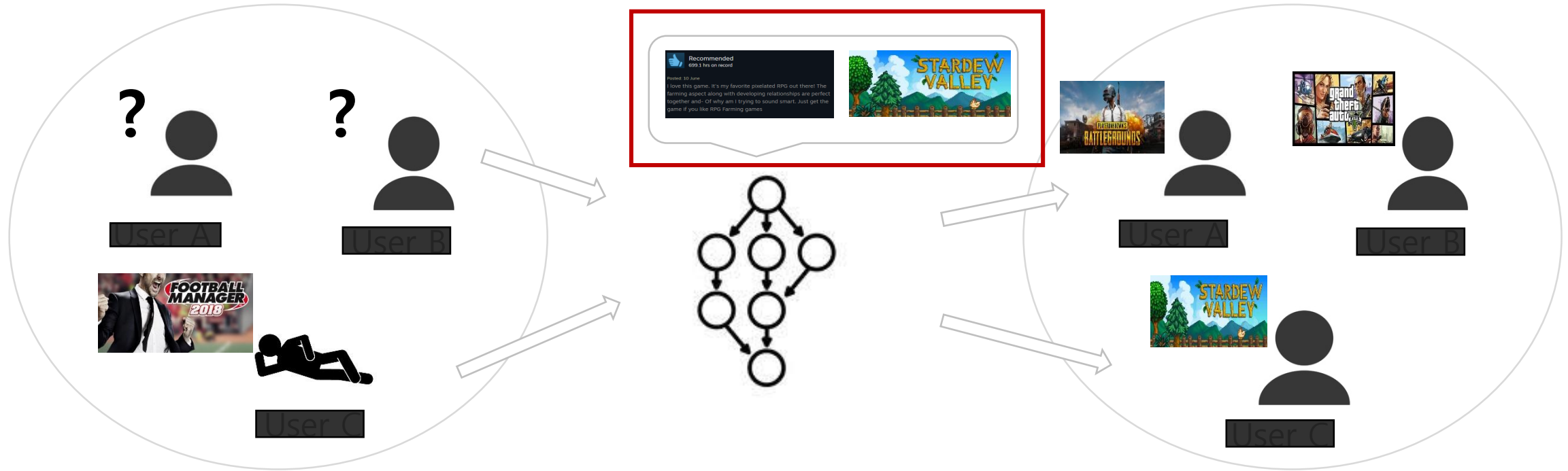
Steam 게임 유저에게 새로운 게임을 추천해 줄 수 있는 **Recommendation algorithm**을 개발

# Introduction – Goals



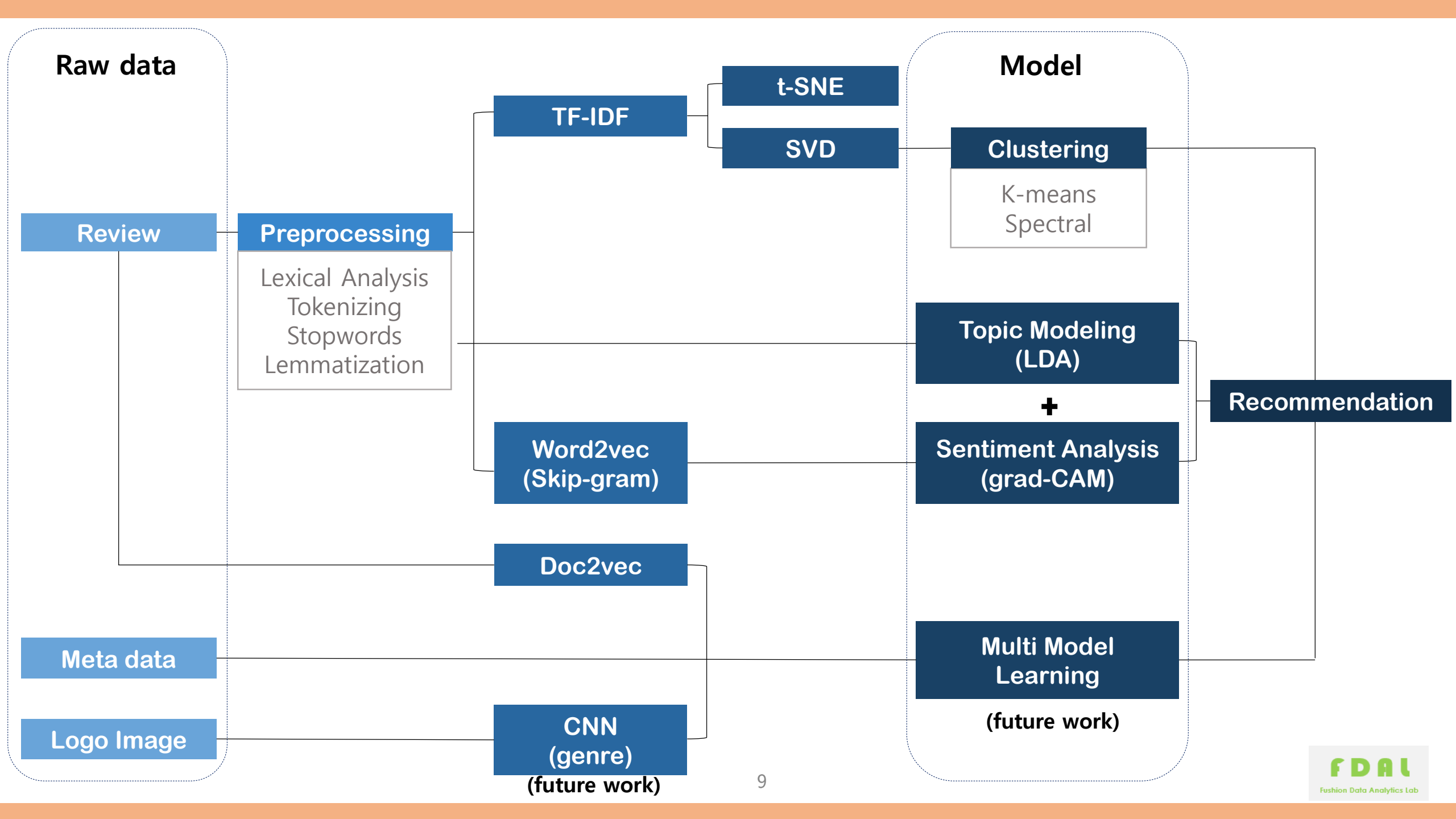
User information이 없어 User based modeling 불가

# Introduction – Goals



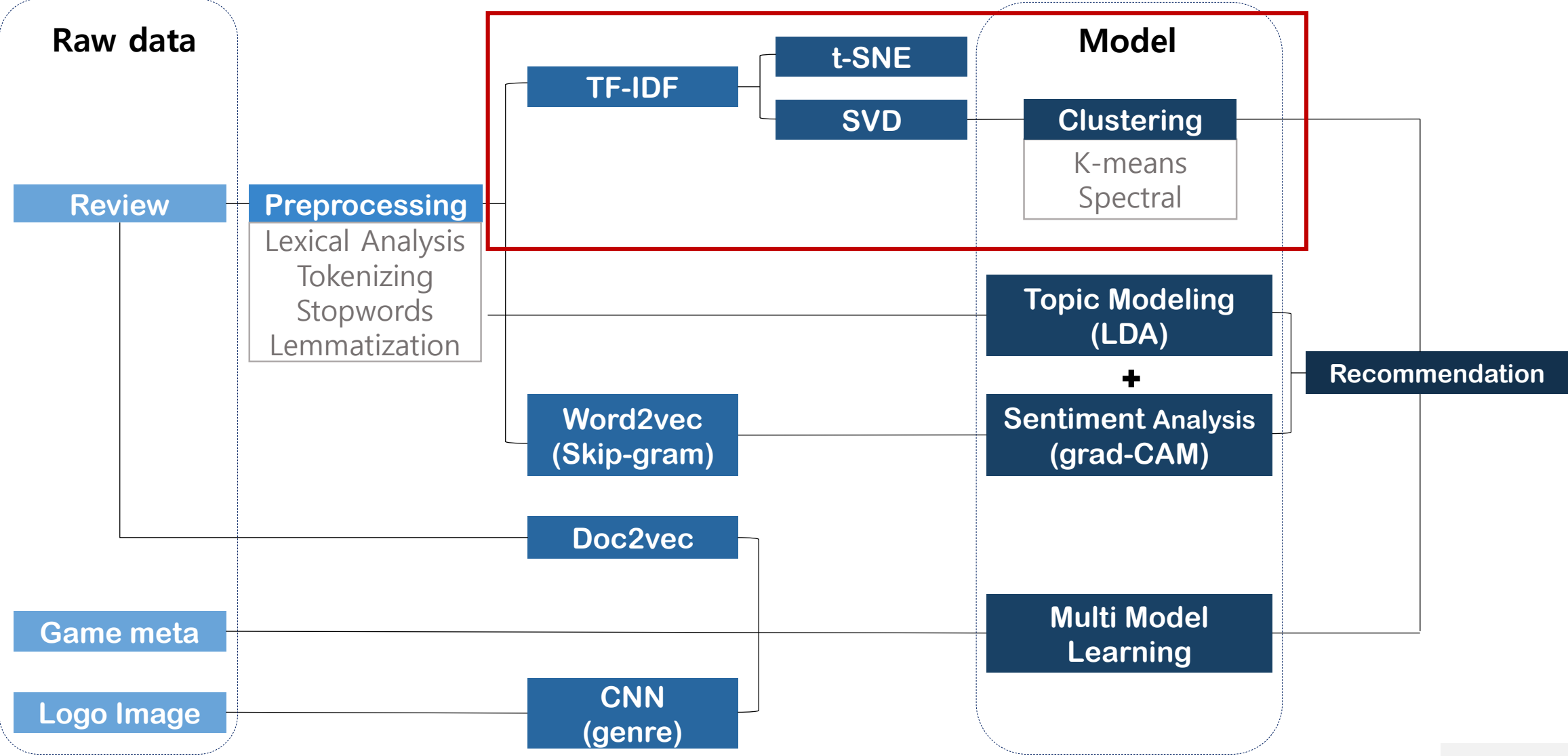
**Content (Text data, Image data) based modeling**을 통해 게임  
Recommendation Algorithm 개발





## Modeling - Clustering

# Modeling – Clustering



# Modeling – Clustering

## 1. Preprocessing

“Yet another fun game as good as the first Minecraft Story Mode I’d recommend this game and I hope this game has a dlc to continue the story .....”

### ➤ Tokenization

[Yet, another, fun, game, as, good, as, the, first, Minecraft, Story, Mode, I’d, recommend, this, game, and, I, hope, this, game, has, a, dlc, to, continue, the, story, .....] ]

### ➤ Tagging

[(Yet, RB), (another, DT), (fun, NN), (game, NN), (as, RB), (good, JJ), (as, IN), (the, DT), (first, JJ), (Minecraft, NNP), (Story, NNP), (Mode, NNP), (I’d, NNP), (recommend, VB), (this, DT), (game, NN), (and, CC), (I, PRP), (hope, VBP), ...]

# Modeling – Clustering

## 1. Preprocessing

### ➤ Delete stopwords

[Yet, another, fun, game, **as**, good, **as**, **the**, first, Minecraft, Story, Mode, I'd, recommend, **this**, game, **and**, I, hope, **this**, game, **has**, **a**, dlc, **to**, continue, **the**, story, funny, game]

### ➤ Lemmatization

[Yet, another, fun, game, good, first, Minecraft, Story, Mode, I'd, recommend, game, I, hope, game, dlc, continue, story, **fun**, game]

# Modeling – Clustering

## 1. Preprocessing

### ➤ Delete stopwords

[Yet, another, fun, game, **as**, good, **as**, **the**, first, Minecraft, Story, Mode, I'd, recommend, **this**, game, **and**, I, hope, **this**, game, **has**, **a**, dlc, **to**, continue, **the**, story, funny, game]

### ➤ Lemmatization

[Yet, another, fun, game, good, first, Minecraft, Story, Mode, I'd, recommend, game, I, hope, game, dlc, continue, story, **fun**, game]

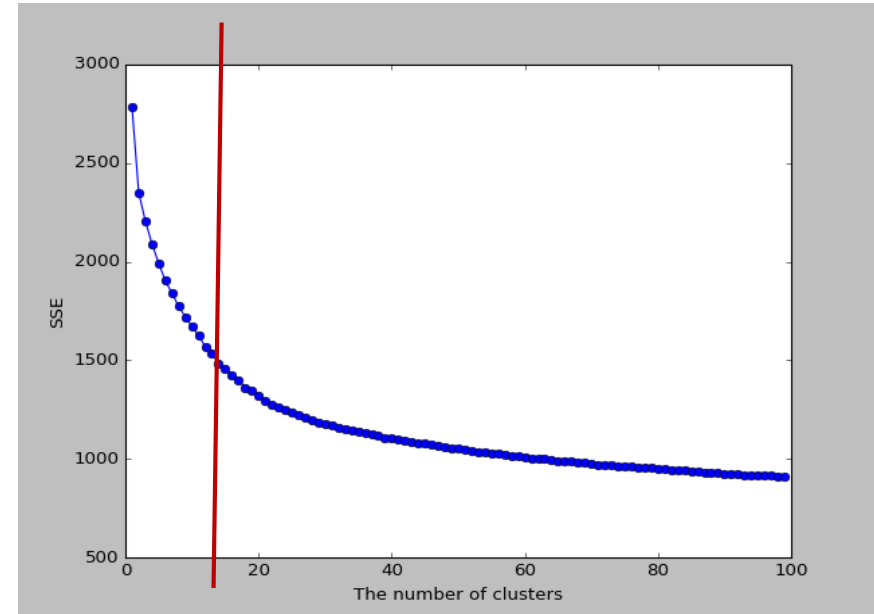
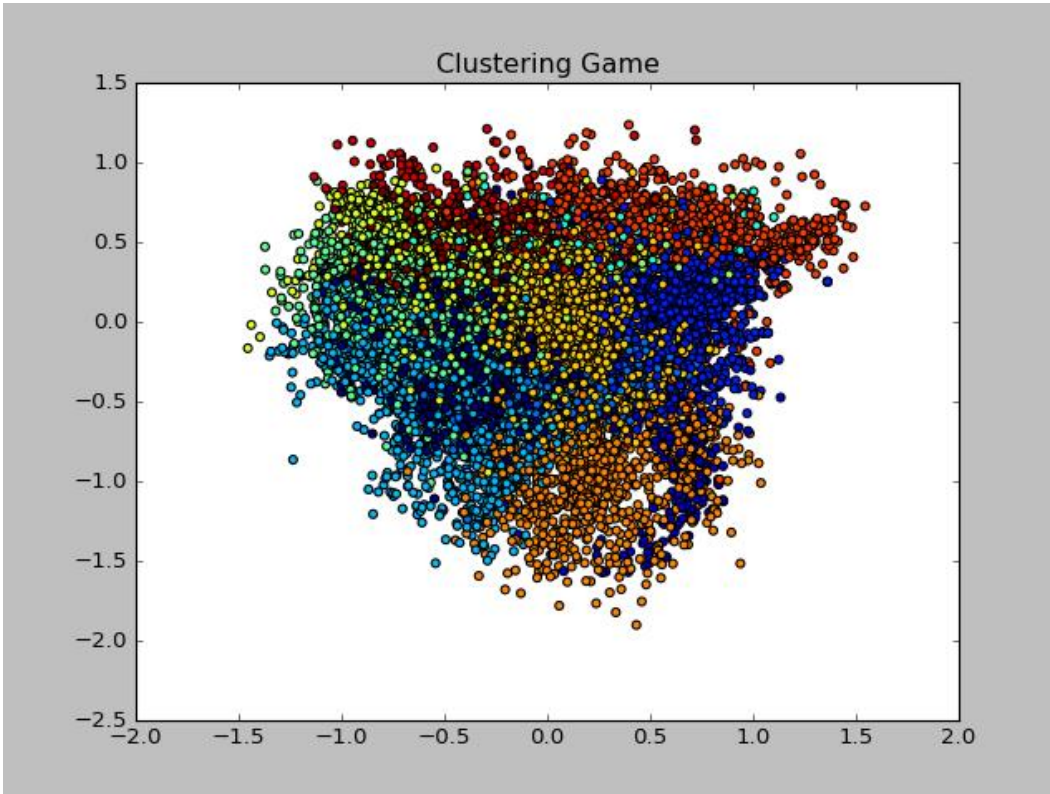
# Modeling – Clustering

## 2. Game Clustering

- **TF-IDF Matrix**를 활용하여 **Game**간 유사도 산출 **frequency**  
→ 40이하 단어 제거
- **Truncated Singular Vector Decomposition**을 활용하여 **Dimensionality Reduction**  
→ 약 300,000개의 단어를 50개의 단어로 **reduction** 진행

# Modeling – Clustering

## 3. K-Means Clustering



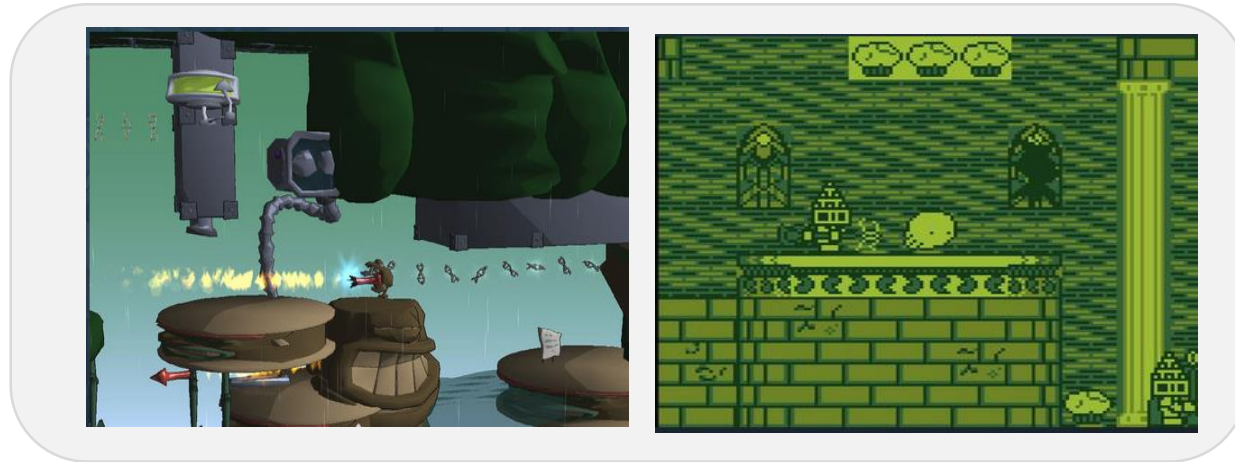
- Elbow를 활용한 최적 cluster개수 산정  
→ **k = 14**로 하여 **K-means clustering**



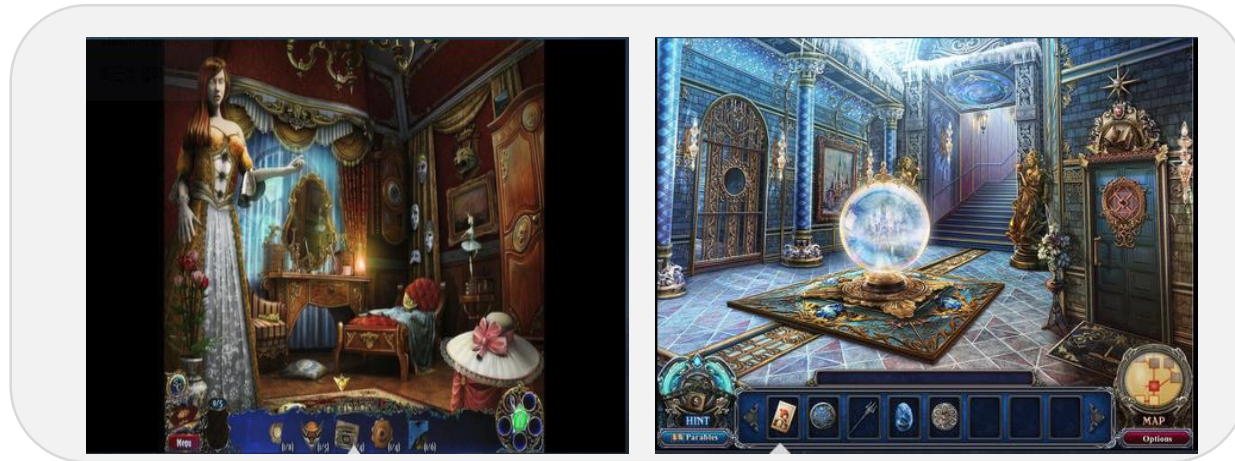
# Modeling – Clustering

## 3. K-Means Clustering

1	2D 게임
2	방탈출, 클릭
3	VR, 3D
4	스토리 진행, 미연시
5	유사도 없음
6	타워 디펜스
7	유사도 없음
8	레이싱 게임
9	FPS, 전략
10	인디
11	아케이드
12	호러
13	비행
14	유사도 없음



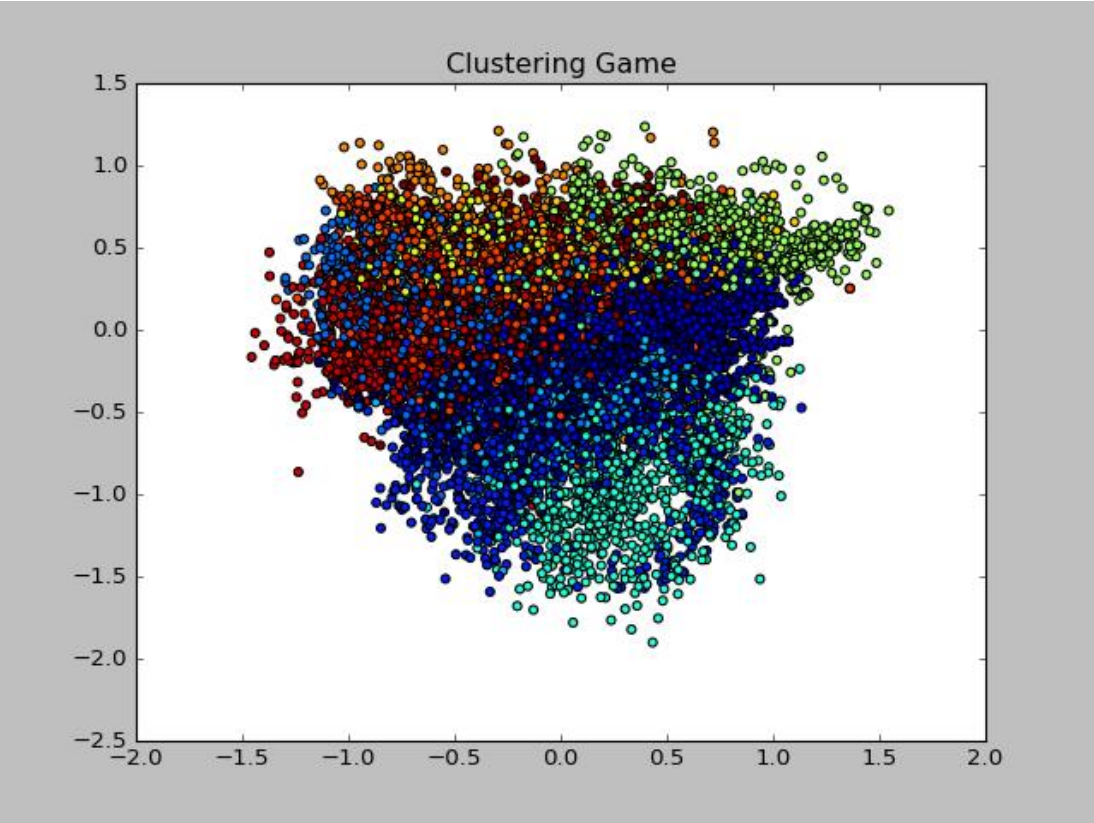
1 Cluster : 2D게임 ( ex : 마리오, 항아리게임 )



2 Cluster : 방탈출 , 원클릭 게임

# Modeling – Clustering

## 3. Spectral Clustering



1	2D 게임
2	외계 생물 관련
3	유사도 없음
4	3D 게임
5	미연시게임
6	퍼즐,전략게임
7	레이싱
8	비행 관련 게임
9	전쟁
10	유사도 없음
11	우주 비행
12	아케이드
13	유사도 없음
14	유사도 없음

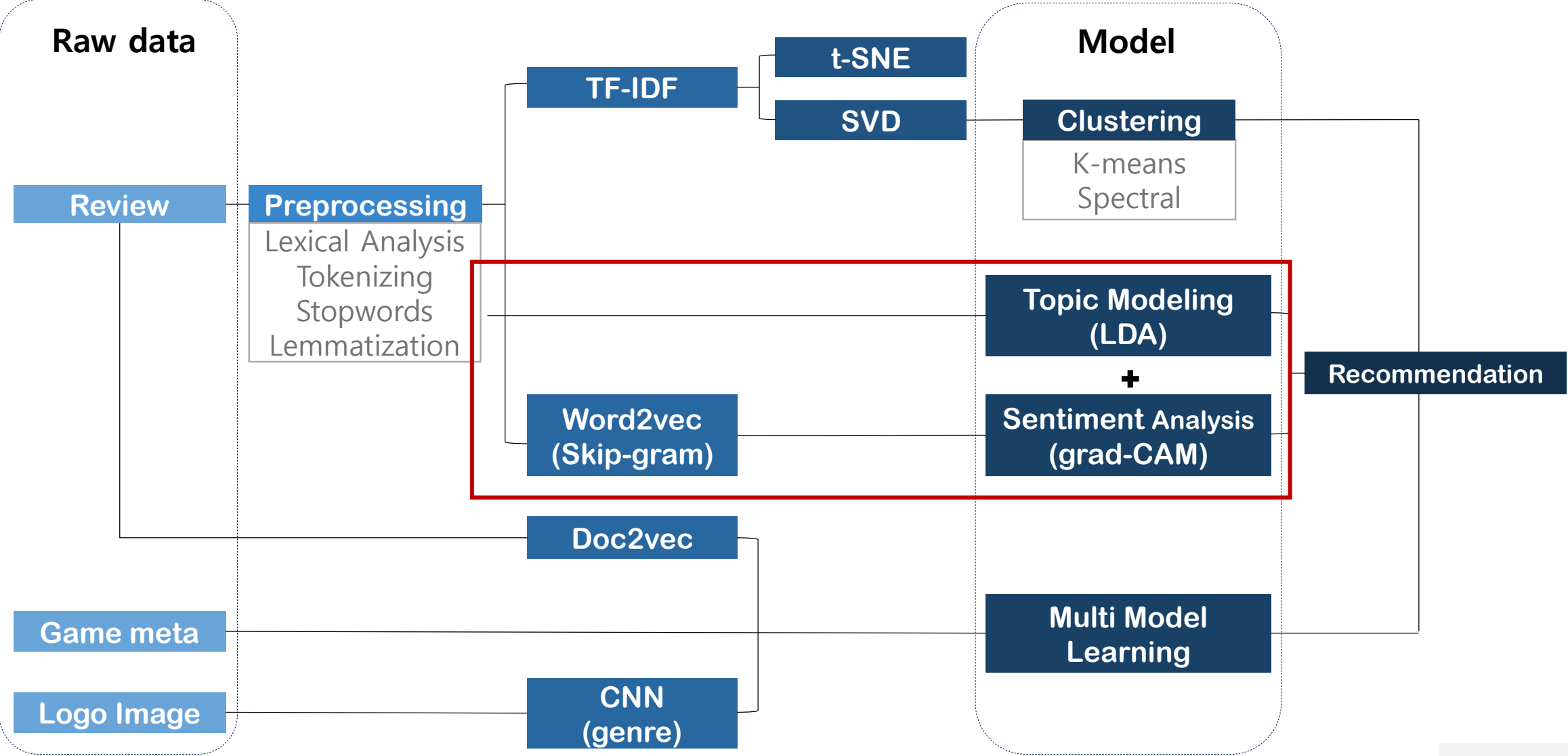
Cluster내 게임을 확인하여  
Cluster에 대한 군집명 정함



## Modeling – Topic Modeling

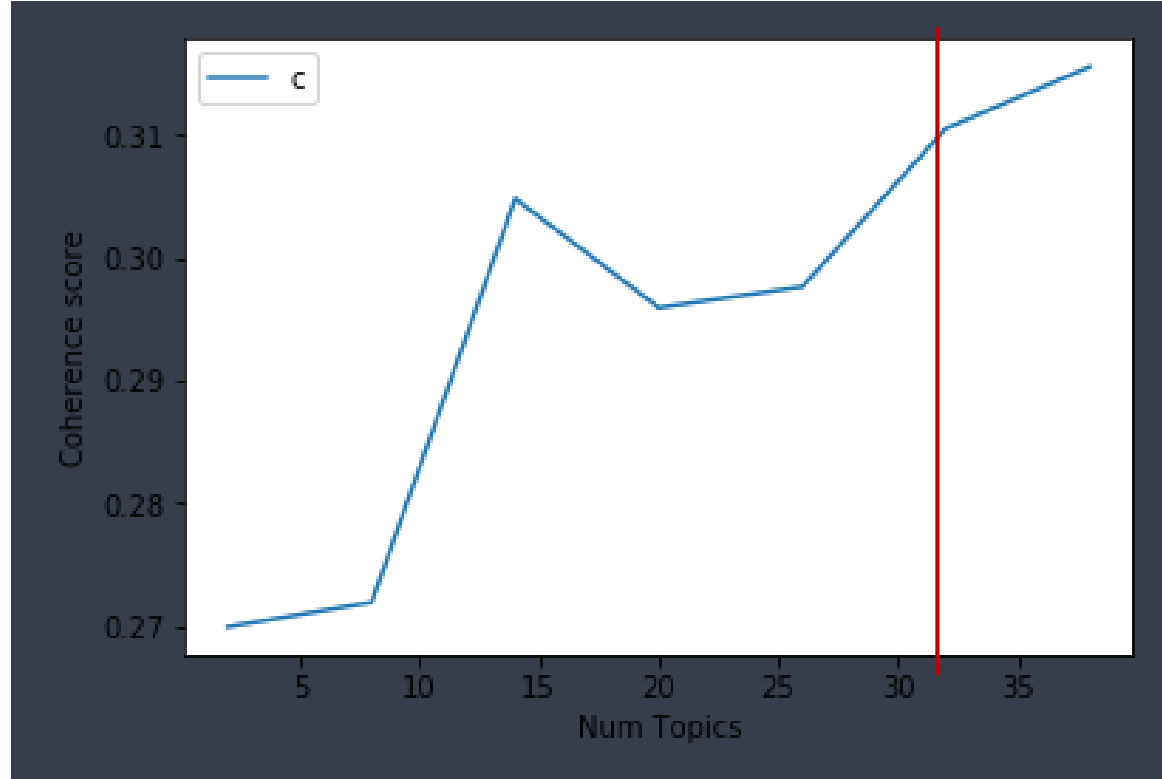


# Modeling – Topic modeling



# Modeling – Topic Modeling

## 1. LDA



Coherence score를 활용 **최적Topic 수 결정** → **32개의 Topic** 설정

# Modeling – Topic Modeling

## 1. LDA

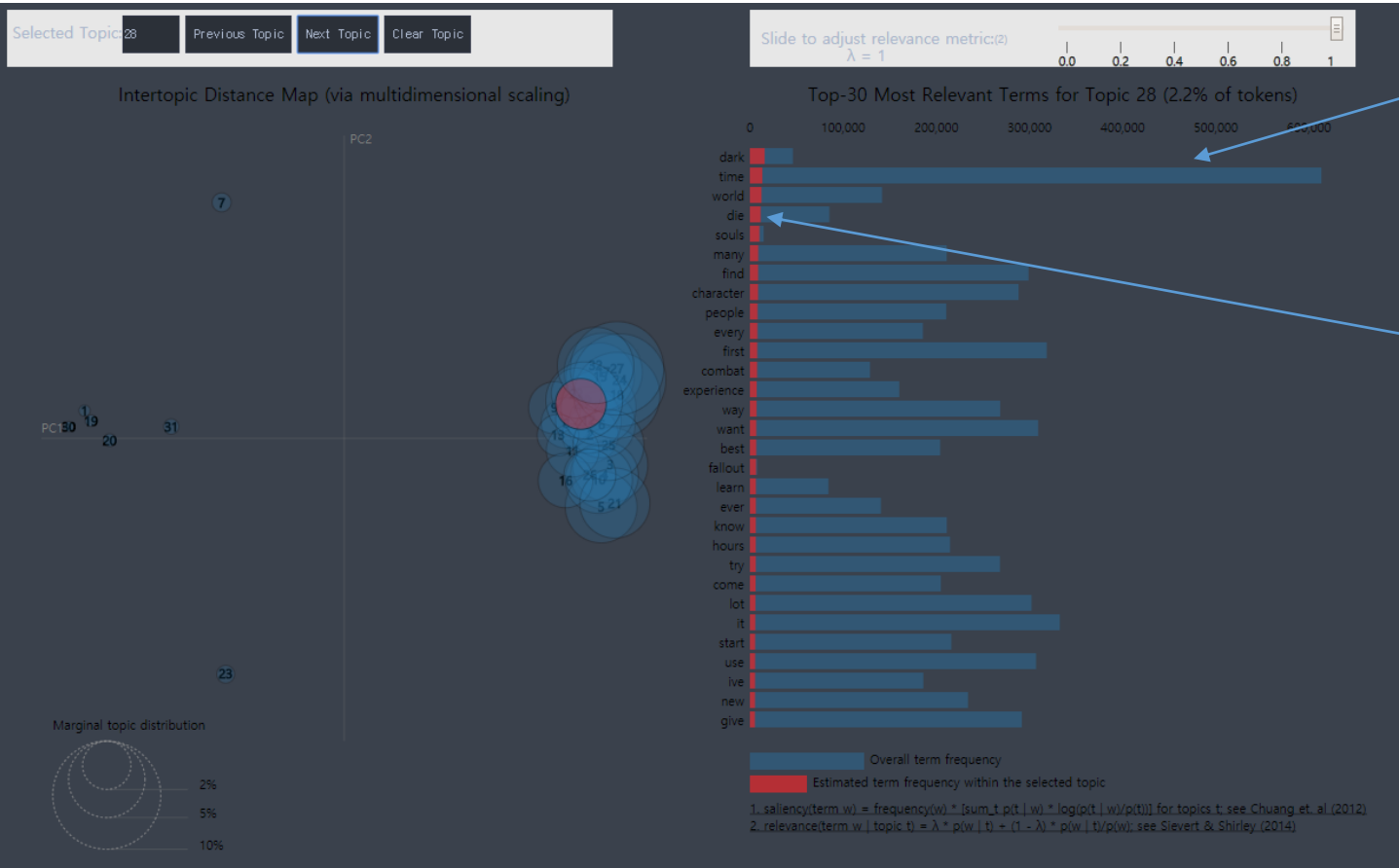
**13 Topic 을 구성하는 단어 People , 가중치: 0.012**

**Topic**

13	0.015*"players" + 0.012*"people" + 0.012*"fun" + 0.009*"time" + 0.008*"free" + 0.007*"pay"+...
1	0.019*"music" + 0.009*"time" + 0.009*"fun" + 0.008*"simple" + 0.007*"relax" + 0.006*"little" +...
14	0.021*"story" + 0.015*"character" + 0.010*"end" + 0.008*"visual" + 0.007*"novel" + 0.006*"read"+...
6	0.020*"quack" + 0.016*"nope" + 0.013*"gta" + 0.010*"hack" + 0.010*"pinball" + 0.010*"police"+...
28	0.010*"kill" + 0.007*"time" + 0.006*"first" + 0.005*"best" + 0.005*"run" + 0.005*"dead" + 0.005*"zombies"+...
11	0.012*"fun" + 0.010*"version" + 0.010*"mode" + 0.009*"original" + 0.009*"online" + 0.009*"multiplayer"+...
3	0.016*"boss" + 0.011*"enemies" + 0.009*"fight" + 0.008*"attack" + 0.007*"combat" + 0.007*"time"+...
24	0.018*"story" + 0.014*"combat" + 0.013*"character" + 0.011*"rpg" + 0.008*"quest" + 0.007*"system"+...
2	0.030*"level" + 0.011*"control" + 0.011*"fun" + 0.011*"time" + 0.008*"jump" + 0.007*"challenge"+...
7	0.021*"race" + 0.013*"car" + 0.013*"drive" + 0.011*"cars" + 0.009*"track" + 0.008*"fun" + 0.006*"time"+...
25	0.069*"puzzle" + 0.011*"level" + 0.010*"solve" + 0.008*"time" + 0.008*"fun" + 0.008*"challenge" + 0.007*"think"+...
12	0.017*"object" + 0.014*"hide" + 0.014*"story" + 0.010*"find" + 0.010*"puzzle" + 0.007*"time" + 0.007*"evil"+...
16	0.016*"character" + 0.015*"story" + 0.009*"first" + 0.008*"love" + 0.008*"time" + 0.006*"adventure"+...
0	0.015*"warriors" + 0.015*"dynasty" + 0.013*"chinese" + 0.010*"ys" + 0.010*"trail" + 0.009*"disgaea"+...
20	0.009*"turn" + 0.009*"time" + 0.008*"strategy" + 0.008*"fun" + 0.006*"base" + 0.005*"different" + 0.005*"battle"+...
29	0.005*"dredd" + 0.003*"feelsfeels" + 0.003*"touhou" + 0.002*"michonne" + 0.002*"clementine" + 0.002*"r"+...
26	0.009*"buy" + 0.008*"work" + 0.008*"bad" + 0.007*"bug" + 0.007*"fix" + 0.007*"time" + 0.007*"look"+...
30	0.048*"flight" + 0.040*"fly" + 0.020*"sim" + 0.020*"quake" + 0.018*"plane" + 0.018*"plan" + 0.016*"aircraft"+...
8	0.012*"war" + 0.006*"time" + 0.006*"best" + 0.005*"total" + 0.004*"ai" + 0.004*"want" + 0.004*"many"+...
5	0.021*"build" + 0.008*"time" + 0.007*"need" + 0.006*"hours" + 0.006*"work" + 0.006*"fun" + 0.005*"want"+...

# Modeling – Topic Modeling

## 1. LDA



전체 단어의 빈도수

Topic에 해당 단어의 가중치

28 Topic 구성



# Modeling – Topic Modeling

## 2.1 Topic별 분석

1	왕궁 RPG	warriors	dynasty	disgaea	sumurai
2	음악	fysuc	fun	relax	soundtrack
3	퀘스트	level	control	time	challenge
4	Laid(Boss)	boss	enemies	fight	attack
5	아케이드	hide	puzzle	time	click
6	타이쿤	build	time	work	city
7	GTA	police	mission	prison	gang
8	Car race	race	car	drive	track
9	전쟁	war	time	strategy	empire
10	던전	dungeon	character	items	quest
11	우주비행선	space	planet	crew	pirate
12	Team	version	online	multiplayer	PC
13	FPS	shoot	weapons	sgiiiter	gun
14	무료	pay	free	money	pay
15	스토리	story	character	end	novel
16	전략시뮬레이션	units	campaign	war	battle

17	시리즈있는 스토리	story	first	adventure	series
18	various(pass)	early	access	see	update
19	외계생명체	alien	caster	park	rebirth
20	레고	lego	star	war	movie
21	히어로전략	time	strategy	battle	heroes
22	가상현실	VR	experience	look	see
23	타워디펜스	tower	defense	worm	upgrade
24	호러스토리	story	honor	find	end
25	RPG	story	combat	character	RPG
26	퍼즐	puzzle	level	solve	time
27	various(pass)	buy	work	bad	long
28	핵전쟁	dark	die	souls	fallout
29	좀비	kill	dead	zombies	weapons
30	various(pass)	redd	touhou	michonne	cleanentine
31	비행기	flight	fly	place	aircraft
32	various(pass)	bad	fun	buy	free

유사도가 낮은 단어들로 이루어진 토픽(various)도 있었으나,  
대부분의 경우 해석이 가능한 것으로 보아 단어들이 토픽 별로 잘 분류되었다고 판단됨



# Modeling – Topic Modeling

## 2.3 게임을 구성하는 Topic 확인

	Document_No	Dominant_Topic	Topic_Perc_Contrib	Keywords	Text
0	0	4.0	0.4796	fun, shoot, weapons, enemies, level, shooter, ...	[favorite, radar, surprise, horizontally, scro...
1	1	12.0	0.2824	object, hide, story, find, puzzle, time, evil,...	[best, far, worst, solitaire, either, wait, sa...
2	2	2.0	0.4744	level, control, fun, time, jump, challenge, pl...	[simply, best, platformers, ive, ever, played,...
3	3	23.0	0.5004	story, find, end, time, short, look, horror, i...	[claire, d, psychological, survivalhorror, pul...
4	4	9.0	0.3549	level, card, dungeon, character, time, items, ...	[ive, run, meet, fantastic, people, awful, peo...

- 각각의 게임을 구성하는 **topic 중 가장 큰 비율을 차지 하는 topic**을 보여줌
- **Ex) 2번** 게임을 구성하는 토픽들 중 **2번 토픽의 비율이 가장 높음 (0.4744)**

# Modeling – Topic Modeling

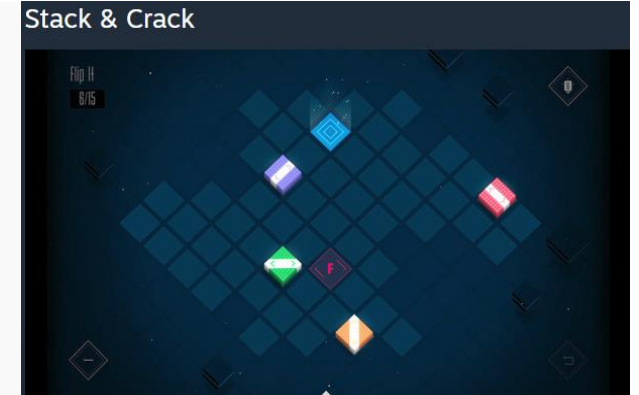
## 2.3 게임을 구성하는 Topic 확인

1	왕궁 RPG	warriors	dynasty	disgaea	sumurai
2	음악	fysuc	fun	relax	soundtrack
3	퀘스트	level	control	time	challenge
4	Laid(Boss)	boss	enemies	fight	attack
5	아케이드	hide	puzzle	time	click
6	타이쿤	build	time	work	city
7	GTA	police	mission	prison	gang
8	Car race	race	car	drive	track
9	전쟁	war	time	strategy	empire
10	던전	dungeon	character	items	quest
11	우주비행선	space	planet	crew	pirate
12	Team	version	online	multiplayer	PC
13	FPS	shoot	weapons	sglitter	gun
14	무료	pay	free	money	pay
15	스토리	story	character	end	novel
16	전략시뮬레이션	units	campaign	war	battle
17	시리즈있는 스토리	story	first	adventure	series
18	various(pass)	early	access	see	update
19	외계생명체	alien	caster	park	rebirth
20	레고	lego	star	war	movie
21	히어로전략	time	strategy	battle	heroes
22	가상현실	VR	experience	look	see
23	타워디펜스	tower	defense	worm	upgrade
24	호러스토리	story	honor	find	end
25	RPG	story	combat	character	RPG
26	퍼즐	puzzle	level	solve	time
27	various(pass)	buy	work	bad	long
28	핵전쟁	dark	die	souls	fallout
29	좀비	kill	dead	zombies	weapons
30	various(pass)	dredd	touhou	michonne	cleanentine
31	비행기	flight	fly	place	aircraft
32	various(pass)	bad	fun	buy	free

Topic 5(아케이드) 의 가중치가 높은 게임의 예시

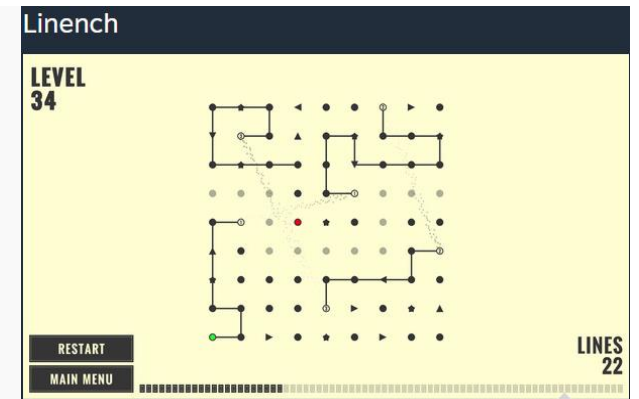
game id: 805420  
game name: Stack & Crack

(0, 0.23685883)  
(4, 0.62882656)  
(10, 0.02633132)  
(16, 0.05302065)  
(31, 0.042071257)



game id: 719090  
game name: Linench

(4, 0.903125)

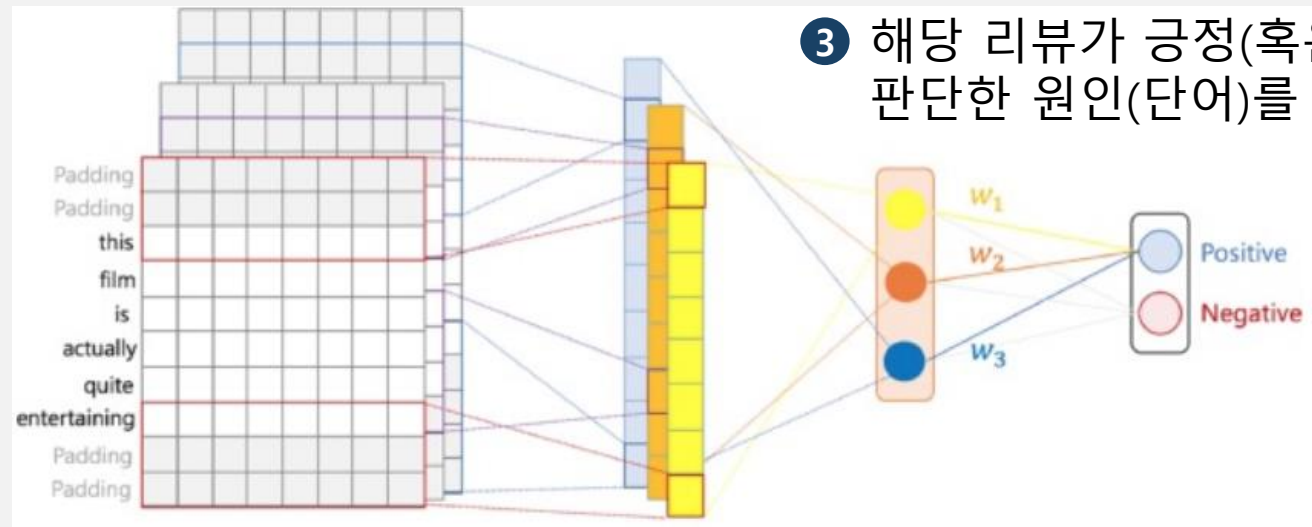


\* Python에서는 0부터 시작하므로 아케이드(5)의 index는 4

# Modeling – Topic Modeling

## 3. Grad CAM을 이용한 감성어 사전 구축

게임 도메인에 특화된 감성어 사전을 구축하고,  
게임이 특정 토픽에 대하여 긍정적인지 부정적인지를 판단하고자 함



- ① Word2vec(Skip-gram) 결과를 load하여 Sentence를 embedding (embedding dimension = 64 max sentence length = 128)

- ② Filter 사이즈 = 3, 4, 5  
Filter 개수 = 64

# Modeling – Topic Modeling

## 3. Grad CAM을 이용한 감성어 사전 구축

### Positive Words

abandoned	0.995536	3
abandonware	0.920624	3
abilities	0.744908	1
Ability	0.831384	2
abit	0.742373	1
able	0.917556	3
absolutely	0.704642	1
ABRAHAM	0.883032	2
⋮		

### Negative Words

controls	0.926128	3
unnecessary	1	3
frustrating	0.758256	1
worse	0.883309	2
the	0.887101	2
brand	0.706599	1
thoughtful	1	3
protagonist's	0.718552	1
boring	0.767905	1
unimaginative	0.865357	2
⋮		

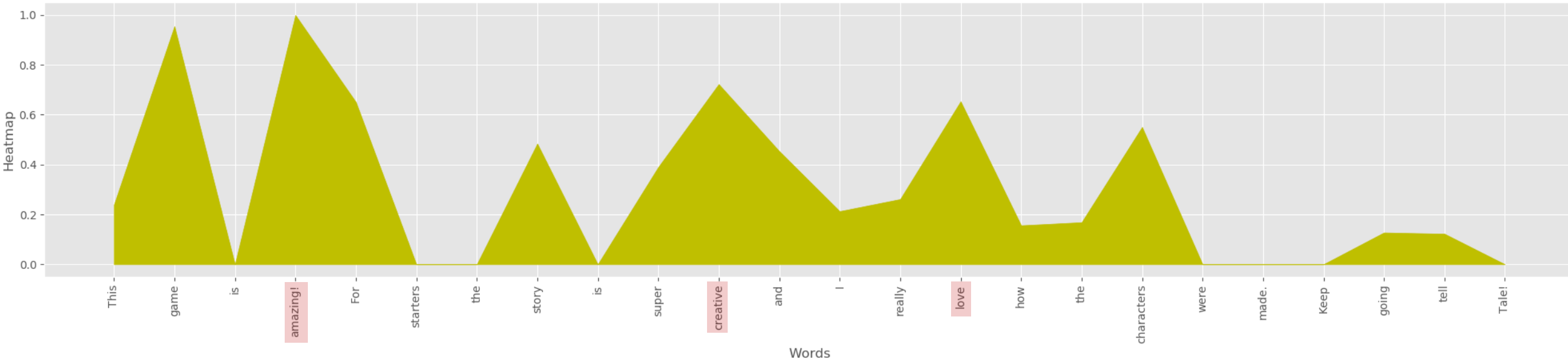
Grad-CAM을 통해 도출된 **filter**를 **ReLU function** 대입하여 상대적으로 높은 **filter**값을 도출하여 감정에 영향을 미치는 긍정 부정 단어를 모아 감성어 사전을 구축

# Modeling – Topic Modeling

## 3. Grad CAM을 이용한 감성어 사전 구축

### ㄱ. Positive review 예시

Heatmap of words in sentiment sentence



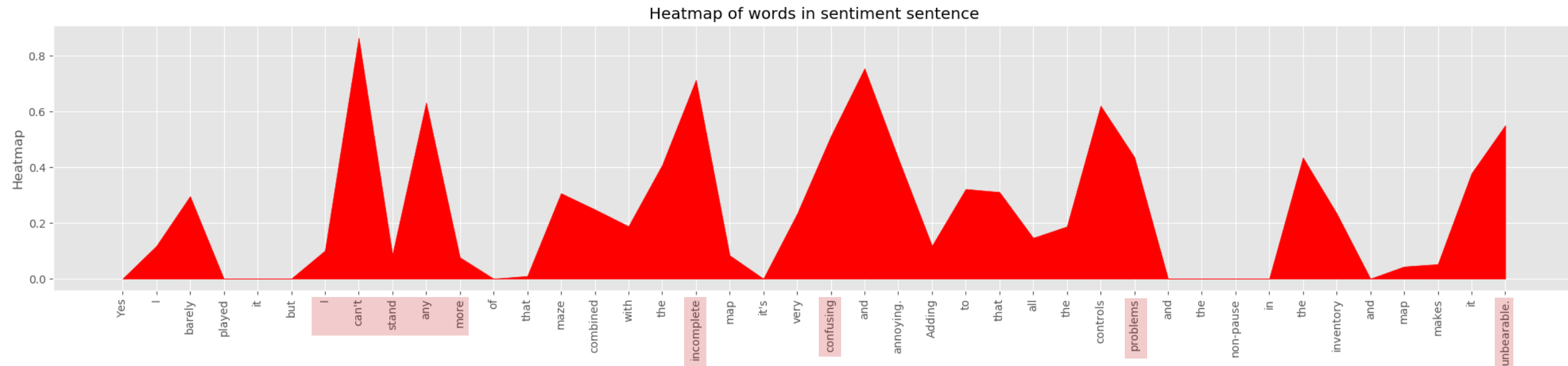
This **game** is **amazing!** For starters the story is super **creative** and I really **love** how the characters were mad keep going tell Tale!

Review를 Positive sentiment sentenc로 판단한 이유는 **game, amazing!, creative**

# Modeling – Topic Modeling

## 3. Grad CAM을 이용한 감성어 사전 구축

### ↳ Negative review 예시



Yes I barely played it but I **can't stand any more** of that maze combined the **incomplete** map it's very **confusing and annoying** adding to that all the **controls** problems and the non-pause in the inventory and map makes it unbearable.

**Review를 Negative sentiment sentence로 판단한 이유는 can't, incomplete, confusing, control problems**

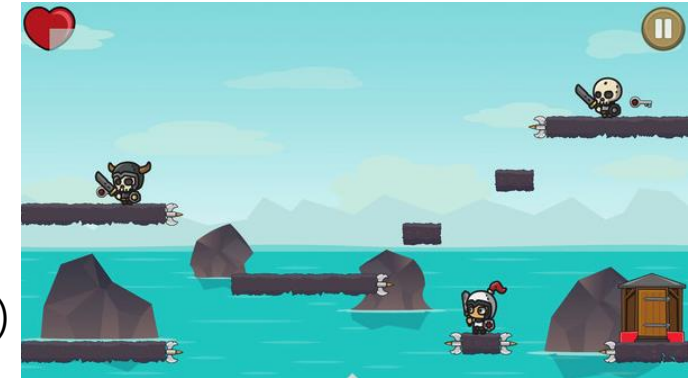
# Modeling – Topic Modeling

## 3. Grad CAM을 이용한 감성어 사전 구축

Game id: 760960

Game name: Only You

Topic: (8, 0.27585763), **(16, 0.5166728)**, (18, 0.11404606), (20, 0.084341414)



760960 게임을 가장 잘 설명하는 토픽  
(16번)에서 설명력이 높은 순위에 있는  
단어를 선택

선택한 단어와 함께 구성된 리뷰 속 단  
어들을 각각 감성어 사전과 비교하여  
긍정 단어, 부정 단어의 빈도를 확인

토픽(16번)속 단어의 긍정,부정 판단을  
통해 토픽의 긍,부정을 판단.

I think I had a stroke playing this game...seriously? No. GAVE ME A SEIZURE AND A BIG BIG  
(IF YA KNOW WHAT I MEAN) EITHER WAY THERE IS FOAM SPEWING OUT OF MY MOUTH AND  
I have no words for this game. Earthlings' words are unfit to describe this.  
Just beat lvl 1. Holy cow this game goes harder than crack cocaine.



## Conclusion





# Conclusion

- 게임 추천시스템의 양질을 높이기 위해 텍스트 데이터 뿐만 아니라 숫자 및 이미지 데이터를 크롤링하여 다양한 모델을 구축함
- 직관적인 이해를 돕기 위해 각 모델을 구축 후 시각화하여 나타냄
- 각각의 모델링한 결과를 비교해 봤을 때 전체적인 정확도는 높지 않은 것으로 나타남

## 개선방안

- 추후 개선사항으로 계속적인 **Grad-CAM**에서 사용되는 감성어사전의 업데이트가 필요하며, **LDA**에서 중복이 많이 되는 단어들의 기준치를 설정하여 제거하고 모델을 구축하여 질을 향상시키는 방향으로 진행할 예정임

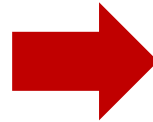
# Conclusion

## 추후 연구



게임 로고 이미지

장르 분류  
CNN



Action, Adventure

① Feature extraction ☒

② 게임의 total review를 사용한(Doc2vec) 게임 vector ☒

③ 게임 메타 정보 (categorical data) ☒

① ② ③ 을 결합한 input으로 Autoencoder model 학습, 게임 clustering ☐



## Q & A

