

# 토픽 모델링과 감정 분석을 이용한 게임 제품 리뷰 분석

정재윤

# 연구 배경 및 목적

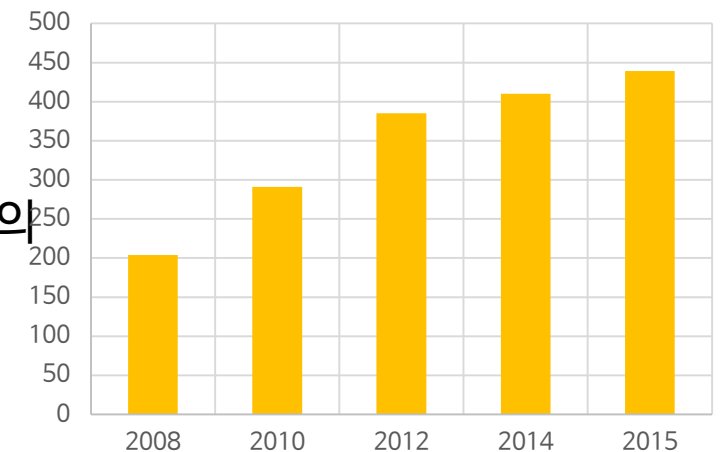
## ■ 게임 시장의 성장과 정보량 증가

- 연간 출시되는 게임의 수 급증
- 게임 시장 트렌드의 지속적인 변화(2016게임산업백서)
- 온라인 상에서 얻을 수 있는 게임 평가 정보의 다양화(리뷰,제품정보,사용자 리뷰 등)
- 게임 분야에서도 제품에 대한 소비자 의견 분석을 활용할 수 있을 것으로 기대됨

## ■ 연구 목적

- ① 각 게임 리뷰를 구성하는 주제를 정의하고 감정의 정도를 측정
  - 토픽 모델링을 통한 주제 정의
  - 감정 사전 구축을 통해 주제 별 감정의 정도 측정
  - 주제를 구성하는 단어에 감정 사전을 적용하여 주제 별 감정 변수 정의
- ② 감정 변수가 실제 리뷰에서 해당 주제의 내용을 반영하는 지 확인
  - 군집 분석을 통해 유사한 게임군으로 분류
  - 각 그룹에 속한 게임들의 특성을 추출하고 비교

연도별 게임 리뷰 제공 매체의 수



## 관련 연구

---

- 사용자 리뷰를 활용한 연구 사례

- 토픽 모델링을 통한 제품 리뷰 분석 연구

- 빅데이터 분석을 활용한 사용자 경험 평가 방법론 탐색 : 아마존 에코에 대한 온라인 리뷰 분석을 중심으로 (황해정, 2016)
    - 제품 특징화를 위한 오피니언 문서의 클러스터링 기법 (장재영, 2014)

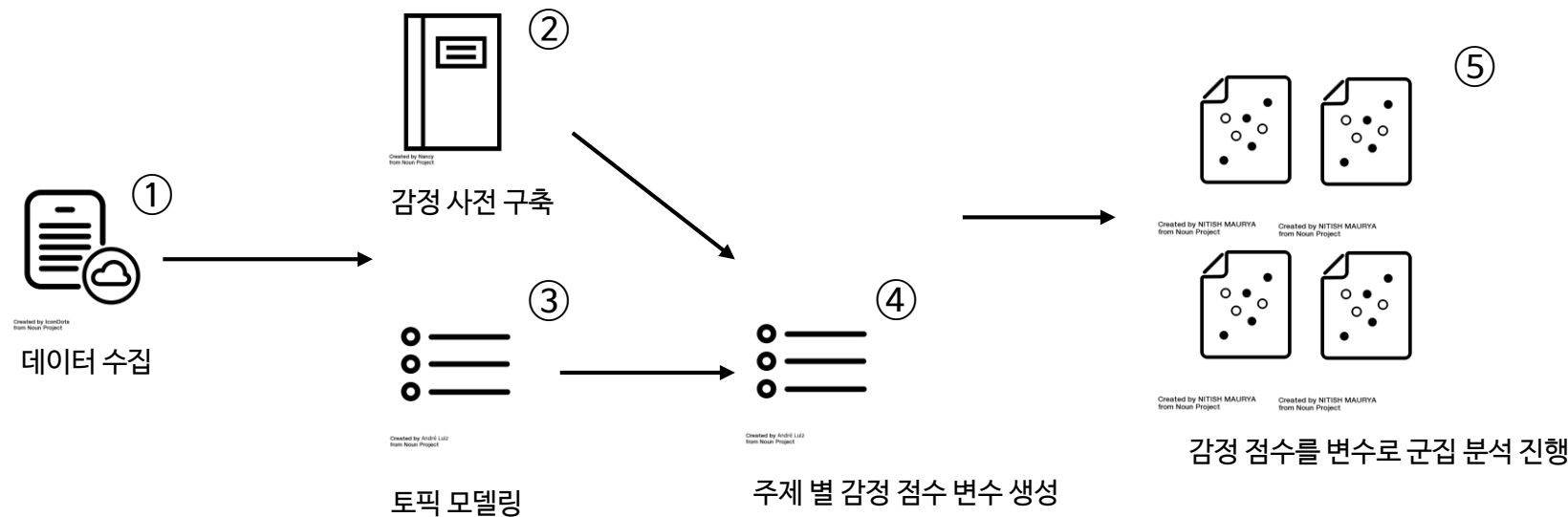
- 감정 사전을 이용한 제품 리뷰 분류

- 상품 리뷰 데이터와 감성 분석 처리 모델링 (연종흠, 2011)

- 게임 데이터를 이용한 제품 분석

- 확산모형과 군집분석을 이용한 게임제품의 흥행요소 분석 (송성민, 2015)

# 연구프레임워크



- ① Metacritic.com에서 972개 액션 게임의 총 33,302건의 리뷰 데이터 수집
- ② 사용자 리뷰 데이터와 게임 평점을 사용한 라쏘 회귀식을 통해 감정 사전 구축
- ③ LDA를 이용한 토픽 모델링 진행
- ④ 정의한 주제들에 감정 사전의 점수를 가중해 주제 별 감정 점수 계산
- ⑤ K-Means 군집화를 통해 군집 간 주제 점수들을 비교 분석

■ 데이터 수집

- 온라인상의 리뷰를 모아 평점을 제공하는 Metacritic.com의 사용자 리뷰를 수집
- 972개 액션 게임의 총 33,302건의 리뷰 데이터 수집

■ 데이터 설명

id	리뷰의 고유 ID 값
kind_of_games	게임 제목
user_name	작성자 ID
review	리뷰 내용
score	평점(10점 만점)
thumb_count	추천 개수



■ 감정 사전 구축

- Lasso 회귀식을 이용해 감정 단어(부사/ 형용사)와 리뷰 점수 간의 회귀식 구축
- 구축된 감정 사전(총 989개 단어)

■ 긍정 / 부정 단어 상위 20개 목록

feature	sen_score	feature	sen_score
econom	1.394457	unplay	-2.529247
awsom	1.185589	greedi	-2.251438
addict	1.111546	nrpa	-2.247656
underr	1.084285	rampant	-2.000497
hook	0.977442	hype	-1.965379
yea	0.972827	unaccept	-1.940549
favourit	0.952611	downhil	-1.866210
outstand	0.944892	paid	-1.864581
fabul	0.910705	worst	-1.721934
awesom	0.910133	micro	-1.614290
unforgett	0.901787	blatant	-1.562597
greatest	0.883936	corpor	-1.508074
genial	0.864944	rubbish	-1.477818
master	0.851198	garbag	-1.434880
med	0.846800	microsoft	-1.424845
skeptic	0.837857	incomplet	-1.419018
brilliant	0.828692	aw	-1.394840
finicki	0.828145	overhyp	-1.392558
intens	0.799160	fatal	-1.384069
favorit	0.786956	fake	-1.383281

긍정 단어

부정 단어

- 토픽 모델링
  - 전체 리뷰 데이터를 이용해 LDA를 이용한 토픽 정의
  - 총 9개의 토픽 정의

주제 1	구매	releas	money	player	review	buy	bug
주제 2	게임플레이	player	get	map	fun	class	lot
주제 3	레벨디자인	level	get	fun	puzzl	gameplay	control
주제 4	조작	pc	control	graphic	look	get	version
주제 5	캐릭터	charact	feel	gameplay	thing	look	voic
주제 6	입소문	get	peopl	fun	say	think	review
주제 7	그래픽	love	graphic	que	life	gameplay	valv
주제 8	멀티플레이	multiplay	graphic	fps	gameplay	player	fun
주제 9	AI	get	shoot	use	take	thing	ai

■ 감정 변수 생성

- 감정 사전을 통해 얻은 감정 단어의 점수와 토픽의 단어를 결합
  - 감정 단어가 등장한 문장의 앞, 뒤 3 단어를 보고 감정 단어 + 주제 단어 형태 구성
  - EX: This game has an awesome graphic! -> awesome graphic
  - 각 주제의 주제 단어들에 대해 주제 단어의 가중 \* 감정 점수의 합을 변수값으로 입력

- 게임의 각 주제들이 리뷰에서 어느 정도로 긍정적으로 받아들여지는지 알고자 함

feature	sen_score	feature	sen_score
econom	1.394457	unplay	-2.529247
awsom	1.185589	greedi	-2.251438
addict	1.111546	irpa	-2.247656
underr	1.084285	rampant	-2.000497
hook	0.977442	hype	-1.965379
yea	0.972827	unaccept	-1.940549
favourit	0.952611	downhil	-1.866210
outstand	0.944892	paid	-1.864581
fabul	0.910705	worst	-1.721934
awesom	0.910133	micro	-1.614290
unforgett	0.901787	blatant	-1.562597
greatest	0.883936	corpor	-1.508074
genial	0.864944	rubbish	-1.477818
master	0.851198	garbag	-1.434880
med	0.846800	microsoft	-1.424845
skeptic	0.837857	incomplet	-1.419018
brilliant	0.828692	aw	-1.394840
finicki	0.828145	overhyp	-1.392558
intens	0.799160	fatal	-1.384069
favorit	0.786956	fake	-1.383281

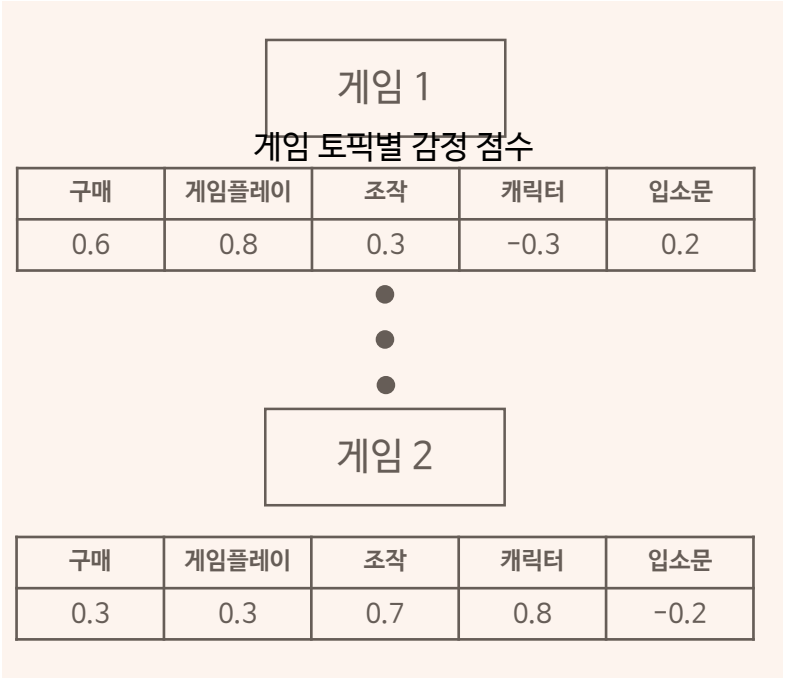
긍정 단어

부정 단어

+

구매	releas	money	player	review	buy	bug
게임플레이	player	get	map	fun	class	lot
레벨디자인	level	get	fun	puzzl	gameplay	control
조작	pc	control	graphic	look	get	version
캐릭터	charact	feel	gameplay	thing	look	voic
입소문	get	peopl	fun	say	think	review
그래픽	love	graphic	que	life	gameplay	valv
멀티플레이	multiplay	graphic	fps	gameplay	player	fun
AI	get	shoot	use	take	thing	ai

=





2.연구 방법		게임 제목		cluster	구매	게임플레이	레벨디자인	조작	캐릭터	입소문	그래픽	멀티플레이	AI	score	리뷰 본문
<div> <div></div> <div> <div></div> <div> <div></div> <div></div> </div> </div> </div> <div> <div></div> <div></div> </div>		BioShock Infinite: Burial at Sea - Episode Two	2		-0.1842	-0.203936718	-0.14356834	-0.2234	-0.1031	-0.3061	-0.0819	-0.204014426	-0.1313	6.88567	Extremely disappointed with the game. Graphics is nice, but game play is rigged with instant kills, and story is just terrible! I wish I never play the DLC at all. Because the story destroys my 80 hours I clocked to play original Bioshock Infinite. It is so bad. So I played it, finished it and I don't recommend it to anybody. Accept my opinion, or delete my review again from metacritics. I don't care.
		Dead Island: Riptide	4		0.02735	0.123908506	0.169150747	0.11838	0.08665	0.18939	0.16215	0.246634523	0.01138	7.01888	Fun addictive melee combat that, atleast from my experience in co-op, carries you through the passable graphics and weak story. Had no major glitches (only stuff like ragdoll spazzing which is pretty funny). Feels more fair than the last game where, especially in the city, you would constantly get attacked by many special zombies at a time. In my experience it is nothing but improvements and whilst the core mechanics are the same it has a lower price point to make up for this.

- “바이오쇼크 인피니트”는 리뷰에서 그래픽이 상대적으로 호평받았으나 입소문 면에서 부정적인 평가를 받음을 확인할 수 있음
- “데드 아일랜드”는 멀티플레이와 그래픽 부분에서 긍정적인 평가를 받았고 가격적인 면과 게임적인 면에서 크게 호평이나 비평을 받지 않았다는 점에서 주제 변수 수치와 어느정도 일치한다 볼 수 있음

- 감정 변수를 이용한 군집분석

- K-means 기법을 이용한 군집분석 결과 k = 5 일 때 최적의 결과로 판단
- 5개 군집에 대해 비교

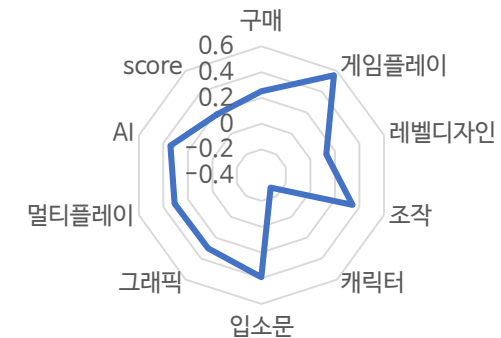
- 군집 1: 전형적 저평가 게임

- 각 주제 별 변수 수치가 낮고 평점이 가장 낮게 나오는 군집으로 전체 게임의 약 14% 차지
- 특히 AI 수치가 낮게 나오는데 대표적으로 팀 플레이가 요구되는 게임인 “Payday2”나 전체적인 평가가 나쁘게 나왔던 “콜 오브 듀티: 고스트” 등이 본 군집에 속해있음

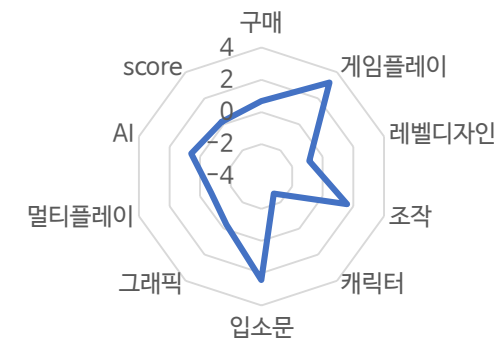
- 군집 2: 대기업 액션 게임

- “파 크라이”, “기어즈 오브 워”, “톰 클랜시의 고스트 리콘” 등 대규모 제작사의 프랜차이즈와 “노 맨즈 스카이”와 같은 유명 인디 제작사의 게임이 속한 군집으로 전체의 약 33% 차지
- 전체적인 주제 변수 수치가 가장 낮지만 평점은 중간으로 나옴, 이는 리뷰에서 부정적인 반응이 있더라도 제작사, 시리즈 등에 따라 기본적으로 얻을 수 있는 지지가 있다고 생각할 수 있음

전형적 저평가 게임

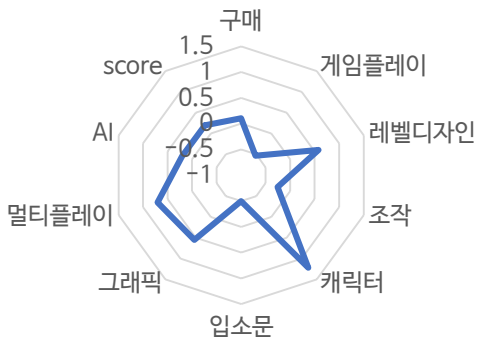


대기업 액션 게임

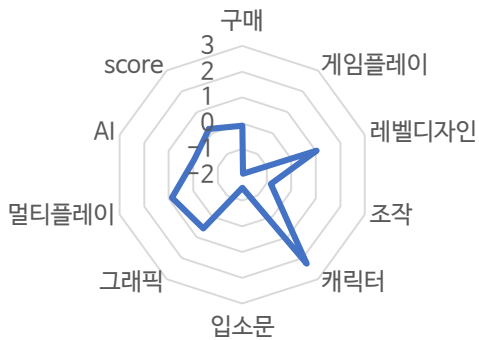


- 군집 3:대전 중심의 멀티플레이 위주 게임
  - 군집 4와 유사한 점수 분포를 보이지만 전반적으로 규모가 작은 제작사에서 제작한 게임이 다수 속해있으며 “스트리트 파이터”와 같은 대전 액션 게임들, “스토커” 시리즈, “F.E.A.R” 같은 비교적 작은 개발사에서 개발한 게임들이 속해있는 특징을 지님
- 군집 4: 고사양 그래픽의 멀티플레이 위주 게임
  - 전체적으로 높은 주제 변수 수치를 지녔으며 특히 그래픽과 멀티플레이에 대한 수치가 높은 군집으로 전체의 43%를 차지함
  - “배틀필드” 시리즈, “콜 오브 듀티” 시리즈, “카운터 스트라이크” 등 현세대를 대표할 수 있는 게임들이 다수 속해있으며 멀티플레이를 지원하는 FPS(1인칭 슈팅 게임)이 대다수를 구성함
- 군집 5: 과거의 명작 게임
  - 8개의 게임으로 구성된 작은 군집으로 “하프 라이프 2”, “워킹 데드” 등이 속해있음
  - 다른 군집들과 비교했을 때 중간 이상의 주제 변수 수치를 지니고 있으며 평점이 가장 높은 군집

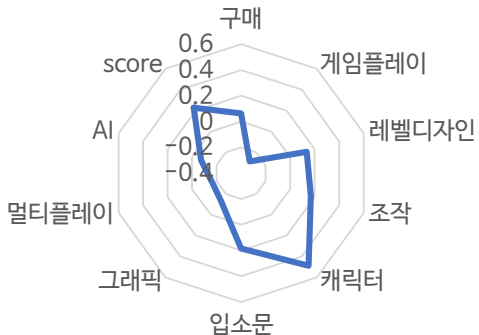
대전 중심의 멀티플레이 위주 게임



고사양 그래픽의 멀티플레이 위주 게임



과거의 명작 게임



### ■ 결론

- 본 연구에서는 리뷰 데이터를 이용해 게임 별 리뷰를 구성하는 주제들을 찾고 그 주제들이 어떻게 인식되는지 알아보고자 함
  - 토픽 모델링을 통해 게임 제품을 구성하는 주제를 정의하였고, 감정 단어들과 평점을 이용한 회귀식을 이용해 감정 사전을 구축함
  - 주제와 감정 사전을 이용해 주제 별 감정 변수를 만들었으며, 계산한 감정 변수가 실제 리뷰 내용을 어떻게 반영하는지 확인할 수 있었음
- 감정 변수들을 이용한 군집 분석을 통해 5개의 군집으로 분류하여 비교
  - AI 주제의 변수 수치가 가장 낮은 전형적 저평가 게임 군집의 게임들은 분대 플레이와 같이 팀 플레이가 요구되는 게임이 많았으며 AI 봇과 플레이하는 오프라인 모드에서의 부정적인 평가를 보임
  - 가장 많은 게임이 고사양 그래픽을 보여주는 멀티플레이 위주 게임 군집에 속하고 있으며 주제 변수 수치 또한 해당 군집이 전반적으로 높은 수치를 보였음
  - 대기업 액션 게임 군집에 속한 게임들은 전세계 유명 제작사에서 제작된 고비용, 프랜차이즈 게임들이 대다수이며 상대적으로 낮은 주제 변수 수치를 보이지만 평균적인 평점을 보여 기본적으로 고려되는 평점의 정도가 다른 군집에 비해 높다고 생각됨