

Music Recommendation Based on Detailed Topic Modeling

19. 06. 21 8조 정승섭, 최민서, 민다빈, 조용원

목치



- I. 문제 정의
- II. 중간 보고 Comment 정리
- III. Term Project 수행 내용 요약
- IV. Term Project 수행 내용
- V. 결론 및 제언

l. 문제 정의

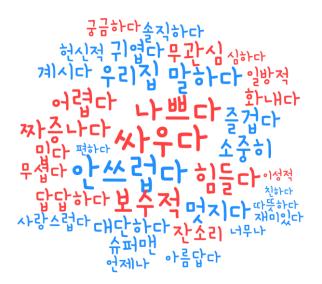


1. 현 음악 추천 시스템에 활용되는 데이터는 메타데이터 위주

- 가사 정보가 사용되지 않거나, **단어수준의 정보**만을 이용함
- 추천시스템의 성능 향상을 위해서 가사의 주제는 매우 **유용한 정보**가 될 수 있음

MP3 File			
1/25: 01 - Marc	he de Trion	nphe et Second Air de Tror	npett
MPEG 1, L Bitrate: ~25 Freq: 441	2 kb/s	Mode: Joint stereo Size: 4.9 MB (124.0 N Time: 2:30 (1:08:37)	1B)
ID3 Tag			
Common Pick	tures (1)		
Title:	Marche de	Triomphe et Second Air d	e T 🗆
Artist:	Musica An	tiqua Köln	
Album Artist:	Musica An	tiqua Köln	0
Album:	Charpenti	er: Musique sacré 🗆 CD: 🛚	1/1
Year:	2003 🔾 🖲	Track #: 01 - / 🗷	26
Genre:	Classical		. 0
Comment:	Amazon.c	om Song ID: 209367230	0
Composer:	Charpenti	er, Marc-Antoine	0
Orig. Artist:			0
Copyright:	(C) 2003 D	eutsche Grammophon Gm	bH □
URL:			0
Encoded by:			0

〈음악 메타데이터〉



〈감정단어사전〉

I. 문제 정의



2. 기존 연구 고도화

- 가사에서 곡의 주제를 도출하였던 기존 연구('15년 수행 과제)는 전체 노래를 오직 **5개의 주제(**자연, 스타일, 클럽 댄스, 간 접적 이별, 직접적 이별)로 분석하였음
- 주제의 범위가 매우 크기 때문에 이를 활용한 추천시스템은 충분히 개인화 될 수 없음
 - → 대량의 데이터를 이용하여 주제의 세분화 필요

바보처럼 **사랑** 때문에 떠난 너 때문에 울지 **않을래** 더 멋진 남잘 만나 꼭 **보여줄게** 너보다 행복한 나

- 에일리, 보여줄게

So, you don't cry for me 세월 지나도 난 변하지 않아 And the I cry for you 이 밤 지나면 이젠 안녕 영원히 널 사랑해 널 사랑해 영원히

- 더 크로스, Don't Cry

좋으니 사랑해서 사랑을 시작할 때 네가 얼마나 예쁜지 모르지 그 모습을 아직도 못 잊어 헤어 나오지 못해 네 소식 들 날은 더 좋으니 그 사람 솔직히 견디기 버거워 너도 조금 더 힘들면 좋겠어 진짜 조금 내 십 분의 일 만이라도 아프다 행복해줘

- 윤종신, 좋니

다 같은 '이별' 노래?

목치



- l. 문제 정의
- II. 중간 보고 Comment 정리
- III. Term Project 수행 내용 요약
- IV. Term Project 수행 내용
- V. 결론 및 제언

II. 중간보고 Comment 요약



1. 최적 토픽 수 탐색

감성 분석을 할 때 세분화된 감성을 명확하게 드러낼 필요

추천 토픽 수 탐색 과정 같은 것 필요

가사의 내용이 유사한 곡을 잘 묶는 방법도 고민 필요

2. LDA 효과

"영어" 곡만 남기는 태스크가 중요 -> 정말로 <mark>영어</mark>만 남아지는가?

여러 번 LDA? Hierarchical LDA 사용?

LDA->LDA 가 정말 감성을 세분화 하나?

3. 추천 시스템의 만족도 / 평가 방법

가사로만 만족감을 주는 추천이 이루어지는 것이 잘 될지는 의문

성능 평가가 직관적으로 전달될 수 있는 방법?

만족도나 정확도에 대한 평가를 어떻게 할 수 있을지?

해결 방안

IV. Term Project 수행 과정에서 설명

(LDA 결과를 통해 1&2번 설명)

(LDA-)LDA vs. Hierarchical LDA)

해결 방안

- 1.설문지를 통한 평가 진행
- 2.시각화 방법을 이용해 가사 기반 추천 결과 설명

목차

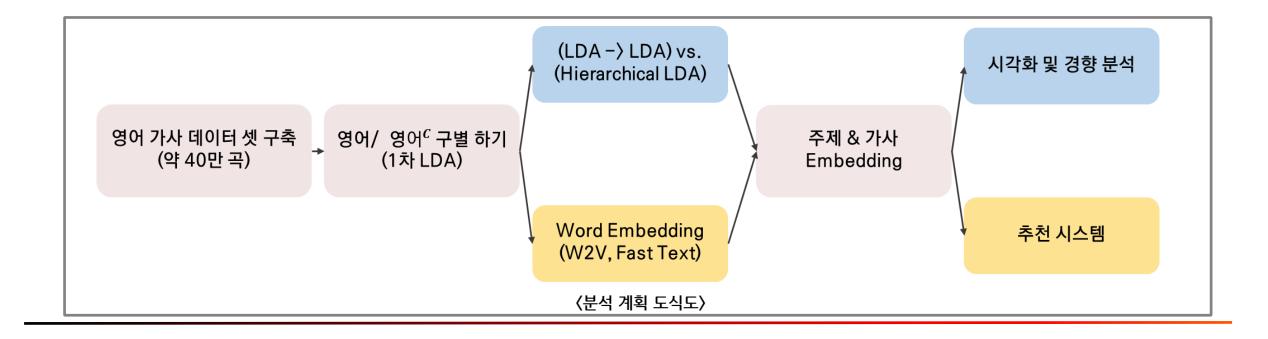


- l. 문제 정의
- II. 중간 보고 Comment 정리
- III. Term Project 수행 내용 요약
- IV. Term Project 수행 내용
- V. 결론 및 제언

III. Term Project 수행 내용 요약



- 1. 약 40만곡의 영어 가사와 메타데이터 수집 및 전처리(Tokenizing, Stop-words&특수문자 제거), Lemmatization 수행
- 2. 영어와 영어가 아닌 가사를 구별하기 위해 1차 LDA 수행
- 3. (1)(2) 영어 가사를 추출 후, (LDA & HLDA & Word Embedding) 수행
- 4. 주제와 가사를 한 공간에 Embedding
- 5. 결과 시각화 및 경향 분석과 추천 시스템 결과 해석



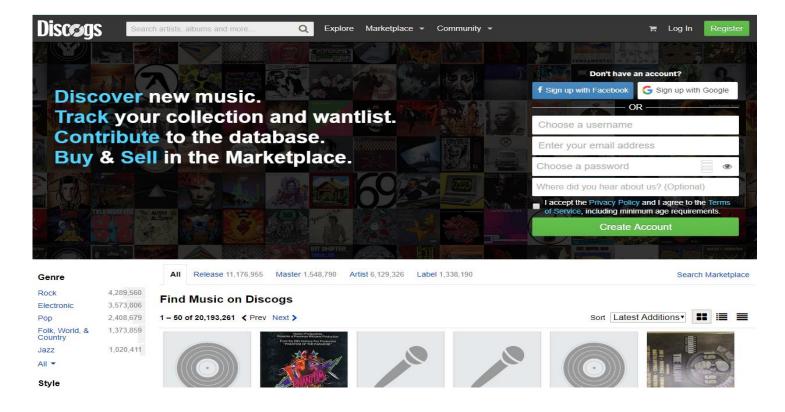
목차



- l. 문제 정의
- II. 중간 보고 Comment 정리
- III. Term Project 수행 내용 요약
- IV. Term Project 수행 내용
- V. 결론 및 제언



- 1. 곡명, 아티스트 명, 곡의 발매 년도
 - discogs.com의 오픈 데이터베이스를 이용하여 음악 메타데이터 수집(1905~2015 전 세계 음악)
 - 가사 검색을 위해 곡명, 아티스트 명을 추출



데이터 출처 : www.discogs.com 데이터 핸들링 : SQLite on Python



- 1. 곡명, 아티스트 명, 곡의 발매 년도
 - 추출한 메타데이터를 이용해 LyricsWikia 에서 가사를 크롤링
 - 약 40만곡의 가사 데이터를 수집 완료

	408489 Ja Ru	le Holla Hol	la (Remix) 1	999 [(Ja Ri	ule) It's murda! (Jay-	-Z) Yeah yeah, Hova Hova We takin' over soldier, told ya it's murdaa I'm here for that paper playa, fuck one time I'm here ta break ya playa one nine Make ya scream a
	408490 Unter	ART Now Or N	Vever 2	008 [Kill m	ne now or never Kill	me now, kill me now This is more than I ever had Kill me now, kill me now I wanna keep it when I'm dead This is devotion not surrender Kill me now, kill me now I b
	408491 Unter	ART Open End	d 2	008 [You o	can't touch glee The	heart's receiving The world is stuffed with modern cries Wasted grieving and believing That's coming down the day I die, Because your fool plan, desperation, And at
	408492 Unter	ART Novalis P	art I (Der I 2	008 [<i>S</i>	anft wird das Ende N	Muss immer der Morgen wieder kommen?)
	408493 Unter	ART The Anta	gonist 2	008 [I look	k into the mirror and	d the wall is still black A stranger turns up with the blame on his back With the bleeding wrists and a crucified 'tat He goes to sleep and never wakes up again But I, I'
	408494 Unter	ART Exit	2	008 [Can y	you see me? I will ta	ake a part of you. Our love is gone. I face reality. I feel that thorn of a rose too dry to bleed. I stare at my trust in your eyes. Let you take me to paradise. And free my s
	408495 Unter	ART Pariah	2	008 [If you	u ever tried to cover	and hide your face If you ever searched for someone of an equal race If you ever gave up the hope to leave your past and cross the line If you ever felt the burden of
	408496 Unter	ART Memento	2	008 [I cut	my name in your he	eart To assure the memories remain I never will release the pain. I cut my name in your heart And I know you never will forget My love will be your deepest bleeding se
	408497 Unter	ART Novalis P	art III (Keir 2	V <i>)] 800</i>	Vo keine G철tter sind	d - walten Gespenster. Oh dass der Geist der Geister euch erf체llte!]
	408498 Unter	ART Perfect W	orld 2	008 [What	t a perfect world Wh	nat a perfect world So disgustingly perfect What a perfect world ???? The end is near we see through this No need to be afraid When we hear the call of the black abys
	408499 Unter	ART Atrophied	1 2	008 [(?)	found my place on	earth No need to take, no need to kill No need to leave The apple is rotten and decayed The garden, forgotten and betrayed No need to fight, no need for fear No n
	408500 Tina	TurneRock Me	Baby 1	993 [Rock	me baby rock me al	Il night long I said I want you to Rock me baby rock me all night long I want you to rock me like my back ain't got no bones Roll me daddy like you roll your flour de
	408501 Tina	Turne (Darlin') Y	ou Know 1	993 [I kno	w you love me baby	y But you never tell me so I know you love me baby But you never tell me so If you don't tell me that you love me I'm gonna pack my bags and go I live across the str
	408502 Biffy	Clyro Umbrella	(BBC Rad 2	008 [You l	had my heart and we	e'll never be worlds apart Maybe in magazines, but you'll still be my star Baby cause in the dark you can see shiny cars And that's when you need me there, with you I'l
	408503 Shae	Jone Talk Show	v Shhh! (C 1	998 [Who	da hell you think yo	ou messin' wit Tryin' to play me like a fool I overheard a conversation Between your ex-girlfriend and you Then I saw it so clearly Suddenly it all made since Before you
	408504 Liars	Danc The Wide	Blue Yon 1	996 []	The Wide Blue Yond	ler Close your eyes and see the land, Where all is well and all is sane. No rocks been moved by human hands, In which the Ruler has no name. 'Cause everybody
40	0540					t now.
40	8512 L	iars Da	Sign /	∖nd	1996	[(b)Sign and Faults(/b) Leaf after leaf, softly sailing to the ground··· not per
	בומוס	- and and - 10.	J			into rea
	408508 Liars	Danc He Sang	In His Batl 1	996 [l	He Sang In His Bath	Before He Died Neighbor John is really sad, because his fucking stinky dog is dead! His shitty cur run through my yard and buried bones in the ground. And no
\	408509 Liars	Danc Always Be	een A Failu 1	996 [/	Always Been A Failur	re Last at night, I'm starring at the ceiling And all my memories pass by. And now it's so conspicuous quiet, I hear a voice somewhere in the night. A cold wind bl
\setminus	408510 Liars	Danc Into The I				ight my day there must be something wrong. Searching for some exit to be free. Have to pray to waste my time along. See the moonlight far behind the tree. What if
$ \cdot $	408511 Liars	Danc Genitals E	Beat Intelle 1	996 [0	Genitals Beat Intelled	ct (explicit) (Man) Look at it my child and tell me what do you think of it. People always talk in riddles, you know, what I mean. Though I think it's high time that

1996 (
b> Sign and Faults Leaf after leaf, softly sailing to the ground. Side by side, dancing in the fog two clowns. But I know it's true, This fall reveals, What I felt for you. Step by step, memo

데이터 출처 : LyricsWikia(www.lyrics.wikia.com)

408512 Liars Dand Sign And Faults

크롤링 툴: BeautifulSoup on Python



2. 사전 LDA 수행

- 영어와 영어 c (영어가 아닌 가사)를 제거하기 위한 LDA 진행
- Topic별 Top20 단어와 해당하는 곡들의 가사를 보고 영어가 아닌 가사를 제거

Topic4			
da	boom		
ah	wa		
dem	fi		
ha	ho		
di	nah		
doo	WO		
mi	ga		
ba	gal		
ni	ay		
yuh	nuh		

Topic7			
nie	je		
na	ja		
que	jest		
si	mi		
eu	za		
se	mnie		
jak	te		
da	VOC		
um	pra		
tak	meu		

Topic8			
la	tu		
le	cest		
et	pour		
je	en		
les	du		
que	ne		
pas	est		
des	mon		
qui	au		
dans	tout		

Topic9		
la	en	
que	il	
che	lo	
non	si	
el	una	
mi	ti	
di	es	
te	yo	
tu	por	
se	le	

Topic17			
la	et		
ja	vaan		
ei	mut		
se	nyt		
en	mit		
kun	ku		
niin	vain		
sen	kuin		
jos	ne		
mun	sun		

영어^C

0 0 0 0



2. 사전 LDA 수행

- 영어와 영어 c (영어가 아닌 가사)를 제거하기 위한 LDA 진행
- 해당 과정을 통해 LDA가 유사한 가사들이 군집(Topic) 이루는 것을 확인 가능
- 수행 전 가사의 수 : 408,513곡에 대한 가사 -> 수행 후 가사의 수 : 326,609곡에 대한 가사

Topic1			
nt	long		
II	said		
home	know		
like	come		
day	way		
time	gone		
night	town		
old	oh		
ve	song		
away	little		

Topic3			
life	pain		
death	die		
dead	eyes		
blood	nt		
world	soul		
like	end		
mind	human		
time	fear		
kill	ii ii		
hate	face		

Topic11			
night	rain		
like	fly		
light	П		
eyes	dark		
sun	love		
sky	moon		
away	world		
come	dream		
heart	life		
time	wind		

Topic18			
let			
like			
make			
come			
time			
ve			
way			
girl			
hold			
say			

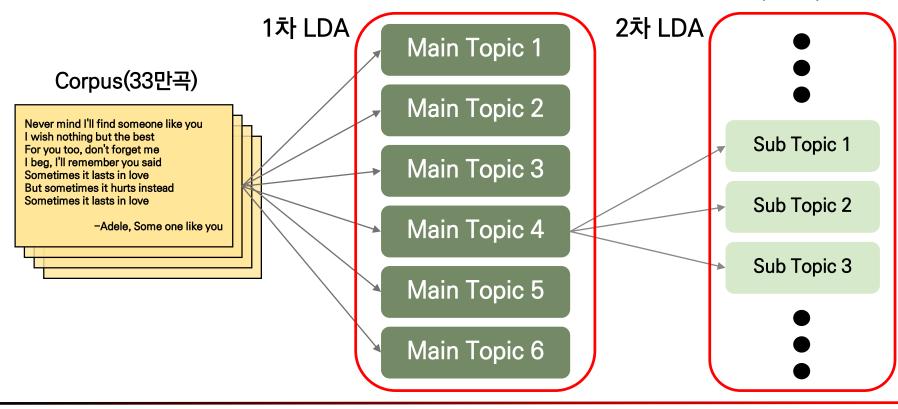
Topic20			
like	girls		
nt	cause		
got	ve		
man	look		
	make		
big	good		
little	right		
said	think		
money	want		
know	girl		

영어

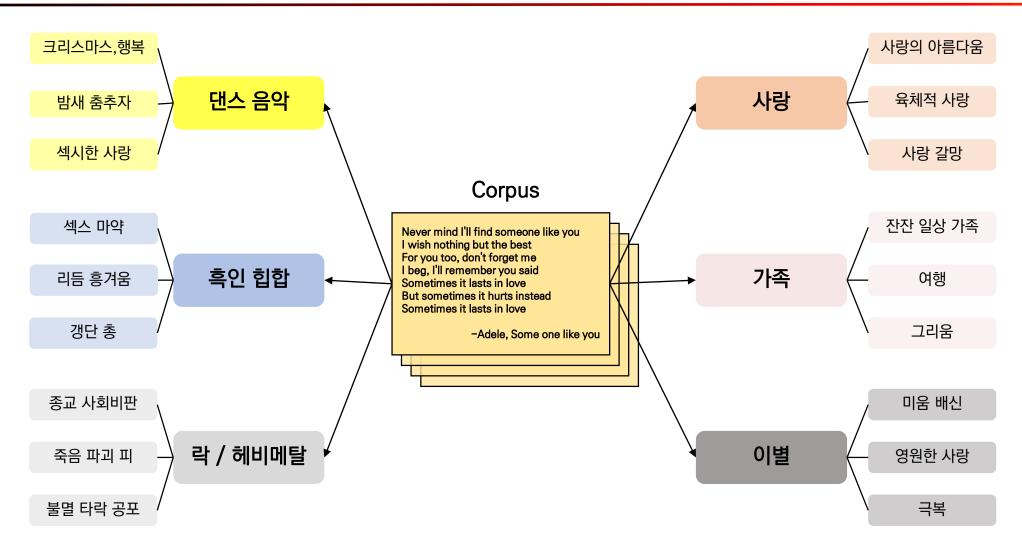
0 0 0 0



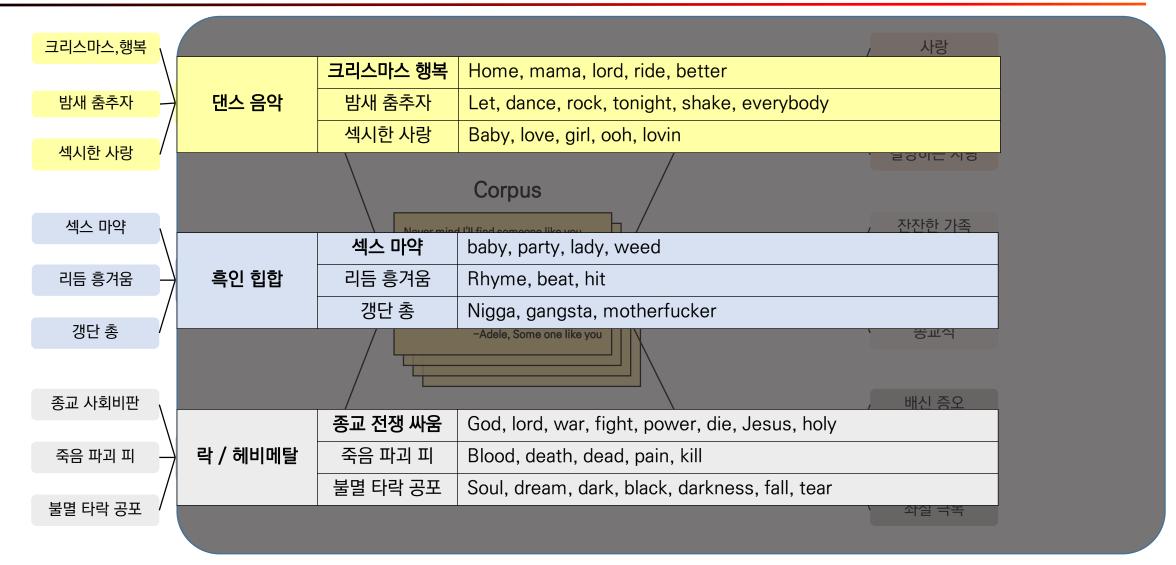
- 3. (1) (LDA -> LDA) vs. (Hierarchical LDA)
 - 메인 토픽과 서브 토픽을 도출하기 위해 계층적으로 LDA를 수행
 - Hyperparameter : 메인 토픽 수 6개, 메인 토픽 당 서브 토픽 수 3개, 반복 수 1000회
 - 다양한 토픽 수 조합에 대해 수행 결과 정성적으로 최적 토픽 수를 선택하였음. (Perplexity 측정 시 30, 9가 최적)









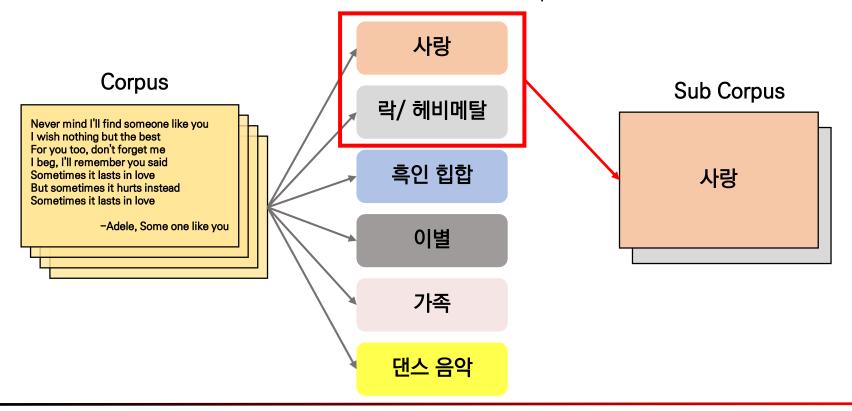




크리스마스,행복 (사랑의 아름다움
	사랑의이름다움	Sun, shine, fly, fall, sea		710-1 11-11
바 사랑	육체적 사랑	Baby, fell, sweet, lover, inside, touch	-	육체적 사랑
	사랑 갈망	Lonely, forever, heaven, sleep, dream		
섹시인 시당 '		Corpus	`	사랑 갈망
섹스와 여자 ,	TITI 0111 71 T	Never mind I'll find company like you		잔잔 일상 가족
	잔잔 일상 가족	Girl, baby, boy, people, kid, Christmas		
마의 가족	여행	Street, sky, round, way, rain, walk	\vdash	여행
	그리움	Said, song, gone		
흑인 갱년 / 죽음 파괴 피 ,		-Adele, Some one like you		그리움
40.44.4	미움 배신	Lie, wrong, believe		이스 메인
불멸 이별 / 슬픔	영원한 사랑	Love, want, gone, stay		영원한 사랑
	극복	Away, run, turn, walk		
종교,선생,싸움 /			\	극복

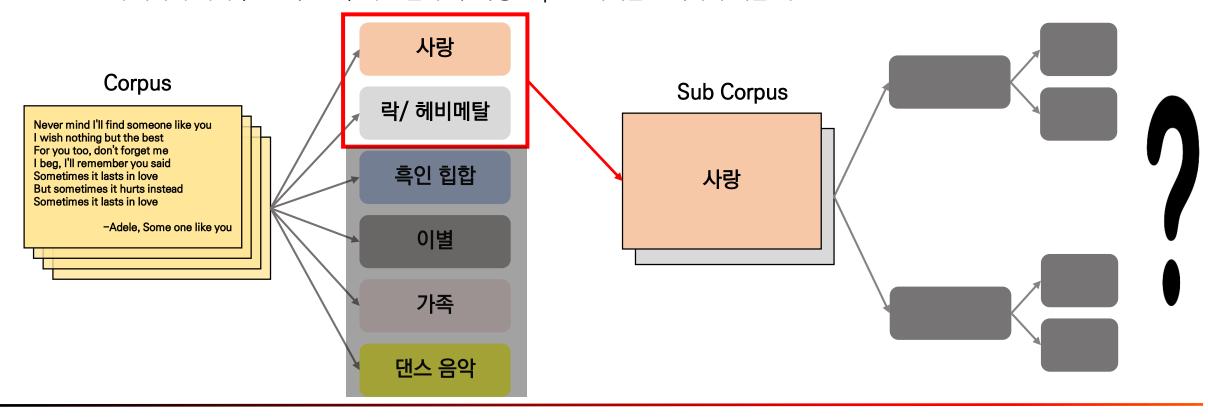


- 3. (1) (LDA -> LDA) vs. (Hierarchical LDA)
 - 개인 컴퓨터: RAM 32GB/ 서버 컴퓨터: RAM 78GB, 현재 가사의 수: 326,609곡
 - 두 컴퓨터 모두 Memory Error 발생
 - 전체 가사에 대해 (LDA->LDA) 비교 불가 -> 특정 Topic 2가지를 선택하여 이를 비교



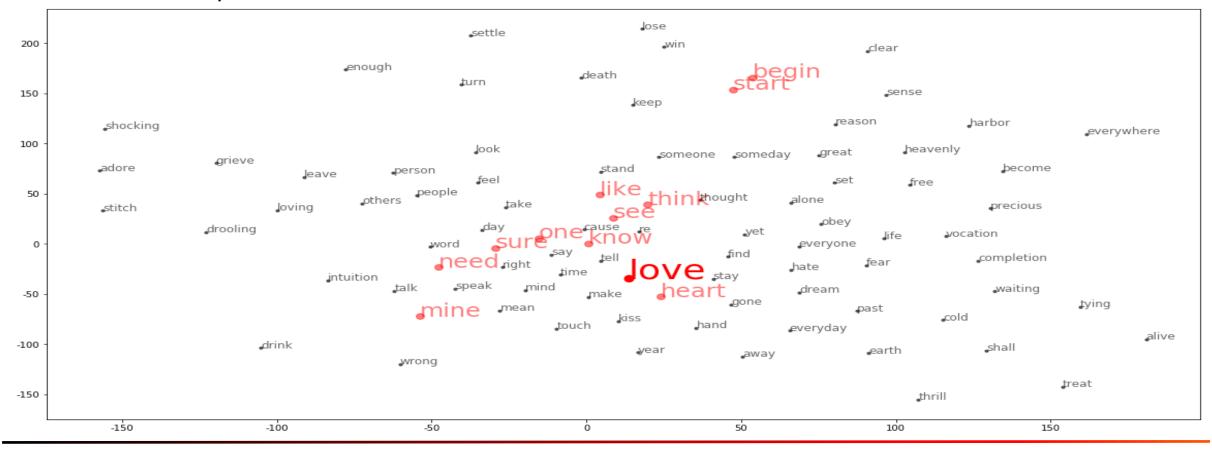


- 3. (1) (LDA -> LDA) vs. (Hierarchical LDA)
 - 개인 컴퓨터: RAM 32GB/ 서버 컴퓨터: RAM 78GB, 현재 가사의 수: 326,609곡
 - 두 컴퓨터 모두 Memory Error 발생
 - 전체 가사에 대해 (LDA->LDA) 비교 불가 -> 특정 Topic 2가지를 선택하여 이를 비교





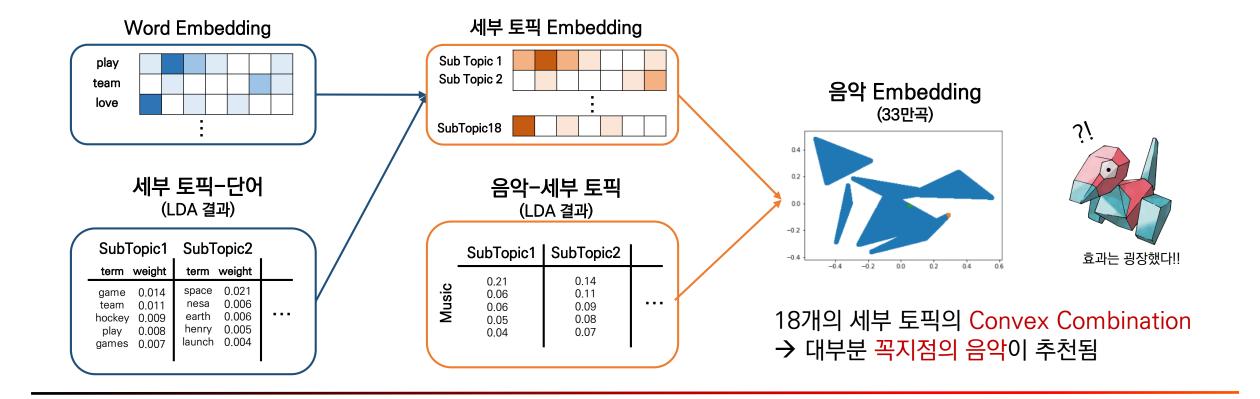
- 3. (2) Word Embedding (W2V & FastText)
 - Word2Vec 과 FastText에 대한 비교를 위해 각각 Embedding 진행(Wrod2Vec약 2.5 시간, FastText 3.5 시간 소요)
 - 문서 집합(Corpus)에 없는 단어에 대해서도 유사도 추출할 수 있는 장점을 가진 FastText 선택





4. 주제 & 가사 Embedding

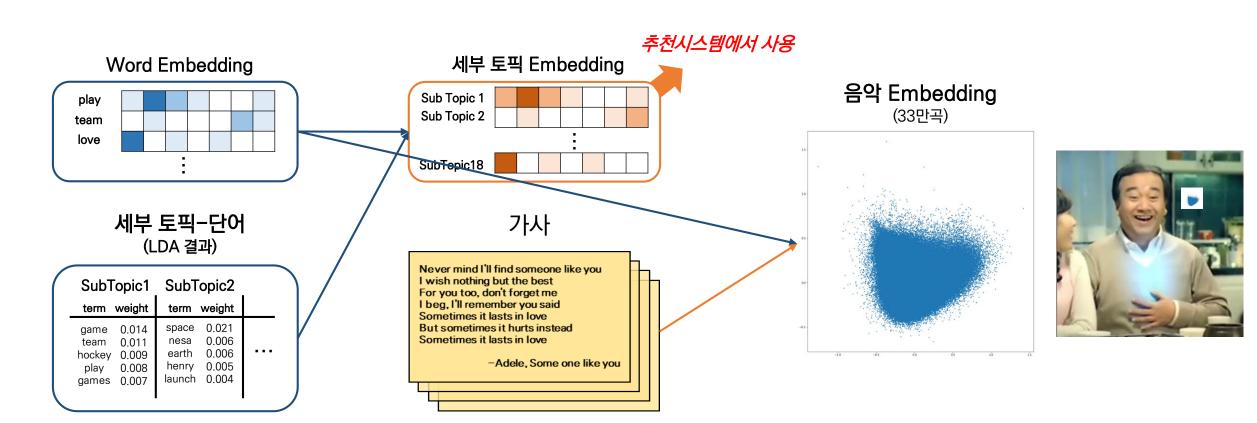
- 세부 토픽 Embedding: 토픽-단어 행렬(LDA결과)으로 Word Embedding 가중 합
- 음악 Embedding : 음악-토픽 행렬(LDA결과)으로 세부 토픽 Embedding 가중 합





4. 주제 & 가사 Embedding

- 세부 토픽 Embedding: 토픽-단어 행렬(LDA결과)으로 Word Embedding 가중 합
- 음악 Embedding : 가사 등장 어휘에 대한 워드 Embedding 평균 (특정 어휘 중복 등장 무시)

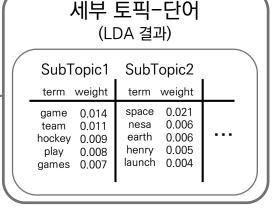


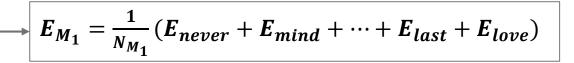


4. 주제 & 가사 Embedding

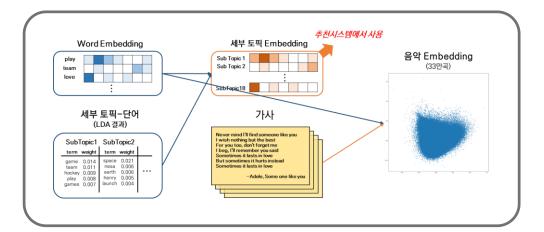
- 세부 토픽 Embedding: 토픽-단어 행렬(LDA결과)으로 Word Embedding 가중 합
- 음악 Embedding: 가사 등장 어휘에 대한 워드 Embedding 평균 (특정 어휘 중복 등장 무시)









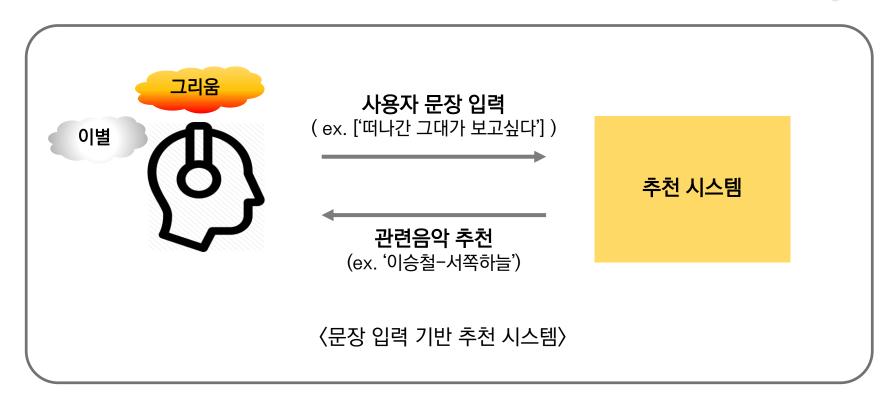


$$E_{SubTopic1} = 0.014 * E_{game} + 0.011 * E_{team} + 0.009 * E_{hockey} + \cdots$$



5. (1) 추천 시스템

- 사용자 문장 입력 시 적절한 추천 음악을 제공하는 추천 시스템 구현
- 사용자의 입력 문장에는 <mark>내재된 주제</mark>가 있다고 가정
- 입력 문장의 주제에 해당되는 음악만을 추천 후보 군으로 설정하고 이를 기반으로 입력 문장의 Embedding과 유사한 음악을 추천





5. (1) 추천 시스템

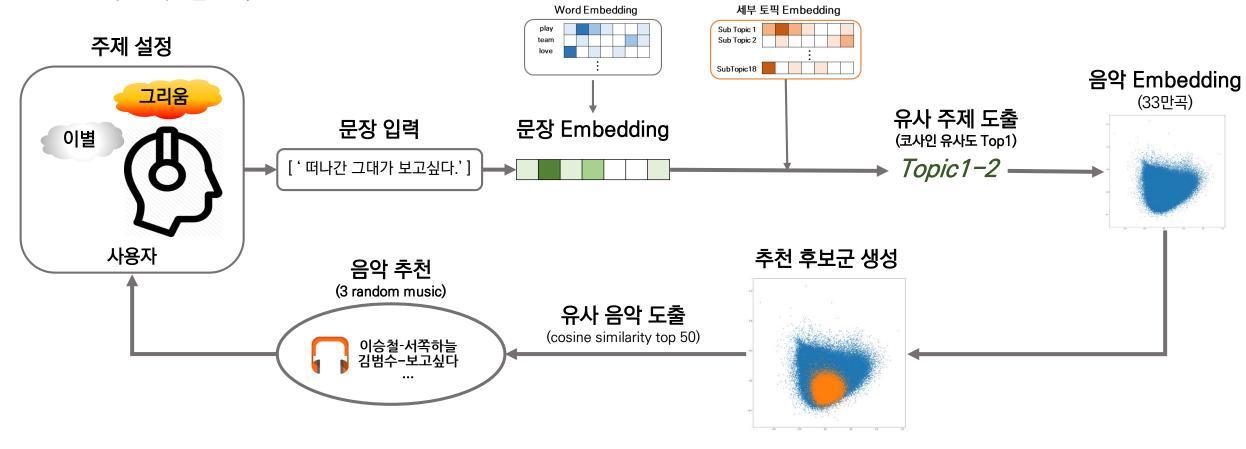
- 사용자 문장 입력 시 적절한 추천 음악을 제공하는 추천 시스템 구현
- 사용자의 입력 문장에는 내재된 주제가 있다고 가정
- 입력 문장의 주제에 해당되는 음악만을 추천 후보 군으로 설정하고 이를 기반으로 입력 문장의 Embedding과 유사한 음악을 추천





5. (1) 추천 시스템

- 추천 시스템 도식도





5. (1) 추천 시스템 실험 결과

- 입력 문장의 주제와 어울리는 음악이 추천됨
- 하나의 메인 주제에서 서로 다른 서브 주제를 잘 구분하여 추천이 이루어짐

입력 문장	"You lied to me. You said you will love me forever."						
주제 분석	메인 토픽: 1. 슬픔 / 서브 토픽: 1. 미움, 배신						
	아티스트	제목	가사(일부)				
추천 곡 1	Johnny Horton	The Mansion You Stole	You said it was love dear but you lied from the start				
추천 곡 2	Seagull Screaming Kiss Her Kiss Her	Think It Over	I'll forget everything for you I know you loved me once But it is over It is over.				
추천 곡 3	Naked Eyes	Promises, Promises	If this is love it breaks my heart You made me promises, You knew you'd never keep				
입력 문장	"You left me. But I'll be waiting for you."						
주제 분석	메인 토픽: 1. 슬픔 / 서브 토픽: 2. 영원한 사랑						
	아티스트	제목	가사(일부)				
추천 곡 1	Not Available	Led By Signs	Here I sit and I am waiting If you'll come				
추천 곡 2	Ringo Starr	Waiting	Till you find your way back to me, I'll be waiting for you.				
추천 곡 3	The Perishers	Still Here	When you've gone I'm still here				

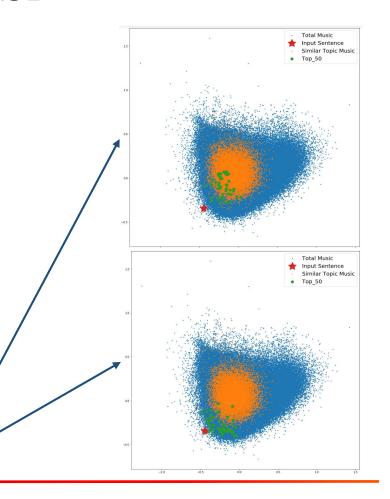


5. (1) 추천 시스템 실험 결과

- 주제를 고려하여 음악을 추천함으로써 입력 문장의 주제에 보다 의미상 근접한 음악을 제공함

입력 문장	"You left me. But I'll be waiting for you."					
주제 분석	메인 토픽: 1. 슬픔 / 서브 토픽: 2. 영원한 사랑					
	아티스트	제목	가사(일부)			
주제 고려 추천 곡1	Not Available	Led By Signs	Here I sit and I am waiting If you'll come			
주제 미 고려 추천 곡2	Feargal Sharkey	Love And Hate	Love and hate. Why should he wait for hours while she is making up her mind			

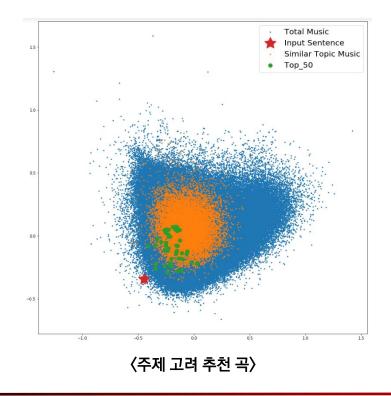
입력 문장	"I miss my home."					
주제 분석	메인 토픽: 3. 가족 / 서브 토픽: 2. 그리움					
	아티스트	제목	가사(일부)			
주제 고려 추천 곡1	Peter & Gordon	Five Hundred Miles	I'm 500 miles from my home 500 miles 500 miles 500 miles 500 miles Lord I'm five hundred miles from my home			
주제 미 고려 추천 곡2	The Cloud Room	O My Love	O my love, O won't you come on home, we miss you so much			

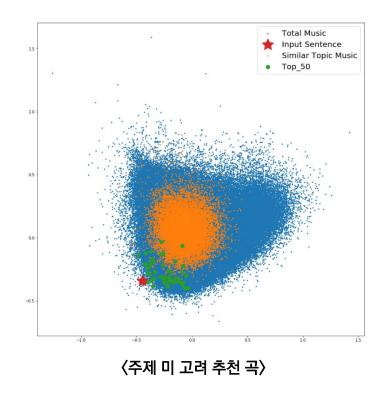




5. (1) 추천 시스템 실험 결과

- 주제를 고려하여 음악을 추천함으로써 입력 문장의 주제에 보다 의미상 근접한 음악을 제공함
- 주제 미 고려 시, 입력 문장과 거리상으로는 더 가까운 음악이 선택됨
 - → 의미상 유사하지 않은 경우가 다수 존재



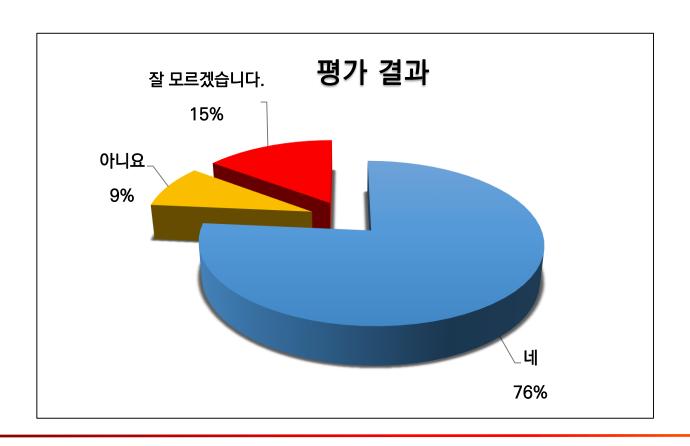




5. (1) 추천 시스템 실험 결과

- 추천 결과에 관한 설문을 시행
- 개별 추천에 관한 결과와 가사 기반 추천의 효율성에 관한 질문을 제공

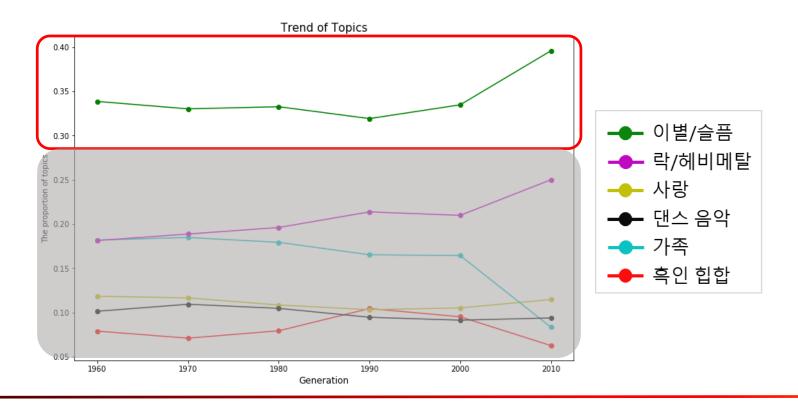
 본 연구를 종합적으로 평가 하였을 때, 음악 추천시스템에 가사와 가사의 주제 를 고려 하는 것이 효과적이라고 생각 하십니까?





5. (2) 시각화 및 경향 분석

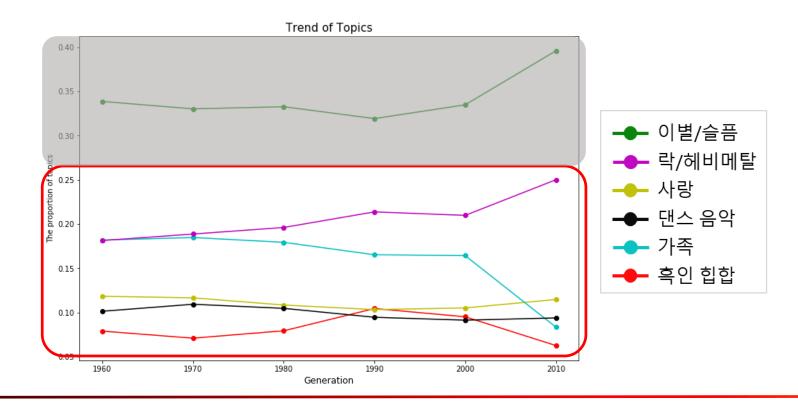
- 이별/슬픔 Topic은 항상 30% 이상 비율을 유지하며, 꾸준히 대중들의 관심을 받아 왔다고 볼 수 있음
- 락/헤비메탈 Topic의 증가 추세와 가족 Topic의 감소 추세를 보임
- 즉, 심오한 주제에 대해 고민하고 개인주의가 증가하는 현대 사회의 모습을 나타내고 있음





5. (2) 시각화 및 경향 분석

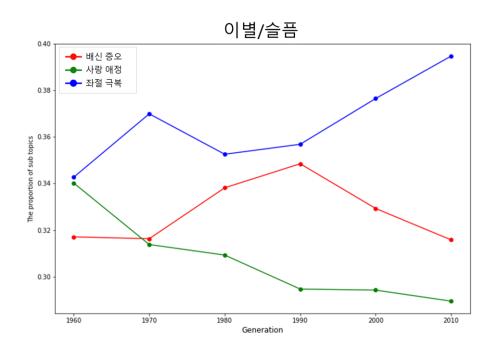
- 이별/슬픔 Topic은 항상 30% 이상 비율을 유지하며, 꾸준히 대중들의 관심을 받아 왔다고 볼 수 있음
- 락/헤비메탈 Topic의 증가 추세와 가족 Topic의 감소 추세를 보임
- 즉, 심오한 주제에 대해 고민하고 개인주의가 증가하는 현대 사회의 모습을 나타내고 있음

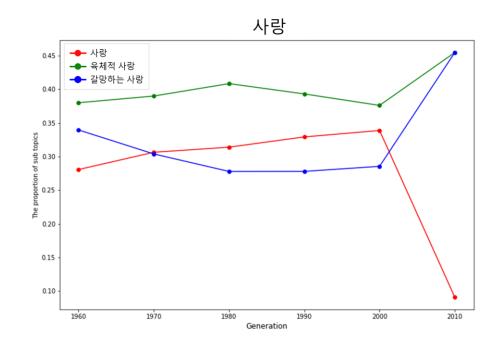




5. (2) 시각화 및 경향 분석

- 총 6가지 Main Topic 각각 Sub Topic에 대한 시각화를 진행하였음
- 이별/슬픔: 슬픔에도 불구하고 사랑 / 증오 를 보이는 곡들은 점점 줄어들고 반대로 극복에 대한 비율이 증가
 - ▶ 과거를 붙잡기보다는 미래를 바라보는 현대 사회의 경향을 알 수 있었음
- 사랑: 사랑에 대해서는 육체적 사랑의 비율이 항상 높으며 단순 사랑 노래는 급격히 감소





목치



- l. 문제 정의
- II. 중간 보고 Comment 정리
- III. Term Project 수행 내용 요약
- IV. Term Project 수행 내용
- V. 결론 및 제언

V. 결론 및 제언



❖ 결론

- 사용자의 감정이 표현된 문장을 입력 받아, 가사 기반의 추천 시스템 구축
- LDA를 통해 음악에 내재된 주제를 추출하여 이를 기반으로 추천 시스템의 효과 증대
- 일반인 설문 결과, 가사를 이용함으로써 사용자에게 보다 만족감을 주는 추천 시스템 개발 가능성 확인
- 기존의 음악추천 방식에 가사를 사용하여 다양한 정보를 이용한 추천 시스템 구현이 가능할 것으로 기대

❖ 한계점

- 대량의 가사들을 사용하여 HLDA를 수행하여 보지 못한 점이 아쉬움
- 많은 학우들의 조언처럼, 멜로디와 메타데이터, 가사를 함께 이용하였을 때의 성능에 대한 궁금증 존재



감사합니다.

Appendix. HLDA 설명 및 결과



- 3. (1) (LDA -> LDA) vs. (Hierarchical LDA)
 - 20번의 반복을 진행하는 데 소요되는 시간 : 약 12시간
 - Hyperparameter 지정 방식이 LDA와 차이 존재(토픽의 개수를 지정 불가)
 - 해당 알고리즘은 대기 행렬 이론과 매우 유사