

# 딥러닝을 활용한 노래 가사 분석 및 추천 시스템

팀 E 김은비 유창현 이은경 이인석

# 발표 내용

이번 주차 진행 사항

다음 주차에 진행해야 할 사항

# 기존 유사도 분석의 문제점 해결 방안

1. 품사 제한 없이 유사도 분석했던 기존 방식 대신 품사를 일반 명사와 형용사로 제한하여 유사도 분석
2. 기존 사전 구축에 사용한 word2vec 방식은 코사인 유사도를 통해 유사도를 산출하기 때문에 다른 유사도 계산 방식 사용
3. 워드 임베딩 방식을 word2vec 대신 다른 방식 사용

# 품사 제한

일반명사와 형용사만을 사용하여 유사도 측정

```
for i in df.loc[:]["가사"]:  
    token = tokenizer.pos(i)  
    picked = []  
    for j in range(len(token)):  
        if (token[j][1] == 'NNG' and len(token[j][0]) > 1) or token[j][1]  
            picked.append(token[j][0])  
    tagging_kkma.append(picked)
```

# 감정 분류 변경

- 기존에는 감정을 7가지 (행복, 편안함, 두려움, 놀람, 분노, 슬픔, 무관심) 로 분류
- '놀람'의 경우 감정 사전 구축이 쉽지 않았고, '사랑'과 적절한 감정이 아니라 생각되어 배제
- '편안함'의 경우 또한 마찬가지로 감정 사전 구축이 쉽지 않았고, 무관심과 겹친다고 생각하여 배제

# 품사 제한 결과

만족할 수 없고, 결과적으로 문맥을 고려하지 못한 결과 생성

행복		슬픔		분노		무관심						두려움					
단어	유사도	단어	유사도	단어	유사도	편안		여유		무심		무섭		두렵		두려움	
						단어	유사도	단어	유사도	단어	유사도	단어	유사도	단어	유사도	단어	유사도
감사	0.91527522	눈물	0.88677794	수룩	0.97509134	이맘때	0.96012634	어이	0.96526027	적응	0.97523189	내기	0.95090091	수룩	0.89540029	지라	0.96836573
고맙	0.89242041	주렴	0.88539439	야만	0.97310942	때쯤	0.95284349	막지	0.96355653	되질	0.97404242	일인	0.94288737	가도	0.89111304	무관심	0.96733677
전부	0.87080586	아픔	0.87642086	놓	0.97310328	숨바꼭질	0.94688785	이하	0.96247303	심하	0.97125399	버겁	0.94090098	모질	0.87767726	타인	0.96281558
벅차	0.83748579	힘겹	0.85673386	어리석	0.96793079	반쯤	0.94320965	땀청	0.96168518	밤길	0.96954489	방황	0.93774867	원망	0.87550509	가나	0.96120876
기도	0.82575685	고통	0.84619629	욕해	0.96269691	갈망	0.94110793	두발	0.96035171	담담	0.96866715	두진	0.93771994	힘겹	0.87078965	족쇄	0.96017402
부족	0.82420254	가요	0.82199097	바람	0.9621951	분하	0.93952721	묘하	0.95967627	의심	0.96828222	지질	0.93651664	헛되	0.86001509	적도	0.95908272
바라	0.78743976	가슴속	0.81564522	뒤늦	0.96080458	잔상	0.93903351	우후	0.95966935	오기	0.96771365	내색	0.93617296	다짐	0.85978848	일생	0.95817423
살기	0.78369892	흔적	0.81378502	나약	0.96057343	겁	0.93805933	이벤트	0.95734578	외	0.96497136	다음날	0.93423218	욕심	0.85746133	절망	0.95758569
기쁘	0.78344989	그리움	0.81155622	니야	0.96025509	채운	0.93748295	난대	0.95696223	헤프	0.9644978	기지	0.93336332	잘못	0.85456336	울음	0.95749056
사람	0.78318763	뒷모습	0.80974662	구도	0.95882589	사라지	0.93431085	정상	0.95659488	고정	0.96289998	기나길	0.93054622	외면	0.85418534	아내	0.95696664
초라	0.78002489	동안	0.80123109	돌리기	0.95490104	시월	0.9342587	각오	0.95501292	친군	0.95719016	감당	0.92916715	애원	0.84925276	냉정	0.95514739
이대	0.77704918	대여	0.79747033	다시	0.95206094	별은	0.930812	정답	0.95306861	별것	0.95618761	애꽃	0.92903268	어리석	0.84145141	천일	0.95470816
사랑	0.7632637	슬프	0.78652585	잔인	0.9517712	유토	0.92920732	장님	0.95205724	억지	0.95595062	정리	0.9287852	없잖	0.84133327	허물	0.95392776
대가	0.75817877	흥터	0.78406024	냉정	0.95129955	그간	0.92917323	환상적	0.9518944	우습	0.95593923	최선	0.92812836	부탁	0.84118676	여섯	0.95349753
인걸	0.75781059	대면	0.77541596	조	0.95010167	오일	0.92876983	달링	0.95103729	동정	0.95556629	만도	0.92773455	감당	0.83713371	구슬	0.95283788
다행	0.75584877	빛속	0.77251244	아시나요	0.94970846	대론	0.92789823	사주	0.95099252	고리	0.95553249	지옥	0.9270035	멀	0.83693558	집착	0.95100433
간직	0.75106323	애원	0.77175283	희생	0.94851035	새	0.92777777	친군	0.9508521	기라	0.95538652	일과	0.92688841	아픔	0.83600509	열쇠	0.95091045
알	0.74670851	지난날	0.7703411	일생	0.9484911	입꼬리	0.92685008	연예	0.95076489	괴로움	0.95496404	혼잣말	0.92589754	인연	0.83243638	사치	0.95058632
평생	0.74147213	아시나요	0.77023876	애처름	0.94817257	노을빛	0.92671001	쿠키	0.95056522	배려	0.95451021	어항	0.92501295	미련	0.8295489	중함	0.95041698
모자	0.73865175	빗물	0.76403785	모질	0.94786328	찬바람	0.92662597	하루쯤	0.95041394	하려	0.95441383	다툼	0.92395443	미움	0.82826561	지나치	0.95007586
만약	0.73688149	이토	0.7629689	안타깝	0.94719714	구멍	0.9260388	여우	0.94936001	난대	0.95440161	속상하	0.92357469	주지	0.82807136	일거	0.9497683
나날	0.73403108	겨울밤	0.76267147	한데	0.94677413	호기심	0.92599523	놀랍	0.94843769	보채	0.95409364	가세	0.92212063	뒤늦	0.82297802	악해	0.94970024
대도	0.72888815	열병	0.76224089	헤어	0.94548029	초승달	0.92597318	소한	0.94794726	자만	0.95348912	막다르	0.92124885	포기	0.81943911	죽음	0.94966781
인연	0.72767437	아푸	0.75668025	말자	0.94466352	과거	0.92553818	크기	0.94781977	저주	0.95339614	안심	0.92122406	빈자리	0.81883013	풋사랑	0.94911969
많	0.72695518	가슴	0.75402242	주기	0.94419491	새벽녘	0.92428637	불가능	0.94770807	인지	0.95229936	단말	0.92036569	서럽	0.81625092	필요없	0.94911259
나지	0.72635627	상처	0.75307357	이율	0.94314373	한구석	0.92417967	칭찬	0.94757736	낭비	0.95190513	각자	0.9193958	외로움	0.81584096	어여쁘	0.94890201
내도	0.726143	외로움	0.75119507	아깝	0.94310856	자석	0.92339218	실없	0.94752526	소심	0.95178843	내야	0.91902006	치료	0.81478506	바라기	0.9475885
약속	0.7245394	깊	0.74381387	열병	0.94260257	이루	0.92323929	이지만	0.94750464	기지	0.95130622	습관	0.91800785	없다	0.81438971	안타깝	0.94640791
세상	0.72169161	아프	0.74149448	애원	0.94139099	달래	0.92316794	외	0.94693804	리지	0.95090002	낯	0.91765231	무디	0.81347018	내안	0.94605058
동안	0.72148162	서글프	0.73799276	낯질	0.94132721	빈틈	0.92297399	엄지	0.94679517	자라	0.95067549	야광	0.91737086	울기	0.81008136	이기심	0.9458586

# 적절하지 않은 결과 도출 해결

## 코드 변경

- 다시 한 번 word2vec 모델의 파라미터 점검
- min\_count = 1, iter = 5 로 하였을 때 원했던 감정 사전과 비슷하게 결과가 도출됨을 확인할 수 있었음
- min\_count 의 경우 1로 설정하면 모든 경우에 대해 무시하지 않으므로 정확한 결과가 도출되었다고 생각

# 적절하지 않은 결과 도출 해결

## 키워드 수정

- 행복->'행복', 슬픔->'슬픔', 무관심->'무심', 분노->'화', 두려움->'두렵'
- 키워드를 가사에서 사용할 가능성이 높은 키워드들로 유사도 분석



# 최종 분석 결과

적절치 않거나 애매한 단어들이 거의 없지만 그러한 단어의 경우 인위적 처리(삭제)해도 될지?

행복		슬픔		무관심		분노		두려움	
단어	유사도	단어	유사도	단어	유사도	단어	유사도	단어	유사도
소중	0.6766625	아픔	0.79957974	담담	0.79819614	티	0.7740007	무섭	0.75231874
부족	0.67472577	흔적	0.71585464	덤덤	0.79390687	화내	0.77310103	변하	0.74434292
사랑	0.656708	외로움	0.67316556	태연	0.78423518	망치	0.76523089	미워지	0.69806772
억울	0.64168227	바람	0.6723215	무관심	0.76668942	욕	0.75582671	잊히	0.67009401
간직	0.62589657	추억	0.64920235	실망	0.76469564	쓰이	0.75535667	겁	0.65137672
완벽	0.61835122	색	0.64481741	아리송	0.75918233	끓	0.74717605	버겁	0.64647341
고맙	0.60794246	어둠	0.63977611	힘차	0.75139546	질리	0.73950529	달라지	0.64578253
아파하	0.58119625	계절	0.63546622	시시	0.75112325	매달리	0.72734559	밉	0.63672245
힘들	0.57923853	고통	0.63125658	시크	0.74774015	보채	0.72300792	좋아지	0.63640726
감사	0.57149899	희망	0.62540185	소심	0.74101937	울먹이	0.71460795	힘들	0.62640613
이해	0.56675023	그리움	0.62530071	잔인	0.74085617	수줍어하	0.71190435	늦	0.6231519
영원	0.5637235	침묵	0.62252235	한심	0.74030709	거절	0.70031011	어렵	0.6207912
기도	0.56199872	병	0.62151134	변화무쌍	0.73682296	회사	0.69570756	아쉽	0.62059814
충분	0.56145269	아쉬움	0.61769199	허술	0.73629373	도망치	0.69254106	낯설	0.61929047
선명	0.54223776	진행	0.61769086	야속	0.73527551	밀	0.68981999	이러	0.61825591
함께	0.53856039	인연	0.61310124	연연	0.73028219	다독이	0.68710136	싫	0.61620784
다정	0.53359425	나날	0.61187863	우월	0.72621226	밀어내	0.68582445	언제쯤	0.61413062
약속	0.5279392	땀방울	0.61071646	진부	0.72596854	쿨하	0.68538696	무너지	0.61305946
미안	0.52466464	우산	0.60980725	심각	0.72528392	오해	0.6852994	망설이	0.60999703
기억	0.52171361	메일	0.60717529	나긋	0.72094852	좋아요	0.68383795	바뀌	0.60065699

# 다음 주차 진행 사항

- 감성사전 최종 확정 및 이를 활용한 곡별 감정 라벨링 진행
- 감정 분류에 활용할 딥러닝 모델 결정 및 설계 진행