

닐로 날로잡기 프로젝트 동향분석서

작성일	'18. 11. 3
작성자	남궁찬(품질 관리자)
참조논문	하정철, 강동훈, 박재모, 장으뜸, 이은영, 임성현, 길준민. (2016). 뉴스 기사와 음원차트 간의 상관관계 추출을 위한 R언어 기반 분석기의 설계 및 구현. 한국정보과학회 학술발표논문집, , 122-124.
결과도출방법	크롤링을 통해 직접 데이터를 수집하고 R 기반의 빅데이터 분석 기술 중 하나인 오피니언 마이닝을 적용하여 비정형 데이터인 뉴스기사에 대한 감성분석 및 뉴스 기사 내용에 대한 감성점수를 측정하여 부정기사가 음원차트에 끼치는 부정적인 영향력을 알아보고, 전체 기사 수와 음원차트 간 상관관계를 분석한다.
사용기술	설명
	적용가능여부
회귀분석	한 변수가 다른 변수에 대해 미치는 영향을 추정할 수 있는 통계 기법, 라쏘, 능형(릿지), 엘라스틱넷 등이 있다. 해당 논문에서는 엘라스틱넷 기술을 사용하여 감성분석을 위한 감성사전을 구축함.
	해당 프로젝트에서는 많은 변수를 사용하지 않으므로 릿지 회귀 방법을 적용하는 것이 적절해 보임. Python Scikit-Learn 라이브러리를 활용.
피어슨 상관계수	두 변수간의 관련성을 얻기 위해 보편적으로 사용되는 방법, 상관계수가 -1.0에 가까우면 강한 음적 선형관계, +1.0에 가까우면 강한 양적 선형관계를 가진다. 해당 논문에서는 뉴스기사 수와 음원차트 순위의 관계를 비교한다. 뉴스기사 수가 많고 음원순위의 숫자가 낮을수록 -1.0에 가까워지며 상관도가 높다고 볼 수 있다.
	변수간의 관련성을 얻기 위해 사용하기 적절해 보임. SPSS에서 지원하고 있음
오피니언 마이닝 (감성분석)	텍스트에 나타난 사람들의 의견이나 성향 같은 주관적인 데이터를 분석하는 자연어 처리 기술. 해당 논문에서는 뉴스 기사의 텍스트를 분석하여 긍정적인 뉴스인지 부정적인 뉴스인지 여부를 판별하기 위해 사용함.
	화제성 변수의 긍정적, 부정적 여부는 음원순위에 영향을 끼치지 않을 것으로 보임. (ex. 노이즈 마케팅) 따라서 적용하기에 부적절.