

Twitter mood predicts the stock market.

1. 요약

목적

트위터에서 수집한 트윗을 바탕으로 대중의 감정이 DJIA(Dow Jones Industrial Average) 주가 변동과 상관관계가 있는지, 주가변동을 예측 가능한지를 알아본다.

방법

'GPOMS와 OF로 측정한 대중의 감정이 미래의 DJIA 가격을 예측하는 데에 효과가 있다'는 가설을 바탕으로 연구를 진행하였다.

2008.2.28-2008.12.19 사이의 트윗을 수집하여 OF(OpinionFinder)와 GPOMS(Google-Profile of Mood States)를 사용해 해당 트윗이 어떤 감정을 내포하는지 평가한다. 그 후 Granger 인과관계 분석 모델을 활용해 수집된 감정 데이터가 주가변동과 관계가 있는지, 변동을 예측할 수 있는지 분석한다. 최종적으로 SOFNN모델로 학습시켜 예측하는 순서로 연구가 진행되었다.

수집한 트윗을 감정 차원으로 분류할 때 두 개의 모델을 사용한 것은, OF는 트윗을 긍정/부정의 이분법적으로 분류하는 것만 가능하기 때문이다. 이는 다차원적인 사람의 감정을 제대로 반영하지 못하는 결과를 도출할 수 있기 때문에 6개의 감정 차원을 가진 GPOMS를 추가로 활용하였다. OF와 GPOMS 각각에서 도출된 time series는 비교를 위해 정규화 과정을 거쳤다.

연구 결과 GPOMS의 'Calm' 감정지수가 DJIA변동 예측에 영향이 있었고, OF를 통해 분류된 '긍정/부정' 감정은 큰 영향을 미치지 않았다. Granger 인과관계 분석 모델을 활용한 결과 마찬가지로 'Calm' 감정이 향후 6일 이내의 DJIA 변동을 예측할 수 있다는 것이 확인되었다. SOFNN모델을 활용한 예측모델의 성능 역시 감정 데이터를 포함한 경우가 예측 정확도를 향상시킨 것이 확인되었고, 그 중에서도 'Calm'과 'Happiness'를 포함한 모델이 87.6%로 가장 높은 예측 정확도를 보였다.

기여점

대중의 감정이 주가변동 예측에 유의미한 영향을 준다는 것을 확인하고, 이를 금융 시장 예측에 활용했을 때의 효과를 기대할 수 있다.

2. 배운점, 장점

첫째로, 머신러닝 분야의 영문으로 작성된 논문을 읽은 것이 처음이기 때문에, 이 분야에서 사용되는 용어들을 학습할 수 있었다.

둘째로, 사람의 감정이라는 모호하고 추상적인 개념을 텍스트 속에서 찾아내고, 분석에 활용 가능한 형태로 만들어내는 과정이 흥미로웠다. 그 과정에서 활용된 복잡한 계산식을 완전히 이해하지는 못했으나, 감성 분석에 대한 힌트를 얻은 것 같았다.

3. 문제점/의문점/제안하고 싶은 것

감정표현을 하는 트윗을 수집하기 위해서 "I feel", "I am feeling", "I'm feeling", "I don't feel", "I'm", "I'm", "I am", "makes me" 등의 표현이 포함되어 있는 트윗 만을 수집하였다. 이 표현을 쓰지 않는 트윗에도 중요한 평가가 있을 수 있는데 그런 트윗까지 포함할 수 있는 방법은 없을까 고민해 보게 되었다. 또한 우리나라에 적용한다면, 문법이나 표현, 신조어의 생성 등이 더 자유로운 한국어에 대해서는 수집 시에 어떻게 필터링을 할 수 있을지, 국내에서 유사한 후속연구가 있는지 궁금했다.