

1. 데이터마이닝

Selenium을 이용해서, 과거 금융관련 자료들을 모은 후 Senti-score을 계산

좀 더 편리한 분석을 위해, 날짜순으로 텍스트를 정렬.

2. 실험 결과 및 정리

1) 선형 회귀분석, 상관계수

(2012.11.9~ 2018/8/17)

Sentiment Factor와 중국시장 인덱스(SSE, SZSE)를 선형 회귀분석한 결과

TABLE II
THE REGRESSION RESULT OF THE STANDARD SENTIMENTAL FACTOR
WITH SSE

standard sentimental factor	11138.27 ***
p	(3.48e-12)

TABLE III
THE REGRESSION RESULT OF THE STANDARD SENTIMENTAL FACTOR
WITH SZSE

standard sentimental factor	39796.6 ***
p	(<2e-16)

피어슨 상관계수는 0.18731(SSE)로 Sentiment Factor와 시장 인덱스가 약한 상관관계를 나타낸다는 것을 알 수 있음. 하지만, 조정된 Sentiment Factor 값을 택했을 때는 향상된 값 0.26119를 얻을 수 있었음.

Pearson correlation coefficient		
	SSE	SZSE
random from uniform	-0.00049399	-0.0005597
random from normal	0.00017451	0.00016782
temperature	-0.025135	-0.063723
standard sentimental factor	0.18731	0.22595
adjusted sentimental factor	0.26119	0.28472

TABLE IV
PEARSON CORRELATION COEFFICIENT FROM 2012/11/6 TO 2018/8/17

(2015.02.11~09.11) 주가 대폭락 기간

동일한 방법으로 피어슨 상관계수를 보면 0.36284로 Sentiment Factor 시장 인덱스가 중간 상관성을 가지는 것으로 나타남. 조정된 값을 택했을 때는 0.58815로 높은 상관관계를 보인다는 것을 알 수 있었음.

Pearson correlation coefficient		
	SSE	SZSE
random from uniform	0.0011922	0.0010882
random from normal	0.00055439	0.00050227
temperature	0.10125	0.064122
standard sentimental factor	0.36284	0.37204
adjusted sentimental factor	0.58815	0.58042

TABLE V
PEARSON CORRELATION COEFFICIENT FROM 2015/2/11 TO 2015/9/11

→ 여론영향을 많이 받은 해당기간(2015/2/11 ~ 9/11)에서 유의미한 결과 얻을 수 있음

2) 주가 대폭락 기간 시계열 분석

시계열 분석을 한 결과, 동일한 기간동안 Sentiment Factor은 시장 인덱스의 움직임과 똑같은 모습을 보인다는 것을 발견.



3) 단어 분석

2781개의 단어에 '1(긍정) -1(부정) 0 보통'을 부과한다음, senti-score가 0.1보다 낮거나 중간값을 보이는 경우 삭제했더니, 452개의 주요 단어에서 73.0088%의 정확도를 얻을 수 있었음.