

테일러 규칙

테일러 규칙(Taylor Rule)은 중앙은행이 경제 상황에 맞춰 기준금리를 얼마나 정해야 하는지를 보여주는 대표적인 경제 공식이다. ($r = p + 0.5y + 0.5(p - 2) + 2$)

r : 적정 명목 기준금리

p : 현재 인플레이션율 (물가 상승률)

y : GDP 캡 (실제 GDP와 잠재 GDP의 차이, 즉 경기가 얼마나 과열/침체되었는지)

$p - 2$: 인플레이션 캡 (실제 물가와 목표 물가(보통 2%)의 차이)

2 : 경제가 안정적일 때의 실질 금리(중립 금리)

논문 간단 정리

2005년 5월 ~ 2017년 12월까지 231,699건의 문서에서 통화 정책 결정에 따른 정량적 정보를 추출하기 위해 텍스트 마이닝 사용. (금통위 의사록 151분, 금리 단어 포함 뉴스 기사 206,223건, 채권 분석 보고서 25,325건)

한국은행(BOK)의 통화정책위원회(MPP) 회의록의 정성적 내용을 정량적 지표로 변환하여 회의록의 감정과 어조를 측정하고, 이것이 실제 금리 결정에 미치는 영향과 추가적인 정보를 담고 있는지 검증. (중앙은행의 소통이 통화정책의 중요한 도구라고 판단하여 정량화 하는 것)

특정 분야 사전과 한국어 원문을 사용하는 것이 영어로 번역된 텍스트를 사용하는 것보다 중요하다는 것을 증명함.

주요 분석 도구 (eKoNLPy) : 한국어 경제·금융 분야에 특화된 자연어 처리 라이브러리를 직접 개발하여 전문 용어를 정확히 인식할 수 있게 함.

분석 기법 : 단어 하나(1-gram)부터 다섯 단어의 연속(5-gram)까지 분석하여 문맥과 미묘한 어조 변화를 포착. (낮은 실업, 부진한 회복과 같이 긍정과 부정을 결합한 단어 파악용)
(의견 형성에 중요한 역할을 하는 명사, 형용사, 부사, 동사, 부정어만을 사용)
(전체 데이터에서 15회 미만으로 등장하는 n-그램은 분석의 정확도를 위해 제외)
(한 문장 내에서 여러 n-그램이 겹칠 경우, 가장 긴 n-gram(가장 높은 n)만을 분석 단위로 인정)

감성 분석 접근법

시장 접근법 : 단어가 발표된 날의 시장 금리 변화와 단어의 출현 빈도를 연결하여 매파 또는 비둘기파 성향을 분류. (연구자의 주관적 판단을 배제하기 위한 방법)

=> 문서 발표일 기준 콜금리의 1개월 변동분이 양수(+)이면 해당 문서를 '매파'로, 음수(-)이면 '비둘기파'로 라벨링.

=> 의미없는 미세한 움직임을 제외하기 위해 +-3bp의 임계값(Threshold)을 적용

사전적 접근법 : 특정 Seed words와의 유사성을 바탕으로 단어의 성향을 결정.

=> 단순한 단어 임베딩을 넘어 n-그램(연속된 단어 뭉치) 간의 관계를 학습하는 'ngram2vec' 기술을 사용.

=> 도메인 특화 감성 유도 알고리즘인 'SentProp' 프레임워크를 채택

- => 이미 성향이 명확하게 알려진 소수의 'Seed Words'로부터 시작하여, 벡터 공간상에서 가까운 위치에 있는 다른 단어들로 성향을 전파
- => 극성 점수가 1.1보다 크면 매파, 1/1.1(약 0.91)보다 작으면 비둘기파로 분류

시장 접근법 기반 라벨링

사람이 직접 라벨링하는 대신 금융 시장의 데이터를 지표로 삼아 기계학습 모델이 스스로 학습하도록 함.

뉴스 기사와 채권 보고서가 발표된 날을 기준으로 콜금리의 1개월 변동분을 확인.

콜금리가 유의미하게 상승(+3bp 초과)한 날 발표된 문서는 매파로 분류.

콜금리가 유의미하게 하락(-3bp 미만)한 날 발표된 문서는 비둘기파로 분류.

이렇게 라벨링된 400만 개 이상의 문장을 나이브 베이즈 분류기(NBM)에 학습.

특정 n-그램이 매파적 문서에 등장할 확률과 비둘기파적 문서에 등장할 확률의 비율을 계산하여 극성 점수를 매김.

점수가 1.3보다 크면 매파, 1/1.3(약 0.77)보다 작으면 비둘기파로 최종 라벨링.

그 사이의 중립적인 단어들은 제외.

사전적 접근법 기반 라벨링

단어 간의 의미적 유사성을 바탕으로 라벨링을 전파하는 방식.

누가 봐도 명확한 성향을 가진 단어들을 먼저 선정(Seed Words).

매파 : 높다, 인상, 팽창, 과열, 긴축 등 25개.

비둘기파 : 낮다, 인하, 축소, 침체, 완화 등 25개.

SentProp이라는 알고리즘을 사용하여, 벡터 공간상에서 seed words와 가까이 위치한 n-그램들을 찾음.

seed words에서 시작한 Random walk이 특정 n-그램에 도달할 확률을 계산하여 극성을 부여.

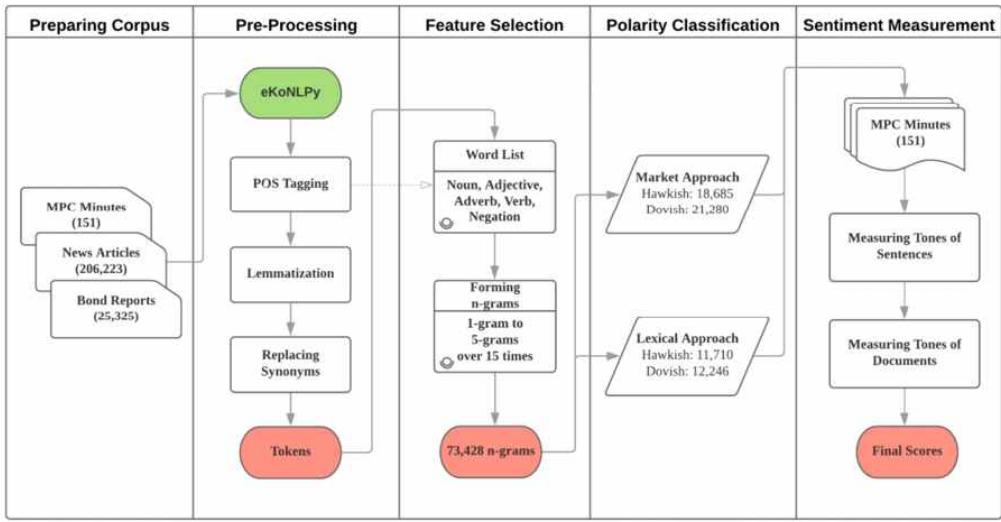
상대적 확률 비율이 1.1보다 크면 매파, 1/1.1(약 0.91)보다 작으면 비둘기파로 라벨링

검증을 위한 수동 라벨링

모델의 정확도를 확인하기 위해 별도로 사람이 직접 라벨링한 데이터를 사용해 테스트.

한국은행 총재의 기자간담회 모두발언에서 추출한 2,341개의 문장을 사람이 직접 매파, 중립, 비둘기파로 분류.

위에서 기계적으로 생성한 단어 사전들이 이 수동 라벨링된 문장들을 얼마나 잘 예측하는지 확인한 결과, 약 67~68%의 정확도를 보임.



감성 분석 절차 (eKoNLPy는 논문 저자가 배포한 라이브러리)

Document type	No. of docs	Average no. of sentences	Max no. of sentences
MPB minutes	151	165	326
News articles	206,223	15	340
Bond reports	25,325	49	2,515
Total	231,699	19	2,515

논문 저자가 사용한 문서 요약

(뉴스 기사는 이데일리, 연합뉴스, 연합 인포맥스만 사용, 월작자의 중복 기사 방지)