

과학 연구를 위한 JSPS 보조금(S)

일본의 지속적인 디스플레이션 이해

워킹 페이퍼 시리즈

002호

2012년 10월

비즈니스 뉴스에서 고품질 주제 추출 비정상적인 금융시장 변동성을 설명하다

히사노 료헤이
디디에 소르네트
미즈노 타카유키
오니시 다카아키
와타나베 츠토무

도쿄대학 가격 프로젝트
702 도쿄대학 경제학부,
113-0033 일본 도쿄도 분쿄구 혼고 7-3-1
전화: +81-3-5841-5595
이메일: watlab@eu-tokyo.ac.jp [http://
www.price.eu-tokyo.ac.jp/english/](http://www.price.eu-tokyo.ac.jp/english/)

작업 보고서는 토론을 위해 공유되는 초안 형식의 일련의 원고입니다.
댓글 목적으로만 사용됩니다. 저자가 명시한 경우를 제외하고는 배포 또는 배포할 수 없습니다. 따라서 작업 보고서는 작성자의 명시적인 동의 없이 복제되거나 배포될 수 없습니다.

비즈니스 뉴스에서 고품질 주제 추출로 비정상적인 금융 시장을 설명합니다. 휘발성

히사노 료헤이^{ab}, 디디에 소르네트^{교류}, 미즈노 타카유키^{dbe}, 오니시 타카아키^{BE}, 와타나베 츠토무^{BE}
†ETH 취리히, 경영학과,
기술과 경제학
비캐논 글로벌 연구소
씨스위스 금융 연구소
다쓰쿠바대학교 SIE 대학원 컴퓨터공학과

이자형도쿄대학 경제학 연구과

(날짜: 2012년 10월 23일)

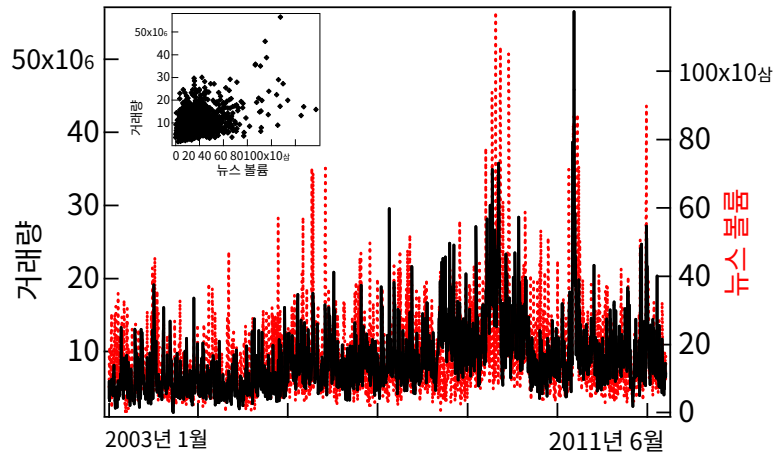
오늘날 사회의 정보 흐름과 사회 활동 사이의 상호 관계를 이해하는 것은 사회 과학의 초석 중 하나입니다. 금융 경제학에서 이와 관련된 핵심 문제는 가능한 모든 유형(지정학적, 환경적, 사회적, 금융, 경제 등)의 뉴스가 조직화된 주식 시장에서 기업의 거래와 가격에 어떻게 영향을 미치는지 이해하고 정량화하는 것입니다. 본 백서에서 우리는 Thompson Reuters가 제공한 2,400만 개 이상의 뉴스 기록과 S&P 미국 주가 지수의 205개 주요 주식에 대한 거래 활동과의 관계를 분석하여 이 문제를 해결하려고 합니다. 우리는 주식 가격 변동에 영향을 미치는 뉴스의 전체 상황이 거래 활동과 분해된 뉴스 정보 사이의 단순 정규 회귀를 통해 자동으로 요약될 수 있음을 보여줍니다. 간단한 주제 모델링 기술을 사용하여 "주제별" 기능을 구현합니다. 이러한 방법을 사용하여 뉴스가 거래에 미치는 영향을 추정하고 정량화할 수 있습니다. 우리는 주식 바구니와 관련된 뉴스 정보의 전체 풍경을 표현하기 위해 네트워크 기반 시각화 기술을 소개합니다. 주제 분포를 대표하는 단어를 조사하면 우리의 방법이 주식 시장에 영향을 미치는 중요한 정보를 추출할 수 있음을 확인할 수 있습니다. 우리의 결과는 금융 경제에서 가장 혼란스러운 정형화된 사실 중 하나, 즉 특정 시점에 거래량이 "비정상적으로 큰" 것처럼 보이는 현상이 뉴스의 흐름으로 설명될 수 있음을 보여줍니다. 이러한 의미에서, 우리의 결과는 뉴스가 진정으로 참신하고 관련 금융 정보를 제공하는 경우 "과잉 거래"가 없음을 입증합니다.

I. 소개

"효율적 시장 가설"(EMH)에 기초한 신고전주의 금융 경제학에서는 가격 변동을 정보 흐름에 대한 거의 완벽한 즉각적인 반응으로 간주합니다. 따라서 EMH에 따르면 가격 변동은 단순히 외생적인 뉴스를 반영할 뿐입니다. 가능한 모든 유형(지정학적, 환경적, 사회적, 금융, 경제 등)의 이러한 뉴스는 투자자들로 하여금 기업의 투자 프로젝트가 미래에 창출할 수 있는 현금 흐름에 대한 기대치를 지속적으로 재평가하게 만듭니다. 이러한 재평가는 재조정된 수요/공급 기능으로 변환되며, 이는 수요와 공급 간의 순 불균형에 따라 가격을 기본 가치로 올리거나 내리게 합니다. 결과적으로 관측된 가격은 미래 현금 흐름의 현재 가치를 가장 잘 나타내는 것으로 간주됩니다. 이 관점에서, 시장 움직임은 내부 피드백 루프 없이 순전히 외생적입니다. 특히, 충돌 시 발생하는 가장 극심한 손실은 오로지 외생적으로 발생하는 것으로 간주됩니다.

이 패러다임의 문제점은 실제로 실제 가격 변동을 특정 뉴스와 연관시키는 것이 매우 어렵다는 것입니다. 낮은 빈도이든 높은 빈도이든 가격 변화를 뉴스와 연관시키려는 많은 시도는 EMH에 대한 설득력 있는 뒷받침 증거를 찾는 데 실패했습니다[1-6]. 더욱이 EMH에서 예측한 것과 비교하여 가격이 너무 크게 움직이고 거래량이 너무 크다는 것이 오랫동안 인식되어 왔습니다[7-9]. 이는 정보의 외생적 흐름보다 가격 역학에 더 많은 것이 있음을 시사합니다. 이러한 배경에서 "재귀성"이라는 개념이 도입되었습니다[10]. 이는 투자자의 과거 행동도 현재 결정에 상당한 영향을 미쳐 피드백 루프와 중요한 내생적 역학을 생성한다는 개념을 구현합니다[11]. 지금까지 해결되지 않은 문제는 외생적 요인과 내생적 요인을 분리하고 어떤 뉴스가 정말로 중요한지, 그리고 그것이 가격에 어떻게 반영되는지를 이해하는 것입니다. 한편으로는 금융경제학의 가격 형성에 대한 뉴스 흐름의 선형적 기본 성격과 다른 한편으로는 이에 대한 실증적 지원이 없다는 점을 감안할 때, 그러한 이해와 그에 따른 통제가 없으면 금융 시장은 그대로 유지될 것입니다. 최근 몇 년간은 물론 대부분의 역사에 걸쳐 그들을 괴롭혔던 과도한 변동성, 급격한 가격 변동, 거품 및 붕괴에 취약합니다[12].

본 논문은 (i) 기관 투자자를 위해 수집된 대규모 비즈니스 뉴스 데이터베이스를 사용하고 (ii) 거래 활동에 영향을 미치는 관련 뉴스를 추출하는 새로운 방법론을 도입함으로써 위의 교착 상태를 타개하려는 시도를 나타냅니다. 이 새로운 방법론을 통해 우리는 가격 역학의 내생적 구성 요소를 대부분 제거할 수 있습니다.



무화과. 1: Toyota의 거래량(검은색 실선)과 뉴스 총량(빨간색 점선)의 시간 변화를 비교합니다. 삽화는 수반되는 뉴스 볼륨의 함수로 거래량을 표시합니다.

중요한 뉴스의 계층 구조를 식별합니다. 우리의 접근 방식은 뉴스가 금융 시장에 미치는 영향을 조사하는 이전 연구(예: [13-19])에서 사용된 접근 방식과 몇 가지 중요한 측면에서 다릅니다. 기존 연구의 한 계열은 구체적인 정보 내용을 고려하지 않고 뉴스가 제공하는 정보만을 집계 방식으로 분석했습니다. 하지만, 무심코 관찰한 바와 같이, 각 뉴스 기록은 투자자에게 서로 다른 의미를 갖기 때문에 가격에 미치는 영향도 다르기 때문에 단순히 특정 기간의 총 뉴스 기록 수를 계산하는 것은 제대로 작동하지 않습니다. 이전의 다른 연구에서는 수익 보고서, 새로운 경제 데이터 발표 등 소수의 제한된 뉴스만 고려했습니다. 따라서 당시에 도착하는 다른 유형의 뉴스가 미칠 수 있는 중대한 영향을 무시하는 심각한 한계에 시달렸습니다. 후자의 문제를 회피하는 한 가지 방법은 귀인 오류를 최소화하기 위해 매우 짧은 시간 간격을 사용하는 것입니다[20]. 그러나 [18, 21]을 포함한 최근 연구에 따르면 뉴스의 영향력은 며칠, 몇 주, 때로는 몇 달에 걸쳐 지속되므로 시간적 분할만으로는 뉴스의 영향력을 추출하는 것이 불가능하지는 않더라도 어렵습니다.

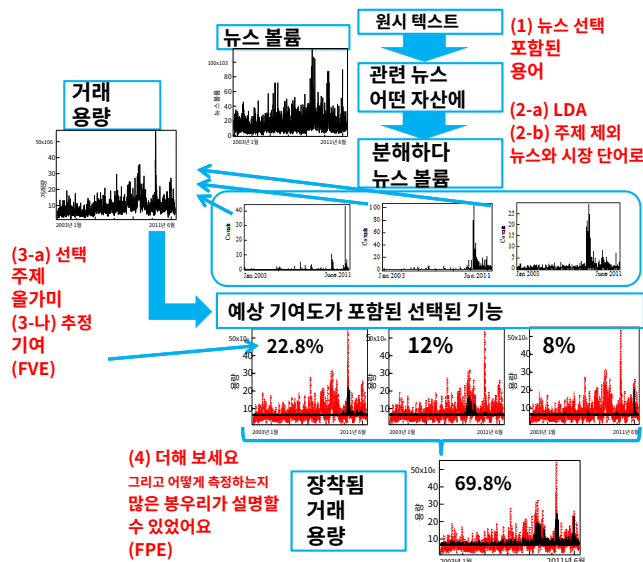
우리는 금융 거래 활동과 관련하여 관련 뉴스 유형을 동시에 세분화하여 평가함으로써 이러한 모든 문제를 해결합니다. 우리는 Thompson Reuters에서 제공한 2,400만 개 이상의 뉴스 기록에 대한 원시 텍스트를 마이닝하고 해당 기간 동안 5,000개 이상의 뉴스 기록이 있었던 S&P 500 미국 주가 지수에 상장된 205개 기업의 주식 거래 활동에 미치는 영향을 조사합니다. 2003년 1월부터 2011년 6월까지, 각 주식의 거래 활동을 설명하는 데 가장 관련성이 높은 정보를 결정하기 위해 정규화된 회귀 분석과 주제 모델링 기술을 조합하여 사용합니다. 이를 통해 다양한 뉴스의 상대적 중요성을 정량적으로 비교할 수 있습니다.

II. 방법론

거래 활동의 시간 변화(일일 거래량으로 측정)와 뉴스 볼륨의 시간 변화 사이에 좋은 상관관계가 존재한다는 것은 잘 알려져 있습니다[13-15]. 이 대응 관계는 그림 1에 설명되어 있습니다. 이는 Toyota 주식의 거래량(일일 거래되는 주식 수)의 시간 변화와 텍스트의 일일 단어 수로 측정되는 뉴스 양의 변화를 보여줍니다. 회사 이름인 Toyota가 포함된 기록입니다. (이 레코드에 있는 총 단어 수 대신) 뉴스 레코드 수만 사용하면 본질적으로 동일한 결과가 나옵니다.

이러한 대략적인 총체적인 서신에서 시작하여 우리의 훨씬 더 야심 찬 목표는 (a) 뉴스의 흐름을 관련 주제 및 관련 단어로 분리하고 (b) 개별 주식의 거래량을 분리하여 완전한 상호 의존 네트워크를 구축하는 것입니다. 그림 2는 (i) 주제 모델링 기술을 적용하여 뉴스의 전체 흐름을 주제별 특징으로 분해하고, (ii) 중요하지 않은 주제를 제거하기 위해 거래 활동에 미치는 영향을 동시에 추정하는 것으로 구성된 방법론의 흐름도를 제공합니다. (iii) 뉴스 충격으로 설명할 수 있는 거래 활동의 최고점 수를 정량화합니다.

용어(예: Toyota)가 선택되고 관련 뉴스 레코드가 수집되면(단계 (1)), 두 번째 단계는 그림 2와 같이 뉴스 정보를 "주제" 기능으로 분해하는 것입니다. 이 작업은 완료됩니다. LDA(Latent Dirichlet Allocation)라는 간단한 주제 모델링 기술을 적용하여 [22, 23]. 주제 모델은 그래픽입니다.



무화과, 2: 분석에서 수행된 절차를 요약한 흐름도. 괄호 안의 숫자는 단계를 나타냅니다. 단계 (1)에서는 특정 용어(여기에서는 Toyota와 같은 회사 이름)와 관련된 뉴스 레코드를 선택합니다. (2-a) 단계에서는 모든 문서를 다양한 주제의 혼합으로 분해하는 LDA(Latent Dirichlet Allocation)를 적용합니다. (3-a) 단계는 제한된 LASSO 회귀를 구현합니다. (3-b) 단계에 표시된 백분율은 각 주제의 예상 영향을 나타냅니다. 단계 (4)에 표시된 백분율은 뉴스에 의한 "설명된 (거래량) 최고점 비율"(FPE)이며, 이는 당사 방법론의 품질을 평가하기 위한 측정 기준입니다(텍스트 참조).

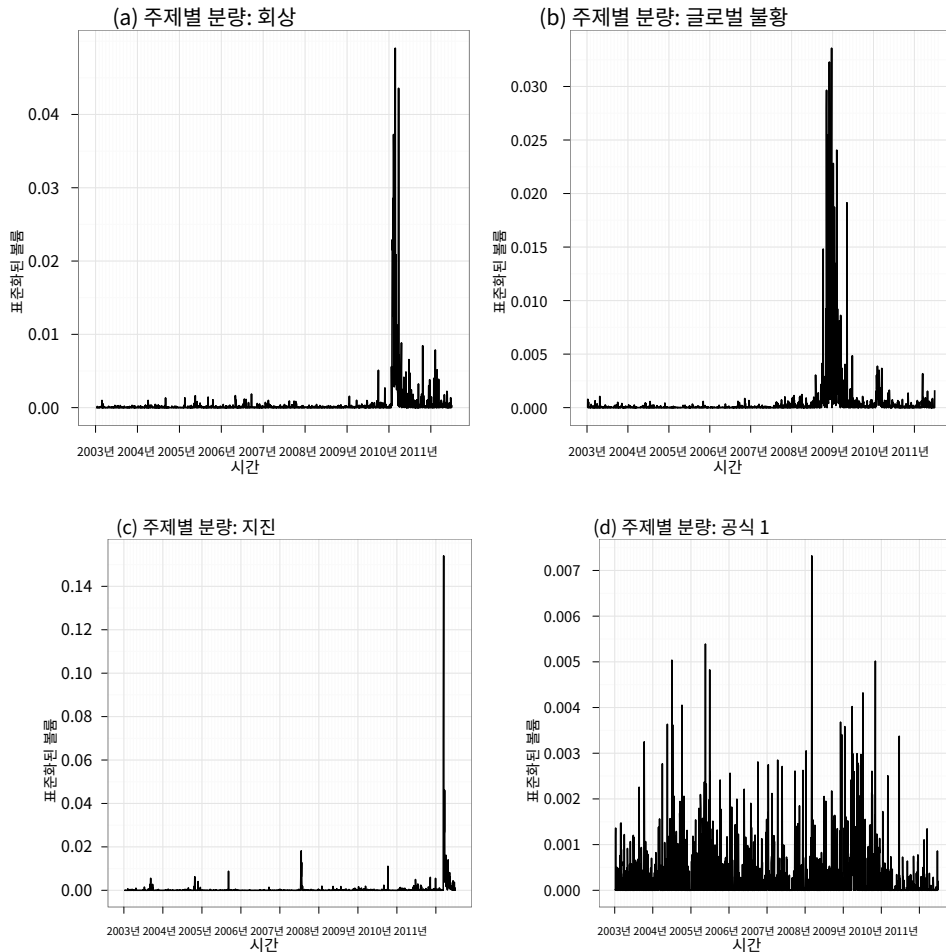
공유된 전역 다항 단어 분포(즉, 주제 분포)가 말뭉치를 지배한다고 가정하는 모델[24]. 특정 문서 내의 단어 빈도는 이러한 글로벌 주제 분포의 혼합으로 생성됩니다. LDA는 가장 간단한 주제 모델이며 기본 다항 분포의 희소성을 보장하기 위해 Dirichlet 사전을 사용합니다. 이렇게 하면 학습된 주제를 더 쉽게 해석할 수 있습니다. LDA는 이미 우수한 결과를 산출했기 때문에 더 정교한 주제 모델을 사용하는 것이 유용하지 않다는 것을 알았습니다. 우리는 원본 뉴스 기록에서 일반적인 불용어를 제거하고 본 논문에서 분석된 모든 주식에 대해 주제 수를 100으로 설정하여 LDA를 실행했습니다. 종목별 뉴스 기록 수에 따라 주제 수를 달리해도 결과는 크게 달라지지 않았습니다.

다음에서는 뉴스 볼륨을 사용합니다. 나케이(티)주어진 주제의케이,이는 주제 번호로 태그된 단어의 총 개수로 정의됩니다.케이낮에 티,

$$\text{나케이(티)} = \sum_{\text{디그것}} \sum_{\text{승}} N(d, w, k), \quad (1)$$

어디 $N(d, w, k)$ 단어의 횟수입니다승주제로 태그됨케이문서에 등장디그리고그것)날짜에 대한 문서 세트의 표시 기능입니다.티.그림 3은 "Toyota"라는 용어에 대한 네 가지 주제에 대한 뉴스 볼륨의 시간 변화에 대한 몇 가지 예를 보여줍니다. 또한 해당 주제 분포의 상위 3개 단어를 나열합니다. 전체 설명은 지원 정보에 제공됩니다.

LDA(및 일반적인 주제 모델링)의 기본 특성은 말뭉치에 나타나는 모든 단어가 특정 주제로 태그가 지정되어 해당 특정 주제 분포에 의해 생성된 것으로 가정된다는 것입니다. 다르게 말하면, 주어진 문서의 단어가 여러 주제를 혼합하여 생성될 수 있더라도 각 단어는 정확히 하나의 주제에서 나온 것으로 가정됩니다. 이 절차를 통해 추정된 주제의 해석을 더 쉽게 이해할 수 있습니다[26]. 그러나 [27]에서 강조한 것처럼 이 구성은 다음과 같은 부정적인 결과를 가져옵니다. 왜냐하면 우리와 같은 뉴스 기록에는 "자세한 내용을 보려면 두 번 클릭하세요", "로이터 메시징 네트워크" 또는 "주요 뉴스"와 같은 반복되는 문구가 많이 있기 때문입니다. 많은 주제 분포는 이러한 반복되는 문구를 단순히 반영합니다. 이 문제를 해결하는 한 가지 방법은 원래 말뭉치에 나타나는 반복되는 문구를 제거하는 것입니다. 그러나 여기에서 분석된 방대한 양의 뉴스 레코드에서 발견된 모든 변형에 대해 잘 작동하는 알고리즘을 구성하기가 어렵기 때문에 다음 절차를 사용하여 주제 분포를 사용하여 주제를 가지치기하기로 결정했습니다. 각 주제에 대해 해당 주제 분포의 상위 6개 단어에 초점을 맞추고 이 상위 6개 단어가 원치 않는 반복 문구의 단어 집합에 포함된 경우 해당 주제를 제거했습니다(그림 1의 2-b 단계). 또한 80일 미만(2003년 1월부터 2011년 6월까지 3,103일) 동안 나타나는 모든 주제를 제거했습니다. 짧은 시간 간격으로 보고된 특정 기호 및 숫자와 같은 주제는 제외됩니다. 또한 주식 시장 활동을 설명하는 주제도 삭제했습니다. 즉, 해당 주식에 영향을 미칠 것으로 예상되는 기본 뉴스 정보에 초점을 맞추기 위해 "인기", "주식", "시장" 등과 같은 단어가 포함됩니다. 이 절차는 내인성을 필터링하는 것에 해당합니다.



무화과. 3: LDA가 학습한 선택된 주제와 "Toyota"라는 용어에 대해 방정식 (1)을 사용하여 추정된 관련 뉴스 양입니다. 이 주제에 대한 상위 3개 단어는 다음과 같습니다: (a) Toyota, 리콜, 안전; (b) 금융, 위기, 경제; (c) 일본, 생산, 지진; (d) 팀, F1, 경주.

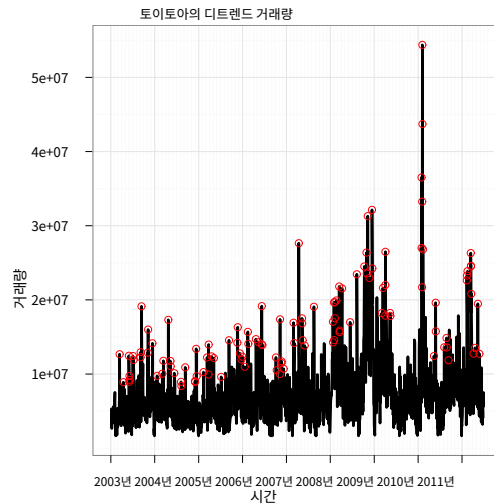
정보 흐름 및 가격 생성 프로세스의 기본 구성 요소입니다. 따라서 예를 들어 "Toyota"의 경우 원래 100개 주제 중 해당 용어와 관련된 작업에 유용한 34개 주제가 남습니다.

특정 주식의 거래량을 설명하는 데 있어 각 주제의 상대적 중요성은 양의 제약 조건이 있는 간단한 LASSO 회귀[28-30]에 의해 결정됩니다.

$$\text{볼륨}(t) = \sum_{k=1}^K \text{승케이나케이}(t) + \alpha(t), \quad (2)$$

어디볼륨(t) 해당 시점의 추세가 해제된 거래량입니다. 티. 거래량의 추세 제거는 지난 100거래일 동안 관찰된 최소 기록 거래량을 빼는 방식으로 수행됩니다(경계 값은 0이 아닌 가장 가까운 값으로 설정됨). 평균 제곱 오차를 사용한 정규 선형 회귀는 대량의 거래 활동 및 뉴스가 있을 때 뉴스 주제와 거래량 간의 관계에 대한 강력한 추정을 제공하므로 더 넓은 범위의 활동 규모가 회귀 결정에 기여할 수 있습니다. 가중치 {승케이}. LASSO 회귀 분석에 사용된 정규화 매개변수는 100회 교차 검증을 통해 정규화 매개변수의 평균값과 동일하게 선택되었습니다. 10겹 교차 검증은 전체 데이터 세트를 무작위로 10개의 하위 집합으로 나누고 10겹 교차 검증에서 각 테스트 세트의 평균 평균 제곱 오차를 측정하여 수행되었습니다. 이 절차는 추정된 정규화 매개변수의 안정성을 보장하기 위해 여러 번 수행되었습니다.

연구자들은 일반적으로 대규모(또는 "비정상") 시장 활동을 설명하는 데 관심이 있기 때문에 일일 거래량의 95번째 백분위수로 정의된 "피크 데이"에 관심을 집중합니다. 성수기보다 규모가 작습니다. 비정상성과 큰 성장을 설명하기 위해



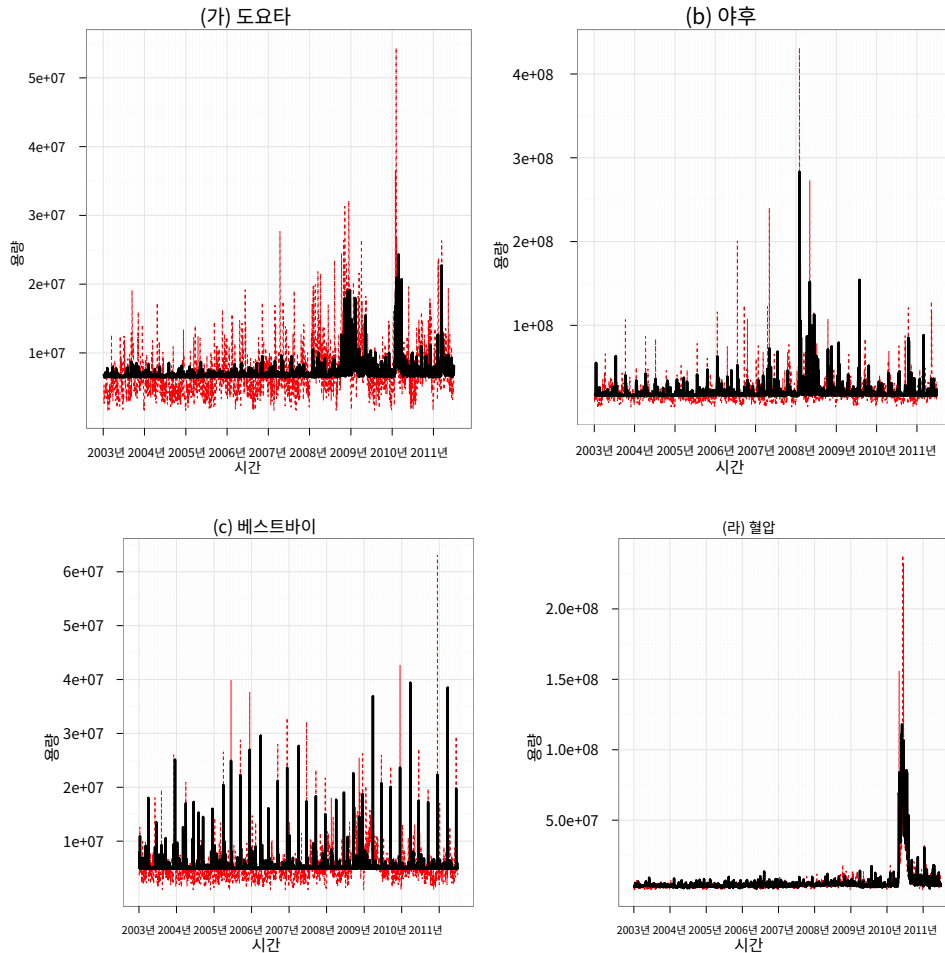
무화과. 4: 검은색 선은 2003년 1월부터 2011년 6월까지의 기간 동안 Toyota 주식의 추세가 해제된 거래량을 나타냅니다. 빨간색 점은 본문에 설명된 방법으로 선택한 "성수기"를 나타냅니다. 2003년 1월부터 2011년 6월까지 전체 기간 동안 '피크데이'는 119일이다.

무화과. 5: "Toyota"에 대해 선택된 12개 주제 목록과 예상 "설명된 볼륨 비율"(FVE). 주제 분포는 가장 자주 사용되는 상위 3~5개 단어로 요약됩니다. 주제 분포에 대한 전체 설명은 지원 정보를 참조하세요.

연구 기간(2003년 1월부터 2011년 6월) 동안의 거래량을 분석하여 전체 기간을 6개월 단위의 17개 기간으로 나누고 17개 기간 각각의 "피크 일자"를 별도로 식별했습니다. 이는 일일 거래량의 17가지 서로 다른 분포를 고려하는 것과 같습니다. 각 분포는 해당 6개월 기간 동안 거의 고정되어 있습니다. 피크 일자의 순서는 그림 4에 나와 있습니다. "Toyota"와 같은 각 용어에 대해 회귀 분석(2)을 통해 각 주제별로 설명할 수 있는 해당 예상 뉴스 볼륨의 비율(뉴스 볼륨에만 관심을 제한) "피크 데이"에 발견된 수치를 "설명된 거래량 비율(FVE)"이라고 합니다. 위에서 정의한 유용한 주제 중 일부만 총 FVE 99%에 도달하는 데 필요하다는 것을 알았습니다. 예를 들어, 케이=34개의 유용한 주제 중 12개는 "Toyota"의 총 FVE 99%에 도달하기에 충분합니다. 그림 5는 이러한 12개 주제 목록과 "Toyota"에 대한 개별 FVE를 제공합니다. 이 목록을 조사하면 우리의 절차가 합리적인 결과를 낳고 그림 3에 표시된 "Formula One"과 같은 중요하지 않은 주제가 올바르게 제거되었음을 알 수 있습니다.

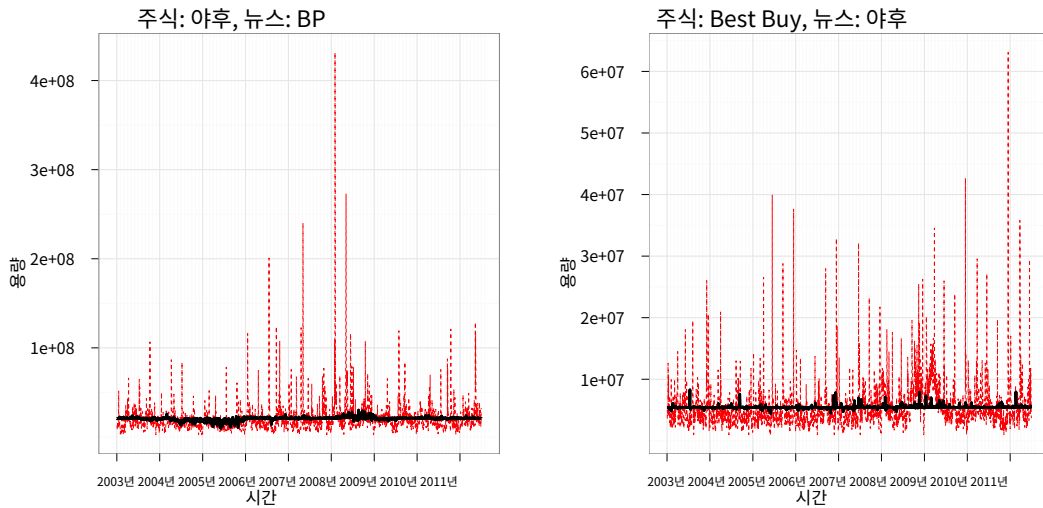
그림 6은 관측된 거래량과 회귀 분석(2)을 사용하여 피팅된 거래량을 비교합니다(잔차 기간 제외). $\hat{Y}(t)$ Toyota, Yahoo, Best Buy 및 BP의 네 가지 주식에 대해. 어떤 부분은 잘 일치하는 반면 다른 부분은 약간의 불일치를 보입니다. 주제의 회귀 품질과 설명력을 정량화하기 위해

분해를 통해 우리는 이전에 정의하고 그림 4에 표시된 "피크일"에 초점을 맞췄습니다. 예측 거래량이 특정 피크일에 관찰된 거래량의 30% 이상인 경우 성공으로 정의합니다. 2003년 1월부터 2011년 6월까지 전체 기간 동안의 총 피크 일수 중 이러한 의미에서 볼륨이 성공적으로 설명된 피크 일수의 비율을 "설명된 피크 일률(FPE)"이라고 합니다. 우리는 다음 값을 얻습니다: Toyota의 경우 $FPE=0.70$ (설명된 피크 일의 총 수는 119일 중 83일입니다), Yahoo의 경우 $FPE=0.94$ (설명된 피크 일의 총 수는 112입니다), $FPE=0.98$ (총 피크 일 수는 112입니다) 설명된 피크 일수는 Best Buy의 경우 117일)이고 BP의 경우 $FPE=0.98$ (설명된 피크 일수의 총 수는 117)입니다.

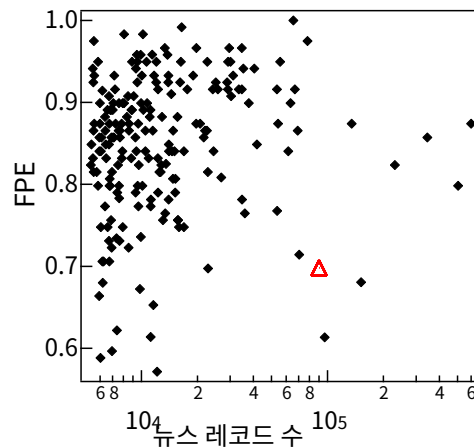


무화과. 6: (a) Toyota, (b) Yahoo, (c) Best Buy, (d) BP 등 4개 회사의 추정(빨간색 점선) 및 실제(검은색 실선) 거래량. 수케이총 "설명된 볼륨 비율"(FVE) 99%에 도달하기 위해 선택된 충분한 주제는 Toyota의 경우 12개, Yahoo의 경우 5개, Best Buy의 경우 4개, BP의 경우 12개입니다.

회귀 분석의 품질은 참조 널을 사용하여 얻은 결과와 결과를 비교하여 추가로 평가할 수 있습니다. 특히, 우리는 다른 회사와 관련된 뉴스를 교환합니다. 예를 들어 Yahoo의 거래량을 설명하기 위해 BP와 관련된 뉴스 레코드를 사용하고 회귀 분석(2)에서 추출된 주제를 사용하고(그림 7의 왼쪽 패널) Yahoo와 관련된 뉴스 레코드를 사용하여 거래를 설명합니다. Best Buy의 거래량(그림 7의 오른쪽 패널). 이는 그림 2의 흐름도에서 단계 (1)만 수정하는 것에 해당하며 다른 모든 단계는 동일하게 유지됩니다. 그림 7에서 볼 수 있듯이 음의 계수를 허용하더라도 설명력은 상당히 감소합니다. 예를 들어 두 경우 모두 FPE가 정확히 0이라는 사실에서 알 수 있습니다. 설명력의 이러한 실질적인 감소는 우리의 모든 테스트에서 발견되었으며 매일 규모로 수행된 회귀가 중요하지 않은 주제를 잘라내고 관련 주제를 식별하는 데 잘 수행된다는 것을 확인합니다. 분명히 (i) 뉴스 기록이 교환된 두 회사에 몇 가지 공통점이 있는 경우(예: 합병 협상에 참여하고 있는 경우) 또는 (ii) 전체 관찰 기간 동안 항상 정확히 같은 날짜에 수익 보고서를 공개하는 경우, 그러면 한 주식에 대해 발견된 일부 주제가 다른 주식의 거래 활동을 설명할 수 있지만, 그런 경우는 거의 없습니다.



무화과. 7: (왼쪽 패널) Yahoo 거래량을 설명할 때 BP의 주제를 사용할 때 예상 거래량과 실제 거래량을 비교합니다. (오른쪽 패널) Best Buy 거래량을 설명할 때 Yahoo의 주제를 사용할 때의 비교. 이러한 회귀 분석에서는 모델의 유연성을 높이기 위해 (비현실적인) 음수 계수도 허용했습니다. 두 경우 모두 정확히 0인 FPE로 표시된 그림 6에 제시된 것과 비교하여 회귀의 품질이 훨씬 감소했음을 알 수 있습니다.

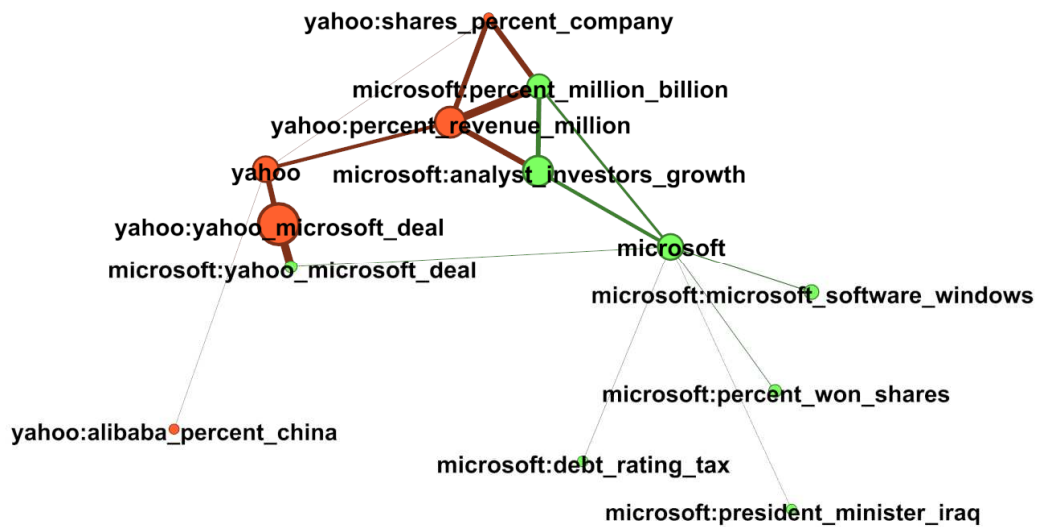


무화과. 8: 2003년 1월부터 2011년 6월까지의 기간 동안 5,000개 이상의 뉴스 기록이 있었던 S&P 500의 205개 주식에 대한 뉴스 기록 수의 함수로서 "설명된 최고점 비율"(FPE). 데이터 물론 외국 기업으로서 S&P 500의 구성 요소가 아닌 Toyota에 대한 포인트가 추가되어 빨간색 삼각형으로 표시됩니다.

III. 결과

우리는 2003년 1월부터 2011년 6월까지의 기간 동안 5,000개 이상의 뉴스 기록이 있는 S&P 500 미국 주식 지수에 상장된 205개 회사에 이전 섹션에서 소개된 방법론을 적용했습니다. 그림 8은 FPE 측정항목을 함수로 표시한 것입니다. 205개 주식에 대한 뉴스 기록의 수입입니다. 전체적으로 12개 회사를 제외하고 얻은 FPE 값은 Toyota(빨간색 삼각형으로 표시)보다 높습니다. 본 논문에서 분석된 뉴스는 모두 영어로 되어 있기 때문에 이는 아마도 국제 기업과 미국 뉴스 간의 일치보다는 미국 기업과 미국 뉴스 간의 더 나은 일치를 의미할 것입니다.

분석된 205개 미국 주식 중 987개 주제가 거래 활동에 중요한 영향을 미치는 것으로 나타났습니다. [24]에서 강조한 대로 주제 모델의 논리는 말뭉치 의미가 전역 다항 단어 분포를 공유하는 주제로 구성된다는 점을 상기하면 주제 간의 유사성을 시각화하는 편리한 방법은 다음을 사용하는 것입니다.

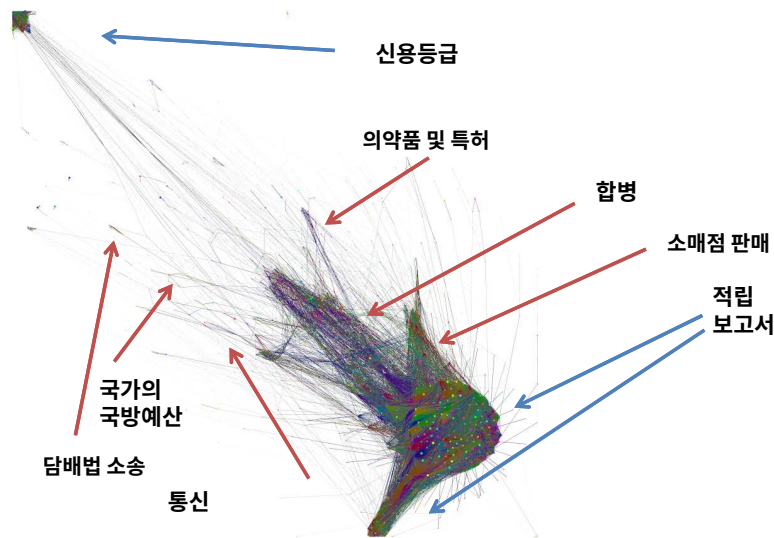


무화과. 9: Microsoft 및 Yahoo에 대해 추출된 네트워크. 여기서 노드는 주제이고 두 주제 사이의 링크는 단어 분포와 관련된 유사성 정도를 정량화합니다.

네트워크 그래프. 따라서 주제를 노드로 사용하여 네트워크를 구성하고 해당 두 주제 분포 간의 Jensen-Shannon Divergence(JSD) [31]가 0.5보다 작을 때 두 주제 사이의 링크가 존재합니다. 노드의 크기는 해당 주제의 "FVE(설명된 볼륨 비율)"에 비례하도록 설정되며 링크의 두께는 1에서 링크된 두 주제에 대한 JSD 측정항목을 뺀 값과 같습니다. 각 주제에는 회사 이름과 함께 주제 분포에 따라 정량화된 가장 자주 사용되는 상위 3개 단어로 레이블이 지정됩니다. 네트워크는 무료로 사용 가능한 소프트웨어인 Gephi[32]를 사용하여 Force Atlas 알고리즘을 사용하여 묘사됩니다.

그림 9는 Microsoft와 Yahoo 두 주식에 대한 주제 네트워크를 보여줍니다. 두 가지 모두 수익 보고서를 반영하는 주제를 가지고 있으며 잠재적인 합병 거래를 반영하는 기능을 보여줍니다. 노드 크기(FVE에 비례)를 보면 두 회사 간의 잠재적인 합병 거래가 Microsoft 주식보다 Yahoo 주식에 더 큰 영향을 미쳤다는 것을 분명히 알 수 있습니다. 이는 Yahoo가 2009년에 어려움을 겪었다는 사실과 일치합니다. 이는 우리 방법의 또 다른 유용한 특성을 보여줍니다. 즉, 두 가지 이상의 외부 영향의 영향을 정량화하고 비교할 수 있다는 것입니다.

그림 10은 우리가 집중하는 205개 주식에 대해 우리의 방법으로 추출한 모든 주제의 전체 네트워크를 보여줍니다. 네트워크는 "본토"와 더 고립된 "섬"으로 구성된 것으로 볼 수 있습니다. 본토는 수익 보고서("이익", "수익", "주", "pct(퍼센트의 줄임말)"), 신용 등급("등급", "부채,")을 반영하는 단어로 생성된 주제 간의 모든 연결로 구성됩니다. "신용"), 합병 거래("병합", "거래") 및 금융 위기("위기", "금융"). 일부 주요 "섬"을 더 잘 식별하기 위해 그림 11은 그림 10의 화살표로 표시된 영역에 대한 6개의 확대/축소를 제공합니다. 관찰된 회사 이름과 주제 분포를 나타내는 단어의 클러스터는 우리 방법이 성공적으로 추출되었음을 확인합니다. 올바른 정보. 구성된 주제 분포의 모든 단어 내용은 투자자의 관점에서 유용한 정보를 전달하는 재정적 및/또는 경제적 의미를 가지며 실제로 기업의 미래 수익에 영향을 미치는 것으로 추측될 수 있습니다. 특히 다음과 같은 단어 내용을 참조합니다: "수익 보고서", "소매업자 이익", "의약품 특허", "국방예산", "신제품", "합병 거래", "세계 경기 침체", "자연 재해, " 등등. 이로써 우리는 금융시장에 영향을 미치는 중요한 정보를 성공적으로 추출했다고 결론을 내렸습니다. "소매업자 이익", "의약품 특허", "국방 예산", "신제품", "합병 거래", "세계 경기 침체", "자연 재해" 등. 이로써 우리는 금융시장에 영향을 미치는 중요한 정보를 성공적으로 추출했다고 결론을 내렸습니다.



무화과. 10: 여기에서 연구된 S&P 500 지수의 미국 205개 기업에 대해 추출된 주제 네트워크. 텍스트에 설명된 대로 단어 분포와 관련된 유사성 정도를 정량화하는 두 주제 간의 링크가 포함되어 있습니다. 6개의 빨간색 화살표는 그림 11에서 확대된 영역을 나타냅니다.

IV. 끝 맺는 말

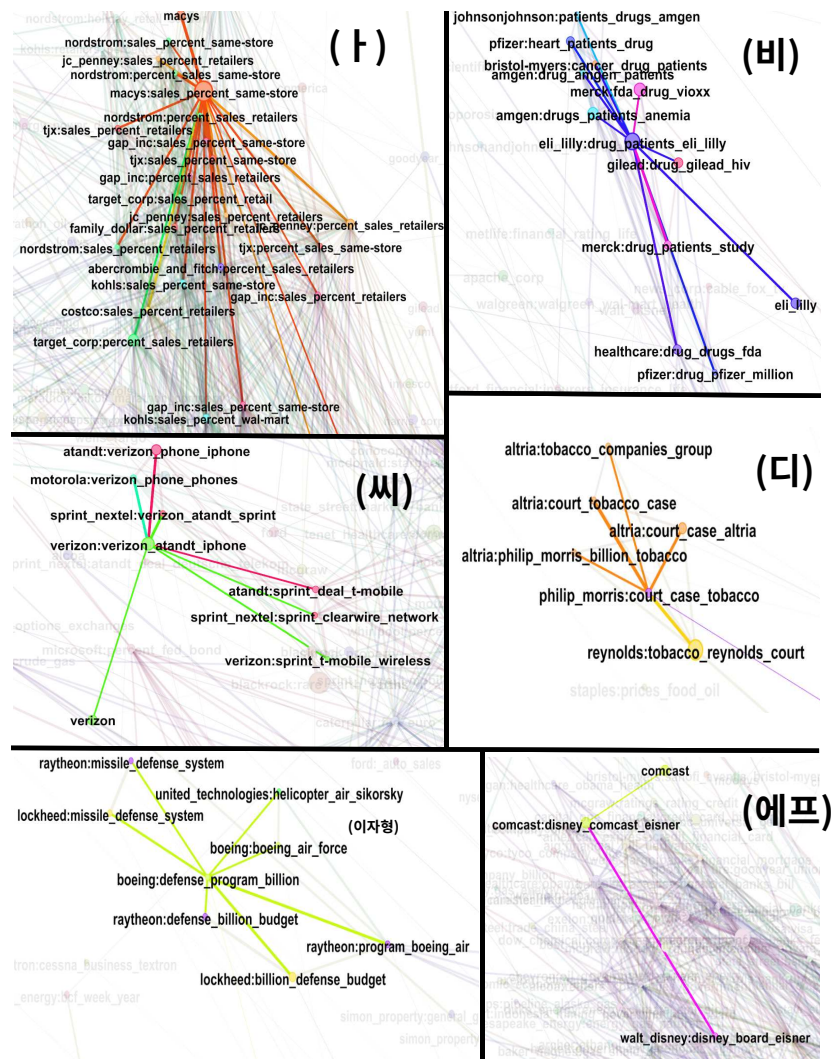
본 연구에서 우리는 Thompson Reuters가 제공한 2,400만 개 이상의 뉴스 기록과 S&P 500 지수에 포함된 205개 주요 기업의 주식 거래 활동과의 관계를 분석했습니다. 우리는 간단한 주제 모델링 기술을 사용하여 거래 활동과 "주제별" 특징으로 분해된 뉴스 정보 사이에 간단한 정규 회귀 분석을 수행함으로써 주가 변동에 영향을 미치는 뉴스의 전체 상황이 자동으로 요약될 수 있음을 보여주었습니다. 이러한 방법을 사용하여 우리는 거래 활동과 잘 동기화되는 정보 조각을 추출할 수 있었을 뿐만 아니라 동시 회귀의 보너스로 다른 방법으로는 수행하기 어려운 영향을 추정하고 정량화할 수도 있었습니다. 또한 네트워크 시각화 기술을 활용하여 주식 바구니와 관련된 뉴스 정보의 전체 풍경을 시각화하는 새로운 방법을 도입했습니다. 주제 분포를 대표하는 단어를 조사한 결과, 우리의 방법이 주식 시장에 영향을 미치는 중요한 정보를 성공적으로 추출했음을 확인했습니다.

분석포인트에서 보여준 주식시장 거래활동에 대한 뉴스의 높은 설명력은 당초 제기된 질문, 즉 어떤 뉴스가 주식시장에 영향을 미치며 주가에 어떻게 소화되는지에 대한 답변에 있다. 특히, 우리의 결과는 대량의 거래량이 뉴스의 흐름으로 설명될 수 있음을 보여준다. 이러한 의미에서, 우리의 결과는 뉴스가 진정으로 참신하고 관련 금융 정보를 제공한다면 “과잉 거래”가 없음을 시사합니다.

지연 효과나 보다 정교한 비선형 동역학을 고려할 필요가 없는 간단한 방법론이 성공한 이유 중 하나는 아마도 높은 신호 대 잡음 비율을 가져온 뉴스 소스의 고품질 때문일 것입니다. 구체적으로 우리가 사용한 뉴스는 상당한 구독료를 지불하여 수집 회사에 인센티브를 제공하는 전문 투자자를 위해 수집되었습니다. 우리의 연구는 트윗이나 블로그와 같은 다른 표준 텍스트 정보와 비교하여 이러한 전문 금융 소스의 탁월한 관련성을 확인합니다. 표준 신문에서 제공되는 것과 비교하여 뉴스 레코드 수 측면에서 우리 데이터베이스의 크기도 금융 시장의 거래 활동에 영향을 미치는 중요한 주제를 추출하는 데 필수적이었습니다. 결론적으로, 우리는 우리의 결과가 금융시장과 관련된 뉴스 정보에서 비롯된 금융시장에 대한 외부 영향의 주요 원천을 요약한다고 믿습니다. 거래 활동을 설명하는 것 이상의 또 다른 과제는 광범위한 뉴스, 주제 및 해당 네트워크를 사용하여 가격 책정 및 재무 평가 전반을 설명하는 것입니다. 이는 향후 작업으로 남겨둡니다.

감사의 말

저자는 이 작업에 관한 유용한 토론과 의견을 주신 Vladimir Filimonov, Georges Harras 및 Ryan Woodard에게 감사드립니다. 히사노 료헤이는 '장기외국어장학금-2010'이라는 장학금을 통해 일본학생지원기구로부터 일부 자금을 지원받고 있습니다.



무화과. 11: 그림 10에 표시된 주제 네트워크에서 화살표로 표시된 "섬"의 6개 확대. 두 주제 사이의 링크는 단어 분포와 관련된 유사성 정도를 정량화합니다. 각 노드에는 회사 이름과 주제 분포로 정량화된 가장 자주 사용되는 상위 3개 단어가 함께 표시됩니다. 노드의 크기는 해당 주제의 "FVE(설명된 볼륨 비율)"에 비례하도록 설정되며 링크의 두께는 1에서 링크된 두 주제에 대한 JSD 측정항목을 뺀 값과 같습니다. 패널 (a)는 의류 회사의 소매 판매와 관련된 네트워크를 보여줍니다. (b) 의약품 및 특허와 관련된 패널; (c) 통신 사업의 제품과 관련된 패널; (d) 담배법 소송과 관련된 패널; (e) 국방예산과 관련된 패널; 2004년 Comcast Disney 합병 가능성과 관련된 패널 (f).

V. 참고문헌

- [1] Cutler D, Poterba J, Summers L (1989), 무엇이 주가를 움직이는가?, Journal of Portfolio Management 15:4-12.
- [2] McQueen G, Roley V (1993), 주가, 뉴스 및 비즈니스 조건, Fin. 연구 6(3):683-707.
- [3] Fleming M, Remolona E (1997), 무엇이 채권 시장을 움직이는가?, FRBNY 경제 정책 검토 3 (4):31-50.
- [4] Fair RC(2002), 시장을 뒤흔든 사건, Journal of Business 75(4):713-731.
- [5] Joulin A, Lefevre A, Grunberg D, Bouchaud JP (2008) 주가 급등: 뉴스와 거래량이 작은 역할을 함, Wilmott Magazine, 9월/10월:46.
- [6] Erdogan O 및 Yezge A (2009), 주식 시장에 무소식이라는 소식, Quantitative Finance 9 (8):897-909.

- [7] Shiller RJ(1981), 주가가 배당금의 후속 변화로 정당화되기에는 너무 많이 움직이는가?, American Economic Review, 71:421-436.
- [8] LeRoy SF 및 Porter RD(1981), 현재 가치 관계: 내재 분산 경계를 기반으로 한 테스트, Econometrica: Journal of the Econometric Society:555-574.
- [9] LeRoy SF(2008), 초과 변동성 테스트, The New Palgrave 경제학 사전, The New Palgrave 경제학 사전, 2판.
- [10] Soros G, 금융의 연금술, Wiley Investment Classics, Wiley(2003년 7월 29일).
- [11] Filimonov V, Sornette D, 금융 시장의 반사성 정량화: 플래시 충돌 예측을 향하여, 물리적 검토 E 85(5): 056108(2012).
- [12] Reinhart CM, Rogoff K, 이번에는 다릅니다: 8세기의 재정적 어리석음, 프린스턴 대학 출판부(2011년 7월 18일).
- [13] Zhi D, Engelberg J, Gao P (2011), 관심을 찾아서, The Journal of Finance, Vol LXVI, No. 5, 2011년 10월.
- [14] Dellavigna S 및 Pollet JM(2009), 투자자 부주의 및 금요일 수익 발표, The Journal of Finance, Vol 64, Issue 2:709-749.
- [15] Engelberg J 및 Parsons C(2011), 금융 시장에서 미디어의 인과적 영향, The Journal of Finance, Vol 66, Issue 1:67-97.
- [16] Tetlock P(2007), 투자 심리에 대한 내용 제공: 주식 시장에서 미디어의 역할, Journal of Finance 62:1139-1168.
- [17] Bollen J, Mao H 및 Zeng XJ (2011), 트위터 기분은 주식 시장을 예측합니다, Journal of Computational Science 2:1-8.
- [18] Uhl MW(2011), 로이터 감정 및 주식 수익률, <http://ssrn.com/abstract=1924867>.
- [19] Gurun UG, Butler AW(2012), 과대 광고를 믿지 마세요: 지역 미디어 경사, 지역 광고 및 기업 가치, The Journal of Finance Volume 67, Issue 2:561-598, 2012년 4월.
- [20] Ito T, Roley VV (1987), 미국과 일본의 뉴스 어느 것이 엔/달러 환율을 움직이는가?, Journal of Monetary Economics 19:255-277.
- [21] Mizuno T, Takei K, Ohnishi T, Watanabe T (2012), 비즈니스 뉴스의 시간적 및 상호 상관 관계, 이론 물리학의 발전 보충 자료, No.194:181-192.
- [22] Blei DM, Ng AY, Jordan MI (2003), 잠재 Dirichlet 할당, 기계 학습 연구 저널 3:993-1022.
- [23] Griffiths T, Steyvers M (2004) 과학적 주제 찾기, PNAS 101:5228-5235.
- [24] Koller D 및 Friedman N (2009), 확률 그래픽 모델, MIT 언론.
- [25] Smola A, Narayanamurthy S(2010), 병렬 주제 모델을 위한 아키텍처, VLDB 기부금 회보, Vol. 3, 1호.
- [26] Hinton GE 및 Salakhutdinov RR(2010), 심층 생성 모델 학습을 통한 문서의 이진 코드 발견, 인지 과학 주제(2010),1-18.
- [27] Mimno D 및 Blei D(2011), 주제 모델에 대한 베이저안 확인, 자연어 처리의 경험적 방법, 2011.
- [28] Tibshirani R(1996), 올가미를 통한 회귀 수축 및 선택, J. Royal. 통계학자. Soc B., Vol. 58, No. 1:267-288.
- [29] Hastie T, Tibshirani R, Friedman J (2009), 통계 학습의 요소: 데이터 마이닝, 추론 및 예측, Springer, 2판.
- [30] Goeman JJ (2010), Cox 비례 위험 모델의 CL-1 페널티 추정, Biometrical Journal 52 (1):70-84.
- [31] Endres DM, Schindelin JE(2003), 확률 분포에 대한 새로운 측정법, IEEE Trans. 정보 이론 49(7):1858-1860.
- [32] 게피, <https://gephi.org/>

6. (2-B) 단계의 정확한 가지치기 절차

가. 반복되는 문구를 반영하는 주제

주제 분포의 상위 6개 단어에 아래 나열된 단어 중 두 개 이상이 포함된 경우 주제 분포를 반복되는 뉴스 문구만 반영하는 것으로 정의했습니다. 즉, "인기", "뉴스", "로이터넷", "메시징", "더블", "클릭", "단어", "이동", "업데이트", "스토리", "언론", "회사"입니다. , "서비스" 및 "상표".

나. 시장 용어를 반영한 주제

주제 분포의 상위 6개 단어에 아래 나열된 단어 중 하나 이상이 포함된 경우 주제 분포를 시장 뉴스만을 반영하는 것으로 정의했습니다. 즉, "주식", "주식", "지수", "선물", "시장", "장미", "하락", "매수", "매도", "나스닥", "NYSE", "다우"입니다. , "target", 모든 종류의 숫자 및 월 이름(예: "jan", "january",...).

VII. "TOYOTA"에 대해 학습한 주제에 대한 전체 설명

"Toyota"에 대해 학습된 100개의 주제 분포 전체에 대해 상위 10개의 단어를 표시합니다. 괄호 안의 숫자는 주제 분포에서 각 단어의 확률을 나타냅니다. 또한 "Toyota"에 대해 학습된 모든 100개 주제에 대한 뉴스 볼륨의 시간 변화를 보여줍니다(그림 1).

주제 0: (에탄올, 0.006) (석유, 0.006) (로이터, 0.005) (정부, 0.004) (설탕, 0.004) (경찰, 0.004) (장관, 0.003) (대통령, 0.003) (태국, 0.003) (이라크, 0.003)

주제 1: (도요타, 0.069) (회상, 0.020) (안전성, 0.015) (자동차, 0.012) (회상, 0.008) (로이터, 0.007) (문제, 0.007) (도요다, 0.006) (가속, 0.005) (백만, 0.005)

주제 2: (통보, 0.025) (9월, 0.016) (8월, 0.012) (기업, 0.010) (2월, 0.010) (단체, 0.009) (3월, 0.008) (취득, 0.007) (공동, 0.007) (10월, 0.007)

주제 3: (퍼센트, 0.038) (엔, 0.029) (니케이, 0.013) (1, 0.010) (주식, 0.010) (주식, 0.009) (메이커, 0.009) (투자자, 0.008) (장미, 0.007) (이익, 0.007)

주제 4: (pct, 0.106) (100, 0.053) (incr, 0.043) (kfw, 0.027) (50, 0.020) (eib, 0.020) (rabobank, 0.013) (5125, 0.010) (70, 0.008) (200, 0.008)

주제 5: (아시아, 0.010) (주식, 0.009) (아시아, 0.008) (홍콩, 0.008) (결과, 0.008) (수익, 0.008) (상위, 0.007) (데이터, 0.007) (싱가포르, 0.006) (시장, 0.006)

주제 6: (% , 0.110) (100, 0.064) (50, 0.030) (incr, 0.016) (9998, 0.016) (475, 0.014) (55, 0.013) (525, 0.012) (단위, 0.009) (200, 0.008)

주제 7: (엔, 0.044) (퍼센트, 0.033) (gmt, 0.016) (이익, 0.009) (1, 0.007) (십억, 0.007) (메이커, 0.007) (회사, 0.006) (법인, 0.006) (보고서, 0.006)

주제 8: (% , 0.046) (엔, 0.040) (닛케이, 0.014) (하락, 0.014) (주, 0.011) (수출, 0.009) (주식, 0.009) (금융, 0.008) (시장, 0.007) (상실, 0.007)

주제 9: (중국, 0.029) (자동차, 0.017) (연도, 0.012) (시장, 0.011) (자동차, 0.011) (퍼센트, 0.011) (자동차, 0.010) (중국, 0.009) (인도, 0.009) (000, 0.009)

주제 10: (% , 0.105) (100, 0.048) (50, 0.046) (incr, 0.045) (575, 0.024) (bng, 0.020) (60, 0.017) (55, 0.013) (200, 0.013) (550, 0.012)

주제 11: (% , 0.095) (100, 0.037) (200, 0.026) (incr, 0.020) (tranche, 0.014) (55, 0.014) (625, 0.012) (50, 0.012) (deal, 0.011) (par, 0.011)

주제 12: (뉴스, 0.016) (퍼센트, 0.016) (클릭, 0.011) (관련, 0.011) (로이터, 0.010) (더블, 0.009) (주, 0.009) (유로, 0.008) (독일, 0.008) (메시지, 0.007)

주제 13: (pct, 0.073) (frn, 0.040) (par, 0.025) (200, 0.023) (incr, 0.020) (300, 0.016) (100, 0.015) (150, 0.014) (50, 0.012) (40, 0.009)

주제 14: (현대, 0.034) (퍼센트, 0.023) (원, 0.022) (매출액, 0.015) (s-korea, 0.014) (시장, 0.009) (주, 0.009) (기아, 0.009) (서울, 0.008) (연도, 0.008)

주제 15: (할당됨, 0.074) (시리즈, 0.037) (신뢰, 0.023) (flind, 0.022) (ptc, 0.020) (flindso, 0.017) (대출, 0.017) (재확인, 0.014) (lt, 0.012) (bbbbind, 0.011)

주제 16: (뉴스, 0.050) (상위, 0.038) (로이터, 0.022) (일본, 0.018) (주식, 0.017) (방문, 0.014) (닛케이, 0.013) (데이터, 0.013) (30, 0.013) (시장, 0.012)

주제 17: (oct, 0.044) (dec, 0.042) (nov, 0.037) (2009, 0.029) (zar, 0.023) (2010, 0.021) (sep, 0.021) (nzd, 0.018) (world-bank, 0.018) (노무라, 0.018)

주제 18: (퍼센트, 0.043) (엔, 0.034) (주, 0.013) (니케이, 0.010) (금융, 0.010) (주식, 0.008) (장미, 0.008) (은행, 0.008) (은행, 0.008) (그룹, 0.007)

주제 19: (pct, 0.108) (100, 0.062) (incr, 0.043) (eib, 0.028) (kfw, 0.020) (65, 0.017) (70, 0.015) (60, 0.014) (200, 0.014) (2009년 7월, 0.013)

주제 20: (경제, 0.012) (일본, 0.012) (퍼센트, 0.011) (엔, 0.011) (boj, 0.011) (일본, 0.010) (일본, 0.008) (연도, 0.008) (시장, 0.007) (정부, 0.006)

주제 21: (나, 0.051) (홍콩, 0.013) (중국, 0.011) (기업, 0.010) (그룹, 0.009) (백만, 0.008) (기업, 0.007) (2008, 0.007) (계획, 0.007) (회사, 0.007)

주제 22: (퍼센트, 0.032) (엔, 0.022) (주식, 0.009) (투자자, 0.009) (시장, 0.008) (니케이, 0.008) (일본, 0.008) (주, 0.008) (도쿄, 0.008) (메이커, 0.005)

주제 23: (07, 0.029) (% , 0.019) (용어, 0.016) (08, 0.016) (06, 0.015) (년, 0.014) (2003, 0.013) (2006, 0.013) (12월, 0.012) (10월, 0.012)

주제 24: (엔, 0.045) (퍼센트, 0.038) (gmt, 0.016) (십억, 0.012) (이익, 0.011) (주, 0.009) (연도, 0.006) (하락, 0.006) (2, 0.005) (메이커, 0.005)

주제 25: (가격, 0.031) (목표, 0.030) (예상, 0.022) (이슈, 0.021) (기업, 0.020) (아웃퍼폼, 0.016) (etr, 0.016) (퍼센트, 0.015) (포인트, 0.015) (등급, 0.013)

주제 26: (na, 0.029) (zar, 0.024) (2012, 0.022) (aud, 0.022) (9월, 0.020) (2011, 0.019) (daiwa, 0.019) (7월, 0.017) (6월, 0.017) (8월, 0.016)

주제 27: (1, 0.073) (jpy, 0.037) (2, 0.024) (3, 0.016) (4, 0.014) (법인, 0.012) (5, 0.010) (h1, 0.010) (6, 0.008) (krw), 0.006)

주제 28: (퍼센트, 0.016) (모터, 0.014) (연도, 0.013) (판매, 0.012) (1, 0.012) (도요타, 0.011) (일본, 0.011) (모터, 0.010) (자동차, 0.010) (중형, 0.010)

주제 29: (팀, 0.014) (f1, 0.011) (레이스, 0.010) (도요타, 0.007) (프릭스, 0.007) (시즌, 0.007) (그랜드, 0.007) (팀, 0.006) (로이터, 0.006) (연도, 0.006)

주제 30: (% , 0.112) (100, 0.083) (60, 0.027) (2004, 0.016) (50, 0.014) (625, 0.013) (2002, 0.012) (2003, 0.011) (75, 0.011) (250, 0.010)

주제 31: (퍼센트, 0.031) (주, 0.016) (로이터, 0.014) (rm, 0.011) (reutersnet, 0.011) (메시징, 0.011) (유로, 0.006) (회사, 0.005) (지수, 0.005) (클릭, 0.005)

주제 32: (4월, 0.028) (채권, 0.020) (엔, 0.019) (10, 0.016) (9월, 0.016) (날짜, 0.016) (미즈호, 0.016) (초, 0.016) (노무라, 0.012) (20, 0.012)

주제 33: (퍼센트, 0.063) (주, 0.028) (지수, 0.024) (adrs, 0.019) (장미, 0.014) (하락, 0.013) (선행, 0.012) (뉴욕, 0.011) (센트, 0.009) (유럽, 0.009)

주제 34: (금, 0.012) (가격, 0.011) (시장, 0.009) (선물, 0.009) (퍼센트, 0.008) (1, 0.008) (고무, 0.007) (jgb, 0.007) (유로, 0.006) (수요, 0.005)

주제 35: (년, 0.005) (중국, 0.005) (명, 0.004) (일본인, 0.004) (근로자, 0.004) (일본, 0.004) (년, 0.003) (시간, 0.003) (사업, 0.003) (일, 0.002)

주제 36: (퍼센트, 0.030) (판매, 0.019) (자동차, 0.018) (시장, 0.016) (연도, 0.012) (브랜드, 0.012) (유럽, 0.010) (자동차, 0.009) (유럽, 0.009) (등록, 0.007)

주제 37: (2010, 0.035) (gmt, 0.033) (제목, 0.030) (reuters, 0.030) (timedate, 0.029) (설명, 0.029) (2월, 0.022) (toyota, 0.013) (wed, 0.012) (내부자, 0.010)

주제 38: (% , 0.062) (frn, 0.048) (100, 0.035) (incr, 0.019) (150, 0.013) (200, 0.013) (500, 0.011) (50, 0.011) (425, 0.010) (2006년 1월), 0.010)

주제 39: (클릭, 0.018) (더블, 0.014) (퍼센트, 0.009) (보고, 0.008) (시장, 0.007) (석유, 0.007) (주식, 0.006) (그룹, 0.006) (최신, 0.006) (유로, 0.005)

주제 40: (pct, 0.048) (frn, 0.043) (deal, 0.017) (incr, 0.016) (500, 0.016) (par, 0.012) (100, 0.012) (tranche, 0.012) (250, 0.011) (10b, 0.010)

주제 41: (부채, 0.015) (등급, 0.015) (백만, 0.012) (moody's, 0.009) (등급, 0.008) (재정, 0.008) (채권, 0.007) (세금, 0.007) (주, 0.006) (신용, 0.006)

주제 42: (% , 0.094) (100, 0.041) (200, 0.028) (incr, 0.017) (625, 0.016) (55, 0.015) (tranche, 0.015) (575, 0.012) (multi, 0.011) (deal), 0.011)

주제 43: (퍼센트, 0.043) (엔, 0.036) (gmt, 0.011) (주식, 0.009) (장미, 0.008) (jal, 0.007) (nikkei, 0.007) (1, 0.006) (도요타, 0.006) (수출업체, 0.005)

주제 44: (pct, 0.076) (incr, 0.031) (frn, 0.021) (reoffer, 0.018) (100, 0.017) (eib, 0.013) (150, 0.013) (multi, 0.012) (tranche, 0.012) (200, 0.012)

주제 45: (% , 0.107) (incr, 0.019) (100, 0.017) (250, 0.012) (25, 0.011) (200, 0.011) (150, 0.011) (500, 0.009) (275, 0.008) (거래, 0.008)

주제 46: (top, 0.021) (news, 0.015) (auf, 0.009) (weitere, 0.009) (der, 0.007) (und, 0.007) (index, 0.007) (firmen, 0.007) (unternehmen, 0.007) (die, 0.006)

Topic 47: (엔, 0.041) (퍼센트, 0.037) (gmt, 0.015) (이익, 0.009) (주, 0.008) (메이커, 0.007) (가격, 0.006) (도요타, 0.006) (예측, 0.006) (주식), 0.006)

주제 48: (rec, 0.059) (000, 0.038) (sm, 0.032) (cargill, 0.019) (sb, 0.018) (mz, 0.015) (bunge, 0.015) (china, 0.014) (sbo, 0.012) (도로, 0.009)

주제 49: (pct, 0.108) (100, 0.054) (incr, 0.043) (kfw, 0.027) (eib, 0.020) (50, 0.020) (rabobank, 0.013) (5125, 0.010) (200, 0.008) (70, 0.008)

주제 50: (1, 0.016) (판매, 0.010) (자동차, 0.008) (2, 0.008) (% , 0.007) (na, 0.007) (자동차, 0.007) (3, 0.006) (4, 0.005) (10, 0.005)

Topic 51: (2월, 0.028) (채권, 0.023) (9월, 0.022) (엔, 0.020) (10, 0.017) (na, 0.015) (날짜, 0.014) (쿠폰, 0.012) (미즈호, 0.012) (신용), 0.011)

주제 52: (백만, 0.078) (10년, 0.021) (십억, 0.018) (채권, 0.017) (회사, 0.016) (amt, 0.016) (rtgs, 0.016) (matdebt, 0.015) (mgrs, 0.015) (매출, 0.015)

주제 53: (pct, 0.109) (incr, 0.049) (rabobank, 0.037) (50, 0.033) (100, 0.028) (75, 0.020) (45, 0.017) (cba, 0.013) (60, 0.013) (kfw, 0.012)

주제 54: (gm, 0.029) (포드, 0.017) (크라이슬러, 0.014) (자동차, 0.011) (십억, 0.008) (자동차 제조업체, 0.008) (자동차 제조업체, 0.006) (파산, 0.006) (노동자, 0.006) (일반-모터, 0.005)

주제 55: (pct, 0.086) (par, 0.028) (재제안, 0.025) (10b, 0.023) (frn, 0.022) (500, 0.018) (300, 0.010) (incr, 0.009) (200, 0.008) (475, 0.008)

주제 56: (하이브리드, 0.017) (자동차, 0.016) (자동차, 0.013) (전기, 0.013) (자동차, 0.011) (도요타, 0.010) (연료, 0.008) (기술, 0.007) (배터리, 0.006) (배터리, 0.005)

주제 57: (단어, 0.035) (이동, 0.022) (업데이트, 0.019) (기대, 0.013) (1, 0.011) (2, 0.009) (비즈니스, 0.006) (600, 0.006) (gmt, 0.006) (3, 0.005)

Topic 58: (엔, 0.043) (퍼센트, 0.035) (gmt, 0.017) (십억, 0.012) (이익, 0.011) (주, 0.009) (연도, 0.008) (예측, 0.006) (메이커, 0.006) (백만, 0.006)

주제 59: (퍼센트, 0.034) (엔, 0.033) (니케이, 0.021) (시장, 0.011) (주식, 0.010) (장미, 0.008) (투자자, 0.007) (수출자, 0.007) (평균, 0.007) (3, 0.006)

주제 60: (금융, 0.013) (위기, 0.011) (경제, 0.010) (글로벌, 0.009) (삭감, 0.009) (은행, 0.008) (경제, 0.008) (경기 침체, 0.007) (십억, 0.007) (퍼센트, 0.006)

주제 61: (일본, 0.026) (생산, 0.020) (지진, 0.011) (공장, 0.011) (공장, 0.009) (원자력, 0.009) (전력, 0.009) (부품, 0.009) (지진, 0.008) (공급, 0.008)

주제 62: (pct, 0.110) (incr, 0.060) (50, 0.049) (100, 0.032) (rabobank, 0.025) (60, 0.023) (bng, 0.016) (575, 0.015) (kfw, 0.015) (jan2015), 0.010)

주제 63: (고정, 0.019) (탭, 0.015) (웨스트팩, 0.014) (채권, 0.014) (mln, 0.013) (rmbs, 0.012) (undisc, 0.012) (frn, 0.012) (유로본드, 0.012) (aaaa1aa, 0.011)

주제 64: (03, 0.030) (02, 0.024) (% , 0.018) (용어, 0.015) (2002, 0.015) (2001, 0.014) (1997, 0.013) (년, 0.013) (1999, 0.012) (12월), 0.012)

주제 65: (pct, 0.107) (100, 0.059) (incr, 0.041) (kfw, 0.025) (eib, 0.023) (60, 0.013) (200, 0.012) (65, 0.011) (70, 0.011) (2009년 7월), 0.010)

주제 66: (자동차, 0.012) (판매, 0.011) (보다, 0.009) (이야기, 0.008) (gm, 0.007) (자동차, 0.007) (유럽, 0.006) (도요타, 0.005) (뉴스, 0.005)

주제 67: (퍼센트, 0.032) (달러, 0.013) (주식, 0.010) (시장, 0.009) (지수, 0.008) (주, 0.008) (투자자, 0.008) (아시아, 0.008) (하락, 0.007) (엔, 0.007)

주제 68: (05, 0.025) (04, 0.024) (% , 0.019) (용어, 0.016) (06, 0.015) (년, 0.014) (2003, 0.013) (2004, 0.013) (03, 0.013) (10월, 0.013)

주제 69: (닛케이, 0.019) (일본, 0.012) (시장, 0.011) (주식, 0.010) (엔, 0.008) (비즈니스, 0.007) (십억, 0.006) (금요일, 0.006) (퍼센트, 0.006) (일일), 0.005)

주제 70: (2010, 0.036) (8월, 0.034) (world-bank, 0.030) (2009, 0.029) (na, 0.026) (jun, 0.024) (rand, 0.022) (mizuho, 0.020) (7월, 0.019) (0.019.02)

주제 71: (로이터, 0.007) (회사, 0.006) (비즈니스, 0.006) (백만, 0.006) (십억, 0.005) (언론, 0.004) (스토리, 0.004) (회사, 0.004) (소니, 0.004) (임원, 0.003)

주제 72: (엔, 0.047) (퍼센트, 0.030) (십억, 0.015) (gmt, 0.015) (이익, 0.009) (연도, 0.008) (백만, 0.007) (주, 0.007) (회사, 0.007) (japans, 0.006)

주제 73: (퍼센트, 0.030) (엔, 0.023) (주식, 0.011) (니케이, 0.011) (투자자, 0.010) (수익, 0.010) (시장, 0.009) (장미, 0.006) (1, 0.006) (주식, 0.006)

주제 74: (엔, 0.045) (퍼센트, 0.038) (gmt, 0.016) (이익, 0.009) (십억, 0.008) (주, 0.007) (1, 0.006) (백만, 0.005) (메이커, 0.005) (예측), 0.005)

주제 75: (nihon, 0.012) (keizai, 0.012) (news, 0.011) (페이지, 0.011) (asahi, 0.011) (mainichi, 0.010) (yomiuri, 0.010) (japan, 0.010) (australian, 0.007) (japanese, 0.007)

주제 76: (% , 0.097) (100, 0.037) (incr, 0.035) (50, 0.035) (575, 0.026) (bng, 0.017) (55, 0.017) (200, 0.014) (frn, 0.011) (550, 0.011)

주제 77: (퍼센트, 0.038) (엔, 0.027) (닛케이, 0.013) (시장, 0.011) (주식, 0.010) (주식, 0.010) (1, 0.008) (투자자, 0.007) (장미, 0.006) (하락, 0.006)

주제 78: (엔, 0.048) (퍼센트, 0.043) (gmt, 0.020) (주, 0.011) (이익, 0.009) (십억, 0.009) (장미, 0.008) (1, 0.008) (니케이, 0.006) (떨어짐, 0.006)

주제 79: (aaaaaa, 0.023) (rmbs, 0.019) (fixed, 0.018) (euro, 0.014) (100, 0.011) (frn, 0.011) (td, 0.010) (sec, 0.009) (aaa, 0.009) (aaaaaaaa, 0.008)

주제 80: (엔, 0.023) (퍼센트, 0.018) (십억, 0.010) (gmt, 0.008) (이익, 0.008) (이익, 0.008) (예측, 0.006) (연도, 0.005) (회사, 0.005) (컷, 0.005)

주제 81: (도요타, 0.022) (퍼센트, 0.021) (연도, 0.019) (000, 0.015) (판매, 0.014) (공장, 0.012) (자동차, 0.012) (백만, 0.010) (생산, 0.010) (대, 0.009)

주제 82: (퍼센트, 0.029) (포인트, 0.015) (지수, 0.012) (뉴욕, 0.010) (금, 0.009) (1, 0.007) (폐쇄, 0.007) (가격, 0.006) (런던, 0.006) (조, 0.006)

주제 83: (% , 0.106) (100, 0.059) (incr, 0.035) (425, 0.017) (50, 0.015) (온타리오, 0.014) (200, 0.013) (525, 0.011) (40, 0.010) (dec2007, 0.010)

주제 84: (판매량, 0.060) (퍼센트, 0.032) (포드, 0.017) (연도, 0.013) (도요타, 0.011) (자동차, 0.011) (시장, 0.010) (gm, 0.009) (차량, 0.009) (백만 대), 0.008)

주제 85: (이익, 0.028) (엔, 0.026) (십억, 0.025) (퍼센트, 0.023) (연도, 0.023) (도요타, 0.018) (매출, 0.015) (운영, 0.013) (예측, 0.011) (수익, 0.009)

주제 86: (엔, 0.045) (퍼센트, 0.035) (이익, 0.016) (gmt, 0.015) (십억, 0.014) (연도, 0.010) (예측, 0.008) (메이커, 0.007) (매출, 0.006) (그룹, 0.006)

주제 87: (% , 0.104) (100, 0.031) (incr, 0.030) (200, 0.019) (50, 0.018) (rabobank, 0.016) (ge-capital, 0.015) (tranche, 0.014) (150, 0.011) (2012년 2월, 0.011)

주제 88: (독일, 0.021) (1, 0.017) (페라리, 0.017) (2, 0.014) (3, 0.014) (르노, 0.014) (도요타, 0.014) (영국, 0.014) (영국, 0.014) (4, 0.014)

주제 89: (rmbs, 0.024) (고정, 0.021) (frn, 0.017) (aaaaaa-, 0.013) (유로본드, 0.013) (aaaaaaaaa, 0.012) (anz, 0.010) (탭, 0.009) (200, 0.008) (2007-1, 0.008)

주제 90: (엔, 0.038) (퍼센트, 0.028) (십억, 0.013) (gmt, 0.012) (이익, 0.008) (그룹, 0.008) (메이커, 0.007) (연도, 0.007) (주, 0.007) (백만, 0.006)

주제 91: (자동차, 0.010) (피아트, 0.010) (유로, 0.008) (르노, 0.008) (퍼센트, 0.008) (유럽, 0.005) (로이터, 0.005) (폭스바겐, 0.005) (독일, 0.005) (산업, 0.005)

주제 92: (퍼센트, 0.020) (엔, 0.020) (전자제품, 0.015) (십억, 0.011) (nec, 0.009) (연도, 0.009) (메이커, 0.008) (산요, 0.007) (주, 0.007) (기업, 0.007), 0.007)

주제 93: (단어, 0.032) (이동, 0.021) (업데이트, 0.016) (1, 0.010) (기대, 0.009) (2, 0.007) (런던, 0.006) (600, 0.006) (gmt, 0.005) (3, 0.005)

주제 94: (말, 0.036) (이동, 0.031) (pix, 0.012) (업데이트, 0.011) (tv, 0.009) (1, 0.006) (600, 0.006) (대통령, 0.005) (2, 0.004) (700, 0.004)

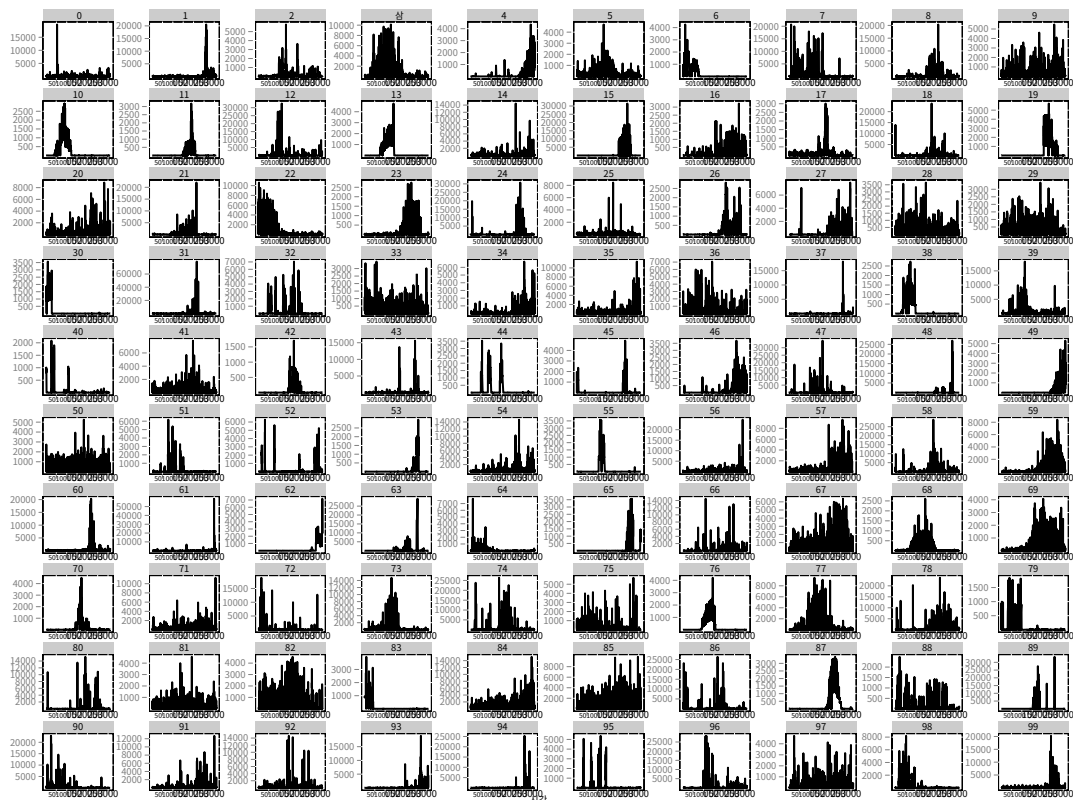
주제 95: (12월, 0.037) (11월, 0.037) (채권, 0.021) (10, 0.018) (엔, 0.017) (3월, 0.013) (노무라, 0.013) (미즈호, 0.013) (날짜, 0.012) (na, 0.011)

주제 96: (로이터, 0.019) (메시징, 0.017) (퍼센트, 0.017) (주식, 0.016) (클릭, 0.014) (주, 0.014) (reutersnet, 0.011) (rm, 0.011) (더블, 0.010) (주식, 0.010)

주제 97: (철강, 0.065) (퍼센트, 0.017) (일본, 0.016) (가격, 0.015) (수요, 0.012) (연도, 0.010) (가격, 0.009) (j5401t, 0.008) (일본제철, 0.007) (일본어, 0.007)

주제 98: (채권, 0.028) (엔, 0.021) (10, 0.017) (na, 0.014) (7월, 0.014) (날짜, 0.013) (10월, 0.012) (20, 0.010) (법인, 0.010) (mizuho, 0.010)

주제 99: (% , 0.035) (기업, 0.024) (수익, 0.020) (보고, 0.019) (도요타, 0.013) (기대, 0.013) (연도, 0.010) (기업, 0.010) (결과, 0.010) (주식), 0.010)



무화과. 12: "Toyota"에 대해 학습된 모든 주제에 대한 뉴스 볼륨의 시간 변화. 가로축은 일수(2003년 1월~2011년 6월)를 나타내고 세로축은 정규화되지 않은 거래량을 나타냅니다.