

소셜 데이터 분석을 통한
벤처기업의 안정성에 영향을 미치는
외부환경요인 분석:
외식 프랜차이즈를 기반으로

KAIST 전산학부 위희주

기존 벤처 기업의 결정요인 연구에서는 기업가의 정신 및 전략 등의 기업 내부 요인이나 정책, 산업 환경 등의 외부 요인만 고려하였다. 이는 무엇보다도 중요한 **소비자 의견**을 반영하지 않았다는 문제점이 있다.

따라서 본 연구에서는 **소셜 데이터**를 이용하여 **대중의 의견을 시간의 흐름에 따른 키워드로** 분석하고, 창업 후 성장 혹은 재정 상태를 유지하고있는 **안정성이 높은 기업**과 창업 후 2-3년 내에 적자전환 혹은 점포 수 감소에 이른 **안정성이 낮은 기업** 사이의 차이점을 비교 및 분석한다.

분석 결과 안정성이 높은 기업의 경우, 대부분의 **키워드들의 변동 폭이 작고, 일정한 빈도를 유지**한다는 것을 확인할 수 있었다.

신생 벤처 기업의 경우 평가할만한 지표의 수가 적거나 그 구간이 짧아 신뢰하지 못 하는 경우가 많은데, 본 연구를 통해 새로운 지표를 제안한다. 또한 소셜 데이터 분석을 벤처기업 연구에 이용한 첫 번째 연구가 되어 이후 관련 연구의 발판이 될 것이다.

정의

1. 안정성

기존 기업의 ‘안정성’ 지표는 기업의 현금흐름과 관련하여 이자를 갚을 수 있는 능력이 되는지 판단하는 데에 쓰인다. 하지만 본 연구의 대상은 대부분 비상장 기업이기 때문에, 이 지표를 파악하기가 힘들었다. 따라서 본 연구에서는 프랜차이즈 지점 수로 기존 안정성 지표를 유추하고자 하였다. **프랜차이즈 지점 수가 유지 및 증가하는 기업을 안정성이 높은 기업, 감소하거나 사라진 기업을 안정성이 낮은 기업**이라고 정의하였다.

2. 빅데이터- Twitter

빅데이터는 기존 데이터의 양을 넘어서는, 대량의 정형 또는 비정형 데이터를 의미하며, 이러한 데이터로부터 필요한 가치를 추출하고 결과를 분석하는 것을 빅데이터 분석이라고 한다. 본 연구에서는 빅데이터 중 하나인 소셜 미디어 데이터인 ‘트위터’ 데이터를 이용하였다. 트위터에서는 140자 이내의 짧은 단문만 작성할 수 있어 다른 미디어에 비해 빠르고 간결하게 자신의 의견을 전달할 수 있다. 또한 실시간 소통을 중요시 여기는 트위터 특성상, 익명이 아닌 자신의 신분을 밝히고 사용하는 경우가 많아, **신뢰도가 높은 솔직한 의견을** 얻기 쉽다. 따라서 국내외의 많은 소셜 데이터 분석 연구에서 트위터를 이용한다.

분석 과정

1. 분석 기업 선정

소셜 데이터를 얻기 쉬운 외식 프랜차이즈 기업
및 2006~2016년 사이에 창업한 기업 대상.

분류	기업명	수집 기간
안정성 낮은 기업	대만카스테라	2016-08-05 ~ 2018-01-05
	망고식스	2012-01-01 ~ 2013-01-16
안정성 높은 기업	명랑핫도그	2016-08-21 ~ 2017-12-07
	백다방	2014-01-27 ~ 2016-07-30
	공차	2012-01-01 ~ 2012-10-03

[표 1] 기업 분류와 기업별 데이터 수집 기간

2. 데이터 수집 및 전처리

(1) 데이터 수집

GetOldTweets3 API와 python 이용하여 각 기업당 9000개의 데이터 수집

(2) 형태소 추출

같은 단어여도 다른 형태로 표현될 수 있기 때문에 형태소 추출 후 기본형으로 표현

(3) 불용어 제거

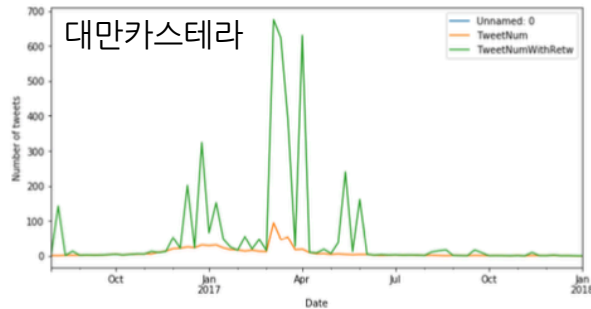
분석에 의미 없는 조사, 부사, 상호명 등의 단어 제거

(4) 유사 의미 단어 그룹화

기존 konlpy의 토큰을 이용한 방식이 아닌, 실제 온라인 한글 텍스트 데이터로 학습된 fasttext 를
이용하여, 단어를 300차원의 벡터로 표현. Cosine 유사도 검사를 통해 유사하다고 판단된 단어를
하나의 단어로 표현하여 의미 분산을 막도록 함.

분석 1. 시간에 따른 관련 데이터 변화량

안정성
낮은
기업



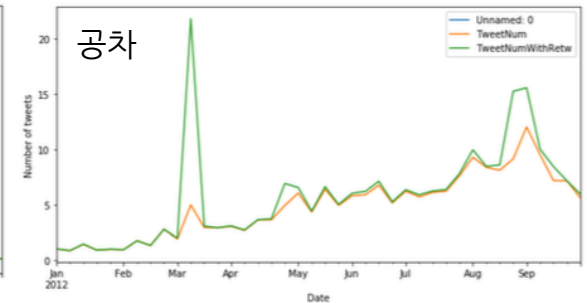
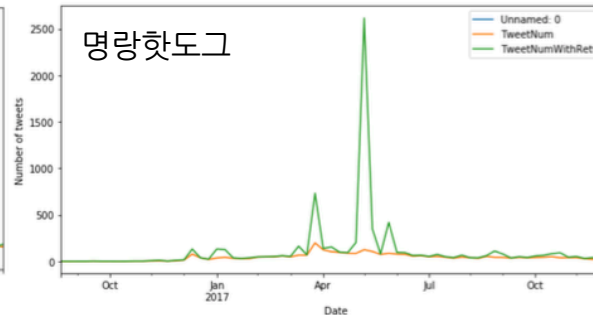
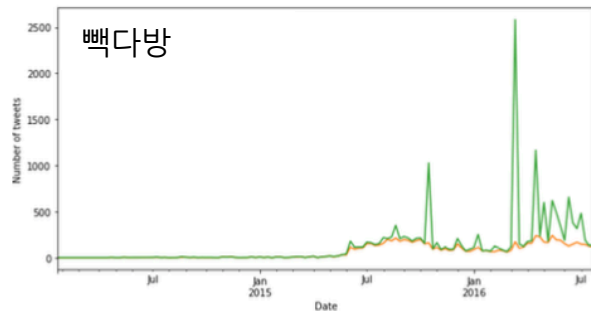
X축: 트윗 150개 구간

Y축:

— 구간당 발생 트윗수

— 리트윗된 트윗 수 포함

안정성
높은
기업

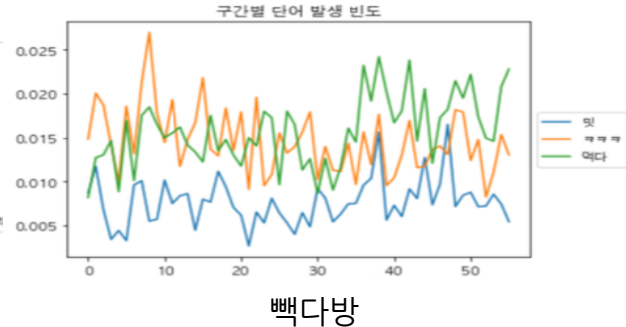
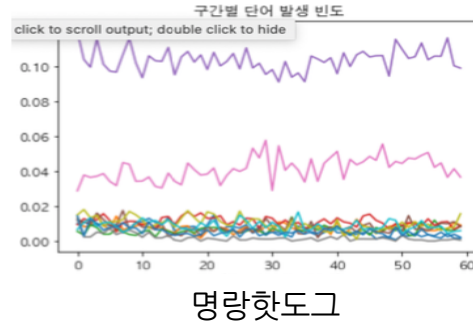
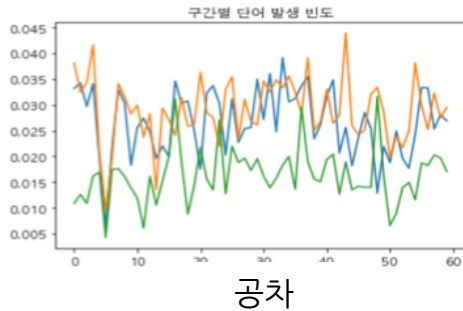


안정성이 높은 기업과 낮은 기업 사이의 그래프의 공통점 및 차이점이 없다.

⇒ 기업 관련 데이터 변화량으로는 기업의 안정성을 유추할 수 없다.

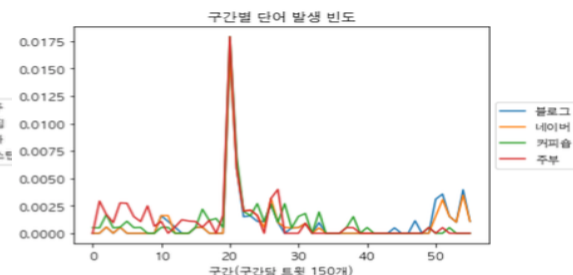
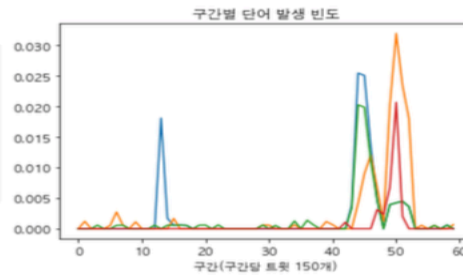
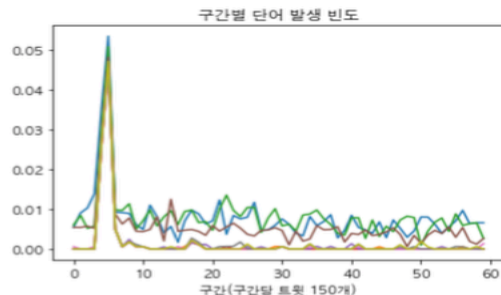
분석 2. 시간에 따른 관련 키워드 흐름 - 안정성 높은 기업

ㄱ. 빈도가 일정한 키워드



1. 관련 상품: 밀크, 티, 버블티, 핫도그, 치즈, 모짜렐라, 빵, 커피, 먹다
2. 이모티콘: ㅋㅋㅋ, ㅌㅌ

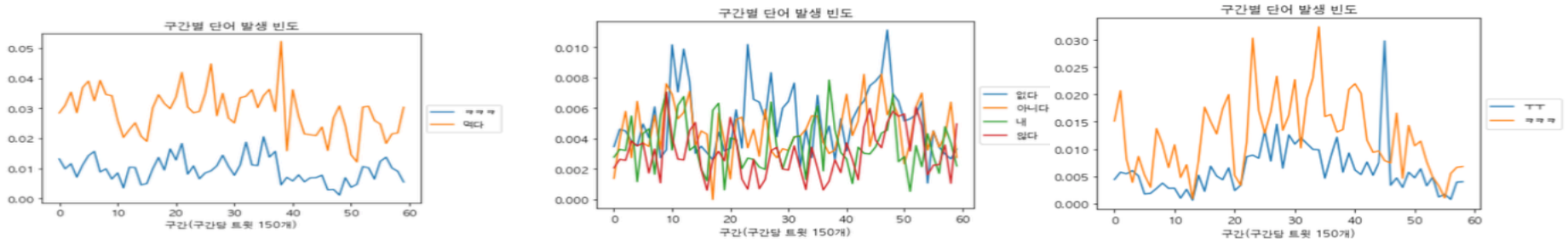
ㄴ. 급증 후 급락한 키워드



1. 긍정적인 어휘: 맛있다, 좋다, 왕, 맛집
2. 관련 장소: 호텔, 웨스턴, 도착
3. 출처: 네이버, 블로그, 커피숍, 주부

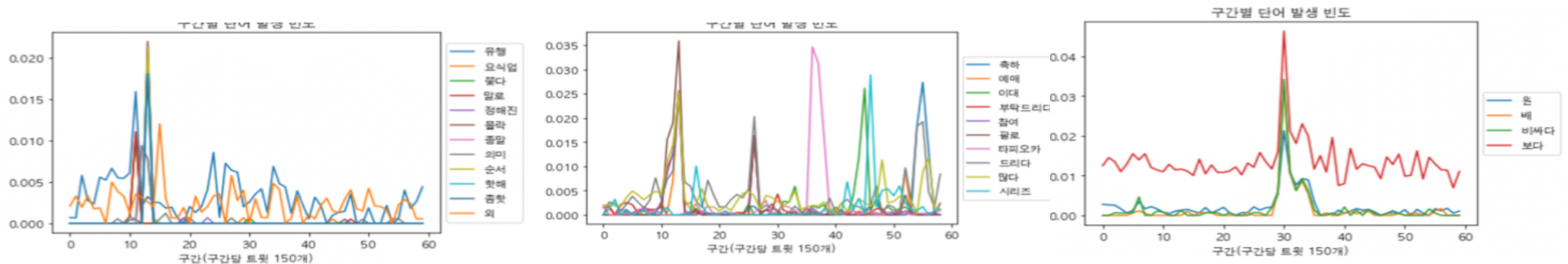
분석 2. 시간에 따른 관련 키워드 흐름 - 안정성 낮은 기업

ㄱ. 빈도가 일정한 키워드



1. 이모티콘 및 관련 동사: ㅋㅋㅋ, ㄷㄷ, 먹다
2. 부정적인 어휘 및 1인칭 어휘: 아니다, 없다, 았다, 내

ㄴ. 급증 후 급락하는 키워드



1. 가격 관련 어휘: 비싸다, 원, 배
2. 대중의 의견을 예측할 수 있는 어휘: 정해진, 몰락, 뚝다
3. 홍보 관련 어휘: 티아라, 박명수, 방은미, 김수로, 축하, 부탁드리다, 시리즈

분석 결과

1. 관련 데이터의 **변화량**이 기업의 안정성에 상관 관계가 없음을 알아냈다.
2. 하나의 기업을 제외하고, 모든 기업들이 **유사한 시기에 급변하는 키워드 그룹**을 가지고 있다.
3. **발생 빈도가 높은 키워드**들의 시간에 따른 경향성은 안정성이 높은 기업과 낮은 사이에 분명한 차이가 있었다.
 - 안정성 높은 기업: 대부분의 키워드의 발생 빈도 폭이 크지 않음. 급변한 키워드도 긍정적인 어휘 혹은 관련 장소를 의미하는 어휘임.
 - 안정성 낮은 기업: 부정적인 어휘들이 상위 빈도를 유지하고 그 변동 폭이 크고 주기적임, 부정적인 어휘나 홍보를 위해 사용된 전혀 연관 없는 어휘들이 급증 후 급락하는 경향을 보임.
4. 특정 단어가 매우 빠르게 급증하고 이후 짧은 기간 내에 급락하고, 해당 단어가 긍정적인 의미를 지니고 있지 않다면 기업의 안정성에 유의해야할 필요가 있다. 대부분의 키워드들이 작은 변동 폭과 일정한 발생 빈도를 가진다면 안정성이 높은 기업이라고 평가할 수 있다.

연구 요약 및 향후 연구 방향

연구 요약

1. 중요한 기업의 외부요인 중 하나인 소비자의 의견을 실시간으로 수집할 수 있는 소셜 데이터를 이용해 소비자의 의견 흐름을 파악하고자 했으며, 이와 기업의 안정성과의 상관성을 파악하고자 하였다.
2. 본 연구는 **최초로 벤처기업의 성공 및 실패 요인 분석에 소셜 빅데이터를 이용한 연구**이며, 추후 관련 연구의 발판이 될 수 있다.
3. **기존의 지표 외에도 기업의 성공과 실패를 예측하는 새로운 지표를 제안**했다는 점에서 의미가 있다. 이를 통해 특정 키워드를 가지는 여론이 기업의 안정성에 미치는 영향력을 파악할 수 있었으며, 여론 분석에 대한 사업가의 주관적인 의견이 아닌 객관적인 수치화된 자료를 제공할 수 있다.

한계점 및 향후 연구 방향 제안

1. **연구대상표본의 수와 특징이 제한적**이다. 따라서 이를 모든 기업에 일반화하기 어렵다. 따라서 충분한 소셜 데이터가 있는 벤처 기업을 대상으로, 외식 프랜차이즈가 아닌 다양한 분야의 벤처기업들을 선정하여 연구를 진행하여야 한다.
2. 소비자들은 트위터 외에도, 페이스북, 인스타그램 등의 **다양한 소셜 미디어**를 이용한다. 각 소셜 미디어의 특징이 다르기 때문에, 사용자의 분포도 차이가 있다. 따라서 다양한 소셜 미디어의 데이터를 이용해 분석을 하면 **더 폭 넓은 사용자들의 의견**을 수집할 수 있다.
3. 분석하려는 대상 기업이 비상장기업인 경우가 많아, 회사의 자세한 재정 및 운영 상태를 알기 어려웠고 따라서 기업의 안정성을 비교할만한 지표가 부족했다는 한계점이 있다. 따라서 추후 연구에서는 재정 및 운영 상태가 투명한 기업에 대해 여러 지표를 구하고, 이 지표들을 이용하면 소셜 데이터의 **직접적인 연관성**을 파악할 수 있을 것이다.