

감정분석을 통한 기업 건전성 평가



유건령 2018020633

장은아 2018021125

박명현 2019020542

INDEX

001 Introduction

002 Feedback & Reflection

003 Dataset

004 Modeling & Analysis

- word2vec : Skip-gram / CBOW
- label propagation
- sentimental analysis

005 Conclusion

001 Introduction

‘나는 시간이 날 때마다
관심있는 회사의 사업보고서와 경쟁사의 사업보고서를 읽는다.
그보다 더 중요한 자료는 없다.’

– 워렌 버핏

워렌 버핏과 같은 Domain knowledge가 없는 사람들은
사업보고서를 어떻게 활용할 수 있을까?

사업보고서

+

기업 관련 기사

I. 회사의 개요

1. 회사의 개요

가. 연결대상 종속회사 개황

상호	설립일	
아시아나IDT(주)	1991.09.02	
금호연건(중국)유한공사	2009.02.23	Huaneng Building, No.399, Xinghuo Road, Changchun, China
아시아나세이버(주)	2004.05.12	서울특별시 종로구 새문안로76 (신문로1가)
Asia Staff Service, Inc	2001.06.22	Tokyo-To, Taito-Ku, Ueno 1-10-12, Shokochuki, Daichi Seir
아		
아시아나개발(주)	1997.05.09	인천광역시 중구 공항동로295번길 77-65 (운서동)

기업 건전성 평가

Domain knowledge가 없는 사람에게 정보 제공

2019.03.28 (목)
15.5°C 청주MK
매일경제 MBN

전체 뉴스 오피니언 프리미엄 우버/사이드 연예 스포츠 증권 부동산 교육 비

제주 농축 바다
없을 것"노후경유차 폐차
때 지원강화"대출 증가세 둔
채 85조 넘어서"임종 과정
투여도 중단 가능열고
처리한국의...
구...7명 후보자

아시아나의 한변... "재무제표 `하절` 의겨우 회계상 차이"

Data Crawling

- Making crawlers
- Collecting news articles
- DART 사업보고서 수집
- 사업보고서 공시 전후 주가 수집



Pre-processing

- Analyzing Morpheme
- Removing stopwords

Modeling & Analysis

- Words embedding
 - word2vec(skip-gram & CBOW)
- Sentimental analysis
 - 감성사전 구축: word polarity(label propagation)
- 분기별 감성점수와 주가 up&down 비교



Input

신문기사
&
사업보고서

Preprocessing
Word2Vec
Label Propagation
Sentimental Scoring

Output

분기별
감성점수



사업보고서 공시
전후 주가의
Up&down과 비교

002 Feedback & Reflection

Feedback & Reflection

Feedback	Reflection
<ul style="list-style-type: none"> Fully connected graph를 그리면 Graph-based learning 관점에서 많은 문제점이 발생함, 한 단어는 일부 관련성이 높은 단어들과만 연결되어야 함 	<ul style="list-style-type: none"> Epsilon-neighborhood로 Graph 구현 및 감성사전 구축, Fully-connected graph와 비교
<ul style="list-style-type: none"> 사업계획서/뉴스의 감성점수와 주가와의 관련성 분석을 하는 궁극적인 목적이 무엇인지 명확하게 정의하고 진행해야 함 	<ul style="list-style-type: none"> 사업보고서/기사(사업보고서 공시 전 10일치)의 감성점수를 통해 사업보고서 공시 전후의 주가 변동(up&down)을 예상을 위한 정보 제공
<ul style="list-style-type: none"> 도메인 지식이 부족한 사용자에 의한 감성사전 기준단어 선정의 신뢰성 문제 	<ul style="list-style-type: none"> Frequency 상위 단어 중 기준단어 선정 → centrality 상위 단어 기준으로 긍정/중립/부정 단어 선정 서울대 언어학과에서 만든 KOSAC기반 Korean Sentiment Lexicon에서 감정일 확률 0.5이상인 단어로 감성점수 산출 후 비교

003 Dataset

기업 기사

매일 & 한국경제

2008 – 2017
경제 기사 55,172건

*제목에 기업 이름을 포함한 기사

사업보고서

정기 사업보고서 (DART)

KOSPI 200
10년 이상 상장
시가 총액 상위 25개 기업
분기 단위 총 939개

Dataset | Preprocessing

Article/Report → Tokenizing / POS tagging → remove stopwords

Komoran

[화학, 388500, 013, 은, 유기, 발광, 다이오드, 의, 인광, 발광,
[3, 일, 오후, 7, 시, 30, 분, 께, 충북, 청원군, 옥산면, 남
[올해, 중, 복수, 의, 중견, 제약사, 인수, 를, 추진, 하, 는
[화학, 의, 전지, 사업, 이, 잇따르, 나, 약재, 로, 흔들리,
[오, 창, 공장, 에, 화재, 가, 나, 나, 화학, 388500, 013, 에,
[화학, 그룹, 등, 주요, 기업, 들, 도, 컨설팅, 회사, 등, 을
[조선, 철강, 화학, 등, 중국, 관련주, 경영진, 이, 자사, 주
[화학, 의, 지인, 창호, 는, 화학, 의, 창호, 제품, 을, 아무

117.097 secs

Okt(Twitter)

화학, 388500, 013, 은, 유기발광다이오드, 의, 인광, 발광,
3일, 오후, 7시, 30분, 께, 충북, 청원군, 옥산면, 남춘, 리,
올해, 중, 복수, 의, 중견, 제, 약사, 인수, 를, 추진, 할, 계
화학, 의, 전, 지, 사업, 이, 잇따른, 약재, 로, 흔들리고, 있
오, 창, 공장, 에, 화재, 가, 난, 화학, 388500, 013, 에, 대
화학, 그룹, 등, 주요, 기업, 들, 도, 컨설팅, 회사, 등, 을
조선, 철강, 화학, 등, 중국, 관련, 주, 경, 영진, 이, 자사,
화학, 의, 지인, 창호, 는, 화학, 의, 창호, 제품, 을, 아무

11.179 secs

Komoran은 Okt 대비 세분화된 결과 제공, 그러나 명사 형태소 분석에서 큰 차이는 없는 것으로 판단
동일한 240개 data에 대한 형태소 분석 시간을 비교했을 때 Komoran이 10배 정도 더 소요되었음
시간과 성능을 고려하여 Okt 형태소 분석기 사용

Article/Report → Tokenizing / POS tagging → remove stopwords

다. 경기변동의 특성 및 계절성

초고속인터넷, 시내전화 서비스는 상대적으로 일반 경기변동의 영향이 미

2012년말 지상파 아날로그 방송 종료에 따른 아날로그 케이블 가입자의

라. 시장점유율 추이

(단위 : %)

구분	2012년 반기	2011년	2010년
초고속인터넷(SKT 재판매 포함)		23.9	23.5 23.2
시내전화(인터넷전화 포함)	15.2	14.6	13.7
IPTV(실시간)	20.7	19.3	23.8

주) 자료 : 방송통신위원회(www.kcc.go.kr) 및 각사

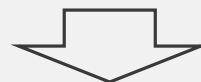
사업보고서

◇텔레콤 = 보통주 1주당 1000원 현금배당 결정. 배당 기준일은 지난달 컨설팅의 채무 280억원 인수 결정. ◇한전산업개발 = 한국전력 검침용 한전 13개 사업본부 중 8개 사업본부 수주. 수주액은 1266억원. ◇한국설 확충 위해 삼척기지에 5537억원 신규 투자 결정. ◇한국단자공업 = 결정. 배당 기준일은 지난달 30일. ◇디스플레이 = 고부가가치 중소형 장에 1조2008억원 신규 투자. ◇한성기업 = 계열 회사인 극동수산이 국내에서 추가 취득. ◇모토닉 = 주가 안정을 위해 씨티은행과 계약한 3 계약 연장. ◇삼성테크윈 = 단기 여유자금 운용 목적으로 다음달 7일 금전신탁을 매수. ◇쌍용자동차 = 우리사주조합을 대상으로 28140000 실시. © 매일경제 .. 무단전재 및 재배포 금

기업 기사

형태소 분석이 제한되는
각종 특수문자, 한글자, 영어 삭제

[화학, 388500, 013, 은, 유기발광다이오드, 의, 인광, 발광, 총, 메
[3일, 오후, 7시, 30분, 계, 충북, 청원군, 옥산면, 남촌, 리, 오,
[올해, 중, 복수, 의, 중견, 제, 약사, 인수, 를, 추진, 할, 계획,
[화학, 의, 전, 지, 사업, 이, 잇따른, 악재, 로, 흔들리고, 있다,
[오, 창, 공장, 메, 화재, 가, 난, 화학, 388500, 013, 메, 대해...
[화학, 그룹, 등, 주요, 기업, 들, 도, 컨설팅, 회사, 등, 을, 통해



[화학, 388500, 013, 유기발광다이오드, 인광, 발광, 대한, 기술, 특
[3일, 오후, 7시, 30분, 충북, 청원군, 옥산면, 남촌, 과학, 산업, 연
[올해, 복수, 중견, 약사, 인수, 추진, 계획, 입니다, 적대, 방식, O
[화학, 사업, 잇따른, 악재, 흔들리고, 있다, 최근, 노트북, 배터리,
[공장, 화재, 화학, 388500, 013, 대해, 저가, 매수, 기회, 활용, ..
[화학, 그룹, 주요, 기업, 컨설팅, 회사, 통해, 새로운, 사업, 전략,

004 Modeling & Analysis

Modeling

	1	2	...	299	300
w_1					
w_2					
w_3					
⋮					
⋮					
⋮					
w_{9054}					
w_{9055}					
w_{9056}					

ngram	freq	COMP	NEG	NEUT	POS	max.value
가*/JKS	1	0	0	0	1	POS
가*/JKS;있/VV	1	0	0	0	1	POS
가*/JKS;있/VV;없/EP	1	0	0	0	1	POS
가*/VV	3	0	0	0	1	POS
가*/VV;ㄴ 다*/EF	1	0	0	0	1	POS
가/JKC	17	0	0.411765	0.117647	0.470588	POS
가/JKC;되/VV	11	0	0.272727	0.181818	0.545455	POS
가/JKC;되/VV;ㄴ /ETM	2	0	0	1	0	NEUT
가/JKC;되/VV;ㄹ/ETM	1	0	0	0	1	POS
가/JKC;되/VV;여/EC	2	0	0.5	0	0.5	NEG
가/JKC;되/VV;여야지요/EF	1	0	0	0	1	POS
가/JKC;아니/VCN	6	0	0.666667	0	0.333333	NEG
가/JKC;아니/VCN;ㄴ 가/EC	1	0	1	0	0	NEG

2013 1분기	0.1391	0.3205	0
2013 2분기	0.0114	0.3148	0
2013 3분기	0.5141	0.2968	1
2013 4분기	0.2768	0.293	
2014 1분기	0.3071	0.2962	1
2014 2분기	0.2064	0.299	0
2014 3분기	0.2359	0.3113	1
2014 4분기	0.4653	0.3055	1
2015 1분기	0.0666	0.2891	0
2015 2분기	0.1517	0.289	0
2015 3분기	0.3059	0.293	1
2015 4분기	0.2471	0.3027	1
2016 1분기	0.1978	0.299	0
2016 2분기	0.3944	0.3207	1
2016 3분기	0.0985	0.3043	0
2016 4분기	0.3475	0.2718	1
2017 1분기	0.1664	0.2803	0
2017 2분기	0.3559	0.2836	1
2017 3분기	0.3415	0.2736	1
2017 4분기	0.2911	0.2772	1

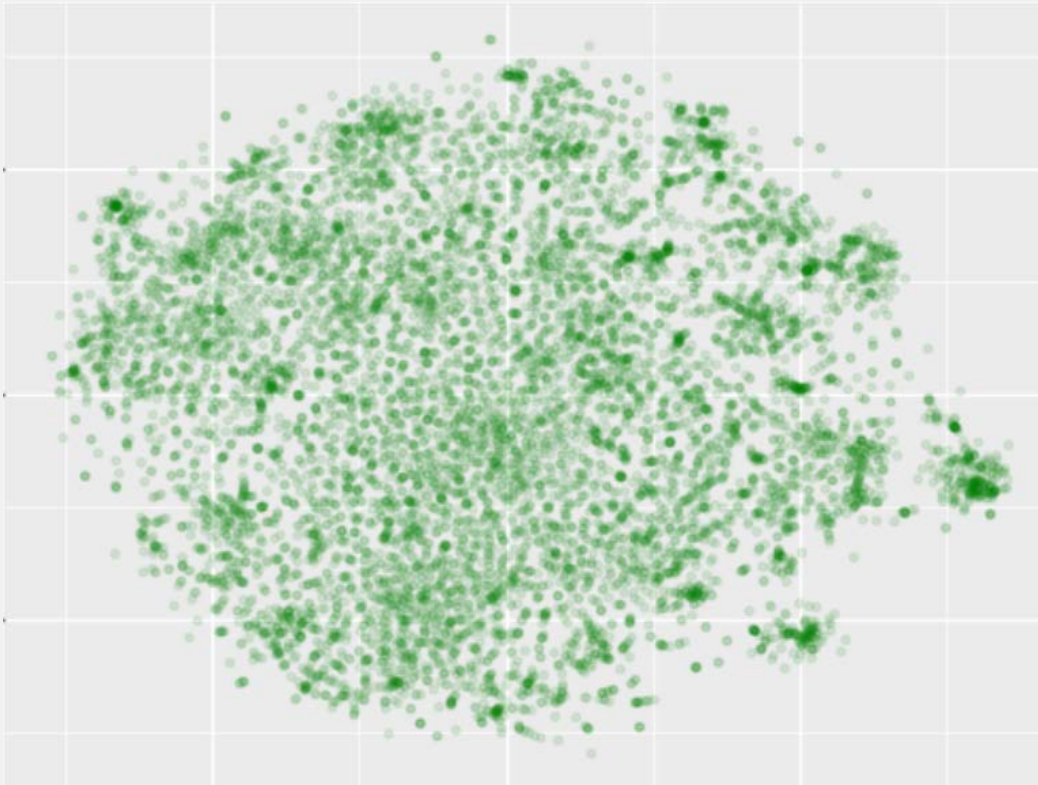
2013 1분기	0
2013 2분기	1
2013 3분기	1
2013 4분기	1
2014 1분기	1
2014 2분기	0
2014 3분기	0
2014 4분기	1
2015 1분기	0
2015 2분기	0
2015 3분기	1
2015 4분기	0
2016 1분기	1
2016 2분기	0
2016 3분기	1
2016 4분기	0
2017 1분기	0
2017 2분기	1
2017 3분기	1
2017 4분기	1

Copyright©. Saebyeol Yu. All Rights Reserved.

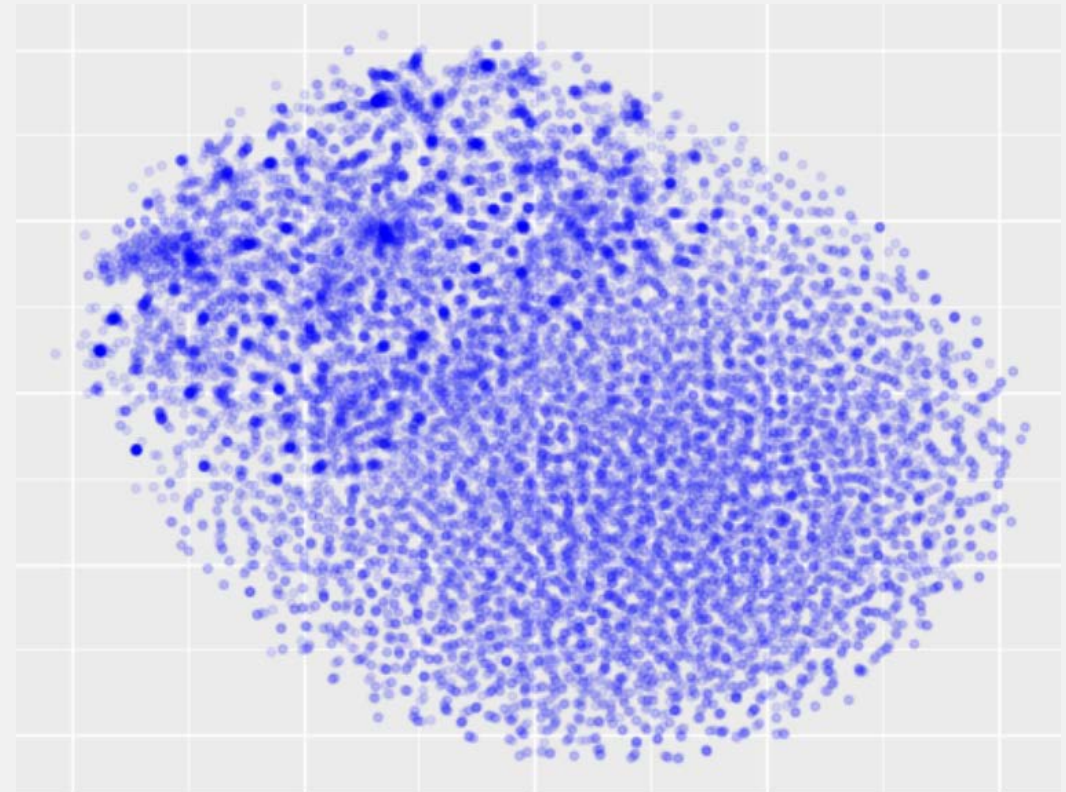
4

Modeling | Word2Vec

신문 기사



사업 보고서



Modeling | 감성사전 구축

9,058개 단어들에 대한 network graph 구축

Fully-connected graph

```
array([2.86687136e-04, 5.63490728e-04, 1.01656209e-04, 1.72252127e-04,
       7.85124139e-05, 9.61022015e-05, 7.93160143e-05, 8.00969065e-05,
       8.43907837e-05, 6.63941973e-05, 8.70377553e-05, 7.46420628e-05,
       6.25180473e-05, 8.40963330e-05, 1.02372556e-04, 8.87539354e-05,
       1.21276069e-04, 1.82835924e-04, 1.17516502e-04, 1.35979586e-04,
       1.23601523e-04, 1.26261599e-04, 5.87845097e-05, 1.69614374e-04,
       2.00583570e-04, 9.46661967e-05, 1.72560423e-04, 7.91719576e-05,
       1.10381530e-04, 8.55658582e-05, 1.59863048e-04, 7.73157954e-05,
       9.84635990e-05, 7.41637123e-05, 7.28001760e-05, 8.79461295e-05,
       2.29248850e-04, 1.25517414e-04, 1.41643934e-04, 1.09731060e-04,
       1.83662996e-04, 9.93698050e-05, 1.48981577e-04, 7.12974870e-05,
       9.00274608e-05, 7.74110304e-05, 8.12759245e-05, 1.41880853e-04,
       1.12505935e-04, 1.49082596e-04, 8.43777962e-05, 1.07962915e-04,
       7.76690358e-05, 1.48075022e-04, 1.02107057e-04, 7.95182495e-05,
       8.98016733e-05, 7.57947128e-05, 7.21007309e-05, 1.04342995e-04,
       7.14947091e-05, 1.16352327e-04, 6.90638990e-05, 6.86557323e-05,
       5.74173173e-05, 7.71817795e-05, 5.19716777e-05, 7.38624221e-05,
       9.61891419e-05, 6.87489446e-05, 6.54704127e-05, 9.13036347e-05,
       9.76081792e-05, 6.97691285e-05, 9.50660033e-05, 6.95700219e-05,
       6.92672766e-05, 7.07703439e-05, 6.53652532e-05, 7.18982483e-05,
```



Epsilon-neighborhood graph

```
array([0.0000000e+00, 5.6349055e-04, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       7.9867972e-20, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 2.3308966e-20, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       1.4558694e-20, 0.0000000e+00, 1.9842740e-21, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 2.0098579e-19, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 0.0000000e+00, 0.0000000e+00,
       0.0000000e+00, 0.0000000e+00, 1.9462828e-19, 0.0000000e+00,
```

CBOW, window=2, iter = 50, gamma = 10	Positive	Neutral	Negative
Positive	10	72	18
Neutral	5	73	22
Negative	14	64	22

105/300



Epsilon-neighborhood로 조금 더 가벼운 graph를 만들었으나
결과가 좋지 않아 기존의 방법을 이용

4

Modeling

RBF kernel을 이용하여 거리가 가까운 노드에 가중치 부여

$$w_{ij} = \exp\left(-\frac{d(x_i, x_j)^2}{2\sigma^2}\right) \text{ where node } i \text{ and node } j \text{ adjacent}$$

$d(x_i, x_j)$: distance between node i and node j

σ : hyperparameter

Before

[-1, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
2, -1, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 2,
-1, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 2,
2, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 2,
1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 3,
1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
1, -1, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 3,
1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, -1, -1, -1, -1, -1, -1, 3, -1, -1, -1, -1, -1,
-1, -1, 2, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, -1, -1, -1, 2, -1, -1, -1, -1, -1, -1, -1,
-1, -1, -1, 3, -1, -1, -1, -1, -1, 2, -1, -1, -1, -1, 2,
1, -1, -1, -1, -1, -1, -1, -1, -1, -1, -1, 3, 3, -1, -1]

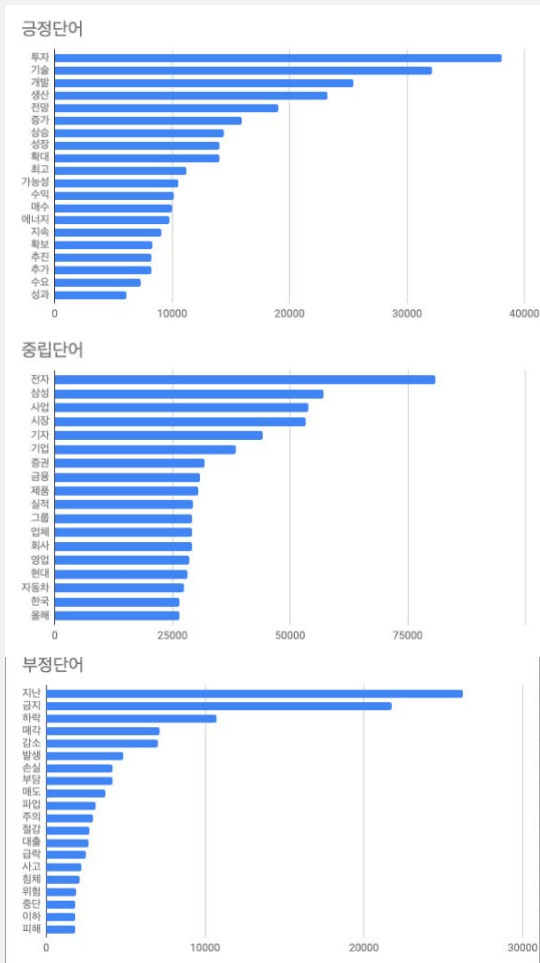


After

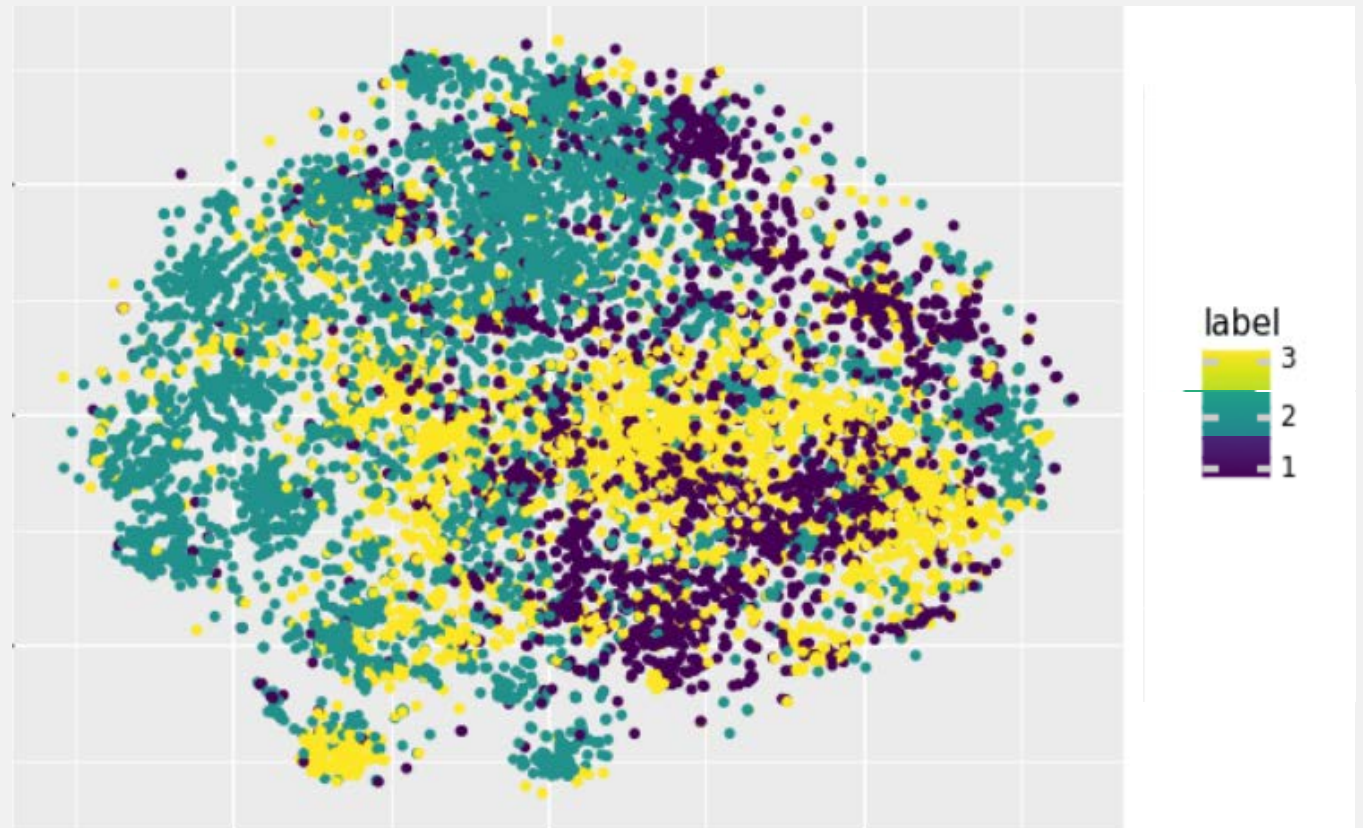
[2 2 2 2 3 3 1 1 2 3 1 2 1 2 2 2 2 3 1 2 2 2 2 3 2 1 2 3
2 2 1 2 2 1 2 2 3 3 2 3 2 2 3 3 2 3 2 2 1 2 2 2 2 2 1 3
3 2 1 2 2 2 2 1 1 3 2 2 2 3 2 2 2 2 2 1 2 2 2 2 2 2 1
2 3 1 2 1 3 2 1 3 2 1 2 2 2 2 3 2 2 2 3 2 3 1 3 1 2 2 2
3 1 2 3 3 3 2 3 2 2 3 1 2 2 2 3 1 1 3 3 2 1 2 3 2 3 3 2
2 3 2 2 2 3 2 3 2 2 2 2 1 2 3 2 3 1 1 1 2 3 3 3 3 2 1 3
1 2 2 3 1 3 3 1 2 2 1 1 1 3 2 2 3 2 2 2 3 2 1 2 3 2 2
3 2 1 2 2 1 3 2 1 1 1 1 1 3 2 2 3 3 3 3 1 1 2 2 1
2 3 3 2 3 3 2 3 2 2 1 2 1 2 2 3 2 1 2 2 2 3 1 1 3 1 1 2
1 3 3 2 3 3 1 1 3 2 2 3 3 3 2 2 3 3 2 2 1 1 3 3 2 2 2 2
1 1 3 1 1 2 3 2 2 1 1 2 2 2 1 1 1 1 1 3 3 2 2 2 2 2 3
3 2 3 3 3 2 3 3 1 1 2 1 3 2 2 2 2 2 2 2 2 3 2 3 2 2 3 2
3 2 2 1 2 1 2 2 2 2 2 3 2 1 2 1 1 1 3 3 3 3 1 3 1 3 3
2 1 2 3 2 1 1 1 3 3 2 2 3 3 2 2 3 3 2 1 3 3 2 1 2 2 1 1
1 3 3 1 2 3 3 1 1 3 1 3 1 1 2 1 2 2 1 3 3 3 1 3 2 2 2 2
2 2 2 2 3 3 2 3 1 3 2 3 2 2 1 2 2 3 3 2 2 3 2 1 1 3 3 3
2 2 3 2 3 2 3 3 2 1 3 2 2 1 2 2 2 2 2 3 2 2 1 2 2 3 2 2]

Modeling | 감성사전 구축 – 사용자 선정 단어

긍정/중립/부정 단어 선정
– Frequency 상위 각 150 단어



레이블 전파된 극성
- 긍정 3 / 중립 2 / 부정 1



4

Modeling | label propagation

Skip-gram

Dimension = 100, min_count = 50, window = 15, iter = 100 vocab_size = 9058

Skip, gamma = 30	Positive	Neutral	Negative
Positive	48	38	14
Neutral	24	68	8
Negative	25	14	61

177/300

CBOW

Dimension = 100, min_count = 50, window = 15, iter = 100 vocab_size = 9058

CBOW, gamma = 30	Positive	Neutral	Negative
Positive	50	30	20
Neutral	17	67	16
Negative	16	4	80

197/300

Modeling | label propagation

CBOW

Dimension = 100, min_count = 50, window = 2, iter = 50 vocab_size = 9058

CBOW, gamma = 30	Positive	Neutral	Negative
Positive	47	30	23
Neutral	21	63	16
Negative	14	3	83

193/300

CBOW, gamma = 20	Positive	Neutral	Negative
Positive	53	27	20
Neutral	19	65	16
Negative	16	4	80

198/300

CBOW, gamma = 10	Positive	Neutral	Negative
Positive	58	25	17
Neutral	19	67	14
Negative	16	3	81

206/300

Modeling | centrality 기반 단어 label propagation

- 중심성을 고려한 단어 재선정
 - 중심단어일수록 다른 단어와의 거리 가중치가 높음
 ➡ 가중치 합이 높을수록 중심도가 높을 것이라고 판단

	cnetrality	단어
1	1490.9595	네온
2	1480.0527	아쿠아
3	1476.6642	영남대
4	1476.3611	김규리
5	1475.148	연극
6	1473.3741	오로라
7	1470.7555	파우더
8	1470.4579	숲속
9	1469.129	티드
10	1468.4908	팟캐스트
11	1467.8584	그릇
12	1467.7085	오렌지
...		



CBOW, window=2, iter = 50, gamma = 10	Positive	Neutral	Negative
Positive	72	12	16
Neutral	5	94	1
Negative	16	12	72

238/300

단어의 centrality를 기반으로 기준 단어를 선정하고
 이를 바탕으로 label propagation을 진행한 정확도가 가장 좋았음
 따라서 centrality 기반 기준 단어로 최종 감성사전 구축

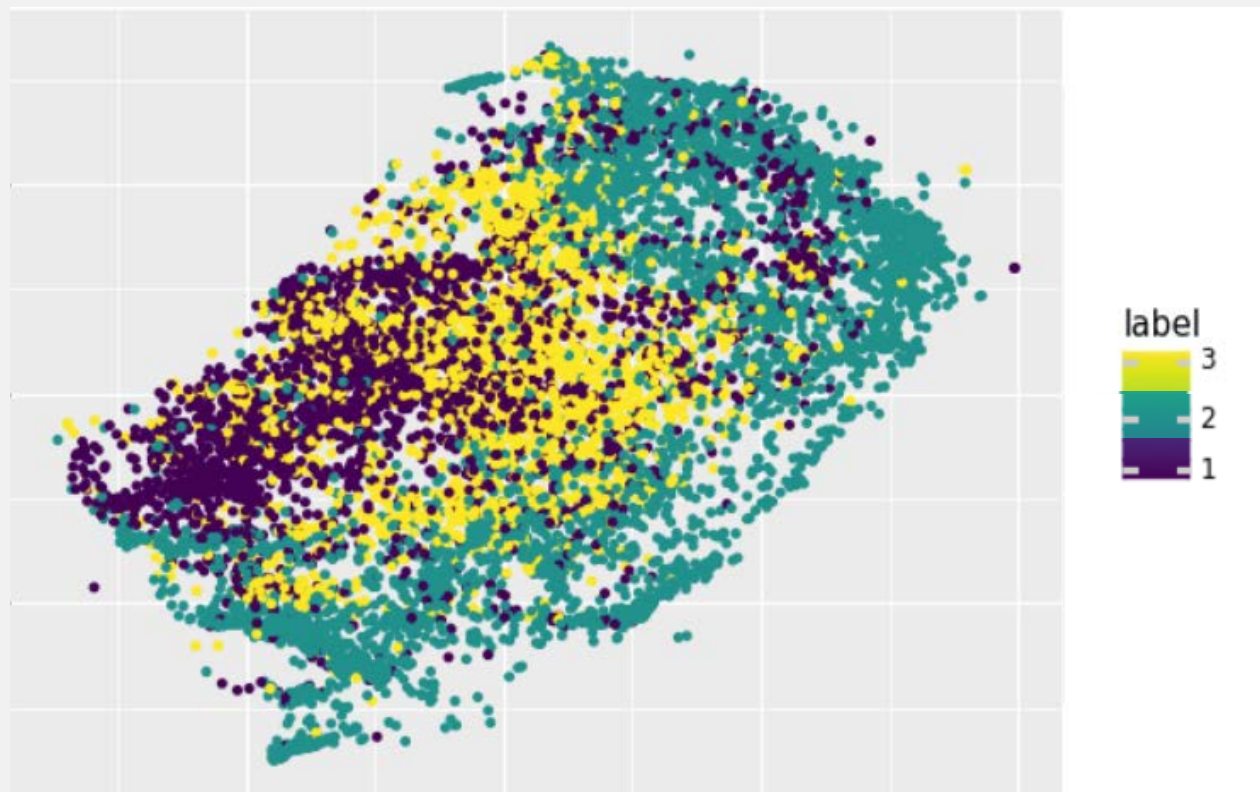
Modeling | 감성사전 구축 결과

9058 words

긍정 2720개 중립 4158개 부정 2830개

레이블 전파된 극성

- 긍정 3 / 중립 2 / 부정 1



Analysis | 감성점수 산출

- 구축된 감성 사전을 이용하여 test기업의 기사 감성 점수 산출
- 사업보고서가 게재된 날짜로부터 최근 10개 기사의 평균값을 산출
- Sentiment Score = $\frac{P - N}{P + N}$ (*P = number of positive words, N = number of negative words)

최근 5년간 현대차의 신문기사의 감성 점수

2013 1분기	0.1391	2015 3분기	0.3059
2013 2분기	0.0114	2015 4분기	0.2471
2013 3분기	0.5141	2016 1분기	0.1978
2013 4분기	0.2768	2016 2분기	0.3944
2014 1분기	0.3071	2016 3분기	0.0985
2014 2분기	0.2064	2016 4분기	0.3475
2014 3분기	0.2359	2017 1분기	0.1664
2014 4분기	0.4635	2017 2분기	0.3559
2015 1분기	0.066	2017 3분기	0.3415
2015 2분기	0.1517	2017 4분기	0.2911

+

최근 5년간 현대차의 사업보고서의 감성 점수

2013 1분기	0.3205	2015 3분기	0.293
2013 2분기	0.3148	2015 4분기	0.3027
2013 3분기	0.2968	2016 1분기	0.299
2013 4분기	0.293	2016 2분기	0.3207
2014 1분기	0.2962	2016 3분기	0.3043
2014 2분기	0.299	2016 4분기	0.2718
2014 3분기	0.3113	2017 1분기	0.2803
2014 4분기	0.3055	2017 2분기	0.2836
2015 1분기	0.2891	2017 3분기	0.2736
2015 2분기	0.289	2017 4분기	0.2772

Analysis | 감성점수 산출

- 사업보고서가 공시된 날과 전날과의 종가의 차이로 up & down을 측정
- $0.7 \times (\text{기사의 감성 점수}) + 0.3 \times (\text{사업 보고서의 감성 점수})$
→ 0.45 이상이면 긍정 1, 미만이면 부정 0
- 전날보다 주가가 상승했으면 1, 동일하거나 하락했으면 0

최근 5년간 현대차의 주가 up&down

2013 1분기	0	2015 3분기	1
2013 2분기	1	2015 4분기	0
2013 3분기	1	2016 1분기	1
2013 4분기	1	2016 2분기	0
2014 1분기	1	2016 3분기	1
2014 2분기	0	2016 4분기	0
2014 3분기	0	2017 1분기	0
2014 4분기	1	2017 2분기	1
2015 1분기	0	2017 3분기	1
2015 2분기	0	2017 4분기	1

최근 5년간 현대차 감성점수의 up&down

2013 1분기	0	2015 3분기	1
2013 2분기	0	2015 4분기	1
2013 3분기	1	2016 1분기	0
2013 4분기	1	2016 2분기	1
2014 1분기	1	2016 3분기	0
2014 2분기	0	2016 4분기	1
2014 3분기	1	2017 1분기	0
2014 4분기	1	2017 2분기	1
2015 1분기	0	2017 3분기	1
2015 2분기	0	2017 4분기	1

- 4개의 test 기업에 대해 up&down의 정확도 측정

Company	Accuracy
현대차	0.55
현대중공업	0.55
한국전력	0.675
LG디스플레이	0.65
average	0.606

- SNU data mining center에서 게재한 논문에서 약 0.62 정도의 정확도가 나왔는데 이에 비해 나쁘지 않은 성능을 보인다.

4 Analysis | 감성점수 산출 – Kosac 감성사전 이용

최근 5년간 현대차의 신문기사의 감성 점수

2013 1분기	0.7331	2015 3분기	0.6174
2013 2분기	0.6429	2015 4분기	0.616
2013 3분기	0.679	2016 1분기	0.5966
2013 4분기	0.704	2016 2분기	0.6101
2014 1분기	0.67	2016 3분기	0.6031
2014 2분기	0.6662	2016 4분기	0.6089
2014 3분기	0.603	2017 1분기	0.6405
2014 4분기	0.6243	2017 2분기	0.616
2015 1분기	0.6678	2017 3분기	0.609
2015 2분기	0.6066	2017 4분기	0.6895

+

최근 5년간 현대차의 사업보고서의 감성 점수

2013 1분기	0.7272	2015 3분기	0.7107
2013 2분기	0.719	2015 4분기	0.7095
2013 3분기	0.7188	2016 1분기	0.7095
2013 4분기	0.7131	2016 2분기	0.7107
2014 1분기	0.7108	2016 3분기	0.7056
2014 2분기	0.7073	2016 4분기	0.7142
2014 3분기	0.7131	2017 1분기	0.6613
2014 4분기	0.7107	2017 2분기	0.6548
2015 1분기	0.7238	2017 3분기	0.658
2015 2분기	0.7238	2017 4분기	0.6651

Analysis | 감성점수 산출 – Kosac 감성사전 이용

최근 5년간 현대차의 주가 up&down

2013 1분기	0	2015 3분기	1
2013 2분기	1	2015 4분기	0
2013 3분기	1	2016 1분기	1
2013 4분기	1	2016 2분기	0
2014 1분기	1	2016 3분기	1
2014 2분기	0	2016 4분기	0
2014 3분기	0	2017 1분기	0
2014 4분기	1	2017 2분기	1
2015 1분기	0	2017 3분기	1
2015 2분기	0	2017 4분기	1

최근 5년간 현대차 감성 점수의 up&down

2013 1분기	1	2015 3분기	0
2013 2분기	1	2015 4분기	0
2013 3분기	1	2016 1분기	0
2013 4분기	1	2016 2분기	0
2014 1분기	1	2016 3분기	0
2014 2분기	1	2016 4분기	0
2014 3분기	0	2017 1분기	0
2014 4분기	1	2017 2분기	0
2015 1분기	1	2017 3분기	0
2015 2분기	0	2017 4분기	1

* 0.65 이상일 경우 1

- 4개의 test 기업에 대해 up&down의 정확도 측정

Company	Accuracy
현대차	0.525
현대중공업	0.525
한국전력	0.55
LG디스플레이	0.525
average	0.531

- 모든 결과가 word embedding를 이용한 결과보다 좋지 않았음.
- 평균 정확도 0.531로 동전 던지기 수준

사업보고서의 상당 부분이 중복된 내용을 포함
따라서 보고서 간 변화된 부분만 집중적으로 분석을 할 필요가 있음

양질의 기사를 이용할 필요가 있음
부고, 승진 기사, 봉사 활동 등 기업의 가치와 관련이 없거나,
기업에 영향을 미치지 않는 기사가 다수 포함

형태소 분석기 자체의 성능 한계

Embedding 방법론에 대한 한계
감성분석에 중점을 둔 embedding 방법이 필요해 보임

신문기사와 사업보고서에 대한 준지도 학습 기반의 그래프 생성과
이를 이용한 감성사전 구축과 감성분석을 수행

경제분야 전문가가 단어에 대하여 극성을 부여하고,
GloVe와 FastText 등 다양한 방법을 적용하면
감성사전 구축 및 감성분석의 성능이 향상 될 것으로 기대

주가는 훨씬 다양한 외생 변수의 영향을 받기 때문에 예측의 정확도가 높지 않았음.
텍스트 마이닝과 기업의 수익률, 성장률 등과 같은 정형데이터를 함께 예측하면 정
확도가 더 상승할 것으로 기대