Word Embeddings

By: Ama, Lu, and Julio

Word Embedding

Julio Cernadas, AMA ADOM,LU CHEN

Collecting Data

Bloomberg

Economics

U.S. GDP Grows at 2.6% Pace as Business Spending Accelerates

By Katia Dmitrieva

February 28, 2019, 8:30 AM EST *Updated on February 28, 2019, 8:54 AM EST*

- Equipment and software investment show bigger contributions
- Consumer spending cools while remaining in solid shape

U.S. GDP Growth Tops Expectations at 2.6% in Fourth Quarter

LISTEN TO ARTICLE

▶ 5:34

SHARE THIS ARTICLE

Share

Tweet

Post

The U.S. economy cooled by less than expected last quarter as business investment picked up, suggesting growth could be stronger for longer as the Federal Reserve takes a patient approach to interest rates.

The 2.6 percent annualized rate of gains in gross domestic product from October to December compared with the 2.2 percent median estimate of economists surveyed by Bloomberg. It followed a 3.4 percent advance in the prior three months, according to a Commerce Department report

Transforming Data

U.S. GDP Grows at 2.6% Pace as Business Spending Accelerates

The U.S. economy cooled by less than expected last quarter as business investment picked up, suggesting growth could be stronger for longer as the Federal Reserve takes a patient approach to interest rates.

The 2.6 percent annualized rate of gains in gross domestic product from October to December compared with the 2.2 percent median estimate of economists surveyed by Bloomberg. It followed a 3.4 percent advance in the prior three months, according to a Commerce Department report Thursday that was delayed a month by the government shutdown.

Growth cooled less than forecast after best two quarters since 2014
Consumption, which accounts for the majority of the economy, grew 2.8 percent, slightly below forecasts, while nonresidential business investment accelerated to a 6.2 percent gain on equipment, software and research spending. Government spending slowed, trade was a minor drag and inventories gave GDP a small boost.

Treasury yields and the dollar rose following the data.

The report shows how Republican-backed tax cuts may have continued to aid growth and help bring the full-year figure to 3.1 percent, just above President Donald Trump's 3 percent goal. While the expansion is poised to become the nation's longest on record at midyear amid a still-healthy consumer, supportive Fed and robust labor market, the pace could cool amid the trade war, slowing global growth and fading impact of fiscal stimulus.

The strength in overall private domestic demand "is good enough to keep the momentum in the economy going," with research and development spending being a "bright spot" in the report, said Neil Dutta, head of economics at Renaissance Macro Research LLC.

A separate report Thursday from the Labor Department showed filings for unemployment benefits rose by more than expected last week to 225,000, still near a five-decade low. The week included the Presidents Day holiday, and claims tend to be more volatile around such events.

Potential Growth

Growth, while slower than the prior two quarters, remains above both the average pace of the expansion and what the Fed sees as the economy's long-run potential of 1.9 percent. Still, surveys and gauges such as Treasury yields indicate chances of a recession have increased in recent months while remaining unlikely for 2019.

Excluding the volatile trade and inventories components of GDP, final sales to domestic purchasers increased at a 2.6 percent pace following 2.9 percent. Economists monitor this measure for a better sense of underlying demand.

What Our Economists Say...

The economy appears to have dodged a bullet at year-end -- but the coast is not clear ... Relative to history, the inventory accumulation in the second half of 2018 is large and threatens to overshadow production schedules (and manufacturing-related employment) in the first half of this year. Analysts should carefully scrutinize industrial surveys, such as the manufacturing ISM, for signs of a production lull intended to work off inventory excess.

-- Carl Riccadonna, Tim Mahedy and Eliza Winger, Bloomberg Economics (read more for the full note)
The increase in consumer spending followed the third quarter's 3.5 percent gain. It contributed 1.92 percentage points to growth. Drivers included health care, financial services and insurance, and other nondurable goods and services, while spending on food services and accommodations fell.

The GDP data may reinforce analyst criticism of the government's recent report on December retail

Text file

0	03-08-2019	Business	Protecting The 'Unbanked' By Banning Cashless Businesses In Philadelphia
1	03-08-2019	Business	Philadelphia just became the first large city in the nation to ban cashless businesses in the city in part to protect people like Dwight Tindal a construction worker who doesn't have a bank or credit card.
2	03-08-2019	Business	Back in December the Philadelphia City Council passed "Fair Workweek" legislation joining a growing national movement aimed at giving retail and fast-food workers more predictable schedules and by
3	03-08-2019	Business	That's typically how it works. Advocates shine a light on a problem. A bill gets introduced.
4	03-08-2019	Business	That's not the way it worked with another new law in Philadelphia. That law can be traced back to one man: City Councilman Bill Greenlee.
5	03-08-2019	Business	Last fall Greenlee introduced a bill outlawing cashless businesses â brick-and-mortar shops and restaurants where customers can only pay with credit and debit cards.
6	03-08-2019	Business	I heard that there started to be some establishments in Center City. Something just didn't sit right with me on that, said Greenlee.
7	03-08-2019	Business	Mayor Jim Kenney signed it into law last week making Philadelphia the first big city in the country to ban cash-free stores. It takes effect July 1.
8	03-08-2019	Business	But anti-poverty advocates say cashless businesses weren't a concern before Greenlee introduced his bill. Many support the bill but they didn't point out the problem to Greenlee nor did they lobby for i
9	03-08-2019	Business	The veteran lawmaker thought it was discriminatory for businesses to turn away low-income residents who don't have bank accounts a population collectively referred to as the "unbanked."
10	03-08-2019	Business	I would have liked to see a measure that more directly impacted poor people as far as getting them out of poverty.
11	03-08-2019	Business	Otis Bullock of Diversified Community Services
12	03-08-2019	Business	It just seems unfair to have that separation. It's almost like it's 'us' and 'them,' said Greenlee.
13	03-08-2019	Business	Nearly 13 percent of Philadelphia's population â close to 200 000 people â are unbanked according to federal banking data. That's more than double the regional average.
14	03-08-2019	Business	Philadelphia construction worker Dwight Tindal is one of them. He had a bank account a few years ago. He closed it because his balance never stayed above zero for very long.
15	03-08-2019	Business	I got a little baby â a son. And bills too. So, that go to all that, said Tindal.
16	03-08-2019	Business	These days the 24-year-old only uses cash. He keeps all his money in a secret location he doesn't share with anyone. He counts his stash two or three times a day every day. And only takes out what h
17	03-08-2019	Business	Tindal said it's a stressful system.
18	03-08-2019	Business	I'm concerned every day ... I just think 'dang, I should move it here' or 'I should move it over here', said Tindal.
19	03-08-2019	Business	Still he's not offended by the handful of businesses in the city that only accept cards including Sweetgreen and Bluestone Lane fast casual chains that sell built-to-order salads and coffee respectively.
20	03-08-2019	Business	That's problematic for Otis Bullock executive director of Diversified Community Services but not because he thinks unbanked Philadelphians should be more outraged. To him the silence shows City Cou
21	03-08-2019	Business	I would have liked to see a measure that more directly impacted poor people as far as getting them out of poverty. You know, a measure that raises their income. Or, at a minimum, a measure that ad
22	03-08-2019	Business	We can do this bill much quicker than I think we're gonna solve the problem of the unbanked.
23	03-08-2019	Business	Bill Greenlee Philadelphia city councilman
24	03-08-2019	Business	Greenlee admits the genesis of his bill was unconventional. He sees value in it nonetheless.
25	03-08-2019	Business	We can do this bill much quicker than I think we're gonna solve the problem of the unbanked, said Greenlee. "If there comes a time in a few years where everybody has the same ability to use some ki
26	03-08-2019	Business	The law not only forces existing cashless businesses to change course but also stops new ones from going that route in the future.
27	03-08-2019	Business	Kerry Smith a staff attorney with Community Legal Services said that safeguard is critical for the future. She's concerned that retailers that offer basic staples such as toiletry items and food may begin
28	03-08-2019	Business	A Sweetgreen spokesman declined comment. Nicholas Stone founder and CEO of Bluestone Lane didn't respond to calls and emails requesting comment on Greenlee's legislation.
29	03-08-2019	Business	In an interview last year Stone told NPR that he made Bluestone cash-free because the overwhelming majority of the company's customers never paid in cash. Stone also said lines move faster when e
30	03-08-2019	Business	The Chamber of Commerce for Greater Philadelphia opposed the bill complaining in a letter to City Council that banning cashless businesses might discourage local entrepreneurs from setting up shop i
31	03-08-2019	Business	Passing such a measure will allow local government to dictate how entrepreneurial business owners will operate versus customers deciding not to shop there, wrote president Rob Wonderling.
32	03-08-2019	Business	Andre Andrews owner of Dre's Homemade Water Ice and Ice Cream is also not a fan of the bill. His is a cash-free business that specializes in non-traditional ice cream flavors.
33	03-08-2019	Business	Andrews sympathizes with the intent of the law but he has a hard time getting past the fact that government is telling him how to operate. "I'm trying to keep up with the times he says. But you're for
34	03-08-2019	Business	New York and San Francisco which each have dozens of cash-free businesses are considering similar legislation.
35	03-08-2019	Business	New Jersey needs the governor's signature to join Massachusetts as the only state that bans cashless businesses.

CSV file

iPython Notebook

Setting Up Environment

```
import re
import nltk
import spacy
from nltk.corpus import stopwords
import matplotlib.pyplot as plt
```

```
import csv
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

Constructing Pandas Dataframe

Get Master Dataframe with All Topics

We parse through each CSV topic article and then read them into a master dataframe consisting a datetime, news type, and text column. Each article row pertains to a paragraph!

```
def get_master_df():
    numbers = ["1","2","3","4","5","6","7","8","9","10"]
    letters = ["B"]
    df_list = list()
    path = "2_csv_text/"
    for letter in letters:
        for num in numbers:
            file = letter + num
            df = pd.read_csv(path+file+".csv",header=None)
            df.drop(df.columns[0], axis=1,inplace=True)
            df_list.append(df)
    df = pd.concat(df_list,ignore_index=True)
    df.columns = ["date","news","text"]
    df["date"] = pd.to_datetime(df["date"], format='%m-%d-%Y')
    count = df['text'].str.split().str.len()
    return df[~(count < 4)]
```

Cleaning the text data

Stop Word Removal & Clean Up

```
# Stop Words Removal & Lemmatizing
spacy_nlp = spacy.load('en_core_web_sm')

remove_stops = spacy.lang.en.stop_words.STOP_WORDS
nltk_stops   = stopwords.words('english')
custom_stops = [".com","say","year","market","a","b","c","d","e","f","g","h","i","j",
                "k","l","m","n","o","p","q","r","s","t","u","v","w","x","y","z"]

for word in nltk_stops:
    spacy_nlp.vocab[word].is_stop = True

def tokenize_removestops(text):
    tokens = [token.text for token in spacy_nlp(text) if not token.is_stop]
    no_spaces = list(filter(lambda word: re.sub(r'[\W_]+','', word), tokens))
    processed = list(filter(lambda word: word not in custom_stops, no_spaces))
    return processed
```

Applying Clean Up

Pre-Processing of Text

```
def generate_tokens(sentence_list):
    final_tokens = []
    for sentence in sentence_list:
        new_text = re.sub("'|\.|\s+|\d+|[^a-zA-Z0-9]"," ", str(sentence).lower())
        tokens = tokenize_removestops(new_text.lower())
        if len(tokens) >= 3:
            final_tokens.append(tokens)
    return final_tokens
```

```
df = get_master_df()
df = df.groupby(["news","date"]).text.agg(sum).reset_index()

df["sentences"] = df.text.apply(lambda x: str(x).split("."))
df["clean_tokens"] = df.sentences.apply(generate_tokens)
df.dropna(inplace=True)

pd.options.display.max_colwidth = 150
df.head()
```

LIST OF LISTS WITH TOKENS

	news	date	text	sentences	clean_tokens
0	Business	2018-06-04	Sharp to buy Toshiba PC business issue $1.8 billion in new sharesTOKYO (Reuters) - Japan's Sharp Corp (6753.T) said it will buy Toshiba Corp's (65...	[Sharp to buy Toshiba PC business issue $1, 8 billion in new sharesTOKYO (Reuters) - Japan's Sharp Corp (6753, T) said it will buy Toshiba Corp's ...	[[sharp, buy, toshiba, pc, business, issue], [billion, new, sharestokyo, reuters, japan, sharp, corp], [said, buy, toshiba, corp], [personal, comp...
1	Business	2018-07-02	4 Stocks With Bounce-Back Potential: Thor Industries Lam Research & Morel love buying stocks when they're down.That's the point of my quarterly Ca...	[4 Stocks With Bounce-Back Potential: Thor Industries Lam Research & Morel love buying stocks when they're down, That's the point of my quarterly ...	[[stocks, bounce, potential, thor, industries, lam, research, morei, love, buying, stocks], [point, quarterly, casualty, list], [start, new, quart...
2	Business	2018-08-13	U.S. stocks close lower as Turkey currency crisis dampens risk appetiteTurkey rattles U.S. markets again.U.S. stocks closed lower Monday with the ...	[U, S, stocks close lower as Turkey currency crisis dampens risk appetiteTurkey rattles U, S, markets again, U, S, stocks closed lower Monday w...	[[stocks, close, lower, turkey, currency, crisis, dampens, risk, appetiteturkey, rattles], [stocks, closed, lower, monday, dow, jones, industrial,...
3	Business	2018-09-20	Record highs — What you need to know in markets on FridayStocks are at record highs.After 164 sessions without a record close the Dow Jones Indust...	[Record highs — What you need to know in markets on FridayStocks are at record highs, After 164 sessions without a record close the Dow Jones Indu...	[[record, highs, need, know, markets, fridaystocks, record, highs], [sessions, record, close, dow, jones, industrial, average, thursday, closed, r...
4	Business	2018-10-28	Alexa for Business opens up to third-party device makersLast year Amazon announced a new initiative Alexa for Business designed to introduce its v...	[Alexa for Business opens up to third-party device makersLast year Amazon announced a new initiative Alexa for Business designed to introduce its ...	[[alexa, business, opens, party, device, makerslast, amazon, announced, new, initiative, alexa, business, designed, introduce, voice, assistant, t...

Word Embedding Set up

Word2Vec with Window Size of 5

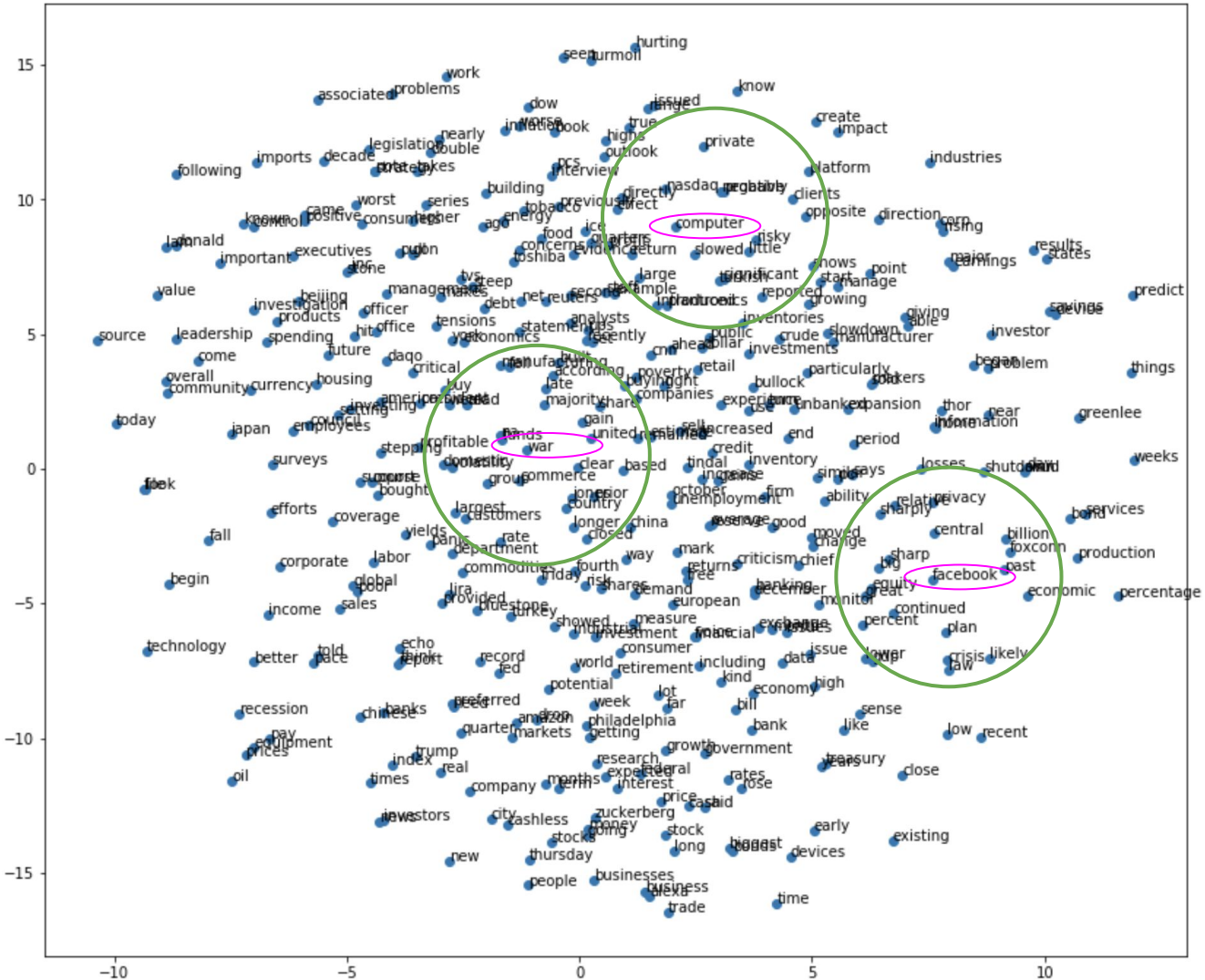
300 dimensional word embeddings

```
from gensim.models import Word2Vec
from sklearn.manifold import TSNE

model = Word2Vec(df["clean_tokens"].sum(), size=300, sg=1, window=5, min_count=3, seed=123)

print(model)
```

```
Word2Vec(vocab=391, size=300, alpha=0.025)
```

```
# Get list of words for annotation of the scatter plot
vocab = list(model.wv.vocab)
X = model[vocab]

# Project them onto the 2 Dimensional space
tsne = TSNE(n_components=2,random_state=123)
X_tsne = tsne.fit_transform(X)

# Create a DataFrame with words as index and 2 dimensions as main columns (x-axis, y-axis)
scatter_df = pd.DataFrame(X_tsne, index=vocab, columns=['x', 'y'])
```

```
# Plot the figure
fig = plt.figure(figsize=(14, 12))
ax = fig.add_subplot(1, 1, 1)
ax.scatter(scatter_df['x'], scatter_df['y'])

# Annotate each point with its word
for word, pos in scatter_df.iterrows():
    ax.annotate(word, pos)

plt.show()
```


computer
war
facebook
15
10
5
0
−5
−10
−15
−10
−5
0
5
10

Word Relationships with vector addition I

```
print(model.wv.most_similar(positive=["analysts","investor"], negative=["market"], topn=2))
```

```
[('change', 0.18723735213279724), ('rising', 0.17341852188110352)]
```

```
print(model.wv.most_similar(positive=["interest","bonds"], negative=["stocks"], topn=3))
```

```
[('market', 0.7553178071975708), ('going', 0.7515588998794556), ('like', 0.7429361343383789)]
```

```
print(model.wv.most_similar(positive=["donald","trump"], topn=2))
```

```
[('government', 0.4335134029388428), ('businesses', 0.4172137975692749)]
```

```
print(model.wv.most_similar(positive=["legislation","american"], topn=4))
```

```
[('going', 0.36799463629722595), ('rates', 0.35892853140830994), ('quarter', 0.35082608461380
005), ('need', 0.33690062165260315)]
```

Word Relationships with vector addition II

```
print(model.wv.most_similar(positive=["zuckerberg","stocks"], topn=4))
```

```
[('term', 0.633190393447876), ('money', 0.5963645577430725), ('interest', 0.5909152626991272)
, ('market', 0.5848158597946167)]
```

```
print(model.wv.most_similar(positive=["oil","inventory"], topn=2))
```

```
[('money', 0.4915943443775177), ('people', 0.4769548177719116)]
```

```
print(model.wv.most_similar(positive=["devices","technology"], negative=["stocks"], topn=5))
```

```
[('existing', 0.23767238855361938), ('staff', 0.21691012382507324), ('zuckerberg', 0.21596953
27281952), ('alexa', 0.2151208221912384), ('provided', 0.21035367250442505)]
```

Word Relationships with vector addition III

```
print(model.wv.most_similar(positive=["dow","jones"], topn=5))
```

```
[('treasury', 0.4103504717350006), ('people', 0.36952513456344604), ('interest', 0.3596799075
603485), ('market', 0.3574627637863159), ('business', 0.35138916969299316)]
```

```
print(model.wv.most_similar(positive=["gdp","economy"], topn=5))
```

```
[('market', 0.6121593713760376), ('stocks', 0.5862126350402832), ('term', 0.5563004612922668)
, ('like', 0.5548907518386841), ('interest', 0.5440171957015991)]
```

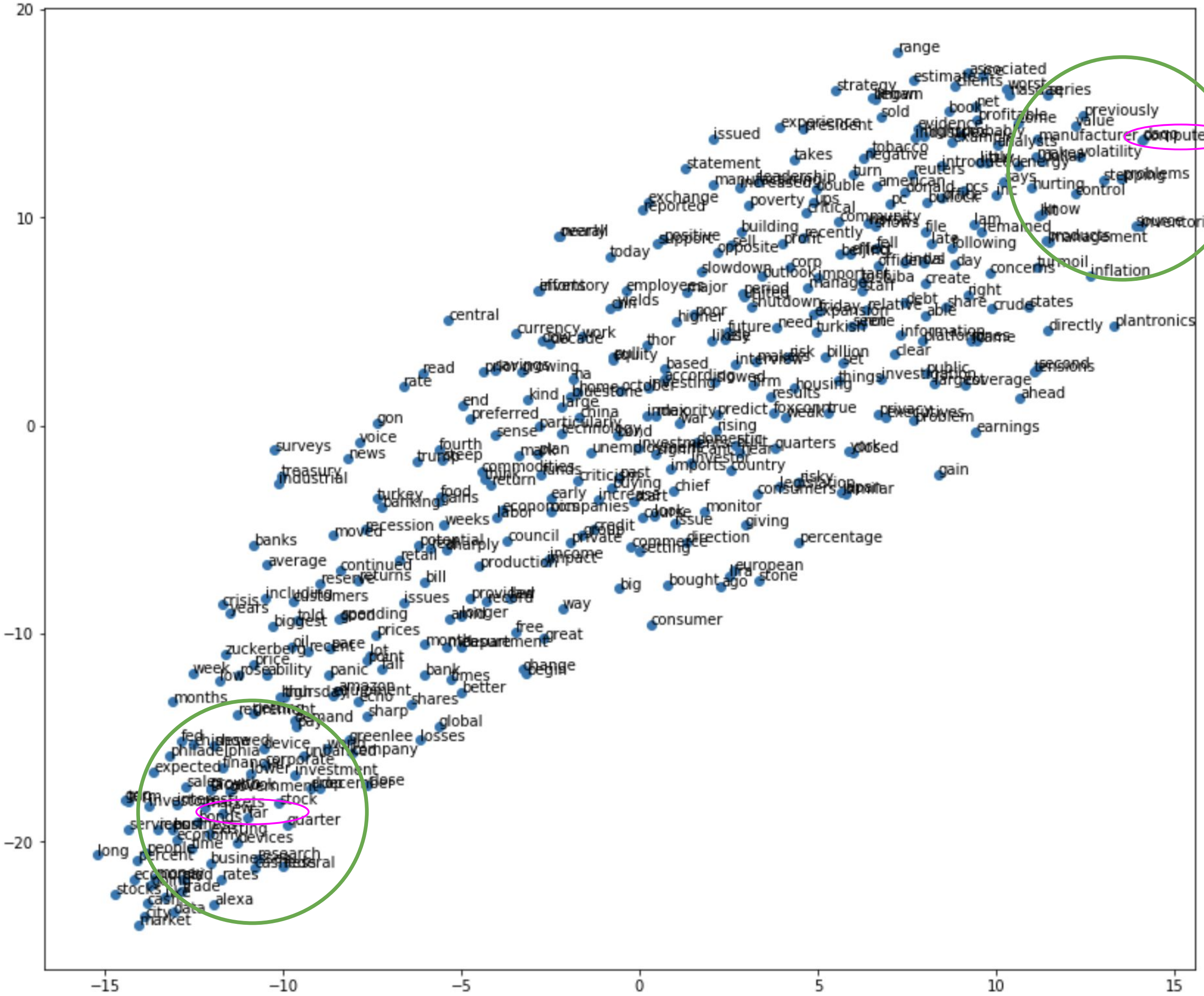
```
print(model.wv.most_similar(positive=["privacy","problems"],negative=["market"], topn=4))
```

```
[('associated', 0.07699616253376007), ('day', 0.062468867748975754), ('computer', 0.061224319
0407753), ('hurting', 0.05847734212875366)]
```

Word2Vec with window = 10

```
from gensim.models import Word2Vec
from sklearn.manifold import TSNE

model = Word2Vec(df["clean_tokens"].sum(), size=300, sg=1, window=10, min_count=3, seed=1)

print(model)
```

```
Word2Vec(vocab=392, size=300, alpha=0.025)
```



20
10
0
−10
−20
−15
−10
−5
0
5
10
15
range
associated
strategy
estimate
clients
series
previously
value
manufacturer
computer
volatility
problems
control
know
inventories
management
inflation
turmoil
concerns
issued
experience
president
statement
takes
tobacco
negative
exchange
reported
poverty
critical
community
building
today
positive
support
opposite
slowdown
outlook
employees
major
period
shutdown
central
currency
decade
work
thor
based
equity
read
rate
end
preferred
kind
large
china
surveys
voice
news
treasury
industrial
turkey
banking
food
gains
return
early
companies
banks
average
recession
moved
retail
continued
returns
bill
council
private
production
impact
crisis
years
including
customers
biggest
told
spending
issues
provided
record
way
big
bought
ago
european
stone
consumer
prices
free
great
zuckerberg
oil
recent
price
possibility
week
low
panic
point
fall
bank
times
better
change
begin
months
amazon
echo
shares
sharp
demand
global
losses
fed
philadelphia
expected
financial
corporate
investment
device
company
close
stock
quarter
devices
research
federal
rates
alexa
long
people
percent
stocks
trade
city
market
earnings
gain
percentage
consumers
popular
chief
monitor
giving
direction
commerce
setting
credit
plantronics
directly
ahead
states
crude
share
right
create
day
following
file
ipad
sold
book
net
profitable
energy
hurting
information
billion
clear
housing
results
public
coverage
privacy
problem
quarters
closed
imports
country

```
model.wv.most_similar(positive=["facebook"],topn=10)
```

```
[('said', 0.848118782043457),
 ('cash', 0.8367924690246582),
 ('trade', 0.8361796140670776),
 ('market', 0.8320250511169434),
 ('stocks', 0.8296152353286743),
 ('city', 0.8266265392303467),
 ('alexa', 0.826267659664154),
 ('time', 0.8260491490364075),
 ('money', 0.8238587379455566),
 ('business', 0.8214989900588989)]
```

```
model.wv.most_similar(positive=["war"],topn=10)
```

```
[('city', 0.7087075710296631),
 ('data', 0.7029075622558594),
 ('new', 0.6955171823501587),
 ('economy', 0.6953791379928589),
 ('going', 0.6933993697166443),
 ('far', 0.6926276683807373),
 ('economic', 0.6922622919082642),
 ('lower', 0.6871247887611389),
 ('cash', 0.6870585680007935),
 ('market', 0.6835224628448486)]
```

```
model.wv.most_similar(positive=["computer"],topn=10)
```

```
[('problems', 0.15980137884616852),
 ('recently', 0.14640238881111145),
 ('treasury', 0.14211420714855194),
 ('daqo', 0.13889500498771667),
 ('building', 0.13340261578559875),
 ('unemployment', 0.12657064199447632),
 ('fourth', 0.1261099874973297),
 ('company', 0.11997942626476288),
 ('important', 0.11415857821702957),
 ('slowdown', 0.1102391704916954)]
```

Word Relationships with vector addition I

```
print(model.wv.most_similar(positive=["analysts","investor"], negative=["market"], topn=5))
```

```
[('change', 0.18723735213279724), ('rising', 0.17341852188110352), ('evidence', 0.15520283579
826355), ('setting', 0.14325203001499176), ('domestic', 0.13980978727340698)]
```

```
print(model.wv.most_similar(positive=["interest","bonds"], negative=["stocks"], topn=3))
```

```
[('market', 0.7553178071975708), ('going', 0.7515588998794556), ('like', 0.7429361343383789)]
```

```
print(model.wv.most_similar(positive=["donald","trump"], topn=3))
```

```
[('trade', 0.7361307144165039), ('market', 0.7355953454971313), ('report', 0.7224865555763245
)]
```

```
print(model.wv.most_similar(positive=["legislation","american"], negative=["market"], topn=4))
```

```
[('return', 0.2812790870666504), ('potential', 0.2806316614151001), ('donald', 0.278567850589
7522), ('risky', 0.27498817443847656)]
```

Word Relationships with vector addition II

```
print(model.wv.most_similar(positive=["zuckerberg","stocks"], negative=["market"], topn=4))
```

```
[('rates', 0.7046475410461426), ('said', 0.7016119360923767), ('cash', 0.7002536654472351), (
'gdp', 0.6995929479598999)]
```

```
print(model.wv.most_similar(positive=["oil","inventory"], negative=["market"], topn=3))
```

```
[('long', 0.4687868356704712), ('research', 0.46835702657699585), ('interest', 0.464967995882
0343)]
```

```
print(model.wv.most_similar(positive=["devices","technology"], negative=["stocks"], topn=5))
```

```
[('market', 0.6350865364074707), ('alexa', 0.6274809837341309), ('corporate', 0.6222401261329
651), ('money', 0.614915668964386), ('like', 0.614204466342926)]
```

Word Relationships with vector addition III

```
print(model.wv.most_similar(positive=["dow","jones"], topn=5))
```

```
[('city', 0.6941426992416382), ('economic', 0.6938381195068359), ('economy', 0.6908421516418457), ('market', 0.687342643737793), ('people', 0.6843339204788208)]
```

```
print(model.wv.most_similar(positive=["gdp","economy"], topn=5))
```

```
[('market', 0.8958903551101685), ('city', 0.8958687782287598), ('trade', 0.8951538801193237), ('going', 0.8946099281311035), ('said', 0.883965015411377)]
```

```
print(model.wv.most_similar(positive=["privacy","problems"],negative=["market"], topn=4))
```

```
[('control', 0.17450252175331116), ('tensions', 0.16042041778564453), ('donald', 0.15823917090892792), ('sense', 0.13482613861560822)]
```

Paper summary

Part 2 : Paper

The Geometry of Culture: Analyzing Meaning through Word Embeddings

Austin C. Kozlowski1 Matt Taddy2,3 James A. Evans

Digitized Text

Sources of Digitized text include:

-Collective activity on the web
-Social media and instant messages as well as online transactions
-Medical records pamphlets, articles, and books

Purpose of paper : In this paper, they demonstrate the utility of a new computational approach - neural-network word embedding models

Word embing

In word embedding models, words are assigned a position in a vector space based on the context that word shares with other words in the corpus. Words that share many contexts are positioned near one another, while words that inhabit very different contexts locate farther apart.

Similar Latent Semantic Analysis (LSA) or Indexing (LSI)

Limitation

- A current limitation for the application of word embeddings is that they require a relatively large body of text if the output vector space is to capture subtle and complex associations of greatest interest to analysts of culture

- However, that word embeddings are tailored to efficiently utilize very large corpora is also a strength

Process

1.Ecological validity of word embedding models

2. Investigate macro-historic cultural trends

3.Conduct a cross-national analysis (between United States and Great Britain)

Findings

-Gender, class and race explain a small percentage of the total variance in the space: gender explains 0.57%, race 0.48%, and class 0.47%

-This demonstrates that, word embedding models are able to faithfully capture complex cultural associations and dimensions from large bodies of text to a degree unapproachable with previous methods.