Trump_Word2Vec

December 11, 2019

[ ]:
```
https://github.com/QuantCS109/TrumpTweets/blob/master/notebooks_features/
↪Trump_Word2Vec.ipynb
```

1 Word2Vec Algorithm

This is an implementation of the Word2Vec algorithm using the skip-gram architecture. I'm adapting code from a course in Udacity to our problem. The original code is here: https://github.com/udacity/deep-learning-v2-pytorch/tree/master/word2vec-embeddings

[1]:
```
import sys
sys.path.append('..') #to add top-level to path
sys.path.append('../modules') #to add top-level to path

from modules.project_helper import TweetData
from modules.skipgram import get_batches, SkipGramNeg, NegativeSamplingLoss,␣
↪subsampling
from modules.skipgram import cosine_similarity, plot_similar_words
import numpy as np

import torch
import torch.optim as optim

%matplotlib inline
%config InlineBackend.figure_format = 'retina'
import matplotlib.pyplot as plt
from sklearn.manifold import TSNE

import warnings
warnings.filterwarnings("ignore")
```

1.1 Loading data

[2]:
```
# time series of trump archives to be able to train the model sequentially
data = TweetData('../data/intermediate_data/trump_archive_ts/
↪trump_archive_db_1910.csv')
# full data set
```

```
# data = TweetData('data/trump_archive_db.csv')
```

[3]:
```
# get list of words
words = data.words
print(words[:30])
```

```
['while', 'the', 'do', 'nothing', 'democrats', 'fail', 'the', 'american',
'people', 'and', 'continue', 'the', 'impeachment', 'scam', 'my',
'administration', 'will', 'continue', 'to', 'deliver', 'real', 'results', 'as',
'seen', 'over', 'the', 'past', 'month', 'below', 'the']
```

[4]:
```
print("Total words in Trump's tweets: {}".format(len(words)))
print("Unique words: {}".format(len(set(words))))
```

```
Total words in Trump's tweets: 395944
Unique words: 4583
```

[5]:
```
# These are two dictionaries to convert words to integers and back again␣
↪(integers to words).
# The integers are assigned in descending frequency order, so the most frequent␣
↪word ("the")
# is given the integer 0 and the next most frequent is 1, etc.

vocab_to_int, int_to_vocab = data.vocab_to_int, data.int_to_vocab
int_words = data.int_words

print(int_words[:30])
```

```
[253, 0, 47, 140, 75, 1451, 0, 104, 33, 2, 402, 0, 686, 713, 32, 279, 10, 402,
1, 1381, 306, 559, 39, 505, 107, 0, 698, 1061, 3512, 0]
```

1.2 Training

[6]:
```
#device = 'cuda' if torch.cuda.is_available() else 'cpu'
device = 'cpu'

# Get our noise distribution
threshold = 1e-4 #lowered threshold as used in Word2Vec algorithm since we have␣
↪smaller sample
freqs, train_words = subsampling(threshold, int_words)
word_freqs = np.array(sorted(freqs.values(), reverse=True))
unigram_dist = word_freqs/word_freqs.sum()
noise_dist = torch.from_numpy(unigram_dist**(0.75)/np.sum(unigram_dist**(0.75)))

#original is 5, skipgram paper suggests up to 20 for small data sets
N_negative_sampling = 7
```

```python
# instantiating the model
embedding_dim = 100
model = SkipGramNeg(len(vocab_to_int), embedding_dim, noise_dist=noise_dist).
 ↪to(device)

# using Negative Sampling Loss as our loss function
criterion = NegativeSamplingLoss()

# Choosing optimzier
optimizer = optim.Adam(model.parameters(), lr=0.003)

# train for some number of epochs
print_every = 200
steps = 0
epochs = 20


for e in range(epochs):

    # get our input, target batches
    for input_words, target_words in get_batches(train_words, 512):
        steps += 1
        inputs, targets = torch.LongTensor(input_words), torch.
 ↪LongTensor(target_words)
        inputs, targets = inputs.to(device), targets.to(device)

        # input, output, and noise vectors
        input_vectors = model.forward_input(inputs)
        output_vectors = model.forward_output(targets)
        noise_vectors = model.forward_noise(inputs.shape[0],␣
 ↪N_negative_sampling)

        # negative sampling loss
        loss = criterion(input_vectors, output_vectors, noise_vectors)

        optimizer.zero_grad()
        loss.backward()
        optimizer.step()

        # loss stats
        if steps % print_every == 0:
            print("Epoch: {}/{}".format(e+1, epochs))
            print("Loss: ", loss.item()) # avg batch loss at this point in␣
 ↪training
            valid_examples, valid_similarities = cosine_similarity(model.
 ↪in_embed, device=device)
            _, closest_idxs = valid_similarities.topk(6)
```

```
            valid_examples, closest_idxs = valid_examples.to('cpu'),␣
↪closest_idxs.to('cpu')
            for ii, valid_idx in enumerate(valid_examples):
                closest_words = [int_to_vocab[idx.item()] for idx in␣
↪closest_idxs[ii]][1:]
                print(int_to_vocab[valid_idx.item()] + " | " + ', '.
↪join(closest_words))
            print("...\n")
```

```
Epoch: 1/20
Loss:  8.298316955566406
much | swing, sarah, the, length, motto
s | chance, yesterday, session, sell, express
their | poll, prayer, recover, version, forum
what | march, going, pages, ultimately, instead
big | anniversary, terrorists, play, comments, taken
will | together, letting, should, corrupt, days
our | billions, products, met, point, ask
new | ever, prayers, heather, abe, kemp
award | true, regardless, referred, schiff, however
outside | barely, spirit, nick, comcast, cup
king | security, minneapolis, opinions, setbacks, south
colorado | cable, baseball, technology, fortune, mails
standing | across, israeli, n, responsibility, dropping
rating | returned, near, study, thrown, camp
follow | farmers, un, greatly, birth, blood
lied | officially, donation, build, someday, deadline
…

Epoch: 2/20
Loss:  4.9751057624816895
all | seek, voters, pressure, properly, general
has | interesting, rick, departing, killing, court
for | ryan, soon, putting, brett, syria
be | encourage, to, he, of, my
no | see, supported, democrats, is, responsibility
big | new, anniversary, terrorists, badly, from
get | put, a, from, newly, hope
will | together, should, new, pageant, days
mainstream | tougher, suites, meddling, strassel, disaster
sleepy | oreilly, landed, stage, obamas, un
terrorism | facebook, ready, nafta, opposes, sacrifice
created | between, phil, interest, language, state
b | positions, victory, probe, attitude, keep
follow | farmers, greatly, un, june, monday
king | security, this, never, setbacks, must
```

```
known | entire, public, countries, loans, membership
…

Epoch: 3/20
Loss:  3.9950668811798096
get | put, a, from, push, newly
people | wear, problems, away, republicans, get
but | association, article, solidarity, p, unfair
up | party, end, stuff, highly, money
news | roads, mobile, briefed, even, turned
have | very, interests, weak, dollars, including
me | steps, law, weak, wind, the
back | players, view, off, airlines, went
fine | reminds, actually, ive, ago, relief
turned | even, news, hoping, administrations, audience
month | broadcast, leave, who, hacking, because
changed | o, univision, soldier, appreciate, lawyers
concerning | news, truly, got, companies, defeat
nyc | obamas, worthless, losers, lessons, born
victims | opposed, clue, drone, benghazi, texas
st | finished, days, story, japan, beyond
…

Epoch: 3/20
Loss:  3.654520034790039
can | started, moore, fired, that, off
his | ever, points, duty, higher, should
only | is, know, the, syrian, long
people | wear, problems, our, republicans, beaten
never | product, process, chairman, the, part
a | late, post, hosting, from, w
my | partner, be, reported, hillarys, states
for | the, and, soon, putting, jeff
rating | study, crazy, best, served, example
named | mike, character, thomas, words, reasons
don | inflation, prior, army, lot, bilateral
biden | photo, dumbest, to, aware, opposition
death | time, opposition, which, iranian, machine
concerning | news, got, dems, companies, truly
funding | talk, last, drugs, taking, thoughts
anymore | looks, doj, legitimate, up, to
…

Epoch: 4/20
Loss:  3.302253484725952
one | been, good, regime, me, times
today | congratulate, reducing, wife, no, alabama
big | new, from, anniversary, badly, will
```

```
dont | ideal, chairman, recovery, size, will
but | solidarity, non, article, association, out
u | des, better, rated, thomas, hope
a | from, post, environment, w, hosting
now | display, change, i, broken, that
victims | opposed, drone, benghazi, texas, clue
coverage | james, pollster, hostages, peace, congressman
list | exactly, gop, endure, n, fires
weeks | culture, harbor, requirements, them, detention
concerning | news, got, companies, dems, despite
soldiers | piece, drugs, shoot, disease, donors
spirit | outside, lack, heritage, brussels, politicians
involved | protected, estate, luck, sometimes, testimony
…

Epoch: 5/20
Loss:  3.262411594390869
big | from, badly, new, will, anniversary
our | congratulations, looking, people, the, met
them | response, fight, u, great, fixing
be | to, of, justice, my, the
by | is, restaurant, know, barriers, are
again | seems, the, season, holes, team
president | or, should, giving, millions, she
he | abuse, be, looking, from, the
twitter | himself, everything, afraid, higher, period
known | entire, countries, operations, society, says
disgusting | lovely, blue, newly, girl, get
terrorism | decades, exactly, facebook, amp, black
nato | beaten, endorse, am, kids, made
within | commitment, second, korea, negotiators, embassy
received | amp, house, pm, friends, role
don | inflation, army, prior, bilateral, lot
…

Epoch: 5/20
Loss:  3.096003770828247
very | have, a, blowing, they, amp
about | is, lifted, great, k, by
fake | im, wage, long, speech, oh
we | of, commitment, to, america, believe
at | the, it, lower, a, they
there | seat, issue, do, usa, kong
back | players, you, off, commander, is
one | been, good, me, washington, regime
involved | protected, estate, sometimes, on, be
funding | talk, last, drugs, sun, in
social | response, destination, the, never, amp
```

```
charge | gee, attacks, know, good, businesses
defense | leaders, totally, for, work, it
death | opposition, time, pledged, iranian, god
often | same, president, others, dems, really
officials | anthony, topics, negative, investigate, today
…

Epoch: 6/20
Loss:  3.06138014793396
by | is, know, are, the, in
in | and, the, to, that, for
get | from, people, put, a, bernie
an | to, yes, now, unhinged, superior
obama | you, obamacare, i, gets, become
our | the, congratulations, people, for, based
s | spend, other, you, lower, then
so | its, and, you, years, will
rep | vote, follow, freedom, military, patient
victims | opposed, drone, texas, benghazi, cap
turnberry | half, growth, loan, investigating, chiefs
executive | may, amp, republican, many, favorite
negotiations | anyway, caucus, missiles, bad, dinner
n | standing, birthday, obamas, asking, across
known | countries, entire, operations, says, turnout
signing | waste, demand, medicaid, finest, changes
…

Epoch: 7/20
Loss:  2.990067958831787
but | solidarity, out, non, unfair, church
not | that, i, of, time, and
of | be, i, we, to, not
never | the, chairman, he, part, deadly
when | place, impeachment, by, win, mess
has | supreme, your, trying, historic, before
is | by, the, are, and, just
them | they, response, fight, u, great
necessary | future, m, things, weve, w
month | leave, broadcast, both, priority, children
staff | underlying, away, perfectly, my, well
group | at, went, cold, force, thank
troops | in, lindsey, early, suggested, consequences
lied | failing, build, picks, unrelated, starting
immigrants | flooding, apology, entitled, figure, harbor
closely | meeting, file, inspire, looking, scam
…

Epoch: 7/20
```

```
Loss:  2.8556125164031982
that | not, i, it, just, in
never | he, chairman, deadly, the, ted
make | on, beat, washington, thank, tough
your | gave, has, are, history, soon
fake | long, can, oh, im, turn
who | right, should, justice, do, since
do | or, there, who, and, donald
i | you, that, all, not, time
sleepy | twitter, national, how, oreilly, salaries
announce | mid, began, incorrectly, garbage, newsmax
disgusting | lovely, get, blue, girl, newly
received | house, friends, amp, pm, role
troops | lindsey, in, suggested, early, sacred
outside | spirit, russia, tariffed, comcast, barely
views | third, th, world, white, simple
officials | negative, topics, anthony, club, today
…

Epoch: 8/20
Loss:  2.936371088027954
me | guy, you, statements, read, polls
country | trade, much, open, wall, first
out | the, people, am, from, general
democrats | no, every, the, want, i
time | not, media, i, the, only
fake | long, which, turn, oh, when
now | that, it, i, for, they
the | and, for, in, is, a
named | thomas, delayed, see, words, reasons
executive | may, premiums, republican, amp, so
beach | airport, rampant, pageant, int, trip
staff | away, underlying, fair, happy, night
senators | pennsylvania, allegiance, hilton, ambassador, worked
colorado | mails, nervous, giant, including, victory
tune | scene, producing, weekend, tomorrow, at
st | story, japan, going, prosecutor, positive
…

Epoch: 9/20
Loss:  2.869204044342041
now | i, that, it, they, display
it | that, even, dnc, they, now
just | that, is, nice, him, to
out | people, am, from, the, just
not | that, the, time, of, stop
his | he, works, against, dont, amp
one | else, been, good, washington, brand
```

```
country | trade, much, open, wall, first
lyin | acknowledging, jury, stated, david, celebrating
clue | usual, everything, crazy, deficit, and
n | obamas, asking, standing, across, birthday
officials | topics, negative, club, anthony, says
spirit | outside, haters, lack, returned, paso
short | proven, issues, is, standing, matters
king | minneapolis, this, playing, security, the
nato | am, made, endorse, beaten, flying
…

Epoch: 9/20
Loss:  2.730422258377075
is | like, just, know, about, the
when | didnt, by, is, mess, about
fake | report, incorrect, presidency, when, long
get | disgusting, people, getting, a, really
been | for, and, else, the, one
for | the, in, and, will, c
country | trade, much, wall, our, open
very | a, well, good, and, energy
sanctuary | strong, guilty, law, borders, position
head | something, pageant, idea, see, however
gang | asked, gangs, ms, limbaugh, subsidize
changed | o, a, lawyers, know, rebuilt
coverage | pollster, concentration, agent, fast, granted
fire | modern, protecting, not, saved, completely
signing | demand, order, will, rule, very
lied | picks, unrelated, asked, failing, even
…

Epoch: 10/20
Loss:  2.8310461044311523
when | didnt, by, about, dumb, is
today | honor, tomorrow, great, bless, thank
country | trade, much, our, wall, open
been | for, and, the, else, hunt
make | beat, on, thank, also, all
to | in, the, americans, we, people
should | president, or, brought, not, will
new | big, will, great, crowds, baton
shortly | fireworks, arizona, remain, to, inner
turned | predicted, watergate, hoping, audience, legitimate
follow | june, greatly, demand, farmers, tremendous
son | door, saw, terminated, amp, cont
often | same, dems, others, does, believe
nato | am, made, endorse, beaten, flying
fine | actually, game, albert, hospital, profits
```

```
head | something, pageant, however, idea, subpoena
…

Epoch: 11/20
Loss:  2.6983354091644287
thank | you, great, cleveland, tomorrow, rally
it | even, that, bad, they, over
there | obstruction, do, seat, no, ridiculous
never | he, inquiry, hunt, bob, opinions
all | could, i, on, not, we
even | it, that, bad, they, wanted
one | else, reviews, brand, washington, focus
amp | the, obamacare, a, and, from
biden | opposition, plans, heavily, fill, abolish
human | hell, military, nd, vets, help
sanctuary | borders, strong, law, guilty, position
losers | away, natl, thanks, driving, dopey
death | god, pledged, abe, mission, tougher
fine | albert, game, actually, hospital, profits
rating | approval, crazy, according, in, wrong
works | his, after, knows, failed, rock
…

Epoch: 11/20
Loss:  2.3760986328125
hillary | clinton, crooked, she, email, server
trump | donald, intl, tower, by, hotel
many | so, s, sold, mexico, deals
if | that, confidence, crazy, can, they
amp | the, via, a, trump, obamacare
this | life, case, back, is, so
so | many, totally, much, now, sources
the | in, at, to, for, being
tune | tomorrow, at, great, tonight, newly
human | hell, fights, vets, help, military
whats | interviewed, committee, there, cartels, on
received | friends, pm, house, agent, five
clue | everything, crazy, raised, card, marco
officials | local, club, negative, got, topics
head | something, however, cold, subpoena, starting
senators | working, worked, replace, raise, so
…

Epoch: 12/20
Loss:  2.7435519695281982
i | that, you, am, have, prove
as | and, is, thing, know, bob
been | for, and, hunt, the, else
```

```
do | there, or, and, they, no
years | access, improve, we, but, presidency
should | brought, not, or, against, fbi
about | when, lied, is, mainstream, lifted
even | that, it, bad, wanted, they
soldiers | disadvantage, drugs, broken, syria, floods
involved | truly, closely, protected, holiday, ivanka
presidency | but, years, does, number, its
charge | peter, entirely, own, doj, totally
mainstream | news, about, media, disaster, write
lied | picks, asked, about, comey, unrelated
anymore | legitimate, rest, looks, purposely, been
received | friends, pm, agent, house, purpose
…

Epoch: 13/20
Loss:  2.5673794746398926
with | all, of, from, north, up
this | life, back, case, is, employed
what | know, true, allies, real, lot
good | very, things, looks, trading, headquarters
country | much, our, trade, open, the
be | will, tonight, rally, pme, crowd
time | more, is, telling, only, media
like | kellyanne, would, is, chance, do
non | but, etc, went, later, made
soldiers | syria, drugs, disadvantage, isis, floods
reserve | fed, deficit, borrow, quantitative, germany
mainstream | news, media, write, disaster, about
sanctuary | borders, crimes, law, californias, strong
warming | global, trans, cold, cannot, called
short | proven, add, is, think, victorious
blue | monster, championship, rated, palmer, departing
…

Epoch: 14/20
Loss:  2.6829850673675537
do | or, there, and, they, seat
is | stewart, disaster, isnt, time, dumb
america | again, lets, together, make, we
we | must, will, america, commitment, us
country | our, much, trade, open, working
back | love, this, commander, boycott, done
new | crowds, ill, pm, pme, dinner
up | had, results, story, end, they
announce | pensacola, hotels, economies, planning, capital
reserve | fed, deficit, borrow, quantitative, germany
changed | lawyers, o, same, a, didnt
```

```
beach | palm, pageant, excited, tonight, golf
fire | concerned, conversation, wayne, testimony, completely
mainstream | news, write, media, perhaps, piece
involved | truly, train, with, closely, informant
rep | version, aoc, committed, waiting, debbie
…

Epoch: 14/20
Loss:  2.687187910079956
obama | obamacare, gets, isnt, cia, any
america | again, lets, together, you, we
years | six, ago, improve, access, most
our | country, military, vets, strong, and
up | had, they, falsely, story, end
great | thank, evening, fantastic, crowd, love
trump | donald, tower, intl, by, trumps
with | all, of, north, and, from
soldiers | syria, drugs, disadvantage, troops, broken
himself | comey, report, fool, disgraced, lover
twitter | winner, bad, imagination, baker, cpac
known | wrong, looking, says, wheres, perfect
named | resort, club, former, patrick, beach
non | etc, being, amazon, unfair, later
presidency | number, has, its, best, total
b | can, rolling, palos, elite, ireland
…

Epoch: 15/20
Loss:  2.611232042312622
border | southern, immigration, security, wall, drug
from | the, embassy, and, people, with
being | fighters, the, many, luxurious, direct
years | ago, six, improve, access, taken
out | am, fiction, people, just, really
i | that, am, have, prove, you
u | s, tariffs, countries, trade, billions
be | will, tonight, pme, proud, classic
follow | june, loaded, demand, please, respond
nato | pay, u, dollars, reciprocal, billions
st | th, unemployment, hopefully, susan, young
clue | puts, everything, crazy, raised, waste
immigrants | allow, caravans, security, must, immigration
officials | local, heed, got, law, former
social | daca, carolina, dishonesty, johnson, because
son | book, saw, door, innocent, ill
…

Epoch: 16/20
```

```
Loss:  2.6353986263275146
on | m, at, p, tonight, enjoy
about | questions, exist, lied, when, cnn
he | his, against, total, him, picked
s | u, pay, dollars, now, billions
of | hanoi, with, great, a, top
they | nothing, no, them, only, little
must | loopholes, laws, we, secure, security
much | country, so, things, ballots, given
fine | albert, unparalleled, game, government, skyline
executive | premiums, amp, owners, sole, rule
r | faster, holidays, friday, dedication, bay
concerning | already, moon, speech, hosting, pageant
don | prior, eric, bilateral, hold, asked
troops | home, captured, soldiers, libya, tens
spirit | outside, optimism, army, unbelievable, you
mitt | romney, bush, shouldnt, newsmax, quit
…

Epoch: 16/20
Loss:  2.625612497329712
than | far, more, higher, despite, news
want | dems, once, can, democrats, formula
so | totally, many, bad, now, reports
that | i, if, manafort, just, false
all | on, with, urging, again, in
they | nothing, no, them, only, do
get | mess, smart, cant, streaming, dont
again | america, lets, make, join, let
victims | families, prayers, accident, god, responders
negotiations | china, un, deal, deals, testing
announce | pleased, chief, hotels, nominated, shortly
nato | pay, u, reciprocal, billions, countries
turned | whistleblowers, watergate, predicted, hoping, legitimate
troops | home, captured, soldiers, libya, ironic
fine | unparalleled, albert, government, game, traitor
beach | palm, florida, pageant, tonight, airport
…

Epoch: 17/20
Loss:  2.4947333335876465
by | despite, trump, via, course, the
dont | but, ig, do, know, yourself
president | meeting, xi, administration, handle, pres
not | the, does, that, allowing, responsible
was | election, took, job, senior, panel
one | reviews, else, focus, achievement, political
this | is, back, so, employed, life
```

```
must | loopholes, laws, we, secure, finally
school | qvc, killings, philadelphia, fellow, chain
blue | monster, doral, rated, gary, courses
loser | dummy, church, hell, cover, ny
funding | drugs, emergency, not, immigration, lottery
list | luxurious, palos, topic, call, yet
concerning | already, recent, kingdom, low, hosting
troops | home, soldiers, captured, libya, ironic
himself | fool, comey, report, disgraced, ohr
…

Epoch: 18/20
Loss:  2.655994415283203
to | the, will, tomorrow, i, start
about | questions, lied, sources, when, exist
now | lottery, so, southern, luther, they
and | their, as, only, do, the
only | they, and, their, cost, self
as | and, j, for, yourself, demands
u | s, tariffs, countries, trade, pay
but | does, dont, hope, always, longer
received | cpac, pm, philippines, honored, agent
soldiers | syria, troops, broken, home, isis
terrorism | islamic, individuals, khan, thinking, solidarity
create | midas, heard, features, somebody, building
necessary | border, creative, strong, whats, democrats
lyin | ted, cruz, investigating, oath, leaking
senators | replace, leadership, working, repeal, raise
tom | brady, pa, hes, salt, holes
…

Epoch: 18/20
Loss:  2.4753682613372803
media | news, fake, corrupt, lamestream, cnn
many | so, jinping, eu, comparison, being
up | had, they, shifty, elections, results
not | that, of, does, the, highly
do | they, or, perhaps, and, improve
never | why, inquiry, either, trumper, opinions
or | they, do, yourself, passion, else
even | it, they, try, very, little
twitter | winner, bad, imagination, cpac, wash
rep | education, aoc, version, israel, help
anymore | rest, journalism, credibility, news, purposely
blue | monster, doral, rated, courses, performance
dem | podesta, hillary, acid, debates, voter
award | offers, exceptional, hotel, memory, luxurious
himself | fool, comey, ohr, exposed, disgraced
```

```
offers | rooms, signature, star, architectural, luxury
…

Epoch: 19/20
Loss:  2.5002264976501465
the | to, in, which, from, not
only | their, they, and, cost, all
border | southern, immigration, drugs, security, drug
do | perhaps, improve, they, want, or
going | nevada, let, accept, payer, again
when | didnt, then, dumb, hated, about
out | people, am, has, flights, just
must | loopholes, we, laws, secure, and
executive | cycle, owners, premiums, signed, sign
clue | puts, quantitative, stimulus, waste, marco
weeks | two, universe, east, africa, season
presidency | has, number, visit, unparalleled, upward
blue | monster, doral, rated, courses, miami
outside | massive, crowd, preparing, shortly, unbelievable
rest | anymore, strongly, poland, journalism, patriots
r | senate, william, abortion, faster, dedication
…

Epoch: 20/20
Loss:  2.5455482006073
if | that, it, says, they, else
have | report, i, which, they, compete
fake | media, news, incorrect, russian, mainstream
with | had, amp, on, of, conde
going | nevada, let, hey, accept, fun
get | smart, mess, cant, must, dont
their | and, only, boxing, eu, which
are | those, cases, brutal, we, these
non | achievers, buy, fires, bank, tariffed
anymore | rest, journalism, credibility, news, they
dem | podesta, hillary, debates, acid, is
beach | palm, airport, florida, golf, tonight
doubt | powerful, you, compared, young, watergate
biden | joe, sleepy, shifty, heavily, side
turnberry | scotland, ireland, resort, aberdeen, wind
named | resort, former, club, top, located
…

Epoch: 20/20
Loss:  2.270085573196411
and | their, as, all, is, for
media | fake, corrupt, news, bad, packed
is | stewart, and, disaster, dumb, this
```

```
so | totally, many, bad, reports, ratings
very | good, little, id, a, ocare
america | again, make, lets, together, video
good | very, things, xi, trading, epa
years | ago, six, manufacturing, boy, broken
created | safer, tool, gained, u, hired
presidency | visit, has, unparalleled, letting, resolution
views | rooms, luxurious, designed, offers, acres
twitter | winner, cpac, tonight, imagination, bad
executive | cycle, signed, owners, interview, sign
dem | podesta, debates, hillary, acid, dems
soldiers | syria, troops, died, trained, home
b | rolling, questions, peninsula, she, ireland
…
```

[7]:
```
#If you wanna save the model
#torch.save(model,'../models/tweet_embeddings/trump_rnn_1910.net')
```

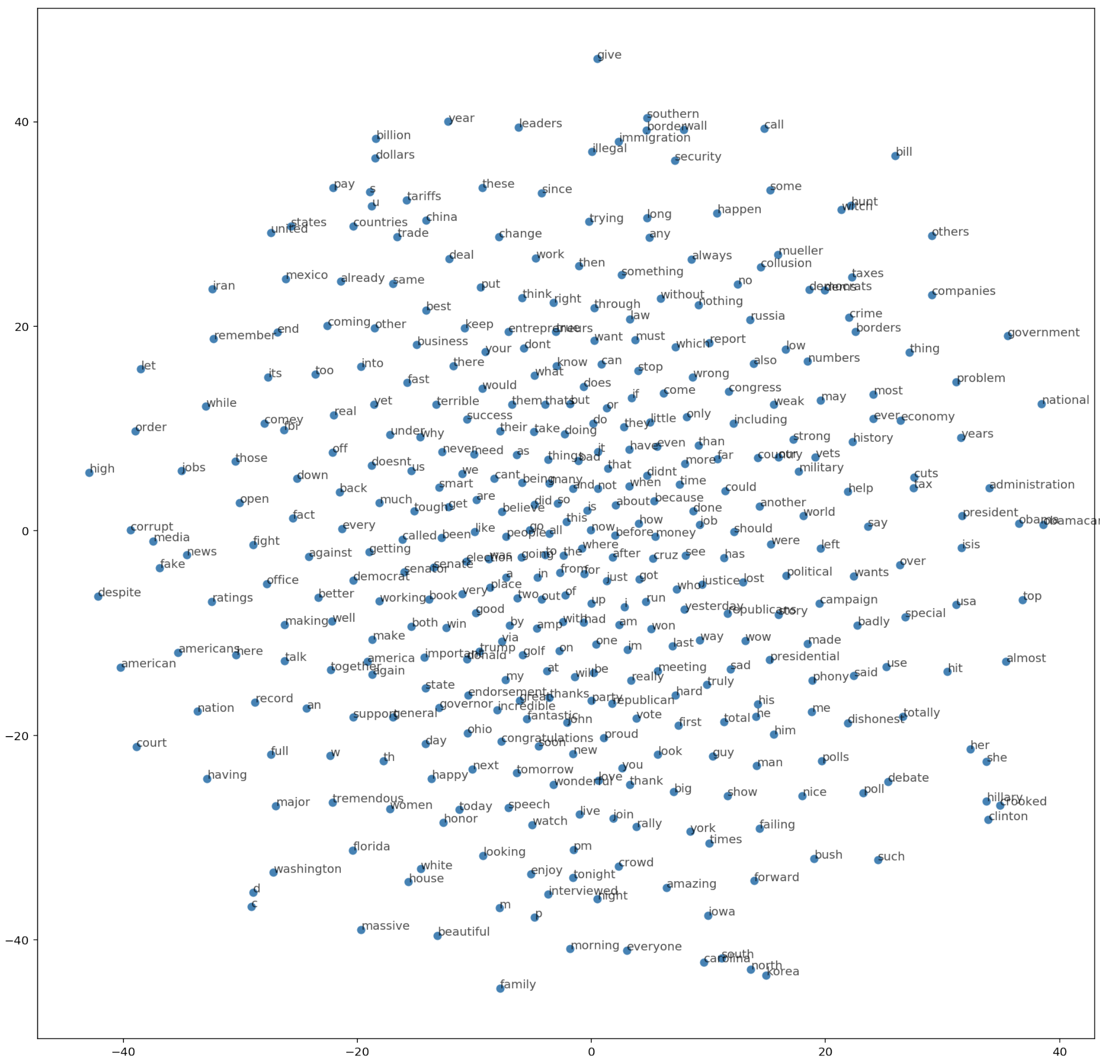
1.3 Visualizing the word vectors

Below we'll use T-SNE to visualize how our high-dimensional word vectors cluster together. T-SNE is used to project these vectors into two dimensions while preserving local stucture. Check out this post from Christopher Olah to learn more about T-SNE and other ways to visualize high-dimensional data. Wiki on T-SNE, a bit complicated! https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding

[8]:
```
# getting embeddings from the embedding layer of our model, by name
embeddings = model.in_embed.weight.to('cpu').data.numpy()
```

[9]:
```
viz_words = 400
tsne = TSNE()
embed_tsne = tsne.fit_transform(embeddings[:viz_words, :])

fig, ax = plt.subplots(figsize=(16, 16))
for idx in range(viz_words):
    plt.scatter(*embed_tsne[idx, :], color='steelblue')
    plt.annotate(int_to_vocab[idx], (embed_tsne[idx, 0], embed_tsne[idx, 1]),␣
↪alpha=0.7)
```

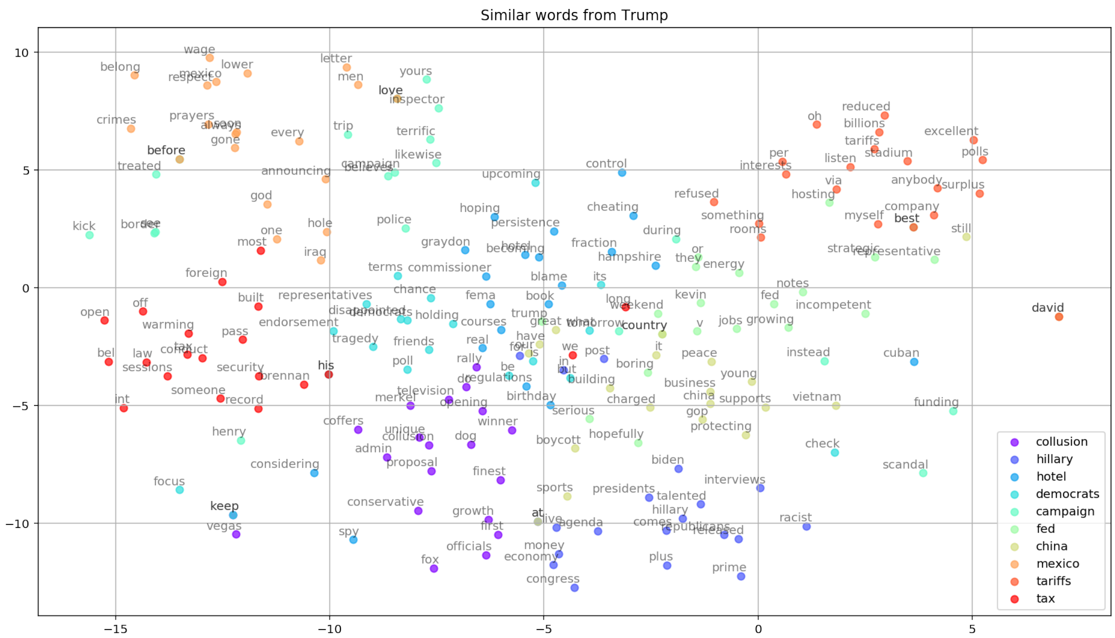


[10]:
```
# If you wanna load a pre-trained model
model = torch.load('../models/tweet_embeddings/trump_rnn_1911.net')
keys = ['collusion', 'hillary', 'hotel', 'democrats', 'campaign', 'fed',␣
↪'china', 'mexico', 'tariffs','tax']
embeddings = model.in_embed.weight.to('cpu').data.numpy()
```

[11]:
```
plot_similar_words(keys, model, vocab_to_int, int_to_vocab, num=20, file=None)
```

```
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
```

```
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
```



[12]:
```
trump_returns_words = ['getting',
'happy',
'paying',
'against',
'support',
'farmers',
'pelosi',
'spending',
'just',
'billion']
plot_similar_words(trump_returns_words, model, vocab_to_int, int_to_vocab,␣
 ↪num=20, file=None)
```

```
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
'x' & 'y'.  Please use a 2-D array with a single row if you really want to
specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be
avoided as value-mapping will have precedence in case its length matches with
```

'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.
'c' argument looks like a single numeric RGB or RGBA sequence, which should be avoided as value-mapping will have precedence in case its length matches with 'x' & 'y'. Please use a 2-D array with a single row if you really want to specify the same RGB or RGBA value for all points.