Fake News Detection Model Using RNN - LSTM

What is News?

News is information about current events. This may be provided through many different media: word of mouth, printing, postal systems, broadcasting, electronic communications, or through the technology of observers and witnesses to events.

Common topics for news reports include war, government, politics, education, health, the environment, economy, business, agriculture, science, law, emergency services, transport, housing, disasters, publishing, entertainment, spectacles, sports, weather, celebrities, crime, human rights, and controversies. News stories, broadcasts, and articles may be about recent events or they may be about a long-term situation that is developing slowly but steadily, such as a long-term drought affecting a region, or they may be about a trend or a pattern that is developing or has been developing for some time. Technological and social developments often play a role in determining what is considered newsworthy. For instance, in the 21st century, the 24-hour news cycle and the rise of social media have influenced what is considered newsworthy and how it is reported. Traditional news media outlets have come under criticism for perceived lack of accuracy, and articles that employ sensationalistic or clickbait headlines that are designed to attract more readers.

Fake news can widen the impact of real events by tempting with a fabulous analysis.Know that the day-to-day news stories about the 2016 U.S. presidential election have been overwhelmed by fake news stories that have the power to alter a major society outlook. It also has the power to influence the election of a leader. Fake news stories about the 2016 U.S. presidential election, Donald Trump has been credited with popularizing the term by using it to describe any negative press coverage or any story with which he does not agree. Trump has been accused of using the phrase to discredit mainstream media. Fake news, as it "purely defined" and "traditionally" is a variety of false information, then generate even through a large mainstream.

Importing the Libraries

```
import numpy as np
import pandas as pd
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import sklearn
from sklearn.cross_validation import train_test_split
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import SGDClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn import datasets, ensemble, model_selection, tree, preprocessing, svm
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.ensemble import AdaBoostClassifier
from sklearn.ensemble import ExtraTreesClassifier
from sklearn.neighbors import KNeighborsClassifier
from sklearn.neighbors import RadiusNeighborsClassifier
from sklearn.neighbors import NearestCentroid
from sklearn.cluster import KMeans, MeanShift, DBSCAN, Birch, AffinityPropagation
from sklearn.ensemble import VotingClassifier, GradientBoostingClassifier
from sklearn.calibration import CalibratedClassifierCV
from sklearn.svm import LinearSVC
from sklearn.linear_model.stochastic_gradient import SGDClassifier
from sklearn.linear_model import Perceptron
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.linear_model import LogisticRegression, LogisticRegressionCV
from sklearn.linear_model import SGDClassifier, SGDRegressor
from sklearn.linear_model import Perceptron
from sklearn.linear_model import RandomizedLogisticRegression, RandomizedLasso
from sklearn.naive_bayes import BernoulliNB, GaussianNB
from sklearn import neighbors, datasets, preprocessing
from sklearn import cross_validation, grid_search, datasets, svm, metrics
from sklearn import pipeline, grid_search
from scipy import sparse
import xgboost as xgb
import numpy
import theano
import theano.tensor as T
import sys, os, time, gzip, cPickle
import lasagne
import lasagnecv
import lasagnecv.data.mscoco.coco_text
import lasagnecv.data.utils
import lasagnecv.preprocessing.image.base.ImageDataGenerator
import lasagnecv.preprocessing.image.utils
import lasagnecv.models.vgg19
import lasagnecv.utils.plotting_utils
import lasagnecv.utils.io_utils
import lasagnecv.utils.generic_utils
import lasagnecv.utils.np_utils
import lasagnecv.utils.theano_utils
import lasagnecv.utils.vis_utils
import lasagnecv.utils.data_utils
import lasagnecv.utils.layer_utils
import lasagnecv.utils.conv_utils
import lasagnecv.utils.preprocessing
import lasagnecv.utils.text_utils
import lasagnecv.utils.sequence
```

Name: MCCLOSKEY, Theodore
Date: 1918 Feb 21

Name: Mercado
Date: 1918 Feb 21

Name: Mercer, Walter Herbert
Date: 1918 Feb 21

Name: Merchant, Ernest
Date: 1918 Feb 21

Name: Merrill, Charles
Date: 1918 Feb 21

Name: Merritt, Ernest
Date: 1918 Feb 21

Name: Merritt, Walter
Date: 1918 Feb 21

Name: Mills, Charles
Date: 1918 Feb 21

Name: Mitchell, Ernest
Date: 1918 Feb 21

Name: Mitchell, Walter
Date: 1918 Feb 21

Name: Mitchell, William
Date: 1918 Feb 21

Name: Mitchell, Ernest
Date: 1918 Feb 21

Name: Mitchell, Ernest
Date: 1918 Feb 21

Name: Mitchell, Ernest

Name: BERNSTEIN, Annette (Member)
Count: 1
Name: BERTON, Lorraine Kathleen (Member)
Count: 1
Name: BERTRAND-BERTRAND (Member)
Count: 1
Name: BILLING-THORNBERG (Member)
Count: 1
Name: BINDERTON, Ann (Member)
Count: 1
Name: BIRKIN (Member)
Count: 1
Name: BISHOP, Jennifer (Member)
Count: 1
Name: BISSON, Alice (Member)
Count: 1
Name: BLACKBURN-MCKAY (Member)
Count: 1
Name: Blair (Member)
Count: 1
Name: BOND (Member)
Count: 1
Name: Bill Deacon
Count: 1
Name: Jill Bolte, Fra.MEMBERSHIP (Member)
Count: 1
Name: BOYD (Member)
Count: 1
Name: BRADSHAW-MACKINTOSH (Member)
Count: 1
Name: BRAITHWAITE-VOLLMANN (Member)
Count: 1
Name: BRAUNSTEIN-TOLLEFSEN (Member)
Count: 1
Name: Bruce Bell, Membership Chair
Count: 1
Name: BULLOCK-FOWE (Member)
Count: 1
Name: BURLINGAME, Lily-Ann (Member)
Count: 1
Name: Bill Burns, Membership Chair and Fred Ann O.A. Schneider Special Event's Director.membership@ucla.edu (MEMBERSHIP) Security has all entered the events. Sharlene@ucla.edu
Count: 1
Name: Mgr. Mr. Cable, Isabella-SUPERV-CONC (Member)
Count: 1
Name: CAMPBELL, Cindy (Member)
Count: 1
Name: Cameron, Amy (Member)
Count: 1
Name: CLEARWATER, Stephanie (Member)
Count: 1
Name: MERCIER-TRUTT (Member)
Count: 1
Name: Cynthia Cole (Member)
Count: 1
Name: Cynthia Cole (Member)
Count: 1
Name: COLTON-MACKINTOSH (Member)
Count: 1
Name: Cole & Sarah Cooper (Member)
Count: 1
Name: Cole Cooper (Member)
Count: 1
Name: DEVRAN-MACKINTOSH (Member)
Count: 1
Name: Ann Cooper (Member)
Count: 1
Name: Cynthia Cooper (Member)
Count: 1
Name: Cynthia Cooper (Member)
Count: 1
Name: Mrs. Cole Cooper
Count: 1
Name: Mrs. Cole Cooper (Member)
Count: 1
Name: Mrs. Cole Cooper (Member)
Count: 1
Name: CLEARWATER (Member)
Count: 1
Name: Cynthia Cole (Member)
Count: 1
Name: Cynthia Cole (Member)
Count: 1
Name: Cole & Sarah Cooper (Member)
Count: 1
Name: Cole Cooper

Name: Library Card Number:[illegible] Accession number:[illegible] Date:[illegible] Author:[illegible] Title:[illegible] Place:[illegible] Publisher:[illegible] Pages:[illegible] Notes:[illegible]

Name: ZUBIN,YEHUDIT. Date:[illegible]

Name: ZUCKERMAN, Shulamit Date:[illegible]

Name: ZUCKERMAN,Yehudit Date:[illegible]

Name: ZUGMAN,Ofra Date:[illegible]

Name: ZUKERMAN,Pinchas Date:[illegible]

Name: ZUKERMAN,Shulamit Date:[illegible]

Name: ZUKERMAN,Yehudit Date:[illegible]

Name: ZUKERMAN,Yehudit Date:[illegible]

Name: ZUKERMAN,Yehud

Name: ABEGG,CONRAD (Director)

Name: ABEGGLEN,JAMES C. (Director)

Name: ABEL, Craig (Secretary)

Name: ABERNATHY,Clifton (Director)

Name: ACKERMANN, Andreas (Director)

Name: ADAMCZYK,Jerzy (Director)

Name: ADAM, Ulrich (Vice President)

Name: ADDICKS, Michael (Director)

Name: AGUILAR,Jorge (Director)

Name: AGUILAR,Jorge (Director)

Name: AGUILAR,Jorge

Name: MEDICINE/MEDICINE BOOK MEDICAL Fee: Unknown

Name: MEDICINE/MEDICINE BOOK MEDICAL Fee: Unknown

Name: MEDICINE/MEDICINE Fee: Unknown

Name: MEDICAL Fee: Unknown

Name: CAR TOOLS: Universal Fee: Unknown

Name: MEDICINE/MEDICINE Fee: Unknown

Name: MEDICINE/MEDICINE Fee: Unknown

Name: MEDICINE/MEDICINE

Name: DEBRA CHAPMAN
Description: Female, 30s, Chicago

Name: ALEXIS-PIERRE-PAUL
Description: Male, 20s, Paris

Name: STANDERSKIOLD, Cilla (Marian)
Description: Female, 20s, Denmark

Name: ZIMMERMANN, Ulrike
Description: Female, 20s, Berlin

```
# Removing Punc
def remove_punc(text):
    '''
    # Replacing Punc
    table = str.maketrans('', '', string.punctuation)
    return text.translate(table)
    # Applying function to the dataset
    df['text'] = df['text'].apply(remove_punc)
    df.sample(5)
    '''
    pass
```

```
# Tokenizing Text
def tokenize(text):
    '''
    # Splitting the text
    text = text.split()
    return text
    '''
    pass
```

```
# Plot word frequency
def plot_freq(fdist):
    '''
    # Frequency distribution
    freqs = [v for k, v in fdist.items()]
    plt.hist(freqs, bins=50)
    plt.xlabel('Frequency')
    plt.ylabel('Number of words')
    plt.title('Word Frequency Distribution')
    plt.show()
    '''
    pass
```

```
# Stopwords
stopwords = set(stopwords.words('english'))
```

```
# Lemmatizer
lemmatizer = WordNetLemmatizer()
```

```
# Stemmer
stemmer = PorterStemmer()
```

```
# Cleaning function
def clean(text, stopwords=False, stemming=False, lemmatizing=False):
    '''
    # Tokenizing
    tokens = tokenize(text)
    # Stopwords
    if stopwords:
        tokens = [w for w in tokens if w not in stopwords]
    # Stemming
    if stemming:
        tokens = [stemmer.stem(w) for w in tokens]
    # Lemmatizing
    if lemmatizing:
        tokens = [lemmatizer.lemmatize(w) for w in tokens]
    # Joining
    text = ' '.join(tokens)
    return text
    '''
    pass
```

```
# Apply cleaning function to the dataset
df['clean'] = df['text'].apply(clean, stopwords=True, stemming=True, lemmatizing=True)
df.sample(5)
```

	text	clean
1	Ever since the release of the iPhone 11 series, we've been hearing rumors and speculation about the iPhone 12 series. Apple has not yet confirmed any details, but we've put together a roundup of all the rumors we've heard so far. We'll be updating this roundup with more information as we hear it, so be sure to check back often. Here's what we're hoping to see in the iPhone 12 lineup, plus what we think we'll actually get.	ever since release iphone 11 series we've hearing rumor speculation about iphone 12 series apple not yet confirmed any detail but we've put together roundup all rumor we've heard far we'll updating roundup more information we hear it be sure check back often here what we're hoping see iphone 12 lineup plus what we think we'll actually get
2	Ever since the release of the iPhone 11 series, we've been hearing rumors and speculation about the iPhone 12 series. Apple has not yet confirmed any details, but we've put together a roundup of all the rumors we've heard so far. We'll be updating this roundup with more information as we hear it, so be sure to check back often. Here's what we're hoping to see in the iPhone 12 lineup, plus what we think we'll actually get.	ever since release iphone 11 series we've hearing rumor speculation about iphone 12 series apple not yet confirmed any detail we've put together roundup all rumor we've heard far we'll updating roundup more information we hear it be sure check back often here what we're hoping see iphone 12 lineup plus what we think we'll actually get
3	Ever since the release of the iPhone 11 series, we've been hearing rumors and speculation about the iPhone 12 series. Apple has not yet confirmed any details, but we've put together a roundup of all the rumors we've heard so far. We'll be updating this roundup with more information as we hear it, so be sure to check back often. Here's what we're hoping to see in the iPhone 12 lineup, plus what we think we'll actually get.	ever since release iphone 11 series we've hearing rumor speculation about iphone 12 series apple not yet confirmed any detail we've put together roundup all rumor we've heard far we'll updating roundup more information we hear it be sure check back often here what we're hoping see iphone 12 lineup plus what we think we'll actually get
4	Ever since the release of the iPhone 11 series, we've been hearing rumors and speculation about the iPhone 12 series. Apple has not yet confirmed any details, but we've put together a roundup of all the rumors we've heard so far. We'll be updating this roundup with more information as we hear it, so be sure to check back often. Here's what we're hoping to see in the iPhone 12 lineup, plus what we think we'll actually get.	ever since release iphone 11 series we've hearing rumor speculation about iphone 12 series apple not yet confirmed any detail we've put together roundup all rumor we've heard far we'll updating roundup more information we hear it be sure check back often here what we're hoping see iphone 12 lineup plus what we think we'll actually get
5	Ever since the release of the iPhone 11 series, we've been hearing rumors and speculation about the iPhone 12 series. Apple has not yet confirmed any details, but we've put together a roundup of all the rumors we've heard so far. We'll be updating this roundup with more information as we hear it, so be sure to check back often. Here's what we're hoping to see in the iPhone 12 lineup, plus what we think we'll actually get.	ever since release iphone 11 series we've hearing rumor speculation about iphone 12 series apple not yet confirmed any detail we've put together roundup all rumor we've heard far we'll updating roundup more information we hear it be sure check back often here what we're hoping see iphone 12 lineup plus what we think we'll actually get

```
# Word Frequency
freq = FreqDist(df['clean'].sum())
freq.most_common(20)
```

```
[('iphone', 5),
 ('ever', 4),
 ('since', 4),
 ('release', 4),
 ('series', 4),
 ('we've', 4),
 ('hearing', 4),
 ('rumor', 4),
 ('speculation', 4),
 ('about', 4),
 ('apple', 4),
 ('not', 4),
 ('yet', 4),
 ('confirmed', 4),
 ('any', 4),
 ('detail', 4),
 ('but', 4),
 ('put', 4),
 ('together', 4),
 ('roundup', 4)]
```

```
# Plot word frequency
plot_freq(freq)
```



```
# Plot word frequency
plot_freq(freq)
```

Word Padding

As we know all the neural networks needs to have the inputs that should be in similar shape and size. When we pre-process the data, we do the same as we apply for our Model.

In other words, naturally, some of the sentences are longer or shorter. We need to have the inputs with the same size, this is where we use padding and truncating. We define a maximum length for all our inputs, then if will pad the short ones by default adding 0 and drop anything that is longer than our maximum length.

```
# Word padding
# Max length
max_length = max(df['clean'].apply(lambda x: len(x.split())))
print('Max length: ', max_length)
# Pad sequences
df['padded'] = df['clean'].apply(lambda x: pad_sequences([x.split()], maxlen=max_length, padding='post', truncating='post')[0])
df.sample(5)
```

	text	clean	padded
1	Ever since the release of the iPhone 11 series, we've been hearing rumors and speculation about the iPhone 12 series. Apple has not yet confirmed any details, but we've put together a roundup of all the rumors we've heard so far. We'll be updating this roundup with more information as we hear it, so be sure to check back often. Here's what we're hoping to see in the iPhone 12 lineup, plus what we think we'll actually get.	ever since release iphone 11 series we've hearing rumor speculation about iphone 12 series apple not yet confirmed any detail but we've put together roundup all rumor we've heard far we'll updating roundup more information we hear it be sure check back often here what we're hoping see iphone 12 lineup plus what we think we'll actually get	[ever, since, release, iphone, 11, series, we've, hearing, rumor, speculation, about, iphone, 12, series, apple, not, yet, confirmed, any, detail, but, we've, put, together, roundup, all, rumor, we've, heard, far, we'll, updating, roundup, more, information, we, hear, it, be, sure, check, back, often, here, what, we're, hoping, see, iphone, 12, lineup, plus, what, we, think, we'll, actually, get, 0, 0, 0