pan-extractive-summary-ex-3-nlp-1

February 8, 2024

0.0.1 Importing Packages

```
[201]: import numpy as np
       import nltk
       from nltk.tokenize import word_tokenize,sent_tokenize
       from sklearn.metrics.pairwise import cosine_similarity
       from heapq import nlargest
       from sklearn.feature_extraction.text import TfidfVectorizer
       import rouge
       from rouge import Rouge
       from sklearn.feature_extraction.text import CountVectorizer
       from gensim.models.word2vec import Word2Vec
       from gensim.models import KeyedVectors
       from gensim.models import FastText
```

0.0.2 Text

```
[2]: text = '''
     Millions go missing at China bank
     Two senior officials at one of China's top commercial banks have reportedly␣
       ↪disappeared after funds
     worth up to $120m (£64m) went missing.
     The pair both worked at Bank of China in the northern city of Harbin, the South␣
       ↪China Morning Post
     reported. The latest scandal at Bank of China will do nothing to reassure␣
       ↪foreign investors that China's
     big four banks are ready for international listings. Government policy sees the␣
       ↪bank listings as vital
     economic reforms. Bank of China is one of two frontrunners in the race to list␣
       ↪overseas. The other is
     China Construction Bank. Both are expected to list abroad during 2005.
     They shared a $45bn state bailout in 2003, to help clean up their balance␣
       ↪sheets in preparation for a
     foreign stock market debut.
     However, a report in the China-published Economic Observer said on Monday that␣
       ↪the two banks may
```

```
have scrapped plans to list in New York because of the cost of meeting␣
 ↪regulatory requirements
imposed since the Enron scandal. Bank of China is the country's biggest foreign␣
 ↪exchange dealer, while
China Construction Bank is the largest deposit holder. China's banking sector␣
 ↪is burdened with at least $190bn of bad debt according to official data,␣
 ↪though most observers believe the true figure is far
higher. Officially, one in five loans is not being repaid. Attempts to␣
 ↪strengthen internal controls and
tighten lending policies have uncovered a succession of scandals involving␣
 ↪embezzlement by bank
officials and loans-for-favours. The most high-profile case involved the␣
 ↪ex-president of Bank of China,
Wang Xuebing, jailed for 12 years in 2003. Although, he committed the offences␣
 ↪whilst running Bank
of China in New York, Mr.Wang was head of China Construction Bank when the␣
 ↪scandal broke. Earlier
this month, a China Construction Bank branch manager was jailed for life in a␣
 ↪separate case.
China's banks used to act as cash offices for state enterprises and did not␣
 ↪require checks on credit
worthiness. The introduction of market reforms has been accompanied by attempts␣
 ↪to modernize the
banking sector, but links between banks and local government remain strong.␣
 ↪Last year, China's
premier, Wen Jiabao, targeted bank lending practices in a series of speeches,␣
 ↪and regulators ordered
all big loans to be scrutinized, in an attempt to cool down irresponsible␣
 ↪lending. China's leaders see
reforming the top four banks as vital to distribute capital to profitable␣
 ↪companies and protect the health
of China's economic boom. But two problems persist. First, inefficient state␣
 ↪enterprises continue to
receive protection from bankruptcy because they employ large numbers of people.␣
 ↪Second, many
questionable loans come not from the big four, but from smaller banks. Another␣
 ↪high-profile financial
firm, China Life, is facing shareholder lawsuits and a probe by the US␣
 ↪Securities and Exchange
Commission following its 2004 New York listing over its failure to disclose␣
 ↪accounting irregularities
at its parent company.
'''
```

```
[11]: ref = '''
The other is China Construction Bank. The latest scandal at Bank of China will↵
  ↪do nothing to reassure
foreign investors that China's big four banks are ready for international↵
  ↪listings. Bank of China is the
country's biggest foreign exchange dealer, while China Construction Bank is the↵
  ↪largest deposit holder.
Bank of China is one of two frontrunners in the race to list overseas.↵
  ↪Although, he committed the
offences whilst running Bank of China in New York, Mr.Wang was head of China↵
  ↪Construction Bank
when the scandal broke. Earlier this month, a China Construction Bank branch↵
  ↪manager was jailed for
life in a separate case. The pair both worked at Bank of China in the northern↵
  ↪city of Harbin, the South
China Morning Post reported. The most high-profile case involved the↵
  ↪ex-president of Bank of China,
Wang Xuebing, jailed for 12 years in 2003.Two senior officials at one of↵
  ↪China's top commercial banks
have reportedly disappeared after funds worth up to $120m (£64m) went missing.↵
  ↪China's banks used
to act as cash offices for state enterprises and did not require checks on↵
  ↪credit worthiness.'''
```

0.0.3 Skeleton Function - Generating Summary

```
[61]: def generate_summary(text,feature_matrix,n):
          sentences = sent_tokenize(text)
          sentence_scores = cosine_similarity(feature_matrix[-1], feature_matrix[:
      ↪-1])[0]
          print(sentence_scores)
          summary_sentences = nlargest(n, range(len(sentence_scores)),␣
      ↪key=sentence_scores.__getitem__)
          summary = ' '.join([sentences[i] for i in sorted(summary_sentences)])
          summary = summary.split('. ')
          formatted_summary = '.\n'.join(summary)
          return formatted_summary
```

0.0.4 Skeleton Function - Evaluation

```
[62]: def calculate_score(ref,generated_summary):
          rouge = Rouge()
          scores = rouge.get_scores(ref,generated_summary)
          score = scores[0]['rouge-1']['f']
          return score
```

0.0.5 Bag of Words

```
[63]: cv = CountVectorizer()
      count_matrix = cv.fit_transform(sent_tokenize(text))
      print(f"Feature Representation Shape: {count_matrix.toarray().shape}")
```

Feature Representation Shape: (24, 272)

```
[64]: bow_summary = generate_summary(text,count_matrix,11)
```

```
[0.1278275  0.16329932 0.14617634 0.04714045 0.1490712  0.18257419
 0.05270463 0.03108349 0.16381203 0.20082136 0.16903085 0.0496904
 0.17888544 0.1490712  0.15569979 0.07698004 0.13679711 0.1754116
 0.12253577 0.21773242 0.         0.03615508 0.0372678 ]
```

```
[65]: bow_summary
```

```
[65]: "The pair both worked at Bank of China in the northern city of Harbin, the South
      China Morning Post\nreported.\nBank of China is one of two frontrunners in the
      race to list overseas.\nThe other is\nChina Construction Bank.\nHowever, a
      report in the China-published Economic Observer said on Monday that the two
      banks may\nhave scrapped plans to list in New York because of the cost of
      meeting regulatory requirements\nimposed since the Enron scandal.\nBank of China
      is the country's biggest foreign exchange dealer, while\nChina Construction Bank
      is the largest deposit holder.\nChina's banking sector is burdened with at least
      $190bn of bad debt according to official data, though most observers believe the
      true figure is far\nhigher.\nAttempts to strengthen internal controls
      and\ntighten lending policies have uncovered a succession of scandals involving
      embezzlement by bank\nofficials and loans-for-favours.\nThe most high-profile
      case involved the ex-president of Bank of China,\nWang Xuebing, jailed for 12
      years in 2003.\nAlthough, he committed the offences whilst running Bank\nof
      China in New York, Mr.Wang was head of China Construction Bank when the scandal
      broke.\nThe introduction of market reforms has been accompanied by attempts to
      modernize the\nbanking sector, but links between banks and local government
      remain strong.\nChina's leaders see\nreforming the top four banks as vital to
      distribute capital to profitable companies and protect the health\nof China's
      economic boom."
```

```
[46]: bow_score = calculate_score(ref,bow_summary)
      print(f"Rouge Score for BOW: {bow_score}")
```

Rouge Score for BOW: 0.5441176421464101

0.0.6 TF-IDF

```python
tfidf = TfidfVectorizer(stop_words='english')
tfidf_matrix = tfidf.fit_transform(sent_tokenize(text))
print(f"Feature Representation Shape: {tfidf_matrix.shape}")
```

Feature Representation Shape: (24, 209)

```python
tf_idf_summary = generate_summary(text,tfidf_matrix,10)
```

```
[0.01808131 0.02302645 0.02416757 0.          0.0180305  0.03742863
 0.          0.          0.07669682 0.07806186 0.00943394 0.
 0.          0.11211433 0.1018688  0.0711365  0.01093595 0.
 0.00837097 0.02138517 0.          0.          0.          ]
```

```python
tf_idf_summary
```

[71]: "\nMillions go missing at China bank\nTwo senior officials at one of China's top commercial banks have reportedly disappeared after funds\nworth up to $120m (£64m) went missing.\nThe pair both worked at Bank of China in the northern city of Harbin, the South China Morning Post\nreported.\nThe latest scandal at Bank of China will do nothing to reassure foreign investors that China's\nbig four banks are ready for international listings.\nThe other is\nChina Construction Bank.\nHowever, a report in the China-published Economic Observer said on Monday that the two banks may\nhave scrapped plans to list in New York because of the cost of meeting regulatory requirements\nimposed since the Enron scandal.\nBank of China is the country's biggest foreign exchange dealer, while\nChina Construction Bank is the largest deposit holder.\nThe most high-profile case involved the ex-president of Bank of China,\nWang Xuebing, jailed for 12 years in 2003.\nAlthough, he committed the offences whilst running Bank\nof China in New York, Mr.Wang was head of China Construction Bank when the scandal broke.\nEarlier\nthis month, a China Construction Bank branch manager was jailed for life in a separate case.\nChina's leaders see\nreforming the top four banks as vital to distribute capital to profitable companies and protect the health\nof China's economic boom."

```python
tf_idf_score = calculate_score(ref,tf_idf_summary)
print(f"Rouge Score for TF-IDF: {tf_idf_score}")
```

Rouge Score for TF-IDF: 0.8031496013243228

0.0.7 CBOW

```python
from nltk.corpus import stopwords
```

```python
new_text = ""
for i in word_tokenize(text):
    if i not in stopwords.words('english'):
```

```
        new_text+=i
        new_text+=" "
```

[133]: `new_text`

[133]: "Millions go missing China bank Two senior officials one China 's top commercial banks reportedly disappeared funds worth $ 120m (£64m) went missing . The pair worked Bank China northern city Harbin , South China Morning Post reported . The latest scandal Bank China nothing reassure foreign investors China's big four banks ready international listings . Government policy sees bank listings vital economic reforms . Bank China one two frontrunners race list overseas . The China Construction Bank . Both expected list abroad 2005 . They shared $ 45bn state bailout 2003 , help clean balance sheets preparation foreign stock market debut . However , report China-published Economic Observer said Monday two banks may scrapped plans list New York cost meeting regulatory requirements imposed since Enron scandal . Bank China country 's biggest foreign exchange dealer , China Construction Bank largest deposit holder . China 's banking sector burdened least $ 190bn bad debt according official data , though observers believe true figure far higher . Officially , one five loans repaid . Attempts strengthen internal controls tighten lending policies uncovered succession scandals involving embezzlement bank officials loans-for-favours . The high-profile case involved ex-president Bank China , Wang Xuebing , jailed 12 years 2003 . Although , committed offences whilst running Bank China New York , Mr.Wang head China Construction Bank scandal broke . Earlier month , China Construction Bank branch manager jailed life separate case . China 's banks used act cash offices state enterprises require checks credit worthiness . The introduction market reforms accompanied attempts modernize banking sector , links banks local government remain strong . Last year , China's premier , Wen Jiabao , targeted bank lending practices series speeches , regulators ordered big loans scrutinized , attempt cool irresponsible lending . China 's leaders see reforming top four banks vital distribute capital profitable companies protect health China 's economic boom . But two problems persist . First , inefficient state enterprises continue receive protection bankruptcy employ large numbers people . Second , many questionable loans come big four , smaller banks . Another high-profile financial firm , China Life , facing shareholder lawsuits probe US Securities Exchange Commission following 2004 New York listing failure disclose accounting irregularities parent company . "

[134]:
```python
cbow = Word2Vec(sent_tokenize(new_text),vector_size=100, window=5, min_count=2,
→sg=0)
vocab = cbow.wv.index_to_key
```

[135]:
```python
def get_mean_vector(model, sentence):
    words = [word for word in sentence if word in vocab]
    if len(words) >= 1:
        return np.mean(model.wv[words], axis=0)
    return np.zeros((100,))
```

```
cbow_array = []
for sentence in sent_tokenize(new_text):
    cbow_array.append(get_mean_vector(cbow, sentence))

cbow_array = np.array(cbow_array)
```

[136]:
```
cbow_array
```

[136]:
```
array([[-0.04307085,  0.10969687,  0.07576666, …, -0.04627705,
         0.01084443,  0.06344573],
       [-0.04334157,  0.10959181,  0.075415  , …, -0.04616618,
         0.0106557 ,  0.06139551],
       [-0.04584682,  0.11412003,  0.07764191, …, -0.04829001,
         0.01143149,  0.06503487],
       …,
       [-0.04625217,  0.11476665,  0.07962143, …, -0.0478877 ,
         0.01082776,  0.06479122],
       [-0.04383902,  0.10899606,  0.07642557, …, -0.04628872,
         0.01152076,  0.0643066 ],
       [-0.04307276,  0.10840796,  0.07429051, …, -0.04578514,
         0.01045212,  0.06130365]], dtype=float32)
```

[137]:
```
cbow_array.shape
```

[137]: `(24, 100)`

[138]:
```
from sklearn.neighbors import NearestNeighbors
model = NearestNeighbors(n_neighbors=10,
                         metric='cosine',
                         algorithm='brute',
                         n_jobs=-1)
model.fit(cbow_array)
```

[138]: `NearestNeighbors(algorithm='brute', metric='cosine', n_jobs=-1, n_neighbors=10)`

[139]:
```
distances, indices = model.kneighbors(cbow_array)
```

[140]:
```
extractive_summary = []

for i in range(len(cbow_array)):
    neighbors_indices = indices[i][1:]  # Exclude the document itself
    summary_embedding = np.mean(cbow_array[neighbors_indices], axis=0)
    extractive_summary.append(summary_embedding)

extractive_summary = np.array(extractive_summary)
```

```python
[141]: top_10_summaries_indices = np.argsort(distances.sum(axis=1))[:10]
       top_10_summaries = extractive_summary[top_10_summaries_indices]
```

```python
[142]: top_10_summaries_indices
```

```
[142]: array([19,  9,  8, 18, 17, 10, 23, 15,  0, 14], dtype=int64)
```

```python
[143]: top_10_sentences = [sent_tokenize(new_text)[index] for index in↵
       ↪top_10_summaries_indices]
```

```python
[144]: top_10_sentences
```

```
[144]: ["China 's leaders see reforming top four banks vital distribute capital
       profitable companies protect health China 's economic boom .",
        "Bank China country 's biggest foreign exchange dealer , China Construction
       Bank largest deposit holder .",
        'However , report China-published Economic Observer said Monday two banks may
       scrapped plans list New York cost meeting regulatory requirements imposed since
       Enron scandal .',
        "Last year , China's premier , Wen Jiabao , targeted bank lending practices
       series speeches , regulators ordered big loans scrutinized , attempt cool
       irresponsible lending .",
        'The introduction market reforms accompanied attempts modernize banking sector
       , links banks local government remain strong .',
        "China 's banking sector burdened least $ 190bn bad debt according official
       data , though observers believe true figure far higher .",
        'Another high-profile financial firm , China Life , facing shareholder lawsuits
       probe US Securities Exchange Commission following 2004 New York listing failure
       disclose accounting irregularities parent company .',
        'Earlier month , China Construction Bank branch manager jailed life separate
       case .',
        "Millions go missing China bank Two senior officials one China 's top
       commercial banks reportedly disappeared funds worth $ 120m ( £64m ) went missing
       .",
        'Although , committed offences whilst running Bank China New York , Mr.Wang
       head China Construction Bank scandal broke .']
```

```python
[145]: sorted(top_10_summaries_indices)
```

```
[145]: [0, 8, 9, 10, 14, 15, 17, 18, 19, 23]
```

```python
[146]: cbow_summary = ' '.join([sent_tokenize(new_text)[i] for i in↵
       ↪sorted(top_10_summaries_indices)])
```

```python
[147]: cbow_summary
```

[147]: "Millions go missing China bank Two senior officials one China 's top commercial banks reportedly disappeared funds worth $ 120m (£64m) went missing . However , report China-published Economic Observer said Monday two banks may scrapped plans list New York cost meeting regulatory requirements imposed since Enron scandal . Bank China country 's biggest foreign exchange dealer , China Construction Bank largest deposit holder . China 's banking sector burdened least $ 190bn bad debt according official data , though observers believe true figure far higher . Although , committed offences whilst running Bank China New York , Mr.Wang head China Construction Bank scandal broke . Earlier month , China Construction Bank branch manager jailed life separate case . The introduction market reforms accompanied attempts modernize banking sector , links banks local government remain strong . Last year , China's premier , Wen Jiabao , targeted bank lending practices series speeches , regulators ordered big loans scrutinized , attempt cool irresponsible lending . China 's leaders see reforming top four banks vital distribute capital profitable companies protect health China 's economic boom . Another high-profile financial firm , China Life , facing shareholder lawsuits probe US Securities Exchange Commission following 2004 New York listing failure disclose accounting irregularities parent company ."

0.0.8 Skipgram

```python
[152]: sg = Word2Vec(sent_tokenize(new_text), vector_size=100, window=5, min_count=2,
       ↪sg=1)
       vocab = sg.wv.index_to_key

       def get_mean_vector(model, sentence):
           words = [word for word in sentence if word in vocab]
           if len(words) >= 1:
               return np.mean(model.wv[words], axis=0)
           return np.zeros((100,))

       sg_array = []
       for sentence in sent_tokenize(new_text):
           sg_array.append(get_mean_vector(sg, sentence))

       sg_array = np.array(sg_array)
```

```python
[154]: sg_array.shape
```

```
[154]: (24, 100)
```

```python
[155]: from sklearn.neighbors import NearestNeighbors
       model = NearestNeighbors(n_neighbors=10,
                                metric='cosine',
                                algorithm='brute',
                                n_jobs=-1)
```

```
model.fit(sg_array)
```

[155]: NearestNeighbors(algorithm='brute', metric='cosine', n_jobs=-1, n_neighbors=10)

[156]: `distances, indices = model.kneighbors(sg_array)`

[157]:
```python
extractive_summary = []

for i in range(len(sg_array)):
    neighbors_indices = indices[i][1:]   # Exclude the document itself
    summary_embedding = np.mean(sg_array[neighbors_indices], axis=0)
    extractive_summary.append(summary_embedding)

extractive_summary = np.array(extractive_summary)
```

[158]:
```python
top_10_summaries_indices = np.argsort(distances.sum(axis=1))[:10]
top_10_summaries = extractive_summary[top_10_summaries_indices]
```

[159]:
```python
top_10_sentences = [sent_tokenize(new_text)[index] for index in
 ↪top_10_summaries_indices]
```

[160]: `top_10_sentences`

[160]: ["China 's leaders see reforming top four banks vital distribute capital
profitable companies protect health China 's economic boom .",
 "Bank China country 's biggest foreign exchange dealer , China Construction
Bank largest deposit holder .",
 'However , report China-published Economic Observer said Monday two banks may
scrapped plans list New York cost meeting regulatory requirements imposed since
Enron scandal .',
 "Last year , China's premier , Wen Jiabao , targeted bank lending practices
series speeches , regulators ordered big loans scrutinized , attempt cool
irresponsible lending .",
 'The introduction market reforms accompanied attempts modernize banking sector
, links banks local government remain strong .',
 "China 's banking sector burdened least $ 190bn bad debt according official
data , though observers believe true figure far higher .",
 'Another high-profile financial firm , China Life , facing shareholder lawsuits
probe US Securities Exchange Commission following 2004 New York listing failure
disclose accounting irregularities parent company .',
 'Earlier month , China Construction Bank branch manager jailed life separate
case .',
 "Millions go missing China bank Two senior officials one China 's top
commercial banks reportedly disappeared funds worth $ 120m (£64m) went missing
.",
 'Attempts strengthen internal controls tighten lending policies uncovered
succession scandals involving embezzlement bank officials loans-for-favours .']

```
[161]: top_10_summaries_indices
```

```
[161]: array([19,  9,  8, 18, 17, 10, 23, 15,  0, 12], dtype=int64)
```

```
[165]: sg_summary = ' '.join([sent_tokenize(new_text)[i] for i in
        ↪sorted(top_10_summaries_indices)])
```

```
[166]: sg_summary
```

[166]: "Millions go missing China bank Two senior officials one China 's top commercial
banks reportedly disappeared funds worth $ 120m (£64m) went missing . However
, report China-published Economic Observer said Monday two banks may scrapped
plans list New York cost meeting regulatory requirements imposed since Enron
scandal . Bank China country 's biggest foreign exchange dealer , China
Construction Bank largest deposit holder . China 's banking sector burdened
least $ 190bn bad debt according official data , though observers believe true
figure far higher . Attempts strengthen internal controls tighten lending
policies uncovered succession scandals involving embezzlement bank officials
loans-for-favours . Earlier month , China Construction Bank branch manager
jailed life separate case . The introduction market reforms accompanied attempts
modernize banking sector , links banks local government remain strong . Last
year , China's premier , Wen Jiabao , targeted bank lending practices series
speeches , regulators ordered big loans scrutinized , attempt cool irresponsible
lending . China 's leaders see reforming top four banks vital distribute capital
profitable companies protect health China 's economic boom . Another high-
profile financial firm , China Life , facing shareholder lawsuits probe US
Securities Exchange Commission following 2004 New York listing failure disclose
accounting irregularities parent company ."

0.0.9   Glove

```
[173]: file_name = "word2vec-glove.6B.300d.txt"
       model = KeyedVectors.load_word2vec_format(file_name, binary=False)
```

```
[176]: sentences = sent_tokenize(new_text)
       sentences
```

[176]: ["Millions go missing China bank Two senior officials one China 's top
commercial banks reportedly disappeared funds worth $ 120m (£64m) went missing
.",
 'The pair worked Bank China northern city Harbin , South China Morning Post
reported .',
 "The latest scandal Bank China nothing reassure foreign investors China's big
four banks ready international listings .",
 'Government policy sees bank listings vital economic reforms .',
 'Bank China one two frontrunners race list overseas .',
 'The China Construction Bank .',

```
 'Both expected list abroad 2005 .',
 'They shared $ 45bn state bailout 2003 , help clean balance sheets preparation
foreign stock market debut .',
 'However , report China-published Economic Observer said Monday two banks may
scrapped plans list New York cost meeting regulatory requirements imposed since
Enron scandal .',
 "Bank China country 's biggest foreign exchange dealer , China Construction
Bank largest deposit holder .",
 "China 's banking sector burdened least $ 190bn bad debt according official
data , though observers believe true figure far higher .",
 'Officially , one five loans repaid .',
 'Attempts strengthen internal controls tighten lending policies uncovered
succession scandals involving embezzlement bank officials loans-for-favours .',
 'The high-profile case involved ex-president Bank China , Wang Xuebing , jailed
12 years 2003 .',
 'Although , committed offences whilst running Bank China New York , Mr.Wang
head China Construction Bank scandal broke .',
 'Earlier month , China Construction Bank branch manager jailed life separate
case .',
 "China 's banks used act cash offices state enterprises require checks credit
worthiness .",
 'The introduction market reforms accompanied attempts modernize banking sector
, links banks local government remain strong .',
 "Last year , China's premier , Wen Jiabao , targeted bank lending practices
series speeches , regulators ordered big loans scrutinized , attempt cool
irresponsible lending .",
 "China 's leaders see reforming top four banks vital distribute capital
profitable companies protect health China 's economic boom .",
 'But two problems persist .',
 'First , inefficient state enterprises continue receive protection bankruptcy
employ large numbers people .',
 'Second , many questionable loans come big four , smaller banks .',
 'Another high-profile financial firm , China Life , facing shareholder lawsuits
probe US Securities Exchange Commission following 2004 New York listing failure
disclose accounting irregularities parent company .']
```

```python
def glove_sentence_embedding(sentence, model):
    words = sentence.split()
    valid_words = [word for word in words if word in model]

    if len(valid_words) > 0:
        embedding = np.mean([model[word] for word in valid_words], axis=0)
    else:
        embedding = np.zeros(model.vector_size)

    return embedding
```

```
[180]: glove_array = [sentence_embedding(sentence, model) for sentence in sentences]
       glove_array = np.array(glove_array)
```

```
[181]: glove_array
```

```
[181]: array([[ 0.22317618,  0.17406385,  0.35028833, …, -0.16210365,
                 0.5665448 , -0.1056913 ],
               [-0.08302357, -0.01501614,  0.20060015, …, -0.02811715,
                 0.5627057 , -0.11961043],
               [ 0.02092777,  0.14262792,  0.37171093, …, -0.04933453,
                 0.3878954 ,  0.32923287],
               …,
               [-0.06292181,  0.24843508,  0.05861692, …, -0.29185206,
                 0.52695715, -0.16893741],
               [-0.00934309,  0.3416618 ,  0.2475369 , …, -0.17657301,
                 0.6884654 ,  0.04916891],
               [ 0.09589174, -0.10853332,  0.06520274, …, -0.13419141,
                 0.5598758 , -0.13724864]], dtype=float32)
```

```
[182]: glove_array.shape
```

```
[182]: (24, 100)
```

```
[183]: from sklearn.neighbors import NearestNeighbors
       model = NearestNeighbors(n_neighbors=10,
                                metric='cosine',
                                algorithm='brute',
                                n_jobs=-1)
       model.fit(glove_array)
```

```
[183]: NearestNeighbors(algorithm='brute', metric='cosine', n_jobs=-1, n_neighbors=10)
```

```
[184]: distances, indices = model.kneighbors(glove_array)
```

```
[185]: extractive_summary = []

       for i in range(len(glove_array)):
           neighbors_indices = indices[i][1:]  # Exclude the document itself
           summary_embedding = np.mean(glove_array[neighbors_indices], axis=0)
           extractive_summary.append(summary_embedding)

       extractive_summary = np.array(extractive_summary)
```

```
[186]: top_10_summaries_indices = np.argsort(distances.sum(axis=1))[:10]
       top_10_summaries = extractive_summary[top_10_summaries_indices]
```

```
[187]: top_10_summaries_indices
```

```
[187]: array([18, 19, 22,  7, 10,  2,  8, 17,  0, 21], dtype=int64)
```

```python
[188]: top_10_sentences = [sent_tokenize(new_text)[index] for index in
        ↪top_10_summaries_indices]
```

```python
[189]: top_10_sentences
```

```
[189]: ["Last year , China's premier , Wen Jiabao , targeted bank lending practices
       series speeches , regulators ordered big loans scrutinized , attempt cool
       irresponsible lending .",
        "China 's leaders see reforming top four banks vital distribute capital
       profitable companies protect health China 's economic boom .",
        'Second , many questionable loans come big four , smaller banks .',
        'They shared $ 45bn state bailout 2003 , help clean balance sheets preparation
       foreign stock market debut .',
        "China 's banking sector burdened least $ 190bn bad debt according official
       data , though observers believe true figure far higher .",
        "The latest scandal Bank China nothing reassure foreign investors China's big
       four banks ready international listings .",
        'However , report China-published Economic Observer said Monday two banks may
       scrapped plans list New York cost meeting regulatory requirements imposed since
       Enron scandal .',
        'The introduction market reforms accompanied attempts modernize banking sector
       , links banks local government remain strong .',
        "Millions go missing China bank Two senior officials one China 's top
       commercial banks reportedly disappeared funds worth $ 120m ( £64m ) went missing
       .",
        'First , inefficient state enterprises continue receive protection bankruptcy
       employ large numbers people .']
```

```python
[190]: glove_summary = ' '.join([sent_tokenize(new_text)[i] for i in
        ↪sorted(top_10_summaries_indices)])
```

```python
[191]: glove_summary
```

```
[191]: "Millions go missing China bank Two senior officials one China 's top commercial
       banks reportedly disappeared funds worth $ 120m ( £64m ) went missing . The
       latest scandal Bank China nothing reassure foreign investors China's big four
       banks ready international listings . They shared $ 45bn state bailout 2003 ,
       help clean balance sheets preparation foreign stock market debut . However ,
       report China-published Economic Observer said Monday two banks may scrapped
       plans list New York cost meeting regulatory requirements imposed since Enron
       scandal . China 's banking sector burdened least $ 190bn bad debt according
       official data , though observers believe true figure far higher . The
       introduction market reforms accompanied attempts modernize banking sector ,
       links banks local government remain strong . Last year , China's premier , Wen
       Jiabao , targeted bank lending practices series speeches , regulators ordered
```

big loans scrutinized , attempt cool irresponsible lending . China 's leaders
see reforming top four banks vital distribute capital profitable companies
protect health China 's economic boom . First , inefficient state enterprises
continue receive protection bankruptcy employ large numbers people . Second ,
many questionable loans come big four , smaller banks ."

0.0.10 FastText

```
[202]: tokenized_sentences = [sentence.split() for sentence in sentences]
       model = FastText(sentences=tokenized_sentences, vector_size=300, window=5,␣
         ↪min_count=1, workers=-1, sg=1)
```

```
[203]: def fast_text_sentence_embedding(sentence, model):
           words = sentence.split()
           valid_words = [word for word in words if word in model.wv]
           if len(valid_words) > 0:
               embedding = np.mean([model.wv[word] for word in valid_words], axis=0)
           else:
               embedding = np.zeros(model.vector_size)

           return embedding
```

```
[204]: fast_embedding = [fast_text_sentence_embedding(sentence, model) for sentence in␣
         ↪sentences]
       fast_embedding_array = np.array(fast_embedding)
```

```
[205]: print(fast_embedding_array)
```

```
[[-2.2793592e-04  1.6232443e-04  6.7126726e-05 …  6.4166503e-05
   1.6630090e-04 -1.2996412e-04]
 [-1.9683743e-05  4.9250887e-04  1.5267938e-04 … -2.9366274e-04
  -6.9838672e-05 -1.1452228e-04]
 [ 5.1756491e-05  1.1708896e-04  1.7875625e-04 …  2.3562390e-05
   6.5808352e-05 -2.0481732e-05]
 …
 [-2.1770365e-04  3.6490077e-04  1.0585137e-04 … -2.1933668e-04
  -1.3737976e-04  1.3439913e-04]
 [-1.5962969e-04  6.7737239e-04  3.9854640e-05 … -6.0914393e-04
   3.5686880e-05  8.1965693e-05]
 [-1.2875089e-04  2.8979307e-04 -1.6815853e-05 … -1.7780946e-04
  -1.6246687e-04 -2.5712408e-05]]
```

```
[206]: fast_embedding_array.shape
```

```
[206]: (24, 300)
```

```
[207]:  from sklearn.neighbors import NearestNeighbors
        model = NearestNeighbors(n_neighbors=10,
                                 metric='cosine',
                                 algorithm='brute',
                                 n_jobs=-1)
        model.fit(fast_embedding_array)
```

[207]: NearestNeighbors(algorithm='brute', metric='cosine', n_jobs=-1, n_neighbors=10)

```
[208]:  distances, indices = model.kneighbors(fast_embedding_array)
```

```
[209]:  extractive_summary = []

        for i in range(len(fast_embedding_array)):
            neighbors_indices = indices[i][1:]  # Exclude the document itself
            summary_embedding = np.mean(fast_embedding_array[neighbors_indices], axis=0)
            extractive_summary.append(summary_embedding)

        extractive_summary = np.array(extractive_summary)
```

```
[210]:  top_10_summaries_indices = np.argsort(distances.sum(axis=1))[:10]
        top_10_summaries = extractive_summary[top_10_summaries_indices]
```

```
[211]:  top_10_summaries_indices
```

[211]: array([14, 13, 18, 22, 23, 11, 15, 21, 9, 1], dtype=int64)

```
[212]:  top_10_sentences = [sent_tokenize(new_text)[index] for index in␣
        ↪top_10_summaries_indices]
```

```
[213]:  top_10_sentences
```

[213]: ['Although , committed offences whilst running Bank China New York , Mr.Wang
        head China Construction Bank scandal broke .',
         'The high-profile case involved ex-president Bank China , Wang Xuebing , jailed
        12 years 2003 .',
         "Last year , China's premier , Wen Jiabao , targeted bank lending practices
        series speeches , regulators ordered big loans scrutinized , attempt cool
        irresponsible lending .",
         'Second , many questionable loans come big four , smaller banks .',
         'Another high-profile financial firm , China Life , facing shareholder lawsuits
        probe US Securities Exchange Commission following 2004 New York listing failure
        disclose accounting irregularities parent company .',
         'Officially , one five loans repaid .',
         'Earlier month , China Construction Bank branch manager jailed life separate
        case .',
         'First , inefficient state enterprises continue receive protection bankruptcy

```

```
      employ large numbers people .',
       "Bank China country 's biggest foreign exchange dealer , China Construction
      Bank largest deposit holder .",
       'The pair worked Bank China northern city Harbin , South China Morning Post
      reported .']
```

[214]: 
```
fast_text_summary = ' '.join([sent_tokenize(new_text)[i] for i in
 ↪sorted(top_10_summaries_indices)])
```

[215]: 
```
fast_text_summary
```

[215]: "The pair worked Bank China northern city Harbin , South China Morning Post
reported . Bank China country 's biggest foreign exchange dealer , China
Construction Bank largest deposit holder . Officially , one five loans repaid .
The high-profile case involved ex-president Bank China , Wang Xuebing , jailed
12 years 2003 . Although , committed offences whilst running Bank China New York
, Mr.Wang head China Construction Bank scandal broke . Earlier month , China
Construction Bank branch manager jailed life separate case . Last year , China's
premier , Wen Jiabao , targeted bank lending practices series speeches ,
regulators ordered big loans scrutinized , attempt cool irresponsible lending .
First , inefficient state enterprises continue receive protection bankruptcy
employ large numbers people . Second , many questionable loans come big four ,
smaller banks . Another high-profile financial firm , China Life , facing
shareholder lawsuits probe US Securities Exchange Commission following 2004 New
York listing failure disclose accounting irregularities parent company ."