

Natural Language Processing Lab

Exercise 2:

Application: Extractive Text Summarization

Task:

- Explain the pipeline for summarizing an input document by identifying informative sentences.
- Perform cleaning and preprocessing of text.
- Generate representations using:
 - Bag of Words
 - TF-IDF
 - Continuous Bag of Words
 - Skip gram
 - Word2Vec.
 - GloVe
 - FastText
- Generate representations for each sentence using the above word embedding techniques.
- Identify the distance between each sentence. Evaluate if semantically similar sentences lie closer to each other.
- Explain the impact of embedding techniques in identifying the distance between sentences.
- Which Embedding technique generates contextual representations? Justify.

Input 1:

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.

Expected Summary:

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.