CS-529

PROJECT REPORT on Fake News Detection using Graph and Summarization Techniques

PHASE -1:

***Introduction**:*

Nowadays, fake news is widely spreading in various media, and this fake information is causing serious damage in many areas. Therefore, there is an increasing need to accurately detect fake news to prevent such damage. This project implements a novel method that uses graph and summarization techniques for fake news detection. This method represents the relationship of all sentences in a graph structure to accurately understand the context information of the document. Accordingly, the relationship between sentences in the graph is calculated as a score through the attention mechanism. Then, the summarization technique is used to reflect the sentence subject information in the graph update process. Our proposed method shows better performance than Karimi's and BERT based models by approximately 10.34%p and 3.72%p, respectively.

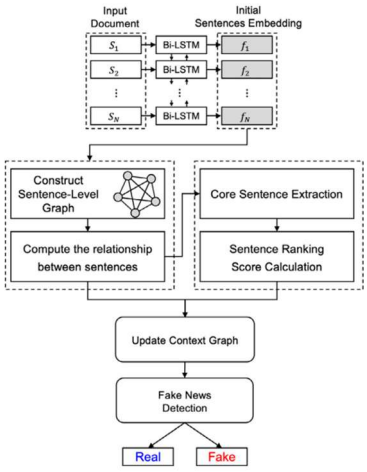
Background and Related work:

Internal information-based approaches rely on the text con-tent to detect the truth of news articles, which are usually comprised of long texts. To learn internal information, various methods have been studied detected fake news by extracting topic information using Term Frequency - Inverse Document Frequency (TF-IDF). External information-based approaches model the process of spreading contents, which typically consists of long text, the user characteristics that propagate the contents, or the response and comments of users who read the contents

3 Challenges in research problem:

The first one is graph construction using the attention mechanism for representing the relationship between all sentences using a graph. The second one is core sentence extraction for ranking sentences with subject information using the summarization technique. The third one is update context graph for calculating more accurate contextual information. The fourth one is fake news detection for discriminating fake news from several documents.

Proposed model:



1.Graph construction using the attention mechanism:

To model the relationship between sentences and their contextual information, to construct a context graph G = (F, E). The graph G is composed of a node (i.e., F = {f1, f2, . . . fN}) and the edge (i.e., E = {e1,2, eij, . . . eN−1,N}). Each node is represented by sentence embedding H, and the edge between the ith node and the jth node is represented by ei,j, and its weight wi,j represents the relational strength of the ith sentence and jth sentence.

xi = ReLU(W1hi + bias1)

xj = ReLU(W2hj + bias2)

wi,j = xiUxj (where W1,W2 ∈ Rm×dim, and x ∈ Rm)

2. Core sentence extraction:

Core_Sent represents the extracted sentence that contains the most subject information of the subject.

$$Core\_Sent = argmax_{1 \le i \le N} \sum_{j=1, j \ne i}^{N} w_{i,j}$$

Based on the cosine similarity values between the core sentence and all other sentences, all sentences in the document are divided into a subject relevant sentence set and an irrelevant sentence set. The subject relevance score of each word is based on the frequency of appearance in the subject relevant and irrelevant sentence sets, and then the word relevance score RSd is calculated by

$$RS_d(word) = \log \frac{p_d \times (1-q_d)}{(1-p_d) \times q_d} = \log \frac{(r_d+0.5)(S_d-s_d+0.5)}{(R_d-r_d+0.5)(s_d+0.5)}$$

where pd and qd are the probabilities that a word appears in the subject relevant and irrelevant sentences set in a document d. Respectively, Rd and Sd are the number of subject relevant sentences and irrelevant sentences in document d, respectively. rd and sd are the number of subject relevant and irrelevant sentences that include the word in a document d, and 0.5 is a naove smoothing factor used to avoid a zero-denominator or log-zero. In this study, the top 30% of sentences with subject similarity were selected as the relevant sentence set and the others as the irrelevant sentence set. Afterward, the sentence score is calculated with the sum of the relevance score of words included in the sentence

$$Score(s_i) = \sum_{word \in s_i} RS_d(word)$$

3.Update context graph:

We use the sentence ranking to calculate the sentence score and then use it to update the weight of the edge in the contextgraph. The weight of the edge represents the reflection rate of the subject information and context information. Sentence_Rank is the order of sentences by the sum of the

attention scores between each sentence and other sentences; the first ranked sentence has the highest sum of attention scores

$$Rank\_Score(s_i) = 1 - \frac{Sentence\_Rank(s_i) - 1}{N}$$

$$w'_{i,j} = w_{i,j} * Rank\_Score(s_i)$$
$$(for\ i, j = 1, 2, \ldots, N\ and\ i \neq j)$$

Subsequently, we can construct subject contextualized sentence embeddings using the updated context graph. The subject contextualized sentence embedding (i.e., h i) is created by the weighted sum of the initial sentence embedding of the current sentence (i.e., hi) and those of the other adjacent sentences (i.e., hj)

$$h'_i = h_i + \sum_{j=1, j\neq i}^{N} w'_{i,j} * h_j$$

4.Fake news detection:

A document embedding doc with the average of all nodes (the contextualized sentence embedding) in the context graph. The document embedding doc is was then fed into a multi-layer feedforward neural network to predict its label for the binary classification, i.e., real or fake, as a vector yˆ.

$$doc = \frac{1}{N}\sum_{i=1}^{N} h'_i$$

$$\hat{y} = Softmax(sigmoid(W_3 doc + bias_3))$$

The cross-entropy loss function was used to optimize our neural network.

loss = −(y log(pf ake) + (1 − y)log(preal)) [where y is the golden standard data to in the document and pf ake and preal are the probabilities of a document being fake or real of yˆ, respectively]

PHASE- 2:

Objective:

A large amount of information through various media. However, some of this information are fake with a different purpose other than telling the truth, and people fail to validate such information and identify it as true. This fake information confuses people and causes social, economic, and national damages. *Therefore, the need to detect fake information is of paramount importance to prevent such damages*.

Datasets which are used:

1. Nela
2. FNC
3. ISOT
4. ClickBait

Results:

Model	Dataset	Accuracy	F-measure		
			macro	micro	weighted
Graph+Summarization	FNC	71.71	0.42	0.71	0.60
	Click Bait	75.25	0.42	0.75	0.64
	Nela	49.38	0.47	0.49	0.47
	ISOT	59.77	0.59	0.59	0.59
BERT	FNC	71.12	0.61	0.61	0.60
	Click Bait	96.19	0.96	0.96	0.96
	Nela	51.28	0.51	0.51	0.51
	ISOT	96.22	0.96	0.96	0.96
Karimi	FNC	72.21	0.42	0.50	0.61
	Click Bait	78.52	0.60	0.65	0.74
	Nela	51.68	0.49	0.49	0.49
	ISOT	72.00	0.71	0.71	0.71

Observation: By seeing the experimental values it is clearly evident that accuracy of nela is poorer than any other dataset and almost all the models give same accuracy on nela. When it comes to clickbait bert and karimi models outperform graph+summarization model. For FNC all models classify with almost same accuracy. BERT performs best for ISOT.