CS 529 - Topics and Tools in Social Media Data Mining

PROJECT TITLE

TI-CNN: Convolutional Neural Networks for Fake News Detection

Submitted by –

Sherin Abraham (214101052)

Sagar Kumar (214101045)

Bhanu Prakash Singh (214101013)

Phase - I

Introduction: -

Fake news dissemination is very common in social networks . Due to the extensive social connections among users, fake news on certain topics, e.g., politics, celebrities and product promotions, can propagate and lead to a large number of nodes reporting the same (incorrect) observations rapidly in online social networks. According to the statistical results reported by the researchers in Stanford University, 72.3% of the fake news actually originates from the official news media and online social networks. The potential reasons are provided as follows. Firstly, the emergence of social media greatly lowered the barriers to enter in the media industry. Various online blogs, "we media", and virtual communities are becoming more and more popular in recent years, in which everyone can post news articles online. Secondly, the large number of social media users provide a breeding ground for fake news. Fake news involving conspiracy and pitfalls can always attract our attention. People like to share this kind of information to their friends. Thirdly, the 'trust and confidence' in the mass media greatly dropped these years. More and more people tend to trust the fake news by browsing the headlines only without reading the content at all.

***Related research*: -**

Deception detection has been a hot topic in the past few years. Deception information includes scientific fraud, fake news, false tweets etc. Fake news detection is a subtopic in this area. Researchers solve the deception detection problem from two aspects:

1) linguistic approach.

2) network approach.

Linguistic approaches:-

Mihalcea and Strapparvva 2009 started to use natural language processing techniques to solve this problem. Bing Liu et.al. analysed fake reviews on Amazon these years based on the sentiment analysis, lexical, content similarity, style similarity and semantic inconsistency to identify the fake reviews. Hai et al. proposed a semi-supervised learning method to detect deceptive text on crowdsourced datasets in 2016.

The methods based on word analysis are not enough to identify deception. Many researchers focus on some deeper language structures, such as the syntax tree. In this case, the sentences are represented as a parse tree to describe syntax structure, for example noun and verb phrases, which are in turn rewritten by their syntactic constituent parts .

Network-based approaches:-

Another way to identify the deception is to analyse the network structure and behaviours, which are important complementary features. As the development of knowledge graph, it will be very helpful to check facts based on the relationship among entities.

Ciampaglia et al. [6] proposed a new concept of 'network effect' variables to derive the probabilities of news. The methods based on the knowledge graph analysis can achieve 61% to 95% accuracy. Another promising research direction is exploiting the social network behaviour to identify the deception.

Neural Network based approaches:-

In the natural language processing (NLP) area, deep learning models are used to train a model that can represent words as vectors. Then researchers propose many deep learning models based on the word vectors and summarization etc.

Researchers solve the deception detection problem from two aspects:

1) Linguistic approach :

- Mihalcea and Strapparvva 2009 started to use natural language processing techniques to solve this problem.
- The methods based on word analysis is not enough to identify deception.
- Many researchers focused on some deeper language structures, such as the syntax tree.

2) Network approach:

- Another way to identify the deception is to analyse the network structure and behaviours, which are important complementary features.
- Graphs help identify the relationship among entities.
- Ciampaglia et al. proposed a new concept of 'network effect' variables to derive the probabilities of news.
- The methods based on the knowledge graph analysis can achieve 61% to 95% accuracy.

Neural Network based approaches:

- In the natural language processing (NLP) area, deep learning models are used to train a model that can represent words as vectors. Then researchers propose many deep learning models based on the word vectors and summarization etc.

Challenges:

Fake news identification from online social media is extremely challenging due to various reasons. Such as :

It is difficult to collect the fake news data, and it is also hard to label fake news manually.

News that appears on Facebook and Twitter news feeds belongs to private data. To this context so far, few large-scale fake news detection public dataset really exists.

Some news datasets available online involve a small number of the instances only, which are not sufficient to train a generalised model for application.

Generally fake news is written by humans. Most liars tend to use their language strategically to avoid being caught. In spite of the attempt to control what they are saying, language leakage occurs with certain verbal aspects that are hard to monitor such as frequencies and patterns of pronoun, conjunction, and negative emotion word usage.

Proposed models in the paper:-

In this paper, we propose a TI-CNN model to consider both text and image information in fake news detection. Beyond the explicit features extracted from the data, as the development of representative learning, convolutional neural networks are employed to learn the latent features which cannot be captured by the explicit features. Finally, we utilise TI-CNN to combine the explicit and latent features of text and image information into a unified feature space, and then use the learned features to identify the fake news. Hence, the contributions of this paper are summarised as follows:

We collect a high quality dataset and take in-depth analysis on the text from multiple perspectives. Image information has proved to be an effective feature in identifying fake news.

Image information has proved to be effective features in identifying fake news.

A unified model is proposed to analyse the text and image information using the convolutional neural networks.

Problem definition:-

Given a set of m news articles containing the text and image information, we can represent the data as a set of text image tuples A = {(AT i , AI i)} m i . In the fake news detection problem, we want to predict whether the news articles in A are fake news or not. We can represent the label set as Y = {[1, 0], [0, 1]}, where [1, 0] denotes real news while [0, 1] represents fake news. Meanwhile, based on the news articles, e.g., (AT i , AI i) $\in$ A, a set of features (including both explicit and latent features to be introduced later in Model Section) can be extracted from both the text and image information available in the article, which can be represented as XT i and XI i respectively. The objective of the fake news detection problem is to build a model f : {XT i , XI i } m i $\in$ X $\rightarrow$ Y to infer the potential labels of the news articles in A.

In this section, we introduce the architecture of the TI-CNN model in detail. Besides the explicit features, we innovatively utilise two parallel CNNs to extract latent features from both textual and visual information. And then explicit and latent features are projected into the same feature space to form new representations of texts and images. At last, we propose to fuse textual and visual representations together for fake news detection.

The overall model contains two major branches, i.e., text branch and image branch. For each branch, taking textual or visual data as inputs, explicit and latent features are extracted for final predictions. To demonstrate the theory of constructing the TI-CNN, we introduce the model by answering the following questions:

1) How to extract the latent features from text?

2) How to combine the explicit and latent features?

3) How to deal with the text and image features together?

4) How to design the model with fewer parameters?

5) How to train and accelerate the training process?

A. Text Branch

For the text branch, we utilise two types of features: textual explicit features XT e and textual latent features XT l. The textual explicit features are derived from the statistics of the news text as we mentioned in the data analysis part, such as the length of the news, the number of sentences, question marks, exclamations and capital letters, etc. The statistics of a single news can be organised as a vector with fixed size. Then the vector is transformed by a fully connected layer to form a textual explicit feature.

B. Image Branch

Similar to the text branch, we use two types of features: visual explicit features XIe and visual latent features XIl. in order to obtain the visual explicit features, We firstly extract the resolution of an image and the number of faces in the image to form a feature vector. And then, we transform the vector into our visual explicit feature with a fully connected layer.

…………………………………….**End of Phase-I**………………………………………

Phase 2 :-

Objective:-

Our main objective is to classify a given news article given along with its image into real or fake using some pretrained model such as vgg-19, resnet-152 (for image) and bert, lstm (for text).

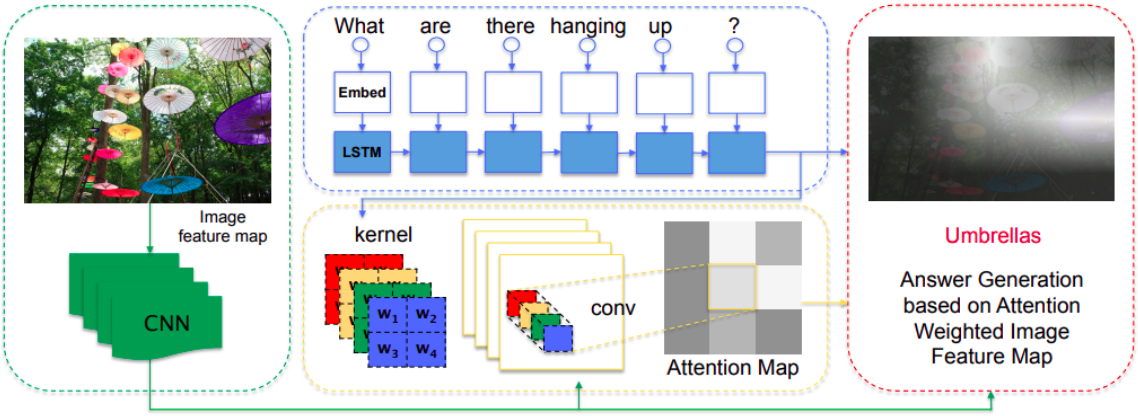

Figure 2. The framework of ABC-CNN. The green box denotes the image feature extraction part using CNN; the blue box is the question understanding part using LSTM; the yellow box illustrates the attention extraction part with configurable convolution; the red box is the answer generation part using multi-class classification based on attention weighted image feature maps. The orange letters are corresponding variables explained in Eq. (1) - (6).

Intuitions behind the objective for phase–II of project**: -**

ABC-CNN is composed of four components:

(1) the image feature extraction part

(2) the question understanding part

(3) the attention extraction part

(4) answer generation

Image Feature Extraction:

The visual information in each image is represented as an N×N×D image feature map. The feature map is generated by dividing an image into an N × N grid, and extracting a D-dimensional feature vector f in each cell of the grid. The VGG-19 deep convolutional neural network extracts a D-dimensional feature vector for each window. The D-dimensional feature vector for each cell is the average of all the 10 D-dimensional feature vectors. The final N ×N ×D image feature map is the concatenation of N × N × D dimensional feature vectors.

Question Understanding:

Question understanding is crucial for visual question answering. The semantic meaning of questions not only provides the most important clue for answer generation. Here we have given the title of the news article as input to our question understanding part.

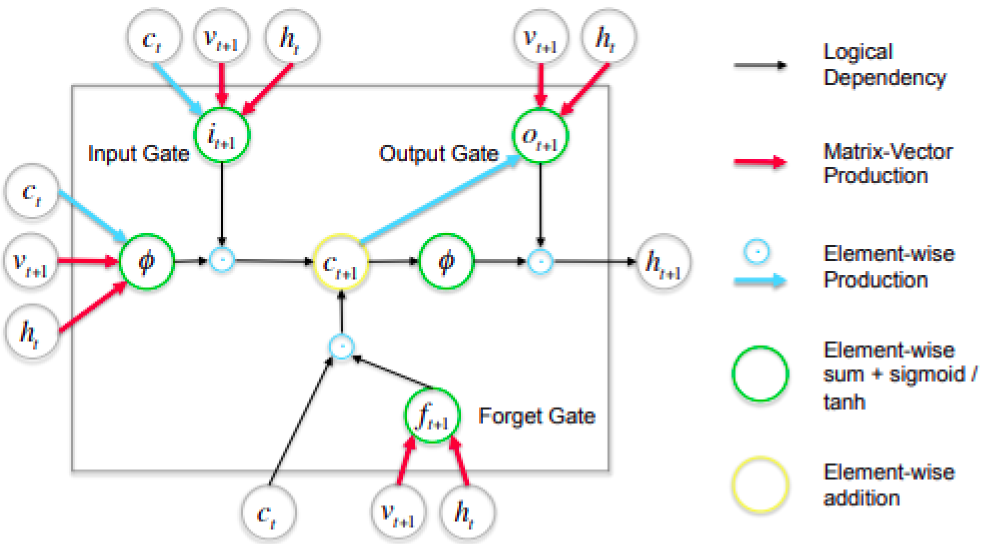

Figure 3. The structure of LSTM [8] for query processing

$$
\begin{aligned}
\boldsymbol{i}_t &= \sigma(\mathbf{W}_{vi}\boldsymbol{v}_t + \mathbf{W}_{hi}\boldsymbol{h}_{t-1} + \boldsymbol{b}_i) \\
\boldsymbol{f}_t &= \sigma(\mathbf{W}_{vf}\boldsymbol{v}_t + \mathbf{W}_{hf}\boldsymbol{h}_{t-1} + \boldsymbol{b}_f) \\
\boldsymbol{o}_t &= \sigma(\mathbf{W}_{vo}\boldsymbol{v}_t + \mathbf{W}_{ho}\boldsymbol{h}_{t-1} + \boldsymbol{b}_o) \\
\boldsymbol{g}_t &= \phi(\mathbf{W}_{vg}\boldsymbol{v}_t + \mathbf{W}_{hg}\boldsymbol{h}_{t-1} + \boldsymbol{b}_g) \\
\boldsymbol{c}_t &= \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \boldsymbol{g}_t \\
\boldsymbol{h}_t &= \boldsymbol{o}_t \odot \phi(\boldsymbol{c}_t)
\end{aligned} \tag{3}
$$

Attention Extraction:-

A QAM, m, capturing the image regions queried by the question, is generated for each image-question pair using a configurable convolutional neural network. The configurable convolution operation can be thought of as searching spatial image feature maps for specific visual features that correspond to the question's intent. The dense question embedding s encodes the semantic object information asked in the question.

$$\boldsymbol{k} = \sigma(\mathbf{W}_{sk}\boldsymbol{s} + \boldsymbol{b}_k), \quad \sigma(x) = \frac{1}{1+e^{-x}} \qquad (1)$$

The projection transforms the semantic information into the corresponding visual information as a CCK, which has the same number of channels as the image feature map I. The QAM is generated by convolving the CCK k with the image feature map I, and applying softmax normalisation:

$$\boldsymbol{m}_{ij} = P(\mathrm{ATT}_{ij}|\mathbf{I}, \boldsymbol{s}) = \frac{e^{\boldsymbol{z}_{ij}}}{\sum_i \sum_j e^{\boldsymbol{z}_{ij}}}, \quad \boldsymbol{z} = \boldsymbol{k} * \mathbf{I} \quad (2)$$

where mij is the element of the QAM at position (i, j), and the symbol * represents the convolution operation. The QAM characterises the attention distribution across the image feature map. The convolution is padded so that the QAM m has the same size as the image feature map I. The QAM corresponds to the regions asked by the question. For example, the question "What is the color of the umbrella?" can generate an attention map focusing on umbrella image regions because the CCK is configured to find umbrella visual features.

Answer Generation:-

The answer generation part is a multi-class classifier based on the original image feature map, the dense question embedding, and the attention weighted feature map. We employ the attention map to spatially weight the image feature map I. The weighted image feature map focuses on the objects asked in the question. The spatial weighting is achieved by the element-wise production between each channel of the image feature map and the attention map.

$$\mathbf{I}'_i = \mathbf{I}_i \odot \boldsymbol{m} \tag{4}$$

where dot represents element-wise production. I 0 i and Ii represent the i-th channel of attention weighted feature map I 0 and original image feature map I, respectively. The attention weighted feature map lowers the weights of the regions that are irrelevant to the meaning of the question. To avoid overfitting, we apply an 1×1 convolution on the attention weighted feature map to reduce the number of channels, resulting in a reduced feature map Ir. The question's semantic information s, the image feature map I and the reduced feature map Ir are then fused by a nonlinear projection.

$$\boldsymbol{h} = g(\mathbf{W}_{ih}\mathbf{I} + \mathbf{W}_{rh}\mathbf{I}_r + \mathbf{W}_{sh}\boldsymbol{s} + \boldsymbol{b}_h) \tag{5}$$

where h denotes the final projected feature, and g(.) is the element-wise scaled hyperbolic tangent function: g(x) = 1.7159 · tanh(2 3 x) . This function leads the gradients into the most non-linear range of value and enables a higher training speed.

A multi-class classifier with softmax activation, which is trained on the final projected features, predicts the index of an answer word specified in an answer dictionary. The answer generated by ABC-CNN is the word with the maximum probability.

$$a^* = \arg\max_{a \in \mathcal{V}_a} \boldsymbol{p}_a \quad \text{s.t.} \quad \boldsymbol{p}_a = g(\mathbf{W}_{ha}\boldsymbol{h} + \boldsymbol{b}_a) \tag{6}$$

Supporting experimental setup:

In experiments, we first choose the resolution of both the image feature map and the attention map to be 3×3, which is called "ATT" model. Each image cell generates a 4096- dimensional image feature vector using a pre-trained VGG network, and we extend each feature vector with the HSV histogram of the cell, resulting in a 4276-dimensional image feature vector for each cell. The image feature vectors from all the image cells constitute an image feature map with dimension 4276×3×3. To avoid overfitting, we reduce the dimension of the feature map to 256×3×3 with an 1×1 convolution. The dimension of the dense question embedding is 256.

Produce result over data provided :

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*

Future work:

Generative Adversarial Networks (GAN) can be used in image to generate captions. It provides a novel way to evaluate the relevance between image and news text.