TI-CNN :
CNN in Fake News Detection

Objective

To identify fake news by using both text and image. Explicit and latent features of both text and image data is used to train the model

Problem Statement

- Given *m* news articles containing the text and image information, represent the data as a set of text image tuples $A = \{(A^T_i, A^I_i)\}^m_i$.
- Output the label set as Y = {[1, 0], [0, 1]}, where ,
  - [1, 0] denotes real news
  - [0, 1] represents the fake news.
- The objective of the fake news detection problem is to build a model f : $\{X^T_i, X^I_i\}^m_i \in X \longrightarrow Y$ to infer the potential labels of the news articles in A.
- X^T_i, X^I_i - are the features extracted from text and image data or news article i

Background/ Related work

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 focus on some deeper language structures, such as the syntax tree.

2) Network approach:

- Another way to identify the deception is to analyze the network structure and behaviors, which are important complementary features.
- Graphs help identify the relationship among entities.
- Ciampaglia et al. proposed a new concept ‘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.

TI-CNN Architecture

- ➢ Two parallel CNN to extract latent features from both textual and visual information
- ➢ Combine the latent and explicit features of both textual and visual data individually and the fuse the result to form the final model

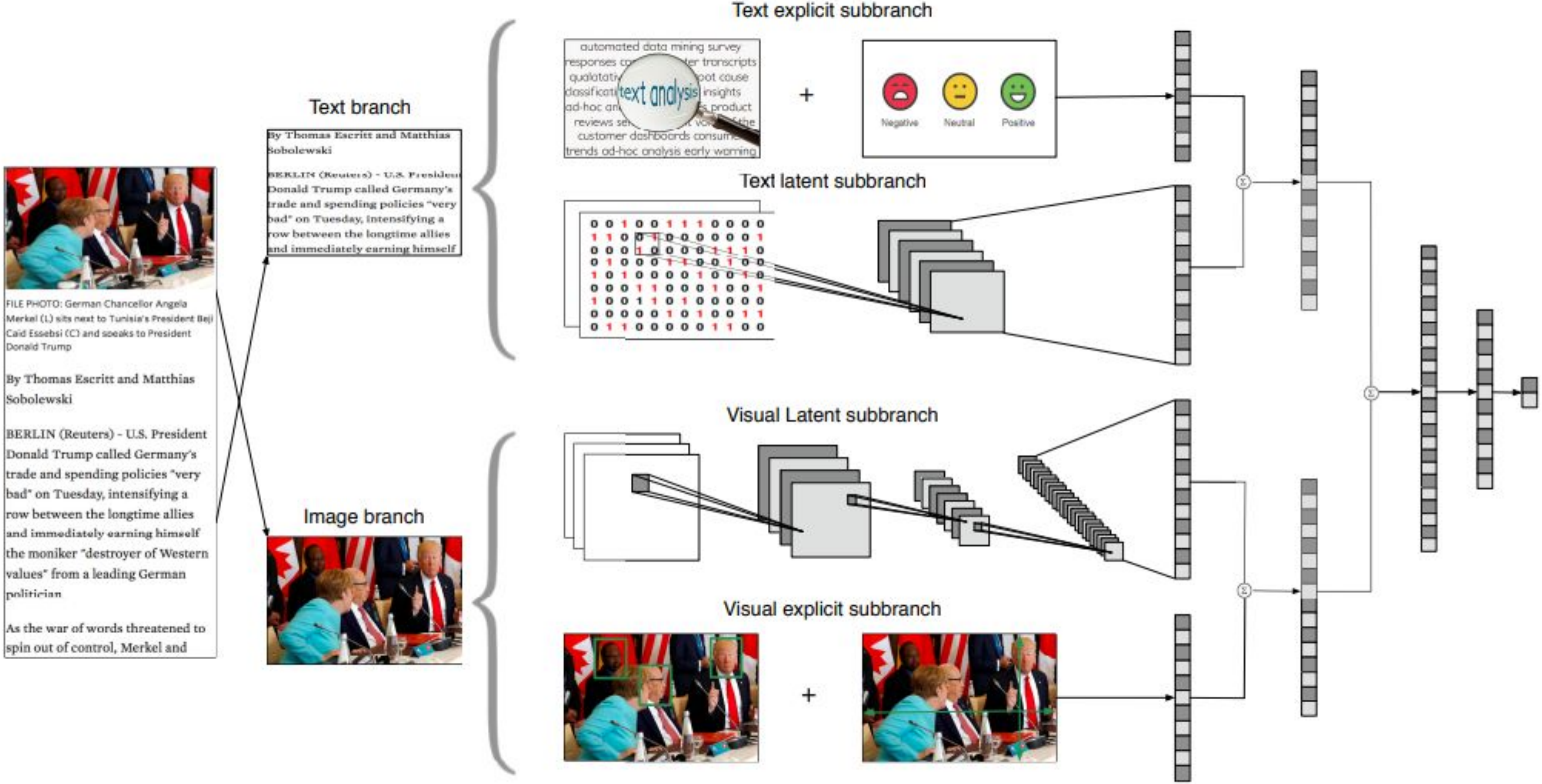
Two branches : Text branch and Image branch

Text branch:

- Explicit features from the statistics of the news text
- Latent features using CNN

Image branch:

- Explicit features from the resolution of the image and number of faces in the image to form a feature vector → fully connected layer
- Latent features:
  - Convolution layer
  - Max Pooling
  - Rectified Linear Neuron
  - Regularization



FILE PHOTO: German Chancellor Angela Merkel (L) sits next to Tunisia's President Beji Caid Essebsi (C) and speaks to President Donald Trump

By Thomas Escritt and Matthias Sobolewski

BERLIN (Reuters) - U.S. President Donald Trump called Germany's trade and spending policies "very bad" on Tuesday, intensifying a row between the longtime allies and immediately earning himself the moniker "destroyer of Western values" from a leading German politician.

As the war of words threatened to spin out of control, Merkel and

Text branch

By Thomas Escritt and Matthias Sobolewski

BERLIN (Reuters) - U.S. President Donald Trump called Germany's trade and spending policies "very bad" on Tuesday, intensifying a row between the longtime allies and immediately earning himself

Image branch

Text explicit subbranch

Text latent subbranch

```
0 0 1 0 0 1 1 1 0 0 0 0
1 1 0 0 1 0 0 0 0 0 0 1
0 0 0 1 0 0 0 0 1 1 1 0
0 1 0 0 0 1 1 0 0 1 0 0
1 0 1 0 0 0 0 1 0 0 1 0
0 0 0 1 1 0 0 0 1 0 0 1
1 0 0 1 1 0 1 0 0 0 0 0
0 0 0 0 0 1 0 1 0 0 1 1
0 1 1 0 0 0 0 0 1 1 0 0
```

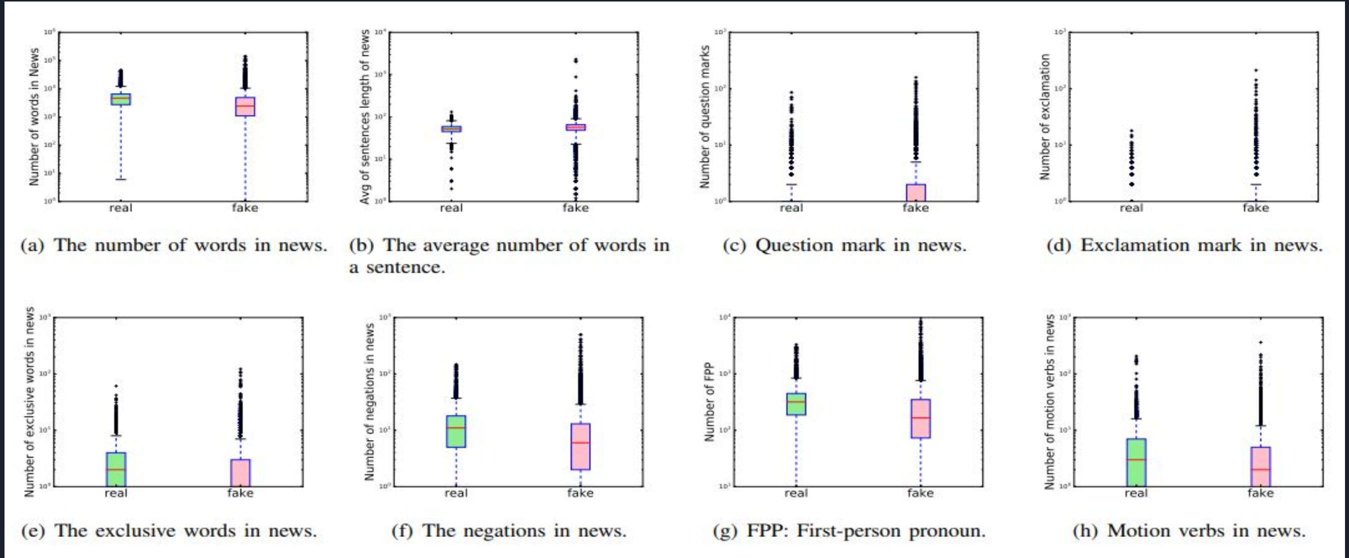
Visual Latent subbranch

Visual explicit subbranch

automated data mining survey
responses co ter transcripts
qualitati oot cause
classificati text analysis insights
ad-hoc an s product
reviews se vo the
customer dashboards consum
trends ad-hoc analysis early warning
+
Negative
Neutral
Positive
+
Σ
Σ
Σ

Data Analysis

- ➢ Dataset
- ➢ Text Analysis
  - ○ Fake news have no titles
  - ○ They have more capital letters
  - ○ Real news contains detailed description



(a) The number of words in news.

(b) The average number of words in a sentence.

(c) Question mark in news.

(d) Exclamation mark in news.

(e) The exclusive words in news.

(f) The negations in news.

(g) FPP: First-person pronoun.

(h) Motion verbs in news.

Dataset

The dataset contains multiple information, such as the title, text, image, author and website.

To reveal the intrinsic differences between real and fake news, we solely use the title, text and image information.

Text Analysis

We can identify fake news by their title such as

- It contains more capital letters.
- It can be title-less news.
- Computational Linguistic like Number of words and sentences, and Question mark,exclamation and capital letters.
- psychology perspective
- Sentiment Analysis

Image Analysis

We can identify fake news by their images :

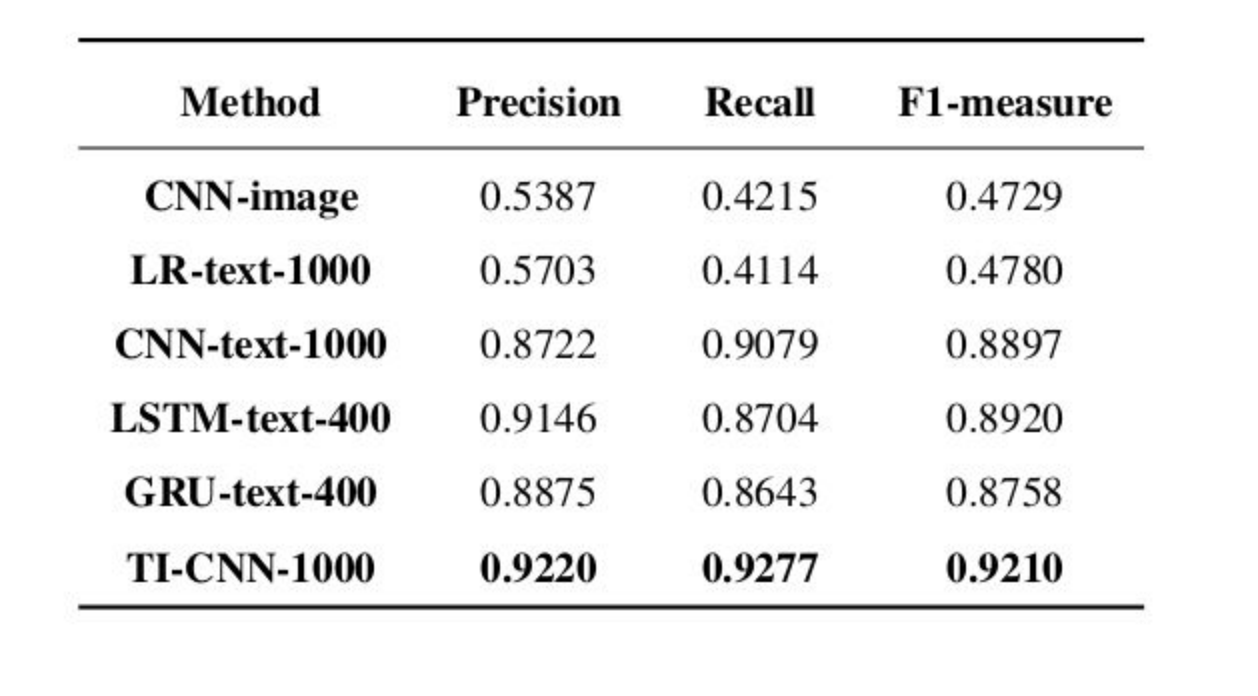
- ➢ It contains less faces than real news like There are 0.366 faces on average in real news, while the number is 0.299 in fake .
- ➢ It contains more irreverent images such as animals and scenes.
- ➢ In addition, real news has a better resolution image than fake like real news has 457 × 277 pixels on average, while the fake news has a resolution of 355 × 228.

Methodology and Experimental setup

- **Experimental Setup**
- **Experimental Results**
- **Sensitivity Analysis**
  - **word embedding dimensions**
  - **batch size**
  - **hidden layer dimension**
  - **Dropout probability and filter size**

Experimental Setup : We use 80% of the data for training, 10% of the data for validation and 10% of the data for testing. All the experiments are run at least 10 times separately.

Experimental Results :



Method	Precision	Recall	F1-measure
CNN-image	0.5387	0.4215	0.4729
LR-text-1000	0.5703	0.4114	0.4780
CNN-text-1000	0.8722	0.9079	0.8897
LSTM-text-400	0.9146	0.8704	0.8920
GRU-text-400	0.8875	0.8643	0.8758
TI-CNN-1000	**0.9220**	**0.9277**	**0.9210**

Conclusion and Future Work

- CNN can see the entire output at once and can be trained much faster than LSTM and RNN
- GAN can be used to find the relevance between the image and the news text