Fake News Spreader Identification in Twitter using Ensemble Modeling

8th Author Profiling Task

PAN Workshop – CLEF 2020

Ahmad Hashemi

Mohammad Reza Zarei

Mohammad Reza Moosavi

Mohammad Taheri

Department of Computer Science and Engineering, Shiraz University, Shiraz, Iran

Why Study Fake News?

- Negative consequences of fake news propagation
  - Political Aspects
  - Economic Aspects
  - Health Related Aspects

Profiling Fake News Spreaders

- Hypothesis: Users who do not spread fake news have a set of different characteristics compared to users who tend to share fake news.

↓

- Identifying fake news spreaders as a first step towards fake news detection

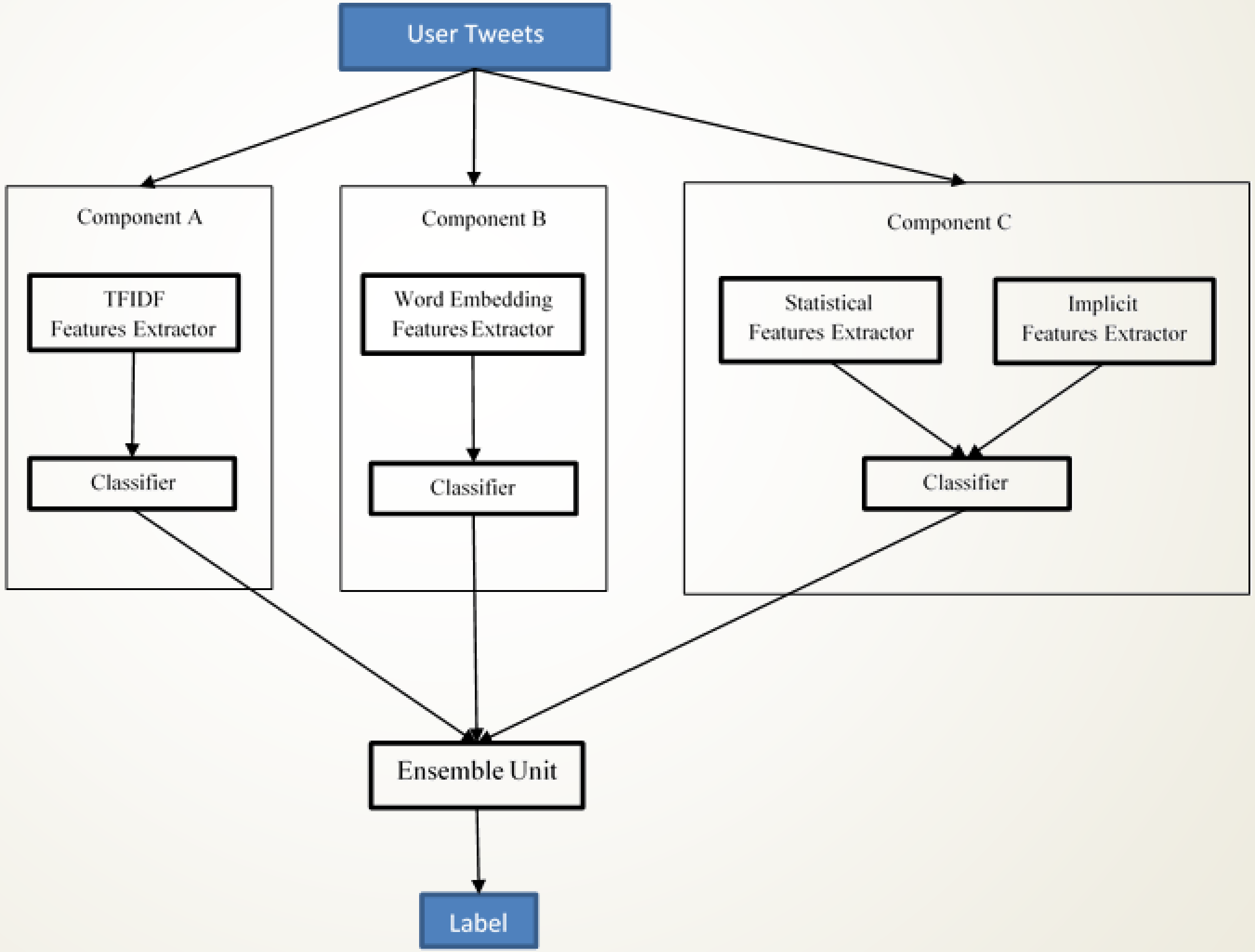
The PAN-AP-20 Provided Corpus

- Number of authors in the competition dataset:

Language	Training	Test	Total
English	300	200	500
Spanish	300	200	500

- For each author, their last 100 tweets have been retrieved

Overview of The Proposed Model



Statistical features

- Fraction of retweets (tweets starting with "RT")
- Average number of mentions per tweet
- Average number of URLs per tweet
- Average number of hashtags per tweet
- Average tweet length

Implicit Features

- Age (English dataset)
- Gender (English dataset)
- Emotional Signals
  - English dataset: anger, anticipation, disgust, fear, joy, sadness, surprise, trust
  - Spanish dataset: joy, anger, fear, repulsion, surprise, sadness
- Personality (English dataset)
  - Agreeableness, conscientiousness, extraversion, neuroticism, openness

Word Embeddings

- Preproccessing
  - Omitting retweet tags, hashtags, URLs and user tag
  - TweetTokenizer module from the NLTK package
- English dataset: pretrained on blogs, news and comments
- Spanish dataset: pretrained on news and media contents

Term Frequency – Inverse Document Frequency (TF-IDF)

- Preproccessing
  - Eliminating punctuations, numbers and stop words
  - Stemming
  - Omit-ting retweet tags, hashtags, URLs and user tag
  - TweetTokenizer module from the NLTK package

Ensembling the Models

Use soft classifiers to obtain the confidence of each model $c_i(u)$

$$c_{out}(u) = \alpha c_1(u) + \beta c_2(u) + \gamma c_3(u)$$

$\alpha + \beta + \gamma = 1$
$c_1(u)$: confidence of the classifier for TFIDF features
$c_2(\mathrm{u})$: confidence of the classifier for Word Embeddings features
$c_3(\mathrm{u})$: confidence of the classifier for implicit+statistical features

The label of the user u is determined as:

$$y(u) = \begin{cases} 0 & if\ c_{out}(u) \leq 0.5 \\ 1 & if\ c_{out}(u) > 0.5 \end{cases}$$

Model Selection

- Accuracy scores of 10-fold cross-validation

Feature groups	Dataset	SVM	Random Forest	Logistic Regression
Statistical + Implicit	English	57.6	**69**	49.6
TF-IDF	English	68.3	**70.3**	68.3
Embedding	English	67.6	**71.3**	67.6
Statistical + Implicit	Spanish	72.6	**73**	56
TF-IDF	Spanish	**82**	80	81.6
Embedding	Spanish	74	**76.3**	76

Ensembling the Models

- Determined weight parameters for merging the individual classifiers

Language	TF-IDF (α)	Embeddings(β)	Statistical+Implicit (γ)
English	0.15	0.45	0.4
Spanish	0.65	0.1	0.25

Local Evaluation

- 10 fold cross validation scores obtained on different components

Features	Accuracy (en)	Accuracy (es)
TF-IDF	70.3	82
Embedding	71.3	76.3
Statistical + Explicit	69	73
Ensembled model (final model)	**74.6**	**82.9**

Final Results

- Accuracy scores obtained on the local evaluation and the official test set

Language	Cross-validation	Official test set
English	74.6	69.5
Spanish	82.9	78.5
Average	78.75	74.0

Future Work

- Extracting more Implicit features and analyzing their discrimination
- Proposing a learning scheme for the ensemble unit
- Using the fake news spreader identification results for fake news detection

Thank You

For Your Attention!

s.ahmad.hmi@gmail.com

mr.zarei@cse.shirazu.ac.ir