Text and Image Synergy with Feature Cross Technique for Gender Identification

Notebook for PAN at CLEF 2018

Takumi Takahashi, Takuji Tahara, Koki Nagatani, Yasuhide Miura, Tomoki Taniguchi, and Tomoko Ohkuma

Fuji Xerox Co., Ltd.
{takahashi.takumi, tahara.takuji, nagatani.koki, yasuhide.miura, taniguchi.tomoki, ohkuma.tomoko}@fujixerox.co.jp

Abstract This paper describes a neural network model for the author profiling task of PAN@CLEF 2018. Traditional machine learning models have shown superior performances for the author profiling task in past PAN series. However, these models often require careful feature-engineering to improve their performance. On the other hand, neural network approaches have recently shown advanced performances in both natural language processing (NLP) and computer vision (CV) tasks. We tackle the author profiling task using neural networks for texts and images. In order to leverage the synergy of the texts and images, we propose Text Image Fusion Neural Network (TIFNN), which considers their interaction. In an in-house experiment, TIFNN achieved accuracies of 84-90% for different languages when used for gender identification.

1 Introduction

Author profiling technologies that extract author profile traits from social media can be applied to some applications, e.g., advertisement, recommendation, and marketing. PAN 2018: Author Profiling Task [13] is identifying the user's gender from tweets that are contained texts and images in three languages (English, Spanish, and Arabic).

In PAN 2017 Author Profiling Task, various approaches based on a deep neural network (DNN) were presented [6,7,9,16,18]. However, such approaches could not outperform traditional machine learning models that were carefully modeled, such as support vector machine. In contrast, neural network approaches have shown superior performances on various NLP tasks, e.g., machine translation, summarization, and information retrieval. In addition, DNN approaches have shown advanced performances in various CV tasks.

Because PAN 2018 Author Profiling Task includes both texts and images, using both texts and images in a neural network will improve the performances. Therefore, we tackle this task using both texts and images in a DNN-based approach.

In order to leverage the synergy of the texts and images, we propose Text Image Fusion Neural Network (TIFNN), which considers their interaction. This paper makes the following contributions.

1. We propose an effective fusion strategy for a neural network to utilize texts and images for gender identification.
2. We show that TIFNN has drastically improved accuracies (3-8pt) compared with both a text-based neural network and an image-based neural network.

In the following section of this paper, we first explain the related work in Section 2. Our neural network model is described in Section 3. The details of the experiments used to confirm the model's performances are described in Section 4. Finally, we conclude the paper and outline future work in Section 5.

2 Related Work

PAN Author Profiling Task was to identify both age and gender from social media text in the past PAN series before 2017 edition [12,15]. In the last year, the task included language variety identification instead of age identification [14]. In PAN 2017 Author Profiling Task, various models that used not only traditional machine learning but also deep neural networks were presented.

Basile et al. [1] used linear support vector machine (SVM) with character 3- to 5-grams and word 1- to 2-grams features and showed that it outperforms other approaches. Martinc et al. [8] explored many approaches (e.g. linear SVM, logistic regression, random forest, XGBoost, and voting classifier combining these models) with various parameters for this task. They finally tested logistic regression because it showed the best performance. Tellez et al. [20] used a generic framework for text classification, as called MicroTC. As shown in these researches, the approaches of traditional machine learning that were carefully designed showed the superior performances in this task.

On the other hand, the approaches based on deep neural networks were also presented [6,7,9,16,18]. Miura et al. [9] used both bi-directional GRU with an attention mechanism to capture the word representations and convolutional neural network (CNN) to capture the character representations. Sierra et al. [18] applied CNN that has a set of convolutional filters of different sizes to capture n-gram features. Although the approaches using deep neural networks are strong model for many NLP tasks, the above approaches could not outperform traditional machine learning approaches in this task.

In author profiling tasks outside of PAN, researches utilizing images or multimodality also exist. The research of [17] utilized images to identify the gender of users and the object of images with a multi-task bilinear model. In addition, the research of [21] presented a state-of-the-art model that utilized both texts and images to predict users' traits such as gender, age, political orientation, and location.

As overviewed in this section, the approaches using traditional machine learning showed the superior performances in past PAN series. Although the approaches based on deep neural networks utilizing only text could not outperform traditional machine learning approaches, the researches of [17,21] indicated that utilizing images is effective in the prediction of author profile traits. Because using images is possible in PAN 2018 Author Profiling Task, utilizing both texts and images would be effective for gender identification.

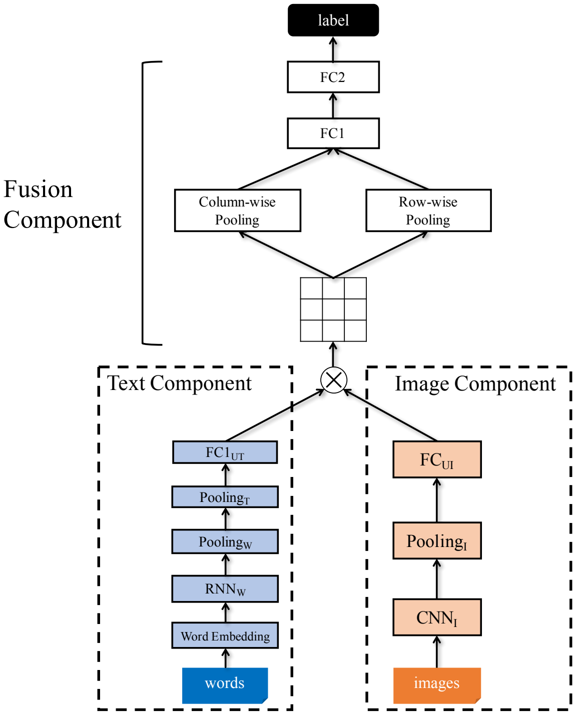

Figure 1. Overview of proposed model. FC denotes a fully connected layer and ⊗ presents the direct-product operation.

3 Model

3.1 Proposed Model

Figure 1 illustrates the proposed model. The model is constructed of a text component, an image component, and a fusion component. The proposed model processes texts and images in their respective components. The fusion component computes the relationship between the texts and images using direct-product, column-wise pooling, and row-wise pooling. Finally, the combination feature of the texts and images is fed to two fully connected layers.

In the following, we first describe each component of the model in Section 3.2 and 3.3. The details of the fusion component are also described in Section 3.4.

3.2 Text Component

This section describes the text component of the model, which is the "Text Component" division in Figure 1. The component is implemented based on the previous models[9]. Figure 2 provides an overview of the text component. This component is constructed of word embedding, recurrent neural network (RNN), pooling, and fully connected (FC)

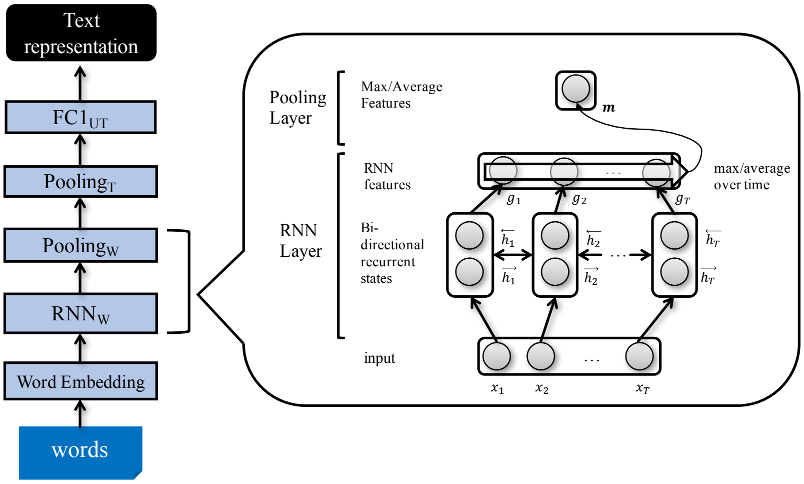

Figure 2. Overview of text component with detailed description of $\mathrm{RNN}_{\mathrm{W}}$ and $\mathrm{Pooling}_{\mathrm{W}}$.

layers. For the RNN, we used Gated Recurrent Unit (GRU) [4] with a bi-directional setting.

First, the input words are embedded to k_w dimensional word embeddings with embedding matrix $\boldsymbol{E}_w$ to obtain $\boldsymbol{x}$ with $\boldsymbol{x}_t \in \mathbb{R}^{k_w}$. $\boldsymbol{x}$ are then fed to $\mathrm{RNN}_{\mathrm{W}}$ with the following transition functions:

$$\boldsymbol{z}_t = \sigma\left(\boldsymbol{W}_z \boldsymbol{x}_t + \boldsymbol{U}_z \boldsymbol{h}_{t-1} + \boldsymbol{b}_z\right) \tag{1}$$

$$\boldsymbol{r}_t = \sigma\left(\boldsymbol{W}_r \boldsymbol{x}_t + \boldsymbol{U}_r \boldsymbol{h}_{t-1} + \boldsymbol{b}_r\right) \tag{2}$$

$$\tilde{\boldsymbol{h}}_t = \tanh\left(\boldsymbol{W}_h \boldsymbol{x}_t + \boldsymbol{U}_h\left(\boldsymbol{r}_t \odot \boldsymbol{h}_{t-1}\right) + \boldsymbol{b}_h\right) \tag{3}$$

$$\boldsymbol{h}_t = \left(\mathbf{1} - \boldsymbol{z}_t\right) \odot \boldsymbol{h}_{t-1} + \boldsymbol{z}_t \odot \tilde{\boldsymbol{h}}_t \tag{4}$$

where $\boldsymbol{z}_t$ is an update gate, $\boldsymbol{r}_t$ is a reset gate, $\tilde{\boldsymbol{h}}_t$ is a candidate state, $\boldsymbol{h}_t$ is a state, $\boldsymbol{W}_z$, $\boldsymbol{W}_r$, $\boldsymbol{W}_h$, $\boldsymbol{U}_z$, $\boldsymbol{U}_r$, $\boldsymbol{U}_h$ are weight matrices, $\boldsymbol{b}_z$, $\boldsymbol{b}_r$, $\boldsymbol{b}_h$ are bias vectors, σ is a logistic sigmoid function, and $\odot$ is an element-wise multiplication operator. The output vectors $\overrightarrow{\boldsymbol{h}}$ and $\overleftarrow{\boldsymbol{h}}$ are concatenated to obtain $\boldsymbol{g}$ as $\boldsymbol{g}_t = [\overrightarrow{\boldsymbol{h}}_t, \overleftarrow{\boldsymbol{h}}_t]$ and are then fed to $\mathrm{Pooling}_{\mathrm{W}}$. In $\mathrm{Pooling}_{\mathrm{W}}$, $\boldsymbol{g}$ are processed to obtain i-th tweet feature $\boldsymbol{m}_i^t$ with max pooling or average pooling over time and are fed to $\mathrm{Pooling}_{\mathrm{T}}$. $\boldsymbol{m}_i^t$ are processed to obtain the j-th user feature $\boldsymbol{m}_j^u$, as well as $\mathrm{Pooling}_{\mathrm{W}}$. Finally, $\boldsymbol{m}^u$ of the user representations are fed to $\mathrm{FC1}_{\mathrm{UT}}$.

3.3 Image Component

This section describes the image component of the model, which is the "Image Component" division in Figure 1. Figure 3 provides an overview of the image component.

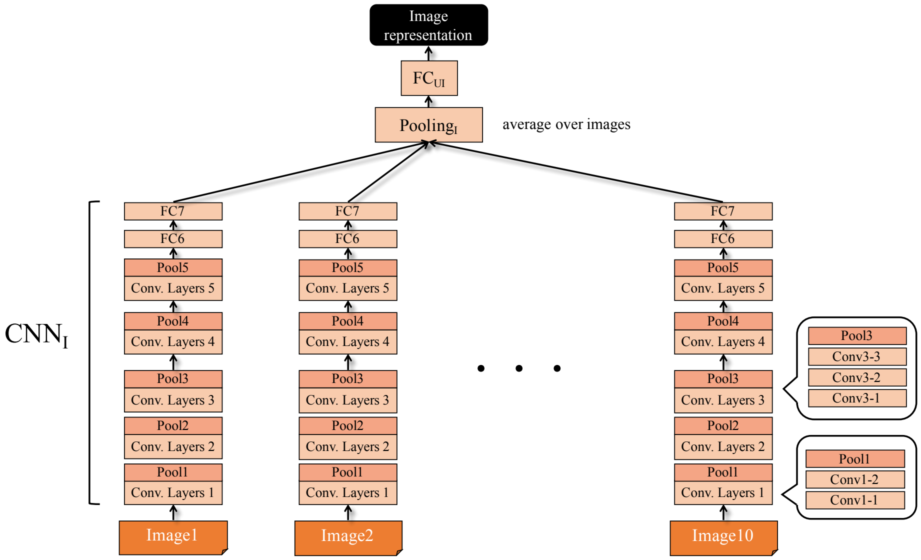

Figure 3. Overview of image component.

This component is constructed of a convolutional neural network architecture ($\mathrm{CNN_I}$), pooling ($\mathrm{Pooling_I}$), and a fully connected layer ($\mathrm{FC_{UI}}$).

It takes the following three steps to utilize multiple posted images:

step 1 The feature representation of each image is extracted using a pre-trained CNN architecture ($\mathrm{CNN_I}$).
step 2 The extracted features are fused ($\mathrm{Pooling_I}$).
step 3 The fused feature is processed with a fully connected layer ($\mathrm{FC_{UI}}$).

$\mathrm{CNN_I}$ in Figure 1 represents the layers from $\mathrm{Conv.Layers1}$ to FC7 in Figure 3. This architecture is implemented based on VGG16 [19]. $\mathrm{CNN_I}$ utilizes the layers from $\mathrm{Conv.Layers1}$ to FC7 to extract each image feature.

$\mathrm{Pooling_{FI}}$ fuses the features extracted from images. The images posted by a single author on social media can be regarded as a kind of time series. However, we cannot know the ground truth of the images' order in time steps and the interval of time between posted images. Therefore, we simply use the average or max operation over image features as $\mathrm{Pooling_I}$.

3.4 Fusion Component

The fusion component is expected to complementarily capture the relationship between the texts and images. User representation $\boldsymbol{r}_{txt} \in \mathbb{R}^M$ is obtained via the text component and $\boldsymbol{r}_{img} \in \mathbb{R}^L$ is obtained via the image component. The relationship between the texts and images is represented as a matrix $\boldsymbol{G} \in \mathbb{R}^{M \times L}$ with the following equation:

$$\boldsymbol{G} = \boldsymbol{r}_{txt} \otimes \boldsymbol{r}_{img} \quad (5)$$

where $\otimes$ is a direct-product operation. We apply column-wise and row-wise max-poolings over $\boldsymbol{G}$ to generate $\boldsymbol{g}_{txt} \in \mathbb{R}^M$ and $\boldsymbol{g}_{img} \in \mathbb{R}^L$, respectively. Formally, the j-th elements of the vector $\boldsymbol{g}_{txt}$ and the j-th elements of the vector $\boldsymbol{g}_{img}$ are computed in the following operation:

$$[g_{txt}]_j = \max_{1 \leq l \leq L} [G_{j,l}] \tag{6}$$

$$[g_{img}]_j = \max_{1 \leq m \leq M} [G_{m,j}] \tag{7}$$

We can interpret the j-th element of the vector $\boldsymbol{g}_{txt}$ as an importance degree for the j-th text feature with regard to image features. Finally, the vectors $\boldsymbol{g}_{txt}$ and $\boldsymbol{g}_{img}$ are concatenated to obtain $\boldsymbol{g}_{comb}$ as $\boldsymbol{g}_{comb} = [\boldsymbol{g}_{txt}, \boldsymbol{g}_{img}]$ and passed to FC1.

4 Experiment

4.1 Data

This section describes two datasets: PAN@CLEF 2018 Author Profiling Training Corpus and streaming tweets. PAN@CLEF 2018 Author Profiling Training Corpus was utilized to train the proposed model and comparison models. Streaming tweets were utilized to pre-train a word embedding matrix $\boldsymbol{E}_w$.

PAN@CLEF 2018 Author Profiling Training Corpus The first dataset we used to train the proposed model was the official PAN@CLEF 2018 Author Profiling Training Corpus. This dataset is constructed of users' tweets in three languages: English, Spanish, and Arabic. There are $3,000$ English language users, $3,000$ Spanish language users, and $1,500$ Arabic language users, with a gender ratio of 1:1. We used random sampling to divide this dataset into $train_8$, dev_1, and $test_1$, with a ratio of 8:1:1, while maintaining the gender ratio of 1:1.

Streaming Tweets The second dataset we used to pre-train the word embeddings was composed of tweets collected by Twitter Streaming APIs [1]. We used the collected tweets to pre-train the word embedding matrix $\boldsymbol{E}_w$ of the proposed model and the comparison models. Table 1 lists the number of resulting tweets. The process of collecting tweets was described in [9]. We will describe the process to pre-train the word embedding matrix in Section 4.3.

4.2 Model Initialization

We pre-trained each component for the proposed model. We used three steps to initialize the proposed model for training according to the following procedure.

[1] https://dev.twitter.com/streaming/overview

Language	#tweet
English	10.72M
Spanish	3.17M
Arabic	2.46M

Table 1. Number of tweets collected for each language with Twitter Streaming APIs. M in the table represents the million unit.

Initialization of text component We first pre-trained a word embedding matrix $\boldsymbol{E}_w$ for the text component. The details of the pre-training of the word embeddings will be described in Section 4.3. The text component was trained using $train_8$ and dev_1.

Initialization of image component The image component was trained by fine-tuning on $train_8$ and dev_1. First, the layers from $\mathrm{Conv.Layers1}$ to FC7 described in Figure 3 (VGG16) were pre-trained on ImageNet [5]. We then initialized $\mathrm{CNN_I}$, as described in Figure 1, using the pre-trained VGG16. Finally, $\mathrm{FC_{UI}}$ was then randomly initialized.

Initialization of TIFNN We described the pre-training procedure for each component using $train_8$ and dev_1 above. This was done because TIFNN could be successfully trained utilizing pre-trained text and image components. Thus, we used the pre-trained text and image components to train TIFNN. Therefore, all of TIFNN parameters except FC1 and FC2 were initialized with the parameters of the pre-trained components.

4.3 Model Configurations

Text pre-processing We applied unicode normalization, user name normalization, URL normalization, and HTML normalization. We used twokenizer [10] for the English text. We used WordPunctTokenizer in NLTK [2] for the other languages for tokenization.

Image pre-processing We applied two resizing methods: direct resizing and resizing-cropping.

- Direct resizing: We resized images to 224 pixels × 224 pixels.
- Resizing-cropping: We resized images to 256 pixels × 256 pixels and then cropped the center of each image to 224 pixels × 224 pixels.

Direct resizing was applied to an image-based neural network and resizing-cropping to TIFNN. After resizing, normalization was applied to all the images by subtracting the average values of the RGB channels for each language.

Initialization of word embeddings We used fastText [3] with the skip-gram algorithm to pre-train a word embedding matrix $\boldsymbol{E}_w$. The pre-training parameters were as follows: dimension = 100, learning rate = 0.025, window size = 5, negative sample = 5, and epoch = 5.

Parameter	# of size (pre-train)	# of size (train)
Word embedding dimension	100	100
$\mathrm{RNN_W}$ units	100	100
$\mathrm{FC1_{UT}}$	100	100
$\mathrm{FC2_{UT}}$	2	-
$\mathrm{CNN_I}$	4096	4096
$\mathrm{FC_{UI}}$	2	100
$\mathrm{FC1}$	-	64
$\mathrm{FC2}$	-	2

Table 2. Sizes of parameters in proposed model.

Parameters and pooling settings for proposed model Table 2 summarizes the number of parameters in the proposed model. In addition, $\mathrm{Pooling_W}$ was applied as a max pooling layer for each language, $\mathrm{Pooling_I}$ was applied as an average operation for each language, and $\mathrm{Pooling_T}$ was applied as a max pooling layer for Arabic or an average pooling layer for the other languages.

Optimization strategies We used cross-entropy loss as an objective function for the models. The objective function of TIFNN was minimized over shuffled mini-batches with SGD. We also used Adam for the text component and SGD for an image component. The initial SGD learning rate for the image component was set at $1e^{-3}$. In addition, we selected the best TIFNN learning rate for each language: $5e^{-3}$ for English and $1e^{-2}$ for the other languages.

Parameter selection The models had l_2 regularization parameter α. We selected the best parameter α of the text component from the following candidates. On the other hand, the parameter α of TIFNN was fixed at $\alpha = 1e^{-5}$.

$$\alpha \in \{1e^{-3}, 5e^{-4}, 1e^{-4}, 5e^{-5}, 1e^{-5}\}$$

We explored the best parameter α for each model using dev_1.

4.4 Comparison Models

We next describe the details of the comparison models used for the in-house experiment. Figure 4 illustrates the following comparison models, except the baseline.

baseline The model was constructed of SVM using TF-IDF uni-gram features.

Text NN The text component in the figure is the same as that for Figure 2 (from $\mathrm{WordEmbedding}$ to $\mathrm{FC1_{UT}}$). The parameter α is set to $1e^{-3}$ for English, $1e^{-4}$ for Spanish, and $5e^{-5}$ for Arabic.

Image NN The image component in the figure is the same as that for Figure 3 (from $\mathrm{Conv.Layers1}$ to $\mathrm{FC_{UI}}$). The model does not apply $l2$ regularization.

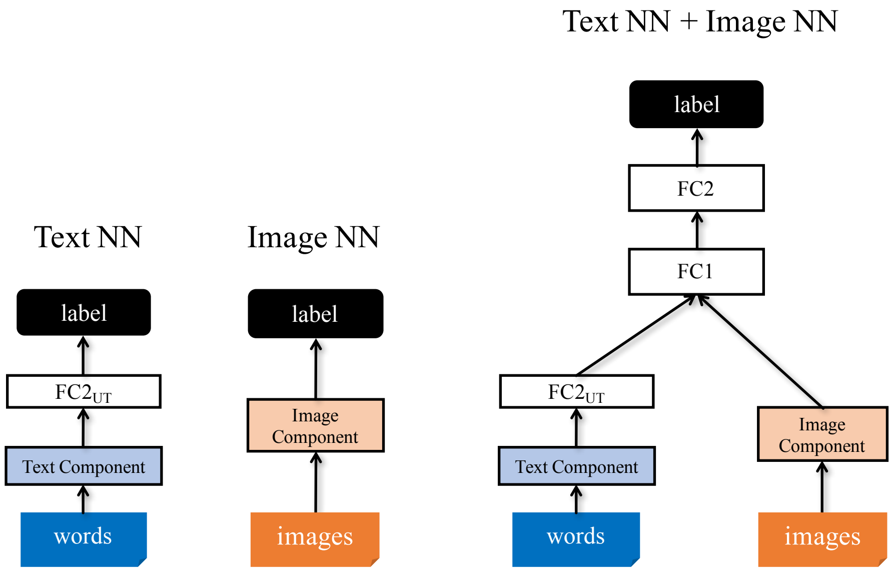

Figure 4. Overview of comparison models. The left model denotes Text NN, the center model denotes Image NN, and the right model denotes Text NN + Image NN.

Text NN + Image NN The model combines the text and image components. Note that the model is different from the proposed model in Figure 1. The details of this model can be described as follows. The user representation $\boldsymbol{r}_{txt} \in \mathbb{R}^M$ is obtained via the text component and $\boldsymbol{r}_{img} \in \mathbb{R}^L$ is obtained via the image component. These features are then concatenated to obtain $\boldsymbol{r}_{comb}$ as $\boldsymbol{r}_{comb} = [\boldsymbol{r}_{txt}, \boldsymbol{r}_{img}]$. Finally, the concatenated feature $\boldsymbol{r}_{comb}$ is fed to FC1, and FC2 is passed to the feature via FC1. The parameter of FC1 is different from that listed in Table 2; we set it to 100.

4.5 In-house Experiment

We evaluated the proposed model and the comparison models using $train_8$, dev_1, and $test_1$. With the exception of the baseline, the models were trained using Titan X GPUs. Table 3 summarizes the gender identification results.

As listed in Table 3, Text NN and Image NN achieved accuracies of 80.0-82.3% for each language. TIFNN drastically improved the accuracies (3-8pt) for each language compared with Text NN and Image NN in this task. Furthermore, TIFNN has also improved the accuracies for English and Spanish compared with Text NN + Image NN. This indicated that obtaining a fusion synergy via the fusion component was an effective approach for this task.

4.6 Submission Run

We chose the best performing models, which were the Text NN, Image NN, and TIFNN, as described in Table 3, for our submission run. The submission run was performed on

Model	Arabic	English	Spanish	Average
baseline	0.760	0.800	0.817	0.792
Text NN	0.813	0.817	0.803	0.811
Image NN	0.800	0.823	0.800	0.808
Text NN + Image NN	**0.840**	0.863	0.850	0.851
TIFNN	**0.840**	**0.903**	**0.863**	**0.869**

Table 3. Performances of proposed model and comparison models for each language on $test_1$. The evaluation metric is accuracy.

Model	Arabic	English	Spanish	Average
Text NN	0.771	0.797	0.786	0.785
Image NN	0.772	0.816	0.773	0.787
TIFNN	**0.785**	**0.858**	**0.816**	**0.820**

Table 4. Performances of our models in submission run. The evaluation metric is accuracy. These results are published in the official website of PAN.

a TIRA virtual machine [11] with CPUs. Table 4 summarizes the performances of the models in the submission run that are published as the official PAN results [2]. Although the models have lower accuracies compared with the in-house experiment, it is observed that TIFNN has better accuracies for each language compared with Text NN and Image NN. They ranked 1st in English ranking, 2nd in Spanish ranking, 7th in Arabic ranking, and 1st in Global ranking.

5 Conclusion

In this paper, we proposed Text Image Fusion Neural Network (TIFNN) for gender identification. In order to leverage the synergy of texts and images, the model computes the relationship between them using the direct-product. In-house experimental results showed that Text NN and Image NN achieved accuracies of 80.0-82.3% for each language in gender identification. TIFNN had drastically improved accuracies (+3-8pt) compared with Text NN and Image NN. Furthermore, TIFNN also had improved accuracies for English and Spanish compared with Text NN + Image NN. In addition to the results of this in-house experiment, we confirmed that TIFNN could improve the accuracy compared with individual models in a submission run.

In future work, we would like to analyze how the proposed model interacts with texts and images. We believe that understanding this interaction will make it possible to improve TIFNN.

[2] https://pan.webis.de/clef18/pan18-web/author-profiling.html

References

1. Basile, A., Dwyer, G., Medvedeva, M., Rawee, J., Haagsma, H., Nissim, M.: N-gram: New groningen author-profiling model. CoRR abs/1707.03764 (2017)
2. Bird, S., Loper, E., Klein, E.: Natural Language Processing with Python. O'Reilly Media Inc. (2009)
3. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016)
4. Cho, K., van Merriënboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder–decoder for statistical machine translation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1724–1734 (2014)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database. In: CVPR09 (2009)
6. Franco-Salvador, M., Plotnikova, N., Pawar, N., Benajiba, Y.: Subword-based deep averaging networks for author profiling in social media. In: CLEF (2017)
7. Kodiyan, D., Hardegger, F., Neuhaus, S., Cieliebak, M.: Author profiling with bidirectional rnns using attention with grus. In: CLEF (2017)
8. Martinc, M., Skrjanec, I., Zupan, K., Pollak, S.: Pan 2017: Author profiling - gender and language variety prediction. In: CLEF (2017)
9. Miura, Y., Taniguchi, T., Taniguchi, M., Ohkuma, T.: Author profiling with word+character neural attention network. In: CLEF (2017)
10. Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., Smith, N.A.: Improved part-of-speech tagging for online conversational text with word clusters. In: Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT). pp. 380–390 (2013)
11. Potthast, M., Gollub, T., Rangel, F., Rosso, P., Stamatatos, E., Stein, B.: Improving the Reproducibility of PAN's Shared Tasks: Plagiarism Detection, Author Identification, and Author Profiling. In: Kanoulas, E., Lupu, M., Clough, P., Sanderson, M., Hall, M., Hanbury, A., Toms, E. (eds.) Information Access Evaluation meets Multilinguality, Multimodality, and Visualization. 5th International Conference of the CLEF Initiative (CLEF 14). pp. 268–299. Springer, Berlin Heidelberg New York (Sep 2014)
12. Rangel, F., Rosso, F.C.P., Potthast, M., Stein, B., Daelemans, W.: Overview of the 3rd author profiling task at pan 2015. In: CLEF 2015 Labs and Workshops, Notebook Papers. CEUR Workshop Proceedings, CEUR-WS.org (Sep 2015), http://www.clef-initiative.eu/publication/working-notes (2015)
13. Rangel, F., Rosso, P., Montes-y-Gómez, M., Potthast, M., Stein, B.: Overview of the 6th Author Profiling Task at PAN 2018: Multimodal Gender Identification in Twitter. In: Cappellato, L., Ferro, N., Nie, J.Y., Soulier, L. (eds.) Working Notes Papers of the CLEF 2018 Evaluation Labs. CEUR Workshop Proceedings, CLEF and CEUR-WS.org (Sep 2018)
14. Rangel, F., Rosso, P., Potthast, M., Stein, B.: Overview of the 5th author profiling task at pan 2017: Gender and language variety identification in twitter. Working Notes Papers of the CLEF (2017)
15. Rangel, F., Rosso, P., Verhoeven, B., Daelemans, W., Potthast, M., Stein, B.: Overview of the 4th author profiling task at pan 2016: Cross-genre evaluations. In: Working Notes Papers of the CLEF 2016 Evaluation Labs. CEUR Workshop Proceedings (2016)
16. Schaetti, N.: Unine at clef 2017: Tf-idf and deep-learning for author profiling. In: CLEF (2017)

17. Shigenaka, R., Tsuboshita, Y., Kato, N.: Content-aware multi-task neural networks for user gender inference based on social media images. 2016 IEEE International Symposium on Multimedia (ISM) pp. 169–172 (2016)
18. Sierra, S., y Gómez, M.M., Solorio, T., González, F.A.: Convolutional neural networks for author profiling in pan 2017. In: CLEF (2017)
19. Simonyan, K., Zisserman, A.: VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION. In: International Conference on Learning Representations (ICLR) (2015)
20. Tellez, E.S., Miranda-Jiménez, S., Graff, M., Moctezuma, D.: Gender and language-variety identification with microtc. In: CLEF (2017)
21. Vijayaraghavan, P., Vosoughi, S., Roy, D.: Twitter demographic classification using deep multi-modal multi-task learning. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers (2017)