INTERNATIONAL JOURNAL OF INFORMATION SECURITY SCIENCE
P. Canbay. *et al.* ,*Vol.9, No.3, pp.154-163*

Deep Combination of Stylometry Features in Forensic Authorship Analysis

Pelin Canbay*‡, Ebru Akcapinar Sezer**, Hayri Sever***

* Department of Computer Engineering, Sutcu Imam University, Kahramanmaras, pelincanbay@ksu.edu.tr

**Department of Computer Engineering, Hacettepe University, Ankara, ebruakcapinarsezer@gmail.com

***Department of Computer Engineering, Cankaya University, Ankara, sever@cankaya.edu.tr

ORCID ID: 0000-0002-8067-3365 0000-0002-9287-2679 0000-0002-8261-0675

Research Paper Received: 11.06.2020 Revised: 14.07.2020 Accepted: 27.08.2020

Abstract- Authorship Analysis (AA) in forensic is a process aim to extract information about an author from his/her writings. Forensic AA is needed for detection characteristics of anonymous authors to make better the security of digital media users who are exposed to disturbing entries such as threats or harassment emails. To analyze whether two anonymous short texts were written by the same author, we propose a combination of stylometry features from different categories in different progress. In the majority of the previous AA studies, the stylometric features from different categories are concatenated in a preprocess. In these studies, during the learning process, no category-specific operations are performed; all categories used are evaluated equally. On the other hand, the proposed approach has a separate learning process for each feature category due to their qualitative and quantitative characteristics and combines these processes at the decision phase by using a Combination of Deep Neural Networks (C-DNN). To evaluate the Authorship Verification (AV) performance of the proposed approach, we designed and implemented a problem-specific Deep Neural Network (DNN) for each stylometry category we used. Experiments were conducted on two English public datasets. The results show that the proposed approach significantly improves the generalization ability and robustness of the solutions, and also have better accuracy than the single DNNs.

Keywords- Forensic Authorship Analysis; Deep Neural Networks; Neural Network Combination; anonymous document pairs.

1. Introduction

The production of textual data is continuously and exponentially increasing in digital environments day by day. The possibility to easily generate data anonymously in this environment has made it easier to commit anonymous cybercrimes, such as threatening or harassing someone by using a pseudonym. Hence, it is of great importance to be able to retrieve information about the authors from digital documents by analyzing the writing style. The writing style of an author is a soft (cognitive or behavioral) biometric [1, 2] and can be extracted with the help of the stylometry features. These features are used to transform the documents into the stylometric representation of an author. They could be lexical, structural, domain/content-specific, syntactic, or semantic [3]. For about 200 years, these features have been used to obtain the personal characteristics of the authors in many studies [4]. In this study, we investigated how these features should be used to increase the generalization ability of the authorship analysis methods.

Many stylometric techniques have been developed to retrieve the authors' writing styles. Due to the problem that researchers are interested in, the stylometric techniques are divided into five significant subtasks in the Authorship Analysis (AA): Authorship attribution, authorship verification, authorship profiling, stylochronometry, and adversarial stylometry [3].

Although researchers in the field of AA are commonly interested in two main subtasks, attribution and verification [5], the problems in this field are interrelated, and the optimal solution offered for one is also useful for the others due to the stylometric infrastructure. Authorship attribution (or Identification) studies try to assign a questioned document to an author. The difference of authorship attribution from verification problems is its candidate set. In authorship attribution studies, it is guaranteed that the author of the questioned document is in the candidate set; therefore, these studies are addressed as a closed set attribution problem [6]. However, there is no candidate set or any background information about the questioned document in authorship verification problems.

Authorship Verification (AV) is one of the main fields of Digital Text Forensic and AA [7, 8]. The goal of AV is to determine whether or not the author of a given set of documents is also the author of a questioned document (one-to-many) [9, 10]. In its most challenging form, which we are interested in, AV aims to decide whether given two anonymous short texts were written by the same author or not (one-to-one) [11, 12]. With this form, authorship of a text can be verified against a text of both specific and anonymous authors. In this study, as a one-to-one AV problem, we dealt with whether the stylometric difference of given two documents represents the style of a single author or not. Our aim with this consideration is to contribute to the solution of the problems such as whether some threat messages are thrown by the same person or whether biased comments for a product are made by the same person or not. Since AV is a challenging problem, it is shown that the models which contain an ensemble of features from different stylometry categories yield more accurate results than that contain single ones [1, 13, 14]. The majority of the AA studies concatenated stylometry features of different categories before the learning processes [14, 15]. Differently, we developed a separate learning process for each stylometry category used due to their qualitative and quantitative characteristics and combined the processes in the decision phase of the solution. By doing so, the generalization ability of the learning model, and representativeness of the stylometry features have been increased.

The problems of AA are classification problems. They can be evaluated in one-class, binary or multi-class classification according to the problem handled. Many supervised classification techniques have been used to solve the kinds of AA problems, including tree-based learnings, logistic regression, Support Vector Machines, Bayesian classifiers, and so on [3]. In clusterings or unsupervised manners, techniques such as K-means or Expectation Maximization have been generally used for finding similar texts in the groups based on their stylistic similarities [8, 16]. Besides, due to the multidimensionality of some features, feature selection, or dimension reduction techniques such as Principal Component Analysis have been used to reduce the dimensionality of the features [17]. In our study, we implemented a supervised learning scheme by using Artificial Neural Networks (ANNs). ANNs with more than one hidden layer and different architectures are ones of the Deep Neural Network (DNN) models. To improve the generalization ability of neural networks, a combination of multiple architectures is a very promising approach [18]. Although designing an efficient combination is more complicated than designing a single network, a combination approach produces a more robust and generalized solution than a single network. Therefore, instead of using the best single DNN architecture for the solutions of AA problems, we proposed a new approach using a combination of different DNN architectures (C-DNN). The DNNs, which were designed for different categories of stylometry features, are combined in the decision stage of the proposed C-DNN architecture. We present the AV performances of both single DNNs and C-DNNs for feature categories used. By doing this, we illustrate the effects of using different feature categories as well as different models together.

Two public, English datasets are employed to test the performances of the proposed C-DNN approach and single DNNs. The experimental results show that the proposed C-DNN approach produced a more robust, more generalized, and more accurate solution than the single DNNs.

2. Problem Statement

2.1. Authorship Analysis in Digital Text Forensic

Forensic AA has been studied for many years in solving many problems such as plagiarism detection or determining the author of the disputed texts [19]. With the digitalization of texts, these studies have evolved as AA of digital text forensic [20]. In addition to the messages that jeopardize personal or corporate reliability and security made by anonymous users, many illegal transactions also have been carried out digitally. For example, external groups communicate via web forum messages [24] or in-company fraud has been done via emails [8]. Forensic AA studies are used to detect these kinds of cybercrimes and criminals. With the help of these studies, information such as age, gender, social or psychological status of an author can be extracted even from short texts. Among the Forensic AA studies, we handled AV which is the most suitable in cases where the candidate criminal set or background information about the criminals is not available.

2.2. Authorship Verification (AV)

One of the fundamental Forensic AA studies is AV. In 2000, Stamatatos et al. dealt with the requirement of the confirmation (or rejection) of the hypothesis that a given person is the author of the questioned text [21]. The study was the first in which the authorship verification problem was addressed. In this case, there exist texts written by an author, and an external text is questioned whether written by this particular author or not. On the other hand, in 2014, Koppel and Winter [22] handled the AV problem as just one relatively short document being available as a known document of an author. If there is only one relatively short known document to verify the authorship of the questioned document, the problem becomes more challenging. The problem, which is considered as authorship verification of two anonymous texts (one to one comparison), forms the basis of many authorship analysis studies [1]. The difficulties of this kind verification problem include the length of the known and questioned documents [22], the selection of the best discriminative feature sets [14] and, finding a successful response function [23].

Studies in AA have shown that the best discriminative features are not from the single stylometry category. Using stylometric features from different categories together increases the success of the many proposed methods [1, 13, 14, 24]. The majority of the studies either used a single stylometry category in a solution [5, 25, 26] or used a single learning model for different feature categories [2, 14, 27, 28]. Unlike the literature, we propose a new approach that evaluates stylometric features according to their qualitative and quantitative characteristics in the learning process. We developed a separate learning process for each stylometry category used and combined these processes in the decision phase with the help of a C-DNN architecture. We used a different process for each feature category to increase the robustness of the learning model and the representativeness of the features used. The proposed approach offers more generalized solution then which using a single stylometry category, or a single model for different category of features.

3. Producing of AV Samples

3.1. Adjustments of Datasets

Two popular datasets, English Blog Corpus [29] and the PAN-2015-English dataset [30], were used to test the performance of the proposed approach. English Blog Corpus consists of the collected posts of 19320 bloggers. From this data set, we obtained a new data set by randomly taking 20 pieces of text containing 500 words from 1000 authors. 20 positive-document-pairs, which means they are taken from the same author, were randomly taken from each author. Similarly, 20 negative-document-pairs, which means they are chosen from the different authors, were obtained by using randomly selected texts of other authors for each author. Thus, we obtained a new dataset with 40,000 samples containing 20,000 positive and 20,000 negative samples.

The PAN-2015-English-Dataset has a collection of dialog lines from plays, excluding the list of speaker names, characters, and so on [30].

This dataset has 100 document pairs in the train set and 500 document pairs in the test set. Since deep architectures would not perform well with a small train set, we gathered the train and test set into one in the experiments.

3.2. *Stylometry Features*

In the literature, more than a thousand stylometry features have been used in various text analysis methods. These features are categorized as lexical, syntactic, semantic, structural, and content-specific [31, 32]. There is no agreement among studies regarding which categories and features yield the best results on authorship analysis. Although they are very different in quantity and quality, the majority of studies used some of the categories together to increase success to be achieved [1, 14, 24].

Three different feature categories were used in the experiments of this study. As lexical features (c1), we used frequencies of token 5 prefixes, which includes all words up to five characters and the first five characters for long words. Punctuation frequencies (c2) were used as syntactical features and, structural features (word-based paragraph length, average word length, number of character per text, number of punctuation per text) were used for the structural feature category (c3). We used these feature sets in two forms. In the first form, we concatenated all features to get the vector representation of documents (dc) before the learning process. Each document is represented by a single vector which contains different features from different stylometry categories. In the second form, we used three different vector representations (dc1, dc2, and dc3) for each document separately. Each representation carries the characteristics of a specific feature category and has a separate process in the learning phase. The representations of these four vectors are shown below.

$$d(c1) = [c1_1, c1_2, c1_3, \ldots, c1_n]$$

$$d(c2) = [c2_1, c2_2, c2_3, \ldots, c2_m]$$

$$d(c3) = [c3_1, c3_2, c3_3, \ldots, c3_k]$$

$$dc = [c1_1, c1_2, c1_3, \ldots, c1_n, c2_1, c2_2, c2_3, \ldots, c2_m, c3_1, c3_2, c3_3, \ldots, c3_k]$$

The c1, c2, and c3 represent the categories of the features, and $c2_m$ represents the m^{th} feature value of the c2 category.

3.3. *Evaluation of Document Pairs*

To evaluate the document pairs in one representation, we used the absolute difference of the pairs in vector space. By doing this, we obtained a new vector representation that has the same number of dimensions of the evaluated pairs. Using this representation, we questioned whether the absolute difference of a document pair represents the style of a single author or not.

We evaluated the document pairs in question to produce a single representation of positive and negative AV samples as implemented in the supervised method of [22]. Each AV sample was obtained from the pairs of document vectors (X and Y) by using Eq. (1).

$$C(X, Y) = [\,|X_1 - Y_1|, |X_2 - Y_2|, |X_3 - Y_3|, \ldots, |X_n - Y_n|\,] \quad (1)$$

Let X and Y be the feature vectors of the two documents, and Xi and Yi be the value of related features. If X and Y were the same author's pair, vector C was labeled as positive or negative in otherwise. By applying this labeling process to the text pairs in both data sets used, they become two-class sets containing positive and negative AV samples.

4. Deep Combination of Stylometry Features

In this study, we applied a new approach employing C-DNN to solve the problem of verifying authorship of two anonymous documents. We designed a single DNNs for each stylometry category and concatenated form of all categories. Then, in the C-DNN architecture, we combined the single DNNs in the decision stage of a deep architecture. Details of the architectures are given below.

4.1. DNN Architectures

While computational techniques evolve rapidly, ANNs with deep architectures provide strong structures for supervised learning methods [33]. In this study, suitable deep architectures with Neural Networks (DNNs) were investigated regarding the AV problem. According to the type of problem and data set used, DNN structures can take many forms in terms of the layer they contain and the hyperparameters used in each layer, such as the number of neurons or type of activation functions. Producing the appropriate architecture for every problem, and every data set is a complicated process. Especially in order to prevent the overfitting of the generated method, it is necessary to determine the appropriate architecture and parameters specific to each problem and each data set. The proposed DNN designs are shown in Figure 1.

The DNN architectures shown in Figure 1 are the representation of the AV models produced for the blog dataset. The first architecture shown in the left side of Figure 1, was used for both c1 (lexical features) and dc (concatenation of all features) samples. The second architecture shown in the middle of Figure 1 was used for c2 (syntactical features) samples. The last architecture shown on the right side of Figure 1 was used for c3 (structural features) samples. The input samples of the blog dataset, in the form of dc, are 35880-dimensional vectors. In all dense layer except from the last one, we used the selu (Scaled Exponential Linear Unit) activation function and lecun_normal distribution for kernel initializer. In the last dense layer, we used the sigmoid activation function to predict the class of the samples. In all the DNN architectures, when we increase the number of layers or increase the number of neurons in each layer, the method's overfitting tendency increases in parallel. Although the blog dataset contains 40,000 samples, it is not necessary to produce a more complex architecture.

On the other hand, although the PAN dataset contains 600 samples, it produced successful results in the same DNN architecture with dc form. The PAN samples are 5551-dimensional vectors, and they were also used in the production of an AV model using the same architectures.

4.2. C-DNN Approach

To increase the representativeness of the different stylometry feature categories and generalization ability of the proposed DNNs, we

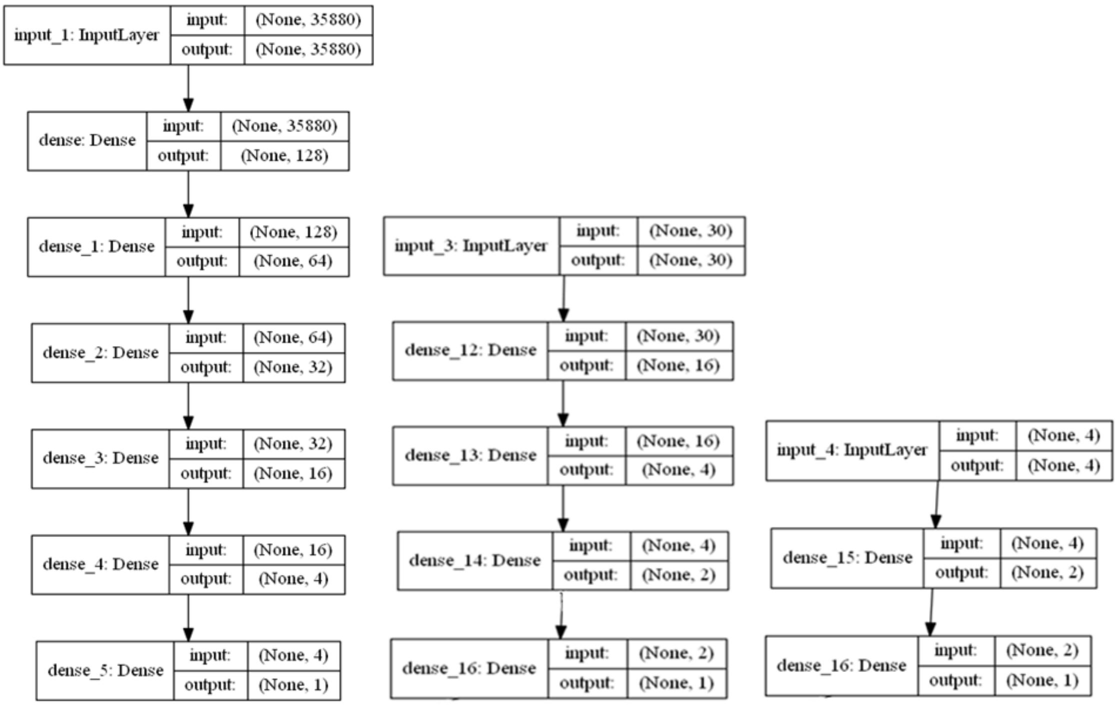

Fig. 1. DNN architectures designed for different stylometry features

designed a C-DNN architecture that has suitable layers and neurons in each layer. The proposed C-DNN architecture has 3 phases; learning or reducing, combining, and deciding. The proposed C-DNN architecture for the AV is shown in Figure 2.

The C-DNN architecture shown in Figure 2 is the representation of the AV model designed for the blog dataset. The first phase of the C-DNN architecture includes three different vector representations of a sample. The representations are based on inputs taken from three different categories of stylometry features used. This phase learns encoded or reduced forms of each input category for the AV problem. The inputs which are encoded due to their qualitative and quantitative characteristics are concatenated in the second phase.

Lexical features are the most used features in the AA studies [9, 22], and according to the literature, the features are more effective than the syntactic features when they are used individually [14]. Therefore, we used 4 neurons for the encoded form of the first category. On the other hand, different syntactic features are also successful in AA studies [26, 27]. The syntactic features we used in this study are shorter than the lexical features in terms of quantity and are used less in the literature; accordingly, we used 2 neurons for the encoded form of the second category. The third category has just 4 features and is generally used for supporting the main feature set. Thus, we used only 1 neuron for the encoded form of the last category. In all layers up to the concatenation layer, the properties of dense layers of DNNs have been preserved.

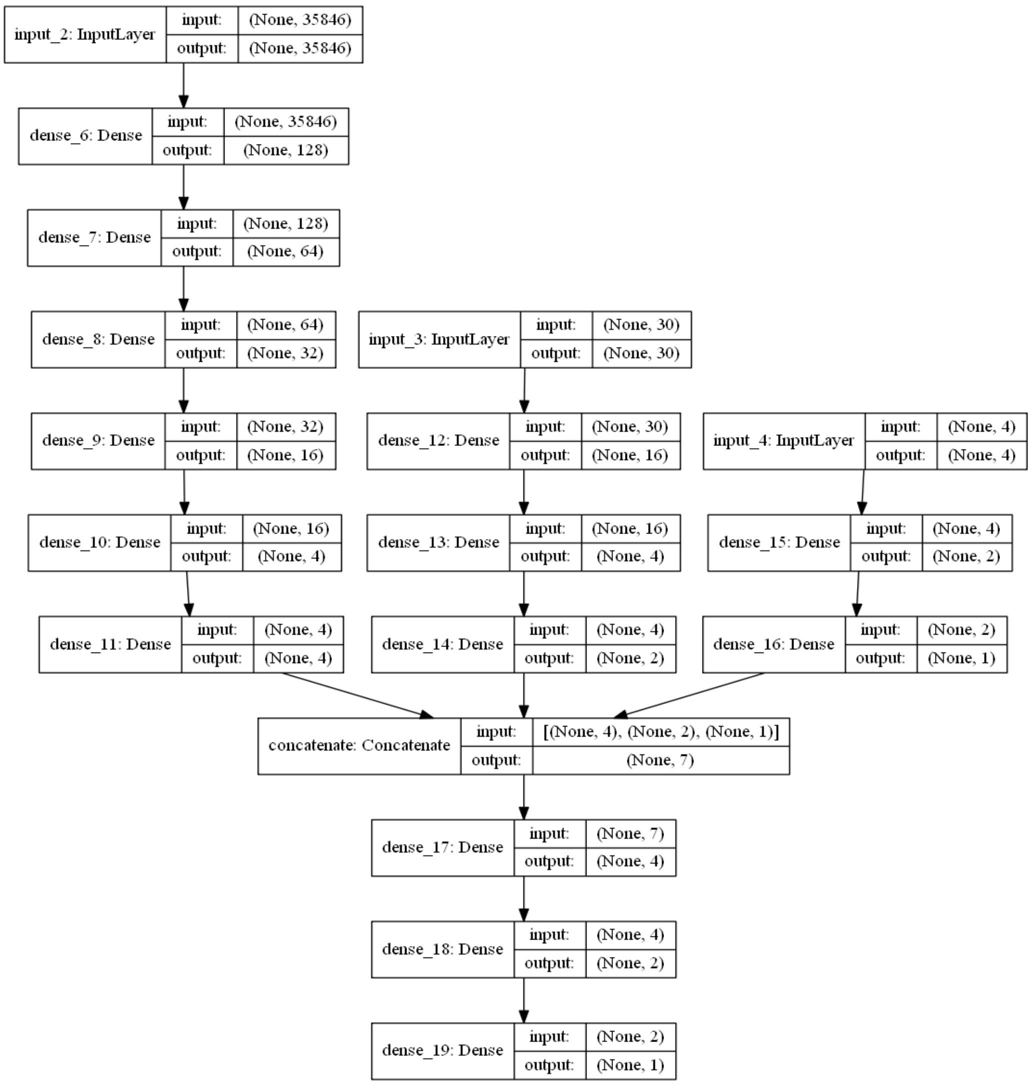

Fig. 2. The proposed C-DNN architecture designed for AV

In order to increase the strength of the combination in the last deciding phase, two more Dense layers were added to the architecture. In these layers, we used the tanh (Hyperbolic tangent) activation function and uniform distribution as the kernel initializer. The final decision was obtained from the last layer, which has the sigmoid activation function. For the PAN samples, similar adjustments were used.

5. Experimental Results

In the experiments, the proposed DNN and C-DNN approaches were tested by using two pre-mentioned datasets to evaluate their AV performances. All experiments were done on Tensorflow 2.0 using Keras in python 3.7. We group the samples into batches of size 100, and we used 50 epochs in all experiments. Since the datasets were balanced in terms of the number of positive and negative samples, we preferred to use accuracy and f-measure for evaluations. These measures are formulated via TP (true positive; the number of samples correctly predicted as positive), TN (true negative; the number of samples correctly predicted as negative), FP (false positive; the number of instances incorrectly predicted as positive), and FN (false negative; the number of instances which are incorrectly predicted as negative). The formulas of these measures are shown in Eq. (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + TN}$$

$$F - measure = \frac{2(Precision\ x\ Recall)}{(Precision\ x\ Recall)} \tag{2}$$

According to the measures given Eq. (2), the AV accuracies obtained from the DNN and C-DNN architectures of the blog and PAN datasets and their comparisons are shown in Figure 3.

The results shown in Figure 3 were taken from the experiments of the blog dataset, except the last column. Using features from a single category, although there are no major differences between the results obtained, the best accurate result was taken from the lexical features. On the other hand, the proposed C-DNN architectures produced more accurate results then DNN architectures in each combination of the stylometry category used. A similar difference of results was also obtained from the PAN dataset using all feature categories with DNN and C-DNN architectures.

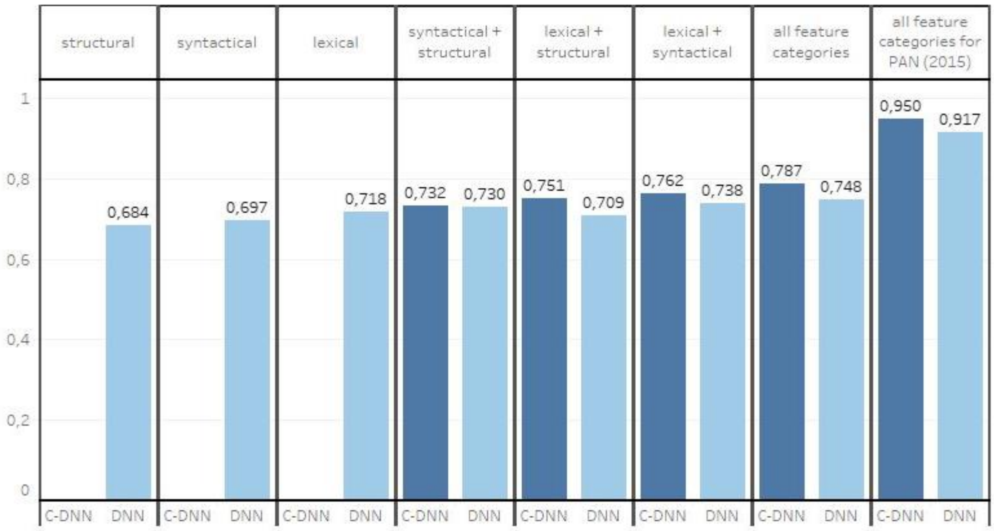

Fig. 3. Accuracy comparison of the AV performances of DNN and C-DNN architectures

The AV f-measures obtained from the DNN and C-DNN architectures of the blog and PAN datasets, and their comparisons are shown in Figure 4.

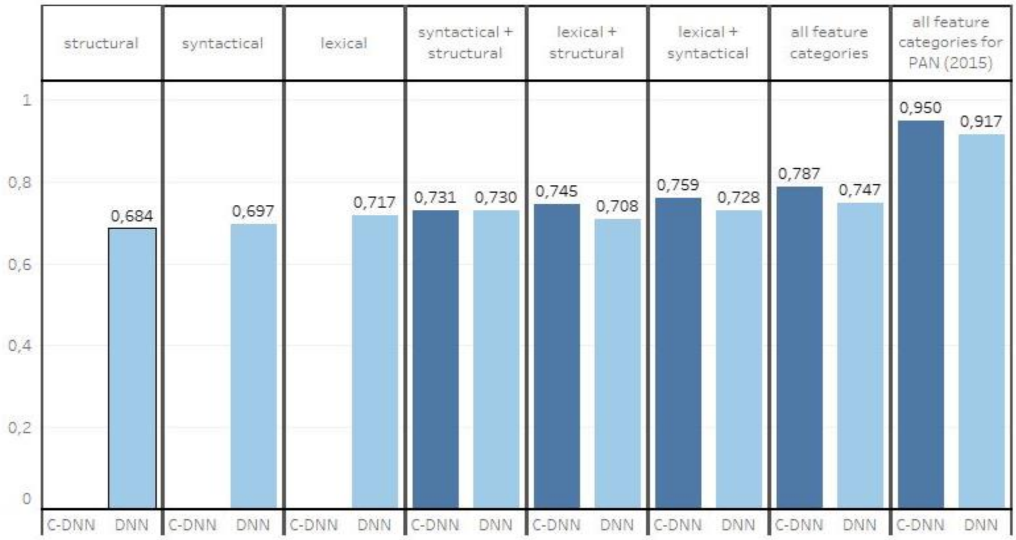

Fig. 4. F-measure comparison of the AV performances of DNN and C-DNN architectures

Although the datasets used in the experiments are balanced, the f-measures were also calculated to show the robustness of the proposed approach. As shown in Figure 4, the results obtained by f-measures are almost the same as those obtained by accuracy measures. According to the results taken from f-measures, the proposed C-DNN architectures produced more robust results then DNN architectures in each combination of the stylometry category used.

In the experiments of which results are given in Figure 3 and Figure 4, 10-fold cross-validation was applied. In the figures, all columns show the AV performances of the blog dataset except the last one. Considering the accuracy obtained using the DNN architecture, high accuracy was obtained from the experiments which evaluate all features together. On the other hand, in addition to increasing the accuracy obtained from the same features, the generalization of the solution has also increased by using C-DNN architecture. The last columns of the figures show the AV performances of the PAN dataset we used. To show the generalization ability of the C-DNN architecture, we present the accuracy performances of the train and test sets of the PAN dataset in Figure 5.

The PAN dataset we used is implemented with the same architectures used with the blog dataset. In order to show the success of the proposed architecture on a different dataset, PAN experiments were implemented with all features from the different categories on DNN and C-DNN architectures. As shown in Figure 5, the difference between the test and train accuracies is smaller in

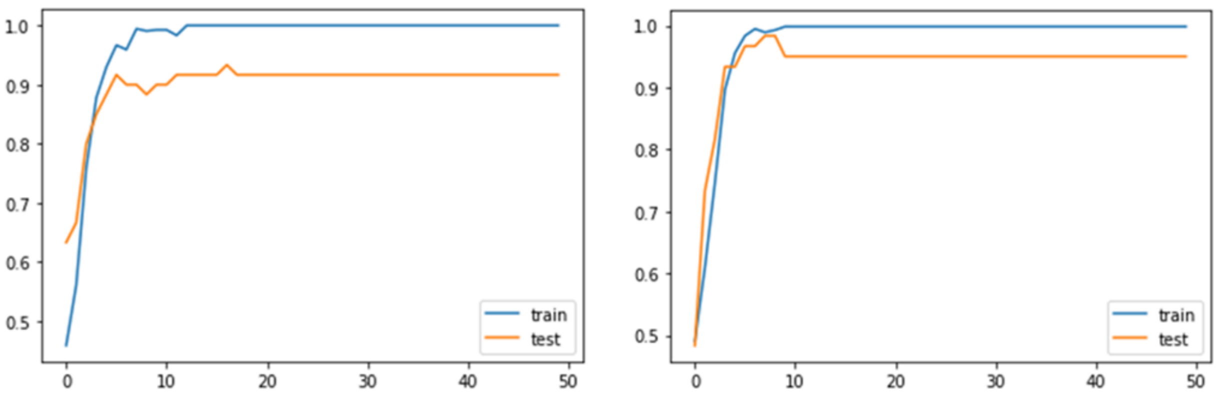

Fig. 5. Accuracies of the PAN dataset using all features in DNN (left) and C-DNN (right) architectures

C-DNN architecture than that of DNN. It means that the C-DNN approach yields a more robust and generalized solution than a single DNN approach.

6. Conclusions and Future Work

Many illegal transactions such as terrorist group communications, internal or external fraud, false news publishing, etc have been carried out anonymously in digital environments. Most of these cyber crimes have been made via textual messages. In such crimes that jeopardize personal or corporate reliability and security, it is very important to have information about the criminals. With Forensic Authorship Analysis, many characteristics such as age, gender, social or psychological status of an author can be extracted from digital texts. The requirement of these forensic-based authorship analysis studies has increased with the increase of anonymous authors or criminals in the digital environment. In this study, we proposed a method of deep combination of stylometry features, that can increase the success and robustness of Forensic Authorship Analysis studies.

Stylometry features are the properties used for extracting information from texts. The majority of the AA studies have combined different categories of stylometry features as a pre-processing in a solution and used them in a single learning process. In this study, a Combination of Deep Neural Network (C-DNN) approach, which combines different categories of stylometry features from different processes, is introduced, implemented, and successfully achieved for the first time.

The proposed approach has three phases; learning, combining and deciding. In the learning phase, we propose to produce an appropriate DNN architecture for each feature category used, according to their qualitative and quantitative characteristics. This phase learns encoded or reduced form of each category. In the combining phase, the encoded forms of different categories of features are concatenated to obtain a single representation of a sample. In the deciding phase, the class of the sample under consideration is decided.

We used the C-DNN approach in the investigation of whether the stylometric difference of given two documents represents the style of a single author or not. As an AV problem, although this investigation is one of the most challenging of authorship analysis studies, we got better accurate and robust results using C-DNN than of the single architectures.

C-DNN is a promising design alternative since the results obtained by combining a set of classifiers tend to be better than a single best classifier. The experimental results confirm that the generalization ability of a solution and the representativeness of stylometry features increase while using the C-DNN approach.

By using sequence information of different stylometry features, effective deep learning methods such as CNN and LSTM will be tested in the C-DNN approach as future works in the solutions of the AA problems.

References

[1] N. Pokhriyal, K. Tayal, I. Nwogu, and V. Govindaraju, "Cognitive-Biometric Recognition from Language Usage: A Feasibility Study," Transactions on Information Forensics and Security, vol. 12, no. 1, pp. 134-143, 2016.

[2] T. Neal, K. Sundararajan, and D. Woodard, "Exploiting linguistic style as a cognitive biometric for continuous verification," in Proceedings - 2018 International Conference on Biometrics, ICB 2018, pp. 270–276, 2018.

[3] T. Neal, K. Sundararajan, A. Fatima, Y. Yan, Y. Xiang, and D. Woodard, "Surveying stylometry techniques and applications," ACM Computing Surveys (CSUR), vol. 50, no. 6, pp. 1-36, 2017.

[4] P. Juola, "Authorship attribution," vol. 3, Now Publishers Inc, 2008.

[5] M. L. Brocardo, I. Traore, S. Saad, and I. Woungang, "Authorship verification for short messages using stylometry," in 2013 International Conference on Computer, Information and Telecommunication Systems, CITS 2013, pp. 1-6, 2013.

[6] M. Koppel, J. Schlier, and S. Argamon, "Computational methods in authorship attribution," Journal of the American Society for information Science and Technology, vol. 60, no. 1, pp. 9–26, 2009.

[7] M. Koppel, J. Schler, S. Argamon, and Y. Winter, "The 'Fundamental Problem' of Authorship Attribution," English Studies, vol. 93, no. 3, pp. 284–291, 2012.

[8] F. Iqbal, H. Binsalleeh, B. C. M. Fung, and M. Debbabi, "Mining writeprints from anonymous e-mails for forensic investigation," Digital Investigation, vol. 7, no. 1-2, pp. 56-64, 2010.

[9] E. Stamatatos, “Authorship verification: a review of recent advances,” Research in Computing Science, vol. 123, pp. 9-25, 2016.

[10] M. Litvak, “Deep dive into authorship verification of email messages with convolutional neural network,” Annual International Symposium on Information Management and Big Data. Springer, Cham, pp. 129-136, 2018.

[11] S. H. H. Ding, B. C. M. Fung, F. Iqbal, and W. K. Cheung, “Learning stylometric representations for authorship analysis,” IEEE transactions on cybernetics, vol. 49, no. 1, pp. 107-121, 2019.

[12] B. Boenninghoff, R. M. Nickel, S. Zeiler, and D. Kolossa, “Similarity Learning for Authorship Verification in Social Media,” in ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2457-2461, 2019.

[13] R. Zheng, J. Li, H. Chen, and Z. Huang, “A framework for authorship identification of online messages: Writing-style features and classification techniques,” Journal of the American society for information science and technology, vol. 57, no. 3, pp. 378-393, 2006.

[14] O. Halvani, C. Winter, and A. Pflug, “Authorship verification for different languages, genres and topics,” Digital Investigation, vol. 16, pp. S33–S43, 2016.

[15] P. Varela, E. Justino, A. Britto, and F. Bortolozzi, “A computational approach for authorship attribution of literary texts using sintatic features,” in Proceedings of the International Joint Conference on Neural Networks (IJCNN). IEEE, pp.4835-4842, 2016.

[16] J. Dunn, S. Argamon, A. Rasooli, and G. Kumar, “Profile-based authorship analysis,” Digital Scholarship in the Humanities, vol. 31, no. 4, pp. 689-710., 2016.

[17] S. Afroz, A. Caliskan-Islam, A. Stolerman, R. Greenstadt, and D. McCoy, “Doppelgänger finder: Taking stylometry to the underground,” in Proceedings - IEEE Symposium on Security and Privacy, pp. 212-226, 2014.

[18] Z. Ahmad and J. Zhang, “Selective combination of multiple neural networks for improving model prediction in nonlinear systems modelling through forward selection and backward elimination,” Neurocomputing, vol. 72, no. 4-6, pp. 1198-1204, 2009.

[19] E. Stamatatos, “A survey of modern authorship attribution methods,” Journal of the American Society for information Science and Technology, vol. 60, no. 3, pp. 538-556, 2009.

[20] P. Rosso, M. Potthast, B. Stein, E. Stamatatos, F. Rangel, and W. Daelemans, “Evolution of the PAN Lab on Digital Text Forensics,” In Information Retrieval Evaluation in a Changing World, Springer, Cham, pp. 461-485, 2019.

[21] E. Stamatatos, G. Kokkinakis, and N. Fakotakis, “Automatic text categorization in terms of genre and author,” Computational Linguistics, vol. 26, no. 4, pp. 471–495, 2000.

[22] M. Koppel and Y. Winter, “Determining if two documents are written by the same author,” Journal of the Association for Information Science and Technology., vol. 65, no. 1, pp. 178–187, 2014.

[23] S. Adamovic, V. Miskovic, M. Milosavljevic, M. Sarac, and M. Veinovic, “Automated language-independent authorship verification (for Indo-European languages),” Journal of the Association for Information Science and Technology, vol. 70, no. 8, pp. 858–871, 2019.

[24] A. Abbasi and H. Chen, “Applying authorship analysis to extremist-group Web forum messages,” IEEE Intelligent Systems., vol. 20, no. 5, pp. 67-75, 2005.

[25] P. Shrestha, S. Sierra, F. A. González, P. Rosso, M. Montes-Y-Gómez, and T. Solorio, “Convolutional neural networks for authorship attribution of short texts,” in 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, pp. 669-674, 2017.

[26] Jafariakinabad, Fereshteh, S. Tarnpradab, and K. A. Hua, “Syntactic Neural Model for Authorship Attribution,” in The Thirty-Third International Flairs Conference, pp. 234-239, 2020.

[27] F. Jafariakinabad and K. A. Hua, “Style-aware neural model with application in authorship attribution,” 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019, pp. 325-328, 2019.

[28] M. L. Brocardo, I. Traore, I. Woungang, and M. S. Obaidat, “Authorship verification using deep belief network systems,” International Journal of Communication Systems, vol. 30, no. 12, e3259, 2017.

[29] M. Koppel and J. Schler, “Authorship verification as a one-class classification problem,” in Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004, pp. 489–495, 2004.

[30] E. Stamatatos et al., “Overview of the author identification task at PAN 2015,” in CEUR Workshop Proceedings, vol. 1391, pp. 1–8, 2015.

[31] S. Seidman, “Authorship verification using the impostors method: Notebook for PAN at CLEF 2013,” in CEUR Workshop Proceedings, vol. 1179, pp. 23-26, 2013.

[32] C. Sanderson and S. Guenter, “Short text authorship attribution via sequence kernels, Markov chains and author unmasking: An investigation,” in COLING/ACL 2006 - EMNLP 2006: 2006 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference, pp. 482–491, 2006.

[33] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, “A survey of deep neural network architectures and their applications,” Neurocomputing, vol. 234, pp. 11-26, 2017.