Author Verification in Stream of Text with Echo State Network-based Recurrent Neural Models

Nils Schaetti
Université de Neuchâtel
Rue Emile-Argand 11
CH-2000 Neuchâtel
Switzerland
nils.schaetti@unine.ch

Abstract

This paper evaluates a type of recurrent neural networks (RNN) named Echo State Network (ESN) on a NLP task referred as author verification. In this case, the model has to identify whether or not a given author has written a specific text. We evaluate these models on a difficult task where the goal is to detect the author in a noisy text stream being the result of a collaborative work of an unknown number of authors. We construct a new dataset (denoted SFGram) composed of science-fiction books, novels and magazines. From this dataset we select three authors, published between the 1952 and the 1974, and we evaluate the effectiveness of ESNs with word and character-based representations to detect these authors in a set of 91 science-fiction magazines (containing around 8M of words).

1 Introduction

In recent years, the need for computer systems able to extract authorship information became important due to the increasing impact of social networks and to the fast growing set of texts available on the internet. In this context, the field of authorship analysis has attracted a lot of attention in the last decade.

Author verification is a well-known task in the authorship attribution domain. In this case, given a single author having written a set of documents, the objective is to determine if a new unseen text has been written or not by this target author. This problem can be viewed as a binary classification problem where the number of candidates is limited to one. This question is harder than traditional attribution tasks because a single author is provided without giving a set of possible impostors.

The motives behind author verification are related to the field of computer security, forensics, law, intelligence, and humanities. For example, forensic experts want to make sure that the author of a given text is not someone under investigation (Olsson and Luchjenbroers, 2013). In humanities, literature experts try to answer the question : *Did Shakespeare write this play ?*

However, the structure of textual data available today on the internet and on social networks does not allow them to be handled as simple documents. Communication systems such as Twitter, Facebook or instant messaging look more like continuous text streams where the segmentation into paragraphs is sometimes problematic as well as identifying the boundaries between two text streams. Consequently, we need systems able to detect such boundaries or events in addition to document classification.

In this paper we propose to evaluate the effectiveness of recurrent neural networks on such a new task where the model has to identify text passages written by a given author, and to detect *points of interest*, defined as positions in a textual stream where authorship is changing. These points can be used thereafter to detect if a given author participated or not to a collaborative work.

Recurrent neural networks are well known for their effectiveness to take the temporal dimension of any length into account. In the NLP field, it means that they are able to take into account word order, a feature ignored with the traditional bag-of-words model.

In this study, we evaluated a specific kind of RNN referred as Echo State Network (ESN) to identify an author in a noisy text stream. The suggested model is based on one-class learning which consists to draw an optimal threshold circumscribing all positives examples of the true author.

(a) Galaxy Magazine, April 1966 (b) Galaxy Magazine March 1972 (c) Lazarus come forth, Robert Silverberg, April 1966 (d) The Gods Themselves, Isaac Asimov, March 1972

FIGURE 1 – Examples of covers and novels extracted from archives.org and included in the SFGram data after digitalisation.

To evaluate our models, we use science-fiction magazines made publicly available through `archive.org`. These texts were digitised with optical character recognition (OCR) and therefore contain a high level of errors and word misidentification. Of all known authors in this corpus, we select three, namely Isaac Asimov, Philip K. Dick, and Robert Silverberg to test our different author verification models. The issues featuring these authors were published between 1952 and 1974. We based our selection on three criteria : their popularity, the number of their contributions, and the fact that they are known not to use pseudonyms.

More precisely, two tasks have been considered. The first task consists to answer the question : *Which text passages have been written by author x ?* The model handles each magazine issue as a stream and must determine an authorship probability at each time step (measured by word-token). The performance is measured using the F_1 score.

In the second task, the model must respond to : *In the collection, which issue contains text written by author x ?* For this task, we used *interest points* to detect whether or not this author wrote a passage (a paragraph or a sequence of paragraphs) in that issue. The final result being also evaluated using F_1.

The rest of this paper is organised as follows. Section 2 introduces related work on author verification, Reservoir Computing and RNN. Section 3 presents the dataset and the evaluation process. Section 4 defines the models and the features used in this paper. Section 5 evaluates the performances of ESNs on these tasks. Finally, section 6 discusses the results of our work and the possibilities of further investigation.

2 Related work

Stamatatos et al. (2000) presents the author verification problem and suggests using multiple regression models to predict whether or not a document was written by a given author. In Van Halteren (2004), a similar model was used in conjunction with a statistical learning approach. The unmasking method based on Support Vector Machines (SVM), is the most known method for this task. It was introduced in Koppel et al. (2007) and used recall, precision and F_1 as evaluation measures, with a corpus of student essays written in Dutch. In Escalante et al. (2009), the same metrics were used to evaluate the application of particle swarm model for determining sections written by a given author. In Koppel and Winter (2014), an effective method to transform this one-class classification task to a multiple-class classification problem was introduced in which additional texts reflecting the style of other possible authors (impostors) are included in the verification procedure.

The author verification task was proposed in different CLEF-PAN evaluation campaigns. In 2011 (Argamon and Juola, 2011), the author identification task included a three authors verification problem with a corpus of emails. The 2013 version of CLEF-PAN was entirely focused on author verification (Juola and Stamatatos, 2013). Precision, recall and F_1 score were used as evaluation measures in these applications.

ESNs have been applied to different scientific fields such as temporal series prediction and classification (Wyffels and Schrauwen, 2010; Coulibaly, 2010), and image classification (Schaetti et al., 2015, 2016). In NLP, they have been applied to cross-domain authorship attribution (Schaetti, 2018) and to author profiling on social network data (Schaetti and Savoy, 2018). The behaviours of ESNs on NLP tasks have been extensively studied in Schaetti (2019) and other recurrent neural network models such as RNNs, LSTMs, and GRUs have been applied to stylometry in Wang (2017) and Qian et al. (2017).

3 Evaluation Methodology

To evaluate author verification models, we generated a dataset named SFGram composed of 1,393 science-fiction books, novels and magazines for a total of 7,067 authors. The books and novels were extracted from the Gutenberg project and the magazines were extracted from `archive.org`.

Our dataset contains two well-known science-fiction magazines namely *Galaxy Science Fiction* and *IF Science Fiction* (IF). The first one was an American science-fiction magazine published from 1950 to 1980 and was the leading science fiction magazine of that time. The IF magazine was also an American magazine published from 1950 to 1974. These magazines contain different sections written by different authors. Along classical science-fiction novels, they contains ads (sometimes in the middle of a novel), editorials and readers' mail which ends up in an unknown number of authors.

For our study, we choose three well-known science-fiction authors who published during the same period : Isaac Asimov, Philip K. Dick, and

Author	Doc.	Words	Ratio
Asimov	22	341,480	4.25%
Dick	25	264,504	3.26%
Silverberg	45	911,219	11.25%
SFGram	91	8,102,853	

TABLE 1 – Number of documents and words per authors included in the dataset. The SFGram line shows the total number of documents and words. The ratio is the percentage of words written by the author in the corpus.

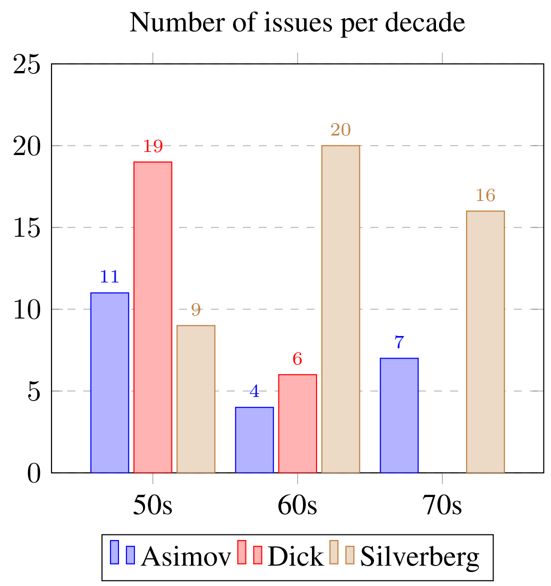

FIGURE 2 – Distribution of the number of issues per decade for each author.

Robert Silverberg. The three authors are American writers who published under their own name with the exception of Silverberg who used different pseudonyms. However, no known pseudonyms of Robert Silverberg appear in our dataset. Figure 1 shows examples of two covers from April 1966 and March 1972 issues of *Galaxy Magazine* and two corresponding pages of novels written respectively by Silverberg and Asimov.

We then extracted from the SFGram dataset magazine issues featuring one of these three authors. It resulted in a set of 89 issues containing a novel (of part of it) written by at least one of these author. In addition, we included two issues that do not contain any text written by any of these three authors. The final resulting dataset is composed 91 issues for a total of 8 millions words. We manually segmented each issue to tag parts written by one of our chosen authors.

Table 1 and figure 2 show respectively the number of documents, words, total ratio and the distribution of issues published over the three decades (50s, 60s, 70s) for each of the three authors. Isaac Asimov is present in 22 issues (or documents) out of the total 91 with 341,480 words for a ratio of 4.25% of the whole dataset. Philip K. Dick wrote in 25 issues with 264,504 words for a ratio of 3.26% of the dataset. Robert Silverberg is present in more issues with a total of 45 issues and 911,219 words for a total ratio of 11.25% of the dataset.

Each selected issue has been transformed to raw text through optical character recognition (OCR). Moreover, each issue includes a significant amount of ads and promotions. As a result, the texts are highly noisy, some random characters being inserted, and novels can be interrupted by illustration with caption and unrelated ads.

For example, the documents are filled with ads and wrongly recognised character such as :

ˆ y AMAZING LOW-PRICE OFFER !
on this Mechanics All-Purpose
/V SOCKET WRENCHnfr^^
p/tvC jl,j cgniplete Workshop That
You've Always Wanted !^^

As each document contains different sections such as indexes and table of contents, each issue is the result of a collaborative work of multiple authors. From the SFGram dataset, we know that there is a minimum of 1,308 authors who contributed to these 91 magazines, for an average of about 14 per document or issue. This does not take into account the participation of unknown authors such as publisher, editors, ads writers and readers intervention in the reader's mail.

To evaluate ESNs we used F_1 score which is a well-known measure of a text accuracy used in the statistical analysis of binary classifiers. It is the harmonic average of the precision and recall and equals one where both precision and recall are perfect, and zero when they both null.

Formally, the F_1 score is defined as,

$$F_1 = 2 * \frac{precision * recall}{precision + recall} \tag{1}$$

where *precision* and *recall* are defined respectively by,

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

Here *TP*, *FP* and *FN* refer respectively to *true positive*, *false positive* and *false negative*. In addition to these measures, we used 5-fold cross validation (5-CV) to compute a fair estimation of ESN's performances.

For the first task and for each fold, we trained an ESN with the desired feature and computed the output stream for the validation and test set. We normalized the output so that they have an average

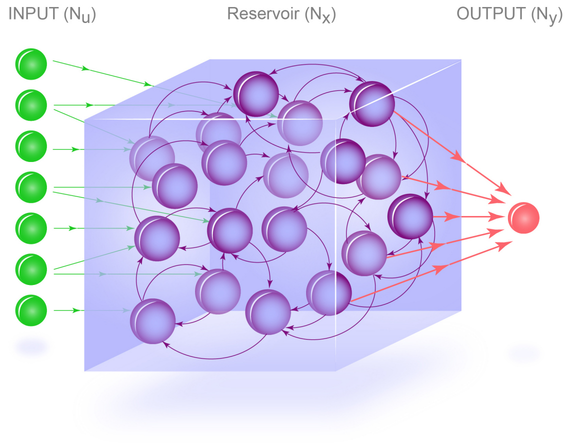

FIGURE 3 – The full reservoir architecture of the *Echo State Network*. The green, purple and red lines represent respectively input connections, internal connections and learnable connections between reservoir's units and outputs. In this study, the input dimension (green dots) is 300 for WV and 60 for C3, the seven green dots being only used to illustrate the ESN architecture.

of zero and a standard deviation of one. We then looked for the best threshold separating points detected as written by the author or not using the validation set. The F_1 score was then computed on the test set. This threshold allow us to separate sections of the stream which the model considers written by the target author.

For the second task, we trained an ESN with the same method as for the first task and computed the output stream for documents contained in the validation set. But in this second case, we looked for a threshold that can separate documents in which the author participated. The classification is then done, not at the token level compared to the first task, but at the document level. This threshold allows us to detect whether the target author collaborated in an issue of not.

4 Echo State Network for Natural Language Processing

4.1 Echo State Networks

Echo State Networks are defined mainly by the following equation :

$$x_t = (1-a)x_{t-1} + a\, f(\overset{in}{W} u_t + W x_{t-1} + \overset{bias}{W}) \tag{4}$$

where x_t is a vector of size N_x (the number of neurons in the reservoir) that represents the highly

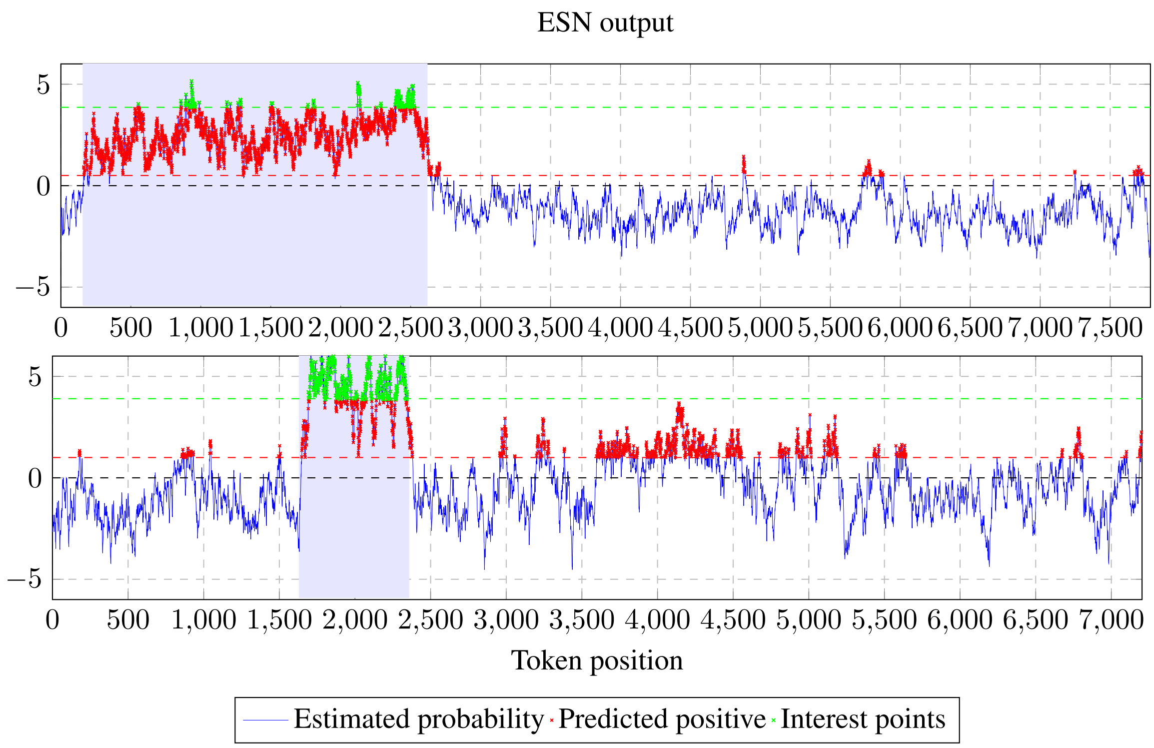

FIGURE 4 – Two examples of output after normalization of an *Echo State Network* with a pre-trained word embedding layer on the author verification task. The blue line shows the estimated authorship probability. The red horizontal dotted line is the threshold and points above it are marked in red. The green points are interest points and the green dotted line is the threshold used to find them.

non-linear reservoir state vector at time t. $\overset{bias}{W}$ and W are respectively biases to the reservoir's units and the matrix of internal weights. $\overset{in}{W}$ represents the input weights applied to u_t, with N_u the dimension of the input signal. a is named the *leak rate* and allows the adaptation of the network's dynamic to the one of the task to learn. Figure 3 shows the complete ESN architecture with inputs, internal connections and outputs. In our study, the reservoir contains 1,000 units ($N_x = 1000$).

The first state x_0 is usually the null state ($x_0 = 0$). The network's output $\hat{y}$ can then be defined as :

$$\hat{y}_t = g(\overset{out}{W} x_t) \tag{5}$$

where the matrix $\overset{out}{W}$ represents the connection between the reservoir's units and the output (with N_y the number of outputs). Here, we have a one-dimensional output, $N_y = 1$, representing the authorship probability at time t. The identity function is usually used as g.

ESN's training consists to solve a system of linear equations to minimise the error $E(Y, \overset{out}{W} X)$ between targets (Y) and outputs ($\overset{out}{W} X$). To compute $\overset{out}{W}$, it is possible to use the well-known ridge regression based on a regularisation factor λ (to minimise the magnitude of output weights).

4.2 Transform text into time series

To use ESNs as a classifier for documents, each text must be first transformed into a time series. In this article, we tested two lexical features based on a word embedding layer (denoted WV) on the one hand, and on the other a pre-trained vector of character trigrams (named C3). For WV, we used Glove vectors of dimension 300 with a vocabulary of 1.1 million words. For each features, the corresponding leak rate value is respectively 0.01 and 0.001.

To test our model with character-based features, we used the character trigrams developed in Schaetti (2019) which was extracted using a bag-of-word model. For this purpose, a feed-forward neural network, with a fully connected layer on top of an embedding layer, and a softmax function as outputs was used to predict a trigram from its surrounding context.

This model was trained on 230 million

Classifier	Asim.	Dick	Silver.	Av.
True	0.08	0.06	0.20	0.11
ESN WV	0.47	**0.73**	0.62	0.61
ESN C3	**0.52**	**0.73**	**0.63**	**0.63**

TABLE 2 – Comparison of F_1 scores (5-fold CV) of ESN models on the task of author verification in streams of text.

examples extracted from Wikipedia. The embedding layer of dimension 60 was finally used as pre-trained vectors for the inputs of our ESN model. To have a fair comparison between word-based an character-based representations, we sub-sample the output obtained with character-based features to have the same length as the character-based outputs.

In figure 3 the input dimension (the number of green dots) is equal to 300 for WV and 60 for C3. During training and classification, we feed the embedding vector from WV or C3 to the ESN as inputs one token at a time.

4.3 Interpreting outputs

The output of the ESN is $\hat{y}_t$, the estimated probability that word-tokens in the ESN's memory at time t have been written by the author. The result is then an output time series of estimated authorship probabilities. Figure 4 shows the output of two issues of *IF Magazine* published respectively in March 1955 and in September 1959. The first contains the novel *War Veteran* and the second *Fair Game* both written by Philip K. Dick. The blue lines show the outputs at time step t (x-axis). The blue areas show the position of the section written by the author. Magazine issues are processed by ESN as text streams and the output is normalised to have mean and variance equal respectively to zero and one. At each time step, the ESN's output represents estimated probabilities that the text currently stored in its memory has been written by the target author.

For the first task, once the output is computed for the whole train dataset, we look for the best threshold which allows to separate the two classes on a validation set. Each position with an output value $\hat{y}_t$ above the threshold was considered part of a section written by the target author. With the final test classification, we computed the F_1 score based on the calibration obtained from the validation set. The red dotted line shows the chosen threshold while the red points represent the position predicted as belonging to the author's section. The training, validation and test sets represent respectively 80%, 10% and 10% of the whole dataset. We used the same principle for the second task (author present in an issue). We define the threshold based on validation set, a threshold that best separate inside an issue where the target author collaborated. In Figure 4, the green dotted line show the threshold used to find the interest points in an issue belonging to the test set. In the current case, the author is considered to have collaborated on that issue.

Classifier	Asim.	Dick	Silver.	Av.
True	0.35	0.36	0.65	0.47
Linear SVR word 3gram	0.35	0.36	0.60	0.44
ESN WV	0.18	0.80	**0.75**	0.58
ESN C3	**0.64**	**0.85**	**0.75**	**0.75**

TABLE 3 – Comparison of F_1 scores (10-fold CV) of ESN models on the task of author verification at the document-level using interest points.

5 Results

Table 2 shows the average F_1 score using 5-fold cross validation (CV) according to the three authors, for word and character-based ESN models and the baseline, for the first task. The baseline is a simple classifier predicting always yes for all points in the stream. For Isaac Asimov, the best F_1 scores are reached by an ESN based on pre-trained vectors of character trigrams (ESN C3) with 0.52. For Philip K. Dick, the best F_1 score is reached by both the word-based and the character-based ESN with 0.73.

The ESN based on Glove pre-trained word vectors (ESN WV) stay behind with respectively 0.47 and 0.62 for Asimov and Silverberg. However, Isaac Asimov seems harder to identify than Philip K. Dick for both models despite a larger training set. The reason for this greater difficulty remains to be identified in future research. As possible explanation, we can assume that Asimov might write with different styles while Dick tends to reuse the same stylistic construction.

For Robert Silverberg, the best F_1 score is achieved by an ESN based on character trigrams embedding with 0.63. Word-based ESN stay just below with a F_1 score of 0.62. To determine which model performs best, we computed the average F_1 score over the three authors. ESN-C3 surpasses the word-based ESN with an average F_1 score of 0.63 against 0.61 respectively.

For the second task (see Table 3), we asked models to determine whether an author's work is present in a document or not. To compare their effectiveness, we added two models as baseline. The first is a simple classifier predicting always yes for all documents (row “True" in Table 3), and the second is a linear Support Vector Regression (SVR) model based on word trigrams (second row in Table 3).

The learning stage of this model is based on a training set where the label indicates the probability that the target author participated to the given issue. This probability is fixed to one if the author is present, zero otherwise. Once trained, we used the same procedure as for ESNs by determining the best threshold to separate both classes.

For Isaac Asimov, the best F_1 score is achieved by a character-based ESN (ESN C3) with 0.64 against 0.18 for word-based ESN (ESN WV). The baseline composed of the true-classifier or based on SVR achieved 0.35. For Philip K. Dick, the best F_1 score is achieved by a character-based ESN with 0.85. The word-based ESN, the linear SVR and the true-classifier got an F_1 score of 0.80, 0.36 and 0.36.

On Silverberg's novels, the best score is achievied by both word-based and character-based ESNs with a F_1 score of 0.75, against 0.60 and 0.65 for respectively the linear SVR and the true-classifier.

The best average F_1 score is achieved by the character-based ESN with 0.75 against 0.58, 0.44 and 0.47 for the word-based ESN, the linear SVR and the true-classifier (see Table 3). The ESN models are the only ones to do better than random on this second task, unlike SVRs which is no better than the true classifier (0.47). In addition, unlike the baseline, ESNs can give the exact position in the text that is considered to be the work of the target author. Of three authors, Philip K. Dick is the easier to detect at the document-level, a surprising result as the training set for Robert Silverberg is much bigger. The character-based ESN even reaches a F_1 of 0.85 on Dick's novels.

This ease to identify Philip K. Dick is an interesting question for future research. Does this author have a particular single style or a specific vocabulary ? In order to have a look at how the model deals with different authors, we extracted a piece of the text of each author with a high estimated probability ($\hat{y}$). The resulting is shown in Table 4.

For Isaac Asimov, the text is extracted from the novel "*The Gods Themselves*" published in the issue of March 1972 of *Galaxy Magazine*. The novel depicts the story of a scientist finding out that a sample of tungsten has been transformed into plutonium 186. This leads to the development of a endless and clean source of energy. The novel is strewn with technological and scientific terms as is often the case in Asimov's novels.

For Philip K. Dick, the text is extracted from the novel *Exhibit Piece* published in the issue of August 1954 in *IF Magazine*. The story told in this novel addresses a common theme in Dick's subsequent works : the concept of shifting realities and time travel.

For Robert Silverberg, the text is extracted from the novel *A Time of Changes* of the issue of May 1971 of *IF Magazine*. It won the Nebula Award that year, the equivalent of the Emmy Awards in science-fiction. It tells the story of a world where words such as *me* and *I* are forbidden and is written from an autobiographical point of view. The hero is telling his own story.

6 Discussions and Conclusion

In this study, we investigated the effectiveness of recurrent neural models referred as Echo State Networks (ESN) to perform authorship verification in noisy text streams. The underlying test-collection was extracted from science-fiction magazines published during 1952 and 1974. We selected three well-known authors and tested word and character-based features with F_1 score as evaluation measure. This study shows that ESNs are able to identify authors in documents resulting from noisy OCR and from the collaboration of tens of authors, where other methods are not better than a random classifier.

Various questions stay without a clear and definitive answer. First, we would like to evaluate recurrent neural networks on a bigger set of authors, especially those using pseudonyms to determine if the proposed models are able to undercover pseudonymous work and are reliable with a higher number of possible authors. For example, the science-fiction writer Randall Garrett is known

Author	Examples
Isaac Asimov	The Hard One to whom Tritt had spoken was agreeing - the other still exuded concern. Dua was looking at Tritt. The first Hard One said, "Where is the food-ball now, Tritt ?" Tritt showed them. It was hidden effectively and the connections were clumsy but serviceable. *The Gods Themselves*, March 1972.
Philip K. Dick	Miller straightened his collar and bright hand-painted necktie. He smoothed down his blue pinstripe coat, expertly lit a pipeful of two-century-old tobacco, and returned to his spools. *Exhibit Piece*, August 1954.
Robert Silverberg	All about me were the things of the gods, and 1 failed to detect the divine presence. Perhaps Schweiz had found the godhood through the souls of other men, but I, dabbling in selfbaring, somehow had lost that other faith and it did not matter to me. *A Time of Changes*, May 1971.

TABLE 4 – Example of a text classified as very probable for each author.

to have used more that ten pseudonyms to publish science-fiction novels. We plan to use this kind of textual data to evaluate the possibility to identify interest points in novels published under pseudonym that could prove the collaboration of an author.

On the model side, we would like to evaluate other recurrent neural models able to handle temporal data such as Long-Short Term Memory (LSTM) and Gated Recurrent Units (GRU). The SFGram dataset fits well the use of this kind of model based on deep learning as it contains thousands of documents and tens of millions of words. We also plan to test deeper models such as stacked-ESN and bidirectional RNNs. We could then use the work presented in this study as a baseline for further investigation.

Other investigations are possible on the use of new additional textual for text representations such as POS tags or sequence of such tags or based on different noun or verb phrases (e.g., adverbial phrase of time, phase of manner, purpose phrase, etc.) or their position inside sentences. These information could be used in combination to the lexical features as used in this study.

We think that this study shows that ESN are interesting models to handle streams of textual data. Consequently, we would like to evaluate these models to detect events in text streams coming from social medias such as Twitter.

References

Shlomo Argamon and Patrick Juola. 2011. Overview of the international authorship identification competition at pan-2011. In *CLEF (Notebook Papers/Labs/Workshop)*.

Paulin Coulibaly. 2010. Reservoir computing approach to great lakes water level forecasting. *Journal of hydrology* 381(1) :76–88.

Hugo Jair Escalante, Manuel Montes, and Luis Villaseñor. 2009. Particle swarm model selection for authorship verification. In *Iberoamerican Congress on Pattern Recognition*. Springer, pages 563–570.

Patrick Juola and Efstathios Stamatatos. 2013. Overview of the author identification task at pan 2013. In *CLEF (Working Notes)*.

Moshe Koppel, Jonathan Schler, and Elisheva Bonchek-Dokow. 2007. Measuring differentiability : Unmasking pseudonymous authors. *Journal of Machine Learning Research* 8(Jun) :1261–1276.

Moshe Koppel and Yaron Winter. 2014. Determining if two documents are written by the same author. *Journal of the Association for Information Science and Technology* 65(1) :178–187.

John Olsson and June Luchjenbroers. 2013. *Forensic linguistics*. A&C Black.

Chen Qian, Tianchang He, and Rao Zhang. 2017. Deep learning based authorship identification .

Nils Schaetti. 2018. Bidirectional echo state network-based reservoir computing for cross-domain authorship attribution : Notebook for PAN at CLEF 2018. *CLEF2018* .

Nils Schaetti. 2019. Behaviors of deep reservoir computing models for textual documents classification. *The 2019 International Joint Conference on Neural Networks (IJCNN)* .

Nils Schaetti, Raphaël Couturier, and Michel Salomon. 2015. Reservoir computing : Étude théorique et pratique en reconnaissance de chiffres manuscrits, mémoire de master .

Nils Schaetti, Michel Salomon, and Raphaël Couturier. 2016. Echo state networks-based reservoir computing for mnist handwritten digits recognition. *19th IEEE International Conference on Computational Science and Engineering (CSE 2016)* .

Nils Schaetti and Jacques Savoy. 2018. Comparison of neural models for gender profiling. In *Proceedings of the 14th international conference on statistical analysis of textual data*.

Efstathios Stamatatos, Nikos Fakotakis, and George Kokkinakis. 2000. Automatic text categorization in terms of genre and author. *Computational linguistics* 26(4) :471–495.

Hans Van Halteren. 2004. Linguistic profiling for author recognition and verification. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, page 199.

LZ Wang. 2017. News authorship identification with deep learning.

Francis Wyffels and Benjamin Schrauwen. 2010. A comparative study of reservoir computing strategies for monthly time series prediction. *Neurocomputing* 73(10) :1958–1964.