A Literature Survey on Writing Style Change Detection Based on Machine Learning: State-Of- The -Art-Review

Article *in* International Journal of Computer Trends and Technology · May 2022

DOI: 10.14445/22312803/IJCTT-V70I5P103

3 authors, including:

Vivian Oloo
Maseno University
2 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Lilian Wanzare
Maseno University
14 PUBLICATIONS **71** CITATIONS
SEE PROFILE

International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V70I5P103

Volume 70 Issue 5, 15-32, May 2022

Review Article

A Literature Survey on Writing Style Change Detection Based on Machine Learning: State- Of-The -Art- Review

Vivian Anyango Oloo[1], Calvins Otieno[2], Lilian Awuor Wanzare[3]

[1,2,3]*Department of Computer Science, Maseno University, Kenya.*

Received: 16 March 2022 Revised: 07 May 2022 Accepted: 12 May 2022 Published: 28 May 2022

Abstract *- The goal of the Style Change Detection task is to detect the stylistic changes in a document and exploit them to determine the number of authors. This study reviewed nineteen (19) state of the art papers and articles on writing style change detection. The papers were identified and selected based on study area, year of publication and the technique proposed for writing style change detection. The focus of this study was to investigate the features used, the techniques and the results obtained by these state of the art studies. Three categories were defined and all papers placed in one of the groups based on the problem it was solving. The study found out that the most commonly used feature category was the lexical features although using feature combinations yields better results. In addition, simple distance measures were shown to outperform other state-of-the art techniques in authorship clustering and style change detection. The use of ensembles of algorithms is recommended for style change detection tasks when the text length is short and the dataset is large.*

Keywords *- Authorship, Clustering algorithms, Multiple authorship, Stylometry, Style change detection.*

1. Introduction

Writing style change detection is a branch of authorship verification focussing on the examination of a document for the different authorial style. The ultimate goal of writing style change detection is identifying the exact number of authors collaborating in writing a text or document [35,43]. Writing style change detection is important because it can be used to determine the number of people behind blog posts especially for renowned and popular people. In addition, determining the exact number of authors participating in anonymously written text from the internet and other sources may be important for forensic investigators [34]. Determining the authors of anonymously written texts has been the focus of authorship verification studies. However, where it is suspected that more than one author wrote the text, the goal is to determine how many writers participated and identify who wrote what portions of texts [3,26].

Focusing on the style change detection, [1]PAN Evaluation Laboratory has organized different competitions requiring participants to judge whether one or more authors wrote a given text; to find out the writing style changes for the multi-author documents, and the need to label the author identifiers. However, determining the exact number of authors in a multi-authored document involves the most critical and challenging task of the writing style change detection tasks [11, 21, 46].

The application areas of writing style change detection range from plagiarism detection, cyber security and forensics and currently in fake news detection [7,29,46]. It has been shown by multiple researchers how word patterns can be used to identify authorial styles and the existence of multiple consistent personal styles indicating the presence of multiple authors [5,6,20]. Endeavors to detect changes in writing styles have been done under author diarization or clustering, and style change detection [31,35,43,44]. This paper takes interest in style change detection.

2. Background Information

Writing style change detection aims at determining the number of authors in a multi-authored document by studying the similarities in styles of writing. This branch of authorship verification has been understudied in literature, although the competitions organized by the PAN clef Laboratories have helped to expand research in this area. Writing style change detection presents two main scenarios; one document one author and one document several authors. In the first scenario, several documents are lumped together and the task is to group together documents written by one author. Although this scenario can be seen as a conventional authorship verification task because it examines whole documents for stylistic similarities, it forms the basis of writing style change detection

[1] PAN is a series of scientific events and shared tasks on digital text forensics and stylometry

which examines a single document for authorial differences. These studies fall under the branch of writing style change detection known as authorship diarization and clustering.

In a single document with multiple authorship scenarios, the task is to determine the number of authors participating in writing the document. Multiple authorship may present itself as in a document having one main author and several small authors, or as one with many small authors. The first scenario where a document has one main author and several small authors has been studied under plagiarism detection, since it is assumed that there is one main author who writes the document and a few pieces of texts from known authors. This scenario of one main author as applied in plagiarism detection has been adapted by some studies [7] to solve the problem of determining the number of authors in multi-author documents although with certain limitations.

The second scenario is the case of a single multi-authored document with several small authors randomly distributed throughout the document [43]. Here, the task is to identify the different writing styles presented in the text, which is representative of the number of authors in a document. Clustering algorithms have been used to solve this problem. The idea is to subdivide the document into sections; sentences, sentence groups or paragraphs, and to generate feature vectors for the various sections. Similarity functions are then used to determine the similarities between the feature vectors and to place the sections into clusters based on the similarity scores with the rest. It is assumed that a cluster contains works of the same author.

Writing style change detection studies have been investigated with both long and short documents, although most studies focus on short text lengths. These studies adapt themselves to the real-life scenario where an author may contribute very short texts in form of sentences or paragraphs [7,29,41]. Whereas most studies focus on short documents, [3] used long documents to determine the number of authors in multi-authored documents. Moreover, authorship clustering and diarization studies are based on longer documents compared to the rest of the other studies in writing style change detection [25, 28,37].

Existing studies have been evaluated with a known or unknown number of authors. In the known authorship case, studies focus on confirming whether the results of the proposed approaches march with what is provided. In other words, there exists labeled data and the task is to counter check the results of the proposed method vis-a-vis the ground truth. Supervised learning methods which work best with labeled data, yield better performance in such cases [5, 6]. In the other scenario, ground truth information is not provided during testing, and the task is to cluster together documents written by the same author. Here, the number of authors is determined by the resultant number of distinct clusters [4,13, 22, 26, 41].

Among the pioneer studies to establish the number of authors in a multi-authored document was the study by [3] which used unsupervised learning to determine the number of authors in a multi-authored document using 500 most common words as the style marker. Most studies in writing style change detection have been a result of the annual competitions by PAN. PAN competitions on writing style change detection are organized annually with increasing complexity. For instance, in 2016 the task was to group together documents written by one author [35]. In 2017 the task was broadened to include linking various parts of a document written by one author, and to find the borders of authorship change. The results posted by the proposed approaches for the 2017 task were below the baseline defined for the task, indicating that the task was difficult [42]. Hence in 2018 the task was relaxed to determining whether a document is single or multi-authored. The preceding years such as 2019, 2020, and 2021 exhibit increasing complexities in terms of the reduction of the document length, training dataset size and the task to be performed [43,44]

The rest of the document is organized as follows; section III outlines the problem statement. Section IV presents the methodology used. The different stylometric features are discussed in section V. Section VI describes the methods used in writing style change detection. Section VII outlines evaluation techniques while section VIII gives the results. Discussion of results is done in section IX, while section X outlines conclusion. Section XI recommendations.

3. Problem Statement

Whereas research in writing style change detection is now the focus of most authorship verification studies because of their varied application, an exhaustive review of the various studies in this field is still lacking. The few review studies that exist focus on the broad area of authorship verification and simply mentions writing style change detection with little emphasis. However, given the importance of this area of study in resolving such cases as in fake news identification and cyber security and forensics, there is a need for a comprehensive review of the proposed methodologies and techniques and the features which have been investigated. Moreover, the evaluation techniques applied and the results obtained is key if these studies are to be fully accepted and deployed. The purpose of this paper is to present a survey of the studies and the techniques used in solving writing style change detection problems together with the most common features used.

4. Methodology

This study reviewed papers and articles published in refereed journals. A systematic way of identifying the most appropriate papers and articles was adopted. The criteria used include the year of publication, area of study and the techniques used. Nineteen papers and articles were identified

using the above criteria, covering authorship clustering and diarization and style change detection techniques. These areas were considered because they focus on determining the change in writing styles within a document, which forms the basis of writing style change detection.

Priority was given to articles published from the year 2016 onwards with the exception of one paper published in 2012. This was the only paper found focussing on determining the number of authors in a multi-author document, done on a different dataset. All the other studies were based on annual PAN competitions using the PAN datasets. The selected papers were reviewed for the stylometric features used for the various tasks, the techniques employed and the results obtained. For ease of results presentations and discussion, we grouped the papers according to the study areas, where three categories were obtained as; authorship clustering, plagiarism detection and style change detection. Studies focussing on grouping documents written by one author together were grouped under authorship clustering, while style change detection group consisted of papers focussing on determining the number of authors in a document.

5. Stylometric Features

Stylometric features have been used to define writing styles by examining characteristics that are persistent through all the works of an author [5,6,20]. Examining the similarities in writing styles presented in a document can be used to determine the number of authors participating in writing the document. Previous studies report a rich set of stylometric features which are applicable in authorship verification studies with different results. For instance, studies in [5,6,20] report over a thousand stylometric features and categorize them into five groups namely; lexical, structural, syntactic, character and content features. However new features continue to be discovered and used with varying effects on performance [2,5,6,8,20,30]. A comprehensive survey of the different categories of stylometric features used in writing style change detection is provided as follows:

5.1. Lexical Features

Lexical features can be used to model author preferences of choice of certain words or character sequences in all their works [3,5,12]. These features are the most commonly used features in previous studies because they can be applied across languages at no extra cost. Most previous studies employed the analysis of lexical features to detect changes in writing styles within documents [3,4,26,37]. Lexical features can be defined at the word level or sentence level. In addition word level statistics such as word length, total number of words, average word length, most frequent words, word pair frequencies, duplicate words, type token ratios have also been investigated. word level features include features such as word n-grams, vocabulary richness, word frequencies, POS words, stop words and word unigrams among others. At the sentence level, lexical features may include repeated sentences, misspellings, sentence length, average sentence length etc.

Previous studies have employed word level features such as word n-grams, vocabulary richness, word frequencies, POS words, stop words, word pair frequencies and word unigram, most frequent punctuations symbols, among others [23,26,37,39,42]. For instance, [23] used fastText word embeddings, deep LSTM and triplet loss to propose a system that is able to learn stylometric embeddings of different documents and measure their stylistic distance. [29] used a Siamese Neural Network on vocabulary richness to compute paragraph similarities for the detection of style changes.

Various word-level statistics have also been used in writing style change detection such as word length, total number of words, average word length, most frequent words, word pair frequencies, etc. [3,11,21]. Some studies have also investigated the applicability of other lexical features such as repeated sentences, duplicate words, misspelling etc, with improved performance.

At the sentence level, studies investigate the use of sentence level statistics such as sentence length, mean sentence length, repeated sentences, misspellings among others [12,21,28,37].
For instance, [30] used a combination of features for the writing style change detection tasks yielding very promising results. However, the study reported a significant performance improvement when duplicate sentences were used. [40] used a ClustDist anomaly detection technique on 15 lexical features to generate a feature vector containing average distances of all sentences from each other.

Combinations of the various lexical feature types have also been used by yet some studies. [9] used mean sentence length in words, mean word length or corrected type-token ratio, and pre-trained FastText embeddings, with multi-layer perceptrons and bidirectional LSTMs for the style change detection task. [41] used two methods to extract features based on Google AI's BERT transformer for generating textual embeddings and extracting textual features and statistics.

The use of lexical features in most previous studies can be attributed to the fact that they provide a good measure of the stylistic differences that are quantifiable into a writing style [2,6,7,10,17,41]. This category of features continues to give promising results not only in the general authorship verification problems, but also in other specialized tasks such as authorship clustering, writing style change detection and change of writing styles with time [26,23,29,42]. Whereas lexical features are the most commonly used features in writing style change detection [3,7,29], the purity of the models based only on these features is still debatable because they are topic dependent and may carry with it effects of topic, genre and domain [1,20,35].

5.2. Character Features

Character features are used to capture variations in lexical information cues of contextual information of capitalization and punctuations [5]. They include the use of character features such as character n-grams, total number of characters, total number of digits, total number of uppercase letters, total number of space characters, and number of tabs and their respective ratios have been used in writing style change detection studies [32,39]. For instance, [21] proposed a method for author clustering and style breach detection based on local-sensitive hashing-based clustering using a bag of n-grams and other stylometric features. A few univariate studies investigated the use of n-grams on authorship verification. For instance, [34] analyzed the effect of n-grams and ensemble of supervised learning algorithms. Better still, [4] investigated the use of different character n-grams and reported that character 2-grams achieve best performance with the top 300 most frequent features. Another study conducted by [18] investigated the author verification using common N-gram profiles of text documents on the PAN 13 dataset.

character frequencies or a measure of character statistics such as most frequent characters n-grams, most frequent punctuations, special character frequencies etc have also been used. [26] used a simple unsupervised author clustering and authorship linking model called SPATIUM on most frequent terms (isolated words and punctuation symbols), and most frequent character n-grams of each text.

Whereas character features form good candidates for authorship verification tasks, and particularly writing style change detection. These features are tolerant to noise from the texts such as grammatical errors, which have been used in some studies to represent the author's traits in the style based categorization of text [35,42]. However, they may not effectively capture stylistic differences in documents with short lengths. Therefore, building an ensemble of these features together with other features may help yield better accuracies in writing style change detection [7,21,23,30,42].

5.3. Syntactic Features

Syntactic features are the only trusted measure of stylistic differences between works of the same or different authors. These features provide a better representation of writing styles in a much easier way because they can be normalized and quantified [5,6]. The most commonly used syntactic features are function words and common words such as nouns, pronouns, prepositions, etc. however, stop words, language parse tree and other syntactic style markers have also been used. Most previous studies have not used syntactic features on their own in writing style change detection, however they have been used together with other features to measure the stylistic differences in documents. In comparison to lexical features, few studies focus on the use of syntactic features partly because of the language dependencies of these features, and the need for a syntactic parser to process specific natural languages [38].

Stop words which form the bulk of the words in any document have been investigated as a measure of authorial styles. For instance, [21] proposed a method for author clustering and style breach detection based on local-sensitive hashing-based clustering using stop word and other stylometric features. In another study, [23] used stop words and other basic stylometric features to develop a comparison model for detecting stylistic changes within a document.

Function words have also been used by some studies. Since there are so many function words that can be applied in writing style change detection, literature confirms that the commonly used function words are between 150 and 675[1,4,7]. For instance, [7] **u**sed characters, lexical and syntactic style markers to build a paragraph representation to establish the number of writers of a document and the corresponding paragraphs authored by each.

Language parse trees have also been used by a few studies. [16] proposed a parallel hierarchical attention network to establish whether a document is multi-authored or not. In this approach, the feature set involved the parse tree features extracted from the tree-based structure of a sentence in order to preserve word order in a sentence. **[41]** used a stacking ensemble of classifiers trained on separately extracted features and BERT embeddings, and combined their predictions by a meta-learner, i.e the stacking ensemble. Other syntactic style markers have also been investigated. For instance, [7] **u**sed characters, lexical and syntactic style markers to build a paragraph representation to establish the number of writers of a document and the corresponding paragraphs authored by each.

The success of using syntactic features depends on the availability and use of a syntactic parser which can process specific natural languages with good accuracy, which is an expensive venture in writing style change detection. However, they are believed to provide the best authorial fingerprinting [4,5,7,35].

5.4. Structural Features

Structural features learn the document organization of different authors and may be used to verify the stylistic differences based on the way every author organizes his/her document. Few studies have investigated the effectiveness of these features on writing style change detection [16,30]. In addition, structural features are not common features even in the general authorship verification studies although they have been used by few studies to determine authorship attribution of emails [6]. Structural features are not good candidates for writing styles change detection when used on their own since they may not provide acceptable stylistic differences between different works.

5.5. Context Features

Context features are those features that signalize the existence of particular key words, interest groups and given activities [10]. They are specific and provide contextual information of the task at hand. For instance, [19] manually observed, analyzed texts written in ancient times and identified some key words in the context of online sales environments such as obbo, windows, hashtags, etc [1,22,31,41]. Few studies use content-based features to detect changes in writing styles within documents. A study by [34] studied the authorship identification of documents with high content similarity. They focused on analyzing how humans judge different writing styles, based on content-agnostic characteristics of authors.

5.6. Combinations of different feature categories

Some studies employ the use of feature combinations to the writing style change detection problem. These studies use a mixture of features cutting across lexical, syntactic, structural and content-based features. Combining features is deemed to have the advantage over small feature sets in writing style change detection involving very short document lengths and larger datasets [30,36]. For instance, [23] proposed a comparison model to detect stylistic changes within a document that answers the question whether or not a document was written by many authors. Their approach was based on the analysis of basic stylometric features such as word frequencies (stop words and other POS words), punctuations, word pair frequencies and POS pair frequencies. [11] used a bag of words on different features on a B-compact graph-based clustering to determine authorship clusters.

[16] proposed a parallel hierarchical attention network to establish whether a document is multi-authored or not. In this approach, the feature set involved the parse tree features extracted from the tree-based structure of a sentence in order to preserve word order in a sentence. Another study [46] used various combinations of stylometric features and an ensemble of clustering algorithms to cluster segments into groups in a multi-authored document. [7] **u**sed characters, lexical and syntactic style markers to build a paragraph representation to establish the number of writers of a document and the corresponding paragraphs authored by each. [21] used a mixture of stylometric features and a bag of n-grams. Tf-idf features and the Wilcoxon signed rank test were computed to determine the style breaches. The use of ensembles of features is recommended to improve the accuracies of the stylometry-based models, and in writing style change detection. This is because of the short document lengths and larger datasets, which may need more than one feature category to satisfactorily discriminate the styles of different authors [5].

6. Methods for Writing Style Change Detection

Writing style change detection began with the simple task of grouping together documents written by one author known as author clustering [35], and author diarization which groups together sections of a document with the same writing style [25,28,37]. Author clustering assumes that an entire document has a single author and exploits the stylistic similarities and differences to group documents with the same writing styles together. Author clustering can be seen as an adaptation of the conventional authorship verification which examines if two documents exhibit similar writing styles [1,5]. Author diarization on the other hand breaks a document into homogenous sections representing similar writing styles. These basic multi-author analysis tasks form the basis of the writing style change detection.

Pioneer studies in the multi-author analysis assume that a document has one main author who writes about 70% of the document and the other authors contributing the rest of the sections. In addition, the first few paragraphs are assumed to be written by the main author. In this scenario, a document is broken down into sentence groups or paragraphs, and a pair of paragraphs/sentence groups compared with each other to determine their similarities. These initial endeavors have been researched further to include the task of style change detection with the basic aim of checking whether a document is single or multi-authored and to determine the borders where authorship changes in multi-authored documents [37,42]. Further explorations on writing style change detection include determining the number of authors in collaborative documents [43], and identifying whether there is style change between consecutive paragraphs. Other tasks of writing style change detection include finding all positions of writing style change detection within a multi-authored document and assigning all paragraphs of the text uniquely to some author out of the assumed number of authors in the document [31,44,45].

Different scenarios can be defined with multi-authored documents; Firstly the case of one main author and several small authors. Since there is one main author contributing a huge portion of the texts in the documents, studies have employed the use of outlier and anomaly detection methods, and hashing-based clustering to determine the number of authors in the document as in style breach detection and intrinsic plagiarism detection [21,28]. Here the task is to find sections of the multi-authored documents which are not written by the main author and to label them as either ‘plagiarized’ or ‘outlier’. The second scenario is the case of several small authors contributing texts randomly in the document. The task in this case is to determine the total number of authors by determining the similarities in the texts such as in paragraphs, sentences or sentence groups [37,42,36]. This task can be challenging if many small authors are contributing relatively short texts, due to similarity overlap [5,6,7]. Attempts to solve this challenge include the use of

clustering algorithms that groups together texts written by the same author. It is believed that a cluster contains only text written by the same author. Determining the optimal value of k (clusters) is the main challenge for these methods [46].

Writing style change detection is based on generation of feature vectors to be used to discriminate or group together documents. Feature vectors can be generated at the document level as in the case of author clustering. Document level feature generation is used in a case where different documents are to be grouped together or compared for similarities [28,37]. Since there is sufficient data, reduced feature sets tend to yield better results in terms of runtime [4,25,32]. Sentence level feature generation can also be used in writing style change detection particularly in multi-authored documents. Feature vectors generated at the sentence level may result in higher purity because they may capture all the stylistic changes within a document including very short text contributions by other authors which may however be ignored [28,37]. The main challenge with this method is that the feature set should be expanded so as to adequately represent an author's writing style. Other studies combine a number of sentences together to form sentence groups, and generate feature vectors based on these groups. This may be seen as the most probable approach as it may provide a sizable amount of data for the style change detection task. However, it's limited since it may ignore very short text contributions made by other authors, such as a sentence contributed by another author leading to reduced reliability of writing style change detection methods [6,20].
Several methods have been proposed to solve the problem of writing style change detection and this paper discusses the different methods under author diarization and clustering and style change detection as indicated below.

6.1. Author diarization/clustering

Author clustering aims to identify and group documents written by the same authors together while author diarization identifies parts of a multi-authored document written by the same author [35]. Simple supervised learning methods such as decision trees have been used to generate feature vectors where labeled data is available [28], while unsupervised learning methods such as k-means are applicable in cases where only unlabeled data is available [40]. To solve the clustering task, feature vectors are generated at the document level so that similarities between document pairs are determined for placement in various clusters. Author diarization on the other hand generates features at either the sentence level or sentence group level [25,28,37]. For author diarization, feature generation at the sentence can be considered ideal since it may take care of even very short text contributions by other authors thereby improving the purity of these methods [33,37]. However, this method may require the use of various combinations of features to be able to distinguish between works of different authors. Paragraph level feature vector generation seems to be practical as it can be assumed that a new author in a multi-authored document may have to contribute a number of sentences summing up to a paragraph for him/her to put across his/her train of thought [25,28].

Several stylometric features types can be used in author diarization and clustering. Most studies employ the use of feature combinations such as lexical, syntactic and character features to analyze the variance in the styles of writing by different authors [5,6,28,37]. In literature stylometric features such as vocabulary richness, word frequencies, sentence length in characters, mean sentence length, average word length, total number of words, ratio of interrogative sentences, character count, digits count, uppercase letters count, spaces count, tabs count, ratio of uppercase letters, ratios of spaces, ratios of tabs, frequent punctuations and Part of speech tags, function words, stop words, spelling mistakes, have been in author diarization and clustering [25,28]. Feature combinations have been shown to produce better results when the text length is short as in author diarization [6,20]. In addition, it has been shown that these features produce the best results in most authorship analysis studies. For instance, lexical features are tokenizable and can be quantified to an author's writing style, while character features can be applied where text length is short. Syntactic features on the other hand are seen as the best feature type as the same attributes are applied subconsciously by a user throughout their writing [1,2,36].

Once the feature vectors have been generated, distance measures are then used to place documents or text in clusters. The idea is to form different clusters representative of the number of authors, and place each document/segment into exactly one cluster [35]. The distance measures are used to calculate the inter-cluster and/or intra-cluster distances for similarities and differences based on a predetermined threshold [4,11,13]. Documents are placed in a cluster if the distance between it and other documents in the cluster does not exceed a predefined value. Several distance measures have been proposed for the authorship clustering and diarization studies. For instance, [25] used a simple distance measure called SPATIUM-L1 based on the L1-norm that has been used to cluster documents and pieces of text together. SPATIUM-L1 calculates the distance between a pair of sentences and places them in the same cluster if the threshold value is not exceeded. Other studies also proposed the use of a cluster distance approach which they referred to as CLUSTDIST. The CLUSTDIST approach calculates the average distance of one portion of text to all other pieces of text, and places the portion in a different cluster if its distance from the other portions is greater than the average distance of all the texts in that cluster [40]. Although the distance measures used in literature are simple, they yield comparable results to state-of-the art methods. For instance, [25] developed an unsupervised technique with a simple distance measure called SPATIUM-L1 based on the L1 norm, to cluster the documents and pieces of texts written by the same author together.

Author diarization and clustering can be solved by using a number of methods; outlier and anomaly detection techniques proposed by [28] and [40]. These methods rely on the assumption that one author writes the better part of the document, upto 70%, and the rest of the document is written by several authors who contribute short texts. In addition, the first few paragraphs in the document are contributed by the main author [35]. These methods generate a feature vector containing the average distances of all groups of texts from each other. The distance between a pair of feature vectors generated at the document or sentence level is calculated to see its deviation from other sentences or documents. For instance [40] used a ClustDist anomaly detection technique on 15 lexical features to generate a feature vector containing average distances of all sentences from each other. The ClustDist method computes the distances between any pair of vectors. The resultant score for each sentence distance from others, generates a ranking which describes the deviation of a sentence from other sentences in the given document.

On the other hand threshold-based outlier detection methods which are based on detecting outliers in an authors' style statistics have been investigated by some studies for their effectiveness in authorship clustering [28]. Here the focus is identifying segments in the document which are not written by the main author [31,33]. For instance, [28] proposed an intrinsic plagiarism detection approach based on gradient boosting regression trees with optimal parameters set at n-estimators= 200 and max-depth = 4. This model is based on threshold-based outlier detection for detecting outliers in an author's style statistics to provide the label "plagiarized" to the outliers.

6.2. Style change Detection

Style change detection is the act of examining a document to identify the different styles of writing present in it [35]. The ultimate goal of style change detection is to determine the number of authors in a document and the various parts of the documents each author has contributed [9,40]. Research in this area is still slow because of the limited benchmark datasets and the limitations of machine learning algorithms on short length text [5,6,7]. However, annual PAN competitions have contributed immensely to the growth of research in this area by providing benchmark datasets and defining tasks to be solved for the style change detection problem. Pioneer studies in style change detection focused on determining the number of authors in a document where it is believed that an author writes a considerably big chunk of text, as in a book chapter or a rather large section of a document [3]. In such cases there is sufficient data for the model to generate feature vectors to discriminate between the works of different authors. However, such studies ignore the contributions of other authors who might have written just a sentence or a paragraph within the document.

State of the art studies are based on reducing text length to identify the change in style in paragraphs, sentence groups or even in sentences although studies determining style changes in a sentence are rare [7,9,37]. The fundamental task in style change detection can be considered as the task of separating single authored from multi-authored documents. It involves examining a document for possible style changes; the existence of style change signifies multiple authorship while the lack of it indicates the presence of only one author [22]. The other tasks of style change detection include finding positions in which authorship changes in a multi-authored document, determining the number of authors in a multi-authored document, and assigning each section of a document to an author. Solutions to these tasks have been systematically sought with increasing complexities. For instance, the first attempt to solve the problem of style change detection sought to determine whether a document is single authored or multi-authored, and for each multi-authored document, determine the position of authorship switches [42]. The proposed approaches yielded poor results which did not meet defined performance baseline defined for the task, hence proving that it was a difficult task. In the following year, the style change detection task was broken down to the fundamental task of style change detection, and the preceding tasks thereafter defined with increasing complexities by combining two or more tasks; a previous successful task and a new more difficult task [23,30]. The style change detection methods are further categorized based on the main task it sought to solve as below.

6.2.1. Determining whether a document is single-authored or multi-authored

Different methods have been proposed by existing studies to solve this task. Some of these methods rely on the analysis of the different stylometric features to detect stylistic changes in a document [16,34], while others adapted the outlier detection methods used in plagiarism detection problems. In addition some studies investigated the use of hierarchical attention networks to solve this problem [16,23,34]. For instance [23] used comparison models on various stylometric features such as word frequencies of stop words and other POS words, punctuations, word pair frequencies and POS pair frequencies. The document is first segmented to various sentence groups and a stylometric match score calculated to check for style changes. The final document score is the sum of the various scores obtained from the sentence groups. This method yields good runtime although it does not produce good accuracy.

[16] proposed a parallel hierarchical attention network to establish whether a document is multi-authored or not. In this approach, the feature set involved the parse tree features extracted from the tree-based structure of a sentence in order to preserve word order in a sentence. To determine style changes in documents, a fusion layer consisting of several similarity functions is used to compute the similarity/differences between the pair of documents. Specifically, they use the weighted vector and its reverse version in the comparison and to check for the existence of

style changes in documents. While [36] approached the style breach detection task by applying a sentence outlier detection commonly used in intrinsic plagiarism detection method.

Although this approach achieves promising results, it took too long to run because of the PTFs whose production is very slow, especially on Stanford stand-alone parser.

6.2.2. Determining whether a document is single or multi-authored, and finding the borders where styles change

This task is an expansion of the fundamental style change detection task which examines a document to determine whether it is single or multi-authored. This task has been solved using various clustering algorithms; to separate single from multi-authored documents, and authorship linking which breaks down a document into smaller sections to establish whether there are authorship changes in the various sections. Literature defines a number of clustering algorithms for the complete authorship clustering and authorship linking; distance measures, B-compact graph-based clustering, compression-based clustering, hierarchical clustering algorithms and local sensitive hashing algorithms [,4,13,15].

Simple distance measures which clusters documents written by the same author together based on the distances between them have been used to solve the problem of complete author clustering and authorship linking. For complete author clustering, this method takes the absolute differences of any two vectors element-wise and sums them up to form summations which are used to check for writing style changes. The summations are transformed to standard deviations, where a high standard deviation score yields more evidence that the pair of documents is written by the same author. For instance [4] used SPATIUM-L1 on character n-grams to solve the problem of authorship clustering. They investigated with different character n-grams and achieved best performance at character 2-grams, with the top 300 most frequent features at threshold of 3.0 symmetrical score. Another study [26] used SPATIUM, on most frequent words, punctuations and character n-grams of each selected text. To measure the distance between a text A and another text B, they used a variant of SPATIUM; L-norm called Canberra in which the absolute differences of the individual features are normalized based on their sum.

The other approach that has been used to solve this problem is the β-Compact graph based clustering [11]. The method is based on defining a threshold function β, which places documents into the various clusters only if the similarity between a pair of documents exceeds the threshold value. For instance [11] proposed a method for discovering author groups using a β-compact graph-based clustering. In this method each document is represented using the classic bag of words tried on different features. Similarity functions are then used to compare the similarity between a pair of documents, using only binary features. A threshold function β is used to place documents into clusters only if the similarity between two pairs exceeds the threshold value of 0.5.

Compression-based models have been proposed to solve the problem of complete author clustering and linking problems. For instance, [15] used compression-based models to perform document clustering into distinct clusters; they modified the k-medoids algorithm using a compression-based dissimilarity measure as opposed to the standard distance measure. The value of k- which represents the number of authors was determined by computing silhouettes coefficients in an iterative manner. N-clustering iterations were performed and the value of k that produced the maximum silhouette coefficient was picked. For the authorship link, they applied a dissimilarity function, compression-based cosine to measure how (dis)similar two documents are to each other. In order to establish authorship links within each cluster, compression-based cosine was modified to calculate similarity score instead of dissimilarity score. This approach does not perform well because compression-based dissimilarity measures do not fulfill even one of the required properties of a real distance-based metric such as identity, symmetry and triangle inequality.

[13] proposed a hierarchical clustering analysis of different document features: typed and untyped character n-grams and word n-grams for the complete author clustering. Hierarchical clustering analysis was used to determine the number of distinct clusters and to place each document in exactly one of the k-clusters. The hierarchical analysis was done using the bottom-up approach where each text starts in its own cluster and after each iteration, a pair of clusters are merged. The average cosine distance is used to decide when to merge pairs. To establish authorship links, pairwise similarity between each pair of documents in each problem was calculated using the cosine similarity metric. The use of the same feature set for all languages may have had a negative effect on the overall performance. Using different features for each language may help improve the problem.

[21] proposed a method for author clustering and style breach detection based on local-sensitive hashing-based clustering of real-valued vectors; a mixture of stylometric features and bag of n-grams. Tf-idf features and the Wilcoxon signed rank test were computed to determine the style breaches. The study investigated two Local-sensitive hashing algorithms; superbit and minHash and found out that superbit, which approximates cosine similarity, yielded the best results in author clustering. Silhouette coefficient was computed to determine the number of clusters. For the style breach detection, a statistical approach- Wilcoxon signed Rank, based on tf-idf features was used to determine the borders of the changing styles within a document.

These approaches used unsupervised techniques and therefore are applicable to solving real-world problems where the number of participating authors is not known in advance. Better still, they employ very simple techniques; distance measures and other clustering algorithms on standalone features thereby yield low runtimes. However, the results posted by these methods are slightly above the baseline and still require strengthening. Expanding feature sets could greatly improve the performance of these methods since features are generated at the sentence level, therefore just one feature type may not adequately represent the writing style of an author.

6.2.3 Is a document multi-authored, if yes determine the number of authors who collaborated

Determining the number of authors in a multi-authored document is the goal of style change detection. However, the need to subject the model only to multi-authored documents necessitates the separation of single authored from multi-authored documents [43]. Whereas this task can be inherent in a model for determining the number of authors in a document, it is essentially done first to minimize the number of documents passing through the model for predicting the number of authors based on similarities in styles of writing and to improve model efficiency [9,39]. Supervised and unsupervised learning techniques have been used to tackle this problem.

[3] used unsupervised learning to determine the number of authors in a multi-authored document using 500 most common words as the style marker. They defined two levels, firstly to cluster the chunks into two, three or four author clusters, using cosine similarity. They then applied supervised learning on an expanded feature set to distinguish between the clusters. They found out that unsupervised learning yielded better results than supervised learning, however as author numbers increased, the model accuracy reduced.

Another study [30] used a combination of features to establish the number of writers in documents. They defined an algorithm using an ensemble of two unsupervised learning algorithms; a threshold based and window merge clustering methods. This study first employed the threshold algorithm to cluster windows based on their closeness. That is, windows with the smallest distances between them, are put in one cluster because it is assumed that such windows have the same author. Then the most similar windows were merged using the distance matrix to calculate the distance between the new windows. The study found out that Threshold Based Clustering outperformed the Window Merge Clustering. Although the use of duplicate sentences improved significantly the accuracy, it also led to an increase in the OCI value. Better still, this scenario was unique to this dataset and may not provide a good generalization.

[46] Defined a two-pipeline for determining style changes in documents. First, they used a feedforward neural network to categorize single authored documents from multi-authored documents. They then applied a 3-model clustering to establish the number of writers in the multi-authored texts. To cluster segments into groups in a multi-authored document, they used various combinations of stylometric features and an ensemble of clustering algorithms. The ensemble consisted of k-means, k-means with similarity and hierarchical clustering. K-means clustering algorithm was used to separate single-authored documents from multi-authored ones. To form the clusters, they employed silhouetting on the k-means algorithm to determine the number of clusters. To establish the number of writers in a document, hierarchical clustering was used on all the features except the tf-idf features, together with the feed forward neural network to determine the exact number of clusters in multiple authored documents. The study noted that classification results varied with an increasing number of authors in a document.

6.2.4. Is a given document multi-authored, if yes is there a style change between consecutive paragraphs?

Determining the change in style between consecutive paragraphs can be approached as a supervised learning problem by generating feature vectors for each paragraph and comparing these feature values [44]. It can be solved using paragraph representations or simply by breaking the document into sentences and generating features at the sentence level [7,29]. For instance,

[7] **u**sed characters, lexical and syntactic style markers to build a paragraph representation to establish the number of writers of a document and the corresponding paragraphs authored by each. The study grouped Paragraphs according to a defined heuristic based on the B0- maximal clustering algorithm. This approach suffers from paragraph overlap. This problem was partly eliminated by considering the order of the paragraph in the document. This method assumes that the writing style in a document is characterized by the style reflected in the first paragraph, and that the main author tends to write the majority of the paragraphs, particularly the first ones. Whereas this assumption may be true the effects of other characteristics such as the size, strength of similarity or the adjacency of the paragraphs ought to have been considered. Paragraph overlap still remains a challenge.

The approach of [29] is based on using Google's BERT language model as a feature extractor, and random forests as a classifier. First, the documents contained in the dataset are split into sentences, and every sentence is fed to BERT, taking the outputs of the last four BERT layers to represent a given sentence. Since the size of the feature matrix produced by this depends on the number of tokens in a sentence, the values along the length dimension are summed to obtain a feature

matrix of a fixed length. After this, representations are formulated for consecutive pairs of paragraphs (to solve the second task), and the whole document (to solve the first task), based on the representations of sentences, by summing (paragraphs) or averaging (whole documents) the feature values of the sentences that make up the paragraph or document. These feature representations are then used to train random forest models for both tasks.

6.2.5. Given a text, find out whether the text is written by a single author or by multiple authors. For each multi-authored text, find the positions of the changes and assign all paragraphs of the text uniquely to some author out of the number of authors you assume for the multi-author document.

This task combines all the other tasks of style change; author clustering, authorship linking, number of authors and finally introduces a new task of assigning paragraphs of the text uniquely to an author. It has been approached using supervised learning techniques, which yield better accuracies, by performing pairwise comparisons of paragraphs. For instance, [9] proposed the use of multi-layer perceptrons and bidirectional LSTMs for the style change detection. Widely used textual features such as mean sentence length in words, mean word length or corrected type-token ratio, and pre-trained FastText embeddings. Multi-layer perceptrons with three hidden layers that are fully connected are used to categorize single authored texts from multi-authored text.

[22] trained a logistic regression classifier on the absolute vector difference between the feature vectors corresponding to each paragraph pair to solve the problem of style change detection. If the average of the classifier scores corresponding to the adjacent paragraph pairs is greater than 0.5, then the document is multi-authored.

[41] used a stacking ensemble of classifiers trained on separately extracted features and BERT embeddings, and combined their predictions by a meta-learner, i.e the stacking ensemble. Classifying single or multi-author documents was achieved by classification on the document level features. A single feature vector per document and label was used to classify each document as being either single or multi-authored.

[45] used google's pre-trained BERT model to determine the style change detection as a binary classification problem based on the similarity of writing style. They modeled the problem of writing style change detection as discovering the similarity of writing styles between different text segments. 'The style changes and the decision of author identifiers were regarded as binary'. They adopted the Bert pre-training model to extract the paragraph features and build a model to solve all the style change detection problems outlined in the competition. In this model they report that if task 2 label includes 1, the corresponding text will at least be two authors, and the corresponding task1 label will be 1. Otherwise task 1 label will be 0. Two paragraphs were presented for similarity measurement where high similarity indicates no change in writing style between the two paragraphs. A low similarity denotes change in writing style. To estimate writing style similarity, BERT together with Fully connected Neural network classifiers were used.

7. Evaluation Metrics

Numerous evaluation metrics exist in literature which can be used to verify the performance of the different methods for writing style change detection. The choice of an evaluation metric is dependent on the task at hand and the desired output. In literature measures such as accuracy score, F1-score, Bcubed-F1 score and mean Average Precision have been used to evaluate the performance of the writing style change detection methods.

7.1. Accuracy

Accuracy is a measure of correctness of the model's predictions of a writing style change detection method. It computes the number of correct predictions in relation to the total number of predictions as follows;

$$Accuracy = \frac{number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \quad [1]$$

For binary classification tasks, accuracy can be calculated in terms of positives and Negatives as follows;

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

The possible scenarios in the binary classification performed in experiment one are; a positive observation predicted as positive known as True Positive (TP), a positive observation predicted as negative referred to as False Negative (FN), a negative observation predicted as negative known as True Negative (TN and a negative observation predicted as positive referred to as False Positive (FP). This measure is widely used in writing style change detection to separate single authored from multi-authored documents, although few studies have also been used to determine the number of authors in a multi-authored document [3]. Accuracy is used to measure the purity of writing style change detection models such that higher accuracy values indicate that the model predicted most values correctly. However, it can not be adequately used alone in cases where the dataset is not balanced [30,35].

7.2. F1-Score

F1-score is defined as the harmonic mean of precision and recall. The harmonic mean is an alternative metric for the more common arithmetic mean. F1-score is computed as follows;

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad [9]$$

This measure is used when computing the average performance rate where there are more than one task. For instance in [25] it was used to combine the precision of the model on grouping documents written by the same author together, and the recall of grouping sections of a document written by the same author together. The overall performance of the method was determined by calculating the F1-score, which is the average performance rate. While a higher F1-score is desirable, a medium F1 value may require scrutiny to identify the type of errors.

Both precision and recall have the same weight in F1 measure. A high F1-score is achievable if both recall and precision are high, while a low F1 value indicates that both recall and precision values are low. A medium F1 value is obtainable if either precision is high and recall is low and vice versa.

7.3. BCubed-F1 measure

Bcubed-F1 scoring is based on performance evaluation of clusters. The precision and recall for each entity are calculated and then combined to produce the final precision and recall for the entire output [28,37]. For an entity i, the precision and recall are defined as follows;

Precision is the ratio of the number of correct elements in the output chain containing i, to the number of elements in the output chain containing entity i. While recall is defined as the ratio of the number of correct elements in the output chain containing i, over the number of elements in true chain containing i.

The final precision and recall for all the entities are;

$$recall = \sum_{i=1}^{N} wi.recalli$$

$$precision = \sum_{i=1}^{N} wi.precisioni$$

Where N is the number of entities in the document, and wi is the weight assigned to entity i in the document.

Bcubed-F1 score is used to overcome the shortcomings of F1-score where both recall and precision have the weight, and therefore considers all types of errors to be equal.

7.4. Ordinal Classification Index (OCI)

Ordinal classification is a form of multiclass classification for which there is an inherent order between the classes, but not a meaningful numeric difference between them. The OCI measure used to measure the error of predicting the number of authors for documents with multiple authors. Since it is a measure of the error rate, it is computed by calculating the Mean Absolute Error (MAE) which addresses the problem of ordinal classification as a regression problem [7,30,46].

$$MAE = \frac{1}{N}\sum_{x\epsilon\sigma} |g(e_x) - g(\hat{e}_x)| \quad [43]$$

Where g(.) corresponds to the number assigned to a class, N = card and e_x and are the true and estimated values.

The OCI value is the inverted value of the MAE. The smaller the OCI value the better the performance.

8. Results

The focus areas for this literature review were the techniques employed in writing style change detection, the features used as well as the results obtained by the various approaches. For ease of results presentations and discussion, all the studies were grouped under three study areas; author diarization and clustering and style change detection. The techniques and features used together with the results obtained by each approach is presented in table 1 below.

SN	Study Area	Task	Features	Techniques	Data set	Results	REFERENCES
1	Author clustering	Complete author clustering and author diarization	Character count, digit counts, uppercase letters counts, spaces count, tabs count, words count, ratio of interrogative	Cluster Distance Approach	PAN	Bcubed- F1 of 0.37	[40]

			sentences, average word length, average sentence length, ratio of digits, ratio of uppercase letters, ratios of spaces, ratios of tabs.				
2	Author clustering	Complete author clustering and author diarization	most frequent words(words and punctuations)	Unsupervised method with a simple distance measure	PAN	F1-score of 0.821	[25]
3	Author clustering	Complete author clustering and author diarization	word frequencies,n-gram frequency count and length, POS, sentence length	Threshold-based outlier detection method	PAN	F1-score of 0.5	[28]
4	Author Clustering	Complete author clustering and authorship link-ranking	no feature engineering	K-medoids and Compression-based Dissimilarity scores	PAN	Average Precision of 0.12	[15]
5	Author clustering	Complete author clustering and authorship link-ranking	typed and untyped character n-grams and word n-gram, most frequent terms	Hierarchical clustering analysis and cosine similarity	PAN	Bcubed-F of 0.57 MAP of 0.455	[12]
6	Author clustering	Complete author clustering and authorship link-ranking	character n-grams	Simple distance measure called SPATIUM-L1	Pan	Bcubed-F of 0.53 and MAP of 0.04	[4]

7	Author clustering	Complete author clustering and authorship link-ranking	Most frequent terms (character n-grams, isolated words and punctuation symbols)	Simple distance measure called SPATIUM-L1	Pan	Bcubed-F of 0.55 MAP of 0.473	[26]
8	Author cluster	Complete author clustering and authorship link-ranking	mixture of stylometric features (Special character frequency, average word length, average sentence length in characters, average sentence length in words and vocabulary richness) and bag of n-grams	Locality sensitive hashing-based clustering and statistical approach	PAN	Bcubed-F of 0.28 MAP of 0.47	[21]
9	Author cluster	Complete author clustering and authorship link-ranking	bag of words on different features	B-compact graph-based clustering	PAN	Bcubed-F of 0.56 MAP of 0.37	[11]
9	Style change detection	Finding number of authors	most common words	unsupervised learning methods and cosine similarity	Own dataset	Accuracy of 74%	[3]
10	Style change Detection	Complete author clustering and finding number of authors	combinations of features and duplicate sentences	Threshold-based and Window Merge clustering techniques	PAN	Accuracy of 0.85 OCI of 0.87	[30]

11	style change detection	Complete author clustering and finding number of authors	Bag of words and combinations of features	ensemble of clustering algorithms	PAN	Accuracy of 0.6 OCI of 0.808	[46]
13	style change detection	Complete author clustering and finding number of authors	combination of features	B0-Maximal clustering	PAN	F1 score of 0.64	[7]
14	Style change detection	Separating single from multi-authored	tf-idf features	Parallel attention networks and similarity functions	PAN	Accuracy 0.8	[16]
15	Style change detection	Combining different tasks	Mean sentence length in words, mean word length, corrected type-token ratio and pretrained fastText embeddings	Multi-layer perceptrons and bidirectional LSTMs	PAN	average F1 score of 0.517	[9]
16	Style change detection	Combining different tasks	Tf-idf features, n-grams of part of speech tags and vocabulary richness	logistic regression classifiers	PAN	average F1 score of 0.574	[22]

17	Style change detection	Combining different tasks	Combination of different stylometric features	Ensemble of classifiers	PAN	average F1 score of 0.642	[41]
18	Style change detection	Combining different tasks	No stylometric features	Pre-trained BERT and Neural Networks	PAN	average F1 score of 0.668	[45]
19	Style change detection	Combinatio n of different tasks	Vocabulary richness	Siamese Neural Networks	PAN	Average F1 Score of 0.450	[31]

9. Discussion

Writing style change detection which aims to determine the number of authors collaborating in a document has not been extensively studied if literature is anything to go by. Apart from a few studies such as the work by [3] and [13] most studies were done as a result of the annual PAN CLEF competitions on author diarization and clustering, and style change detection. The various tasks defined under the style change detection are; to determine whether a given text is multi-authored or not; determining the positions of style change in multi-authored documents; determining the number of authors in multi-author documents and assigning sections of multi-authored documents to a probable author.

Author clustering studies seek to group together documents written by the same author together by examining documents for similarities in writing styles. Various stylometric features exist that can be used to discriminate between the works of different authors such as lexical, character, syntactic, content-based and structural features. These features have been applied either as standalone or as combinations. However, for the author clustering task reduced feature sets tends to be more effective as opposed to using a number of features [25]. This is so because of the sufficiency of the authors' data which can distinguish between different writing styles. Moreover, model overfitting on the data and runtime are reduced when a few features are used in long length documents. Because the document length is long, sometimes increasing the number of features does not result in improved performance.

In terms of methods used, simple distance measures are still effective in author clustering tasks as they yield comparable, if not better, results to state of the art studies. Simple distance measures methods called SPATIUM based on the L1 norm or other variants of SPATIUM known as L-norm are shown to perform better than other approaches.

Since they are simple methods, they also yield low runtimes which is beneficial to the author's clustering problems. The use of compression-based dissimilarity methods did not yield good performance.

For the author's diarization tasks, studies investigated the use of anomaly detection methods and detection of outliers. These methods are applicable in the intrinsic plagiarism detection which assumes that there is one main author (writing upto 70% of the document), and other authors contributing the rest of the text in the document. Whereas these methods yield promising results, they are only applicable where this heuristic holds true. In addition, they are effective only when the first so many consecutive paragraphs are written by the main author and not vice-versa. Simple distance measures (SPATIUM-L1) and hierarchical clustering yield the best results for the complete authorship clustering and authorship linking task. [26] used SPATIUM-L1 on most frequent terms and obtained a MAP of 0.473 while [21] used Locality sensitive hashing-based clustering and statistical approach on a mixture of stylometric features and obtained a MAP of 0.470.

Determining the number of authors in multi-authored documents has been solved by a number of studies. These studies approach this task as a clustering problem where the number of authors is not known in advance. Clustering algorithms, which perform best with unlabeled data scenarios are used to determine the number of authors in a document. Most studies propose the use of ensembles of clustering algorithms which have been reported to produce better performance [5,6]. Since the text length is short (paragraph), the use of feature combinations is preferred as they offer more attributes to define an author's style. For instance, [46] used three algorithms; K-means, K-means with similarity and Hierarchical clustering. K-means was used to separate single-authored from multi-authored documents, while K-means with similarity was used to form clusters while hierarchical clustering was used to determine the number of clusters obtaining an accuracy of 0.604 and OCI of 0.809. Similarly, [30] used two algorithms; threshold-based and Window merge algorithms. This study used the two algorithms independently and found out that the window merge method yields better results. They obtained an accuracy of 0.848 and OCI of 0.865.

On the other hand, various stylometric features have been used across all studies. This study found out that lexical feature category is the most commonly used feature as style markers in previous studies. However, the use of feature combinations is highly recommended when text length is small, since they produce better accuracies. Although there is no optimal feature set combination that can be applied across all the problems and with different techniques, feature selection is still important in writing style change detection.

Supervised and unsupervised learning are applied in most of the writing style change detection with varying results. Although unsupervised learning produces the best results in cases where only unlabeled data is available, this study reports the success of supervised learning on clustering tasks. Clustering tasks can be modeled as binary classification problems and supervised learning used to detect the change in styles at the document or even paragraph level with better performance. For instance, [44] uses a pre-trained BERT and Neural Networks and obtains an average F1 of 0.668 while [29] uses ensembles of supervised classifiers and obtains an average F1 of 0.642.

10. Conclusion

Writing style change detection which aims to determine the number of authors in a multi-authored document is still considered a difficult task. various approaches and techniques proposed in literature continue to yield promising results, further research on alternative techniques and strengthening of the existing techniques is still required. For instance, only three techniques were identified under author diarization and clustering; cluster distance measure, outlier anomaly detection methods and a simple distance measure called the SPATIUM-L1 approach. The best results were achieved by the outlier anomaly detection method which yielded a Bcubed-F of 0.5. The distance measure yielded a Bcubed-F of 0.37 while the SPATIUM-L1 approach achieved an f-score of 0.82 on the same dataset.

The approaches identified for authorship clustering were able to group documents of the same author together with varying precision. For instance, compression-based dissimilarity scores achieved a precision of 0.12, hierarchical clustering analysis obtained a final MAP of 0.455, SPATIUM-L1 on singleton feature obtained a MAP of 0.04. When SPATIUM-L1 was investigated on an expanded feature set, a significant improvement on the precision level was noted at a final MAP of 0.47. The local sensitive hashing-based clustering achieved a MAP of while the B-compact graph-based clustering reported a MAP of 0.37 on author clustering. This study found out that simple distance measures and Google's Pre-trained BERT algorithms are still effective approaches in writing style change detection tasks. In addition, the use of singleton features may yield poor precision levels when the document length is short.

Two techniques were identified and reviewed for determining the number of authors within a multi-authored document; an ensemble of three clustering algorithms, and a window merge and Threshold-based clustering algorithm. The ensemble of clustering algorithms technique outperformed the other technique achieving an Ordinal Classification Index (OCI) of 0.808. The Threshold-based

clustering algorithm achieved an OCI of 0.87 in determining the number of authors participating in writing a text. Although these are promising results, they are barely above the baseline and therefore still need strengthening. It is important to note that few studies have investigated the use of ensembles of algorithms implementing consensus clustering.

11. Recommendation

This study notes the success of the state of the art studies on writing style change detection. However, it proposes the use of combinations of features with various techniques as this is shown to have a significant effect on the accuracy of these models especially with short text length.

The study also recommends the use of ensembles of clustering algorithms in writing style change detection studies. It is believed that ensembles of clustering algorithms can help solve the problem of similarity overlap, which is the main challenge of style change detection on larger datasets.

The use of more than one dataset is also recommended to clear the issue of biases in datasets, which hinders the applicability and standardization of writing style change detection approaches.

References

[1] Abbasi, A., & Chen, H. (2005). Applying authorship analysis to extremist-group Web forum messages. IEEE Intelligent Systems, 20(5), 67–75. https://doi.org/10.1109/MIS.2005.81

[2] Ahmed, H. (2018). The Role of Linguistic Feature Categories in Authorship Verification. Procedia Computer Science, 142, 214–221. https://doi.org/10.1016/j.procs.2018.10.478

[3] Akiva, N., & Koppel, M. (2012). Identifying distinct components of a multi-author document. Proceedings - 2012 European Intelligence and Security Informatics Conference, EISIC 2012, 205–209. https://doi.org/10.1109/EISIC.2012.16

[4] Alberts, H. (2017). Author clustering with the aid of a simple distance measure: Notebook for PAN at CLEF 2017. CEUR Workshop Proceedings, 1866.

[5] Brocardo, M. L., Traore, I., Saad, S., & Woungang, I. (2013). Authorship verification for short messages using stylometry. 2013 International Conference on Computer, Information and Telecommunication Systems, CITS 2013. https://doi.org/10.1109/CITS.2013.6705711

[6] Brocardo, M. L., Traore, I., & Woungang, I. (2015). Authorship verification of e-mail and tweet messages applied for continuous authentication. Journal of Computer and System Sciences, 81(8), 1429–1440. https://doi.org/10.1016/J.JCSS.2014.12.019

[7] Castro-Castro, D., Alberto Rodríguez-Losada, C., & Muñoz, R. (n.d.). Mixed Style Feature Representation and B 0-maximal Clustering for Style Change Detection Notebook for PAN at CLEF 2020.

[8] Daelemans, W., Verhoeven, B., Potthast, M., Stamatatos, E., Stein, B., Juola, P., Sanchez-Perez, M. A., & Barrón-Cedeño, A. (n.d.). Overview of the Author Identification Task at PAN 2014.

[9] Deibel, R., & Löfflad, D. (2021). Style change detection on real-world data using an LSTM-powered attribution algorithm. CEUR Workshop Proceedings, 2936, 1899–1909.

[10] Ding, S. H. H., Fung, B. C. M., Iqbal, F., & Cheung, W. K. (2016). Learning Stylometric Representations for Authorship Analysis.

[11] García-mondeja, Y., Castro-castro, D., & Lavielle-castro, V. (2017). Discovering Author Groups using a β-compact. 1–6. http://ceur-ws.org/Vol-1866/

[12] Gómez-Adorno, H., Aleman, Y., Vilariño, D., Sanchez-Perez, M. A., Pinto, D., & Sidorov, G. (2017). Author clustering using hierarchical Clustering analysis: Notebook for PAN at CLEF 2017. CEUR Workshop Proceedings, 1866.

[13] Gómez-Adorno, H., Posadas-Duran, J. P., Ríos-Toledo, G., Sidorov, G., & Sierra, G. (2018). Stylometry-based approach for detecting writing style changes in literary texts. Computacion y Sistemas, 22(1), 47–53. https://doi.org/10.13053/CyS-22-1-2882

[14] Gorman, R. (2020). Author identification of short texts using dependency treebanks without vocabulary. Digital Scholarship in the Humanities, 35(4), 812–825. https://doi.org/10.1093/LLC/FQZ070

[15] Halvani, O., & Graner, L. (2017). Author Clustering using compression-based dissimilarity scores: Notebook for PAN at CLEF 2017. CEUR Workshop Proceedings, 1866.

[16] Hosseinia, M., & Mukherjee, A. (2018). A parallel hierarchical attention network for style change detection: Notebook for PAN at CLEF 2018. CEUR Workshop Proceedings, 2125.

[17] Howedi, F., Mohd, M., Aborawi, Z. A., & Jowan, S. A. (2020). Authorship Attribution of Short Historical Arabic Texts using Stylometric Features and a KNN Classifier with Limited Training Data. Journal of Computer Science, 16(10), 1334–1345. https://doi.org/10.3844/jcssp.2020.1334.1345

[18] Jankowska, M., Milios, E., & Kešelj, V. (n.d.). Author Verification Using Common N-Gram Profiles of Text Documents.

[19] Jiexu L. I., Zheng, R., & Chen, H. (2006). From fingerprint to writeprint. In Communications of the ACM (Vol. 49, Issue 4, pp. 76–82). Association for Computing Machinery. https://doi.org/10.1145/1121949.1121951

[20] Juola P. (2006). Authorship attribution for electronic documents. IFIP International Federation for Information Processing, 222, 119–130. https://doi.org/10.1007/0-387-36891-4_10

[21] Karaś,D., Śpiewak, M., & Sobecki, P. (2017). OPI-JSA at CLEF 2017: Author clustering and style breach detection: Notebook for PAN at CLEF 2017. CEUR Workshop Proceedings, 1866.

[22] Kaur R., Singh, S., & Kumar, H. (2020). TB-CoAuth: Text based continuous authentication for detecting compromised accounts in social networks. Applied Soft Computing Journal, 97.

[23] Kestemont, M., Tschuggnall, M., Stamatatos, E., Daelemans, W., Specht, G., Stein, B., & Potthast, M. (2018). Overview of the Author Identification Task at PAN-2018 Cross-domain Authorship Attribution and Style Change Detection.

[24] Khan, J. A. (2018). A model for style change detection at a glance: Notebook for PAN at CLEF 2018. CEUR Workshop Proceedings, 2125.

[25] Kocher, M. (2016). UniNE at CLEF 2016: Author Clustering. CEUR Workshop Proceedings, 1609, 895–902.

[26] Kocher, M., & Savoy, J. (2017). UniNE at CLEF 2017: Author clustering: Notebook for PAN at CLEF 2017. CEUR Workshop Proceedings, 1866.

[27] Koppel, M., & Schler, J. (2004). Authorship verification as a one-class classification problem. Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004, 489–495. https://doi.org/10.1145/1015330.1015448

[28] Kuznetsov, M., Motrenko, A., Kuznetsova, R., & Strijov, V. (2016). Methods for intrinsic plagiarism detection and author diarization. CEUR Workshop Proceedings, 1609, 912–919.

[29] Iyer, A., & Vosoughi, S. (2020). Style Change Detection Using BERT. Clef 2020. http://ceur-ws.org/Vol-2696/paper_232.pdf

[30] Nath, S. (2019). Style Change Detection by Threshold Based and Window Merge Clustering Methods (Notebook paper) Style Change Detection by Threshold Based and Window Merge Clustering Methods. September.

[31] Nath,S. (2021). Style change detection using Siamese neural networks.

[32] Potha, N., & Stamatatos, E. (2018, July). Intrinsic author verification using topic modeling. ACM International Conference Proceeding

Series. https://doi.org/10.1145/3200947.3201013
[33] Ramnial, H., Panchoo, S., & Pudaruth, S. (2016). Authorship attribution using stylometry and machine learning techniques. Advances in Intelligent Systems and Computing, 384, 113–125. https://doi.org/10.1007/978-3-319-23036-8_10
[34] Rexha, A., Kröll, M., Ziak, H., & Kern, R. (2018). Authorship identification of documents with high content similarity. Scientometrics, 115(1), 223–237. https://doi.org/10.1007/s11192-018-2661-6
[35] Rosso, P., Rangel, F., Potthast, M., Stamatatos, E., Tschuggnall, M., & Stein, B. (2016). Overview of PAN'16: New challenges for authorship analysis: Cross-genre profiling, clustering, diarization, and obfuscation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9822 LNCS, 332–350. https://doi.org/10.1007/978-3-319-44564-9_28
[36] Safin, K., & Ogaltsov, A. (2018). Detecting a change of style using text statistics: Notebook for PAN at CLEF 2018. CEUR Workshop Proceedings, 2125.
[37] Safin, K., & Kuznetsova, R. (2017). Style breach detection with neural sentence embeddings: Notebook for PAN at CLEF 2017. CEUR Workshop Proceedings, 1866.
[38] Sari, Y. (2018). Neural and Non-neural Approaches to Authorship attribution.
[39] Sari, Y., & Stevenson, M. (2016). Exploring Word Embeddings and Character N -Grams for Author Clustering Notebook for PAN at CLEF 2016. Working Notes for CLEF.
[40] Sittar, A., Iqbal, H. R., & Nawab, R. M. A. (2016). Author diarization using cluster-distance approach. CEUR Workshop Proceedings, 1609, 1000–1007.
[41] Str, E. (2021). Multi-label Style Change Detection by Solving a Binary Classification Problem.
[42] Tschuggnall, M., Stamatatos, E., Verhoeven, B., Daelemans, W., Specht, G., Stein, B., & Potthast, M. (2017). Overview of the author identification task at PAN-2017: Style breach detection and author clustering. CEUR Workshop Proceedings, 1866.
[43] Zangerle, E., Tschuggnall, M., Specht, G., Stein, B., & Potthast, M. (2019). Overview of the Style Change Detection Task at PAN 2019. September, 9–12.
[44] Zangerle, E., Mayerl, M., Specht, G., Potthast, M., & Stein, B. (2020). Overview of the Style Change Detection Task at PAN 2020. CEUR Workshop Proceedings, 2696.
[45] Zhang, Z., Han, Z., Kong, L., Miao, X., Peng, Z., Zeng, J., Cao, H., Zhang, J., Xiao, Z., & Peng, X. (2021). Style change detection based on writing style similarity. CEUR Workshop Proceedings, 2936, 2208–2211.
[46] Zuo, C., Zhao, Y., & Banerjee, R. (2019). Style Change Detection with Feed-forward Neural Networks. September, 9–12.