

Temporal positional lexicon expansion for federated learning based on hyperpartism detection

Usman Ahmed¹  | Jerry Chun-Wei Lin¹  | Gautam Srivastava^{2,3,4} 

¹Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway

²Department of Mathematics & Computer Science, Brandon University, Brandon, Canada

³Research Centre for Interneural Computing, China Medical University, Taichung, Taiwan

⁴Department of Computer Science and Math, Lebanese American University, Beirut, Lebanon

Correspondence

Jerry Chun-Wei Lin, Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway.

Email: jerrylin@ieee.org

Abstract

Internet-based information exchange has resulted in the propagation of false and misleading information, which is highly detrimental to individuals and humankind. Due to the speed and volume of social media news production, supervised artificial intelligence algorithms require many annotated data, which is difficult, costly, and time-consuming. To address this issue, we offer a novel federated semi-supervised framework based on self-ensembling that utilizes the linguistic and stylometric information of annotated news articles and searches for hidden patterns in unlabeled data to denoise labels. Self-ensembling predicts the labels of unlabeled data by using the outcomes of network-in-training from earlier epochs. These cumulative predictions should be a stronger predictor for unknown labels than the output of the most recent training epoch; hence, they may be utilized as a substitute for the labels of unlabeled data. The approach is distinctive in collecting all of the outputs from the neural network's past training periods. It utilizes them as an unsupervised target against which to assess the current output prediction of unlabeled articles. We intend to create a dataset centred on denoising to forward the study. The dataset is mapped using (1) the shifting focus time from published news articles and (2) the semi-supervised method based on coincidence contexts for a neural contrast embedding model for learning low-dimensional continuous vectors that generate a focus time-based query in sequential news articles for temporal comprehension. The model achieved 0.83% F-measure with lexicon expansion semi-supervised learning.

KEYWORDS

fake news detection, Infodemic, self-ensembling, semi-supervised, temporal ensembling

1 | INTRODUCTION

Disinformation that is hyper-partisan has emerged as a severe issue for the public, contributing to the escalation of political differences over even the most fundamental truths (Lazer et al., 2018). Due to confirmation bias, partisans are more likely to trust news that confirms their ideas, regardless of the truth of the information. It has been demonstrated that confirmation bias is a significant aspect that plays a role in the belief that social media users have in news items (Kim et al., 2019). On social media, partisans may spread information from a political standpoint. However, it may be inaccurate from a factual viewpoint, mainly since the attention is focused on the political alignment of the content rather than its validity. An initial set of studies concluded that partisanship is a better predictor of the distribution of false information than the truth among users (Moravec

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Expert Systems* published by John Wiley & Sons Ltd.

et al., 2020). In traditional machine learning, data must be delivered from several sources to a central site before machine learning models can be trained to discover patterns in the data. This training takes place at a central location. Sensitive data created by Internet of Things (IoT) devices is being transferred to the cloud due to the increased use of applications based on IoT. Due to the sensitive nature of data, there is a possibility that hostile individuals will attempt to hack these devices. In Federated Learning, the machine learning model is brought to the data source rather than transmitting data to a centralized cloud. This eliminates the need to transport data. Federated Learning can solve various obstacles related to cyber security in applications based on IoT.

Techniques from Natural Language Processing (NLP) are increasingly being used in various real-world applications (Clark et al., 2020). These applications include natural language inference, natural language comprehension, named entity recognition (NER), question answering, and generation (Tian et al., 2022). According to numerous research findings, most deep learning-based natural language processing tasks are driven by a large quantity of labelled data. This leads to poor task performance in many situations due to the restrictions put on labelled data (Radford et al., 2019). In order to circumvent this limitation, pre-trained models use a self-supervised representation learning method to automatically extract a vast amount of linguistic information and patterns from sizeable unlabeled text corpora. This method has become the preeminent approach in this area of research (Clark et al., 2020; Yao et al., 2021).

All of the pre-trained models make use of the same learning paradigm, which consists of the following steps: (1) training on unlabeled text data by performing a task that is similar to language modelling; and (2) fine-tuning on a smaller quantity of downstream data that is tagged for a specific purpose (Ahmed, Lin Jerry, & Srivastava, 2022a). Regarding accuracy, pre-trained models often perform better than traditional models when tested using the NLP benchmarks. However, for such a method of language development to become widespread, it is anticipated that two significant obstacles will be overcome. First, pre-training techniques require a large-scale dataset, yet, this can only be afforded by a few groups of significant high-tech businesses, which has led to the dominance of Artificial Intelligence (AI).

In addition, because of privacy concerns and legislative limitations, most sensitive data, such as medical health records and personal financial data, are kept in separate locations and are not allowed to be directly exchanged in practice (Ahmed, Srivastava, & Lin Jerry, 2022). It might be challenging for individuals and smaller enterprises to pre-train models on sensitive information. Second, a large number of computational resources are required for the pre-training phase of an NLP model (Ahmed, Srivastava, & Lin Jerry, 2021). This may be extremely expensive for certain small businesses, particularly those concerned about their data security and are unwilling to utilize cloud services (Ahmed, Lin Jerry, et al., 2021). Because of this need, it will be challenging for people or devices with low processing capabilities to perform such labour-intensive pre-trained model learning. This research presents two possible solutions for the abovementioned problems to challenge the dominance of AI and help resource-constrained devices while contributing to the pre-training work for NLP.

Because of privacy concerns, most customers or users keep their confidential information on local devices. When there is inadequate data for training, it is impossible for each customer to pre-train a language model, even if they have access to significant computer resources. Through Federated learning (FL), it is feasible to train a global model across numerous datasets without sharing the raw data to fully use the vast quantity of scattered, diverse, and privately owned data. This is done in order to maximize efficiency. FL is a collaborative learning paradigm that also considers privacy concerns. It learns a global model by aggregating the models learned on local devices. Customers that use FL do not have to worry about their sensitive data being shared with other customers. However, they still have the opportunity to collaborate to develop a pre-trained model.

1.1 | Motivation

Given that any system for identifying misinformation must operate in such a dynamic environment, wherein naturally occurring changes to the news are possible, it is unclear how such techniques should be constructed and maintained over time. It is feasible that both natural and intentional changes to the news might affect input data distribution, resulting in a decline in the performance of previously trained models.

The problem of fast-spreading misleading information has attracted a lot of attention due to the growing prevalence of internet usage. A kind of reporting known as “fake news” is journalism that is intentionally designed to incite controversy, be excessively prejudiced (sometimes known as “hyperpartisan”), or contain outright untruths in order to deceive its audience into holding a preconceived viewpoint. Firstly, our primary focus was on gaining knowledge and capitalizing on the opportunities presented by pertinent qualities such as topic and sentiment data. We reasoned that, given the purpose of hyperpartisan journalism, the stance taken on politically charged matters would be an essential factor in determining whether or not anything is hyperpartisan. Experiments, on the other hand, indicated that the dataset had intrinsic noise that acted as a considerable barrier to the development of a robust classifier: (1) Text inputs from an article that are noisy because they contain domain-specific (i.e., political) phrases, slang, and words that might be misspelt. (2) Labels that are primarily the result of applying publisher-level information in order to label articles (for example, all articles published by left or right-wing newspapers are labelled as “hyper-partisan”).

However, the development of human-labelled large-scale datasets is an activity that is both extremely costly and time-consuming. As a result, it is vital to identify a more efficient strategy to use this information that is poorly labelled. As a direct result of this, we conducted research and

experiments on noise reduction to advance our model learning. In this investigation, in order to de-noise the dataset, we make use of semi-supervised pseudo-labelling. In order to get a more accurate representation of the noisy input, we pre-train the BERT algorithm.

1.2 | Contribution

Active learning (AL) is a technique for machine learning that efficiently collects data for labelling from an (often enormous) pool of unlabeled data. The objective is to concentrate human labelling efforts on the most valuable data points to enhance model performance and decrease labelling expenses. This study recommends that data points close to the model's feature space with varying prediction probabilities be collected. In order to do this, the paper presents a novel acquiring algorithm that searches the pool of unlabeled data for instances of contrast high-class weights. The method picks unlabeled data points from the pool whose prediction probabilities differ from their training set neighbours. The method is comparable to diversity sampling, except that it employs feature space to generate neighbourhoods as opposed to clusters. Additionally, the proposed model uses uncertainty by sorting unlabeled data based on prediction probabilities.

This paper provides an approach for integrating NLP, attention-based active learning and Federated learning to acquire knowledge for hyper-partisan. A method of temporal context extraction was used to represent semantic vectors. Relevant boundary elements are extracted from unlabeled text and incorporated into the active learning process. The technique enhances model training with additional examples. The cycle will continue until an optimal solution is discovered. The pool of unlabeled text is introduced to the training set at this stage. This study aims to increase knowledge of the text data addition learning process. The suggested method can reduce data annotation tasks and enhance the learning system's generalization. The network's semantic vectors and synonym expansion aid in achieving high precision while minimizing the effects of data annotation. The findings of this research propose a FL architecture for the pre-training of language models. The pre-training task is split among several clients as specified by the FL structure. The clients and the server work together to train a pre-trained NLP model. After that, every client will train and update its embedding layers, then will give the server the embedded data. The server will compile the information provided by the clients and bring the Transformer layers up to date while it is doing training and communication activities. After the trained model has converged, the training will be complete. Every client can personalize the job to their specifications.

Specifically, the following contributions are included in this paper:

1. For the pre-training of language models, our research came up with the idea of the FL framework. For the proposed framework to make a meaningful contribution to a pre-training activity, it must be able to successfully service all devices or clients with low computing capabilities and a small amount of data.
2. To extract relevant and contextual information, the proposed approach focuses on accurately determining the focused expressions found to denoise the labelled data.
3. The learned representation is used to train and classify the input text using the attention-based method.

The rest of the paper is organized as follows. Section 2 discusses the state of the art. Section 3 describes the approach, the proposed algorithm, the data collection, and the training phase of the model. The results and improvements of the proposed model are detailed in the Section 4. The predicted prospects is summarized in Section 5.

2 | RELATED WORK

Current research concentrates on investigating two characteristics of veracity: the dependability of the news and its bias. Even though these ideas are linked, they highlight two distinct ways the media may misinform customers. When it comes to reporting information, a biased source is more likely to do so in an unbalanced or politically charged way. In contrast, reputable sources are more likely to offer content that is both factually true and well-researched (Baly et al., 2018). For example, an incorrect article can feature material in whole or made up, and it would do it maliciously most of the time. In contrast, a biased article (also known as hyperpartisan or strongly prejudiced) may contain factually correct information (Horne et al., 2018). However, it may be presented inaccurately, taken out of context, or incomplete. The two concepts are commonly confused with one another in news articles, even though one may be the cause of the other (e.g., political partisan motivating a false claim; Ferrara et al., 2016).

Automated tools use the characteristics of news articles (Horne et al., 2018). Using language characteristics as the basis, Horne et al. wanted to ascertain the reliability of news items and any inherent bias. Their method offered a comparative evaluation of the automatic classification of dependability and bias. However, the primary focus is on classification at the source level rather than at the article level.

Baly et al. provided fact-checked news article analysis, have been taken in order to concentrate on-ground realities of the news articles (Baly et al., 2018). In another research, the authors examined the content feature space used by fact-checked fake and actual publications, in addition

to sarcasm. Another research followed a similar line of thought, building fake news classifiers using n-gram features taken from the article's text. They then used a meta-learning technique to fact-check fake news items published by hyperpartisan news sites.

Horne and Adali evaluated fact-checked statements as the ground truth for remotely supervised credibility identification leveraging both content and source attributes (Horne & Adali, 2017). This allowed them to determine whether or not a source was credible. Their system was able to identify fake news reports with an accuracy rate of 96% by employing a deep learning model based on words, phrases, and the title of the news article. In place of simple labelling, the experimental design used fact-checked true stories sourced from politifact.com.

A second technique, which studied claim-based classification, used deep learning on labelled statements from politifact.com. This method was quite similar to the first (Horne & Adali, 2017). These labels had varying degrees of truthfulness. In addition to these machine learning methods, there have been several articles that use knowledge graphs to check the truth of claims automatically. Despite the facts, knowledge graphs, and learning from fact-checked articles providing helpful answers, do not scale well with methods that require continuous retraining, are not very good at capturing newly discovered information, and do not adequately capture article bias. In all of these tests, regardless of the labelling granularity, the most effective strategy for news categorization was supervised machine learning models employing features taken from the article content. These models achieved an accuracy ranging from 70% to well over 90% in each laboratory scenario.

Numerous Internet-based information retrieval applications employ time as their primary search parameter (Rosin et al., 2022). Temporal information retrieval is a subfield of information retrieval that is only beginning to develop as a distinct area of research. It considers both the textual and temporal relevance of the material to satisfy user requests for timely data. When looking at a temporal document, the document's creation date is typically a key consideration, and the majority of commercial search engines sort results according to the document's creation date (Khan Shafiq Ur et al., 2018). However, two concerns are associated with a document's creation time: (1) The document's creation date is not always accessible, and (2) the document's creation time may not correspond with the period covered in the text. If the user is more concerned with the document's focus time than its creation time, ranking papers by creation time may diminish the efficacy of an IR system (Khan Shafiq Ur et al., 2018).

Event detection (ED) attempts to identify event triggers in unstructured text and then automatically classify these events using a set of preset event types (Chen et al., 2017). This is one of the most critical steps to perform when attempting to collect data about organized events from unstructured texts. Dynamic Memory Networks (DMN) are used to generate not just more robust sentence encodings for event mentions, but also to learn better prototypes for the many sorts of events. Deng et al. (2020) suggested a more robust model that can extract contextual information from many references of events due to the multi-hop mechanism of DMN. On the other hand, prototypical vanilla networks only calculate event prototypes by averaging, and these networks only take up event references once. Experiments reveal that the network copes better with sample scarcity than a collection of baseline models, and performs more consistently in circumstances with a high diversity of event kinds. Simultaneously, the number of occurrences is exceedingly low. The network can handle the fact that there are fewer samples available.

Recent research on event detection (ED) has demonstrated that graph-based convolution neural networks (GCN) employing syntactic dependency graphs may attain peak performance when used for ED. In graph-based models, the computation of hidden vectors is independent of the triggering candidate words, removing irrelevant data for event prediction. In addition, the models currently utilized for ED do not take advantage of the correlation information that may be acquired via the dependency tree. The gating approach suggested by Lai Viet et al. (2020) is used to filter out noisy information concealed in the vectors of GCN models for ED using information collected from a trigger candidate. In addition, distinctive strategies for achieving contextual diversity for the gates and consistency in the essential ratings for graphs and models in ED are described. This is accomplished by adding unique techniques for generating significant scores. The experimental findings demonstrate that the suggested model has the highest performance on both ED datasets.

Suggested a hierarchical encoder that can capture both local and global text exchanges (Yang et al., 2021). The authors introduce an additional task to anticipate event-relevant context for multitask learning. This helps an argument-linking model understand event context. Our model performs substantially better than baseline solutions on a dataset often used for connecting arguments, confirming the strategy's efficiency.

Event extraction (EE) identifies event triggers and their rationales. Most recent event extraction research involves pattern- or feature-based algorithms trained on annotated corpora to discriminate triggers, arguments, and context. Due to the unequal distribution of event occurrences in the ACE corpus, many commonly used ACE event trigger phrases are absent in the training set. This affects system functioning. Cao et al. (2018) highlighted that expert-level TABARI patterns could increase EE performance. Experiments reveal that their pattern-based system with upgraded patterns may obtain an F-measure of 69.8% (with an absolute improvement of 1.9%) over the baseline, which is better than state-of-the-art systems in this field.

Hyper-partism event recognition is complex since it requires storing word meanings in diverse situations. Previous techniques emphasized language-specific information and natural language processing technology. Other languages are not as well-equipped as English. A technique that automatically learns from data is more promising than using language-specific resources. Feng et al. (2018) created a neural network that did not need a specific language to collect sequences. The authors utilized this information to train an event detector without explicitly encoding characteristics. Their technique provided robust, efficient, and accurate solutions for many languages. The technique scored 73.4% on the English ACE 2005 event detection issue, an absolute increase of 3.0%. Both Chinese and Spanish experiment outcomes were equivalent.

Alonso et al. split the approaches utilized in their study into two categories: content-based and non-content-based (Alonso Omar, 2008). The content-based technique used the document's content to identify its date; a dependent document collection with a timestamp is necessary to create a model. In contrast, non-content-based document dating relies on superfluous information. The primary disadvantage of these methods is the restricted availability and doubtful dependability of external sources.

Using a statistical language model, Rosin et al. assessed the time necessary to generate a document in a previous study (Rosin et al., 2022). In this procedure, the reference data is separated into several time granularities, and then unique temporal language models are developed for each time granularity. The temporal language model of each division is subsequently compared to the language model of the updated content. Wang et al. have expanded this model by including temporal entropy, Google Zeitgeist, and semantic preprocessing (Wang et al., 2021). Filannino gathered expressions of time from document texts. Then, they developed a chronology connected with a particular entity, calculating its maximum and lower limits (Filannino, 2016). Suissa et al. (2022) investigate the problems by assessing different application cases of Digital Humanities (DH) research published in recent scholarly works and their potential remedies. In addition, they provide DH professionals with a realistic decision model for determining when and how to pick the most relevant deep learning algorithms for their study.

Traditional lexicography involves a substantial investment of time and effort. Several languages have analogous constraints. As a result, their lexicographic representation is insufficient. Now, the same digital procedures that have hastened the creation of dictionaries for languages with abundant resources may be used for these languages. Low-resource languages have several traits, including a paucity of data and resources ranging from machine translation corpora to grammars and spell-checkers (Yeh et al., 2015). Limited-data synthetic languages are computationally inefficient. The first step in enhancing languages with limited resources is constructing bilingual dictionaries. Large volumes of annotated data are necessary to train machine-learning models or assess rule-based techniques for languages with limited resources. Transfer learning decreases the demand for labelled target data by transferring learned models and representations, whereas data augmentation and remote monitoring produce and increase task-specific training data (Kumar et al., 2021). Automated knowledge extraction in natural language processing would be suitable for obtaining automatically acceptable knowledge assertions from big datasets to resolve the issues above. Therefore, the automated extraction of knowledge employs a two-dimensional strategy. The first is the autonomous extraction of superficial knowledge from vast document collections, and the second is the statistical accumulation of the machine-collected superficial understanding, which requires additional semantics. In addition, semantic annotation, classical information extraction, ontology-based information extraction, and linguistic annotation would aid in automated knowledge extraction and natural language processing for the lexicography of resource-poor languages (Bhowmick Rajat et al., 2021; Jain et al., 2021).

Ahmed et al. proposed a structural hypergraph and an intuitive vocabulary for word representation (Ahmed, Lin Jerry, & Srivastava, 2022a). An embedded method was proposed based on federated learning for symptom identification in mental health. The model treats textual data as a sequence of words. Afterwards, the model can learn a low-dimensional continuous vector while preserving contextual association. The resulting models incorporating attention-based processes and federated learning are then empirically evaluated. In addition, the designed approach is appropriate for vocabulary expansion, grammatical word representation, and dynamic lexicon analysis. The objective is to generate semantic word representations using a model of attention networks. Eventually, clinical methods are applied to embed the content. The experimental results demonstrated that emotional words are encoded utilizing structural hypergraphs. Using a bidirectional LSTM architecture with an attention mechanism, a 0.86 ROC was attained.

Transaction classification and clustering activities in Cyber-Physical Systems have benefited from effective vector representation. Traditional methods employ heuristic-based approaches and various pruning techniques to locate the needed patterns (Ahmed, Srivastava, & Lin Jerry, 2021) quickly. With the broad and high-dimensional availability of transactional data in cyber-physical systems, classic approaches employing frequent itemsets (FIs) as features are challenged by dimensionality, sparsity, and privacy concerns. An embedding model for transaction categorization based on federated learning. The model utilizes a collection of frequent item sets to represent transaction data. The model may then learn low-dimensional continuous vectors by retaining the contextual link between frequent items and sets. To validate the created models with attention-based mechanisms and federated learning, the authors conducted an extensive experimental study on many high-dimensional transactional data. The findings indicated that the suggested model may aid and enhance the decision boundaries by minimizing the global loss function while preserving security and privacy.

3 | METHODOLOGY

Figure 1 outlines the detailed strategy for performing federated learning-based denoising of label data. This section details the document preprocessing, lexicon expansion, and temporal lexicon profiling processes. Obtaining the dataset, annotating the data, and developing a ranking function are also covered here.

Recently, a lot of interest has been shown in the utilization of knowledge-based (KB) systems. A lexicon is made up of several components, the most important of which are a vocabulary of word meanings and an acquired variety of context anchors. Affective knowledge is made up of words that can communicate the context of the situation. After that, we provide an embedding method that emphasizes contributions made in

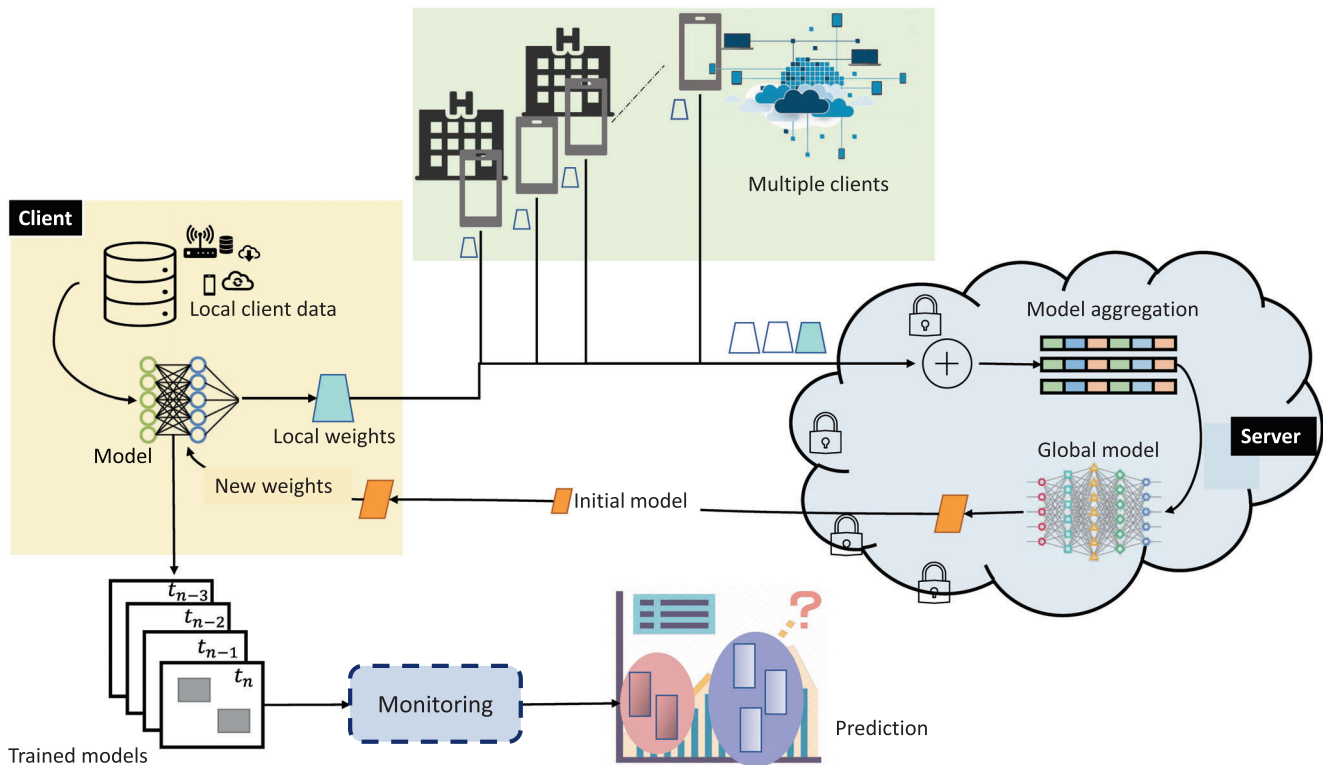


FIGURE 1 A workflow of the proposed federated learning approach

online discussion forums and uses keywords from the lexicon that are contextually flexible (depending on the meaning of the words). We made use of a pre-trained model for a global vector for the representation of words (Glove) that has 300 dimensions (Pennington et al., 2014). The word embedding for every single word token is present in the focus text. A glove-based vector embedding is utilized to project the context into the vector space. The new structure of the sentence taught is highlighted by the embedded component. Once the embedding has been retrieved, it is feasible that the semantic structure of the text will remain intact (Charles, 2000).

Our goal is to map the lexicon data by learning a function $f: \mathcal{D} \rightarrow \mathbb{R}^d$ such that every instance $l_i \in \mathcal{D}$ is mapped to a d -dimensional continuous vector. The semi-supervised learning requires having the denoising labelled based on the similar embedding. The embedding $\mathbf{X} = [f(l_1), f(l_2), \dots, f(l_N)]$ can be used by embedding.

The federated learning method with multi-client data shown in Figure 1 is used in the construction of the proposed framework mentioned in Figure 2. We sent the information to six different consumers in the same way so that we could perform equivalent testing. In addition, the data for each customer consists of data that does not overlap with other customers' data, the local model, and the database. The distribution of data across customers is completely random and without patterns. Because of these contrasts, the framework often shows real-world data where a large number of branches are connected to the server to explore supply and demand. Moreover, the configuration shows that the environment is not independent and has a homogeneous distribution throughout (Algorithm 1, line 1). The federated learning algorithm operates in a server-client environment to perform its tasks. Client information is stored locally on each model's system (Algorithm 1, lines 2–6). The initialization model is trained using locally collected data (Algorithm 1, lines 7–8). After each iteration, the client is required to submit either the locally learned weights or the gradient to the server. The server is responsible for collecting and computing the weights. This is to convey training of the client model without actually providing any data or weights.

After that, the global model updates automatically using the aggregated weights. Another cycle of local client models begin using global weights once the aggregation step is complete. After a predetermined number of rounds of federated learning, the client will eventually reach a convergence point (Algorithm 1, line 10). To find the convergence point experimentally, we relied on the early termination method. Throughout the empirical inquiry, the value for an early termination patient was consistently set to 10. The client can then select the most appropriate model for each iteration depending on the data held. The client records the validation loss on the local test set. The best global aggregated model and the best local iteration model are available for the client to choose from. We chose federated averaging because it converges quickly and reduces the risk of overfitting the model. However, our empirical studies have shown that the embedding size parameter should be increased to achieve the best possible performance. This is because the decoder model has a larger vector space in which to map positional attention.

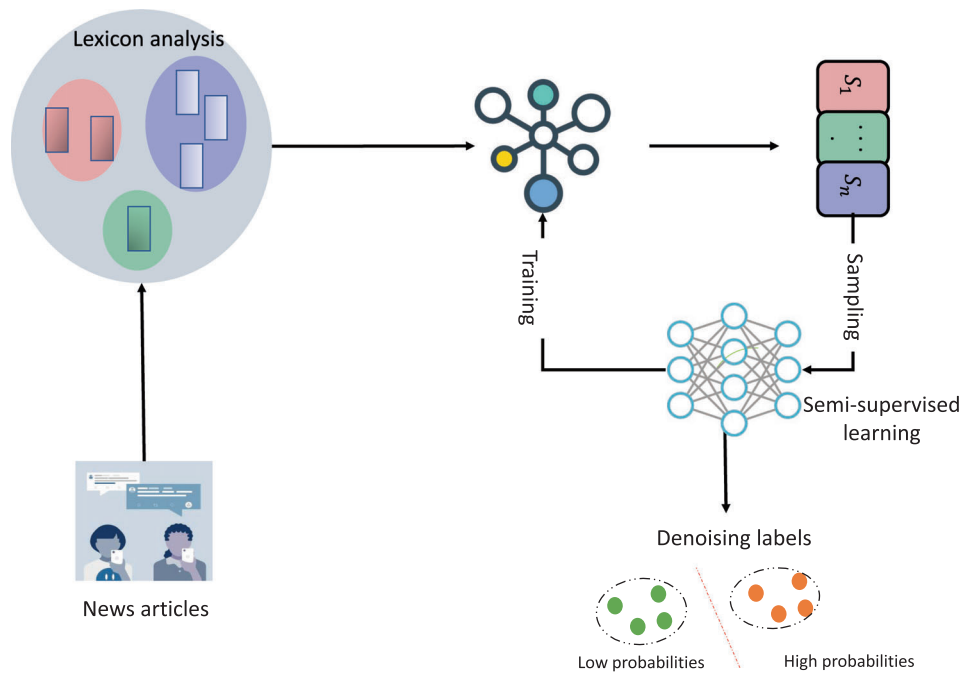


FIGURE 2 Client side model training

ALGORITHM 1 Attention-based federated averaging method

INPUT: I Instance, r number of rounds, k number of clients, ϕ_k model gradient, n new training gradient of clients and $\phi^{(r)}$ aggregate model

OUTPUT: Optimize gradient

- 1: $Weights \leftarrow \phi^{random}$ ▷ Initialize weights randomly
- 2: **for all** $r \in R$ **do**
- 3: **for all** client $\in n$ **do**
- 4: Send();
- 5: Receive();
- 6: **end for**
- 7: $\phi_k^{(r)} \leftarrow \phi^{(r-1)} + \Delta\phi_k^{(r)}$;
- 8: $\phi^{(r)} \leftarrow \frac{1}{\sum_k n_k} \sum_k (n_k \cdot \phi_k^{(r)})$ ▷ Aggregate()
- 9: **end for**
- 10: **Return** $\phi^{(r)}$

The global loss function that is the outcome of the weighted combination of the distributed aggregated function is what the federated averaging technique is trying to bring down, and this is the goal of the technique. The model might discover an embedding by combining local data sets with a representation of the embedding and using the parameter that minimizes its impact on the local data.

3.1 | Dataset

We utilize a publicly available dataset titled “SemEval 2019 Task 4 - Hyperpartisan News Detection” (Kiesel et al., 2019) that is made up of labels arranged in two separate ways: at the publisher level and the article level, as mentioned in Table 1. Level of the publisher (by publisher): 750 K items are tagged depending on the publisher’s political viewpoint, without regard for the content. It contains an equal number of hyperpartisan and non-hyper partisans. In the hyperpartisan samples, the ratio of right to left political inclination is equal. On an article-by-article basis: 645 articles were categorized as article-level by examining the actual content. Only articles for consensus among crowdsourcing employees are included

TABLE 1 Dataset characteristics

Labels	Training (%)	Validation (%)
Right	25	25
Right-centre	7.1	8.8
Left	25	25
Left-centre	11.7	15.7
Least bias	31.2	25.5

in the dataset. 238 (37%) of these individuals are hyperpartisan, whereas 407 (63%) are not. Through the utilization of human judgement, we observed that not all of the required labels are correct. This labelling noise is not uncommon since the political leaning of the publisher was the primary factor used to choose the labels. It is not possible to know with certainty that every essay published by a hyperpartisan publisher is also biased in the same direction. It is possible that there are just not enough newspapers that are not overtly prejudiced, and that is another issue that contributes to this cacophony.

3.2 | Content segmentation

This method divides the news article into three sections based on the number of terms and words it includes, with each section holding a particular proportion of the overall number of words. The first segment is made up of 50% of all words, the second 30%, and the third is the remaining 20%. This indicates that the first part is the most significant since it provides useful news-related information. This section appears at the start of the article. We have made the first portion extremely lengthy to reduce the possibility of missing vital information. Because the last portion of a news piece often does not contain much background information, we opted to maintain the original length of the third section despite its lower length. The text of news items is separated into phrases before breaking into three sections. The first segment has 40% of the entire phrases, while the second and third sections have 40% and 20%, respectively.

3.3 | Temporal profiling and expansion

The idea of a pyramid serves as the conceptual foundation for the method applied when determining how to classify various news items into one of three separate categories. In the first portion, all of the questions “what,” “when,” and “where,” as well as “who,” are addressed. This is seen in Figure 2. In order to get information that may be utilized for the questions that came before it, the first thing that is done is a search of the text of the news report for phrases that can be translated as “what,” “when,” “where,” and “who.” Who exactly are the people and groups going to be attending this event? When does the event start, and what time is it? Where exactly will the event be taking place across the world? What is the connection between the question “What is the true event?” and a person who took part in the event? Because the title of the article, which also functions as a definition of the occurrence, provides a response to the first part of the question, “what,” the keywords are extracted from the title. If we want to be able to extract information about when something took place, we have to run the text through a temporal tagger first. This tagger locates and normalizes any temporal expressions in the text. In order to respond to questions about where and who, the Stanford NER system is utilized to tag geographical locations, persons, and organizations. In this way, our algorithm searches for the phrases “entity,” “title,” “time,” and “geographic place” (Khan Shafiq Ur, Islam Muhammad, et al., 2018; Khan Shafiq Ur, Islam Muhammd, et al., 2018; Metzler et al., 2009).

3.4 | Lexicon expansion

The words containing the tagged part of speech (*noun*, *verb*, *adverb*, and *adjective*) are retrieved using the part of speech tagging system. We utilized WordNet to extract synonyms, hyponyms, morphemes, and physical meanings for each recovered word element from the corpus, consisting of a series of texts. As a result, we directly obtain temporal keywords for each paragraph. The phrase used to train the model is then used to build the vocabulary. The word dimension denotes the newly obtained vector. The trained model converts the lexicons to vector format (pre-trained network). Cosine similarity is a method for calculating the similarity between the focus time and the creation time of the suggested quarry. We employ training to embed to turn texts into vectors with semantic awareness.

Each expression will appear in all three logical divisions for each instance-assigned content block. In order to compute the weights for each segment, the positional attention network is used to determine the total amount of words in 918 pertinent documents, considering their temporal

characteristics. Since the weight of a temporal expression reflects the segment's importance about the amount of information, segments with more information are given greater weight than those with less. After assigning a score to each temporal expression in the document, the scores are arranged in decreasing order based on the scores.

3.5 | Temporal focus time attention network

The shortcomings of supervised and unsupervised learning are resolved by semi-supervised learning. A few labelled samples are used to train a model and apply to unlabeled data. This approach uses both labelled and unlabeled data, reducing the cost and time of human annotation and data preparation. Using theory-based machine learning, it is possible to synthesize multimodal data at different levels for predictive purposes. Probabilistic formulations can also measure prediction uncertainty and motivate the collection of extensive evidence for dynamic model improvement. As shown in Figure 3, in the age of the digital twin, this integration offers new potential for incorporating data into predictive models that analyse emerging hazards and suggest tailored prevention recommendations and strategies. Humans are capable of recognizing and using shape qualities and contextual relationships. The ability of machine learning to use contextual information that is not accessible to humans could be the reason for its comparable or better performance and lower agreement with human readers in studies. Robust and reproducible machine learning can help overcome human biases and errors.

The active learning process is shown in Figures 1, 2, and 3. The proposed technique first trains a classifier (model) for each labelled dataset using the temporally pre-trained network (Almasian et al., 2021). The pseudo-labels are then applied to the broader pool of unlabeled data to create highly reliable pseudo-labels. Then, limited pseudo-labels based on the original labelled data are produced using the partially trained model (e.g., unequal representation of classes leading to bias). Then, the most reliable predictions of the model are selected. If any pseudo-labels meet this confidence level, they are added to the labelled dataset to train a more accurate model. Suppose the first classifier (attention network) correctly predicts the actual label for a data sample, and the second classifier makes an incorrect prediction. Then, the pseudo-labels of the first classifier are used to update the second classifier and vice versa. Each cycle adds more pseudo-labels (usually 10). If the data is accurate, the model improves with each repetition, as shown in Figure 2, where the sampling method is mentioned. The predictions of the two updated classifiers are integrated in the final stage. The proposed technique creates a labelled dataset from unlabeled inputs over multiple rounds.

3.6 | Attention network

The attention strategy offered makes use of the word meaning of the text, which can be found (Zichao et al., 2016). After adding the LSTM layer, the attention approach was created. This asset employed the extraction of valuable terms. The dropout layer requires the attention output vector as input to function correctly. Supervised learning typically requires a sizeable labelled dataset to train massive networks properly. Transfer learning was utilized to broaden the area of lexical analysis and to label the dataset. In this work, the lexicon is constructed utilizing the previously disclosed method. Attention-based contrast phrases were used to map the continuous vector to its related label. We utilized the attentional weights

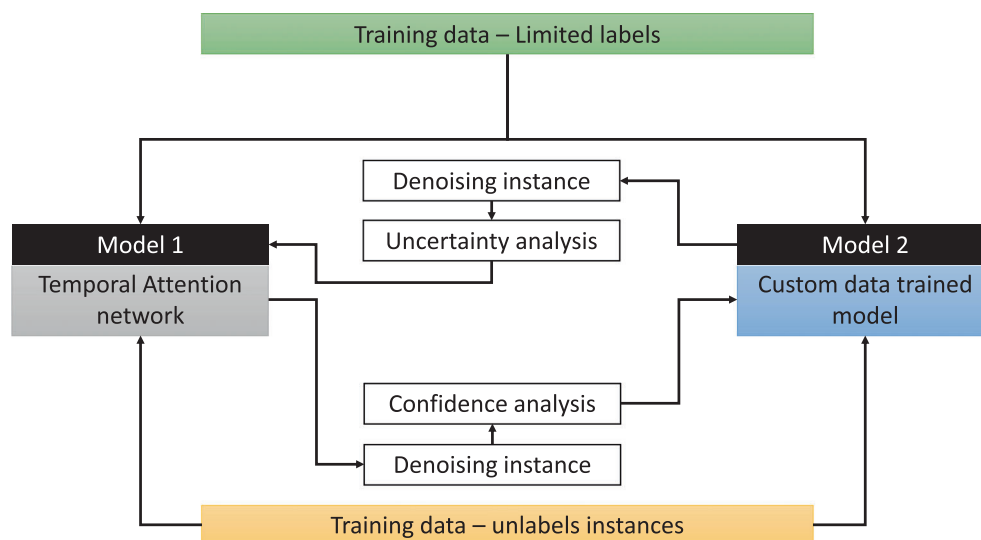


FIGURE 3 A workflow of the proposed semi-supervised learning method

for each labelled dataset to establish the classification of temporal focus throughout the year. In extraction, detection, and classification operations, attention-learning techniques create contrast sets.

4 | EXPERIMENTAL RESULT AND ANALYSIS

First, the content text from news articles is preprocessed, and then it is turned into a lexicon expansion. Then, we trained a few different networks. For transfer learning, we used a Glove network that had already been trained (Pennington et al., 2014). A new lexicon is added to the network so that the learned embedding can store a more extensive set of lines. Once this step is done, the text model is changed into sequential lexicons. Then, vectors are used to label the data that did not have labels before. Then, the labelled data is trained with several patterns and then compared. We used Precision, Recall, and F-measure to determine how well they did. We used an Adam optimizer to cut down on the time needed to train.

We demonstrated the effectiveness of our proposed method using a variety of pre-train models of varying sizes. Centralised (Ahmed, Lin Jerry, & Srivastava, 2022b), FedAVG (Li et al., 2019) and Embedded averages were used to evaluate the BERT model (Almasian et al., 2021). According to Figures 4, 5, and 6, the embedding model performs better since it is trained on the entire dataset. High performance is possible with our proposed lexicon.

The embedding averaging was handled effectively by the LSTM model, as mentioned in Figure 4. There are fewer data points to work with. Therefore, the centralized model is more tolerant of being trapped in its local optimum. The evaluation results should be close to centralized training since learning with embedding layers sharing is conceptually equivalent to centralized training. However, we discovered that BERT's learning rate began to decline, and the training process was halted. Because this data is not being used for training, these positive tasks worsen.

We designed and carried out the experiments to test if a large-scale contextual language representation model such as BERT can be pre-trained and fine-tuned utilizing data from several sources. We first looked at the cases when there was no pre-training BERT model for the domain in question. The original BERT base model checkpoints were utilized for training the parameters in these circumstances. To understand how various BERT models function, we examined how the attention head behaves. First, we calculated the entropy of the attention distribution

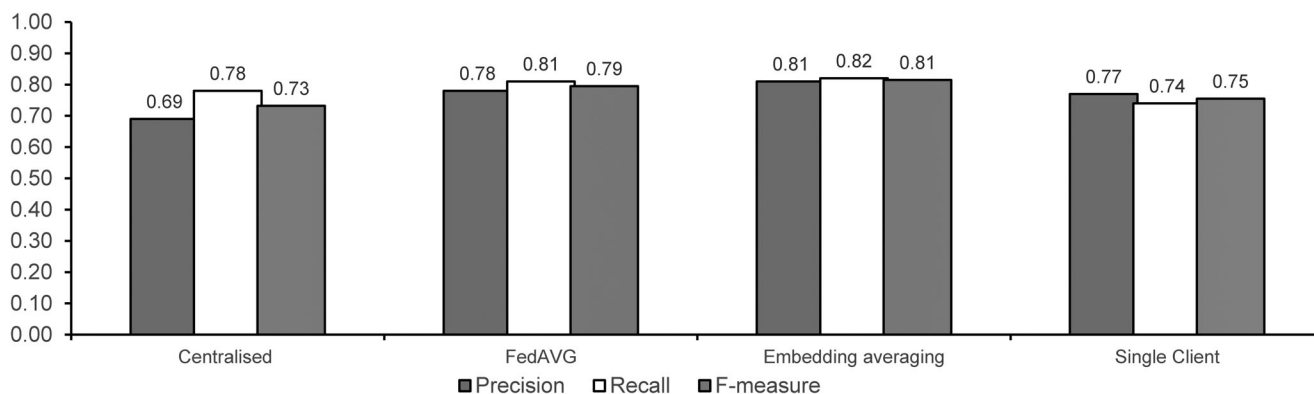


FIGURE 4 LSTM model with proposed lexicon-based denoising

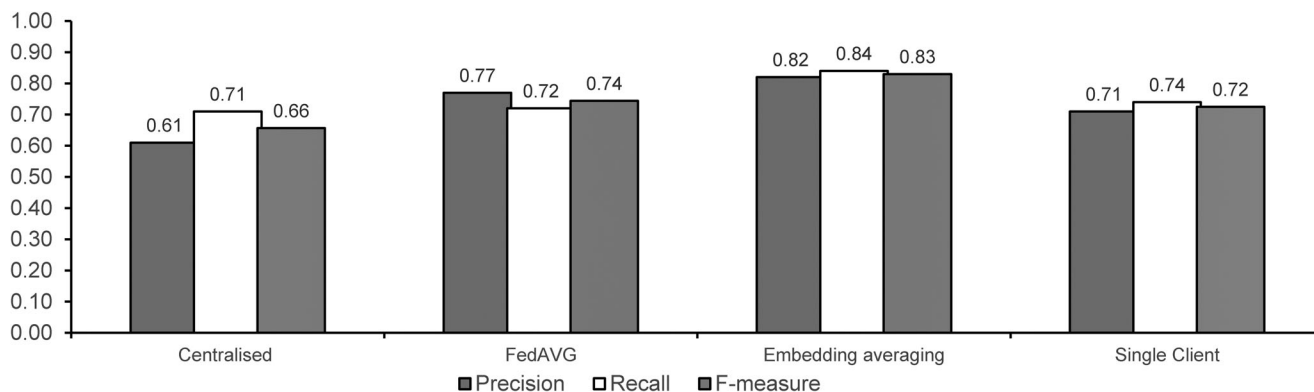


FIGURE 5 Bert model with proposed lexicon-based denoising

in each of the four heads. BERTbase and BERT federated training models were not fine-tuned for this investigation. Each model has a different rate at which entropy decreases. By analyzing the correlation of attention entropy between heads in each model, we discovered that the federated clinical BERT model is more similar to the BERTbase model than the centralized news BERT model.

For the experimental setting, we divided the dataset for clients, active learning, and the semi-supervised cycle is set to 10. We use the Jensen-Shannon distance as a unit of measurement. Multivariate comparisons of attention head divergence matrices from different models were used. The news-trained models are relatively similar, while the BERT basic model is significantly different.

Based on the results of our experiments, we found that training with embedded average federated data from multiple clients leads to better performance than training with centralized data, as shown in Figures 5 and 6. It is important to note that loss of federated communication is a phenomenon that reduces performance when data is stored in multiple places simultaneously. Pre-training on lexical analysis and fine-tuning the model for the central-level event detection tasks yielded few performance improvements on the downstream tasks. During fine-tuning, data were stored in a single location for the purpose of pre-training, while fine-tuning was performed in a decentralized manner. It was found that fine-tuning in a distributed manner resulted in a performance difference of less than 2%. When both pre-training and pre-tuning are performed independently, the results are less than satisfactory.

The average embedding technique generates a sentence vector representation as part of the federated learning method. This improves the performance compared to previous methods that rely only on feature averaging and feature embedding to achieve their results. As a result of the proposed method, sentence representations are trained using lexicon expansion and semi-supervised techniques to achieve good sentence representation. It has been shown that the semi-supervised method has the highest performance and the most significant gains by the developed method. This led to the conclusion that the proposed method could be extended as a data-driven approach to any application as a result of the research.

In Figure 7, we included autonomously produced and tagged data in the aforementioned training sets. With semi-supervised extended training sets, we evaluated whether the performance of the event extractor has improved after training these extended training sets. In this way, it is possible to demonstrate the effectiveness of the proposed technique in an automated manner. As a result of temporal expansion, distracting verbal triggers were eliminated and nominal triggers were expanded to clarify the task. By using the feature set to identify temporal information and adopting expressive neural techniques, event identification results are improved. Since there are significant differences between words, it is

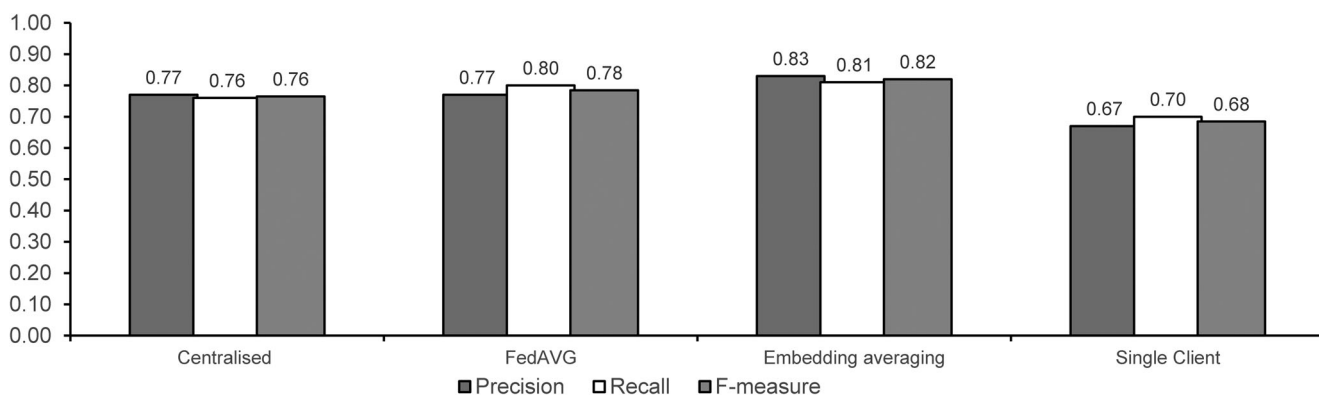


FIGURE 6 Bert fine-tune model with proposed lexicon-based denoising

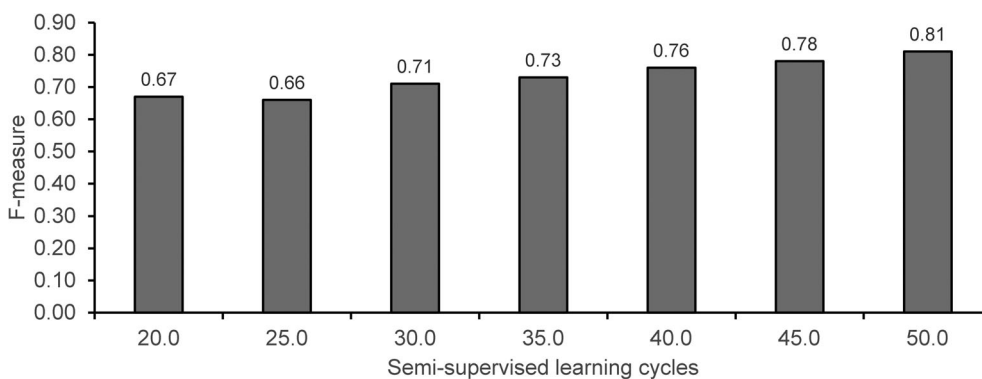


FIGURE 7 Semi-supervised learning cycles improvement

TABLE 2 Model comparison

Model	LSTM	BERT	Bert fine-tune + denoise
Centralised (Usman, Chun-Wei, & Gautam, 2022c)	0.66	0.73	0.76
FedAVG (Li et al., 2019)	0.74	0.79	0.78
Embedding averaging (Li et al., 2019)	0.83	0.81	0.82
Single client	0.72	0.75	0.68

obvious that the identified features will be useful for sentence expansion. This is in addition to the large differences between the standard and the extended lexicons.

The model comparison is described through Table 2. The centralized method used the weighted methods (Ahmed, Lin Jerry, & Srivastava, 2022b). Weights are assigned to nodes within a neighbourhood based on their properties or emotions without requiring complex matrix operations, such as similarity or architectural knowledge. However, our method required quality training instances. The FedAVG method uses communications broadcast and aggregation methods. Our method scheme converges at a slower speed even if its learning rate is fine-tuned (Li et al., 2019). The embedding was utilized averaging of the lexicon expansion to get closer to semantic meaning using the federated average method. Our model achieved the highest score of 0.83.

5 | CONCLUSION

Cyberspace has become a familiar and widespread way to spread false or misleading information. This is often called an “Info demic” in modern times, and it is a severe threat to human intelligence and actions. In this paper, we made a semi-supervised text fake news classification system based on temporal architecture, considering how hard and inconsistent it is to annotate web data. After the system has been trained using the title and body of news stories with different size BERT models, the retrieved feature vectors are added together. We show how lexicon expansion approaches using a variety and semantic exploration can be used to add syntactic data and improve hidden vectors for focal time retrieval. These methods can also be used to add more words to the lexicon. The proposed method uses a temporally pre-trained network and trains a classifier (model) for each labelled dataset. Then, high-confidence fake labels are added to the data pool that has not been labelled. Pseudo-labels are made from the partially trained model and the original labelled data. When a classifier (attention network) correctly predicts the label of one data sample, the pseudo-labels of the other data samples are changed, and vice versa. In the last step, the predictions made by the changed classifiers are put together. In this method, the unlabeled inputs are given labels repeatedly. Temporal expansion and profiling worked better when co-training and semi-supervised learning were used. This gave the proposed model a high BERT score of 81%. The suggested method will be used in the future for things like content-based news searches and time-relevant suggestions. We were able to get rid of noise at both the data level and model level by using pseudo-labels and the latest version of BERT. Compared to other baselines, our model with the noise taken out performed better than all of them. Given how expensive it is to manually sort false news data into classification, our plan is to use a smaller but clean dataset to make a generalized model. Using semi-supervised learning techniques, we will extend lexicon analysis in the future. We can adapt, improve, and extend our proposed method to other domains, such as audio and health applications.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in SemEval-2019 Task 4: Hyperpartisan News Detection at <https://aclanthology.org/S19-2145/>, reference number 35.

ORCID

Usman Ahmed  <https://orcid.org/0000-0002-3933-4273>

Jerry Chun-Wei Lin  <https://orcid.org/0000-0001-8768-9709>

Gautam Srivastava  <https://orcid.org/0000-0001-9851-4103>

REFERENCES

- Ahmed, U., Lin Jerry, C.-W., & Srivastava, G. (2022a). Hyper-graph attention based federated learning method for mental health detection. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1.
- Ahmed, U., Lin Jerry, C.-W., & Srivastava, G. (2022b). Social media multiaspect detection by using unsupervised deep active attention. *IEEE Transactions on Computational Social Systems*, pp. 1–9.
- Ahmed, U., Srivastava, G., & Lin Jerry, C.-W. (2021). A federated learning approach to frequent Itemset mining in cyber-physical systems. *Journal of Network and Systems Management*, 29(4), 1–17.
- Alonso Omar, R. (2008). *Temporal information retrieval*. University of California.

- Almasian, S., Aumiller, D., & Gertz, M. (2021). BERT got a date: Introducing transformers to temporal tagging. *arXiv preprint arXiv:2109.14927*.
- Ahmed, U., Srivastava, G., & Lin Jerry, C.-W. (2022). Reliable customer analysis using federated learning and exploring deep-attention edge intelligence. *Future Generation Computer Systems*, 127, 70–79.
- Ahmed, U., Lin Jerry, C.-W., Srivastava, G., & Fournier-Viger, P. (2021). A transaction classification model of federated learning (pp. 509–518). Springer.
- Baly, R., Karadzhov, G., Alexanedrov, D., Glass, J., & Nakov, P. (2018). Predicting factuality of reporting and bias of news media sources. *arXiv preprint arXiv:1810.01765*.
- Bhowmick Rajat, S., Ganguli, I., Paul, J., & Sil, J. (2021). A multimodal deep framework for derogatory social media post identification of a recognized person. *Transactions on Asian and Low-Resource Language Information Processing*, 21(1), 1–19.
- Cao, K., Li, X., Weicheng, M., & Grishman, R. (2018). Including new patterns to improve event extraction systems (pp. 487–492). AAAI Press.
- Charles, W. G. (2000). Contextual correlates of meaning. *Applied Psycholinguistics*, 21(4), 505–524.
- Chen, Y., Liu, S., Zhang, X., Liu, K., & Zhao, J. (2017). Automatically labeled data generation for large scale event extraction. In *Proceedings of the 55th annual meeting of the association for computational linguistics (Volume 1: Long Papers)*. (pp. 409–419). Association for Computational Linguistics.
- Clark, K., Luong, M.-T., Le Quoc, V., & Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.
- Deng, S., Zhang, N., Kang, J., Zhang, Y., Zhang, W., & Chen, H. (2020). Meta-learning with dynamic-memory-based prototypical network for few-shot event detection.
- Feng, X., Qin, B., & Liu, T. (2018). A language-independent neural network for event detection. *Science China Information Sciences*, 61(9), 1–12.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104.
- Filannino, M. (2016). Data-driven temporal information extraction with applications in general and clinical domains.
- Horne, B., & Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1), 759–766.
- Horne, B. D., Dron, W., Khedr, S., & Adali, S. (2018). Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news.
- Jain, R., Jain Deepak, K., & Sharma, N. (2021). Fake news classification: A quantitative research description. *Transactions on Asian and Low-Resource Language Information Processing*, 21(1), 1–17.
- Khan Shafiq Ur, R., Islam Muhammad, A., Aleem, M., Iqbal Muhammad, A., & Ahmed, U. (2018). Section-based focus time estimation of news articles. *IEEE Access*, 6, 75452–75460.
- Khan Shafiq Ur, R., Islam Muhammd, A., Aleem, M., & Iqbal Muhammad, A. (2018). Temporal specificity-based text classification for information retrieval. *Turkish Journal of Electrical Engineering & Computer Sciences*, 26(6), 2915–2926.
- Kumar, A., Esposito, C., & Karras, D. A. (2021). Introduction to special issue on misinformation, fake news and rumor detection in low- resource languages. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 21(1), 1–3.
- Kiesel, J., Mestre, M., Shukla, R., et al. (2019). *SemEval-2019 task 4: Hyperpartisan news detection* (pp. 829–839). Association for Computational Linguistics.
- Kim, A., Moravec, P. L., & Dennis, A. R. (2019). Combating fake news on social media with source ratings: The effects of user and expert reputation ratings. *Journal of Management Information Systems*, 36(3), 931–968.
- Laser, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096.
- Lai Viet, D., Nguyen, T. N., & Nguyen, T. H. (2020). Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics.
- Li, X., Huang, K., Wenhao, Y., Shusen, W., & Zhang, Z. (2019). On the convergence of FedAvg on non-IID data. *arXiv*.
- Metzler, D., Jones, R., Peng, F., & Zhang, R. (2009). Improving search relevance for implicitly temporal queries. In *Proceedings of the 32nd international ACM SIGIR conference on research and development in information retrieval* (pp. 700–701). ACM Press.
- Moravec, P. L., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media. *Information Systems Research*, 31(3), 987–1006.
- Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics.
- Radford, A., Wu, J., Child, R., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8), 9.
- Rosin, G. D., Guy, I., & Radinsky, K. (2022). Time masking for temporal language models (pp. 833–841). ACM.
- Suissa, O., Elmalech, A., & Zhitomirsky-Geffet, M. (2022). Text analysis using deep neural networks in digital humanities and information science. *Journal of the Association for Information Science and Technology*, 73(2), 268–287.
- Tian, Y., Wan, Y., Lyu, L., Yao, D., Jin, H., & Sun, L. (2022). FedBERT: When federated learning meets pre-training. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(4), 1–26.
- Wang, J., Jatowt, A., & Yoshikawa, M. (2021). Event occurrence date estimation based on multivariate time series analysis over temporal document collections.
- Yang, H., Chen, Y., Liu, K., Zhao, J., & Wang, T. (2021). Multi-sentence argument linking via an event-aware hierarchical encoder. In *Proceedings of the 30th acm international conference on information & knowledge management* (pp. 3578–3582). ACM.
- Yao, D., Pan, W., Wan, Y., Jin, H., & Sun, L. (2021). FedHM: Efficient federated learning for heterogeneous models via low-rank factorization. *arXiv*.
- Yeh, J.-F., Su, Chen, W.-Y., & Su, M.-C. (2015). Chinese spelling checker based on an inverted index list with a rescoring mechanism. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 14(4), 1–28.
- Zichao, Y., Diyi, Y., Dyer, C., Xiaodong, H., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification (pp. 1480–1489). Springer Science and Business Media (LLC).

AUTHOR BIOGRAPHIES

Usman Ahmed is a data scientist and PhD candidate at the Western Norway University of Applied Sciences (HVL). As a research scientist, he has published more than 50 research papers in peer-reviewed journals with more than 300 citations on Google Scholar (H-index 11, I-10

index 12). He has rich experience building and scaling high-performance systems based on data mining, natural language processing, and machine learning. The research interests of Usman include sequential data mining, heterogeneous computing, uncertainty analysis, semi-supervised learning, natural language processing, recommendation systems, and machine learning.

Jerry Chun-Wei Lin (Senior Member, IEEE) received the PhD degree from the Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently a Full Professor with the Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway. He has published more than 500 research papers in refereed journals and international conferences. His research interests include data mining, soft computing, artificial intelligence and machine learning, and privacy preserving and security technologies. Dr. Lin is a fellow of IET (FIET) and an ACM Distinguished Member (the Scientist). He is the Editor-in-Chief of the *International Journal of Data Science and Pattern Recognition*, and an Associate Editor of several top-tier journals, including *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, *IEEE Transactions on Cybernetics (TCYB)*, *IEEE Transactions on Dependable and Secure Computing (TDSC)*, and *Information Sciences*.

Gautam Srivastava (Senior Member, IEEE) received the BSc degree from Briar Cliff University, Sioux City, IA, USA, in 2004, and the MSc and PhD degrees from the University of Victoria, Victoria, BC, Canada, in 2006 and 2012, respectively. In 2014, he joined a tenure-track position at Brandon University, Brandon, MB, Canada, where he is currently active in various professional and scholarly activities. He was promoted to the rank of an associate professor in January 2018. As he is popularly known, he is active in research in the field of cryptography, data mining, security and privacy, and blockchain technology. In his 5 years, as a research academic, he has published a total of 200 papers in high-impact conferences in many countries and in high-status journals (SCI, SCIE) and has also delivered invited guest lectures on big data, cloud computing, the Internet of Things, and cryptography at many universities worldwide. Dr. Srivastava is an editor of several SCI/SCIE journals. He has active research projects with other academics in Taiwan, Singapore, Canada, Czech Republic, Poland, and USA. He is an Associate Editor of the world-renowned *IEEE Access Journal*.

How to cite this article: Ahmed, U., Lin, J. C.-W., & Srivastava, G. (2022). Temporal positional lexicon expansion for federated learning based on hyperpartism detection. *Expert Systems*, e13183. <https://doi.org/10.1111/exsy.13183>