A co-training model based in learning transfer for the classification of research papers

Alex Cevallos-Culqui
LIFIA - Facultad de Informática
Universidad Nacional de La Plata
La Plata, Argentina
alex.cevallosc@info.unlp.edu.ar

Claudia Pons
LIFIA - Facultad de Informática
Universidad Nacional de La Plata
CAETI - Facultad de Tecnología Informática
Universidad Abierta Interamericana
Buenos Aires, Argentina
claudia.pons@uai.edu.ar

Gustavo Rodríguez
Departamento de Tecnologías de Información y Comunicación
UTC
Latacunga, Ecuador
gustavo.rodriguez@utc.edu.ec

Abstract—**A multitude of scholarly papers can be accessed online, and their continual growth poses challenges in categorization. In diverse academic fields, organizing these documents is important, as it assists institutions, journals, and scholars in structuring their content to improve the visibility of research. In this study, we propose a co-training model based on transfer learning to classify papers according to institutional research lines. We utilize co-training text processing techniques to enhance model learning through transformers, enabling the identification of trends and patterns in document texts. The model is structured with two views (titles and abstracts) for data preprocessing and training. Each input employs different document representation techniques that augment its training using BERT's pre-trained scheme. For evaluating the proposed model, a dataset comprising 898 institutional papers is compiled. These documents undergo classification prediction in five or eleven classes, and the model performance is compared with individually trained models from each view using the BART pre-trained scheme and combined models. The best precision level of 0,87 has been achieved, compared to BERT pre-trained model's metric of 0,78 (five classes). These findings suggest that co-training models can be a valuable approach to improve the predictive performance of text classification.**

Keywords - text classification; co-training; transformer; pre-trained

I. INTRODUCTION

The large amount of digital information and the centralization of documents in digital repositories generate the need to organize and classify a large number of documents [1]. In the academic field, the correct classification of documents is crucial to improve accessibility to relevant information that facilitates the exchange of knowledge [2]. The automation of this classification would facilitate the search and retrieval of information by researchers, as well as the identification of patterns in scientific [3] production that would support decision making in the research field of organizations [4].

For adequate automation of document classification, a set of labeled documents is required to support the training of the model; the greater the number of labeled documents available, the better the effectiveness of the classification [5]. However, the availability of these labeled documents is scarce due to limited resources and time [6]. For this reason, co-training and transfer strategies are effective in addressing this problem, as they allow for the improvement of label generation through self-labeling and the use of pre-trained data sets [7].

Both models minimize reliance on extensive labeled datasets. In the case of co-training, combining multiple views with different perspectives and decisions helps mitigate the influence of poorly labeled data by reducing noise, bias, and variance in the final model, this improves the labeling process even with a limited number of labels [8]. On the other hand, transfer models benefit from the learning derived from previously labeled datasets in pre-trained models. Among these types of architectures are BERT and BART, which consistently enhance their qualities of pre-trained data volume and processing speed through transformers, these factors contribute to reducing the dependency on labeled data [9].

In recent years, there has been significant progress in the development of co-training and transfer models, in the analysis of the literature review on learning models in [10], it is evident that among the most notable models in terms of their precision levels for a sample of works focused on document classification are the co-training (0,79) and transfer models (0,79). In this study, these models are applied to datasets composed mainly of user review texts, with an average of five classes to be classified.

The classification of documents through co-training or transfer models in independent configurations is a topic of study in the field of research, however, there are limited studies with experiments in which the two models are configured in fusion to achieve the classification [11]. Several studies have been conducted on these two models individually. For instance, in [12] a co-training model is constructed using an MDDRL (Multi-view Deep Discriminant Representation Learning) architecture with two views. The model engages in data processing, representing documents from news web pages using orthogonality constraint extraction techniques. This framework supports multiple views, each one receiving specific and shared treatment during training, prediction, and classification. The views are co-trained to establish a labeling process through neural networks.

In [13], a co-training model is proposed for processing product review documents using two views: fine resolution and coarser resolution. The first relies on representations of characteristics with a low level of granularity, such as individual words; while the second receives representations with a higher

979-8-3503-5098-2/24/$31.00 ©2024 IEEE

2024 IEEE 12th International Conference on Intelligent Systems (IS) | 979-8-3503-5098-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/IS61756.2024.10705226

Authorized licensed use limited to: Universidad Abierta Interamericana. Downloaded on April 16,2025 at 17:18:29 UTC from IEEE Xplore. Restrictions apply.

level of granularity, such as words sets or complete phrases. Both undergo training and to integrate the various resolutions, align instances, and predict labels, attention mechanisms are employed. While the model effectively utilizes a variety of document representations, it tends to overfit due to the abundance of elements in the fine resolution, resulting in a loss of performance.

In the realm of document classification models employing transfer learning, we have [14], which utilizes a structure combined with generative models (DGMs) via the NX-VAE decoder (Variational autoencoder) and extracts knowledge from the pre-trained BERTW to classify news documents into four categories. Another study is [15], in this model an ontology is built to organize the classification categories in a hierarchical structure. Documents from Google Scholar are collected and processed, with text extracted using n-grams. This set of features is then trained and labeled using zero-shot classification pipeline, which is a classifier that use the pre-trained BART transformer to label documents according to the categories defined by the ontology.

Several studies have explored document classification in co-training and transfer learning environments. However, there is limited research specifically focusing on classifying research papers using co-training models with pre-trained data sources [11]. In the case of transfer models, the utilization of zero-shot classification pipeline is important [16], this multiclass classification method involves receiving a set of labels that are then inputted into the pre-trained BART model to classify a predetermined text. The limitations of this technique include its high sensitivity to noise, in both, the categories and the text being classified. When descriptions are ambiguous, precision in the classification is lost.

For this reason, the present study aims to design a model that enhances the performance of research paper classification through a combination of co-training and transfer learning, with two views, which extract knowledge from pre-trained models such as BERT using Transformer architectures. For training, the first view utilizes the text from the research paper titles, while the second view employs the text from the document reviews. To evaluate the model's performance, the proposal is compared with the performance of zero-shot classification pipeline, which employs a transfer model to extract pre-trained knowledge from BART using a Transformers scheme. It is anticipated to advance the understanding of research paper classification amidst the challenges of enhancing the training efficiency of co-training models. This involves improving their ability to handle large, complex, heterogeneous, and unbalanced datasets, while also mitigating noise and errors in labeled data.

II. PROPOSED APPROACH

A co-training model is also referred to as a multi-view model because it adopts a training structure with diverse approaches to its characteristics [11]. Labeled documents can be treated specifically or shared between views to reinforce the training process [12]. The evaluation of predictions becomes a more complex process due to the diversity of existing views and decision strategies [17]. Accordingly, the model's structure is examined. Figure 1 illustrates the operational scheme outlining the stages of the learning process of a co-training model.

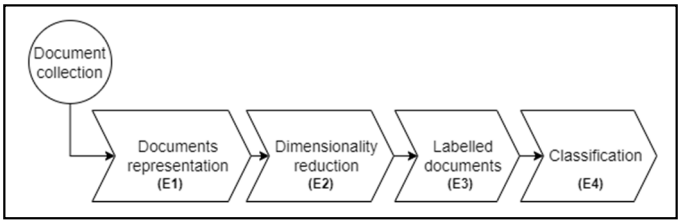


Figure 1. Learning process of a co-training model

On the other hand, transfer models allow us to take advantage of the knowledge acquired from a source domain (SD). That is, there is pre-trained learning that can be transferred and adapted to a target domain (TD) for the training of a new model with a set of reduced labeling [10]. This transfer learning is conducted using Transformer structures, which receive input text to be processed with embeddings, encoders, decoders, and pre-trained data to generate an output (Figure 2).

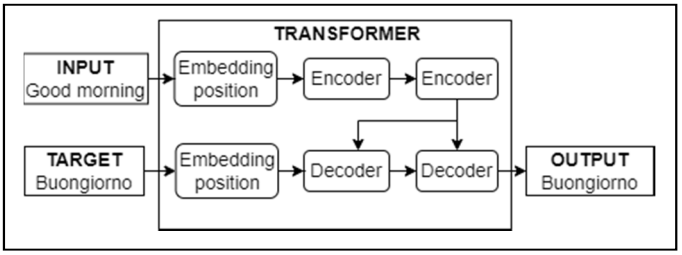


Figure 2. Learning process of a co-training model

This section outlines the structure of the proposed co-training and transfer model combined (COTRA). The model integrates two views to enhance its training through transfer learning, supported by a transformer structure with pre-trained BERT data. The first view is trained using the titles of research papers, while the second view is trained using the reviews. The training of both views is coordinated to select the best predictions. An overview of the model can be seen in Figure 3.

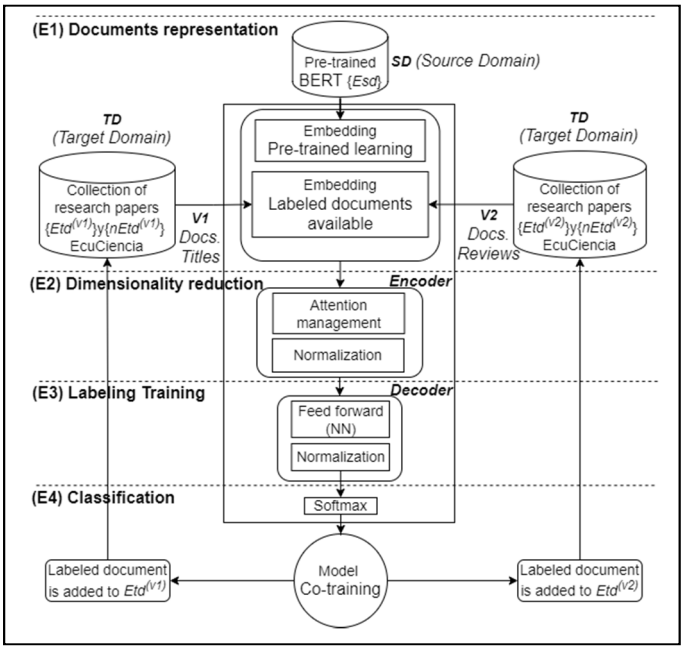


Figure 3. Structure of the COTRA model

Authorized licensed use limited to: Universidad Abierta Interamericana. Downloaded on April 16,2025 at 17:18:29 UTC from IEEE Xplore. Restrictions apply.

The two views proposed follow the classification process of co-training models. Additionally, when considering the pre-trained BERT dataset, the transformer offers the following advantages for both views:

- *Documents representation*. - The BERT structure enables bidirectional tokenization of both the titles and reviews of research papers. For their representation, the transformer utilizes embeddings that by their bidirectionality, generate an attention mechanism. This mechanism establishes semantic relationships between tokens, enabling the model to focus on the textual meaning's relevance.
- *Dimensionality reduction*. - The encoders adjust the dimensionality of the feature representation through a normalization process. This ensures compatibility between input and output ranges, thereby enhancing the training process.
- *Labeled.* - The training process utilizes the labeled documents available in the TD. BERT's pre-trained knowledge is incorporated through decoders, and both sets of knowledge are merged and trained in a neural network.
- *Classification*. - The decoders have softmax as their last layer, which has the function of distributing probabilities in labels set to define which is the most suitable category for the input text and define the classification task.

Generally, text processing is sequential. In this model, text processing is parallel, which has enabled the parallelization of the training process and can be executed using GPUs (Graphics processing unit). This allows for better resource availability and accelerates the process.

As input, COTRA receives: both views (title and reviews) $\{v\}=\{(v_i)|i=1,2\}$, the set $\{Esd\}=\{(d_i)|i=1,...,n\}$ containing documents labeled with BERT pre-training (SD), the set $\{Etd^{(v1)}\}=\{(d_j)|j=1,...,n\}$ containing view 1 with the labeled document titles from TD, the set $\{Etd^{(v2)}\}=\{(d_j)|j=1,...,n\}$ contains view 2 with the labeled documents reviews from TD, the set $\{nEtd^{(v1)}\}=\{(d_k)|k=n+1,..,n+m\}$ contains unlabeled documents from view 1, the set $\{nEtd^{(v2)}\}=\{(d_l)|l=n+1,..,n+m\}$ contains unlabeled documents from view 2, the set $\{Cl\}=\{(c_k)|k=1,...,n\}$ contains the distribution of classes or labels, two document representation strategies dr_1(SD), dr_2(TD) and the classifier C.

The details of the COTRA performance method are presented in Algorithm 1. At first, documents are represented based on their text type (title or reviews) from $\{Esd\}$, $\{Etd\}$ and $\{nEtd\}$ into two sets of features represented by embeddings. After defining the representation, the classifier is trained using the set of labeled documents and the pre-trained BERT dataset. Next, the trained classifier is applied to the unlabeled documents to predict their distribution class label $\{Cl\}$. The training of the two views is shared to retrain the model with a larger set of labeled documents. As a result, the model gradually accumulates a greater number of labels, enhancing its prediction performance.

Algorithm 1: COTRA Co-Training Model

Input:

$\{v1, v2\}$ → Instances per view;
$\{Esd\}=\{(d_i)|i=1,...,n\}$ → Documents labeled SD;
$\{Etd^{(v1)}\}=\{(d_j)|j=1,...,n\}$ → Documents (titles) labeled TD;
$\{Etd^{(v2)}\}=\{(d_j)|j=1,...,n\}$ → Documents (reviews) labeled TD;
$\{nEtd^{(v1)}\}=\{(dj)|j=n+1,...,n+m\}$ → Documents unlabeled v1;
$\{nEtd^{(v2)}\}=\{(dj)|j=n+1,...,n+m\}$ → Documents unlabeled v2;
$\{Cl\}=\{(c_k)|k=1,...,n\}$ → Class-Label list;
$\{dr\}$ → embeddings;
$\{C\}$ → NN;
$\{dr^{(v1)}, dr^{(v2)}\}$ → Document representations;
$\{C^{(v1)}, C^{(v2)}\}$ → Classifiers;

1: ***repeat***
2: $dr^{(v1)}$ and $dr^{(v2)}$, in each view for the documents of SD → $Esd^{(v)}$, TD → $Etd^{(v)}$ and $nEtd^{(v)}$;
3: Training the classifier of each view, $C^{(v1)}$ and $C^{(v2)}$, with the documents labeled $Esd^{(v1)}$, $Etd^{(v1)}$ and $Esd^{(v2)}$, $Etd^{(v2)}$;
4: Classify $nEtd^{(v1)}$ and $nEtd^{(v2)}$ with the classifiers, $C^{(v1)}$ and $C^{(v2)}$, assigning a label Cl;
5: The documents that are labeled and classified by $C^{(v1)}$ and $C^{(v2)}$ are added to the set of labeled documents, $Etd^{(v1)}$ and $Etd^{(v2)}$;
6: The labeled documents are removed from $nEtd^{(v1)}$ and $nEtd^{(v2)}$;
7: ***until*** $\{nEtd^{(v1)}\}$ && $\{nEtd^{(v2)}\} = \emptyset$

Output: Documents labeled $\{E\}$

III. EXPERIMENTS

For the evaluation of COTRA, an experimental setup was designed using Python and the PyCharm framework. This implementation is part of the EcuCiencia scientific platform (ecuciencia.utc.edu.ec), which aggregates scientific documents from researchers at the Universidad Técnica de Cotopaxi (UTC).

In this study, we evaluate the classification efficiency of the proposed model using precision performance metrics. The dataset comprises titles and reviews of research papers collected from the EcuCiencia platform. Table 1 shows information about the EcuCiencia dataset, which consists of 898 research papers collected from the repository. We select 90% of the data for training, 5% for validation and 5% for testing. Additionally, we select 10%, 20% and 30% of the training data as labeled data (when calculating the percentage of documents, it is approximated to the nearest higher or lower number).

TABLE I. DATASET FEATURES AND PARAMETERS

Source	*Dataset*	*Class*	*10%*	*20%*	*30%*
EcuCiencia	810 docs. (90% training)	5	80*	165*	245*
	44 docs. (5% validation)				
	44 docs. (5% testing)	11	77*	165*	242*

*Labeled documents in the training set

Authorized licensed use limited to: Universidad Abierta Interamericana. Downloaded on April 16,2025 at 17:18:29 UTC from IEEE Xplore. Restrictions apply.

These documents belong to five UTC departments, representing class numbers at a first level (C). These class labels are C1: Department of Agricultural Sciences, C2: Department of Environmental Sciences, C3: Department of Engineering and Applied Sciences, C4: Department of Administrative and Economic Sciences, C5: Department of Social Sciences, Arts and Education. Each department hosts various research lines, amounting to eleven categories, representing class numbers at a second level (SC). The experimentation of this model involves the utilization of both class sets C and SC (Table 2) for the individual training of co-training and transfer models with each view, in comparison with combined models.

TABLE II. LIST OF CLASSES AND SUBCLASSES

C	*SC*	
C1	SC1	Development and food security
	SC2	Animal health
C2	SC3	Analysis, conservation, and use of local biodiversity
	SC4	Planning and management of sustainable tourism
C3	SC5	Alternative and renewable energies, energy efficiency, and environmental protection
	SC6	Industrial processes
	SC7	Information and communication technologies
C4	SC8	Administration and economy for human and social development
	SC9	Quality management and occupational safety
C5	SC10	Education, communication and graphic design for human and social development
	SC11	Culture, heritage and ancestral knowledge

Table 3 presents the experimental characteristics for the proposed classification model. The performance of COTRA is compared with the individual performance of the views by title (V1) and by reviews (V2). Also, it is compared with a transfer model that uses zero-shot classification pipeline with the BART transformer, which incorporates a set of pre-trained data and complements its training with the titles (PIP1) and reviews (PIP2) of research papers. Finally, it is compared with the co-training of PIP1 and PIP2, as well as with the DGMs model [14] that combines generative models and transfer learning to extract knowledge from the pre-trained BERTW.

TABLE III. EXPERIMENTAL MODEL FEATURES

Id	*Model*	*Dataset*	*Docs. Repr.*	*Classifier*
V1	view	title	embeddings	NN
V2	view	Review		
PIP1	pipeline	Title		
PIP2	pipeline	Review		
COPIP	Co-training	title/review		
DGMs	Combined	title/review		
COTRA	Co-training	title/review		

Figure 4 depicts the workflow of preprocessing and representation of the titles and reviews of research papers. The figure shows how the data preparation process is carried out, which includes various stages such as text tokenization, removal of unwanted characters or irrelevant words, and the normalization of terms. Afterward, the document representation stage demonstrates the use of embedding techniques to convert titles, reviews, and pre-trained BERT data into numerical vectors.

After representing the documents, the model needs to identify the position of the tokens and their relationships, which is achieved through positional encoding. Furthermore, before the training process, the attention mechanism is utilized to assess the level of relevance between the tokens to enhance the semantics of the model. It is an essential vector representation for applying classification algorithms and categorizing research papers into established classes.

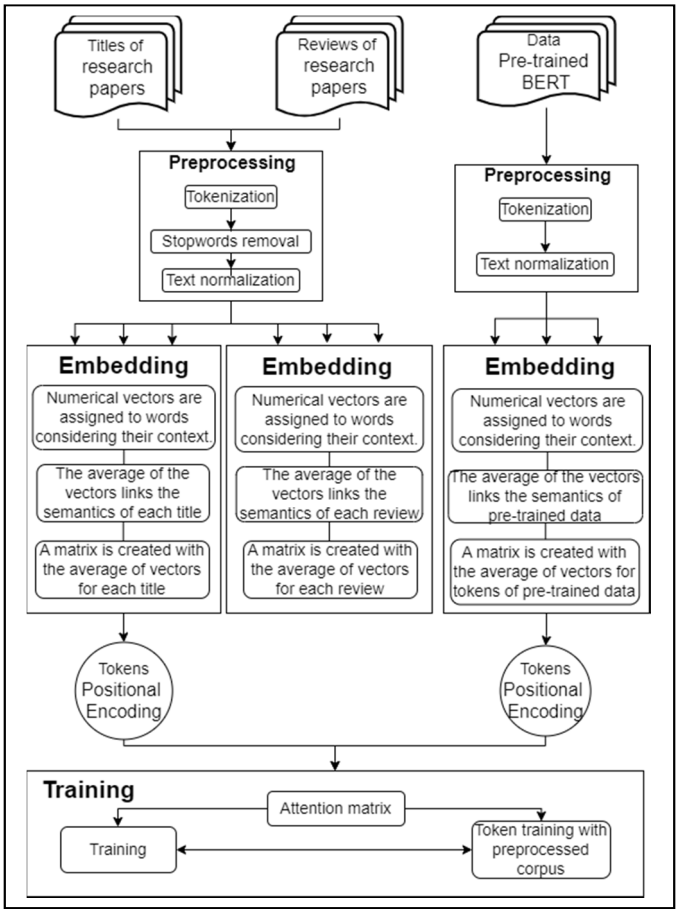


Figure 4. Preprocessing and documents representation of model

Table 4 shows the values of the configuration parameters for the algorithms used in the models under study. In models V1, V2, and COTRA, configuration parameters are shared. To prevent overfitting of the model, a dropout rate of 50% is set. In the Linear layer, the output characteristics of the BERT transformer, which are 768, are configured to be adapted to [5/11] dimensions corresponding to the number of classes. The softmax activation function is defined to distribute the probability of each class proportionally to 1. On the other hand, in the PIP1, PIP2 and COPIP models, their parameters are configured to utilize the zero-shot classification pipeline with

Authorized licensed use limited to: Universidad Abierta Interamericana. Downloaded on April 16,2025 at 17:18:29 UTC from IEEE Xplore. Restrictions apply.

the transformer employing the pre-trained BART dataset. DGMs use two latent spaces, z1 and z2, each with 768 dimensions, for feature representation.

TABLE IV. ALGORITHM PARAMETER CONFIGURATION

Algorithm	*Configuration*
V1	*Parameters:* dropout=0.5; nn.Linear(768, [5/11]); nn.LogSoftmax(dim=1). *Hyperparameters:* padding='max_length'; max_length= 20; truncation=True; return_tensors='pt'; batch_size=16; Epochs=10; LR = 5e-5.
V2	*Parameters:* dropout=0.5; nn.Linear(768, [5/11]); nn.LogSoftmax(dim=1). *Hyperparameters:* max_length= 150; truncation=True; return_tensors='pt'; batch_size=16; Epochs=15; LR = 5e-5.
PIP1	*Parameters:* task=classification; model="bart-large-mnli". *Hyperparameters:* padding='max_length'; max_length= 20; truncation=True; return_tensors='pt'.
PIP2	*Parameters:* task=classification; model="bart-large-mnli". *Hyperparameters:* padding='max_length'; max_length= 150; truncation=True; return_tensors='pt'.
COPIP	*Parameters:* dropout=0.5; nn.Linear(768, [5/11]); task=classification; model="bart-large-mnli". *Hyperparameters:* max_length= 150; truncation=True; return_tensors='pt'; batch_size=16; Epochs=7; LR = 5e-5.
DGMs	*Parameters:* dropout=0.5; classifier (768, [5/11]); z1_dim=768; z2_dim=768. *Hyperparameters:* padding='max_length'; max_length= 200; truncation=True; batch_size=4; Epochs=50; LR = 2e-5; vocabulary_size=1e-5; tie_embedding=true.
COTRA	*Parameters:* dropout=0.5; nn.Linear(768, [5/11]); nn.LogSoftmax(dim=1). *Hyperparameters:*max_length= 150; truncation=True; return_tensors='pt'; batch_size=16; Epochs=5; LR = 5e-5.

Regarding the models hyperparameters, they are presented in Table 4. V1, V2, PIP1, PIP2, COPIP and COTRA maintain a similar configuration with their tensor type defined as 'pt' (return_tensors='pt') after tokenization or processing, utilizing PyTorch. The batch size for training the models is set to batch_size=16 with a learning rate of LR = 5e-5. For tokenization, it is established to process the text by sequence sizes (padding='max_length'). If it exceeds its limit, it is truncated (truncation=True). For the models that process the titles of the documents (V1 and PIP1) a max_length=20 is set, and for the models that process reviews (V2 and PIP2), a max_length=150 is defined. Regarding the epochs for the neural network, in V1, optimal training is obtained with Epochs=10, for V2, due to processing a greater amount of text in the reviews, it is defined as Epochs=15, while COTRA has better processed text and requires Epochs=5.

IV. EVALUATION AND RESULTS

The performance measure used to evaluate the classification of the dataset is precision. This metric is obtained by relating true positives (TP) and false positives (FP) as follows: TP / (TP + FP). The classification performance of research papers from the EcuCiencia platform has been evaluated across 11 research lines using this metric.

Table 5 presents the precision values of the classification models V1, V2, PIP1, PIP2, COPIP, DGMs and COTRA trained with different percentages (10%, 20%, 30%) of labeled documents, for classification into 5 and 11 classes. The precision of the V1 model, which utilizes the BERT architecture and trains with document titles, ranges from 0.68 (10%) to 0.73 (30%) for five classes (C), and from 0.35 (10%) to 0.64 (30%) for eleven classes (SC). The V2 model, which also uses BERT and trains with document reviews, exhibits the third highest precision, ranging from 0,71 to 0,78 for C and from 0,48 to 0,69 for SC.

Regarding the PIP1 model, trained with the document's titles in a BART structure, it has a precision ranging from 0,71 to 0,78 for C and from 0,30 to 0,45 for SC, while the PIP2 model, also trained with BART using the document reviews, exhibits a precision of 0,51 to 0,61 for C and 0,32 to 0,48 for SC. With the combined models, COPIP ranges from 0.62 to 0.71 for C, and from 0.4 to 0.6 for SC, while that DGMS presents the second highest precision ranging from 0,63 to 0,82 for C and from 0,52 to 0,73 for SC. Finally, it can be observed that the COTRA co-training model exhibits the highest precision among all the models, with values ranging from 0,77 to 0,87 for C and from 0,56 to 0,79 for SC.

TABLE V. PRECISION VALUES OF THE MODELS

	C (5 classes)			*SC (11 classes)*		
Model	***10%***	***20%***	***30%***	***10%***	***20%***	***30%***
V1	0,68	0,7	0,73	0,35	0,50	0,64
V2	0,71	0,75	0,78	0,48	0,65	0,69
PIP1	0,48	0,53	0,57	0,30	0,34	0,45
PIP2	0,51	0,54	0,61	0,32	0,41	0,48
COPIP	0,62	0,65	0,71	0,4	0,47	0,6
DGMs	0,63	0,71	0,82	0,52	0,62	0,73
COTRA	**0,77**	**0,83**	**0,87**	**0,56**	**0,70**	**0,79**

The comparison of COTRA's performance is illustrated with various traditional individual models (Figure 5) and combined models (Figure 6), considering different percentages of labeled documents and numbers of classes. The precision of all models increases as the size of the labeled document set increases. It is also noted that precision increases when the number of classes is fewer. Furthermore, the curve corresponding to the COTRA model exhibits the highest precision across all labeling percentages, followed by the curves for DGMs and V2. In contrast, the curve corresponding to the PIP models demonstrates the lowest precision compared to the other models.

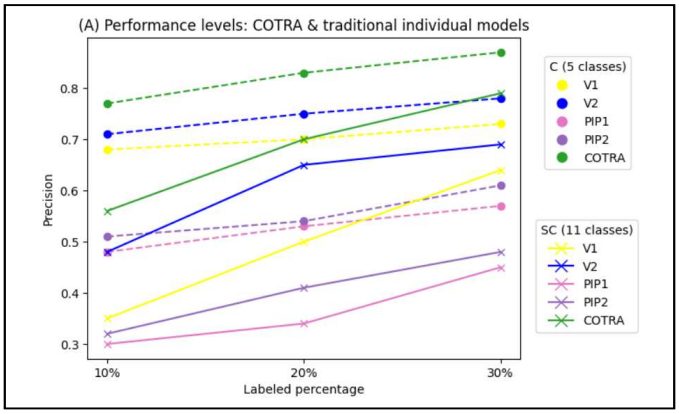


Figure 5. Performance levels with traditional individual models

Authorized licensed use limited to: Universidad Abierta Interamericana. Downloaded on April 16,2025 at 17:18:29 UTC from IEEE Xplore. Restrictions apply.

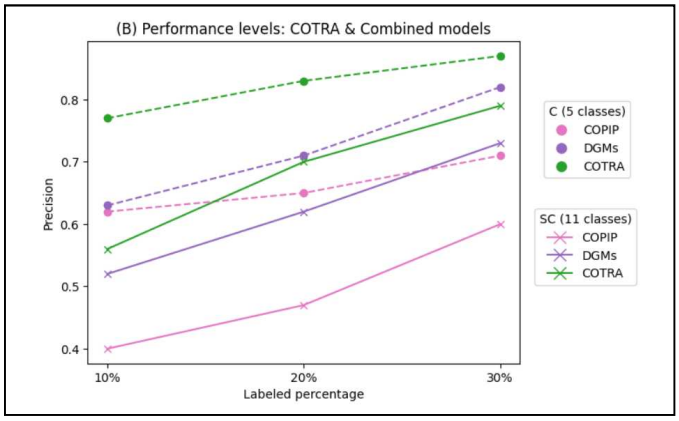


Figure 6. Performance levels with combined models

V. CONCLUSIONS AND FUTURE WORK

In the present study, the COTRA co-training model has been proposed to enhance the classification efficiency of research papers. Considering that the titles and reviews of papers consist of unstructured data, this model architecture shares two views with different features from dataset, that enriching the documents representation and their training. In addition, the model's training structure has been enhanced with pre-trained data from the BERT and BART transformers. This redundancy of co-training and transfer learning has reduced the error margin in labeled documents, resulting in more efficient training. The label compartment in the co-training and transfer schemes mitigates the challenge of the scarcity of labeled documents.

The experimental results have confirmed that the classification efficiency of the proposed COTRA model achieved the best precision metrics in all cases, outperforming individual training for each view, zero-shot classification pipeline models and combined models (V1, V2, PIP1, PIP2, COPIP, DGMs). Different percentages of labeled documents were used to train the model, ranging from 10% to 30%, with the objective of analyzing the model's behavior as the number of labeled documents increases. In a general context, it can be observed that the significant increase in the precision of the model is directly proportional to the increase in the percentage of labeled documents used for training. This demonstrates the importance of increasing the number of labeled documents for training the model, improving its classification performance.

However, COTRA has its limitations; the redundancy of labeled documents among views and pre-trained sets results in a model with high computational resource requirements. As future work, the model's efficiency could be enhanced by developing an automated space to better distinguish specific labels from shared ones. Additionally, other types of learning could be explored to improve classification efficiency. For example, exploring model ensemble techniques could improve prediction management, leading to enhanced overall model efficiency. It is also important to evaluate the model in comparison with the performance of other co-trained models for document classification. Furthermore, it is suggested that the use of COTRA could be useful not only in the classification of research papers but also in other application domains.

REFERENCES

[1] A. Q. Wynne Harlen OBE, *The Teaching of Science in Primary Schools*, Sixth. England: Routledge, 2014.

[2] N. Pulido and A. Mata, *El documento digital: aspectos para garantizar su integridad por la ciudadanía*, Astro data. Potosí, México, 2022. [Online]. Available: https://www.researchgate.net/publication/365839063

[3] A. Bhavani and B. Santhosh, "A Review of State Art of Text Classification Algorithms," *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, no. Iccmc, pp. 1484–1490, 2021, doi: 10.1109/ICCMC51019.2021.9418262.

[4] S. Franko, "Multiclass analysis of automatic text classification techniques," Galatasaray University, 2018.

[5] V. Prabhu, S. Khare, D. Kartik, and J. Hoffman, "SENTRY: Selective Entropy Optimization via Committee Consistency for Unsupervised Domain Adaptation," *Computer Vision Foundation*, 2021, [Online]. Available: https://github.com/virajprabhu/SENTRY.

[6] K. Watanabe and Y. Zhou, "Theory-Driven Analysis of Large Corpora: Semisupervised Topic Classification of the UN Speeches," *Soc Sci Comput Rev*, pp. 1–21, 2020, doi: 10.1177/0894439320907027.

[7] M. Mirończuk and J. Protasiewicz, "A recent overview of the state-of-the-art elements of text classification," *Expert Syst Appl*, vol. 106, pp. 36–54, 2018, doi: 10.1016/j.eswa.2018.03.058.

[8] A. K. Shrivas, A. K. Dewangan, S. M. Ghosh, and D. Singh, "Development of proposed ensemble model for spam e-mail classification," *Information Technology and Control*, vol. 50, no. 3, pp. 411–423, 2021, doi: 10.5755/j01.itc.50.3.27349.

[9] A. Mustar, S. Lamprier, B. Piwowarski, and B. Piwowarski Using BERT, "Using BERT and BART for Query Suggestion," Samatan, París, 2020. [Online]. Available: https://hal.sorbonne-universite.fr/hal-02989015

[10] A. Cevallos-Culqui, C. Pons, and G. Rodriguez, "Semi-supervised learning models for document classification: A systematic review and meta-analysis," *Inteligencia Artificial*, vol. 26, no. 72, pp. 81–111, Jun. 2023, doi: 10.4114/intartif.vol26iss72pp81-111.

[11] A. Masmoudi, H. Bellaaj, K. Drira, and M. Jmaiel, "A co-training-based approach for the hierarchical multi-label classification of research papers," *Expert Syst*, vol. 38, no. 4, pp. 1–19, 2021, doi: 10.1111/exsy.12613.

[12] X. Jia *et al.*, "Semi-supervised multi-view deep discriminant representation learning," *IEEE Trans Pattern Anal Mach Intell*, vol. 43, no. 7, pp. 2496–2509, 2021, doi: 10.1109/TPAMI.2020.2973634.

[13] G. Nayak, R. Ghosh, X. Jia, V. Mithal, and V. Kumar, "Semi-supervised Classification using Attention-based Regularization on Coarse-resolution Data," *Proceedings of the 2020 SIAM International Conference on Data Mining, SDM 2020*, pp. 253–261, 2020, doi: 10.1137/1.9781611976236.29.

[14] Y. Zhu, E. Shareghi, Y. Li, R. Reichart, and A. Korhonen, "Combining deep generative models and multi-lingual pretraining for semi-supervised document classification," *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, no. 1, pp. 894–908, 2021, doi: 10.18653/v1/2021.eacl-main.76.

[15] Z. Pan, P. Soong, S. Rafatirad, Z. Pan, P. Soong, and S. Rafatirad, "Ontology-Driven Scientific Literature Classification using Clustering and Self-Supervised Learning Ontology-Driven Scientific Literature Classification using Clustering and Self-Supervised Learning," *Easy Chair*, 2022.

[16] W. Alhoshan, A. Ferrari, and L. Zhao, "Zero-shot learning for requirements classification: An exploratory study," *Inf Softw Technol*, vol. 159, p. 107202, Jul. 2023, doi: 10.1016/j.infsof.2023.107202.

[17] W. Jia, X. Liu, Y. Wang, W. Pedrycz, and J. Zhou, "Semisupervised Learning via Axiomatic Fuzzy Set Theory and SVM," *IEEE Trans Cybern*, vol. 52, no. 6, pp. 4661–4674, Jun. 2022, doi: 10.1109/TCYB.2020.3032707.

Authorized licensed use limited to: Universidad Abierta Interamericana. Downloaded on April 16,2025 at 17:18:29 UTC from IEEE Xplore. Restrictions apply.