Business Knowledge and Neural Learning: organisation-specific transformer via semantic pre-training

Daniele Margiotta[1,2], Danilo Croce[2], Marco Rotoloni[3], Barbara Cacciamani[3] and Roberto Basili[2]

[1]Reveal s.r.l., Italy
[2]University of Roma, Tor Vergata, Italy
[3]ABI Lab, Italy

Abstract

AI approaches to business knowledge management have often neglected the role of documents, which are the backbone of expertise, norms, and optimal practices that every organisation implicitly encodes in its large-scale document collections. Banks make no exception and have to deal with operational documents on business process engineering, as well as norms on legal compliance aspects. They are thus particularly interested in the mining of the huge body of knowledge implicitly stored in their text archives, i.e. in their document assets. Extracting semantic metadata from raw bank documents is therefore central for supporting effective governance, business engineering as well as legal monitoring processes in an accurate and profitable manner. In this paper, a weakly-supervised neural methodology for creating semantic metadata from bank documents and its application to different banking organisations is presented. Based on a neural pre-training methodology driven by knowledge models of individual banks, it is shown to improve with respect to inductive approaches previously presented, that are domain specific, but organisation independent. The application to business process design in different Italian banks has been here tested and the observed impact through measurements confirms its wide applicability at the level of banks, as well as to other business organisations.

Keywords

Domain-specific neural learning, Domain Knowledge Modeling, Zero-shot Learning in NLP, BERT-based NL Inference

1. Introduction e Motivations

Traditional banking technologies focus on transaction processing and data analysis. Artificial Intelligence is promoting the adoption of data-driven methods that can induce expert rules and accurate predictions for financial forecasting tasks, such as the estimation of future values for bonds and equities, identifying market opportunities, or anti-money laundering decisions [1, 2, 3, 4]. However, dealing with massive unstructured information poses challenges, especially with non-numerical data. Financial information management applications are responding by transforming unstructured into structured data to support information labeling, searching, and promoting industry development. The banking and financial industry heavily relies on internal documentation to record and regulate, processes and organisational units. These texts include regulatory documents, reference models, and terminologies, making up a valuable repository of core data for business analysis and strategic planning. The organisational regulations are expressed in a semi-formal or textual way, making it a suitable domain for natural language processing (NLP) approaches. NLP has been used since the 1990s, and data-driven technologies have shown their value. While there are general-purpose tools available and neural techniques that have demonstrated accurate language modeling and inference capabilities [5], the application of such methods in business process mining scenarios is still limited. Structured representations of legacy models such as terminologies or ontological resources, e.g., Process Hierarchies, are not directly exploitable in supervised learning tools, as they are difficult to integrate with the unstructured information counterparts, e.g. linguistic concept descriptions, defined informally. NLP models require manually annotated examples in the target domain, making the process *time-consuming* and *costly*. The use of machine learning approaches, such as BERT [5], is appealing in discovering and classifying many-to-many associations among nodes in a Process Tree and thousands of documents produced daily by an organisation, but the effort required to provide examples of such complex associations is significant.

In a previous work ([6]), a neural architecture based on BERT was demonstrated to be effective in automatically classifying texts into Process hierarchy classes. This architecture, known as ABILaBERT, was able to associate texts with nodes in Banking Process Tree provided by ABI Lab[1]. Most notably, ABILaBERT was trained without

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
margiotta@revealsrl.it (D. Margiotta); croce@info.uniroma2.it (D. Croce); m.rotoloni@abilab.it (M. Rotoloni); b.cacciamani@abilab.it (B. Cacciamani); basili@info.uniroma2.it (R. Basili)
0000-0001-9111-1950 (D. Croce); 0000-0001-5140-0694 (R. Basili)

[1]ABI Lab (https://www.abilab.it) is the technology monitoring division of ABI, the Italian Banking Association.

the need for any labeled text, but rather through a process called *textification*, where the target taxonomy and its semantic relations between concepts (i.e. processes) was used to generate a large-scale corpus made of their corresponding textual descriptions. Notice that the ABI Lab taxonomy is representative of a generic bank and can be used for pre-training BERT before fine-tuning is carried out for text classification. However, while it was shown effective in classifying bank-specific texts through the neutral taxonomy (i.e. the ABILaB one, [6]) the above method was never applied to bank-specific taxonomies.

In this paper, we aim at answering the following Research Questions: "*Is a unified ABILaBERT model sufficiently accurate for a set of different banks $B_1, \ldots, B_n$?*"; "*Is fine-tuning of the ABILaBERT model possible against a bank-specific knowledge model?*", or in other words, "*Is a specialization of ABILaBERT towards a bank through pre-training possible and effective to induce a bank-specific model, such as $B_i - BERT$?*"; "*Which kind of fine-tuning is applicable to $B_i - BERT$ in order to get specific and optimal classifiers for the individual banks?*".

The experimental evaluation confirms that the combination of pre-training on bank-specific taxonomies and fine-tuning over (semi-automatically annotated) documents is highly beneficial, demonstrating that a bank-specific ABILaBERT can be extremely effective in the automatic classification with respect to different and heterogeneous taxonomies.

In the rest of the paper, Section 2 summarized the ABILaBERT approach. Section 3 shows how it was applied to different banks, Section 4 reports the experimental evaluation while Section 5 derives the conclusions.

2. The ABILaBERT approach

The timely and precise sharing of information is crucial for business-related problems in banks like Legal Governance, Financial Planning, or Risk Assessment. This is usually ensured through rigorous Business Process Management (BPM) frameworks. Processes are thus defined by specialists, consultants, and banking leaders, mainly designed through unstructured, or semi-structured data, such as documents or process case templates. Maintaining an efficient BPM system is a crucial activity for banks. Typically, they obtain machine-readable forms of processes through semi-formal specifications, then document them in process management platforms. Bank analysts use process-related information, such as norms or activity obligations, in their document and information management processes. The overall BPM system gives rise to a hierarchy of processes that formalize tasks and obligations at different abstraction levels.

The ABI Lab Process Tree Taxonomy[2] is a bank-independent formalization of processes active in the Italian bank eco-system that aims to map all areas of activity at a common level of detail across different banks and financial organisations without referencing organisational structures, products, or delivery channels. The process taxonomy defines process types and their subsumption relation, with specific properties of each process including a label and textual description. The processes naming and descriptions are in Italian, even though all examples will be reported through their English translations in the rest of the paper.

More formally, the process taxonomy $\mathscr{P}$ defines conceptualized process types, i.e., taxonomy nodes $p \in \mathscr{P}$, and a subsumption relation $\sqsubseteq$ in $\mathscr{P} \times \mathscr{P}$. Specific properties of a process p include at least the label, i.e., the process naming term $\texttt{label}(p)$, and its textual description, namely $\texttt{desc}(p)$. As an example, a process p has $\texttt{label}(p)$: *"Definition of the Company Vision"*, while its description $\texttt{desc}(p)$: "*The process of Defining, at an abstract level, some company objectives towards the different stakeholder, the expected company positioning and the policies to be adopted to achieve them*".

The automatic association of a text t (e.g., a paragraph from a document or the entire document itself) to nodes in this Taxonomy is traditionally modeled as a text classification task $f : t \rightarrow \mathscr{P}$. However, in order to train a classifier that approximates f, a training set of texts manually associated with nodes in the taxonomy is required. Unfortunately, this manual annotation is a costly activity, especially when the size of $\mathscr{P}$ grows.

In [6] a Zero-Shot Learning technique (ZSL) is proposed to inject information from $\mathscr{P}$ directly into a text classifier without the need for annotated documents. In particular, an approach based on *textification* is applied. The idea is to capitalize the (textual) information about the nodes of the taxonomy $\mathscr{P}$ to initialize a neural-based classifier, similar to the pre-trained stages, as discussed in [5].

Language Modeling (e.g., [5, 7]) has been largely used as an effective pre-training method for large-scale neural networks. However, the auxiliary tasks adopted (such as Masked Language Modeling) just emphasize language general properties, and models *de facto*, task- and, more importantly, domain-independent information. The domain-specific knowledge is particularly important in certain inferences, such as entity recognition and metadata creation in the financial domain: in our case, the use of a process tree as a source of information for pre-training neural networks has been shown beneficial. Since all the nodes and properties of the process tree have a linguistic nature, they can be mapped into text units ussefull to tagger inference tasks. Specifically, a subsumption relation $p_1 \sqsubset p_2$ between two processes $p_1, p_2 \in \mathscr{P}$ can be mapped into the text classification

[2]It's available at: https://www.abilab.it/tassonomia-processi-bancari

binary task of accepting a sentence like:

"$label(p_1)$ is a process more specific than $label(p_2)$"

or rejecting its inverse statement:

"$label(p_2)$ is a process more specific than $label(p_1)$"

Notice that the different information explicit in the taxonomy gives rise to auxiliary text classification tasks that can be seen as a form of pre-training of neural transformer models.

Positive and negative examples for the task can be automatically derived from the taxonomy and its related textual properties. Training the neural network to understand how processes are defined and how they subsume other processes corresponds to injecting domain-specific knowledge through a stage of domain-specific pre-training.

Once the model is pre-trained on thousands of statements automatically derived from $\mathscr{P}$, it can be used in a ZSL fashion (i.e., no text is labeled in the process) to classify texts, i.e., prompting the model with the question if a text "*t is a valid association to a node label*(p)" is true. This approach aims to avoid the manual labeling stage typical of supervised learning, and at the same time fosters different auxiliary tasks, sensitive to the knowledge implicit in the process tree.

The objective is to allow the system to encode free sentences from domain documents in an informed manner and support classification, i.e. the association of the proper processes from $\mathscr{P}$ to input texts. The resulting model is called ABILaBERT: given an incoming text t, it exploits the Transformer-based architecture to first generate an embedding for t (contained in the vector in the first position `[CLS]`) and then make it available for the classification step, possibly fine-tuned with labeled examples. For every sentence, or paragraph, t and process $p \in \mathscr{P}$ the system can estimate an auxiliary function, such as Definition Recognition $f_{desc}(t, \texttt{label}(p))$, that corresponds to accept (or reject) a sentence s_p such as:

s_p : "t is a valid description for the process $label(p)$"

In this way, the training of a classifier corresponds to learning the function

$$f_{desc}(t, \texttt{label}(p)) \textbf{ iff } \text{sentence } s_p \text{ is True}$$

These promote the node p as a good candidate to represent the semantics of a sentence t with respect to the process taxonomy $\mathscr{P}$. Note that a document is usually made of complex textual units (e.g., paragraphs) made of more than one sentence. As a consequence, ABILaBERT can be used to automatically extract rich metadata from a document by applying $f_{desc}(t, \texttt{label}(p))$ to individual paragraphs t.

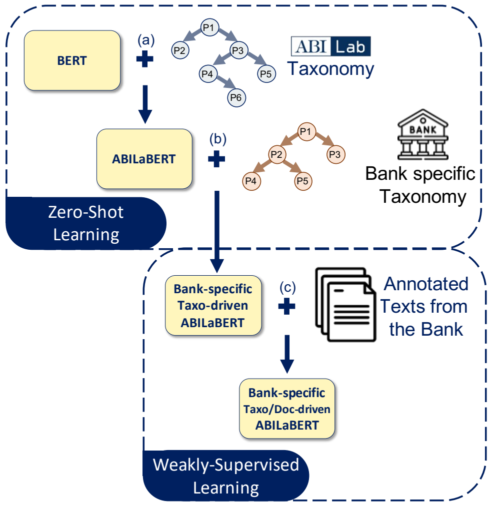

Figure 1: Adapting ABILaBERT to specific banks.

In [6], ABILaBERT was demonstrated to be effective in associating texts with the ABI Lab Process Taxonomy. In the remainder of the paper, we explore how ABILaBERT can be successfully adapted to individual bank-specific process taxonomies while maintaining its ZSL approach. We also investigate the possibility of extending the training process with a set of labeled examples in a weakly supervised manner. By exploring these avenues, we aim to enhance the applicability of ABILaBERT to a wider range of banking-specific domains, while also improving its performance in classifying texts with respect to their associated process trees.

3. Adapting ABILaBERT to specific banks

To specialize the ABILaBERT model for a specific bank, denoted by B_i, we developed a strategy outlined in Figure 1. We began by pre-training a standard BERT-based model using information derived from the ABI Lab Process Taxonomy, resulting in the "*ABILaBERT*" model, as demonstrated in [6] (step **(a)** in Figure 1).

Next, we utilized the Process Taxonomy specific to a bank B_i to create a ZSL approach for deriving a bank-specific ABILaBERT model (step **(b)** in Figure 1), denoted by "$ABILaBERT_{taxo}$". This specialized model is *taxonomy-driven*, meaning it is exclusively exposed to information derived from the taxonomy. The proposed strategy offers a straightforward way to tailor the language of processes that ABILaBERT was pre-trained on, which were specific to ABI Lab, to the language, definitions, and semantic relationships that are specific to a particular bank.

Table 1
Statistics on taxonomies and documents shared by banks.

Bank	Num. Processes	Num. Documents
$Bank_1$	25	250
$Bank_2$	15	30
$Bank_3$	10	48
$Bank_4$	28	236

Finally, when annotated documents becomes available, we can fine-tune the model in a "supervised" manner by incorporating the labeled examples (step **(c)** in Figure 1), denoted by "$ABILaBERT_{taxo,doc}$".

It also improves the model's performance by fine-tuning with labeled examples, so that the model is exposed to the "language" used in a specific bank. However, we can still avoid requiring that the annotation is completely made manually. We instead refer to this latter approach as *weakly supervised* because we do not require all paragraphs from a document to be labeled with a specific process. A subset of paragraphs can be manually annotated by the analysts, but we can also adopt ABILaBERT to annotate paragraphs within a document, thus avoiding the need for costly manual annotations. This strategy allows us to limit the need for annotated examples, by mining the bank-independent ABILaBERT as an already available supervised classifier.

It's worth noting that in some cases, the weakly-supervised strategy used by ABILaBERT may not be immediately applicable when the bank-specific taxonomy contains processes with different names or descriptions compared to the ABI Lab taxonomy, even if they express the same process. As an example, the ABI Lab tree defines the process called "*Gestione servizi di banca virtuale*"[3] while in the bank-specific hierarchy the process is called "*Gestione Digital banking e servizi remoti alla clientela*"[4]. However, ABILaBERT can be used to support this mapping process. As a text encoder [5], ABILaBERT can be used to support the mapping between the ABI Lab and bank-specific taxonomies, in order to reuse paragraphs labeled by ABILaBERT, by simply assigning the corresponding process in the targeted bank-specific taxonomy. This mapping is derived by exploiting process definitions: first, ABILaBERT can be applied to derive the embeddings of the process definitions of both taxonomies (i.e., extracting the embeddings that encode the respective `[CLS]` token), and then the semantic similarity between individual nodes of the two process trees are estimated through the cosine similarity between such embeddings. For each process in a bank's taxonomy, ABILaBERT is used to select the most similar candidate processes from the ABI Lab taxonomy. The resulting pairings are not as numerous (up to 5/10 candidates for each input process) and can be validated by the bank's expert analysts. The analysts can prompt ABILaBERT with the definition of a process from ABILaBERT, rank all the processes according to their cosine similarity, and easily retrieve the corresponding one. In this way, all labels associated with the ABI Lab taxonomy are translated to the processes of the new taxonomy: this enables their reuse to fine-tune the final transformer against the bank-specific knowledge model. This approach is cost-effective, assuming that the BERT-based model is robust enough against the potential noise introduced by the proposed automatic labeling process, more details are in [6].

[3] In English: "*Management of Virtual Banking Services*".
[4] In English: "*Digital banking and remote customer services management*".

4. Experimental Evaluation

In this section, the experimental evaluation is reported. The objective is to study the effects of tuning ABILaBERT on both bank-specific taxonomies and documents and measure its benefits.

4.1. Data and Hyperparameters

We worked with documents and process trees provided by four banks, referred to as $Bank_1, Bank_2, Bank_3$, and $Bank_4$ for sake of simplicity. While $Bank_4$ uses the same process taxonomy as ABI Lab, the internal process taxonomy of $Bank_1, Bank_2$, and $Bank_3$ are different from that of ABI Lab. Therefore, ABILaBERT will be fine-tuned only on internal documents from $Bank_4$, whereas for the other three banks, both pre-training on the taxonomy and fine-tuning on the documents will be performed. Table 1 provides a summary of the number of processes considered within the process tree from each bank, along with the corresponding number of provided documents.

ABILaBERT is a language model that is based on BERT, a popular transformer-based model used for natural language processing tasks. According to [6], ABILaBERT is a model that has been built on top of GilBERTo, then pre-trained on the texts expressing knowledge from the ABI Lab taxonomy.

To tailor ABILaBERT to document classification to a specific bank, it was further tuned, for each bank, on texts derived from the respective internal taxonomy (pre-training) as well as on annotated documents. This latter fine-tuning process involved training the specialized ABILaBERT model on the bank documents for 10 epochs, with a learning rate of $5e^{-7}$.

Differences in the internal taxonomies of the involved banks resulted in different pre-training and fine-tuning stages. As summarized in Table 1, each bank makes reference to a different number of processes as targets

Table 2
Statistics on pre-training datasets.

Bank	Positive	Negative
$Bank_1$	305	300
$Bank_2$	178	179
$Bank_3$	256	252

Table 3
Statistics on fine-tuning datasets.

Bank	Positive	Negative
$Bank_1$	14,247	341,928
$Bank_2$	3,265	45,710
$Bank_3$	2,831	25,479
$Bank_4$	11,522	311,094

Table 4
Experimental results.

Bank	Model	R@1	R@2	R@3
$Bank_1$	*GilBERTo*	.043	.087	.348
	ABILaBERT	.609	.739	.783
	$ABILaBERT_{taxo}$	.696	.783	.826
	$ABILaBERT_{taxo,doc}$	.826	.870	.913
$Bank_2$	*GilBERTo*	.000	.167	.167
	ABILaBERT	.333	.667	1.00
	$ABILaBERT_{taxo}$	.536	.722	1.00
	$ABILaBERT_{taxo,doc}$	.950	.980	1.00
$Bank_3$	*GilBERTo*	.100	.130	.230
	ABILaBERT	.427	.571	.632
	$ABILaBERT_{taxo}$	.495	.594	.632
	$ABILaBERT_{taxo,doc}$	.672	.710	.710
$Bank_4$	*GilBERTo*	.007	.012	.012
	ABILaBERT	.364	.632	.632
	$ABILaBERT_{doc}$	.510	.795	.795

for the document classification stage. Specifically, $Bank_1$ provided 250 documents that were representative of 25 internal macro-processes, $Bank_2$ provided 30 documents that represented 15 macro-processes, $Bank_3$ provided 48 documents, with reference to 10 macro-processes, and finally, $Bank_4$ provided 236 documents, and in this case, the reference taxonomy was the same as the ABI Lab one, with 28 activated macro-processes.

Using these documents and macro-processes, a dataset was generated for each bank. One dataset refers to the pre-training phase on the taxonomy (Table 2), and one dataset refers to the fine-tuning phase of the model on document paragraphs (Table 3). In Table 2 the number of "textified" examples derived from each bank-specific taxonomy are reported. It should be noted that the auxiliary relationship task used for the generation of the pre-training dataset on the taxonomy is only that of Definition Recognition of a process, as it will then be the one used (and the most effective) in the classification phase (as described in [6]). Each process generates a positive example when paired with its definition while the association with a random incorrect definition generates negative examples. Moreover, for each process, we added also examples derived by considering all its subsumed nodes. In Table 3, data referring to the fine-tuning of the ABILaBERT model on the bank-specific documents is presented. In particular, for each bank, documents were split in a percentage of 90% for training and 10% for testing. In the training documents, paragraphs were labeled by ABILaBERT and each paragraph represents a positive example for the associated process p_i. Specifically, for each positive example $f_{desc}(s, \mathtt{label}(p_i))$, where p_i are the macro-processes. When the process assigned by ABILaBERT from the ABI Lab Process Tree does not exist in the target taxonomy, this is derived using the "mapping" strategy described in Section 3. For each paragraph, t_i that showed a positive association with the macro-process $p_j \in \mathscr{P}$, positive examples $\langle t_i, p_j \rangle$ and several negative examples $\langle t_i, p_z \rangle$ were generated, where p_z are all other macro-processes present in the taxonomy (so excluding the correct macro-process p_j), in this way for each positive example we have $MP - 1$ negative examples, where MP is the number of different macro-processes in the reference taxonomy $\mathscr{P}$.

Although the paragraphs in the training set were automatically annotated, the processes associated with the paragraphs in the test dataset were manually checked by bank analysts to ensure reliable measurements.

4.2. Cross-bank Evaluation of Organisation-specific Transformers

The results are presented in Table 4, and the classification process used is the same as the one described in [6]. To summarize, ABILaBERT$_b$ first applies a filtering phase to identify a subset of processes that may be evoked by the input paragraph. The filtered subset is then classified using ABILaBERT$_b$, and the resulting candidates are ranked by their classification confidence. This ranking enables the selection of the top k ordered processes.

For each bank, the recall at k ($R@k$) is reported as a measure of classification performance. $R@k$ represents the percentage of paragraphs that were correctly associated with a process downstream of the k processes proposed by the model. By reporting recall at k for each bank, we gain insight into how well ABILaBERT$_b$ performs for different banking domains. First of all, table 4 confirm the outcomes in [6]: the original GilBERTo diverges on bank-specific documents, with a low $R@k$ comparable to a baseline where processes are randomly assigned. ABILaBERT pre-trained in [6] shows significant improvements, suggesting that the pre-training step on the ABI Lab taxonomy is highly beneficial[5]. The row $ABILaBERT_{taxo}$ show the systematic improvement due to the specific pre-train. On $ABILaBERT_{taxo,doc}$) model a more significant boost is obtained, with an average improvement of 44% in terms of $R@1$. This improvement is confirmed also for $Bank_4$ where no additional taxonomy is provided. The experimental $R@3$ results from different banks indicate that, on average, over 81% of texts can be accurately assigned to the correct process within the bank when three processes are suggested, even without prior document labeling in the overall process. We believe that this approach would be effective in supporting an annotation process to scale up effectively to fully supervised ones.

Error Analysis. A manual error analysis shows that a non-fine-tuned model such as ABILaBERT provides processes that are topically related to the texts, but these are in general too vague (i.e., "too high" in the process tree). For example, a text like '*Il codice interno della nuova linea è il '234', l'importo minimo conferibile è pari a 12.000 € e le commissioni di gestione si attestano all'1,40% + Iva (in base all'aliquota tempo per tempo vigente)*"[6] is incorrectly classified by the original ABILaBERT with the process "*Amministrazione*"[7] while the fine-tuned model correctly associates "*Credito*"[8]. The process "*Amministrazione*" seems indeed topically related to the text, but it is too vague. The fine-tuned model $ABILaBERT_{taxo,doc}$ provides a more specific and consistent labeling.

5. Conclusions

This paper has presented a novel weakly-supervised neural methodology for creating semantic metadata from bank documents and its successful application to various banking organisations. Our approach is based on a neural pre-training methodology that is driven by knowledge models specific to individual banks, and it has been shown to outperform inductive approaches that are merely specialized to the domain but independent from the organisation. Our experiments on four different Italian banks have demonstrated that the proposed methodology can significantly impact the design of business processes within an organisation. The observed results suggest that our methodology has wide applicability to other banks, as well as to other types of business organisations. This work highlights the potential of deep learning-based techniques to cost-effectively automate the process of extracting semantic information from business documents, thereby reducing the manual effort required to design effective machine reading tools beneficial to the overall efficiency of business operations.

[5]Since ABILaBERT returns processes consistent with the ABI Lab taxonomy, the mapping procedure is applied to derive the bank-specific ones.

[6]In English: "*The internal code of the new line is '234', the minimum amount that can be confirmed is €12,000, and management fees are set at 1.40% + VAT (based on the prevailing rate at the time)*".

[7]In English: "*Financial reporting*" described as: "*Management of accounting, tax, and reporting requirements borne by the bank and its group*".

[8]In English: "*Credit management process*" described as: "*Process of credit management, in its funding and origination components, to different recipients (businesses, households, and public administration), in the different types (land credit, agricultural, ...*

Acknowledgments

The authors would like to thank the Special Interest Group of "ABI Lab" for actively supporting the research and experimentation presented in this paper. In particular, we thank the following banks: Monte dei Paschi di Siena (in particular, Ugolini Massimiliano), Mediolanum (Paolo Crocè, Francesco Fasano, Demetrio Migliorati, Gaetano Silletti), Banca Nazionale del Lavoro (Emanuele Tango, Ciro Esposito) and Banca Popolare di Sondrio (Roberta Besseghini, Gianpaolo Mura, Sergio Pozzi). We acknowledge financial support from the PNRR MUR project PE0000013-FAIR.

References

[1] N. Cohen, T. Balch, M. Veloso, Trading via image classification, CoRR abs/1907.10046 (2019). arXiv:1907.10046.

[2] J.-H. Chen, Y.-C. Tsai, Encoding candlesticks as images for patterns classification using convolutional neural networks, 2020. arXiv:1901.05237.

[3] X. Li, J. Saúde, P. Reddy, M. Veloso, Classifying and understanding financial data using graph neural network, in: AAAI-20 Workshop on Knowledge Discovery from Unstructured Data in Financial Services, 2020.

[4] D. Borrajo, M. Veloso, S. Shah, Simulating and classifying behavior in adversarial environments based on action-state traces: an application to money laundering, CoRR abs/2011.01826 (2020). arXiv:2011.01826.

[5] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, CoRR abs/1810.04805 (2018).

[6] D. Margiotta, D. Croce, M. Rotoloni, B. Cacciamani, R. Basili, Knowledge-based neural pre-training for intelligent document management, in: 20th International Conference of the Italian Association for Artificial Intelligence, Virtual Event, December 1-3, 2021, volume 13196 of *Lecture Notes in Computer Science*, Springer, 2021, pp. 564–579.

[7] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, ArXiv abs/1907.11692 (2019).