Quantum Transfer Learning for Sentiment Analysis: an experiment on an Italian corpus

Giuseppe Buonaiuto*
giuseppe.buonaiuto@icar.cnr.it
ICAR-CNR
Naples, Italy

Raffaele Guarasci
raffaele.guarasci@icar.cnr.it
ICAR-CNR
Naples, Italy

Massimo Esposito
massimo.esposito@icar.cnr.it
ICAR-CNR
Naples, Italy

ABSTRACT

Exploiting quantum properties to improve performance of different tasks in Natural Language Processing (NLP) and other domains has increasingly becoming a successful trend to deal complex language phenomena or to fill task or domain-specific gaps with an approach that needs less data and minor computational resources. The field that has, to date, yielded more quantum-based attention is the retrieval and classification of textual data. This work aims to replicate the excellent results of hybrid quantum approaches for syntactic tasks on semantic classification tasks. In detail, a quantum machine learning algorithm, namely, the Variational Quantum Classifier (VQC), is used to perform sentiment analysis classification tasks. This algorithm can deduce the relationships between input features and their corresponding class affiliations us ing a parametrized quantum circuit and an encoding layer that translates classical data into quantum states. The approach has been tested on a well-known benchmark annotated dataset used for the Italian language, and the results have been compared to existing baselines, pointing out state-of-the-art scores.

KEYWORDS

Variational Quantum Classifier, Quantum Transfer Learning, QNLP

ACM Reference Format:
Giuseppe Buonaiuto, Raffaele Guarasci, and Massimo Esposito. 2018. Quantum Transfer Learning for Sentiment Analysis: an experiment on an Italian corpus. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3660318.3660325

1 INTRODUCTION

NLP has experienced exponential growth in recent years, primarily driven by the emergence of Neural and Large Language Models (NLMs/LLMs). The introduction of Transformer-based models, notably exemplified b y p ioneering m odels s uch a s B ERT [15], has steered into a new era in NLP methodologies. These advancements have consistently elevated the attainable performance across many tasks, spanning various domains and languages [20–22], ranging from classical applications such as information extraction or text classification[18, 38], to highly specialized verticalized approaches [23, 27, 32, 34, 37, 47, 48]. The increasing demand for resources needed to tow and fine-tune these new models has led to the release of more and more resources and datasets optimized for diverse applications [6, 42].

However, all these factors have led to a complexity of models, increasingly greedy for data and computational power for training and fine-tuning [7, 16].

For these reasons, quantum machine learning (QML)[5] and its derivative sub-field Natural Language Processing (QNLP) [13] have been considered as a viable alternative, using properties derived from quantum theory.

Various approaches have been proposed, from those that are only valid theoretically, to others tested on classical hardware to those launched on the current available noisy quantum hardware (NISQ). Although the latter offer the most promising performance, due to the current immaturity of the NISQ machines, they are limited to easy tasks and small scale datasets.

Therefore, in the current scenario, the most exploitable avenue is hybrid classical-quantum approaches. Firstly proposed in [28], this approach is based on a quantum self-attention neural network (QSANN) and it introduces the possibility of non-linearity, achieving best performances over other QNLP models[31]. In [29], the approach has been extended, addressing the low non-linearity issue for QNLP models using the classical-quantum transfer learning paradigm [33]. This mechanism, combined with pre-trained quantum encodings, has demonstrated its effectiveness in classification tasks [9, 17], open up for the possibility of being implemented on real quantum hardware.

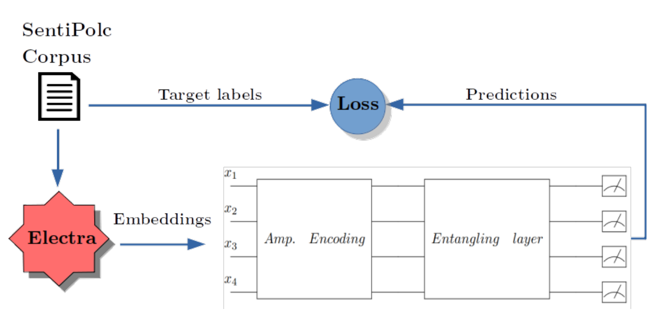

Figure 1: A schematic of the computational pipeline. The corpus SentiPolc is processed via a pre-trained Electra model, which is used to extract vector embeddings that are then processed by a variational quantum circuits. This is trained for performing a binary classification.

QUASAR '24, June 3, 2024, Pisa, Italy

ACM ISBN 979-8-4007-0646-2/24/06.
https://doi.org/10.1145/3660318.3660325

Starting from these premises, this work aims to use a hybrid pipeline based on the Electra model and Variational Quantum Classifier (VQC) [11] to perform a semantic-based classification task. The task chosen is Sentiment Analysis, which has a long and prolific history in the NLP field. The choice of this specific task is due to multiple factors. First, it has been a task used in recent years in the Quantum NLP field [35]; furthermore, the pipeline proposed here has already proven effective in a classification task that involved mainly syntactic aspects [9], so it can be attractive to investigate the performance about tasks based on different levels of linguistic knowledge[10].

The paper is organized as follows: in section 2, the research works available in the literature related to what is presented are described, while in section 3 the dataset and applied methodologies are presented, then in section 4, the results obtained are exposed, and relevant aspects are discussed, and finally, overall conclusions are drawn in section 5.

2 RELATED WORK

Although in recent years, quantum-based approaches have invested in all NLP tasks, and many solutions have been proposed, in this section, we focus only on the hybrid approaches closest to the one presented here and on the task. For a more in-depth review of work that has addressed other tasks, see [19].

2.1 Quantum Transfer Learning

Numerous examples have received increasing interest in recent years concerning the adaptation of classical machine learning algorithms through the use of properties and techniques borrowed from quantum mechanics. In tasks like sentiment analysis or document classification, great research instruments could be Quantum Support Vector Machines (QSVMs), which utilize quantum algorithms to improve traditional Support Vector Machines (SVMs), or the quantum-inspired algorithms for text classification, as Quantum-Inspired Genetic Algorithm (QGA) and Quantum-Inspired Particle Swarm Optimization (QPSO) which borrow principles from quantum mechanics to enhance optimization techniques like feature selection and parameter optimization for text classification. To represent and analyze textual data to capture more nuanced and context-dependent information, have been recently (?) developed Quantum Embeddings. Although still in its early stages, quantum machine learning algorithms, such as Quantum Neural Networks (QNNs) and Quantum Boltzmann Machines (QBMs), utilize quantum systems' computational capabilities to efficiently handle complex computations for text classification tasks. One notable approach is Quantum Support Vector Machines (QSVMs), which aim to enhance the performance of traditional Support Vector Machines (SVMs) by utilizing quantum algorithms. QSVMs have shown promising results in various text classification tasks, including sentiment analysis, topic classification, and document classification. Another area of research is quantum-inspired algorithms for text classification. Quantum-inspired algorithms, such as Quantum-Inspired Genetic Algorithm (QGA) and Quantum-Inspired Particle Swarm Optimization (QPSO), draw inspiration from quantum mechanics and apply quantum-like principles to improve traditional optimization techniques. These algorithms have been explored in the context of feature selection and parameter optimization for text classification, demonstrating their potential to enhance classification accuracy and efficiency. Moreover, quantum embeddings have gained attention as a means to represent and analyze textual data. Quantum embeddings leverage the concepts of quantum superposition and entanglement to capture semantic relationships between words or documents. These embeddings aim to capture more nuanced and context-dependent information compared to traditional word embeddings. By utilizing quantum representations, classification models can benefit from enhanced semantic understanding, improving performance in various NLP tasks. Furthermore, quantum machine learning algorithms, such as Quantum Neural Networks (QNNs) and Quantum Boltzmann Machines (QBMs), have been explored in text classification. These quantum-inspired models leverage the unique computational capabilities of quantum systems to perform complex computations efficiently. Although still in its early stages, quantum machine learning holds the potential to address the computational challenges associated with large-scale language data and to provide more powerful models for classification tasks.

2.2 Quantum-based Sentiment Analysis

Sentiment analysis is a task in NLP used to determine the leading emotional tone behind a text (positive, negative, neutral). During the last decade, it has experienced enormous popularity for allowing us to determine the polarity orientation of a wide span of textual data, ranging from user-generated-content social media posts to online reviews. The task can be considered a sub-task of Information Retrieval, in particular, derived from text classification-based tasks [39]. While information retrieval focuses on extracting relevant information from a large corpus of data, sentiment analysis aims to filter and rank information extracted during retrieval. In [40], sentiment analysis techniques have been used to support document filtering and classification within information retrieval systems. In contrast, [49] have exploited sentiment information contained in documents to enhance relevance ranking algorithms. Moreover, analyzing user feedback has been applied in different domains, ranging from organizations [24] to user personalization [14]. [30] has proposed a sentiment-based approach to improve the performance of information retrieval systems. Sentiment information extracted by documents (i.e., user intent and emotion) can help the system to perform a more accurate retrieval and recommendation of relevant information. In summary, integrating sentiment analysis into information retrieval systems has transformed them from mere content retrievers to intelligent platforms that understand and respond to user emotions and preferences. This fusion improves the quality and relevance of retrieved content and enhances the overall user experience in navigating the digital landscape.

Concerning recent approaches, sentiment analysis has been addressed using several machines and deep learning methods, ranging from LSTM [**?**] to different neural language models [1]. The situation is much more complex when trying to approach low-resource languages such as Italian, for which few works achieve the same performance as English, and most approaches are pretty dated due to the lack of available resources[44]. Typically, the task has been addressed, focusing on employing popular textual feature representation methods to construct vector representations of documents.

Although these approaches can model linguistic information, they can fail to capture the sentiment information [25], and many issues still need to be solved.

Moving to quantum-inspired approaches, sentiment Analysis is one of the tasks that has benefited the most from such advancements. In [52], a method based on features from quantum probability theory has been proposed. This unsupervised approach is based on a density matrix using two custom sentiment dictionaries. The sentiment is obtained using quantum relative entropy, calculated due to the similarity between dictionaries and documents. To obtain a better representation of relations between words composing sentences, [51] has developed a novel approach exploiting quantum-inspired interactive networks, which merge quantum theory and the long short-term memory (LSTM) neural network. Word relations are identified using the density matrix and then used as input for LSTM. This approach has been compared to different baselines ranging from CNN [26] to attention-based LSTM and Contextual/Hierarchical biLSTM [43]. Experiments have been conducted on two annotated datasets. An extension of this work has been proposed in [50]. This work is structured on an architecture based on a tensor network that is able to improve the performance and interpretability of the results. This model encodes high-dimensional word vectors in a probabilistic space using a generative tensor network to classify texts. This approach has been evaluated against most used sentiment analysis benchmark models showing comparable performances.

3 MATERIALS AND METHODS

3.1 Dataset

The resource chosen for this work is the most famous Italian dataset in the literature for the sentiment analysis task, SENTIPOLC 2016 [1]. It is a collection of Italian tweets annotated with sentiment polarity and other related information. It has been originally developed for the sentiment analysis task at EVALITA 2016, the fourth evaluation campaign for natural language processing and speech tools for Italian [2].

The dataset has been developed for four different subtasks: subjectivity classification, polarity classification, topic-based polarity classification, and irony detection. In detail it contains:

- 9,000 tweets total (7,000 for training and 2,000 for testing)
- The tweets have been randomly sampled from a larger corpus of 100,000 tweets collected between January and April 2016 using the Twitter Streaming API and a set of keywords related to politics, economics, and social issues.
- annotation has been performed by three experts. The inter-annotator agreement has been measured using Krippendorff's alpha and ranged from 0.64 to 0.82 depending on the subtask.

Although the dataset has intrinsic and extrinsic limitations, primarily concerning its data source (Twitter), it proves to be the most suitable for the purpose of this study. This choice is mainly motivated by the opportunities it offers for comparison with other models, hence the release of baselines[36, 41].

[1] https://github.com/evalita2016

3.2 Model

The model chosen for this work is ELECTRA [12]. This model has gained much success in recent years due to its ability to better capture contextual word representations compared to the widely-used BERT, given the same model size, data, and compute [45].

Electra is a transformer model composed of a generator and a discriminator. The generator is trained as a masked language model and attempts to replace tokens in a given sequence. Instead, the discriminator attempts to identify which token has been modified. Specifically, for a sentence, the tokens are replaced randomly by a mask, and the generator is trained to predict the original tokens from the masked ones. Then, the generator outputs a fake sentence for the discriminator. The discriminator then is trained to decide if the tokens provided are fake or real. With this approach, the number of examples required for training is reduced significantly compared to other models like Bert.

In detail, for a given input sequence, in which some tokens are randomly replaced with a [MASK] token, the generator G is trained to predict the original tokens for all masked ones. On the other hand, G is given input sequences built by replacing [MASK] tokens with *fake* ones produced by G, and it is trained to predict whether they are original or *fake*.

More formally, given an input sentence s of raw text χ, composed by a sequence of tokens $s = w_1, w_2, \ldots, w_n$ where w_t $(1 \leq t \leq n)$ represents the generic token, both G and D firstly encode s into a sequence of contextualized vector representations $h(s) = h_1, h_2, \ldots, h_n$.

Then, for a given position t so that the corresponding $w_t = [MASK]$, the generator outputs the probability to have a token w_t, with a softmax layer:

$$p_G(w_t|s) = \frac{e(w_t)^T h_G(s)_t}{\sum_{w'} exp(e(w')^T h_G(s)_t)} \tag{1}$$

where $e(\cdot)$ represents the embedding function.

The discriminator predicts whether w_t is the original or "fake", using a sigmoid layer:

$$D(s,t) = sigmoid(e(w_t)^T h_D(s)_t) \tag{2}$$

During the pre-training, G employs the following loss function:

$$\mathcal{L}_{Gen} = \mathcal{L}_{MLM} = \mathbb{E}(\sum_{i \in m} -\log p_G(w_i|s^{masked})) \tag{3}$$

where $m = m_1, m_2, \ldots, m_k$ are k random selected words and s^{masked} is the sentence with the masked words.

On the other hand, D uses the following loss function:

$$\begin{aligned}\mathcal{L}_{Dis} = \mathbb{E}(\sum_{t=1}^{n} -\mathbb{I}(w_t^{corrupt} = x_t) \log D(s^{corrupt}, t) + \\ -\mathbb{I}(w_t^{corrupt} \neq x_t) \log D(s^{corrupt}, t))\end{aligned} \tag{4}$$

where $w_t^{corrupt}$ is the corrupted word within the corrupted sentence $s^{corrupt}$.

Finally, the following combined loss is minimized:

$$\min_{\theta_G, \theta_D} \sum_{s \in \chi} \mathcal{L}_{Gen}(s, \theta_G) + \lambda \mathcal{L}_{Dis}(s, \theta_D) \tag{5}$$

At the end of the pre-training, G is discarded, and only D is used.

The main reason behind ELECTRA's improved results is that predictions are calculated not only over masked tokens but also for each token, and the discriminator loss can be calculated over all input tokens. Several studies in the literature have already pointed out that a massive amount of computational resources are needed to train ELECTRA. For such reasons, a pre-trained version of the model is used. In particular, the cased and XXL version of the *dbmdz* Italian ELECTRA model[2] has been selected for the embedding extraction, i.e., the model is taken without the final neural network used for classification.

3.3 Quantum Pipeline

The numerical representation of sentences, obtained from the pre-trained Electra, must be encoded in a quantum state for a proper manipulation in a quantum computing algorithm. In this work the quantum amplitude encoding has been used: here the classical real vectors of data are encoded in the amplitudes of a quantum superposition. In particular, classical data are assigned to the complex amplitudes of specific computational basis states. In a formal way, given the classical feature vector $\mathbf{v} = (v_1, v_2, \ldots, v_N)$, where N represents the dimension of the feature space, and a set of $n = \lceil \log_2 N \rceil$ qubits, the encoded quantum state $|\mathbf{v}\rangle$ is given by:

$$|\mathbf{v}\rangle = \sum_{i=0}^{N-1} \beta_i |i\rangle, \tag{6}$$

where $|i\rangle$ is the computational basis state for the n qubits, and β_i are complex amplitudes related to the encoded data v_i via:

$$\beta_i = \frac{v_i}{\sqrt{\sum_{j=0}^{N-1} |v_j|^2}}, \tag{7}$$

where the normalization ensures that the quantum state $|\mathbf{v}\rangle$ is normalized. The constructed quantum states are then used as initial state of a quantum circuits with parametrized gates [3]. The parameters are updated via a classical optimizer, with the goal of finding the optimal configuration that gives, after the measurement, a minimum of the objective function evaluated on the target labels. The performances of such parametrized circuits are highly dependent on the ansatz in use and on the characteristic of the quantum computer itself [8]. Although there is no universal rule prescribed for the selection of the best ansatz for a specific task, it is crucial to put the qubit involved in a quantum superposition, i.e. an entangled state, to best exploit the learning capabilities of a variational quantum learning algorithm: here a **Basic Entangling Layer** provided by Pennylane [4] is used. This ansatz is made of a sigle qubit rotational gates with one trainable parameter and a closed ring of CNOTs, which are two qubit gate that build up entanglement and classical correlation between each qubit in the computation. The last step of the quantum pipeline is the measurements of the final quantum state: measurements are taken in multiple shots, over a single quantum operator, in order to collect a meaningful statistics of the outcome. More formally, given the parametrized quantum circuit U_θ, with parameters θ and a measurement operator M, the outcome, p, is given by:

$$p = |\langle \mathbf{v}|U_\theta^\dagger M U_\theta|\mathbf{v}\rangle|^2 \tag{8}$$

where $U_\theta|\mathbf{v}\rangle$ is the state obtained from the application of the quantum circuit on the initial state. In particular p is a real number, or a real vector-depending on the composition of the measurements and on the target variable, which is thus used to evaluate the objective function together with the training labels. It is hence crucial to select the proper measurement operators too: in the experiment here described, given the selected ansatz, the measurement M, used for extracting information out of the parametrized quantum circuit, is the projection of each qubit onto the z axis of rotation for each qubit, i.e. the expectation value of the *Pauli-Z* operator.

3.4 Experimental Assessment

For the specific problem here assessed, the Sentiment Analysis on Italian dataset SentiPolc, the modules specifications and the steps of the experimental phase are the followings:

- Pre-trained Electra used for word embedding with max length per sentence of 152. The length of each embedding is of 768
- The embeddings are encoded in a quantum state via amplitude encoding with 10 qubit, as $\lceil log_2(768) \rceil = 10$, where the rest of the 1024 amplitudes are padded with a value of 0.001. After the encoding phase, the state obtained is used as an input for a 8 layers of parametrized.
- After the encoding, the state is used as an input of a parametrized quantum circuit. The ansatz in use, i.e. the structure and the nature of the quantum gates composing the parametrized cirucit, is the *BasicEntangledLayer*, which is made of single parameter rotations on each qubit, and a chain of CNOT gate with cyclic boundary condition. In particular 6 layers of this ansatz have been used in the experiment.
- As the desired output is binary, the measurements are applied on two qubits out of ten. Specifically, the *PauliZ* operator have been measured on the final state emerging from the quantum circuit.
- The loss function evaluated for each run of the computation is the binary cross entropy loss, which is standard for binary classification. Furthermore, the optimizer used for the update rule of the parameter is the $AdamW$ optimizer, with a learning rat of 10^{-5} and numerical stabilization term $\epsilon = 10^{-6}$

The outcome are compared with the ground truth data, as customary in supervised learning. While the accuracy of the classifier is evaluated on the test set.

4 RESULTS

The training of the Electra-quantum hybrid model has been performed for 8 epochs, a small number that serves to avoid over-fitting, a learning rate of 10^{-5}, on a training set made of 4476 sentences. Conversely, the test set is composed of 500 sentences. The quality of the binary classification is assessed through the F1 score, which provides a robust metric in this case, given the possible class unbalance within the training set.

The model evaluation results on the test are shown in 1: the F1 score of the Electra-quantum model is compared with other

[2] https://huggingface.co/dbmdz/electra-base-italian-xxl-cased-discriminator

Model	F1-Score
Electra-quantum	**0.77**
BERT-based system (with pre-processing)	0.75
BERT-based system (without pre-processing)	0.73
Alberto	0.72
LSTM-based	0.69

Table 1: Comparison of classification results on the test set of SentiPolc dataset.

available classical models in the literature, evaluated on the same dataset. The result shows a prominent improvement compared to previously tested models: this is a rather significant result that serves as a first hint of the possibilities offered by quantum classifiers. Regarding several parameters, the quantum circuit attached to the pre-trained Electra model has a reduced size comparable to a simple feed-forward neural network. Nevertheless, the results outperform mode complicated models, stating that the quantum algorithm learning abilities can outperform their classical counterparts on some specific tasks. Notice that 30 experiments have been performed with a random distribution of the initial parameters of the quantum circuit, thus obtaining a statistically significant dispersion of the results. The obtained final F1-score for the ensemble evaluation returns a value of 0.7775 ± 0.009, hence proving the outperformance on the specific task compared to other models, well above the standard deviation. Compared to those obtained in [44], the results are particularly interesting (second and third row of the table). In fact, in that work, a BERT-based architecture has been implemented and tested using two different approaches: the first using the raw dataset, the second performing targeted pre-processing consisting mainly of data-cleansing as often carried out in user-generated-content from Twitter in order to ensure a better result[46]. Such a language-independent phase of pre-processing has allowed to clean up the text from the Twitter jargon and meta-textual elements such as emoji, URLs, and hashtags by shifting the F1 value by two percentage points (from 0.73 to 0-75). The performance of the presented quantum transfer learning approach achieves superior results using a comparable architecture, Electra, also based on Transformers. Indeed, a higher score (0.77) is shown even using just the original dataset, thus without cleaning the raw data and without pre-processing, paving the way for extensive testing that may include additional models for comparison.

5 CONCLUSIONS

This work has presented a quantum transfer-learning approach for semantic tasks in the Italian language, focusing on sentiment analysis. In detail, the methodology is based on a quantum classifier, trained using a sentence embedding strategy provided by Electra. Concerning the dataset used to test this approach, the choice has fallen on one previously used for the EVALITA evaluation campaign.

Results have pointed out that this type of pipeline can achieve performances that outperform other NLMs, particularly BERT-based models, which achieve state-of-the-art performance on this dataset. Besides the achieved scores, another strong point of the proposed approach is the compactness of the model, based on a relatively simple ansatz. The lack of data preprocessing in the experimental phase, compared to the most performing classical model, points towards a promising direction of quantum advantage on this specific task.

Since this work has been tested on a strictly semantic task, it stands as a complement to the experiments that have already verified the model's applicability to syntactic tasks. Future developments have two planned directions. From a technical perspective, experimenting with different possible embeddings in order to capture more knowledge about the sentences will offer different insights. From a linguistic point of view, it will be interesting to combine the two tasks to develop a pipeline capable of approaching both the syntactic and semantic levels and extending the analysis by taking into account multilingual models.

ACKNOWLEDGMENTS

We acknowledge financial support from the project PNR MUR project PE0000013-FAIR.

REFERENCES

[1] Shivaji Alaparthi and Manit Mishra. 2021. BERT: A sentiment analysis odyssey. *Journal of Marketing Analytics* 9, 2 (2021), 118–126.

[2] Francesco Barbieri, Valerio Basile, Danilo Croce, Malvina Nissim, Nicole Novielli, Viviana Patti, et al. 2016. Overview of the evalita 2016 sentiment polarity classification task. In *CEUR Workshop Proceedings*, Vol. 1749. CEUR-WS.

[3] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. 2019. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology* 4, 4 (2019), 043001.

[4] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, Shahnawaz Ahmed, Vishnu Ajith, M. Sohaib Alam, Guillermo Alonso-Linaje, B. AkashNarayanan, Ali Asadi, Juan Miguel Arrazola, Utkarsh Azad, Sam Banning, Carsten Blank, Thomas R Bromley, Benjamin A. Cordier, Jack Ceroni, Alain Delgado, Olivia Di Matteo, Amintor Dusko, Tanya Garg, Diego Guala, Anthony Hayes, Ryan Hill, Aroosa Ijaz, Theodor Isacsson, David Ittah, Soran Jahangiri, Prateek Jain, Edward Jiang, Ankit Khandelwal, Korbinian Kottmann, Robert A. Lang, Christina Lee, Thomas Loke, Angus Lowe, Keri McKiernan, Johannes Jakob Meyer, J. A. Montañez-Barrera, Romain Moyard, Zeyue Niu, Lee James O'Riordan, Steven Oud, Ashish Panigrahi, Chae-Yeun Park, Daniel Polatajko, Nicolás Quesada, Chase Roberts, Nahum Sá, Isidor Schoch, Borun Shi, Shuli Shu, Sukin Sim, Arshpreet Singh, Ingrid Strandberg, Jay Soni, Antal Száva, Slimane Thabet, Rodrigo A. Vargas-Hernández, Trevor Vincent, Nicola Vitucci, Maurice Weber, David Wierichs, Roeland Wiersema, Moritz Willmann, Vincent Wong, Shaoming Zhang, and Nathan Killoran. 2018. PennyLane: Automatic differentiation of hybrid quantum-classical computations. *arXiv e-prints*, Article arXiv:1811.04968 (Nov. 2018), arXiv:1811.04968 pages. https://doi.org/10.48550/arXiv.1811.04968 arXiv:1811.04968 [quant-ph]

[5] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. 2017. Quantum machine learning. *Nature* 549, 7671 (2017), 195–202.

[6] Federico Bonetti, Elisa Leonardelli, Daniela Trotta, Raffaele Guarasci, and Sara Tonelli. 2022. Work Hard, Play Hard: Collecting Acceptability Annotations through a 3D Game. *2022 Language Resources and Evaluation Conference, LREC 2022* (2022), 1740 – 1750. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85144445524&partnerID=40&md5=9e8e4addd231c367f1897d91c1d974b4

[7] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[8] Giuseppe Buonaiuto, Francesco Gargiulo, Giuseppe De Pietro, Massimo Esposito, and Marco Pota. 2024. The effects of quantum hardware properties on the performances of variational quantum learning algorithms. *Quantum Machine Intelligence* 6, 1 (Feb. 2024). https://doi.org/10.1007/s42484-024-00144-5

[9] Giuseppe Buonaiuto, Raffaele Guarasci, Aniello Minutolo, Giuseppe De Pietro, and Massimo Esposito. 2024. Quantum transfer learning for acceptability judgements. *Quantum Machine Intelligence* 6, 1 (2024), 13.
[10] Elena Cardillo, Alessio Portaro, Maria Taverniti, Claudia Lanza, and Raffaele Guarasci. 2024. Towards the Automated Population of Thesauri Using BERT: A Use Case on the Cybersecurity Domain. In *International Conference on Emerging Internet, Data & Web Technologies*. Springer, 100–109.
[11] Samuel Yen-Chi Chen, Chih-Min Huang, Chia-Wei Hsing, and Ying-Jer Kao. 2020. Hybrid quantum-classical classifier based on tensor network and variational quantum circuit. *arXiv preprint arXiv:2011.14651* (2020).
[12] Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators. In *ICLR*. https://openreview.net/pdf?id=r1xMH1BtvB
[13] Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394* (2010).
[14] Kushal Dave, Steve Lawrence, and David M Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*. 519–528.
[15] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423
[16] Luciano Floridi and Massimo Chiriatti. 2020. GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines* 30, 4 (2020), 681–694.
[17] Raffaele Guarasci, Giuseppe Buonaiuto, Giuseppe De Pietro, and Massimo Esposito. 2023. Applying Variational Quantum Classifier on Acceptability Judgements: a QNLP experiment. *Numerical Computations: Theory and Algorithms NUMTA 2023* (2023), 116.
[18] Raffaele Guarasci, Emanuele Damiano, Aniello Minutolo, Massimo Esposito, and Giuseppe De Pietro. 2020. Lexicon-grammar based open information extraction from natural language sentences in Italian. *Expert Systems with Applications* 143 (2020), 112954.
[19] Raffaele Guarasci, Giuseppe De Pietro, and Massimo Esposito. 2022. Quantum natural language processing: Challenges and opportunities. *Applied Sciences* 12, 11 (2022), 5651.
[20] Raffaele Guarasci, Stefano Silvestri, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito. 2021. Assessing BERT's ability to learn Italian syntax: a study on null-subject and agreement phenomena. *Journal of Ambient Intelligence and Humanized Computing* (2021), 1–15.
[21] Raffaele Guarasci, Stefano Silvestri, Giuseppe De Pietro, Hamido Fujita, and Massimo Esposito. 2022. BERT syntactic transfer: A computational experiment on Italian, French and English languages. *Computer Speech & Language* 71 (2022), 101261.
[22] Raffaele Guarasci, Stefano Silvestri, and Massimo Esposito. 2024. Probing Cross-lingual Transfer of XLM Multi-language Model. In *International Conference on Emerging Internet, Data & Web Technologies*. Springer, 219–228.
[23] Anshika Gupta, Shaz Furniturewala, Vijay Kumari, and Yashvardhan Sharma. 2023. Steno AI at SemEval-2023 Task 6: Rhetorical Role Labelling of Legal Documents using Transformers and Graph Neural Networks. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*. 1858–1862.
[24] Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. 168–177.
[25] Shofiqul Islam, Ngahzaifa Ab Ghani, and Manjur Ahmed. 2020. A review on recent advances in Deep learning for Sentiment Analysis: Performances, Challenges and Limitations. *Compusoft* 9, 7 (2020), 3775–3783.
[26] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, Doha, Qatar, 1746–1751. https://doi.org/10.3115/v1/D14-1181
[27] Tharindu Kumarage and Huan Liu. 2023. Neural Authorship Attribution: Stylometric Analysis on Large Language Models. *arXiv preprint arXiv:2308.07305* (2023).
[28] Guangxi Li, Xuanqiang Zhao, and Xin Wang. 2022. Quantum self-attention neural networks for text classification. *arXiv preprint arXiv:2205.05625* (2022).
[29] Qiuchi Li, Benyou Wang, Yudong Zhu, Christina Lioma, and Qun Liu. 2023. Adapting Pre-trained Language Models for Quantum Natural Language Processing. *arXiv preprint arXiv:2302.13812* (2023).
[30] Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In *Mining text data*. Springer, 415–463.
[31] Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran. 2020. Quantum embeddings for machine learning. *arXiv preprint arXiv:2001.03622* (2020).
[32] Alessandro Maisto and Raffaele Guarasci. 2016. Morpheme-based recognition and translation of medical terms. *Communications in Computer and Information Science* 607 (2016), 172 – 181. https://doi.org/10.1007/978-3-319-42471-2_15
[33] Andrea Mari, Thomas R Bromley, Josh Izaac, Maria Schuld, and Nathan Killoran. 2020. Transfer learning in hybrid classical-quantum neural networks. *Quantum* 4 (2020), 340.
[34] Fiammetta Marulli, Marco Pota, Massimo Esposito, Alessandro Maisto, and Raffaele Guarasci. 2018. Tuning syntaxnet for pos tagging italian sentences. *Lecture Notes on Data Engineering and Communications Technologies* 13 (2018), 314 – 324. https://doi.org/10.1007/978-3-319-69835-9_30
[35] Konstantinos Meichanetzidis, Alexis Toumi, Giovanni de Felice, and Bob Coecke. 2020. Grammar-Aware Question-Answering on Quantum Computers. *arXiv preprint arXiv:2012.03756* (2020).
[36] Luca Molinaro, Rosalia Tatano, Enrico Busto, Attilio Fiandrotti, Valerio Basile, and Viviana Patti. 2022. DelBERTo: A Deep Lightweight Transformer for Sentiment Analysis. In *International Conference of the Italian Association for Artificial Intelligence*. Springer, 443–456.
[37] Mario Monteleone, Raffaele Guarasci, and Alessandro Maisto. 2018. NooJ Morphological Grammars for Stenotype Writing. *Communications in Computer and Information Science* 811 (2018), 200 – 212. https://doi.org/10.1007/978-3-319-73420-0_17
[38] Smruthi Mukund, Rohini Srihari, and Erik Peterson. 2010. An information-extraction system for Urdu—a resource-poor language. *ACM Transactions on Asian Language Information Processing (TALIP)* 9, 4 (2010), 1–43.
[39] Georgios Paltoglou and Mike Thelwall. 2010. A study of information retrieval weighting schemes for sentiment analysis. In *Proceedings of the 48th annual meeting of the association for computational linguistics*. 1386–1395.
[40] Bo Pang, Lillian Lee, et al. 2008. Opinion mining and sentiment analysis. *Foundations and Trends® in information retrieval* 2, 1–2 (2008), 1–135.
[41] Marco Polignano, Pierpaolo Basile, Marco de Gemmis, Giovanni Semeraro, and Valerio Basile. 2019. AlBERTo: Italian BERT Language Understanding Model for NLP Challenging Tasks Based on Tweets, In Proceedings of the Sixth Italian Conference on Computational Linguistics (CLiC-it 2019). *CEUR Workshop Proceedings* 2481. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85074851349&partnerID=40&md5=7abed946e06f76b3825ae5e294ffac14
[42] Marco Polignano, Pierpaolo Basile, Marco De Gemmis, Giovanni Semeraro, Valerio Basile, et al. 2019. Alberto: Italian BERT language understanding model for NLP challenging tasks based on tweets. In *CEUR Workshop Proceedings*, Vol. 2481. CEUR, 1–6.
[43] Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. 2017. Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)*. 873–883.
[44] Marco Pota, Mirko Ventura, Rosario Catelli, and Massimo Esposito. 2020. An Effective BERT-Based Pipeline for Twitter Sentiment Analysis: A Case Study in Italian. *Sensors* 21, 1 (Dec. 2020), 133. https://doi.org/10.3390/s21010133
[45] Anna Rogers, Olga Kovaleva, and Anna Rumshisky. 2020. A primer in bertology: What we know about how bert works. *Transactions of the Association for Computational Linguistics* 8 (2020), 842–866.
[46] Amrita Shelar and Ching-Yu Huang. 2018. Sentiment analysis of twitter data. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 1301–1302.
[47] Daniela Trotta, Raffaele Guarasci, Elisa Leonardelli, and Sara Tonelli. 2021. Monolingual and Cross-Lingual Acceptability Judgments with the Italian CoLA corpus. *Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021* (2021), 2929 – 2940. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129142333&partnerID=40&md5=dc1e208191ead8adbcfe66e0fdda33e8
[48] Daniela Trotta, Michele Stingo, Raffaele Guarasci, Annibale Elia, and Teresa Albanese. 2018. Multi-word expressions in spoken language: PoliSdict. *CEUR Workshop Proceedings* 2253 (2018). https://www.scopus.com/inward/record.uri?eid=2-s2.0-85057730561&partnerID=40&md5=d1aa2eb532830d71c8eb0dddbacf3a09
[49] Peter D Turney and Michael L Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *acm Transactions on Information Systems (tois)* 21, 4 (2003), 315–346.
[50] Peng Zhang, Jing Zhang, Xindian Ma, Siwei Rao, Guangjian Tian, and Jun Wang. 2020. TextTN: Probabilistic Encoding of Language on Tensor Network. (2020).
[51] Yazhou Zhang, Qiuchi Li, Dawei Song, Peng Zhang, and Panpan Wang. 2019. Quantum-Inspired Interactive Networks for Conversational Sentiment Analysis. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 5436–5442. https://doi.org/10.24963/ijcai.2019/755
[52] Yazhou Zhang, Dawei Song, Xiang Li, and Peng Zhang. 2018. Unsupervised sentiment analysis of twitter posts using density matrix representation. In *European Conference on Information Retrieval*. Springer, 316–329.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009