QuantumCLEF - Quantum Computing at CLEF

Andrea Pasin[1](✉), Maurizio Ferrari Dacrema[2], Paolo Cremonesi[2], and Nicola Ferro[1]

[1] University of Padua, Padua, Italy
andrea.pasin.1@phd.unipd.it, nicola.ferro@unipd.it
[2] Politecnico di Milano, Milan, Italy
{maurizio.ferrari,paolo.cremonesi}@polimi.it

Abstract. Over the last few years, *Quantum Computing (QC)* has captured the attention of numerous researchers pertaining to different fields since, due to technological advancements, QC resources have become more available and also applicable in solving practical problems. In the current landscape, *Information Retrieval (IR)* and *Recommender Systems (RS)* need to perform computationally intensive operations on massive and heterogeneous datasets. Therefore, it could be possible to use QC and especially *Quantum Annealing (QA)* technologies to boost systems' performance both in terms of efficiency and effectiveness. The objective of this work is to present the first edition of the QuantumCLEF lab, which is composed of two tasks that aim at:

– evaluating QA approaches compared to their traditional counterpart;
– identifying new problem formulations to discover novel methods that leverage the capabilities of QA for improved solutions;
– establishing collaborations among researchers from different fields to harness their knowledge and skills to solve the considered challenges and promote the usage of QA.

This lab will employ the QC resources provided by CINECA, one of the most important computing centers worldwide. We also describe the design of our infrastructure which uses Docker and Kubernetes to ensure scalability, fault tolerance and replicability.

1 Introduction

In the current challenging scenario where *Information Retrieval (IR)* and *Recommender Systems (RS)* systems face ever increasing amounts of data and rely on computational demanding approaches, *Quantum Computing (QC)* can be used to improve their performance. Although QC has already been applied in several domains, limited work has been done specifically for the IR and RS fields [6,10,13]. Indeed, despite there is an area of IR called Quantum IR [9,16,17], it consists of exploiting the concepts of quantum mechanics to formulate IR models and problems but it does not deal with implementing IR and RS models and algorithms via QC technologies.

N. Goharian et al. (Eds.): ECIR 2024, LNCS 14612, pp. 482–489, 2024.
https://doi.org/10.1007/978-3-031-56069-9_66

In this work we focus on *Quantum Annealing (QA)*, which exploits special-purpose devices able to rapidly find optimal solutions to optimization problems by leveraging quantum-mechanical effects. Our goal is to understand if QA can improve the efficiency and effectiveness of IR and RS systems. So, we present a new evaluation lab called *QuantumCLEF (qCLEF)*[1] [12], which aims at:

- evaluating the performance of QA with respect to traditional approaches;
- identifying new ways of formulating IR and RS algorithms and methods, so that they can be solved with QA;
- growing a research community around this new field in order to promote a wider adoption of QC technologies for IR and RS.

Working with QA does not require particular knowledge about how quantum physics works underneath it. There are in fact available tools and libraries that can be easily used to program and solve problems through this paradigm.

The paper is organized as follows: Sect. 2 introduces related works; Sect. 3 presents the tasks in the qCLEF lab; Sect. 4 considers some critical evaluation aspects; Sect. 5 shows the design of the infrastructure for the lab; finally, Sect. 6 draws some conclusions and outlooks some future work.

2 Related Works

What is Quantum Annealing. QA is a QC paradigm that is based on special-purpose devices (quantum annealers) able to tackle optimization problems. A quantum annealer represents a problem as the energy of a physical system and then leverages quantum-mechanical phenomena to let the system find a state of minimal energy, corresponding to the solution of the original problem.

These problems need to be formulated as minimization ones using the *Quadratic Unconstrained Binary Optimization (QUBO)* formulation, defined as follows:

$$\min \quad y = x^T Q x$$

where x is a vector of binary decision variables and Q is a matrix of constant values representing the problem to solve. The QUBO formulation is very general and can represent many problems [8]. Then, the *minor embedding* step maps the QUBO problem into the quantum annealer hardware, accounting for its topology. This can be done automatically, relying on some heuristics. A QUBO problem is usually solved by quantum annealers in few *milliseconds*.

Applications of Quantum Annealing. QA can have practical applications in several fields due to its ability to tackle *NP-Hard* integer optimization problems.

QA has been previously applied to tackle IR and RS tasks such as feature selection [10], showing feasibility and promising improvements in efficiency and effectiveness. QA has also been applied to *Machine Learning (ML)* tasks. For example, Willsch et al. [19] proposes a formulation of kernel-based *Support Vector Machine (SVM)* on a D-Wave 2000Q quantum annealer, while Delilbasic

[1] https://qclef.dei.unipd.it/.

et al. [4] proposes a quantum multiclass SVM formulation aiming to reduce the execution time for large training sets. Other works explore the application of QA to clustering; for example, Zaiou et al. [21] applies it to a balanced K-means method showing better performance according to the Davies-Bouldin Index.

3 Tasks

In the qCLEF lab there are two tasks, each with the following goals:

- find one or more possible QUBO formulations of the problem;
- evaluate the quantum annealer approach compared to a corresponding traditional approach to assess both its efficiency and its effectiveness.

In general, we expect QA to solve problems more quickly than traditional approaches, achieving results that are similar better in terms of effectiveness.

3.1 Task 1 - Quantum Feature Selection

This task focuses on formulating the well-known *NP-Hard* feature selection problem to solve it with a quantum annealer, similarly to other previous works [6,10].

Feature selection is a widespread problem for both IR and RS which requires to identify a subset of the available features (e.g., the most informative, less noisy etc.) to train a learning model. This problem is very impacting, since many IR and RS systems involve the optimization of learning models, and reducing the dimensionality and noise of the input data can improve their performance.

If the input data has n features, we can enumerate all the possible sets of input data having a fixed number k of features, thus obtaining $\binom{n}{k}$ possible subsets. Therefore, to find the best subset of k features the learning model should be trained on all the subsets of features, which is infeasible even for small datasets. So, in this task we want to understand if QA can be used to solve this problem more efficiently and effectively. Feature selection fits very well the QUBO formulation: there is one variable x per feature indicating whether it should be selected or not. The challenge lies in designing the objective function, i.e., matrix Q.

We have identified some possible datasets such as MQ2007 or MQ2008 [15] and The Movies Dataset[2] which have already been used in previous works [6, 10], LETOR4.0 and MSLR-WEB30K [15]. These datasets contain pre-computed features and the objective is to select a subset of these features to train a learning model, such as LambdaMART [3] or a content-based RS, and to achieve best performance according to metrics such as nDCG@10.

3.2 Task 2 - Quantum Clustering

This task focuses on the formulation of the clustering problem to solve it with a quantum annealer. Clustering is a relevant problem for IR and RS which involves grouping items together according to their characteristics.

[2] https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset.

Clustering can be helpful for organizing large collections, helping users to explore a collection and providing similar results to a query. It can also be used to divide users according to their interests or build user models with the cluster centroids [20] boosting efficiency or effectiveness for users with limited data.

There are different clustering problem formulations, such as centroid-based Clustering or Hierarchical Clustering. In this task, each document or user can be represented as a vector in a similarity space and it is possible to cluster documents based on the similarity between each other.

Clustering fits very well with a QUBO formulation and various methods have already been proposed [1,2,18]. Most of them use variables x to represent the associated cluster to a datapoint, hence the number of points and clusters is the main limitation. There are ways to overcome this issue which can result in approximate solutions but allow the use of quantum annealers for large datasets.

For this task, we have identified MSMARCO [11] as a possible dataset, but due to the high number of documents in MSMARCO, we have identified an alternative smaller dataset such as 20 Newsgroups[3] or Wikipedia Movie Plots[4]. From the dataset we will produce embeddings using models such as BERT [5]. The cluster quality will be measured with user queries that undergo the same embedding process. These queries will match only the most representative embeddings of the clusters, avoiding computing similarities on the whole collection. For the recommendation task, we will generate user and item embeddings using state-of-the-art collaborative recommendation algorithms such as graph neural networks, on datasets Yelp and Amazon-Books. The cluster quality will be measured based on whether the centroids can be used to improve the efficiency and effectiveness of the user modeling similarly to what done in [20]. In this case the cluster quality will be measured according to the Silhouette coefficient and P@10.

3.3 Additional Challenges

Even though State-of-the-art quantum annealers nowadays have thousands of qubits (e.g., the D-Wave Advantage has $\sim$ 5000 qubits), one crucial limitation is that each qubit is physically connected only to a limited number of other qubits (15–20) in a graph of a certain topology. The process of *minor embedding* transforms the QUBO formulation in an equivalent one that fits in the particular hardware topology. This process may require to use multiple qubits to represent a single problem variable. Therefore, even if the quantum annealer has $\sim$ 5000 qubits, in practice it is possible to consider problems with at most hundreds of variables. If the problem does not fit, hybrid traditional-quantum methods exist to split the problem in smaller ones that can be solved on the quantum annealer and then combine the results. This is usually done in a general way, so a possible further challenge consists in finding better ways to split a problem in sub-problems exploiting its structure, as well as developing new problem formulations that account for the limited connectivity of quantum annealers.

[3] http://qwone.com/~jason/20Newsgroups/.
[4] https://www.kaggle.com/datasets/jrobischon/wikipedia-movie-plots.

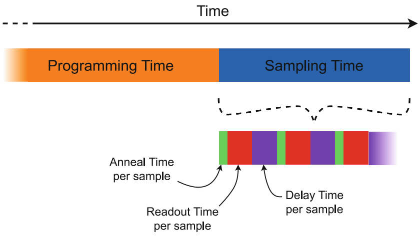

Fig. 1. The quantum annealer access time split in several steps.

4 Evaluation of Quantum Annealing

Using a quantum annealer requires several stages:

Formulation: compute the QUBO matrix Q;
Embedding: generate the *minor embedding* of the QUBO for the hardware;
Data Transfer: transfer the problem and the embedding to the datacenter that hosts the quantum annealer;
Annealing: run the quantum annealer itself.

Considering effectiveness, there are at least two layers of stochasticity. First, the embedding phase in which heuristic methods transform the QUBO formulation in an equivalent one that will fit in the hardware. This process is not deterministic: it could produce different embeddings for the same problem, that are in principle equivalent but in practice may affect the result. Second, the annealing phase that samples a low-energy solution. In some cases, many samples might be needed to get a reliable solution. Usually one selects the best solution found, but this may result in experiments with high variance. Therefore statistical evaluation measures are essential.

Considering efficiency, while the annealing phase in which the quantum annealer is actually used may last in the range of *milliseconds*, transferring the problem on the network introduces large delays and generating the minor embedding may require even minutes for particularly large problems. Furthermore, the runtime can be split in several phases, see Fig. 1: first the device needs to be programmed for the problem, then the quantum-mechanical annealing process is run and lastly the result is read. The annealing process is extremely fast, requiring *few microseconds*, but it is repeated multiple times due to the stochastic nature of the device. It is indeed necessary to consider the time requirements of all the steps involved to measure efficiency.

5 QuantumCLEF Infrastructure

We present our custom infrastructure that is required since participants cannot have direct access to quantum annealers and we want measurements to be as fair and reproducible as possible. As depicted in Fig. 2, it is composed of several components with specific purposes:

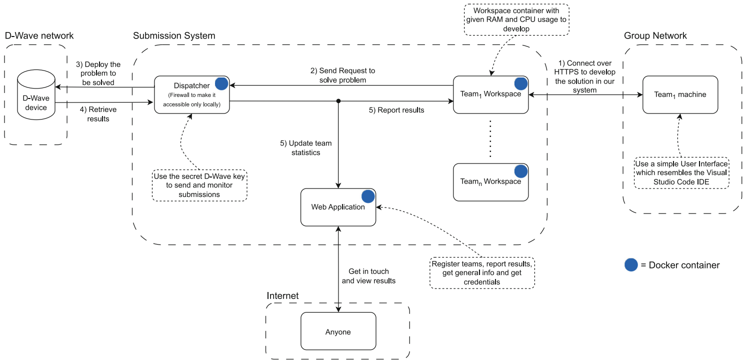

Fig. 2. High-level representation of the infrastructure.

- **Workspace**: each team has its own workspace which is accessible through the browser by providing the correct credentials. The workspace has a pre-configured git repository that is fundamental for reproducibility reasons.
- **Dispatcher**: it manages and keeps track of all the teams' submissions. It also holds the secret API Key that is used to submit problems to the quantum annealer. In this way, participants will never know the secret Key used.
- **Web Application**: it is the main source of information to the external users about the ongoing tasks. Moreover, it allows teams to view their quotas and some statistics through a dashboard. Also organizers have their own dashboard through which it is possible to manage teams and tasks.

The system can be deployed on cloud making use of different physical machines to handle several teams working together. Our infrastructure plays for QA a role similar to others, such as TIRA [14] or TIREx [7], for more general evaluation purposes. We will use the QC resources provided by CINECA that will make available D-Wave's cutting-edge quantum annealers thanks to an agreement already met.

6 Conclusions and Future Work

In this paper we have discussed the qCLEF lab, a new lab composed of two practical tasks aiming at evaluating the performance of QA applied to IR and RS. We have also discussed about the potential benefits that QA can bring to the IR and RS fields and we have highlighted how the evaluation of both efficiency and effectiveness should be performed. Finally, we have presented an infrastructure designed and implemented to satisfy both participants and organizers' needs.

qCLEF can represent a starting point for many researchers worldwide to know more about these new cutting-edge technologies that will likely have a big impact on the future of several research fields. Through this lab it will be also possible to assess whether QA can be employed to improve the current state-of-the-art approaches, hopefully delivering new performing solutions.

References

1. Arthur, D., Date, P.: Balanced k-means clustering on an adiabatic quantum computer. Quantum Inf. Process. **20**(9), 294 (2021). https://doi.org/10.1007/s11128-021-03240-8
2. Bauckhage, C., Piatkowski, N., Sifa, R., Hecker, D., Wrobel, S.: A QUBO formulation of the k-medoids problem. In: Lernen, Wissen, Daten, Analysen, Berlin, Germany, CEUR Workshop Proceedings, vol. 2454, pp. 54–63, CEUR-WS.org (2019). https://ceur-ws.org/Vol-2454/paper_39.pdf
3. Burges, C.J.C.: From RankNet to LambdaRank to LambdaMART: an overview. Technical report, Microsoft Research, MSR-TR-2010-82 (2010)
4. Delilbasic, A., Saux, B.L., Riedel, M., Michielsen, K., Cavallaro, G.: A single-step multiclass SVM based on quantum annealing for remote sensing data classification. arXiv preprint arXiv:2303.11705 (2023)
5. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
6. Ferrari Dacrema, M., Moroni, F., Nembrini, R., Ferro, N., Faggioli, G., Cremonesi, P.: Towards feature selection for ranking and classification exploiting quantum annealers. In: Proceedings of 45th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022), pp. 2814–2824. ACM Press, New York (2022)
7. Fröbe, M., et al.: The information retrieval experiment platform. In: Chen, H.H., Duh, W.J., Huang, H.H., Kato, M.P., Mothe, J., Poblete, B. (eds.) Proceedings of 46th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2023), pp. 2826–2836. ACM Press, New York (2023)
8. Glover, F.W., Kochenberger, G.A., Du, Y.: Quantum bridge analytics I: a tutorial on formulating and using QUBO models. 4OR **17**(4), 335–371 (2019). https://doi.org/10.1007/s10288-019-00424-y
9. Melucci, M.: Introduction to Information Retrieval and Quantum Mechanics. The Information Retrieval Series, vol. 35. Springer, Heidelberg (2015). https://doi.org/10.1007/978-3-662-48313-8
10. Nembrini, R., Ferrari Dacrema, M., Cremonesi, P.: Feature selection for recommender systems with quantum computing. Entropy **23**(8), 970 (2021)
11. Nguyen, T., et al.: MS Marco: a human generated machine reading comprehension dataset. Choice **2640**, 660 (2016)
12. Pasin, A., Ferrari Dacrema, M., Cremonesi, P., Ferro, N.: qCLEF: a proposal to evaluate quantum annealing for information retrieval and recommender systems. In: Arampatzis, A., et al. (eds.) CLEF 2023. LNCS, vol. 14163, pp. 97–108. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-42448-9_9
13. Pilato, G., Vella, F.: A survey on quantum computing for recommendation systems. Information **14**(1), 20 (2023). https://doi.org/10.3390/info14010020
14. Potthast, M., Gollub, T., Wiegmann, M., Stein, B.: TIRA integrated research architecture. In: Ferro, N., Peters, C. (eds.) Information Retrieval Evaluation in a Changing World. The Information Retrieval Series, vol. 41, pp. 123–160. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22948-1_5
15. Qin, T., Liu, T.Y.: Introducing LETOR 4.0 Datasets. arXiv org, Information Retrieval (cs.IR) arXiv:1306.2597 (2013)

16. van Rijsbergen, C.J.: The Geometry of Information Retrieval. Cambridge University Press, Cambridge (2004)
17. Uprety, S., Gkoumas, D., Song, D.: A survey of quantum theory inspired approaches to information retrieval. ACM Comput. Surv. (CSUR) **53**(5), 1–39 (2020)
18. Ushijima-Mwesigwa, H., Negre, C.F.A., Mniszewski, S.M.: Graph partitioning using quantum annealing on the d-wave system. CoRR abs/1705.03082 (2017). https://arxiv.org/abs/1705.03082
19. Willsch, D., Willsch, M., De Raedt, H., Michielsen, K.: Support vector machines on the d-wave quantum annealer. Comput. Phys. Commun. **248**, 107006 (2020)
20. Wu, Y., Cao, Q., Shen, H., Tao, S., Cheng, X.: INMO: a model-agnostic and scalable module for inductive collaborative filtering. In: SIGIR 2022: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, pp. 91–101. ACM (2022). https://doi.org/10.1145/3477495.3532000
21. Zaiou, A., Bennani, Y., Matei, B., Hibti, M.: Balanced k-means using quantum annealing. In: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–7 (2021). https://doi.org/10.1109/SSCI50451.2021.9659997