qCLEF: a Proposal to Evaluate Quantum Annealing for Information Retrieval and Recommender Systems

Andrea Pasin[1], Maurizio Ferrari Dacrema[2,3],
Paolo Cremonesi[2], and Nicola Ferro[1]

[1] University of Padua, Italy
pasinandre@dei.unipd.it, nicola.ferro@unipd.it
[2] Politecnico di Milano, Italy
{maurizio.ferrari, paolo.cremonesi }@polimi.it
[3] ICSC, Italy

Abstract. *Quantum Computing (QC)* has been a focus of research for many researchers over the last few years. As a result of technological development, QC resources are also becoming available and usable to solve practical problems in the *Information Retrieval (IR)* and *Recommender Systems (RS)* fields. Nowadays IR and RS need to perform complex operations on very large datasets. In this scenario, it could be possible to increase the performance of these systems both in terms of efficiency and effectiveness by employing QC and, especially, *Quantum Annealing (QA)*. The goal of this work is to design a Lab composed of different Shared Tasks that aims to:

- compare the performance of QA approaches with respect to their counterparts using traditional hardware;
- identify new ways of formulating problems so that they can be solved with quantum annealers;
- allow researchers from to different fields (e.g., Information Retrieval, Operations Research...) to work together and learn more about QA technologies.

This Lab uses the QC resources provided by CINECA, one of the most important computing centers worldwide, thanks to an already met agreement. In addition, we also show a possible implementation of the required infrastructure which uses Docker containers and the Kubernetes orchestrator to ensure scalability, fault tolerance and that can be deployed on the cloud.

1 Introduction

Information Retrieval (IR) and *Recommender Systems (RS)* play a fundamental role in providing access to and retrieving relevant resources to address our information needs. To this end, they face ever increasing amounts of data and rely on more and more computational demanding approaches. For example, search engines have to deal with the estimated 50 billions indexed pages[4] of the Web.

[4] https://www.worldwidewebsize.com/

In this challenging scenario, *Quantum Computing (QC)* can be employed to improve the performance of IR and RS methods, thanks to the development and implementation of more and more powerful QC devices, now able to tackle realistic problems. Although QC has been applied to many mathematical problems with applications in several domains, limited work has been done specifically for the IR and RS fields [6, 9, 11]. In particular, we focus on *Quantum Annealing (QA)*, which exploits a special-purpose device able to rapidly find an optimal solution to optimization problems by leveraging quantum-mechanical effects. Therefore, the goal of this work is to better understand if QA can be used to improve the efficiency and effectiveness of IR and RS systems. In particular, the contribution of this work is the design of an evaluation lab, called *Quantum CLEF (qCLEF)*, aimed at:

- evaluating the efficiency and effectiveness of QA with respect to traditional approaches;
- identifying new ways for 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.

qCLEF will consist of different tasks, each one specifically focused on a computationally-intensive problem related to IR and RS that is solvable with quantum annealers, namely:

1. **Feature selection**: identify the subset of the most relevant features that can be used to optimize a learning model.
2. **Clustering**: group items based only on their characteristics and similarities.
3. **Boosting**: find the optimal subset of *weak* predictors that can be combined together to form a *strong* predictor which performs better according to the considered dataset.

To run QA algorithms developed by qCLEF participants, we will use the QA resources provided by CINECA[5], one of the most important computing centers worldwide, located in Italy. Since participants cannot have direct access to quantum annealers, we will design and develop a dedicated infrastructure to sandbox participants' systems and execute them on the CINECA resources.

The paper is organized as follows: Section 2 discusses related works; Section 3 presents the tasks which will constitute the qCLEF lab while Section 4 introduces the design and implementation of the infrastructure for the lab; Section 5 shows a practical example of how to solve the feature selection problem, using the developed infrastructure; finally, Section 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 with

[5] https://www.cineca.it/en

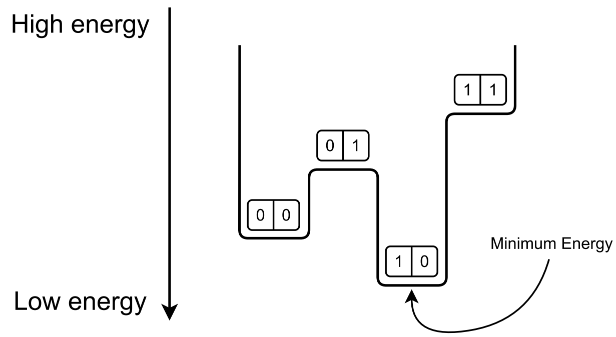


Fig. 1: Example of the Quantum Annealing process with two qubits. The optimal solution is **10**, representing the lowest point in the energy landscape.

a certain structure, such as the famous *Travelling Salesman Problem (TSP)*. The basic idea of a quantum annealer is to represent a problem as the energy of a physical system and then leverage quantum-mechanical phenomena, i.e. superposition and entanglement, to let the system find a state of minimal energy, which corresponds to the solution of the original problem. This can be seen in Figure 1, where we consider an example of the energy landscape obtained by two entangled qubits after the annealing process.

In order to use a quantum annealer, you need to formulate the optimization problem as a minimization one using the *Quadratic Unconstrained Binary Optimization (QUBO)* formulation, which is 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 we wish to solve. Note that the QUBO formulation is very general and can be used to represent many interesting problems [7]. Once the problem has been formulated as QUBO, a further step called *minor embedding* is required to map the general mathematical formulation into the physical quantum annealer hardware, accounting for the limited number of qubits and the physical connections between them. This step can be done automatically, relying on some heuristic methods. Generally, a QUBO problem can be solved by a quantum annealer in a few *milliseconds*.

Applications of Quantum Annealing. QA can have practical applications in several fields thanks to its ability to tackle integer optimization problems which are *NP-Hard*. These problems can be found in different areas such as IR, RS, banking, finance, chemistry, drug development, and many others.

Quantum annealers have been previously applied to tackle IR and RS tasks such as feature selection [9], showing the feasibility of the task and promising improved efficiency and effectiveness. Indeed, as the technology matures, these devices have the potential to offer significant speedups for *NP-Complete* and *NP-Hard* problems that are difficult to tackle on traditional hardware.

QA has also been applied for *Machine Learning (ML)* tasks. For example, Willsch et al. [16] proposes a formulation of kernel-based *Support Vector Machine (SVM)* on a D-Wave 2000Q Quantum Annealer, while Delilbasic et al. [4] proposes a quantum multiclass SVM formulation aiming to reduce the execution time as the training set size increases. Other works explore the application of QA to clustering; for example, Zaiou et al. [18] applies it to a balanced K-means method which showed better efficiency and effectiveness, according to the *Davies–Bouldin Index (DBI)*.

3 The qCLEF Proposal

In this section, we describe 3 different problems that can be solved with a quantum annealer and that correspond to different tasks in qCLEF. Each task has 2 main 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 that quantum annealers can solve problems in a shorter amount of time compared to traditional approaches obtaining results that are similar, or even better, in terms of effectiveness.

The evaluation of efficiency and effectiveness is further discussed in Section 3.5. Moreover, effectiveness will be measured according to different evaluation measures specific to each task.

3.1 Task 1 - Quantum Feature Selection

This task focuses on formulating the well-known *NP-Hard* feature selection problem in such a way that it can be solved with a quantum annealer, similarly to what has already been done in previous works [6, 9].

Feature selection is a widespread problem for both IR and RS which requires to identify a subset of the available features with certain characteristics (e.g., the most informative, less noisy etc.) to train a learning model. This problem is very impacting, since many of IR and RS systems involve the optimization of learning models, and reducing the dimensionality 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 obtain the best subset of k features we should train our learning model on all the possible $\binom{n}{k}$ subsets of features, which is infeasible even for small datasets. There are nowadays heuristics to find good solutions in a short amount of time, but they do not guarantee to find the optimal one.

Therefore, in this task, we aim to understand if QA can be applied to solve this problem more efficiently and effectively. Feature selection fits very well the QUBO formulation, in which there is one variable x per feature and its value indicates 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 [13] and The Movies Dataset[6] which have already been used in previous works [6, 9], LETOR4.0 and MSLR-WEB30K [14]. 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, in order 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 in such a way that it can be solved with a quantum annealer. Clustering is a relevant problem for IR and RS and it involves grouping the items together according to their characteristics. In this way, "similar" items fall in the same group while different items will belong to different groups. Clustering can be helpful for organizing large collections, helping users to explore a collection and providing similar search results to a given query. Furthermore, it can be helpful to divide users according to their interests or build user models with the cluster centroids [17] speeding up the runtime of the system or its effectiveness for users with limited data.

There are different clustering problem formulations, such as centroid-based Clustering or Hierarchical Clustering. In this task, we focus on centroid-based clustering, since each document can be seen as a vector in the space and it is possible to cluster points based on their distances, which can be interpreted as a dissimilarity function: the more distant two vectors are, the more different the corresponding documents are likely to be. A similar reasoning can be applied in the case of features corresponding to users.

In this context, k-means clustering has a formal definition as an optimization problem but is known to be an *NP-Hard* problem. The Lloyd's algorithm is usually employed to return an approximation of the optimal solution. However, Lloyd's algorithm does not guarantee to return the optimal clustering solution, even though it provides some computational guarantees [8]. In addition, the number of iterations needed to compute the final clusters can still be exponential.

Clustering fits very well with a QUBO formulation and various methods have already been proposed [1, 2, 15]. Most of these methods use variables x to indicate in which cluster should the data point be put, hence the number of points in the space is the main limitation. There are ways to overcome this issue, such as by applying a weighted-centroid approach, which results in an approximate solution but allows to use quantum annealers for large datasets.

For this task, we have identified as a possible dataset the MSMARCO dataset [10]. In addition, since the high number of documents in MSMARCO could be

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

an issue, we have identified a smaller dataset such as 20 Newsgroups [7]. From the considered dataset we will produce embeddings using powerful 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 found clusters, avoiding having to compute the similarity between 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 [17]. In this case the cluster quality will be measured according to the Silhouette coefficient and P@10.

3.3 Task 3 - Quantum Boosting

This task focuses on the formulation of the boosting problem for a quantum annealer. This is the most challenging task in our proposal.

Boosting is another problem that finds wide application in IR and RS. It involves identifying the best subset of *weak* predictors that can be combined together to form a *strong* predictor which performs better. A possible application of boosting is LambdaMART [3], which is a combination of LambdaRank and *Multiple Additive Regression Trees (MART)*. It uses gradient boosted decision trees with a cost function derived from LambdaRank to order documents.

Similarly to feature selection and clustering, also boosting is a combinatorial problem that cannot be solved easily. In fact, it would require to try $\binom{n}{k}$ possible subsets of weak classifiers to find the optimal one, where n is the total number of classifiers and k is the desired number of classifiers to employ. QA can provide a boost in terms of both efficiency and effectiveness, allowing to retrieve the optimal solution in few microseconds if the size of the problem is small enough to fit the quantum annealer.

Also in this case, we consider as a viable dataset the LETOR4.0 dataset [14]. The aim here is to build a strong predictor which performs the best according to the dataset itself and evaluation measures such as nDCG@10.

3.4 Additional Challenges

When using quantum annealers to solve optimization problems, identifying an appropriate QUBO formulation is only part of the challenge. State-of-the-art quantum annealers nowadays have thousands of qubits (e.g., the D-Wave Advantage has ~ 5000 qubits) and more powerful devices are planned to become available in the near future.

One crucial limitation of currently available quantum annealers is that each qubit is physically connected only to a limited number of other qubits (15-20) in a graph of a certain topology.

[7] http://qwone.com/~jason/20Newsgroups/

The process of *minor embedding* transforms the QUBO formulation in an equivalent one that fits in the particular topology of the quantum annealer. This process may require to use multiple physical qubits to represent a single problem variable, therefore even if the quantum annealer has ~ 5000, qubits in practice one can fit in its topology only problems with at most hundreds of variables.
Furthermore, if the problem does not fit on the device, 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 independently on the specific problem, thus not exploiting its possible structure and properties.

One possible further challenge consists in finding a 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 the quantum annealer.

3.5 Evaluation of Quantum Annealing

Evaluating QA approaches is not straightforward. Using a quantum annealer requires several stages:

Formulation: compute the QUBO matrix Q;
Embedding: generate the *minor embedding* of the QUBO for the quantum annealer hardware;
Data Transfer: transfer the problem and the embedding on the global network to the datacenter that hosts the quantum annealer;
Annealing: run the quantum annealer itself. This is an inherently stochastic process, therefore it is usually run a large number of times (hundreds).

Considering effectiveness, one must account for the fact that there are at least two layers of stochasticity, the embedding phase and the annealing phase. First, in the embedding phase heuristic methods transform the QUBO formulation of the problem in an equivalent problem that accounts for the limited connectivity of the physical qubits. This process includes some randomization steps and therefore may result in different embeddings for the same problem. Different embeddings will create different physical systems that are, in principle, equivalent but in practice may affect the final result.
Second, the annealing phase is a highly stochastic process and operates by *sampling* a low-energy solution, therefore depending on the problem one may require a large number of samples to obtain a good solution with sufficient probability. Usually one selects the best solution found, but this may result in experiments with high variance. Due to this, statistical evaluation measures are essential to account for the inherent stochastic behaviour of the quantum device.

Considering efficiency, while the annealing phase in which the quantum annealer is actually used may last in the range of *tens of milliseconds*, transferring the problem on the global network will introduce a delay of seconds and generating the minor embedding may require minutes for particularly large problems. Furthermore, the total quantum annealer runtime can be split in several phases,

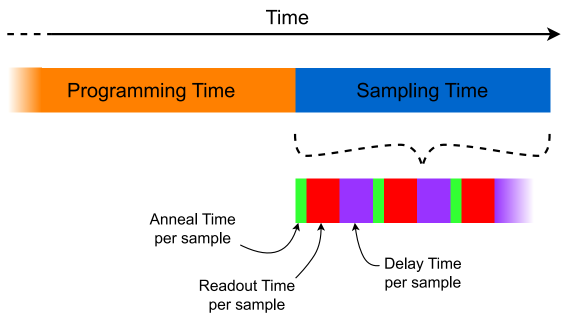


Fig. 2: The quantum annealer access time split in several steps.

see Figure 2: first the device needs to be programmed for the specific problem, then the quantum-mechanical annealing process is run and lastly the result is read. The annealing process is extremely fast, requiring in the range of *20 microseconds*, but because the device is inherently stochastic the annealing process is repeated multiple times. Clearly it is unfair to evaluate the efficiency based on a single annealing step, it is instead necessary to consider the time requirements of all the steps involved.

4 Implementation of the Infrastructure

Since participants cannot have direct access to the quantum annealers and we want the measurements to be as fair and reproducible as possible, we provide here a possible design and implementation of the infrastructure required to carry out the Lab. This infrastructure has been designed following the principles of scalability, availability, security and fault-tolerance. As depicted in Figure 3, our infrastructure is composed of several components which have specific purposes:

- **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.
  There is a custom library installed in the workspace which allows the communication with the dispatcher to submit problems to the actual quantum annealer.
- **Dispatcher**: it manages and keeps track of all the submissions done by the teams. It also holds the secret API Key that is used to submit problems to the quantum annealer. In this way, participants will never know what is the actual secret Key used.
  The dispatcher is accessible only from inside the system so that attackers cannot reach it.
- **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.

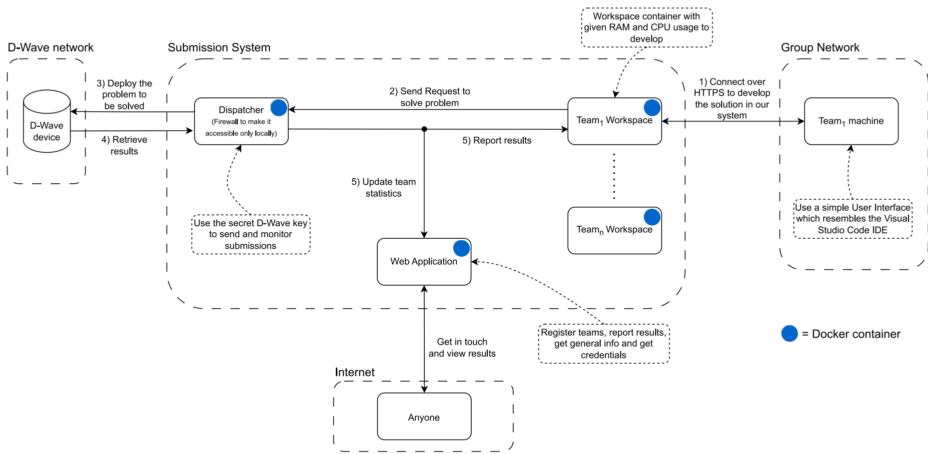


Fig. 3: High-level representation of the infrastructure.

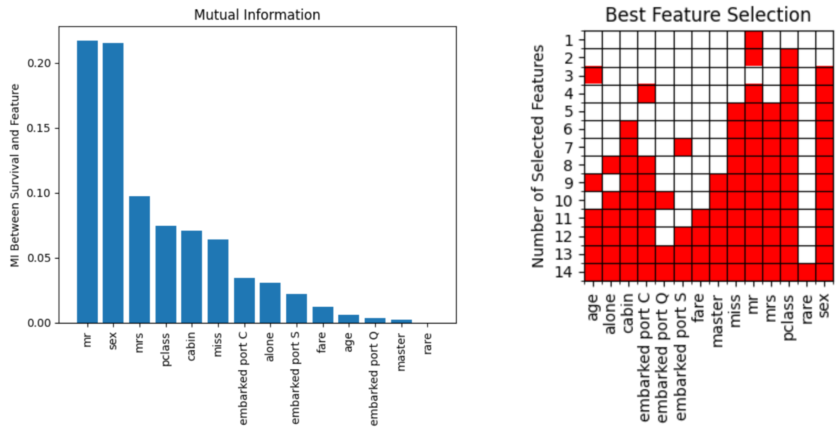
We employ technologies such as Docker containers and Kubernetes in order to make the system scalable and fault-Tolerant. In fact, the system can be deployed on cloud making use of different physical machines to handle several teams working together. We ensure to apply the correct security measures to handle possible vulnerabilities such as SQL-Injection and Cross-Site Scripting. Note that our infrastructure plays for QA a role similar to other infrastructures, such as TIRA [12], for more general evaluation purposes.

5 Feature Selection in Practice

In this section we will show how the feature selection problem can be solved with a quantum annealer using our infrastructure. This is based on an example taken from the D-Wave tutorials[8].

The task is to identify the subset of the most relevant k features that can be used to predict the survival of Titanic passengers, using the D-Wave quantum annealer. The problem is formulated as a QUBO, where the matrix Q contains the *Conditional Mutual Information* of the features associated to the row and column, and the *survival* feature. This approach is called MIQUBO and in Figure 4a it is possible to see the *Mutual Information (MI)* values considering each feature and the *survival* feature. The method further requires to define the number of features to select, k, by introducing penalties to the QUBO model so

[8] https://github.com/dwave-examples/feature-selection-notebook



(a) The representation of the Mutual Information calculated between the *survival* feature and the others present in the dataset.

(b) Subsets of selected relevant features according to subset sizes highlighted in red.

Fig. 4: Representations of the Mutual Information values and the features chosen with the QA approach.

that solutions with a different number of selected features are penalized. In total the dataset contains 15 features and 1045 rows representing passengers. We can solve the problem directly on the quantum annealer without applying hybrid approaches by calling the D-Wave APIs. Considering different values of $k \in [1, 14]$ and applying the opportune penalties, it is possible to identify which are the most relevant features that can be used to establish the survival of a passenger. This is seen in Figure 4b, where the subsets of selected relevant features for each k are highlighted in red.

Using a number of samples $n_{samples} = 100$ and a number of features $k = 4$, we report here in Table 1 the times required for some of the steps:

Table 1: Timings to solve the considered problem on the QPU for $k = 4$ features.

Access time	**Sampling time**	**Programming time**	**Anneal time per sample**
28615.97 μs	12856.0 μs	15759.97 μs	20.0 μs

Once the most relevant features for each k are identified, we can train a Tree Classifier and evaluate its classification accuracy. In order to do this, the data is split selecting 90% for training and the remaining 10% for testing. Table 2 reports the results in terms of Accuracy for $k = \{2, 3, 4, 5, 6\}$ for the Tree Classifier trained respectively on the most relevant subset of features found with

Table 2: Comparison of the Accuracy of the Tree Classifier trained on different subsets of features.

	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$
Selected Features	0.621	0.766	0.828	0.786	0.793
Highest MI features	0.621	0.621	0.786	0.793	0.793

the quantum annealer and on the subset of features that had the highest MI values with respect to the *survival* feature (e.g., for $k = 3$ *mr*, *sex*, *mrs* have the highest corresponding MI values as in Fig. 4a).
It is possible to see that the Accuracy measured on the subsets obtained through the feature selection process is always similar or even better than the Accuracy measured on the subsets having the highest MI values.

6 Conclusions and Future Work

In this paper we have proposed qCLEF, a new lab composed of 3 different tasks that aims at evaluating the performances of QA applied to IR and RS. These tasks represent some practical problems that are often faced by these systems. 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 proposed an infrastructure that has been 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 using quantum annealers.

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), URL https://ceur-ws.org/Vol-2454/paper_39.pdf
3. Burges, C.J.C.: From RankNet to LambdaRank to LambdaMART: An Overview. Tech. rep., Microsoft Research, MSR-TR-2010-82 (June 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: Proc. 45th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2022), pp. 2814–2824, ACM Press, New York, USA (2022)
7. 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
8. Lu, Y., Zhou, H.H.: Statistical and computational guarantees of lloyd's algorithm and its variants (2016)
9. Nembrini, R., Ferrari Dacrema, M., Cremonesi, P.: Feature selection for recommender systems with quantum computing. Entropy **23**(8), 970 (2021)
10. Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., Deng, L.: Ms marco: A human generated machine reading comprehension dataset. choice **2640**, 660 (2016)
11. Pilato, G., Vella, F.: A survey on quantum computing for recommendation systems. Inf. **14**(1), 20 (2023), https://doi.org/10.3390/info14010020
12. 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 – Lessons Learned from 20 Years of CLEF, The Information Retrieval Series, vol. 41, pp. 123–160, Springer International Publishing, Germany (2019)
13. Qin, T., Liu, T.Y.: Introducing letor 4.0 datasets. arXiv preprint arXiv:1306.2597 (2013)
14. Qin, T., Liu, T.Y.: Introducing LETOR 4.0 Datasets. arXiv.org, Information Retrieval (cs.IR) **arXiv:1306.2597** (June 2013)
15. 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), URL http://arxiv.org/abs/1705.03082
16. Willsch, D., Willsch, M., De Raedt, H., Michielsen, K.: Support vector machines on the d-wave quantum annealer. Computer physics communications **248**, 107006 (2020)
17. Wu, Y., Cao, Q., Shen, H., Tao, S., Cheng, X.: INMO: A model-agnostic and scalable module for inductive collaborative filtering. In: SIGIR '22: 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
18. 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