Towards a Hyperparameter-Free QUBO Formulation for Feature Selection in IR

Notebook for the QuantumCLEF Lab at CLEF 2024

Tiago Almeida[1], Sérgio Matos[1]

[1]IEETA/DETI, LASI, University of Aveiro, Aveiro, Portugal

Abstract

Feature selection is a crucial step in Learning to Rank systems, focused on identifying the most relevant features to either maintain or enhance model effectiveness while minimizing computational costs. Traditional methods frequently struggle with the NP-Hard nature of feature selection. In response, Quantum Computing, especially Quantum Annealing, has emerged as a promising alternative by recasting the problem as a Quadratic Unconstrained Binary Optimization problem.

This paper presents the participation of the University of Aveiro Biomedical Informatics and Technologies (BIT.UA) group in Task 1A of the QuantumCLEF challenge, focusing on feature selection for training LambdaMART models on the MQ2007 and ISTELLA datasets. In this work, we propose a hyperparameter-free QUBO formulation that automatically balances redundancy and relevance. Our validation on MQ2007 demonstrates competitive performance compared to traditional methods, and our application to ISTELLA shows the robustness of our approach.

Keywords

Quantum Computer, Quantum Annealer, QUBO, Feature Selection, Information Retrieval

1. Introduction

In the landscape of Information Retrieval (IR), machine learning (ML) has become instrumental in the development of Learning to Rank (LTR) systems. These systems automatically learn a ranking model, using training data, enabling them to sort new objects based on their relevance to a user query. However, the effectiveness and efficiency of these models is often closely linked to the number and quality of features they utilize. Consequently, when the feature space increases, the computational complexity also grows, often rendering these systems impractical.

Feature selection thus becomes an important step in LTR systems, aiming to identify the most relevant subset of features that maintain or enhance the model's effectiveness while reducing its computational cost. Despite its promise, feature selection is recognized as an NP-Hard problem [1] due to the exponential number of feature combinations that must be evaluated to find the optimal subset. As such, traditional methods of feature selection struggle to scale effectively with the size and complexity of modern datasets. In this context, Quantum Computing (QC) emerges as a promising alternative to explore these large search spaces. More precisely, Milne et al. [2] demonstrated that feature selection can be reformulated as a quadratic unconstrained binary optimization (QUBO) problem, making it ideally suited for execution on a Quantum Annealer (QA).

Although Milne et al. [2] have paved the way for exploring quantum solutions for feature selection, QC remains a relatively unexplored topic. In this context, the QuantumCLEF [3, 4] challenge serves as a platform to deepen the understanding of QC, with a particular focus on quantum annealers, through a series of tasks. More precisely, the challenge organizers have proposed two main tasks: Task 1, which focuses on feature selection and is further divided into feature selection for IR (1A) and feature selection for Recommender Systems (1B), and Task 2, which is centred on clustering.

CLEF 2024: Conference and Labs of the Evaluation Forum, September 9–12, 2024, Grenoble, France

tiagomeloalmeida@ua.pt (T. Almeida); aleixomatos@ua.pt (S. Matos)

0000-0002-4258-3350 (T. Almeida); 0000-0003-1941-3983 (S. Matos)

In this paper, we outline the participation of the University of Aveiro Biomedical Informatics and Technologies (BIT.UA) group in Task 1A. Specifically, this task required participants to select the most appropriate features for training a LambdaMART [5] model on the MQ2007 and ISTELLA datasets. Our approach enhances the foundational QUBO optimization framework established by Rodriguez-Lujan et al. [6] by automatically adjusting the redundancy and relevance components of the original equation. This refinement makes our method hyperparameter-free, improving its applicability.

2. Related Work

The exploration of Quantum Computing (QC) in feature selection, specially through quantum annealing, has gained momentum due to the limitations of traditional methods in handling high-dimensional datasets [7]. Traditional feature selection approaches, such as Recursive Feature Elimination (RFE) [8] and LASSO [9], often struggle with scalability and efficiency, especially in the context of high-dimensional datasets used in Learning to Rank (LTR) systems.

Early explorations in this area include a notable work by Milne et al. [2], who utilized a quantum annealer for optimal feature selection in credit prediction, marking one of the initial applications of quantum annealers to solve QUBO problems. Subsequent research expanded the use of quantum annealers across various domains: Nath et al. [10] for automated feature selection in stress detection, Otgonbaatar and Datcu [11] for feature selection and classification in hyperspectral imaging, and Qiao et al. [12] for optimizing feature selection in predicting building energy consumption, highlighting the versatility of quantum annealing.

In the context of information retrieval, Ferrari Dacrema et al. [13] have introduced three distinct QUBO formulations for feature selection: MI-QUBO, QUBO-Correlation, and QUBO-Boosting. Each model is tailored to optimize the selection process by utilizing different strategies: mutual information for MI-QUBO, Pearson correlation for QUBO-Correlation, and boosting techniques for QUBO-Boosting. Specifically, MI-QUBO enhances the feature selection process by maximizing both mutual and conditional mutual information between features and the target. QUBO-Correlation focuses on maximizing feature relevance and minimizing redundancy, whereas QUBO-Boosting assesses feature importance through classifier predictions. These methods, according to the author, showcase the potential of quantum annealing to equal or potentially exceed the performance of traditional solvers.

Despite these advances, most QUBO formulations rely on some form of hyperparameter, requiring extra experimental steps to determine optimal settings. Following this, Rodriguez-Lujan et al. [6] proposed a method to estimate the hyperparameter α used for balancing the linear and quadratic terms of the QUBO equation. Consequently, our work closely aligns with this approach, which we refer to as Auto-α throughout this paper.

3. Methodology

3.1. Feature selection as QUBO problem

Feature selection can be defined as a combinatorial optimization problem where the goal is to identify the most beneficial subset of features from a broader set. This task can be effectively modelled using Quadratic Unconstrained Binary Optimization (QUBO), a method that focuses on minimizing a quadratic objective function comprised of binary variables. In the context of feature selection, each binary variable directly corresponds to the inclusion or exclusion of a feature, where the aim is to minimize the overall cost associated with various combinations of features, thereby determining the optimal selection according to specific criteria.

More precisely, consider $x = (x_1, x_2, ..., x_n)$ as a vector of binary variables, where each x_i indicates whether a feature i is selected (1) or not selected (0). Additionally, let Q be a $n \times n$ symmetric matrix of coefficients that quantifies the impact of including each feature and their interactions. Then the objective function is given as,

$$Y(x) = \sum_{i=1}^{n} Q_{i,i} x_i + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} Q_{i,j} x_i x_j, \tag{1}$$

by leveraging the property that $x_j x_j = x_j$ for binary variables, this allows to rewrite the summation in a vectorized form:

$$Y(x) = x^T Q x, \tag{2}$$

where the QUBO model is formalized by the following optimization problem:

$$\min_x Y(x) = x^T Q x. \tag{3}$$

In practice, solving the QUBO problem effectively, especially for large datasets, requires advanced computational techniques. We focus on the use of annealers, including both simulated annealing and quantum annealing, both provided by the challenge organizers. Simulated annealing is a probabilistic technique inspired by the physical process of annealing in materials science [14]. It uses a cooling schedule to minimize energy states and find an optimal solution, allowing significant changes at high temperatures and smaller adjustments as the temperature decreases. Conversely, QA navigates a problem-encoded energy landscape to locate the minimal energy state, representing the optimal solution. In other words, while SA is based on the physical process of thermal fluctuations, QA uses the concept of quantum fluctuations which enables quantum tunnelling between states [15], resulting in faster and greater movements [16].

Regarding the definition of the coefficients of the matrix Q, Rodriguez-Lujan et al. [6], Milne et al. [2], Ferrari Dacrema et al. [13] have suggested a model-independent approach of selecting features based on their relevance and redundancy. Essentially, the goal is to include features that are highly relevant for making predictions while excluding those that are redundant. This can be achieved by employing correlation-based measures, denoted as r, both between each feature and the target variable, and among the features themselves. The former assesses the relevance of each feature in relation to the target, while the latter helps identify redundant features. Consequently, the coefficients of Q can naturally be defined as:

$$Q = \begin{cases} -r(f_i, t) & if\ i = j \\ r(f_i, f_j) & if\ i \neq j \end{cases} \tag{4}$$

Note that in this formulation, the relevant terms correspond to the linear components in Equation 1, while the redundancy terms make up the quadratic components. Due to the typically high number of quadratic terms compared to linear ones, the redundancy component tends to dominate the relevant component. This dominance results in an imbalance in Equation 1, leading to feature selection solutions that include only a small number of features. To address this issue, a common strategy involves introducing a hyperparameter to balance both components. The most typical method is a linear weighting with α:

$$Y(x) = \alpha \sum_{i=1}^{n} Q_{i,i} x_i + (1 - \alpha) \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} Q_{i,j} x_i x_j. \tag{5}$$

Another widely used alternative involves adding a penalty term to Equation 1:

$$Y(x) = \sum_{i=1}^{n} Q_{i,i} x_i + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} Q_{i,j} x_i x_j + \gamma \text{penalty}(x), \tag{6}$$

where, in the case of feature selection, the penalty is typically defined as $(\sum_{i=1}^{n} x_i - k)^2$. This term constrains the objective function to solutions that contain exactly k features, penalizing any deviation from this count.

3.2. Hyperparameter-Free QUBO

In this work, we focus on exploring a complete model-independent and hyperparameter-free QUBO formulation. We build on the methodologies of Rodriguez-Lujan et al. [6], Milne et al. [2], Ferrari Dacrema et al. [13], leveraging the notion of relevance and redundancy in the construction of Q to achieve a model-independent solution. To circumvent the need for hyperparameters, which typically address the imbalance between linear and quadratic terms, we define coefficients that are transformed and scaled to counter this imbalance. As a result, we utilize Equation 1 as our objective function and propose the following generic definition for the coefficients of Q:

$$Q = \begin{cases} \phi(r(f_i, t)) \times s & if\ i = j \\ \psi(r(f_i, f_j)) \times \frac{1}{s} & if\ i \neq j, \end{cases} \tag{7}$$

where ϕ and ψ represent non-linear transformations applied to the relevance of individual features and the redundancy between feature pairs, respectively, enabling us to modify the impact of the correlation coefficients for relevance and redundancy. For instance, they can be used to diminish the effects of lower correlation values while amplifying higher ones. On the other hand, the scalar s is a scaling factor designed to balance the weights of the linear and quadratic components within the QUBO framework. This adjustment is crucial because, in QUBO models, the redundancy (quadratic) terms often outnumber the relevance (linear) terms, potentially causing the model to overlook important features. Thus, we define our scaling factor based on the ratio of the number of quadratic pairs to linear terms, as shown in the equation:

$$s = \frac{|\{Q_{i,j} \mid i, j = 1, \ldots, n, i > j\}|}{|\{Q_{i,i} \mid i = 1, \ldots, n\}|} = \frac{\frac{n^2-n}{2}}{n} = \frac{n-1}{2}. \tag{8}$$

The choice of non-linear transformation is contingent upon the specific correlation measure used, and as such, we conduct experiments, described below, to gauge the effectiveness of different transformations.

4. Experimental Setup

In this section, we begin by outlining the datasets adopted by the QuantumCLEF [3, 4] challenge. Next, we detail our validation experiments and present our initial findings.

4.1. Datasets

For Task 1A, the organizers selected two well-established datasets in the field of Information Retrieval (IR): MQ2007 [17] and ISTELLA [18]. Both are Learning to Rank datasets, which are specifically designed to develop and test ranking algorithms through a collection of features. MQ2007 is a smaller dataset, comprising 41,955 samples and 46 features, while ISTELLA presents a more significant challenge with its 2,043,304 samples and 220 features.

4.2. Validation results

Due to severe time constraints and limited availability of quantum computing resources, we confined our validation to using only the MQ2007 dataset and simulated annealing. The smaller size of MQ2007 allowed us to efficiently test various configurations, making it a practical choice under these constraints. In terms of experimental setup, we performed feature selection, using simulation annealing, over the MQ2007 dataset using our hyperparameter-free QUBO formulation under different configurations. Subsequently, we ran a LambdaMART model with the selected features on the validation set of the MQ2007 dataset to obtain the ndcg@10 metric. This was compared to the metric achieved using all features. The objective was to achieve similar or better results while utilizing fewer features.

Regarding the configurations, we retained the scaling factor as presented in Equation 8, which corresponds to 22.5 for the MQ2007 dataset. In terms of the correlation function, r, the literature predominantly employs statistical-based measures such as Pearson and Spearman correlations, in addition to information theory measures like Mutual Information [13]. The former are fundamentally used for measuring linear and rank-based relationships, respectively. In the realm of information theory, Mutual Information is commonly utilized to quantify the amount of shared information between variables. In this work, we opted to exclusively use statistical-based measures. This decision was influenced by the well-defined nature of these measures, which are confined within the $[-1, 1]$ range. This bounded range facilitates its interpretability, where a value of 1 indicates a positive correlation, -1 indicates a negative correlation, and 0 signifies no correlation. Additionally, this range also facilitates the understanding of the behaviour of the coefficient when non-linear transformation are applied.

Lastly, we explored three specific non-linear transformations to enhance the synergy between correlation measures and the scaling factor in our model. Importantly, in our analysis, both positive and negative correlations are equally significant. To align with this approach, each transformation we selected maps negative values to positive, thereby treating all correlations with the same weight regardless of their direction:

- **Quadratic**: x^2 - This transformation dampens lower values, ensuring a smooth escalation in influence as correlations become stronger. By squaring the correlation coefficient, smaller correlations yield significantly smaller values, progressively enhancing the weight as the correlation strength approaches 1 or -1. It is important to note that the values produced by this transformation are lower than those from a linear transformation, which should harmonize effectively with the scaling factor, s.
- **Log-Quadratic**: $log(-x^2 + 1 + \epsilon)$ - This function applies a more aggressive dampening effect at lower correlation values, while sharply increasing as the correlation approaches 1/-1. This is particularly useful in emphasizing features that have very strong correlations, either positive or negative.
- **Absolute** value: $|x|$ - This linear function acts as a control to see the importance of using non-linear functions.

We present the results of our validation experiments in Table 1, which outlines the performance of various non-linear transformations and correlation measures, with a focus on the ndcg@10 metric and the number of features selected. Additionally, for further context, we included comparisons against an automated method for selecting the value of α in Equation 5, proposed by Rodriguez-Lujan et al. [6], as well as against a baseline model that uses all available features.

Table 1
Validation experiments. Note: The symbol "(-)" indicates the inversion of the signal of the function when integrated into matrix Q. This inversion is employed because we desire ϕ to reduce our objective, whereas ψ is intended to increase it.

ϕ	ψ	r	ndcg@10	Number of Features
(-) Quadratic	Quadratic	Pearson	0.5876	19
(-) Quadratic	Quadratic	Spearman	0.5864	21
Log-Quadratic	(-) Log-Quadratic	Pearson	0.5897	13
Log-Quadratic	(-) Log-Quadratic	Spearman	0.5895	15
Log-Quadratic	Quadratic	Pearson	**0.5934**	16
Log-Quadratic	Quadratic	Spearman	0.5864	19
(-) Absolute	Absolute	Pearson	0.5870	38
(-) Absolute	Absolute	Spearman	0.5882	38
Auto-α [6]			0.5761	7
All features			0.5870	46

As shown, all configurations outperformed the **Auto-α** approach, suggesting that the scaling factor, s, serves as a more effective mechanism to balance both relevance and redundancy components. Furthermore, when examining the number of features, it appears that the **Auto-α** method is more biased towards controlling redundancy, which may explain its inferior results. Pearson correlation demonstrates a preferable advantage, especially when combined with non-linear transformations, as it achieved better scores with fewer selected features. Additionally, non-linear transformations have demonstrated a positive contribution, as evidenced by their performance compared to the Absolute transformation. This is noteworthy because non-linear transformations selected fewer features while achieving comparable results. Notably, the best configuration involved using log-quadratic function as ϕ and quadratic function as ψ with Pearson correlation, achieving a ndcg@10 score of 0.5934, even surpassing the baseline that uses all available features.

To further explore the MQ2007 feature search space, we decided to conduct an additional experiment, which consisted of performing a linear search using the MI-QUBO [13] model with a penalty-k to try and identify the optimal set of features for different values of k. We varied k from 2 to 45 in single incremental steps, resulting in 44 unique runs. For each run, we trained a LambdaMART model and computed the ndcg@10 score. The best result obtained was 0.5916, using 18 features. This result was surprising, as our method was able to identify a shorter yet more effective set of features that contributed to a slightly higher score on ndcg@10, while using less resources since we do not need to perform any linear search.

5. Results

In this section, we mainly focus on describing our submission to the QuantumCLEF [3, 4] challenge, as well as the analysis of its results.

5.1. Submissions

We leveraged insights from our previous experiments to construct our submission runs. Notably, for all submissions, we defined ϕ as Log-Quadratic function, ψ as Quadratic function, $s = \frac{n-1}{2}$ and r as Pearson correlation. Additionally, when utilizing the Quantum Annealer, we used all the default parameters from D-Wave.

5.1.1. MQ2007

For MQ2007 we submitted the following four runs:

- **Run 0**: Utilizes our hyperparameter-free formulation on a quantum annealer.
- **Run 1**: Utilizes our hyperparameter-free formulation on a simulated annealer.
- **Run 2**: MI-QUBO [13] with linear search on quantum annealer (best $k = 20$).
- **Run 3**: MI-QUBO [13] with linear search on simulated annealer (best $k = 18$).

5.1.2. ISTELLA

Due to the higher number of features and their fully connected nature, running on a quantum annealer presented significant challenges, as there are fewer fully connected variables available in a quantum annealer than the total number of variables in this dataset. To circumvent this problem, we primarily divided the runs into two steps: first, a selection step that identifies the best k features, followed by the application of our hyperparameter-free formulation over this subset. Regarding our runs, we submitted the following five:

- **Run 0**: Conducted in two stages, our process began with using a simulated annealer that implemented our hyperparameter-free formulation with a penalty of $k = 110$ to select the top 110 features. Subsequently, from these 110 features, we identified the most important using our hyperparameter-free formulation on a quantum annealer.

- **Run 1**: Similar to Run 1, but after the initial selection, the rerun over the top 110 features was performed using simulated annealing.
- **Run 2**: Manually eliminated features with an absolute correlation higher than 0.85, then applied our hyperparameter-free model to the remaining features using a quantum annealer
- **Run 3**: Similar to Run 3, but after reducing the feature set, the hyperparameter-free model was run on a simulated annealer.
- **Run 4**: Utilizes our hyperparameter-free formulation on a simulated annealer.

Note that we did not utilize the MI-QUBO [13] formulation on ISTELLA, since it would be required to run a linear search for the parameter k, which becomes unfeasible to do for the 220 features.

5.2. Official results

In the QuantumCLEF [3, 4] challenge, although 25 teams registered, only 7 submitted valid entries across the different tasks. Specifically, 5 teams submitted results for the MQ2007 dataset, and 3 teams submitted for the ISTELLA dataset. The primary evaluation metric used for this challenge was ndcg@10, obtained by the LambdaMART model trained by the organizers over the features selected by each team.

5.2.1. MQ2007

The results for the MQ2007 dataset are shown in the Table 2. The performances of our runs were quite consistent, with scores hovering around the mid-44s in terms of ndcg@10, except for Run 2, which slightly outperformed the others.

Table 2
Official results on the MQ2007 dataset.

System	ndcg@10	Annealer type	Time (us)	Number of Features
Run 0	0.4410	QA	273682	18
Run 1	0.4410	SA	1351082	16
Run 2	0.4497	QA	269805	20
Run 3	0.4446	SA	3606880	18
Median	0.4410	QA	273682	18
Top-performing competitor	**0.4515**	SA	2752363	25
All features	0.4473	-	-	46
Auto select	0.4450	-	-	23

Furthermore, contrary to our validation results, our best submission was Run 2, which utilized a MI-QUBO [13] on a quantum annealer. This suggests that employing quantum technology for feature selection might offer a slight advantage over simulated annealing in certain configurations, especially when considering time efficiency. However, it is important to note that all our runs fell short of the top-performing competitor's score of 0.4515. Nonetheless, our solution managed to stay competitive while using fewer features.

5.2.2. ISTELLA

For the ISTELLA dataset, we adopted a zero-shot approach, applying the experimental setup refined on MQ2007 directly to ISTELLA without prior validation. The results from our five runs are summarized in Table 3.

The ndcg@10 scores for ISTELLA ranged from 0.6699 to 0.7081, with Run 4 achieving the highest score among our submissions. Impressively, our results significantly surpassed both the median and the top-performing competitor. However, they still fell short of the baseline that utilizes all features.

Despite this, the performance across both datasets demonstrates the robustness and adaptability of our proposed hyperparameter-free QUBO method, which consistently delivered highly competitive

Table 3
Official results on the ISTELLA dataset.

System	ndcg@10	Annealer type	Time (us)	Number of Features
Run 0	0.6699	SA + QA	16324680	92
Run 1	0.6814	SA	19071071	90
Run 2	0.6905	QA	551363	82
Run 3	0.7029	SA	5403909	72
Run 4	0.7081	SA	13827390	161
Median	0.6541	SA	3727877	25
Top-performing competitor	0.6566	SA	3874706	25
All features	**0.7146**	-	-	220

scores. Furthermore, the variability in the number of features selected for different datasets accentuates our method's ability to dynamically adapts to different sized datasets. This flexibility is likely one of the key reasons why our results on ISTELLA were considerably higher than those of other competitors.

6. Limitations

This work was conducted under severe time constrains, which naturally resulted in some limitations:

- **Lack of Rigorous Mathematical Analysis**: The methodologies and transformations used in this work were selected primarily based on empirical evaluations conducted on the MQ2007 dataset. As a result, the choices were not supported by a rigorous mathematical framework or analysis. This may limit the generalizability of the findings to other datasets.
- **Limited Dataset Testing**: The validation and testing of our method were confined to just two datasets: MQ2007 and ISTELLA. Although these datasets are well-establish within the field of information retrieval, the limited diversity in test environments restricts our ability to fully assess the model's effectiveness and adaptability.

7. Conclusions and Future Work

In this work, we proposed a hyperparameter-free QUBO formulation for feature selection, which we extensively validated on the MQ2007 dataset and subsequently tested under the QuantumCLEF challenge. Our approach demonstrated competitive results on both datasets, coupled with the advantage of not requiring any hyperparameter tuning or prior data knowledge.

For future work, it would be beneficial to pursue a more rigorous mathematical foundation to support the hyperparameter-free QUBO formulation. Additionally, testing the methodology across a larger number of datasets would be advantageous to further validate its effectiveness and adaptability.

Acknowledgments

This work was funded by the Foundation for Science and Technology (FCT) in the context of the project doi.org/10.54499/UIDB/00127/2020. Tiago Almeida is funded by the grant doi.org/10.54499/2020.05784.BD.

References

[1] W. J. Welch, Algorithmic complexity: three np- hard problems in computational statistics, Journal of Statistical Computation and Simulation 15 (1982) 17–25.

URL: https://doi.org/10.1080/00949658208810560. doi:10.1080/00949658208810560. arXiv:https://doi.org/10.1080/00949658208810560.

[2] A. Milne, M. Rounds, P. Goddard, Optimal feature selection in credit scoring and classification using a quantum annealer, 2017. URL: https://api.semanticscholar.org/CorpusID:203690617.

[3] A. Pasin, M. Ferrari Dacrema, P. Cremonesi, N. Ferro, QuantumCLEF 2024: Overview of the Quantum Computing Challenge for Information Retrieval and Recommender Systems at CLEF, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, September 9th to 12th, 2024, 2024.

[4] A. Pasin, M. Ferrari Dacrema, P. Cremonesi, N. Ferro, Overview of QuantumCLEF 2024: The Quantum Computing Challenge for Information Retrieval and Recommender Systems at CLEF, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9-12, 2024, Proceedings, 2024.

[5] C. J. C. Burges, From RankNet to LambdaRank to LambdaMART: An Overview, Technical Report, Microsoft Research, 2010. URL: http://research.microsoft.com/en-us/um/people/cburges/tech_reports/MSR-TR-2010-82.pdf.

[6] I. Rodriguez-Lujan, R. Huerta, C. Elkan, C. S. Cruz, Quadratic programming feature selection, Journal of Machine Learning Research 11 (2010) 1491–1516. URL: http://jmlr.org/papers/v11/rodriguez-lujan10a.html.

[7] A. Vlasic, A. Pham, Understanding the mapping of encode data through an implementation of quantum topological analysis, Quantum Information and Computation 23 (2023) 1091–1104. URL: http://dx.doi.org/10.26421/QIC23.13-14-2. doi:10.26421/qic23.13-14-2.

[8] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (2002) 389–422. URL: https://doi.org/10.1023/A:1012487302797. doi:10.1023/A:1012487302797.

[9] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society: Series B (Methodological) 58 (1996) 267–288.

[10] R. K. Nath, H. Thapliyal, T. S. Humble, Quantum annealing for automated feature selection in stress detection, in: 2021 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), 2021, pp. 453–457. doi:10.1109/ISVLSI51109.2021.00089.

[11] S. Otgonbaatar, M. Datcu, A quantum annealer for subset feature selection and the classification of hyperspectral images, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing PP (2021) 1–1. doi:10.1109/JSTARS.2021.3095377.

[12] Q. Qiao, A. yunusa kaltungo, R. Edwards, Feature selection strategy for machine learning methods in building energy consumption prediction, Energy Reports (2022). doi:10.1016/j.egyr.2022.10.125.

[13] M. Ferrari Dacrema, F. Moroni, R. Nembrini, N. Ferro, G. Faggioli, P. Cremonesi, Towards feature selection for ranking and classification exploiting quantum annealers, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 2814–2824. URL: https://doi.org/10.1145/3477495.3531755. doi:10.1145/3477495.3531755.

[14] P. J. M. van Laarhoven, E. H. L. Aarts, Simulated annealing: Theory and applications, in: Mathematics and Its Applications, 1987. URL: https://api.semanticscholar.org/CorpusID:61815519.

[15] T. Kadowaki, H. Nishimori, Quantum annealing in the transverse ising model, Physical Review E 58 (1998) 5355–5363. URL: http://dx.doi.org/10.1103/PhysRevE.58.5355. doi:10.1103/physreve.58.5355.

[16] C. C. McGeoch, Adiabatic quantum computation and quantum annealing: Theory and practice, in: Adiabatic Quantum Computation and Quantum Annealing, 2014. URL: https://api.semanticscholar.org/CorpusID:13341621.

[17] T. Qin, T.-Y. Liu, Introducing letor 4.0 datasets, 2013. arXiv:1306.2597.

[18] D. Dato, S. MacAvaney, F. M. Nardini, R. Perego, N. Tonellotto, The istella22 dataset: Bridging traditional and neural learning to rank evaluation, in: Proceedings of the 45th International ACM

SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, Association for Computing Machinery, New York, NY, USA, 2022, p. 3099–3107. URL: https://doi.org/10.1145/3477495.3531740. doi:10.1145/3477495.3531740.