Team R2AI at QuantumCLEF 2026: Interaction-Aware Feature Selection and Complementary-Signal Instance Selection via QUBO Optimization

Notebook for the QuantumCLEF Lab at CLEF 2026

Jiayang Niu, Yan Wang, Jie Li, Ke Deng, Azadeh Alavi and Yongli Ren

School of Computing Technologies, RMIT University, Melbourne, Victoria 3000

Abstract

This paper presents the participation of Team R2AI in QuantumCLEF 2026, focusing on Task 1B (Feature Selection for Recommender Systems) and Task 2 (Instance Selection for Noisy Text Classification). For Task 1B, we investigate the role of pairwise feature interactions in recommender-system feature selection. Building on a performance-driven QUBO framework, we derive interaction signals from recommendation-performance changes caused by jointly removing feature pairs and encode these effects as quadratic terms in the optimization objective. For Task 2, we formulate instance selection using multiple complementary signals derived from document embeddings, including representativeness, noise robustness, boundary awareness, and class balance, which are jointly incorporated into a unified QUBO objective. Furthermore, beyond the official QuantumCLEF evaluation, we conduct additional local validation experiments using observable training-data splits to better understand the behaviour of the proposed feature-selection approach. The analysis disentangles the effects of individual feature importance, pairwise interaction modelling, and annealing-based optimization, providing empirical evidence that pairwise interaction information contributes additional value beyond first-order feature effects. The official QuantumCLEF evaluation places our submissions among the top entries on both tasks, with Hybrid QA achieving the highest Macro-F1 on Task 2 and the largest NDCG@10 gains on the 400-feature benchmark.

Keywords

Pairwise Feature Interaction, Complementary Signals, Feature Selection, Instance Selection

1. Introduction

QuantumCLEF 2026, the third Quantum Computing Challenge for Information Retrieval and Recommender Systems at CLEF [1, 2], investigates the applicability of quantum and quantum-inspired optimization techniques to practical machine learning and information access problems. The benchmark provides a common evaluation framework for studying Simulated Annealing, Quantum Annealing, and hybrid quantum solvers on real-world optimization tasks. Beyond retrieval and recommendation, quantum and quantum-inspired optimization has been explored across a broad range of machine-learning problems, including quantum architecture search [3, 4], which further motivates studying annealing-based approaches to practical subset-selection tasks. In this work, Team R2AI participated in Task 1B (Feature Selection for Recommender Systems) and Task 2 (Instance Selection for Noisy Text Classification), both of which can be naturally formulated as subset selection problems.

Feature and instance selection are fundamental operations in modern information systems [5, 6, 7, 8]. In recommender systems, redundant or noisy item attributes may degrade recommendation quality while increasing computational cost. Likewise, noisy text datasets often contain uninformative or unreliable instances that hinder efficient model adaptation [9, 10]. Quadratic Unconstrained Binary Optimization (QUBO) [11] provides a natural framework for modeling such combinatorial selection problems by encoding selection decisions as binary variables and explicitly representing interactions among selected elements. Recent studies have demonstrated the effectiveness of QUBO-based formulations for recommender-system feature selection [12, 13, 14, 15] and instance selection for efficient

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany
s4068570@student.rmit.edu.au (J. Niu); s3646791@student.rmit.edu.au (Y. Wang); hey.jieli@gmail.com (J. Li); ke.deng@rmit.edu.au (K. Deng); azadeh.alavi@rmit.edu.au (A. Alavi); yongli.ren@rmit.edu.au (Y. Ren)

Transformer fine-tuning [16].

While subset selection is often driven by individual feature or instance importance, the usefulness of an element frequently depends on its relationship with other selected elements. In feature selection, two features may be complementary, redundant, or jointly influential in ways that cannot be captured through independent importance estimates alone. Similarly, effective instance selection often requires balancing multiple criteria, such as representativeness, robustness to label noise, proximity to decision boundaries, and class distribution. Consequently, selection quality may depend not only on the importance of individual elements but also on the interactions and complementary signals present within the candidate set.

Motivated by this observation, we investigate two related questions in the context of QuantumCLEF 2026. For Task 1B, we examine whether pairwise feature-interaction signals can improve recommender-system feature selection beyond individual feature importance. Specifically, we employ a performance-driven QUBO framework [12] and explicitly model the recommendation-performance changes induced by jointly removing feature pairs. These pairwise effects are encoded as quadratic interaction terms, enabling the optimization process to account for feature complementarities and redundancies when constructing the final subset.

For Task 2, we investigate how multiple complementary signals can be integrated into a unified instance-selection framework. We formulate noisy-instance selection over document embeddings by jointly considering representativeness, noise robustness, boundary awareness, and class balance. We combine these four signals in a single QUBO objective so that selection reflects multiple notions of instance quality rather than any one criterion.

Both tasks are formulated within a common QUBO optimization framework and solved using Simulated Annealing (SA) and Hybrid Quantum Annealing (Hybrid QA). Beyond the official QuantumCLEF evaluation, which is conducted on hidden test sets inaccessible to participants, we perform additional local validation experiments using observable training-data splits. These experiments provide further insight into the mechanisms underlying feature-selection performance by disentangling the contributions of individual feature importance, pairwise interaction modelling, and annealing-based optimization. In particular, the analysis enables us to examine whether pairwise interaction information contributes value beyond first-order feature effects.

On the official leaderboard, our Hybrid QA submission attains the top Macro-F1 on Task 2, while on the 400-feature Task 1B benchmark it improves NDCG@10 by roughly 20% over the organizer baseline.

2. Methodology

2.1. QUBO Formulation

Both tasks considered in this work can be formulated as subset selection problems and represented within a Quadratic Unconstrained Binary Optimization (QUBO) framework [11]. Given binary decision variables $x \in \{0, 1\}^n$, we consider the objective

$$\min_x \sum_{i=1}^{n} q_i x_i + \sum_{i<j} q_{ij} x_i x_j, \tag{1}$$

where the unary coefficients q_i capture the individual contribution of each variable and the pairwise coefficients q_{ij} model interactions between variables. To enforce a target subset size k, we incorporate an exact-cardinality penalty,

$$\min_x \sum_{i=1}^{n} q_i x_i + \sum_{i<j} q_{ij} x_i x_j + \lambda \left(\sum_{i=1}^{n} x_i - k \right)^2, \tag{2}$$

where λ controls the strength of the cardinality constraint. The resulting QUBO formulations are optimized using Simulated Annealing (SA) and, when computationally feasible, Hybrid Quantum Annealing (Hybrid QA).

2.2. Task 1B: Performance-Driven QUBO for Feature Selection

Task 1B aims to identify a subset of item-content features that improves the recommendation quality of an Item-Based K-Nearest Neighbors (Item-KNN) recommender. Rather than relying on static feature statistics, we adopt a performance-driven strategy in which QUBO coefficients are derived directly from recommendation-quality perturbations.

A locally observable validation split is first constructed from the provided training data. Let M_0 denote the recommendation performance obtained using the full Item Content Matrix and let M_i denote the performance after removing feature i. The corresponding single-feature degradation is

$$\Delta_i = M_0 - M_i. \tag{3}$$

Similarly, for a feature pair (i, j), we define

$$\Delta_{ij} = M_0 - M_{ij}, \tag{4}$$

where M_{ij} denotes the recommendation performance after jointly removing features i and j. Note that constructing these coefficients is computationally expensive, as each Δ_i and Δ_{ij} requires retraining and evaluating the Item-KNN recommender on the corresponding $(n-1)$- and $(n-2)$-feature subsets.

We introduce binary deletion variables $z_i \in \{0, 1\}$, where $z_i = 1$ indicates that feature i is removed. Let r denote the target number of removed features. The resulting performance-driven QUBO is

$$\min_{z \in \{0,1\}^n} \sum_{i=1}^{n} \Delta_i z_i + \sum_{i<j} \Delta_{ij} z_i z_j + \lambda \left(\sum_{i=1}^{n} z_i - r \right)^2 . \tag{5}$$

The unary terms quantify the recommendation-quality degradation caused by removing individual features, while the pairwise terms capture the effect of jointly removing feature pairs. After optimization, the retained feature subset corresponds to the complement of the removed-feature set, i.e., the features with $z_i = 0$.

SA is applied to the complete QUBO formulation for both ICM sizes. For the 400-feature case, the interaction matrix is too large for a single Leap Hybrid query; Hybrid QA is therefore applied to a reduced sub-BQM constructed via diagonal pre-selection, as described in the experimental setup.

2.3. Task 2: QUBO-Based Instance Selection

Task 2 focuses on selecting representative training instances from noisy document embeddings. Let $x_i \in \{0, 1\}$ indicate whether instance i is retained. We first L2-normalize all document embeddings and compute pairwise cosine similarities. Based on local neighborhood statistics and noisy class labels, we construct a fold-level QUBO that combines four complementary signals:

- **Noise robustness.** Each instance is scored by the label agreement among its k nearest neighbors, so that samples surrounded by inconsistent labels are down-weighted as likely noise.
- **Representativeness** rewards instances lying in dense, same-class neighborhoods, identifying points that summarize the bulk of their class.
- **Boundary awareness.** Instances close to reliable cross-class boundaries receive higher scores, preserving the samples that carry the most discriminative information.
- **Class balance** counteracts the over-selection of majority classes by penalizing instances from already well-represented labels.

The unary coefficient of each instance is obtained as a weighted combination of these four signals,

$$q_i^{(2)} = \alpha_{noise} s_i^{noise} - \alpha_{repr} s_i^{repr} - \alpha_{bnd} s_i^{bnd} + \alpha_{bal} s_i^{bal}, \tag{6}$$

where the α coefficients control the relative importance of the different criteria.

The pairwise coefficients are constructed from sparse neighborhood graphs in the embedding space. Positive couplings are assigned to highly similar same-label pairs in order to discourage redundant co-selection, whereas negative couplings are assigned to reliable cross-label boundary pairs to encourage the preservation of informative decision-boundary instances.

The resulting optimization problem is

$$\min_{x \in \{0,1\}^m} \sum_{i=1}^{m} q_i^{(2)} x_i + \sum_{i<j} q_{ij}^{(2)} x_i x_j + \lambda \left(\sum_{i=1}^{m} x_i - k \right)^2, \tag{7}$$

where k denotes the target number of retained instances.

Because each fold contains several thousand instances, directly applying Hybrid QA to the full problem is impractical. We therefore adopt a two-stage optimization strategy. First, SA is applied to the complete fold-level QUBO to obtain a globally consistent selection pattern. The fold is then partitioned into smaller subproblems, and each subproblem inherits its target keep count from the global SA solution. Hybrid QA is subsequently used as a local refinement procedure on eligible subproblems, while unsuitable subproblems retain their SA assignments. The refined subproblem solutions are finally merged to produce the fold-level instance selection.

3. Experimental Setup

3.1. Datasets and Evaluation

Task 1B uses the official feature-selection benchmark consisting of a User Rating Matrix (URM) and two Item Content Matrices (ICMs) containing 100 and 400 features, respectively. The URM contains 20,428 users, 14,607 items, and over 3.2 million interactions. To support local model selection, we construct an observable 80/20 holdout split using random seed 42. Local evaluation is performed with an Item-Based K-Nearest Neighbors recommender and NDCG@10, while final rankings are determined by the organizer-side hidden evaluation.

Task 2 uses the official noisy text classification benchmark, provided as five fold-wise datasets of 768-dimensional document embeddings. Each fold is processed independently and produces one Simulated Annealing (SA) submission and one Hybrid Quantum Annealing (Hybrid QA) submission.

3.2. Task 1B Configuration

For Task 1B, recommendation performance is evaluated using `ItemKNNCBFRecommender` with cosine similarity, `topK=100`, `shrink=5`, and normalization enabled. We evaluate multiple target feature counts on both ICM variants and optimize the corresponding PDQUBO formulations using SA and Hybrid QA.

The 100-feature formulation is solved directly by both SA and Hybrid QA. For the 400-feature formulation, the full QUBO contains 400 binary variables whose interaction matrix exceeds the capacity of a single Leap Hybrid submission. We therefore apply a *diagonal pre-selection* step: features are ranked in ascending order of their individual counterfactual delta Δ_i (the diagonal of the Q matrix), and the 150 features with the most negative values—those whose individual removal is least harmful (most beneficial) to performance—are retained as optimization candidates. The remaining 250 variables are fixed to zero (i.e., always retained), reducing the effective QUBO to a 150-variable sub-BQM. Hybrid QA is then applied to this reduced problem, with the cardinality constraint adjusted to the candidate pool size.

3.3. Task 2 Configuration

Task 2 uses the official fold-wise noisy text classification benchmark, where each fold consists of approximately 4,000 document embeddings of dimension 768. For each fold, we construct a QUBO

Table 1
Key statistics of the QuantumCLEF 2026 tasks.

	Task 1B	Task 2
Data type	URM + ICM	Embeddings
Users	20,428	–
Items / Instances	14,607 items	≈4,000/fold
Feature dimension	100 / 400	768
Evaluation	NDCG@10	Organizer protocol
Optimizers	SA, Hybrid QA	SA, Hybrid QA

formulation based on local neighborhood statistics in the embedding space, combining noise robustness, representativeness, boundary awareness, and class-balance signals. Exact-cardinality constraints are incorporated during optimization to control the number of retained instances.

Because each fold contains several thousand instances, directly solving the full problem with Hybrid QA is impractical. We therefore adopt the two-stage optimization strategy described in Section 2.3. A full-fold SA solve is first used to obtain a globally consistent selection pattern, after which the problem is partitioned into smaller subproblems and Hybrid QA is applied as a local refinement procedure under inherited cardinality constraints. Subproblems that are trivial or unsuitable for hybrid optimization retain their SA assignments.

3.4. Implementation Details

All experiments use a fixed random seed of 42. Simulated Annealing is implemented using the `neal` sampler, while Hybrid QA uses D-Wave's Leap Hybrid solver. Unless otherwise specified, default solver parameters are adopted. The balancing coefficients α_{noise}, α_{repr}, α_{bnd}, and α_{bal} in Section 2.3 are tunable hyper-parameters; while they can be adjusted (e.g., via validation-based tuning), in all our experiments we set them to 1, weighting each signal equally as a simple unbiased default.

4. Results

The official QuantumCLEF 2026 evaluation results were released by the organizers and form the primary basis for assessing our submissions. In this section, we first report the official leaderboard results achieved by Team R2AI on Task 1B and Task 2 before presenting additional analyses.

4.1. Official Results for Task 1B

Table 2 summarizes the official results obtained on the two feature-selection benchmarks. For the 100-feature ICM, the best SA submission achieved an NDCG@10 of 0.0256 with 97 retained features, while the best Hybrid-QA submission achieved 0.0245 with the same feature count. Both approaches improved upon the all-feature baseline reported by the organizers.

For the 400-feature ICM, Hybrid QA consistently outperformed SA across all submitted configurations. The best overall result was obtained by `1B_400_ICM_QA_r2ai_004`, which achieved an NDCG@10 of 0.0396 with 380 retained features. This result exceeded both the corresponding SA submission (0.0341) and the organizer baseline (0.0328). Overall, the results indicate that Hybrid QA becomes increasingly beneficial as the feature dimensionality grows, achieving highly competitive performance on the 400-feature benchmark.

4.2. Official Results for Task 2

Table 3 reports the official Task 2 results averaged over the five evaluation folds. The Hybrid-QA submission achieved a Macro-F1 score of 75.7(3.7), substantially outperforming the SA submission, which achieved 67.8(8.9). Both methods operated at the same average reduction level of 26.02%.

Table 2
Official Task 1B results reported by QuantumCLEF 2026. H denotes Hybrid QA and S denotes Simulated Annealing.

Dataset	Submission ID	Type	Features	NDCG@10	Annealing Time (μs)
100 ICM	1B_100_ICM_QA_r2ai_001	H	95	0.0238	2,990,292
100 ICM	1B_100_ICM_QA_r2ai_002	H	92	0.0224	2,997,680
100 ICM	1B_100_ICM_QA_r2ai_003	H	97	0.0245	2,994,913
100 ICM	1B_100_ICM_QA_r2ai_004	H	90	0.0233	2,995,671
100 ICM	1B_100_ICM_QA_r2ai_005	H	96	0.0243	2,989,746
100 ICM	1B_100_ICM_SA_r2ai_001	S	95	0.0241	24,468,772
100 ICM	1B_100_ICM_SA_r2ai_002	S	92	0.0211	25,387,836
100 ICM	1B_100_ICM_SA_r2ai_003	S	97	0.0256	22,948,788
100 ICM	1B_100_ICM_SA_r2ai_004	S	90	0.0218	26,344,972
100 ICM	1B_100_ICM_SA_r2ai_005	S	96	0.0244	21,390,295
400 ICM	1B_400_ICM_QA_r2ai_001	H	370	0.0374	2,994,624
400 ICM	1B_400_ICM_QA_r2ai_002	H	375	0.0382	2,985,272
400 ICM	1B_400_ICM_QA_r2ai_003	H	390	0.0376	2,985,798
400 ICM	1B_400_ICM_QA_r2ai_004	H	380	0.0396	2,998,937
400 ICM	1B_400_ICM_QA_r2ai_005	H	385	0.0363	2,988,981
400 ICM	1B_400_ICM_SA_r2ai_001	S	370	0.0344	703,619,903
400 ICM	1B_400_ICM_SA_r2ai_002	S	375	0.0340	678,943,036
400 ICM	1B_400_ICM_SA_r2ai_003	S	390	0.0316	229,985,415
400 ICM	1B_400_ICM_SA_r2ai_004	S	380	0.0341	651,264,506
400 ICM	1B_400_ICM_SA_r2ai_005	S	385	0.0338	651,264,506

Table 3
Official Task 2 results averaged over five folds.

Submission	Type	Macro-F1	Reduction	FT+Pred (s)	Annealing (μs)
QA_r2ai_001	H	75.7(3.7)	26.02%	1458.3	191,481,640
SA_r2ai_001	S	67.8(8.9)	26.02%	1517.7	976,908,763

Hybrid QA also reduced both the average fine-tuning time and the annealing time relative to the SA baseline. At the time of evaluation, `QA_r2ai_001` achieved the highest Macro-F1 score among the participant submissions reported on the official leaderboard, demonstrating the effectiveness of combining a global SA solution with local Hybrid-QA refinement.

4.3. Local Validation Analysis

The official evaluation is conducted on the organizers' private test set, which was not available to participants. To complement the official results, we constructed an 80 / 20 holdout split from the provided training interactions and evaluated all submitted feature subsets on the local test portion. The local baseline NDCG@10 with all features retained is 0.0343 for ICM100 and 0.0478 for ICM400, both of which differ from the organizer's reported baseline (0.0328 for ICM400) owing to the different test distributions. Local and official rankings are therefore not directly comparable in absolute terms.

For ICM400, we additionally evaluate a *diagonal greedy* baseline that removes the r features with the most negative individual delta Δ_i without solving any QUBO, isolating the contribution of first-order feature importance from the pairwise interaction modelling.

Tables 4 and 5 report the local NDCG@10 results. Table 6 compares the QUBO objective energies achieved by SA and Hybrid QA.

ICM100. SA and Hybrid QA achieve comparable performance on the local split, with SA averaging $+4.1\%$ and Hybrid QA $+5.7\%$ over the all-feature baseline. The energy comparison in Table 6 confirms that Hybrid QA finds lower-energy QUBO solutions than SA in four out of five ICM100 configurations,

Table 4
Local holdout evaluation results on the 100-feature ICM. Baseline NDCG@10 (all 100 features) = 0.0343. Type: S = SA, H = Hybrid QA.

Submission	Type	Features	NDCG@10	Δ (%)
1B_100_ICM_SA_r2ai_001	S	95	0.0351	$+2.30$
1B_100_ICM_SA_r2ai_002	S	92	0.0337	-1.61
1B_100_ICM_SA_r2ai_003	S	97	0.0396	$+15.55$
1B_100_ICM_SA_r2ai_004	S	90	0.0337	-1.62
1B_100_ICM_SA_r2ai_005	S	96	0.0363	$+5.91$
SA average			*0.0357*	*+4.11*
1B_100_ICM_QA_r2ai_001	H	95	0.0372	$+8.49$
1B_100_ICM_QA_r2ai_002	H	92	0.0341	-0.45
1B_100_ICM_QA_r2ai_003	H	97	0.0382	$+11.49$
1B_100_ICM_QA_r2ai_004	H	90	0.0356	$+3.87$
1B_100_ICM_QA_r2ai_005	H	96	0.0361	$+5.22$
QA average			*0.0362*	*+5.72*

consistent with its marginally higher NDCG. The modest gap between the two methods on ICM100 is in line with the official evaluation trend.

ICM400. A markedly different pattern emerges for ICM400, where the three methods exhibit a clear hierarchy: SA ($+2.5\%$) $<$ Diagonal Greedy ($+9.5\%$) $<$ Hybrid QA ($+15.1\%$). The energy table reveals a paradox: SA consistently achieves lower (better) QUBO objective energies than Hybrid QA across all five configurations, yet SA produces substantially worse recommendation performance. This indicates that without pre-selection the solver is free to remove features anywhere in the full 400-variable space, optimizing a second-order proxy that diverges from the true nDCG objective at large removal counts.

The diagonal greedy baseline ($+9.5\%$), which selects features purely on individual Δ_i without any pairwise modelling, substantially outperforms SA ($+2.5\%$). This demonstrates that restricting removal to the most individually impactful features is a critical prerequisite for effective large-scale feature selection, regardless of whether pairwise interactions are modelled.

Crucially, Hybrid QA ($+15.1\%$) significantly outperforms the diagonal greedy baseline ($+9.5\%$), confirming that pairwise interaction information in the Q matrix *does* provide useful signal once the search space is appropriately constrained. The diagonal pre-selection aligns the candidate pool with the true objective, after which the pairwise QUBO refinement within the 150-variable sub-BQM delivers meaningful additional gains. The combination of diagonal filtering and quadratic optimization is therefore what drives Hybrid QA's advantage on the large-scale feature set.

5. Conclusion

This paper presented Team R2AI's participation in QuantumCLEF 2026 on Task 1B (feature selection) and Task 2 (instance selection) through QUBO-based annealing optimization. For Task 1B, the PDQUBO formulation derives optimization coefficients directly from counterfactual recommendation-quality measurements, enabling annealing-based feature selection that consistently improves NDCG@10 over the all-feature baseline. For Task 2, a two-stage SA–Hybrid QA pipeline combines global consistency from SA with local quantum refinement, achieving the highest reported Macro-F1 among submitted systems.

Local validation experiments on Task 1B revealed a scale-dependent distinction. On the 100-feature problem, SA and Hybrid QA achieve similar performance with QA showing a slight advantage, consistent with its marginally better QUBO solution quality. On the 400-feature problem, a diagonal greedy ablation reveals that restricting removal to the most individually impactful features is a critical prerequisite: pure SA on the full 400-variable QUBO achieves only $+2.5\%$ despite lower objective energies, while diagonal selection alone reaches $+9.5\%$. Hybrid QA further improves to $+15.1\%$, confirming that

Table 5
Local holdout evaluation results on the 400-feature ICM. Baseline NDCG@10 (all 400 features) = 0.0478. Type: S = SA, H = Hybrid QA, D = Diagonal Greedy.

Submission	Type	Features	NDCG@10	Δ (%)
1B_400_ICM_SA_r2ai_001	S	370	0.0497	+3.88
1B_400_ICM_SA_r2ai_002	S	375	0.0493	+2.99
1B_400_ICM_SA_r2ai_003	S	390	0.0464	−2.95
1B_400_ICM_SA_r2ai_004	S	380	0.0502	+4.82
1B_400_ICM_SA_r2ai_005	S	385	0.0497	+3.85
SA average			*0.0490*	*+2.52*
1B_400_ICM_QA_r2ai_001	H	370	0.0540	+12.92
1B_400_ICM_QA_r2ai_002	H	375	0.0570	+19.08
1B_400_ICM_QA_r2ai_003	H	390	0.0552	+15.27
1B_400_ICM_QA_r2ai_004	H	380	0.0563	+17.74
1B_400_ICM_QA_r2ai_005	H	385	0.0528	+10.31
QA average			*0.0551*	*+15.06*
Diag. Greedy	D	370	0.0550	+14.93
Diag. Greedy	D	375	0.0540	+12.85
Diag. Greedy	D	380	0.0505	+5.61
Diag. Greedy	D	385	0.0514	+7.39
Diag. Greedy	D	390	0.0509	+6.45
Diag. average			*0.0524*	*+9.45*

Table 6
QUBO objective energies for all feasible Task 1B solutions. Lower energy indicates a better QUBO solution. For ICM400 Hybrid QA, the energy is reported on the 150-variable sub-BQM, which equals the full 400-variable objective energy at the selected solution since the 250 fixed variables contribute zero.

Dataset	Features	Removed	SA Energy	QA Energy
100 ICM	90	10	−0.0062	−0.0068
100 ICM	92	8	−0.0100	−0.0088
100 ICM	95	5	−0.0053	−0.0071
100 ICM	96	4	−0.0052	−0.0068
100 ICM	97	3	−0.0059	−0.0079
400 ICM	370	30	≈0.0000	+0.3196
400 ICM	375	25	−0.0007	+0.2126
400 ICM	380	20	−0.0122	+0.0959
400 ICM	385	15	−0.0073	+0.0489
400 ICM	390	10	−0.0095	+0.0028

pairwise interaction information in the Q matrix remains effective once the search space is appropriately constrained by diagonal pre-selection.

6. Declaration on Generative AI in the Writing Process

During the preparation of this work, the author(s) used ChatGPT, Grammarly in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

7. Acknowledgements

This work was supported by the ARC Discovery Project (DP210100743) and RACE(RMIT Advanced Computing Ecosystem). We sincerely thank RACE for providing access to quantum computing hardware and computational infrastructure, which was essential for conducting our experiments and advancing

this research.

References

[1] A. Pasin, M. F. Dacrema, W. Cunha, M. A. Gonçalves, P. Cremonesi, N. Ferro, Quantumclef 2026: Overview of the third quantum computing challenge for information retrieval and recommender systems at CLEF, in: Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2026, Jena, Germany, 21-24 September 2026, CEUR Workshop Proceedings, CEUR-WS.org, 2026.

[2] A. Pasin, M. F. Dacrema, W. Cunha, M. A. Gonçalves, P. Cremonesi, N. Ferro, Overview of quantumclef 2026: The third quantum computing challenge for information retrieval and recommender systems at CLEF, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 17th International Conference of the CLEF Association, CLEF 2026, Jena, Germany, September 21-24, 2026, Proceedings, Lecture Notes in Computer Science, Springer, 2026.

[3] J. Niu, Y. Wang, J. Li, K. Deng, A. Alavi, M. Usman, Y. Ren, Hybrid action reinforcement learning for quantum architecture search, arXiv preprint arXiv:2511.04967 (2025).

[4] J. Niu, A. Karim, Y. Wang, J. Li, K. Deng, A. Alavi, M. Usman, Y. Ren, HamQASBench: A Hamiltonian-informed diagnostic benchmark for evaluating quantum architecture search, arXiv preprint arXiv:2607.04845 (2026).

[5] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, H. Liu, Feature selection: A data perspective, ACM computing surveys (CSUR) 50 (2017) 1–45.

[6] Y. Guo, Z. Liu, J. Tan, C. Liao, S. Yang, L. Yuan, D. Kong, Z. Chen, J. Liu, Lpfs: Learnable polarizing feature selection for click-through rate prediction, arXiv preprint arXiv:2206.00267 (2022).

[7] P. Jia, Y. Wang, Z. Du, X. Zhao, Y. Wang, B. Chen, W. Wang, H. Guo, R. Tang, Erase: Benchmarking feature selection methods for deep recommender systems, in: Proceedings of the 30th ACM SIGKDD conference on knowledge discovery and data mining, 2024, pp. 5194–5205.

[8] S. Mücke, R. Heese, S. Müller, M. Wolter, N. Piatkowski, Feature selection on quantum computers, Quantum Machine Intelligence 5 (2023) 11.

[9] J. A. Olvera-López, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, J. Kittler, A review of instance selection methods, Artificial Intelligence Review 34 (2010) 133–143.

[10] H. Brighton, C. Mellish, Advances in instance selection for instance-based learning algorithms, Data mining and knowledge discovery 6 (2002) 153–172.

[11] A. P. Punnen, The quadratic unconstrained binary optimization problem, Springer International Publishing 10 (2022) 978–3.

[12] J. Niu, J. Li, K. Deng, M. Sanderson, N. Ferro, Y. Ren, Performance-driven qubo for recommender systems on quantum annealers, ACM Transactions on Recommender Systems (2026).

[13] J. Niu, J. Li, K. Deng, Y. Ren, Cruise on quantum computing for feature selection in recommender systems, arXiv preprint arXiv:2407.02839 (2024).

[14] J. Niu, Q. Zou, J. Li, K. Deng, M. Sanderson, Y. Ren, Estimating quantum execution requirements for feature selection in recommender systems using extreme value theory, in: Proceedings of the Nineteenth ACM Conference on Recommender Systems, 2025, pp. 587–592.

[15] A. Alavi, F. Kouchmeshki, A. Alavi, Y. Ren, J. Niu, Quantum semi-random forests for qubit-efficient recommender systems, in: 2025 IEEE International Conference on Quantum Artificial Intelligence (QAI), IEEE, 2025, pp. 401–406.

[16] A. Pasin, W. Cunha, M. A. Gonçalves, N. Ferro, A quantum annealing instance selection approach for efficient and effective transformer fine-tuning, in: Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval, 2024, pp. 205–214.