Team SYZYGY at QuantumCLEF 2026: Feature Selection Using Quantum Annealing

Task 1 Evaluation of Team SYZYGY for the QuantumCLEF Lab at CLEF 2026

Dhruv Khemani[1], Benedikt Kantz[1,*]

[1]Graz University of Technology, Rechbauerstrasse 12, Graz, Austria

Abstract

Feature selection, the task of identifying the most informative subset of attributes from a large feature pool, is an NP-hard problem central to Information Retrieval (IR) and Recommender Systems (RS). This paper presents Team SYZYGY's submission to Task 1 of the QuantumCLEF 2026 Lab, formulating feature selection as a Quadratic Unconstrained Binary Optimization (QUBO) problem solved on a D-Wave quantum annealer. For Task 1A (IR), we introduce five information-theoretic QUBO formulations that organically balance feature relevance and redundancy without hard cardinality constraints, combining mutual information, conditional mutual information, interaction information, and two normalized weighted mutual information variants. On the MQ2007 benchmark, our best configurations reduce the feature count from 46 to just 21 to 24 while surpassing both the full feature and recursive feature elimination baselines. For Task 1B (RS), a simplified collaborative driven QUBO aligns content based similarity with real user behavior, with the simulated annealer successfully improving over the full feature baseline while the physical quantum annealer is hampered by embedding complications. Across both tasks, quantum hardware completes sampling an order of magnitude faster than simulated annealing, though run to run inconsistencies on the physical QPU remain a key open challenge.

Keywords

Information Retrieval, Recommender Systems, Feature Selection, Quantum Annealing, Simulated Annealing, Quadratic Unconstrained Binary Optimization, Shannon Entropy, Mutual Information, Conditional Mutual Information, Interaction Information

1. Introduction

Information Retrieval (IR) and Recommender Systems (RS) are fundamentally important areas of computer science focused on providing users with relevant information and assistance in the exploration of large collections of digital data [1, 2]. To facilitate this, these systems provide functions such as ranking, classification, feature selection, amongst many others [1]. As data complexity and scale grows, the computational challenge of these tasks also increases, leading to the fact that the optimal solutions to these tasks become very resource-intensive [3]. Many problems in this field can be categorized as NP-hard problems, such as feature selection, which aims to isolate a subset of features to improve a model's effectiveness or reduce computational cost [1]. To address these computational bottlenecks, evaluation forums like QuantumCLEF explore the applicability of quantum computing architectures to traditional data discovery problems. In this context, this lab provides a standardized benchmarking environment to investigate how quantum approaches can optimize key tasks like feature selection under realistic experimental conditions.

This paper presents Team SYZYGY's contribution to this benchmarking effort by formalizing the feature selection tasks as Quadratic Unconstrained Binary Optimization (QUBO) problems. For the IR task, we introduce five different information-theoretic formulations using mutual information, conditional mutual information, and interaction information to capture feature relevance, identify synergistic pairings, and penalize redundancy without specifying a cardinality constraint. For the RS task, we apply a collaborative-driven approach that aligns content-based similarity with user behavior to

CLEF 2026 Working Notes, 21 – 24 September 2026, Jena, Germany

dhruv.khemani@student.tugraz.at (D. Khemani); benedikt.kantz@tugraz.at (B. Kantz)

https://github.com/DhruvKhemani (D. Khemani); https://dakantz.at/ (B. Kantz)

0009-0005-4681-6253 (D. Khemani); 0000-0003-3294-8421 (B. Kantz)

identify optimal feature subsets. We then present an empirical evaluation of both tasks on the MQ2007 and 100_ICM datasets, comparing the performance and computational efficiency of our formulations against classical benchmarks on both simulated and physical quantum annealing hardware. Finally, we discuss observed hardware sensitivities and the potential for future research in this domain.

2. Task overview

For the third edition of QuantumCLEF in 2026, Pasin et al. [4, 5] introduced three overarching tasks. Team SYZYGY's scope was confined to the first, which evaluated Feature Selection across two distinct subtasks: Task 1A (The IR Task) and Task 1B (The RS Task).

The primary objective of both tasks is to extract the most relevant feature subset from the base dataset. This is accomplished by transforming the initial input data into a QUBO formulation, which is then solved on a quantum annealer. For evaluation, simulated annealing is run alongside it to serve as a classical benchmark. The resulting feature selection then acts as the input for training specialized machine learning models, including those optimized for ranking tasks. As Pasin et al. [6] point out in the first edition of QuantumCLEF in 2024, minimizing both dimensionality and noise in the input data is key to potentially improving model performance.

2.1. Task 1A: Information Retrieval

This subtask applies feature selection to information retrieval, focusing specifically on learning to rank applications. Participants should isolate the most informative features within the provided training dataset to optimize a downstream LambdaMART model. The experimental setup uses two benchmark datasets: MQ2007 and ISTELLA, with ISTELLA presenting a notable scalability challenge since its total feature count exceeds the native capacity of the quantum hardware. The evaluation assesses both algorithm efficiency and model effectiveness by measuring the total annealing time required by the solver alongside the final ranking quality on unseen test data via the Normalized Discounted Cumulative Gain metric at a threshold of ten, denoted as NDCG@10.

2.2. Task 1B: Recommender Systems

This subtask applies feature selection to recommender systems, focusing on isolating a subset of features that produces the best recommendation quality. Participants should identify the most informative features within the provided dataset to optimize an Item-Based K-nearest neighbors (KNN) recommendation model. The setup utilizes a private dataset consisting of a User Rating Matrix (URM) alongside two sparse Item Content Matrix (ICM) containing 100 and 400 features respectively. For this configuration, item similarity is calculated using cosine metrics on the feature vectors, applying a shrinkage factor of five to the denominator and setting the number of neighbors to 100. The performance of the feature selection method is evaluated against two classical benchmarks: the identical Item-Based KNN model trained using all original features, and the same model trained using features selected via a Bayesian search strategy. The evaluation again uses both algorithm efficiency and model effectiveness by measuring the total annealing time required by the solver alongside the final recommendation quality on unseen test data via the NDCG@10 score.

3. Background

Before discussing the specific methodology employed in this work, it is necessary to introduce several fundamental concepts that form the theoretical foundation of the approach. As mentioned in section 2, to be able to solve a problem on a quantum annealer, it needs to first be expressed as a QUBO. This is because a quantum annealer (QA) is a type of Quantum Processing Unit (QPU) built specifically to find the lowest energy state of a physical system, which mathematically corresponds to finding the minimum of a QUBO function. However, physical QPUs have a fixed network of qubits with limited

physical connections between them. To map the QUBO to the physical quantum hardware, the problem must go through a process known as minor embedding. Introduced by Choi [7], this technique maps the logical variables onto the physical qubits, which requires careful parameter tuning to ensure the quantum hardware accurately represents the original problem during the computation.

3.1. Quadratic Unconstrained Binary Optimization (QUBO)

The Quadratic Unconstrained Binary Optimization is a way of writing optimization problems where the goal is to minimize or maximize a quadratic function of binary variables. The goal for Quantum Annealing is to minimize the energy state of the system, which means that the optimal solution should be a global minimum of a given function. Glover et al. [8] formulate QUBO as follows:

$$\text{minimize} \quad y = x^\intercal Q x,$$

where $x \in \{0,1\}^n$ and Q is a square $(n \times n)$ matrix of constants. The matrix Q is called the QUBO matrix, and it encodes the entire problem structure inside it. It can either be encoded into its symmetric form ($Q = Q^\intercal$), where off-diagonal elements q_{ij} are replaced by the average $\frac{1}{2}(q_{ij} + q_{ji})$, or the upper triangular form, where the below-diagonal entries $q_{ji} = 0$ for $i < j$ and the combined coefficients reside in the above-diagonal entries q_{ij} for $i < j$. However, in this work, we employed a lower triangular form, where the above-diagonal entries $q_{ij} = 0$ for $i < j$ and the active coefficients reside in the below-diagonal entries q_{ji} for $i < j$.

Besides the binary constraint of the x variables, QUBO is, by definition, unconstrained, meaning that there are no additional constraints or conditions given. This is often not how many real-world problems are formulated. To address this, Glover et al. [8] show that this can be worked around, by adding penalty terms to the original function. Because of the inherent versatility of QUBO, the applied penalties depend on the problem at hand. These penalty terms are designed to penalize any solutions that violate the original constraints, effectively transforming a constrained problem into an unconstrained one. This is possible because the penalties can be directly absorbed into the Q matrix.

For instance, consider a constraint requiring exactly k variables to be selected from a set of n binary variables, expressed as $\sum_{i=1}^{n} x_i = k$. This constraint can be enforced by adding a penalty term of the form:

$$P\left(\sum_{i=1}^{n} x_i - k\right)^2, \tag{1}$$

where P is a sufficiently large penalty coefficient [2]. When exactly k variables are selected, this penalty evaluates to zero and does not contribute to the objective function. However, when the number of selected variables deviates from k, the squared difference is multiplied by P, adding a significant penalty.

Designing a balanced QUBO matrix that effectively represents both the optimization goals and these structural penalties without destabilizing the hardware solver represents the core engineering challenge of the feature selection task.

3.2. Information Theory and Shannon Entropy Derivatives

Introduced by Shannon [9], entropy serves as a fundamental metric in information theory to quantify the uncertainty associated with a random variable. For a discrete random variable X with a probability function $P(x)$, the Shannon entropy $H(X)$ is mathematically defined as:

$$H(X) = -\sum_{x \in X} P(x) \log_2 P(x),$$

where the choice of a base-2 logarithm measures the resulting uncertainty in bits. This concept naturally extends to scenarios involving multiple variables via the joint Shannon entropy $H(X, Y)$.

This quantifies the total uncertainty of a paired system of random variables X and Y, and is calculated over their joint probability distribution $P(x, y)$ as follows:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} P(x, y) \log_2 P(x, y).$$

By subtracting the individual Shannon entropy of a known variable from the overall joint uncertainty of the system, we can show how the uncertainty of one variable is reduced by the knowledge of another. This combination is known as the conditional Shannon entropy $H(X|Y)$, originally formalized by Shannon [9]:

$$H(X|Y) = H(X, Y) - H(Y).$$

In the context of feature selection, these entropy metrics provide the essential mathematical foundation needed to evaluate the raw informational capacity of individual dataset variables and their combinations.

3.3. Mutual Information and Conditional Mutual Information

Building upon these entropy frameworks, mutual information $I(X; Y)$ quantifies the amount of information shared between two random variables. Instead of measuring raw uncertainty, it determines how much knowing one variable reduces the uncertainty of the other, establishing a formal measure of statistical dependence. As defined by Cover and Thomas [10], the mutual information between two discrete random variables X and Y can be expressed as:

$$I(X; Y) = H(X) - H(X|Y). \tag{2}$$

By substituting the algebraic definition of conditional entropy into Equation 2, this shared information can be equivalently calculated directly from the individual and joint entropies as $I(X; Y) = H(X) + H(Y) - H(X, Y)$. To account for the presence of other features, conditional mutual information $I(X; Y|Z)$ measures the remaining mutual information between X and Y when a third random variable Z is already known. This is formulated by conditioning the individual entropy terms on the known context variable Z:

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z).$$

This conditional variant serves as an essential tool for the feature selection task. It allows us to evaluate whether a candidate feature X provides genuine, novel information about a target class Y, or if that information is entirely redundant given a currently selected subset of features Z.

3.4. Interaction Information

Interaction information $I(X; Y; Z)$ extends mutual information to a three-variable system by quantifying how the presence of a context variable alters the informational dependence between a pair of attributes. As detailed by Perrone and Ay [11], this metric evaluates the net balance between redundant and synergistic dependencies within the trio, and is mathematically defined as:

$$I(X; Y; Z) = I(X; Y|Z) - I(X; Y).$$

Under this standard sign convention, Perrone and Ay [11] classify a positive outcome ($I(X; Y; Z) > 0$) as synergy, indicating that the context variable Z enhances the shared information between X and Y. Conversely, the authors define a negative outcome ($I(X; Y; Z) < 0$) as redundancy, meaning that the information provided by the variables overlaps. Because a QUBO objective function is structured as a minimization problem, the matrix elements must be aligned to penalize redundant pairs with a positive cost value. To directly map these relationships into a minimization framework, this work utilizes the inverted interaction information term $-I(X; Y; Z)$ and defines $II(X; Y; Z)$ as:

$$II(X; Y; Z) = -I(X; Y; Z) = I(X; Y) - I(X; Y|Z). \tag{3}$$

This new formulation serves as the active definition of interaction information throughout our methodology, allowing the final objective function to penalize redundant features via positive additive costs and reward synergistic features via negative weights.

4. Methodology

It is paramount to design the Q matrix to be in line with the objective which is set out by the tasks in section 2. In our case, this means finding the most relevant features for our given problems.

To achieve more consistent and reliable results from the quantum hardware, we utilized a custom minor embedding via the minorminer library[1] rather than relying on the automatic default settings. This dedicated embedding search was recomputed for each formulated QUBO matrix to optimize the layout for that specific problem structure, then fixed across all corresponding sampling reads for consistency. Additionally, the native chain strength parameter was tuned to maintain the integrity of the logical variable chains during execution on the quantum annealer.

4.1. Task 1A: Information Retrieval

As a preprocessing step, all constant features were removed from the dataset. A feature with zero variance carries no information about any other variable. Its mutual information with the target is identically zero, meaning it provides absolutely no value to the model.

Our primary design goal for this task was to build a QUBO formulation that naturally balances feature relevance and subset size on its own. By letting the informational structure of the data drive the selection, it completely avoids the need to force a fixed, artificial constraint (like Equation 1) on how many features to include.

4.1.1. Standard QUBO Matrix Encoding

The diagonal elements Q_{ii} encode the standalone relevance of each candidate feature X_i with respect to the target feature T, defined as:

$$Q_{ii} = -I(T; X_i).$$

A highly negative entry for these linear elements points to a feature sharing significant information with the target, making its selection energetically favorable for the annealer.

Looking at the off-diagonal elements, we write these as Q_{ji} to match our choice of a lower triangular matrix form. These entries capture the pairwise interaction between candidate features X_i and X_j. In mutual information focused feature selection, a well-established approach is to utilize conditional mutual information to measure the residual relevance of X_i after conditioning on X_j. This exact technique was widely adopted by participants in previous editions of the QuantumCLEF Lab [12, 13, 14, 15, 16, 17]. By fundamental definition, conditional mutual information is symmetric with respect to its first two variables, meaning $I(X_i; T \mid X_j) = I(T; X_i \mid X_j)$. While changing the position of the target variable does not alter the underlying information slice, swapping the candidate features themselves does change the result since $I(T; X_i \mid X_j) \neq I(T; X_j \mid X_i)$ in general. Evaluating only a single conditioning perspective would therefore discard exactly half of the available joint information. To maintain notation consistency, we choose to keep the target variable T first in all terms. We then preserve both directions of conditioning within a single term by explicitly summing both directional terms to define the combined CMI term:

$$\text{CMI}(T; X_i; X_j) = -\big(I(T; X_i \mid X_j) + I(T; X_j \mid X_i)\big). \tag{4}$$

The off-diagonal entry of the standard baseline formulation is subsequently mapped as $Q_{ji} = \text{CMI}(T; X_i; X_j)$ for $j > i$. Because this quantity is negative, it actively rewards feature pairs that provide unique, complementary information about the target when selected together.

[1] https://docs.dwavequantum.com/en/latest/ocean/api_ref_minorminer/source/index.html

4.1.2. Penalizing Redundancy with Weighted Mutual Information

While the baseline conditional mutual information term offers a robust metric for pairwise relevance, it alone does not fulfill our optimization objectives. Both the diagonal relevance terms and the off-diagonal conditional terms are strictly non-positive, meaning every included feature simply drops the overall energy of the system. Since adding variables can only decrease the total objective value, the QUBO is minimized trivially when all features are active. The annealer consequently has no mathematical incentive to exclude overlapping attributes, completely preventing meaningful dimensionality reduction. This structural limitation demands the introduction of positive penalty terms to actively punish redundancy.

A direct and aggressive way to introduce this penalty is by including the mutual information between the candidate features themselves:

$$\mathrm{MI}(X_i; X_j) = +I(X_i; X_j).$$

Adding this positive value to Q_{ji} imposes a direct penalty based purely on the shared information between the two attributes, regardless of how relevant they are to the target class. In practice, this raw mutual information operates as an excessive penalty. Every additional feature pair accumulates an independent positive cost, creating an overly restrictive optimization landscape that drives the annealer to select at most a single feature.

To properly balance this redundancy penalty between candidate features, we evaluated multiple normalization approaches mentioned by Vinh et al. [18] before selecting two distinct weighted variants through empirical testing. Both methods scale the raw mutual information $I(X_i; X_j)$ by a normalized coefficient that strictly ranges within the interval $[0, 1]$. This modifier ensures that the final penalty remains strictly proportional to the relative overlap of the features.

The first variant scales the mutual information by its ratio to the joint entropy of the pair. This term, which we define as $\mathrm{WMI}_1(X_i; X_j)$ calculates the precise fraction of the total information pool that is duplicated between both attributes:

$$\mathrm{WMI}_1(X_i; X_j) = I(X_i; X_j) \ \cdot \ \frac{I(X_i; X_j)}{H(X_i, X_j)}.$$

By evaluating the overlap against the entire joint system, this scaling factor naturally drops when the features introduce a large amount of unique data alongside their shared information.

The second variant instead normalizes the mutual information against the individual entropy of the less informative attribute. This approach directly measures information containment by evaluating the shared data relative only to the uncertainty of the smaller feature:

$$\mathrm{WMI}_2(X_i; X_j) = I(X_i; X_j) \ \cdot \ \frac{I(X_i; X_j)}{\min\big(H(X_i), H(X_j)\big)}.$$

Unlike the first variant which considers both features together, this scaling factor reaches its maximum value of one as soon as the smaller feature is completely absorbed by the information content of the larger feature.

These two scaling methods yield distinct off-diagonal formulations, designated as Weighted_MI_1 and Weighted_MI_2 respectively:

$$\begin{aligned} Q_{ji}^{\mathrm{WMI}_1} &= \mathrm{WMI}_1(X_i; X_j) + \mathrm{CMI}(T; X_i; X_j), \\ Q_{ji}^{\mathrm{WMI}_2} &= \mathrm{WMI}_2(X_i; X_j) + \mathrm{CMI}(T; X_i; X_j). \end{aligned}$$

The first complete formulation is:

$$Q_{ji}^{\mathrm{WMI}_1} = \begin{cases} -I(T; X_i) & \text{if } j = i \\ \mathrm{WMI}_1(X_i; X_j) + \mathrm{CMI}(T; X_i; X_j) & \text{if } j > i \\ 0 & \text{otherwise} \end{cases}.$$

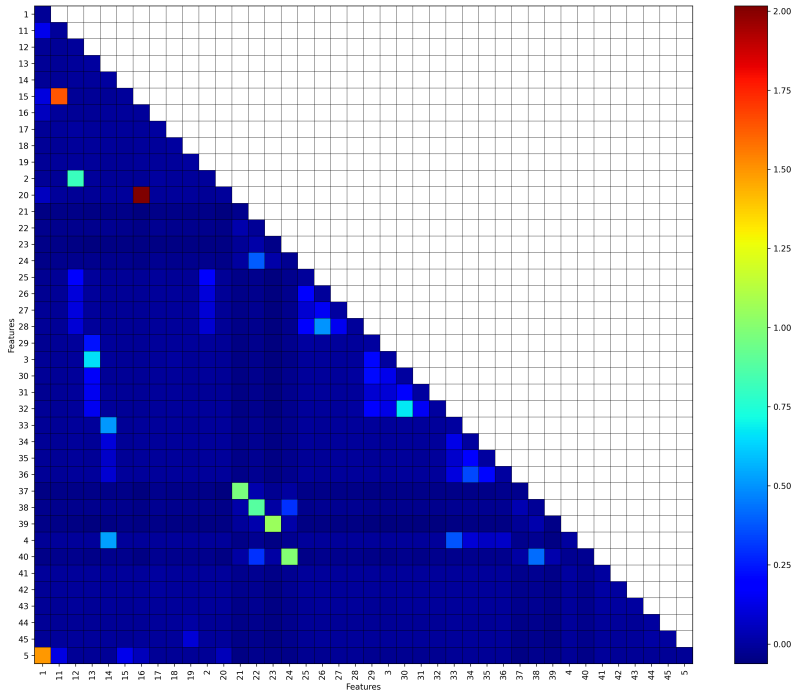

Figure 1: QUBO matrix Q^{WMI_1} illustrating the interaction weights across candidate features when the quadratic redundancy coefficient is normalized by joint entropy.

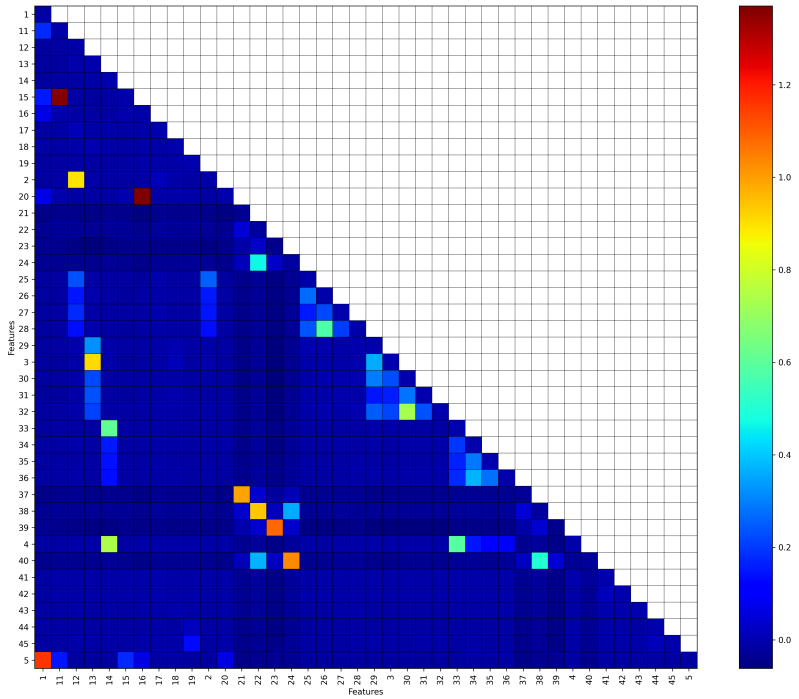

Figure 2: QUBO matrix Q^{WMI_2} illustrating the interaction weights across candidate features when the quadratic redundancy coefficient is normalized by the minimum individual entropy.

The resulting QUBO matrix for Q^{WMI_1} is illustrated in Figure 1. Notice that features 6 through 10, alongside feature 46, are absent from the matrix because of the initial preprocessing step. The second

formulation is defined similarly, substituting WMI_2 for WMI_1 in the off-diagonal entries:

$$Q_{ji}^{\mathrm{WMI}_2} = \begin{cases} -I(T;X_i) & \text{if } j = i \\ \mathrm{WMI}_2(X_i;X_j) + \mathrm{CMI}(T;X_i;X_j) & \text{if } j > i \\ 0 & \text{otherwise} \end{cases} .$$

The corresponding matrix Q^{WMI_2} is presented in Figure 2. Comparing it to Figure 1 reveals the structural impact of these distinct normalization strategies on the quadratic matrix entries. While both formulations advance the same overarching optimization objective, they reshape the quadratic penalty distribution in different ways. In the second figure (Q^{WMI_2}), more quadratic terms are visibly highlighted, showing that it penalizes a broader selection of candidate feature pairs that remain muted or relatively favored under the first formulation. At the same time, the maximum penalty applied is far higher in the Q^{WMI_1} formulation than in Q^{WMI_2}. These differences demonstrate that the two weighting approaches produce different energy landscapes for the solver to navigate, even though both work toward the same goal of reducing redundancy.

4.1.3. Capturing Synergy between Features

While the weighted mutual information variants successfully control redundancy, they focus entirely on penalizing overlapping information. We wanted to implement a mechanism to reward feature synergy, which occurs when two attributes provide more information when observed together than they do individually.

To capture these cooperative dynamics, we use the inverted interaction information $II(T;\, X_i;\, X_j)$ as established in Equation 3. This metric carries a positive value for redundant feature pairs and a negative value for synergistic ones. We visualize this synergistic and redundancy behavior in Figure 3.

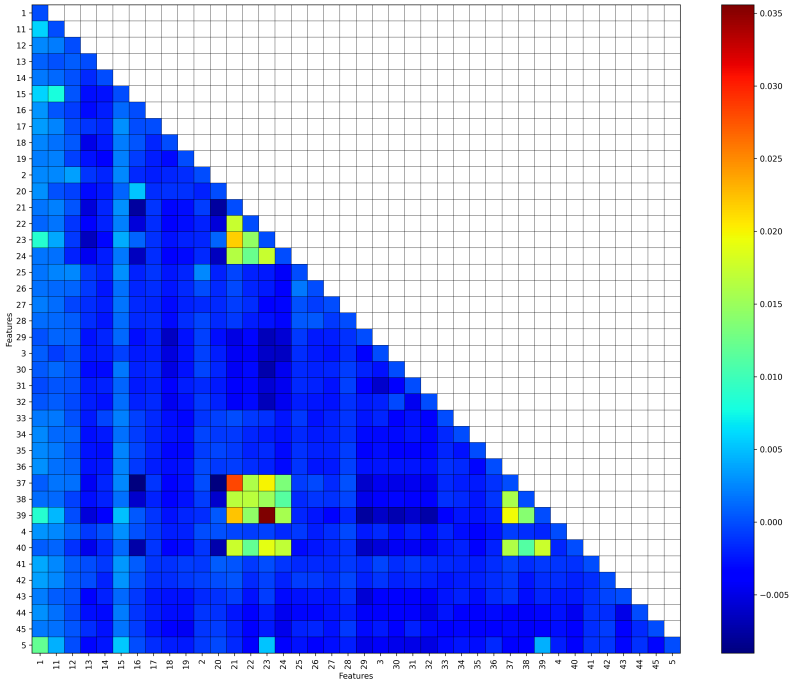

Figure 3: QUBO matrix displaying the isolated quadratic entries using only the $II(T;\, X_i;\, X_j)$ coefficient.

For this specific plot, the diagonal elements are explicitly set to zero to highlight the visual impact of the interaction information across the off-diagonal entries. Integrating this $II(T;\, X_i;\, X_j)$ term directly into our baseline off-diagonal elements introduces a dynamic cost for overlapping information

alongside a critical reward for synergy between two candidate features:

$$Q_{ji}^{II} = \begin{cases} -I(T;\, X_i) & \text{if } j = i \\ II(T;\, X_i;\, X_j) + \text{CMI}(T;\, X_i;\, X_j) & \text{if } j > i \\ 0 & \text{otherwise} \end{cases} .$$

By contrast, Figure 4 illustrates the full Q^{II} formulation. Comparing this visualization to Figure 3

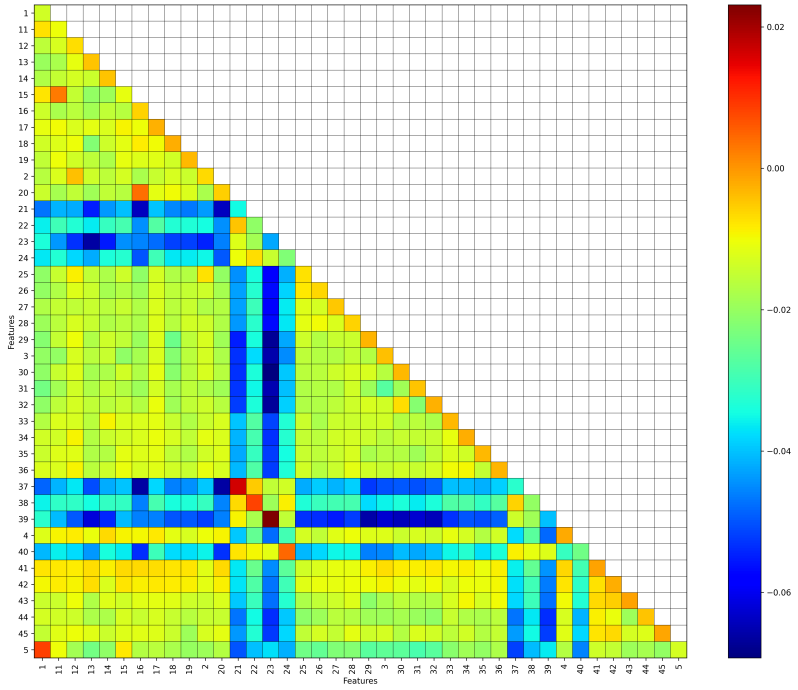

Figure 4: QUBO matrix Q^{II} illustrating the interaction landscape using interaction information and conditional mutual information.

demonstrates that the addition of the conditional mutual information imposes an entirely different pattern over the resulting matrix entries. This confirms that both mathematical components contribute different information to the QUBO formulation. Both this visual comparison and our empirical testing reveal that while the interaction information term provides a vital synergy coefficient, it acts only as a soft penalty against redundancy. This positive penalty applied by the interaction information is frequently overwhelmed by the accumulated negative weights of the linear terms and the conditional mutual information, causing the solver to still favor oversized feature subsets.

To solve this limitation, we can directly combine this synergy metric with our previous normalization approaches to construct two new QUBO matrices. Combining the interaction information with the weighted mutual information variants gives us the unified formulations $Q^{II+\text{WMI}_1}$ and $Q^{II+\text{WMI}_2}$:

$$Q_{ji}^{II+\text{WMI}_1} = \begin{cases} -I(T;\, X_i) & \text{if } j = i \\ II(T;\, X_i;\, X_j) + \text{WMI}_1(X_i;\, X_j) + \text{CMI}(T;\, X_i;\, X_j) & \text{if } j > i \\ 0 & \text{otherwise} \end{cases} .$$

$$Q_{ji}^{II+\text{WMI}_2} = \begin{cases} -I(T;\, X_i) & \text{if } j = i \\ II(T;\, X_i;\, X_j) + \text{WMI}_2(X_i;\, X_j) + \text{CMI}(T;\, X_i;\, X_j) & \text{if } j > i \\ 0 & \text{otherwise} \end{cases} .$$

These final two formulations bring together all our core optimization objectives. The linear elements and quadratic conditional mutual information terms consistently prioritize features that carry high

individual or conditional relevance to the target feature. Simultaneously, the interaction information term actively rewards cooperative pairings that exhibit genuine information synergy. Finally, the normalized mutual information weights protect the energy landscape from excessive penalization, ensuring that redundant overlap is punished proportionally without suppressing the selection of uniquely informative attributes. With these formulations, the QUBO is explicitly engineered for effective dimensionality reduction on quantum annealing hardware for feature selection applications.

4.2. Task 1B: Recommender Systems

Heavily inspired by the collaborative driven QUBO framework introduced by Nembrini et al. [2], Task 1B encodes feature selection for an ItemKNN recommender system as a binary optimization problem. Rather than adopting their full formulation, we employ a simplified variant that retains the core alignment signal while reducing the number of required hyperparameters.

The core idea is to identify item features that are consistent with how users actually interact with items. This is achieved by comparing two sources of item similarity. The first is a collaborative similarity matrix, computed directly from the URM introduced in subsection 2.2:

$$S^{CF} = \text{URM}^\top \cdot \text{URM}. \tag{5}$$

The second is a content based similarity matrix, computed from the ICM also established in that same subsection:

$$S^{CBF} = \text{ICM} \cdot \text{ICM}^\top.$$

Both matrices are then converted to binary form, reducing each entry to a simple indicator of whether any similarity exists between a given pair of items. An agreement mask K is then constructed by retaining only those item pairs for which both sources simultaneously report a positive similarity:

$$K_{ij} = \begin{cases} 1 & \text{if } S^{CF}_{ij} > 0 \text{ and } S^{CBF}_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}.$$

Intuitively, K captures the item pairs whose content based similarity is actually supported by real user behavior. This agreement mask is then projected back into the feature space via the ICM to produce the Feature Penalization Matrix (FPM):

$$\text{FPM} = \text{ICM}^\top \cdot K \cdot \text{ICM}.$$

Each entry FPM_{fg} accumulates the agreement signal across all item pairs that share both feature f and feature g, directly encoding how much selecting that feature pair is supported by collaborative evidence. The FPM is then normalized to the interval $[-1, 1]$ before further processing.

Unlike Nembrini et al., who additionally penalize item pairs with content similarity that is absent from the collaborative model, our formulation retains only the positive alignment signal. This reduces the hyperparameter space to two quantities: the target feature count k and the penalty strength P, which are incorporated via the size constraint established in the Equation 1:

$$Q = \text{FPM}_{\text{norm}} + P\left(\sum_{i=1}^{n} x_i - k\right)^2.$$

The optimal values of k and P are identified via an iterative, parallelized grid search, with each candidate configuration evaluated against the NDCG@10 of the resulting ItemKNN model on the validation split. To efficiently navigate the hyperparameter space, we progressively refined the search boundaries and step sizes from an initially broad range. The final localized grid search evaluated the target feature count $k \in \{60, 65, \ldots, 95, 100\}$ in increments of 5, alongside the penalty values $P \in \{5, 6, 7, 8, 9\}$. Through this empirical evaluation on the 100_ICM dataset, the optimal configuration was determined to be $k = 85$ with a penalty strength of $P = 7$.

5. Results

This section presents the empirical evaluation and comparative analysis of our custom QUBO formulations across both experimental tasks. As mentioned in section 2, all submissions were benchmarked against established baselines using the validation splits of the MQ2007 and 100_ICM datasets to determine the overall impact of our optimization models on retrieval and recommendation quality.

5.1. Task 1A: Information Retrieval

The evaluation of the MQ2007 dataset demonstrates the effectiveness of our custom QUBO formulations in achieving dimensionality reduction without sacrificing retrieval performance. As detailed in Table 1, the results highlight a clear success for the weighted mutual information models, which effectively halved the dataset while outperforming the baselines.

Table 1
MQ2007 feature selection results and baselines, sorted by NDCG@10.

Team	Submission ID	NDCG@10	Anneal Time (μs)	Type	Features
SYZYGY	1A_MQ2007_SA_SYZYGY_Weighted_MI_2	0.4519	4280651	S	21
SYZYGY	1A_MQ2007_QA_SYZYGY_II+weighted_MI_1	0.4513	418802	Q	23
SYZYGY	1A_MQ2007_QA_SYZYGY_Weighted_MI_2	0.4505	403362	Q	21
SYZYGY	1A_MQ2007_SA_SYZYGY_II+weighted_MI_1	0.4498	5426991	S	24
SYZYGY	1A_MQ2007_SA_SYZYGY_II+weighted_MI_2	0.4482	5778965	S	21
SYZYGY	1A_MQ2007_SA_SYZYGY_Weighted_MI_1	0.4479	4626496	S	24
SYZYGY	1A_MQ2007_QA_SYZYGY_Weighted_MI_1	0.4477	267361	Q	22
SYZYGY	1A_MQ2007_SA_SYZYGY_II	0.4473	3775185	S	40
SYZYGY	1A_MQ2007_QA_SYZYGY_II	0.4450	347840	Q	38
SYZYGY	1A_MQ2007_QA_SYZYGY_II+weighted_MI_2	0.4373	380762	Q	22
BASELINE	BASELINE_ALL	0.4473	-	-	46
BASELINE	RFE_BASELINE_HALF	0.4450	-	-	23

The absolute highest performance was achieved by the Q^{WMI_2} formulation executed on the simulated annealer, which delivered an NDCG@10 score of 0.4519 while aggressively reducing the feature subset to just 21 variables. Right behind this top result, our most successful quantum hardware run was achieved by the combined $Q^{II+\mathrm{WMI}_1}$ formulation on the quantum annealer, scoring a highly competitive 0.4513 and selecting a compact subset of 23 features. Both of these leading configurations represent a substantial improvement over the full 46 feature baseline score of 0.4473. Eight of our ten submitted configurations successfully outperformed the classical Recursive Feature Elimination baseline, proving that these optimization approaches are highly viable for this task. Furthermore, these same eight configurations achieved the highest scores among all participants in this edition of the QuantumCLEF evaluation forum.

When examining the result for the interaction information Q^{II} in isolation, the results empirically validate our theoretical observations in the methodology. Without the weighted mutual information components to enforce a strict penalty, the standalone Q^{II} models only managed a soft reduction to 38 and 40 features on the quantum and simulated annealers respectively, with the SA variant matching the baseline and the QA variant falling marginally below it at 0.4450. However, combining these approaches generally yielded the excellent results observed in our top quantum runs.

There is one notable anomaly in the results: The execution of the $Q^{II+\mathrm{WMI}_2}$ formulation on the QA resulted in a significant performance drop to 0.4373, despite its simulated counterpart performing exceptionally well. This isolated weak result on the QPU points toward a potential suboptimal minor embedding or run to run inconsistency, highlighting the sensitivity of physical qubits to complex matrix structures.

Finally, in terms of computational efficiency, the physical quantum hardware demonstrated a substantial structural speed advantage across all formulations. While the quantum annealer completed its sampling runs in under half a second, requiring roughly 260 to 420 milliseconds per submission, the

classical simulated annealer required a full order of magnitude longer, demanding between 3.7 and 5.8 seconds to traverse the exact same energy landscapes.

5.2. Task 1B: Recommender Systems

The evaluation of the collaborative driven formulation on the 100_ICM dataset yielded highly contrasting results between the simulated and quantum environments, as shown in Table 2.

Table 2
ICM_100 Submission and Baseline Performance, sorted by NDCG@10.

Team	Submission ID	NDCG@10	Anneal Time (μs)	Type	Features
SYZYGY	1B_100_ICM_SA_SYZYGY_Standard	0.0243	14244704	S	85
SYZYGY	1B_100_ICM_QA_SYZYGY_Standard	0.0152	550841	Q	62
BASELINE	ALL_FEATURES	0.0226	-	-	100

The simulated annealer successfully validated the mathematical formulation, reducing the feature set to the targeted configuration of 85 features while achieving an NDCG@10 of 0.0243. This represents a clear improvement over the 100-feature baseline score of 0.0226, proving that the simplified QUBO formulation correctly identifies and removes noisy or irrelevant content features. Conversely, the physical quantum annealer exhibited a severe decline in performance, yielding an NDCG@10 of only 0.0152 and selecting an overly aggressive subset of just 62 features.

This drop in performance on the physical quantum hardware is likely due to the fact that embedding a highly dense recommender system matrix forced the creation of exceptionally long qubit chains. While our chain strength tuning (section 4) successfully prevented chain breaks, the resulting chain lengths significantly increase the system's sensitivity to environmental factors, such as thermal fluctuations and flux noise[2] during the annealing cycle. Instead of breaking the chains apart, this ambient noise may act as a collective disturbance that steers the synchronized qubit chains into incorrect energy valleys. This mechanism could ultimately trap the system into suboptimal local energy minima that do not reflect the true mathematical objective of the QUBO.

6. Conclusion

We successfully designed and evaluated various custom QUBO formulations to optimize feature selection tasks for Information Retrieval and Recommender Systems. The empirical findings confirm that our combinatorially weighted frameworks can effectively reduce feature dimensionality while simultaneously enhancing or maintaining downstream model performance. Notably for Task 1A, eight of our ten submitted configurations outperformed the classical baseline, ultimately achieving the highest overall scores among all participants in the QuantumCLEF evaluation forum this year. Deploying these complex optimization problems directly on real quantum hardware, as can be seen in Figure 5, has proven to be an incredibly fascinating area to explore.

The differences observed between classical and quantum solutions highlight the unique, intricate nature of working with quantum systems. While physical quantum processors offer undeniable speed advantages, the observed run to run inconsistencies present critical areas for future investigation. Systematically uncovering exactly where and why these hardware failures occur, remains a compelling direction for future research. Furthermore, there is also plenty of room to expand our core feature selection methodology. Future research should focus on refining these formulations, for instance by amplifying the influence of the synergistic coefficients. A comparison of the scales between Figure 3 and Figure 4 reveals that while the interaction term provides synergistic information, its numerical magnitude is heavily overshadowed by the CMI term (Equation 4). Subsequent work could also explore alternative information-theoretic metrics or implement the additional normalization variants outlined by Vinh et al. [18] to construct even more robust QUBO models.

[2]Flux noise describes small magnetic interference on the chip that causes random shifts in qubit energy levels.

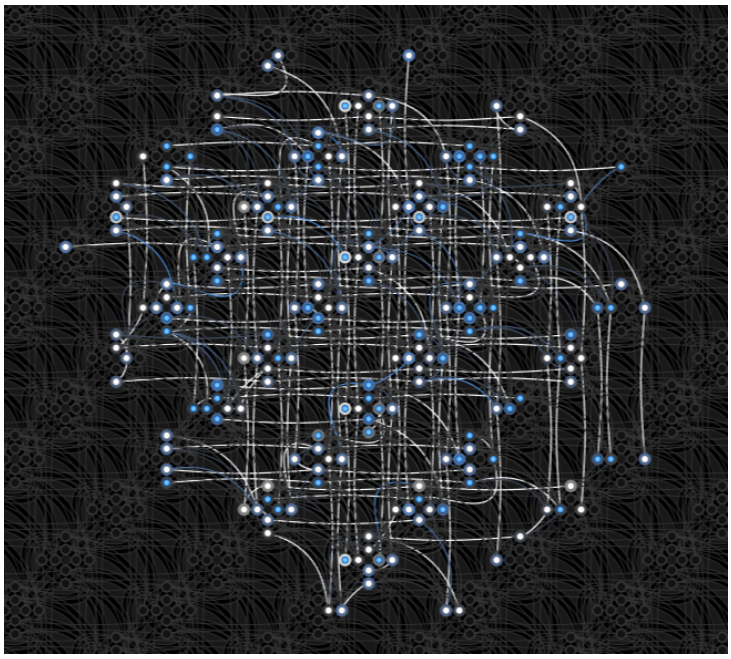

Figure 5: Screenshot from the D-Wave problem inspector showing the qubit mapping on the Advantage2 hardware during our very first successful run.

Acknowledgments

This work is partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074. The authors would also like to thank Markus Aichhorn[3] for insightful discussions regarding the physics behind quantum computing.

Declaration on Generative AI

During the preparation of this work, the authors used Gemini in order to: Grammar and spell check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. F. 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 International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, ACM, 2022, pp. 2814–2824. doi:10.48550/arXiv.2205.04346.

[2] R. Nembrini, M. F. Dacrema, P. Cremonesi, Feature selection for recommender systems with quantum computing, Entropy 23 (2021) 1–22. doi:10.48550/arXiv.2110.05089.

[3] A. Pasin, M. F. 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, Springer Nature Switzerland, 2024, pp. 260–282. doi:10.1007/978-3-031-71908-0_12.

[4] 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

[3] Markus Aichhorn ORCID: https://orcid.org/0000-0003-1034-5187

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.
[5] 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.
[6] A. Pasin, M. F. Dacrema, P. Cremonesi, N. Ferro, Quantumclef - quantum computing at CLEF, in: Advances in Information Retrieval, Springer Nature Switzerland, 2024, pp. 482–489. doi:10.1007/978-3-031-56069-9_66.
[7] V. Choi, Minor-embedding in adiabatic quantum computation: I. the parameter setting problem, Quantum Information Processing 7 (2008) 193–209. doi:10.1007/s11128-008-0082-9.
[8] F. Glover, G. Kochenberger, R. Hennig, Y. Du, Quantum bridge analytics i: a tutorial on formulating and using qubo models, Annals of Operations Research 314 (2022) 141–183. doi:10.1007/s10479-022-04634-2.
[9] C. E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (1948) 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x.
[10] T. M. Cover, J. A. Thomas, Elements of Information Theory, 2nd ed., John Wiley & Sons, Inc., 2005. doi:10.1002/047174882X.
[11] P. Perrone, N. Ay, Hierarchical quantification of synergy in channels, Frontiers in Robotics and AI 2 (2016) 1–10. doi:10.3389/frobt.2015.00035.
[12] M. Fröbe, D. Alexander, G. Hendriksen, F. Schlatt, M. Hagen, M. Potthast, Team openwebsearch at clef 2024: Quantumclef, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), CEUR Workshop Proceedings, CEUR-WS.org, 2024, pp. 2481–2496. URL: https://ceur-ws.org/Vol-3740/paper-300.pdf.
[13] A. Naebzadeh, S. Eetemadi, Nica at quantum computing clef tasks 2024, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), CEUR Workshop Proceedings, CEUR-WS.org, 2024, pp. 2779–2793. URL: https://ceur-ws.org/Vol-3740/paper-302.pdf.
[14] J. Niu, J. Li, K. Deng, Y. Ren, Cruise on quantum computing for feature selection in recommender systems, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), CEUR Workshop Proceedings, CEUR-WS.org, 2024, pp. 3269–3281. URL: https://ceur-ws.org/Vol-3740/paper-303.pdf.
[15] E. Payares, E. Puertas, J. C. Martínez-Santos, Team qtb on feature selection via quantum annealing and hybrid models, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), CEUR Workshop Proceedings, CEUR-WS.org, 2024, pp. 3694–3714. URL: https://ceur-ws.org/Vol-3740/paper-304.pdf.
[16] L. Molino-Piñar, J. Collado-Montañez, A. Montejo-Ráez, Sinai team at quantumclef 2025: Quantum feature selection based on energy with d-wave, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025), CEUR Workshop Proceedings, CEUR-WS.org, 2025, pp. 4130–4139. URL: https://ceur-ws.org/Vol-4038/paper_341.pdf.
[17] C. Pomeroy, A. Pramov, K. Thakrar, L. Yendapalli, Quantum annealing for machine learning: Applications in feature selection, instance selection, and clustering, in: Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2025), CEUR Workshop Proceedings, CEUR-WS.org, 2025, pp. 4682–4693. URL: https://ceur-ws.org/Vol-4038/paper_342.pdf.
[18] N. X. Vinh, J. Epps, J. Bailey, Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance, Journal of Machine Learning Research 11 (2010) 2837–2854. URL: http://jmlr.org/papers/v11/vinh10a.html.