Max-Sum Dispersion via Quantum Annealing

Christian Bauckhage[1,2,3], Rafet Sifa[1,2], Dirk Hecker[1,2], and Stefan Wobel[1,2,3]

[1] Fraunhofer Center for Machine Learning, Sankt Augustin, Germany
[2] Fraunhofer IAIS, Sankt Augustin, Germany
[3] B-IT, University of Bonn, Bonn, Germany

Abstract. We devise an Ising model for the max-sum dispersion problem which occurs in contexts such as Web search or text summarization. Given this Ising model, max-sum dispersion can be solved on adiabatic quantum computers; in proof of concept simulations, we solve the corresponding Schrödinger equations and observe our approach to work well.

1 Introduction

Quantum computing exploits quantum mechanical phenomena for information processing and is about to become a technical reality [7]. This will likely impact machine learning and data mining because quantum computing promises efficient solutions to certain optimization problems encountered there [2, 4, 12]. Here, we explore quantum computing for unsupervised learning and discuss how to solve the max-sum dispersion problem via adiabatic quantum computing (AQC).

Adiabatic quantum computers determine minimum energy states of Ising models, that is they solve combinatorial optimization problems of the form

$$\boldsymbol{s}^* = \underset{\boldsymbol{s}\in\{-1,+1\}^n}{\operatorname{argmin}} \; \boldsymbol{s}^\intercal \boldsymbol{Q}\, \boldsymbol{s} + \boldsymbol{s}^\intercal \boldsymbol{q}. \tag{1}$$

The 2^n vectors $\boldsymbol{s}$ over which to minimize represent possible global states of a system of n entities each of which can be in one of two local states; $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{q} \in \mathbb{R}^n$ model internal and external dependencies, respectively. Ising models as in (1) occur in various data science settings [5, 6] which thus stand to benefit from quantum computing because it promises an efficiency unreachable by digital computers.

2 An Ising Model for the Max-Sum Dispersion Problem

The max-sum dispersion problem occurs in where data mining practitioners are interested in identifying diverse or mutually far apart elements of a set of objects or observations [1, 8, 10, 11]. Given a finite data set $\mathcal{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ and a distance measure $d(\cdot, \cdot)$, the problem is to determine a subset of size $k < n$ of

maximum dispersion. In other words, the max-sum dispersion problem consists in solving the following constrained optimization problem

$$\mathcal{S}^* = \underset{\mathcal{S}\subset\mathcal{X}}{\operatorname{argmax}} \sum_{\boldsymbol{x}_i\in\mathcal{S}} \sum_{\boldsymbol{x}_j\in\mathcal{S}} d\big(\boldsymbol{x}_i, \boldsymbol{x}_j\big) \qquad \text{s.t. } |\mathcal{S}| = k. \tag{2}$$

Introducing a distance matrix $\boldsymbol{D} \in \mathbb{R}^{n\times n}$ where $D_{ij} = d\big(\boldsymbol{x}_i, \boldsymbol{x}_j\big)$ and a binary indicator vector $\boldsymbol{z} \in \{0,1\}^n$ where $z_i = 1$ if $\boldsymbol{x}_i \in \mathcal{S}$, (2) can be written as

$$\boldsymbol{z}^* = \underset{\boldsymbol{z}\in\{0,1\}^n}{\operatorname{argmax}} \ \boldsymbol{z}^\intercal \boldsymbol{D}\, \boldsymbol{z} \qquad \text{s.t. } \ \boldsymbol{z}^\intercal \mathbf{1} = k. \tag{3}$$

Equation (3) reveals max-sum dispersion to be an integer programming problem and thus to be NP-hard. State of the art solutions involve greedy approximations which achieve $O(nk)$ efficiency but lack optimality guarantees. More sophisticated algorithms provide optimality guarantees of $O(1 - 1/k)$ and require $O(nk^2 \log k)$ runtime or achieve $O(1 - 1/\epsilon)$ at $O(n/\epsilon \log k)$ runtime where the constant $\epsilon \ll 1$. Interestingly, however, problem (3) resembles the Ising energy minimization problem in (1). This suggests that it might be amenable to quantum computing.

Using the equivalence $\boldsymbol{z}^\intercal \mathbf{1} = k \ \Leftrightarrow \ \big(\boldsymbol{z}^\intercal \mathbf{1} - k\big)^2 = 0$, the problem in (3) can also be written as an unconstrained maximization problem

$$\boldsymbol{z}^* = \underset{\boldsymbol{z}\in\{0,1\}^n}{\operatorname{argmax}} \ \boldsymbol{z}^\intercal \boldsymbol{D}\, \boldsymbol{z} - \lambda \big(\boldsymbol{z}^\intercal \mathbf{1} - k\big)^2 \tag{4}$$

where $\lambda \in \mathbb{R}$ is a Lagrange multiplier. Treating this multiplier as a constant and expanding the expression on the right hand side, we find

$$\boldsymbol{z}^* = \underset{\boldsymbol{z}\in\{0,1\}^n}{\operatorname{argmin}} \ -\boldsymbol{z}^\intercal \big(\boldsymbol{D} - \lambda\, \mathbf{1}\mathbf{1}^\intercal\big)\, \boldsymbol{z} - \lambda\, 2\, k\, \boldsymbol{z}^\intercal \mathbf{1} + const \tag{5}$$

where we used that (4) is quadratic form in $\boldsymbol{z}$ so that the maximization problem can equivalently be cast in terms of a minimization problem.

Written like this, max-sum dispersion is recognizable as a problem similar to the one in (1). However, while (5) minimizes over binary vectors $\boldsymbol{z} \in \{0,1\}^n$, Ising models as in (1) involve bipolar vectors $\boldsymbol{s} \in \{-1,+1\}^n$. Yet, since $\boldsymbol{s} = 2 \cdot \boldsymbol{z} - \mathbf{1}$ is a bipolar vector so that we can write binary vectors as $\boldsymbol{z} = \big(\boldsymbol{s} + \mathbf{1}\big)/2$. Hence, using the substitutions $\boldsymbol{Q} = -\frac{1}{4}(\boldsymbol{D} - \lambda\, \mathbf{1}\mathbf{1}^\intercal)$ and $\boldsymbol{q} = -\frac{1}{2}(\boldsymbol{D} - \lambda\, \mathbf{1}\mathbf{1}^\intercal)\mathbf{1} - \lambda k \mathbf{1}$, the problem in (5) can also be written as

$$\boldsymbol{s}^* = \underset{\boldsymbol{s}\in\{-1,+1\}^n}{\operatorname{argmin}} \ \boldsymbol{s}^\intercal \boldsymbol{Q}\, \boldsymbol{s} + \boldsymbol{s}^\intercal \boldsymbol{q} \tag{6}$$

which constitutes an appropriate Ising model for max-sum dispersion. Once the solution $\boldsymbol{s}^*$ is available, entries $s_i^* = +1$ indicate which $\boldsymbol{x}_i \in \mathcal{X}$ to select into $\mathcal{S}^*$.

3 Max-Sum dispersion via AQC

When using adiabatic quantum computing in order to identify k diverse elements among a set of n elements, we consider a *time dependent* system of n qubits

$$\big|\boldsymbol{\psi}(t)\big\rangle = \sum_{i=0}^{2^n-1} a_i(t)\,\big|\boldsymbol{\psi}_i\big\rangle. \tag{7}$$

The key idea is to understand each basis state $|\boldsymbol{\psi}_i\rangle$ as an indicator vector that represents one of the 2^n subsets of our data. The task is then to manipulate the system such that, when measured afterwards, it will likely collapse to a state which encodes the sought after solution.

Manipulations that obey the laws of quantum mechanics are modeled in terms of Hamiltonian operators, and, if a qubit system as in (7) evolves under the influence of a time-dependent Hamiltonian $H(t)$, its behavior is governed by the Schrödinger equation $i\,\hbar\,\frac{d}{dt}\,\big|\boldsymbol{\psi}(t)\big\rangle = H(t)\,\big|\boldsymbol{\psi}(t)\big\rangle$. To use this for computing, one prepares a qubit register in the ground state of a problem independent *beginning Hamiltonian* H_B and evolves it towards a *problem Hamiltonian* H_P whose ground state represents a solution to the problem at hand.

If the sought after solution is known to correspond to a minimum energy configuration of an Ising model, a seminal paper by Farhi et al. [6] proposed a simple recipe for how to set up appropriate problem Hamiltonians. To set up H_P for the Ising model in (6), we therefore follow standard suggestions and define

$$H_P = \sum_{i=1}^{n}\sum_{j=1}^{n} Q_{ij}\,\sigma_z^i\,\sigma_z^j + \sum_{i=1}^{n} q_i\,\sigma_z^i \tag{8}$$

where σ_z^i denotes the Pauli spin matrix σ_z acting on the i-th qubit. Likewise and again following standard suggestions, we choose

$$H_B = -\sum_{i=1}^{n} \sigma_x^i \tag{9}$$

where σ_x^i denotes the Pauli spin matrix σ_x acting on the i-th qubit. Considering an evolution from $t=0$ to $t=T$, the Hamiltonian of the system

$$H\big(t\big) = \big(1-\tfrac{t}{T}\big)\cdot H_B + \tfrac{t}{T}\cdot H_P \tag{10}$$

can be used to let $|\boldsymbol{\psi}(t)\rangle$ evolve from the ground state $|\boldsymbol{\psi}(0)\rangle$ of H_B to a final state $|\boldsymbol{\psi}(T)\rangle$ where $|\boldsymbol{\psi}(0)\rangle$. Upon termination of this evolution, a measurement of the qubit system causes it to collapse to one of its 2^n basis states and the probability for this state to be $|\boldsymbol{\psi}_i\rangle$ is given by the amplitude $|a_i(T)|^2$. However, since the adiabatic evolution was steered towards the problem Hamiltonian H_P, states $|\boldsymbol{\psi}_i\rangle$ that correspond to ground states of H_P are more likely to be found.

The computational efficiency of adiabatic quantum computing will depend on the choice of T in (10) which is known to depend on the minimum energy gap

between the ground state and the first excited state of $H(t)$. This gap is inversely proportional to the square root of the number of basis states that have energies close to global minimum [3]. For problems such as max-sum dispersion where the number of potential solutions is typically small, $T \in O\left(\sqrt{2^n}\right)$ is the smallest possible runtime Hence, AQC can perform exhaustive searchers for max-sum dispersion solutions quadratically faster than classically possible.

4 Proof of Concept Simulations

Next, we present didactic simulation experiments to illustrate the feasibility of AQC for max-sum dispersion. In each case, we numerically solve the governing Schrödinger equations using the *Python* toolbox *QuTip* [9].

Figure 1 shows $n = 12$ monthly weather measurements in Hamburg. We normalize them to unit variance so that far points are $O(2)$ apart. Setting $\lambda = 2\,n$ will then cause neither of the terms in (6) to dominate. We consider Euclidean distances and try to extract $k \in \{4, 5, 6\}$ diverse elements. In each setting, we work with 12 qubit systems in a superposition of $2^{12} = 4096$ basis states. Using Hamiltonians as discussed above, we evolve them over $T = 100 \in O\left(\sqrt{2^{12}}\right)$ steps.

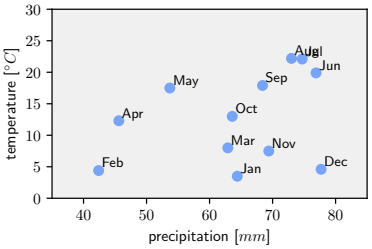
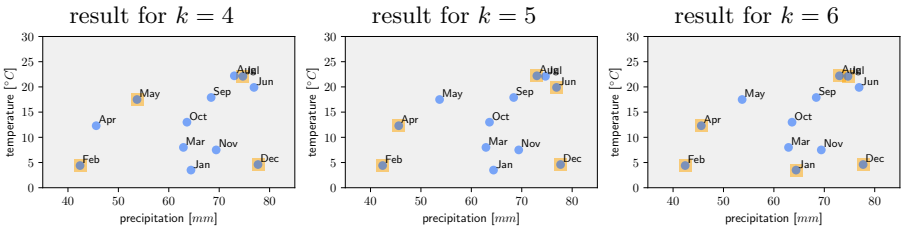


Fig. 1: Data of monthly climatic conditions in Hamburg.

The top row of Fig. 2 shows the results for each setting. The middle row visualizes how each qubit system evolves over time. At $t = 0$, all basis states are equally likely but over time their amplitudes begin to diverge; amplitudes of basis states that correspond to low energy states of our Ising model increase while amplitudes of basis states that could hardly be considered a solution to our problem decrease. At $t = T$, certain basis states are therefore more likely to be measured and the tables in the bottom row of Fig. 2 rank the 5 most likely ones. Looking at these tables, the most likely final states for the qubit systems to be found in are $|010010100001\rangle$, $|010101010001\rangle$, and $|110100110001\rangle$ for $k \in \{4, 5, 6\}$, respectively. Understood as indicator vectors, these qubit configurations represent the sets {February, May, July, December}, {February, April, June, August, December}, and {January, February, April, July, August, December}.

Our experiments thus suggest that AQC can solve max-sum dispersion. Classical algorithms typically resort to greedy heuristics. AQC, on the other hand, (implicitly) performs exhaustive searches over all 2^n subsets of the n given data points. Our choice of $T \in O\left(\sqrt{2^n}\right)$ suggests that it can accomplish this quadratically faster than classically possible. Finally, in contrast to classical algorithms, the runtime of the AQC approach does *not* depend on the size k of the subset to be selected.



$\|\boldsymbol{\psi}_i\rangle$	$\|a_i\|^2$
$\|010010100001\rangle$	0.002463
$\|010100100001\rangle$	0.002404
$\|010010010001\rangle$	0.002388
$\|010011000001\rangle$	0.002357
$\|010100010001\rangle$	0.002324

$\|\boldsymbol{\psi}_i\rangle$	$\|a_i\|^2$
$\|010101010001\rangle$	0.002161
$\|010101100001\rangle$	0.002159
$\|010100110001\rangle$	0.002092
$\|010100101001\rangle$	0.001935
$\|010100011001\rangle$	0.001850

$\|\boldsymbol{\psi}_i\rangle$	$\|a_i\|^2$
$\|110100110001\rangle$	0.001763
$\|110101100001\rangle$	0.001719
$\|110101010001\rangle$	0.001711
$\|010100110011\rangle$	0.001617
$\|011100110001\rangle$	0.001593

Fig. 2: AQC for max-sum dispersion (see the explanation in the text).

References

1. Abassi, Z., Mirrokni, V.S., Thakur, M.: Diversity Maximization under Matroid Constraints. In: KDD (2013)
2. Aïmeur, E., Brassard, G., Gambs, S.: Quantum Clustering Algorithms. In: ICML (2007)
3. Amin, M.H.S.: Effect of Local Minima on Adiabatic Quantum Optimization. Physical Review Letters 100(13), 130503 (2008)
4. Bauckhage, C., Brito, E., Cvejoski, K., Ojeda, C., Sifa, R., Wrobel, S.: Ising Models for Binary Clustering via Adiabatic Quantum Computing. In: EMMCVPR (2017)
5. Bauckhage, C., Ojeda, C., Sifa, R., Wrobel, S.: Adiabatic Quantum Computing for Kernel k=2 Means Clustering. In: KDML-LWDA (2018)
6. Farhi, E., Goldstone, J., Gutmann, S., Sipser, M.: Quantum Computation by Adiabatic Evolution. arXiv:quant-ph/0001106 (2000)
7. Gibney, E.: Quantum Computer Gets Design Upgrade. Nature 541(7638) (2017)
8. Gollapudi, S., Sharma, A.: An Axiomatic Approach for Result Diversification. In: WWW (2009)
9. Johansson, J., Nation, P., Nori, F.: QuTiP 2: A Python Framework for the Dynamics of Open Quantum Systems. Computer Physics Communications 184(4) (2013)
10. Santos Rodrygo, L.T., Macdonald, C., Ounis, I.: Intent-aware Result Diversification. In: SIGIR (2011)
11. Sifa, R., Bauckhage, C.: Online k-Maxoids Clustering. In: DSAA (2017)
12. Wiebe, N., Kapoor, A., Svore, K.M.: Quantum Algorithms for Nearest-Neighbor Methods for Supervised and Unsupervised Learning. Quantum Information & Computation 15(3–4) (2015)