2023 Asia and Pacific Mathematical Contest in Modeling

Wuyue Cup Quantum Computing Challenge

The optimization of the layout of Computing Power Network (CPN)

1. Introduction

The Computing Power Network (CPN) is a novel information infrastructure that allocates and schedules computing resources based on business needs, which are typically comprised of end users, edge servers and cloud servers. The network aims to meet the demands of various computational tasks. Strategically distributing computing resources based on the spatial distribution of computational needs contributes to reducing latency, minimizing costs, and enhancing overall network efficiency and user experience.

Utilizing operations research and optimization techniques for the location selection and layout of information infrastructure facilitates a more scientifically grounded decision-making process from a global optimization perspective. This approach serves to improve decision-making efficiency and planning effectiveness. As the scale of computational challenges grows, the computational complexity of accurately solving combinatorial optimization problems increases exponentially. Existing solving techniques often struggle to complete these solutions within a reasonable timeframe. Moreover, with traditional computing power nearing the limits defined by Moore's Law, quantum computing emerges as a promising technology, breaking through the computational bottlenecks of conventional computers. Coherent Ising machines (CIM), as specialized quantum computing devices that introduce a new perspective for efficiently solving operational optimization problems in various industries, have garnered extensive attention due to their exceptional performance in solving combinatorial optimization problems.

Due to its close connection to Ising models, QUBO (Quadratic Unconstrained Binary Optimization) model constitutes a central problem class for quantum computation, where it can be solved by the use of Coherent Ising Machines (CIM). QUBO model is formulated as $min\, x^T Q x,\ \ x \in \{0,1\}^n$, where Q represents a $n \times n$ matrix. In the context of this competition, which primarily focuses on optimizing the layout of computational network infrastructure, the problem is modeled using the QUBO form, and the solution is implemented using the Kaiwu SDK. Kaiwu SDK is a software development kit designed for solving QUBO models using

Coherent Ising machines (CIM). You can access the Kaiwu SDK by visiting this link[1] and clicking on the '资料下载' option to obtain the necessary resources.

2. Problem Description

This competition addresses the optimization of the layout of computational network infrastructure within a specific region. The region is divided into several adjacent square grids, and the computational demand distribution data provides the aggregated computational demands within each grid. The coordinates in the data represent the central coordinates of each grid. To simplify the problem, the computational demand points within each grid are consolidated to the central point of the grid (i.e., treating each grid as corresponding to a single demand point).

The computational demand within the grid is generated by users' devices, which are terminals connected to the network, such as sensors, smartphones, industrial robots, and the like. The computational needs within the computing network are met by edge servers and cloud servers. Edge servers are positioned at the 'edge' of the network, typically in proximity to end users or devices. Their role is to process data closer to the users, thereby enhancing response speed and efficiency. Edge servers can handle requests more swiftly as they are closer to the users, alleviating the burden on the core cloud infrastructure and improving overall operational efficiency. Cloud servers, on the other hand, are located in data centers distant from users, boasting robust computing and storage capabilities. When the capacity of edge servers is insufficient, cloud servers can serve as a supplementary resource. The collaborative interaction between edge servers and cloud servers optimizes the performance and reliability of the entire system.

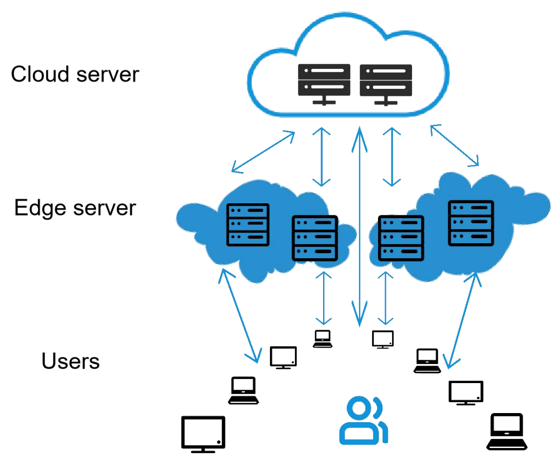

Figure 1: Illustration of Computational Power Network

2.1 Problem 1

The task is to deploy two edge servers within the grid area based on the computational demand distribution. Each edge server has a coverage radius of 1. As an example, in Figure 2, we present an illustration of the coverage effect of edge server with a coverage radius of 2. The goal is to determine the placement of the two edge servers that can cover the maximum computational demand. Assuming the positions of the edge servers are located at the centers of the grids, the computational demands within each grid are provided in the attachment (Attachment 1_Computational Demand Distribution Data.csv).

Please formulate a QUBO model for the problem, and solve it using both the simulated annealing solver and the CIM simulator of the Kaiwu SDK. Provide the coordinates for deploying the edge servers that cover the maximum computational demand, along with the corresponding total computational demand covered.

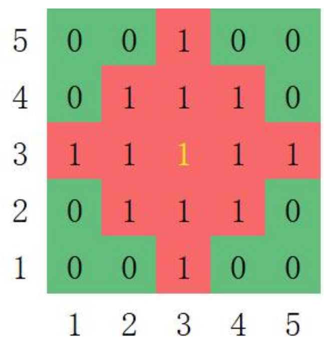

Figure 2: Schematic Diagram of Coverage Effect by an Edge Server with a Radius of 2 at Coordinates [3,3].(Regarding Coverage Determination: If the central distance(Euclidean distance) of a grid is less than or equal to the coverage radius of an edge server, it is considered covered by that edge server.)

2.2 Problem 2

When edge servers cannot meet the computational demands, computational services will be provided by upstream cloud servers. Edge servers and end users can choose to connect to the cloud servers. When the computational demands received by edge servers exceed their capacity limits, the excess demands of the edge servers will be directly allocated to cloud servers. **Each end user's demand must be fulfilled and can only connect to one server, which can be either a cloud or an edge server.** Computing servers have a resource capacity limit. Suppose each edge server has an available computational resource capacity of 12, while the cloud server has an infinite computational resource capacity (ignoring any resource limitations for cloud servers). servers also have a coverage radius; assume the coverage radius for edge servers is 3, while the coverage radius for cloud servers is infinite.

The deployment of edge servers typically incurs costs. This cost comprises fixed costs, computational costs, and transmission costs. The fixed cost is solely dependent on whether an edge server is deployed at the candidate location. Computational costs are directly proportional to the requested computational resource quantity, calculated as the unit computational cost multiplied by the computational load. The unit computational cost is 1 for cloud server and 2 for edge servers. Additionally, transmission costs exist for transfers between user end to edge, edge to cloud, and user end to cloud, calculated as the computational demand multiplied by the transmission distance multiplied by the unit transmission cost. The calculation of the Euclidean distance for transmission distance retains two decimal places, computed as a one-way distance (ignoring round-trip transmission). The unit transmission cost is 1 for both user end to edge and edge to cloud, while it is 2 for user end to cloud.

Please provide the computational network layout with the minimum overall cost that satisfies all user end computational demands within the region. This includes the positions and quantities of edge servers, as well as the connectivity between user end to edge, edge to cloud, and user end to cloud server. Formulate a QUBO model with as few qubits as possible (The SDK only supports problem instances with up to 100 bits), and solve it using the simulated annealing solver and the CIM simulator of the Kaiwu SDK. The computational demand for each user, corresponding to its geographical location, is provided in the user data file (Attachment 2_Computational Demand Distribution Data.csv). The Edge data file (Attachment 3_Candidate Edge Facilities Data.csv) offers candidate coordinates for deploying edge servers, along with fixed costs associated with each location. The Cloud data file (Attachment 4_Cloud Facilities Data.csv) provides the coordinates of the cloud server.

2.3 Problem 3

Please propose a potential decision optimization application scenario that can construct a suitable QUBO model. Recommended scenario domains include, but are not limited to, artificial intelligence, big data, cloud computing, and edge computing. The scenario should have practical value, scalability, real business needs, and demonstrate the advantages of Coherent Ising machines (CIM). Provide necessary background information, research methods, approaches, anticipated research results, a technical roadmap, and supporting references or materials. The report should be between 500 to 1000 words. We have provided a template for this problem (Attachment 5_Template of Problem 3.pptx). You can answer this problem in either Chinese or English.

3. Submission Instructions

The submission should include a paper, QUBO models, executable code, computational results, CIM simulator parameters, SA parameters, solving time, scenario description document, technical roadmap, relevant references, etc. The necessary content and related formats for each problem are outlined as follows:

3.1 Problem 1

QUBO model, executable code, computational results, CIM simulator parameters, SA parameters, solving time. Please provide the computational results in tables as per the following format (the data in the table are examples, you should fill in your own results according to the specified format):

Table 1-1 Demand Coverage Table（Example）

Total Computational Demand Covered
14

Table 1-2 Edge Activation Status Table（Example）

Number	**Edge Coordinates**	**Covered Grid Coordinates**	**Covered Computational Demand**
1	[1,2]	[1,2],[2,1]	6
2	[2,3]	[1,3],[3,3],[2,2],[2,4]	8

3.2 Problem 2

QUBO model, executable code, computational results, CIM simulator parameters, SA parameters, solving time. Please provide the computational results in tables as per the following format (the data in the table are examples, you should fill in your own results according to the specified format):

Table 2-1 Total Cost Table（Example）

<table>
<tr><th colspan="5">Total Cost</th></tr>
<tr><td colspan="5">100</td></tr>
<tr><th>Fixed Cost</th><th>Computational Cost</th><th colspan="3">Transmission Cost</th></tr>
<tr><td>30</td><td>30</td><td colspan="3">40</td></tr>
<tr><td></td><td></td><th>User-Edge</th><th>Edge-Cloud</th><th>User-Cloud</th></tr>
<tr><td></td><td></td><td>10</td><td>10</td><td>20</td></tr>
</table>

Table 2-2 Edge Activation Status Table(Example)

Number	Edge Coordinates	Activation Status
1	[1,2]	1
2	[2,3]	1
3	[3,4]	1
4	[4,5]	0
5	[5,6]	0

Note: The 'Activation Status' field in the table: 1 represents the establishment of an edge server at that location, and 0 represents non-activation.

Table 2-3 User-to-Edge Transmission of Computational Demand Table(Example)

Number	Edge Coordinates	User Coordinates	Amount of Computational demand for transmission
1	[1,2]	[2,2]	13
2	[2,3]	[3,3]	20
3	[3,4]	[4,4]	20
4	[4,5]	[5,5]	0
5	[5,6]	[6,6]	0

Table 2-4 User-to-Cloud Transmission of Computational Demand Table(Example)

Number	Cloud Coordinates	User Coordinates	Amount of Computational demand for transmission
1	[1,0]	[2,2]	0
2	[1,0]	[3,3]	0
3	[1,0]	[4,4]	0
4	[1,0]	[5,5]	12
5	[1,0]	[6,6]	12

Table 2-5 Edge-to-Cloud Transmission of Computational Demand Table(Example)

Number	Edge Coordinates	Cloud Coordinates	Amount of Computational demand for transmission
1	[1,2]	[1,0]	1
2	[2,3]	[1,0]	8
3	[3,4]	[1,0]	8

3.3 Problem 3

Please submit a document according to the given template, including a scenario description, technical roadmap, and relevant references. The length of the document should be between 500-1000 words.

Attachments:

[1]Link:https://ecloud.10086.cn/api/query/developer/user/home.html#match@matchDetails?id=c0466578cfeb4369905c9657614f71f5&utm_source=WUYUEBEI&_channel_track_key=dpPYo2aG