T5.4. Informe de Evaluación de las estrategias en División/Distribución/Paralelización de Modelos

Ibermática – Quantumi3b

[15/07/2023]

XUNTA DE GALICIA
Unión Europea
Unha maneira de facer Europa.
Fondos Europeos
CESGA
GALICIA SUPERCOMPUTING CENTER

Despregamento dunha infraestructura baseada en tecnoloxías cuánticas da información que permita impulsar a I+D+i en Galicia.

Apoiar a transición cara a una economía dixital.

Operación financiada pola Unión Europea, a través do FONDO EUROPEO DE DESENVOLVEMENTO REXIONAL (FEDER) como parte da resposta da Unión á pandemia da COVID-19

DATA	AUTOR	CAMBIOS	VERSIÓN
18/09/2023	Ibermática	Inicial	1

Tabla de contenidos

Objetivo y alcance del proyecto 7

Introducción 7

Caso QAOA 7

Caso QUBO 9

Caso PUBO 9

Caso JSP 9

Lista de figuras

No se encuentran elementos de tabla de ilustraciones.

Lista de Tablas

No se encuentran elementos de tabla de ilustraciones.

Lista de acrónimos

TDC *Táboa de Contidos*

Objetivo y alcance del proyecto

Introducción

La computación cuántica tiene un gran potencial en la resolución de problemas, tanto en investigación básica como en casos industriales. Sin embargo, debido a la situación actual del hardware cuántico, Noisy intermediate-scale quantum (NISQ) era, los errores debidos a la aplicación de operaciones cuánticas, además del error de lectura y el tiempo de coherencia, son un factor limitante para la aplicación de algoritmos cuánticos. Otra gran limitación es la disponibilidad de un conjunto suficientemente grande de qubits de calidad para realizar dichos algoritmos.

Una propuesta para resolver el problema de la escala de qubits es utilizar una técnica de división de circuitos, mediante técnicas de máximo likelihood, para ejecutar subcircuitos en diferentes dispositivos cuánticos con menor profundidad y número de qubits. Estas técnicas consisten en 'cortar' las puertas CNOT que unen distintas zonas del circuito, ejecutando 2 circuitos diferentes. El número de circuitos a ejecutar aumenta exponencialmente con el número de cortes a realizar. Estas técnicas permiten ejecutar algoritmos cuánticos mucho más complejos en ordenadores cuánticos más pequeños y con más ruido.

A continuación, se detallan distintas aproximaciones para los distintos sistemas propuestos:

Caso QAOA

En relación con la divisibilidad del circuito QAOA, hay que considerar que los términos no diagonales de la matriz del QUBO, las relaciones entre variables añaden términos ZZ al hamiltoniano de Ising del problema. Estos términos *ZZ* se convierten en *RZZ*, lo cual implica una conexión entre dos zonas del circuito por dos puertas *CNOT*. Por lo tanto, por cada término no local habrá que cortar el circuito QAOA 2 veces por cada capa que pongamos. Además, no podemos evitar estos términos *ZZ*, debido a que sin ellos no habría relaciones entre variables y el problema sería trivial, solo hay que activar los qubits cuyo coste asociado sea negativo.

Se han sometido a análisis las siguientes consideraciones:

- Por un lado, si queremos estudiar el caso donde no hay términos *RZZ* entre diferentes bloques del circuito, tenemos que ser conscientes de que eso implica un problema donde la matriz del QUBO está formada por bloques. Esos problemas pueden resolverse separando cada uno de los bloques. De este modo, no hay que dividir el circuito. Se ejecutan dos circuitos directamente.

- Por otro lado, si consideramos una matriz QUBO tridiagonal, que es el caso sparse útil más simple, nos encontramos con que para un problema 3x3 ya tenemos 2 términos no diagonales, 4 CNOTS. Para una matriz NxN tridiagonal tendremos *N-1* términos no diagonales, lo cual implica *2N-2* puertas CNOT como mínimo (sin contar las SWAPS).
- Por último, para un problema combinatorio general, debido a que el QAOA tendrá su hamiltoniano del problema compuesto por operadores *X*, *Y*, *Z* y combinaciones de ellos y que los términos de interacción son los términos que hacen que el problema sea complicado clásicamente (con solamente términos locales se puede resolver el problema de forma cuadrática), siempre necesitaremos puertas *RXX*, *RYY*, *RZZ* y las que combinan *X*, *Y*, *Z*. Estas puertas se implementan usualmente con al menos una puerta CNOT, por lo que todo problema con términos de interacción tendrá al menos una puerta CNOT por dimensión y capa.

Superado el problema 2x2, se hace inviable dividir el circuito por el coste asociado a cada división, el cual será del nivel de comprobar a fuerza bruta todas las combinaciones.

En definitiva, para un QAOA general, la división de circuitos es inviable, puesto que es necesario aumentar exponencialmente el número de circuitos a ejecutar. En el algoritmo QAOA una parte clave es el uso del hamiltoniano de problema en forma exponencial $U_j = e^{i\gamma_j H}$, siendo γ_j el parámetro de ajuste en la capa *j* y H el hamiltoniano del problema a resolver. Esta exponencial en general, necesitará una Trotterización, lo cual implica una conexión extensa entre todos los qubits. Por ello, dependiendo de la Trotterización utilizada, cada capa estará compuesta de un gran número de operadores con un gran número de conexiones CNOT. Por ello mismo, es inviable para una buena aproximación realizar el cortado de circuitos, ya que tendremos que cortar un número excesivo de CNOTs, lo cual implicaría ejecutar una cantidad irrealizable de circuitos.

Caso QUBO

En el caso QUBO, la exponencial del hamiltoniano del problema estará compuesta de puertas identidad, *RZ* y *RZZ*. Esto implica que solo habrá que cortar las puertas *RZZ*, que se corresponden con los términos no diagonales del QUBO. Para cualquier problema que queramos resolver y no sea directamente divisible en dos subproblemas, habrá que tener al menos un término no diagonal en cada columna. Esto implica que habrá al menos una puerta *RZZ* conectando cada par adyacente de qubits. Debido a que cada puerta *RZZ* está compuesta de 2 CNOTS, necesitamos hacer 2 cortes del circuito por cada capa del QAOA. Esto es, el número mínimo de cortes a realizar entre 2 partes de un circuito QAOA para un problema QUBO es 2 por capa.

Teniendo en cuenta que un problema QUBO tridiagonal puede ser resuelto en un tiempo $O(N^2D^3)$ y que los problemas QUBO generales aplicados a casos reales no suelen ser densos por bloques con un único término no diagonal entre ellos, esto es,

$$\begin{pmatrix} a & a & a & & & \\ a & a & a & & & \\ a & a & a & c & & \\ & & c & b & b & b \\ & & & b & b & b \\ & & & b & b & b \end{pmatrix}$$

No vemos viable el cortado de circuitos para el QAOA en problemas QUBO.

Caso PUBO

En el caso PUBO tendremos que la exponencial del problema estará compuesta por puertas *I*, *RZ*, *RZZ*, *RZZZ*, *RZZZZ*, etc. Cada una de estas conecta varios qubits, por lo que nuestro problema de cortado es el mismo o peor que el del QUBO, ya que tendremos términos al menos tan conectados como los del QUBO. Esto es, es menos factible que el caso QUBO.

Caso JSP

La matriz de pesos del problema QUBO del Job Shop Scheduling Problem es una matriz no tridiagonal, con un número amplio de términos no diagonales. Esto implica que este problema es ampliamente conexo, por lo que su cortabilidad no es factible.