People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University of Setif 1 – Ferhat Abbas
Faculty of Sciences
Department of Computer Science

Master's Thesis

Presented By :

Zaidi Celia Ourida

Program: Computer Science
Specialization: Quantum Computing

Title :

Unsupervised Machine Learning with Adiabatic Quantum Computing

Submitted to the Examination Committee :

Pr. Achiri Leila	University of SETIF -1- FERHAT ABBAS	President
Dr. Rouabah Mohamed Taha	University of CONSTANTINE -1-	Reviewer
Dr. Talbi Hicham	University of CONSTANTINE -2-	Examiner

Academic Year : 2024-2025

Contents

General Introduction **8**

1 Chapter 1: Classical Clustering **10**
1.1 Introduction ... 10
1.2 Clustering in Machine Learning ... 10
1.3 Classical Clustering Methods ... 12
1.3.1 Partitioning Methods ... 14
1.3.2 Hierarchical Methods ... 17
1.3.3 Graph-based Methods ... 19
1.3.4 Density-based Methods ... 22
1.3.5 Model-based Methods ... 23
1.4 Limitations of Classical Clustering Approaches ... 24
1.5 Conclusion ... 27

2 Chapter 2: Clustering with Adiabatic Quantum Computing **28**
2.1 Introduction ... 28
2.2 Fundamentals of Adiabatic Quantum Computing ... 29
2.2.1 Adiabatic Evolution ... 29
2.2.2 Quantum Annealing ... 29
2.2.3 Current Hardware Implementations ... 33
2.2.4 Comparison with Gate-based Quantum Computing ... 35
2.3 Formulating Clustering Problems for Quantum Annealing ... 36
2.3.1 Binary Quadratic Models: QUBO and Ising Model ... 36
2.4 QUBO Formulations for Clustering Algorithms ... 40
2.4.1 QUBO K-Means Clustering ... 40
2.4.2 QUBO MST-based Clustering ... 41
2.5 Considerations for Quantum Clustering Algorithm Selection ... 43
2.5.1 Hierarchical Clustering and QUBO Incompatibility ... 43
2.5.2 Density-Based Clustering Challenges ... 44
2.6 Practical Limitations of Quantum Annealing-Based Clustering ... 45

3 Chapter 3: Implementation of QUBO-based Clustering Algorithms **47**
3.1 Introduction ... 47
3.2 Development Environment ... 47
3.2.1 Programming Language and Libraries ... 47

3.2.2 Quantum Annealing Platform 48
3.3 Dataset Description 48
3.4 Implementation Details 49
3.4.1 QUBO for K-Means Clustering 49
3.4.2 QUBO for MST-based Clustering 51
3.5 Parameter Tuning 53
3.6 Classical Baseline Implementations 54
3.6.1 Classical K-Means Clustering 54
3.6.2 Classical MST-based Clustering 54

4 Chapter 4: Results and Discussion 56
4.1 Introduction 56
4.2 Evaluation Metrics 56
4.3 Parameter Optimization 56
4.4 Results and Analysis for QUBO K-Means 57
4.4.1 Results of Clustering Performance 57
4.4.2 Discussion and Comparative Analysis with Classical K-Means 57
4.5 Results and Analysis for QUBO MST-Based Clustering 60
4.5.1 Results of Clustering Performance 60
4.5.2 Discussion and Comparative Analysis with Classical MST Clustering 60
4.6 Challenges Encountered 64

Conclusion and Future Directions 66

Résumé

Ces dernières années, l'informatique quantique est apparue comme un paradigme prometteur pour résoudre des problèmes d'optimisation combinatoire inaccessibles aux algorithmes classiques. Ce mémoire étudie l'application de méthodes inspirées du quantique à des tâches de partitionnement de données, en s'appuyant sur les principes de l'informatique quantique adiabatique (AQC). L'étude se concentre sur deux techniques majeures : K-Means et le partitionnement basé sur l'arbre couvrant de poids minimal (Minimum Spanning Tree – MST), toutes deux implémentées dans leurs versions classiques et formulées en QUBO.

Le recuit simulé est utilisé pour imiter le comportement du recuit quantique, en raison d'un accès limité au matériel quantique réel.

Une série d'expériences a été menée sur des jeux de données de référence standards, tels que Iris, Blobs et Moons, afin de comparer les performances des approches classiques et basées sur QUBO à l'aide de métriques d'évaluation. Les résultats révèlent que, bien que les méthodes de partitionnement QUBO puissent atteindre des performances compétitives, elles sont très sensibles aux paramètres de pénalité, et leur efficacité diminue à mesure que la taille du jeu de données augmente. Ces observations mettent en évidence des limites clés liées à la scalabilité et au réglage des paramètres dans les formulations QUBO actuelles.

Malgré ces défis, l'étude démontre le potentiel du partitionnement de données inspiré du quantique et constitue une base pour de futures recherches. Des pistes d'amélioration incluent des approches hybrides quantique-classique, un ajustement adaptatif des pénalités, et des codages QUBO plus efficaces. Ce travail contribue au développement croissant de la recherche à l'interface entre l'informatique quantique et l'apprentissage automatique, et ouvre la voie à des cadres de partitionnement quantique plus robustes.

الملخص

في السنوات الأخيرة، ظهرت الحوسبة الكمومية كمنهج واعد لحل مسائل التحسين التوافقي التي يصعب حلها باستخدام الخوارزميات التقليدية. يتناول هذا البحث تطبيق الأساليب المستوحاة من الحوسبة الكمومية في مهام التحليل العنقودي ضمن إطار التعلم الاستنتاجي، بالاعتماد على مبادئ الحوسبة الكمومية الآدياباتية(AQC).

تركّز الدراسة على تقنيتين بارزتين في التحليل العنقودي: خوارزمية "كاي-مينز (K-Means) "و التحليل العنقودي المعتمد على شجرة الإمتداد الدنيا .(MST)وقد تم تنفيذ كل منهما بصيغتين: تقليدية وأخرى معتمدة على نموذجQUBO.

ونظرًا لمحدودية الوصول إلى العتاد الكمومي الحقيقي، تم استخدام التلدين المحاكى لمحاكاة سلوك التلدين الكمومي.

تم إجراء سلسلة من التجارب على مجموعات بيانات معيارية مثل "Iris" و "Blobs"و"Moons"، وذلك لمقارنة أداء الأساليب التقليدية وتلك المبنية على نموذج QUBO باستخدام مقاييس تقييم مناسبة. أظهرت النتائج أن الأساليب المعتمدة على QUBO يمكن أن تحقق أداءً تنافسياً، لكنها شديدة الحساسية لقيمة وزن العقوبة، كما أن فعاليتها تنخفض مع زيادة حجم البيانات. وتبرز هذه النتائج التحديات المرتبطة بقابلية التوسع وصعوبة ضبط المعاملات في النماذج الحالية.

رغم هذه التحديات، تُظهر الدراسة الإمكانيات الواعدة للتحليل العنقودي المستوحى من الحوسبة الكمومية، وتوفّر أساسًا مهمًا لأبحاث مستقبلية. من بين مسارات التحسين المقترحة: استخدام منهج هجين يجمع بين الأساليب التقليدية والكمومية، وضبط تكيّفي لأوزان العقوبة، وتطوير نماذج QUBO أكثر كفاءة. يسهم هذا العمل في دعم النمو المتزايد للبحث في تقاطع الحوسبة الكمومية وتعلم الآلة، ويمهد الطريق نحو اطار تحليل عنقودي كمومي أكثر قوة وموثوقية.

Abstract

In recent years, quantum computing has emerged as a promising paradigm for solving combinatorial optimization problems that are intractable for classical algorithms. This thesis investigates the application of quantum-inspired methods to unsupervised clustering tasks using adiabatic quantum computing (AQC) principles. The research concentrates on two prominent clustering techniques: K-Means and Minimum Spanning Tree (MST)-based clustering, both implemented in their classical and QUBO-based forms. Simulated annealing is used to mimic quantum annealing behavior due to limited access to real quantum hardware.

A series of experiments were conducted on standard benchmark datasets, including Iris, Blobs, and Moons, to compare the performance of classical and QUBO-based approaches using evaluation metrics. The results reveal that while QUBO clustering methods can achieve competitive performance, they are highly sensitive to penalty weight parameters, and their effectiveness diminishes as dataset size increases. These results show important limitations related to scalability and parameter tuning in current QUBO formulations.

Even with these challenges, the study illustrates the capability of quantum-inspired clustering and sets the stage for future research. Opportunities for improvement include hybrid quantum-classical approaches, adaptive penalty tuning, and more efficient QUBO encodings. This work helps to the growing of research at the intersection of quantum computing and machine learning, and lays the groundwork for more robust quantum clustering frameworks.