الجمهورية الجزائرية الديمقراطية الشعبية
The People's Democratic Republic of Algeria
وزارة التعليم العالي والبحث العلمي
The Ministry of Higher Education and Scientific Research
المدرسة العليا للإعلام الآلي - 8 ماي 1945 - سيدي بلعباس
The Higher School of Computer Science -08 May 1945- Sidi Bel Abbes

MASTER THESIS

To obtain the diploma of **Master**
Field: **Computer Science**
Specialty: **Artificial Intelligence And Data Science**

Theme

Multi-label Classification Using Multi-objective Optimization and Metaheuristics: State of the Art

---

Presented By
Adel Abdelkader MOKADEM

Defended on September 24^{th}, 2025 In front of the jury composed of

Dr. Bachir Nadir MAHAMMED	President
Dr. Nassima DIF	Supervisor
Pr. Sidi Mohammed BENSLIMANE	Co-supervisor
Pr. Marie-Eléonore KESSACI	Co-supervisor
Dr. Julie JACQUES	Co-supervisor
Dr. Samiha MEZRAR	Examinator

Academic year : 2024/2025

Abstract

Multi-label classification (MLC) is a growing research area in machine learning that allows each instance to be associated with multiple labels simultaneously. While this paradigm offers a richer representation of complex data, it also introduces major challenges, such as correlations between labels, high dimensionality, severe class imbalance, and the exponential growth of possible label combinations. To address these difficulties, researchers have increasingly framed the problem as combinatorial optimization task, motivating the use of multi-objective optimization and metaheuristic techniques, which are capable of efficiently navigating large search spaces and generating diverse sets of solutions. Current research has shown that metaheuristic approaches, particularly evolutionary algorithms and ant colony optimization, are effective in addressing the computational complexity of MLC and in producing solutions that balance multiple objectives. However, despite these advances, a persistent limitation lies in the lack of explainability, as most optimization-driven methods prioritize predictive performance while providing little interpretability. This shortcoming restricts their adoption in sensitive domains such as healthcare, where transparency is essential. These observations highlight the necessity of developing new approaches that are not only robust and accurate but also interpretable, thereby enhancing their applicability in real-world decision-making. In this context, the objective of this thesis is to present a comprehensive review of state-of-the-art multi-objective optimization approaches and metaheuristics for multi-label classification, to analyze and compare recent contributions, and to highlight current challenges and future research directions.

Keywords : Multi-label Classification, Combinatorial Optimization, Multi-objective Optimization, Metaheuristics, Evolutionary Algorithms, Ant Colony Optimization.

Résumé

La classification multi-label (MLC) est un domaine de recherche en plein essor dans l'apprentissage automatique, qui permet d'associer chaque instance à plusieurs labels simultanément. Bien que ce paradigme offre une représentation plus riche des données complexes, il introduit également des défis majeurs tels que les corrélations entre labels, la haute dimensionnalité, le déséquilibre marqué des classes et la croissance exponentielle du nombre de combinaisons possibles de labels. Pour relever ces difficultés, les chercheurs ont de plus en plus formulé le problème comme une tâche d'optimisation combinatoire, motivant l'utilisation de l'optimisation multi-objectif et des techniques métaheuristiques, capables de parcourir efficacement de grands espaces de recherche et de générer des ensembles de solutions diversifiés. Les recherches actuelles ont montré que les approches métaheuristiques, en particulier les algorithmes évolutionnaires et l'optimisation par colonie de fourmis, sont efficaces pour traiter la complexité computationnelle de la MLC et produire des solutions équilibrant plusieurs objectifs. Toutefois, malgré ces avancées, une limite persistante réside dans le manque d'explicabilité, la plupart des méthodes basées sur l'optimisation privilégiant la performance prédictive tout en offrant peu d'interprétabilité. Cette lacune restreint leur adoption dans des domaines sensibles tels que la santé, où la transparence est essentielle. Ces observations soulignent la nécessité de développer de nouvelles approches à la fois robustes, précises et interprétables, afin de renforcer leur applicabilité dans la prise de décision en contexte réel. Dans ce contexte, l'objectif de ce mémoire est de présenter une revue complète de l'état de l'art des approches d'optimisation multi-objectifs et des métaheuristiques appliquées à la classification multi-label, d'analyser et de comparer les contributions récentes, et de mettre en évidence les défis actuels ainsi que les perspectives de recherche futures.

Mots-clés **:** Classification multi-label, Optimisation combinatoire, Optimisation multi-objectif, Métaheuristiques, Algorithmes évolutionnaires, Optimisation par colonie de fourmis.

ملخص

يُعدّ التصنيف متعدد العلامات مجالًا بحثيًا متناميًا في مجال التعلم الآلي، إذ يسمح بربط كل حالة بعلامات متعددة في آن واحد. وبينما يُقدّم هذا النموذج تمثيلًا غنيًا للبيانات المعقدة، فإنه يطرح أيضًا تحدياتٍ كبيرة، مثل الترابط بين العلامات، وارتفاع أبعادها، واختلال التوازن الحاد بين الفئات، والنموّ الأسّي لمجموعات العلامات المُحتملة. ولمعالجة هذه الصعوبات، عمد الباحثون بشكل متزايد إلى صياغة المشكلة كمهمة تحسين توافقي، مما حفّز استخدام التحسين متعدد الأهداف وتقنيات الارشاد العام، القادرة على التنقل بكفاءة في مساحات بحث واسعة وتوليد مجموعاتٍ متنوعة من الحلول. وقد أظهرت الأبحاث الحالية أن مناهج الارشاد العام، وخاصةً الخوارزميات التطورية و خوارزمية مستعمرة النمل، فعّالة في معالجة التعقيد الحسابي لتصنيف العلامات وفي إنتاج حلول تُوازن بين الأهداف المتعددة. ومع ذلك، وعلى الرغم من هذه التطورات، يكمن أحد القيود الدائمة في عدم القدرة على التفسير، حيث تُعطي مُعظم الطرق القائمة على التحسين الأولوية للأداء التنبئي مع توفير قدرٍ ضئيل من قابلية التفسير. هذا القصور يُقيّد اعتمادها في مجالات حساسة كالرعاية الصحية، حيث تُعدّ الشفافية أمرًا بالغ الأهمية. تسلّط هذه الملاحظات الضوء على ضرورة تطوير مناهج جديدة، ليست متينة ودقيقة فحسب، بل قابلة للتفسير أيضًا، مما يُعزّز قابلية تطبيقها في صنع القرار على أرض الواقع. في هذا السياق، يهدف هذا العمل إلى تقديم استعراض شامل لأحدث تقنيات تحسين متعدد الأهداف و تقنيات الإرشاد العام المستخدمة في تصنيف البيانات متعددة العلامات، بالإضافة إلى تحليل ومقارنة الإسهامات البحثية الحديثة، وتسليط الضوء على التحديات الحالية واتجاهات البحث المستقبلية.

الكلمات المفتاحية: التصنيف متعدد العلامات، التحسين التوافقي، التحسين متعدد الأهداف، الارشاد العام، الخوارزميات التطورية، خوارزمية مستعمرة النمل