الجمهورية الشعبية الديمقراطية الجزائرية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
المدرسة العليا للإعلام الآلي ٠08 ماي 1945٠ بسيدي بلعباس
École Supérieure en Informatique
-08 Mai 1945- Sidi Bel Abbès

THESIS

To obtain the diploma of **Engineer**
Field: **Computer Science**
Specialty: **Intelligence artificielle et Science des Données (IASD)**

Theme

Multimodal Data Fusion using MLP-based Architectures

---

presented by:

- CHERGUI Abdelmadjid

Submission Date: **23/09/2024** In front of the jury composed of:

Ms. BOUSMAHA Rabab	President
Mr. BENSLIMANE Sidi Mohamed	Supervisor
Ms. NAOUM Hanae	Examiner

Academic Year: 2023/2024

Abstract

Multimodal machine learning is an active interdisciplinary field of research that focuses on developing intelligent computer systems capable of understanding, reasoning, and learning by integrating multiple modalities. It also presents unique computational and theoretical challenges due to the heterogeneity of data sources and the often complex interconnections between different modalities. These challenges make the selection of an effective deep learning architecture for multimodal data fusion and learning a non-trivial task.

This Thesis primarily focuses on addressing these challenges. And to propose a novel framework that tackles the challenge of selecting an appropriate architecture for a multimodal learning tasks. Our approach leverages the strengths of Multi-Layer Perceptrons (MLP)-based architectures, which are known for their computational efficiency, simple design, and robustness in handling various data types and tasks.

In our research, we delve into the basic principles of multimodal machine learning, with a particular emphasis on the role of MLP-based architectures. We provide a detailed exploration of our proposed framework, discussing its design, implementation, and the advantages it offers. This work represents a significant step towards more efficient and effective multimodal learning systems.

Keywords : Multimodal Machine Learning, MLP-based Architectures, Multimodal Data Fusion, Deep Learning.

Résumé

L'apprentissage automatique multimodal est un domaine de recherche interdisciplinaire actif qui se concentre sur le développement de systèmes informatiques intelligents capables de comprendre, de raisonner et d'apprendre en intégrant plusieurs modalités. Il présente également des défis calculatoires et théoriques uniques en raison de l'hétérogénéité des sources de données et des interconnexions souvent complexes entre les différentes modalités. Ces défis font de la sélection d'une architecture d'apprentissage profond efficace pour la fusion de données multimodales et l'apprentissage une tâche non triviale.

Cette thèse se concentre principalement sur la résolution de ces défis. Et de proposer un nouveau framework qui s'attaque au défi de la sélection d'une architecture d'apprentissage profond. qui relève le défi de la sélection d'une architecture appropriée pour une tâche d'apprentissage multimodale. Notre approche s'appuie sur les forces des architectures basées sur les perceptrons multicouches (MLP), qui sont connues pour leur efficacité de calcul, leur simplicité de conception, et leur robustesse dans la gestion de divers types de données et de tâches.

Dans notre recherche, nous nous concentrons sur les principes de base de l'apprentissage automatique multimodal, en mettant particulièrement l'accent sur le rôle des architectures basées sur les MLP. Nous explorons en détail le cadre que nous proposons, en discutant de sa conception, de sa mise en œuvre et des avantages qu'il offre. Ce travail représente un pas important vers des systèmes d'apprentissage multimodaux plus efficaces et plus performants.

Keywords : Apprentissage automatique multimodal, architectures basées sur les MLP, fusion de données multimodales, apprentissage profond.

الملخص

التعلم الآلي متعدد الوسائط هو مجال بحثي نشط متعدد التخصصات يركز على تطوير أنظمة حاسوبية ذكية قادرة على الفهم والاستدلال والتعلم من خلال دمج أنماط بيانات متعددة. كما أنه يمثل تحديات حسابية ونظرية فريدة من نوعها بسبب عدم تجانس مصادر البيانات والترابطات المعقدة في كثير من الأحيان فيما بينها. هذه التحديات تجعل اختيار بنية فعالة للتعلم العميق لدمج البيانات متعددة الوسائط والتعلم مهمة صعبة.

تركز هذه الأطروحة في المقام الأول على معالجة هذه التحديات. واقتراح إطار عمل جديد يعالج مشكلة اختيار بنية مناسبة لمهمة التعلم متعدد الوسائط. ويستفيد إطار عملنا من نقاط القوة في البنى القائمة المدركات متعددة الطبقات, والتي تُعرف بكفاءتها الحسابية وتصميمها البسيط وقدرتها على التعامل مع أنواع البيانات والمهام المختلفة.

في بحثنا هذا، نتعمق في المبادئ الأساسية للتعلم الآلي متعدد الوسائط، مع التركيز بشكل خاص على دور البنى القائمة على المدركات متعددة الطبقات. كما نقدم شرحًا تفصيليًا لإطار عملنا المقترح، ونناقش تصميمه وتطبيقاته والمزايا التي يقدمها. يمثل هذا العمل خطوة مهمة نحو أنظمة تعلم متعددة الوسائط أكثر كفاءة وفعالية.

الكلمات المفتاحية: المدركات متعددة الطبقات , لتعلم الآلي متعدد الوسائط, التعلم العميق.