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 **Master**
Field: **Computer Science**
Specialty: **Intelligence artificielle et Science des Données (IASD)**

Theme

Multimodal Data Fusion using MLP-based Architectures: A Comparative Study

---

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

This thesis provides an in-depth exploration of Multimodal Machine Learning (MML), a rapidly evolving field with significant implications for technology and society. The study delves into the architecture, variations, and extensions of Transformer models and MLP Mixers, highlighting their strengths and weaknesses.

This thesis is centered around research, with no practical implementation undertaken at this stage. Multimodal Machine Learning holds paramount significance across various applications, encompassing areas like multisensory wearable devices, Robotics, and more. Deep learning (DL) techniques, particularly Transformer models and MLP Mixers, have exhibited commendable accuracy and adaptability in handling the intricacies of multimodal data.

This thesis has the potential to improve the effectiveness and accessibility of multimodal machine learning, by providing a comprehensive understanding of the latest advancements and challenges in the field.

Keywords : Multimodal Machine Learning, Transformers, MLP Mixers, Deep Learning.

Résumé

Cette thèse propose une exploration approfondie de l'apprentissage automatique multimodal (AMM), un domaine qui évolue rapidement et qui a des implications significatives pour la technologie et la société. L'étude se penche sur l'architecture, les variations et les extensions des modèles transformateurs et mixeurs MLP, en soulignant leurs forces et leurs faiblesses.

Cette thèse est centrée sur la recherche, aucune mise en œuvre pratique n'étant entreprise à ce stade. L'apprentissage automatique multimodal revêt une importance capitale dans diverses applications, des domaines tels que les dispositifs portables multisensoriels, la robotique, et bien d'autres encore. Les techniques de l'apprentissage profond (AP), en particulier les modèles de transformateurs et les mixeurs MLP, ont fait preuve d'une précision et d'une efficacité remarquables, ont fait preuve d'une précision et d'une adaptabilité remarquables dans le traitement des complexités des données multimodales.

Cette thèse a le potentiel d'améliorer l'efficacité et l'accessibilité de l'apprentissage automatique multimodal, en fournissant une compréhension complète des dernières avancées et des défis dans ce domaine.

Mots clés : Apprentissage Automatique Multimodal, Apprentissage Profond, Transformateurs , Mixeurs MLP.

الملخص

تقدم هذه الأطروحة استكشافًا متعمقًا للتعلم الآلي متعدد الوسائط , وهو مجال سريع التطور له آثار كبيرة على التكنولوجيا والمجتمع.

تتعمق الدراسة في بنية نماذج المحولات وتنويعاتها وتوسعاتها اضافة الى خلاطات المدركات متعددة الطبقات، مع تسليط الضوء على نقاط قوتها وضعفها.

تتمحور هذه الأطروحة حول الجانب البحثي، مع عدم وجود تطبيق عملي في هذه المرحلة. يحظى التعلم الآلي متعدد الوسائط بأهمية قصوى عبر تطبيقات مختلفة، تشمل مجالات مثل الأجهزة متعددة الحواس القابلة للارتداء والروبوتات وغيرها. تقنيات التعلّم العميق ، وخاصةً تقنيات المحولات وخلاطات المدركات متعددة الطبقات، فقد أظهرت دقة جديرة بالثناء وقدرة على التكيف في التعامل مع تعقيدات البيانات متعددة الوسائط. تعد هذه الأطروحة بمثابة محاولة لتحسين فعالية وإمكانية الوصول إلى للتعلم الآلي متعدد الوسائط، من خلال توفير فهم شامل لأحدث التطورات والتحديات في هذا المجال.

الكلمات المفتاحية: خلاطات لمدركات متعددة الطبقات,المحولات , لتعلم الآلي متعدد الوسائط, التعلم العميق.