

Offre de stage de M2 recherche : Réduction des bandits contextuels aux modèles d'apprentissage en flux (Stream learning)

Novembre 2025

1 Contexte du stage

La croissance exponentielle des données produites par l'économie numérique et la dépendance au stockage de données posent des problèmes majeurs : coût, sécurisation et mise en conformité, risque d'approvisionnement, impact écologique.

Alors que les GAFAMs dominent toujours plus l'économie numérique, outre notre dépendance technologique à ces acteurs, ils ne devraient pas assurer seuls la prise en compte des enjeux de maîtrise de l'empreinte écologique de l'IA, de la protection des données personnelles et sensibles, des biais algorithmiques, de la démocratisation de son usage dans notre société.

Les entreprises et les collectivités veulent extraire de la connaissance à partir des données à leur disposition pour créer de la valeur et innover, mais d'un point de vue technique cette création de valeur peut être difficile à construire et orchestrer. L'IA décuple à la fois les attentes et les coûts de cette création de valeur.

Une toute nouvelle approche algorithmique, le stream learning ou machine learning en temps-réel [1], promet de grandement simplifier l'intégration des modèles d'IA en offrant aux entreprises et collectivités des modèles jamais obsolètes, avec des coûts de production réduits, tout en prenant naturellement en compte les enjeux de protection des données personnelles et d'impact écologique. Cette innovation de rupture permet d'apprendre en continu, de s'adapter aux changements en temps-réel, directement à partir des flux de données pris à leur source, sans recours au stockage.

Le problème du multi-armed bandit (MAB en abrégé) est un problème d'apprentissage en flux et constitue l'une des premières techniques d'apprentissage automatique à avoir rencontré un succès industriel réel. Les MAB ont trouvé des applications dans les essais cliniques, les systèmes de recommandation, la tarification dynamique et le placement publicitaire [2]. Ils ont gagné en popularité récemment grâce à leur utilisation dans le fine-tuning des grands modèles de langage (LLM en abrégé) [3].

L'objectif de ce stage est de résoudre le problème du bandit contextuel en le réduisant à un problème d'apprentissage en flux.

2 L'objectif

Ce stage constitue le point de départ d'un projet plus vaste portant sur l'utilisation des bandits contextuels pour résoudre des problèmes réels à l'aide d'une réduction générale en apprentissage en flux, incluant des modèles linéaires et non linéaires. Cela implique également l'étude de contextes plus réalistes et complexes de bandits, où différents types de bandits — tels que les feedback-graphs [4], les bandits à espace d'actions continu [5] ou bien encore les bandits multiples agrégés [6] — doivent être combinés.

L'objectif principal du stage est d'étudier la réduction des bandits contextuels à des modèles d'apprentissage en flux (stream learning). La première étape de ce travail consiste à créer un benchmark similaire à celui de [7], mais utilisant des modèles d'apprentissage en flux non linéaires, et à le comparer aux réductions précédentes utilisant des modèles linéaires. En particulier vous vous intéresserez aux tâches suivantes :

- Bibliographie sur l'apprentissage en flux [1], plus particulièrement les MAB [2,8] et sur les réductions des MAB contextuels.
- Étudier du point de vue théorique la réduction des MAB contextuels à des modèles d'apprentissage en flux: proposition d'algorithmes, études de complexité et obtention de bornes théoriques sur le regret...
- Créer un répertoire de jeux de données variés pour les MAB contextuels.
- Produire un benchmark de MAB contextuels utilisant des modèles d'apprentissage en flux sur des jeux de données sélectionnés, en utilisant les métriques pertinentes (regret, temps CPU, mémoire).
- Réaliser une étude comparative des MAB contextuels utilisant l'apprentissage en flux et d'autres réductions disponibles dans la littérature.

Ce que vous allez apprendre

- Les bases des MAB contextuels et de l'apprentissage en flux.
- La conduite d'un travail bibliographique sur un sujet spécifique.
- La collecte de jeux de données pertinents.
- Les outils mathématiques utiles à l'étude du comportement des MAB.

- L'application de tests statistiques pour comparer les performances d'algorithmes.
- L'utilisation des bibliothèques d'apprentissage en flux les plus populaires en *Python* : [River](#), [CapyMoa](#), [VowpalWabbit](#).

3 Déroulement du stage

Le stage est proposé par Ubiquity AI, en relation avec l'Institut de Mathématiques de Bordeaux (IMB). Ubiquity AI est une deeptech bordelaise, essaimage du LIPN-CNRS de l'Université Sorbonne Paris Nord avec plus de 20 ans de recherche fondamentale et 10 ans de R&D travaillant sur le stream learning. Sa mission est de démocratiser l'intelligence artificielle en la rendant accessible grâce au stream learning qui est la forme la plus pure et naturelle d'apprentissage. L'IMB est le laboratoire regroupant l'essentiel de la recherche en mathématiques du site bordelais.

- Lieu : Bordeaux centre-ville dans les locaux d'Ubiquity AI et une journée par semaine au sein de l'IMB sur le campus de Talence
- Encadrants Ubiquity : Subhy Albakour et Erick Alphonse
- Encadrants IMB : Bernard Bercu et Adrien Richou
- Période : Février-Mars à Août-Septembre 2026 (6 mois)
- Niveau : Master 2
- Gratification : 1000€ brut par mois
- Candidature : Envoyez votre CV ainsi que vos relevés de notes de master à l'adresse suivante : subhy.albakour@idaaas.com.
- La suite : Le stage peut être suivi par une thèse CIFRE pour continuer le travail sur le même projet, et étendre à des contextes plus réalistes et complexes de bandits. Pour ce poste, il est attendu un très bon niveau en probabilités-statistique et en algorithmique.

References

- [1] Albert Bifet, Ricard Gavalda, Geoffrey Holmes, and Bernhard Pfahringer, Machine Learning for Data Streams: with Practical Examples in MOA. Cambridge, Massachusetts: The MIT Press, 2018.

REFERENCES

- [2] T. Lattimore and C. Szepesvári, *Bandit Algorithms*, 1st edition. Cambridge New York, NY Melbourne New Delhi Singapore: Cambridge University Press, 2020.
- [3] D. Bouneffouf and R. Feraud, “Survey: Multi-Armed Bandits Meet Large Language Models,” Sept. 30, 2025, arXiv: arXiv: 2505.13355. doi: 10.48550/arXiv.2505.13355.
- [4] S. Mannor and O. Shamir, “From Bandits to Experts: On the Value of Side-Observations,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2011.
- [5] A. Krishnamurthy, J. Langford, A. Slivkins, and C. Zhang, “Contextual Bandits with Continuous Actions: Smoothing, Zooming, and Adapting,” *Journal of Machine Learning Research*, vol. 21, no. 137, pp. 1–45, 2020.
- [6] A. Agarwal, H. Luo, B. Neyshabur, and R.E. Schapire . “Corralling a Band of Bandit Algorithms”. *Annual Conference Computational Learning Theory*. (2016)
- [7] A. Bietti, A. Agarwal, and J. Langford, “A Contextual Bandit Bake-off,” *Journal of Machine Learning Research*, vol. 22, no. 133, pp. 1–49, 2021.
- [8] Sébastien Bubeck and Nicolo Cesa-Bianchi. *Regret analysis of stochastic and nonstochastic multi-armed bandit problems*. Foundations and Trends in Machine Learning, 2012.