



Offre n°2024-07203

PhD Position F/M Data Injection Attacks in Supervised Machine Learning Systems

Type de contrat : Fixed-term contract

Niveau de diplôme exigé : Graduate degree or equivalent

Fonction : PhD Position

A propos du centre ou de la direction fonctionnelle

The Inria Université Côte d'Azur center counts 37 research teams as well as 8 support services. The center's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The majority of the center's research teams are located in Sophia Antipolis and five of them are based in an Inria antenna in Montpellier. The Inria branch in Montpellier is growing in size, in accordance with the strategy described in the institution's Contract of Objectives and Performance (COP).

Contexte et atouts du poste

This PhD position is funded by the French Ministry of Defense via the "agence de l'innovation de défense (AID)", which gives the final word on the acceptance of the candidate. This position is exclusive for holders of a European, UK, or Swiss nationality.

The PhD candidate is hosted by INRIA at Sophia Antipolis. The PhD degree is granted by the Université Côte d'Azur (UniCA) and it develops within a close collaboration between INRIA, Princeton University, and the University of Sheffield. The position is jointly supervised by Samir M. Perlaza (Inria) and Iñaki Esnaola (University of Sheffield, UK). Research stays in the University of Sheffield and Princeton University might be envisioned.

Mission confiée

Recently, we have introduced the notion of worst-case data-generating (WCDG) probability measure [1, 2], which has been a key instrument to the study of generalization capabilities of machine learning algorithms [3]. We have come to the conclusion that this work has set a fruitful mathematical theory that has already led to important results: (i) An analytical characterization of the generalization error of machine learning algorithms; and (ii) The identification of the Gibbs algorithm as an instrument for the characterization of the generalization capabilities of any machine learning algorithm. The advantages of pairing any algorithm with a particular Gibbs algorithm is that, the latter is well understood and known to have mathematical properties that ease the analysis of generalization [4, 5, 6, 7, 8, 9].

The WCDG probability measure also models data-injection attacks to machine learning systems that are the most difficult to detect. Essentially, the WCDG probability measure describes the probability distribution of datasets after a malicious modification aiming at tampering with the model selection. Such a malicious intervention on the datasets is said to be difficult to detect because the WCDG probability measure is sufficiently close to the original distributions of the datasets. Interestingly, how close the WCDG probability measure is to the original measure is quantified via relative entropy (or Kullback-Leibler divergence) via a parameter, which remains part of the design.

REFERENCES

- [1] X. Zou, S. M. Perlaza, I. Esnaola, and E. Altman, "Generalization analysis of machine learning algorithms via the worst-case data-generating probability measure," in Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, Feb. 2024.
- [2] —, "The worst-case data-generating probability measure," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9515, Aug. 2023.
- [3] X. Zou, S. M. Perlaza, I. Esnaola, E. Altman, and H. V. Poor, "An exact characterization of the generalization error of machine learning algorithms," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9539, Jan. 2024.
- [4] S. M. Perlaza, G. Bisson, I. Esnaola, A. Jean-Marie, and S. Rini, "Empirical risk minimization with relative entropy regularization: Optimality and sensitivity," in Proceedings of the IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, Jul. 2022, pp. 684–689.
- [5] F. Daunas, I. Esnaola, S. M. Perlaza, and H. V. Poor, "Analysis of the relative entropy asymmetry in the

regularization of empirical risk minimization," in Proceedings of the IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, Jun. 2023.

[6] S. M. Perlaza, I. Esnaola, G. Bisson, and H. V. Poor, "On the validation of Gibbs algorithms: Training datasets, test datasets and their aggregation," in Proceedings of the IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, Jun. 2023.

[7] S. M. Perlaza, G. Bisson, I. Esnaola, A. Jean-Marie, and S. Rini, "Empirical risk minimization with generalized relative entropy regularization," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9454, Feb. 2022.

[8] F. Daunas, I. Esnaola, S. M. Perlaza, and H. V. Poor, "Empirical risk minimization with relative entropy regularization type-II," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9508, May. 2023.

[9] —, "Empirical risk minimization with f-divergence regularization in statistical learning," INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, Tech. Rep. RR-9521, Oct. 2023.

Principales activités

The objectives of this thesis are the following.

- To characterize the fundamental trade-off between generalization error and detection probability that governs data-injection attacks onto supervised machine learning systems;
- To identify algorithm design guidelines that increase the robustness of machine learning algorithms to data-injection attacks, e.g., conditions on the minimum sample size, assumptions on the sets of labeled patterns, etc.; and
- To construct prototypes of algorithms over which data-injection attacks can be implemented in a controlled manner such that the above fundamental limits can be studied in specific practical cases.

Avantages

- Subsidized meals
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Social, cultural and sports events and activities
- Access to vocational training
- Contribution to mutual insurance (subject to condition)

Rémunération

Gross Salary per month: 2010€ brut per month (year 1 & 2) and 2190€ brut per month (year 3)

Informations générales

- **Thème/Domaine** : Optimization, machine learning and statistical methods
Information system (BAP E)
- **Ville** : Sophia Antipolis
- **Centre Inria** : [Centre Inria d'Université Côte d'Azur](#)
- **Date de prise de fonction souhaitée** : 2024-09-01
- **Durée de contrat** : 3 years
- **Date limite pour postuler** : 2024-04-28

Contacts

- **Équipe Inria** : [NEO](#)
- **Directeur de thèse** :
Medina Perlaza Samir / samir.perlaza@inria.fr

A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 215 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3900 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 200 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

L'essentiel pour réussir

Candidates are expected to have a strong background in mathematics. Previous knowledge on information theory, and game theory is desirable. Abilities in algorithm design and computer programming are also essential. The candidate must have a provable level of written and spoken english. Skills in french language are not required.

Attention: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

Consignes pour postuler

Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.