2019-01575 - Doctorant F/H PhD: Machine Learning meets Data Mining: Towards sparse and interpretable deep neural networks

Type de contrat : CDD de la fonction publique
Niveau de diplôme exigé : Bac + 5 ou équivalent
Autre diplôme apprécié : Master’s or Diploma degree in Computer Science, strong experience machine learning and data mining.
Fonction : Doctorant

Contexte et atouts du poste

Machine Learning meets Data Mining: Towards sparse and interpretable deep neural networks

Context: This PhD thesis is part of the research project “Academics” within the IDEX Lyon Scientific Breakthrough program. The project aims to bring together data, social and climate scientists to better understand and predict complex dynamical systems. In that context, the PhD student will particu- larly be interested in the improvement of the Deep Neural Networks (DNN) by working on their simplification and the increase of their interpretability.

Predictive models based on DNN have become very powerful and are now widely used in a large variety of applications. However, the underlying deep learning methods tend to produce models that are very large as well as difficult to interpret and understand, which raises two issues. Firstly, the size of these models may jeopardize their scalability and, therefore, has to be reduced without altering their performance. Secondly, the use of such methods in a data science context implies that users are scientists (social and climate scientists in this project) who aim to acquire new knowledge in their own field. Therefore, the models must be not only effective but also understandable. Descriptive data mining techniques can help achieve this goal.

Mission confiée

Subject: The aim of this PhD thesis is to tackle these two ambitious challenges, at the intersection of machine learning and data mining. The young researcher will work on the development of sparse neural networks based on new data mining approaches to analyze, understand and compress the produced models.

− A first task will consist in developing novel approaches for the simplification and compression of DNNs. Compressing models is very important as they can be very large (up to several Gigabytes) and require a great amount of computation that cannot be parallelized (e.g. very deep architectures). However, there is a high redundancy in the computations and model parameters. It has been shown [7, 2] that compression can save 10 to 100 times memory, while keeping almost the same prediction ability. Other approaches drastically reduce the precision of weight parameters [5], factorize the weight matrices [3], or perform optimizations on the network structure [1, 8]. Our previous research work [10, 4] showed the potential of such optimization techniques for DNNs. Most importantly, compressing the network by removing spurious information can help in its understanding and interpretation [9], which previous work has mostly tried to tackle by specific visualization techniques [12, 13, 11] or by explicitly learning automatically extracted concepts [6]. Pattern mining algo- rithms can also be used for analyzing neural activations, by identifying blocks in the weight matrices (or tensors) which represent noise and thus do not con- tribute to the target and final output activations, and by identifying paths of neuron activations strongly correlated with an output. We will consider neural network architectures that facilitate this mining process, e.g. by preferring partially-connected structures and sub-modules and by avoiding fully-connected parts as much as possible as they “diffuse” the extracted information across the whole neural network.

− A better understanding of DNNs requires also to work on both the input and the output of the models in several ways:
  • Integrating the priors to the models: data mining approaches can be used to characterize non-explicit priors;
  • Using the results from data mining or other unsupervised learning tech- niques as priors;

Characterizing prediction errors and learning specific models for these “extreme” cases, e.g. geographically localized errors, or to bias training samples to better handle these errors.

− A third task will be to investigate formalisms (i.e. languages) that make DNNs interpretable or partially interpretable, and the definition of algorithms that make possible the discovery or learning of DNN descriptions and parameterizations with regard to a related language.

References


Principales activités

Subject: The aim of this PhD thesis is to tackle these two ambitious challenges, at the intersection of machine learning and data mining. The young researcher will work on the development of sparse neural networks based on new data mining approaches to analyze, understand and compress the produced models.

A first task will consist in developing novel approaches for the simplification and compression of DNNs. Compressing models is very important as they can be very large (up to several Gigabytes) and require a great amount of computation that cannot be parallelized (e.g. very deep architectures). However, there is a high redundancy in the computations and model parameters. It has been shown [7, 2] that compression can save 10 to 100 times memory, while keeping almost the same prediction ability. Other approaches drastically reduce the precision of weight parameters [5], factorize the weight matrices [3], or perform optimizations on the network structure [1, 8]. Our previous research work [10, 4] showed the potential of such optimization techniques for DNNs. Most importantly, compressing the network by removing spurious information can help in its understanding and interpretation [9], which previous work has mostly tried to tackle by specific visualization techniques [12, 13, 11] or by explicitly learning automatically extracted concepts [6]. Pattern mining algorithms can also be used for analyzing neural activations, by identifying blocks in the weight matrices (or tensors) which represent noise and thus do not contribute to the target and final output activations, and by identifying paths of neuron activations strongly correlated with an output. We will consider neural network architectures that facilitate this mining process, e.g. by preferring partially-connected structures and sub-modules and by avoiding fully-connected parts as much as possible as they “diffuse” the extracted information across the whole neural network.

A better understanding of DNNs requires also to work on both the input and the output of the models in several ways:

- Integrating the priors to the models: data mining approaches can be used to characterize non-explicit priors;
- Using the results from data mining or other unsupervised learning techniques as priors;
- Characterizing prediction errors and learning specific models for these “extreme” cases, e.g. geographically localized errors, or to bias training samples to better handle these errors.

A third task will be to investigate formalisms (i.e. languages) that make DNNs interpretable or partially interpretable, and the definition of algorithms that make possible the discovery or learning of DNN descriptions and parameterizations with regard to a related language.

References


Informations générales

- **Thème/Domaine**: Optimisation, apprentissage et méthodes statistiques
  Statistiques (Big data) (BAP E)
- **Ville**: Lyon
- **Centre Inria**: CRI Grenoble - Rhône-Alpes
- **Date de prise de fonction souhaitée**: 2019-09-02
- **Durée de contrat**: 3 ans
- **Date limite pour postuler**: 2019-07-08

Contacts

- **Equipe Inria**: DANTE
- **Directeur de thèse**: Gonzalves Paulo / paulo.gonzalves@inria.fr

A propos d'Inria

Inria, l'institut national de recherche dédié aux sciences du numérique, promeut l'excellence scientifique et le transfert pour avoir le plus grand impact. Il emploie 2400 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3000 scientifiques pour relever les défis des sciences informatiques et mathématiques, souvent à l'interface d'autres disciplines. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 160 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

Advisors: The PhD candidate will work in strong collaboration with Stefan Duffner, Marc Plantevit, Celine Robardet and Christophe Garcia. She/he will also work in a fruitful environment in contact to other members of Academics.

Required qualifications: Master's or Diploma degree in Computer Science, strong experience machine learning and data mining.

Duration: 3 years.
Monthly net salary: 1700 e. Possibility to teach starting the 2nd year of PhD (+200 e/ month).

Location LIRIS, Lyon, France – http://liris.cnrs.fr

Starting date: the position will start in September 2018.

Contacts To apply, the candidate will send an email including a motivation letter, a CV, copies of diplomas, to celine.robardet@insa-lyon.fr.

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.

**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.