Offer #2024-08415

Master 2 internship: Development of a deep latent block model for co-clustering

Contract type : Internship

Level of qualifications required : Graduate degree or equivalent

Fonction : Internship Research

About the research centre or Inria department

The Inria center at Université Côte d'Azur includes 42 research teams and 9 support services. The center's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRAE, INSERM ...), but also with the regional economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaborative robotics, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

Context

The proposed internship is in the context of co-clustering which consists in simultaneously clustering the rows and the columns of an array of data [1], this is particularly useful to summarize large datasets (see Figure 1). A popular probabilistic co-clustering model is the latent block model [3](LBM), it assumes that the clusters in each row and each column are drawn independently from two multinomial distributions and that given these clusters all the entries of the data array are independent, and that each entry follows a distribution only depending on its clusters in row and column. In the internship, we propose to develop an extension of the LBM in the case of binary data by assuming that each row and each column can be encoded by a latent position in an Euclidean space and that the parameter of the distribution of each entry only depends on these latent positions similarly to [5]. This model will allow to perform both co-clustering and visualization of the data through the latent positions as in [2]. For the parameters inference we will consider a variational approach as in [2] by making use of a neural network architecture for the approximate posterior distribution of the latent variables.

References

[1] Christophe Biernacki, Julien Jacques, and Christine Keribin. A survey on model-based co-clustering: High dimension and estimation challenges. 2022.

[2] Reⵎmi Boutin, Pierre Latouche, and Charles Bouveyron. The deep latent position topic model for clustering and representation of networks with textual edges, 2024.

[3] Vincent Brault and Mahendra Mariadassou. Co-clustering through latent bloc model: A review. Journal de la Socieⵎteⵎ Francⵎaise de Statistique, 156(3):120–139, 2015.

[4] Geⵎrard Govaert and Mohamed Nadif. Block clustering with bernoulli mixture models: Comparison of different approaches. Computational Statistics and Data Analysis, 52(6):3233–3245, 2008.

[5] Mark S Handcock, Adrian E Raftery, and Jeremy M Tantrum. Model-based clustering for social networks. Journal of the Royal Statistical Society: Series A (Statistics in Society), 170(2):301–354, 2007.

Assignment

The main mission of the internship will be to write the mathematical model and its parameters inference, and perform its implementations on Python. Moreover, the accuracy of the proposed methodology will also be studied on real data sets.
A thesis subject may be proposed as a continuation of this internship.

Main activities

- Bibliographic research
- Mathematical calculations
- State-of-the-art writing
- Programming
- Interpretation of results

Skills

Technical skills and level required:

- Python programming: Advanced level, with experience in libraries such as NumPy, Pandas, Scikit-learn, PyTorch, or TensorFlow.
- Experience with machine learning frameworks and tools.
- Knowledge of statistical modeling, optimization techniques, and data preprocessing.

Languages:

- English: Professional working proficiency (for documentation and collaboration in an international team).
- French: Optional but appreciated

Relational skills:

- Ability to work collaboratively in a multidisciplinary team.
- Strong problem-solving mindset and critical thinking.
- Good communication skills for presenting findings and writing reports.

Other valued appreciated:

- Prior experience with research projects or internships in AI/ML.
- Interest in contributing to open-source projects.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- 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 (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Contribution to mutual insurance (subject to conditions)

Remuneration

Traineeship grant depending on attendance hours.

General Information

- **Theme/Domain** : Optimization, machine learning and statistical methods
  Statistics (Big data) (BAP E)
- **Town/city** : Sophia Antipolis
- **Inria Center** : Centre Inria d'Université Côte d'Azur
- **Starting date** : 2025-03-01
- **Duration of contract** : 6 months
- **Deadline to apply** : 2025-02-28

Contacts

- **Inria Team** : MAASAI
- **Recruiter** :
  Vandewalle Vincent / Vincent.Vandewalle@inria.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

Instruction to apply

Applications must be submitted online on the Inria website. Collecting applications by other channels is not guaranteed.

Defence Security :
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST).Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :
As part of its diversity policy, all Inria positions are accessible to people with disabilities.