Contexte et atouts du poste

Mission confiée

Contexte: Representation learning has long been an important step in machine learning and signal processing [1]. Over the past few years, deep latent variable generative models, including variational autoencoder (VAE) and normalizing flow (NF), have been extensively utilized to model data generative processes and to learn a representation of data. A representation vector provides a concise description of data and is desired to capture and disentangle interpretable factors of variation in data. This enables learning explainable, robust, and universal representations that generalize well across different domains and tasks, and provides human-controllable data generation [2].

In particular, speech as a complex signal carries various sources of information, ranging from (speaker-independent) linguistic content to identity, age, emotion, and health status of the speaker, which are entangled and combined in a sophisticated way to produce speech. In privacy-preserving speech processing, for instance, disentangling the speech representation in order to keep only the required attributes and to remove the identity-related ones is of paramount importance. In some other applications, e.g. expressive text-to-speech and voice conversion, it is desired to manipulate the generative factors of speech.

Unsupervised disentangled representation learning has been proven to be impossible without inductive bias [3], and supervised learning requires expensive labeled data. Nevertheless, weakly supervised and semi-supervised approaches have shown promising performance recently [4]. In weak supervision, one has access to pairs of data that share some subsets of generative factors, and in semi-supervision, a limited set of data with known ground-truth generative factors is available.

Principal activités

Project description: Despite several attempts to learn disentangled and interpretable representations for sequential data, including speech, they are still far from achieving satisfactory performance. Furthermore, the majority of the existing works assume independence on the latent dimensions, whereas in real-world applications, subsets of latent codes (generative factors) might be correlated.

On the other hand, human perception is inherently multimodal, yet there is only limited research on multimodal disentangled representation learning. Specifically, the visual modality, e.g. jaw and lip movements, which provides complementary information about speech [5], has not been properly investigated in this context, neither as a stand-alone signal (e.g. for lip-reading tasks) nor in combination with the acoustic modality (e.g. for audio-visual speech processing).

In this post-doc project, we are going to develop efficient frameworks to learn disentangled and interpretable representations for audio-visual speech. To achieve this objective, weakly/semi-supervised approaches along with flexible (hierarchically structured) latent variable models will be investigated. We will also explore transfer learning techniques to enable learning with the help of second dataset (synthetic or real) that has ground truth annotations for generative factors. We will build our models using probabilistic generative modeling, in particular, flow-based latent variable models will be investigated in this context, neither as a stand-alone signal (e.g. for lip-reading tasks) nor in combination with the acoustic modality (e.g. for audio-visual speech processing).

Informations générales

- **Thème/Domaine:** Langue, parole et audio
- **Ville:** Villers-lès-Nancy
- **Centre Inria:** CRI Nancy - Grand Est
- **Date de prise de fonction souhaitée:** 2021-10-04
- **Durée de contrat:** 1 an, 4 mois
- **Date limite pour postuler:** 2021-05-30

Contacts

- **Equipe Inria:** MULTISPEECH
- **Recruteur:** Sadeghi Mostafa

A propos d’Inria

Inria est l’institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 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 personnelles d’appui à la recherche, la innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria est associé avec de nombreuses entreprises et a accompagné la création de plus de 180 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

Application deadline:

May 30th, 2021 (Midnight Paris time)

How to apply

Upload your file on jobs.inria.fr in a single pdf or zip file, and send it as well by email to mostafa.sadeghi@inria.fr, romain.serizel@loria.fr and emmanuel.vincent@inria.fr. Your file should contain the following documents:

- CV including a description of your research activities (2 pages max) and a short description of what you consider to be your best contributions and why (1-page max and 3 contributions max), the contributions could be theoretical or practical. Weblinks to the contributions should be provided.
- Include also a brief description of your scientific and career projects, and your scientific positioning regarding the proposed subject.
- The report(s) from your Ph.D. external reviewer(s), if applicable.
- If you haven’t defended yet, the list of expected members of your Ph.D. committee (if known) and the expected date of defense (the defense, not the manuscript submission).

In addition, at least one recommendation letter from your Ph.D. advisor should be sent directly by their author(s) to mostafa.sadeghi@inria.fr, romain.serizel@loria.fr, and emmanuel.vincent@inria.fr.

Applications are to be sent as soon as
- Ph.D. degree in the field of speech/audio processing, computer vision, machine learning, or in a related field,
- Ability to work independently as well as in a team,
- Solid programming skills (Python, PyTorch),
- Decent level of written and spoken English.

Avantages
- 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
- Social security coverage

Rémunération
Salary: 2653€ gross/month

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.