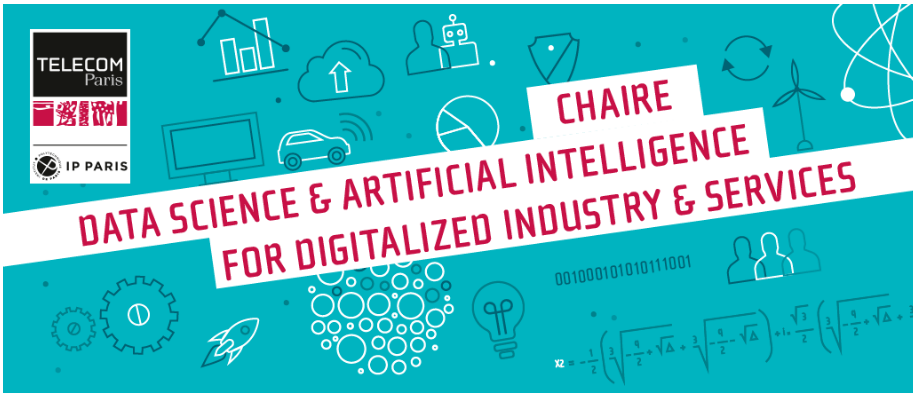

Offre de stage

Sujet : Apprentissage dynamique de caractéristiques par champs aléatoires conditionnels pour la détection des sentiments dans les conversations téléphoniques

Possibilité de poursuivre sur une thèse

La Chaire Data Science and Artificial Intelligence for Digitized Industry and Services

Portée par Florence d'Alché-Buc, enseignante-chercheur dans le département Image, Données, Signal de Télécom ParisTech, la chaire DSAI réunit cinq partenaires industriels : Airbus Defence & Space, Engie, Idemia, Safran et Valeo Finance. Son objectif général est de développer, en liaison étroite entre les Parties, une formation et une recherche de niveau international.

Ses quatre principaux axes de recherche sont :

1. Analyse et prévision de séries temporelles (Predictive Analytics on Time Series) ;

2. Exploitation de données hétérogènes, massives et partiellement étiquetées (Exploiting Large Scale and Heterogeneous, Partially Labelled Data) ;

3. Apprentissage pour une prise de décision robuste et fiable (Learning for Trusted and Robust Decision) ;

4. Apprentissage dans un environnement dynamique (Learning through Interactions with a Changing Environment).

Description du stage

Encadrement
Ekhine Irurozki, Ons Jelassi

Lieu et dates du stage
Telecom Paris, 19 Place Marguerite Perey, 91120 Palaiseau

Date de début du stage : début 2021

Équipe(s) d'accueil de la thèse
Laboratoire LTCI, équipes Signal, Statistique et Apprentissage (S^2A) et MultiMedia (MM).

Mots clés
Deep learning, Combinatorial optimization

Sujet détaillé
The **motivation** of this proposal is to bridge the gap between two major areas in applied mathematics: **Deep learning (DL)** and **combinatorial optimization (CO)**. DL is the hottest topic in research and industry. Because of their impressive results in applications such as image recognition or text translation DL is the state-of-the-art in most applied problems.

In CO the goal is to find an optimal object among a finite set of such objects. One of the best-known problems is the Traveling salesman problem, TSP. In the TSP, the input to the problem is a set of cities and the distances among them (represented by a graph) and the goal is to find the tour (a permutation of cities) that travels to all the cities in which the total distance is minimized. In a regular computer, a problem with 50k cities takes more than 20 years! Since huge problems of this kind arise in many areas, for example network planning, the most common approach when solving CO problems is to obtain a 'good' solution in a 'reasonable' time along with an estimation/bound on how far our solution is from the optimal solution. However, this approach lack any transferability property.

A big step in CO towards transferability has been taken recently transforming the combinatorial optimization problem into a prediction problem [1, 2, 10] . In this way, we can predict a good solution from previous data. This groundbreaking methodology has been carried out by using known, standard methodologies for different kinds of data (such as text) for combinatorial data (such as graphs or permutations). However, it is known that specific data requires specific techniques. Thus, in this proposal, **we will develop specific methodologies for combinatorial data**.

The proposal is divided into two **tasks**: Extracting information from combinatorial data and analyzing which information is relevant for which problem. Extracting relevant information from the data can be trivial for numerical data but it is far obvious for combinatorial data. We can find in the literature different embeddings for graphs [6] and permutations [4, 7, 9]. In this task, we review and propose new embeddings for graphs and permutations. A theoretical analysis of the convergence properties for existing learning algorithms is in order.

In the second part of the internship, we will analyze which embeddings are appropriate for each problem. This analysis will be carried out experimentally and it is based on related previous results [3, 5, 8]. The intern will code an end-to-end Neural Network using the embedding studied in the first task, he/she will test it on benchmarking problem instances. We will compare the results to baseline methods and state-of-the-art methods, elaborating the results in terms of quality of the solution, transferability and time performance. For the simulations, there will have access to computational resources.

Interest in a PhD thesis on the continuation of the internship is highly wishable. This proposal describes some first steps in a hot area of research in which both methodological and practical lines follow naturally.

Profil du candidat

Student holding a Master 2 researchM

- Statistical learning, interest in combinatorics

- Good level in programming (Java, C/C++, Python)

- Good level of English

Candidatures

Send to irurozki@telecom-paris.fr

- Curriculum Vitae

- Personalized cover letter explaining the candidate's interest in the subject (directly in the body of the email)

- Transcripts from previous years

- Contact a reference person

Incomplete applications will not be considered.

Références

[1] Irwan Bello, Hieu Pham, Quoc V. Le, Mohammad Norouzi, and Samy Bengio. Neural combinatorial optimization with reinforcement learning. In 5th International Conference on Learning Representations, ICLR 2017 - Workshop Track Proceedings, 2019.

[2] Yoshua Bengio, Andrea Lodi, and Antoine Prouvost. Machine learning for combinatorial optimization: A methodological tour d'horizon. European Journal of Operational Research, 2020.

[3] J. Ceberio, E. Irurozki, A. Mendiburu, and J.A. Lozano. A distance-based ranking model estimation of distribution algorithm for the flowshop scheduling problem. IEEE Transactions on Evolutionary Computation, 18(2), 2014.

[4] J. Ceberio, E. Irurozki, A. Mendiburu, and J.A. Lozano. A review of distances for the Mallows and Generalized Mallows estimation of distribution algorithms. Computational Optimization and Applications, 62(2), 2015.

[5] Josu Ceberio, Ekhine Irurozki, Alexander Mendiburu, and Jose A. Lozano. A review on estimation of distribution algorithms in permutation-based combinatorial optimization problems. Progress in Artificial Intelligence, 1(1):103–117, 2012.

[6] Hanjun Dai, Elias B. Khalil, Yuyu Zhang, Bistra Dilkina, and Le Song. Learning combinatorial optimization algorithms over graphs. In Advances in Neural Information Processing Systems, 2017.

[7] E. Irurozki, B. Calvo, and J.A. Lozano. Learning probability distributions over permutations by means of Fourier coefficients, volume 6657 LNAI. 2011.

[8] M. Malagon, E. Irurozki, and J. Ceberio. Alternative Representations for Codifying Solutions in Permutation-Based Problems. In 2020 IEEE Congress on Evolutionary Computation, CEC 2020 - Conference Proceedings, 2020.

[9] Arjun Seshadri, Stephen Ragain, and Johan Ugander. Learning Rich Rankings. In Neural Information Processing System (NeurIPS), volume 5, pages 1–12, 2020.

[10] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In Advances in Neural Information Processing Systems, 2015.