

Title of the project: Statistical Modeling and Feature Extraction for Large-scale Functional Data

Name, affiliation, and contact information for the supervisor and co-supervisor

Supervisor: Chen Xu

Associate Professor

Department of Mathematical and Statistics, University of Ottawa

cx3@uOttawa.ca

<https://science.uottawa.ca/mathstat/en/people/xu-chen>

Co-supervisor: Peijun Sang

Assistant Professor

Department of Statistics and Actuarial Sciences, University of Waterloo

psang@uwaterloo.ca

<https://uwaterloo.ca/scholar/psang>

Abstract [0.25 pages]

In modern scientific research, data scientists and statisticians often need to deal with large-scale functional data, which contain a large number of functional covariates (features) possibly measured from different resources or coming from a data stream. The complex data structure brings serious challenges for traditional statistical modeling in terms of statistical accuracy, computational efficiency, and algorithm stability. This proposal aims to provide a qualified postdoc fellow an excellent opportunity to explore this emerging research area with two experienced experts in the field. In particular, the fellow is encouraged to work on one or more of the following specific topics. 1) Screening and selection the important features in high-dimensional functional regression; 2) Dimensionality reduction and feature extraction for streaming functional data; 3) Distributed modeling for functional data collected from different resources. With the above research objectives, this project will build a theoretical foundation and develop effective implementation procedures for large-scale functional data.

Interdisciplinary/applied experience [0.5 pages]

The postdoc fellow will be strongly encouraged to apply the developed methodology to solving real world problems. Benefited from the location of the University of Ottawa (uOttawa), the postdoc fellow will be given good opportunities to gain hands-on data analytic experience via existing collaboration agreements between uOttawa and its research partners such as Statistics Canada and Kanata North, the home to Canada's largest tech-park.

In particular, the methodological results from the proposed research topic 1) can be readily used to identify relevant wavelength numbers that are predictive of progression of the gluing process

from sequential NIR spectra along with wood panel compression. The postdoc fellow will work closely with the co-supervisor Dr. Peijun Sang at the University of Waterloo (uWaterloo); he has extensive experience in functional data analysis and has a long-term collaboration with FPInnovations to tackle important scientific problems arising from lumber industry. The postdoc fellow will be encouraged to justify the developed methods using the real lumber datasets accessible by the supervisory team. The related computing will be conducted on servers at the High Performance Computing Virtual Laboratory, which is accessible for students and research personals at both uOttawa and uWaterloo.

Teaching/training/education [0.5 pages]

The postdoctoral fellow will teach one three-credit course at uWaterloo in the first year and one three-credit course at uOttawa in the second year. The postdoctoral fellow will have great opportunities to participate in various activities held by SSC, CANSSI, and CANSSI Ontario, like presenting at SSC annual meetings and CANSSI workshops.

In addition, the fellow is encouraged to mentor undergraduate and MSc projects while staying at the two hosting institutions. The fellow is also encouraged to actively involve in coaching student teams for various modeling and data analytic competitions, such as the annual MathClub at uOttawa, SSC case study, and the ASA DataFest.

Mentoring of the postdoctoral fellow [0.5 pages]

The two mentors share common research interest in theoretical developments on feature screening and selection in high dimensional data. The CANSSI Distinguished Postdoctoral Fellow will stay at the University of Waterloo in the first year and the University of Ottawa in the second year. Therefore, the candidate will be exposed to training by both De. Chen Xu and Dr. Peijun Sang in dealing with large-scale functional data.

The postdoctoral fellow will have regular interactions with us through weekly meetings, and will be encouraged to make presentations at professional meetings like SSC annual meeting and CANSSI sponsored workshops. Though the mentors will provide necessary guidance in the aforementioned projects, the successful candidate is expected to make independent contributions in the collaborative research. We will help the trainee to better appreciate the motivation of these projects, and understand computational and theoretical challenges to tackle them. More specifically, in our regular interactions, both supervisors will provide guidance for the postdoctoral fellow on understanding the research problem, proposing sensible statistical models to address the problem, developing efficient algorithms to fit the model, establishing statistical theories, conducting numerical studies, and writing a research manuscript. Besides the training on the hard technical skills, both supervisors will help with enhancing soft skills of the postdoctoral fellow such as oral and written communications skills, and critical and independent thinking. Regardless of future plans of the trainee, we believe that the training on both hard technical skill and soft skills will help him/her to achieve great success.

A list of qualifications of suitable candidates [0.25 pages]

Successful candidates are expected to work with Professor Chen Xu and Professor Peijun Sang

on topics including but not limited to large-scale functional and high dimensional data. The proposed research will be viewed and investigated jointly from the perspectives of mathematical statistics, machine learning, optimization, and approximation theory. The scientists trained by the project are urgently needed for analyzing big functional data and high dimensional data in industry, research institutes, and government agencies.

Position Qualifications: A Ph.D. in Statistics, Biostatistics, Mathematics, Computer Science, or other relevant discipline is required.

Proposed schedule for postdoctoral fellowship [0.5 pages]

Schedule for year 1: Research: Develop efficient algorithms in the first and second themes and derive theoretical properties of the estimators, write the manuscripts; Teaching: Teach one three-credit course at the University of Waterloo; Activities: get involved in the collaboration with FPInnovation, present the research work on professional conferences like SSC and CANSSI workshops, coach students teams in the ASA DataFest hosted by University of Waterloo.

Schedule for year 2: Research: Develop efficient algorithms in the second and third themes and derive theoretical properties of the estimators, write the manuscripts; Teaching: Teach one three-credit course at the University of Ottawa; Activities: join the collaborative research with Statistics Canada and Kanath North, present the research work on professional conferences like SSC and CANSSI workshops, coach students teams in SSC case studies and the annual MathClub at uOttawa.