



NLHPC: La experiencia de trabajar usando recursos de computación de alto desempeño

Paola Bordón, Ph.D. in Economics, University of Wisconsin-Madison, EE. UU. Directora Magíster en Economía Aplicada en Políticas Públicas, FEN-UAH; y Fernando Crespo, Doctor en Ciencias de la Ingeniería de la Pontificia Universidad Católica de Chile, académico FEN-UAH.



La computación de alto desempeño (HPC, en su acrónimo en inglés) se entiende por compartir simultáneamente: tener recursos con varias horas de disponibilidad, infraestructuras de redes, seguridad y ejecución de programas computacionales en diferentes lenguajes usando programación paralela de alta velocidad, todos los cuales se encuentran disponibles en el Laboratorio Nacional de Computación de Alto Rendimiento, NLHPC, alojado en el Centro de Modelamiento Matemático de la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile.

La supercomputación, otra forma de denominarle, sirve para varios propósitos, entre los que se encuentran realizar simulaciones cuya carga sería imposible de hacerlo en computadores domésticos, ejemplo de estos se relacionan con desarrollo de nuevos materiales, nanotecnología, fluido dinámica, simulaciones para búsqueda de nuevas drogas y medicamentos, investigaciones que utilizan y procesan grandes cantidades de datos como en genómica y astronomía, otras que buscan soluciones en problemas que requieren alta velocidad para

encontrar respuestas en tiempo razonable, cargas de entrenamiento de modelos intensivos en cálculo como lo son los casos de ejecución para entrenamiento de aprendizaje profundo (Deep Learning).

Como consecuencia natural, disponer de los recursos antes mencionados, nos permite situarnos con desarrollos e investigaciones de punta, y potencia la articulación del desarrollo de tecnologías que requieran en alguna de sus etapas de desarrollo de la anterior necesidad de dichos recursos tecnológicos.

La infraestructura del NLHPC ofrece recursos computacionales y de fomento a la colaboración de la comunidad científica del país para potenciar el impacto de la investigación, desarrollo e innovación. Está abierta para investigadores alocados en Chile, y pertenecientes a Universidades o Institutos de Investigación del País, para ello es necesario crear una cuenta que permita acceder a sus servidores. En otros casos, hay convenios con instituciones y empresas privadas que requieren de este tipo de infraestructura.

Las cifras que exhibe el NLHPC muestra que tiene más de 838 registros asociados a publicaciones con participación de investigadores nacionales que agradecen haber utilizado los servicios disponibles y atender a más de 456 usuarios activos, con un total de 31.233.207 horas de procesamiento.

Nosotros hemos tenido acceso al NLHPC para realizar investigación de punta intensiva en procesamiento de datos. Paola Bordón y coautoras lo utilizaron para la estimación de modelos de elección de carreras universitarias en Chile y cuantificar la brecha de género. Para ello, se estima un modelo logit anidado para predecir las preferencias de estudiantes al momento de postulación. Utilizando información de postulaciones del Departamento de Evaluación, Medición y Registro Educacional (DEMRE) se tienen 76,680 estudiantes que se graduaron de la enseñanza media el año 2014 y postularon a una carrera universitaria a través del sistema único de admisión.

Para la estimación y calibración del modelo usando el software Stata y el NLHPC, incluimos características de los estudiantes, tales como, el género, la ocupación de los padres, tipo de colegio, ubicación geográfica, puntajes en las pruebas de admisión universitaria, notas por curso en la enseñanza media, situación socioeconómica del hogar, ayuda financiera; y características de las carreras como requisitos de postulación, puntajes de corte, aranceles y vacantes disponibles. Así, estimamos los efectos marginales por área de estudio, universidad y género. Los errores de predicción fueron menores al 1%, lo que nos permitió analizar de manera fidedigna como el género afecta las postulaciones a las distintas carreras universitarias en Chile.

Nuestros resultados mostraron que los hombres postulan a programas más selectivos, incluso cuando es probable que no sean admitidos, en mayor proporción que las mujeres de similares características.

Finalmente, realizamos una serie de contrafactuales para ver que escenario reduciría en mayor medida la brecha de género. Esto significó nuevamente estimar para todo el set de estudiantes las probabilidades de postulación

para cada área y universidad, considerando tres escenarios: que pasaría si las mujeres postularan como los hombres, si los hombres postularan como las mujeres, y si los estudiantes postularan con preferencias promedio entre hombres y mujeres, lo que es altamente intensivo en computo por la cantidad de estudiantes y variables consideradas. Estos ejercicios contrafactuales nos permitieron concluir que para disminuir considerablemente la brecha de género en las áreas STEM (ciencias, tecnología, ingeniería y matemáticas) se debe no solo promover la participación femenina de estas carreras desde la niñez, sino que también

Gracias a la disponibilidad del NLHPC, pudimos correr más de 120 modelos, con el fin de elegir el modelo más representativo, con más de 10 mil tweets asociados a diferentes usuarios, ajustados por el comportamiento de tres días, el día anterior a la protesta, el día de la protesta, y el día posterior, lo que generó una carga enorme de datos, donde en algunos casos los modelos no convergen, con todo la ejecución de cada modelo tardaba alrededor de 72 horas, gracias a la disponibilidad de procesamiento paralelo, en un semana completa de 7 días pudimos completar la ejecución de los modelos. El mejor modelo explica la densidad



La supercomputación, otra forma de denominarle, sirve para varios propósitos, entre los que se encuentran realizar simulaciones cuya carga impide hacerlo en computadores domésticos”

incentivar a que los hombres consideren carreras no STEM. Para mayor detalle, se sugiere revisar el artículo Bordón, P., Canals, C. y Mizala, A. The gender gap in college major choice in Chile, *Economics of Education Review*, Volume 77, 2020, 102011.

Por su parte, Fernando Crespo y coautores lograron ajustar modelos de captura recaptura (spatial capture-recapture modeling) que se utilizan en biología, para seguir la pista de los desplazamientos de especies que recorren sus territorios en busca de comida, como una forma de ver el patrón de comportamiento de sus desplazamientos habituales, los que se denominan como sus centros de actividad. El modelo estaba programado en el software R. Para ajustar el modelo, cada punto representa un evento en una trayectoria, y cada individuo puede estar además caracterizado por otras variables explicativas. Esto hace que cuando existe un grupo de individuos, el modelo es altamente intensivo en cálculo. Nosotros lo aplicamos para estudiar el comportamiento de viandantes en protestas espontáneas sobre información georreferenciada en Ciudad de México, con datos obtenidos de la plataforma Twitter.

espacial de los usuarios, una alta variabilidad en la probabilidad de detectar los usuarios favorables a la protesta, y que el ritmo del posteo colectivo y el día de observación son factores explicatorios. Se pudo observar como cambiaban los centros de actividad de los individuos que participaban de la protesta, no sólo por un efecto de proximidad a ésta, ya que ese efecto no explica adecuadamente la variación espacial en la probabilidad de detección y densidad durante la protesta. Nuestra investigación está referenciada como Masías, Víctor Hugo; Crespo R, Fernando A.; Navarro R, Pilar; Masood, Razan; Krämer, Nicole C.; Hoppe, H. Ulrich (2021). On spatial variation in the detectability and density of social media user protest supporters. *Telematics and Informatics*, 65, 2021, 101730. <https://doi.org/10.1016/j.tele.2021.101730>.

Está por demás, agradecer tener acceso y disponibilidad al NLHPC, y su importancia para poder participar en investigación compatible con la realizada en condiciones de mejores recursos en países desarrollados. Y obviamente, esperamos seguir ocupando esta capacidad operativa disponible. 