

A) Objetivos del proyecto de investigación.

1. Comportamiento de un modelo de red neuronal dado, a medida que se varía el tamaño del conjunto de entrenamiento.
2. Métodos de regularización y su efecto para controlar el sobre-entrenamiento cuando se usan modelos relativamente complejos y flexibles.

B) Justificación que fundamente por qué los recursos de supercómputo solicitados son necesarios para desarrollar el proyecto.

Dentro de una colaboración entre la Facultad de Ciencias de la UNAM (Dr. Humberto Andrés Carrillo Calvet) y el Instituto Nacional de Investigaciones Nucleares (Dr. Juan Jaime Vega Castro), hemos llevado a cabo el estudio (ver referencias) de redes neuronales de alimentación hacia adelante, las cuales las entrenamos usando el algoritmo de retro-propagación para ser usadas como reconocedoras de patrones. Estos patrones son curvas de Bragg digitalizadas (curvas de poder de frenado específico de iones en un medio gaseoso). Esto es de interés para la espectroscopia de curva de Bragg, la cual es ampliamente usada en física nuclear experimental para la identificación de las partículas o iones emergentes de una reacción nuclear. Esta técnica espectroscópica esencialmente consiste en determinar dos cantidades: la energía del ion incidente determinada por el área bajo la curva de Bragg, y el número atómico del ion determinado por la amplitud del pico de Bragg o máximo de la curva de Bragg.

Como todas las señales que son obtenidas como resultado de una medición, las curvas de Bragg no son curvas ideales sino que presentan cierta cantidad de ruido experimental. En vista de esto, hemos estudiado el efecto del ruido presente en los patrones o curvas de Bragg sobre el desempeño de las redes neuronales como reconocedoras de curvas de Bragg. Esto se ha hecho mediante el estudio de la dependencia del error mínimo alcanzado por las curvas del error de validación como función de la razón de señal a ruido presente en las curvas de Bragg. Han sido 10 las razones de señal a ruido estudiadas: 1%, 2%, ..., 10%. Esto lo hemos realizado hasta ahora usando redes neuronales de retro-propagación con aprendizaje en línea y aprendizaje en mini-lotes, usando con una arquitectura de red neuronal de 81 neuronas en la capa de entrada, 5 capas de neuronas ocultas de 9 neuronas cada una, y una capa de salida de 2 neuronas. Uno de los problemas que se ha observado como resultado de este estudio es el de sobre-entrenamiento de las redes neuronales. Dos alternativas para remediar este problema observado serían:

- 1) Para la misma arquitectura de red neuronal hasta ahora estudiada, emplear conjuntos de entrenamiento más grandes, lo cual reduciría el sobre-entrenamiento de la red neuronal.

- 2) Utilizar algún método de regularización como por ejemplo decaimiento de pesos, lo cual contribuiría a que la complejidad del modelo no crezca innecesariamente dando lugar a sobre-entrenamiento.

Este tipo de estudios basado en redes neuronales son altamente consumidor de recursos computacionales, en vista de lo cual requerimos el soporte de supercómputo para llevarlos a cabo.

Referencias

- J. J. Kolata, J. J. Vega, B. B. Back, C. N. Davids y D. J. Henderson, "Bragg curve detector for the FMA focal plane", ANL-90/18, Physics Division Annual Review, Abril 1, 1989 - Marzo 31, Argonne National Laboratory (1990) 108.
- J. J. Vega, J. J. Kolata, W. Chung, D. J. Henderson and C. N. Davids, "Bragg curve spectroscopy and digital processing", en: (Ed.), M. Brandan, Memorias del XIV Symposium on Nuclear Physics, Cuernavaca, Mexico, 1991, World Scientific, Singapore, 1991, 221-230.
- J. J. Vega, "Redes neuronales y física nuclear", conferencia invitada al II Ciclo de Conferencias Sobre Neurocomputación (aplicaciones y tendencias), Centro Nacional de Cálculo, IPN, Mayo-Julio, 1992.
- M. R. Reynoso, J. J. Vega, J. Fernández, E. Belmont, R. Policroniades y E. Moreno, "Una red neuronal para el reconocimiento de curvas de Bragg sintéticas", VIII Seminario IMP-IIE-ININ sobre Especialidades Tecnológicas, Cuernavaca, Junio 26, 1996.
- M. R. Reynoso, J. J. Vega, J. Fernández, M. Torres, R. Policroniades y E. Moreno, "Utilización de reconocimiento de patrones para el procesamiento de señales obtenidas con un espectrómetro de Curva de Bragg", Simposio Internacional de Computación "Nuevas Aplicaciones e Innovaciones Tecnológicas en Computación", Noviembre 1997.
- J. J. Vega, M. R. Reynoso, M. A. Estrada y L. A. Robles, "Bragg Curve Identification Using a Neural Network", IEEE-INNS-ENNS International Joint Conference on Neural Networks 2000, Como, Italy, Julio 24-27, (2000) 4379-4383.
- J. J. Vega y M. R. Reynoso, "Instrumentación para la Espectroscopia de Curva de Bragg Utilizando Redes Neuronales Artificiales", conferencia invitada a SOMI XVII Congreso de Instrumentación, Mérida, Yucatán, México, Octubre 2002.
- J. J. Vega y M. R. Reynoso, "Redes Neuronales Artificiales Aplicadas a la Espectroscopia de Curva de Bragg", conferencia invitada al XLV Congreso Nacional de Física, León, Guanajuato, Noviembre 2002.
- J. J. Vega y M. R. Reynoso, "Application of Neural Networks to Pulse-Shape Analysis of Bragg Curves", J. The 7th World Multiconference on Systemics, Cybernetics and Informatics SCI 2003, Orlando Julio 27-30 2003.
- M. R. Reynoso y J. J. Vega, "Effect of the size of an artificial neural network used as pattern identifier", I International Symposium on Radiation Physics, Diciembre 1-5, 2003, México, D.F.
- J. J. Vega y M. R. Reynoso, "Application of Neural Networks to Pulse-Shape Analysis of Bragg Curves", CAARI 2004: 18th International Conference on the Application of Accelerators in Research and Industry, Octubre 10-15, 2004, CAARI 2004, Fort Worth TX, USA.
- J. J. Vega y M. R. Reynoso, "Effect of noise on the identification of digitized Bragg Curves", II International Symposium on Radiation Physics, Febrero 27- Marzo 2, 2006, Chihuahua, Chihuahua, México.
- J. J. Vega, R. Reynoso, "Application of neural networks to pulse-shape analysis of Bragg curves", NIM B 243 (2006) 232-240.
- J. J. Vega, R. Reynoso, "Effect of noise on the identification of digitized Bragg Curves". Rev. Mex. de Fís. 53 S3 (2007) 118-124.
- J. J. Vega, R. Reynoso, "Learning limits of an artificial neural network", Rev. Mex. de Fís. S 54 (1) (2008) 22-29.
- J. J. Vega, R. Reynoso, H. Carrillo Calvet, "Effect of signal noise on the learning capability of an artificial neural network", NIM A 606 (2009) 693-699.

C) Cantidad de los recursos de supercómputo en horas CPU, número de núcleos de trabajo requeridos y, en su caso, datos específicos como uso de GPUs, tamaño de memoria, recursos de disco, etc.

Horas CPU = 500,000 horas

Número de núcleos = 60 núcleos (60 procesos)

Tamaño de memoria = cada uno de los 60 proceso requiere 500 MB de memoria dando un total de 30 GB

Recurso de disco = 300 GB

D) Listado de productos a obtener: publicaciones, patentes y software, o bien una relación que señale el impacto del proyecto para la formación de recursos humanos.

Publicaciones y formación de recursos humanos.

E) Descripción general que explique la forma en que se utilizarán los recursos de supercómputo requeridos. Sólo aplica a las propuestas de la categoría Proyecto Semilla.

F) Descripción cualitativa y cuantitativa de los procesos de cómputo que se ejecutarán y descripción del software a utilizar, incluyendo sus características numéricas y computacionales. Aplica a las propuestas de las categorías Proyecto de Investigación Regular, Proyecto de Investigación Grande y Megaproyecto.

Se llevará a cabo el entrenamiento de 60 redes neuronales como reconocedoras de curvas de Bragg. 30 de estas corridas corresponden al entrenamiento de redes neuronales para las cuales el conjunto de entrenamiento consistirá de 451,000 patrones o curvas de Bragg. Las otras 30 corridas corresponden al entrenamiento de redes neuronales usando decaimiento de pesos como método regularizador. En relación al software que se utilizará para el entrenamiento de las redes neuronales, este será el mismo que hasta ahora hemos utilizado en los trabajos que hemos hecho hasta ahora y que se describen en el inciso B. Se trata del paquete SNNS - Stuttgart Neural Network Simulator. Éste es un simulador para redes neuronales para plataforma Unix. El objetivo de SNNS es crear un ambiente de simulación eficiente y flexible para la investigación y aplicación de redes neuronales.

El simulador SNNS consta de dos partes principales:

- 1) El núcleo del simulador en C.
- 2) Interfaz gráfica de usuario bajo X11R4 o X11R5.

G) Síntesis curricular del (los) solicitante (s). Debe expresar de forma clara su experiencia en relación con el tema del proyecto de investigación, el uso de equipos de supercómputo y del software que se empleará.

Síntesis Curricular

Profesor Titular "C" del Departamento de Matemáticas de la Facultad de Ciencias de la UNAM. Fundador y líder académico del Laboratorio de Dinámica no Lineal de esta Facultad. Investigador Nacional Nivel 3, PRIDE "D".

Obtuvo el grado de licenciado, maestro y doctor en Ciencias (Matemáticas) en la Facultad de Ciencias, UNAM. Realizó una estancia de investigación de dos años en el *Mathematics Department of the University of Utah, USA* y una estancia posdoctoral de un año en el *Lefschetz Center for Dynamical Systems, Division of Applied Mathematics, Brown University*.

Ha recibido distinciones internacionales, como el *Fogarty Research Award* de los *National Institutes of Health, USA*, donde tuvo una estancia como investigador invitado por un año. La Academia de Ciencias de Cuba lo reconoció explícitamente en el Premio Nacional de Ciencias 2002, por su colaboración para el desarrollo de "Herramientas para el Análisis de Oportunidades Científico-Tecnológicas" y el Consejo Nacional de Sociedades Científicas de la Salud de Cuba le otorgó el Premio Anual de Salud 2011, por la coautoría del libro *The Art and Science of Tuberculosis Vaccine Development*, publicado por *Oxford University Press*.

Su obra ha abarcado diversos aspectos del trabajo universitario: docencia, divulgación, investigación (básica y aplicada), desarrollo de tecnología, liderazgo de grupos académicos, fundación de sociedades científicas y la dirección de proyectos de vinculación de la academia con la sociedad.

Ha destacado como profesor y conferencista, pero su actividad docente no se ha restringido a la impartición de clases y la formación de profesionistas y científicos de la manera tradicional. Desde principio de la década de los 90's, con visión y audacia, incursionó en la experimentación de nuevas tecnologías y recursos computacionales para la enseñanza y la experimentación científica, esfuerzos que han resultado en la formación de varias generaciones de profesionistas, docentes y técnicos, capaces de crear y utilizar estas tecnologías. Por estos esfuerzos se le reconoce como uno de los pioneros en la generación de sistemas de software libre para la docencia en nuestra Máxima Casa de Estudios. Bajo su dirección la Facultad de Ciencias ha logrado consolidar una colección de 11 productos de software que son valiosos como herramientas de apoyo a la docencia y a la investigación en los temas de: Sistemas Dinámicos y Ecuaciones Diferenciales, Oscilaciones no Lineales, Redes Neuronales y Bayesianas, así como herramientas de visualización de información para la Cienciometría.

Los sistemas desarrollados están disponibles en la página web del Laboratorio de Dinámica no Lineal de la Facultad de Ciencias. Solamente de mayo de 2007 a junio de 2014, el sitio de este laboratorio recibió 58,548 visitas desde 115 países y se han registrado más de 2,700 descargas de los sistemas de software. Durante este período 56 textos de apoyo a la docencia, incluyendo

manuales y textos de difusión o divulgación científica, disponibles en este sitio han sido objeto de 40,375 descargas. Su labor de divulgación, además de la publicación de artículos, ha abarcado también la participación en medios periodísticos, programas de radio, televisión y la impartición de conferencias de difusión científica.

Siendo uno de los primeros mexicanos que realizó investigación en Biología Matemática, ha sido activo promotor en la Facultad de Ciencias de esta temática, pero también de otros temas científicos emergentes, como la Ciencia y Minería de Datos con Inteligencia Artificial, la Bioinformática y el estudio de Sistemas Dinámicos y Fenómenos no Lineales, tema primordial de su formación y del laboratorio que fundó en la Facultad de Ciencias.

Consecuentemente, con un enfoque transdisciplinar en su laboratorio se han trabajado diversos temas: tecnologías de la información, nuevos paradigmas de la Inteligencia Artificial, como las redes neuronales artificiales y su aplicación a la Cienciometría y la Física Nuclear, temas de Biología Teórica y Bioinformática, incluyendo la modelación biofísica-matemática de las células nerviosas y más recientemente el análisis computacional de genomas, con miras al diagnóstico de enfermedades como la tuberculosis y el cáncer de mama.

Por haber producido algoritmos eficientes para la alineación óptima de secuencias biológicas en colaboración con el médico David Lipman, Director y fundador del National Center for Biotechnology Information, [SIAM Journal of Applied Mathematics, Vol.48, No. 5, pp. 1073-1082, 1988], ha sido reconocido como uno de los pioneros de la Bioinformática [Early Bioinformatics: the birth of a discipline—a personal view. CA Ouzounis, A Valencia - Bioinformatics, Vol. 19, No. 17, Pp. 2176-2190, 2003].

Tiene contribuciones científicas en varios campos de conocimiento: Sistemas Dinámicos no Lineales, Inteligencia Artificial, Biología Matemática, Bioinformática, Cienciometría y Ciencias de la Información. Estas han sido reconocidas mundialmente con más de 700 citas incluyendo 70 en libros especializados, 58 tesis doctorales, además de artículos de investigación y patentes. Solamente en la base de patentes de los Estados Unidos (*US Patent and Trademark Office, USPTO*) el profesor Carrillo aparece citado en 120 patentes.

Ha sido miembro del Comité Editorial de la Revista Mexicana de Ingeniería Biomédica y actualmente es miembro del Comité editorial de la revista cubana Ciencia de la Información.

Es miembro fundador de 3 asociaciones científicas nacionales: la Academia Mexicana de Tecnología, la Sociedad de Neurociencias y Neurobiología de México y la Academia Mexicana de Ciencias, Artes, Tecnología y Humanidades.

Además de su obra científica, cuenta con 12 productos de ingeniería (sistemas de software) registrados ante el Instituto Mexicano de Derechos de Autor y ha impartido más de 100 cursos a nivel de licenciatura y posgrado. Bajo su dirección se han graduado 25 alumnos en las áreas de Matemáticas, Física, Actuaría,

Ciencias de la Información y Ciencias de la Computación (17 de licenciatura, 5 de maestría y 3 de doctorado).

Actualmente es miembro de la Comisión Universitaria de Vinculación Investigación Docencia de la UNAM.

En relación a la experiencia con el tema del proyecto de investigación, en el punto B) se presenta una descripción de esa experiencia.

En relación a la experiencia con el uso del software que se usará, en el punto F) se presenta una descripción de esa experiencia.

En relación a la experiencia con el uso de equipos de supercómputo, contamos con la experiencia de haber realizado proyectos en supercomputadoras anteriores de la DGESCA. Varios de nuestros alumnos fueron becarios en el programa de supercómputo.

H) Comprobante de adscripción a una entidad o dependencia de la UNAM, o bien con la que se tiene convenio para el uso de la supercomputadora.

Se entregará posteriormente en la coordinación de supercómputo.