

GA²LA: Grupo de Aprendizaje Automático, Lenguajes y Autómatas

Juan C. Vázquez¹, Julio Castillo¹, Leticia Constable¹, Marina Cardenas¹

Depto. de Ingeniería en Sistemas de Información / Facultad Regional Córdoba / Universidad Tecnológica Nacional

¹{jcjvazquez, jotacastillo, leticiaconstable, angelaesmeralda}@gmail.com

RESUMEN

Durante más de catorce años de desarrollo de proyectos de investigación en las áreas de autómatas, lenguajes y redes neuronales artificiales, entre otras muchas temáticas afines, se ha conformado una masa crítica de recursos humanos capacitados, dotados de los conocimientos, experiencia y trayectoria para generar un consistente equipo de trabajo. La necesidad de conformar un ámbito más estructurado, concentrado y formal donde se establezca la discusión científica, académica y práctica de los distintos temas involucrados, de las líneas de acción a emprender a futuro, de las formas de financiamiento a utilizar y del modo trabajo que promueva la sinergia de equipo, guían la propuesta de conformación de un Grupo UTN de I+D en temáticas afines a los proyectos ya realizados y en marcha de estos investigadores. El “*Grupo de Estudio, Investigación, Desarrollo y Transferencia en Aprendizaje Automático, Lenguajes y Autómatas*” (GA²LA) es presentado aquí con sus antecedentes, expectativas y líneas de investigación.

Palabras clave: redes neurales artificiales, autómatas, lenguaje formal, lenguaje natural.

CONTEXTO

En 2002, se crea el Laboratorio de Investigación de Software (LIS) en la Facultad Regional Córdoba de la Universidad Tecnológica Nacional, con aportes de Microsoft Research de E.E.U.U. y de la Universidad. Las actividades de investigación y desarrollo inician formalmente en el ámbito académico, aproximadamente por el año 2004. Desde entonces los sucesivos integrantes del LIS han definido y llevado adelante proyectos de investigación acreditados por la Secretaría de Cien-

cia y Técnica de la UTN y otros organismos (Agencia Córdoba Ciencia, Ministerio de Ciencia y Tecnología de la Prov. de Córdoba, CONICET, ANPCYT), relacionados al estudio, diseño, desarrollo, implementación y transferencia, en las temáticas de ingeniería de software, modelos evolutivos y conexionistas, máquinas abstractas/autómatas, procesamiento de lenguaje natural y de lenguajes formales. Estos proyectos que se encuadran en líneas afines de investigación son los que dan origen a la idea de conformar el grupo que presentamos en este artículo.

En UTN, “Grupo” y “Centro” son entidades formales de I+D con características especiales definidas en sus estatutos.

Algunos proyectos de los integrantes y exintegrantes del LIS han formado parte durante años del Grupo de Inteligencia Artificial (GIA) desde su constitución en 2007 y del CIDS (Centro de Investigación, Desarrollo y Transferencia de Sistemas de Información, ex grupo GIDTSI) desde su creación en 2014 (entidades de las cuales el director propuesto para el nuevo grupo GA²LA es miembro fundador). Así éstos, se constituyen en el origen natural del nuevo grupo en formación.

Es oportuno señalar, que quien se propone como director del grupo GA²LA fue el primer jefe a cargo del LIS y el codirector propuesto es el actual jefe a cargo del LIS, por lo que seguimos en contacto con el mismo.

1. INTRODUCCIÓN

Como producto de las actividades de I+D indicadas, en la temática específica del grupo GA²LA se desarrollan y se han desarrollado los siguientes proyectos:

- a. Relaciones entre los momentos de aprendizaje y de reconocimiento de las redes neuronales artificiales y la evolución espacio-temporal de autómatas celulares (2004-2005): este proyecto estudió la historia de los pesos sinápticos de una red neuronal artificial tipo perceptrón multicapa durante su entrenamiento, intentando determinar si la misma podía ser construida por algún autómata celular del tipo unidimensional. También se estudió la historia de los estados de las neuronas de redes Hopfield durante la fase de reconocimiento de patrones, con el mismo fin.
- b. Construcción de herramientas didácticas para la enseñanza y ejercitación práctica en laboratorio de la asignatura Sintaxis y Semántica del Lenguaje en la carrera de Ingeniería en Sistemas de Información (2006-2008): este proyecto sistematizó el conocimiento previsto para la asignatura SSL, generando una herramienta de autoevaluación y simuladores de máquinas abstractas y procedimientos sobre gramáticas formales.
- c. Evaluación de arquitecturas y técnicas de entrenamiento de redes neuronales multicapa de perceptrones (2006-2008): su intención fue determinar cómo lograr el afinado del conjunto de parámetros que definen una red neuronal artificial, para obtener el mejor desempeño posible de la misma en sus tareas de reconocimiento.
- d. Redes neurales artificiales y autómatas celulares. Estudio y aplicaciones (2008-2011): seguido de (a), además de seguir con su objetivo inicial, aquí se desarrollaron aplicaciones de software bajo convenios, para empresas de desarrollo de software (VATES S.A.), para reconocimiento de imágenes de preparados de laboratorio de la Facultad de Veterinaria de la UCC y para el Instituto de Investigaciones Geo-Históricas (IIGHI-CONICET) sobre la metodología para el cálculo del riesgo para la salud de las viviendas urbanas. Este último fue además transferido a centros de investigación en Argentina, Paraguay, Brasil, Cuba y Colombia.
- e. Transferencia del sistema de software RVS 2.1 para el cálculo de riesgo de la vivienda saludable en la Provincia de Córdoba (2008-2009): con subsidio del Gobierno de Córdoba, se transfirieron productos de software y metodología diagnóstica a la Municipalidad de Villa del Totoral, Córdoba, Argentina.
- f. El riesgo de la vivienda urbana para la salud desde la perspectiva de la salud comunitaria y la vulnerabilidad sociodemográfica en las ciudades Resistencia (Chaco) y Córdoba (2009-2010): este es un proyecto PICT del IIGHI-CONICET del cual se participó como parte del grupo responsable, para reforzar las transferencias realizadas.
- g. Modelado para la predicción de incendios forestales en la Provincia de Córdoba (2010-2012): utilizando datos históricos e índices específicos de incendios, se desarrolló un modelo y herramienta para predecir dónde y de qué envergadura podrían producirse incendios en la zona serrana de Córdoba.
- h. Aseguramiento de la trazabilidad en proyectos de desarrollo de sistemas de software (2010-2014): se estudió la aplicabilidad de los procesos de trazabilidad a los distintos tipos de sistemas desarrollados en distintos tipos de empresas. Se condujeron encuestas en empresas del medio para obtener datos históricos que, además de métodos estadísticos que fueron utilizados, alimentaron una red neural para efectuar predicciones sobre trazabilidad.
- i. Redes neurales artificiales y autómatas celulares. Productos y aplicaciones (2012-2014): seguido desde (d), este proyecto actualizó las herramientas de software construidas en anteriores proyectos y realizó difusión y divulgación, en conjunto con investigadores de IIGHI-CONICET. Se realizó una transferencia a la Municipalidad de Benevides en el estado de Belem-Pará, Brasil y se estudian nuevas transferencias.
- j. Construcción de un modelo de pronósticos para predicción de incendios forestales en la Provincia de Córdoba, (2013-2016): seguido de (g), se profundizó el estudio del modelo, se generaron herramientas para obtención de información histórica por Internet y se entrenaron nuevas redes neurales para efectuar predicciones.
- k. Desarrollo de un sistema de análisis de tex-

to no estructurado (2012-2013) / Fase II (2014-2015) / Fase III (2016-2017/18): estos proyectos (en relación con la tesis doctoral de uno de los integrantes de GA²LA), mediante técnicas de inteligencia artificial para el procesamiento de textos escritos en español e inglés, intenta determinar implicaciones lógicas entre distintas partes del mismo o entre distintos textos.

l. Evaluación del impacto de variantes no convencionales en el desempeño de autómatas finitos con memoria de pila (2015-2017/18): se definieron máquinas con memoria de pila con accesos no usuales, para determinar su efectividad en la resolución de distintos tipos de problemas. Se construyeron simuladores de estas máquinas que ayudaron en la experimentación.

m. Detección de errores sintácticos bajo el algoritmo de Earley (2014-2017/18): aquí se intenta determinar cómo se pueden informar adecuadamente (localización y tipo) sobre errores sintácticos de sentencias analizadas con el algoritmo de Earley. Se desarrollaron analizadores léxicos y sintácticos para experimentación y se estudia la relación entre el funcionamiento del algoritmo Earley y las redes de Petri, entre muchos otros aspectos.

n. Redes neurales artificiales con aprendizaje profundo (2018-2020): este es un nuevo proyecto iniciado muy recientemente, que se interesa primeramente por la parte teórica del aprendizaje profundo para establecer cómo se visualizan las características de distinto nivel que esta técnica extrae de los datos de entrenamiento. Por otro lado, trabajará con las herramientas disponibles y posiblemente genere nuevas, para actualizar las transferencias de redes neurales entrenadas efectuados en anteriores proyectos.

En todos estos proyectos, además de los logros específicos alcanzados según sus objetivos de investigación y desarrollo, se formaron numerosos recursos humanos de grado (ya sea mediante prácticas supervisadas de alumnos próximos a recibirse, como a becarios alumnos) y de posgrado (becarios graduados, especialidades, maestrías y doctorados).

Se efectuaron además numerosas transferencias tanto a centros de investigación, como a empresas o entes oficiales de gobierno.

Por otro lado, los integrantes de estos proyectos se desempeñan en numerosas cátedras de distintas carreras de la Facultad Córdoba de la UTN y de otras universidades (Nacional de Córdoba, Instituto Universitario Aeronáutico, Blas Pascal, Católica de Córdoba, Nacional de Villa María) y en carreras de posgrado de las mismas. Por ello, siempre parte de los proyectos contienen algún objetivo académico, por lo cual a las transferencias anteriores se deben agregar las transferencias efectuadas a las mismas cátedras de las cuales son parte los investigadores, mediante artículos, conferencias y seminarios.

Finalmente, todos los proyectos han realizado difusión de los conocimientos logrados en los más diversos congresos nacionales e internacionales de informática y otros relacionados, participación en la redacción de libros y capítulos de libros, y también se ha realizado divulgación de las investigaciones en medios periodísticos, charlas en escuelas de nivel medio y artículos en páginas web.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

De los proyectos indicados en la Introducción, se desprenden claramente al menos tres líneas de investigación y desarrollo interrelacionadas que el grupo GA²LA tendrá como ejes para su funcionamiento, a saber:

- Autómatas y Lenguajes Formales
- Procesamiento de Lenguaje Natural
- Aprendizaje Automático

Dentro del grupo propuesto, sus integrantes participan de entre muchas, de las cátedras de Sintaxis y Semántica de los Lenguajes, Investigación Operativa, Inteligencia Artificial, Algoritmos y Estructuras de Datos, Paradigmas de Programación, Simulación, Matemática Discreta y Análisis Matemático, por lo que las líneas se mezclan “adecuadamente” con sus inquietudes y trabajo académico habitual.

Además, han efectuado o tienen en marcha, posgrados relacionados con las mismas.

3. RESULTADOS OBTENIDOS / ESPERADOS

Los integrantes del grupo GA²LA en conjunto, como resultado de los proyectos realizados entre otras cosas, a la fecha han:

- Realizado más de 70 presentaciones y publicaciones en congresos nacionales e internacionales.
- Obtenido derechos de autor por productos de software en cuatro oportunidades.
- Escrito 11 libros o capítulos de libros.
- Participado en la organización de al menos siete congresos y workshops nacionales.
- Participado en la evaluación de: artículos en congresos, revistas y journals nacionales e internacionales, proyectos en el Ministerio de Ciencia y Tecnología de Córdoba, categorización de investigadores en UTN, competencias de programación y de desarrollo de software, etc.
- Participado de numerosas transferencias bajo convenio a centros de investigación, empresas y entes de gobierno local.
- Dirigido tesis de maestría, especialidad, numerosas prácticas supervisadas y becarios tanto de grado como graduados.

Se espera que trabajando organizados en un grupo formal, estos logros puedan no solo mantenerse en el tiempo, sino incrementarse.

4. FORMACIÓN DE RECURSOS HUMANOS

El grupo que se presenta está conformado por seis investigadores formados, seis investigadores noveles y alumnos de grado y graduados (como becarios) de la Universidad Tecnológica Nacional, Facultad Regional Córdoba.

Actualmente, en los proyectos en proceso, se tienen en formación:

- Dos doctorandos

- Dos maestrandos
- Un becario graduado
- Siete becarios de grado

Se piensa sumar en el corto plazo un nuevo doctorando y becarios de grado para los proyectos vigentes durante 2018.

5. BIBLIOGRAFÍA

1. Castillo, J., Cardenas, M., Medel, R., Casco, O., Navarro, M. & Gutierrez, S. "SISTEMA PARA PREDICCIÓN DE INCENDIOS APLICADO A LA PROVINCIA DE CÓRDOBA". (2016). XVII Workshop de Investigadores en Ciencias de la Computación, WICC 2016, Entre Ríos, Argentina.
2. Castillo, Julio J.; Cardenas, Marina E.; Curti, Adrián; Casco, Osvaldo. "Software para asistencia en la creación de corpus para sistemas de análisis de texto no estructurado". WICC 2015. XVII Workshop de Investigadores en Ciencias de la Computación (Salta, 2015).
3. Marina Cardenas, Juan C. Vázquez, Julio Castillo, Sebastián Villena Ruiz. "SISTEMA DE PREDICCIÓN DE INCENDIOS FORESTALES BASADO EN EL ÍNDICE FWI PARA LA PROVINCIA DE CÓRDOBA". XV Workshop de Investigadores en Ciencias de la Computación. WICC 2013. Paraná, Entre Ríos, Argentina. Mayo de 2013.
4. Juan C. Vázquez, Julio Javier Castillo, Marina E. Cardenas, Romina Gordillo, Sebastián Villena Ruiz. "Predicción de Incendios Forestales en la Provincia de Córdoba". Lugar: Posadas, Misiones, Argentina. Mayo de 2012. WICC 2012.
5. Juan C. Vázquez, Julio Javier Castillo, Marina E. Cardenas, María del Carmen Rojas. "Modelo Computacional empleando Redes Neuronales Artificiales para la estimación del Riesgo para la Salud de la Vivienda Urbana". WICC 2011 – XIII Workshop de Investigadores en Ciencias de la Computación. Mayo de 2011. Fecha: 5 Y 6 DE

MAYO DE 2011 Rosario, SANTA FE, ARGENTINA.

6. Juan C. Vázquez, Julio Castillo, María del C. Rojas, Marina Cardenas. “Modelización y Estimación del Riesgo para la Salud de la Vivienda Urbana empleando Redes Neuronales Artificiales”, WICC 2010, Mayo de 2010. ISBN: 978-950-34-0652-6. 5 Y 6 DE MAYO DE 2010 CALAFATE, SANTA CRUZ, ARGENTINA.
7. Vázquez Juan, Castillo Julio J., Rojas M. del Carmen, Marciszack Marcelo. Redes Neuronales Artificiales aplicadas a Ciencias Sociales. WICC 2008 - X Workshop de Investigadores de Ciencias de la Computación. Entidad Responsable: Red UNCI. Ciudad: General Pico, La Pampa, Argentina. Idioma: Español. ISBN: 978-950-863-101-5.
8. Una estrategia didáctica para el dictado de Informática Teórica; X Workshop de Investigadores en Ciencias de la Computación (WICC 2008); 2008; General Pico, La Pampa, Argentina., Autores: Marciszack M., Cárdenas M., Vázquez J., Castillo J., ISBN: 978-950-863-101-5 (CD)
9. Construcción de herramientas didácticas para la enseñanza y la ejercitación práctica en laboratorio de “Informática Teórica” en las carreras con Informática; VIII Workshop de Investigadores en Ciencias de la Computación (WICC 2006); Junio de 2006; Morón, Buenos Aires, Argentina. Autores: Marciszack M., Vázquez J., Serrano D., Pérez R., ISBN: 950-9474-35-5, ISBN13: 978-950-9474-35-2 (CD), ISBN: 950-9474-34-7, ISBN13: 978-950-9474-34-5 (Impreso)
10. Detección de errores sintácticos bajo el algoritmo Earley; XVII Workshop de Investigadores en Ciencias de la Computación (WICC 2015); Abril de 2015; Salta, Salta, Argentina. Autores: Vázquez J., Constable L., Brenda M., Jornet W., Arcidiácono M., Parisi G., ISBN: 978-987-633-134-0 – Red UNCI/UNSA
11. Tratamiento de Imágenes de Fibras Animales; Constable L., Arcidiacono M., Vázquez J., Picco J.; XV Workshop de Investigadores en Ciencias de la Computación (WICC 2013); Red UNCI, Universidad Autónoma de Entre Ríos, Facultad de Ciencia y Tecnología; Paraná, Entre Ríos, Argentina; 2013. ISBN: 978-987-28179-6-1 (CD)
12. Verificación estática de confidencialidad en un sistema de múltiples niveles de seguridad basado en Java bytecodes; Medel R., Spessot C., Vázquez J., Giagante I.; XI Workshop de Investigadores en Ciencias de la Computación (WICC 2009), Universidad Nacional de San Juan; San Juan, San Juan, Argentina; 2009. ISBN: 978-95060-557-0-7 (CD)
13. Redes Neuronales Artificiales aplicadas a Ciencias Sociales; Vázquez J., Castillo J.; X Workshop de Investigadores en Ciencias de la Computación (WICC 2008), Universidad Nacional de La Pampa; General Pico, La Pampa, Argentina. Mayo de 2008. ISBN: 978-950-863-101-5 (CD)
14. Redes Neuronales Artificiales vs. Autómatas Celulares; Martínez F., Vázquez J., Marciszack M.; VII Workshop de Investigadores en Ciencias de la Computación (WICC 2005); Río Cuarto, Córdoba, Argentina. Mayo de 2005. ISBN: 950-665-337-2 (CD)