

Redes Neuronales Artificiales aplicadas a Ciencias Sociales

Ing. Juan C. Vázquez¹, Sr. Julio Castillo¹, Dra. María del C. Rojas², Ing. Marcelo Marciszack¹
jcvazquez@acm.org, jotacastillo@gmail.com, rojas_herrera@arnet.com.ar, marciszack@gmail.com

¹ Universidad Tecnológica Nacional – Facultad Córdoba – Dpto. Ing. en Sistemas de Información
Maestro Marcelo López y Cruz Roja Argentina, Ciudad Universitaria, 5016, Córdoba, Argentina.

² Instituto de Investigaciones Geohistóricas (IIGHI-CONICET)
Av. Castelli 930, Resistencia, 3500, Chaco, Argentina.

Resumen: Como modelo computacional que implementa la metodología de estimación del riesgo para la salud de la vivienda urbana desde el enfoque de la vulnerabilidad social, propuesta por el IIGHI-CONICET, el software desarrollado de redes neuronales artificiales emula una red conceptual de relación de factores sociales, económicos y demográficos que no ajusta a los modelos clásicos. Los demógrafos del IIGHI suponen el dominio de problema como complejo y no lineal; indican además que aún no existen mediciones precisas de causa y efecto para el mismo, por lo que su tratamiento por ecuaciones matemáticas no es posible. Se han utilizado cinco redes de tipo perceptrón multicapa trabajando en forma conjunta para calcular un índice de riesgo que permitirá a las autoridades de salud dirigir más acertadamente los fondos disponibles, en la medida que la metodología sea confirmada por trabajo de campo, tarea que está en marcha en algunos países de América Latina.

Palabras clave: Redes Neuronales; riesgo en salud; sistema complejo; sistema no lineal.

1. Introducción

Las autoridades gubernamentales ocupadas en temas de salud, tienen siempre un presupuesto acotado para una problemática sencilla y compleja, por lo que la decisión de dónde, cuándo y cuántos fondos asignar no es una tarea trivial y es de gran importancia para el bienestar de la población que sea acertadamente llevada a cabo.

Un grupo de investigadores del Instituto de Investigaciones Geohistóricas (IIGHI) del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), bajo la dirección de la Dra. Norma Meichtry y la Dra. María del Carmen Rojas, ha desarrollado una metodología para la estimación del riesgo de la vivienda urbana para la salud humana (Rojas, 2004 y 2006); utilizando un enfoque holístico, se estudian los factores que determinan la vulnerabilidad social, considerando a la vivienda como un todo con su entorno, los servicios públicos disponibles, los aspectos sociales y económicos de sus residentes, la capacidad de reacción frente a emergencias, etc.; los factores involucrados en la determinación del riesgo se han seleccionado premeditadamente de aquellos que pueden encontrarse usualmente en los censos de población, con lo cual el cálculo puede realizarse hacia atrás en el tiempo para estudiar evolución y verificar el ajuste del modelo frente a estadísticas de salud existentes.

Como es usual en las Ciencias Sociales, es prácticamente imposible efectuar mediciones del impacto sobre la salud de un determinado factor de riesgo en presencia de todas las otras variables constantes (¡no se puede parar el mundo!), por lo que es difícil atacar el modelado del dominio con herramientas matemáticas (por ejemplo, ecuaciones diferenciales o algebraicas); por otro lado, si bien la metodología propone la incidencia de los factores tenidos en cuenta sobre conceptos demográficos no mensurables como resiliencia, exposición y fragilidad social (Rojas, 2006), los expertos del IIGHI-CONICET estiman que estas relaciones son en general complejas y no lineales, aunque por su experiencia pueden armar “ejemplos” de valores de los factores y los resultados esperados que los mismos tienen sobre el riesgo de salud.

La Dra. Rojas llegó al Laboratorio de Investigación de Software del Departamento de Ingeniería en Sistemas de Información (Facultad Córdoba – UTN – Argentina) con estas ideas, solicitando un modelo computacional para la adquisición de datos y el cálculo del riesgo según la metodología desarrollada. Allí, trabajamos desde hace algunos años con proyectos de redes neuronales y autómatas celulares, intentando encontrar semejanzas entre estos dos modelos al pensar en las redes neuronales artificiales como sistemas evolutivos.

El presente artículo describe brevemente la problemática que presenta el método del IIGHI-CONICET desde el punto de vista informático/computacional y la solución que pudimos ofrecer desde el software; además, se comentan las posibles mejoras al esquema desarrollado y los próximos pasos previstos en ese sentido.

2. Metodología de estimación del riesgo para la salud

La metodología pertenece al campo del conocimiento y a la práctica de la Medicina Social en América Latina y reconoce como fuente a la tesis doctoral sobre *Estimación holística del riesgo sísmico utilizando sistemas dinámicos complejos*¹, desarrollada por Omar Darío Cardona Arboleda.

Actualmente constituye uno de los elementos básicos para el desarrollo de la línea de investigación sobre vigilancia ambiental que se está trabajando conjuntamente con investigadores de Paraguay, Brasil, Cuba y Venezuela en el marco de la Red Interamericana de Vivienda Saludable de la Organización Panamericana de la Salud (OPS).

Pensando en espacios de intervención donde poder crear herramientas sistemáticas de valoración de procesos protectores y deteriorantes de las condiciones de salud-enfermedad, se presenta como un interesante ámbito de trabajo el estudio y la valoración de la vivienda y su influencia en la salud de sus residentes.

La idea básica surge de la necesidad de fortalecer los sistemas nacionales y locales de vigilancia de los factores de riesgo y protección a la salud asociados con la vivienda, a través del diseño de nuevos modelos que contribuyan a generar alternativas para el desarrollo saludable y sustentable regional latinoamericano.

Se intenta mutar el concepto de factores linealmente relacionados con el riesgo en salud, por el de proceso dinámico; se podría decir, que los modos de devenir que determinan la salud se desarrollan mediante un conjunto de procesos y estos procesos adquieren proyecciones distintas frente a la salud, de acuerdo a los condicionantes sociales de cada espacio y tiempo, es decir de acuerdo a las relaciones sociales en que se desarrollan.

Consecuentemente, es necesario aproximarse a la vivienda no como un reservorio estático de contaminantes, parásitos, vectores de transmisión de enfermedades infecciosas, sino como un espacio históricamente estructurado donde también se expresan las consecuencias benéficas y destructivas de la organización social, donde los procesos del espacio construido llegan a ser mediadores necesarios y donde se transforman las condiciones de reproducción social dominantes en bienes o soportes que favorecen la salud, o en fuerzas destructivas que promueven la enfermedad.

Los investigadores del IIGHI-CONICET han propuesto un modelo que trabaja con información de censos de población para la evaluación sociodemográfica y planillas de recuento de datos de gobiernos locales, para la valoración de la resiliencia. Se procura establecer una tipología de

¹ Omar Darío Cardona Arboleda es ingeniero de la Universidad Nacional de Colombia. Fue distinguido por el premio Sasakama de Prevención de desastres de Naciones Unidas.

viviendas urbanas en relación con la salud humana, con validez para los diferentes niveles sociales, pero su aplicación concreta inmediata, se referirá a las viviendas urbanas marginales. Se presentan los componentes de la amenaza y la vulnerabilidad a partir de las variables de la vivienda que constituyen la amenaza para la salud y variables sociodemográficas que conforman la vulnerabilidad.

EL modelo de determinación del índice de riesgo para la salud que implica la vivienda urbana se muestra en la figura 1.

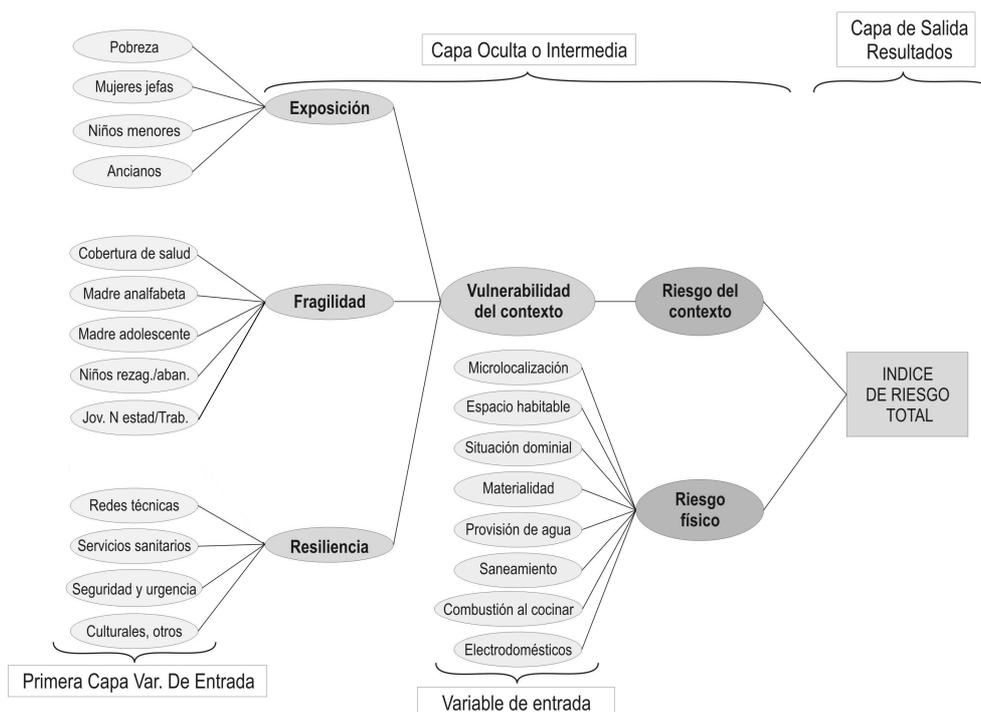


Figura 1 – Modelo conceptual de estimación del riesgo propuesto por IIGHI-CONICET.

Como puede verse, los demógrafos han pensado el esquema teniendo en mente la teoría de redes neuronales artificiales (por la nomenclatura de capas empleada) pero sin ajustarse a un modelo conocido; se incluyen por ejemplo, entradas intermedias dentro de la denominada capa oculta para el cálculo del riesgo físico y se omiten capas intermedias reales necesarias para el cálculo de sub-conceptos.

3. Modelo computacional desarrollado

3.1. Supuestos.

No siendo demógrafos sino ingenieros en sistemas de información, hicimos las preguntas de rigor sobre la independencia de los factores propuestos como entrada y sobre el tipo de valores que se podían esperar recibir a los mismos, sin cuestionar la validez del modelo conceptual planteado.

Si bien al parecer los factores no son absolutamente independientes, los expertos del IIGHI consideran que la influencia mutua que pudiera existir se encuentra reflejada en las interrelaciones planteadas en el modelo conceptual.

Por otro lado, los valores de los factores tomados como entrada siempre se dan como porcentajes, ya que sobre un total de habitantes de una región, se informan cantidad de hogares que detentan el

cumplimiento de determinado factor, por lo que trabajamos en todos los casos con porcentajes (enteros de cero a cien).

Para los conceptos intermedios y finales no mensurables (resiliencia, exposición, fragilidad, vulnerabilidad, riesgo físico y riesgo total), los expertos indicaron que podrían aproximar una estimación porcentual basada en su experiencia, según los valores de factores mensurables tabulados de cero a cien. Sin embargo, esta estimación sería *difusa*, ya que proporcionaron *intervalos* en vez valores individuales, por lo que se procedió a establecer una escala de conjuntos borrosos triangulares, basados en los datos suministrados.

3.2. Diseño del modelo computacional.

Ante la imposibilidad de establecer un modelo matemático mediante el uso de ecuaciones algebraicas o diferenciales, ya que se desconoce la relación formal existente entre los factores y los subconceptos y entre los subconceptos y el índice total, se pensó en modelar computacionalmente el modelo conceptual demográfico con redes neuronales artificiales (RNA).

Esta decisión se basó también en el formato del modelo propuesto por el IIGHI-CONICET y por la complejidad y no linealidad declarada por sus investigadores, del dominio de problema.

En este sentido se pensaron dos alternativas:

- Una sola red neuronal en la que todos los factores mensurables fueran puestos al mismo nivel, a modo capa de entrada, con una sola salida compuesta por el riesgo total (figura 2).
- Un conjunto de varias redes neuronales que calculasen en forma independiente cada subconcepto y finalmente el índice de riesgo total, cada una con sus capas intermedias. (figura 3)

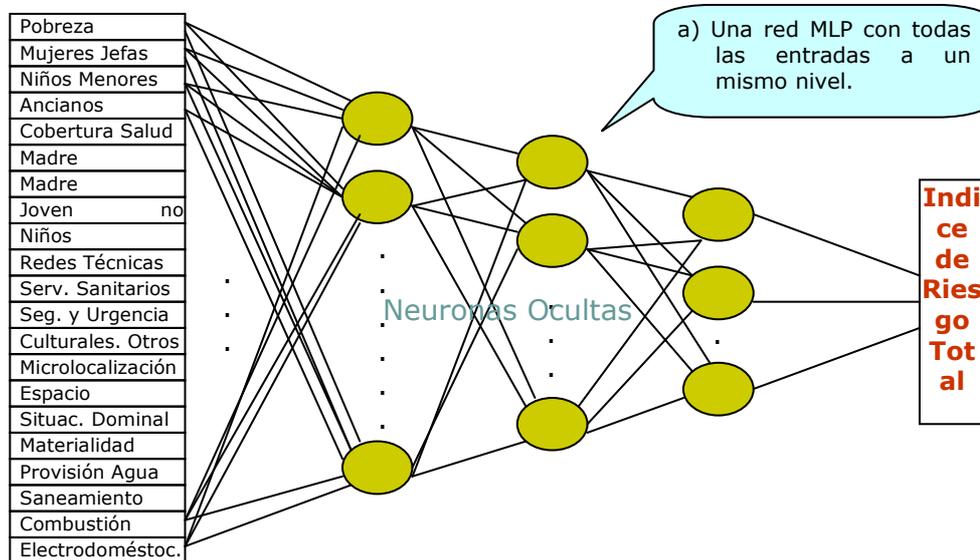


Figura 2 – Primera propuesta para el modelo computacional.

La primera alternativa fue desechada ya que se perdían los conceptos intermedios (resiliencia, exposición, fragilidad y vulnerabilidad) que son considerados de importancia por los investigadores de IIGHI y que podrían ser utilizados para otras investigaciones posteriores.

Por otro lado, los expertos tenían ya dificultades para estimar los valores de los subconceptos dados uno conjunto de valores de prueba para los factores de entrada, por lo que la estimación del índice

de riesgo total teniendo que considerar todos los factores sociodemográficos al mismo tiempo se complicaba o al menos era muy poco confiable.

La segunda alternativa (figura 3) resultó ser mejor acogida ya que:

- Los conceptos intermedios serían calculados para su posterior tabulación y utilización en otras investigaciones.
- El criterio de los expertos resultaba más claro teniendo que hacer estimaciones de problemas más acotados; estas estimaciones serían de vital importancia para el entrenamiento de las redes neuronales ya que proporcionarían los *ejemplos* necesarios para el método de retro-propagación de errores.
- El modelo computacional se asemeja más al modelo conceptual del IIGHI.

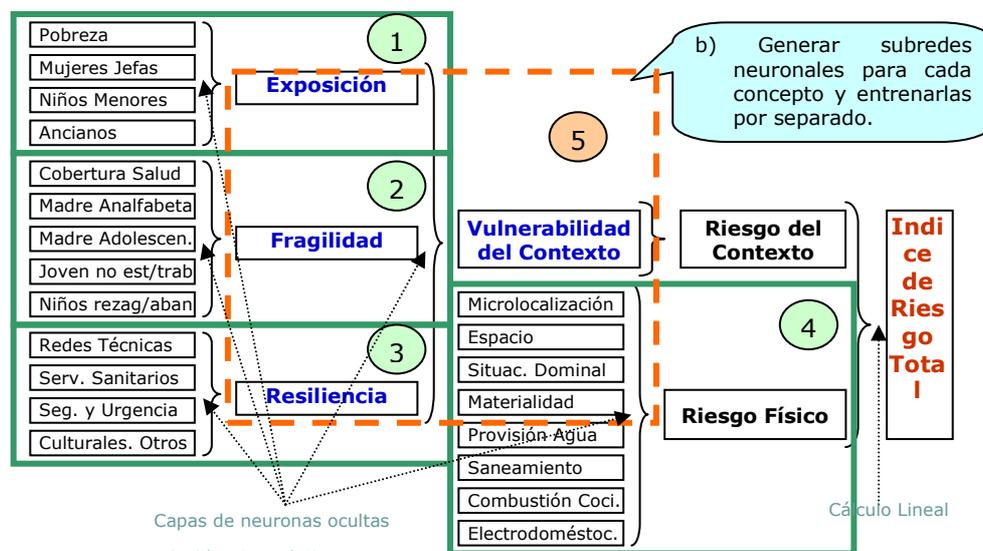


Figura 3 – Segunda propuesta para el modelo computacional.

Se diseñaron así, finalmente, cinco redes neuronales artificiales tipo perceptron multicapa que se entrenarían con el algoritmo de retropropagación de errores; el cálculo final del índice de riesgo se haría en forma lineal según ponderaciones indicadas por los expertos.

Las redes neuronales se diseñaron como perceptrón multicapa puro, con neuronas de capa oculta gobernadas por tangente hiperbólica y neuronas de salida lineales; las neuronas de la capa de entrada sólo normalizarían los valores de ingreso para que pertenezcan al rango real [0,1].

La construcción fue emprendida por parte del equipo de trabajo del proyecto RNA-AC – 25/E078 – (Martínez, Vázquez & Marciszack, 2005) de la Facultad Córdoba de la UTN; por la experiencia del equipo y por estar disponible en el Laboratorio de Investigación de Software, se optó por el lenguaje C# de Microsoft para la implementación y la orientación a objetos para el diseño de los programas.

3.3. Módulos construidos en C#.

En principio, se construyeron dos módulos bien diferenciados:

1. **Módulo de entrenamiento:** este módulo configura las redes neuronales en base a archivos de configuración textuales, que permiten definir la cantidad de capas y la cantidad de neuronas por capa para cada subred. Se fijan además por programación, el error medio cuadrático aceptado como razonable para terminar los entrenamientos. Al finalizar, el

módulo crea un archivo descriptivo de la topología neuronal utilizada, el tipo de subred, la fecha de entrenamiento, la cantidad de ejemplos tomados y el valor de los pesos sinápticos obtenidos como resultado del entrenamiento.

2. **Módulo de producción:** este módulo (RVS 2.1) es el que se dispone para transferencia a distintos centros de investigación de estudios de población y salud comunitaria. Leyendo el archivo descriptivo generado por el módulo de entrenamiento, configura la **red entrenada** en memoria y permite el ingreso de valores reales de poblaciones; efectúa el cálculo por propagación sobre la red de cada subconcepto y del índice total, informando en cada caso los porcentajes de participación según los conjuntos borrosos definidos. Genera además informes de valores exportables a planillas Excel para su registro y posterior estudio.

Por razones de espacio, no mostramos aquí las distintas pantallas que demuestran el funcionamiento de estos módulos, pero los mismos pueden consultarse en el informe final del proyecto 25/E078, disponible en la Secretaría de Ciencia, Técnica e Innovación Productiva (hoy Ministerio) de la República Argentina o en la Secretaría de Ciencia y Técnica del Rectorado de la Universidad Tecnológica Nacional en Buenos Aires, Argentina.

4. Pruebas, discusión y difusión.

El software desarrollado está siendo probado por los investigadores en demografía del IIGHI-CONICET utilizando datos reales censados en regiones de las ciudades de Córdoba y Resistencia de Argentina, con resultados favorables hasta el momento; en las pruebas iniciales se detectó un problema con el tratamiento de valores nulos de entrada que fue reparado en el software; las neuronas de la red perceptron no aprendían su tarea adecuadamente cuando propagaban sus entradas en cero, por lo que procedió a desfazar los valores en 0,01 en caso de valores nulos sin pérdida de las cualidades del modelo conceptual.

El software y la metodología base han sido presentado en distintos foros:

- Congresos de la Asociación de Estudios de Población de Argentina.
- Seminario Iberoamericano de Hábitat Popular, Centro de Estudios Avanzados de la Universidad Nacional de Córdoba.
- II Congreso de la Asociación Latinoamericana de Estudios de Población, Guadalajara, México.
- Primer Congreso Internacional Mexicano de Vivienda Saludable, Pachuca de Soto Hidalgo, México.
- ... y otros.

Por otro lado, la Organización Panamericana de la Salud ha manifestado su interés en efectuar pruebas de metodología y programas en su red latinoamericana de vivienda saludable; la primer transferencia será realizada a la Facultad de Arquitectura y Urbanismo de la Universidad Católica de Nuestra Señora de la Asunción, en Paraguay (para lo cual el software ha sido registrado con titularidad conjunta de CONICET y UTN-Facultad Córdoba) y están previstos similares convenios con Universidades de Brasil, Cuba y Venezuela.

En los casos estudiados, los resultados del software coinciden con la opinión de los expertos en cuanto al riesgo previsible para la salud. Sin embargo es temprano aún para declarar éxito; la metodología debe verificar estadísticamente sus previsiones, el software debe ser llevado a situaciones límites para comprobar su corrección, incluso los ejemplos propuestos por los demógrafos están bajo revisión permanente.

