



Redes neuronales para el tratamiento de agua potable en zona de altitud del Perú*

Neural Networks for Drinking Water Treatment in the Altitude Zone of Peru

Réseaux de neurones pour le traitement de l'eau potable dans la zone d'altitude du Pérou

Anieval Peña-Rojas**, Lisveth Flores-del-Pino***

Recibido: 2014-02-30 // Aprobado: 2014-04-30 // Disponible en línea: 2014-09-10

Cómo citar este artículo: Peña-Rojas, A. & Flores-del-Pino, L. (2014). Redes neuronales para el tratamiento de agua potable en zona de altitud del Perú. *Ambiente y Desarrollo*, 18(35), 109-116. <http://dx.doi.org/10.11144/Javeriana.AyD18-35.rnta>
doi:10.11144/Javeriana.AyD18-35.rnta

Resumen

Se evaluó la dosificación de coagulantes utilizando redes neuronales artificiales, en una planta de agua potable ubicada en Perú (12° 00' 00" S y 75° 11' 00" W) a 3420m, donde la calidad al ingreso es inestable debido a precipitaciones fluviales que alteran sus propiedades fisicoquímicas. Se utilizó la "prueba de jarras" para evaluar la dosis de coagulante (Al_2SO_4), considerando turbiedad, pH, conductividad y sólidos disueltos, con temperatura y caudal promedio de 10° C y 148 L/s. Los datos de once meses fueron usadas para entrenar diversas tipologías de redes, siendo la de regresión generalizada la que presentó un rendimiento de 97,77% y evidenció que los factores más influyentes son los sólidos disueltos, la turbidez, la conductividad y el pH, con 42,79, 32,43, 10,33 y 7,26%, respectivamente.

Palabras clave: red neuronal; agua; optimización

* Este artículo se deriva del proyecto de investigación doctoral *Aplicación de redes neuronales para la optimización y control de la dosificación de coagulantes en el tratamiento de agua de calidad inestable*, del Doctorado en Ingeniería Ambiental de la Universidad Nacional Agraria La Molina, de Lima-Perú.

** Ingeniero químico de la Universidad Nacional del Centro del Perú, Magíster en Minería y Medio Ambiente de la Universidad Nacional Daniel A. Carrión. Correo electrónico: anieval@hotmail.com

*** Ingeniera química de la Universidad Nacional Mayor de San Marcos, Ph.D. en Michigan State University, Estados Unidos. Correo electrónico: lisveth@lamolina.edu.pe

Abstract

We assessed the dosing of coagulants using artificial neural networks in a drinking water plant located in Peru (12°00'00" S and 75°11'00" W), 3,420 m.a.s.l., where the quality upon intake is unstable due to fluvial rainfall that alters its physical and chemical properties. We used the "jar test" to assess the coagulant dose (Al_2SO_4) taking into account turbidity, pH, conductivity and dissolved solids with an average temperature and flow rate of 10 °C and 148 Lt/s. We used data from 11 months to train different network typologies, with the generalized regression one recording a performance of 97.77% and showed that the factors with the biggest influence are dissolved solids, turbidity, conductivity and pH, with 42.79%, 32.43%, 10.33% and 7.26% respectively.

Keywords: neural network; water; optimization

Résumé

On a évalué la posologie de coagulants en utilisant les réseaux de neurones artificielles en une plante d'eau potable situé au Pérou (12°00'00" S y 75°11'00" W) a 3420 m.s.n.m. où la qualité à l'entrée est instable pour les précipitations fluviales qui altèrent ses propriétés physicochimiques. On a utilisé le « test de jarras » pour évaluer la dose de coagulant (Al_2SO_4) en considérant l'eau turbide, pH, conductivité et solides dissoudre avec une température et courant d'environ de 10 °C y 148 L/s. Les données de 11 mois ont été utilisées pour dresser divers typologies de réseaux, étant-donné la de régression généralisée qui a présenté un rendement de 97.77% et a montré que les facteurs les plus présents sont les solides dissous, turbidité, conductivité et pH avec 42.79%, 32.43%, 10.33% y 7.26% respectivement.

Mots clés: réseaux neuronale; eau; optimisation.

Introducción

El agua potable para consumo humano es uno de los elementos más importantes para la vida, en la mayoría de poblaciones del Perú esta es tratada a partir de efluentes lénticos (ríos o riachuelos).

Un problema recurrente en las ciudades de la sierra peruana (sobre todo durante el invierno) es la alta concentración de sólidos disueltos y no disueltos en el agua que va a tratarse, así como la alteración del pH y de la conductividad, debido a las precipitaciones intempestivas que se producen.

El proceso de coagulación es una de los más importantes cuando se trata de remover estos sólidos del agua —que se presentan también como partículas coloidales—, por ser difícil el suministro de la dosis de coagulante en función de los parámetros físicos y químicos de calidad de agua cruda; además, se ve influido por factores tales como: el corto tiempo de reacción del proceso de coagulación, el largo tiempo de reacción del proceso de floculación y el caudal de agua que ingresa al tratamiento. Todo esto hace que el control de este proceso requiera técnicas de análisis químico como la prueba de jarras en la estimación de la dosis óptima, así como la acción de un operador que supervise el proceso y haga los ajustes necesarios para garantizar una remoción de turbiedad efectiva, lo cual lleva tiempo de ejecución y permite que en ese intervalo se genere agua de mala calidad. La calidad del agua está determinada por límites permisibles en parámetros físicos y químicos (tabla 1).

Tabla 1. Límites permisibles de factores estudiados para el diseño de la red neuronal

<i>Parámetro</i>	<i>Unidad</i>	<i>LMP (*)</i>	<i>Fuente</i>
Turbiedad	UNT	5	1
Ph	Unidad	6,5-8,5	1
Conductividad - 25° C	uS/cm	1500	1
Color-platino/cobalto	UCV	20	1
Total de solidos disueltos (TDS)	mg/L (ppm)	500	2

(*) Límite máximo permisible

Fuente: 1.) Legislación SUNASS; 2.) Environmental Protection Agency (EPA).

Redes neuronales artificiales en aplicaciones industriales

La aplicación de la inteligencia artificial al control de procesos es cada vez mayor, ejemplo de ello son: el mejoramiento en el control de los procesos mediante la introducción de técnicas tales como los sistemas expertos, redes neuronales, algoritmos genéticos y lógica difusa. Todo esto conlleva oportunidades para el desarrollo de soluciones óptimas, en procesos cuyo control mediante el uso de técnicas convencionales resulta muy difícil.

A este respecto se han desarrollado investigaciones de aplicación a la evaluación de la calidad del agua subterránea, considerando diversos parámetros fisicoquímicos (Kuo et ál., 2003) y prediciendo la distribución de nitrato en aguas subterráneas (Wang et ál., 2006; Yesilnacar et ál., 2007), así como en hidráulica, prediciendo porosidades, permeabilidades y conductividades hidráulicas en suelos (Helle, et ál., 2001; García y Shigidi, 2006), además de fluctuaciones del nivel superficial de lagos (Altunkaynak, 2006).

Desde el punto de vista de la ingeniería, las redes neuronales artificiales (RNA) pueden ser vistas como un gran sistema dinámico en paralelo, que puede realizar transformaciones por medio de la

respuesta de estado de su información de entrada. Cómo se lleva a cabo la transformación, depende del modelo de la RNA y de su forma de aprehender la transformación. El área de aplicación más natural de las RNA son obviamente las tareas en las cuales se deben obtener las transformaciones apropiadas de ciertas entradas a ciertas salidas, pero las transformaciones no pueden ser descubiertas analíticamente, debido a la variedad de factores involucrados en ellas.

No es sorprendente que la mayoría de las aplicaciones exitosas de RNA puedan ser encontradas en áreas de visión de máquinas, reconocimiento de patrones, control de motores, procesamiento de señales, donde tales transformaciones “entradas a salidas” dominan la solución del problema (Hinton, 1992).

La red neuronal deberá predecir la dosis adecuada de sulfato de aluminio, dependiendo de algunas de las propiedades del agua cruda; por lo tanto, se utiliza el programa Neural Tools versión 6.0, que maneja algoritmos de eficiencia demostrada.

Una red neuronal es un sistema que imita al cerebro humano en la transformación de entradas en una o más salidas, con un conjunto de neuronas. Para encontrar el modelo más eficiente, en este estudio se utilizaron las redes multinivel alimentadas hacia adelante (MLFN) y las redes neuronales probabilísticas (PN). De modo específico, la función tangente hiperbólica se usa como función de activación en las neuronas de la capa oculta para la red MLFN.

La neurona de salida emplea la identidad como función de activación, la cual simplemente devuelve la suma ponderada de sus entradas. Las redes PN fueron propuestas primero por Specht (1991) y aproximan cualquier función arbitraria entre los vectores de entrada y salida con el trazo de la función estimada, de manera directa desde los datos de entrenamiento (Specht, 1991). El entrenamiento de la red busca identificar el valor del conjunto de pesos o sinapsis que, minimizando el error de predicción, logra reproducir las salidas incluidas en el conjunto de datos de entrenamiento.

Neural Tools respalda diferentes configuraciones de red neuronal para obtener las mejores predicciones posibles. Para la predicción de clasificaciones /categorías (en las cuales la variable dependiente es de tipo categoría), se ofrecen dos tipos de redes: PN y MLFN. La predicción numérica se puede realizar usando MLFN, así como redes neuronales de regresión generalizada (GRN), que están estrechamente relacionadas con las PN.

Materiales y métodos

Se utilizaron datos de las pruebas de jarras realizadas en el laboratorio de aguas de la planta referida para los parámetros del agua de entrada, tales como: pH, turbiedad (NTU), color (UCV), conductividad y total de sólidos disueltos (TDS). Cabe señalar que la temperatura promedio del agua en las operaciones fue de 10 °C y el promedio del caudal de ingreso 148 L/s. Estos datos se preprocesaron a diario durante el periodo de julio del 2012 a mayo del 2013 (tabla 2). El río Shullcas es el proveedor del agua cruda.

Se desarrollaron pruebas de optimización de aplicación de coagulantes, para lo cual se aplicaron dosis diferentes de coagulante en el laboratorio, en los diferentes tipos de muestras de agua de la planta, controlando turbiedad, pH, color, conductividad y TDS. La prueba de jarras fue la elegida como alternativa principal, y los resultados obtenidos se derivaron al proceso de entrenamiento de la red neuronal.

El programa usa un algoritmo de aprendizaje supervisado, el cual necesita un conjunto de datos, donde en cada grupo de entrada existe una salida esperada, para así poder aproximar una función que los relacione. A este conjunto de datos se le da el nombre de datos de entrenamiento. En este caso en particular, los datos de entrenamiento provendrán de los resultados de pruebas de jarras que se llevarán a cabo, así como de las dosis históricas registradas en pruebas de laboratorio de julio del 2012 a mayo del 2013. Las propiedades físico-químicas del agua cruda seleccionadas como entrada son: turbiedad, pH y conductividad, por su importancia en el proceso de la cinética de coagulación. Naturalmente, la variable de salida es la dosis óptima de sulfato de aluminio (mg/l).

Tabla 2. Análisis de los parámetros para las pruebas de jarras realizadas

Fecha	pH		Turbiedad (NTU)				Color (UCV)		Conductividad ($\mu\text{S}/\text{cm}$)		TDS (ppm)	
	Ingreso	Salida	Rio Shullcas	Agua decantada	Salida	Ingreso	Salida	Ingreso	Salida	Ingreso	Salida	
Julio	7,76	7,38	1,93	1,92	0,57	13	0	296	301	148	150	
Agosto	7,75	7,36	2,14	1,88	0,76	17	0	285	283	143	139	
Septiembre	7,7	7,32	12,43	6,67	0,69	66	1	251	255	126	128	
Octubre	7,81	7,35	12,41	7,01	0,72	70	1	250	256	125	128	
Noviembre	7,74	7,35	19,22	8,68	0,92	78	2	264	267	136	137	
Diciembre	7,94	7,48	48,22	14	8,34	103	3	212	218	107	108	
Enero	8,04	7,33	27,09	13,04	2,83	117	5	192	203	96	101	
Febrero	7,75	7,17	29,51	12,5	2,47	134	9	170	172	85	87	
Marzo	7,68	7,18	19,05	11	2,31	134	9	165	171	82	85	
Abril	7,98	7,35	17,25	10,9	2,86	110	8	173	178	86	89	
Mayo	8,05	7,59	4,75	7,84	1,08	35	1	228	233	114	116	

Los datos se procesaron en hojas de cálculo y el software usado para el entrenamiento de las RNA fue NeuroTools versión 6.0 de PalisadeSM
Fuente: elaboración propia

Resultados

El programa Neural Tools facilitó la selección de una configuración de red, al ofrecer una búsqueda de mejor red, opción en la cual este *software* entrena y prueba todas las configuraciones de red marcadas, considerando el porcentaje de errores y el índice de determinación (variables estadísticas que nos hicieron elegir el mejor modelo). Para el estudio se incluyeron las redes GRN, PN y las redes MLFN, con cuentas de distintos nodos en la capa oculta. Así se logró disponer de modelos de los cuales Neural Tools selecciona el mejor.

Los tres mejores modelos de acuerdo con el error cuadrático generado y el porcentaje de predicciones incorrectas se exponen en la tabla 3, y la selección del mejor modelo se efectuó con base en los datos diarios desde el mes de julio del 2012 a mayo del 2013. La mejor red resultó ser la de regresión generalizada (GRN), cuya arquitectura se presenta en la figura 1.

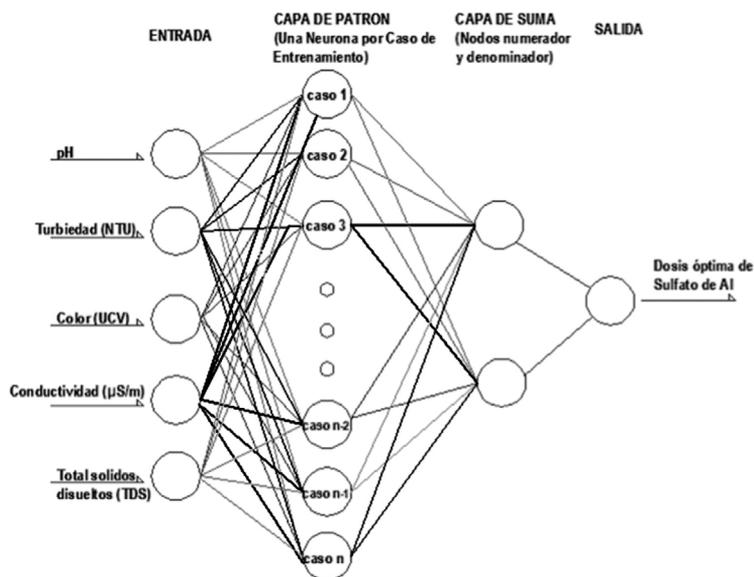


Figura 1. Arquitectura de la red neuronal de regresión generalizada (GRN)
Fuente: elaboración propia

Tabla 3. Los mejores modelos de acuerdo con el error cuadrático

Características	Con cinco nodos capa oculta	Con seis nodos capa oculta	Con predicción GRN
Nombre	Red entrenada en datos para red Sedam (2)	Red entrenada en datos para red Sedam (3)	Red entrenada en datos para red Sedam (4)
Configuración	Predicción numérica MLFN (cinco nodos)	Predicción numérica MLFN (seis nodos)	Predicción numérica GRN
Variable de categoría independiente	0	0	0

<i>Características</i>	<i>Con cinco nodos capa oculta</i>	<i>Con seis nodos capa oculta</i>	<i>Con predicción GRN</i>
Variables numéricas independientes	5 (agua decantada NTU, Ingreso Color, Ingreso pH, Ingreso Conduct, Ingreso TDS)	5 (Agua decantada NTU, Ingreso Color, Ingreso pH, Ingreso Conduct, Ingreso TDS)	5 (Agua decantada NTU, Ingreso Color, Ingreso pH, Ingreso Conduct, Ingreso TDS)
Variable dependiente	Var. numérica (Sulf. al PPM)	Var. numérica (Sulf. al PPM)	Var. numérica (Sulf. al PPM)
Entrenando	Var. numérica (Sulf. al PPM)	Var. numérica (Sulf. al PPM)	Var. numérica (Sulf. al PPM)
Número de casos	268	268	268
Tiempo de entrenamiento	01:14:12	01:42:00	00:00:02
Número de pruebas	4049256	5000931	90
Razón de la parada	Parado por el usuario	Auto-parada	Auto-parada
Porcentaje de predicciones incorrectas	5,22	5,60	2,24
Error cuadrático medio	2244	2147	1065

Fuente: elaboración propia

En la figura 2 se ilustran los factores de mayor preponderancia que influyen en la red neuronal.

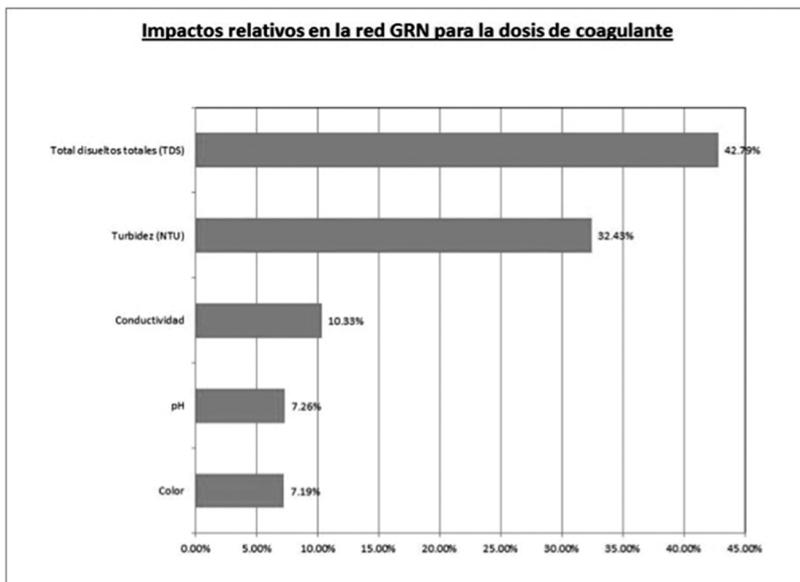


Figura 2. Factores que influyen en la red neuronal

Fuente: elaboración propia

Conclusiones

La red neuronal de regresión generalizada es la que tiene mejor rendimiento, con un porcentaje de error de 2,24, es decir, con una asertividad de 97,77% en la fase de entrenamiento para una adecuada respuesta de dosificación del coagulante. Los modelos de red neuronal pueden probarse constantemente y entrenarse de manera continua, con mayores datos históricos, en tanto que el modelo configurado permite hacer predicciones para nuevas entradas de las variables independientes, prescindiendo en momentos críticos de la prueba de jarras.

La red neuronal artificial GRN permite tomar decisiones rápidas de dosificación ante cambios inesperados de turbiedad o pH que se presentan cuando el afluente se altera, debido a las precipitaciones fluviales intempestivas que caen en la zona. Los factores más influyentes en la dosificación, de acuerdo con el modelo GRN, son los sólidos disueltos (42,79%), la turbidez (32,43%), la conductividad (10,33%), el pH (7,26%) y el color (7,19%). Estos factores influyen, según su variabilidad y la desviación estándar de sus medidas, en el ingreso a la planta de agua.

La temperatura es un factor importante en el tratamiento de coagulación y por las condiciones de altitud de la zona de estudio es de 10° C en promedio anual.

Referencias

- Altunkaynak, A. (2006). Forecasting Surface Water Level Fluctuations of Lake Van by Artificial Neural Networks. *Water Resource Manage*, 21, 399-408.
- García, I. A. & Shigidi, A. (2006). Using Neural Networks for Parameter Estimation in Ground Water. *Journal of Hydrology*, 318, 215-231.
- Helle, H. B., Alpana B. & Bjorn U. (2001). Porosity and Permeability Prediction from Wireline Logs Using Artificial Neural Networks: a North Sea Case Study. *Geophysical Prospecting*, 49, 431-444.
- Hinton, G. E. (1992). Redes neuronales que aprenden de la experiencia. *Investigación y Ciencia*, 194, 105 – 112
- Kuo, Y., Chen-Wuing, L. & Kao-Hung, L. (2003). Evaluation of the Ability of an Artificial Neural Network Model to Assess the Variation of Groundwater Quality in an area of Blackfoot Disease in Taiwan. *Water Research*, 38, 148-158.
- Neural Tools (2013). *Guía de uso versión 6.0*. Recuperado el 13 de agosto del 2013, de http://www.palisade.com/downloads/manuals/612/ES/NeuralTools6_ES.pdf
- Salini, G. & Pérez J. P. (2006). Estudio de series temporales de contaminación ambiental mediante técnicas de redes neuronales artificiales. *Ingeniare, Revista Chilena de Ingeniería*, 14(3), 284-290.
- Specht, D. (1991). A General Regression Neural Network. *IEEE Transactions on Neural Networks*, 2, 568-576.
- Superintendencia Nacional de Servicios de Saneamiento (SUNASS) (2000). *Parámetros de calidad y límites máximos permisibles de agua potable*. Recuperado el 21 de septiembre del 2013, de <http://www.sunass.gob.pe/doc/normas%20legales/legisla%20web%28cambio%29/normas/calidad%20de%20agua/Oficio%20677.pdf>
- United States Environment Protection Agency (EPA) (2000). *Estándares del Reglamento Nacional Primario de Agua Potable*. Recuperado el 21 de septiembre del 2013, de <http://water.epa.gov/drink/agua/estandares.cfm>
- Wang, M. X., Liu, G. D., Wu, W. L., Bao, Y. H. & Liu W. N. (2006). Prediction of Agriculture Derived Groundwater Nitrate Distribution in North China Plain with GIS-based BPN. *Environment Geology*, 50, 637-644.
- Yesilnacar, M. I., Erkan, S., Naz, M. & Bestamin, O. (2007). Neural Network Prediction of Nitrate in Groundwater of Harran Plain, Turkey. *Environmental Geology*, 56(1), 19 – 25.