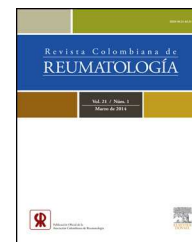




Revista Colombiana de REUMATOLOGÍA

www.elsevier.es/rcreuma



Editorial

Modelos de aprendizaje computacional en reumatología

Machine learning models in rheumatology



El aprendizaje computacional (*machine learning* en inglés) estudia la construcción de sistemas capaces de aprender a partir de datos¹. Esto incluye una gran variedad de sistemas, desde sistemas de visión por ordenador hasta sistemas para detectar correo no deseado (*spam*). En todos los casos, un sistema que aprende debe ser capaz de generalizar, es decir, de encontrar patrones y regularidades en los datos que le permitan desempeñarse bien en datos que no ha observado previamente. Existen 2 tipos principales de modelos de aprendizaje computacional, modelos de aprendizaje supervisado y modelos de aprendizaje no supervisado. En el aprendizaje supervisado se busca inducir modelos capaces de predecir el valor de ciertas variables dependientes a partir de variables independientes. Un ejemplo de problema de aprendizaje supervisado es el problema de clasificación, en el cual la variable dependiente corresponde a un atributo que indica a qué clase (por ejemplo, enfermo o control en el caso de un problema de diagnóstico médico) pertenece una muestra particular. En los modelos de aprendizaje no supervisado no hay una distinción entre variables dependientes y no dependientes, en este caso se pretende encontrar la estructura subyacente que explique la estructura de los datos. El ejemplo más representativo de aprendizaje no supervisado es el análisis de conglomerados (*clustering* en inglés) en el cual el objetivo es encontrar grupos de datos que compartan características similares.

En medicina los modelos de aprendizaje computacional se han aplicado con éxito tanto a problemas motivados por la práctica clínica, como el diagnóstico asistido por ordenador, como a problemas de análisis de datos de investigación médica básica. En los últimos años ha habido un gran auge en la investigación y desarrollo de modelos de aprendizaje computacional aplicados al diagnóstico médico de diversas enfermedades y condiciones médicas.

La [figura 1](#) muestra el número de artículos publicados entre 1969 y 2014 sobre aplicaciones de técnicas de aprendizaje

computacional al diagnóstico médico. Como se puede observar, hay una tendencia creciente, la cual se ha acelerado durante los últimos 10 años, alcanzando un volumen de cerca de 100 artículos sobre el tema por año. Los tipos de problemas de diagnóstico médico abordados con estas técnicas cubren prácticamente todas las especialidades de la medicina², algunos ejemplos incluyen: diagnóstico de glaucoma³, identificación de enfermedades cardiovasculares⁴, detección de la enfermedad de Alzheimer⁵ y detección del cáncer de próstata⁶.

La reumatología no ha sido ajena a este fenómeno, pues también ha habido un incremento en las investigaciones sobre la aplicación de modelos de aprendizaje computacional al diagnóstico de enfermedades reumáticas. La [figura 2](#), muestra el número de publicaciones sobre el tema desde 1986 hasta el 2014, evidenciando un aumento de las mismas en los últimos años. Ejemplos de aplicaciones incluyen: diagnóstico de la artritis reumatoide^{7,13}, predicción de la respuesta al tratamiento con certolizumab pegol en pacientes con artritis reumatoide⁸, diagnóstico de lupus eritematoso sistémico⁹, predicción de la respuesta terapéutica en pacientes con artritis poliarticular juvenil idiopática¹⁰, diagnóstico del síndrome de Sjögren¹¹, predicción de complicaciones pulmonares y supervivencia de largo plazo en esclerosis sistémica¹², entre otros.

Siguiendo esta línea de trabajo, la presente edición de la REVISTA COLOMBIANA DE REUMATOLOGÍA incluye un artículo de investigación que aborda la aplicación de modelos de aprendizaje computacional para la clasificación de pacientes con artritis reumatoide y controles, a partir de datos genéticos, serológicos y clínicos¹³. Además de los modelos de clasificación, el trabajo logra establecer relaciones entre marcadores genéticos y fenotipos (endofenotipos).

Los autores aplican métodos de aprendizaje supervisado como las redes neuronales y las redes bayesianas, y de aprendizaje no supervisado como el método de agrupamiento

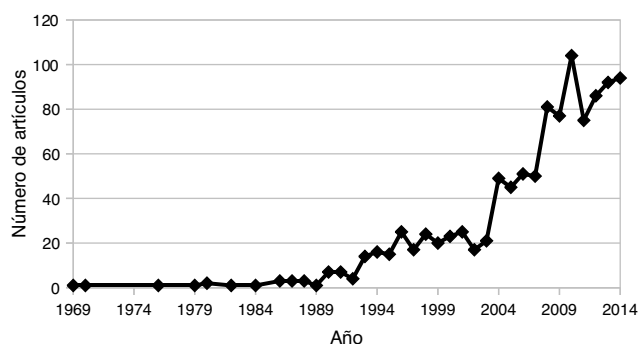


Figura 1 – Artículos publicados por año desde 1969 hasta 2014 incluyendo temas relacionados con aprendizaje computacional y diagnóstico médico. Fuente: Scopus.

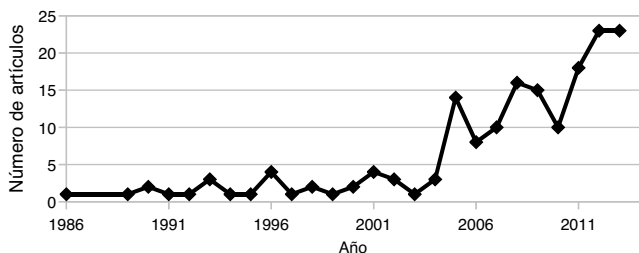


Figura 2 – Artículos publicados por año desde 1986 hasta 2014 incluyendo temas relacionados con aprendizaje computacional y reumatología. Fuente: Scopus.

k-medias. Adicionalmente, para la identificación de endofenotipos los autores utilizaron alineamiento de secuencias de aminoácidos.

Los resultados presentados en el artículo son bastante prometedores: los modelos predictivos logran un alto desempeño en la tarea de discriminar entre pacientes con artritis reumatoide y controles, con sensibilidad y especificidad de hasta el 92,3% y el 93,3%, respectivamente. Por otro lado, el análisis de conglomerados basado en k-medias permitió establecer 2 grupos de pacientes con diferentes niveles de severidad. Finalmente, el análisis de secuencias de aminoácidos reveló secuencias comunes en pacientes con desenlaces iguales.

BIBLIOGRAFÍA

1. Alpaydin E. *Introduction to machine learning*. 2.^a ed MIT press; 2014.
2. Sajda P. Machine learning for detection and diagnosis of disease. *Annu Rev Biomed Eng*. 2006;8:537–65.
3. Chan K, Lee T-W, Sample PA, Goldbaum MH, Weinreb RN, Sejnowski TJ. Comparison of machine learning and traditional classifiers in glaucoma diagnosis. *IEEE Transactions on Bio-Medical Engineering*. 2002;49:963–74. Disponible en: <http://doi.org/10.1109/TBME.2002.802012>.

4. Chui KT, Tsang KF, Wu CK, Hung FH, Chi HR, Chung HS, et al. Cardiovascular diseases identification using electrocardiogram health identifier based on multiple criteria decision making. *Expert Systems with Applications*. 2015;42:5684–95. Disponible en: <http://doi.org/10.1016/j.eswa.2015.01.059>.
5. Escudero J, Ifeachor E, Zajicek JP, Green C, Shearer J, Pearson S. Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease. *IEEE Transactions on Bio-Medical Engineering*. 2013;60:164–8. Disponible en: <http://doi.org/10.1109/TBME.2012.2212278>.
6. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Transactions on Medical Imaging*. 2014;33:1083–92. Disponible en: <http://doi.org/10.1109/TMI.2014.2303821>.
7. Conaghan PG. Predicting outcomes in rheumatoid arthritis. *Clin Rheumatol*. 2011;30 Suppl 1:S41–7. Disponible en: <http://doi.org/10.1007/s10067-010-1639-4>.
8. Curtis JR, Luijckens K, Kavanaugh A. Predicting future response to certolizumab pegol in rheumatoid arthritis patients: features at 12 weeks associated with low disease activity at 1 year. *Arthritis Care Res (Hoboken)*. 2012;64:658–67. Recuperado de: <http://www.scopus.com/inward/record.url?eid=2-s2.0-84861127529&partnerID=tZOtx3y1>.
9. Huang Z, Shi Y, Cai B, Wang L, Wu Y, Ying B, et al. MALDI-TOF MS combined with magnetic beads for detecting serum protein biomarkers and establishment of boosting decision tree model for diagnosis of systemic lupus erythematosus. *Rheumatology (Oxford)*. 2009;48:626–31. Disponible en: <http://doi.org/10.1093/rheumatology/kep058>.
10. Jiang K, Sawle AD, Frank MB, Chen Y, Wallace CA, Jarvis JN. Whole blood gene expression profiling predicts therapeutic response at six months in patients with polyarticular juvenile idiopathic arthritis. *Arthritis Rheumatol*. 2014;66:1363–71. Disponible en: <http://doi.org/10.1002/art.38341>.
11. Li Y, Sun X, Zhang X, Yang Y, Jia R, Liu X, et al. Establishment of a novel diagnostic model for Sjögren's syndrome by proteomic fingerprinting. *Clin Rheumatol*. 2014;33:1745–50. Disponible en: <http://doi.org/10.1007/s10067-014-2762-4>.
12. Nihtyanova SI, Schreiber BE, Ong VH, Rosenberg D, Moynzadeh P, Coghlan JG, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. *Arthritis Rheumatol*. 2014;66:1625–35. Disponible en: <http://doi.org/10.1002/art.38390>.
13. Modelo computacional para la identificación de endofenotipos y clasificación de pacientes con artritis reumatoide a partir de datos genéticos, serológicos y clínicos utilizando técnicas de inteligencia computacional. *Rev Colomb Reumatol*. 2015;22:90–103. Disponible en: <http://dx.doi.org/10.1016/j.rcreu.2015.05.005>

Fabio A. González
Grupo de investigación MindLab, Universidad Nacional de
Colombia, Bogotá, Colombia

Correo electrónico: fagonzalezo@unal.edu.co

<http://dx.doi.org/10.1016/j.rcreu.2015.06.001>
0121-8123/© 2015 Asociación Colombiana de Reumatología.
Publicado por Elsevier España, S.L.U. Todos los derechos reservados.