Revista Colombiana de
REUMATOLOGÍA
www.elsevier.es/rcreuma

Editorial

Machine Learning Models in Rheumatology☆

Modelos de aprendizaje computacional en reumatología

Machine learning explores the construction of systems capable to learn from data.[1] This includes a great variety of systems, from computer vision systems to systems that detect unwanted email (spam). In all cases, a system that learns must be able to generalize, i.e., to find patterns and regularities in the data that enable it to have a good performance with data not observed previously. There are two main types of machine learning models, supervised learning models and unsupervised learning models. The supervised learning searches to induce models capable to predict the value of certain dependent variables based on independent variables. An example of supervised learning is the issue of classification, in which the dependent variable corresponds to an attribute that indicates to which class (for example, patient or control in the case of a problem of medical diagnosis) belongs a particular sample. In the unsupervised learning models there is no distinction between dependent and non-dependent variables, in this cases is intended to find the underlying structure that explains the structure of the data. The most representative example of unsupervised learning is the cluster analysis (clustering) in which the objective is to find datasets that share similar characteristics.

In medicine the machine learning models have been successfully applied both to issues motivated by the clinical practice, such as computer-aided diagnosis, and to issues of data analysis of basic medical research. In the last years there has been a boom in the research and development of machine learning models applied to the medical diagnosis of various diseases and medical conditions.

Figure 1 shows the number of articles published between 1969 and 2014 on the application of machine learning techniques to medical diagnosis. As it can be seen, there is an increasing trend, which has been accelerated in the last 10 years, reaching a volume of nearly 100 articles on the subject per year. The types of medical diagnosis issues addressed by this techniques cover virtually all specialties of medicine,[2] some examples include: diagnosis of glaucoma,[3] identification of cardiovascular diseases,[4] detection of Alzheimer's disease[5] and detection of prostate cancer.[6]

Rheumatology has not been alien to this phenomenon, as there has also been an increase in research on the application of machine learning models to the diagnosis of rheumatic diseases. Figure 2 shows the number of publications on the subject since 1986 until 2014, evidencing an increase thereof in the last years. Examples of applications include: diagnosis of rheumatoid arthritis,[7-13] prediction of the response to treatment with certolizumab pegol in patients with rheumatoid arthritis,[8] diagnosis of systemic lupus erythematosus,[9] prediction of the therapeutic response in patients with polyarticular juvenile idiopatic arthritis,[10] diagnosis of Sjögren syndrome,[11] prediction of pulmonary complications and long-term survival in systemic sclerosis,[12] among others.

Following this line of work, this edition of the Revista Colombiana de Reumatología (Colombian Journal of Rheumatology) includes a research article addressing the application of machine learning models for the classification of patients with rheumatoid arthritis and controls, based on genetic, serological and clinical data.[13] In addition to the classification models, the work could establish relationships between genetic markers and phenotypes (endophenotypes).

The authors apply supervised learning methods such as the neural networks and the Bayesian networks, and or unsupervised learning such as the k-means clustering method. Additionally, for the identification of endophenotypes the authors used alignment of amino acid sequences.

The results presented in the article are quite promising: the predictive models achieve high performance in the task of discriminating between patients with rheumatoid arthritis and controls, with a sensitivity and specificity up to 92.3% and 93.3%, respectively. On the other hand, the cluster analysis based in k-means made possible to establish 2 groups of patients with different levels of severity. Finally, the analysis of amino-acid sequences revealed common sequences in patients with the same outcomes.

☆Please cite this article as: González FA. Modelos de aprendizaje computacional en reumatología. Rev Colomb Reumatol. 2015. http://dx.doi.org/10.1016/j.rcreu.2015.06.001

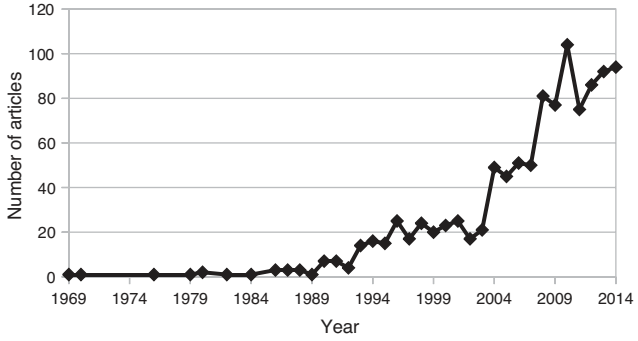

Figure 1 – Articles published per year since 1969 until 2014 including topics related with machine learning and medical diagnosis. Source: Scopus.

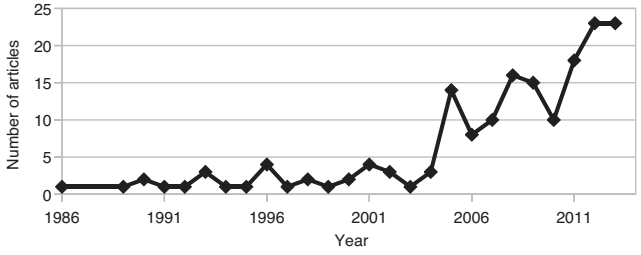

Figure 2 – Articles published per year since 1986 until 2014 including topics related with machine learning and Rheumatology. Source: Scopus.

REFERENCES

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. Available from: 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. Available from: 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. Available from: 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. Available from: http://doi.org/10.1109/TMI.2014.2303821
7. Conaghan PG. Predicting outcomes in rheumatoid arthritis. Clin Rheumatol. 2011;30 Suppl 1:S41-7. Available from: http://doi.org/10.1007/s10067-010-1639-4
8. Curtis JR, Luijtens 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. Retrieved from: 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. Available from: 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. Available from: 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. Available from: http://doi.org/10.1007/s10067-014-2762-4
12. Nihtyanova SI, Schreiber BE, Ong VH, Rosenberg D, Moinzadeh P, Coghlan JG, et al. Prediction of pulmonary complications and long-term survival in systemic sclerosis. Arthritis Rheumatol. 2014;66:1625-35. Available from: 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. Available from: http://dx.doi.org/10.1016/j.rcreu.2015.05.005

Fabio A. González

Grupo de investigación MindLab, Universidad Nacional de Colombia, Bogotá, Colombia

E-mail address: fagonzalezo@unal.edu.co
2444-4405/© 2015 Asociación Colombiana de Reumatología. Publised by Elsevier España, S.L.U. All rights reserved.