

## Predicen el número de citas que tendrán los artículos científicos

Investigadores de la [Universidad Politécnica de Madrid \(UPM\)](#) han elaborado un sistema predictivo para calcular el número de citas futuras que tendrán los artículos científicos publicados.

UPM

8/3/2010 10:52 CEST



Fuente: (cc) Julio Pesquero.

A través de un estudio sobre artículos publicados en la revista *Bioinformatics*, investigadores del [departamento de Inteligencia Artificial](#) de la [UPM](#) han construido una serie de modelos con los que se puede predecir, en el mismo momento de su publicación, el número de citas que recibirá un artículo publicado en dicha revista.

Debido al apogeo de la ciencia y a la multitud de artículos que se publican sobre ella, los editores de las revistas científicas se encuentran con la dificultad de saber elegir, entre una gran variedad, los artículos de mayor calidad y, al mismo tiempo, con el mayor número de lectores posibles.

Con el fin de evaluar los resultados de las investigaciones científicas de una manera objetiva, se han desarrollado una serie de medidas cuantitativas asociadas a las publicaciones de las mismas. Una de estas medidas es el número de citas recibidas por un artículo en un periodo determinado de tiempo. Aunque el número de citas recibidas es una medida de visibilidad, podría ser considerada como una medida indirecta de la calidad del artículo.

Relacionado con esto, investigadores de la [Facultad de Informática](#) de la UPM han construido unos modelos inteligentes(1) capaces de predecir el

número de citas que, en cada uno de los primeros cuatro años, recibirán los artículos publicados en la revista *Bioinformatics* (revista concreta donde han aplicado sus modelos).

Para la construcción de estos modelos se han utilizado como variables predictivas una selección de ciertas palabras clave identificadas de forma automática a partir de los resúmenes de los artículos, junto con variables referidas a las áreas de publicación de los artículos y a la fecha de publicación de los mismos. Gracias a las variables predictivas utilizadas se podrán realizar en el mismo momento de la publicación predicciones sobre el número de citas que recibirá un artículo determinado.

El porcentaje de acierto de los modelos depende del área de publicación del artículo que se quiera predecir, del año que se quiera pronosticar y del método utilizado para realizar la predicción. [Tras realizar distintas predicciones](#) con cada combinación de área, año y método, se obtiene un porcentaje de acierto medio del 91.8% en cada una de las posibles áreas y años de predicción.

Gracias a estos modelos predictivos se demuestra que la aparición de ciertas palabras en el resumen del artículo está correlacionada con el número de citas recibidas. Estas palabras pueden ser usadas como un punto de referencia para identificar temas que están de actualidad.

Finalmente, se debe resaltar la ventaja de estos modelos frente a los existentes ya que ofrecen la posibilidad de predecir en el mismo momento de la publicación el número de citas que obtendrá un artículo en los primeros años.

*(1) Bioinformatics, 25, 3303-3309 (2009). "Predicting citation count of Bioinformatics papers within four years of publication". Ibáñez, A., Larrañaga, P. y Bielza, C.*

Derechos: **Creative Commons**

**TAGS**

PREDICCIÓN DEL NÚMERO DE CITAS | REGRESIÓN LOGÍSTICA |  
REDES BAYESIANAS | APRENDIZAJE AUTOMÁTICO | MINERÍA DE TEXTO |  
CLASIFICACIÓN |

Creative Commons 4.0

Puedes copiar, difundir y transformar los contenidos de SINC. [Lee las condiciones de nuestra licencia](#)