Aplicación de la Redes Neuronales en la Bibliometria

Susana Beatriz Miranda Asturizaga
Simulación de Sistemas Inf-391
su_ma84@hotmail.com

RESUMEN

El desarrollo tecnológico ha hecho surgir nuevas disciplinas y ha potenciado otras, en medio de ello la observación de la convergencia entre esas disciplinas emergentes y tradicionales resulta interesante.

Las redes neuronales artificiales (RNA) son aplicadas en diversos ámbitos de la actividad humana. Una de sus aplicaciones es como herramienta de análisis de información, específicamente dentro de la Bibliometría.

La bibliometría ha sido utilizada durante décadas para medir la actividad y repercusión científica.

En este trabajo se hace una introducción sobre las particularidades de las RNA, específicamente las basadas en el modelo Kohonen (Mapas auto-organizativos). Se exponen los elementos que la integran y se vincula su principio de funcionamiento con la Bibliometría.

Se ejemplifica la utilidad de las RNA, dentro de la Bibliometría, a través de casos prácticos.

Palabras Clave

Bibliometría, Ciencia, Herramientas, Información.

1. INTRODUCCIÓN

Las técnicas computacionales, desde sus inicios, marcaron un paradigma en la creación. Sus aplicaciones, hoy día, van desde la industria de los juegos hasta las cadenas de producción de varias empresas. Para ello se han desarrollado múltiples técnicas como las relacionadas con la inteligencia artificial. Las más conocidas son la lógica difusa (aprendizaje inductivo), algoritmos genéticos y redes neuronales.

Actualmente las redes neuronales se emplean en diferentes campos, estos se agrupan según varios criterios. Uno de ellos es el propuesto por Deboeck, quien los agrupa en:

- Modelación financiera y económica.
- Perfiles de mercado y clientes.
- Aplicaciones médicas.
- Gerencia del conocimiento y “descubrimiento de datos”.
- Optimización de procesos industriales y control de calidad.
- Investigación científica.

La medición del esfuerzo y repercusión de la actividad científica se basa, hoy en día en la bibliometría. Los indicadores que se construyen a partir de técnicas bibliométricas cuantifican el número de documentos publicados por un país, institución, grupo de investigación o individuo, así como las citas recibidas por dichos documentos.

Por lo tanto, las medidas bibliométricas más comunes son las basadas en las publicaciones y en las citas.

El objetivo del presente trabajo es utilizar la base teórica de las redes neuronales artificiales como una herramienta práctica que permita realizar análisis exploratorio de datos o minería de datos vinculados con los indicadores bibliométricos [1].

2. REDES NEURONALES ARTIFICIALES

Las redes neuronales artificiales o RNA nacen de la necesidad de crear sistemas que trabajen igual que el cerebro humano, es decir que tengan la capacidad de poder adaptarse ante cualquier situación que se presente en la vida real, situaciones que el ser humano puede reconocer y manejar; por ejemplo reconocer a una persona, cosa o adaptarse a un timbre de vos humana nuevo, reconocer daños físicos, etc.

Lo importante en el desarrollo de la técnica de las RNA es su útil comportamiento al aprender, reconocer y aplicar relaciones entre objetos y tramas de objetos propios del mundo real. En este sentido, se utilizan las RNA como una herramienta que podrá utilizarse para resolver problemas difíciles.

La posibilidad de resolver problemas difíciles es dable gracias a las ventajas de las redes Neuronales.

Estos son enunciados a continuación:

2.1. Aprendizaje adaptativo

La capacidad de aprendizaje adaptativo es una de las características más atractivas de redes neuronales. Esto es, aprenden a llevar a cabo ciertas tareas mediante un entrenamiento con ejemplos ilustrativos. Como las redes neuronales pueden aprender a diferenciar patrones mediante ejemplos y entrenamientos, no es necesario elaborar modelos a priori ni se tiene la necesidad de especificar funciones de distribución de probabilidad.

Las redes neuronales son sistemas dinámicos autoadaptativos. Son adaptables debido a la capacidad de autoajuste de los elementos procesales (neuronas) que componen el sistema. Son

dinámicos, pues son capaces de estar constantemente cambiando para adaptarse a las nuevas condiciones.

En el proceso de aprendizaje, los enlaces ponderados de las neuronas se ajustan de manera que se obtengan ciertos resultados específicos. Una red neuronal no necesita un algoritmo para resolver un problema, ya que ella puede generar su propia distribución de pesos en los enlaces mediante el aprendizaje. También existen redes que continúan aprendiendo a lo largo de su vida, después de completado su período de entrenamiento.

2.2. Auto-organización

Las redes neuronales emplean su capacidad de aprendizaje adaptativo para autoorganizar la información que reciben durante el aprendizaje y/o la operación. Mientras que el aprendizaje es la modificación de cada elemento procesal, la autoorganización consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico.

Esta autoorganización provoca la generalización: facultad de las redes neuronales de responder apropiadamente cuando se les presentan datos o situaciones ha las que no había sido expuesta anteriormente. El sistema puede generalizar la entrada para obtener una respuesta. Esta característica es muy importante cuando se tiene que solucionar problemas en los cuales la información de entrada no es muy clara; además permite que el sistema dé una solución, incluso cuando la información de entrada está especificada de forma incompleta.

2.3. Tolerancia a fallos

Las redes neuronales fueron los primeros métodos computacionales con la capacidad inherente de tolerancia a fallos. Comparados con los sistemas computacionales tradicionales, los cuales pierden su funcionalidad cuando sufren un pequeño error de memoria, en las redes neuronales, si se produce un fallo en un número no muy grande de neuronas y aunque el comportamiento del sistema se ve influenciado, no sufre una caída repentina.

Hay dos aspectos distintos respecto a la tolerancia a fallos:

a) Las redes pueden aprender a reconocer patrones con ruido, distorsionados o incompletos. Esta es una tolerancia a fallos respecto a los datos.

b) Las redes pueden seguir realizando su función (con cierta degradación) aunque se destruya parte de la red.

La razón por la que las redes neuronales son tolerantes a los fallos es que tienen su información distribuida en las conexiones entre neuronas, existiendo cierto grado de redundancia en este tipo de almacenamiento. La mayoría de los ordenadores algorítmicos y sistemas de recuperación de datos almacenan cada pieza de información en un espacio único, localizado y direccionable. En cambio, las redes neuronales almacenan información no localizada. Por lo tanto, la mayoría de las interconexiones entre los nodos de la red tendrán sus valores en función de los estímulos recibidos, y se generará un patrón de salida que represente la información almacenada.

2.4. Operación en tiempo real

Una de las mayores prioridades, casi en la totalidad de las áreas de aplicación, es la necesidad de realizar procesos con datos de forma muy rápida. Las redes neuronales se adaptan bien a esto debido a su implementación paralela. Para que la mayoría de las redes puedan operar en un entorno de tiempo real, la necesidad de cambio en los pesos de las conexiones o entrenamiento es mínima.

2.5. Fácil inserción dentro de la tecnología existente

Una red individual puede ser entrenada para desarrollar una única y bien definida tarea (tareas complejas, que hagan múltiples selecciones de patrones, requerirán sistemas de redes interconectadas). Con las herramientas computacionales existentes (no del tipo PC), una red puede ser rápidamente entrenada, comprobada, verificada y trasladada a una implementación hardware de bajo coste. Por lo tanto, no se presentan dificultades para la inserción de redes neuronales en aplicaciones específicas, por ejemplo de control, dentro de los sistemas existentes. De esta manera, las redes neuronales se pueden utilizar para mejorar sistemas en forma incremental y cada paso puede ser evaluado antes de acometer un desarrollo más amplio [4].

3. MODELO KOHONEN

Este modelo surge ante la curiosidad de Teuvo Kohonen, quien interesado en comprender la clasificación natural que hace el cerebro en cuanto a su funcionalidad, ideó el algoritmo SOM. Una definición simplificada sobre los mapas topológicos podría ser que, en una correspondencia que respecte la topología, las unidades que se encuentran físicamente próximas entre sí van a responder a una clase de vectores de entrada que, análogamente, se encuentren cerca unos de otros.

Los vectores de entrada de muchas dimensiones son representados sobre el mapa bidimensional, de tal manera que se mantenga el orden natural de los vectores de entrada.

Estos mapas presentan la característica de organizar la información de entrada, de entre un gran volumen de datos, clasificándola automáticamente, esto permitirá visualizar relaciones importantes entre datos. Este modelo es muy útil para establecer relaciones desconocidas previamente.

Este modelo de RNA es del tipo “aprendizaje autoorganizado” que como se explicó con anterioridad es un proceso donde las neuronas aprenden mediante la autoorganización.

Durante el proceso de aprendizaje, al ingresar un dato solo una neurona que tenga una actividad positiva dentro de la vecindad será activada en la capa de salida.

La arquitectura SOM se ha extendido a infinidad de aplicaciones, es por ello que este algoritmo y sus modelos se han automatizado para dar vida a varios software como son el Viscovery SOMine y el WEBSOM. Ambos son utilizados en el análisis y filtrado de información, el Viscovery ha sido validado tanto en estudios de mercado como análisis financiero o proyecciones urbanísticas [2].

Actualmente está incursionando en el tratamiento de la información usando herramientas diseñadas por la bibliometría.

3.1. Los mapas autoorganizados aplicados a la Bibliometría.

La Bibliometría es una disciplina que estudia los aspectos cuantitativos de la información registrada, para ello se han creado una serie de modelos estadísticos que aportan datos numéricos sobre el comportamiento de la actividad científica. También se han adaptado modelos de otras disciplinas para facilitar los análisis y representar los resultados desarrollados a partir de la Bibliometría. Los mapas auto-organizados (SOM) o modelo de Kohonen (basado en las RNA) es una de estas herramientas [3].

En los estudios métricos la aplicación de las redes neuronales, y específicamente los SOM, están asociados en lo fundamental con la clasificación de información, o sea, la formación de cluster y su representación en mapas bidimensionales de conceptos y más específicamente con el descubrimiento de información (*data mining*). Este último vinculado con la recuperación de la información con "ruido" e incompleta o con el tratamiento de información que incluye diferentes tipos de datos (números, texto, registros estructurados, etc.). Los SOM facilitan que el conocimiento tácito se haga explícito, a partir de la extracción no-trivial (a partir de los datos) de conocimientos implícitos potencialmente útiles desconocidos previamente. Se podrán encontrar patrones o estructuras en el conocimiento tácito.

Las investigaciones bibliométricas, a través de la utilización de las redes neuronales, incursionan en:

- la selección de variables,
- clasificación de información o formación de cluster,
- regresión,
- relaciones entre variables,
- cambios y desviaciones,
- representación de las variables.

Lo anterior se puede ejemplificar a partir de algunas aplicaciones prácticas relacionadas con la evaluación de páginas web y trabajos relacionados con la clasificación de revistas en un determinado campo temática. Se conocen, además, investigaciones relacionadas con la minería de textos (*text mining*) sobre todo aplicado a la asociación de palabras o co-word.

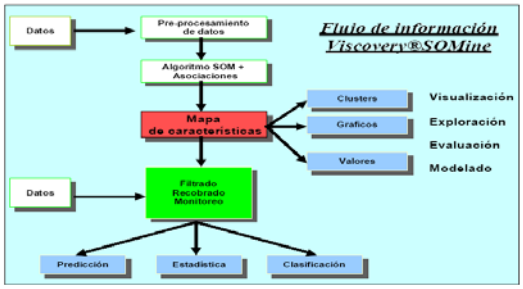
En todos estos ejemplos se utiliza como variante de las RNA el modelo de los mapas autoorganizativos (*self-organizing map*, SOM).

Un ejemplo de SOM podría ser el estudio de una temática determinada, para este caso en un mapa cada documento (artículo de revista, podría ser una patente, una tesis, etc.) va a ocupar un lugar en el espacio, en función de su contenido temático. Cada área del mapa va a reflejar un contenido específico y los tópicos van variando levemente a lo largo del mismo.

Las diferentes tonalidades indican la densidad de documentos, cuanto más oscura más documentos se encuentran.

Este uso frecuente de los SOM quizás se deba a lo amigable de la interfaz de los mapas para los usuarios finales y a la diversidad de sus utilidades prácticas, estas representaciones son válidas para poder identificar, además de los desarrollos temáticos antes mencionados, relaciones entres áreas temáticas y publicaciones, alianzas estratégicas y características de la cooperación. Permite, también, visualizar los avances tecnológicos que tienen lugar en un período, conocer la evolución de una tecnología a través del tiempo e identificar campos emergentes.

La lógica de funcionamiento del ViscoverySOMine se muestra en la Figura siguiente.



El punto de partida, para el uso de esta herramienta, es la entrada de un conjunto de datos numéricos (datos multivariables, variables, “nodos”). Estos datos necesitan ser preprocesados con el objetivo de “organizarlos” en forma de matrices. Los datos son convertidos hasta obtener una información visual en forma de mapa, para ello se aplica un número de técnicas de evaluación como coeficientes de correlación entre variables o factores discriminantes. Los mapas serán amigables a la vista del usuario final, en ellos se identificarán dependencias entre parámetros, cluster y gráficos que facilitarán diferentes predicciones o el proceso de monitoreo.

4. APLICACIONES

Se estima que a pesar de las limitaciones técnicas, las redes neuronales aplicadas a la Bibliometría constituyen un campo de investigación muy prometedor. Un ejemplo es presentado a continuación.

Esta aplicación se presenta en un campo diferente del conocimiento: la agricultura. En este ejemplo no solo se tomó como elemento de entrada la producción documental de determinados países, también se consideraron otros tipos de variables como "gastos en I+D" en cada uno de los países analizados, "personal dedicado a actividades de I+D".

Mapa de América Latina y el Caribe (países seleccionados) dacuerdo con la actividad en ciencia agrícola.

En la figura anterior se presenta un mapa auto-organizado sobre la actividad en ciencia agrícola en América Latina y el Caribe (se seleccionaron algunos países según los datos disponibles).

Con ello se pretende lograr una representación de la región, teniendo en cuenta los indicadores de insumo y de resultados más significativos. En la figura aparecen, en dos dimensiones, 22 países. La semejanza de los países, considerando de forma simultánea los 20 indicadores, se expresa mediante la cercanía de estos en el mapa.

Los 22 países se agrupan (autorganizan) en 3 grupos: El grupo C1, que aparece en la esquina inferior izquierda, es seguido por una banda de 4 países correspondientes al grupo C2 (Cuba, Colombia, Chile y Venezuela) y otro grupo (C3) con el resto de los países.

En el ámbito regional e internacional hay un grupo de países formado por Brasil, México y Argentina que tiene una investigación en la temática con mayor solidez y con parecidos niveles de desarrollo. Esto coincide con los países que presentan mayor nivel regional en el desarrollo agrícola.

5. CONCLUSIONES

La bibliometría y las redes neuronales artificiales como vimos en el ejemplo anterior se convierten en una herramienta de estudio muy valiosa, dado que gracias a las ventajas que las redes neuronales nos brindan podemos representar datos coleccionados durante un período tiempo sobre un país determinado, el software mencionado en articulo es una herramienta muy útil para genera un mapa en el cual se representan el conjunto de los datos.

El vínculo de la Bibliometría con las redes neuronales se muestra muy fructífero. Esta herramienta de análisis debe ser estudiada no solo por la Bibliometría, sino también por otras disciplinas, las cuales pueden potenciar con ella su propio desarrollo.

La revolución tecnológica esta ahí y continuará adelante, su desarrollo implica la convergencia de disciplinas y requiere de equipos multidisciplinares de biólogos, médicos y... especialistas en información que dominen las tecnologías de gestión del conocimiento, minería de datos, lenguajes avanzados de Internet, sistemas distribuidos de información y servicios Web, flujos de información, procesos de producción del conocimiento, etc.

6. REFERENCIAS

[1] Norton, MJ. Knowledge Discovery in Database. *Library Trends*. 48(1):9-21; 1999.

[2] StatSoft, Inc. Neural Networks. Statistica for Windows [Computer program manual]..<<http://www..statsoft.com>>.

[3] Sotolongo, G; Suárez, CA.; Guzmán, MV. Modular Bibliometrics Information System withhttp://libres.curtin.edu.au/

[4] www.wikipedia.com