

Reporte de Lectura

Cómo funciona Google: cadenas de Markov y valores propios

Resumen

El artículo escrito por Nectoux (2012) es un texto bastante denso en el que se nos permite conocer las herramientas matemáticas que el algoritmo de búsqueda de Google utiliza para proveer a los usuarios de dicho buscador las páginas y sitios webs más relevantes o con una mayor relación con lo que se ha buscado.

En un principio se nos explica que se trata a los sitios web en general como una especie de nodos pertenecientes a una red más grande de nodos interconectados mediante vértices dirigidos que en conjunto generan un grafo cuya magnitud es similar al tamaño total de páginas web indexadas en el buscador. Dicho ordenamiento permite representar cada nodo mediante las entradas y salidas que posee mediante una especie de matriz de adyacencia conocida como matriz de proceso de cadena de Markov, la cual nos presenta la posibilidad que una página o sitio web posee de ser dirigido (o que el siguiente estado del sistema) sea una de las páginas con las que la página actual es adyacente (tiene aristas dirigidas hacia él).

A partir de su representación como matriz de Markov, los datos que permiten conocer las relaciones y posibilidades de estado de todas las páginas, se pueden utilizar una serie de teoremas probabilísticos un tanto complejos que relacionan los valores y vectores característicos de determinada página web con las probabilidades relacionadas a esta, generando un método en el que no se depende de la cantidad de sitios web indexados en la matriz de Markov, sino en las coordenadas determinadas de la página web dentro de la matriz.

Por último se obtiene una denominada distribución estacionaria (la cual puede ser obtenida directamente o mediante manipulaciones de la matriz original) a partir de la cual se utiliza una especie de modificación del algoritmo del punto fijo denominado "punto fijo de Banach", en base al cual se obtiene una sucesión de páginas web, las cuales podrían ser interpretadas como las páginas web con mayor relación a lo buscado en un principio.

Opinión

Me parece que el texto de Nectoux (2012) explica de forma bastante detallada la teoría matemática detrás del algoritmo de búsqueda de Google, sin embargo, me hubiese gustado que la explicación tuviese un mayor contenido de analogías hacia el proceso que se sigue realmente dentro de la aplicación de la teoría en la consecución del algoritmo de PageRank; claramente la implementación del algoritmo no podría ser explicado de forma directa debido a la propiedad que Google posee sobre él, sin embargo, un poco más de acercamiento al proceso de obtención inicial del grafo dirigido que representa al conjunto de la World Wide Web indexada al buscador de Google hubiera sido bastante útil al momento de comprender cómo es que a partir de los datos que el buscador posee, se establece el grafo dirigido de una forma eficiente que no ocasione que el algoritmo explicado a través del artículo sea inservible (es decir, conocer el Big O de la consecución del grafo dirigido y del algoritmo de PageRank), así también la explicación de si dicho grafo es estático y va aumentando conforme se indexan nuevas páginas al buscador, o si es generado especialmente cada vez que se realiza una búsqueda.

A pesar de lo anterior, la explicación es bastante buena en el sentido que explica de forma detallada los distintos procesos que se siguen, así como los teoremas que se invocan para permitir al algoritmo funcionar de forma eficiente y elegante. El siempre exponer qué significa cada nuevo término introducido en relación a la clasificación y obtención de las páginas web requeridas es bastante valioso, pues permite trasladar lo teórico a lo práctico de forma sutil.

Una de las cosas que encontré un tanto interesante fue el hecho de que, a pesar de tratar temas bastante avanzados en cuestión matemática, fui capaz de comprender la idea general de la gran mayoría de lo que se estaba explicando gracias a los conocimientos que he aprendido durante el curso de esta materia, así como las relacionadas con las matemáticas más propias de mi ingeniería, como lo son Matemáticas Discretas y Estructuras de Datos, lo cual me permitió comprender que soy capaz de entender el significado general de los teoremas y demás constructos matemáticos mencionados, dentro de la aplicación puntual que se está explicando y entender la importancia que poseen dentro de la implementación de dicho algoritmo.

Por lo tanto, hay varias cuestiones que resultaron familiares debido a que fueron tratadas durante el semestre: la eliminación gaussiana así como los vectores y valores característicos fueron cuestiones que tratamos en clase y que, después de haber leído sobre ellos siendo aplicados de forma práctica me permite reconocer aún más el valor que poseen y el papel importante que juegan en mi vida diaria sin que yo sea consciente de ello; dicho reconocimiento me ha motivado bastante a seguir investigando respecto a los temas tratados en clase, los expuestos en el artículo y en aquellos que sé son útiles en el desarrollo de herramientas que facilitan la realización de tareas y la generación de conocimiento más complejo mediante aplicación del conocimiento ya existente.

Para concluir, solo me gustaría mencionar el hecho que me resulta bastante sorprendente el hecho de que Google haya conseguido desarrollar su algoritmo PageRank de forma tan elegante y eficiente; digo esto debido a que las ideas que debieron de haber transitado por las mentes de sus desarrolladores fueron algo completamente innovador, debido a que lograron explotar de forma práctica, conocimientos que parecen no estar relacionados en nada con lo que terminaron consiguiendo, por lo tanto, resulta evidente el hecho que en cualquier caso de innovación y desarrollo de tecnología capaz de revolucionar toda una industria, es necesario que aquellos encargados de desarrollarla (es decir, ingenieros de la misma rama de ingeniería que mi carrera) posean un entendimiento bastante profundo no solo de las bases y fundamentos de los principios matemáticos que rigen el funcionamiento de lo que ya ha sido creado, sino que seamos capaces de transformar

esos conocimientos en nuevas herramientas mediante su correcta aplicación fuera de lo establecido conceptualmente en la teoría.

En base a lo anterior, me parece que mi opinión podría resumirse en las siguientes ideas PNI:

- P El artículo explica de forma bastante densa, pero detallada, los fundamentos matemáticos teóricos que permiten que millones de personas alrededor del mundo realicen búsquedas casi instantáneas de forma sencilla y eficaz; dichos fundamentos incluyen una serie de conocimientos y teoremas que permiten reconocer aún más el valor y utilidad que el conocimiento y correcto dominio de las matemáticas como una herramienta tiene, al momento de desarrollar más y mejores herramientas que permitan producir más conocimiento más sencillamente.
- N El artículo posee teoría matemática bastante densa y complicada en el sentido que la gran mayoría de los teoremas referenciados son bastante difíciles de demostrar sin bases lo suficientemente fuertes en la respectiva área en la que fueron derivados y aplicados.
- I La implementación de Google de su algoritmo es bastante interesante en el sentido que lograron traducir la complejidad de la teoría matemática abstracta hacia un resultado físico que les permitiera obtener ganancias substanciales; es impresionante que los ingenieros y científicos de datos que desarrollaron esta tecnología hayan tenido semejantes ideas para adaptar conocimientos que en un principio parecen no tener relación alguna con la creación de un motor de búsqueda, y convertirlos en los pilares del funcionamiento del mismo.

Referencias

Nectoux, A. (2012). *Cómo Funciona Google: Cadenas De Markov Y Valores Propios*. Recuperado de: Blog Proyecto Klein. En: <http://blog.kleinproject.org/?p=1605&lang=es>