

Cálculo del ranking de los países sudamericanos que participaron en las eliminatorias para el mundial Rusia 2018 usando el algoritmo PageRank

Alex Avila, Bazalar Rommel, Konrad Trejo

aavilas@uni.pe, rbazalarc@uni.pe, konrad.trejo.c@uni.pe

1. Introducción

El objetivo de este proyecto es lograr un ranking parecido al de FIFA, pero solo con las selecciones sudamericanas tomando en cuenta los partidos de la eliminatoria para el mundial Rusia 2018.

Con el fin de cumplir el objetivo mencionado, surge la idea de la creación de un nuevo ranking basándonos en el método de PageRank (generado por Larry Page y Sergey Brin), generando así un nuevo ranking en que se destacan los países sudamericanos.

Basándonos en un historial de los partidos, nos proponemos evaluar cómo se comporta el nuevo método creado en comparación de otros rankings existentes y así observar cual de los métodos disponibles refleja mejor la realidad.

Una vez obtenido el ranking, nos proponemos a, dado un partido, evaluar el ranking de los países que lo disputan así poder indicar con que probabilidad saldrá victorioso el equipo con mejor posicionamiento. Para lograr todo esto tendremos que usar nuestros conocimientos en el curso de Probabilidades y también aprovechar nuestro conocimiento en el lenguaje R.

2. Estado del arte

- "TennisRank: Un Nuevo Ranking De Jugadores De Tenis basado en PageRank", Analizan un nuevo ranking de jugadores de tenis debido a las críticas existentes del ranking actual. Para esto utilizan el famoso algoritmo de google "PageRank".
- "PageRank Approach to Ranking National Football Teams". Analizan los datos disponibles sobre los campeonatos mundiales de fútbol desde 1930 hasta hoy. El objetivo del paper es clasificar los equipos nacionales en función de todos los partidos durante los campeonatos. Para este propósito, aplican el PageRank creado a partir de los partidos jugados durante los torneos.

3. Diseño del experimento

3.1. Objetos, funciones y técnicas utilizadas

- Vectores : Son una de las unidades básicas de trabajo en R y muchos de los métodos de trabajo con este tipo de objetos se aplica para otras clases de objetos. Por ejemplo, casi todos los objetos pueden ser indexados usando `[]`. Un vector es una colección de uno o más objetos del mismo tipo (caracteres, números, etc); esta es una restricción importante a la hora de crear un vector.
- Matrices : Una matrix es un array con dos dimensiones. Tienen una funcionalidad muy parecida a la del vector.
- Plot : La función plot es una función genérica para la representación gráfica de objetos en R. Los gráficos más sencillos que permite generar esta función son nubes de puntos (x,y).
- Grafos con Igraph : El paquete para R Igraph, necesita que se le presenten los datos de la matriz de adyacencia por parejas. Es decir, una matriz de doble entrada convencional (también llamada sociomatrix, tabla de confundido o tabla de concordancia) ha de pasarse al formato de Igraph.
- Eigen : Calcula valores propios y vectores propios de matrices numéricas (dobles, enteras, lógicas) o complejas.

4. Experimentos y resultados

- Línea base :

País	Ranking por FIFA
Brasil	1
Uruguay	2
Argentina	3
Colombia	4
Perú	5
Chile	6
Paraguay	7
Ecuador	8
Venezuela	9
Bolivia	10

- Evaluación del rendimiento de los modelos ensayados:

Ranking	País	PageRank
1	Brasil	0.4738853
2	Uruguay	0.3731390
3	Argentina	0.3793522
4	Colombia	0.3153363
5	Perú	0.2846076
6	Chile	0.2163052
7	Paraguay	0.2669837
8	Ecuador	0.1977172
9	Venezuela	0.1546235
10	Bolivia	0.3616417

- Comparación de línea base y resultados propios :
Podemos ver que el primer puesto lo ocupa Brasil en ambos ranking's. El segundo puesto y tercer puesto sí han cambiado con respecto del Ranking FIFA, en el que Uruguay y Argentina los ocupan respectivamente; en cambio, con el algoritmo PageRank, Argentina y Uruguay ocupan esos puestos respectivamente. En el cuarto puesto sí hay un gran cambio pues Bolivia ascendió del último puesto en el Ranking FIFA al cuarto puesto en el de PageRank. Otro cambio más sería que Paraguay está por encima de Chile en el nuevo Ranking desarrollado con PageRank .

5. Discusión

Interpretación de los resultados obtenidos.

- El cambio más notorio se nota en Bolivia, esto se debe porque Bolivia le ganó o empató a países con pagerank altos, entonces su pagerank subirá considerablemente. Acá no solo se considerará partidos ganados sino también a quién gana.

¿Cómo podría ser mejorado sus resultados?

- Una mayor cantidad de datos mejoraría la precisión de nuestros resultados, obviamente; aún así, es importante señalar que actualmente no se puede realizar un análisis completo de la eficacia del algoritmo, dado que la muestra de todos los equipos sudamericanos es muy pequeña, por lo tanto se necesitaría su desempeño a través de los años en eliminatorias y copas mundiales para compensar.

6. Conclusiones y trabajos futuros

Al inicio del paper se identificó una problemática con la metodología con la cual se obtiene el ranking oficial de la Fifa. Para solucionar esta problemática, hemos desarrollado un algoritmo de clasificación basado en PageRank, aunque no se aplicó a todos los equipos nacionales clasificados en el mundial, sino que se redujo la muestra a tan solo los países sudamericanos. Una vez obtuvimos nuestra propia clasificación, se pasó a compararla con el ranking de la FIFA y se determinó una gran similitud, excepto en un caso especial de gran desviación (Bolivia). Se concluye que el algoritmo presentado tiene gran potencial, y es capaz de reproducir los resultados de otros métodos de clasificación que poseen más data con la que determinar sus resultados.

En el futuro se puede extender este proyecto para calcular un ranking con los países en su totalidad y considerando como antecedentes unos 15-20 años de historia.

7. Bibliografía

- <http://es.fifa.com/worldcup/preliminaries/index.html>
- M. Eisermann, Comment Google classe les pages webb, <http://images.math.cnrs.fr/Comment-Google-classe-les-pages.html>, 2009.
- A.N. Langville and C. D. Meyer, A Survey of Eigenvector Methods for Web Information Retrieval, SIAM Review, Volume 47, Issue 1, (2005), pp. 135–161.
- <http://www.economicas.uba.ar/wp-content/uploads/2016/04/2-¿Cómo-funciona-google-el-algoritmo-pagerank-diagramas-de-grafos-y-cadenas-de-markov.-Juan-Manuel-Barriola-y-Milena-Dotta-2.pdf>
- <https://www.youtube.com/watch?v=u8HtO7Gd5q0>
- C. Rousseau and Y. Saint-Aubin, Mathematics and technology, SUMAT Series, Springer-Verlag, 2008.