

Modelo Predictivo para el problema de las condiciones de surf en Pichilemu en base a Procesos de Decisión de Markov

Gonzalo Gálvez, Felipe Romero y Joaquin Seguel

1. INTRODUCCIÓN

Se sabe que constantemente las condiciones climáticas en Chile pueden variar de un momento a otro. Esta problemática afecta constantemente al surfista que practica dicha disciplina en la ciudad de Pichilemu, ubicada en la zona central de Chile. Esta ciudad es conocida como la capital del surf y, mundialmente conocida también por sus olas y hermosos paisajes, que atraen a miles de turistas chilenos y extranjeros durante todo el año. Para llevar a cabo una práctica ideal de esta disciplina se deben considerar varios climáticos, dentro de los cuales para el caso de estudio, se considerarán relevante el periodo de la ola, la dirección de las olas y finalmente, el tamaño de la ola. Estos factores son los más influyentes en cuanto a que se generen las condiciones ideales para surfear.

De esta forma estudiaremos a través de un análisis de cadena de Markov el comportamiento del mar, clasificándolo en estados y las probabilidades de transición entre estos.

2. REVISIÓN BIBLIOGRÁFICA

En la teoría de probabilidades las cadenas de Markov son procesos estocásticos discretos en los cuales la probabilidad de que ocurra un evento depende solamente del evento inmediatamente anterior, esto se conoce como falta de memoria y fue introducido por Andrei Markov en el año 1907 [Markov, 1906]. De esta forma comienza un amplio campo de estudio el cual se deriva en varios temas, dentro de estos esta el estudio de las condiciones meteorológicas. Ejemplo de esto son los trabajos de [Carpinone et al., 2015] el cual se basa en cadenas de Markov para modelar el comportamiento del viento y [POSADA,] que de igual forma utiliza cadenas de Markov para modelar indicadores de calidad del tiempo. Finalmente el trabajo realizado por [Córdoba et al., 2009] plantea una forma del uso de cadenas de Markov para realizar predicciones del fenómeno meteorológico del niño.

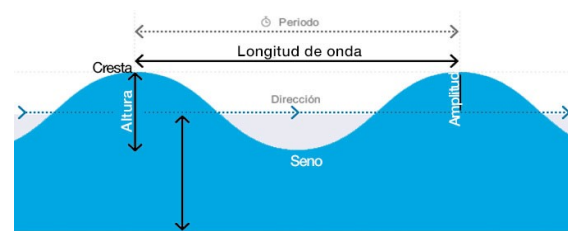
3. METODOLOGÍA

En la presente sección se explicaran los aspectos teóricos en los cuales se sustenta la investigación realizada.

3.1. Datos y Clasificación

Los datos utilizados en este trabajo fueron extraídos desde el sitio web *Windguru*, el cual es un sitio especializado en el gestión de datos climáticos enfocados en condiciones de surf. De esta forma para el estudio se decidió analizar los datos correspondientes al año 2017, los cuales fueron recopilados cada 3 horas, obteniendo un total de 2912 observaciones.

Para la cadena de Markov se decidió crear 5 estados cada uno con una asignación de estrellas, de esta forma el peor mar para realizar surf seria el de 1 estrella y el mejor el de 5 estrellas. Para realizar esta valoración consideramos los 3 principales aspectos que influyen en la calidad de las olas, los cuales son dirección, tamaño y periodo. Estos parámetros fueron escogidos ya que son los principales factores que influyen en las condiciones de una buena ola [Mead and Black, 2015].



Cada uno de las observaciones obtenidas fueron clasificadas en algún estado según la siguiente tabla:

Cuadro 1. Parámetros de Clasificación

Características Ola			Calificación
Dirección	Tamaño	Periodo	
0	no aplica	no aplica	1 estrella
1	≥ 2.5	≥ 13	5 estrellas
	≥ 2.5	$[10,13[$	4 estrellas
	≥ 2.5	<10	2 estrellas
	$[1.5,2.5[$	≥ 13	4 estrellas
	$[1.5,2.5[$	$[10,13[$	3 estrellas
	$[1.5,2.5[$	<10	2 estrellas
	<1.5	≥ 13	3 estrellas
	<1.5	$[10,13[$	2 estrellas
	<1.5	<10	1 estrella

En la tabla anterior se destaca que la dirección del mar se considera como un parámetro binario, ya que si consideramos la dirección del mar como un vector, podemos clasificarlo en 360 grados, por lo cual dado la geografía de Pichilemu, solo si el vector se encuentra entre los 0 y 90 grados es posible realizar surf, de lo contrario, a pesar de las demás condiciones, no se puede realizar.

Esto se puede apreciar de mejor forma en la siguiente imagen satelital de Pichilemu.



3.2. Cadenas de Markov

Si consideramos el comportamiento del océano de una región a través de distintos días, es plausible asumir que el estado actual sólo depende del último estado y no de toda la historia en sí, de modo que se pueden usar cadenas de Márkov [Sawhney, 1962] para formular modelos basados en modelos climáticos usados para el estudio de la dinámica del sistema meteorológico [Gabriel and Neumann, 1962].

Una cadena de Markov se define como un tipo especial de proceso estocástico discreto, en el que la probabilidad que ocurra un evento, depende solamente del evento inmediatamente anterior, por lo tanto se habla de

una falta de memoria en el proceso.

La definición formal de un proceso de Markov se encuentra dado por una serie X_1, X_2, X_3, \dots de variables aleatorias y un X_n , el cual es el estado del proceso en el periodo n , entonces si la probabilidad de transición de un estado a otro depende solamente del estado previo, se tiene:

$$P(X_{n+1}|X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_2 = x_2, X_1 = x_1) \\ = P(X_{n+1}|X_n = x_n)$$

Así la probabilidad de transitar de un estado i a un estado j , se encuentra dada por:

$$P_{ij} = P(X_{n+1} = j|X_n = i)$$

Es así como la matriz de transición se encuentra dada por:

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & \dots & P_{0j} \\ P_{10} & P_{11} & \dots & P_{1j} \\ \dots & \dots & \dots & \dots \\ P_{i0} & P_{i1} & \dots & P_{ij} \end{bmatrix}$$

Con esta matriz definida, podemos predecir el comportamiento del sistema en distintas situaciones:

- Probabilidad de transición en N pasos
- Análisis de periodicidad
- Distribución de transición en el límite

3.3. Transición en N pasos

La probabilidad de transición en N pasos se encuentra dado por:

$$P_{ij}^{(n)} = P(X_{m+n} = j|X_m = i)$$

Entonces,

$$P_{ij}^{(n)} = \sum_{k=0}^{\infty} P_{ik} P_{kj}^{(n-1)} \\ t_i = 1 + \sum_{j=1}^i P_{ij} t_j$$

3.4. Análisis de periodicidad

Considerando el estado arbitrario pero fijo:

$$f_{ii}^{(n)} = P(X_n = i, X_v \neq i; \forall v = 1, 2, \dots, n-1 | X_0 = i)$$

Sea E_k el primer retorno al estado i ;

$$P(E_k) = P(\text{El primer retorno sea en la } k\text{-ésima transición} | X_0 = i)$$

$$*P(X_n = i | X_k = i)$$

$$P(E_k) = f_{ii}^{(k)} P_{ii}^{(n-k)}$$

Considerando $P_{ii}^{(0)} = 1$, Entonces;

$$P(X_n = i | X_0 = i) = \sum_{k=1}^n P(E_k) = \sum_{k=0}^n f_{ii}^{(k)} P_{ii}^{(n-k)}$$

Finalmente por el teorema básico del limite se tiene

$$\lim_{n \rightarrow \infty} P_{ii}^{(n)} = \frac{1}{\sum_{n=0}^{\infty} n * f_{ii}^{(n)}} = \frac{1}{m_i}$$

3.5. Distribución de transición en el límite

La distribución en el limite se puede obtener resolviendo un sistema con la forma:

$$\pi_j = \sum_{k=0}^N \pi_k P_{kj}, \forall j = 0, 1, \dots, N$$

De esta forma podemos estudiar el comportamiento del sistema en el largo plazo.

4. RESULTADOS

4.1. Clasificación

De acuerdo a los datos obtenidos en *Windguru*, cada periodo de 3 horas puede ser clasificado de acuerdo a los parametros definidos anteriormente, a continuación se presenta un extracto de un día de clasificación.

Cuadro 2. Clasificación en un día

Dia	Hora	Direccion Ola	Periodo Ola	Tamaño Ola	Clasificación
01/01/2017	0:00	1	14	2	4
	3:00	1	14	2	4
	6:00	1	14	2	4
	9:00	1	13	2	4
	12:00	1	14	2	4
	15:00	1	14	1.9	4
	18:00	1	14	1.8	4
	21:00	1	14	1.8	4

De esta misma forma se clasificaron las 2.912 observaciones.

4.2. Cadena de Markov

Basándonos en la clasificación mostrada anteriormente podemos obtener la matriz de frecuencia.

$$\begin{bmatrix} 151 & 0 & 1 & 22 & 3 \\ 1 & 64 & 3 & 18 & 0 \\ 0 & 3 & 40 & 7 & 0 \\ 23 & 19 & 6 & 1414 & 87 \\ 3 & 0 & 0 & 86 & 964 \end{bmatrix}$$

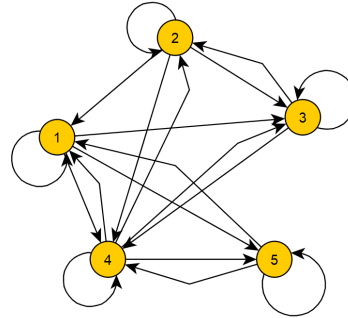
Para el calculo de las probabilidades de transición en cada caso, por ejemplo, para $p_{11} = 151$, significa que se observaron 151 transiciones del estado 1 a si mismo, por lo tanto la probabilidad de transicion se estima como

$$\frac{151}{151 + 0 + 1 + 22 + 4} \approx 0,8531$$

De acuerdo con lo anterior la matriz de transición se presenta a continuación, esta es irreducible y aperiódica,

$$\begin{bmatrix} 0,8531 & 0 & 0,0056 & 0,1242 & 0,0169 \\ 0,0116 & 0,7441 & 0,0348 & 0,2093 & 0 \\ 0 & 0,06 & 0,8 & 0,14 & 0 \\ 0,0142 & 0,0122 & 0,0038 & 0,9132 & 0,0563 \\ 0,0028 & 0 & 0 & 0,0816 & 0,9154 \end{bmatrix}$$

A continuación se presenta una representación gráfica de la cadena.



4.3. Transición en N pasos

La probabilidad de que partiendo en el estado 1 lleguemos al estado 5 en 3 pasos, esta dada por:

$$P_{1,5}^{(3)} = P(X_{m+n} = 5 | X_m = 1)$$

Entonces,

$$P_{1,5}^{(3)} = \sum_{k=0}^5 P_{1k} P_{k5}^{(2)}$$

De esta forma $P^{(2)}$ es:

$$\begin{bmatrix} 0,7293 & 0,0018 & 0,0098 & 0,2217 & 0,0369 \\ 0,0215 & 0,5584 & 0,0547 & 0,3532 & 0,0119 \\ 0,0026 & 0,0943 & 0,6426 & 0,2524 & 0,0078 \\ 0,0254 & 0,0206 & 0,0071 & 0,8435 & 0,1032 \\ 0,0062 & 0,0010 & 0,0003 & 0,1497 & 0,8427 \end{bmatrix}$$

es así como la probabilidad $P_{1,5}^{(3)} = 0,77$ y además la probabilidad $P_{1,4}^{(3)} = 0,91$, los cuales son valores bastante buenos ya que indican que hay elevadas probabilidades de que estando el día en un estado malo para

surfear, las condiciones mejoren durante el día para que pase a un buen estado.

4.4. Distribución de transición en el límite

A cotinuación se presenta las probabilidades asociadas en el sistema cuando ha pasado un gran número de periodos.

$$\pi_1 = 0,036 \Rightarrow 3,59 \%$$

$$\pi_2 = 0,027 \Rightarrow 2,72 \%$$

$$\pi_3 = 0,033 \Rightarrow 3,3 \%$$

$$\pi_4 = 0,511 \Rightarrow 51,1 \%$$

$$\pi_5 = 0,393 \Rightarrow 39,3 \%$$

Por lo tanto, se espera que en el largo plazo exista con una probabilidad de 51,15% un mar con oleaje tipo 4 y también, un mar tipo 5 con una probabilidad de 39,26% la mayor parte del tiempo en la playa de Pichilemu.

4.5. Análisis de Periodicidad

Los tiempos medios de retorno a cada estado son

$$E_1 = 27,77$$

$$E_2 = 37,03$$

$$E_3 = 30,3$$

$$E_4 = 1,96$$

$$E_5 = 2,54$$

Como se puede ver en los resultados anteriores los estados E_4 y E_5 son los que tienen el tiempo de retorno más bajo, por lo tanto el sistema va a tender a retornar a estos constantemente.

5. Conclusiones

Como se puede apreciar en los resultados, se comprueba de forma empírica que la playa de Pichilemu es una de las mejores de Chile para realizar la disciplina del surf, ya que las probabilidades de tener un buen mar son siempre muy elevadas y su tiende a mantenerse con buenas condiciones, un aspecto relevante es que en el largo plazo estas condiciones se mantendrían constantes, por lo cual es un muy buen incentivo a potenciar esta localidad en torno a la practica de este deporte. Por otro lado sería muy interesante como investigación futura, expandir el horizonte de estudio, además se puede agregar el factor del viento como un parámetro que determine la calidad del mar para surfear.

Referencias

- [Carpinone et al., 2015] Carpinone, A., Giorgio, M., Lange-lla, R., and Testa, A. (2015). Markov chain modeling for very-short-term wind power forecasting. *Electric Power Systems Research*, 122:152–158.
- [Córdoba et al., 2009] Córdoba, K., Delgado, L., Palacios, J., and Aguilar, V. (2009). Análisis de patrones climáticos de eventos enso mediante cadenas de markov y modelos de grafos. *Acta Científica Venezolana*, 60(1-2):5–10.
- [Gabriel and Neumann, 1962] Gabriel, K. and Neumann, J. (1962). A markov chain model for daily rainfall occurrence at tel aviv. *Quarterly Journal of the Royal Meteorological Society*, 88(375):90–95.
- [Markov, 1906] Markov, A. (1906). Rasprostranenie zakona bol'shih chisel na velichiny, zavisyaschie drug ot druga, izvestiya fiziko-matematicheskogo obschestva pri kazanskom universitete, 2-ya seriya 15, 135–156. *Russian. English translation: 'Extension of the law of large numbers to dependent quantities.*
- [Mead and Black, 2015] Mead, S. and Black, K. (2015). *Coastal Research on Surfing*.
- [POSADA,] POSADA, R. O. Las cadenas de markov como medio alterno de proyectar indicadores de calidad en el tiempo.
- [Sawhney, 1962] Sawhney, M. (1962). A study of ocean wave amplitudes in terms of the theory of runs and a markov chain process. Technical report, NEW YORK UNIV BRONX DEPT OF METEOROLOGY AND OCEANOGRAPHY.