Pronósticos con Recomendaciones de Expertos

Quantil S.A.S.

Adrian Visbal - Quantil S.A.S.

Abril - 2015

Introducción

Problema: predecir secuencias individuales $\{y_1, y_2, ...\}$ que son el resultado de algún mecanismo desconocido que puede ser:

- Determinístico.
- Estocástico.
- Resultado de una reacción o adaptación al comportamiento del individuo que está prediciendo la secuencia.

Este análisis se aleja de la aproximación tradicional en la que la secuencia observada es la realización de un proceso estocástico. No tengo una distribución de probabilidad para la secuencia.

- Descripción problema:
  - Predecir secuencia $\{y_1, y_2, ...\}$ de bits $y_t \in \{0, 1\}$.
  - En cada momento t, N expertos hacen sus pronósticos: vector binario $(f_{1,t}, ..., f_{N,t})$, donde $f_{i,t} \in \{0, 1\}$.
  - En cada momento t, el jugador hace su pronóstico $\widehat{p}_t \in \{0, 1\}$, basado en las recomendaciones de los N expertos.
  - Objetivo: acotar el número de errores del jugador $\widehat{p}_t \neq y_t$.
- Suponemos que hay un experto i que no comete errores: $f_{i,t} = y_i, \forall i, t$. No sabemos cuál experto es.
- El jugador empieza asignando un peso $w_j = 1$ a cada experto $j = 1, ..., N$.
- En cada momento t, el jugador predice $\widehat{p}_t = 1$ si el número de expertos j con $w_j = 1$ y $f_{j,t} = 1$ es mayor que el de aquellos con $w_j = 1$ y $f_{j,t} = 0$.

- Al ser revelado y_t: si $\widehat{p}_t \neq y_t$, entonces el jugador asigna el peso $w_k \longleftarrow 0$ a todos aquellos expertos k tales que $f_{k,t} \neq y_t$. El jugador predice teniendo en cuenta los expertos que, hasta el momento t, nunca se han equivocado.
- Sea W_m la suma de los pesos de los expertos después de que el jugador ha cometido m errores. En $t = 0$: $m = 0$ y $W_0 = N$.
- Cuando el jugador comete su error m_{th}, al menos la mitad de los expertos que no se habían equivocado, cometen su primer error. Por lo tanto: $W_m \leq W_{m-1}/2$.
- Esta desigualdad se mantiene $\forall m \geq 1 \Rightarrow W_m \leq W_0/2^m$.

- Experto i no comete errores $(w_i = 1) \Rightarrow W_m \geq 1$.
- $W_0 = N \Rightarrow 1 \leq N/2^m$.
- Solucionando para m se obtiene: $m \leq log_2 N$.
- Se obtiene una cota superior para el número de errores, la cual depende del número de expertos N que se consideren para hacer las pronósticos.

Caso General

- Mismo problema pero sin el supuesto de tener un experto que no cometa errores.
- Nuevo objetivo: relacionar el número de errores que comete el jugador con el número de errores que comete el mejor experto. Y acotar errores.
- Ya no tiene sentido asignar un peso $w_k \longleftarrow 0$ al experto k que cometa un error (no tengo la certeza de que hay uno que nunca se equivocará).
- Cambio: $w_k \longleftarrow \beta w_k$, cada vez que el experto k cometa un error, donde $0 < \beta < 1$ (parámetro libre). Cada vez que un experto cometa un error, se reduce su peso en un factor constante β.

- En cada momento t, el jugador predice $\widehat{p}_t = 1$ si el peso asignado a los expertos j con $f_{j,t} = 1$ es mayor que el de aquellos con $f_{j,t} = 0$.
- Cuando el jugador comete su error m_{th}, el peso de los expertos incorrectos es al menos $W_{m-1}/2$; mientras que el de los expertos correctos es como máximo $W_{m-1}/2$.
- El peso de los expertos incorrectos es multiplicado por β, mientras que el de los correctos permanece igual. Por lo tanto: $W_m \leq W_{m-1}/2 + \beta W_{m-1}/2$.
- Esta desigualdad se mantiene $\forall m \geq 1$, por lo tanto: $W_m \leq W_0(1+\beta)^m/2^m$.

- Sea l el experto que ha cometido la menor cantidad de errores cuando el jugador comete su error m_{th}. Su número mínimo de errores se denota por m^* y, por lo tanto, su peso actual es $w_l = \beta^{m^*}$. Entonces, $W_m \geq \beta^{m^*}$.
- Esto lleva a la desigualdad $\beta^{m^*} \leq W_0(1+\beta)^m/2^m$. Con $W_0 = N$ y resolviendo para m, se obtiene:

$$m \leq \frac{log_2 N + m^* log_2(1/\beta)}{log_2(2/(1+\beta))}$$

- Para un valor dado de β, esta desigualdad establece una relación lineal entre el número de errores m cometidos por el jugador, después de cierta cantidad de pronósticos, y el número de errores cometidos por el mejor experto después de esa misma cantidad de pronósticos.

- Parámetros:
  - $\mathcal{D}$: espacio de decisión convexo.
  - $\mathcal{Y}$: espacio de resultados .
  - $l : \mathcal{D} \times \mathcal{Y} \rightarrow R$: función de perdida (no negativa).
  - $\mathcal{E}$: conjunto de expertos. Número finito de expertos $(1, ..., N)$.
- Para cada período $t = 1, 2, ...$:
  1. Los expertos hacen sus pronósticos $\{f_{E,t} \in \mathcal{D} : E \in \mathcal{E}\}$ y son revelados al jugador (agente).
  2. El jugador hace su predición $\widehat{p}_t \in \mathcal{D}$ una vez conocidos los pronósticos de los expertos.
  3. La naturaleza (adversario) escoge el próximo resultado $y_t \in \mathcal{Y}$.
  4. El jugador incurre en una pérdida $l(\widehat{p}_t, y_t)$ y cada experto E en una pérdida $l(f_{E,t}, y_t)$.

- La pérdida acumulada del experto E hasta el período n ($L_{E,n}$) se define como:

$$L_{E,n} = \sum_{t=1}^{n} l(f_{E,t}, y_t)$$

Análogamente, la pérdida acumulada del jugador es: $\widehat{L}_n$.

- El arrepentimiento (acumulado) o *regret* $R_{E,n}$ del jugador con respecto al experto E, hasta el período n, se define como:

$$R_{E,n} = \widehat{L}_n - L_{E,n}$$

Y el arrepentimiento instantáneo se define como:

$$r_{E,n} = l(\widehat{p}_t, y_t) - l(f_{E,t}, y_t) \Rightarrow R_{E,n} = \sum_{t=1}^{n} r_{E,t}$$

- Objetivo del jugador: minimizar el arrepentimiento acumulado $R_{E,n}$ respecto a cada experto E, para toda realización posible de resultados.
- Por ejemplo: el objetivo podría ser que para toda secuencia de resultados:

$$\lim_{n\to\infty} \max_{i=1,\dots N} \frac{R_{i,n}}{n} = 0$$

O de forma equivalente:

$$\lim_{n\to\infty} \frac{1}{n} \left(\widehat{L}_n - \min_{E\in\mathcal{E}} L_{E,n} \right) \to 0$$

Predicciones Ponderadas

Una estrategia natural del jugador para hacer sus predicciones es asignar un promedio ponderado a los pronósticos de los expertos:

$$\widehat{p}_t = \frac{\sum_{i=1}^{N} w_{i,t-1} f_{i,t}}{\sum_{i=1}^{N} w_{i,t-1}}$$

donde $w_{i,t-1} \geq 0$ son los pesos que se le asocian a cada experto en el momento t. La notación hace referencia a que los pesos deben escogerse con base en la información revelada hasta $t-1$.

$\widehat{p}_t \in \mathcal{D}$: combinación de $f_{i,t} \in \mathcal{D}$.

Si $R_{i,t-1}$ es grande entonces el experto i tiene una pérdida acumulada pequeña, por lo tanto, su peso asignado ($w_{i,t-1}$) debe ser grande. En particular (por conveniencia), vamos a suponer que:

$$w_{i,t-1} = \phi'(R_{i,t-1})$$

donde $\phi : R \to R$ es una función no negativa, convexa y creciente.

$$\Rightarrow \widehat{p}_t = \frac{\sum_{i=1}^{N} \phi'(R_{i,t-1}) f_{i,t}}{\sum_{i=1}^{N} \phi'(R_{i,t-1})}$$

Alternativamente, se definirán cierto tipo de predicciones basadas en una función potencial.

Definición (Predicciones Basadas en un Potencial)

Definamos la función potencial $\Phi : R^N \to R$ *de la forma:*

$$\Phi\left(\mathbf{u}\right) = \psi\left(\sum_{i=1}^{N}\phi(u_i)\right)$$

donde $\phi : R \to R$ *es una función no negativa, convexa, creciente y dos veces diferenciable y* $\psi : R \to R$ *es una función no negativa, estrictamente creciente, cóncava y dos veces diferenciable. Una predicción basada en el potencial* Φ *es una predicción de la forma:*

$$\widehat{p}_t = \frac{\nabla\Phi\left(\mathbf{R}_{t-1}\right)\cdot f_t}{\sum_{i=1}^{N}\nabla\Phi\left(\mathbf{R}_{t-1}\right)_i} \;, donde\ \mathbf{R}_{t-1} = \left(R_{1,t-1}, ..., R_{N,t-1}\right).$$

$$\Phi(\mathbf{R}_{t-1}) = \psi\left(\sum_{i=1}^{N}\phi(R_{i,t-1})\right)$$

$$\Rightarrow \nabla\Phi\left(\mathbf{R}_{t-1}\right)_j = \frac{\partial\Phi(\mathbf{R}_{t-1})}{\partial R_{j,t-1}}$$

$$\nabla\Phi\left(\mathbf{R}_{t-1}\right)_i = \psi'\left(\sum_{i=1}^{N}\phi(R_{i,t-1})\right)\phi'(R_{j,t-1})$$

$$\Rightarrow \widehat{p}_t = \frac{\sum_{i=1}^{N}\phi'(R_{i,t-1})f_{i,t}}{\sum_{i=1}^{N}\phi'(R_{i,t-1})} \rightarrow \text{Independiente de } \psi$$

Lema (Condición de Blackwell)

Si la función de pérdida es convexa en el primer argumento entonces:

$$\sup_{y_t \in \mathcal{Y}} \mathbf{r}_t \cdot \nabla\Phi\left(\mathbf{R}_{t-1}\right) \leq 0$$

Prueba.

$$\sup_{y_t \in \mathcal{Y}} \mathbf{r}_t \cdot \nabla\Phi\left(\mathbf{R}_{t-1}\right) \leq 0 \Leftrightarrow \sup_{y_t \in \mathcal{Y}} \sum_{i=1}^{N} r_{i,t} \phi'(R_{i,t-1}) \leq 0$$

Usar desigualdad de Jensen para llegar a la segunda expresión. ■

El siguiente teorema aplica a cualquier tipo de pronóstico que satisfaga la condición de Blackwell y no únicamente a pronósticos ponderados. No requiere que ϕ sea convexa.
Sea $\mathbf{r}_t = (r_{1,t}, ..., r_{N,t})$ el vector de arrepentimiento instántaneo.

Teorema

Supongamos que los pronósticos de un jugador satisfacen la condición de Blackwell para cierto potencial Φ, entonces:

$$\Phi\left(\mathbf{R}_n\right) \leq \Phi(\mathbf{0}) + \frac{1}{2}\sum_{i=1}^{n} C(\mathbf{r}_t)$$

donde,

$$C(\mathbf{r}_t) = \sup_{u \in R^N} \psi'\left(\sum_{i=1}^{N} \phi(u_i)\right) \sum_{i=1}^{N} \phi''(u_i) r_{i,t}^2$$

Una aplicación del Teorema 1 es: Supongamos que ϕ es invertible. Entonces, como ψ es invertible por definición de función potencial, entonces:

$$\begin{aligned}
\psi\left(\phi\left(\max_{i=1,\dots,N} R_{i,n}\right)\right) &\leq \psi\left(\sum_{i=1}^{N}\phi\left(R_{i,n}\right)\right) = \Phi(\mathbf{R}_n) \\
&\Rightarrow \\
\max_{i=1,\dots,N} R_{i,n} &\leq \phi^{-1}\left(\psi^{-1}(\Phi(\mathbf{R}_n))\right)
\end{aligned}$$

y $\Phi(\mathbf{R}_n)$ se puede limitar por la cota que arroja el teorema.

Ejemplo (Potencial Polinomial)

Considere el siguiente potencial denominado el potencial **polinomial***:*

$$\Phi_p(\mathbf{u}) = \left(\sum_{i=1}^{N} (u_i)_+^p \right)^{2/p} = \|\mathbf{u}\|_+^p$$

donde $p \geq 0$*. Si calculamos los pesos, podemos ver que la predicción satisface:*

$$\widehat{p}_t = \frac{\sum_{i=1}^{N} (R_{i,t-1})_+^{p-1} f_{i,t}}{\sum_{i=1}^{N} (R_{i,t-1})_+^{p-1}}$$

Corolario (Cotas para predictor polinomial)

Supongamos que la función de pérdida l es convexa en su primer argumento y toma valores en $[0,1]$. Entonces para todo $n \geq 1$, $p \geq 2$ y realización $y_1, ..., y_n \in \mathcal{Y}$, el arrepentiemiento del predictor polinomial satisface:

$$\widehat{L}_n - \min_{i=1,...N} L_{i,n} \leq \sqrt{n(p-1)N^{2/p}}$$

Si $p = 2 \Rightarrow \widehat{L}_n - \min_{i=1,...N} L_{i,n} \leq \sqrt{nN}$

Optimizar esta cota superior para el predictor polinomial, sugiere usar $p = 2ln(N)$

$$\Rightarrow \widehat{L}_n - \min_{i=1,...N} L_{i,n} \leq \sqrt{n(2ln(N)-1)e}$$

Ejemplo (Potencial Exponencial)

Considere el siguiente potencial denominado el potencial **exponencial***:*

$$\Phi_\eta(\mathbf{u}) = \frac{1}{\eta} \ln\left(\sum_{i=1}^{N} e^{\eta\mu_i}\right)$$

donde $\eta > 0$*. Si calculamos los pesos, podemos ver que la predicción satisface:*

$$\widehat{p}_t = \frac{\sum_{i=1}^{N} e^{-\eta L_{i,t-1}} f_{i,t}}{\sum_{i=1}^{N} e^{-\eta L_{i,t-1}}}$$

- Ventajas:
  - La predicción solo depende del rendimiento pasado de los expertos y no de las predicciones pasadas del jugador como sí es el caso con predictores basados en otro potenciales.
  - Los pesos satisfacen una recursión simple.

$$w_{i,t} = \frac{w_{i,t-1} e^{-\eta l(f_{i,t}, y_{i,t})}}{\sum_{i=1}^{N} w_{i,t-1} e^{-\eta l(f_{i,t}, y_{i,t})}}$$

- Desventajas:
  - La elección óptima del parámetro η depende de n (número total de rondas de predicción).

Corolario (Cotas para predictor exponencial)

Supongamos que la función de pérdida l es convexa en su primer argumento y toma valores en $[0,1]$. Entonces para todo $n \geq 1$, $\eta > 0$ y realización $y_1, ..., y_n \in \mathcal{Y}$, el arrepentiemiento del predictor exponencial satisface:

$$\widehat{L}_n - \min_{i=1,\ldots N} L_{i,n} \leq \frac{\ln(N)}{\eta} + \frac{n\eta}{2}$$

Optimizar esta cota superior para el predictor exponencial, sugiere usar $\eta = \sqrt{\frac{2\ln(N)}{n}}$

$$\Rightarrow \widehat{L}_n - \min_{i=1,\ldots N} L_{i,n} \leq \sqrt{2n\ln(N)}$$

La cota óptima encontrada para el predictor exponencial, puede ser mejorada a través del siguiente teorema:

Teorema

Supongamos que la función de pérdida l es convexa en su primer argumento y toma valores en $[0,1]$. Entonces para todo $n \geq 1$, $\eta > 0$ y realización $y_1, ..., y_n \in \mathcal{Y}$, el arrepentiemiento del predictor exponencial satisface:

$$\widehat{L}_n - \min_{i=1,\dots N} L_{i,n} \leq \frac{\ln(N)}{\eta} + \frac{n\eta}{8}$$

$$\text{Si } \eta = \sqrt{\frac{8\ln(N)}{n}} \Rightarrow \widehat{L}_n - \min_{i=1,\dots N} L_{i,n} \leq \sqrt{\frac{n}{2}\ln(N)}$$

"Double-tricking"

La cota del predictor exponencial depende del parámetro η, por lo tanto no se mantiene uniforme para secuencias de resultados de cualquier longitud, sino para una longitud determinada n. Esta desventaja puede solucionarse a través de una técnica conocida como "double-tricking":

- Dividir el tiempo en períodos cuyas longitudes sean crecientes exponencialmente. Dentro de cada período se tienen varias rondas de predicción.
- En cada período se optimiza η para la longitud correspondiente.
- Los pesos pueden hallarse recursivamente dentro de cada período. Al finalizar el período, se reinician los pesos y se vuelve a optimizar η para el intervalo correspondiente.

Usando la técnica de "double-tricking", para todo $n \geq 1$ y realización $y_1, ..., y_n \in \mathcal{Y}$, el arrepentiemiento del predictor exponencial satisface:

$$\widehat{L}_n - \min_{i=1,\dots N} L_{i,n} \leq \frac{\sqrt{2}}{\sqrt{2}-1}\sqrt{\frac{n}{2}\ln(N)}$$

Esta cota es mayor que la cota óptima encontrada en el Teorema 2, por el factor $\frac{\sqrt{2}}{\sqrt{2}-1} \approx 3{,}41$.
Para mejorar esta cota, se evita la técnica de "double-tricking", y se recurre a funciones potenciales cuyos parámetros varían en el tiempo.

Potencial variable en el tiempo

Cuando $\eta = \sqrt{\frac{8\ln(N)}{n}}$ obtengo la mejor cota no-uniforme para el predictor exponencial. Por lo tanto, una forma de escoger el potencial exponencial con parámetro variable en el tiempo es:

$$\eta_t = \sqrt{\frac{8\ln(N)}{t}}$$

donde t es el número de la ronda de predicción en que nos encontramos. La cota óptima está dada por el siguiente teorema:

Teorema

Supongamos que la función de pérdida l es convexa en su primer argumento y toma valores en $[0,1]$. Entonces para todo $n \geq 1$, $\eta_t = \sqrt{\frac{8\ln(N)}{t}}$ y realización $y_1, ..., y_n \in \mathcal{Y}$, el arrepentiemiento del predictor exponencial satisface:

$$\widehat{L}_n - \min_{i=1,...N} L_{i,n} \leq 2\sqrt{\frac{n}{2}\ln(N)} + \sqrt{\frac{\ln(N)}{8}}$$

Esta cota es menor que la encontrada a través de la técnica de "double-tricking".

Aplicación en Quantil

Modelo de optimización de portafolio (CARM): generación de views como uno de los inputs fundamentales del modelo.

- Variable a pronosticar: precio objetivo de cada acción (view).
- Expertos: analistas (empresas) que reportan precios objetivos para cada acción.

Por definir:

- Tipo de predictor.
- Función de pérdida para evaluar precisión de los pronósticos.
- Horizonte de evaluación de los pronósticos (¿Precio objetivo a cuánto tiempo?).

Tipo de predictor

Predictor exponencial con parámetro variable en el tiempo:
El precio objetivo en cada ronda de predicción (p_t) estará dado por:

$$\widehat{p}_t = \frac{\sum_{i=1}^{N} e^{-\eta_t L_{i,t-1}} f_{i,t}}{\sum_{i=1}^{N} e^{-\eta_t L_{i,t-1}}}, \; \eta_t = \sqrt{\frac{8\ln(N)}{t}}$$

Donde,

- N: número total de analistas (empresas) por acción.
- $f_{i,t}$: precio objetivo del analista i en cada ronda de predicción.
- $L_{i,t-1}$: pérdida acumulada del analista i hasta la ronda de predicción $t-1$.

Función de pérdida

Según revisión de literatura, algunas medidas propuestas para determinar la precisión en los precios objetivos de los analistas son:

- Bilinski et al.(2012) - Target Price Accuracy: International Evidence.
  - $Met - Any$: variable indicadora que es igual a 1 si el precio de la acción toca el precio objetivo en cualquier momento hasta un período de 12 meses.
  - $aTPE = \frac{|TP - P_{rev}|}{P_0}$,
    donde TP:precio objetivo, P_{12}:precio de cierre en 12 meses, P_0:precio de cierre en fecha de emisión del precio objetivo.

- $Met - Any - rev$: variable indicadora que es igual a 1 si el precio de la acción toca el precio objetivo desde la emisión del precio objetivo hasta el momento de la actualización.
- $aTPE - rev = \frac{|TP - P_{12}|}{P_0}$,
  donde TP:precio objetivo, P_{rev}:precio de cierre en fecha de actualización, P_0:precio de cierre en fecha de emisión del precio objetivo.

- Bradshaw et al.(2012) - Analyst Target Prices and Forecast Accuracy around the World.
  - Clasifican los pronósticos de precio objetivo en: compra (si $TP > P_0$) y venta (si $TP \leq P_0$).
  - $Acc12$: porcentaje de días, entre la fecha de emisión del precio objetivo y la fecha en la que se cumpla el horizonte de 12 meses, en los que el precio de la acción es mayor que el precio objetivo si es señal de compra (o menor si es señal de venta).
  - También lo analizan para un horizonte de 3 y de 6 meses (para tener en cuenta las actualizaciones del precio objetivo).
  - $TPMet = Met - Any$.

- Kerl (2010) - Target Price Accuracy.
  - $AM = \frac{|TP - P_{12}|}{TP}$
  - $AM - Adj = 1 - \frac{\frac{|TP - P_{12}|}{TP}}{Volatility}$
  - Estas medidas pueden ajustarse en caso de una actualización del precio objetivo del analista, según lo propuesto por Bilinski et al.(2012).

Sugerencia: realizar ejercicio de pronóstico con las medidas: $aTPE$(Bilinski), AM y $AM - Adj$ (Kerl), teniendo en cuenta el ajuste en caso de actualización del precio objetivo antes del horizonte de 12 meses, propuesto en Bilinski (2012).

- Cesa-Bianchi N., Lugosi G., (2006). Prediction, Learning, and Games.
- Bilinski P., Lyssimachou D., Walker M., (2012). Target Price Accuracy: International Evidence.
- Kerl A., (2010) - Target Price Accuracy
- Bradshaw M., Huang A., Tan H., (2012) - Analyst Target Prices and Forecast Accuracy around the World