

Comparison of Classic Bootstrap Regression and Bayesian Bootstrap Regression Methods

-JAYO ACUÑA BRUCE

Asignatura: Laboratorio de Control
Escuela Profesional de Ingeniería Industrial
Facultad de Ingeniería Industrial y de Sistemas
Universidad Nacional de Ingeniería

Resumen

Se quiere saber como influyen las variables que se dan en el campo de juego en la diferencia de goles entre el local y el visitante en un partido de la liga BBVA. Se construirá dos modelos con los enfoques señalados en función a la data de la liga 2012-2013 y se vera la performance de este modelo para la otras temporadas. El modelo del enfoque clásico aplicara bootstrap a una regresión robusta. El modelo bootstrap bayesiano nos dara fuertes evidencias para el valor explicativo del modelo y el modelo bajo el enfoque clásico será superior en performance de predicción

INTRODUCCIÓN

Se sabe que el futbol es un deporte muy dinámico y que lo que más influye en el resultado de un partido es lo que se da en el mismo campo de juego. Por eso hemos tomado las estadísticas de cada partido de la liga BBVA temporada 2012-2013 y ver como estas predicen y explican la diferencia de goles entre el local y el visitante. Trabajos anteriores también han modelado este diferencia, e incluso Eva María García Quintero(2014) ha modelado esto en la Liga BBVA bajo una regresión Poisson Bivariada. El enfoque bootstrap que plantearemos, tiene como ventaja que esta libre de supuestos de alguna distribución debido a que estima una distribución empirica de los datos, y con el enfoque bayesiano que se le dara, veremos como puede cambiar los resultados a priori.

PRESENTACIÓN DEL PROBLEMA

Al final de cada partido por lo general sus estadísticas del juego respectivas no son muy analizadas o son tomadas para el anécdota , pero ahora se busca saber realmente como influyen estas en el resultado del mismo. Por tanto nos plantemos ¿Cuáles son las Variables que se dan en el campo de juego que más influyen en el resultado y que información nos pueden proveer en un partido determinado en vista de que estas cifras por lo general son poco analizadas?.

OBJETIVOS

Objetivo General:

Explicar y predecir la Diferencia de Goles entre un Equipo Local y un Equipo Visitante en la Liga BBVA tomando como referencia la temporada 2012-2013 mediante los Modelos de Regresión Lineal Bootstrap y Bootstrap Bayesiana.

DESCRIPCIÓN DE LA SOLUCIÓN

-Se empleara una regresión robusta de estimadores tipo MM y luego el respectivo procedimiento bootstrap bajo el enfoque clásico, la regresión robusta se usa con el objetivo de controlar el efecto de los outlier.

-En el caso del bootstrap bayesiano, se empleara distribuciones a priori para los coeficientes con distribución multinormal

RESULTADOS

- El bootstrap clásico regresión robusta da un error cuadrático medio para la data de prueba de 0.084 y el del enfoque bayesiano es 1.726316. Los signos de los coeficientes son similares.

-Los signos de los coeficientes resultan acorde a sus distribuciones a priori, excepto para las variables Tiros al arco del local, faltas del local y corners del local. El valor negativo a posteriori del número de tiros al arco del local es incluso un valor de muy baja probabilidad para su distribución a priori, lo cual nos indica una fuerte evidencia para la negatividad de este coeficiente. Se podría explicar que esto sucede porque en el caso del local, mientras este dispare mas al arco y como el gol no esta contabilizado como tira al arco, esto denota poca efectividad. En el caso de las faltas del local, este también es un valor de muy baja probabilidad para su distribución a priori y por tanto representa una fuerte evidencia, que a medida que aumentan las faltas del local, esto juega a favor del local. En el caso del corners del local, este también es un valor extremo y también se podría explicar que más corners marcados para el local es un indicador de ineffectividad.

- En vista de que por partido casi siempre se dan más tiros a puerta que tarjetas rojas y los coeficientes de tarjetas rojas tanto como para el local y visitante comparados con los coeficientes de tiros a puerta tanto para el local como y visitante son similares. Entonces la variable tiros a puerta es la que más influye en el resultado de un partido, y no los tiros al arco por las razones explicadas arriba.

- Vemos que la bondad de nuestro modelo reside en partidos donde hay ganador, en ese caso el modelo es bien eficiente para predecir si ganara el local o el visitante.

CONCLUSIONES

-Los coeficientes para las regresiones bootstrap clásica y bayesiana son iguales en signo. En consecuencia, el bootstrap regresión bayesiana nos da fuerte evidencia en las variables que sus valores observados a priori eran valores de muy poca probabilidad.

-Con respecto a predicción nos inclinamos por el bootstrap clásico con regresión robusta debido a que arroja un error cuadrático medio en la data de prueba de 0.084. Esto posiblemente sea, es que en el efecto de los outliers en las muestras bootstrap haya afectado a las estimaciones, y justamente la regresión robusta controla este efecto.

- La bondad del modelo reside en partidos donde ha habido un ganador, sobre todo si queremos saber si gana el local o la visita , más allá de la diferencia de goles.

BIBLIOGRAFÍA

<http://www.sciencedirect.com/science/article/pii/S0304407602001963>

<http://ageconsearch.umn.edu/bitstream/18786/1/dp960002.pdf>

<http://statweb.stanford.edu/~tibs/sta305files/FoxOnBootingRegInR.pdf>

<https://socserv.socsci.mcmaster.ca/jfox/Books/Companion/appendix/Appendix-Bootstrapping.pdf>

http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_751.pdf

http://projecteuclid.org/download/pdf_1/euclid.aos/1021379865

INDICACIONES GENERALES

- El informe se presenta en doble columna y con una longitud entre 6 y 10 páginas.
- Tamaño de hoja: A4.
- El tipo de la letra es arial.
- El tamaño de la letra:
 - Título: 16, en negritas.
 - Nombre de autores, asignatura, facultad y universidad: 10
 - Título de cada sección: 11, en mayúsculas y en negritas.
 - Texto general: 10.
- Todas las figuras, tablas, gráficos debes estar centrados y explicados en el texto. Deben tener una leyenda numerada y explicativa en 1 o 2 líneas.
- Todas las ecuaciones debe estar centradas y numeradas.
- La redacción debe realizarse en tercera persona. Ejemplo: se diseñó, se analizó, se probó, se construyó. No debe redactar en primera persona: diseñé, diseñamos, analicé, analizamos, construí, construimos.
- Márgenes:
 - Superior : entre 1.5 y 2.0 cm. Página inicial: 3.0 cm.
 - Inferior: entre 1.0 y 1.5 cm.
 - Izquierda: entre 1.0 y 1.5 cm.
 - Derecha: entre 1.0 y 1.5 cm.
 - Separación entre columnas: entre 0.75 y 1.0 cm

Entregar a los miembros del jurado una copia impresa. Entregar además, un CD con el informe en formato WORD y PDF, así como las diapositivas, fotos y videos de la presentación.

EJEMPLOS DE GRÁFICOS Y FIGURAS

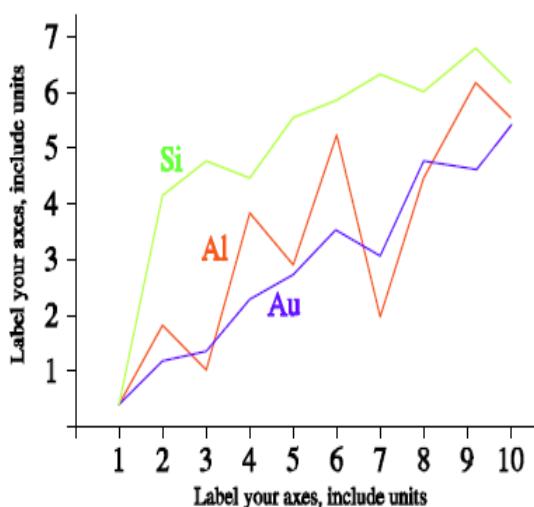


Fig. 1 El ejemplo de un gráfico con colores sólidos que resaltan sobre el fondo blanco.



Fig. 3 Ejemplo de figura con buena resolución.