



**FACULTAD
DE INGENIERIA**

Universidad de Buenos Aires

Teoria del Lenguaje:

Informe individual

Ignacio Janeiro - 103550



Paquete AnomalyDetection R

Introducción:

AnomalyDetection es un paquete de Código abierto desarrollado por twitter, que tiene el objetivo de detectar anomalías. Cabe destacar que este paquete desde un punto de vista estadístico es de lo más robusto que se puede encontrar hoy en día.

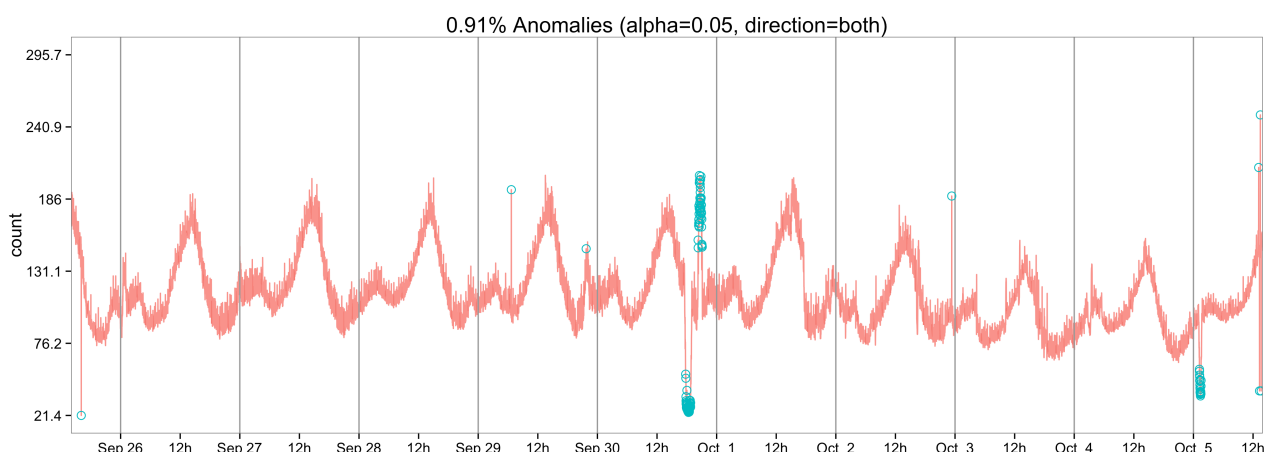
Este paquete se puede utilizar en una amplia variedad de contextos. Por ejemplo, la detección de anomalías de las métricas del sistema luego de una nueva version de software, la participación del usuario después de una prueba A/B o por problemas en econometría, ingeniería financiera, ciencias políticas y sociales.

Como funciona el paquete:

El algoritmo conocido como SH-ESD, se basa en la prueba ESD generalizada para detectar anomalías. Tenga en cuenta que SH-ESD se puede utilizar para detectar anomalías tanto globales como locales. Esto se logra empleando la descomposición de series de tiempo y utilizando métricas estadísticas sólidas, a saber, la mediana junto con la ESD.

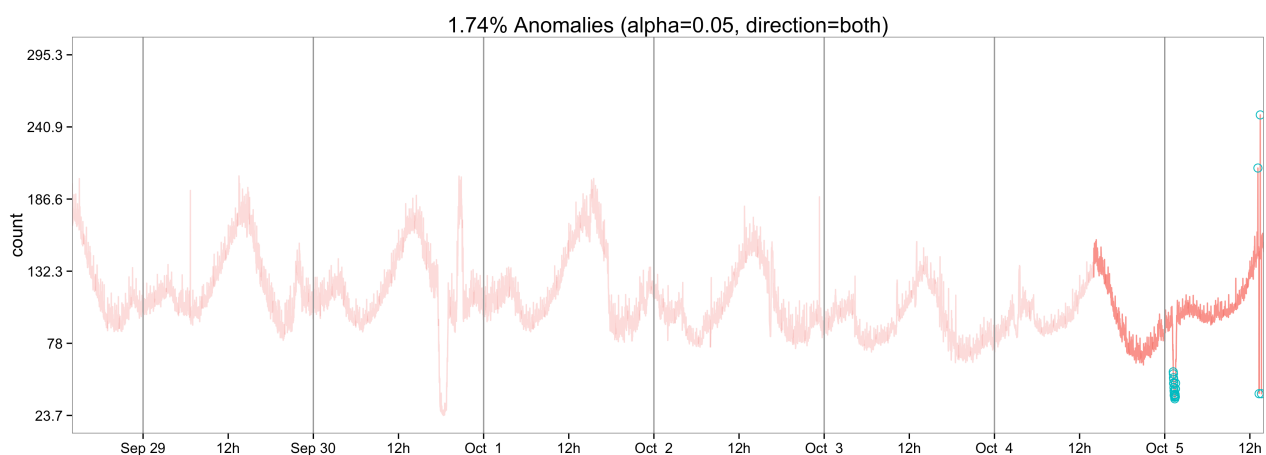
Un ejemplo:

Para este ejemplo, utilizamos un conjunto de datos que ya viene de ejemplo en este paquete. Y aplicándole la función AnomalyDetectionTs a este mismo, podemos llegar a detectar los diferentes tipos de anomalías.



A partir del gráfico, observamos que la serie temporal de entrada experimenta tanto anomalías positivas como negativas. Además, estas anomalías son locales (por eso, no se puede detectar utilizando los enfoques tradicionales). Para finalizar el análisis de la primera figura cabe aclarar que las anomalías detectadas mediante esta propuesta son las marcadas en el gráfico anteriormente mostrado.

También es importante mencionar que a menudo, la detección de anomalías se lleva a cabo de forma periódica. Por ejemplo, a veces, uno puede estar interesado en determinar si ayer hubo alguna anomalía. Para eso la función `AnomalyDetection` contiene un parámetro específico llamado `only_last`, el cual si le asignamos el valor "day", nos puede devolver las últimas anomalías ocurridas en el último día.



Acá lo que podemos observar que solo se han anotado las anomalías ocurridas durante el último día. Pero, también muestra en un segundo plano el resto de la serie para poder tener un panorama más general.

Conclusión:

Como vimos en el trabajo y en el ejemplo recién mencionado. R es una herramienta muy utilizada para el análisis estadístico, usada por importantes compañías que manejan grandes volúmenes de datos. Cómo puede llegar a ser este ejemplo, el cual estamos analizando un paquete desarrollado por la empresa Twitter.