1. Motivación

El análisis de regresión es un conjunto de técnicas de modelado y análisis numérico la cual consiste en determinar el valor de una variable dependiente a partir de un conjunto de variables independientes. Esta técnica es utilizada en pronósticos, inferencia, testeo de hipótesis y evaluacion de causalidad.

El aprendizaje supervisado es una técnica de aprendizaje de maquina la cual consiste en aprender una función a partir de un conjunto de datos de entrenamiento. Cada elemento de este conjunto consiste en un par (datos de entrada(X), salida esperada(Y)). La función aprendida es utilizada luego solo con datos de entrada(X) para que la misma produzca una salida(Y). Cuando el valor de la función de salida es continuo la tarea de aprendizaje se denomina regresión y cuando el valor es una etiqueta se denomina clasificación. Existe una basta cantidad de algoritmos de regresión. Ejemplos de alguno de ellos son: .

La eficacia de los diferentes algoritmos esta estrechamente con el problema subyacente que se quiere abarcar.

Nuestra propuesta consiste en la evaluación de un conjunto de algoritmos de regresión en el marco de un caso de estudio especifico. A continuación describiremos el mismo.

1. Caso de estudio

En la actualidad existen una variedad de deportes acuaticos que se ven beneficiados por los pronósticos acerca del oleaje en una zona cercana a la costa. Ejemplos de estos pueden ser el surf, windsurf, kitesurf, kayac… Los deportistas están sumamente atentos a las condiciones que presenta el mar para evaluar la playa en la que van a practicar el deporte.

Utilizando un pronostico de oleaje el deportista puede planificar sus actividades para lograr estar presente en el lugar con mejores condiciones. Eventualmente el deportista puede evitarse recorridas extensas buscando el sitio con mejores condiciones del dia.

Actualmente los pronósticos de oleaje son derivados a partir de la resolución de un modelo matemático global sumamente complejo. Este provee de la predicción de la altura de las olas la dirección de las mismas y el tiempo entre una ola y otra. Si bien este pronóstico tiene buen rendimiento en alta mar, no sucede lo mismo en la cercanía de la costa. Esto se debe a que la interacción de las olas con el lecho marino cercano no es contemplado por estos modelos. De acuerdo a la geografía de la costa puede suceder que en un rango de 15 km un lugar presente condiciones apropiadas para la práctica del deporte mientras que en otro las condiciones sean desfavorables. Por esta razón es que la utilización de un pronóstico preciso en las cercanías de la costa es de suma utilidad. Existen modelos para contemplar el comportamiento de las olas en las cercanías de la costa. Los mismos reciben como entrada los datos de pronostico de alta mar e información acerca del lecho marino. Utilizando estos datos los mismos proveen un refinamiento al pronostico inicial. Si bien estos modelos poseen un buen comportamiento, su limitación esta relacionada con la obtención de los datos del lecho marino. Usualmente la obtención de estos datos es costosa y no se encuentra disponible en todos los lugares.

Nuestra propuesta busca generar un pronostico de oleaje con precisión en la cercanía de la costa que no necesite utilizar información del lecho marino para su construccion.

Trabajos relacionados podemos encontrar ( ), en ambos trabajos se utilizan datos del pronostico de alta mar junto con observaciones de la altura de la olas para entrenar una red neuronal que pueda refinar el pronostico de alta mar. En 1) Se entrena una red neuronal del tipo feed forward y se realiza una comparación entre un método numérico para refinamiento del modelo de alta mar y un método de redes neuronales, dando este ultimo una mejor precisión en los pronósticos. En 2) se experimenta con un tipo de redes neuronales para poder calcular la distribución de probabilidad de la altura de la ola, en vez de un valor único.

Esto está bien, pero tienen que contar los otros trabajos que hay sobre el tema. Los artículos que vimos al prencipio que hacian predicción de oleaje en base a redes neurales los tienen que contar muy brevemente y referenciarlos.

Nuestra propuesta consiste en utilizar modelos de aprendizaje de maquina para aprender de casos pasados como un determinado pronostico de alta mar se vio reflejado en las cercanías de la costa. De esta manera ante un nuevo pronostico de alta mar se podrá predecir cual será el efecto del mismo en las cercanías de la costa. Para realizar esta predicción utilizaremos métodos de regresión los cuales tendrán como datos de entrada: el pronostico de alta mar y como valor esperado, la observación de la altura en la cercanía de la costa. La altura de la ola será medida por un observador desde la costa.

Referencias

[ ] **Predicting probability distributions for surf height using an ensemble of mixture density networks,** Michael Carney, P´adraig Cunningham, Jim Dowling and Ciaran Lee, **ACM International Conference Proceeding Series; Vol. 119.**

[ ]