**Conclusiones**

**Resumen**

El aprendizaje de maquina es una disciplina de la inteligencia artificial dedicada al desarrollo de algoritmos que permiten a la computadora aprender a partir de la experiencia. Dentro del área de aprendizaje de máquina el aprendizaje supervisado es una técnica que consiste en aprender una función a partir de un conjunto de datos de entrenamiento. El conjunto de entrenamiento consiste en pares de objetos de entrada y salida esperada, es decir cada par es una instancia y una clase a la que esa instancia pertenece. Una instancia representa un conjunto de observaciones sobre un hecho determinado. Una clase representa la clasificación que se le da al hecho. A partir de un conjunto de ejemplos y un algoritmo de aprendizaje supervisado podemos entrenar un clasificador para que el mismo pueda predecir cuál será la clase de una instancia que desconocemos.

El resultado del proceso de aprendizaje es un clasificador, el cual representa un concepto aprendido a partir de los ejemplos. El clasificador obtenido es resultado de un proceso inductivo y como tal su efectividad no está garantizada. Además de evaluar el clasificador con el sentido crítico del usuario es necesario establecer métricas claras del desempeño del mismo. Para esto se prueba el clasificador con instancias cuya clase se conoce previamente y se comparan la clase verdadera y la predicha.

Dentro del aprendizaje supervisado existen diferentes métodos de predicción, los cuales son más o menos eficientes de acuerdo a la naturaleza de los datos del problema en cuestión. Ejemplos de ellos son las maquinas de soporte vectorial, redes neuronales, árboles de decisión, regresión lineal, etc.

La herramienta WEKA contiene una colección de algoritmos para el análisis de datos y modelado predictivo para hacer experimentos de aprendizaje automático. Provee, entre otros, los métodos de predicción mencionados en el párrafo anterior, para experimentación e investigación. Dicha herramienta toma como entrada un conjunto de instancias (cada instancia se compone de un conjunto de variables predictoras y uno o más atributos de clase que representan las variables a predecir) y ejecuta el algoritmo correspondiente. El resultado de este proceso es un clasificador entrenado el cual es capaz de predecir el atributo de clase (variable a predecir) suministrándole únicamente las variables predictoras que componen una instancia. Esta herramienta fue utilizada en este proyecto para el entrenar los clasificadores que utiliza el sistema desarrollado para realizar predicciones.

**Conclusiones del proyecto**

En el proyecto presentado utilizamos el técnicas de aprendizaje supervisado para desarrollar un sistema capaz de pronosticar olas en las cercanías de la costa de cualquier lugar del mundo basándonos en lecturas de altamar provistas por un sistema de pronósticos global denominado WAVEWATCH III desarrollado por el NOAA (National Oceanic and Atmospheric Administration) y en observaciones visuales suministradas por un observador experto en la ubicación exacta de la que se espera obtener pronósticos. El resultado del estudio fue un sistema denominado “Surf-Forecaster” en el cual los usuarios pueden dar de alta sus olas preferidas en cualquier lugar y obtener pronósticos en dichas ubicaciones con hasta una semana de anticipación en incrementos de tres horas, siempre y cuando se le provea al sistema de un histórico de observaciones visuales tomadas en dicha ubicación.

El sistema cuenta con un clasificador entrenado generado con WEKA para poder realizar las predicciones en la ubicación deseada por el usuario. El algoritmo utilizado fueron las Maquinas de Soporte Vectorial debido a que fue el que mejor rendimiento demostró luego de varias etapas de evaluaciones y experimentaciones (Ver capitulo 4 - Evaluaciones).

Finalmente para evaluar el sistema desarrollado utilizamos información de diferentes playas ubicadas alrededor de la isla de Oahu en Hawái, más específicamente en las playas Sunset, Diamond Head, Makapu, Makaha y Ala Moana, de las cuales disponíamos de una base de datos histórica de observaciones visuales con más de seis años de registros con periodicidad diaria. Los resultados obtenidos fueron comparados, con muy buenos resultados, con los publicados en diferentes sitios web locales de la isla, orientados a deportes acuáticos como puede ser el surf, windsurf, kite-surf, etc. Estos sitios dan información al lector de las condiciones del mar para la práctica de estos deportes en distintas playas de Oahu. Con este enfoque logramos demostrar que el uso de algoritmos de aprendizaje de maquina son una herramienta efectiva y viable para realizar pronósticos de oleaje y del comportamiento del mar, mejorando también los resultados e información ofrecidos por modelos matemáticos desarrollados para el mismo fin.

**Ventajas del enfoque**

El sistema desarrollado, Surf-Forecaster, presenta un gran número de ventajas respecto de los sistemas de pronósticos de oleaje existentes actualmente:

* Se logro predecir la altura de las olas exactamente en la ubicación del observador, lo que le da al usuario mayor certeza de las condiciones del mar donde a este realmente le interesa, dado que otros pronosticadores dan esta información pero en una zona alejada de la costa ya sea a partir de mediciones de boyas en altamar o de mediciones brindadas por algún sistema como puede ser WAVEWATCH III en un GridPoint especifico.
* Con Surf-Forecaster llevamos la predicción a una granularidad de “Ola” a diferencia de los pronosticadores existentes más usados los cuales tienen una granularidad de “Localidad” o “Zona”, es decir que logramos diferenciar la predicción de la altura de dos olas que pueden romper por ejemplo a una distancia de no más 100 metros entre ambas. Aunque la distancia sea mínima, estas dos olas pueden ser totalmente diferentes dependiendo de la geografía de la zona, de la arquitectura de la playa o del fondo del mar entre otros parámetros. Tal vez una escollera puede separar una ola que rompe en mar abierto de una ola que rompe a tan solo 50 metros de esta última, pero esta se encuentra rodeada entre escolleras, estas condiciones hacen que la predicción de esta ola sea distinta a la que rompe en mar abierto. Con este nivel de granularidad trabaja Surf-Forecaster, permitiendo definir múltiples olas a lo largo de la costa de una localidad específica. Los pronosticadores comunes, generalmente obtienen la información de la boya más cercana en altamar o del GridPoint más cercano si estos utilizan un sistema como WAVEWATCH III y dan el mismo pronóstico para toda la costa de una localidad, lo cual es irreal.
* Surf-Forecaster es un sistema web muy sencillo de entender y utilizar. Si es un usuario registrado puede crear las olas que desee y si dispone de un histórico de observaciones visuales para esas olas podría entrenar varios clasificadores para comenzar a recibir pronósticos exactamente en esos lugares, si no dispone de esta historia, entonces solo recibirá los pronósticos de los GridPoint elegidos como fuente de medición en altamar, en este último caso el sistema se comportara como cualquier otro pronosticador existente basado en WAVEWATCH III.
* Si el usuario interesado depende de encontrar un lugar a lo largo de toda la costa de una localidad que presente las mejores condiciones de oleaje y vientos para desarrollar una actividad determinada, Surf-Forecaster le ahorra recorrer largas distancias en busca de tales condiciones, permitiéndole ir directamente al lugar de su interés.
* En lugares donde la población se encuentra ubicada cerca de la costa el sistema puede predecir posibles catástrofes con olas de gran tamaño, con hasta una semana de anticipación. Esta es una posible ventaja de utilización de un sistema de este tipo, aunque para detectar este tipo de catástrofes, el clasificador debe ser entrenado con datos excepcionales, ya que el aprendizaje es en base a experiencia de casos pasados, y estos hechos son poco frecuentes.
* Este enfoque hace que no necesitemos recolectar datos costosos de conseguir como la batimetría marina. La batimetría es la ciencia que mide las profundidades marinas para determinar la topografía del fondo del mar, actualmente las mediciones son realizadas por GPS diferencial para una posición exacta, y con sonares hidrográficos mono o multihaz para determinar la profundidad exacta. Recolectar esta información es sumamente costoso, requiere de equipamiento especial y de expertos en la materia para analizar los resultados y poder cargarlos una vez procesados en un sistema que haga uso de ellos para poder pronosticar el oleaje. Gracias al uso de algoritmos de aprendizaje de máquina para entrenar un clasificador, los datos suministrados por las boyas en altamar, mas las observaciones visuales costeras son suficientes para poder pronosticar olas en aguas poco profundas.
* Surf-Forecaster puede predecir olas sin importar el tipo de fondo que la geografía de esa costa presenta. Es indistinto para el sistema si este es de arena, piedra, arcilla, arrecife de corales, etc. Aunque el tipo de fondo no afecta a las predicciones, si el fondo no es de un material solido como la piedra o un arrecife, este puede requerir una actualización de las observaciones visuales cada determinado tiempo (dependiente de cuán rápido la actividad marina pueda remover el fondo cambiando su topografía).
* Es posible pronosticar olas en cualquier lugar del mundo. Si se dispone de un histórico suficiente de observaciones visuales en cualquier lugar se podrán obtener pronósticos exactamente en esa ubicación, caso contrario el sistema nos mostrará los pronósticos capturados por WAVEWATCH III de un GridPoint cercano a ese punto elegido por el usuario al momento de dar de alta la ola. Cualquiera sea el caso, se recibirán pronósticos con una semana de anticipación en incrementos de tres horas.
* El sistema provee una sección de comparaciones, lo cual puede resultarle útil al usuario a la hora de decidir en qué lugar va a practicar su actividad. En dicha sección puede ver gráficos y tablas comparativas con las condiciones del mar actuales y durante las siguientes 180hs que el sistema alcanza a pronosticar. Así mismo puede resultar útil a usuarios con intenciones de realizar algún viaje y todavía no han elegido el destino, dado que esta elección depende de las condiciones que el mar presente en cada lugar.

**Limitaciones**

La siguiente lista muestra algunas de las limitaciones, o posibles implementaciones que podrían ser agregadas a Surf-Forecaster en el futuro a fin de tener un sistema pronosticador aun más robusto:

* Solo predice la altura de la ola en la costa y nos da información del periodo y dirección de la misma y a su vez velocidad y dirección del viento en esa zona, pero el sistema no indica si estas condiciones son óptimas, buenas o malas para una actividad. Supongamos que al usuario le interesa practicar surf, el sistema puede indicar que la altura de las olas en un determinado lugar va a ser de 3 metros (lo cual le puede resultar una altura ideal), pero esto a veces no es suficiente, dado que si, por ejemplo, el periodo de ola es pequeño (3 o 4 segundos) y los vientos son fuertes y tienen una dirección determinada, esto puede ser un indicador de que el mar estará revuelto, generando olas grandes de altura como predijo pero desparejas y sin forma, que no son buenas para la práctica del surf, pero tal vez lo sean para la práctica de windsurf o kite-surf. En definitiva el sistema puede predecir la altura de las olas en la costa con seguridad, pero aun requiere de un mínimo conocimiento del usuario en cuanto al análisis del resto de las variables pronosticadas para saber cual ola será la más apropiada para practicar su actividad favorita.
* Surf-Forecaster requiere de una historia de al menos 60 observaciones visuales en el punto exacto en el que se quiere obtener pronósticos. Solo requiere de la altura de la ola más grande de ese día junto con la fecha de la observación lo cual no es difícil de recolectar, pero si el usuario desea obtener pronósticos de todas las olas que rompen en la costa de su ciudad tendrá que dedicar al menos dos meses (aunque es recomendable más tiempo para mejorar los pronósticos) para recorrer la costa y recolectar la información necesaria, para luego entrenar cada clasificador en el sistema.
* Si el fondo del mar varia (por ejemplo en playas donde el fondo es de arena), hay que recolectar datos actualizados cada N cantidad de meses u años (según la frecuencia con que el fondo cambia lo suficiente como para que la ola que rompe en esa zona cambie su comportamiento) para mantener entrenado con información actualizada al clasificador que pronostica dicha ola.
* El sistema requiere de hardware con grandes capacidades de almacenamiento ya que está continuamente descargando información de las olas para todos los GridPoint definidos por WAVEWATCH III del planeta los cuales reportan lecturas actualizadas cada 6 horas. Toda esta información es persistida en la base de datos que utiliza luego Surf-Forecaster como historia para futuras predicciones.