Capítulo 4 - Evaluaciones

En el presente capítulo se detallan las pruebas realizad as para encontrar los mejores clasificadores que permitan predecir el tamaño que tendrá la ola al momento de romper en la costa, a partir de la información de alta mar brindada por el modelo de olas WAVEWATCH III. En especial se analizan los resultados obtenidos de experimentar con diferentes conjuntos de entrenamiento así como también con diferentes algoritmos de regresión.

**Consideraciones generales**

Todas las pruebas fueron realizadas utilizando el soporte de la herramienta **Weka**. La misma es una herramienta gráfica y librería escrita en Java para realizar minería de datos, proveyendo soporte para pre-procesamiento de los datos, algoritmos de clasificación, y visualización de los datos entre otras cosas.

Durante las pruebas se utilizaron diversos algoritmos con el objetivo de evaluar su desempeño en el problema en cuestión, los mismos son listados en la tabla 2.

|  |  |
| --- | --- |
| Algoritmo de Regresión | Clase |
| Regresión Lineal | weka.classifiers.functions.LinearRegression |
| Red Neuronal Multicapa | weka.classifiers.functions.MultilayerPerceptron |
| Soporte de Vectores | weka.classifiers.functions.SVMreg |
| Arboles de Regresión | weka.classifiers.trees.M5P |

Una vez entrenado un clasificador es necesario evaluar su desempeño. Para este fin se utilizan instancias de prueba cuya clase se conoce, el clasificador realiza la predicción de la clase y luego se compara el valor predicho y el valor real. Este proceso es repetido para un conjunto de instancias significante y luego diversos indicadores estadísticos son extraídos.

En este trabajo se utilizó la técnica de validación cruzada de 10 conjuntos (ver capítulo 2), y se utilizaron dos indicadores para medir el desempeño de los clasificadores: la *correlación* y el *error absoluto promedio.*

La *correlación* indica cual es el grado de relación entre el valor predicho y el valor verdadero. El valor de correlación se encuentra en el intervalo [-1,1], siendo los extremos indicadores de buen comportamiento del clasificador, no siendo así los valores cercanos a 0.

El *error absoluto promedio*, indica el promedio entre la diferencia del valor predicho y el valor verdadero en cada instancia de prueba. Mientras menor se este indicador mejor clasificador tendremos.

**Conjuntos de entrenamiento**

Como se mencionó en el capítulo 3, los atributos que forman las instancias de entrenamiento son de vital importancia para el rendimiento del clasificador obtenido.

Para determinar los atributos que generaban los mejores resultados, se experimentó generando conjuntos de entrenamiento a partir de los modelos de instancias descriptos en el capitulo 3. Estos conjuntos fueron utilizados como entrada a un algoritmo de aprendizaje de máquina y los clasificadores resultantes fueron comparados.

En la tabla a continuación se detalla el desempeño de los clasificadores obtenidos a partir de los diferentes modelos de instancias.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Playa | #Instancias | Modelo | Corr(Lin) | Mae | Corr(Nolin) | Mae |
| Nshore | 600 | 1 |  |  | 0.886 | 0.8529976 |
| Nshore | 600 | 2 |  |  | 0.88 | 0.881307052 |
| Nshore | 600 | 3(N=2) |  |  | 0.892 | 0.822519007 |
| Nshore | 600 | 3(N=3) |  |  | 0.887 | 0.876755772 |
| Nshore | 600 | 3(N=4) |  |  | 0.872 | 0.881098463 |
| Wshore | 600 | 1 |  |  |  |  |
| Wshore | 600 | 2 |  |  |  |  |
| Wshore | 600 | 3(N=2) |  |  |  |  |
| Wshore | 600 | 3(N=3) |  |  |  |  |
| Wshore | 600 | 3(N=4) |  |  |  |  |
| DH | 600 | 1 |  |  |  |  |
| DH | 600 | 2 |  |  |  |  |
| DH | 600 | 3(N=2) |  |  |  |  |
| DH | 600 | 3(N=3) |  |  |  |  |
| DH | 600 | 3(N=4) |  |  |  |  |
| Almo | 600 | 1 |  |  |  |  |
| Almo | 600 | 2 |  |  |  |  |
| Almo | 600 | 3(N=2) |  |  |  |  |
| Almo | 600 | 3(N=3) |  |  |  |  |
| Almo | 600 | 3(N=4) |  |  |  |  |
| WWard | 600 | 1 |  |  |  |  |
| WWard | 600 | 2 |  |  |  |  |
| WWard | 600 | 3(N=2) |  |  |  |  |
| WWard | 600 | 3(N=3) |  |  |  |  |
| WWard | 600 | 3(N=4) |  |  |  |  |

Se puede apreciar que en las diferentes playas el modelo de instancias que predomino en términos de correlacion y MAE ha sido el numero 3. Esto quiere decir que tener un conjunto de entrenamiento formado de la siguiente manera: genera los mejores clasificadores.

La siguiente tabla lista los conjuntos de entrenamiento generados para estas

Curva de aprendizaje

En el capítulo 3 se describieron los diferentes conjuntos de entrenamiento generados, los mismos varían en la composición de las instancias, la cantidad de instancias, y la selección de los atributos que los componen. Cada conjunto de entrenamiento fue utilizado como entrada a los diferentes algoritmos de regresión para obtener diferentes clasificadores.

Conjunto de entrenamiento

Para entrenar un clasificador es necesario especificar el problema en cuestión. Esta especificación se realiza en forma de instancias. Las instancias estarán compuesta por diferentes atributos. Luego un conjunto de instancias, denominado conjunto de entrenamiento, serán suministradas a un algoritmo de aprendizaje de maquina para obtener un clasificador del problema en cuestión. Luego el rendimiento del clasificador será evaluado.

El investigador tendrá a su cargo la definición de las instancias, el armado de los conjuntos de entrenamiento, la elección del algoritmo de aprendizaje, la parametrizacion del algoritmo y el análisis de la evaluacion del clasificador.

Conjuntos de entrenamiento

Para obtener un clasificador es necesario entrenarlo a partir de un grupo de instancias pasadas. La cantidad y los tipos de instancias utilizados afectan directamente la perfomance del clasificador, independientemente del algoritmo de aprendizaje utilizado.

A continuacion se muestran los resultados de experimentar con conjuntos de entrenamiento de diferente tamaño asi como también hemos generado conjuntos de entrenamiento a partir de instancias.

denominado conjunto de entrenamiento. Las instancias de este conjunto se caracterizan por tener todas el mismo numero y tipo de atributos, asi como también un atributo de clase. La elección de los atributos que componen una instancia dada afectan directamente el rendimiento del clasificador, y representa una fase muy importante en el proceso de obtención de un clasificador.

Otro factor importante es la cantidad de instancias en el conjunto de entrenamiento, ya que un numero de instancias inadecuado no permitirá obtener el rendimiento optimo del clasificador, asi como también un numero muy elevado de las mismas

**Algoritmos de regresión**

Weka es una herramienta gráfica y librería escrita en Java para realizar minería de datos, la misma provee soporte para realizar pre-procesamiento de los datos, algoritmos de clasificación, y visualización de los datos entre otras cosas. Las pruebas fueron realizadas utilizando esta herramienta. En la tabla 2 se muestran los diferentes algoritmos utilizados en las evaluaciones.

|  |  |
| --- | --- |
| Algoritmo de Regresión | Clase |
| Regresión Lineal | weka.classifiers.functions.LinearRegression |
| Red Neuronal Multicapa | weka.classifiers.functions.MultilayerPerceptron |
| Soporte de Vectores | weka.classifiers.functions.SVMreg |
| Arboles de Regresión | weka.classifiers.trees.M5P |

**Parametrización**

Los algoritmos de regresión utilizados son parametrizables. Siendo que no existe una base teórica completa para determinar que parámetros son apropiados para cada tipo de problema, la configuración optima suele realizarse a prueba y error.

Para obtener los parámetros óptimos de cada clasificador se fueron probando diferentes combinaciones de los mismos, aquellas configuraciones que lograban indicadores destacados, resultantes de la validación cruzada, fueron utilizadas para el entrenamiento de los clasificadores de las diferentes playas.

**Medidas de evaluación utilizadas**

Una vez entrenado un clasificador es necesario evaluar su desempeño. Para este fin se utilizan instancias de prueba cuya clase se conoce, el clasificador realiza la predicción de la clase y luego se compara el valor predicho y el valor real. Este proceso es repetido para un conjunto de instancias significante y luego diversos indicadores estadísticos son extraídos.

En este trabajo se utilizó la técnica de validación cruzada de 10 conjuntos (ver capítulo 2), para realizar la *optimización de los parámetros*(explicado a continuación) de los clasificadores. Mientras que para el entrenamiento y *validación* final de cada clasificador se utilizó un conjunto de entrenamiento independiente y se realizó 10 veces 10 fold-cross validation. En ambas casos se calcularon dos indicadores para medir el desempeño de los clasificadores: la *correlación* y el *error absoluto promedio.*

La *correlación* indica cual es el grado de relación entre el valor predicho y el valor verdadero. El valor de correlación se encuentra en el intervalo [-1,1], siendo los extremos indicadores de buen comportamiento del clasificador, no siendo así los valores cercanos a 0.

El *error absoluto promedio*, indica el promedio entre la diferencia del valor predicho y el valor verdadero en cada instancia de prueba. Mientras menor se este indicador mejor clasificador tendremos.

**Resultados**

A continuación se describe y se analiza el desempeño de los diferentes clasificadores obtenidos. Debido a la gran cantidad de experimentos los mismos fueron divididos por playa y son resumidos en tablas como la de la figura 4.1. En la columna 1 se hace referencia al número de conjunto descripto en el capítulo 3, las columnas 2 y 3 representan el desempeño del clasificador utilizando los parámetros del algoritmo por defecto, mientras que las columnas 3 y 4 representan el desempeño utilizando los parámetros óptimos encontrados.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Conjunto** | **Algoritmo** | **Correlación** | **Error Absoluto** | **Correlación**  **(Opt)** | **Error Absoluto(Opt)** |

Fig.4.1 – Tabla de resultados

Playa North Shore

Resultados.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Conjunto** | **Algoritmo** | **Correlación** | **Error Absoluto** | **Correlación**  **(Opt)** | **Error Absoluto(Opt)** |
| **1** | **Regresión Lineal** | **0.8098** | **1.0932** | **0.8098** | **1.0932** |
| **1** | **Red Neuronal** | **0.7857** | **1.2161** | **0.8466** | **0.9659** |
| **1** | **SVM** | **0.8101** | **1.0693** | **0.8499** | **0.9207** |
| **1** | **Arbol regresión** | **0.848** | **0.9496** | **0.848** | **0.9496** |
| **2** | **Regresión Lineal** | **0.7958** | **1.1255** | **0.7958** | **1.1255** |
| **2** | **Red Neuronal** | **0.772** | **1.2201** | **0.8331** | **1.0022** |
| **2** | **SVM** | **0.7955** | **1.096** | **0.8365** | **0.9496** |
| **2** | **Arbol regresión** | **0.8331** | **0.9924** | **0.8331** | **0.9924** |
| **3** | **Regresión Lineal** | **0.811747752** | **1.093111** | **0.811747752** | **1.093111** |
| **3** | **Red Neuronal** | **0.84822391** | **1.146125** | **0.847763577** | **0.964612** |
| **3** | **SVM** | **0.811342373** | **1.071063** | **0.850122646** | **0.919618** |
| **3** | **Arbol regresión** | **0.846815747** | **0.957505** | **0.846815747** | **0.957505** |
| **4** | **Regresión Lineal** | **0.8304** | **1.0294** | **0.8304** | **1.0294** |
| **4** | **Red Neuronal** | **0.8296** | **1.0395** | **0.863** | **0.8986** |
| **4** | **SVM** | **0.8304** | **1.0046** | **0.867** | **0.8519** |
| **4** | **Arbol regresión** | **0.8622** | **0.9031** | **0.8622** | **0.9031** |
| **5** | **Regresión Lineal** | **0.8349** | **1.0184** | **0.8349** | **1.0184** |
| **5** | **Red Neuronal** | **0.7726** | **1.2985** | **0.8651** | **0.8938** |
| **5** | **SVM** | **0.8347** | **0.9917** | **0.8702** | **0.8399** |
| **5** | **Arbol regresión** | **0.8619** | **0.9015** | **0.8619** | **0.9015** |
| **6** | **Regresión Lineal** | **0.835** | **1.014** | **0.835** | **1.014** |
| **6** | **Red Neuronal** | **0.8561** | **0.9814** | **0.866** | **0.8907** |
| **6** | **SVM** | **0.8348** | **0.9883** | **0.8696** | **0.8433** |
| **6** | **Arbol regresión** | **0.86255** | **0.9047** | **0.86255** | **0.9047** |
| **7** | **Regresión Lineal** | **0.863109** | **0.979726** | **0.863109** | **0.979726** |
| **7** | **Red Neuronal** | **0.876382** | **1.152842** | **0.888932** | **0.865886** |
| **7** | **SVM** | **0.862481** | **0.962793** | **0.89143** | **0.82562** |
| **7** | **Arbol regresión** | **0.88279** | **0.882462** | **0.88279** | **0.882462** |
| **8** | **Regresión Lineal** | **0.8013** | **1.2876** | **0.8013** | **1.2876** |
| **8** | **Red Neuronal** | **0.8262** | **1.3232** | **0.8309** | **1.1976** |
| **8** | **SVM** | **0.8014** | **1.2551** | **0.8306** | **1.1679** |
| **8** | **Arbol regresión** | **0.8006** | **1.2963** | **0.8006** | **1.2963** |
| **9** | **Regresión Lineal** | **0.6714** | **0.7505** | **0.6714** | **0.7505** |
| **9** | **Red Neuronal** | **0.7357** | **0.8020** | **0.7425** | **0.6750** |
| **9** | **SVM** | **0.6674** | **0.7154** | **0.7811** | **0.6023** |
| **9** | **Arbol regresión** | **0.7606** | **0.6396** | **0.7606** | **0.6396** |

Análisis……

Pruebas:

Se disponían 7 anios de datos. Se utilizaron 3 años para la optimización de los parámetros. 4 anios para la validación final.

Tanto en la optimización de los parámetros como en la validación final se utilizó la técnica de cross validations en 10 carpetas. A continuacion se detallan en tablas los resultados de la validación de los diferentes clasificadores obtenidos. Como se menciono anteriormente los diferentes clasificadores varian en el algoritmo de regresion utilizado y en la composición de las instancias utilizadas.

Problema como comparo resultados de dos conjuntos diferentes y que tienen diferente cantidad de instancias?

Result set 1: correlación, error absoluto.

Result set 2: correlación, error absoluto.

Diferentes modelos de instancias….

* Obtener una curva de entrenamiento a partir de la cantidad de instancias utilizadas.

1, 2, 3,…8 años.

* Aplicar la optimización de los parámetros.

Utilizando n años.

* Obtener las validaciones finales.

Utilizando n años.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Un conjunto de entrenamiento está compuesto de instancias. Estas instancias se caracterizan por tener el mismo tipo y cantidad de atributos. Los atributos elegidos para un determinado conjunto de entrenamiento determinan la información que recibirá el algoritmo de aprendizaje de maquina. La selección de estos atributos es de vital importancia para cualquier problema de aprendizaje de maquina, siendo que estos afectan directamente la performance del clasificador obtenido, independientemente del algoritmo de aprendizaje utilizado.

Para determinar los atributos que generaban los mejores resultados, se experimentó generando conjuntos de entrenamiento a partir de los modelos de instancias descriptos en el capitulo 3. Estos conjuntos de entrenamiento fueron utilizados como entrada a un algoritmo de aprendizaje de máquina y los clasificadores resultantes fueron comparados.

Tba

Como se eligio la mejor Estrategia

Porque se eligio el mejor algoritmo

Tablas con breve explocación cada uno. Parametros optimos y no optimos. En caso de no dar dibujar

Playas = Rendimiento con la estategia y algoritmo seleccionado.