**Una comparación experimental de las medidas de rendimiento para la clasificación**

En nuestro caso nos enfocamos en resaltar lo que consideramos nos genera mas valor y aporta en nuestro proceso de formación.

Teniendo en mente que el objetivo más importante de esta metodología es que nos permite evaluar la calidad de los métodos de aprendizaje y los modelos aprendidos, es fundamental no olvidar que debemos considerar que afuera hay una búsqueda constante en otros métodos que nos permitan hacer mejores elecciones mas enfocadas en ares de espacialidad específicas, por ejemplo, TI o manufactura entre muchas otras.

Después de los análisis realizados con clústeres, relaciones, análisis de sensibilidad, umbral de clase, calidad de clasificación/separabilidad, rendimiento en C/S etc. , y de las correspondientes definiciones y experimentos , además de las relaciones entre métrica , una taxonomía y ordenación evidenciamos que en la mayoría de los casos lo que nos permite elegir la medida más adecuadas para la aplicación especifica de nuestro caso de estudio o análisis, apoyar estrategias de innovación en el momento de una toma de decisión .

Queremos resaltar la importancia de hacer una evaluación correcta de los modelos, identificar las diferencias que existen entre evaluar un modelo de regresión con error absoluto o con error cuadrático. y definitivamente lo más importantes es que contamos con un número importante de medidas para evaluar los clasificadores.

Para este caso se definió en tomar métricas para evaluar calificadores con precisión, Medida F, Tasa de rango, AUC, Brier entre otras y que se usaran 18 métricas diferentes clasificadas en 3 familias ,Se buscaron métricas con las que buscaron: Un modelo que minimice el número de errores y que sean las más apropiadas para conjuntos de datos balanceados o desbalanceados, para detección de señales o fallas. Otras que fueran más útiles cuando queremos evaluar la fiabilidad de los clasificadores y que esto es fundamental para los modelos realicen correctamente una fusión ponderadora de los modelos. Entre muchas otras que nos permiten que lo clasificadores se utilicen para seleccionar los mejores Instancias de un conjunto de datos, los sistemas de recomendación, la detección de fraudes, el filtrado de spam, Al final de todo este análisis lo más importante es interpretar los resultados que en este caso nos muestran que la mayoría de estas métricas realmente miden cosas diferentes y en muchas situaciones la elección realizada con una métrica puede ser diferente de la elección realizada con otra , los resultados obtenidos en una medida podrían extrapolarse a otras medidas.

Otro aspecto a considerar que no debemos olvidar es que el análisis se completa con un conjunto de experimentos para cuantificar la sensibilidad a cuatro rasgos importantes que están presentes en algunas medidas, pero no están presentes en otras. Estos rasgos son la elección óptima del umbral de clase, la calidad de la clasificación/separabilidad, el rendimiento de la calibración y la sensibilidad (o, por el contrario, la solidez) a los cambios en la distribución de la clase anterior. A partir de este análisis, podemos cuantificar las relaciones sobre estas 'dimensiones', lo que es un complemento muy útil para los resultados del análisis de correlación.

Lo que se encontró en las conclusiones más interesantes de este primer trabajo experimental es que comparo las métricas de evaluación de los clasificadores mas utilizadas, obteniendo conclusiones que involucran interdependencia y sensibilidad de las medidas, aunque sabemos que los estudios continúan constantemente, lo más importante es mantenernos actualizados con lo último que vamos teniendo como recurso.

Finalmente, frente a las conclusiones del estudio de relaciones, es que estructuraron un proceso progresivo y evolutivo para analizar las correlaciones que los llevo a validar la existencia de similitudes entre las medidas sin embargo las diferencias también son determinantes en el mismo. Lo que nos permite identificar que no hay nada concluyente, simplemente debemos buscar aplicar la que mas convenga de acuerdo a la data y caso de uso que vayamos a analizar.

C. Ferri, J. Hernández-Orallo, R. Modroiu, An experimental comparison of performance measures for classification, Pattern Recognition Letters, Volume 30, Issue 1, 2009, Pages 27-38.