

RESUMEN DE PROPUESTA DE TOG

APLICACIÓN DE LOS TESTORES EN TAREAS DE CLASIFICACIÓN

Proponente: Nombre, Empresa o institución de Adscripción, correo-e

Proponente: Guillermo Sánchez Díaz.
Universidad Autónoma de San Luis, gsanchezd@iteso.mx

1) Descripción general del proyecto propuesto

En el campo del aprendizaje máquina, los testores irreducibles son una técnica de selección de variables que nos permiten reducir la dimensionalidad de grandes conjuntos de datos, en particular, para tareas de clasificación. Es particularmente útil cuando los datos son mezclados, es decir, la información se compone de datos numéricos (entero, real) y categóricos (lógico, caracter, enumeración, conjunto de cadenas de texto, etc).

En este trabajo se propone la aplicación de los testores irreducibles en la resolución de un subconjunto específico de problemas de clasificación. Se implementarán/adaptarán dos sistemas clasificadores que utilicen el conjunto completo de variables/atributos en su entrenamiento y prueba de al menos dos conjuntos de datos. Se obtendrán los testores irreducibles de cada conjunto de datos y, a partir de ellos, se calculará la relevancia de cada atributo. La relevancia puede manifestarse como un valor numérico en el rango [0.0...1.0] o un valor lógico [Considerar, Ignorar]. Se ejecutarán los dos sistemas clasificadores tomando como insumo la relevancia de cada atributo. Finalmente, se contrastarán los resultados obtenidos sin utilizar y utilizando testores irreducibles, en cuanto a eficiencia computacional y precisión en la clasificación.

Para un mejor entendimiento de la selección de variables en tareas de clasificación, considérese el siguiente ejemplo. La tabla muestra los síntomas observados en cinco pacientes. Como es aprendizaje supervisado, contamos también con la información de la clase: el diagnóstico.

Paciente	Dolor de garganta	Tos	Resfriado	Fiebre	Diagnóstico
Hugo	SI	SI	NO	NO	Gripa
Paco	SI	NO	SI	NO	Gripa
Luis	NO	NO	SI	SI	Faringitis
Donald	SI	NO	SI	SI	Faringitis
Mickey	NO	NO	SI	NO	Faringitis

Un sistema clasificador deberá predecir qué enfermedad tiene un paciente nuevo proporcionando sólo sus síntomas. Es muy claro que el nombre del paciente no es relevante para el diagnóstico, sin embargo, forma parte de la tabla de datos y podría meter ruido en el proceso de predicción. Nos podemos dar cuenta que los síntomas Dolor de garganta y Fiebre son suficientes para no confundir entre un paciente con Gripa y otro con Faringitis: “si tiene dolor de garganta y no presenta fiebre, definitivamente es gripa”. Por lo tanto, esos atributos tienen mayor relevancia que Tos y Resfriado.

RESUMEN DE PROPUESTA DE TOG

2) Vinculación o Colaboración

Dr. Guillermo Sánchez Díaz, Universidad Autónoma de San Luis Potosí, coasesor.

3) Asignaturas de la MSC relacionadas con el desarrollo del proyecto

Fundamentales: Matemáticas Avanzadas para Computación, Análisis y Diseño de Algoritmos.

Electivas: Aprendizaje Automático, Estadística Aplicada, Aprendizaje Profundo.

4) Estudiante participante en la propuesta

La propuesta está dirigida para 1 estudiante.

Bibliografía relacionada

Ivan Piza-Davila, Guillermo Sanchez-Diaz, Manuel S. Lazo-Cortes, Cristina Noyola-Medrano. "Enhancing the Performance of YYC Algorithm Useful to Generate Irreducible Testors". International Journal of Pattern Recognition and Artificial Intelligence, vol. 32, no. 1, 2018.

Ivan Piza-Davila, Guillermo Sanchez-Diaz, Manuel S. Lazo-Cortes, Luis Rizo-Dominguez. "A CUDA-based hill-climbing algorithm to find irreducible testors from a training matrix". Pattern Recognition Letters, vol. 95, pp. 22-28, 2017.

Ivan Piza-Davila, Guillermo Sanchez-Diaz, Carlos A. Aguirre-Salado, Manuel S. Lazo-Cortes. "A parallel hill-climbing algorithm to generate a subset of irreducible testors". Applied Intelligence, vol. 42, no. 4, 2015, pp. 622-641, 2015.

Guillermo Sanchez-Diaz, Manuel Lazo-Cortes, Ivan Piza-Davila. "A Fast Implementation for the Testor Property Identification Based in an Accumulative Binarycation Based in an Accumulative Binary Tuple". International Journal of Computational Intelligence Systems, vol. 5, no 6, 2012.

CV breve del proponente

El Dr. Guillermo Sánchez Díaz es profesor-investigador adscrito a la Facultad de Ingeniería de la Universidad Autónoma de San Luis Potosí. Sus líneas de investigación son reconocimiento de patrones, minería de datos y aplicaciones de la inteligencia artificial en la geoinformática. Es autor y coautor de diferentes artículos publicados en revistas internacionales indexadas, capítulos de libros y memorias de congresos. Ha sido árbitro revisor de distintas revistas y congresos nacionales e internacionales. Es miembro del Sistema Nacional de Investigadores del CONACYT, nivel I.