

Medidas de complejidad de grafos para clasificación en múltiples dominios

El objetivo de este trabajo fue ofrecer nuevas ideas y métodos para clasificar redes complejas basadas en sus propiedades topológicas. Estudiamos las propiedades de redes de diferentes dominios (biológicos, sociales, infraestructura, etc.), para encontrar características comunes compartidas por las redes de cada uno de ellos. El paso siguiente fue proponer técnicas que aprovechen estos elementos comunes, para automatizar la clasificación de las mismas en sus respectivos dominios. Esta información también es de utilidad para determinar si una red artificial presenta las características esperadas para el dominio al que se pretende que pertenezca.

Analizamos alrededor de 100 fórmulas de complejidad encontradas en la literatura. De ellas, elegimos 13 que luego derivan en un total de 43 métricas. Seleccionamos un conjunto de 240 redes de distintas fuentes distribuidas en 11 dominios. Hicimos un análisis exhaustivo de las medidas obtenidas para ellas, analizamos su correlación y distribución estadística, e intentamos predecir las posibilidades de clasificación que tendría un algoritmo automático a la hora de identificar correctamente sus respectivos dominios. Para encarar la tarea de clasificación, recurrimos a 3 tipos de algoritmos de clasificación de la literatura (K Nearest Neighbors, Support Vector Machines y el sistema de Evolutionary Rule Learning Bio-HEL) usando las medidas obtenidas como valores de entrada.

Para entender el potencial que tienen estos métodos usamos el generador de grafos CiGRAM para crear un conjunto de 160 redes artificiales distribuidas en 8 dominios. Realizando pruebas preliminares con estos grafos artificiales logramos una precisión en la clasificación de hasta 80%. Observamos además que es fácil interpretar los resultados que ofrece el sistema BioHEL, ya que su salida está compuesta de reglas legibles por una persona, que explican como clasificar redes en base a los valores de las métricas calculadas.

Finalmente aplicamos las técnicas de clasificación a las redes reales elegidas. Obtuvimos una clasificación de menor precisión que la obtenida para las redes artificiales, por lo cual realizamos experimentos adicionales para estudiar posibles mejoras, como eliminación de outliers y particionamiento manual de los conjunto de entrenamiento y prueba. En promedio los resultados siguen siendo peores a los obtenidos para las redes artificiales, pero un subconjunto de los mismos presenta una precisión de 80% como en el caso anterior.

Nuestras conclusiones permiten confirmar que es posible clasificar redes de distintos dominios considerando únicamente las mediciones obtenidas de un conjunto específico de métricas de complejidad.

Palabras clave: grafos, redes, complejidad, clasificación, métricas, automatización, algoritmos, múltiples dominios