UNIVERSIDAD DE GRANADA

FACULTAD DE CIENCIAS

Propuesta TFG. Curso 2026/2027

GRADO: Grado en Matemáticas

CÓDIGO DEL TFG: 270-065-2026/2027

1. DATOS BÁSICOS DEL TFG:

Título: Modelado causal mediante algoritmos genéticos

Descripción general (resumen y metodología):

La inferencia causal se ocupa de entender las relaciones de causa y efecto en distintos escenarios y estudios experimentales. A diferencia del análisis estadístico tradicional, que se centra en identificar patrones de asociación entre variables, la inferencia causal busca responder a preguntas contrafactuales: ¿cómo cambiaría un resultado si alteráramos deliberadamente un factor del sistema? Esta distinción es fundamental, ya que dos variables pueden estar correlacionadas sin que exista una relación causal entre ellas, como puede ocurrir cuando intervienen factores no observados.

La capacidad de inferir relaciones causales sólidas es esencial en contextos donde se requiere tomar decisiones informadas o intervenir en sistemas complejos. Por ejemplo, en medicina, para evaluar la efectividad real de un tratamiento; en políticas públicas, para anticipar el impacto de una reforma; o en economía y negocios, para optimizar estrategias ante variaciones del entorno. Sin un análisis causal riguroso, se corre el riesgo de tomar decisiones basadas en correlaciones espurias, lo que puede conducir a acciones ineficaces o incluso perjudiciales.

Una pieza fundamental para las técnicas de inferencia causal son los denominados modelos causales que constituyen una representación formal de la relación e interacción entre las distintas variables a considerar en el proceso de inferencia. Estos modelos causales pueden venir determinados por asesoramiento experto, aunque en otras ocasiones no se cuenta con tal conocimiento. Por ello, otros algoritmos para construir el mejor modelo causal a partir de los datos observacionales disponibles han sido propuestos en la literatura.

A priori, sin otro conocimiento, y pensando en que un modelo causal pueda definirse mediante un grafo dirigido acíclico, donde los nodos son las variables consideradas en el análisis estadístico, y las aristas las relaciones causales entre estas, el espacio de posibles modelos causales —de posibles grafos— a considerar puede ser muy grande, lo que se conoce como una explosión combinatoria. Esto hace que determinar el grafo óptimo para un determinado conjunto de datos sea un problema NP duro.

Así, en este trabajo se explora la aplicación de algoritmos genéticos a la exploración del espacio de posibles grafos causales para, a través de una función de adecuación de cada posible modelo causal a los datos observados, encontrar bien el óptimo o un modelo suficientemente bueno. Los algoritmos genéticos son una estrategia evolutiva de gran relevancia en problemas de optimización en los que el vasto espacio de búsqueda imposibilita la exploración exhaustiva. Se estudiarán desde el punto de vista teórico-matemático, y también desde el punto de vista práctico, para su posterior aplicación al problema que ocupa a este trabajo. Del mismo modo se estudiarán los fundamentos de la inferencia causal; en particular los relacionados con el modelado causal.

Finalmente, se llevará a cabo una experimentación con datos reales que pongan a prueba los conocimientos adquiridos en la realización de este trabajo.

Tipología: Estudio de casos, teóricos o prácticos, relacionados con la temática del Grado.

Objetivos planteados:

1. Estudiar y comprender los fundamentos teóricos de la inferencia causal.
2. Estudiar y comprender los fundamentos teóricos y matemáticos de los algoritmos genéticos.

3. Estudiar y explorar la propuesta de aplicación de una estrategia evolutiva en la búsqueda del modelo causal más adecuado para un conjunto de datos observado.
4. Aplicación de esta metodología a conjuntos de datos reales empleando el software R y/o el lenguaje de programación Python.

Bibliografía básica:

- Wang, H., et al. (2024). Causal Structure Learning Based on Genetic Algorithm and Markov Blanket. 2024 IEEE 14th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER), 720-725.
- Goldberg, D. E. (1989). Genetic algorithms in search, optimization, and machine learning (Reprint. with corr). Addison Wesley.
- Holland, J. H. (1992). Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). Causal inference in statistics: A primer. John Wiley & Sons.
- Rosenbaum, P. R. (2023). Causal inference. MIT Press.
- Smith, M. J., Mansournia, M. A., Maringe, C., Zivich, P. N., Cole, S. R., Leyrat, C., Belot, A., Rachet, B., & Luque-Fernandez, M. A. (2022). Introduction to computational causal inference using reproducible Stata, R, and Python code: A tutorial. Statistics in Medicine, 41(2), 407-432.

Recomendaciones y orientaciones para el estudiante:

Haber cursado las siguientes asignaturas obligatorias del Grado en Matemáticas: Estadística Descriptiva e Introducción a la Probabilidad, Probabilidad e Inferencia Estadística.

Plazas: 1

2. DATOS DEL TUTOR/A:

Nombre y apellidos: GUSTAVO RIVAS GERVILLA

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: griger@ugr.es

3. COTUTOR/A DE LA UGR (en su caso):

Nombre y apellidos: MIGUEL ÁNGEL LUQUE FERNÁNDEZ

Ámbito de conocimiento/Departamento: ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Correo electrónico: mluquefe@ugr.es

4. COTUTOR/A EXTERNO/A (en su caso):

Nombre y apellidos:

Correo electrónico:

Nombre de la empresa o institución:

Dirección postal:

Puesto del tutor en la empresa o institución:

Centro de convenio Externo:

5. DATOS DEL ESTUDIANTE:

Nombre y apellidos: ALICIA GORDO MORAL

Correo electrónico: aliciagmm@correo.ugr.es