

Estudio de problemas de programación matemática mediante técnicas estocásticas y neuronales

Investigadores responsables: Lic. José Luis Espinoza B., Lic. Walter Mora F., M.Sc. Geovanni Figueroa M., M.Sc. Alcides Astorga M., Dr. Javier Trejos Zelaya, Dra. Ileana Castillo A.
Departamento de Matemática

Resumen

En este reporte técnico se hace una descripción general del avance que hemos tenido en el proyecto. En éste hemos estudiado algunas teorías fundamentales de programación matemática usando técnicas estocásticas y neuronales y se han implementado algunas de las metodologías, principalmente las relacionadas con los conjuntos aproximados. Actualmente estamos implementando una versión más eficiente para poder trabajar con tablas de datos más grandes y con mejoras en las interfaces con el usuario.

Abstract

This technical report makes a general description of the advance we have been having in our research. We have studied some important theories about mathematical programming by using stochastic and neural technics, and we have implemented some methodologies, mainly those related with the rough sets. At the present time we are programming a more efficient version in order to work with larger tables of data, with a better user's interface.

Problema

Estudiar la solución de algunos problemas de programación matemática a través de técnicas

estocásticas de optimización y de algunas otras técnicas propias de la inteligencia artificial. Específicamente, el proyecto se ha enfocado a la clasificación y discriminación de datos mediante conjuntos aproximados, con miras a aplicarlos a la caracterización y simplificación de datos reales. Se pretende además dejar sentadas las bases para la obtención de reglas de decisión y, a partir de éstas, generar sistemas expertos. Ligado a esto, se investiga el comportamiento de algunas redes neuronales y la obtención de árboles de decisión, vistos como mecanismos de clasificación de datos.

Objetivos del proyecto

- Estudio e implementación de técnicas previsionales como los conjuntos aproximados y evaluar los resultados sobre datos reales (médicos) y simulados.
- Caracterización de tipos de pacientes de cáncer de mama y cáncer de pulmón por un lado, y de úlcera sometidos a una operación de vagotomía altamente selectiva, por otro; esta caracterización se pretende obtener mediante la elección de un número reducido de variables discriminantes obtenidas de la aplicación de la metodología de conjuntos aproximados.
- Plantear el problema de programación lineal mediante redes neuronales y comparar la solución con la obtenida mediante el algoritmo del simplex y el de Karmakar.

- Plantear problemas de programación no lineal, programación entera y programación dinámica mediante sobrecalentamiento simulado, algoritmos genéticos y búsqueda tabú, y comparar los resultados con las técnicas clásicas en el campo.
- Plantear la optimización de una función en una red de *Hopfield* usando un algoritmo de tipo genético para adaptar los pesos de la red y compararlo con uno de tipo sobrecalentamiento simulado (Máquina de Boltzmann).
- Escribir un sistema informático experimental sobre redes neuronales y optimización combinatoria, con flexibilidad para plantear las redes que el usuario necesita y que éste lo pueda usar de manera experimental en sus investigaciones.

Metodología

El proyecto tiene una parte teórica, que consiste en el estudio de metodologías novedosas, y una parte aplicada, que consiste en usar los métodos propuestos o estudiados.

En general, como en la mayoría de las investigaciones matemáticas, la metodología consiste en:

- Investigación bibliográfica.
- Reuniones periódicas de tipo seminario, con exposiciones al resto del grupo por parte de cada investigador.
- Planteamiento de métodos y algoritmos para resolver los problemas estudiados.
- Estudio de las propiedades teóricas de los métodos y algoritmos que se plantean.
- Implementación computacional de las técnicas estudiadas.

Los datos concretos que se han venido estudiando provienen de investigadores médicos del Hospital R.A. Calderón Guardia, así como otros datos citados con frecuencia en la literatura y que sirven de base para hacer las comparaciones con los métodos existentes.

Financiamiento del proyecto

El proyecto ha sido financiado por la Vice-rectoría de Investigación y Extensión del Instituto Tecnológico de Costa Rica.

Resultados alcanzados hasta el momento

- Se estudió con mucho detalle la teoría de conjuntos aproximados. Esta es útil para caracterizar una partición conocida de un conjunto de objetos a los que se les ha medido un cierto número de atributos, determinando cuáles de éstos son relevantes en los datos y cuáles se pueden omitir. Lo que se obtiene es una o varias reducciones del conjunto de atributos.
- La metodología dada por los conjuntos aproximados parte de tablas de datos cualitativos así que hemos trabajado con el Algoritmo de Fisher. Este lo que produce es una partición óptima de un conjunto de datos cuantitativos y, por lo tanto, sirve para recodificar los datos pertenecientes a atributos cuantitativos para convertirlos a cualitativos.
- Se hizo una implementación en lenguaje Pascal de los conjuntos aproximados. Haciendo uso del programa, se realizaron experimentos con grupos de datos simulados y con tablas clásicas de datos tomados de la literatura, tales como datos zoológicos, datos médicos y datos de pintores, entre otros. Los resultados de estos experimentos fueron muy satisfactorios. La teoría, el código-fuente del programa y los experimentos pueden consultarse en el primer informe presentado a la Vice-rectoría de Investigación y Extensión.
- Actualmente se está haciendo una implementación mejorada mediante un lenguaje orientado a objetos, con el que se pretende, por un lado, mejorar la interfaz con el usuario y por otro, mejorar la eficiencia y robustez del programa y agregarle algunos módulos para la recodificación de tablas de datos cuantitativos a

datos cualitativos, para el manejo de tablas de decisión y para hacer reducciones aproximadas basándose en un umbral de calidad de las representaciones. En este momento se están realizando pruebas del programa para determinar su funcionalidad. Al final de este informe aparece un apéndice con unas pantallas que ilustran el ambiente de trabajo que presenta el prototipo del programa al usuario.

- Hemos estudiado también las tablas de decisión como un paso intermedio entre la reducción de atributos y la creación de un sistema experto basado en reglas. Esto último significa una rama sobre la que vale la pena profundizar en un futuro: la generación automática de un sistema experto a partir de la tabla original de datos.
- Otro de los campos en que se ha incursionado es en las redes neuronales, estudiando e implementando los principales algoritmos. Pueden consultarse los detalles de este trabajo en el tercer informe del proyecto.
- Se está estudiando la forma de construir un árbol de decisiones usando la teoría de conjuntos aproximados, para luego generar una red neuronal y usarla para la clasificación de datos.

Impacto del proyecto

Ya se manifiestan algunas situaciones y necesidades en las que puede tener impacto si se le da el seguimiento y continuidad requeridos.

En primer lugar, podemos contribuir mucho a la simplificación de mediciones y datos tomados para el estudio de ciertas enfermedades. Aunque actualmente tenemos algunas tablas de datos reales provenientes de estudios de enfermedades a nivel nacional, no se les ha hecho a todas el cálculo de reducciones debido a limitaciones de la implementación que se había hecho en la primera fase del proyecto. Tal implementación, pese a que quedó muy bien programada y calculó apro-

piadamente las reducciones en los datos con que fue probada, tiene la restricción de trabajar sólo con tablas de datos relativamente pequeñas y algunas de estas tablas tienen muchos datos incompletos, por lo que requerirían compartir trabajo con expertos, ya sea para eliminar información innecesaria o para compensar la ausencia de datos con algún criterio de reposición.

En segundo lugar, tenemos conocimiento de grupos de investigadores en nuestro país, quienes tienen recopilada información que necesitan recodificar para luego hacerles algún tratamiento estadístico.

En general, el impacto que puede tener el proyecto se resume en la contribución a otras ciencias, en la simplificación y sintetización de la información recolectada sobre diversos estudios realizados, y en el mejoramiento de la recolección de datos futuros, de manera que sólo la información pertinente e indispensable sea tomada y procesada.

Referencias bibliográficas

- Diday, E.; Lemaire, J.; Pouget, J.; Testu, F. (1982). *Eléments d'Analyse de Données*. Ed. Dunod, París.
- Espinoza, J. L.; Figueroa, G.; Mora, W.; Trejos, J. (1996) *Primer informe del proyecto "Estudio de problemas de programación matemática mediante técnicas estocásticas y neuronales"*. ITCR, Costa Rica.
- Espinoza, J. L.; Figueroa, G.; Mora, W.; Trejos, J. (1996) *Segundo informe del proyecto "Estudio de problemas de programación matemática mediante técnicas estocásticas y neuronales"*, ITCR, Costa Rica.
- Castillo, I.; Espinoza, J. L.; Figueroa, G.; Mora, (1997) *Tercer informe del proyecto "Estudio de problemas de programación matemática mediante técnicas estocásticas y neuronales"*. ITCR, Costa Rica.
- Pawlak, Z. (1985) *Rough Sets*. Kluwer Publ. Co., Dordrecht.

- Pawlak, Z. (1986) "Rough classification of patients after highly selective vagotomy for duodenal ulcer" *Int. J. Man-Machine Studies*, 24, 413-433.
- Slowinsky, R.; Stefanowski, J. (1989) "Rough classification in incomplete information systems", *Mathl. Comput. Modelling*, Vol. 12 (10-11),
- Trejos, J. (1994) *Contribution à l'Acquisition Automatique de Connaissances à Partir de Données Qualitatives*. Thèse de Doctorat, Université Paul Sabatier, Toulouse.