

Técnicas de Optimización. Aplicaciones en Minería de Datos, y Robótica Evolutiva

Laura Lanzarini¹, Waldo Hasperue², Leonardo Corbalán³, Javier López⁴, César Estrebou⁵, Franco Ronchetti⁶, Juan Maulini⁷, Augusto Villa Monte⁸

Instituto de Investigación en Informática LIDI (III-LIDI)⁹ -Facultad de Informática. UNLP

{laural, whasperue, corbalan}@lidi.info.unlp.edu.ar, lopezjavierhernan@gmail.com, {cesarest, fronchetti, jmaulini, avillamonte}@lidi.info.unlp.edu.ar

Contexto

Esta presentación corresponde al Subproyecto “Sistemas Inteligentes” perteneciente al Proyecto “Procesamiento paralelo y distribuido. Fundamentos y aplicaciones en Sistemas Inteligentes y Tratamiento de imágenes y video” acreditado por la UNLP.

Resumen

Esta línea de investigación se centra en el estudio y desarrollo de Técnicas de Optimización basadas en diferentes mecanismos de adaptación.

La métrica para determinar la eficiencia y eficacia de los métodos propuestos es la optimización de funciones complejas. Interesa especialmente la reducción del tiempo de convergencia con el objetivo de transferir los resultados obtenidos.

En el área de Minería de Datos, se ha trabajado fuertemente en la adaptación de distintas metaheurísticas poblacionales con el objetivo de representar y obtener por adaptación, reglas de clasificación con antecedentes formados por atributos numéricos y nominales. Se ha trabajado con técnicas de optimización mono-objetivo utilizando poblaciones de tamaño variable.

En el área de la robótica evolutiva, el énfasis está puesto en investigar diferentes estrategias evolutivas aplicables al control de robots autónomos en tiempo real dentro de un ambiente experimental permitiendo especificar distintos recorridos y objetivos así como la función a optimizar. Se busca obtener controladores neuronales adaptados por evolución. Esto implica un detallado análisis de los operadores genéticos involucrados y el estudio e investigación de técnicas de procesamiento paralelo que permitan obtener soluciones en períodos de tiempo acotados.

Palabras claves: Redes Neuronales, Algoritmos Evolutivos, Minería de Datos, Optimización mediante Cúmulos de Partículas, Robótica Evolutiva.

1. Introducción

Las técnicas de optimización han demostrado ser herramientas sumamente útiles en la resolución de problemas complejos. Su capacidad de adaptación al entorno de información les permite brindar buenos resultados en distintas áreas.

En el Instituto de Investigación en Informática LIDI se está trabajando en este tema desde hace varios años. Inicialmente se desarrollaron estrategias basadas en Redes Neuronales y Algorit-

¹ Profesor Titular. Facultad de Informática. UNLP

² Becario de Postgrado Tipo II CONICET – Ayudante Diplomado Facultad de Informática. UNLP

³ Profesor Adjunto. Facultad de Informática. UNLP

⁴ Alumno del Doctorado en Cs. Informática. Facultad de Informática. UNLP

⁵ Jefe de Trabajos Prácticos SD. Facultad de Informática. UNLP

⁶ Becario UNLP Tipo A. Ayudante Alumno. Facultad de Informática. UNLP

⁷ Ayudante Diplomado SD. Facultad de Informática. UNLP

⁸ Becario CIN. Ayudante Alumno. Facultad de Informática. UNLP

⁹ Calle 50 y 120 2do Piso, (1900) La Plata, Argentina, TE/Fax +(54) (221) 4227707. <http://weblidi.info.unlp.edu.ar>

mos Evolutivos aplicables al Reconocimiento de Patrones así como al control de agentes autónomos. Se definieron y desarrollaron varias estrategias basadas en redes neuronales competitivas con aplicaciones concretas en distintas áreas.

A continuación se detallan en forma breve los avances realizados últimamente.

1.1. Técnicas de Optimización basadas en metaheurísticas poblacionales

En el III-LIDI se han investigado en profundidad dos metaheurísticas poblacionales: GA (Genetic Algorithms) y PSO (Particle Swarm Optimization) [1] [2] [3] [4] [5].

Los GA son ampliamente conocidos por su capacidad de evitar caer en óptimos locales. Sin embargo, este comportamiento se encuentra condicionado a una selección adecuada de los individuos a reproducir y a la definición correcta de los operadores genéticos. En el III-LIDI actualmente se está trabajando en un nuevo método evolutivo GAAP (Genetic Algorithm with Auxiliary Populations) que combina múltiples poblaciones auxiliares con la población principal de un GA clásico. GAAP realiza evoluciones cortas e independientes de pequeñas poblaciones que aportan material genético a la población principal. La función de las poblaciones auxiliares es doble: impide o dificulta la convergencia prematura a subóptimos locales y al mismo tiempo provee un mecanismo de búsqueda local para intensificar la explotación de las regiones más prometedoras del espacio de búsqueda.

La experimentación llevada a cabo muestra que GAAP, aplicado a problemas de optimización continua, mejora ampliamente el rendimiento de un GA clásico tanto en funciones unimodales como multimodales, sean éstas separables o no separables. Los resultados sugieren además que GAAP presenta un buen equilibrio entre exploración global y explotación de las mejores regiones del espacio de búsqueda.

Por otro lado, en [6] se ha desarrollado una nueva técnica de optimización basada en cúmulos de partículas inspirada en PSO. El funcionamiento de esta técnica, denominada lpPSO, consiste en determinar, para cada partícula, un

camino a seguir tal que ésta recorra el espacio de búsqueda de un extremo a otro. De esta manera cuando una partícula alcanza un límite del espacio se le asigna un nuevo camino para recorrer. Esta técnica presenta tres grandes ventajas: el cúmulo nunca queda estancado en zonas con mínimos locales, el espacio de búsqueda se recorre de una manera expansiva y, una zona con una solución óptima prometedoras es recorrida por las partículas en diferentes direcciones. lpPSO fue comparada con PSO demostrando encontrar mejores soluciones utilizando para ello una menor cantidad de evaluaciones de partículas, lo que produce una solución más eficiente, en términos de recursos computacionales, para resolver problemas de optimización.

1.2 Minería de Datos

Otra línea de investigación que se lleva a cabo en el III-LIDI se ha centrado en la Minería de Datos (Data Mining) obteniendo buenos resultados al aplicar variantes de PSO (Particle Swarm Optimization) para la obtención de reglas de clasificación.

En [3] se ha definido una nueva versión de PSO Binario combinando una representación discreta (o binaria) con otra continua que al ser utilizada para resolver problemas de optimización de funciones ha demostrado ser capaz de brindar resultados superadores.

Esta misma técnica está siendo adaptada para poder obtener una lista de reglas de clasificación a partir de información estructurada formada por atributos nominales y numéricos. Para lograrlo se ha extendido la representación de cada partícula, combinando información real y binaria en cada individuo.

Cada individuo de la población sólo lleva codificado el antecedente de la regla y su valor de aptitud depende de la clase cuyos elementos se busquen cubrir.

Se trata de un proceso iterativo que a medida que cubre los ejemplos, los retira de la entrada para dar lugar a la generación de nuevas reglas.

En lo que se refiere a la representación, cada atributo nominal es representado por una cadena de dígitos binarios cuya longitud queda determinada por la cantidad de valores distintos que puede tomar y se calcula según [3]. Como mecanismo de simplificación del antecedente de la

regla se utiliza una selección por ruleta del valor más probable. Esto último hace uso de la representación real de la partícula. Para los atributos numéricos se utiliza PSO convencional que determina el intervalo de valores que puede tomar. Cada atributo lleva asociado un límite inferior y un límite superior. La parte binaria de la partícula indica si dicho límite debe ser considerado o no. Si para ambos, su valor es cero, el atributo numérico no formará parte del antecedente de la regla.

Por otro lado, se está analizando la posibilidad de extender el área de aplicación de las técnicas desarrolladas, llevándolas a la Minería de Textos. Sin embargo, esta etapa implica una cuidadosa definición del espacio de entrada ya que en general se trata de patrones con una dimensionalidad alta lo que dificulta la evaluación de la correspondiente función de aptitud. Es importante tener en cuenta que la dimensionalidad del problema guarda una estrecha relación con la cantidad de ejemplos disponible y cuando se trata de texto, si se utiliza el concepto de diccionario convencional, estas relaciones se encuentran invertidas dando lugar a una relación 1 a 10 entre la cantidad de ejemplos disponibles y la longitud de cada ejemplo.

1.3. Robótica Evolutiva

En el III-LIDI se ha trabajado en este tema desde hace varios años [7] [8] [9] [10] [11]. En general se trata de procesos que requieren un tiempo computacional elevado. Actualmente, se ha investigado el problema del estacionamiento automático de vehículos por ser de gran interés tanto en la comunidad científica como en la industria. Si bien en la actualidad existen diversos estudios e implementaciones de este problema, las soluciones no son del todo satisfactorias quedando aun muchos aspectos por resolver. Para ejecutar el proceso evolutivo se adaptó un entorno en Player/Stage, un simulador de acceso libre muy utilizado en la comunidad científica.

En robótica evolutiva, uno de los mayores desafíos es la detección de los diferentes espacios temporales. Pocos trabajos existentes plantean el concepto de *memoria* y no existe una definición que provea un mecanismo para redefinir tareas en función del tiempo. En los traba-

jos desarrollados se definió una arquitectura neuronal que posee como memoria temporal las salidas de la red en cierta cantidad de pasos previos. Estos pasos previos, junto con un cuentakilómetros son utilizados como nuevas entradas de la red neuronal.

Uno de los aspectos en los que se prestó mayor atención fue en la correcta definición de la función de aptitud para lograr reducir los óptimos locales. Los resultados obtenidos lograron resolver el problema de estacionar un vehículo autónomo en forma paralela en un espacio reducido [12].

2. Líneas de investigación y desarrollo

- Desarrollo e implementación, a partir de los métodos existentes, de estrategias adaptativas capaces de construir y mantener modelos adecuados en entornos de información dinámicos.
- Desarrollo de nuevas variantes de algoritmos tipo PSO y evaluación de rendimiento.
- Estudios de performance de los algoritmos desarrollados. Análisis de eficiencia en la resolución de problemas concretos.
- Estudio de distintos métodos de caracterización de textos haciendo énfasis en su longitud.
- Análisis comparativo de GAAP con otras metaheurísticas tanto poblacionales como basadas en trayectoria. Estudio minucioso sobre los parámetros del método para fijar con precisión cuáles son los rangos de valores más convenientes y en qué casos deben ser aplicados.
- Integración de la metaheurística PSO con GAAP. Desarrollo de una solución que combine múltiples procesos PSO en auxilio del GA principal.
- Estudio de diferentes arquitecturas neuronales aplicables a la resolución del problema del estacionamiento automático de vehículos, enfocado en la correcta definición de las funciones de aptitud.

3. Resultados obtenidos

- Desarrollo e implementación de estrategias basadas en cúmulos de partículas (PSO)

aplicables a la optimización de funciones Se trabaja sobre los siguientes conceptos

- Población variable.
- Movimiento de partículas discretas (representación binaria) utilizando vectores de velocidad continuos.
- Desarrollo e implementación de una nueva técnica de optimización utilizada para ajustar modelos de simulación biológicos.
- Desarrollo de un controlador neuronal que permita resolver el problema del estacionamiento automático en un entorno adaptado en el simulador Player/Stage.
- Estudio de distintas métricas para medir la separación de clases en el espacio de entrada.

4. Formación de Recursos Humanos

Dentro de los temas involucrados en esta línea de investigación se están desarrollando actualmente 2 tesis de doctorado, 1 de maestría y al menos 4 tesinas de grado de Licenciatura. También participan en el desarrollo de las tareas becarios y pasantes del III-LIDI.

Referencias

- [1] "Mixed Alternative in Assignment by Classes vs. conventional methods for Calculation of individual lifetime in GAVaPS ". Lanzarini, Sanz, Naiouf, Romero. ITI 2000. 22nd International Conference on Information Technology Interfaces. IEEE Cat. No.00EX411, Croatia, Junio 13-16,2000. Pags 383-390
- [2] "Particle Swarm Optimization with Variable Population Size". Lanzarini L., Leza V., De Giusti A. Lecture Notes in Computer Science. Vol 5097/2008. Artificial Intelligence and Soft Computing. Pags. 438-449. Junio de 2008. ISBN 987-3-540-69572-1. Springer.
- [3] "A new Binary PSO with velocity control". Lanzarini L., López J., Maulini J., De Giusti A. Advances in Swarm Intelligence. Lecture Notes in Computer Science. 2011, Volume 6728/2011, 111-119. Springer 2011.
- [4] "Particle Swarm Optimization with Oscillation Control". López J., Lanzarini L., De Giusti A. ACM Special Interest Group on Genetic and Evolutionary Computation. Pages: 1751-1752. 2009. ISBN:978-1-60558-325-9
- [5] "VarMOPSO: Multi-Objective Particle Swarm Optimization with Variable Population Size". López J., Lanzarini L., De Giusti A. Lecture Notes in Computer Science. Vol 6433/2010, pp.60-69. ISBN 978-3-642-16952-6. Springer. Noviembre 2010.
- [6] "lpPSO - New optimization strategy inspired by PSO". Hasperué W., Corbalan L. XII Workshop de Agentes y Sistemas Inteligentes, XVII Congreso Argentino de Ciencias de la Computación. CACIC 2011. Pages 131-140. Octubre 2011.
- [7] "Modular Creation of Neuronal Networks for Autonomous Robot Control". Osella Massa G., Vinuesa H., Lanzarini L. 3er IEEE Latin American Robotics Symposium. Chile. Octubre 2006. ISBN: 1-4244-0537-8. Pág. 66-73.
- [8] "Improving Controllers based on Neural Networks obtained by Layered Evolution Strategy". Vinuesa, Lanzarini, Osella Massa. Conferencia Latinoamericana de Informática. CLEI 2008.
- [9] "Continuous Evolution of Neural Modules of Autonomous Robot Controllers". Vinuesa, Osella Massa, Corbalán y Lanzarini. Jornadas Chilenas de Computación 2007. Iquique, Chile. Noviembre de 2007.
- [10] "Controlador Robótico obtenido a través de una metaheurística de población variable". Ronchetti F, Lanzarini L. Workshop de Agentes y Sistemas Inteligentes (WASI). XVI Congreso Argentino de Ciencias de la Computación CACIC 2010.
- [11] "Robotic controller obtained through a speciation based metaheuristic". Ronchetti F., Lanzarini L. 2011 INNS International Educational Symposium on Neural Networks. Lima, Perú. 2011.
- [12] "Automatic Vehicle Parking Using an Evolution-Obtained Neural Controller". Ronchetti F., Lanzarini L. Workshop de Agentes y Sistemas Inteligentes (WASI). XVII Congreso Argentino de Ciencias de la Computación CACIC 2011.