

Sistemas Inteligentes. Aplicaciones en Minería de Datos, Procesamiento de Señales y Redes de Computadoras

Laura Lanzarini¹, Waldo Hasperue², Leonardo Corbalán³, César Estrebou⁴, Juan La Battaglia⁵

Instituto de Investigación en Informática LIDI (III-LIDI)⁶
Facultad de Informática. UNLP

CONTEXTO

Esta presentación corresponde al Subproyecto “Sistemas Inteligentes” perteneciente al Proyecto “Algoritmos Distribuidos y Paralelos. Aplicación a Sistemas Inteligentes y Tratamiento Masivo de Datos” del Instituto de Investigación en Informática LIDI.

RESUMEN

Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes basados en mecanismos de adaptación. Actualmente el énfasis está puesto en la transferencia de tecnología a las áreas de minería de datos, procesamiento de señales y redes de computadoras.

En el área de la Minería de Datos, los temas centrales se encuentran relacionados con la investigación de nuevas estrategias adaptativas que generen modelos a partir de grandes volúmenes de datos y además sean capaces de acomodar su estructura ante algún cambio en los datos y/o nueva información que se obtenga, reflejando estos cambios en el conocimiento actual adquirido.

En el área del procesamiento de señales, el énfasis está puesto en la combinación de las técnicas habitualmente utilizadas con estrategias adaptativas inteligentes que permitan desarrollar aplicaciones con capacidades para reconocer e identificar objetos o personas en tiempo real. Se trabaja en el desarrollo de métodos adaptativos factibles de ser aplicables en distintos contextos.

La aplicación de distintas metaheurísticas sobre redes Peer-to-Peer (P2P), es otra línea de investigación que se está llevando a cabo en el III-LIDI. El objetivo central es mejorar la capacidad de adecuación del sistema a los cambios rápidos del entorno de información.

Palabras claves : Redes Neuronales, Algoritmos Evolutivos, Minería de Datos, Técnicas de Optimización, Procesamiento de Señales.

1. INTRODUCCIÓN

Los Sistemas Inteligentes 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 Algoritmos Evolutivos aplicables al Reconocimiento de Patrones [Lan00, Lan04] así como al control de agentes autónomos [Cor03a, Cor03b, Oli05]. Luego se profundizó en el estudio de la capacidad de caracterización de este tipo de estrategias a fin de establecer un modelo de la

¹ Profesor Titular. Facultad de Informática. UNLP

² Becario de Perfeccionamiento de la CIC – Ayudante Diplomado - Facultad de Informática. UNLP

³ Becario de Formación Superior de la UNLP. Jefe de Trabajos Prácticos - Facultad de Informática. UNLP

⁴ Jefe de Trabajos Prácticos DS - Facultad de Informática. UNLP

⁵ Becario de Iniciación de la UNLP - Facultad de Informática. UNLP

⁶ Calle 50 y 115 1er Piso, (1900) La Plata, Argentina, TE/Fax +(54) (221) 422-7707. <http://weblidi.info.unlp.edu.ar>

información disponible. Se definieron y desarrollaron varias estrategias basadas en redes neuronales competitivas con aplicaciones concretas en distintas áreas [Has07a, Cor06a, Cor06b]. Actualmente, con el objetivo de mejorar la transferencia tecnológica de los resultados obtenidos, el énfasis está puesto por un lado, en la definición de estrategias que faciliten la interpretación del modelo y por otro en la incorporación de mecanismos de adaptación a entornos dinámicos.

A continuación se detallan brevemente los avances realizados últimamente.

1.1. Minería de Datos

Las investigaciones realizadas en el III-LIDI relacionadas con diferentes mecanismos de aprendizaje y adaptación de redes neuronales competitivas aplicables a la Minería de Datos han facilitado la construcción de modelos de la información disponible principalmente a partir de reglas de asociación y clasificación [Has05, Has06, Has07]. Sin embargo, existen situaciones en las cuales el modelo por si solo es incapaz de transmitir al usuario el conocimiento adquirido así como su importancia y relación con los datos subyacentes. En esta dirección, las técnicas visuales resultan una manera intuitiva de presentación, dándole una perspectiva exploratoria y contextual.

El gran obstáculo que actualmente presenta la transferencia tecnológica en temas referidos a estrategias de minería de datos, es el desconocimiento por parte del usuario acerca de la forma de adaptar el modelo a sus necesidades. Como forma de ayudar a revertir esta situación, se han desarrollado estrategias que definen, a partir del modelo basado en reglas de clasificación, las acciones a seguir para lograr el beneficio esperado [Has07b].

Otro aspecto importante a considerar es el dinamismo de la información. Hoy en día, la tecnología posibilita el almacenamiento de enormes volúmenes de datos. Gran parte de las estrategias inteligentes existentes permiten modelizarlos para un instante de tiempo dado. Sin embargo, existe una amplia gama de problemas que requiere disponer de mecanismos capaces de adaptar el modelo existente a los cambios del entorno. En [Has08] se ha generado un método de obtención de reglas difusas las cuales facilitan su adaptación ante nueva información.

Siguiendo con esta línea de trabajo se espera conseguir nuevos métodos y técnicas de minería de datos que sean capaces de generar conocimiento útil, produciendo resultados que sean de provecho al usuario final.

1.2. Procesamiento de Señales

En el III-LIDI interesa especialmente la combinación de técnicas de procesamiento de señales con estrategias adaptativas inteligentes que permitan desarrollar aplicaciones con capacidades para reconocer e identificar objetos o personas en tiempo real. Este tipo de aplicaciones resulta de sumo interés en área tales como seguridad y telefonía celular, entre otras.

En esta línea, se han realizado algunas investigaciones sobre señales de voz a fin de verificar la identidad de la persona que habla. Este es un problema difícil de resolver debido a la variabilidad de la información de entrada; no sólo porque dos personas distintas pueden tener un registro parecido sino porque una misma persona puede generar dos registros diferentes en distintos instantes de tiempo. Esto motiva la necesidad de utilizar herramientas adaptativas capaces de procesar información dinámica.

En el III-LIDI se han utilizado redes neuronales competitivas dinámicas para modelizar las voces de un conjunto de personas a reconocer. El entrenamiento se realizó a partir de los coeficientes cepstrales de las señales de voz logrando un buen reconocimiento al aplicarse tanto a voces sintéticas como reales.

También se está trabajando en el desarrollo de un protocolo de comunicación para teléfonos celulares con capacidad de transmisión de voz y audio en un formato propietario reducido y eficiente.

1.3. Redes de Computadoras

La aplicación de distintas metaheurísticas sobre redes Peer-to-Peer (P2P), es otra línea de investigación que se está llevando a cabo en el III-LIDI. Las redes P2P constituyen un sistema distribuido, descentralizado y sumamente dinámico, donde la información disponible cambia con mucha rapidez, al igual que la topología misma de la red. El objetivo central de esta investigación es mejorar la capacidad de adecuación del sistema a los cambios rápidos del entorno de información utilizando metaheurísticas poblacionales.

Si bien existen soluciones previas [Shi07] basadas en cúmulos de partículas para determinar la interconexión de los nodos en redes P2P con algunas restricciones, investigaciones llevadas a cabo en el III-LIDI han permitido definir y desarrollar una extensión original de PSO que, a diferencia de [Ken95][Shi99], permite trabajar con una población de tamaño variable. De esta forma, no es necesario definir a priori la cantidad de soluciones a utilizar, evitando así condicionar la calidad de la solución a obtener [Lan08].

Para la búsqueda de recursos en sistemas P2P se ha propuesto una estrategia que utiliza redes neuronales locales para asistir a los nodos en la propagación selectiva de solicitudes dirigidas sólo a los vecinos más adecuados. El algoritmo de aprendizaje tradicional de las redes neuronales que se utilizan ha sido reemplazado por un apropiado intercambio de información entre vecinos que permite acelerar los tiempos de entrenamiento y mantener actualizado el conocimiento adquirido por cada nodo [Cor09].

Actualmente se está trabajando en la reducción del costo computacional de la solución propuesta. Cuando la información disponible se encuentra distribuida, los modelos globales resultan ineficientes. Por tal motivo se propone dividir el proceso en dos partes, una primera parte que utiliza un algoritmo local muy eficiente para modelizar la información cercana y luego una segunda etapa que utiliza la información generada localmente para monitorear el sistema completo. Con esto se espera lograr una reducción considerable en el tiempo de adaptación y una mayor flexibilidad a los cambios de la información en cada nodo.

2. TEMAS DE INVESTIGACIÓN Y DESARROLLO

- Estudio y análisis de diferentes estructuras de modelización dinámicas. Interesa especialmente la obtención de árboles de decisión incluyendo los árboles difusos e incrementales.
- Desarrollo y aplicación de diferentes métricas que permitan analizar el conjunto de reglas a utilizar para representar el modelo. Esto incluye considerar distintas estrategias de poda que permitan maximizar la representación obtenida.
- Desarrollo e implementación, a partir de los métodos existentes, de estrategias adaptativas capaces de construir y mantener modelos adecuados en entornos de información dinámicos.
- Análisis de los distintos tipos de Redes Neuronales competitivas dinámicas. Estudio de las estrategias existentes que permiten determinar, durante la adaptación, el tamaño de la arquitectura y forma de conexión de los elementos que componen la red neuronal.

- Estudio y aplicación de diferentes de métricas que permitan analizar la preservación de la topología de los datos tanto en el espacio de los patrones de entrada como en el espacio de salida de la red.
- Estudio de distintas metaheurísticas de optimización basadas en trayectoria y en población aplicables al problema de ruteo en redes P2P.
- Análisis de la importancia de los parámetros de las distintas metaheurísticas en el funcionamiento y eficiencia de la estrategia seleccionada. Análisis de la función de aptitud a utilizar en el caso de redes P2P.
- Estudios de performance de los algoritmos desarrollados. Análisis de eficiencia en la resolución de problemas concretos.

3. RESULTADOS OBTENIDOS/ ESPERADOS.

- Desarrollo e implementación de una estrategia para la obtención de reglas de clasificación a partir de una matriz de co-asociación.
- Desarrollo e implementación de mecanismos que permitan la especificación de un conjunto de acciones a seguir a fin de objetivar la interpretación de la información modelizada.
- Desarrollo e implementación de una red neuronal entrenada a partir del método AVGSOM como herramienta para reconocer patrones de señales de voz.
- Desarrollo e implementación de un sistema de votación a partir de varias redes neuronales competitivas dinámicas que permite reducir el error en la etapa de reconocimiento.
- Desarrollo e implementación de una estrategia basada en cúmulos de partículas (PSO) con tamaño de población variable basado en los conceptos de edad y vecindario. Se ha comprobado que el mecanismo utilizado para incorporar nuevos individuos así como la forma de calcular el tiempo de vida preserva la diversidad de la población.
- Resolución de problemas concretos, tanto en ambientes simulados como en el mundo real. En este último caso, resulta de fundamental importancia la optimización del algoritmo propuesto.

4. FORMACIÓN DE RECURSOS HUMANOS

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

5. REFERENCIAS

- [Cor03a] Corbalán, Lanzarini. GNE. Grupos Neuronales Evolutivos. *XXIX Conferencia Latinoamericana de Informática*. CLEI 2003. Bolivia. 2003
- [Cor03b] Corbalán, Lanzarini. Evolving Neural Arrays A new mechanism for learning complex action sequences. *Special Issue of Best Papers presented at CLEI'2002*. Volumen 6. Number 1, December 2003. Montevideo, Uruguay.
- [Cor05] Corbalán, Lanzarini, De Giusti A. ALENA. Adaptive-Length Evolving Neural Arrays". *Journal of Computer Science and Technology*. Vol 5, nro. 4. 2005. Pags. 59-65.

- [Cor06a] Corbalán, Osella Massa, Russo, Lanzarini. Image recovery using a new nonlinear Adaptive Filter based on Neural Networks. *CIT - Journal of Computing and Information Technology*. Vol. 14, No.4. December 2006 - Pág. 315-320.
- [Cor06b] Corbalán, Hasperue, Osella Massa, Lanzarini. BPNn-CPN. Nuevo método para segmentación de Imágenes basado en Redes Neuronales Artificiales. *IV Workshop de Computación Gráfica, Imágenes y Visualización (WCGIV)*. CACIC 2006. San Luís. Argentina. Octubre de 2006.
- [Cor09] Corbalán, Lanzarini, De Giusti. Resources NeuroSearch in Peer-to-Peer Networks. *31th International Conference of Information Technology Interfaces (ITI)*. Cavtat, Croacia. June 2009
- [Has05] Hasperué, Lanzarini. Dynamic Self-Organizing Maps. A new strategy to enhance topology preservation. *XXXI Conferencia Latinoamericana de Informática*. CLEI 2005.
- [Has06] Hasperué, Lanzarini. Classification Rules obtained from Dynamic Self-organizing Maps. *VII Workshop de Agentes y Sistemas Inteligentes*. CACIC 2006. San Luis. Argentina. Octubre de 2006.
- [Has07a] Hasperue, Corbalán, Lanzarini, Bría. Skeletonization of Sparse Shapes using Dynamic Competitive Neural Networks. *Ibero-American Journal of Artificial Intelligence*. Vol.11. Nro.35. pp.33-42 2007.
- [Has07b] Hasperué, Lanzarini. Extracting Actions from Classification Rules. *Workshop de Inteligencia Artificial. Jornadas Chilenas de Computación 2007*. Iquique, Chile. Noviembre de 2007.
- [Has08] Hasperué, W., G. Osella Massa & L. Lanzarini. Obtaining a Fuzzy Classification Rule System from a non-supervised Clustering. *30th International Conference of Information Technology Interfaces (ITI)*. Cavtat, Croacia. June 2008.
- [Ken95] Kenedy, Eberhart. Particle Swarm Optimization. *IEEE International Conference on Neural Networks*. Vol IV, pp.1942-1948. Australia 1995.
- [Lan00] Lanzarini. Reconocimiento de Patrones en Imágenes Médicas utilizando Redes Neuronales. *Journal of Computer Science and Technology*. Vol.4 . Dic 2000.
- [Lan04] Lanzarini, Yanivello. Reconocimiento de Comandos Gestuales utilizando GesRN. *V Workshop de Agentes y Sistemas Inteligentes*. CACIC 2004. Bs.As. Argentina. 2004. ISBN 987-9495-58-6
- [Lan08] Lanzarini, Leza, De Giusti. Particle Swarm Optimization with Variable Population Size. *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.
- [Oli05] Olivera y Lanzarini. Cyclic Evolution. A new strategy for improving controllers obtained by layered evolution. *Journal of Computer Science and Technology*. Vol 4, nro. 1. 2005. Pags 211-217.
- [Shi07] Shichang, Ajito, Guiyong, Hongbo. A Particle Swarm Optimization Algorithm for Neighbor Selection in Peer-to-Peer Networks. *6th International Conference on Computer Information. Systems and Industrial Management Applications (CISIM'07)*. Pp. 166-172. ISBN 0-7695-2894-5. June 2007.
- [Shi99] Shi Y., Eberhart R. An empirical study of particle swarm optimization. *IEEE Congress Evolutionary Computation*. pp.1945-1949. Washington DC, 1999.