

# Predicción de fenómenos naturales mediante paralelismo, algoritmos evolutivos y búsqueda por novedad

Caymes-Scutari Paola<sup>1,2</sup>, Bianchini Germán<sup>1</sup>, Strappa Jan<sup>1,2</sup>

<sup>1</sup>Laboratorio de Investigación en Cómputo Paralelo/Distribuido,  
Departamento de Ingeniería en Sistemas de Información. Facultad Regional  
Mendoza/Universidad Tecnológica Nacional  
Rodríguez 273 (M5502AJE) Mendoza, +54 261 5244579

<sup>2</sup>Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)

pcaymesscutari@frm.utn.edu.ar, gbianchini@frm.utn.edu.ar, jstrappa@frm.utn.edu.ar

## RESUMEN

La predicción de diversos fenómenos, como las avalanchas, los incendios, las inundaciones, etc., constituye un proceso complejo en el cual interviene una gran cantidad de información a considerar y numerosas acciones, operaciones, y procedimientos a realizar sobre dicha información. Ello permite modelar el fenómeno, predecir su comportamiento a un cierto plazo, y tomar las decisiones que se consideren más adecuadas a fin de prevenir, mitigar, o paliar los efectos negativos que pueden provocarse en el ecosistema, la población, la sociedad, y la economía a raíz del fenómeno. Nuestra línea de investigación propone abordar la predicción de fenómenos de propagación a través de la integración de tres elementos: la potencia que ofrece el cómputo paralelo/distribuido, la capacidad de los algoritmos poblacionales para orientar la búsqueda de soluciones, y la oportunidad que ofrecen los algoritmos basados en búsqueda por novedad (NS o novelty search). Dichos elementos serán considerados para una familia de métodos predictivos denominada DDM-MOS (*Data Driven Methods with Multiple Overlapping Solutions*), la cual se trata de

una familia de métodos de ayuda a la decisión. Las características intrínsecas de la búsqueda basada en novedad privilegian la evolución u obtención de nuevos resultados (predicciones) en función de la novedad y no solamente por aptitud, como sucede en los algoritmos evolutivos puros. Dicha característica permitiría prevenir o disminuir la aparición de situaciones de convergencia prematura o estancamiento, en beneficio de la calidad de los resultados obtenidos al realizar la predicción.

**Palabras clave:** Metaheurísticas, Predicción, Novelty, Incertidumbre.

## CONTEXTO

Los DDM-MOS consideran un modelo de simulación que toma como entrada un conjunto de parámetros, los cuales describen características que afectan a la propagación, y en base a la combinación de diferentes técnicas computacionales arrojan una predicción sobre el comportamiento del fenómeno. Esta línea se ha desarrollado a lo largo de los años en el LICPaD (Laboratorio de Investigación en Cómputo Paralelo/Distribuido), en la UTN-FRM,

aplicándose al caso particular de la predicción de incendios forestales. Los métodos desarrollados por el grupo tienen como objetivo la reducción de incertidumbre en las simulaciones de esta naturaleza, dado que las limitaciones para proveer parámetros de entrada correctos para el modelo conducen a errores en la precisión de la predicción. En general, los métodos desarrollados involucran un componente o metaheurística evolutiva para reducir el espacio de búsqueda y a la vez orientar la búsqueda. En este contexto, la incorporación de búsqueda por novedad o *novelty search*, tiene como objetivo modificar el funcionamiento clásico de las metaheurísticas, otorgando importancia a la riqueza que puedan brindar al proceso aquellos puntos que resulten más novedosos, y así privilegiar no sólo a los sectores aparentemente más aptos, sino también poniendo en valor a los que resulten más diferentes y ofrezcan posibilidades renovadas de exploración. Actualmente, la línea de investigación cuenta con dos líneas de financiamiento homologadas, a través de un proyecto PID UTN que dará inicio en abril de 2023 (SITCME0009816TC), y un proyecto PICTO de FONCyT (UUMM-2019-00042).

## 1. INTRODUCCIÓN

Existen diversos simuladores cuyo objetivo es predecir la evolución de la línea de fuego a lo largo de un período de tiempo [1]. Para esto, los simuladores requieren un conjunto de parámetros de entrada, llamados también escenarios, que describen características que afectan a la propagación, tales como el tipo de material combustible, su humedad, pendiente y relieve del terreno, temperatura, y la velocidad y dirección del viento. Desde el punto de vista computacional, este problema de

predicción es desafiante debido a la complejidad de los modelos involucrados y a las fuentes de incertidumbre existentes en los datos de entrada. Este último aspecto es de gran importancia dado que las limitaciones para proveer parámetros de entrada correctos para el modelo conducen a errores en la predicción. Esta incertidumbre se debe a la dificultad o imposibilidad de obtener los valores de las variables, tanto por limitaciones de recursos (por ejemplo, cantidad de sensores), porque las mediciones son indirectas (por ejemplo, la medición de la humedad de la vegetación), o porque las variables tienen un comportamiento dinámico y no es viable su observación en tiempo real (por ejemplo, características del viento). Esto hace que la predicción basada en una solución única no sea confiable [1].

Para enfrentar esta incertidumbre, se han propuesto métodos que combinan resultados de múltiples simulaciones para mitigar los efectos negativos de la misma. Este tipo de métodos son los denominados DDM-MOS [1], dos de los cuales han sido desarrollados en los últimos años: ESSIM-EA [5,6] y ESSIM-DE [10]. Su funcionamiento involucra la combinación de una Etapa de Optimización, para obtener escenarios cuya simulación permita obtener buenas predicciones, con una Etapa de Análisis Estadístico, que construye una matriz de probabilidades a partir de la agregación de mapas obtenidos de las múltiples simulaciones paralelas, para así poder predecir el comportamiento de la línea de fuego. En el caso de ESSIM-EA, el método de optimización empleado es un algoritmo genético, mientras que en ESSIM-DE se utiliza un algoritmo de evolución diferencial [8]. Debido a la carga de cómputo de este proceso, ambos sistemas utilizan cómputo paralelo; concretamente, se basan en un modelo de

islas [9], utilizando una doble jerarquía de procesos que les permite abarcar mejor el espacio de búsqueda y acelerar la optimización. Es importante señalar que, además de emplearse para la prevención y predicción de incendios forestales, estos métodos pueden aplicarse a otros modelos de propagación, tales como inundaciones, avalanchas o corrimientos de terreno.

Dado que los mencionados métodos se basan en metaheurísticas, se pueden ver afectados por algunas limitaciones intrínsecas de los mismos, como son el estancamiento o la convergencia prematura. Para el caso particular de ESSIM-DE, se propusieron diferentes aproximaciones que permitieron sintonizar su comportamiento y lograr mejoras tanto en calidad de resultados como en tiempo de ejecución.

Más recientemente, se consideró la posibilidad de utilizar un paradigma alternativo al de las metaheurísticas tradicionales: la *búsqueda basada en novedad* o *Novelty Search* (NS) [2-4]. Este es un paradigma de búsqueda que ignora el objetivo como guía para la exploración y, en su lugar, recompensa a soluciones candidatas que presentan comportamientos novedosos (diferentes a los anteriormente descubiertos), con el fin de maximizar la exploración y evitar óptimos locales. La búsqueda basada en novedad es una alternativa prometedora frente a las limitaciones de las metaheurísticas aplicadas previamente al problema de la predicción, debido a que, en este paradigma alternativo, la búsqueda es dirigida por una caracterización del comportamiento de los individuos, cambiando así el *paisaje de la función de fitness* (*fitness landscape*), previniendo de manera directa los problemas de convergencia prematura y estancamiento.

## 2. LÍNEAS DE INVESTIGACIÓN y DESARROLLO

Los sistemas de predicción existentes han obtenido resultados satisfactorios para la reducción de incertidumbre en el problema de aplicación. En particular, ESSIM-EA demostró obtener predicciones de buena calidad, mientras que ESSIM-DE redujo significativamente los tiempos de respuesta, pero sin obtener mejoras en calidad. Por esta razón, posteriormente se ha trabajado en mejorar el rendimiento del método ESSIM-DE mediante estrategias denominadas de sintonización [7]. Se desarrollaron dos de estas métricas [11, 12], ambas orientadas a mitigar los problemas de convergencia prematura y estancamiento de la población presentes en el caso de aplicación del algoritmo, obteniendo mejoras en calidad y tiempos de respuesta.

Dadas las características del problema de la predicción y los resultados previos obtenidos mediante los DDM-MOS ya desarrollados, recientemente se ha llevado a cabo una primera aproximación para incorporar NS al proceso de reducción de incertidumbre y predicción guiado por Evolución Diferencial. Los resultados obtenidos son prometedores y señalan un camino a seguir para continuar explorando en las posibilidades que puede proveer, en este caso, a los DDM-MOS. Asimismo, los resultados han permitido confirmar que el enfoque NS es adecuado para este tipo de problema, dado que, por un lado, resuelve por diseño los problemas propios de algoritmos explotativos; y por otro lado, porque presenta múltiples oportunidades de paralelización y de hibridación con enfoques basados en *fitness*. También es posible adaptar el comportamiento del

algoritmo de acuerdo a las características del problema.

Otra de las líneas de investigación del grupo se centra en la sintonización de aplicaciones paralelas, motivo por el que también es factible plantear a un plazo más largo, el análisis y modelado del comportamiento de aquellos DDM-MOS que incorporen búsqueda por novedad, y así contribuir a lograr predicciones más precisas y/o en menor tiempo.

### **3. OBJETIVOS TEÓRICOS Y EXPERIMENTALES**

El principal objetivo del proyecto consiste en considerar nuevas estrategias para mejorar la calidad de predicción alcanzada por los DDM-MOS, como métodos computacionales para la reducción de incertidumbre en la predicción de fenómenos de propagación. Más allá de la utilidad y beneficios que los algoritmos evolutivos poblacionales aportan a los DDM-MOS para orientar la búsqueda, son algoritmos que intrínsecamente traen aparejados ciertos problemas de rendimiento (como convergencia prematura o estancamiento) que afectan a las características de la población a lo largo del tiempo. Es por ello que se propone potenciar las capacidades de los métodos predictivos basados en algoritmos evolutivos poblacionales mediante la integración de características basadas en novedad que privilegien la evolución según este aspecto y no sólo por aptitud. La incorporación de la búsqueda por novedad permitirá, además, conocer mejor las potencialidades de las metaheurísticas para aplicarlas en la

predicción de diferentes fenómenos de propagación, como se mencionó anteriormente. Como parte de este objetivo general pueden señalarse algunos objetivos más específicos. Por un lado, identificar los beneficios de utilizar búsqueda por novedad, y las posibilidades de adaptarla e integrarla a algoritmos predictivos y de reducción de incertidumbre basados en metaheurísticas y paralelismo. Y, por otro lado, analizar los resultados obtenidos, comparándolos y contrastándolos con los métodos existentes, a fin de considerar las fortalezas a sostener, y las debilidades a tener en cuenta respecto de la utilización de NS. De esta manera, el principal aporte que se espera alcanzar es la integración de NS a los DDM-MOS, bajo la hipótesis de que por sus características, NS permitiría flexibilizar y mejorar la exploración del espacio de búsqueda, y con ello mejorar los resultados de las predicciones ya sea en términos de tiempo de ejecución o de precisión de la predicción. Alcanzar al menos uno de los dos objetivos constituirá un logro importante, dado que aportará una herramienta de ayuda a la decisión más fiable, que podría utilizarse para el manejo de incendios forestales, tanto a nivel local, como nacional o internacional, en caso de catástrofe.

### **4. FORMACIÓN DE RECURSOS HUMANOS**

La temática propuesta por este plan de trabajo permite continuar con la formación de los distintos integrantes del grupo de trabajo, de forma complementaria a la formación adquirida hasta el momento en el campo de los

métodos predictivos, del paralelismo, y de reducción de incertidumbre. En el caso particular de Dr. Strappa, esta temática se relaciona con su plan de trabajo postdoctoral (se incorporó al laboratorio en 2021 con una beca postdoctoral de CONICET, dirigida por la Dra. Paola Caymes Scutari y codirigida por el Dr. Germán Bianchini,), lo cual constituye un aporte a su formación en investigación.

En el marco de la línea general de los DDM-MOS, en 2020 se defendieron dos tesis doctorales relacionadas con esta línea de investigación, desarrolladas por la Dra. Laura Tardivo (bajo la dirección de la Dra. Paola Caymes Scutari) y el Dr. Miguel Méndez Garabetti (bajo la dirección del Dr. Germán Bianchini y la codirección de la Dra. Paola Caymes Scutari).

El grupo también está integrado por estudiantes de grado que están comenzando a introducirse en los temas atinentes a la línea de investigación, y asimismo las puertas del grupo de trabajo siempre están abiertas a la incorporación de nuevos integrantes (de grado o postgrado) que deseen familiarizarse con las temáticas con las que se trabaja dentro del mismo.

## 5. BIBLIOGRAFÍA

- [1] Bianchini, Germán, Denham, Mónica, Cortés, Ana, Margalef, Tomàs, Luque, Emilio. (2010). "Wildland Fire Growth Prediction Method Based on Multiple Overlapping Solution". *Journal of Computational Science Vol 1 Issue 4*, pp. 229-237. Elsevier.
- [2] Lehman J. and Stanley K. O.. (2011). "Abandoning Objectives: Evolution Through the Search for Novelty Alone," *Evolutionary Computation*, vol. 19, no. 2, pp. 189–223.
- [3] Lehman J. and Stanley K. (2008). "Exploiting Open-Endedness to Solve Problems Through the Search for Novelty," *Artificial Life*, p. 8.
- [4] Lehman J. and Stanley K. (2013). "Evolvability Is Inevitable: Increasing Evolvability without the Pressure to Adapt," *PLoS ONE*, v. 8, 4, pp. 2–10.
- [5] Méndez Garabetti, Miguel, Germán Bianchini, María Laura Tardivo, and Paola Caymes Scutari. (2015). "Comparative Analysis of Performance and Quality of Prediction Between ESS and ESS-IM." *Electronic Notes in Theoretical Computer Science* 314 (June): 45–60.
- [6] Mendez Garabetti, Miguel, Bianchini, Germán, Caymes Scutari, Paola, Tardivo María Laura. (2016). "Increase in the quality of the prediction of a computational wildfire behavior method through the improvement of the internal metaheuristic". *Fire Safety Journal*, pp. 49-62. Elsevier.
- [7] Naono, K., Teranishi, K., Cavazos, J. y Suda, R. (2010). *Software Automatic Tuning: From Concepts to State-of-the-Art Results*, Springer, New York.
- [8] Price K., Storn R., Lampinen J. (2005). *Differential Evolution - A practical approach to global optimization*. Springer-Verlag New York.
- [9] Talbi, E. (2009). *Metaheuristics: From Design to Implementation*. John Wiley & Sons, Hoboken, New Jersey.
- [10] Tardivo, María Laura, Paola Caymes Scutari, Germán Bianchini, and Miguel Méndez Garabetti. (2017). "Hierarchical Parallel Model for Improving Performance on Differential Evolution" *Concurrency and Computation: Practice and Experience* 29 (10): e4087.
- [11] Tardivo M. L., Caymes Scutari P., Méndez Garabetti M., and Bianchini G. (2018). "Optimization for an Uncertainty Reduction Method Applied to Forest Fires Spread Prediction," in *Computer Science 2017*, vol. 790, A. E. De Giusti, Ed. Cham: Springer International Publishing. pp. 13–23.
- [12] Tardivo M. L., Caymes Scutari P., Bianchini G., and Méndez Garabetti M. (2019). "Sintonización Dinámica del Método Paralelo de Predicción de Incendios Forestales ESSIM-DE," *XXV Congreso Argentino de Ciencias de la Computación*, p. 10, 2019.

# SIMULACIÓN Y ANÁLISIS EFICIENTE DE GRANDES VOLÚMENES DE DATOS

Mercedes Barrionuevo, LIDIC, FCFMyN, UNSL

Alicia Castro, LIDIC, FCFMyN, UNSL

Gil Costa G. Veronica LIDIC, FCFMyN, UNSL

Danilo Labella, FCFMyN, UNSL

Fernando Loor LIDIC, FCFMyN, UNSL

A.Marcela Printista, LIDIC, FCFMyN, UNSL

Cristian Tissera, LIDIC, FCFMyN, UNSL

Guillermo Trabes LIDIC, FCFMyN, UNSL

## CONTEXTO

En este trabajo se presentan líneas de investigación que combinan “Simulación y análisis eficiente de grandes volúmenes de datos”, las cuales pertenecen al proyecto PROICO “Tecnologías Avanzadas aplicadas al Procesamiento de Datos Masivos” de la UNSL. El proyecto tiene como metas principales investigar nuevas metodologías para el procesamiento eficiente y eficaz de grandes volúmenes de datos y proponer nuevas tecnologías para la obtención de información útil. Por otro lado, pretende avanzar en temas de investigación actuales, incorporando el análisis y uso de técnicas de ciencia de datos. El procesamiento eficiente de grandes volúmenes de datos (Big Data) y la ciencia de datos son dos áreas de la informática que han crecido y llamado la atención de diferentes disciplinas, tanto académicas como de negocios. En este sentido será posible detectar anomalías, descubrir patrones en los conjuntos de datos, encontrar posibles relaciones entre los datos, etc. y por supuesto obtener análisis que sirvan como respaldo para la toma de decisiones.

## RESUMEN

El crecimiento en el volumen de datos, así como en la heterogeneidad de los tipos de datos que existen en el mundo ha crecido exponencialmente en la última década. Esto implica dos grandes desafíos para la comunidad científica: (1) cómo procesar esos datos lo más rápido y eficiente posible y (2) cómo obtener información a través de dichos datos. Por otro lado, este gran volumen de datos brinda una oportunidad única en el área de modelado y simulación. Estos datos pueden ser utilizados para generar modelos de diferentes situaciones y aplicaciones que puedan ser validados y posteriormente escalados.

En este trabajo, se presentan los objetivos, trabajo realizado y desafíos que aborda el grupo de investigación de la Universidad Nacional de San Luis, para abordar los temas que involucran el procesamiento eficiente de grandes volúmenes de datos, para realizarlo en forma eficiente de forma tal de reducir los tiempos de ejecución, evitar saturar los recursos, balancear la carga de trabajo y reducir consumos de energía, así como para obtener información efectiva de esos datos que puedan ser utilizados para la toma de decisiones utilizando técnicas de ciencias de datos y finalmente utilizar esos datos para generar modelos de simulación que permitan analizar situaciones que no son factibles de ser implementadas en

la realidad por temas de costos, disponibilidad de recursos, etc.

*Palabras clave:* Grandes volúmenes de datos, procesamiento eficiente, análisis de datos, simulación, resiliencia.

## 1. INTRODUCCIÓN

Durante la última década, Big Data (Grandes Datos) se ha convertido en un fuerte foco de interés mundial, que atrae cada vez más la atención de la academia, la industria, el gobierno y otras organizaciones. El término "Big Data" apareció por primera vez en las comunidades científicas a mediados de la década de 1990, y actualmente se sugiere como fuente predominante de innovación, competencia y productividad. El creciente flujo de datos, que proviene de los diferentes tipos de sensores, sistemas de mensajería y redes sociales, además de sistemas de medición y observación más tradicionales, ya ha invadido muchos aspectos de nuestra existencia cotidiana. Por otro lado, tanto en las ciencias como en la industria, los proyectos a gran escala representados como sistemas complejos se han convertido en una tangible realidad, siendo la naturaleza de los mismos descentralizada, no lineal y generalmente compuestos por una multitud de unidades autónomas, que pueden interactuar, evolucionar, y adaptar su comportamiento a cambios en el entorno; lo que implica que en la mayoría de los casos es muy difícil (sino imposible) conseguir soluciones analíticas resolubles que sean capaces de describir un sistema complejo de manera adecuada. Utilizadas de manera apropiada, las diferentes técnicas de simulación proporcionan un método poderoso e indispensable para analizar una diversidad de problemas, ya que permite una comprensión cualitativa y cuantitativa de muchos fenómenos que son demasiados complejos para ser tratados con métodos analíticos o cuyos experimentos tienen limitantes temporales, económicas o éticas. La combinación de estos dos aspectos, Simulación y Grandes Volúmenes de Datos, nos permitirá abordar el estudio de diferentes escenarios de una aplicación social, incluyendo un gran número de personas y una importante variedad de características ambientales, haciendo más sencillo el estudio de comportamientos complejos que surgen de la interacción de las multitudes. Nuestra propuesta podría ser utilizada para instaurar políticas de seguridad ante posibles desastres y contribuir con acciones de prevención para lograr formas eficientes de evacuar ciudades o espacios

cerrados a través del diseño de políticas activas que minimicen el tiempo de evacuación y pongan al resguardo rápidamente a las personas.

## 2. INVESTIGACIÓN Y DESARROLLO

### 2.1. MODELOS COMPUTACIONALES PARA EL PROCESAMIENTO EFICIENTE CON SoCs

Hoy en día, el desarrollo de algoritmos se focaliza en efectuar los cálculos de manera tal de conseguir un buen rendimiento, logrando a su vez una eficiencia energética. Tecnologías como arreglo de compuertas lógicas programables en campo (field programmable gate array - FPGA) y los sistemas en chip (SoC) basados en FPGA (FPGA/SoC) han demostrado que es factible la aceleración de aplicaciones computacionalmente intensivas, disminuyendo el consumo de potencia gracias a la capacidad del alto paralelismo que puede obtenerse con estas plataformas y a la reconfiguración de la arquitectura.

Como se ha mostrado en trabajos anteriores [M22, GC22,GC22\_1,M21], las FPGAs pueden ser exitosamente utilizados para acelerar el procesamiento de grandes volúmenes de datos. En particular, los trabajos citados se centran en la ejecución de un algoritmo de re-ranking de documentos para la Web.

Sin embargo, estas arquitecturas tienen la limitación de la cantidad de recursos que pueden utilizarse para ejecutar los algoritmos y además, el tiempo de sintetización suele ser elevado. Con lo cual hace que estas tecnologías sean difíciles de ser ampliamente utilizadas. En otras palabras, los ciclos de diseño existentes para FPGA/SoC son largos debido a la complejidad de la arquitectura.

Para reducir los costos de desarrollo y despliegue de algoritmos sobre las FPGAs, una alternativa es utilizar un modelo computacional como un estimador de rendimiento e integral para la aceleración por hardware, el cual tiene como objetivo predecir eficazmente la utilización de recursos y la latencia para FPGA/SoC, reduciendo la brecha entre las aplicaciones y las arquitecturas FPGA/SoC. Integrar modelos de costo para la predicción del rendimiento y un motor de exploración del espacio de diseño permite proporcionar información de alto nivel a los desarrolladores de hardware.

La Fig. 1 muestra el diseño de un modelo para estimar el rendimiento integral de aplicaciones para la aceleración por hardware. El modelo trabaja con micro-benchmarks y restricciones (constraints) para ajustar los parámetros de la placa a utilizar, también utiliza diferentes enfoques como autómatas y algunas técnicas que pueden ser definidas por el usuario para realizar la exploración (DSE) de todas las directivas posibles que pueden ser utilizadas para mejorar la ejecución del algoritmo.

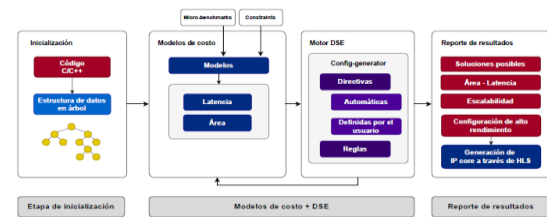


Fig. 1: Esquema general de modelo para estimar el rendimiento de aplicaciones en SoC.

### 2.2. CROWDSOURCING COMO BASE PARA EL PROCESAMIENTO ESCALABLE

Las plataformas de crowdsourcing han sido ampliamente utilizadas para realizar tareas en forma colaborativa. Cuanto mayor es el impacto de la situación o desastre ocurrido, mayor es la respuesta de la gente para cooperar [TW17]. A estas personas se las denomina voluntarios digitales. Los voluntarios digitales han ayudado a recopilar información pertinente y a procesarla mucho más rápido de lo que podrían hacerlo solos los funcionarios o personas a cargo de la coordinación de actividades en las situaciones particulares o, inclusive, que los algoritmos de aprendizaje que deben ser entrenados para tales fines. Es decir, en caso de que por ejemplo se deban identificar o taggear elementos de imágenes, videos o audio, los voluntarios digitales pueden hacerlo a medida que se generan los datos a procesar, mientras que los algoritmos de aprendizaje automático deben tener un conjunto de datos suficientemente grande y luego ser entrenados, lo cual puede llevar mucho tiempo.

En este contexto, se propone investigar una solución que combina técnicas de modelado y simulación discreta, análisis de sensibilidad y optimización dinámica de parámetros, para clasificar en forma eficiente imágenes y videos utilizando una plataforma de crowdsourcing que se construye sobre una red P2P.

En este sentido se ha avanzado en el diseño de la plataforma incluye un servidor conectado a una red de voluntarios a través de dispositivos set-top-box (STB) de TV digital, que constituyen una red P2P. El servidor envía tareas con imágenes e información adicional relacionada a desastres naturales a los usuarios de la red P2P y recupera los resultados. Las tareas consisten en una imagen y una lista de opciones o categorías, sobre las cuales los voluntarios deben enviar una respuesta. El servidor reúne las respuestas y evalúa si hay consenso para cada tarea, existiendo la posibilidad de reenviar las tareas en caso de ser necesario. La conexión entre la red P2P y el servidor se realiza mediante un proveedor de servicios de internet (ISP) [LOOR22][MAN19].

Además, la solución sobre la que se desea avanzar incluye: (I) El desarrollo de un simulador de una plataforma de crowdsourcing desplegada sobre una red P2P, (II) El análisis y la caracterización de los parámetros más relevantes a través de métodos de análisis de sensibilidad, (III) El diseño de algoritmos de

planificación de distribución de tareas y (IV) La optimización dinámica de parámetros de la plataforma para garantizar una ejecución eficiente. En particular la Fig. 2 muestra un esquema de cómo se pueden usar índices métricos para cumplir con este último punto. Sin embargo, también es posible llevarlo a cabo utilizando técnicas de aprendizaje por refuerzo u otras técnicas.

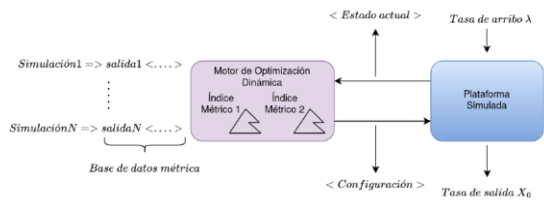


Fig. 2: Esquema general del método de ajuste de parámetros.

### 2.3 SIMULACIÓN DE EVACUACIONES EN CIUDADES COSTERAS

Existen diversas técnicas reportadas en la literatura para la simulación de evacuación de personas en situaciones de emergencias, como incendios, terremotos y tsunamis. La mayor parte de los trabajos presentan escenarios donde se estudian principalmente los tiempos de evacuación y densidad de personas en las zonas de encuentro [AG19]. En base a estos resultados, se crean, diseñan y evalúan los planes de evacuación que deberán ser socializados a los habitantes de la zona respectiva. Por lo tanto, la generación de escenarios con distintos parámetros, por ejemplo, diferentes tipos de población, intensidades de sismos y niveles de inundación, puede ser una herramienta de gran utilidad. Esto permitiría crear planes de evacuación efectivos y que se acerquen a la realidad, luego de analizar los resultados de las simulaciones de los distintos escenarios. En este sentido, se pretende investigar el desarrollo de un método para realizar un análisis de múltiples escenarios en forma eficiente utilizando herramientas como análisis geoespacial y minería de datos para generar mapas de predicción que apoyen los respectivos planes de evacuación. El trabajo presentado en [AST22] presenta una primera aproximación para obtener esta metodología mediante un simulador basado en agentes denominado DEMPS. Este trabajo utiliza el protocolo ODD [GRIMM20] para describir el modelo desarrollado. El simulador desarrollado tiene un entorno visual que permite visualizar el avance de los agentes, definir las zonas seguras de evaluación, las zonas iniciales donde están los agentes y además reportar datos estadísticos como se muestra en la Fig. 3.

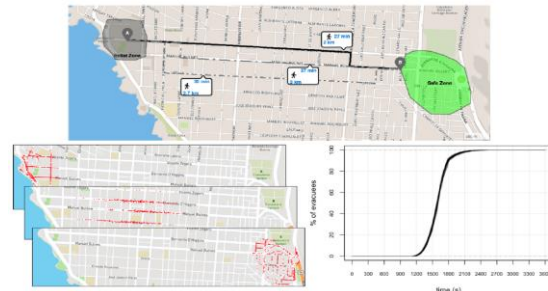


Fig 3. Simulador basado en agentes.

### 2.4. SIMULACIONES DE EVACUACIONES EN AMBIENTES CERRADOS

Otra área donde el grupo de investigación tiene experiencia es en simulaciones de evacuaciones en ambientes cerrados. En este contexto, a pesar del surgimiento de numerosas reglamentaciones, aún existen numerosas edificaciones con capacidad de alojar cientos o miles de individuos en su interior, donde es posible observar que cualquier incidente podría comprometer la seguridad de sus ocupantes. La tarea de predecir la performance de una evacuación para grandes y complejas estructuras con cientos de individuos representa una labor compleja, la cual se ve incrementada cuando se considera la interacción entre los individuos y la presencia de fuego y humo en el ambiente. Realizar estudios preventivos que permitan analizar las consecuencias de futuros siniestros permitiría tomar las medidas de seguridad necesarias para evitar que tales catástrofes ocurran. Desde este punto de vista, las simulaciones de las dinámicas pedestres en situaciones de emergencia son una herramienta de vital importancia, ya que son capaces de mostrar un gran número de características existentes en el tráfico de peatones, permitiéndonos de esta manera tener una mejor comprensión de los principios fundamentales que rigen este fenómeno. Además, dan la posibilidad de explorar los efectos que tienen en el modelo el cambio de reglas o de la geometría del espacio físico, sin la necesidad de realizar costosos experimentos, los cuales incluso pueden ser peligrosos o hasta imposibles de realizar. Mediante la modelización, es posible estudiar nuevos diseños o políticas en el sistema sin la necesidad de construir los ambientes físicos y las situaciones por ellas modeladas, permitiendo detectar errores en los diseños. Además, cuentan con las ventajas de evitar los riesgos de experimentación, y de poder analizar su comportamiento y resultados de forma segura.

De acuerdo con lo expuesto, se pretende dar continuidad a la investigación y desarrollo de un nuevo framework de simulación tipo cloud "Software as a Service" (SaaS), basado en el modelo de interacción de autómatas celulares y agentes inteligentes, presentado en [TI14] que permite instanciar múltiples realidades o escenarios a ser estudiados mediante la experimentación simulada.

Se pretende además en este contexto, investigar tanto impacto como las relaciones de los paradigmas y técnicas



de simulación asociadas al campo de la ciencia de datos, haciendo especial hincapié en la utilización de las mismas en las etapas tempranas de preprocesamiento. Finalmente, se desea investigar el uso y la aplicación de técnicas y herramientas de visualización de datos para el análisis de información obtenida mediante simulación utilizando el framework propuesto.

## 2.5. LA VISUALIZACIÓN COMO HERRAMIENTA PARA LA INTERPRETACIÓN DE DATOS

Como se mencionó anteriormente, la ciencia de datos es un área que ha llamado la atención en muchas disciplinas y está siendo usada en diferentes ámbitos tanto académicos como industriales. Los mercados financieros, en particular, el mercado de las criptomonedas es un área de interés para la comunidad científica teniendo en cuenta dos puntos de vista [BARI]: tecnológico y económico/financiero.

La ciencia de datos permite usar técnicas para procesar los datos relacionados con el problema a estudiar, en tiempo real y para eso hay que buscar la forma de leer datos y extraer conocimiento en un tiempo razonable. Aquí es un punto crucial donde se une ciencia de datos con diseño de algoritmos y infraestructuras eficientes para grandes volúmenes de datos. Por ejemplo, desde el punto de vista económico y financiero, las investigaciones tienden a analizar todas las características financieras de las criptomonedas no sólo por su volumen de negociación sino también por su comportamiento distintivo frente a activos tradicionales y por el hecho de que las criptomonedas se comercializan las 24 horas del día, los 7 días de la semana, planteando nuevos desafíos sobre cómo procesar de manera eficiente un flujo continuo de grandes volúmenes de datos. Dada su reciente popularidad, estos conjuntos de datos son cada vez más grandes. La existencia de más de 1.000 tipos de criptomonedas en el mundo, y la certeza de que seguirán creciendo, tanto en número como en popularidad, es muestra de su auge y su entrada activa en el mercado financiero internacional.

Por lo tanto, el tratamiento de dichos datos involucra la aplicación de una gran variedad de herramientas para su análisis y exploración. Una de las técnicas de interés, dentro de las etapas de la ciencia de datos, es la visualización de datos, que permite transmitir información por medio de representaciones visuales interactivas de datos, siendo una herramienta adecuada para enfrentarnos a la gran cantidad de datos que se manejan actualmente como así también para su correcta interpretación. En este contexto, el objetivo general es contribuir al desarrollo de soluciones en torno al análisis visual de indicadores relevantes para el mercado de las criptomonedas.

En este sentido, se ha desarrollado un prototipo con interacciones dinámicas que busca responder preguntas iniciales que dan el puntapié inicial para seguir trabajando con dichos datos.

## 2.6. PRESTACIONES SOCIALES Y RESILIENCIA

Este proyecto integra el Consorcio de I+D+I en Cloud Computing-Big Data & Emergent Topics, el cual propone un modelo de trabajo inclusivo y orientado a adaptarse a las necesidades de una sociedad que requiere incorporar los Objetivos de Desarrollo Sustentable. En esa dirección, el presente proyecto se plantea dos objetivos específicos: a) analizar cómo las tecnologías de transformación digital que se utilizan en sus distintas líneas (tales como datos, información y conocimiento) tienen implicancia en las prestaciones sociales (human centric) que se abordarán; y b) analizar el grado de Resiliencia de las aplicaciones descritas en los apartados anteriores y desarrollar para éstas un *indicador de performance clave (KPI)* que muestre como un sistema, cuyo estudio se realiza por medio de simulación, está alcanzando sus objetivos.

## 3. RESULTADOS OBTENIDOS/ESPERADOS

Las líneas de investigación y las técnicas descritas en la sección anterior, involucran una serie de desarrollos individuales que en su conjunto logran obtener el objetivo planteado. Este objetivo contempla por un lado el modelado y diseño de sistemas complejos mediante diferentes herramientas, algunas de las cuales han sido desarrolladas por los integrantes del proyecto. Mediante estas herramientas se pretende facilitar el entendimiento del sistema, realizar *profiling* a la ejecución de los algoritmos y mejorar el análisis de resultados. Por otro lado, se propone utilizar técnicas de ciencias de datos para analizar los datos obtenidos y encontrar comportamientos o tendencias que puedan ser utilizadas en diferentes situaciones que requieran soporte a la toma de decisiones. Finalmente, el procesamiento de este gran volumen de datos debe realizarse en forma eficiente y en el menor tiempo posible. Esto implica investigar formas de despliegue de los algoritmos en diferentes plataformas y tecnologías que permitan lograr el objetivo deseado.

Los resultados obtenidos hasta el momento son:

- El equipo de investigadores y estudiantes, han implementado herramientas de simulación que se encuentran disponibles en GitHub.
- Se han obtenido publicaciones en revistas indexadas y congresos internacionales.
- Se han dictado cursos de posgrado relacionados con la temática (en Argentina y en Chile).
- Se ha participado en proyectos de cooperación internacional como Stic-Amsud y CNR-CONICET.
- Se obtuvo el premio: Industry Impact Award. ECIR 2022. Paper: Ensemble Model Compression for Fast and Energy-Efficient Ranking on FPGAs.
- En marzo del 2023 finalizó una co-tutela con la Universidad de Trieste, Italia, para una tesis de doctorado en Ciencias de la Computación.
- En 2022, finalizó una tesis de doctorado y una tesis de maestría en colaboración con la Universidad de Santiago de Chile.

Los resultados esperados son:

- Avanzar en formular nuevas metodologías y técnicas para la adquisición, tratamiento y análisis de datos masivos.
- Investigar el uso de simuladores eficientes y eficaces para evaluar el rendimiento de nuevos algoritmos para aplicaciones de sistemas de gran escala con datos masivos.
- Abordar problemas reales que involucren datos reales, desarrollando soluciones eficientes para el análisis de grandes volúmenes de datos.
- Continuar formando estudiantes de posgrado en la temática.
- Capacitar profesionales en la temática mediante talleres y cursos de posgrado.

#### 4. FORMACION DE RECURSOS HUMANOS

Actualmente, se cuenta con tres doctores en ciencias de la computación realizando la investigación y dirección de los doctorandos involucrados en las líneas. Se cuenta con dos estudiantes de doctorado, los cuales se encuentran en la etapa final de su trabajo. El primer estudiante realiza su tesis en el desarrollo en el análisis, diseño y evaluación de factibilidad de plataformas distribuidas eficientes para el procesamiento de grandes cantidades de datos en el menor tiempo posible. En este trabajo se combinan técnicas de simulación y aprendizaje por refuerzo. El tema de estudio del segundo estudiante de doctorado, incluye el modelo, diseño e implementación eficiente de simuladores para procesar aplicaciones de gran escala en arquitecturas heterogéneas. También se cuenta con una integrante que ha concluido el trabajo final de la especialización en Inteligencia de Datos Orientada a Big Data utilizando técnicas de visualización para el tratamiento de datos financiero y se cuenta con un estudiante de la Lic. en Ciencias de la Computación. Mediante este trabajo de investigación se podrán formar profesionales que puedan modelar, diseñar e implementar algoritmos eficientes (previamente evaluados y analizados con un mínimo porcentaje de error) que se ejecuten en sistemas complejos donde intervienen una gran cantidad de variables y requieren el procesamiento de datos masivos. Por otro lado, se investigará en el uso de técnicas de ciencia de datos para obtener información a partir de los datos procesados y que puedan ser utilizados en la toma de decisiones. Finalmente, como se comentó anteriormente, los datos obtenidos para el procesamiento eficiente y obtención de patrones mediante ciencia de datos, pueden ser utilizados para generar modelos y simular escenarios alternativos.

#### 5. BIBLIOGRAFÍA

[AG19] Aguilar, L., Wijerathne, L., Jacob, S., Hori, M., & Ichimura, T. (2019). Mass evacuation simulation considering detailed models: behavioral profiles, environmental effects, and mixed-mode evacuation. *Asia Pacific Management Review*.

[AST22] Efficient simulation of natural hazard evacuation for seacoast cities Astudillo Muñoz, G., Gil-

Costa, V., Marin, M. *International Journal of Disaster Risk Reduction*, 2022

[GC22] V. Gil-Costa, F. Loor, R.Molina, F. Maria Nardini, R.Perego, S. Trani. Ensemble Model Compression for Fast and Energy-Efficient Ranking on FPGAs. *ECIR 2022*: 260-273

[GC22\_1] V. Gil-Costa, F. Loor, R. Molina, F. Maria Nardini, R. Perego, S.Trani: Energy-Efficient Ranking on FPGAs through Ensemble Model Compression . *IIR 2022*

[LOOR22] F. Loor, M. Manriquez, V. Gil-Costa, M.Marin. Feasibility of P2P-STB based crowdsourcing to speed-up photo classification for natural disasters. *Clust. Comput.* 279-302 (2022)

[M21] R. Molina, F. Loor, V. Gil-Costa, F.Maria Nardini, R.Perego, S.Trani. Efficient traversal of decision tree ensembles with FPGAs. *J. Parallel Distributed Comput.* 155: 38-49 (2021)

[M22] R. Soledad Molina, V.Gil-Costa, M.L. Crespo, G.Ramponi. High-Level Synthesis Hardware Design for FPGA-Based Accelerators: Models, Methodologies, and Frameworks. *IEEE Access* (2022).

[MAN19] M. Manriquez, F. Loor, V. Gil-Costa and M. Marin. "A Digital Tv-based Distributed Image Processing Platform for Natural Disasters" *WinterSim 2019*

[GRIMM20] V. Grimm, S.F. Railsback, C.E. Vincenot, U. Berger, C. Gallagher, D.L. DeAngelis, B. Edmonds, J. Ge, J. Giske, J. Groeneveld, et al., The ODD protocol for describing agent-based and other simulation models: A second update to improve clarity, replication, and structural realism, *J. Artif. Soc. Soc. Simul.* (2020).

[BARI] I.Merediz-Solà and A.Bariviera. A bibliometric analysis of bitcoin scientific production. *Research in Int. Business and Finance*, 2019

[TW17] Twigg and Mosel(2017). Emergent groups and spontaneous volunteers in urban disaster response.*Environment and Urbanization*,29(2), 443-458.

[TI14] (2014) Tissera C, Modelo Basado en Autómatas Celulares extendidos para Diseñar Estrategias de Evacuaciones en Casos de Emergencias. UNSL.