

TFG: Echoes of Gaia

Framework de modelado evolutivo basado en
simulación de biomas artificiales enfocado en sistemas
multiagente e Inteligencia Artificial

Aingeru García Blas

Índice

0. Reflexión inicial y agradecimientos	2
1. Introducción	2
2. Módulos, agentes y fases	3
2.1. Módulos	3
2.1.1. Simulación y Bioma	3
2.1.2. Research Hub	3
2.1.3. Entorno de entrenamiento	4
2.1.4. Visor de datos	4
2.2. Agentes	4
2.3. Fases	5
2.4. Precauciones y contingencias	5
2.4.1. Extensión	5
2.4.2. Backups	5
3. Objetivos	6
3.1. Objetivos simplificados ordenados.	6
3.2. Milestones	6
3.2.1. Milestone 1: Implementación del módulo de simulación y bioma	7
3.2.2. Milestone 2: Implementación del visor de datos básico	7
3.2.3. Milestone 3: Algoritmos genéticos y entrenamiento de agentes Reinforcement	7
3.2.4. Milestone 4: Research Hub: Training, generación y sistema de ingesta	7
3.2.5. Milestone 5: Desarrollo del modelo basado en Deep Learning	8
3.2.6. Milestone 6: Inferencia y conexión con la demo-videojuego	8
3.2.7. Milestone 7: Análisis, redacción y presentación de la memoria	8

0. Reflexión inicial y agradecimientos

Me gustaría dar las gracias a **GeoAI Tech Global**, la empresa con la que realicé las prácticas y he acordado la realización de mi TFG, ya que me han brindado la oportunidad de elegir, diseñar, planificar y llevar a cabo el proyecto por mi cuenta - gracias por la confianza. Considero que aunque el proyecto no responde a un problema de la empresa directamente, puede servir como herramienta experimental o prototipo para la generación y análisis de datos en entornos simulados, ya que como se explicará posteriormente la base consiste en un framework que permite disponer de un espacio controlado para probar hipótesis sobre interacciones ecológicas, evolución de especies, fenómenos medioambientales, etc.

También quiero expresar mi gratitud a **Iñaki Inza** por su siempre-constante amabilidad y por su apoyo en la decisión de centrarme en un proyecto que realmente me apasione - considero que trabajar en algo que me haga disfrutar tantísimo como este proyecto - me impulsa a intentar dar lo mejor de mí.

-
- **Empresa:** GeoAI Tech Global.
 - **Director de TFG:** Iñaki Inza.

1. Introducción

La idea central del proyecto es que haga la labor de un framework de simulación para la generación y análisis de biomas artificiales inteligentes, donde el eje principal es un sistema multiagente. Pero el objetivo se enfoca principalmente en **modelar la evolución de los ecosistemas de estos biomas** - de esta manera me permitirá tanto automatizar y observar, como evaluar las interacciones entre las diferentes entidades y componentes, esto es: flora, fauna, factores ambientales y climáticos...etc. En él, se introducirá también el componente humano, cuya idea es la de poder estudiar su impacto en el bioma y así observar los patrones emergentes.

Para llevar a cabo esto, emplearé diversas técnicas avanzadas de **Inteligencia Artificial**; desde redes neuronales profundas (**secuenciales** y **multi-label**, probablemente), *Reinforcement Learning* (DQN), hasta *algoritmos genéticos*.

Mi objetivo principal es hacer una especie de '**oda a la Inteligencia Artificial**'; aprovechar esta oportunidad y culminar mi viaje académico experimentando en un contexto que me apasiona para aplicar los conocimientos que he adquirido en relación a diferentes áreas de Inteligencia Artificial y Machine Learning, pero produciendo una herramienta basándome en mi experiencia en el mundo del software de manera que sea lo más modular y extensible posible; con objeto de que pueda ser de utilidad para otros programadores. Como nota extra, también será un entorno altamente configurable por el usuario.

Los datos producidos por la herramienta, así como los modelos e incluso el motor de simulación, son versátiles y no están ligados a un dominio específico, así que son extrapolables a multitud de áreas diferentes. Con lo que, considero que es un proyecto de grandísimo interés, del cual se podrá extraer información interesante con la predicción medioambiental como eje principal.

2. Módulos, agentes y fases

Para contextualizar un poco mejor, explico brevemente los módulos principales que lo componen:

2.1. Módulos

2.1.1. Simulación y Bioma

Este módulo está compuesto a su vez de dos submódulos: por un lado el **motor de simulación** (**SimPy** utilizaré internamente, para gestión del tiempo y sus escalas) y por otro lado todos los componentes que constituyen el **Bioma** y su entorno. Será un sistema reactivo - de esta manera los agentes y cualquier entidad del bioma puede reaccionar sin necesidad de disponer de eventos programados, para ello he diseñado una arquitectura basada en eventos - EDA (Event Driven Architecture, concretamente la vertiente inspirada en publisher-subscriber, mediante hooks, event dispatcher etc).

En cuanto a la representación del bioma - no será 3D, sino un simple mapa basado en un grid 2D - la idea es poder enfocar el esfuerzo tanto en desarrollar un framework robusto como en trabajar las técnicas de IA/ML y algoritmos que lo hacen inteligente; y no perderme en las complejidades añadidas de un entorno 3D.

Los biomas dispondrán de mapas procedurales los cuales generaré (en principio) con **Perlin Noise** o **Wave Function Collapse** (WFC). Como se ha mencionado anteriormente, estoy enfocándolo desde un punto de vista de diseño muy modular, lo cual, junto con un sistema de eventos reactivo, intento facilitar las cosas a la hora de introducir nuevos escenarios, variaciones climáticas, diferentes comportamientos en los agentes, etc.

Entre los múltiples paradigmas y patrones de diseño que utilizaré, en este punto merece la pena mencionar el **ECS (Entity Component System)**, para poder inyectar comportamientos, estados, tipos de movimientos e incluso agentes, entre otros - de manera sencilla y además dinámica y en tiempo de ejecución. Estos se comunicarán únicamente mediante exposición de APIs.

2.1.2. Research Hub

Entorno *base* de experimentación y entrenamiento de modelos - la idea es que su función principal sea la de lanzar experimentos y generar datos masivamente con diferentes configuraciones, para posteriormente analizarlos.

A parte de las estadísticas tradicionales, también se implementará un visor para poder observar la evolución de los biomas en sus diferentes eras y/o episodios. Un episodio, estará constituido por muchas eras.

2.1.3. Entorno de entrenamiento

Principalmente se centrará en el entrenamiento de dos áreas distintas: los agentes de **Reinforcement Learning** por un lado, y por otro, en el modelo predictivo de **red neuronal** que finalmente escoja.

Con respecto a *Reinforcement*, utilizaré un modelo **Deep Q-Network (DQN)**: o bien **Stable Baselines 3** o bien uno implementado por mí mismo **from scratch**, se decidirá más adelante. En este caso, el objetivo es el de hacer que los agentes aprendan por sí mismos a interactuar con el entorno y seguir su instinto de supervivencia - la fase de entrenamiento se dará durante la simulación en modo **train**. En una primera instancia y por tener cuanto antes una versión simple y funcional del entorno, crearé 2-3 individuos por especie. De ésta manera, tendré 2-3 modelos y con ello las políticas aprendidas para, posteriormente en inferencia, poder distribuir las aleatoriamente entre el número x de individuos de esa especie. Si aprenden decentemente, entonces pasaré a un entrenamiento más complejo: **Especies clusterizadas**; un modelo por especie, donde cada individuo contribuye con sus experiencias y todos los individuos poseen la política del modelo compartido.

Para el modelo de red neuronal, en un principio, tengo en mente utilizar un modelo basado en secuencias, dada la naturaleza del histórico de datos secuenciales del Bioma. Seguramente utilice una **LSTM**, ya que tengo algo de experiencia reciente con ellas (**Lore Nexus**) - aunque esta vez seguramente fuese **multi-label** (gracias a Iñaki por la sugerencia). No obstante, también es posible que finalmente me decante por un modelo **Transformer**. Sea como fuere, la red aprenderá mediante ingesta de los datos producidos por la simulación en modo **inferencia** y el objetivo es el de **predecir futuros estados del Bioma**.

2.1.4. Visor de datos

Por un lado, se podrán visualizar estadísticas y métricas de manera tradicional. Por otro, existirá un visor para poder observar la evolución del bioma y analizar sus estados individualmente. Por supuesto, la idea es poder analizar tendencias y patrones en la interacción entre los agentes y el entorno del bioma.

2.2. Agentes

Existen varios agentes que conforman los pilares de la plataforma, y varios agentes que únicamente se introducirían una vez disponga de una versión mínima funcional:

- **Agente de fauna:** Interactúa con el entorno. Toma de decisiones basada en **Reinforcement**.
- **Agente evolutivo:** Modifica atributos de flora y fauna. **Algoritmos genéticos**.
- **Agente de navegación:** Cálculo de rutas óptimas - en principio, únicamente algoritmo **A***. Guía a la entidad *host* del agente.
- **Agente de predicción:** Red neuronal; el **Cerebro** que, durante inferencia o simulación normal, será capaz de predecir estados del bioma futuros (información con respecto a el clima, la flora y fauna, especies...etc).

-
- **Agente climático:** Simula y predice variaciones ambientales.
 - **Agente de adaptación:** Para demo de videojuego - basado en reglas, en función de predicción dada por el **Cerebro**, capacidad de tomar decisiones y adaptación.

- **Agente humano:** En un principio, únicamente expansión. Implica impacto negativo en medioambiente, deforestación etc.

2.3. Fases

Con objeto de exponer y las diferentes fases del **pipeline** que el framework gestionará, veamos el siguiente listado:

- **Simulación + train:** Se entrenan los agentes de Reinforcement dentro del bioma.
- **Simulación en inferencia:** Lanzar simulaciones con los agentes ya entrenados para generar datos.
- **Entrenamiento del *Cerebro* (modelo predictivo):** Nutriéndolo de los datos generados, entrenar el modelo para predecir estados futuros del bioma.
- **Prototipo-Demo en Videojuego:** La idea es demostrar las posibilidades del framework mediante una demo interactiva, con los modelos cargados y simulación en modo inferencia. Para un estado dado, basándose en el histórico, el jugador puede recibir *insights* por parte del **Cerebro** con respecto al futuro del Bioma.

2.4. Precauciones y contingencias

2.4.1. Extensión

Conforme vaya avanzando en el desarrollo, evaluaré elevar la complejidad de los sistemas - se hará gradualmente para no dificultar el aprendizaje de los modelos. Estoy diseñando el framework desde el inicio con vistas a esto, de manera que disponga de soporte para una sencilla gestión. Es decir, en una primera instancia intentaré disponer de un **MVP** (*minimum viable product*), con versiones del entorno y agentes lo más simplistas posibles. Una vez funcional esta versión, entonces será cuando se puedan tomar decisiones de cuán complejos pueden ser los cambios climáticos, qué nuevos comportamientos base implementar, otorgar *rewards* más avanzados que involucren más estados, etc.

2.4.2. Backups

Es un sistema complejo, con lo que es posible que, dadas las sinergias entre diferentes agentes, resulte complicado de entrenar y llevar a cabo. Es por ello que en el caso de bloqueo en cuanto al aprendizaje de los agentes de fauna que utilicen políticas obtenidas con *Reinforcement Learning*, su lógica sea reemplazada por un sistema **basado en reglas**. Tengo diseñado a modo de bosquejo la interacción entre el entorno de RL (basado en *gym* si utilizo SB3) y el API del motor de simulación. No obstante, hasta que no me centre en la implementación es complicado dilucidar cómo aprenderán las políticas los agentes. Reinforcement Learning, es un tema delicado.

Considero que, en el caso de tener problemas y, si voy excediendo los tiempos estimados para los hitos (expuestos posteriormente), me veré forzado a omitir alguna parte - como la vertiente de demo a modo de videojuego, entre otros. No obstante, me gustaría poder implementar esta parte.

3. Objetivos

3.1. Objetivos simplificados ordenados.

1. Idear, investigar, planificar. Decidir tecnologías, módulos, paradigmas, patrones de diseño y estructurar el proyecto.
2. Tener funcional el motor de simulación y el bioma (mapas, clima, eventos reactivos, flora con algoritmos genéticos, sin agentes de fauna (*R.L*)).
3. Sobre ese bioma, una vez funcional - entrenar la fauna con *R.L*.
4. Mediante **Research Hub** ejecutar simulaciones para producir datos (*mediante lanzamiento de episodios*). Logs exhaustivos. Los datos serán la **representación del estado** del bioma en un punto dado - además serán autoetiquetados.
5. Entrenamiento del modelo basado en secuencias para intentar que comprenda las relaciones que ocurren en el bioma, su evolución y patrones emergentes.
6. En este punto se podría - gracias a los resultados ya disponibles - sacar conclusiones, métricas, gráficas etc.
7. Una vez funcional, comenzar a introducir factores que introduzcan complejidad elevada. El primero será introducir factor evolutivo en la fauna (hasta este momento, solo existirá en flora), pero después se pasará otros, como introducir el impacto humano, crear un nuevo comportamiento, ampliar los diferentes tipos de clima base, introducir nuevas posibilidades en el sistema evolutivo...etc.
8. Crear un módulo de videojuego simplista basado en la simulación, a modo de demo. El cual puede reutilizar y gestionar simulación a tiempo real gracias al API del motor de simulación, y cargar los modelos entrenados. De ésta manera, la demo puede ofrecer *insights* sobre futuros estados del bioma.

3.2. Milestones

He trazado un **roadmap** a modo de tentativa, en función a los tiempos estimados. No expongo en detalle, omito número de horas estimadas etc, con lo que para simplificar redacto incluyendo fines de semanas.

Como nota extra, he realizado varias tareas ya relacionadas al Milestone 1, tras el listado se podrán visualizar dos imágenes relacionadas, a modo de orientación.

3.2.1. Milestone 1: Implementación del módulo de simulación y bioma

- **Fecha:** 1 - 15 de febrero 2025
- Architecture base. Construcción del framework y su infraestructura: foco en arquitectura modular y escalable. Sobre todo, extensible.
- Implementación del motor de simulación (SimPy para la gestión del tiempo virtual). Exponer API.
- Desarrollo del bioma y el esqueleto de sus componentes. Exponer API.
- Mapas generados proceduralmente (Wave Function Collapse u otro algoritmo).
- Eventos reactivos e integración con los diferentes componentes y agentes (EventDispatcher, hooks).
- Definición de agentes e interacciones. Únicamente a nivel de arquitectura en este punto.
- Tests de integración y flujo para comprobar interconexión entre submódulos.

3.2.2. Milestone 2: Implementación del visor de datos básico

- **Fecha:** 16 - 20 de febrero 2025
- Programación del visor inicial - representación básica del bioma y navegación **secuencial**.
- Logs exhaustivos. Formato fichero y probablemente BD (considerar series temporales, me gusta InfluxDB. Si InfluxDB, quizá Grafana).
- Pruebas iniciales con simulaciones para validar resultados.

3.2.3. Milestone 3: Algoritmos genéticos y entrenamiento de agentes Reinforcement

- **Fecha:** 21 de febrero - 18 de marzo 2025
- Impl. de los algoritmos genéticos para sistema evolutivo de flora.
- Implementación de modelos DQN (RL) para los agentes de fauna.
- Refinamiento del sistema de rewards.
- Con esto, ya se puede extender y mejorar el visor de datos.
- Si todo correcto, inyectar agente evolutivo especializado a las entidades de fauna.

3.2.4. Milestone 4: Research Hub: Training, generación y sistema de ingesta

- **Fecha:** 19 de marzo - 2 de abril 2025
- Soporte para entrenamiento de los modelos e interconexión entre módulos para lanzamiento de simulaciones.
- Infraestructura de generación y almacenamiento para los datos.
- Para sistema de ingesta, algo simple, no es a tiempo real así que con generadores y carga en batches puede funcionar muy bien.

3.2.5. Milestone 5: Desarrollo del modelo basado en Deep Learning

- **Fecha:** 3 - 21 de abril 2025
- Diseño e implementación de la red. Validación.
- Aplicaré diferentes técnicas para mejorar el modelo, aunque esto es dependiente del modelo final que escoja. Utilizaré Pytorch y, si LSTM, implementaré **from Scratch**. Si Transformer, utilizaré modelo pre-entrenado de Huggingface y fine-tunearé.
- Análisis de métricas y recolección de las mismas para memoria.
- Más mejoras al visor de datos para representar predicciones de modelo, a parte de métricas.
- Si todo correcto, introducir nuevos comportamientos, agentes etc. Elevar complejidad del bioma.

3.2.6. Milestone 6: Inferencia y conexión con la demo-videojuego

- **Fecha:** 22 de abril - 20 de mayo 2025
- Impl. el sistema de inferencia del modelo en la simulación.
- Desarrollo de la demo interactiva. Base ya está hecha, implementé a ratos libres sistema modular de escenas y esqueleto base, utilizando **pygame**.
- Gracias a las APIs expuestas, una vez cargados los modelos y entornos, el juego puede comunicarse sencillamente y obtener información del bioma a través del motor de simulación.

3.2.7. Milestone 7: Análisis, redacción y presentación de la memoria

- **Fecha:** 20 de mayo - 22 de junio 2025 (último día para entrega de documentación).
- Análisis final de la rendimiento de modelos y evolución del bioma.
- Documentación del proceso, experimentos, resultados etc. Redacción de memoria.

Como se ha mencionado anteriormente, pongo dos imágenes para ilustrar tímidamente, a modo de ejemplo, el formato en el que definiré el resto de milestones. Relaciones y roadmap.

Estructura base del framework y Bootstrapping #1

geru-scotland opened 2 days ago

Crear la estructura base del framework, principalmente de la parte del `Bioma`, un primer brochazo.

También crear la inicialización, pero con el pipeline y sus capas (comunicación entre APIs de `bioma` y `simulación`, para poder ser utilizadas, desde módulo de `Research`).

Sub-issues 2 of 4

- Mapas procedurales #2
- Logger específico para el bootstrapping #3
- Refactorizar bootstrap. #4
- Flujo base con Simpy #7

geru-scotland self-assigned this 2 days ago

geru-scotland moved this to Backlog in TFG - Echoes of Gaia 2 days ago

geru-scotland added this to TFG - Echoes of Gaia 2 days ago

geru-scotland added this to the Fase 1: Setup Framework completo + generación de mapas procedurales. milestone 2 days ago

geru-scotland 2 days ago

Update db62332

geru-scotland added a sub-issue 2 days ago

- Mapas procedurales #2

geru-scotland added core high-priority yesterday

geru-scotland added a sub-issue yesterday

- Logger específico para el bootstrapping #3

geru-scotland moved this from Backlog to Ready in TFG - Echoes of Gaia yesterday

geru-scotland added sub-issues yesterday

- Refactorizar bootstrap. #4
- Flujo base con Simpy #7

geru-scotland moved this from Ready to In progress in TFG - Echoes of Gaia yesterday

TFG - Echoes of Gaia

Kanban Roadmap Backlog + New view

Filter by keyword or by field

January 2025 February 2025

Item	Priority	Assignee	Due Date
1	High	geru-scotland	Feb 15, 2025
2	High	geru-scotland	Feb 15, 2025
3	High	geru-scotland	Feb 15, 2025
4	High	geru-scotland	Feb 15, 2025
5	High	geru-scotland	Feb 15, 2025
6	High	geru-scotland	Feb 15, 2025
7	High	geru-scotland	Feb 15, 2025

4 de Febrero de 2025