



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA INFORMÁTICA

**MÁSTER UNIVERSITARIO EN COMPUTACIÓN LÓGICA E
INTELIGENCIA ARTIFICIAL**

**ANÁLISIS DE LAS TRANSICIONES DE
COMPORTAMIENTO EN CRECIMIENTO DE TUMORES
USANDO UNA SIMULACIÓN CON AUTÓMATA
CELULAR**

**Realizado por
Sergio Rodríguez Calvo
XXXXXXXXXL**

**Supervisado por
Dr. Miguel Ángel Gutiérrez Naranjo**

**Departamento
Ciencias de la Computación e Inteligencia Artificial**

Sevilla, octubre 2017

Realizado por
Sergio Rodríguez Calvo
sergiorodriguezcalvo@gmail.com

Support: Memoria para el trabajo fin de máster

ÍNDICE GENERAL

| | |
|--|-----------|
| 1. Resumen | 5 |
| 2. Introducción | 7 |
| 2.1. Objetivos | 8 |
| 2.2. Alcance | 8 |
| 2.3. El cáncer | 8 |
| 2.4. Los autómatas celulares | 10 |
| 3. Modelo de eventos | 11 |
| 3.1. Las células, genoma y propiedades | 11 |
| 3.2. Parámetros globales | 12 |
| 3.3. Pruebas previas a la mitosis | 13 |
| 3.4. Mitosis | 13 |
| 3.5. Eventos mitoticos | 14 |
| 3.6. Equivalencia temporal | 14 |
| 3.7. Bucle principal del modelo de eventos | 15 |
| 4. Implementación | 17 |
| 4.1. Módulo <i>genome.py</i> | 17 |
| 4.2. Módulo <i>simulation_globals.py</i> | 18 |
| 4.3. Módulo <i>cell.py</i> | 18 |

| | |
|--|-----------|
| 4.4. Módulo <i>experiments.py</i> | 18 |
| 4.5. Módulo <i>grid.py</i> | 18 |
| 4.6. Módulo <i>automata.py</i> | 19 |
| 5. Experimentos de los autores | 21 |
| 5.1. Influencia del parámetro <i>Tasa de mutación base (m)</i> | 21 |
| 5.1.1. Experimento 1: Tasa de mutación base igual a 10.000 | 22 |
| 5.1.2. Experimento 2: Tasa de mutación base igual a 1.000 | 22 |
| 5.1.3. Experimento 3: Tasa de mutación base igual a 100 | 22 |
| 5.2. Influencia del resto de marcadores | 22 |
| 6. Resultados propios | 23 |

ÍNDICE DE FIGURAS

RESUMEN

El cáncer es un nombre genérico que agrupa más de 200 enfermedades que causan proliferación descontrolada de células provocando la aparición de masas anormales. A toda masa anormal se le conoce como neoplasia y, según su inavisiibilidad y factor de crecimiento, puede ser maligna (carcinoma) o benigna (adenoma).

Esta enfermedad puede ser vista como un comportamiendo emergente, que puede ser explicado a partir de la presencia de ciertos marcadores cancerosos a nivel local.

José Santos y Ángel Monteagudo, autores de una serie de artículos y, entre ellos, el artículo en el que se centra este trabajo, proponen el uso de autómatas celulares como técnica para simular dicho comportamiendo dada su similitud e idoneidad.

Se propone una modelización en el cual se utiliza una rejilla en tres dimensiones donde, sin tener en cuenta el tamaño de la célula, se presenta una única célula en el centro de dicha rejilla. Cada célula, cuenta con un genoma asociado que representa la aparición de mutaciones, así como, dos propiedades: tamaño del telomero, y tasa de mutación base.

Todo ello, es utilizado dentro de un modelo de eventos, en el cual, se programan eventos mitóticos en el futuro y en el que se consideran una serie de pruebas previas a la mitosis. Dichas pruebas, modelizan el comportamiendo individual de cada célula a la hora de ejecutar la división, así como, todos aquellos factores que pueden derivar en la muerte de la célula y factores provocados por las mutaciones adquiridas.

Finalmente, se realizan una serie de simulaciones, utilizando una serie de parámetros globales concretos en cada caso, para estudiar la evolución y el comportamiento del sistema.

INTRODUCCIÓN

Desde el punto de vista de la biología computacional el cáncer puede ser visto como un sistema ecológico y de comportamiento emergente. Lo cual, hace apropiado el uso de autómatas celulares para simular su comportamiento. Es por ello que, cada vez más se opta por el uso de simulaciones por ordenador como la que se presenta en este trabajo ayudando, desde el estudio y comprensión del cáncer, hasta el estudio y desarrollo de nuevas terapias contra él.

El cáncer es una enfermedad cada vez más común debido a sus muy diversos factores que propician su aparición, esto es, mayor esperanza de vida de la población y, por tanto, individuos cuyas células han sufrido más mutaciones, factores ambientales y/o formas de vida poco saludables, mayor estrés de la población, y hasta exposición a fuentes de radiación artificiales.

A pesar de todo, el cuerpo tiene una serie de mecanismos para prevenir la aparición y la proliferación de esta enfermedad. Estos, son muchos y muy diversos, así como, los mecanismos propios de las células que intervienen en este proceso. En esta simulación, por tanto, se pretenden tener en cuenta el enfoque de los autores del trabajo original, eligiendo sólo un subconjunto de todas estas características y comportamientos.

La elección de un autómata celular como enfoque para modelizar el comportamiento de esta enfermedad proviene por la relativa facilidad de obtener comportamientos complejos a partir de una reglas relativamente sencillas a nivel local en cada célula.

Este trabajo, como se ha comentado previamente, pretende reproducir el trabajo de José Santos y Ángel Monteagudo de Julio de 2014, uno de los muchos trabajos que dichos autores han realizado siguiendo la idea de realizar simulaciones sobre esta enfermedad y, por ello, aquí se realiza una nueva implementación con el fin de reproducir los resultados obtenidos en el citado artículo.

La implementación consiste en seguir un modelo de eventos, es decir, se pretende simular el proceso que siguen las células para su división, que consiste en, la realización de una serie de pruebas para comprobar si la célula debe reproducirse, la reali-

zación de la propia división y, la programación de un nuevo evento de división en el futuro. A lo largo de la ejecución del programa en cada iteración puede haber ninguna, una o varias células que deben seguir el proceso que se acaba de describir brevemente.

Finalmente, para estudiar el comportamiento del cáncer se realizan diversas ejecuciones utilizando diferentes parámetros y configuraciones, y observando en cada caso qué factor es el que provoca el cambio en el comportamiento global.

2.1 OBJETIVOS

En este trabajo, se esperan alcanzar los siguientes objetivos:

- Encontrar artículos científicos de calidad sobre esta temática.
- Extraer información del artículo (o los artículos).
- Elegir un artículo y hacer un modelado e implementación propia.
- Reproducir los resultados del artículo.
- Comparar los resultados obtenidos con los del artículo.

2.2 ALCANCE

El presente trabajo tiene como finalidad reproducir el trabajo de José Santos y Ángel Monteagudo de Julio de 2014, realizando una implementación propia del sistema que se describe en el presente documento.

No obstante, el comportamiento que se pretende simular cuenta con una serie de pasos en los cuales el comportamiendo dependerá de ciertos sorteos con una probabilidad dada. Esta es una limitación a tener en cuenta a la hora de obtener los mismos resultados que los autores de dicho trabajo.

2.3 EL CÁNCER

El cáncer es el nombre genérico usado para referirse a más de 200 enfermedades cuyo resultado es la proliferación descontrolada de las células que se ven afectadas.

Aunque no era común encontrar signos de cáncer en los seres vivos en la antigüedad debido a una menor esperanza media de vida, lo cierto es que es inherente al reino animal.

El origen etimológico del nombre proviene de la antigua Grecia, donde un médico llamado Hipócrates utilizaba los términos *carcino* y *carcinoma* para referirse a dicha enfermedad, y las cuales hacían referencia al cangrejo debido a la forma en la que se proyectaban los cánceres.

El cáncer se clasifica como:

- Una **enfermedad genética**, debido a, que es causada por un cambio en el material genético (ADN) de las células.
- Una **enfermedad multigénica**, ya que, afecta a diferentes genes.
- Una **enfermedad multifactorial**, es decir, tiene causas muy diversas.
- Una **enfermedad multiorgánica**, por tanto, afecta a diferentes tejidos y órganos.

Se conoce como neoplasia a toda masa anormal que tenga lugar en el cuerpo. Esto es, la división descontrolada de la célula y el aumento de tamaño de cada una de ellas, y ocurre por un problema en el ADN de la célula, la cual, contiene la información acerca de cuáles son las funciones de la misma.

Al producirse en todo tipo de tejidos la evolución y el tratamiento de esta enfermedad varía. El factor determinante entre que este tipo de masas anormales sea benigna (adenoma) o maligna (carcinoma) es el factor de crecimiento y la invasividad de tejido adyacente.

Presenta diferentes fases, partiendo de un crecimiento anormal en tamaño, en la cual, cada vez necesitará más nutrientes. Esa necesidad de nutrientes hace que la célula entre en una segunda fase conocida como *angiogénesis*, en la cual, crea nuevos vasos sanguíneos para lograr su objetivo. Finalmente, la célula se vasculariza, es decir, partes de ella pasan a la sangre, continuando con este comportamiento allí donde se depositen y, además, en muchas ocasiones se observa metástasis.

Es obvio, que la célula para llegar a la última fase va a necesitar cierto grado de daño genético, así como, superar determinadas barreras. Por un lado, obtener todas las capacidades necesarias para poder dividirse sin control, crear sus propios vasos sanguíneos, etc. dependerá de la obtención de mutaciones en su ADN que permitan

todo esto. Por otro lado, el cuerpo tiene mecanismos para evitar que esto ocurra, como por ejemplo, muerte por daño genético, evitar enviar orden de división e, incluso, el propio sistema inmune interviene para intentar evitar su crecimiento entre otros.

Por tanto, esta es una enfermedad compleja de tipo ecológico y emergente, lo cual, dificulta su estudio, así como, su tratamiento.

2.4 LOS AUTÓMATAS CELULARES

Un autómata celular es una de las formas más adecuadas de simular comportamientos emergentes, al igual que los sistemas de Lindenmayer y, por tanto, adecuado para este tipo de problemas. José Santos y Ángel Monteagudo optaron por utilizar autómatas celulares para sus trabajos.

Consiste en modelo matemático que modela sistemas dinámicos. Está inspirado en cómo ciertas especies han resuelto el problema de qué color o forma tomar en su aspecto exterior, como por ejemplo, es el caso de ciertas conchas marinas. Este mecanismo, al ser favorable desde el punto de vista evolutivo, les ha hecho sobrevivir gracias a una mejor mimetización con su entorno.

Este modelo necesita de unas propiedades, tales como, dimensión, estados posibles de sus celdas, qué se considera vecindario de una célula y, por último, cuales son las reglas de evolución. Además, hay que definir cómo es su frontera.

Es, por tanto, muy adecuado para ser implementado en ordenadores.

MODELO DE EVENTOS

Los autores, José Santos y Ángel Monteagudo, decidieron utilizar un autómatas celular siguiendo un modelo de eventos. Es decir, por un lado, la simulación se realiza sobre una rejilla en tres dimensiones, comenzando con una única célula en el centro de la misma. Por otro lado, se programan una serie de eventos para cada una de las células de la rejilla, aleatoriamente entre 5 y 10 (ambos inclusive) iteraciones en el futuro.

Las características y propiedades del sistema se describen en las secciones de este capítulo que se muestran a continuación.

3.1 LAS CÉLULAS, GENOMA Y PROPIEDADES

Cada célula estará alojada en una única posición del autómatas. En esta simulación, no se modeliza el tamaño de las células, es decir, aunque en las células cancerosas se observa, además de un comportamiento replicativo sin control, un crecimiento en su tamaño sin control, los autores no han tenido en cuenta esto.

El genoma de cada célula presente en la rejilla cuenta con un genoma y unas propiedades únicas para cada una de ellas. En cuanto a su genoma, cuenta con 5 variables binarias que representan la presencia o no de una mutación, la cual, define un determinado comportamiento canceroso. Estas mutaciones son las siguientes:

- **SG:** Autogeneración de los mensajes de crecimiento. Esto es, la mutación que permite que la célula genere sus propios mensajes para ejecutar la división con independencia externa.
- **IGI:** Inhibición de las señales de anticrecimiento. Esto es, ante la recepción de una orden de detener su crecimiento, la célula tiene una mutación que le permite un mecanismo de ignorancia de los mismos.
- **EA:** Evasión de apoptosis. Esto es, la célula puede, mediante mutación, no hacer caso ante una orden de apoptosis, o muerte celular controlada.

- **EI:** Inmortalidad efectiva. Esto es, la célula adquiere una mutación que permite evitar un límite replicativo existente, entre otros factores, por el tamaño del telómero.
- **GI:** Inestabilidad genética. Esto es, una mutación que permite a la célula acumular más daño genético, es decir, la tasa de mutación base se va incrementando con el paso del tiempo.

Además, cada célula tiene una tasa de mutación base y un tamaño de telómero. El primero, es utilizado a la hora de añadir nuevas mutaciones a la célula. Y la segunda, es un límite replicativo debido a que el ADN queda desprotegido para futuras mutaciones y podrían ocurrir errores.

3.2 PARÁMETROS GLOBALES

De cara a la simulación, existen una serie de parámetros globales que inciden en determinados mecanismos y, por tanto, afectan en su evolución. Esto son:

- **t:** Tamaño de la rejilla.
- **m:** Valor por defecto que indica la tasa de mutación base de cada célula al inicio.
- **tl:** Valor por defecto que indica el tamaño del telómero de cada célula al inicio.
- **e:** Valor por defecto utilizado para definir la probabilidad de una célula de morir por daño genético.
- **i:** Valor por defecto utilizado como factor de aumento de la tasa de mutación base de las células.
- **g:** Valor por defecto para ver qué probabilidad hay de que una célula cancerosa mate a un vecino para poder reproducirse.
- **a:** Valor por defecto para ver con qué probabilidad una célula muere de forma aleatoria. Este parámetro se introduce como forma de simular las muy diversas causas que pueden originar en la muerte de la célula, por ejemplo, recibir una alta dosis de radiación entre otras.

3.3 PRUEBAS PREVIAS A LA MITOSIS

El modelo de eventos, a la hora de realizar la mitosis (división celular), realiza tres pruebas:

- **Prueba de muerte aleatoria:** La célula muere con una probabilidad dada ($1/a$).
- **Prueba de daño genético:** A mayor cantidad de mutaciones, mayor probabilidad de que la célula muera (n/e , donde n es el número de mutaciones de la célula). Excepto, que la célula tenga la mutación *EA* presente en su genoma.
- **Prueba de mitosis:** En realidad, se trata de tres pruebas cuyo resultado debe ser positivo para que la célula pueda ejecutar la mitosis:
 - **Comprobación del factor de crecimiento:** La célula sólo puede realizar la mitosis si se encuentra dentro de un límite espacial predefinido. Es decir, hay suficiente factor de crecimiento. Fuera de este área no podrá realizar la mitosis, excepto si el marcador *SG* está activo.
 - **Comprobación de ignorancia de inhibición de crecimiento:** Si no hay espacio en el vecindario, la célula no podrá realizar la división. Excepto, que la mutación *IGI* esté presente, en cuyo caso, reemplazará a un vecino con una probabilidad dada ($1/g$).
 - **Comprobación de potencial replicativo sin límites:** Si el tamaño del telómero es 0, la célula muere. Excepto si el marcador *EI* está presente.

En estas pruebas, se utilizan los parámetros globales, además, de las propiedades y el genoma de la célula.

Por tanto, si una célula pasa todas las pruebas podrá entrar en fase de mitosis.

3.4 MITOSIS

La mitosis es la fase donde se realiza la división celular, lo cual origina una nueva célula, llamada célula hija, que es una copia exacta de la célula original.

Durante este proceso, se pueden producir una serie de errores, bien de forma natural, o bien, debido mutaciones presentes en la célula original que hacen más probable un error.

En este trabajo, el proceso de mitosis cuenta con los siguientes pasos:

- Incremento de la tasa de mutación base m (de acuerdo al parámetro global i) si el marcador GI está presente.
- Añadir nuevas mutaciones a la célula de acuerdo a la tasa base de mutación $(1/m)$.
- Decrementar el telómero en ambas células en una unidad, siempre que su tamaño no sea 0.

A continuación, se programan nuevos eventos mitóticos para ambas hijas como se describe a continuación.

3.5 EVENTOS MITOTICOS

Un evento mitotico se puede ver como la posibilidad de que una célula ejecute o no la división (o mitosis), en el futuro.

Todas las células en la rejilla contarán con un evento mitótico en el futuro. Cuando se selecciona una de las candidatas en cada iteración la célula candidata debe pasar las pruebas previas a la mitosis para ver si la misma puede o no ejecutar la mitosis. Si el resultado es positivo, se realiza la mitosis y se crean nuevos eventos mitóticos en el futuro para la célula original y, también, para la célula hija. Si el resultado es negativo, pero la célula no ha muerto, se programa un nuevo evento mitótico en el futuro.

La programación de eventos mitóticos futuros consiste en la generación, de forma aleatoria, de un valor entre 5 y 10 iteraciones (ambas inclusive) del bucle principal. Estos valores tienen que ver con el ciclo de vida temporal de las células, sobre el cual, se detalla más información en la siguiente sección.

3.6 EQUIVALENCIA TEMPORAL

El ciclo de vida de las células biológicas es de entre 15 y 24 horas. Este ciclo se divide en 5 fases, que son: fase G0, fase G1, fase S, fase G2 y fase M.

Una célula parte en fase de reposo (G0). Si hay espacio suficiente a su alrededor (vecindario) automáticamente pasa a fase G1. En la simulación, G1 se simula mediante

el paso del tiempo (iteraciones) y la programación de eventos mitóticos en el futuro (entre 5 y 10 iteraciones, ambas inclusive). Además, en la simulación no se tiene en cuenta el crecimiento celular.

La fase S es cuando tiene lugar la replicación del ADN. La cual, puede introducir una mutación ocasionalmente. La célula entra en una última fase previa a la mitosis, llamada fase G2, en la cual se producen una serie de comprobaciones sobre el daño genético. Esto, puede provocar la apoptosis (muerte celular programada) en la célula.

Finalmente, la célula entra en la fase M o de mitosis. Todo este ciclo, que en la simulación toma entre 5 y 10 iteraciones (15 y 24 horas en células biológicas) da una media de 2,6 horas por iteración. Por ejemplo, para 5000 iteraciones se tienen unas 77,4 semanas aproximadamente.

3.7 BUCLE PRINCIPAL DEL MODELO DE EVENTOS

Prueba:

$$T_0 \quad j \leftarrow 1 \text{ in } n - 1 \text{ copiar}(T_0^{j-1}, \{T_1^j, \dots, T_n^j\}) \quad i \leftarrow 1 \text{ in } n \quad \bar{T}_i^j \leftarrow \text{quitar}(T_i^j, \{p_j r : r \neq i\} \cup \{p_k i : j + 1 \leq k \leq n\}) \text{ uni'on}(\{\bar{T}_1, \dots, \bar{T}_n\}, T^j) \quad T_{n-1}$$

IMPLEMENTACIÓN

Para realizar la implementación se ha optado por utilizar el lenguaje de programación *Python*, en su versión 3, concretamente, en el momento de la realización de este trabajo el desarrollo se ha realizado bajo la versión 3.6.2 de dicho lenguaje.

Python es un lenguaje multiparadigma que es ampliamente utilizado en ámbitos científicos y de investigación. Cuenta con multitud de herramientas para la realización de *tests*, gestión de dependencias, etc., entre las que se encuentran *Unittest* y *Pip*, que han sido empleadas en este trabajo.

Como librerías se han empleado *Numpy* y *Matplotlib*, las cuales, son casi un estándar dentro del mundo *Python*.

Para el desarrollo de este proyecto se ha optado por implementar la solución en orientado a objeto en su mayoría. El programa cuenta con un único punto de entrada, en este caso, siguiendo el patrón *Singleton*.

En este punto de entrada, se establecen los valores de los parámetros por defecto que se desean para cada ejecución. Por defecto, el programa utiliza los valores por defectos establecidos en uno de los módulos en los que se divide el proyecto, y que se comentarán más adelante. Dichos valores, coinciden con los valores por defectos que se presentan en el artículo.

En este caso, el programa cuenta con 6 módulos. El módulo principal, *automata.py* presenta una clase que sigue el patrón *Estrategia* y, el cual, engloba la función principal de ejecución, así como, el resto de funciones necesarias para su correcta ejecución. Además, importa y utiliza el resto de módulos, los cuales, se describen a continuación.

4.1 MÓDULO *genome.py*

En este módulo se presenta la clase que contienen como propiedades las cinco variables binarias que representan la activación o no de determinadas mutaciones cancero-

sas, así como, algunas funciones de utilidad para la simulación referentes al genoma.

Completar.

4.2 MÓDULO *simulation_globals.py*

Este módulo, contiene los valores, así como, la clase necesaria para su almacenaje, para los parámetros que intervienen en la simulación. Contiene la definición de constantes que representan dichos valores junto a sus valores por defecto.

Además, contiene otra serie de valores como constantes que son necesarios y que no forman parte de la clase *SimulationGlobals* y que no son susceptibles de variar.

4.3 MÓDULO *cell.py*

El módulo *cell.py* cuenta con una clase que modela la célula. Utiliza la composición para incluir el objeto *Genome* frente al uso de la herencia, lo cual, presenta ventajas de cara a la refactorización.

Cuenta, además de con el objeto *Genome*, con algunas propiedades extra, como por ejemplo, el tamaño del telómero (*tl*) y la tasa de mutación base (*m*) para cada célula.

Finalmente, presenta algunos métodos necesarios en la simulación para interactuar con la célula.

4.4 MÓDULO *experiments.py*

Completar.

4.5 MÓDULO *grid.py*

Completar.

4.6 MÓDULO *automata.py*

Completar.

EXPERIMENTOS DE LOS AUTORES

José Santos y Ángel Monteagudo realizaron una serie de simulaciones en su artículo donde intentan estudiar la influencia de los marcadores en la evolución de la simulación. Todas las simulaciones, y sus resultados, se comentan en esta sección.

El tamaño de rejilla utilizada para todos los experimentos es de 50^5 . Esto son, 125,000 células posibles en la rejilla.

5.1 INFLUENCIA DEL PARÁMETRO *Tasa de mutación base* (m)

Los autores en su artículo presentan 3 experimentos utilizando los valores por defecto y variando el parámetro *Tasa de mutación base* (m) para estudiar como afecta dicho parámetro en la proliferación del cáncer.

Cada experimento se muestra en una serie de gráficas en las cuales se presenta como resultado la media de 5 ejecuciones diferentes.

A continuación, se muestra cada uno de los experimentos especificando en cada caso el valor m utilizado. El resto de parámetros de la simulación se mantiene constante con los valores considerados por defectos, que son:

Cuadro 5.1: Valores de los parámetros, excepto m .

| Nombre | Símbolo | Valor |
|---|---------|-------|
| Tamaño del telómero | tl | 50 |
| Muerte por daño genético | e | 10 |
| Factor de incremento de tasa de mutación base | i | 100 |
| Muerte de un vecino | g | 30 |
| Muerte aleatoria | a | 1000 |

5.1.1 Experimento 1: Tasa de mutación base igual a 10.000

Pruebas.

5.1.2 Experimento 2: Tasa de mutación base igual a 1.000

Pruebas.

5.1.3 Experimento 3: Tasa de mutación base igual a 100

Pruebas.

5.2 INFLUENCIA DEL RESTO DE MARCADORES

Para observar el comportamiento del resto de marcadores, los autores, realizan una variación de los parámetros respecto al experimento anterior. Los valores utilizados son los siguientes:

Cuadro 5.2: Valores de los parámetros.

| Nombre | Símbolo | Valor |
|---|---------|---------|
| Tasa de mutación base | m | 100.000 |
| Tamaño del telómero | tl | 35 |
| Muerte por daño genético | e | 20 |
| Factor de incremento de tasa de mutación base | i | 100 |
| Muerte de un vecino | g | 10 |
| Muerte aleatoria | a | 400 |

Prueba.

RESULTADOS PROPIOS