Universidad Nacional Autónoma de México

Posgrado en Ciencia e Ingeniería de la Computación

RACIONALIDAD ACOTADA, LÓGICA Y PROGRAMACIÓN DINÁMICA APLICADAS A TEORÍA DE JUEGOS

T E S I S

QUE PARA OPTAR POR EL GRADO DE:
Doctor en Ciencias (Computación)

PRESENTA:
Eduardo Espinosa Avila

DIRECTOR DE TESIS
Dr. Francisco Hernández Quiroz
Facultad de Ciencias

COMITÉ TUTOR
Dra. Atocha Aliseda Llera
Instituto de Investigaciones Filosóficas
Dr. David Rosenblueth Laguette
Instituto d eInvestigaciones en Matemáticas Aplicadas y en Sistemas

Ciudad Universitaria, Cd.Mx. Noviembre 2017

Índice general

Índice de figuras

Resumen

El objetivo principal de esta tesis es buscar posibles respuestas a la pregunta: ¿Qué puede ser considerada una elección racional dadas ciertas limitaciones en las capacidades de un agente? Dicha pregunta es además el hilo conductor en el desarrollo de los trabajos presentados en esta tesis. Se proponen tres enfoques en la búsqueda de estas respuestas. El primer enfoque es una extensión de un modelo de racionalidad acotada aplicada a la clase de juegos alternados repetidos. El modelo extendido es aplicable a la clase de juegos dinámicos alternados y para obtenerlo se sigue una metodología estándar en teoría de juegos: establecer las características de la clase de juegos a la cual se aplica y definir el concepto de solución. El segundo enfoque utiliza una extensión de la lógica proposicional combinada con una semántica de vecindarios y una lógica de preferencias para obtener la lógica para racionalidad acotada $\mathcal{BRL}$. Con este lenguaje es posible expresar fórmulas que incluyan restricciones computacionales de forma concreta. El tercer enfoque es una aplicación de optimización acotada a teoría de juegos. Utilizando el principio de optimalidad se obtiene la ecuación de Hamilton–Jacobi–Bellman para el caso discreto del juego y, posteriormente, se obtienen soluciones explícitas con ayuda de la versión continua, es decir, haciendo que la cantidad de pasos del juego crezca indefinidamente. De esta forma es posible calcular estrategias óptimas acotadas. Más aún, el último enfoque proporciona una forma alternativa de cálculo para el equilibrio perfecto por subjuegos.

Agradecimientos

Agradezco a la UNAM por seguir cobijándome, permitiéndome además desarrollar proyectos que aporten a la comunidad universitaria y espero a lo sociedad en general.

Agradezco a mi tutor Francisco Hernández Quiroz por su guía, consejos, apoyo y paciencia para realizar y terminar esta tesis. A los miembros de mi comité tutor Atocha Aliseda y David A. Rosenblueth por el tiempo y consejos para obtener un mejor trabajo. Así mismo, a mis sinodales Pablo Padilla y Héctor Sánchez por los atinados comentarios y sugerencias sobre este escrito.

Agradezco tanto al CONACyT como a los proyectos PAPIIT IN113013 e IN400514 por los fondos brindados para la realización de este proyecto.

Agradezco y dedico este trabajo a mi familia, amigos y especialmente a Viri sin lo cuales no podría haber culminado este trabajo, gracias por su compañía, su respaldo, sus consejos y por ser fuente de inspiración durante todo este tiempo.

Hacer predicciones es muy difícil,
especialmente cuando se trata del futuro.
Niels Bohr[1]

[1]Prediction is very difficult, especially about the future. Niels Bohr

Capítulo 1

Introducción

Desde el nacimiento de la teoría matemática de juegos con el trabajo de von Neumann y Morgenstern [NM44], se han desarrollado diversos trabajos buscando conceptos de solución para diferentes clases de juegos. Entre los conceptos más conocidos y utilizados, se encuentran el equilibrio de Nash y el de equilibrio perfecto por subjuegos [Os04]. Un juego es una descripción de la interacción que se da entre agentes racionales que buscan maximizar su ganancia.

Estos conceptos pertenecientes a la llamada teoría *clásica* de juegos suelen hacer una suposición muy fuerte sobre las características de los agentes involucrados: poseen omnisciencia lógica. Esta característica implica que los agentes conocen (o pueden inferir) todas las implicaciones de su conjunto base de *conocimientos* (creencias), lo cual, desde el punto de vista computacional, puede a su vez traducirse en que los agentes poseen capacidad infinita, tanto de cómputo como de memoria.

Sin embargo, pronto fue notorio que esta suposición resultaba poco realista y se vio la necesidad de buscar formas alternativas para modelar situaciones de interacción entre agentes en competencia. Entre los pioneros, se encuentra Herbert A. Simon [Si55] quien introdujo por primera vez la idea de *racionalidad acotada* para analizar situaciones en las que no existe certeza sobre las situaciones futuras.

La guía principal de este trabajo de tesis es utilizar diferentes ideas de racionalidad acotada desarrolladas recientemente para aplicarlas a la teoría de juegos. Por otro lado, una pregunta fundamental para este trabajo es ¿Qué puede ser considerada una elección racional dadas ciertas limitaciones en las capacidades de un agente? Para tratar de responderla se utilizan tres enfoques de racionalidad acotada y se presentan ejemplos que ilustran su aplicación.

El primer enfoque toma en cuenta qué tanto tiene presente (*is aware*) el agente de ciertas características del juego y a partir de esto, desarrollo un modelo para la clase de juegos dinámicos de movimientos alternados, definiendo el concepto de solución asociado que toma en cuenta la racionalidad acotada a la capacidad de cómputo del agente que lo implemente.

La relación entre lógica y teoría clásica de juegos ha sido muy prolífica desde los primeros trabajos que se ocuparon de ella. Se han obtenido múltiples resultados, especialmente aplicando lógica epistémica, pero poco existe en la dirección de racionalidad acotada. En el segundo enfoque desarrollo un lenguaje de lógica para racionalidad acotada que permite expresar fórmulas que incluyan nociones de racionalidad acotada como las del modelo del primer enfoque. Combina ideas de lógicas de vecindarios y preferencias para poder expresar este tipo de fórmulas.

El tercer enfoque contiene la idea más novedosa de las aquí presentadas: aplicar optimización vía programación dinámica a teoría de juegos. Primeramente se obtienen versiones discretas para determinar las ganancias en los juegos analizados, para después obtener soluciones explícitas al resolver las versiones *continuas* de los mismos. Este enfoque fue muy bien aceptado y obtuvo buenos comentarios ya que contiene un algoritmo que permite calcular un equilibrio perfecto por subjuegos desde la programación dinámica.

La estructura de este trabajo de tesis sigue el desarrollo cronológico de trabajos desarrollados durante mis estudios de doctorado. El capítulo 2 presenta un marco genérico del estado del arte al iniciar este proceso. En el capítulo 3 está el desarrollo de mi primer modelo de aplicación de racionalidad acotada a teoría de juegos. El capítulo 4 contiene la lógica para racionalidad acotada que sirve para expresar algunas propiedades deseables. En el capítulo 5 aplico optimización vía programación dinámica en teoría de juegos. Por último, el capítulo 6 incluye conclusiones generales y posibles caminos para trabajos posteriores.

Con el fin de evitar que el presente trabajo se extienda demasiado, se ha intentado incluir los conceptos básicos mínimos para hacerlo autocontenido y las referencias pertinentes para aquellos lectores que deseen profundizar en dichos conceptos y temas relacionados.

Capítulo 2

Estado del arte

Resumen

En este capítulo se presenta un marco teórico de los conceptos que se utilizan en el desarrollo de cada uno de los enfoques mencionados. La sección de teoría de juegos 2.1 en conjunto con la sección de racionalidad acotada 2.3 son el fundamento para el modelo extendido del primer enfoque. Por otro lado, la sección Lógica y teoría de juegos 2.2 en combinación con las secciones 2.1 y 2.3 sirven de base para el desarrollo del lenguaje de la lógica para racionalidad acotada $\mathcal{BRL}$. Finalmente, la sección Optimización: programación dinámica 2.4 junto con las anteriores fundamentan el desarrollo del tercer enfoque.

2.1. Teoría de juegos

La teoría de juegos es un conjunto de herramientas analíticas que nos ayudan a entender los fenómenos que se observan cuando interactúan varios tomadores de decisiones. Esta teoría asume que los tomadores de decisión persiguen objetivos bien definidos y toman en cuenta su conocimiento (o suposiciones) sobre el comportamiento de los otros participantes.

2.1.1. Juegos y soluciones

Un juego es una descripción de la interacción estratégica que incluye las restricciones en las acciones que los jugadores pueden realizar, así como sus intereses, pero no especifica las acciones que los jugadores realmente toman. Una solución, es una descripción sistemática de los resultados que pueden surgir en un conjunto de juegos. La teoría de juegos sugiere soluciones racionales para clases de juegos y examina sus propiedades.

2.1.1.1. Juegos cooperativos y no cooperativos

En todos los modelos teóricos de juegos, la entidad básica es el jugador. Un jugador puede ser un individuo o un grupo de individuos tomando decisiones. Una vez definido el conjunto de jugadores, se pueden distinguir dos tipos de modelos: aquellos en los que el conjunto de posibles acciones de jugadores individuales son primitivas (no cooperativos), y aquellos en los que el conjunto de posibles acciones conjuntas de grupos de jugadores son primitivas (cooperativos).

2.1.1.2. Juegos estratégicos y extensivos

Un juego estratégico es un modelo de una situación en la que cada jugador elige su plan de acciones al principio del juego (y no cambia durante el desarrollo), y todas las decisiones de los jugadores se hacen simultáneamente (es decir, cuando eligen su plan de acción, cada jugador no está informado del plan de acción de ningún otro jugador). En contraste, el modelo de un juego extensivo, especifica el posible orden de eventos, cada jugador selecciona su plan de acción no solamente al principio del juego, sino cada vez que debe tomar una decisión.

2.1.1.3. Juegos con información perfecta e imperfecta

En un juego con información perfecta, los participantes están completamente informados acerca de los movimientos realizados por todos los jugadores, mientras que en un juego con información imperfecta no sucede así.

2.1.2. Comportamiento racional

Los modelos *clásicos* utilizados asumen que cada participante (jugador) es "racional" en el sentido de que está consciente de sus alternativas, formula expectativas sobre suposiciones, tiene claramente definidas sus preferencias y elige su acción deliberadamente después de algún proceso de optimización. Los siguientes elementos constituyen un modelo de elección racional:

- Un conjunto $\mathcal{A}$ de posibles *acciones* para los jugadores.
- Un conjunto $\mathcal{C}$ de posibles *consecuencias* de esas acciones.
- Una *función de consecuencia* $g : \mathcal{A} \to \mathcal{C}$ que asocia una consecuencia a cada acción.

• Una *relación de preferencia* (una relación binaria completa, transitiva y reflexiva) $\succeq$ sobre el conjunto $\mathcal{C}$.

Usualmente las preferencias de los tomadores de decisiones se especifican dando una *función de utilidad* (ganancia) $\mathcal{U} : \mathcal{C} \to \mathbb{R}$, que define una relación de preferencia $\succeq$ con la condición que $x \succeq y$ si y solo si $\mathcal{U}(x) \geq \mathcal{U}(y)$.

2.1.3. Juegos estratégicos y equilibrio de Nash

Ahora se presenta el modelo de interacción estratégica conocido como *juego estratégico* (juego en forma normal en la terminología de von Neumann y Morgensten). Este modelo especifica para cada jugador un conjunto de posibles acciones y un orden de preferencia sobre el conjunto de posibles perfiles de acción. El equilibrio de Nash es uno de los conceptos básicos en teoría de juegos.

2.1.3.1. Juego estratégico

Definición

Un juego estratégico es un modelo de toma de decisiones interactivas en el cual cada participante elige su plan de acciones al inicio, no cambia durante el desarrollo y las decisiones se realzan simultáneamente. El modelo consiste en:

- ♢ Un conjunto finito $\mathcal{N}$: el **conjunto de jugadores**.
- ♢ Para cada jugador $i \in \mathcal{N}$, un conjunto no vacío $\mathcal{A}_i$: el **conjunto de acciones** disponibles para el jugador i.
- ♢ Para cada jugador $i \in \mathcal{N}$, una relación de preferencia $\succeq_i$: la **relación de preferencia** del jugador i.

Si el conjunto $\mathcal{A}_i$ para cada jugador i es finito, entonces el juego es finito.

Interpretación

Una interpretación común de un juego estratégico es que es un modelo de un evento que ocurre sólo una vez; cada jugador conoce los detalles del juego y el hecho de que todos los jugadores son racionales y que los jugadores eligen sus acciones simultánea e independientemente. Bajo esta interpretación, cada jugador no sabe las decisiones tomadas por los otros jugadores al momento de elegir su acción; no hay información sobre si un jugador puede basar sus especulaciones acerca del comportamiento de los otros jugadores.

Otra interpretación es que un jugador puede formar sus especulaciones del comportamiento de los otros jugadores sobre la base de información acerca de la forma en que el juego o juegos similares fueron jugados en el pasado. Una secuencia de jugadas del juego puede modelarse por un juego estratégico solo si no existe relación entre las jugadas, es decir, un individuo que juega varias veces el juego solo estará preocupado de su pago en el juego actual e ignora los efectos que su acción actual tiene sobre el comportamiento futuro de los otros jugadores.

2.1.3.2. Equilibrio de Nash

El concepto de solución más usado comúnmente el teoría de juegos, es el de equilibrio de Nash. Esta noción captura un estado estable de una partida en un juego estratégico en la que cada jugador tiene especulaciones correctas acerca del comportamiento de los otros jugadores y actúa racionalmente. No intenta examinar el proceso por el cual se alcanzó el estado estable.

Definición

Un equilibrio de Nash de un juego estratégico $\langle \mathcal{N}, (\mathcal{A}_i), \succeq_i \rangle$ es un perfil de acciones $a^* \in \mathcal{A}$ con la propiedad de que para cada jugador $i \in \mathcal{N}$ se tiene:

$$\left(a^*_{-i}, a^*_i\right) \succeq_i \left(a^*_{-i}, a_i\right) \text{ para todo } a_i \in \mathcal{A}_i$$

Es decir, para que a^* sea un equilibrio de Nash no debe suceder que ningún jugador i tenga una acción que da un resultado que él prefiere al generado cuando elige a^*_i, dado que cualquier otro jugador j elige a^*_j. En resumen, ningún jugador puede desviarse rentablemente, dadas las acciones de los otros jugadores.

La siguiente reformulación de la definición resulta útil algunas veces. Para cualquier $a_{-i} \in \mathcal{A}_{-i}$, se define $B_i(a_{-i})$ como el conjunto de mejores acciones del jugador i dado a_{-i}:

$$B_i(a_{-i}) = \left\{a_i \in \mathcal{A}_i | \, (a_{-i}, a_i) \succeq_i \left(a_{-i}, a_i^{'}\right)\right\} \text{ para todo } a_i^{'} \in \mathcal{A}_i$$

El conjunto evaluado B_i se llama función de mejor respuesta del jugador i. Un equilibrio de Nash es un perfil de acciones a^* para las cuales:

$$a^*_i \in B_i\left(a^*_{-i}\right) \text{para todo } i \in \mathcal{N}$$

Esta formulación alternativa de la definición nos lleva a un método (no necesariamente eficiente) para obtener equilibrios de Nash: primero, calcular la función de mejor respuesta para cada jugador, después, encontrar un perfil de acciones a^* para el cual $a_i^* \in B_i\left(a_{-i}^*\right)$ para todo $i \in \mathcal{N}$. Si las funciones B_i son evaluadas a conjuntos unitarios, entonces el segundo paso es resolver $|\mathcal{N}|$ ecuaciones para las $|\mathcal{N}|\,(a_i^*)_{i\in\mathcal{N}}$ desconocidas.

2.1.3.3. Juegos estrictamente competitivos

Los juegos estrictamente competitivos son una clase de juegos en los cuales existen dos jugadores cuyas preferencias son diametralmente opuestas. Por facilidad, los nombres de los jugadores son "1" y "2" ($\mathcal{N} = \{1, 2\}$).

Definición

Un juego $\langle\{1,2\}, (\mathcal{A}_i), (\succeq_i)\rangle$ es estrictamente competitivo, si para cualquier $a \in \mathcal{A}$ y $b \in \mathcal{A}$, se tiene $a \succeq_1 b$ si y solo si $b \succeq_2 a$.

Un juego estrictamente competitivo también se llama suma cero porque si la relación de preferencia del jugador 1 se representa por la función de utilidad u_1 y la relación de preferencia del jugador 2 se representa por la función de utilidad u_2 entonces $u_1 + u_2 = 0$.

Se dice que el jugador i *max-minimiza* si elige una acción que es la mejor para él, asumiendo que cualquier cosa que él haga, el jugador j elegirá su acción con el fin de *dañarlo* tanto como le sea posible. En un juego estrictamente competitivo que posee un equilibrio de Nash, un par de acciones es un equilibrio de Nash si y solo si la acción de cada jugador es un *max-minimizador*.

Sea $\langle\{1,2\}, (\mathcal{A}_i), (u_i)\rangle$ un juego estrictamente competitivo. La acción $x^* \in \mathcal{A}_1$ es un max-minimizador para el jugador 1 si $min\, u_1(x^*, y) \geq min\, u_1(x, y)$; para $y \in A_2$ y para toda $x \in \mathcal{A}_1$. Similarmente, $y^* \in \mathcal{A}_2$ es un max-minimizador para el jugador 2 si $min\, u_2(x, y^*) \geq min\, u_1(x, y)$; para $x \in A_1$ y para toda $y \in \mathcal{A}_2$.

En palabras, un max-minimizador para el jugador i, es una acción que maximiza la utilidad que el jugador i puede garantizar. Un max-minimizador para el jugador 1 resuelve el problema $max_x min_y u_1(x, y)$ y un max-minimizador para el jugador 2 resuelve el problema $max_y min_x u_2(x, y)$.

2.1.4. Juegos extensivos con información perfecta

Un juego extensivo es una descripción detallada de la estructura secuencial de los problemas de decisión que enfrentan los jugadores en una situación estratégica.

Hay *información perfecta* en tales juegos si cada jugador, al tomar una decisión, tiene información perfecta de todos los eventos que ocurrieron previamente. Por simplicidad, inicialmente se restringe el estudio a juegos en los que los jugadores no toman decisiones al mismo tiempo y todos los movimientos relevantes los realizan los jugadores (no interviene el azar). Posteriormente se eliminan estas restricciones.

2.1.4.1. Definición

Un juego extensivo con información perfecta tiene los siguientes componentes.

♢ Un conjunto finito $\mathcal{N}$: el conjunto de **jugadores**.

♢ Un conjunto $\mathcal{H}$ de secuencias (finito o infinito) que satisface las siguientes tres propiedades.

- La secuencia vacía $\emptyset$ es miembro de $\mathcal{H}$.
- Si $\left(a^k\right)_{k=1,\ldots,K} \in \mathcal{H}$ (donde K puede ser infinita) y $L < K$, entonces $\left(a^k\right)_{k=1,\ldots,L} \in \mathcal{H}$.
- Si una secuencia infinita $\left(a^k\right)_k^\infty$ satisface $\left(a^k\right)_{k=1,\ldots,L} \in \mathcal{H}$ para cualquier entero positivo L, entonces $\left(a^k\right)_k^\infty \in \mathcal{H}$.

Cada miembro de $\mathcal{H}$ es una **historia**; cada componente de una historia es una **acción** tomada por un jugador. Una historia $\left(a^k\right)_{k=1,\ldots,K} \in \mathcal{H}$ es *terminal* si es finita o si no existe a^{K+1} tal que $\left(a^k\right)_{k=1,\ldots,K+1} \in \mathcal{H}$. El conjunto de historias terminales se denota por Z.

♢ Una función P que asigna a cada historia no terminal (cada miembro de $\mathcal{H} - Z$) un miembro de $\mathcal{N}$, P es la **función de jugador** y $P(h)$ es el jugador que elige una acción después de una historia h.

♢ Para cada jugador $i \in \mathcal{N}$, se tiene una **relación de preferencia** sobre Z.

En algunas ocasiones resulta conveniente especificar la estructura de un juego extensivo sin especificar las preferencias de los jugadores. Una tupla $\langle \mathcal{N}, \mathcal{H}, P \rangle$ cuyos componente satisfacen las tres primeras condiciones se conoce como **juego en forma extensiva con información perfecta**.

Si el conjunto $\mathcal{H}$ de historias posibles es finito, entonces el juego es finito. Si la historia más larga es finita, entonces el juego tiene un horizonte finito. Sea h

una historia de longitud k; entonces (h, a) denota la historia de longitud $k+1$ que consiste en h seguida de a.

Un juego extensivo con información perfecta (*juego extensivo* en el resto de la sección) se interpreta de la siguiente manera. Después de cualquier historia no terminal h, el jugador $P(h)$ elige una acción del conjunto $A(h) = \{a : (h, a) \in \mathcal{H}\}$.

La historia vacía es el punto inicial del juego algunas veces se refiere como *historia inicial*. En este punto, el jugador $P(\emptyset)$ elige algún miembro de $A(\emptyset)$; para cada posibilidad a^0 de este conjunto, el jugador $P(a^0)$ elige un miembro de $A(a^0)$; esta elección determina el movimiento del siguiente jugador y así sucesivamente. Una historia que no tiene más elecciones posibles es terminal; nótese que una historia puede ser una secuencia infinita de acciones. Frecuentemente se especifican las preferencias de los jugadores sobre historias terminales mediante funciones de utilidad que representan dichas preferencias.

2.1.4.2. Estrategias

Una estrategia de un jugador en un juego extensivo es un plan que especifica la acción elegida por el jugador para cada historia después de la cual es su turno para mover.

Definición

Una estrategia del jugador $i \in \mathcal{N}$ en un juego extensivo con información perfecta $\langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$ es una función que asigna una acción en $A(h)$ a cada historia no terminal $h \in \mathcal{H} - Z$ para la cual $P(h) = i$.

Nótese que la noción de estrategia de un jugador en un juego $\langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$ depende solo de la forma del juego $\langle \mathcal{N}, \mathcal{H}, P \rangle$.

Un hecho importante es que una estrategia especifica la acción a tomar para toda historia después de la cual es su turno, aún para historias que, si se sigue la estrategia, nunca se alcanzan.

Ahora, para cada perfil de estrategia $s = (s_i)_{i \in \mathcal{N}}$ en un juego extensivo $\langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$ se define la salida $O(s)$ de s como la historia terminal que resulta cuando cada jugador $i \in \mathcal{N}$ sigue los preceptos de s_i, es decir, $O(s)$ es la historia (posiblemente infinita) tal que para $0 \leq k < K$ se tiene $s_{P\left(a^1, \ldots, a^k\right)}\left(a^1, \ldots, a^k\right) = a^{k+1}$.

Al igual que en los juegos estratégicos, se puede definir una estrategia mixta como una distribución de probabilidad sobre el conjunto de estrategias (puras); sin embargo con información perfecta se agrega muy poco al considerar dichas estrategias. Por tanto, las estrategias mixtas se presentarán posteriormente junto con los juegos extensivos con información imperfecta, en los que esta noción tiene mayor significado.

2.1.4.3. Equilibrio de Nash

El primer concepto de solución que se define para juegos extensivos ignora la estructura secuencial del juego; considera las estrategias como elecciones que se toman una sola vez (y para siempre) antes de que inicie el juego.

Definición

Un equilibrio de Nash de un juego extensivo con información perfecta $\Gamma = \langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$, es un perfil de estrategia s^* tal que para todo jugador $i \in \mathcal{N}$ se tiene $O\left(s^*_{-i}, s^*_i\right) \succsim O\left(s^*_{-i}, s_i\right)$ para toda estrategia s_i del jugador i.

2.1.4.4. Equilibrio perfecto de subjuegos

Ahora se define la noción de equilibrio perfecto de subjuegos; primero se define la noción de subjuego. El subjuego de un juego extensivo con información perfecta $\Gamma = \langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$ que sigue a la historia h, es el juego extensivo $\Gamma(h) = \langle \mathcal{N}, \mathcal{H}|_h, P|_h, (\succeq_i |_h) \rangle$, donde $\mathcal{H}|_h$ es el conjunto de secuencias de acciones $h^{'}$ para las cuales $\left(h, h^{'}\right) \in \mathcal{H}$ está definida por $h^{'} \in \mathcal{H}|_h$ para cada $\succeq_i |_h$ y $h^{'} \succeq_i |_h h^{''}$ está definida por $\left(h, h^{'}\right) \succeq_i \left(h, h^{''}\right)$ si y solo si $h^{'} \in \mathcal{H}|_h$.

Esta noción de equilibrio requiere que la acción prescrita por cada estrategia de jugador sea óptima, dadas las estrategias de los otros jugadores, después de toda historia. Dada la estrategia s_i del jugador i y una historia h en un juego extensivo Γ, $s_i|_h$ denota la estrategia que s_i induce en el subjuego $\Gamma(h)$ (es decir, $s_i|_h = i\left(h, h^{'}\right)$, para cada $\Gamma = \langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$, O_h denota la función de salida de $\Gamma(h)$).

Un **equilibrio perfecto de un subjuego** de un juego extensivo con información perfecta $\Gamma(h) = \langle \mathcal{N}, \mathcal{H}|_h, P|_h, (\succeq_i |_h) \rangle$ es un perfil de estrategias s^* tal que para todo jugador $i \in \mathcal{N}$ y para toda historia no terminal $h \in \mathcal{H} - Z$ para la cual $P(h) = i$ se tiene $O_h\left(s^*_{-i}|_h, s^*_i|_h\right) \succeq_i |_h O_h\left(s^*_{-i}|_h, s_i|_h\right)$ para toda estrategia s_i del jugador i en el subjuego $F(h)$.

De forma equivalente, se puede definir equilibrio perfecto de un subjuego como un perfil de estrategias s^* en Γ en el cual, para cada historia h, el perfil de estrategia $s^*|_h$ es un equilibrio de Nash del subjuego $\Gamma(h)$. Esta noción de equilibrio perfecto de un subjuego elimina los equilibrios de Nash es los que las amenazas de los jugadores no son creíbles.

Lema: La propiedad de desviación

Sea $\Gamma = \langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$ un juego extensivo con información perfecta con horizonte finito. El perfil de estrategias s^* es un equilibrio perfecto de subjuego de Γ, si y solo si para todo jugador $i \in \mathcal{N}$ y para toda historia $h \in \mathcal{H}$ para la cual $P(h) = i$, se tiene $O_h\left(s^*_{-i}|_h, s^*_i|_h\right) \succeq_i |_h O_h\left(s^*_{-i}|_h, s_i\right)$ para toda estrategia s_i del jugador i en el subjuego $\Gamma(h)$ que difiere de $s^*_i|_h$ solo en la acción que prescribe después de la historia inicial de $\Gamma(\emptyset)$.

Proposición: Teorema de Kuhn

Todo juego extensivo con información perfecta tiene un equilibrio perfecto de subjuegos [Os04].

Prueba

Sea $\Gamma = \langle \mathcal{N}, \mathcal{H}, P, (\succeq_i) \rangle$ un juego extensivo con información perfecta. Se construye un equilibrio perfecto de un subjuego de Γ por inducción sobre $l(\Gamma(h))$ (la longitud de la historia más larga en Γ); y al mismo tiempo se define una función R que asocia una historia terminal con cada historia $h \in \mathcal{H}$ y se muestra que esta historia es una salida de equilibrio perfecto del subjuego $\Gamma(h)$.

- ♢ Si $l(\Gamma(h)) = 0$ (es decir, h es una historia terminal de Γ), entonces $R(h) = h$.

- ♢ Ahora, supóngase que $R(h)$ está definida para toda $h \in \mathcal{H}$ con $l(\Gamma(h)) \leq k$ para algún $k \geq 0$. Sea h^* una historia para la cual $l(\Gamma(h^*)) = k + 1$ y sea $P(h^*) = i$. Como $l(\Gamma(h^*)) = k + 1$ se tiene $l(\Gamma(h^*, a)) \leq k$ para toda $a \in A(h^*)$. Definiendo $s_i(h^*)$ como una $\succeq_i -maximizadora$ de $R(h^*, a)$ sobre $a \in A(h^*)$ y definiendo $R(h^*, s(h^*))$. Por inducción se ha definido un perfil de estrategia s en Γ; y por la propiedad de desviación (lema anterior) este perfil de estrategia es un equilibrio perfecto de un subjuego de Γ.

El procedimiento usado en esta prueba se conoce frecuentemente como inducción hacia atrás. Además de servir como prueba de la proposición, este procedimiento es un algoritmo para obtener el conjunto de equilibrios perfectos de subjuegos de un juego finito. Parte del atractivo de la noción de equilibrio perfecto de subjuegos deriva del hecho de que el algoritmo describe lo que parece ser una forma natural en la que los jugadores analizan tales juegos mientras el horizonte está relativamente cerca.

Existen buenas referencias para continuar revisando otros conceptos de teoría de juegos. Una de ellas, de la cual se tomaron las ideas de esta sección, es [Os04].

2.2. Lógica y teoría de juegos

Tal como mencionan Bonanno y Dégremont [BD13] en las últimas dos décadas la relación entre Lógica y Teoría de Juegos ha crecido de forma acelerada. Entre los más activos en el desarrollo de esta relación, se encuentran el mismo Giacomo Bonanno y Johan van Benthem. Al utilizar lógica modal encontraron formas de modelar juegos (para una descripción mas detallada, se puede consultar [BS14]).

2.2.1. Lógica modal

La lógica modal es el estudio de proposiciones modales y las relaciones lógicas que tienen entre ellas. Las proposiciones lógicas mejor conocidas son proposiciones acerca de aquello que es *necesario* y aquello que es *posible*. Los operadores de *posibilidad* y *necesidad*, se llaman operadores modales, ya que especifican una manera o un modo por el cual el resto de la proposición puede ser verdadera.

2.2.1.1. Lenguaje modal básico

Las fórmulas que conforman el lenguaje básico de la lógica modal $\mathcal{LM}$ se construyen a partir de un conjunto no vacío P numerable de proposiciones atómicas. Utilizando notación BNF, las fórmulas de $\mathcal{LM}$ se definen como:

$$\varphi ::= \top \mid P \mid \neg\varphi \mid (\varphi \vee \psi) \mid \Box\varphi$$

El lenguaje $\mathcal{LM}$ es un conjunto que se forma a partir de la proposición $\top$, las proposiciones atómicas P, es cerrado bajo la negación, la disyunción y el operador modal $\Box$ (llamado necesitación). Los conectivos lógicos $\wedge, \rightarrow, \leftrightarrow$ (conjunción, implicación y doble implicación respectivamente) y el operador modal $\Diamond$ (posibilidad), se definen como:

- $\bot \equiv \neg\top$
- $(\varphi \wedge \psi) \equiv \neg(\neg\varphi \vee \neg\psi)$
- $(\varphi \rightarrow \psi) \equiv (\neg\varphi \vee \psi)$
- $(\varphi \leftrightarrow \psi) \equiv (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$

$\bullet \Diamond\varphi \equiv \neg\Box\neg\varphi$

Los conectivos $\wedge, \rightarrow, \leftrightarrow$ tienen el significado habitual. Las fórmulas de tipo $\Diamond\varphi$ tienen el significado intuitivo “es posible que φ ” o “no es cierto que no φ sea necesario”.

2.2.1.2. Semántica

Un modelo estándar (estructura) $\mathcal{M}$ para un conjunto de fórmulas atómicas P, es una tripleta $\langle W, R, V\rangle$ que satisfaga las siguientes condiciones:

1. W es un conjunto no vacío, cuyos elementos son llamados *mundos*.

2. R es una relación binaria sobre W, $(R \subseteq W \times W)$, llamada *relación de accesibilidad*.

3. V es una función que asigna a cada elemento de P un subconjunto $V(p)$ de W, esto es $V : P \rightarrow 2^W$ (donde 2^W es el conjunto potencia de W).

Estos modelos usualmente se conocen como *modelos de Kripke*, en honor a S. Kripke quien los inventó y trabajó en lógica modal entre los años 1950s y 1960s. Intuitivamente, $\omega \in W$ indica un mundo posible y $\omega R\omega'$ quiere decir que ω' es accesible desde el mundo ω.

La relación de satisfactibilidad se obtiene recursivamente sobre la estructura de φ. Dado un modelo $\mathcal{M}$ y una evaluación e para cada variable proposicional en φ:

1. Las conectivas proposicionales se interpretan de la forma usual en algún mundo ω.

2. Para el operador $\Box$, $\omega \vDash \varphi \Leftrightarrow (\forall\omega') [\omega R\omega' \Rightarrow \omega' \vDash \varphi]$

2.2.1.3. Bisimulación

Definición Una bisimulación entre dos estructuras puntuales $(\mathcal{M}, \omega)$ y $(\mathcal{M}', \omega')$, con $\mathcal{M} = \langle W, R, V\rangle$ y $\mathcal{M}' = \langle W', R', V'\rangle$, es una relación binaria $Z \subseteq W \times W'$ tal que $\omega Z\omega'$ y para cualquier par de mundos $(x, x') \in W \times W'$ siempre que xZx' tenemos:

1. x, x' verifican las mismas variables proposicionales

2. si xRu está en $\mathcal{M}$ entonces existe $u' \in W'$ con $x'R'u'$ y uZu'

3. si $x'R'u'$ está en $\mathcal{M}'$ entonces existe $u \in W$ con xRu y uZu'

$\mathcal{M} = \langle W, R, V \rangle$ y $\mathcal{M}' = \langle W', R', V' \rangle$ son bisimilares ($\mathcal{M} \underline{\leftrightarrow} \mathcal{M}'$) si existen $\omega \in W$ y $\omega' \in W'$ tales que ($\mathcal{M}, \omega \underline{\leftrightarrow} \mathcal{M}', \omega$').

2.2.2. Lógica epistémica

La lógica epistémica permite razonar acerca del conocimiento. En situaciones de competencia resulta muy útil poder razonar acerca de los hechos y sobre lo que los otros agentes (posiblemente de *software*) conocen acerca de estos hechos. Para el proceso de razonamiento y toma de decisiones, un agente debe considerar tanto los hechos que conoce, como los hechos que conocen los demás.

La lógica epistémica, tal como se conoce actualmente, está muy influenciada por el desarrollo de la lógica modal; en particular, por su semántica, basada en el trabajo de Kripke. El sistema lógico $\mathcal{S}5$, es la lógica epistémica más popular y aceptada.

2.2.2.1. Sintaxis de lógica epistémica

Sea P un conjunto de proposiciones atómicas, y A un conjunto de agentes. El lenguaje lógico espistémico para multiagentes $\mathcal{L}_K$ es generado siguiendo las reglas, que se muestran a continuación en notación BNF:

$$\varphi ::= \top \mid p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_a\varphi$$

donde $p \in P$ y $a \in A$. Para cualquier agente a, $K_a\varphi$ se interpreta como “el agente a sabe φ ”. Además la noción intuitiva “el agente a considera posible φ” se escribe como $\hat{K}_a\varphi \equiv \langle a \rangle \varphi \equiv \neg K_a \neg\varphi$.

2.2.2.2. Semántica

Dado que cada agente tiene su modalidad, la lógica epistémica es una lógica multimodal. Algo crucial en la lógica epistémica es su semántica, que utiliza un caso especial de estructuras de Kripke.

Definición Sea P un conjunto contable de proposiciones atómicas y A un conjunto finito de agentes, un modelo de Kripke, es una estructura $\mathcal{M} = \langle S, \sim, V \rangle$, donde: (1) S es un conjunto de estados. Este conjunto es también conocido como el dominio $D(\mathcal{M})$, (2) $\sim$ es una función que asigna a cada agente $a \in A$ una

PL	Todas las instancias de las tautologías proposicionales
MP	Si $\vdash \varphi$ y $\vdash \varphi \to \psi$, entonces $\vdash \psi$
Nec	Si $\vdash \varphi$, entonces $\vdash K_a\varphi$
K	$\vdash K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$
T	$\vdash K_a\varphi \to \varphi$
4	$\vdash K_a\varphi \to K_aK_a\varphi$
5	$\vdash \neg K_a\varphi \to K_a\neg K_a\varphi$

Tabla 2.1: Sistema axiomático de $\mathcal{L}_K$ ($\mathcal{S}5$)

relación de accesibilidad $\sim_a \subseteq S \times S$, también se puede escribir como R_a y (3) $V : P \to 2^S$ es una función de evaluación que para cada $p \in P$ contiene el conjunto $V(p) \subseteq S$ de los estados en los cuales p es verdadero.

Definición: Satisfactibilidad y validez Las fórmulas epistémicas se interpretan en pares $(\mathcal{M}, s)$, donde $\mathcal{M}$ es un modelo de Kripke $\mathcal{M} = \langle S, \sim, V\rangle$ y un estado $s \in S$. algunas veces se refiere al par $(\mathcal{M}, s)$ como un estado. Si $\mathcal{M}$ es un modelo epistémico, el par $(\mathcal{M}, s)$ se llama estado epistémico. Este par también es llamado *modelo puntual*.

En la semántica de $\mathcal{L}_K$, las fórmulas *booleanas* se interpretan en la forma habitual. Para el caso modal, tenemos:

- $\mathcal{M}, s \models K_a\varphi$ sii para todo t tal que $(s,t) \in \sim_a$ y $\mathcal{M}, t \models \varphi$

La cláusula K_a es el operador de necesidad con respecto a $\sim_a$. El dual $\hat{K}_a$ es el operador de posibilidad con respecto a $\sim_a$.

Cuando $\mathcal{M}, s \models \varphi$ en todos los estados, se representa como $\mathcal{M} \models \varphi$ y se dice que φ es ***verdadera*** en $\mathcal{M}$. Si $\mathcal{M} \models \varphi$ para todos los modelos $\mathcal{M}$ de una determinada clase $\mathcal{C}$ (por ejemplo, todos los modelos epistémicos), se dice que φ es ***válida en*** $\mathcal{C}$ y se escribe como $\mathcal{C} \models \varphi$. Si $\mathcal{M} \models \varphi$ para todos los modelos de Kripke $\mathcal{M}$, se dice que φ es ***válida*** y se escribe como $\models \varphi$ o $\mathcal{K} \models \varphi$.

La lógica epistémica está completamente axiomatizada con el sistema de la tabla 2.1.

2.2.2.3. Noción de conocimiento común

Sea $K_a[s] = \{t \in S | s \sim_a t\}$. Para cualquier grupo no vacío de agentes $G \subseteq A$, escribimos $R_G^*[s] = \{t \in S | s \sim_G^* t\}$. Sea φ una fórmula de lógica epistémica. Si $R_G^*[s]$ es un subconjunto de los estados en los que φ es verdadera, entonces

C_G**FP**	$\vdash C_G\varphi \to \bigwedge_{a\in G} K_a\,(\varphi \wedge C_G\varphi)$
C_G**IR**	Si $\vdash \varphi \to \bigwedge_{a\in G} K_a\,(\varphi \wedge \psi)$, entonces $\vdash \varphi \to C_G\psi$

Tabla 2.2: Sistema axiomático de *MEL*

todos los agentes del grupo tienen acceso a φ. Intuitivamente se dice que φ es *conocimiento común* en el estado s. Se puede agregar una nueva fórmula $C_G\varphi$ para cada $G \subseteq A$, con la semántica:

$\diamondsuit$ $\mathcal{M}, s \models C_G\varphi$ sii para todo t tal que $(s,t) \in \sim_G^*$, tenemos $\mathcal{M}, t \models \varphi$

El lenguaje resultante se conoce como lógica epistémica multiagente (*MEL*). La lógica epistémica multiagente puede ser axiomatizada extendiendo $\mathcal{S}5$ con los axiomas de la tabla 2.2.

En [FHMV95] Fagin y sus coautores presentan una prueba de la completitud de *MEL*.

2.2.3. Topología y lógica modal

Parikh, Moss y Steinsvold [PMS07] analizan, entre otros temas, la conexión entre lógica modal y topología presentada inicialmente por Tarski y McKinsey [MT44], así como la interpretación epistemológica de la topología.

Muestran que la *tensión intuitiva* entre los dos paradigmas se desvanece cuando notamos que gran parte de la topología puede verse como una lógica modal, en particular como una lógica epistémica que combina las nociones de *conocimiento* y *esfuerzo*; y, aprovechando el hecho que la lógica modal es un fragmento decidible de la lógica de primer orden, se puede utilizar para resolver problemas topológicos que, en principio se encuentran en lógica de segundo orden.

Como ejemplo, suponen que se cuenta con cierta *medida*: una velocidad v que tiene cierto valor $s \pm e$. Esto puede interpretarse diciendo que v está en el conjunto abierto $(s-e, s+e)$, es decir, cualquier cosa que *sepamos* de v debe ser cierta *no solamente* en sí misma, sino también en cualquier v' dentro del intervalo $(s-e, s+e)$. Por tanto, se convierte en una relación de equivalencia para una lógica de conocimiento $\mathcal{S}5$ apropiada.

A su vez, la noción de *esfuerzo* también puede modelarse, dado que v puede medirse de manera más precisa con mayor esfuerzo. Con estas dos intuiciones, se desarrolló una lógica bimodal llamada *topologic* que está axiomatizada y es decidible.

2.3. Racionalidad acotada

La motivación principal para obtener modelos de racionalidad acotada se presenta por la insatisfacción que producen los modelos que resultan del paradigma del humano racional perfecto, el cual tiene una habilidad extraordinaria para optimizar: una vez que se proporcionan sus preferencias (bien definidas) siempre puede obtener y seleccionar el mejor resultado de entre todas las opciones posibles. La falta de satisfacción que produce este paradigma se debe a los resultados que se obtienen al comparar las hipótesis con las observaciones del comportamiento humano. Aunque estos resultados fueron presentados desde la década de los 50 por Herbert Simon [Si55], durante muchos años no se obtuvieron avances para desarrollar modelos alternativos para toma de decisiones. Este tema no fue retomado sino hasta la década de los 90 y desde entonces ha mostrado ser una opción viable para atacar este tipo de problemas.

Desde este punto de vista, el objetivo principal de la teoría de decisiones es obtener relaciones interesantes entre conceptos que surgen durante el razonamiento en situaciones de interacción. Con este enfoque se da importancia a examinar la plausibilidad de las hipótesis y no solo de las conclusiones. El énfasis en modelar los procesos no minimiza la importancia del objetivo, que es construir modelos que sean herramientas útiles para explicar fenómenos que no puedan explicarse de otra forma (idealmente, comparables con resultados observados).

2.3.1. El humano "racional"

En la teoría económica (clásica) [Ru98], un tomador de decisiones racional es un agente que debe elegir una opción después de realizar un proceso de deliberación (reflexión) en el cual responde a tres preguntas:

- ♢ ¿Qué es posible?
- ♢ ¿Qué es deseable?
- ♢ ¿Cuál es la mejor alternativa, de acuerdo con sus preferencias (deseos) y dadas ciertas restricciones en las opciones posibles?

En esta descripción se observa una suposición importante del modelo del humano racional: el procedimiento de determinar las alternativas viables es totalmente independiente del procedimiento para *depurar* sus preferencias. Es decir, si el agente califica una opción sobre la otra cuando observa un conjunto de opciones que incluya ambas, entonces las clasificará de igual forma cuando las encuentre en cualquier otro problema de decisión en el que ambas alternativas estén disponibles.

Definición: El humano "racional"

El elemento principal del proceso que realiza el humano racional, es una relación de preferencia $\succeq$ sobre un conjunto $\mathbf{A}$. Dado un problema de elección $A \subseteq \mathbf{A}$, el agente elige un elemento $x^* \in A$ el cual es $\succeq$ $-$óptimo (es decir, $x^* \succeq x$ para todo $x \in A$). Por simplicidad, se asume que la relación es asimétrica: $a \succeq b \Rightarrow b \not\succeq a$.

Algunos de los supuestos en el procedimiento que realiza el humano racional son los siguientes:

- ♢ *Conocimiento del problema*. El agente tiene claro todo el panorama del problema de elección que enfrenta: está completamente consciente del conjunto de alternativas entre las cuales debe elegir.

- ♢ *Preferencias claras*. El agente tiene un ordenamiento completo sobre todo el conjunto de alternativas.

- ♢ *Habilidad para optimizar*. El agente tiene la capacidad requerida para realizar cualquier cálculo necesario para encontrar el curso de acción óptimo. Su capacidad de cálculo es ilimitada y nunca comete errores.

Además, en ocasiones se requiere manejar situaciones en las que existe cierta incertidumbre entre la opción elegida y el resultado que obtendrá. Para poder realizar este manejo el modelo se enriquece con un *espacio de estados* $\mathbf{\Omega}$. Cada elemento de Ω representa una lista de factores exógenos que son relevantes para el agente, pero de los cuales no tiene control alguno. Ahora, un problema de decisión es un par (A, Ω) , donde $A \subseteq \mathbf{A}$ y $\Omega \in \mathbf{\Omega}$ es un conjunto de estados no excluidos por la información recibida por el agente.

2.3.2. La posición tradicional de la teoría de decisiones

La teoría económica usualmente se muestra permisiva sobre la suposición que los agentes se comportan como el humano racional y acepta que esa suposición es poco realista. Por esta razón desde hace algunos años se argumenta que el paradigma del humano racional debe ser tomado de forma *menos literal*.

El argumento tradicional *grosso modo* es el siguiente: en economía (en general teoría de decisiones), el interés principal es el comportamiento de los agentes y no el proceso que lo lleva a tomar cierta decisión. Aún si el agente no se comporta de la manera descrita por el procedimiento del humano racional, puede ser que su comportamiento pueda ser descrito como si siguiera dicho comportamiento: esto sería suficiente para nuestro propósito.

Una condición suficiente y necesaria para una función de elección inducida por un agente que se comporta como el humano racional es que debe satisfacer la *condición de consistencia* (también conocida como "independencia de alternativas irrelevantes"): para toda $A_1 \subseteq A_2 \subseteq \mathbf{A}$ si $C(A_2) \in A_1$ entonces $C(A_1) = C(A_2)$, donde $C(A)$ es la opción elegida (***C****hoice*) por el agente cuando enfrenta el problema de elección A. Esta condición indica que si el elemento elegido en un conjunto más grande está a su vez contenido en un conjunto más pequeño, entonces el agente elegirá este mismo elemento del conjunto más pequeño.

Las funciones de elección que satisfacen la condición de consistencia (aún cuando no resulten de un procedimiento del humano racional), pueden describirse como si se derivaran de algún humano racional. La importancia de este resultado depende de la existencia procedimientos plausibles que satisfagan la condición de consistencia, aún cuando no correspondan con la definición de humano racional. Un ejemplo clásico se conoce como *procedimiento satisfactorio*.

Definición: procedimiento satisfactorio

Los elementos básicos del procedimiento satisfactorio son un ordenamiento O del conjunto $\mathbf{A}$ y un conjunto $S \subseteq \mathbf{A}$ (así como una regla de desempate). Para cualquier problema de decisión A, secuencialmente examinan las alternativas de A de acuerdo con O, hasta encontrar un elemento que sea miembro de S (el conjunto de alternativas satisfactorias). Una vez encontrado dicho elemento, elegirlo y terminar. En el caso en que ningún elemento de A pertenezca a S, utilizar la regla de que satisface el requerimiento de consistencia (por ejemplo, elegir el último elemento de A).

Cualquier procedimiento acorde con esta definición satisface la condición de consistencia.

2.3.3. Racionalidad mínima

El concepto de racionalidad mínima implica que el agente no tiene habilidad perfecta para elegir las acciones apropiadas [Ch81]. Se presentarán condiciones de racionalidad relacionadas con conjuntos de creencias y las habilidades deductivas del agente.

2.3.3.1. Teoría de creencias

Las condiciones de racionalidad pueden ser muy débiles o muy fuertes para una teoría predictiva satisfactoria.

La teoría de creencias más rudimentaria es la *teoría de asentimiento*, esta teoría no requiere racionalidad:

◇ Un agente cree única y exclusivamente las declaraciones que puede afirmar.

Esta teoría tiene una atractiva simplicidad y parece indicar que el agente tiene autoridad total sobre sus creencias; sin embargo, tiene un defecto crucial: no impone ninguna restricción de racionalidad sobre el conjunto de creencias ni restricciones sobre el conjunto de creencias-deseos de agente. De acuerdo con esta teoría no se realizan inferencias a partir de las creencias, sin importar si son obvias o útiles; y el conjunto de creencias puede incluir todas las inconsistencias posibles. Una teoría que no incluye restricciones de racionalidad no tiene contenido predictivo; usarlo prácticamente no proveerá expectativas sobre el comportamiento del agente.

Al contrario de la permisividad de la teoría de asentimiento, las teorías predominantes incluyen condiciones que requieren racionalidad ideal del agente. En teoría de juegos y de decisiones, el principio de que el agente elegirá la acción que maximiza la utilidad esperada se clasifica dentro de este tipo. Existen serias dificultades al asumir racionalidad o información perfecta e idealizada.

En filosofía se utiliza (de forma implícita), una *condición de racionalidad ideal*, que puede formularse *grosso modo* como:

◇ Si un agente tiene un conjunto de creencias-deseos particular, tomará *todas* las acciones que son aparentemente apropiadas.

Una simplificación conveniente: se dice que una acción aparentemente es apropiada si y solo si, de acuerdo con sus creencias, estaría orientada a satisfacer sus deseos.

Esta idealización tiene cierto valor, pero es demasiado restrictiva. Por ejemplo, excluye situaciones en las que los agentes olvidan o no tienen cuidado al elegir sus acciones. Además, esta condición de racionalidad requiere de una habilidad deductiva idealizada.

La falta de satisfacción de esta condición de racionalidad idealizada, surge de la negación de una característica fundamental humana: que se encuentra en la situación finita de tener un límite en su capacidad cognoscitiva y en el tiempo disponible.

La importancia de idealizar siempre es relativa a los objetivos planteados; la condición de racionalidad ideal puede ser útil en ciertas condiciones. Una motivación para idealizar una teoría puede ser obtener una teoría más manejable de lo que sería una teoría enteramente correcta o completa. Además se pueden definir diferentes grados de idealización; con esto, simplicidad y manejabilidad pueden combinarse para tener más campo de aplicación.

2.3.3.2. Racionalidad, inferencia y conveniencia mínimas

La teoría de creencias basada en condiciones de racionalidad mínima es un intento para obtener mayor aplicabilidad a cambio de una teoría más compleja. Por el momento, esta teoría seguirá siendo idealizada respecto a la concepción de inferencia.

Las condiciones mostradas a continuación son solamente condiciones necesarias para tener creencias y deseos. Las condiciones siguientes pueden derivarse de la *condición general de racionalidad mínima*:

- ◇ Si un agente tiene un conjunto particular de creencias-deseos, entonces intentaría realizar sólo algunas de las acciones que son aparentemente apropiadas.

El argumento por el cual esta condición en efecto es utilizada (y debería ser utilizada) por agentes con creencias, se compone de tres partes:

1. Ningún agente con recursos finitos puede satisfacer la condición de racionalidad ideal.

2. Se puede predecir el comportamiento a partir de las creencias y los deseos del agente y, tanto la condición de racionalidad ideal como la nula excluyen esta predicción.

3. Si a un agente no se le solicita que intente tomar alguna de las acciones aparentemente apropiadas, entonces no se podría obtener ninguna predicción sobre su comportamiento.

Adicionalmente, para tener una teoría predictiva de creencias, existe una condición de racionalidad general más fuerte sobre el conjunto de creencias-deseos: el agente no debe intentar utilizar alguna de las acciones que son inapropiadas.

Por otro lado, la condición general de racionalidad mínima implica que el agente debe tener una habilidad deductiva mínima (y posiblemente también una habilidad inductiva mínima). La *condición de inferencia deductiva mínima* es:

- ◇ Si un agente tiene un conjunto particular de creencias-deseos, entonces realizaría algunas, pero no necesariamente todas las inferencias sensatas a partir de su conjunto de creencias que sean aparentemente apropiadas.

Si el agente realiza una acción apropiada de acuerdo con sus creencias y deseos es porque llegó a la conclusión de que esa acción es deseable. La condición de inferencia mínima requiere que el agente realice algunas de las inferencias racionales que sean aparentemente útiles. Si no satisface al menos esta condición mínima, entonces no será capaz de reconocer y tomar acciones que son apropiadas dadas sus creencias.

Esta condición de habilidad lógica mínima puede describirse en términos de un *requerimiento de conveniencia mínima*:

- ♢ El agente llevaría a cabo algunas de las inferencias sensatas de su conjunto de creencias que serían aparentemente apropiadas realizar.

Es decir, las inferencias resultantes ayudarían al agente a elegir las acciones que tenderían a satisfacer sus deseos, de acuerdo con sus propias creencias. Para evitar que un agente solo obtenga conjunciones a partir de una de sus creencias, es necesario un *requerimiento de consecuencias mínimas*:

- ♢ El agente debe tener éxito en realizar algunas de las inferencias racionales aparentemente apropiadas que ha tomado.

Nótese que al definir requerimientos de conveniencia y consecuencia por separado no hay implicación de que dos procesos distintos siempre se lleven a cabo cuando se realiza una inferencia.

2.3.3.3. Consistencia mínima

La habilidad deductiva requerida para satisfacer la condición general de racionalidad mínima debe incluir, además de la habilidad de realizar inferencias útiles, la habilidad de eliminar inconsistencias en el conjunto de creencias. El conjunto de creencias está sujeto a la condición de consistencia mínima:

- ♢ Si un agente tiene un conjunto particular de creencias-deseos y si algunas inconsistencias (pero no necesariamente todas) se generan en el conjunto, entonces el agente las eliminará.

Esta condición aplica tanto para inconsistencias explícitas como $\{p, \neg p\}$, como para inconsistencias implícitas como $\{p, p \rightarrow q, \neg q\}$.

Un agente no puede permitir todas las inconsistencias que aparecen en su conjunto de creencias, pero no se requiere que elimine todas las inconsistencias que pueden generarse en dicho conjunto. Si en el sistema de creencias de un agente no se restringe a alguna de las restricciones de consistencia presentadas y, por tanto pudiera contener un número ilimitado de inconsistencias, entonces este sistema no tendría valor alguno para predecir el comportamiento del agente.

2.3.3.4. Condiciones normativas

Finalmente, se identifica qué es la racionalidad mínima; se muestra que el conjunto de inferencias requeridas por la condición de inferencia mínima solamente es un subconjunto propio del conjunto de inferencias que un agente debe realizar para ser *pragmáticamente racional*.

- ◇ La tesis *descriptiva* es la condición de inferencia mínima: un agente debe realizar algunas de las inferencias a partir de sus creencias que tenderían a satisfacer sus deseos.
- ◇ La tesis *normativa* es que el agente debe hacer todas (y solo) las inferencias sensatas factibles a partir de sus creencias que tenderían a satisfacer sus deseos.

Entonces, la noción de racionalidad pragmática se explica de la siguiente manera: "si p implica q y el agente cree que p, entonces *debe* creer q".

Solo un subconjunto *pequeño* de las inferencias sensatas que sería prácticamente posible para ser realizado por el agente sería útil para él en un momento dado. Una inferencia puede ser sensata, pero puede no ser razonable llevarla a cabo, porque no es de valor previsible al momento y evita que el agente lleve a cabo otras acciones que son obviamente valiosas al momento, dados sus recursos cognoscitivos limitados. Para determinar si un agente debe realizar la inferencia de q a partir de p, debe tener en cuenta:

1. la sensatez de la inferencia,
2. la factibilidad de realizarla, y
3. la utilidad aparente, de acuerdo con sus creencias y deseos.

La tabla 2.3 muestra el esquema general para los conceptos de racionalidad descritos. El punto importante de las condiciones de racionalidad mínima es que al parecer son indispensables para cualquier teoría de creencias que resulte satisfactoria.

2.3.4. Omnisciencia lógica como problema de complejidad computacional

La omnisciencia lógica asume que un agente epistémico conoce todas las consecuencias lógicas de sus supuestos. En [AK09] Artemov y Kuznets presentan un

Todas las inferencias del conjunto de creencias
⋮
Todas las creencias sensatas (condición de cerradura deductiva)
⋮
Todas las inferencias aparentemente deseables (condición de inferencia ideal)
⋮
Todas las inferencias factibles (condición de inferencia normativa)
⋮
Inferencias requeridas para creencias (condición de inferencia mínima)
⋮
Sin inferencias (teoría de asentimiento)

Tabla 2.3: Esquema general de racionalidad

marco general que analiza la omnisciencia lógica desde el punto de vista de la complejidad computacional.

Se plantean las siguientes definiciones:

1. *Test* de Omnisciencia Lógica (LOT): un sistema epistémico E *no es lógicamente omnisciente* si para cualquier afirmación de conocimiento $\mathcal{A}$ del tipo *se sabe que* F, existe una prueba de F en E, cuya complejidad está acotada por algún polinomio del tamaño de $\mathcal{A}$.

2. *Test* de Omnisciencia Lógica Fuerte ($SLOT$): un sistema epistémico E *no es lógicamente omnisciente* si para cualquier afirmación de conocimiento $\mathcal{A}$ del tipo *se sabe que* F, una demostración de F en E puede ser encontrada en tiempo polinomial en el tamaño de $\mathcal{A}$.

Tanto LOT como $SLOT$ conectan el tamaño de una afirmación sobre el conocimiento de F con la habilidad del sistema para proveer de forma factible una evidencia adecuada de F. En LOT, la medida de factibilidad es la longitud de la prueba; mientras que en $SLOT$, es el tiempo requerido para obtener dicha prueba.

Además, muestran que las lógicas modales de conocimiento y creencia más utilizadas son lógicamente omniscientes con respecto a LOT y $SLOT$; mientras que sistemas lógicos de justificación, los cuales contienen evidencia de afirmaciones del tipo *t es una justificación de F*, no son lógicamente omniscientes. Los autores argumentan que estos resultados son acordes con la intuición.

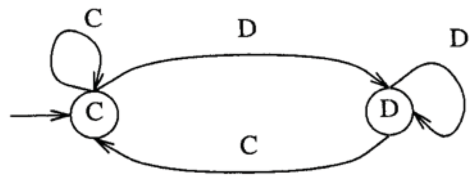

Figura 2.1: Equilibrio *tit-for-tat*

2.3.5. Memoria limitada

Desde el punto de vista computacional, una cota natural para la racionalidad es la memoria: ¿Cuánta información puede almacenarse? En [PYa94], se presenta un modelo de memoria limitada que utiliza un autómata en el que el número de estados representa la cantidad de memoria disponible. En este modelo, los estados representan las acciones del agente, mientras las transiciones corresponden con las acciones del oponente. Con este modelo, los autores analizaron y obtuvieron resultados interesantes utilizando el juego del dilema del prisionero.

Este juego ha sido ampliamente estudiado y se sabe que el equilibrio de Nash es que ambos prisioneros confiesen. Esta misma situación es también el único equilibrio si el juego se repite n veces. La prueba es por inducción hacia atrás: en la última ronda, el mejor movimiento (para ambos) es confesar; y por tanto no existe razón por la cual coopere con el otro prisionero en la penúltima ronda y así sucesivamente. Este resultado parece aún *menos racional* que el de una sola ronda. Nótese que existen dos versiones del dilema en los llevan a jugar de forma cooperativa, el juego repetido no acotado y el juego acotado por una constante que no se conoce.

2.3.5.1. Autómata con menos de n estados

Limitando a los agentes a estrategias que pueden modelarse usando autómatas finitos de tamaños dependientes del número de rondas es posible establecer equilibrios más satisfactorios. Para el caso del dilema del prisionero jugado n veces, si las estrategias de ambos agentes está limitada a menos de n estados, ninguno puede contar hasta n y por tanto, no pueden aplicar inducción hacia atrás. En este caso, una estrategia razonable básica conocida como *tit-for-tat* (ojo por ojo) es un equilibrio: una estrategia de siempre cooperar (C) a menos que el otro confiese puede jugarse con el autómata mostrado en la figura 2.1, si el otro jugador confiesa (D), entonces confiesa una ronda como castigo.

2.3.5.2. Autómata con un número de estados subexponencial

El caso más interesante es cuando el autómata tiene más de n estados, pero al menos uno de los jugadores tiene un número de estados menor que exponencial en el número de rondas.

Teorema [PYa94]: para cualquier $\epsilon > 0$, en la n-ésima ronda jugada por autómatas con tamaños acotados; si al menos una de las cotas es $< 2^{c_\epsilon n}(c_\epsilon = \epsilon/6(1+\epsilon))$, entonces existe un equilibrio de estrategia mixta con ganancia promedio para cada jugador de al menos $3 - \epsilon$.

En otras palabras, es posible construir un equilibrio de estrategia mixta tal que ambos jugadores obtendrán ganancias arbitrariamente cercanas a la ganancia que obtendrían si cooperan en todas las rondas. Este equilibrio estratégico incluye varios componentes de ciclos. En las primeras d rondas, los jugadores muestran una serie de movimientos (de acuerdo a su estrategia); d depende del tamaño de la cota de los jugadores. Posteriormente, un agente debe recordar los movimientos de su oponente, por tanto debe almacenar los 2^d posibles movimientos. Después de este intercambio, los agentes juegan de acuerdo con ciertas reglas que equilibran cualquier ventaja que alguno pudiera obtener sobre el otro en las primeras d rondas. Después ambos caen en un ciclo de dos estados: el primero es colaboración por un gran número de rondas (determinado por su tamaño); el segundo es una serie de acciones coordinadas que dependen de los primeros movimientos mostrados. Esta etapa tiene dos propósitos: equilibrar movimientos ventajosos y mostrar que los agentes aún pueden recordar los movimientos del oponente. Si en algún momento un jugador realiza un movimiento inesperado, entonces el oponente entrará en un estado de confesión perpetua. Esta construcción es tal que los jugadores deben utilizar todos los estados para jugar su estrategia y, por tanto no tiene estados para contar y aplicar inducción hacia atrás. De hecho, aún si uno de los jugadores tuviera un número exponencial de estados y pudiera contar, la única diferencia sería que confiesa en la última ronda (dado que hacerlo antes no incrementa su ganancia).

2.3.6. Razonamiento de previsión limitada

Un ejemplo que motiva el estudio de razonamiento de predicción limitada es el juego de ajedrez. Cada juego consiste en una serie de movimientos que termina en uno de tres resultados (gana blanco, gana negro o empate). Sin embargo, este valor no se puede determinar desde el principio, ya que esto requeriría enumerar todas las historias posibles (más de 10^{40}). Aún más, debido a esta dificultad, un jugador no puede seleccionar el movimiento que le entregue la mejor ganancia posible. Sin embargo, los jugadores aún son capaces de jugar de manera sensata.

Los jugadores de ajedrez emplean diversas estrategias para determinar nuevos movimientos; una muy común es seleccionar movimientos candidato razonables (posiblemente basados en experiencia y conocimiento del juego) y después tratar de predecir como podría evolucionar el juego.

Un par de modelos sencillos de este comportamiento de *predicción* se presentan en [Ru98]. Ambos modelos asumen que el jugador conoce tanto su equilibrio como el de su oponente durante los siguientes k movimientos; el objetivo del agente es maximizar la ganancia obtenida durante esos k movimientos. En un modelo, el agente puede cambiar solamente su movimiento actual, pero no los futuros; en el segundo, se asume que las acciones del oponente son fijas pero el agente puede optimizar cambiando los siguientes k movimientos. Ambos modelos producen definiciones razonables de equilibrio; sin embargo, intuitivamente no son satisfactorias. El primero por que un agente debe tomar su futuro como *determinado* aún cuando el agente podría influenciarlo. En ambos, los movimientos del oponente deben ser tratados como dados, aunque es probable que la estrategia del oponente cambie de acuerdo a la evolución del juego.

El valor principal de estos modelos es que al no producir descripciones satisfactorias de razonamiento de predicción limitada presentan el reto de construir mejores modelos.

2.3.7. Optimización acotada

Russell y Subramanian definieron un modelo que parece apropiado para determinar si un sistema de inteligencia artificial (AI) exhibe comportamiento racional [?]. Presentan cuatro posibles definiciones de racionalidad; las primeras tres son utilizadas actualmente en sistemas de AI y la cuarta es su propuesta de definición.

1. Racionalidad perfecta: el agente racional *siempre* actúa de modo que maximiza su ganancia.

2. Racionalidad de cálculo: el agente racional *tarde o temprano* actúa de tal modo que maximiza su ganancia.

3. Racionalidad en metanivel: el agente racional *calcula la mejor serie de acciones*, teniendo en cuenta que el cálculo debe seleccionar la acción.

4. Racionalidad con optimización acotada: el agente especifica programas óptimos. Un agente es un agente racional *si lo hace tan bien como le sea posible*, dados sus recursos computacionales.

La propiedad deseada de sistemas inteligentes tradicionalmente corresponde con la definición 1, pero en general no es factible computacionalmente. La definición 2 es interesante dado que muestra que en principio tomaría la acción correcta, pero que generalmente en la práctica no es útil debido a que podría tomar tiempo exponencial para determinar una solución. Por otro lado, la definición 3 frecuentemente es más difícil que el problema original. Sin embargo, un enfoque común en AI es diseñar sistemas de racionalidad de cálculo y después utilizar técnicas de aceleración con aproximaciones para acercarse al valor óptimo.

Una función de agente es una relación entre sus percepciones y sus acciones. Un programa de agente es aquel que genera sus acciones. La función de agente definida por Russell y Subramanian es tal que la acción puede ser *nula* si el agente aún está realizando algún cálculo. Un agente óptimo acotado se define como un agente que maximiza su utilidad sobre todas las funciones factibles, dadas las restricciones temporales y una clase determinada de máquinas. Su noción se extiende a agentes óptimos asintóticamente, utilizando la misma noción de asintótico que se presenta en complejidad computacional.

Este artículo formaliza algunas ideas sobre qué es racional cuando las restricciones no permiten tener racionalidad perfecta. Su modelo no es útil para describir racionalidad humana, dado que en general se requiere realizar una gran cantidad de cálculos, pero presenta una forma plausible de definir sistemas de AI que son racionales de forma acotada.

2.3.8. Teoría computacional de conciencia y toma de decisiones

En [DF09], Nikhil y Lance presentan una definición de conciencia desde el punto de vista computacional. En general, el nivel de conciencia que tenemos sobre un objeto es la cantidad de tiempo necesario para generar ese objeto dentro de cierto ambiente. La intuición es que los objetos de los cuales tenemos mayor conciencia serán los primeros en ser generados.

Además, muestran su relación con toma de decisiones y cómo otros agentes pueden manipular las decisiones utilizando la *publicidad* apropiada.

2.3.8.1. El modelo

El modelo que presentan se compone de:

- ♢ Un alfabeto fijo $\Sigma = \{0, 1\}$, generalizable a alfabetos más grandes.

- ♢ El ambiente es una función $E : \Sigma^* \to \Sigma^*$.

- ♢ El contexto y es una cadena en Σ^*.

- ♢ Un proceso de enumeración es una máquina de Turing con oráculo M, tal que dado un oráculo A y una entrada w, $M^A(w)$ producirá una serie posiblemente infinita de cadenas $z_1, z_2, \ldots$ El oráculo A puede verse como una función cuyo resultado aparecerá en una cinta especial. Cada consulta al oráculo se contará como un solo paso temporal, de modo que tomará más tiempo escribir cada bit de la consulta y alcanzar cada bit de la respuesta.

- ♢ La falta de conciencia de la cadena x en el ambiente E con el contexto y usando el proceso de enumeración M. $U_M^E(x|y)$ se define como la cantidad de tiempo necesario para que $M^E(y)$ enumere a x. Es importante notar que se cuenta el tiempo, no el índice en el cual la cadena es enumerada. Además, no se requiere que $M^E(y)$ se detenga después de producir la cadena x en su lista de salida. Si $M^E(y)$ nunca enumera a x, se dice que la falta de conciencia es infinita.

Esta definición no incluye ninguna acción ni decisión hecha por ningún agente. Un posible enfoque es combinar conciencia con una *función de conveniencia*. Algunas funciones candidato son:

- ♢ $f : \Sigma^* \to \{0, 1\}$, que mide la “conveniencia” de una cadena. Esta función modela un proceso de decisión binario, en el que la cadena es conveniente o no, el objetivo es encontrar alguna cadena que lo sea.

- ♢ $f : \Sigma^* \times \Omega \to \{0, 1\}$, donde Ω es un espacio de probabilidad. f se interpreta como una variable aleatoria que mide la conveniencia. Un caso especial se presenta cuando la conveniencia de las cadenas es independiente de Ω, entonces, se tiene $f : \Sigma^* \to [0, 1]$, que mide la probabilidad de que una cadena satisfaga los criterios. Este tipo de funciones aleatorias permiten modelar múltiples agentes con un solo proceso.

- ♢ $f : \Sigma^* \to \mathbb{R}$. En este caso, la conveniencia es una medida continua y un agente busca una cadena que la maximice.

Para obtener un proceso de toma de decisiones se puede combinar la función de conveniencia con un procedimiento de enumeración; por ejemplo, para la primera función candidato, elegir un objeto x si es el primero listado por $M^E(y)$ tal que

$f(x) = 1$. Además, se puede incorporar de forma explícita el costo asociado al cálculo realizado por el proceso de toma de decisión.

Por otro lado, se puede manipular el proceso de toma de decisiones manipulando el ambiente E: cambiando la salida de la función E para algunas entradas (por ejemplo, si se manipulan los resultados de motores de búsqueda). Sea A algún conjunto de acciones (subconjunto de todas las acciones posibles) y sea $c(A)$ el costo de cambiar todas las entradas de E correspondientes a A. Es decir, el costo de hacer publicidad a las acciones de A es $c(A)$. Además un anunciante w recibe una ganancia $p_w(x)$ si x es la decisión tomada por el agente. Por tanto, es rentable para un anunciante hacer publicidad a las acciones de A siempre que el resultado de modificar la decisión del agente devuelva ganancia mayor que $c(A)$.

2.3.8.2. Aplicaciones

Una vez definido el modelo, los autores [DF09] describen algunas situaciones en las que la noción computacional de conciencia surge de forma natural, en escenarios en los que se organiza, se accede y se transfiere información. Estas descripciones se basan en un conjunto de preguntas que son contestadas en cada situación.

- ◇ ¿Conciencia de qué? Generalmente resulta útil definir un conjunto particular de cadenas de interés, dado el contexto.

- ◇ ¿Cuál es el ambiente? Puede darse una definición contextual a un conjunto particular de cadenas de interés. Además, resulta útil definir las reglas de cómo puede modificarse dicho ambiente.

- ◇ ¿Cuál es el proceso de enumeración? Usualmente este proceso se modela de forma estadística sin considerar la “sensatez” algorítmica del proceso. Buscar un modelo algorítmico diferente a los tradicionales modelos en estadística puede mostrar nuevas propiedades interesantes.

- ◇ ¿Cuál es el proceso de toma de decisión? Esto incluye la forma de tomar decisiones y la definición de la conveniencia de las cadenas; además, también puede incluir el costo de su cálculo.

2.4. Optimización: programación dinámica

2.4.1. Optimización en tiempo discreto

Considérese un problema de múltiples decisiones, en el que varias decisiones relacionadas se eligen de forma secuencial. Elegir la última decisión de manera óptima usualmente es sencillo: tomar la mejor del conjunto de acciones posibles en ese punto. Por otra parte, si conocemos la solución para un problema de múltiples decisiones en el que faltan r elecciones, entonces esta solución puede usarse para resolver el problema de $r+1$ decisiones por tomar. Esto se debe al *principio de optimalidad* [BT05]:

- ♢ Dada una secuencia óptima de decisiones, el final de cada secuencia también es óptimo para el problema correspondiente.

Este método de solución se llama *programación dinámica*. Fue desarrollado por Richard Bellman en los 1950s. La versión de tiempo continuo puede rastrearse hasta los trabajos de Hamilton y Jacobi, e incluso hasta Huygens y su caracterización como onda para la luz.

2.4.1.1. Ecuación de Hamilton–Jacobi–Bellman

Ahora se presenta formalmente el principio de optimalidad, llamado ecuación de Hamilton–Jacobi–Bellman o en breve ecuación de Bellman.

Sean los conjuntos U y X, $\mathcal{T} \subset X$, $a \in X$, y las funciones $(x_k, u_k) \to f_k(x_k, u_k)$, $(x_k, u_k) \to g_k(x_k, u_k)$, (*mapeos* de valores en X) tales que $x_k \in X$, $u_k \in U$, $k = 0, \ldots n$.

Definición El siguiente problema se conoce como *un problema de programación dinámica determinístico y de tiempo discreto.*

$$F(\overline{x}, \overline{u}) = \sum_{k=0}^{n} f_k(x_k, u_k) \to \min,$$

$$\overline{x} = (x_k)_{0 \leq k \leq n+1}, \overline{u} = (u_k)_{0 \leq k \leq n+1},$$

$$x_k \in X, u_k \in U, k = 0, \ldots, n,$$

$$x_{k+1} = g_k(x_k, u_k), k = 0, \ldots, n,$$

$$x_0 = a, x_{n+1} \in \mathcal{T}$$

Considérese la familia de problemas $P_{r,x}$ con $x \in X$ y $r \in \{0, 1, \ldots, n\}$: r decisiones a tomar, con estado inicial a.

$$F_{r,x}(\overline{x}, \overline{u}) = \sum_{k=n-r}^{n} f_k(x_k, u_k) \to \min,$$

$$\overline{x} = (x_k)_{n-r \leq k \leq n+1}, \overline{u} = (u_k)_{n-r \leq k \leq n},$$

$$x_k \in X, k = n-r, \ldots, n+1, u_k \in U, k = n-r, \ldots, n,$$

$$x_{k+1} = g_k(x_k, u_k), k = n-r, \ldots n,$$

$$x_{n-r} = x, x_{n+1} \in \mathcal{T}$$

Se escribe $J_r(x)$ para el valor del problema $P_{r,x}$, donde $x \in X$ y $r = 0, \ldots, n+1$. El hecho siguiente es la formalización del principio de optimalidad.

Teorema ***Ecuación de Hamilton–Jacobi–Bellman, condiciones necesarias y suficientes***. La función J puede caracterizarse recursivamente con las siguientes propiedades [BT05]:

$$J_{r+1}(x) = \inf_{u \in U} (f_{n-r}(x, u) + J_r(g_{n-r}(x, u)))$$

para $r = 0, \ldots, n, x \in X$ y

$$J_0(x) = 0 \text{ para } x \in \mathcal{T},\ J_0(x) = \infty \text{ en otro caso.}$$

2.4.2. Optimización en tiempo continuo

El método de programación dinámica, la ecuación de HJB, en tiempo continuo es simple intuitiva y elegante. Tiene muchas aplicaciones de interés y puede extenderse a problemas estocásticos.

Considérese el problema de minimizar costo. Si iniciamos en tiempo t_0 con cierto costo x_0, denotado como $V(t_0, x_o)$, y el costo en algún tiempo t como $x(t)$. Entonces, la función $t \mapsto V(t, x(t))$ es monotónica no decreciente y constante por tramos.

Ahora, tomando en cuenta que la función $t \mapsto V(t, x(t))$ es constante por tramos, entonces sabemos que su derivada

$$\frac{dV}{dt}(t, x(t)) = \frac{\partial V}{\partial t}(t, x(t)) + \frac{\partial V}{\partial x}(t, x(t))\, \dot{x}(t)$$

es cero en dichos tramos [BT05]. Este hecho es la esencia de la ecuación de Hamilton–Jacobi–Bellman en tiempo continuo.

Capítulo 3

Racionalidad acotada y teoría de juegos

Resumen

Este capítulo contiene el primero de los enfoques tomados en la búsqueda de responder a la pregunta guía. Comienza con el cálculo de posibles reparticiones iniciales para el juego del dominó. Es importante notar que el número de reparticiones resulta extremadamente alto para suponer que se puede disponer de todas las historias posibles, además este total sólo incluye las posibles reparticiones iniciales 3.1. Después, se presenta un modelo para la clase de juegos dinámicos de movimientos alternados. Este modelo es una extensión del trabajo de Jehiel [Je95] para juegos alternados repetidos 3.2. Posteriormente se contrasta el modelo obtenido con un trabajo similar desarrollado prácticamente al mismo tiempo por Mirrokni y su equipo [Mi12] esta comparación se encuentra en la sección 3.2.2. Además, se presentan ejemplos de aplicación del modelo extendido 3.2.1.2 y 3.2.2.1.

3.1. Motivación

La suposición de racionalidad perfecta, que, aplicada a la teoría de juegos, puede traducirse como la posibilidad de generar *todo* el árbol asociado al juego, es poco realista. Dependiendo de las limitaciones de recursos que tenga un agente, pueden existir diferentes estrategias que se consideren racionales. El estudio y análisis de las limitaciones permitirá construir estrategias racionales bajo determinadas condiciones. La mayoría de los trabajos relacionados a racionalidad acotada y estrategias, se han desarrollado para el equilibrio de Nash. Dada la naturaleza de la clase de juegos dinámicos alternados, se requiere un modelo de racionalidad acotada que tome en cuenta equilibrio perfecto por subjuegos.

La razón principal para utilizar este punto de vista es tratar de determinar equilibrios y estrategias de forma realista. Como ejemplo, utilizamos el juego de dominó. Para determinar una aproximación del número total de posibles historias de una partida de dominó, se utiliza el número total de reparticiones posibles que se calcula como:

$$\begin{aligned} r \;=\; & \binom{ntf}{f_{j_1}} \times \binom{ntf - f_{j_1}}{f_{j_2}} \times \binom{ntf - f_{j_1} - f_{j_2}}{f_{j_3}} \\ & \times \binom{ntf - f_{j_1} - f_{j_2} - f_{j_3}}{f_{j_4}} \end{aligned}$$

donde:

- ♢ ntf es el número total de fichas a repartir y
- ♢ f_{j_i} es el número de fichas que recibe el jugador i al inicio de la partida.

Como todos los jugadores reciben la misma cantidad de fichas, la expresión puede reducirse a:

$$r = \prod_{i=1}^{n} \binom{ntf - \sum_{k=0}^{i-1} f_{j_k}}{f_{j_i}}$$

Esta última resulta en la forma más útil para nuestro análisis. Para el caso de un juego de 28 fichas y cuatro jugadores, tendremos:

$$r = \binom{28}{7} \times \binom{21}{7} \times \binom{14}{7} \times \binom{7}{7} = 472,518,347,558,400$$

Por otro lado, en la variante de dominó cubano, se utilizan 55 fichas (doble nueve). Para este caso, las reparticiones (distribuyendo el mayor número de fichas disponibles a cada jugador) posibles son:

$$r = \binom{55}{13} \times \binom{42}{13} \times \binom{29}{13} \times \binom{16}{13}$$

$$r \approx 10^{33}$$

Si se reparten siete fichas a cada jugador en esta variante, obtenemos:

$$r \approx 10^{30}$$

Por tanto, el número de historias posibles para un juego de 28 fichas y cuatro jugadores es aproximadamente 4×10^{14}, pero si se toma la versión cubana con 55 fichas y cuatro jugadores, el número de historias posibles crece al orden de 10^{30}.

Inicialmente se tomarán los enfoques de conciencia (*awareness*) restringida y toma de decisiones [DF09] así como el razonamiento de previsión limitada [Ru98, Je95] con optimización acotada [RS93] en la capacidad del cómputo, para modelar la racionalidad en el juego de dominó.

3.2. Desarrollo del modelo

La presente sección está basada en el trabajo [EH13]. Como primer paso se responden las preguntas planteadas en [DF09]; para el caso particular del dominó:

- ¿*Conciencia*[1] de qué? El agente debe estar al tanto (*consciente*) de las acciones realizadas hasta el momento actual, así como de las fichas que posee. Además, debe estar consciente de las acciones que le son posibles cuando es su turno de tirar.

- ¿Cuál es el ambiente? El entorno se compone del estado actual del juego: el número de fichas que tiene cada jugador y el tren del juego sobre la mesa. Una consulta al entorno es tratar de reconstruir el historial del juego con ayuda del turno actual, siguiendo el tren de juego.

- ¿Cuál es el proceso de enumeración? La obtención de las posibles acciones se realiza recorriendo el árbol de decisiones del juego mientras la partida evoluciona. Este recorrido puede devolver un conjunto o un estado particular. Sin embargo, construir el árbol completo requiere espacio exponencial en el número de fichas totales disponibles.

- ¿Cuál es el proceso de toma de decisión? Ya sea del estado o del conjunto obtenido en la enumeración, se puede seleccionar la mejor acción posible de entre sus aristas de salida.

Una vez hecho lo anterior, el siguiente paso es definir un modelo del juego considerando las limitaciones de previsión mencionadas en [Ru98, Je95] en función de algún tipo de optimización acotada [RS93].

[1]En este contexto, *conciencia* hace referencia a la intuición de tener presente, un agente tiene mayor *conciencia* de un objeto si lo tiene más presente que otro.

En [Je95], Philippe Jehiel presenta un modelo de previsión con horizonte limitado, aplicado a *juegos repetidos de movimientos alternados*. Las características principales de esta clase de juegos son que existen dos jugadores que mueven secuencialmente en pasos discretos de tiempo. En cada periodo t, la ganancia actual del jugador i está en función de la acción que realiza el jugador en ese tiempo, y la acción realizada por el otro jugador en el tiempo anterior. Los espacios de acciones son finitos y no cambian durante el juego.

Jehiel asume que cada jugador tiene *habilidad limitada para prever el futuro*. El jugador i es caracterizado por la longitud de su previsión n_i (un valor constante). En cada periodo t, el jugador i formula predicciones para los siguientes n_i movimientos después de su turno. Por tanto, debe elegir su acción en el turno actual solamente de acuerdo a su previsión limitada. Esto se debe a que:

1. El jugador i no puede construir sobre lo que sucederá después de n_i pasos, ya que no puede predecirlo (no tiene idea o *awareness* de ello) y,

2. Dado que este tipo de juegos son estacionarios, la ganancia promedio puede considerarse como una buena aproximación de la ganancia final real.

También se define un concepto de solución llamado (n_1, n_2)–*solución* que requiere dos nociones preliminares:

1. Una estrategia para el jugador i es justificada por una secuencia de previsiones si la estrategia maximiza la ganancia promedio obtenida en la longitud de su previsión n_i. y,

2. Una secuencia de previsiones para el jugador i es consistente con un perfil de estrategia si la previsión coincide con el truncamiento a las primeras n_i acciones de la ruta respectiva inducida por el perfil de estrategia.

Una (n_1, n_2)–*solución* se define como un perfil de estrategia que puede ser justificado por secuencias consistentes de previsiones para los jugadores 1 y 2. En otras palabras, en una (n_1, n_2)–*solución*:

1. Las acciones son elegidas para maximizar la ganancia promedio de la longitud de su previsión y,

2. En cualquier periodo en el que el jugador i debe mover, las previsiones de las siguientes n_i acciones en función de la acción actual son correctas.

Es importante notar que las predicciones de las siguientes n_i acciones del agente i incluyen sus propias acciones y que las previsiones de equilibrio acerca de todas estas acciones se asumen como correctas sin importar su acción actual y no solo para la acción en la ruta de equilibrio.

Otra cuestión importante señalada por Jehiel es el hecho de que no se presenta mejora para el jugador en caso de incrementar su previsión, dado que es cíclico.

Para el caso de *juegos dinámicos*, como el dominó, se tienen diferencias sustanciales:

- ♢ Si bien los movimientos son alternados, el juego es dinámico, en el sentido de que el espacio de acciones se actualiza en cada tirada y, en general se reduce.

- ♢ De acuerdo con la clasificación presentada en [Al94], el dominó es un juego convergente, de información perfecta y de muerte súbita. Dado esto, la longitud de la previsión no es constante. Una mejor forma para la previsión es una función dependiente de la capacidad de cómputo[2] del agente.

- ♢ Dado que en juegos de naturaleza similar al dominó se utiliza el concepto de equilibrio perfecto de subjuegos, un enfoque sensato para el concepto de solución podría ser un *equilibrio perfecto de subjuegos de previsión acotada*, en el cual se tendría que calcular en cada paso un nuevo equilibrio de acuerdo con algún beneficio deseado (función de ganancia).

- ♢ Un problema para aplicar previsión acotada al juego de ajedrez es la dificultad de determinar una *función sensata* para determinar la ganancia al final del horizonte acotado por la previsión. En el caso del dominó podemos usar las *heurísticas* o guías conocidas del *folclor* del juego para determinar esta función de ganancia e indicar que un jugador novato puede dar diferente valor a estás guías que un jugador experimentado como algunas de las presentadas en [Te01].

3.2.1. Juegos dinámicos de movimientos alternados

Ahora definimos un modelo de previsión con horizonte limitado por la capacidad de cómputo de cada agente, aplicado a *juegos dinámicos de movimientos alternados*. Las características principales de esta clase de juegos son que existen

[2]Capacidad de cómputo, en este contexto indica la capacidad que se tiene para generar y visitar cierta cantidad de nodos futuros.

dos (o más) jugadores que mueven secuencialmente en pasos discretos de tiempo. En cada periodo t, la ganancia actual del jugador i está en función de la acción que realiza el jugador en ese tiempo, y la acción realizada por el(los) otro(s) jugador(es) en tiempos anteriores. Los espacios de acciones son finitos y en general, se reducen durante el juego.

Esta última característica garantiza que en los últimos pasos de tiempo, el número de nodos para los subárboles finales sea muy reducido y pueda determinarse en tiempo razonable. Dicho de otra forma, el número de hojas es muy pequeño comparado con el número de ramas en los primeros turnos del juego.

Cada jugador tiene *habilidad limitada para predecir cierta cantidad de estados futuros*. El jugador i es caracterizado por su capacidad de generar estados c_i. En cada periodo t, el jugador i formula predicciones para los siguientes $n_i = f(c_i)$ movimientos después de su turno. Por tanto, debe elegir su acción en el turno actual solamente de acuerdo a su previsión limitada. Esto se debe a que:

1. El jugador i no puede construir sobre lo que sucederá después de n_i pasos, ya que no puede predecirlo (no tiene *awareness* de ello) y,

2. La ganancia promedio puede considerarse como una buena aproximación de la ganancia final real.

Para esto, la función de ganancia debe estar bien definida y aplicada al subárbol generado por la capacidad de cómputo del agente permitirá poder calcular la ganancia promedio entre los nodos terminales de *ese* subárbol.

El jugador i elige acciones a_i de un espacio finito de acciones A_i. Los jugadores toman sus acciones en pasos de tiempo discreto y el horizonte del juego es finito. Los periodos de tiempo son indexados como $t = 1, 2, 3, \ldots$

Si solo existen dos jugadores, realizan sus movimientos de forma alternada y el jugador 1 realiza el primer movimiento. En cada periodo impar ($t = 1, 3, 5, \ldots$), el jugador 1 elige una acción de su conjunto. De forma similar, el jugador 2 realiza movimientos en cada periodo par ($t = 2, 4, 6, \ldots$). En ambos casos, la acción elegida modifica la acción inmediata siguiente del otro jugador. En el tiempo t, la ganancia del periodo de cada jugador está en función de las acciones actuales a_i^t de ambos jugadores.

Un flujo de perfiles de acción $\{q_i^t\}_{t=1}^{n_{max}} = \left\{q_1^{2k-1}, q_2^{2k}\right\}_{t=1}^{n_{max}}$, con n_{max} el número de pasos necesarios para llegar al final del juego, $q_1^{2k-1} \in A_1$ y $q_2^{2k} \in A_2$ se conoce como un *camino* y se denota como Q. Como los jugadores realizan acciones cada dos pasos temporales, una acción elegida en un periodo t se combina con la acción tomada por el jugador contrincante en el periodo anterior $t - 1$ para modificar la estructura del árbol de juego (realiza una poda) y, por tanto, la ganancia del jugador i inducida por el camino Q.

3.2.1.1. El concepto de solución

Definiciones auxiliares Sea $\mathcal{H}(N_i)$ el conjunto de historias de acciones alternadas de longitud N_i, cuya última acción es un elemento de $A_j\,(j \neq i)$ y h un elemento de $\mathcal{H}(N_i)$.

1. Una predicción (pura) de longitud $n_i = f(c_i)$ para el jugador i es una serie de movimientos alternados de longitud n_i, que inicia con una acción que pertenece a $A_j\,(j \neq i)$. El conjunto de predicciones de longitud n_i (o menor para las últimas acciones) a partir del nodo siguiente, se denota como P_{n_i}.

2. Una previsión para el jugador i en el periodo t en el cual es su turno de mover, se denota como f_i^t; mapea para cada historia de longitud N_i, $h \in \mathcal{H}(N_i)$, el conjunto de acciones A_i que pueden elegirse para el conjunto de predicciones P_{n_i}.

3. $f_i = \{f_i^t\}_t$ denota una secuencia arbitraria de previsiones f_i^t para cada periodo t en el que el jugador i debe mover. el conjunto de f_i se denota como $\mathcal{F}_i$. Un par $(f_1, f_2) \in \mathcal{F}_1 \times \mathcal{F}_2$ se denota como f y el conjunto de f se denota como $\mathcal{F}$.

4. Una estrategia pura para el jugador i, se denota como σ_i, es una secuencia de funciones σ_i^t, una para cada periodo en el que el jugador i debe elegir una acción. La función en el periodo t, σ_i^t, es la estrategia de comportamiento del jugador i en ese periodo. Determina la acción del jugador i en el periodo t en función de las últimas N_i acciones. Formalmente, $\sigma_i^t : \mathcal{H}(N_i) \to A_i$. El conjunto de estrategias del jugador i se denota como Σ_i. Un perfil de estrategias (σ_1, σ_2) se denota como σ y el conjunto de perfiles de estrategias $\Sigma_1 \times \Sigma_2$ es denotado por Σ.

La idea principal de esta construcción es que las estrategias del jugador, es decir su elección de acciones, se basa en la previsión acotada en función a su capacidad de cómputo. Por tanto, para definir formalmente el concepto de solución, es necesario (1) especificar un criterio basado en la previsión como función de la capacidad y (2) mostrar cómo el equilibrio acotado se relaciona con las estrategias.

El criterio para determinar la acción a tomar consiste en determinar la mayor ganancia entre las ramas del subárbol que le es visible. Este criterio resulta natural, dadas las características del juego. En las primeras etapas (en las que el jugador no puede ver el horizonte del juego) esta ganancia le indicará con qué acción obtiene la mejor ganancia aunque esta ganancia puede no resultar la mejor en el árbol completo. En las etapas finales (una vez que el jugador puede ver todos los subárboles a partir del nodo actual) el agente puede determinar el *equilibrio real* y tomar le mejor *acción real* de acuerdo con el equilibrio perfecto por subjuegos. Dicho de otra forma, el jugador asume que puede ver todo el árbol y con base en esta suposición puede calcular un equilibrio por subjuegos en cada turno. Formalmente:

Definición 1 Una estrategia $\sigma_i \in \Sigma_i$ es justificada por una secuencia de previsiones $f_i = \{f_i^t\} \in \mathcal{F}_i$ si en cada turno el jugador i elige la acción que le entrega la mayor ganancia aplicando equilibrio perfecto por subjuegos para el árbol generado por su previsión acotada en función de la capacidad del jugador para generar y revisar nodos.

Por otro lado, se pueden relacionar los equilibrios de las previsiones con estrategias de equilibrio para el jugador i mediante una relación de *consistencia*, definida de la siguiente manera. Dada una historia h^* de longitud $t-1$, y una acción a_i del periodo actual t, $Q\left(\sigma|_{h^* a_i}\right)$ es el camino inducido por σ después de $(h^* a_i)$. En el periodo t, las N_i últimas acciones son $h = [h^*]^{N_i}$. La consistencia exige que para cualquier (h^*, a_i), la predicción $f_i^t\left(a_i|h\right)$ coincida con el truncamiento a las primeras $n_i = f\left(c_i\right)$ acciones del camino inducido por σ, $\left[Q\left(\sigma|_{h^* a_i}\right)\right]_{n_i}$. Es decir, la consistencia implica que las previsiones son correctas dentro y fuera del camino de equilibrio. Formalmente:

Definición 2 Una secuencia de previsiones $f_i = \{f_i^t\} \in \mathcal{F}_i$ es consistente con $\sigma \in \Sigma$ si para todo periodo t en el que el jugador i debe mover: $\forall a_i \in A_i$, $\forall h^* \in \mathcal{H}^{t-1}$, $f_i^t\left(a_i|h\right) = \left[Q\left(\sigma|_{h^* a_i}\right)\right]_{n_i}$ con $h = [h^*]^{N_i}$.

Ahora, el concepto de solución se puede definir como: una (n_1, n_2)–*solución* es un perfil de estrategia que puede *justificarse* con previsiones *consistentes* por los jugadores 1 y 2, es decir, un perfil de estrategia es asociado con secuencias de previsiones tales que (1) los jugadores eligen sus acciones para maximizar la ganancia obtenida aplicando equilibrio perfecto por subjuegos sobre la longitud de su previsión y (2) las previsiones del jugador i para las siguientes n_i acciones después de su movimiento son correctas dentro y fuera del camino de equilibrio. Formalmente:

Definición 3 El concepto de solución Un perfil de estrategia $\sigma = (\sigma_1, \sigma_2) \in \Sigma$ es una $(n_1, n_2) - solution$ si y solo si existen secuencias de previsiones $f = (f_1, f_2) \in \mathcal{F}$ tales que para $i = 1, 2$.

1. σ_i es justificada por f_i y
2. f_i es consistente con σ.

En [Je-98], Jehiel discute el proceso de aprendizaje y justifica que los jugadores eventualmente aprenden a realizar previsiones de forma correcta para calcular equilibrios con previsión acotada.

3.2.1.2. Ejemplo

El ejemplo que estaremos revisando en el desarrollo de esta y las siguientes secciones es una instancia del juego conocido como *draw*, perteneciente a un grupo más grande llamado *blocking games* y que se realiza con fichas de dominó. En estos juegos, las fichas son repartidas aleatoriamente boca abajo entre los jugadores al inicio del juego y el jugador que tiene el doble más grande (ficha con el mismo número en ambos extremos) es el que inicia la partida, colocando esa ficha sobre la mesa, dando inicio al llamado tren del juego. Posteriormente, los jugadores colocan de forma alternada en cada turno una ficha que corresponda con alguno de los extremos libres del tren. El juego termina cuando un jugador ha colocado todas sus fichas, o bien, cuando el juego está bloqueado, lo que ocurre si ningún jugador puede colocar otra ficha que corresponda con alguno de los extremos libres.

Mientras que en un *blocking game*, si quedan fichas después de la repartición inicial no se utilizarán durante la partida; en *draw*, las fichas restantes pueden ser tomadas de la mesa (aleatoriamente) si algún jugador las requiere. Al finalizar una partida, la suma de los puntos restantes es la ganancia obtenida por el ganador.

Consideraremos una instancia particular del juego de dominó para dos jugadores con seis fichas, de tal forma que cada jugador recibe tres, para nuestro ejemplo, consideramos que el jugador 1 recibe las fichas $\{(0,0), (0,1)\,(2,2)\}$ y el jugador 2 $\{(0,2), (1,1)\,(1,2)\}$; además, la capacidad de cómputo de cada agente es $2n$ con n el número de fichas que cada jugador tiene al inicio.

Además, como heurística, se utiliza solo la guía de jugar conservativamente (deshacerse de las fichas altas). El objetivo principal del juego es obtener la mayor cantidad posible de puntos y después se toma en cuenta la guía mencionada.

Es importante mencionar que cuando los nodos generados y visitados por un agente no cubren completamente un nivel, se considera que no posee información alguna sobre el nivel trunco, dado que no se puede determinar de antemano qué nodos serán generados para obtener la ganancia hasta ese punto.

El juego descrito anteriormente genera el árbol de la figura 3.1. Entre otras cosas, se observa que sí el jugador 1 tira la ficha $(0,0)$ al inicio existe una ruta que le entrega una ganancia de 3. Sin embargo, como su previsión está limitada a que puede generar solamente seis estados a partir del actual, no tiene *awareness* de esta posibilidad.

Al inicio, el número de niveles que puede visitar el jugador 1 es $n_1 = 1$; por tanto, su ganancia la evalúa en función de la guía de deshacerse de las fichas más altas y decide tirar la ficha $(2,2)$.

En esta sección, en las figuras se muestran encerrados los niveles que el agente puede generar y visitar. Además, en gris están las ramas que ya no se pueden jugar. En la figura 3.2, se observa que el jugador 2 tiene ahora una previsión de $n_2 = 2$.

Ahora es el turno del jugador 2, quien tiene la opción de tirar las fichas $(0,2)$ y $(1,2)$; la ficha $(0,2)$ puede darle el triunfo; sin embargo, no puede darse cuenta de este hecho, dada la limitación de su previsión y, por tanto elige tirar su ficha $(1,2)$, la más alta.

En este punto, la figura 3.3 muestra que el jugador 1 solo puede tirar su ficha $(0,1)$, ya puede prever el resto del árbol, $n_1 = 4$ y sabe que ganará, obteniendo una ganancia de 2.

La figura 3.4 muestra las últimas tiradas de los jugadores. El jugador 2 solo puede tirar la ficha $(0,2)$, lo puede hacer dejando en los extremos libres $(0,0)$ o $(2,2)$, en ambos casos pierde entregando 2 puntos al oponente, pero decide dejar $(2,2)$ para *ahorcar* la ficha $(0,0)$ del oponente.

Este ejemplo muestra que si el jugador 2 tuviera mayor capacidad para generar y visitar estados futuros habría elegido tirar la ficha $(0,2)$ para ganar la partida. Además, también muestra que la profundidad de la previsión crece conforme evoluciona el juego, dado que el número de estados futuros se reduce con cada ficha tirada. Esta característica junto con la finitud intrínseca del juego hacen que sea factible construir un equilibrio (y estrategia) por subjuegos *más cercano* al perfecto en cada paso temporal.

3.2.1.3. Propiedades

Algunas de propiedades interesantes y útiles del modelo son:

- ♢ Un juego dinámico de movimientos alternados siempre tiene al menos un *equilibrio perfecto por subjuegos con previsión acotada*. Utilizando el Co-

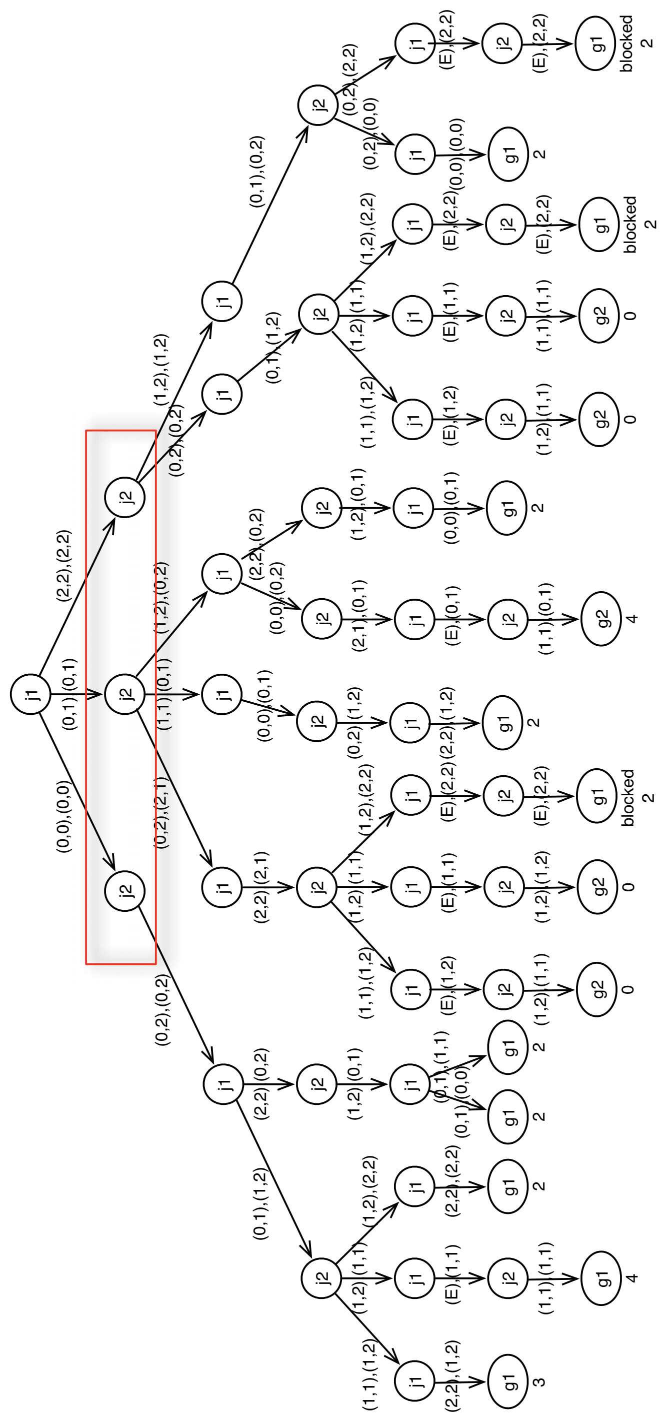

Figura 3.1: Árbol completo

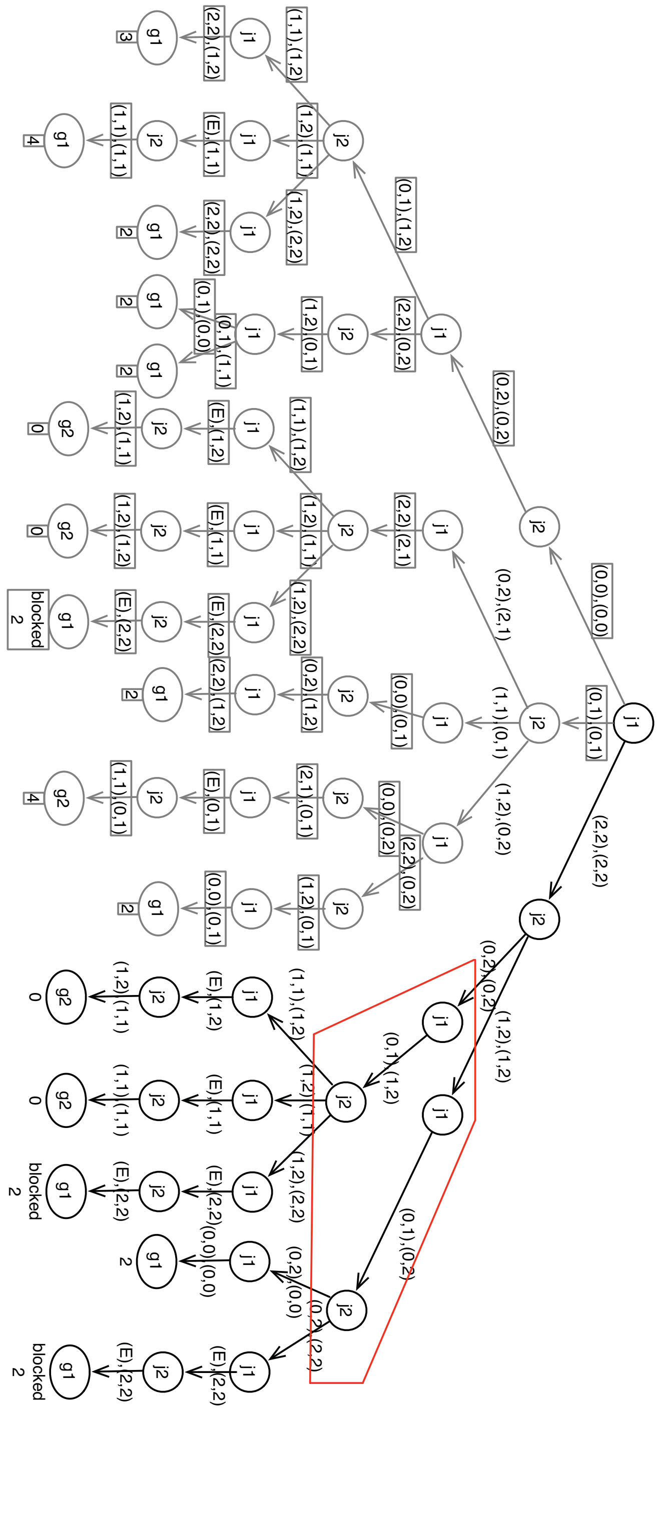

Figura 3.2: Árbol después de la primera ficha tirada por el jugador 1

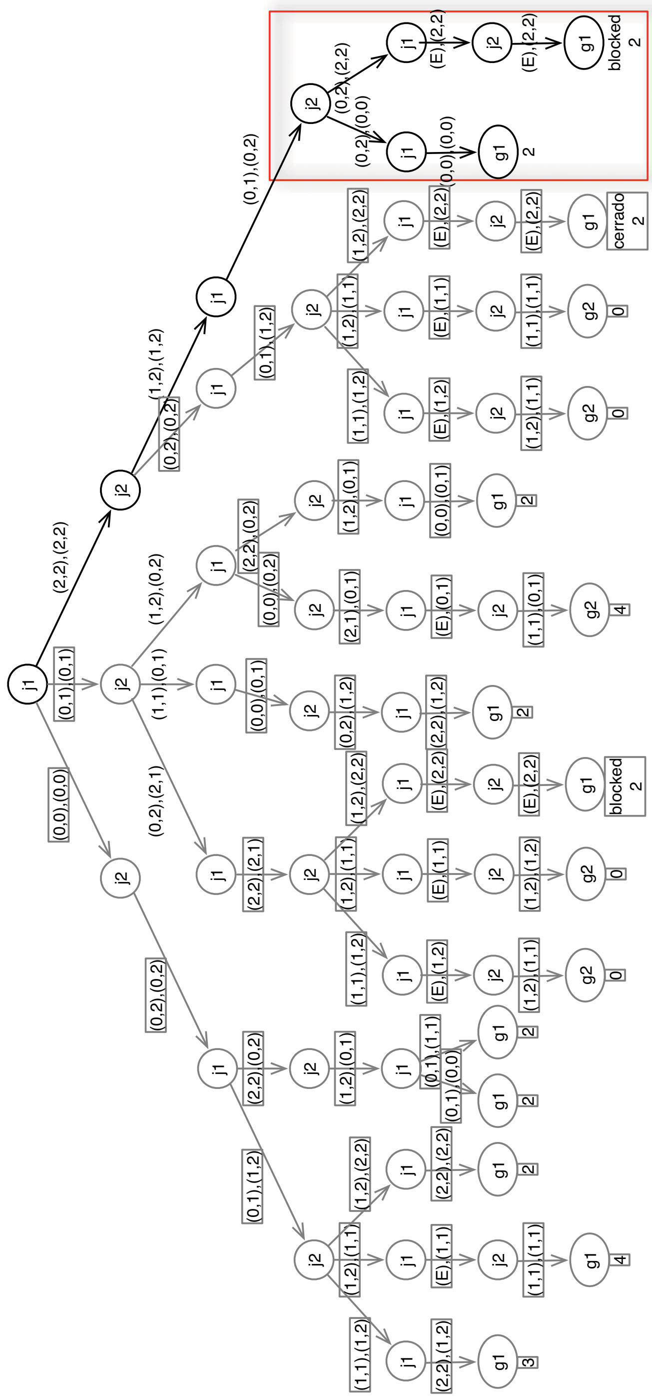

Figura 3.3: Árbol después de la primera ficha tirada por el jugador 2

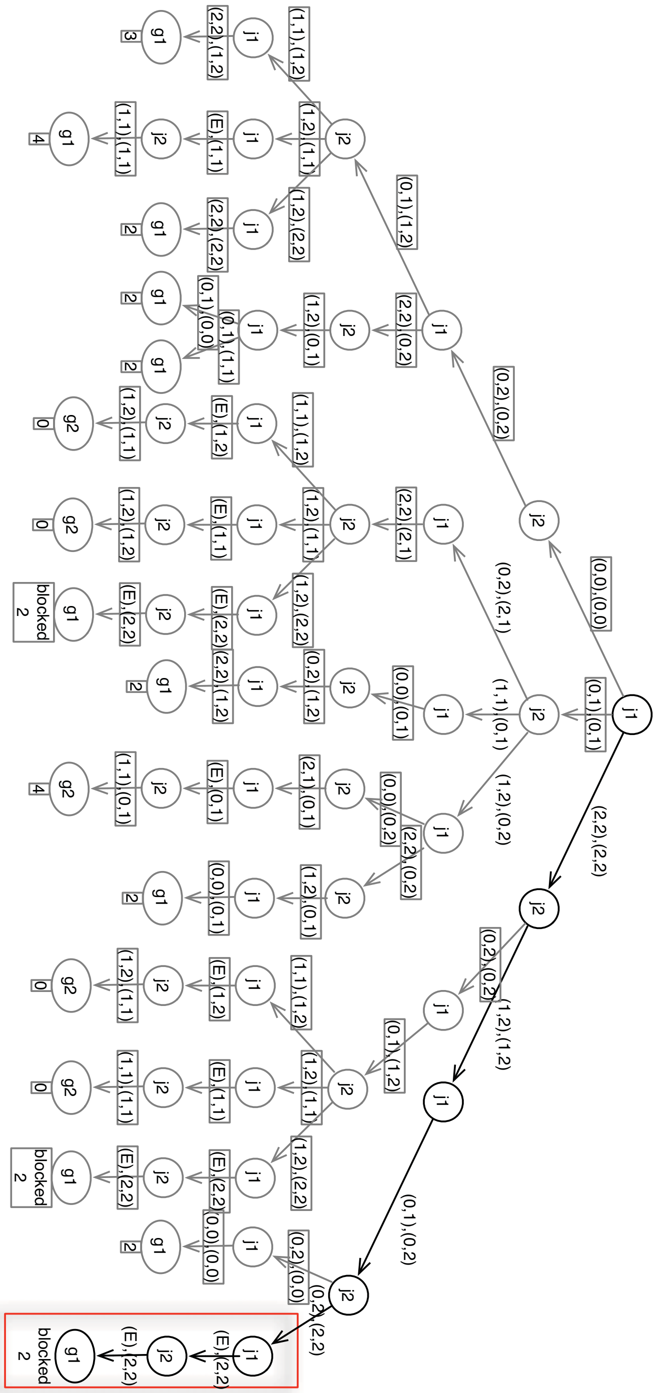

Figura 3.4: Árbol en los últimos pasos del juego

rolario de Kuhn, se garantiza que para cada subárbol existe al menos un equilibrio perfecto por subjuegos, entonces se puede construir un equilibrio perfecto por subjuegos con previsión acotada.
Corolario de Kuhn del Teorema de Zermelo-von Neumann
Todo juego genérico Γ de *n–jugadores* con información perfecta siempre tiene un punto de equilibrio en estrategias puras [Ku-97].

◇ El modelo es independiente de la historia: las decisiones se toman a partir de la última acción (realizada por el oponente) y el estado actual del juego.

◇ Contrario al modelo presentado por Jehiel [Je95], dado que el juego es dinámico y finito (de muerte súbita), en este tipo de juegos sí puede haber beneficio al incrementar la capacidad de cómputo (previsión) del jugador.

3.2.2. Un modelo similar

Por otro lado, Mirrokni et al. [Mi12] presentan un modelo que puede aplicarse a juegos similares. Mencionan que su modelo es muy flexible y puede usarse en diversos contextos:

◇ Puede usarse en árboles dinámicos de búsqueda. Ellos asumen que T_i es un árbol de búsqueda primero en amplitud con profundidad fija k_i.

◇ Una función de evaluación que, usualmente será diferente para distintos tipos de jugadores. En sus ejemplos utilizan la función estándar *max* (o *min*).

◇ Los autores distinguen entre dos grandes clases de juegos: modelo de *hojas* y modelo de *caminos*. En el primer caso las ganancias se determinan al final del juego; por tanto, los valores de *hojas* en árboles *locales* son estimaciones del resultado final. El ajedrez y el dominó entran en esta categoría. En la segunda categoría, las ganancias se acumulan conforme avanza el tiempo. Juegos repetidos como los juegos industriales de múltiples periodos entran en este grupo.

◇ Además, discuten cómo el orden de los movimientos en el juego afecta el proceso de toma de decisiones. Examinan dos situaciones: peor caso del *lookahead* y caso promedio del *lookahead*.

Nuestro modelo se adapta bien a los tres primeros contextos:

- ♢ En árboles dinámicos, nuestro modelo puede comportarse mejor, dado que la profundidad no es fija: depende de la capacidad de cómputo del agente y es fácilmente intercambiable por otro tipo de límite; por ejemplo, tiempo disponible.
- ♢ La función de evaluación la utilizan de forma similar a nuestra guía heurística, que también es muy flexible.
- ♢ También es aplicable a ambos tipos de modelos. Ya mostramos el caso del modelo de hojas. Para el modelo de caminos, utilizaremos uno de los juegos presentados por Jehiel.

3.2.2.1. Ejemplo

Para mostrar que nuestro modelo es también aplicable al llamado *de caminos*, lo aplicaremos a uno de los juegos definidos en [Je95]. La descripción de este juego particular es: ambos jugadores tienen la misma longitud de previsión $n_1 = n_2 = 1$ y la misma capacidad de memoria $N_1 = N_2 = 2$. Jehiel supone que existe un equilibrio de tipo $(n_1, N_1; n_2, N_2)$–*solución* σ, tal que la previsión asociada al jugador 1 en el periodo $2k+1$ está dada por:

- ♢ $f_1^{2k+1}(U|\,UL) = L$, $f_1^{2k+1}(U|\,UR) = R$, $f_1^{2k+1}(U|DL) = R$, $f_1^{2k+1}(U|DR) = L$ y
- ♢ $f_1^{2k+1}(D|\,UL) = R$, $f_1^{2k+1}(D|\,UR) = R$, $f_1^{2k+1}(D|DL) = L$, $f_1^{2k+1}(D|DR) = L$.

Con esta suposición, construye la previsión del jugador 2 las siguientes ganancias: $UL = (3,2)$, $UR = (0,3)$, $DL = (2,1)$, $DR = (2,0)$. En el periodo $2k$, con una historia arbitraria h^* de tamaño $(2k-1)$ en el que la última acción es U (tomada por el jugador 1), si el jugador 2 elige la acción L en el periodo $2k$, entonces la historia en $(2k-1)$ es (h^*, L) el truncamiento, dada la memoria también limitada, es $h = [h^*, L]^{N_2} = (U, L)$. A continuación define que el jugador *debe* elegir U en lugar de D dado que le entrega una mejor ganancia $v_1\left(hUf_1^{2k+1}(U|h)\right) > v_1\left(hDf_1^{2k+1}(D|h)\right)$ $(3+3+3 > 3+2+2)$. Por la consistencia, las predicciones de equilibrio hechas por el jugador 2 en el periodo $2k$ son correctas y *debe* prever que si elige L, entonces el jugador 1 elegirá U en el periodo $(2k-1)$: $\forall a_2 \in A_2$, $f_2^{2k}(L|\,(a_2, U)) = U$. Utilizando el mismo argumento, encuentra que $\forall a_2 \in A_2$, $f_2^{2k}(R|\,(a_2, U)) = D$, $f_2^{2k}(L|\,(a_2, D)) = D$ y $f_2^{2k}(R|\,(a_2, D)) = D$. Finalmente demuestra que la solución es independiente de la historia y utiliza un equilibrio de tipo $(n_1, n_2) - \textit{solution}$ sin historial.

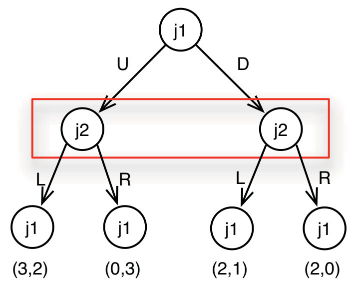

Figura 3.5: Árbol inicial para el juego alternado repetido

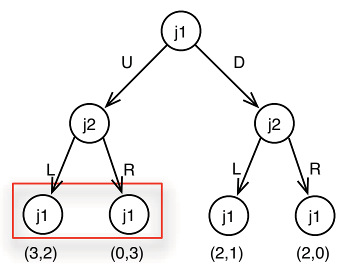

Figura 3.6: Árbol después del primer turno de 1

El resultado de Jehiel puede verse más claramente si se utiliza el árbol del juego (figura 3.5). Suponiendo capacidad de cómputo para cada jugador es $c_1 = c_2 = 2n$, con n el número de acciones disponibles en cada periodo. Note que esta capacidad computacional entrega la misma previsión inicial que el ejemplo original $n_1 = n_2 = 1$.

Ahora, suponiendo que el jugador 1 implementa nuestra modelo y que su guía es una base de conocimientos[3] inicialmente vacía. Con esto, no puede saber que eligiendo D tiene ganancia asegurada de 2 entonces puede elegir U. En el periodo siguiente, el jugador 2 (también con base de conocimientos vacía) elegirá R porque es la que le entrega mejor ganancia, como se ve en la figura 3.6.

En este punto, el jugador 1 actualiza su base de conocimientos agregando la entrada $(U, 0)$, indicando que la acción U le entrega ganancia 0. Además, esto significa que esta acción le resulta *irracional*. Por tanto, en su turno actual no considera dicha rama. Con esta actualización, el jugador 1 ya puede ver el subárbol de la acción D completo y *aprende* que obtendrá ganancia de 2 (figura 3.7).

[3]En el mismo sentido usado en sistemas expertos.

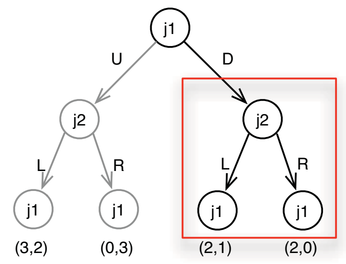

Figura 3.7: Árbol en el segundo turno de 1

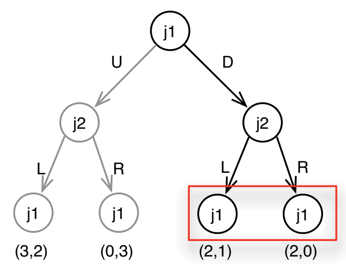

Figura 3.8: Árbol en el segundo turno de 2

Finalmente, el jugador 2 elige L dado que le entrega mejor ganancia como se observa en la figura 3.8. Y el jugador 1 agrega la entrada $(D, 2)$.

Ahora supongamos que inicialmente el jugador 1 elige D; entonces el jugador 2 elegirá L y el jugador 1 agrega $(D, 2)$ a su base de conocimientos. Si en su siguiente turno elige U, el jugador 2 selecciona R y el jugador 1 agrega $(U, 0)$ en su base. Este segundo razonamiento lleva a la misma conclusión: elegir U es irracional para el jugador 1.

El equilibrio tipo (c_1, c_2)–*solución* así construido, es diferente al (n_1, n_2)–*solución* obtenido por Jehiel. Sin embargo, es posible obtener las mismas ganancias para ambos jugadores. Dos maneras de hacerlo suponen que el jugador 2 tiene (o puede obtener) más información que el jugador 1:

- $\diamondsuit$ Si el jugador 2 tiene capacidad de cómputo suficientemente grande para reconstruir el árbol completo, entonces verá que obtendrá mejor ganancia al elegir L cuando el jugador 1 juega U que en cualquier otro caso.

- $\diamondsuit$ Si la base de conocimientos del jugador 2 no está inicialmente vacía y contiene información (guías) que le indica que cuando el jugador 1 juega U

obtendrá mejor ganancia global eligiendo L que en cualquier otro caso.

Existe además otra forma de obtener el mismo resultado. Es aún más interesante, dado que no supone ningún tipo de información adicional para ninguno de los jugadores; solo requiere que ambos puedan actualizar sus bases de conocimientos:

- ◇ Ambos jugadores *solo* deben jugar todas sus posibilidades y agregar las entradas correspondientes en sus bases. Esta situación es realista, dado que el juego se repite de forma indefinida.

Aplicando esta idea, después de una ronda completa, ambos jugadores tendrán en sus bases de conocimiento la información necesaria para que sea utilizada por la función de guías. Con esto, la *eventualidad* de aprender a jugar se reduce a un juego completo por ambos jugadores.

3.3. Conclusiones

En este capítulo se presenta un modelo de previsión limitada como función de la capacidad de cómputo del agente y lo aplicamos tanto a la clase de juegos dinámicos alternados como a la de juegos alternados repetidos. Además, definimos el concepto de solución que llamamos *equilibrio perfecto por subjuegos con previsión acotada* que se sustenta en el bien establecido equilibrio perfecto por subjuegos aplicado a cada subárbol generado en cada periodo del juego.

Adicionalmente, analizamos algunas propiedades de este concepto de solución. Para mostrar su funcionamiento, presentamos un ejemplo concreto de su aplicación en una instancia del juego de dominó. De igual forma, lo utilizamos para analizar un juego presentado originalmente por Jehiel [Je95] y mencionamos la forma en que se puede obtener el mismo resultado utilizando el mismo modelo para determinar las ganancias parciales en este juego repetido.

Capítulo 4

Lógica y racionalidad acotada

Resumen

En el presente capítulo se muestra el desarrollo de un lenguaje lógico que sirve para razonar sobre situaciones en las que los jugadores eligen sus estrategias basados en una búsqueda incompleta del árbol del juego. lógica para racionalidad acotada ($\mathcal{BRL}$) extiende a la lógica proosicional utilizando dos ideas ya conocidas. Por un lado, una semántica de vecindarios [PMS07] y por el otro una lógica de preferencias [BS11]. Sin embargo, La forma aquí utilizada es novedosa porque ambas se actúan de forma distinta a las anteriormente presentes. Como motivación e ilustración del uso de este lenguaje, utilizamos el juego de dominó. Este capítulo se basa en el trabajo [FEH14].

4.1. Teoría de juegos y dominó

En años recientes, ha crecido el interés en desarrollar marcos lógicos prácticos para modelar la teoría matemática de juegos. Existen diversas posibilidades para llevarlo a cabo, pero muchos sistemas tienen limitaciones cuando se desea razonar directamente sobre el conocimiento estratégico de los jugadores, que es relevante cuando los agentes no tienen conocimiento completo del juego o cuando tienen recursos computacionales limitados para elegir sus acciones. Una cuestión particular es describir cómo fluyen tanto la información como el conocimiento durante la evolución del juego.

Un primer intento de utilizar lógica para estudiar dominó es el de Schwarz. En [Sc01] analizó un juego de parejas utilizando lógica de primer orden. Otro trabajo relacionado es el presentado por Velázquez-Quesada y Hernández-Quiroz en [VH05, VH06] quienes describen el conocimiento de cada jugador usando un tipo de lógica dinámica epistémica. De esta forma, no es necesario utilizar fórmulas para describir propiedades del conocimiento modelado: están implícitas en la relación de accesibilidad del modelo en el que las fórmulas serán interpretadas. Al añadir acciones a su lenguaje, pudieron expresar la forma en la que el conocimiento de cada jugador es actualizado. Con este trabajo se puede describir:

- ♢ Escenarios para el desarrollo de un juego.
- ♢ El conocimiento que los jugadores tienen acerca de la situación del juego.
- ♢ Cómo se modifica este conocimiento por las acciones realizadas.

Sin embargo, dejan como trabajo futuro:

- ♢ La representación del conocimiento estratégico.
- ♢ La representación del conocimiento histórico.

En esta sección presentamos un desarrollo en la dirección del primer punto: representación de conocimiento estratégico utilizando un tipo de lógica modal. Para esto tomamos de ejemplos los presentados por Jehiel [Je95] y por Hernández-Quiroz y Espinosa-Avila en [EH13]; en ambos modelos se utiliza un método de razonamiento heurístico, dado que las estrategias no son optimas dadas las restricciones en la previsión de los jugadores.

4.2. Lógica modal y razonamiento estratégico

Con el fin de modelar el comportamiento estratégico de los agentes bajo condiciones de racionalidad acotada, tomamos prestadas dos ideas ya conocidas en la literatura. La primera de ellas es la semántica de *vecindades* para modelar conocimiento incompleto como el usado en *topologic* [PMS07]. En este, las vecindades modelan la cantidad de "esfuerzo" que un agente invertirá para determinar el estado real del mundo. La segunda es una lógica de preferencias [BS11], en el que los agentes razonan acerca de *mundos maximales* bajo cierto orden de preferencias.

En nuestra propuesta un agente elige una opción deseable de entre cierta cantidad de posibilidades; sin embargo, no puede comparar todas las posibles ganancias *reales*, dadas sus limitaciones en recursos computacionales. Por tanto, sólo examinará una muestra de las posibilidades y elegirá entre los elementos de esa muestra. Utilizamos las vecindades para modelar dicha muestra, similar a como se utiliza en *topologic*, pero con diferencias sustanciales.

En ambas propuestas, el estado actual del mundo (digamos w) está ligado a un conjunto de estados posibles (digamos, V), y sólo los estados pertenecientes a V se toman en cuenta. En *topologic*, el objetivo del agente es distinguir el estado real de las cosas de entre todas las posibilidades. Por tanto, un vecindario más pequeño $V' \subseteq V$ le entregará mayor información. Se puede decir que V representa un "margen de error". En cambio, en nuestra propuesta el objetivo del agente es elegir una de varias posibles ganancias, por tanto un vecindario mayor $V' \supsetneq V$ le entregará *mayor* información. Sin embargo, el tratamiento formal de vecindarios es similar (pero no idéntico). En lugar del operador B, añadimos una variable proposicional best que es cierta *solo* en puntos $\preccurlyeq$*–maximales*. Esto permitirá que los agentes planeen una estrategia con la que puedan alcanzar estos estados preferidos y no solo razonar acerca del estado de las cosas en dichos puntos.

Además, existe otra novedad en cómo el orden de preferencias y los vecindarios interactúan y que es clave para modelar heurísticas como la de Jehiel. Se da por la interacción entre preferencias y vecindades: como el agente solo considera estados en V, sólo maximizará sus ganancias *sobre* V. Por tanto, best no solo será cierta en los estados que son $\preccurlyeq$*–maximales*, sino también en aquellos que son $\preccurlyeq$*–maximales dentro de* V.

4.3. El modelo formal

Ahora precisamos las nociones mencionadas anteriormente.

4.3.1. Sintaxis

Sea *Act* el conjunto finito de acciones que puede realizar un agente. El lenguaje básico de la lógica para racionalidad acotada $\mathcal{BRL}$ (*bounded rationality logic*) incluye la lógica proposicional con la variable proposicional especial best y los operadores unarios $\Diamond$ y $\langle i \rangle$ para $i \in Act$. Además, incluye un conjunto infinito y contable de proposiciones atómicas At.

Formalmente, la gramática de $\mathcal{BRL}$ está dada por:

$$\varphi ::= \top \mid p \mid \mathsf{best} \mid \neg\varphi \mid (\varphi \vee \psi) \mid \Diamond\varphi \mid \langle i \rangle\varphi$$

Para un orden total $\preccurlyeq$, escribimos $v \prec w$ si $v \preccurlyeq w$ pero $w \not\preccurlyeq v$; y $v \approx w$ si $v \preccurlyeq w$ y $w \preccurlyeq v$. Además, un orden parcial $\prec$ es *Noethereano* si no existe una secuencia infinita del tipo $a_0 \prec a_1 \prec a_2 \prec \cdots$.

Definición 1 *Un marco de preferencias acotadas*, es una tupla:

$$\mathcal{X} = \left\langle W, \preccurlyeq, \{R_i\}_{i \in Act}, \mathcal{N} \right\rangle,$$

donde W es un conjunto de estados, $\preccurlyeq$ es un orden total Noethereano y $\mathcal{N} \subseteq W \times \mathcal{P}(W)$ es una relación de vecindarios.

Un *modelo de preferencias acotadas* es un marco de preferencias acotado equipado con una evaluación $l : At \to P(W)$.

Definición 2 Sea $\mathcal{M} = \langle W, \preccurlyeq, \{R_i\}_{i \in Act}, \mathcal{N}, l \rangle$ un modelo de preferencias acotadas, definimos la relación de satisfacción sobre (w, V), donde $w \in W$ y $V \subseteq W$ como:

- ♢ Las fórmulas proposicionales de la forma común
- ♢ $(w, V) \models \langle i \rangle \varphi$ si existe v tal que wR_iv y $(v, V) \models \varphi$
- ♢ $(w, V) \models \Diamond\varphi$ si y solo si existe $V' \in \mathcal{N}(w)$ tal que $(w, V') \models \varphi$
- ♢ $(w, V) \models \mathsf{best}$ si y solo si, siempre que $v \in V$, se cumple $v \preccurlyeq w$ (i.e., w es $\preccurlyeq$–*maximal* dentro de V)

Los operadores duales $\Box$ y $[i]$ se pueden introducir de la forma usual. Las versiones correspondientes a los axiomas K de estos operadores son también válidas.

4.3.2. Modelado lógico de juegos alternados

4.3.2.1. Juegos alternados repetidos

Con el objeto de mostrar algunas proposiciones expresables con nuestro lenguaje, primero lo haremos con un ejemplo introducido por Jehiel en [Je95]. Recordemos que el árbol para dicho juego es el mostrado en la figura 4.1 (eliminando las cajas usadas en el capítulo anterior). Recordemos también que en la clase de juegos alternados repetidos no existen estados finales y que la ganancia global se calcula acumulando las ganancias de cada iteración. El mejor estado global es en el que las ganancias son $(3, 2)$ (estado 4), es decir, si el juego se repite indefinidamente, la secuencia de acciones que entrega las mejores ganancias es U, L. Pero si se trata de un juego sencillo, desde el punto de vista del jugador 2 (estados cuadrados), es mejor elegir R cuando el jugador 1 eligió U porque le entrega mejor ganancia.

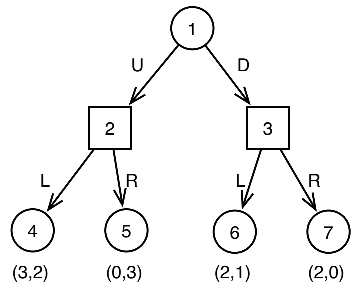

Figura 4.1: Árbol para el juego alternado repetido

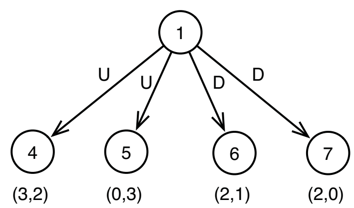

Figura 4.2: Árbol colapsado para el juego alternado repetido

Como realizaremos un análisis desde el punto de vista de un solo agente, obtenemos el árbol *colapsado* del juego (4.2).

Considerando solo una ronda de juego, tenemos que el conjunto de estados es $W = \{1, 4, 5, 6, 7\}$. La relación de preferencias está dada por la ganancia del jugador 1. Además, se puede considerar que las ganancias en el estado 1 son cero. Con esto, obtenemos que $1 \approx 5 \prec 6 \approx 7 \prec 4$. Finalmente, para los vecindarios, tenemos: $\mathcal{N}(1) = \big\{\{4, 5, 6, 7\}\big\}$ y $\mathcal{N}(w) = \varnothing$ para $w \neq 1$, dado que no consideramos más movimientos. Sea $V = \{4, 5, 6, 7\}$. Entonces tenemos las siguientes proposiciones expresables en $\mathcal{BRL}$:

- $\Diamond$ $(4, V) \models \mathsf{best}$, dado que $4 \in V$ y 4 es $\preccurlyeq$–*maximal* dentro de V.
- $\Diamond$ $(1, V) \models \langle U \rangle\, \mathsf{best}$, intuitivamente, cuando el jugador 1 elige la acción U, existe al menos un camino hacia su estado preferido (4 en el análisis de Jehiel).
- $\Diamond$ $(1, V) \models \neg \langle D \rangle\, \mathsf{best}$; no importa lo que haga el jugador 2, después de elegir D, el jugador 1 no alcanza su estado preferido (equivalente a $(1, V) \models [D]\, \neg\mathsf{best}$).

Ahora, consideremos un vecindario diferente, inspirado en la heurística presentada en [EH13], en la que el agente tiene capacidad computacional para considerar cierto número de estados, digamos dos. Al igual que en el caso anterior, tenemos: $\mathcal{N}(1) = \{\{4,5,6,7\}\}$ y $\mathcal{N}(w) = \varnothing$ para $w \neq 1$, pero, ahora $\mathcal{N}'(1)$ es el conjunto de todos los subconjuntos de V que tienen dos elementos. Entonces, podemos expresar:

- ♢ $(1,\{4,6\}) \models [U]\mathsf{best}$; de acuerdo a nuestra semántica, como el agente no toma en cuenta al estado 5, la única salida esperada al elegir U es 4, que es su estado preferido.

- ♢ $(1,\{5,6\}) \models \diamond[U]\neg\mathsf{best}$; esto es porque $\{5,6\} \in \mathcal{N}'(1)$, pero entre $\{5,6\}$, el estado $\preccurlyeq -maximal$ es 6. Si, el agente ya eligió $\{5,6\}$ como sus estados vecinos, debería elegir D en lugar U.

Este ejemplo muestra que nuestro sistema no solo permite razonar acerca de un conjunto de estados que el jugador considera, sino también puede servir para mostrar cómo los cambios en dicho conjunto lo llevan a cambiar sus decisiones.

4.3.2.2. Juegos dinámicos alternados

Regresando al juego que motivó este análisis y con el objeto de mostrar el tipo de proposiciones que deseamos expresar en nuestro lenguaje, nos referiremos a una instancia concreta del juego de dominó presentada en [EH13] y descrita en la sección 3.2.1.2 de este trabajo. El árbol de juego asociado se muestra en la figura 4.3 (sin las cajas del análisis anterior).

Siguiendo la estrategia del análisis del juego de Jehiel, modelaremos este juego desde el punto de vista del jugador 1. La figura 4.4 muestra la versión colapsada de esta partida de dominó desde su perspectiva.

En dicha figura, los estados intermedio (circulares) representan las acciones que puede elegir por el jugador 1. Las transiciones corresponden con la ficha tirada por el jugador 1 y los extremos libres después de la acción tomada por 2. Los números ubicados cerca de los estados son la suma total de puntos que el jugador 1 conserva al ubicarse en ese estado. Es importante recordar que se ha considera la heurística simple de *deshacerse de las fichas más pesadas*. Los estados finales tienen doble línea, si son cuadrados, ha ganado el jugador 2 y la ganancia del jugador 1 es cero o negativa.

Por tanto, el conjunto de estados ahora está formado por $W = \{1,\ldots,46\}$. La relación de preferencia, dada por $v \prec w$ si el jugador 1 tiene al menos tantos puntos en el estado v como en el w. Dado un estado no terminal w, $\mathcal{N}(w)$ es el

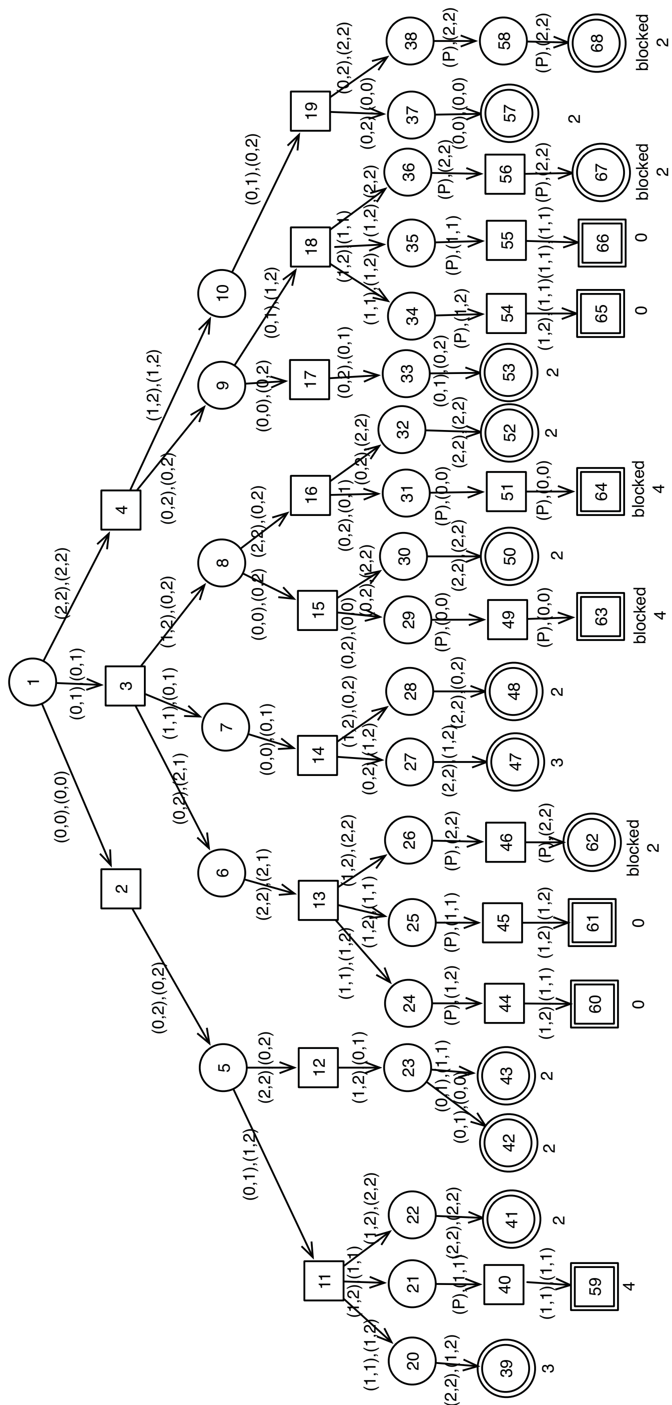

Figura 4.3: Árbol para la partida de dominó

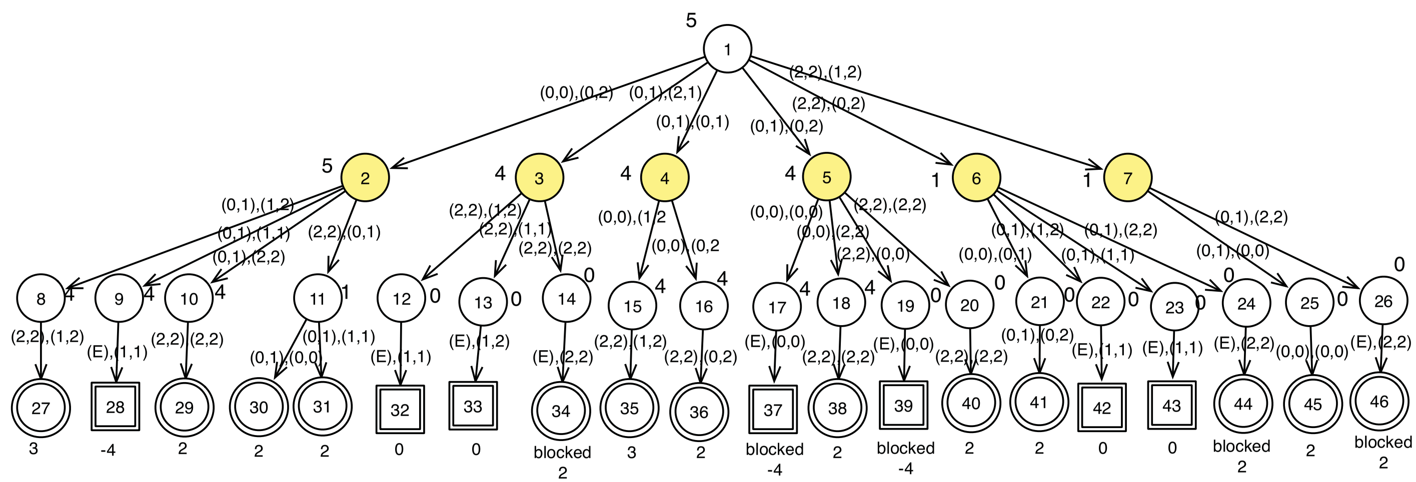

Figura 4.4: Árbol de juego del dominó desde la perspectiva del jugador 1. Los estados amarillos conforman al vecindario Y

conjunto de estados que pueden alcanzarse desde w en un solo movimiento. Por ejemplo, $Y \in \mathcal{N}(1)$, donde $Y = \{2, 3, 4, 5, 6, 7\}$ está marcado en color amarillo (gris si se imprime sin color) en la figura 4.4. Con este árbol de juego, podemos expresar fórmulas como las siguientes:

- $(1, Y) \models [(2, 2)]\,\mathsf{best}$. Esta fórmula indica que, sin importar lo que elija el jugador 2, después de tirar la ficha $(2, 2)$, el jugador 1 alcanzará su estado preferido en el conjunto Y; es decir, donde conserva un punto.

- $(1, U) \models \langle(2, 2)\rangle\,[(0, 1)]\,\mathsf{best}$, donde U incluye los estados accesibles en dos pasos. La intuición detrás de esta fórmula es que después que el jugador ha tirado la ficha $(2, 2)$, existe al menos un camino que lo lleva a su estado preferido: aquel en el que ya no tiene puntos.

- $(1, U) \not\models \langle(2, 2)\rangle\,\langle(0, 1)\rangle\,\top$. Intuitivamente, esta fórmula muestra un caso en que el jugador no tiene capacidad de ver dos pasos en el futuro y, desde su punto de vista, es imposible realizar dos movimientos seguidos.

- $(1, U) \models \langle(2, 2)\rangle\,\Box\,\langle(0, 1)\rangle\,\mathsf{best}$. En comparación con el ejemplo anterior, mientras que el agente 1 sólo puede ver un paso hacia adelante, nuestro lenguaje permite que razonemos acerca de lo que sucederá si repite su heurística: después de jugar $(2, 2)$, el jugador 1 volverá a analizar el árbol del juego en un paso hacia adelante, eligiendo ahora jugar la ficha $(0, 1)$.

4.4. Conclusiones

La lógica para racionalidad acotada $\mathcal{BRL}$ (*bounded rationality logic*) permite expresar las elecciones tomadas por un jugador bajo ciertas restricciones computacionales. Aún cuando las restricciones mostradas se basan en el modelo de *limited horizon forecast* de Jehiel, en principio no parece haber mayor restricción para que otras formas de racionalidad acotada puedan ser utilizadas.

Como trabajo posterior, sería deseable tener un cálculo completo y consistente, además de un procedimiento efectivo para verificación de modelos para esta lógica: como es usual, se puede asumir que el árbol existe (o al menos secciones de él) y los jugadores deben razonar sobre esta estructura *fija*.

Otra extensión de interés, es modelar razonamiento simultáneo con varios agentes, incluyendo razonamiento sobre conocimiento de los otros, que llevaría a un modelado mucho más completo sobre agentes que interactúan estratégicamente.

Capítulo 5

Teoría de juegos y optimización

Resumen

Ahora presentamos otra manera de analizar diversas clases de juegos. El objetivo es proveer un marco conceptual metodológico que combine ideas de varias perspectivas, entre ellas, racionalidad acotada, teoría de juegos y programación dinámica, con mayor énfasis en las dos últimas. Para esto, retomamos ideas de algunos trabajos realizados anteriormente y aplicamos optimización que considere que un agente tiene racionalidad acotada.

En particular estudiamos la relación entre juegos de tiempo discreto (espacio de estados) y su versión de *tiempo continuo*; y como la segunda puede entenderse como en límite de la primera. Además, mostramos casos en los que la ecuación de Hamilton–Jacobi–Bellman (HJB) para el caso discreto del juego puede conducir a su correspondiente ecuación diferencial parcial de HJB en tiempo continuo. El paso de la versión discreta a la continua es útil y de interés porque permite determinar soluciones explícitas de la ecuación de HJB asociada, con lo cual es posible obtener información útil sobre estrategias *óptimas*. Finalmente, desarrollamos ejemplos en los que se observa este procedimiento. Este análisis también muestra una manera distinta de calcular un equilibrio perfecto por subjuegos. El presente capítulo está basado en el artículo [EPH17].

5.1. Motivación

Con el fin de analizar diferentes clases de juegos y, en general, la solución de problemas, se han utilizado diversos enfoques, entre otros teoría de juegos, lógica, racionalidad acotada y programación dinámica.

Por ejemplo, [MB97] menciona: "*Problem solving in AI often reduces to a process of tackling three main issues: representing knowledge, searching for solutions, and using knowledge to direct the search*". Por otro lado, el grupo de IBM a cargo de este desarrollo escribió en su sección de preguntas frecuentes (*faq*) que el «*Deep Blue's "chess knowledge" has improved since last year*» [DB], indicando que una de las mejoras importantes es justamente en cuanto al conocimiento que el sistema tiene del área. Otro ejemplo interesante, es Watson, definido por IBM como «*an application of advanced Natural Language Processing, Information Retrieval, Knowledge Representation and Reasoning, and Machine Learning technologies to the field of open-domain question answering. At its core, Watson is built on IBM's DeepQA technology for hypothesis generation, massive evidence gathering, analysis, and scoring*» [DQA]. En estos ejemplos puede observarse que se siguen buscando formas de determinar ganancias aún sin tener racionalidad perfecta, es decir, con capacidad de cómputo limitada.

En esta sección, revisaremos algunos enfoques con el fin de proveer un marco de trabajo que combina algunas de las perspectivas mencionadas, aplicándolas a diferentes tipos de juegos. El objetivo principal es utilizar programación dinámica y teoría de juegos aplicando sus ideas a racionalidad acotada con juegos alternados, tanto dinámicos como repetidos. También consideramos la relación con teoría de juegos clásica, desde la perspectiva de programación dinámica, usando la derivación estándar de la ecuación de Hamilton–Jacobi–Bellman (HJB).

Para el caso de los juegos repetidos, mostramos la aplicación a uno de los juegos presentados por Jehiel [Je95]; mientras que para mostrar la aplicación en los juegos dinámicos, lo hacemos con la instancia del juego de dominó presentada en [EH13] y en [FEH14]. Planteamos una forma genérica de HJB aplicable a ambos casos. para posteriormente encontrar una posible solución en el caso continuo.

Además, presentamos un ejemplo interesante relacionado con contar caminos posibles para llegar desde una esquina hasta la opuesta en una malla rectangular, derivando su ecuación de HJB. Posteriormente, llevamos al límite el tamaño de la malla, obteniendo así una ecuación diferencial parcial del tipo HJB continuo. Lo anterior permite obtener estimaciones rigurosas sobre el crecimiento de estrategias óptimas. Pretendemos mostrar cómo ideas diferentes que provienen de áreas diversas aplicadas a juegos pueden combinarse para obtener una perspectiva más rica sobre los problemas asociados y sus soluciones.

5.2. Ecuación HJB para juegos alternados

Con el fin de obtener una ecuación general de HJB para esta clase de juegos, consideramos el número de opciones posibles que tiene el jugador en su turno actual:

- ♢ El mejor caso es cuando el jugador sólo tiene una opción disponible (aún si esa opción es no realizar movimiento alguno). En este caso, no debe realizar cómputo para elegir su jugada. Por ejemplo en la figura 5.1, cuando el jugador 2 debe elegir por primera vez (nodos cuadrados en el nivel 2), el mejor caso se encuentra en el nodo número 2: solo tiene una opción.

- ♢ El peor caso sucede cuando un jugador debe elegir entre todas las ramas posibles del juego: debe calcular la mejor opción para el turno actual. Cuando tiene muchas opciones y no le es posible observar el fin del juego, utiliza una función heurística para evaluar la mejor entre estas opciones. Por ejemplo, en la misma figura 5.1 cuando el jugador 2 debe elegir por primera vez, el peor caso se encuentra en el nodo número 3: debe elegir una de todas sus fichas disponibles sin alcanzar a ver el final del juego.

Una buena forma de derivar la ecuación de HJB es utilizar el promedio de opciones posibles para un jugador y con esto determinar la profundidad promedio del árbol de juego. En general, si un jugador tiene n posibles opciones, la profundidad será $2n$, dado que el otro jugador tiene también n posibles opciones.

Ahora podemos proceder a obtener la ecuación. Un primer enfoque es maximizar la suma de las ganancias obtenidas por el jugador analizado (o minimizar sus pérdidas).

Sea G_t la ganancia esperada en el tiempo actual y k_t el número de pasos que el jugador puede prever. Entonces podemos escribir una ecuación en diferencias en la que la ganancia actual $J(n)$ depende de los últimos k_t pasos $J(n-k_t)$ como:

$$J(n) = \sum_{\forall t \in [0,\dots,n]} max\left(G_t \cdot J(n-k_t)\right)$$

con:

$$J(0) = 0$$

5.2.1. Ejemplos

Ahora como aplicación práctica de la ecuación de HJB obtenida en el punto anterior, la usamos en casos particulares de juegos dinámicos y repetidos.

5.2.1.1. Juegos dinámicos alternados

Nos enfocamos en una instancia particular del juego *draw*, que utiliza fichas de dominó. Esta instancia está descrita en la sección 3.2.1.2 de este trabajo. El árbol de juego asociado se muestra en la figura 5.1 (sin las cajas de ese análisis).

En este juego, el peor caso es aquel en el que el jugador puede jugar todas sus n fichas en alguno de los dos extremos. Entonces la aproximación de la profundidad promedio es $2n$. Cuando un jugador no puede ver el final del juego, utiliza la heurística de deshacerse de la ficha más pesada.

De acuerdo a lo anterior, y definiendo W_t como el peso de una ficha, la ecuación de HJB para este juego es:

$$J(n) = \sum_{\forall t \in [0,\ldots,n]} max\left(W_t \cdot J(n - k_t)\right)$$

y

$$J(0) = 0$$

Como modelamos y analizamos este juego desde el punto de vista del jugador 1, el árbol del juego será *colapsado* a los nodos en los que ese jugador elige su acción. Aplicando esta idea, se obtiene el árbol de la figura 5.2.

En este árbol, los estados intermedios son aquellos en los que el turno es para el jugador 1. Las transiciones corresponden a una ficha elegida por el jugador y los extremos libres después de la ficha tirada por el jugador 2. Los números cercanos a cada estado representan la suma de puntos que mantiene el jugador 1 después de la última ficha usada. Es importante recordar que la función utilizada es la guía simple de deshacerse de la ficha más pesada: mientras no pueda ver el final del juego, el objetivo principal es conservar la menor cantidad de puntos en la jugada actual. Los estados finales están marcados con doble línea: cuando gana el jugador 2, la ganancia del jugador es 0 o negativa y los estados correspondientes son cuadrados.

Con esta configuración estamos listos para analizar el comportamiento del jugador 1 desde el punto de vista del resultado anterior. Supongamos que puede prever un nivel completo en el árbol colapsado (esto podría no ser el caso en el árbol completo, dado que en realidad serían 2 niveles). Tal como lo indica la ecuación HJB obtenida, el jugador desea maximizar la suma de puntos de las fichas que ha tirado hasta el tiempo actual; es decir, desea alcanzar el estado en el que conserve la menor cantidad de puntos. Para nuestro ejemplo concreto, en el primer turno elegirá la ficha $(2, 2)$ para llegar a un estado entre $(6, 7)$. Es

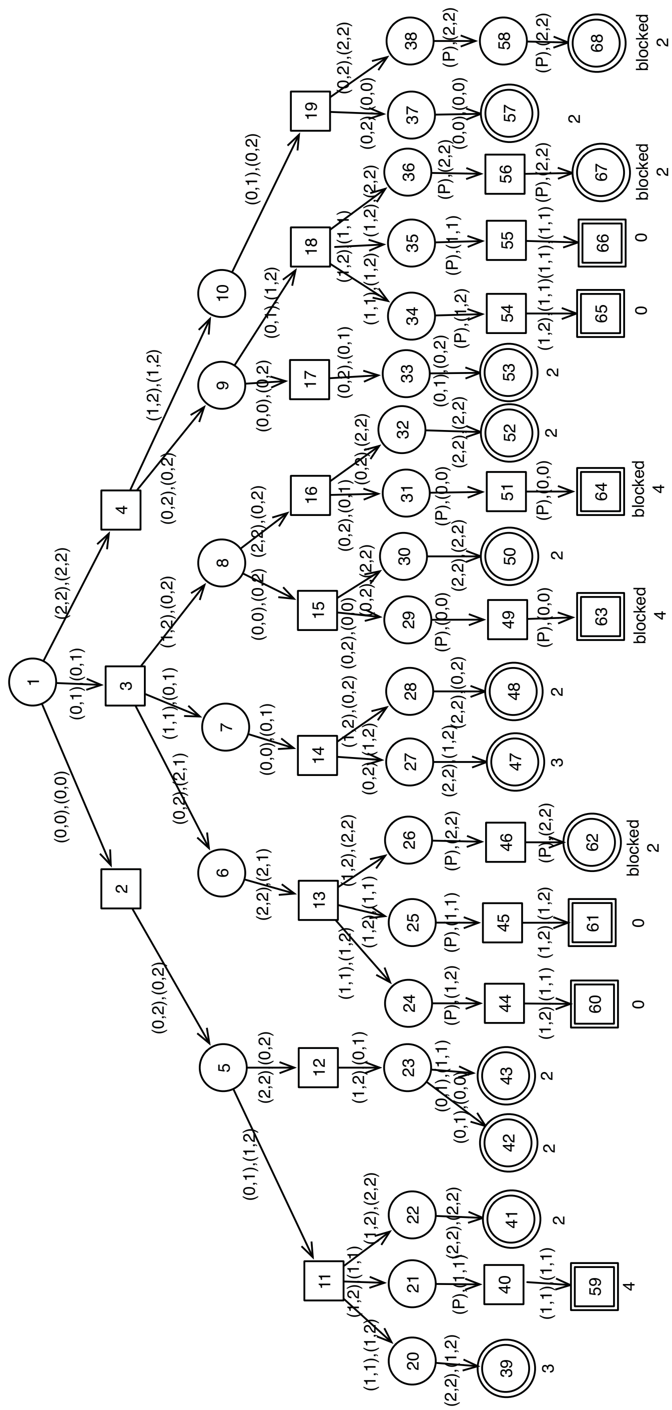

Figura 5.1: Árbol para la partida de dominó

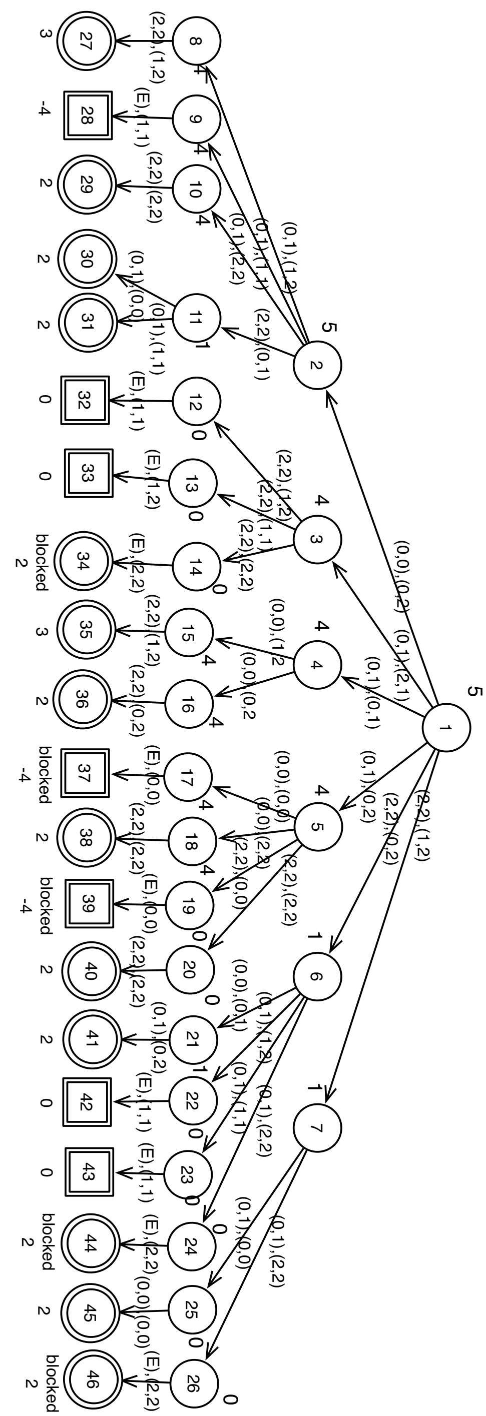

Figura 5.2: Árbol de juego del dominó desde la perspectiva del jugador 1

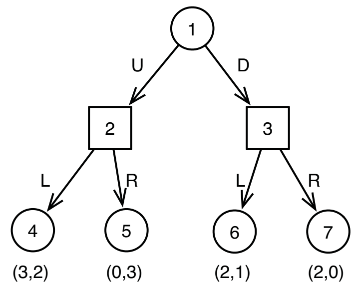

Figura 5.3: Árbol del juego alternado repetido

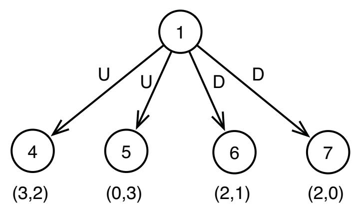

Figura 5.4: Árbol colapsado del juego alternado repetido

importante notar que no puede determinarse de antemano el estado al que llegará después de ese movimiento, ya que debe esperar el movimiento elegido por el oponente para saber el estado *real*.

Al final de la partida, podrá realizar el cómputo de la suma de puntos usados. Este será el resultado de aplicar la ecuación HJB en cada estado visitado y su ganancia *final* será la suma de los puntos restantes (a favor o en contra) tal como se acostumbra en el dominó. Este análisis lleva al mismo resultado que el obtenido en [EH13].

5.2.1.2. Juegos alternados repetidos

Ahora presentamos un ejemplo para la clase de juegos presentados en [Je95]. Aplicando la misma idea de colapsar el árbol del juego para analizar la dinámica del juego desde la perspectiva de un solo jugador. La figura 5.3 muestra el árbol completo, mientras que en la figura 5.4 se presenta el árbol colapsado al punto de vista del jugador 1.

De acuerdo con el análisis realizado por Jehiel, el mejor estado global es en el que las ganancias de los jugadores son (3.2) (estado 4). Esto implica que cuando el juego se repite de forma indefinida, la secuencia de acciones que entrega las mejores ganancias globales para ambos jugadores, es U, L.

Sin embargo, en el caso de una partida simple, desde la perspectiva del jugador 2, debería elegir R cuando el jugador 1 eligió U, dado que le entrega mejor ganancia $(0, 3)$ (y la peor para el jugador que analizamos). Posteriormente, el jugador 1 *aprende* que su oponente prefiere L después de haber elegido D donde las ganancias son $(2, 1)$.

Con este razonamiento, podemos aplicar la ecuación HJB obtenida a este juego. Como el jugador 1 puede prever el siguiente nivel del juego colapsado, *sabe* que el jugador 2 prefiere al estado 4 sobre el 5; por tal motivo elegirá la acción D que le garantiza una ganancia de 2, maximizando así su ganancia global.

Sin embargo, recordando los argumentos de la sección 3.2.2.1 existen tres formas de obtener las mismas ganancias que presenta Jehiel para ambos jugadores. A grandes rasgos, las dos primeras suponen que el jugador 2 tiene o puede obtener de alguna forma mayor información que el jugador 1, de forma que pueda saber de antemano que en el juego repetido le resulta más conveniente elegir L cuando el jugador 1 juega U que en cualquier otro caso de todo el árbol del juego.

La tercera forma de obtener el mismo resultado es más interesante, dado que no supone ningún tipo de información adicional para ninguno de los jugadores, sólo requiere que ambos puedan actualizar sus bases de conocimientos. Si ambos jugadores *ensayan* todas sus posibilidades y agregan las entradas correspondientes en su base de conocimientos, después de una ronda completa, tendrán la información necesaria para obtener la mejor secuencia global de acciones U, L: con ganancias $(3, 2)$. Insistimos en que esta situación es realista debido a que el juego se repite indefinidamente.

Al aplicar la ecuación HJB con cualquiera de los tres resultados anteriores, la máxima ganancia para nuestro jugador analizado se obtendrá utilizando la transición U.

5.2.1.3. Contando caminos

Ahora presentamos otro juego que puede analizarse desde el punto de vista de la teoría de juegos, racionalidad acotada y programación dinámica.

La descripción de este juego es sencilla. Se tiene una malla rectangular simulando un mapa. El juego comienza en una esquina, digamos la superior izquierda, y el objetivo es obtener el máximo número de caminos posibles para alcanzar alguna otra posición en la malla. En cada tiempo, el agente puede elegir moverse hacia la derecha o abajo. Una instancia de este juego se puede observar en la

i, j	0	1	2	3
0	1	1	1	1
1	1	2	3	4
2	1	3	6	10
3	1	4	10	20

Figura 5.5: Ejemplo del juego contando caminos

figura 5.5.

Dada esta definición del juego, la función de ganancia puede escribirse como:

$$paths\,(i, j) = paths\,(i-1, j) + paths\,(i, j-1)$$

con:

$$paths\,(i, 0) = 1 \text{ y } paths\,(0, j) = 1$$

Ahora, con el objeto de analizar este juego desde el punto de vista de la optimización acotada, obtenemos su ecuación de HJB asociada. Primero reescribimos la función original en una forma equivalente, pero más adecuada:

$$J\,(i, j) = J\,(i+1, j) + J\,(i, j+1)$$

con:

$$J\,(i, 0) = 1 \text{ y } J\,(0, j) = 1$$

Sustrayendo $J(i\Delta, j\Delta)$ en ambos lados de la ecuación y aplicando el teorema del valor medio, tenemos:

$$0 = \Delta \frac{\partial J}{\partial x}(i\Delta, j\Delta) + o(\Delta) J(i\Delta, (j+1)\Delta)$$

$$0 = \Delta \frac{\partial J}{\partial y}(i\Delta, j\Delta) + o(\Delta) J((i+1)\Delta, j\Delta)$$

Sumando los dos expresiones anteriores, aplicando nuevamente el teorema del valor medio y dividiendo entre 2Δ, obtenemos:

$$0 = \frac{\partial J}{\partial x} + \frac{\partial J}{\partial y} + \frac{1}{\Delta} J(i\Delta, j\Delta) + \frac{1}{\Delta} o(\Delta)$$

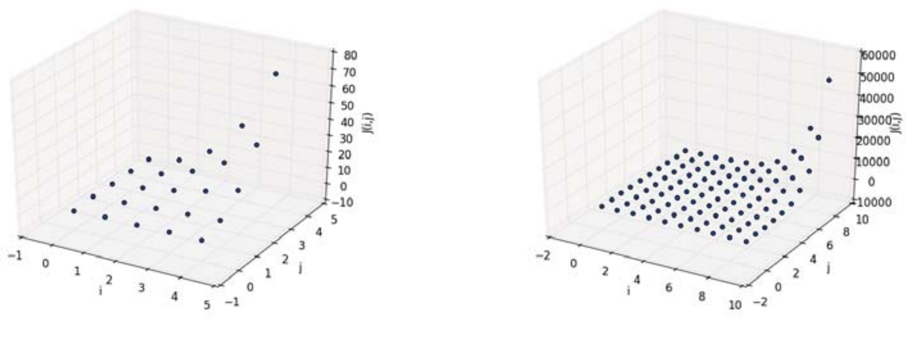

Figura 5.6: Dos mallas cuadradas para el juego de caminos: 5×5 y 10×10

Suponiendo que cuando $\Delta \to 0$, $J(i\Delta, j\Delta)$ puede expresarse como:

$$\Delta J(i/j) + o(\Delta)$$

Esto es válido si, por ejemplo, J depende de la distancia en forma superlineal. Si este es el caso, es posible obtener el límite cuando $\Delta \to 0$:

$$\nabla J + J = 0$$

donde $\nabla J = \dfrac{\partial J}{\partial x} + \dfrac{\partial J}{\partial y}$ y es justamente la ecuación de HJB que estábamos buscando.

Esta ecuación puede interpretarse como una estrategia para un agente que se encuentra decidiendo en este juego: la elección óptima es buscar incrementar los índices tanto a i como a j y no maximizar sólo uno de ellos. Esto se relaciona directamente con racionalidad acotada puesto que, en principio, no es necesario conocer toda la malla sino solamente los valores actuales de ambos índices y verificar que se pueda seguir aumentando uno o ambos, con el fin de tomar una elección que pueda considerarse como racional.

En la figura 5.6 se presentan dos instancias del juego con malla cuadrada en el eje horizontal, mientras que la figura 5.7 muestra también dos instancias, pero con malla rectangular. El eje vertical es el resultado de contar los caminos sobre cada una de dichas mallas. En ambas figuras es notorio que contar los caminos entrega una salida con crecimiento exponencial al incrementar ambos índices.

Lo anterior es acorde con el resultado esperado de la solución de la ecuación de HJB obtenida previamente. Al resolver la forma continua se obtendrá la solución explícita para dicha ecuación, esta solución confirma el crecimiento exponencial.

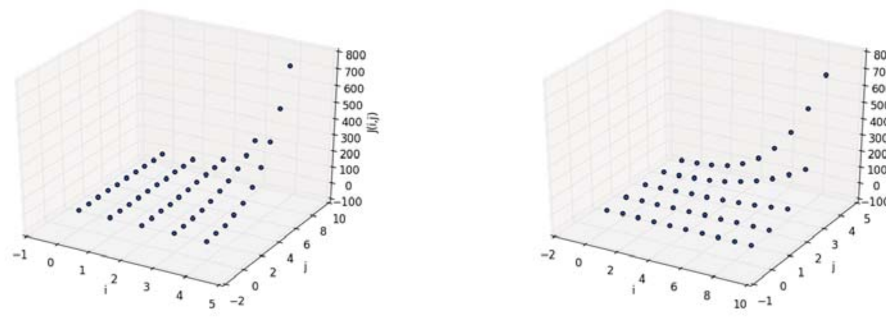

Figura 5.7: Dos mallas rectangulares para el juego de caminos: 5×10 y 10×5

5.3. Versión continua

Con el fin de calcular soluciones explícitas para los juegos presentados en la sección anterior, ahora desarrollamos modelos continuos para ellos. Esto se puede conseguir permitiendo que n crezca indefinidamente. Para un juego alternado repetido es totalmente natural, dado que una de sus características es justamente que el juego se repite de forma *indefinida*. En un juego dinámico, como el dominó, se puede incrementar el número de fichas a repartir, tal como sucede en la versión cubana; o también es posible hacer repeticiones indefinidas del juego. Finalmente, para el ejemplo de contar caminos, podemos pensar en una malla en la que se permite que ambos índices sigan creciendo cuando se requiera.

5.3.1. Juegos alternados

Ahora desarrollamos una función que ayudará en cálculo de la ganancia en juegos alternados tanto repetidos como dinámicos cuando el número de pasos crece indefinidamente. Para poder obtener dicha función haremos uso de la media aritmética en varios pasos del juego.

Consideremos un intervalo de tiempo genérico, digamos $[a, b]$ y

$$f(t) = \max\left(G_t \cdot J\left(n - k_t\right)\right)$$

Primero, necesitamos dividir el intervalo en n subintervalos de tamaño $\Delta t = \dfrac{b-a}{n}$, de forma que la media pueda calcularse como:

$$\frac{1}{n}\sum_{k=0}^{n} f\left(t\right)$$

Por tanto:

$$\frac{1}{b-a}\sum_{k=0}^{n} f\left(t\right)\frac{b-a}{n}$$

Finalmente:

$$\frac{1}{b-a}\sum_{k=0}^{n} f\left(t\right)\Delta t$$

La expresión anterior indica que la ganancia promedio de n pasos del juego puede ser calculada multiplicando $\frac{1}{b-a}$ por la suma de Riemann de esos n pasos.

Ahora, si permitimos que el número de subintervalos es decir, el número de pasos crezca indefinidamente, obtenemos:

$$\frac{1}{b-a}\lim_{n\to\infty}\sum_{k=0}^{n} f\left(t\right)\Delta t$$

De lo anterior, se sigue que:

$$\frac{1}{b-a}\int_a^b f\left(t\right)dt$$

Consecuentemente, si permitimos juegos con gran cantidad de pasos, podremos determinar la ganancia esperada como la media de las ganancias en todos los pasos.

5.3.2. Contando caminos

Ahora desarrollamos la función que realizará el cómputo de caminos cuando la malla crece de manera indefinida. Recordando la ecuación de HJB asociada a este juego:

$$\nabla J + J = 0$$

Y sus dos casos base:

$$J(i,0) = 1 \text{ y } J(0,j) = 1$$

Por consiguiente, el sistema característico asociado es:

$$\frac{dx}{ds} = 1$$

$$\frac{dy}{ds} = 1$$

$$\frac{dJ}{ds} = -J$$

donde s es el parámetro sobre la curva característica. En este caso una línea recta sobre la diagonal principal y que está relacionada con el número de movimientos. De hecho, es su versión continua.

La última ecuación es una ecuación diferencial ordinaria de primer orden. Al resolverla, vemos que $J(s)$ es de la forma:

$$J(s) = Ce^{-s}$$

En otras palabras, se espera que el número de caminos crezca exponencialmente cuando s decrece, que coincide y confirma el resultado obtenido en el caso discreto. C es un parámetro de normalización determinado por las condiciones iniciales. Aún más, $C = J(0)$ está completamente relacionado con el tamaño de la malla.

La solución mencionada es la formulación de una ecuación diferencial ordinaria de primer orden en términos de un sistema de ecuaciones diferenciales. Es un enfoque estándar y puede consultarse en la literatura, por ejemplo en [Ev10].

5.4. Conclusiones

En la sección presente obtuvimos la ecuación HJB para estudiar el comportamiento de juegos desde la perspectiva de programación dinámica y mostramos su aplicación con algunos ejemplos. Además, los analizamos tanto desde el punto de vista discreto como su correspondiente límite continuo. Pensamos que este enfoque puede resultar útil cuando el procedimiento lo justifique y parece razonable que varios juegos puedan estudiarse utilizando estas ideas.

Es importante notar que esta solución es equivalente al equilibrio perfecto por subjuegos (eps), dado que ambos calculan las mismas ganancias cuando se aplican bajo las mismas condiciones. Además recordemos que el eps es un refinamiento del equilibrio de Nash.

También vale la pena recalcar que trabajos previos muestran soluciones para juegos muy particulares y que no existe un marco conceptual genérico. Particularmente, para juegos complejos no existe un entendimiento real de ellos. Este trabajo es un intento orientado en la búsqueda de dicho marco ya que permite obtener estimaciones asintóticas, tal como se muestra en el ejemplo del conteo de caminos.

Finalmente, otros juegos discutidos en [Sm07] que son objeto de investigación, podrían ser considerados para aplicar nuestro modelo, formalizándolos para obtener soluciones explícitas para sus versiones *continuas*.

Capítulo 6

Conclusiones generales

La guía principal de este trabajo de tesis es la aplicación de racionalidad acotada a la teoría de juegos.

La relación existente entre estos dos tópicos no es nueva [RS93, Je95, Je-98], pero sí ha tomado nuevo auge en años recientes [Mi12, EH13, BS14, FEH14, EPH17], debido a que se desea modelar las interacciones entre agentes en los que no es realista pensar que posee racionalidad perfecta, que, en términos de teoría de juegos, puede verse como la habilidad de tener disponible todo el árbol asociado al juego en cuestión y, en términos computacionales, se traduce como una capacidad de cómputo y memoria infinitas.

El objetivo primordial es responder a la pregunta ¿Qué puede ser considerada una elección racional dadas ciertas limitaciones en las capacidades de un agente? A manera de guía (heurística), esta pregunta se trata de responder mediante tres propuestas realizadas desde tres diferentes enfoques.

En el capítulo 3 presentamos el primero de dichos enfoques, comienzamos con el cálculo de posibles reparticiones iniciales para el juego del dominó, en función del número de fichas disponibles para ser repartidas. Es notorio que el número de reparticiones resulta extremadamente alto para suponer que se disponga de todas las historias posibles, además en esa cuenta aún falta los árboles de cada juego particular: solamente son las posibles reparticiones iniciales.

Posteriormente, extiendemos el modelo desarrollado por Jehiel [Je95] para juegos alternados repetidos, de forma que el modelo extendido se aplica a esta clase de juegos, pero también a los juegos dinámicos de movimientos alternados. Este modelo extendido sigue una metodología estándar en teoría de juegos: se establece la clase de juegos a la cual se aplica, y se define el concepto de solución asociado al modelo.

Por último, mostramos la aplicación del modelo extendido, tanto a un juego dinámico como a un juego repetido presentado en el trabajo de Jehiel. Esto último con el objeto de mostrar que el modelo es realmente una extensión y puede aplicarse a la clase de juegos anterior.

En el capítulo 4 mostramos el segundo enfoque, en el cual desarrollo un lenguaje lógico denominado lógica para racionalidad acotada: $\mathcal{BRL}$ (*bounded rationality logic*), el cual extiende a la lógica proposicional combinando dos ideas ya presentes en la literatura: una semántica de vecindarios [PMS07] y una lógica de preferencias [BS11]. Pero esta combinación es novedosa ya que tanto la semántica de vecindarios como las preferencias se comportan de forma distinta a las anteriormente presentes.

En cuanto a la primera, en ambos el estado actual está asociado a un vecindario particular. En *topologic*, el objetivo del agente es determinar el estado real de entre todo ese vecindario y, por tanto, un conjunto menor le entregará mayor información; es decir, en este modelo, el vecindario representa un "margen de error". En cambio, en nuestra propuesta el agente busca elegir la mejor ganancia entre las opciones posibles (el vecindario); por tanto, un conjunto más grande entrega mayor información. Por lo demás, el tratamiento de los vecindarios es similar.

En lo referente a las preferencias, en lugar de un operador B, se añade una variable proposicional best, la cual es verdadera sólo en los nodos que son *maximales* para el agente. Esta variable permitirá a los agentes, no solamente razonar acerca del estado del juego en esos estados, sino también planear una estrategia para llegar a dichos nodos preferidos.

Aunado a lo anterior, presentamos otra novedad clave para modelar heurísticas acotadas y es debida a la forma en la que interactúan entre los vecindarios y las preferencias: como el agente sólo tiene acceso a estados dentro del vecindario actual, sólo puede maximizar sus ganancias dentro de dicho vecindario. Es decir, best no sólo es cierta en estados *maximales*, sino también en aquellos que son *maximales dentro del vecindario actual*.

En este enfoque hay varios puntos que pueden extenderse. Si bien el lenguaje permite expresar fórmulas que toman en cuenta restricciones computacionales genéricas (es decir, previsión, tiempo, espacio, etc.), es deseable obtener un cálculo que sea completo y consistente, así como un procedimiento efectivo para realizar verificación de modelos para este lenguaje. Además sería interesante modelar razonamiento entre varios jugadores que podría llevar a un modelo mucho más completo sobre la interacción estratégica de los agentes.

Finalmente, en el capítulo 5 desarrollamos el tercer enfoque, el cual es una aplicación de optimización acotada a la teoría de juegos utilizando programación dinámica. Revisamos la relación entre juegos en versión de tiempo discreto (representación de espacio de estados) y la versión *continua* de los mismos al llevar la primera al límite para obtener la segunda.

La metodología para realizar el análisis es la siguiente; utilizamos la ecuación de Hamilton–Jacobi–Bellman (HJB) aplicada al caso discreto de los juegos, obteniendo así una forma de determinar las ganancias en juegos discretos. Una vez obtenida la ecuación HJB discreta, se procede a obtención de soluciones explícitas para la versión *continua* del juego en cuestión con ayuda de una ecuación diferencial parcial asociada a dicho juego. De esta forma, se pueden calcular estrategias *óptimas acotadas*.

Aunado a lo anterior, este último enfoque entrega una forma alternativa para calcular equilibrio perfecto por subjuegos, que es un concepto de solución muy importante en la teoría clásica de juegos.

Entre los trabajos a futuro que podrían desarrollarse está aplicar esta metodología, que pretende ser un marco genérico a otros juegos que son objeto de investigación, formalizándolos con ecuaciones de HJB para después obtener soluciones explícitas obteniendo su límite. Algunos de los juegos que pueden ser tomados en cuenta se presentan en [Sm07].

Bibliografía

[NM44] von Neumann, John and Morgenstern, Oskar. *Theory of games and Economic Behavior*. Princeton University Press, 1944.

[Os04] Osborne, Martin. *An introduction to game theory*. Oxford University Press, New York, 2004.

[BD13] Bonanno, Giacomo; Dégremont, Cédric (2013): *Logic and game theory*, Working Papers, University of California, Department of Economics, No. 13-4.

[BS14] Alexandru Baltag and Sonja Smets: *Johan van Benthem on Logic and Information Dynamics*, Springer International Publishing Switzerland 2014.

[FHMV95] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, Moshe Y. Vardi. *Reasoning About Knowledge*. MIT Press, Cambridge, 1995.

[PMS07] Rohit Parikh, Lawrence S. Moss, Chris Steinsvold. *Topology and epistemic logic*. In Marco Aiello, Ian Pratt-Hartmann, and Johan van Benthem, editors, Handbook of Spatial Logics, pages 299–341. Springer, 2007.

[MT44] McKinsey, J. C. C., A. Tarski. *The algebra of topology*. Annals of Mathematics, 45:141–191, 1944.

[Si55] Herbert A. Simon, *A Behavioral Model of Rational Choice*, The Quarterly Journal of Economics, Vol. 69, No. 1. (Feb., 1955), pp. 99–118.

[Ru98] Rubinstein, A. *Modeling Bounded Rationality*. MIT Press, 1998.

[Ch81] Christopher Cherniak. *Minimal Rationality*. Mind, New Series, Vol. 90, No. 358 (Apr., 1981), pp. 161–183. Oxford University Press on behalf of the Mind Association.

[AK09] Sergei Artemov, Roman Kuznets, *Logical Omniscience as a Computational Complexity Problem*, Proceedings of TARK XII (2009), Stanford, California.

[PYa94] Papadimitriou, C. H., y Yannakakis, M. 1994. On complexity as bounded rationality (extended abstract). Procedings of the twenty-sixth annual ACM symposium on Theory of computing 726–733.

[RS93] Russell, S., and Subramanian, D. 1993. Provably bounded-optimal agents. Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence.

[DF09] Nikhil R Devanur y Lance Fortnow, *A Computational Theory of Awareness and Decision Making*, Proceedings of TARK XII (2009), Stanford, California, 99–107.

[BT05] Jan Brinkhuis y Vladimir Tikhomirov. *Optimization: Insights and Applications*. Princeton Series in Applied Mathematics. Princeton University Press, 2005, pp. 441–502.

[Je95] Jehiel, P. *Limited Horizon Forecast in Repeated Alternate Games*. Journal of Economic Theory 67. 1995, 497–519.

[EH13] Espinosa-Avila, Eduardo and Hernández-Quiroz, Francisco. *Bounded Rationality in a Dynamic Alternate Game*. Theoretical Aspects of Rationality and Knowledge (TARK, actas del congreso). 2013 pp. 71–77.

[Al94] Allis, L. V., *Searching for Solutions in Games and Articial Intelligence*, Phd Thesis, 1994, University of Limburg, http://fragrieu.free.fr/SearchingForSolutions.pdf.

[Te01] Tejeiro-Arias, G. A. "How to play latin partnership dominoes" Hats Off Books, 2001.

[Je-98] Jehiel, P. "Learning to play limited forecast equilibria." Games and Economic Behavior 22. 1998, 274–298.

[Ku-97] Kuhn, H. W. "Extensive Games and the Problem of Information", pp 46–68 in Classics in Game Theory, Princeton University Press, 1997.

[Mi12] Mirrokni, V., Thain, N. and Vetta, A. *A Theoretical Examination of Practical Game Playing: Lookahead Search*, Proceedings of the 5th International Symposium, SAGT 2012, Barcelona, Spain, October 2012.

[Sc01] Schwarz, Eric. *An instance of a complete communication cycle within cooperative games: the case of domino*. 2001.

[FEH14] D. Fernández-Duque, E. Espinosa-Avila and F. Hernández-Quiroz *A logic for bounded rationality*, Proceeding of the short presentations of Advances in Modal Logics, AiML'14, 2014, pp. 34–38.

[VH05] Velázquez-Quesada R. Fernando and Hernández-Quiroz, Francisco. *A Logical Language for Dominoes*. Logic for Programming in Artificial Intelligence and Reasoning (LPAR 2005), Short Papers. 2005, pp. 38–42.

[VH06] Velázquez-Quesada R. Fernando and Hernández-Quiroz, Francisco. *Some Semantics for a Logical Language for the Game of Dominoes*. Artificial Intelligence and Applications (AIA 2006), Acta Press. 2006, pp. 293–298.

[BS11] Baltag, Alexandru and Smets, Sonja. *Keep Changing Your Beliefs, Aiming for the Truth*. Erkenntnis, Springer Netherlands, Vol. 75, Num. 2. 2011, pp. 255–270.

[EPH17] Espinosa-Avila, Eduardo, Padilla Longoria Pablo and Hernández-Quiroz, Francisco. *Game theory and dynamic programming in alternate games*, Journal of Dynamics and Games (JDG), American Institute of Mathematical Sciences, Volume 4, Number 3, July 2017, pp. 205–216.

[MB97] T. A. Marsland and Y. Björnsson, *From minimax to manhattan*, Proceedings of Deep Blue Versus Kasparov: The Significance for Artificial Intelligence, Collected Papers from the 1997. AAAI Workshop, 1997, 31–36.

[DB] IBM DeepBlue, *What are the differences between last year's rendition of Deep Blue and this year's model*?, IBM DeepBlue faq, https://www.research.ibm.com/deepblue/meet/html/d.3.3a.html. IBM. Retrieved May 12, 2017.

[DQA] IBM DeepQA, *What kind of technology is Watson based on*?, DeepQA faq, https://www.research.ibm.com/deepqa/faq.shtml. IBM. Retrieved May 12, 2017.

[Ev10] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, Second Edition, Graduate Studies in Mathematics, Volume: 19, Providence, RI, 2010.

[Sm07] D. K. Smith, *Dynamic programming and board games: A survey*, European Journal of Operational Research, 2007, pp. 1299–1318.