
PROGRAMACIÓN MATEMÁTICA PARA ASESORAR A UN ENTRENADOR DE FÚTBOL: UN JUEGO DE FANTASÍA COMO CASO DE ESTUDIO

F. BONOMO *
G. DURÁN **
J. MARENCO ***

Resumen

La película “Moneyball” dejó el mensaje de que la matemática podía ser de gran ayuda para tomar decisiones en el campo del deporte. Un club de béisbol de los Estados Unidos utilizaba herramientas matemáticas para mejorar su rendimiento deportivo. En este trabajo enfocamos este mismo problema y utilizamos como caso de estudio un juego de fantasía organizado por un diario argentino. Presentamos modelos de programación matemática para el desarrollo de un director técnico virtual de fútbol. Diseñamos modelos a priori para tener equipos más robustos para el juego, y a posteriori, para conocer cuál hubiera sido la configuración óptima de equipos durante todo el campeonato, una vez conocidos los resultados. Nuestro jugador virtual a priori se posiciona habitualmente entre los mejores de la competencia. La expansión de este

*Departamento de Computación, FCEN, UBA, Argentina. CONICET, Argentina.

**Instituto de Cálculo y Departamento de Matemática, FCEN, UBA, Argentina. CONICET, Argentina. Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas. Universidad de Chile, Santiago, Chile.

***Instituto de Ciencias, Universidad Nacional de General Sarmiento, Argentina. Departamento de Computación, FCEN, UBA, Argentina

tipo de desarrollos debería ser útil para asesorar en la toma de decisiones a entrenadores o gerentes de equipos en deportes reales.

Palabras Clave: Fútbol, Juego de Fantasía, Programación Matemática.

1. Introducción

El “Gran DT” es un juego de fantasía creado por un diario argentino en los '90 y que retomó su acción en 2008. El juego ha contado desde su reinicio con más de 1 millón de participantes en cada una de sus ediciones. Consiste en que cada uno de los participantes toma el rol de un director técnico de fútbol a lo largo del campeonato argentino de Primera División, con el objetivo de formar el mejor equipo posible, combinando a jugadores de los diferentes equipos que participan del campeonato. Cada participante suma o resta puntos por la actuación semanal de sus jugadores, existiendo datos objetivos (goles convertidos, valla invicta, tarjetas amarillas, tarjetas rojas) y datos subjetivos aportados por el diario (puntaje sobre la actuación de cada jugador en el partido, figura de la cancha).

El reglamento del juego exige que los equipos virtuales cumplan una serie de restricciones (presupuesto máximo, número de jugadores por puesto, número de jugadores por equipo), y su constitución es dinámica dado que fecha a fecha se pueden hacer una cierta cantidad de modificaciones en el equipo.

El “Gran DT” se inspiró en el “Fantacalcio” [10], juego de fantasía italiano dedicado a su Serie A de fútbol. El “Fantacalcio” fue inventado por Riccardo Albini [12] a fines de los '80 y es también organizado por uno de los principales diarios de Italia. Según relata Albini, se inspiró para la creación del “Fantacalcio” en el juego de fantasía de la Liga de Béisbol de los Estados Unidos [14]. La reglamentación del juego italiano es muy similar a la que terminó adoptando el “Gran DT” en la Argentina. Actualmente existen también otros juegos de fantasía en diferentes lugares del mundo, como el basado en la Premier League inglesa de fútbol [9], o el de la NBA o el fútbol americano en los Estados Unidos de América [15, 16]. Existen también otros juegos similares con mucha aceptación en el público, pero sobre fútbol virtual [13, 17]. El uso de juegos de fanta-

sía para incentivar la enseñanza de la matemática en la escuela ha sido ampliamente estudiado en los Estados Unidos [21].

En este trabajo presentamos un modelo de programación matemática diseñado a priori, que llamamos prescriptivo, para tener un equipo más robusto para el juego (teniendo en cuenta las actuaciones de los jugadores en torneos anteriores y en el mismo torneo, más ciertas características de las fechas a disputarse); y dos modelos diseñados a posteriori, que llamamos descriptivos, para conocer cuál hubiera sido el equipo óptimo a lo largo del torneo, una vez conocidos los resultados.

Los modelos descriptivos consiguen mostrar el equipo ideal que se debió haber armado fecha a fecha, a lo largo de todo el torneo, para obtener el mayor puntaje posible, cumpliendo con las restricciones del juego.

Testeamos el modelo prescriptivo haciendo competir a nuestro jugador virtual en el juego. Los resultados marcan que el mismo califica habitualmente en el 3 % mejor del juego, llegando en alguna oportunidad a estar en el mejor 1 por 1000 del certamen. Si tomamos los 6 campeonatos en los que el modelo participó del juego como un único torneo, nuestro jugador virtual se posiciona en el mejor 2 por 1000 de la competición.

La literatura de la disciplina conocida como “Sports Analytics” (SA), que engloba desarrollos matemáticos y computacionales orientados a problemas del deporte, ha aumentado notablemente en los últimos años. Importantes revistas de investigación operativa, matemática aplicada, estadística, gestión y economía han publicado numerosos artículos en estos temas [1]. En lo que hace a métodos de programación matemática aplicados a cuestiones deportivas, el mayor desarrollo se ha dado en problemas de confección de fixtures para diferentes competencias. Esta subdisciplina del SA se la conoce como “Sports Scheduling” (SS). Excelentes resúmenes sobre el estado del arte en SS y análisis de distintos problemas abiertos en el tema se encuentran en [2, 7]. Un análisis de las principales instancias para distintos deportes estudiadas en la literatura aparece en [6].

A pesar de que en los últimos años ha surgido software que asiste a entrenadores en diferentes deportes en la tarea de recopilar, almacenar y consultar datos, hasta lo mejor de nuestro conocimiento, no existen aplicaciones de algoritmos de programación matemática para asesoramiento a directores técnicos, reales o virtuales, del estilo de lo que se presenta en este trabajo. Lo más cercano podría ser el “Fantarobot” [11], una

funcionalidad opcional en la página web del “Fantacalcio”, que elige el “mejor” equipo dentro del plantel de titulares y suplentes que un determinado jugador seleccionó, pero sólo considera la actuación promedio de cada jugador hasta ese momento, a fin de dar una recomendación. En lo que hace a aplicación de técnicas matemáticas para asesorar en la toma de decisiones en deportes reales, el caso más conocido se ha dado en el béisbol de los Estados Unidos [3], donde se aplican fundamentalmente herramientas estadísticas.

Este trabajo está organizado de la siguiente manera. En la Sección 2 se describen en detalle las características del juego. En la Sección 3 se presentan dos modelos descriptivos, modelos de programación lineal entera que encuentran el equipo ideal a lo largo de todo el torneo, si se pudiera prepararlo en cada fecha conociendo los resultados de todo el campeonato. Se muestran allí los resultados alcanzados en distintos torneos. En la Sección 4 se presenta el modelo prescriptivo, que incluye un modelo de programación lineal entera que busca maximizar el puntaje del equipo, una vez que fueron estimados los puntajes de cada jugador en cada fecha. Se exhiben en esta sección resultados obtenidos por el jugador virtual en diferentes torneos. Por último, en la Sección 5 se presentan las conclusiones y el posible trabajo futuro.

2. Descripción del juego

El juego empieza en la fecha 4 o 5 del Torneo Argentino de Primera División (habitualmente en los torneos Clausura empieza en la fecha 5 y en los Apertura, en la 4). De este modo, el juego se desarrolla a través de 15 o 16 fechas (dado que la Primera División del fútbol argentino la disputan 20 equipos en un sistema de todos contra todos, a lo largo de 19 fechas).

Cada competidor participa del juego con su número de documento nacional y debe armar un equipo integrado por jugadores de la Primera División del fútbol argentino, compuesto por 11 titulares y 4 suplentes.

Dentro de los 11 titulares, se pueden armar 3 tácticas de juego diferentes: 1 arquero, 4 defensores, 4 volantes y 2 delanteros; 1 arquero, 4 defensores, 3 volantes y 3 delanteros; o 1 arquero, 3 defensores, 4 volantes y 3 delanteros. Los 4 suplentes son un arquero, un defensor, un volante

y un delantero. Los suplentes sólo son considerados para el juego en una determinada fecha si alguno de los titulares de su misma posición no juega, o juega menos de 20 minutos, y por lo tanto no recibe puntuación por su actuación de parte del diario.

Cada jugador tiene un valor monetario simbólico, que va de los \$300.000 (los que aún no debutaron en primera), hasta más de \$10.000.000 (los jugadores de mejor nivel de los torneos anteriores). Los equipos no pueden superar los 65 millones de pesos (este valor ha variado entre 60 y 70 millones a lo largo de los torneos). No puede haber más que 3 jugadores del mismo club en un equipo.

Cada jugador titular suma o resta puntos en cada fecha por criterios subjetivos (la nota que le pone el diario, o si es catalogado como la figura de la cancha) y objetivos (si convierte goles; si no le convierten, en el caso de arqueros o defensores; si es expulsado; si es amonestado). La nota que el diario le pone a cada jugador es un valor numérico entero entre 1 y 10. Por convertir un gol en jugada de campo, un arquero recibe 10 puntos extras; un defensor, 9 puntos; un volante, 6 puntos y un delantero, 4 puntos. Un gol de penal le aporta 3 puntos al jugador que lo convierte, cualquiera sea su posición en la cancha. La figura de la cancha, determinada por el diario, recibe 4 puntos extra. Un arquero al que no le hacen goles recibe 3 puntos extra, y un defensor que termina el partido con la valla invicta, recibe 2 puntos extra. Un arquero que recibe goles, resta tantos puntos como la cantidad de goles que le hicieron; mientras que un arquero que ataja un penal suma 4 puntos extra (puntos que le son descontados al jugador que falló el penal). Una tarjeta amarilla resta 2 puntos, y una tarjeta roja resta 4 puntos. El jugador que juega menos de 20 minutos se lo considera como que no participó de esa fecha, y si es titular en el equipo del juego, le deja su lugar al suplente de la misma posición. En caso que haya más de un titular de una misma posición que no participa de la fecha, el equipo en cuestión juega con menos de 11 jugadores.

Fecha a fecha se pueden cambiar titulares por suplentes de manera ilimitada y se pueden realizar hasta 4 transferencias, reemplazos donde se incorpora a un jugador que no pertenecía al equipo y se retira a uno que sí pertenecía, manteniendo el equipo las restricciones básicas en cuanto a su conformación y su presupuesto (este parámetro fue modificado, dado que en las primeras ediciones del juego se admitían hasta 3 transferencias

por fecha).

El juego fue presentado en los '90, época en que se realizaron 3 ediciones del mismo. En agosto de 2008 fue reiniciado, y desde entonces fue jugado de manera ininterrumpida todos los campeonatos de Primera División que se han disputado hasta la fecha. La máxima cantidad de participantes se dio en el primer semestre de 2009, con casi 2 millones (cerca del 5% de la población de Argentina), mientras que la más baja se dio en el primer semestre de 2012, con un poco más de 1 millón de participantes.

3. Modelos descriptivos

Los modelos “a posteriori”, o descriptivos, se ejecutan al finalizar el campeonato (también podrían ejecutarse después de cada fecha), y permiten encontrar una configuración de equipos óptima una vez que se conocen los resultados (es decir, toma como dato los puntajes obtenidos por cada jugador en cada fecha).

Desarrollamos dos modelos descriptivos. El primero arma un equipo fijo con 11 titulares, cumpliendo con las restricciones del juego, y no lo modifica a lo largo de todo el torneo. No considera a los jugadores suplentes para nada, sólo guarda \$1.200.000 para colocar 4 jugadores suplentes de \$300.000 cada uno, de modo que el equipo sea válido. Busca maximizar el puntaje a lo largo de todo el torneo considerando sólo a esos 11 jugadores. El segundo modelo arma lo que llamamos el “equipo perfecto”. Empieza por un equipo inicial en la primer fecha del juego, y nos dice fecha a fecha que intercambios entre titulares y suplentes debe hacerse, y que jugadores deben incorporarse al equipo, de modo de obtener el mayor puntaje final posible.

Formulamos ambos modelos a través de programación lineal entera en el que la función objetivo maximiza el puntaje total, y las restricciones garantizan que fecha a fecha se cumple con las condiciones del juego, en cuanto a conformación del equipo, número de transferencias permitidas y presupuesto máximo.

3.1. Equipo fijo de titulares: formulación matemática

Este modelo busca los 11 jugadores fijos que maximizan el puntaje total, cumpliendo las restricciones del juego. Llamamos E al conjunto de equipos, J al conjunto de jugadores y $P = \{\text{arquero, defensor, volante, delantero}\}$ a las posiciones dentro del campo de juego. Para formular el modelo, para cada jugador $j \in J$ introducimos la variable binaria x_j que vale 1 si y sólo si el jugador j se incluye en el equipo. Los parámetros del modelo incluyen los siguientes datos:

- Para cada jugador $j \in J$, el parámetro equipo $_j \in E$ representa el equipo al que pertenece el jugador j , el parámetro posicion $_j \in P$ especifica la posición del jugador j dentro de la cancha, el parámetro precio $_j \in \mathbb{R}_+$ representa el precio del jugador, y puntaje $_j \in \mathbb{Z}$ informa el puntaje acumulado total del jugador j a lo largo del campeonato.
- Para cada posición $p \in P$, los parámetros máx $_p$ y mín $_p$ especifican la cantidad mínima y máxima de jugadores en esa posición, de acuerdo con las tres estrategias de juego permitidas.

Con estas definiciones, el modelo se plantea del siguiente modo:

$$\begin{aligned} & \text{máx} && \sum_{j \in J} \text{puntaje}_j x_j \\ & \sum_{j \in J} \text{precio}_j x_j & \leq & 63,800,000 \end{aligned} \quad (1)$$

$$\sum_{j \in J: \text{equipo}_j = e} x_j \leq 3 \quad \forall e \in E \quad (2)$$

$$\sum_{j \in J} x_j = 11 \quad (3)$$

$$\sum_{j \in J: \text{posicion}_j = p} x_j \geq \text{mín}_p \quad \forall p \in P \quad (4)$$

$$\sum_{j \in J: \text{posicion}_j = p} x_j \leq \text{máx}_p \quad \forall p \in P \quad (5)$$

$$x_j \in \{0, 1\} \quad \forall j \in J \quad (6)$$

La función objetivo maximiza el puntaje total de los jugadores. Las restricciones (1) solicitan que el monto total desembolsado por los jugadores no exceda \$63.800.000 (lo cual corresponde al límite superior de \$65.000.000, descontando cuatro suplentes de \$300.000 cada uno, el menor valor posible). Las restricciones (2) impiden que haya más de tres jugadores de un mismo equipo. La restricción (3) pide que el equipo se componga de 11 jugadores titulares, mientras que las restricciones (4) y (5) piden que se respeten las cantidades mínima y máxima de jugadores por posición (notar que estas restricciones, más el hecho de pedir que el equipo está compuesto por 11 jugadores, garantizan que se elige alguna de las tres estrategias permitidas). Finalmente, las restricciones (6) especifican la naturaleza de las variables.

Es interesante analizar la estructura de este modelo. Las restricciones (1) son restricciones de tipo *knapsack*, una estructura que se encuentra muy estudiada en los paquetes comerciales para resolver modelos de programación entera. El resto de las restricciones puede complicar un poco la resolución –particularmente las restricciones (4) y (5), que imponen un rango de jugadores por posición–, pero globalmente el modelo es una versión levemente complicada del problema de la mochila. Esto sugiere que en la práctica la resolución computacional del modelo puede no ser complicada. Efectivamente comprobamos que esta cuestión estructural, sumado a que la instancia considerada es pequeña, da como resultado tiempos de resolución muy bajos para los casos considerados en este trabajo.

3.2. Equipo Perfecto: formulación matemática

Este segundo modelo busca la combinación óptima del equipo inicial y las modificaciones a realizar en cada fecha, de modo tal de sumar la mayor cantidad posible de puntos. Nuevamente, llamamos E al conjunto de equipos, J al conjunto de jugadores y P a las posiciones dentro del campo de juego. Además, definimos F como el conjunto de fechas, y llamamos $F' = F \setminus \{\text{mín}(F)\}$ al conjunto de todas las fechas excepto la primera (necesitamos definir este subconjunto de fechas dado que las modificaciones al equipo se hacen a partir de la segunda fecha del juego). Los parámetros del modelo son los mismos que en el modelo anterior, a excepción del puntaje: para cada jugador $j \in J$ y cada fecha $k \in F$, el parámetro $\text{puntaje}_{jk} \in \mathbb{Z}$ especifica el puntaje que obtuvo el jugador j

en la fecha k .

Para cada jugador $j \in J$ y cada fecha $k \in F$, introducimos las variables binarias x_{jk} , que informan si en la fecha k el jugador j es titular, e y_{jk} , que informan si en la fecha k el jugador j es suplente. Además, para $j \in J$ y $k \in F'$ introducimos la variable binaria z_{jk} , de modo tal que $z_{jk} = 1$ si y sólo si el jugador j se incorpora al equipo a partir de la fecha k . Con estas definiciones, el modelo es el siguiente:

$$\text{máx} \quad \sum_{j,k \in J \times F} \text{puntaje}_{jk} x_{jk}$$

$$x_{jk} + y_{jk} \leq 1 \quad \forall j, k \in J \times F \tag{7}$$

$$\sum_{j \in J} x_{jk} = 11 \quad \forall k \in F \tag{8}$$

$$\sum_{j \in J: \text{posicion}_j = p} y_{jk} = 1 \quad \forall p, k \in P \times F \tag{9}$$

$$\sum_{j \in J} \text{precio}_j (x_{jk} + y_{jk}) \leq 65,000,000 \quad \forall k \in F \tag{10}$$

$$\sum_{j \in J: \text{equipo}_j = e} (x_{jk} + y_{jk}) \leq 3 \quad \forall e, k \in E \times F \tag{11}$$

$$\sum_{j \in J: \text{posicion}_j = p} x_{jk} \geq \min_p \quad \forall p, k \in P \times F \tag{12}$$

$$\sum_{j \in J: \text{posicion}_j = p} x_{jk} \leq \max_p \quad \forall p, k \in P \times F \tag{13}$$

$$x_{jk} + y_{jk} - x_{j,k-1} - y_{j,k-1} \leq z_{jk} \quad \forall j, k \in J \times F' \tag{14}$$

$$x_{jk} + y_{jk} \geq z_{jk} \quad \forall j, k \in J \times F' \tag{15}$$

$$1 - (x_{j,k-1} + y_{j,k-1}) \geq z_{jk} \quad \forall j, k \in J \times F' \tag{16}$$

$$\sum_{j \in J} z_{jk} \leq 4 \quad \forall k \in F' \tag{17}$$

$$x_{jk}, y_{jk} \in \{0, 1\} \quad \forall j, k \in J \times F \tag{18}$$

$$z_{jk} \in \{0, 1\} \quad \forall j, k \in J \times F' \tag{19}$$

Nuevamente, la función objetivo maximiza el puntaje total obtenido por los jugadores titulares del equipo a lo largo de todo el torneo (dado que podemos intercambiar titulares y suplentes sin límite no se pierde generalidad considerando sólo a los titulares). Las restricciones (7) especifican que cada jugador puede ser titular o suplente, o bien no estar

seleccionado para el equipo en cada fecha, y las restricciones (8) piden que el equipo tenga exactamente 11 jugadores titulares. Además, las restricciones (9) solicitan exactamente un suplente por posición, totalizando así los cuatro suplentes del equipo. Las restricciones (10) y (11) especifican los límites de presupuesto total y cantidad máxima de jugadores por equipo, respectivamente. Por su parte, las restricciones (12) y (13) imponen los límites inferior y superior a la cantidad de jugadores en cada posición dentro del campo de juego.

Las restricciones (14), (15) y (16) relacionan las variables x e y con las variables z , de modo tal que $z_{jk} = 1$ si y sólo si el jugador j se incorpora al equipo (como titular o suplente) en la fecha k . Esta definición les permite a las restricciones (17) poner como límite un máximo de 4 incorporaciones en cada fecha. Finalmente, las restricciones (18) y (19) especifican la naturaleza de las variables.

Este modelo es sensiblemente más complicado que el anterior, dado que incorpora un número máximo de transferencias entre cada fecha y la siguiente. En cada fecha se sigue teniendo la estructura de *knapsack* del modelo anterior, pero la introducción de las variables z para representar los jugadores transferidos y de las restricciones (17) que imponen un máximo a estas transferencias, hace que la resolución computacional se complique. En sintonía con estas observaciones, más el hecho de que las instancias resueltas ahora son de tamaño mediano, los tiempos de resolución de este modelo son un poco más altos que para el modelo anterior.

3.3. Resultados

En la Tabla 1 se muestran los resultados obtenidos por los 2 modelos en los cuatro torneos disputados entre los años 2009 y 2010, y los tiempos de corrida de cada uno de ellos. Se exhibe también el puntaje del ganador del juego.

Cabe notar que en los cuatro torneos se da prácticamente un empate entre el ganador del juego y el modelo del equipo fijo con sólo titulares (el modelo gana por muy poco en 3 de los 4 torneos, y pierde por muy poco en el restante). Estos resultados muestran que la actuación del ganador del juego en todos los casos es muy meritoria ya que obviamente compite sin conocer los resultados que se darán a posteriori. Claramente los buenos jugadores aprovechan la naturaleza dinámica del equipo para

ir mejorándolo fecha a fecha, y de este modo competir a la par con un modelo que deja fijo el equipo a lo largo de todo el torneo, pero que juega con los resultados conocidos.

El Equipo Perfecto, que marca una cota superior en puntaje a lo que cualquier jugador puede alcanzar, supera al ganador del juego con puntajes que están entre un 50 % y un 70 % por encima, según el campeonato. Esta gran diferencia se debe fundamentalmente a que el Equipo Perfecto captura a jugadores que tienen grandes actuaciones de manera bien esporádica, lo que habitualmente no es encontrado ni siquiera por jugadores expertos.

La diferencia de puntajes entre los torneos Clausura y los torneos Apertura se explica en que en estos últimos el juego comienza una fecha antes.

Los resultados obtenidos por nuestros modelos fueron publicados en el diario en diversas oportunidades [18, 19, 20].

En lo que respecta a los tiempos de corrida, el primer modelo corre muy rápidamente, mientras que la obtención del Equipo Perfecto puede demorarse hasta un poco más que 20 minutos. El modelo del equipo fijo tiene alrededor de 500 variables (el número de jugadores total del campeonato) y poco más de 20 restricciones. El modelo del Equipo Perfecto tiene alrededor de 2000 variables y 3500 restricciones. En todos los casos los modelos se resuelven a optimalidad en los tiempos reportados en la Tabla 1. Los experimentos se realizaron con Cplex 12.2 en una PC con 2 GB de memoria RAM y un procesador con una frecuencia de 1.6 GHz.

Torneo	Ganador GDT	Equipo Titulares	Tiempo Corrida	Equipo Perfecto	Tiempo Corrida
Cl. 09	1279	1318	2 seg	1990	26 seg
Ap. 09	1375	1336	1 seg	2173	9 min
Cl. 10	1227	1232	1 seg	2027	20 min
Ap. 10	1394	1412	2 seg	2168	21 min

Tabla 1: Resultados de los modelos descriptivos en los torneos de 2009 y 2010

4. Modelo prescriptivo

El modelo “a priori”, o prescriptivo, presenta un mayor desafío, dado que ahora debemos encontrar equipos óptimos sin conocer el desempeño futuro de cada jugador. Para ello, confeccionamos un índice para cada jugador, que es una predicción del puntaje que va a alcanzar en la fecha siguiente. Una vez obtenidos dichos índices, corremos un modelo similar al de la sección anterior, donde ahora en vez de tener los puntajes de cada jugador tenemos los índices. La pregunta es cómo armar dichos índices, de modo de que sean una representación razonable de lo que irá a pasar en la realidad.

Tras algunas pruebas iniciales, llegamos a la conclusión de que el promedio de puntos que cada jugador sacó en las últimas fechas no es un buen predictor del puntaje que el mismo jugador sacará el siguiente fin de semana, ya que no tiene en cuenta consideraciones claves como el rival al que va a enfrentar, si es local o visitante, la actualidad de su equipo, etc.

Decidimos entonces construir el índice considerando el puntaje de cada jugador en las fechas que se han jugado hasta ese momento en el torneo en cuestión, pero ponderando dicho promedio por 3 factores: la condición de local o visita (usamos un ponderador de 1,05 para los locales, y de 0,95 para las visitas), la posición en la tabla del rival que va a enfrentar (ponderamos entre 1 y 1,05 si se enfrenta a los últimos de la tabla, y entre 0,95 y 1 si se enfrenta a los primeros), y la actualidad del jugador o de su equipo (hasta un 5% más a jugadores o equipos que vienen en buenas rachas, y hasta un 5% menos a jugadores o equipos que vienen en malas rachas, o que llegan cansados por venir por ejemplo de una disputa cercana en un torneo internacional). Los promedios obtenidos por dicho jugador en los 2 últimos torneos se consideran como si fueran una actuación más en una fecha del torneo actual.

Por último, hay un factor más de ponderación en cada índice, que llamamos “juega o no juega”, y que es un 1 para aquellos jugadores que se anuncian en la prensa o por parte de los directores técnicos como titulares en la fecha que va a venir, y un 0 para el resto. De esta manera, tratamos de garantizar que nuestros 11 titulares disputen la fecha. Este

es un dato crucial, y que a veces no se conoce a priori. Hay que tener en cuenta que la constitución de los equipos del juego para cada fecha cierra media hora antes del inicio del primer partido. Como las fechas se suelen jugar entre viernes y lunes, hay veces que uno puede no conocer el viernes la formación titular que un equipo tendrá el día domingo o el día lunes. Ahí entran a tener importancia los jugadores suplentes, por ello también es aconsejable tener “buenos” jugadores suplentes, aunque puede ser importante no gastar un gran presupuesto en ellos porque en la mayoría de los casos no serán utilizados. Utilizamos con esta idea un ponderador de 0,1 para el índice de los jugadores suplentes en la función objetivo, valor que fijamos tras realizar previamente algunas pruebas. Este ponderador podría ajustarse haciendo un estudio estadístico de en cuántas oportunidades los jugadores suplentes terminan participando efectivamente del equipo, o incluso corriendo el modelo para varios campeonatos con diferentes ponderadores para deducir cuál entrega mejores resultados habitualmente.

Notar que si un jugador jugó menos que k partidos en el torneo actual (y solemos trabajar con $k = 3$), no se lo considera como potencial candidato a formar parte del equipo por más que en la fecha siguiente se lo vuelva a anunciar como titular (para que el modelo no se vea tentado a poner a un jugador que tuvo una gran actuación pero después no se consolidó como titular en su equipo).

El modelo prescriptivo tiene entonces dos etapas: la constitución del equipo inicial y las modificaciones sugeridas fecha a fecha.

4.1. El equipo inicial

Una vez que tenemos definidos los índices de cada jugador, preparamos un equipo inicial que intenta sacar el mayor puntaje posible en la primer fecha del juego. Nuevamente, el conjunto E representa a los equipos, el conjunto J corresponde a los jugadores y P representa las cuatro posiciones en el campo de juego. Los parámetros son los mismos que para los modelos descriptivos, a excepción del parámetro índice $j \in \mathbb{R}$ asociado con cada jugador $j \in J$, que representa el índice descrito más arriba. Para cada jugador $j \in J$, introducimos una variable binaria x_j que vale $x_j = 1$ si y sólo si el jugador j es titular en el equipo inicial, y una variable binaria y_j que vale $y_j = 1$ si y sólo si el jugador j es suplente en el equipo inicial. Con estas definiciones, el modelo es el siguiente:

$$\begin{aligned} & \text{máx} && \sum_{j \in J} \text{indice}_j x_j + 0,1 \text{ indice}_j y_j \\ & x_j + y_j &\leq 1 & \quad \forall j \in J & (20) \\ & \sum_{j \in J} \text{precio}_j (x_j + y_j) &\leq 65,000,000 & (21) \\ & \sum_{j \in J: \text{equipo}_j = e} (x_j + y_j) &\leq 3 & \quad \forall e \in E & (22) \\ & \sum_{j \in J: \text{posicion}_j = p} x_j &\leq \max_p & \quad \forall p \in P & (23) \\ & \sum_{j \in J: \text{posicion}_j = p} x_j &\geq \min_p & \quad \forall p \in P & (24) \\ & \sum_{j \in J} x_j &= 11 & (25) \\ & \sum_{j \in J: \text{posicion}_j = p} y_j &= 1 & \quad \forall p \in P & (26) \\ & x_j, y_j &\in \{0, 1\} & \quad \forall j \in J & (27) \end{aligned}$$

La función objetivo maximiza el índice total del equipo, ponderando con un 10 % a los suplentes. Las restricciones (20) piden que cada jugador sea titular o suplente, o que no esté en el equipo, mientras que las restricciones (21)-(26) establecen las condiciones que el equipo debe cumplir dadas por las reglas del juego. Finalmente, las restricciones (27) especifican la naturaleza de las variables.

4.2. Cambios y transferencias

A partir de la segunda fecha del juego se puede ir actualizando el equipo, intercambiando titulares por suplentes y realizando hasta las 4 transferencias que el juego permite. Obviamente el índice de cada jugador se actualiza con lo ocurrido en la fecha que pasó, sumado a las características de la fecha que viene (localía, actualidad, rival, juega o no juega).

El modelo para hacer las modificaciones en el equipo fecha a fecha utiliza los mismos conjuntos, parámetros y variables que el modelo para determinar el equipo inicial. Para cada jugador $j \in J$, se tienen las

variables binarias x_j e y_j , que representan si el jugador j es titular o suplente en el equipo, respectivamente. Además, definimos $A \subseteq J$ como el equipo actual, sobre el cual se realizarán las modificaciones. Con estas definiciones, el modelo es el siguiente:

$$\text{máx} \quad \sum_{j \in J} \text{indice}_j x_j + 0,1 \text{ indice}_j y_j$$

$$x_j + y_j \leq 1 \quad \forall j \in J \quad (28)$$

$$\sum_{j \in J} \text{precio}_j (x_j + y_j) \leq 65,000,000 \quad (29)$$

$$\sum_{j \in J: \text{equipo}_j = e} (x_j + y_j) \leq 3 \quad \forall e \in E \quad (30)$$

$$\sum_{j \in J: \text{posicion}_j = p} x_j \leq \max_p \quad \forall p \in P \quad (31)$$

$$\sum_{j \in J: \text{posicion}_j = p} x_j \geq \min_p \quad \forall p \in P \quad (32)$$

$$\sum_{j \in J} x_j = 11 \quad (33)$$

$$\sum_{j \in J: \text{posicion}_j = p} y_j = 1 \quad \forall p \in P \quad (34)$$

$$\sum_{j \in A} x_j + y_j \geq 11 \quad (35)$$

$$x_j, y_j \in \{0, 1\} \quad \forall j \in J \quad (36)$$

El modelo es similar al presentado en la sección anterior, buscando maximizar el índice del nuevo equipo (otra vez ponderando con un 10 % a los suplentes) y respetando las restricciones impuestas por el juego. Como única diferencia, la restricción (35) solicita que el nuevo equipo tenga al menos 11 jugadores que estaban presentes en el equipo anterior (lo cual corresponde a haber realizado a lo sumo 4 transferencias).

Tanto este modelo como el de la subsección anterior cuentan con alrededor de 1000 variables y 500 restricciones, y se ejecutan a optimalidad en un par de segundos.

4.3. Resultados

Analizamos la actuación de nuestro modelo en los dos torneos del año 2010, Clausura y Apertura (en ese orden, dado que en la Argentina en la primera mitad del año se juega el torneo Clausura y en la segunda mitad, el Apertura). En las Tablas 2 y 3 se puede ver fecha a fecha la predicción global de puntaje para el equipo elegido por el modelo y el puntaje real que obtuvo, para cada uno de los dos torneos analizados. Como era de esperar, dado que el modelo selecciona a los mejores jugadores hasta el momento y la variabilidad de los rendimientos de los jugadores de fútbol es bastante amplia, la predicción de puntajes suele estar por encima del puntaje real obtenido. Para confirmar esta idea durante el Clausura 2010 también armamos un equipo “random”, es decir, un equipo conformado por jugadores elegidos al azar, con la única condición de que fecha a fecha los 11 titulares participaran de los partidos del fin de semana. El equipo random sacó 899 puntos, con una predicción de puntaje de 934 puntos, es decir el puntaje real se acerca mucho más a la predicción: menos de un 4% de diferencia (contra un 21% y un 14% en los equipos generados por el modelo, para el Clausura y el Apertura de 2010, respectivamente).

En el Clausura 2010 participaron del juego 1.442.682 jugadores. El ganador del juego sacó 1227 puntos. Nuestro modelo, con 1070 puntos, se ubicó en la posición 13.547, quedando en el mejor 1% de la competencia. El equipo random, con 899 puntos, terminó en la posición 498.726. La posición del equipo random da también una idea de cuántos equipos “activos” existen (consideramos activo a un equipo que se va actualizando fecha a fecha), alrededor de 1 millón, ya que es esperable que el equipo random finalice por la mitad de la tabla de posiciones, dentro de los equipos activos. Un equipo no activo suele terminar el torneo participando con menos que 11 jugadores en cada fecha, ya que no logra reemplazar a aquellos jugadores que a lo largo del torneo se fueron lesionando o perdiendo la titularidad. La estimación de 1 millón de equipos activos coincide con la estimación de los organizadores del juego, que tienen observado que $2/3$ de los equipos se actualizan fecha a fecha. Si comparamos entonces la actuación de nuestro modelo contra el total de los equipos activos, tenemos que el mismo se ubica en el mejor 1,5% de la competencia.

En el Apertura 2010 participaron del juego 1.445.531 jugadores. El ganador del juego sacó 1394 puntos. Nuestro modelo, con 1322 puntos, se

Fecha	Predicción	Puntaje Real
5	97,84	48
6	90,47	76
7	90,06	59
8	86,40	63
9	86,84	67
10	95,64	77
11	87,05	81
12	91,27	81
13	90,12	88
14	94,81	69
15	86,51	69
16	90,31	73
17	91,01	72
18	91,72	59
19	83,68	88
Total	1353,73	1070

Tabla 2: Resultados del modelo prescriptivo en el Clausura 2010

ubicó en la posición 643, quedando en el mejor 1 por mil de la competencia, incluso considerando sólo a los equipos activos. Esta fue claramente la mejor actuación del modelo, considerando los 6 campeonatos en los que participó.

Cabe destacar que en ambos casos, la predicción de puntaje global superó incluso al ganador del juego, pero como ya mencionamos, no era esperable conseguir en nuestro modelo un puntaje similar a dicha predicción.

Una posible explicación de por qué el modelo suele tener mejores actuaciones en el segundo torneo del año que en el primero es la existencia de la Copa Libertadores en el primer semestre de cada año. Los mejores equipos disputan esta Copa, y por lo tanto los mejores jugadores lo hacen. Esto hace que los mejores jugadores jueguen a veces cansados en el torneo local, o directamente no jueguen. También suele pasar que los entrenadores anuncien su formación inicial horas antes de los partidos. Todo esto complica la constitución del equipo virtual, dado que genera una mayor incertidumbre, y puede afectar a los resultados finales.

Fecha	Predicción	Puntaje Real
4	89,85	91
5	101,77	77
6	92,64	105
7	97,07	95
8	93,9	65
9	89,6	83
10	102,44	82
11	95,85	97
12	95,01	84
13	89,25	88
14	100,34	55
15	101,16	90
16	95,58	80
17	96,73	78
18	98,03	67
19	92,56	85
Total	1531,19	1322

Tabla 3: Resultados del modelo prescriptivo en el Apertura 2010

Un dato adicional es que si consideramos a los 6 campeonatos en los que el modelo participó del juego como un único campeonato, hay 343.017 competidores que participaron en todos ellos, estando nuestro modelo en la posición 530. Es decir, se ubica en el mejor 2 por mil de la competencia.

5. Conclusiones y trabajo futuro

El objetivo de este trabajo es analizar el aporte de la programación matemática a la hora de diseñar un entrenador virtual, o asesorar a un entrenador deportivo real. Utilizamos como caso de estudio un juego de fantasía realizado en el marco del torneo argentino de fútbol.

Presentamos modelos de programación matemática diseñados a priori, que llamamos prescriptivos, y diseñados a posteriori, que llamamos descriptivos, en la búsqueda de conseguir equipos óptimos para este jue-

go organizado por un diario argentino. El juego moviliza a más de un millón de jugadores en cada edición.

Los modelos descriptivos aquí desarrollados consiguen mostrar el equipo ideal que se debió haber armado a lo largo de todo el torneo, para obtener el mayor puntaje posible, cumpliendo con las restricciones del juego.

El modelo prescriptivo utiliza datos históricos y características de la próxima fecha a disputar, a fin de armar un equipo competitivo para el juego. Lo testeamos haciéndolo participar de la competencia. Los resultados marcan que nuestro modelo se posiciona habitualmente en el 3 % mejor del juego, llegando en una oportunidad a estar en el mejor 1 por 1000 del certamen. Tomando los 6 campeonatos en los que el modelo participó del juego como un único torneo, nuestro jugador virtual se posiciona en el mejor 2 por 1000 de la competición. Era esperable que cuánto más largo sea el torneo, más chances hay de que nuestras herramientas de estadística y optimización funcionen mejor.

El modelo prescriptivo desarrollado fue usado siguiendo sus indicaciones al 100 %. Se podría pensar al modelo como asistente de un jugador experto. Por ejemplo, proponiendo los k mejores conjuntos de transferencias para el equipo en cada fecha y que el experto elija uno de ellos. Esto se puede hacer corriendo k veces el modelo que sugiere las modificaciones, prohibiendo en cada caso las soluciones óptimas que se van obteniendo. Otra posibilidad sería decidir de manera externa al modelo la inclusión o exclusión de un determinado jugador, y correr después el modelo de cambios y transferencias para determinar el resto de las modificaciones.

En cuanto a qué se podría hacer para intentar mejorar a nuestro jugador virtual, surgen diferentes ideas. Una de ellas es que la optimización sea más global y no tan golosa. Es decir, cuando se van a hacer las modificaciones al equipo en una fecha dada, no sólo se mire dicha fecha sino también una o dos más para adelante. E incluso pensar en una optimización más global para armar el equipo inicial. También se podría armar un equipo de “alto riesgo”, con jugadores que presenten altas varianzas en sus puntajes aunque no tengan los mejores índices, tal cual se definieron en nuestro modelo. Esto podría llevar a peores resultados en general, pero podría otorgar cada una cierta cantidad de campeonatos una muy buena performance.

Otra alternativa sería intentar encontrar a los mejores jugadores de

cada fecha, para formar al equipo virtual. Dado que los puntajes de los jugadores suelen tener grandes variabilidades, puede ser interesante, más que predecir su puntaje, utilizar herramientas estadísticas para predecir quienes serán los mejores jugadores del fin de semana que se avecina. Para ello se podrían implementar modelos estadísticos sofisticados, utilizando datos históricos, para intentar encontrar cuáles son las variables que determinan en forma más fidedigna quienes serán los mejores jugadores en un determinado fin de semana (localía, rival, estadio, árbitro, etc).

Cabe destacar también que el problema aquí encarado presenta algunas similitudes con el problema de armar un portafolio de acciones empresariales, a fin de maximizar las ganancias de un inversor. Por ello creemos que algunos de los modelos que se usan en finanzas para predecir el comportamiento futuro de determinadas acciones, como los modelos de tipo CAPM [4, 8], basados en la teoría del portafolio de Markowitz [5], también podrían ser útiles a fin de conseguir equipos robustos para nuestro director técnico virtual.

Herramientas como las desarrolladas en este trabajo deberían servir para asesorar a entrenadores en deportes reales. La combinación entre deportes y matemática con este fin es muy bien tratada en la película de Bennett Miller, “Moneyball”, basada en [3] e interpretada por Brad Pitt y Jonah Hill. Esta película trata sobre la historia real de Billy Beane y Peter Brand, gerentes deportivos de un club de béisbol modesto de los Estados Unidos, quienes cambiaron radicalmente la forma de actuar, incorporando la matemática en la toma de decisiones deportivas, obteniendo de esta manera excelentes resultados.

Agradecimientos: A Carlos Prieto, que lamentablemente ya no está con nosotros, y acompañó este proyecto con gran entusiasmo. A Diego Javier Romero y Jorge Blanco, responsables del juego en el diario, por su ayuda permanente para la concreción de este trabajo. A Andrés Farall, Leonardo Faigenbom, Leonel Spett y Pablo Groisman, por las interesantes discusiones mantenidas sobre este proyecto. A Mario Guajardo por la lectura detallada del trabajo y sus sugerencias para mejorar la versión final. A Sebastián Ceria y Gustavo Braier, por sus comentarios y observaciones. El presente trabajo fue parcialmente financiado por los proyectos ANPCyT PICT-2012-1324 (Argentina), CONICET PIP 112-

200901-00178 (Argentina), UBACyT 20020100100980 (Argentina), y por el Instituto Milenio “Sistemas Complejos de Ingeniería” (Chile). El segundo autor es parcialmente financiado por el proyecto FONDECyT 1110797 (Chile).

Referencias

- [1] Coleman B.J., Identifying the “Players” in Sports Analytics Research, *Interfaces* 42 (2) (2012) , 109-118.
- [2] Kendall G., Knust S., Ribeiro C.C., Urrutia S., Scheduling in sports: An annotated bibliography, *Computers & Operations Research* 37 (1) (2010), 1-19.
- [3] Lewis M., *Moneyball: The Art of Winning an Unfair Game*, Norton & Company (2003).
- [4] Lintner J., The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Review of Economics and Statistics* 47 (1) (1965), 13-37.
- [5] Markowitz H.M., Portfolio Selection, *The Journal of Finance* 7 (1) (1952), 77-91.
- [6] Nurmi, K., Goossens, D., Bartsch, T., Bonomo, F., Briskorn, D., Durán, G., Kyngas, J., Marenco, J., Ribeiro, C.C., Spieksma, F., Urrutia, S., Wolf-Yadlin, R., 2010. A framework for scheduling professional sports leagues. In Ao, S-I., Katagir, H., Xu, L., Chan, A.H-S. (eds) *IAENG Transactions on Engineering Technologies*, American Institute of Physics. Vol. 5, pp. 1428.
- [7] Ribeiro C.C., Sports scheduling: Problems and applications, *International Transactions in Operational Research* 19 (2012), 201-226.
- [8] Sharpe W.F., Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19 (3) (1964), 425-442.
- [9] <http://fantasy.premierleague.com/>
- [10] <http://www.fantacalcio.kataweb.it/>

- [11] http://fantacalcio.repubblica.it/index.php?page=faq&ck_fantacalcio#16
- [12] <http://www.fantagazzetta.com/esclusive-fg/come-inventai-il-fantacalcio-intervista-esclusiva-a-riccardo-albini-inventore-del-fantacalcio-165805>
- [13] <http://www.hattrick.org/>
- [14] <http://mlb.mlb.com/mlb/fantasy/>
- [15] <http://www.nba.com/fantasy/>
- [16] <http://msn.foxsports.com/fantasy/football/>
- [17] <http://www.xperteleven.com/>
- [18] <http://edant.clarin.com/diario/2009/07/09/deportes/d-01955551.htm>
- [19] <http://edant.clarin.com/diario/2009/12/20/deportes/d-02104838.htm>
- [20] <http://edant.clarin.com/diario/2010/05/20/deportes/d-02197713.htm>
- [21] <http://www.fantasysportsmath.com/>