



# **Trabalho Final de Curso**

**Eng. Electrotécnica e de Computadores**

**Ramo de Sistemas Electrónicos e Computadores**

## **Totobola por AE**

1999/2000

**Nº 26518 - Paulo J Silva**

# Totobola por Algoritmos Evolutivos

## 1. Introdução

O Totobola por algoritmos evolutivos é um programa baseado em algoritmos genéticos destinado a prever os resultados dos jogos do boletim do Totobola. O programa pode ter uma maior variedade de aplicações: para prever resultados de diversos campeonatos e de jogos; ou para classificar e pontuar os jogadores de um campeonato; ou outros problemas que envolvam uma tabela e se queira converter para uma lista com um determinado e específico significado.

O programa analisa os resultados efectuados nas jornadas anteriores (nas primeiras jornadas a previsão é errónea e insuficiente). Para isso constrói uma lista ordenada, com as equipas, de maneira a respeitarem as vitórias, empates e derrotas obtidas anteriormente.

Um algoritmo genético é responsável pela optimização do número de jogos certos para cada lista. As listas são portanto o nosso cromossoma. Depois de obtida a melhor lista, fazem-se as previsões para os jogos que foram introduzidos previamente.

No segundo capítulo, deste trabalho, fala-se sobre o Totobola, como sistema em análise, sua história e o motivo deste trabalho. No terceiro capítulo explica-se o programa na sua concepção e o algoritmo usado. Por fim, no último capítulo interpretam-se os resultados, tiram-se conclusões e apresentam-se propostas de melhorias do sistema.

## 2. O Jogo

O Totobola é um jogo social, conhecido por todos nós, que consta em apostar e acertar nos catorze resultados, previamente escolhidos, de jogos de futebol da primeira (1ª divisão) e segunda liga (2ª divisão – divisão de honra). Para cada jogo referido no boletim temos a equipa que joga em casa versus a equipa que joga fora, ao qual é preenchido com uma aposta no resultado do jogo. Essa aposta é feita sob a forma dos símbolos:

“1” - vitória para a equipa que joga em casa;

“X” – empate das equipas;

“2” – derrota para equipa da casa.

Na Figura 2.1 é apresentado um boletim do Totobola já preenchido. Este boletim corresponde ao vaticínio da trigésima quarta jornada (a última) do campeonato de 1999/2000.

MATRIZ totobola 20 2000.05.14 NORMAL

1. GIL VICENTE-PORTO  
2. BOAVISTA-BRAGA  
3. BENFICA-MARITIMO  
4. SALGUEIROS-SPORTING  
5. SANTA CLARA-GUIMARÃES  
6. SETÚBAL-ULLEIRA  
7. FARENSE-BELELENSES  
8. AMADORA-CAMPOMAIENSE  
9. RIO AVE-ALVESCA  
10. CHAVES-P. FERREIRA  
11. MAIA-ACADÉMICA  
12. FREAMUNDE-VARZIM  
13. FELGUEIRAS-BEIRA-MAR  
14. LEÇA-ESPINHO

MULTIPLAS - ASSINALE ACUM. O N.º DE APOSTAS (ver tabela)

7940473

NOME PAULO SILVA  
MORADA  
CODIGO POSTAL

SE DESEJAR ANONIMATO EM CASO DE PREMIO, MARQUE 'X'

Figura 2.1

Pode-se observar que em certos jogos (1, 2, 7, 8, e 13) foram colocados dois símbolos, que significa uma dupla, ou seja aposta-se nos dois resultados em simultâneo o que

corresponde ao equivalente de 2 apostas individuais, cada uma com o seu símbolo. Como temos 5 duplas (e nenhuma tripla – também possível – 3 apostas em simultâneo) corresponde a 32 apostas ( $2^5$ ). Podemos calcular o número de apostas feitas, pela seguinte fórmula:

$$(1) \quad \text{N}^\circ \text{ de apostas} = 2^{(\text{n}^\circ \text{ duplas})} * 3^{(\text{n}^\circ \text{ triplas})}$$

Os prémios são para os apostadores que acertarem nos 14 jogos sendo-lhes dado a sua quota-parte do dinheiro atribuído ao primeiro prémio. Aos que acertarem em 13, 12 jogos ficam na categoria do segundo e terceiro prémio, respectivamente.

Fazer vaticínios sobre os resultados dos jogos é coisa velha, tão velha seguramente como os próprios jogos. No caso do Totobola temos, de seguida um pouco da história.

## 2.1 A História

O Totobola teve início em 24 Setembro de 1961, criado pelo decreto-lei nº 43 377. de 3 de Julho de 1961, como um jogo de apostas mútuas.

Parece, porém, que só por volta dos anos trinta se deu estrutura às previsões dos resultados dos jogos de futebol, estabelecendo com essas previsões um sistema de apostas.

Foi na Inglaterra que tudo teria começado quanto à complexa realidade das apostas mútuas sobre jogos de futebol. Mas não tardou que a experiência frutificasse e se alargasse a outros países, conquistando sem grande demora vários países da Europa Continental. A Suécia foi o primeiro país a contar com um sistema organizado de apostas mútuas, o qual, divergindo embora do inglês em diversos aspectos, estabeleceu bases e orientações que têm de ser consideradas fundamentais.

Com nomes diversos, mas sempre em ligação íntima com o futebol, as apostas mútuas cimentaram o seu interesse junto de públicos muito diversos, fazendo criar uma aura de novo agrado para o jogo que cada vez tomava mais consciência da sua posição de *"desporto-rei"* e que não tardava a ser, em muitos países, o *"rei dos desportos"*. Portugal, onde o futebol tem uma história de muita curiosa rapidez na conquista da simpatia popular (embora tivesse começado por ser, nitidamente, um desporto de elites), não tardaria a ser contagiado pelo gosto das apostas mútuas. Vários sectores de entre os adeptos de futebol praticavam, de quando em quando, mas quase sempre por ocasião de um jogo ou de um torneio com maior

importância, apostas sobre os resultados, mas sempre de uma forma embrionária, em que faltava tudo excepto o gosto pelo jogo e o desejo de adivinhar.

Fizeram-se diversas tentativas para dar a estas apostas um carácter permanente e, até, uma existência legal. Por mais de uma vez, entidades ligadas ao futebol pretenderam introduzir no nosso País a modalidade das apostas mútuas, muito embora fossem já bem conhecidas, pelos resultados da experiência, a complexidade e a dificuldade que a montagem de um sistema impiedosamente impunham. O desejo da introdução das apostas mútuas era já muito vivo em 1948, a avaliar pelas exposições apresentadas e pelos ensaios que, de vez em quando e sem qualquer carácter oficial, se iam realizando. Mas o terreno que teria de pisar-se mostrava-se particularmente melindroso. Havia muitas circunstâncias a acautelar, desde a intangível dignidade do desporto ao interesse nacional. A primeira dessas razões era o receio de que a nova modalidade de apostas mútuas desportivas viesse a prejudicar a existência e os benefícios da lotaria nacional.

Procedeu-se então a um estudo das condições do «mercado» e sobre a compatibilidade de interesses entre a lotaria e as apostas mútuas – apostas que merecedoras da simpatia popular, pois se praticavam, nos pontos mais diversos do País, clandestinamente, promovidas a título esta ou aquela comemoração ou sempre que um clube desportivo desejava, com elas, angariar receitas para fazer face às suas necessidades.

As sondagens efectuadas mostraram que era possível a coexistência, em termos satisfatórios, das duas realidades. Outros estudos demonstravam, ao mesmo tempo, que interesses futuros importaria acautelar. Tinha sido cometido à Santa Casa da Misericórdia de Lisboa a faculdade de realizar concursos públicos de prognósticos ou apostas mútuas sobre competições desportivas. Através de bem ponderadas decisões e após insistentes exigências de estudos de inatacável seriedade, admitia-se, assim, a possibilidade da criação, entre nós, dos populares concursos de prognósticos e atribuía-se a competência da nova modalidade à Misericórdia de Lisboa. O primeiro e decidido passo estava dado. Com que extremos de cuidado! Com que imperioso rigor de prudência e de dignidade sob todos os aspectos!

Entretanto, começa a sentir-se em Portugal necessidade de se criarem serviços para reabilitação dos diminuídos físicos. Como angariar receitas para uma actividade assistencial tão onerosa, mas tão humanamente imperiosa e urgente como essa reabilitação? No espírito do então provedor da Misericórdia de Lisboa, Dr. José Guilherme de Mello e Castro, começa

a criar corpo a ideia de se assentar nas apostas mútuas, como meio de financiamento dos serviços de reabilitação a nível nacional, ao mesmo tempo que (e isso era do mesmo modo importante e justo) ao desporto seria facultada uma nova fonte de receita, e que em muito iria, com certeza, fomentar todas as actividades desportivas. Estas circunstâncias aliadas ao facto de popularem por toda a parte, com forte incidência em vários sectores da sociedade, as apostas mútuas clandestinas, levaram o Governo, após um atento estudo económico sobre o mercado, a dar em 30 de Novembro de 1960, «luz verde» à criação entre nós das apostas mútuas desportivas.

Numa memorável conferência de imprensa, além de se ter anunciado a decisão governamental, transmitiu-se uma primeira ideia, muito ténue, ainda, da estrutura a emprestar à futura organização. A notícia encontrou, como não podia deixar de ser, um eco extraordinário em todos os órgãos de informação do País, especialmente, como era de prever, nos jornais desportivos. Naquele mesmo dia foi nomeado o director do departamento das Apostas Mútuas Desportivas, o Dr. Ricardo Jorge Correia da Fonseca. E, como urgia estudar para se pisar terreno firme, logo passados poucos dias o director do novo serviço seguia para o estrangeiro, afim de tomar contacto íntimo com as várias organizações europeias do «Toto». Entre as organizações visitadas, contaram-se a espanhola, a italiana, a suíça, a sueca e algumas alemãs.

Procurava-se, deste modo, tomar conhecimento seguro dos vários sistemas existentes nos diversos centros da Europa, especialmente o sistema mecânico e o sistema de selos. O primeiro apresentava sobre o segundo vantagens extraordinárias, até sem sombra de dúvida, o mais evoluído de todo o mundo.

Numa organização acabada de criar-se e que ia começar a sua vida, parecia que tudo apontava o sistema mecânico. Portugal devia ter uma organização de harmonia com as mais evoluídas, embora colhendo os ensinamentos e os frutos das experiências de outras organizações como, por exemplo, as da Espanha e Itália.

Em Abril de 1961, deu-se início aos trabalhos práticos de organização e de montagem da nova e complexa «máquina» que constituiria, no futuro, o Totobola. Trabalho esse que se desenvolveria, de dia e de noite, com a colaboração entusiástica e infatigável dos primeiros funcionários que iriam constituir a equipa responsável pelo empreendimento. Os fundos

originários da exploração do Totobola, são distribuídos pela assistência e pelo desporto, ver Fig. 2.2.



Figura 2.2

## 2.2 A Ideia

O Totobola, pode ser considerado como um sistema estocástico, em que, há partida, as probabilidades de acertar no resultado de um jogo são de 1 para 3. Poder-se-ia pensar que estava tudo dito, quanto às hipóteses de acertar num jogo, mas como veremos mais adiante as coisas não se passam bem assim. Sabe-se por exemplo, que certas equipas jogam muito melhor em casa do que quando jogam fora. E entre os muitos factores que influenciam o resultado dum jogo como o tempo, o espírito dos jogadores, o árbitro, etc... temos aqueles que podem ser quantificáveis como as vitórias, os golos, e pouco mais.

No início do campeonato, a tabela classificativa das equipas (3 pontos por vitória e 1 ponto por empate) dá-nos uma ideia falsa da realidade. De facto uma equipa pode ter os mesmos pontos que outra, mas ter jogado com as melhores equipas (mais pontuadas), enquanto que a outra pode ter jogado com as do fim da tabela, sendo por isso, à partida, melhor a primeira equipa do que a segunda. Contudo à medida que o campeonato avança

podemo-nos basear na tabela classificativa para prever o resultado de um jogo entre duas equipas dadas.

É o que muita gente faz e basta subtrair as duas classificações das equipas para ver quem poderá ganhar ou perder. No caso de se querer prever o empate já se tem de entrar com uma margem de empate, à nossa escolha, para esse cálculo. Tudo funciona por meio de cálculos aritméticos, como por intuição e preferências de valores qualitativos vários.

O Totobola, como um sistema que não tendo solução (previsível), torna-se num sistema ideal para estes tipos de estratégias de procura de soluções. De facto não se procura a solução, mas sim a melhor resposta possível, para uma possível solução. Não se pretende neste trabalho dar uma chave do Totobola, que seja vencedora inevitavelmente mas sim dar uma chave que qualquer bom entendedor de futebol considere boa.



## 3. O Programa

Com o aparecimento dos algoritmos genéticos podemos utiliza-los para a resolução deste problema sem ocorrer em programações, normalmente tendenciosas nos resultados (normalmente tudo para o lado dos um's, ou para o lado dos dois e poucos ou nenhuns empates). Para isso começou-se por definir qual o valor das equipas a que deve ser dada toda a importância. Avançamos para o número de jogos, já efectuados, certos pela lista, obtida no algoritmo. Mas analisemos melhor o que é um algoritmo genético e a sua aplicação neste caso.

### 3.1 O Algoritmo Genético

Os algoritmos genéticos constam de várias etapas repetidas de uma forma cíclica, à semelhança da natureza, para produzir um genoma evoluído dos seres humanos, animais, plantas, vírus, fungos, etc.

Assim como na natureza, os animais nascem, crescem e reproduzem-se segundo o seu genoma, que herdado dos pais é transmitido para os filhos e destes para os seus filhos, criando as chamadas gerações. Cada geração é então, atravessada no tempo, por uma população de indivíduos, identificados pelo seu(s) cromossoma(s). Cada cromossoma possui genes que são de início inicializados por exemplo aleatoriamente. A esse conjunto de cromossomas é-lhes assim aplicado, em cada geração, as seguintes etapas de funções genéticas:

- **Avaliação**

Cada cromossoma dá origem a um genoma ou entidade real (indivíduo) que caracteriza, na sociedade dos indivíduos, um valor próprio dado pela chamada função de *mérito*. Essa função pega nos genes que constituem o cromossoma e devolve um determinado valor por cada cromossoma (ver Fig. 3.1).

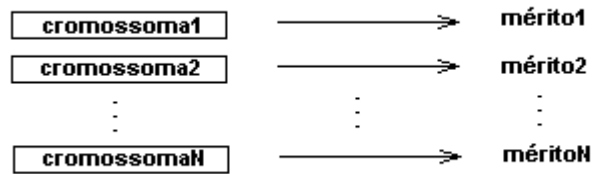


Figura 3.1

- **Seleção**

Através do valor do *mérito* é feita a selecção dos cromossomas que passam de geração, para gerar filhos. De facto os melhores são privilegiados com uma fatia maior na roda da sorte que é feita à população de cromossomas. Assim aleatoriamente é criada a nova população, de cromossomas, que vai estar sujeita às próximas etapas. Os melhores poderão aparecer repetidos. Esta operação é semelhante a uma cópia dos cromossomas, seleccionados.

- **Cruzamento**

Por cada 2 cromossomas escolhidos aleatoriamente dentro da população é-lhes feita a operação do cruzamento. Esta operação (tal como na vida temos a reprodução entre macho e fêmea), consta da criação de um novo cromossoma com genes oriundos do par que lhe deu origem (pais). Pega-se em pares de cromossomas e troca-se-lhes os genes, originando assim na nova população elementos diferentes dos originais. Esses genes do filho são escolhidos aleatoriamente de um ou do outro cromossoma seu superior hierárquico, ou através de uma estratégia diferente para a escolha dos genes do filho. Normalmente, para a população de cromossomas se manter do mesmo tamanho, procede-se à criação de dois filhos com características distintas.

- **Mutação**

Por fim, tem-se a mutação que consiste em alterar, numa determinada percentagem da população, um ou mais genes desses

cromossomas. A alteração dos genes é feita de uma maneira aleatória, tanto nos que vão ser escolhidos como na alteração em si. Outras formas podem ser encontradas para alterar uma pequena parte da população ao nível do gene, tal como acontece na natureza, explicando o aparecimento de novas espécies.

Este esquema cíclico mantém-se até se atingir uma condição de paragem. Essa condição pode ser por, se ter atingido uma determinada geração ou por se ter atingido um nível de avaliação de um cromossoma suficiente para a solução ou por qualquer outro motivo.

### 3.2 A Aplicação

No caso da aplicação do algoritmo genético ao Totobola, temos como dados, os jogos das jornadas anteriores, em termos de vitória, empate e derrota. Esses dados estão guardados num ficheiro juntamente com o nome das equipas e dos jogos a realizar na época e vão ser comparados posteriormente com as previsões feitas pelas listas construídas pelo algoritmo genético.

No campeonato Português temos principalmente, sem contar com a 2ª e 3ª Divisão, duas ligas ou seja 2 campeonatos distintos. São chamadas a primeira e segunda liga (antiga 2ª Divisão de Honra). Cada liga é constituída por 18 equipas cada. Como cada equipa joga em sua casa com todas as outras 17, faz com que, existam 306 ( $18 \times 17$ ) jogos. Como também, em cada jornada todas as equipas jogam, realizam-se assim 9 jogos, e ao todo temos 34 ( $306/9$ ) jornadas.

Os genes do algoritmo genético representam os valores próprios das diversas equipas. As equipas são colocadas posicionalmente no cromossoma. Assim temos no primeiro gene o valor da equipa A, no segundo o valor da equipa B, etc. Esses valores são atribuídos de início de maneira aleatória, com números inteiros entre 0 e 511 (9 bits). O conjunto desses valores constitui o nosso cromossoma e representa uma lista ordenada por ordem de grandeza dos valores desses genes. Ficamos assim com uma lista que define melhores e piores equipas.

Ao todo temos uma população de 100 cromossomas. Para cada um deles, vai ser calculado o seu *mérito*, da seguinte maneira: por cada jogo, já previamente jogado, o seu resultado (vitória, empate ou derrota) vai ser comparado com o que a lista ditar como sendo o

resultado estimado, entre os três previstos, uma vitória, empate ou derrota; a contabilização de igualdades das vitórias, empates e derrotas entre os jogos todos, já jogados, e as estimativas efectuadas pelo cromossoma totalizam o *mérito* ; a estimativa dada pelo cromossoma é por sua vez, analisado da seguinte maneira, ver figura 3.2.

Dada uma equipa A, com o seu valor dentro do cromossoma  $a$ , é-lhe acrescentada a margem de empate  $mar\_emp$ , quer-lhe seja somada quer subtraída. A margem de empate é também um gene dentro do cromossoma que é inicializado aleatoriamente e entra em todas as operações do algoritmo genético, como os outros genes.

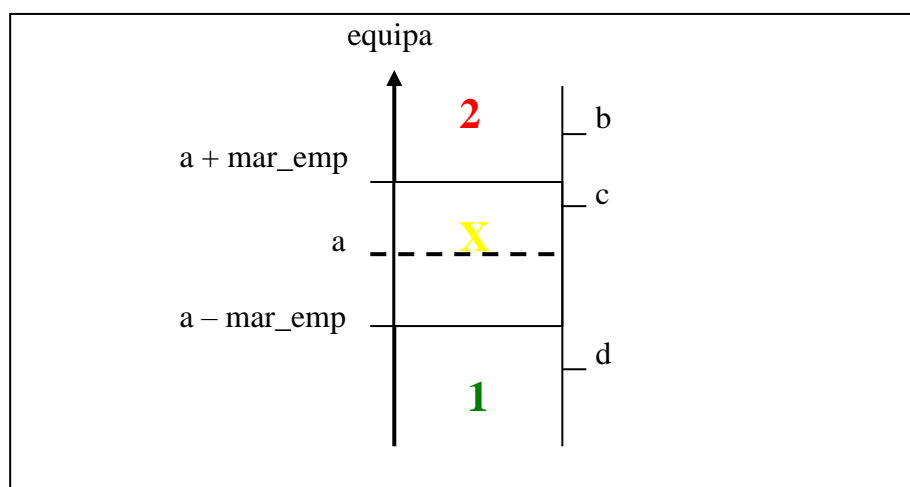


Figura 3.2 – Decisão da vitória, empate e derrota.

Todas as equipas que estejam para além da soma de  $a+mar\_emp$  ganham à equipa A. As equipas que estejam entre  $a-mar\_emp$  e  $a+mar\_emp$ , inclusive empatam com A. Restando as equipas que se encontram abaixo a  $a-mar\_emp$  que perdem com A. Na figura 3.2 pode-se ver que a equipa B ganha a A, que a equipa C empata com A e D perde com A.

O *mérito* representa então o número de jogos, já efectuados e certos. Por jogos certos quer-se dizer que são jogos em que a estimativa dada pela lista é igual ao resultado real do jogo. Cada lista tem o seu *mérito*. A melhor lista é a que vai ser usada para a previsão dos resultados da(s) próxima(s) jornada(s), pelo mesmo processo que o anterior usado na estimativa.

Calculado o *mérito*, termina também a operação de avaliação do algoritmo genético e passa-se de seguida para a operação de selecção. Na selecção obtém-se uma nova população,

esquecendo-nos da população anterior à excepção do melhor elemento, ou seja a melhor lista dessa mesma população, tendo-se assim uma selecção elitista.

Essa nova população é obtida pela escolha aleatória dos cromossomas da população a esquecer. Para essa escolha cria-se um género de roda da sorte em que os cromossomas de melhor *mérito* têm uma fatia directamente proporcional, proporcionando assim uma maior probabilidade de ser seleccionado pela escolha aleatória. De facto, o tamanho da fatia que cada cromossoma tem, é obtido pelo que se chamou de *mérito* ponderado, dado pela função (2), que apesar do aspecto não deixa de ter um comportamento equivalente ao exponencial, favorecendo os melhores (*mérito* mais alto). Foi escolhida por questões de rapidez de cálculos, evitando-se a operação exponencial. A escolha do valor para o parâmetro  $\alpha$  pode ser ponderada para termos curvas diferentes, mas no nosso caso vale 5.

$$(2) \quad \boxed{\text{mérito\_ponderado} = \frac{1}{\alpha * \left(1 - \frac{\text{mérito} - \text{mérito}_{\min} + 1}{\text{mérito}_{\max} - \text{mérito}_{\min} + 1}\right) + 1} - \frac{1}{\alpha + 1}}$$

Devido à repetida escolha aleatória (até completar o tamanho da nova população) feita à população antiga, podem aparecer, na nova população, cromossomas repetidos, uma ou várias vezes, com tendência a terem sido os melhores a ser seleccionados. Tendo nós uma nova população de tamanho idêntico à anterior e que inclui o melhor elemento da anterior e outros tantos seleccionados, terminamos a operação de selecção e entramos na operação de cruzamento.

No cruzamento a população sofre uma transformação própria de maneira que, pode-se considerar que ficamos com uma população diferente. Com efeito são escolhidos de maneira aleatória pares de cromossomas para serem cruzados. Todos os cromossomas cruzam, formando 50 pares. A cada par é criada uma máscara de bits representativa dos genes que se vão cruzar. Ver figura 3.3.

A máscara M é criada aleatoriamente e para cada bit posicional a 1, significa uma troca de genes na mesma posição. Tendo então, por exemplo, um par de cromossomas C1 e C2 com os respectivos genes contendo valores (a, b, c, ...) e (a', b', c', ...) dão origem a outros

dois cromossomas C1' e C2' preenchidos pelos genes quer de C1 quer de C2 conforme ditou a máscara estabelecida das trocas de genes.

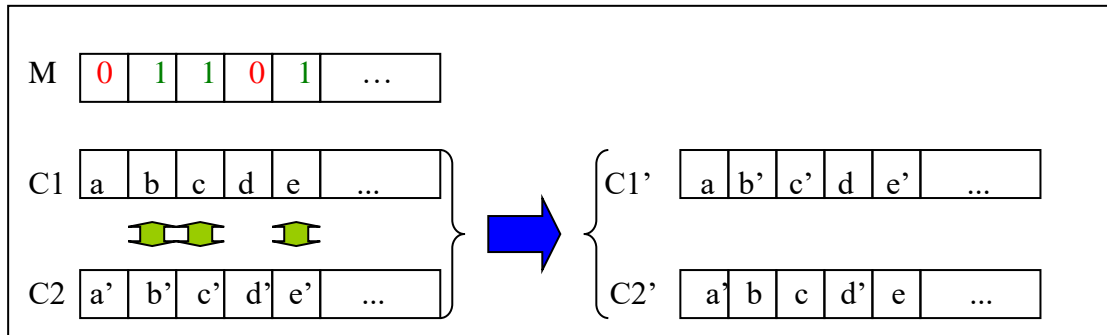


Figura 3.3 – Cruzamento

No caso da selecção a nova população criada não acrescenta informação nenhuma à antiga população, pois os cromossomas já existiam e não foram alterados. Já no cruzamento isso não é verdade. São criados cromossomas novos que herdam as características dos seus antepassados. E com isto terminamos o cruzamento e passamos à mutação.

A mutação altera determinados genes trocando-lhes o valor. No nosso caso essa alteração resume-se a uma inicialização do gene de forma aleatória. Os genes que são alterados são escolhidos de maneira aleatória. No nosso caso são alterados um determinado número de genes por cromossoma. Esse número está definido pela constante  $Nmut$  e vale 1, o que significa que é alterado só um gene por cromossoma. Os cromossomas que são alterados pela mutação são escolhidos segundo uma percentagem de mutação dada por  $Pmut * 100$ . O valor de  $Pmut$  é de 0.1, o que significa que só cerca de 10% da população de cromossomas muta.

Assim para cada cromossoma é calculado um valor aleatório entre 0 e 1 (este valor e todos os outros valores aleatórios referidos obedecem a uma distribuição uniforme), se calhar abaixo da percentagem de mutação o cromossoma é escolhido para mutar e assim é escolhido aleatoriamente um gene para ser reinicializado. A população de cromossomas é alterada pela mutação e acaba assim um ciclo ou seja uma geração.

O ciclo recomeça, com a uma nova geração de cromossomas que irão ser avaliados. A inicialização aleatória dos genes, no início do ciclo, já não é efectuada, dando origem ao longo das gerações, a novos e melhores cromossomas.

Ao fim de 10 000 gerações obtém-se uma lista já melhor do que a melhor dada aleatoriamente no início do algoritmo. Com essa lista vai ser feita a previsão dos jogos da próxima jornada. A previsão é feita da mesma maneira que são feitas as estimativas para o cálculo da função de *mérito*.

### 3.3 As melhorias à aplicação

Feita uma primeira aplicação da forma descrita no parágrafo anterior, foram feitas novas inovações ao algoritmo na tentativa de produzir melhores resultados. Entende-se por melhores resultados as vezes em que se obtêm melhores *méritos* alguma vez alcançados numa execução.

Uma das estratégias que foi acrescentada, ao algoritmo anterior, foi o “factor casa”, uma outra foi a “protecção dos genes bons” e ainda uma outra foi a das “equipas dinâmicas”. A primeira estratégia foi bem sucedida, como se vai ver na análise dos resultados do quarto capítulo, a segunda a sua utilidade depende da jornada, e a terceira não deu resultados melhores, pelo contrário baixou a média do *mérito* alcançado.

O “factor casa” parte da suposição de que as equipas não jogam da mesma maneira em casa como noutros campos menos conhecidos. Verifica-se que no Totobola 50% dos resultados são vitórias em casa, símbolo 1. Os outros 50% são igualmente repartidos pelos símbolos X e 2. Para o efeito vamos considerar que cada equipa tem associado a ela mais um gene. Poder-se-ia considerar que os dois valores de cada equipa são, um para quando a equipa joga em casa e outro para quando a equipa joga fora. No entanto vamos considerar que um dos valores, o antigo, permanece como estava e o outro é-se-lhe somado para dar o valor para quando a equipa joga em casa. Esse outro valor entra nas contas com um *offset* de  $-256$ , em relação aos seus permitidos de 0 a 511, de maneira a que o valor de uma equipa quando joga em casa poder ser superior, ou inferior, ao valor de quando esta joga fora num máximo de 256 valores. Assim se o segundo gene somado com o *offset*, for positivo significa principalmente que a equipa joga melhor em casa do que fora. Se for negativo a equipa joga pior em casa do que fora.

A segunda estratégia aplicada foi a da “protecção dos genes bons”. Como o nome indica vamos considerar que os genes das equipas (á excepção do gene que define a margem de empate), podem ser bons ou maus. Um gene é qualificado como bom se acertar em todos os jogos já efectuados quer sejam jogos em casa quer sejam fora. Um gene ao ser considerado bom não é alterado nas operações de cruzamento nem de mutação. Fica assim o valor do gene constante variando sim os outros genes com o decorrer do algoritmo até vir a ser considerado gene mau.

Se, na operação de cruzamento, a máscara escolher que sejam trocados um gene bom e um mau o bom sobrepõe o mau. Se forem escolhidos dois bons permanecem como estão, e se forem escolhidos dois maus então trocam. Ver figura 3.4.

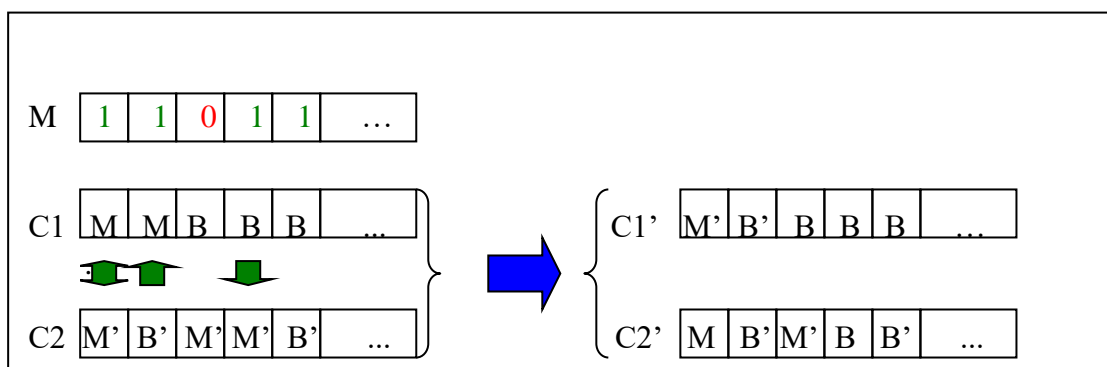


Figura 3.4 – Propagação dos genes bons no cruzamento.

Como se pode ver pela figura, os cromossomas criados têm um maior número de genes bons, que apesar de bom gene no contexto de um cromossoma poder não o ser no novo contexto e alterar assim a qualificação do que são os genes bons. Tem-se pois que recalculer a qualificação dos genes bons e maus. Podendo os bons virem a ser a maus e vice-versa.

Esta estratégia exige mais de cada equipa à medida que o campeonato evolui, devido às situações não cobertas por uma lista, que são as situações que obedecem à propriedade transitiva. Esta propriedade dita que se a equipa A ganha à equipa B e que se B ganha a C então pelo sistema linear seria lógico que A ganhasse a C, o que nem sempre acontece. Dai se optar por se considerar que o gene é bom se acertar acima de uma determinada percentagem do total dos jogos efectuados (essa constante recebeu o nome de *ppbom* – percentagem para bom).



A terceira melhoria “equipas dinâmicas” pressupõe que as equipas evoluem ao longo das jornadas, de uma maneira linear. Assim temos que, o valor da equipa é utilizado para os jogos da ultima jornada efectuados, enquanto que nos jogos das outras jornadas é acrescentado um valor equivalente ao declive da recta indicado na figura 3.5.

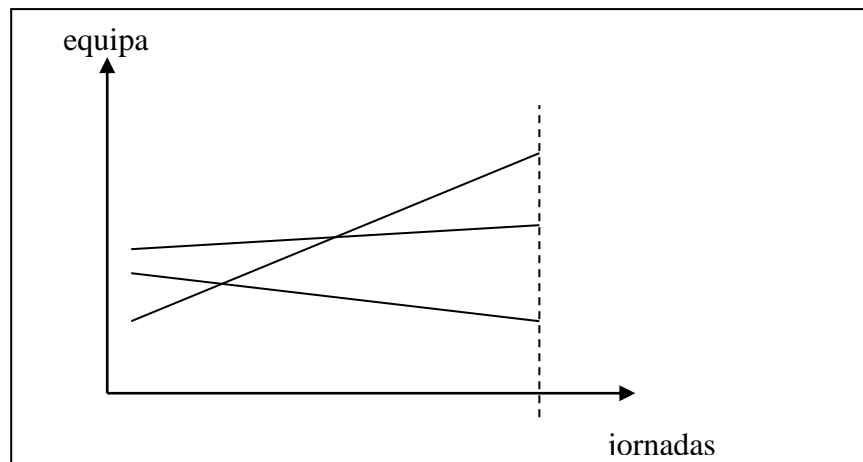


Figura 3.5 – Equipas dinâmicas. O tracejado é a jornada actual.

Ficamos então com valores diferentes para as equipas quanto maior for a distância entre a jornada actual e a jornada do jogo a considerar. É com esse novo valor que vão ser feitas as estimativas dos jogos e por conseguinte calculado o *mérito*. O declive da recta é dado por um novo gene associado a cada equipa. O declive toma valores entre 25,5 e -25,6 valores/jornada.

## 4. A Conclusão

O algoritmo exemplificado foi executado 10 vezes seguidas em diversas situações e com parâmetros diferentes ( ver tabela 4.1).

Primeiro foi aplicado o algoritmo normal isto é, sem as melhorias. Considerando a mesma jornada (31) temos o número de mutações por cromossoma  $Nmut$  diferentes, assim como a percentagem de cromossomas que mutam  $Pmut$ , diferentes.

Na tentativa de se atingir o *mérito* máximo ou óptimo de 279 (31\*9) jogos efectuados, foram alcançados diversos valores situados entre 158 e 171 de média. Não são apresentados todos os valores calculados pela execução do algoritmo, nomeadamente os valores das execuções #3 a #9, para não perturbar a legibilidade do que é importante na tabela, tornando-a menos exaustiva. O melhor resultado foi portanto, nesta fase, para o valor de  $Nmut = 1$  e  $Pmut = 0,1$ , que equivale a que 10% dos cromossomas mutam (em média).

Descrição	J	Nmut	Pmut	ppbom	óptim	#1	#2	...	#10	média	%
normal	31	3	0,5	--	<b>279</b>	159	158	...	157	<b>158</b>	<b>56,6</b>
normal	31	1	0,25	--	<b>279</b>	171	170	...	168	<b>169</b>	<b>60,7</b>
normal	31	1	0,1	--	<b>279</b>	173	171	...	172	<b>171</b>	<b>61,3</b>
normal	31	0	0,05	--	<b>279</b>	170	171	...	172	<b>168</b>	<b>60,4</b>
normal+casa	31	1	0,25	--	<b>279</b>	187	185	...	185	<b>184</b>	<b>66,0</b>
normal+casa	31	2	0,1	--	<b>279</b>	190	191	...	185	<b>190</b>	<b>68,1</b>
normal+casa	31	1	0,05	--	<b>279</b>	189	191	...	183	<b>189</b>	<b>67,7</b>
normal+casa	31	1	0,1	--	<b>279</b>	193	198	...	190	<b>194</b>	<b>69,4</b>
	31				<b>279</b>	192	185	...	197	<b>191</b>	<b>68,3</b>
normal+casa+bons	31	1	0,1	90	<b>279</b>	196	188	...	194	<b>192</b>	<b>68,9</b>
normal+casa+bons	31	1	0,1	80	<b>279</b>	194	196	...	186	<b>192</b>	<b>69,0</b>
normal+casa+bons	31	1	0,1	75	<b>279</b>	199	188	...	191	<b>191</b>	<b>68,5</b>
normal+casa+bons	31	1	0,1	65	<b>279</b>	191	190	...	194	<b>192</b>	<b>68,9</b>
normal+casa+bons	31	1	0,1	50	<b>279</b>	181	186	...	181	<b>184</b>	<b>66,1</b>
normal+casa+bons+ dinâmica	31	1	0,1	100	<b>279</b>	195	189	...	193	<b>191</b>	<b>68,4</b>

Tabela 4.1 – Vários testes para a mesma jornada

Numa segunda fase aplicou-se o “factor casa” e, para a mesma jornada, alterou-se os valores de  $Nmut$  e  $Pmut$ , verificando-se que os parâmetros melhores foram os mesmos que os anteriormente alcançados. No entanto atingiu-se um novo patamar no *mérito*. Cresceu em

mais de 8%. Conseguiu-se *méritos* máximos que anteriormente eram de 173 e que a gora são de 198.

A seguir acrescentou-se a “protecção dos genes bons”, que como se pode ver na tabela acima, não melhorou em nada os resultados anteriores já alcançados inclusivamente piorando no caso de percentagens para tornar o gene bom, *ppbom* baixas. Podemos verificar que nesta jornada 31, pelo menos a partir dos 50% de percentagem para bons, todos os genes do melhor genoma cumpriram essa fasquia, ficando todos bons. Acima dos 80-90% nenhuma equipa/genes o atinge, permanecendo todos maus.

Este método deve ser executado com  $ppbom = 100$ , significando assim que só se os jogos já efectuados foram todos acertados pela estimativa desse gene em relação aos outros genes/equipas, é que é considerado gene bom. Só assim se pode ter resultados no mínimo equivalentes à execução do programa só com o “factor casa”. Como se vai ver a seguir, esse método serve basicamente para acelerar a procura de maneira que fixando uns genes como bons, tornando-os imunes ao cruzamento e à mutação, os outros genes tendem a estabilizar mais depressa. Além de que, se propagam os genes bons. Alcançando-se mais rapidamente melhores cromossomas, apesar do que o que é bom para um cromossoma poder não o ser para outro.

Descrição	J	Nmut	Pmut	ppbom	óptim	#1	#2	...	#10	média	%
1ª Liga 1998/99	34	1	0,1	80	<b>306</b>	206	200	...	213	<b>206</b>	<b>67,4</b>
1ª Liga 1999/00	6	1	0,1	100	<b>54</b>	48	48	...	48	<b>48</b>	<b>88,3</b>
1ª Liga 1999/00	7	1	0,1	100	<b>63</b>	53	53	...	49	<b>52</b>	<b>83,0</b>
1ª Liga 1999/00	8	1	0,1	100	<b>72</b>	59	60	...	59	<b>58</b>	<b>81,1</b>
1ª Liga 1999/00	26	1	0,1	80	<b>234</b>	161	161	...	161	<b>164</b>	<b>70,1</b>
1ª Liga 1999/00	31	1	0,1	80	<b>279</b>	194	196	...	186	<b>192</b>	<b>69,0</b>
1ª Liga 1999/00	32	1	0,1	80	<b>288</b>	193	182	...	198	<b>195</b>	<b>67,7</b>
1ª Liga 1999/00	33	1	0,1	80	<b>297</b>	202	207	...	203	<b>203</b>	<b>68,4</b>
1ª Liga 1999/00	34	1	0,1	80	<b>306</b>	210	207	...	212	<b>208</b>	<b>67,8</b>
2ª Liga 1999/00	6	1	0,1	100	<b>54</b>	47	48	...	46	<b>48</b>	<b>88,1</b>
2ª Liga 1999/00	7	1	0,1	100	<b>63</b>	47	55	...	54	<b>52</b>	<b>82,4</b>
2ª Liga 1999/00	8	1	0,1	100	<b>72</b>	59	56	...	57	<b>56</b>	<b>77,9</b>
2ª Liga 1999/00	26	1	0,1	80	<b>234</b>	162	164	...	157	<b>161</b>	<b>68,6</b>
2ª Liga 1999/00	31	1	0,1	80	<b>279</b>	186	189	...	176	<b>184</b>	<b>65,8</b>
2ª Liga 1999/00	32	1	0,1	80	<b>288</b>	184	190	...	190	<b>188</b>	<b>65,2</b>
2ª Liga 1999/00	33	1	0,1	80	<b>297</b>	201	201	...	200	<b>200</b>	<b>67,3</b>
2ª Liga 1999/00	34	1	0,1	80	<b>306</b>	200	203	...	198	<b>203</b>	<b>66,2</b>

Tabela 4.2 – Vários testes para várias jornadas

No fim da tabela 4.1 podemos ainda ver os resultados da estratégia das “equipas dinâmicas” que não se revelou a ser uma boa estratégia a ver pelo *mérito* alcançado.

Numa outra tabela 4.2 podemos ver a variação percentual do *mérito* em função das jornadas. Primeiro temos a análise feita para o fim do campeonato da primeira liga de 1998/99, atingindo-se uma percentagem de 67,4%.

Aqui os parâmetros do algoritmo permanecem inalterados à excepção da percentagem para bom *ppbom* que no início da época é de 100%, e que várias equipas (ou todas, nas primeiras jornadas) a atingem. No fim do campeonato podemos descer essa percentagem para os 80% em que quase nenhuma equipa o atinge equivalendo-se assim a um algoritmo sem a “protecção dos genes bons”.

De seguida temos uma análise feita para a primeira e segunda liga do campeonato de 1999/00. Como se pode ver pelo gráfico 4.1 e 4.2 os resultados apresentam uma lacuna no tempo (em jornadas), pelo que foram apresentados em dois gráficos, no entanto pode-se ter uma ideia da evolução dos dados intermédios, ou seja a descida do *mérito* ao longo das jornadas.

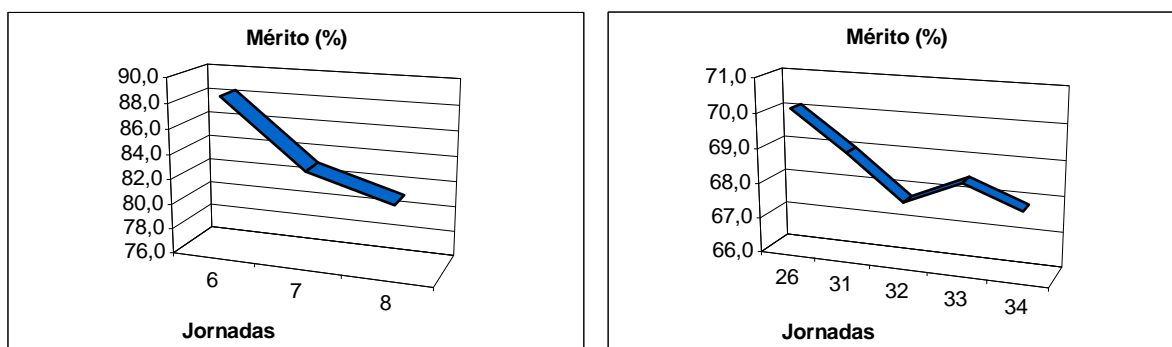


Gráfico 4.1 – Campeonato da 1ª Liga.

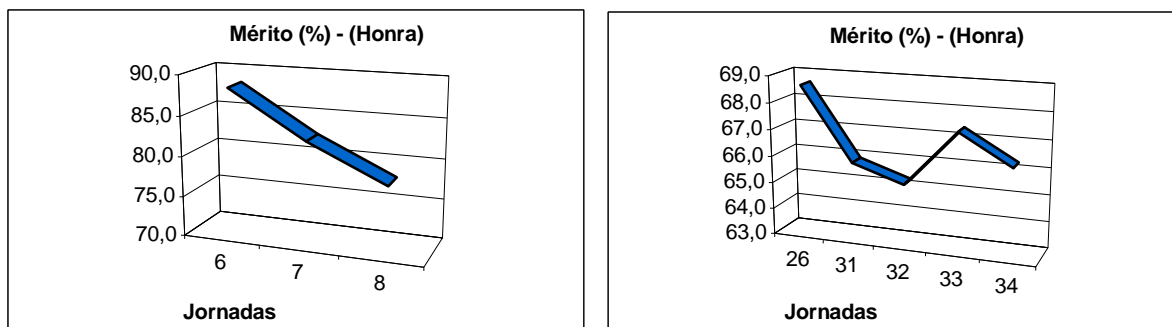


Gráfico 4.2 – Campeonato da 2ª Liga.

Nestes gráficos podemos ver que a percentagem média do *mérito* atingido desce desde os 100% ou perto disso para estabilizar por volta dos 67%. Este último valor é válido para os 3 campeonatos analisados, na tabela 4.2.

Como dissemos no início do relatório a percentagem aleatória de acertarmos num jogo é de 1/3 ou seja 33,3% aqui duplicou-se essa percentagem. Isto pode não ser muito explícito, mas em termos de probabilidades de ganhar no Totobola temos  $3^{14} \left( \frac{100}{33,3} \right)^{14}$  ou seja 4 782 969 hipóteses de acertar num 14. Por outro lado com 67% de hipóteses de acertar num jogo as coisas ficam diferentes, pois ficamos apenas com  $\left( \frac{100}{67} \right)^{14}$  ou seja 272 hipóteses. Comparativamente ao Totoloto que são de 13 983 816  $\left( \frac{49!}{6! \cdot 43!} \right)$  hipóteses de acertar num 6, podemos dizer que é preferível apostar no Totobola usando o programa.

O programa pode ser executado várias vezes para obter uma aposta com duplas e triplas, já que nem sempre o algoritmo constrói a mesma lista apesar de o *mérito* alcançado poder ser o mesmo.

Restam sempre os jogos que não respeitam a propriedade transitiva que acabam por ser as chamadas surpresas no campeonato já que as listas criadas pelo algoritmo revelam algo que a tabela classificativa também revela que são uma ordem para as equipas tornando-as em melhores e piores equipas comparativamente.

Pode-se concluir que, sendo o objectivo do Totobola o de acertar em 14 jogos, ou de 13 e 12 jogos no caso do segundo e terceiros prémios, e do facto de termos 67% dos jogos

certos, em média, garantindo-se portanto 9 jogos certos com a utilização do programa, não nos é suficiente para garantir um prémio. Estamos a este nível na presença de um sistema discreto a onde, ou se acerta num 12, obtendo-se o 3º prémio, ou não se ganha nada, pelo contrário perde-se o dinheiro da aposta. Para alcançar este nível seria necessário termos uma garantia de acertar de 86%, o que ainda se encontra muito longe dos 67% conseguidos pelo programa.

Os quase 20% que falta para obter um prémio não se conseguiram contornar o que não significa que a utilização do programa não se venha a verificar a ser uma boa ideia, pelo menos para se ter uma segunda opinião.

De notar por fim, que no boletim que aparece na figura 2.1 se obteve 10 jogos certos, 8 dos quais respeitantes à 1ª Liga.

## 4.2 Melhorias futuras

Colocar o factor motivação da equipa, que dependeria da previsão dos resultados possíveis na parte final do campeonato, quer para não descer de divisão, como para ficar nas primeiras posições afim de poderem ir às taças europeias ou/e ganharem o campeonato.

Acertar em jogos do estilo da propriedade transitiva. Se a equipa A ganha à B e a B ganha à C, então pelo sistema linear temos que A ganharia à C, o que pode não acontecer. Talvez se considerar-mos as equipas distribuídas num plano.

Podia-se também introduzir o conceito, existente no Xadrez , chamado Elo. Neste sistema cada jogador possui um valor (Elo) de entrada e conforme ganha, empata ou perde assim também lhe é adicionado, ou subtraído uma quantidade dependente da diferença entre os elos dos dois jogadores.

# Índice

Totobola por Alorítmos Evolutivos.....	1
1. Introdução.....	1
2. O Jogo .....	2
2.1 A História .....	3
2.2 A Ideia .....	6
3. O Programa .....	8
3.1 O Algoritmo Genético.....	8
3.2 A Aplicação.....	10
3.3 As melhorias à aplicação.....	14
4. A Conclusão .....	17
4.2 Melhorias futuras.....	21