

G-Loto

Abstract

Este documento é sobre o projecto «G-Loto», em que se procura gerar apostas de *Totoloto*, que aumentem a possibilidade de prémio de um concorrente. Entende-se por *Totoloto*, um jogo em que são sorteadas sequências de números inteiros aleatórios, sem repetições. O projecto «G-Loto» é válido para qualquer concurso de *Totoloto*, independentemente do tamanho ou composição da sequência aleatória.

1. Introdução

Uma vez que qualquer *Totoloto* é, por natureza, *absolutamente aleatório*, as soluções informáticas clássicas para apostar, são meros geradores de números inteiros, quase sempre limitadas a cenários concretos, como o *Totoloto* inglês ou o *Totoloto* português, por exemplo. Esta aproximação ao problema não tem nada de errado, se se fizer fé limitada ao que está *estabelecido*, para o conceito de *absolutamente aleatório*, que é a ausência total de previsibilidade.

O projecto «G-Loto» tem uma nova atitude perante o conceito de *absolutamente aleatório*...

Em Informática digital não existem geradores genuínos de números aleatórios; toma-se sempre uma raiz de geração, que é normalmente o estado de um relógio suportado em hardware. Ao fim de intervalos de tempo muito precisos, os resultados tendem a repetir-se.

A nova atitude do projecto «G-Loto», consiste em acreditar que também na vida quotidiana, os processos aleatórios têm uma origem ou relação tal com a dimensão Tempo, que, em certas circunstâncias, tendem a repetir a sua história passada. *Exactamente* quais essas origens e/ou relações, é coisa que permanece sem resposta e que fica muito para lá do objectivo modesto e focado do «G-Loto».

A Natureza está recheada de demonstrações de fenómenos periódicos, desde as estações meteorológicas, às ondas sinusoidais que neste momento sobrepovoam o espaço hertziano... O ser humano, por mais que o esqueça, não passa de uma entidade natural, restrita a características que são produto de uma evolução muito específica. A sua percepção do meio condiciona grandemente as suas respostas, para os seus problemas. Por exemplo, só muito recentemente, com Sir Isaac Newton, coisas que hoje temos como triviais, como o efeito da força da gravidade, foram efectivamente *observadas*, por omnipresentes que tenham sido, e sejam, desde o momento zero, se é que o houve...

O projecto «G-Loto», assume pois que *absolutamente aleatório* também pode significar, «capaz de se repetir, total ou parcialmente», que é, aliás, qualquer coisa de implícita à definição clássica do termo «aleatório», por contraditório que pareça. Esta assumpção acontece sem demonstração.

2. Esforço de Observação

Num esforço para diferenciar e valorizar distintamente sequências de números inteiros aleatórios, o projecto «G-Loto» utiliza uma função matemática, totalmente baseada num historial de «chaves», entendendo-se por «chaves», sequências de números inteiros aleatórios já sorteados, no âmbito de um concurso de *Totoloto*.

Esta função matemática será a função objectivo de um algoritmo genético simples, que terá como ponto de partida, ou população inicial, uma família de sequências ou «apostas», que serão avaliadas e seleccionadas, de forma a envolverem-se num processo de evolução, do qual se espera que resulte uma população mais provável.

São consideradas 34 classes de informação, a partir do historial de chaves.

Por cada chave que conste do historial de chaves já sorteadas:

- A soma dos números da chave.
- A média dos números da chave.
- O número mínimo da chave.
- O número máximo da chave.
- A amplitude matemática da chave (MAX-MIN).
- O número de colunas diferentes que a chave exige, num boletim físico de *Totoloto*.
- O número de linhas diferentes que a chave exige, num boletim físico de *Totoloto*.
- O número de números pares, na chave.
- O número de números ímpares, na chave.
- O ano, o número de concurso e o dia, da chave.
- A lista de número propriamente ditos, que compõem a chave.
- A distância cronológica (em Semanas / Dias), da chave, em relação à chave mais recente do historial.

(12)

Para o conjunto de TODAS as chaves:

- A média da soma das chaves.
- A média da média das chaves.
- A média dos máximos das chaves.
- A média dos mínimos das chaves.
- A média da amplitude matemática das chaves.
- A média do número de colunas diferentes, que as chaves exigem, em boletins físicos de *Totoloto*.
- A média do número de linhas diferentes, que as chaves exigem, em boletins físicos de *Totoloto*.
- A média dos números pares, nas chaves.
- A média dos números ímpares, nas chaves.
- A lista de todos os números propriamente ditos, que alguma vez fizeram parte de alguma chave.
- A lista de quantas vezes saiu cada número que alguma vez foi sorteado, em alguma chave.
- A média de quantas vezes foram sorteados os números, que foram sorteados em chaves.
- A lista de todos os números sorteados mais vezes do que a média.
- A lista de todos os números sorteados menos vezes do que a média.
- A chave média, de todas as chaves sorteadas, ponderada por índice.
- A informação cronológica de todos os números que alguma vez foram sorteados.

(16)

Para cada número, alguma vez sorteado:

- Quantas vezes o número foi sorteado.
- A lista da evolução das ocasiões em que o número foi sorteado, na forma de uma recta temporal crescente, indexada por Semanas / Dias.
- A lista da evolução das distâncias das ocasiões em que o número foi sorteado, relativamente ao sorteio mais recente do historial. Esta lista está ordenada, na forma de uma recta temporal de recência crescente; isto é, as menores distâncias cronológicas, para a chave mais recente, vêm no final.
- A soma das ocasiões em que o número foi sorteado, de acordo com a respectiva lista de sorteios. Um valor absoluto elevado desta soma *pode* significar que o

número foi sorteado muitas vezes. Um valor relativo elevado desta soma é indicador de que o número tem um historial de sucessos.

- A soma das distâncias das ocasiões em que o número foi sorteado, relativamente à chave mais recente. Um valor modesto desta soma, *pode* significar que o número foi sorteado muito recentemente. Um valor relativo elevado desta soma, é indicador da frequência recente de sucessos do número.

(6)

Este total de 34 informações de natureza distinta, é ponderado por uma fórmula de pesos variáveis. Por defeito, utiliza-se a seguinte função objectivo F , no caso da valoração de um sequência de n números inteiros aleatórios:

$$F = 0.5 * F1 + 0.5 * F2$$

$$F1 = \left(\frac{f1 + f2 + \dots + fn}{n} \right)$$

$$F2 = 0.2 * Sum + 0.1 * (Max + Min + Amp + Evens + Odds + Average + Cols + Lines)$$

As parcelas $f1$ a fn , consistem na avaliação de cada número individual da sequência, e são calculadas da seguinte forma:

$$fn = 0.2 * A1 + 0.2 * A2 + 0.1 * A3 + 0.1 * A4 + 0.25 * A5 + 0.25 * A6$$

$$A1 = \frac{\sum Tn}{Semana} \quad A2 = \frac{\sum Dn}{Semana} \quad A3 = \frac{\sum Tn}{Dia} \quad A4 = \frac{\sum Dn}{Dia}$$

$$A5 = \frac{\sum T}{D \max(semanas)} \quad A6 = \frac{\sum D}{D \max(dias)}$$

$A1$ representa a percentagem de sucesso, em semanas, do número n , relativamente a um número hipotético que tivesse saído em TODAS as semanas que constam do historial. O T representa a recta temporal de chaves.

$A2$ representa a percentagem de distância para a chave mais recente, em semanas, do número n , relativamente a um número hipotético, que tivesse saído em TODAS as semanas, que constam do historial. O D representa a evolução das distâncias, da recta temporal de chaves, relativamente à chave mais recente do historial.

$A3$ e $A4$ são análogos de $A1$ e $A2$, mas relativamente ao dia da semana.

$A5$ e $A6$ representam o valor relativo das distâncias médias, em semanas e dias, do número n , relativamente às distâncias máximas, na mesma unidade.

$A1$, $A2$, $A3$ e $A4$, tendem a avaliar um número n , pelo seu sucesso; isto é, números que constem frequentemente do historial de chaves disponível, serão melhor cotados.

$A5$ e $A6$, tendem a avaliar um número n , pelo seu insucesso; isto é, número invulgares, relativamente ao historial de chaves disponível, tendem a ser melhor cotados.

Sum , Max , Min , Amp , $Evens$, $Odds$, $Average$, $Cols$ e $Lines$, são válidos, não para um número n singular, mas para o conjunto de números da aposta. Estes parâmetros medem a

percentagem de proximidade da aposta em causa, relativamente aos valores médios obtidos para o total do historial de chaves.

Qualquer um destes parâmetros, vê a sua percentagem de proximidade, medida pela seguinte expressão, indicada para um parâmetro genérico P:

$$P = 1 - \frac{|\bar{P} - V(\text{aposta})|}{\bar{P}}$$

V(aposta) representa a valoração da aposta, para o parâmetro.

Por exemplo, se o valor médio para um parâmetro, relativamente ao historial de todas as chaves disponíveis, é 150, e se uma aposta aleatória em particular, é medida também em 150 para esse parâmetro, então o valor percentual da sua proximidade à média é:

$$P = 1 - |(150-150)/150| = 1 - 0 = 1$$

Ou seja, essa aposta seria medida como absolutamente coincidente, relativamente à média do parâmetro.

Sum representa a proximidade, relativamente à soma média.

Max representa a proximidade, relativamente ao valor máximo médio.

Min representa a proximidade, relativamente ao valor mínimo médio.

Amp representa a proximidade, relativamente ao valor da amplitude matemática média.

Evens representa a proximidade, relativamente ao valor médio de números pares.

Odds representa a proximidade, relativamente ao valor médio de números ímpares.

Average representa a proximidade, relativamente ao valor médio das médias.

Cols representa a proximidade, relativamente ao valor médio do número de colunas exigido.

Lines representa a proximidade, relativamente ao valor médio do número de linhas exigido.

3. Flexibilidade e Genética

O projecto «G-Loto» foi concebido de forma a ser o mais versátil possível. O programa é totalmente dinâmico, com alocação de memória em tempo real, sendo pois capaz de representar e trabalhar com um historial de chaves potencialmente ilimitado. Não há limites para o número de números por aposta, e a configuração gráfica das matrizes físicas do *Totoloto*, é definível, de forma a que parâmetros como *Cols* e *Lines* não sejam afectados por boletins futuros, ou de diferentes partes do mundo.

Todas as opções que afectem qualquer algoritmo do programa são parametrizáveis [também] por ficheiros externos à aplicação.

Todo os *outputs* do programa podem acontecer para ficheiros, e algumas operações geram automaticamente ficheiros.

A forma como o programa foi escrito, permite o *design* fácil de interfaces gráficas de alto nível, que apenas precisarão de conhecer as possíveis entradas e saídas do «G-Loto», na forma de ficheiros, de forma a permitirem toda a interacção tolerada pelo programa.

Na sua versão final, o «G-Loto» apresentará uma interface textual e uma interface gráfica para *MS-Windows*.

Para lá da geração de apostas aconselhadas, relativamente a um historial de chaves, o «G-Loto» permite ainda:

- A correcção automática de apostas, relativamente a chaves.
- A geração de estatísticas e avaliações cronológicas, para leitura, pelo apostador.

- Editores de alto nível para ficheiros que contêm apostas, chaves e configurações do utilizador.

Utiliza-se um algoritmo genético, para chegar ao objectivo máximo do programa, que é a geração de uma família de apostas, potencialmente mais prováveis, relativamente ao historial de chaves disponibilizado:

- Parte-se de uma população inicial aleatória...
- ...cujos elementos são avaliados pela função objectivo já descrita.
- Os melhores 30% elementos estão garantidos na geração seguinte.
- Os restantes 70% elementos da próxima geração, são obtidos em sorteio por roleta. Neste processo são escolhidos os «pais» das apostas que seguem, e que serão cruzados em 50% da dimensão do respectivo cromossoma. Por exemplo, para sequências de tamanho 6 (seis números), serão cruzados 3 dos números que compõem a aposta. As posições que cruzam são aleatórias, o que significa que se utiliza uma máscara *random*.
- Daqui se depreende que também o cromossoma com que o algoritmo genético trabalha é flexível; isto é, não se está limitado a apostas de X números; pode-se trabalhar com apostas (e chaves), de qualquer dimensão.
- O algoritmo pára ao fim de $100 * \text{Número_de_chaves_no_historial}$ gerações. Por exemplo, para 84 chaves conhecidas, a família de apostas aconselhadas, será gerada ao cabo de 8400 gerações / iterações. Todavia, este valor (100) também é definível. Nada é estático no «G-Loto».
- Optou-se por uma condição de paragem deste género, em alternativa a um certo valor médio do *fitness* da população, porque, para o problema em causa, é impossível saber quais as apostas que merecem um 100% de avaliação. Deposita-se confiança na função objectivo, mas sem esquecer o enquadramento da questão, e procurando manter a aleatoriedade em diversos aspectos.

4. Porquê? Crítica.

Os pesos escolhidos por defeito, o critério de paragem e diversos outros detalhes, foram implementados da forma descrita, pela mais natural das justificações: o raciocínio subjacente ao processo permitiu a um dos autores do programa ser totalista (6 acertos, *TotoLoto* português, 1988)! Curiosamente, o prémio atribuído foi o menor de sempre, da história do concurso...

Registaram-se ainda dois terceiros prémios (5 acertos), noutras tantas ocasiões.

É evidente que os autores do «G-Loto» não acreditam que exista um algoritmo para a *Sorte*, e não estão convencidos de estarem na posse de uma «galinha dos ovos de ouro». Nada disso. O interesse pelo assunto é apenas a concretização de um interesse mais lato nas diversas disciplinas científicas envolvidas.

Coincidência, ou não, o algoritmo genético desenvolvido já mostrou resultados por diversas ocasiões, que mais do que justificam dar-se continuidade ao projecto «G-Loto».

5. Futuro

O futuro do «G-Loto» passa pela concepção de um meta-G-Loto; isto é, de um meta-Algoritmo Genético (meta-AG), que avalie populações de soluções, e gere automaticamente parametrizações aconselhadas, para a função objectivo.

A forma como o «G-Loto» está escrito facilita o caminhar neste sentido, mas a prioridade é agora consolidar o producto actual, principalmente ao nível da interface para com o utilizador.

A última versão do «G-Loto» e actualizações a este documento (eventualmente noutras línguas), estão disponíveis em www.ArturMarques.com, a partir de 9 de Março de 1999.

6. Referências

"Genetic Algorithms + Data Structures = Evolution Programs"
Zbigniew Michalewicz
1996, 3ª edição, Artificial Intelligence series, Springer

"Machine Learning"
Tom M. Mitchell
1997, Computer Science series, McGraw-Hill