

4 Experimentos

A parte experimental deste trabalho visa a conduzir uma avaliação de desempenho do modelo aqui proposto. Para tal, são executados experimentos comparativos entre os modelos PGLIQ e AIMGP. Estes experimentos abordam problemas de Regressão Simbólica e de Classificação Binária, uma vez que estes são os tipos de problemas mais comumente utilizadas para se avaliar o desempenho de sistemas de PG em geral [6].

A Regressão Simbólica pode ser descrita como um processo de determinação da equação que melhor se ajuste a uma amostra de dados de um processo ou sistema. Por sua vez, pode-se dizer que Classificação Binária é a tarefa de se classificar os elementos de um determinado conjunto de objetos em dois grupos, tomando como base o fato destes objetos possuírem ou não uma determinada propriedade. Ou seja, em problemas de classificação binária, a tarefa de aprendizado consiste em classificar exemplos de dados em duas categorias.

Para os problemas de Regressão Simbólica, o valor da aptidão de um indivíduo é determinado pela computação do seu erro absoluto médio (*mean absolute error* – MAE) ou do seu erro percentual absoluto médio (*mean absolute percentage error* – MAPE) relativo aos casos de aptidão, dependendo do experimento. O erro absoluto médio sobre os n casos de aptidão é definido como

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |t_i - V[0]_i|, \quad (4-1)$$

onde t_i é o valor alvo (real) do caso i e $V[0]_i$ é o valor da saída do indivíduo para este mesmo caso (valor da aproximação). Por sua vez, o erro percentual absoluto médio é determinado pela seguinte equação:

$$\text{MAPE} = 100 \times \frac{1}{n} \sum_{i=1}^n \left| \frac{t_i - V[0]_i}{t_i} \right|. \quad (4-2)$$

O conjunto de funções utilizado nos problemas de regressão simbólica é composto pelas 18 primeiras instruções relacionadas na tabela 3.1, as quais efetuam adições, subtrações, multiplicações, divisões e transferências de dados, assim como operações aritméticas e trigonométricas.

Nos problemas de classificação binária, cada caso de aptidão (amostra) da

base de dados tem sua classe definida pelos valores zero ou um. Sendo assim, a aptidão de cada indivíduo pode ser calculada a partir da taxa de acerto (TA), cujo valor percentual é calculado pela equação:

$$TA = 100 \times \frac{1}{n} \sum_{i=1}^n a_i, \quad (4-3)$$

onde a_i vale um se o indivíduo classifica corretamente o caso de aptidão i , ou zero, caso contrário. A classe que uma saída de um indivíduo representa é determinada a partir de um valor de limiar, que se encontra exatamente entre os valores zero e um, ou seja, 0,5. Sendo assim, se o valor dessa saída for maior ou igual a 0,5, considera-se que tal saída representa a classe um. Caso contrário, se a saída for menor que 0,5, essa saída representa a classe zero.

Uma outra forma de se calcular a aptidão é por meio do “erro de classificação” (EC), definido pela equação:

$$EC = 1 - \frac{1}{n} \sum_{i=1}^n a_i, \quad (4-4)$$

a partir da qual pode-se notar que, neste caso, a evolução deve minimizar a aptidão dos indivíduos, contrariamente ao caso da taxa de acerto, onde o objetivo evolutivo é a maximização da aptidão.

A função de avaliação adotada pelo sistema de referência (Discipulus™) para problemas de classificação é denominada “Acertos-Então-Erro” (*Hits-Then-Error*) [69], que é composta pelos seguintes passos:

1. Calcula-se a taxa de acerto de ambos os indivíduos envolvidos em uma comparação, conforme a equação 4-3.
2. Se os dois indivíduos “empatarem” com relação às suas taxas de acerto, o modelo prossegue para o passo 3. Se não houver empate, o passo 3 é ignorado e o indivíduo com a maior taxa de acerto é considerado o melhor.
3. Calcula-se o erro absoluto médio das saídas de ambos os indivíduos, a partir da equação 4-1, e o indivíduo que apresentar o menor valor para este erro é considerado o melhor.

De forma a se manter a neutralidade na comparação de resultados, a PGLIQ também utiliza a mesma função de avaliação do sistema Discipulus, com apenas uma modificação: a aptidão é calculada pelo erro de classificação (equação 4-4), ao invés da taxa de acerto. Tal modificação visa a uniformizar os resultados oriundos dos estudos de caso de regressão simbólica, onde a aptidão deve ser minimizada, com os de classificação binária. Em função desta alteração, faz-se necessário um

pós-processamento dos resultados do sistema Discipulus, de forma a convertê-los de taxa de acerto (TA) para erro de classificação (EC), da seguinte forma:

$$EC = 1 - \frac{TA}{100}. \quad (4-5)$$

O conjunto de funções utilizado nos problemas de classificação binária é uma extensão do conjunto para regressão simbólica, uma vez que também inclui as três últimas instruções relacionadas na tabela 3.1, as quais possibilitam a execução condicional de instruções de um programa. Isso se deve ao fato de que, normalmente, problemas de classificação são resolvidos com mais sucesso, ou até mesmo somente resolvidos, se forem fornecidas à PG instruções condicionais ou de desvio do fluxo da execução [6].

O conjunto de terminais de todos os experimentos é composto pelas respectivas variáveis de entrada e, exceto para o estudo de caso da seção 4.2.3, pelo mesmo conjunto de constantes adotado em [6]: {1, 2, 3, 4, 5, 6, 7, 8, 9}.

Cada experimento é composto por cem execuções dos modelos, de forma a se obter as curvas evolutivas com a média das aptidões dos melhores indivíduos por geração, assim como as aptidões dos melhores indivíduos dentre as cem execuções. A aptidão representada por cada curva é a média dos valores obtidos a partir dos conjuntos de treinamento e de validação.

Objetivando uma análise comparativa mais ampla da qualidade das soluções, o desempenho de ambos os modelos é aferido com base em duas referências: tempo de execução e número de indivíduos avaliados. Esta abordagem é importante porque o custo computacional por indivíduo é significativamente diferente entre os modelos, uma vez que possuem diferentes algoritmos evolutivos, representações e operadores.

Os experimentos foram conduzidos em um PC com a seguinte configuração básica:

- Processador: AMD Phenom™ II X4 965;
- Frequência de *clock*: 3,41 GHz;
- Capacidade de memória RAM: 8 GB;
- Sistema Operacional: Windows 7 Professional.

4.1 Configuração dos Modelos

4.1.1 AIMGP

Conforme mencionado anteriormente, o *software* Discipulus™ é utilizado nos experimentos deste trabalho por implementar o modelo AIMGP no seu núcleo. Sendo assim, a configuração do modelo adotada nos experimentos é descrita na tabela 4.1. Os parâmetros recebem seus valores padrão, conforme recomendado pelo manual do usuário do Discipulus™ [69]. O número de gerações por execução e o tamanho da população também foram escolhidos por se situarem em faixas de valores típicos de experimentos com PG em geral [6, 12].

Parâmetro	Valor
Número de gerações (G)	1.000
Tamanho da população (M)	1.000
Número de <i>demes</i>	10
Taxa de migração entre <i>demes</i>	1%
Comprimento máximo inicial do programa	20
Comprimento máximo do programa	128
Frequência de mutação	95%
– Taxa de mutação de bloco	30%
– Taxa de mutação de instrução	30%
– Taxa de mutação de dados	40%
Frequência de cruzamento	50%
– Taxa de cruzamento homólogo	95%
Número de registradores da FPU	8

Tabela 4.1: Configuração do modelo AIMGP para os experimentos.

Todos os estudos de caso são executados com e sem o uso de *demes* (vide seção 2.4.3). Conforme mostrado na tabela 4.1, nos experimentos em que são utilizados, os 10 *demes* possuem 100 indivíduos cada. Por sua vez, a taxa de migração define o percentual de torneios, entre dois indivíduos, que resultam na migração de indivíduos entre *demes* adjacentes. Quando um torneio é selecionado para migração, o modelo executa os seguintes passos:

- um *deme* é selecionado aleatoriamente;
- um dos dois *demes* adjacentes é selecionado aleatoriamente, com 50% de probabilidade para ambos;
- dois indivíduos são escolhidos aleatoriamente, com distribuição uniforme, de cada um dos *demes* selecionados anteriormente;
- o melhor indivíduo de cada *deme* substitui o pior indivíduo do outro *deme*.

Uma vez que o modelo PGLIQ possui apenas uma população, composta por indivíduos quânticos e clássicos, pode-se concluir que a comparação de desempenho entre os dois modelos é mais neutra ao se executar o modelo AIMGP sem *demes*, ou seja, também com uma única população. Entretanto, de forma a ampliar a análise de desempenho da PGLIQ, também são conduzidos experimentos onde a AIMGP é executada com *demes*.

A frequência de mutação determina a probabilidade de se aplicar um dos operadores de mutação aos indivíduos selecionados como vencedores em um torneio. Uma vez que esta frequência seja configurada, a aplicação da mutação é controlada por outros parâmetros.

Os programas evoluídos mantêm suas instruções dentro de blocos com 32 *bits* de comprimento. A taxa de mutação de bloco determina o percentual de operações de mutação que substituem um bloco de instruções completo por um novo bloco gerado aleatoriamente, com distribuição uniforme. Por sua vez, a taxa de mutação de instrução representa o percentual de operações de mutação que resultam na substituição de uma única instrução por uma nova instrução, de mesmo comprimento, escolhida aleatoriamente. Finalmente, a taxa de mutação de dados determina a probabilidade de que a operação de mutação substitua apenas o argumento de uma instrução, que pode ser um registrador ou uma posição de memória (entradas ou constantes).

O cruzamento homólogo é um operador que busca reproduzir a evolução natural de uma forma mais próxima que o cruzamento tradicional o faz. No cruzamento homólogo, os dois indivíduos evoluídos são alinhados entre si, a partir das suas respectivas primeiras instruções. O cruzamento ocorre então pela troca de grupos de blocos de instrução contíguos entre os dois indivíduos. Estes grupos são escolhidos de forma que os grupos de cada genitor tenham o mesmo comprimento e sejam extraídos a partir da mesma posição em ambos os genitores evoluídos. Sendo assim, os descendentes resultantes de um cruzamento homólogo possuem o mesmo comprimento dos seus genitores.

A taxa de cruzamento homólogo determina o percentual de eventos de cruzamento que são homólogos, em oposição ao percentual dos que não são homólogos (tradicionais). O cruzamento tradicional é o único recurso que o modelo possui para alterar o tamanho dos indivíduos evoluídos, uma vez que os grupos de blocos de instruções são trocados entre os dois indivíduos sem nenhuma referência quanto aos seus tamanhos e localizações, o que geralmente acarreta na criação de descendentes de tamanhos diferentes dos de seus genitores.

4.1.2 PGLIQ

A configuração do modelo PGLIQ é mostrada pela tabela 4.2, onde os parâmetros da segunda à sexta linha recebem seus melhores valores, determinados a partir de experimentos preliminares.

Parâmetro	Valor
Número de gerações (G)	300.000 a 350.000
Tamanho da população (M)	6
Número máximo de gerações sem melhoria ($gsm_{máximo}$)	20.000
Probabilidade inicial de NOP ($p_{0,0}$)	0,9
Tamanho do passo de reinício (s_r)	1,0
Tamanho do passo do operador P (s)	0,004
Comprimento máximo do programa	128
Número de registradores da FPU	8

Tabela 4.2: Configuração do modelo PGLIQ para os experimentos.

O número de indivíduos avaliados por execução em um algoritmo evolutivo é, geralmente, calculado de maneira trivial, através do produto do número de indivíduos da população (M) pelo número de gerações (G). É importante citar que este cálculo é válido inclusive para as publicações de Nordin, nas quais o modelo AIMGP foi originalmente apresentado [54, 60, 17] e aprimorado [55, 71].

Entretanto, a partir dos experimentos realizados neste trabalho, pode-se constatar que, no caso do sistema Discipulus, o número de indivíduos avaliados por execução não corresponde ao produto $M \times G$. O que ocorre é que, apesar de $M = 1.000$ e $G = 1.000$ (vide tabela 4.1), ao final de cada execução, o sistema Discipulus™ sempre indica ter avaliado um número de indivíduos significativamente superior ao valor esperado de $M \times G = 1.000.000$. Este valor varia entre as execuções de um mesmo experimento e, com maior amplitude, entre experimentos com estudos de casos diferentes. Dessa forma, pode-se concluir que o número de indivíduos avaliados por geração pelo Discipulus™ é variável, não havendo nenhuma explicação a respeito deste fato em nenhum manual do produto.

A forma aqui adotada para contornar essa situação consta na variação do número de indivíduos avaliados pela PGLIQ, variando-se o número de gerações G , em função do estudo de caso. Dessa forma, garante-se que a PGLIQ avalie, no mínimo, o mesmo número de indivíduos que o Discipulus™ para aquele mesmo estudo de caso.

4.2 Estudos de Caso

4.2.1 Distância Euclideana

Este problema de regressão, utilizado como *benchmark* em [6], requer que a programação genética gere um programa que calcule a distância euclideana entre dois pontos a e b no espaço tridimensional, o que representa um problema de 6 variáveis de entrada, conforme mostra a equação:

$$f_{dist}(a_x, b_x, a_y, b_y, a_z, b_z) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2 + (a_z - b_z)^2}. \quad (4-6)$$

Cada um dos três conjuntos de casos de aptidão (treinamento, validação e teste) é composto por 300 amostras, onde os valores das coordenadas dos pontos ($\in \mathbb{R}^3$) são aleatórios e se situam no intervalo $[0, 1]$, com distribuição uniforme.

Diferentemente dos demais estudos de caso de regressão simbólica, este primeiro utiliza um conjunto de funções reduzido, composto pelas seguintes instruções (vide tabela 3.1): FADD, FSUB, FMUL, FXCH, FABS e FSQRT.

A figura 4.1 mostra as curvas evolutivas referentes à média do melhor indivíduo por geração para os modelos AIMGP, com e sem o uso de *demes*, e PGLIQ. Este gráfico permite a comparação dos resultados em função do tempo médio de cada execução dos modelos.

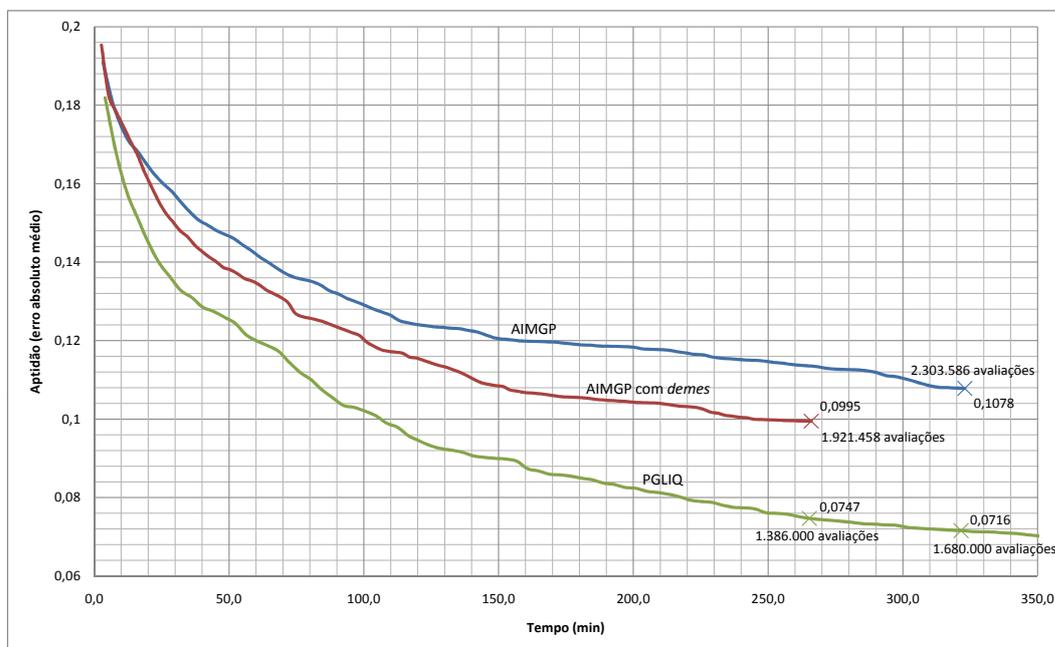


Figura 4.1: Gráfico evolutivo do estudo de caso “Distância Euclideana” em função do tempo médio de cada execução dos modelos.

Também a partir da figura 4.1, pode-se observar que a execução de 1.000 ge-

rações com 1.000 indivíduos pela AIMGP acarreta na avaliação média de 2.303.586 indivíduos, em torno de 325 minutos, sem *demes*, e de 1.921.458 em aproximadamente 265 minutos, com *demes*. No ponto da curva da PGLIQ equivalente ao ponto da AIMGP sem *demes* (325 minutos), a PGLIQ tem avaliados 1.680.000 indivíduos, enquanto que no ponto equivalente ao da AIMGP com *demes* (265 minutos), a PGLIQ tem 1.386.000 indivíduos avaliados.

Pela análise destes primeiros valores, nota-se que o custo computacional da PGLIQ para avaliar um indivíduo é mais elevado em relação ao da AIMGP. Entretanto, neste estudo de caso, pode-se notar também que, apesar da PGLIQ conseguir avaliar um número menor de indivíduos por unidade de tempo que a AIMGP, a eficiência da PGLIQ é maior, pois obtém resultados consideravelmente melhores, com um número menor de avaliações.

A figura 4.2, assim como a figura 4.1, também mostra as curvas evolutivas referentes à média do melhor indivíduo por geração, mas toma como referência para a comparação o número de indivíduos avaliados pelos modelos.

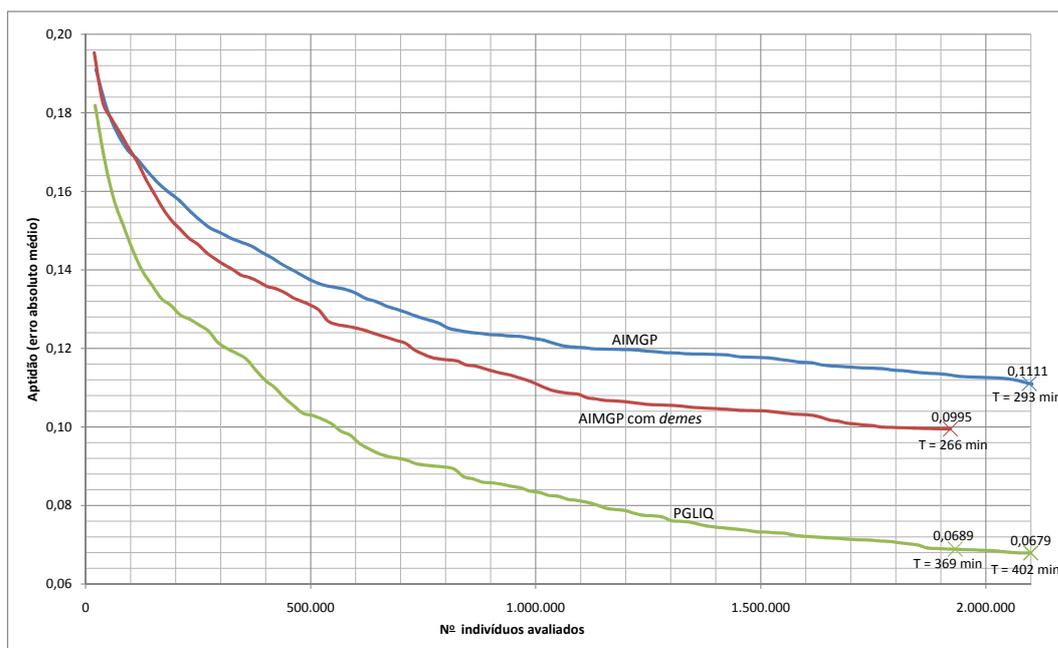


Figura 4.2: Gráfico evolutivo do estudo de caso “Distância Euclideana” em função do número médio de indivíduos avaliados a cada execução dos modelos.

A tabela 4.3 concentra os resultados obtidos a partir dos dados ilustrados pelas figuras 4.1 e 4.2, apresentando também o desvio padrão σ relativo às aptidões dos melhores indivíduos por geração, para as 100 execuções de cada modelo. A penúltima coluna à direita mostra os valores percentuais das reduções dos erros médios obtidos pela PGLIQ sobre o modelo AIMGP, que variam de 24,9 a 38,8%.

A tabela 4.4 mostra, para ambos os modelos, o número de indivíduos que, dentre as 100 execuções, representam a solução exata do problema (erro $< 10^{-6}$).

Critério de comparação	Demes	Aptidão					
		AIMGP		PGLIQ		Variação (%)	
		Média	σ	Média	σ	Média	σ
Tempo	Não	0,1078	0,0490	0,0716	0,0470	-33,6	-4,1
Nº de avaliações	Não	0,1111	0,0490	0,0679	0,0471	-38,8	-4,0
Tempo	Sim	0,0995	0,0495	0,0747	0,0474	-24,9	-4,2
Nº de avaliações	Sim	0,0995	0,0495	0,0689	0,0476	-30,7	-3,8

Tabela 4.3: Comparação de média e desvio padrão das aptidões dos melhores indivíduos por geração para o estudo de caso “Distância Euclideana”.

A última coluna à direita mostra em quantas vezes o número destes acertos obtidos pela PGLIQ supera os obtidos pela AIMGP, cujos valores variam de 4,3 a 8,5 vezes.

Critério de comparação	Demes	Número de acertos		
		AIMGP	PGLIQ	Variação (×)
Tempo	Não	2	13	6,5
Nº de avaliações	Não	2	17	8,5
Tempo	Sim	3	13	4,3
Nº de avaliações	Sim	3	17	5,7

Tabela 4.4: Comparação do número de acertos obtidos no estudo de caso “Distância Euclideana”.

4.2.2 Chapéu Mexicano

Neste problema de regressão, que também é *benchmark* em [6], a PG tem como tarefa a reconstrução de uma superfície a partir de um dado conjunto de pontos. A superfície é dada pela função bidimensional denominada “Chapéu Mexicano”, definida pela equação:

$$f_{chapmex}(x,y) = \left(1 - \frac{x^2}{4} - \frac{y^2}{4}\right) \times e^{\left(-\frac{x^2}{8} - \frac{y^2}{8}\right)}. \tag{4-7}$$

A figura 4.3 mostra um gráfico tridimensional da função, fornecendo uma visualização da superfície que deve ser aproximada.

As coordenadas x e y do conjunto de treinamento (entradas) constituem uma grade com $17 \times 17 = 289$ pontos uniformemente espaçados (casos de aptidão) no intervalo $[-4, 4]$, conforme mostra a figura 4.3. Os conjuntos de validação e de treinamento possuem, cada um, $16 \times 16 = 256$ casos de aptidão, cujas coordenadas também são uniformemente espaçadas dentro do mesmo intervalo.

A figura 4.4 mostra as curvas evolutivas dos modelos em função do tempo de duração médio de cada execução, enquanto que a figura 4.5 o faz em função do número médio de avaliações.

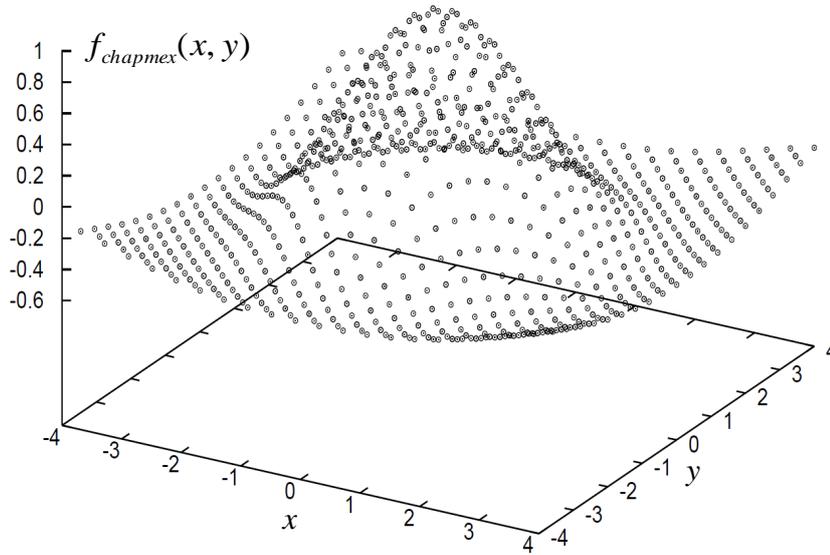


Figura 4.3: Superfície representando a função “Chapéu Mexicano”.

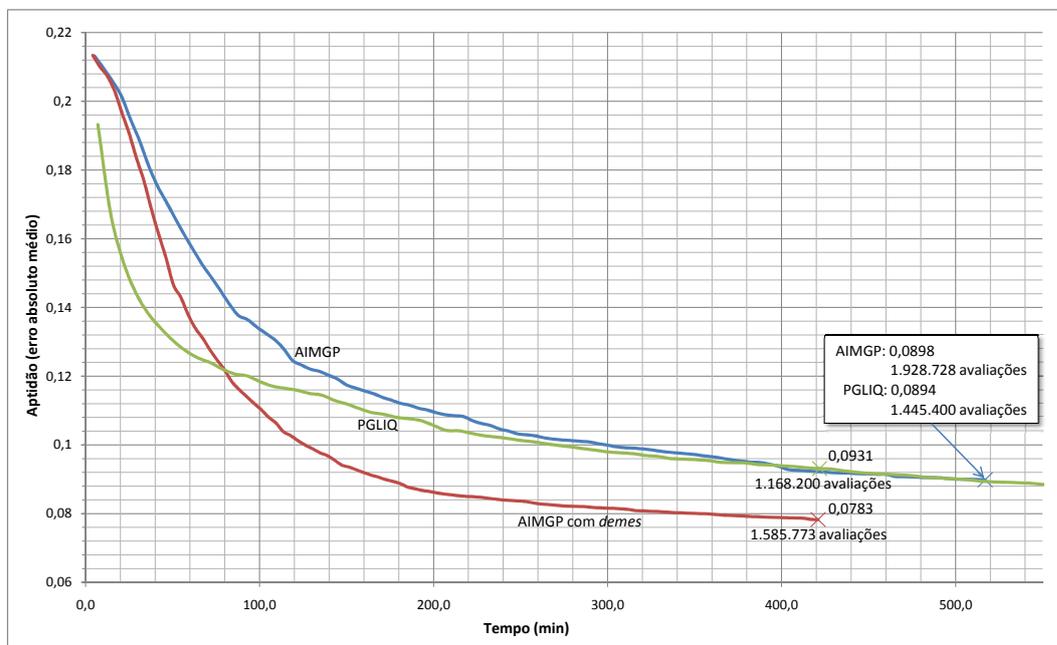


Figura 4.4: Gráfico evolutivo do estudo de caso “Chapéu Mexicano” em função do tempo médio por execução.

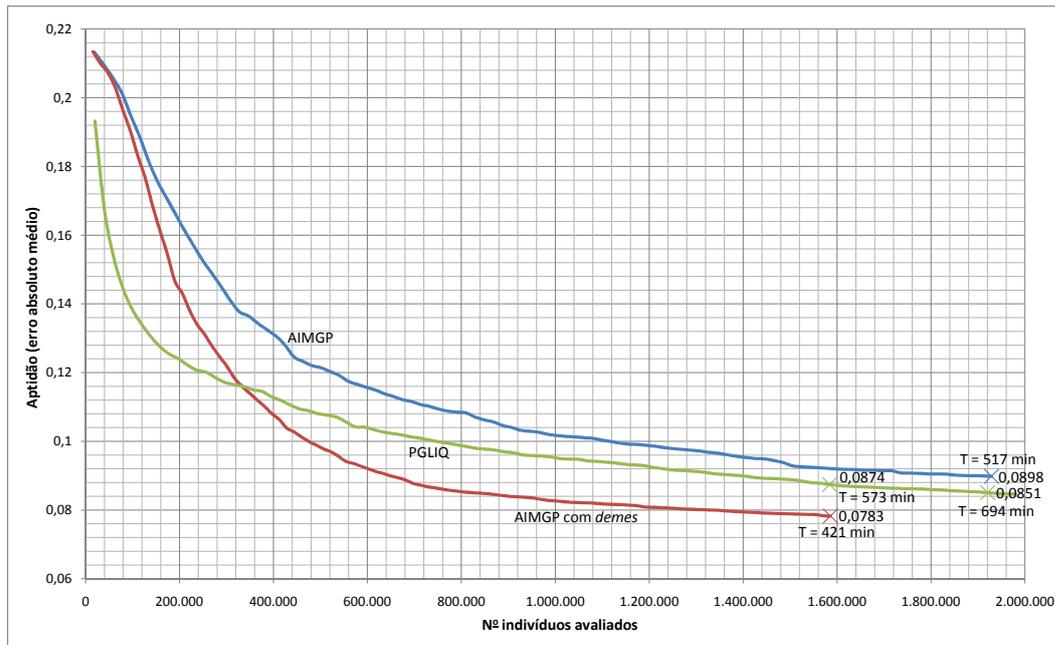


Figura 4.5: Gráfico evolutivo do estudo de caso “Chapéu Mexicano” em função do número médio de indivíduos avaliados por execução.

A tabela 4.5 concentra os resultados obtidos a partir dos dados ilustrados pelas figuras 4.4 e 4.5, apresentando também o desvio padrão σ relativo às aptidões dos melhores indivíduos por geração. A penúltima coluna à direita mostra que, quanto à qualidade média das soluções, a PGLIQ supera a AIMGP em 5,6 e 0,6%, quando esta é executada sem *demes* (apenas uma população única com 1.000 indivíduos). Entretanto, o desempenho da PGLIQ é inferior ao da AIMGP em 11,7 e 19,0%, quando é executada com *demes* (1.000 indivíduos divididos em 10 *demes* com 100 indivíduos cada).

Critério de comparação	Demes	Aptidão					
		AIMGP		PGLIQ		Variação (%)	
		Média	σ	Média	σ	Média	σ
Tempo	Não	0,0898	0,0490	0,0893	0,0248	-0,6	-49,4
Nº de avaliações	Não	0,0898	0,0490	0,0848	0,0240	-5,6	-51,0
Tempo	Sim	0,0783	0,0495	0,0931	0,0255	19,0	-48,4
Nº de avaliações	Sim	0,0783	0,0495	0,0874	0,0246	11,7	-50,4

Tabela 4.5: Comparação de média e desvio padrão das aptidões dos melhores indivíduos por geração para o estudo de caso “Chapéu Mexicano”.

Quanto às aptidões dos melhores indivíduos encontrados pelos modelos, e considerando-se os quatro critérios de comparação, o desempenho da PGLIQ, neste estudo de caso, é significativamente inferior ao da AIMGP. Os erros obtidos pelos melhores indivíduos da PGLIQ, para os três conjuntos de casos de aptidão, são de 99 a 219% maiores que os erros dos melhores indivíduos da AIMGP.

Apesar do desempenho inferior apresentado neste estudo de caso, é importante citar que se encontra em desenvolvimento uma versão da PGLIQ que, a exemplo do modelo AIMGP, também faz uso de *demes* para melhorar seu desempenho. No caso da PGLIQ, cada *deme* tem a mesma estrutura ilustrada pela figura 3.8, possuindo indivíduos clássicos e quânticos. Em linhas gerais, o funcionamento deste modelo pode ser descrito da seguinte forma:

- Cada *deme* possui um índice de identificação, de forma a se estabelecer uma vizinhança limitada e unidirecional entre os mesmos. Como consequência, estabelece-se um grafo fechado e unidirecional (em anel), onde os vértices são os *demes*, e cujas arestas definem a vizinhança para a migração unidirecional de indivíduos entre os *demes*.
- Cada *deme* é iniciado e executado paralela e independentemente, segundo o mesmo pseudocódigo da figura 3.9.
- Quando um dos *demes* atinge a condição de reinício por estagnação evolutiva ($gsm = gsm_{máximo}$), este *deme* é reiniciado de forma semelhante à descrita pelas linhas 29 a 32 do pseudocódigo da figura 3.9 e pela seção 3.6.7, com duas diferenças básicas:
 1. Além do indivíduo C_1 da sua população clássica receber uma cópia do seu melhor indivíduo clássico evoluído até então (C_m), o indivíduo clássico C_2 recebe uma cópia do melhor indivíduo C_m do seu *deme* vizinho. Portanto, a consequência deste processo é que a população clássica do *deme* é semeada por dois indivíduos clássicos, um local e outro imigrado.
 2. Após a migração descrita no item anterior e o reinício da sua população quântica, segundo o pseudocódigo supracitado, seus indivíduos quânticos Q_1 e Q_2 são semeados, respectivamente, pelos indivíduos C_1 e C_2 da sua população clássica (vide “semeadura quântica” na seção 3.6.7). Aqui, os demais indivíduos da população quântica do *deme* não são semeados.

Os primeiros resultados obtidos por esta versão da PGLIQ são promissores, uma vez que a melhoria de desempenho obtida pela adoção dos *demes* é significativa. A tabela 4.6 mostra os resultados obtidos, quanto à média de execuções, ao se empregar o modelo com 3 *demes*, contendo 3 indivíduos clássicos e quânticos cada. Os resultados são comparados com aqueles obtidos pela AIMGP com *demes*, replicados da tabela 4.5.

Ao se comparar os resultados das tabelas 4.5 e 4.6, com a referência no tempo de execução, pode-se perceber que a PGLIQ apresenta um desempenho médio

Critério de comparação	Aptidão					
	AIMGP		PGLIQ		Variação (%)	
	Média	σ	Média	σ	Média	σ
Tempo	0,0783	0,0495	0,0792	0,0227	1,1	-54,1
Nº de avaliações	0,0783	0,0495	0,0759	0,0223	-3,0	-55,0

Tabela 4.6: Comparação de média e desvio padrão das aptidões dos melhores indivíduos por geração para o estudo de caso “Chapéu Mexicano” (ambos os modelos executados com *demes*).

apenas 1,1% pior que o da AIMGP quando ambos os modelos são executados com *demes*, contra 19,0% quando a PGLIQ é executada sem *demes*. Tomando-se como referência o número de avaliações por execução, a melhoria de desempenho é ainda maior: o desempenho médio da PGLIQ, que é inferior em 11,7% sem os *demes*, passa a ser superior em 3% ao ser executado com *demes*.

4.2.3

Composição Química

Este estudo de caso tem origem em dados laboratoriais de uma composição química (saída) que, apesar do elevado custo de obtenção, são significativamente ruidosos. Portanto, o objetivo da PG é gerar um modelo de regressão simbólica que relacione esses dados a 57 medições de processo de baixo custo, tais como temperaturas, pressões e vazões (entradas).

Trata-se de um caso baseado em dados de uma aplicação industrial real da empresa “Dow Chemical”, tendo sido utilizado em uma competição de Regressão Simbólica que ocorreu no congresso “EvoStar 2010”.

A base de dados original [72] é composta por um conjunto de treinamento com 747 amostras e um conjunto de teste com 319 amostras. Para os experimentos aqui conduzidos, o conjunto de treinamento original foi dividido, acarretando na sua redução para 374 amostras e na criação de um conjunto de validação com 373 amostras. O tamanho do conjunto de constantes utilizados neste estudo de caso é menor, uma vez que o sistema Discipulus™ tem capacidade de acessar no máximo 64 posições de memória. Como este problema possui 57 variáveis de entrada e uma de saída, só restam $64 - 57 + 1 = 6$ posições de memória para os valores das constantes. Portanto, o conjunto de constantes definido para este estudo de caso é o seguinte: $\{1,2,3,5,7,11\}$.

As figuras 4.6 e 4.7 mostram as curvas evolutivas dos modelos em função do tempo médio de duração e do número médio de avaliações de cada execução, respectivamente.

A tabela 4.7 agrupa os resultados obtidos a partir das curvas evolutivas ilustradas nas figuras 4.6 e 4.7, assim como o desvio padrão σ entre as aptidões dos

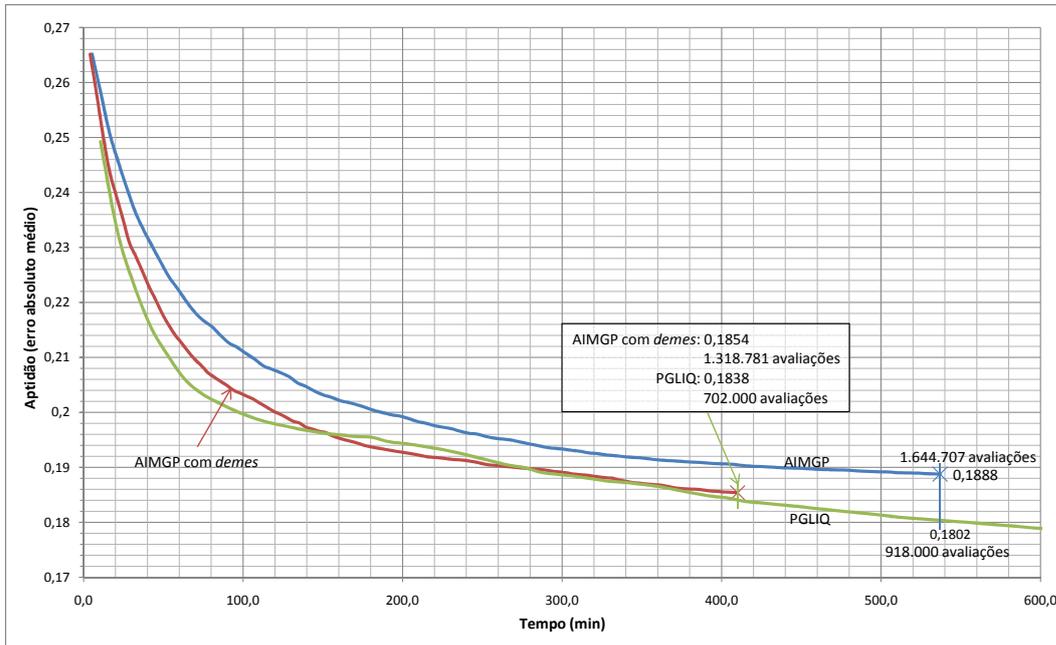


Figura 4.6: Gráfico evolutivo do estudo de caso “Composição Química” em função do tempo médio de cada execução dos modelos.

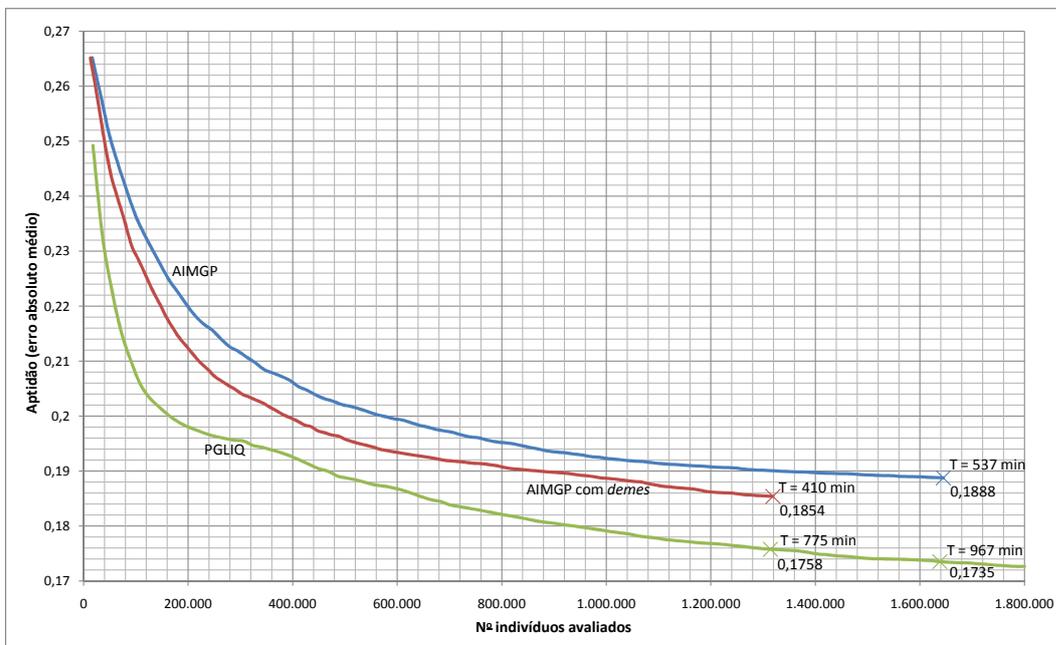


Figura 4.7: Gráfico evolutivo do estudo de caso “Composição Química” em função do número médio de indivíduos avaliados a cada execução dos modelos.

melhores indivíduos por geração. A penúltima coluna à direita mostra que, quanto à qualidade média das soluções, a PGLIQ supera a AIMGP em 4,5 e 8,1%, quando essa é executada sem *demes*, e em 0,9 e 5,2%, quando executada com *demes*.

Critério de comparação	<i>Demes</i>	Aptidão					
		AIMGP		PGLIQ		Variação (%)	
		Média	σ	Média	σ	Média	σ
Tempo	Não	0,1888	0,0246	0,1802	0,0206	-4,5	-16,3
Nº de avaliações	Não	0,1888	0,0246	0,1735	0,0210	-8,1	-14,7
Tempo	Sim	0,1854	0,0234	0,1838	0,0206	-0,9	-12,1
Nº de avaliações	Sim	0,1854	0,0234	0,1758	0,0208	-5,2	-11,0

Tabela 4.7: Comparação de média e desvio padrão das aptidões dos melhores indivíduos por geração para o estudo de caso “Composição Química”.

A tabela 4.8 compara o desempenho dos melhores indivíduos obtidos pelos modelos, em função do número médio de avaliações e do tempo médio de duração de cada execução, mostrando a aptidão dos indivíduos com relação aos conjuntos de treinamento, validação e teste. As duas últimas colunas à direita mostram que o desempenho da PGLIQ é superior ao da AIMGP para os conjuntos de validação e teste, apresentando soluções com erros de 0,8 a 11,0% menores que os da AIMGP.

Critério de comparação	<i>Demes</i>	Aptidão								
		AIMGP			PGLIQ			Variação (%)		
		Trein.	Valid.	Teste	Trein.	Valid.	Teste	Trein.	Valid.	Teste
Tempo	Não	0,126	0,131	0,140	0,129	0,129	0,131	2,0	-1,4	-5,9
Nº avals.	Não	0,126	0,131	0,140	0,116	0,121	0,124	-7,9	-7,1	-11,0
Tempo	Sim	0,120	0,139	0,143	0,130	0,130	0,142	8,6	-6,3	-0,8
Nº avals.	Sim	0,120	0,139	0,143	0,126	0,126	0,130	1,9	-9,7	-9,2

Tabela 4.8: Comparação da aptidão dos melhores indivíduos obtidos pelos modelos para o estudo de caso “Composição Química”.

4.2.4 Pontos Quânticos

Pontos quânticos são estruturas nanométricas, que confinam os elétrons nas três dimensões, acarretando na quantização completa dos níveis eletrônicos de energia. O desempenho de diversos dispositivos optoeletrônicos, tais como fotodetectores de infravermelho e emissores de laser, pode ser significativamente beneficiado pelo fato dos níveis de energia se apresentarem totalmente quantizados [73]. Entretanto, o desempenho de tais dispositivos depende fortemente do crescimento (síntese) de estruturas com alta densidade de pontos, com pontos de tamanho pequeno e que apresentem baixa dispersão quanto ao tamanho dos pontos [73].

Este é um estudo de caso de regressão simbólica, que se baseia nos trabalhos [73, 74], onde os autores utilizaram a técnica de Redes Neurais para inferir tendências no comportamento da altura média dos pontos quânticos obtidos (saída), em função de diferentes parâmetros de síntese.

A base de dados com os parâmetros de síntese é composta por 67 amostras, todas obtidas a partir de experimentos realizados no laboratório de semicondutores (LABSEM) da PUC-Rio. Os seis parâmetros (entradas) são: a vazão do elemento índio no reator, a temperatura de crescimento, o tempo de deposição, a espessura da camada sobre a qual os pontos são crescidos e as concentrações de índio e alumínio desta camada. Os conjuntos de treinamento, validação e teste possuem, respectivamente, 47, 14 e 6 amostras, como nos trabalhos de referência.

As figuras 4.8 e 4.9 mostram as curvas evolutivas dos modelos, em função do tempo médio de duração e do número médio de avaliações de cada execução, respectivamente. Também indicam os valores resultantes da comparação de desempenho dos modelos quanto à média das aptidões dos melhores indivíduos por geração, onde as soluções da PGLIQ apresentam valores de erro médio inferiores aos da AIMGP, resultando em uma superioridade de desempenho de 7,1% a 23,5%.

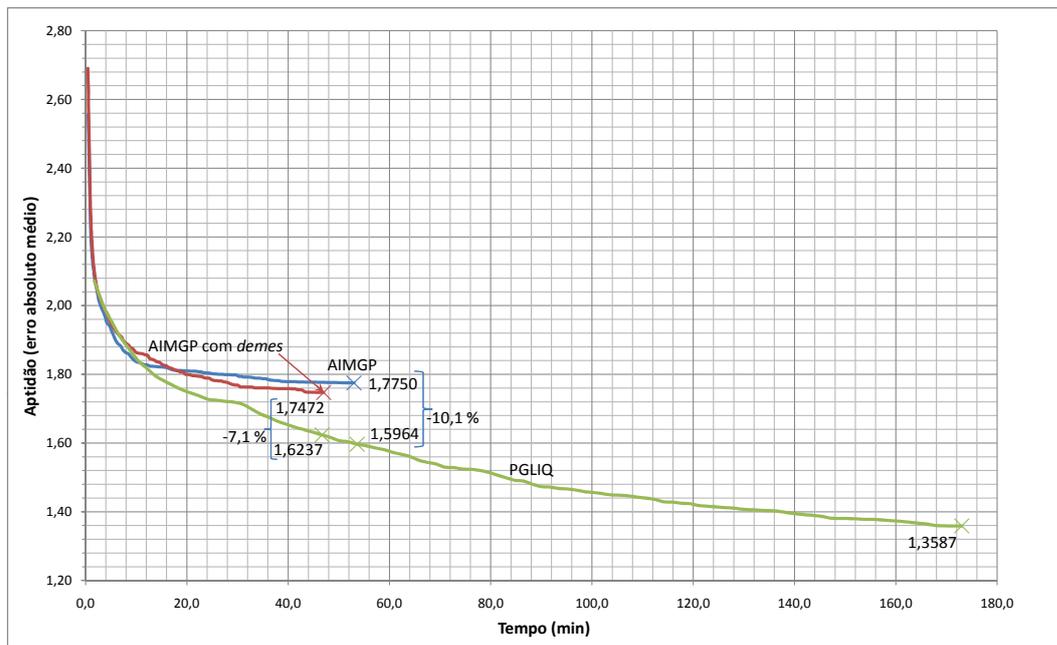


Figura 4.8: Gráfico evolutivo do estudo de caso “Pontos Quânticos” em função do tempo médio de cada execução dos modelos.

A tabela 4.9 concentra os valores de aptidão e de comparação mostrados nas figuras 4.8 e 4.9, além de apresentar os valores de desvio padrão e de suas variações percentuais decorrentes da comparação dos dois modelos.

O modelo PGLIQ também alcança um desempenho superior ao das redes neurais de [73] e [74] para o conjunto de teste. O melhor indivíduo evoluído pela

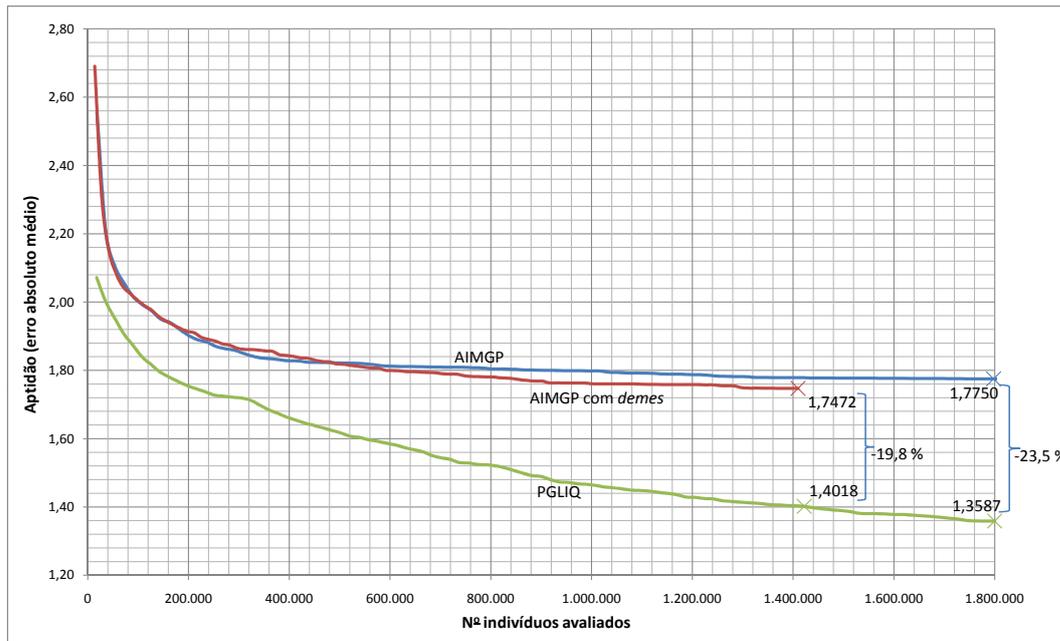


Figura 4.9: Gráfico evolutivo do estudo de caso “Pontos Quânticos” em função do número médio de indivíduos avaliados a cada execução dos modelos.

Critério de comparação	<i>Demes</i>	Aptidão				Variação (%)	
		AIMGP		PGLIQ		Média	σ
		Média	σ	Média	σ		
Tempo	Não	1,7750	0,1890	1,5964	0,1764	-10,1	-6,7
Nº Avaliações	Não	1,7750	0,1890	1,3587	0,1680	-23,5	-11,1
Tempo	Sim	1,7472	0,2616	1,6237	0,1795	-7,1	-31,4
Nº Avaliações	Sim	1,7472	0,2616	1,4018	0,1719	-19,8	-34,3

Tabela 4.9: Comparação de média e desvio padrão das aptidões dos melhores indivíduos por geração para o estudo de caso “Pontos Quânticos”.

PGLIQ apresenta um erro percentual absoluto médio de 5,2%, contra 8,3% da melhor rede neural, o que representa uma redução do erro em 37,3%.

4.2.5 Duas Espirais

Este estudo de caso é um problema de classificação bastante conhecido na área de aprendizado de máquina [5, 6], onde duas espirais tem que ser distinguidas em um espaço bidimensional (\mathbb{R}^2). Todas as coordenadas dos pontos de uma espiral pertencem à mesma classe, conforme ilustrado pela figura 4.10.

Os conjuntos de treinamento, validação e teste contêm, cada um, 194 pontos (casos de aptidão) cujos valores das coordenadas (entradas) pertencem ao intervalo $[2\pi, -2\pi]$, conforme pode-se perceber pela figura 4.10.

As figuras 4.11 e 4.12 mostram as curvas evolutivas dos modelos em função do tempo médio de duração e do número médio de avaliações de cada execução,

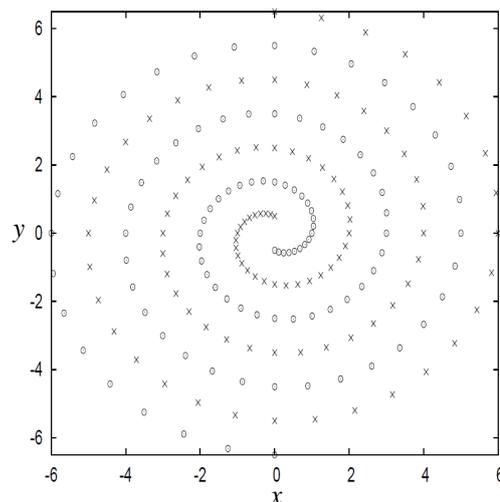


Figura 4.10: Representação gráfica do problema “Duas Espirais”.

respectivamente. Também mostram a comparação de desempenho dos modelos quanto à média das aptidões dos melhores indivíduos por geração, onde as soluções da PGLIQ apresentam valores de erro médio inferiores aos da AIMGP, resultando em um desempenho de 31,5% a 40,1% superior.

A tabela 4.10 concentra os valores de aptidão e de comparação mostrados nas figuras 4.11 e 4.12, além de apresentar os valores de desvio padrão, assim como os de suas variações percentuais decorrentes da comparação dos dois modelos.

Critério de comparação	<i>Demes</i>	Aptidão				Variação (%)	
		AIMGP		PGLIQ		Média	σ
		Média	σ	Média	σ		
Tempo	Não	0,2477	0,0475	0,1509	0,0518	-39,1	9,1
Nº avaliações	Não	0,2477	0,0475	0,1486	0,0520	-40,0	9,6
Tempo	Sim	0,2340	0,0389	0,1604	0,0527	-31,5	35,5
Nº avaliações	Sim	0,2340	0,0389	0,1506	0,0518	-35,6	33,1

Tabela 4.10: Comparação de média e desvio padrão das aptidões dos melhores indivíduos por geração para o estudo de caso “Duas Espirais”.

Os valores de aptidão dos melhores indivíduos obtidos pelos modelos são exibidos pela tabela 4.11. Estes valores, obtidos pelos conjuntos de treinamento, validação e teste são comparados em função do tempo médio de duração e do número médio de avaliações de cada execução. Pela análise das três últimas colunas das duas últimas linhas, nota-se que a PGLIQ apresenta melhor desempenho, ao obter erros de 12,5 a 88,0% inferiores aos obtidos pela AIMGP executada com *demes*. Neste caso, contrariamente ao esperado, o modelo AIMGP apresenta desempenho pior, quanto aos melhores indivíduos obtidos, quando executado com

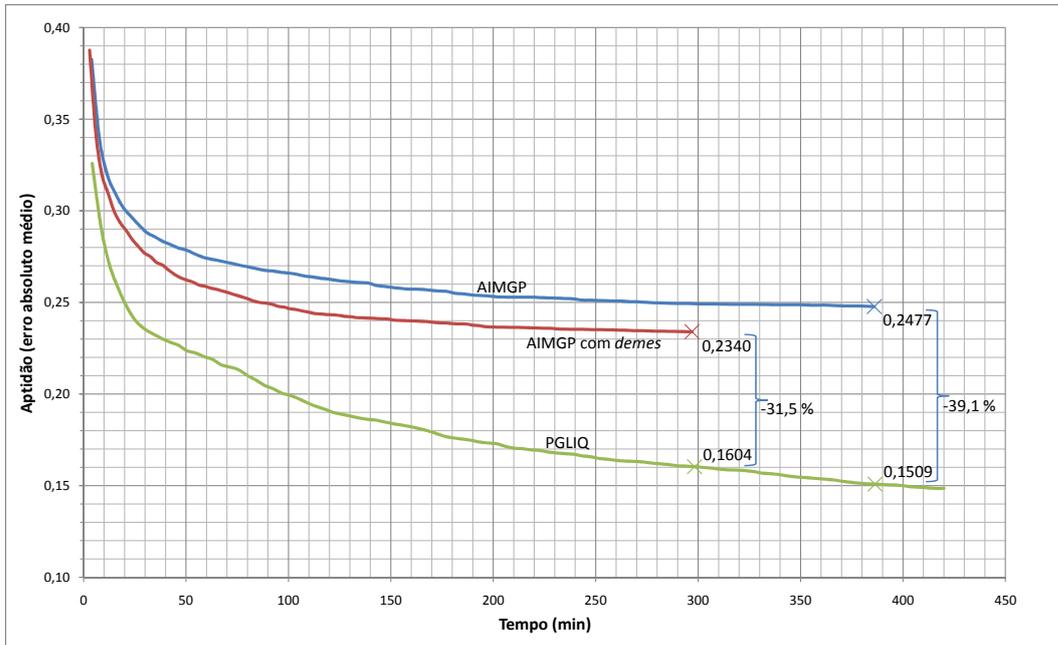


Figura 4.11: Gráfico evolutivo do estudo de caso “Duas Espirais” em função do tempo médio de cada execução dos modelos.

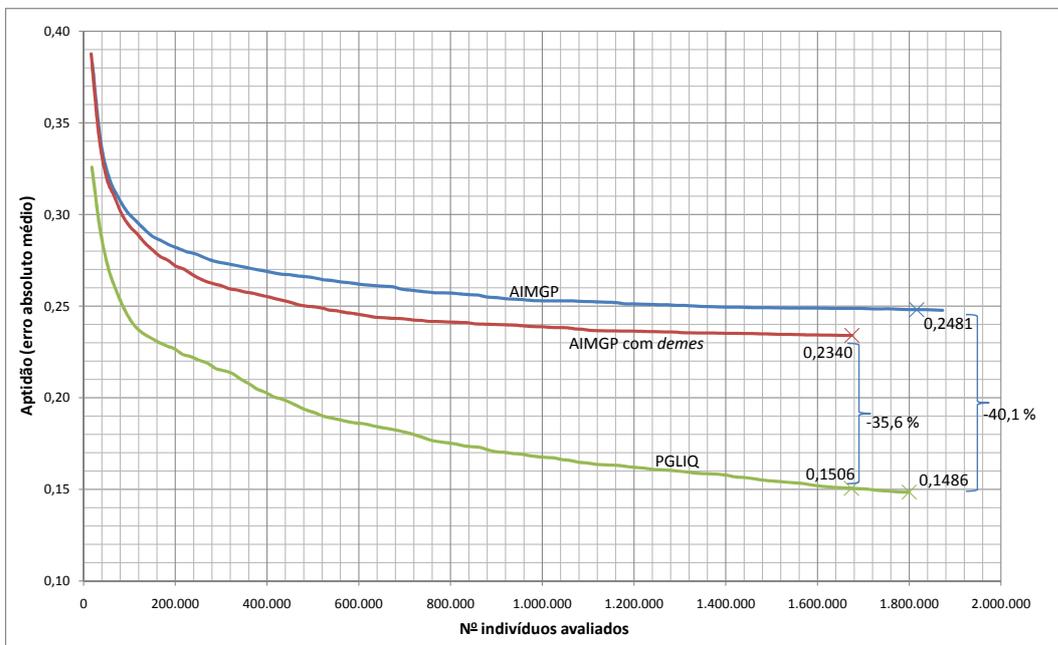


Figura 4.12: Gráfico evolutivo do estudo de caso “Duas Espirais” em função do número médio de indivíduos avaliados a cada execução dos modelos.

demes. Também pode-se notar que um dos melhores indivíduos obtidos pela PGLIQ acerta todos os casos de aptidão do conjunto de validação (erro = 0).

Critério de comparação	<i>Demes</i>	Aptidão								
		AIMGP			PGLIQ			Variação (%)		
		Trein.	Valid.	Teste	Trein.	Valid.	Teste	Trein.	Valid.	Teste
Tempo	Não	0,005	0,067	0,041	0,026	0,015	0,021	395,6	-76,9	-49,0
Nº avals.	Não	0,005	0,067	0,041	0,005	0,000	0,005	-0,9	-100,0	-87,5
Tempo	Sim	0,088	0,129	0,124	0,041	0,021	0,108	-52,9	-84,0	-12,5
Nº avals.	Sim	0,088	0,129	0,124	0,026	0,015	0,098	-70,6	-88,0	-20,8

Tabela 4.11: Comparação da aptidão dos melhores indivíduos obtidos pelos modelos para o estudo de caso “Duas Espirais”.

4.2.6 Diagnóstico Cardíaco

Este estudo de caso é baseado em um problema de classificação binária extraído de [6]. Trata-se de um problema de diagnóstico médico, onde a tarefa é determinar se o diâmetro de um vaso cardíaco encontra-se reduzido em mais de 50% ou não. O problema apresenta 13 variáveis de entrada sobre o paciente, sendo 6 contínuas, como idade, pressão sanguínea em repouso etc., e 7 discretas, como o sexo e tipo de dor no peito (4 tipos diferentes) etc. Esta e outras bases de dados sobre diagnósticos médicos, assim como suas respectivas documentações e detalhes, podem ser encontradas no repositório “UCI Repository of Machine Learning Databases” [75].

A base de dados original possui um total de 303 amostras, aqui divididas em 151, 76 e 76 amostras, para a composição dos conjuntos de treinamento, validação e teste, respectivamente. Cada um dos três conjuntos possui, aproximadamente, o mesmo número de amostras pertencentes às duas classes.

As curvas evolutivas dos modelos, em função do tempo médio de duração e do número médio de avaliações de cada execução, são mostradas, respectivamente, pelas figuras 4.13 e 4.14. Estas figuras também indicam os valores resultantes da comparação de desempenho dos modelos quanto à média das aptidões dos melhores indivíduos por geração, onde as soluções da PGLIQ apresentam valores de erro médio inferiores aos do modelo AIMGP, resultando em uma superioridade de desempenho de 32,4% a 41,2% sobre tal modelo.

Os valores de aptidão e de comparação mostrados nas figuras 4.13 e 4.14 são reunidos na tabela 4.12. Também são apresentados os valores de desvio padrão, além dos valores de sua variação percentual resultantes da comparação dos dois modelos.

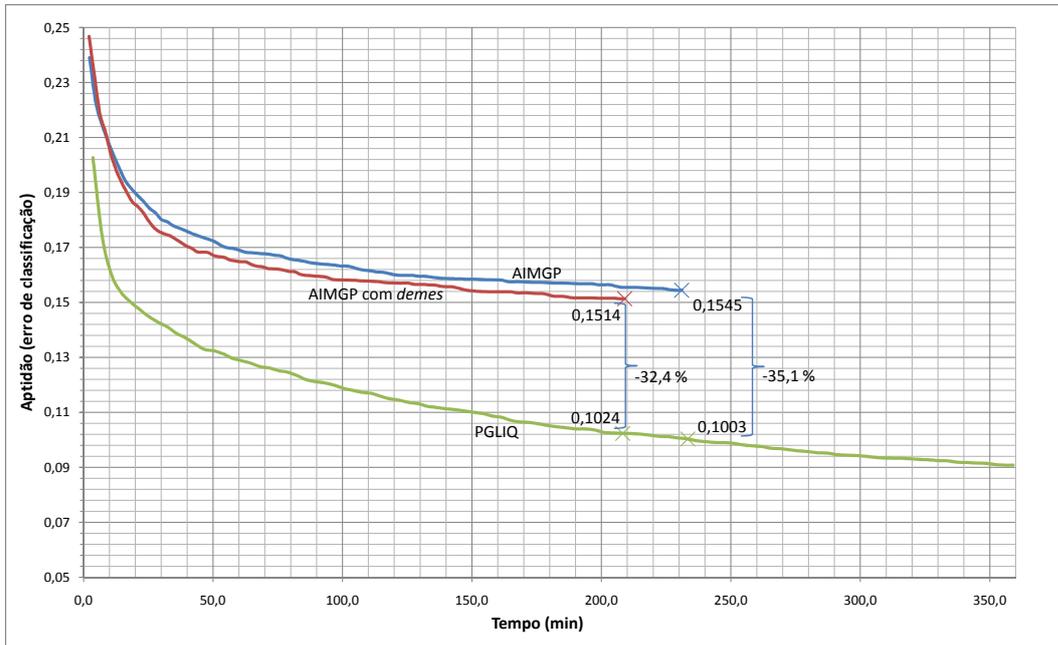


Figura 4.13: Gráfico evolutivo do estudo de caso “Diagnóstico Cardíaco” em função do tempo médio de cada execução dos modelos.

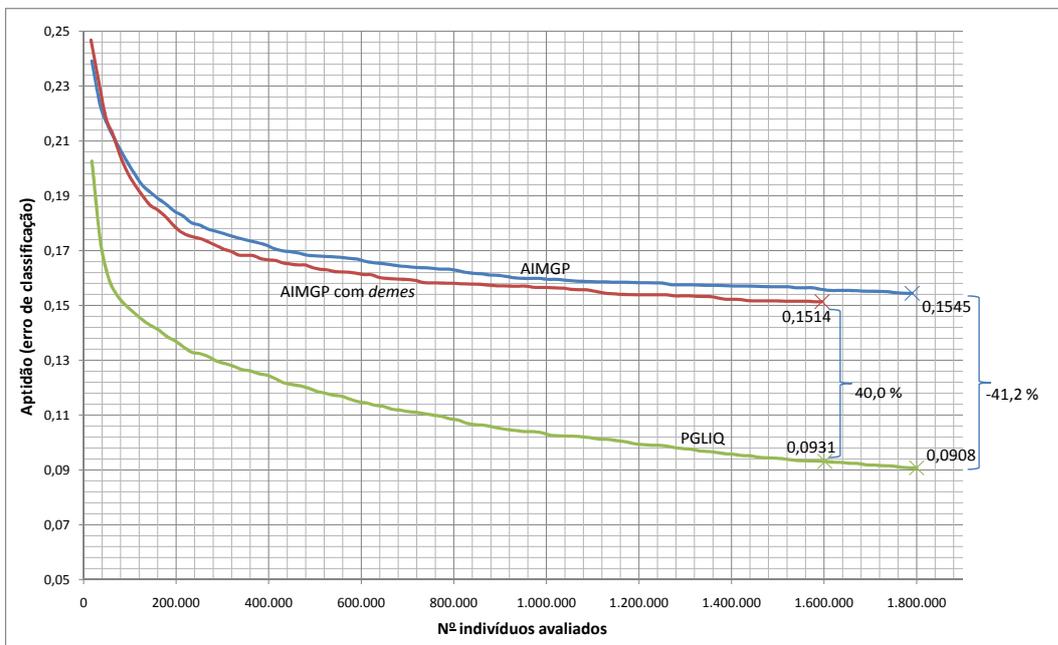


Figura 4.14: Gráfico evolutivo do estudo de caso “Diagnóstico Cardíaco” em função do número médio de indivíduos avaliados a cada execução dos modelos.

Critério de comparação	<i>Demes</i>	Aptidão					
		AIMGP		PGLIQ		Variação (%)	
		Média	σ	Média	σ	Média	σ
Tempo	Não	0,1545	0,0237	0,1003	0,0137	-35,0	-42,2
Nº avaliações	Não	0,1545	0,0237	0,0908	0,0154	-41,2	-35,2
Tempo	Sim	0,1514	0,0275	0,1024	0,0136	-32,4	-50,5
Nº avaliações	Sim	0,1514	0,0275	0,0931	0,0152	-38,5	-44,9

Tabela 4.12: Comparação de média e desvio padrão das aptidões dos melhores indivíduos por geração para o estudo de caso “Diagnóstico Cardíaco”.

A tabela 4.13 compara o desempenho dos melhores indivíduos evoluídos pelos modelos, em função do número médio de avaliações e do tempo médio de duração de cada execução, mostrando a aptidão dos indivíduos com relação aos conjuntos de treinamento, validação e teste. As três últimas colunas à direita mostram que o desempenho da PGLIQ é superior ao da AIMGP para os três conjuntos de dados e para os dois critérios de avaliação, apresentando soluções com erros de 2,4 a 69,2% menores em relação ao modelo AIMGP.

Critério de comparação	<i>Demes</i>	Aptidão								
		AIMGP			PGLIQ			Variação (%)		
		Trein.	Valid.	Teste	Trein.	Valid.	Teste	Trein.	Valid.	Teste
Tempo	Não	0,080	0,171	0,276	0,066	0,053	0,213	-16,7	-69,2	-22,9
Nº avals.	Não	0,080	0,171	0,276	0,060	0,053	0,211	-25,0	-69,2	-23,8
Tempo	Sim	0,086	0,158	0,237	0,073	0,079	0,231	-15,4	-50,0	-2,4
Nº avals.	Sim	0,086	0,158	0,237	0,073	0,066	0,221	-15,1	-58,3	-6,7

Tabela 4.13: Comparação da aptidão dos melhores indivíduos obtidos pelos modelos para o estudo de caso “Diagnóstico Cardíaco”.