

## 4. Resultados Modelo NEIQ-BR

### 4.1. Casos para Avaliação

O objetivo desta seção será a descrição de todos os problemas de classificação a serem utilizados para aplicação do algoritmo NEIQ-BR, e posterior comparação com outros pesquisadores na literatura. Em 4.1.1 os casos benchmark serão descritos; em 4.1.2 problemas reais de uma cooperativa de crédito brasileira serão apresentados; e em 4.1.3 detalhes sobre o pré-processamento a ser executado em todas as bases de dados.

#### 4.1.1. Problemas Benchmark

A seguir, uma breve descrição dos casos benchmark para problemas de classificação de dados utilizados nesta dissertação. Para mais detalhes, recomenda-se o acesso de informações em [26].

**Caso “Australian Credit”:** O banco de dados consiste em uma aplicação típica de tomada de crédito, que tem como objetivo prever bons e maus pagadores. São um total de 14 variáveis preditivas, sendo 8 categóricas e 6 numéricas. Existe um total de 690 registros, sendo 307 para bons pagadores, e 383 para maus pagadores. Em 37 registros existe pelo menos um valor faltante em uma das variáveis. Na oportunidade, os resultados obtidos pelo modelo NEIQ-BR serão comparados com o melhor obtido por outros autores em [10], [33] e [35].

**Caso “German Credit”:** Com o mesmo objetivo do banco de dados australiano apresentado anteriormente, o crédito alemão busca identificar bons e maus pagadores. São um total de 20 variáveis preditivas, sendo 7 numéricas e 13 categóricas. Tem-se um total de 1.000 registros, sendo 700 para bons pagadores, e 300 para maus pagadores. Nenhum valor faltante pode ser identificado. Na oportunidade, os resultados obtidos pelo modelo NEIQ-BR serão comparados com o melhor obtido por outros autores em [35], [36], e [37].

**Caso “Iris Data”:** Apresentado por Fisher (1936), o banco de dados consiste de uma amostra de 3 tipos de plantas Iris, cada tipo contendo 50 amostras. Uma das classes é completamente separável de todas as outras, enquanto que as duas

restantes não. As três classes a serem previstas são descritas como: Iris Setosa, Iris Versicolour, e Iris Virginica. Existe um total de 4 variáveis numéricas preditivas, e nenhum valor faltante. Na oportunidade, os resultados obtidos pelo modelo NEIQ-BR serão comparados com o melhor obtido por outros autores em [35], [37], [38], [39], [40] e [41].

**Caso “Wisconsin Diagnostic Breast Cancer”:** O principal objetivo deste banco trata-se de diferenciar um tumor de mama do tipo benigno ou maligno. São 569 registros, onde 357 são para o tumor benigno e 212 para o maligno. Existem 30 variáveis numéricas preditivas, medidas através de imagens digitalizadas. Nenhum valor faltante pode ser identificado. Na oportunidade, os resultados obtidos pelo modelo NEIQ-BR serão comparados com o melhor obtido por outros autores em [42] e [43].

**Caso “Wisconsin Prognostic Breast Cancer”:** Diferentemente do banco de dados anterior, onde existe o diagnóstico de um câncer de mama, este objetivará a identificação prévia da possibilidade de recorrência da doença, ou seja, um prognóstico. Existem 30 variáveis numéricas preditivas, medidas através de imagens digitalizadas para 198 registros de pacientes. Destes, 47 são casos recorrentes e 151 são não recorrentes. Valor faltante pode ser encontrado em 4 registros. Na oportunidade, os resultados obtidos pelo modelo NEIQ-BR serão comparados com o melhor obtido por outros autores em [42].

**Caso “Wine Data”:** O principal objetivo deste banco trata-se em determinar a origem do vinho através de uma análise química de 13 variáveis numéricas preditivas, e um total de 3 origens distintas. A primeira origem tem um total de 59 registros, a segunda origem tem 71, e a terceira origem com 48. Nenhum valor faltante pode ser observado. Na oportunidade, os resultados obtidos pelo modelo NEIQ-BR serão comparados com o melhor obtido por outros autores em [38], [39], [40], [44] e [45].

Um resumo de fontes e modelos utilizados por outros autores pode ser obtido pela tabela 4.1.

#### 4.1.2. Problemas Reais

Além dos casos benchmark vastamente estudados por outros pesquisadores, alguns problemas reais inéditos de classificação também serão estudados. Tratam-

se de dois exemplos de problemas provenientes de uma cooperativa de crédito brasileira.

Pela definição do SEBRAE SP, uma cooperativa de crédito trata-se de “uma associação formada por no mínimo 20 pessoas, que possui legislação própria e oferece aos associados serviços equivalentes ao de uma instituição financeira”. O SEBRAE ressalta que “um dos maiores benefícios de uma cooperativa de crédito são as vantajosas taxas e os encargos oferecidos aos associados, por ser esta uma entidade sem fins lucrativos, e conseqüentemente, com baixa tributação”.

Modelos	Fonte
<b>Australian Credit Approval</b>	
RBF – GA	Carvalho and Lacerda [10]
C4.5 Trees	Carvalho and Lacerda apud Jones and Quinlan [10]
Composite Rule Sets	J. R. Quinlan [33]
EDRL	Kwedlo and Kretowski [46]
<b>German Credit</b>	
EDRL	Kwedlo and Kretowski [35]
Selective Bayesian Classifier	Hatanamahatana and Gunopulos [36]
Refined GP Gain (k=3)	EGGERMONT et al. [37]
<b>Iris Data</b>	
EDRL	Kwedlo and Kretowski [35]
Clustering GP (k=3)	EGGERMONT et al. [37]
Depuration (3,2)	Jiang and Zhou [38]
RelabelOnly (3,2)	Jiang and Zhou [38]
v-K-SVCR	Zhong and Fukushima [39]
FSS-Kmeans-ML-STD-2	Dy and Brodley [40]
LDC-LLE	Ridder et al [41]
LDC-alfa-SLLE	Ridder et al [41]
<b>Diagnostic Breast Cancer</b>	
SMO	Potter [42]
MSM-T	Mangasarian and Street [43]
<b>Prognostic Breast Cancer</b>	
ClassificationViaRegression	Potter [42]
Logistic	Potter [42]

Tabelas 4.1 – Resumo de casos benchmark.

Por estes motivos, é grande a demanda dos associados por produtos e serviços ofertados pela cooperativa, dentre os quais se destacam dois deles, disponíveis a maioria dos sócios:

- i. Obtenção de empréstimos financeiros;
- ii. Utilização do crédito especial de limite de conta corrente.

Uma instituição deste tipo necessita se apoiar em boas práticas de mercado no relacionamento com o cliente, evitando altos riscos de operação consequente da alta oferta de crédito e possível aumento das taxas de inadimplência. Para que isto não ocorra, esforços são direcionados de tal forma que: (i) haja o cumprimento das dívidas obtidas no médio e longo prazo; (ii) no curto prazo, clientes deverão honrar os créditos tomados.

Para que isto ocorra, a cooperativa necessita dispor de mecanismos eficientes de oferta de crédito e obtenção de recursos financeiros. Isto garantirá de certa forma a sustentabilidade do negócio. Uma maneira eficiente para que tais mecanismos estejam disponíveis na empresa, seria através da *aplicação de técnicas estatísticas ou de inteligência artificial na modelagem* destes problemas tipicamente bancários. Mais detalhes a seguir.

**Caso “Empréstimos Financeiros”:** O principal problema aqui se trata de separar-se sócios inadimplentes e adimplentes, após ter sido concedido o empréstimo financeiro. Diferentemente de um clássico modelo de concessão de crédito, nosso objetivo aqui não se trata de prever futuros clientes inadimplentes antes da concessão. Tratou-se o evento de concessão como feito, e o que se deseja é obter uma nota (ou *score*, em inglês), associado ao risco de inadimplência futura para cada sócio ainda adimplente.

Para análise, utilizou-se um histórico de empréstimos concedidos por um período de 12 meses. *É considerado inadimplente o sócio com algum contrato que no final de algum mês estava com parcela vencida.* Em caso de períodos consecutivos ou alternados de contrato inadimplente, considerou-se o mês mais recente do sócio. Analogamente, *adimplente é o sócio com algum contrato que no final de todos os meses não apresentou nenhuma inadimplência.* Resumidamente tem-se:

Sócio Inadimplente (Algum mês com contrato inadimplente) = 82.  
Sócio Adimplente (Nenhum mês com contrato inadimplente) = 2.537.

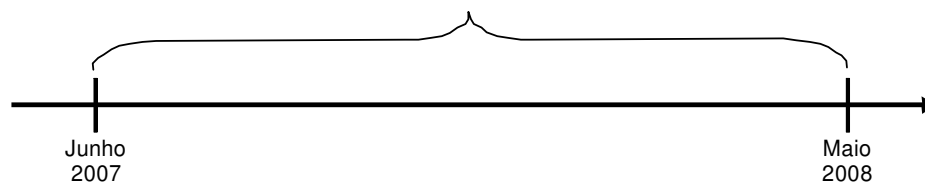


Figura 4.1 – Extração dos dados de empréstimos financeiros.

**Caso “Crédito de Limite Especial”:** A utilização do crédito de limite especial é de grande interesse para instituições financeiras, pois gera lucratividade pela aplicação das taxas de juros sobre o montante utilizado. Porém, podem gerar prejuízos no curto prazo pelo não cumprimento das dívidas, e conseqüentemente, menor crédito rotativo disponível ao consumidor.

Como não há um prazo determinado para o cumprimento deste tipo de dívida quando utilizada, o que se faz é considerar um limiar no percentual utilizado do

total disponível para um cooperado, como sendo o separador das fronteiras de alto e baixo risco de utilização. É consenso geral de o mercado financeiro considerar como sendo de *alta utilização* o cliente cuja utilização exceda oitenta por cento do limite disponível, portanto, altamente sujeito a inadimplência. Por outro lado, é considerado de *baixa utilização*, o cliente cuja utilização não exceda oitenta por cento do limite disponível, neste caso, baixa chance de inadimplência. Desta forma, o principal interesse aqui será separar clientes com alta utilização dos de baixa utilização do limite especial. Isto permitirá obter-se um *score* associado ao risco de alta utilização.

Para a análise, se utilizará também um histórico de 12 meses de sócios os quais utilizaram o limite especial pelo menos uma vez. No nosso problema, *sócio de alto risco será o que apresentou algum mês com alta utilização do limite especial*. Em caso de mais de uma ocorrência, considera-se o mês com alto risco mais recente. Analogamente, *baixo risco é o sócio que apresentou nenhum mês com alto risco*. Resumidamente tem-se:

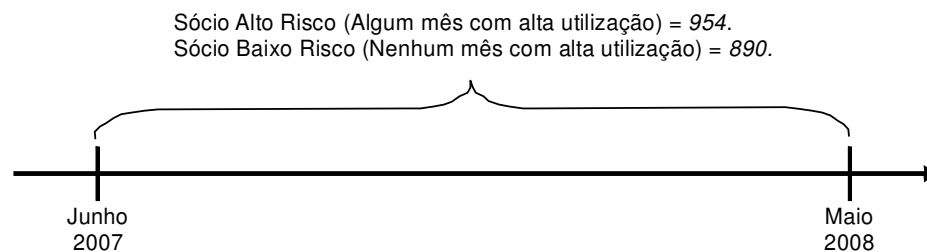


Figura 4.2 – Extração dos dados de limite especial.

Para consulta básica de informações dos sócios, a cooperativa dispõe de um banco de dados com informações do tipo: sócio-demográfica; cadastrais; e de movimentação em conta-corrente, ou transacional. A tabela 4.2 a seguir oferece um resumo destas informações. Variáveis iniciadas com “\*” representam as variáveis dependentes, interesse de modelagem. São elas:

- A “flgInadimplenteUM” para *empréstimos financeiros*;
- A “flgUtilizacaoLimiteAlta” para *crédito de limite especial*.

Variáveis explicativas, ou preditivas, serão todas as variáveis que não a variável dependente (dependendo do problema) e excluindo o grupo de variáveis explicativas a qual esta última pertence.

Grupos	Variáveis	Descrição	Informação
Transacional (Investimentos)	flgInvestido	Código 1 se o sócio possuir dinheiro investido, 0 caso contrário.	{x ∈ I [0,1]}
	vlrInvestido	Valor financeiro que o sócio possui investido na cooperativa.	{x ∈ R [0,2.000.000]}
	flgInvestiuUM	Código 1 caso o sócio tenha investido algum valor no mês anterior, 0 caso contrário.	{x ∈ I [0,1]}
Transacional (Limite Especial)	vlrLimiteEspecial	Valor financeiro que a cooperativa disponibiliza para a utilização do limite especial.	{x ∈ R [0,50.000]}
	flgLimiteEspecial	Código 1 caso a cooperativa disponibilize limite especial, 0 caso contrário.	{x ∈ I [0,1]}
	vlrDevedorUM	Valor financeiro utilizado pelo cooperado do limite especial no último mês.	{x ∈ R [0,40.000]}
	flgDevedorUM	Código 1 caso o cooperado tenha utilizado limite especial no último mês.	{x ∈ I [0,1]}
	vlrExcessoLimiteUM	Valor financeiro de excesso de limite no cheque especial no último mês.	{x ∈ R [0,11.000]}
	flgExcessoLimiteUM	Código 1 se excedeu limite do cheque especial no último mês, 0 caso contrário.	{x ∈ I [0,1]}
	*flgUtilizacaoLimiteAlta	Código 1 se utilização excedeu 80% do disponível, 0 caso contrário.	{x ∈ I [0,1]}
Transacional (Seguros)	flgSeguro	Código 1 se possui algum tipo de seguro com a cooperativa, 0 caso contrário.	{x ∈ I [0,1]}
	vlrSeguro	Valor pago a cooperativa mensalmente pelos seguros contratados.	{x ∈ R [0,1.600]}
	flgInicioSegurosUM	Código 1 caso o sócio tenha iniciado algum tipo de seguro no último mês.	{x ∈ I [0,1]}
Transacional (Movimentação C.c.)	vlrPagamentoUM	Valor recebido como crédito de serviços prestados no último mês.	{x ∈ R [0,350.000]}
	vlrDebitadoUM	Valor financeiro total debitado em conta no último mês.	{x ∈ R [-677.000,0]}
	vlrCreditadoUM	Valor financeiro total creditado em conta no último mês.	{x ∈ R [0,723.000]}
	flgDccUM	Código 1 se teve débito automático em conta corrente no último mês.	{x ∈ I [0,1]}
	qtdeDocTedSaq-ChUM	Quantidade total de transações financeiras efetuados no último mês.	{x ∈ I [0,47]}
	flgDocTedSaq-ChUM	Código 1 caso tenha efetuado alguma transação financeira no último mês.	{x ∈ I [0,1]}
Transacional (Empréstimos)	*flgInadimplenteUM	Código 1 caso tenha algum empréstimo inadimplente no último mês.	{x ∈ I [0,1]}
	flgEmprestimoUM	Código 1 caso apresente algum empréstimo aberto no último mês.	{x ∈ I [0,1]}
	flgTomouEmprestimoUM	Código 1 caso tenha tomado algum empréstimo no último mês.	{x ∈ I [0,1]}
	qtdeEmprestimosAbertosUM	Quantidade total de empréstimos em aberto que possui no último mês.	{x ∈ I [0,4]}
	vlrEmprestimosAbertosUM	Valor a pagar de empréstimos em aberto que o sócio possui no último mês.	{x ∈ R [0,694.000]}
Sociodemográficas	vlrIdade	Idade em anos do sócio.	{x ∈ I [17,97]}
	nomSexo	Sexo do sócio: masculino ou feminino.	{x ∈ I [1,0]}
	nomEstadoCivil	Estado civil do sócio: casado; divorciado; separado; solteiro; viúvo.	{x ∈ I [1,5]}
	nomCidade	Cidade do sócio: Rio de Janeiro ou outra cidade.	{x ∈ I [1,20]}
	vlrRendaLiquidaMensal	Valor de renda líquida mensal declarada pelo usuário.	{x ∈ R [0,915.000]}
Cadastral	nomAgenciaCadastro	Agência de cadastro do sócio: até 8 possíveis agências.	{x ∈ I [1,8]}
	TempoDesdeMatriculaDias	Tempo (em dias) desde a matrícula até a data de retirada das informações.	{x ∈ I [340,10.206]}

Tabelas 4.2 – Variáveis para os casos da cooperativa de crédito.

### 4.1.3. Pré-processamento

Para simplificar e unificar o pré-processamento, os mesmos procedimentos de **normalização, transformação e tratamento** foram executados em todos os problemas.

Por exemplo, no caso do *tratamento de observações faltantes*, quando variáveis numéricas, substituiu-se pela média da variável em questão. Já quando variáveis categóricas, observações faltantes foram substituídas pela moda. Na *transformação de variáveis*, variáveis categóricas com duas classes eram substituídas por variáveis binárias. Já outras variáveis categóricas com mais de duas classes, todas foram transformadas conforme *codificação 1 de N-categorias*. Por exemplo, se uma variável tinha originalmente 10 categorias, esta seria transformada em 10 novas variáveis binárias, onde quando uma variável estiver igual a 1, representando a sua categoria correspondente, todas as demais estarão com valores nulos.

Na *normalização de variáveis* numéricas com alguma interpretação direta de tempo decorrido (por exemplo: idade), utilizava-se a *faixa dinâmica* conforme (4-1). Esta normalização garantirá, por exemplo, que idades próximas umas as outras, continuem próximas após a transformação.

$$x_{ij}^{normalizado} = \frac{x_{ij} - \text{mínimo}(x_i)}{\text{máximo}(x_i) - \text{mínimo}(x_i)} \quad (4-1)$$

Onde  $i$  é a  $i$ -ésima variável, e  $j$  a  $j$ -ésima observação. Demais variáveis numéricas foram normalizadas pela *normalização padrão* conforme (4-2).

$$x_{ij}^{normalizado} = \frac{x_{ij} - \hat{X}}{\hat{\sigma}_X} \quad (4-2)$$

A tabela a seguir resume todos os bancos de dados após sofrer processos de normalização, transformação e tratamento dos dados.

Base de Dados	Amostras	Variáveis Binárias	Variáveis Contínuas	Número Classes
Australian Credit Approval	690	32	6	2
German Credit	1.000	52	7	2
Iris Data	150	-	4	3
Diagnostic Breast Cancer	569	-	30	2
Prognostic Breast Cancer	198	-	33	2
Wine Data	178	-	13	3
Empréstimos Financeiros	2.619	19	13	2
Crédito de Limite Especial	1.844	16	13	2

Tabela 4.3 – Resumo descrição dos bancos de dados.

## 4.2. Descrição dos Experimentos

Conforme algoritmo detalhado no capítulo 3, diversos parâmetros devem ser ajustados pelo usuário conforme tabela 3.6 no capítulo 3.7.2. São eles  $N_Q$ ,  $N_C$ ,  $C_Q$ ,  $C_{Cb}$ ,  $C_{Cr}$ ,  $T$ ,  $\theta$ ,  $minT$ ,  $updateT$  e  $TI$ .

Objetivando a comparação do desempenho do algoritmo, tanto na comparação com outros autores dentre problemas, quanto na comparação dos resultados entre problemas, optar-se-á por variar os parâmetros *igualmente* em cada um dos casos apresentados na tabela 4.3. Isto tem como motivação nossa crença de que o ajuste de parâmetros dependerá mais do tipo de problema que se deseja otimizar – a evolução de uma RNA – do que características gerais de um banco de dados. Não que as características de um banco de dados não influencie o ajuste de parâmetros, pois devem sim influenciar. *O que se deseja é obter ajuste de parâmetros que sejam eficientes para a maioria dos problemas de classificação.* Esta decisão de se restringir a variação igualmente entre problemas se baseia em alguns fundamentos técnicos a serem detalhados, mas, sobretudo, tem como objetivo principal tornar viável computacionalmente a realização de todos os experimentos desta dissertação.

Todos os bancos de dados descritos na seção anterior serão avaliados através de um processo de validação cruzada particionado em 3, para cada uma das estratégias acima. Cada partição terá exatamente 33% da amostra, e em cada uma das validações, duas partições serão utilizadas para treinamento, e o grupo remanescente será utilizado para testar-se a generalização do modelo. Neste caso, uma estratégia seria avaliada através do desempenho médio obtido entre as 3 partições do grupo de teste, através de alguma função de avaliação conforme 3.7.3. Este tipo de procedimento é também chamado de “3-fold cross-validation”.

Suponham algumas estratégias evolutivas do algoritmo NEIQ-BR onde se queira testar dois ajustes de cada um dos parâmetros, e que também se deseja testar se estes parâmetros têm alguma dependência de variação conjunta entre si. Neste caso, pretender-se-ia avaliar estas estratégias em todos os bancos de dados apresentados. Considere também um processo de validação cruzada em 3 partições. Sendo assim, neste caso hipotético, seria necessário que se executasse pelo menos 24.576 experimentos, que é a conta de  $2^{10} \times 8 \times 3$ , ou seja, ajustes<sup>parâmetros</sup>  $\times$  problemas  $\times$  validações cruzadas. O que tornaria inviável o processamento de to-



dos estes experimentos em tempo aceitável, visto que algoritmos evolutivos para problemas de classificação tomam um tempo e memória consideráveis de processamento.

Entendam que a cada geração de um único experimento, para cada um dos indivíduos clássicos, toda uma base de padrões deverá ser lida e propagada pela rede neural obtida pelo indivíduo, até que se obtenha um único valor de aptidão. Isto sem dúvida consumiria algum tempo de processamento, e de certa forma, tem sido uma das duras críticas que algoritmos evolutivos têm recebido na literatura<sup>7</sup>.

Adicionalmente aos parâmetros do algoritmo NEIQ-BR, se incluirá um novo elemento a ser variado e testado nas estratégias, que é o *tipo de função de avaliação*. A função de avaliação não faz parte explicitamente dos parâmetros de ajuste do algoritmo NEIQ-BR. Neste momento, tem-se o interesse de estudar se uma ou outra função de avaliação seria mais apropriada para processos de neuroevolução. Na tabela 4.4 a seguir, considere: (1) o percentual total de acerto ponderado pela classe 1; (2) o percentual total de acerto; e (3) a área por debaixo da curva ROC. Sendo assim, optou-se pelas seguintes estratégias evolutivas:

Estratégia	TI	NQ	NC	T	C <sub>cb</sub>	C <sub>cr</sub>	$\theta$	C <sub><math>\alpha</math></sub>	minT	updateT	Função aptidão
1	30	30	30	200	0,92	0,08	0,050*pi	0.20	2	4	(1)
2	20	15	60	200	0,95	0,05	0,080*pi	0.10	2	4	(1)
3	40	10	100	200	0,98	0,02	0,020*pi	0.30	2	4	(1)
4	30	30	30	200	0,99	0,01	0,050*pi	0.20	2	4	(1)
5	20	15	15	200	0,95	0,05	0,080*pi	0.10	2	4	(1)
6	40	5	200	200	0,98	0,02	0,020*pi	0.30	2	4	(1)
7	30	30	30	200	0,92	0,08	0,050*pi	0.20	2	4	(2)/(3)
8	20	15	60	200	0,95	0,05	0,080*pi	0.10	2	4	(2)/(3)
9	40	10	100	200	0,98	0,02	0,020*pi	0.30	2	4	(2)/(3)
10	30	30	30	200	0,99	0,01	0,050*pi	0.20	2	4	(2)/(3)
11	20	15	15	200	0,95	0,05	0,080*pi	0.10	2	4	(2)/(3)
12	40	5	200	200	0,98	0,02	0,020*pi	0.30	2	4	(2)/(3)

Tabela 4.4 – Estratégias evolutivas analisadas.

<sup>7</sup> Nesta dissertação, não investigaremos eficiências de custo computacional ganho ou perdido com a utilização do algoritmo NEIQ-BR, como já mencionado. Esta discussão sobre custo computacional nos servirá para planejarmos melhor os experimentos desta dissertação, tornando-os viáveis computacionalmente. É sabido por Abs da Cruz [6] e Han e Kin [7] que os algoritmos, respectivamente, AEIQ-R e AEIQ-B são mais eficientes em custo computacional. Estudos mais aprofundados sobre custo computacional dos novos algoritmos AEIQ-BR e NEIQ-BR ficarão para estudos futuros.

Sendo assim, com todas estas 12 estratégias na tabela 4.4 e todos os 8 bancos de dados da tabela 4.3, considerando-se todo o processo de validação cruzada, ao final se realizaria 288 experimentos, que é a conta de  $12 \times 8 \times 3$ .

Reparem que alguns parâmetros permanecem constantes entre estratégias. Isto foi feito intencionalmente, pois se consideraram menos importantes analiticamente se estudar a variação e impacto destes parâmetros, deixando isto para estudos futuros. São eles:  $T$ , o número de gerações total;  $minT$ , o número mínimo de gerações que um neurônio permanecerá ativo até que este possa ser desativado por um cruzamento clássico; e  $updateT$ , o intervalo entre gerações para que os indivíduos quânticos sejam atualizados.

Detalhes sobre o ajuste e variação de cada um destes parâmetros podem ser obtidos a seguir:

- **Número de indivíduos quânticos  $N_Q$  e clássicos  $N_C$ :** Quando utilizados *conjuntamente valores bem baixos*, se estará interessado em estudar a capacidade do algoritmo em obter ótimas soluções através da representação probabilística do espaço de busca pelos indivíduos quânticos, sem que se tenha a possibilidade de armazenar muitas soluções promissoras (indivíduos clássicos). E neste caso, se os parâmetros de atualização quântica não estiverem bem ajustados, existiria o risco de convergir prematuramente para uma solução promissora de  $B(t)$ , já que esta não disporá de muitos indivíduos. Por outro lado, quando utilizados *inversamente valores mais altos* para  $N_C$  e mais baixo para  $N_Q$ , se estaria interessado em disponibilizar para o processo evolutivo, um conjunto de soluções mais promissoras, diversificando assim as regiões de buscas dos indivíduos clássicos, pelos indivíduos quânticos. Isto porque a cada geração indivíduos clássicos poderão trocar de posição em ordem de aptidão em  $B(t)$  e com isto, direções de atualização de cada indivíduo em  $Q(t)$  poderiam ser modificadas. Observa-se que em caso de mais indivíduos clássicos do que quânticos, somente os melhores clássicos seriam utilizados para atualização dos indivíduos quânticos. Por outro lado, é importante observar que quanto mais elevado forem ambos os parâmetros, maior será o tempo computacional exigido para se executar um experimento, o que seria de nosso desgosto, pois o que se deseja é uma convergência rápida e eficiente para alguma solução ótima.

- **Total de neurônios na camada intermediária  $TI$ :** Embora muito se discuta na literatura o número de neurônios ótimo para uma camada escondida (ver tabela 2.5), não se pode aproveitar desta discussão para uma consideração inicial mais empírica deste parâmetro. Isto porque o número ótimo de neurônio será definido pelo processo evolutivo, e não pelo usuário. Este parâmetro foi feito – e por opção de evolução codificará no cromossoma a camada intermediária da rede – com o propósito de permitir que a rede neural navegue livremente pelo espaço de soluções de possíveis neurônios considerados na topologia. Por exemplo, não se pode considerar que uma solução acima de 25 neurônios possa ser ótima somente no caso que soluções com neurônios inferiores sejam sub-ótimas também, pois é possível que haja alguma descontinuidade no número de neurônios, e ganho de desempenho da rede. Por exemplo, se um número entre 20 e 25 neurônios não seja comprovadamente melhor, há ainda a possibilidade de que acima de 25 neurônios o algoritmo encontre soluções mais promissoras. Resumidamente,  $TI$  determinará o número máximo de neurônios permitidos. Isto não significa que qualquer solução terá necessariamente este número máximo de neurônios. O parâmetro  $C_{Cb}$  será o que mais impactará na definição de  $TI$ , pois se  $TI$  for muito alto, talvez o algoritmo não consiga evoluir o suficiente para que se obtenha o número ótimo de neurônios. Por outro lado, se  $TI$  for muito baixo, talvez não esteja explorando com eficácia o espaço de soluções mais promissoras em termos de topologia ótima. Por fim, optou-se por valores entre um mínimo  $TI$  de 20 neurônios, e um máximo o dobro do mínimo de 40 neurônios.
- **Total de gerações  $T$ :** O número de gerações permanecerá constante, pois se tratando de todas as características do algoritmo AEIQ-BR e as vantagens de se aplicá-lo na evolução de RNAs, estas não impactarão sobremaneira no número de gerações total ideal. Na verdade, demais parâmetros deverão ser ajustados em concomitância com o número de geração utilizado. É importante dizer que parâmetros que controlam exploração e aproveitamento estão sendo variados, e neste caso, seria importante que o ambiente que estes estão sujeitos fosse semelhante. Em outras palavras, o número de gerações constante nos ajudará a controlar melhor as condições que populações em

diversas estratégias evoluirão, e conseqüentemente, tornará mais justa uma comparação de resultados. O que se fez foi ter alguma consonância na determinação deste parâmetro com base em experiências anteriores em estudos sobre computação evolucionária (ver [31]), e posterior ajuste de demais parâmetros do NEIQ-BR neste ambiente temporalmente estático. É mais razoável entender o comportamento de outros parâmetros de exploração e aproveitamento quando o número de gerações permanece constante, do que quando se varia as etapas de evoluções. Pois o que parecia ser ótimo em termos de ajustes de outros parâmetros, poderá não ser mais quando se tem um tempo de evolução alterado.

- **Taxa de crossover clássico binário  $C_{Cb}$  e real  $C_{Cr}$ :** Nesta parte, todas as 12 estratégias foram configuradas com altas taxas para  $C_{Cb}$  e baixas taxas para  $C_{Cr}$  conforme tabela 4.4. A taxa  $C_{Cb}$  atuará sobre a parte do cromossoma responsável por seleção de neurônios na camada de entrada e intermediária, e seleção de funções de ativação. Este alto ajuste definido para  $C_{Cb}$  ocasionará um ambiente mais dinâmico de ativação e desativação de neurônios, com mudanças mais brusca na direção do algoritmo sendo mais interessante nos passos iniciais do problema, funcionando como um “guia numa multidão de cegos”. Estas mudanças provocadas por  $C_{Cb}$  funcionarão como uma análise de sensibilidade de neurônios durante todo o processo evolutivo, ativando e desativando neurônios mais ou menos importantes para a aptidão de cada indivíduo. Já o  $C_{Cr}$ , o baixo ajuste justifica uma maior precisão na soma ponderada de um neurônio. Pense num neurônio que em determinada geração apresenta alguma relevância em se modelar um comportamento. É importante que a atualização em seus pesos seja a mais fina possível, *do contrário*, mudanças mais bruscas tornariam a busca por pesos ótimos mais árdua, uma vez que a cada atualização dos pesos deste neurônio, dedutivamente se teria todo um conjunto de novos pesos. Em outras palavras, o que se objetiva nesta parte é uma evolução lenta e gradual, semelhante ao que seria o treinamento tradicional de uma rede neural pelo algoritmo de retropropagação.

- **Ângulo de rotação para velocidade de atualização dos genes clássicos binários,  $\theta$ :** A definição de  $\theta$  bem baixo como apresentado na tabela 4.4 foi feita para que o processo de convergência para uma probabilidade máxima igual a 1 de se observar um neurônio ativado/desativado (identicamente para o tipo de função de ativação), seja feito *lentamente* explorando eficientemente todo o espaço de busca. Tem-se que o algoritmo NEIQ-BR se demonstrará mais promissor com  $C_{Cb}$  ativando e desativando dinamicamente neurônios. Para algum neurônio específico,  $\theta$  bem baixo fará com que  $C_{Cb}$  unicamente ative ou desative este neurônio com o passar das gerações, isto porque somente será possível a observação de um dos casos em um gene de um indivíduo quântico, uma vez que a probabilidade estaria caminhando geração a geração para uma convergência máxima de algum dos casos. Além destas justificativas, estas decisões de ajuste também se apóiam em sugestões feitas por Han e Kin [8] na otimização de problemas combinatórios, com ângulos de atualização  $\theta$  bem baixos.
- **Taxa de crossover quântico  $C_Q$ :** Quanto mais baixo for este parâmetro, mais lenta e gradual também será a atualização dos centros e amplitudes observadas nos genes reais dos indivíduos quânticos. Em outras palavras, cada peso de um neurônio será vastamente explorado em todo o seu domínio em gerações iniciais, e, como visto no gráfico 3.1, a partir de gerações mais elevadas e, conseqüente, diminuição da amplitude  $\sigma$ , se contribuiria para a capacidade de aproveitamento do algoritmo. Isto porque a amplitude de  $\sigma$  seria mínima e não nula, e pesos observados de indivíduos quânticos estariam bem próximos – não exatamente iguais – aos pesos dos melhores indivíduos clássicos. Ou seja, contribuindo mais uma vez para ajustes bem finos nos pesos de uma rede neural. Pode-se observar que  $C_Q$  nas estratégias apresentadas na tabela 4.4 esta variando entre 10% a 30%. Conforme explicado anteriormente, quanto mais alto for este parâmetro, mais se desistiria de uma grande exploração do espaço de busca em gerações iniciais, com um alto aproveitamento em passos imediatamente posteriores, devendo ter-se cautela com alguma possível convergência prematura para algum centro de variação.

- **Número mínimo de gerações que um neurônio permanecerá ativo,  $\min T$ , e intervalo entre gerações até que os operadores quânticos sejam executado,  $\text{update} T$ :**  $\min T$  foi definido constante igual a 2. Uma vez que não se dispõe de resultados que indiquem o número necessários de gerações para que se observe os pesos de um neurônio qualquer convergirem para um conjunto ótimo, esta decisão foi a mais conservadora possível. Perceba que se  $\min T$  igual a 1, este parâmetro perderia sua real utilidade. Sendo assim, após *duas rodadas* de possíveis atualizações nos genes reais associados a um neurônio de um indivíduo clássico, este gerará filhos cujo neurônio correspondente a estes pesos poderá ser desativado por um cruzamento. Já  $\text{update} T$ , optou-se por 4 gerações como sendo o intervalo de gerações para atualização dos indivíduos quânticos. A definição de  $\text{update} T$  igual a 4 significa que durante cada 4 gerações o algoritmo estaria explorando o espaço de soluções, e após 4 permitira a atualização dos indivíduos quânticos buscando um aproveitamento do espaço.
- **Função de avaliação:** o que se decidiu foi que nas primeiras 6 estratégias, todos os problemas seriam avaliados conforme percentual de acerto total ponderado pela classe 1. Isto faria com que o algoritmo buscasse soluções que classificassem parcimoniosamente todas as classes do problema. Já nas estratégias seguinte, 7 até a 12, soluções poderiam ser obtidas desbalanceadamente conforme desempenho em se prever uma ou outra determinada classe. Isto porque, no caso de problemas com *mais de duas classes*, seria utilizada somente a função percentual total de acerto, já que a curva ROC não se aplicaria a estes problemas. Enquanto que em problemas com somente duas classes, nas estratégias 7 a 12 seriam utilizados a curva ROC. Resumidamente, para cada um dos 8 problemas apresentados tem-se 6 estratégias com juízo de parcimônia entre classes, e outras 6 estratégias sem.

### 4.3. Resultados Obtidos NEIQ-BR

Na ocasião, o algoritmo NEIQ-BR foi implantado em ambiente Matlab®, e os experimentos delineados no capítulo anterior foram executados. Resultados

para diversos experimentos e problemas serão comparados em termos da média do percentual de padrões classificados erroneamente, denotados pela sigla MPPCE, e desvio padrão de padrões classificados erroneamente, ou DPPCE, conforme tabelas 4.5 e 4.6. Em vermelho, o menor MPPCE por problema, e em azul, o maior. Já em cinza, a média de MPPCE por problema através das 12 estratégias.

Por exemplo, através da estratégia 7 para o problema Iris data, a média do percentual total de padrões classificados erroneamente obtida através do processo de validação cruzada particionado em 3 foi de 4,7%, conforme tabela 4.5. Já o DPPCE foi de 3,1% conforme tabela 4.6. Com isto, a evolução das melhores estratégias geração a geração, por problema, por partição de validação cruzada, pode ser visualizada na figura 4.3.

Observa-se que a estratégia 9 foi a que teve uma maior quantidade de problemas com *melhor desempenho*. Entre as características que esta estratégia possui que a torna diferente das demais, pode-se destacar: a curva ROC como função de avaliação; alta taxa de cruzamento quântico  $C_Q$ , o que fará com que o algoritmo tenha uma fase de exploração mais curta, e aproveitamento mais longo; e ainda, diferentemente da estratégia 12, apresentará um menor desbalanceamento entre número de indivíduos quânticos, e números de indivíduos clássicos. Esta última característica sugere que uma alta taxa de cruzamento quântico seria apropriada, quando houver sim diversos indivíduos quânticos, de tal forma que cada um se aproveitará de um espaço de soluções, podendo mudar sua direção de representação devido a uma grande diversidade populacional presente nos melhores clássicos.

Já as estratégias com *pior desempenho* em um maior número de problemas foram a estratégia 4 e 10. O que diferencia uma da outra é somente a função de avaliação utilizada, percentual total ponderado para estratégia 4, e curva ROC para a 10. Com tudo, a diferenciação das demais estratégias estará no baixo número de indivíduos clássicos e alto de quânticos. Esta última característica parece estar impactando numa alta exploração do espaço de busca, porém sujeita a alguma convergência equivocada, visto que os espaços de busca visitados não estão sendo diversificadamente guardados pelos indivíduos clássicos (baixo). Soma-se a isto também uma convergência lenta, com menor taxa de cruzamento clássico real.

Problema / Estratégia	1	2	3	4	5	6	7	8	9	10	11	12	Média
Australian Credit Approval	12,8%	11,2%	13,0%	16,5%	13,0%	12,3%	6,2%	6,6%	6,1%	6,8%	6,2%	6,2%	9,7%
German Credit	29,8%	24,1%	29,1%	29,9%	25,0%	26,5%	24,9%	26,6%	24,8%	26,0%	26,6%	26,3%	26,6%
Iris Data	3,3%	2,7%	1,3%	2,0%	2,0%	2,0%	4,7%	3,3%	2,7%	3,3%	2,0%	4,7%	2,8%
Diagnostic Breast Cancer	2,1%	1,8%	2,8%	2,6%	2,1%	1,8%	1,9%	2,3%	1,3%	2,8%	1,9%	1,8%	2,1%
Prognostic Breast Cancer	18,2%	22,7%	21,7%	22,2%	18,2%	15,7%	20,7%	17,7%	15,6%	17,2%	21,2%	16,2%	18,9%
Wine Data	3,4%	1,1%	5,0%	5,6%	2,8%	2,8%	4,5%	1,7%	3,4%	6,2%	1,7%	1,1%	3,3%
Empréstimos Financeiros	7,1%	7,2%	8,1%	5,1%	6,1%	5,8%	5,1%	6,6%	8,7%	8,0%	9,5%	7,8%	7,1%
Crédito de Limite Especial	28,7%	29,0%	28,4%	31,0%	26,6%	30,2%	28,9%	27,9%	27,6%	28,7%	28,3%	28,7%	28,7%

Tabela 4.5 – Média do percentual total de padrões classificados erroneamente através de processo validação cruzada em 3.

Problema / Estratégia	1	2	3	4	5	6	7	8	9	10	11	12	Média
Australian Credit Approval	3,5%	1,1%	2,3%	3,9%	1,2%	1,3%	1,9%	0,6%	0,1%	1,1%	2,0%	0,7%	1,6%
German Credit	4,5%	2,1%	1,7%	3,8%	0,5%	0,7%	2,8%	2,1%	0,3%	2,3%	2,2%	0,9%	2,0%
Iris Data	1,2%	1,2%	2,3%	2,0%	0,0%	2,0%	3,1%	2,3%	1,2%	3,1%	0,0%	1,2%	1,6%
Diagnostic Breast Cancer	1,1%	1,1%	1,1%	0,9%	0,9%	1,3%	0,3%	1,1%	0,3%	1,1%	0,3%	1,7%	0,9%
Prognostic Breast Cancer	1,5%	9,9%	3,8%	3,8%	6,9%	4,9%	0,9%	4,6%	4,9%	4,9%	2,6%	3,2%	4,3%
Wine Data	2,9%	1,0%	3,3%	6,9%	3,5%	2,0%	1,0%	0,0%	1,7%	2,6%	0,0%	1,0%	2,2%
Empréstimos Financeiros	2,5%	1,4%	2,8%	1,6%	1,1%	0,8%	1,8%	0,9%	4,8%	1,9%	3,4%	2,9%	2,2%
Crédito de Limite Especial	0,7%	1,7%	1,8%	0,5%	1,7%	2,4%	1,3%	1,1%	0,4%	1,3%	3,0%	0,8%	1,4%

Tabela 4.6 – Desvio padrão do percentual total de padrões classificados erroneamente através de processo validação cruzada em 3.



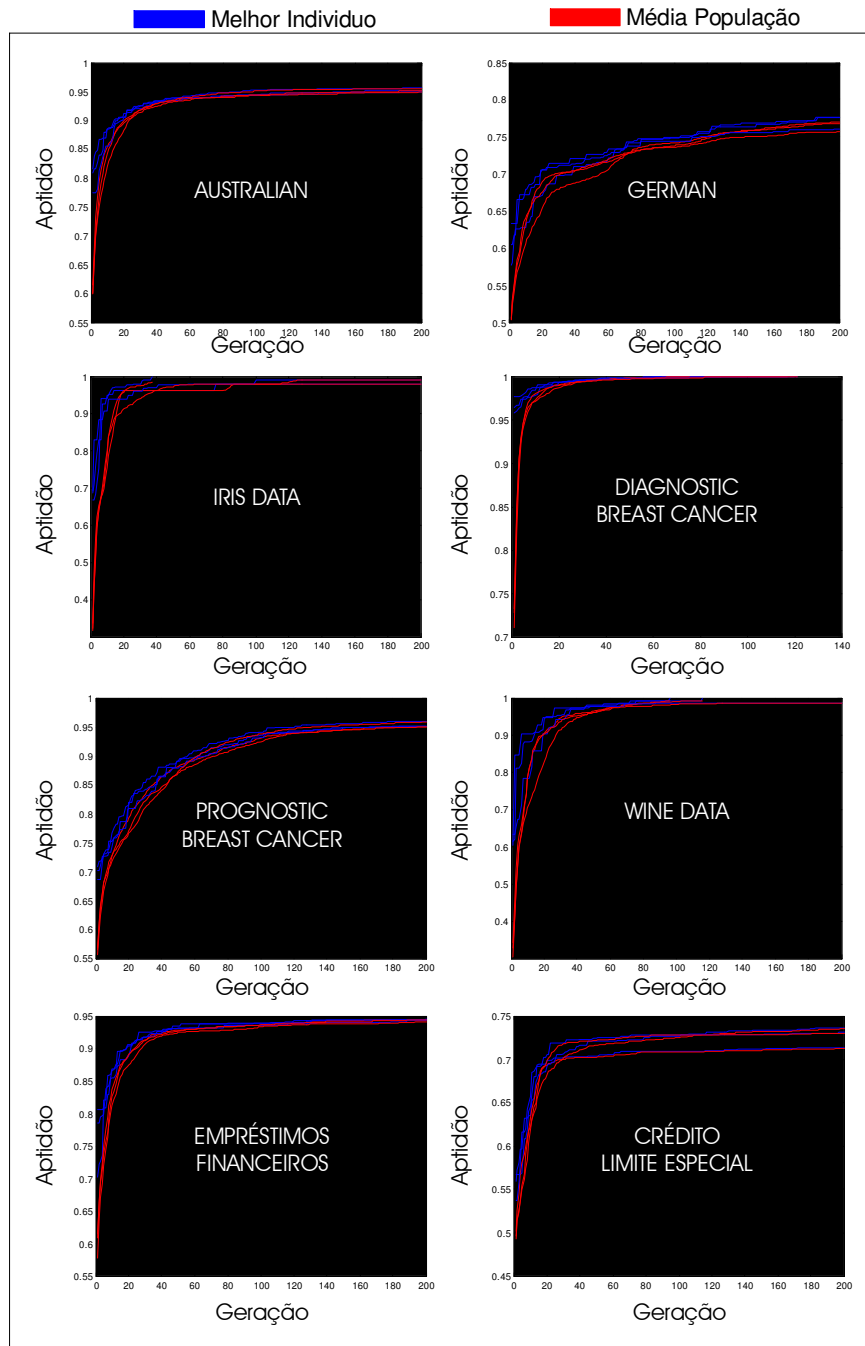


Figura 4.3 – Evolução de populações por melhor estratégia, por problema, por partição de validação cruzada.

Analisando a figura 4.3, pode-se entender as diferenças no nível de dificuldade encontrando pelo algoritmo NEIQ-BR em se otimizar uma rede para classificação de padrões, em cada um dos problemas. Isto pode ser feito através da curvatura dos melhores indivíduos e da média populacional ao longo das gerações, e os ganhos obtidos em termos de aptidão. Ao que sugere, o banco de dados Iris,

Wine, e Diagnostic Breast Cancer apresentaram um baixo nível de dificuldade com resultados da curva ROC atingindo valores próximos de 1 rapidamente. Enquanto que o German Data, um alto grau de dificuldade com valores próximos de 0.75, apresentando ainda uma tendência positiva de crescimento para gerações mais elevadas.

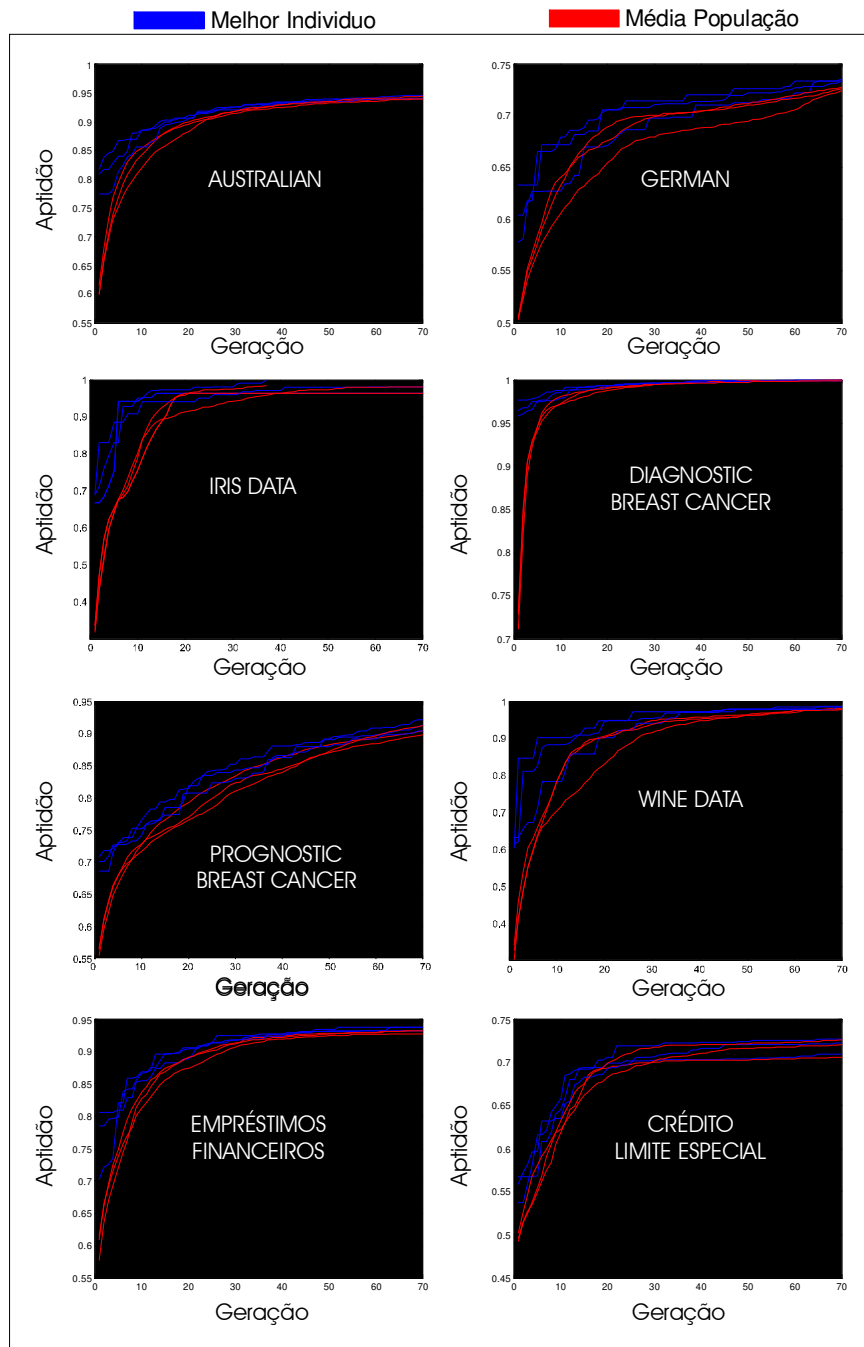


Figura 4.4 – Evolução de populações por melhor estratégia, por problema, por partição de validação cruzada, para as 70 primeiras gerações.

A figura 4.4 fornece uma visualização ampliada para as primeiras 70 gerações, quando a variação dos melhores indivíduos e da média populacional é maior.

Embora estes apresentados até então possam parecer interessantes, não se dará ou tirará o mérito a alguma estratégia analisada na tabela 4.5, pois estas diferenças em termos de MPPCE foram até certo ponto sutis. Deve-se investigar com maior rigor através de procedimentos estatísticos mais adequados o quão significativo foram estas diferenças obtidas por uma ou outra configuração de parâmetros em específico. Neste sentido, estatísticas de MPPCEs e DPPCEs serão submetidas a até quatro tipos de testes, conforme figura 4.5. O descritivo teórico dos testes pode ser encontrado em anexo.

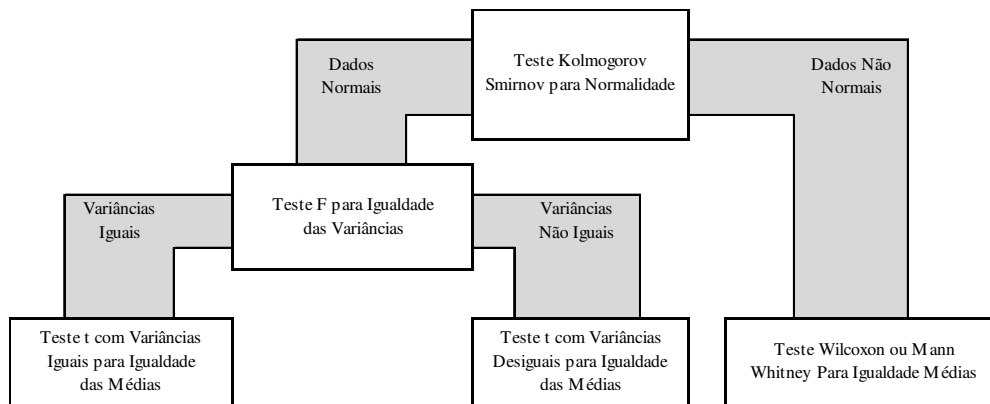


Figura 4.5 – Raciocínio analítico para comparação de duas médias quaisquer.

O primeiro trata-se do *teste de Kolmogorov Smirnov* (KS) para amostras independentes testando se uma amostra em termos de MPPCE apresenta-se como uma distribuição normal (ver Bussab [48], página 406). Este teste é necessário, pois para aplicação do teste *t* para igualdade de médias, uma premissa básica é de que os dados seguem uma distribuição normal de probabilidade, pois em caso negativo, este teste não seria mais apropriado.

O segundo, o de teste de igualdade de variância para duas populações normais ou *teste F* (ver Bussab [48], página 358). Tendo-se aceito a hipótese de normalidade pelo teste KS, a fórmula mais apropriada do teste *t* será utilizada para verificação de diferença entre médias, de acordo com considerações sobre a variância dos dados de duas populações que se deseja comparar.

O terceiro, o *teste t-student* para diferença entre média conforme explicado as diferenças na formulação quando variância das duas populações iguais ou diferentes (ver Bussab [48], página 363 e 364). E o quarto, o *teste Wilcoxon ou*

*Mann-Whitney* para diferença entre médias, porém, quando a normalidade não é verificada nos dados (ver Bussab [48], página 366).

Para todos os testes se utilizará um *nível de significância de 5%*. Sendo assim, foram testadas as hipóteses de diferenças estatísticas significativas nos MPPCE, em termos de:

- Estratégias utilizadas;
- Total de neurônios na camada intermediária,  $TI$ ;
- Total de indivíduos quânticos,  $N_Q$ ;
- Total de indivíduos clássicos,  $N_C$ ;
- Taxa de cruzamento clássico binário;
- Taxa de cruzamento clássico real;
- Velocidade de atualização quântica real,  $\theta$ ;
- Taxa de atualização quântica;
- E por fim, funções de avaliação.

O procedimento para aplicação destes testes seguirá conforme figura 4.5. Por exemplo, quando desejado verificar se experimentos utilizando a curva ROC foram estatisticamente diferentes (entende-se superiores ou inferiores) dos utilizando percentuais de acerto total ponderado, deverá se executar o seguinte procedimento:

- (1) Aplica-se primeiramente o teste de Kolmogorov Smirnov em MPPCE para verificar a normalidade dos dados;
- (2) Caso a normalidade seja verificada, os dados estão aptos para serem submetidos ao testes estatísticos t-student para diferenças entre médias. Neste caso, a formulação mais adequada do teste *t-student* deve ser utilizada, segundo verificação de variâncias iguais ou desiguais através do teste F;
- (3) Do contrário, caso os dados não sejam normalmente distribuídos, ou esta normalidade seja passível de ser verificada, o teste não paramétrico Wilcoxon seria mais apropriado.

Em cada uma das variações dos parâmetros utilizadas nas estratégias 1 a 12 foi aplicado o teste estatístico Kolmogorov-Smirnov para verificar a normalidade dos dados em termos de uma amostra de MPPCE. Resumidamente, tivemos um

96 amostras de dados – 12 estratégias x 8 problemas, cada uma compondo uma estatística de MPPCE – e a normalidade foi aceita para qualquer conjunto, e resultados estão em apêndice. Em seguida, cada amostra de um total de 96 foi submetida ao teste F, onde a hipótese de variância igual foi aceita, com exceção das funções de avaliação onde a variância é estatisticamente diferente.

Posteriormente, o teste  $t$  mais apropriado foi aplicado, onde a igualdade de médias também foi aceita para qualquer comparação de parâmetros. Ou seja, *não existem diferenças significativas em termos de MPPCE, quando se opta pela variação de um ou outro parâmetro do algoritmo NEIQ-BR no conjunto de estratégias analisadas*. Isto não significa que a variação em conjunto de algum grupo de parâmetro possa surtir impacto no resultado de MPPCE, pois isto não foi testado. Enfatiza-se que o que foi testado foi variação unilateral de cada um dos parâmetros configurados.

Estes resultados para igualdade de variância e igualdade de médias também se encontram em apêndice.

Vale ressaltar que a *diferença entre funções de avaliação* seria significativa ao nível de significância  $\alpha$  igual a 10%. Ou seja, a curva ROC apresentaria resultados significativamente superiores que o de percentual total de acerto ponderado, ao nível de significância de 10%.

*Resumidamente temos que todas as 12 estratégias analisadas são robustas o suficiente, tal forma que diferenças significativas nos resultados em termos de MPPCE não são verificadas.*

Resultados adicionais podem ser encontrados nas tabelas 4.7 e 4.8. São eles:

- (1) Percentual de acerto obtida dos melhores indivíduos, das melhores estratégias, por problemas, por grupos de teste e treinamento, por partição de validação cruzada, e por classe predita;
- (2) Topologia ótima obtida dos melhores indivíduos, das melhores estratégias, por problema, por partição de validação cruzada.

Sobre o número de neurônios na camada de entrada, não se observa nenhuma uniformidade por problema, por partição de banco de dados. Ou seja, cada partição apresentará um conjunto distinto de variáveis de entrada.

Problema	Partição Dados	Treino					Teste					Função Avaliação
		Acerto Classe 1 %	Acerto Classe 2 %	Acerto Classe 3 %	Acerto Total %	Aptidão	Acerto Classe 1 %	Acerto Classe 2 %	Acerto Classe 3 %	Acerto Total %	Aptidão	
Australian Credit Approval	1	84,2	90,3	-	93,8	0,87	90,0	87,0	-	94,0	0,89	Curva Roc
	2	86,5	91,0	-	94,9	0,89	83,7	91,6	-	93,8	0,87	
	3	85,1	86,8	-	94,0	0,86	87,5	87,3	-	94,0	0,87	
German Credit	1	77,2	74,8	-	76,5	0,76	74,5	71,4	-	73,6	0,73	% Acerto Ponderado
	2	74,3	74,5	-	74,4	0,74	80,3	71,0	-	77,5	0,76	
	3	74,3	73,8	-	74,2	0,74	76,9	76,2	-	76,6	0,77	
Iris Data	1	100,0	93,9	100,0	98,0	0,98	100,0	94,1	91,7	96,0	0,95	% Acerto Ponderado
	2	100,0	97,1	91,4	96,0	0,96	100,0	100,0	100,0	100,0	1,00	
	3	100,0	94,4	94,4	96,0	0,96	100,0	100,0	100,0	100,0	1,00	
Diagnostic Breast Cancer	1	98,6	99,0	-	98,2	1,00	97,4	100,0	-	98,4	1,00	Curva Roc
	2	98,6	99,0	-	98,7	1,00	98,4	100,0	-	98,9	1,00	
	3	100,0	99,0	-	98,7	1,00	99,2	98,6	-	98,9	1,00	
Prognostic Breast Cancer	1	80,6	88,5	-	86,4	0,87	89,1	81,8	-	87,9	0,87	Curva Roc
	2	88,5	73,6	-	76,5	0,87	73,3	90,5	-	78,8	0,87	
	3	76,5	86,7	-	84,1	0,87	84,9	92,3	-	86,4	0,90	
Wine Data	1	97,2	98,0	93,9	96,6	0,96	100,0	100,0	93,3	98,3	0,98	% Acerto Ponderado
	2	100,0	97,8	97,1	98,3	0,98	95,0	100,0	100,0	98,3	0,98	
	3	100,0	94,3	100,0	97,5	0,98	100,0	100,0	100,0	100,0	1,00	
Empréstimos Financeiros	1	94,1	93,2	-	94,1	0,87	93,5	82,6	-	93,2	0,89	Curva Roc
	2	93,4	94,4	-	93,4	0,87	94,4	100,0	-	94,6	0,87	
	3	95,9	81,8	-	95,5	0,87	97,2	85,2	-	96,8	0,89	
Crédito de Limite Especial	1	70,0	72,5	-	71,3	0,71	75,2	67,9	-	71,3	0,72	% Acerto Ponderado
	2	63,0	76,7	-	70,0	0,70	69,3	80,4	-	75,2	0,75	
	3	70,0	71,2	-	70,6	0,71	70,6	76,2	-	73,5	0,73	

Tabela 4.7 – Percentual de acerto obtida dos melhores indivíduos, das melhores estratégias, por problema, por partição de validação cruzada, e por classe.

Problema	Partição Dados	Número de variáveis camada de entrada (sem o viés)	Número de neurônios camada escondida (sem o viés)	Neurônios com ativação "Tangente Hiperbólica" na escondida	Neurônios com ativação "Sigmóide" na escondida	Neurônios com ativação "Tangente Hiperbólica" na saída	Neurônios com ativação "Sigmóide" na saída
Australian Credit Approval	1	15	21	10	11	0	1
	2	20	27	11	16	0	1
	3	14	24	10	14	0	1
German Credit	1	25	12	6	6	1	1
	2	34	10	5	5	2	0
	3	30	8	5	3	1	1
Iris Data	1	1	16	8	8	2	1
	2	1	21	11	10	2	1
	3	2	21	13	8	1	2
Diagnostic Breast Cancer	1	11	18	9	9	1	0
	2	20	21	8	13	0	1
	3	13	21	12	9	0	1
Prognostic Breast Cancer	1	24	22	14	8	1	0
	2	19	19	10	9	1	0
	3	18	22	12	10	1	0
Wine Data	1	5	11	6	5	1	2
	2	9	9	7	2	1	2
	3	5	12	11	1	2	1
Empréstimos Financeiros	1	13	12	7	5	0	1
	2	13	13	8	5	1	0
	3	14	15	4	11	0	1
Crédito de Limite Especial	1	10	11	5	6	2	0
	2	13	12	8	4	1	1
	3	16	9	4	5	1	1

Tabela 4.8 – Topologia obtida dos melhores indivíduos, das melhores estratégias, por problema, por partição de validação cruzada.

O mesmo acontece para o número de neurônios na camada de entrada. Porém, é interessante perceber que para cada problema, parece existir uma linha base, ou número médio ótimo de neurônios na camada escondida. Conclui-se isso, pois apesar de diferentes estes neurônios na escondida por partição, a variabilidade é mais alta entre problemas do que dentro problemas. Por exemplo, percebam o caso Empréstimos Financeiros com 12, 13 e 15 neurônios na escondida em comparação com o Australian Credit com 21, 27 e 24. Isto aumenta a crença da existência de um número ótimo de neurônios na camada escondida por problema.

Na próxima seção comparações do resultado do NEIQ-BR com outros autores serão apresentadas.

#### 4.4. Comparação com Outros Autores

Neste capítulo, uma série de resultados obtidos por outros autores para os problemas benchmark serão utilizados, em termos de MPPCE. Sendo assim, diversos autores foram selecionados (vide seção 4.1.1) de tal forma que a aplicação de seus modelos desenvolvidos apontaria para pelo menos um dos problemas aqui estudados. Esta seleção de outros autores, para os casos benchmark, foi feita com base na bibliografia recomendada por Murphy e Aha [26].

Em cada um destes autores dos artigos comparados, selecionou-se o melhor resultado obtido por problema distinto. Ou seja, para cada problema e por autor, será possível a visualização e comparação nesta seção de somente um resultado em termos de MPPCE e DPPCE.

É importante mencionar que todos os autores selecionados utilizaram partições no banco de dados do tipo *10-fold-cross-validation*, em contrapartida a nossa seleção do tipo *3-fold-cross-validation*.

Já para os casos reais, se comparará o modelo NEIQ-BR com o ajuste manual de redes neurais artificiais através do algoritmo de retropropagação, visto que não se dispõe até então de estudos publicados sobre estas bases de dados. Para cada um dos casos reais, foi-se obtida uma rede neural final da seguinte forma:

- (1) Com época fixa igual a 50, o número de neurônios na camada escondida foi variado em:
  - a) Sugestões presentes no software de datamining Weka®.



- igual ao número total de atributos;
  - número total de classes;
  - número de atributos mais o número de classes;
  - número de atributos mais o número de classes, divididos por dois;
- b) Números próximos aos encontrados na literatura para modelos neuroevolutivos conforme [13], [14], [21] e [29], uma vez que se compararão os resultados destas RNAs, com as obtidas pelo NEIQ-BR.
- 4
  - 6 neurônios;
  - e 8 neurônios.

Para cada uma das variações obtinha-se a MPPCE no grupo de teste através de um processo de validação cruzada particionado em 3.

- (2) Com o melhor resultado de neurônios na camada escondida do passo anterior, passava-se a variar o número épocas, ou seja, durações na fase de treinamento. Esta variação era feita com: 25; 50; 100; 200; 400; e 800 épocas. A utilização de outros procedimentos para o tempo de treinamento destas redes como, por exemplo, a parada por aumento do erro no grupo de validação, exigiria um novo particionamento no banco de dados. Não se seguiu este caminho, pois desejou-se que a mesma divisão dos bancos de dados utilizada para o algoritmo NEIQ-BR também seja utilizada para esta modelagem manual. Sendo assim, *obtinha-se uma rede final com o par de neurônios ótimos do passo anterior e duração do treinamento ótima do passo atual, medido através da MPPCE no grupo de teste*. Nosso objetivo aqui foi explorar um número pequeno, porém diversificado de infinitas possibilidades de se submeter uma rede ao treinamento.

Sendo assim, uma rede neural treinada seria obtida para cada caso real da cooperativa de crédito, a partir da variação de parâmetros de número de épocas de treinamento, e número de neurônios na camada escondida. Nestes casos, todas variáveis preditivas estariam sendo utilizadas na rede neural. Um resumo dos resultados obtidos pode ser visto na tabela 4.9 para casos reais e benchmark.

Modelos	MPPCE	DPPCE	Fonte
<b>Australian Credit Approval</b>			
RBF - GA	14,0%	3,5%	Carvalho and Lacerda [10]
C4.5 Trees	15,1%	Não Disponível	Carvalho and Lacerda [10]
Composite Rule Sets	14,4%	1,9%	J. R. Quinlan [33]
EDRL	13,9%	0,4%	Kwedlo and Kretowski [46]
Média	14,3%	1,9%	-
<b>German Credit</b>			
EDRL	29,9%	0,8%	Kwedlo and Kretowski [35]
Selective Bayesian Classifier	23,8%	Não Disponível	Hatanamahatana [36]
Refined GP Gain (k=3)	27,1%	0,8%	Eggermont et al. [37]
Média	26,9%	0,8%	-
<b>Iris Data</b>			
EDRL	4,0%	0,0%	Kwedlo and Kretowski [35]
Clustering GP (k=3)	2,1%	0,2%	Eggermont et al. [37]
Depuration (3,2)	4,3%	4,8%	Jiang and Zhou [38]
RelabelOnly (3,2)	4,3%	4,8%	Jiang and Zhou [38]
v-K-SVCR	1,3%	Não Disponível	Zhong and Fukushima [39]
FSS-Kmeans-ML-STD-2	2,0%	3,1%	Dy and Brodley [40]
LDC-LLE	1,0%	1,6%	Ridder et al [41]
LDC-alfa-SLLE	1,0%	1,6%	Ridder et al [41]
Média	2,5%	2,3%	-
<b>Diagnostic Breast Cancer</b>			
SMO	2,1%	Não Disponível	Potter [42]
MSM-T	3,0%	Não Disponível	Mangasarian and Street [43]
Média	2,6%	-	-
<b>Prognostic Breast Cancer</b>			
ClassificationViaRegression	20,2%	Não Disponível	Potter [42]
Logistic	20,2%	Não Disponível	Potter [42]
MultiClassClassifier	20,2%	Não Disponível	Potter [42]
Média	20,2%	-	-
<b>Wine Data</b>			
RemoveOnly (3,2)	5,1%	2,9%	Jiang and Zhou [38]
NNEE (5,5)	5,1%	2,9%	Jiang and Zhou [38]
v-K-SVCR	3,3%	Não Disponível	Zhong and Fukushima [39]
Kmeans-STD-2	0,0%	0,0%	Dy and Brodley [40]
Média	3,4%	1,9%	-
<b>Empréstimos Financeiros</b>			
Multilayer Perceptron - 32-2-1	4,9%	2,0%	Apêndice
Média	4,9%	2,0%	-
<b>Crédito de Limite Especial</b>			
Multilayer Perceptron - 29-2-1	29,8%	1,1%	Apêndice
Média	29,8%	1,1%	-

Tabela 4.9 – Média e desvio padrão percentual do total de padrões classificados erroneamente para outros autores de casos benchmark e modelagem manual de casos reais.

O próximo passo agora se trata da comparação do melhor resultado obtido por outros autores, com a *melhor estratégia* obtida pelo modelo NEIQ-BR, em cada um dos problemas pelas 12 configurações de parâmetros propostas.

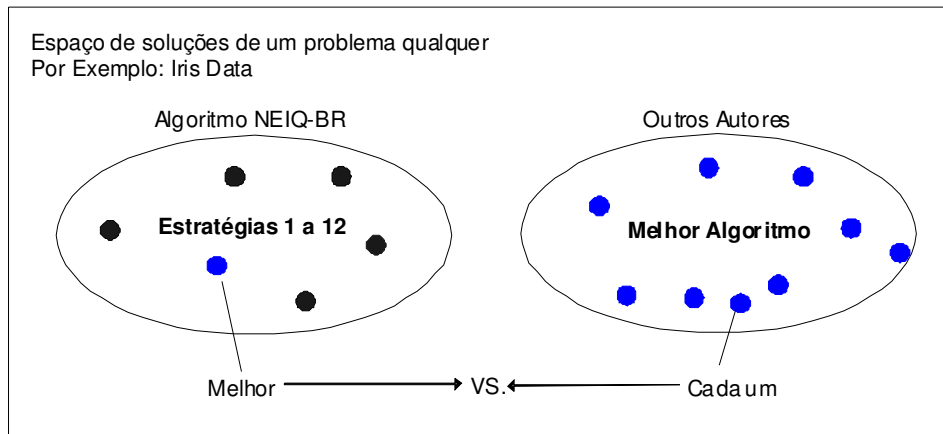


Figura 4.6 – Comparação de melhores estratégias contra melhores algoritmos para um problema qualquer.

Todo o processo de verificação de normalidade, e posteriores procedimentos mais adequados conforme descrito em figura 4.5 deveria ter sido utilizado. Contudo, não se dispõe de resultados de outros autores individualmente por cada uma das 10 validações cruzadas, inviabilizando a aplicação de teste nos dados.

Sendo assim, se assumirá como verdadeira a normalidade nos dados, para então prosseguir com a aplicação do teste F para igualdade de variância, e posterior utilização do teste  $t$  mais adequado (considerando ou não variâncias iguais).

*Diferentemente da tabela 4.9, o melhor obtido por cada algoritmo estudado por cada autor (podem ser vários algoritmos), encontra-se em anexo. Pode-se encontrar neste mesmo anexo que algoritmos consagrados como o SVM ou o C4.5 estarão mais presentes, mas que certamente não foram os vencedores na maioria dos artigos aqui estudados, e, portanto, a ausência na tabela 4.9. Comparações com outros autores, os quais haviam disponibilizado o desvio padrão conforme tabela 4.9, podem ser vistos na tabela 4.10.*

Pelos resultados da tabela 4.10 observa-se que o algoritmo NEIQ-BR alcançou ótimo desempenho. Em 18 das comparações onde se tinha o desvio padrão disponível, viabilizando assim os testes estatísticos, um aproveitamento de 8 superações, 9 empates, e 1 derrota foram observados. Além disso, o modelo NEIQ-BR superou resultados de pesquisas de diversos outros autores em 44% dos casos. Estes resultados são ainda mais otimistas se considerar-se que não fomos tão restritivos na variação dos parâmetros, buscando estratégias que fossem comuns a todos os problemas, e não ajustes individuais por casos.

Modelos	Outros Autores	NEIQ-BR	Teste F	Teste T	Vencedor
<b>Australian Credit Approval</b>					
Composite Rule Sets	14,4%	6,1%	Rejeita H0	Rejeita H0	NEIQ-BR
RBF - GA	14,0%	6,1%	Rejeita H0	Rejeita H0	NEIQ-BR
C4.5 Trees	15,1%	6,1%	-	-	-
EDRL	13,9%	6,1%	Aceita H0	Rejeita H0	NEIQ-BR
<b>German Credit</b>					
Selective Bayesian Classifier	23,8%	24,1%	-	-	-
Refined GP Gain (k=3)	27,1%	24,1%	Aceita H0	Rejeita H0	NEIQ-BR
EDRL	29,9%	24,1%	Aceita H0	Rejeita H0	NEIQ-BR
<b>Iris Data</b>					
Depuration (3,2)	4,3%	1,3%	Aceita H0	Aceita H0	Empatado
RelabelOnly (3,2)	4,3%	1,3%	Aceita H0	Aceita H0	Empatado
v-K-SVCR	1,3%	1,3%	-	-	-
FSS-Kmeans-ML-STD-2	2,0%	1,3%	Aceita H0	Aceita H0	Empatado
LDC-LLE	1,0%	1,3%	Aceita H0	Aceita H0	Empatado
LDC-alfa-SLLE	1,0%	1,3%	Aceita H0	Aceita H0	Empatado
Clustering GP (k=3)	2,1%	1,3%	Aceita H0	Aceita H0	Empatado
EDRL	4,0%	1,3%	Aceita H0	Rejeita H0	NEIQ-BR
<b>Diagnostic Breast Cancer</b>					
SMO	2,1%	1,3%	-	-	-
MSM-T	3,0%	1,3%	-	-	-
<b>Prognostic Breast Cancer</b>					
ClassificationViaRegression	20,2%	15,6%	-	-	-
Logistic	20,2%	15,6%	-	-	-
MultiClassClassifier	20,2%	15,6%	-	-	-
<b>Wine Data</b>					
RemoveOnly (3,2)	5,1%	1,1%	Aceita H0	Rejeita H0	NEIQ-BR
NNEE (5,5)	5,1%	1,1%	Aceita H0	Rejeita H0	NEIQ-BR
v-K-SVCR	3,3%	1,1%	-	-	-
KmeansSTD2	0,0%	1,1%	Aceita H0	Rejeita H0	KmeansSTD2
<b>Empréstimos Financeiros</b>					
Multilayer Perceptron -32-2-1	4,9%	5,1%	Aceita H0	Aceita H0	Empatado
<b>Crédito de Limite Especial</b>					
Multilayer Perceptron - 29-2-1	29,8%	26,6%	Aceita H0	Aceita H0	Empatado

Tabela 4.10 – Comparação de melhores médias MPPCE entre outros autores e NEIQ-BR.

No banco Iris Data, onde se obteve um maior número de empates com outros autores, acredita-se que estes resultados estejam associados com a facilidade em se discriminar as classes deste problema, com resultados próximos de 100% de acurácia total. Problemas como o Australian Credit e German Credit, onde o nível de dificuldade é um pouco maior, e ainda não são totalmente separáveis, o melhor resultado do NEIQ-BR obteve um ótimo desempenho, superando estatisticamente outros autores. Além do mais, Australian e German apresentarão elevado número de variáveis preditivas e registros se comparados com o Iris com 4 variá-

veis. Já no banco de dados Wine, obteve-se uma derrota para o modelo KmeansSTD2 de Dy e Brodley [40], já que este último conseguiu 0% de erro.

Nos problemas reais da cooperativa de crédito, o NEIQ-BR obteve desempenho estatisticamente igual à modelagem manual de RNAs. Contudo, vale ressaltar que o algoritmo NEIQ-BR apresenta algumas vantagens adicionais como configuração otimizada de topologia de rede, e, desta forma, pode ser preferido em contrapartida à modelagem manual de uma rede neural. Em termos de neurônios na camada escondida, os resultados do NEIQ-BR tiveram uma variação entre 9 a 15 neurônios dependendo da partição e problema real, enquanto que na modelagem manual, foram 2 o total nesta camada.

Visando uma maior robustez na média dos resultados do algoritmo NEIQ-BR, uma vez que foi-se utilizada para cada estratégia um 3-fold-cross-validation, se considerará uma média geral por problema. Sendo assim, para cada problema teremos uma única média, dada por 3x12 observações da média entre experimentos, e posterior comparação com o melhor obtido por outros autores. A figura a seguir representa esta nova comparação:

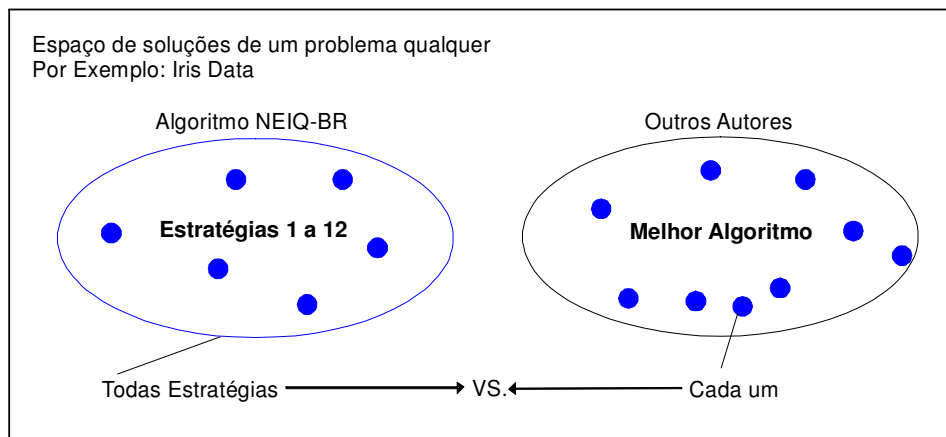


Figura 4.7 – Comparação de todas estratégias contra melhores algoritmos para um problema qualquer.

Observa-se que a consideração de todas as estratégias do NEIQ-BR elevou a média de MPPCE observados anteriormente, porém, a significância estatística com favoritismo para o modelo NEIQ-BR continuar presente em muitos problemas. Isto significa dizer que mesmo com a variabilidade dos resultados imposta por uma ou outra configuração de parâmetros do algoritmo evolutivo, este se apresentou robusto o suficiente no alcance de resultados estatisticamente significantes.

Resultados a seguir:

Modelos	Outros Autores	NEIQ-BR	Teste F	Teste T	Vencedor
<b>Australian Credit Approval</b>					
Composite Rule Sets	14,4%	9,7%	Aceita H0	Rejeita H0	NEIQ-BR
RBF - GA	14,0%	9,7%	Aceita H0	Aceita H0	Empatado
C4.5 Trees	15,1%	9,7%	-	-	-
EDRL	13,9%	9,7%	Aceita H0	Rejeita H0	NEIQ-BR
<b>German Credit</b>					
Selective Bayesian Classifier	23,8%	26,6%	-	-	-
Refined GP Gain (k=3)	27,1%	26,6%	Aceita H0	Aceita H0	Empatado
EDRL	29,9%	26,6%	Aceita H0	Rejeita H0	NEIQ-BR
<b>Iris Data</b>					
Depuration (3,2)	4,3%	2,8%	Rejeita H0	Aceita H0	Empatado
RelabelOnly (3,2)	4,3%	2,8%	Rejeita H0	Aceita H0	Empatado
v-K-SVCR	1,3%	2,8%	-	-	-
FSS-Kmeans-ML-STD-2	2,0%	2,8%	Rejeita H0	Aceita H0	Empatado
LDC-LLE	1,0%	2,8%	Rejeita H0	Aceita H0	Empatado
LDC-alfa-SLLE	1,0%	2,8%	Rejeita H0	Aceita H0	Empatado
Clustering GP (k=3)	2,1%	2,8%	Aceita H0	Rejeita H0	NEIQ-BR
EDRL	4,0%	2,8%	Aceita H0	Rejeita H0	NEIQ-BR
<b>Diagnostic Breast Cancer</b>					
SMO	2,1%	2,1%	-	-	-
MSM-T	3,0%	2,1%	-	-	-
<b>Prognostic Breast Cancer</b>					
ClassificationViaRegression	20,2%	18,9%	-	-	-
Logistic	20,2%	18,9%	-	-	-
MultiClassClassifier	20,2%	18,9%	-	-	-
<b>Wine Data</b>					
RemoveOnly (3,2)	5,1%	3,3%	Rejeita H0	Aceita H0	Empatado
NNEE (5,5)	5,1%	3,3%	Rejeita H0	Aceita H0	Empatado
v-K-SVCR	3,3%	3,3%	-	-	-
KmeansSTD2	0,0%	3,3%	Aceita H0	Rejeita H0	Kmeans-STD-2
<b>Empréstimos Financeiros</b>					
Multilayer Perceptron -32-2-1	4,9%	7,0%	Aceita H0	Aceita H0	Empatado
<b>Crédito de Limite Especial</b>					
Multilayer Perceptron - 29-2-1	29,8%	28,7%	Aceita H0	Aceita H0	Empatado

Tabela 4.11 – Comparação de melhores médias MPPCE entre outros autores contra todas estratégias NEIQ-BR.

Outra análise de diferenças de médias que será realizada é a comparação das médias obtidas por todas as estratégias avaliadas de um problema qualquer, contra a média de todos os melhores resultados obtida por outros autores, conforme figura 1.8. No caso das estratégias do NEIQ-BR, se obterá 12 observações para cada problema para computação da média. E, no caso de outros autores, a depender da quantidade de que se tem disponível para um problema qualquer.

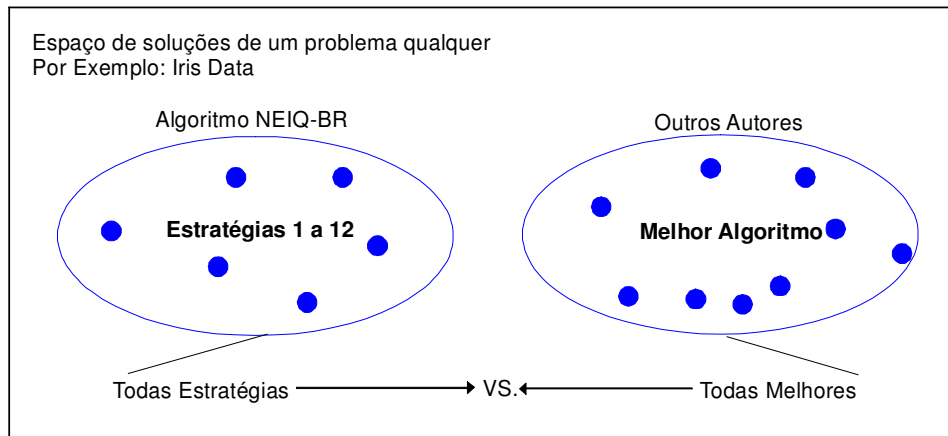


Figura 4.8 – Comparação de todas as estratégias contra todos melhores algoritmos para um problema qualquer.

No caso do teste *t* especificamente, não se conseguirá avaliar de maneira correta se os dados de outros artigos seguem, ou não, uma distribuição de probabilidade normal, premissa básica para aplicar o teste. Isto porque, em alguns problemas se dispõe de alguns 2 ou 3 registros de outros autores. Desta forma, prosseguir-se-á com análises e conclusões do teste Wilcoxon, que segundo investigações nesta dissertação será mais apropriado, pois é um teste não paramétrico e não faz suposições de distribuição de probabilidade. Os resultados a seguir:

Base de Dados	MPPCE		Z Wilcoxon	Z Tabelado	Resultado
	NEIQ-BR	Outros Algoritmos			
Australian Credit	9,7%	14,3%	-2,304	1,960	Rejeito H0
German Credit	26,6%	26,9%	-0,144	1,960	Aceito H0
Iris	2,8%	2,5%	1,080	1,960	Aceito H0
Diagnostic Breast Cancer	2,1%	2,6%	1,732	1,960	Aceito H0
Prognostic Breast Cancer	18,9%	20,2%	-0,144	1,960	Aceito H0
Wine	3,3%	3,4%	-0,105	1,960	Aceito H0
Empréstimos Financeiro	7,0%	4,9%	1,612	1,960	Aceito H0
Crédito Limite Especial	28,7%	29,8%	-0,802	1,960	Aceito H0

Tabela 4.12 – Comparação de médias MPPCE entre outros autores e estratégias NEIQ-BR através do teste de Wilcoxon.

Conclui-se que para o “Australian Credit”, a média entre todas as 12 estratégias conseguiu ser melhor que a média de outros algoritmos. Para os demais problemas, o algoritmo NEIQ-BR conseguiu eficientemente substituir outros algoritmos no geral, pois apresentou média estatisticamente igual pelo teste de Wilcoxon. Sendo assim, pode-se dizer que o algoritmo NEIQ-BR é robusto o bastante dentro do conjunto de estratégias que se utilizou nesta dissertação.

Os bancos de dados utilizados nesta dissertação foram particionados em 3, conforme descrições de experimentos na seção 4.2. Esta decisão de usar-se 3-fold visava diminuir o impacto em tempo consumido para execução dos programas desta dissertação. Por outro lado, todos os autores em algoritmos comparativos utilizaram 10-fold. Para termos mais um indicativo de que os resultados com 3-fold desta dissertação *são robustos*, uma verificação com 10-fold foi executada.

Para tanto, escolheu-se o problema benchmark com menor número de padrões, *o Iris*, e uma das estratégias que menos consumiria tempo de processamento, *a estratégia 5*, devido ao baixo número de avaliações do cromossoma. Os novos resultados mostram um MPPCE e DPPCE no grupo de teste, respectivamente, de 3% e 1%. Resultados bem próximos do observado nas tabelas 4.5 e 4.6, com MPPCE e DPPCE, respectivamente, de 2% e 0% para o Iris e estratégia 5. Através de uma comparação pelo teste t, conclui-se tratarem de médias estatisticamente iguais, e, portanto, indicando a robustez do processo 3-fold-cross-validation.

É fácil perceber pela tabela 4.13 que este experimento com o problema Iris e estratégia 5 seria uma das combinações que menos consumiria tempo de processamento. Considerando que a cada geração de um experimento ocorrerão um número de avaliações igual a 2 vezes o número de indivíduos clássicos configurado para cada estratégia – a 1ª vez para a avaliação dos indivíduos observados, e a 2ª vez para a avaliação dos filhos gerados, conforme figura 3.10 – pode-se obter um número total de avaliações conforme a seguir:

Estratégias	NC	T	Avaliações
1	30	200	11.970
2	60	200	23.940
3	100	200	39.900
4	30	200	11.970
5	15	200	5.985
6	200	200	79.800
7	30	200	11.970
8	60	200	23.940
9	100	200	39.900
10	30	200	11.970
11	15	200	5.985
12	200	200	79.800

Tabela 4.13 – Número de avaliações necessárias por cada experimento.



Onde avaliações= $NC*(2*T-1)$ . Faz-se o desconto de “-1”, pois na primeira geração ocorrerá somente a avaliação de indivíduos quânticos observados, conforme mesma figura 3.10.

Neste processamento em particular para o problema Iris, foram necessárias 5.985 avaliações com a estratégia 5. Um total de 8.977.500 propagações de redes neurais ( $5.985*150*10 = \text{avaliações}*\text{padrões}*\text{folds}$ ). O tempo necessário para este total de propagações foi mensurado igual a 27 minutos, ou 0,45 horas.

Considerando outras variáveis impactantes no tempo de processamento *constantes entre problemas*, através da relação acima de horas por propagações, pode-se obter uma aproximação para o tempo gasto com cada um dos problemas e estratégias nesta dissertação, conforme tabela a seguir:

Horas de Tempo Processamento	Australian Credit	German Credit	Iris	Diagnostic Breast Cancer	Prognostic Breast Cancer	Wine	Empréstimos Financeiros	Crédito Limite Especial
E1	1,24	1,8	0,27	1,02	0,36	0,32	4,71	3,32
E2	2,48	3,6	0,54	2,05	0,71	0,64	9,43	6,64
E3	4,14	6,0	0,90	3,41	1,19	1,07	15,71	11,06
E4	1,24	1,8	0,27	1,02	0,36	0,32	4,71	3,32
E5	0,62	0,9	0,14	0,51	0,18	0,16	2,36	1,66
E6	8,28	12,0	1,80	6,83	2,38	2,14	31,43	22,13
E7	1,24	1,8	0,27	1,02	0,36	0,32	4,71	3,32
E8	2,48	3,6	0,54	2,05	0,71	0,64	9,43	6,64
E9	4,14	6,0	0,90	3,41	1,19	1,07	15,71	11,06
E10	1,24	1,8	0,27	1,02	0,36	0,32	4,71	3,32
E11	0,62	0,9	0,14	0,51	0,18	0,16	2,36	1,66
E12	8,28	12,0	1,80	6,83	2,38	2,14	31,43	22,13

Tabela 4.14 – Tempo aproximado gasto para processar cada problema benchmark por estratégia, em horas.

Somando-se todas as células da tabela acima, tem-se que foram gastos aproximadamente 378,35 horas ou 15,76 dias para se processar todos os experimentos presentes nesta dissertação.

Uma comparação mais aprofundada será fornecida com autores que tenham utilizado de algoritmos neuroevolutivos em seus estudos, conforme tabela 2.5 do capítulo estado da arte. Percebe-se que de todos os artigos, somente os [10], [12], [13], [14], [19], [21], [22], e [29], foram estudados pelo menos 1 dos casos ben-

chmark para classificação apresentados nesta dissertação. A tabela 4.15 oferece um resumo dos resultados encontrados nestes artigos.

As informações coletadas destes outros artigos com algoritmos neuroevolutivos foram: (1) tamanho da população inicial de indivíduos; (2) número de gerações utilizado; (3) MPPCE e DPPCE através de processos de validação cruzada; (4) neurônios otimizado na camada escondida; (5) e tempo computacional de execução do algoritmo em segundos. Em caso de indisponibilidade de resultados, a denominação “-“ era atribuída.

Em alguns artigos o processo evolutivo continuava até que algum critério de minimização de erro era satisfeito, e não como acontece comumente em AE, quando um número de gerações era atingido. Neste último caso, a coluna Geração estará indicada pela denominação de "*Minimum Improvement*".

A primeira conclusão que se chega é que é difícil a comparação entre artigos com algoritmos neuroevolutivos, uma vez que não existe uma grande uniformidade na condução dos experimentos e apresentação dos resultados por diversos autores. Sobre o tamanho da população, observa-se valores próximo de 100, enquanto nesta dissertação, algumas estratégias com a 5 e a 11 utilizaram apenas 15 indivíduos clássicos com desempenho satisfatório já com provado através de testes estatísticos .

Não será possível a comparação de parâmetros quânticos, como por exemplo, o número de indivíduos quânticos  $N_Q$ , pois não se dispõe de nenhum artigo na literatura com abordagem neuroquântica em problemas de classificação.

Em termos de gerações, observam-se valores bem próximos aos utilizados nesta dissertação.

Para os artigos que apresentaram números de neurônios otimizados na camada escondida, e, sobretudo, existia uma orientação para este tipo de otimização, observou-se valores bastante inferiores ao desta dissertação. Percebe-se nestes artigos que a evolução do número de neurônios iniciava com valores baixos próximos de 5, e aumentavam conforme processo evolutivo. A mecânica de nosso algoritmo, o NEIQ-BR, em permitir que soluções caminhassem livremente por um domínio alto de possíveis neurônios na camada escondida pode ter acarretado em algum ganho significativo no poder de aprendizado e generalização da rede.

Problema	Algoritmo	População	Geração	MPPCE	DPPCE	Neurônios na Camada Escondida	Tempo Processamento (segundos)	Artigo
Australian	RBF-Batch	100	200	16,67	3,87	-	-	[10]
Australian	RBF-DF	100	200	16,28	2,54	-	-	[10]
Australian	RBF-DFIO	100	200	16,67	4,28	-	-	[10]
Australian	RBF-GA	100	200	13,95	3,53	-	-	[10]
Australian	RBF-IO	100	200	17,83	3,96	-	-	[10]
Australian	RBF-IODF	100	200	17,25	4,44	-	-	[10]
Australian	RBF-On-line	100	200	16,86	4,39	-	-	[10]
Australian	RBF-Optimal	100	200	15,89	4,66	-	-	[10]
Australian	MOBNET	100	"Minimum Improvement"	10,05	0,65	20	-	[12]
Australian	CCONN	100	20	12,17	0,01	7	-	[13]
Australian	COVNET	100	"Minimum Improvement"	11,57	0,01	4	-	[14]
Australian	GA	120	100	15,29	2,33	-	12	[19]
Australian	BKS+KL	20	200	9,6	1,4	5	-	[21]
Australian	GA	12	3000	12,21	0,65	3	853	[29]
Diagnostic BC	GA	12	3000	1,15	0,34	3	510	[29]
German	GA	120	100	29,0	3,16	-	12	[19]
Iris	GA	120	100	6,40	2,41	-	12	[19]
Wine	GA	120	100	13,94	10,59	-	12	[19]

Tabela 4.15 – Resultados de outros autores com algoritmos neuroevolutivos para os problemas estudados nesta dissertação.

Sobre o tempo de processamento, para os casos de comparação disponíveis, observa-se que o algoritmo NEIQ-BR é bastante superior, mesmo considerando que os valores na tabela 4.14 são aproximações. Percebeu-se através da leitura dos artigos que se tratando especificamente da área de AE, a sociedade se preocupa em criar algoritmos onde a sinergia entre indivíduos e geração seja a mais eficiente possível – como foi verificado para o algoritmo NEIQ-BR –, enquanto que o tempo computacional consumido passa a ser um aspecto secundário. Nossa análise sobre o tempo de processamento, ou custo computacional terminará aqui, pois como defendido em capítulos anteriores, este não foi o objetivo central desta dissertação.