

Algoritmo evolucionário com inspiração quântica para síntese de sistemas de classificação fuzzy

Waldir Nunes, Marley Vellasco, Ricardo Tanscheit

Departamento de Engenharia Elétrica,
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio),
Rua Marquês de São Vicente, 255 - 22451-900, Rio de Janeiro – RJ – Brazil

wgalluzzi@gmail.com; {marley, ricardo}@ele.puc-rio.br

Resumo. Sistemas de Inferência Fuzzy, dentre as ferramentas capazes de representar sistemas reais, são considerados instrumentos muito úteis em comparação com outras técnicas, por viabilizarem a representação de conhecimento de forma humanamente compreensível, por meio de suas regras de inferência. Contudo, quando empregados para representar sistemas complexos, sua fidelidade em relação ao original (exatidão) ou sua compreensibilidade são comprometidas. Procurando identificar métodos de síntese desses sistemas que atendam a ambos os objetivos (exatidão e compreensibilidade), este trabalho propõe a utilização de Algoritmos Evolucionários com Inspiração Quântica. Trata-se de um caso particular de Algoritmo de Estimação de Distribuição, aplicável a este processo de síntese, como uma alternativa viável pelo seu bom desempenho frente a técnicas evolutivas mais tradicionais. Além do uso dessa técnica inovadora para a síntese de Sistemas de Inferência Fuzzy, os Algoritmos Evolucionários com Inspiração Quântica propostos também representam uma contribuição na área de Algoritmos de Estimação de Distribuição e são aplicáveis a outros problemas que não o desíntese referenciado.

Palavras-chave: Sistemas-fuzzy; Algoritmos evolucionários com inspiração quântica; Algoritmos de estimação de distribuição; Sistemas fuzzy genéticos.

1 Introdução

Sistemas de inferência fuzzy (SIF) são ferramentas interessantes para tarefas de classificação, pois, geralmente apresentam bom desempenho e podem ser usados como demonstradores de conceito. Quando as regras de inferência são significativas para os seres humanos, o conhecimento que elas contêm pode ser facilmente compreendido. Uma das contribuições pioneiras [1] para a construção de sistemas classificadores construídos por meio de algoritmos genéticos considera duas etapas: (i) o particionamento do universo de discurso (UoD) de variáveis em subespaços e (ii) a determinação de uma regra de inferência para cada subespaço obtido na etapa anterior. Na citada contribuição, é utilizado particionamento fixo e simétrico do espaço de variáveis em N conjuntos (termos), resultando em $(N)^{\text{Número de variáveis}}$ regras.

O desempenho do classificador assim obtido depende de uma escolha adequada do particionamento: se for muito grosseiro (poucos conjuntos fuzzy por variável), muitos padrões podem ser erroneamente classificados; se for demasiado refinado (muitos conjuntos por variável), muitas regras deixarão de ser geradas devido à falta de exemplos para a fase de aprendizado. Uma abordagem para superar este problema é a utilização de vários SIF em paralelo, cada um com um nível de particionamento, de forma que se tenha $\sum_i (N_i)^{\text{Número de Variáveis}}$ regras, onde i varia de poucas a muitas partições por variável. Esquemáticamente, a ideia é ilustrada na Fig. 1. Isto resulta em classificadores eficientes, mas com um número excessivo de regras.

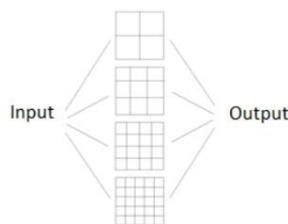


Fig. 1. Processo de inferência utilizando o espaço de variáveis (com apenas duas variáveis de entrada) dividido em 2, 3, 4 e 5 conjuntos fuzzy para cada variável. Com base em [1].

Em [2], uma evolução da contribuição anterior, aplica-se um algoritmo genético sobre o conjunto original de $\sum_i (N_i)$ Número de Variáveis, para seleção de regras, obtendo-se um subconjunto menor e razoavelmente eficiente de regras para composição de um classificador. Essas duas contribuições são conceitualmente importantes, por indicarem a influência do particionamento na acuracidade do classificador e por expressarem a preocupação com o tamanho da base de regras (visando à redução dos requisitos de armazenamento). Este problema também é relevante nas discussões sobre compreensibilidade do processo de classificação, como observado em [3]. Deve-se ressaltar que nas conclusões de [2] menciona-se que o método proposto seleciona regras altamente significativas e pode, portanto, ser considerado como um método de aquisição de conhecimento para problemas de classificação, onde o conceito de classificação estaria incorporado nas regras selecionadas.

Num estudo subsequente ([4]), regras são geradas por meio de um algoritmo genético (ao contrário do método anterior, que foi baseado em um conjunto pré-definido e exaustivo de regras sobre o qual é aplicada seleção). Neste caso, se utiliza particionamento fixo do espaço de variáveis e, para evitar a explosão combinatória da base de regras, um algoritmo genético produz combinações de conjuntos de condições antecedentes, que são em seguida transformadas em regras.

Cada indivíduo do algoritmo genético representa uma regra: a sequência Term1,1 | Term1,2 | ... Term1,N, forma um indivíduo, onde Termi,j corresponde a um conjunto fuzzy (que pode ser Don'tCare - DC) associado à variável j , do indivíduo Regrai. O algoritmo genético fornece, para cada indivíduo, uma combinação de antecedentes, a qual é utilizada para realizar um processo estatístico de aprendizado do melhor

consequente para a regra, bem como a identificação de uma medida da sua qualidade (grau de certeza da regra).

O classificador final é composto pela união de regras que formam o todo (ou parte) da população (arquitetura Michigan). Esse método visa à compreensão semântica, uma vez que a compartimentação do espaço de variáveis é pré-definido e formado com labels compreensíveis ("pequeno", "médio" e "grande", por exemplo). Além disso, ele também admite DC, o que torna cada regra potencialmente menor.

No entanto, a utilização de particionamento predefinido pode comprometer a eficiência do classificador devido à rigidez dos limites de decisão. É por isso que [5] enfatiza o processo de aprendizagem da base de dados e da base de regras simultaneamente, embora isto tenha um custo computacional bastante elevado.

Uma questão igualmente importante é o fato de que esta técnica usa a arquitetura Michigan ([5] e [6]), na qual cada indivíduo representa uma regra. Isto é necessário porque, quanto maior for o número de variáveis de entrada, maior será a representação genética de uma regra e ainda maior a representação de uma base de regras completa. Isso faz com que a adoção da arquitetura Michigan seja uma escolha natural. No entanto, tal como referido em [6], apesar dos sistemas baseados na arquitetura Michigan tenderem a produzir boas regras, eles podem não conduzir a um bom classificador quando as boas regras obtidas são combinadas.

Subsequentemente, [7] propõe a utilização da técnica apresentada em [4], mas com um número pré-definido (e pequeno) de regras, de modo que é possível representar toda a base de regras em um indivíduo (arquitetura Pittsburgh). Esta opção permite que as regras geradas possam ser avaliadas em conjunto, conduzindo a uma avaliação mais realista em termos de classificadores finais, com um pequeno número de regras, em benefício da compreensibilidade do classificador.

A técnica apresentada em [7] resulta em classificadores com menos regras, cada uma com poucos antecedentes e um bom nível de eficiência. Mas sua obtenção envolve um problema de otimização em um grande espaço de busca, em geral com alto custo computacional. Esta foi a principal motivação para o desenvolvimento deste trabalho, que propõe uma significativa evolução sobre Algoritmos Evolucionários com Inspiração Quântica (QIEA) [8], que são um caso particular de Algoritmo de Estimativa de Distribuição (EDA), conforme será apresentado na seção 2. Trata-se de uma alternativa viável pelo seu bom desempenho frente a técnicas evolutivas mais tradicionais, acrescido de capacidade de tratamento de múltiplos objetivos e de uma heurística em que se trabalha com regiões promissoras do espaço de buscas, como será apresentado na seção 3. Além do uso dessa técnica para a síntese de SIF, os QIEA propostos também representam uma contribuição na área de EDA e são aplicáveis a outros problemas que não o de síntese referenciado.

A seção 4 apresenta os resultados da aplicação da técnica em casos de referência encontrados na literatura, para fins de comparação de desempenho com outras técnicas, ao que se seguem conclusões.

2 Algoritmos Evolucionários com Inspiração Quântica e Algoritmos de Estimação de Distribuição

Um QIEA [8] é um tipo de algoritmo evolucionário que faz uso de conceitos de computação quântica e que pode ser executado em computadores convencionais, em vez de máquinas de processamento quântico. Inspirado nos conceitos de computação quântica, tais como q-bits e portas quânticas, os QIEA têm as três características principais que se seguem: (i) um QIEA adota um tipo de representação baseada na utilização de Q-bits (inspirados em q-bits da computação quântica) para descrever os indivíduos numa população. Um Q-bit permite representar probabilisticamente uma superposição de estados simultâneos; (ii) Um QIEA emprega uma emulação de portas quânticas (Q-gates) para gerar indivíduos de uma próxima geração, as quais orientam indivíduos na direção das melhores soluções; e (iii) Um QIEA pode explorar o espaço de busca para uma solução global com um pequeno número de indivíduos.

No primeiro QIEA proposto [9], indivíduos são representados por Q-bits, Q-gates são usados como operadores evolucionários para fornecer os indivíduos com melhores habilidades, e um processo de "observação" é empregado, a fim de conectar a representação em Q-bits com a capacidade de processamento dos computadores clássicos. Esse algoritmo pioneiro pode ser considerado um Algoritmo de Estimação de Distribuição ([10]), uma vez que manipula distribuições de probabilidades em busca de regiões promissoras, levando a soluções ótimas ou quase ótimas.

Uma análise comparativa entre EDA e QIEA pode ser encontrada em [8]. O desempenho dos QIEA é considerado melhor do que o dos EDA que não tratam dependências [11], porque operam com mais de um modelo probabilístico simultaneamente.

De acordo com [8], alguns algoritmos evolucionários classificados como QIEA apresentam diferenças conceituais, quando comparados com o conceito original dos QIEA. Um deles é o referenciado em [12], mas que manipula vários modelos probabilísticos como indivíduos (conforme [11]). Em [12], um par de valores constituído pela média e a largura de um pulso quadrado é utilizado para representar um gene, permitindo a sua associação com uma função de densidade de probabilidade uniforme, capaz de modelar uma variável aleatória real. O processo evolucionário é realizado em três etapas: crossover, translação e escalonamento. O operador de crossover permite a troca de características entre indivíduos da população clássica e os melhores indivíduos já produzidos (elite). Os operadores de translação e escalonamento são aplicados para modificar a média ρ e a largura σ do pulso quadrado respectivamente. Trata-se de um algoritmo muito parecido com um EDA, mas, como emprega vários modelos probabilísticos (indivíduos quânticos), a sua classificação como QIEA parece mais apropriada.

Este trabalho propõe a utilização dos benefícios do uso de vários modelos probabilísticos no processo evolucionário, como visto nos QIEA, em vez de EDA. Além disso, considera o uso de um modelo probabilístico com representação mais sofisticada, como proposto em [12], que é estendido para a representação de elementos categóricos, valendo-se de uma heurística que faz uso de relações conhecidas entre esses elementos categóricos.

3 QIEA categórico com o tratamento de afinidades

Uma etapa muito importante na síntese de SIF é a definição dos antecedentes das regras. Considere a Fig. 2, em que um determinado atributo de entrada é dividido em três ("baixo", "médio" e "alto"), cinco ("muito baixo", "baixo", "médio", "alto" e "muito alto") e sete conjuntos simétricos ("muito baixo", "baixo", "médio-baixo", "médio", "média-alto", "alto" e "muito alto").

Se uma regra de inferência tem bom desempenho e faz referência a um antecedente com o valor "muito baixo", no particionamento de cinco conjuntos de uma determinada variável, ela provavelmente terá um bom desempenho para valores "baixo", na divisão em três conjuntos, ou "muito baixo" ou "baixo", na divisão em sete conjuntos.

Em outras palavras, na síntese de sistemas de inferência fuzzy, a informação sobre as regiões de interesse é muito relevante. No exemplo apresentado acima, a região promissora citada abrangea região coberta pelo conjunto "baixo" do particionamento em três termos, "muito baixo" e "baixo", do particionamento em cinco termos, e "muito baixo", "baixo" e "médio-baixo", quando sete termos são considerados. A Fig. 2 mostra três regiões de interesse.

O método proposto aqui considera as regiões de interesse ou regiões promissoras em benefício do processo de otimização. O uso de regiões promissoras para otimização usando estratégias evolucionárias é uma característica inerente aos QIEA, que foi empregada em [12]. Esse conceito é expandido aqui para a formulação de um QIEA para otimizar combinações de atributos categóricos (conjuntos fuzzy utilizados como antecedentes de regras) que mantêm alguma afinidade entre si.

Note-se que os SIFa serem produzidos fazem uso de mais de um particionamento por variável (pré-definido e com diferentes granularidades), e favorece a geração de regras que cubram hiper-volumes promissores de variáveis de entrada (promissores em termos de estarem ligados a regras que compõem bons classificadores), bem como aqueles com poucos antecedentes (usando DC) em benefício da compreensibilidade obtida por poucas regras e regras curtas.

O algoritmo proposto pode selecionar elementos categóricos do conjunto {"low" (3), "medium" (3), "high" (3), "verylow" (5), "low" (5), "medium" (5), "high" (5), "very high" (5), "verylow" (7), "low" (7), "mediumlow", (7), "medium" (7), "medium high" (7), "high" (7), very high" (7), DC}, em que o número entre parênteses corresponde ao número de partições. A seleção de um antecedente, entre os possíveis da base de dados, ocorre através da observação dos genes quânticos, que armazenam estes elementos possíveis e suas probabilidades de ocorrência.

A Fig. 3 mostra a associação entre os elementos que podem ser observados num gene quântico (os elementos categóricos acima mencionados), e as suas probabilidades iniciais de ocorrência, como usado na inicialização do algoritmo. O elemento DC também pode ser observado e é inicialmente favorecido com uma alta probabilidade (p) de observação, de modo a proporcionar um elevado número de ocorrências (por uma questão de facilidade de interpretação do sistema final). Os outros elementos observáveis são inicialmente equiprováveis. Assim, um gene quântico consiste dos seguintes campos: (i) um conjunto de elementos observáveis, (ii) as probabilidades de ocorrência de cada um destes elementos observáveis, e (iii)

Categorical element	Probability
don't care	P
"low" (3)	(1 - p) / 15
"medium" (3)	(1 - p) / 15
"high" (3)	(1 - p) / 15
"very low" (5)	(1 - p) / 15
"low" (5)	(1 - p) / 15
"medium" (5)	(1 - p) / 15
"high" (5)	(1 - p) / 15
"very high" (5)	(1 - p) / 15
"very low" (7)	(1 - p) / 15
"low" (7)	(1 - p) / 15
"medium-low" (7)	(1 - p) / 15
"medium" (7)	(1 - p) / 15
"medium-high" (7)	(1 - p) / 15
"high" (7)	(1 - p) / 15
"very high" (7)	(1 - p) / 15

Fig. 3. Associação elemento observável x probabilidade inicial, para um gene quântico.

Da mesma forma, se depois de algumas gerações se observa que um determinado antecedente para uma determinada regra apresenta um valor diferente de DC para os classificadores de boa qualidade, é provável que a região de interesse daquela variável naquela regra seja composta por todos os elementos possíveis, exceto DC. Isto é repetido várias vezes e, assim, a sub-região mais adequada é determinada, por meio do estreitamento gradual das regiões de interesse (reduzindo-se a sua cardinalidade).

a)

$q_{1,1}$	$q_{1,2}$...	$q_{1,N}$	$q_{2,1}$	$q_{2,2}$...	$q_{2,N}$...	$q_{M,1}$	$q_{M,2}$...	$q_{M,N}$
-----------	-----------	-----	-----------	-----------	-----------	-----	-----------	-----	-----------	-----------	-----	-----------

b)

Quantum Gene $q_{1,1}$	Quantum Gene $q_{1,2}$	Quantum Gene $q_{M,N}$
Don't Care $\rightarrow p$	Don't Care $\rightarrow p$	Don't Care $\rightarrow p$
"low" (3) $\rightarrow (1 - p) / 15$	"low" (3) $\rightarrow (1 - p) / 15$	"low" (3) $\rightarrow (1 - p) / 15$
"medium" (3) $\rightarrow (1 - p) / 15$	"medium" (3) $\rightarrow (1 - p) / 15$	"medium" (3) $\rightarrow (1 - p) / 15$
"high" (3) $\rightarrow (1 - p) / 15$	"high" (3) $\rightarrow (1 - p) / 15$	"high" (3) $\rightarrow (1 - p) / 15$
"very low" (5) $\rightarrow (1 - p) / 15$	"very low" (5) $\rightarrow (1 - p) / 15$	"very low" (5) $\rightarrow (1 - p) / 15$
"low" (5) $\rightarrow (1 - p) / 15$	"low" (5) $\rightarrow (1 - p) / 15$	"low" (5) $\rightarrow (1 - p) / 15$
"medium" (5) $\rightarrow (1 - p) / 15$	"medium" (5) $\rightarrow (1 - p) / 15$	"medium" (5) $\rightarrow (1 - p) / 15$
"high" (5) $\rightarrow (1 - p) / 15$	"high" (5) $\rightarrow (1 - p) / 15$	"high" (5) $\rightarrow (1 - p) / 15$
"very high" (5) $\rightarrow (1 - p) / 15$	"very high" (5) $\rightarrow (1 - p) / 15$	"very high" (5) $\rightarrow (1 - p) / 15$
"very low" (7) $\rightarrow (1 - p) / 15$	"very low" (7) $\rightarrow (1 - p) / 15$	"very low" (7) $\rightarrow (1 - p) / 15$
"low" (7) $\rightarrow (1 - p) / 15$	"low" (7) $\rightarrow (1 - p) / 15$	"low" (7) $\rightarrow (1 - p) / 15$
"medium-low" (7) $\rightarrow (1 - p) / 15$	"medium-low" (7) $\rightarrow (1 - p) / 15$	"medium-low" (7) $\rightarrow (1 - p) / 15$
"medium" (7) $\rightarrow (1 - p) / 15$	"medium" (7) $\rightarrow (1 - p) / 15$	"medium" (7) $\rightarrow (1 - p) / 15$
"medium-high" (7) $\rightarrow (1 - p) / 15$	"medium-high" (7) $\rightarrow (1 - p) / 15$	"medium-high" (7) $\rightarrow (1 - p) / 15$
"high" (7) $\rightarrow (1 - p) / 15$	"high" (7) $\rightarrow (1 - p) / 15$	"high" (7) $\rightarrow (1 - p) / 15$
"very high" (7) $\rightarrow (1 - p) / 15$	"very high" (7) $\rightarrow (1 - p) / 15$	"very high" (7) $\rightarrow (1 - p) / 15$

Fig. 4. Representação de indivíduo quântico. Em 4(a), $q_{i,j}$ representa um gene quântico associado a uma distribuição de probabilidades discreta (elementos categóricos aplicáveis à condição antecedente j da regra i). Em 4(b), são apresentados os valores inicialmente possíveis e suas probabilidades iniciais de observação para $q_{i,j}$.

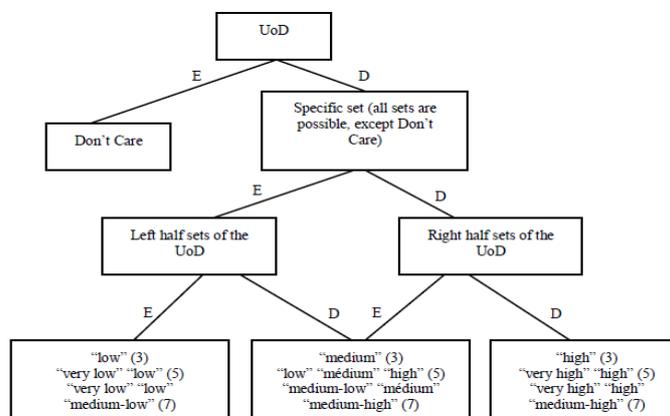


Fig. 5. Árvore mostrando a convergência das regiões promissoras. Uma árvore como esta está associada a cada gene quântico (ou, um antecedente de uma regra).

Os passos seguintes descrevem todo o processo:

Passo 1: Na inicialização do algoritmo, cada um dos genes quânticos aponta para a raiz da árvore de regiões promissoras (a árvore mostrada na Figura 5). Quando observado, pode produzir um elemento clássico do tipo *DC* ou qualquer outro termo;

Passo 2: A partir da observação de indivíduos quânticos, indivíduos clássicos são produzidos, que correspondem a *M* combinações de antecedentes das variáveis de entrada. Cada combinação se tornará uma regra. Para cada combinação de antecedentes, um processo de aprendizagem de consequente para a regra é realizado, com o mesmo método empregado em [7] –processo que resulta em um classificador;

Passo 3: O desempenho do classificador é avaliado para cada indivíduo clássico, e os melhores indivíduos clássicos indicam quais sub-regiões promissoras ocorreram mais vezes nos classificadores de boa qualidade. Por exemplo, se um bom classificador tem uma regra (*i*) com antecedente "muito alto" (5) para uma determinada variável de entrada (*j*), um contador denominado *OcorrenciaEmBonsClassificadores* do conjunto "muito alto" (5) para o antecedente *j* da regra *i* é incrementado;

Passo 4: Depois de algumas gerações, a sub-região promissora com maior número de ocorrências é identificada e cada um dos genes quânticos aponta para esta nova sub-região (a transição ocorre a um nível inferior da árvore apresentada na figura 5);

Passo 5: A transição para o nível mais baixo da árvore ocorre como indicado no passo anterior, até que atinja o seu fim (que corresponde sempre a um subconjunto relativamente menor do conjunto original de categorias);

Passo 6: os melhores indivíduos clássicos são preservados (pelo processo de elitismo), e, periodicamente, após várias gerações, aqueles indivíduos quânticos que não geraram indivíduos clássicos de boa qualidade têm seus genes quânticos reinicializados (passam a apontar novamente para a raiz da árvore da Fig.5), de modo a expandir a capacidade de exploração do algoritmo, processo que guarda similaridade com a operação de mutação dos algoritmos genéticos.

Em resumo, inicialmente, a região promissora considerada em cada gene quântico é todo o universo de discurso de cada variável. Depois de algum tempo, é possível restringir esta região, tal como sugerido na Fig. 5: se, depois de algumas gerações, observa-se que nos classificadores de boa qualidade houve um grande número de

ocorrências de *DC* em um antecedente de regra, é provável que a região de interesse adequada para o teste daquela variável naquela regra é um único elemento (*DC*).

3.1. Tratamento de Múltiplos Objetivos

O algoritmo também contempla o tratamento de múltiplos objetivos. Para isso, o processo de identificação de fronteiras de Pareto, proposto em NSGA-2 [13], é mesclado com a QIEA. O NSGA-2 é um algoritmo genético que, para cada geração, produz uma nova população, resultante de operações de mutação e de cruzamento de pais selecionados a partir da geração anterior, fundindo-se duas gerações (indivíduos antigos e novos). Dessa junção se aproveitam os melhores indivíduos, considerando simultaneamente os múltiplos objetivos e que se preserve um bom nível de diversidade de soluções, formando-se fronteiras de Pareto progressivamente refinadas a cada geração. Ao final do processo, se tem identificado um conjunto diverso e de boa qualidade de soluções, frente a todos os objetivos, compondo uma boa aproximação da fronteira teórica de Pareto.

O conceito apresentado em [14], que mescla o NSGA-2 com QIEA binário, foi empregado na presente proposta, sendo que os indivíduos clássicos que compõem a fronteira de Pareto (ou seja, os classificadores de melhor qualidade) são considerados para a evolução dos indivíduos quânticos (conforme 3º passo apresentado acima).

A avaliação de cada indivíduo clássico é realizada com três objetivos: (i) máxima eficiência na classificação, (ii) mínima quantidade de antecedentes de regras, e (iii) mínima quantidade de regras ativas (uma regra ativa apresenta ao menos uma condição antecedente com teste diferente de *DC*).

3.2. A implementação completa

Ao indivíduo quântico apresentado na seção anterior também foi incorporado um gene quântico com representação de reais para cada particionamento de cada variável de entrada, para implementação de um desvio (offset). Por exemplo, se o classificador a ser construído é composto por quatro variáveis de entrada, e cada variável é particionada em 3, 5 e 7 termos linguísticos, existem 4×3 genes quânticos com representação de reais (conforme [12]), agregados ao indivíduo quântico, de tal maneira que cada número real será aplicado como um offset nos particionamentos de uma variável.

$q_{1,1}$	$q_{1,2}$...	$q_{1,N}$	$q_{2,1}$	$q_{2,2}$...	$q_{2,N}$...	$q_{3,1}$	$q_{3,2}$...	$q_{3,N}$
-----------	-----------	-----	-----------	-----------	-----------	-----	-----------	-----	-----------	-----------	-----	-----------

Fig. 6. Representação de genes quânticos utilizados para o ajuste da base de dados.

A representação final é constituída pela concatenação do indivíduo da Figura 4 (a) e da Fig. 6. A Fig. 7 apresenta o efeito da aplicação do offset. Como mencionado, o tratamento de representação quântica de valor real foi feito pela incorporação da técnica proposta em [12], resultando numa representação de QIEA híbrido (categórica

com afinidades de tratamento, e real), mas também incluindo o tratamento multiobjetivos.

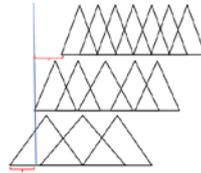


Fig. 7. No alto, um valor de offset é aplicado ao particionamento em 7 conjuntos de uma variável e na parte de baixo, um deslocamento negativo ao particionamento em 3 conjuntos.

4 Resultados experimentais

Para avaliação do desempenho do algoritmo proposto foram utilizados os dados de duas publicações, [7] e [15]. Ambas utilizam a mesma forma de avaliação proposta em [7] e utilizada em [15] para comparação de seus resultados com os resultados da primeira.

Foram considerados seis conjuntos de dados com atributos numéricos, disponíveis em <http://www.ics.uci.edu/~mllearn/>. A Tabela 1 apresenta informações sobre esses conjuntos. Alguns deles incluem padrões incompletos, que foram desconsiderados.

Tabela 1. Conjuntos utilizados nas simulações computacionais (fonte [7])

Conjunto	Conjuntos utilizados		
	Atributos (n)	Padrões (m)	Classes (M)
Breast W	9	683*	2
Diabetes	8	768	2
Glass	9	214	6
Hart C	13	297*	5
Sonar	60	208	2
Wine	13	178	3

*Padrões com valores inexistentes foram desconsiderados

O desempenho de cada classificador, para os conjuntos de dados acima citados, tanto para [16], quanto para [7] e [15], foi avaliado com 10 repetições de validação cruzada do tipo k -fold (para $k = 10$), ou seja, 10×10 -fold cross validation ([17]).

Ressalte-se que [16] produziu classificadores com o uso do C4.5 (não se tratavam de SIF para classificação), [7] produziu SIF para classificação, partindo de regras inicialmente formadas pelo aprendizado estatístico da combinação de antecedentes, e [15] produziu SIF para classificação, partindo de um classificador construído com o uso de C4.5, que foi perturbado e utilizado como população inicial em um processo evolucionário. Ou seja, todos os métodos apresentados utilizaram uma forma de inicialização que direcionava fortemente a construção do classificador para um desempenho bastante elevado, para depois, em [7] e [15], submetê-lo a processo evolutivo.

O algoritmo proposto neste trabalho não parte de qualquer indivíduo previamente conformado para ter alto desempenho. Esta premissa foi adotada para submetê-lo ao elevado esforço de aprendizado da base de dados e da base de regras simultaneamente, e com os mesmos tamanhos de população e número de épocas.

Ainda, em [7], na construção das regras da população inicial, por meio de aprendizado estatístico, foi possível introduzir um esquema de priorização probabilística para o uso de DC, em que se atribuiu $p(\text{DC})$ igual a 0,95 para o conjunto de dados Sonar e 0.8 para os demais conjuntos de dados. Por fim, em [7] e [15] foi utilizada população de 200 indivíduos e 5000 gerações (perfazendo um total de 1.000.000 avaliações por execução), havendo 100 execuções para treinamento e 100 execuções para teste, para cada conjunto de exemplos, em decorrência do uso de 10×10 -fold crossvalidation. Resultados são apresentados nas tabelas que se seguem.

Tabela 2. Melhores taxas percentuais de erro médio para classificadores atuando nos conjuntos de treinamento

Melhores taxas médias de erro em conjuntos de treinamento								
Data	[15]	MOP-I	MOP-2	MOP-3	SOP-1	SOP-2	SOP-3	Proposta
Glass	9,50	25,11	27,08	25,94	17,81	21,92	22,36	35,60
Sonar	1,95	8,55	8,69	8,42	3,55	5,89	5,82	14,88
Wine	0,60	0,01	0,10	0,03	0,00	0,00	0,00	1,97
Cleve	23,18	33,43	35,05	34,59	25,72	29,65	29,98	40,75
Wisc	1,94	1,59	1,71	1,74	1,08	1,44	1,51	2,70
Pima	16,17	19,48	19,79	19,59	17,74	18,37	18,41	22,82

Tabela 3. Melhores taxas percentuais de erro médio para classificadores atuando nos conjuntos de teste

Melhores taxas médias de erro em conjuntos de teste									
Data	[15]	MOP-I	MOP-2	MOP-3	SOP-1	SOP-2	SOP-3	C4.5[16]	Proposta
Glass	24,05	35,55	33,93	34,05	35,76	39,21	38,36	27,30	38,00
Sonar	16,73	23,18	17,32	17,51	24,04	23,47	24,29	24,60	10,00
Wine	2,98	3,99	3,65	3,04	7,30	6,49	6,52	5,60	3,76
Cleve	38,22	42,57	42,85	42,64	44,83	45,80	45,44	46,30	37,93
Wisc	2,95	2,93	2,74	2,66	3,88	3,69	3,56	5,10	2,76
Pima	21,78	23,27	22,32	21,80	25,26	25,00	24,20	25,00	22,89

Em [7], foram utilizadas seis diferentes abordagens para o tratamento de múltiplos objetivos. Nas três primeiras foi empregado algoritmo baseado no NSGA-II (multiobjetivos: MOP-1, 2 e 3) e, nas três últimas, tratamento de apenas um objetivo por vez, ponderado com pesos previamente estabelecidos.

A partir das tabelas, é possível observar que o desempenho dos classificadores obtidos com o uso do algoritmo proposto, para conjuntos de treinamento, foi inferior

ao das demais técnicas, embora próximo. Contudo, a obtenção de classificadores com o algoritmo proposto não partiu de nenhum classificador ou conjunto de regras previamente aprendidas, muito menos utilizando ferramentas estatísticas que descaracterizam completamente a base de dados (o que é o caso de [15], que parte de um classificador aprendido com C4.5, ou mesmo do classificador formado diretamente com o C4.5, que também consta das tabelas). É notável o bom desempenho do algoritmo proposto para conjuntos de teste, frente às demais técnicas.

Ou seja, a aplicação de algoritmos evolucionários nos outros trabalhos pode, de certa forma, ser considerada uma aplicação de ajuste de classificadores existentes, o que não foi o caso do algoritmo proposto, que foi aplicado no aprendizado simultâneo da base de dados e da base de regras, sem qualquer composição preliminar dos indivíduos.

5. Conclusões

Este artigo apresentou um novo algoritmo evolutivo do tipo QIEA, adaptado para lidar com elementos categóricos que apresentam uma relação conhecida de afinidade, tendo se obtido um desempenho comparável e em alguns casos superior a outras técnicas, para a síntese de classificadores construídos por meio de SIF.

As contribuições deste trabalho, em relação ao QIEA original, partiram do uso da representação real (conforme [12]), agregando-se a representação categórica com o uso da heurística de afinidades, resultando em um QIEA com representação híbrida e tratamento de múltiplos objetivos, capaz de aprender a base de dados (por meio de sintonia) e a base de regras, simultaneamente.

O algoritmo proposto, como outros EDA e QIEA, pode lidar com espaços de busca complexos, como é o caso da síntese de SIFclassificadores, pois se trata de um problema que envolve um grande espaço de busca e para o qual não é viável a identificação de dependências entre as variáveis.

Referências

1. Ishibuchi, H., Nozaki, K., Tanaka, H.: Distributed representation of fuzzy rules and its application to pattern classification. *Fuzzy Sets and Syst.*, vol. 52, pp. 21-32 (1992)
2. Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms. *IEEE Transactions on Fuzzy Systems*, vol. 3, No 3 (1995)
3. Gacto, M. J., et al.: Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. *Inform. Sci.* (2011)
4. Ishibuchi, H., Nakashima, T., Murata, T.: Performance Evaluation of Fuzzy Classifier Systems for Multidimensional Pattern Classification Problems. *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 29, No 5 (1999)
5. Herrera, F.: Genetic Fuzzy Systems: Taxonomy, Current Research Trends and Prospects. *Evolutionary Intelligence* 1 27-46(2008)

6. Ishibuchi, H., Nakashima, T., Tanaka, T.: Comparison of the Michigan and Pittsburgh Approaches to the Design of Fuzzy Classification Systems. *Electronics and Communications in Japan, Part 3, Vol 80, No 12* (1997)
7. Ishibuchi, H., Nojima, Y.: Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning. *International Journal of Approximate Reasoning*, vol. 44, no. 1, pp. 4-31 (2007)
8. Zhang, G.: Quantum-inspired evolutionary algorithms: a survey and empirical study. *Journal of Heuristics*, pages 1-49 (2010)
9. Han, K.-H., Kim, J.-H.: Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. *IEEE Trans. Evol. Comput.*, vol. 6, pp. 580-593 (2002)
10. Pelikan, M., Hauschild, M. W., Lobo, F. G.: Introduction to Estimation of Distribution Algorithms. Missouri Estimation of Distribution Algorithms Laboratory (MEDAL) Report No. 2012003 (2012)
11. Platel, M. D., Schliebs, Kasabov, S., N.: Quantum-Inspired Evolutionary Algorithm: A Multimodel EDA. *IEEE Transactions on Evolutionary Computation*, Vol. 13, No. 6 (2009)
12. Abs da Cruz, A. V., Vellasco, M.M.B.R., Pacheco, M.A.C.: Quantum-Inspired Evolutionary Algorithm for Numerical Optimization. *Book Series Studies in Computational Intelligence, Book Quantum Inspired Intelligent Systems*, Vol. 121, 156 p., Publisher Springer Berlin / Heidelberg, pp. 115-132 (2008)
13. Deb, K., Pratap, A., Agarwal, S., Meyer, T.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 2 (2002)
14. Kim, Y., Kim, J.-H., Han, K.-H.: Quantum-inspired Multiobjective Evolutionary Algorithm for Multiobjective 0/1 Knapsack Problems. *IEEE Congress on Evolutionary Computation*. Vancouver, BC, Canada (2006)
15. Pulkkinen, P., Hytonen, J., Kolvisto, H.: Developing a bioaerosol detector using hybrid genetic Fuzzy systems. *Eng. Appl. Artif. Intell.*, vol. 21, no. 8, pp. 1330-1346 (2008)
16. Elomaa, T., Rousu, J.: General and efficient multisplitting of numerical attributes. *Mach Learning*, 36(3): 201-244 (1999)
17. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. *International joint Conference on artificial intelligence*. [S.l.: s.n.], v. 14, p. 1137-1145 (1995)