



Using the Choquet Integral in the Fuzzy Reasoning Method of Fuzzy Rule-Based Classification Systems

Gabriel Rosa

Foco do artigo e Objetivos



- O objetivo do artigo é apresentar um novo Método de Raciocínio Fuzzy(FRM) com a integral de Choquet sendo usada como função de agregação;
- É proposto também um método de aprendizado através de um *algoritmo genético*(GA) que se mostrou o mais apto para a Medida Fuzzy em cada classe computada nos testes;
- Testagem desse novo FRM em diferentes datasets e trazer uma análise de desempenho na utilização da integral de Choquet como função de agregação, 3ª etapa da FRM.

Preliminares



- Fuzzy Sets -> As variáveis linguísticas que compõem os antecedentes das regras;
- Operador T-norma Triangular-> É o operador que aplicará a operação de conjunção nos antecedentes das regras;
- Funções de Agregação -> Necessária para a combinação de vários valores em um único
- Medidas Fuzzy -> A partir dela é possível modelar a relação entre os elementos a serem agregados;
- Integral Discreta de Choquet -> A fórmula geral desenvolvida por Gustave Choquet, função que será desenvolvida no artigo para atuar em problemas de classificação;
- Fuzzy Rule-Based Classification Systems(FRBCSs) -> Modelo de aprendizagem de máquina que utiliza de regras de associação fuzzy, utilizado no artigo;
- Fuzzy Reasoning Method (FRM) -> Técnica utilizada nos FRBCS para conseguir trabalhar em cima de dados fuzzy e conseguir estipular, nesse caso de classificação, uma classe C para um novo dado;

Preliminares: Algoritmo de Chi

- É uma extensão do algoritmo proposto por Wang e Mendel;
- Muito utilizado devido a sua simplicidade na formulação de regras fuzzy;
- Utiliza de um método para determinar uma relação entre atributos e classes do problema, faz isso seguindo os seguintes passos:
 - Estabelecimento de partições linguísticas
 - Geração de uma regra fuzzy para um exemplo $x_p = (x_{p1}, \dots, x_{pn}, C_p)$, seguindo os passos:
 - Computa o grau de pertinencia $\mu(x_p)$ de um exemplo com todas as regiões fuzzy com o operador de conjunção (mínimo ou Produto t-norma)
 - Assinala o exemplo x_p com a região de maior grau correspondente
 - Gera as regras para o exemplo, com o antecedente sendo a região fuzzy designada e o consequente a classe designada para aquele exemplo
 - Utiliza de uma regra de peso para avaliação das regras:

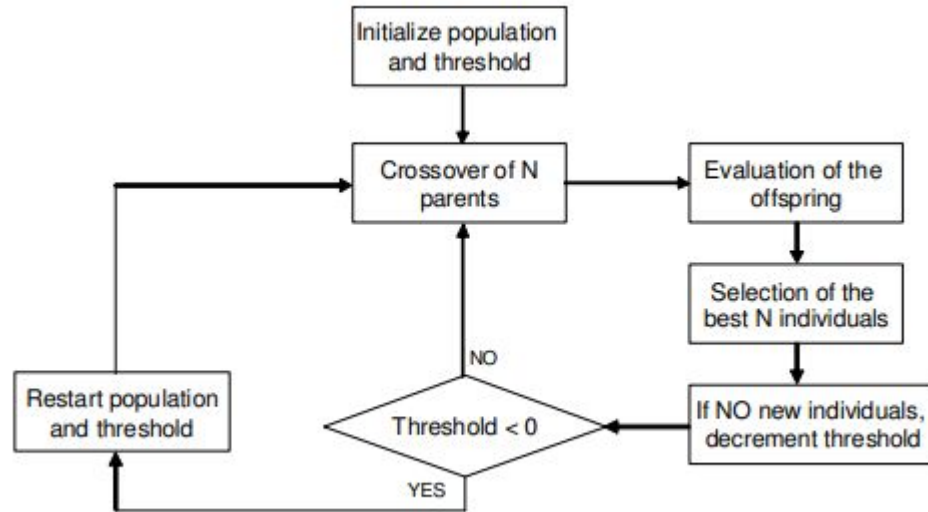
$$RW_j = PCF_j = \frac{\sum_{x_p \in Class C_j} \mu_{A_j}(x_p) - \sum_{x_p \notin Class C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^m \mu_{A_j}(x_p)}$$

Preliminares: Modelo Evolucionário



- Nos experimentos realizados, foi utilizado o modelo evolucionário CHC para realizar a etapa de aprendizado das medidas fuzzy;
- É um modelo com uma abordagem de *Population-Based Selection*, ou seja, é formado por N parentes e seus descendentes correspondentes são combinados para encontrar os melhores N indivíduos para formar a próxima população;
- Como mecanismo de prevenção de incesto, foi definido no modelo apresentado que dois parentes só serão cruzados se metade de suas distâncias (medida pela distância Hamming) é acima de um limite predeterminado $Th(threashold)$;
- Para esse limite foi criado um esquema de codificação, em que cada gene é considerado um *Gray Code* (codificação binária) com um número fixo de bits por gene ($BITSGENE$), estipulado por um expert. O valor inicial do Th é ditado pela fórmula: $Th = (\#Genes * BITSGENE) / 4.0$, sendo $\#Genes$ o tamanho total de um cromossomo;
- O Th é decrementado por um $BITSGENE$ quando não há novos indivíduos na próxima geração;
- O algoritmo reinicia quando $Th < 0$;
- De maneira visual esse modelo evolucionário é construído da seguinte forma:

Preliminares: Modelo Evolucionário



Desenvolvimento da utilização da Integral

- A integral é alterada de maneira a utilizar as regras que estão no RB do modelo para prover as informações locais a serem agregadas ficando da seguinte forma:

$$Y_k = C_{m_k}(b_j^k | j = 1, \dots, L \text{ and } b_j^k > 0), \quad k = 1, \dots, M$$

fuzzy considerada para a k-th classe, M o número de classes e L o número de regras compondo o RB.

- Na FRM proposta no artigo, primeiramente foi testado a utilização da integral de Choquet com a mesma medida fuzzy para cada classe do problema. Isso não foi tão eficaz pois cada conjunto de regras interagem de maneiras diferentes. Isso é observado na medida fuzzy de cardinalidade exponencial;
- Então a partir desse problema o CHC evolucionário foi usado para encontrar a melhor medida fuzzy a ser utilizada em cada classe;
- Os passos desse modelo são:
 - Esquema de Codificação
 - Gama de Genes Iniciais
 - Validação de Cromossomos
 - Operador de Cruzamento
 - Abordagem de Reinício

Descrição do Modelo Evolucionário

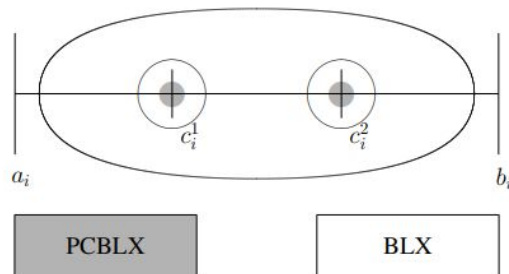
- Esquema de Codificação -> A partir de um set de parâmetros reais a ser otimizado (q_k , $k = 1, \dots, M$), com um intervalo de variação sugerido de $[0.01, 100]$. A codificação não é feita diretamente em um cromossomo, mas sim é adaptada usando os cromossomos na forma de: $C_{\text{Choquet}} = \{G_1, \dots, G_M\}$, em que $G_k \in [0.01, 1.99]$, com $k = 1, \dots, M$. Para encontrar os valores reais ($[1, 100]$) a equação a seguir é aplicada:

$$q_k = \begin{cases} G_k, & \text{if } 0 < G_k \leq 1 \\ \frac{1}{2-G_k}, & \text{if } 1 < G_k < 2 \end{cases}$$

- Gama de Genes Iniciais -> É incluso um indivíduo que tiver todos os genes no valor 1. Desse jeito é possível obter os resultados providos pela medida de cardinalidade
- Validação de Cromossomos → Utiliza da taxa de acurácia para a classificação;

Descrição do Modelo Evolucionário

- Operador de Cruzamento → Utiliza dos conceitos de ambientes (a prole é gerada ao redor de seus pais). Utiliza o operador PCBLX (Parent Centric BLX), com a variação de BLX- α (prole ao redor dos dois pais).
 - PCBLX -> Assumindo que $X(x_1, \dots, x_n)$ e $Y = (y_1, \dots, y_n)$ e $(x_i, y_i \in [a_i, b_i] \subset R, i = 1 \dots n)$ como dois cromossomos reais-codificados que irão ser cruzados. o PCBLX tem dois resultados:
 - $O_1 = (o_{11}, \dots, o_{1n})$, em que o_{1i} é um número escolhido aleatoriamente (uniformemente) do intervalo $[l_i^1, u_i^1]$, com $l_i^1 = \max\{a_i, y_i - l_i\}$ e $u_i^1 = \min\{b_i, x_i + l_i\}$ e $l_i = |x_i - y_i|$
 - $O_2 = (o_{21}, \dots, o_{2n})$, em que o_{2i} é um número escolhido aleatoriamente (uniformemente) do intervalo $[l_i^2, u_i^2]$, com $l_i^2 = \max\{a_i, y_i - l_i\}$ e $u_i^2 = \min\{b_i, x_i + l_i\}$ e $l_i = |x_i - y_i|$
 - Representação visual desse operador PCBLX em comparação com o BLX:



Descrição do Modelo Evolucionário



- Abordagem de Reinício -> esse algoritmo usa de uma abordagem de reinício já que ele não aplica a mutação durante a fase de recombinação. Entretanto, quando o $\text{threshold} < 0$, todos os cromossomos são re-gerados randomicamente para introduzir uma nova diversidade na busca.
 - A melhor solução encontrada é incluir na população para melhorar a convergência do algoritmo.

Testagem do Modelo

- Foram selecionados 17 datasets, retirados do repositório KEEL data-set.
- Algumas etapas de preparo foram realizadas: retirando dados que tinha valores faltando e em alguns datasets foi necessário uma redução de tamanho para não ficarem extremamente grandes.
- O método de validação do modelo utilizado foi o 5-fold Cross-Validation, sendo 20% para teste e 80% para treino.
- Datasets utilizados:

Id.	Data-set	#Ex.	#Atts.	#Class.
bal	Balance	625	4	3
ban	Banana	5300	2	2
eco	Ecoli	336	7	8
gla	Glass	214	9	6
iri	Iris	150	4	3
led	Led7digit	500	7	10
mag	Magic	1902	10	2
new	Newthyroid	215	5	3
pho	Phoneme	5404	5	2
pim	Pima	768	8	2
rin	Ring	740	20	2
seg	Segment	2310	19	7
tit	Titanic	2201	3	2
two	Twonorm	740	20	2
veh	Vehicle	846	18	4
win	Wine	178	13	3
wis	Wisconsin	683	11	2

Configurações do modelo na testagem



Para a utilização do algoritmo de Chi *et al.* foram consideradas as seguintes configurações::

- Operador de conjunção → Produto t-norma
- Regra de Peso → Penalized Certainty Factor
- Numero de labels linguísticos → 3

Para o processo genético foram utilizados os valores:

- Tamanho de população: 50 individuos
- Numero de avaliações: 20000
- Bits por gene para a codificação Gray (prevenção de incesto): 30 bits

Na proposta de utilização da medida fuzzy de Dirac, o valor selecionado como i é associado com a média, ou seja, se o número de elementos é impar $i = (n+1)/2$, se o número de elementos é par $i = n/2 + 1$

Resultados Obtidos

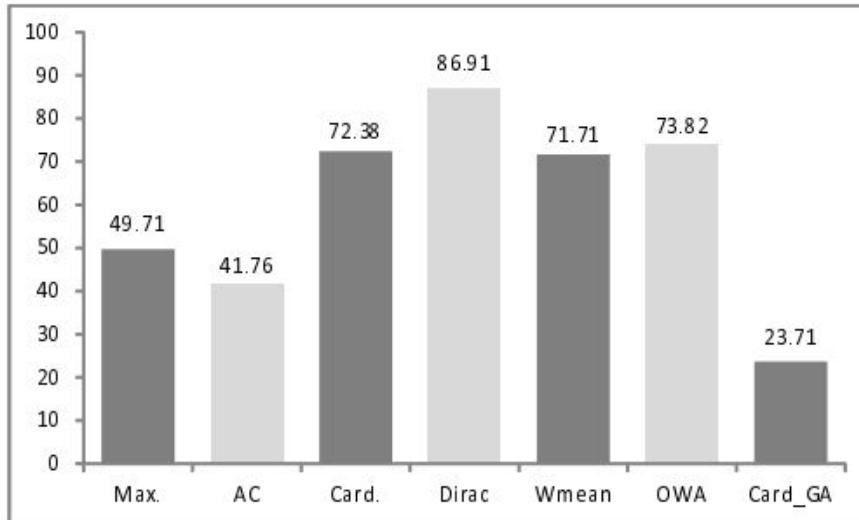
Data Set	Max.		AC		Card.		Dirac		WMean		OWA		Card_GA	
	Tr.	Tst	Tr.	Tst	Tr.	Tst	Tr.	Tst	Tr.	Tst	Tr.	Tst	Tr.	Tst
Bal	91.52 ± 0.23	90.56 ± 1.04	91.52 ± 0.23	90.56 ± 1.19	89.96 ± 1.11	86.72 ± 2.57	89.48 ± 0.41	87.04 ± 1.04	90.20 ± 0.93	86.72 ± 1.93	90.24 ± 1.07	86.88 ± 2.37	91.52 ± 0.23	89.92 ± 0.91
Ban	60.36 ± 0.39	60.32 ± 1.33	59.83 ± 0.28	60.02 ± 1.30	60.68 ± 0.60	60.47 ± 1.62	62.00 ± 1.18	61.77 ± 0.79	60.34 ± 0.63	60.08 ± 1.44	60.55 ± 0.71	60.36 ± 1.34	63.11 ± 0.90	62.36 ± 1.28
Eco	76.27 ± 1.34	71.76 ± 6.69	78.42 ± 0.86	73.53 ± 5.40	75.97 ± 2.03	69.71 ± 5.94	73.36 ± 2.57	67.94 ± 6.01	75.60 ± 1.45	70.59 ± 6.33	75.45 ± 1.73	70.88 ± 6.36	83.26 ± 2.12	75.00 ± 9.24
Glu	66.01 ± 2.58	57.67 ± 1.04	66.47 ± 1.84	59.07 ± 2.65	62.61 ± 3.39	58.60 ± 3.82	62.15 ± 4.82	57.21 ± 2.08	61.09 ± 3.97	57.21 ± 4.53	63.31 ± 2.86	59.53 ± 2.65	68.23 ± 1.36	59.53 ± 3.12
Iri	93.00 ± 0.75	92.67 ± 1.49	95.33 ± 1.51	94.67 ± 4.47	88.50 ± 1.09	87.33 ± 1.49	84.67 ± 1.73	84.00 ± 4.35	88.67 ± 1.92	87.33 ± 2.79	87.00 ± 1.26	86.67 ± 2.36	96.67 ± 0.83	94.67 ± 2.98
Led	75.90 ± 2.96	64.20 ± 5.63	75.90 ± 2.96	64.20 ± 5.63	75.90 ± 2.96	64.20 ± 5.63	75.90 ± 2.96	64.20 ± 5.63	75.90 ± 2.96	64.20 ± 5.63	75.90 ± 2.96	64.20 ± 5.63	75.90 ± 2.96	64.20 ± 5.63
Mag	76.00 ± 0.61	74.75 ± 1.85	76.49 ± 0.56	75.38 ± 1.51	74.11 ± 0.22	72.65 ± 1.50	71.65 ± 0.43	70.81 ± 1.04	74.01 ± 0.31	72.97 ± 1.65	74.29 ± 0.32	73.07 ± 1.40	79.31 ± 0.93	77.80 ± 3.62
New	86.40 ± 1.06	85.12 ± 3.53	87.44 ± 1.40	86.51 ± 4.16	87.79 ± 1.23	86.05 ± 3.68	89.19 ± 0.88	86.98 ± 4.22	87.91 ± 1.61	87.44 ± 3.53	88.02 ± 1.52	86.51 ± 3.45	94.42 ± 1.27	92.56 ± 4.47
Pho	71.91 ± 0.11	71.91 ± 0.37	72.82 ± 0.09	72.62 ± 0.64	71.54 ± 0.18	71.23 ± 0.80	72.20 ± 0.36	72.16 ± 0.71	71.79 ± 0.27	71.56 ± 0.60	72.03 ± 0.26	71.93 ± 1.01	76.16 ± 0.25	75.39 ± 1.28
Pim	75.46 ± 0.70	72.99 ± 0.98	74.64 ± 0.50	73.25 ± 1.55	75.75 ± 0.40	73.77 ± 1.75	75.36 ± 0.49	73.38 ± 1.66	75.98 ± 0.48	74.55 ± 2.53	75.46 ± 0.28	74.03 ± 2.43	78.39 ± 0.71	75.06 ± 1.18
Rin	59.39 ± 0.44	52.70 ± 0.83	57.70 ± 0.44	52.03 ± 0.48	53.75 ± 0.44	51.08 ± 0.37	51.11 ± 0.28	50.68 ± 0.48	53.99 ± 0.64	51.08 ± 0.37	53.72 ± 0.41	51.22 ± 0.30	81.35 ± 1.69	77.70 ± 1.85
Seg	86.01 ± 1.31	85.02 ± 2.26	86.03 ± 0.95	84.81 ± 1.84	84.40 ± 0.92	83.51 ± 1.93	78.40 ± 1.93	77.92 ± 3.93	84.43 ± 1.22	83.38 ± 2.16	83.99 ± 0.91	82.94 ± 2.62	87.18 ± 0.74	85.06 ± 2.23
Tit	78.33 ± 0.41	78.32 ± 1.71	78.33 ± 0.41	78.32 ± 1.71	78.33 ± 0.41	78.32 ± 1.71	78.33 ± 0.41	78.32 ± 1.71	78.33 ± 0.41	78.32 ± 1.71	78.33 ± 0.41	78.32 ± 1.71	78.33 ± 0.41	78.32 ± 1.71
Two	87.13 ± 0.77	83.78 ± 1.72	94.97 ± 0.51	93.24 ± 1.85	73.34 ± 1.23	70.41 ± 4.39	63.68 ± 1.44	62.57 ± 4.91	73.01 ± 1.33	69.19 ± 4.29	74.02 ± 2.23	70.68 ± 3.59	91.45 ± 0.44	87.57 ± 1.83
Veh	66.11 ± 0.80	61.41 ± 3.66	64.16 ± 0.70	61.29 ± 3.26	64.21 ± 0.56	60.94 ± 4.29	57.54 ± 1.25	53.41 ± 3.29	64.18 ± 0.85	60.12 ± 3.56	63.95 ± 0.58	59.65 ± 3.26	67.85 ± 0.48	60.59 ± 3.53
Win	98.74 ± 0.58	92.78 ± 5.41	98.74 ± 0.59	93.33 ± 5.76	96.21 ± 0.62	91.11 ± 5.69	89.61 ± 1.32	85.56 ± 4.12	95.93 ± 0.76	91.11 ± 5.69	96.21 ± 1.07	90.00 ± 4.65	99.86 ± 0.31	92.22 ± 4.56
Wis	98.17 ± 0.29	95.62 ± 1.37	97.99 ± 0.45	95.77 ± 1.31	98.21 ± 0.27	96.06 ± 1.42	98.13 ± 0.24	95.91 ± 1.60	98.24 ± 0.28	95.77 ± 1.66	98.21 ± 0.27	96.06 ± 1.42	98.54 ± 0.13	95.33 ± 1.42
Mean	79.22 ± 0.90	75.98 ± 2.41	79.81 ± 0.84	76.98 ± 2.63	77.13 ± 1.04	74.24 ± 2.86	74.87 ± 1.33	72.34 ± 2.80	77.03 ± 1.18	74.21 ± 2.96	77.10 ± 1.11	74.29 ± 2.74	83.03 ± 0.93	79.02 ± 2.99

Resultados Obtidos



- Foram utilizados algumas técnicas de validação de hipóteses como suporte estatístico a análise dos dados. Utilizando de métodos como:
- Teste de postos alinhados de Friedman: para detectar diferenças estatísticas entre um grupo de resultados;
- Teste post-hoc de Holm para encontrar os algoritmos que rejeitam a hipótese de igualdade em relação a um método de controle selecionado, permitindo identificar se uma hipótese de comparação pode ser rejeitada em um nível de significância especificado α ;
- Cálculo do valor de p ajustado (APV) para levar em conta o fato de que vários testes são realizados, permitindo a comparação direta do APV com o nível de significância α para poder rejeitar a hipótese nula.
- O método de postos alinhados dos algoritmos para mostrar graficamente quão bom um método é em relação aos seus parceiros.
 - O primeiro passo para calcular esse ranking é obter o desempenho médio dos algoritmos em cada conjunto de dados
 - Após calcula-se as subtrações entre a precisão de cada algoritmo menos o valor médio para cada conjunto de dados e os classifica em ordem decrescente;
 - Por fim, é feita a média dos rankings obtidos por cada algoritmo, logo, o algoritmo que tiver a menor média é o melhor

Resultados Obtidos pelo suporte estatístico



<i>i</i>	Algorithm	APV
1	Dirac	5.52E−7
2	OWA	1.14E−4
3	Card.	1.56E−4
4	WMean	1.56E−4
5	Max.	0.06
6	AC	0.13