Agregação de elenco de classificadores utilizando integral de Choquet com respeito a medida λ-*fuzzy*

André G. C. Pacheco
UFES - Universidade Federal do Espírito Santo
PPGI - Programa de Pós-Graduação em Informática
Av. Fernando Ferrari 514 - CEP 29060-270, Vitória, ES, Brasil
pacheco.comp@gmail.com

Renato A. Krohling
UFES - Universidade Federal do Espírito Santo
PPGI - Programa de Pós-Graduação em Informática
Av. Fernando Ferrari 514 - CEP 29060-270, Vitória, ES, Brasil
krohling.renato@gmail.com

RESUMO

Classificação de dados está presente em diversos problemas reais, tais como: reconhecer padrões em imagens, diferenciar espécies de plantas, classificar tumores benignos e malignos, dentre outros. Muitos desses problemas possuem padrões de dados difíceis de serem identificados, o que requer, consequentemente, técnicas mais avançadas para sua resolução. Por ser uma metodologia simples e eficaz o uso de um elenco de classificadores tem sido bastante explorado para atacar estes problemas. Todavia, para que o elenco seja capaz de promover melhorias, uma boa metodologia de agregação deve ser aplicada. Neste trabalho é apresentado uma abordagem de agregação de classificadores baseada na integral de Choquet com respeito a uma medida *fuzzy*. A metodologia é aplicada à diversas bases de dados e seus resultados se mostraram promissores.

PALAVRAS CHAVE. Classificação de dados, Integral de Choquet, Elenco de classificadores.

Tópicos: Estatística

ABSTRACT

Data classification appear in many real-world problems, e.g., recognize image patterns, differentiate plants species, classify benign and malignant tumors, among others. Many of these problems present data patterns, which are difficult to be identified, thus requiring more advanced techniques. As a simple and effective approach, the use of an ensemble of classifiers has been extensively explored to tackle these problems. However, in order to improve the performance of the ensemble, an adequate aggregation methodology must be developed. In this paper we present a classifier aggregation approach based on Choquet integral regarding to a fuzzy measure. The methodology is applied to different data bases and the results are promising.

KEYWORDS. Data classification, Choquet integral, Ensemble classifiers

Paper topics: Statistics

1. Introdução

Em aprendizado de máquina, sistemas de aprendizagem por elencos são inspirados no comportamento humano em tomadas de decisões. É inerente ao ser humano buscar mais de uma fonte de opinião especialista ao se deparar com uma decisão importante a ser tomada, principalmente nas áreas: médica, financeira e social [Štefka e Holeňa, 2010]. Um exemplo clássico é a coleta de mais de uma opinião médica, neste caso o especialista, relacionado a uma possível intervenção cirúrgica. O indivíduo ouve cada uma das fontes, considera a opinião de cada uma delas e toma a decisão final: realizar ou não o processo cirúrgico. Esse é um processo bastante comum para um ser humano e vem sendo aplicado em diversos problemas que envolvem inteligência computacional [Polikar, 2006].

Informações obtidas por fontes diferentes possuem características diferentes. Partindo deste princípio, a fusão de informações é uma tarefa que consiste em produzir, a partir de um conjunto de informações relacionadas, uma única informação que é capaz de extrair as principais características das demais. Atualmente, abordagens que utilizam fusão de informação vêm sendo utilizadas largamente em problemas como: classificação de dados [Štefka e Holeňa, 2010; Mousavi e Eftekhari, 2015; Seokho et al., 2015; Krawczyk e Woźniak, 2016], tomada de decisão [Polikar, 2006; Tapia-Rosero et al., 2016], predição de séries temporais [Merigó et al., 2015; Kourentzes e Petropoulos, 2015], dentre outros.

A agregação de classificadores tem como objetivo obter uma classificação mais confiável por meio de um conjunto de diferentes algoritmos de classificação atuando sobre a mesma base de dados. Ao invés de buscar o melhor classificador, busca-se um conjunto de classificadores combinados por um algoritmo de agregação. De maneira geral, um conjunto de classificadores simples pode ser melhor do que um único classificador complexo [Chen et al., 2010]. Baseado nisso, diversos algoritmos de agregação vem sendo aplicados em conjuntos de classificadores a fim de se obter melhores resultados. Xu et al. [1992] combinaram a teoria de Dempster-Shafer, formalismo Bayesiano e o princípio de votação para reconhecimento de escrita. Krawczyk e Woźniak [2016] alcançaram agregação por meio de um procedimento para determinar pesos de uma média ponderada. Štefka e Holeňa [2015] utilizaram a integral *fuzzy* para realizar agregação. Cao [2012] e He et al. [2007] obtiveram a agregação por meio da integral de Choquet.

Uma das técnicas utilizadas para agregação de classificadores é a integral de Choquet [Choquet, 1954; Höhle, 1982], que é capaz de realizar a fusão de um conjunto de classificadores levando em consideração a saída individual de cada classificador com respeito a uma medida *fuzzy*. A medida *fuzzy* foi proposta por Sugeno [1974] e é uma forma natural para realizar avaliações envoltas de incerteza e subjetividade. A grande dificuldade para utilizar a integral de Choquet é justamente determinar quais são as medidas *fuzzy* relacionada a cada classificador. Devido a isso, com intuito de simplificar a teoria da medida *fuzzy*, Sugeno [1974] propôs a medida λ-*fuzzy*, na qual deve-se determinar as densidades *fuzzy* referentes a medida. Essa simplificação vem sendo utilizada em diversas aplicações [He et al., 2007; Chen e Wang, 2001; Krishnan et al., 2015].

Mesmo com o modelo simplificado da medida *fuzzy*, ainda é necessário determinar as densidades *fuzzy* para a medida λ-*fuzzy*. Essas densidades podem ser interpretadas como a importância de um determinado elemento dentro de um conjunto. Diversas abordagens foram desenvolvidas para determinar as densidades *fuzzy*, utilizando por exemplo: Redes Neurais Artificiais (ANN) [Jia e Zhenyuan, 1997] ou Algoritmo Genético (GA) [Chen e Wang, 2001; Lee e Leekwang, 1995]. Neste trabalho é proposto uma abordagem utilizando Análise de Componentes Principais (PCA) para determinar as densidades *fuzzy*, e consequentemente a medida λ-*fuzzy*, com objetivo de agregar um elenco de classificadores, atuando em determinadas bases de dados, por meio da integral de Choquet. O restante deste artigo está organizado da seguinte maneira: na seção 2 os conceitos básicos referentes a integral de Choquet, medida *fuzzy* e análise de componentes principais são discutidos. Na seção 3 são apresentados conceitos relacionados ao elenco de classificadores e sua agregação. Na seção 4 são apresentados os resultados experimentais. Por fim, na seção 5 é realizada

uma breve conclusão.

2. Conceitos básicos

2.1. Medida *Fuzzy*

A medida *fuzzy* [Sugeno, 1974], nada mais é do que uma generalização da medida aditiva, no qual a condição de aditividade é substituída por uma condição mais fraca, a monotonicidade [Pedrycz e Gomide, 1998]. Nesta subseção serão abordadas as definições básicas relacionadas à medida *fuzzy*.

> **Definição 1:** [Sugeno, 1974] Dado um conjunto finito $X = (x_1, x_2, ..., x_r)$ e seja $F(X)$ uma função *power set* de X, uma função $\mu : F(X) \rightarrow [0, 1]$ é uma medida *fuzzy* se ela satisfaz as seguintes condições:
>
> 1. $\mu(\varnothing) = 0$ e $\mu(X) = 1$ (condição de fronteira).
> 2. $\mu(A) \leq \mu(B)$ se $A \subseteq B, \forall A, B \in X$ (condição de monotonicidade).

A definição de medida *fuzzy* nem sempre é a mesma na literatura. Por exemplo, é comum exigir apenas que tal medida seja monótona e positiva. No entanto, entre as diversas definições, o que há em comum é a exigência de monotonicidade e que a medida do conjunto vazio seja zero [de Barros e Bassanezi, 2006].

> **Definição 2**: [Sugeno, 1974] Dado um conjunto finito $X = (x_1, x_2, ..., x_r)$ e seja $F(X)$ uma função *power set* de X, uma função $g : F(X) \rightarrow [0, 1]$ é uma medida λ-*fuzzy* se ela satisfaz as seguintes condições:
>
> 1. $g(\varnothing) = 0$ e $g(X) = 1$.
> 2. Se $A, B \in F(X)$ então $g(A \cup B) = g(A) + g(B) + \lambda g(A)g(B)$, para algum $\lambda > -1$, se $A \cap B = \varnothing$.

O valor de λ pode ser encontrado a partir de $g(X) = 1$, que é equivalente a solucionar a equação 1, no qual $g_i = g(x_i)$, $x = (1, 2, ..., r)$, conhecido como densidade *fuzzy*, que pode ser interpretado como a importância do subconjunto x_i.

$$1 + \lambda = \prod_{i=1}^{n} (\lambda g_i + 1) \tag{1}$$

A partir da propriedade 2 da definição 2, concluí-se:

- Se $\lambda = 0$, então a medida é aditiva.
- Se $\lambda \neq 0$, então a medida não é aditiva.

Se a medida *fuzzy* não for aditiva isso expressa uma interação entre os diferentes conjuntos pertencentes ao universo de discurso. Caso contrário, então isso denota a independência entre esses conjuntos [He et al., 2007].

2.2. A integral de Choquet

A integral de Choquet foi introduzida, inicialmente no contexto da teoria da capacidade por Choquet [1954], foi utilizada como integral com respeito a uma medida *fuzzy* por Höhle [1982] e redescoberta mais tarde por Murofushi e Sugeno [1989] e Murofushi e Sugeno [1991]. De maneira geral, a integral de Choquet é um operador de agregação com respeito a qualquer medida *fuzzy* [Grabisch, 1995] e a grande utilização dessa técnica se deve, principalmente, ao fato de seu modelo considerar não só a importância de cada atributo individual a ser agregado, mas também a interação entre eles [Cao, 2012]. A definição da integral de Choquet é descrita da seguinte forma:

> **Definição 3:** [Murofushi e Sugeno, 1989] Seja μ uma medida *fuzzy* em um conjunto finito X, então a integral de Choquet da função $f : X \to [0, \infty]$, com respeito a medida *fuzzy* μ é definida como:
>
> $$\int f d\mu = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\mu(A_i) \tag{2}$$
>
> onde $A_i \subset X$ para $i = 1, 2, ..., n$, $f(x_1) \leq f(x_2) \leq f(x_3) \leq ... \leq f(x_n)$ e $f(x_0) = 0$.

2.3. Análise de Componentes Principais (PCA)

A análise de componentes principais (PCA) é um método de análise de dados multivariados que pode ser usada para interpretar bases de dados multivariados a fim de reduzir sua dimensionalidade. Análise de dados multivariados é utilizada em diversas áreas de pesquisa como alternativa para tratar grandes quantidade de dados [Jolliffe, 2002]. Matematicamente, PCA é definido como uma transformação linear ortogonal que transforma um conjunto de variáveis correlatas em um outro conjunto menor de variáveis não correlacionadas denominadas componentes principais. O processo de PCA é descrito nos passos a seguir [Jolliffe, 2002]:

> *Passo 1:* Dado um conjunto de dados, $X = (x_1, x_2, ..., x_n)$, com $x_i = (x_i^1, x_i^2, ..., x_i^d)$, determinar a matriz de variância-covariância de X, denominada P.
>
> *Passo 2:* Calcular os autovalores $\lambda = (\lambda_1, \lambda_2, ..., \lambda_d)$ e os autovetores $v = (v_1, v_2, ..., v_d)$ relacionados a P.
>
> *Passo 3:* Calcular as componentes principais por meio de uma combinação linear entre os autovetores e os dados de entrada da seguinte forma: $CP_i = v' x_i$.

As componentes principais apresentarão as seguintes propriedades [Jolliffe, 2002]:

1. A variância da *i-ésima* componente principal CP_i, é igual ao valor do autovalor λ_i relacionado a ela.
2. A componente principal com maior autovalor representará a componente com maior variância.
3. O total de variância das variáveis originais é igual ao somatório dos autovalores, que por sua vez é igual ao total de variância das componentes principais.
4. As componentes principais não têm correlação entre si.

Considerando as propriedades acima, calcula-se a contribuição C_i de cada componente de acordo:

$$C_i = \frac{\lambda_i}{\sum_{i=1}^{d} \lambda_i} \times 100 \tag{3}$$

Para realizar a redução de dimensionalidade, escolhe-se as componentes principais com as maiores contribuições. Todavia, neste trabalho a técnica de PCA não é utilizada com este propósito.

3. Agregação classificadores

3.1. Elenco de classificadores

Um elenco de classificadores simples pode ser melhor do que um único classificador complexo, pois um classificador sozinho, quase sempre, não é capaz de identificar todas as correlações entre os atributos de entrada e de saída da base de dados [Chen et al., 2010]. Como diversos algoritmos funcionam realizando buscas que podem levar a um ótimo local, e consequentemente prejudicar o resultado final, computacionalmente e estatisticamente a agregação de um conjunto de classificadores reduz o risco de escolher uma hipótese ruim. Todavia, nada impede de que um único classificador seja melhor do que todo o elenco em determinada base de dados [Polikar, 2006].

Um classificador pode ser definido como do tipo abstrato ou do tipo medida de nível. Classificadores abstratos já retornam, como resultado da classificação, o rótulo de saída, ou seja, a classe que determinada amostra pertence. Por outro lado, classificadores de medida de nível retornam como resposta um vetor, de dimensão igual ao número de classes do problema, de medidas no intervalo [0,1], no qual cada uma dessas medidas representam a aceitação da hipótese de que o dado de entrada pertença a uma determinada classe. Além disso, elencos de classificadores podem ser do tipo homogêneo ou híbrido. Elencos homogêneos são aqueles compostos pelos mesmos algoritmos de classificação, alterando alguns parâmetros dos mesmos. Já os elencos híbridos, são formados por diferentes tipos de classificadores. Neste trabalho são utilizados classificadores do tipo de medida de nível e elencos híbridos.

Métodos baseados em agregação de classificadores se tornaram ferramentas padrões para melhorar a qualidade da classificação de dados [Štefka e Holeňa, 2015]. Em vez de utilizar apenas um classificador, um elenco de classificadores é criado e cada avaliação individual é considerada para formar a classificação final. Na Figura 1 é exibido um elenco de classificadores no qual o operador de agregação será a integral de Choquet.

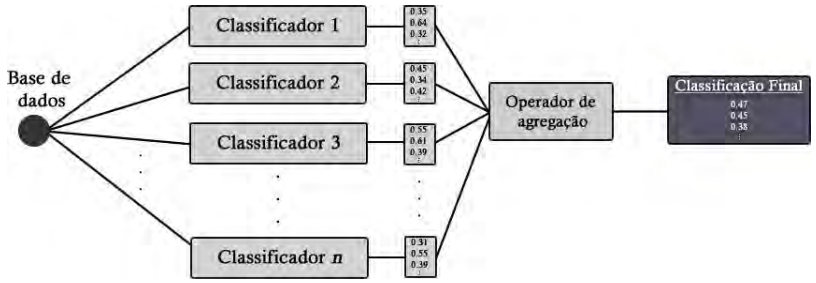


Figura 1: Elenco de classificadores

3.2. Metodologia de agregação

Dado um elenco de classificadores, como mostrado na Figura 1, o primeiro passo para realizar a agregação dos mesmos é organizar a saída de cada classificador na matriz de decisão M_n.

$$M_k = \begin{bmatrix} c_{11}^k & c_{12}^k & \cdots & c_{1j}^k \\ c_{21}^k & c_{22}^k & \cdots & c_{2j}^k \\ \vdots & \vdots & \ddots & \vdots \\ c_{i1}^k & c_{i2}^k & \cdots & c_{ij}^k \end{bmatrix} \tag{4}$$

onde cada coluna representa um classificador do elenco ($j = 1, 2, ..., J$), sendo J o nº de classificadores, cada linha representa um rótulo da base de dados ($i = 1, 2, ..., I$), sendo I o nº de rótulos, k é uma determinada amostra da base de dados e c_{ij}^k é a hipótese de classificação, também conhecida como função suporte referente à amostra k da base dados, ao rótulo i e ao classificador j. Cada

amostra da base de dados possuirá sua própria matriz de decisão M_k. Essa matriz considera cada classificação individual do elenco e será nela que o operador de agregação atuará.

Montada a matriz de decisão o próximo passo é realizar a agregação das funções suporte de classificação para cada rótulo a fim de obter o resultado final em um vetor coluna denominado C_f, como mostrado na Figura 2.

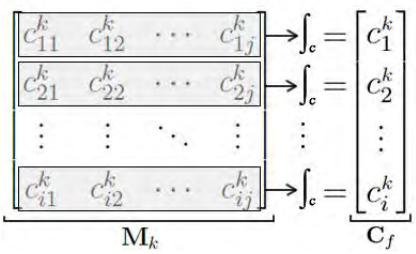


Figura 2: Agregando a matriz de decisão dos classificadores

O rótulo da classificação final do elenco será aquele que possui a maior valor em C_f. Portanto, calcula-se o *arg max*(C_f).

3.3. Determinação da medida *fuzzy*

Ao utilizar a integral de Choquet para realizar a agregação das hipótese de classificação, é necessário determinar as medidas *fuzzy* dos conjuntos de classificadores. A medida escolhida neste trabalho é a medida λ-*fuzzy*. Sendo assim, o desafio aqui é determinar as densidades *fuzzy*, referentes a cada classificador do elenco, para então determinar as medidas λ-*fuzzy* entre os conjuntos.

Ao longo dos anos, diversos trabalhos surgiram com objetivo de determinar os valores das densidades *fuzzy* [Leszczynski et al., 1985; Takahagi, 2007; Larbani et al., 2011], porém esses métodos necessitam de informações subjetivas fornecidas por algum especialista. Outros métodos que excluem a subjetividade de um especialista, utilizando RNA e GA, foram propostos [Lee e Leekwang, 1995; Jia e Zhenyuan, 1997; Chen e Wang, 2001]. Todavia, esses métodos são supervisionados, sendo necessário conhecimento prévio do resultado, além de sofrer com problemas de convergência lenta [Krishnan et al., 2015]. A proposta deste trabalho é utilizar PCA para determinar as densidades *fuzzy* de maneira não-supervisionada.

Zhang e Sun [2009] propuseram o uso do método PCA para ponderar uma base de dados para treinamento de uma rede neural. A ideia básica deste método é de que quanto mais útil uma componente principal, mais importante deve ser o papel dela. Para isso, os dados são ponderados de acordo com o vetor de autovalores λ. Os autovalores, como mostrado na seção 2.3, contêm a informação de variância da base de dados e quanto maior a variância, segundo PCA, mais informação a componente é capaz de prover. Sendo assim, basedo no trabalho de Zhang e Sun [2009], é aplicado PCA nos resultados obtidos por cada classificador. O procedimento é descrito a seguir.

Passo 1: Obter a matriz de resultados R_j da base de dados para cada classificador do elenco.

Passo 2: Aplicar PCA para cada matriz R_j.

Passo 3: Calcular a variância total fazendo $\alpha_j = \sum_{k=1}^{j} \lambda_k$.

Passo 4: Normalizar α entre $[0, 1]$.

Na técnica PCA os dados mais importantes são aqueles que concentram as maiores variâncias. Sendo assim, os classificadores mais importantes do elenco são aqueles que possuem maior valor α. Portanto, as densidades *fuzzy* serão iguais a α.

4. Resultados experimentais

Nesta seção são apresentados os resultados experimentais para a agregação de um elenco composto pelos seguintes classificadores:

- Uma rede neural *feedforward* (FF) [Haykin, 2007] com duas camadas escondidas contendo 20 neurônios cada, com algoritmo de treinamento Levenberg-Marquardt utilizando *mini-batch* de 10 partições, para bases acima de 1000 amostras, 1000 iterações e taxa de aprendizagem igual a 0.01.
- Uma rede neural usando *Extreme Learning Machine* (ELM) [Huang et al., 2006] com 30 neurônios na camada escondida para bases com menos de 1000 amostras e com 300 neurônios para bases com mais.
- Um K-*Nearest Neighbor* (KNN) [Keller et al., 1985] com k escolhido como $\sqrt{T}$, sendo T o tamanho da base de dados utilizada.

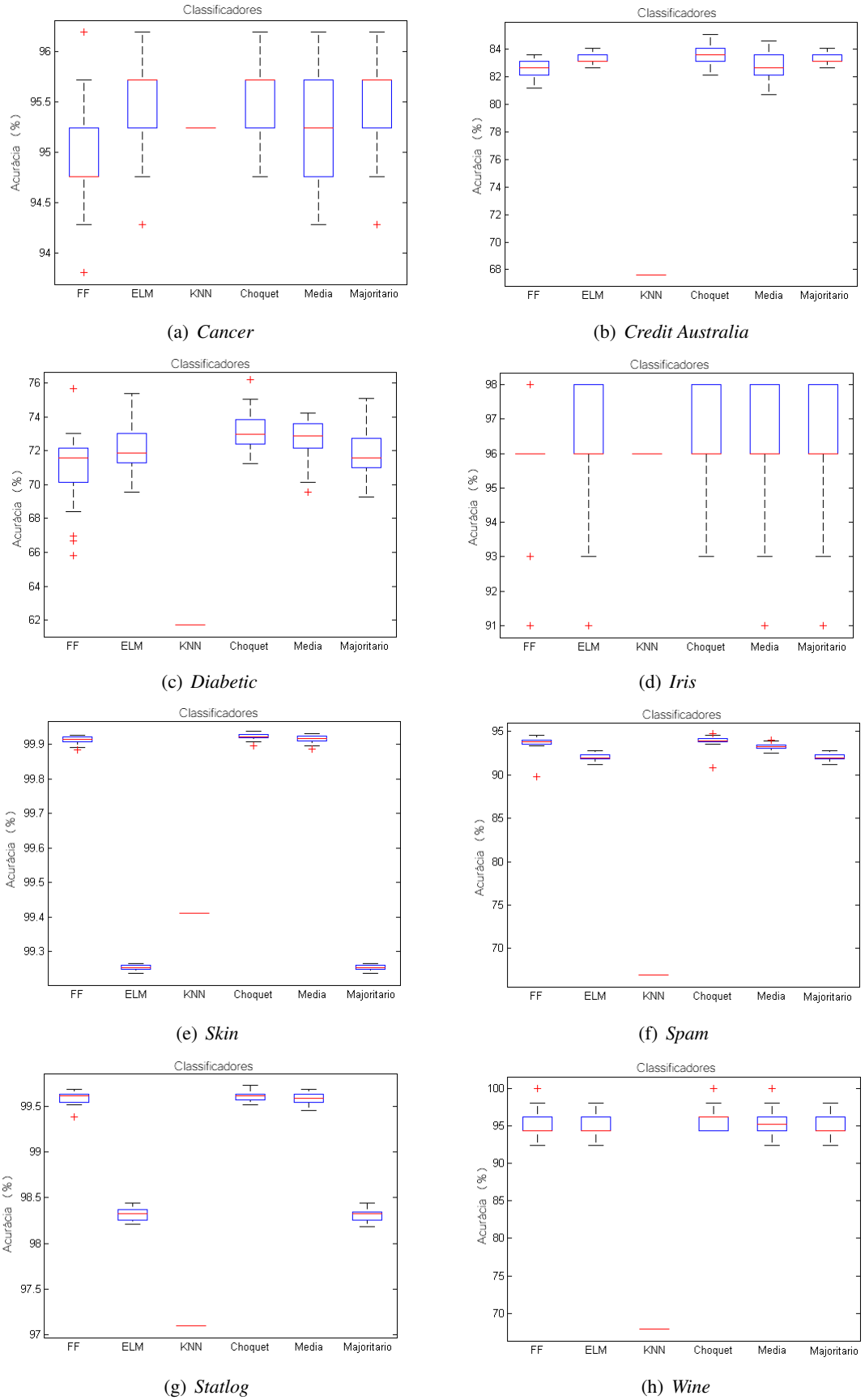
Para executar o experimento foi utilizado ambiente de desenvolvimento MATLAB e a *toolbox* de rede neural nativa do ambiente e implementações próprias. Cada classificador, com exceção do KNN, que não é um algoritmo estocástico, foi executado 30 vezes para cálculo de estatísticas dos resultados. Foram utilizadas oito bases de dados do repositório UCI [Lichman, 2013], que são descritas na Tabela 1. Cada base foi particionada em 70% para treinamento dos algoritmos e 30% para testes dos mesmos.

Base de Dados	# de amostras	# de atributos	# de classes
Cancer	699	9	2
Credit Australia	690	14	2
Diabetic	1151	19	2
Iris	150	4	3
Skin	245057	3	2
Spam	4601	57	2
Statlog	14500	9	7
Wine	178	13	3

Tabela 1: Bases de dados utilizadas nos experimentos

Como mostrado na Tabela 1 as bases de dados possuem configurações diferentes em relação a número de amostras, atributos e classes. As características de cada conjunto de dados também possuem suas peculiaridades, como a base *iris*, que possui dados bem comportados e distribuídos; a base *statlog* que é mal distribuída, 80% dos dados estão na classe 1, e o desafio é conseguir acurácia acima de 99%; a base *spam*, que possui 57 atributos de entradas; e por fim a base *skin*, que possui mais de 245 mil amostras. O desempenho dos classificadores e do elenco em cada base de dados da Tabela 1 são mostrados na Tabela 2, que contém os resultados em termos de acurácia de classificação para cada método. Na Figura 3, são exibidos os *boxplots*, também para cada método. Além da agregação via integral de Choquet, foi utilizado, para fins de comparação, dois métodos de agregação bastante utilizados: a agregação pela média dos suportes e a agregação por do voto majoritário dos classificadores [Krawczyk e Woźniak, 2016].

Analisando os resultados da Tabela 2 e na Figura 3 é possível notar que a agregação dos classificadores por meio da integral de Choquet obteve o melhor desempenho em todas as bases de dados. Apesar da melhoria não ser extremamente significativa, a principal vantagem da abordagem é garantir que o melhor desempenho, dentre os classificadores do elenco, seja transmitido para a classificação final. Em outras palavras, o elenco garantiu uma classificação mais confiável ao se



(a) *Cancer* (b) *Credit Australia*

(c) *Diabetic* (d) *Iris*

(e) *Skin* (f) *Spam*

(g) *Statlog* (h) *Wine*

Figura 3: Boxplot do resultado das classificações para cada base de dados

Acurácia em %						
Base de Dados	**FF**	**ELM**	**KNN**	**Choquet**	**Média**	**Majoritário**
Cancer	$94,98 \pm 0,610$	$95,45 \pm 0,57$	95.24 ± 0	$\mathbf{95,68 \pm 0,29}$	95.09 ± 0.53	$95.47 \pm 0,52$
Credit Australia	$82,62 \pm 0,65$	$83,33 \pm 0,41$	$67,63 \pm 0$	$\mathbf{83,62 \pm 0,62}$	$82,61 \pm 0,97$	$83,12 \pm 0,44$
Diabetic	$70,97 \pm 2,09$	$72,10 \pm 1,36$	$61,74 \pm 0$	$\mathbf{72,98 \pm 1,04}$	$72,71 \pm 1,15$	$71,40 \pm 1,35$
Iris	$95.71 \pm 1,42$	$\mathbf{95,93 \pm 1.79}$	95.55 ± 0	$\mathbf{95.93 \pm 1.31}$	95.77 ± 1.40	95.83 ± 1.75
Skin	$99,91 \pm 0,01$	$99,25 \pm 0,08$	$99,41 \pm 0$	$\mathbf{99,92 \pm 0,01}$	$99,91 \pm 0,01$	$99,35 \pm 0,01$
Spam	$93,68 \pm 0,80$	$92,00 \pm 0,40$	$66,96 \pm 0$	$\mathbf{93,78 \pm 0,64}$	$93,23 \pm 0,34$	$92,35 \pm 0,41$
Statlog	$99,59 \pm 0,06$	$98,32 \pm 0,05$	$97,10 \pm 0$	$\mathbf{99,61 \pm 0,05}$	$99,58 \pm 0,06$	$98,43 \pm 0,06$
Wine	$95,20 \pm 2,07$	$95,15 \pm 1,54$	$67,92 \pm 0$	$\mathbf{96,04 \pm 1,50}$	$95,34 \pm 1,83$	$95,17 \pm 1,52$

Tabela 2: Desempenho dos classificadores para cada base de dados. Em negrito o algoritmo com melhor desempenho considerando a acurácia média

basear no melhor classificador da execução. Em determinadas bases, como *cancer* e *iris*, o ELM foi o classificador individual de melhor desempenho e a agregação o seguiu. Já em outras bases, como *statlog* e *skin*, a rede FF obteve o melhor desempenho individual e a agregação também a acompanhou. Caso fosse utilizado cada um desses classificadores de maneira isolada, a melhor classificação não teria sido garantida.

Dentre as metodologias de agregação a integral de Choquet se destaca perante a média dos classificadores do elenco. Todavia, ambas estão próximas, o que significa que a integral ainda pode ser melhorada. A agregação por voto majoritário foi bastante influenciada pelo desempenho ruim do KNN nas bases maiores e mais difíceis, o que fez com que a mesma tivesse um desempenho ligeiramente abaixo em relação à integral de Choquet e média, que se mostraram mais robustos em relação a isso. Os gráficos da Figura 3 mostram que a integral de Choquet leva vantagem sobre as demais metodologias.

5. Conclusão e trabalhos futuros

Neste trabalho foi apresentada uma metodologia de agregação baseada em integral de Choquet com respeito a medida λ-*fuzzy*. A técnica de Análise de Componentes Principais foi utilizada para determinar as densidades *fuzzy* pertencentes à medida. Um elenco de três classificadores, contendo uma rede neural *feedforward*, uma rede *extreme learning machine* e um K-*nearest neighbor* foi aplicado à oito bases de dados do repositório UCI. Cada classificador individual foi agregado, por meio da integral de Choquet, média e voto majoritário, e seus resultados foram apresentados e discutidos. O uso da integral de Choquet assegurou certa melhora na classificação final e garantiu um resultado mais confiável.

Apesar da integral de Choquet apresentar um bom resultado, as análises mostraram que a técnica ainda tem potencial para ser melhorada. Sendo assim, os próximos passos visam utilizar outras metodologias não supervisionadas para determinar a medida *fuzzy*, sem a simplificação de Sugeno, e incluir mais classificadores no elenco a fim de obter mais diversidade, auxiliando na agregação.

Referências

Cao, Y. (2012). Aggregating multiple classification results using Choquet integral for financial distress early warning. *Expert Systems With Applications*, 39(2):1830–1836.

Chen, J., He, Q., e Li, Y. (2010). Different types of classifiers combination based on Choquet integral. In *IEEE International Conference on Systems Man and Cybernetics (SMC)*, p. 4056–4061.

Chen, T.-Y. e Wang, J.-C. (2001). Identification of λ-fuzzy measures using sampling design and genetic algorithms. *Fuzzy Sets and Systems*, 123(3):321–341.

Choquet, G. (1954). Theory of capacities. *Annales de l'institut Fourier*, 5:131–295.

de Barros, L. C. e Bassanezi, R. C. (2006). *Tópicos de lógica fuzzy e biomatemática*. Grupo de Biomatemática, Instituto de Matemática, Estatística e Computação Científica (IMECC), Universidade Estadual de Campinas (UNICAMP).

Grabisch, M. (1995). Fuzzy integral in multicriteria decision making. *Fuzzy sets and Systems*, 69 (3):279–298.

Haykin, S. (2007). *Neural Networks: A Comprehensive Foundation (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

He, Q., Chen, J.-F., Yuan, X.-Q., e Li, J. (2007). Choquet fuzzy integral aggregation based on g-lambda fuzzy measure. In *International Conference on Wavelet Analysis and Pattern Recognition, ICWAPR'07*, volume 1, p. 98–102.

Höhle, U. (1982). Integration with respect to fuzzy measures. In *Proc. IFAC Symposium on Theory and Applications of Digital Control*, p. 35–37.

Huang, G.-B., Zhu, Q.-Y., e Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1):489–501.

Jia, W. e Zhenyuan, W. (1997). Using neural networks to determine Sugeno measures by statistics. *Neural Networks*, 10(1):183–195.

Jolliffe, I. (2002). *Principal component analysis (2nd Edition)*. Springer, New York, USA.

Keller, J. M., Gray, M. R., e Givens, J. A. (1985). A fuzzy k-nearest neighbor algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, 15(4):580–585.

Kourentzes, N. e Petropoulos, F. (2015). Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics, in press*.

Krawczyk, B. e Woźniak, M. (2016). Untrained weighted classifier combination with embedded ensemble pruning. *Neurocomputing*, 196:14–22.

Krishnan, A. R., Kasim, M. M., e Bakar, E. M. N. E. A. (2015). A short survey on the usage of Choquet integral and its associated fuzzy measure in multiple attribute analysis. *Procedia Computer Science*, 59:427–434.

Larbani, M., Huang, C.-Y., Tzeng, G.-H., et al. (2011). A novel method for fuzzy measure identification. *International Journal of Fuzzy Systems*, 13(1):24–34.

Lee, K.-M. e Leekwang, H. (1995). Identification of λ-fuzzy measure by genetic algorithms. *Fuzzy Sets and Systems*, 75(3):301–309.

Leszczynski, K., Penczek, P., e Grochulski, W. (1985). Sugeno's fuzzy measure and fuzzy clustering. *Fuzzy Sets and Systems*, 15(2):147–158.

Lichman, M. (2013). UCI machine learning repository. URL `http://archive.ics.uci.edu/ml`.

Merigó, J. M., Palacios-Marqués, D., e Ribeiro-Navarrete, B. (2015). Aggregation systems for sales forecasting. *Journal of Business Research*, 68(11):2299–2304.

Mousavi, R. e Eftekhari, M. (2015). A new ensemble learning methodology based on hybridization of classifier ensemble selection approaches. *Applied Soft Computing*, 37:652–666.

Murofushi, T. e Sugeno, M. (1989). An interpretation of fuzzy measures and the Choquet integral as an integral with respect to a fuzzy measure. *Fuzzy sets and Systems*, 29(2):201–227.

Murofushi, T. e Sugeno, M. (1991). A theory of fuzzy measures: representations, the Choquet integral, and null sets. *Journal of mathematical Analysis and Applications*, 159(2):532–549.

Pedrycz, W. e Gomide, F. (1998). *An introduction to fuzzy sets: analysis and design*. Mit Press, Cambridge, USA.

Polikar, R. (2006). Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 6(3):21–45.

Seokho, K., Cho, S., e Kang, P. (2015). Multi-class classification via heterogeneous ensemble of one-class classifiers. *Engineering Applications of Artificial Intelligence*, 43:35–43.

Štefka, D. e Holeňa, M. (2010). Dynamic classifier aggregation using fuzzy integral with interaction-sensitive fuzzy measure. In *10th International Conference on Intelligent Systems Design and Applications (ISDA)*, p. 225–230.

Štefka, D. e Holeňa, M. (2015). Dynamic classifier aggregation using interaction-sensitive fuzzy measures. *Fuzzy Sets and Systems*, 270:25–52.

Sugeno, M. (1974). *Theory of fuzzy integrals and its applications*. Ph.D thesis. Tokyo Institute of Technology, Japan.

Takahagi, E. (2007). A fuzzy measure identification method by diamond pairwise comparisons: AHP scales and Grabish's graphical interpretation. In *Knowledge-Based Intelligent Information and Engineering Systems*, p. 316–324. Springer, Berlin, Germany.

Tapia-Rosero, A., Bronselaer, A., De Mol, R., e De Tré, G. (2016). Fusion of preferences from different perspectives in a decision-making context. *Information Fusion*, 29:120–131.

Xu, L., Krzyżak, A., e Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 22(3):418–435.

Zhang, Q. e Sun, S. (2009). Weighted data normalization based on eigenvalues for artificial neural network classification. In *Neural Information Processing*, p. 349–356. Springer, Berlin, Germany.