Agregação de classificadores

Aluno: André Pacheco
Orientador: Renato A. Krohling

Sumário

- Introdução
- Algoritmo de Classificação
- Fusão de classificadores
- Integral de Choquet
- Medida Fuzzy
- Abordagem de Rowley et. al
- Metodologia de agregação
- Resultados
- Conclusão

Introdução

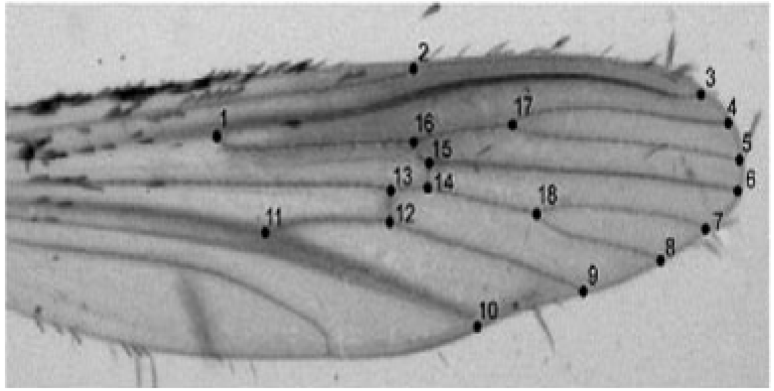
- O problema de classificação de dados consiste em mapear um conjunto de dados de entrada em um número finito de classes (grupos ou rótulos).

- Classificar dados é tarefa tão corriqueira que desde criança realizamos classificações.

- Além disso é um problema que abrange as mais diversas áreas: Biologia, medicina, automobilística, entretenimento, financeira etc.

Introdução

- Um algoritmo de classificação tem como objetivo encontrar correlação entre os atributos de entradas, de uma determinada base de dados, e as respectivas classes em que cada amostra pertence



-0.4933	0.0130
-0.0777	0.0832
0.2231	0.0861
0.2641	0.0462
0.2645	0.0261
0.2471	0.0003
0.2311	-0.0228
0.2040	-0.0452
0.1282	-0.0742
0.0424	-0.0966
-0.0674	-0.1108
-0.4102	-0.0163
-0.3140	0.0318
-0.1768	0.0341
0.0715	0.0509
-0.0540	0.0238
0.0575	-0.0059
-0.1401	-0.0240

127 ESPÉCIES

11 GÊNEROS

Introdução

- Um algoritmo de classificação tem como objetivo encontrar correlação entre os atributos de entradas, de uma determinada base de dados, e as respectivas classes em que cada amostra pertence

- Mas por que usar apenas um algoritmo? Não seria melhor usar um conjunto de algoritmos para classificar uma amostra?

Introdução

- Atualmente, abordagens que fazem uso de fusão de informação vem sendo utilizadas largamente em problemas como: classificação de dados, tomada de decisão, predição de séries temporais, dentre outros.

- Informações obtidas por fontes diferentes, possuem características diferentes. Com isso, a ideia de agregar essas informações visa maximizar a confiabilidade das mesmas.

- Exemplos do cotidiano:
  - Buscar um diagnóstico médico
  - Realizar um investimento

Introdução

- O objetivo de agregar classificadores é buscar uma classificação mais confiável pagando o preço de uma maior complexidade.

- Ao invés de desenvolver **o melhor classificador** de todos os tempos, busca-se **o melhor conjunto de classificadores** com um método de fusão dos mesmos.

- Estatisticamente, utilizar mais de um classificar diminui o risco de uma má classificação.
  - Mas, isso não impede que a classificação, considerando o conjunto, seja pior que um classificador individual.
  - Todavia, a fusão diminui o risco da escolha de um classificador inadequado para o problema

Algoritmo de classificação

LABCIN

- Existem vários algoritmos de classificação e a maioria deles se baseiam em treinamento por exemplos para alcançar a correlação entre atributos e classes.

- Exemplo de classificadores: KNN (K – nearest neighbor), árvores de decisão, Support Vector Machine (SVM), algoritmos baseados em **redes neurais artificiais.**

- Após o treinamento o modelo obtido estará apto a rotular classes para novas entradas.

Algoritmo de classificação

- Classificadores podem ser diferenciados em abstratos e de nível de medida.

Algoritmo de classificação

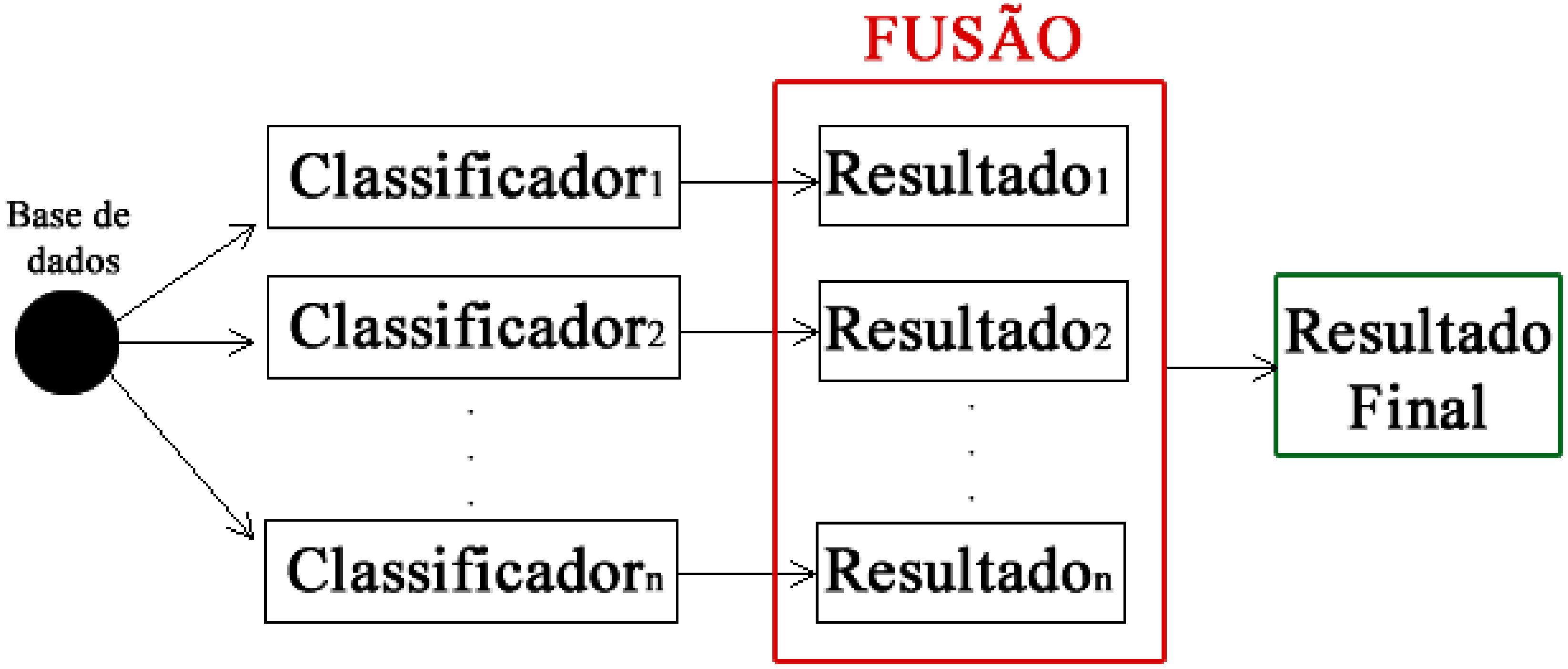
- Classificadores podem ser diferenciados em abstratos e de nível de medida.

Algoritmo de classificação

- **Importante!**
- **Classificar é diferente de *clusterizar*!**
- Algoritmos de classificação em geral, já parte do princípio que a quantidade de grupos é conhecida

Fusão de classificadores

- A ideia é montar um elenco (*ensemble*) de classificadores, atuando em uma base de dados, e ao final realizar a fusão de todos os resultados da classificação.



Fusão de classificadores

LABCIN

- A ideia é montar um elenco (*ensemble*) de classificadores, atuando em uma base de dados, e ao final realizar a fusão de todos os resultados da classificação.

- Para realizar a fusão de informações, existem alguns métodos, como: teoria de Dempster-Shafer, Integral Fuzzy, Integral de Choquet, dentre outros.

- O foco aqui é a **Integral de Choquet**.

Integral de Choquet

- A integral de Choquet foi introduzida, primeiramente, pela teoria da capacidade de Gustave Choquet [1].
- De maneira geral, a integral de Choquet é definida como uma função de integração com respeito a qualquer **medida fuzzy.**
- A integral de Choquet vem sendo utilizada como uma ferramenta de agregação em diversas aplicações: tomada de decisão, predição de séries temporais, **classificação de dados.**

Integral de Choquet

Medida Fuzzy:

Uma função $\mu : \mathbb{A} \to [1, 0]$ é uma medida fuzzy se ela satisfaz:

1. $\mu(\varnothing) = 0$ e $\mu(\Omega) = 1$.
2. $\mu(A) \leq \mu(B)$ se $A \subseteq B, \forall A, B \in \mathbb{A}$ $(monotonicidade)$.

Integral de Choquet

Medida fuzzy de Sugeno [2]:

Uma função $g_\lambda : \mathbb{A} \rightarrow [1,0]$ é uma medida fuzzy de Sugeno se ela satisfaz:

1. $g_\lambda(\Omega) = 1$
2. $g_\lambda(A \cup B) = g_\lambda(A) + g_\lambda(B) + \lambda g_\lambda(A) g_\lambda(B)$,

Em geral, os valores da constante λ podem ser determinadas pela equação a seguir:

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

Integral de Choquet

Seja μ uma medida fuzzy em X, então a integral de Choquet da função $f : X \to [0, \infty]$, com respeito a medida fuzzy μ é definida como:

$$\int f d\mu = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))\mu(A_i)$$

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 $f(x_0) = 0$.

- **Grande problema: Encontrar a medida fuzzy!**
- **No caso da medida de Sugeno: Como encontrar as densidades fuzzy?**

Integral de Choquet

- **Grande problema: Encontrar a medida fuzzy!**
  - Se um humano fosse definir, seriam (2^n-2) coeficientes
- **No caso da medida de Sugeno: Como encontrar as densidades fuzzy?**
  - Melhora a situação anterior, mas ainda assim, temos que definir os valores das densidades

Integral de Choquet

- **A primeira abordagem foi utilizar o *Differential Evolution* para determinar as densidades fuzzy da medida fuzzy de sugeno.**
  - Abordagem supervisionada, ou seja, precisa de saber o resultado final.

  - Precisa de calcular raízes da equação: $1 + \lambda = \prod_{i=1}^{n} (\lambda g_i + 1)$. Quanto mais classificadores, mais raízes.
  - Nem sempre, o conjunto de raízes encontrados satisfaz as condições da medida fuzzy e isso é um grande problema.
  - O algoritmo é muito lento.

Integral de Choquet

- **A segunda abordagem foi utilizar *Principal Componets Analyses* (PCA) para determinar as densidades fuzzy da medida de Sugeno.**
  - Abordagem não-supervisionada
  - Precisa de calcular raízes da equação: $1 + \lambda = \prod_{i=1}^{n}(\lambda g_i + 1)$. Quanto mais classificadores, mais raízes.
  - Nem sempre, o conjunto de raízes encontrados satisfaz as condições da medida fuzzy e isso é um grande problema.
  - Utiliza PCA em matrizes que podem ser muito grandes.

Integral de Choquet

- **A terceira abordagem foi utilizar a abordagem de Rowley et. Al [3] para estimar as medidas fuzzy diretamentes**
  - Abordagem não-supervisionada
  - Não utiliza a simplificação da medida fuzzy de Sugeno, consequentemente, não é necessário calcular raízes
  - Como não calcula raízes, não cai no problema de satisfazer as condições da medida de Sugeno.
  - Continua utilizando "PCA", mas com matrizes pequenas.

Abordagem Rowley et. al

- Rowley et. al propuseram, em 2015, uma nova abordagem para estimar a medida fuzzy.

- Essa nova abordagem foi baseada no trabalho de Kojadinovic [4]. Ambas são não supervisionadas.

- Kojadinovic propôs um método probabilístico baseado em entropia para determinar a medida fuzzy. Todavia, este método necessita de uma grande quantidade de dados para ser realizado com sucesso.

- Portanto, a grande vantagem da nova abordagem é a não necessidade de grandes quantidades de dados.

- **A seguir será apresentado o algoritmo passo a passo.**

Abordagem Rowley et. al

- **Passo 1:**
  - Determinar a matriz de a avaliações de todos os classificadores para cada amostra do conjunto de dados.

$$
M = \begin{array}{c} \begin{matrix} C_1 & C_2 & \cdots & C_j \end{matrix} \\ \begin{bmatrix} c_{11}^{n} & c_{12}^{n} & \cdots & c_{1j}^{n} \\ c_{21}^{n} & c_{22}^{n} & \cdots & c_{2j}^{n} \\ \vdots & \vdots & \vdots & \vdots \\ c_{i1}^{n} & c_{i2}^{n} & \cdots & c_{ij}^{n} \end{bmatrix} \end{array} \begin{matrix} G_1 \\ G_2 \\ \vdots \\ G_i \end{matrix}
$$

$i = 1$ ... nº de grupos da base

$j = 1$ ... nº de classificadores

$n = 1$... nº de amostras da base

$c =$ resultado da classificação

Abordagem Rowley et. al

- **Passo 2:**
  - Construir sub-matrizes com conjuntos $A_x \in M$.
    - Por exemplo, se tivermos 3 classificadores, teríamos os seguintes conjuntos: $\{C_i\},\{C_2\},\{C_3\},\{C_1C_2\},\{C_1C_3\},\{C_2C_2\},\{C_1C_2C_3\}$.
  - Aplicar PCA para cada $A_x \in M$.
    - Calcular matriz de variância-covariância.
    - Determinar os autovalores λ_k da matriz de variância-covariância.

Abordagem Rowley et. al

- **Passo 3:**
  - Para cada conjunto $A_x \in M$ determinar:

$$J^*(A_x) = \sum_{K;\lambda_k<1} \lambda_k + |\{\lambda_k \mid \lambda_k \geq 1\}|$$

  - E por fim, calcula-se a medida fuzzy:

$$\mu(A_x) = \frac{J^*(A_x)}{J^*(M)}$$

Agregando os classificadores

- **Após determinar as Medidas Fuzzy de cada conjunto, podemos realizar a agregação dos classificadores utilizando a Integral de Choquet.**

$$M = \begin{bmatrix} \boxed{c_{11}^n \quad c_{12}^n \quad \cdots \quad c_{1j}^n} \\ \boxed{c_{21}^n \quad c_{22}^n \quad \cdots \quad c_{2j}^n} \\ \vdots \qquad \vdots \qquad \vdots \qquad \vdots \\ \boxed{c_{i1}^n \quad c_{i2}^n \quad \cdots \quad c_{ij}^n} \end{bmatrix} \begin{matrix} \rightarrow \int_{\text{Choquet}} = \\ \rightarrow \int_{\text{Choquet}} = \\ \\ \rightarrow \int_{\text{Choquet}} = \end{matrix} \begin{bmatrix} c_1^n \\ c_2^n \\ \vdots \\ c_i^n \end{bmatrix}$$

$$\max(c_i^n) = \text{grupo da amostra } n$$

Resultados

LABCIN

- **Resultados obtidos para 4 bases de dados:**

Desempenho dos classificadores					Desempenho das agregações		
Base (amostras)	FF	FFDBN	ELMDBN	ELM	DE	PCA	Rowley et. al
Iris (45)	2	2	3	6	1	1	1
Wine (53)	2	15	5	15	1	1	0
Cancer (210)	11	10	10	30	11	11	10
Susy (9000)	1816	1822	1869	1822	1816	1810	1793

Valores em números de erros de classificação

Conclusões

- A nova abordagem de Rowley et. al se mostrou promissora. A metodologia é mais adequada, além de se basear no trabalho de Kojadinovic.

- Nos resultados atuais, a nova abordagem obteve os melhores resultados.

- As atividades atuais consistem em aumentar o número de bases, consolidar um grupo de classificadores e aumentar o número de testes a fim de realizar uma estatística mais confiável.

Principais referências

[1] Choquet, G. Theory of capacities. Annales de l'Institut Fourier, vol. 5, p. 131–295, 1953.

[2] Sugeno M. Theory of fuzzy integrals and its applications. Ph.D. thesis, Tokyo Institute of Technology, Tokyo, Japan, 1974.

[3] Rowley, H. V., Geschke, A., Lenzen, M. A practical approach for estimating weights of interacting criteria from profile sets, Fuzzy Sets Systems, vol. 272, p. 70-88, 2015.

[4] Kojadinovic, I. Unsupervized aggregation of commensurate correlated attributes by means of the Choquet integral and entropy functionals, Int. J. Intell. Syst., vol. 23, p. 128–154, 2008.

Obrigado

Aluno: André Pacheco
Orientador: Renato A. Krohling