

## Sistemas Multi-classificadores ou Ensembles

Prof. Carlos Alberto de Araújo Padilha

10 de Abril, 2018

# Sumário

- 1 Introdução
- 2 Geração de Componentes
- 3 Diversidade
- 4 Práticas

# Sumário

1 Introdução

2 Geração de Componentes

3 Diversidade

4 Práticas

- *Ensembles*, também conhecidos como *máquinas de comitê*, são estruturas baseadas no princípio de dividir para conquistar.
- São constituídos por um conjunto de estimadores individuais organizados de forma paralela, que recebem padrões de entrada e geram saídas que são combinadas de alguma maneira em uma resposta final.
- Segundo [Kuncheva 2004], a saída do *ensemble* é teoricamente superior às respostas geradas individualmente.
- Essa estratégia de combinação visando melhoria de performance não é nova, o trabalho de Nilsson [Nilsson 1965] utilizava uma rede formada por uma camada de perceptrons seguida de um perceptron de votação na camada seguinte.



- Existem 3 razões fundamentais que motivam a utilização de *ensembles*.
  - **Motivo estatístico:** quando temos dados insuficientes, o algoritmo de aprendizagem pode encontrar várias hipóteses diferentes no espaço de busca  $H$ . Com o uso de um *ensemble* composto de bons classificadores, o algoritmo pode realizar uma média das decisões dos componentes, reduzindo a chance de fazer a escolha por uma hipótese errada.
  - **Motivo computacional:** muitos algoritmos de aprendizagem trabalham na busca por um mínimo local. Assim, usando um comitê formado por estimadores obtidos com diferentes pontos iniciais de busca, permite uma melhor aproximação da função desejada.
  - **Motivo representacional:** em muitas casos, uma função  $f$  não pode ser representada fora do espaço de hipóteses  $H$ . Dessa forma, ensembles permitem obter uma representação dessa função fora de  $H$ , permitindo uma melhor aproximação de  $f$ .

- Na construção de *ensembles*, há dois pontos importantes a serem considerados:
  - **Base de estimadores**
    - Um comitê efetivo deve balancear a acurácia individual contra a diversidade entre seus componentes; ou seja, o comitê deve consistir de um conjunto de classificadores que não possuem apenas alta acurácia, mas cujos erros são descorrelacionados [Kuncheva 2004].
    - A escolha correta é fundamental para o desempenho do *ensemble* como um todo.
  - **Método de combinação**
    - Não há consenso na literatura sobre qual é a melhor forma de combinar as respostas dos componentes de um ensemble.
    - Exigindo um teste exaustivo para selecionar o método que melhor atende as necessidades do problema em questão.

- Na literatura, há vários métodos de combinação diferentes, que devido às suas características, podem ser agrupadas em três estratégias principais [Kuncheva 2000]
  - a) **Baseados em fusão**: consideram as respostas de todos os estimadores do ensemble.
  - b) **Baseados em seleção**: escolhem qual estimador vai ser responsável para dar a resposta para uma dada entrada.
  - c) **Métodos híbridos**: mesclam as duas abordagens anteriores, usando a estratégia baseada em seleção somente se o melhor classificador é bom o suficiente. Caso contrário, a estratégia baseada em fusão é usada.



# Sumário

1 Introdução

2 Geração de Componentes

3 Diversidade

4 Práticas

- A geração de componentes para constituir um ensemble deve ter a preocupação de manter/proporcionar um certo nível de diversidade entre eles.
- Dentre os métodos mais importantes encontrados na literatura, os algoritmos de *Bagging* [Breiman 1996] e *Boosting* [Schapire 1990], são os mais mencionados.
- **Bagging** (também conhecido como *Bootstrap Aggregating*)
  - Cada membro do ensemble é treinado com um conjunto de dados gerado através de uma amostragem com reposição dos dados de entrada.
  - Como a probabilidade de cada exemplo ser escolhido é igual para todos, alguns podem ser repetidos e outros deixados de fora no treinamento geral.
  - A combinação de resultados é feita através de uma *votação simples*.

- **Boosting** [Schapire 1990]

- Diferentemente do método *Bagging*, a probabilidade de um padrão ser escolhido depende da participação deste para o erro de treinamento dos classificadores treinados anteriormente.
- Isto é, a probabilidade de uma amostra ser escolhida aumenta em relação as outras, caso ela tenha sido classificada incorretamente antes.
- Sequência genérica
  - O primeiro classificador é treinado seguindo uma amostragem uniforme como no *Bagging*.
  - As instâncias classificadas de maneira incorreta recebem um peso maior para que, no próximo conjunto de treinamento, tenham uma maior probabilidade de serem escolhidos e, assim, sucessivamente.

- Um dos problemas no algoritmo proposto por Schapire [Schapire 1990], conhecido como *Boosting*, é assumir que se dispõe de um grande conjunto de dados.
- Dessa forma, Freund e Schapire [Freund e Schapire 1996] propuseram um novo método chamado de *AdaBoost* (*Adaptive Boosting*), que reúne as principais qualidades dos algoritmos *Bagging* e *Boosting*.
- O *AdaBoost* realiza uma reamostragem com reposição, mas a escolha das amostras é feita de maneira adaptativa.

**Input:** Conjunto de treinamento  $\{(x_i, d_i)\}_{i=1}^N$ , Distribuição  $D$  sobre os  $N$  exemplos rotulados, Inteiro  $T$  especificando o número de iterações e o Modelo de aprendizagem fraca

Inicialização:  $D_1(i) = \frac{1}{N}$  para todo  $i$

**for**  $n = 1$  até  $T$  **do**

    Chame o modelo de aprendizagem fraca, utilizando a distribuição  $D_n$

    Retorne a hipótese  $F_n : \mathbf{X} \rightarrow Y$

    Calcule o erro da hipótese  $F_n$

$$E_n = \sum_{i: F_n(x_i) \neq d_i} D_n(i) \quad (1)$$

    Faça  $\beta_n = \frac{1}{2} \log \frac{1-E_n}{E_n}$

    Atualize a distribuição  $D_n$ :

$$D_{n+1}(i) = \frac{D_n(i) * e^{-\beta_n d_i F_n(x_i)}}{Z_n} \quad (2)$$

    onde  $Z_n$  é uma constante de normalização.

**end**

**Output:** A hipótese final é

$$H(\mathbf{x}) = \text{sign}\left(\sum_{n=1}^T \beta_n F_n(\mathbf{x})\right) \quad (3)$$

# Sumário

1 Introdução

2 Geração de Componentes

**3 Diversidade**

4 Práticas

- Para que um comitê seja efetivo, ele não pode ser formado por classificadores idênticos, mas por aqueles que apresentam erros descorrelacionados.
- Há basicamente quatro formas de criar diversidade em comitês:
  - **Diferentes conjuntos de treinamento**
    - Múltiplos conjuntos de treinamento são criados por reamostragem dos dados originais seguindo alguma distribuição e cada classificador é treinado usando um desses conjuntos.
    - *Bagging* [Breiman 1996] e *Boosting* [Schapire 1990]
  - **Diferentes atributos de entrada**
    - Somente um subconjunto dos atributos dos dados é selecionado para cada conjunto de treinamento.
    - Essa seleção de atributos pode ser feita utilizando as abordagens filtro e *wrapper*.
    - Métodos filtro: são independentes de qualquer algoritmo de aprendizado, suas funções de avaliação consideram apenas as propriedades dos dados.
    - Métodos *wrapper*: utilizam um modelo para avaliar os subconjuntos de atributos.

- **Diferentes configurações de parâmetros**

- A diversidade pode ser alcançada através do uso de diferentes configurações de parâmetros iniciais nos métodos de classificação.
- Como exemplo, no caso de redes neurais, isso significaria ter uma variação no conjunto de pesos e na topologia do modelo.

- **Diferentes algoritmos de aprendizagem**

- A escolha de diferentes tipos de algoritmos para constituir um ensemble (comitê heterogêneo) pode contribuir com a geração de diversidade.
- Isso significa que em um ensemble composto por SVM, árvores de decisão e redes neurais é mais provável que haja desacordo entre os classificadores do que em um composto por somente um desses algoritmos.









# Sumário

- 1 Introdução
- 2 Geração de Componentes
- 3 Diversidade
- 4 Práticas**

# Hora das Práticas!

- 1 Implementação do método Bagging.
- 2 Implementação do método AdaBoost.
- 3 Comparação de resultados dos métodos anteriores com outros métodos ensemble (como o Random Forest, por exemplo) e também com métodos individuais.

-  BREIMAN, L. Bagging predictors. *Machine Learning*, Springer, v. 24, n. 2, p. 123–140, 1996. ISSN 08856125.
-  FREUND, Y.; SCHAPIRE, R. E. Experiments with a New Boosting Algorithm. *Update*, Citeseer, pages, p. 148–156, 1996.
-  KUNCHEVA, L. I. *Combining Pattern Classifiers: Methods and Algorithms*. [S.l.: s.n.], 2004. v. 47. 517–518 p. ISSN 0040-1706. ISBN 0-471-21078-1.
-  NILSSON, N. J. *Learning machines: foundations of trainable pattern-classifying systems*. [S.l.]: McGraw-Hill, 1965. xi, 137 p. p. ISBN 0070465703.
-  SCHAPIRE, R. E. The strength of weak learnability. *Machine Learning*, v. 5, n. 2, p. 197–227, 1990. ISSN 08856125.
-  SCHAPIRE, R. E. A brief introduction to boosting. In: *IJCAI Int. Jt. Conf. Artif. Intell.* [S.l.: s.n.], 1999. v. 2, p. 1401–1406. ISBN 3540440119. ISSN 10450823.