

Estudo de Caso 02: Comparação entre algoritmos de Classificação

Equipe 04

15 de maio de 2017

Coordenador: Danny Tonidandel

Relator: Alessandro Cardoso Verificador: Gustavo Vieira Monitor: Bernardo Marques

O Experimento

Este estudo de caso consiste na comparação de algoritmos de classificação, em que um deles consiste no *padrão atual* e um novo algoritmo proposto, que utiliza uma técnica de simplificação baseada em inferência estatística.

Segundo os pesquisadores responsáveis pelo algoritmo proposto, este apresenta uma melhora significativa frente ao padrão atual com relação ao tempo requerido para a classificação e que essa nova abordagem não resulta em grandes perdas de desempenho em termos de acurácia da classificação.

O experimento deste trabalho foi desenvolvido no intuito de verificar as afirmações acima. Este consiste na comparação dos dois algoritmos citados baseando-se nos questionamentos abaixo:

Questionamentos

1. O método proposto realmente apresenta ganhos em relação ao tempo de execução, quando comparado ao método padrão?
2. O método proposto realmente não resulta em variações consideráveis de acurácia?

Características desejadas

Para que sejam investigados os questionamentos acima são desejadas as seguintes características para os testes estatísticos:

- Probabilidade de rejeitar a hipótese nula (i.e., de incorrer em erro do tipo I), também chamado de nível de significância: $\alpha = 0.05$;
- O menor valor de interesse prático para os ganhos em tempo de execução: $d_{tempo}^* = 1.5$;
- A faixa de valores para os quais as acurácias estão, relativamente, “próximas” para serem consideradas equivalentes, ou margem de equivalência para a acurácia: $\delta_{acuracia}^* = 0.05$.
- Potência estatística: $\pi = 0.8$;

Planejamento experimental

Os dados experimentais para os testes foram gerados por meio de uma aplicação disponibilizada na web em <http://orcslab.cpdee.ufmg.br/3838/classdata>. Os dados são referentes ao tempo necessário para classificação (*Time.s*) e acurácia (*Accuracy*) por cada um dos algoritmos em cada instância e em cada

execução. O grupo de trabalho deve fornecer a data de nascimento do membro mais jovem da equipe e selecionar um número de instâncias e execuções por instâncias.

Estes dados são apresentados conforme seleção de número de instâncias e número de repetições das mesmas. Segundo [1] quando cada par de observação é coletado sobre condições homogêneas, ainda assim estas podem variar de um par para outro. Conforme apresentado em [2], a variabilidade devida aos diferentes problemas de teste é uma forte fonte de Variação espúria que pode e deve ser controlada e que uma solução elegante para eliminar a influência deste inconveniente é o pareamento das medições por problema.

Conforme a apresentação das amostras, verificou-se a necessidade de efetuar os testes estatísticos com as amostras pareadas. Conforme [1] o procedimento experimental mais adequado quando os dados são coletados aos pares é o *t-test pareado*. Para tal, seja $(X_{11}, X_{21}), (X_{12}, X_{22}), \dots, (X_{1n}, X_{2n})$ um conjunto de n observações pareadas onde assumimos que a média e a variância da população representada por X_1 são μ_1 e σ_1^2 , e a média e a variância da população representada por X_2 são μ_2 e σ_2^2 . Defini-se as diferenças entre cada par de observações como $D_j = X_{1j} - X_{2j}$, $j = 1, 2, p, \dots, n$. Os D_j 's são assumidos como sendo normalmente distribuídos na média.

Os testes de hipótese elaborados são:

- Referente ao questionamento 1 definiu-se o teste de hipótese convecional como abaixo:

$$\begin{cases} H_0 : \mu_p - \mu_a = 0 \\ H_1 : \mu_p - \mu_a < 0 \end{cases}$$

- Referente ao questionamento 2 definiu-se o teste de não-inferioridade como abaixo:

$$\begin{cases} H_0 : \mu_p - \mu_a = -\delta_{acurácia}^* \\ H_1 : \mu_p - \mu_a < 0 \end{cases}$$

Sendo μ_p e μ_a , as médias de tempo de execução dos algoritmos *proposto* e *padrão atual* respectivamente.

Tamanho amostral

Afim de se atingir as características desejadas exigidas para o experimento é necessário definir o tamanho amostral a ser utilizado.

No caso de amostras pareadas, entende-se por amostras o números de instâncias conforme [3], ou problemas a serem resolvidos pelos algoritmos deste trabalho. Outra consideração é o número de repetições a serem adotadas por instância. Para este estudo adotou-se o valor de 30 repetições por instância seguindo recomendação expressão em [2] de que um valor heurístico para esta finalidade deve ser maior ou igual a 30.

Para a determinação do tamanho amostral foi realizado um *Estudo Piloto*, de forma a determinar uma estimativa inicial para os testes atendendo aos parâmetros desejados:

Como não há informação histórica para a determinação inicial da variância dos dados, e supondo amostras de mesmo tamanho, dado que dispõe-se do menor tamanho de interesse prático para os ganhos de tempo

$$n = 2 \left(\frac{1}{d_{tempo}^*} \right)^2 (t_{\alpha/2} + t_{\beta})^2$$

,

- Nascimento: 21/11/1992;
- Número de instâncias de cada problema: 17;
- Número de execuções: 30;

```
#dados <- read.csv("1992-11-21_17_30.csv", header = T)
#head(dados)
```

Para o problema proposto, são desejadas as seguintes características (parâmetros dos testes estatísticos):

Mostra-se algumas características do conjunto de dados gerado

```
# str(dados)
```

Referências

- [1] D. C. Montgomery and G. C. Runger, *Applied statistics and probability for engineers*, vol. 5. John Wiley; Sons, 2011.
- [2] F. Campelo, “Lecture notes on design and analysis of experiments.” <https://github.com/fcampelo/Design-and-Analysis-of-Experiments>, 2015.
- [3] E. Walker and A. S. Nowacki, “Understanding equivalence and noninferiority testing.” *Journal of general internal medicine*, vol. 26 2, pp. 192–6, 2011.