



Universidade do Estado do Amazonas
Escola Superior de Tecnologia
Algoritmos e Estrutura de Dados II (Code: ESTCMP012)

17 de novembro de 19

Tempo: 05 dias

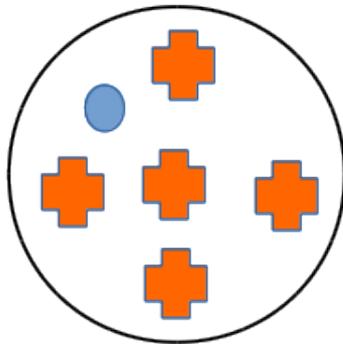
Trabalho Prático II - Parte II

Especificação do Trabalho

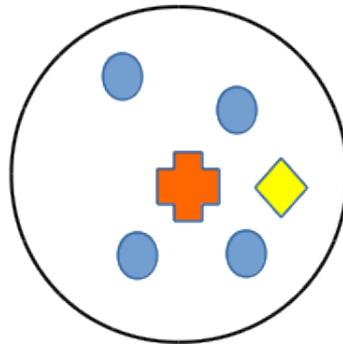
O trabalho é uma continuação do Trabalho Prático II apresentado no dia 17 de outubro de 2019. Portanto, o domínio do problema será o mesmo. A forma de avaliação dos métodos é que será diferente. A forma de avaliação será explicada adiante.

Deve-se implementar o mesmo método de agrupamento de grafos da Parte I do Trabalho Prático II. Este método estava organizado em quatro etapas. Porém, agora deve-se considerar o novo dataset *datav2.json*. Os atributos são os mesmos, mas houve uma redução do número de aspectos. São apenas quatro aspectos para cada atributo. O objetivo é que os aspectos estejam mais relacionados com os atributos. Além desse método, a dupla deverá implementar um segundo método. Esse segundo método será uma variação do primeiro e ficará a cargo da dupla. Sempre com o objetivo de melhorar o resultado da primeira versão do método. Em caso de dúvidas, deve-se procurar o professor em tempo hábil para maiores detalhes.

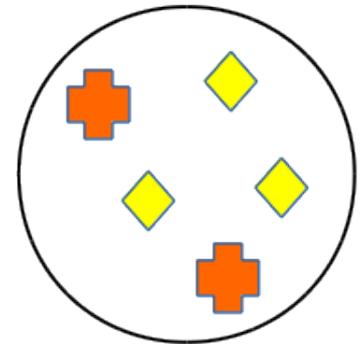
De posse do resultado dos grupos (*clusters*), deve-se avaliar se os aspectos foram corretamente agrupados. Para isso, deve-se implementar os métodos de avaliação explicados a seguir:



Grupo 1



Grupo 2



Grupo 3

ENTROPIA

A entropia define a homogeneidade dos grupos encontrados, ou seja, de que forma que as classes dos objetos estão distribuídas nos grupos, sendo que baixa entropia indica grupos mais homogêneos.

Passo 1

$$H(G1) = 5/6 \cdot \log(5/6) + (1/6) \cdot \log(1/6) + 0/6 \cdot \log(0/6) = -0,21919533819482817 - 0.430827083453526 + 0 = -0.6500224216483541$$

$$H(G2) = 1/6 \cdot \log(1/6) + 4/6 \cdot \log(4/6) + 1/6 \cdot \log(1/6) = -0.430827083453526 -0.38997500048077083 -0.430827083453526 = -1.2516291673878228$$

$$H(G3) = 2/5 \cdot \log(2/5) + 0/5 \cdot \log(0/5) + 3/5 \cdot \log(3/5) = -0.5287712379549449 + 0 -0.44217935649972373 = -0.9709505944546686$$

Passo 2

$$H(\text{Final}) = (-1) \cdot [H(G1) \cdot 6/17 + H(G2) \cdot 6/17 + H(G3) \cdot 5/17]$$

$$H(\text{Final}) = (-1) \cdot [-0.22941967822883086 -0.44175147084276095 -0.28557370425137313]$$

$$H(\text{Final}) = 0.956744853322965$$

PUREZA

A pureza fornece a razão da classe dominante no grupo com relação ao tamanho do próprio grupo. Valores de pureza próximos a 1 indicam um subconjunto puro da classe dominante no grupo obtido.

$$P(G1) = 5$$

$$P(G2) = 4$$

$$P(G3) = 3$$

$$\text{Pureza (Final)} = 1 / \text{TOTAL} \times [P(G1) + P(G2) + P(G3)]$$

$$\text{Pureza (Final)} = 1 / 17 \times [5 + 4 + 3]$$

$$\text{Pureza (Final)} = 0.7058823529411764$$

Regras:

- O professor irá detalhar e tirar dúvidas sobre o trabalho em sala dia no dia 21 de novembro no horário da aula.
 - Os trabalhos deverão ser realizados em dupla.
 - O início do trabalho será no dia **17 de novembro de 2019** (15:00 horas) e a entrega será no dia **25 de novembro de 2019** (até as 10 horas).
 - Cada dupla terá **até 5 minutos** para apresentar o trabalho. A apresentação deverá ser preparada para esse tempo. Nesta defesa, a dupla deverá apresentar e justificar os algoritmos escolhidos para geração do resultado. Além disso, deverá apresentar o resultado e explicar o quanto ele ficou próximo do gabarito. Deve-se considerar as métricas de **pureza e entropia**.
 - Deverá ser encaminhado para o e-mail `tmelo@uea.edu.br` do professor o relatório, os slides e o código-fonte. Os três arquivos deverão estar compactados (.zip) e o título do e-mail deverá ser *“Trabalho Pratico II - Parte II - AED 2”*.
 - A defesa do trabalho deverá acontecer no dia **26 de novembro de 2019** no horário da aula. A ordem das apresentações será por sorteio. Portanto, todas as duplas deverão estar presentes no início das apresentações.
 - Esse trabalho será composto por duas notas: a implementação e a defesa. Estilo de programação e técnicas adotadas serão consideradas na avaliação da implementação. Segurança e conhecimento demonstrado na defesa serão considerados como critérios de avaliação.
 - A dupla deverá descrever no relatório o comando para executar o programa. Além disso, se for necessário algum arquivo auxiliar, esse arquivo deverá ser encaminhado junto com os demais. É responsabilidade da dupla orientar o professor na execução do programa enviado para avaliação.
 - A discussão sobre o trabalho deverá acontecer **somente** entre a dupla.
 - A versão padrão do Python será a 2.7. Caso a implementação tenha sido em outra versão, a dupla deverá informar no relatório.
-