

PGCA008 - Mineração de Dados
Prof. Rodrigo Tripodi Calumby
Exercício Computacional

**Estudo Experimental Comparativo Qualitativo de Algoritmos de
Agrupamento de Dados**

Objetivos

1. Aplicar algoritmos de agrupamento de dados sobre diferentes datasets.
2. Realizar estudo de otimização paramétrica para cada algoritmo e analisar seu impacto sobre a qualidade do resultado produzido.
 1. Computar métricas de eficácia e gerar gráficos demonstrativos para as variações de cada algoritmo, por exemplo, o impacto da variação do número de grupos gerados na qualidade dos agrupamentos.
 2. Apresentar discussão sobre o impacto dos parâmetros dos algoritmos nos resultados gerados.
3. Analisar a eficácia dos algoritmos considerando as características de cada dataset
4. Propor e aplicar um protocolo de avaliação e comparação da eficácia dos algoritmos de modo a garantir uma análise adequada, detalhada e rigorosa.
 1. Considerando os dados disponíveis, deve ser definida qual a melhor forma de analisar a qualidade do resultado.
5. Apresentar estudo comparativo para as variações de um mesmo algoritmo e comparativo dos melhores resultados de cada algoritmo.

Requisitos

1. Deve ser utilizado pelo menos um algoritmo de cada categoria: particional, hierárquico e baseado em densidade. Outros tipos de algoritmos ou algoritmos híbridos podem ser utilizados também.
2. Avaliar a eficácia de cada algoritmo em cada um dos datasets considerando o ground-truth fornecido e medidas intrínsecas.
3. Apresentar resultados visuais dos melhores agrupamentos encontrados.
4. As ferramentas, bibliotecas e fontes de informação utilizadas para realização dos experimentos devem ser citadas adequadamente. Em caso de utilização de códigos/scripts próprios, os mesmos devem ser disponibilizados junto ao relatório.

Bases de Dados

Os experimentos devem ser realizados utilizando os 14 datasets fornecidos. Estes datasets são conjuntos de dados disponíveis na literatura e já utilizados em trabalhos anteriores de avaliação de algoritmos de agrupamento.

Observações

1. Diferentes medidas de similaridade podem ser consideradas.
2. Os experimentos devem ser realizados considerando que a quantidade de grupos não

é conhecida. Experimentos comparativos adicionais podem ser realizados utilizando a quantidade de grupos informada no ground-truth.

3. Não é preciso descrever os algoritmos no artigo, mas referências devem ser fornecidas para algoritmos pouco conhecidos.
4. Os experimentos podem ser realizados utilizando ferramentas de mineração de dados ou utilizando códigos próprios com ou sem o uso de bibliotecas especializadas.
5. Etapas de pré-processamento podem ser aplicadas, quando apropriadas e devem ser citadas no artigo.
6. Trabalhos fraudulentos serão anulados, acarretando nota zero.
7. Em caso de entrega de trabalhos replicados, mesmo que parcialmente, os trabalhos envolvidos serão anulados, acarretando nota zero.

Relatório

O relatório deve ser produzido em formato de artigo científico, seguindo o modelo de formatação *IEEE Transactions* (disponível no link abaixo). Utilize preferencialmente o template LaTeX. O relatório deve ter no máximo 6 páginas (incluindo imagens e referências).

http://www.ieee.org/publications_standards/publications/authors/author_templates.html#sect1

Duas versões do artigo devem ser enviadas. Uma versão com nomes de autores e outra versão sem nomes de autores no cabeçalho. Os nomes dos arquivos devem ser, respectivamente, artigo_autor1_autor2.pdf e artigo.pdf.

Alguns Critérios de Avaliação

1. Atendimento ao prazo de entrega estabelecido.
2. Atendimento aos objetivos, requisitos e orientações desta proposta.
3. Adequação do protocolo experimental aplicado aos dados utilizados.
4. Amplitude dos experimentos realizados.
5. Rigor científico dos experimentos realizados e resultados apresentados.
6. Qualidade da discussão dos resultados e conclusões.
7. Apresentação.
 1. Qualidade da apresentação visual dos experimentos e resultados.
 2. Atendimento à formatação indicada.
8. Qualidade do texto do relatório.
 1. Clareza e objetividade.
 2. Atendimento às normas de ortografia e gramática.
 3. Adequação das citações.
 4. Organização das informações.