

Trabalho de Agrupamento de Dados

Elaine Faria

elaine@ufu.br

Agrupamento de Dados - 2024
Pós-graduação em Ciência da Computação
Faculdade de Computação
Universidade Federal de Uberlândia

1. Introdução

O objetivo principal do trabalho é realizar um estudo detalhado sobre métodos de agrupamento e sua utilização em diferentes bases de dados.

- Cada grupo poderá ser composto por no máximo 3 alunos.
- Valor: 30 pontos.
- Data de Entrega do Artigo Final: 17/06/24

2. Descrição do Trabalho

A. Para alunos que não possuem projeto de mestrado/doutorado definido ou que não pretendem usar técnicas de agrupamento nos seus projetos

Os alunos deverão implementar ou buscar o código fonte na Internet de 3 algoritmos de agrupamento, sendo que pelo menos um deles deve ser um algoritmo que não foi visto em sala de aula.

Os algoritmos escolhidos deverão ser executados em 4 bases de dados diferentes. As bases de dados podem ser obtidas na Internet (por exemplo, no repositório UCI). Os pré-processamentos necessários deverão ser executados em cada base de dados.

O resultado do agrupamento em cada base de dados deve ser analisado usando medidas de validação vistas em sala de aula. Pelo menos uma medida de validação interna (ou relativa) e uma medida de validação externa devem ser usadas.

Cada grupo produzirá um relatório (ver Seção 3) que apresentará os algoritmos, bases de dados e resultados encontrados.

B. Para alunos que pretendem usar técnicas de agrupamento nos seus projetos ou que gostariam de pelo menos investigar uma base de dados relacionada ao seu tema e investigar algoritmos de agrupamento neste contexto.

Neste caso, o aluno poderá usar uma base própria. A base não precisa ser divulgada, mas a sua estrutura sim, com informações tais como nro de atributos, tipo de atributos, se tem classe ou não, se tiver a distribuição de instâncias por classe. O aluno também poderá usar uma base da Internet que é relativa ao seu projeto.

O objetivo aqui é aplicar o algoritmo de agrupamento em um contexto mais específico do projeto do aluno. Nesse caso, agende uma conversa com o professor antes, para definirem um novo escopo para o projeto.

3. Artigo

Um artigo deve ser produzido com os resultados do desenvolvimento do trabalho. O artigo deve conter uma descrição sucinta de cada algoritmo utilizado e o seu respectivo pseudocódigo, uma descrição sucinta de cada base de dados utilizada, uma descrição sucinta de cada medida de validação. O documento também deve conter os resultados da execução de cada algoritmo em cada base de dados, bem como uma discussão dos resultados obtidos. Lembre-se, se algum pré-processamento foi feito na base de dados, ele deve ser descrito no artigo. Lembre-se de adicionar ao artigo uma introdução, discutir os resultados obtidos e conclusões.

Forneça detalhes sobre o código usado, indicando onde ele está disponível (*framework*, ferramenta, biblioteca, etc.) e como ele foi configurado para execução. Se a implementação foi feita pelo grupo, dê detalhes do código.

O artigo deverá ser enviado pelo Microsoft Teams no formato .pdf.

Para a opção B, o artigo também precisa contextualizar a importância do uso de técnicas de agrupamento no problema proposto.

Lembre-se, se o artigo for publicado em alguma conferência ou periódico, o nome do professor da disciplina NÃO deverá ser listado como autor do artigo.

4. Regras

- Não serão aceitos trabalhos atrasados. Se o grupo não entregar o trabalho no dia combinado, ele receberá nota zero.

- Em caso de projetos copiados de colegas todos os envolvidos recebem nota zero. Lembre-se é muito improvável que haja trabalhos iguais, afinal há várias bases de dados e diferentes algoritmos de agrupamento. Se os alunos usarem códigos disponíveis na Internet, é preciso entendê-lo antes da apresentação. É preciso também citar a referência a eles.
- O professor em hipótese alguma verificará ou ajudará na construção do código fonte.
- O professor poderá tirar dúvidas conceituais sobre o trabalho em horário de aula ou horário extra-classe.
- A interpretação dos resultados e o entendimento dos algoritmos fazem parte da avaliação e devem ser realizados pelos alunos.