

Especificação do Trabalho

Retificado (última alteração) em 05/07/2016

GCC128 - Inteligência Artificial - 2016/1

Vinicius Ruela Pereira Borges

vinicius.borges@dcc.ufla.br

Objetivo

Projetar uma técnica de busca, de classificação ou de agrupamento de dados considerando um domínio do conhecimento específico. O trabalho da disciplina envolve a implementação e realização de experimentos com a técnica projetada e a escrita de um artigo descrevendo a técnica e os trabalhos “estado-da-arte”.

Como proceder

Escolha um problema do domínio do conhecimento e defina a tarefa a ser realizada. Procure por conjuntos de dados:

1. UCI Machine Learning Repository: <https://archive.ics.uci.edu/ml/datasets.html>
2. Delve datasets: <http://www.cs.toronto.edu/~delve/data/datasets.html>
3. UCI Knowledge Discovery: <http://kdd.ics.uci.edu/>

... e decida se a tarefa será de busca, classificação ou clustering.

Um exemplo de um problema de domínio do conhecimento seria: “Análise de clima no mundo”, em que os dados são informações de temperatura para diversas cidades ao redor do mundo. Para isso, utiliza-se uma técnica de agrupamento para agrupar as cidades mais semelhantes considerando a temperatura, a umidade relativa, radiação UV etc. Dados sobre o domínio do conhecimento escolhido encontram-se nas respectivas páginas citadas acima.

Você pode optar por modificar ou adaptar uma técnica de classificação ou agrupamento já existente na literatura atual. Lembre-se que você precisará de ajustar uma medida de dissimilaridade entre dos dados e optar por uma estratégia de avaliação da técnica a ser implementada.

Para o artigo, vocês deverão fazer uma revisão de literatura considerando **a tarefa e o domínio do conhecimento** escolhido para o trabalho. Por exemplo, se o tema escolhido foi “Análise de clima no mundo”, vocês deverão pesquisar em ferramentas disponíveis na Web por **trabalhos científicos** que estudem/analise/implementem outras técnicas de “Análise de clima no mundo”. **Não serão aceitas referências providas de sites de internet, Wikipédia** ou fontes relacionadas. Pesquisem em artigos científicos publicados em anais de congressos, monografias de TCC, dissertações de mestrado e teses de doutorado. Sugestões:

Google Scholar: <http://scholar.google.com.br/>

CAPES Periódicos: <http://www.periodicos.capes.gov.br/>

Além disso, o artigo segue um formato:

- No máximo 4 páginas
- Língua portuguesa ou inglesa

- Deve conter as Seções: Introdução, Revisão de literatura, Metodologia proposta, Resultados experimentais, Conclusão e referências.
- No início do documento, escrever um resumo do trabalho utilizando no **mínimo de 150 palavras** e **máximo 200 palavras**.
- Template disponível (em formato .doc e L^AT_EX):
http://www.ieee.org/conferences_events/conferences/publishing/templates.html

Parte I

Na primeira parte do trabalho, os módulos de leitura de dados a partir de arquivos em disco (a ser visto em laboratório) e de pré-processamento e tratamento de dados deverão ser apresentados.

Para o módulo de pré-processamento e tratamento de dados, pelo menos uma das funcionalidades abaixo de cada tópico deverá ser implementada:

1. Pré-processamento de dados: tratamento de dados, limpeza de dados, discretização ou normalização.
2. Redução de dimensionalidade: seleção de características ou transformação do espaço de atributos.

Resumindo, deverá ser entregue na primeira parte do trabalho via moodle o arquivo .PDF ou .DOC do artigo, contendo **ao menos** as Seções de **Introdução** e **Revisão bibliográfica**.

Na apresentação do trabalho, devem ser apresentados os módulos de leitura de dados a partir de arquivos em disco, e:

- ao menos um método de pré-processamento e tratamento de dados;
- ao menos um método de redução de dimensionalidade;

Parte II

Na segunda parte do trabalho, os módulos de leitura de dados e de pré-processamento deverão ser integrados com a técnica proposta de classificação / agrupamento, formando um projeto completo. Também será necessário um módulo que componha a estratégia de avaliação da técnica proposta considerando o domínio do conhecimento, sendo desejável a comparação com algum outro trabalho da literatura.

Resumindo, deverá ser entregue na segunda parte do trabalho:

- A implementação da técnica proposta (integrada com o primeiro trabalho) e o módulo da avaliação dos resultados experimentais
- O artigo completo.

Como proceder

Realizar o trabalho em grupos de **no máximo 3 alunos**. Enviar email para vinicius.borges@dcc.ufba.br informando o nome e a turma de cada um. Por isso, é permitido um trio conter alunos das turmas 10A e 14A. No entanto, a presença de todos é **obrigatória** na apresentação do trabalho.

Deadlines

Parte I:

~~8 de julho de 2016, 23:55~~, 11 de julho de 2016, 23:55

Parte II:

12 de agosto de 2016, 23:55

Apresentações

Parte I:

12, 13 e 15 de julho de 2016.

Parte II:

16, 17 e 19 de agosto de 2016.

Caso queiram apresentar antes das datas propostas, agendar horário via e-mail com o professor para apresentação e entrega de trabalho em horário extra-aula.