

Tarefa 2: Classificação, Clustering e Trajetória

Resgate de Vítimas de Catástrofes Naturais, Desastres ou Grandes Acidentes

1 Versão

v1 14/10/2023	original
v2 6/11/2023	explicações adicionais sobre os arquivos de teste cego e arquivos CSV de saída

2 Objetivos da tarefa

No estágio atual, os agentes socorristas conhecem total/parcialmente o mapa do local do acidente e a localização das vítimas com seus sinais vitais que foram encontradas pelos exploradores.

Na presente tarefa, os socorristas devem socorrer as vítimas. Cada um deve socorrer um grupo de vítimas segundo critérios estabelecidos pelos projetistas do sistema multiagente. Dentre os critérios, a classe de gravidade deve obrigatoriamente ser utilizada e calculada por alguma técnica de classificação a partir dos sinais vitais das vítimas localizadas. Com os grupos de vítimas definidos, cada agente socorrista fica responsável por um deles. Conhecendo o grupo e a localização das vítimas, cada agente socorrista define uma trajetória para deixar o kit de socorro. A trajetória deve buscar minimizar o custo do percurso saindo da base e voltando a ela.

Portanto, esta tarefa possui os seguintes objetivos:

- 1) **classificar a gravidade das vítimas;**
- 2) **agrupá-las e atribuí-las aos socorristas;**
- 3) **calcular a trajetória para cada socorrista salvar seu grupo de vítimas.**

As seções seguintes detalham cada um dos objetivos e o que deve ser entregue.

3 Classificação

Os agentes socorristas devem aprender a classificar o estado de gravidade das vítimas utilizando obrigatoriamente um algoritmo árvore de decisão indutiva ou um Sistema de Inferência Fuzzy com Wang-Mendel.

Nesta tarefa, vocês terão acesso a um histórico de sinais vitais que foram coletados por um corpo médico de outros acidentes contendo os seguintes dados:

- pSist: **-- não utilizar ---** pressão diastólica
- pDiast: **-- não utilizar ---** pressão diastólica
- **qPA**: qualidade da pressão arterial; resulta da avaliação da relação entre a pressão sistólica e a diastólica;
- **pulso**: pulsação ou Batimento por Minuto (pulso);
- **frequência respiratória**: frequência da respiração por minuto;
- gravidade: **-- não utilizar ---** valor calculado em função dos sinais vitais acima
- **classes de gravidade**: são 4 classes que apresentam o estado de saúde do acidentado.

O corpo médico construiu uma fórmula para determinar a classe de gravidade de uma vítima em função da qPA, pulso e frequência respiratória:

- 1 = crítico,
- 2 = instável,
- 3 = potencialmente estável e
- 4 = estável.

O problema é que a fórmula de cálculo foi perdida. Portanto, você deve utilizar técnicas de aprendizado de máquina (árvore de decisão ou fuzzy com Wang-Mendel) para que os agentes socorristas aprendam a classificar as vítimas nas quatro classes.

Com base nos modelos aprendidos, o desempenho do classificador deve ser calculado com as seguintes métricas: **precisão, recall, *f-measure* e acurácia**. O sistema deve ser capaz de imprimir estas métricas.

Importante: treinar e validar os classificadores com diferentes estruturas e parametrizações. Por exemplo, se você utilizar um sistema de inferência fuzzy (SIF) poderá mudar as variáveis linguísticas que caracterizam as entradas (o total de termos linguísticos, as funções de pertinência que os definem). Se utilizar árvore de decisão indutiva, escolher diferentes parametrizações de poda.

Para extrair um comportamento médio independente dos dados de treinamento/validação, **você deve utilizar o método de validação cruzada para cada configuração tentada**. A partir daí, você seleciona o modelo que gerou o melhor resultado para utilizá-lo na fase de testes.

No dia da apresentação, o sistema deve estar pronto para **executar com dados de teste** que não foram utilizados nem no treinamento, nem na validação. Também, deve ser capaz de imprimir as métricas acima listadas.

4 Agrupamento e Atribuição

Com as vítimas localizadas e um modelo de classificação de gravidade, os agentes socorristas podem decidir qual agente salva qual grupo de vítimas. Os grupos de vítimas devem ser definidos obrigatoriamente por um algoritmo de *clustering* (e.g. K-means). É obrigatório utilizar a classe de gravidade como um dos critérios. Além deste, os projetistas podem adicionar um ou mais critérios.

O sistema deve ser capaz de imprimir os grupos atribuídos a cada agente. Para cada grupo, imprimir o **id da vítima, as coordenadas, o valor de gravidade e a classe de gravidade predita** no formato CSV: <id, x, y, grav, classe>. Nesta tarefa, como vocês não geram o valor de gravidade, deixá-lo zerado. A classe de gravidade é o valor predito pelo método escolhido (Fuzzy ou árvore de classificação).

5 Trajetória e Salvamento

Estabelecer, para cada agente socorrista, uma trajetória para salvar as vítimas do grupo pelo qual é responsável de forma a minimizar o custo do percurso. O socorrista deve sair da base, salvar todas as vítimas possíveis dentro do tempo limite de socorro (Ts) e voltar à base. Para salvar uma vítima, é necessário deixar o kit de salvamento junto dela. Ao terminar o socorro de um grupo, o agente socorrista pode ajudar outros (opcional).

Para definir a trajetória, é obrigatório utilizar um algoritmo evolutivo (e.g. Algoritmos Genéticos, *Particle Swarm Optimization*).

O sistema deve ser capaz de imprimir uma lista de todas as vítimas salvas por todos os agentes socorristas contendo o **id da vítima, as coordenadas x e y, o valor de gravidade e a classe de gravidade predita** no formato CSV: <id, x, y, grav, classe>. Nesta tarefa, como vocês não geram o valor de gravidade, deixá-lo zerado. A classe de gravidade é o valor predito pelo método escolhido (Fuzzy ou árvore de classificação).

6 Arquivos de Treinamento e Testes

6.1 Arquivo de Treinamento do Classificador

Este arquivo contém os dados históricos de sinais vitais de vítimas de outros acidentes. Cada linha representa uma vítima.

Para uma vítima i do histórico temos 5 sinais vitais (s_1 até s_5) que resultam a gravidade g_i da vítima (valor real) e o rótulo que identifica a classe de gravidade y_i

$i \quad s_{i1} \quad s_{i2} \quad s_{i3} \quad s_{i4} \quad s_{i5} \quad g_i \quad y_i$

i : identificação da vítima (número sequencial)

s_{i1} : não utilizar - pressão sistólica (pSist): [5, 22] - não usar, é utilizada no cálculo de s_{i3}

s_{i2} : não utilizar - pressão diastólica (pDiast): [0, 15] - não usar, é utilizada no cálculo de s_{i3}

s_{i3} : **qualidade da pressão (qPA)**: [-10,10] onde 0 é a qualidade máxima -10 é a pior qualidade quando a pressão está excessivamente baixa, +10 é a pior qualidade quando a pressão está excessivamente alta

s_{i4} : **pulso**: [0,200] bpm

s_{i5} : **respiração**: [0,22] FpM (frequência de respiração)

g_i : não utilizar – valor da gravidade

y_i : **classe de saída**: deve ser inferida pelo classificador

Exemplo:

i	s _{i1}	s _{i2}	s _{i3}	s _{i4}	s _{i5}	g _i	y _i
	pSist	pDiast	qPA	pulso	resp	gravid	classe
1,	8.5806,	2.2791,	-8.4577,	56.8384,	9.2229,	33.5156,	2

Há dois *datasets* disponíveis para treinamento e validação:

1) da tarefa 1 (132 vítimas):

https://github.com/tacla/VictimSim/blob/main/datasets/data_100x80_132vic/sinais_vitais.txt

2) um maior (800 vítimas):

https://github.com/tacla/VictimSim/blob/main/datasets/data_800vic/sinais_vitais_com_label.txt

Arquivo de Teste Cego

O *dataset* para o **teste cego** segue quase o mesmo formato dos dados de treinamento. No entanto, retiramos g_i (o valor de gravidade) e y_1 (a classe). Este arquivo será utilizado somente na fase de teste cego do projeto. O cálculo de y_1 ficará a cargo do classificador implementado no projeto. O valor de g_i não pode ser utilizado pelo classificador.

i	s _{i1}	s _{i2}	s _{i3}	s _{i4}	s _{i5}
	pSist	pDiast	qPA	pulso	resp
1,	8.5806,	2.2791,	-8.4577,	56.8384,	9.2229

7 ENTREGA

A apresentação e a entrega final se darão em dois momentos distintos:

MOMENTO 1: Apresentação - durante a aula

As soluções devem estar implementadas e funcionando para que vocês possam:

- apresentar a solução e mostrar os resultados obtidos para as métricas e saídas impressas que estão na descritas ao longo do documento;
- rodar a solução com os arquivos de teste que serão passados pelo professor no momento 1;
- comparar com as soluções de outras equipes com base nos testes realizados em sala de aula e discutir melhorias;
- não é necessário fazer upload no Moodle.

MOMENTO 2: Entrega - fora do horário de aula (semana seguinte ao do momento 1)

- Corrigir a solução implementada
- Revisar o artigo da tarefa que engloba ambas as partes da tarefa: treinamento e testes
- Carregar a versão final do artigo e do código no Moodle.

7.1 Artefatos da Entrega

- 1) Os códigos fonte na linguagem que desejar com as instruções para rodar.
- 2) Um artigo PDF de até 3 páginas, sem contar os apêndices, no [formato da SBC](#) com a estrutura abaixo e fonte tamanho 11.

Introdução: dentro do problema como um todo, quais subproblemas atacará e por quais razões: quais são as motivações e justificativas para resolvê-los.

Fundamentação Teórica: as técnicas escolhidas com uma breve descrição

Metodologia: descreva como procedeu para avaliar cada uma das técnicas escolhidas, salientando as variações de parametrização e de estrutura (e.g. regras fuzzy). Explicar a razão de tentar uma nova parametrização e/ou estrutura. Explicar como procedeu a validação cruzada (quantas execuções, qual critério de seleção do modelo a ser usado nos testes)

Resultados e análise: mostrar os resultados numéricos das métricas de desempenho para as etapas de treinamento/avaliação e para a etapa de testes.

Conclusões: atingiu os objetivos, a solução de classificação está sobreadaptada ou generalizou bem para o cenário de teste. Há algo a ser melhorado nas soluções apresentadas? Há problemas vieses? Há problemas éticos?

---- 3 páginas no máximo até aqui ----

Apêndices:

- 1. obrigatório:** instruções claras de como executar o código (deve respeitar os formatos de arquivos de entrada e de configuração), print das telas se desejar (não colocar print das telas no corpo do artigo)
- 2. opcionais:** (outras informações, trechos relevantes do código, gráficos, dados, cenários de treinamento).