

UNIVERSIDADE FEDERAL DE VIÇOSA
Campus Florestal

Trabalho Prático III

Valor: 24 pontos

Disciplina: Meta-heurísticas – CCF-480

Curso: Ciência da Computação

Professor: Marcus Henrique Soares Mendes

Data: 10/09/2021

O trabalho é em dupla. A entrega do trabalho será via PVANet Moodle até o dia 15/10/21 às 23:59.

Implementar um Algoritmo Genético (AG) **ou** um algoritmo baseado em Programação Genética (PG) para resolver um problema de classificação que visa inferir regras para saber se um cliente de uma empresa de telecom irá cancelar (churn) sua assinatura ou não. A base de dados¹ utilizada está disponível no PVANet Moodle, bem como um notebook do google colab, o qual faz um pequeno ajuste na base de dados (exclui uma coluna indesejada e 12 registros incompletos).

Os indivíduos **podem** ser representados por regras do tipo se-então. Por exemplo:

Ex 1: se (cliente é casado e cliente tem filho) então classe churn.

Ex 2: se (cliente é aposentado e cliente paga no débito automático ou no cartão de crédito) então classe não churn.

Baseado na representação do indivíduo **deve-se** propor como será o cruzamento e a mutação. A função fitness **deve ser** baseada na acurácia, calculada da seguinte forma:

$$Acurácia = \frac{TP+TN}{TP+TN+FP+FN}$$

Sendo os valores de TP, TN, FP e FN fornecidos pela matriz de confusão abaixo:

		Classe predita pelo indivíduo	
		Churn	Não churn
Classe real do cliente	Churn	TP	FN
	Não churn	FP	TN

Assim, quanto mais próximo de 1, melhor será a acurácia do indivíduo (regra).

Nos experimentos deve-se separar parte dos registros para o conjunto de treinamento (algo entre 70% e 80%) e parte dos indivíduos para o conjunto de teste (algo entre 20% e 30%). A soma dos conjuntos de treinamento e teste deve ser 100%. Quem quiser pode aplicar a

¹¹ Extraída de <https://www.kaggle.com/radmirzosimov/telecom-users-dataset>

validação cruzada com k-fold (cross-validation, mas não é obrigatório). Portanto, cada indivíduo irá possuir um valor de fitness de treinamento e um valor de fitness de teste. Execute o algoritmo proposto 30 vezes de modo independente utilizando uma configuração A e uma configuração B (valores dos parâmetros definidos usando qualquer critério). E baseado no valor final da fitness de treinamento e da fitness de teste retornados pelo melhor indivíduo em cada uma das 30 execuções faça uma tabela que mostre: média, valor mínimo, valor máximo e desvio padrão do valor da função objetivo retornada pelo algoritmo. Mostre também o resultado graficamente com boxplot. **Faça um relatório** que explique como os algoritmos foram implementados (**pode ser feito em qualquer linguagem de programação**), quais foram as configurações A e B utilizadas e como foi feita a representação do indivíduo e as demais etapas do algoritmo. Envie também o código fonte. **Para a melhor solução encontrada para a base de dados considerando o valor da fitness de teste em cada configuração explicita a regra encontrada.** Apresente as seguintes tabelas e discuta os resultados obtidos.

Resultado da fitness de treinamento referente às 30 execuções independentes

Algoritmo	Mínimo	Máximo	Média	Desvio-padrão
AG ou PG configuração A				
AG ou PG configuração B				

Resultado da fitness de teste referente às 30 execuções independentes

Algoritmo	Mínimo	Máximo	Média	Desvio-padrão
AG ou PG configuração A				
AG ou PG configuração B				

Melhor regra encontrada pelo algoritmo com a configuração A (considerar fitness de teste)

--

Melhor regra encontrada pelo algoritmo com a configuração B (considerar fitness de teste)

--

Bom Trabalho!

Leitura sugerida:

Matos, Maicon D. dos S. Aplicação de Algoritmos Genéticos para a Mineração de Regras SE-ENTÃO em Conjuntos de Dados com Distribuição Não-Linear. Dissertação de mestrado em Ciência da Computação na UFU, 2018. Acesso em:

<https://repositorio.ufu.br/bitstream/123456789/20950/1/AplicacaoAlgoritmosGeneticos.pdf>