LVI SBPO SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL
DE 04 ATÉ 07 NOVEMBRO 2024
FORTALEZA | CEARÁ

O uso da integral de Choquet como classificador em dados binários: Resultados preliminares na detecção da COVID-19

Guilherme Dean Pelegrina
Universidade Presbiteriana Mackenzie
Rua da Consolação, 930, CEP 01302-907, São Paulo - SP
guilherme.pelegrina@mackenzie.br

RESUMO

Em diversos problemas envolvendo a agregação de informações complexas, é comum utilizar funções que permitem modelar interações entre atributos. Em problemas de ordenamento, por exemplo, é comum o uso da integral de Choquet como agregador das informações dos critérios. No entanto, embora a integral de Choquet seja muito utilizada em decisão multicritério, há poucos estudos que aplicam tal função para a construção de um classificador. Neste trabalho, consideramos seu uso como classificador em cenários onde os dados são binários. Essa situação é comum em dados médicos, uma vez que muitas informações são do tipo "tem" ou "não tem" um sintoma. O interessante é que esse cenário resulta em um caso particular da integral de Choquet, que facilita o ajuste dos parâmetros e que simplifica a transformação dos mesmos para os chamados valores e índices de interação de Shapley, os quais são amplamente utilizados em aprendizado de máquina para extrair interpretação dos modelos. Basicamente, esses parâmetros indicam tanto o impacto de cada atributo na classificação e quanto efeitos sinérgicos ou redundantes entre eles. A fim de ilustrar o uso da integral de Choquet como classificador, utilizamos uma base de dados acerca da detecção da COVID-19. Nessa base de dados binária, há informações de pessoas que apresentaram diversos sintomas (como febre, tosse, dentre outros) e que realizaram o teste da COVID-19, sendo o resultado obtido positivo ou negativo. A partir do modelo proposto, foi possível compreender quais sintomas estão mais associados à detecção da doença nos indivíduos testados.

PALAVRAS CHAVE. Integral de Choquet, classificação com dados binários, COVID-19.

IC – Inteligência Computacional

ABSTRACT

In several problems involving the aggregation of complex information, it is common to use functions that allow modeling interactions between attributes. In ranking problems, for example, one generally uses the Choquet integral to aggregate the criteria information. However, although the Choquet integral is widely used in multi-criteria decision problems, there are few studies that apply this function as a classifier. In this work, we consider its use as a classifier in scenarios where the dataset is binary. This situation is typical in medical data, since much information is of the type "has" or "does not have" a symptom. An interest in this scenario is that it results in a particular case of the Choquet integral, which makes it easier to adjust the parameters and simplify the transformation from them to the so-called Shapley values and interaction indices, which are widely used in machine learning to extract interpretation from models. Basically, these parameters indicate both the impact of each attribute on the classification and the synergistic or redundant effects between them. In order to illustrate the use of the Choquet integral as a classifier, we used a COVID-19 detection dataset. This binary dataset contains information on people who presented various symptoms (such as fever, cough, among others) and who took the COVID-19 test, with the result being positive or negative. Using the proposed model, we could understand which symptoms are most associated with the detection of this disease in the tested individuals.

KEYWORDS. Choquet integral, classification with binary data, COVID-19.

CI – Computational intelligence