

The 5th Brazilian Competition on Knowledge Discovery in Databases (KDD-BR 2021)

Ana C. Lorena¹, Filipe A. N. Verri¹, Tiago A. Almeida²

¹Divisão de Ciência da Computação (IEC), Instituto Tecnológico de Aeronáutica (ITA)
12228-900 – São José dos Campos – SP – Brazil

²Departamento de Computação (DComp), Universidade Federal de São Carlos (UFSCar)
18052-780 – Sorocaba – SP – Brazil

{aclorena,verri}@ita.br, talmeida@ufscar.br

Abstract. *This editorial paper describes the Brazilian Competition on Knowledge Discovery in Databases Brazilian (KDD-BR) and summarizes the contributions of the three top-ranked solutions obtained in its fifth edition. The 2021 competition involved solving instances of the Travelling Salesman Problem (TSP), of different sizes, using an edge prediction approach.*

Resumo. *Este artigo editorial descreve a Competição Brasileira de Descoberta de Conhecimento em Bancos de Dados (KDD-BR 2021) e resume as contribuições das três melhores soluções obtidas em sua quinta edição. A competição de 2021 envolveu a resolução de instâncias do Problema do Caixeiro Viajante, de diferentes tamanhos, usando uma abordagem de previsão de arestas.*

1. Introduction

The Brazilian competition on Knowledge Discovery in Databases (KDD-BR) was firstly launched in the 2017 joint editions of the Brazilian Conference on Intelligent Systems (BRACIS), the National Meeting of Artificial and Computational Intelligence (ENIAC), the Brazilian Symposium on Databases (SBBD) and the Symposium on Knowledge Discovery, Mining and Learning (KDMiLe), at Uberlândia-MG, Brazil. Since then, the competition has been held annually as a satellite event of BRACIS and ENIAC. At each edition, the top three teams are invited to present their solutions in a special session dedicated to KDD-BR.

The first edition (KDD-BR 2017)¹ involved classifying images captured by one of the monitoring stations of the EXOSS Citizen Science organization, which monitors meteors crossing the southern skies [de Cicco et al. 2018, Lorena et al. 2018]. The images were captured by a simple and low-cost video monitoring camera placed at a monitoring station located in the Observatory of Astronomy and Space Physics of the University of Vale do Paraíba (UNIVAP), at São José dos Campos-SP-Brazil. Each time a moving object is detected by the camera, it starts shooting. Since the interest is to monitor the occurrence of meteors, the competition aimed to build an automated system for classifying the recorded images into two classes: meteor vs non-meteor. A set of features have been extracted from regions of interest of the images and provided to the competitors. A total

¹<https://www.kaggle.com/c/can-i-make-a-wish-detecting-shooting-stars>

of 32 teams with up to five members participated. One of the top-ranked teams published their solution at the ENIAC 2018 edition [Silva et al. 2018].

The second edition (KDD-BR 2018)² was held in partnership with the IBM Research Center in São Paulo, as part of the BRACIS, ENIAC, and KDMile events. The objective was to predict the production of palm oil harvests from the company Agropalma. The dataset had information about palm trees, harvest dates, atmospheric data during plant development, and soil characteristics of cultivated fields. This edition had the highest number of participating teams, 50 teams with up to five members.

In the KDD-BR 2019 edition³, the competition involved predicting the similarity between the partitions produced by the manual clustering of a set of molecular markers and those obtained by an auto-clustering tool, using a total of 100,000 scatter plots provided by the Corteva Agriscience company. A total of 38 teams with up to five members participated. The Corteva Agriscience company provided a monetary prize to the top-three teams.

In KDD-BR 2020⁴, the challenge involved predicting the unavailability of cars in a car rental agency using temporal data provided by the Localiza Hertz company. Car rental rates were offered to the top-three teams. The teams were limited to at most three participants each, and a total of 22 teams competed. All competition, including the presentation session, was held online due to the COVID-19 pandemic restrictions.

This year, the fifth edition of the KDD-BR competition (KDD-BR 2021)⁵ involved solving the traveling salesman problem (TSP) instances through a link prediction approach. The dataset had 100,000 training examples of TSP problem-solution pairs ranging from 50 to 200 city nodes and was provided by the Loggi company. Eight teams with up to three members participated. The top three teams received free registration to the BRACIS 2021 event, a monetary prize by the Loggi company, and an invitation to write a short paper about their solutions, presented in this section of the ENIAC 2021 proceedings.

2. The top-three solutions

The TSP is a classic optimization problem from the literature. However, the top-three solutions are very distinct and reveal the potentialities of different Artificial Intelligence (AI) approaches for solving this problem.

The top-ranked solution was offered by the PaRaNa (Pattern Recognition and Network Analysis) team. They processed the data using a graph-based neural network with edge convolution operators. Then, the results fed a Multilayer Perceptron (MLP) classifier, which predicts the presence or absence of an edge between a given pair of vertices (cities). The solution is described in the paper “*KDD-BR 2021: Using Graph Neural Networks for Link Prediction in TSP problem*”, authored by Bruno Klaus de Aquino Afonso, Willian Dihanster Gomes de Olivera, Jessica Domingues Lamosa and Lilian Berton.

From the Touring Around team, the second-place solution employed an iterated

²<https://www.kaggle.com/c/kddbr-2018>

³<https://www.kaggle.com/c/kddbr-2019>

⁴<https://www.kaggle.com/c/kddbr-2020>

⁵<https://www.kaggle.com/c/kddbr-2021>

local search meta-heuristic adapted for link prediction. The solution was implemented from scratch by Pedro B. Castellucci, who authors the paper “*An iterated local search for the traveling salesman problem*”.

Finally, the third position solution employed a Self-Organizing Map (SOM) neural network with a dedicated hyperparameter tuning procedure. The solution is described in the paper “*Enhanced Self-Organizing Map Solution for the Travelling Salesman Problem*”, authored by João P. A. Dantas, André N. Costa, Marcos R. O. A. Máximo, and Takashi Yoneyama.

3. Acknowledgments

We thank Loggi for the prizes and for providing the dataset for this edition of KDD-BR, particularly to Gabriela Surita, Ângelo Gregório Lovatto, Regiane Nascimento, Tiago Leite, Fillipe Goulart, Juan Camilo Fonseca-Galindo, and José Carlos Maldonado. We also thank the organizers of BRACIS, Reinaldo Bianchi and Zhao Liang, for supporting the competition and providing free registrations for the top-three teams. Moreover, for Leliane Barros for the communication with the Loggi company and support to make the competition possible. Finally, we extend our acknowledgments to the ENIAC chairs, Elaine Ribeiro de Faria Paiva and Paulo de Tarso Guerra Oliveira, for publishing the KDD-BR 2021 short papers as part of the ENIAC proceedings.

Finally, we thank very much all competitors from this and previous KDD-BR editions and organizers of previous editions of the competition which are not part of the current team, namely: Elaine Ribeiro de Faria Paiva (Computer Science Professor at Universidade Federal de Uberlândia – UFU), Ricardo Cerri (Computer Science Professor at Federal University of São Carlos – UFSCar), Vinicius Veloso de Melo (former Computer Science Professor at Universidade Federal de São Paulo – UNIFESP), Daniel dos Santos Kaster (Computer Science Professor at Universidade Estadual de Londrina – UEL), Francisco Carlos Rocha Fernandes (Physics Professor at UNIVAP), Irapuan Rodrigues (Physics Professor at UNIVAP), Jennifer Nielsen (UNIVAP), Luis Fernando Milano Oliveira (UEL), Rodrigo Kuerten (UEL), Renato Luiz de Freitas Cunha (Researcher at IBM Research), Igor Oliveira (Researcher at IBM Research), Eduardo de Paula Costa (Associate Research Scientist at Corteva Agriscience), Ricardo Araújo Rios (Computer Science Professor at Universidade Federal da Bahia – UFBA), Angelo Ciarlini (Data Science Director at Localiza Hertz), Daniel Almeida (Data Scientist at Localiza Hertz), Caio Gomes (Pricing Coordinator at Localiza Hertz), Gustavo Borges (Pricing and Fleet Management Director at Localiza Hertz), João Maciel (Pricing Intern at Localiza Hertz), and Ricardo Stary (Senior Pricing Analyst at Localiza Hertz).

References

- de Cicco, M., Lazzaro, D., Santiago, E. P., and Team, E. (2018). Brazilian video monitoring meteors network exoss: Status and first results—a citizen science project. In *Proceedings of the International Meteor Conference*, pages 162–164.
- Lorena, A. C., Kaster, D. S., Cerri, R., Paiva, E. R. F., and Melo, V. V. (2018). Can i make a wish?: a competition on detecting meteors in images. In *Proceedings of the Symposium on Knowledge Discovery, Mining and Learning (KDMiLe)*, pages 1–8.

Silva, R. M., Lorena, A. C., and Almeida, T. A. (2018). Detecting the presence of meteors in images: new collection and results. In *Anais do XV Encontro Nacional de Inteligência Artificial e Computacional*, pages 128–139. SBC.