


Variant of Northern Bald Ibis Algorithm for Unmasking Outliers

Ravi Kumar Saidala, Acharya Nagarjuna University, Guntur, India

 <https://orcid.org/0000-0002-3658-6651>

ABSTRACT

Clustering, one of the most attractive data analysis concepts in data mining, are frequently used by many researchers for analysing data of variety of real-world applications. It is stated in the literature that traditional clustering methods are trapped in local optima and fail to obtain optimal clusters. This research work gives the design and development of an advanced optimum clustering method for unmasking abnormal entries in the clinical dataset. The basis is the NOA, a recently proposed algorithm, driven by mimicking the migration pattern of Northern Bald Ibis (*Threskiornithidae*) birds. First, we developed the variant of the standard NOA by replacing C_1 and C_2 parameters of NOA with chaotic maps turning it into the VNOA. Later, we utilized the VNOA in the design of a new and advanced clustering method. VNOA is first benchmarked on a 7 unimodal (F_1-F_7) and 6 multimodal (F_8-F_{13}) mathematical functions. We tested the numerical complexity of proposed VNOA-based clustering methods on a clinical dataset. We then compared the obtained graphical and statistical results with well-known algorithms. The superiority of the presented clustering method is evidenced from the simulations and comparisons.

KEYWORDS

Chaotic Maps, Clinical Dataset, Clustering Problems, Nature Inspired Metaheuristics, NOA, NP Hard Problems, Unmasking Outlier

1. INTRODUCTION AND BACKGROUND

Data mining is a subfield of computer science which finds the valuable hidden information in the data. It is the investigations of data for hidden information and the use of software techniques for discovering patterns, regularities and relationships in datasets. It involves methods at the intersection of database systems and mathematics (Aggarwal, 2014). Partitioning the input dataset into meaningful groups is one of the fundamental data analysis concepts of understanding and learning the valuable information. Clustering and classification are the two important concepts in data analysis to extract the useful covered information from the dataset. Clustering is one of the unsupervised machine learning technique and it is defined as the process of identifying the similar groups of data of a dataset. When groups are formed in clustering process, it should be externally isolated and internally interconnected (i.e.) degree of homogeneity within clusters should be high and degree of heterogeneity between clusters also should be high (Han et al., 2011; Kharat et al., 2018; Witten et al., 2016). The key challenge in the data analysis is to categorize patterns or useful hidden knowledge into a set of meaningful groups without the knowledge of class labels (Larose & Larose, 2014). Since the decades, cluster analysis has contributed a vital role in numerous fields such include Artificial Intelligence, Machine Vision and Machine Learning, pattern Recognition, Image Processing etc. (Aggarwal, 2014; Celebi, 2014;

DOI: 10.4018/IJSSCI.2020010102

Celebi & Aydin, 2016). Unmasking outliers in a particular dataset are a pertinent problem in the data mining area. Outlier mining is the process of distinguishing ambient events, irregular or divergent elements, sudden changes in system generated data and exceptions (Campos et al., 2016; Roopa et al., 2020). Outlier analysis is an important data mining issue in knowledge extraction. Clustering based outlier unmasking is one of the best approaches to detect and eliminating such irregular or divergent data elements (Hodge & Austin, 2004). Classical clustering techniques are found inaccurate when data size increases (Nikbakht & Mirvaziri, 2015). Many researchers stated that the clustering performance of the classical clustering techniques can be increased with the help of nature inspired Metaheuristics.

The term optimum means the ultimate ideal or the best. Hence, optimization denotes to find the best solution for the real-world NP hard problems. Optimization is repetitive procedure to obtain the desired outcome to the formulated problem by satisfying all its constraint or bounded conditions (Mirjalili et al., 2017). Nature has always been a key source for the researchers to get inspired for developing many algorithms to solve the real-world complex problems. The algorithms developed by taking inspiration from the nature, can mimic certain phenomena from nature (Mirjalili et al., 2018), referred as Nature Inspired Metaheuristics. Over the past few decades, nature has paved the path to design and develop numerous efficient algorithms to solve complex optimization problems. Recently many Metaheuristic algorithms have gain the researchers attention from the field of pattern recognition and clustering. Many Nature-inspired Metaheuristics have been developed recently and successfully applied for solving complex problems for instance recently invented algorithms Whale Optimization Algorithm (WOA) (Mirjalili & Lewis, 2016), Tornadogenesis optimization algorithm (TOA) (Saidala & Devarakonda, 2017a), New class topper optimization algorithm (NCTOA) (Das et al., 2018), etc., and new versions of these algorithms have also been applied in solving many real-world application problems (Moein, 2014; Esmin et al., 2015; Saidala & Devarakonda, 2017b; Saidala & Devarakonda, 2017c; Li et al., 2015; Saidala & Devarakonda, 2018a; Saidala & Devarakonda, 2019; Lin et al., 2018; Saidala & Devarakonda, 2018b; Saidala & Devarakonda, 2018c; Saidala et al., 2018e; Saidala et al., 2018f). The most popular theorem, i.e., no free lunch theorem (Wolpert & Macready, 1997) justifies that the number of new optimization algorithms needed to come into existence to solve complex problems with low solution cost.

Outlier unmasking can be done in two different ways. They are distance based outlier unmasking (Radovanović et al., 2015) and angle based outlier unmasking (Zhang et al., 2015). Authors of (Alam et al., 2010) proposed outlier detection using PSO (Particle Swarm Optimization) based clustering method. Authors of (İnkaya et al., 2015) have developed ant colony algorithm-based clustering method for solving outlier unmasking problem. Authors of (Saidala & Devarakonda, 2018b) have proposed improved version of WOA for designing optimal clustering method and used it for unmasking outlier in clinical dataset which have different clinical attributes of pregnant woman for instance Hb, HCT, WBCs, RBCs, Platelets, Trimester, etc. As far as we investigated numerous research articles regarding clustering-based outlier unmasking methods, Optimization algorithm-based clustering methods, still there is a room for designing optimal clustering method for outlier unmasking. In this work, a new outlier unmasking approach using Nature inspired Metaheuristic optimization algorithm-based clustering method is proposed.

The structural organization of the remaining paper is framed as follows. The existing approach is presented in Section II. Section III presents the proposed approach where the standard Northern Bald Ibis Algorithm and chaotic maps are presented. The experiments and result analysis have been furnished in Section IV and the last section concludes the whole paper.

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/variant-of-northern-bald-ibis-algorithm-for-unmasking-outliers/250858

Related Content

Evaluating the Effects of Size and Precision of Training Data on ANN Training Performance for the Prediction of Chaotic Time Series Patterns

Lei Zhang (2019). *International Journal of Software Science and Computational Intelligence* (pp. 16-30).

www.irma-international.org/article/evaluating-the-effects-of-size-and-precision-of-training-data-on-ann-training-performance-for-the-prediction-of-chaotic-time-series-patterns/227734

Context-Aware Mobile and Wearable Device Interfaces

Claas Ahlrichs, Hendrik Iben and Michael Lawo (2015). *Recent Advances in Ambient Intelligence and Context-Aware Computing* (pp. 308-321).

www.irma-international.org/chapter/context-aware-mobile-and-wearable-device-interfaces/121782

A PSO-Based Framework for Designing Fuzzy Systems from Noisy Data Set

Satvir Singh, J. S. Saini and Arun Khosla (2012). *Machine Learning Algorithms for Problem Solving in Computational Applications: Intelligent Techniques* (pp. 210-228).

www.irma-international.org/chapter/pso-based-framework-designing-fuzzy/67705

The Formal Design Models of a Set of Abstract Data Types (ADTs)

Yingxu Wang, Xinming Tan, Cyprian F. Ngolah and Philip Sheu (2010). *International Journal of Software Science and Computational Intelligence* (pp. 72-100).

www.irma-international.org/article/formal-design-models-set-abstract/49133

The Kolmogorov Spline Network for Authentication Data Embedding in Images

Pierre-Emmanuel Leni, Yohan D. Fougerolle and Frédéric Truchetet (2013). *Efficiency and Scalability Methods for Computational Intellect* (pp. 96-114).

www.irma-international.org/chapter/kolmogorov-spline-network-authentication-data/76471