Comparison of Normalization Techniques on Data Sets With Outliers

Nazanin Vafaei, Nova University of Lisbon, Portugal*

https://orcid.org/0000-0002-7820-7437

Rita A. Ribeiro, Nova University of Lisbon, Portugal

Luis M. Camarinha-Matos, Nova University of Lisbon, Portugal

https://orcid.org/0000-0003-0594-1961

ABSTRACT

With the fast growth of data-rich systems, dealing with complex decision problems with skewed input data sets and respective outliers is unavoidable. Generally, data skewness refers to a non-uniform distribution in a dataset (i.e., a dataset which contains asymmetries and/or outliers). Normalization is the first step of most multi-criteria decision-making (MCDM) problems to obtain dimensionless data, from heterogeneous input data sets, that enable aggregation of criteria and thereby ranking of alternatives. Therefore, when in the presence of outliers in criteria datasets, finding a suitable normalization technique is of utmost importance. As such, in this work, the authors compare seven normalization techniques (max, max-min, vector, sum, logarithmic, target-based, and fuzzification) on criteria datasets, which contain outliers to analyse their results for MCDM problems. A numerical example illustrates the behaviour of the chosen normalization techniques and an (ongoing) evaluation assessment framework is used to recommend the best normalization technique for this type of criteria.

KEYWORDS

Data Set, Decision Making, Fuzzification, MCDM, Normalization, Outliers, Skewed Data, Target Value

1. INTRODUCTION

Human beings use multi-criteria decision-making methods (sometimes also called multiple attribute decision-making, MADM) in many daily activities to solve decision problems and find an optimum decision, in face of several criteria and alternatives (Zavadskas and Turskis 2010). A Multi-Criteria Decision-Making (MCDM) problem can be defined by a decision matrix, composed of a finite set of alternatives Ai (i=1, …, m), a set of criteria Cj (j=1,…, n), the relative importance of the criteria (or weights) Wj, and the matrix cell elements, rij, representing the rating for alternative i with respect to criteria j (Jahan, Edwards, and Bahraminasab 2016; Triantaphyllou 2000). In most MCDM problems, the used criteria can be expressed either as qualitative or quantitative, usually expressed in different scales, which is an obstacle for the aggregation/ ranking process (Zavadskas and Turskis 2010). Hence, there is a need to use normalization to

*Corresponding Author

prepare dimensionless and comparable criteria values (Jahan et al. 2016; Triantaphyllou 2000; Zavadskas and Turskis 2010). Using different normalization techniques may cause changes on the ranking of alternatives in decision problems, therefore, it is of paramount importance to ensure a proper normalization technique is selected (one objective of this paper).

In recent years, with the advent of data science and data analysing contexts (Chen, Chiang, and Storey 2012), many datasets with outliers emerged (i.e. criterion values skewness), which may greatly influence the aggregation/ranking process. Barnett and Lewis (Barnett and Lewis 1974) defined outlier as "*an observation (or subset of observations) which appears to be inconsistent with the remainder of the data set*". Kennedy et al. (1992) stated that "*an outlier is not an "incorrect" observation but is a realization from a distribution that is in general highly skewed…. One reason for these extreme observations is that some popular variables, such as size, have skewed distributions.*"

So far, there is very little research about the effect of skewed datasets (i.e. criterion values) on decision problems especially from the normalization point of view. To the best of our knowledge there is a big gap in the literature about MCDM methods for selecting suitable normalization techniques especially when there are outliers in the input data.

Therefore, in this study, we discuss the effect of outliers in criteria values and recommend the most suitable normalization technique for MCDM problems that contain skewed criteria values. To this aim we compare seven normalization techniques, using a numerical example that contains outliers in criteria datasets, and use an (on-going) evaluation framework to recommend the best normalization technique.

The major contributions of this study are; (i) addressing the gap in literature about existence of outliers in input data sets in MCDM methods; (ii) comparing the effects of different normalization techniques in MCDM methods with outliers in input data sets; (iii) continue to develop an evaluation assessment framework by adding more metrics to the previous designed frame work (Vafaei et al. 2019; Vafaei, Rita. A. Ribeiro, and Camarinha-Matos 2018); (iv) discussing our contribution with a small illustrative example to exemplify the proposed framework.

This article first presents a brief overview of normalization techniques and assessment frameworks (section 2). Then, it addresses the suitability of the proposed framework for dealing with outliers and choosing the best normalization technique in MCDM problems, using an illustrative example (section 3). Finally, the conclusion and future work on the topic is presented (section 4).

2. OVERVIEW OF NORMALIZATION TECHNIQUES AND ASSESSMENT FRAMEWORK

MCDM methods help decision makers to solve complex decision problems including several criteria with different units. Several MCDM methods are introduced and developed in the literature such as TOPSIS (The Technique for Order of Preference by Similarity to Ideal Solution), SAW (Simple Additive Weighting), AHP (Analytic Hierarchy Process), ANP (Analytic Network Process), ELECTRE (ELimination Et Choix Traduisant la REalité), VIKOR (VIseKriterijumska Optimizacija I Kompromisno Resenje), WASPAS (Weighted Aggregated Sum Product Assessment), PROMETHEE (Preference Ranking Organization METHod for Enrichment of Evaluations) and etc. For more details about MCDM methods please see (Hwang and Kwangsun Yoon 1981; Tzeng and Huang 2011). All methods lead to comparison of the finite alternatives with respect to related criteria for ranking alternatives from the best to worst. However, for ranking alternatives all criteria should be in the same units which is done with normalization techniques (Hwang and Kwangsun. Yoon 1981). So, normalization is a crucial step of most MCDM methods to ensure the input values are between [0-1] (Hwang and Kwangsun. Yoon 1981). Hence below, we discuss some normalization techniques in the context of MCDM methods specifically for skewed data sets/criteria.

Related Content

Successful Implementation of Six Sigma Considering Management Styles
Kouroush Jenaband Selva Staub (2014). *Analytical Approaches to Strategic Decision-Making: Interdisciplinary Considerations (pp. 59-76).*
www.irma-international.org/chapter/successful-implementation-of-six-sigma-considering-management-styles/102150

Is the Character of Institutional Leadership Central to the Quality of Higher Education (HE) Management?
Nwachukwu Prince Ololube (2017). *International Journal of Strategic Decision Sciences (pp. 46-64).*
www.irma-international.org/article/is-the-character-of-institutional-leadership-central-to-the-quality-of-higher-education-he-management/181063

An Empirical Investigation of Extensible Information Sharing in Supply Chains: Going Beyond Dyadic
InduShobha Chengalur-Smithand Peter Duchessi (2017). *Decision Management: Concepts, Methodologies, Tools, and Applications (pp. 1625-1649).*
www.irma-international.org/chapter/an-empirical-investigation-of-extensible-information-sharing-in-supply-chains/176823

Quantitative Concession Behavior Analysis and Prediction for Decision Support in Electronic Negotiations
Réal A. Carbonneau, Rustam Vahidovand Gregory E. Kersten (2014). *International Journal of Decision Support System Technology (pp. 16-30).*
www.irma-international.org/article/quantitative-concession-behavior-analysis-and-prediction-for-decision-support-in-electronic-negotiations/124310

Fuzzy Logic in Multi-Criteria Decision Making: Recent Advances and Applications
C. V. Suresh Babu, K. Bhavanaand M. Samnatha (2025). *Recent Theories and Applications for Multi-Criteria Decision-Making (pp. 399-424).*
www.irma-international.org/chapter/fuzzy-logic-in-multi-criteria-decision-making/361532