Expert Systems with Applications 41 (2014) 4414–4421

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

# The Hellinger distance in Multicriteria Decision Making: An illustration to the TOPSIS and TODIM methods

Rodolfo Lourenzutti [a,*], Renato A. Krohling [b,1]

[a] Graduate Program in Computer Science, PPGI, UFES Federal University of Espirito Santo, Av. Fernando Ferrari, 514, CEP 29075-910 Vitória, Espírito Santo, ES, Brazil
[b] Department of Production Engineering & Graduate Program in Computer Science, PPGI, UFES Federal University of Espirito Santo, Av. Fernando Ferrari, 514, CEP 29075-910 Vitória, Espírito Santo, ES, Brazil

A R T I C L E I N F O

*Keywords:*
Hellinger distance
Multicriteria Decision Making (MCDM)
Probability distribution
Stochastic dominance degree
TOPSIS
TODIM

A B S T R A C T

Due to the difficulty in some situations of expressing the ratings of alternatives as exact real numbers, many well-known methods to support Multicriteria Decision Making (MCDM) have been extended to compute with many types of information. This paper focuses on the information represented as probability distribution. Many of the methods that deal with probability distribution use the concept of stochastic dominance, which imposes very strong restrictions to differentiate two probability distributions, or uses the probability distributions to obtain a quantity that will be used to rank the alternatives. This paper brings the Hellinger distance concept to the MCDM context to assist the models to deal with probability distributions in a direct way without any transformation. Transformations in the data or summary quantities may miss represent the original information. For direct comparisons among probability distributions we use the stochastic dominance degree (SDD). We illustrate how simple it can be to adapt the existing methods to deal with probability distributions through the Hellinger distance and SDD by adapting the TOPSIS and TODIM (an acronym in Portuguese of Interactive and Multicriteria Decision Making) methods.



## 1. Introduction

Multicriteria Decision Making (MCDM) problems have becoming increasingly complicated over the years. It is necessary to consider plenty of alternatives and criteria which make the problems easily overwhelming to human reasoning. Methods to support decision making are now essential and a lot of efforts has been made in the past few decades to advance the field.

The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS), proposed by Hwang and Yoon (1981), is one of the most used methods to support MCDM. The main idea of TOPSIS is that the solution should be as far as possible from the worst possible solution and as close as possible from the best possible solution. This method is quite simple and intuitive, presenting a satisfactory performance in many applications. For a broad survey about the TOPSIS method we refer the interest reader to Behzadian, Otaghsara, Yazdani, and Ignatius (2012).

Despite the importance of the TOPSIS method, there are some other difficulties in the MCDM problems. As shown by Kahneman and Tversky (1979), human thinking is not completely rational presenting strong bias in some situations. For instance, people are more sensitive to losses than they are to gains. In this same work, the Prospect Theory, which describes how individuals make decisions in situations involving risks, was proposed.

In order to consider the human bias in the MCDM methods, Gomes and Lima (1992) proposed the TODIM (an acronym in Portuguese of Interactive and Multicriteria Decision Making) method, one of the first MCDM methods based on the Prospect Theory. The TODIM method makes use of the prospect function to calculate the dominance of one alternative over another. This method has been successfully applied to MCDM problems, e.g., Gomes and Rangel (2009) and Gomes, Rangel, and Maranhão (2009).

In the standard formulations, the TOPSIS and TODIM methods deal with crisp numbers only. This is a serious drawback. Crisp numbers are very precise information and sometimes this desired accuracy is not possible to achieve. Several types of information have been considered in the MCDM models. For example, there are many TOPSIS adaptations to deal with interval numbers (Dymova, Sevastjanov, & Tikhonenko, 2013; Jahanshahloo, Lotfi, & Davoodi, 2009; Yue, 2011), probability distribution information

* Corresponding author. Tel.: +55 27 8135 0081.
*E-mail addresses:* rodolfo.lourenzutti@gmail.com (R. Lourenzutti), krohling.renato@gmail.com (R.A. Krohling).
[1] Prof. Krohling had originally the idea of bringing the Hellinger Distance into Multicriteria Decision Making.

http://dx.doi.org/10.1016/j.eswa.2014.01.015
0957-4174/

(Wentao & Huan, 2010), fuzzy information (Chen, 2000, Krohling & Campanharo, 2011, Lee, Jun, & Chung, 2014), intuitionistic fuzzy information (Boran, Gen, Kurt, & Akay, 2009), interval-valued intuitionistic fuzzy information (Park, Park, Kwun, & Tan, 2011; Tan, 2011; Ye, 2010), and so on. On the other hand, the TODIM method was extended first to deal with fuzzy numbers (Krohling & de Souza, 2012), intuitionistic fuzzy information (Krohling, Pacheco, & Siviero, 2013) and intuitionistic fuzzy information in a random environment (Lourenzutti & Krohling, 2013). Also, a hybrid TODIM was proposed in Fan, Zhang, Chen, and Liu (2013) which deals with crisp numbers, interval-valued numbers and fuzzy numbers at the same time.

A problem that arises with many methods that deal with more complex types of information or which consider heterogeneous information is that they do not use a direct approach to analyze the data in hands. Usually, or they use some heuristics, sometimes unjustified, or they perform a transformation in the dataset. For example, Wentao and Huan (2010) summarize the probability distribution into a confidence interval. Fan et al. (2013) perform transformations in the data types: the interval-valued number is transformed into a uniform distribution and the fuzzy information into a probability distribution. While this may be a helpful approach, it may cause some problems. Transformations in the data or summary quantities may miss represent the original information. For instance, the interval of confidence used in Wentao and Huan (2010) to summarize the probability distribution is centered in the expected value. But what if the expected value does not exist? This is the case for the Cauchy distribution. We could choose infinitely many other confidence intervals. What about for a very asymmetric distribution? In this case the median should be preferred to the mean? And for a multi modal distribution? We choose the mean to center the interval or we choose one of the peaks? In the latter case, which peak should we choose? Also, Dymova et al. (2013) pointed out some problems with many extensions of the TOPSIS method that deal with interval-valued numbers and proposed a direct approach.

In some situations, the ratings of the alternatives with respect to the criteria are best described as probability distributions. This could occur because there are some random factors that could affect the performances or simply because the expert's knowledge is provided in a probability distribution form. For example, if we are evaluating the time of arrival of an elevator in a building, it is natural to express this evaluation as a probability distribution since this time is not deterministic. Also, when analyzing investment projects we do not know exactly how the projects will perform in the future, so in this case probability distributions are very helpful to express the expert's knowledge.

Probability distributions are a complex type of information to work with in MCDM. This is because MCDM methods are always comparing alternatives' ratings to determine which one is preferable, but comparisons between probability distributions are not a trivial task. Given two probability distributions it is hard to state which one is higher/preferable. One of the most used concepts to address this problem is the stochastic dominance (Nowak, 2004a; Nowak, 2004b; Zaras, 2004; Zhang, Fan, & Liu, 2010). However, the stochastic dominance is too restrictive, i.e., given two probability distributions, hardly one of them will dominate the other. Recently, Liu, Fan, and Zhang (2011) proposed a very interesting, intuitive and easy to use method, called stochastic dominance degree. In Liu et al. (2011), the stochastic dominance degree was combined with the PROMETHEE II to rank the alternatives. Their method first calculates the stochastic dominance degree matrices for each criterion; second, these matrices are aggregated into the overall dominance matrix; third, the outgoing flow, entering flow and net flow are calculated using this matrix; finally, the final ranks are determined based on the obtained net flows.

In PROMETHEE II we first calculate the deviation based on pairwise comparisons (the difference of the ratings) and then apply this to a preference function (Behzadian, Kazemzadeh, Albadvi, & Aghdasi, 2010). While the method in Liu et al. (2011) is useful and interesting, they applied the preference function to the stochastic dominance degrees instead of using the expert's information in the original form. Therefore, it is an indirect use of the original information.

As mentioned above, many MCDM methods that deal with probability distributions are based on a too restrictive concept, which is the stochastic dominance, or use an indirect approach by transforming the original information. To avoid some drawbacks of using these approaches (as already mentioned) we want to adapt well-established methods to be capable of dealing with this type of information in a direct way.

In this paper we introduce new versions of the TOPSIS and TODIM methods that are capable of dealing directly with the probability distributions as alternatives' ratings. In order to do so, it is necessary to calculate distances between these probability distributions. The distances will play the role of the difference in the standard formulation. Here we will use the Hellinger distance (Hellinger, 1909). The Hellinger distance is a well established metric for calculating distance between probability distributions and has been broadly applied over the years, making it well tested and reliable.

The fundamental difference, besides the underlying methods, between our approach and the Liu et al. (2011) method is that we will use the stochastic dominance only to determine which alternative is preferable under a certain criterion. The TOPSIS and TODIM method are applied directly to the probability distributions ratings. We emphasize that this paper is not about comparisons between methods. Instead it is about bringing a very useful metric, the Hellinger distance, to the context of MCDM.

The remaining of the paper is organized as follows: In Section 2 we present some necessary concepts. The methods are introduced in Section 3. In Section 4 some applications are discussed. The final remarks are presented in Section 5.

## 2. Preliminary concepts

Consider the problem of selecting one between $m$ alternatives. Each alternative is evaluated with respect to $n$ criteria. Let $A = \{A_1, A_2, \ldots, A_m\}$ be a set with the $m$ alternatives and $C = \{C_1, C_2, \ldots, C_n\}$ be a set with the $n$ criteria. We can summarize the Multicriteria Decision Making (MCDM) problem into the following matrix:

$$DM = \begin{matrix} \\ A_1 \\ A_2 \\ \vdots \\ A_m \end{matrix} \begin{matrix} C_1 & C_2 & \ldots & C_n \\ \left( \begin{matrix} X_{11} & X_{12} & \ldots & X_{1n} \\ X_{21} & X_{22} & \ldots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \ldots & X_{mn} \end{matrix} \right) \end{matrix} \tag{1}$$

where $X_{ij}$ represents the rating of the $i$th alternative evaluated with respect to the $j$th criterion. In this work $X_{ij}$ is considered to be a random variable with distribution $f_{ij}$, denoted by $X_{ij} \sim f_{ij}$. Since the distribution of $X_{ij}$ provides a complete characterization of the random variable $X_{ij}$, in order to simplify the formulation of the problem, the decision matrix will be written in terms of the distributions instead of random variables.

In order to define a method that deals directly with information in the form of probability distributions without the need of any kind of transformation, we need to answer two questions: (i) given two probability distributions which one is higher/preferable? (ii)