Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 61 (2015) 24 – 31

Procedia Computer Science

Complex Adaptive Systems, Publication 5
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2015-San Jose, CA

Analyzing Responses from Likert Surveys and Risk-Adjusted Ranking: A Data Analytics Perspective

Abhijit Gosavi*

Missouri University of Science and Technology, 219 Engineering Management,Rolla, MO 65409, USA

Abstract

We broadly consider the topic of ranking entities from surveys/opinions. Often, numerous ranks from different respondents are available for the same entity, e.g., a candidate from a pool, and yet an averaging of those ranks may not serve the purpose of identifying a consensus candidate. We first consider a risk-adjusted paradigm for ranking, where the rank is defined as the average (mean) rank plus a scalar times the risk in the rank; we use standard deviation as a risk metric. In case of a candidate being ranked either on the basis of opinions of a selection committee's members or on social interactions in a social network such as Facebook, risk-adjusted ranking can result in selecting a consensus candidate who/which does not secure the best average rank, but is acceptable to a large number of the opinion providers. Second, we present an approach to develop the margin of error in Likert surveys, which are increasingly being used in data analytics, where the responses are on a five-point scale, but one is interested in a binary response, e.g., yes-no, agree-disagree. Computing the margin of error in Likert surveys is an open problem.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: risk; ranking; Likert scale; surveys; margin of error; social media

1. Introduction

Ranking of entities is a common activity undertaken when a candidate is to be selected from a pool. Consider for instance the scenario where a number of applicants have applied for the same position, and members from the selection committee rank the different candidates. Ranks provided by the different committee members are then often combined to determine the top candidate(s). Another scenario encountered is that of selecting a site (location) for a new facility, e.g., a manufacturing plant. Different alternative locations are usually evaluated and ranked to form a

* Corresponding author. Tel.: 573-341-4624
E-mail address: gosavia@mst.edu

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2015.09.139

preference. Projects are often ranked for value (societal in case of government projects and monetary for industrial projects) to determine which ones need to be prioritized.

The scenarios described above have been encountered for many years now in industry and government. A more recent application of ranking is that of "friends" in social-media networks, e.g., Facebook, where the selected group of top friends appear on the page of the user; this ranked group may also be used to drive the news feed of the user in the social network. In the Facebook or any other social-media setting, where there are millions of users, such ranking is performed by computer algorithms. Ranking is also an important marketing tool for amazon.com, which provides a Sales Rank to every book, and to Google, which uses the famous PageRank algorithm for ranking web-pages based on traffic. And, finally, before major elections, candidates are ranked based on opinion polls to predict who will win.

In this paper, we study two different aspects of ranking and surveys. The first aspect is that of *risk-adjusted* ranking, where a rank that takes risk into consideration is proposed as an alternative to the *mean rank*; the latter is simply the average of the ranks provided by the different members of the selection committee (or the respondents to a survey). A risk-adjusted rank, we will show, has certain advantages over the mean rank. The other aspect that we will consider here is that of determining binary responses from Likert surveys in which a five-point scale is used to answer a question. Despite the use of the five-point scale, oftentimes, what is needed in practice from the survey is the value of the actual proportion that voted on one side and the proportion that voted on the other and whether there is a statistical tie. We provide a mechanism for calculating the margin of error on such surveys. It is to be noted that in recent times, because of the internet, it is becoming possible to gather voluminous amounts of data, which has led to the birth of the field called data analytics. Tools such a SurveyMonkey (https://www.surveymonkey.com) are becoming increasingly popular for surveying. Further, Likert surveys, where the response is on a five-point scale, are commonly used in industry.

2. Risk, Ranking, and Surveys: A Short Review

Risk is a widely studied topic—especially from the viewpoint of decision-making. While risk has many definitions and dimensions, the aspect we are interested in here is that of uncertainty or variability in data. In other words, data oftentimes do not provide concrete answers in the form or "yes" or "no." Another aspect that interests us here is that of ranking, where there is difference of opinion amongst respondents (or rank-givers/rating-providers)—leading to grey areas. Under these circumstances, one can use statistical properties of the data to reveal whether the data says "yes" or "no," or in the case of ranks, which individual is ranked highest, who is ranked second and so on. We further note that at times, the responses cannot be described in the yes-no format, but as like-dislike etc. Nonetheless, from the five-point Likert survey, it is usually possible to extract a $(-1, +1)$ pattern. We now review a subset of contributions to the literature on ranking.

Adler et al. [1] provide an excellent review of ranking methods in the context of data envelopment analysis (DEA). The Delphi method [2] is a widely used method that uses stages in ranking to obtain consensus. The analytic hierarchy process (AHP) is also widely used in developing ranks. Schmidt [3] discusses the use of the Delphi method via non-parametric statistical techniques. Layton [4] discusses a random coefficients model, and also provides a discussion in the context of rank-ordered logit model. Fuzzy set theory has also been used widely in ranking when human subjectivity is involved [5]. A comparison of ratings and rankings has been performed in [6]. Langville and Meyer [7] provide a detailed analysis of different rating methods.

3. Risk-Based Ranking of Surveys

We first present the theory underlying risk-based ranking and then provide a numerical example to illustrate the concept.

3.1. Risk-Adjusted Mean

We first motivate the need to modify the mean rank, which is often used in project management for ranking government transportation projects [8], where the rank from each respondent is a weighted mean of some ranks. For instance, it is common to develop a series of weights for characteristics on which the candidate is ranked, and then a weighted sum of those ranks is provided as the mean rank from the respondent. To be more formal, let $w(i)$ denote the