Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2015) 000–7

Procedia Computer Science

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2015 - San Francisco, 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.

Keywords: risk; ranking; Likert scale; surveys; margin of error; risk, 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

*Tel: 573-341-4624; Email: gosavia@mst.edu

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 weight associated to the ith characteristic (where examples of characteristics are viability, societal impact, long-term value etc), where there are I different characteristics and C candidates. Then, if $r_c(i)$ denotes the rank assigned to the ith characteristic for the cth candidate, where we have C candidates, then the average rating for a given candidate, c, is given by:

$$R_c = \sum_{i=1}^{I} w(i) r_c(i),$$

in which $0 \leq w(i) \leq 1$ for every i and $\sum_{i=1}^{I} w(i) = 1$. Further note that typically each candidate is evaluated by more than one evaluator or respondent. In order to account for the respondent, we will further modify our notation by using the index j for the jth respondent. Then, let R_{cj} denote the mean rank assigned to the cth candidate by the jth respondent, where we have J respondents. The overall mean rank for the cth candidate will then be given by:

$$\mu_c = \frac{\sum_{j=1}^{J} R_{cj}}{J},$$

assuming that the opinion of each evaluator is weighed equally.

The mean rank, defined above, does not take into account the variability in the ranks provided by the different respondents. For instance, the candidate with the best mean rank may be one who (or which) received quite a few poor ranks from some respondents. It may be possible that a candidate with a worse mean rank than the best actually displays lower variability, and hence may be more acceptable to many respondents. If one is interested in building consensus, it can be useful to go down the order of mean ranks and select a candidate who is viewed favorably by many but is not at the top of the lists of the majority. In other words, we may be interested in selecting a candidate who is acceptable to many but may not be the top choice overall. This situation is commonly experienced when the candidate with the best mean rank (top candidate) is ranked at the bottom by some. Those respondents, who ranked this candidate at the bottom, may be more accepting of a candidate who does not get the best mean rank but is quite high on their own lists. In order to identify such a rank order, we propose the rank-adjusted mean approach below.

The standard deviation of the rank for the cth candidate will be given by σ_c, where

$$\sigma_c^2 = \frac{\sum_{j=1}^{J} \left(R_{cj} - \mu_c\right)^2}{J-1}.$$

The risk-adjusted mean (RAM) for the cth candidate will then be defined as:

$$RAM_c = \mu_c + \theta\sigma_c \tag{1}$$

where θ is a small positive constant, usually in the range $(0.1, 0.3)$. The term σ_c is a measure of the risk/variability in the rank provided by the different respondents, while $\theta\sigma_c$ is the risk-adjusting term. We will assume that the lower the rank, the higher the preference, and our ranking mechanism in Equation (1) works under this convention. Note that if a higher rank is preferred, then the risk-adjusted mean would be as follows:

$$RAM_c = \mu_c - \theta\sigma_c.$$

The underlying idea here is that the risk-adjusting term should *penalize* the mean rank, and hence if higher ranks are preferred, the risk-adjusting term will lower the rank, while if lower ranks are preferred, the risk-adjusting term will increase the rank. The increase or decrease will be in proportion to the risk/variability in the rank. Thus, what we propose here is to use RAM, defined in Equation (1), instead of the mean rank to define the rank order. It should be noted here that a candidate whose rank shows a lot of volatility/risk/variability, which we measure via the standard deviation, will be penalized in our ranking system.

Obviously, the value of θ can influence the rank order obtained via RAM and must be chosen suitably. A value too small may produce an order which is identical to that produced by the mean, while too large a value may produce a list where a candidate who is ranked uniformly poorly gets to the top of the list.

Another advantage of using the standard deviation is that it often allows us to distinguish between two or more candidates that have the same mean rank. Using RAM, if more than one candidate has the same mean rank, the one with the least standard deviation would be given a higher preference and so on.

At this point, it is necessary to point out one deficiency of adjusting for risk: the candidate who will emerge at the top of our RAM ranking mechanism may *not* be the top candidate for *any* respondent; it is somebody to whom no one objects to much. This is an unfortunate consequence of trying to seek consensus, which was vividly described by the late Margaret Thatcher in the following words [9] as: "the process (consensus) of abandoning all beliefs, principles, values, and policies in search of something in which no one believes, but to which no one objects..." At the same time, it must be noted that in the real world, many decisions have to be made on the basis of consensus, and our model is designed to help in that direction. The concept of Nash equilibrium [10] also revolves around the idea that often parties settle down on a solution that is optimal for neither party, and yet no party is willing to move away from it.

3.2. Numerical Results

We now provide a simple example to illustrate the idea discussed above. Consider the ranks provided by eight respondents on ten candidates ($C = 10$ and $J = 8$) in a pool in Table 1. The lower the rank the higher the preference, i.e., the top candidate is ranked 1, and the second is ranked 2, and so on. Ranks do not have to be unique; in case, n candidates are given the same rank, the next $(n - 1)$ ranks are skipped. Table 2 shows the mean rank, the standard deviation, the RAM value, and the rank orders using the mean and RAM. We use $\theta = 0.2$ in the calculations below. Note that the orders produced by the two are not identical. The top rank via mean is ranked second on the RAM list, while the top rank on the RAM list is ranked second on the other list. This is due to the difference in the standard deviations. Further note that the rank order list of the mean has many ties which are broken in the list produced by RAM. This kind or ranking is also helpful in screening from a large pool, where ties can create problems.

Table 1. Ranks for 10 candidates from 8 respondents: Resp denotes Respondent #

Candidate	*Resp 1*	*Resp 2*	*Resp 3*	*Resp 4*	*Resp 5*	*Resp 6*	*Resp 7*	*Resp 8*
1	9	6	5	2	6	6	6	1
2	3	3	1	3	1	1	1	5
3	2	2	2	1	3	2	3	3
4	1	4	3	5	4	4	2	2
5	4	5	4	4	5	7	4	4
6	7	8	8	8	7	3	7	6
7	8	10	9	7	8	8	5	7
8	6	9	7	9	2	10	10	8
9	5	7	10	10	10	9	9	10
10	10	1	6	6	9	5	8	9

Table 2. Mean and standard deviations of the ranks, as well as the rank orders by the mean and RAM

Candidate	*Mean Rank*	*Standard Deviation*	*RAM*	*Rank order by Mean*	*Rank Order by RAM*
1	4.666666667	2.738612788	5.214389224	4	5
2	2.222222222	1.394433378	2.501108898	1	2
3	2.333333333	0.707106781	2.47475469	2	1
4	3.222222222	1.301708279	3.482563878	3	3
5	4.666666667	1	4.866666667	4	4
6	6.666666667	1.58113883	6.982894433	6	6
7	7.666666667	1.414213562	7.949509379	8	8
8	7.666666667	2.5	8.166666667	8	9
9	8.777777778	1.715938357	9.120965449	9	10
10	7.111111111	2.934469477	7.698005006	7	7

We now provide a simple example from a social-media setting. Let $X(i, j, t)$ be the number of interactions between friends i and j for the time slice t. A time slice could be a week, for instance, and thus $X(i, j, t)$ would count the total number of interactions that occur during the week. A social network is likely to weigh different interactions using dissimilar weights, e.g., likes, comments, replies, checking profile, and private messages are all some kind of interactions, but private messages are likely to have higher weights than a simple like. We will assume that $X(i, j, t)$ is the *weighted sum* of the interactions that occur during the time slice t. Let us further assume that the history of the last T time slices is used to generate the rank of friends, which as we noted above, is also used to drive the news feed. Let N_i denote the number of friends for user i. Then, if we use the mean rank, we would use the following quantity to determine the rank for all friends of user i:

$$Y(i) = \frac{\sum_{j=1;j\neq i}^{N_i} \sum_{t=1}^{T} X(i, j, t)}{T}.$$

Thus, for instance, consider the simple example in which user 1 has four friends, $T = 3$, and the values of $X(.,.,.)$ are as follows:

$$\text{Interactions with 2: } X(1, 2, 1) = 12; X(1, 2, 2) = 12; X(1, 2, 3) = 11;$$

$$\text{Interactions with 3: } X(1, 3, 1) = 36; X(1, 3, 2) = 17; X(1, 3, 3) = 1.0;$$

$$\text{Interactions with 4: } X(1, 4, 1) = 14; X(1, 4, 2) = 18; X(1, 4, 3) = 18.5;$$

$$\text{Interactions with 5: } X(1, 5, 1) = 7.0; X(1, 5, 2) = 4.0; X(1, 5, 3) = 3.0.$$

From the above, we have that:

$$Y(1) = 11.66666667; Y(2) = 18; Y(3) = 16.83333333; Y(4) = 4.666666667.$$

If we wish to select the top friend of user 1 on the basis of the mean, it would clearly be user 2 since the mean interactions with 2 are the most. We now look at the variability in the values, and define the standard deviation as follows:

$$\sigma(i) = \sqrt{\frac{\sum_{j=1;j\neq i}^{N_i} \sum_{t=1}^{T} (X(i, j, t) - Y(i))^2}{T - 1}}.$$

For the example above, the standard deviations are:

$$\sigma(1) = 0.577350269; \sigma(2) = 17.52141547; \sigma(3) = 2.466441431; \sigma(4) = 2.081665999.$$

The risk-adjusted score for user i would then be:

$$RAM(i) = Y(i) - \theta\sigma(i),$$

the values of which are: $RAM(1) = 11.55119661; RAM(2) = 14.49571691; RAM(3) = 16.34004505; RAM(4) = 4.250333467$. Clearly, the top friend based on the RAM value will be user 3; this is due to the fact that although the mean value of user 3 is lower, because of its lower standard deviation, user 3 acquires a higher RAM score than user 2.

Ranking the friends using the risk-adjusted score will lead to a more *stable* ranking system that is less susceptible to sudden fluctuations in interactions. In other words, friends who *consistently* interact with each other will maintain a higher rank. This is important in a social media setting where the social network is interested in keeping users engaged, since the network's revenues are proportional to the time spent in engagement and the news feed, which affects the degree of engagement, is driven by these ranks. A user who *regularly* interacts with others will be interested in seeing them often in his/her own news feed, and is hence a risk-adjusted ranking mechanism is likely to be of value.

4. A Binary Response from Likert Surveys

Likert surveys, as discussed above, provide a 5-point scale. They are used widely, and although the five-point scale is used because respondents are often not comfortable with providing a solid yes or no (or like/dislike), but like shades in their answers. And yet one is often interested in determining how many are on one side and how many are on the other, or if there is a statistical tie. We now provide a mechanism, rooted in statistics, to help answer this question. Our work here is a part of our ongoing study [11], but the test shown here was originally developed and used by the author for analyzing a pre-election survey for an Indian business magazine Moneylife [12].

4.1. Statistical Test

We will employ the binomial distribution to develop a test of statistical significance. For categorial data, this test is used widely when the data is of a binary nature. Usually, the Likert survey asks a question, and the respondents either strongly agree, agree, are neutral, disagree or strongly disagree. Thus, in other words, there are shades of yes and no, where if yes goes with agree (disagree), then no goes with disagree (agree).

For the sake of our analysis, those who agree and strongly agree will be combined into one group, and the respondents who disagree and strongly disagree will be merged into another group. And then we have the neutral responses, and the critical question is: what should one do with them? We will address this via our test.

We now need to discuss how we can extract the binomial distribution here. The two groups will be treated as two different responses, and the proportions of respondents in each group will be computed, which allow us to use the binomial distribution. It turns out that for measuring proportions of populations, we can use the normal distribution as an approximation for the binomial under certain conditions that we describe later.

Let p_i denote the estimated proportion of population belonging to Group i, and L_i denote the number in the ith group. Note that we do not include the neutral responses at this stage in any group. Here i can take only 2 values: 1 and 2. Let $n = L_1 + L_2$ denote the total number of respondents. Then, the estimated proportion of the ith group should be in %:

$$p_i = \frac{L_i}{n} \times 100.$$

It is the *margin of error* (in the survey) that can be estimated via the normal approximation to the binomial distribution. The margin of error is a concept used widely in analyzing surveys. It helps determine if the survey shows a clear winner in case of two choices (binary situation) or if there is a statistical tie. A statistical tie indicates that although proportion for one choice is higher than that of the other, because of statistical variation, one cannot draw definite conclusions about who the winner is.

The standard error for the binomial distribution is [13]:

$$\sqrt{\frac{p_i(1-p_i)}{n}}.$$

If α is the level of significance (probability of rejecting the null hypothesis when it is true), then the $100(1-\alpha)$ % margin of error is given by

$$\text{margin of error } = z_{\alpha/2}\sqrt{\frac{p_i(1-p_i)}{n}} \times 100. \tag{2}$$

This yields the following confidence interval for p_i:

$$\left(p_i - z_{\alpha/2}\sqrt{\frac{p_i(1-p_i)}{n}}, p_i + z_{\alpha/2}\sqrt{\frac{p_i(1-p_i)}{n}}\right) \times 100. \tag{3}$$

We will number the groups such that $L_1 \geq L_2$. The hypothesis that needs to be tested here is:

- $H_0 : p_1 \leq p_2$
- $H_1 : p_1 > p_2$

We use a step-by-step procedure to test the above hypothesis. The first stage of this test is a preliminary one devised to determine if there is any possibility of detecting conclusive information from the data. If the data fails this test, no further computations are necessary.

Stage I: Ignore the neutral responses and compute L_1 and L_2. Without loss of generality, let us assume that $L_1 \geq L_2$ (if it turns out that $L_1 < L_2$, then reverse the numbering of the groups, i.e., rename the original Group 1 as Group 2 and Group 2 as Group 1). Add the neutral observations to Group 2 and recompute L_2. If $L_2 \geq L_1$, no conclusions about the statistical winner can be drawn from the test, and there is no need to proceed to Stage II.

Stage II:

Step 0: Ignore the neutral responses and compute L_1 as the number of respondents in Group 1 and L_2 as the same in Group 2. Set a value for α.

Step 1: Merge the neutral responses with Group 1 to recompute L_1. Compute the confidence intervals for p_1 and p_2 using Equation (3). Determine if the confidence intervals overlap. If they overlap, go to Step 3. If they do *not* overlap, go to Step 2.

Step 2: Merge the neutral responses with Group 2 and recompute L_1 and L_2. Re-compute the confidence intervals using the new values of L_1 and L_2. If the confidence intervals overlap, go to Step 3. If the confidence intervals do *not* overlap, then one can *reject* the null hypothesis, and we are done with the test here.

Step 3: We *cannot reject* the null hypothesis, and we are done with the test here.

The principle underlying the above is as follows: Let us refer to Group 1 as the "larger" group, since it has more respondents (ignoring the neutral respondents). (Of course, larger does not mean it is the statistical winner, and our test is designed to determine if there is one.) Now if the neutral observers are added to the larger group, it will become even larger, and hence we do not perform this calculation. We will first see what happens when we add the neutral observers to the smaller group, Group 2. If that makes Group 2 larger, the data will not be able to provide any conclusions either way, and there is no need to got to the next stage.

If the first stage is cleared, we will add the neutral observers to Group 1 and compute confidence intervals. At this time, if the confidence intervals overlap, then it means that we have a statistical tie and therefore we cannot reject the null hypothesis. If they do not overlap, we then combine the neutral with the smaller group (Group 2) and recompute the confidence intervals. At this point, if the confidence intervals do not overlap once again, then we can reject the null hypothesis and declare Group 1 to be the winner in a statistical sense. If the confidence intervals overlap, on the other hand on either occasion, our conclusion will be that we cannot reject the null hypothesis. We illustrate these ideas with examples in the next subsection. We now present the two conditions that need to be met in order to use the normal approximation to the binomial [13]:

- The sample size is large enough, e.g., about 100 or more, but is a small proportion of the population.
- The samples are drawn randomly from the population and are spread out within the population.

4.2. Numerical Results

We now provide examples to illustrate how the test above can be used. Each example is designed to work for the different scenarios that can emerge in a Likert survey.

Example 1: Strongly Agree = 20; Agree = 28; Neutral = 8; Disagree = 14; Strongly Disagree = 30

Here, if we combine those who strongly agree and agree into Group 1 and those who disagree and strongly disagree into Group 2, we have $L_1 = 48$ and $L_2 = 44$. In Stage I, combining the neutral with Group 2, we have that $L_2 = 44 + 8 = 52 > L_1$, and thus the data fails in Stage I, i.e., we cannot derive any definitive conclusions here.

Example 2: Strongly Agree = 40; Agree = 28; Neutral = 8; Disagree = 14; Strongly Disagree = 10

Combining strongly agree and agree observations into Group 1 and the disagree and strongly disagree observations into Group 2, we have $L_1 = 68$ and $L_2 = 24$. Here, the first stage will be cleared since $68 > 8 + 24$. We now go to Step 1 of Stage II in which $L_1 = 68 + 8 = 76$ and $L_2 = 24$, which leads to $p_1 = 76\%$ and $p_2 = 24\%$. Using $\alpha = 0.05$, we have the margin of error via Equation (2) as 8.3707% and the confidence intervals do not overlap. Then, we add the neutral observations to Group 2, to obtain $L_1 = 68$ and $L_2 = 32$, which leads to $p_1 = 68\%$ and $p_2 = 32\%$. Using $\alpha = 0.05$, we have the margin of error via Equation (2) as 9.1428% and the confidence intervals again do not overlap. This implies that we can reject the null hypothesis, and that at α level of confidence, we can claim that the number of those who agree and strongly agree exceeds the others.

Example 3: Strongly Agree = 20; Agree = 33; Neutral = 2; Disagree = 20; Strongly Disagree = 25

We first combine the strongly agree and agree into Group 1 and the disagree and strongly disagree into Group 2, we have $L_1 = 53$ and $L_2 = 45$. Here, the first stage will be cleared since $53 > 2 + 45$. We now go to Step 1 of Stage II in which $L_1 = 53 + 2 = 55$ and $L_2 = 45$, which leads to $p_1 = 55\%$ and $p_2 = 45\%$. Using $\alpha = 0.05$, we have the margin of error via Equation (2) as 9.7507% and the confidence intervals overlap. Hence, we cannot reject the null hypothesis, and no definitive conclusions can be drawn.

Example 4: Strongly Agree = 20; Agree = 38; Neutral = 2; Disagree = 20; Strongly Disagree = 20

Combining strongly agree and agree into Group 1 and disagree and strongly disagree into Group 2, we have $L_1 = 58$ and $L_2 = 40$. The first stage will be cleared since $58 > 2 + 40$. We now go to Step 1 of Stage II in which $L_1 = 58 + 2 = 60$ and $L_2 = 40$, which leads to $p_1 = 60\%$ and $p_2 = 40\%$. Using $\alpha = 0.05$, we have the margin of error via Equation (2) as 9.6018% and the confidence intervals do not overlap. Then, we add the neutral observations to Group 2, to obtain $L_1 = 58$ and $L_2 = 42$, which leads to $p_1 = 58\%$ and $p_2 = 42\%$. Using $\alpha = 0.05$, we have the margin of error via Equation (2) as 9.6736% and the confidence intervals do overlap. Hence, we cannot reject the null hypothesis, and no definitive conclusions can be drawn.

Although in three out of the four examples above, no definitive conclusions were drawn, we note that those three examples were intentionally created to demonstrate the different scenarios that might arise in the test leading to situations when definitive conclusions cannot be drawn. In practice, unless the data are within the margin of error, the test results will be of the form shown in Example 2.

5. Conclusions and Future Work

Likert surveys are being used widely and yet there seems to be a lack of statistical tools that can be used to draw conclusions from them. They are being routinely used via SurveyMonkey, as the business of online surveys has magnified enormously in this decade. The test we developed is a part of our ongoing work to better understand data from Likert surveys. In the future, we expect to perform empirical work on larger data sets. The model we presented for risk-adjusted ranking is inspired by a similar model we have used in artificial intelligence/reinforcement learning [14, 15] for capturing risk in decision-making; our risk-adjusted model here is proposed to help in arriving at a consensus in candidate selection. Future work on this topic will involve other metrics of risk, and a comparison of the behavior of the resulting algorithms.

References

[1] N. Adler, L. Friedman, Z. Sinuany-Stern, Review of ranking methods in the data envelopment analysis context, European Journal of Operational Research 140 (2002) 249–265.

[2] N. Dalkey, O. Helmer, An experimental application of the Delphi method to the use of experts, Management Science 9 (3) (1963) 458–467.

[3] R. Schmidt, Managing Delphi surveys using nonparametric statistical techniques, Decision Sciences 28 (3) (1997) 763–774.

[4] D. Layton, Random coefficient models for stated preference surveys, Journal of Environmental Economics and Management 40 (1) (2000) 21–36.

[5] P.-T. Chang, E. Lee, Ranking of fuzzy sets based on the concept of existence, Computers & Mathematics with Applications 27 (9) (1994) 1–21.

[6] D. Alwin, J. Krosnick, The measurement of values in surveys: A comparison of ratings and rankings, Public Opinion Quarterly 49 (4) (1985) 535–552.

[7] A. Langville, C. Meyer, Who's #1?, Princeton University Press, 2012.

[8] C.-W. Su, M.-Y. Cheng, F.-B. Lin, Simulation-enhanced approach for ranking major transport projects, Journal of Civil Engineering and Management XII(4) (2006) 285–291.

[9] M. Thatcher, Speech at Monash University (Sir Robert Menzies Lecture), Melbourne, Australia, October 6, 1981 http://www.margaretthatcher.org/document/104712.

[10] J. Nash, Equilibrium points in n-person games, Proceedings, Nat. Acad. of Science, USA 36 (1950) 48–49.

[11] L. Blenke, A. Gosavi, W. Daughton, Attitudes towards face-to-face meetings in virtual engineering teams: Perceptions from a survey of defense projects, Working paper at Missouri Univeristy of Science and Technology, Department of Engineering Management and Systems Engineering, 2015.

[12] M. D. Team, General Elections 2014: The Concerns and Sentiments of Urban Voters, MoneyLife Magazine Online (Edts Sucheta Dalal and Debashis Basu), September, 17 (2013).

[13] R. Johnson, G. Bhattacharyya, Statistics: Principle and Methods, 7th Edition, Wiley, 2014.

[14] A. Gosavi, Variance-penalized Markov decision processes: Dynamic programming and reinforcement learning techniques, International Journal of General Systems 43 (6) (2014) 649–669.

[15] A. Gosavi, A risk-sensitive approach to total productive maintenance, Automatica 42 (2006) 1321–1330.