

Young People’s Priorities: An Illustrative Example of Scaling Skew-Symmetric Data

Ehsan Bokhari Lawrence Hubert

Abstract

A scaling procedure for skew-symmetric data is illustrated using results from a survey by the Institute of Politics at Harvard University. These data were discussed in an April 27, 2012 op-ed column from the *New York Times* by Charles Blow.

The op-ed columnist for the *New York Times*, Charles Blow, is known for his creative graphics and visual representation of data. One of his more recent columns, entitled “Young People’s Priorities” (Blow, 2012), presented two tables of percentages representing young people’s views on a number of real-world issues. One table was labeled *Domestic Affairs* and the second *Foreign Affairs*. These data were collected by the Institute of Politics (IOP) at Harvard University (Harvard University, 2012); the respondents were between the ages of 18 and 29. This 21st edition released in 2012 is available on their website. The two tables presented by Mr. Blow were selected from the larger IOP survey that compared all of the issues together.

The display of the raw data in table form and the information is certainly helpful for interpretation. But however informative, there is more that can be done to add clarity. For example, one might note the scattering of the three colors representing

preferences, but there is no apparent ordering of the issues (in fact, they are just presented in the order provided by the IOP). If readers want to know the most important issue among the respondents, they must sift through the information provided. This may not be that difficult in some cases. For instance, in the *Domestic Affairs* table it is fairly obvious that “Creating jobs and lowering the unemployment rate” is the most important item for the young people surveyed. This is clear because the entire row for this issue is green (green indicates the issue in the row is favored “significantly” more than the issue in the column). Such relationships may not always be so easily detectable; in the *Foreign Affairs* table there is no single issue that is entirely green. To determine which is the most important, the reader must go deeper into the information provided. And what about the next most important issue? That is even more difficult to determine, and it is more than just simply tallying up the number of times an issue was rated higher. These tables would be more informative if the rows and columns followed some kind of rank ordering. This is where methods of scaling can be of assistance.

1 A method for reordering and scaling skew-symmetric data

As mentioned, each cell in Mr. Blow’s tables displays a number representing the proportion of respondents rating the issue in the row more important than the issue in the column. In other words, these data are proximities indicating how similar two issues are in terms of importance. If $\mathbf{D}_{N \times N} = \{d_{ij}\}$ represents the complete table in matrix form among these N issues, d_{ij} pro-

vides proximity information for the issue in row i and the issue in column j . Furthermore, because these entries, d_{ij} , are described as the respondents' rating of issue i over issue j in terms of importance, the data in \mathbf{D} are asymmetric (and nonnegative). This asymmetric matrix can be decomposed as

$$\mathbf{D} = \frac{1}{2} [\mathbf{D} + \mathbf{D}'] + \frac{1}{2} [\mathbf{D} - \mathbf{D}'],$$

where $\mathbf{P} \equiv \frac{1}{2} [\mathbf{D} + \mathbf{D}']$ is an $N \times N$ nonnegative symmetric proximity matrix and $\mathbf{P}^{(s)} \equiv \frac{1}{2} [\mathbf{D} - \mathbf{D}']$ is an $N \times N$ skew-symmetric proximity matrix such that $p_{ij}^{(s)} = -p_{ji}^{(s)}$ for all $1 \leq i, j \leq N$. There are several items of information contained in this latter matrix worth noting. First, there is a directionality component with respect to the dominance relationship between the two objects (or in our case, issues). Thus, if $p_{ij}^{(s)}$ is positive (or equivalently, if $p_{ji}^{(s)}$ is negative), then issue i is rated more important than issue j . Second, the magnitude of the dominance relationship (i.e., the absolute value, $|p_{ij}^{(s)}|$) provides an indication of how often issue i is rated more important than issue j (assuming $p_{ij}^{(s)} > 0$; if not, then $|p_{ij}^{(s)}|$ represents how *less* often issue i is rated more important than issue j). (For a thorough discussion of methods for modeling asymmetric proximities, the reader is referred to Zielman and Heiser, 1996.)

Given this skew-symmetric matrix, $\mathbf{P}^{(s)}$, there is a least squares solution for determining a coordinate representation (i.e., a placement of the issues along a line that also induces a rank ordering) that can be given by a simple formula (see, for example, Hubert, Arabie, and Meulman, 2001, pp. 62–66). For the given objects, the set of N coordinates x_1, \dots, x_N that minimizes the least-

squares criterion

$$\sum_{i < j} \left(p_{ij}^{(s)} - (x_j - x_i) \right)^2,$$

is obtained by averaging the proximities within columns, $x_j = (1/N) \sum_i p_{ij}^{(s)}$; these coordinate values give a minimal loss value of $\sum_{i < j} \left(p_{ij}^{(s)} \right) - (1/N) \sum_j \left(\sum_i p_{ij}^{(s)} \right)$.

Because the sum of the entries in $\mathbf{P}^{(s)}$ is zero, the variance-accounted-for (VAF) can be expressed as

$$\frac{(1/N) \sum_j \left(\sum_i p_{ij}^{(s)} \right)^2}{\sum_{i < j} \left(p_{ij}^{(s)} \right)^2}.$$

If we want to consider a coordinate representation that is only equally-spaced, the loss function is adjusted as follows:

$$\sum_{i < j} \left(p_{ij}^{(s)} - \alpha(x_j - x_i) \right)^2,$$

where x_1, \dots, x_N are now the integers $1, \dots, N$ (not necessarily ordered) and α is an estimated constant. In both scenarios, the optimal reordering of the objects (issues) is found by ordering the coordinates from smallest (most negative) to largest (most positive).

2 Results

We now demonstrate this technique with the IOP dataset that contains the twenty issues that Mr. Blow separated as *Domestic Affairs* and *Foreign Affairs*. All analyses were done using the MATLAB software and an M-file called

`skew_symmetric_scaling.m`

is available for download at

http://cda.psych.uiuc.edu/skew_symmetric_scaling

In addition to the MATLAB M-file, an equivalent function file for the open-source statistical software R is also given at this web site.

Figure 1 displays the twenty issues plotted with the vertical axis representing the coordinates. The two types of issues are separated, with the foreign issues being on the left. The issues are in order of importance with the most important on top and the least important on the bottom. The distance between issues is representative of how much those two items differ in terms of importance. Figure 1 displays more information about the relationships between the issues than can the raw tables presented by Mr. Blow.

The VAF for this scaling solution is 90.2%. It is important to note that if this table was separated into two groups, domestic and foreign, as done by Mr. Blow, the ordering may not hold within each group (and, in fact, does not). Remember that these orders are based on minimizing a loss function taking into account all issues, so that some of the larger p_{ij} 's may not be included. Although we should not expect the ranking within each group to be exactly the same, they should be similar. As an illustration, consider the two issues of “Reducing the federal deficit” and “Lowering the tax burden for all Americans.” The first issue (“Reducing the federal deficit”) was rated more important than all of the foreign affairs issues. In addition, these proportions were larger than the proportions for the second issue (“Lowering the tax burden for all Americans”) for six of the eight foreign affairs items (the issue “Preventing the

spread of terrorism” was actually rated as more important than this second issue). When looking at the two issues compared to the other domestic affairs issues, however, we see the second item is rated as more important than nine of the eleven others whereas the first issue is only rated more important than six of the eleven domestic issues. Figure 2 (*Domestic Affairs*) and Figure 3 (*Foreign Affairs*) (given at the aforementioned web site in the file: `skew_symmetric_paper_tables_figures.pdf`) plot the issues when analyzed separately. The VAFs are 87.5% for domestic affairs and 93.0% for foreign affairs. As expected, the results are very similar, though not identical as when compared across all issues. For the latter comparison, the issue “Reducing the federal deficit” is more important than “Lowering the tax burden for all Americans”; when compared with just the domestic issues, however, this relationship is reversed. A few other issues are interchanged with one another, but for the most part the ordering is the same when compared comprehensively. Finally, Tables 1 and 2 display the issues in a table similar to Mr. Blow’s but with the issues now ranked accordingly; Tables 3 and 4 display, respectively, the skew-symmetric matrices for the domestic issues and foreign issues with their coordinates listed in the column heading. In the skew-symmetric matrices, there is a fairly clear gradient across the rows. Aside from two cases in the foreign affairs data, all entries above the diagonal are positive and all entries below the diagonal are negative in both matrices (these four tables are in the file: `skew_symmetric_paper_tables_figures.pdf`).

Young People's Ratings on World Affairs

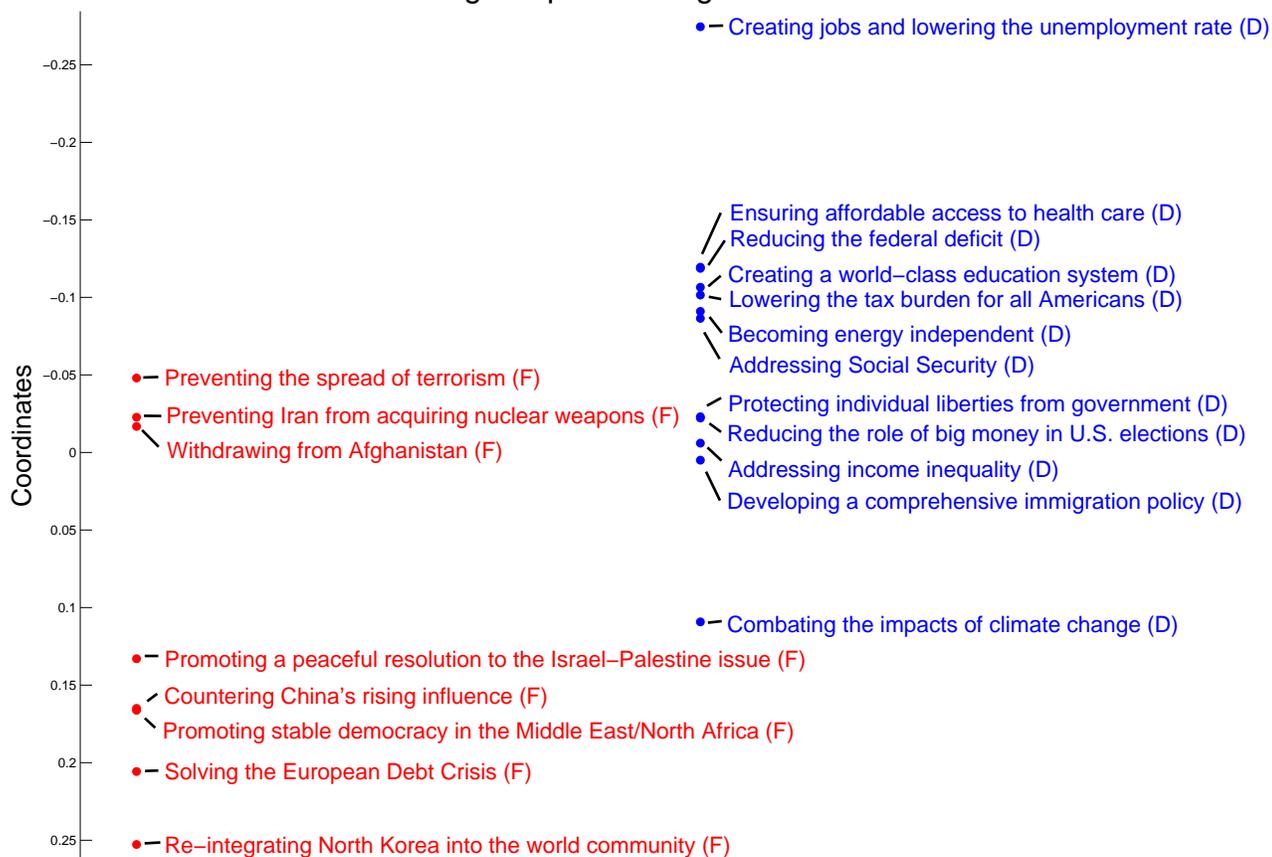


Figure 1: Young people's ratings on domestic and foreign issues ranked in terms of importance. The vertical axis represents the coordinates for the issues.

3 Discussion

The skew-symmetric scaling approach to ordering the issues provides a representation of how young Americans view domestic and foreign issues. It is evident that the most important issue among this population is “Creating jobs and lowering the unemployment rate,” and it is much more important than any other issue. The least important is “Re-integrating North Korea into the world community.”

Probably the most interesting thing to note from Figure 1 is that domestic issues appear to be more important than foreign issues. Because Mr. Blow uses separate tables for the types of issues, it is not possible to infer this from his presentation; thus, this information is lost. The least important domestic issue to young Americans is “Combating the impacts of climate change,” but compared to the foreign issues, the issue of climate change is actually more important than most. Many issues appear to be similar in terms of importance. For instance, “Ensuring affordable access to health care,” “Reducing the federal deficit,” “Creating a world-class education system,” “Lowering the tax burden for all Americans,” “Becoming energy independent,” and “Addressing Social Security” are essentially equally important among young Americans. The most important foreign issue is “Preventing the spread of terrorism,” and it is not as important as the Social Security issue. “Preventing Iran from acquiring nuclear weapons” and “Withdrawing from Afghanistan” are the next two most important foreign issues. These two issues are similar to each other in terms of importance and similar to the two domestic items of “Protecting individual liberties from government” and “Reducing the role of big money in U.S. elections.”

The type of comparisons just given between domestic and foreign issues would be difficult to make with the two tables as presented by Mr. Blow. Skew-symmetric scaling provides a useful tool for summarizing asymmetric proximity data. Given a series of similar surveys conducted over several periods of time, one can even compare how issues have changed over time. The most important issue in 2012 may not necessarily have been that important several years ago or remain important several years from now. Similarly, certain issues not important at this time may have been more important in the past.

References

- [1] Blow, C. M. (2012, April 27). Young People's Priorities. *The New York Times*. Retrieved from <http://www.nytimes.com>.
- [2] Hubert, L., Arabie, P., & Meulman, J. (2001). *Combinatorial data analysis: Optimization by dynamic programming*. Philadelphia, PA: Society for Industrial and Applied Mathematics.
- [3] Institute of Politics, Harvard University (2012). *Survey of Young Americans' Attitudes Toward Politics and Public Service: 21st Edition*.
- [4] MATLAB [Computer Software]. Natick, Massachusetts: The MathWorks Inc., 2013.
- [5] Zielman, B., & Heiser, W. J. (1996). Models for asymmetric proximities. *British Journal of Mathematical and Statistical Psychology*, 49, 127–146.