Thurstone's Law of Comparative Judgment for Group Support

Zheng Li
New Jersey Institute of Technology

Kung-E Cheng
Rutgers University

Yuanqiong Wang
New Jersey Institute of Technology

Starr Roxanne Hiltz
New Jersey Institute of Technology, roxanne.hiltz@gmail.com

Follow this and additional works at: http://aisel.aisnet.org/amcis2001

Thurstone's Law of Comparative Judgment for Group Support

Zheng Li
CIS Department
New Jersey Institute of Technology
zxl8078@njit.edu

Kung-E Cheng
Management Program
Graduate School-Newark
Rutgers University
kecheng@pegasus.rutgers.edu

Yuanqiong Wang
CIS Department
New Jersey Institute of Technology
yxw9836@njit.edu

Starr R. Hiltz
CIS Department
New Jersey Institute of Technology
hiltz@adm.njit.edu

Murray Turoff
CIS Department
New Jersey Institute of Technology
turoff@adm.njit.edu

Abstract

Thurstone's law of Comparative Judgment has the important quality of being able to transform rank order data or comparative preference data by individuals in a group to a single group composite interval scale. In so doing it can convey to the group information on the group position than many other ways of processing rank order data. We discuss our current implementation of this tool on the Web and our plans for extending it as a group support tool.

Keywords: Thurstone scaling, rank order, paired comparison, GDSS, voting tool, dynamic voting

Introduction

Scaling is the science of determining measuring instruments for human judgment (McIVER 1981). Clearly, one needs to make use of appropriate scaling methods to aid in improving the accuracy of subjective estimation and voting procedures (Turoff & Hiltz 1997). Torgerson (1958) pointed out that scaling, as a science of measuring human judgment, is as fundamental as collecting data on well-developed natural sciences. Nobody would deny the fact that all science advances by the improvement of its measurement instruments. And researchers are constantly attempting to obtain more effective scaling methods that could be applied to the less well developed yet more complicated social sciences. Of the many methods that have been developed to measure human judgment, there are some very basic methods, such as Thurstone scaling, Likert scaling, and Guttman scaling etc.

Scaling models can be distinguished according to whether they are intended to scale persons, stimuli, or both (McIVER 1981). For example, Likert scale is a subject-centered approach since only subjects receive scale scores. Thurstone scaling is considered a method to evaluate the stimuli with respect to some designated attributes. It is the stimuli rather than the persons that are scaled (Togerson 1958). Guttman scaling is an approach in which both subjects and stimuli can be assigned scale values (McIVER 1981).

In the research of group decision support system (GDSS), in many cases, researchers are interested in soliciting individual and/or group preferences (Fjermestad, J. & Hiltz, S.R. 1999). In these cases, people need to vote on a set of items to determine their preferences and help make group decisions. During this process, there are a couple of questions we need to discuss further: How

to collect, select/filter and present the items to be voted for preferences? What scaling methods are to be employed, and what are their characteristics and limitations? What kind of voting tool is to be used? What is the principle and potential limitations of such voting tool? … These are the research issues we are currently working on. In this paper, we discuss our current implementation of Thurstone's scaling tool on the Web and our plans for extending it as a group support tool within our collaborative voting tool, which is under developing now.

Thurstone's Law of Comparative Judgment and the Scaling Tool

Thurstone scaling represented a major advancement in the scaling of psychological stimuli (Thurstone, 1927). Building on his law of comparative judgment, Thurstone introduced three methods of scaling: paired comparisons, successive intervals, and equal-appearing intervals. While Thurstone scaling techniques were quite popular during the 1920s and 1930s, they are not employed widely today because of a number of limitations (Togerson 1958; Luce 1994). However, many methods that have been used today are extensions of Thurstone's scaling method (White et al. 1999).

One main characteristic of Thurstone scaling method is that: Thurstone's law of Comparative Judgment has the important quality of being able to transform rank order data or comparative preference data by individuals in a group to a single group composite interval scale. In so doing it can convey to the group information on the group position than many other ways of processing rank order data. When direct scales (such as a physical scale) are not available, by using human judgment, one could use either rank order or paired comparison to obtain an ordering of the preference of the objects based on the comparative judgments of the group of individuals. Then, by applying a set of mathematic equations, such binary or ordinal scale data could be turned into interval scale data, which can indicate the relative distances between the objects that been judged. When applying Thurstone scaling method, it is not assumed that each stimulus always evokes the same discriminal process for different individuals or even for the same individual at different times.

There are very important practical reasons to employ the method. In many GDSS researches, it is comparatively more accurate to ask individuals to rank order items, such as objectives or goals, than to ask for interval or ratio measures. For example, in many cases the appropriate judgment we wish to solicit from an individual is a ranking (i.e. ordinal scale measurement) of individual items. A person can estimate that a particular goal is more important than another one; however, it is much more difficult to estimate consistently how much more important among a group of individuals. A scaling method such as Thurstone's Law of Comparative Judgment (Thurstone 1927; Torgerson 1958) can transform individual ranking judgments and produce analytically a group result, which is an interval scale rather than a rank ordered scale. Providing the group the results in terms of this interval scale allows the individuals to detect in a much more reliable manner the extent to which certain objectives are clearly distinct from other objectives, and which are considered in closer proximity. Merely providing an averaging of the ranking scale does not contribute this added insight to the group as a whole.

Furthermore, standard averaging approaches can lead to inconsistencies in group judgments (i.e. Arrow's Paradox). This can occur when there are disagreements underlying the averaging and when there is a lack of appropriate "anchoring" of the scales. Thurstone's law of comparative judgment provides a way to look into the degree of agreement/disagreement in a new point of view.

In addition, traditionally the development of scaling methods has been to determine an absolute measurement of human response where one assumes the human does not change. For example, the use of surveys for psychological scales or consumer preferences in marketing studies never took changes in their consideration while it is quite possible the user attitudes/preferences could be fluid and fluctuated in many cases. Thurstone scaling has the advantage of admitting and allowing such changes.

In group support systems and the original Delphi method (Linstone & Turoff 1975), there is a completely different objective: How to use these human judgment measuring instruments to provide feedback to the individual on the implications of their judgment and feedback to the group to expose consistencies (e.g. agreement) and inconsistencies (e.g. disagreement) in the resulting group judgment. Note that consistency is a broader concept than agreement. We identified that after about 80 years, Thurstone's law of comparative judgment is still the sort of right tools we are looking for. And we are working on developing and integrating the Thurstone's law of comparative judgment scaling method into our future group support tool kits—the collaborative voting tool kits — in serving such purposes.

Arrow's Theorem (Arrow 1951) says that decision-making is impossible only in the sense that it is impossible to obtain a group ranking based on the input of individual rankings (Roberts, 1976). Arrows paradox may limit the ability to define a single perfect measurement instrument in a group setting, but it does not prevent the use of multi-instruments that complement one another and exhibit different properties of the group judgment under different conditions. In a group support system, we propose one solution

to look at group consistencies: allow the group consensus function to choose several possible consensus rankings, rather than just one. This means Thurstone scaling method needs to be used accompany with other measurements, such as Condorcet's law, Borda count etc., to gain more insight than just one, and to overcome the certain limitations and possible manipulatability that single measurement may introduce. Yet the group needs to understand these methods and understand further whether there were small or large differences underlying the resulting same point on the scale.

Our current research involves developing a web-based collaborative voting tool. It is not a simple tool that just provides majority voting or simple ranking, but integrated several major voting and scaling methods. For example, the voting tool includes:

• Voting interface (scaling methods): support "yes/no" voting, Likert scale, rank order, paired comparison, interval scale, ratio scale, etc.

• Measurements: support Thurstone's law of comparative judgment, simple majority voting, Condorcet's law, Borda count, approval voting etc.

The tool of Thurstone's scaling which we discussed above is part of this collaborative voting tool. Right now we have finished this part. A web site has been set up on the Thurstone's Law of comparative judgment alone with a Java applet demo of the tool of the method. Interested readers please refer to the following URL: http://www-ec.njit.edu/~zxl8078/research/Thurstone.html

This site is open to all researchers solely for research purposes. Our first version will be a single user tool for people needing to analysis the data from groups, and we will be working on the Web-based multi-user version where the users can enter and discuss their rankings via the Web.

Another major feature of the tool is to provide human dynamic voting, that is, allow group users to change their mind and vote repeatedly until certain criteria is met. This provides users a way to look into the dynamics of group decision-making. Related research on this area will be reported later.

Developing such a composite toolkit is not only helpful to test and explore better ways of measurement of human judgment as we discussed above, but also have significant practical values for real problem solving.

Conclusions

From the work we've been done, we come to the following conclusions:

First, the use of a complex set of analysis results poses two major challenges. First, we may visualize the relative comparison of alternative results on the same data set, which may present opposing differing views of the group results. Second, the individual and the group cannot be expected to immediately make use of such a tool. It would be similar to providing an Ohm meter to someone who never had a basic course in electricity. This means the group has to go through a learning phase. However, this is much easier to do in the context of asynchronous and continuous group support systems than in short term face-to-face meeting exercises. In any case the design of the system has to incorporate learning material on the meaning of the method, and what the inferences are that can be made by individuals and groups from the results of the analysis.

Second, the future of CMC in group communications will be the incorporation of these tool kits as an integral part of the communication process. One would expect that even before the preference is done, there will be a process of just voting on which items and which version of items will be included in the final list to go to the use of this tool. We call this pre-voting activity "List Gathering" activity, and some of our colleagues are working hard on developing such tool in the meantime.

Finally, the next generation of group support systems has to allow the groups to evolve the use of the tools they want for their application and the nature of their group. For everyday groups in organizations it has to be a process of learning and evolving on the specific configuration of how to deal with the underlying problems in their application domain. It will no longer be a fixed system created by programmers and facilitated by external facilitators. The user group needs to become the definitive designers of their communication process.

Acknowledgment

References

Arrow, K. *Social Choice and Individual Values*, Cowles Commission Monograph 12, New York: John Wiley & Sons, 1951

Fjermestad, J. and Hiltz, S.R. "An Assessment of Group Support System Experimental Research: Methodology and Results", *JMIS Winter 1998-99* (15:3), pp.7-149.

Linstone, H. and Turoff, M., *The Delphi Method: Techniques and Applications*, Addison-Wesley, 1975.

Luce, R. D. "Thurstone and Sensory Scaling: Then and Now." *Psychological Review,* 1994, pp.271–277.

McIver, J.P. and Carmines, E.G. *Unidimensional Scaling*, CA: Sage Publications, 1981

Roberts, F.S. *Discrete Mathematical Models: With Application to Social, Biological, and Environmental Problems*, New Jersey: Prentice-Hall, 1976

Thurstone, L. L. "A law of comparative judgment." *Psychological Review*, 34, pp.273–286, 1927

Togerson, W.S. *Theory and Methods of Scaling*, New York: Wiley, 1958

Turoff, M. and Hiltz, S.R., "Computer Based Delphi Processes", CACM 1997

White, K.G. and Wixted, J.T. "Psychophysics of Remembering," *Journal of the Experimental Analysis of Behavior* (71:1), 1999, pp.91–113