Journal of Uncertain Systems
Vol.4, No.4, pp.301-305, 2010
Online at: www.jus.org.uk

How to Estimate Individual Contributions to a Group Project

Jeerapa Sappakitkamjorn[1], Wannaporn Suriyakat[1],
Wichai Suracherdkiati[1], Olga Kosheleva[2,*]

[1] *Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, 1518 Pibulsongkram Rd., Bangsue, Bangkok 10800 Thailand*

[2] *Department of Teacher Education, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA*

Received 18 January, 2009; Revised 30 January, 2010

Abstract

In many real-life situation, the only information that we have for estimating the individual contributions E_j to a group project consists of individual estimates e_{ij} of contributions of other participants j. In this paper, we describe a new faster algorithm for estimating individual contributions to a group project based on the estimates e_{ij}.

Keywords: group projects, estimating individual contributions

1 Formulation of the Problem

Formulation of the practical problem. How can we estimate individual contributions to a group project? This problem is important in education when several students work together on a project, it is important in the business environment when several people work together on a joint project.

In all such situations, we need to know the relative contributions $E_1, \ldots, E_n$ of all n participants, relative in the sense that they represent the fraction of the overall credit – and thus, the sum of these contributions should be equal to 1:

$$\sum_{i=1}^{n} E_i = 1. \tag{1}$$

Available information for solving the problem. In many practical situation, the only available information for estimating contributions consists of the estimates that different participants give to each other's contribution. In this case, we have n^2 values e_{ij} $(1 \le i, j \le n)$ – estimates made by the i-th participate of the contribution of the j-th participant.

Ideal case when all estimates are unbiased. In the ideal case when all estimates are unbiased, for each participant j, we have n estimates $e_{1j}, \ldots, e_{nj}$ for the desired value E_j. In this case, we have n approximate equalities to find E_j:

$$E_j \approx e_{1j}, \quad \ldots, \quad E_j \approx e_{nj}. \tag{2}$$

To find a reasonable estimate E_j from these approximate equalities, a natural idea is to use the Least Squares technique and find the value E_j for which the sum

$$\sum_{i=1}^{n} (E_j - e_{ij})^2 \tag{3}$$

is the smallest possible. This is a textbook use of the Least Squares method, to combine several estimates of the same quantity, and the solution to this optimization problem is well known – it is the arithmetic average of these estimates:

$$E_j = \frac{e_{1j} + \ldots + e_{nj}}{n}. \tag{4}$$

*Corresponding author. Email: olgak@utep.edu (O. Kosheleva).

Problem: self-estimates are biased. In practice, estimates of others' contributions are often unbiased, but it is very difficult to get an unbiased estimate of one's own contribution in comparison with the contributions of others.

In other words, while the estimates e_{ij} for $i \neq j$ are really unbiased, the estimates e_{ii} are too subjective and biased to be useful.

Available information for solving the problem: revisited. Because of the bias, we can say that the only available information about each value E_j consists of estimates e_{ij} with $i \neq j$.

What was known before. The problem of estimating the values E_j based on the estimates e_{ij} with $i \neq j$ was considered earlier; see, e.g., [1, 2] and references therein. In particular, [2] describes an algorithm for estimating E_j.

Limitations of the previous approaches. The algorithm from [2] is based on solving a highly non-linear optimization problem and is, therefore, reasonably time-consuming.

What we do in this paper. In this paper, we use ideas from [1] to come up with a much faster algorithm for estimating E_j, an algorithm whose most time-consuming step is solving a system of linear equations.

2 Towards the Formulation of the Problem in Precise Terms

In principle, each participant i provides *relative* estimates $e_{i1}, \ldots, e_{in}$ which add up to 1. Once we have *absolute* estimates $a_{i1}, \ldots, a_{in}$, we get the relative contributions e_{ij} by dividing each value a_{ij} by the total contribution $a_{i1} + \ldots + a_{in}$:

$$e_{ij} = \frac{a_{ij}}{a_{i1} + \ldots + a_{in}}. \tag{5}$$

If all these estimates were unbiased, then each estimate e_{ij} would provide an unbiased estimate of E_j.

In particular, in the idealized case when all estimates are exact, we have $a_{ij} = c_i \cdot E_j$ for some value c_i. Thus, for the normalized values, we get

$$e_{ij} = \frac{c_i \cdot E_j}{c_i \cdot E_1 + \ldots + c_i \cdot E_n} = \frac{E_j}{E_1 + \ldots + E_n}, \tag{6}$$

hence $e_{ij} = E_j$ and

$$e_{i1} + \ldots + e_{in} = E_1 + \ldots + E_n = 1. \tag{7}$$

In practice, as we have mentioned, each self-estimate a_{ii} is biased: $a_{ii} \neq c_i \cdot E_i$. As a result, even when the i-th participant provides the exact *absolute* values of the contributions $c_i \cdot E_j$ for $i \neq j$, the corresponding *relative* (normalized) contribution

$$e_{ij} = \frac{a_{ij}}{a_{i1} + \ldots + a_{in}} =$$

$$\frac{c_i \cdot E_j}{c_i \cdot E_1 + \ldots + c_i \cdot E_{i-1} + a_{ii} + c_i \cdot E_{i+1} + \ldots + c_i \cdot E_n} = \frac{E_j}{s_i} \tag{8}$$

(where

$$s_i = E_1 + \ldots + E_{i-1} + \frac{a_{ii}}{c_i} + E_{i+1} + \ldots + E_n,) \tag{9}$$

is different from E_j: instead of the correct value

$$E_j = \frac{E_j}{E_1 + \ldots + E_{i-1} + E_i + E_{i+1} + \ldots + E_n} = \frac{E_j}{1}, \tag{10}$$

we get the new value

$$e_{ij} = \frac{E_j}{s_i}, \tag{11}$$

where, due to $a_{ii} \neq c_i \cdot E_i$, we have

$$s_i \stackrel{\text{def}}{=} E_1 + \ldots + E_{i-1} + \frac{a_{ii}}{c_i} + E_{i+1} + \ldots + E_n \neq 1. \tag{12}$$

Thus, the relative estimates e_{ij} of j-th contribution will be equal not to E_j, but to the normalized value (11). Since the estimates are only approximate, we have an approximate equality

$$e_{ij} \approx \frac{E_j}{s_i}, \tag{13}$$

or, equivalently,

$$E_j \approx e_{ij} \cdot s_i. \tag{14}$$

Here, the values e_{ij} are known, while the estimates E_j and the values s_i are unknown. The Least Squares approach now means that we minimize the sum of the squares of the discrepancies

$$\sum_{i \neq j} (E_j - e_{ij} \cdot s_i)^2 \tag{15}$$

under the constraint (1). Thus, we arrive at the following problem.

Precise formulation of the problem. Minimize the expression (15) under the constraint (1).

3 Solving the Problem

Towards solving the problem. First, let us use the Lagrange multiplier method to reduce the above constrained optimization problem to the following un-constrained one: minimize

$$J \stackrel{\text{def}}{=} \sum_{i \neq j} (E_j - e_{ij} \cdot s_i)^2 + \lambda \cdot \left(\sum_{i=1}^{n} E_i - 1 \right), \tag{16}$$

where the Lagrange multiplier λ must be chosen in such a way that the constraint (1) is satisfied.

Since the function J attains minimum, its partial derivatives must be equal to 0. Differentiating J with respect to E_j and equating the derivative to 0, we conclude that

$$\frac{\partial J}{\partial E_j} = 2 \sum_{i \neq j} (E_j - e_{ij} \cdot s_i) + \lambda = 0. \tag{17}$$

Dividing both sides of this equality by 2 and taking into account that E_j is repeated for each $i \neq j$, i.e., $n-1$ times, we conclude that

$$(n-1) \cdot E_j - \sum_{i \neq j} e_{ij} \cdot s_i = -\frac{\lambda}{2}. \tag{18}$$

This equation can be somewhat simplified if we take $e_{ii} \stackrel{\text{def}}{=} 0$; then, the sum over all $i \neq j$ can be simply described as the sum over all i:

$$(n-1) \cdot E_j - \sum_{i=1}^{n} e_{ij} \cdot s_i = -\frac{\lambda}{2}. \tag{19}$$

It should be mentioned that the values $e_{ii} = 0$ are selected *only* for the purpose of simplifying computations; these values *do not* mean that we somehow think that each participate estimates his or new own contribution as 0.

Differentiating J with respect to s_i and equating the derivative to 0, we conclude that

$$\frac{\partial J}{\partial s_i} = 2 \sum_{j \neq i} (E_j - e_{ij} \cdot s_i) \cdot e_{ij} = 0. \tag{20}$$

Dividing both sides of this equality by 2, we get

$$s_i \cdot \sum_{j \neq i} e_{ij}^2 = \sum_{j \neq i} E_j \cdot e_{ij}, \tag{21}$$

or, equivalently, that

$$s_i = \frac{\sum\limits_{\ell \neq i} E_\ell \cdot e_{i\ell}}{\sum\limits_{m \neq i} e_{im}^2}. \tag{22}$$

(Here, for convenience of the following transformations, we renamed the indices in the two sums into two different ones.) By taking $e_{ii} = 0$, we can simplify this expression into the following one:

$$s_i = \frac{\sum\limits_{\ell=1}^{n} E_\ell \cdot e_{i\ell}}{\sum\limits_{m=1}^{n} e_{im}^2}. \tag{23}$$

Substituting the expression (23) into the formula (19), we conclude that

$$(n-1) \cdot E_j - \sum_{i=1}^{n} e_{ij} \cdot \frac{\sum\limits_{\ell=1}^{n} E_\ell \cdot e_{i\ell}}{\sum\limits_{m=1}^{n} e_{im}^2} = -\frac{\lambda}{2}, \tag{24}$$

i.e.,

$$(n-1) \cdot E_j - \sum_{\ell=1}^{n} c_{j,\ell} \cdot E_\ell = -\frac{\lambda}{2}, \tag{25}$$

where

$$c_{j,\ell} \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{e_{ij} \cdot e_{i\ell}}{c_i}, \tag{26}$$

where

$$c_i \stackrel{\text{def}}{=} \sum_{m=1}^{n} e_{im}^2. \tag{27}$$

The values $c_{j,\ell}$ can be explicitly computed from the known values e_{ij}. To solve the resulting system of linear equations (25) with the unknown value λ, it is sufficient to solve it for $\lambda = -2$, when $-\dfrac{\lambda}{2} = 1$, and then multiply the resulting values e_j by a common factor in such a way that the new values add up to 1: i.e., take

$$E_j = \frac{e_j}{S}, \tag{28}$$

where

$$S \stackrel{\text{def}}{=} \sum_{i=1}^{n} e_i. \tag{29}$$

Thus, we arrive at the following algorithm.

New algorithm for solving the problem. Given the values e_{ij} for $i \neq j$, we first take $e_{ii} \stackrel{\text{def}}{=} 0$. Then, we compute the values

$$c_i \stackrel{\text{def}}{=} \sum_{m=1}^{n} e_{im}^2 \tag{30}$$

and

$$c_{j,\ell} \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{e_{ij} \cdot e_{i\ell}}{c_i}, \tag{31}$$

and solve the following system of linear equations

$$(n-1) \cdot e_j - \sum_{\ell=1}^{n} c_{j,\ell} \cdot e_\ell = 1. \tag{32}$$

After this, we compute the sum

$$S = \sum_{i=1}^{n} e_i, \tag{33}$$

and then, the desired estimates as

$$E_j = \frac{e_j}{S}. \tag{34}$$

Discussion. In this algorithm, the most time-consuming step is solving a system of linear equations. Thus, this algorithm is indeed much faster than the algorithm from [2] that requires a solution of the non-linear optimization problem.

Acknowledgments

The authors are thankful to Sa-aat Niwitpong and to Vladik Kreinovich for their guidance and help.

References

[1] Kreinovich, V., *How to Pay for Collective Creative Activity?*, Technical Report, Informatika Research Center, St. Petersburg, Russia, 1989.

[2] Leung, Y.-W., Least-squar-error estimate of individual contribution in group project, *IEEE Transactions on Education*, vol.41, no.4, pp.282–284, 1998.