Matching Index Calculation:

12/2021

This document describes in detail how the mentor-mentee matching algorithm uses survey response data to generate a summary metric (M) that describes the estimated compatibility of a mentor and mentee. I first present the equation behind the algorithm and break it down afterwards.

Eq1: Full equation

Eq2: Moderately simplified equation

Eq3: Very simplified equation

First, for every potential mentor and mentee pair, the following equation identifies the count of the intersection ( of mentor (o)/mentee (e) responses (R) for a question (i) and divides by the average number of entries selected by mentors and mentees . We call the resulting number the “similarity score” . This number is calculated for each question and potential pairing. A perfect intersection of mentor and mentee response entries yields a maximum of 1 while 0 indicates no intersection between mentor and mentee responses for that question.

For example:

Mentor A is interested in:

**Chromatin/Epigenetics**

**Computational Biology/Genomics**

**Gene Expression**;**Genetics**

RNA Biology

(5 responses total)

Mentee B is interested in:

**Chromatin/Epigenetics**

**Computational Biology/Genomics**

Developmental Biology

Evolution;Evolutionary Developmental Biology

**Gene Expression**

Genetics

(6 responses total)

They share 3 responses so their similarity score ( is calculated to be 3/((5+6)\*0.5) = 0.55. They are pretty similar, but not an exact match. If they matched more, their score increases.

Additionally, each question is assigned a submitter-determined weight. Mentors and mentees prioritize sharing different traits to differing degrees. We divide the weight assigned for a given question (1 being very important mentors and mentees share similar responses and 0 being indifferent) by the total amount of weight assigned by the submitter . The resultant metric is an adjusted weight . This number is specific to each submitter’s response to each question and thus may differ between mentors and mentees. is multiplied by the similarity score to yield a matching score for each question. Again, the mentee and mentor may have differing opinions of question importance. If a submitter thinks every question is the most important, then it dilutes the significance of each singular question. These scores are summed for both the mentee and mentor . These numbers are simplified to and respectively. and may range from 0 (poor match) to 1 (perfect match).

For example:

Mentor A rates sharing research interests with their mentee as being **3** out of 10 in importance.

= 0.3

Mentee B rates sharing research interests with their mentor as being 9 out of 10 in importance.

= 0.9

Throughout the survey, Mentor A rated many other things to be important

and thus, the total amount of weight (including = 0.3) assigned was 5.3.

Mentee B felt that matching research interests was the only important quality of their mentee

and thus, the total amount of weight (including = 0.9) assigned was 1.4.

Therefore, the adjusted weight for the question “ for Mentor A () is 0.3/5.3 = 0.057. Essentially, the similarity with a mentee’s response to this question accounts for 5.7% of .

The adjusted weight for the question “ for Mentee B () is 0.9/1.4 = 0.643. Essentially, the similarity with a mentor’s response to this question accounts for 64.3% of . The adjusted weights to all questions add to 1 for a given survey-taker.

Thus, for Mentor A, = 0.55 \* 0.057 = 0.031

And for Mentee B, = 0.55 \* 0.643 = 0.354

The different weights between Mentor A and Mentee B responses to question “Research Interests” scales the similarity score according to how they each feel about the relative importance of the question.

Before merging and , we re-scale these metrics by , the fraction of the final metric determined by mentors matching mentee expectations. This allows us to prioritize the match of the mentor to the mentee while not entirely discounting the matchedness from of the mentee to the mentor. should always yield a matching score between 0-1, which we present on a 1-100 scale by multiplying by 100. This final score, , reflects the overall matchedness of a mentor/mentee pairing.

For example:

If the different weights of Mentor A lead to a score of (the mentee somewhat matches the traits desired by a mentor) while Mentee has a score of (the mentor highly matches in traits desired by the mentee), the final score is = . That is a fairly high score, mostly driven by mentor matches for questions deemed important by the mentee, but not discounting the mentor’s preferences.

Note: In my current iteration of the analysis ≥ 3 matches between mentor and mentee responses is rounded up to the same value as a full match. For instance, if a mentor is very familiar with 7 labs, they will be maximally compatible with someone who is only interested in 3 labs.

Interesting properties:

1. Maximally matching pairs have a total of 100. Completely incompatible pairs have a total of 0.
2. Can score an arbitrarily large number of mentors or mentees.
3. The equation is fully compatible with adding or removing questions from the survey
4. Poor matches on questions that don’t matter to either mentee or mentor carry no penalty. The mentors and mentees can answer as many questions as they wish. Skipping questions or providing low weights to the responses increases the relative weight for higher priority questions. However, answering each question is still important in case you share qualities that your potential match finds important.
5. Can be adjusted to decrease or increase to adjust the amount the analyzer values the mentee’s input. > 0.5 means that a mentor matching a mentee’s preferences is more important to the final score than the mentee matching mentor preferences and vice versa.
6. If a mentee has little preference in mentor, they will end up scoring well with many mentors; i.e. they are easy to assign to other mentors.