Proposal a New Scoring Method for FractalForum.com Contests

1. Previous methods

In the past, the FF contest used the **average-score** of each entry. Although this appeared to be adequate method, it had a few drawbacks. Here are examples:

- An entry with [99 ★★★★★ + 1 ★★★★] has an average of is 4.99. It is outranked by an entry with a single [★★★★★], which has average 5.00.
- A no-vote is ignored by the scoring. Thus, voting [★] will worsen the ranking. For an entry with [10 ★★★★★], one no-voting is effectively equivalent to voting [★★★★★].

The current scoring method uses **sum-score** of each entry. Although this resolved the problems with the past method, it introduced new problems. A notable example is this:

- An entry with [10 ★★★★★] with score 50 will be outranked by an entry with [13 ★★★★] with score 52.

This method treats no-votes as voting with 0 stars.

2. Proposal for new method

This section describes a proposal for a scoring method, that complies with the following requirements:

- The scoring uses only already existing statistical data (like total sum of stars and total number of votes), i.e. it is not required to gather and analyze more complex data.
- The scoring is calculated by a single mathematical expression, i.e. it is not required to implement a programming algorithm to calculate the scoring.
- The scoring supports natural order of impact of votes (towards the final rating), i.e. [★★★★★]>[★★★★]>[★★★]>[★★]>[★]>[no-vote].
- The scoring protects ranking from deviation of entries, which have very few votes (as described in the average-score).
- The scoring protects ranking from deviation of entries, which have massive amount of not-so-good votes (as described in the sum-score).

The proposed expression is this:

Score = Average*Confidence

where *Average* is the average score, i.e. (number-of-stars)/(number-of-votes); and *Confidence* is a coefficient, which converges to 1 as the number of votes increases.

A suitable expression for the confidence is $(1-\alpha^{\text{number-of-votes}})$, where α is close to, but less than 1. The actual value of α will affect the sensitivity of the scoring method, i.e. how quickly the scores of the entry make us confident that they are adequate. The unrolled formula is:

Score = Stars/Votes*(1- α^{Votes})

3. Examples

Here are examples that explain the behavior of the proposed scoring method. All calculations are based on the assumption of α =0.9.

- An entry has one max vote [★★★★★]. The average is 5.00, but the confidence is 0.10 (i.e. one vote does not make us truly believe that the entry is excellent). As a result, the score is 0.5.
- An entry has two max votes [2★★★★★]. The average is 5.00, the confidence is 0.19, and the score is 0.95.
- An entry with 10 max votes [10★★★★★]. The average is 5.00, the confidence is 0.65, and the score is 3.26 (for the record: confidence 0.99 is reached at 44 votes).
- The problem case for average-score method: an entry with [99 ★★★★★ + 1 ★★★★] will have score 4.96, while an entry with a single [★★★★★], which have score 0.5.
- The problem case for sum-score method: an entry with [10 ★★★★★] will have score 3.26 (average 5, confidence 0.65); while an entry with [13 ★★★★] with have score 2.98 (average 4, confidence 0.75).

4. Conclusion

The proposed algorithm resolves all the problems described in this document. The *confidence* component of the scoring could be also considered as *popularity*. More popular contest entries use the full potential of their average score.

The concrete value of parameter α should be decided so that to provide morally acceptable balance of the importance of the average scores against the importance of the popularity. For example, if α=0.9 then [10★★★★★] ≈[16★★★★]. However, if we increase α to α=0.98 then we will make confidence harder to gain and [10★★★★★] ≈ [84★★★★].

Additionally, it is suggested that scoring is rounded to 2 or 3 digits after the point.

Appendix 1

The behavior of the formula can be fine-tuned by finding a suitable value for α. Thus, neither the value of α=0.90, nor α=0.98 should be accepted as a must. They are used only to demonstrate the proposal.

As a demonstration, here is the ranking of the latest contest (section FILMS) using the current and the proposed methods. This is only a demonstration and the results are not to be used for any contestation.

FILMS (α=0.90)

Current ranking	Stars	Votes	Average	Confidence Popularity	New score (Average*Confidence)	New ranking
1	113	26	4.346	0.935	4.065	2
2	112	26	4.308	0.935	4.029	3
3	110	25	4.400	0.928	4.084	1
4	80	19	4.211	0.865	3.642	5
5	78	18	4.333	0.850	3.683	4
6	76	19	4.000	0.865	3.460	6
7	75	19	3.947	0.865	3.414	7
8	73	19	3.842	0.865	3.323	8
9	64	20	3.200	0.878	2.811	10
10	63	18	3.500	0.850	2.975	9

As seen by the last column, the current 3rd place (green) will outrank 1st (orange) and 2nd places (red), because its average is better, and its popularity is almost the same high level – the first three entries are so popular, that their relative ranking is controlled by the average.

Current 5th place (blue) has worse popularity than 2nd place (red), but better average. Yet, it is unable to utilize the full potential of its average, because of the poor popularity. In this case the popularity overpowers the average.