

1 Approach

To address above mentioned issues a statistical approach is considered where first step is to identify the features which are relevant for the model. Secondly, model needs to learn the parameters which govern the interaction between these features. Below I describe the model.

1.1 Subscripts

- i is subscript for question.
- j is subscript for answer.
- k is subscript for user.

1.2 Notations

1. u_k : Quality measure of the k th user.
2. q_i : Quality measure of the i th question.
3. va_{ij} : Normalized votes corresponding to j th answer of i th question.
Calculated as: $va_{ij} = \frac{|sa_{ij}|}{\sum_j |sa_{ij}|}$
where sa_{ij} is the actual votes(score) read from data dump.
4. a_{ijk} : Quality measure of a_{ij} th answer given by the k th user.
5. f_{acc}^{ij} : Boolean flag telling if this answer was Accepted, read from data dump.
6. r_k : Reputation of the k th user, read from data dump.
7. N_a^i : Number of answer to i th question, read from data dump.
8. vq_i : Number of votes to i th question, read from data dump.

1.3 Equations

Below equations model the relation/dependence between the above defined parameters. Bold values are known features. All the features were scaled between 0 & 1

1. $a_{ijk} = f_a(u_k, \mathbf{va}_{ij}, \mathbf{f}_{acc}^{ij})$
 $a_{ijk} = 1/3 * u_k + 1/3 * \mathbf{va}_{ij} + 1/3 * \mathbf{f}_{acc}^{ij}$
2. $u_k = f_u(\{a_{ijk}\}_{ij}, \{q_i\}_k, \mathbf{r_k})$, where $\{a_{ijk}\}_{ij}$ is set of all answers by user k
 $u_k = 1/3 * \text{mean}\{a_{ijk}\}_{ij} + 1/3 * \text{mean}\{q_i\}_k + 1/3 * \mathbf{r_k}$
3. $q_i = f_q(u_k, \mathbf{N_a^i}, \mathbf{vq_i}, \sum_j |\mathbf{sa}_{ij}|)$, where $\sum_j |sa_{ij}|$ is sum of votes for all the answers of i th question
 $q_i = 1/4 * u_k + 1/4 * \mathbf{N_a^i} + 1/4 * \mathbf{vq_i} + 1/4 * \sum_j |\mathbf{sa}_{ij}|$

2 Status

Currently, the model is very simple with $va_{ij} = \frac{|sa_{ij}|}{\sum_j |sa_{ij}|}$, where sa_{ij} is the actual votes(score) read from data dump.

and $a_{ijk} = w_1 * u_k + w_2 * q_i + w_3 * va_{ij} + w_4 * f_{acc}^{ij}$

As of now weights w_1, w_2, w_4 are 0 and w_3 is 1, so $a_{ijk} = va_{ij}$

User quality is being modeled as: $u_k \sim \mathcal{N}(\text{mean}(\{a_{ijk}\}_{ij}), \text{Var}(\{a_{ijk}\}_{ij}))$

Question quality is still not modeled.