Gmail by Google

Daniel Wong <wonglkd@gmail.com>

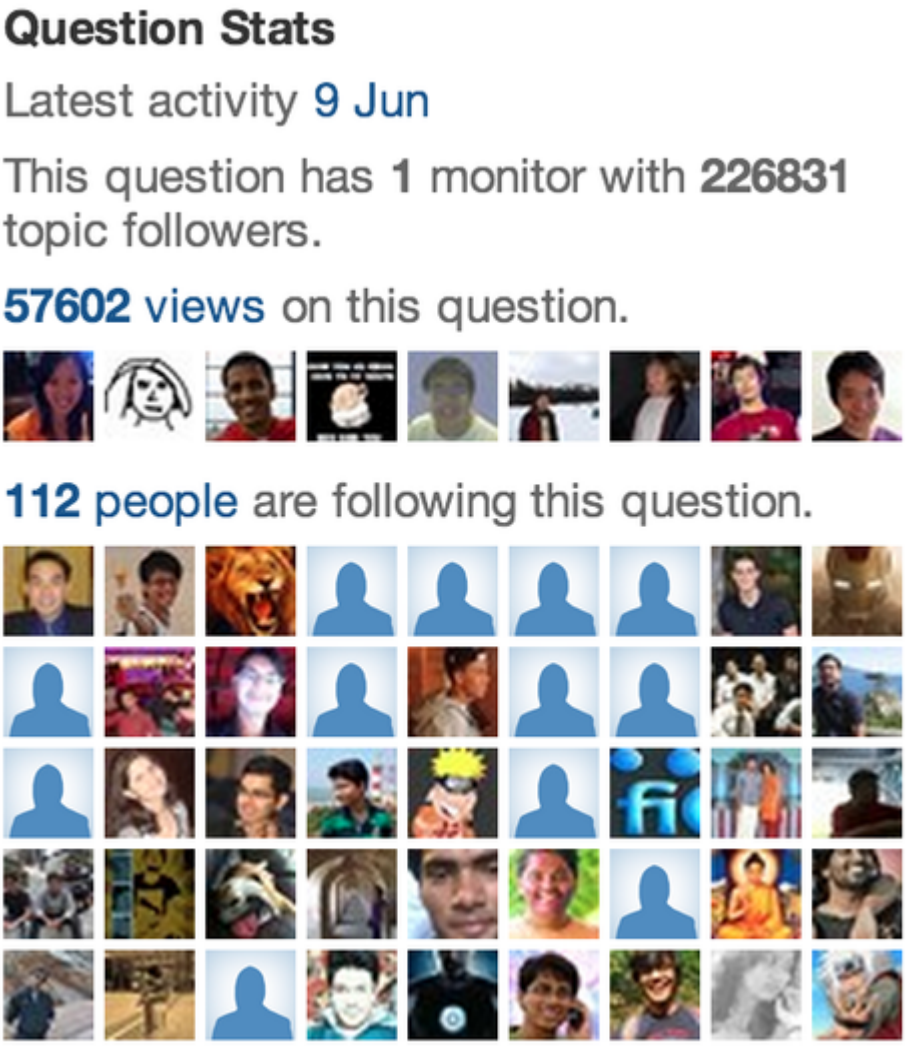
[Codesprint Quora] Quora ML Problem: INTEREST

Quora <noreply@quora.com> Sun, Jul 21, 2013 at 5:59 AM
To: wonglkd@gmail.com

Quora

[Codesprint Quora] Quora ML Problem: INTEREST

Kah Seng Tay



Quora uses machine learning algorithms to try to generate interesting news feeds and digest emails for people.

Before a question gets an answer, we'd like to be able to make use of all the

available information we have to be able to predict how interesting or relevant the question is to people. Ideally, we'd like to be able to tell this in real-time as soon as a few people have viewed it, by measuring the people following the question as a proxy of interest. Can you tell what questions will get the most followers?

For this task, given Quora question text and topic data for questions with 0 visible answers, predict the ratio of viewers to followers.

Input Format

The first line contains N. N questions follow, each being a valid json object. The following fields of raw data are given in json.

- question_key (string): Unique identifier for the question.
- question_text (string): Text of the question.
- context_topic (object): The primary topic of a question, if present. Null otherwise. The topic object will contain a name (string) and followers (integer) count.
- topics (array of objects): All topics on a question, including the primary topic. Each topic object will contain a name (string) and followers (integer) count.
- anonymous (boolean): Whether the question was anonymous.
- __ans__ (float): The ratio of viewers to followers of the question.

This is immediately followed by an integer T.
T questions follow, each being a valid json object.
The json contains all but one field *__ans__*.

Output Format

T rows of JSON encoded fields, with the *question_key* key containing the unique identifier given in the test data, and the predicted value keyed by *__ans__*.

Constraints

question_key is of ascii format.
question_text, name in topics and context_topic is of UTF-8 format.
0 <= followers <= 106
9000 <= N <= 45000
1000 <= T <= 5000

Training Data

Sample testcases can be downloaded here and used for offline training if desired.

Scoring

Your solution is evaluated by a Root Mean Squared Logarithmic Error (RMSLE) metric. We then calculate a final score by how close it reaches a target score of 0.5, and scale that by 100%.

$$\frac{0.5}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log(x_i)-\log(y_i))^2}} \times 100\%$$

Your score will be based only on the hidden input. The sample input is only for your convenience.

Submit your solutions here: Quora ML Problem: INTEREST

To see the post, visit:
http://www.quora.com/login/auto_login?...

To stop receiving emails on Codesprint Quora, visit: http://codesprintquora.quora.com/auto_disable_email_alerts?key=3707b9424e0e29b5bea2cd0678ea7fe8

To control which emails we send you, visit: Email Settings

This email was sent by Quora (650 Castro Street #450, Mountain View, CA 94041). Quora is your best source for knowledge. To disable all email from Quora, visit the following link: Unsubscribe.