Annotation Guidelines

March 2022

1 Overview

Nowadays, People love to ask and answer questions in social media platforms like Quora, StackOverflow and Reddit, etc. However, the notion of a helpful answer can hardly be quantified and determined.
Determining an answer is helpful or not could be highly subjective, but we do believe that there are a few key standards that can help readers better quantify the helpfulness of answers from various perspectives.
To rate the helpfulness of replies to answers using a 5-point Likert scale from (1) Not Helpful to (5) Very helpful, the annotation guidelines in this essay would point out well-defined standards and rules on rating, along with understandable examples that can help readers to better comprehend the rules.
The main features that we are going to consider the most to rate the helpfulness are: accuracy of the content, clarity of writing, completeness of the answer, examples, and generality of the answer.

2 Definition of question type

To give better annotations of the problems, we roughly classify the problems in to two types: Opinion-type problem and Experience-type problem.

2.1 Opinions

Questions asking people to share their opinions about something.

2.2 Experiences

Questions asking people to share their experiences.

The specific judging criteria are dependent on the type of questions thus might vary from one to one. If we think the define of features are too different for those two type of questions, we would provide independent explanations.

3 5 key features for measuring the helpfulness

3.1 Accuracy

Accuracy of content is the key feature that can distinguish (2) from (1).When the answer contains wrong answers that make no sense at all or is completely off the target, we should regard the reply as lacking accuracy. There are mainly two kinds of answers should be conceived as lacking accuracy.

3.1.1 Wrong answers

For the replies that give both completely invalid explanations(process) and wrong conclusions, the answers could be unhelpful or even misleading, thus should be rated as (1) Not helpful.

3.1.2 Off the target

To be at least (2)'somehow helpful' or above, the reply should at least provide some additional information that is essentially related to the question and helpful to solve the problem.

It's not always the case that the reply can always perfectly answer is applicable under any circumstances. In this case, as long as the reply contains information, or even clues that can help partially or fully answer this question, the reply would deserve a rate that is no less than (2).

On the other hand, if the answer is completely reasonable, but is targeting at another question that is barely related to the one we are concerned about, then this answer should be annotated as not helpful.

3.2 Clarity

Clarity in both logic and writing is the basic component of a good expression, which is also a crucial feature for determining how helpful an answer could be. Generally speaking, an answer has to be readable and contains less grammar mistakes to be considered as "clear answer"

Audience matters. A helpful answer should not only convey accurate and sufficient information that is easy to absorb for the audience,but also shows a clear logic thread without ambiguities, subjective opinions, errors or careless speech.

For experience type questions, clarity is more about whether the narrative of the story is clearly written.

For the opinion type questions, the answers should provide clear opinions and also clear explanations on its validity and logic to satisfy the clarity standard.

3.3 Completeness

When the "question" contains multiple questions, multiple meanings or multiple constraints, completeness of the answer is also an important criteria to consider.

Ideally, a complete reply should give answers to all the questions raised in the problem, discuss all the meanings that the problem contains, and should state the logic to prove that the answer satisfies all the constraints indicated by the key words in the problem.

3.4 Concrete Examples

Helpful answers could be abstract and short. However, an outstanding answer should include concrete examples with sufficient details or substantive support materials. In most cases, answer of such quality gets more accessible and better accepted by a larger audience group.

For experience type questions, a concrete example could be a story with detailed narratives that can provide sufficient supports to prove its helpfulness .

For opinion type questions, a concrete example could be a code snippet, , or a reference article. It depends on what the question is asking for. Inappropriate examples or examples lack details could weaken the power of the answer. In this case, we should regard the reply as lacking concrete examples.

3.5 Generality and Consistency

Suppose an answer has satisfied all the features we've mentioned before, how to make it the best answer?

From my perspective, the best answer should be a more generic answer that further extends the scope of the question and provide more insights for the posted problem.

In other words, it should not only solve the problem but also helps all readers who may face the similar issue in the future.

4 How to distinguish 5-point Likert scales?

4.1 5-point Likert scale

We define the 5-point scales as:
(1) Not Helpful
(2) Somehow Helpful
(3) Helpful
(4) More helpful
(5) Very helpful

4.2 Scaling Criteria

The table below specifies the basic standards of the 5 scales.
It's worth of note that the listed standards are the least that the answer should meet to stay 'positive' about a specific scale.
For a scale S and an answer Ans:

If there exists one unsatisfied condition, then the answer Ans should be rated no more than $S-1$ but no less than 1;
If all the conditions of scale S are satisfied, then the answer Ans should be rated no less than S but no more than 5.

Points	Accuracy	Clarity	Completeness	Examples	Generality
1	Not satisfied	/	/	/	/
2	Satisfied	/	/	/	/
3	Satisfied	Satisfied	Satisfied	/	/
4	Satisfied	Satisfied	Satisfied	Satisfied	/
5	Satisfied	Satisfied	Satisfied	Satisfied	Satisfied

4.3 Examples

- (1)Not Helpful

  Question : Humans can only see 0.0035% of all light, what are we not seeing?
  Answer: *THE OTHER LIGHT?!*

  Explanation: This answer does not satisfy the accuracy condition because of 'off the target' rule. To be more specific, it contains no additional information and only repeats the information that the question has conveyed. Thus, it shall be rated as (1).

  Question : What was a moment when the paparazzi truly crossed their line with a celebrity?
  Answer: Good morning.

  Explanation: The question ask you to share a experience. But this answer does not satisfy the accuracy condition because of 'off the target' rule and totally unrelated with the question. To be more specific, it contains no additional information and only repeats the information that the question has conveyed. Thus, it shall be rated as (1).

- (2)Somehow Helpful

  Question: If you could bring animal back from extinction, what would it be and why?
  Answer: The Dodo. Pointless animals, but aren't we all?

  Explanation: This answer satisfies the Accuracy condition but doesn't satisfy Clarity. The readers can't really understand why the Dodo are 'Pointless animals' based on the answer without referring to other answers or information. In other word, the answer is ambiguous and lacks full logic. Thus, it should be rated as (2).

  Question: Would you pay for personal snail mail?
  Answer: : Isn't it better to get proper pen pal?

  Explanation: This answer satisfies the Accuracy condition but doesn't satisfy Clarity. The readers can't get the idea of the answer easily but need to do think more about the relationship and logic between them. In other word, the answer is ambiguous and lacks full logic. Thus, it should be rated as (2).

- (3)Helpful

  Question: What is the scariest thing you have ever witnessed?
  Answer: The extraction of a 19 years old girl out of the water after she was in the river for about 8 weeks.
  Explanation: This answer is accurate, clear and complete. However, the description is too short to understand why the thing is 'the scariest thing you have ever witnessed'. Therefore, the rate is (3).

  Question: How different do you think the world would be today if dinosaurs, like the T-Rex, still existed in the wild?
  Answer: id be better at standing still.

  Explanation: This answer is clear and complete but does not provides enough reasons and examples to make it more helpful.

- (4)More helpful

  Question: What petty revenge are you currently seeking?
  Answer: I'll go first. Capital One gave me this dumb credit card when I was 20 and too stupid to understand interest properly. Anyway I had a 26% interest rate and it put me into some debt foe a while. I've worked to get out of it but even penny by penny every time their stock is low I buy it and when it's high I sell it. I'm going to make my money back cent by cent.

  Explanation: This answer satisfies all the first 4 conditions. The answer is accurate, clear and complete for the question, and it gives detailed descriptions on the what the revenge is and how the revenge is generated from. But in terms of generality, the answer didn't extend the answer beyond the scope of the question
  Thus, it should be rated as (4).

Question: What was your worst date and why?
Answer: Went out with a girl, should've know from the start of her responding in like 1 word messages and hardly ever also, went out, she barely talked, I had to initiate every conversation, totally not worth it, I could've gotten more conversation out of a rock

Explanation: This answer satisfies all the first 4 conditions. The answer is accurate, clear and complete for the question, but it is not concist enough.

- (5)Very helpful

  Question: ELI5: How does a salt water rinse work for toothache/inflamed gums?
  Answer: Saltwater is simply a more gentle form of antiseptic rinse. It also acts as a mild anti-inflammatory, as well. If you cannot tolerate the Clorhexidene, then saltwater is the way to go. I do it whenever I happen to eat something too hot (temperature) that burns the inside of my mouth. Soothing to the skin.

  Explanation: This answer satisfies all 5 conditions. Accuracy, Clarity and Completeness are obvious. In terms of examples, the answer gives an example of how he/she feels when using a salt water rinse for toothache/inflamed gums. As for Generality, the answer further explains what salt water can be used for and how it compares with other antiseptic rinses.Thus, it should be rated as (5).

  Question: What are some embarrassing things your parents did in public?
  Answer: My mom always argues with cashiers and salesmen even when there is no reason for it. She's the kind of person who gets offended when they're offered a bigger size of clothes. So when we have to go shopping together I have to control her all the time.

  Explanation: This answer satisfies all 5 conditions. Accuracy, Clarity and Completeness are obvious. This answer provides a good case that is "embarrassing" that answers the question. I could read it easily and get the feeling of the author.