Supplementary Material: Measuring Similarity of Opinion-bearing Sentences

Wenyi Tay[1,2]**, Xiuzhen Zhang**[1]**, Stephen Wan**[2] and **Sarvnaz Karimi**[2]

[1]RMIT University, Australia
[2]CSIRO Data61, Australia
{wenyi.tay, xiuzhen.zhang}@rmit.edu.au
{Stephen.Wan, Sarvnaz.Karimi}@data61.csiro.au

1 Instructions for Annotations

Table 1 presents the instructions provided to annotators to annotate sentence pairs on Amazon Mechanical Turk[1].

2 Quality Control of Annotations

We calculate the accuracy of every annotator's annotations on a pre-defined set of quality control sentence pairs. A quality control sentence pair consists of a review sentence and a copy of itself with its sentiment bearing word swapped to its synonym or antonym according to Wiktionary. A quality control sentence pair is inserted at random positions to each annotation page. We keep annotations by annotators whose accuracy on quality control pairs is above 70%.

Anomalous annotators are annotators who consistently disagree with other annotators. This disagreement is measured as follows. For every annotated sentence pair of an annotator, we calculate the mean score by all other annotators. We then calculate the Root Mean Square Error (RMSE) of all the sentence pairs of each annotator. We choose RMSE to penalise large differences between an annotation and the mean. Higher RMSE suggests that the annotator disagrees more frequently with the other annotators. We consider annotators to be anomalous when the RMSE is above the threshold of 1.0, which is set empirically. We exclude all annotations by anomalous annotators.

---

[1]https://www.mturk.com/

Level	Explanation	Example
4	Completely same opinion	S1: Serves really good sushi S2: The sushi here is delicious!
3	Mostly similar opinion on the same topic(s)	S1: and the waiter suggested a perfect sake!! S2: The service was courteous and attentive.
2	Partially similar opinion, but some topic(s) differs or missing	S1: Great value sushi with high quality nice setting. S2: This is some really good, inexpensive sushi.
1	Different opinion but on the same topic(s)	S1: The service ranges from mediocre to offensive. S2: The service was courteous and attentive.
0	Completely different opinion	S1: I am never disappointed with there food. S2: Our server was very professional and friendly.

Table 1: Explanation and examples of opinion similarity levels between sentence pairs S1 and S2.