

Instructions for creating “consensus” categories

(v9)

Background

We are analyzing a dataset of texts in order to identify categories of things being discussed. Examples of texts that can be analyzed in this way might include sets of open-ended text survey responses, social media posts, conversational turns in interview transcripts, or public comments to a federal agency.

The analysis we’ve done so far involves two steps. First, there was an automatic computational analysis of the whole dataset to produce “topics” (sometimes also referred to as “themes”) that represent categories or concepts found in the collection of text. This first step is fast and automatic, but it doesn’t give names or labels to the topics that were found, and it also doesn’t provide useful, readable descriptions of what the categories are about.

In the second step, two subject matter experts looked independently at the automatic analysis from the first step. They both followed a set of step-by-step instructions in order to assign a name or label to each category, along with writing down free-form descriptions and additional notes about each one to describe what the category was about. If you have prior experience with coding and codeset development, you can think of what they did as a machine-assisted process of codeset discovery, in order to independently produce a set of meaningful codes.

Objective

The next step in the process is to look at what the two subject matter experts (SMEs) produced independently, and combine their results into a single “consensus” analysis of the categories that were discovered. Sometimes we ask the original SMEs to work together to do this, sometimes we ask someone new to do this part, and sometimes we ask two new people to work together. By the time you’re reading this, we should have already let you know what the plan is.

More specifically, the spreadsheet you’re getting contains the two independent sets of category labels along with corresponding descriptions/notes. Based on these, for every category, the goal is to come up with a single name or label that captures the topic or concept, and to provide a free-form description or notes including any compromises or considerations you made during the process. Optionally, as discussed below, you may also choose to group the topics into higher-level categories.

As an example, in a previous study, we analyzed writing by people related to their mental health, talking about how they managed to get through difficult times. Here is the analysis for topic category #14 in that study by two independent mental health subject matter experts.

Topic Name	Description
Little Pleasures of life	People really focused on small, tactile or even mundane things here. Garlic bread, hot chocolate, ice cream, watching Bob ross, pizza, the sunset, spicy food, flowers

Topic Name	Description
Anticipation of food	Looking forward to pleasure related to food. Favorite food, new food, food as a reliable pleasure.

A good consensus topic name/label for this topic might be “Food and sensory pleasures”.

Optionally, there will be the opportunity to create a higher-level category that contains multiple topics. For instances, in the mental health study the people doing what you’re doing decided that topic #14 in the above example, together with some other topics, grouped naturally into a higher-level category they called “Activities or events that provide enjoyment, distraction, anticipation, an alternative focus”.

The materials you’ll be working with

Input: two labeling spreadsheets from the two independent subject-matter experts (SMEs). Along with these instructions, you should have received two spreadsheets with topic names and descriptions similar in format to the example above.

Topic	Coherence (1=low, 3=high)	Topic Name	Description
Topic 1			

Each labeled topic name and description is followed by information that was used by the independent SMEs to curate the topics to produce names and descriptions: a bar chart showing words and phrases that are characteristic of the topic, and after that a set of rows showing texts (documents, verbatims) that are characteristic of the topic.

Output spreadsheet: You should also have a spreadsheet (probably named coding_scheme_template.xlsx) that you will fill in as part of this consensus process.

Topic	Code Name/Label	Description	Notes/Comments
1			
2			
3			
4			

Note that the independent SMEs have already done their work creating labels and descriptions for the topic, and your job here is not to do the same work over again. You should only drill

down into the bar charts and documents/texts *if you need to* in order to help you understand what the previous SMEs were getting at with particular topics, e.g. if their labels or descriptions are very unclear or hard to understand or if they seem to conflict significantly with each other. Otherwise please focus on the labels and descriptions, not the original materials from which they were created. We recommend deciding in advance on a maximum amount of time you'll spend per topic, and setting a timer when you start working on each topic. In past experience we've found that 2 to 5 minutes per topic on average is sufficient to arrive at good results.

What you should do

First, please keep track of the time you spend working on this (excluding breaks). We don't expect that it should take you more than a few hours, but knowing how long you actually took will be useful information for other studies.

Please follow these steps:

1. Open the spreadsheets with the topic names/labels from the two independent SMEs. Optionally but recommended: open a document (Word, Google Doc, etc.) to keep track of notes from your discussion, e.g. topic-by-topic notes where you summarize what made it easy/difficult to reach consensus.
2. First pass. This step is designed to *efficiently* identify the cases where it's easy to spot that the SMEs were agreeing with each other. Possibly they've both agreed that topic is incoherent and should be discarded, or possibly they've both identified basically the same concept, and it's easy to come up with a consensus name and description.

Go through the topics one by one. For each one:

- a. Set a timer for 3 minutes
- b. Read the two SMEs' topic labels, descriptions, and coherence scores for this topic.
- c. See if you can construct a single label and a single description/notes to capture the consensus of what the two SMEs were seeing in this category, informed by that information. (Again, note that your consensus label might simply be DISCARD to say that this topic is not providing any sensible code to include.)
- d. If it's hard to understand what the SMEs were getting at, or if they seem like they're significantly in conflict with each other and more discussion is needed, make the consensus code/label UNSURE and move on to the next topic. (For this try hard to *not* go beyond the timer, even if you're tempted to. You'll be looking at the UNSURE cases again on the next pass.)

At the end of this pass, every topic should *either* have a consensus code name and description, *or* it should be labeled as UNSURE. It is *very* typical to have unsure topics,

sometimes more than a few, although if you chose a reasonable granularity you should also have quite a few where it was relatively fast and easy to establish agreement.

(If most of your topics are UNSURE or the SMEs thought they mixed different concepts into the same topic too much of the time, that's an indication that the granularity in the automatic analysis might have been too coarse grained, and/or that the dataset was too small or too varied to find good topics.)

3. Second pass. Now you're going to go through the UNSURE topics to get to consensus on what to do with them.
 - a. Set a timer for N minutes, where N is roughly the amount of time you want to allocate to sorting out what to do with the UNSURE topics, divided by how many of those there are. For example, if you have 6 unsure topics and you want to spend no more than an hour, set your timer for $60/6 = 10$ minutes. Unlike the first pass, you don't *have* to be strict about respecting the timer, but having a timer remains a good way to at least be aware of when discussions are taking longer than you wanted them to, and contributes to efficiency.
 - b. Same as the first pass: Read the two SMEs' topic labels, descriptions, and coherence scores for this topic.
 - c. Same as the first pass: See if you can construct a single label and a single description/notes to capture the consensus of what the two SMEs were seeing in this category, informed by that information.
 - d. If it's hard to understand what the SMEs were getting at, or if they seem like they're significantly in conflict with each other:
 - i. Look at the words associated with the topic (the bars) and also at the most representative documents for this topic.
 - ii. Update your own understandings of what this topic might be about, *informed by* what the previous SMEs thought. Note that it's important that you do not start from scratch here – the goal is to get agreement on what the topic represents, not to re-do the original process.
 - iii. Finally, discuss and see if you are *now* able to construct a single label and a single description that captures what the two SMEs were seeing in this category.
 - e. If you are still unable to arrive at a single consensus label and description that seem reasonable, use the label DISCARD and if appropriate update the description of the problem you had.
 - f. Your coding scheme is now the set of codes that were not labeled DISCARD.
4. Optional reliability check and refinement. Often the steps above will yield categories that are suitable for downstream use, e.g. reporting or coding. However, if you wish to take further steps to ensure rigor, you could use these codes as a head start on the traditional inductive coding process, where you've already efficiently generated the first-pass code set. For example, one would typically code a sample of items, compute inter-coder agreement (e.g. Krippendorff's alpha), and iterate on your definitions of the codes and/or

expand your coding guidelines until a sufficiently high level of agreement is reached. Or, of course, one could simply make a judgment call for unsure cases, informed by the requirements of your specific analysis.

5. Optional grouping into themes. After you've created your single-level coding scheme, you may choose to add a further level of hierarchical organization by grouping codes together. The illustration above provided an example where could both be assigned a higher-level category you might name "Activities or events that provide enjoyment, distraction, anticipation, an alternative focus".

Thank you for your help on this project!