Annotation guideline Relation Extraction ADR-Coping Strategy

By: Femke Voorwinden, Dorien Renting, Jelle de Boer

Date: Jan 2021

The dataset that will be annotated comes from the GIST International Support Facebook forum. GIST stands for gastrointestinal stromal tumor and is a type of tumor that occurs in the gastrointestinal tract, most commonly in the stomach or small intestine. Experiences from patients and relatives are shared on this forum about the disease in general, but there are also many reports about medication and possible side effects. These messages were already made suitable for annotation at an earlier stage. The names of the people posting the message have been anonymized, but names in the messages themselves have not been removed. Only typos have been corrected in the messages.

From this dataset, 4195 text snippets with annotations, 808 text snippets have a identified strategy for side effects and how they were dealt with. The focus will now be specifically on which strategy, according to the contributors on the forums, was effective for which side effects.

The main goal of the research is to develop a good method for automatic linking of the side effects to the coping strategies. Human annotation of a subset of the data (100 text units) is used as training data for the computer. These guidelines specify the rules the annotators need to follow in order to annotate correctly and reliably. The inter-annotator agreement (IAA) will be calculated afterwards to see whether these guidelines are reliable.

Based on a small test sample of 12 text units of our dataset, we decided on several rules for the annotation guideline. These are based on the annotations previously made and we have composed several possible situations that arise when linking a side effect to a strategy.

How to handle previous included messages

In most cases, the text is split up into different messages to provide context before the person mentions the strategy. We have decided to consider the entire text including context when linking ADRs to the strategy. This is a maximum of 5 messages together, that are providing context to the message where the strategy is mentioned. All side effects that can be linked to the strategy in the text will be linked. Example:

"P1: I am on imatinib 400 mg . I live 20 miles away from Charlotte P2:I take my medicines at night so I sleep through most of the nausea. […] I do have ondansetron but like to find other ways to deal with side effects before I absolutely have to take it."

- In this case, there are two messages (P1 and P2), and there is only one side effect in the text to which the strategy can be linked. So we will annotate this to be linked.

Different wordings of the same side effects will also be linked.

How to handle no fitting side effects

When there are no side effects mentioned in the text that are solved or helped by the strategy, we don't link any side effects to the strategy. We determine whether a side effect can be linked to the strategy pretty strictly. We do not base this on our own interpretations, but try to stay as close to the text as possible. We link a side effect to the strategy when someone explicitly indicates that the side effect is helped or solved by the strategy.

More than one fitting side effect

In some cases, there are multiple side effects that can be traced back to the same strategy. Since we do not want to distinguish between whether one side effect is more important than the other, we decide to annotate all side effects helped by the strategy. Even if someone indicates that one is solved and one only helped slightly, we don't distinguish between this.

Example:

" P5: I took my imatinib at bedtime also ... It helped with being tired and the nausea."

- In this case, the strategy taking imatinib at bedtime helped with both tiredness and nausea. We will annotate both to be linked to the strategy.

Wrongfully annotated data

Since we are working with an already annotated dataset, it can be the case that we do not agree with an annotation that is made. We can disagree on two points; on the content of the annotation and on small errors. For both we have decided to not change the annotations, even if we think there was more potential to link a side effect to a strategy if for example that side effect was not annotated as an ADR. We decided upon this after quite some annotating and finding that allowing ourselves to change the annotations made the accuracy much lower. Therefore, we will only use the annotations as they are presented to us in the data.

Example:

"Lately though my stomach just feels miserable at night before I even take my medicines"

- In this case, we did not agree with the annotation of this being the side effect. In our opinion it should be that the person's stomach feels miserable. However, this is a change in the content of the annotation so we will not change the annotation.

Example two:

"I am less tired on it than I was in the early years."

- Even though it looks cleaner to only annotat 'tired', we have decided to not change the annotation and just link 'tired on' as a whole.