

Guidelines for Annotating Biomedical Tweets

Amelie Wühl

1 Introduction

We use the following guidelines to annotate biomedical claims in tweets. The annotation procedure and the resulting corpus is described in Wühl and Klinger (2021).

- Tweets are annotated with a binary label that indicates whether or not the tweet contains (a) claim(s) or not.
- Explicitness/Implicitness is annotated if the tweet is labeled to contain a tweet.
 - Explicit cases: mark the token sequence within the tweet that makes up the claim.
 - Implicit cases: the claim which can be inferred from the implicit utterance is stated alongside the implicitness annotation. The claim should be relatively easy to infer and no excessive amount of external knowledge should be needed.

2 Guidelines and Examples

1. Characteristics of the claim:
 - The claim is the central statement of the argument structure.
 - It is the conclusion of the voiced thoughts or opinions.
 - It is stated as true, but objectively can be either true or false.
 - It contains opinion and clear stance (support/oppose) about the discussed controversy.
 - The tweet containing a claim has a relatively clear argumentative intent, meaning the user has written the tweet intending to make an argument or share an argumentative opinion.
 - Claims can be voiced implicitly, e.g. by using a rhetorical question, sarcasm or irony.
 - Argumentation on Twitter can often be incomplete, but traditionally the claim is accompanied by a premise element. Premises are used to substantiate the claim, they give reasons/evidence for the conclusion. The presence of premise elements can be a hint that there's also a claim:
 - Types of evidence/premises: studies, testimony by an expert or authority, anecdotal evidence, concrete examples
2. We assign the claim label to a tweet, if it contains a claim about a biomedical issue. For those instances we further assign the claim type that is used –explicit or implicit. For tweets that contain both an explicit and implicit claim: If what appears to be the main claim can be categorized as mostly explicit, it is annotated as explicit. In other words, if the claim conveyed is not completely implicit, we annotate it as explicit.

3. We consider each tweet as a self-contained document. However, for tweets that are obviously part of a thread or longer discussion, tweets that explicitly reference e.g., an attached image and the statement can not be understood without looking at the item referenced, we follow the URL to the original tweet before annotating. Whenever possible, the tweet under consideration/the annotated claim should be as self-contained as possible.
4. Sometime it can be helpful to check who the author is or who the author is mentioning in their tweet. This can help determine if there is an argumentative intention. If someone is tweeting at a politician, it might be more likely that they are being argumentative. But generally, there should be no extensive research necessary to understand the tweet.
5. When in doubt, we annotate as claim. Borderline cases in which maybe both cases (claim or non-claim) can be argued are annotated to be claims.
6. If a claim spans a whole sentence, punctuation, emoticons etc. should be included in the token level annotation. When in doubt about the span of the claim, we include introductory words and phrases like 'I guess'.
7. According to the definition of the claim being the conclusive part of an argument, what presents itself as the 'main point' of a tweet is most likely to be the claim. In some cases, there seem to be multiple sentences or subsentences sounding like a claim when looked at in isolation, but they are in fact used to premise, lead up to or back up the conclusive claim. These elements should not be annotated as explicit claims or added to the span of the explicit claim. In the examples below the conclusive claim in italics. Even though in both cases the other sentences in the tweet read like claims at first, they are serving as premises to support the conclusive part of the arguments brought forward in the tweets.
 - Ex. 1: "[@Michael_NYC1024](https://twitter.com/Michael_NYC1024) [@KTHopkins](https://twitter.com/KTHopkins) *Just because ONE scientist says different, doesn't make it true.* In 1998 Dr Andrew Wakefield stated there was a direct link between MMR vaccine and Autism. The work was later discredited. Look it up or don't bother - I don't care (to misquote you)"¹
 - Ex. 2: "[@Raysie_1](https://twitter.com/Raysie_1) [@KillAuDeepState](https://twitter.com/KillAuDeepState) Sorry, I don't agree. I have only seen manipulated data regarding measles vaccine for example. *Actually measles became more prevalent in the 90s since MMR started.* Peace"²
 - Ex. 3: "[@jenndawn](https://twitter.com/jenndawn) [@LARideShareGuy](https://twitter.com/LARideShareGuy) Wrong! Vaccine does not equal safe! Many people refuse vaccines now which has led to resurgence of illnesses such as measles which were eradicated for decades. *A COVID 19 vaccine will help to decrease the spread but will not guarantee 100 % safety.*"³
8. In case there are in fact multiple conclusive independent claims in one tweet, we annotate from the onset of the first claim to the last token of the last claim.
9. In general, what appears to be the intention of the author should be taken into account. This means considering whether or not a tweet seems to have been written with the intent to be argumentative or to convince the reader with their post. This often is the case when there is an argument or claim being made about a personal experience. In Ex. 4 the last sentence could be a potential claim, but the fact the author used the hashtag 'MondayMotivation' makes the tweet appear more like an encouraging message than a claim.

¹<https://twitter.com/karldm1/status/1262039297999413248>

²<https://twitter.com/twinturbowp99/status/1263625748876673025>

³<https://twitter.com/jleighz/status/1263093182478385152>

In contrast to that, the author of Ex. 5 appears to have written their posting with the intention of making claims about the side effects of their medication for depression:

- Ex. 4: "Been aff antidepressants a hale month iday after bein on em for 2 year Been a challenging couple year, looked my happiest fan ive been at rock bottom. Still get my off days, but at least ivery day isna a bad day ony mare #MondayMotivation"⁴
- Ex. 5: "@commiepeasant Have you been playing SSRI bingo? I started on prozac but it gave me heartburn, tried stratterra but it made me sweaty and angry, ended up with paxil that's been working all right for awhile now. maybe worth looking at different instead of more"⁵

10. Some tweets contain quotes which are or appear to be claims. Generally, quotes can be annotated as claims if they represent the conclusive element of the Twitter user's argument or if by using the quote someone is making an implicit claim. Often times, however, quotes are used to support or substantiate the tweet author's opinion, in which case the quote should not be annotated as claim even though it might be formulated as one:

- Ex. 6: "' coffee acts as antidepressants ' ok great i'm on my second cup now"⁶

11. In general, argumentative tweets which are off topic but are part of the dataset because they contain a certain keyword, should be annotated with claim = False. If the conclusive claim is not on topic, but there is a strong, directly formulated and highly relevant claim made along the way – maybe used as a premise element – which is on topic, we can annotate it as an explicit claim.

- Ex. 7: "@millerniumtwit @BBCNews Unfortunately people don't like unpalatable truths (e.g. man-made climate change is very real and an existential threat to humanity, the *MMR vaccine does not cause autism but does prevent three potentially severe illnesses*) and prefer to look for reassuring lies and conspiracies."⁷

12. Hashtags can be explicit claims if they are the overarching conclusion to the tweet.

- Ex. 8: "TBT to 1963, when measles caused approximately 2.6 million deaths each year. Thanks to vaccination efforts, measles deaths dropped by 80% between 2000 and 2017. #VaccinesWork #vaxfacts #worldimmunizationweek"⁸

References

Wühl, A., & Klinger, R. (2021). Claim detection in biomedical twitter posts [accepted at BioNLP 2021].

⁴<https://twitter.com/paiigediixon/status/1259784027353092096>

⁵<https://twitter.com/SAmalgamated/status/1262980894899367937>

⁶<https://twitter.com/higabear/status/1262966299531120642>

⁷<https://twitter.com/FolkestoneRoss/status/1265630577769754625>

⁸<https://twitter.com/MHCCMD/status/1255928695270891520>