Claim Detection in Biomedical Twitter Posts as a Prerequisite for Fact-Checking

Amelie Wührl and Roman Klinger
Institut für Maschinelle Sprachverarbeitung, University of Stuttgart, Germany
{amelie.wuehrl, roman.klinger}@ims.uni-stuttgart.de

Social media contains unfiltered and unique information, which is potentially of great value. Biomedical insights and discussions are no exception here: patients report on their experiences with particular medical conditions and drugs, discuss and hypothesize about the potential value of a treatment, and doctors share insights from their everyday life.

However, with regards to such topics, false information, unproven claims or even intentionally spread misinformation can be particularly dangerous [3]. It is therefore essential that social media posts are contextualized e.g., by providing additional information. This could help inform people if a medical statement can actually be proven with a reference to a reliable source. For example, in the tweet in Fig. 1 the user claims that the drug Ivermectin helps treating Covid. Ideally, we want to provide readers of this post with a trustworthy source that substantiates or in this case refutes this claim [4]. Methods of automatic fact-checking and fake news detection address this problem, but have not been applied to the biomedical domain in social media yet [2].

With our contribution [5], we aim to fill this research gap and annotate a corpus of 1200 tweets for implicit and explicit biomedical claims – the latter also with span annotations for the claim phrase. We sample the corpus to be related to COVID-19, measles, cystic fibrosis, and depression, and subsequently develop baseline models which detect tweets that contain a claim automatically. With this dataset we contribute the first resource for claim detection in biomedical tweets. Claims are considered the conclusive and central statements in arguments [1], consequently making them the most valuable information to extract. This is a central task in argument mining and an essential prerequisite for fact-checking or hypothesis generation.

Our analyses reveal that biomedical tweets are densely populated with claims (45 % in a corpus sampled to contain 1200 tweets focused on the domains mentioned above). Table 1 provides examples from the dataset. The majority of claims (68 %) in our corpus are explicit like Ex. 1 and 2 in Table 1. The other instances of the claim class express the claim implicitly. They often use irony or sarcasm like Ex. 3 and 4 in Table 1.

Baseline classification experiments with embedding-based classifiers and BERT-based transfer learning show acceptable performance for detecting claim tweets (.70 F_1). When predicting the claim type, we find that detecting tweets with implicit claims is substantially more challenging (.36 F_1) than detecting explicit claim tweets (.59 F_1). Further, we find in a cross-corpus study that a generalization across domains is challenging and that biomedical tweets pose a particularly difficult environment for claim detection.

Figure 1: Tweet claiming Ivermectin successfully treats Covid. Source: https://twitter.com/Griff_GOP/status/1441154024774914054

id	Instance
1	Latest: Kimberly isn't worried at all. *She takes #Hydroxychloroquine and feels awesome the next day.* Just think, it's more dangerous to drive a car than to catch corona
2	Lol exactly. It's not toxic to your body idk where he pulled this information out of. *Acid literally cured my depression/anxiety I had for 5 years in just 5 months (3 trips).* It literally reconnects parts of your brain that haven't had that connection in a long time.
3	Hopefully! The MMR toxin loaded vaccine I received many years ago seemed to work very well. More please!
4	Wow! Someone tell people with Cystic fibrosis and Huntington's that they can cure their genetics through Mormonism!

Table 1: Examples of explicit and implicit claim tweets from the collected dataset. Explicit claims are in italics.

References

[1] Johannes Daxenberger et al. "What is the Essence of a Claim? Cross-Domain Claim Identification". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Copenhagen, Denmark: Association for Computational Linguistics, Sept. 2017, pp. 2055–2066. DOI: 10.18653/v1/D17-1218. URL: https://aclanthology.org/D17-1218.

[2] Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. *A Survey on Automated Fact-Checking*. 2021. arXiv: 2108.11896 [cs.CL].

[3] Neema Kotonya and Francesca Toni. "Explainable Automated Fact-Checking for Public Health Claims". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, Nov. 2020, pp. 7740–7754. DOI: 10.18653/v1/2020.emnlp-main.623. URL: https://aclanthology.org/2020.emnlp-main.623.

[4] U.S. Food & Drug Administration. *Why You Should Not Use Ivermectin to Treat or Prevent COVID-19*. Sept. 2021. URL: https://www.fda.gov/consumers/consumer-updates/why-you-should-not-use-ivermectin-treat-or-prevent-covid-19 (visited on 09/27/2021).

[5] Amelie Wührl and Roman Klinger. "Claim Detection in Biomedical Twitter Posts". In: *Proceedings of the 20th Workshop on Biomedical Language Processing*. Online: Association for Computational Linguistics, June 2021, pp. 131–142. DOI: 10.18653/v1/2021.bionlp-1.15. URL: https://aclanthology.org/2021.bionlp-1.15.