LCHQA-Summ: Multi-perspective Summarization of Publicly Sourced Consumer Health Answers

Anonymous ACL submission

Abstract

Community question answering forums provide a convenient platform for people to source answers to their questions including those related to healthcare from the general public. The answers to user queries are generally long and contain multiple different perspectives, redundancy or irrelevant answers. This presents a novel challenge for domain-specific concise and correct multi-answer summarization which we propose in this paper.

1 Introduction

Community Question Answering (CQA) platforms like Yahoo!Answers, Stack Exchange, Reddit, Quora, etc., are vast repositories of question-answer pairs where common people ask questions as well as contribute answers across various domains. One such domain is healthcare. People not only seek answers from experts but also from the general public which is facilitated by these websites. The reasons for sourcing laymen contributed answers could be to avoid the use of medical jargon in the language used by the experts (Boyd et al., 2018), opportunity to freely express themselves (Park and Conway, 2018) and share their experiences (Alvaro et al., 2015). The posts also give a fair idea of public opinion on specific health issues (Odlum and Yoon, 2015). However, often these answers are long-winded and irrelevant. These challenges necessitate summarization of answers in CQA forums, especially for healthcare domain which directly impacts the well-being of people.

Majority of the existing works in answer summarization is in the general domain (Liu et al., 2008; Fabbri et al., 2019, 2021). There has been a limited study towards summarizing answer in the healthcare domain (Savery et al., 2020; Abacha et al., 2021; Demner-Fushman et al., 2020), which is confined to expert sourced answers. To the best of our knowledge healthcare related question-answers from CQA forums have not been harnessed yet. To bridge the gap, we bring forward an abstractive multi-document summarization approach for consumer health answer summarization. We also observe that these answers present several perspectives. For example, in Table 1, Answers 1 and 3 describe the the cause of hay fever symptoms while Answer 2 shares a personal experience and possible treatments. Answer 4 provides some suggestions that can potentially solve the problem. This motivates us to tackle the summarization problem while covering the different perspectives as done by (Fabbri et al., 2021).

Towards this, we frame our research objectives as follows: (i) Develop a novel gold standard Laymen-sourced Consumer Health Question Answer Summaries (LCHQA-Summ) dataset with summaries covering the breadth of perspectives across various healthcare topics. (ii) Propose an automated health answer summarization pipeline to generate perspective-specific answer summaries.

2 Proposed Plan of Research

2.1 Data Collection and Annotation

We begin by collecting dataset from popular CQA forum –Yahoo! Answers[1]. In particular, we plan to use Yahoo! L6 corpus that consist of 4.5 million questions across different topics, their answers and metadata such as question categories, number of answers, best answer, date, language etc. Since, our goal is focused on consumer healthcare domain, we selected the “Health" category which has 21 subcategories like Allergies, Diabetes, Heart Diseases and so on. It is also necessary to remove outliers in terms of number of answers which can range from as low as zero and as high as 2235 answers in response to a single query. We finally retain posts where number of answers range from $4 - 6$. The final data includes 77K question-answer pairs. To curate a gold dataset of manually written multi-

[1] http://answers.yahoo.com

Proceedings of the First Workshop on Natural Language Generation in Healthcare (NLG4Health), pages 23 - 26

Question	Why are my hay fever symptoms worse early in the morning and how I can stop suffering the first two hours after I wake up?
Context	Allergies
Answer 1	It's because the pollen counts are higher in the morning. Plants release their pollen earlier in the day, thus anyone with hayfever will find this part of the day more annoying.
Answer 2	I have similar problems. When I wake up I have a stuffy nose but then in like an hour or two and I'm fine. I take Zyrtec every morning but before I go to bed I take a Benadryl and that seems to help.
Answer 3	Because pollen is released early in the day, rises with the warm air and falls again in the evening.
Answer 4	It may help if you wash your hair in the evening to get rid of any pollen that might be left in there.
Summary Perspectives:	
Perspective 1	Plants release pollen early in the day.
Perspective 2	Pollen counts are higher in the morning.
Perspective 3	I have similar problems for an hour or two after I wake up.
Perspective 4	Taking Benadryl before bed and Zyrtec in the morning has helped me.
Perspective 5	Washing your hair at night can get rid of any left-over pollen.

Table 1: An example illustrating question, context and answers from Yahoo! L6 dataset. This is followed by an abstractive summary of the answers showcasing 5 different perspectives.

perspective abstractive summaries from the data we sample a subset of the data and put forward the following annotation strategy:
(1) Validate if a question is related to medical domain or not, that is if it pertains to diseases or conditions, drug or treatment, medical diagnosis or therapeutic procedure, any other related medical topic. This helps to weed out any irrelevant question, especially in more generic sub-categories like "Other-Health" and "Other-Health & Beauty".
(2) The next step is to generate abstractive multi-perspective summaries of answers to valid medical question. Based on our preliminary dataset analysis, we have identified 6 major perspectives—*information*, *cause*, *treatment*, *suggestion*, *experience* and *clarification* that describes most of the consumer answers. Example of such summary is shown in Table 1, where perspective 1 and 2 describe cause of the problem, 3 and 4 narrates experience as well as treatment and 5 suggests solution.

2.2 Automated Summarization Pipeline

For obtaining system generated multi-perspective summaries of consumer health answers, we devise a three-step pipeline described next.

Relevant Sentence Extraction: This step is to be applied at the sentence level with the goal of finding the answer sentences that are relevant to the question. As a baseline we would use BM25 (Robertson et al., 1994) to compute relevance of each answer sentence to a given question and retain those with score above a threshold as relevant. A similar approach is measuring semantic similarity between the embeddings of each answer sentence and question using cosine similarity or mutual information. For this we propose to use Sentence-BERT(Reimers and Gurevych, 2019) (SBERT) and UmlsBERT (Michalopoulos et al., 2021) representations.

Perspective Type Identification: Allocating perspective labels to a relevant answer sentence can be treated as a multi-label classification problem (For example, Perspective 4 in Table 1 can be both an *experience* and a *treatment*). Given the success of transfer learning along with zero-shot and few-shot approaches in text classification (Chalkidis et al., 2020; Zhang et al., 2019), we propose to adapt a Natural Language Inference (NLI) based transfer learning approach as done by (Yin et al., 2019) for assigning perspective labels to the sentences. Based on the performance of this method we would also experiment with more refined rules to improve performance across specific classes.

Summarization of answers: In the final stage of the pipeline, we aim to propose a perspective-guided multi-document answer summarization approach focusing on answer summary generation conditioned over the given perspective. We plan to infuse the perspective in terms of the external knowledge to the pre-trained encoder decoder models such as BART (Lewis et al., 2020) and T5 (Raffel et al., 2020) which has shown state-of-the-art performance on the answer summarization task(Yadav et al., 2021; Mrini et al., 2021). Towards this, we will begin by inducing perspective information into the encoder as well as decoder to train the model which incorporates the underlying perspective while generating the summary.

3 Conclusion

Overall in this paper we present the novel problem of multi-perspective abstractive answer summarization from CQA forums focusing on the healthcare domain. We outline a data annotation process, followed by a three-step approach for automatic summary generation with a focus on the perspectives present in these answers.

References

Asma Ben Abacha, Yassine M'rabet, Yuhao Zhang, Chaitanya Shivade, Curtis Langlotz, and Dina Demner-Fushman. 2021. Overview of the mediqa 2021 shared task on summarization in the medical domain. In *Proceedings of the 20th Workshop on Biomedical Language Processing*, pages 74–85.

Nestor Alvaro, Mike Conway, Son Doan, Christoph Lofi, John Overington, and Nigel Collier. 2015. Crowdsourcing twitter annotations to identify first-hand experiences of prescription drug use. *Journal of biomedical informatics*, 58.

Andrew D Boyd, Karen Dunn Lopez, Camillo Lugaresi, Tamara Macieira, Vanessa Sousa, Sabita Acharya, Abhinaya Balasubramanian, Khawllah Roussi, Gail M Keenan, Yves A Lussier, et al. 2018. Physician nurse care: A new use of umls to measure professional contribution: Are we talking about the same patient a new graph matching algorithm? *International journal of medical informatics*, 113:63–71.

Ilias Chalkidis, Manos Fergadiotis, Sotiris Kotitsas, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. 2020. An empirical study on large-scale multi-label text classification including few and zero-shot labels. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7503–7515, Online. Association for Computational Linguistics.

Dina Demner-Fushman, Yassine Mrabet, and Asma Ben Abacha. 2020. Consumer health information and question answering: helping consumers find answers to their health-related information needs. *Journal of the American Medical Informatics Association*, 27(2):194–201.

Alexander Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir Radev. 2019. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1074–1084, Florence, Italy. Association for Computational Linguistics.

Alexander R Fabbri, Xiaojian Wu, Srini Iyer, and Mona Diab. 2021. Multi-perspective abstractive answer summarization. *arXiv preprint arXiv:2104.08536*.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Yuanjie Liu, Shasha Li, Yunbo Cao, Chin-Yew Lin, Dingyi Han, and Yong Yu. 2008. Understanding and summarizing answers in community-based question answering services. In *Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008)*, pages 497–504, Manchester, UK. Coling 2008 Organizing Committee.

George Michalopoulos, Yuanxin Wang, Hussam Kaka, Helen Chen, and Alexander Wong. 2021. Umls-BERT: Clinical domain knowledge augmentation of contextual embeddings using the Unified Medical Language System Metathesaurus. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1744–1753, Online. Association for Computational Linguistics.

Khalil Mrini, Franck Dernoncourt, Seunghyun Yoon, Trung Bui, Walter Chang, Emilia Farcas, and Ndapa Nakashole. 2021. UCSD-adobe at MEDIQA 2021: Transfer learning and answer sentence selection for medical summarization. In *Proceedings of the 20th Workshop on Biomedical Language Processing*, pages 257–262, Online. Association for Computational Linguistics.

Michelle Odlum and Sunmoo Yoon. 2015. What can we learn about the ebola outbreak from tweets? *American journal of infection control*, 43 6:563–71.

Albert Park and Mike Conway. 2018. Tracking health related discussions on reddit for public health applications. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, 2017:1362–1371.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.

Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford. 1994. Okapi at trec-3. In *TREC*.

Max Savery, Asma Ben Abacha, Soumya Gayen, and Dina Demner-Fushman. 2020. Question-driven summarization of answers to consumer health questions. *Scientific Data*, 7(1):1–9.

Shweta Yadav, Mourad Sarrouti, and Deepak Gupta. 2021. NLM at MEDIQA 2021: Transfer learning-based approaches for consumer question and multi-answer summarization. In *Proceedings of the 20th Workshop on Biomedical Language Processing*, pages 291–301, Online. Association for Computational Linguistics.

Wenpeng Yin, Jamaal Hay, and Dan Roth. 2019. Benchmarking zero-shot text classification: Datasets, evaluation and entailment approach. In *Proceedings of*

the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 3914–3923.

Jingqing Zhang, Piyawat Lertvittayakumjorn, and Yike Guo. 2019. Integrating semantic knowledge to tackle zero-shot text classification. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1031–1040, Minneapolis, Minnesota. Association for Computational Linguistics.