Brianna Murray
CS 4395.001

NLP Based Doctor Recommendations

AUTHORS AND AFFILIATIONS

- Xiaoxin Lu from Hong Kong University of Science and Technology has 67 citations
- Yubo Zhang from Hong Kong Polytechnic University has 4 citations
- Ying Li from Hong Kong Polytechnic University has 1109 citations
- Shi Zong

PROBLEM

There are many people around the world who seek out medical care through online portals and platforms. With the lasting effects of COVID-19 on society, searching for medical professions virtually has become more and more common as time progresses. It is quite expensive to manually find a doctor due to the sheer amount of overhead required to check the credentials of each available doctor with the needed expertise to assist a patient. The researchers of this study took on the challenge of atomizing the process of matching patients to their needed doctor of expertise instead of manually being assigned a doctor based on the doctor's capabilities/specialties rather than based on the patient's information and symptoms. This is helpful because all of the patient information has to be anonymous leaving only the symptoms as available reference points.

In past studies of this topic, doctor recommendations were done by creating a modeling target of users that would use past behavior to predict future behavior. However, the issue with this way of doing things is that because the models were being created based on patients, limited information was allowed to be used in order to protect patient privacy. Due to this limitation, making an accurate doctor recommendation was challenging. It should also be noted that the researchers mentioned how the solution they developed is quite unique, which makes it hard to find more information about it within mainstream media.

SUMMARY

It is important to note that the people conducting this research pointed out that, as far as they know, they are the "first to study doctor recommendation to automate the pairing of doctors and patients in online health forums, where the joint effects of doctor profiles and their previous interrogation dialogues are explored to learn what a doctor is good at and how they are able to help handle a patient's request" (Xiaoxin et al).

All data that was used came from Chunyu Yisheng, a Chinese online health forum. Patients are able to use this platform to communicate with doctors virtually. To automate doctor recommendations, researchers took chat conversations from Chunyu Yisheng and made queries that an expertise learning model would use in order to help speed up the process in getting people the care they need. In order to create the proper model, a query was needed. The query represented the patient's ailments which were analyzed to find the semantic relations between patient queries and a doctor's given specialty. Doctor's specialties contained within their profile were then compared with the query in likeness to find the best fit between a single patient and multiple doctors. The profiles of each doctor are composed of their specialities along with all of their dialog history. In order to obtain this history, "a self-learning task is used to predict if a profile and dialog are from the same doctor" using BERT (Bidirectional Encoder Representations from Transformers) which matches every output element to every input, then weights weighs between the elements are dynamically calculated based on their connection (Xiaoxin et al). The dataset being used is composed of 119K patient-doctor dialogues and 359 doctors from 14 departments within Chunyu Yisheng, and all collected with an HTML crawler (refer to Table 1). When evaluating the collected data, they noticed that given dialogues were typically more lengthy than doctor profiles. THUOCL (Tsinghua University Open Chinese Lexicon) medical lexicon was used to evaluate doctors' language styles and to count how many medical terms were used within all conversations.

The framework was built of three modules: a query encoder, doctor encoder, and prediction layer (refer to Figure 3). The query encoder is based on patient queries, the doctor encoder is built off doctor expertise from profiles and dialogues, the prediction layer combines output for both encoders to come up with a recommendation.

# of dialogues	119,128
# of doctors	359
# of departments	14
# of tokens in vocabulary	8,715
Avg. # of dialogues per doctor	331.83
Avg. # of doctors per department	25.64
Avg. # of tokens in a query	89.97
Avg. # of tokens in a dialogue	534.28
Avg. # of tokens in a profile	87.53

Table 1: Data statistics. Each dialogue starts with a patient query and each doctor is associated with a profile.

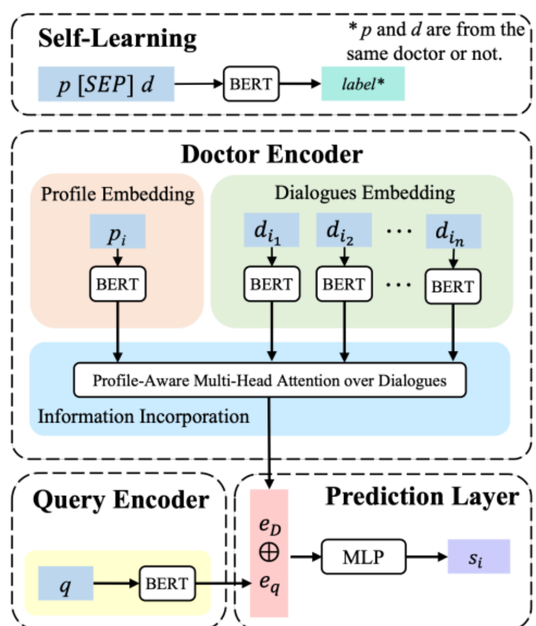

Figure 3: Overview of our framework. The doctor encoder first has its embedding layer (pre-trained BERT) fine-tuned via self-learning. It then employs profile-aware multi-head attention over dialogues to explore doctor expertise and works with the query encoder (to capture patient needs) to pair doctors with queries.

EVALUATION

Before testing the model, the data was preprocessed using an "open-source toolkit jieba for Chinese word segmentation" on non-neural models; neural models were tokenized with "MC-BERT, a pre-trained BERT for biomedical language" (Xiaoxin et al). Then the data was split where 80% of dialogues are randomly selected from each doctor and 20% of dialogues are from the patient query. The model also had a limit of 512 on its input length due to BERT.

Work was evaluated based on weak and neural baselines. Weak baselines for doctor rankings were based on randomness, frequency, KNN, cosine similarity, and gradient boosted decision trees with TF-IDF. Neural baselines were compared with MLP which matches query embeddings with profile embeddings, dialogue embeddings, and average embeddings of profiles and dialogue. Metrics are lastly evaluated with mean average precision.

When analyzing the results, three main observations came about. The first being that the deep semantics required to match patient needs with doctor specialties cannot solely rely on heuristic rules. Secondly, dialogues contain richer content compared to doctor profiles because the language used is more similar to patient queries. Third, there needs to be certain methods in place that effectively join the two writing styles to better characterize a doctor.

Across the board, multi-head attention outperformed its other counterparts but, "self multi-head attention over profile performs much worse than other ablations" (Xiaoxin et al). In addition, query, dialogue, and profile length also resulted in varying results. When it came to query length, researchers noticed that across the board all models performed better with longer queries. Yet, their "full model consistently outperforms" its counterparts along with "showing a relatively smaller performance gain for smaller queries" (Xiaoxin et al). Dialogue length had similar results to query length. However, model performance suffered when profile length was too long.

CHALLENGES

One of the challenges that the researchers faced pertained to the linguistic differences between patients and doctors. They found that doctors tend to use more technical terms within their profiles but switch to speak in layman's terms when using chat messages with patients. In addition, patients also tend to give extra information that a doctor with professional knowledge can skip over but, their created model does recognize these extra words as unimportant details.

An additional challenge arose when reviewing model results that pertained to doctors within certain departments. All models struggled with internal medicine compared to neurology due to the major overlap internal medicine has with other categories of medicine. Yet, on the other hand many models struggled with categories of medicine like Otolaryngology because there were fewer than average results to train on.

Lastly, it should also be noted that some doctors “profile themselves generally from experience” instead of specific specialties which blurs the accuracy of the models. This is because some doctors have a more wide range of general skills while others focus more on their own specialty.

CONCLUSION

The current system is severely outdated primarily because the risk of patient data being exposed through data collection is needed to create an accurate patient model, which is then used to accurately assign a patient to their needed care specialist. This research is important because it will help a vast majority of people to get the care they need as fast as possible. The newly created platform would also be more cost effective than manually assigning doctors to patients. The platform is also revolutionary because of how it is structured. By pulling information from the doctor’s background rather than the patient, the platform is able to gain exponentially more information for a better frame of reference in order to meet demands.

WORK CITED

Xiaoxin Lu, Yubo Zhang, Jing Li, and Shi Zong. 2022. Doctor Recommendation in Online Health Forums via Expertise Learning. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1111–1123, Dublin, Ireland. Association for Computational Linguistics.