

THE RESEARCH ON THE DETECTION OF NOTEWORTHY SYMPTOM DESCRIPTIONS

Yu-Ling Chen, Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan, yulingchen08@gmail.com

Shanlin Chang, Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan, d004020002@student.nsysu.edu.tw

San-Yih Hwang, Department of Information Management, National Sun Yat-sen University, Kaohsiung, Taiwan, syhwang@mis.nsysu.edu.tw

Kai-Sheng Hsieh, Department of Pediatrics, Kaohsiung Chang-Gung Hospital, Taiwan, kshsieh@cgmh.org.tw

Abstract

The advance of mobile devices and communication technologies enable patients to communicate with their doctors in a more convenient way. We have developed an App that allows patients to record their symptoms and submit them to their doctors. Physicians can keep track of patients' conditions by looking at the self-report messages. Nevertheless, physicians are usually busy and may be overwhelmed by the large amount of incoming messages. As a result, critical messages may not receive immediate attentions, and patient care is compromised. It is imperative to identify the messages that require physicians' attention, called noteworthy messages.

In this research, we propose an approach that applies text-mining technologies to identify medical symptoms conveyed in the messages and their associated sentiment orientation, as well as other factors. Noteworthy messages are subsequently characterized by symptom sentiment and symptom change features. We then construct a prediction model to identify messages that are noteworthy to the physicians. We show from our experiments using data collected from a teaching hospital in Taiwan that the different features have different degrees of impact on the performance of the prediction model, and our proposed approach can effectively identify noteworthy messages.

Keywords: Text-mining, Sentiment analysis, Medical symptoms, Noteworthy messages.

1 INTRODUCTION

In traditional clinical settings, the interaction between a patient and a physician occur at the physician's office when the patient visits the physician. However, symptoms of patients may have persisted for quite some time before physicians ever know them, causing the deterioration of their health condition. The advance of computer and communication technologies inspires the emergence of Personal Health Record (PHR). Patients can report and/or share their health conditions and experiences. Some doctors also use PHR to keep track of the health conditions of patients under his care. Of the various types of health data reported by the patients, textual data plays an important role as patients tend to report their detailed symptoms in writing. We have developed an App via smart cell phone with the function of patient-initiated self-reporting about their symptoms/activities related to their own disease management. This App has been used by physicians in a teaching hospital in southern Taiwan for more than two years to enhance patient-physician interaction, and the satisfaction is remarkably high (Hsieh et al. 2014). However, in view of the large number of patients each physician needs to tend to and their busy schedules, messages from patients were seldom checked by their physicians in time. Thus, it is imperative to identify the messages that require the attention of physician, called noteworthy messages in subsequent discussion.

Yu et al. (2011) illustrate compatibility between text mining and grounded theory. Text mining is similar to content analysis from grounded theory and aims to develop a fully automated procedure to extract representative themes. There have been several proposed works that use text-mining technologies to extract information from PHR textual data (Meystre et al. 2008) and predict disease risk (Lan et al. 2012). Lan et al. (2012) propose a disease risk prediction model which predicts chronicle diseases from both health examination records and the life-style datasets. However, this research does not utilize the textual data in PHR. In Zhou et al. (2006), the authors propose an approach to extract the values of some health items, such as blood pressure, pulse, temperature from clinical medical records. However, identification of important messages in a real time manner has not been explored.

We collect more than 10,000 messages from the patients' self-report messages in the division of pediatric cardiology of a hospital in Kaohsiung city, Taiwan. Based on these messages we built a prediction model. Incoming messages that are noteworthy can be identified by the model and provided to the physicians for immediate attention. Our preliminary experimental results show that noteworthy messages can be distinguished by three types of features, namely symptom sentiment features, symptom change features and similarity features. The constructed prediction model is shown to effectively identify noteworthy messages for physicians.

This paper is structured as follows. In Section 2, we describe our data set and the addressed problem in more detail. In Section 3, we present our proposed approach. In Section 4 experiments and the preliminary results are discussed. Finally, Section 5 summarizes this research and describes our ongoing work.

2 DATA DESCRIPTION

We examined the content of some sample messages and found that there are generally four types of information contained in these messages, namely symptoms, administration, therapy, and disease, which are described in more detail below.

2.1 Symptom

This type of information is about the current health conditions of patients. Some symptom descriptions could be important and need attention from physicians. For example, negative symptoms could be noteworthy. However, physicians that we interviewed cautioned us that repeated messages, even with negative symptoms, are less important. For instance, the following message “今天胸口很悶，而且會咳嗽、流鼻水。7/03/2012” (Today I feel chest congestion, and I cough and have a runny nose.) may need attention because it describes some negative symptoms. However, the next message which appears in the next day “今天還是會咳嗽、流鼻水。7/04/2012” (Today I still cough and have a runny nose.), may not deserve attention, because no new symptoms are developed. Thus, symptom data as well as their sentiment and dates, could be critical in determining the noteworthiness of messages.

2.2 Administration

Patients need to make an appointment before seeing a doctor. Sometimes, patients remind their physicians of their future visit in the messages. For instance: “我是王喬，5/4 將會回診” (I’m Joe Wang. I’ll come visit on 4th of May.), “醫生您好，李珍妮 11/23 日要給主任照超音波” (Hello, Jenny Lee will do heart echocardiography on 23th of Nov.). While this kind of messages could be useful to hospital staff, physicians usually disregard them.

2.3 Disease

Some patients describe their disease in their messages. For example, “王喬，3 歲，心雜音(Joe Wang, three years ago, Heart Murmur.)”，“李珍妮，肺動脈狹窄”(Jenny Lee, Pulmonary stenosis.). This type of information often appears at the beginning of the messages that serves as a reminder to the physicians. This type of information is useful only when the doctor determines to read the message.

2.4 Therapy

Patients sometimes describe their therapy progress. Based on the patient’s therapy progress, the physician can adjust the dose of medicine. For instance: “王喬上周已做過胸部 X 光檢查，目前正在接受藥物治療”(Joe Wang had a chest X-ray last Sunday. He is currently undergoing drug treatment.)

After interviewing physicians, we found that the importance of these four types of information vary with symptoms seemingly being the most important. Thus, in designing our method, we focus on the symptoms described in messages.

3 THE PROCESS OF THE RESEARCH

The procedure of our approach is described in Figure 1. The App we developed stores messages sending from patients to their physicians in a central database, from which we collect our data. We apply natural language process (NLP) techniques for cleaning the messages, segmenting Chinese messages into terms, and getting part of speech (POS) for terms. In our experiments, we choose CKIP as our NLP tool (Ma & Chen 2003). To determine the type of symptoms contained in a message, we prepare a symptom lexicon. In addition, we decide on the sentiment associated to a symptom based on a sentiment lexicon. The construction of the two lexicons will be described later. Following our discussion in the previous section, we intend to identify three types of information, namely symptoms, sentiment, and sentiment change features. Figure 1 shows the procedure of our approach.

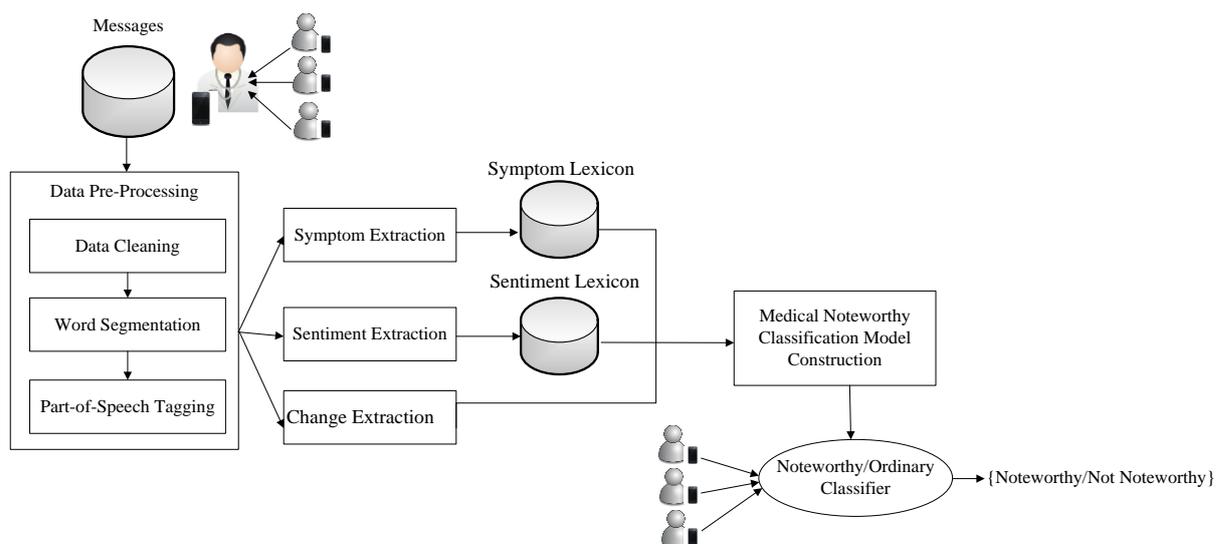


Figure 1. The procedure of our approach

3.1 Construction of Sentiment Lexicon and Symptom Lexicon

As no suitable Chinese sentiment dictionaries are available in the medical domain, we decide to create our own sentiment lexicon. We observe from our collected messages that Taiwanese people tend to describe their symptoms in a more gentle way. These gentle adverbs (ADV) include “有點(a little)” and “還是(still)”, which are often followed by negative sentiment words with POS being intransitive verb or noun. Examples include “胸口有點痛” (My chest is a little pain.) and “還是會咳嗽” (Still cough.). We thus tag the words 痛(pain) (Vi) and 咳嗽(cough)(Vi) as negative sentiment terms. Another term “一切(everything)” (DET) is often followed by positive sentiment words. For example:“一切都很ok” (Everything is O.K.) Thus, “ok” (FW) is regarded as a positive sentiment term. We then manually remove some inappropriate words.

Then we use these collected sentiment terms as seed words for further expansion. We observe that terms that share some character with an existing sentimental term often have the same sentiment orientation. For example, we can use term “痛(pain)” to find more sentiment words like “胸痛(Chest pain)”, “胃痛(Stomach ache)”, and “頭痛(Head ache)”. After expansion, we obtain 184 negative sentiment terms and 32 positive sentiment terms.

Symptom extraction is another important process in our work. Liu (2012) proposed a dependency approach to find feature terms decorated by sentiment words. We follow their approach by using a set of rules to identify a set of symptoms decorated by some sentiment terms. Example rules include (1)*Symptom* (N) + *Sentiment*(Vi)/(Vt). (2)*Symptom*(N) + (ADV) + ... + *Sentiment*(Vi)/(Vt) (3)*Symptom* (N) + 有(Vt) (do,can) /會(ADV) (well) + *Sentiment* (V)/(Vt). For example, “食慾很好”(Appetite is very good.) is a sentimental sentence with 食慾(appetite)(N)+很(very)(ADV)+好 (is good) (Vi). We thus identify “食慾” as a possible symptom. Finally we obtain 193 symptom terms for our symptom lexicon, which are then manually divided into six categories as follows:

- ENT: These terms describe symptoms related to ear, nose and throat.
- Chest: The terms describe symptoms related to chest.
- Cardiology: These terms describe symptoms related to heart.
- Breath: These terms include the symptoms about breathing.

- Gastroenterology: These terms include the symptoms about stomach and intestine system like diarrhea.
- Other: These terms include any other symptoms such as swollen feet, allergies, blood pressure motility, sweating, headache, and edema.

3.2 Representing Messages

Following the Lexicon-based approach proposed in Liu (2012), for each message we determine the sentiment associated with each of the six types of symptoms, namely *ENT*, *Chest*, *Cardiology*, *Breath*, *Gastroenterology*, and *Others*. The domain of each feature is {positive, neutral, negative}. For example, if a message m contains the sentence “今天咳嗽好多了”(Cough is better.), we may derive its sentiment on ENT as $m.ENT_sentiment = \text{'positive'}$. If m does not include any description about a particular symptom, say Chest, then $m.Chest = \text{'neutral'}$.

In addition, as mentioned in Section 2, symptom change seems to be important for physicians. For symptom change, we use six features to express the variations for the six types of symptoms. The domain for the six variation features is {"Unchanged", "Getting better", "Getting worse", "Unknown"}. If a sentiment of a symptom turns from positive/neutral to negative in three days, we mark the symptom change as "Getting worse". On the contrary, if a sentiment of a symptom changes from negative to positive or neutral in three days, we indicate the symptom change as "Getting better". Finally, we add another feature "*sim*" to measure the textual difference between the current message and the last message issued by the same patient, which is measured using the edit distance between two messages.

4 PRELIMINARY PERFORMANCE EVALUATION

We design and conduct some experiments to evaluate the performance of the proposed method. The experimental design and the results are described in the following subsections.

4.1 Experiment Design

We collect more than 10,000 messages sending from patients to physicians in the pediatric cardiology department of a hospital using our developed App. From the 1478 patients who have written messages, we randomly choose 100 individuals who have written more than ten messages. For each message, we randomly choose a starting point and retrieve the subsequent 10 messages. Totally we have 1,000 messages in our training data set. A web site is designed to allow experts to label the messages as noteworthy or not-noteworthy. Three nurses who serve as assistants for the physicians of the patients are employed to classify the 1,000 messages. The nurses use their medical experience and knowledge to classify the same 1,000 messages. Afterward, we choose the results upon the majority rule, i.e., the label voted by two or three nurses become the dominating label of the corresponding message. Of the 1,000 messages, 43 messages are voted noteworthy by all three nurses, 143 messages are voted as noteworthy by two nurses, 118 messages are voted not-noteworthy by two nurses, and 696 messages are voted not-noteworthy by all three nurses. We perform the reliability analysis for the labels given by the three nurses. The reliability analysis value of Cronbach's Alpha is 0.716, which means the reliability of the manual classification is believable.

4.2 Vectorization and Classification

In our research, we have three different types of features for messages of patients, namely symptom sentiment, symptom change, and similarity, as shown in Figure 2.

Symptom sentiment						Symptom change						Similarity
ENT	Che	Car	Bre	Gas	Ext	D_ENT	D_Che	D_Car	D_Bre	D_Gas	D_Ext	Sim

Figure 2. Feature vector of each message

Our data set is skewed in that the numbers of noteworthy messages and non-noteworthy messages are 182 and 818, respectively. To solve the skewed data problem, we adjust the parameter “Weight” of LIBSVM (Chang & Lin 2011) when building the prediction model using SVM. The proportion of noteworthy to negligible messages is approximately 2:8, thus we set the parameter “Weight” 4:1. We choose training data and testing data via 10-fold cross validation.

We use sensitivity, specificity and F-measure as the performance measures. Sensitivity and specificity are statistical measures of performance popularly used in medicine domain. F-measure serves as a combinational measure of the two measures. The definitions of the measures are shown in the following, where TP, TN, FP and FN denote true positive, true negative, false positive, false negative respectively:

$$\text{Sensitivity} = \frac{TP}{TP+FN}, \text{ Specificity} = \frac{TN}{TN+FP}, \text{ and F-measure} = \frac{2TP}{2TP+FP+FN}.$$

We first perform logical regression to compare the three types of features, and similarity is found to be closely related to symptom change. We thus focus our attention on the two types of features, namely symptom sentiment and symptom change. The comparison is shown in Figure 3. It shows that “Symptom sentiment” alone yields much better performance than “Symptom change” alone. However, if we apply both types of features, the sensitivity and specificity can reach 93% and 98% respectively, outperforming either type of features.

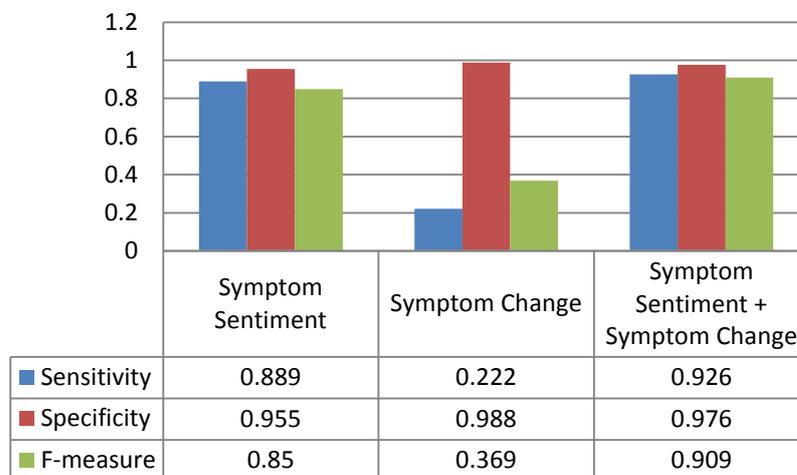


Figure 3. Prediction Performance

5 CONCLUSION

We have proposed an approach to identify noteworthy messages of patients for the physicians. In this research, text mining techniques are employed to analyze the textual messages. We proposed a method to construct lexicons for the symptom and sentiment terms. We then use lexicon-based method to determine the sentiment of a given symptom for each message.

Each message is represented by 13 features, including symptom sentiment, symptom change, and similarity. SVM was employed to train a data set with 1000 messages. The experimental results showed that the combined features reach the greatest sensitivity and specificity.

The good performance of our proposed method is highly impacted by the symptom lexicon and sentiment lexicon. However, the current method of building symptom lexicon and sentiment lexicon involves quite some human effort. We are in the process of developing automatic method for the construction of both lexicons. In addition, more feedback from the hospital professionals is planned.

References

- Chang, C.-C. and Lin, C.-J.. (2011) Libsvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2(3),27.
- Hsieh, K.-S., Hsu, C.-C., Chiang, C.-C., Lee, Y.-B., Lin C. C. and Hwang, S.-Y., (2014) Patient-initiated self-reporting through mobile device---a new mode of physician-patient interaction, In: 142nd APHA Annual Meeting and Exposition, APHA.
- Lan, G.-C., Lee, C.-H., Lee, Y.-Y., Tseng, V. S., Chin, C.-Y., Day, M.-L., Wang, S.-C., Chang, C.-N., Cheng, S.-Y. and Wu, J.-S., (2012) Disease risk prediction by mining personalized health trend patterns: A case study on diabetes, In: *Technologies and Applications of Artificial Intelligence*, IEEE, p. 27-32.
- Liu, B. (2012) Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5(1),1-167.
- Ma, W.-Y. and Chen, K.-J. (2003) Introduction to ckip chinese word segmentation system for the first international chinese word segmentation bakeoff. *Proc. Proceedings of the second SIGHAN workshop on Chinese language processing*. Sapporo, Japan, 168-171.
- Meystre, S. M., Savova, G. K., Kipper-Schuler, K. C. and Hurdle, J. F. (2008) Extracting information from textual documents in the electronic health record: A review of recent research. *Yearb Med Inform* 35),128-144.
- Yu, C. H., Jannasch-Pennell, A. and DiGangi, S. (2011) Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *The Qualitative Report* 16(3),730-744.
- Zhou, X., Han, H., Chankai, I., Prestrud, A. and Brooks, A. (2006) Approaches to text mining for clinical medical records, In: *Proceedings of the 2006 ACM symposium on Applied computing*, ACM, p. 235-239.