Towards Generating Comprehensible Hospital-Stay Summaries for Patients

Sabita Acharya, Barbara Di Eugenio, Andrew D. Boyd and Karen Dunn Lopez
University of Illinois at Chicago
Chicago, Illinois
Email: sachar4, bdieugen, boyda, kdunnl2@uic.edu

I. Abstract

When patients are discharged from hospitals, they are provided with discharge instructions and patient education materials. However, if patients are not able to understand the content of the materials, they will not follow the recommended treatment plans. In order to address this issue, we aim to generate concise and comprehensible hospital-stay summaries. Below are some of the tasks that we have accomplished so far.

Summary Generation: Patients are usually handed over the doctors' discharge notes while the perspective of nurses who monitor the health of the patient throughout their hospital stay are not considered. Since our previous study has shown that nurses and doctors talk about different aspects of patient's health, we include both the perspectives in our summaries. We extract the medical concepts present in the nursing documentation and *hospital course* section of the doctor's discharge notes and query UMLS for any intermediate concepts that are needed to form a connection between the doctor and nurse concepts. These information are couched as features of phrasal constituents and supplied to SimpleNLG API, which then assembles the grammatical phrases in right order.

Identifying follow-up components: Even though discharge notes contain several distinct sections, there is no uniform structure that is followed by all physicians. In particular, the patient follow-up information may appear as a separate section or may be spread across various other sections. We set out to algorithmically recognize follow-up information so that they can be appended to our hospital-stay summaries. For this task, we used a set of 749 de-identifed discharge notes (80% for training and 20% for testing). After thorough inspection of the training set, we came up with an algorithm that consists of 15 regular expressions and a list of 67 keywords. Results obtained after applying the algorithm on test data were compared against the human annotated results. The accuracy of the algorithm was 92.66%, while the precision, recall and fscore values were 0.93, 0.96, and 0.95 respectively.

Aggregation: In order to make the summaries more natural, we performed aggregation based on the type of diagnosis. We noted that categories of diagnosis that involve *risk of a condition*, or *impairment of a body part*, or *ineffective status of some mechanisms of the body*, frequently occur together in the nursing documentation for most of the patients . We mention such terms together as *During your hospitalization, you were monitored for chances of ineffective cerebral tissue perfusion, risk for falls, problem in verbal communication and walking*. We also group the topics for which education was provided to the patient and mention them in a single sentence.

Providing definitions of complex terminologies: Since our summaries are aimed for patients, we decided to provide definitions for difficult medical terms. However, all of the existing tools for assessing reading level (Flesch, SMOG, Fry Graph) and health literacy (REALM, TOFHLA, NAALS) work only on sentences and cannot be used to assess the complexity of terms. In order to develop our metric for complexity, we randomly selected 300 terms from Dale-Chall list, which is known to be understood by more than 80% of 4th grade students, and labeled them as *Simple*. We also randomly chose 300 words from our database and asked two non-native undergraduate students who have never had any medical condition to annotate them as *Simple* or *Complex* (Cohen's Kappa k=0.786). Several features like number of vowels, consonants, prefix, suffix; counts of the number of nouns, verbs, adjectives; whether the term is present in Wordnet; semantic types of the terms were extracted for each of the 600 terms. Linear regression was performed on the terms with *Complexity* as the dependent variable. Feature values for each of the terms were then supplied to the linear regression function and the resulting score was obtained. We found out that 88% of the terms labeled as *Simple* had scores below 0.4 while 96% of the terms labeled as *Complex* had scores above 0.7. Hence, we use these thresholds for determining the complexity of terms. For a term identified as *Complex* by our metric, we look for its definition in Wordnet, Wikipedia (the first sentence), and UMLS. Medical terminologies present in each of the definitions are extracted and the metric is used to find the scores for the terminologies, which are eventually added up to get a single score for the definition. Definition with lowest score is selected and presented to the user.

Current and Future Work: We are currently working on including patient's perspective in the summaries. For this purpose, we have interviewed four patients so far and are transcribing the recordings. We expect these recordings to guide us about the kind of terminologies and sentence structure that are used by patients to describe their disease and physical conditions. Once this is done, we plan to evaluate the effectiveness our system.