Probabilistically Populated Medical Record Templates: Reducing Clinical Documentation Time Using Patient Cooperation

Tristan Naumann, MS,[1] Marzyeh Ghassemi, MS,[1] Andreea Bodnari, MS,[1] Rohit Joshi, PhD[1]
[1] CSAIL, MIT, Cambridge, MA, USA

Abstract

The adoption of electronic medical records was a milestone in healthcare, but they have proven tedious to maintain. Increasingly, medical records contain information that is repeated within individual and even across multiple records. Such redundancy is frustrating and contributes to poor documentation as studies have shown that clinicians often copy-and-paste old, potentially out-of-date information into patients' records. Moreover, it requires clinicians to spend additional time with documentation that could otherwise be spent with patients who often experience unreasonably long wait times. Therefore, we propose a solution to reduce the time clinicians must spend creating and editing medical records by generating patient-specific medical record templates from similarities found among patients in large medical data warehouses. Specifically, this work predicts the most likely fill-in-the-blank sentences to add to patients' records based on information the patient provided through customized surveys during hospital wait time. In this way, our solution decreases the entry of unneeded information by clinicians, replaces patient idle time with relevant activity, and at the same time predicts information needed for improvements in patient's health quality indices, for better reimbursements, and for enhanced medical legal-justifications.

Introduction

Electronic medical records (EMRs) are useful tools in day-to-day health care, but record editing is a burdensome and time-consuming process. Patient records are generally obtained through manual data entry, whereby patients fill out paper forms that are in turn partially transcribed by a healthcare professional into their medical record. During this process, patients typically interact with various care staff and provide answers to the same questions multiple times. Thus, the documentation process is not only repetitive and tedious, but even wastes patients' time.

Many clinicians see similarly afflicted patients over the course of their day, and thus often create similar content across several patients' medical records. This is exemplified in specialities and during outbreaks of common illness (e.g. flu, colds, etc.). At the same time, outpatients wait 15 minutes, on average, before seeing a clinician.[8] Our goals are to (1) reduce the time doctors spend with documentation in order to permit additional patient interaction, and (2) provide patients with relevant, informative activity during their stay while implicitly reducing their wait times.

The main contributions of our work are twofold: (1) we develop a probabilistic algorithm for generating patient-specific templates; and (2) we propose to leverage the patient wait time in order to reduce the burden of documentation on clinicians. Specifically, we propose surveying patients for their medical information and using the results for generating relevant medical record content. The surveyed information is used to select medical records from similarly afflicted patients. The common content from the selected records is abstracted in order to generate a pre-populated medical record template. This is accomplished in four steps by (1) learning a patient-relevant medical context, (2) learning a medically similar patient cohort, (3) learning a medical record template and (4) customizing templates by specialities or health professionals' roles.

Our solution aims to transform the nature of the clinical documentation by introducing dynamic personalized templates and by increasing the patient involvement in this process. The goal is to remove as much unnecessary input of information as possible using the old records of similar patients. For example, we aim to successfully predict the most likely additions or deletions or repetitions to patients history, family history, physical examination, allergies, review of systems, management options, medications order sets, and counseling information based on the existing data. If successful, we believe our solution can reduce the unneeded information input by about 30-40%.

Background

In order to make EMRs faster to create and consume, clinicians often repeat content within each record and across multiple records. A recent study demonstrated that the majority of medical records consist of more than 20% duplicated content.[1] As early as 2006, there were calls for an EMR system that could " automatically and intelligently delete 90% of text processed by copy-and-paste."[2]

Prior work suggests that information duplication introduces errors in medical documentation. For example, in a 2007 case study[3] a 77 year-old woman who was hospitalized with an initial note to receive heparin for venous thromboembolism prevention had her note copied and pasted for four days, and was ultimately discharged without receiving heparin. Two days later she was rehospitalized with a pulmonary embolism. Siegler et al. noted that copy-and-paste had caused "a number of unexpected problems and concerns about electronic note writing … including reducing the credibility of the recorded findings, clouding clinical thinking, limiting proper coding, and robbing the chart of its narrative flow and function."[4]

At the same time, patients are frustrated by long wait times to see care staff, and patient satisfaction decreases as a result.[5,6] This is especially true in non-elderly (<65 years) patients, for whom shorter wait times are more significantly associated with treatment satisfaction.[7] In contrast to these expectations for more expedient care, the mean wait time in U.S. emergency departments (EDs) increased 25%, from 46.5 minutes to 58.1 minutes from 2003 through 2009.[8] This is often worse in EDs located in urban areas (62.4 minutes), or with 50,000 or more annual visits (69.8 minutes). Engaging patients in their medical treatment should also be a priority. Prior work has suggested that having highly-involved patients promotes better alignment of patients' and physicians' goals and agendas by influencing the extent and type of information physicians provide.[9]

One common solution for reducing clinical documentation time is to use hard coded medical record templates. This solution has focused on using established expert knowledge of patient care to create condition-specific hard-coded templates.[10] Another solution used to diminish clinical documentation time is to hire a transcriber to capture medical dictations or to use sophisticated speech to text applications, but recent work has suggested that physicians who dictate their notes appeared to have worse quality of care than physicians who used structured EMR documentation.[11]

Our method focuses on creating patient templates, but differs from past template-based work by generating a flexible and patient-customized template. Instead of cloning sentences or paragraphs blindly, we use natural language processing and statistical techniques for predicting, constructing, and updating these standard templates with most likely sentences based on the information from similar patients.

Methods

We propose generating a pre-populated medical record template by first learning a patient-relevant medical context. The patient medical context is then used to retrieve a medically similar patient cohort. We use the medical records of the patient cohort to generate a medical record template using statistical and language processing techniques. Our solution is depicted in Figure 1 and we further give a step-by-step methodology description. In addition, a working coded example of our solution together with an interactive presentation of the methodology is available at http://web.mit.edu/andreeab/www/AMIAStudentChallenge2013.

(1) *Context Learning***:** Learning a patient's context refers to using a patient's wait time to gather relevant medical history and current symptoms. Each patient is presented with an electronic survey which captures information regarding current symptoms, together with the medical, family, and social history. The survey questions are dynamically generated based on previous answers such that only relevant information is collected. Dynamic

question generation is done via a rule-based approach,[12] and is customized per each clinic and speciality. The patient answers are automatically recorded into a centralized EMR system and can be accessed in real-time by doctors. We further refer to the patient survey answers as the survey answers.

(2) *Cohort Learning*: We define a medically similar patient cohort as a set of patients with similar health conditions, as captured in their medical records.[13] In order to retrieve the patient cohort of interest, we start with all the patient records from an institutional EMR system. We treat each patient record as a word blob and count the occurrence rate ***R*** of the survey answers inside each blob. We create a similarity ranking between the current patient and the remaining patient population by scoring the patient records inside the EMR system based on the highest occurrence rate ***R***. We define the top 20% of the ranked patient records as the medically similar patient cohort.

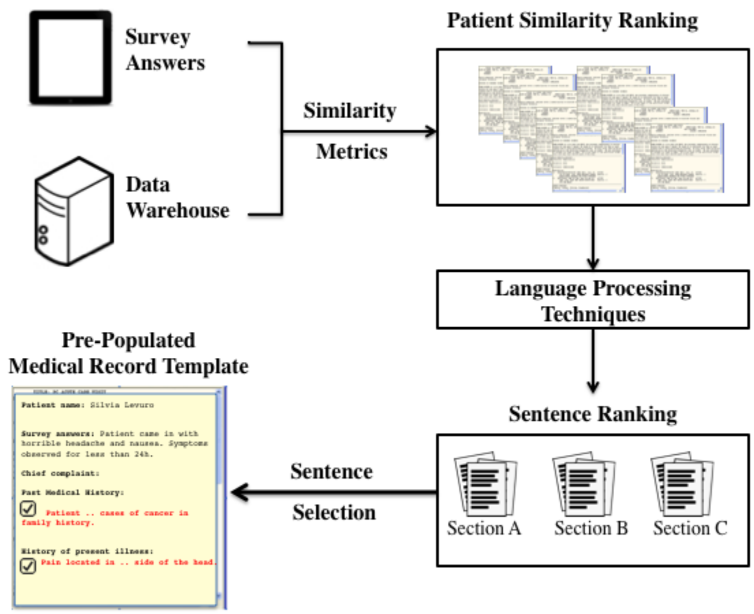

Figure 1. System overview

(3) *Medical Record Template Learning*: We assume that each medical record is organized into medically relevant sections (e.g., current symptoms, medical history, follow-up). We use a rule-based language processing system in order to identify the sentences and sections of each medical record inside the patient cohort.[14,15,16] The rule-based system is developed on and optimized for the medical domain. We cluster together the common sections, and discard the sections occurring only one time. For each section in a section cluster we generate a ranked list of the sentences with the most commonly occurring n-grams (i.e., sequences of n consecutive words). The ranked sentences represent the most commonly used sentences inside that specific section, or the sentences that are most likely to be copy-pasted into a new patient record. We refer to the ranking score as the copy-paste score of each sentence. By ranking sentences through their n-gram frequency we can capture the common content while still allowing for variation (e.g., “Patient was given 10 ml of insulin”, “Patient was given 15 ml of insulin”). In our list of ranked sentences we only record the common content of each sentence and mark the content variation as a fill-in area (e.g., “Patient was given … of insulin”).

Loading..., where K is the length of the set of n-grams in a sentence

In order to generate a pre-populated medical record, we use the common sections identified by our language processing system. Under each section, we include the top k% (k=3) of the ranked sentences from the specific section cluster. We propose to use a specific user-interface where each of the pre-populated sentences can be deleted by a one-click, thus allowing the doctor to easily alter the pre-populated content. We also create a special section that records the survey questions and patient answers.

(4) *Customizing Templates by Specialities*: Finally, each template is customized for distributed care teams by assigning an accessibility score to each learned sentence based on the specialities of healthcare professionals who authored the corresponding documents in the patient cohort from Step (2).

Discussion

Our solution leverages two untapped resources to generate relevant, contextual templates: large volumes of existing patient data, and patient waiting time. In using patient input, we not only encourage patients to become more involved in their care process but also save the care staff time spent querying, and often re-querying, each patient about their condition and medical history. This information also provides context to learn medically similar cohort. In reusing existing patient data, our approach saves clinicians additional time filling out documentation by predicting and pre-populating the repeated patterns of information in patient history, physical examination, clinical assessment, management plan, managing health quality indicators, medical legal justification and reimbursement.

A primary challenge in creating patient-specific templates by reusing information from existing records is how accurate the template should be. We expect to reduce the irrelevant or noisy results retrieved by language processing techniques by utilizing only sentences with frequent occurrence; however, properly abstracting these sentences may very sensitive to the available data. We expect that our system will occasionally retrieve partially irrelevant content for a given patient. To reduce the errors that may be introduced by inaccurate templates, we implement a "one-click delete" functionality for each template sentence in order to quickly remove the irrelevant content,. We rely on the healthcare professional to review the automatically generated template and filter out the irrelevant content. Other considerations include how many templates should be provided, how to generate templates for rarely encountered symptoms, and how to best present these templates to physicians. Similarly, our current solution assumes a single template is generated per patient visit, but it is also reasonable to assume that multiple templates could be generated in an alternative solution which varies the number of top sentences used per template.

Often patients are unsure about their current condition. Therefore, our solution accommodates for their uncertainty, or inability to express the current symptoms, using explained content during the dynamically generated survey and the option to receive clarification from a health professional. Likewise, we encourage unique symptom descriptions through the use of text fields rather than strictly multiple choice selection. In the future, we also plan to integrate the survey generation with the information already available inside the patient's medical record, in order to only capture relevant, contextual information regarding a patient's previous illnesses or history which is not already stored in the EMR system.

Expert-guided templates have demonstrated improvements in clinical authoring time and are more timely than other methods like transcription services. However, these methods are tightly coupled with the specific healthcare organization and with the healthcare staff that define them, and do not take advantage of large amounts of repetitive patient data which is already available. Our solution proposes a more personalized approach to the medical documentation process, where the key element is patient involvement. A variation of our current implementation would be to integrate the probabilistically chosen sentences with content from hard-coded medical record templates. This variation would allow for the integration of expert knowledge with knowledge retrieved from the patient.

We propose to evaluate our solution in a real-life setting. Specifically, our evaluation plan will recruit subjects -- both patients and doctors -- in order to determine overall satisfaction. This will be achieved by evaluating the satisfaction rate of patients taking the survey, how likely they are to fill in a survey on a second visit, and whether the survey was a relevant or informative activity for them. Further, the usage of pre-populated templates will be evaluated through doctor satisfaction studies and analyzing the amount of time spent on medical documentation when the pre-populated template is available versus when the entire medical record is written up by the doctor. Finally, we are also interested to assess whether the medical records based on the pre-populated templates are more complete and present a lower rate of errors compared to regular records.

Finally, our solution requires minimal changes to the medical records system, and therefore can be easily implemented inside hospitals where the surveys can be distributed on mobile devices.

Conclusion

We propose a solution to transform the siloed nature of clinical documentation by integrating the most useful data from the prior notes into the new notes without repeated care staff transcription or literal 'copy-and-paste'. By searching through patient records for similar records, our system supports a "kernel method" of medical reasoning, where other patients' records are weighted according to their relevance to the current patient. The patient wait time is leveraged and transformed into an activity relevant to the healthcare process. Overall, we expect this solution to diminish the time doctors spend documenting and increase the doctor-patient interaction time, to transform the patient wait time into an informative and relevant activity, and to reduce both the doctor and the patient frustrations with the healthcare system.

References

1. Thornton, J. Daryl, et al. "Prevalence of Copied Information by Attendings and Residents in Critical Care Progress Notes." *Critical care medicine* 41.2 (2013): 382-388.
2. Hirschtick, Robert E. "Copy-and-paste." *JAMA: the journal of the American Medical Association* 295.20 (2006): 2335-2336.
3. Hersh, William. "Copy and Paste." AHRQ WebM&M [online journal]. July/August 2007. Available at: http://www.webmm.ahrq.gov/case.aspx?caseID=157
4. Siegler, Eugenia L., and Ronald Adelman. "Copy and paste: a remediable hazard of electronic health records." *The American journal of medicine* 122.6 (2009): 495-496.
5. Leddy, Kelly M., Dennis O. Kaldenberg, and Boris W. Becker. "Timeliness in ambulatory care treatment: an examination of patient satisfaction and wait times in medical practices and outpatient test and treatment facilities." *The Journal of ambulatory care management* 26.2 (2003): 138-149.
6. Anderson, Roger T., Fabian T. Camacho, and Rajesh Balkrishnan. "Willing to wait?: The influence of patient wait time on satisfaction with primary care." *BMC Health Services Research* 7.1 (2007): 31.
7. Kong, Meg C., et al. "Correlates of patient satisfaction with physician visit: Differences between elderly and non-elderly survey respondents." *Health Qual Life Outcomes* 5 (2007): 62
8. Hing E, Bhuiya F. Wait time for treatment in hospital emergency departments: 2009. NCHS data brief, no 102. Hyattsville, MD: National Center for Health Statistics. 2012.
9. Cegala, Donald J., Richard L. Street Jr, and C. Randall Clinch. "The impact of patient participation on physicians' information provision during a primary care medical interview." *Health communication* 21.2 (2007): 177-185.
10. Schnipper, Jeffrey L., et al. ""Smart Forms" in an electronic medical record: Documentation-based clinical decision support to improve disease management." *Journal of the American Medical Informatics Association* 15.4 (2008): 513-523.
11. Linder, Jeffrey A., Jeffrey L. Schnipper, and Blackford Middleton. "Method of electronic health record documentation and quality of primary care." *Journal of the American Medical Informatics Association* 19.6 (2012): 1019-1024.
12. Mittal, Vibhu O., and Johanna D. Moore. "Dynamic generation of follow up question menus: facilitating interactive natural language dialogues."*Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., 1995.
13. Cao, Hui, et al. "Use abstracted patient-specific features to assist an information-theoretic measurement to assess similarity between medical cases." *Journal of biomedical informatics* 41.6 (2008): 882-888.
14. Savova, Guergana K., et al. "Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications." *Journal of the American Medical Informatics Association* 17.5 (2010): 507-513.
15. Tibor Kiss and Jan Strunk. Unsupervised multilingual sentence boundary detection. In Computational Linguistics, volume 32, pages 485–525, 2006
16. [Nltk, 2013]: Nltk toolkit: http://nltk.org/