HARVEST, a Holistic Patient Record Summarizer at the Point of Care

Noémie Elhadad, PhD¹, Sharon Lipsky Gorman, MA¹, Jamie S. Hirsch, MD, MS¹,
Connie Liu, MS¹, David K. Vawdrey, PhD¹, Marc Sturm²
¹Columbia University, New York, NY; ²NewYork-Presbyterian Hospital, New York, NY

Introduction

In many care settings, clinicians are faced with an overwhelming amount of complex information about their patients, with little time for chart review [1]. Failure to digest patient data may result in errors in diagnosis and delayed care [2,3]. The need for better health information management and visualization tools has long been recognized [4,5], yet current electronic health records (EHRs) do not yet provide the cognitive support necessary for effectively and efficiently reviewing patient data. EHR displays are plagued by low information content, do not honor established interface design principles, and cannot be readily customized without imposing a considerable burden on clinicians and information technology professionals [6,7]. Important research has been undertaken on visualization of patient histories [8–10] and domain-specific summarizers [11], motivating our hypothesis that a holistic patient-record summarizer can impact care in a beneficial fashion. In this abstract, we describe the design, implementation, and deployment of HARVEST, a longitudinal patient record summarization system.

Methods

HARVEST is an interactive, problem-oriented patient record summarization system (see Figure below) [12]. The summarizer differs from previous work in three critical ways: (i) it extracts content from the patient notes, where key clinical information resides; (ii) it aggregates and presents information from multiple care settings, including inpatient, ambulatory, and emergency department encounters; and (iii) it is integrated into two commercial EHR systems, and is available for all patients in our institution, not just a curated dataset or for specific patient cohorts.

The natural language processing (NLP) of clinical notes is carried out through a named-entity recognition system that indexes concept mentions. Because HARVEST aggregates problems, NLP was constrained to extracting concepts from the UMLS semantic group “Disorder” only, restricted to the SNOMED-CT Core Problem List. Concept salience weights are computed dynamically to reflect both the frequency of the concept in the patient notes in a given time slice of the record, and the prevalence of the concept across all patients in the institution. To enable parsing and salience computation at scale, we created a distributed computing infrastructure (using Apache Hadoop) and implemented a map-reduce version of our NLP system to parse the notes from a variety of HL7 interface feeds.

Results

The Hadoop infrastructure enabled us to accommodate the large volume of clinical documents generated in our institution (650,000 notes per month): a small four-node cluster processed 20,000 notes/second compared to 500 notes/second in a non-distributed computing environment. The infrastructure also permitted us to experiment with different parsing and salience computation strategies through short development cycles.

HARVEST was deployed within NewYork-Presbyterian’s iNYP clinical information review system as a beta release in September 2013. As a first phase of deployment and to study the impact of HARVEST on clinical care, access was limited to internal medicine residents and residents and attending physicians in the emergency department (ED). Three primary use cases for the summarizer have emerged thus far: (i) in the ED, as a way to capture the essential knowledge about a patient’s history in an efficient fashion, including previous emergency visits and patterns of visits; (ii) in internal medicine, as a support tool for performing an in-depth patient chart review before admitting a patient to the hospital; and (iii) in the primary care clinic, for chief residents as an education tool during precepting hours with residents.

Discussion

HARVEST addresses an unmet need for clinicians at the point of care, facilitating effective and efficient review of essential patient information. The deployment of HARVEST in our institution allows us to study patient record summarization as an informatics intervention in a real-world setting. It also provides an opportunity to learn how clinicians use the summarizer, enabling informed interface and content iteration and optimization to improve patient care. Future work includes customizing the content selected by HARVEST for different types of clinicians.
De-identified HARVEST screenshot for a sample patient, part of iNYP. For the selected time range (3 months), stable angina, pulmonary hypertension, end-stage renal disease, and dyspnea are the most prominently documented problems. HARVEST also identified diabetes mellitus, hypertension, and dyslipidemia as important problems. The problem “dyspnea” was selected. The Notes panel lists all notes in the selected time range that mention this problem. A cardiology consult note is selected and displayed in the lower right panel, with all mentions of dyspnea (and synonyms) highlighted. On the timeline, documentation of dyspnea is highlighted by purple bars, indicating that dyspnea was a particularly salient issue at that time, as well as 6 months later.

References


