Workshop on using natural language processing applications for enhancing clinical decision making: an executive summary

Vinay M Pai,1 Mary Rodgers,1,2 Richard Conroy,1 James Luo,1 Ruixia Zhou,1 Belinda Seto1

1National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, Maryland, USA
2School of Medicine, University of Maryland, Baltimore, Maryland, USA

Correspondence to Dr Vinay Pai, National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, 6707 Democracy Blvd., Ste 200, Bethesda, MD 20892, USA; paiv@mail.nih.gov

ABSTRACT
In April 2012, the National Institutes of Health organized a two-day workshop entitled ‘Natural Language Processing: State of the Art, Future Directions and Applications for Enhancing Clinical Decision-Making’ (NLP-CDS). This report is a summary of the discussions during the second day of the workshop. Collectively, the workshop presenters and participants emphasized the need for unstructured clinical notes to be included in the decision making workflow and the need for individualized longitudinal data tracking. The workshop also discussed the need to: (1) combine evidence-based literature and patient records with machine-learning and prediction models; (2) provide trusted and reproducible clinical advice; (3) prioritize evidence and test results; and (4) engage healthcare professionals, caregivers, and patients. The overall consensus of the NLP-CDS workshop was that there are promising opportunities for NLP and CDS to deliver cognitive support for healthcare professionals, caregivers, and patients.

INTRODUCTION
With the explosion of new biomedical data, knowledge, and guidelines, clinical information has far exceeded human cognitive capacity. Clinical decision support (CDS) systems have great potential to make this information accessible and readily comprehensible to humans. CDS systems are computer-based software systems designed to help health professionals, patients, and care providers make informed clinical decisions, to provide rapid access to evidence-based guidance, and to suggest when additional information is needed or alternative hypotheses need to be considered. Natural language processing (NLP), with its purpose of enabling computers to derive meaning from natural language, has the potential to greatly enhance the function of the CDS systems.

In April 2012, the National Institutes of Health organized a two-day workshop entitled ‘Natural Language Processing: State of the Art, Future Directions and Applications for Enhancing Clinical Decision-Making’.1 This workshop was designed to assess the current state of the art, challenges, and opportunities of NLP and CDS. The discussion of challenges and opportunities involved a wide range of stakeholders, including clinicians, academicians, and representatives from federal agencies, health insurance organizations, and industry. The viewpoints from the first day of the meeting, on the state-of-the-art and future directions for NLP, are being published elsewhere.2

During the second day of the workshop, the participants discussed how NLP and CDS could be harnessed to:
▸ incorporate legacy and unstructured clinical notes,
▸ develop longitudinal models for interpreting patient’s progression in health and wellness,
▸ enhance medical reasoning
▸ provide trusted and reproducible clinical advice,
▸ prioritize evidence and test results, and
▸ engage healthcare professionals, patients, and caregivers to promote effective communication and coordination of care.

STRUCTURED VERSUS UNSTRUCTURED DATA
A major theme in the workshop was the need to address barriers to incorporating machine-readable unstructured notes into the CDS process. For the most part, traditional medical records have been unstructured notes that include medical history, detailed profiles of patients, medical examinations, pertinent interactions, and the clinician’s thought process. Roth et al3 indicated that qualitative measures (ie, disease-specific history, family history, patient education, and social history) affecting improvement in quality of care are difficult to capture in a structured note framework.

The limitations of CDS systems as rule-based solutions that act on constrained ontologies may be addressed with the integration of NLP in CDS. NLP is instrumental in using free-text information to drive automated decision support, representing clinical knowledge interventions in standardized formats, and leveraging unstructured narrative.4

The workshop attendees agreed that it would be preferable to have the data structured after it is captured in the electronic health records (EHRs) rather than having it structured as part of the capturing process. One way to address this conundrum is by using NLP to extract context and meaning from the narrative text content of EHR with clinical information extraction (CIE) tools. Speakers noted the following issues with structured note entry: (1) there is a lack of uniformity in clinical definitions and procedures; (2) not all variables are captured to provide the best course of action under complex guidelines; (3) integration of a constrained system into the workflow of busy clinicians seeing complex patient cases is non-trivial; and (4) it is hard to train people to use encoded text systems consistently.

Participants noted that the development and refinement of these tools in the clinical
environment has been restricted due to the limited availability of training datasets because of patient confidentiality and privacy concerns and variability in text quality. In order for CIE tools to progress towards applications in the clinic, there is a need to develop trust to promote data sharing and to use human experts to verify extracted information.

PERSONALIZED LONGITUDINAL HEALTHCARE

Since current EHRs are typically document-based and episode-based, they do not readily capture elements of the patient’s data that transcend care transitions or tease apart longitudinal chronic-care management of multiple co-morbidities. Elderly patients, in particular, tend to see a large number of different providers, who may not all be at the same institution, making it difficult to access complete medical records. Decision rules are primarily cross-sectional and not longitudinal because clinical evidence guidelines are most commonly developed based on episodic datasets or simple changes without taking into account complex personal histories.

Radiology was presented as an example where decision support based on longitudinal analysis could be enhanced. While radiology reports are text-based, they usually have a constrained vocabulary and a limited number of concepts for each imaging modality. An NLP-CDS system would need to: (a) determine whether prior tests were positive or negative, (b) find previous recommendations, (c) identify any unexpected or unresolved findings in the past, and (d) consider the value of a new test, particularly one involving ionizing radiation. Ideally, an NLP-CDS system should be capable of detecting any unexpected findings in a summary report based on contextual analysis. Additionally, such a system should be able to automatically generate protocol guidance based on current findings and medical history and flag details for further attention.

ENHANCING MEDICAL REASONING

Rule-based and statistical systems are two different techniques that have been developed to prioritize evidence and enhance medical reasoning using NLP. Participants noted that both approaches can trade-off sensitivity and specificity, an important characteristic to be able to accommodate a variety of applications. For example, specificity is a priority when triggering automated, patient-specific notices because of potential alert fatigue, while sensitivity is more important for identifying alternative diagnoses.

Participants discussed the rule-based languages, including the Arden syntax\(^1\) promoted by HL7,\(^6\) and its application in ambulatory care. For example, in such care settings, drug–drug interactions can be handled for patients allergic to certain drugs, alternative medications can be recommended, abnormal lab results can be flagged, and providers can be alerted to opportunities for immunizations and preventive services. Rule-based systems have been developed using clinical practice guidelines formulated from clinical trials. However, since the clinical trials typically do not enroll patients with comorbidities, these rule-based systems fail to appropriately prioritize evidence and test results for patients with polypharmacy and/or comorbidity.

Probabilistic analyses based on statistical models and machine learning approaches were discussed, including the recently demonstrated IBM DeepQA ‘Watson’ system.\(^7\)\(^8\) According to Duda and Shortliffe, a knowledge-based system can be described as an artificial intelligence (AI) program whose performance depends more on the explicit presence of a large body of knowledge than on the presence of ingenious computational procedures.\(^9\) Participants recognized that considerable research and implementation work has been done in the development of computational approaches and knowledge-based systems utilizing NLP for supporting clinical decision-making.\(^10–27\) One of the current approaches that was discussed was the Lexicon-Mediated Entropy Reduction (LEXIMER) system,\(^28\) which extracts recommendations from a database of millions of clinical reports based on whether the reports are positive or negative. Thus, beside expert advice or literature, data mining could be used to determine percentage of positive studies or to evaluate previous recommendations.

The DeepQA system used both structured and unstructured text to create a large body of knowledge on which statistical methods could be used. While this approach is not expected to comprehend very complex medical guidelines, there are plans to provide it with shallow semantics and reasoning tools to interpret a large number of evidence-based medical guidelines. Participants recognized that for the DeepQA system to be successful, it would need to be enhanced with multimodal analytics, provided by a framework like the Unstructured Information Management Architecture (UIMA).\(^29\)

Three concerns were noted: (1) eventually machine learning saturates, (2) errors can propagate downstream, and (3) a good generalized knowledge base is difficult to generate for use by different AI systems. Participants observed that small doses of knowledge could inform and optimize statistical processes in ways that would be challenging for any amount of computation.

EVALUATING CLINICAL DECISION SUPPORT ENGINES

Development of measures for evaluating CDS engines within the same clinical context is hindered by the lack of access to a standardized corpus of data. The challenge for researchers is the need to tackle legal, privacy, and institutional review board concerns for enabling access to the colossal amount of data currently available. The participants indicated that a federated database of anonymized medical data would be useful to enable the evaluation of CDS engines.

Regarding the metrics for evaluating these engines, participants considered two approaches: (1) whether CDS systems generate advice that follows evidence-based guidelines reproducibly; or (2) whether the outcome is as expected. The latter can be difficult to analyze because it is idiosyncratic based on a particular patient and a particular situation. Even for a trial comparing CDS on a cohort of patients, the time required for determining whether different outcomes are reached can be so long that it may be difficult to draw adequate conclusions about the relative efficacy of a CDS system. While the ultimate correct outcome for decision support is the outcome and not the process change, the complexity in analyzing outcomes means that the optimal approach for evaluating CDS engines may be to determine how well the advice they provide follows evidence-based guidelines.

PRIORITIZING EVIDENCE AND TEST RESULTS

There is a need for a framework, which could be ontological,\(^9\) that permits guidelines with specifications for recommendations and actions as well as algorithms for suggesting the temporal order of interventions. This would enable the CDS system to go beyond identifying very simple health problems and manage complex clinical scenarios that unfold in a complicated temporal sequence.

However, this type of framework requires a lot of information which is not accessible through the coded data but is primarily in the narrative text. In addition to polypharmacy and comorbidities, the information needed includes currently not documented information such as patient preferences, provider preferences, and social support. The validity of secondary use of
large clinical datasets in retrospective studies should be considered carefully in view of the potential for ‘missing’ coded data and the multi-faceted nature of human disease. Further work needs to be done to incorporate the non-coded data that exists in EHRs, not in the context of clinical trials but of a compendium of medical treatment of similar patients, in order to enable decision-making for complex patients with polypharmacy and/or comorbidities.

STAKEHOLDER ENGAGEMENT
There are multiple stakeholders in CDS systems. These systems assist in making decisions that generally distinguish between three zones: do-not-treat, collect more knowledge through testing, and treating the patient. The creation of a knowledge base, or information repository, is essential for any CDS to be successful in this clinical process. However, the speakers noted that populating the knowledge base with the appropriate set of structured data is essential to a strong statistically-based NLP system.

Most EHRs do not include medical knowledge-base in one package. Medical knowledge-base is usually added as a component either by the users or implementers of EHR systems or by the medical knowledge-base vendors. There is a need for unifying the teams that are creating the medical knowledge and the clinical teams that are using that knowledge to support patient care. Another challenge for any good knowledge base is that it should be flexible enough to be compatible with different EHR systems and a variety of CDS engines.

One of the concerns raised during the workshop was the need to expand the medical knowledge base to include information for and about patients with low locus of control, with limited education or limited English proficiency (LEP), or with low interest in maintaining personal health. In other words, how could the medical knowledge base generalize to incorporate information about patients who may or may not be within the healthcare system? There is also a need to incorporate patient preferences and utilities within the CDS system. CDS systems need to be informed about patients’ interest in their genetic testing, and reasoning about their diseases and treatments.

SUMMARY
Participants agreed that there were many challenges to be addressed, including importing hand-written notes, capturing oral dictation, and seamless implementations in clinical and non-clinical environments. However, participants believed that NLP and CDS were promising technologies for enabling delivery of cognitive support to healthcare professionals, patients, and caregivers by providing easily understandable synthesis and summary of the ever-expanding medical evidence and knowledge-base. Participants expressed optimism that NLP-enhanced CDS systems could become a ubiquitous tool in providing improved personalized healthcare.

Acknowledgements We would like to thank the organizing committee of the NLP-CDS workshop for setting up an excellent discussion forum. Besides the authors, the organizing committee included: Dr Blackford Middleton, Dr Olga Bzazhnik, Dr Elaine Collier, Dr Milton Com, Dr Dina Fushman, Dr Mike Huerta, Dr Thomas Rindflesch, Dr Steven Hirschfeld, Dr George Reddmond, Dr Abdul Shaik, Dr Jennie Larkin, Dr Rongling Li, Dr Peter Lyter, and Dr James DeLeo.

Contributors VMP extracted the notes from the transcribed pages of the workshop and drafted the early version of the manuscript with MR. All authors contributed to editing the drafts of the manuscript and approved the final paper.

Funding National Institutes of Health.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The PowerPoint presentations as well as the transcribed notes for the workshop can be found at: http://www.nibib.nih.gov/NewsEvents/MeetingsEvents/MeetingSummaries/LP2012.

REFERENCES
10 Begg CB, Greensa RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 1983;207–15.

