Technology Strategies Notes

Volume 1, Issue 10
October 2014

Office of Technology Strategies (TS), Architecture, Strategy & Design (ASD)

A VA Executive's Guide to Natural Language Processing

INTRODUCTION

The most recent TS Note on Health Informatics briefly mentioned clinical decision support systems (CDSS). Natural language processing is one of the key concepts behind the benefit of using CDSS's. This TS Note will explore natural language processing and artificial intelligence in healthcare. Health technology is a major site of human-computer interaction, be it patients interacting with personal devices to monitor their health, hospitals managing administrative processes (e.g., scheduling), or doctors using the latest technology to improve health outcomes. As VA adopts (and adapts to) leading-edge technology trends in healthcare, natural language processing will be one of the key innovations that will impact Veteran health outcomes, as well as the underlying processes in VA's health and benefits delivery.

BACKGROUND

Artificial intelligence (AI) has been a popular science-fiction concept for years, but in reality it has been built into the computers and applications we interact with every day (e.g., email spam filters). One big AI innovation that's been in tech news lately is IBM's Watson supercomputer. Although Watson was initially created to answer trivia questions, IBM has recently partnered with Memorial Sloan Kettering Cancer Center to apply Watson to cancer diagnoses and treatment options. Watson is also being used by WellPoint to assist in utilization management, and USAA will be using a Watson-backed app to help Service members transition into civilian lives. These applications all rely specifically on Watson's natural language processing, which is a key function of artificially intelligent systems.

Natural language processing describes a computer's ability to understand or generate human or natural languages. One simple case of this function is in Google's email service, which can "read" incoming messages, understand whether an email is spam, and act accordingly (i.e., file the message as "spam"). The "understanding" step is where most of the work is done, and it requires more than fast processors. There are generally two ways in which an artificially intelligent application learns or comprehends language (known as 'machine learning'):

- Rule-based model: the computer relies on an expert user to input rules about the data it is computing.

Continued on Page 2

Technology Strategies

Defining OI&T's "To Be" Technology Vision

This newly established office within OI&T's Architecture, Strategy & Design (ASD) interacts not only with the ASD pillar offices, but also with multiple stakeholders within OI&T and with strategic offices across the enterprise. TS works closely with IT and business owners to capture business rules and provide technical guidance as it relates to Data Sharing across the enterprise, specifically for inter-agency operability.

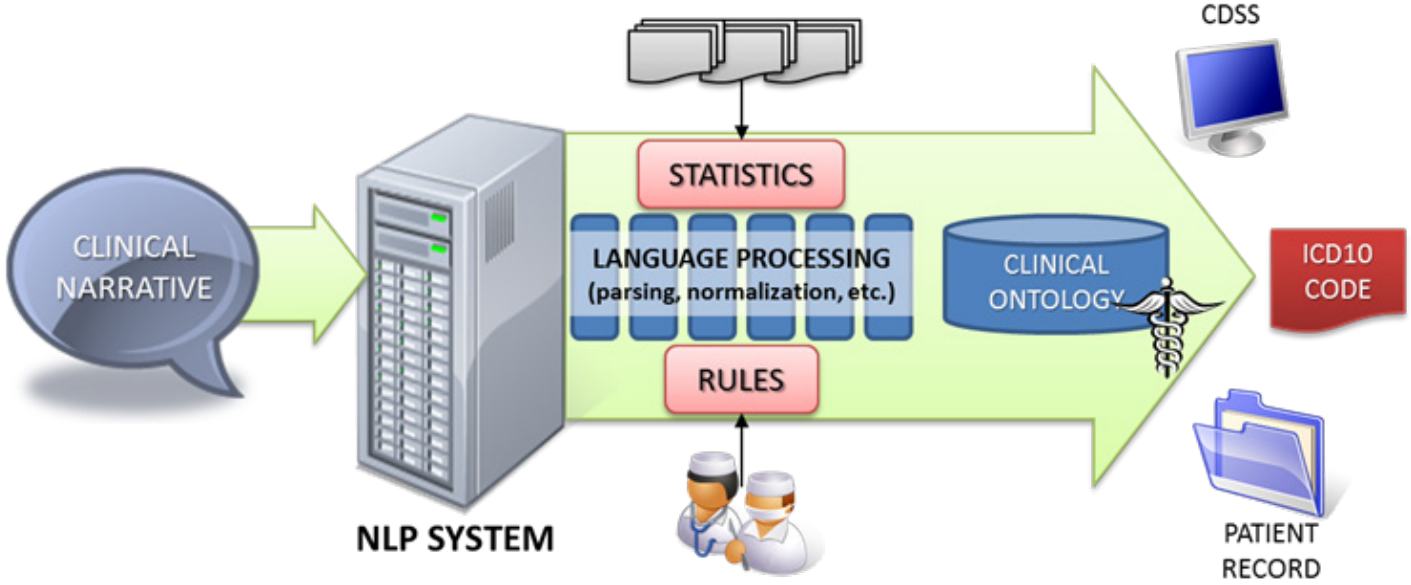

Figure: Using Natural Language Processing to Translate Clinical Narrative

Technology Strategies Notes

Volume 1, Issue 10
October 2014

A VA Executive's Guide to Natural Language Processing

Continued from Page 1

- Statistical model: the computer examines the statistical relationships within large sets of data in order to create "rules" and outputs itself.

In both models, there are key components or steps in the machine learning process:

- Normalization: how the computer knows how to define terms
- Parsing: sentence structure and context for a given term
- Word-sense disambiguation: choosing from multiple meanings of a single word in context
- Named entity recognition: identifying all proper nouns, such as medications; and
- Semantics: using the syntax-focused steps above to create a meaningful output.

In essence, natural language processing gives computers on-demand grammar lessons in order to comprehend anything from email spam to trivia questions to medical diagnoses or codes.

NATURAL LANGUAGE PROCESSING IN HEALTHCARE

Natural language processing is more than just question-answering or filtering—it can also perform knowledge extraction. In a healthcare environment (see Figure), this means analyzing medical data and extracting information from related material like lab results or clinical notes to deliver meaningful outputs.

As the healthcare industry adapts to such challenges as the explosion of health data and the complexities of adopting the 10th revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), it must find ways to make all this information support positive health outcomes. Much of the newly generated data isn't structured or easy to use—some estimates suggest that up to 80% of the medical record is unstructured narrative. ICD-10 contains more than five times the number of diagnostic codes of ICD-9.

Natural language processing provides a number of solutions to these healthcare challenges by "reading" such unwieldy data sets and structuring or standardizing them for use by clinicians or other systems. Some of the current innovative applications of natural language processing in healthcare are listed below:

- *Auto-coding* – one of the greatest challenges in healthcare is the combined increase in unstructured medical data and diagnostic codes. A computer must analyze the linguistic information (clinician's notes, other lexical medical data) and deliver an output that is useful to a human user of that information. Natural language processing can perform this linguistic analysis on patient data, while also using statistical machine learning to return multiple diagnostic codes based on relative certainty.
- *Clinical Decision Support* – traditionally, CDSS's have had limited amounts of computable data to work with, which reduces their effectiveness and has slowed adoption across the industry. Natural language processing allows CDSS's to mine massive amounts of unstructured data from EHRs, patient apps, or medical studies. (For more on clinical decision support and health informatics, see TS Note 9)
- *Concept processing* – patient data also poses challenges to clinician workflows, as the same information tends to be entered multiple times. Concept processing uses natural language processing to help eliminate this redundancy by auto-populating a clinician's current case based on similar elements from past cases.

NATURAL LANGUAGE PROCESSING AT VA

Natural language processing features in the OneVA Enterprise Technology Strategic Plan's (ETSP) Information Management Technology section, specifically in the near- and long-term vision for Data Management and Business Intelligence & Data Warehouse Platforms. At VA, natural language processing offers more than just health technology solutions. As the agency moves towards more unstructured data, it should invest in "automated capabilities to analyze a problem, generate hypotheses, test possible solutions, and assign a confidence rating in its answer." While this is commonly applied to differential diagnoses in medical care, there are other areas across the enterprise where natural language processing can be applied to complex data to assist in decision-making. Ultimately, in its IT vision, VA has identified natural language processing as a strategic solution for enterprise-wide adoption.

If you have any questions about natural language processing, don't hesitate to ask CTS (askCTS@va.gov) for assistance or more information.

Check out earlier TS Note editions here

(http://www.techstrategies.oit.va.gov/docs_ctsnotes.asp)