Natural Language Processing Provides Foundation for AI in Medical Diagnoses

Developing technology could lead to AI-based treatment plans for many common orthopaedic problems

● ALAN M. REZNIK, MD, MBA, FAAOS, AND KEN URISH, MD, PHD

The first article in this series ("Understanding the Impact of Artificial Intelligence on Orthopaedic Surgery," *AAOS Now*, September 2018) discussed the basic concepts and history of artificial intelligence (AI). The second installment ("How Would a Computer Diagnosis Arthritis on a Radiograph?" *AAOS Now*, December 2018) described how AI can distinguish groups of items by features or local texture. This article explores the AI approach to spoken words and how it may be used in medical diagnoses.

For many, AI remains a mysterious black box with disruptive potential in every direction. One of the more challenging areas of AI is understanding spoken language and answering human queries. Because most interactions that relate to patients are verbal or written and published data/research is printed in text, understanding the meaning of language has become a major focus of AI endeavors that interact with human behavior.

This article provides an overview of several approaches to the use of language in understanding medical conditions, charting, conducting research, and making diagnoses. It discusses natural language processing (NLP) tools, including stemming, rare token analysis, edit distance, and mathematical word representation vectors, as they apply to orthopaedics.

What is NLP?

NLP is used in understanding what words and sentences mean. It is of little value unless the effort is associated with useful actions. Its capabilities range from simple

Editor's note: This article is the third installment of an ongoing series about artificial intelligence. The first two installments are available online at www.aaosnow.org.

tasks such as "send a text message to" (understand me and then do something) to IBM's Watson playing Jeopardy. The ability of a computer to play Jeopardy against humans was a big leap for AI. It showed an understanding of subtleties of language, such as the clue rhyming with the answer or the category title implying irony or humor in the question. In addition, the computer had to make judgments whether it should buzz in or risk a penalty for a wrong answer.

Today, NLP challenges include translating local language that may contain idioms or sarcasm, reading with a comparative understanding of scientific literature, and taking a medical history to make a diagnosis.

To understand the general concepts, this article examines a few basic tools now used to break down language, analyze it, and use it to find value in the data. The goal is an understanding of how it works, possible applications, and the pitfalls of this type of analysis as it applies to medical records.

Words, stemming, and rare word analysis

NLP can start with the simplest of processed words. The first step may be a word dictionary—a catalog of the roughly 177,000 words available in English, their correct spellings, and maybe even the most common misspellings. It may include words that have been "stemmed" and grouped together. Stemming is reducing a word to its stem word or "token" (for example, cars and car, walking and walk, obstruction and obstruct).

Stemming

Computer stemming (mapping words to roots) was accomplished in 1968 by Lovins, who was the first to publish a working "stemmer" algorithm. Although mapping words to their simpler stems

may miss subtle meanings, it can be a great first pass. Word stems or tokens can relate topics, sentences, or articles in a series of papers. In rare token analysis, NLP will drop common words (e.g., and, the, it, or, a) and count the frequency of the remaining "rare" words or tokens. For example, in a search for data on metastatic osteosarcoma, the rare token frequencies of words such as metastatic, tumor, osteosarcoma, limb, and salvage may match in papers of interest. Tokens can be taken one at a time, in clusters of words, or in any order to aid in the analysis.

To allow a computer to read the literature and make inferences, deep learning can be added to the AI process. As examples accumulate, the neural network would learn the value of "limb salvage" as a word pair in osteogenic sarcoma treatment as it relates to other tokens or word sequences. It may "decide" to add "five-year survival rate" to the list of phrases of higher value. It might prioritize physical findings over vague symptoms. A key feature of AI and machine learning is training a neural network. Any network can be trained to be smarter, just like Google learns when you click on a search result.

An AI engine might look at a constellation of medical histories and find that a patient with a herniated disk and a foot drop might have the following in common: rare words such as numbness, tingling, weakness, tripping, and pain in the history; altered reflexes, paresthesia, weak dorsiflexion, or weak extensor hallucis longus in the physical examination; and loss of disk space on the X-ray report. NLP would process the words and find a constellation of tokens based on stemming and rare word frequencies to create a diagnostic framework. It could look at combinations close to the answers and evaluate a likely answer based on how close it is to

the idea presentation. This may be taken as an edit distance—a measure of how far away one situation is from another. In its own way, it gauges how much work it takes to match things up.

Edit distance

Edit distance is used frequently to complete spell check or provide alternative spelling options. If a word is misspelled, the processor looks for near matches. For example, when it sees "thar," it knows that the correct word may be "there," "that," "the," or "tear." It offers possible edits for the mistaken word, perhaps offering to change "r" to "t" because those letters are next to each other on the keyboard. It may rank "that" as the highest or most likely edit because even though "there" sounds similar, it would require two changes (edits). Similarly, "tear" may be ranked high on the list of possibilities because it requires only one change. The effort needed to correct the word is measured and becomes the edit distance. The AI engine would then rank the edits in order ("that," "tear," "there," "the"), demonstrating an understanding of the work needed to make the correction.

It does not take much imagination to see how the edit distance strategy could be beneficial for medical information. Diagnostic medical information (symptoms, findings, history, and labs) that is one edit away from a known diagnosis or treatment plan would be preferred over another choice that is three edits away. The idea of edit distance or closeness can be extended to help find the meaning of words when vectors are used to place words in a mathematical space where distances have meaning. The process of creating the vectors is called Word2Vec (W2V), available inside most AI software packages.

W2V

Although mathematical algorithms can calculate edit distance to perform spell check, that is considered shallow AI and does not involve any true word understanding. Understanding meaning is much more difficult. That is where mathematical word vectors, or W2V, come into play.

Imagine creating vectors (multidimensional matrices of numbers) that help define a word. Like stemming on steroids, it includes values that relate the stems to all the words they come from, their
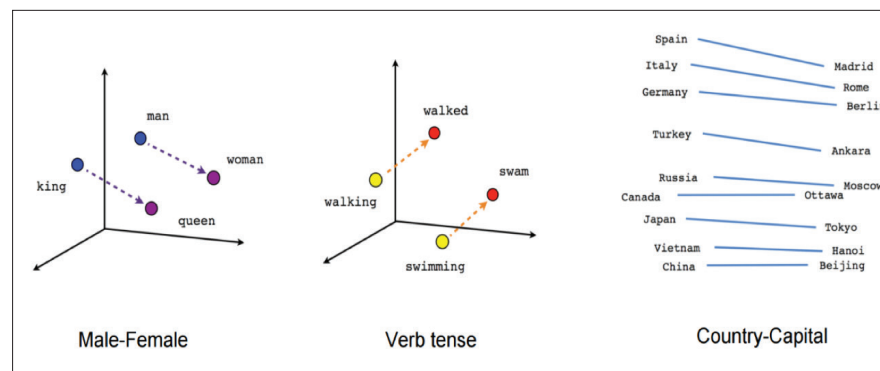

Fig. 1 Word vectors are also useful as features for many canonical natural language processing prediction tasks, such as part-of-speech tagging or named entity recognition.
COURTESY OF TENSORFLOW™ WEB TUTORIAL

synonyms, and other related words. Take the vector for "prince" subtract "male" and add "female," and the vector for "princess" should be the result. Other W2V concepts are graphically represented in Fig. 1 with well-described model concepts.

Standard libraries for W2V are part of tools that are available to AI programmers, and they are being refined every day. With W2V, a deep learning network could process a string of sentences and "know" whether they follow each other, say the same thing in different words, or are opposites. One also can use the vector to define an angle or direction of a word and seek vectors that have similar angles or directions to further understand the relationships between words.

Mathematical models such as edit distance and W2V can help computers map words for comprehension, but for medical diagnosis, a training set of real patients with known outcomes and test cases with known results are needed to validate the AI before it can be used on true unknowns. For many AI projects, how programmers train the AI engine and the mathematical models can create unforeseen biases, while equally discovering information or relationships never before considered.

Pitfalls of AI, machine learning, and medical records

Many electronic medical records (EMRs) are computer-filled by menu instead of by free text. There are also many that are geared to electronic population of forms. In those EMRs, many differences are lumped together in one International Classification of Diseases, Tenth Revision, Clinical Modifications code, and finer points are lost forever in the record. In a free text document, abbreviations can overlap in meaning. For example, "PE" can mean pulmonary embolism, physical exam, pectus excavatum, pleural effusion, proton emission,

professional engineer, and more.

Differences in medical education and training can make the situation worse. For example, what is a 1+ Lachman test when most normal knees have 5 mm to 9 mm anterior draw on KT-1000 testing? Or what is the value of studies that show there is a large interobserver difference in the Neer classification for multipart humeral head fractures, as well as many other fracture rating systems?

Moreover, many findings are subjective and vary from one patient to another (e.g., referred pain, pain levels, nausea, location or point tenderness). Verbal and visual pain scales often do not correlate as numerical representations and make little sense to patients who lack the simplest of quantitative skills. In the medical world, we may have "garbage in, garbage out" more often that we realize.

The future

Pitfalls aside, understanding the

basic tools outlined here and knowing there are many more being developed, it is not difficult to imagine AI taking medical histories, formulating AI diagnoses, and creating AI-based treatment plans for many common orthopaedic problems.

In a world were Alexa can order a pizza, Siri can give you directions or call a friend, and Google can find the top papers on a topic in a fraction of a second, a good history and differential diagnosis may be a short edit distance from AI's reach.   Ⓝ

Next month in *AAOS Now*

- Sex-based Differences in Shoulder Instability
- Frostbite Image Modalities and Treatment
- Medicare's "Relaxed" Documentation Rules
- Workplace Leadership Ensures Quality and Safety
- Vitamin D Research and Controversies
- Lasers in Spine Surgery