Natural Language Processing: A Coding Professional's Perspective

by Gregory L. Schnitzer, RN, CCS-P, CPC, CPC-H

Despite understandable skepticism from the medical community, natural language processing (NLP) is slowly emerging as a viable rival to human coding professionals' speed and accuracy. NLP autocoding software teaches computers to "understand" English well enough to "read" physicians' dictated reports and notes and then code these encounters with no involvement from human coding professionals. Empirical studies comparing NLP coding software to coding professionals are still few and far between, although early results lend support to this promising new technology.[1] Certainly, some providers have already reduced their coding staffs, relying on NLP "autocoding" software.

As the manager of coding compliance for an NLP autocoding software company, I work closely with a team of coding professionals, computer programmers, and computational linguists to constantly improve the coding ability of our software. From the company's inception, our efforts have had one primary guiding principle: accurate, compliant coding. Although NLP could also ultimately be cost-effective, in our efforts to design NLP autocoding software, cost-effectiveness has always taken a back seat to accuracy. This article offers a brief overview of NLP from a coding professional's perspective that may help to explain why medical coding is one of the few places where NLP technology may enjoy a great deal of success.

"If It Wasn't Documented, It Wasn't Done"

Communication isn't just about language; much of human language often involves understanding what isn't spoken. Knowledge about the world around us and about our fellow human beings enables us to make inferences and assumptions so that what we understand is much more than what is actually said. However, computers don't possess that kind of background knowledge, and giving it to them is notoriously difficult. Normally, the inability to infer or assume might be an obstacle, but from a strict coding compliance standpoint, this is an advantage. When it comes to coding or reviewing a medical chart, inferring or assuming content that is not actually present is frowned on.

So, "if it wasn't documented, it wasn't done" is more than a compliance slogan. If a service isn't appropriately documented, there's no way for NLP to find it, assume it, or infer it. Consequently, NLP has the potential of being far more compliant than human coding professionals ever could. That's what makes medical coding and NLP such a beautiful marriage of a problem and a solution: even if a human coding professional could infer it, if it's not in the chart explicitly, it can't be considered in coding or auditing a medical chart. This requirement that information be explicitly stated is the conservative coding approach that compliance auditors are seeking to enforce these days.

Consistently Right and Wrong

One of the advantages of NLP is its ability to recognize essentially the same content expressed in different ways. Physicians have their own charting and dictation preferences. For example, "gadolinium was used," "T1 sagittal scan with Gd," "views obtained using paramagnetic contrast material," and "gadolinium-enhanced images" all convey substantially the same idea. And although resources like the Unified Medical Language System (UMLS) and other nosological efforts to standardize medical nomenclature can provide a tremendous amount of information on linguistic variation, no single resource could possibly handle the wide range of expression found in real medical charts from real providers.

Fortunately, NLP software works by "reading" a document and finding certain key concepts (not merely words) by identifying the many different phrasings of the concept. By reducing (or "normalizing") these concepts, different phrasings of the same concept can all be compared with one another for statistical purposes. By employing statistical algorithms, NLP can then compare and code these similar expressions. Plus, the computer will code the same expression the same way every time. It

won't forget some arcane coding rule once in a while, it won't have a "bad day," and it won't rush through a chart because it has to meet a coding quota. Unlike most human coding professionals, for any given coding situation, it will code the same way every time.

With NLP autocoding software, even coding mistakes are consistent-so any pattern of error will be consistent and easy to detect. These errors are easy to identify and correct and the coding software can then be improved so that those particular mistakes will never be repeated. Conversely, human coders don't always make mistakes consistently or predictably, making them difficult to spot and correct. As a result, it may actually be easier to audit, monitor, and "reeducate" NLP autocoding software than to do the same for human coding professionals.

Combining Accuracy with Speed

Coding has become increasingly complicated over the years. The amount of coding rules and guidelines, plus third-party payer preferences with which coding professionals must be familiar, is enormous. Clearly, the data storage and computational power of a computer are now required to store these numerous coding rules; it's also becoming apparent that the computational power of NLP software may actually be required to apply these coding rules accurately and efficiently.

Early studies show that NLP autocoding rivals the accuracy of coding professionals for some medical domains. However, even in those medical domains where NLP autocoding software isn't as accurate as coding professionals, the software still has the advantage of its sheer computational speed. NLP software, combined with a computer's ability to make thousands of computational decisions in a split second, is a very valuable, indeed formidable, coding tool.

A Language of Numbers

Although using NLP to code medical records is fairly recent, NLP itself is not new. Natural language processing has been around since the introduction of computers, and commercial-quality software for machine translation, speech recognition, and other tasks has been honed by decades of research. Although this is not a perfect analogy, NLP autocoding software is akin to translation software that automatically translates English documents into Russian documents, or French into Italian, and so forth. Obviously, reducing any language-especially medical English-into a collection of concepts and phrase-mapping rules is no small matter. And having to do the same thing for the language being translated into, known as the "target language," is just as difficult. However, the advantages of using NLP to code medical charts are twofold: first, the task being performed is not strictly "translation," but rather, the somewhat simpler task of "classification," and second, the target language of CPT-4 codes, HCPCS codes, and ICD-9-CM codes is merely numbers.

With medical coding, NLP, in a broad sense, "translates" medical English into the rather simple and straightforward "language" of CPT-4, ICD-9-CM and HCPCS codes. They are straightforward because the very point of classification and coding systems is to bring structure to unstructured data. Collectively, between CPT-4, HCPCS, and ICD-9-CM, there are approximately 26,000 "words" or codes, which constitutes a fairly small vocabulary. The small size, well-defined quality, and lack of grammatical combinations of the "words" in this target language make the "translation" (or, rather, "classification") job of NLP autocoding software easier and more accurate.

Getting Smarter Each Time

Advances in NLP in the last decade involve a statistical, data-driven, "machine learning" approach rather than a strict adherence to more traditional rules-based approaches to dealing with language. Using statistics allows NLP autocoding software to "predict" how the best coding professionals would code a chart. It's hard for a rigid set of rules to work given the subtlety, variability, and sheer "noise" found in large samples of a real language. In contrast, statistical methods work well when dealing with large quantities of noisy, variable data-which is precisely the kind of problem faced in medical coding. So, rather than simply coding a chart based on some rigid computer programming rule, making use of numerous statistics allows the computer to emulate proper human coding and improve its coding ability through machine learning techniques.

The upshot is that as NLP coding software codes, its body of data grows, thus improving the software's accuracy. Like the best human coding professionals, the statistical-based, data-driven NLP autocoding software improves with experience. Moreover, the use of statistics not only permits the NLP autocoding software to reach its coding decisions with confidence, but also permits it to know when it's uncertain. Finally, machine learning approaches offer statistical-based NLP software

programs the ability to constantly improve themselves over time in an automatic way, rather than through an ongoing process of "tweaking" a large set of coding rules.

Currently, NLP autocoding software is extracting CPT-4 codes and ICD-9-CM codes to code physician charts. However, as more patients' medical charts become available in an electronic format via a computerized patient record (CPR), it will eventually be possible for NLP software to code or "translate" hospital charts using ICD-9-CM procedure codes and simply link the NLP software's output to commercial DRG and/or APC grouper software. NLP autocoding software can then become a valuable tool for hospital coding professionals, easing their burden and allowing them to concentrate on more complex and challenging HIM tasks.

New Roles for HIM Professionals

NLP autocoding software is not developed in a vacuum, nor does it function in a vacuum. NLP autocoding software is developed, audited, and monitored by HIM professionals. Improvements are made to it by HIM professionals, and as new coding rules are developed or as we migrate to new code sets, it will be up to HIM professionals to retrain these NLP software programs. HIM professionals are needed to perform all the necessary compliance auditing and reviews-required under any compliance plan-and required whether the coding professional is human or software. And just as skilled nurses have had to learn to read EKG monitors and work ventilators, those HIM professionals who can embrace new technologies-NLP, computer-based records, and voice recognition software-will be at the leading edge of HIM. The HIM professional who excels at coding, who is knowledgeable on nosology and nomenclature issues, and who is well versed in all the major code sets will be best prepared and will always have a role to play as HIM moves into the 21st century.

Learn More About NLP

Books

Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing*. Upper Saddle River, NJ: Prentice-Hall, Inc., 2000.

Manning, Christopher D., and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: The MIT Press, 1999.

Articles

Aller, Kathleen C. "Information Systems for the Outcomes Movement." *Journal of Healthcare Information Management* 10, no. 1 (1996).

Beinborn, Julie. "Automated Coding: The Next Step?" *Journal of AHIMA* 70, no. 7 (1999): 38-42.

Johns, Merida. "A Crystal Ball for Coding." *Journal of AHIMA* 71, no. 1 (2000): 26-30.

Warner, Homer R., Jr. "Can Natural Language Processing Aid Outpatient Coders?" *Journal of AHIMA* 71, no. 8 (2000) 78-81.

Weber, Joe. "Improving Health Care Through Clinical Documentation." *Journal of Healthcare Information Management* 11, no. 2 (1997).

Professional Organizations

The Association for Computational Linguistics Web site: www.aclweb.org

The North American Chapter of the Association for Computational Linguistics Web site: www.cs.cornell.edu/info/people/cardie/naacl/

Note

1. Warner, Homer R., Jr. "Can Natural Language Processing Technology Aid Outpatient Coders?" *Journal of AHIMA* 71, no. 8 (2000): 78-81.

***Gregory L. Schnitzer**, is vice president of coding compliance at CodeRyte, Inc. in Bethesda, MD. He can be reached at GSchnitzer@coderyte.com.*

Article citation:
Schnitzer, Gregory L. "Natural Language Processing: a Coding Professional's Perspective." *Journal of AHIMA* 71, no.9 (2000): 95-98.
