Solution Brief

Healthcare
Natural Language Processing

John Snow Labs' Spark NLP for Healthcare Library Speeds Up Automated Language Processing with Intel® AI Technologies

Intel optimizations and 2nd Gen Intel Xeon® Scalable processors deliver up to 116 percent faster performance for the healthcare-specific Natural Language Processing library

INTEL® AI BUILDERS MEMBER

John Snow LABS

Advances and breakthroughs in medicine and public health are built on research and prior learnings. Understandings are contained in a wide range of content, such as the following:

- Patient records
- Imaging, genomic, and lab reports
- Medical billing records
- Research reports
- White papers and articles
- Clinical trial results
- Medical and healthcare regulatory filings

Petabytes of new information are added every year, which is searched, culled, and perused by researchers, analysts, and data scientists across the entire healthcare sector. They rely on automated systems that leverage artificial intelligence (AI) and Natural Language Processing (NLP) libraries to search for and analyze selected content to locate data they need.

Understanding Enhanced by Spark NLP

Spark NLP, developed by John Snow Labs, is an open source NLP library that provides state-of-the-art natural language understanding at scale with easy to use Python, Java, and Scala libraries. Built on top of Apache Spark, it is the only distributed and natively scalable NLP library today. It implements the latest breakthroughs in deep learning, transfer learning, and transformers, providing practitioners and enterprises with production-grade, scalable, and trainable versions of novel research for the first time (Figure 1).

Spark NLP was adopted by 16 percent of enterprises within 18 months of its first release.[1] It has remained the most widely used NLP library in the enterprise.[2] It continues to improve rapidly, with 30 new releases in 2019.

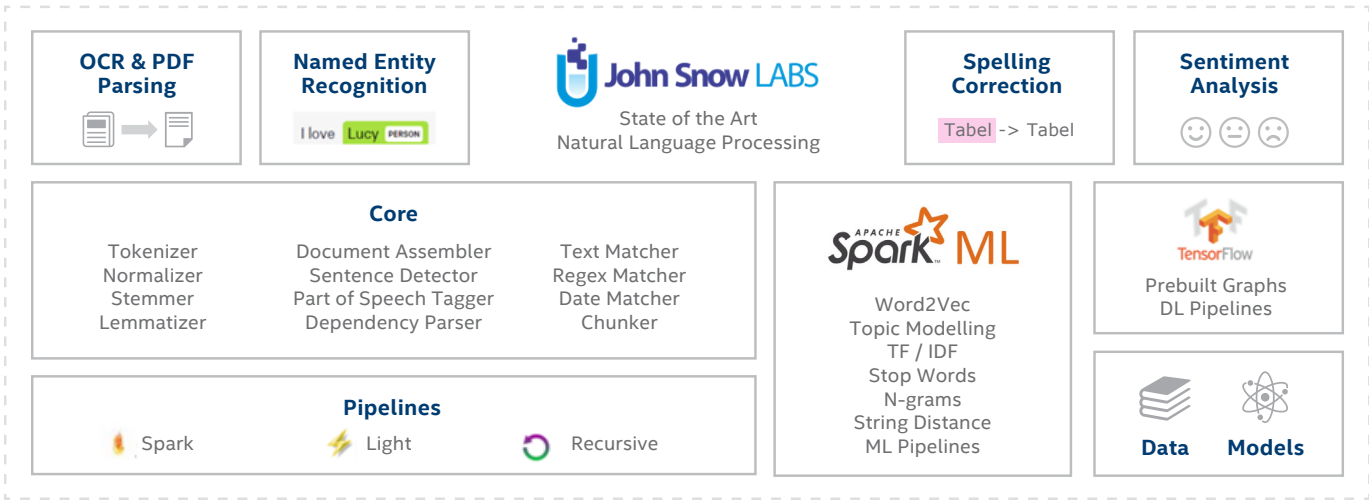

Figure 1. Spark NLP library components leveraging machine learning methods.

Spark NLP for Healthcare Enhances Clinical and Life Sciences Research

In addition to massive amounts of data, medicine and public health are filled with thousands of unique terms and identifiers. Many critical facts required by healthcare AI applications such as patient risk prediction, cohort selection, automated clinical coding, and clinical decision support, are locked in unstructured free-text data.

John Snow Labs offers a commercial version of their library for healthcare and life science data scientists, called Spark NLP for Healthcare. This version of the popular library provides a production-grade, scalable, and trainable implementation of novel healthcare-specific NLP algorithms and models. It includes pre-trained models for the most common medical NLP tasks.

Spark NLP for Healthcare extends the open-source library, raising the bar on achievable accuracy for tasks like clinical named entity recognition (NER), assertion status detection, entity resolution, de-identification, and Optical Character Recognition (OCR). The industry-specific library enables easy and automated access to information hidden across a broad range of documents—both physical and electronic.

Spark NLP Runs Faster for Lower Cost on Intel AI Technologies

Spark NLP is designed from the ground up to make the most from modern hardware that is optimized for deep learning and AI. The library uses Intel® Math Kernel Library (Intel MKL) and Intel Optimizations for TensorFlow to deliver unmatched performance from the latest Intel Xeon® Scalable processors.

Benchmarks were done on a 2 GB training dataset with ~108K sentences for a named entity recognition task in French. Testing demonstrated 116 percent improvement through use of Intel MKL libraries and Intel Xeon Scalable processors.[3] Additionally, when compared against previous generation Intel Xeon processors, the data was recorded as shown in Figure 2.

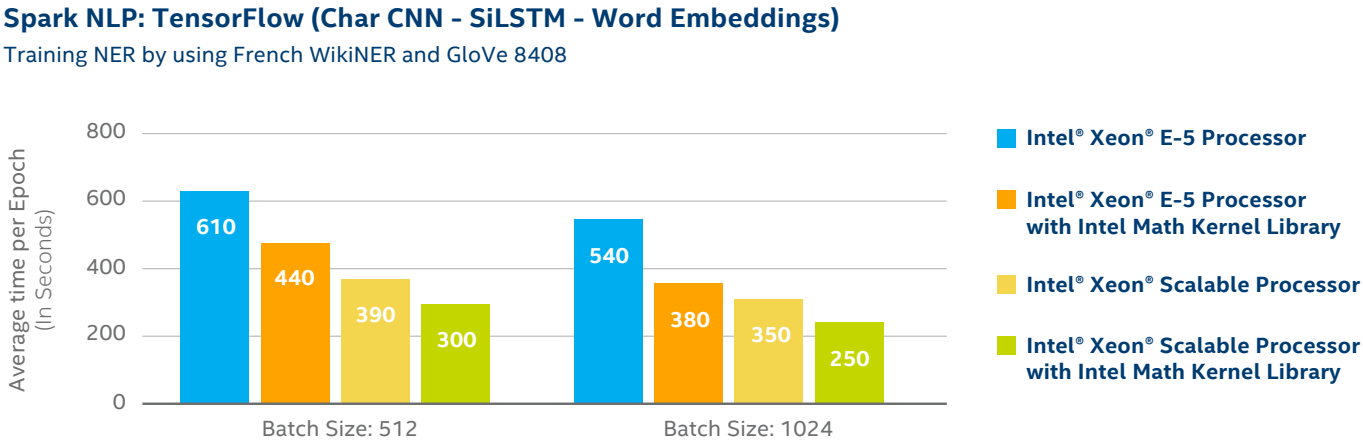

Figure 2. Intel technologies and processors deliver over 2X speedup of Spark NLP.

Summary

John Snow Labs' Spark NLP for Healthcare implements the latest breakthroughs in deep learning, transfer learning, and transformers, providing researchers and practitioners with production-grade, scalable, and trainable versions of novel NLP research for the first time. Designed to run on the latest processing technologies and with Intel AI optimizations and software, Spark NLP for Healthcare can deliver up to 116 percent faster learning on Intel Xeon Scalable processors with Intel MKL and Intel Optimizations for TensorFlow.

To learn more about John Snow Labs' clinical NLP software, visit Spark NLP for Healthcare.

Learn more about the Intel AI Builders program at https://builders.intel.com/ai/membership.

INTEL® AI BUILDERS MEMBER

John Snow Labs is an award-winning AI and NLP company, specializing in healthcare and life science and accelerating progress in data science. It is the developer of Spark NLP and provides state-of-the-art AI models, data, and software. John Snow Labs helps the world's largest healthcare and pharma companies get AI from concept to production.

---

1. Based on O'Reilly's "AI Adoption in the Enterprise" survey of 1,300 practitioners published in February 2019. https://www.oreilly.com/data/free/ai-adoption-in-the-enterprise.csp

2. Based on O'Reilly's "AI Adoption in the Enterprise" 2020 survey. https://www.oreilly.com/radar/ai-adoption-in-the-enterprise-2020/

3. Configuration Intel Xeon processor: Tested by Intel as of 09/13/2019. 2 socket Intel® Xeon® E5-2698 v4 @ 2.20 GHz, 20 cores per socket, Virtualization: VT-x Framework TensorFlow 1.12, Library Spark 2.4, Topology FP32 CharCNN-BiLSTM, NUMA node0 CPU(s): 0-15,32-47, NUMA node1 CPU(s): 16-31,48-63

Configuration Intel Xeon Scalable processor: Tested by Intel as of 09/13/2019. Intel® Xeon® Platinum 8175M CPU @ 2.50 GHz, 24 cores per socket, Memory type DDR4 2400 MHz, Framework TensorFlow 1.12, Library Spark 2.4, Topology FP32 CharCNN-BiLSTM, NUMA node0 CPU(s): 0-47

Performance results are based on testing as of dates shown in configuration and may not reflect all publicly available security updates. No product or component can be absolutely secure. See configuration disclosure for details.

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit: http://www.intel.com/performance

 0820/AU/HBD/PDF ♻ Please Recycle 343846-001US

intel®