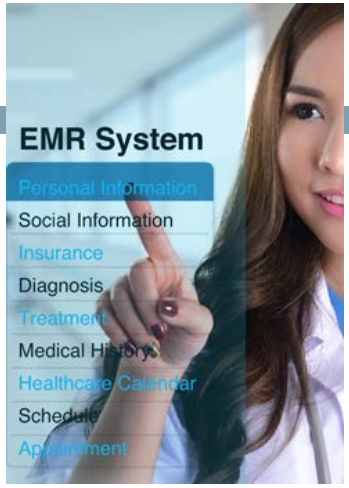

CASE STUDY

Improving Patient Health through Text Analysis of Electronic Medical Records

Developed by Megaputer Intelligence

Background

A prestigious teaching hospital in the U.S. is a global leader in biomedical research and one of the premier hospitals in the world. It has an annual research budget of over 500 million dollars. The hospital treats tens of thousands of patients every year and is committed to providing the highest quality patient care.

According to a recent Johns Hopkins study, medical errors and misdiagnoses are the 3rd leading cause of deaths in the U.S., with 10% of all deaths attributed to these errors. However, a multihospital study concluded that 83% of diagnosis errors can be prevented by leveraging the data contained in patient Electronic Medical Records (EMRs). In another study, a hospital discovered that 25% of the EMRs contained key disease hallmarks well before a patient diagnosis was established. In fact, these clinical warning signs were detected on average 19 days before breast cancer patients were diagnosed and 2 *years* before bladder cancer patients were diagnosed[1]. Early diagnostics has the potential to improve 5-year cancer patient survival rates by 90%. Correspondingly, the hospital decided to examine and improve its own EMR analysis process.

Challenge

EMRs represent a wealth of information for hospitals and physicians in improving patient care and public health at large. A large portion of this EMR data is composed of doctor and nurse notes as well as lab test results, making these EMRs a valuable resource for correct diagnoses. Thus, correct and early diagnosis of diseases relies on careful scrutiny of patient EMRs.

Typically, medical professionals are tasked with performing these manual reviews of EMRs. Often overworked, these medical professionals generally do not have enough time to scrutinize every EMR carefully. Hence, they may occasionally miss key signs and symptoms present in the EMRs, leading to an overall higher rate of misdiagnosis, delayed diagnosis, and litigation. Due to these factors, the hospital turned to automated text analysis for diagnostic decision auditing and support.

Automated text analysis solutions often struggle to accurately analyze textual data due to the complexity of medical text. EMR's are written in a unique language style, which is a combination of the vernacular, scientific terms, medical jargon, shorthand abbreviations, and medical coding. For example, a patient complaining of abdominal pain could be recorded as: Abdominal pain, bellyache, stomach ache, abd pain, RLQ [Right Lower Quadrant] pain, LLQ [Left Lower Quadrant] discomfort, or ICD10 Code-R10.10. The complex and evolving nature of medicine itself adds to the challenge of automated text analysis with thousands of potential symptoms, diagnoses, medications, and lab tests. Furthermore, an intelligent text analysis solution must be able to perform context-based

Megaputer
Your Knowledge Partner™

For more information:
Megaputer Intelligence | info@megaputer.com | 812-330-0110
www.megaputer.com

disambiguation of acronyms, such as for LBP, which can refer to “Lower Back Pain”, “Low Blood Pressure” or several other terms.

Solution

Megaputer used its proprietary data and text analytics software, PolyAnalyst™, to develop an automated solution for extraction of clinical findings from EMRs. Focusing on high risk patient diagnoses, Megaputer successfully implemented a system that extracts and structures the presence or absence of symptoms and findings that point to breast, colon, lung, and ovarian cancer in addition to lumbar disc disease, appendicitis, and myocardial infarction. The following were outputs of the system:

Context Based Disambiguation. PolyAnalyst performed automated context-based disambiguation of medical acronyms and diagnostic coding. This was achieved through machine learning of hand annotated medical records as well as millions of medical journal articles obtained from PubMed. The system builds a predictive model to determine the most likely expansion based on the words in context surrounding the acronyms. For example, LBP is more likely to be interpreted as “Lower Back Pain” if “back”,”pain”,”leg” or “numbness” is mentioned in the surrounding paragraph.

Interpretation of vital signs and lab tests. PolyAnalyst was easily trained to recognize dozens of lab tests utilized by the hospital. Using metrics provided by the hospital and statistical analysis, the system interprets the data and indicates the presence or absence of one of the diseases in focus.

Extraction of medication and dosages. The solution extracted thousands of medications from the hospital’s EMR data as well as included attributes such as the dosage and frequency of use. In order to accomplish this task, PolyAnalyst draws from RxNorm, a prescription drug database. Additionally, the system’s medical domain dictionary enables seamless brand-to-generic name mapping for the medications extracted from EMRs.

Extraction of key clinical findings. Using existing medical ontologies as a reference, such as the DxPlain Diagnoses ontology and the MeSH (Medical Subject Headings) dictionary, PolyAnalyst successfully extracted over 600 key symptoms associated with the diseases for diagnostic purposes. This was achieved through the system’s proprietary linguistic parsing language and classification of the medical terms, common vernacular forms, disambiguated abbreviations and interpretation of lab tests and vital signs.

Indication of true or negated findings. In addition to the extraction of the key clinical findings, PolyAnalyst also interpreted whether the finding was present or negated since the absence of a symptom can be important information for diagnosis. The system achieved this through deep linguistic parsing and analyzing complex negation patterns based on the syntactic relations of the words.

Results & Benefits

In order to verify the performance of Megaputer’s automated solution, the hospital conducted a comparison of the solution’s results against its own physician-generated manual annotations. As seen in Figure 1, Megaputer’s solution outperforms the doctor’s annotations in the percentage of symptoms extracted.

Megaputer
Your Knowledge Partner™

The benefits of this solution include:

- **Earlier, more accurate diagnoses**. The consistency and transparency of the solution system enhances diagnostic efficiency by quickly and accurately capturing indications within text notes that might otherwise go unnoticed by medical professionals, thereby reducing the number of misdiagnoses and delayed diagnoses. For diseases such as cancer, early detection increases the chances of patient survival significantly.
- **Financial Savings**. As a result of reduced medical errors, the solution also decreases the chances of malpractice litigation against the hospital that is engendered by misdiagnoses and delayed diagnoses.
- **Improved patient satisfaction and brand reputation.** Reduction of diagnosis errors, and thus, bad patient experiences and medical expenses, will have a positive impact on the overall customer satisfaction and brand reputation of the hospital.

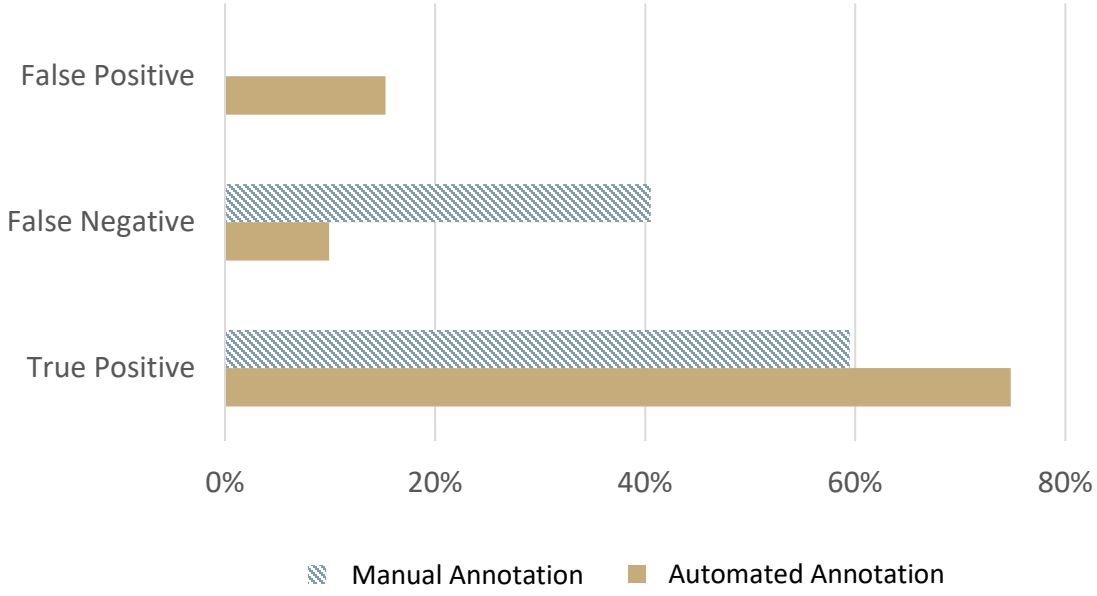

***Figure 1**. Percent of key clinical findings extracted*

References

[1]Feldman, M. J., Hoffer, E. P., Barnett, G. O., Kim, R. J., Famiglietti, K. T., & Chueh, H. (2012). Presence of key findings in the medical record prior to a documented high-risk diagnosis. Journal of the American Medical Informatics Association, 9 (4), 591–596. http://doi.org/10.1136/amiajnl-2011-000375

Megaputer
Your Knowledge Partner™