Name:	Lewis J Frey
Affiliation:	Medical University of South Carolina
Email:	frey@musc.edu
Presentation type (select one):	Poster

Clinical Personalized Pragmatic Predictions of Outcomes Discovered by Yotta-Scalable Secure Elastic Ubiquitous Search (C3PODYSSEUS)

Lewis J. Frey, PhD[1], Patrick Mauldin, PhD[2], Jihad Obeid, MD[1], William Moran, MD MS FACP[3], William Weintraub, MD, MACC, FAHA, FESC[4]
[1] Biomedical Informatics Center, Medical University of South Carolina (MUSC), Charleston, SC; [2] Pharmacy and Clinical Sciences, MUSC. Charleston, SC; [3] Division of General Internal Medicine and Geriatrics, MUSC, Charleston, SC; [4] Department of Cardiology, Center for Outcomes Research, Value Institute, Christiana Care Health System, Newark, DE

Abstract

We report on an elastically scalable system used for the prediction of outcomes involving emergency room utilization. The Observational Medical Outcomes Partnership common data model is used to support interoperable search across different institutions using a secure Accumulo/Hadoop platform. A risk stratification and clustering method is run on the platform in order to determine risk of readmission to the emergency department for patients at the Medical University of South Carolina. An open source ubiquitously deployable virtualized system has been deployed to Christiana Care in Delaware to replicate the risk stratification and clustering model developed at MUSC. Results obtained using the system will be presented along with a description of the system architecture.

Introduction

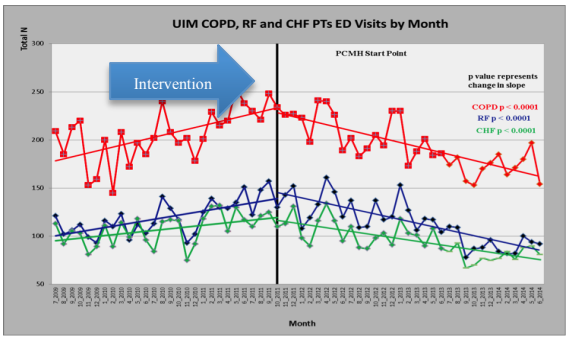
Big data by definition pose challenges to traditional storage and analysis approaches[1]. The challenges arise from data complexity, size and velocity and result in an inability of traditional methods to scale with the increasing complexity of big data. New methodologies have been developed to manage data complexity by parallelizing storage and analysis over large numbers of commodity computers through bringing the computation to the data instead of moving data to the computation. Clinical Personalized Pragmatic Predictions of Outcomes Discovered by Yotta-Scalable Secure Elastic Ubiquitous Search (C3PODYSSEUS) is such a system with its architecture being built on top of the National Security Agency's (NSA) open source system, Accumulo[2], presumably used at the NSA's reportedly yotta-scale data center in Utah[3]. Accumulo extends HBase to include a visibility field Boolean logic that can be specified to control which records are recalled in a secure fashion based on the access logic. The environment stores data from patient-physician encounters in an interoperable Observational Medical Outcomes Partnership/Observational Health Data Sciences and Informatics (OMOP/OHDSI) common data model[4].

Using a BigTable[5] Hadoop architecture consisting of a sparse, distributed, persistent multidimensional sorted map that increases performance by taking a large number of records and parallelizing their processing over many machines, the system can search for relevant features in OMOP over the Hadoop data store. The system can elastically scale up or down when system commodity hardware is added or removed to increase or decrease storage and computational power. C3PODYSSEUS (C3PO) was designed to have a secure, flexible, dynamic system that handles large amounts of data processed from the Veterans Affairs Informatics and Computing Infrastructure (VINCI), with teams at VINCI transforming their data model to OMOP. The opportunity for a virtualized ubiquitously deployable platform for interoperable search across different institutions other than the VA is currently

being investigated in a pilot grant. Through the Big Data Pilot Grant Program in Delaware we focus on the reduction of unnecessary acute care utilization for patients with risk factors of stroke and Cardiovascular Disease (CVD) at Medical University of South Carolina (MUSC) and at Christiana Care in Delaware.

Results

Our team has deployed C3PO at MUSC and performed risk stratification with clustering for the MUSC University Internal Medicine (UIM) quality improvement initiative to identify subgroups of complex patients who have similar co-morbidities and are at higher risk for hospitalization or ED visits. The sample included all UIM patients age ≥18 years who were seen at least once in UIM during January 1, 2008 through April 30, 2011. A total sample of 10,663 patients met eligibility criteria. Risk factor variables included age, gender, race, marital status, primary doctor type, insurance type, 35 diagnoses (ICD-9-CM code), visit compliance, place of residence, abnormal laboratory values /vital signs, and methodology as proposed in this application. Figure 1a shows that following the implementation of risk stratification and clustering based interventions at the beginning of October 2011, statistically significant decreases have occurred in the number of ED visits. From our risk-stratification / clustering analyses, we determined that certain disease clusters had a significant number of high risk patients, and we have demonstrated statistically significant decreases in ED visits and hospitalizations for our patients with COPD, Renal Failure (RF), Congestive Heart Failure (CHF) through alerting based on risk-stratification and clustering analysis. We tested the risk stratification and cluster translational application using data and analysis encoded in the C3PO system. Risk stratification and clustering analysis were conducted with traditional and the C3PO system using the original SAS risk and clustering compared to C3PO's R code implementation. The system was evaluated by comparing the two data representations (traditional versus C3PO) with the clusters having 94.7% overlap between Patient Centered Medical Homes (PCMH) and C3PO. The risk stratification has 96.2% overlap between the two methods. Figure 1b shows the similarity and difference between the two data representation in detail. C3PO is also deployed at Christiana Care in Delaware and is being loaded with a 40,000 patient cohort for the same time period as MUSC.



Cluster ID	Cluster Name	N of Pts		Cluster Inter-section	%		N of High Risk (Top 20%) Pts		Inter-section
		PCMH	C3PO		PCMH	C3PO	PCMH	C3PO	
1	CV Risk & Obesity	1865	1944	1800	17.9	18.7	33	27	27
2	Healthy	931	931	931	9	9	12	15	12
3	Hyperlipidemia Only	298	298	298	2.9	2.9	0	0	0
4	Multiple Chronic Conditions	4728	4723	4397	45.4	45.4	1110	1090	1053
5	Obesity Only	445	445	445	4.3	4.3	6	6	6
6	COPD/Asthma	1059	1073	1059	10.2	10.3	196	220	192
7	Renal Disease	1082	994	927	10.4	9.6	724	723	711
	SUM	10408	10408	9857	100	100	2081	2081	2001

Figure 1. (a) ED visits by month for COPD, RF and CHF. Vertical line indicates intervention based on risk stratification and clustering findings. (b) The clusters intersection (PCMH & C3PO)

Conclusion

The C3PO system has been successfully generalized outside of the VA to typical healthcare environments at MUSC and Christiana Care. The risk stratification and clustering model were similar for the traditional method compared with C3PO with risk stratification have 96.2% overlap and the clusters having 94.7% overlap. The work of replicating the analysis at Christiana Care in Delaware is currently ongoing. The research was supported in part by National Institutes of Health (NIH) grants 1R01GM108346-01 and U54-GM104941.

References

1. Frey LJ, Lenert L, Lopez-Campos G. EHR Big Data Deep Phenotyping. Yearb Med Inform 2014;9:206-211.
2. Sawyer SM, O'Gwynn BD. Evaluating Accumulo Performance for a Scalable Cyber Data Processing Pipeline. arXiv preprint arXiv 2014;1407.5661.
3. Bamford, J. The NSA is Building the Country's Biggest Spy Center (Watch What You Say). 2012. https://www.wired.com/2012/03/ff_nsadatacenter/
4. Stang PE, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Annals of internal medicine 2010;153.9:600-606.
5. Chang F, Dean J, et al. Bigtable: A Distributed Storage System for Structured Data. OSDI '06: 7th USENIX Symposium on Operating Systems Design and Implementation 2006;205-218, USENIX Association.