Convention ID: 1208
Submitting author: Dr Anna TONG
Post title: Health Informatician, NULL, NULL

Flu Activity Monitoring from Automatic Identification of Positive Laboratory Results with Free Text Laboratory Result Data

Y H Anna TONG (1), Lap Yi LAI (1), Ping Chong HO (1), Chris WONG (2); C Y Marita CHENG (2), Ivy CHENG (2),Wai Kit FUNG (2), Florence CHANG (2), Sam YEUNG (2)
(1) Health Informatics Section, Hospital Authority Head Office Information Technology and Health Informatics Division   (2) Information Technology Department, Authority Head Office Information Technology and Health Informatics Division

Keywords:
Machine Learning
Prediction
Infectious disease
outbreak

Introduction

Timely identification of potential outbreaks is important for prompt infectious disease outbreak prevention and control. Abnormal upsurge of laboratory isolations of alerting micro-organisms may indicate potential outbreaks of communicable diseases. Standard Reports to facilitate identification of potential abnormalities for infection management teams have been developed at CDARS with structured laboratory result data over these years. However, there are laboratory results are still in free text format because of clinical or laboratory workflow matters and this makes automatic aberration detection difficult.   Laborious and error prone manual works have been done to pick up those isolations with free text laboratory results of important micro-organisms.   With the recent IT advancement on big data analytics, a simple business case - automatic identification of positive laboratory results from free-text lab result data in HA database to start exploring the concept and application of Big Data Analytics, in particular the features of the machine learning and predictive analysis.

Objectives

(1) To explore the precision of automatic result identification by Proof of Concept (POC) of the machine learning and predictive analysis of big data analytics technology, i.e. reports are same as the previous flu activities monitoring report by automatic identification of positive results from structured flu lab result data & unstructured free text flu lab data   (2) To examine the value of adopting the new technology in the Hospital Authority for similar business cases

Methodology

Hadoop database has been set up for a small scale pilot study.   Sample dataset of HA CDR M&V test results (excluding C&ST), with test name and test result from Jan 2011 – Aug 2015 (2015 data for prediction) and rules of positive flu results have been

provided for the "supervised" machine learning.

Result
Flu Positive results identification (Prediction rate > 0.9) with sensitivity of 0.999970791. 68469 positive records identified in 2015 data versus 68471 positive records in the dataset retrieved by golden rule of CDARS standard report #46   Conclusion Machine learning and prediction of big data analytics technology have been shown to be helpful for the simple business case of automatic identification of positive laboratory results from free-text lab result data in HA database. Further exploration will be done for more complicated business cases with more data types & rules to further test out the big data analytics concept for HA use.